In the Final Analysis

The 2006 L S Theobald Lecture

delivered at the University of Plymouth
on 03/05/06 by

Michael Thompson

Birkbeck College
University of London
m.thompson@bbk.ac.uk

The three essentials of quality

- What accuracy does the customer ***NEED***?
  Fitness for purpose (Decision theory)
- What accuracy ***CAN*** I achieve?
  Single laboratory validation
  Collaborative trials
- What accuracy ***DO*** I achieve?
  Internal quality control
  Proficiency testing

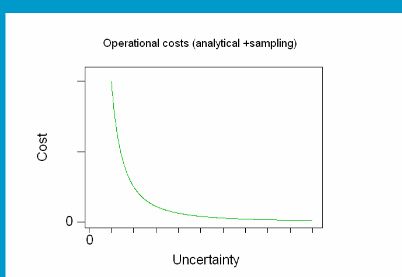
Three issues relating to quality

- Fitness for purpose (What is it?)
- Statistics (Can we do it?)
- Metrology (Do we need it?)

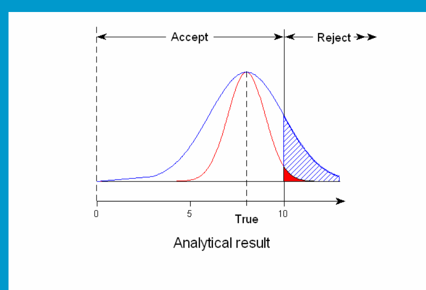
Fitness for purpose

- A result is fit for purpose when it maximises its expected utility.
- This means roughly that we need to minimise expected costs in the long term.
- There are operational costs of sampling and analysis.
- There are potential costs resulting from incorrect decisions based on the result.
- Both of these costs depend on uncertainty.

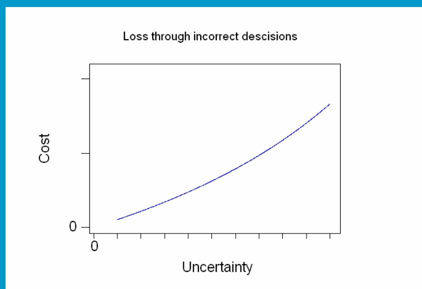
"Thompson's rule" $L \propto 1/u^2$



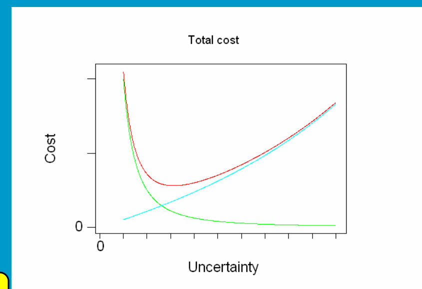
Cost of incorrect decisions—probabilities of false rejection



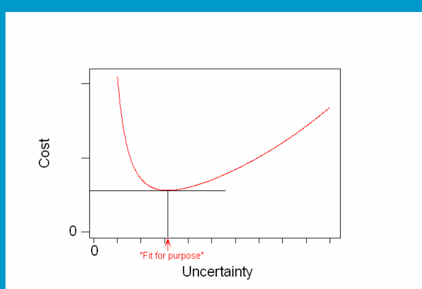
Typical loss function



Long-term loss



Fit-for-purpose uncertainty



Balancing sampling and analytical uncertainties

$$u = \sqrt{u_{sam}^2 + u_{an}^2} \qquad u_{sam} = u_{an}\ ?$$

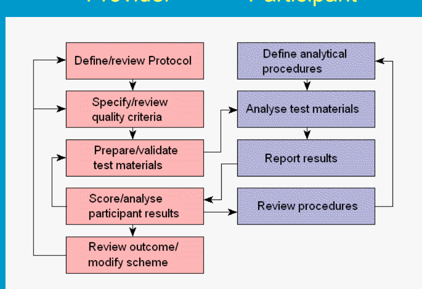
$$\frac{u_{an}}{u_{sam}} = \left(\frac{L_{sam}}{L_{an}}\right)^{1/4}$$

L_{sam}, L_{an}, are unit costs for a given uncertainty.

