soup vary slightly from can to can; the time required to assign a seat at an airline check-in counter varies from passenger to passenger. To disregard the existence of variation (or to rationalize falsely that it is small) can lead to incorrect decisions on major problems. Statistics helps to analyze data properly and draw conclusions, taking into account the existence of variation.

Statistical variation—variation due to random causes—is much greater than most people think. Often, we decide what action to take based on the most recent data point, and we forget that the data point is part of a history of data.

In order to make decisions and improve processes, statistical variation must be taken into account. Variation can be visualized through the use of histograms, box plots, and similar tools. Frequently, such tools are sufficient to draw practical conclusions because differences in central tendency are large and variation is relatively small. However, statistical tools become necessary when the picture (quite literally) is less clear.

Building on the foundation of descriptive statistics, we start with an overview of the probability distributions that underlie many statistical tools and are used to model data and allow estimation of probabilities. Terms are defined as they are encountered, including further discussion of enumerative and analytical studies. Following an introduction to statistical inference and hypothesis testing, specific methods are discussed by way of example.

Probability Distributions

Before diving in, we should make a distinction between a sample and a population. A population is the totality of the phenomenon under study. A sample is a limited number of items taken from that population. Measurements are made on the smaller subset of items, and we can calculate a sample statistic (e.g., the mean). A sample statistic is a quantity computed from a sample to estimate a population parameter. Samples for statistics must be random. Simple random samples require that every element of the population have the same equal probability of selection for the sample. More complex sampling, such as stratified sampling, requires still requires that each element have a known, but not necessarily equal, chance of selection.

A probability distribution function is a mathematical formula that relates the values of the characteristic with their probability of occurrence in the population. The collection of these probabilities is called a probability distribution. The mean (μ) of a probability distribution often is called the expected value. Some distributions and their functions are summarized in Figure 19.11. Distributions are of two types:

Continuous (for "Variable" Data). When the characteristic being measured can take on any value (subject to the fineness of the measuring process), its probability distribution is called a "continuous probability distribution." For example, the probability distribution of the resistance data in Table 19.2 is an example of a continuous probability distribution because the resistance could have any value, limited only by the fineness of the measuring instrument. Most continuous characteristics follow one of several common probability distributions: the normal distribution, the exponential distribution, or the Weibull distribution.

Discrete (for "Attribute" Data). When the characteristic being measured can take on only certain specific values (e.g., integers 0, 1, 2, 3), its probability distribution is called a "discrete probability distribution." For example, the distribution of the number of defects r in a sample of five items is a discrete probability distribution because r can be only 0, 1, 2, 3, 4, or 5 (and not 1.25 or similar intermediate values). The common discrete distributions are the Poisson and binomial.

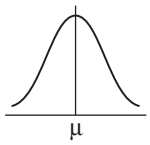
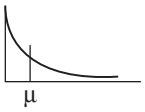
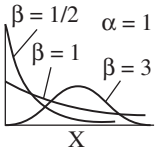
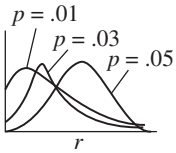
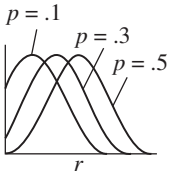
Distribution	Form	Probability function	
Normal		$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ μ = Mean σ = Standard deviation	Applicable when there is a concentration of observations about the average and it is equally likely that observations will occur above and below the average. Variation in observations is usually the result of many small causes.
Exponential		$y = \frac{1}{\mu} e^{-\frac{x}{\mu}}$	Applicable when it is likely that more observations will occur below the average than above.
Weibull		$y = \alpha\beta(X-\gamma)^{\beta-1} e^{-\alpha(X-\gamma)^{\alpha}}$ α = Scale parameter β = Shape parameter γ = Location parameter	Applicable in describing a wide variety of patterns in variation, including departures from the normal and exponential.
Poisson*		$y = \frac{(np)^r e^{-np}}{r!}$ n = Number of trials r = Number of occurrences p = Probability of occurrence	Same as binomial but particularly applicable when there are many opportunities for occurrence of an event but a low probability (less than .10) on each trial.
Binomial*		$y = \frac{n!}{r!(n-r)!} p^r q^{n-r}$ n = Number of trials r = Number of occurrences p = Probability of occurrence $q = 1 - p$	Applicable in defining the probability of r occurrences in n trials of an event that has constant probability of occurrence on each independent trial.

FIGURE 19.11 Summary of common probability distributions. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

3.37	3.34	3.38	3.32	3.33	3.28	3.34	3.31	3.33	3.34
3.29	3.36	3.30	3.31	3.33	3.34	3.34	3.36	3.39	3.38
3.35	3.36	3.30	3.32	3.33	3.35	3.35	3.34	3.32	3.38
3.32	3.37	3.34	3.38	3.36	3.37	3.36	3.31	3.33	3.30
3.35	3.33	3.38	3.37	3.44	3.32	3.36	3.32	3.29	3.35
3.38	3.39	3.34	3.32	3.30	3.39	3.36	3.40	3.32	3.33
3.29	3.41	3.27	3.36	3.41	3.37	3.36	3.37	3.33	3.66
3.31	3.33	3.35	3.34	3.35	3.34	3.31	3.36	3.37	3.35
3.40	3.35	3.37	3.35	3.32	3.36	3.38	3.35	3.31	3.34
3.35	3.36	3.39	3.31	3.31	3.30	3.35	3.33	3.35	3.31

(*Source: Quality Planning and Analysis,* Copyright 2007. Used by permission.)

TABLE 19.2 Resistance of 100 Coils, Ω

Statistical Inference

Statistical inference is the process of estimating, through sampling and application of statistical methods, certain characteristics of a population. In the world of quality, these estimates and statistical conclusions are used to draw practical conclusions, typically providing the practitioner confidence in taking subsequent action (or inaction) to improve a process.

Sampling Variation and Sampling Distributions

Suppose that a battery is to be evaluated to ensure that life requirements are met. A mean life of 30 hours is desired. Preliminary data indicate that the life follows a normal distribution and that the standard deviation is equal to 10 hours. A sample of four batteries is selected at random from the population and tested. If the mean of the four is close to 30 hours, it is concluded that the population of batteries meets the specification. Figure 19.12 plots the distribution of individual batteries from the population, assuming that the true mean of the population is exactly 30 hours.

If a sample of four is life-tested, the following lifetimes might result: 34, 28, 38, and 24, giving a mean of 31.0 hours. However, this random sample is selected from the many batteries made by the same process. Suppose that another sample of four is taken. The second sample of four is likely to be different from the first sample. Perhaps the results would be 40, 32, 18, and 29, giving a mean of 29.8 hours. If the process of drawing many samples (with four in each sample) is repeated over and over, different results would be obtained in most samples. The fact that samples drawn from the same process can yield different sample results illustrates the concept of sampling variation.

Returning to the problem of evaluating the battery, a dilemma exists. In the actual evaluation, let's assume only one sample of four can be drawn (e.g., because of time and cost limitations). Yet the experiment of drawing many samples indicates that samples vary. The question is, How reliable is the single sample of four that will be the basis of the decision? The final decision can be influenced by the "luck" of which sample is chosen. The key point is that the existence of sampling variation means that any one sample cannot always be relied upon to give an adequate decision. The statistical approach analyzes the results of the sample, taking into account the possible sampling variation that could occur.

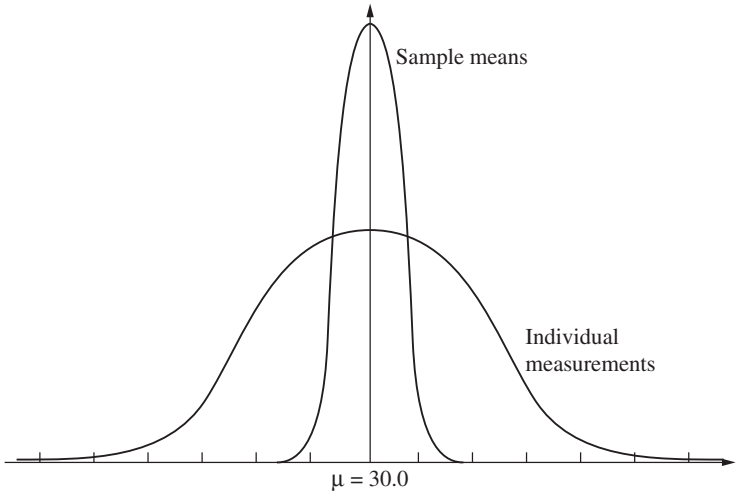


Figure 19.12 Distributions of individual measurements and sample means. (*Juran Institute, Inc., 1994.*)

Formulas have been developed to define the expected amount of sampling variation. In particular, the central limit theorem states that if $x_1, x_2, \ldots x_n$ are outcomes of a sample of n independent observations of a random variable x, then the mean of the samples of n will approximately follow a normal distribution, with mean μ and standard deviation $\sigma\bar{X} = \sigma\sqrt{n}$. When n is large ($n > 30$), the normal approximation is very close. For smaller samples, a modified Student-T distribution applies. The central limit theorem is very helpful to much practical statistical work. First, the variation of means is smaller than the variation of the underlying population, which makes conclusions easier. Second, because means are approximately normally distributed, we can apply the wide variety of techniques that rely on the assumption of normality.

Statistical Tools for Improvement

This concept of a sampling distribution is fundamental to the two major areas of statistical inference, estimation and tests of hypotheses, which are discussed next.

Statistical Estimation: Point Estimation and Confidence Intervals

Estimation is the process of analyzing a sample result to predict the corresponding value of the population parameter. In other words, the process is to estimate a desired population parameter by an appropriate measure calculated from the sample values. For example, the sample of four batteries previously mentioned had a mean life of 31.0 hours. If this is a representative sample from the process, what estimate can be made of the true average life of the entire population of batteries? The estimation statement has two parts:

1. The point estimate is a single value used to estimate the population parameter. For example, 31.0 hours is the point estimate of the average life of the population.
2. The confidence interval is a range of values that include (with a preassigned probability called a confidence level*) the true value of a population parameter. Confidence limits are the upper and lower boundaries of the confidence interval. Confidence limits should not be confused with other limits (e.g., control limits, statistical tolerance limits).

Table 19.3 summarizes confidence limit formulas for common parameters. The following example illustrates one of these formulas.

Problem Twenty-five specimens of brass have a mean hardness of 54.62 and an estimated standard deviation of 5.34. Determine the 95 percent confidence limits on the mean. The standard deviation of the population is unknown.

Solution Note that when the standard deviation is unknown and is estimated from the sample, the t distribution in Table 19.4 must be used. The t value for 95 percent confidence is found by entering the table at 0.975 and 25 – 1, or 24, degrees of freedom† and reading a t value of 2.064.

*A confidence level is the probability that an assertion about the value of a population parameter is correct. Confidence levels of 90, 95, or 99 percent are usually used in practice.

†A mathematical derivation of degrees of freedom is beyond the scope of this book, but the underlying concept can be stated. Degrees of freedom (DF) is the parameter involved when, for example, a sample standard deviation is used to estimate the true standard deviation of a universe. DF equals the number of measurements in the sample minus some number of constraints estimated from the data to compute the standard deviation. In this example, it was necessary to estimate only one constant (the population mean) to compute the standard deviation. Therefore, DF = 25 – 1 = 24.

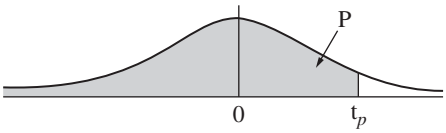
Mean of a normal population (standard deviation known)	$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ where $\bar{X}$ = sample average Z = normal distribution coefficient σ = standard deviation of population n = sample size
Mean of a normal population (standard deviation unknown)	$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ where t = distribution coefficient (with $n-1$ degrees of freedom) s = estimated σ (s is the sample standard deviation)
Standard deviation of a normal population	Upper confidence limit $= s\sqrt{\frac{n-1}{x^2_{\alpha/2}}}$ Lower confidence limit $= s\sqrt{\frac{n-1}{x^2_{1-\alpha/2}}}$ where x^2 = chi-square distribution coefficient with $n-1$ degrees of freedom $1-\alpha$ = confidence level
Population fraction defective	See charts: *Ninety-five percent confidence belts for population proportion* and *Binomial Distribution* at the end of this chapter, pages 670-672.
Difference between the means of two normal populations (standard deviations σ_1 and σ_2 known)	$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Difference between the means of two normal populations ($\sigma_1 = \sigma_2$ but unknown)	$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $\times\sqrt{\frac{\Sigma(X-\bar{X}_1)^2 + \Sigma(X-\bar{X}_2)^2}{n_1+n_2-2}}$
Mean time between failures based on an exponential population of time between failures	Upper confidence limit $= \frac{2rm}{x^2_{\alpha/2}}$ Lower confidence limit $= \frac{2rm}{x^2_{1-\alpha/2}}$ where r = number of occurrences in the sample (i.e., number of failures) m = sample mean time between failures DF = $2r$

(*Source: Quality Planning and Analysis,* Copyright 2007. Used by permission.)

TABLE 19.3 Summary of Confidence Limit Formulas ($1-\alpha$) (Confidence Level

Distribution of *t*

Value of *t* corresponding to certain selected probabilities (i.e., tail areas under the curve). To illustrate: the probability is .975 that a sample with 20 degrees of freedom would have $t = +2.086$ or smaller.



DF	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819

TABLE 19.4 Distribution of *t*

23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

(*Source: Introduction to Statistical Analysis*, Copyright 1969, Used by permission.)

TABLE 19.4 (*Continued*)

$$\text{Confidence limits} = \bar{X} = \pm t\frac{s}{\sqrt{n}}$$

$$= 54.62 \pm (2.064)\frac{5.34}{\sqrt{25}}$$

$$= 52.42 \text{ and } 56.82$$

There is 95 percent confidence that the true mean hardness of the brass is between 52.42 and 56.82.

Determination of Sample Size

The only way to obtain the true value of a population parameter such as the mean is to measure (with a perfect measurement system) each and every individual within the population. This is not realistic (and is unnecessary when statistics are properly applied), so samples are taken instead. But how large a sample should be taken? The answer depends on (1) the sampling risks desired (alpha and beta risk, discussed further below and defined in Table 19.5), (2) the size of the smallest true difference that is desired to be detected, and (3) the variation in the characteristic being measured.

For example, suppose it was important to detect that the mean life of the battery cited previously was 35.0 hours (recall that the intended value is 30.0 hours). Specifically, we want to be 80 percent certain of detecting this difference (this is the "power" of the test, and has a corresponding risk of $\beta = 0.2$; this means we are willing to take a 20 percent chance of failing to detect the five-hour difference when, in fact, it exists). Further, if the true mean was

Null hypothesis (H_0): Statement of no change or no difference. This statement is assumed true until sufficient evidence is presented to reject it.
Alternative hypothesis (H_a): Statement of change or difference. This statement is considered true if H_0 is rejected.
Type I error: The error in rejecting H_0 when it is true or in saying there is a difference when there is no difference.
Alpha risk: The maximum risk or maximum probability of making a type I error. This probability is preset, based on how much risk the researcher is willing to take in committing a type I error (rejecting H_0 wrongly), and it is usually established at 5% (or .05). If the *p*-value is less than alpha, reject H_0.
Significance level: The risk of committing a type I error.
Type II error: The error in failing to reject H_0 when it is false or in saying there is no difference when there really is a difference.
Beta risk: The risk or probability of making a type II error or overlooking an effective treatment or solution to the problem.
Significant difference: The term used to describe the results of a statistical hypothesis test where a difference is too large to be reasonably attributed to chance.
p-value: The probability of obtaining different samples when there is really no difference in the population(s)—that is, the actual probability of committing a type I error. The *p*-value is the actual probability of incorrectly rejecting the null hypothesis (H_0) (i.e., the chance of rejecting the null when it is true). When the *p*-value is less than alpha, reject H_0. If the *p*-value is greater than alpha, fail to reject H_0.
Power: The ability of a statistical test to detect a real difference when there really is one, or the probability of being correct in rejecting H_0. Commonly used to determine if sample sizes are sufficient to detect a difference in treatments if one exists. Power = $(1 - \beta)$, or 1 minus the probability of making a type II error.

