The AAFCO Proficiency Testing Program Statistics and Reporting

Program Chair: Dr. Victoria Siegel
Statistics and Reports: Dr. Andrew Crawford

Contents

The AAFCO Proficiency Testing Program

Program Model

Return To Contents

Program Model

Based Closely On:

"The International Harmonized Protocol for the Proficiency Testing of Analytical Laboratories", 2006 (IHP), **MICHAEL THOMPSON, STEPHEN L. R. ELLISON AND ROGER WOOD**

- AMC supported (Analytical Methods Committee of the RSC)
- Uses ISO statistical models - ISO 13528, 2005 and ISO 5725-2, 1994
- Robust statistics used as described in the IHP and ISO 13528
- Duplicate analysis supports method precision calculations.
- Proficiency testing often required for Laboratory Accreditation.
- Independent documentation on how it all works.
- Makes full use of Web based data transfer.

To view a pdf version of the IHP click here.

Post Chemical Analysis Data Flow for One Sample

Collect Data

- Web server based collection of analytical data from Labs.
- Lab submits duplicate analysis and method used for each analyte run.
- Preliminary data review by Chair.
- Chair delivers raw data for statistical review.

Statistical Review

- ❑ Screen for poor duplicates, extreme outliers and data distribution shape.
- ❑ Perform Robust Stats calculations for individual methods and group analytes.
- ❑ Establish Consensus Values, Robust SD’s and Uncertainties.
- ❑ Calculate Z scores and supporting Stats.
- ❑ Calculate method precision parameters.
- ❑ Expert review to handle anomalies.

Report to Labs

- Create report cards, general reports and Sample run reports. Report run observations and provide all reports to chair.
- Deliver Reports for Web based distribution to labs.

Proficiency Testing

From the IHP, 2006

1.1 Rationale for proficiency testing

- For a laboratory to produce consistently reliable data, it must implement an appropriate program of quality-assurance and performance-monitoring procedures. **Proficiency testing is one of these procedures**.

2.10 Choice of analytical method by participant

- Participants shall **normally use the analytical method of their choice**. In some instances, however, for example, where legislation so requires, participants may be instructed to use a specific documented method.

Proficiency Testing

Data Analysis

- Use Robust Statistics to estimate Consensus Value and fit-for-purpose sigma (σ_{rob}).
- Mean of Lab duplicates used for Robust statistics.
- The different methods used for a single analyte can be grouped and used for true **Proficiency Testing**.
- Individual methods are still handled separately and called **Proficiency Testing for Individual Methods**.
- Duplicates are required to calculate individual method precision for each Sample run.

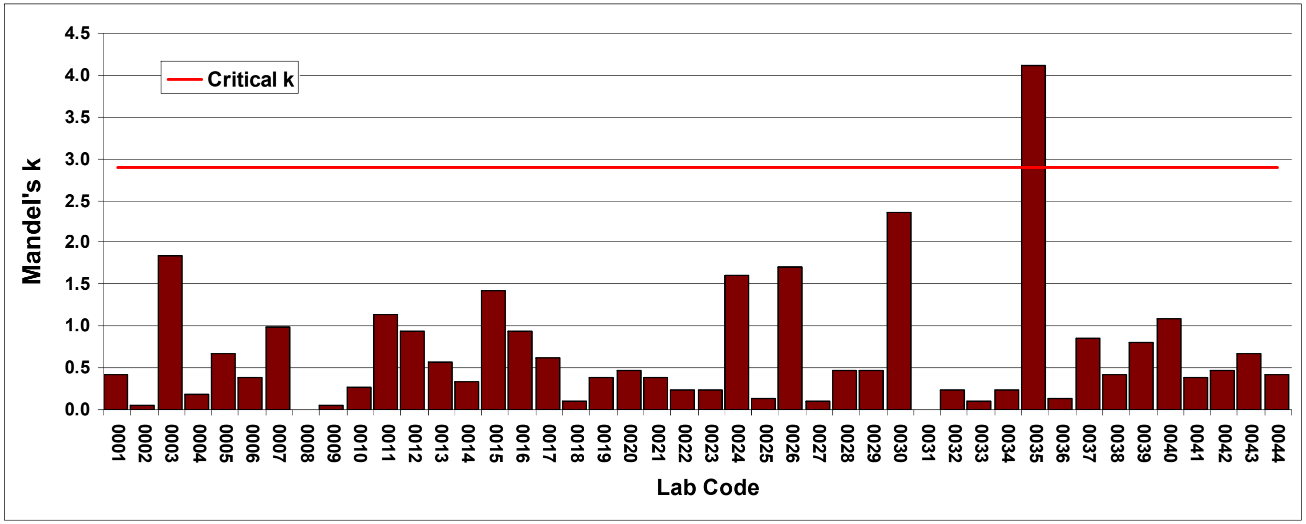
Data Pre-Screening and Just Looking at the Data

