MA4605 Lecture 7B Statistical Process Control

Statistical process control refers to the application of the methods of statistical quality control to the monitoring of processes (and not just, as in the earlier practice, to the inspection of the final outputs of the processes).

The purpose is to control the quality of product or service outputs from a process by maintaining control of the process.

When a process is described as being "***in control***," it means that the amount of variation in the output is relatively constant and within established limits that are deemed acceptable.

There are two kinds of causes of variation in a process:
- ***Common causes***, or ***chance causes***, of variation are due to factors that are inherent in the design of the system, and reflect the usual amount of variation to be expected.
- ***Assignable causes***, or ***special causes***, of variation are due to unusual factors that are not part of the process design and not ordinarily part of the process.

A stable process is one in which only common causes of variation affect the output quality. Such a process can also be described as being in a state of statistical control.

An unstable process is one in which both assignable causes and common causes affect the output quality. (Note that, by definition, the common causes are always present.)

Such a process can also be described as being out of control, particularly when the assignable cause is controllable.

The way we set out to improve the quality of output for a process depends on the source of process variation.

For a process that is stable, improvement can take place only by improving the design of the process. A pervasive error in process management is ***tampering***, which is to take actions (such as making machine adjustments) that presume that a process is not in control, when in fact it is stable.

Such actions only increase variability, and are analogous to the over-correcting that new drivers do in learning to steer a car. For a process that is unstable, improvement can be achieved by identifying and correcting assignable causes.

Control Charts

A run chart is a time series plot that plots levels of output quality on the vertical axis, with respect to a sequence of time periods on the horizontal axis. In the context of statistical process control the measurements that are graphed are typically sample data that have been obtained by the method of rational subgroups.

A control chart is a run chart that includes the lower and upper control limits that identify the range of variation that can be ascribed to common causes.

Any outputs that are outside of the control limits suggest the existence of assignable-cause variation.

The control limits are determined either by process parameters having been specified, or by observing sample outcomes during a period of time in which the process is deemed to be in a stable condition.

The statistical methods of process control are based on the concepts of hypothesis testing (see Table)

- The null hypothesis is that the process is stable and only common causes of variation exist.

- The alternative hypothesis is that the process is unstable and includes assignable-cause variation.

Thus, the lower and upper control limits on a control chart are the critical values with respect to rejecting or not rejecting the null hypothesis that the process is stable and in control.

		Condition of Process	
		H_0 True: Stable	H_0 False: Unstable
Decision	Continue Process	Correct decision	Type II error: Allowing an unstable process to continue
	Adjust Process	Type I error: Adjusting a stable process	Correct decision

The standard practice is to place the control limits at three standard error units above and below the hypothesized value. This 3-sigma rule is conventionally applied for most control chart procedures.

If a process is stable, not only should all sample statistics be within the control limits, but there also should be no discernable pattern in the sequence of the sample statistics.

We look at the methods for determining control limits and we describe the interpretation of control charts for the

- process mean,
- process range
- process standard deviation,
- process range, process proportion.

Of these four, control charts for the mean, standard deviation, and range are designated as **control charts for variables**, because measurements are involved.

Control charts for the proportion are designated as **control charts for attributes**, because counts (which are discrete rather than continuous variables) are involved.

1) *Process Mean*

Process Mean and Standard Deviation Known

The process mean and standard deviation would be known either because they are process specifications or are based on historical observations of the process when It was deemed to be stable. The centreline for the **X.bar** chart is set at the process mean.

Using the 3-sigma rule, the control limits are defined by the same method as that used for determining critical values for hypothesis testing.

$$\text{Control limits} = \mu \pm 3\frac{\sigma}{\sqrt{n}}$$

Process Mean and Standard Deviation Unknown.

When the process mean and standard deviation are unknown, the required assumption is that the samples came from a stable process. Thus, recent sample results are used as the basis for determining the stability of the process as it continues. First the overall mean of the **k** sample means and the mean of the **k** sample standard deviations are determined:

X.double-bar	*s.bar*
$\bar{\bar{X}} = \dfrac{\Sigma \bar{X}}{k}$	$\bar{s} = \dfrac{\Sigma s}{k}$

Although **X.double-bar** is an unbiased estimator of **μ**, **s.bar** is a biased estimator of σ. This is true even though **s²** is an unbiased estimator of **σ²**.

