

## Reliability and Quality Control

46.1	Reliability	2
46.2	Quality Control	21

### *Learning outcomes*

*You will first learn about the importance of the concept of reliability applied to systems and products. The second Section introduces you to the very important topic of quality control in production processes. In both cases you will learn how to perform the basic calculations necessary to use each topic in practice.*

# Reliability

# 46.1



## Introduction

Much of the theory of reliability was developed initially for use in the electronics industry where components often fail with little if any prior warning. In such cases the hazard function or conditional failure rate function is constant and any functioning component or system is taken to be 'as new'. There are other cases where the conditional failure rate function is time dependent, often proportional to the time that the system or component has been in use. The function may be an increasing function of time (as with random vibrations for example) or decreasing with time (as with concrete whose strength, up to a point, increases with time). If we can develop a lifetime model, we can use it to plan such things as maintenance and part replacement schedules, whole system replacements and reliability testing schedules.



## Prerequisites

Before starting this Section you should ...

- be familiar with the results and concepts met in the study of probability
- understand and be able to use continuous probability distributions



## Learning Outcomes

On completion you should be able to ...

- appreciate the importance of lifetime distributions
- complete reliability calculations for simple systems
- state the relationship between the Weibull distribution and the exponential distribution

# 1. Reliability

## Lifetime distributions

From an engineering point of view, the ability to predict the lifetime of a whole system or a system component is very important. Such lifetimes can be predicted using a statistical approach using appropriate distributions. Common examples of structures whose lifetimes we need to know are airframes, bridges, oil rigs and at a simpler, less catastrophic level, system components such as cycle chains, engine timing belts and components of electronic systems such as televisions and computers. If we can develop a lifetime model, we can use it to plan such things as maintenance and part replacement schedules, whole system replacements and reliability testing schedules.

We start by looking at the length of use of a system or component prior to failure (the age prior to failure) and from this develop a definition of reliability. Lifetime distributions are functions of time and may be expressed as probability density functions and so we may write

$$F(t) = \int_0^t f(t) dt$$

This represents the probability that the system or component fails anywhere between 0 and  $t$ .



### Key Point 1

The probability that a system or component will fail only after time  $t$  may be written as

$$R(t) = 1 - F(t)$$

The function  $R(t)$  is usually called the reliability function.

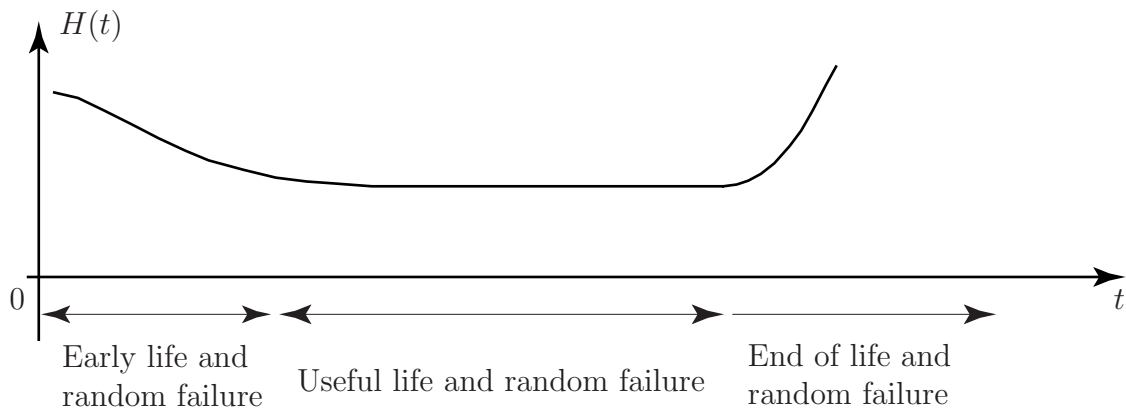
In practice engineers (and others!) are often interested in the so-called hazard function or conditional failure rate function which gives the probability that a system or component fails after it has been in use for a given time. This function may be defined as

$$\begin{aligned} H(t) &= \lim_{\Delta t \rightarrow 0} \left( \frac{P(\text{failure in the interval } (t, t + \Delta t)) / \Delta t}{P(\text{survival up to time } t)} \right) \\ &= \frac{1}{R(t)} \lim_{\Delta t \rightarrow 0} \left( \frac{F(t + \Delta t) - F(t)}{\Delta t} \right) \\ &= \frac{1}{R(t)} \frac{d}{dt} F(t) \\ &= \frac{f(t)}{R(t)} \end{aligned}$$

This gives the conditional failure rate function as

$$\begin{aligned}
 H(t) &= \lim_{\Delta t \rightarrow 0} \left( \frac{\int_t^{t+\Delta t} f(t) dt / (\Delta t)}{R(t)} \right) \\
 &= \frac{f(t)}{R(t)}
 \end{aligned}$$

Essentially we are describing the rate of failure per unit time for (say) mechanical or electrical components which have already been in service for a given time. A graph of  $H(t)$  often shows a high initial failure rate followed by a period of relative reliability followed by a period of increasingly high failure rates as a system or component ages. A typical graph (sometimes called a bathtub graph) is shown below.



**Figure 1**

Note that 'early life and random failure' includes failure due to defects being present and that 'end of life and random failure' includes failure due to ageing.

The reliability of a system or component may be defined as the probability that the system or component functions for a given time, that is, the probability that it will fail only after the given time. Put another way,  $R(t)$  is the probability that the system or component is still functioning at time  $t$ .

### The exponential distribution

We have already met the exponential distribution in the form

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0$$

However, one of the simplest distributions describing failure is the exponential distribution

$$f(t) = \frac{1}{\mu} e^{-t/\mu}, \quad t \geq 0$$

where, in this case,  $\mu$  is the mean time to failure. One property of this distribution is that the hazard function is a constant independent of time - the 'good as new' syndrome mentioned above. To show that the probability of failure is independent of age consider the following.

$$F(t) = \int_0^t \frac{1}{\mu} e^{-t/\mu} dt = \frac{1}{\mu} \left[ -\mu e^{-t/\mu} \right]_0^t = 1 - e^{-t/\mu}$$

Hence the reliability (given that the total area under the curve  $F(t) = 1 - e^{-t/\mu}$  is unity) is

$$R(t) = 1 - F(t) = e^{-t/\mu}$$

Hence, the hazard function or conditional failure rate function  $H(t)$  is given by

$$H(t) = \frac{F(t)}{R(t)} = \frac{\frac{1}{\mu}e^{-t/\mu}}{e^{-t/\mu}} = \frac{1}{\mu}$$

which is a constant independent of time.

Another way of looking at this is to consider the probability that failure occurs in the interval  $(\tau, \tau+t)$  given that the system is functioning at time  $\tau$ . This probability is

$$\begin{aligned} \frac{F(\tau+t) - F(\tau)}{R(\tau)} &= \frac{(1 - e^{-(\tau+t)/\mu}) - (1 - e^{-\tau/\mu})}{e^{-\tau/\mu}} \\ &= 1 - e^{-t/\mu} \end{aligned}$$

This is just the probability that failure occurs in the interval  $(0, t)$  and implies that ageing has no effect on failure. This is sometimes referred to as the 'good as new syndrome.'

It is worth noting that in the modelling of many complex systems it is assumed that only random component failures are important. This enables us to assume the use of the exponential distribution since initial failures are removed by a 'running-in' process and the time to ultimate failure is usually long.



### Example 1

The lifetime of a modern low-wattage electronic light bulb is known to be exponentially distributed with a mean of 8000 hours.

Find the proportion of bulbs that may be expected to fail before 7000 hours use.

#### Solution

We know that  $\mu = 8000$  and so

$$\begin{aligned} F(t) &= \int_0^t \frac{1}{8000} e^{-t/8000} dt \\ &= \frac{1}{8000} \left[ -8000 e^{-t/8000} \right]_0^t \\ &= 1 - e^{-t/8000} \end{aligned}$$

Hence  $F(7000) = 1 - e^{-7000/8000} = 1 - e^{-0.675} = 0.4908$  and we expect that about 49% of the bulbs will fail before 7000 hours of use.



A particular electronic device will only function correctly if two essential components both function correctly. The lifetime of the first component is known to be exponentially distributed with a mean of 6000 hours and the lifetime of the second component is known to be exponentially distributed with a mean of 7000 hours. Find the proportion of devices that may be expected to fail before 8000 hours use. State clearly any assumptions you make.

### Your solution

### Answer

The assumption made is that the components operate independently.

For the first component  $F(t) = 1 - e^{-t/6000}$  so that  $F(8000) = 1 - e^{-8000/6000} = 1 - e^{-4/3} = 0.7364$

For the second component  $F(t) = 1 - e^{-t/7000}$  so that  $F(8000) = 1 - e^{-8000/7000} = 0.6811$

The probability that the device will continue to function after 8000 hours use is given by an expression of the form  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Hence the probability that the device will continue to function after 8000 hour use is

$$0.7364 + 0.6811 - 0.7364 \times 0.6811 = 0.916$$

and we expect just under 92% of the devices to fail before 8000 hours use.

An alternative answer may be obtained more directly by using the reliability function  $R(t)$ :

The assumption made is that the components operate independently.

For the first component  $R(t) = e^{-t/6000}$  so that  $R(8000) = e^{-8000/6000} = e^{-4/3} = 0.2636$

For the second component  $R(t) = e^{-t/7000}$  so that  $R(8000) = e^{-8000/7000} = 0.3189$

The probability that the device will continue to function after 8000 hour use is given by

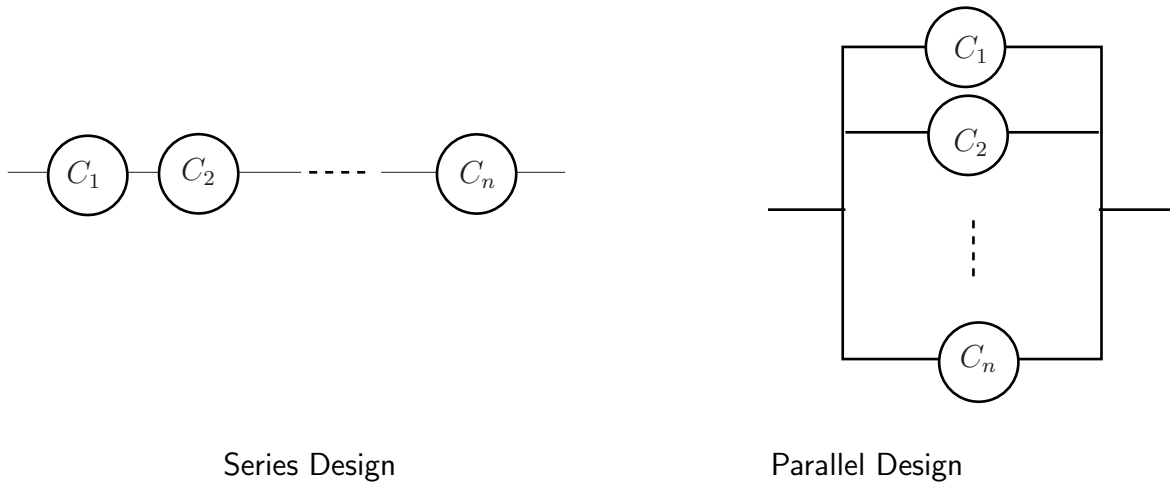
$$0.2636 \times 0.3189 = 0.0841$$

Hence the probability that the device will fail before 8000 hours use is  $1 - 0.0841 = 0.916$

and we expect just under 92% of the devices to fail before 8000 hours use.

## 2. System reliability

It is reasonable to ask whether, in designing a system, an engineer should design a system using components in series or in parallel. The engineer may not have a choice of course! We may represent a system consisting of  $n$  components say  $C_1, C_2, \dots, C_n$  with reliabilities (these are just probability values)  $R_1, R_2, \dots, R_n$  respectively as series and parallel systems as shown below.



**Figure 2**

With a series design, the system will fail if any component fails. With a parallel design, the system will work as long as any component works.

Assuming that the components are independent, we can express the reliability of the series design as

$$R_{\text{Series}} = R_1 \times R_2 \times \dots \times R_n$$

simply by multiplying the probabilities.