Tools to Identify and Remove the Clearly Bad Data

Return To Contents

Mandel's k to Flag for Duplicates Too Far Apart (k_{crit} set at $\alpha = 0.0025$)

$k_i = \frac{S_i}{S_r}$ ≡ A ratio of the i^{th} Lab SD to the within Lab SD

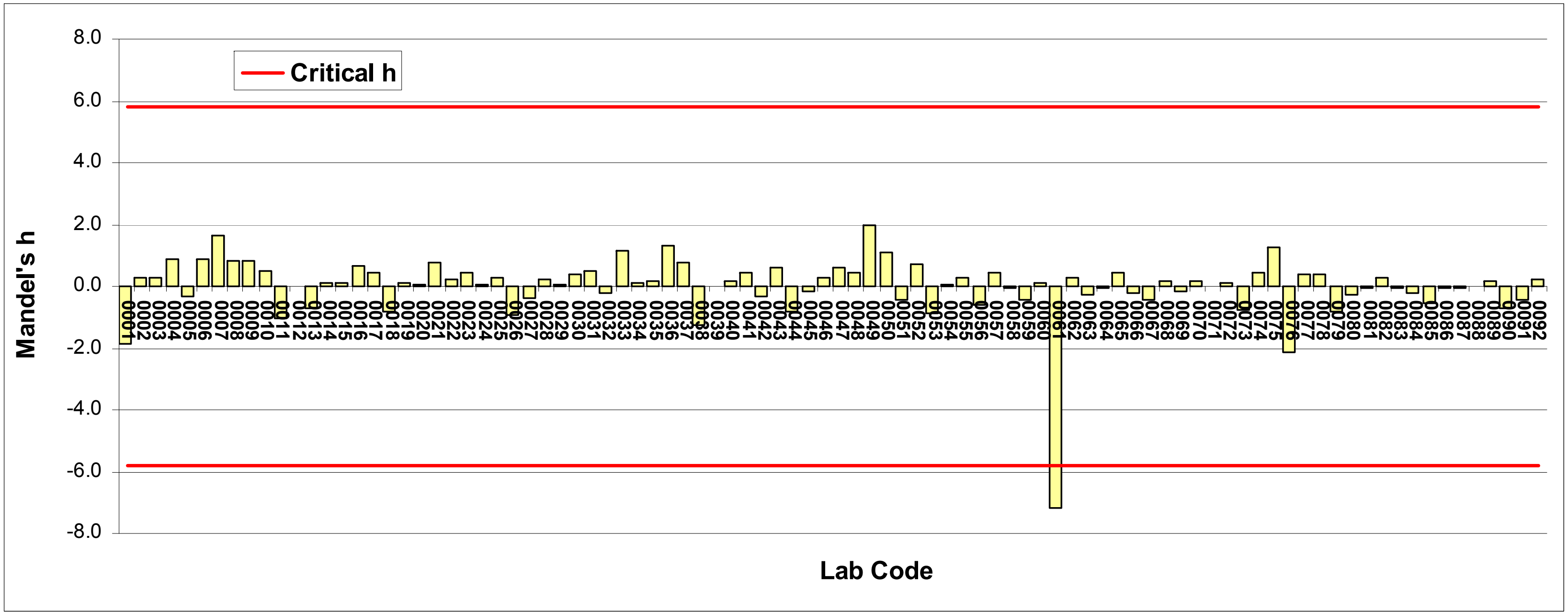


44 Labs perform Acid Detergent Fiber analysis in duplicate. The duplicates for Lab # 35 are too far apart. Data for this Lab may not be included in calculations.

Mandel’s h to Flag for Extreme Outliers
(h_{crit} set at α = 1.0E-10)

$$h_i = \frac{\overline{X}_i - \overline{\overline{X}}}{S_{\overline{X}}}$$

The difference between the ith Lab Mean and the Grand Mean as it relates to the SD of all the Lab means.