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.5 Hypothesis Testing Definitions

30.0 hours, we want to have only a 5 percent risk of wrongly concluding it is not 30.0 hours (a risk of $\alpha = 0.05$). Then, using the following formula:

$$n = \left[\frac{(Z_{\alpha/2} + Z_{\beta})_{\sigma}}{\mu - \mu_o}\right]^2$$

we plug in our values to obtain

$$n = \left[\frac{(1.96 + 0.84)10}{35 - 30}\right]^2 = 31.4$$

The required sample size is 32 (Gryna et al., 2007, p. 605).

Note that sample size sometimes is constrained by cost or time limitations; in addition, rules of thumb exist to estimate sample size. However, these potentially lead to

gross under- or oversampling, with wasted time and effort. The recommended approach is to use power and sample size calculators (available online and in statistical software; these readily apply formulas appropriate for different sampling situations) in order to enter data collection and hypothesis testing with full knowledge of the statistically appropriate sample size.

Hypothesis Testing

A hypothesis, as used here, is an assertion about a population. Typically, the hypothesis is stated as a pair of hypotheses as follows: the null hypothesis (H_0) and an alternative hypothesis, H_a. The null hypothesis, H_0, is a statement of no change or no difference—hence, the term "null." The alternative hypothesis is the statement of change or difference—that is, if we reject the null hypothesis, the alternative is true by default.

For example, to test the hypothesis that the mean life of a population of batteries equals 30 hours, we state:

$$H_0: \mu = 30.0 \text{ hours}$$

$$H_a: \mu \neq 30.0 \text{ hours}$$

A hypothesis test is a test of the validity of the assertion, and is carried out by analyzing a sample of data. Sample results must be carefully evaluated for two reasons. First, there are many other samples that, by chance alone, could be drawn from the population. Second, the numerical results in the sample actually selected can easily be compatible with several different hypotheses. These points are handled by recognizing the two types of sampling errors, already alluded to above.

The Two Types of Sampling Errors. In evaluating a hypothesis, two errors can be made

- Reject the null hypothesis when it is true. This is called a type I error, or the level of significance. The maximum probability of a type I error is denoted by α.
- Fail to reject the null hypothesis when it is false. This is called type II error, and the probability is denoted by β.

These errors are defined in terms of probability numbers and can be controlled to desired values. The results possible in testing a hypothesis are summarized in Table 19.6. Definitions are found in Table 19.5. For additional detail on sampling errors in the context of quality, see Gryna at al (2007).

	Suppose the H_0 Is	
Suppose Decision of Analysis Is	**True**	**False**
Fail to reject H_0	Correct decision $p = 1 - \alpha$	Wrong decision $p = \beta$
Reject H_0	Wrong decision $p = \alpha$	Correct decision $p = 1 - \beta$

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.6 Type I (α) Error and Type II (β) Error

Steps to Hypothesis Testing. As emphasized earlier, it is important to plan for data collection and analysis; an investigator ideally should arrive at the point of actual hypothesis testing with elements such as sample size already defined. Hypothesis testing often is an iterative process, however, and as mentioned above in the opening discussion of data collection, further data may be needed after initial collection, for example, to bolster sample sizes to obtain the desired power so that both type I and type II errors are defined in advance.

Generally, then, the steps to test a hypothesis are as follows:

1. State the practical problem.
2. State the null hypothesis and alternative hypothesis.
3. Choose a value for α (alpha). Common values are 0.01, 0.05, and 0.10.
4. Choose the test statistic for testing the hypothesis.
5. Determine the rejection region for the test (i.e., the range of values of the test statistic that results in a decision to reject the null hypothesis).
6. Obtain a sample of observations, compute the test statistic, and compare the value to the rejection region to decide whether to reject or fail to reject the hypothesis.
7. Draw the practical conclusion.

Common Tests of Hypotheses. No single means of organizing hypothesis tests can convey all the information that may be of interest to an investigator. Table 19.7 summarizes some common tests of hypotheses in terms of the formulas. Table 19.8 categorizes tests according to the question being asked and type of data. Figure 19.13 provides similar information but in the form of a roadmap to assist in deciding what hypothesis test(s) are appropriate. Readers may find that the combination of these presentations will provide the best understanding of what is a multifaceted topic.

The hypothesis testing procedure is illustrated through the following example.

1. State the practical problem. To investigate a problem with warping wood panels, it was proposed that warping was caused by differing moisture content in the layers of the laminated product before drying. The sample data shown in Table 19.9 were taken between layers 1-2 and 2-3. Is there a significant difference in the moisture content?
2. State the null hypothesis and alternative hypothesis:

$$H_o: \mu 1\text{-}2 = \mu 2\text{-}3$$

$$H_a: \mu 1\text{-}2 \neq \mu 2\text{-}3$$

3. Choose a value for α. In this example, a type I error (α) of 0.05 will be assumed.
4. Choose the test statistic for testing the hypothesis.

Because we have two samples and desire to test for a difference in the means, a two-sample t-test is appropriate. (Note: A probability plot or test for normality will confirm the assumption of normality in the data. Also, an equal variance test concludes variances are approximately equal.)

1. Determine the rejection region for the test.

Hypothesis	Test Statistic and Distribution
H_o: $\mu = \mu_0$ (the mean of a normal population is equal to a specified value μ_0; σ is known)	$Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ Standard normal distribution
H_o: $\mu = \mu_0$ (the mean of a normal population is equal to a specified value μ_0; σ is estimated by s)	$t = \dfrac{\bar{X} - \mu_0}{s / \sqrt{n}}$ t distribution with $n - 1$ degrees of freedom (DF)
H_o: $\mu_1 = \mu_2$ (the mean of population 1 is equal to the mean of population 2; assume that $\sigma_1 = \sigma_2$ and that both populations are normal)	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{1/n_1 + 1/n_2}\sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right] / (n_1 + n_2 - 2)}}$ t distribution with DF = $n_1 + n_2 - 2$
H_o: $\sigma = \sigma_0$ (the standard deviation of a normal population is equal to a specified value σ_0)	$X^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$ Chi-square distribution with DF = $n - 1$
H_o: $\sigma_1 = \sigma_2$ (the standard deviation of population 1 is equal to the standard deviation of population 2; assume that both populations are normal)	$F = \dfrac{s_1^2}{s_2^2}$ F distribution with $DF_1 = n_1 - 1$ and $DF_2 = n_2 - 1$
H_o: $\hat{p} = p_0$ (the fraction defective in a population is equal to a specified value p_0; assume that $np_0 \geq 5$) $\hat{p}$ = sample proportion	$Z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$ Standard normal distribution
H_o: $p_1 = p_2$ (the fraction defective in population 1 is equal to the fraction defective in population 2; assume that n_1p_1 and n_2p_2 are each ≥ 5)	$Z = \dfrac{X_1/n_1 - X_2/n_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$ $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$ Standard normal distribution
To test for independence in a J × K contingency table that cross-classifies the variable A and B H_o: A is independent of B H_a: A is dependent on B	$X^2 = \sum_{j=1}^{J}\sum_{k=1}^{K} \dfrac{(f_{jk} - e_{jk})^2}{e_{jk}}$ Chi-square distribution with DF = (J − 1) (K − 1) where f_{jk} = the observed frequency of data for category j of variable A and to category k of variable B e_{jk} = the expected frequency = $f_{j0}f_{0k}/f_{00}$ f_{j0} = frequency total for category j for variable A f_{0k} = frequency total for category k of variable B f_{00} = frequency total for J × K table

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.7 Summary of Formulas on Tests of Hypotheses

Tests of hypotheses organized by the question being asked. All tests assume a categorical X in the Y= *f* (X) format. For example, X might be manufacturing plant, and there could be 1, 2 or more than two plants of interest in terms of output, Y. A continuous Y might be mean or standard deviation of daily units produced, a categorical Y might be proportion defective units produced in a single day.

Question: Is There a Difference in the Parameter	Number of Sample Groups	Continuous Y (Normal)		Categorical Y	
		Parameter of Interest	Test	Parameter of Interest	Test
Compared to a target?	1	μ σ	1-sample *t* Chi- square	Proportion	1-proportion test
between two groups?	2	μ σ	2-sample *t* F-test	Proportion	2-proportion test
among all groups?	≥2	μ σ	ANOVA* Bartlett's	Proportion	Chi-square test of Independence

*ANOVA assumes both equal variances and normality.
(*Source:* Juran Institute, Inc., Used by permission.)

Table 19.8 Hypothesis Testing Table

The critical value defining the rejection region is approximately 2.0 (see Table 19.4); if the absolute value of the calculated *t* is larger than the critical value, then we reject the null hypothesis.

1. Obtain a sample of observations, compute the test statistic, and compare the value to the rejection region to decide whether to reject or fail to reject the hypothesis.

A box plot (remember to plot the data!) suggests that the moisture content in Layer 1-2 tends to be higher than in Layer 2-3. Minitab output (see Figure 19.14) shows that the calculated *t* is 4.18, which is in the rejection region.

Because the calculated *t* is larger than the critical value, the associated *p*-value is < α, and we reject the null hypothesis, H_0.

N	Mean	StDev	SE Mean	
Layer 1-2	25	5.350	0.613	0.12
Layer 2-3	25	4.689	0.499	0.10

Difference = μ (Layer 1-2) – μ (Layer 2-3)

Estimate for difference: 0.660901

95 percent CI for difference: (0.343158, 0.978644)

T-test of difference = 0 (vs. not =): *t*-value = 4.18 *p*-value = 0.000 DF = 48

Both use pooled StDev = 0.5587

1. Draw the practical conclusion. We conclude that the moisture content in Layer 1-2 is higher than the moisture content of Layer 2-3.

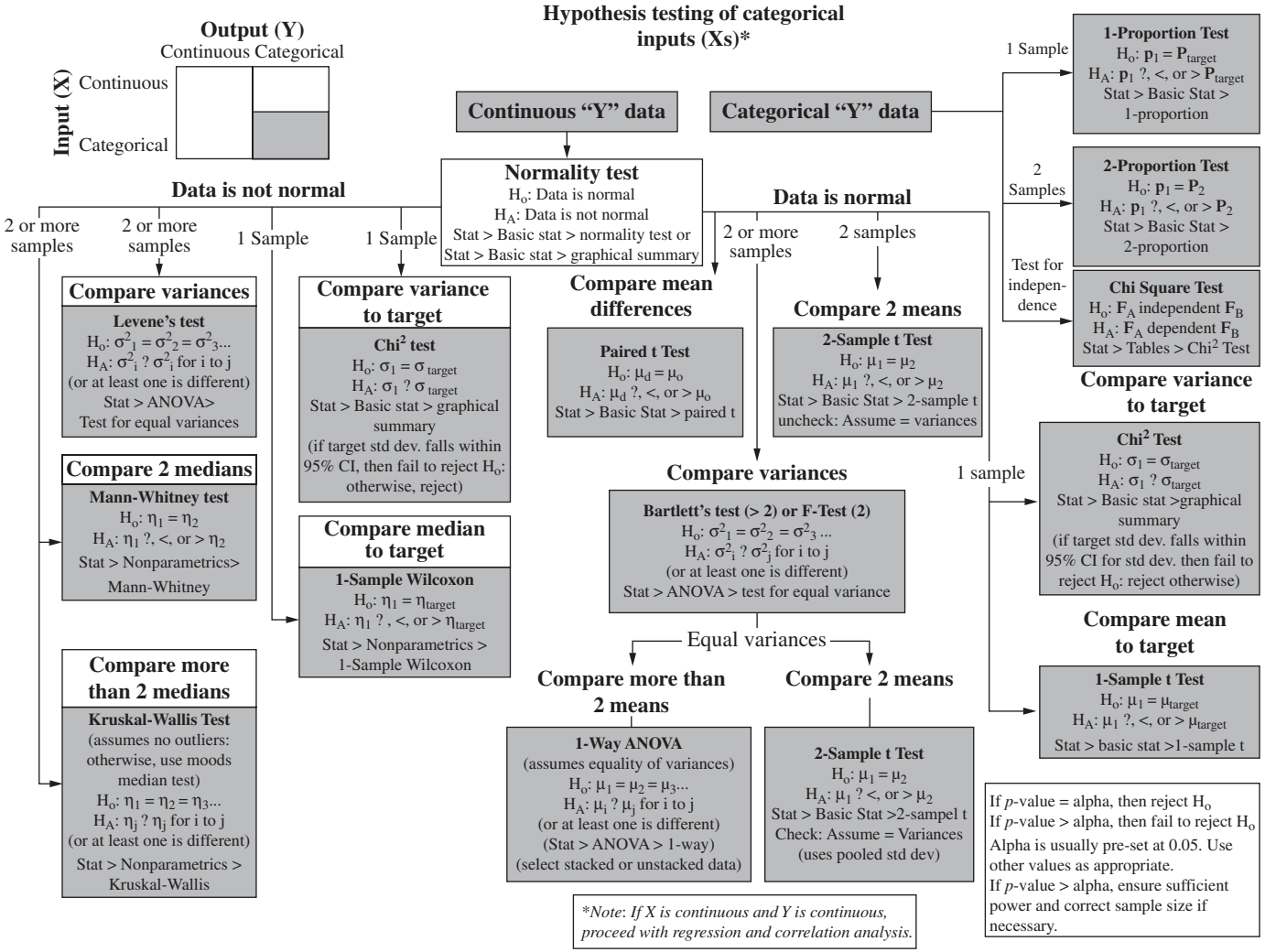


FIGURE 19.13 Hypothesis testing.

Layer 1-2		Layer 2-3	
4.43	4.40	3.74	5.14
6.01	5.99	4.30	5.19
5.87	5.72	5.27	4.16
4.64	5.25	4.94	5.18
3.50	5.83	4.89	4.78
5.24	5.44	4.34	5.42
5.34	6.15	5.30	4.05
5.99	5.14	4.55	3.92
5.75	5.72	5.17	4.07
5.48	5.00	5.09	4.54
5.64	5.01	4.74	4.23
5.15	5.42	4.96	5.07
5.64		4.21	

(*Source: Quality Planning and Analysis,* Copyright 2007. Used by permission.)

TABLE 19.9 Moisture Content

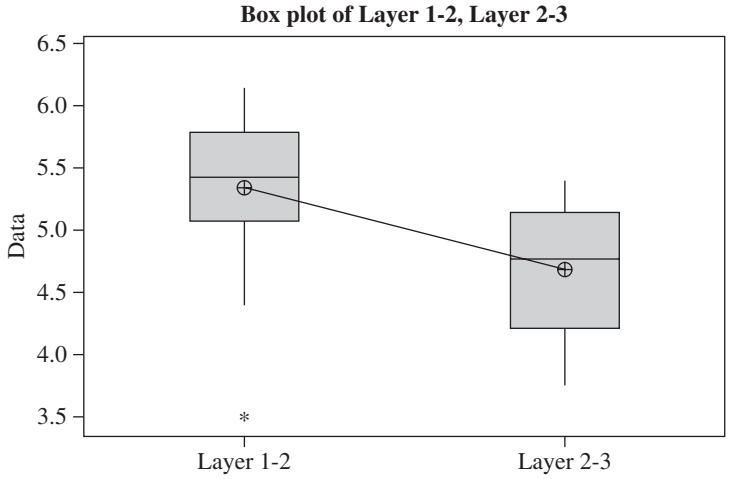


FIGURE 19.14 Box plot of Layer 1-2, Layer 2-3. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

Nonparametric Hypothesis Tests, Data Transformation, and Bootstrapping

The preceding discussion has focused on "parametric" hypothesis tests (so-called because they rely on parameter estimation). Often, it is the case that one or more of the assumptions underlying the parametric tests are violated. In particular, practitioners frequently face skewed or otherwise nonnormal data, and application of parametric tests that assume

bell-shaped data distribution may lead to erroneous conclusions and inappropriate action. Fortunately, options are available; these include nonparametric tests, data transformation, and bootstrapping.

Nonparametric hypothesis tests avoid violating key assumptions by virtue of being "distribution-free"; that is, they are not strictly dependent on particular distributions (such as a normal distribution); however, nonparametric tests have their own set of assumptions of which investigators should be aware). In effect, these methods typically transform the original data into ranks, and hypothesis tests then are carried out on the ranked data. Although nonparametric methods are not nearly as well developed and frequently are statistically less powerful compared to parametric tests, they are available for basic one-, two-, and two or more sample tests (see the bottom of Table 19.7 and the left side of the roadmap in Figure 19.13). See Sprent and Smeeton (2001) for more on traditional nonparametric methods. New methods continue to emerge, for example, wavelets and nonparametric Bayesian techniques; see Kvam and Vidakovic (2007).