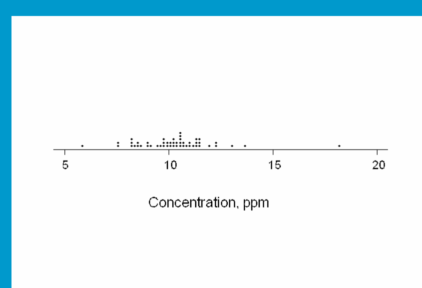
Fearn et al, *Analyst*, 2002, **127**, 818-824.
AMC Technical Brief No. 20.

Proficiency tests - organisation

Provider Participant



Participants' raw results



The z-score

$$z = (x - x_a)/\sigma$$

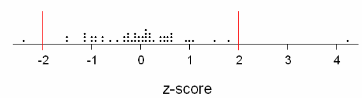
x = participant's result;

x_a = the "assigned value", the scheme provider's best estimate of the true value;

σ = the "standard deviation for proficiency", a scaling factor.

Using simple statistics for x_a and σ

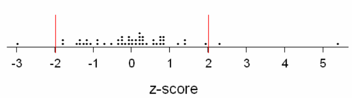
$$z = \left(x - \bar{x}\right)/s$$



96% of z-scores within the range -2<z<2

Using robust statistics

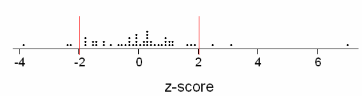
$$z = \left(x - \hat{\mu}_{rob}\right)/\hat{\sigma}_{rob}$$



94% of z-scores within the range -2<z<2

Using a fitness-for-purpose criterion σ_f

$$z = \left(x - \hat{\mu}_{rob}\right)/\sigma_f$$



88% of z-scores within the range -2<z<2

Statistics—some 'advanced' methods useful for analytical scientists

- Robust methods.
- Test for "sufficient homogeneity".
- Kernel densities.
- Maximum likelihood (mixture models etc).

Robust methods

- The statistics (*e.g.*, mean and standard deviation) are defined by an ***algorithm*** (a process), not by an equation.
- A commonly used robust method for the estimation of mean and standard devation is "Huber's H15" .

  Details and references can be found in:
  AMC Technical Brief No. 6.
  Analyst, 1989, **114**, 1689 and 1693.

Huber's H15

$$\mathbf{x}^{\mathrm{T}} = [x_1 \quad x_2 \quad \cdots \quad x_n]$$

Set $1 < k < 2, \quad p = 0, \quad \hat{\mu}_0 = \text{median}, \quad \hat{\sigma}_0 = 1.5 \times \text{MAD}$

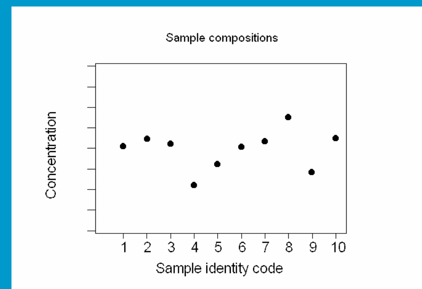
$$\tilde{x}_i = \begin{cases} x_i & \text{if } \hat{\mu}_p - k\hat{\sigma}_p < x_i < \hat{\mu}_p + k\hat{\sigma}_p \\ \hat{\mu}_p - k\hat{\sigma}_p & \text{if } x_i < \hat{\mu}_p - k\hat{\sigma}_p \\ \hat{\mu}_p + k\hat{\sigma}_p & \text{if } x_i < \hat{\mu}_p + k\hat{\sigma}_p \end{cases}$$

$$\hat{\mu}_{p+1} = \text{mean}(\tilde{x}_i)$$

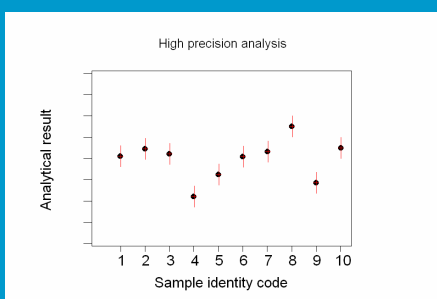
$$\hat{\sigma}^2_{p+1} = f(k)\,\text{var}(\tilde{x}_i)$$