Therefore, the formula for the control limits incorporates a correction for the biasedness in **s.bar.**

$$\text{Control limits} = \bar{\bar{X}} \pm 3\frac{\bar{s}}{c_4\sqrt{n}}$$

Quality Control Charts

In the Lab classes, we will look at using the **R** package **qcc**, which stands for quality control charts. This package can be used to create the following.

- Shewhart quality control charts for continuous, attribute and count data.
- CUSUM and EWMA charts.
- Operating characteristic curves.
- Process capability analysis.
- Pareto chart and cause-and-effect chart.
- Multivariate control charts

(We won't have time to get through all of this material)

There are two inbuilt data sets, available from the **qcc** package, that we will use

- pistonrings
- orangejuice

The "pistonrings" data set (diameter)

The first data set describes piston rings for an automotive engine are produced by a forging process.

The inside diameter of the rings manufactured by the process is measured on 25 samples, each of size 5, drawn from a process being considered 'in control'.

There are further 15 sets of observation (i.e. 40 batches in total) to demonstrate alternate outcomes.

The dataset is restructured as a data set called "diameter".

The "orangejuice" data set

The second data set describes frozen orange juice concentrate is packed in 6-oz cardboard cans. These cans are formed on a machine by spinning them from cardboard stock and attaching a metal bottom panel.

A can is then inspected to determine whether, when filled, the liquid could possible leak either on the side seam or around the bottom joint. If this occurs a can is considered nonconforming.

The data were collected as 30 samples of 50 cans each at half-hour intervals over a three-shift period in which the machine was in continuous operation. From sample 15 used a new batch of cardboard stock was punt into production.

Sample 23 was obtained when an inexperienced operator was temporarily assigned to the machine. After the first 30 samples, a machine adjustment was made. Then further 24 samples were *taken* from the process.

Interpreting X.bar charts

There are eight standard tests that are used to detect the presence of assignable-cause variation in *X.bar* charts. Before considering these tests, it is useful to understand the logic of the zones that are identified on *X.bar* charts
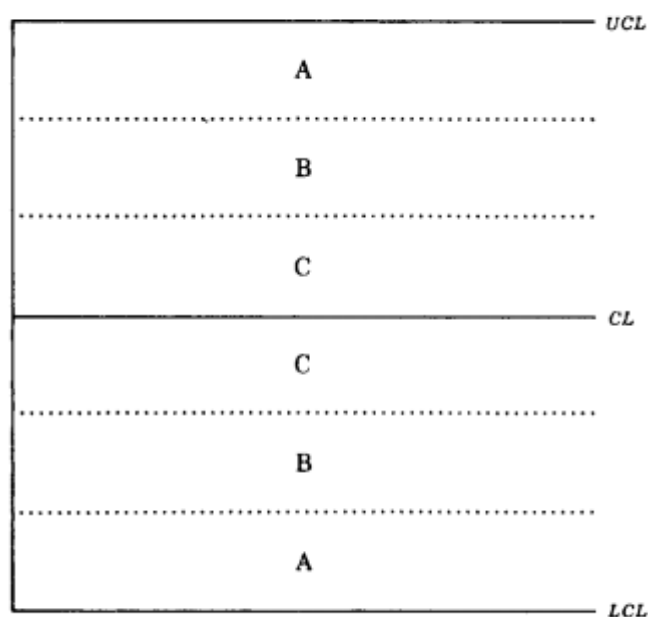
Because the sample means are approximately normally distributed if the process is stable, about 68 percent, 95 percent, and 99 percent of the sample means should be located within one, two, and three standard deviations (i.e., standard error units) of the centre-line, respectively.

Zone C is the zone within one standard deviation of the centre-line, *Zone B* is the zone between one and two standard deviations from the centreline, and *Zone A*, the outermost labelled zone, is between two and three standard deviations from the centre-line

The following eight example tests identify patterns are very unlikely to occur in stable processes.