Data transformation allows one to take data that violate some assumption of a parametric test and change them so that the assumption no longer is violated. For example, nonnormal data, or sample data with unequal variances can be changed to new numbers that are normal or have equal variances. Three common methods are

Power Functions. Traditionally, standard functions such as taking the square (x^2), square root ($x^{1/2}$), log (log10(x)), natural log (ln(x)), or inverse (x^{-1}) were used because they could easily be done with a calculator. Trial and error often is needed to find a function that appropriately transforms the data to meet the test assumptions.

Box-Cox Transformation. This method provides simultaneous testing of power functions to find an optimum value λ that minimizes the variance. Typically, one selects a power (value of λ) that is understandable and within a 95 percent confidence interval of the estimated λ (e.g., square: $\lambda = 2$; square root: $\lambda = 0.5$; natural log: $\lambda = 0$; inverse: $\lambda = -1$). The Box-Cox transformation does not work with negative numbers.

Johnson Transformation. This method selects an optimal function among three families of distributions (bounded, unbounded, lognormal). While effective in situations where Box-Cox does not work, the resulting transformation is not intuitive.

These methods are easy to apply (with software), and allow use of the more powerful parametric tests. However, the transformed data do not necessarily have intuitive meaning.

Bootstrapping is one of a broader class of computation-intensive resampling methods. Rather than assuming any particular distribution of a test statistic (such as normal), the distribution is determined empirically. More specifically, a statistic of interest (such as the mean) is repeatedly calculated from different samples drawn themselves, with replacements, from a sample. The distribution of these calculated statistics then is used as the basis for determining the probability of obtaining any particular value by chance. Itself a nonparametric approach, bootstrapping is a flexible method that gradually is gaining acceptance. For more information on the method and applications, see Davison and Hinkley (2006).

Correlation and Regression Analysis

Correlation and regression analysis help us understand relationships. More specifically, regression analysis is the modeling of the relationships between independent and dependent variables, while correlation analysis is a study of the strength of the linear relationships among variables. From a practical perspective, simple linear regression examines the distribution of one variable (the response, or dependent variable) as a function of one or more independent variables (the predictor, or independent variable) held at each of several levels.

X	Y	X	Y	X	Y	X	Y
90	41	100	22	105	21	110	15
90	43	100	35	105	13	110	11
90	35	100	29	105	18	110	6
90	32	100	18	105	20	110	10

(X, in feet per minute versus tool life; Y, in minutes)
(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.10 Cutting Speed

Note that the cause-and-effect relationship is stated explicitly, and it is this relationship that is tested to determine its statistical significance. In addition, regression analysis is used in forecasting and prediction based on the important independent variables, and in locating optimum operating conditions. In contrast, correlation typically looks at the joint variation of two variables that have not been manipulated by the experimenter, and there is no explicit cause-and-effect hypothesis.

For example, suppose that the life of a tool varies with the cutting speed of the tool and we want to predict life based on cutting speed. Thus, life is the dependent variable (Y) and cutting speed is the independent variable (X). Data are collected at four cutting speeds (Table 19.10).

Remembering to always plot the data, we note that a scatter plot (Figure 19.15) suggests that life varies with cutting speed (specifically, life decreases with an increase in speed) and also varies in a linear manner (i.e., increases in speed result in a certain decrease in life that is the same over the range of the data). Note that the relationship is not perfect—the points scatter about the line.

Often, it is valuable to obtain a regression equation. In this case, we have a linear relationship in the general form provided by

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

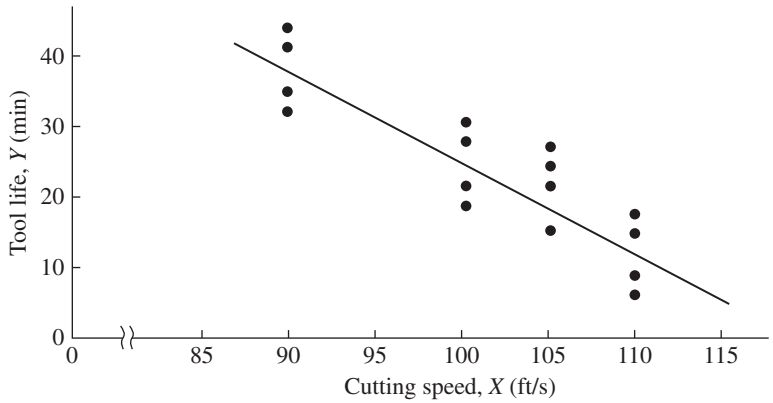


FIGURE 19.15 Tool life (Y) versus cutting speed (X). (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

where β_0 and β_1 are the unknown population intercept and slope, and ε is a random-error term that may be due to measurement errors and/or the effects of other independent variables. This model is estimated from sample data by the form

$$\hat{Y} = b_0 + b_1 X$$

where $\hat{Y}$ is the predicted value of Y for a given value of X and b0 and b1 are the sample estimates of β_0 and β_1. Estimates usually are found by least-squares methods; formulas can be found in statistics books such as Kutner et al. (2004).

For this example, the resulting prediction equation is

$$\text{Tool life} = 106.90 - 1.3614\ (\text{cutting speed})$$

This equation can be used to predict tool life by plugging in values of cutting speed. Extreme caution should be used in making predictions outside the actual sample space (e.g., for cutting speeds above or below the tested maximum or minimum), however, as these are tenuous without confirmation by observation.

Although a prediction equation can be found mathematically, it should not be used without knowing how "good" it is. A number of criteria exist for judging the adequacy of the prediction equation. One common measure is R_2, the proportion of variation explained by the prediction equation. R_2, or the coefficient of determination, is the ratio of the variation due to the regression to the total variation. The higher R_2, the greater the probable utility of the prediction equation in estimating Y based on X.

Another measure of the degree of association between two variables is the simple linear correlation coefficient, *r*. This is the square root of the coefficient of determination, so that the values of *r* range from –1 to +1. A positive *r* is consistent with a positive relationship (an increase in one variable is associated with an increase in the other), whereas the opposite is true of a negative *r* (an increase in one variable is associated with a decrease in the other). Scatter plots are strongly recommended when interpreting correlations, especially as very different patterns can result in identical values of *r*. The significance level of *r* varies with sample size; statistical software is recommended to obtain exact significance levels.

The above discussion introduces simple linear correlation and regression—the direction and strength of a relationship between two variables, or prediction of a dependent variable, Y, from a single predictor variable, X. A natural extension of this is multiple regression that allows for two or more independent variables. For a discussion of how to estimate and examine a multiple regression prediction equation, see Kutner et al. (2004).

Analysis of Variance

Analysis of Variance (ANOVA) is an approach related to linear regression, falling into the class of what are called general linear models. However, unlike regression, the X is discrete rather than continuous (noting that general linear models actually can blend characteristics of both regression and ANOVA). In ANOVA, the total variation of all measurements around the overall mean is divided into sources of variation that are then analyzed for statistical significance. It is used in situations where the investigator is interested in comparing the means among two or more discrete groups. For example, an investigator may be interested in comparing performance among three different machine configurations. The ANOVA analysis detects a difference somewhere among the means (i.e., at least one mean is different from the others), and confidence intervals or follow-up tests such as pairwise comparisons can be applied to determine which mean (or means) is different. ANOVA is the basis for design of experiments, discussed next.

Design of Experiments

With origins in the pioneering work in agriculture of Sir Ronald A. Fisher, designed experiments have taken on an increasingly significant role in quality improvement in the business world. This section will first compare the classical and designed approaches to experimentation, thereby providing the reader with an understanding as to the limitations of traditional methods and the power of contemporary methods. Next, basic concepts and terminology will be introduced in the context of an example improvement problem, followed by an overview of different types of designs and the typical progression through a series of designed experiments. The section finishes with the related topic of Taguchi designs.

Contrast between the Classical and Contemporary Methods of Experimentation. The classical method of experimentation is to vary one factor at a time (sometimes called OFAT), holding everything else constant. By way of example, and to illustrate the need for designed experiments, consider the case of a certain fellow who decided he wanted to investigate the causes of intoxication. As the story goes, he drank some whiskey and water on Monday and became highly inebriated. The next day, he repeated the experiment holding all variables constant except one… he decided to replace the whiskey with vodka. As you may guess, the result was drunkenness. On the third day, he repeated the experiment for the last time. On this trial, he used bourbon in lieu of the whiskey and vodka. This time it took him two days just to be able to gather enough of his faculties to analyze the experimental results. After recovering, he concluded that water causes intoxication. Why? Because it was the common variable!

The contrast between this traditional method and the designed approach is striking. In particular, a designed approach permits the greatest information to be gained from the fewest data points (efficient experimentation), and allows the estimation of interaction effects among factors. Table 19.11 compares these two approaches in more detail for an experiment in which there are two factors (or variables) whose effects on a characteristic are being investigated (the same conclusions hold for an experiment with more than two factors).

Concepts and Terminology—An Example Designed Experiment. Suppose that three detergents (A, B, C) are to be compared for their ability to clean clothes in an automatic washing machine. The "whiteness" readings obtained by a special measuring procedure are the dependent, or response, variable. The independent variable under investigation (detergent) is a factor, and each variation of the factor is called a level; in this case, there are three levels. A treatment is a single level assigned to a single factor, detergent A. A treatment combination is the set of levels for all factors in a given experimental run. A factor may be qualitative (different detergents) or quantitative (water temperature). Finally, some experiments have a fixed-effects model (i.e., the levels investigated represent all levels of concern to the investigator—for example, three specific washing machines or brands). Other experiments have a random effects model, that is, the levels chosen are just a sample from a larger population (e.g., three operators of washing machines). A mixed-effects model has both fixed and random factors.

Figure 19.16 outlines six possible designs of experiments, starting with the classical design in Figure 19.16a. Here, all factors except detergent are held constant. Thus, nine tests are run, three with each detergent with the washing time, make of machine, water temperature, and all other factors held constant. One drawback of this design is that the conclusions about detergent brands apply only to the specific conditions of the experiment.

Figure 19.16b recognizes a second factor at three levels (i.e., washing machines brands I, II, and III). However, in this design, it would not be known whether an observed difference was due to detergents or washing machine (they are said to be confounded).

Criteria	Classical	Modern
Basic procedure	Hold everything constant except the factor under investigation. Vary that factor and note the effect on the characteristic of concern. To investigate a second factor, conduct a separate experiment in the same manner.	Plan the experiment to evaluate both factors in one main experiment. Include in the design measurements to evaluate the effect of varying both factors simultaneously.
Experimental conditions	Care should be taken to have material, workers, and machine constant throughout the entire experiment.	Realizes difficulty of holding conditions reasonably constant throughout an entire experiment. Instead, experiment is divided into several groups or blocks of measurements. Within each block, conditions must be reasonably constant (except for deliberate variation to investigate a factor).
Experimental error	Recognized but not stated in quantitative terms.	Stated in quantitative terms.
Basis of evaluation	Effect due to a factor is evaluated with only a vague knowledge of the amount of experimental error.	Effect due to a factor is evaluated by comparing variation due to that factor with the quantitative measure of an experimental error.
Possible bias due to sequence of measurements	Often assumed that sequence has no effect.	Guarded against by randomization.
Effect of varying both factors simultaneously ("interaction")	Not adequately planned into experiment. Frequently assumed that the effect of varying factor 1 (when factor 2 is held constant at some value) would be the same for any value of factor 2.	Experiment can be planned to include an investigation for interaction between factors.
Validity of results	Misleading and erroneous if interaction exists and is not realized.	Even if interaction exists, a valid evaluation of the main factors can be made.
Number of measurements	For a given amount of useful and valid information, more measurements are needed than in the modern approach.	Fewer measurements needed for useful and valid information.
Definition of problem	Objective of experiment frequently not defined as necessary.	Designing the experiment requires defining the objective in detail (how large an effect do we want to determine, what numerical risks can be taken, etc.).
Application of conclusions	Sometimes disputed as applicable only to the controlled conditions under which the experiment was conducted.	Broad conditions can be planned in the experiment, thereby making conclusions applicable to a wider range of actual conditions.

(*Source: Quality Planning and Analysis,* Copyright 2007. Used by permission.)

TABLE 19.11 Comparison of Classical and Modern Methods of Experimentation

A	B	C
-	-	-
-	-	-
-	-	-

(a)

I	II	III
A	B	C
A	B	C
A	B	C

(b)

I	II	III
C	B	B
A	C	B
A	A	C

(c)

I	II	III
B	A	C
C	C	A
A	B	B

(d)

	I	II	III
1	C	A	B
2	B	C	A
3	A	B	C

(e)

I	II	III
ABC	ABC	ABC
1 ---	---	---
2 ---	---	---
3 ---	---	---

(f)

Figure 19.16 Some experimental designs. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

In Figure 19.16c, the nine tests are assigned completely at random, thus the name "completely randomized design." However, detergent A is not used with machine brand III, and detergent B is not used with machine brand I, thus complicating the conclusions.

Figure 19.16d shows a randomized block design. Here each block is a machine brand, and the detergents are run in random order within each block. This design guards against any possible bias due to the order in which the detergents are used and has advantages in the subsequent data analysis and conclusions. First, a test of hypothesis can be run to compare detergents and a separate test of hypothesis run to compare machines; all nine observations are used in both tests. Second, the conclusions concerning detergents apply for the three machines and vice versa, thus providing conclusions over a wider range of conditions.

Now suppose that another factor such as water temperature is also to be studied, using the Latin square design shown in Figure 19.16e. Note that this design requires using each

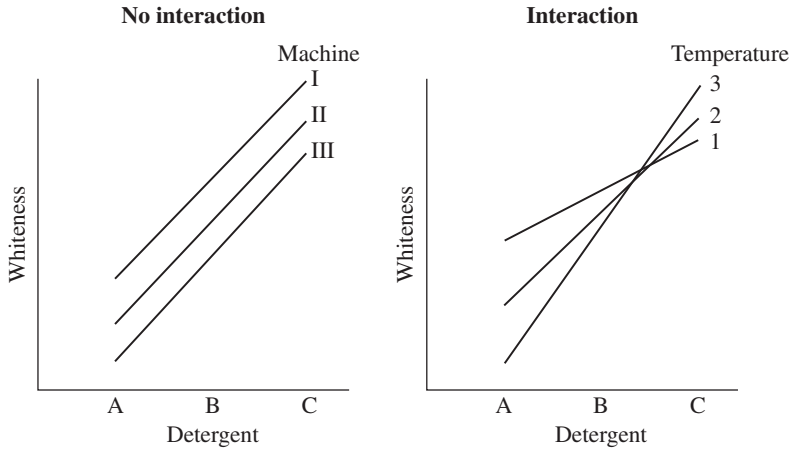


Figure 19.17 Interaction. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

detergent only once with each machine and only once with each temperature. Thus, three factors can be evaluated (by three separate tests of hypothesis) with only nine observations. However, there is a danger. This design assumes no interaction among the factors. No interaction between detergent and machine means that the effect of changing from detergent A to B to C does not depend on which machine is used, and similarly for the other combinations of factors. The concept of interaction is shown in Figure 19.17. There is no interaction among the detergents and the machines. But the detergents do interact with temperature. At high temperatures, C is the best performer. At low temperatures, A performs best.

Finally, the main factors and possible interactions could be investigated by the factorial design in Figure 19.16f. Factorial means that at least one test is run for every combination of main factors, in this case $3 \times 3 \times 3$ or 27 combinations. Separate tests of hypothesis can be run to evaluate the main factors and also possible interactions. Again, all the observations contribute to each comparison. When there are many factors, a portion of the complete factorial (i.e., a "fractional factorial") is useful when experimental resources are limited (see its application in a sequential testing approach, below).

Most problems can be handled with one of the standard experimental designs or a series of these. Designs can be classified by the number of factors to be investigated, the structure of the experimental design, and the kind of information the experiment is intended to provide (Table 19.12). For a description of both the design and analysis of various design structures, see Box et al. (2005). Another excellent general reference is Myers et al. (2009) for a detailed look at response surface designs.

A sequential approach to experimentation often can be helpful. Briefly, a typical sequence of designed experiments will allow an experimenter to quickly and efficiently narrow down a large number of possible factors (or X's in the Y = f(X) terminology of Lean Six Sigma) to find out which are most important, and then refine the relationships to find optimal settings for each of the vital few factors. The steps might be as follows:

1. *Screening experiment*. In this stage, a fractional factorial design may be applied that does not allow interactions to be detected, but can ferret out which of many factors have the greatest main effect.
2. *Fractional factorial design*. The smaller number of factors identified in the screening experiment are tested to allow detection of interaction effects.

