Control Charts, Out-of-Control Signals, and Process Capability

Statistical Process Control in Semiconductor Manufacturing

ECE 8803 · AI for Semiconductor Manufacturing

Asif Khan

Associate Professor, School of Electrical and Computer Engineering

School of Materials Science and Engineering · Georgia Institute of Technology

akhan40@gatech.edu

Georgia Tech

What you should be able to do by the end of today

1. Build an $\bar{X}$–R chart from raw subgrouped data, and read the two charts in the right order
2. Choose the right chart for the data you actually have — continuous, n = 1, or attribute
3. Interpret the run rules, and name the physical mechanism each pattern points at
4. Explain what it costs to add a rule, and decide which ones to switch on
5. Compute Cp and Cpk, and say precisely what each one does and does not tell you

Georgia Tech

Where we left off

Lecture 1 in four lines

Variation has two causes

Common cause is the system. Special cause is an event. They demand opposite responses.

σ is a distance

The typical gap between a measurement and the mean, in the units you measured.

z is a universal ruler

Distance from centre in units of σ. Control limits, capability and ppm are all z-scores.

3σ is a choice

0.27% false alarms, one per 370 points. An economic compromise, not a law.

Today we turn the last two into an actual chart, and then ask whether the process is good enough.

Georgia Tech

Three questions, in this order

And you may not skip one

1. Is it stable?

Does the process behave today the way it behaved yesterday?

Control charts. Segments 2 to 5 of today.

2. What is it telling me?

A signal is not a diagnosis. Which pattern, and what physical mechanism does that pattern point at?

Run rules.

3. Is stable good enough?

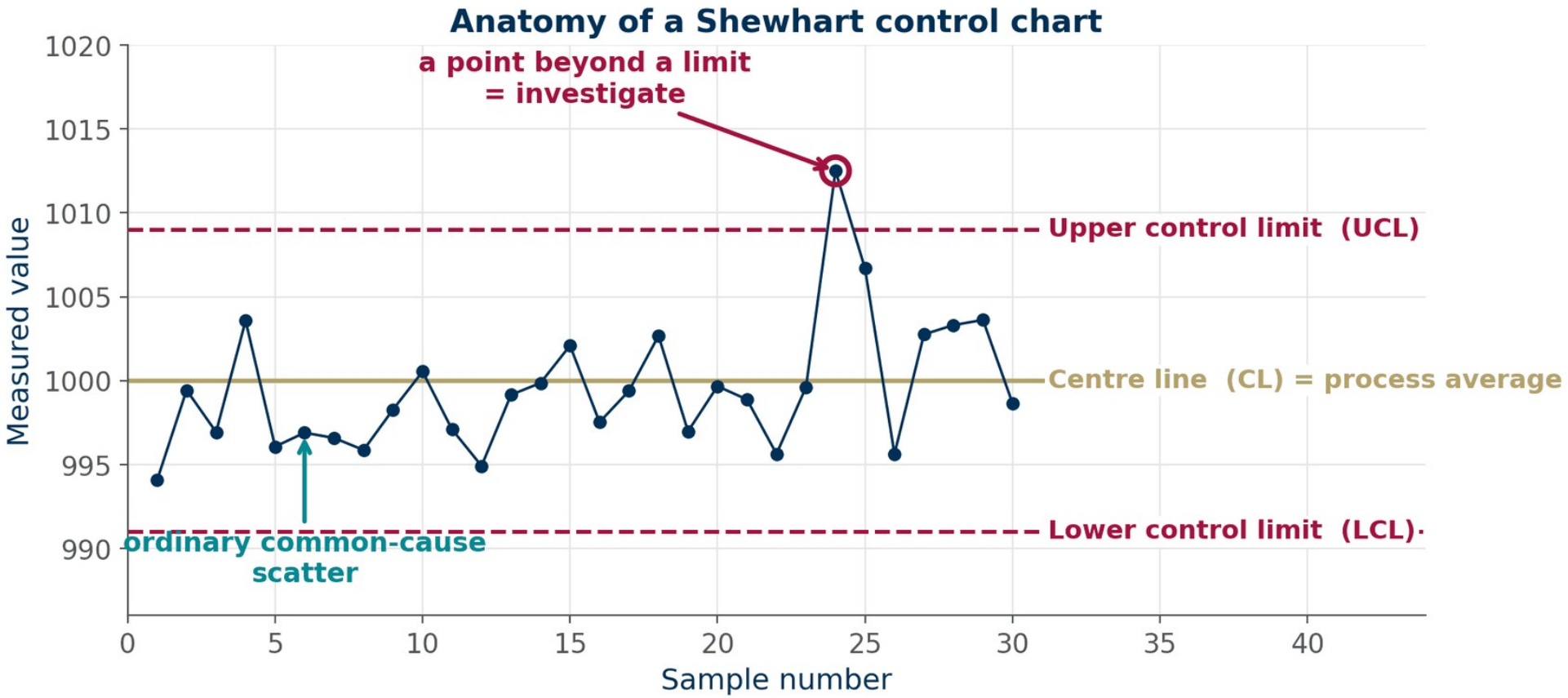
A perfectly stable process can produce nothing but scrap. Only now can you ask this.

Capability: Cp and Cpk.

Computing a capability index for an unstable process is meaningless — there is no single distribution to compute it from.

Georgia Tech

Anatomy of a control chart



Three lines and a rule for what to do when a point falls outside them.

What a control chart quietly assumes

Four things. All of them fail somewhere in a fab.

The plotted statistic is roughly normal

True for subgroup means by the central limit theorem. Much shakier for individuals, and for counts at low defect rates.

Consecutive points are independent

Frequently false. Chamber state carries over, and run-to-run control introduces correlation on purpose.

σ is constant while you are watching

The R chart is precisely the test of this — which is why you read it first.

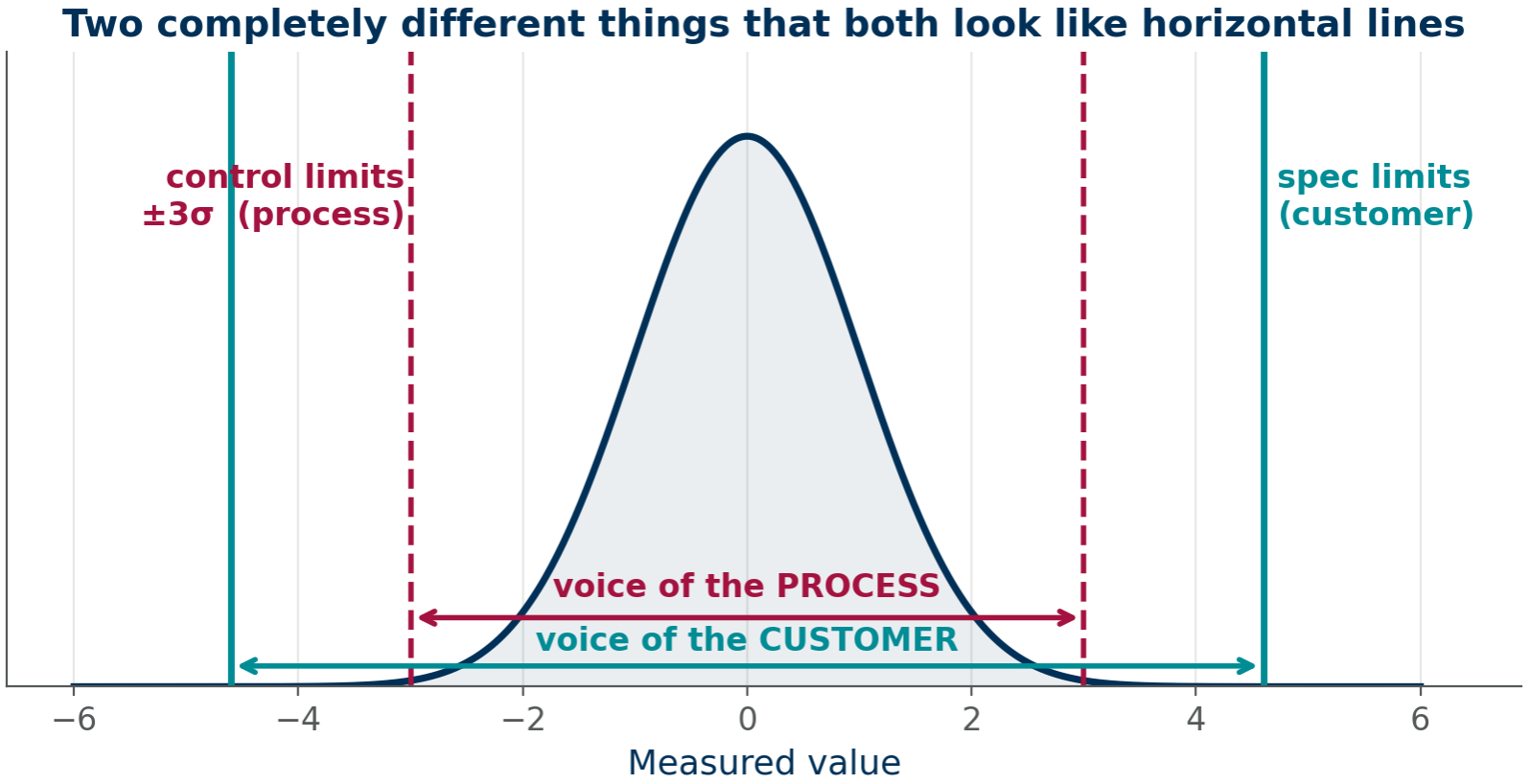
The measurement system is trustworthy

Every σ on today's slides includes the metrology tool's own variation. Lecture 3.

The chart still works when these are approximately true. Knowing which one is failing is how you debug a chart that misbehaves.

Georgia Tech

Control limits are not specification limits



One is computed from the process. The other is handed to you by the customer.

Georgia Tech

Four situations, four different jobs

In control · Capable

The goal. Stable, and comfortably inside spec.

Job: leave it alone and monitor.

In control · Not capable

Predictable, and predictably making scrap.

Job: a system change — tool, recipe, or spec.

Out of control · Capable

Getting away with it. Wide enough margin to absorb the excursions — for now.

Job: find the special cause before the margin runs out.

Out of control · Not capable

Firefighting. And you cannot even predict how bad tomorrow will be.

Job: stabilise first. Capability is not yet a meaningful question.

Stability and capability are independent axes. Students lose marks every year for collapsing them into one.

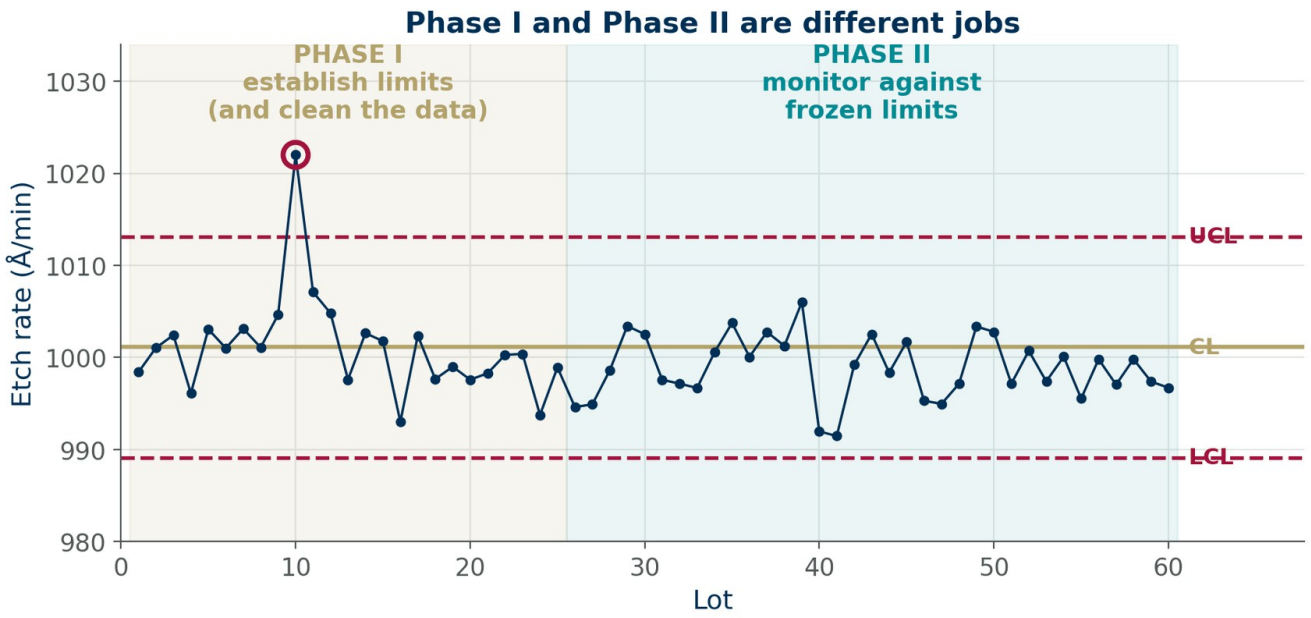
Georgia Tech

THE MOST COMMON ERROR IN THIS FIELD

A point inside the control limits can still be out of specification.

And a point outside the control limits can be perfectly saleable.

Phase I and Phase II



- **Phase I — establish the limits.**
- Collect 20–25 subgroups from a period you believe was stable. Compute trial limits. Investigate any point that fails, and remove it only if you find an assignable cause.
- Recompute. Iterate until what remains is clean.
- **Phase II — monitor.**
- Freeze those limits and plot new data against them. Do not recompute because the process 'moved' — that is the signal you built the chart to see.

Limits computed from contaminated data are the most common reason a chart never fires again.

Georgia Tech

Phase I in practice

Where the judgement lives

How much data?

20–25 subgroups is the conventional minimum. Fewer and your limits are so uncertain they are nearly arbitrary — the same confidence-interval problem we will hit again with Cpk at the end of today.

What may I delete?

A point with a documented assignable cause: a logged recipe error, a failed qualification, a known bad material lot.

Never a point that is merely inconvenient.

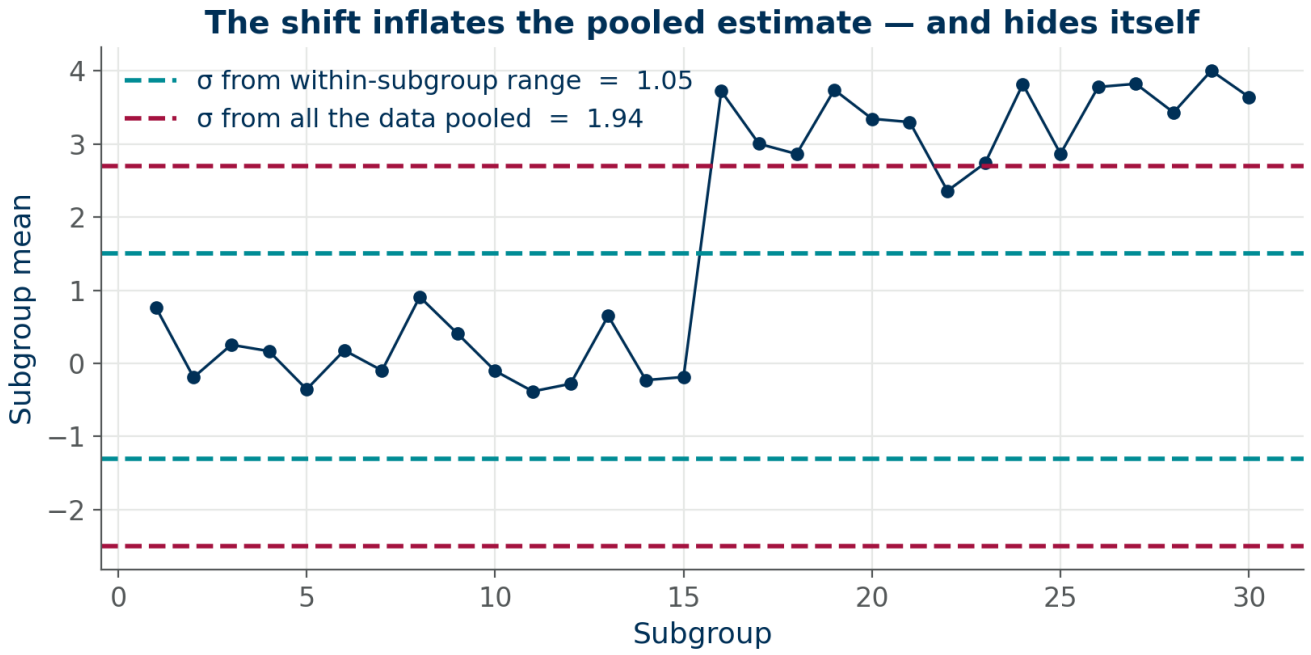
When do I redo them?

After a deliberate process change — new tool, new recipe, new chamber. Never because the chart started alarming.

'The chart keeps alarming, so we widened the limits' is the single most destructive sentence in industrial SPC.

Georgia Tech

Estimating σ: why not just take the standard deviation?



- The obvious move is to pool all the data and compute σ directly.
- **It is wrong, and it fails in the worst possible direction.**
- If the process shifted during your baseline, the pooled estimate absorbs the shift. σ comes out too large, the limits come out too wide, and the chart cannot see the very thing that inflated it.
- **So we estimate σ from within-subgroup variation only —** $\hat{\sigma} = \bar{R} / d_2$ — which is blind to shifts between subgroups.