Since each reliability value is less than one, we may conclude that a series design is less reliable than its *least reliable* component.

Similarly (although by no means as clearly!), we can express the reliability of the parallel design as

$$R_{\text{Parallel}} = 1 - (1 - R_1)(1 - R_2) \dots (1 - R_n)$$

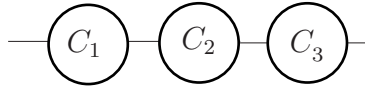
The derivation of this result is illustrated in Example 3 below for the case  $n = 3$ . In this case, the algebra involved is fairly straightforward. We can conclude that the parallel design is at least as reliable as the most reliable component.

Engineers will sometimes include 'redundant' components in parallel to improve reliability. The spare wheel of a car is a well known example.



## Example 2 Series design

Consider the three components  $C_1, C_2$  and  $C_3$  with reliabilities  $R_1, R_2$  and  $R_3$  connected in series as shown below



Find the reliability of the system where  $R_1 = 0.2$ ,  $R_2 = 0.3$  and  $R_3 = 0.4$ .

### Solution

Since the components are assumed to act independently, we may clearly write

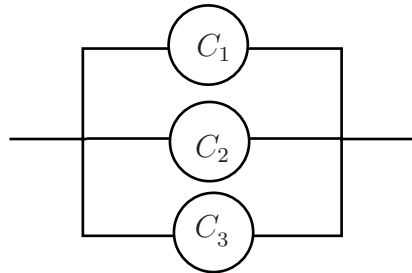
$$R_{\text{Series}} = R_1 \times R_2 \times R_3$$

Taking  $R_1 = 0.2$ ,  $R_2 = 0.3$  and  $R_3 = 0.4$  we obtain the value  $R_{\text{Series}} = 0.2 \times 0.3 \times 0.4 = 0.024$



## Example 3 Parallel design

Consider the three components  $C_1, C_2$  and  $C_3$  with reliabilities  $R_1, R_2$  and  $R_3$  connected in parallel as shown below



Find the reliability of the system where  $R_1 = 0.2$ ,  $R_2 = 0.3$  and  $R_3 = 0.4$ .

### Solution

Observing that  $F_i = 1 - R_i$ , where  $F_i$  represents the failure of the  $i$ th component and  $R_i$  represents the reliability of the  $i$ th component we may write

$$F_{\text{System}} = F_1 F_2 F_3 \rightarrow R_{\text{System}} = 1 - F_1 F_2 F_3 = 1 - (1 - R_1)(1 - R_2)(1 - R_3)$$

Again taking  $R_1 = 0.2$ ,  $R_2 = 0.3$  and  $R_3 = 0.4$  we obtain

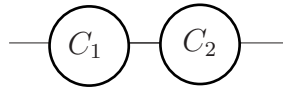
$$F_{\text{System}} = 1 - (1 - 0.2)(1 - 0.3)(1 - 0.4) = 1 - 0.336 = 0.664$$



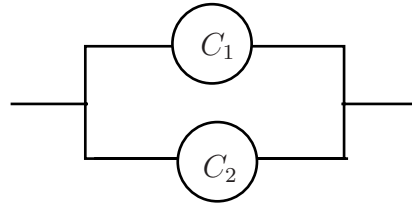
Hence series system reliability is *less* than any of the component reliabilities and parallel system reliability is *greater* than any of the component reliabilities.



Consider the two components  $C_1$  and  $C_2$  with reliabilities  $R_1$  and  $R_2$  connected in series and in parallel as shown below. Assume that  $R_1 = 0.3$  and  $R_2 = 0.4$ .



Series configuration



Parallel configuration

Let  $R_{\text{Series}}$  be the reliability of the series configuration and  $R_{\text{Parallel}}$  be the reliability of the parallel configuration

- Why would you expect that  $R_{\text{Series}} < 0.3$  and  $R_{\text{Parallel}} > 0.4$ ?
- Calculate  $R_{\text{Series}}$
- Calculate  $R_{\text{Parallel}}$

### Your solution

### Answer

- You would expect  $R_{\text{Series}} < 0.3$  and  $R_{\text{Parallel}} > 0.4$  because  $R_{\text{Series}}$  is less than any of the component reliabilities and  $R_{\text{Parallel}}$  is greater than any of the component reliabilities.
- $R_{\text{Series}} = R_1 \times R_2 = 0.3 \times 0.4 = 0.12$
- $R_{\text{Parallel}} = R_1 + R_2 - R_1 R_2 = 0.3 + 0.4 - 0.3 \times 0.4 = 0.58$

### 3. The Weibull distribution

The Weibull distribution was first used to describe the behaviour of light bulbs as they undergo the ageing process. Other applications are widespread and include the description of structural failure, ball-bearing failure and the failure of a variety of electronic components.



#### Key Point 2

If the random variable  $X$  is defined by a probability density function of the form

$$f(x) = \alpha\beta(\alpha x)^{\beta-1}e^{-(\alpha x)^\beta}$$

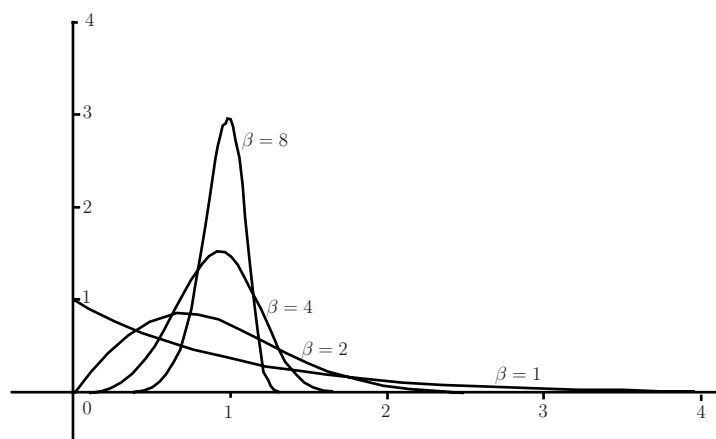
then  $X$  is said to follow a Weibull distribution.

The hazard function or conditional failure rate function  $H(t)$ , which gives the probability that a system or component fails after it has been in use for a given time, is constant for the exponential distribution but for a Weibull distribution is proportional to  $x^{\beta-1}$ . This implies that  $\beta = 1$  gives a constant hazard,  $\beta < 1$  gives a reducing hazard and  $\beta > 1$  gives an increasing hazard. Note that  $\alpha$  is simply a scale factor while the case  $\beta = 1$  reduces the Weibull distribution to

$$f(x) = \alpha e^{-\alpha x}$$

which you may recognize as one form of the exponential distribution.

Figure 3 below shows the Weibull distribution for various values of  $\beta$ . For simplicity, the graphs assume that  $\alpha = 1$ . Essentially the plots are of the function  $f(x) = \beta x^{\beta-1} e^{-x^\beta}$



The Weibull distribution with  $\beta = 1, 2, 4$  and  $8$

Figure 3

In applications of the Weibull distribution, it is often useful to know the cumulative distribution function (cdf). Although it is not derived directly here, the cdf  $F(x)$  of the Weibull distribution, whose probability density function is  $f(x) = \alpha\beta(\alpha x)^{\beta-1}e^{-(\alpha x)^\beta}$ , is given by the function

$$F(x) = 1 - e^{-(\alpha x)^\beta}$$

The relationship between  $f(x)$  and  $F(x)$  may be seen by remembering that

$$f(x) = \frac{d}{dx}F(x)$$

Differentiating the cdf  $F(x) = 1 - e^{-(\alpha x)^\beta}$  gives the result

$$f(x) = \frac{d}{dx}F(x) = \frac{(\alpha x)^\beta \beta e^{-(\alpha x)^\beta}}{x} = \alpha^\beta \beta x^{\beta-1} e^{-(\alpha x)^\beta} = \alpha \beta (\alpha x)^{\beta-1} e^{-(\alpha x)^\beta}$$

## Mean and variance of the Weibull distribution

The mean and variance of the Weibull distribution are quite complicated expressions and involve the use of the Gamma function,  $\Gamma(x)$ . The outline explanations given below defines the Gamma function and shows that basically in terms of the application here, the function can be expressed as a simple factorial.

It can be shown that if the random variable  $X$  follows a Weibull distribution defined by the probability density function

$$f(x) = \alpha \beta (\alpha x)^{\beta-1} e^{-(\alpha x)^\beta}$$

then the mean and variance of the distribution are given by the expressions

$$E(X) = \frac{1}{\alpha} \Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{and} \quad V(X) = \left(\frac{1}{\alpha}\right)^2 \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2 \right\}$$

where  $\Gamma(r)$  is the gamma function defined by

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$$

Using integration by parts gives a fairly straightforward identity satisfied by the gamma function is:

$$\Gamma(r) = (r-1)\Gamma(r-1)$$

Using this relationship tells us that

$$\Gamma(r) = (r-1)(r-2)(r-3)\dots\Gamma(1) \quad \text{for integral } r$$

but

$$\Gamma(1) = \int_0^\infty x^{1-1} e^{-x} dx = \int_0^\infty e^{-x} dx = \left[ -e^{-x} \right]_0^\infty = 1$$

so that

$$\Gamma(r) = (r-1)(r-2)(r-3)\dots(3)(2)(1) = (r-1)! \quad \text{for integral } r$$

Essentially, this means that when we calculate the mean and variance of the Weibull distribution using the expressions

$$E(X) = \frac{1}{\alpha} \Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{and} \quad V(X) = \left(\frac{1}{\alpha}\right)^2 \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2 \right\}$$

we are evaluating factorials and the calculation is not as difficult as you might imagine at first sight.

Note that to apply the work as presented here, the expressions

$$\left(1 + \frac{1}{\beta}\right) \quad \text{and} \quad \left(1 + \frac{2}{\beta}\right)$$

must both take positive integral values (i.e.  $\frac{1}{\beta}$  must be an integer) in order for the factorial calculation to make sense. If the above expressions do not assume positive integral values, you will need to know much more about the values and behaviour of the Gamma function in order to do calculations. In practice, we would use a computer of course. As you might expect, the Gamma function is tabulated to assist in such calculations but in the examples and exercises set in this Workbook, the values of  $\frac{1}{\beta}$  will always be integral.



#### Example 4

The main drive shaft of a pumping engine runs in two bearings whose lifetime follows a Weibull distribution with random variable  $X$  and parameters  $\alpha = 0.0002$  and  $\beta = 0.5$ .

- (a) Find the expected time that a single bearing runs before failure.
- (b) Find the probability that any one bearing lasts at least 8000 hours.
- (c) Find the probability that both bearings last at least 6000 hours.

#### Solution

(a) We know that  $E(X) = \frac{1}{\alpha} \Gamma\left(1 + \frac{1}{\beta}\right) = 5000 \times \Gamma(1 + 2) = 5000 \times 2 = 10000$  hours.

(b) We require  $P(X > 8000)$ , this is given by the calculation:

$$\begin{aligned} P(X > 8000) &= 1 - F(8000) = 1 - (1 - e^{-(0.0002 \times 8000)^{0.5}}) \\ &= e^{-(0.0002 \times 8000)^{0.5}} = e^{-1.265} \\ &= 0.282 \end{aligned}$$

(c) Assuming that the bearings wear independently, the probability that both bearings last at least 6000 hours is given by  $\{P(X > 6000)\}^2$ . But  $P(X > 6000)$  is given by the calculation

$$P(X > 6000) = 1 - F(6000) = 1 - (1 - e^{-(0.0002 \times 6000)^{0.5}}) = 0.335$$

so that the probability that both bearings last at least 6000 hours is given by

$$\{P(X > 6000)\}^2 = 0.335^2 = 0.112$$



A shaft runs in four roller bearings the lifetime of each of which follows a Weibull distribution with parameters  $\alpha = 0.0001$  and  $\beta = 1/3$ .

- Find the mean life of a bearing.
- Find the variance of the life of a bearing.
- Find the probability that all four bearings last at least 50,000 hours. State clearly any assumptions you make when determining this probability.

