



Statistics and Metrology

Alessandro Anzalone, Ph.D.

Hillsborough Community College, Brandon Campus



Statistics and Metrology

1. Basic Statistics
2. Probability
3. Acceptance Sampling
4. Summary



Statistical Quality Control



The concept of statistical quality control (SQC) became widely used in the late 1970s and 1980s and is the driving force behind most quality programs today. SPC deals specifically with the use of statistics to control a process, typically some type of control chart or precontrol. The term statistical quality control, or SQC, deals with the wider use of statistics to control the product from the design stages to the final product, which is shipped to the customer. SQC includes such activities as:

- Process capability studies
- SPC
- Statistical based sample inspection
- Statistical design of experiments

Basic Statistics



Descriptive statistics consists of four general areas: central tendency or what is the predominate value of the data, dispersion or the spread of the data, position of the data with respect to other data or data points, and graphic presentation.

Central Tendency. The four measures of central tendency are mean, median, mode, and midrange. We frequently call this value the "average," "mean," or "typical" value. The terms average and mean have very specific statistical definitions.

Basic Statistics



Mean. The arithmetic mean or mean \bar{x} (read as x bar”), we commonly call the average, is the summation of all the values divided by the total number of values.

This is the value you receive when you ask for your average grade or the average high or low temperature for a given month in the city or town you live in.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Basic Statistics



Median. The median, M_d (read M sub d), is the value of the data that occupies the middle of the data when arranged in ascending order. When the data set contains an odd number of data points then the value is one of the data points. When the data set contains an even number of data points then the value is the “middle” point between the two numbers. Note that the median “breaks” the data set into two subsets. The depth (number of positions from either end), or position, of the median is determined by the formula:

$$d(M_d) = \frac{n + 1}{2}$$

where 1 is the position of the smallest value, and n is the number of pieces of data.

Basic Statistics



Mode. The mode, M_o (read "M sub o"), is the value that occurs with the greatest frequency, the most common value. A mode may not exist or it may not be unique.

Midrange. The midrange is midway between the highest and lowest values. It is found by averaging the high and low values:
The midrange is the numerical value halfway between the two extreme values, the high H and low L.

$$\text{midrange} = \frac{H + L}{2}$$

The four measures of central tendency represent four different methods of describing the middle value. The four values may be the same but more likely will result in different values.

Basic Statistics



Dispersion

The three measures of dispersion are range, variance, and standard deviation. These terms assign a numerical value to the amount of spread of the data set. The variance and standard deviation are deviations from the mean.

Range. The range is the simplest measure of dispersion. It is the difference between the largest (highest) valued data (H) and the smallest (lowest) valued data (L) in the data set.

$$\text{Range} = H - L$$

Basic Statistics



Variance. The variance, s^2 , is calculated by using the following formula:

$$s^2 = \frac{\sum_{i=1}^{i=n} (x - \bar{x})^2}{n - 1}$$

where x is the individual value, \bar{x} is the mean, and n is the sample size. A simpler formula to use is:

$$s^2 = \frac{\sum_{i=1}^{i=n} x^2 - \frac{\left[\sum_{i=1}^{i=n} x \right]^2}{n}}{n - 1}$$

Basic Statistics



Standard Deviation. The standard deviation of a sample is the positive square root of the variance.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{i=n} (x - \bar{x})^2}{n - 1}}$$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{i=n} x^2 - \frac{\left[\sum_{i=1}^{i=n} x \right]^2}{n}}{n - 1}}$$

Basic Statistics



Position

Measures of position allow us to compare one piece of data to the set of data. The two most frequently used forms are *quartiles and percentiles*.

Quartiles are numbers that divide the ranked data into quarters; each set of data has four quartiles. The first quartile, Q_1 , is a number such that, at most, one fourth of the data are smaller in value than Q_1 . The second quartile is the median. The third quartile, Q_3 , is a number that, at most, 75% of the data are less than or not more than 25% of the data are greater than.

Percentiles are determined in the same manner as the quartiles except the data are ranked into 100 parts.