Short-term variation sets the limits. Between-subgroup variation is what the chart is trying to detect.

Georgia Tech

The constants, and where they come from

You look them up. But you should know what they are.

n	d_2	A_2	D_3	D_4	What it does
2	1.128	1.880	0	3.267	$\hat{\sigma} = \bar{R}/d_2$
3	1.693	1.023	0	2.574	$\bar{X}$ limits = $\bar{\bar{X}} \pm A_2\bar{R}$
4	2.059	0.729	0	2.282	R limits = $D_3\bar{R}$, $D_4\bar{R}$
5	2.326	0.577	0	2.114	the fab default
6	2.534	0.483	0	2.004	
7	2.704	0.419	0.076	1.924	LCL on R stops being zero

d_2 is the expected range of n draws from a standard normal. That is all. Divide your average range by it and you have an unbiased estimate of σ — computable by an operator with a pencil.

Georgia Tech

Which chart?

Answer two questions and the chart is chosen for you

Continuous, n > 1

$\bar{X}$ and R (n ≤ 10)

$\bar{X}$ and S (n > 10)

Subgrouped CD, thickness, overlay. The most sensitive option — use it whenever the metrology budget allows.

Continuous, n = 1

Individuals and Moving Range

One site, one wafer, one lot. The fab workhorse — and the chart with the weakest assumptions.

Attribute

p, np (proportions)

c, u (counts)

Probe yield, particle excursions, visual defects. Cheap data, but you need a great deal of it.

The chart follows from the data you can afford to collect — which is a metrology decision, not a statistics one.

Georgia Tech

First, the subgroup

Rational subgrouping, from Lecture 1

The rule

Whatever varies inside the subgroup becomes the noise the chart is measured against.

Sources of variation inside the subgroup set the limits. Sources between subgroups are what the chart detects.

So form subgroups deliberately

Five sites on one wafer → limits calibrated against within-wafer signature.

One site on each of five wafers → limits calibrated against wafer-to-wafer.

Same n. Same arithmetic. Different chart, different blind spots.

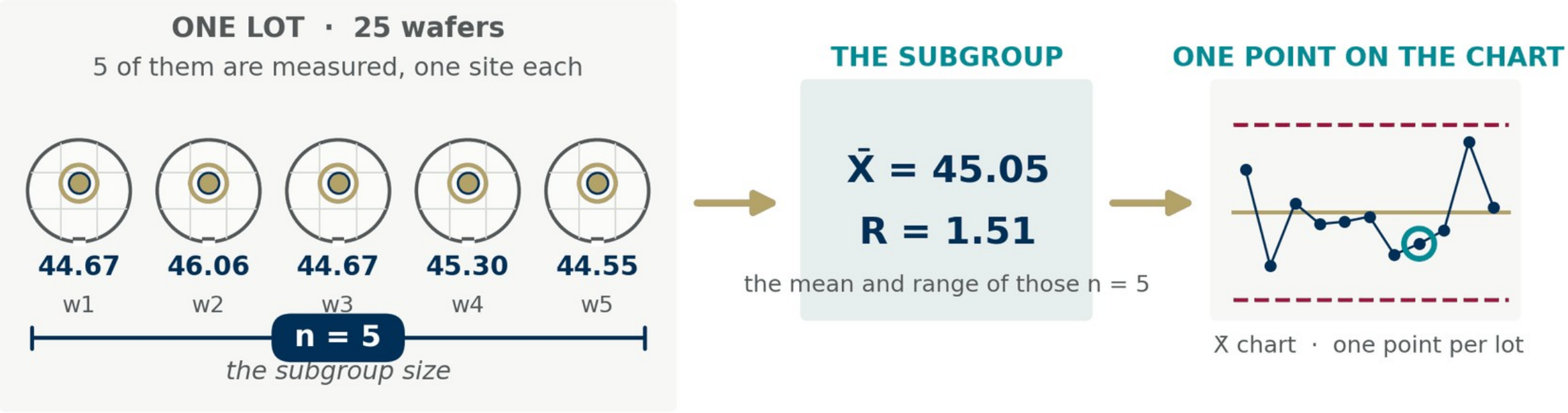
Get this wrong and every number downstream is answering a question you did not ask.

Decide what you want the chart to be blind to. That decision is the subgroup.

Georgia Tech

What a subgroup is

n = the subgroup size — how many measurements make one point on the chart



125 wafer measurements → 60 plotted numbers → two charts you can actually read

A subgroup is the set of measurements treated as one sample. Its size is n — here n = 5. One subgroup in, one plotted point out.

Georgia Tech

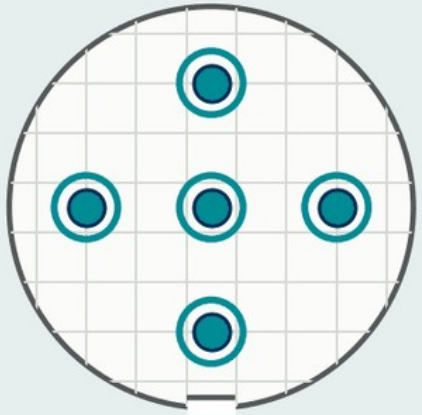
Same n, two different subgroups

n is only half the specification. What goes into the n is the other half.

OPTION 1

n = 5

5 sites on one wafer



one wafer, five measurements

Within the subgroup:

centre-to-edge non-uniformity

the tool's radial signature

Blind to: a change in that signature

OPTION 2

n = 5

1 site on each of five wafers

five wafers, one measurement each

Within the subgroup: wafer-to-wafer variation

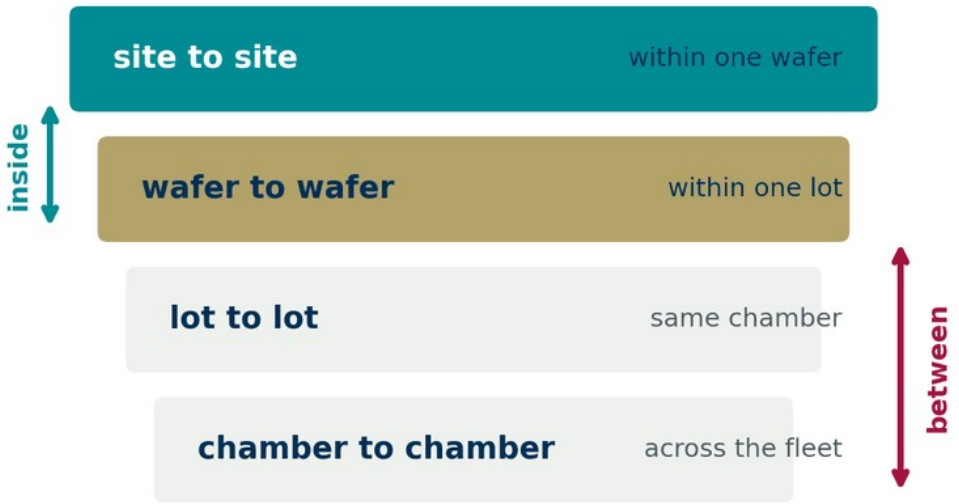
Blind to: what happens across a wafer

Georgia Tech

Why the choice matters

Rational subgrouping — the most consequential design decision in SPC

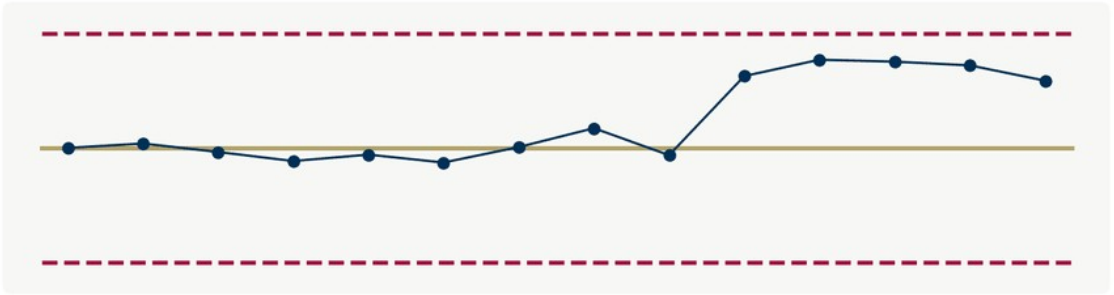
The variation is nested



What is INSIDE sets the limits.

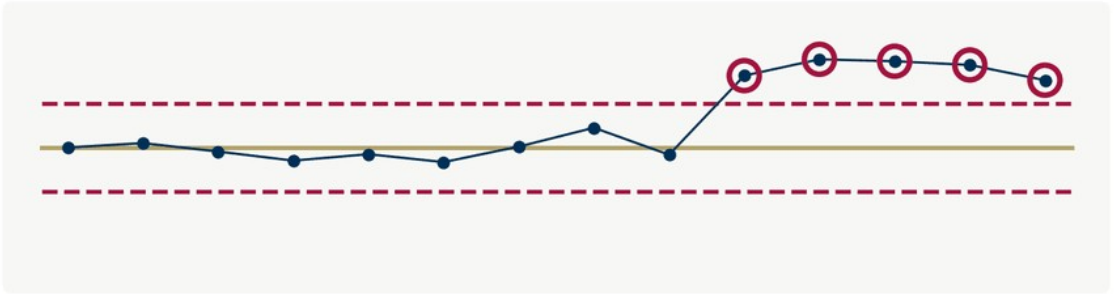
What is BETWEEN is what the chart detects.

Subgroup = 5 sites on one wafer (n = 5)



wide limits — the 0.55 nm shift at lot 10 is absorbed

Subgroup = 5 wafers from one lot (n = 5)



tight limits — the same shift is caught immediately

Same wafers. Same n. Same arithmetic. Different chart.

Georgia Tech

Building an $\bar{X}$–R chart

Six steps, and you will do this in the homework

Step	What you compute	Formula
1	Range of each subgroup	$R_i = \max - \min$
2	Average range	$\bar{R} = \Sigma R_i / k$
3	Grand average	$\bar{\bar{X}} = \Sigma \bar{X}_i / k$
4	R chart limits	$UCL = D_4\bar{R}$ $LCL = D_3\bar{R}$
5	Check the R chart is stable	if not, stop — step 6 is invalid
6	$\bar{X}$ chart limits	$\bar{\bar{X}} \pm A_2\bar{R}$

Notice step 5. The $\bar{X}$ limits are built from $\bar{R}$, so an unstable R chart makes them meaningless. This is why you always read the R chart first.

Georgia Tech

Worked example

Gate CD, 25 subgroups of 5 wafers, target 45.0 nm

The data

25 lots. Five wafers measured per lot, one site each — so within-subgroup variation is wafer-to-wafer within a lot.

Specification: 43.0 to 47.0 nm.

What we compute

$\bar{R} = 0.98$ nm → $\hat{\sigma} = \bar{R}/d_2 = 0.98 / 2.326 = 0.42$ nm

$\bar{\bar{X}} = 45.19$ nm

$A_2 = 0.577$, $D_3 = 0$, $D_4 = 2.114$

$\bar{X}$ limits: 45.19 ± 0.577 × 0.98 = 44.62 to 45.75 nm

R limits: 0 to 2.114 × 0.98 = 2.07 nm

Note how far inside the spec window those control limits sit. That gap is capability, and we will measure it at the end of today.

Georgia Tech

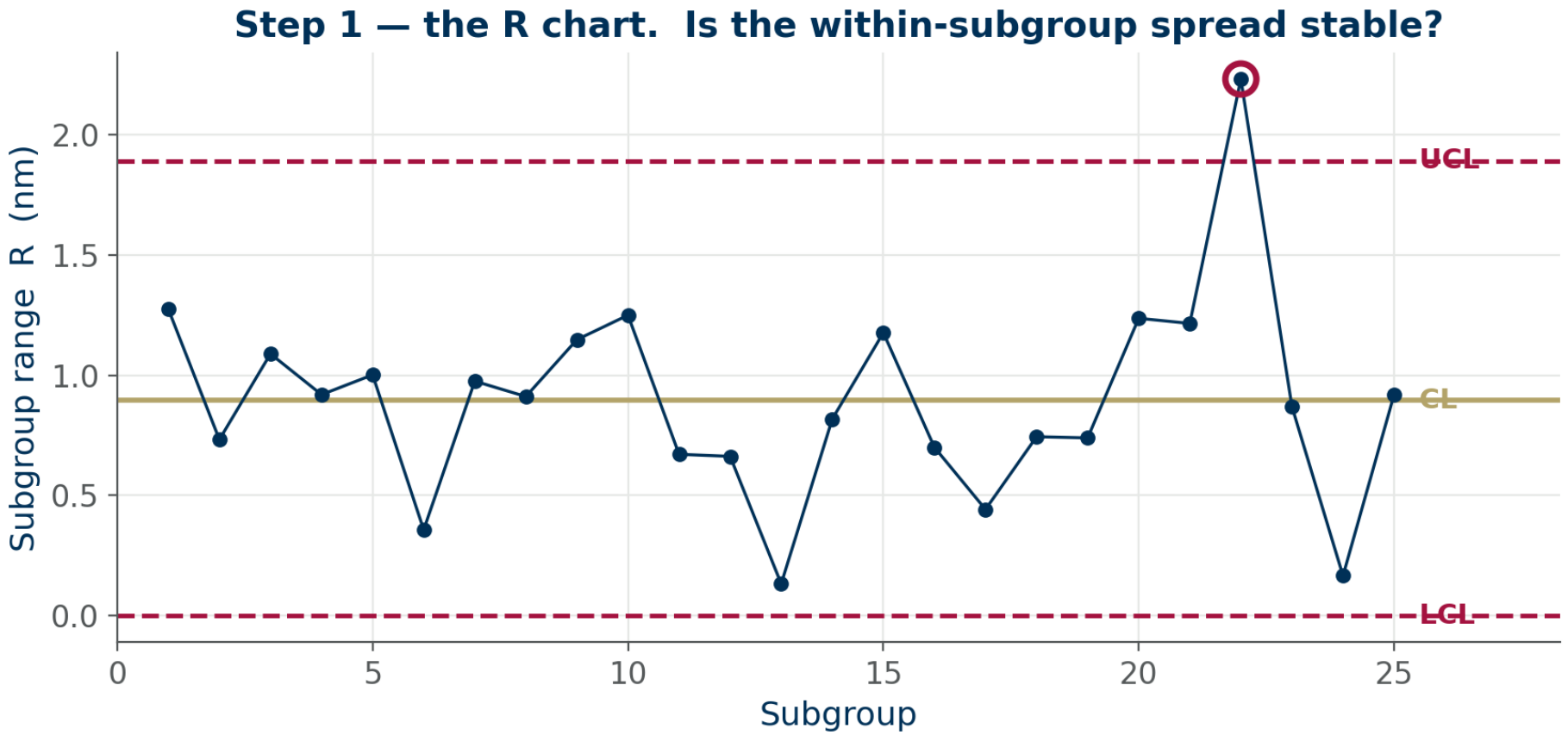
The first ten subgroups

Five wafers per lot, gate CD in nm

Lot	w1	w2	w3	w4	w5	$\bar{X}$	R
1	45.02	44.71	45.38	44.93	45.20	45.05	0.67
2	45.31	44.88	45.07	45.44	44.79	45.10	0.65
3	44.66	45.29	45.11	44.82	45.35	45.05	0.69
4	45.44	45.02	44.58	45.19	44.91	45.03	0.86
5	44.87	45.51	45.13	44.70	45.26	45.09	0.81
6	45.18	44.63	45.40	45.05	44.84	45.02	0.77
7	44.95	45.22	44.68	45.47	45.09	45.08	0.79
8	45.36	44.79	45.14	44.92	45.28	45.10	0.57
9	44.74	45.33	45.02	45.41	44.88	45.08	0.67
10	45.09	44.85	45.44	44.97	45.21	45.11	0.59

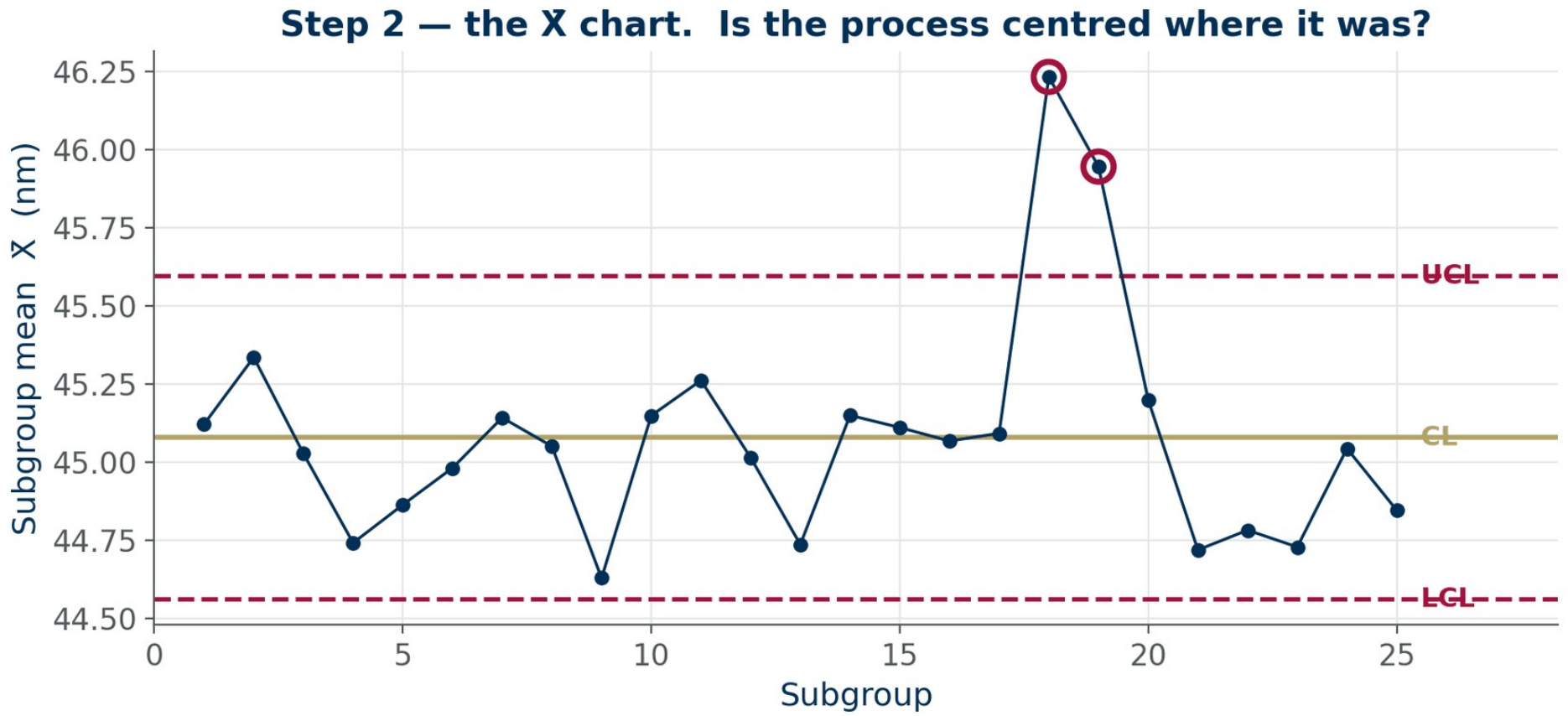
Two columns is all a Shewhart chart ever uses. $\bar{X}$ goes on one chart, R on the other, and the raw wafers are never plotted again.

