Understanding Proficiency Testing Statistical Analysis and Evaluation

Presented by Craig Huff

Statistical Analysis and Evaluation of Population Data- Commonly Utilized Models

- NELAP (TNI)*
- Z- Scoring
- Pure Study/Population Approach

* Most recognized in US and focus of today's presentation

NELAP (TNI)

- Accepted by most states – wholly or in part
- Data evaluated using regression equations or fixed limits
- Information published in the TNI FoPT tables
- NPW Acceptance Limits represent ~ ±3 standard deviations
- DW Acceptance Limits represent ~±2 standard deviations
- Utilizes robust mean & robust standard deviation

Mean and Standard Deviation- Robust vs Arithmetic Techniques

- Arithmetic = Simple "Average" and "Standard Deviation"
  - Used for sample sizes of 7 to 20 samples
- Robust = Multi-iterative, bi-weighted[1] mean and standard deviation
- What does "bi-weighted" really mean?
  - Begins with the median of the data population
  - Assigns a weighting factor to each data point with each iteration based on "distance" from the median
  - 15 iterations conducted
  - Utilized for sample sizes of 20 or more values
- Why use Robust technique?
  - Minimizes the effect of data outliers on the mean and standard deviation

1. "A Bi-weight Approach to the One-Sample Problem"- Dr. Karen Kafadar

Outliers – Determination and Treatment

- Grubb's Test (Grubbs 1969 and Stefansky 1972) is used to detect a single outlier in a univariate data set that follows an approximately normal distribution.

Grubbs' test is defined for the hypothesis:

H0: There are no outliers in the data set

Ha: There is exactly one outlier in the data set– (*multiple iterations may be conducted*)

Test Statistic: The Grubbs' test statistic is defined as:

$$G=\frac{\max|Y_i-\bar{Y}|}{s}$$

with $\bar{Y}$ and s denoting the sample mean and standard deviation, respectively. The Grubbs' test statistic is the largest absolute deviation from the sample mean in units of the sample standard deviation.

Http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h1.htm

Note: Outlier testing is utilized only when Arithmetic techniques are used to determine population means and standard deviations.

Used for sample sizes of 7 to 20 samples- **No more than 20% of the values in a data set may be classified as outliers**.

PT Regression Equations vs. Fixed Limits: Where do they come from and how were they derived?

- Exist within the TNI FoPT Tables (excerpt from TNI NPW FoPT table below)

Matrix	EPA Analyte Code	NELAC Analyte Code	Analyte [1,2]	Conc Range	Acceptance Criteria [3,4,5,6] a	b	c	d	NELAC PTRL [7]
			Nutrients						
NPW	0031	1515	Ammonia as N	1.0 to 20	0.9923	0.0567	0.0583	0.0914	0.60
NPW	0032	1810	Nitrate as N	2.0 to 25	0.9975	-0.0005	0.0506	0.0642	1.50
NPW		1820	Nitrate-nitrite as N	2.5 to 25	0.9957	-0.0010	0.0509	0.0400	1.99
NPW		1840	Nitrite as N	0.4 to 4.0	1.0017	-0.0030	0.0377	0.0250	0.28
NPW	0033	1870	Orthophosphate as P	0.5 to 5.5	±15% fixed acceptance limit				0.42
NPW	0034	1795	Total Kjeldahl-Nitrogen [10f]	3.0 to 35	0.9701	0.2283	0.0680	0.1906	1.95
NPW	0035	1910	Total Phosphorus	0.5 to 10	0.9932	0.0084	0.0506	0.0254	0.35

- Developed using data from multiple PT providers, EPA and method specific data – multiple methods combined

How are PT Acceptance Limits derived from TNI Regression Equations

- Are your results "Acceptable" or "Not Acceptable"
- For NPW… Acceptance Limits are set at ± 3 Std Dev as calculated from the "predicted mean"

Eg. ***Nitrite as N***; a = 1.0017, b = -0.0030, c = 0.0377, d = 0.0250

Assume PT sample assigned value = 1.00 mg/L

Predicted Mean= (1.00*1.0017)+(-0.0300) = **0.999 mg/L**

Predicted Std Dev = (1.00*0.0377)+0.0250 = **0.0627 mg/L**

Acceptance Limits = 0.999 ±(3*0.0627) or **0.811 – 1.19 mg/L**

Note: Analytical method bias is accounted for where regression equations are prescribed.

Assigned Values—How are they determined?

- Actual "made-to" value as determined by weights/measurements (taking into account substrate purities).
- Measured means (established by the PT provider)
- PT Study mean (eg…where only "c" & "d" factors are supplied on the FoPT table)
- Must be compliant with Verification, Homogeneity and Sample Stability criteria (VHS)

Data Modality – What is it and how is it handled?

- Multi-modal distributions can occur where two or more data distribution scenarios are exhibited within a data set.
- Methods for detecting and treating these situations must be approved by the PT provider's Proficiency Testing Provider Accreditor (PTPA)
- When detected, the PT provider must assess the cause, segregate the data and evaluate separately…or invalidate the analyte/sample in that PT study

Some Potential Causes of Multi-Modality:

-Prep/analytical method bias (i.e., two or more methods may not be equivalent)

-PT sample(s) inhomogeneity (within and/or between the samples)

-PT sample(s) may have exhibited instability during the course of the study

PT Sample Concentration and it's Impact on Acceptance Limits

- Regression-based acceptance limits:

  Typically change as a percentage of the assigned value over the PT concentration range (generally widen as the concentration approaches the PTRL)

- Fixed acceptance limits…need I say more

Monitoring and Trending PT Performance-Tools You Can Utilize

- PT Performance and Exception Reports
  - Summarizes overall performance -by analyte or by comparison to study population
- Custom Export Generator
  - Define and save the data you want…when you want it
- Z-Scores (a powerful trending tool)
  - Know when you have opportunity for improvement—before you experience a "not acceptable" evaluation

Other Sources of Valuable Information

- TNI Website (FoPT Tables, Laboratory Accreditation, PT Program Info.)
  www.nelac-institute.org

- ISO 17025
- ISO Guide 34
- ISO Guide 43
- ERA
  www.eraqc.com

Thank You

Questions?