92 Labs perform Copper analysis. The value for Lab # 61 is extremely different from the other 91 values.
A review of this data did exclude it from Robust calculations.

AAFCO
Association of American Feed Control Officials

View Data Distribution Shape
Kernel Density Plots

$$f(X,h)=\frac{1}{nh}\sum_{i=1}^{n}\Phi\left(\frac{X-X_i}{h}\right)$$

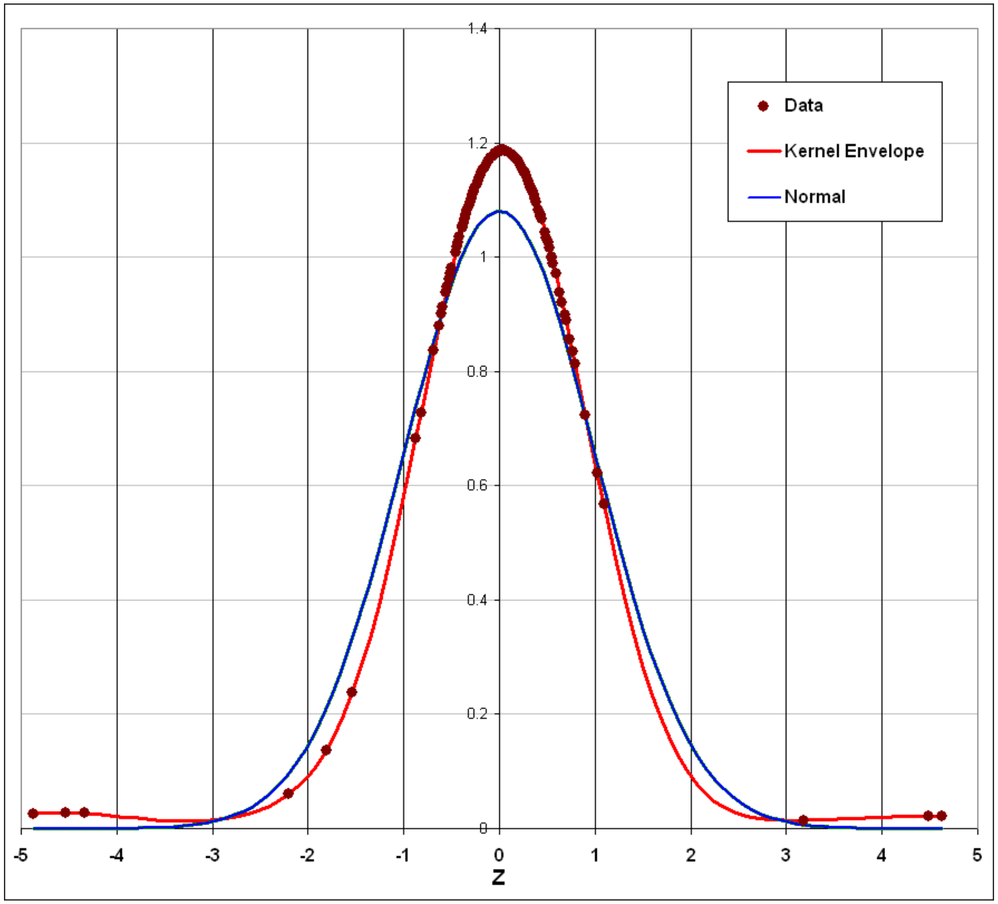
IHP recommended bandwidth, $h=0.75\times\sigma_n$

$\Phi=$ Standard Normal density function.

For a more complete description of how a Kernel Density Plot is formed from the summation of all the “Normal kernels” please click here.

150 Labs perform Crude Ash analysis. Here you can see the Lab means (**Brown**) distributed on a Kernel Density Plot (**Red**) compared with the Normal curve for this data (**Blue**).

This Kernel Density plot compares quite well with the shape of a Normal curve



AAFCO Proficiency Testing Program

Calculating Robust Statistics

“The International Harmonized Protocol For The Proficiency Testing Of Analytical Chemistry Laboratories”, 2006

“ISO 13528 Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons”, 2005 – Algorithm A

Return To Contents

Why Robust Statistics?

