

# Control Charts and Process Capability

Prof. Sayak Roychowdhury

# Sampling

- It is not always possible to measure quality characteristics of each item in a population.
- Samples are used to provide information about process or product characteristics at a fraction of cost.
- Necessary for destructive tests
- A **sampling design** is a procedure by which the observations in a sample are chosen from the population
- An **element** is an object for which data are gathered
- A **sampling unit** is an individual element or a collection of elements from a population
- A sampling frame is a list of **sampling units**



# Sampling Errors

Random  
Variation

- Inherent sampling variability, e.g. due to instrument, people etc.

Misspecification

- Happens in opinion polling, customer satisfaction survey, incorrect listing of sampling frame

Non responses

- Happens in sample surveys, cases where measurements not possible

# Sampling Methods



# Simple Random Sampling

- A sample of size  $n$  is chosen from a fixed population of size  $N$ . In SRS, each possible sample of size  $n$  has **equal chance** of being selected.
- Random number tables may be used for sampling
- In estimating population mean  $\mu$  by the sample mean  $\bar{X}$ , the variance of the estimator is given by

$$\hat{\sigma}_{\bar{x}}^2 = \frac{s^2(N-n)}{N} \quad (s^2 \text{ sample variance})$$

$(N - n)/N$  is called finite population correction factor

Precision is inverse of variance.





# Stratified Random Sample

- Useful when the population is heterogeneous, e.g. production from multiple machines, under multiple operators, samples from different geographical regions.
- **Stratified random samples** are obtained by separating the elements of the population in nonoverlapping groups (**strata**).
- Proportional allocation of sample size, for  $k$  strata, let  $N_i$  be the population size of the  $i^{th}$  strata, and  $\sum_{i=1}^k N_i = N$ .

Sample size from each strata:

$$n_i = \frac{nN_i}{N} \quad i = 1, 2, \dots, k$$



# Stratified Random Sample

- Sample mean and variance of estimator are given by

$$\bar{x}_{st} = \frac{1}{N} \sum_{i=1}^k N_i \bar{x}_i$$
$$Var(\bar{x}_{st}) = \frac{1}{N^2} \sum_{i=1}^k N_i^2 \left( \frac{(N_i - n_i)}{N_i} \right) \left( \frac{s_i^2}{n_i} \right), \quad i = 1, 2, \dots, k$$

Where  $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ ,  $s_i^2 = \sum_{j=1}^{n_i} \frac{(x_{ij} - \bar{x}_i)^2}{n_i - 1}$



# Cluster Sampling

- When a sampling frame is not available or obtaining samples from all segments of the population is not feasible due to geographical reasons, cluster sampling is used.
- Population is divided into groups of elements, called clusters
- Clusters are randomly selected and a census data is obtained.
- Sampling error maybe reduced by choosing many small clusters rather than choosing a large cluster





# Example of Cluster Sampling

A researcher wants to conduct a study to judge the performance of sophomore's in business education across the India.

By using cluster sampling, the researcher can club the universities from each city into one cluster (North, South, East, West regions).

These clusters then define all the sophomore student population in India.

Next, either using simple random sampling or systematic random sampling, **randomly pick clusters** for the research study.

Subsequently, by using simple or systematic sampling, the **sophomore's from each of these selected clusters can be chosen on whom to conduct the research study.**



# How to choose sample size?

- Bound on error estimation on population mean:

Let there is  $(1 - \alpha)$  probability that the difference between the estimated mean and the actual mean is not greater than  $B$  (*tolerable error bound*).

$$B = z_{\frac{\alpha}{2}} \sigma_{\bar{x}} = z_{\frac{\alpha}{2}} \left( \frac{\sigma}{\sqrt{n}} \right) \Rightarrow$$
$$n = z_{\frac{\alpha}{2}}^2 \left( \frac{\sigma^2}{B^2} \right)$$

An analyst wishes to estimate the average bore size of a large casting. Based on historical data, it is estimated that the standard deviation of the bore size is **4.2 mm**. If it is desired to estimate with a probability of **0.95** the average bore size to within **0.8 mm**, find the appropriate sample size.



# How to choose sample size?

- Bound on error estimation on population proportion:

Let there is  $(1 - \alpha)$  probability that the difference between the estimated proportion  $\hat{p}$  and the actual proportion  $p$  is not greater than  $B$  (*tolerable error bound*).  
*E.g. proportion of satisfied customers, prop of non conforming etc.*

$$B = z_{\frac{\alpha}{2}} \sigma_{\hat{p}} = z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \Rightarrow$$
$$n = z_{\frac{\alpha}{2}}^2 \left( \frac{p(1-p)}{B^2} \right)$$

*Either put  $p = \hat{p}$  (sample proportion) or  $p = 0.5$  for conservative estimate*



# Example

We want to estimate with a probability of 0.90 the proportion of nonconforming tubes to within 4%. How large a sample should be chosen if no prior information is available on the process?





# How to choose sample size?

- Estimating difference between 2 population means:

$$n = z_{\frac{\alpha}{2}}^2 \left( \frac{\sigma_1^2 + \sigma_2^2}{B^2} \right)$$

*Where  $B$  is the tolerance of error for estimating the difference in population means with sample means*

- Estimating difference between 2 population proportions:

$$n = z_{\frac{\alpha}{2}}^2 \left( \frac{p_1(1 - p_1) + p_2(1 - p_2)}{B^2} \right)$$



