Statistical Process Control

MSc: Statistics and Actuarial-Financial Mathematics

Konstantinos Bourazas
kbourazas@aegean.gr

Department of Statistics and Actuarial-Financial Mathematics

Course 2

February 26, 2026

$\bar{X}$ and R charts

- Many quality characteristics are **measured on a numerical scale**: length, weight, strength, concentration, temperature, ...
- We typically do not inspect items one by one
  - Instead, we collect small **rational subgroups** of size m at regular time points
  - Within each subgroup we summarise the data by a few statistics that we monitor over time
- The $\bar{X}$ **chart** monitors the process **mean level**
- The R **chart** monitors the process **short-term variability**
- In practice both charts are used **together** as a stable mean is meaningless without stable variability, and vice versa

Data collected in rational subgroups

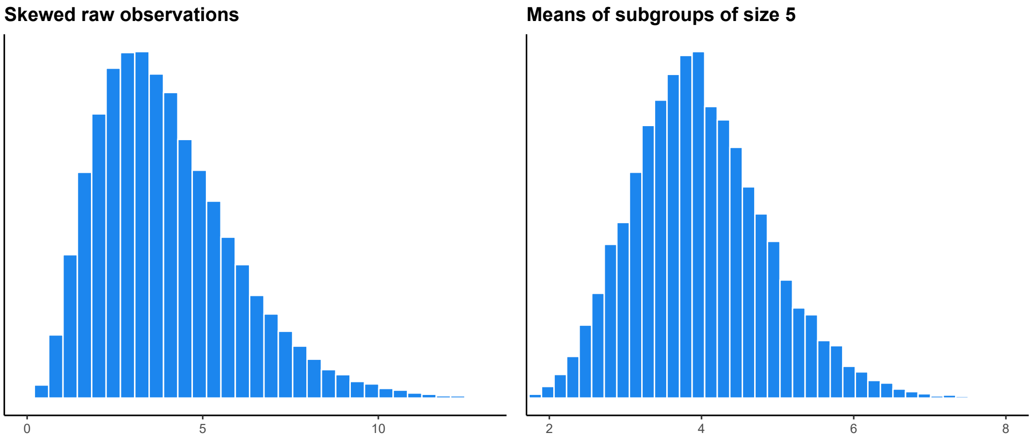
- At time point i we observe m items

$$X_{i1}, X_{i2}, \ldots, X_{im}, \qquad i = 1, 2, \ldots, n$$

- **Rational subgrouping**: items within a subgroup should be produced under **nearly the same conditions**, so that between-subgroup differences can reveal process changes
- In practice m is typically **4 or 5**
  - The subgroup mean $\bar{X}_i$ is approximately normal by the **Central Limit Theorem**, even when individual items are not
  - Larger m is rarely needed for this purpose
- Phase I uses n subgroups collected when the process is believed to be **In Control (IC)**. Then, Phase II follows and it monitors future subgroups using the limits established in Phase I

Central Limit Theorem and subgroup means

Even if individual observations X_{ij} are skewed, the subgroup mean $\bar{X}_i$ becomes much closer to normal for $m = 5$



Subgroup summaries $\bar{X}_i$ and R_i

- For each subgroup i we compute two basic summaries

$$\bar{X}_i = \frac{1}{m}\sum_{j=1}^{m} X_{ij}, \qquad R_i = \max_{1\le j\le m} X_{ij} - \min_{1\le j\le m} X_{ij}$$

- Across all n Phase I subgroups we also compute the overall averages

$$\bar{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}_i, \qquad \bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

- Interpretation
  - $\bar{X}_i$ tracks the **current process level** at time i
  - R_i tracks the **short-term spread** within the subgroup at time i

If the IC parameters were known

- Assume for a moment that both μ_0 and σ are **known**
- To check whether the process mean is IC at time i, we can test

