MAT 2377
Probability and Statistics for Engineers

Chapter 8
Statistical Process Monitoring

P. Boily (uOttawa)

Winter 2021

Based on course notes by Rafał Kulik

Contents

Statistical Process Monitoring

A production process is often subject to **variability**. There are 2 types:

- variability due the cumulative effect of many small, essentially unavoidable causes (a process that only operates with such **common causes** is said to be **in (statistical) control**);
- variability due to **special causes**, such as improperly adjusted machines, operator errors, defective materials, etc. (the variability is typically much larger than for common causes, and the process is said to be **out of (statistical) control**).

The aim of **statistical process monitoring** (SPM) is to identify occurrence of special causes.

Why Do We Need SPM?

From *Vaughan 1997, p.383*:

> NASA engineers did not identify the association between unexpectedly low launch-pad temperatures and O-ring failures in the space shuttle booster rockets.
>
> They interpreted this critical signal as simply chance variation in the failure of the joints.
>
> Lack of this insight was critical in the decision to launch the *Challenger* on its final and disastrous flight.

Time Series

So far, we have treated samples $\{X_1, \ldots, X_n\}$ as if they arose as a result of a random experiment, i.e. X_i is drawn from some distribution with population mean μ and population variance σ^2, and we use $\overline{X}$ and S^2 as estimates of μ and σ^2.

In practice, the index i is often a **time index**, which is to say that the X_i are observed in **sequence**. In this case, we say that the sample is a **time series**.

If distribution changes over time due to external factors (war, pandemic, election, etc.) or internal factors (modification of the manufacturing process, policy change, etc.), the sample mean and the sample variance might not provide a useful summary of the situation.

To get a sense of what is going on, it is preferable to plot the data in the order that it has been collected, where the horizontal coordinate is the time of collection t (order, day, week, quarter, year, etc.) and the vertical coordinate is the observation x_t. We look for trends, cycles, shifts, etc.

Examples: the following time series record sales x (in $10,000\$$) for 3 different products, against the passage of time t (a, in years, c in weeks). Is any action necessary?

a)

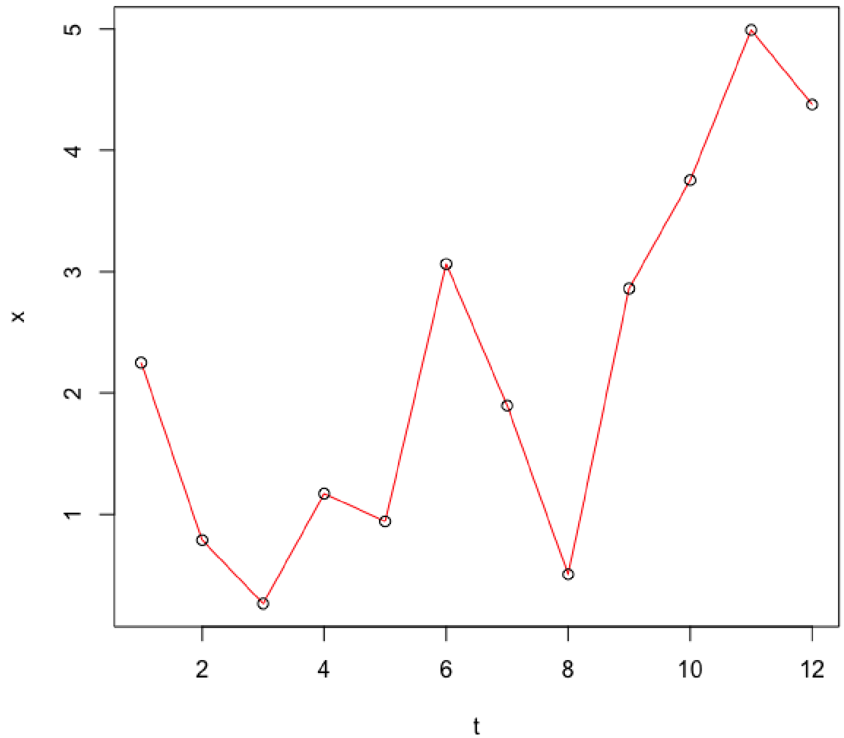
t	1	2	3	4	5	6	7	8	9	10	11	12
x	2.3	0.8	0.3	1.2	0.9	3.1	1.9	0.5	2.9	3.8	5.0	4.4