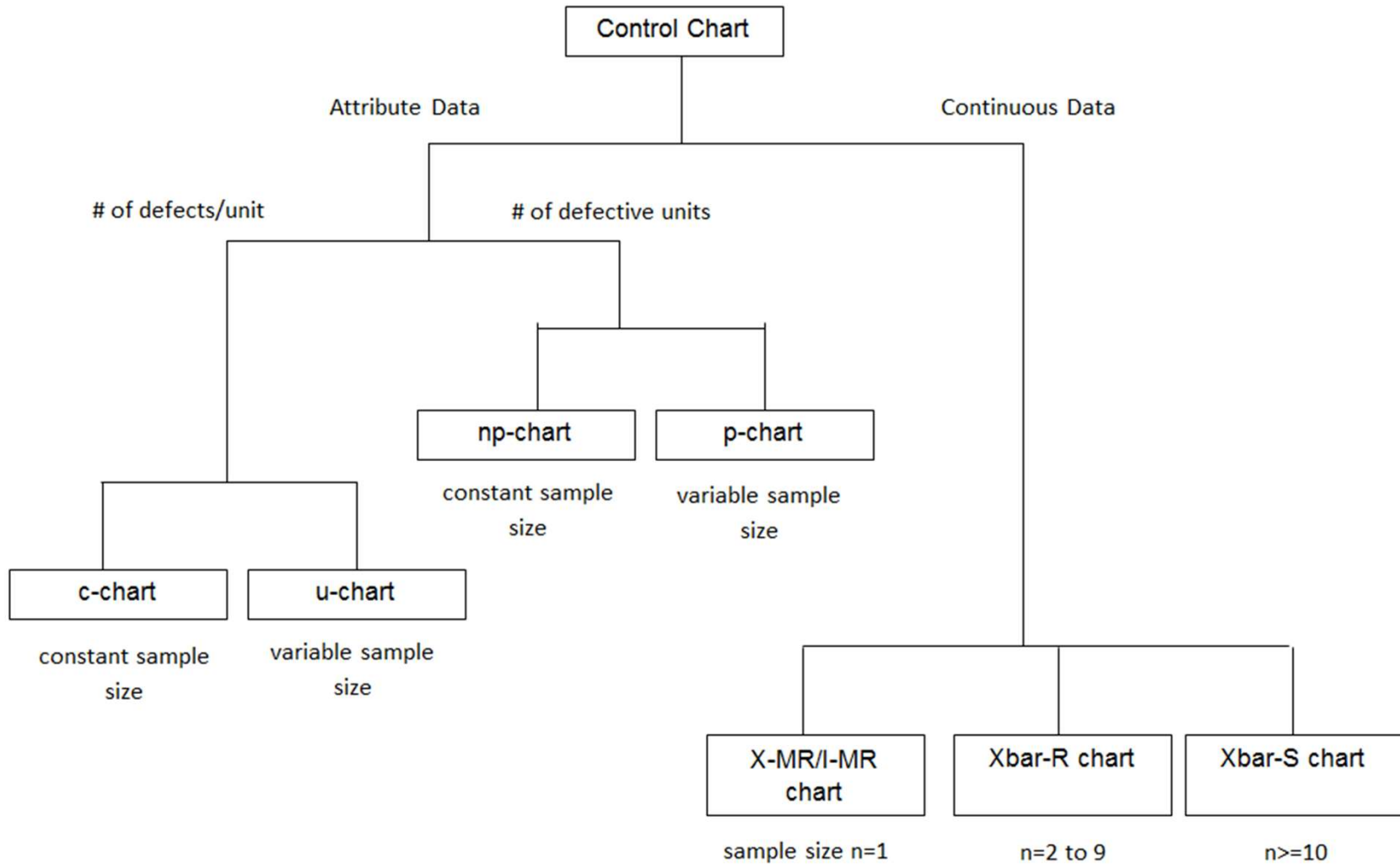


# 7 QC Tools

- Cause and Effect Diagram
- Check Sheet
- **Control Chart**
- Histogram
- Pareto Chart
- Scatter Diagram
- Stratification / Defect Concentration Diagram



# Control Charts



# Utility of Control Charts

- Control charts are proven techniques to improve productivity
- Effective in defect identification and prevention
- Control charts prevent unnecessary process adjustments
- Diagnostic information
- Process capability information

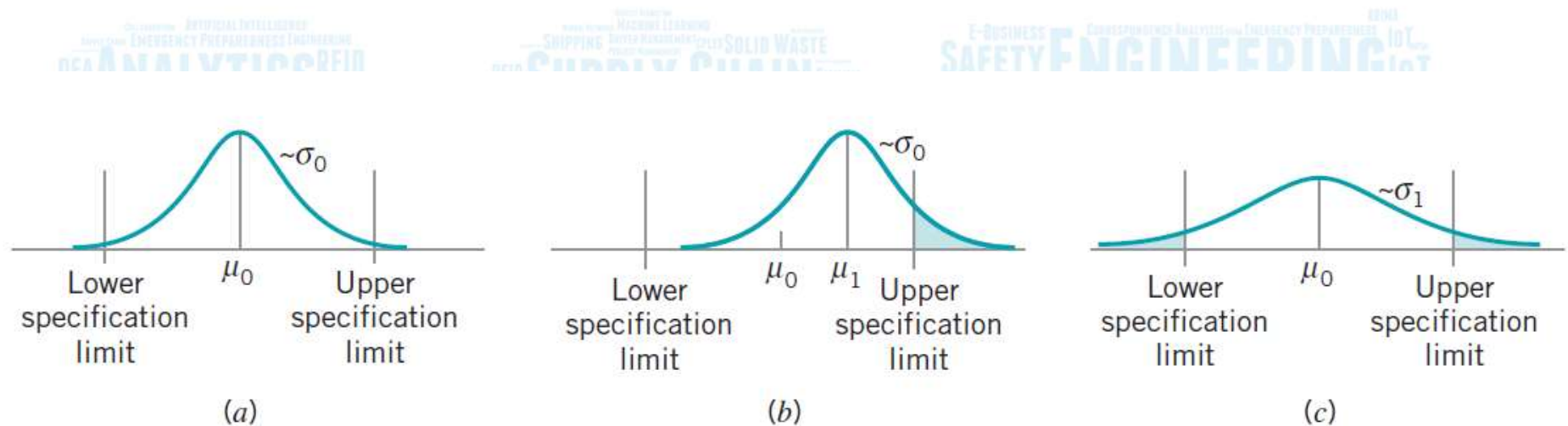


# Rational Subgroups (Section 5.3.4 Montgomery)

- Sampling procedure to ensure that the variation within the group is only due to chance causes.
- Lots from which the subgroups are chosen should be homogeneous, e.g. same machine, same operator, same mold cavity etc.
- Items of any one subgroup should be produced under essentially same conditions.
- Instant time method: Parts in the subgroup are chosen in the same time-instant. The next subgroup is picked after a certain time interval. (Maximum variation among subgroup)
- Period of time method: Subgroups are sampled from parts produced since the last sample was taken. It is used to make decisions about acceptance of products produced since the last inspection.



# In-control or Out of Control



■ **FIGURE 6.1** The need for controlling both process mean and process variability. (a) Mean and standard deviation at nominal levels. (b) Process mean  $\mu_1 > \mu_0$ . (c) Process standard deviation  $\sigma_1 > \sigma_0$ .





# Causes of Variation

Samples maybe out-of-control due to:

- **Special Cause or Assignable Cause:** Not inherent in the process, does not affect all the time. It could be the use of a wrong tool, tool damage, operator mistake, incorrect measurement etc. Control charts are used to detect the presence of special causes as soon as possible. Special causes need to be removed to get the process back to normality.
- **Common Cause:** Variability due to common or chance causes, inherent to a process. It is inherent part of process design and affects all time. Process need not be changed due to common cause variations.