If not converged, $p = p + 1$

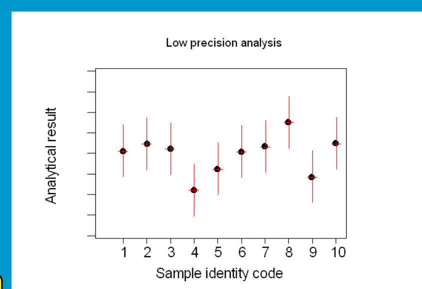
Testing for heterogeneity



Heterogeneity detected



Heterogeneity not detected



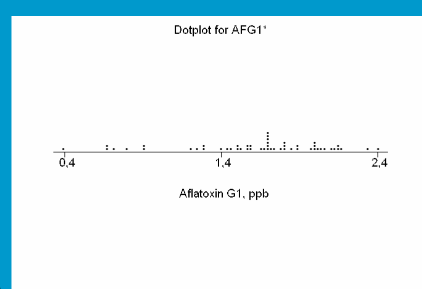
Testing for heterogeneity

- The problem—what is the correct analytical precision?
- The solution—don't test for significant heterogeneity, test for 'sufficient homogeneity'.

$$H_0 : \sigma^2_{sam} = \sigma^2_f \qquad H_A : \sigma^2_{sam} \le \sigma^2_f$$

Fearn and Thompson, *Analyst*, 2001, **126**, 1414-1417.
AMC Recommendation No. 1

Graphical representation of sample data



The normal kernel density

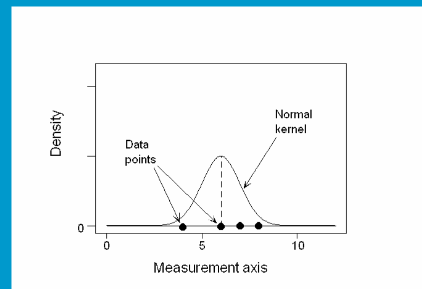
$$y = \frac{1}{nh}\sum_{i=1}^{n}\Phi\left(\frac{x - x_i}{h}\right)$$

where Φ is the standard normal density,

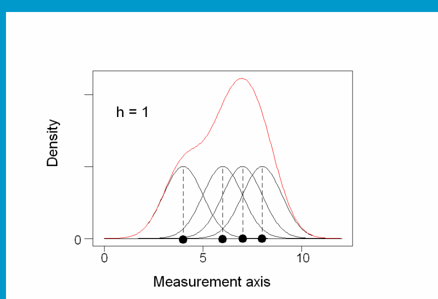
$$\Phi(a) = \frac{\exp(-a^2/2)}{\sqrt{2\pi}}$$

AMC Technical Brief No. 4

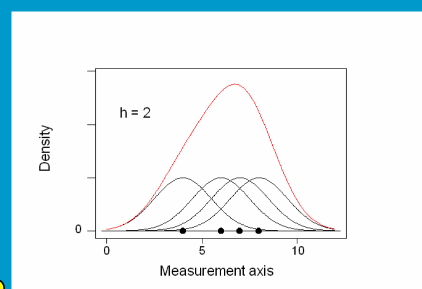
A normal kernel



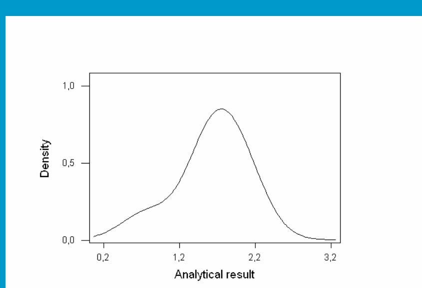
A kernel density



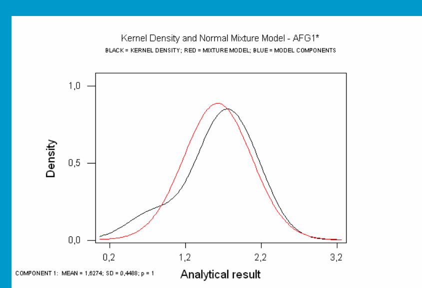
Another kernel density



Kernel density of the aflatoxin data



"Fit" of normal model



The normal mixture model

$$f(y)=\sum_{j=1}^{m} p_j f_j(y), \quad \sum_{j=1}^{m} p_j = 1$$

$$f_j(y)=\frac{\exp(-(y-\mu_j)^2/2\sigma^2}{\sqrt{2\pi}\sigma}$$

AMC Technical Brief No 23, and *AMC Software*. Thompson, *Acc Qual Assur*, 2006, **10**, 501-505.

