Quality Data Analysis

Statistical Quality Monitoring - Introduction

Bianca Maria Colosimo – biancamaria.colosimo@polimi.it

Reference:
Introduction to Statistical Quality Control – D.C. Montgomery

1

Process monitoring – Statistical Process Monitoring

Model of the process output
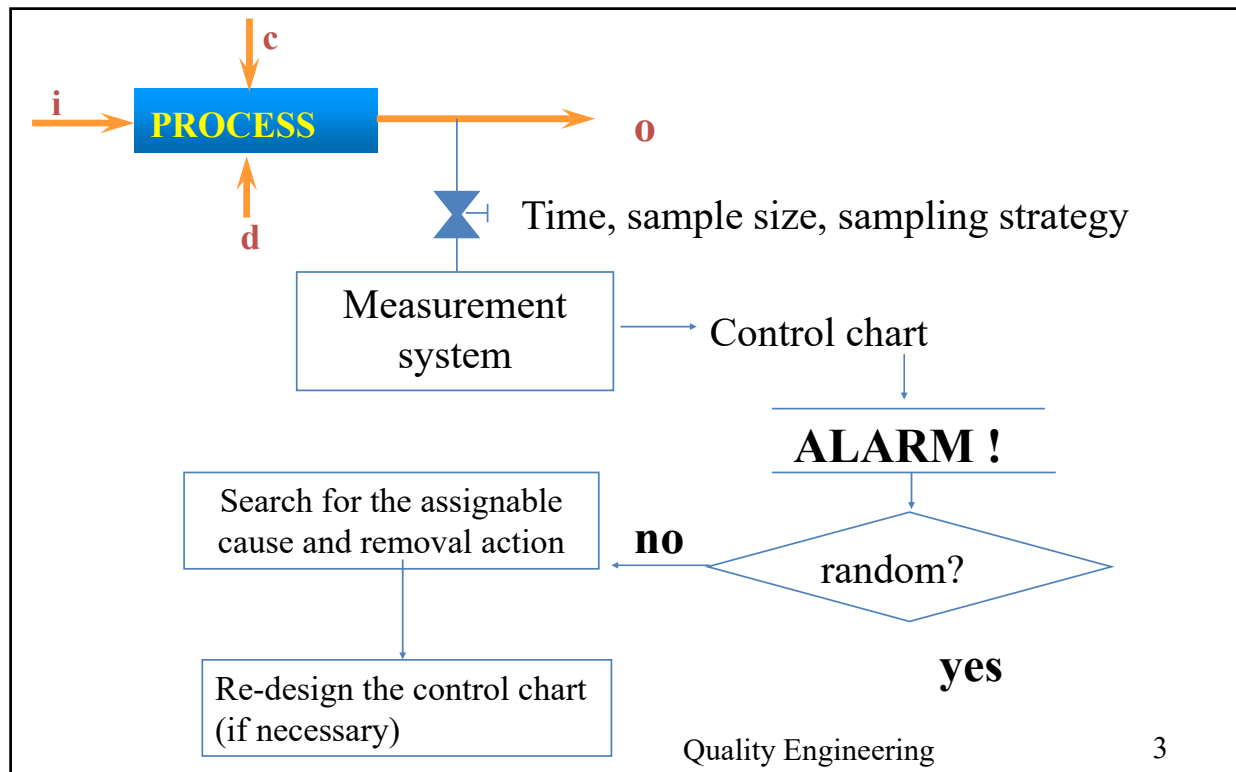
$$Y_t \overset{iid}{\sim} (\mu, \sigma_\varepsilon^2)$$

Basic model
(in-control process)

Is the process stable with time?
- Is the mean μ stable with time?
- Is the variance σ_ε^2 (standard deviation σ_ε) stable with time?
- (is the distribution stable with time)

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

2

2

Slide 3

i → **PROCESS** → o

c ↓ (into PROCESS)

d ↓↑ (into PROCESS)

Time, sample size, sampling strategy

Measurement system → Control chart

ALARM !

random?

no → Search for the assignable cause and removal action

yes

Re-design the control chart (if necessary)

Quality Engineering   3

3

Slide 4

Control chart

statistical procedure –a systematic analysis of the process output is used to identify out of controls as effect of special causes

1: success depends on the detection of the out-of-control states but especially on the identification of the assignable cause (even this is an *improvement step*)

2: Schonberger (1986): People collecting data are the ones that can analyze those data and think to possible solutions. The success depends on the ability to collect the right data at the right time. Operators tend to collect data after a problem has been detected. Usually, the relevant part of data are is related to the time before the problem detection.