# Control Limits of Shewhart Control Chart

## ■ X bar Chart

- $UCL_{\bar{X}} = \bar{\bar{X}} + 3\sigma_{\bar{X}}$  ( $\sigma_{\bar{X}}$  population standard deviation of subgroup averages)  
 $\approx \bar{\bar{X}} + A_2\bar{R}$  (used in practice)
- $LCL_{\bar{X}} = \bar{\bar{X}} - 3\sigma_{\bar{X}}$   
 $\approx \bar{\bar{X}} - A_2\bar{R}$
- $CL_{\bar{X}} = \bar{\bar{X}}$  Center line

## • R chart

- $UCL_{\bar{R}} = \bar{R} + 3\sigma_{\bar{R}} \approx D_4\bar{R}$
- $LCL_{\bar{R}} = \bar{R} - 3\sigma_{\bar{R}} \approx D_3\bar{R}$

# Derivation

- $\bar{\bar{X}} = \frac{1}{m} \sum_{i=1}^m \bar{X}_i = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n X_{ij}$  Centre Line
- UCL and LCL these are 3 sigma
- $X \sim N(\mu, \sigma), \bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$
- $UCL = \bar{\bar{X}} + 3\sigma_{\bar{X}} = \bar{\bar{X}} + 3 \frac{\sigma}{\sqrt{n}}$
- $LCL = \bar{\bar{X}} - 3\sigma_{\bar{X}} = \bar{\bar{X}} - 3 \frac{\sigma}{\sqrt{n}}$

# Derivation

- Relative Range  $W = R/\sigma$  a random variable
- Parameters of distribution of  $W$  depend on sample size  $n$
- Mean of  $W$  is  $d_2$
- Estimator of  $\sigma$  is  $\hat{\sigma} = R/d_2$ , we may use  $\hat{\sigma} = \bar{R}/d_2$  as  $\bar{R}$  is the average range of  $m$  preliminary samples
- $UCL = \bar{\bar{x}} + 3\hat{\sigma}/\sqrt{n} = \bar{\bar{x}} + 3\bar{R}/(d_2\sqrt{n}) = \bar{\bar{x}} + (A_2)\bar{R}$
- $CL = \bar{\bar{x}}$
- $LCL = \bar{\bar{x}} - 3\hat{\sigma}/\sqrt{n} = \bar{\bar{x}} - 3\bar{R}/(d_2\sqrt{n}) = \bar{\bar{x}} - (A_2)\bar{R}$



# Derivation

- Standard deviation of  $W$  is  $d_3$
- Estimator of  $\sigma$  is  $\hat{\sigma} = R/d_2$
- Standard deviation of  $R$  can be written as  $\sigma_R = d_3\sigma$  as  $R = W\sigma$
- $\hat{\sigma}_R = d_3\bar{R}/d_2$
- For R chart
  - $UCL = \bar{R} + 3\hat{\sigma}_R = \bar{R} + \frac{3d_3\bar{R}}{d_2} = \mathbf{D_4\bar{R}}$
  - $LCL = \bar{R} - 3\hat{\sigma}_R = \bar{R} - 3d_3\bar{R}/d_2 = \mathbf{D_3\bar{R}}$
  - $CL = \bar{R}$





# Revised Control Limits

Discard out of control samples with assignable causes.

Revised control limits are calculated as below, for total number of samples  $m$  and number of defective samples  $d$ :

$$\bar{\bar{X}}_{new} = \frac{m\bar{\bar{X}} - \sum_d \bar{X}_d}{m-d} = \bar{\bar{X}}_0$$

$$\bar{R}_{new} = \frac{m\bar{R} - \sum_d R_d}{m-d} = \bar{R}_0$$

$$\sigma_0 = \frac{\bar{R}_0}{d_2} \quad (d_2 \text{ can be found in table})$$

$$UCL_{\bar{X}} = \bar{\bar{X}}_0 + A \sigma_0 ; LCL_{\bar{X}} = \bar{\bar{X}}_0 - A \sigma_0$$

$$UCL_{\bar{R}} = D_2 \sigma_0 ; LCL_{\bar{R}} = D_1 \sigma_0 ;$$

# X-bar and R chart

- $\bar{X}$  chart monitors between sample variability,  $R$  chart monitors within sample variability
- To design  $\bar{X} - R$  chart, the following must be specified:
  - Sample size
  - Control limit width
  - Frequency of sampling
- $\bar{X}$  chart is capable to signal moderate to large process shifts ( $2\sigma$  or larger)
- $R$  chart is relatively insensitive to shift in process standard deviation for small samples *e.g.*  $n = 5$



# Error in Making Inference

- **Type I Error:** This error results from inferring a process is out of control when it is not. It is denoted by  $\alpha$ . This happens due to chance causes, when a control chart falls outside control limits. For  $3\sigma$  limits, probability of type I error is 0.0027.
- **Type II Error:** This error results from inferring a process is in control when it is out of control. It is denoted by  $\beta$ . This can happen when the process mean or the process variability or both have changed.



# Xbar & R Charting

- *Step 1.* (Startup) Collect data for 25.  $\bar{X}_{barbar}$  is grand average,  $R_{bar}$  is...
- *Step 2.* (Startup) “Trial” limits:
- *Step 3.* (Startup) Find out of control signals. Remove if assignable causes are found
- *Step 4.* (Startup) Revise limits.
- *Step 5.* (Steady State) Plot and local authority investigates if out-of-control signals occur (can act).

Subgp.	X1	X2	X3	X4	Xbar	R
1	20.50	3.10	2.10	4.00	7.43	18.40
2	1.20	2.40	2.40	1.40	1.85	1.20
3	5.40	2.20	2.30	0.20	2.53	5.20
4	1.10	11.00	3.10	1.50	4.18	9.90
5	1.40	6.50	2.20	6.50	4.15	5.10
6	2.30	2.30	0.30	19.40	6.08	19.10
7	13.10	3.10	2.40	0.40	4.75	12.70
8	0.60	2.10	3.30	5.30	2.82	4.70
9	1.60	1.60	13.10	0.50	4.20	12.60
10	8.10	1.00	2.20	0.10	2.85	8.00
11	3.20	5.30	9.20	1.30	4.75	7.90
12	4.50	5.40	4.50	14.70	7.28	10.20
13	2.40	1.30	0.20	10.30	3.55	10.10
14	1.60	5.40	3.10	7.20	4.32	5.60
15	3.20	9.30	4.00	2.10	4.65	7.20
16	0.20	0.60	1.30	2.60	1.17	2.40
17	1.30	1.30	5.10	0.30	2.00	4.80
18	2.00	6.10	5.10	5.20	4.60	4.10
19	2.30	1.10	6.10	5.20	3.67	5.00
20	2.40	4.60	0.60	6.20	3.45	5.60
21	3.00	1.60	6.50	1.20	3.07	5.30
22	5.60	14.10	6.50	2.30	7.12	11.80
23	0.50	1.10	2.10	1.10	1.20	1.60
24	1.40	3.00	4.40	2.20	2.75	3.00
25	2.30	2.20	1.40	2.40	2.08	1.00
				Xbarbar=	3.86	Rbar=7.3