- Most “real world” data distributions do not follow the Normal Gaussian Model, they are more like “contaminated” Normals.
- Distributions have “Fat Tails” and Outliers that skew the Mean and inflate the Standard Deviation (Normal estimators are very sensitive!).
- Even Outliers contain information.
- We need a Robust estimate of the Location of the data center.
- We need a Robust estimate of the data Dispersion.
- We need to identify and weight the “Reliable” data.

John Tukey, Peter Huber and Frank Hampel credited with founding the discipline.

All since Tukey’s landmark paper in 1960

Tukey, J. W. (1960). “A survey of sampling from contaminated distributions.”

Calculating Robust Statistics

Starting values are calculated:
Robust Mean (X*) = Median(LAB()) {Median value of All Lab X's}

Robust Std (S*) = 1.483 * MAD(LAB()) {MAD of Lab deviations}
δ = 1.5 S*
The dataset is calculated for each i:

$$x_i^* = \begin{cases} \mathbf{x^* - \delta \ \text{when } x_i < x^* - \delta} \\ \mathbf{x^* + \delta \ \text{when } x_i > x^* + \delta} \\ \mathbf{x_i \quad \text{otherwise}} \end{cases}$$

The Median is a Robust measure of Location.

The Median Absolute Deviation (MAD) is a Robust measure of Dispersion.

Intermediate values for x* and s* are calculated for the next iteration as follows:

$$x^* = \frac{\sum_{i=1}^{P} x_i^*}{P} \qquad s^* = 1.134\sqrt{\frac{\sum_{i=1}^{P}\left(x_i^* - x^*\right)^2}{(P - 1)}}$$

Where P is # Labs.

When there is no further change in the 6th decimal place of X* the iteration is stopped and the following values are assigned:
Robust Assigned Value (X_{rob}) = X*
Robust Fit for Purpose Sigma (SD_{rob}) = S*

Uncertainty in the Assigned value (U_{rob}) is calculated as follows:

$$U_{rob} = 1.25 \times \frac{SD_{rob}}{\sqrt{P}}$$

Graphical Analysis Review

Data points (**Red**) on Kernel Density Envelope.

Normal Curve (**Grey**)

Winsorizing Squeezes outer Data Points In (**Green Points**)

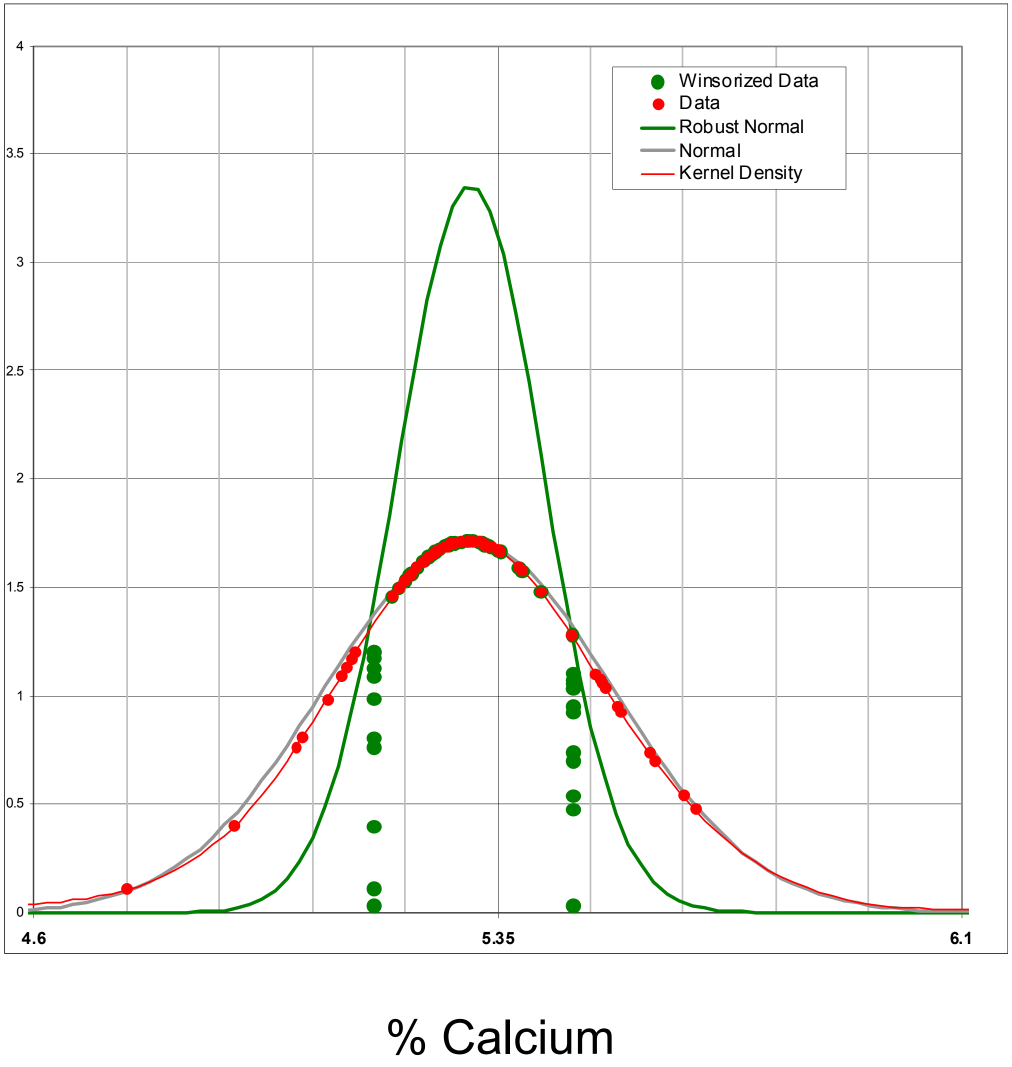
A Robust Normal Is Calculated (**Green Curve**)

