Chapter 6: Variables Control Chart

Learning Outcomes

After careful study of this chapter Students should be able to do the following:

• Know how to design variables control charts,
• Know how to set up and use x-bar and R control charts,
• Know how to set up and use x-bar and S or S² control charts,
• Know how to interpret patterns on x-bar and R control charts,
• Understand the rational subgroup concept for variables control charts,
• Determine the average run length for variables control charts,
• Understand the importance of the normality assumption for individuals control charts and know how to check this assumption,
**Discrete data:**
A data which is based on information such as pass / fail. In data you cannot be more specific.
Ex1. Do you like my PPP?
The answer can only be yes or no. The data is discrete.
Ex2. AC Gas leakage can be either OK or NOK type.

**Continuous data:**
The data which uses a measurement scale of length, time or any scale.
The continuous data contains more information than discrete data.
Ex1. How much do you like my PPP?
The answer can be anything, for example on a scale of 1-10 you can say 1 if you don’t like them much, but you can give 10, if you are deeply like them. This is an example of continuous data.
Ex2. Height, Length, weight, diameter etc..

---

**Control Chart Selection**

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>Quality Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>n&gt;1?</td>
<td>x and MR</td>
</tr>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>n&gt;=10 or</td>
<td>x and R</td>
</tr>
<tr>
<td>computer?</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>- x and s</td>
</tr>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>p-chart with variable sample size</td>
</tr>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>p or np</td>
</tr>
<tr>
<td></td>
<td>constant sampling unit?</td>
</tr>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>u</td>
</tr>
</tbody>
</table>
```
Comparison of Variables v. Attributes

- **Variables**
  - Fit certain cases.
  - Both mean and variation information.
  - More expensive!
  - Identify mean shifts sooner before large number nonconforming.

- **Attributes**
  - Fit certain cases – taste, color, etc.
  - Larger sample sizes.
  - Provides summary level performance.
  - Must define nonconformity.

**Statistics – Background (cont.)**

**MEAN:**

The mean (also called the average) is a measure of where the center of your distribution lies. It is simply the sum of all observations divided by the number of observations. (Arithmetic mean)

The mean is strongly influenced by extreme values.

Even though most cities (7 out of 11) had 3 mm or less of rainfall, the mean is close to 4. The extreme value of 10 mm with rainfall for Gaza is affecting the mean quite a bit. Without this observation, the mean would be exactly 3. On the other hand, if you include Gaza with 30 mm of rain instead of 10 in the calculations, the mean would be 5.455, a value that is greater than all but one observation!
Islamic University, Gaza - Palestine

Statistics – The Mean

Suppose that $x_1, x_2, \ldots, x_n$ are the observations in a sample. The most important measure of central tendency in the sample is the sample average.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n} \quad (2.1)$$

Note that the sample average $\bar{x}$ is simply the arithmetic mean of the $n$ observations. The sample average for the metal thickness data in Table 2-2 is:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{45.001}{100} = 450.01 \mu$$

Refer to Fig. 2-3 and note that the sample average is the point at which the histogram exactly “balances.” Thus, the sample average represents the center of mass of the sample data.

Islamic University, Gaza - Palestine

Statistics – Background (cont.)

MEDIAN

