Probability Distributions

- A random variable, X, is a numerical description of the outcome of an experiment. Formally, a random variable is a function that assigns a numerical value to every possible outcome in a sample space.
- A probability distribution, f(x), is a characterization of the possible values that a random variable may assume along with the probability of assuming these values.
- The cumulative distribution function, F(x), specifies the probability that the random variable X will assume a value less than or equal to a specified value, x, denoted as $P(X \leq x)$.

Important Probability Distributions

- Discrete
  - Binomial
  - Poisson
- Continuous
  - Normal
  - Exponential

2

Binomial Distribution

- The binomial distribution describes the probability of obtaining exactly x "successes" in a sequence of n identical experiments, called trials.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \ldots, n \qquad (6.3)$$

$$E[X] = \mu = np \qquad (6.4)$$

$$\sigma^2 = np(1-p) \qquad (6.5)$$

$$\sigma = \sqrt{np(1-p)} \qquad (6.6)$$

3

EXAMPLE 6.4 Using the Binomial Distribution

If the probability that a process produces a defective part is 0.2, then the probability distribution that x parts out of a sample of 10 will be defective is described using formula (6.3) with $n = 10$ and $p = 0.2$:

$$f(x) = \begin{cases} \binom{10}{x} (0.2)^x (0.8)^{(10-x)} & \text{for } x = 0, 1, 2, \ldots 10 \\ 0, \textit{otherwise} \end{cases}$$

Thus, to find the probability that 3 parts among a sample of 10 will be defective, we compute

$$f(3) = \binom{10}{3} (0.2)^3 (0.8)^{10-3} = \frac{10!}{3!7!}(0.008)(0.2097152)$$

$$= 120(0.008)(0.2097152) = 0.20133$$

4

Poisson Distribution

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{6.7}$$

λ = expected value or average number of occurrences,
x = 0, 1, 2, 3, … , and e = 2.71828…

5

Continuous Probability Distributions

- A curve that characterizes outcomes of a continuous random variable is called a probability density function, and is described by a mathematical function *f(x)*.
- Probabilities are only defined over intervals.
- The cumulative distribution function, F(x), represents the probability $P(X \leq x)$.

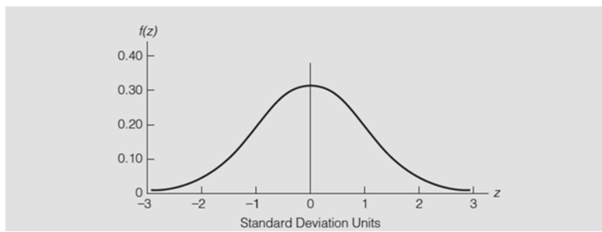
$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) \tag{6.8}$$

6

Normal Distribution

- Familiar bell-shaped curve.
- If a normal random variable has a mean μ = 0 and a standard deviation σ = 1, it is called a standard normal distribution, represented by z.

FIGURE 6.5
Standard Normal Distribution



7

Calculating Normal Probabilities

- If x is any value from a normal distribution with mean μ and standard deviation σ, we may easily convert it to an equivalent value from a standard normal distribution using:

$$z = \frac{x - \mu}{\sigma} \qquad (6.11)$$

8

Statistical Quality Control

Eugene Lodewick Grant, Richard S. Leavenworth

7th edition

New York: McGraw-Hill, 1996

The 7 M's

- MEN
- MACHINES
- MATERIALS
- METHODS
- MANAGEMENT
- MONEY
- MOTHER NATURE

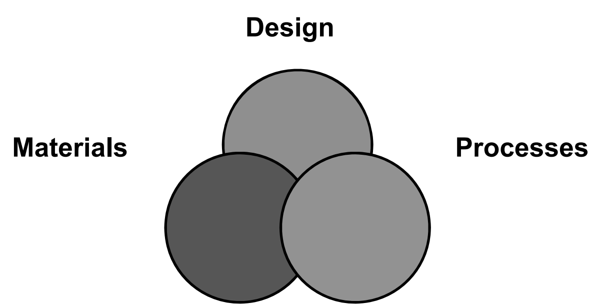
Variation in Manufacturing

Undesirable variation due to:

–Internal processes

–Supplier variation

–Unreasonably tight specifications



Consequences of variation

- Inspection needs (routine burden)
- Reject, rework, downgrade, recycle, scrap
  (internal burden)
- Return, repair and warranty costs (external burden)
- Field repairs (present loss)
- Lost customers (future loss)
- Late design changes (short term loss)
- Damaged business reputation (long term loss)

The Need for SQC

Quantification – Wastes Understood

Detection – Wastes Tolerated

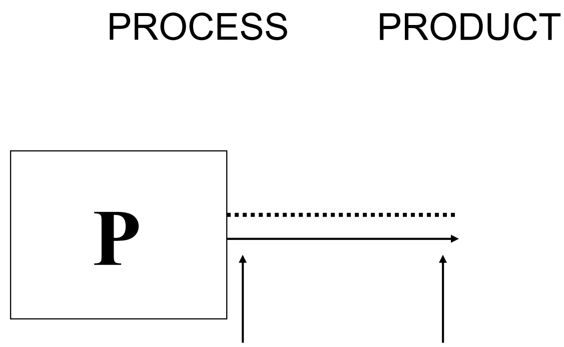
Prevention – Wastes Avoided

Methods of SQC

- Management of variation
  - Quantification (Analysis)
  - Detection (Decision)
  - Prevention (Action)

- Medium of information
  - Collection
  - Analysis
  - Interpretation

Information driven improvement:
How is SQC applied?