Thus the existence of any of these patterns in an *X.bar* chart indicates that the process may be unstable, and that one or more assignable causes may exist.
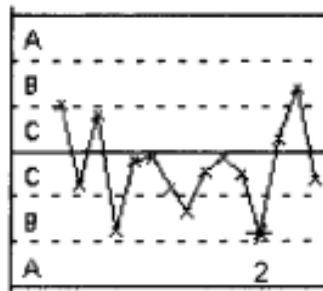
Test 1
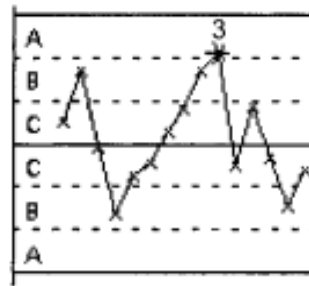One point more than 3 sigmas from center line
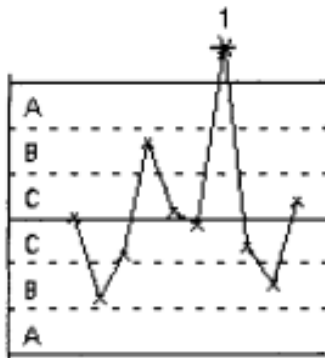
Test 2
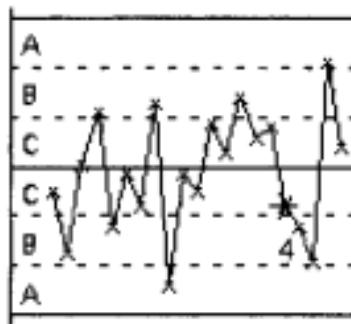Nine points in a row on same side of center line

Test 3
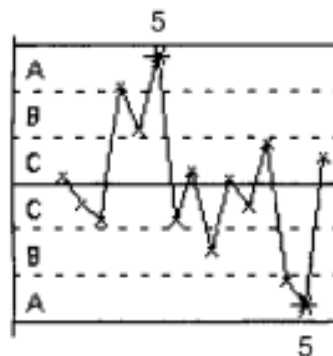Six points in a row, all increasing or all decreasing

Test 4
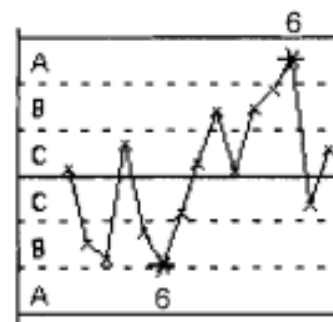Fourteen points in a row, alternating up and down

Test 5
Two out of three points in a row more than 2 sigmas from center line (same side)
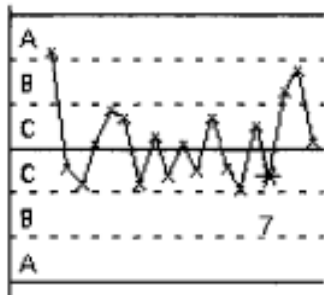
Test 6
Four out of five points in a row more than 1 sigma from center line (same side)
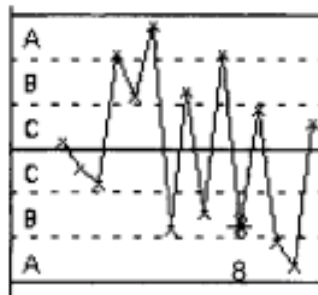
Test 7

Fifteen points in a row within 1 sigma of center line (either side)

Test 8

Eight points in a row more than 1 sigma from center line (either side)

The most obvious test in terms of its rationale is Test 1, which simply requires that at least one value of *X.bar* be beyond Zone A. For a sample mean to be beyond three standard errors from the centre-line is of course very unlikely in a stable process. The other seven patterns also are very unlikely.

The one test whose rationale is not obvious is Test 7, which requires that 15 consecutive values of *X.bar* be in Zone C, which is within one standard error of the centre-line.