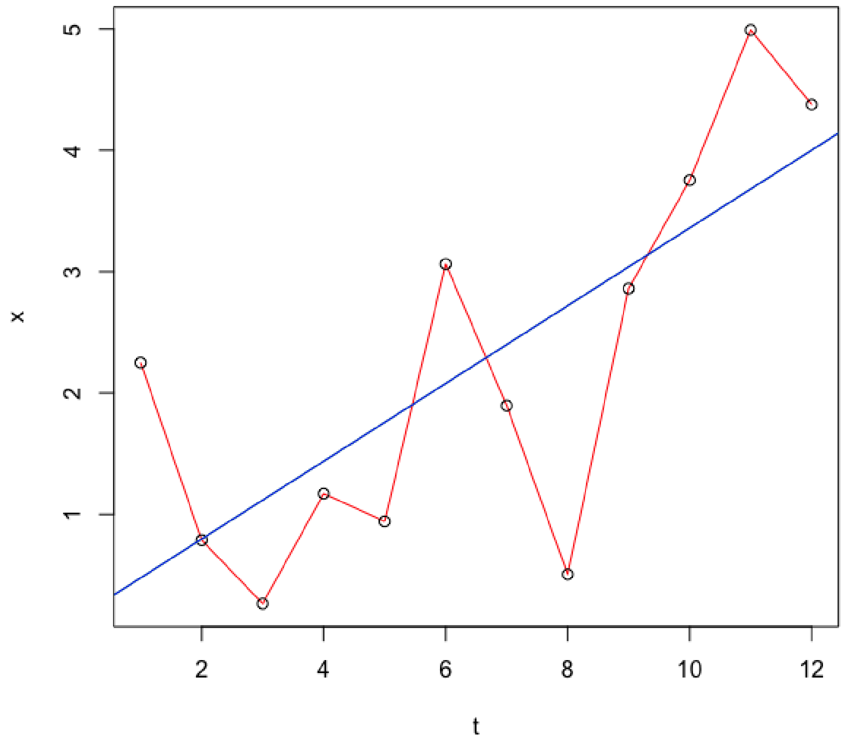
b)

t	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
x	10	25	28	13	10	12	29	25	20	31	30	29	25	28	33	31

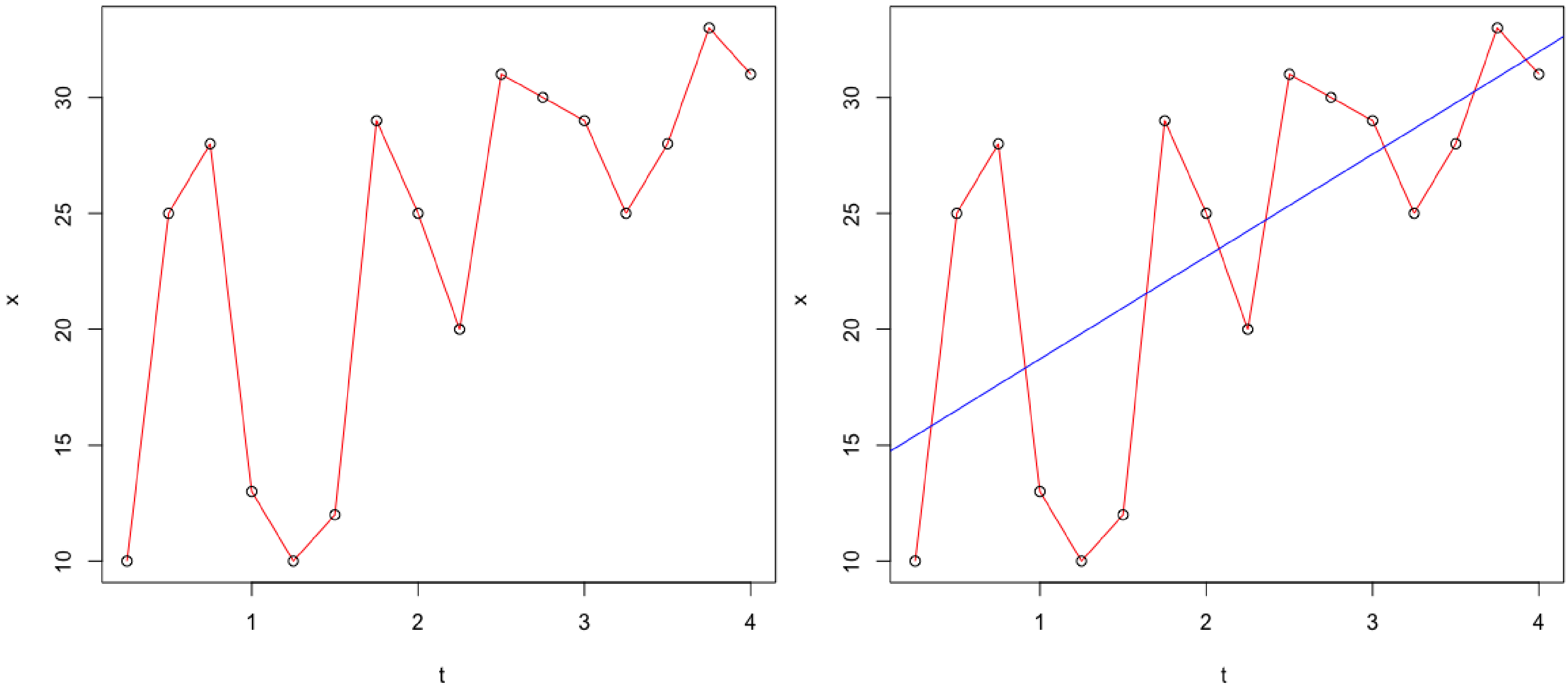
c)

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	10.8	10.1	10.2	9.4	9.9	9.8	10.1	9.6	10.0	9.7	3.2	3.4	3.4	5.5	3.4