Design	Type of Application
Completely randomized	Appropriate when only one experimental factor is being investigated
Factorial	Appropriate when several factors are being investigated at two or more levels and interaction of factors may be significant
Blocked factorial	Appropriate when number of runs required for factorial is too large to be carried out under homogeneous conditions
Fractional factorial	Appropriate when many factors and levels exist and running all combinations is impractical
Randomized block	Appropriate when one factor is being investigated and experimental material or environment can be divided into blocks or homogeneous groups
Balanced incomplete block	Appropriate when all the treatments cannot be accommodated in a block
Partially balanced incomplete block	Appropriate if a balanced incomplete block requires a larger number of blocks than is practical
Latin square	Appropriate when one primary factor is under investigation and results may be affected by two other experimental variables or by two sources of nonhomogeneity. It is assumed that no interactions exist.
Youden square	Same as Latin square, but number of rows, columns, and treatments need not be the same
Nested	Appropriate when objective is to study relative variability instead of mean effect of sources of variation (e.g., variance of tests on the same sample and variance of different samples)
Response surface	Objective is to provide empirical maps (contour diagrams) illustrating how factors under the experimenter's control influence the response
Mixture designs	Use when constraints are inherent (e.g., the sum of components in a paint must add to 100%)

(*Source*: Adapted from JQH5, Table 47.3.)

TABLE 19.12 Classification of Designs

3. *Full factorial design*. A small number of factors (usually no more than five) are tested to allow all main effects and higher-order (e.g., three-way, four-way) interactions to be detected and accounted for. Such designs also can detect curvature that indicates a potential optimum.
4. *Response surface design*. By adding data points in particular ways (e.g., a composite design), an experimenter can build on earlier experiments to fully characterize nonlinear relationships and pinpoint optimal settings.
5. *EVOP*. Once an improved process is in production mode, evolutionary operation techniques can be used to conduct many small experiments on production units over time. Although individual changes are small, the cumulative effect over time can be quite large, and exemplifies the power of continuous improvement. See Box and Draper (1969) for a classic text on this subject.

For a series of four papers on sequential experimentation, see Carter (1996). Emanuel and Palanisamy (2000) discuss sequential experimentation at two levels and a maximum of seven factors.

Taguchi Approach to Experimental Design

Professor Genichi Taguchi uses an approach to experimental design that has three purposes:

- Design products and processes that perform consistently on target and are relatively insensitive ("robust") to factors that are difficult to control.
- Design products that are relatively insensitive (robust) to component variation.
- Minimize variation around a target value.

Thus, although cited in this "improvement tools" section because of its association with DOE, the approach is meant to provide valuable information for product design and development (see "Statistical Tools for Designing for Quality" in this chapter). Taguchi divides quality control into online control (e.g., diagnosing and adjusting a process during production) and offline control that encompasses the engineering design process and its three phases: systems design, parameter design, and tolerance design. For an extensive bibliography and a summary of some controversial aspects of the Taguchi approach, see Box and Draper (1969, pp. 47.58 and 47.59).

Many books are available that cover DOE for engineering and manufacturing applications. For readers in nonmanufacturing environments, Ledolter and Swersey (2007) may be of interest. A recent text readers may find useful for not only classical but more contemporary techniques (e.g., Bayesian inference, kriging) is del Castillo (2007).

Discrete Event and Monte Carlo Simulation

Advances in user-friendly software make computer simulations increasingly accessible to quality practitioners that do not have a strong background in mathematics, programming, or modeling. Numerous types of simulation models exist, but two that may be of most interest to readers are discrete event and Monte Carlo simulations. These can be powerful methods for making process improvements; in particular, modeling provides a means of asking "what if?" questions and rapidly testing the effects of process changes and potential solutions in a safe, low-risk environment.

Discrete Event Simulation. Discrete event simulation (DES) attempts to mimic situations in which there are distinct, recognizable events and transactions. In a hospital, for example, arrival of patients at an emergency department and subsequent steps in patient care represent specific events that combine into a flow of transactions: arrival, registration, triage, nursing assessment, physician assessment, etc., through inpatient admission, discharge, or transfer. Discrete event simulation enables system components to be changed and tracks the resulting process flow over time to help understand the relationships among inputs, outputs, and process variables.

Typically, a process flow diagram (or process "map") that graphically displays the sequence and flow of activities forms the basis for a discrete event simulation. A discrete event simulation takes this basic flow diagram and adds inputs and process variables that govern the flow of transactions. Following on the hospital example, these include inputs (such as patient arrivals), human resources (e.g., number of nurses, physician schedules, overtime availability, skill levels, pay rates, etc.), equipment resources (e.g., types and number of beds, imaging equipment, etc.), rules for flow (the required sequence of steps, batching of inputs or outputs, priority rules, exceptions, decisions), resource acquisition (what resources are needed to complete an activity (e.g., one RN or one physician's assistant; two RNs; one RN and one physician, etc.), activity cycle times (work time, wait time), and similar details.

Once these details are built into the model, it "runs" by tracing the path of units (patients, in the hospital example) from arrival through to exit from the process. Patients are processed in accordance with the activities, rules, and constraints, and any relevant attributes (patient-specific characteristics) that may be assigned to them (e.g., acuity level, age, gender). The

output consists of a multitude of descriptive statistics and measures that portray the collective behavior of the process as the various players interact and move through time.

Although every model is different and details vary, there are basic steps that should be a part of every simulation study. These steps and related questions are (adapted from Law and Kelton 2000):

1. State the problem and question(s) being asked. What is the business need for the simulation? What problem is to be fixed? What answers are being sought?
2. Prepare a plan for the simulation study. Who needs to be involved? What data are needed and how will data be collected? What alternative scenarios are to be tested? What are the milestones and timeline for completion?
3. Collect data. What is my current state? What are the data for alternative scenarios? Are there gaps in the data, and how will they be handled?
4. Build and validate a conceptual model. Given available data, what is the general structure of the model? What will be the inputs, process variables, and outputs? What statistical accumulators are needed, and where? If the model is built, will it provide the answers to the questions?
5. Build and validate an operational model. Are the model components necessary and sufficient? Does the model produce results consistent with the current state?
6. Design scenarios or experiments needed to answer the questions. What model parameters will be changed? Which are fixed? What combinations of factors need to be tested?
7. Run the scenarios or experiments to obtain the needed outputs. Are the results reproducible? Are additional scenarios or experiments suggested?
8. Analyze and interpret the data. What are the statistical results? Do the descriptive statistics and/or statistical tests indicate meaningful effects? What are the answers to the original questions? Are additional questions raised?

As emphasized at the beginning of this chapter, formulation of the question(s) being asked is a critical first step to the successful application of simulation modeling. Failure to have a clear understanding of what the model is being asked to do leads to poorly constructed models, models with insufficient inputs or process detail, or overly complicated models that take unnecessary time and effort to build and run. In addition, a clearly communicated business need will garner the stakeholder support needed to collect data, evaluate the model, and implement suggested changes.

Monte Carlo. Named after the famed gambling destination, this method seeks to account for uncertainty (variability) in inputs and carry this forward into probability distributions of outcomes. Essentially, instead of using single, fixed values in equations [such as $Y = f(X)$], distributions are used for the inputs (X's), and samples repeatedly are drawn from the distributions, yielding a distribution of outputs (Y values) instead of a single value. For example, while the forecasted net return on a new product could simply be stated as an expected $10 million, it would be useful to know the probability of achieving this, or that the uncertainty in the forecast is such that there is a high probability of a negative return.

By way of illustration, assume we have three components, A, B, and C that are assembled end-to-end to create a final product. If the mean lengths are 5, 10, and 15 mm, then we can simply add these together to arrive at an expected mean combined total length of 5 mm + 10 mm + 15 mm = 30 mm. However, we know from the concept of statistical variation that there will be variation in the components. Assuming we sample populations of each component and find the respective distributions for each of A, B, and C, what can we expect the overall distribution of assembled product length to look like? By repeatedly taking a random sample from each distribution and adding the lengths, Monte Carlo simulation generates a distribution of the total length Figure 19.18 shows the relative frequency distribution of the combined

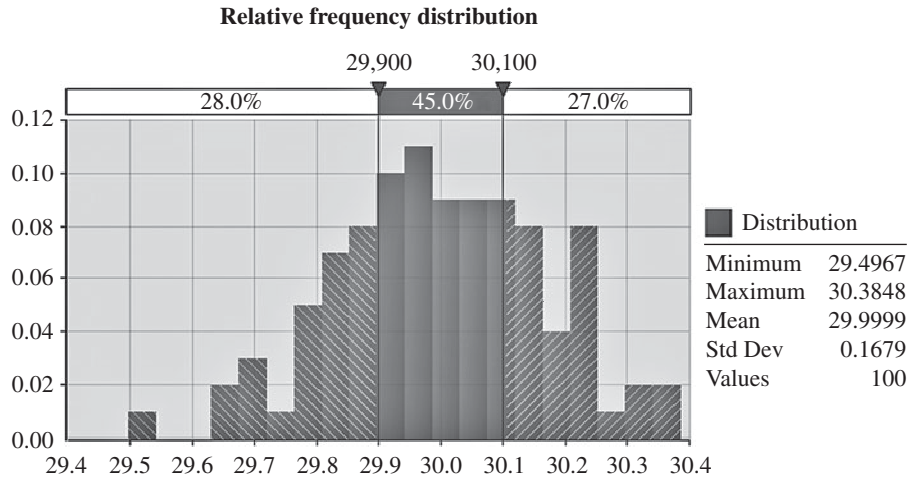


Figure 19.18 Result of Monte Carlo simulation showing a relative frequency distribution of combined total length of three components A, B and C that individually have normal distributions of 5, 10 and 15 mm, respectively, each with a standard deviation of 0.1 mm. The mean expected combined total length is approximately 30 mm, but the simulation shows the variation around this, e.g., that only 45% of assembled components are expected to be within +/− 0.1 mm of this mean value.

lengths of the three components from a Monte Carlo simulation with each of the three components having a standard deviation of of 0.1 mm. The mean expected combined total length is almost exactly 30 mm, but the simulation shows the variation around this, with only 45% of assembled components expected to be within +/− 0.1 mm of the total mean value. This approach provides substantially more information than the single estimate of 30 mm.

Simulated DOE. As tools evolve, they are being combined in new ways. One example is the combination of Monte Carlo, discrete event simulation, and DOE. Briefly, this approach involves a discrete event simulation (DES) that uses probability distributions for the input and/or process variables (Monte Carlo), and the investigator changes these variables (as factors) following a structured, designed approach (DOE). While any results and conclusions should be treated as preliminary until verified by actual experimentation, this can be particularly useful in environments where real-life changes may be difficult or dangerous to make.

Additional Advanced Analysis Tools

For practitioners faced with more complex scenarios such as multiple variables (more than one y and/or x), nonlinear data, or categorical outputs, extensions of the general linear models and other alternatives are available. In particular are methods for multivariate analysis; this refers to statistical techniques that simultaneously analyze multiple measurements on subjects. Many techniques are extensions of the univariate (single-variable distributions) and bivariate (correlation, regression) methods dealt with above. Beyond the scope of this chapter, these include:

- *Multiple regression*. Applies when the investigator has a single, continuous dependent variable and multiple, continuous independent variables (X's) of interest.
- *Nonlinear regression*. Useful when data cannot easily be treated by standard linear methods (note that curvilinear data do not necessarily require nonlinear methods).
- *Nonparametric linear regression*. Applies when the usual assumptions of regression are violated.

- *Multiple discriminant analysis*. Used in situations with a single, categorical (dichotomous or multichotomous) dependent variable (Y) and continuous independent variables (X's).
- *Logistic regression*. Also known as logit analysis, this is a combination of multiple regression and multiple discriminant analysis in which one or more categorical or continuous independent variables (X's) are used to predict a single, categorical dependent variable (Y). Odds ratios often are computed with this method.
- *Multivariate analysis of variance and covariance (MANOVA, MANCOVA)*. Dependence techniques that extend ANOVA to allow more than one continuous, dependent variable (Y) and several categorical independent variables (X's).
- *Principal component analysis (PCA) and common factor analysis*. These methods analyze interrelationships among a large number of variables and seek to condense the information into a smaller set of factors without loss of information.
- *Cluster analysis*. An interdependence technique that allows mutually exclusive subgroups to be identified based on similarities among the individuals. Unlike discriminant analysis, the groups are not predefined.
- *Canonical correlation analysis*. An extension of multiple regression that correlates simultaneously several continuous dependent variables (Y's) and several continuous independent variables (X's).
- *Conjoint analysis*. Often used in marketing analyses, this method helps assess the relative importance of both attributes and levels of complex entities (e.g., products). It is useful when trade-offs exist when making comparisons.
- *Multidimensional scaling*. An interdependence method (also called perceptual mapping), this seeks to transform preferences or judgments of similarity into a representation by distance in multidimensional space.
- *Correspondence analysis*. Another interdependence technique; this accommodates the perceptual mapping of objects (such as products) onto a set of categorical attributes. This method allows both categorical data and nonlinear relationships.

Readers are encouraged to research any techniques that appear to fit their need; although complex, these are powerful means of getting useful information from data. Some useful references include

Multivariate techniques:

Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., and Tatham, R. L. (2006). *Multivariate Data Analysis*. Pearson Prentice-Hall, Upper Saddle River, NJ.

Affifi, A., Clark, V. A., and May, S. (2004). *Computer-Aided Multivariate Analysis* (4th ed.). Chapman and Hall/CRC Press, Boca Raton, FL.

Coleman, S, Greenfield, T., Stewardson, D., and Montgomery, D. C. (2008). *Statistical Practice in Business and Industry*. John Wiley & Sons, Hoboken, NJ. (see Chapter 13).

Hypothesis testing and DOE:

Box, G. E. P., Hunter, J. S., and Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation and Discovery* (2nd ed.). Wiley-Interscience, Hoboken, NJ.

Logistic regression, Poisson regression, odds ratios:

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley & Sons, New York.

Nonparametric:

Sprent, P., and Smeeton, N. C. (2001). *Applied Nonparametric Statistical Methods* (3rd ed.). Chapman and Hall/CRC Press, Boca Raton, FL.

Statistical Tools for Designing for Quality

Statistical tools for quality in the design and development process include techniques such as graphical summaries, probability distributions, confidence limits, tests of hypotheses, design of experiments, regression, and correlation analysis. These topics are covered in earlier sections of this chapter. To supplement these techniques, this section explains some statistical tools for reliability and availability, and tools for setting specification limits on product characteristics.