Read the R chart first



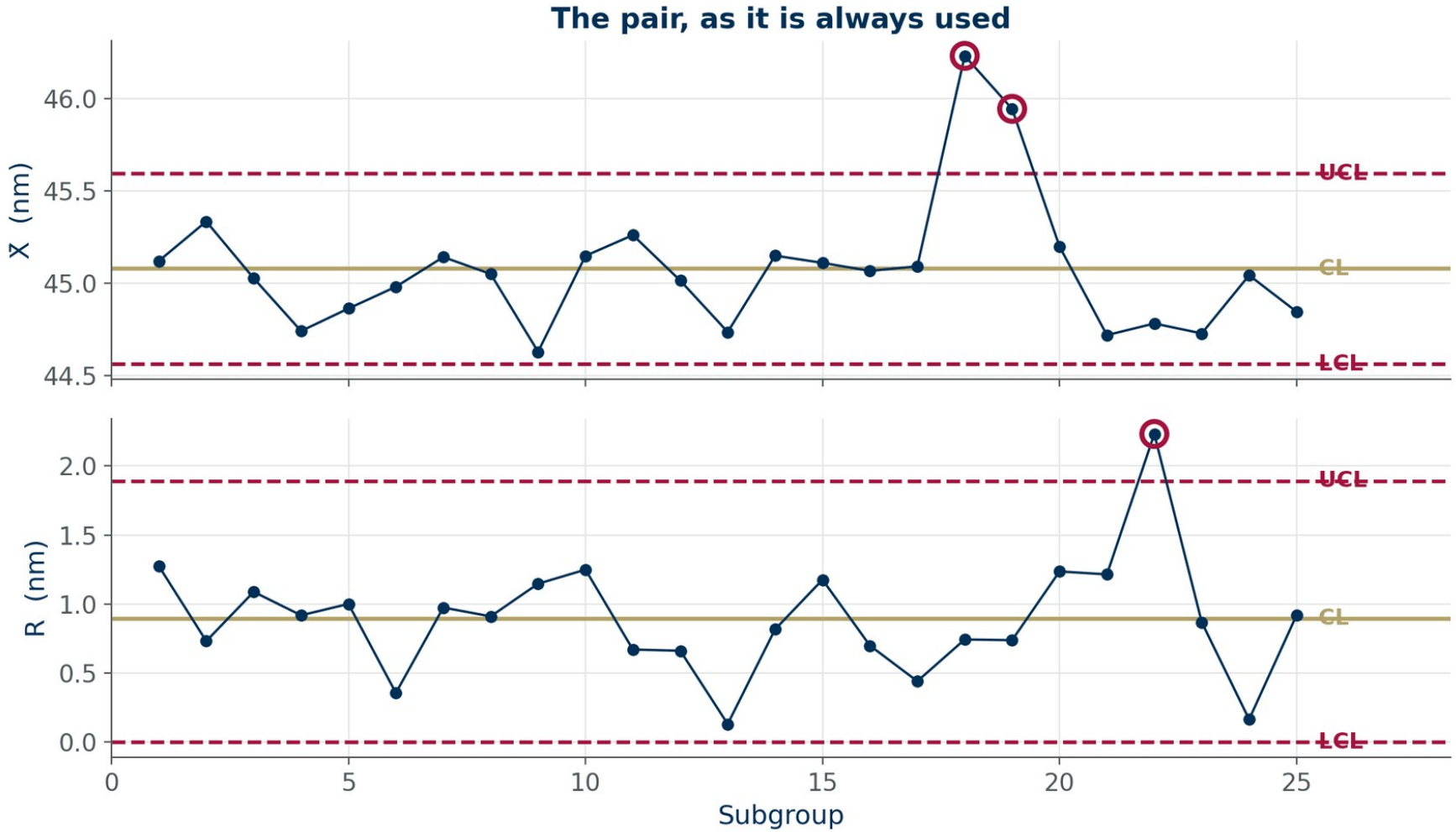
Subgroup 22 fails. Something happened to the within-lot spread — before we look at the mean at all.

Then the $\bar{X}$ chart



Subgroups 18 and 19 are above the upper limit. The process centre moved.

The pair, as it is always used



What each chart catches

The $\bar{X}$ chart — location

Has the process centre moved?

Catches: recipe drift, a wrong offset, chamber-to-chamber mismatch, a controller over-correcting, temperature drift.

Blind to: a change in spread, provided the mean stays put.

The R chart — spread

Has the within-subgroup variability changed?

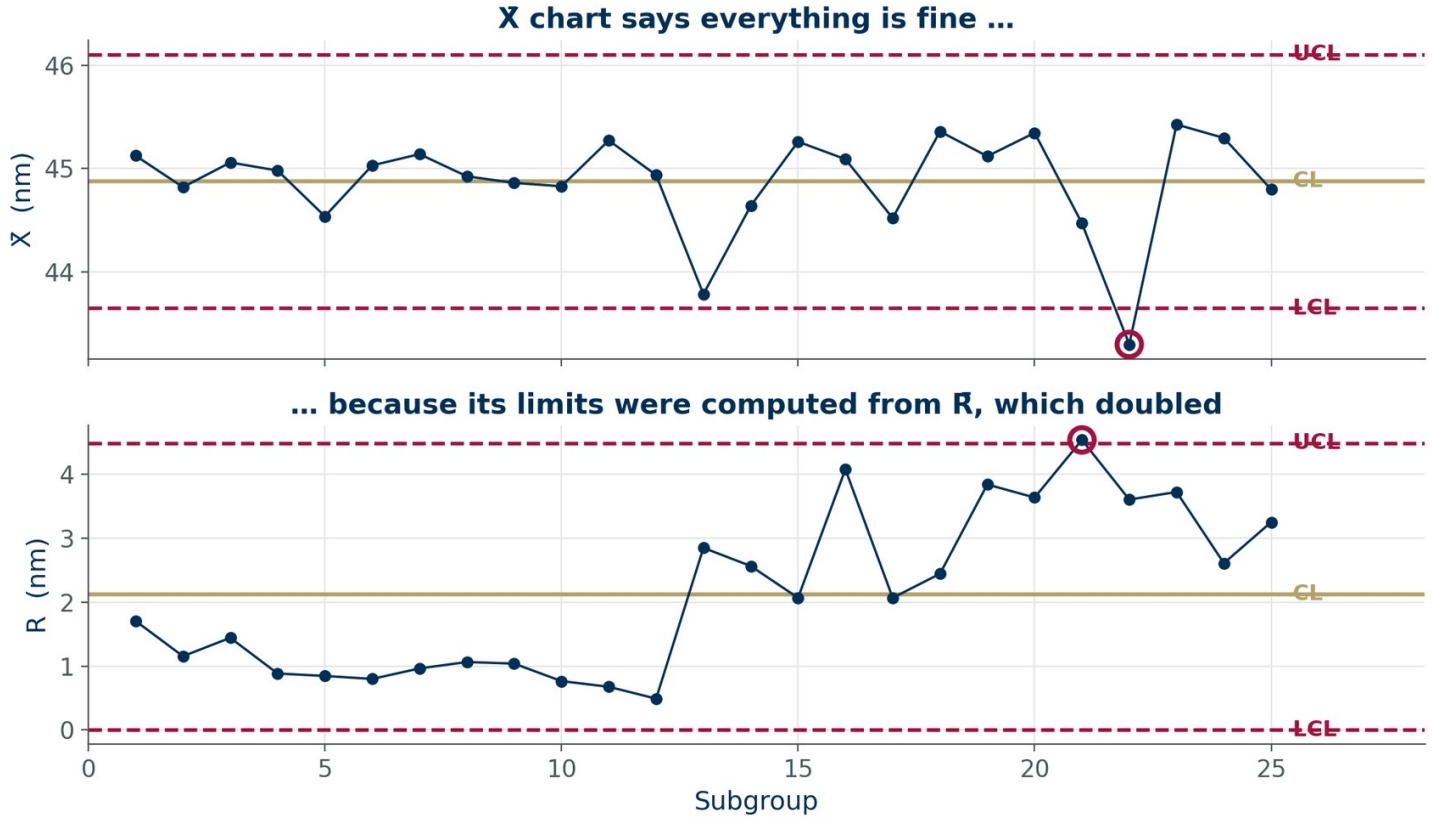
Catches: a degrading chamber, non-uniformity across the wafer, an unstable metrology tool, one bad wafer in a lot.

Often moves first. A tool frequently gets noisier before it shifts.

The R chart is the early warning. The $\bar{X}$ chart is the consequence.

Georgia Tech

What happens if you read them backwards



The spread more than tripled at subgroup 13 — and the $\bar{X}$ chart shows nothing, because its limits were computed from the inflated $\bar{R}$.

Georgia Tech

$\bar{X}$–S charts

When the subgroup gets big

Why the range works at all

With n ≤ 10 the range uses most of the information in the subgroup — it is about 95% as efficient as the standard deviation at n = 5.

And it can be computed in your head.

Why it stops working

As n grows, the range only ever uses two of the n values, so it throws away more and more. By n = 15 it is markedly worse.

Above about n = 10, switch to S: $\hat{\sigma} = \bar{S}/c_4$.

In a fab you will rarely need it — metrology cost keeps n small.

Where you do see large n is automated inline metrology, where measuring 49 sites costs no more than measuring 5.

Georgia Tech

Four ways to get an $\bar{X}$–R chart wrong

All four are common, and all four are quiet

Reading the $\bar{X}$ chart first

Its limits are built from $\bar{R}$. If the spread moved, the limits are wrong and the $\bar{X}$ chart is decorative.

Putting spec limits on the chart

Now every conversation is about spec, the run rules become invisible, and operators adjust to the wrong lines.

Recomputing limits when it alarms

You have deleted the signal. Limits change after a deliberate process change — never in response to data.

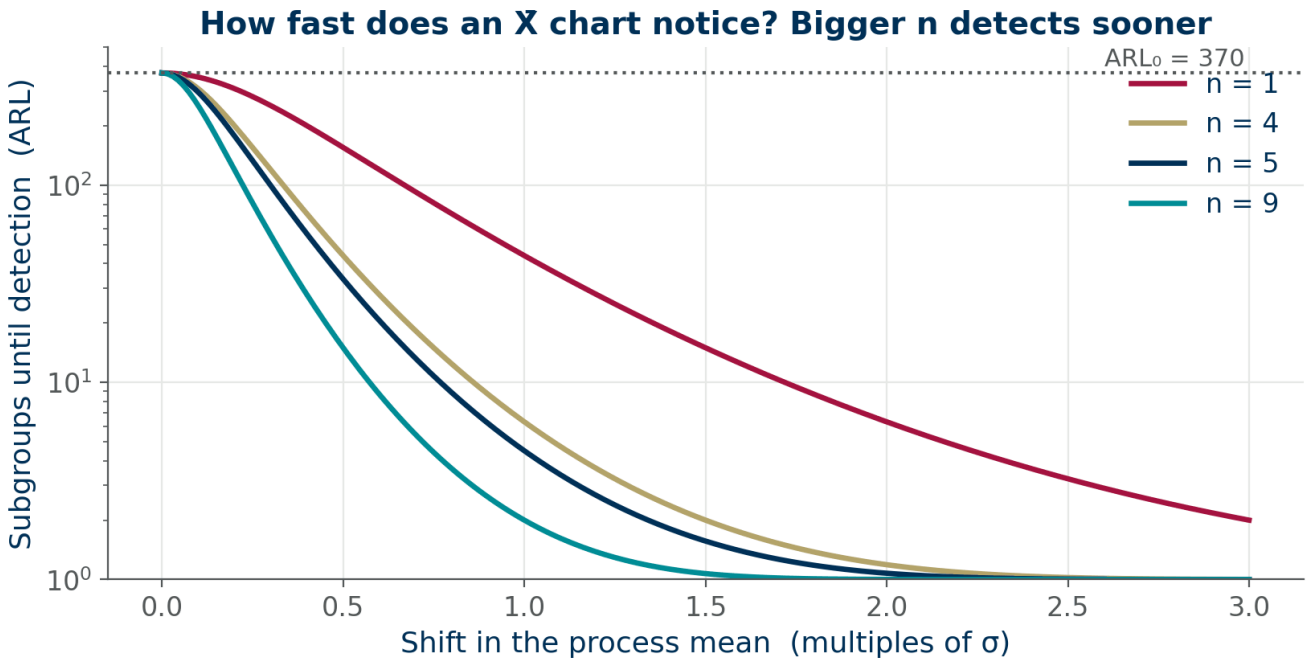
Subgrouping by convenience

Whatever the metrology recipe happened to measure is not a rational subgroup. Decide what you want the chart to be blind to.

Three of these four make the chart less sensitive without making it look broken.

Georgia Tech

How fast will it notice?



- Average run length: how many subgroups until the chart fires.
- **A 1σ shift with n = 5 takes about 5 subgroups to detect.**
- A 0.5σ shift takes over 30 — the Shewhart chart is genuinely poor at small sustained shifts, and this is its main weakness.
- Bigger subgroups detect sooner, which is the sensitivity half of the subgroup-size trade from Lecture 1.
- **Two fixes, both later:** run rules today, and the EWMA and CUSUM charts in Lecture 3.

Know the weakness: a single Shewhart point is nearly blind to shifts under 1σ.

Georgia Tech

When n = 1

Which, in a fab, is most of the time

Metrology is the constraint

A CD-SEM measurement takes minutes and the tool has a queue. Measuring five wafers per lot instead of one is a five-fold increase in the scarcest resource on the floor.

Some measurements are destructive

Cross-sections, SIMS profiles, some electrical tests. You get one number and the wafer is gone.

Some quantities are one per lot

Bath chemistry. Chamber pressure at a setpoint. There is no subgroup to form — the quantity itself is singular.

No subgroup means no $\bar{R}$, and no $\bar{R}$ means we need another way to estimate σ.

Georgia Tech

The moving range trick

$$MR_i = | x_i - x_{i-1} | \qquad \hat{\sigma} = \bar{MR} / d_2 = \bar{MR} / 1.128$$

The idea

Treat each consecutive pair as a subgroup of two. The difference between neighbouring lots is a short-term estimate of variation — it has no time to absorb a drift.

d_2 for n = 2 is 1.128.