Basic Statistics

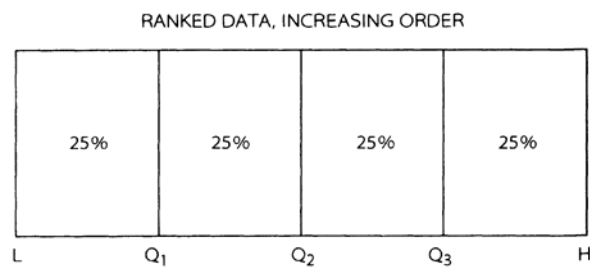


FIGURE 4-1 Quartiles

Basic Statistics



Example:

A set of five measurements was taken from a sample of five shifts. The values in millimeters are 25.06, 25.03, 25.08, 25.05, and 25.03.

Mean:

$$\begin{aligned}\bar{x} &= \frac{25.06 + 25.03 + 25.08 + 25.05 + 25.03}{5} \\ &= \frac{125.25}{5} \\ &= 25.05\end{aligned}$$

Median:

STEP 1. Rank the data in descending order.

25.08, 25.06, 25.05, 25.03, 25.03

STEP 2. Calculate the depth of median.

$$d(M_d) = \frac{5 + 1}{2} = 3$$

STEP 3. Determine the third position. This is the median.

25.05

Basic Statistics



Example:

A set of five measurements was taken from a sample of five shifts. The values in millimeters are 25.06, 25.03, 25.08, 25.05, and 25.03.

Mode:

STEP 1. Rank the data in descending order.

25.08, 25.06, 25.05, 25.03, 25.03

STEP 2. Determine the frequency of each value.

Value	Frequency
25.08	1
25.06	1
25.05	1
25.03	2

STEP 3. Determine the mode(s).

25.03

Basic Statistics



Example:

A set of five measurements was taken from a sample of five shifts. The values in millimeters are 25.06, 25.03, 25.08, 25.05, and 25.03.

Midrange:

STEP 1. Rank the data in ascending order.

25.08, 25.06, 25.05, 25.03, 25.03

STEP 2. Select the highest value, H, and the lowest value, L.

H = 25.08

L = 25.03

STEP 3. Determine the midrange.

$$\text{midrange} = \frac{25.08 + 25.03}{2}$$

$$\text{midrange} = \frac{50.11}{2}$$

$$\text{midrange} = 25.05$$

Basic Statistics



Example:

A set of five measurements was taken from a sample of five shifts. The values in millimeters are 25.06, 25.03, 25.08, 25.05, and 25.03.

Range:

STEP 1. Rank the data in ascending order.

25.08, 25.06, 25.05, 25.03, 25.03

STEP 2. Determine the highest value (H).

H = 25.08

STEP 3. Determine the lowest value (L).

L = 25.03

STEP 4. Determine the range (difference) between the highest and the lowest values.

$$\text{Range} = 25.08 - 25.03$$

$$\text{Range} = 0.05$$

Basic Statistics



Example 2:

Given a set of dimensions from a machined block, 25.6 mm, 25.3 mm, 25.8 mm, 25.8 mm, 25.5 mm, 25.5 mm, and 25.3 mm, find the variance and the standard deviation:

Basic Statistics



Example 2: Variance:

STEP 1. List the data using the following format: where
n = number, x = value, x_2 = value squared.

n	x	x^2
1	25.6	655.36
2	25.3	640.09
3	25.8	665.64
4	25.8	665.64
5	25.5	650.25
6	25.5	650.25
7	25.3	640.09

STEP 2. Find the sum of columns x and x^2 .

178.8 4567.32

These are the sum of x (Σx) and the sum of x^2 (Σx^2)
in Equation 4-6.

