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 $\sim \pm 3$ standard deviations
- DW Acceptance Limits represent $\sim \pm 2$ 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-weighted¹ 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:
  
  \[ G = \max |Y_i - \bar{Y}| \]
  
  \[ s \]

  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.


  **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)

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

• 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 = \((1.00*1.0017)+(-0.0300) = 0.999 \text{ mg/L}\)
Predicted Std Dev = \((1.00*0.0377)+0.0250 = 0.0627 \text{ mg/L}\)

**Acceptance Limits =** \(0.999 \pm (3*0.0627)\) 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 Accréditor (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?