### Your solution

### Answer

(a) We know that  $E(X) = \frac{1}{\alpha} \Gamma(1 + \frac{1}{\beta}) = 10000 \times \Gamma(1 + 3) = 60000 \times 2 = 10000$  hours.

(b) We know that  $V(X) = \left(\frac{1}{\alpha}\right) \left\{ \Gamma(1 + \frac{2}{\beta}) - \left(\Gamma(1 + \frac{1}{\beta})\right)^2 \right\}$  so that

$$\begin{aligned} V(X) &= (10000)^2 \{ \Gamma(1 + 6) - (\Gamma(1 + 3))^2 \} \\ &= (10000)^2 \{ \Gamma(7) - (\Gamma(4))^2 \} \\ &= (10000)^2 \{ 6! - (3!)^2 \} = 684(10000)^2 = 6.84 \times 10^{10} \end{aligned}$$

(c) The probability that one bearing will last at least 50,000 hours is given by the calculation

$$\begin{aligned} P(X > 50000) &= 1 - F(50000) \\ &= 1 - (1 - e^{-(0.0001 \times 50000)^{0.5}}) \\ &= e^{-5^{0.5}} = 0.107 \end{aligned}$$

Assuming that the four bearings have independent distributions, the probability that all four bearings last at least 50,000 hours is  $(0.107)^4 = 0.0001$

## Exercises

1. The lifetimes in hours of certain machines have Weibull distributions with probability density function

$$f(t) = \begin{cases} 0 & (t < 0) \\ \alpha\beta(\alpha t)^{\beta-1} \exp\{-(\alpha t)^\beta\} & (t \geq 0) \end{cases}$$

- (a) Verify that  $\int_0^t \alpha\beta(\alpha u)^{\beta-1} \exp\{-(\alpha u)^\beta\} du = 1 - \exp\{-(\alpha t)^\beta\}$ .
- (b) Write down the distribution function of the lifetime.
- (c) Find the probability that a particular machine is still working after 500 hours of use if  $\alpha = 0.001$  and  $\beta = 2$ .
- (d) In a factory,  $n$  of these machines are installed and started together. Assuming that failures in the machines are independent, find the probability that all of the machines are still working after  $t$  hours and hence find the probability density function of the time till the first failure among the machines.

2. If the lifetime distribution of a machine has hazard function  $h(t)$ , then we can find the reliability function  $R(t)$  as follows. First we find the "cumulative hazard function"  $H(t)$  using

$$H(t) = \int_0^t h(t) dt \quad \text{then the reliability is} \quad R(t) = e^{-H(t)}.$$

- (a) Starting with  $R(t) = \exp\{-H(t)\}$ , work back to the hazard function and hence confirm the method of finding the reliability from the hazard.
- (b) A lifetime distribution has hazard function  $h(t) = \theta_0 + \theta_1 t + \theta_2 t^2$ . Find
- (i) the reliability function.
  - (ii) the probability density function.
- (c) A lifetime distribution has hazard function  $h(t) = \frac{1}{t+1}$ . Find
- (i) the reliability function;
  - (ii) the probability density function;
  - (iii) the median.

What happens if you try to find the mean?

- (d) A lifetime distribution has hazard function  $h(t) = \frac{\rho\theta}{\rho t + 1}$ , where  $\rho > 0$  and  $\theta > 1$ . Find
- (i) the reliability function.
  - (ii) the probability density function.
  - (iii) the median.
  - (iv) the mean.

## Exercises continued

3. A machine has  $n$  components, the lifetimes of which are independent. However the whole machine will fail if any component fails. The hazard functions for the components are  $h_1(t), \dots, h_n(t)$ . Show that the hazard function for the machine is  $\sum_{i=1}^n h_i(t)$ .
4. Suppose that the lifetime distributions of the components in Exercise 3 are Weibull distributions with scale parameters  $\rho_1, \dots, \rho_n$  and a common index (i.e. "shape parameter")  $\gamma$  so that the hazard function for component  $i$  is  $\gamma\rho_i(\rho_i t)^{\gamma-1}$ . Find the lifetime distribution for the machine.
5. (*Difficult*). Show that, if  $T$  is a Weibull random variable with hazard function  $\gamma\rho(\rho t)^{\gamma-1}$ ,
- (a) the median is  $M(T) = \rho^{-1}(\ln 2)^{1/\gamma}$ ,
  - (b) the mean is  $E(T) = \rho^{-1}\Gamma(1 + \gamma^{-1})$
  - (c) the variance is  $V(T) = \rho^{-2}\{\Gamma(1 + 2\gamma^{-1}) - [\Gamma(1 + \gamma^{-1})]^2\}$ .

Note that  $\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx$ . In the integrations it is helpful to use the substitution  $u = (\rho t)^\gamma$ .

## Answers

1.

$$(a) \quad \frac{d}{du}(-\exp\{-(\alpha u)^\beta\})$$

so

$$\int_0^t \alpha\beta(\alpha u)^{\beta-1} \exp\{-(\alpha u)^\beta\} du = \left[ -\exp\{-(\alpha u)^\beta\} \right]_0^t = 1 - \exp\{-(\alpha t)^\beta\}.$$

(b) The distribution function is  $F(t) = 1 - \exp\{-(\alpha t)^\beta\}$ .

$$(c) \quad R(500) = \exp\{-(0.001 \times 500)^\beta\} = 0.7788$$

(d) The probability that all of the machines are still working after  $t$  hours is

$$R_n(t) = \{R(t)\}^n = \exp\{-(\alpha t)^{n\beta}\}.$$

Hence the time to the first failure has a Weibull distribution with  $\beta$  replaced by  $n\beta$ . The pdf is

$$f_n(t) = \begin{cases} 0 & (t < 0) \\ \alpha n\beta(\alpha t)^{n\beta-1} \exp\{-(\alpha t)^{n\beta}\} & (t \geq 0) \end{cases}$$

2.

(a) The distribution function is

$$F(t) = 1 - R(t) = 1 - e^{-H(t)} \quad \text{so the pdf is} \quad f(t) = \frac{d}{dt}F(t) = h(t)e^{-H(t)}.$$

Hence the hazard function is

$$h(t) = \frac{f(t)}{R(t)} = \frac{h(t)e^{-H(t)}}{e^{-H(t)}} = h(t).$$

(b) A lifetime distribution has hazard function  $h(t) = \theta_0 + \theta_1 t + \theta_2 t^2$ .

(i) Reliability  $R(t)$ .

$$\begin{aligned} H(t) &= \left[ \theta_0 t + \theta_1 t^2/2 + \theta_2 t^3/3 \right]_0^t \\ &= \theta_0 t + \theta_1 t^2/2 + \theta_2 t^3/3 \\ R(t) &= \exp\{-(\theta_0 t + \theta_1 t^2/2 + \theta_2 t^3/3)\}. \end{aligned}$$

(ii) Probability density function  $f(t)$ .

$$\begin{aligned} F(t) &= 1 - R(t) \\ f(t) &= (\theta_0 + \theta_1 t + \theta_2 t^2) \exp\{-(\theta_0 t + \theta_1 t^2/2 + \theta_2 t^3/3)\}. \end{aligned}$$



**Answers**

(c) A lifetime distribution has hazard function  $h(t) = \frac{1}{t+1}$ .

(i) Reliability function  $R(t)$ .

$$H(t) = \int_0^t \frac{1}{u+1} du = \left[ \ln(u+1) \right]_0^t = \ln(t+1)$$

$$R(t) = \exp\{-H(t)\} = \frac{1}{t+1}$$

(ii) Probability density function  $f(t)$ .  $F(t) = 1 - R(t)$ ,  $f(t) = \frac{1}{(t+1)^2}$

(iii) Median  $M$ .  $\frac{1}{M+1} = \frac{1}{2}$  so  $M = 1$ .

To find the mean:  $E(T+1) = \int_0^\infty \frac{1}{t+1} dt$  which does not converge, so neither does  $E(T)$ .

(d) A lifetime distribution has hazard function

$$h(t) = \frac{\rho\theta}{\rho t + 1}, \quad \text{where } \rho > 0 \text{ and } \theta > 1.$$

(i) Reliability function  $R(t)$ .

$$H(t) = \int_0^t \frac{\rho\theta}{\rho t + 1} dt = \left[ \theta \ln(\rho t + 1) \right]_0^t = \theta \ln(\rho t + 1)$$

$$R(t) = \exp\{-\theta \ln(\rho t + 1)\} = (\rho t + 1)^{-\theta}$$

(ii) Probability density function  $f(t)$ .

$$F(t) = 1 - R(t)$$

$$f(t) = \theta\rho(\rho t + 1)^{-(1+\theta)}$$

(iii) Median  $M$ .

$$(\rho M + 1)^{-\theta} = 2^{-1}$$

$$\rho M + 1 = 2^{1/\theta} \quad \therefore \quad M = \rho^{-1}(2^{1/\theta} - 1)$$

(iv) Mean.

$$E(\rho T + 1) = \theta\rho \int_0^\infty (\rho t + 1)^{-\theta} dt$$

$$= \frac{\theta}{\theta - 1} \left[ -(\rho t + 1)^{1-\theta} \right]_0^\infty = \frac{\theta}{\theta - 1}$$

$$E(T) = \rho^{-1} \left( \frac{\theta}{\theta - 1} - 1 \right) = \rho^{-1}(\theta - 1)^{-1}$$

## Answers

3. For each component

$$H_i(t) = \int_0^t h_i(t) dt$$
$$R_i(t) = \exp\{-H_i(t)\}$$

For machine

$$R(t) = \prod_{i=1}^n R_i(t) = \exp\left\{-\sum_{i=1}^n H_i(t)\right\}$$

the probability that all components are still working.

$$F(t) = 1 - R(t)$$
$$f(t) = \exp\left\{-\sum_{i=1}^n H_i(t)\right\} \sum_{i=1}^n h_i(t)$$
$$h(t) = f(t)/R(t) = \sum_{i=1}^n h_i(t).$$

4. For each component

$$h_i(t) = \gamma \rho_i (\rho_i t)^{\gamma-1}$$
$$H_i(t) = \int_0^t \gamma \rho_i (\rho_i t)^{\gamma-1} dt = (\rho_i t)^\gamma$$

Hence, for the machine,

$$\sum_{i=1}^n H_i(t) = t^\gamma \sum_{i=1}^n \rho_i^\gamma$$
$$R(t) = \exp\left(-t^\gamma \sum_{i=1}^n \rho_i^\gamma\right)$$
$$F(t) = 1 - R(t)$$
$$f(t) = \gamma t^{\gamma-1} \sum_{i=1}^n \rho_i^\gamma \exp\left(-t^\gamma \sum_{i=1}^n \rho_i^\gamma\right)$$
$$h(t) = \gamma \rho (\rho t)^{\gamma-1}$$

so we have a Weibull distribution with index  $\gamma$  and scale parameter  $\rho$  such that

$$\rho^\gamma = \sum_{i=1}^n \rho_i^\gamma.$$

**Answers**

5. Suppose the distribution function is

$$F(t) = 1 - e^{-(\rho t)^\gamma}.$$

Then the pdf is

$$f(t) = \frac{d}{dt}F(t) = \gamma\rho(\rho t)^{\gamma-1}e^{-(\rho t)^\gamma}$$

and the hazard function is

$$H(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)} = \gamma\rho(\rho t)^{\gamma-1}$$

as required.