The median (also called the 2nd quartile or 50th percentile) is the middle observation in the data set. It is determined by ranking the data and finding observation number $\lfloor N + 1 \rfloor / 2$. In the rainfall data set, there are 11 (non-missing) observations. Thus, the median is the value of the 6th highest (or 6th lowest) observation, which is 3: 1 2 2 3 3 3 4 4 5 10

If there are an even number of observations, the median is extrapolated as the value midway between that of observation numbers $N / 2$ and $[N / 2] + 1$.

The median is less sensitive to extreme values than the mean. For example, the median of this data set would be 3 even if there were 30 mm with rainfall in Gaza instead of 10. Therefore, the median is often used instead of the mean when data contain outliers, or are skewed. (at an angle)

Always arrange the given data in ascending or descending order
The Median Value

The median of a group of numbers is the number in the middle, when the numbers are in order of magnitude. For example, if the set of numbers is 4, 1, 6, 2, 6, 7, 8, the median is 6: 1, 2, 4, 6, 6, 7, 8 (6 is the middle value when the numbers are in order)

If you have n numbers in a group, the median is the \((n + 1)/2\) th value. For example, there are 7 numbers in the example above, so replace n by 7 and the median is the \((7 + 1)/2\) th value = 4th value. The 4th value is 6.

MODE:

The mode is the value in an array of data that is repeated the most. The mode is also a measure of central tendency but is rarely used, as in some cases, chances are there that a single unrepresentative value is also the one that is repeated most often.

For example if our data had been 0, 2, 5, 7, 15, 1, 4, 6, 8, 15
Then our mode would have been 15.

RANGE:

The range is defined as the difference between the highest and the lowest observed values, in a given array of data.

It is very easy to understand but has very limited usefulness as a measure of dispersion.
VARIANCE
The variance of a population signifies the deviation of the data values from the mean value, it is symbolized by . To calculate the population variance we have the following formula:

STANDARD ERROR OF THE MEAN (SE MEAN)
The standard error of the mean (SE Mean) is not often used as a descriptive statistic, but it is important in hypothesis testing. It is an estimate of the dispersion that you would observe in the distribution of sample means, if you continued to take samples of the same size from the population.

The standard error of the mean is the standard deviation divided by N

Statistics - Variance

The variability in the sample data is measured by the sample variance.

\[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \]  

(2-2)

Note that the sample variance is simply the sum of the squared deviations of each observation from the sample average \( \bar{x} \), divided by the sample size minus one. If there is no variability in the sample, then each sample observation \( x_i = \bar{x} \), and the sample variance \( s^2 = 0 \). Generally, the larger is the sample variance \( s^2 \), the greater is the variability in the sample data.
Statistics – Standard deviation

The units of the sample variance $s^2$ are the square of the original units of the data. This is often inconvenient and awkward to interpret, and so we usually prefer to use the square root of $s^2$, called the **sample standard deviation** $s$, as a measure of variability.

It follows that:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (2-3)$$

The primary advantage of the sample standard deviation is that it is expressed in the original units of measurement. For the metal thickness data, we find that

$$s^2 = 180.2928 \text{ Å}^2$$

and

$$s = 13.43 \text{ Å}$$

---

Statistics – Background (cont.)

Cumulative Frequency Distribution is the frequency of a random variable below a particular level. It tells how often the value of the random variable is less than or equal to a particular reference value.

Relative Frequency Distribution is the frequency for that cell divided by the total number of observations.