# Trial Limits

$$d_2(n=4) = 2.059$$

$$\sigma_{est} = 7.300/2.059 = 3.55$$

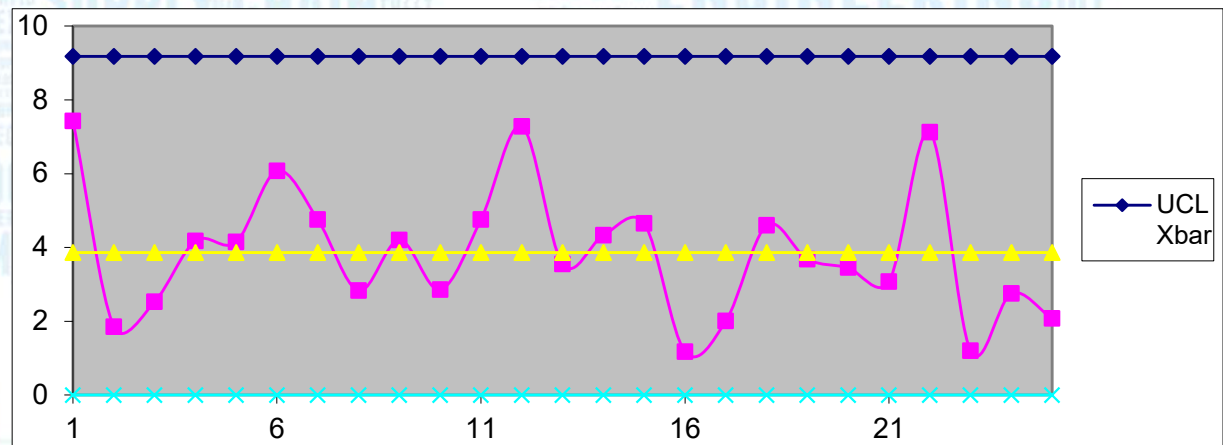
$$D_2 = 4.698$$

$$D_1 = 0.000$$

$$UCL_{\bar{X}} = \bar{X}_{barbar} + 3.0 \times \frac{\sigma_0}{\sqrt{n}}$$

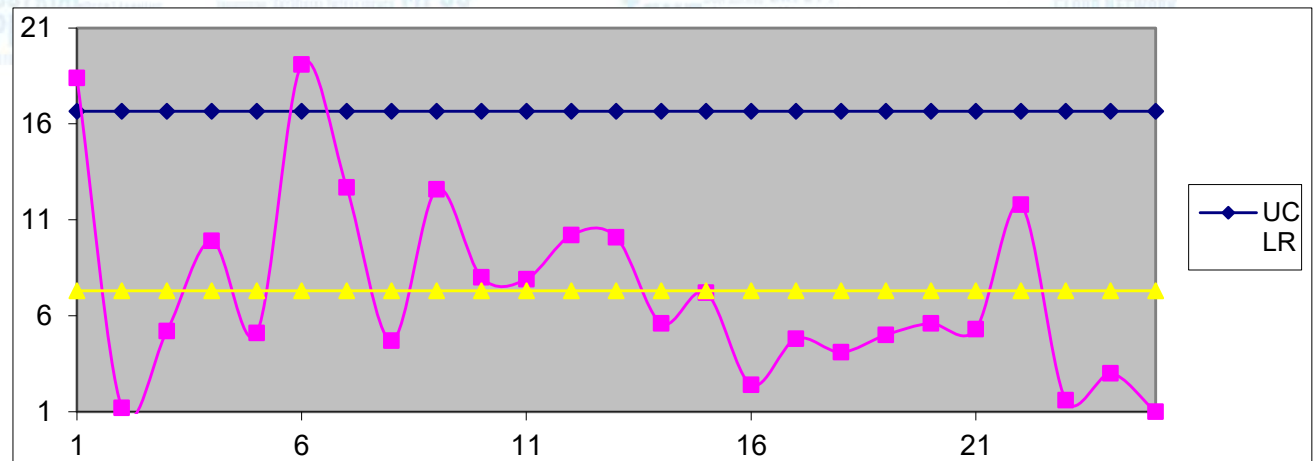
Signals on the R chart. Do  
detective work. Overnight  
stays. Not fair to keep.

So remove.



Subgroup	UCL $\bar{X}$	$\bar{X}$	CL $\bar{X}$	LCL $\bar{X}$
1	9.1781	7.43	3.86	0
2	9.1781	1.85	3.86	0

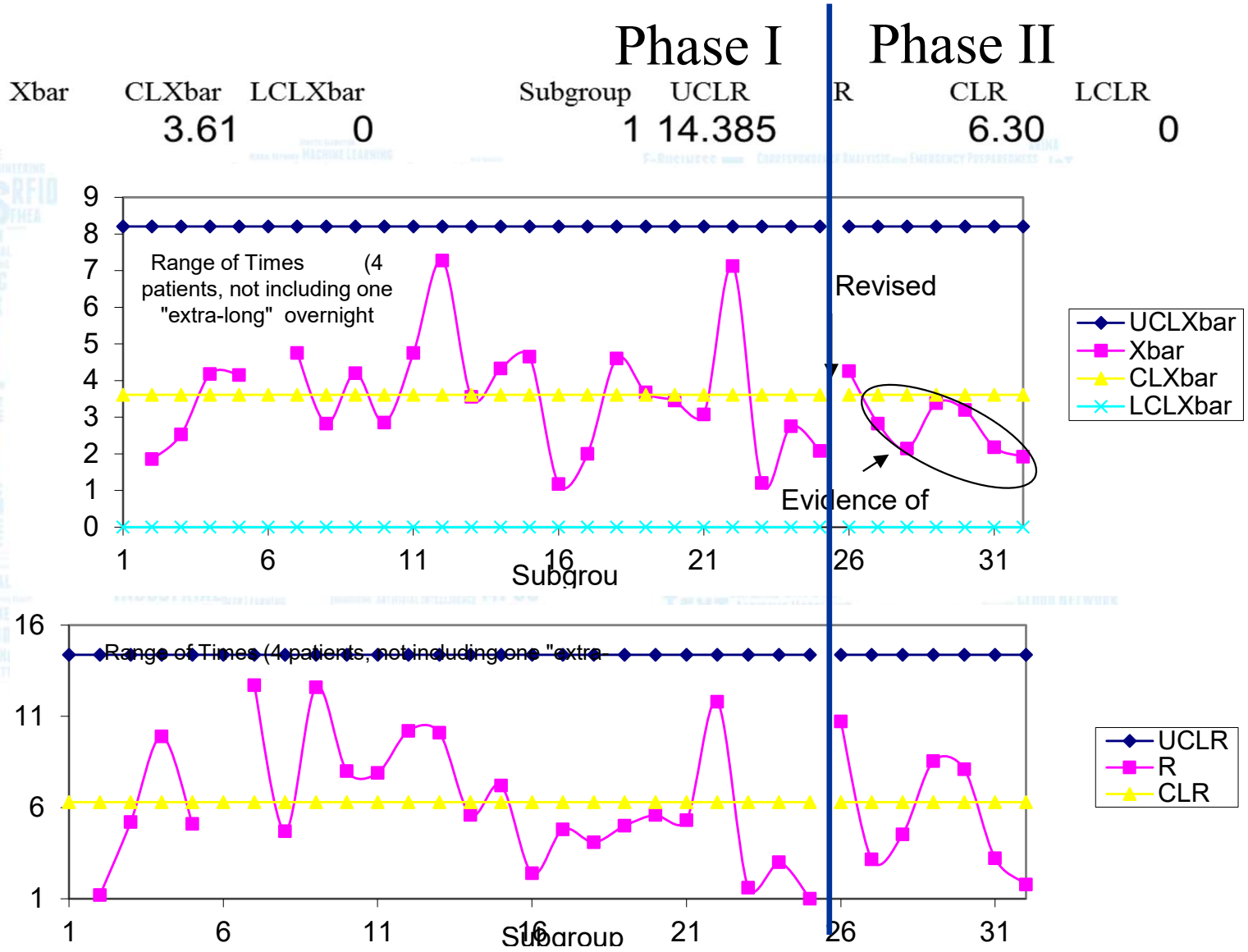
Subgroup	UCLR	R	CLR	LCLR
1	16.657	18.40	7.30	0
2	16.657	1.20	7.30	0





