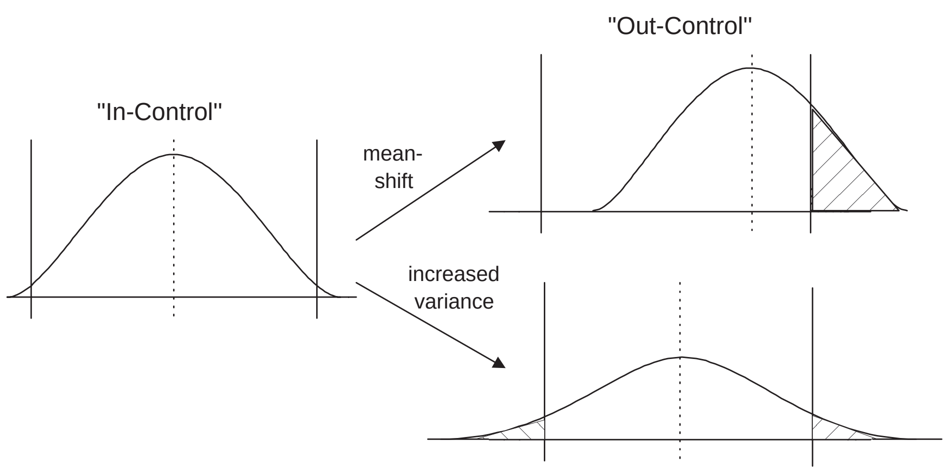
quickly into a position to challenge and then surpass American automobile and electronic equipments production.

1.2.2 SHEWART CHART

Basic Idea:

The basic idea is that there are switches in time which transfer the generating process into a distribution not typical of the dominant distribution. These switches manifest themselves into different average measurements and variances of the products.



Procedure

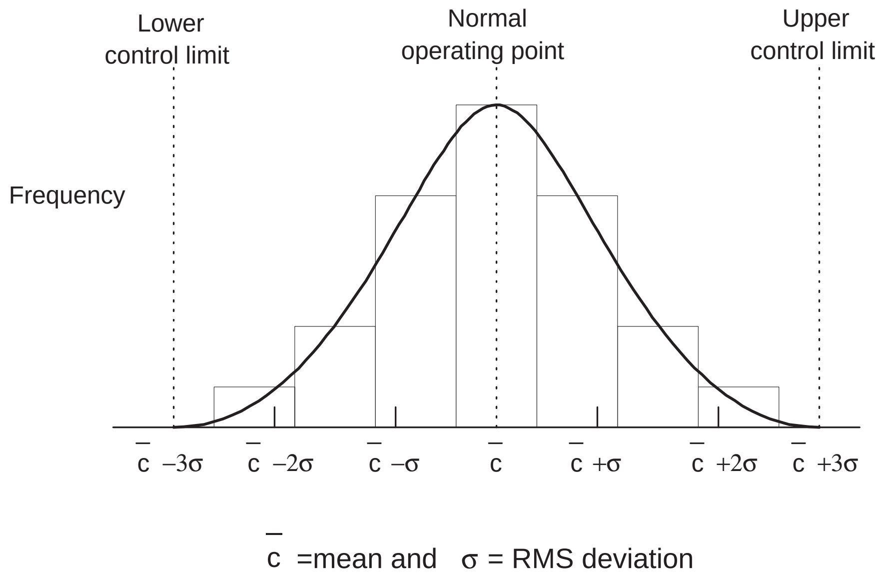
The procedure for constructing and using the Shewart chart is as follows:

- Collect samples during normal (*in-control*) epochs of the operation.
- Compute the sample mean and standard deviation.

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}(i) \tag{1.1}$$

$$\sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}(i)-\bar{y})^2} \tag{1.2}$$

You may also want to construct the histogram to see if the distribution looks close to being normal.

