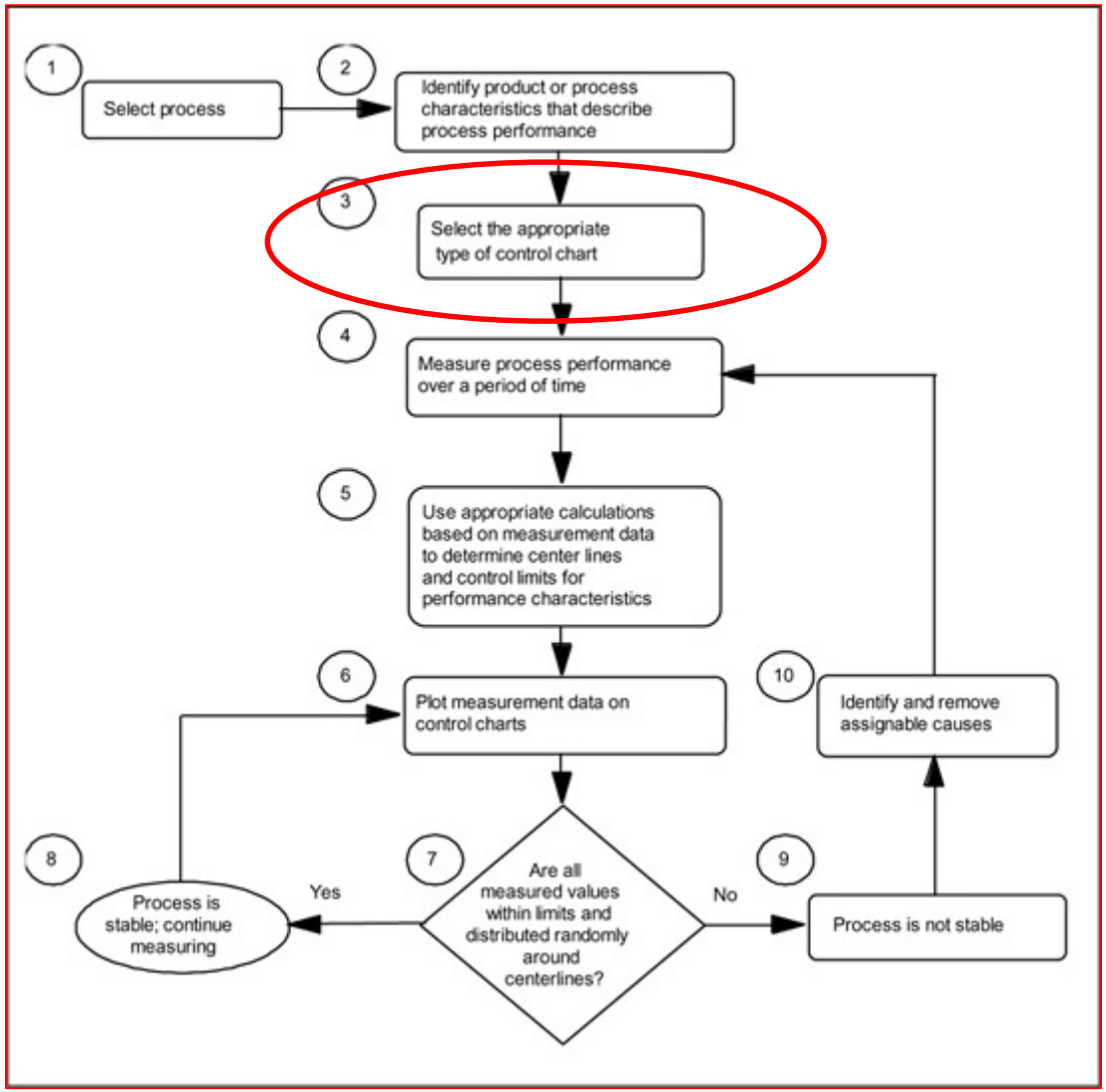
CALIFORNIA STATE UNIVERSITY
FULLERTON™

Control Charts for Variables and Discrete Data

Paulo de Oliveira
paulohgo@csu.fullerton.edu

Using Control Charts



Types of data

- Variables
- Discrete data (Attributes)

Unless you have a clear understanding of the distinctions between the two kinds of data, you can easily fall victim to inappropriate control-charting methods

Types of data - Variables

- Also called *Measurement data*
- They are usually measurements of *continuous phenomena*
- Examples:
  - Memory utilization
  - CPU utilization
  - Cost of rework