(a) At the median,  $M$ ,  $F(M) = 0.5$ . So

$$\begin{aligned} 1 - e^{-(\rho M)^\gamma} &= 0.5 \\ e^{-(\rho M)^\gamma} &= 0.5 \\ (\rho M)^\gamma &= -\ln(0.5) = \ln(2) \\ \rho M &= (\ln 2)^{1/\gamma} \\ M &= \rho^{-1}(\ln 2)^{1/\gamma} \end{aligned}$$

(b) The mean is

$$\begin{aligned} E(T) &= \int_0^\infty t f(t) dt = \int_0^\infty t \gamma\rho(\rho t)^{\gamma-1}e^{-(\rho t)^\gamma} dt \\ &= \int_0^\infty \gamma(\rho t)^\gamma e^{-(\rho t)^\gamma} dt \end{aligned}$$

Let  $u = (\rho t)^\gamma$ , so  $du/dt = \gamma\rho(\rho t)^{\gamma-1}$  and  $du = \gamma\rho(\rho t)^{\gamma-1} dt$ . Then

$$\begin{aligned} E(T) &= \int_0^\infty \rho^{-1}u^{1/\gamma}e^{-u} du \\ &= \rho^{-1} \int_0^\infty u^{1+1/\gamma-1}e^{-u} du \\ &= \rho^{-1}\Gamma(1 + 1/\gamma) \end{aligned}$$

## Answers

5. continued

(c) To find the variance we first find

$$\begin{aligned} E(T^2) &= \int_0^{\infty} t^2 f(t) dt = \int_0^{\infty} t^2 \gamma \rho (\rho t)^{\gamma-1} e^{-(\rho t)^\gamma} dt \\ &= \int_0^{\infty} \gamma \rho^{-1} (\rho t)^{\gamma+1} e^{-(\rho t)^\gamma} dt \end{aligned}$$

Let  $u = (\rho t)^\gamma$ , so  $du/dt = \gamma \rho (\rho t)^{\gamma-1}$  and  $du = \gamma \rho (\rho t)^{\gamma-1} dt$ . Then

$$\begin{aligned} E(T^2) &= \int_0^{\infty} \rho^{-2} u^{2/\gamma} e^{-u} du \\ &= \rho^{-2} \int_0^{\infty} u^{1+2/\gamma-1} e^{-u} du \\ &= \rho^{-2} \Gamma(1 + 2/\gamma) \end{aligned}$$

So

$$\begin{aligned} V(T) &= E(T^2) - \{E(T)\}^2 \\ &= \rho^{-2} \Gamma(1 + 2/\gamma) - \rho^{-2} \{\Gamma(1 + 1/\gamma)\}^2 \\ &= \rho^{-2} \{\Gamma(1 + 2/\gamma) - \{\Gamma(1 + 1/\gamma)\}^2\} \end{aligned}$$

# Quality Control

## 46.2

### Introduction

Quality control via the use of statistical methods is a very large area of study in its own right and is central to success in modern industry with its emphasis on reducing costs while at the same time improving quality. In recent times, the Japanese have been extremely successful at applying statistical methods to industrial quality control and have gained a significant advantage over many of their competitors. One need only think of the reputations enjoyed by Japanese motor, camera, TV, video, DVD and general electronics manufacturers to realize just how successful they have become. It is because of the global nature of modern industrial competition that quality control, or more precisely, statistical quality control has become an area of central importance to engineers. Manufacturing a product that the public wants to buy is no longer good enough. The product must be of sufficiently high quality and sufficiently competitive price-wise that it is preferred to its competitors. Without statistical quality control methods it is extremely difficult, if not impossible, to either attain or maintain a truly competitive position.



### Prerequisites

Before starting this Section you should ...

- be able to calculate means and standard deviations from given data sets
- be able to estimate population means and standard deviations from samples



### Learning Outcomes

On completion you should be able to ...

- state what is meant by the term statistical quality control.
- explain why different types of control charts are necessary.
- construct and interpret a small variety of control charts, in particular those based on means and ranges.
- describe in outline the relationship between hypothesis testing and statistical quality control.

# 1. Quality control

## Background

Techniques and methods for checking the quality of materials and the building of houses, temples, monuments and roads have been used over the centuries. For example, the ancient Egyptians had to make and use precise measurements and adopt very high standards of work in order to build the Pyramids. In the Middle Ages, the Guilds were established in Europe to ensure that new entrants to the craft/trade maintained standards. The newcomer was required to serve a long period of apprenticeship under the supervision of a master craftsman, and had to demonstrate his ability to produce work of the appropriate quality and standard before becoming a recognized tradesman. In modern times the notion of quality has evolved through the stages outlined below.

## Inspection

The Industrial Revolution introduced mass production techniques to the workplace. By the end of the 19th century, production processes were becoming more complex and it was beyond the capabilities of a single individual to be responsible for all aspects of production. It is impossible to inspect quality into a product in the sense that a faulty product cannot be put right by means of inspection alone. Statistical quality control can and does provide the environment within which the product is manufactured correctly the first time. A process called acceptance sampling improves the average quality of the items accepted by rejecting those items which are of unacceptable quality. In the 1920s, mass production brought with it the production line and assembly line concepts. Henry Ford revolutionised car production with the introduction of the mass production of the 'Model T.'

Mass production resulted in greater output and lower prices, but the quality of manufactured output became more variable and less reliable. There was a need to tackle the problem of the production of goods and parts of a fixed quality and standard. The solution was seen to be in the establishment of inspection routines, under the supervision of a Quality Inspector. The first inspection procedures required the testing of the entire production - a costly, time consuming and inefficient form of sorting out good and defective items.

## Quality Control

The Second World War brought with it the need for defect-free products. Inspection departments now came to control the production process, and this resulted in the conformance to set specifications (the reduction of variability and the elimination of defects) being monitored and controlled throughout production. Quality Control departments were separate from, and independent of, the manufacturing departments.

## Quality Assurance

In turn, Quality Control evolved into Quality Assurance. The function of Quality Assurance is to focus on assuring process and product quality through operational audits, the supplying of training, the carrying out of technical analysis, and the giving of advice on quality improvement. The role of Quality Assurance is to consult with the departments (design and production for example) where responsibility for quality actually rests.

## Total Quality Management

Quality Assurance has given way to Company Wide Quality Management or Total Quality Management. As the names imply, quality is no longer seen to be the responsibility of a single department, but has become the responsibility and concern of each individual within an organisation. A small executive group sets the policy which results in targets for the various sections within the company.

For example, line management sets 'do-able' objectives; engineers to design attractive, reliable, and functional products; operators to produce defect-free products and staff in contact with customers to be prompt and attentive. Total Quality Management aims to measure, detect, reduce, eliminate, and prevent quality lapses, it should include not only products but also services, and address issues like poor standards of service and late deliveries.

### **Statistical Quality Control**

As a response to the impracticality of inspecting every production item, methods involving sampling techniques were suggested. The behaviour of samples as an indicator of the behaviour of the entire population has a strong statistical body of theory to support it.

A landmark in the development of statistical quality control came in 1924 as a result of the work of Dr. Walter Shewhart during his employment at Bell Telephone Laboratories. He recognised that in a manufacturing process there will always be variation in the resulting products. He also recognized that this variation can be understood, monitored, and controlled by statistical procedures. Shewhart developed a simple graphical technique - the control chart - for determining if product variation is within acceptable limits. In this case the production process is said to be 'in control' and control charts can indicate when to leave things alone or when to adjust or change a production process. In the latter cases the production process is said to be 'out of control.' Control charts can be used (importantly) at different points within the production process.

Other leading pioneers in the field of quality control were Dr. H. Dodge and Dr. H. Romig. Shewhart, Dodge and Romig were responsible for much of the development of quality control based on sampling techniques. These techniques formed what has become known as 'acceptance sampling.'

Although the techniques of acceptance sampling and control charts were not widely adopted initially outside Bell, during the 1920s many companies had established inspection departments as a means of ensuring quality.

During the Second World War, the need for mass produced products came to the fore and in the United States in particular, statistical sampling and inspection methods became widespread. After the Second World War, the Japanese in particular became very successful in the application of statistical techniques to quality control and in the United States, Dr. W.E. Deming and Dr. J.M. Duran spread the use of such techniques further into U.S. industry.

### **Statistical Process Control**

The aim of statistical process control is to ensure that a given manufacturing process is as stable (in control) as possible and that the process operates around stated values for the product with as little variability as possible. In short, the aim is the reduction of variability to ensure that each product is of a high a quality as possible. Statistical process control is usually thought of as a toolbox whose contents may be applied to solve production-related quality control problems. Essentially the toolbox contains the major tools called:

- the Histogram;
- the Pareto chart;
- cause-and-effect digrams;
- defect-concentration diagrams;
- sampling inspection.
- scatter diagrams;
- check sheets;
- control charts;
- experimental design methods;

Note that some authors argue that experimental design methods should not be regarded as a statistical process control tool since experimental design is an active process involving the deliberate change of

variables to determine their effect, while the other methods are based purely on the observation of an existing process.

Control charts are a very powerful tool in the field of statistical process control. Before looking in some detail at control charts in general we will look at the relationship between specification limits and tolerance limits since this can and does influence the number of defective items produced by a manufacturing process. It is this quantity that control charts seek to minimize.

### **Tolerance Limits and Specifications**

An example of a specification for the manufacture of a short cylindrical spacer might be:

diameter:  $1 \pm 0.003$  cm;                      length:  $2 \pm 0.001$  cm.

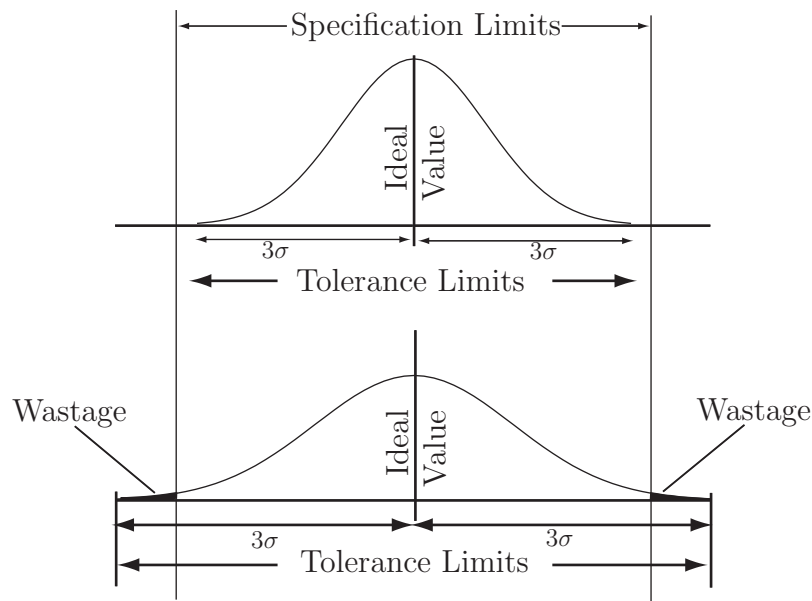
Even though the  $\pm$  differences here are the same ( $\pm 0.003$  in the case of the diameters and  $\pm 0.002$  in the case of the lengths), there is no reason why this should always be so. These limits are called the specification limits. During manufacture, some variation in dimensions will occur by chance. These variations can be measured by using the standard deviation of the distribution followed by the dimensions produced by the manufacturing process. Figure 1 below shows two normal distributions each with so-called **natural tolerance limits** of  $3\sigma$  either side of the mean.

Taking these limits implies that virtually all of the manufactured articles will fall within the natural tolerance limits. Notice that in the top part of the figure the specification limits are rather wider than the natural tolerance limits and that little if any wastage will be produced. One could argue that in this case the manufacturing process is actually better than it needs to be and that this may carry unnecessary costs. In the lower part of the figure the specification limits are narrower than the natural tolerance limits and wastage will occur. In general, a production process should aim to equate the two types of limits to minimize both costs and wastage.

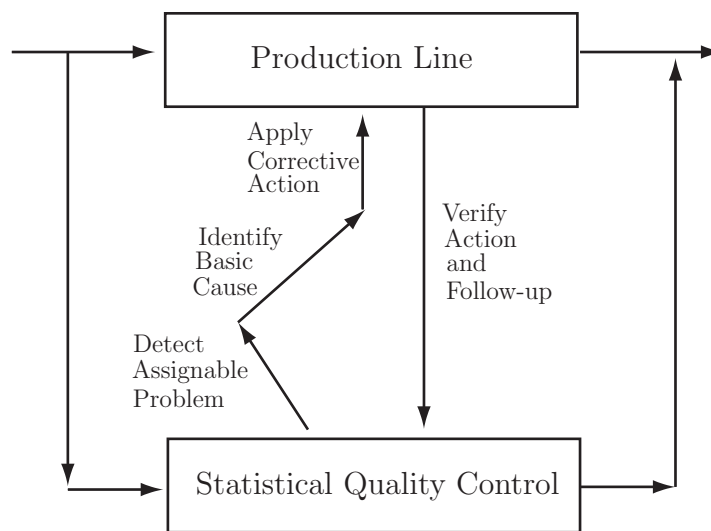
### **Control Charts**

Any production process is subject to variability. Essentially, this variability falls into two categories called **chance causes** and **assignable causes**. A production process exhibiting only chance causes is usually described as being 'in control' or 'in statistical control' whereas a production process exhibiting assignable causes is described as being 'out of control.' The basic aim of control charts is the improvement of production processes via the detection of variability due to assignable causes. It is then up to people such as process operators or managers to find the underlying cause of the variation and take action to correct the process. Figure 4 below illustrates the point.





