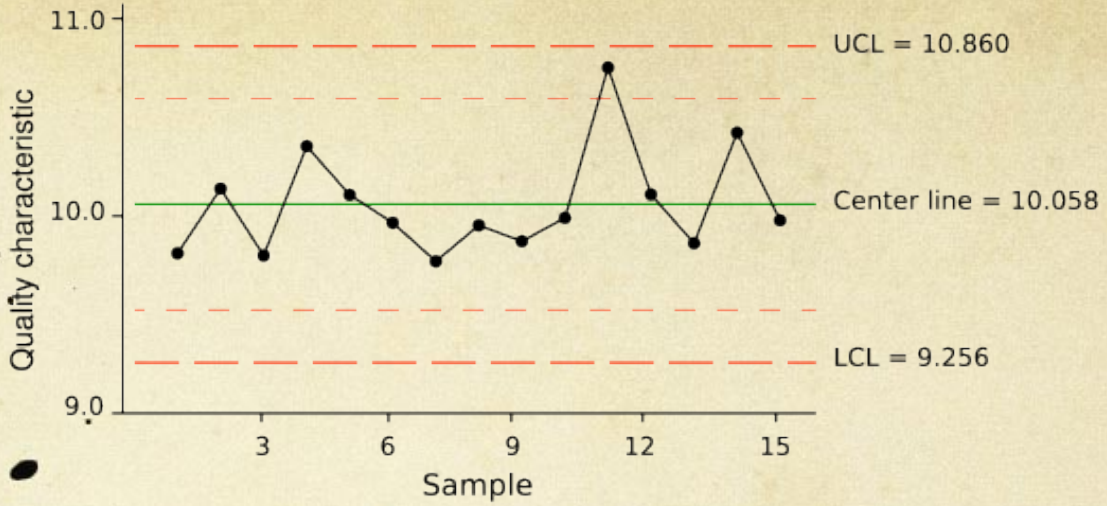


Session on Novel Contexts for SPC Methodology and Applications: Discussion

Ron Fricker

2014 Joint Statistical Meetings

Session Outline

- **Statistical Process Control for Reliability Assurance of Devices in Long Term Storage**
  - Alix Robertson, Sandia National Laboratories
- **SPC Methods for Non-Stationary Correlated Count Data with Applications to Network Surveillance**
  - Daniel Jeske & Yingzhuo Fu, University of California, Riverside
- **A Robust Phase I EWMA Chart for Dispersion**
  - Ronald Does, Inez Zwetsloot & Marit Schoonhoven, University of Amsterdam
- **Generalized Exponential Percentile Control Charts**
  - Trenton Brown, Yuhlong Lio & Nan Jiang, University of South Dakota

Novel Contexts

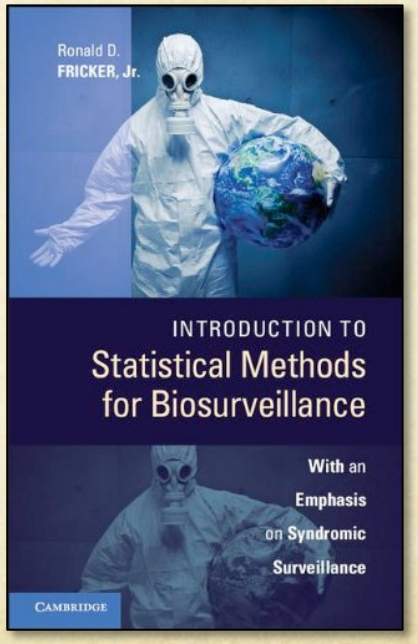
- This session:
  - Specific applications: stockpile & computer network monitoring
  - General applications: variability & lifetime monitoring
- Other specific applications:
  - Healthcare
    - Google "statistical process control new applications"
  - Biosurveillance
    - Detecting changes in disease prevalence
  - Social media sentiment analysis
    - Are population attitudes about something changing?
  - SPC & Big Data
    - Per Dennis Lin*:
      - Time series 101—each observation is a number (scalar)
      - Time series 201—each observation is a vector
      - Time series 301—each observation is a function (between response and covariates)
      - Time series 401—each observation is a network
      - Time series 501—each observation is a graphic

* Lin, D.K.L. (2014). Comments: Some Challenges for Multivariate Statistical Quality Control, *Quality Engineering*, **26**, 96-98.