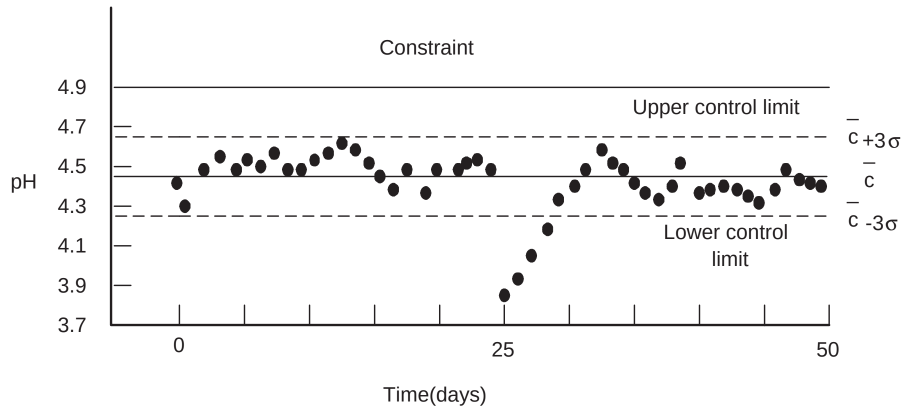


- Establish control limits based on the desired probability level γ. For normal distribution,

b	$P_r\{(\frac{y-\bar{y}}{\sigma_y})^2 \leq b\}$
0.02	0.1
1	0.632
2.71	0.9
3.84	0.95
5.02	0.975
6.63	0.99
7.88	0.995
9.0	0.997

For instance, the bound $\bar{y} \pm 3\sigma_y$ corresponds to the 99.7% probability level.

- Plot the quality data against the established limits. If data violate the limit (repeatedly), *out of control* is declared.



Assessment

The typicallly used 3 σ bound is thought to be both too small and too large, i.e.,

- too small for rejecting outliers (i.e., to prevent false alarms).
- too large for (quickly) catching small mean-shifts and other time-correlated changes.

The first problem can be solved by using the following modification.

Modification: "q-In-A-Row" Implementation

A useful modification is the *q-in-a-row* implementation. In this case,

> *out-of-control* is declared only when the bound is violated q times in a row.

q is typically chosen as 2-3.

Note that, assuming *samples during normal operation are independent in time*, the probability of samples being outside the γ probability bound q consecutive times during the *in-control* period is $(1-\gamma)^q$. For instance, with $\gamma = .9$ and $q = 3$, $(1-\gamma)^q = 0.001$. Hence, in effect, one is using 99.9% confidence limit.

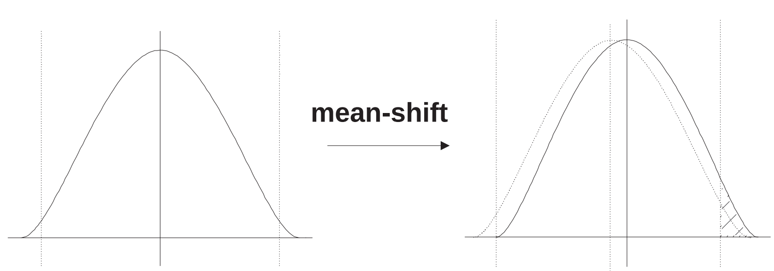
This way, the bound can be tightened without dropping the probability level (or the probability level can be raised without enlarging the bound). The q-in-a-row concept is very effective in rejecting measurement outliers or other short-lived changes for an obvious reason.

On the other hand, with a large q, the detection time is slowed down. In addition, the above concept relies on the fact that *sample data during normal operation are independent.* This may not be true. In this case, false alarms can result using a bound computed under the assumption of independence.

1.2.3 CUSUM CHART

Basic Idea

As shown earlier, with the Shewart Chart, one may have difficulty (quickly) detecting *small, but signficant* mean shifts.