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$$

- A natural test statistic is

$$Z_i = \frac{\bar{X}_i - \mu_0}{\sigma / \sqrt{m}} \overset{H_0}{\sim} N(0, 1)$$

- We declare OOC at time i when $|Z_i| > Z_{1-\alpha/2}$, which is equivalent to

$$\bar{X}_i > \mu_0 + Z_{1-\alpha/2} \frac{\sigma}{\sqrt{m}} \quad \text{or} \quad \bar{X}_i < \mu_0 - Z_{1-\alpha/2} \frac{\sigma}{\sqrt{m}}$$

- In practice μ_0 and σ are unknown and must be estimated from Phase I data
- This is why $\bar{\bar{X}}$, $\bar{R}$, and the constants $d_1(m)$ and $d_2(m)$ appear in the final control limits

Why estimate σ with the range instead of the sample SD

- A natural estimator of σ is the pooled sample standard deviation $\bar{s}$
- So why do $\bar{X}$ and R charts use $\bar{R}$ instead
- **Historical reason**
  - When Shewhart charts were introduced, computing a square root for every subgroup by hand was time consuming
  - The range requires only a **simple subtraction**, so it was much easier to compute routinely
- **Statistical reason**
  - For small subgroup sizes, especially $m = 4$ **or** 5, the range retains nearly all the information about σ that the sample SD contains
  - The relative efficiency of $\bar{R}$ compared with $\bar{s}$ is **above** 85% for $m \leq 6$
  - For larger m the efficiency drops, and $\bar{X}$ **and** s **charts** are preferred
- Today the computational advantage is less important, but $\bar{X}$ **and** R **charts** remain a common industry default for $m \leq 6$

Estimating σ from the average range

- In an $\bar{X}$ **and** R **chart** we need an estimate of the process standard deviation σ
  - The $\bar{X}$ chart limits depend on $\sigma/\sqrt{m}$, the standard error of the subgroup mean
  - The R chart limits depend on how much the range R_i varies when the process is **IC**
- Instead of estimating σ from subgroup standard deviations, Shewhart charts estimate it from the **average range** $\bar{R}$
- To turn $\bar{R}$ into an estimate of σ, we need to know what the range looks like under **IC** normal sampling
- Specifically, we need the **mean** and **standard deviation** of the standardised range R_i/σ — and these depend only on m

The constants $d_1(m)$ and $d_2(m)$

- When the process is **IC** and normal, $X_{ij} \sim N(\mu_0, \sigma^2)$, the standardised range R_i/σ has a distribution that depends **only on** m
- We define two constants that capture this distribution

$$d_1(m) = E\left[\frac{R_i}{\sigma}\right], \qquad d_2(m) = \sqrt{\mathrm{Var}\left[\frac{R_i}{\sigma}\right]}$$

- Because they depend only on m, they can be computed once and tabulated (see next slide)
- They give us plug-in estimators based on $\bar{R}$

$$\hat{\sigma} = \frac{\bar{R}}{d_1(m)}, \qquad \widehat{\mathrm{SD}}(R_i) = \frac{d_2(m)}{d_1(m)}\,\bar{R}$$

- $\hat{\sigma}$ feeds into the $\bar{X}$ chart limits; $\widehat{\mathrm{SD}}(R_i)$ feeds into the R chart limits

Table of constants $d_1(m)$ and $d_2(m)$ (Qiu, Table 3.1)

m	$d_1(m)$	$d_2(m)$	m	$d_1(m)$	$d_2(m)$
2	1.128	0.853	14	3.407	0.763
3	1.693	0.888	15	3.472	0.756
4	2.059	0.880	16	3.532	0.750
5	2.326	0.864	17	3.588	0.744
6	2.534	0.848	18	3.640	0.739
7	2.704	0.833	19	3.689	0.734
8	2.847	0.820	20	3.735	0.729
9	2.970	0.808	21	3.778	0.724
10	3.078	0.797	22	3.819	0.720
11	3.173	0.787	23	3.858	0.716
12	3.258	0.778	24	3.895	0.712
13	3.336	0.770	25	3.931	0.708