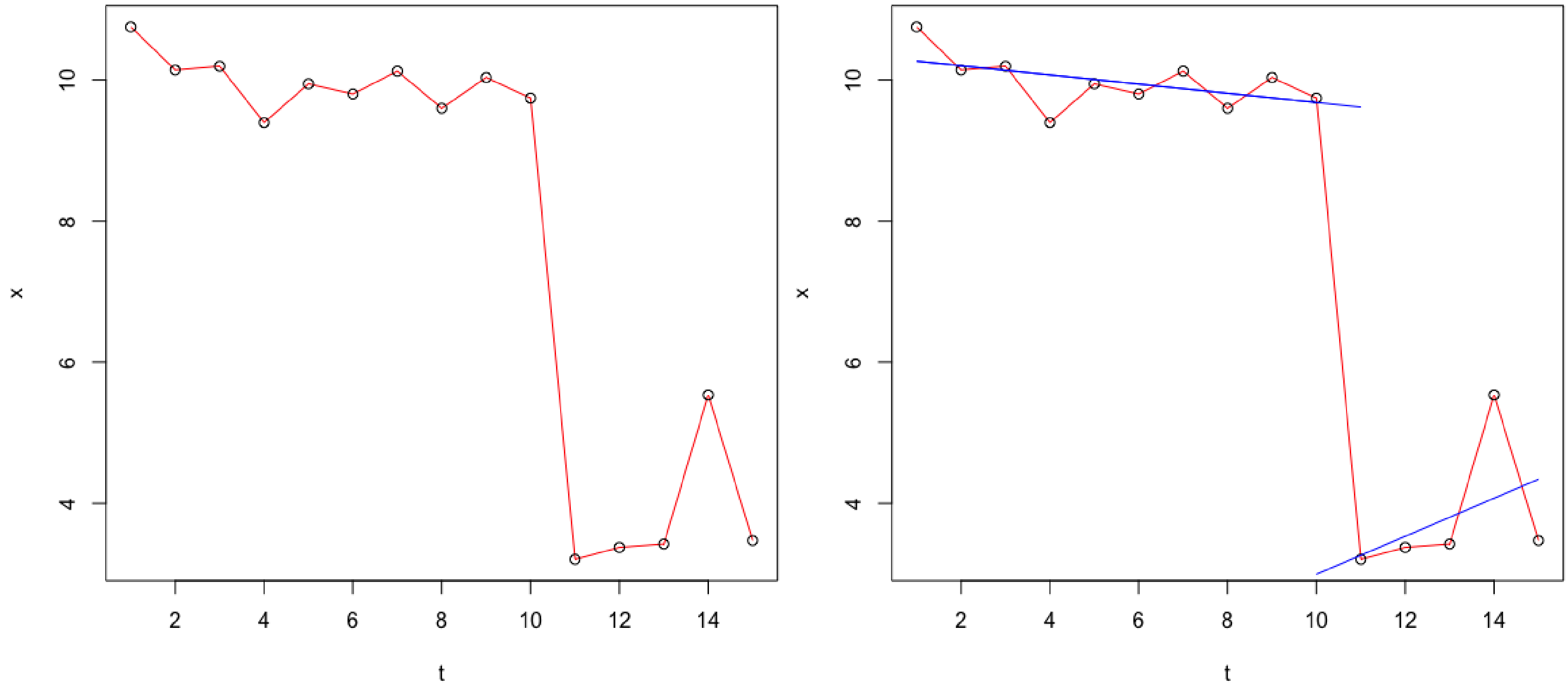




a) There are occasional drops in sales from one year to the next, but a clear upward trend. If only the last two points are presented to stockholders, they might think that there are issues and that changes have to be made.



b) There is a cyclic effect with increases from Q1 to Q2, and from Q2 to Q3, but decreases from Q3 to Q4, and from Q4 to Q1. Overall, there seems to be an upward trend, as indicated by the line of best fit.



c) Clearly, something happened after the tenth week. Whether the special causes are internal or external depend on the context (which we do not have at our disposal). Action seems to be needed.

Control Charts

We assume that the quantity of interest is defined in terms of a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. For example, the inside diameter of a piston ring is supposed to be 74mm; the standard deviation of the manufactured ring diameter is 0.01mm. Then we would have $X =$ actual ring diameter.

At given points $t_1, \ldots, t_N$ (often representing time, but not always), we sample n observations $\{X_{i,1}, \ldots, X_{i,n}\}$. Let $\overline{X}_i$ and S_i be the sample mean and the sample standard deviation, respectively, for $i = 1, \ldots, N$.

The **grand mean** and **mean standard deviation** are

$$\overline{\overline{X}} = \frac{1}{N}\sum_{i=1}^{N}\overline{X}_i \quad \text{and} \quad \overline{S} = \frac{1}{N}\sum_{i=1}^{N} S_i.$$

$\overline{X}$ Chart

A **control chart** consists of observed values of a statistic, such as $\overline{x}$ or s, plotted as a time series.

If the true mean μ and the true standard deviation σ of the process are known, then the CLT implies that

$$\overline{X}_i \sim \mathcal{N}(\mu, \sigma^2/n) \quad \text{for all } i,$$

and one would expect that the observed sample means $\overline{x}_i$ would lie in the interval $\mu \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ roughly $100(1-\alpha)\%$ of the time.

The **upper control limit** (UCL) is the upper end of the interval, the **lower control limit** (LCL) is the lower end of the interval, and the **central line** (CL) is μ.