The Robust curve provides a better estimate of the location of the mean.

In this case the dispersion is reduced to better represent the “reliable” Normal data in the dataset.

σ_{rob} provides a more realistic fit-for-purpose measure of dispersion.

Example 135 Labs run % Calcium Analysis



QQ Plots are created for each analyte method in a Sample run.

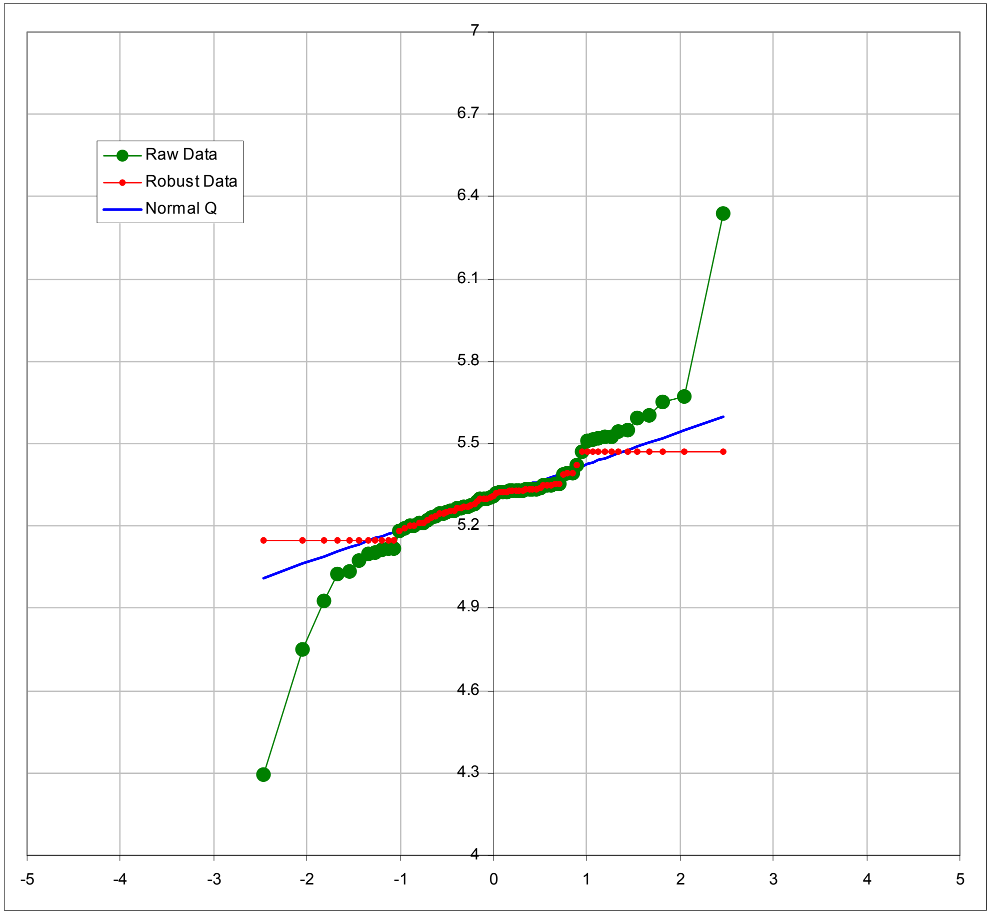
Graphical Analysis Review

% Calcium quantiles (y axis) are plotted against Normal scores (x axis) for the ordered data (**green**). The Winsorized data (**red**) and standard Normal (**blue**) are plotted on the same chart.