The assumption hiding inside it

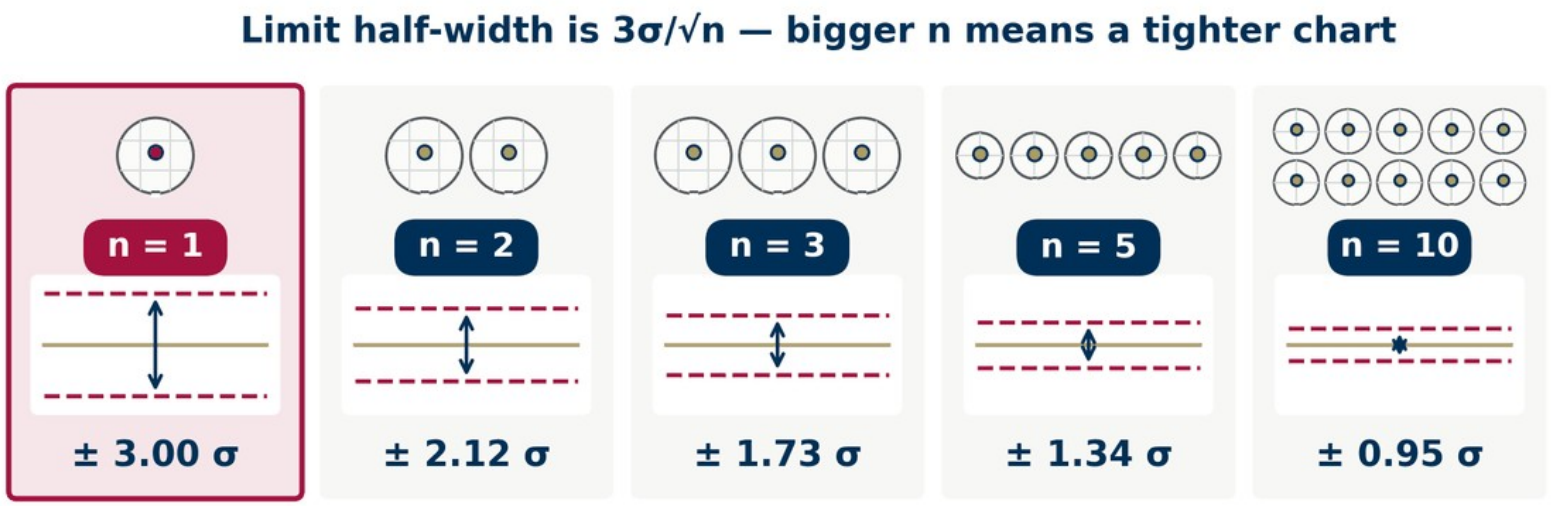
That consecutive lots are independent.

If they are correlated — and in a fab they very often are — the moving range is too small, $\hat{\sigma}$ is too small, and the limits are far too tight. We will see this shortly.

Georgia Tech

So what does n actually buy you?

Bigger n, tighter chart — and the case at the far left is the one you will meet most often



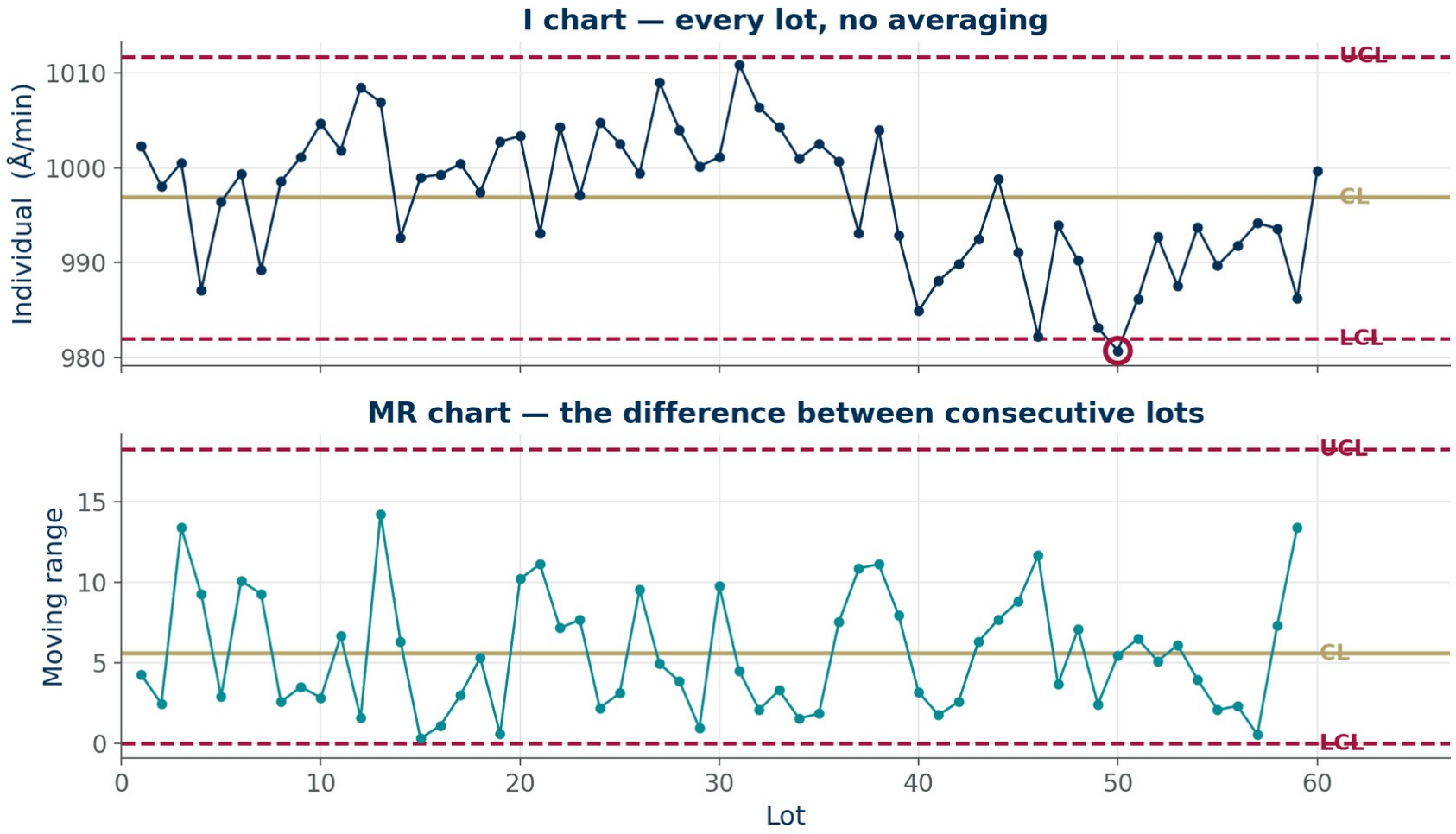
n = 1 is not just a small subgroup — it is no subgroup at all

One measurement has no range. No $\bar{R}$, and no $\hat{\sigma}$ from within a subgroup — so the moving range between consecutive lots stands in: $\hat{\sigma} = \overline{MR} / 1.128$

And it is the fab default: metrology is expensive, sometimes destructive, and some quantities are one per lot. That is why the I–MR chart exists.

Georgia Tech

An I–MR chart



Etch rate, one measurement per lot. The shift at lot 39 is real — and only one point actually breaks a limit.

I–MR worked numbers

Etch rate, 60 lots, one measurement each

From the data

$\bar{x}$ = 996.1 Å/min

$M\bar{R}$ = 5.24 Å/min

$\hat{\sigma} = M\bar{R} / 1.128 = 4.65$ Å/min

The limits

I chart: 996.1 ± 3 × 4.65

= 982.2 to 1010.1

MR chart: 0 to 3.267 × 5.24 = 17.1

Compare with a subgrouped chart on the same process.

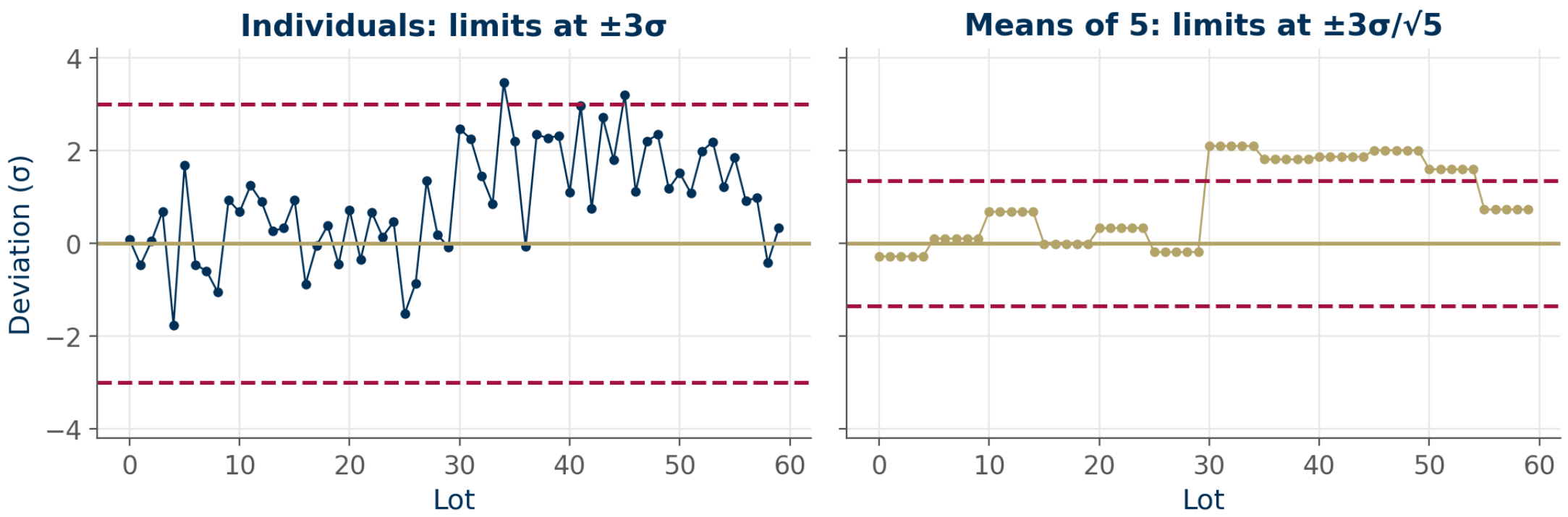
With n = 5 the $\bar{X}$ limits would sit at $\pm 3\sigma/\sqrt{5}$ — about ±6.2 instead of ±14. The individuals chart is more than twice as wide, and that is the whole cost of n = 1.

Same process, same σ. The chart you can afford is the chart that detects least.

Georgia Tech

The price you pay

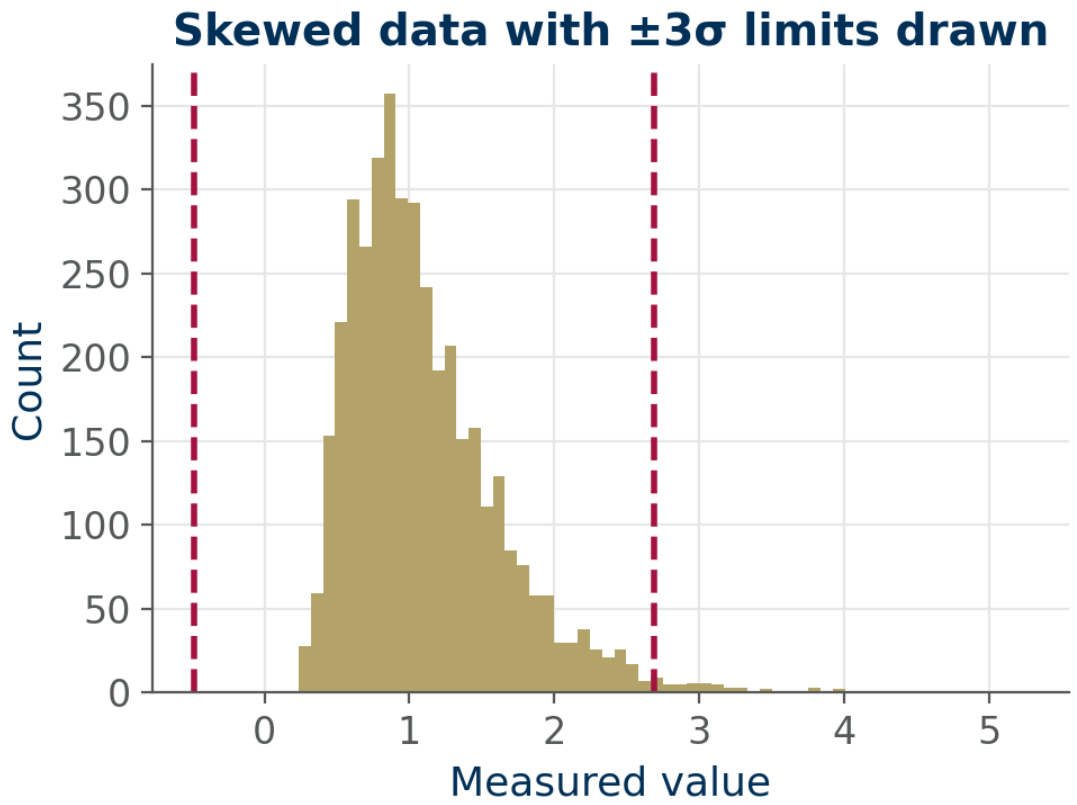
Same 1.5σ shift. The individuals chart barely reacts.

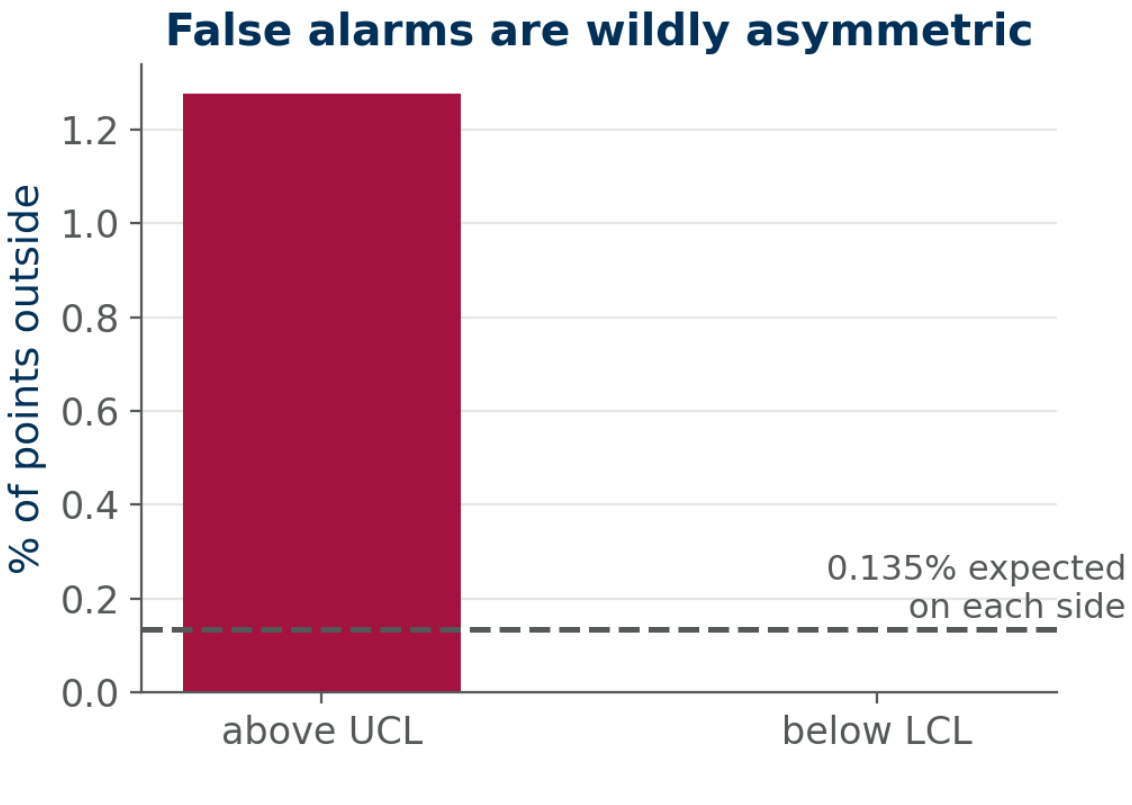


A 1.5σ shift. Obvious on the subgrouped chart, nearly invisible on the individuals chart — the limits are √5 times wider.

Georgia Tech

And the assumption you inherit





Skewed data, ±3σ limits. Almost all the false alarms land on one side.

Georgia Tech

What to do about non-normality

Three legitimate options, one wrong one

Transform

Log, square-root, or Box–Cox the data, chart the transformed value, and convert the limits back for reporting.

Works well for skewed positive quantities like particle counts.

Fit the right distribution

Compute limits from the actual percentiles of a fitted distribution instead of assuming ±3σ.

Honest, but you need enough data to fit the tails.