Failure Patterns for Complex Products

Methodology for quantifying reliability was first developed for complex products. Suppose that a piece of equipment is placed on test, is run until it fails, and the failure time is recorded. The equipment is repaired and again placed on test, and the time of the next failure is recorded. The procedure is repeated to accumulate the data shown in Table 19.13. The failure rate is calculated, for equal time intervals, as the number of failures per unit of time. When the failure rate is plotted against time, the result (Figure 19.19) often follows a familiar pattern of failure known as the bathtub curve. Three periods are apparent that differ in the frequency of failure and in the failure causation pattern:

- *The infant mortality period*. This period is characterized by high failure rates that show up early in use (see the lower half of Figure 19.18). Commonly, these failures

Time of Failure, Infant Mortality Period		Time of Failure, Constant Failure Rate Period		Time of Failure, Wear-Out Period	
1.0	7.2	28.1	60.2	100.8	125.8
1.2	7.9	28.2	63.7	102.6	126.6
1.3	8.3	29.0	64.6	103.2	127.7
2.0	8.7	29.9	65.3	104.0	128.4
2.4	9.2	30.6	66.2	104.3	129.2
2.9	9.8	32.4	70.1	105.0	129.5
3.0	10.2	33.0	71.0	105.8	129.9
3.1	10.4	35.3	75.1	106.5	
3.3	11.9	36.1	75.6	110.7	
3.5	13.8	40.1	78.4	112.6	
3.8	14.4	42.8	79.2	113.5	
4.3	15.6	43.7	84.1	114.8	
4.6	16.2	44.5	86.0	115.1	
4.7	17.0	50.4	87.9	117.4	
4.8	17.5	51.2	88.4	118.3	
5.2	19.2	52.0	89.9	119.7	
5.4		53.3	90.8	120.6	
5.9		54.2	91.1	121.0	
6.4		55.6	91.5	122.9	
6.8		56.4	92.1	123.3	
6.9		58.3	97.9	124.5	

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.13 Failure History for a Unit

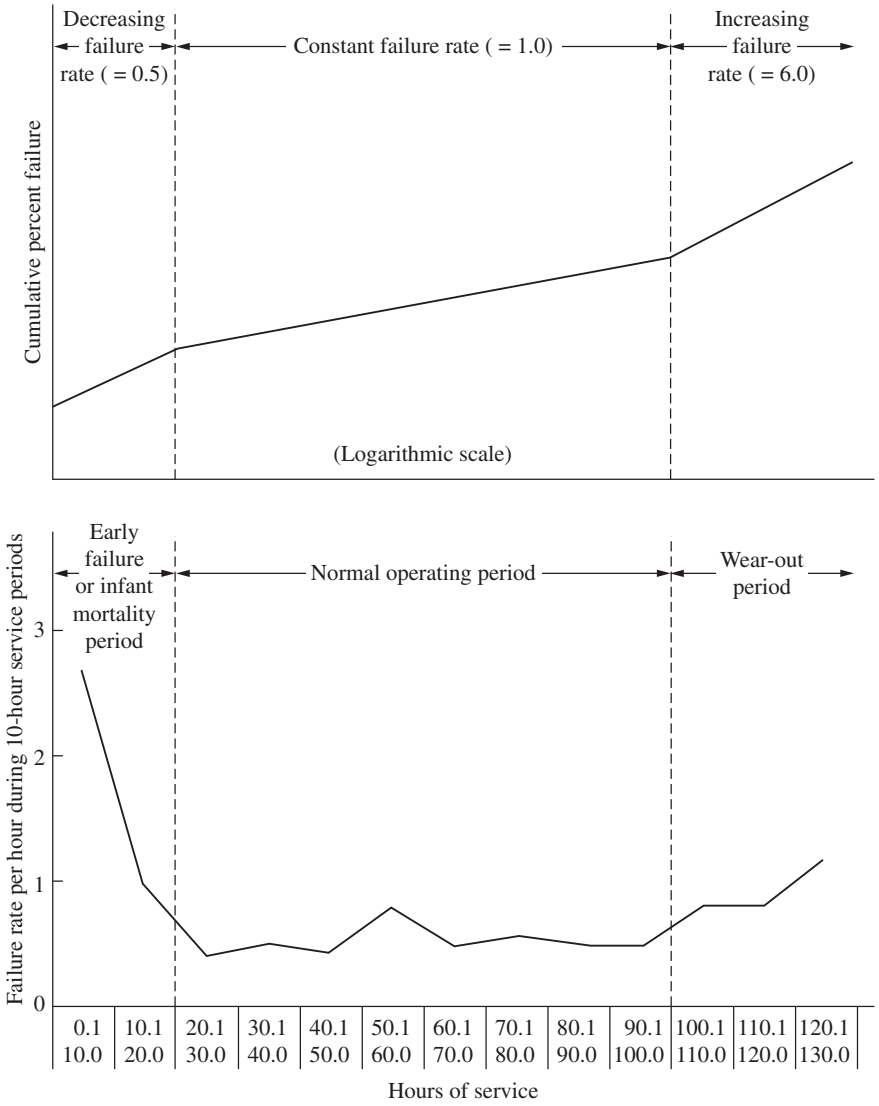


Figure 19.19 Failure rate vs. time. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

are the result of blunders in design or manufacture, misuse, or misapplication. Once corrected, these failures usually do not occur again (e.g., an oil hole that is not drilled). Sometimes it is possible to "debug" the product by a simulated use test or by overstressing (in electronics this is known as burn-in). The weak units still fail, but the failure takes place in the test rig rather than in service. O'Connor (1995) explains the use of burn-in tests and environmental screening tests.

- *The constant-failure-rate period*. Here the failures result from the limitations inherent in the design, changes in the environment, and accidents caused by use or maintenance.

The accidents can be held down by good control of operating and maintenance procedures. However, a reduction in the failure rate requires basic redesign.

- *The wear-out period*. These failures are due to old age (e.g., a metal becomes embrittled or insulation dries out). A reduction in failure rates requires preventive replacement of these dying components before they result in catastrophic failure.

The top portion of Figure 19.19 shows the corresponding Weibull plot when $\alpha = 2.6$ was applied to the original data (Table 19.14). The values of the shape parameter, β, were approximately 0.5, 1.0, and 6.0, respectively. A shape parameter less than 1.0 indicates a decreasing failure rate, a value of 1.0 a constant failure rate, and a value greater than 1.0 an increasing failure rate.

The Distribution of Time Between Failures. Users desire low failure rates during the infant mortality period, and after this are concerned with the length of time that a product will perform without failure. Thus, for repairable products, the time between failures (TBF) is a critical characteristic. The variation in time between failures can be studied statistically. The corresponding characteristic for nonrepairable products is usually called the time to failure.

When the failure rate is constant, the distribution of time between failures is distributed exponentially. Consider the 42 failure times in the constant failure rate portion of Table 19.13. The time between failures for successive failures can be tallied, and the 41 resulting TBFs can be formed into the frequency distribution shown in Figure 19.20a. The distribution is roughly exponential in shape, indicating that when the failure rate is constant, the distribution of time between failures (not mean time between failures) is exponential. This distribution is the basis of the exponential formula for reliability.

The Exponential Formula for Reliability

The distribution of TBF indicates the chance of failure-free operation for the specified time period. The chance of obtaining failure-free operation for a specified time period or longer can be shown by changing the TBF distribution to a distribution showing the number of intervals equal to or greater than a specified time length (Figure 19.20b). If the frequencies are expressed as relative frequencies, they become estimates of the probability of survival. When the failure rate is constant, the probability of survival (or reliability) is

$$P_s = R = e^{-t/\mu} = e^{-t\lambda}$$

where $P_s = R$ = probability of failure-free operation for a time period equal to or greater than t
$e = 2.718$
t = specified period of failure-free operation
μ = mean time between failures (the mean of TBF distribution)
λ = failure rate (the reciprocal of μ)

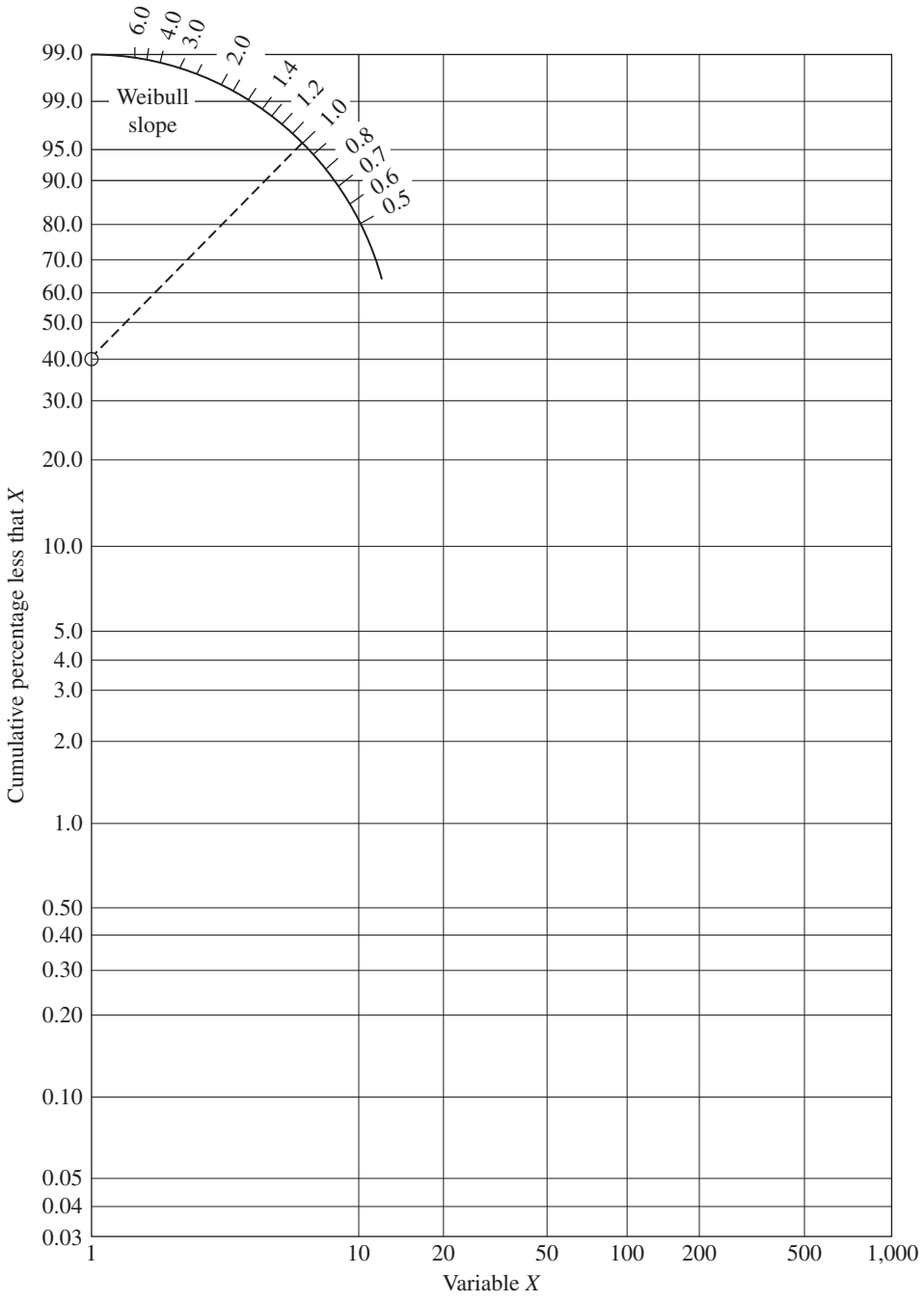
Note that this formula is simply the exponential probability distribution rewritten in terms of reliability.

Problem A washing machine requires 30 minutes to clean a load of clothes. The mean time between failures of the machine is 100 hours. Assuming a constant failure rate, what is the chance of the machine completing a cycle without failure?

Solution Applying the exponential formula, we obtain

$$R = e^{-t/\mu} = e^{-0.5/100} = 0.995$$

There is a 99.5 percent chance of completing a washing cycle.



(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.14 Weibull Paper

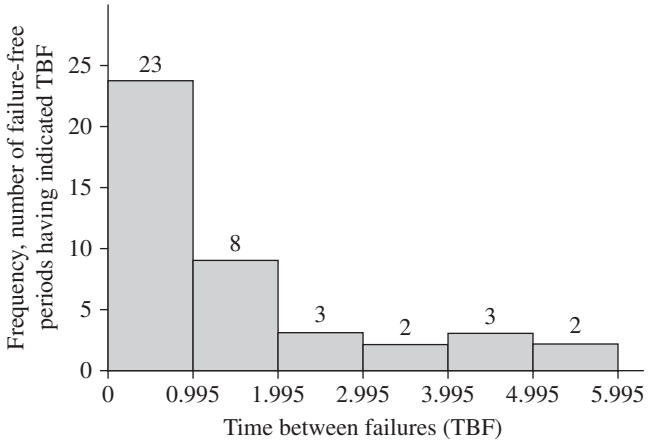


Figure 19.20*a* Histogram of TBF. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

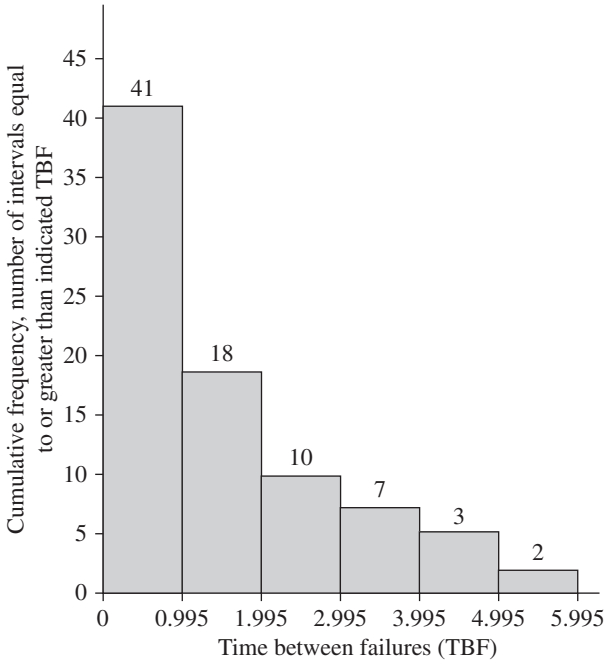


Figure 19.20*b* Cumulative histogram of TBF. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

How about the assumption of a constant failure rate? In practice, sufficient data usually are not available to evaluate the assumption. However, experience suggests that this assumption often is true, particularly when (1) infant mortality types of failures have been eliminated before delivery of the product to the user, and (2) the user replaces the product or specific components before the wear-out phase begins.

The Meaning of Mean Time Between Failures. Confusion surrounds the meaning of mean time between failures (MTBF). Further explanation is warranted:

- The MTBF is the mean (or average) time between successive failures of a product. This definition assumes that the product in question can be repaired and placed back into operation after each failure. For nonrepairable products, the term "mean time to failure" (MTTF) is used.
- If the failure rate is constant, the probability that a product will operate without failure for a time equal to or greater than its MTBF is only 37 percent. This outcome is based on the exponential distribution (R is equal to 0.37 when t is equal to the MTBF). This result is contrary to the intuitive feeling that there is a 50-50 chance of exceeding an MTBF.
- MTBF is not the same as "operating life," "service life," or other indexes, which generally connote overhaul or replacement time.
- An increase in an MTBF does not result in a proportional increase in reliability (the probability of survival). If $t = 1$ hour, the following table shows the MTBF required to obtain various reliabilities.

MTBF	*R*
5	0.82
10	0.90
20	0.95
100	0.99

A fivefold increase in MTBF from 20 to 100 hours is necessary to increase the reliability by 4 percentage points compared with a doubling of the MTBF from 5 to 10 hours to get an 8 percentage point increase in reliability.

MTBF is a useful measure of reliability, but it is not correct for all applications. Other reliability indexes are listed in Chapter 28, Research & Development: More Innovation, Scarce Resources.

The Relationship Between Part and System Reliability

It often is assumed that system reliability (i.e., the probability of survival, P_s) is the product of the individual reliabilities of the n parts within the system:

$$P_s = P_1 P_2 \ldots P_n$$

For example, if a communications system has four subsystems with reliabilities of 0.970, 0.989, 0.995, and 0.996, the system reliability is the product, or 0.951. The formula assumes that (1) the failure of any part causes failure of the system and (2) the reliabilities of the parts are independent of one another (i.e., the reliability of one part does not depend on the functioning of another part).

These assumptions are not always true, but in practice, the formula serves two purposes. First, it shows the effect of increased complexity of equipment on overall reliability. As the number of parts in a system increases, the system reliability decreases dramatically (see Figure 19.21). Second, the formula often is a convenient approximation that can be refined as information on the interrelationships of the parts becomes available.

When it can be assumed that (1) the failure of any part causes system failure, (2) the parts are independent, and (3) each part follows an exponential distribution, then

$$P_s = e^{-t_1\lambda_1} e^{-t_2\lambda_2} \ldots e^{-t_n\lambda_n}$$

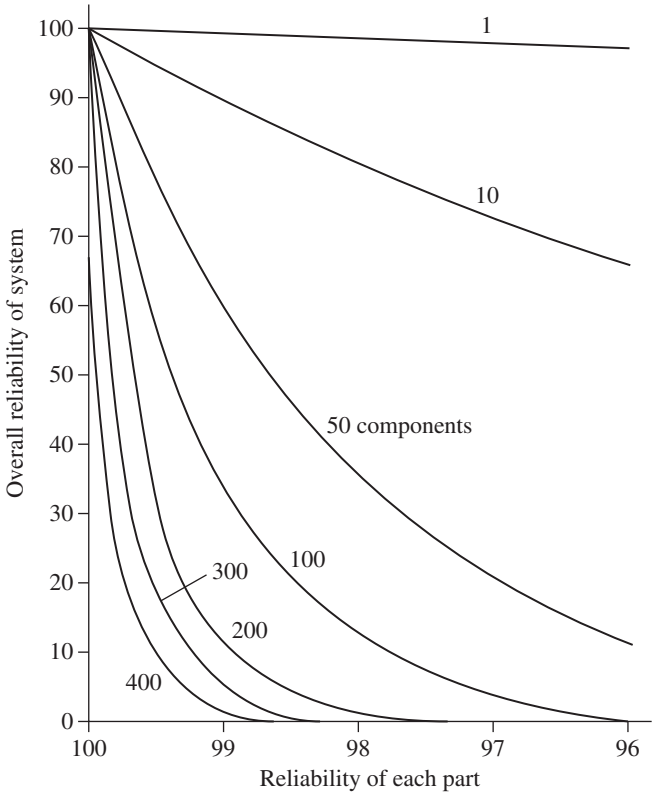


Figure 19.21 Relationship between part and system reliability. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

Further, if t is the same for each part,

$$P_s = e^{-1\Sigma\lambda}$$

Thus, when the failure rate is constant (and therefore the exponential distribution can be applied), the reliability of a system can be predicted based on the addition of the part failure rates (see the section "Predicting Reliability during Design," next).