Mixture models found by the maximum likelihood method (the EM algorithm)

- The M-step

$$\hat{p}_j = \sum_{i=1}^{n} \hat{P}(j|y_i)/n$$

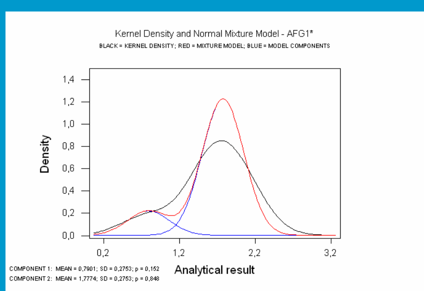
$$\hat{\mu}_j = \sum_{i=1}^{n} y_i \hat{P}(j|y_i) \Big/ \sum_{i=1}^{n} \hat{P}(j|y_i)$$

$$\hat{\sigma}^2 = \sum_{j=1}^{n}\sum_{i=1}^{m}\left((y_i-\hat{\mu}_j)^2 \hat{P}(j|y_i)\right) \Big/ \hat{P}(j|y_i)$$

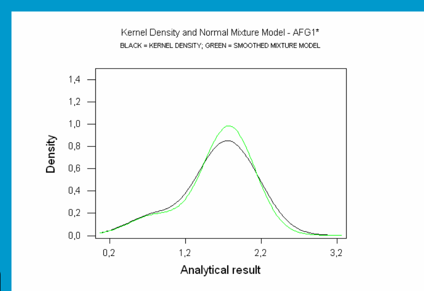
- The E-step

$$\hat{P}(j|y_i) = \hat{p}_j f_j(y_i) \Big/ \sum_{j=1}^{m} \hat{p}_j f_j(y_i)$$

Kernel density and fit of 2-component normal mixture model



Kernel density and variance-inflated mixture model



Find out more?

AMC Technical Briefs and Software on

www.rsc.org/amc/

Statistics

- Lies, damned lies, and statistics!

Metrology

- Fiction, science fiction, and metrology!

Metrologist's creed

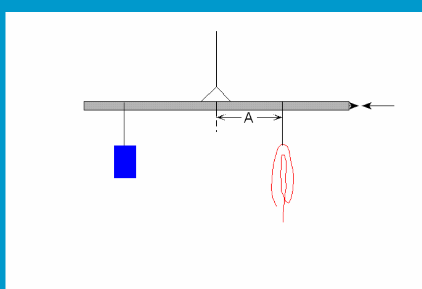
- Uncertainty is important.
- Analytical chemists are not good at estimating uncertainty.
- All results of chemical measurement are traceable to SI units, in particular the mole, the kilogramme, the metre.
- Analytical chemists don't worry about traceability, that's why their results are questionable.

Metrological false premise 1

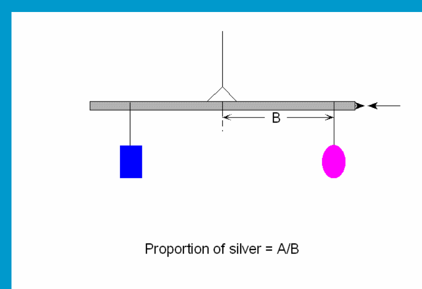
- All analytical results are traceable to SI units, in particular the mole, the kilogramme, and the metre.

- *NO!* The majority of analytical measuremnts made for commercial purposes are mass fractions, not traceable to *any* unit.
  Corollory: expressions such as %, ppm, ppb, etc are perfectly correct.

False premise No 1 contd. – Silver content of silver solder



False premise No 1 contd. – Silver content of silver solder



False premise No 1 contd. – Silly or what!

- Is the concentration of silver, A/B, traceable to the metre ?
- Should we express the result as (say) 70 cm m^{-1} ?
- Or 700 mg g^{-1} (when no mass standard is involved)?