**Diagram 1 - Specification and Tolerance Limits**



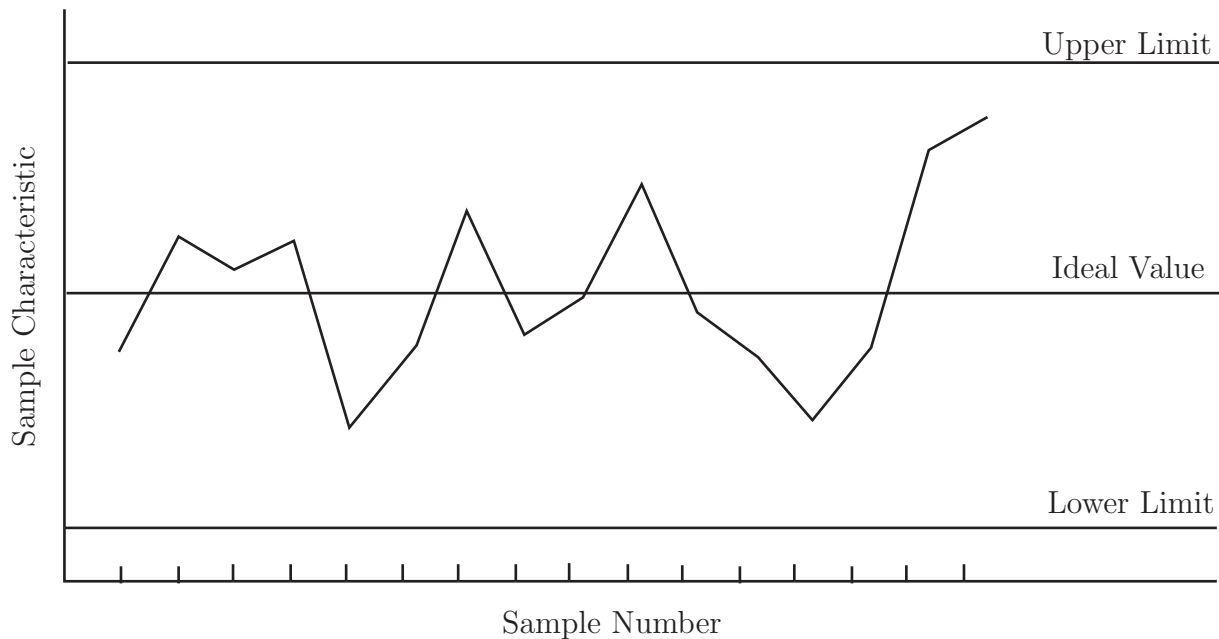
**Diagram 2 - Correcting Assignable Variation**

**Figure 4**

A control chart basically consists of a display of ‘quality characteristics’ which are found from samples taken from production at regular intervals. Typically the mean, together with a measure of the variability (the range or the standard deviation) of measurements taken from the items being produced may be the variables under consideration. The appearance of a typical control chart is shown below in Figure 5.

The centre line on Figure 5 represents the ideal value of the characteristic being measured. Ideally, the points on the figure should hover around this value in a random manner indicating that only chance variation is present in the production process. The upper and lower limits are chosen in such a way that so long as the process is in statistical control, virtually all of the points plotted will fall between these two lines. Note that the points are joined up so that it is easier to detect any trends present in the data. At this stage, a note of caution is necessary. The fact that a particular control

chart shows all of its points lying between the upper and lower limits does not necessarily imply that the production process is in control. We will look at this aspect of control charts in some depth later in this booklet. For now simply note that even points lying within the upper and lower limits that do not exhibit random behaviour can indicate that a process is either out of control or going out of control.



**Figure 5:** A typical control chart

It is also worth noting at this stage that there is a close connection between control charts and hypothesis testing. If we formulate the hypotheses:

$H_0$  : the process is in control

$H_1$  : the process is not in control

then a point lying with the upper and lower limits is telling us that we do not have the evidence to reject the null hypothesis and a point lying outside the upper and lower limits is telling us to reject the null hypothesis. From previous comments made you will realize that these statements are not an absolute truth but that they are an indicative truth.

Control charts are very popular in industry. Before looking at particular types of control charts, the following reasons are given for their popularity.

- (a) They are simple to use. Production process operators do not need to be fluent in statistical methods to be able to use control charts.
- (b) They are effective. Control charts help to keep a production process in control. This avoids wastage and hence unnecessary costs.
- (c) They are diagnostic in the sense that as well as indicating when adjustments need to be made to a process, they also indicate when adjustments do not need to be made.
- (d) They are well-proven. Control charts have a long and successful history. They were introduced in the 1920s and have proved themselves over the years.

- (e) They provide information about the stability of a production process over time. This information is of interest to production engineers for example. A very stable process may enable fewer measurements to be taken - this has cost implications of course.

### Control chart construction

In order to understand how control charts are constructed, consider the following scenario.

Part of a production line used by a manufacturer of instant coffee involves the use of a machine which fills empty packets with coffee. The filler is set to fill each empty packet with 200 grams of coffee. As this is an electro-mechanical process, repeated over a long period of time, it is inevitable that there will be some degree of variation in the amounts of coffee dispensed. The question that production engineers will need to answer is "Is the variation sufficient to result in a significant 'underfilling', or 'overfilling', of the packets?"

The following data (which represent the weights of a sample of 50 packets of 200 grams of coffee) are obtained from the production line:

200.0	197.8	200.1	200.6	198.7
198.7	203.4	201.3	199.8	200.6
202.2	196.4	199.2	199.9	200.3
198.2	201.4	202.8	201.3	203.6
201.9	199.7	198.5	199.8	202.4
201.3	201.2	198.5	196.4	199.2
202.4	202.4	197.9	201.3	200.8
197.8	202.0	200.7	201.1	199.7
201.3	197.6	200.9	201.6	196.2
199.7	203.1	203.1	200.9	199.2

In this example, the data should be read from top left to bottom right, down the columns, that is from 200.0, 198.7, 202.2 and so on to 199.2.

Using Microsoft Excel, the data give the following numerical summaries:

$$\bar{x} = 200.298 \approx 200.30 \text{ grams,}$$

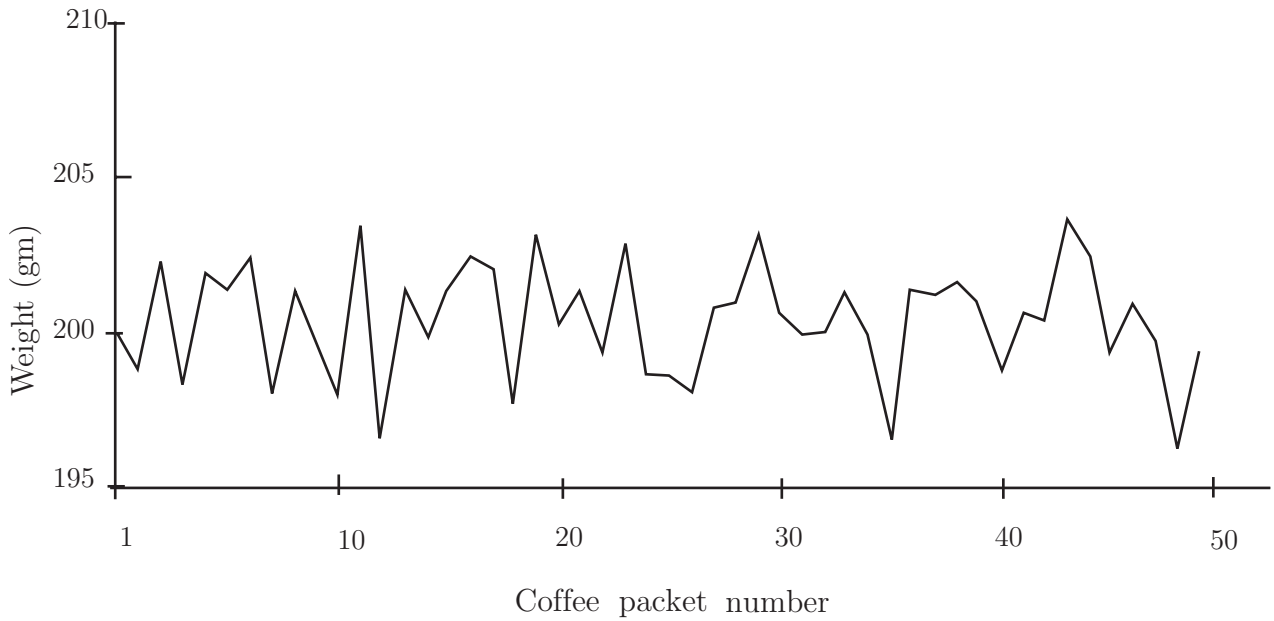
$$s = 1.842 \approx 1.84 \text{ grams}$$

Note that the summaries refer to a sample but, as the sample size is large ( $n = 50$ ), these values may be taken as estimates for the population parameters and we shall take their values as:

$$\mu = 200.30 \text{ grams,}$$

$$\sigma = 1.84 \text{ grams}$$

A plot of the data from Excel is shown in Figure 5.



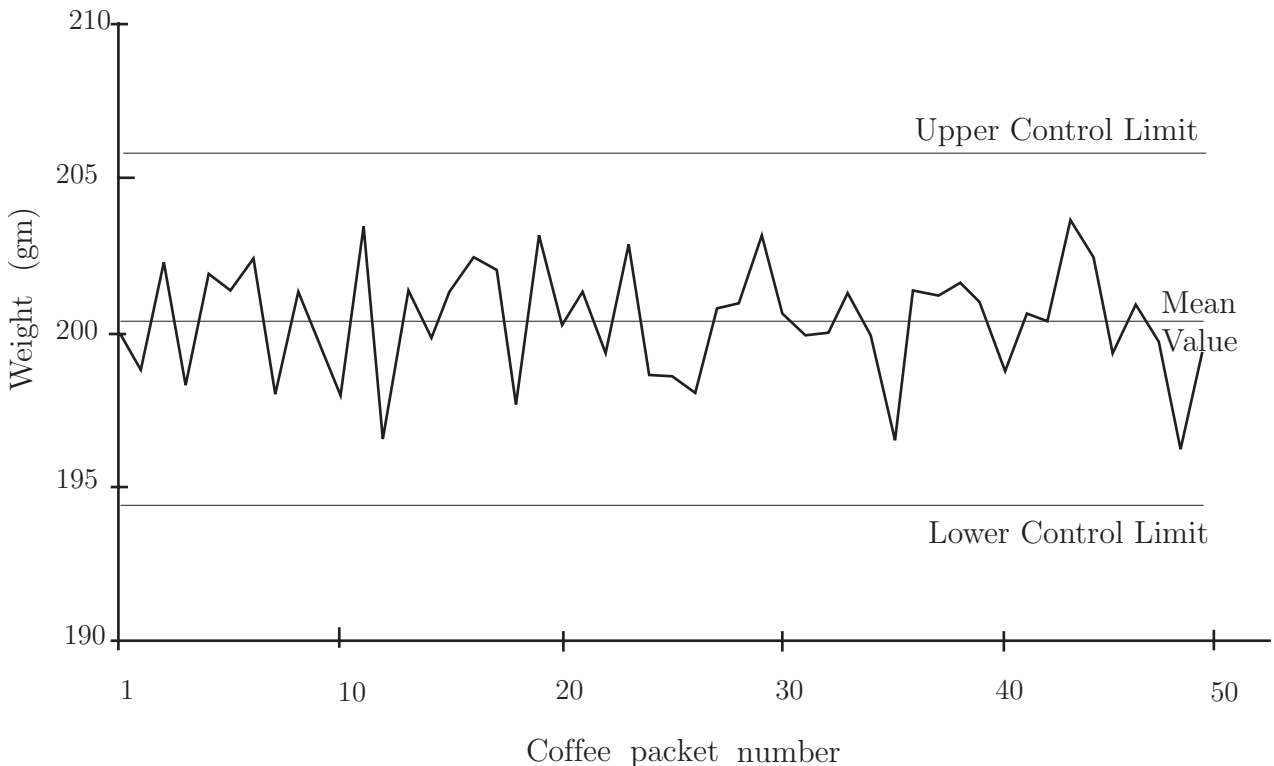
**Figure 6:** Data plot chart

To the plotted series of data are added:

- (a) a horizontal line representing the process mean of 200.30 grams
- (b) two further horizontal lines representing the upper and lower control limits. The values of these limits are given by  $\bar{x} \pm 3\sigma$  and are calculated as 205.8 and 194.8

These lines are called the **Upper Control Limit (UCL)** and the **Lower Control Limit (LCL)**.