Sometimes designs are planned with redundancy so that the failure of one part will not cause system failure. Redundancy is an old (but still useful) design technique invented long before the advent of reliability prediction techniques. However, the designer can now predict the effect of redundancy on system reliability in quantitative terms.

Redundancy is the existence of more than one element for accomplishing a given task, where all elements must fail before there is an overall failure of the system. In parallel redundancy (one of several types of redundancy), two or more elements operate at the same time to accomplish the task, and any single element is capable of handling the job itself in case of failure of the other elements. When parallel redundancy is used, the overall reliability is calculated as follows:

$$P_s = 1 - (1 - \mathrm{P1})n$$

where P_s = reliability of the system
P1 = reliability of the individual elements in the redundancy
n = number of identical redundant elements

Problem Suppose that a unit has a reliability of 99.0 percent for a specified mission time. If two identical units are used in parallel redundancy, what overall reliability will be expected?

Solution Applying the formula above, we obtain

$$R = 1 - (1 - 0.99)(1 - 0.99) = 0.9999, \text{ or } 99.99 \text{ percent}$$

Predicting Reliability during Design

Reliability prediction methods continue to evolve, but include such standards as failure mode and effects analysis (FMEA) and testing. Ireson et al. (1996) provide an extensive discussion of reliability prediction, and should be consulted beyond the methods discussed in this handbook.

The following steps make up a reliability prediction method:

1. Define the product and its functional operation. The system, subsystems, and units must be precisely defined in terms of their functional configurations and boundaries. This precise definition is aided by preparation of a functional block diagram that shows the subsystems and lower-level products, their interrelationships, and the interfaces with other systems. Given a functional block diagram and a well-defined statement of the functional requirements of the product, the conditions that constitute failure or unsatisfactory performance can be defined.
2. Prepare a reliability block diagram. For systems in which there are redundancies or other special interrelationships among parts, a reliability block diagram is useful. This diagram is similar to a functional block diagram, but the reliability block diagram shows exactly what must function for successful operation of the system. The diagram shows redundancies and alternative modes of operation. The reliability block diagram is the foundation for developing the probability model for reliability. O'Connor (1995) provides further discussion.
3. Develop the probability model for predicting reliability. A simple model may add only failure rates; a complex model can account for redundancies and other conditions.
4. Collect information relevant to parts reliability. The data include information such as parts function, parts ratings, stresses, internal and external environments, and operating time. Many sources of failure-rate information state failure rates as a function of operating parameters. For example, failure rates for fixed ceramic capacitors are stated as a function of (1) expected operating temperature and (2) the ratio of the operating voltage to the rated voltage. Such data show the effect of derating (assigning a part to operate below its rated voltage) on reducing the failure rate.
5. Select parts reliability data. The required parts data consist of information on catastrophic failures and on tolerance variations with respect to time under known operating and environmental conditions. Acquiring these data is a major problem for the designer because there is no single reliability data bank comparable to handbooks such as those for physical properties of materials. Instead, the designer must build a data bank by securing reliability data from a variety of sources:

Field performance studies conducted under controlled conditions:

- Specified life tests
- Data from parts manufacturers or industry associations

- Customers' parts qualification and inspection tests
- Government agency data banks such as the Government Industry Data Exchange Program (GIDEP) and the Reliability Information Analysis Center (RIAC)

Combine all of the above to obtain the numerical reliability prediction.

Ireson et al. (1996) and O'Connor (1995) are excellent references for reliability prediction. Included are the basic methods of prediction, repairable versus nonrepairable systems, electronic and mechanical reliability, reliability testing, and software reliability. Box and Draper (1969) provides extensive discussion of reliability data analysis, including topics such as censored life data (not all test units have failed during the test) and accelerated-life test data analysis. Dodson (1999) explains how the use of computer spreadsheets can simplify reliability modeling using various statistical distributions.

Reliability prediction techniques based on component failure data to estimate system failure rates have generated controversy. Jones and Hayes (1999) present a comparison of predicted and observed performance for five prediction techniques using parts count analyses. The predictions differed greatly from observed field behavior and from each other. The standard ANSI/IEC/ASQC D60300-3-1-1997 (Dependability Management—Part 3: Application Guide—Section 1—Analysis Techniques for Dependability) compares five analysis techniques: FMEA/FMECA, fault tree analysis, reliability block diagram, Markov analysis, and parts count reliability prediction.

The reliability of a system evolves during design, development, testing, production, and field use. The concept of reliability growth assumes that the causes of product failures are discovered and action is taken to remove the causes, thus resulting in improved reliability of future units ("test, analyze, and fix"). Reliability growth models provide predictions of reliability due to such improvements. For elaboration, see O'Connor (1995). Also, ANSI/IEC/ASQC D601164-1997 (Reliability Growth—Statistical Test and Estimation Methods) and the related IEC 61164 Ed. 2.0 (2004) (Reliability growth—Statistical test and estimation methods) describe methods of estimating reliability growth.

Predicting Reliability Based on the Exponential Distribution

When the failure rate is constant and when study of a functional block diagram reveals that all parts must function for system success, then reliability is predicted to be the simple total of failure rates. An example of a subsystem prediction is shown in Table 19.15. The prediction for the subsystem is made by adding the failure rates of the parts; the MTBF is then calculated as the reciprocal of the failure rate.

For further discussion of reliability prediction, including an example for an electronic system, see Gryna et al. (2007).

Predicting Reliability Based on the Weibull Distribution

Prediction of overall reliability based on the simple addition of component failure rates is valid only if the failure rate is constant. When this assumption cannot be made, an alternative approach based on the Weibull distribution can be used.

1. Graphically, use the Weibull distribution to predict the reliability R for the time period specified. $R = 100 - \%$ failure. Do this for each component (Table 19.14).
2. Combine the component reliabilities using the product rule and/or redundancy formulas to predict system reliability.

Predictions of reliability using the exponential distribution or the Weibull distribution are based on reliability as a function of time. Next we consider reliability as a function of stress and strength.

Part Description	Quantity	Generic Failure Rate per Million Hours	Total Failure Rates per Million Hours
Heavy-duty ball bearing	6	14.4	86.4
Brake assembly	4	16.8	67.2
Cam	2	0.016	0.032
Pneumatic hose	1	29.28	29.28
Fixed displacement pump	1	1.464	1.464
Manifold	1	8.80	65.0
Guide pin	5	13.0	65.0
Control valve	1	15.20	15.20
Total assembly failure rate			273.376

MTBF = 1/0.000273376 = 3.657.9 hours

(*Source:* Adapted from Ireson et al., p. 19.9. *Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.15 Example of Mechanical Parts and Subsystem Failure Rates

Reliability as a Function of Applied Stress and Strength

Failures are not always a function of time. In some cases, a part will function indefinitely if its strength is greater than the stress applied to it. The terms "strength" and "stress" here are used in the broad sense of inherent capability and operating conditions applied to a part, respectively.

For example, operating temperature is a critical parameter, and the maximum expected temperature is 145°F (63°C). Further, capability is indicated by a strength distribution having a mean of 172°F (78°C) and a standard deviation of 13°F (7°C) (Figure 19.22). With knowledge of only the maximum temperatures, the safety margin is

$$\frac{172-145}{13}=2.08$$

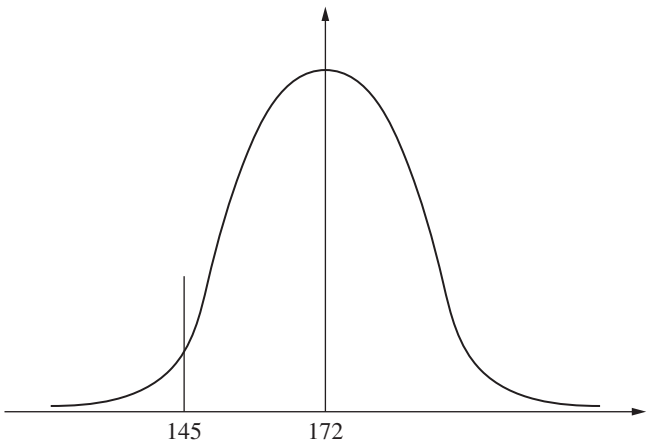
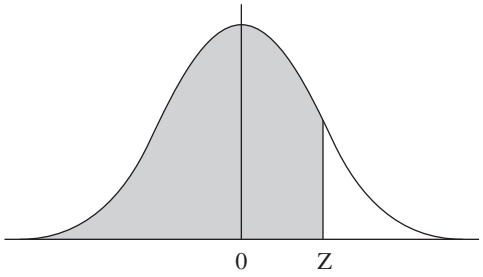


FIGURE 19.22 Distribution of strength.

The safety margin says that the average strength is 2.08 standard deviations above the maximum expected temperature of 145°F (63°C). Table 19.16 can be used to calculate a reliability of 0.981 [the area beyond 145°F (63°C)].

This calculation illustrates the importance of variation in addition to the average value during design. Designers have always recognized the existence of variation by using a safety factor in design. However, the safety factor is often defined as the ratio of average strength to the worst stress expected.

TABLE A
Normal distribution



Proportion of total areas under the curve from $-\infty$ to $Z = \frac{X-\mu}{\sigma}$, To illustrate when $Z = 2$, the probability is .9773 of obtaining a value equal to or less then X.

Z	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
–3.0	.00100	.00104	.00107	.00111	.00114	.00118	.00122	.00126	.00131	.00135
–2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0017	.0018	.0019
–2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
–2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
–2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
–2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062
–2.4	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082
–2.3	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107
–2.2	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139
–2.1	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179
–2.0	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228
–1.9	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287
–1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359
–1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446
–1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548
–1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668
–1.4	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808
–1.3	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968
–1.2	.0985	.1003	.1020	.1038	.1057	.1075	.1093	.1112	.1131	.1151
–1.1	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.16 Normal Distribution

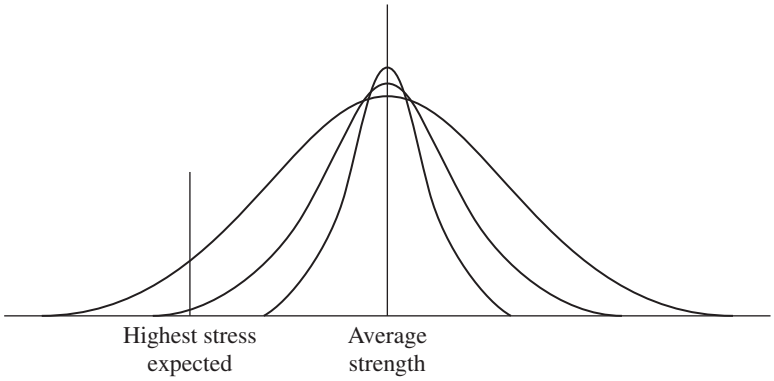


Figure 19.23 Variation and safety factor. (*Quality Planning and Analysis*, *Copyright 2007. Used by permission.*)

Note that in Figure 19.23, all designs have the same safety factor. Also note that the reliability (probability of a part having a strength greater than the stress) varies considerably. Thus, the uncertainty often associated with this definition of safety factor is, in part, due to its failure to reflect the variation in both strength and stress. Such variation is partially reflected in a safety margin, defined as

$$\frac{\text{Average strength} - \text{worst stress}}{\text{Standard deviation of strength}}$$

This recognizes the variation in strength but is conservative because it does not recognize a variation in stress.

Availability

Availability has been defined as the probability that a product, when used under given conditions, will perform satisfactorily when called upon. Availability considers the operating time of the product and the time required for repairs. Idle time, during which the product is not needed, is excluded.

Availability is calculated as the ratio of operating time to operating time plus downtime. However, downtime can be viewed in two ways:

- *Total downtime*. This period includes active repair (diagnosis and repair time), preventive maintenance time, and logistics time (time spent waiting for personnel, spare parts, etc.). When total downtime is used, the resulting ratio is called operational availability (A_0).
- *Active repair time*. The resulting ratio is called intrinsic availability (A_i). Under certain conditions, availability can be calculated as:

$$A_0 = \frac{\text{MTBF}}{\text{MTBF+MDT}} \quad \text{and} \quad A_i = \frac{\text{MTBF}}{\text{MTBF+MTTR}}$$

where MTBF = mean time between failures
MDT = mean downtime
MTTR = mean time to repair

This is known as the steady-state formula for availability. The steady-state formula for availability has the virtue of simplicity. However, the formula is based on several assumptions that are not always met in the real world. The assumptions are

- The product is operating in the constant failure rate period of the overall life. Thus, the failure-time distribution is exponential.
- The downtime or repair-time distribution is exponential.
- Attempts to locate system failures do not change the overall system failure rate.
- No reliability growth occurs (such growth might be due to design improvements or through debugging of bad parts).
- Preventive maintenance is scheduled outside the time frame included in the availability calculation.

More precise formulas for calculating availability depend on operational conditions and statistical assumptions. These formulas are discussed by Ireson et al. (1996).

Setting Specification Limits

A major step in the development of physical products is the conversion of product features into dimensional, chemical, electrical, and other characteristics of the product. Thus, a heating system for an automobile will have many characteristics for the heater, air ducts, blower assembly, engine coolant, etc.

For each characteristic, the designer must specify (1) the desired average (or "nominal value") and (2) the specification limits (or "tolerance limits") above and below the nominal value that individual units of product must meet. These two elements relate to parameter design and tolerance design, as discussed in Gryna et al. (2007).

The specification limits should reflect the functional needs of the product, manufacturing variability, and economic consequences. These three aspects are addressed in the next three sections. For greater depth in the statistical treatment of specification limits, see Anand (1996).

Specification Limits and Functional Needs

Sometimes data can be developed to relate product performance to measurements of a critical component. For example, a thermostat may be required to turn on and shut off a power source at specified low and high temperature values, respectively. A number of thermostat elements are built and tested. The prime recorded data are (1) turn-on temperature, (2) shut-off temperature, and (3) physical characteristics of the thermostat elements. We can then prepare scatter diagrams (Figure 19.24) and regression equations to help establish critical component tolerances on a scientific basis within the confidence limits for the numbers involved. Ideally, the sample size is sufficient, and the data come from a statistically controlled process—two conditions that are both rarely achieved. O'Connor (1995) explains how this approach can be related to the Taguchi approach to develop a more robust design.

Specification Limits and Manufacturing Variability

Generally, designers will not be provided with information on process capability. Their problem will be to obtain a sample of data from the process, calculate the limits that the process can meet, and compare these to the limits they were going to specify. If they do not have any limits in mind, the capability limits calculated from process data provide a set of limits that are realistic from the viewpoint of producibility. These limits must then be evaluated against the functional needs of the product.