→ Data gathering should be systematic

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

4

4

Traditional control charts (Shewhart)

1. Process output O: take a sample (*subgroup*) of data (n) at regular intervals
2. From all the samples: sample statistic V (eg. Sample mean, sample standard deviation,, *sample* range)
3. If O **is caused by a** random IID process, any sample statistic computed from these data will be random IID

(Pay attention: distribution of the individual measurements is usually different from the distribution of the sample statistic)

Out-of-control alarm

LCS

LC

LCI

t

distribution of V
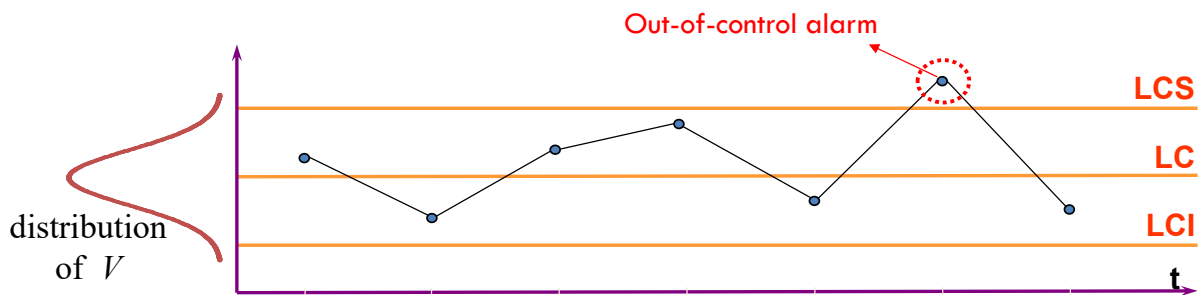
Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

5

5

Basic scheme Shewhart's control chart

$$UCL = \mu_V + K\sigma_V$$
$$CL = \mu_V$$
$$LCL = \mu_V - K\sigma_V$$

where
- $\mu_V = E(V)$
- $\sigma_V = \sqrt{V(V)}$
- K = positive coeff (usually K = 3)

Note:

– Symmetry
– (quantitative) out-of-control rule: alarm if points outside the limits
– (qualitative) out-of-control rule: look for strange data pattern for V (trend, cicles, etc.)

Quality Data Analysis- BM Colosimo
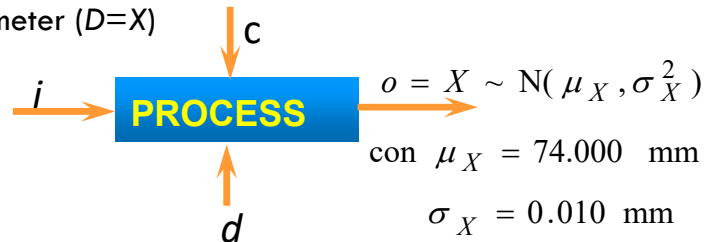
POLITECNICO MILANO 1863

6

6

Example: Shewhart control chart

Assume that the process produces shafts:
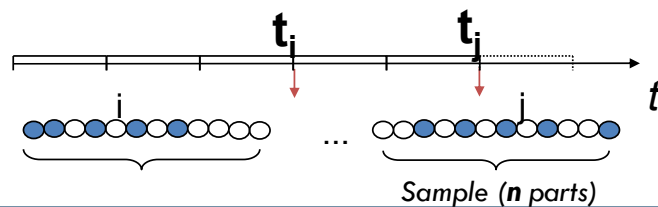
we are interested to the outer diameter (*D=X*)

Stable process:

c

i → **PROCESS**

d

$o = X \sim N(\mu_X, \sigma_X^2)$

con $\mu_X = 74.000$ mm

$\sigma_X = 0.010$ mm

Monitoring the process mean:

t_i t_j

t

1. Take at regular interval

i

...

i

Samples n=5

*Sample (**n** parts)*

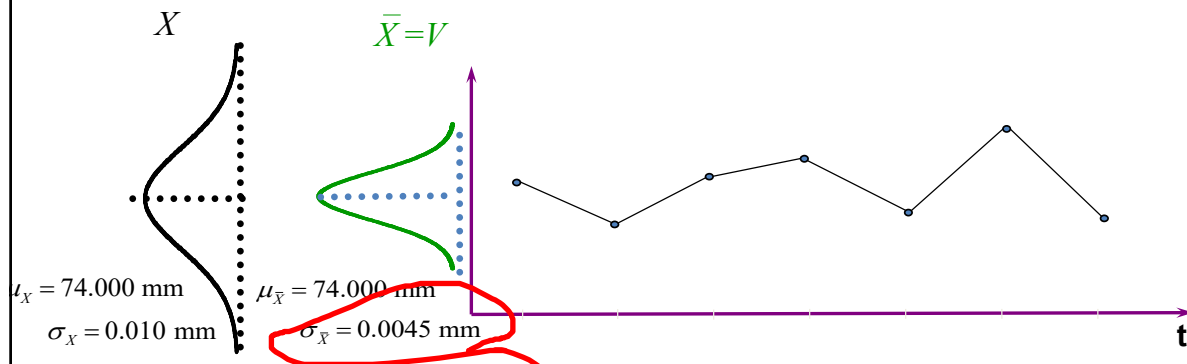Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