Statistical Approach to Problem Solving

THE PROBLEM SOLVING PROCESS

a. Detect problems
b. Narrow problem area
c. Assess factors which seem to cause the problem
d. Determine whether the assumed causes of defects are true or not
e. Prevent errors due to omission, haste or carelessness
f. Confirm the effects of improvement
g. Detect outliers

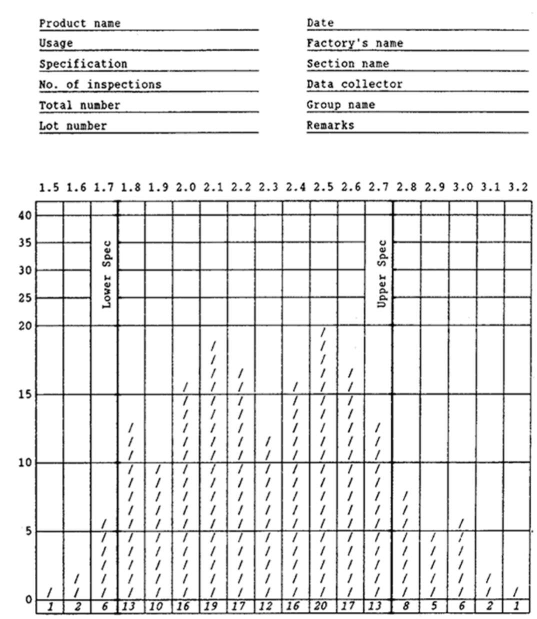
SEVEN COMMON PROBLEM-SOLVING TOOLS IN QUALITY CONTROL

- CHECK SHEET
- HISTOGRAM
- PARETO DIAGRAM
- CAUSE-AND-EFFECT DIAGRAM
- STRATIFICATION
- SCATTER DIAGRAM
- GRAPH/CONTROL CHART

THE SEVEN TOOLS

	TOOL	*FUNCTION*
1.	CHECK SHEET	Easy data collection
2.	HISTOGRAM	Grasp shape of distribution of quality characteristic
3.	PARETO DIAGRAM	Narrow problem area
4.	CAUSE & EFFECT DIAGRAM	Assess factors which seem to cause quality characteristic
5.	STRATIFICATION	Verify effect of assumed discrete causes on quality characteristic
6.	SCATTER DIAGRAM	Verify effects of assumed continuous causes on quality characteristic
7.	GRAPH/CONTROL	Grasp dynamic change and confirm CHART observation of standards

CHECK SHEET



CHECK SHEET

Product: ______________________ Date: ______________________

Manufacturing stage: *final insp.* Factory: ______________________

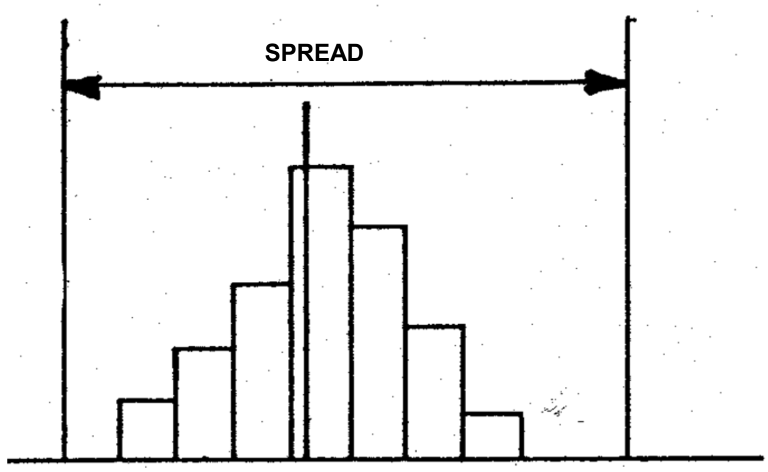
Type of defect: *scars, cracks, incomplete, dirt* Section: ______________________

Inspector's name: ______________________

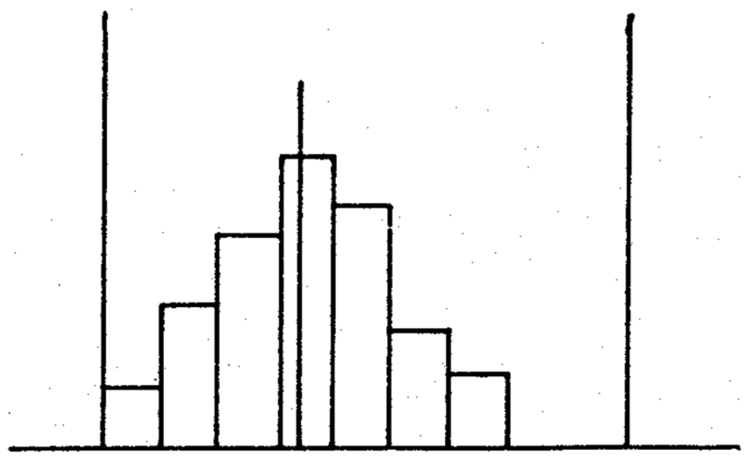
Total no. inspected: *2530* Lot no: ______________________

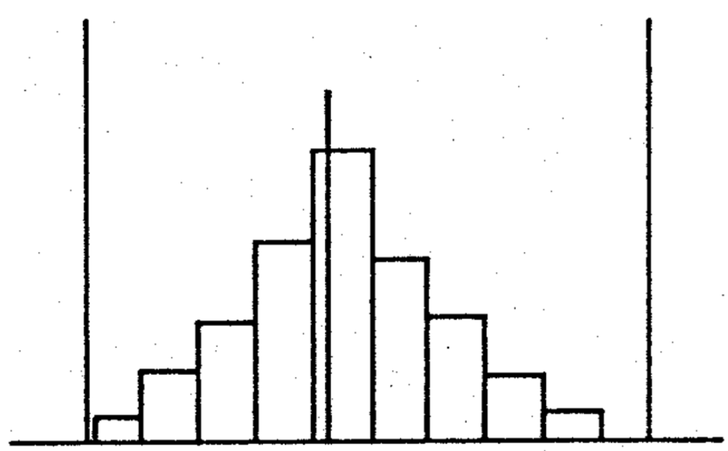
Remarks: *all items inspected* Order no: ______________________

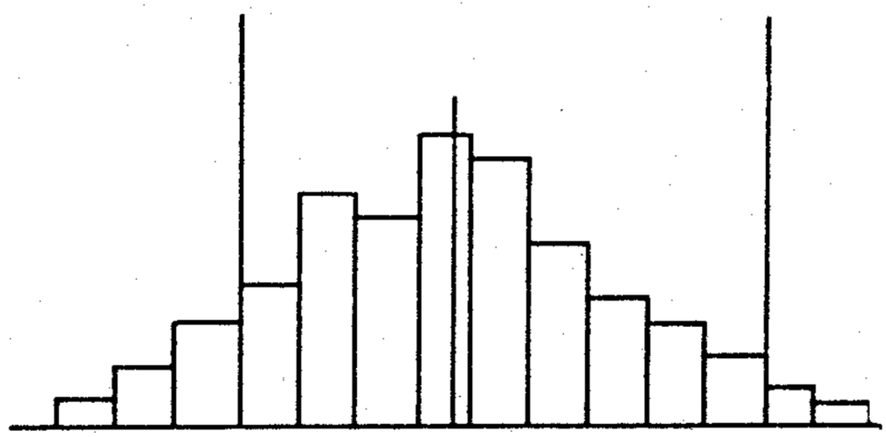
Type	Check	Sub-total
Surface scars	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ //	32
Cracks	~~////~~ ~~////~~ ~~////~~ ~~////~~ ///	23
Incomplete	~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ~~////~~ ///	48
Dirt	////	4
Others	~~////~~ ///	8
	Grand total:	115