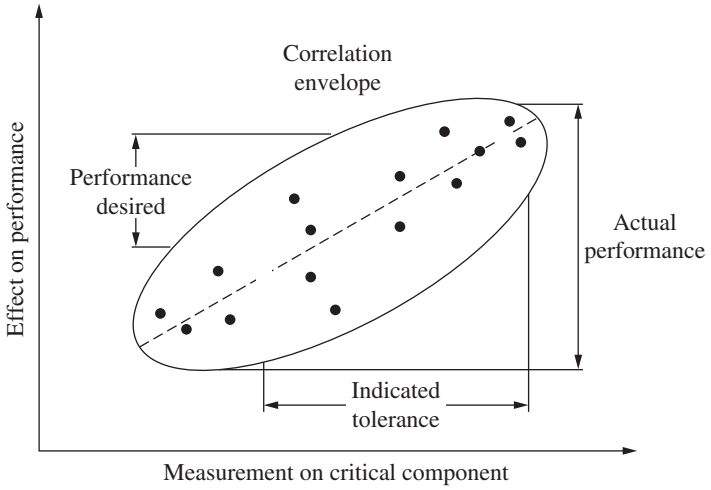


Figure 19.24 Approach to functional tolerancing. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

Statistically, the problem is to predict the limits of variation of individual items in the total population based on a sample of data. For example, suppose that a product characteristic is normally distributed with a population average of 5.000 in (12.7 cm) and a population standard deviation of 0.001 in (0.00254 cm). Limits can then be calculated to include any given percentage of the population. Figure 19.25 shows the location of the 99 percent limits. Table 19.16 indicates that 2.575 standard deviations will include 99 percent of the population. Thus, in this example, a realistic set of tolerance limits would be

$$5.000 \pm 2.575(0.001) = \frac{5.003}{4.997}$$

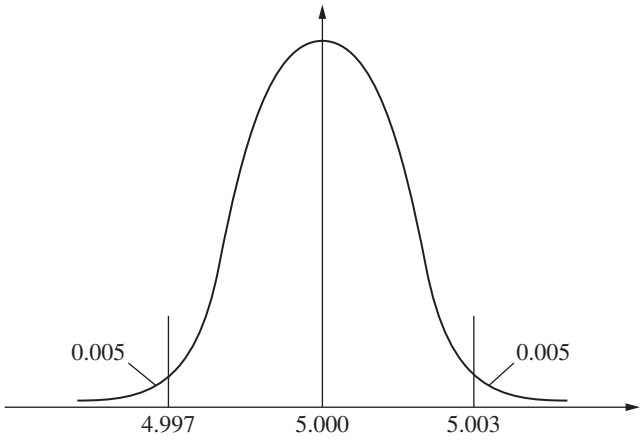


Figure 19.25 Distribution with 99 percent limits. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

Ninety-nine percent of the individual pieces in the population will have values between 4.997 and 5.003.

In practice, the average and standard deviation of the population are not known but must be estimated from a sample of product from the process. As a first approximation, tolerance limits are sometimes set at

$$\overline{X} \pm 3s$$

Here, the average $\bar{X}$ and standard deviation s of the sample are used directly as estimates of the population values. If the true average and standard deviation of the population happen to be equal to those of the sample, and if the characteristic is normally distributed, then 99.73 percent of the pieces in the population will fall within the limits calculated. These limits are frequently called natural tolerance limits (limits that recognize the actual variation of the process and therefore are realistic). This approximation ignores the possible error in both the average and standard deviation as estimated from the sample.

Methodology has been developed for setting tolerance limits in a more precise manner. For example, formulas and tables are available for determining tolerance limits based on a normally distributed population. Table 19.17 provides factors for calculating tolerance limits that recognize the uncertainty in the sample mean and sample standard deviation. The tolerance limits are determined as

$$\overline{X} \pm Ks$$

The factor K is a function of the confidence level desired, the percentage of the population to be included within the tolerance limits, and the number of data values in the sample.

For example, suppose that a sample of 10 resistors from a process yielded an average and standard deviation of 5.04 and 0.016, respectively. The tolerance limits are to include 99 percent of the population, and the tolerance statement is to have a confidence level of 95 percent. Referring to Table 19.17, the value of K is 4.433, and tolerance limits are then calculated as

$$5.04 \pm 4.433(0.016) = \frac{5.11}{4.97}$$

We are 95 percent confident that at least 99 percent of the resistors in the population will have resistance between 4.97 and 5.11 Ω. Tolerance limits calculated in this manner are often called statistical tolerance limits. This approach is more rigorous than the $3s$ natural tolerance limits, but the two percentages in the statement are a mystery to those without a statistical background.

For products in some industries (e.g., electronics), the number of units outside of specification limits is stated in terms of parts per million (ppm). Thus, if limits are set at three standard deviations, 2700 ppm (100 to 99.73 percent) will fall outside the limits. For many applications (e.g., a personal computer with many logic gates), such a level is totally unacceptable. Table 19.18 shows the ppm for several standard deviations. These levels of ppm assume that the process average is constant at the nominal specification. A deviation from the nominal value will result in a higher ppm value. To allow for modest shifts in the process average, some manufacturers follow a guideline for setting specification limits at $\pm 6\sigma$.

Designers often must set tolerance limits with only a few measurements from the process (or more likely from the development tests conducted under laboratory conditions).

Tolerance Factors for Normal Distributions (Two Sided)										
P / N	γ = 0.75					γ = 0.90				
	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
2	4.498	6.301	7.414	9.531	11.920	11.407	15.978	18.800	24.167	30.227
3	2.501	3.538	4.187	5.431	6.844	4.132	5.847	6.919	8.974	11.309
4	2.035	2.892	3.431	4.471	5.657	2.932	4.166	4.943	6.440	8.149
5	1.825	2.599	3.088	4.033	5.117	2.454	3.494	4.152	5.423	6.879
6	1.704	2.429	2.889	3.779	4.802	2.196	3.131	3.723	4.870	6.188
7	1.624	2.318	2.757	3.611	4.593	2.034	2.902	3.452	4.521	5.750
8	1.568	2.238	2.663	3.491	4.444	1.921	2.743	3.264	4.278	5.446
9	1.525	2.178	2.593	3.400	4.330	1.839	2.626	3.125	4.098	5.220
10	1.492	2.131	2.537	3.328	4.241	1.775	2.535	3.018	3.959	5.046
11	1.465	2.093	2.493	3.271	4.169	1.724	2.463	2.933	3.849	4.906
12	1.443	2.062	2.456	3.223	4.110	1.683	2.404	2.863	3.758	4.792
13	1.425	2.036	2.424	3.183	4.059	1.648	2.355	2.805	3.682	4.697
14	1.409	2.013	2.398	3.148	4.016	1.619	2.314	2.756	3.618	4.615
15	1.395	1.994	2.375	3.118	3.979	1.594	2.278	2.713	3.562	4.545
16	1.383	1.977	2.355	3.092	3.946	1.572	2.246	2.676	3.514	4.484
17	1.372	1.962	2.337	3.069	3.917	1.552	2.219	2.643	3.471	4.430
18	1.363	1.948	2.321	3.048	3.891	1.535	2.194	2.614	3.433	4.382
19	1.355	1.936	2.307	3.030	3.867	1.520	2.172	2.588	3.399	4.339
20	1.347	1.925	2.294	3.013	3.846	1.506	2.152	2.564	3.368	4.300
21	1.340	1.915	2.282	2.998	3.827	1.493	2.135	2.543	3.340	4.264
22	1.334	1.906	2.271	2.984	3.809	1.482	2.118	2.524	3.315	4.232
23	1.328	1.898	2.261	2.971	3.793	1.471	2.103	2.506	3.292	4.203
24	1.322	1.891	2.252	2.950	3.778	1.462	2.089	2.480	3.270	4.176
25	1.317	1.883	2.244	2.948	3.764	1.453	2.077	2.474	3.251	4.151
26	1.313	1.877	2.236	2.938	3.751	1.444	2.065	2.460	3.232	4.127
27	1.309	1.871	2.229	2.929	3.740	1.437	2.054	2.447	3.215	4.106
30	1.297	1.855	2.210	2.904	3.708	1.417	2.025	2.413	3.170	4.049
35	1.283	1.834	2.185	2.871	3.667	1.390	1.988	2.368	3.112	3.974
40	1.271	1.818	2.166	2.846	3.635	1.370	1.959	2.334	3.066	3.917
100	1.218	1.742	2.075	2.727	3.484	1.275	1.822	1.172	2.854	3.646
500	1.177	1.683	2.006	2.636	3.368	1.201	1.717	2.046	2.689	3.434
1000	1.169	1.671	1.992	2.617	3.344	1.185	1.695	2.019	2.654	3.390
∞	1.150	1.645	1.960	2.576	3.291	1.150	1.645	1.960	2.576	3.291

TABLE 19.17 Tolerance Factors for Normal Distributions

γ = 0.95					γ = 0.99				
0.75	**0.90**	**0.95**	**0.99**	**0.999**	**0.75**	**0.90**	**0.95**	**0.99**	**0.999**
22.858	32.019	37.674	48.430	60.573	114.363	160.363	188.491	242.300	303.054
5.922	8.380	9.916	12.861	16.208	13.378	18.930	22.401	29.055	36.616
3.779	5.369	6.370	8.299	10.502	6.614	9.398	11.150	14.527	18.383
3.002	4.275	5.079	6.634	8.415	4.643	6.612	7.855	10.260	13.015
2.604	3.712	4.414	5.775	7.337	3.743	5.337	6.345	8.301	10.548
2.361	3.369	4.007	5.248	6.676	3.233	4.613	5.488	7.187	9.142
2.197	3.136	3.732	4.891	6.226	2.905	4.147	4.936	6.468	8.234
2.078	2.967	3.532	4.631	5.899	2.677	3.822	4.550	5.966	7.600
1.987	2.839	3.379	4.433	5.649	2.508	3.582	4.265	5.594	7.129
1.916	2.737	3.259	4.277	5.452	2.378	3.397	4.045	5.308	6.766
1.858	2.655	3.162	4.150	5.291	2.274	3.250	3.870	5.079	6.477
1.810	2.587	3.081	4.044	5.158	2.190	3.130	3.727	4.893	6.240
1.770	2.529	3.012	3.955	5.045	2.120	3.029	3.608	4.737	6.043
1.735	2.480	2.954	3.878	4.949	2.060	2.945	3.507	4.605	5.876
1.705	2.437	2.903	3.812	4.865	2.009	2.872	3.421	4.492	5.732
1.679	2.400	2.858	3.754	4.791	1.965	2.808	3.345	4.393	5.607
1.655	2.366	2.819	3.702	4.725	1.926	2.753	3.279	4.307	5.497
1.635	2.337	2.784	3.656	4.667	1.891	2.703	3.221	4.230	5.399
1.616	2.310	2.752	3.615	4.614	1.860	2.659	3.168	4.161	5.312
1.599	2.286	2.723	3.577	4.567	1.833	2.620	3.121	4.100	5.234
1.584	2.264	2.697	3.543	4.523	1.808	2.584	3.078	4.044	5.163
1.570	2.244	2.673	3.512	4.484	1.795	2.551	3.040	3.993	5.098
1.557	2.225	2.651	3.483	4.447	1.764	2.522	3.004	3.947	5.039
1.545	2.208	2.631	3.457	4.413	1.745	2.494	2.972	3.904	4.985
1.534	2.193	2.612	3.432	4.382	1.727	2.460	2.941	3.865	4.935
1.523	2.178	2.595	3.409	4.353	1.711	2.446	2.914	3.828	4.888
1.497	2.140	2.549	3.350	4.278	1.668	2.385	2.841	3.733	4.768
1.462	2.090	2.490	3.272	4.179	1.613	2.306	2.748	3.611	4.611
1.435	2.052	2.445	3.213	4.104	1.571	2.247	2.677	3.518	4.493
1.311	1.874	2.233	2.934	3.748	1.383	1.977	2.355	3.096	3.954
1.215	1.737	2.070	2.721	3.475	1.243	1.777	2.117	2.783	3.555
1.195	1.709	2.036	2.676	3.418	1.214	1.736	2.068	2.718	3.472
1.150	1.645	1.960	2.576	3.291	1.150	1.645	1.960	2.576	3.291

*Table H—Tolerance factors for normal distributions" from *Selected Techniques of Statistical Analysis—OSRD* by C. Eisenhart, M. W. Hastay, and W. A. Wallis, Copyright 1947 by The McGraw-Hill Companies, Inc. Reprinted by permission of The McGraw-Hill Companies, Inc.

γ = confidence level

P = percentage of population within tolerance limits

N = number of values in sample

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.17 (*Continued*)

Number of Standard Deviations	Part per Million (ppm)
±3σ	2700
±4σ	63
±5σ	0.57
±6σ	0.002

*If the process is not centered and the mean shifts by up to 1.5σ, then ±6σ will be 3.4 ppm.
(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.18 Standard Deviations and PPM (centered process)*

In developing a paint formulation, for example, the following values of gloss were obtained: 76.5, 75.2, 77.5, 78.9, 76.1, 78.3, and 77.7. A group of chemists was asked where they would set a minimum specification limit. Their answer was 75.0—a reasonable answer for those without statistical knowledge. Figure 19.26 shows a plot of the data on normal probability paper. If the line is extrapolated to 75.0, the plot predicts that about 11 percent of the population will fall below 75.0, even though all of the sample data exceed 75.0. Of course, a larger

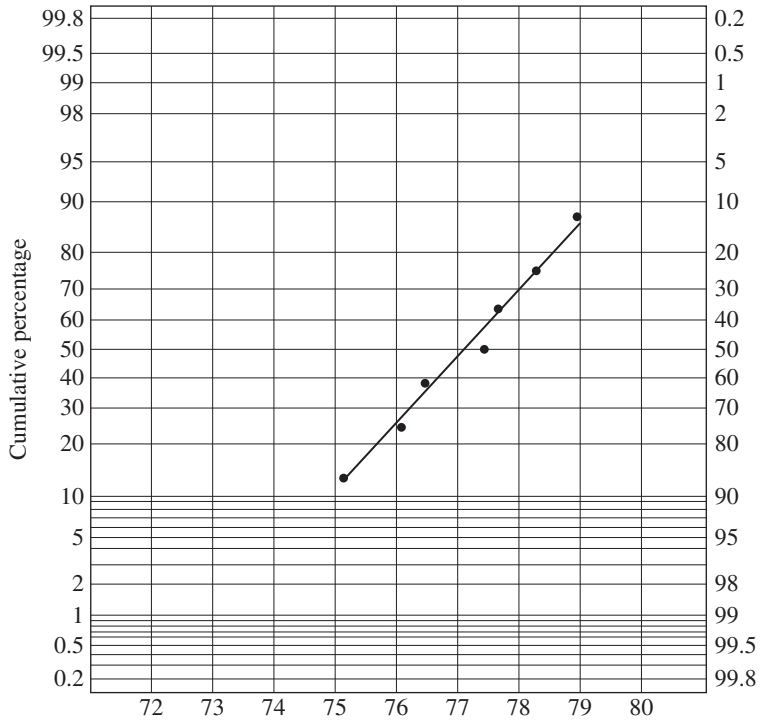


FIGURE 19.26 Probability plot of development data. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

Name of Limit	Meaning
Tolerance	Set by the engineering design function to define the minimum and maximum values allowable for the product to work properly
Statistical tolerance	Calculated from process data to define the amount of variation that the process exhibits; these limits will contain a specified proportion of the total population
Prediction	Calculated from process data to define the limits which will contain all of *k* future observations
Confidence	Calculated from data to define an interval within which a population parameter lies
Control	Calculated from process data to define the limits of chance (random) variation around some central value

(*Source: Quality Planning and Analysis*, Copyright 2007. Used by permission.)

TABLE 19.19 Distinctions Among Limits

sample size is preferred and further statistical analyses could be made, but the plot provides a simple tool for evaluating a small sample of data.

All methods of setting tolerance limits based on process data assume that the sample of data represents a process that is sufficiently stable to be predictable. In practice, the assumption is often accepted without any formal evaluation. If sufficient data are available, the assumption should be checked with a control chart.

Statistical tolerance limits are sometimes confused with other limits used in engineering and statistics. Table 19.19 summarizes the distinctions among five types of limits (see also Box, pp. 44.47–44.58).

Specifications Limits and Economic Consequences

In setting traditional specification limits around a nominal value, we assume that there is no monetary loss for product falling within specification limits. For product falling outside the specification limits, the loss is the cost of replacing the product.

Another viewpoint holds that any deviation from the nominal value causes a loss. Thus, there is an ideal (nominal) value that customers desire, and any deviation from this ideal results in customer dissatisfaction. This loss can be described by a loss function (Figure 19.27).