**Example:**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Cumulative frequency:</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>10 \quad (4 + 6)</td>
</tr>
<tr>
<td>3</td>
<td>13 \quad (4 + 6 + 3)</td>
</tr>
<tr>
<td>2</td>
<td>15 \quad (4 + 6 + 3 + 2)</td>
</tr>
<tr>
<td>6</td>
<td>21 \quad (4 + 6 + 3 + 2 + 6)</td>
</tr>
<tr>
<td>4</td>
<td>25 \quad (4 + 6 + 3 + 2 + 6 + 4)</td>
</tr>
</tbody>
</table>
**Control charts applications**

Either $\bar{X}$ and $R$ or Xbar and $s$ are always used for variables,

- To establish a state statistical process control,
- To monitor a process and a signal when the process goes out of control,
- To determine the process capability,
- To provide a basis for current decisions during production and also to provide the basis for current decision on acceptance or rejection of manufactured product.

---

**Setting up and operating control charts ($\bar{X}$ and $R$)**

Control Charts are often used by management, show what has happening over specified period of time. Control charts have a control limits and central lines on them.

They are accompanied by decision rules to signal to the user when action should be taken on process.

$\bar{X}$ is used to analyze the average of process, and $R$ chart is used to analyze its spread or dispersion.

Together ($\bar{X}$ and $R$ chart) they tell us whether or not the process is stable.

This pair of charts is the most frequently used for monitoring variables data.
After Shewhart, FOUR is the ideal subgroup size but in the industrial use, five seems to be the most common size.

In special cases, subgroups of two or three may often be used, where the cost of measurement is very high.

In other cases, a subgroup of 10 or 20 may be used, where it is desired to make the control chart sensitive to small variation in the process average.

It is practical to use Xbar and s charts rather than Xbar and R charts whenever the subgroup size is greater than 15.

5.2 CONTROL CHARTS FOR $\bar{x}$ AND R

Suppose that a quality characteristic is normally distributed with mean $\mu$ and standard deviation $\sigma$, where both $\mu$ and $\sigma$ are known. If $x_1, x_2, \ldots, x_n$ is a sample of size $n$, then the average of this sample is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

and we know that $\bar{x}$ is normally distributed with mean $\mu$ and standard deviation $\sigma_i = \frac{\sigma}{\sqrt{n}}$. Furthermore, the probability is $1 - \alpha$ that any sample mean will fall between

$$\mu + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

and

$$\mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

(5-1)
Subgroup Data with Unknown $\mu$ and $\sigma$

\[
\bar{y} = \frac{\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n}{n}
\]  

(5.2)

\[
R = \bar{r}_{\text{max}} - \bar{r}_{\text{min}}
\]

Let $R_1, R_2, \ldots, R_n$ be the ranges of the $n$ samples. The average range is

\[
\bar{R} = \frac{R_1 + R_2 + \cdots + R_n}{n}
\]

(5.3)

Control Limits for the $\bar{X}$ Chart

\[
\text{UCL} = \bar{x} + A_2 \bar{R}
\]

Center line $= \bar{x}$

\[
\text{LCL} = \bar{x} - A_2 \bar{R}
\]

(5.4)

The constant $A_2$ is tabulated for various sample sizes in Appendix Table VI.

Control Limits for the $R$ Chart

\[
\text{UCL} = D_4 \bar{R}
\]

Center line $= \bar{R}$

\[
\text{LCL} = D_3 \bar{R}
\]

(5.5)

The constants $D_3$ and $D_4$ are tabulated for various values of $n$ in Appendix Table VI.
\[ \hat{\sigma} = \frac{\bar{R}}{d_2} \quad (5.6) \]

If we use \( \bar{x} \) as an estimator of \( \mu \) and \( \bar{R}/d_i \) as an estimator of \( \sigma \), then the parameters of the \( \bar{x} \) chart are

\[
\begin{align*}
UCL &= \bar{x} + \frac{3}{d_i \sqrt{n}} \bar{R} \\
Center \ line &= \bar{x} \\
LCL &= \bar{x} - \frac{3}{d_i \sqrt{n}} \bar{R}
\end{align*}
\quad (5.7)
\]

If we define

\[ \lambda_2 = \frac{3}{d_i \sqrt{n}} \quad (5.8) \]

then equation 5.7 reduces to equation 5.4.

---

Islamic University, Gaza - Palestine

