Quality control charting

The Norwegian X-ray conference
13.09.10
Hege Zahl

I shall……..

- Define the analytical system
- Briefly discuss common pit falls when establishing the reference values for the analytical system
- Present Quesenberry-statistics and provide the rationale for using Q-charts in quality control charting
- Introduce the ”Statistics Mill”, created by Bente Kroka
- Show the calculation of control limits for Shewhart charts

The analytical system

- The set of samples of a given matrix
- The exactly defined analytical procedure
- The instrument
- The calibration equations enabling the calculations of analyte concentrations

The control sample

- Prepared and presented to the instrument as a pressed powder pellet, a liquid or a glass bead
- Criterion: the control sample material must be **homogeneous**
- Stored in temperature and humidity controlled conditions in order to remain stable

The control process

- Error is inherent in any analytical system, even during stable and well controlled conditions!

→ The control process should detect

Human errors

Changes in the analytical system

Establishing the reference values

- Use a CRM or inhouse control material

- CRM values are known, however the **true value μ** that the measurements will fluctuate around, is unknown

→ the analytical system introduces bias!

True value versus measured value

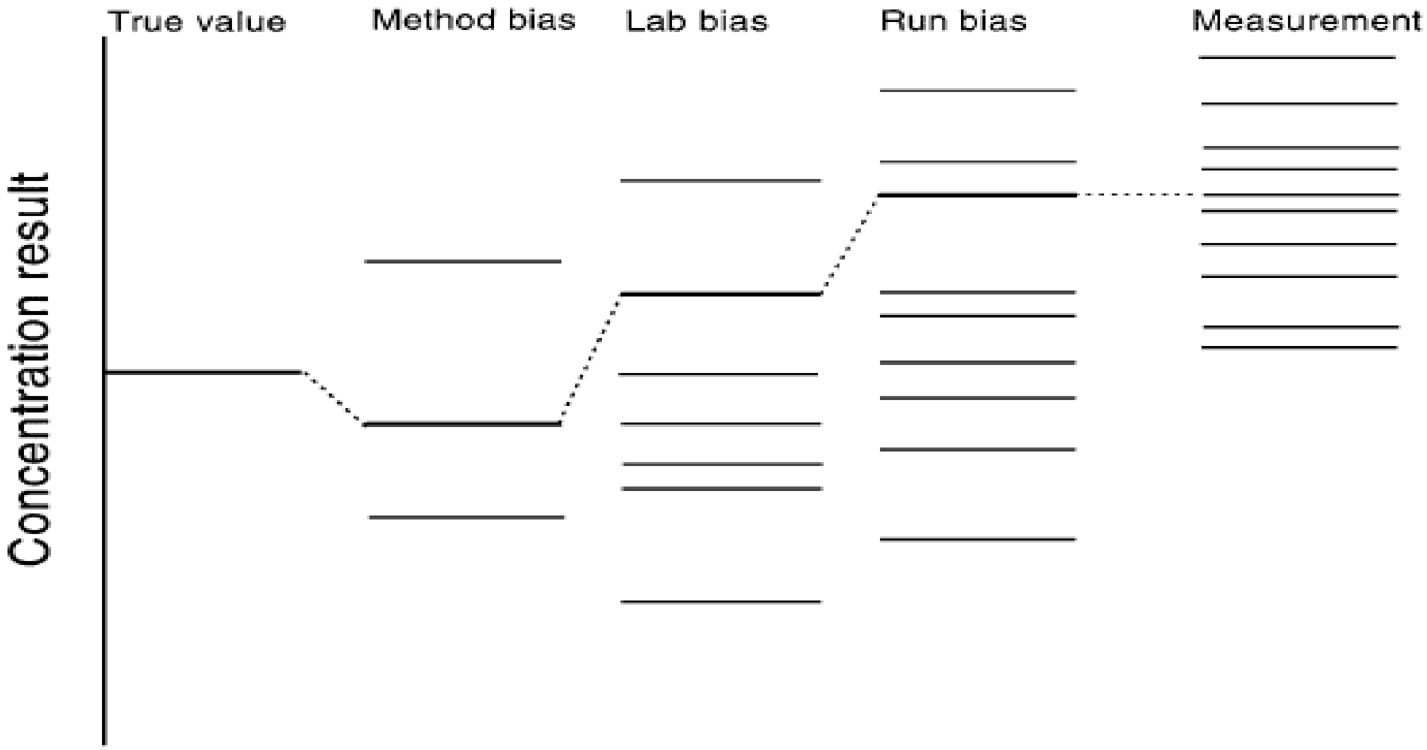

Figure reproduced from Thompson (2000).