Although this kind of pattern would seem to be very desirable, it is also very unlikely. It may, for example, indicate that the process standard deviation was overstated in the process specifications or that the sample measurements are in error. Whatever the cause, the results are too good to be true and therefore require investigation.
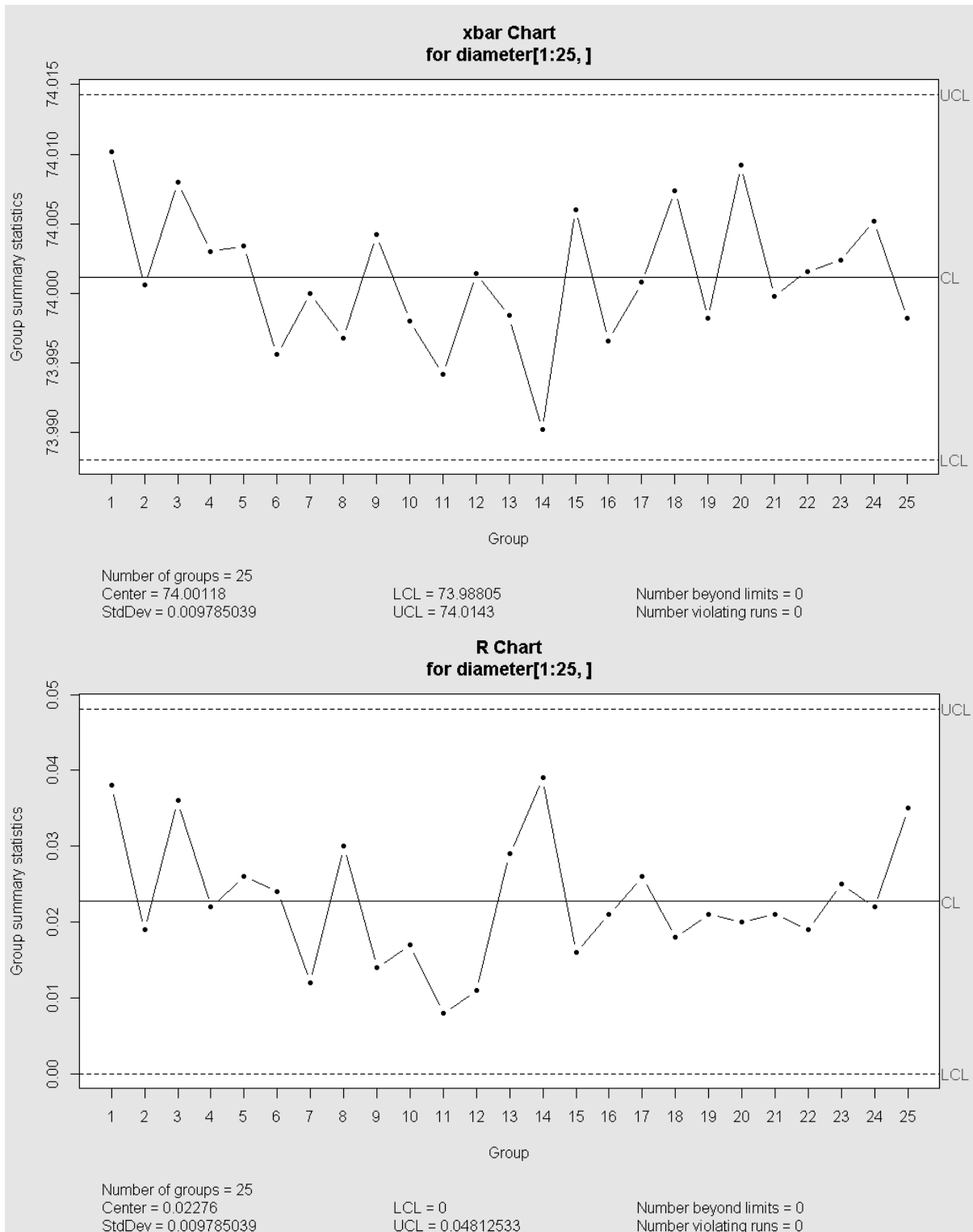
Diameter (Piston Rings)