STEP 3. Calculate the variance using Equation 4-6.

$$\begin{aligned}s^2 &= \frac{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}{n - 1} \\&= \frac{4567.32 - \frac{31969.44}{7}}{7 - 1} \\&= \frac{4567.32 - 4567.06}{6} \\&= \frac{0.26}{6} \\&= 0.043\end{aligned}$$

Basic Statistics



Example 2:

Standard Deviation:

STEP 1: Use Equation 4-8.

$$\begin{aligned}s &= \sqrt{s^2} \\ &= \sqrt{0.043} \\ &= 0.21\end{aligned}$$



Probability



Nature of Probability

TABLE 4-1 Coin Toss

OUTCOME	FREQUENCY
0 Tails	1
1 Tail	5
2 Tails	4

TABLE 4-2 Results of 100 Trials in a Coin Toss

OUTCOME	1	2	3	4	5	6	7	8	9	10	TOTAL
0 Tails	1	2	1	4	2	2	5	3	0	4	24
1 Tail	5	5	6	5	5	7	5	4	8	4	54
2 Tails	4	3	3	1	3	1	0	3	2	2	22

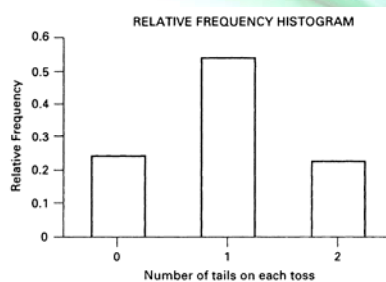


FIGURE 4-3 Relative frequency histogram of a coin toss



Probability



Probability of Events

From the previous example we can now define what is meant by probability and specifically what is meant by the “probability of an event.” The probability of an event is the relative frequency with which an event will occur. Using the previous example, the probability of tossing the coins and getting zero tails is one in four or 25%. Notice that the total for all of the events is 100%. In all cases the total of all events is 100%. The way the value of probability is expressed is in the following equation:

$$P'(\text{event}) = \frac{n(\text{event})}{n_{\text{total}}}$$

Probability Distributions



There are two types of probability distributions: discrete and continuous.

Discrete Probability Distributions. A discrete distribution is a sample of any set of possible outcomes that have a unique result. The previous example of the coin toss is an example of a discrete probability distribution. The results of the coin toss can only be one of three outcomes. The outcomes are all “discrete” or separate results. The outcome is always one of two discrete results: head or tail. There cannot be any other result; if there is, then the experiment or trial is invalid. The rolling of dice is another example of a discrete distribution because there are six possible outcomes from this distribution. In metrology the results of a go/no-go gage would represent a discrete distribution.

Probability Distributions



Binomial Probability Distribution.

$$P(D = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, 2, \dots, n$$

where p is the fraction of the population that contains a defect. Because we are working with a sample, we use the term p' (p prime) to differentiate the difference between a sample and population. To calculate p' we use

$$p' = \frac{D}{n}$$

$$\text{Mean} = \mu = np$$

$$\text{Standard Deviation} = \sigma = \sqrt{np(1 - p)}$$

Probability Distributions



Poisson Distribution. The Poisson distribution, named after French mathematician S. D. Poisson (1781–1840)

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

where λ (the Greek letter lambda) is the term for the mean or average number of defects per unit. The unique item regarding the Poisson distribution is that the mean and the variance are always the same value.

$$\lambda' = \frac{D}{n}$$

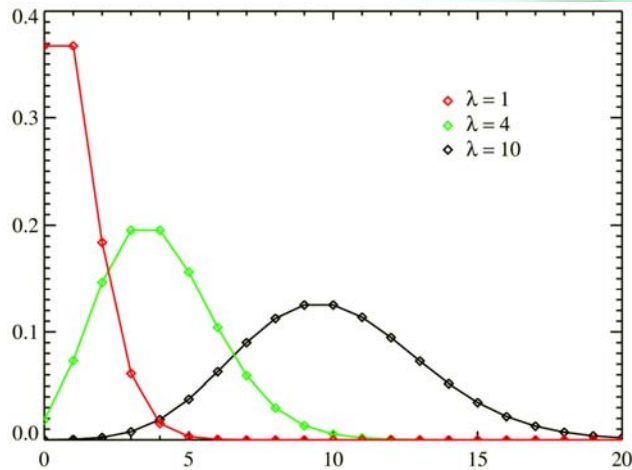
$$\text{Mean} = \mu = \lambda$$

$$\text{Standard Deviation} = \sigma = \sqrt{\lambda}$$

Probability Distributions



Poisson Distribution.