For such charts, if we observe $\overline{x}_i > \text{UCL}$ or $\overline{x}_i < \text{LCL}$, we have an indication that the process is **instable** and potentially out of (statistical) control.

The parameter α is again interpreted as the probability of a type I error:

$$\alpha = P(\text{signal of instability} \mid \text{process is stable}).$$

Typically, we use $z_{\alpha/2} = 3$, i.e. $\alpha = 0.9973$. If $N \leq 30$, that means that even one value outside the control limits is enough to make us suspect that something is off.

In practice, however, μ and σ^2 are not known. In that case, we estimate μ by the observed grand mean $\overline{\overline{x}}$, $3\frac{\sigma}{\sqrt{n}}$ with the help of the observed mean of standard deviations $\overline{s}$:

$$\mu \approx \overline{\overline{x}}, \quad 3\frac{\sigma}{\sqrt{n}} \approx A_3(n)\overline{s}.$$

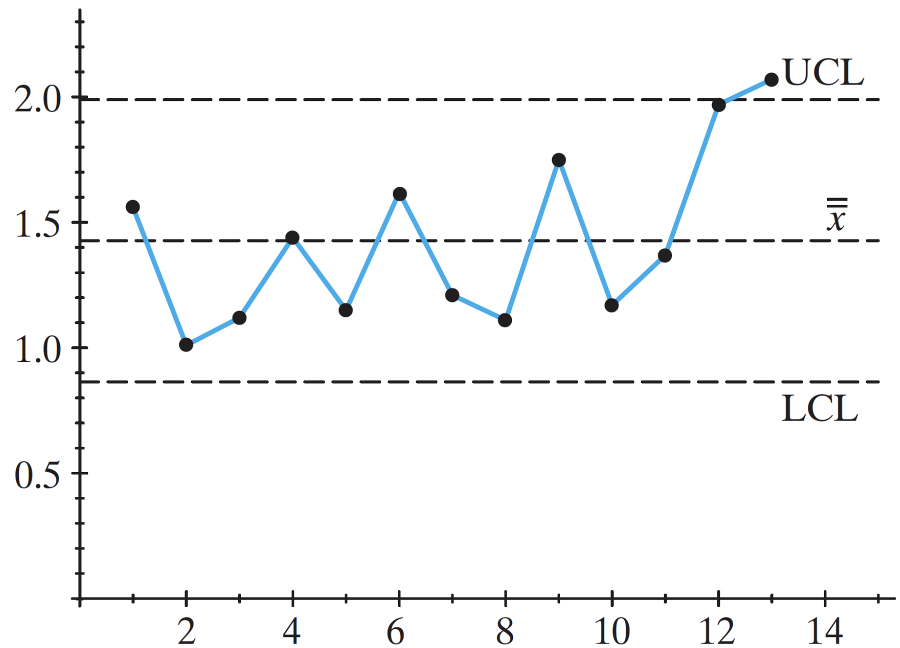
In this case, the UCL, LCL, and CL are, respectively:

$$\text{UCL} = \overline{\overline{x}} + A_3(n)\overline{s}, \ \text{LCL} = \overline{\overline{x}} - A_3(n)\overline{s}, \ \text{CL} = \overline{\overline{x}}.$$

Control Chart Constants

n	A_3	B_3	B_4	A_2	D_3	D_4
4	1.63	0	2.27	0.73	0	2.28
5	1.43	0	2.09	0.58	0	2.11
6	1.29	0.03	1.97	0.48	0	2.00
8	1.10	0.185	1.815	0.37	0.14	1.86
10	0.98	0.28	1.72	0.31	0.22	1.78
20	0.68	0.51	1.49	0.18	0.41	1.59

Example: in the chart below, $\overline{x}_{13}$ is outside the control limits in the 13th sampling period. We suspect that the process has changed and some investigation/action is needed to correct this change (upward shift?).



Control Charts (Reprise)

A **control chart** consists of:

- points representing a sample statistic taken from the process at different times
- the grand mean and the mean standard deviation of the sample statistic, which is computed using all observations, and is used to determine
  - the **center line**, which is drawn at the value of the grand mean, and
  - the **upper and lower control limits** which indicate the threshold at which the process output is considered statistically unlikely (typically three standard deviations away from the central line).

Optional features include:

- **upper and lower warning limits**, drawn as separate lines, typically two standard deviations from the central line;
- division into **zones**, with the addition of rules governing frequencies of observations in each zone;
- annotation for **events of interest**, from the point of view of process quality, and
- action on special causes

Several rule sets exist for flagging and detecting process instability and process control loss; it should be clearly stated **prior** to the start of SPM. Here is one example of such a rule set:

- a sample mean found outside the warning limits $\Rightarrow$ flag as warning
- a sample mean found outside the control limits $\Rightarrow$ process out of control
- a run of 7 successive sample means all above or all below the central line
  1. stop the production, quarantine and check
  2. adjust process and check 5 new successive samples
  3. if all good, continue process; if not $\Rightarrow$ return to step 1
- a run of 7 successive sample means all showing improvement or all showing a decrease $\Rightarrow$ follow instructions as above

$\overline{X}$ Chart from $\overline{R}$

The standard deviation σ may also be estimated by the **range** of the observations at each sampling time.

