The Paradata Concept in Survey Research

Lars Lyberg

Stockholm University

CSDI

London, UK, March 24, 2011

The general concept of quality

- Fitness for use
- Something inversely proportional to variability

Assuring and controlling quality

Quality Level	Main stake-holders	Control instrument	Measures and indicators
Product	User, client	Product specs, SLA, evaluation studies	Frameworks, compliance, MSE, user surveys
Process	Survey designer	SPC, acceptance sampling, CBM, SOP, paradata, checklists, verification	Variation via control charts, other paradata analysis, outcomes of evaluation studies
Organization	Agency, owner, society	Excellence models, ISO, CoP, reviews, audits, self-assessments	Scores, strong and weak points, user surveys, staff surveys

Quality improvement

- The reduction of variability in processes and products
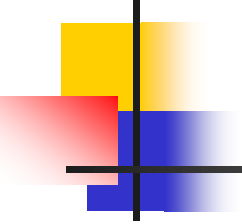- The reduction of waste

Definitions

- Process is a series of actions or steps towards achieving a particular end

- Process quality is an assessment of how far each step meets defined criteria

- Process variables are factors that can vary with each repetition of the process

- Key process variables are factors that have a large effect on process end result

- Observations of process variables result in paradata

Some terminology

- Data, Metadata, Paradata
- Macro paradata– global process data such as response rates, coverage rates, edit failure rates, sometimes broken down
- Micro paradata– process data that concern individual records such as flagged imputed records, keystroke data
- Formal selection, collection, and analysis of key process variables that have an effect on a desired outcome, e.g., increased productivity

The roots of paradata

- 1924 Shewhart's control chart
- 1940 the US War Department's guide for analyzing process data
- 1960 the zero defects program
- 1960's QC programs at statistical agencies

- 1975-89 TQM, Malcolm Baldrige, EFQM, Six Sigma
- 1997 Marker and Morganstein, Bristol monograph
- 1998 Couper paradata
- 2006 Groves and Heeringa responsive design

Issues

- Paradata are a subset of process data
- Selecting key process variables
- Collecting and analyzing paradata
- Diagnosing the variability pattern
- Is the variation due to common or special causes?
- Paradata are multivariate by nature
- Action or no action?
- Risks (a lot of data, ghost chasing)

Control chart (example)

c Control Chart
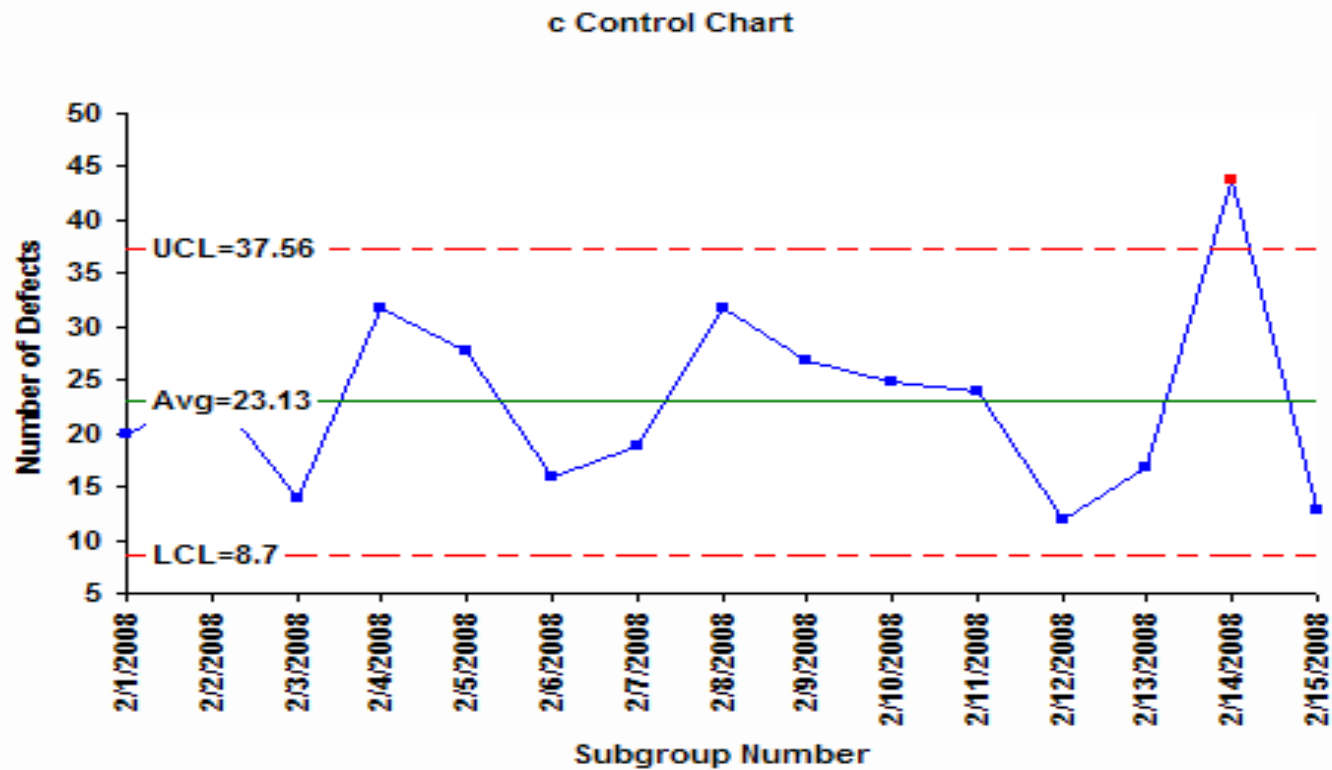
Marker and Morganstein

- Never collect process data that are not related to quality

- Collecting data on processes related to quality without using SPC and other proper analysis methods is extremely wasteful

- If you don't know how to analyze don't collect

Importance of paradata (I)

- Continuous updates of progress and stability checks (monitoring)
  - Control charts, standard reports
  - Managers choose to act or not to act
  - Early warning system
- Input to long-run process improvement of product quality
  - Analysis of special and common cause variation
- Input to methodological changes
  - Finding and eliminating root causes of problems
  - Research

Importance of paradata (II)

- Responsive designs
  - Simultaneous monitoring of paradata and regular survey data to improve efficiency and accuracy
- Input to organizational change
  - E.g., centralization, decentralization, standardization
- Quality profiles, client communication, public use paradata files, inference, picturing quality over time

Paradata in a 3 M perspective

- Overwhelming evidence that process specifications are not uniformly adhered to across countries
- QC using spec checks and SPC necessary not only for data collection to preserve comparability
- Capacity building necessary
- Need for specific 3 M QC approaches