Control limits for the $\bar{X}$ chart

- From the previous slide, we declare OOC when $\bar{X}_i$ falls outside $\mu_0 \pm Z_{1-\alpha/2}\,\sigma/\sqrt{m}$
- In practice both μ_0 and σ are unknown, so we replace them with their Phase I estimates

$$\mu_0 \;\to\; \bar{\bar{X}}, \qquad \frac{\sigma}{\sqrt{m}} \;\to\; \frac{\bar{R}}{d_1(m)\sqrt{m}} = \hat{\sigma}_{\bar{X}}$$

- The Upper Control Limit (UCL), the Centre Line (CL), and the Lower Control Limit (LCL) are:

$$\text{UCL} = \bar{\bar{X}} + \frac{Z_{1-\alpha/2}}{d_1(m)\sqrt{m}}\,\bar{R}, \qquad \text{CL} = \bar{\bar{X}}, \qquad \text{LCL} = \bar{\bar{X}} - \frac{Z_{1-\alpha/2}}{d_1(m)\sqrt{m}}\,\bar{R}$$

- The structure is simply $\bar{\bar{X}} \pm Z_{1-\alpha/2}\,\hat{\sigma}_{\bar{X}}$, identical to the known-parameter case with both parameters estimated
- In practice $Z_{1-\alpha/2} = 3$ is the universal convention
  - It corresponds to $\alpha = 2\,\Phi(-3) \approx 0.0027$, i.e. a **false alarm rate of** 0.27% **per subgroup** when IC
  - Equivalently, under IC the chart signals on average once every $1/0.0027 \approx 370$ subgroups, a concept we return to shortly
- A point outside the limits is evidence that the process mean is **OOC**

Control limits for the R chart

- We monitor R_i around its IC baseline $\bar{R}$
- The estimated standard deviation of R_i is $\widehat{\mathrm{SD}}(R_i) = d_2(m)\,\bar{R}\,/\,d_1(m)$
- The R chart limits are therefore

$$\mathrm{UCL} = \left(1 + \frac{Z_{1-\alpha/2}\, d_2(m)}{d_1(m)}\right)\bar{R}, \quad \mathrm{CL} = \bar{R}, \quad \mathrm{LCL} = \left(1 - \frac{Z_{1-\alpha/2}\, d_2(m)}{d_1(m)}\right)\bar{R}$$

- Since $R_i \geq 0$, a negative LCL is set to 0
  - For $m = 5$: $\mathrm{LCL} = (1 - 3 \times 0.864/2.326)\,\bar{R} = -0.114\,\bar{R} < 0$, so $\mathrm{LCL} = 0$
  - The LCL becomes positive only for $m \geq 7$
- A point above the UCL signals **increased variability**; a point below a positive LCL signals **decreased variability** (often a sign of quality improvement)

Phase I workflow for $\bar{X}$ and R charts

- Collect n rational subgroups of size m under conditions believed to be **IC**
- Compute the subgroup summaries $\bar{X}_i$ and R_i for each $i = 1, \ldots, n$
- Compute $\bar{\bar{X}}$ and $\bar{R}$ and obtain $d_1(m)$ and $d_2(m)$ from a table
- Choose $Z_{1-\alpha/2}$ for the desired false alarm rate
  - A common default is $Z_{1-\alpha/2} = 3$, which makes false alarms rare under IC
  - The exact choice depends on the cost of false alarms versus missed detection
- **First** plot the R chart and check that variability is stable
  - The $\bar{X}$ chart limits depend on $\bar{R}$, so if variability is unstable those limits are not trustworthy
  - Only if the R chart looks stable do we proceed to the $\bar{X}$ chart
- If special causes are found, investigate, adjust, and rebuild the Phase I baseline