# Revised Limits

$d_2(n=4) = 2.059$   
 $\sigma_{est} = 6.300/2.059 = 3.062$   
 $D_2 = 4.698$   
 $D_1 = 0.000$   
 $6\sigma_0$  = the process capability  
 = 18.4 hours  
 (range for hospital)  
 No specs. so no  $C_{pk}$



# Revision Formula

$$X_{barbar, revised} = [25 X_{barbar, trial} - (\text{removed})] \div (25 - \# \text{ removed})]$$

$$(25 * 3.86 - 7.43 - 6.08) \div 23 = 3.6 \text{ hours}$$

Makes a small difference but it is fair as long as we clarify we are not considering overnight stays

Only remove if an assignable cause was found and eliminated.

Otherwise leave data in (common or chance causes).

Process capability: Measurement of the Common Cause variation/system quality:

→  $6 \sigma_0$  derive fairly measures common cause variation

→ 18.4 hours (you can pretty much count on range less than that)



# X bar and S chart

- It is occasionally desirable to monitor process standard deviation directly, rather than indirectly as done in  $R$  chart.
- $\bar{X}$  and  $S$  chart are preferable when
  - The sample size  $n$  is moderately large for  $n > 10$  or 12.
  - The sample size  $n$  is variable
- The unbiased estimator of population variance  $\sigma^2$  is sample variance  $s^2$
- The sample sd  $s$  estimates  $c_4\sigma$ , sd of  $s$  is  $\sigma\sqrt{1 - c_4^2}$

\* The sample standard deviation  $s$  is *not* an unbiased estimator of the population standard deviation  $\sigma$ . It can be shown that

$$\begin{aligned} E(s) &= \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma[(n-1)/2]} \sigma \\ &= c_4\sigma \end{aligned} \quad (4.17)$$

# Xbar and S Chart

- Since  $E(s) = c_4\sigma$  the center line is  $c_4\sigma$ . The 3 sigma limits of the s-chart is given by

$$\begin{aligned}UCL &= c_4\sigma + 3\sigma\sqrt{1 - c_4^2} \\CL &= c_4\sigma \\LCL &= c_4\sigma - 3\sigma\sqrt{1 - c_4^2}\end{aligned}$$

# X bar and S chart with Sample Estimators

- Consider  $s/c_4$  as an unbiased estimator of  $\sigma$
- For  $m$  preliminary samples with sd  $s_i$ , the average of  $m$  standard deviation is given by  $\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i$

$$UCL = \bar{s} + \frac{3\bar{s}}{c_4} \sqrt{1 - c_4^2} = B_4 \bar{s}$$

$$CL = \bar{s}$$

$$LCL = \bar{s} - \frac{3\bar{s}}{c_4} \sqrt{1 - c_4^2} = B_3 \bar{s}$$

$$\text{Where } B_3 = 1 - \frac{3}{c_4} \sqrt{1 - c_4^2} \text{ and } B_4 = 1 + \frac{3}{c_4} \sqrt{1 - c_4^2}$$



# X bar and S chart with Sample Estimators

- Control limit for corresponding  $\bar{X}$  chart is given by

$$UCL_{\bar{X}} = \bar{\bar{x}} + \frac{3\bar{s}}{c_4\sqrt{n}} = \bar{\bar{x}} + A_3\bar{s}$$

$$CL_{\bar{X}} = \bar{\bar{x}}$$

$$LCL_{\bar{X}} = \bar{\bar{x}} - \frac{3\bar{s}}{c_4\sqrt{n}} = \bar{\bar{x}} - A_3\bar{s}$$

# Probability of an Xbar False Alarm

Assume that the subgroups are rational (skipping and representative of a homogeneous group) and the system is under control (no assignable causes so distribution is the same), what is probability of a false alarm on the next subgroup from the Xbar chart?

CLT  $\rightarrow$   $\bar{X} \sim N[\mu, \sigma/\sqrt{n}]$

$$1 - \Pr\{\text{LCL}_{\bar{X}} = \mu - 3\sigma/\sqrt{n} \leq \bar{X} \leq \text{UCL}_{\bar{X}} = \mu + 3\sigma/\sqrt{n}\} = 1 - \Pr\{-3 \leq Z \leq 3\} = 1 - 2 \Pr\{Z \leq -3\} = 0.0027$$

Even if you are doing everything correctly, you have a 0.0027 chance of a false alarm.

(Average Run Length (ARL) in control  $1 \div 0.0027 = 370.4$ )