This process results in the Figure 7 below.



**Figure 7:** Control chart with centre line and limits

In a production process, samples are taken at regular intervals, say every 10 mins., 15 mins., 30 mins, and so on, so that at least 25 samples (say) are taken. At this stage, we will be able to detect trends - looked at later in this workbook - in the data. This can be very important to production engineers who are expected to take corrective action to prevent a production process going out of control. The size of the sample can vary, but in general  $n \geq 4$ . Do not confuse the number of samples taken, with the sample size  $n$ .

The centre line, denoted by  $\bar{\bar{x}}$  is given by the mean of the sample means

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_k}{k}$$

where  $k$  is the number of samples.  $\bar{\bar{x}}$  is used as an estimator for the population mean  $\mu$ .

Given that the sample size is  $n$  the upper and lower control limits are calculated as

$$\text{UCL} = \bar{\bar{x}} + \frac{3s}{\sqrt{n}} \quad \text{LCL} = \bar{\bar{x}} - \frac{3s}{\sqrt{n}}$$

These limits are also known as **action limits**; if any sample mean falls outside these limits, action has to be taken. This may involve stopping the production line and checking and/or re-setting the production parameters.

In the initial example, a single sample of size  $n = 50$  was used. In reality, with samples being taken at short time intervals, e.g. every 10 min., there is not sufficient time to weigh this many packages. Hence, it is common to use a much smaller sample size, e.g.  $n = 4$  or  $5$ .

Under these circumstances, it is not sensible (very small sample size may give inaccurate results) to use  $s$  as an estimate for  $\sigma$  the standard deviation of the population. In a real manufacturing process, the value of  $\sigma$  is almost always unknown and must be estimated.

### Estimating the population standard deviation.

It is worth noting that there is, in practice, a distinction between the initial setting up of a control chart and its routine use after that. Initially, we need a sufficiently large sample to facilitate the accurate estimation of the population standard deviation  $\sigma$ . Three possible methods used to estimate are:

- Calculate the standard deviations of several samples, and use the mean of these values as an estimate for  $\sigma$ .
- Take a single, large size sample from the process when it is believed to be in control. Use the standard deviation  $s$  of this sample as an estimate for  $\sigma$ . Note that this was the method adopted in the initial example.
- In industry, a widely used method for obtaining an unbiased estimate for  $\sigma$  uses the calculation of the average range  $\bar{R}$  of the sample results. The estimate for  $\sigma$  is then given by the formula:

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

where  $d_2$  is a constant that depends on the sample size. For example, the Table given at the end of this Workbook, show that for  $n = 5$ ,  $d_2 = 2.326$ .

The control limits can now be expressed in the form:

$$\bar{\bar{x}} \pm \frac{3\bar{R}}{d_2\sqrt{n}}$$

and, by a further simplification as:

$$\bar{\bar{x}} \pm A_2 \bar{R}$$

A table for the constant  $A_2$  are given at the end of this Workbook (Table 1).



### Example 5

The following table shows the data obtained from 30 samples, each of size  $n = 4$ , taken from the 'instant coffee' production line considered above. The table has been extended to show the mean  $\bar{x}$  and the range  $\bar{R}$  of each sample. Find the grand mean  $\bar{\bar{x}}$ , the upper and lower control limits and plot the control chart. You are advised that all of the necessary calculations have been done in Excel and that you should replicate them for practice. You can, of course, use a dedicated statistics package such as MINITAB if you wish.

Sample	Weight	Weight	Weight	Weight	Mean	Range	Sample	Weight	Weight	Weight	Weight	Mean	Range
1	202.3	199.8	201.4	200.3	200.95	2.5	16	195.5	202.0	199.2	202.3	199.75	6.8
2	198.6	201.3	199.7	201.8	200.35	3.2	17	199.0	200.7	200.1	199.3	199.78	1.7
3	197.9	199.8	200.1	189.8	196.90	10.3	18	197.6	188.9	201.1	200.3	196.98	12.2
4	202.4	205.1	199.6	189.2	199.08	15.9	19	200.6	199.4	201.8	202.0	200.95	2.6
5	194.7	201.2	197.5	201.1	198.63	6.5	20	199.4	201.0	200.8	197.6	199.70	3.4
6	200.8	199.8	200.3	196.2	199.28	4.6	21	197.8	205.1	203.6	204.3	202.70	7.3
7	198.5	201.4	200.6	199.2	199.93	2.9	22	201.3	196.8	197.6	199.6	198.83	4.5
8	202.3	204.1	199.2	201.3	201.73	4.9	23	200.6	199.4	200.4	201.8	200.55	2.4
9	205.4	198.3	194.9	198.5	199.28	10.5	24	198.9	204.2	202.0	201.1	201.55	5.3
10	199.0	202.2	197.1	202.8	200.28	5.7	25	203.0	201.6	201.4	200.8	201.70	2.2
11	189.7	200.1	202.6	201.9	198.58	12.9	26	200.1	194.8	201.7	198.3	198.73	6.9
12	203.6	197.5	204.5	196.4	200.50	8.1	27	198.2	201.4	197.6	201.1	199.58	3.8
13	198.6	197.8	199.7	200.4	199.13	2.6	28	200.6	202.8	210.0	202.8	201.80	2.2
14	202.6	199.2	199.0	199.2	200.00	3.6	29	200.3	201.3	201.6	201.3	201.13	1.3
15	203.3	203.1	200.8	201.7	202.23	2.5	30	195.9	203.3	196.3	203.4	199.73	7.5

#### Solution

Using Excel, the following results are obtained:

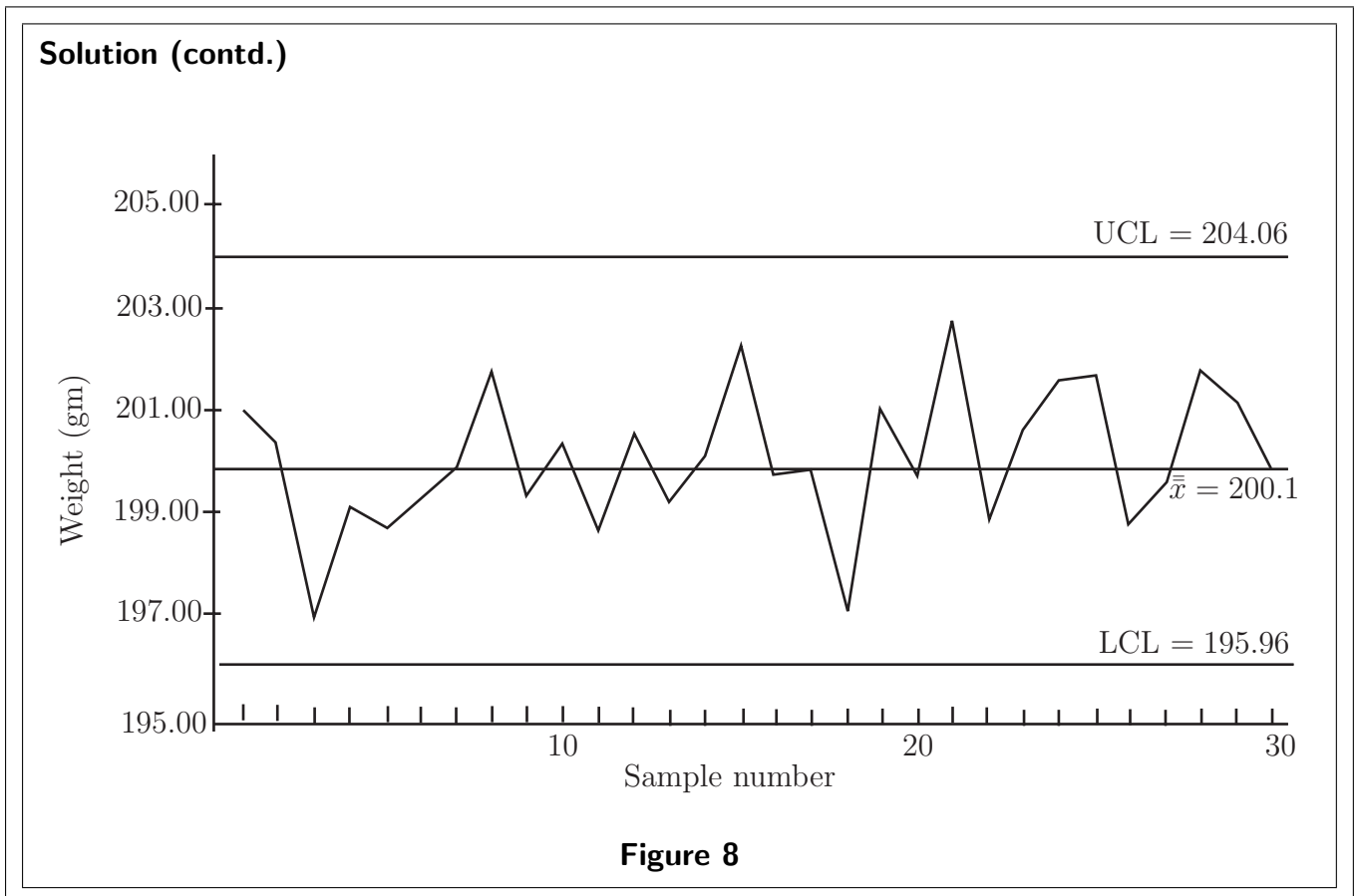
$$\bar{\bar{x}} = 200.01 \text{ and } \bar{R} = 5.56$$

Using these values, the control limits are defined as:

$$\text{UCL} = \bar{\bar{x}} + A_2 \bar{R} = 200.01 + 0.729 \times 5.56 = 204.06$$

and

$$\text{LCL} = \bar{\bar{x}} - A_2 \bar{R} = 200.01 - 0.729 \times 5.56 = 195.96$$



Note that if the process is in control, the probability that a sample mean will fall outside the control limits is very small. From the central limit theorem we know that the sample means follow (or at least approximately follow) a normal distribution and from tables it is easy to show that the probability of a sample mean falling outside the region defined by the  $3\sigma$  control limits is about 0.3%. Hence, if a sample mean falls outside the control limits, there is an indication that there could be problems with the production process. As noted before the process is then said to be out of control and process engineers will have to locate the cause of the unusual value of the sample mean.

### Warning limits

In order to help in the monitoring of the process, it is common to find additional **warning limits** which are included on the control chart. The settings for these warning limits can vary but a common practice involves settings based upon two standard deviations. We may define warning limits as:

$$\bar{\bar{x}} \pm \frac{2\sigma}{\sqrt{n}}$$

Using the constant  $d_2$  defined earlier, these limits may be written as

$$\bar{\bar{x}} \pm \frac{2\bar{R}}{d_2\sqrt{n}}$$



### Example 6

Using the data from Example 5, find the  $2\sigma$  warning limits and revise the control chart accordingly.