Many formulas can predict loss as a function of deviation from the target. Taguchi proposes the use of a simple quadratic loss function:

$$L = k(X - T)^2$$

where L = loss in monetary terms
k = cost coefficient
X = value of quality characteristic
T = target value

Ross (1996) provides an example to illustrate how the loss function can help to determine specification limits. In automatic transmissions for trucks, shift points are designed to

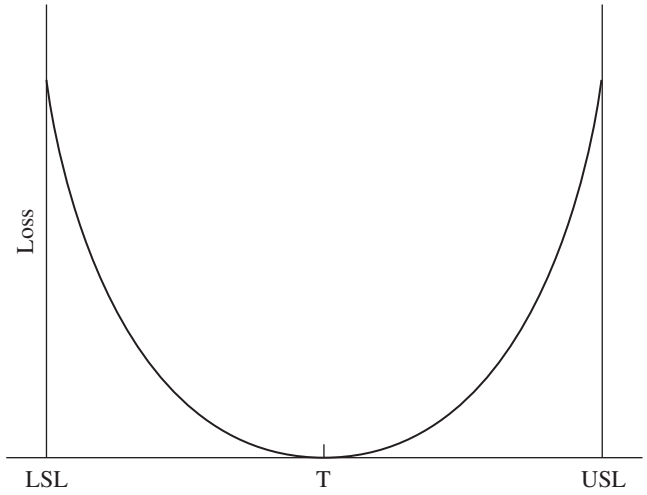


Figure 19.27 Loss function. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

occur at a certain speed and throttle position. Suppose it costs the producer $100 to adjust a valve body under warranty when a customer complains of the shift point. Research indicates that the average customer would request an adjustment if the shift point is off from the nominal by 40 rpm transmission output speed on the first-to-second gear shift. The loss function is then

$$\text{Loss} = k(X - T)^2$$
$$100 = k(40)^2$$
$$k = \$0.0625$$

This adjustment can be made at the factory at a lower cost, about $10. The loss function is now used to calculate the specification limits:

$$\$10 = 0.0625(X - T)^2$$
$$(X - T) = \pm 12.65 \text{ or } \pm 13 \text{ rpm}$$

The specification limits should be set at 13 rpm around the desired nominal value. If the transmission shift point is further than 13 rpm from the nominal, adjustment at the factory is less expensive than waiting for a customer complaint and making the adjustment under warranty in the field. Ross (1996) discusses how the loss function can be applied to set one-sided specification limits (e.g., a minimum value or a maximum value).

Specification Limits for Interacting Dimensions

Interacting dimensions mate or merge with other dimensions to create a final result. Consider the simple mechanical assembly shown in Figure 19.28. The lengths of components A, B, and C are interacting dimensions because they determine the overall assembly length.

Suppose the components were manufactured to the specifications indicated in Figure 19.28. A logical specification for the assembly length would be 3.500 ± 0.0035, giving limits of

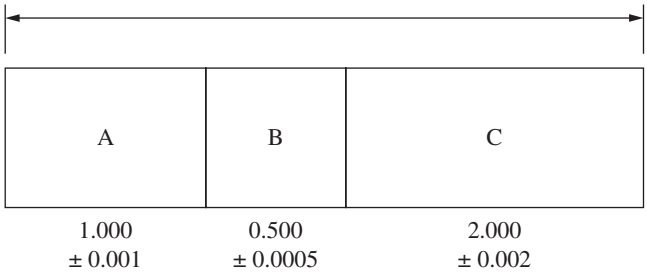


Figure 19.28 Mechanical assembly. (*Quality Planning and Analysis, Copyright 2007. Used by permission.*)

3.5035 and 3.4965. This logic may be verified from the two extreme assemblies shown in the following table.

Maximum	Minimum
1.001	0.999
0.5005	0.4995
2.002	1.998
3.5035	3.4965

The approach of adding component tolerances is mathematically correct, but is often too conservative. Suppose that about 1 percent of the pieces of component A are expected to be below the lower tolerance limit for component A and suppose the same for components B and C. If a component A is selected at random, there is, on average, 1 chance in 100 that it will be on the low side, and similarly for components B and C. The key point is this: If assemblies are made at random and if the components are manufactured independently, then the chance that an assembly will have all three components simultaneously below the lower tolerance limit is

$$\frac{1}{100}\times\frac{1}{100}\times\frac{1}{100}=\frac{1}{1,000,000}$$

There is only about one chance in a million that all three components will be too small, resulting in a small assembly. Thus, setting component and assembly tolerances based on the simple addition formula is conservative in that it fails to recognize the extremely low probability of an assembly containing all low (or all high) components.

The statistical approach is based on the relationship between the variances of a number of independent causes and the variance of the dependent or overall result. This may be written as

$$\sigma_{\text{result}}=\sqrt{\sigma^2_{\text{causeA}}+\sigma^2_{\text{causeB}}+\sigma^2_{\text{causeC}}+\ldots}$$

In terms of the assembly example, the formula is:

$$\sigma_{\text{assembly}}=\sqrt{\sigma_A^2+\sigma_B^2+\sigma_C^2}$$

Now suppose that for each component, the tolerance range is equal to three standard deviations (or any constant multiple of the standard deviation). Because σ is equal to T divided by 3, the variance relationship may be rewritten as

$$\frac{T}{3} = \sqrt{\left(\frac{T_A}{3}\right)^2 + \left(\frac{T_B}{3}\right)^2 + \left(\frac{T_C}{3}\right)^2}$$

or

$$T_{\text{assembly}} = \sqrt{T_A^2 + T_B^2 + T_C^2}$$

Thus, the squares of tolerances are added to determine the square of the tolerance for the overall result. This formula compares to the simple addition of tolerances commonly used.

The effect of the statistical approach is dramatic. Listed below are two possible sets of component tolerances that will yield an assembly tolerance equal to 0.0035 when used with the previous formula.

Component	Alternative 1	Alternative 2
A	±0.002	±0.001
B	±0.002	±0.001
C	±0.002	±0.003

With alternative 1, the tolerance for component A has been doubled, the tolerance for component B has been quadrupled, and the tolerance for component C has been kept the same as the original component tolerance based on the simple addition approach. If alternative 2 is chosen, similar significant increases in the component tolerances may be achieved. This formula, then, may result in a larger component tolerance with no change in the manufacturing processes and no change in the assembly tolerance.

The risk of this approach is that an assembly may fall outside the assembly tolerance. However, this probability can be calculated by expressing the component tolerances as standard deviations, calculating the standard deviation of the result, and finding the area under the normal curve outside the assembly tolerance limits. For example, if each component tolerance is equal to 3s, then 99.73 percent of the assemblies will be within the assembly tolerance, that is, 0.27 percent, or about 3 assemblies in 1000 taken at random would fail to meet the assembly tolerance. The risk can be eliminated by changing components for the few assemblies that do not meet the assembly tolerance.

The tolerance formula is not restricted to outside dimensions of assemblies. Generalizing, the left side of the equation contains the dependent variable or physical result, and the right side of the equation contains the independent variables of physical causes. If the result is placed on the left and the causes on the right, the formula always has plus signs under the square root—even if the result is an internal dimension (such as the clearance between a shaft and hole). The causes of variation are additive wherever the physical result happens to fall.

The formula has been applied to a variety of mechanical and electronic products. The concept may be applied to several interacting variables in an engineering relationship. The nature of the relationship need not be additive (assembly example) or subtractive (shaft-and-hole example). The tolerance formula can be adapted to predict the variation of results that are the product and/or the division of several variables.

Assumptions of the formula. The formula is based on several assumptions:

- The component dimensions are independent and each component to be assembled is chosen randomly. These assumptions are usually met in practice.
- Each component dimension should be normally distributed. Some departure from this assumption is permissible.
- The actual average for each component is equal to the nominal value stated in the specification. For the original assembly example, the actual averages for components A, B, and C must be 1.000, 0.500, and 2.000, respectively. Otherwise, the nominal value of 3.500 will not be achieved for the assembly and tolerance limits set at about 3.500 will not be realistic. Thus it is important to control the average value for interacting dimensions. Consequently, process control techniques are needed using variables measurement.

Use caution if any assumption is violated. Reasonable departures from the assumptions may still permit applying the concept of the formula. Notice that in the example, the formula resulted in the doubling of certain tolerances. This much of an increase may not even be necessary from the viewpoint of process capability.

Bender (1975) has studied these assumptions for some complex assemblies and concluded, based on a "combination of probability and experience," that a factor of 1.5 should be included to account for the assumptions:

$$T_{\text{result}} = 1.5\sqrt{T_A^2 + T_B^2 + T_C^2 + \cdots}$$

Graves (1997) suggests developing different factors for initial versus mature production, high versus low volume production, and mature versus developing technology and measurement processes.

Finally, variation simulation analysis is a technique that uses computer simulation to analyze tolerances. This technique can handle product characteristics with either normal or nonnormal distributions. Dodson (1999) describes the use of simulation in the tolerance design of circuits; Gomer (1998) demonstrates simulation to analyze tolerances in engine design. For an overall text on reliability, see Meeker and Escobar (1998).

Statistical Tools for Control

In addition to the fundamental control charts introduced in Chapter 18, Core Tools to Design, Control, and Improve Performance, there are some special-purpose methods for control that are sometimes helpful.

PRE-Control

PRE-Control is a statistical technique for detecting process conditions and changes that may cause defects (rather than changes that are statistically significant). PRE-Control focuses on controlling conformance to specifications, rather than statistical control. PRE-Control starts a process centered between specification limits and detects shifts that might result in making some of the parts outside a specification limit. It requires no plotting and no computations, and it needs only three measurements to give control information. The technique uses the normal distribution curve to determine significant changes in either the aim or the spread of a production process that could result in increased production of defective work.

The relative simplicity of PRE-Control versus statistical control charts can have important advantages in many applications. The concept, however, has generated some controversy. For a comparison of PRE-Control versus other approaches and the most appropriate applications of PRE-Control, see Ledolter and Swersey (1997) and Steiner (1997). For a complete story, also see the references in both of these papers.

Short-Run Control Charts

Some processes are carried out in such short runs that the usual procedure of collecting 20 to 30 samples to establish a control chart is not feasible. Sometimes these short runs are caused by previously known assignable causes that take place at predetermined times (such as a frequent shift in production from one product to another, as may be the case in lean production systems). Hough and Pond (1995) discuss four ways to construct control charts in these situations:

1. Ignore the systematic variability, and plot on a single chart.
2. Stratify the data, and plot them on a single chart.
3. Use regression analysis to model the data, and plot the residuals on a chart.
4. Standardize the data, and plot the standardized data on a chart.

The last option has received the most consideration. It involves transforming the data via the Z-transformation:

$$Z = \frac{X - \mu}{\sigma}$$

to remove any systematic changes in level and variability (thereby normalizing the data to a common baseline). This standardization of Shewhart charts has been discussed by Nelson (1989), Wheeler (1991), and Griffith (1996). Pyzdek (1993) also provides a good discussion of short and small runs.

Cumulative Sum Control Chart

The cumulative sum (CUMSUM or CUSUM) control chart is a chronological plot of the cumulative sum of deviations of a sample statistic (e.g., $\bar{X}$, p, number of nonconformities) from a reference value (e.g., the nominal or target specification). By definition, the CUMSUM chart focuses on a target value rather than on the actual average of process data. Each point plotted contains information from all observations (i.e., a cumulative sum). CUMSUM charts are particularly useful in detecting small shifts in the process average (say, 0.5σ to 2.0σ). The chart shown in Figure 19.29 is one way of constructing CUMSUM charts. The method is as follows:

1. Compute the control statistic (x-bar for the example in Figure 19.29).
2. Determine the target value T (10 in Figure 19.29).
3. Compute the standard deviation s (1.96 in Figure 19.29).
4. Draw a reference line at zero and upper and lower control limits (UCL and LCL respectively) at ±4s.
5. Compute the upper cumulative sum C_U for each sample point k as follows:

$$C_{U,k} = \text{Maximum}\left\{0, \sum_{i=1}^{k} [\bar{x}_i - (T + s/2)]\right\}$$

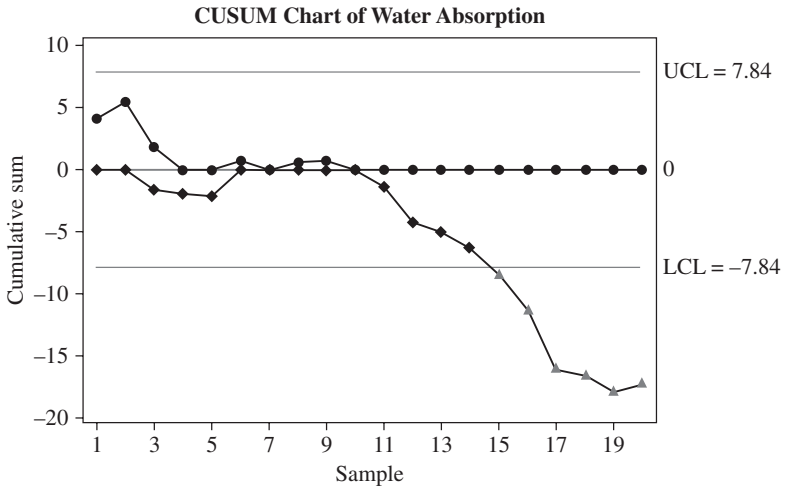


FIGURE 19.29 Cumulative sum control chart. (*Juran Institute, Inc. Copyright 1994. Used by permission.*)

6. Compute the upper cumulative sum C_L for each sample point k as follows:

$$C_{L,k} = \text{Minimum}\left\{0, \sum_{i=1}^{k} [\bar{x}_i - (T - s/2)]\right\}$$

7. Plot C_U and C_L as two separate lines.
8. When C_U exceeds the UCL, then an upward shift has occurred. When C_L drops below LCL, then a downward shift has occurred.

Moving Average Control Charts

Another special chart is the moving average chart. This chart is a chronological plot of the moving average, which is calculated as the average value updated by dropping the oldest individual measurement and adding the newest individual measurement. Thus, a new average is calculated with each individual measurement. A further refinement is the exponentially weighted moving average (EWMA) chart. In the EWMA chart, the observations are weighted, and the highest weight is given to the most recent data. Moving average charts are effective in detecting small shifts, highlighting trends, and using data in processes in which it takes a long time to produce a single item.

Box-Jenkins Manual Adjustment Chart

Still another chart is the Box-Jenkins manual adjustment chart. The average and range, CUMSUM, and EWMA charts for variables focus on monitoring a process and reducing variability due to special causes of variation identified by the charts. Box-Jenkins charts have a different objective: to analyze process data to regulate the process after each observation and thereby minimize process variation. For elaboration on this advanced technique, see Box and Luceño (1997).

Multivariate Control Charts

Finally, we consider the concept of multivariate control charts. When there are two or more quality characteristics on a unit of product, these could be monitored independently with separate control charts. Then the probability that a sample average on either control chart

exceeds three sigma limits is 0.0027. But the joint probability that both variables exceed their control limits simultaneously when they are both in control is (0.0027)(0.0027) or 0.00000729, which is much smaller than 0.0027. The situation becomes more distorted as the number of characteristics increases. For this and other reasons, monitoring several characteristics independently can be misleading. Multivariate control charts and statistics (e.g., Hotelling's T^2 charts, multivariate EWMA) address this issue. See Montgomery (2000, Section 8.4) for a highly useful discussion.

Process Capability

In planning the quality aspects of operations, nothing is more important than advance assurance that the processes will meet the specifications. In recent decades, a concept of process capability has emerged to provide a quantified prediction of process adequacy. This ability to predict quantitatively has resulted in widespread adoption of the concept as a major element of quality planning. Process capability is the measured, inherent variation of the product turned out by a process.

Basic definitions. Each key word in this definition must itself be clearly defined because the concept of capability has an enormous extent of application, and nonscientific terms are inadequate for communication within the industrial community.

- Process refers to some unique combination of machine, tools, methods, materials, and people engaged in production. It is often feasible and illuminating to separate and quantify the effect of the variables entering this combination.
- Capability refers to an ability, based on tested performance, to achieve measurable results.
- Measured capability refers to the fact that process capability is quantified from data that, in turn, are the results of measurement of work performed by the process.
- Inherent capability refers to the product uniformity resulting from a process that is in a state of statistical control (i.e., in the absence of time-to-time "drift" or other assignable causes of variation). "Instantaneous reproducibility" is a synonym for inherent capability.
- The product is measured because product variation is the end result.

Uses of process capability information. Process capability information serves multiple purposes:

- Predicting the extent of variability that processes will exhibit. Such capability information, when provided to designers, provides important information in setting realistic specification limits.
- Choosing from among competing processes that are most appropriate to meet the tolerances.
- Planning the interrelationship of sequential processes. For example, one process may distort the precision achieved by a predecessor process, as in hardening of gear teeth. Quantifying the respective process capabilities often points the way to a solution.
- Providing a quantified basis for establishing a schedule of periodic process control checks and readjustments.
- Assigning machines to classes of work for which they are best suited.