7

7

Example: Shewhart control chart

2. Compute

$$\Rightarrow \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu_X, \frac{\sigma_X^2}{n})$$

X

$\overline{X} = V$

$\mu_X = 74.000$ mm

$\sigma_X = 0.010$ mm

$\mu_{\overline{x}} = 74.000$ mm

$\sigma_{\overline{x}} = 0.0045$ mm

t

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

8
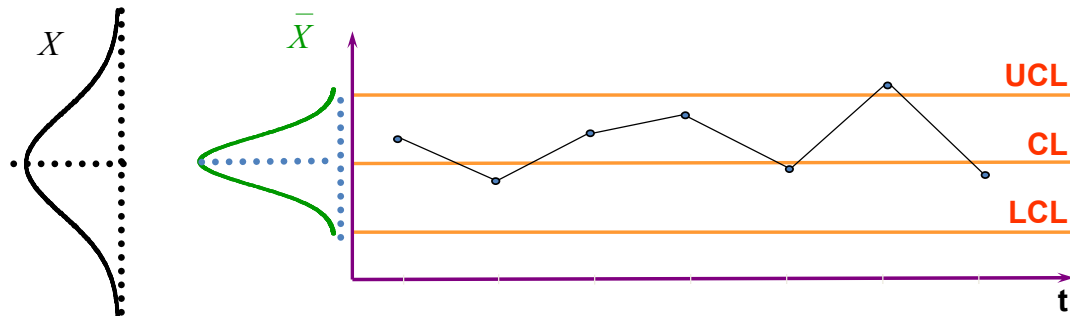
8

4

Example: Shewhart control chart

3. Control limits

$$K = 3$$

$$UCL = \mu_{\bar{X}} + K\sigma_{\bar{X}} = 74.000 + 3(0.0045) = 74.0135$$

$$CL = \mu_{\bar{X}} \qquad = \mu_{\bar{X}} = 74.000$$

$$LCL = \mu_{\bar{X}} - K\sigma_{\bar{X}} = 74.000 - 3(0.0045) = 73.9865$$

X $\bar{X}$

UCL

CL

LCL

t

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

9

9

Connection with hypothesis testing
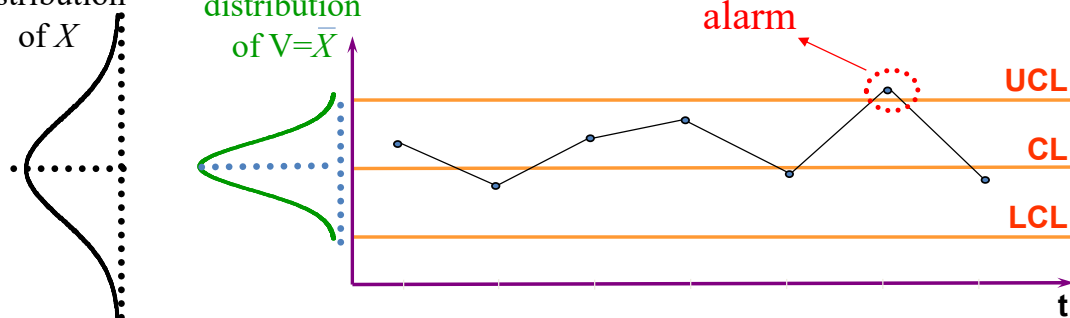
test

H_0: process is in control (IID)

H_1: process is out of control

Under H_0

distribution of X

distribution of V=$\bar{X}$

alarm

UCL

CL

LCL

t

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

10

10

5

Sources of variability

Process variability $\quad=\quad$ **(1)** Common causes $\quad+\quad$ **(2)** Special causes

(1): due to many different sources (material, machines, tools, operators, measurement systems, …). NB: anche in Business process

(2): due to some specific issues that can be easily identified (batch of material of poor quality, operator error, wrong setup, wear, etc.). Sometimes positive effect (serendipity)

