A critical view on ISO standard 13528

Wim Coucke

EQALM symposium, Dublin,

20th of October 2017

.be

ISO 13528(2012): aims

- Detailed descriptions of statistical methods

- Procedures can be applied to demonstrate that the measurements meet specified criteria for acceptable performance

- Applicable to either quantitative measurements or qualitative observations

- Applicable, especially for newly established proficiency testing schemes.

.be

Sample quality: Homogeneity

ISO 13528:

- At least 10 samples at random

- Analyse samples in random order in duplicate

- Analytical variability of method$<0.5\sigma_{pt}$

- inter-sample standard deviation $<0.3\sigma_{pt}$

- Inferential test, reject sample if evidence of heterogeneity

Sample quality: homogeneity

Probability of non-estimable sample standard deviation increases with ratio analytical variability/sample variability

If $\sigma_s = 0.3\sigma_{pt}$ and $\sigma_a = 0.5\sigma_{pt}$, duplicates: 21%

If $\sigma_s = 0.4\sigma_{pt}$ and $\sigma_a = 0.5\sigma_{pt}$, duplicates: 11%

Probability can be reduced by taking more replicates per sample.

Sample quality: homogeneity

Probability of non-estimable sample standard deviation increases with ratio analytical variability/sample variability

If $\sigma_s = 0.3\sigma_{pt}$ and $\sigma_a = 0.5\sigma_{pt}$, triplicates: 8%

If $\sigma_s = 0.4\sigma_{pt}$ and $\sigma_a = 0.5\sigma_{pt}$, triplicates: 2%
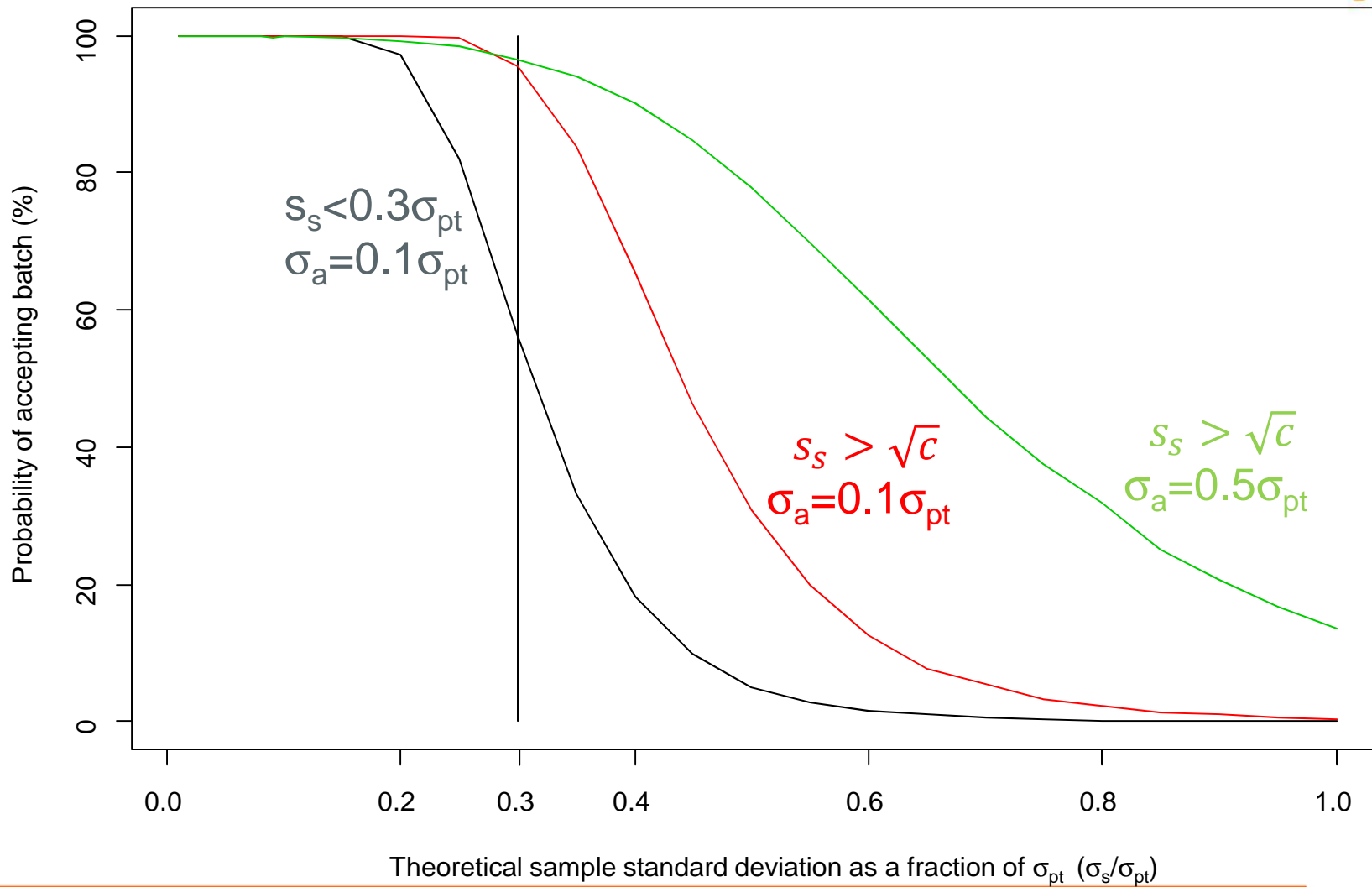
Sample quality: homogeneity