Now consider the \( R \) chart. The center line will be \( R \). To determine the control limits, we need an estimate of \( \sigma_R \). Assuming that the quality characteristic is normally distributed, \( \sigma_R \) can be found from the distribution of the relative range \( W = R/\sigma \). The standard deviation of \( W \), say \( d_3 \), is a known function of \( n \). Thus, since

\[ R = W \sigma \]

the standard deviation of \( R \) is

\[ \sigma_R = d_3 \sigma \]

Since \( \sigma \) is unknown, we may estimate \( \sigma_R \) by

\[ \hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2} \quad (5.9) \]
Consequently, the parameters of the $R$ chart with the usual three-sigma control limits are:

$$\text{UCL} = \bar{R} + 3\hat{\sigma}_R = \bar{R} + 3d_3 \frac{R}{d_2}$$

Center line $= \bar{R}$  \hspace{1cm} (5-10)

$$\text{LCL} = \bar{R} - 3\hat{\sigma}_R = \bar{R} - 3d_3 \frac{R}{d_2}$$

If we let

$$D_1 = 1 - \frac{d_1}{d_2} \quad \text{and} \quad D_4 = 1 + \frac{d_1}{d_2}$$

equation 5-10 reduces to equation 5-5.

---

**Control vs. Specification Limits**

- Control limits are derived from natural process variability, or the natural tolerance limits of a process.
- Specification limits are determined externally, for example by customers or designers.
- There is no mathematical or statistical relationship between the control limits and the specification limits.
Revision of Control Limits and Center Lines

• Effective use of control charts requires periodic review and revision of control limits and center lines
• Sometimes users replace the center line on the control limit chart with a target value
• When $R$ chart is out of control, out-of-control points are often eliminated to re-compute a revised value of which is used to determine new limits and center line on $R$ chart and new limits on $s$ chart
• Use of control chart for monitoring future production, once a set of reliable limits are established.
• A run chart showing individuals observations in each sample, called a tolerance chart or tier diagram, may reveal patterns or unusual observations in the data

Guidelines for Control Chart Design

• Control chart design requires specification of sample size, control limit width, and sampling frequency
  – Exact solution requires detailed information on statistical characteristics as well as economic factors
  – The problem of choosing sample size and sampling frequency is one of allocating sampling effort
• For $R$ chart, choose as small a sample size is consistent with magnitude of process shift one is trying to detect. For moderate to large shifts, relatively small samples are effective. For small shifts, larger samples are needed.
• For small samples, $R$ chart is relatively insensitive to changes in process standard deviation. For larger samples ($n > 10$ or $12$), $s$ or $s^2$ charts are better choices.
### Control Chart Formulas

<table>
<thead>
<tr>
<th>Type of Chart</th>
<th>LCL</th>
<th>CL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$ (with $R$)</td>
<td>$\bar{x} - A_R \bar{R}$</td>
<td>$\bar{x}$</td>
<td>$\bar{x} + A_R \bar{R}$</td>
</tr>
<tr>
<td>$R$</td>
<td>$D_R \bar{R}$</td>
<td>$\bar{R}$</td>
<td>$D_R \bar{R}$</td>
</tr>
<tr>
<td>$p$</td>
<td>$\bar{p} - 3\sqrt{\bar{p}(1 - \bar{p})/n}$</td>
<td>$\bar{p}$</td>
<td>$\bar{p} + 3\sqrt{\bar{p}(1 - \bar{p})/n}$</td>
</tr>
<tr>
<td>$\bar{x}$ (with $s$)</td>
<td>$\bar{x} - A_s \bar{s}$</td>
<td>$\bar{x}$</td>
<td>$\bar{x} + A_s \bar{s}$</td>
</tr>
<tr>
<td>$s$</td>
<td>$\bar{s}$</td>
<td>$s$</td>
<td>$\bar{s}$</td>
</tr>
<tr>
<td>$\bar{x}$</td>
<td>$\bar{x} - 3\bar{R}_{d_2}$</td>
<td>$\bar{x}$</td>
<td>$\bar{x} + 3\bar{R}_{d_2}$</td>
</tr>
<tr>
<td>$np$</td>
<td>$np - 3\sqrt{np(1 - \bar{p})}$</td>
<td>$np$</td>
<td>$np + 3\sqrt{np(1 - \bar{p})}$</td>
</tr>
<tr>
<td>$c$</td>
<td>$\bar{c} - 3\sqrt{\bar{c}}$</td>
<td>$\bar{c}$</td>
<td>$\bar{c} + 3\sqrt{\bar{c}}$</td>
</tr>
<tr>
<td>$u$</td>
<td>$\bar{u} - 3\sqrt{\bar{u}/n}$</td>
<td>$\bar{u}$</td>
<td>$\bar{u} + 3\sqrt{\bar{u}/n}$</td>
</tr>
</tbody>
</table>

### End of Chapter 6