Common causes (natural variability): management is responsible but contributions from all the people in the organization are needed to improve… "
Management may be the only ones who can really do something about the opportunity for improvement, but they cannot act if they really do not know about it" (Scherkenbach, 1986)
Special causes: identification is often linked to the way in which data are gathered and analized - they can be removed by the operator (setup) or from a decision at a higher level (input material)
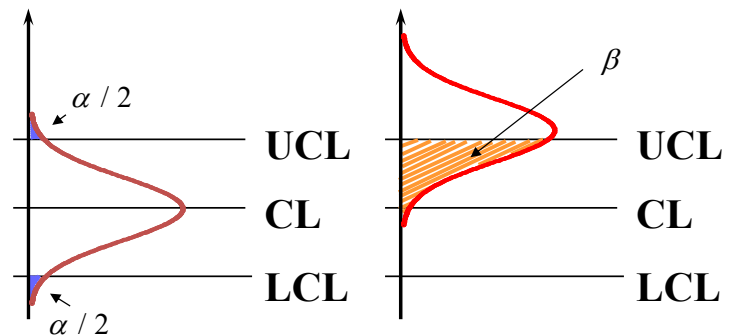
Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863
11

11

Connecton with hypothesis testing

Out of control: it can be due to an assignable cause … but it can be also due to the chance

In control: natural variability… but it can also be that an assignable cause is present

(Dynamic) test

H_0: process is in control

H_1: process is out of control

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863
12

12

Slide 13

In the previous example:
  in-control state: $\quad H_0$
  out-of-control state $\quad H_1$

$\mu_X = 74.000$ mm, $\sigma_X = 0.010$ mm
$\mu'_X = 74.020$ mm, $\sigma_X = 0.010$ mm

$$X_i \overset{iid}{\sim} N(\mu_X, \sigma_X^2) \Rightarrow \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu_X, \frac{\sigma_X^2}{n})$$
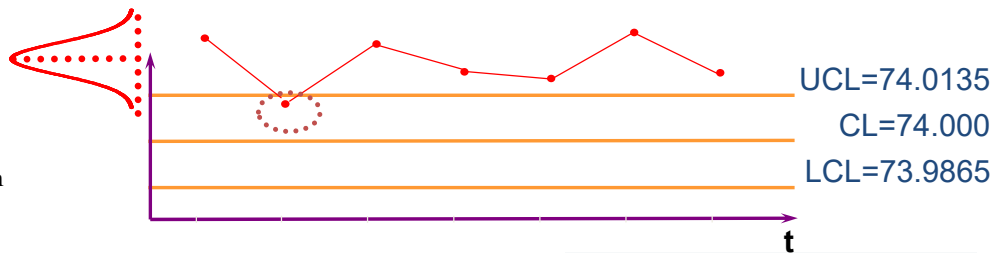
H_0 true

$\mu_{\bar{x}} = 74.000$ mm

$\sigma_{\bar{x}} = 0.0045$ mm

UCL=74.0135
CL=74.000
LCL=73.9865
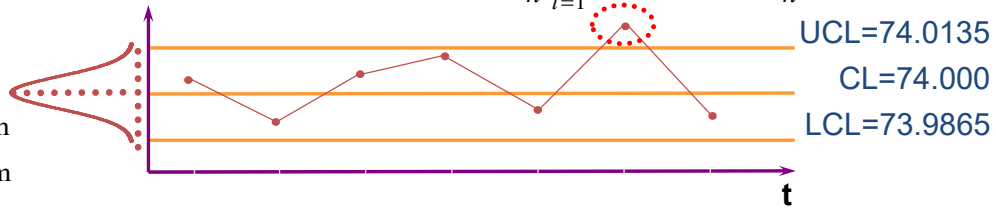
H_1 true

$\mu_{\bar{x}} = 74.020$ mm

$\sigma_{\bar{x}} = 0.0045$ mm

UCL=74.0135
CL=74.000
LCL=73.9865

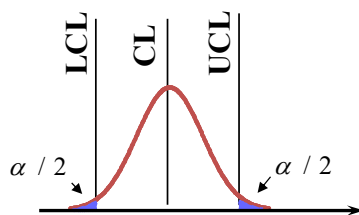Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

13

13

Slide 14

Control chart: K

Assume $\quad \boxed{V \sim NID(\mu_V, \sigma_V^2)}$
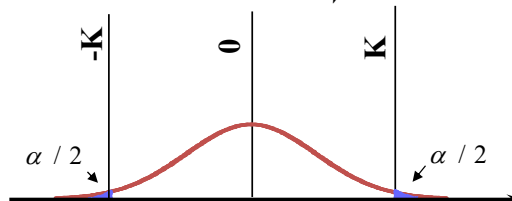
$$\frac{V - \mu_V}{\sigma_V} = Z \sim N(0,1)$$

$$\frac{LCL - \mu_V}{\sigma_V} = \frac{\mu_V - K\sigma_V - \mu_V}{\sigma_V} = -K$$