# Average Run Length (ARL)

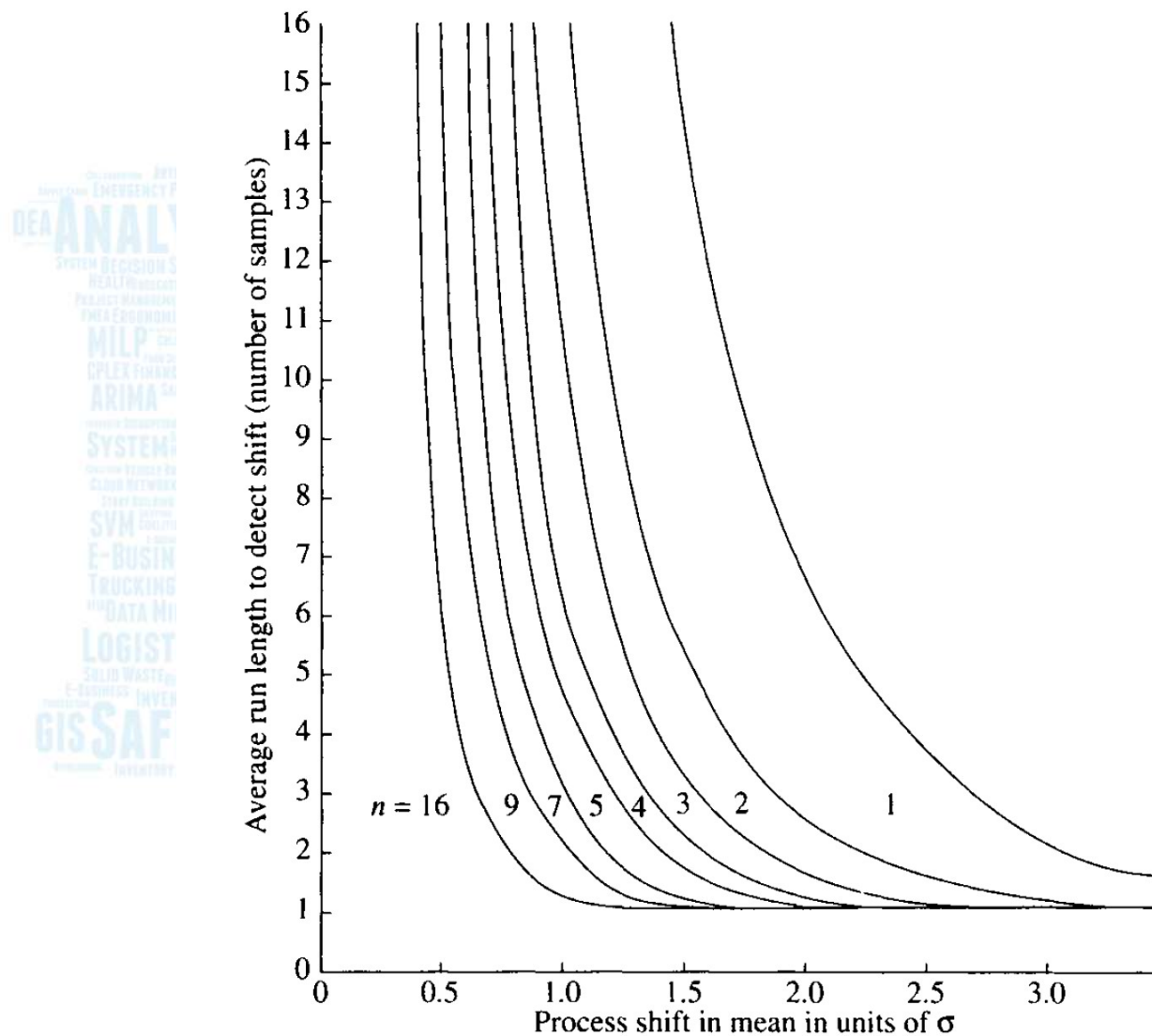
- To measure the performance of a control chart, ARL is used.
- ARL denotes the number of samples, on average, required to detect and out of control signal.
- If  $P_d$  is the probability that a process is out of control then run length is 1 with probability  $P_d$ , 2 with probability  $(1 - P_d)P_d$ , 3 with  $(1 - P_d)^2 P_d$ . Hence

$$ARL = \sum_{j=1}^{\infty} j(1 - P_d)^{j-1} P_d = \frac{P_d}{[1 - (1 - P_d)]^2} = \frac{1}{P_d}$$

- For a process in control,  $P_d$  is  $\alpha$  (probability of type I error)
- For an in-control process, ARL should be as large as possible.
- For an out of control process  $P_d = 1 - \beta$ .  $\beta$ : Probability of Type II error.  $ARL = \frac{1}{1 - \beta}$
- For out of control process, ARL should be as small as possible.



# ARL Curve



**FIGURE 6-8** ARL curves for control charts for the mean.

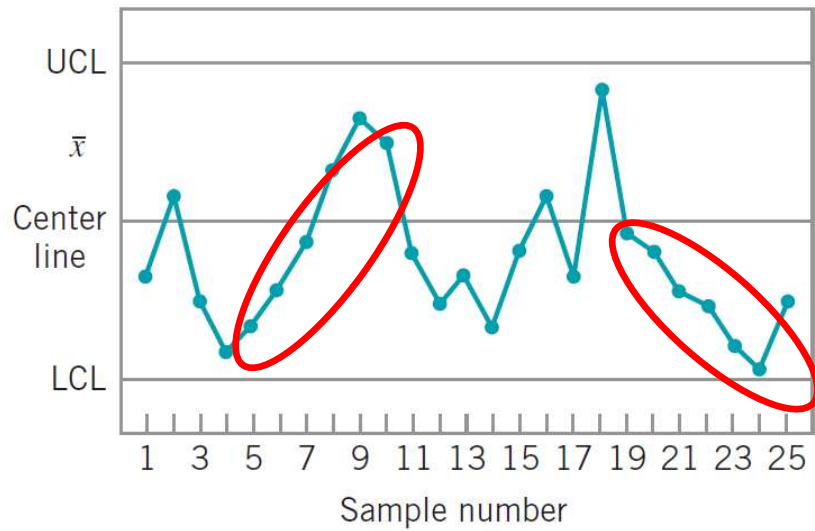
Source: QCI, Amitava Mitra





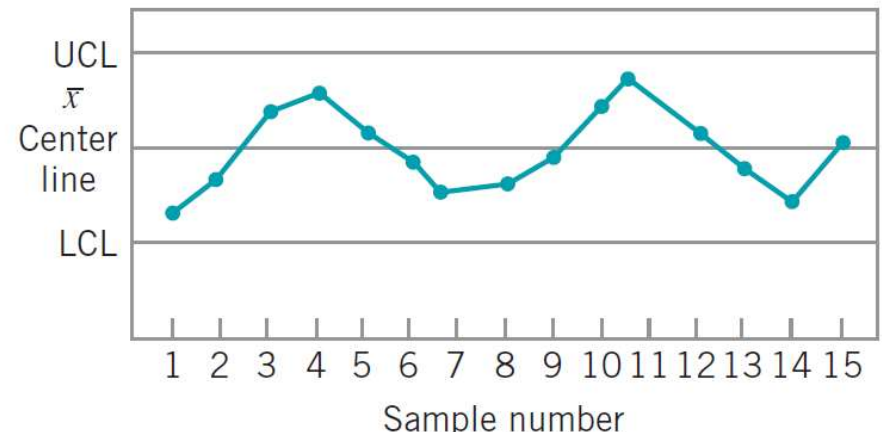
# Analysis of Pattern in Control Chart

- Process should be investigated when there is non-random pattern in control chart



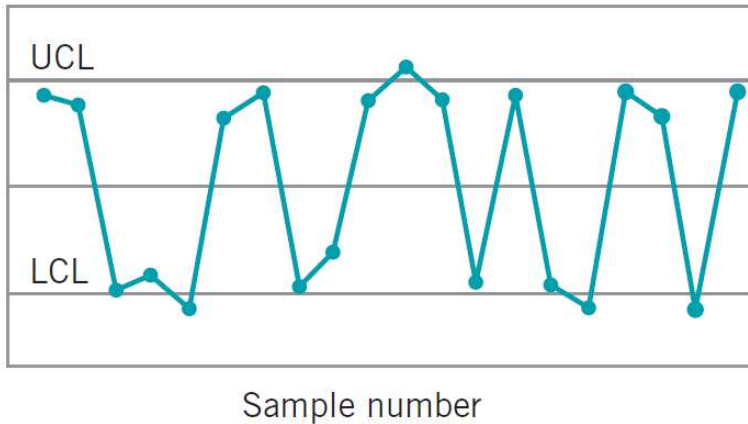
- Most sample averages are below centre line !
- Continuous rise (*run-up*) or fall (*run-down*)