The “reliable” data for a Normal distribution exists where the 3 curves overlap.

The effect of Winsorizing clearly shows how the data in “fat tails” is drawn into the standard Normal.

Example: 135 Labs run % Calcium Analysis



In summary:

from the Huber H15 Process we now have:

- An Assigned Value X_a (robust measure of location). This is a participant Consensus Value.
- A “fit for purpose” σ_{rob} standard deviation (robust measure of dispersion) based on participants in the round.
- An estimate of uncertainty in the assigned value U_a.

Proficiency Testing Program

Z Statistics & Fitness for Purpose

Return To Contents

Calculating a Z Score

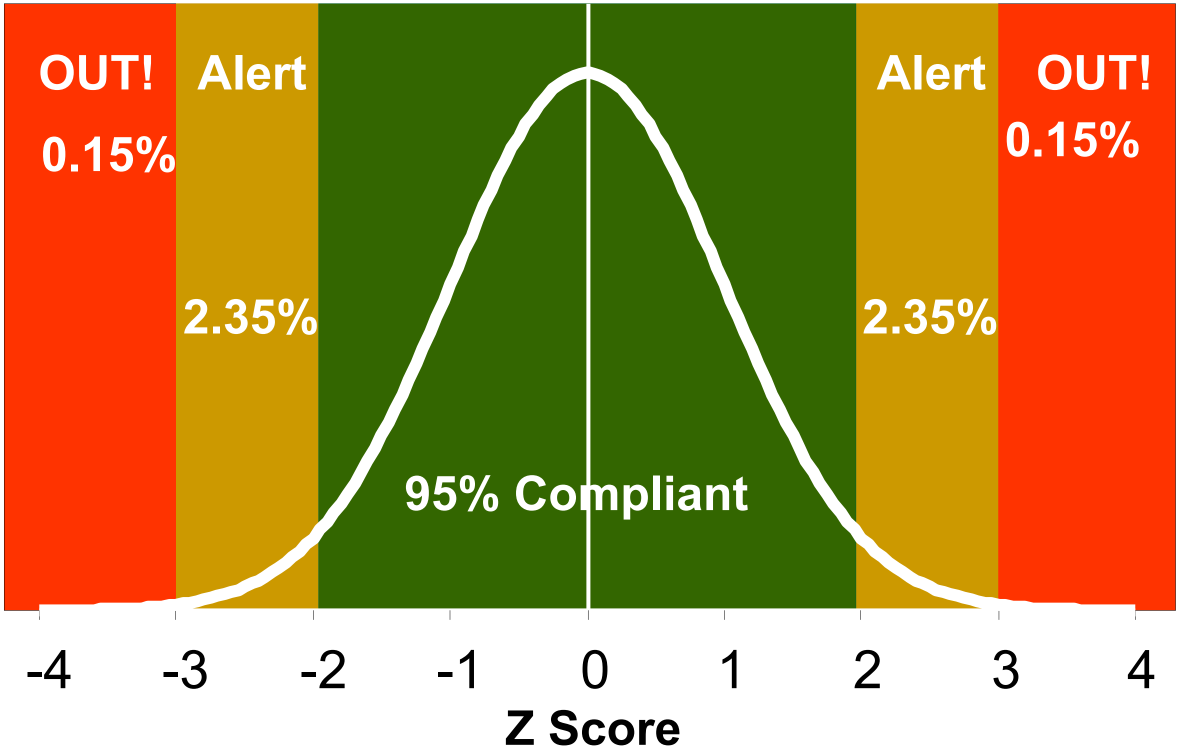
This is the classical Z score where we expect about 95% of the participants to fall between ± 2 and 99.7% to fall between ± 3.