For $1 \leq i \leq N$, let $R_i = \max\{X_{i,1}, \ldots, X_{i,n}\} - \min\{X_{i,1}, \ldots, X_{i,n}\}$.

The mean of the ranges is denoted by $\overline{R}$.

If $\{r_1, \ldots, r_N\}$ are the observed values for R_i and $\overline{r}$ is the observed value of $\overline{R}$, then the UCL, LCL, and CL are, respectively:

$$\text{UCL} = \overline{\overline{x}} + A_2(n)\overline{r}, \ \text{LCL} = \overline{\overline{x}} - A_2(n)\overline{r}, \ \text{CL} = \overline{\overline{x}}.$$

This approach is preferable when $1 \leq n \leq 10$. Otherwise use the sample variance.

Other Charts

S **chart**: $\mathsf{UCL} = B_4(n)\overline{s},\ \mathsf{LCL} = B_3(n)\overline{s},\ \mathsf{CL} = \overline{s}.$

R **chart**: $\mathsf{UCL} = D_4(n)\overline{r},\ \mathsf{LCL} = D_3(n)\overline{r},\ \mathsf{CL} = \overline{r}.$

Observation chart from R:

$$\mathsf{UCL} = \overline{\overline{x}} + \sqrt{n}A_2(n)\overline{r},\ \mathsf{LCL} = \overline{\overline{x}} - \sqrt{n}A_2(n)\overline{r},\ \mathsf{CL} = \overline{\overline{x}}.$$

Observation chart from S:

$$\mathsf{UCL} = \overline{\overline{x}} + \sqrt{n}A_3(n)\overline{s},\ \mathsf{LCL} = \overline{\overline{x}} - \sqrt{n}A_3(n)\overline{s},\ \mathsf{CL} = \overline{\overline{x}}.$$

Examples:

a) A company produces a storage console. Twice a day, nine critical characteristics are tested on five consoles that are selected randomly from the production line. One of these characteristics is the time it takes the lower storage component door to open completely.

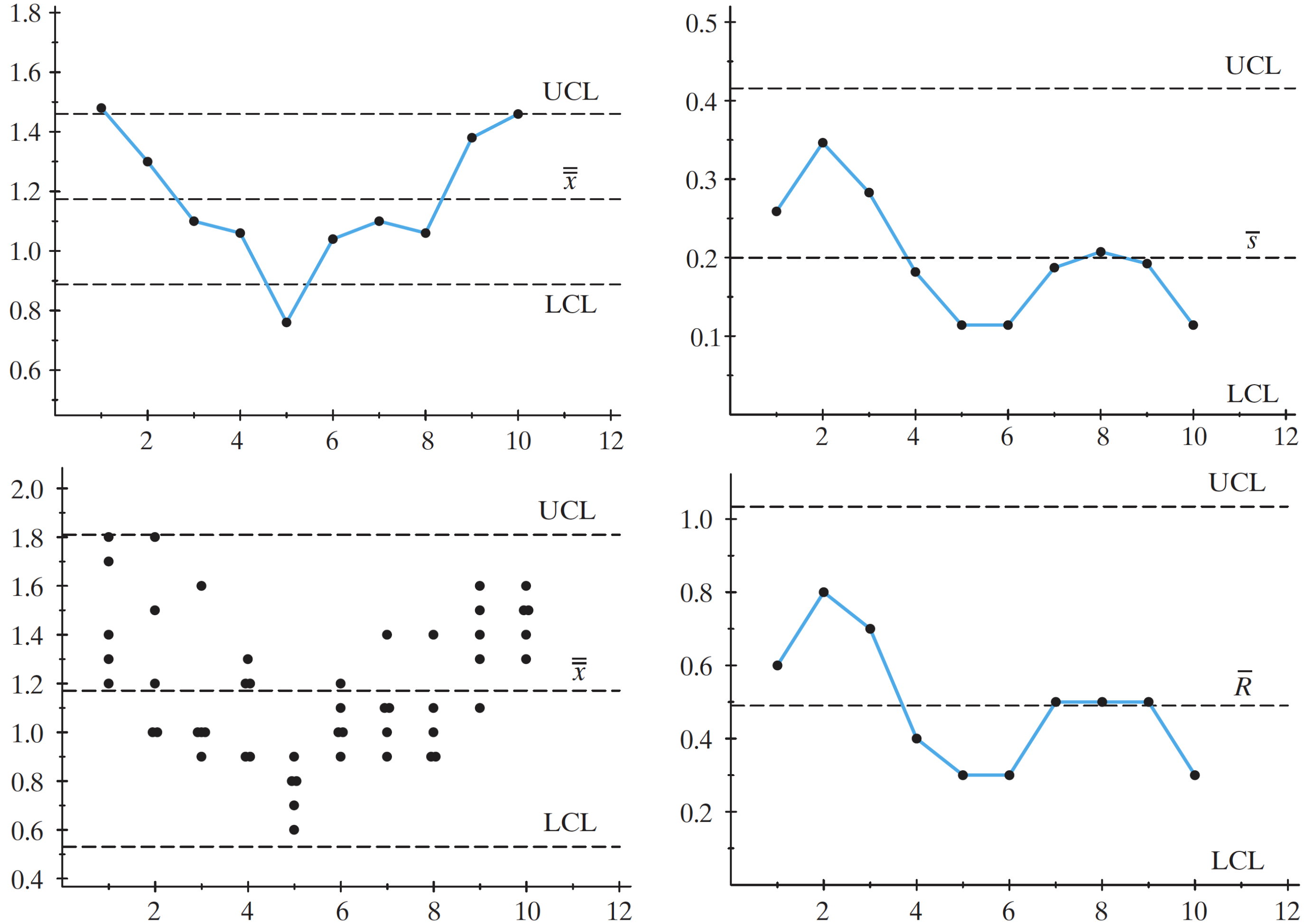
The table on the next slide lists the opening times in seconds for the consoles that were tested during one week. We have $N = 10$ and $n = 5$. Also included in the table are the sample means, sample standard deviations, and the sample ranges.