Novel Context Challenges

	Classical ~~Typical~~ SPC Data Characteristics		Typical Biosurveillance Data Characteristics
1.	The in-control distribution (F_0) is (or can reasonably be assumed to be) stationary.	1.	There is little to no control over disease incidence and the disease incidence distribution is usually non-stationary.
2.	Observations can be drawn from the process so they are independent (or nearly so).	2.	Autocorrelation and the need to monitor all the data results in dependence.
3.	The asymptotic distributions of the statistics being monitored are known and thus can be used to design control charts.	3.	Individual observations being monitored, so asymptotic sampling distributions not relevant.
4.	Monitoring the process mean and standard deviation is usually sufficient.	4.	Little is known about which statistics are useful; often looking for anything unusual.
5.	Out-of-control condition remains until it is detected and corrective action is taken.	5.	Outbreaks are transient, with disease incidence returning to its original state when the outbreak has run its course.
6.	Temporal detection is the critical problem.	6.	Detecting both temporal and spatial anomalies are critical.

Table 6.1. Characteristics of typical SPC data compared to biosurveillance data.



Implications

"This brings us to another important challenge, that of the complexity of many ~~big data~~ [novel] *applications. SPC applications have a tradition of* back of the napkin *methods. The custom within SPC practice is the use of simple methods that are easy to explain like the Shewhart control chart. These are often the best methods to use to gain credibility because they are easy to understand and easy to explain to a non-statistical audience. However, ~~big data~~* [the novel application] *often does not lend itself to easy-to-compute or easy-to-explain methods."**

- Suggests:
  - More sophisticated / complicated process models required
  - Users may have to be more skilled to apply SPC in these contexts
  - Or perhaps we need better / more automated software
    - Perhaps some intersection of SPC and machine learning?

* Megahed, F.M. and L.A. Jones-Farmer (2013). Statistical Perspectives on "Big Data," 11th International Workshop on Intelligent Statistical Quality Control, Sydney, Australia. Accessed on-line at www.academia.edu/4357171/Statistical_Perspectives_on_Big_Data_.

Important SPC Research Principles

- Compare proposed new methods against existing standards / methods
  - Just proposing a method is not sufficient
- Assess performance under a variety of conditions
  - Don't just use scenarios that favor your method
- Provide guidelines for choosing parameters and other algorithmic settings
  - In the end, a layperson is likely going to have to apply the method
- Distinguish between Phase I and Phase II operations
  - In particular, avoid overly simplistic comparisons (e.g., actual parameter values assumed known)

Some Opening Questions for Our Speakers

- Alix:
  - An assumption seems to be that EWMA performs best, but that likely depends on rate of degradation.* Have you compared the EWMA to other options?
  - I would think that aging degradation goes in only one direction. If so, have you looked at using one-sided control charts?
- Dan:
  - How might one do Phase I monitoring in the real world to identify in-control observations (from which to fit the GLMM)?
  - What is the potential to automate the GLMM fitting process, and even the control chart (AILR, JLR) calculations, so that a non-statistician network monitor / layperson could implement?

* E.g., see Chang & Fricker (1999). Detecting When a Monotonically Increasing Mean has Crossed a Threshold, *Journal of Quality Technology*, **31**, 217-233.

Some Opening Questions for Our Speakers

- Ron:
  - In comparison to Dan's GLMM methodology, the robust EWMA chart for dispersion seems simpler to implement. What are your thoughts on the potential to automate the Phase I analysis and Phase II implementation to facilitate layperson application?
  - For some applications (network monitoring, biosurveillance), the change is sustained but transient. Any idea how your proposed method would work under these conditions?
- Trenton:
  - Could you give an example of how lifetime monitoring using control charts might be conducted in a manufacturing setting?
  - Do you have any thoughts on how one might do Phase I monitoring to identify in-control observations?