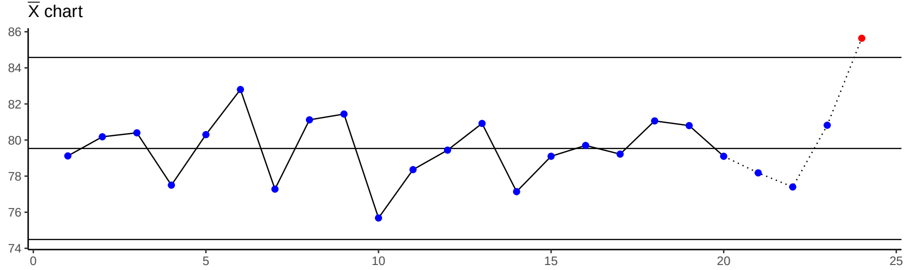
Phase II interpretation using the established limits

- In Phase II, new subgroups arrive sequentially and we compute $\bar{X}_i$ and R_i each time
- We compare each new point against the Phase I control limits
  - If a point crosses a limit, we treat it as an **alarm** and investigate the process
  - If points remain within limits, the data are consistent with the IC baseline
- When an alarm occurs, the practical response is also sequential
  - Stop and diagnose when appropriate
  - Remove the special cause and adjust the process
  - Restart monitoring after we are confident the process is back to IC

Example 3.1 Injection molding process

- We monitor an injection molding process where the quality characteristic is the **compressive strength** of manufactured parts
- $n = 20$ Phase I subgroups are collected, each with subgroup size $m = 5$
- For each subgroup, we compute $\bar{X}_i$ and R_i
- We then build Phase I limits using $Z_{1-\alpha/2} = 3$ and the constants for $m = 5$
- Finally, we monitor four additional Phase II subgroups and check whether the charts signal

Example 3.1 $\bar{X}$ and R charts



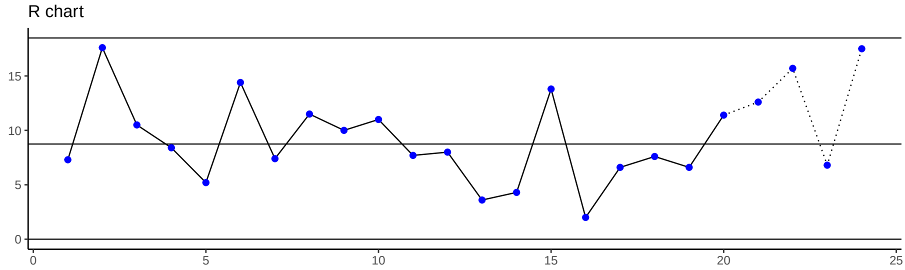


Chart performance: the Average Run Length

- How do we measure how well a control chart behaves when the process is **IC**?
- At each time point, the probability of a false alarm is α independently of all other time points, so the number of subgroups until the first false alarm follows a **Geometric**(α) distribution
- Its mean is the **IC Average Run Length**

$$\mathrm{ARL}_0 = \frac{1}{\alpha}$$

- For the conventional choice $Z_{1-\alpha/2} = 3$, i.e. $\alpha = 0.0027$

$$\mathrm{ARL}_0 = \frac{1}{0.0027} \approx 370$$

- Interpretation: when the process is IC, the chart signals on average once every **370 subgroups**
- A larger ARL_0 means fewer false alarms. In practice, we want ARL_0 to be large under IC and small when the process shifts

Multiple testing in Phase I: the overall false alarm rate

- In Phase I we apply the chart to n subgroups **simultaneously**
- Even if the process is IC throughout, each subgroup has probability α of triggering a false alarm independently, so

$$\begin{aligned} P(\text{at least one false alarm}) &= 1 - P(\text{no false alarm in any of the } n \text{ subgroups}) \\ &= 1 - (1-\alpha)^n \;=\; \tilde{\alpha} \end{aligned}$$