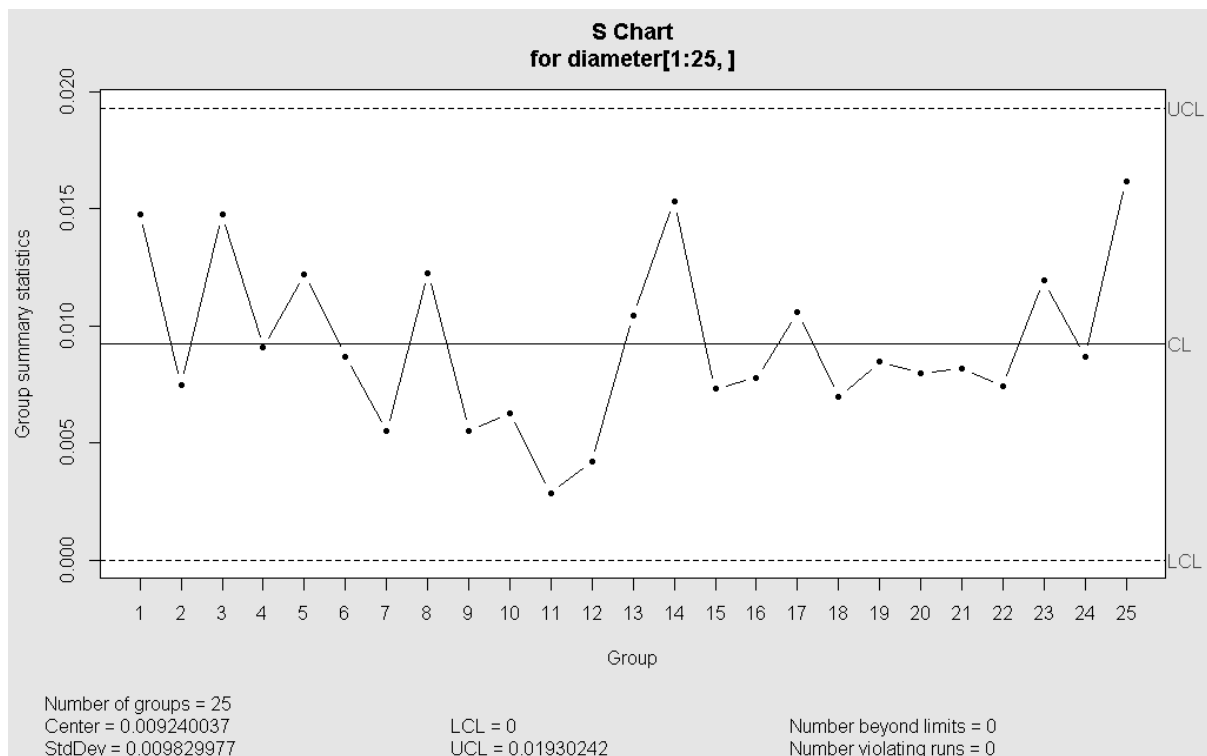Let us consider the mean, range and standard deviation for each batch.

```
> head(diameter)
  item 1 item 2 item 3 item 4 item 5
1 74.030 74.002 74.019 73.992 74.008
2 73.995 73.992 74.001 74.011 74.004
3 73.988 74.024 74.021 74.005 74.002
4 74.002 73.996 73.993 74.015 74.009
...

...

...
38 74.035 74.010 74.012 74.015 74.026
39 74.017 74.013 74.036 74.025 74.026
40 74.010 74.005 74.029 74.000 74.020
>
>
> cbind(diameter, apply(diameter,1,mean))
   item 1 item 2 item 3 item 4 item 5
1   74.030 74.002 74.019 73.992 74.008 74.0102
2   73.995 73.992 74.001 74.011 74.004 74.0006
3   73.988 74.024 74.021 74.005 74.002 74.0080
4   74.002 73.996 73.993 74.015 74.009 74.0030
...

...

...
39 74.017 74.013 74.036 74.025 74.026 74.0234
40 74.010 74.005 74.029 74.000 74.020 74.0128
>
> cbind(diameter, apply(diameter,1,max)-apply(diameter,1,min))
   item 1 item 2 item 3 item 4 item 5
1   74.030 74.002 74.019 73.992 74.008 0.038
2   73.995 73.992 74.001 74.011 74.004 0.019
3   73.988 74.024 74.021 74.005 74.002 0.036
4   74.002 73.996 73.993 74.015 74.009 0.022
5   73.992 74.007 74.015 73.989 74.014 0.026
...

...
39 74.017 74.013 74.036 74.025 74.026 0.023
40 74.010 74.005 74.029 74.000 74.020 0.029
>
> cbind(diameter, apply(diameter,1,sd))
   item 1 item 2 item 3 item 4 item 5
1   74.030 74.002 74.019 73.992 74.008 0.014771594
2   73.995 73.992 74.001 74.011 74.004 0.007503333
3   73.988 74.024 74.021 74.005 74.002 0.014747881
4   74.002 73.996 73.993 74.015 74.009 0.009082951
...

...
38 74.035 74.010 74.012 74.015 74.026 0.010597169