Probability Distributions



Continuous Probability Distributions.

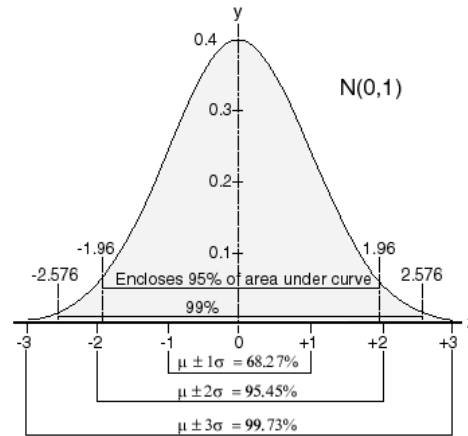
In metrology we normally work with continuous distributions. A continuous distribution is only limited by our ability to discern between two different measurements. The continuous probability distribution that is used the most in SPC and in inspection is the normal probability distribution.

Normal Probability Distribution. The Gaussian distribution or normal distribution dates back to eighteenth century Germany where the mathematician, Gauss, found that repeated measurements of the same astronomical quantity produced a pattern similar to the continuous curve in the following figure.

Probability Distributions



Normal Probability Distribution.



Probability Distributions



The formula for the Gaussian or normal curve is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]} \quad \text{for } -\infty < x < \infty$$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{i=n} x^2 - \frac{\left[\sum_{i=1}^{i=n} x\right]^2}{n}}{n}}$$

Acceptance Sampling



Sampling and Sample Variability.

Because we normally select a sample of the parts we are manufacturing, we need to understand a little more about sampling and sampling variability. When we did the initial coin toss we obtained the results shown in Table 4-1. When we did the coin toss a second time we did not get the exact same results of the first; in fact, when we did the coin toss a total of ten times (see Table 4-2), we obtained the same result only on outcomes 2 and 5. In other words, 9 of the 10 out- comes were different. The difference is due to sample variability. This is why it is so important to obtain a random or representative sample of the work being performed.

TABLE 4-1 Coin Toss

OUTCOME	FREQUENCY
0 Tails	1
1 Tail	5
2 Tails	4

TABLE 4-2 Results of 100 Trials in a Coin Toss

OUTCOME	1	2	3	4	5	6	7	8	9	10	TOTAL
0 Tails	1	2	1	4	2	2	5	3	0	4	24
1 Tail	5	5	6	5	5	7	5	4	8	4	54
2 Tails	4	3	3	1	3	1	0	3	2	2	22

Acceptance Sampling



Acceptance Sampling is used to determine if the material produced is acceptable for use. Because this process is conducted after the material is produced, any unacceptable material would have to be reworked or scrapped. There is a major difference when comparing acceptance sampling with statistical process control. SPC allows the process to be controlled so all material is acceptable.

The advantages of acceptance sampling are primarily economical because of the following:

1. Only part of the material needs to be inspected.
2. There is less handling damage during inspection.
3. Fewer inspectors are needed.

Acceptance Sampling



The second advantage is motivational:

1. Removal of the monotonous 100% inspection.
2. Applicability to destructive inspection or testing.
3. The lot or batch is rejected; therefore, stronger motivation is placed on the supplier or manufacturer to improve.

The **disadvantages** are:

Accepting "bad" lots and rejecting "good" lots of material.

There is additional planning and documentation needed.

The sample usually provides less information than 100% inspection.

Acceptance Sampling



When a sampling plan is being developed the first consideration is to determine if the sampling plan is to accept or reject the lot of material immediately being evaluated, or to determine which of the lot of material produced is acceptable.