$$\frac{UCL - \mu_V}{\sigma_V} = K$$

standardize

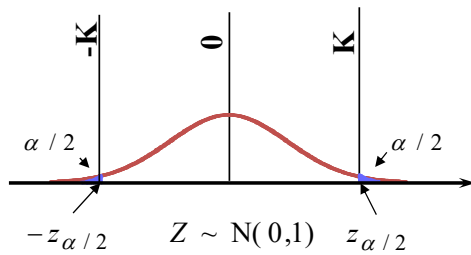LCL CL UCL

$\alpha / 2$ $\quad$ $\alpha / 2$

$Z \sim N(0,1)$

-K $\quad$ 0 $\quad$ K

$\alpha / 2$ $\quad$ $\alpha / 2$

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

14

14

K

$$K = z_{\alpha/2}$$

$Z \sim N(0,1)$ $z_{\alpha/2}$ $-z_{\alpha/2}$

$V \sim N(\mu_V, \sigma_V^2)$ $\alpha/2 = \Pr(V \le LCI) = \Pr(V \ge LCS)$

$$\alpha/2 = \Pr\left(\frac{V - \mu_V}{\sigma_V} \le \frac{LCI - \mu_V}{\sigma_V}\right) = \Pr(Z \le -K) = \Pr(Z \ge K) \Rightarrow K = z_{\alpha/2}$$

Cumulative Distribution Function

```
Normal with mean = 0 and standard
deviation = 1.00000

      x      P( X <= x )

  -3.0000        0.00135
```

$$\alpha = 0.0027 = 2.7\ ^0/_{00}$$

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

15

15

Link K-α

Assumptions: *V* is approximately distributed as a normal variable **(from IID to NID)**

- If variability due to a lot of causes
- Central limit Theorem

$$V \sim \cancel{N}ID(\mu_V, \sigma_V^2)$$

If ?

1. **Tchebyscheff inequality**: *X* follows any distribution

$$\Pr\left(|X - E(X)| > K\sqrt{V(X)}\right) \le 1/K^2$$

se $K = 3$ $\alpha \le \frac{1}{9} = 11\%$

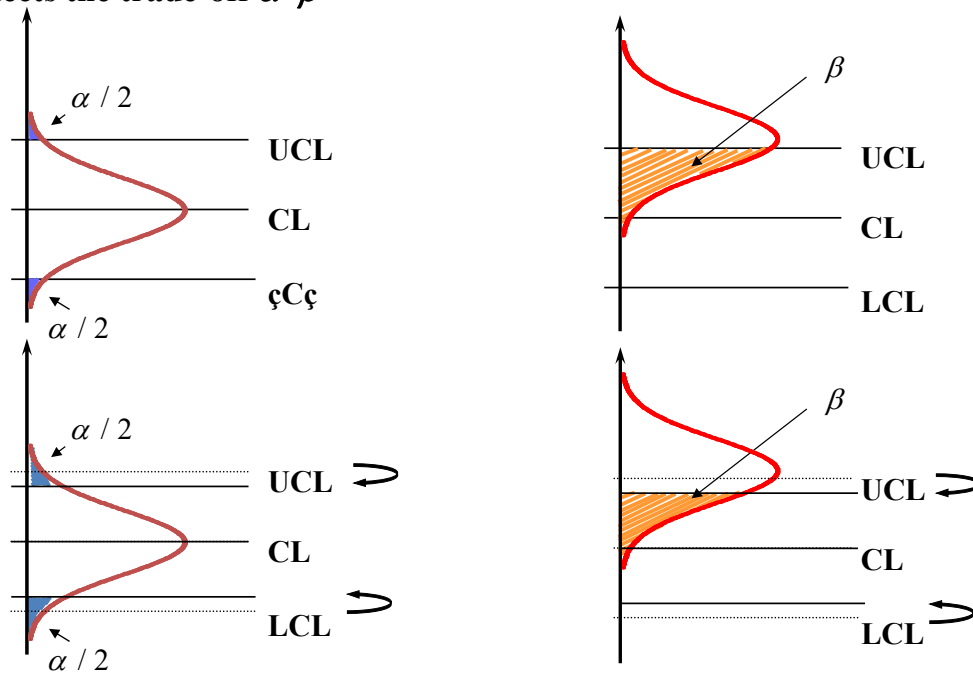Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

16

16

8

2. *Camp-Meidell* inequality (X follows a unimodal distribution in μ_0):

$$\Pr\left(|X - E(X)| > K\sqrt{V(X)}\right) \leq \frac{4}{9}\frac{1+f^2}{(K-f)^2} \qquad K > f \qquad f = \frac{|E(X) - \mu_0|}{\sqrt{V(X)}}$$

$$if \quad E(X) = \mu_0 \quad f = 0 \qquad and \ if \ K = 3 \qquad \boxed{\alpha \leq \frac{4}{81} = 4.9\%}$$