Widen deliberately

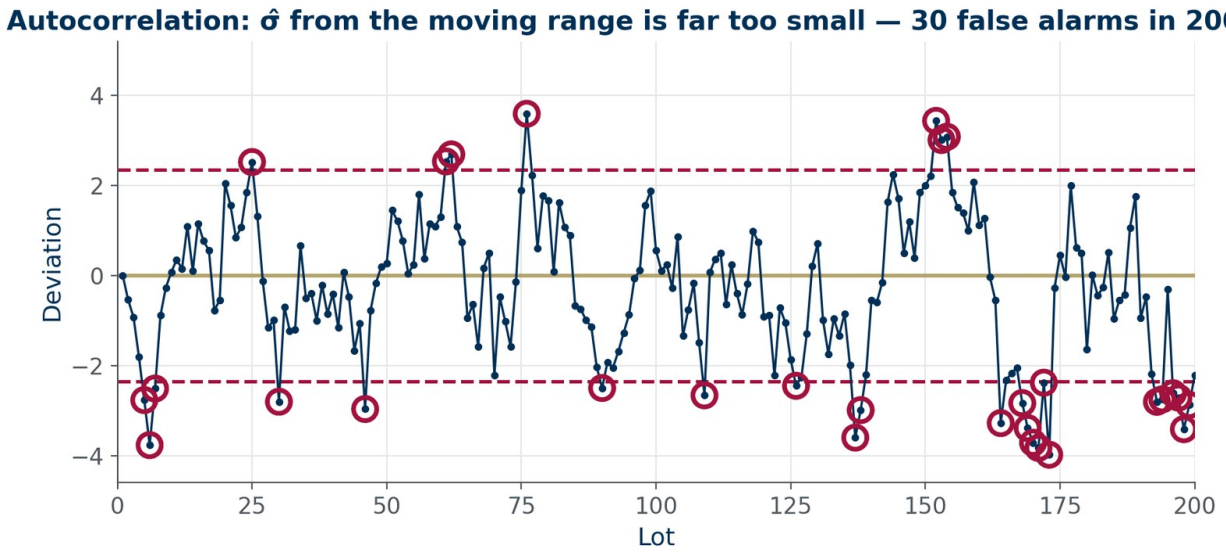
Keep the simple chart and set the limits at a stated false-alarm rate instead of at 3σ.

Crude, but transparent — and everybody still understands the chart.

The wrong option is to apply ±3σ to visibly skewed data and not mention it.

Georgia Tech

The fab's real problem with I–MR



- Consecutive lots through the same chamber are not independent.
- Chamber state carries over between runs. A run-to-run controller creates correlation deliberately. Bath chemistry evolves slowly.
- **So the moving range understates σ,** the limits come out far too tight, and the chart alarms constantly on a process that is behaving exactly as designed.
- Here φ = 0.8 and the chart fires dozens of times on a stationary process.
- **The fix is not wider limits.** It is to model the correlation and chart the residuals — Lecture 3.

Before you trust an I chart, plot the data against itself one lag back.

Georgia Tech

I–MR checklist

Run this before you trust the chart

1 Is the data roughly symmetric?

Normal probability plot. Skew makes the false alarm rate asymmetric.

2 Are consecutive points independent?

Plot x_i against x_{i-1}. Structure means the moving range is lying to you.

3 Is there a known deterministic pattern?

Seasoning and PM sawtooth is not a special cause. Chart the residual from the expected cycle, not the raw value.

4 Did you compute limits from a clean baseline?

Phase I discipline matters more here, because with n = 1 a single excursion moves $\hat{\sigma}$ a long way.

Georgia Tech

Attribute charts

When all you have is pass or fail

Why a different chart

For continuous data we estimated the spread from the data. For counts we do not have to — the distribution supplies it.

If each die fails independently with probability p, the count of failures is binomial, and its standard deviation is determined by p and n. Nothing to estimate.

Which is a blessing and a trap

Blessing: the limits come free from the model.

Trap: if the model is wrong — if p varies from lot to lot — the limits are too tight and the chart alarms constantly. That failure has a name, and we will get to it.

With variables data you measure the noise. With attribute data you assume it.

The chart family

Pick by what you are counting

Chart	You are counting	Constant n?	Limits
p	Fraction of units defective	not required	$\bar{p} \pm 3\sqrt{(\bar{p}(1-\bar{p})/n_i)}$
np	Number of units defective	required	$n\bar{p} \pm 3\sqrt{(n\bar{p}(1-\bar{p}))}$
c	Defects per unit	required	$\bar{c} \pm 3\sqrt{\bar{c}}$
u	Defects per unit area	not required	$\bar{u} \pm 3\sqrt{(\bar{u}/n_i)}$

p and np — binomial

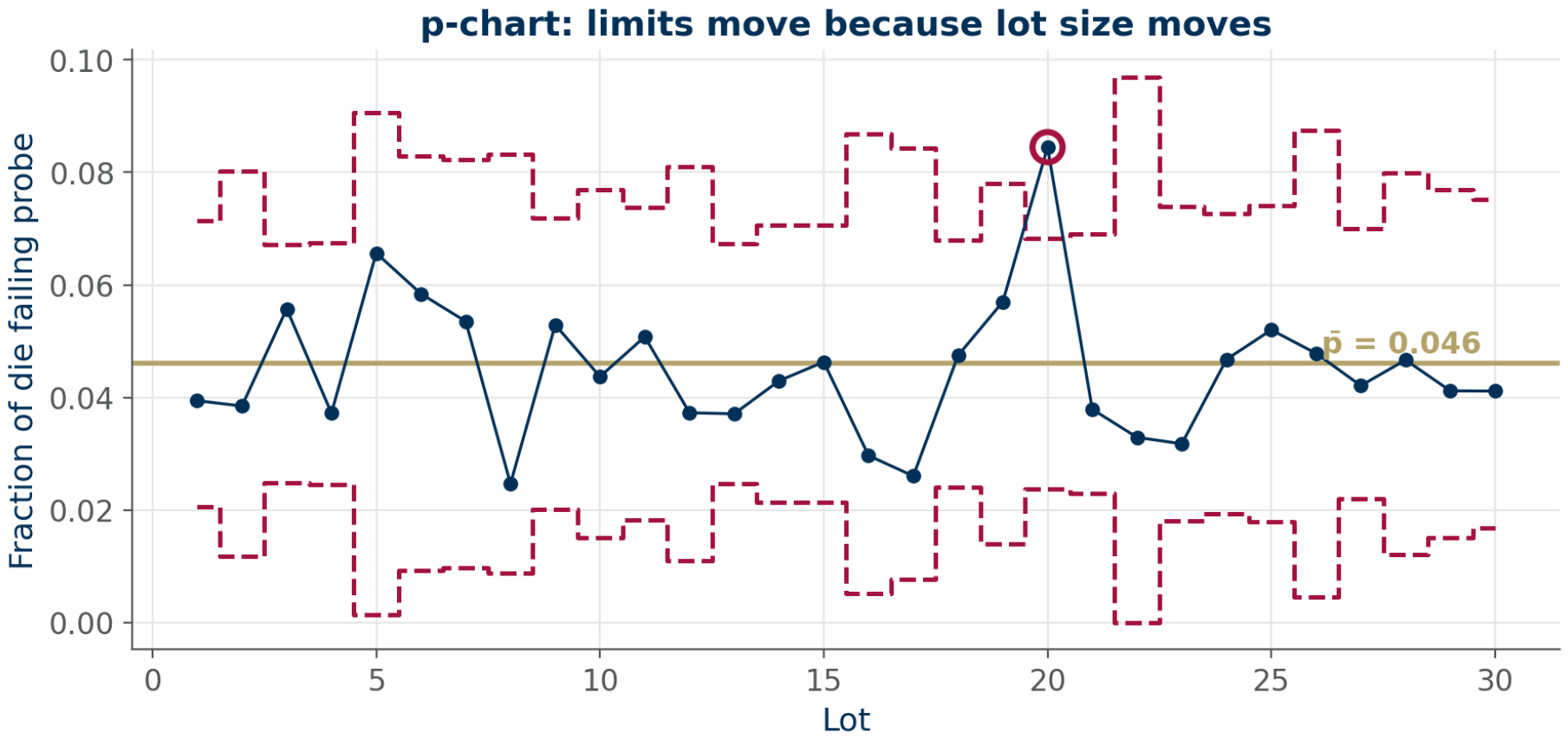
A unit either passes or fails. One die, one verdict. Probe yield is the classic case.

c and u — Poisson

A unit can carry several defects. Particles on a wafer, or visual defects per reticle field.

Georgia Tech

A p-chart with variable lot size



The limits move because the limits depend on n. A big lot gets a tighter test than a small one — as it should.

Georgia Tech

p-chart worked numbers

Probe yield, 30 lots of varying size

The centre line

Pool everything — do not average the lot fractions.

$\bar{p}$ = total die failed / total die tested

= 1 384 / 30 100 = 0.046

Averaging $\hat{p}_i$ instead would over-weight the small lots.

The limits — one pair per lot

$UCL_i = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p}) / n_i}$

n = 200 → 0.046 ± 0.044

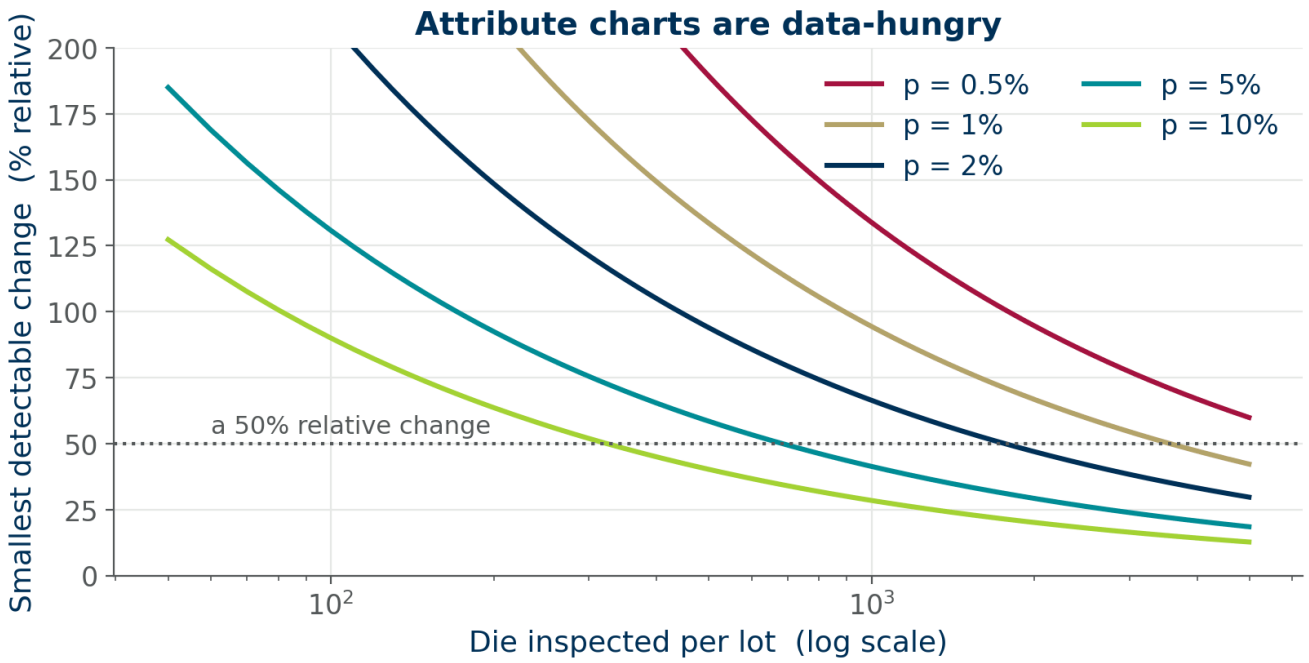
n = 800 → 0.046 ± 0.022

The big lot gets the tighter test, because you know more about it.

When the lower limit computes to a negative number, it is set to zero — you cannot have a negative fraction defective.

Georgia Tech

Attribute charts are data-hungry

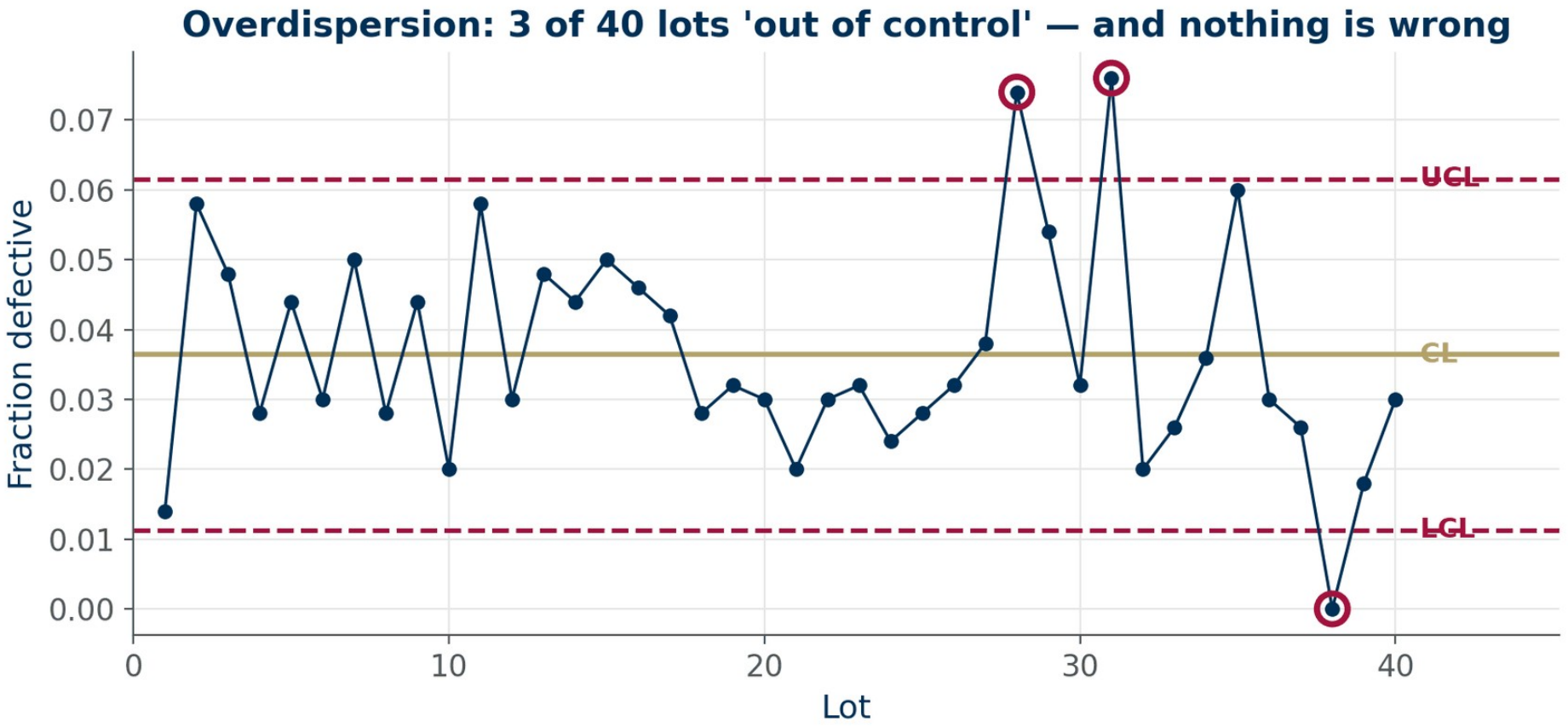


- To see a change, the change must exceed 3 standard errors — and for a proportion that means $3\sqrt{(p(1-p)/n)}$.
- **At p = 1% and 500 die per lot,** you cannot reliably detect anything smaller than a doubling.
- Rare events are the worst case: the rarer the defect, the more units you need to see any change at all.
- This is the quantitative reason fabs chart continuous parameters wherever they can and fall back to yield only when they must.
- **It is also why yield alone is a terrible process monitor —** by the time yield moves, the parametric charts moved weeks ago.

Yield is the outcome you care about and the worst signal to steer by.

Georgia Tech

Overdispersion



Forty lots from a healthy line. Three 'out of control' — because p genuinely varies from lot to lot and the binomial model does not allow for it.

Georgia Tech

Fixing an over-alarming p-chart

Test the assumption first

Compare the observed scatter of $\hat{p}$ to what the binomial predicts. If the ratio is much above one, you have overdispersion — not a process problem.

Then either stratify …

Often the extra variation is real structure you pooled away: two products, two probe testers, two chambers. Chart them separately and the overdispersion disappears.

… or widen the model

If the variation is genuinely lot-to-lot, use limits based on the observed between-lot scatter — a Laney p′ chart — rather than the binomial.

Always try stratifying before you widen. A chart that alarms may be telling you your population is not one population.