Type A sampling plans specify the quality level for each lot in terms of percent defective at a given risk of being accepted by the customer. They are known as lot tolerance percent defective plans or LTPD plans.

Type B sampling plans specify the quality level in terms of percentage of accepting the submitted lots at a given quality level. They are known as acceptable quality levels or AQL plans.

Both LTPD and AQL plans provide lot by lot protection. A second type of protection is the limiting average percentage of defective items. These types of plans are known as average outgoing quality limit or AOQL plans.

Acceptance Sampling



Key Definitions in Sampling Plans

Producer's Risk. The producer's risk is the probability that a "good" lot will be rejected by the sampling plan. The producer's risk traditionally will vary from 0.001 to 0.10, depending on the individual sampling plan. The risk is stated in conjunction with the maximum quality level that the plan will accept. This is normally called the AQL.

Consumer's Risk. The consumer's risk is the probability that a "bad" lot will be accepted by the sampling plan. The risk is stated in conjunction with the definition of rejectable quality such as lot tolerance percent defective. Traditionally, only 0.10 is used for the consumer's risk.

Acceptance Sampling



Key Definitions in Sampling Plans

Operating Characteristic Curve. The operating characteristic (OC) curve is a graph of the lot fraction defective versus the probability of accepting the lot.

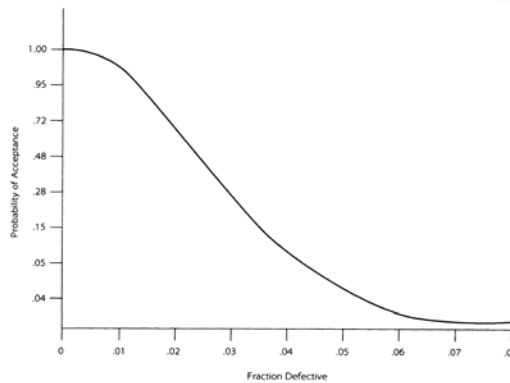


FIGURE 4-6 Operating characteristic (OC) curve for sampling inspection plan with $n = 100$, $c = 2$.

Acceptance Sampling



Key Definitions in Sampling Plans

Average Outgoing Quality Limit. We stated that the AQL and LTPD are two common quality indices for sampling plans. A third quality index is the AOQL, which is the worst case of quality level a consumer would receive when an inspection plan is in effect and all rejected lots have been 100% inspected with the removal of defective material.

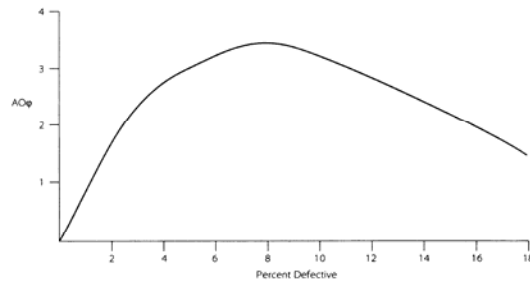


FIGURE 4-7 Average outgoing quality limit (AOQL) curve for 2.5% average quality level (AQL), normal inspection

Acceptance Sampling



Key Definitions in Sampling Plans

Average Total Inspection. The average total inspection (ATI) is the average inspection required when 100% inspection of rejected lots is included.

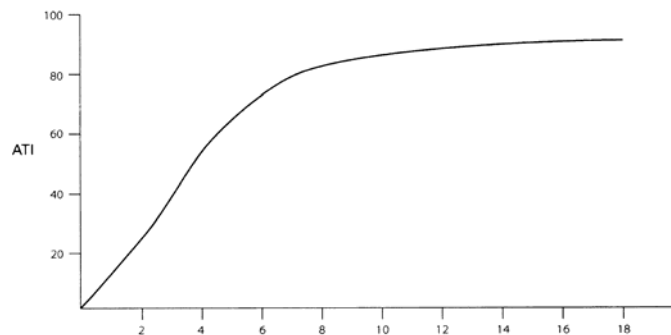


FIGURE 4-8 Average total inspection (ATI) for 2.5% average quality level (AQL), normal inspection