The control chart coefficients of interest are

$$A_2(5) = 0.58,\ A_3(5) = 1.17,\ B_3(5) = D_3(5) = 0,\ B_4(5) = 2.09,\ D_4(5) = 2.11.$$

We show the $\overline{X}$ chart from S, the S chart, an observation chart from R, and the R chart. The 5th group should be investigated further.

Group	x_1	x_2	x_3	x_4	x_5	$\overline{x}$	s	R
1	1.2	1.8	1.7	1.3	1.4	1.480	0.259	0.60
2	1.5	1.2	1.0	1.0	1.8	1.300	0.346	0.80
3	0.9	1.6	1.0	1.0	1.0	1.100	0.283	0.70
4	1.3	0.9	0.9	1.2	1.0	1.060	0.182	0.40
5	0.7	0.8	0.9	0.6	0.8	0.760	0.114	0.30
6	1.2	0.9	1.1	1.0	1.0	1.040	0.104	0.30
7	1.1	0.9	1.1	1.0	1.4	1.100	0.187	0.50
8	1.4	0.9	0.9	1.1	1.0	1.060	0.207	0.50
9	1.3	1.4	1.1	1.5	1.6	1.380	0.192	0.50
10	1.6	1.5	1.4	1.3	1.5	1.460	0.114	0.30
						$\overline{\overline{x}} = 1.174$	$\overline{s} = 0.200$	$\overline{R} = 0.49$


UCL
$\bar{\bar{x}}$
LCL
UCL
$\bar{s}$
LCL
UCL
$\bar{\bar{x}}$
LCL
UCL
$\bar{R}$
LCL

b) Consider the $N = 5$ samples of size $n = 4$ shown in the table below. Build various control charts for the underlying process.

i	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$x_{i,4}$	$\overline{x}_i$	r_i	s_i
1	27.1	29.4	27.2	30.0	28.43	2.90	1.49
2	30.6	32.5	32.4	31.9	31.85	1.90	0.87
3	25.7	35.5	30.0	34.8	31.50	9.80	4.57
4	31.1	23.2	25.0	22.3	25.40	8.80	3.96
5	24.1	34.2	15.2	22.0	23.88	19.00	7.86
					$\overline{\overline{x}} = 28.21$	$\overline{r} = 8.48$	$\overline{s} = 3.75$

Solution: if you want to do it by hand, use

$$A_2(4) = 0.73,\ A_3(4) = 1.63,\ B_3(4) = D_3(4) = 0,\ B_4(4) = 2.27,\ D_4(4) = 2.28.$$

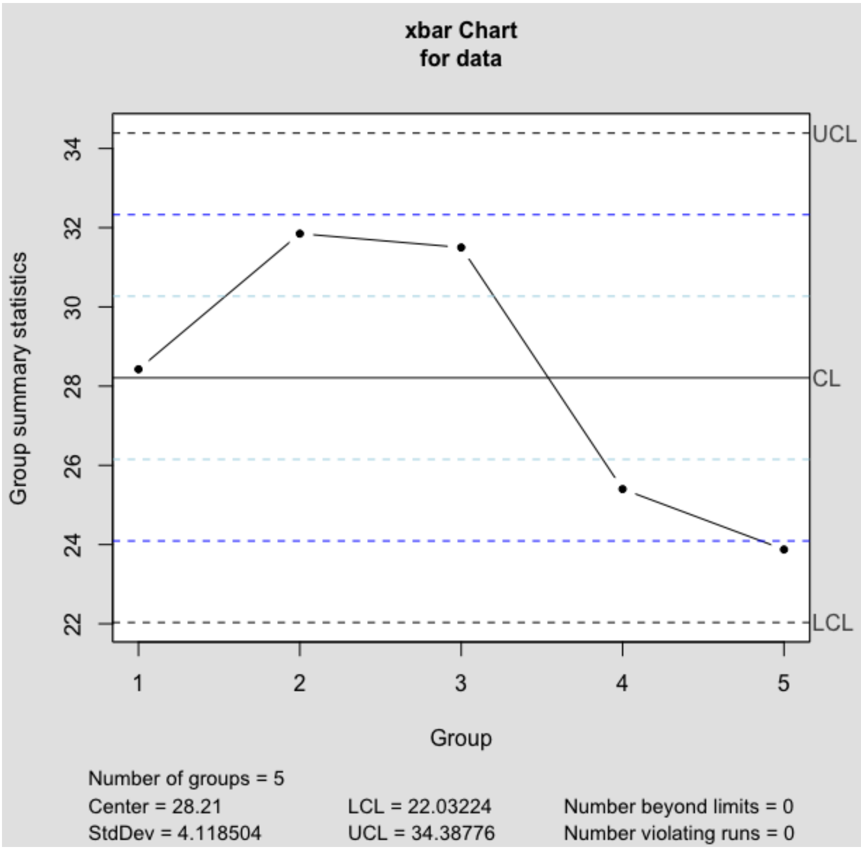
We will use the R library qcc and the following code:

```
# Import the qcc library (might need to be installed first)
library(qcc)

# Create a data frame with the measures and sample columns
measures = c(27.1,29.4,27.2,30.0,
             30.6,32.5,32.4,31.9,
             25.7,35.5,30.0,34.8,
             31.1,23.2,25.0,22.3,
             24.1,34.2,15.2,22.0)
sample = c(rep(1,4),rep(2,4),rep(3,4),rep(4,4),rep(5,4))
df <- data.frame(measures, sample)

# Group the measures per sample
measure <- with(df, qcc.groups(measures, sample))
```