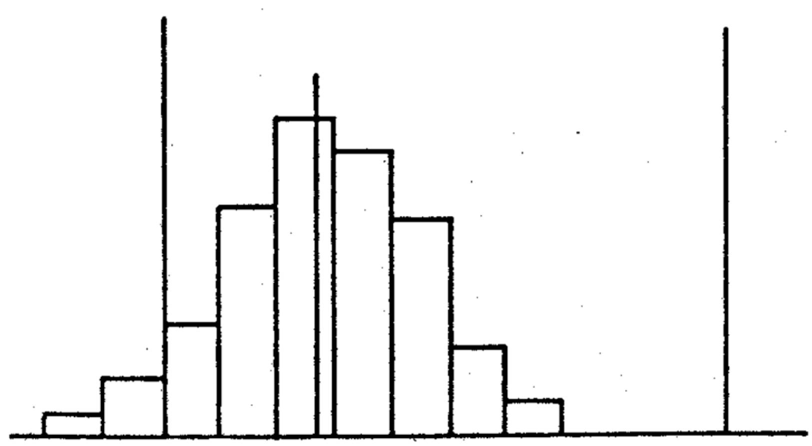


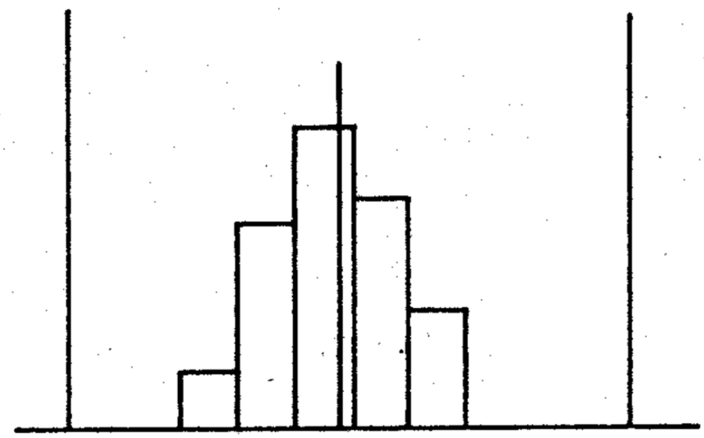
LOCATION

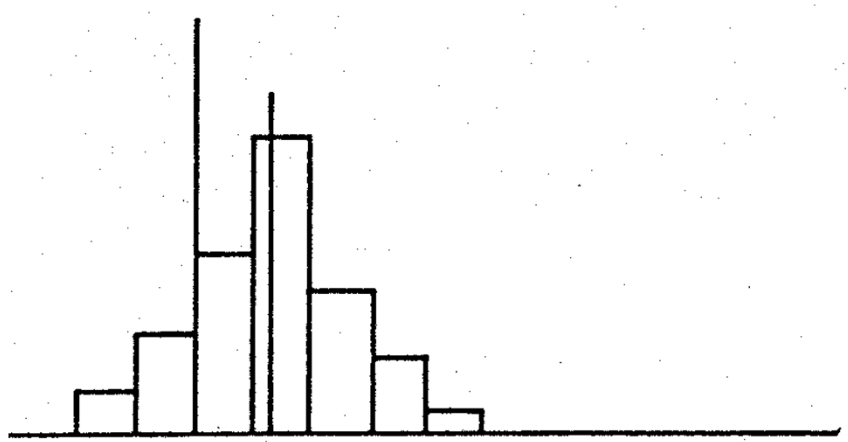


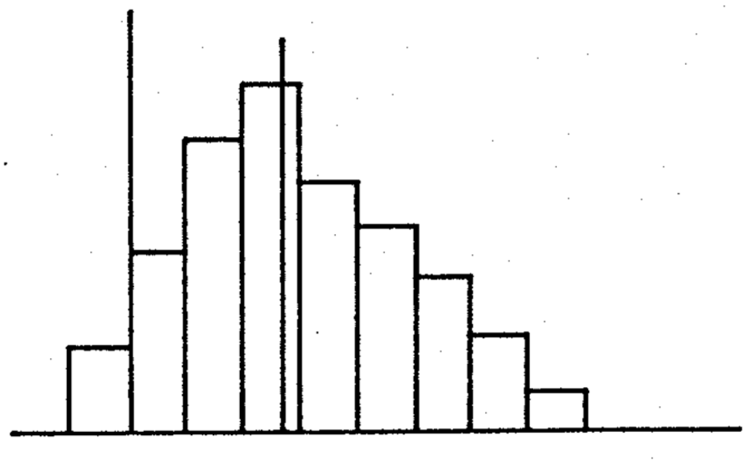


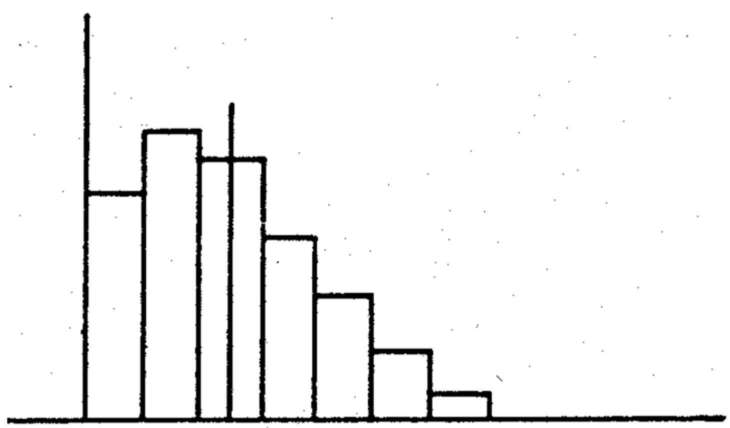


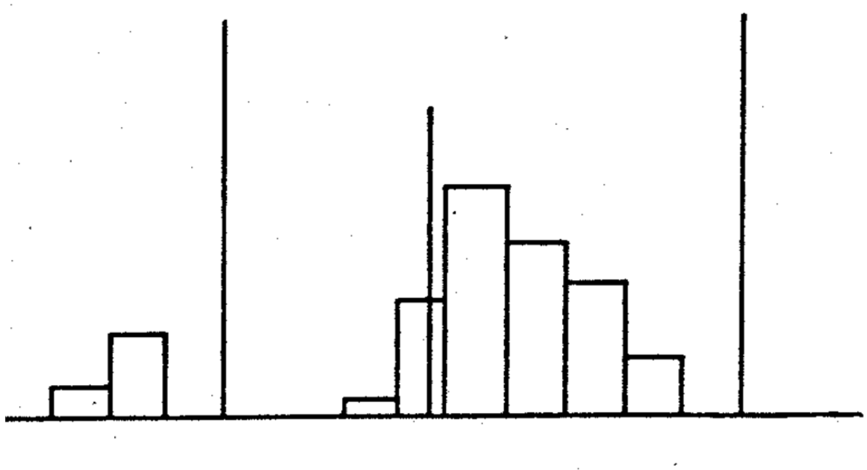












Central Tendency/Position

- Mean
- Median
- Mode

Variability/Spread

- Range
- Standard Deviation
- Variance

Pareto Analysis

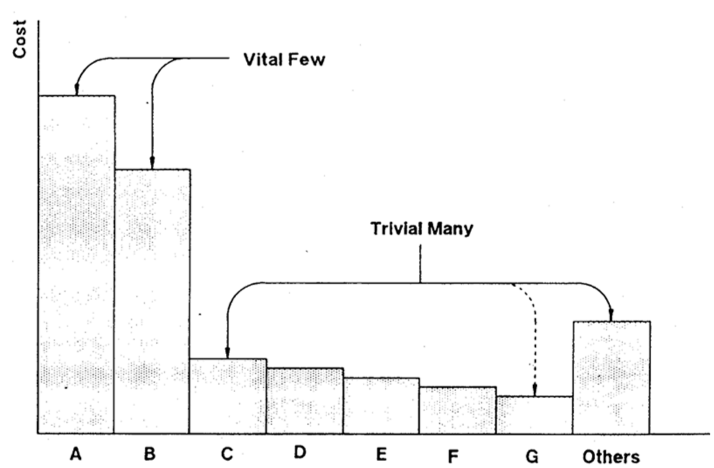