Sampling Plans



TABLE 4-3 Sampling Plans and Uses

TYPE OF SAMPLING PLAN	TYPICAL USE OF SAMPLING PLAN
Single	General use
Double	General use with known good quality
Multiple	General use with known excellent quality
Sequential	General use with optimal sample size
Bulk	Bulk material sampling
Rectification	No lot can exceed a given percent defective
Continuous	Used when no natural lot occurs
Compliance	Maximizes consumer protection
Reliability	Controls reliability of product

Sampling Plans



ANSI/ASQC Z1 .4. Sampling Procedures and Tables for **Inspection by Attributes**, Z1.4 (MIL-STD-105), provides items to consider.

1. A choice of 26 AQL values ranging from 0.010 to 1000.0 (values of 10 or less can be considered as percent defective or defects per hundred units; values of 10 or greater must be considered as defects per hundred units).
2. The probability of accepting at AQL quality varies from 89 to 99.5%.
3. Seven inspection levels, 3 general and 4 special levels.
4. Defect classification as critical, major, or minor.
5. The customer may specify separate AQLs for each class or specify an AQL for each kind of defect that a product may show.

Sampling Plans



You receive 450 pieces of a new machined part. What is the correct sample size to determine the acceptance of the part using a go/no-go snap gage?

Step 1. Using Table 4-4, determine the sample code.

281 to 500—general inspection level II is H

Step 2. Using Table 4-5, determine the sampling size.

H = 50

Therefore, the sample size would be 50 units for attribute inspection.

Sampling Plans



TABLE 4-4 Sample Size Codes

Lot Size	SPECIAL INSPECTION LEVEL				GENERAL INSPECTION		
	S1	S2	S3	S4	I	II	III
1-8	A	A	A	A	A	A	B
9-15	A	A	A	A	A	B	C
16-25	A	A	B	B	B	C	D
26-50	A	B	B	C	C	D	E
51-90	B	B	C	C	C	F	F
91-150	B	B	C	D	D	F	G
151-280	B	C	D	E	E	G	H
281-500	B	C	D	E	F	H	J
501-1200	C	C	E	F	G	J	K
1201-3200	C	D	E	G	H	K	L
3201-10000	C	D	F	G	J	L	M
10001-35000	C	D	F	H	K	M	N
35001-150000	D	E	G	J	L	N	P
150001-500000	D	E	G	J	M	P	Q
500001-OVER	D	E	H	K	N	Q	R

Sampling Plans



TABLE 4-5 ANSI/ASQC Z1.4 contains 15 sampling plans. The plan shown here is the most frequently used. In these tables, Ac is the acceptable number and Re is the rejection number. When the sample size (by code letter, see Table 4-4) and the desired acceptable quality level (AQL) are known, the table may be used to find the sample size and acceptable number. When the arrow is pointing downward, the first sampling plan (size and acceptance number) below the arrow is used. When pointed upward, the first sampling plan above the arrow is used. For example, for an AQL of 1.0, the smallest sample size is 13 units. Conversely, with an AQL of 15 defectives per hundred parts, the largest sample required is 80 parts.

Single Sampling Plans for Normal Inspection (Master Table)																																						
Sample Size Code Letter	Sample Size	Acceptable Quality Levels (Normal Inspection)																																				
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000											
A	2																																					
B	3																																					
C	5																																					
D	8																																					
E	13																																					
F	20																																					
G	32																																					
H	50																																					
J	80																																					
K	125																																					
L	200																																					
M	315																																					
N	500																																					
P	800																																					
Q	1250																																					
R	2000																																					

- = Use first sampling plan below arrow. If sample size equals or exceeds lot or batch size, do 100 percent inspection.
- = Use first sampling plan above arrow.
- = Acceptance number.
- = Rejection number.