Establishing the reference values

- The spread around μ, which is the **standard deviation σ**, is also unknown
- Need to estimate μ and σ correctly for the analytical system
- It is not correct to estimate μ to be equal to the CRM value

Quesenberry chart

- Method development, validation and documentation
  → analytical system is ready to use

- Common situation: wish to monitor the analytical system before the data set is large enough

- At this stage, it is not possible to estimate μ and σ for the analytical system

Quesenberry chart

- Assumption: no autocorrelation (refer to Miller and Miller 2005 for more information on tests)
- Prepare and measure the control sample, set the values for the first observation to $X_1 = x_1$ og $S^2_1 = 0$
- Sequential estimates of X and S^2 are obtained run-by-run using standard updating methods:
  - $X_i=(1/i)\mathbf{[}(i-1)X_{i-1} + x_i\mathbf{]}$, i = 2, 3,……….., N
  - $S^2_i= (i-2)/(i-1)\ S^2_{i-1} + 1/i\ (x_i – X_{i-1})^2$, i = 3, 4,……….., N

Q-statistics for the mean

- Calculation of mean $Q_i(X_i)$:
  - $Q_i(X_i) = \varphi^{-1}\{G_{i-2}[\sqrt{(((i-1)/i)*((X_i - X_{i-1})/(S_{i-1})))}]\}$, $i = 3, 4, \ldots\ldots, N$
  - $G_v(x)$ = single sided Student-*t* distribution, v = d.o.f
  - $\varphi^{-1}(x)$ = inverse of the single sided standard normal distribution function

- The Q-statistics for the individual observations, $Q_i(X_i)$, are a sequence of independent and identically distributed standard normal variates

  → can be plotted in a $Q_i(X_i)$ chart with a centre line at zero and control limits at ± 2 and ± 3, beginning with the third observation

Q-statistics for the variance

- Calculation of variance $Q_i(R^2_i)$:
  - $Q_i(R^2_i) = \varphi^{-1}\{F_{1,v}(vR^2_i/(R^2_2 + R^2_4 + \ldots + R^2_{i-2}))\}$, $i = 4, 6, \ldots, N$
  - $R_i = |x_i - x_{i-1}|$
  - $v = i/2 - 1$ degrees of freedom
  - $F_{v1,v2}(x)$ = one sided Snedecor-*F* distribution function with (v_1, v_2) degrees of freedom
  - $\varphi^{-1}(x)$ = inverse of the single sided standard normal distribution function

- The Q-statistics are sequences of independent and identically distributed standard normal variates, but are calculated for even-numbered observations only

  → can be plotted in a $Q(R^2)$ -chart with a centre line at zero and control limits at ± 2 and ± 3, beginning with the fourth observation

Elkem

Q-chart advantages

- An outlier can immediately be eliminated from the data set, and is excluded from subsequent Q-calculations
- The power of detecting an outlier increases as the amount of past data grows
- The control process is continuously updated
- The standardisation of the presentation makes for easy comparison where control of more than one analyte is needed, as the Q(X) and Q(R^2) statistics for several analytes can be plotted on the same respective Q-chart

Size of data set for calculation of S-chart bounds

- Prior to the introduction of Q-chart, it was often acceptable to estimate μ and σ based on m = 20-30 sub-groups composed of n = 4-5 measurements per sub-group
- **In order to secure a reliable control system, use *minimum* m = 100 sub-groups composed of n = 5 measurements each**
- **$m \geq 400/(n-1)$**