Inferential test:

$$c = F_1 \sigma^2_{allow} + F_2 \sigma^2_w, \text{ reject batch if } s_s > \sqrt{c}$$

Test only rejects if sample heterogeneity is too large

Accepts batch in all other cases

Sample quality: homogeneity

Probability of accepting batch (%)

$s_s < 0.3\sigma_{pt}$
$\sigma_a = 0.1\sigma_{pt}$

$s_s > \sqrt{c}$
$\sigma_a = 0.1\sigma_{pt}$

$s_s > \sqrt{c}$
$\sigma_a = 0.5\sigma_{pt}$

Theoretical sample standard deviation as a fraction of σ_{pt} (σ_s/σ_{pt})

Sample quality: homogeneity

What about qualitative schemes ?

ISO 13528: Appropriate sample of proficiency test items, all of which should demonstrate the expected property value

What is the frequency of nonconforming units in a batch if a certain number of samples are all conforming and we don't have any prior knowledge about the effectiveness of our preparation procedure ?

→ 95% lower limit of confidence:

Sample size	95% upperlimit of confidence of rate of nonconforming units
10	23.8
50	5.7
57	5
100	2.9

Use acceptance sampling theory or Bayesian approach with informative prior

Sample quality: Stability

ISO 13528: take at least 2 samples, measure them in the beginning and at the end.

If difference of means $< 0.3\sigma_{pt}$, accept batch for stability

Comment:

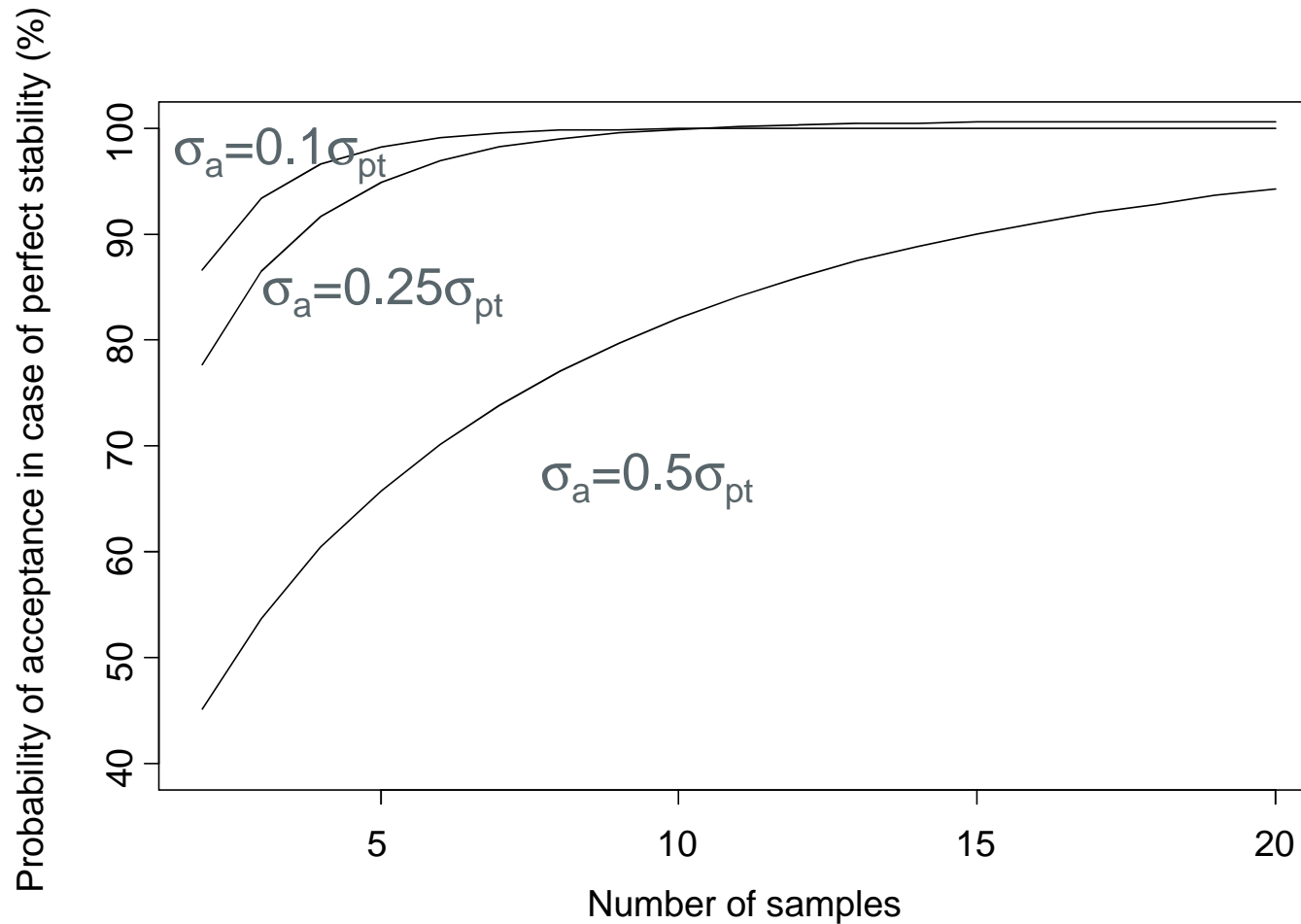Probability of meeting criterion in case of perfect stability:

$$\bar{Y}_1 \sim N(AV, \frac{0.5\sigma_{pt}}{\sqrt{2}}), \qquad \bar{Y}_2 \sim N(AV, \frac{0.5\sigma_{pt}}{\sqrt{2}})$$

$$\bar{Y}_1 - \bar{Y}_2 \sim N(0, \sqrt{\left(\frac{0.5\sigma_{pt}}{\sqrt{2}}\right)^2 + \left(\frac{0.5\sigma_{pt}}{\sqrt{2}}\right)^2}), \bar{Y}_1 - \bar{Y}_2 \sim N(0, 0.5\sigma_{pt})$$

$$P(|\bar{Y}_1 - \bar{Y}_2| < 0.3\sigma_{pt}) =$$

$$P(\bar{Y}_1 - \bar{Y}_2 < 0.3\sigma_{pt}|\bar{Y}_1 - \bar{Y}_2 > 0) + P(\bar{Y}_1 - \bar{Y}_2 > -0.3\sigma_{pt}|\bar{Y}_1 - \bar{Y}_2 < 0) \quad \textbf{=0.45}$$

Sample quality: stability

Solution: use very precise method or
assess relation between date of analysis and reported value

Normal distribution of data

- ISO 13528 not very clear:

  - 6.6.1: not always necessary to verify normal distribution, approximate symmetry is enough, only if outlier searching techniques are used (point 6.6.1).
  - 9.4.4: proficiency testing provider may wish to check the normality of the distribution if there is a very large number of participants.

- In fact:
  - Threshold values of Z scores are based on the normal distribution, so better to check.
  - Robust estimators of variability assume a normal distribution with outliers

- Recommended way of checking: normal quantile plot after removing outliers.

Data quality: multimodality

- ISO 13528:

  - mentions unimodality two times and evaluates robust estimation procedures with respect to their robustness against multimodality

- In fact:

  - Multimodality may show bad repartitioning of peer groups or lack of harmonisation

  - Multimodality is an interesting finding, not something we should protect ourselves against

.be

Data quality: multimodality

Histogram or density plot for multimodality ?

Enough data are needed

Hartigan's dip test

Silverman test (more powerful)

After adjusting for outliers and same values appearing multiple times !

Estimates of deviation:

Based on reported data:

MAD / nIQR / Algorithm A / Qn / Q

Comment:

- Algorithm A: originally designed for finding assigned value.