Sampling Plans

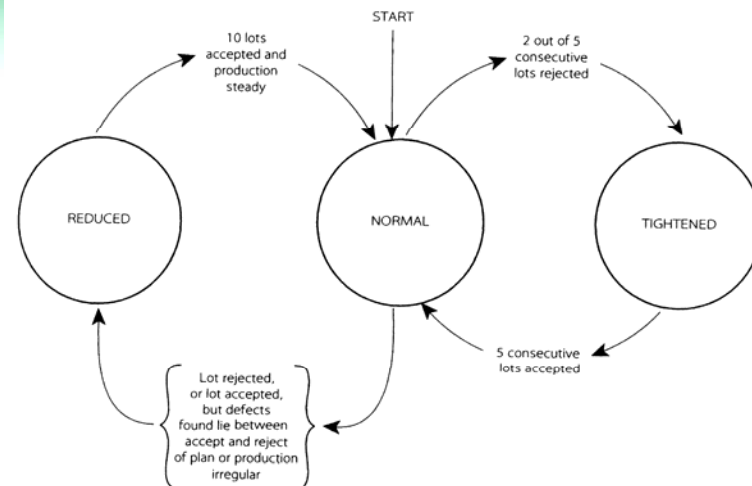


FIGURE 4-9 Switching rules for ANSI/ASQC Z1.4

Sampling Plans



You receive 450 pieces of a new machined part. What is the correct sample size to determine the acceptance of the part using a go/no-go snap gage?

Assume an engineer determines the product can have an AQL of 0.65% and the material is received in 1,500-piece lots. What would the inspection sample size and accept-reject level be?

By using Table 4-4 we would determine that the correct sample code would be "K," because the lot size is between 1,201 and 3,200 units for level II inspection. Then going to Table 4-5, we would find sample size code letter "K" and determine the sample size would be 125 units. Then going to the 0.65 column, we would find the accept-reject values of accept on 2 and reject on 3.

The inspector would obtain a 125-piece random sample of material and do an attribute inspection of all pieces. After determining the number of rejectable units, and if the number of rejectable units is 2 or less, then the lot would be accepted.



Sampling Plans



ANSI/ASQC Z1.9. Sampling Procedures and Tables for **Inspection by Variables** for Percent Nonconforming, ANSI/ASQC Z1.9 (MIL-STD-414), is an AQL type of sampling plan that assumes the individual measurements being taken are normally distributed. Series errors can be introduced if this is not the case. The sampling plan allows for the use of three different measures of variability:

- Known population standard deviation, μ
- Sample standard deviation, s
- The average range

The choice between μ , s , and average range is a matter of economics. The average range method requires a larger sample size but is easier to use. The OC curves are based on the sample standard deviation, s , while the population standard deviation, σ , and average range are matched as closely as possible to those using s .



Summary



The concept of statistical quality control was introduced in the 1940s at Bell Labs by Dr. W. A. Shewhart. Today, the concept of statistical process control (SPC) is a requirement for almost any business to be successful.

One main area of statistics is descriptive statistics. Here we are describing the data we have. In describing data we need two key components: the central tendency and the dispersion. The mean, median, and mode are all terms for the central tendency. The variance, standard deviation, and range are names for the dispersion.

Probability plays a key part in SPC. It allows us to predict what a population will look like. Probability is a key component of sampling.

By tying the concepts of probability and descriptive statistics together we can develop probability distributions. These are mathematical models of the distributions. There are two types of probability distributions: attribute and variable distribution. Both play key roles in SPC.

Acceptance sampling plans are used to predict the acceptability of lots of material without having to inspect the complete lot. There are several types of sampling plans but they fall into the two general groups of attribute acceptance sampling and variable acceptance sampling. The most widely used plans are ANSI/ASQC Z1.4 and ANSI/ASQC Z1.9 or their military equivalents of MIL-STD-105 and MIL-STD-414.

References



<http://www.thefreedictionary.com/statistics>
<http://www-math.bgsu.edu/~albert/m115/probability/outline.html>
<http://mathforum.org/dr.math/faq/faq.prob.intro.html>
<http://www.itl.nist.gov/div898/handbook/>