- Origin
- Principles
- Applications
- Pitfalls

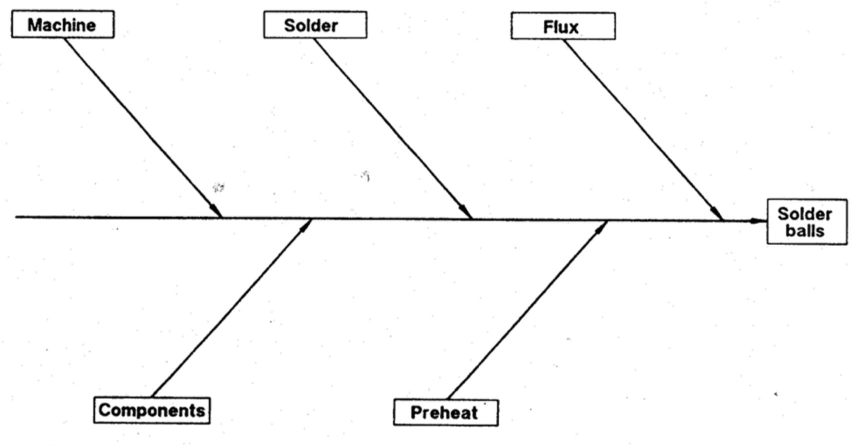
PARETO ANALYSIS

1. List the elements
2. Measure the elements
3. Rank the elements
4. Create Cumulative Distributions
5. Draw the Pareto Curve
6. Interpret the Pareto Curve

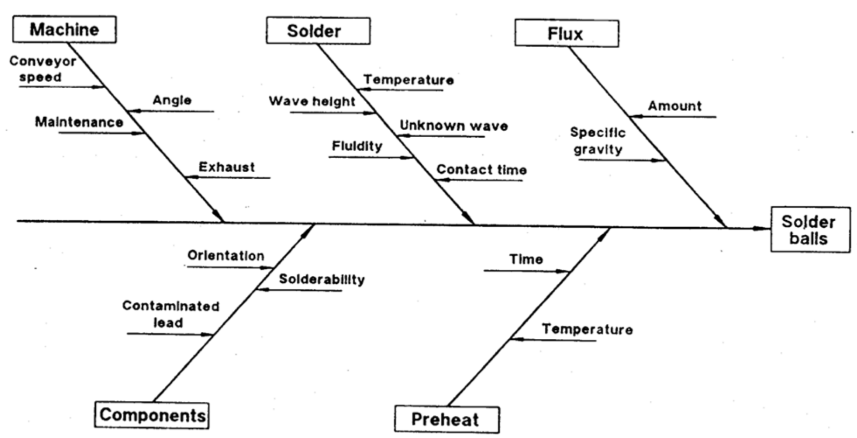
PARETO ANALYSIS

The technique identifies the ‘Vital Few’ high-priority items from the ‘Trivial Many’ in the problem-solving process.