Types of data - Attributes

- Different origin than Variables
- Information is recorded when an item conforms to a specified criteria
- Almost originate as *counts*
- They are usually measurements of *continuous phenomena*
- Examples:
  - Number of design defects found
  - Number of lines of code comment

Understanding your data type

- Rule of thumb:
  - Variables data => Continuous phenomena
  - Attributes => Counts
- DANGER! That's not always the case though!!
  - Counts get used as measures of size instead => Variables. Example: total lines of code, backlogged items…

Exercise

- What type pf data is this?
  1. Daily count of bugs for a product
  2. Team velocity (per sprint) over time
  3. Number of localizability bugs in a product being developed
  4. Number of active defects for a product
  5. Support calls from customers

Exercise - Solution

- What type of data is this?
  1. Daily count of bugs for a product (variables)
  2. Team velocity (per sprint) over time (variables)
  3. Number of localizability bugs in a product being developed (attribute)
  4. Number of active defects for a product (atribute)
  5. Support calls from customers (variables)

Why do we want to know?

- There are different types of control charts that can be employed in different situations, depending on your data type and how you'd like to use it
- It's important to clearly identify what type of data you have to map it to the different options of control charts

Types of Control Charts

- Xbar and R (Average and Range)
- Xbar and S (Average and Standard deviation)
- XmR (Individuals and Moving Range)
- MAMR (Moving average and Moving Range)
- c
- u
- Z

Data Types and Control Charts

Chart type	Discrete	Variable	Subgroups
Xbar and R	x	x	x
XmR	x	x	
Xbar and S	x	x	x
c	x		
u	x		
Z	x		

Xbar and R

- Charts for averages (X-bar charts) and range (R) charts are used to portray process behavior when you have the option of collecting multiple measurements within a short period of time under basically the same conditions
- Variables data, mostly
- Subgrouping takes place
- Example 5.1 from the textbook is a great exercise

Xbar and R

- Xbar chart answers the following question:
  - What is the central tendency of the process?
- R chart answers:
  - What is the variation within the subgroups?

- Average and Range are used together!
- Maximum 10 rounds of measurement recommended

Xbar and S

- The efficiency of Xbar and R charts falls off rapidly as the size (n) of the subgroups increases beyond *n = 10*
- Most experts advise that range charts be used only when there are 10 or less observations in each subgroup
- For subgroups larger than 10, S charts based on averages of the standard deviation within subgroups give tighter control limits

Xbar and S

- Why SD instead of range?
  - Too many items in a subgroup make the Range information less meaningful
  - Standard deviation provides a more formal statistical analysis to the *Range* data
- Example 5.2 from the textbook

XmR for Continuous Data

- When measurements are spaced widely in time or when each measurement is used by itself to evaluate or control a process, a time-sequenced plot of individual values, rather than averages, may be all that is possible
- The subgroup size *n =1*
- We use the short-term variation between adjacent observed values to estimate the natural (inherent) variation of the process

XmR for Continuous Data

- When measurements are spaced widely in time or when each measurement is used by itself to evaluate or control a process, a time-sequenced plot of individual values, rather than averages, may be all that is possible
- The subgroup size *n =1*
- We use the short-term variation between adjacent observed values to estimate the natural (inherent) variation of the process

XmR for Discrete Data

- XmR charts will be the choice in 99% of the cases
- u, c, p, np charts can be used in special occasions. We will not discuss them in this course.

Conclusion

- The X-bar and R charts apply to variables or discrete data, as do the X-bar and S charts.
- The individuals and moving range charts may be used for either variables or attributes data and are particularly appropriate to use with periodic data (subgroups of n = 1).
- The c, u, and Z charts apply to attributes or count data, provided the underlying data distribution conforms to the Poisson distribution.

Group Exercise

- Get together with your teams and solve the following problems. For each problem:
  - Decide on the best chart to be used
  - Come up with a table of fictitious measurements
  - Based on your data, draw the chosen control chart, decide if the process is under control or not

Group Exercise

1) *Company "Supersoft" produces a product that takes a long time to compile. Recently, the team leader receives a complaint from one of the team members that builds are taking too long to compile. That value (time to build) can be measured daily and the team wishes to understand if the variations being experienced are within normal rates or not.*

Group Exercise

2) *An Agile team runs on weekly sprints. An automated set of test cases intended to verify that the product is stable is run every night and produces several bugs based on a daily build of a software product. This process runs regularly for 2 months and is part of the team's quality control processes. The team wishes to know how much instability (defects) the weekly sprints are causing in the main product with the added work per sprint.*