#### Solution

We know that  $\bar{\bar{x}} = 200.01$ ,  $\bar{R} = 5.56$ , and that the sample size  $n$  is 4. Hence:

$$\bar{\bar{x}} \pm \frac{2\bar{R}}{d_2\sqrt{n}} = 200.01 \pm \frac{2 \times 5.56}{2.059 \times \sqrt{4}} = 200.01 \pm 2.70$$

The warning limits clearly occur at 202.71 and 197.31. The control chart now appears as shown below. Note that the warning limits in this case have been shown as dotted lines.

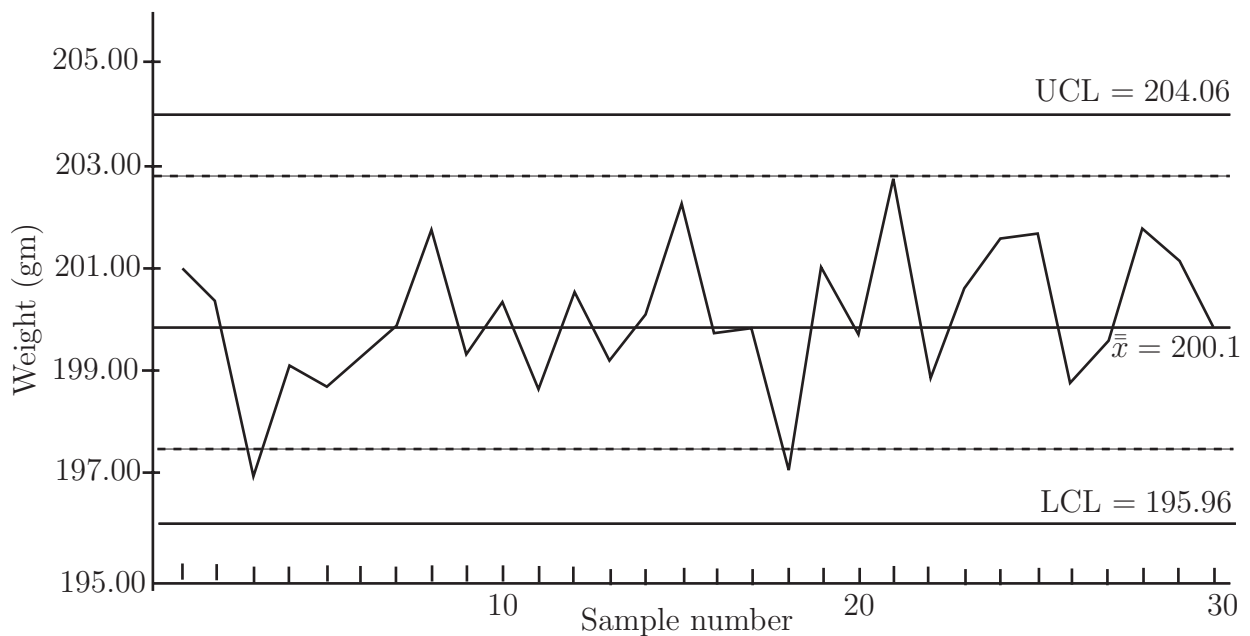


Figure 9

#### Detecting trends

Control charts can also be used to detect trends in the production process. Even when all of the sample means fall within the control limits, the existence of a trend can show the presence of one, or more, assignable causes of variation which can then be identified and brought under control. For example, due to wear in a machine, the process mean may have shifted slightly.

Trends are detected by considering **runs** of points above, or below, the centre line. A **run** is defined as a sequence of consecutive points, all of which are on the same side of the centre line.

Using a plus (+) sign to indicate a point above, and a minus (-) sign to indicate a point below, the centre line, the 30 sample means used as the worked examples above would show the following sequence of runs:

+ + - - - - + - + - + - + + - - - + - + - + + + - - + + -



Considerable attention has been paid to the study of runs and theory has been developed. Simple rules and guidelines, based on the *longest run* in a chart have been formulated.

Listed below are eight tests that can be carried out on the chart. Each test detects a specific pattern in the data plotted on the chart. Remember that a process can only be considered to be in control if the occurrence of points above and below the chart mean is random and within the control limits. The occurrence of a pattern indicates a special cause for the variation, one that needs to be investigated.

### The eight tests

- One point more than 3 sigmas from the centre line.
- Two out of three points in a row more than two sigmas from the centre line (same side)
- Four out of five points in a row more than 1 sigma from centre line (same side)
- Eight points in a row on the same side of the centre line.
- Eight points in a row more than 1 sigma either side of the centre line.
- Six points in a row, all increasing or all decreasing.
- Fourteen points in a row, alternating up and down.
- Fifteen points in a row within 1 sigma of centre line (either side)

The first four of these tests are sometimes referred as the Western Electric Rules. They were first published in the Western Electric Handbook issued to employees in 1956.

### The $R$ -Chart

The control chart based upon the sample means is only one of a number of charts used in quality control. In particular, to measure the variability in a production process, the chart is used. In this chart, *the range of a sample* is considered to be a random variable. The range has its own mean and standard deviation. The average range  $\bar{R}$  provides an estimate of this mean, while an estimate of the standard deviation is given by the formula:

$$\hat{\sigma}_R = d_3 \times \frac{\bar{R}}{d_2}$$

where  $d_2$  and  $d_3$  are constants that depend on sample size. The control limits for the  $R$ -chart are defined as:

$$\bar{R} \pm 3\hat{\sigma}_R = \bar{R} \pm 3d_3 \times \frac{\bar{R}}{d_2}$$

By the use of further constants, these control limits can be simplified to:

$$\text{UCL} = \bar{R} \times D_4 \text{ and } \text{LCL} = \bar{R} \times D_3$$

Values for  $D_3$  and  $D_4$  are given in the table at the end of this workbook. For values of  $n \leq 6$  the value of  $D_3$  is zero.

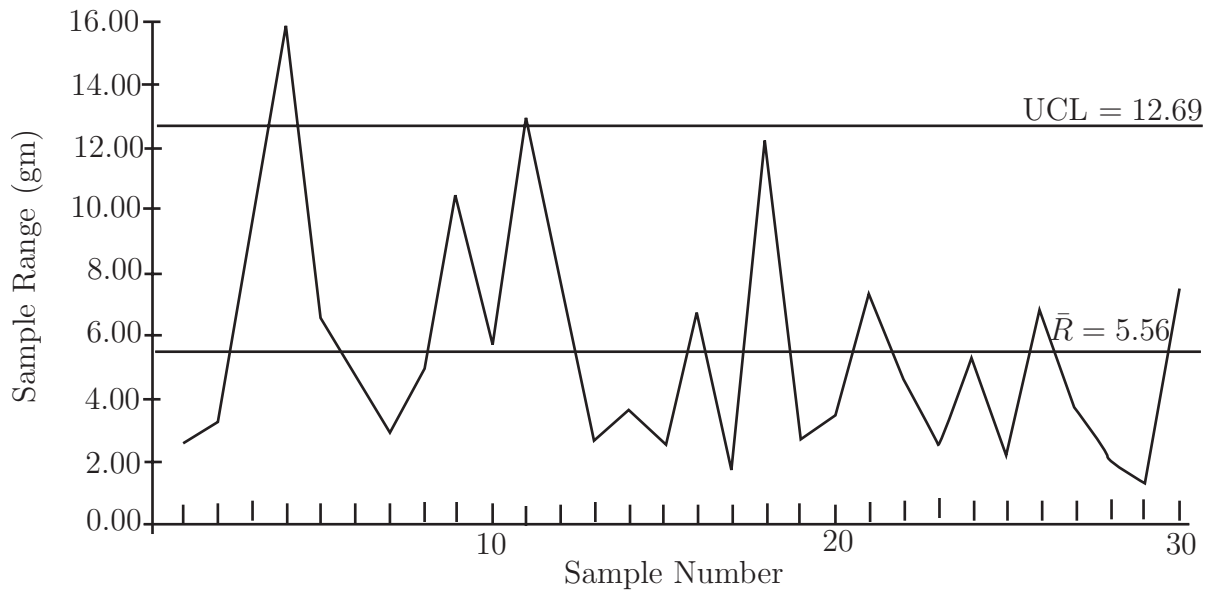
Note that if the  $R$ -chart indicates that the process is *not* in control, then there is no point in trying to interpret the  $\bar{x}$ -chart.

Returning to the data we have used so far, the control limits for the chart will be given by:

$$\text{UCL} = \bar{R} \times D_4 = 5.56 \times 2.282 \simeq 12.69$$

$$\text{LCL} = \bar{R} \times D_3 = 0$$

since the sample size is only 4. Effectively, the LCL is represented by the horizontal axis in the following chart.



**Figure 10:** Control Chart

Note that although the  $\bar{x}$ -chart indicated that the process was in control, the  $\bar{R}$ -chart shows that in two samples, the ranges exceed the UCL. Hence, no attempt should be made to interpret the  $\bar{x}$ -chart. Since the limits used in the  $\bar{x}$ -chart are based upon the value of  $\bar{R}$ , these limits will have little meaning unless the process variability is in control. It is often recommended that the  $R$ -chart be constructed first and if this indicates that the process is in control then the  $\bar{x}$ -chart be constructed. However you should note that in reality, both charts are used simultaneously.

### Further control charts

The  $\bar{x}$ -charts and the  $R$ -charts considered above are only two of the many charts which are used in quality control. If the sample is of size  $n > 10$  then it may be better to use a chart based on the standard deviation, the  $s$ -chart, for monitoring process variability. In practice, there is often little to choose between the  $s$ -chart and the  $R$ -chart for samples of size  $n < 10$  although it is worth noting that the  $R$ -chart is easier to formulate thanks to the easier calculations.

Many control charts have a common appearance with a central line indicating a mean, a fixed setting, or perhaps a percentage of defective items, when the process is in control. In addition, the charts include control limits that indicate the bounds within which the process is deemed to be in control. A notable exception to this is the CUSUM chart. This chart is not considered here, but see, for example, *Applied Statistics and Probability for Engineers*, Montgomery and Runger, Wiley, 2002.