Size of data set for calculation of S-chart bounds

- % increased false alarm rate during the 20 subsequent measurements if m sub-groups are used:

m	% increased false alarm rate
20	39
30	24
50	16
100	0

The Statistics Mill

- An excel macro which
  - Performs Q-statistics
  - Displays updated Q-charts
  - Calculates Shewhart-chart control limits at m = 30
  - Displays S-charts
  - Tests the Q- and S-chart observations against the SPC rules
  - Logs SPC rule violations automatically
- The purpose of the Statistics Mill was to make quality control charting of our XRF-methods easy, fast and reliable

- Made by Bente Kroka at Elkem Solar Research

Demonstration

Q data only - Shortcut.lnk

Calculation of μ for S-chart

- Best estimate of **μ** is the grand mean for k runs:
  - $\mu = \frac{1}{k} \sum_{j=1,k} \mathbf{x}_j$
- Mean for sub-groups:
  - $\mathbf{x}_j = \frac{1}{n} \sum_{i=1,n} x_i$

Calculation of σ for S-chart

- Pooled estimate of the standard deviation for k runs:
  - $\sigma = \sqrt{\{\sum_{j=1,k} s_j^2/k\}}$
- Standard deviation for sub-groups:
  - $s = \sqrt{\{\sum_{i=1,n} (x_i - x_{mean})^2/(n-1)\}}$
- Alternatively: use MS (Within groups) from ANOVA single factor:
  - $\sigma = \sqrt{MS}$

Calculating S-chart **μ** control limits

- Upper/lower warning limit
  - $\mu \pm 1.96\sigma/\sqrt{n}$

- Upper/lower action limit
  - $\mu \pm 3.09\sigma/\sqrt{n}$

- Acceptable to use 2σ og 3σ

Calculating S-chart σ control limits

- Upper/lower warning limit
  - $\{1 \pm (1.96/C_4)\sqrt{(1-C_4^2)}\}\sigma$
- Upper/lower action limit
  - $\{1 \pm (3.09/C_4)\sqrt{(1-C_4^2)}\}\sigma$
- The correction factor C_4 can be found in Howarth (1995) (table 3)

n	2	3	4	5	10
C_4	0.7979	0.8862	0.9213	0.9400	0.9727

The SPC rules (Western Electric rules)

- Action should be taken if one or more of the following conditions occur:
    1. 1 point beyond the upper or lower 3 sigma limit
    2. 2 out of 3 successive points beyond the upper or lower 2 sigma limit
    3. 4 out of 5 points 1 sigma or more away from the centreline
    4. 8 successive points on one side of the centreline
    5. 8 successive points on both sides of the centreline, but at least 1 sigma away

Summary

- The Q-chart is
  - capable of reliable monitoring of mean and variance during data set acquisition
  - capable of signalling a shift in the analytical system
- When a sufficiently large data set consisting of **$m \geq 400/(n-1)$** has been obtained, this can be used as basis for *any* classic control chart
- The Statistics Mill
  - an excel macro which makes Q-charting quick and reliable
  - fulfils the data set size criterion: $m = 30$
  - automatically calculates Shewhart chart limits at $m = 30$

Take home message

- Q-charting is the most suitable and reliable choice during the data acquisition phase
- Successful S-charting requires the use of correct formulae and a sufficiently large data set for control limit calculations, and knowledge of the SPC-rules
- Laboratories that take quality control seriously will become very proficient!

Thank you….

- Bente Kroka
- Ole Petter Thorstensen and Gro Eide
- …to all of you for listening!

References

- Bhattacharyya, G. K. og Johnson, R. A. (1977) *Statistical Concepts and Methods*, 1st edition, John Wiley & Sons, Inc
- Howarth, R. J. (1995) Quality Control Charting for the Analytical Laboratory Part1. Univariate Methods, *Analyst*, **120** s.1851-1873
- Miller, J. N. og Miller, J. C. (2005) *Statistics and Chemometrics for Analytical Chemistry*, 5th edition, Pearson Education Ltd
- Thompson, M. (2000) Towards a unified model of errors in analytical measurement, *Analyst*, **125**, s. 2020-2025