- This is directly analogous to the **Family-Wise Error Rate (FWER)** in multiple testing
- For $\alpha = 0.0027$ and typical Phase I sample sizes

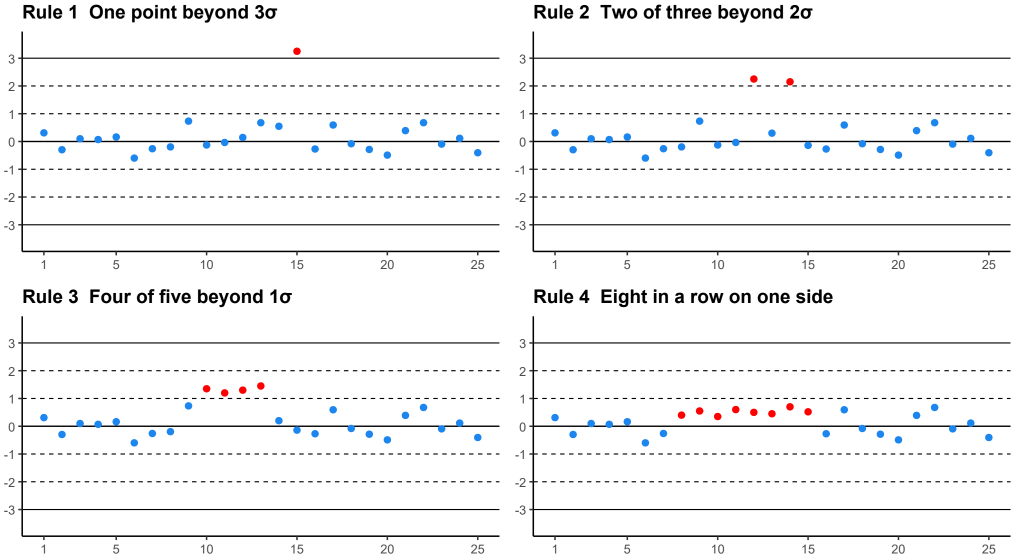
$$n = 10 \;\Rightarrow\; \tilde{\alpha} \approx 2.7\%, \qquad n = 20 \;\Rightarrow\; \tilde{\alpha} \approx 5.2\%, \qquad n = 50 \;\Rightarrow\; \tilde{\alpha} \approx 12.6\%$$

- Consequence: even with the conservative 3σ limits, spurious signals in Phase I are not negligible for large n
- If tighter control of $\tilde{\alpha}$ is needed, choose α via $\alpha = 1 - (1-\tilde{\alpha})^{1/n}$

Beyond the 3σ rule: run patterns

- A control chart can reveal problems even when **no point crosses a limit**
- Non-random patterns such as long runs, trends, or periodic behaviour are unlikely under IC and suggest special causes
- The **Western Electric Handbook (1956)** codified four supplementary rules; the process is declared OOC if any of the following occur
  - One point outside the 3σ limits
  - Two out of three consecutive points beyond the 2σ limits
  - Four out of five consecutive points more than 1σ from the centre line
  - Eight consecutive points on the same side of the centre line
- The motivation is clear: these patterns are unlikely under IC and may indicate a small or gradual shift that a single point cannot detect

Western Electric rules in four examples



Limitations of the Western Electric rules

- The rules are largely **ad hoc**, as they were motivated by intuition rather than derived from a formal statistical framework
- Applying multiple rules simultaneously **inflates the Type I error rate**
  - Each additional rule increases the chance of a false alarm
  - The rules are correlated, so the overall $\tilde{\alpha}$ is difficult to compute or control
- We discuss them here for **didactic purposes**, as they build intuition about what non-random behaviour looks like on a control chart
- In modern practice, **CUSUM and EWMA charts** offer a principled and statistically rigorous solution to detecting small or gradual shifts, with explicit control of the false alarm rate