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

17

17

K affects the trade-off α–β

Quality Engineering

18

18

9

Second-type error β:

$$H_0 : \mu_X = 74.000$$
$$H_1 : \mu_X \neq 74.000 (= \mu_X + \delta)$$

$\alpha / 2$

β

UCL

δ_1

CL

LCL

$\alpha / 2$

t^*

$\alpha / 2$

β

UCL

δ_2 CL

LCL

$\alpha / 2$

t^*

$$\beta = \beta(\delta)$$

$$\beta(\delta = 0) = 1 - \alpha$$
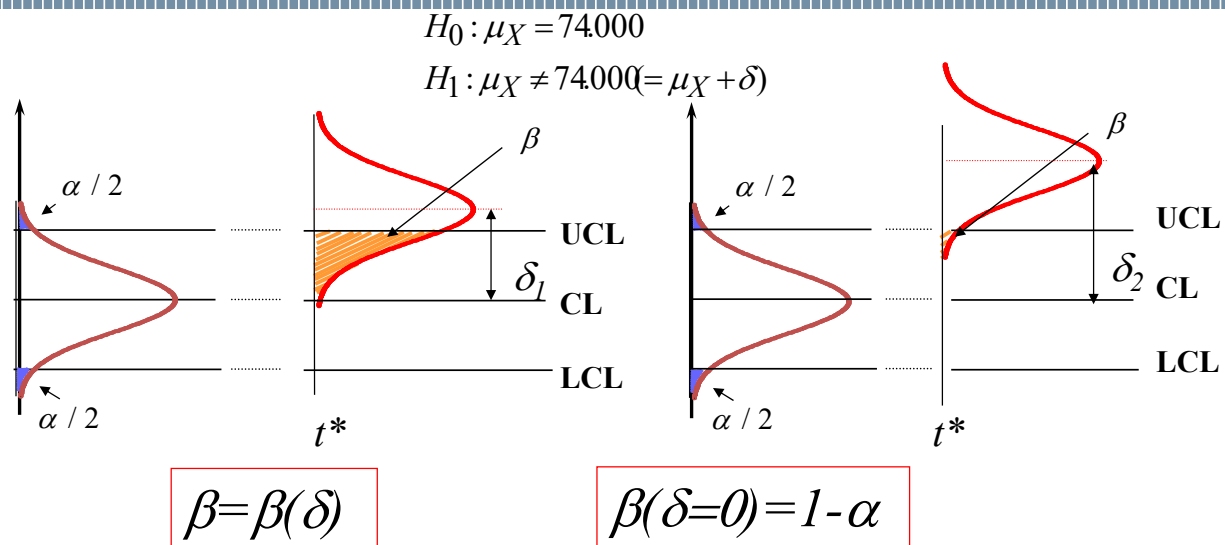
Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

19

19

Average Run Length (ARL)

ARL= expected value of points to be plotted on the control chart before an alarm

Assume shift $(H_0 \rightarrow H_1)$

$$\text{ARL}_1 = \sum_{j=1}^{\infty} j \; p(x = j)$$

$\alpha / 2$

β

UCL

δ_1 CL

LCL

$\alpha / 2$

h

t^*
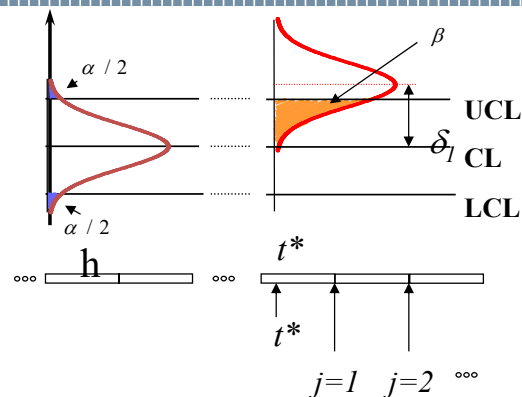
t^*

$j=1$ $j=2$

but:

$$p(x = 1) = 1 - \beta$$
$$p(x = 2) = \beta(1 - \beta)$$
...
$$p(x = j) = \beta^{j-1}(1 - \beta)$$

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

20

ARL (Average Run Length) يا متوسط طول دنباله در واقع متوسط تعداد نقاطي است كه داخل حدود كنترل قرار مي گيرند، قبل از اينكه يك نقطه خارج از كنترل مشاهده شود. از اين معيار معمولا براي مقايسه روشهاي مختلف كنترل آماري يك فرايند در فاز 2 استفاده مي شود. بگونه اي كه ابتدا با تغيير حدود كنترل ARL تحت كنترل روشها را يكسان نموده (معمولا 200 قرار مي دهند) و سپس به پارامترهاي فرايند شيفت مي دهند و ARL خارج از كنترل هر روش را محاسبه مي كنند. روشي كه داراي ARL خارج از كنترل كمتري باشد، به طور متوسط تغييرات پارامتر فرايند را سريعتر كشف مي كند.