Georgia Tech

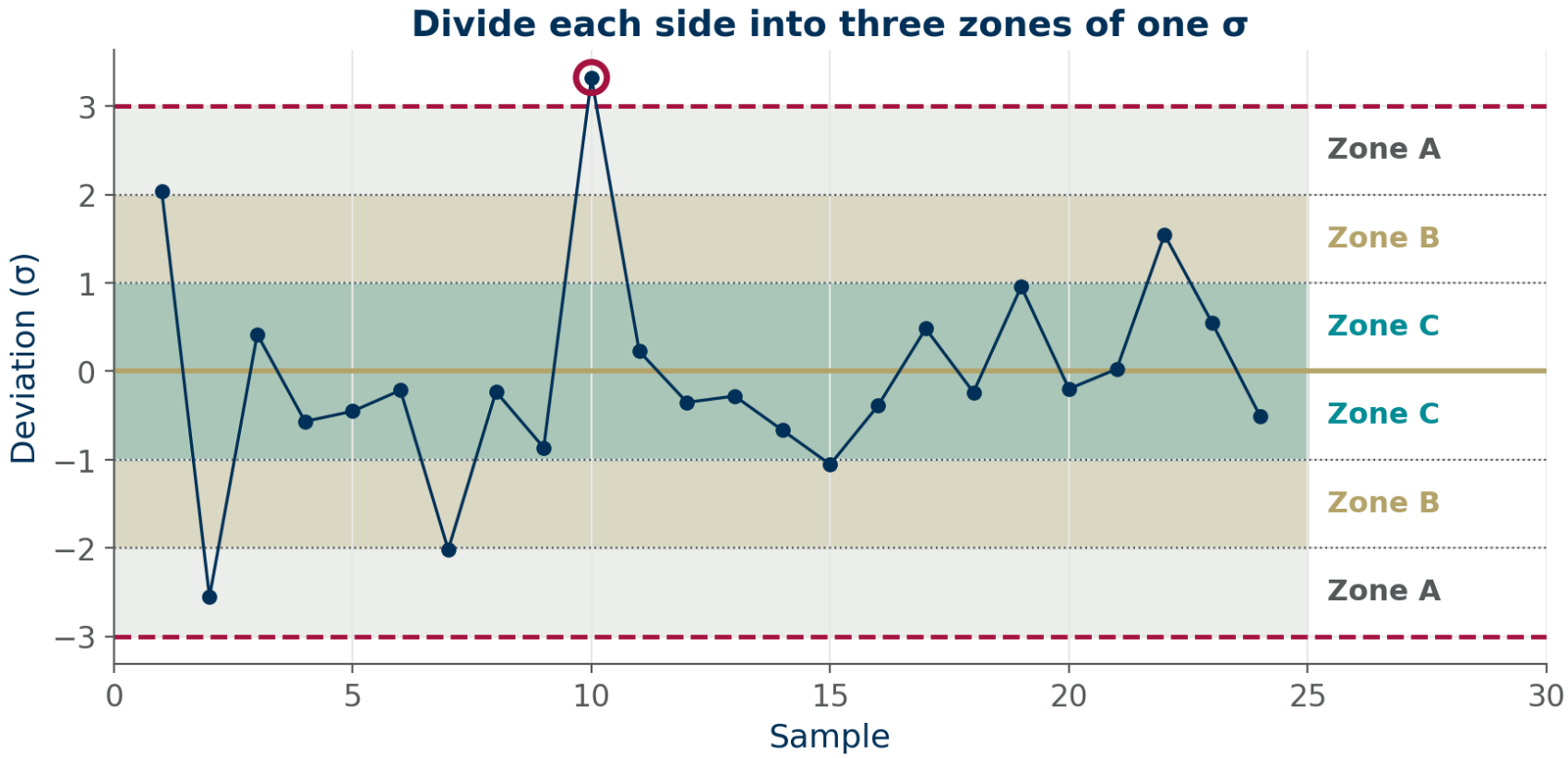
RUN RULES

The points inside the limits are talking to you.

A chart that only checks whether points fall outside the limits is using a fraction of the information in front of it.

Georgia Tech

Divide each half into three zones



Zone C is within 1σ, Zone B between 1σ and 2σ, Zone A between 2σ and 3σ. Every run rule is a statement about zones.

Georgia Tech

The Western Electric rules

Published 1956, still the default today

#	The pattern	Chance of it happening anyway	What it usually means
1	1 point beyond Zone A (>3σ)	1 in 370	A discrete event — recipe, material, hardware
2	2 of 3 consecutive in Zone A	1 in 1000	A large shift, caught early
3	4 of 5 consecutive in Zone B or beyond	1 in 400	A moderate sustained shift
4	8 or 9 in a row on one side	1 in 256	A small sustained shift
5	6 in a row steadily increasing or decreasing	1 in 700	Drift — wear, erosion, depletion
6	14 in a row alternating up and down	1 in 4000	Two populations, or over-adjustment
7	15 in a row inside Zone C	1 in 300	Stratification, or edited data
8	8 in a row outside Zone C on both sides	1 in 500	Mixture of two distributions

Georgia Tech

From pattern to physical cause

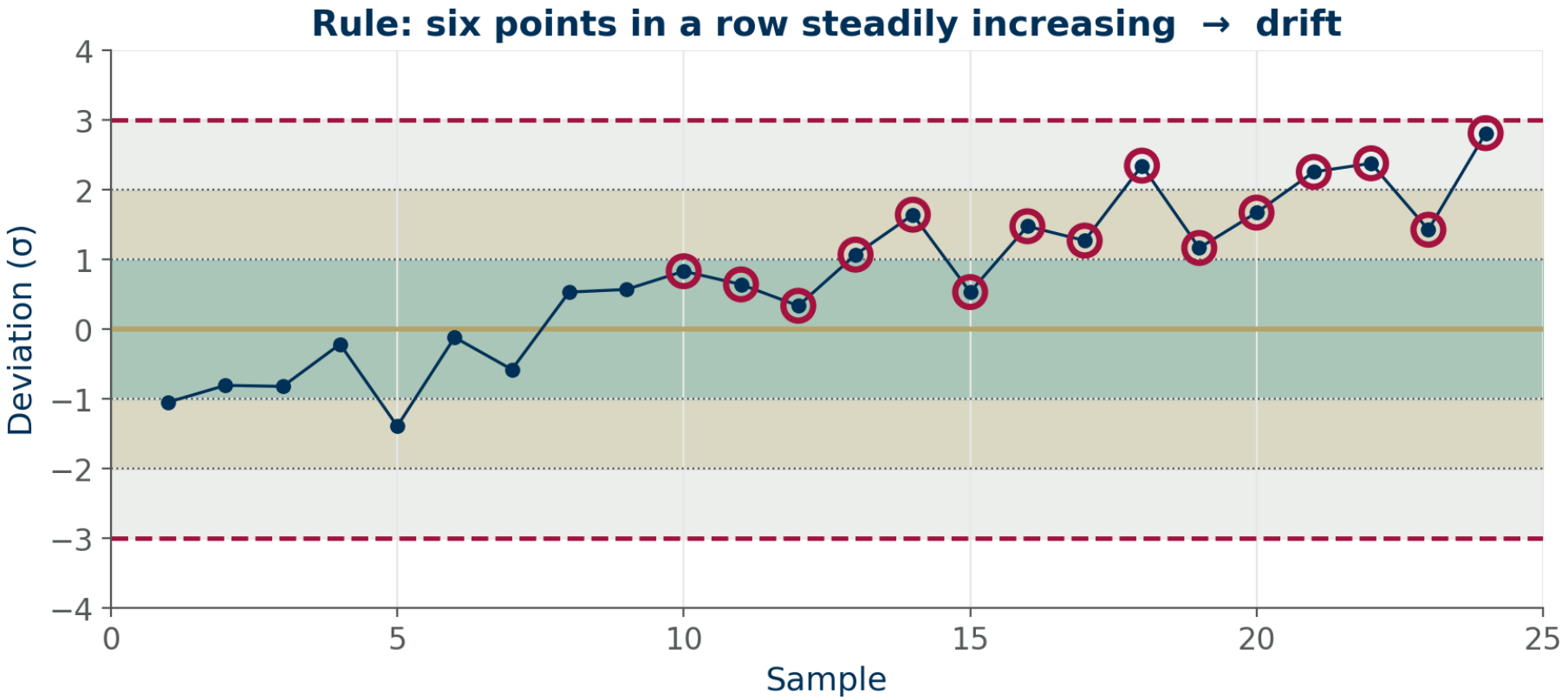
The diagnostic value is here, not in the alarm itself

Pattern on the chart	What to go and look at
Single wild point	Recipe log, material lot, operator actions on that shift
Sustained shift, step-like	Chamber swap, recipe edit, PM event, new consumable
Steady trend	Anything that wears: pad, target, filament, wall coating, filter
Sawtooth with resets	Seasoning cycle — probably not a fault at all
Alternating up and down	Two chambers, two testers, two shifts — or an over-tuned controller
Hugging the centre line	Data problem: over-stratified subgroups, edited numbers, or a stale sensor

The chart never tells you the cause. It tells you which log to open and roughly when.

Georgia Tech

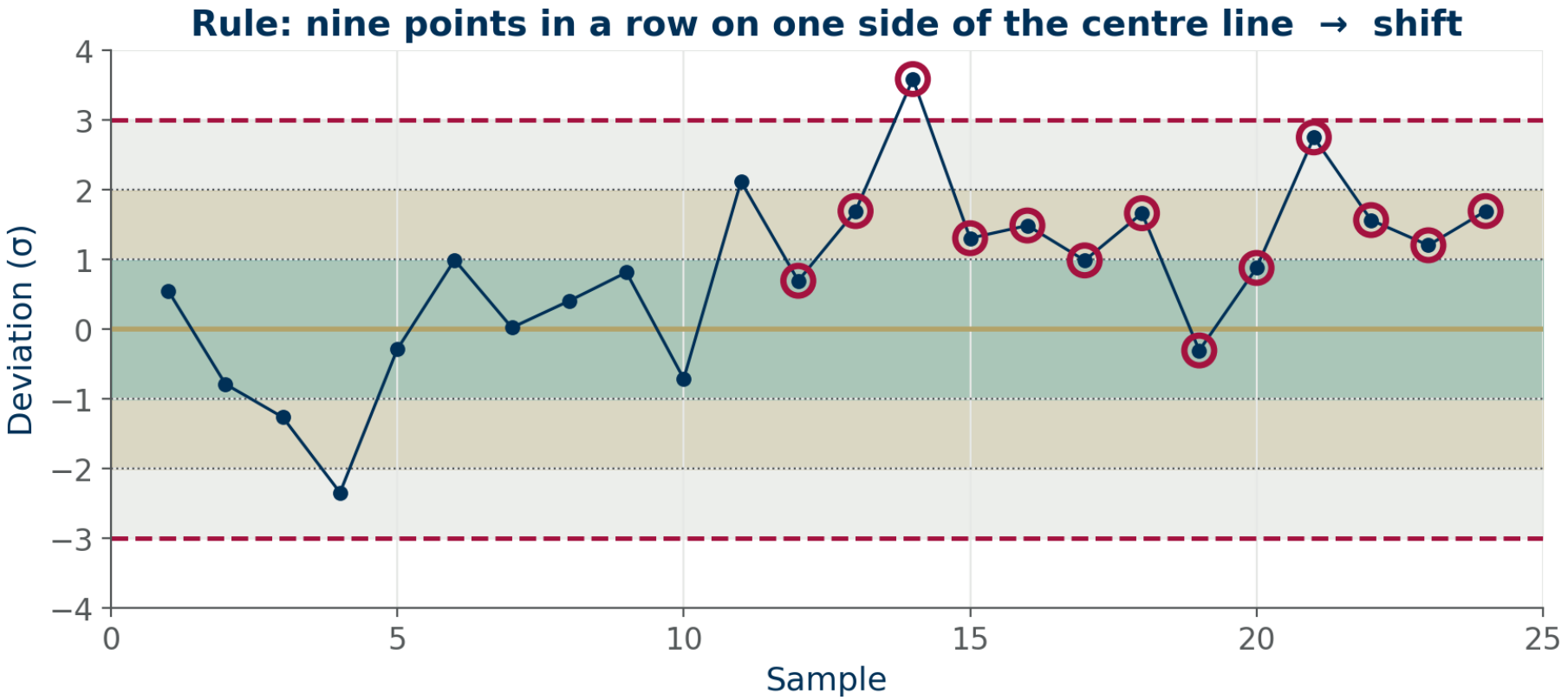
Rule 5 — a trend



Six points climbing in a row. Nothing has crossed a limit yet, and you already know something is wearing out.

Georgia Tech

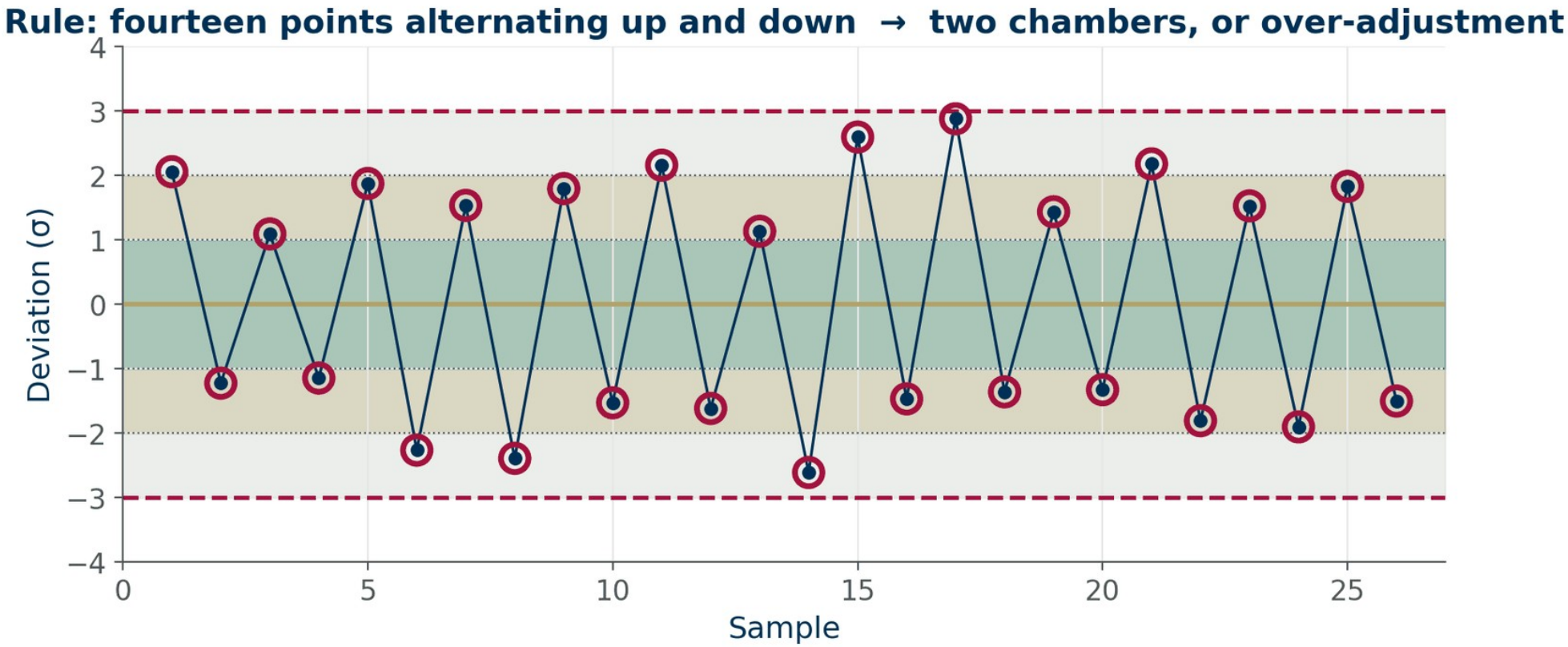
Rule 4 — a run on one side



This is how you catch the 1.5σ shift the individuals chart could not see.

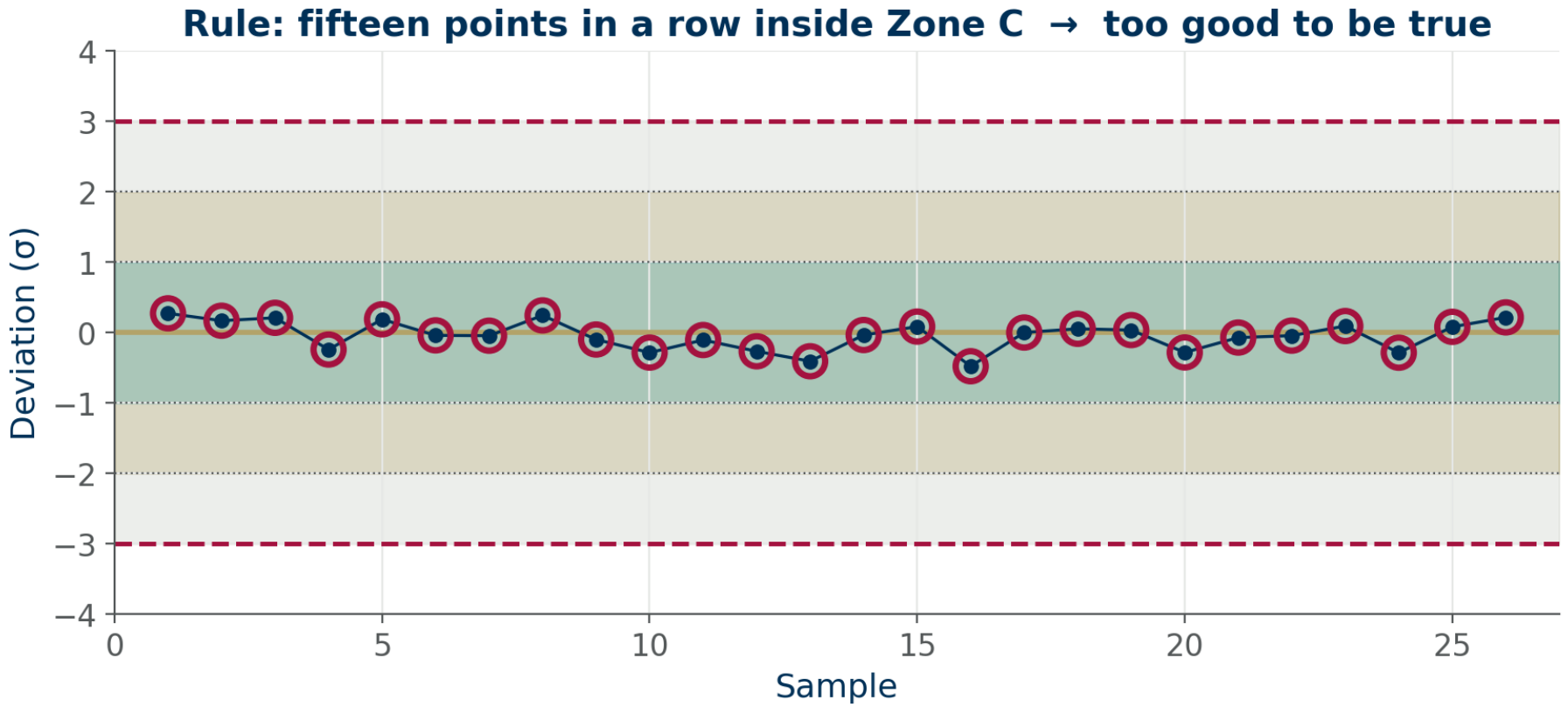
Georgia Tech

Rule 6 — alternating



Sawtooth from point to point almost always means two populations interleaved — or a controller correcting every lot.