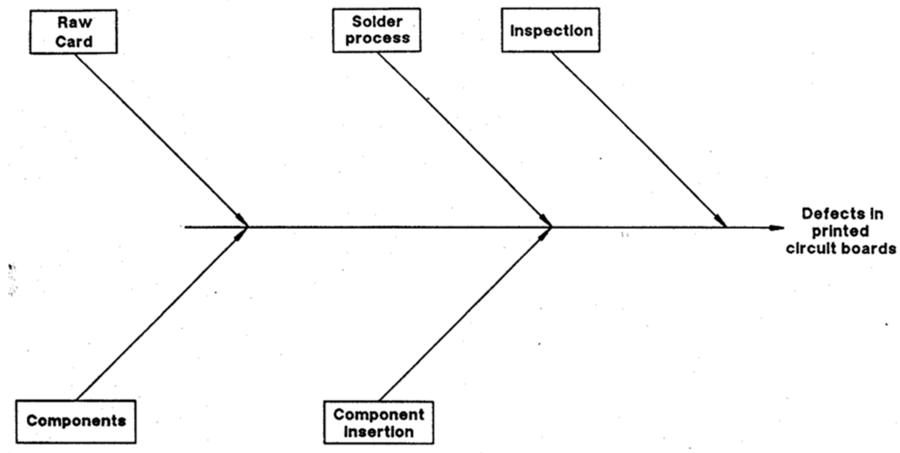




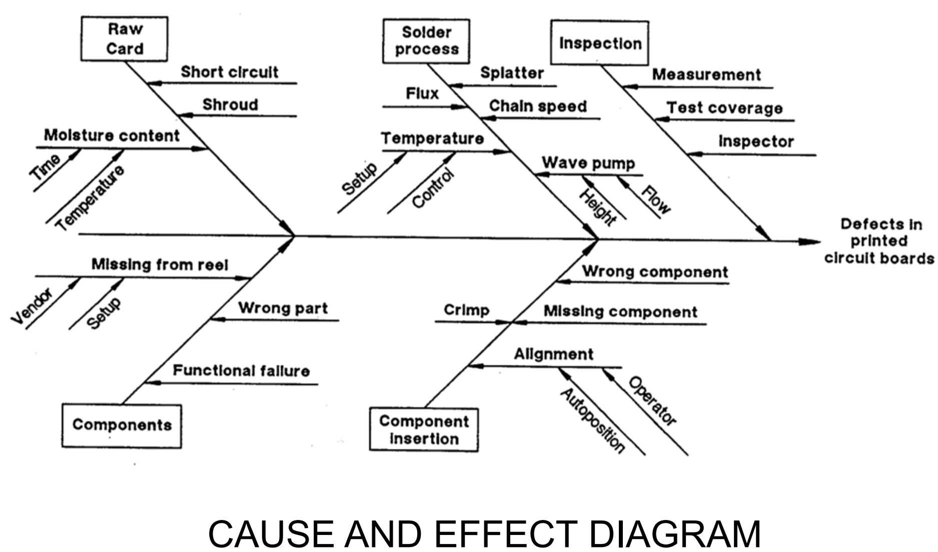
CAUSE-AND-EFFECT DIAGRAM



CAUSE-AND-EFFECT DIAGRAM



CAUSE-AND-EFFECT DIAGRAM

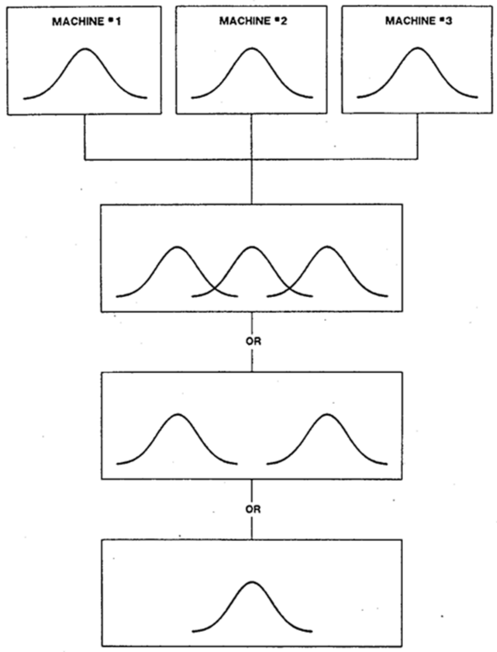


CAUSE AND EFFECT DIAGRAM

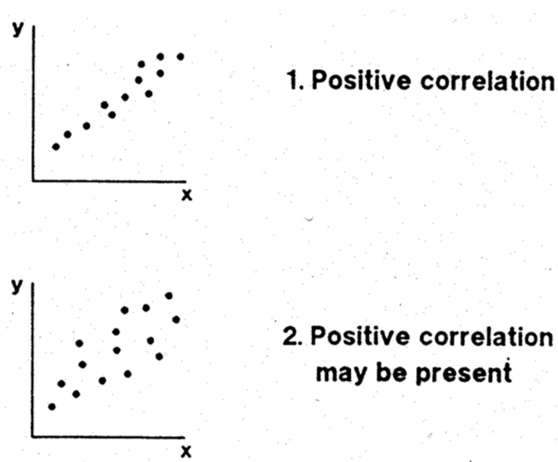
- Cause-and-Effect Diagram
- Fishbone Diagram
- Ishikawa Diagram

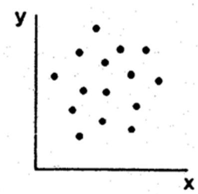
Application Example:

Large observed variation

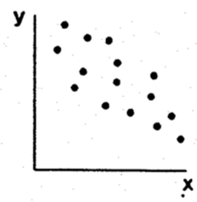


SCATTER DIAGRAM

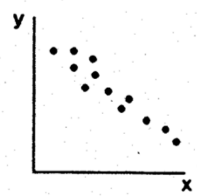




3. No correlation



4. Negative correlation may be present



5. Negative correlation

CORRELATION

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2\right]}}$$

$$r > \frac{2}{\sqrt{n}}$$

Coefficient of Correlation

Interpret the following results in a correlation analysis:

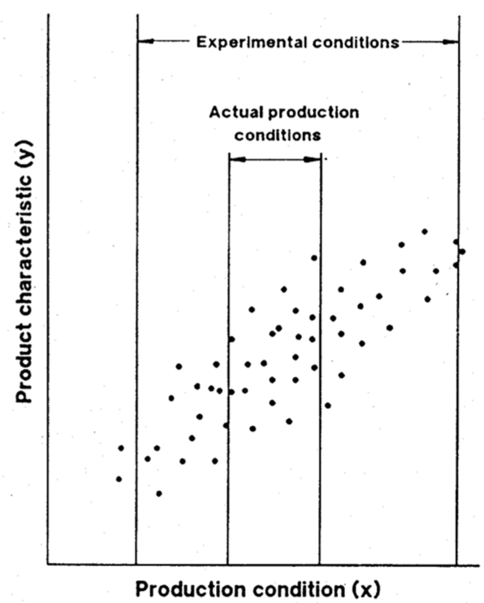
a. $r = 0.27$, $n = 100$

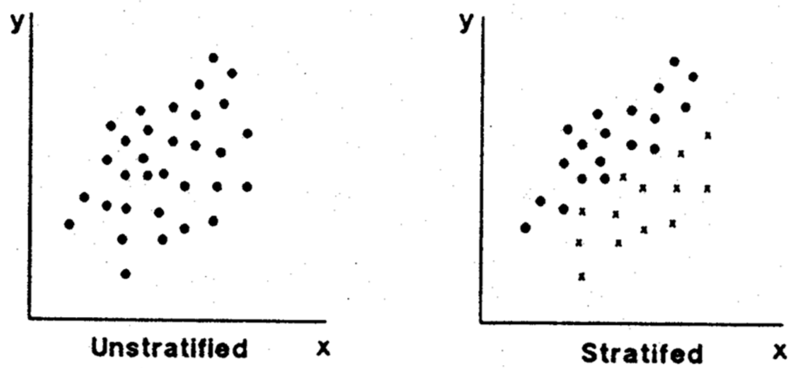
b. $r = 0.27$, $n = 16$

c. $r = -0.75$, $n = 16$

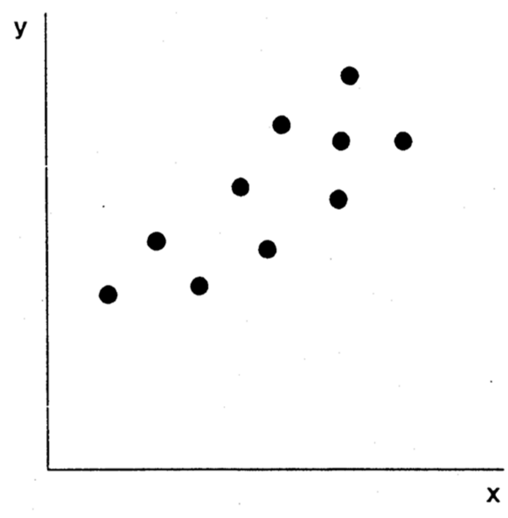
d. $r = 0.02$, $n = 100$

STRATIFICATION IN A SCATTER DIAGRAM

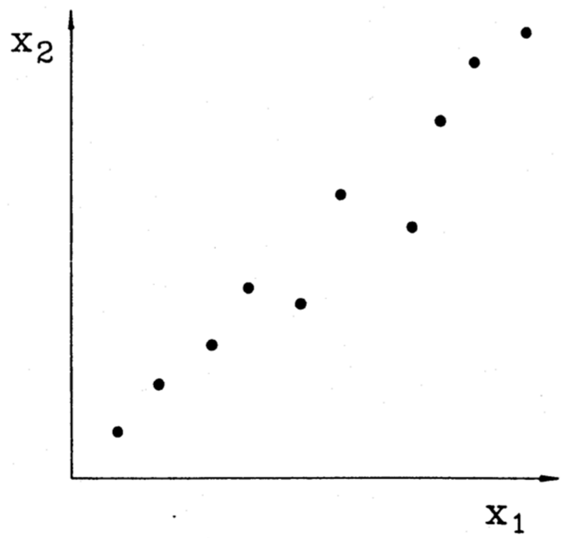




ASSOCIATION
CAUSATION



REPRODUCIBILITY
REPEATABILITY



SQC TEST I
(24 Questions on Data Analysis)

1. Accuracy is getting an unbiased value in a measurement.
2. Precision is getting consistent results repeatedly.
3. Random selection of a sample assures that the sample average will equal the population average.
4. Once randomness in sampling is truly achieved, correct statistical conclusions can be guaranteed.
5. It is more important to have a random sample than to have a large sample size.
6. Quality control has been labeled as the science and art of identifying and controlling variability. One measure of variability as used in this context is the variance.
7. It is possible to have a measure of variability that is not dependent on the exact value of every measurement.

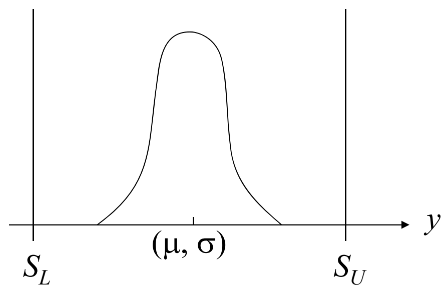
8. Histograms are used to identify major variables.
9. Histograms involve percentage plots of categories.
10. Important factors can usually be identified through a Pareto diagram.
11. Pareto analysis is a method of identifying systematic variations.
12. Pareto analysis involves cause-and-effect investigations.
13. Positive correlation is an indication of the cause-and-effect relationship between two variables.
14. Negative correlation is an indication of absence of any cause-and-effect relationship between two variables.
15. A non-uniform pattern is called a random pattern in statistical analysis.
16. Regression analysis should not be used when there is negative correlation between two variables.
17. Regression analysis can express non-linear relationships between two variables.
18. Coefficient of correlation can be expressed in the form of a percentage.
19. Generally, the larger the sample size, the higher is the confidence one can have in the results of a statistical analysis.
20. A large sample size can compensate for the effect of biasedness in a sample.
21. Statistics enables us to make a general statement based on what we know of a particular sample.
22. A scatter diagram is a form of descriptive statistics.
23. A Pareto diagram is a form of histogram.
24. A Pareto diagram is a form of stratification.

Process Capability Analysis

Static and Dynamic Assessments

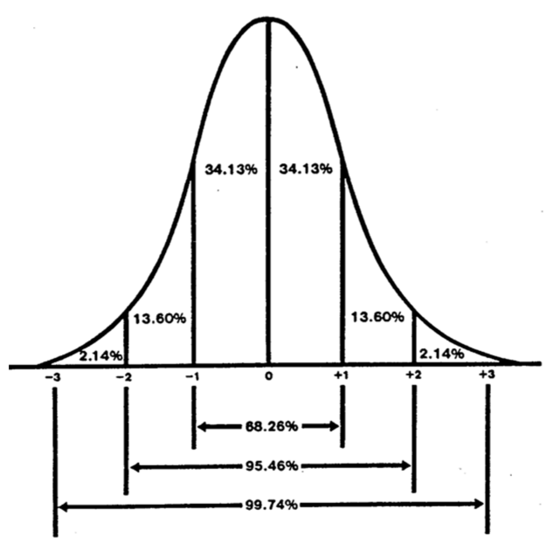
of Variation

Static Analysis of a Process



Process capability analysis

THE NORMAL DISTRIBUTION CURVE
APPROXIMATE AREAS UNDER THE CURVE

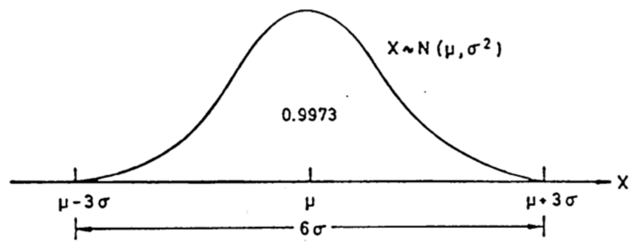


PROCESS CAPABILITY

Reflects the extent of natural, random variations of a stable process.