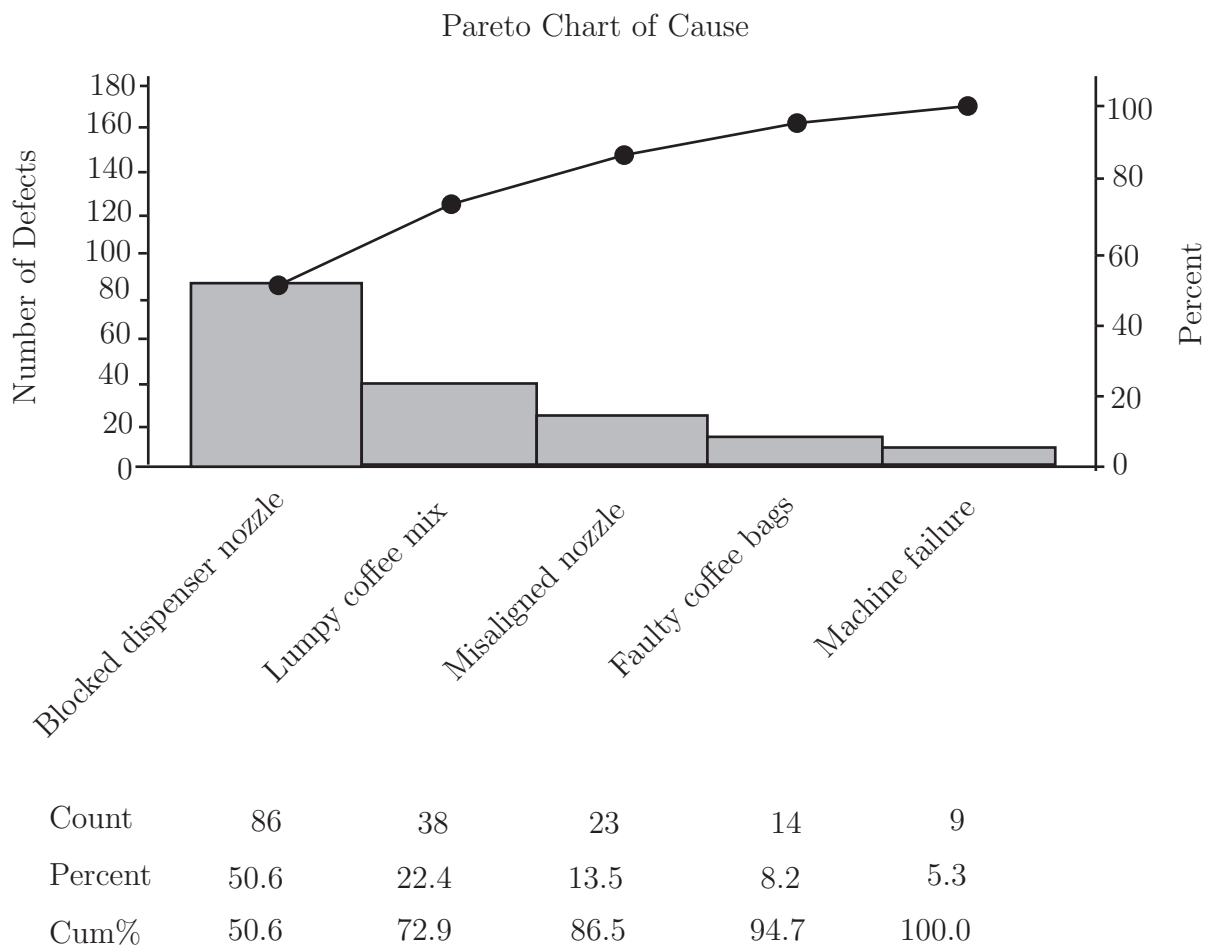
### Pareto charts

A Pareto diagram or **Pareto chart** is a special type of graph which is a combination line graph and bar graph. It provides a visual method for identifying significant problems. These are arranged in a bar graph by their relative importance. Associated with the work of Pareto is the work of Joseph Juran who put forward the view that the solution of the ‘vital few’ problems are more important than solving the ‘trivial many’ . This is often stated as the ‘80 20 rule’ ; i.e. 20% of the problems (the vital few) result in 80% of the failures to reach the set standards.

In general terms the Pareto chart is not a problem-solving tool, but it is a tool for analysis. Its function is to determine which problems to solve, and in what order. It aims to identify and eliminate major problems, leading to cost savings.

Pareto charts are a type of bar chart in which the horizontal axis represents categories of interest, rather than a continuous scale. The categories are often types of defects. By ordering the bars from largest to smallest, a Pareto chart can help to determine which of the defects comprise the vital few and which are the trivial many. A cumulative percentage line helps judge the added contribution of each category. Pareto charts can help to focus improvement efforts on areas where the largest gains can be made.

An example of a Pareto chart is given below for 170 faults found in the process considered above where coffee is dispensed by an automatic machine.



**Figure 12:** A Pareto chart



The following table shows the data obtained from 30 samples, each of size  $n = 4$ , taken from a production line whose output is nominally 200 gram packets of tacks. Extend the table to show the mean  $\bar{x}$  and the range  $\bar{R}$  of each sample. Find the grand mean  $\bar{\bar{x}}$ , the upper and lower control limits and plot the  $\bar{x}$ - and  $\bar{R}$ -control charts. Discuss briefly whether you agree that the production process is in control. Give reasons for your answer. You are advised to do all of the calculations in Excel or use a suitable statistics package.

| Sample | Weight | Weight | Weight | Weight | Sample | Weight | Weight | Weight | Weight |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1      | 201.3  | 199.8  | 201.4  | 200.3  | 16     | 196.5  | 202.0  | 199.2  | 202.3  |
| 2      | 198.6  | 201.3  | 199.7  | 201.8  | 17     | 199.0  | 200.7  | 200.1  | 199.3  |
| 3      | 198.9  | 199.8  | 200.1  | 199.8  | 18     | 196.6  | 198.9  | 201.1  | 200.3  |
| 4      | 202.4  | 203.1  | 199.6  | 199.2  | 19     | 200.6  | 199.4  | 201.8  | 201.0  |
| 5      | 194.7  | 201.2  | 197.5  | 201.1  | 20     | 199.4  | 201.0  | 200.8  | 197.6  |
| 6      | 200.8  | 199.8  | 200.3  | 196.2  | 21     | 199.8  | 203.1  | 201.6  | 204.3  |
| 7      | 199.5  | 201.4  | 200.6  | 199.2  | 22     | 201.3  | 196.8  | 197.6  | 199.6  |
| 8      | 202.3  | 203.1  | 199.2  | 201.3  | 23     | 200.6  | 199.4  | 200.4  | 200.8  |
| 9      | 205.4  | 198.3  | 197.9  | 198.5  | 24     | 197.9  | 202.2  | 201.0  | 200.1  |
| 10     | 199.0  | 202.2  | 197.1  | 202.8  | 25     | 203.0  | 201.6  | 201.4  | 200.8  |
| 11     | 189.7  | 200.1  | 202.6  | 201.9  | 26     | 200.1  | 199.8  | 201.7  | 199.3  |
| 12     | 201.6  | 197.5  | 204.5  | 196.4  | 27     | 198.2  | 201.4  | 198.6  | 201.1  |
| 13     | 198.6  | 198.8  | 199.7  | 200.4  | 28     | 200.6  | 202.8  | 210.0  | 202.8  |
| 14     | 202.6  | 199.2  | 199.0  | 199.2  | 29     | 200.3  | 201.3  | 201.6  | 201.3  |
| 15     | 203.3  | 203.1  | 200.8  | 201.7  | 30     | 197.9  | 201.3  | 198.3  | 201.4  |

### Your solution

**Answer**

The solution is given in two parts, first the  $\bar{x}$ -control chart and secondly the  $\bar{R}$ -control chart.

**Part (i)**

Using Excel, the following results are obtained:  $\bar{\bar{x}} = 200.24$  and  $\bar{R} = 3.88$

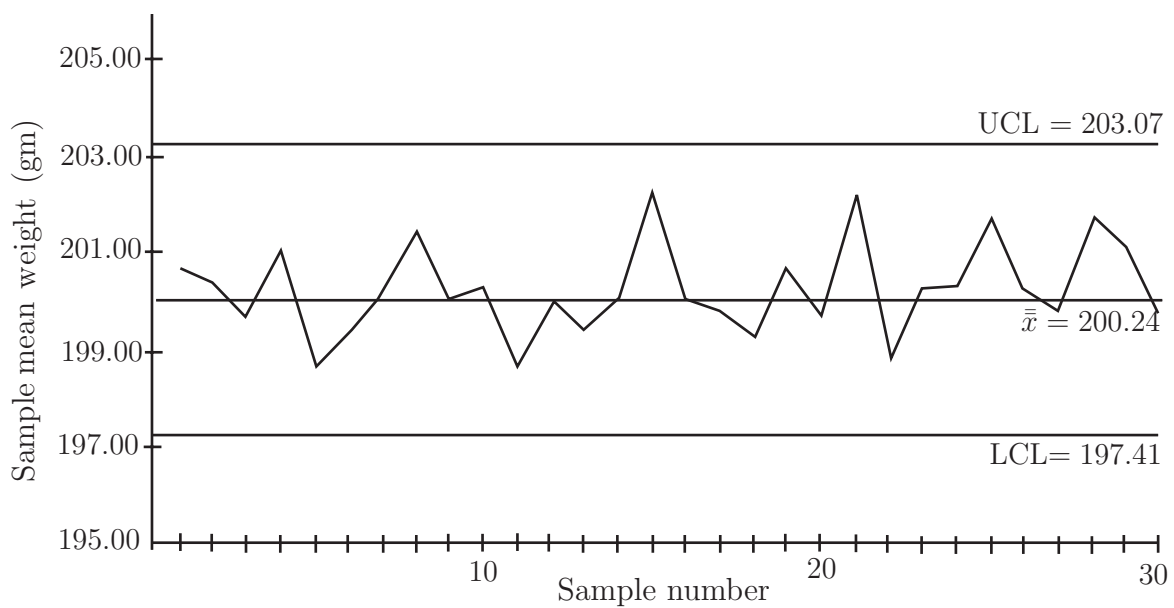
Using these values, the control limits are defined as:

$$UCL = \bar{\bar{x}} + A_2\bar{R} = 200.24 + 0.729 \times 3.88 = 203.07$$

and

$$LCL = \bar{\bar{x}} - A_2\bar{R} = 200.24 - 0.729 \times 3.88 = 197.41$$

The  $\bar{x}$ -chart is shown below.



**Answer**

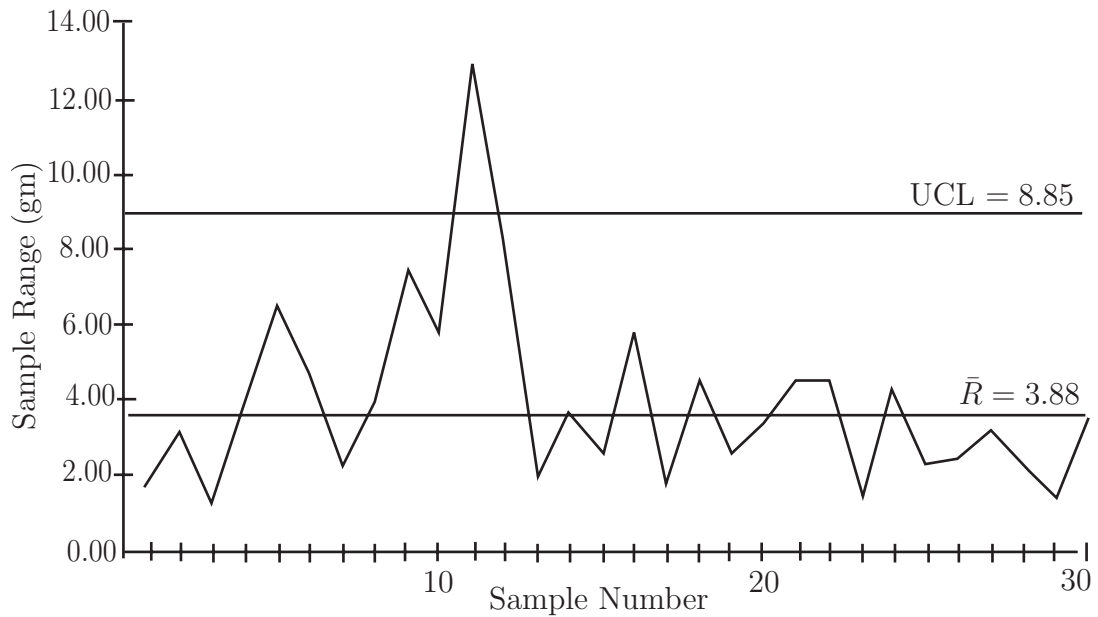
**Part (ii)**

The control limits for the chart are given by:

$$UCL = \bar{R} \times D_4 = 3.88 \times 2.282 = 8.85$$

$$LCL = \bar{R} \times F_3 = 0$$

since the sample size is 4, so that the LCL is represented by the horizontal axis in the following *R*-chart.



Since one member of the *R*-chart exceeds the UCL we *cannot* say that the process is in control and conclude that the source of the error should be investigated and corrected.

**Table 1: Some constants used in the construction of control limits**

| Sample Size ( <i>n</i> ) | $d_2$ | $A_2$ | $D_3$ | $D_4$ |
|--------------------------|-------|-------|-------|-------|
| 2                        | 1.128 | 1.880 | 0     | 3.267 |
| 3                        | 1.693 | 1.023 | 0     | 2.574 |
| 4                        | 2.059 | 0.729 | 0     | 2.282 |
| 5                        | 2.326 | 0.577 | 0     | 2.114 |
| 6                        | 2.543 | 0.483 | 0     | 2.004 |
| 7                        | 2.704 | 0.419 | 0.076 | 1.924 |
| 8                        | 2.847 | 0.373 | 0.136 | 1.864 |
| 9                        | 2.970 | 0.337 | 0.184 | 1.816 |
| 10                       | 3.078 | 0.308 | 0.223 | 1.777 |
| 15                       | 3.472 | 0.223 | 0.347 | 1.653 |
| 20                       | 3.735 | 0.180 | 0.415 | 1.585 |
| 25                       | 3.931 | 0.153 | 0.459 | 1.541 |