Rule 7 — hugging the centre line



Too good to be true, and usually is. Real processes do not behave this well.

Why 'too good' is a problem, not a triumph

Over-stratified subgroups

If your subgroup deliberately spans every source of variation, $\bar{R}$ is huge, the limits are huge, and nothing can ever escape them.

The data has been edited

Someone is rounding, re-measuring failures, or reporting the median of a few tries. It happens, and the chart is how you find out.

The sensor has stopped

A frozen reading, a stale cached value, or a gauge that quietly failed to a nominal number. Very common with automated data collection.

A chart that never alarms is not evidence of a good process. It is usually evidence of a broken chart.

Georgia Tech

What an out-of-control signal is not

Not a defect

The lot may be entirely within specification and perfectly saleable. The chart is telling you the process changed, not that the product is bad.

Not proof

An in-control process throws a point past 3σ about three times in a thousand. Roughly one alarm in every few hundred is genuinely nothing.

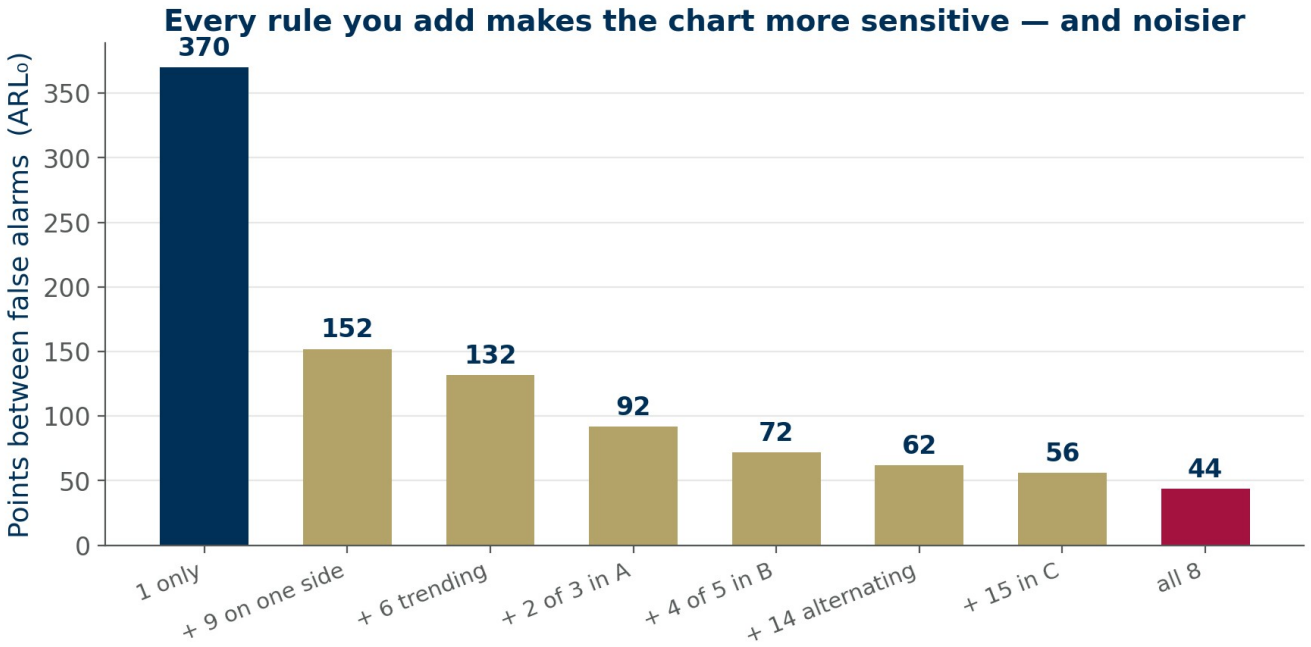
Not a diagnosis

The chart identifies a moment and a pattern. It has no idea what physically happened. That is your job, and the pattern is the hint.

It is a prompt to go and look, with a timestamp attached.

Georgia Tech

Every rule you add has a price



- Each rule is another chance to be wrong. They compound.
- **Rule 1 alone: one false alarm per 370 points.**
- **All eight rules: one per 44.**
- Chart daily with all eight and you are investigating something every six weeks that is not there — per chart. Multiply by the number of charts in a fab.
- You are buying sensitivity to small shifts with false alarms. That is the same trade as the 3σ decision, just spent differently.

Turn on the rules that map to failure modes you actually have. Not all eight, by default, everywhere.

Georgia Tech

So which rules do you turn on?

A defensible default

Almost always

Rule 1 — a point beyond 3σ. Non-negotiable.

Rule 4 — a run on one side. This is what catches small sustained shifts, which is exactly where Rule 1 is weakest. The two are complementary and together cost you roughly one false alarm per 150 points.

Add deliberately, per process

Rule 5 (trend) where you have a known wear mechanism — CMP, sputter targets.

Rule 6 (alternating) where you interleave chambers or testers.

Rule 7 (hugging) as a data-integrity check on any automated feed.

Choose rules from the failure modes of that specific process — not from the textbook's list.

Georgia Tech

The alarm is worthless without an OCAP

Out-of-Control Action Plan

Who does what, now

Named owner. What to hold, what to keep running, who to call. Decided in advance, in writing, before anyone is under pressure.

What to check, in order

A decision tree for that specific signal on that specific tool. Rule 1 on an etch chamber is a different investigation from Rule 5.

What gets recorded

The signal, the finding, the action, the outcome. This record is what turns a chart into a learning system rather than an alarm bell.

A signal with no defined response trains everybody to ignore signals.

Georgia Tech

Now — and only now — the second question

Is stable good enough?

Capability compares what the process naturally does to what the customer will accept.

Why stability comes first

An unstable process has no single distribution. Its mean and spread are moving. Any capability number you compute describes a mixture that will not exist tomorrow.

You would be forecasting from a process with no forecast.

What capability is not

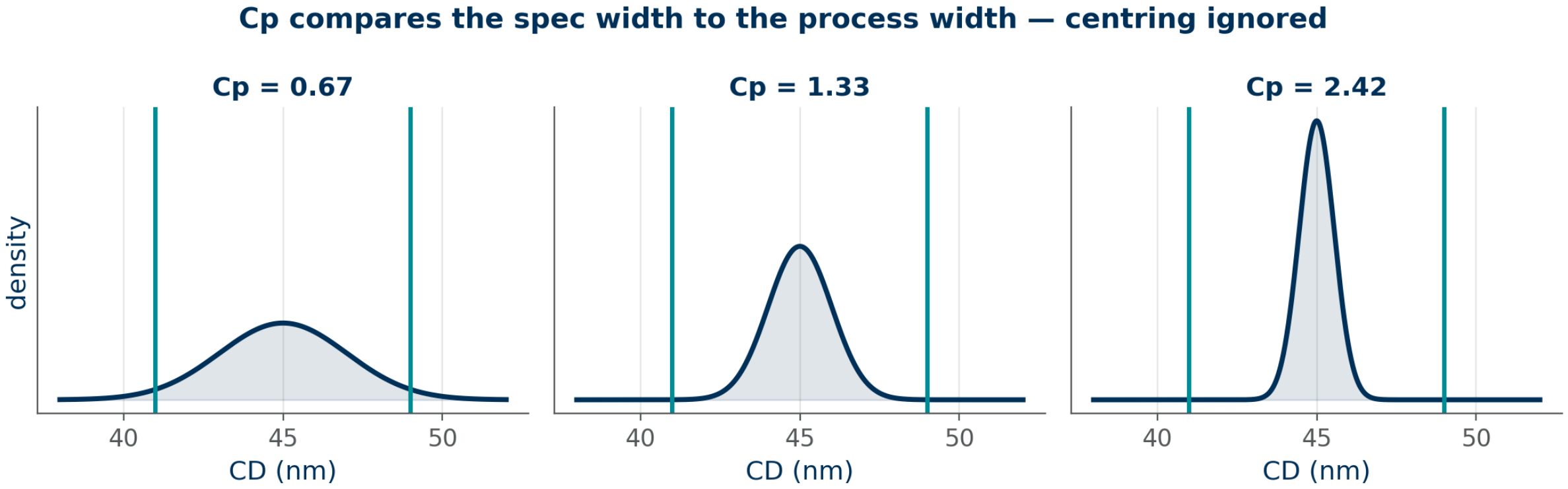
It is not a measure of how good your product is — that is yield.

It is not a statement about this lot. It is a statement about the process's ability to keep producing lots like this one.

Chart first. Stabilise. Then measure capability.

Georgia Tech

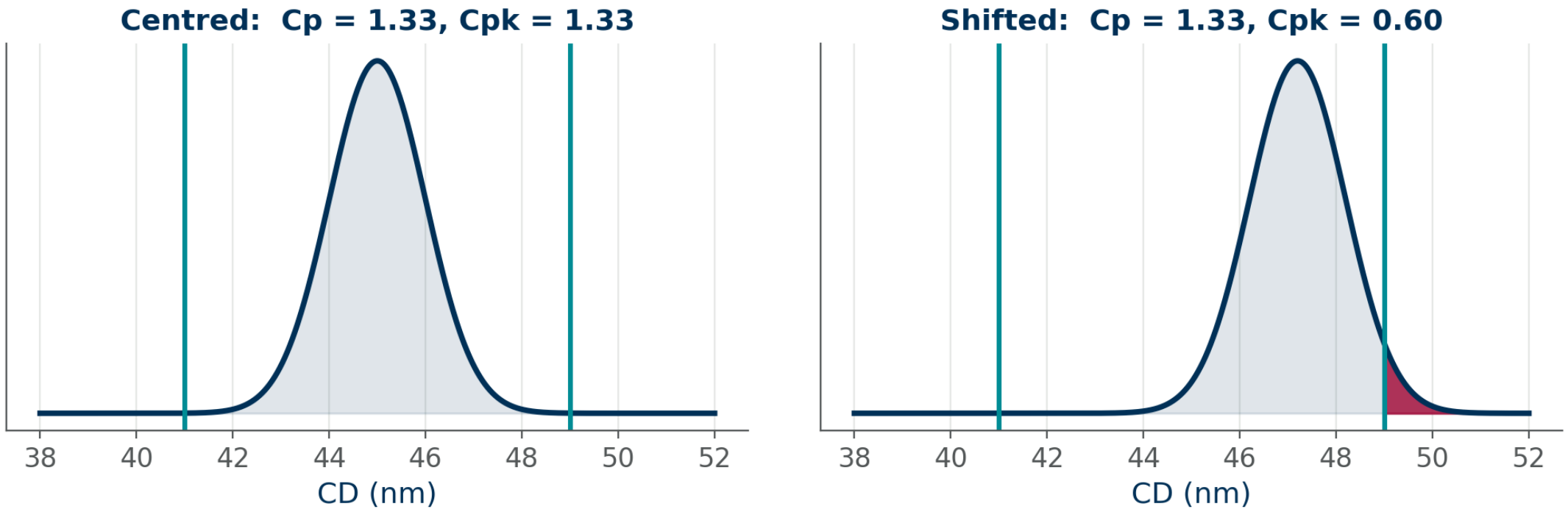
Cp — is the process narrow enough?



Cp = (USL − LSL) / 6σ. Spec width divided by process width. Centring is not considered at all.

Cpk — and is it in the right place?

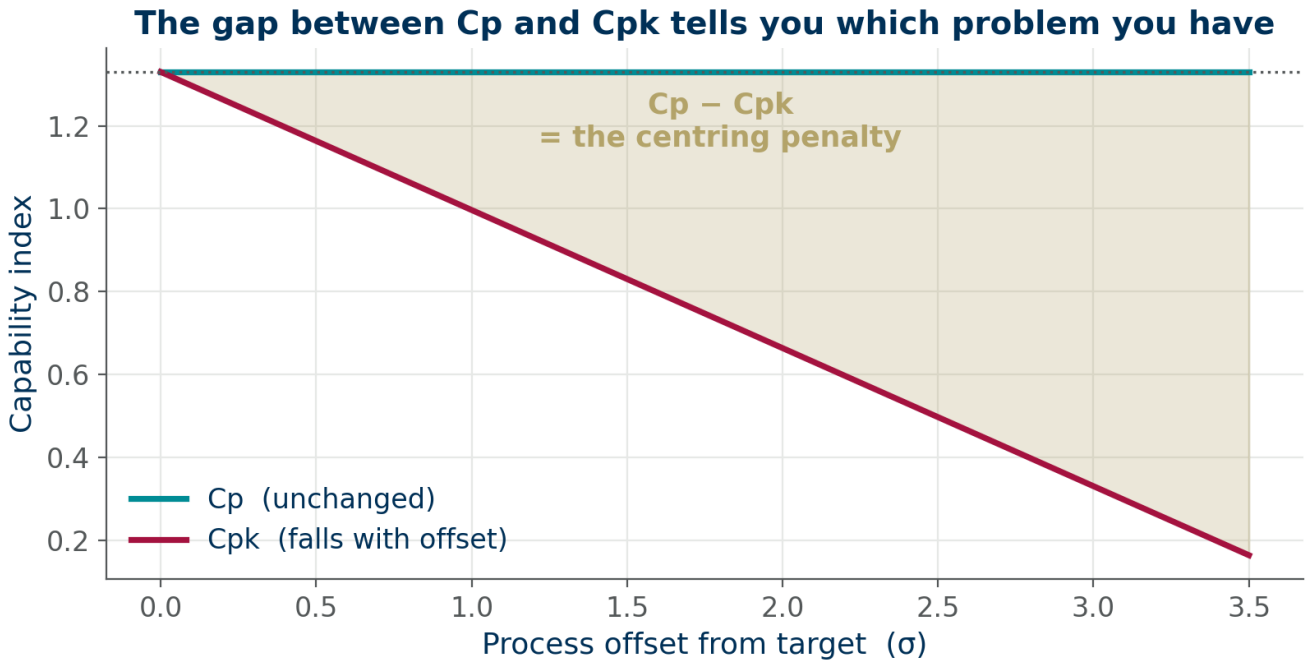
Same spread. Same Cp. Completely different defect rate.



Cpk uses the distance to the nearer spec limit: min(USL − x̄, x̄ − LSL) / 3σ.

Georgia Tech

The gap between them is the diagnosis



- **Cp is what you could achieve if you centred the process.**
- **Cpk is what you are achieving now.**
- Cp ≈ Cpk → the process is centred. If capability is poor, it is a spread problem, and that is a project: tool, recipe, or measurement system.
- Cp ≫ Cpk → the process is off target. That is usually a recipe offset and an afternoon's work.
- **This is the asymmetry from Lecture 1,** made quantitative: centring is cheap, spread is expensive.

Always report both. Cpk alone hides which of the two problems you actually have.