PROCESS CAPABILITY

Can be defined mathematically as the interval of 99.73% of measurements of a product characteristic under the influence of only random variations; for normally distributed measurements, this interval is six standard deviations (6σ).



Quick Process Capability Estimation

$$n = 10$$

$$\begin{aligned} 6\sigma &= 6\frac{R}{d_2} \\ &= \frac{6R}{3.078} \\ &= 2R \end{aligned}$$

Example:	0.826	0.837
	0.833	0.825
	0.835	0.833
	0.842	0.830
	0.840	0.836

Capability = 2 (0.842 − 0.825)
= 0.034

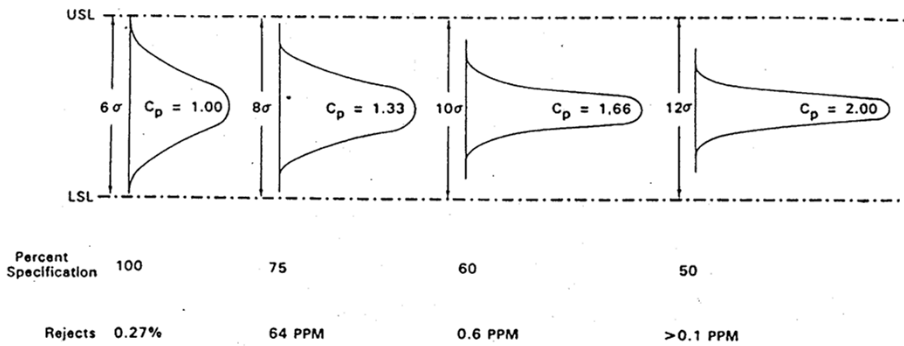
PROCESS CAPABILITY

- Is meaningful only when the process is stable.
- Is stated in absolute measurement units, e.g. ohms, millimeters.
- Does NOT involve any product specifications.

PROCESS CAPABILITY INDEX

$$PCI = \frac{\text{Allowed Variation}}{\text{Actual Variation}}$$

PROCESS CAPABILITY INDICES



C_p indices for varying widths of the process distribution

C_p	Defective ppm (process fallout)	
	One-sided specifications	Two-sided specifications
0.80	8,198	16,395
0.90	3,467	6,934
1.00	1,350	2,700
1.10	484	967
1.20	159	318
1.30	48	96
1.40	14	27
1.50	4	7
1.60	1	2
1.70	0.17	0.34
1.80	0.03	0.06
2.00	0.0009	0.0018

Defective ppm corresponding to a given C_p value

From C_p to C_{pk}

- All the above assume that the process is centered, i.e. the mean is situated at the mid-point of the specification range.
- What if the mean is closer to one of the specifications?

$$c_p = \frac{USL - LSL}{6\sigma} \qquad \hat{c}_p = \frac{USL - LSL}{6s}$$

$$C_{pk} = \min\left\{\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\}$$

$$\hat{C}_{pk} = \min\left\{\frac{USL - \overline{X}}{3s}, \frac{\overline{X} - LSL}{3s}\right\}$$

Questions:

•Why is C_p sometimes preferred to C_{pk}?

•What are the C_{pk} values of

$\mu_A = 15 \quad \sigma_A = 1.25$

$\mu_B = 18 \quad \sigma_B = 0.5$

$\mu_C = 15.5 \quad \sigma_c = 1.125$

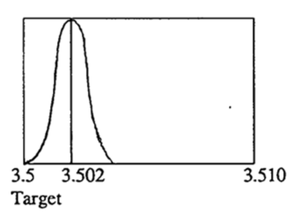
for a product with LSL = 10, USL = 20, and a target value of 15?

(A : μ ; B : μ ; C : μ and σ if necessary)

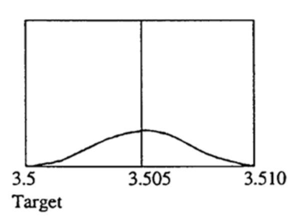
A limitation of C_{pk}

The C_{pk} index is NOT able to discriminate between a good process condition and a poor one in the case of unilateral tolerance, as illustrated here:

$$x = 3.5 + 0.010/-0.000$$



$$C_{pk} = \frac{0.002}{0.002} = 1.0$$



$$C_{pk} = \frac{0.005}{0.005} = 1.0$$

Note: $C_p \geq C_{pk}$

C_p : Process capability index; process potential

C_{pk} : Process performance

C_r : Process capability ratio

k : Off-centre index

Derivation:

$$C_p(1-k) = \frac{S_U - S_L}{6\sigma}\left[1 - \frac{\frac{1}{2}(S_U + S_L) - \mu}{\frac{1}{2}(S_U - S_L)}\right]$$
$$= \frac{1}{6\sigma}\left[S_U - S_L - (S_U + S_L) + 2\mu\right]$$
$$= \frac{2\mu - 2S_L}{6\sigma}$$
$$= C_{pk}$$

Summary:

$$C_p = \frac{S_U - S_L}{6\sigma}$$
$$C_{pk} = min\left\{\frac{S_U - \mu}{3\sigma}, \frac{\mu - S_L}{3\sigma}\right\}$$
$$C_{pk} = C_p(1-k)$$
$$C_r = \frac{1}{C_p}$$

where $k = \left|\frac{m-\mu}{1/2(S_U - S_L)}\right|$

and $m = 1/2(S_U + S_L)$

Important points:

PROCESS CAPABILITY INDICES relate
process capability to the specifications.

C_p shows whether specifications CAN be met.

C_{pk} shows whether specifications ARE being met.

The larger the process capability index, the lower is the percentage of product not conforming to specifications.