Robust statistics will usually cause slightly fewer labs to fall within accepted limits.

$$Z = \frac{X_{LAB} - X_{a}}{\sigma_{rob}}$$

This is fine if you want to score yourself against the other participants in that round.

Interpreting Z Scores for Proficiency Testing



Red indicates a normally distributed Z value >3 or <-3 and usually requires action. About 0.3 % fall in this range. **Orange** indicates a Z score between 2 and 3 or -2 and -3. This is a warning and roughly 4.7 % lie in this region. **Green** indicates a Z score < 2 and >-2 and is considered in compliance.

Calculating A Proficiency Z Score That is Fit-For-Purpose (σ_{ffp})

We can calculate a Normally distributed 0 centered Z score using the σ_{ffp} based on %RSD or other pertinent sigma rather than σ_{rob} derived from participants in that round.

$$Z = \frac{X_{LAB} - X_a}{\sigma_{ffp}}$$

If you wish you can substitute your own fit-for-purpose standard deviation (σ_{ffp}) to obtain an appropriate Z score.

Calculating A Proficiency Z Score Based on % RSD as Fitness For Purpose (σ_{ffp})

It may be more important to your client, a regulatory agency, a legal position or even to you that you are compliant to a predetermined level.

So we establish a “Fitness for Purpose” sigma to reflect this predetermined level (ie: %RSD).

$$\sigma_{ffp} = X_a \times \frac{\%RSD}{100}$$

Calculating a Threshold %RSD
Which is Independent of the variability in the run

$$Z = \frac{X_{LAB} - X_a}{\sigma_{ffp}}$$

Substituting for σ_{ffp}

$$Z = \frac{X_{LAB} - X_a}{\frac{\%RSD}{100} \times X_a}$$

Substituting Z = 2

$$\%RSD_{Z=2} = \frac{\left|X_{LAB} - X_a\right|}{2 \times X_a} \times 100$$

The %RSD is the relative standard deviation as a percent of the Assigned value and is a popular way to express variability. We cater to well over 300 labs in several different countries with different client, legal and regulatory requirements . Consequently there is no single fit-for-purpose sigma (σ_{ffp}) we can realistically report.

We offer the Threshold %RSD as a single fit-for-purpose parameter that can be compared with the individual requirements of your lab.

Fitness for Purpose Examples

This Table demonstrates some of the dilemmas that can arise if you rely solely on Z scores derived from participant variation in the round and how the Threshold %RSD can alert you to the problem. The Table shows Z scores for five analytes at six different between lab %RSD's (1% to 50%), the corresponding AAFCO Z score using σ_{rob} and the Threshold %RSD at Z = 2 where n is the number of participating Labs.

For example, 140 labs perform Protein analysis by N_2 combustion and you receive your Z score of **-3.02**. This is quite disturbing and could possibly trigger some action. The Table below indicates that you become Z compliant somewhere between 1% RSD and 2% RSD of the assigned value. This is acceptable to you and your client. So, just because 95% of the participants generated compliant Z scores does not necessarily mean your result is unacceptable.

Conversely, for a different sample 28 labs run Fiber analysis using the Fibertec system. This example shows that blindly accepting the compliant Z score of **-1.02** could actually represent a 33% discrepancy from the Assigned value.

	Z Scores Based on % RSD Fitness for Purpose (σ_{ffp})							
AAFCO CS Z Score (σ_{rob})	**% RSD (σ_{ffp})**	**1%**	**2%**	**5%**	**10%**	**20%**	**50%**	**Threshold %RSD$_{(Z=2)}$**
-3.02	**Protein (N_2 Comb. n = 140)**	-3.08	-1.54	-0.62	-0.31	-0.14	-0.06	**1.5%**
-6.12	**Crude Ash (n = 110)**	-5.18	-2.59	-1.04	-0.52	-0.26	-0.10	**2.6%**
-1.64	**Copper (ICP, n = 30)**	-16.46	-8.23	-3.29	-1.65	-0.82	-0.33	**8.2%**
-1.42	**Vitamin A (HPLC, n = 17)**	-32.50	-16.25	-6.50	-3.25	-1.63	-0.65	**16%**
-1.02	**Fiber (Fibertec, n = 28)**	-65.75	-32.87	-13.15	-6.57	-3.29	-1.13	**33%**

In summary we now have:

- A Check Sample Z Score where Red indicates a normally distributed value >3 or <-3 and requires action. An Orange value between 2 and 3 or -2 and -3 provides a warning and a Green value < 2 and >-2 indicates compliance and is within 95% of the other Lab values.
- A Threshold %RSD which provides a personalized operating parameter for your Lab. This parameter is dependant only on your bias from the assigned value and not on the variability of the other labs and is designed to help address the “Fitness for Purpose” concerns of the IHP.