- Qn: only when all values are different

Estimates of deviation

Z scores: $\dfrac{x_i - x_{pt}}{\sigma_{pt}}$ ---- Z' scores: $\dfrac{x_i - x_{pt}}{\sqrt{\sigma^2_{pt} + u^2(x_{pt})}}$

Comments:

Z' scores are corrected for uncertainty in estimation of central location

Why is there no corrected for uncertainty in estimation of deviance when standard deviation is calculated based on the reported results ?

Estimates of deviation

Z scores: $\dfrac{x_i - x_{pt}}{\sigma_{pt}}$ ---- Z' scores: $\dfrac{x_i - x_{pt}}{\sqrt{\sigma^2{}_{pt} + u^2(x_{pt})}}$

Possible solution:

compare Z score with quantile of t distribution
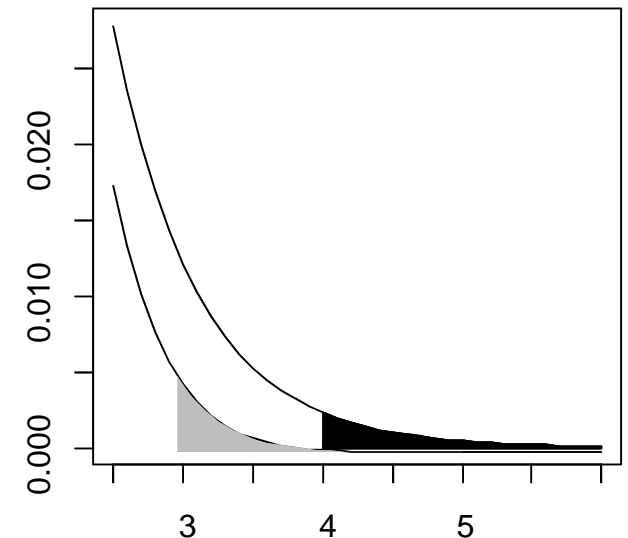Give corresponding inverse quantile of normal distribution

eg:
Z-score of 4, n=10.
Quantile of t-distribution: 0.998448;
corresponding inverse quantile of normal Distribution: 2.96

Z-score of 4, n=20. Corresponding value: 3.36

# Assessing quality checks for estimates of location and deviation

- ISO 13528:
    - The proficiency testing provider should apply a procedure to monitor interlaboratory agreement, to track changes in performance and ensure the reasonableness of statistical procedures.

      ...

    - Estimates of variability should be plotted on graphs sequentially or as a time-series

How to do as well: monitor interlaboratory agreement

Characteristic function:

$$SD = \sqrt{\alpha^2 + \beta^2 AV^2}$$

95% confidence bounds
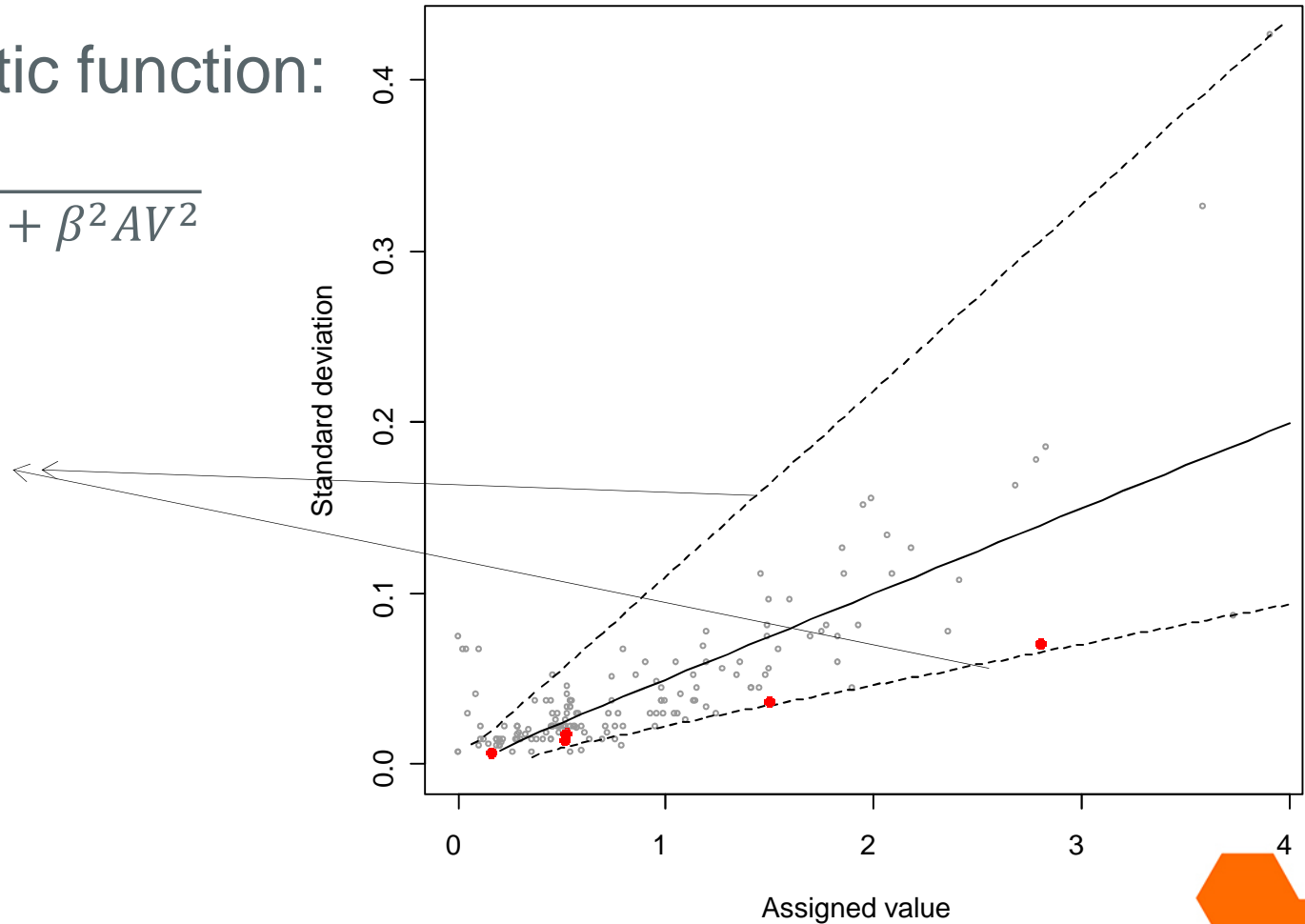
ISO 13528(2012): aims

- Detailed descriptions of statistical methods

- Procedures can be applied to demonstrate that the measurements meet specified criteria for acceptable performance

- Applicable to either quantitative measurements or qualitative observations

- Applicable, especially for newly established proficiency testing schemes.

Conclusion

- Very complete in some cases, too superficial in others

  - ISO 13528 lacks sufficient description for

    - long-term evaluation
    - qualitative schemes

Literature

Cárdenas S, Valcárcel M. Analytical features in qualitative analysis. TrAC Trends Anal Chem. 2005;24:477–87.

Coucke W, China B, Delattre I, Lenga Y, Van Blerk M, Van Campenhout C, et al. Comparison of different approaches to evaluate External Quality Assessment Data. Clin Chim Acta. 2012;413:582–6.

Coucke W, Van Blerk M, Libeer J-C, Van Campenhout C, Albert A. A new statistical method for evaluating long-term analytical performance of laboratories applied to an external quality assessment scheme for flow cytometry. Clin Chem Lab Med. 2010;48:645–50.

Coucke, W., Charlier, C., Lambert, W., Martens, F., Neels, H., Tytgat, J., ... & Verstraete, A. G. (2015). Application of the characteristic function to evaluate and compare analytical variability in an external quality assessment scheme for serum ethanol. Clinical chemistry, 61(7), 948-954.

Duewer DL. A comparison of location estimators for interlaboratory data contaminated with value and uncertainty outliers. Accreditation Qual Assur. 2008;13:193–216.

Ehrmeyer S, Laessig R. Alternative statistical approach to evaluating interlaboratory performance. Clin Chem. 1985;31:106–8.

Literature

Ellison SLR, Fearn T. Characterising the performance of qualitative analytical methods: Statistics and terminology. TrAC Trends Anal Chem. 2005;24:468–76.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The annals of Statistics*, 70-84.

Langton SD, Chevennement R, Nagelkerke N, Lombard B. Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. Int J Food Microbiol. 2002;79:175–81.

Ram B. J. A recursive version of Grubbs' test for detecting multiple outliers in environmental and chemical data. Clin Biochem. 2010;43:1030–3.

Schilling, E. G., & Neubauer, D. V. (2017). *Acceptance sampling in quality control*. CRC Press.

Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components* (Vol. 391). John Wiley & Sons.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 97-99.