- A run length of 8, or consecutive 8 points above or below centre line may indicate out-of-control
- Cycles are another type of pattern



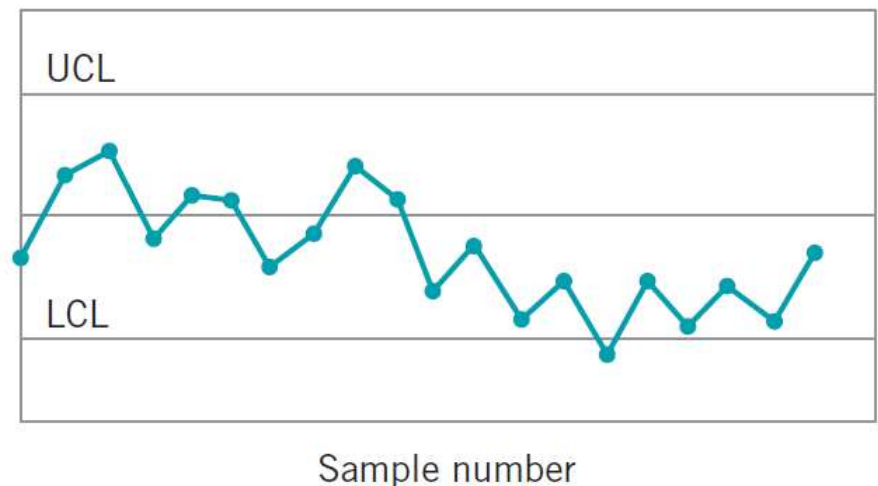


# Analysis of Pattern in Control Chart

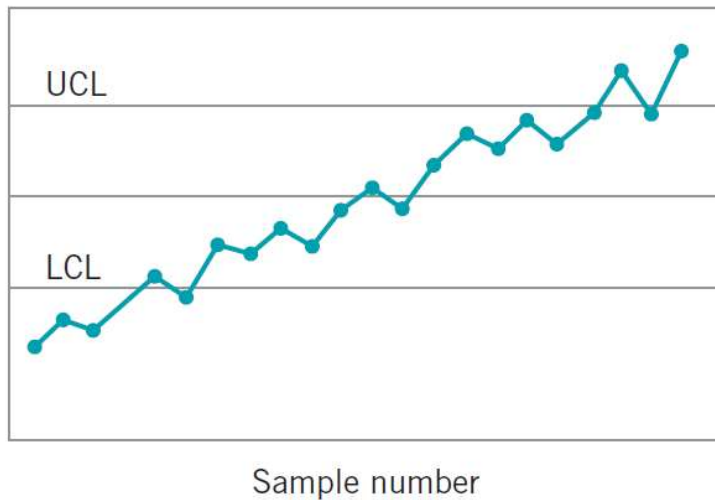


- A **mixture pattern** when most points are near control limits, result of two or more distributions

- A **shift** in the process may occur due to introduction of new operator, material, machine, inspection method etc.

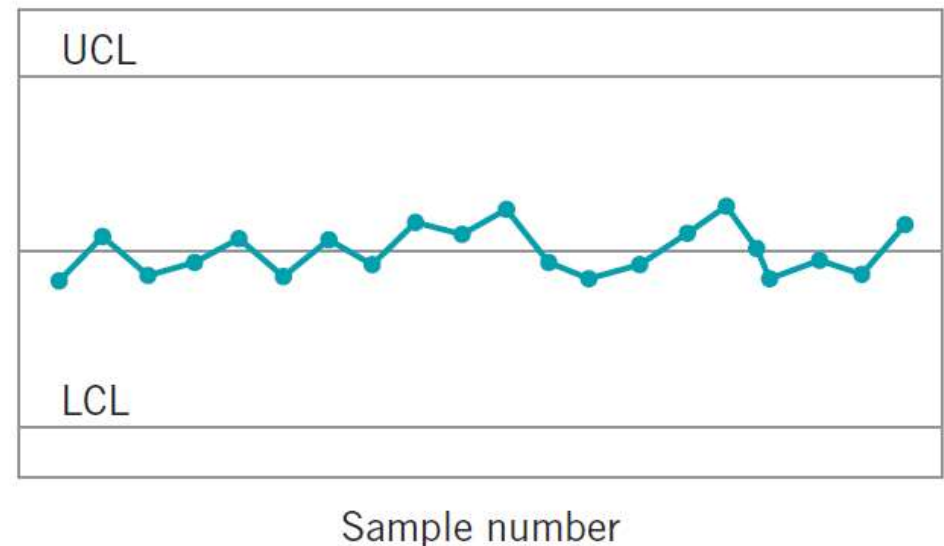


# Analysis of Pattern in Control Chart



- A **trend** occurs when there is continuous deterioration of tools. In chemical processes they occur due to settling or separation of components.

- **Stratification** is when points cluster artificially near centre line. This may happen when rational subgrouping is not done.



# Rules of Identifying Out of Control Points

1. If single point plots are outside control limits
2. If 2 out of 3 consecutive points plots fall outside  $2\sigma$  warning limits on the same side of centre line
3. If 4 out of 5 consecutive points fall beyond  $1\sigma$  limit on the same side of centre line
4. If 9 or more consecutive points fall on one side of centre line
5. If 6 or more consecutive points steadily increases or decreases