$$\sum_{k=0}^{\infty} \beta^k = \frac{1}{1-\beta} \qquad \sum_{k=0}^{\infty} k\beta^k = \frac{\beta}{(1-\beta)^2}$$

J-1=k

$$ARL\ (H_1) = (1-\beta)\sum_{J=1}^{\infty} J\beta^{J-1} = (1-\beta)\sum_{k=0}^{\infty}(k+1)\beta^k =$$

$$= (1-\beta)\left[\sum_{k=0}^{\infty} k\beta^k + \sum_{k=0}^{\infty}\beta^k\right] = (1-\beta)\left[\frac{\beta}{(1-\beta)^2} + \frac{1}{1-\beta}\right]$$

$$= (1-\beta)\left[\frac{\beta+1-\beta}{(1-\beta)^2}\right] = \frac{1}{1-\beta}$$

$$ARL_1 = \sum_{j=1}^{\infty} j\,\beta^{j-1}(1-\beta) = \frac{1}{1-\beta}$$

Quality Data Analysis- BM Colosimo

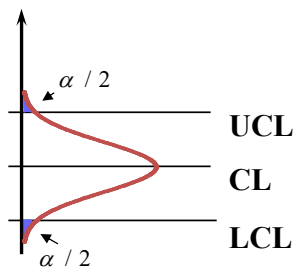POLITECNICO MILANO 1863

21

21

Shewhart control charts:

$$ARL_1 = \sum_{j=1}^{\infty} j\,\beta^{j-1}(1-\beta) = \frac{1}{1-\beta}$$

In-control process

$$ARL_0 = \frac{1}{\alpha}$$
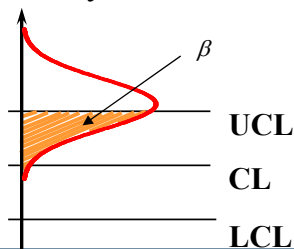
$$\alpha = 0.0027 \Rightarrow ARL_0 = 370.4$$

- Out-of-control process (shift)

$$ARL_1 = \frac{1}{1-\beta}$$
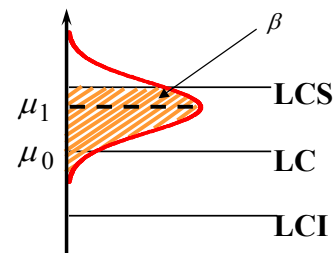
con $1-\beta$ = power =

= probability to detect the shift

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

22

22

11

Example

$$X \overset{iid}{\sim} \mathrm{N}(\mu_0,\sigma) \qquad \mu_0 \rightarrow \mu_1 = \mu_0 + 2\sigma$$

$$\mathrm{LCS} = \mu_0 + 3\sigma$$

$$\mathrm{LC} = \mu_0$$

$$\mathrm{LCI} = \mu_0 - 3\sigma$$
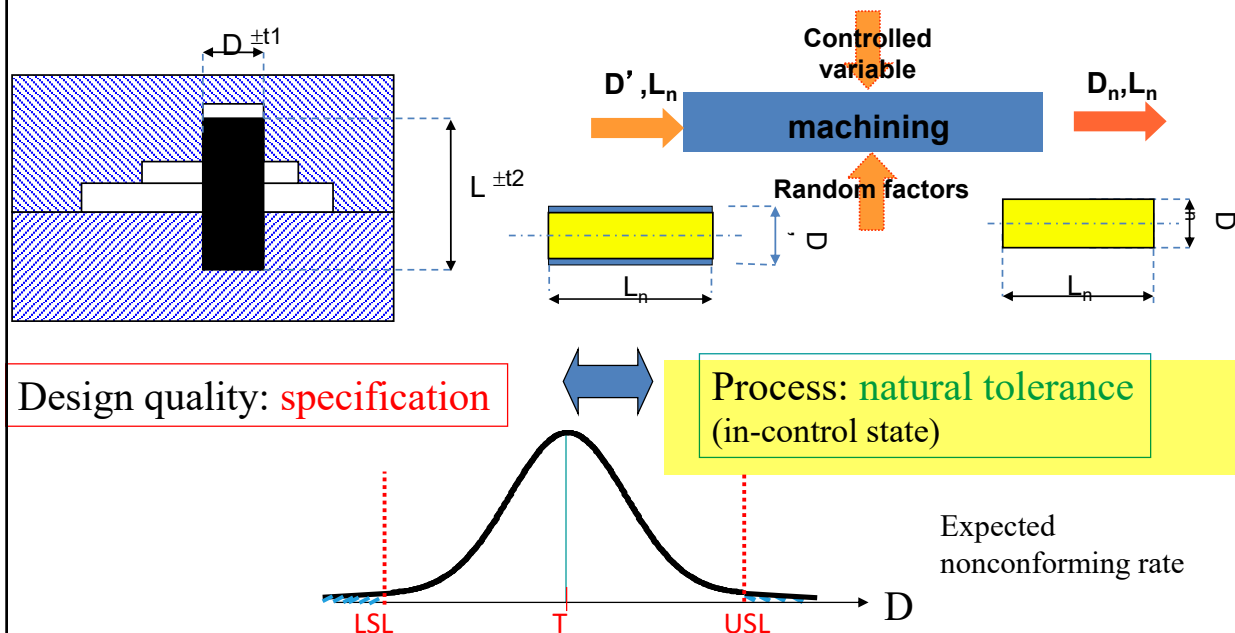
$$1 - \beta = \Pr(X < LCI) + \Pr(X > LCS) =$$