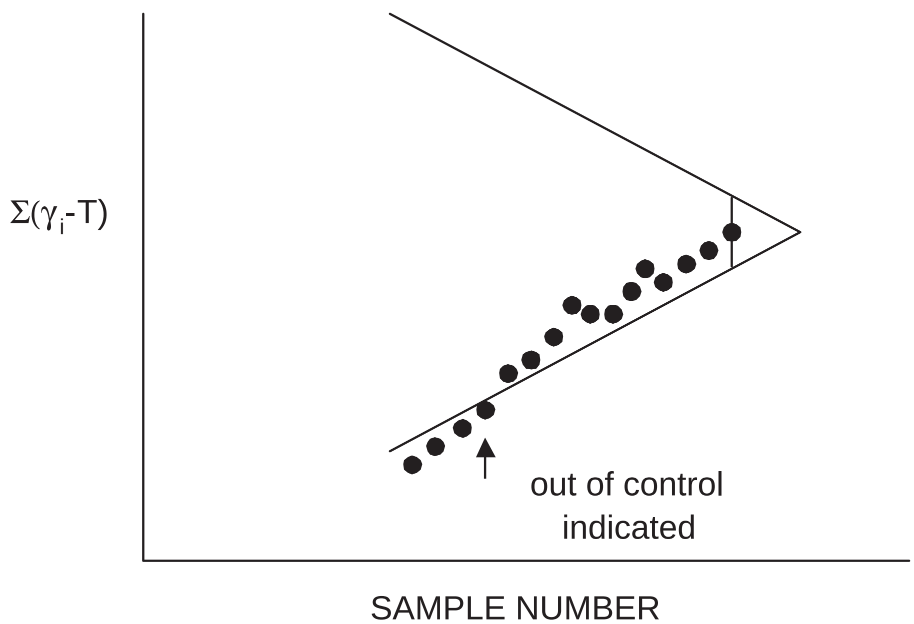
To alleviate this problem, the Shewart Chart can be used in conjunction with the CUSUM Chart. The CUSUM chart plots the following:

$$s(k) = \sum_{i=1}^{k} (\hat{y}(i) - \bar{y}) = s(k-1) + (\hat{y}(i) - \bar{y}) \qquad (1.3)$$

DuPont has more than 10,000 CUSUM charts currently being used.

Graphical Procedure

In this procedure, $s(k)$ is plotted on a chart and monitored. Any mean shift should show up as a change in the slope in the CUSUM plot. To test the statistical signficance of any change, a V mask is often employed:



Here one can set some tolerance on the rate of rise / fall. If any leg of the V-mask crosses the plotted CUSUM values, a *mean-shift* is thought to have occurred.

Non-Graphical Procedure

In practice, rather than using the graphical chart, one computes the two cumulative sums

$$s(k) = \max\{0, s(k-1) + (y(k) - \bar{y}) - \epsilon\} \tag{1.4}$$

$$t(k) = \min\{0, t(k-1) + (y(k) - \bar{y}) + \epsilon\} \tag{1.5}$$

where ϵ is the allowable slack.

One then tests:

$$\text{whether } s(k) \geq h \text{ or } h(k) \leq -h$$

If either is true, *out-of-control* is declared.

Two parameters need to be decided:

- The allowed slack ϵ is chosen as one-half of the smallest shift in the mean onsidered to be important.
- h is chosen as the compromise that will result in an acceptable long average run length (ARL) in normal situations, but an acceptably short ARL when the process mean shifts as much as 2ϵ units.

1.2.4 EWMA CHART

Basic Idea

The exponentially weighted moving average control chart plots the following exponentially weighted moving average of the data:

$$\begin{aligned} z(i) &= (1-\alpha)(y(i) + \alpha y(i-1) + \alpha^2 y(i-2) + \cdots\cdots + \alpha^k y(0)) \\ &= (1-\alpha)y(i) + \alpha z(i-1) \end{aligned} \tag{1.6}$$

Procedure

- As before, collect operational data from the *in-control* epochs.

- Establish its mean and standard deviation. This gives the target line and upper / lower bounds for the EWMA Chart.
- Compute on-line the sequence $z(k)$ according to $z(k) = y(k) + \alpha z(k-1)$ and plot it on the chart. As before, *out of control* is declared when either of the bounds is exceeded.

Advantages / Justification
The advantages of the EWMA are as follows:

- Note that the above amounts to first-order filtering of the data. The sensitivity to measurement outliers and other high-frequency (short-lived) variations is thus reduced.
- With $\alpha = 0$, one gets the Shewart Chart. As $\alpha \rightarrow 1$, it approaches the CUSUM chart. Hence, the parameter α affords the user some flexibility. (The choice of $\alpha = 0.8$ is the most common in practice.)
- This filtering is thought to provide one-step ahead prediction of $y(k)$ in many situations (more on this later).

On the other hand, one does lose some high frequency information through filtering, so it can slow down the detection.

1.3 MULTIVARIATE ANALYSIS

1.3.1 MOTIVATION

Main Idea / Motivation
A good way way to speed up the detection of abnormality and reduce the frequency of false alarm is **utilize more measurements**. This may mean