Georgia Tech

What the numbers mean

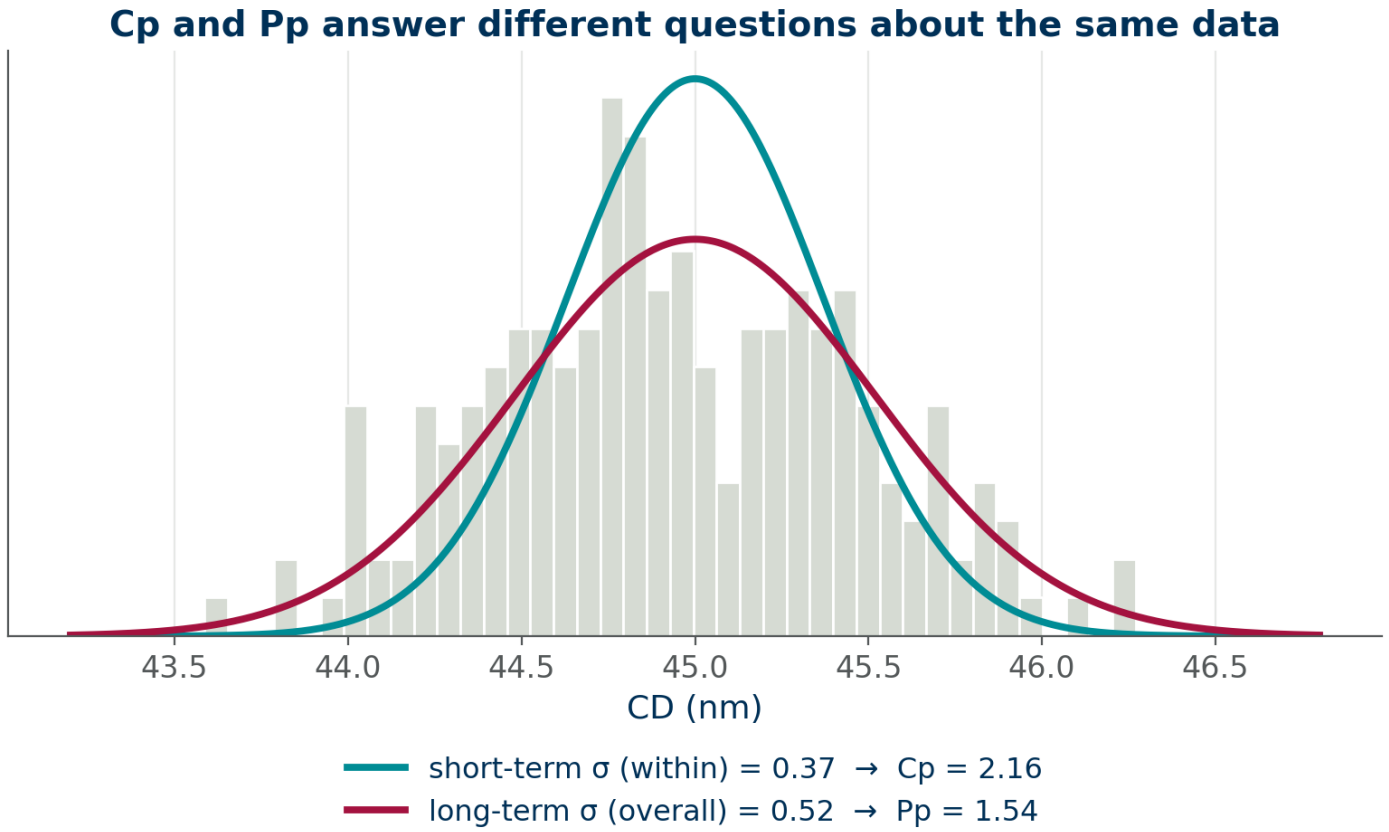
Cpk, defect rate, and what a fab expects

Cpk	Nearest limit at	ppm out of spec	Typical verdict
0.67	2σ	22 750	Not viable
1.00	3σ	1 350	Marginal — the historical minimum
1.33	4σ	32	Common contractual requirement
1.67	5σ	0.3	Automotive / medical
2.00	6σ	0.002	'Six sigma' — before any drift allowance

These are one-sided figures for a perfectly centred, perfectly normal process. Every one of those assumptions fails a little in practice, and the tails are where it shows.

Georgia Tech

Cp and Pp: short term and long term



Same data. σ from within-subgroup range gives Cp. σ from all the data pooled gives Pp. The gap is everything that drifts between subgroups.

Georgia Tech

Which one should you quote?

Cp and Cpk — short term

$\hat{\sigma}$ from within-subgroup variation, $\bar{R}/d_2$. The same estimate the control chart uses.

Answers: what is this process capable of when it is behaving?

Use it for process development and improvement work.

Pp and Ppk — long term

σ from all the data pooled, including everything that drifted between subgroups.

Answers: what did the customer actually receive over this period?

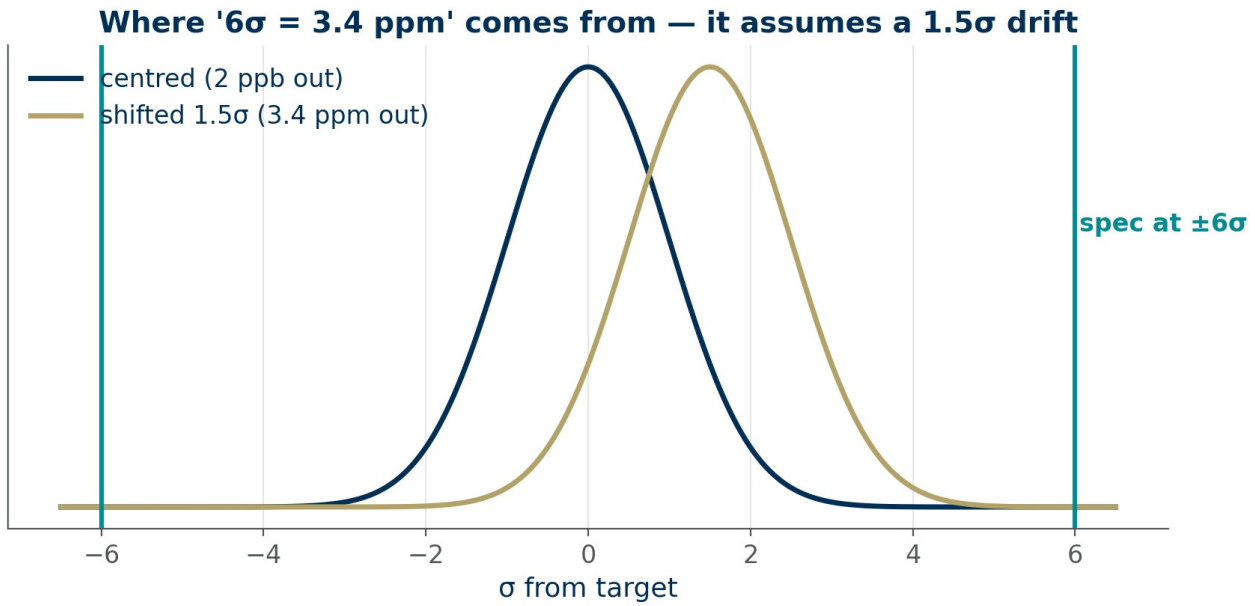
Use it for qualification, reporting, and contracts.

$Cp \gg Pp$ means the process is stable in the short run and wanders in the long run — which is a control problem, not a capability problem.

Quoting the flattering one without saying which is the oldest trick in quality reporting.

Georgia Tech

Where 'six sigma = 3.4 ppm' comes from



- A perfectly centred process with spec limits at ±6σ produces about two defects per billion — not 3.4 per million.
- **The famous number assumes the mean drifts by 1.5σ.**
- Motorola's empirical observation was that real processes wander over the long run, and 1.5σ was their allowance for it.
- So '6σ quality' means: 6σ of margin, minus a 1.5σ drift budget, giving an effective 4.5σ and 3.4 ppm.
- **It is a convention with an empirical justification —** not a property of the normal distribution. Your process may drift more, or less.

When someone quotes 3.4 ppm, ask whether their 1.5σ drift allowance matches their actual data.

Georgia Tech

What capability is expected of, in a real fab

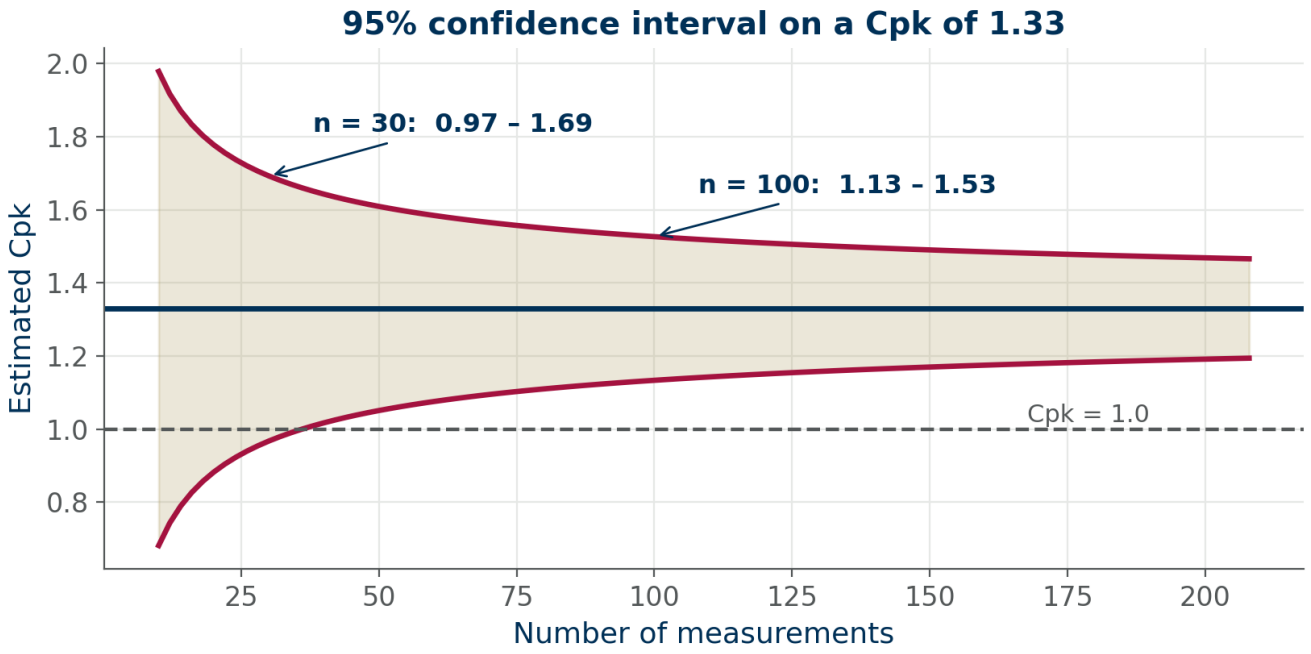
The target is negotiated per layer, not set globally

Parameter	Why it is hard	Typical Cpk target
Gate CD	Directly sets drive current and leakage; tolerance shrinks every node	1.33 – 2.00
Overlay	Multivariate and multi-layer; budget shared across the stack	1.33 – 1.67
Gate oxide thickness	Exponential effect on leakage; very tight window	1.67+
Metal line width	More forgiving at upper levels	1.00 – 1.33
Implant dose	Well controlled by the tool; rarely the limiter	1.67+

Notice there is no single number. Capability targets are set by how much the parameter matters to the device and how much margin the design gave you — which is why process and design teams argue about them.

Georgia Tech

How much do you trust that Cpk?



- Cpk is an estimate from a sample, so it has a confidence interval — and it is much wider than anyone expects.
- **With 30 measurements, a computed Cpk of 1.33 means the true value is somewhere between 0.97 and 1.69.**
- You cannot even distinguish 'marginal' from 'excellent'.
- With 100 measurements it narrows to about 1.13 to 1.53 — better, still not tight.
- **So a supplier certificate quoting 'Cpk = 1.35' from 25 parts** is telling you almost nothing. Ask for n.

Report Cpk with its sample size, or do not report it.

Georgia Tech

Where capability goes wrong

Four ways to produce a confident number that is not true

The process is not stable

The number describes a mixture that will never recur. This is the big one, and it is why stability comes first.

The data is not normal

Cpk converts a distance into a tail probability using the normal model. Skew changes that probability by orders of magnitude.

Too few measurements

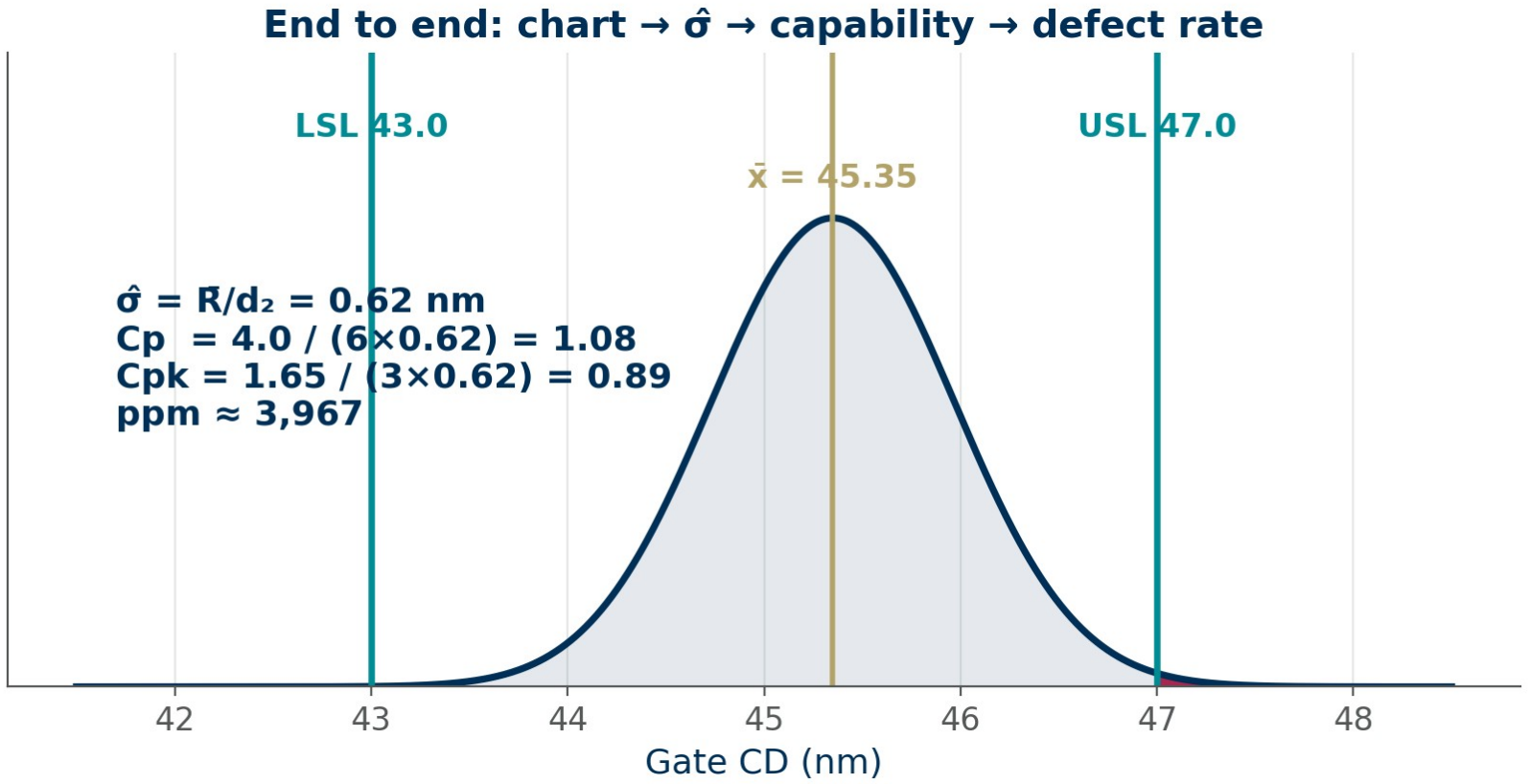
Thirty points gives you a range of ±0.35 on Cpk. The point estimate looks authoritative and is not.

Measurement variation is inside σ

Your σ includes gauge error, so you are penalising the process for your metrology. This is Lecture 3.

Georgia Tech

End to end



The same CD data from this morning: chart → σ̂ from R̄/d₂ → Cp and Cpk → parts per million → yield.

The three questions, answered

Is it stable?

$\bar{X}$–R for subgrouped data. I–MR when n = 1, with the assumptions checked. p, np, c, u for attributes.

Read the spread chart first.

What is it telling me?

Zones and run rules. Each pattern points at a class of physical cause.

Turn on the rules that match your failure modes, and pay the false-alarm price knowingly.

Is stable good enough?

Cp for spread, Cpk for spread and centring, Pp and Ppk for the long run.

Always with a sample size attached.

What is still missing: do you believe your measurements at all?

Every σ today included the metrology tool's own variation. Lecture 3 opens there.

Georgia Tech

Homework

Due before Lecture 3 · datasets on the course site

1. Build an $\bar{X}$–R chart from the supplied 30-subgroup CD dataset. Report $\bar{R}$, $\hat{\sigma}$, and both sets of limits. Identify every out-of-control subgroup and state which rule it violates.

2. Take the same data as individuals — first wafer of each lot only — and build an I–MR chart. Compare the limits and the signals you detect. Explain the difference using $\sqrt{n}$.

3. For the etch rate dataset, plot x_i against x_{i-1}. Estimate the lag-1 correlation, and state what it does to the validity of the I chart you would build.

4. Compute Cp, Cpk, Pp and Ppk for the CD data against spec 43.0–47.0 nm. Explain what the gap between Cp and Cpk tells you, and what the gap between Cp and Pp tells you.

Georgia Tech

References and pre-reading

Core reading

Montgomery, Introduction to Statistical Quality Control — Ch. 5 ($\bar{X}$–R, I–MR), Ch. 7 (attribute charts), Ch. 8 (capability).

May & Spanos — Ch. 6 for the semiconductor treatment of SPC.

Western Electric Statistical Quality Control Handbook (1956) — the original rules, and still remarkably readable.

Before Lecture 3

Read Montgomery Ch. 8.7 on gauge capability.

Bring your homework Cpk numbers — we will recompute them after removing measurement variation, and the answers change more than you expect.

Next: Lecture 3 — Measurement Systems, Fab Case Studies, and SPC inside APC

Gauge R&R · etch, CD and overlay · run-to-run control · multivariate SPC and ML

Georgia Tech