Metrological false premise 2

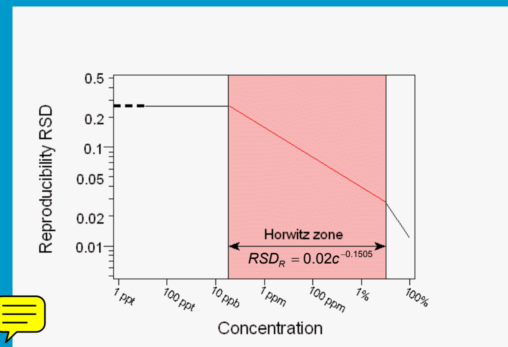
- Chemical measurement results are not accurate enough, and that is because of a lack of traceability to SI units.

- *NO!* Most chemical measurement results are fit for purpose or more accurate.
- Where results are not accurate enough—it sometimes happens—the shortfall is often irreducible and traceability to SI units does not help.

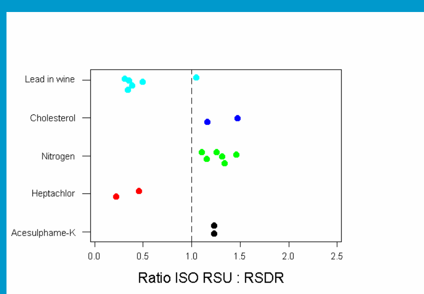
Metrological false premise 3

- Uncertainty is under-estimated by interlaboratory studies: only “bottom-up” models with clear traceability to SI units give the correct answer.
- *NO!* When proper comparisons are made, we mostly find that (say) reproducibility standard deviations from collaborative trials give equal or greater uncertainty estimates than “bottom-up”.

Reproducibility relative standard deviations

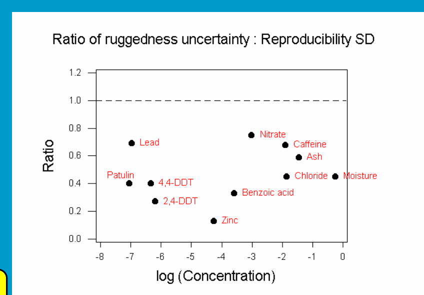


“Top-down versus bottom-up”

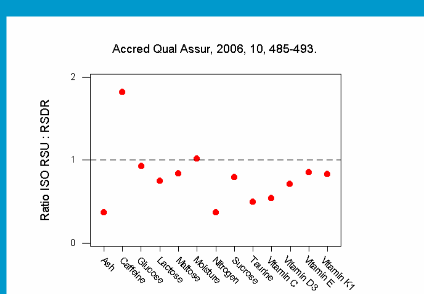


Thompson *et al* data

Analyst, 2002, **117**, 1669-1675.



Populaire & Gimínez data



Metrological false premise 4

- Chemical measurements have a larger relative uncertainty in comparison with most physical measurements. (True)
- That is because they are not traceable to SI units.
- *NO!* The traceability chain to SI units contributes almost nothing to the combined uncertainty of analytical results.

Metrological false premise 4 contd.

- Realistic relative uncertainties in analytical results are mostly in the approximate range 1-30%.
- Relative uncertainties in tranferring SI units (such as mass and volume) to the analytical laboratory bench are less than 0.1%.

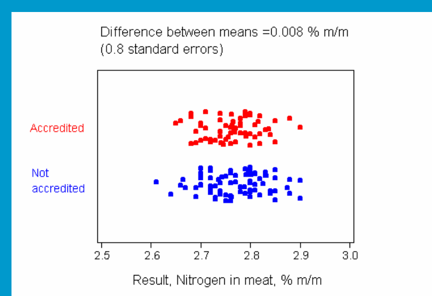
Metrological false premise 5

- Terms such as “true value”, “trueness”, and “bias” have no proper place in metrology (because we can’t know them).
- *NO!* “True value” (and its dependent terms) are readily defined.
- The whole of statistics is based on the idea of unknown population values, a concept logically isomorphic with “true value”.

Metrological false premise 6

- Only accredited laboratories can produce reliable results.
- ***No!*** Evidence from proficiency tests contradicts this idea.

Metrological false premise 6



Metrological false premise 6