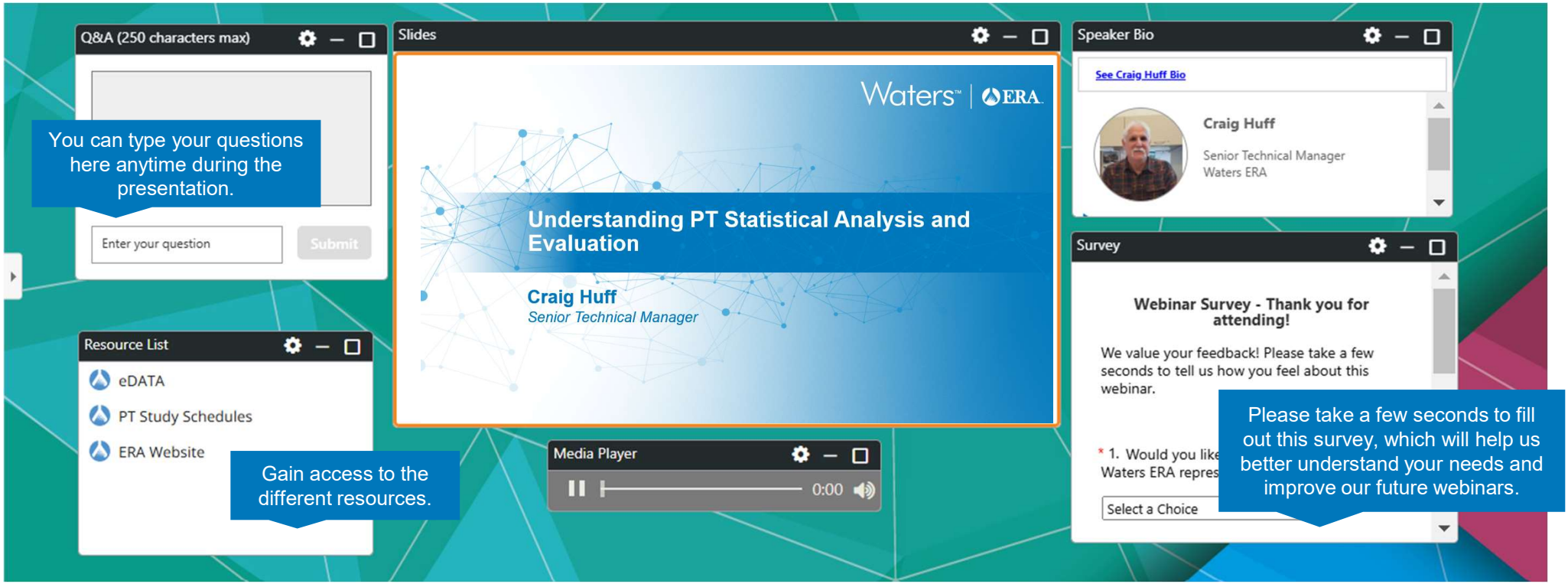
Understanding PT Statistical Analysis and Evaluation

Craig Huff
Senior Technical Manager

Webinar survey – your feedback is appreciated!



Understanding PT Statistical Analysis and Evaluation

- Key Learning Topics
  - Use of Z-scoring models
  - Peer study consensus approaches
  - Application of regression equations and fixed limits
  - Robust statistical techniques
  - Tools available for monitoring PT performance

- Speaker – Craig Huff
  - Senior Technical Manager

Outline

- Peer study consensus approaches
- Application of regression equation-based limits and fixed limits
- Assigned values
- Robust statistical techniques
- Proficiency testing reporting limits (PTRLs)
- Tools available for monitoring PT performance
- Multi-modality
- Z-scores a simple way to trend PT results

Statistical Analysis and Evaluation of PT Study Data- Commonly Utilized Models

- The NELAC Institute (TNI)*
- Z- Scores
- Study Consensus Approach

* Most recognized in US and primary focus of this presentation

Robust vs Arithmetic Statistical Techniques

Robust	Arithmetic
= Multi-iterative, bi-weighted[1] mean and standard deviation *Utilized for sample sizes of 20 or data points*	= "Simple" Average and "Simple" Standard Deviation *Used for sample sizes of 7 to 19 data points*

What is "bi-weighted[1]" and how is it calculated?

- Begins by calculating the median of the data set
- Assigns a weighting factor to each data point with each iteration based on "distance" from the median
- 15 iterations conducted to arrive at the robust mean and standard deviation

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

Waters™ | ERA

Grubb's Test[2] →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: $G=(\max |Yi-\bar{Y}|)/s$

Where $\bar{Y}$ and s denote 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.

- *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 19 samples- No more than 20% of the values in a data set may be classified as outliers.*

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

PT Regression Equations vs Fixed Acceptance Limits

- Prescribed by NELAC and reside within NELAC Fields of Proficiency (FoPT) tables
- Developed from historical data provided by multiple PT providers

Example of a TNI NPW FoPT table

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

How are PT Acceptance Limits Calculated?

- Example-> Using the regression equation for Nitrite as N from the previous slide
  - For NPW… Acceptance Limits are set at ± 3 “predicted standard deviations” as calculated from the “predicted mean”

Nitrite as N regression equation; a = 1.0017, b = -0.0030, c = 0.0377, d = 0.0250

Assume PT sample assigned (made-to) 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.