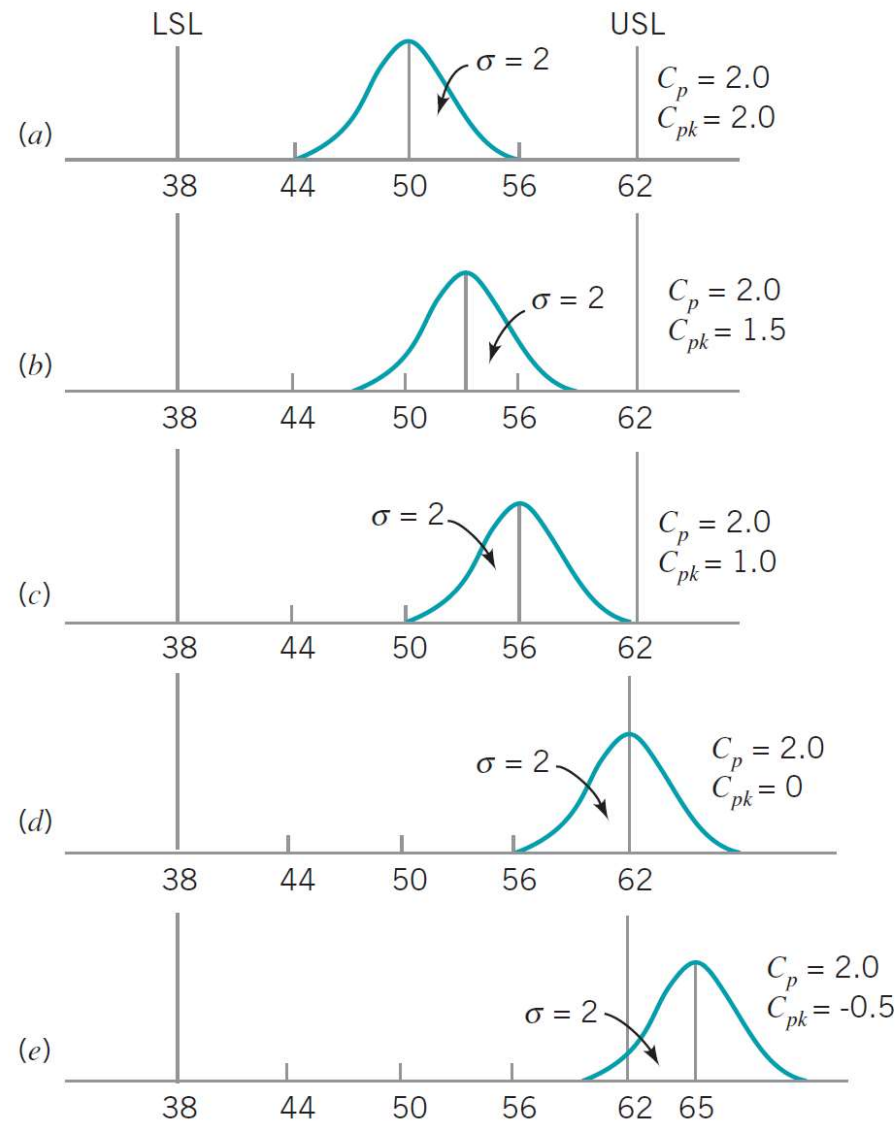


# Process Capability

- Assumptions:
  - The quality characteristic has normal distribution
  - Process is in statistical control
- Process Capability Ratio (PCR) :  $\widehat{C}_p = \frac{USL - LSL}{6\hat{\sigma}}$
- $P = \left(\frac{1}{C_p}\right) 100$  gives the percentage of the specification band used by the process.
- $C_{pk} = \min\left(C_{pu} = \frac{USL - \mu}{3\sigma}, C_{pl} = \frac{\mu - LSL}{3\sigma}\right)$  is used to determine if the process is centered.



# Process Capability



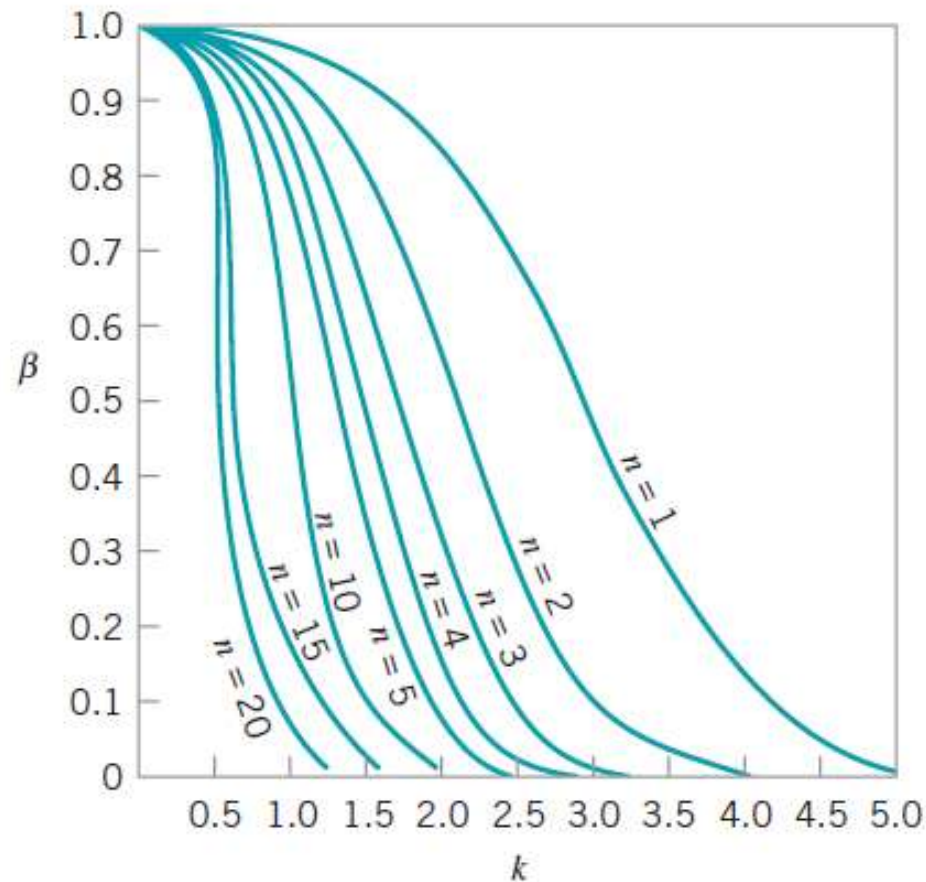


# Type I and Type II Error in CC

- Type I Error ( $\alpha$ ): Detecting a shift in process when there is no shift (False Alarm)
  - Wider the control limits, lower is the probability of Type I Error
- Type II Error ( $\beta$ ): Not detecting a shift in process when there is a shift
  - Closer the control limits, smaller is the probability of Type II Error
  - Decreases with increase in sample size
  - OC-curve is used to as a visual tool to analyze the change in probability of Type II Error with the change in process parameter.
- Sample size should be chosen judiciously, so that a balance in probabilities of Type I and Type II Error is maintained.



## OC-curve for $\bar{x}$ chart



# OC-curve for $\bar{x}$ chart

- Consider the process mean has shifted from  $\mu_0$  to

- $\mu_0 + k\sigma$

- $\beta = P(UCL \leq \bar{x} \leq LCL | \mu = \mu_1 = \mu_0 + k\sigma)$

- $\beta = \Phi\left(\frac{UCL - (\mu_0 + k\sigma)}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{LCL - (\mu_0 + k\sigma)}{\frac{\sigma}{\sqrt{n}}}\right)$

$$= \Phi\left(\frac{\mu_0 + \frac{L\sigma}{\sqrt{n}} - (\mu_0 + k\sigma)}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{\mu_0 - \frac{L\sigma}{\sqrt{n}} - (\mu_0 + k\sigma)}{\frac{\sigma}{\sqrt{n}}}\right)$$
$$\beta = \Phi(L - k\sqrt{n}) - \Phi(-L - k\sqrt{n})$$