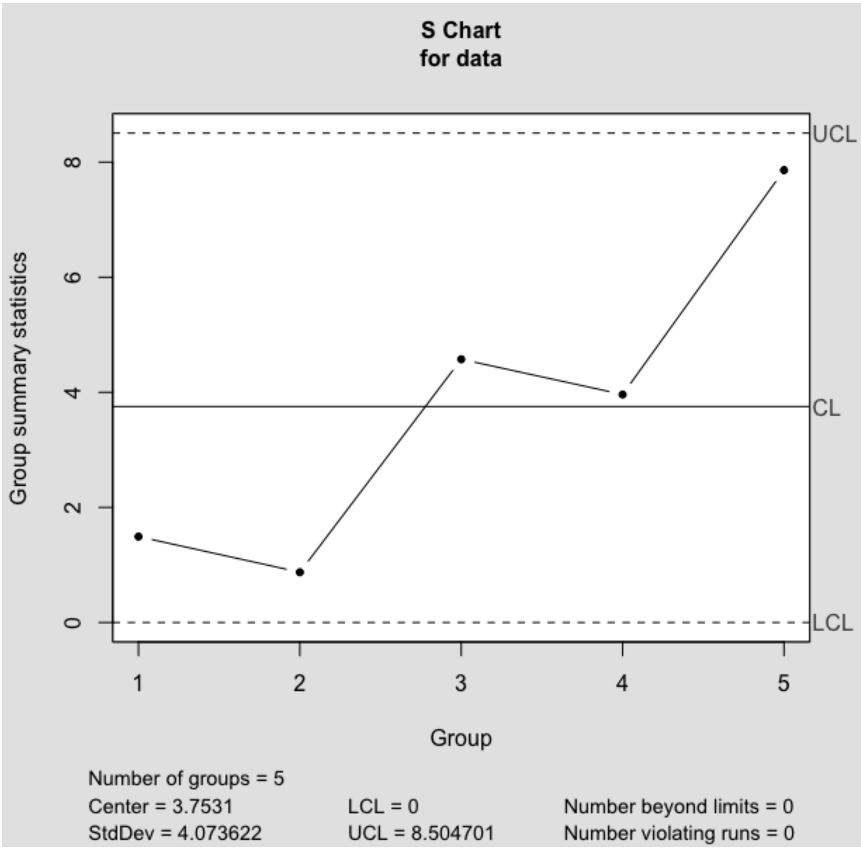
```
# Specify the measures unit or name of variable
measure_unit <- "data"

# Create the x-bar chart
xbar <- qcc(measure, type = "xbar", data.name = measure_unit)

# Specify the warning limits (2 sigmas)
(warn.limits.2 = limits.xbar(xbar$center, xbar$std.dev, xbar$sizes, 2))

# Specify the warning limits (1 sigmas)
(warn.limits.1 = limits.xbar(xbar$center, xbar$std.dev, xbar$sizes, 1))

# Plot the x-bar chart and warning limit lines
plot(xbar, restore.par = FALSE)
abline(h = warn.limits.2, lty = 2, col = "blue")
abline(h = warn.limits.1, lty = 2, col = "lightblue")
```

```
# Get the summary for x-bar chart
summary(xbar)

# Create the R-chart
r_chart <- qcc(measure, type = "R", data.name = measure_unit)

# Get the summaries for R-chart
summary(r_chart)

# Create the S-chart
s_chart <- qcc(measure, type = "S", data.name = measure_unit)

# Get the summaries for R-chart
summary(s_chart)
```