PT Sample Concentration, PTRLs and Impacts 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 low end of the prescribed concentration range)
- Fixed acceptance limits…Yield the same relative percentage across concentration range
- Proficiency Testing Reporting Limit (PTRL)
  - As defined in Volume 3 of the NELAC standard (2016):
    - "A statistically derived value that represents the lowest acceptable concentration for an analyte in a proficiency test sample, if the analyte is spiked into the proficiency test sample. The PTRLs are specified in the TNI Field of Proficiency Testing tables"
    - Note: PTRLs are not the same as Method Reporting Limits, LODs or MDLs
- Key consideration for PTRLs:
  - You are not required to be able to quantitate down to the PTRL for a given analyte. However, your analytical method should be able to quantitate down to these levels for added assurance that you can properly report a result should the PT provider have an assigned value at or very close to the lower end of the prescribed concentration range. Noting that acceptance limits can extend below the lower concentrations in these situations.

Assigned Values - How are They Determined?

Assigned values for each analyte are determined accordingly:

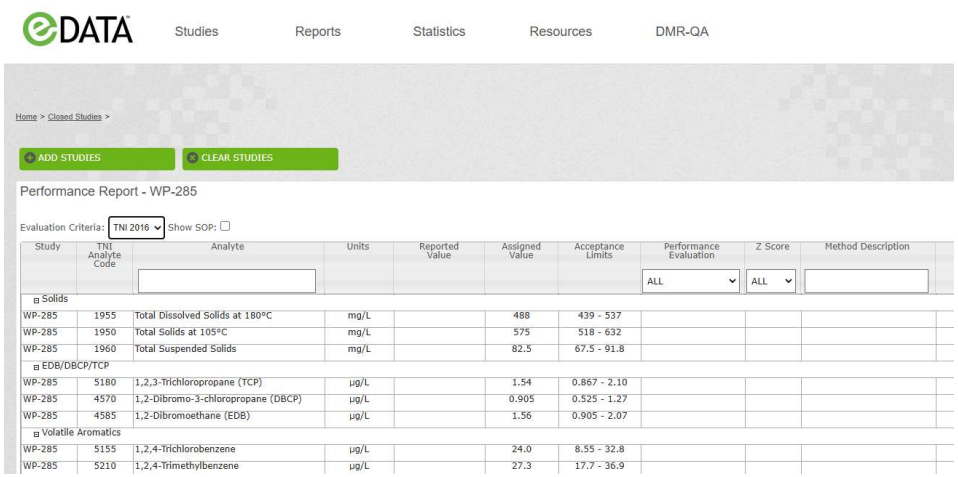
- Actual “made-to” value as determined by mass/volumetric measurements (taking into account chemical substrate purities).
- Measured means (established by the PT provider through internal analyses)
- PT study mean (eg…where only “c” & “d” factors are supplied on the FoPT table)
- Must be compliant with Verification, Homogeneity and Stability criteria (VHS)

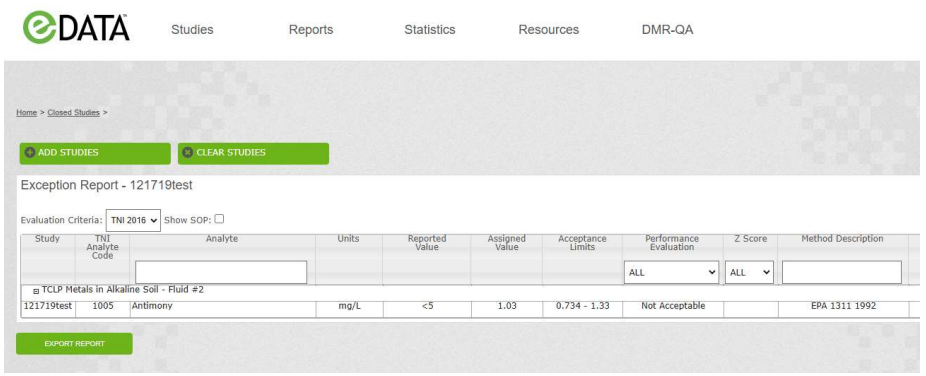
Multi-Modal Data: What is it and how is it Handled?

- 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:
  - Preparatory and/or analytical method bias (i.e., two or more methods may not yield equivalent performance characteristics)
  - PT sample(s) inhomogeneity (within sample and/or between the samples)
  - PT sample(s) may have exhibited instability during the course of the study

PT Monitoring and Trending Tools Available in eDATA

- PT Performance and Exception Reports
- 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





Z-Scores: A Simple Trending Tool

- What is a Z-score?
  - Quite simply, a Z-score represents the distance of a result from the mean of the data set, expressed as a standard deviation

$$Z = \frac{x - \mu}{\sigma}$$

  - Where:
    - X = Your PT result
    - μ = The mean of the data set
    - σ = Standard deviation of the data set
- Negative Z-score represents a result that is below the mean
- Positive Z-score represents a result that is above the mean
- For evaluation purposes, Z-scores $\leq$ 2.00 and sometimes $\leq$ 3.00 are applied for acceptable performance…depending on the study type

Additional Sources of Information

- TNI Website (FoPT Tables, Laboratory Accreditation, PT Program Info.)
  - www.nelac-institute.org
- ISO 17025
- ISO 17043
- ISO 17034
- ERA
  - www.eraqc.com

Thank You

Live Q&A Session: Understanding PT Statistical Analysis and Evaluation

Please submit your questions into the Q&A Chatbox

https://www.eraqc.com