  For example, if your Threshold %RSD is 3% then you are in compliance with a minimum threshold of 3% RSD at Z = 2 (95%).

Method Precision Data from The AAFCO Check Sample Program

Statistical Model Based on ISO 5725-2 Accuracy (Trueness and Precision) of Measurement Methods and Results, 1994

For more information click on link - Methodology of Inter-comparison Tests and Statistical Analysis of Test results – Nordtest project No. 1483-99, 2000

Return To Contents

Outliers and Poor Duplicates

- Mandel’s h for Outliers (h_{crit} set at $\alpha = 0.01$)
- Mandel’s k for Duplicates Too Far Apart (k_{crit} set at $\alpha = 0.01$)

Precision

- The closeness of agreement between independent test results obtained under stipulated conditions.
- Dependent on the distribution of random errors.
- Repeatability and Reproducibility are 2 commonly defined stipulated conditions.
- We quantify precision by measuring:

Between Labs SD (s_L)
Repeatability SD (s_r) ≡ Within Labs SD
Reproducibility SD (s_R) ≡ Combined Variance

Computational formulas for calculating critical precision variances.

$$s_L = \sqrt{\left[\frac{n\left(\sum_{i=1}^{n} X_{LAB(i)}^{2}\right) - \left(\sum_{i=1}^{n} X_{LAB(i)}\right)^{2}}{n(n-1)}\right] - \frac{s_r^2}{2}}$$

Between Labs SD (s_L)
Repeatability SD (s_r) ≡ Within Labs SD
Reproducibility SD (s_R) ≡ Combined Variance

$$s_r = \sqrt{\frac{\sum_{i=1}^{n} (X_{i1} - X_{i2})^2}{2n}}$$

$$s_R = \sqrt{s_L^2 + s_r^2}$$

Example Method Precision Report

Method # 019.00
Calcium, Ox-MnO_4 Vol.
Sample # 201321
Dry Dog Food

- The method precision report includes Robust parameters for comparison (**Green**).
- Here we see 3 labs removed from calculations after examining Mandel statistics.
- The Robust measure of location has shifted to 1.34 % from 1.32 % and the Robust measure of dispersion is substantially reduced (0.14 to 0.09).
- A Robust measure of the uncertainty in the assigned value is provided.
- The Between, Within and Reproducibility standard deviations and CV’s are given (**Blue**).
- The S_R/S_r ratio of 5.9 is somewhat higher than the ~ 3 expected for ordinary lab bias.
- The average range is usually a very good estimate of S_r.
- The Horwitz %RSD for the Assigned value is given based on the 0.1505 exponent. Along with the assigned value this can be used to determine a σ_{ffp} if desired.

Total # Labs Submitting	**26**
# Labs Included in Calculations	**23**
Mean	**1.316**
SD	**0.137**
Assigned Value - Robust Mean	**1.340**
AAFCO CS ffp - Robust sd	**0.089**
Uncertainty (U_a) - Robust	**0.013**
% RSD - Robust	**6.66%**
Between Labs s_L	**0.136**
Within Labs s_r	**0.023**
Reproducibility s_R	**0.138**
Between Labs %RSD	**10.34%**
Within Labs %rsd	**1.78%**
Reproducibility %RSD	**10.50%**
s_R/s_r	**5.907**
Average Range (R-bar)	**0.026**
Horwitz %RSD	**3.83%**

In summary we now have:

- A measure of the between labs variability (S_L)
- A measure of the within labs variability (S_r) called repeatability.
- A combined measure of the variability (S_R) called reproducibility.
- Using these standard deviations and the ordinary (non robust) mean of the dataset we can calculate the corresponding %rsd's which are very useful for comparing variability in samples with different analyte concentrations.
- If we look at S_R/S_r we create a new parameter which describes the between lab variability in terms of the within lab variability. Large ratios indicate possible lab generated method bias. Small values ~ 3 are indicative of the expected lab bias.
- The monthly Method report also includes other summary information as well as the average range for duplicates (R-Bar).