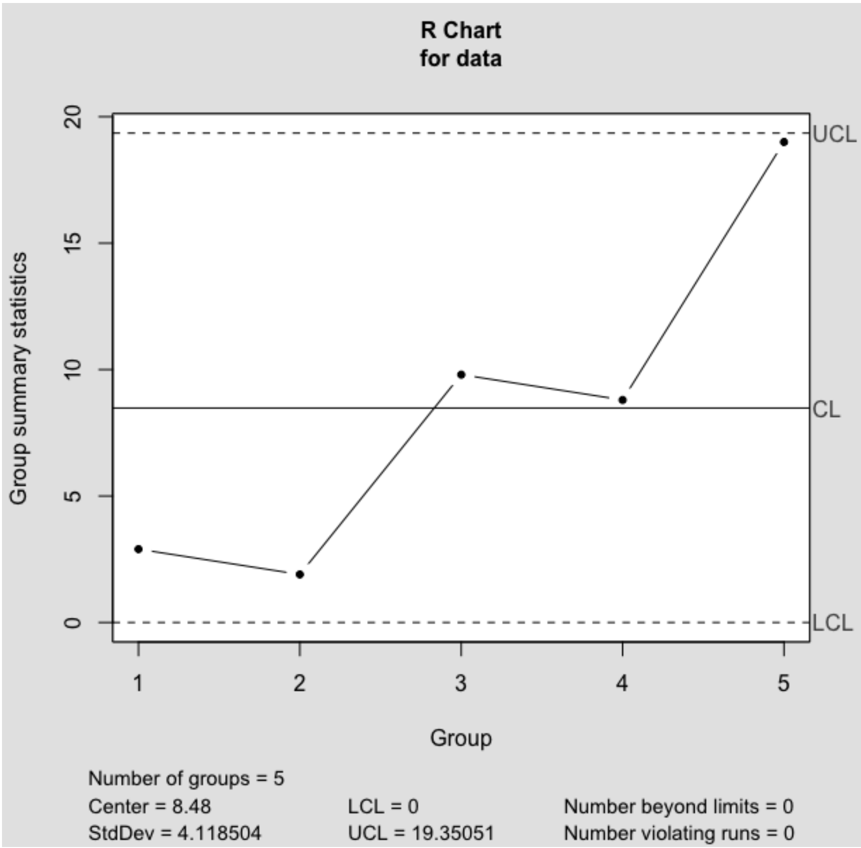
The charts are shown in the following slides, together with the summaries.




```
xbar chart for data

Summary of group statistics:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 23.875  25.400  28.425  28.210  31.500  31.850

Group sample size:  4
Number of groups:  5
Center of group statistics:  28.21
Standard deviation:  4.118504

Control limits:
      LCL      UCL
 22.03224 34.38776
```

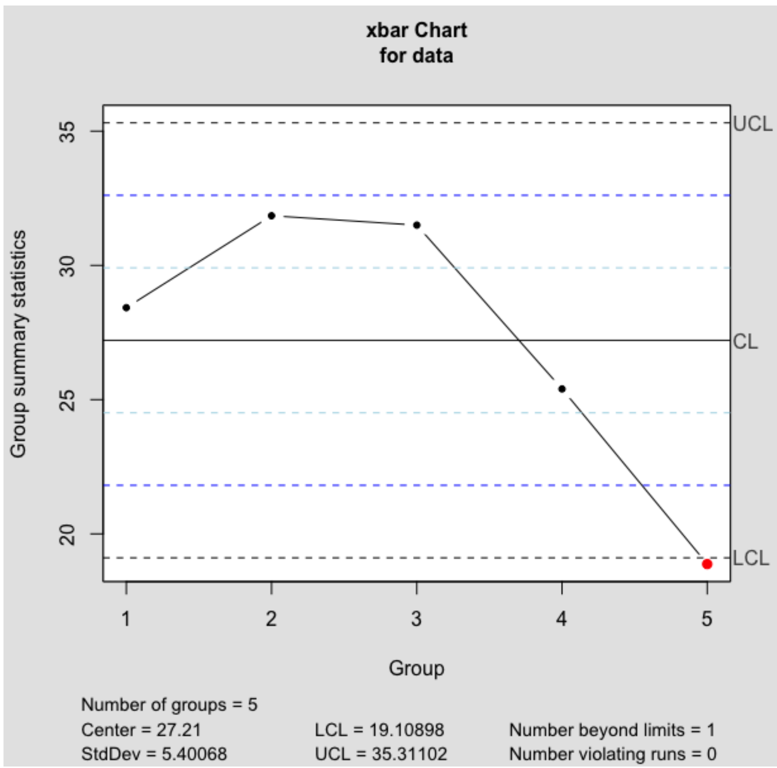



```
S chart for data

Summary of group statistics:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
0.873689 1.493039 3.962323 3.753100 4.574567 7.861881 

Group sample size:  4
Number of groups:  5
Center of group statistics:  3.7531
Standard deviation:  4.073622 

Control limits:
 LCL      UCL
   0 8.504701
```




```
R chart for data

Summary of group statistics:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.90    2.90    8.80    8.48    9.80   19.00

Group sample size:  4
Number of groups:  5
Center of group statistics:  8.48
Standard deviation:  4.118504

Control limits:
 LCL      UCL
   0 19.35051
```

c) What if the very last entry is changed to $x_{5,4} = 2.0$?




```
xbar chart for data

Summary of group statistics:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 18.875  25.400  28.425  27.210  31.500  31.850

Group sample size:  4
Number of groups:  5
Center of group statistics:  27.21
Standard deviation:  5.40068

Control limits:
      LCL      UCL
 19.10898 35.31102
```