$$= \Pr\left(\frac{X - \mu_1}{\sigma} < \frac{\mu_0 - 3\sigma - \mu_1}{\sigma}\right) + \Pr\left(\frac{X - \mu_1}{\sigma} > \frac{\mu_0 + 3\sigma - \mu_1}{\sigma}\right) =$$

$$= \Pr(Z < -5) + \Pr(Z > 1) = 0.000000 + 0.158655 = 0.158655$$

$$\mathrm{ARL}_1 = \frac{1}{0.158655} = 6.3$$

Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

23

23

1. SPC and process capability

Design quality: specification

Process: natural tolerance (in-control state)

Expected nonconforming rate

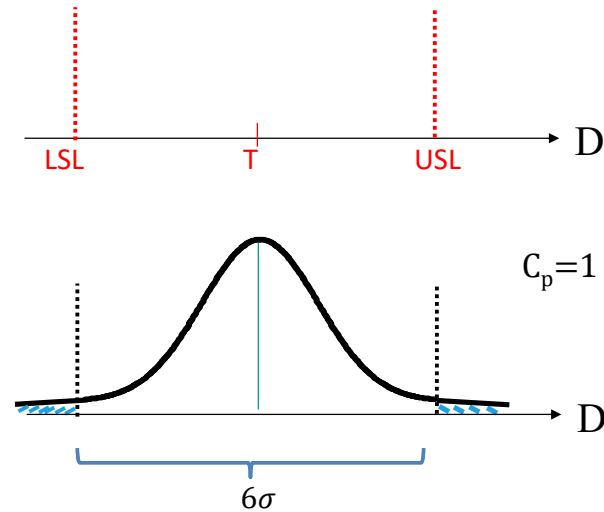Quality Data Analysis- BM Colosimo

POLITECNICO MILANO 1863

25

25

Process capability

$$C_p = \frac{USL - LSL}{6\sigma}$$

$$C_{pu} = \frac{USL - \mu}{3\sigma} \qquad C_{pl} = \frac{\mu - LSL}{3\sigma}$$
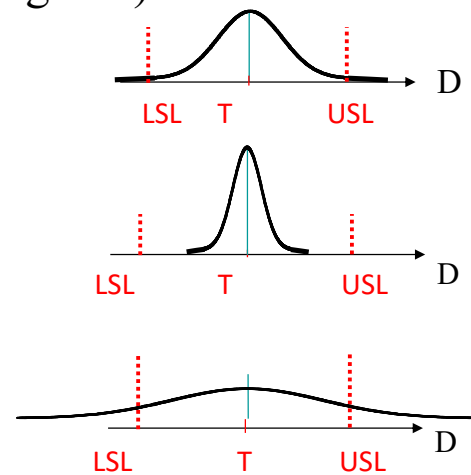
$$C_{pk} = \min\left(C_{pu}; C_{pl}\right)$$

LSL   T   USL   D

$C_p = 1$

D

6σ

POLITECNICO MILANO 1863

26

---

1. SPC and process capability

Process capability (expected nonforming rate):

1. Process capability - std

   LSL   T   USL   D

2. process capability - high
   - I can accept process changes

   LSL   T   USL   D

3. processo capability - low
   Reduce variability due to common causes:
   - Process improvement
   - Change the process

   LSL   T   USL   D

POLITECNICO MILANO 1863

27

27

Case 2 – high capability

USL

USL

(a+b)

(a)

LSL

LSL

==Special cause:== trend t

==Special cause:==- shift mean (a) t
- shift dispersion (b)

POLITECNICO MILANO 1863

28

28