39 74.017 74.013 74.036 74.025 74.026 0.008905055
40 74.010 74.005 74.029 74.000 74.020 0.011691878
```
We can plot the control charts using the `qcc()` command, specifying the type of chart required.
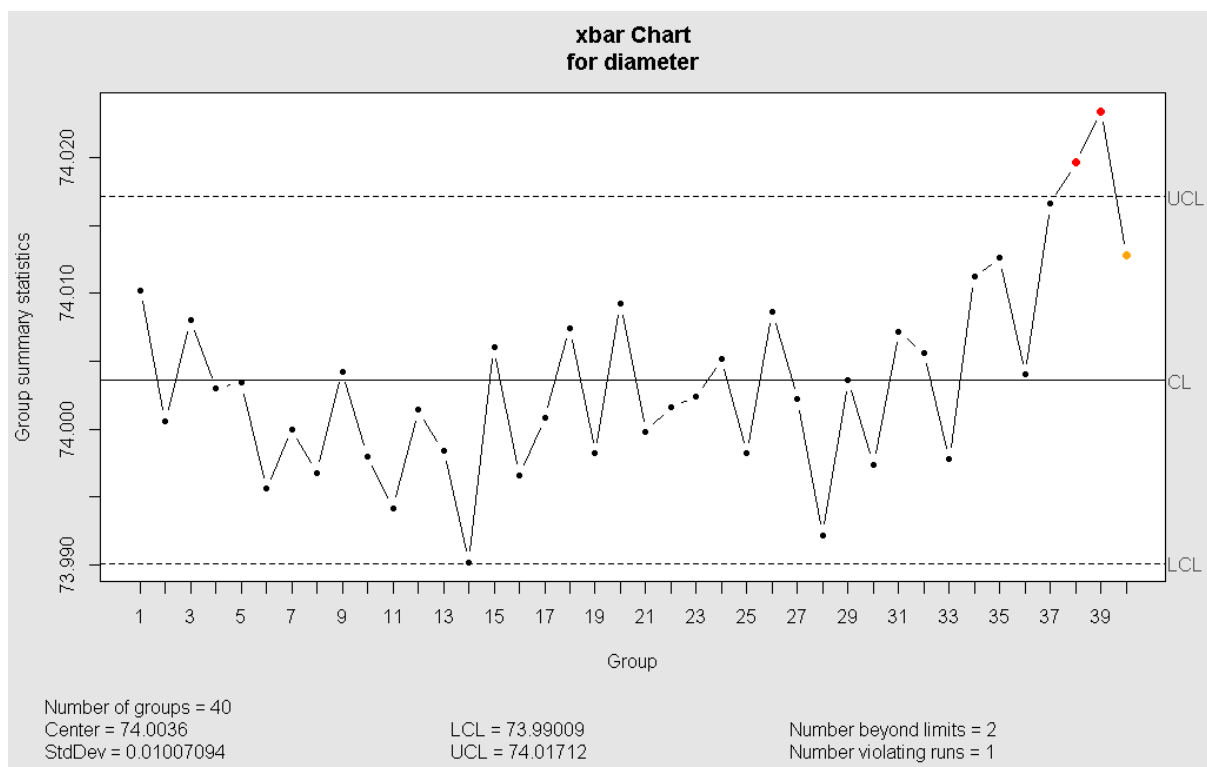
We will compute the charts for the first 25 batches (i.e. when the process is in control)

```
> plot1 = qcc(diameter[1:25,], type="xbar")
> plot2 = qcc(diameter[1:25,], type="R")
> plot3 = qcc(diameter[1:25,], type="S")
```

xbar Chart
for diameter[1:25,]

Number of groups = 25
Center = 74.00118   LCL = 73.98805   Number beyond limits = 0
StdDev = 0.009785039   UCL = 74.0143   Number violating runs = 0

R Chart
for diameter[1:25,]

Number of groups = 25
Center = 0.02276   LCL = 0   Number beyond limits = 0
StdDev = 0.009785039   UCL = 0.04812533   Number violating runs = 0

S Chart
for diameter[1:25,]

Number of groups = 25
Center = 0.009240037
StdDev = 0.009829977

LCL = 0
UCL = 0.01930242

Number beyond limits = 0
Number violating runs = 0

Now we will compute the charts again for the full set of data. In the X-bar chart, how many batches are beyond the control limits?

xbar Chart
for diameter

Number of groups = 40
Center = 74.0036
StdDev = 0.01007094

LCL = 73.99009
UCL = 74.01712

Number beyond limits = 2
Number violating runs = 1

1) Differentiate common (or chance) causes of variation in the quality of process output from assignable (or special) causes.

2) Differentiate a stable process from an unstable process.

3) Describe how the output of a stable process can be improved. What actions do not improve a stable process, but rather, make the output more variable?