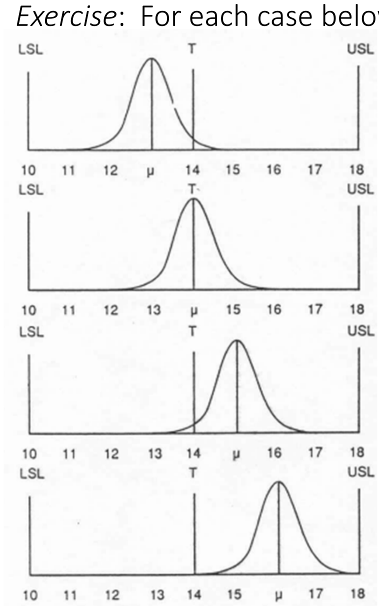
SOME USES OF PROCESS CAPABILITY STUDY

1. Predicting whether design tolerances can be met
2. Assigning equipment to production
3. Evaluating new equipment purchases
4. Estimating fraction defective to be expected
5. Making adjustments during manufacture
6. Setting specifications
7. Costing out contracts

Some REMEDIAL STEPS when a
poor Process Capability Index is discovered

1. Center process at mid-specs
2. Review specifications
3. Apparent ~ True process capability study
4. Change variance (process)
5. 100% inspection

Exercise: For each case below, calculate the C_p and C_{pk}



Further Applications

Process capability index and "Sigma level"

Sigma Level of a Process

Margin is needed to counter:

- piece-to-piece (component) variations
- manufacturing process variations
- deterioration (operating/shelf life) variations
- environmental variations
  (refer to "noise classification" in Robust Design)

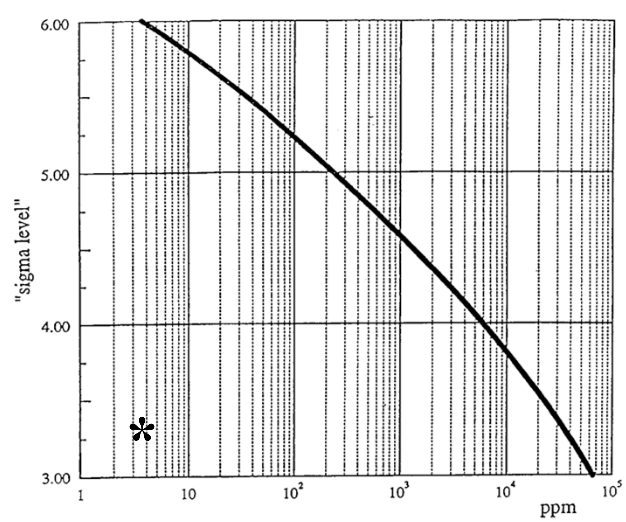
For a product property distribution with mean μ and standard deviation σ, if the closest specification limit is $k\sigma$ away from the mean, then it is said that the process or the product is at a $k\sigma$ quality level.

Less sensitive components or subsystems need not be at 6σ quality level to lead to a 6σ level for the overall system.

Motorola's "sigma level" concept

Tolerance width	C_p	ppm	after 1.5σ shift
$\mu \pm 2\sigma$	0.67	45500	308770
$\mu \pm 3\sigma$	1.00	2700	66811
$\mu \pm 4\sigma$	1.33	63	6210
$\mu \pm 5\sigma$	1.67	0.57	232
$\mu \pm 6\sigma$	2.00	0.002	3.4

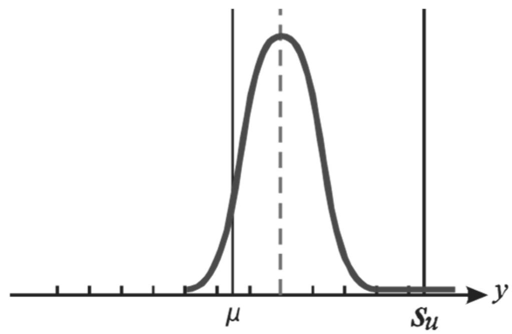
Conversion of dpmo ~ Sigma Level



Conversion of defective ppm to process "sigma level"

A Six Sigma Process

After 1.5 Sigma Shift: 3.4 ppm



WHAT DOES '6σ' IMPLY?

The area under the normal curve outside the $\pm 6\sigma$ regions is about 0.0000000018.

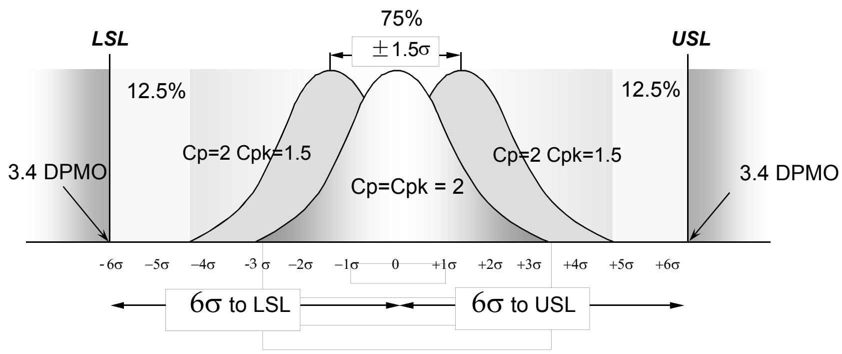
This corresponds to 1.8 ppb, i.e. under 2 per 1,000,000,000 units if

(a) the specifications are set at $\pm 6\sigma$ lines

(b) the product property variations follow the normal distribution

(c) the product units have an average property equal to the mid-specification value.

Technical Definition of Six Sigma (Motorola)



With conditions (a) and (b) above, then even if the average value of product property is 1.5σ off the target set at the center of the ±6σ range, the proportion of nonconforming product is about 0.0000034. The nonconformance is later known as DPMO, or "defects per million opportunities".

WHAT CAN BE EXPECTED FROM '6σ' PROCESS?

At the time of analysis, a process with C_{pk} of 2 is said to be **'6σ'** in the short term. In the long term, even if the average value of product property is 1.5σ off the target, the proportion of nonconforming product is about 0.0000034, or no more than 4 ppm.

WHAT SHOULD BE THE **'6σ'** OBJECTIVE?

It does not mean setting wide specifications.

It is meant to reduce or control variation (σ) such that the specifications given by customers become ±6σ points or even farther out.

'6σ' means not only meeting specifications but also with tremendous margin to spare (i.e. up to 1.5σ shift in center value and yet and yet less than 4 ppm conforming).

The Process Improvement Cycle

- Analyze the process
- Maintain the process
- Improve the process

Analyze the Process

- What should the process be doing?
- What can go wrong?
- What is the process doing?
- Achieve a state of statistical control
- Determine capability

Maintain the Process

- Monitor process performance
- Detect special cause variation and act upon it

Improve the Process

- Change the process to better understand common cause variation
- Reduce the common cause variation