Treating Non-Detects

Kirk Cameron, Ph.D.
kcmacstat@gmail.com
719-532-0453

Just For Fun

- “There is no truth to the allegation that statisticians are mean. They are just your standard normal deviates.”

‘Bug-a-Boos’ of Data Analysis

Key Questions

- What is the ‘ranking’ problem?
- Is simple substitution ‘dead’?
- How do NDs impact common statistical methods?
- How can NDs affect site or regulatory decisions?
- What are proper strategies for handling NDs?
Models of Non-Detects

- Non-detects known as ‘censored values’
  - **True concentration between (0, MDL/RL)**
- Aitchison’s model: non-detects are true 0s (delta model)
- **Modified-delta model:** uniformly distributed between (0, MDL/RL)
  - \( \frac{1}{2} \text{MDL/RL} \) = expected average (no bias)

Models of Non-Detects (cont.)

- Censored model: non-detects behave like detected data, only ‘hidden’
  - Assumes common joint distribution for NDs and detects
  - Simple substitution ‘biased’ in this case

Modified Delta Model

Censored Data Model
Which Model When?
- Aitchison’s model OK when NDs represent true absence (or at most a very minimal reading)
  - Example: NDs generated by distinct physical process; seasonal pattern at plume boundary due to water table fluctuations

Which Model (cont.)
- Modified-delta model
  - Non-detects contain ‘something’ (non-zero), but may not fit censored model
  - Most statistical comparisons between two populations (t-tests, prediction limits, control charts)
    - Non-detects must be treated equivalently in each group (same direction of bias)

Which Model (cont.)
- Censored model
  - Compound present but not (always) measurable below MDL/RL
  - Summary statistics from single population (mean, standard deviation)
  - Single group tests (confidence intervals)
  - ‘Interweaving’ of detects and non-detects
    - Multiple detection limits mixed with low-level detects

Model Upshot
- Aitchison’s and modified-delta both lead to simple substitution (i.e., replacing NDs by 0 or 1/2 MDL/RL)
- Censored model requires more sophisticated adjustments and more data (but less bias)
- If not enough data, simple substitution is the only option
When To (Simple) Substitute

- Sample size too small (e.g., \( n \leq 5 \))
- Plotting cumulative sums (CUSUM) on control charts
  - Only 1 new measurement added per sampling event
- Intrawell prediction limits with small BG size
- Low fraction of NDs (10-15%)
- Any group comparison where all MDL/RLs < minimum detect

Prediction Limits/Control Charts

- Up to 50% NDs —
  - Simple substitution might be OK since comparison groups will have same bias
  - Under censored model, use adjusted estimates of mean, standard deviation in BG limit equation
    \[ \text{Limit} = \bar{x} + k\hat{s} \]
  - May still need to substitute for ND comparison data

Prediction/Control (cont.)

- More than 50% NDs —
  - Very hard to normalize data
  - Switch to nonparametric prediction limit
    - Need at least 1 detect to compute limit

Non-Detects in Confidence Intervals

- Easy button — replace each ND by 1/2 RL or MDL
  - ‘OK’ if \( n \) is small OR all reporting limits < all detected values
  - Not so good if reporting limits, detects intermixed
Non-Parametric Interval?

- Too many NDs negate parametric model fitting (e.g., normal, lognormal)
- Can use nonparametric confidence interval, but only if all NDs can be explicitly ranked

<table>
<thead>
<tr>
<th>Observation</th>
<th>Rank?</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1</td>
<td>1 (?)</td>
</tr>
<tr>
<td>&lt; 1</td>
<td>2 (?)</td>
</tr>
<tr>
<td>2.3</td>
<td>3 (?)</td>
</tr>
<tr>
<td>3.7</td>
<td>4 (?)</td>
</tr>
<tr>
<td>&lt; 5</td>
<td>5 (?)</td>
</tr>
<tr>
<td>&lt; 5</td>
<td>6 (?)</td>
</tr>
<tr>
<td>&lt; 5</td>
<td>7 (?)</td>
</tr>
<tr>
<td>8.1</td>
<td>8</td>
</tr>
<tr>
<td>12.5</td>
<td>9</td>
</tr>
</tbody>
</table>

Adjusting for NDs (Censored Model)

- Two strategies outlined in EPA's Unified Guidance
  - Kaplan-Meier (KM), Regression on Order Statistics (ROS)
  - Both used to build more accurate parametric confidence intervals
- BUT, not always necessary or impactful

Manganese Example

<table>
<thead>
<tr>
<th>Event</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
<th>W5</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>6.3</td>
<td>17.9</td>
</tr>
<tr>
<td>s2</td>
<td>12.1</td>
<td>7.7</td>
<td>5.3</td>
<td>11.9</td>
<td>22.7</td>
</tr>
<tr>
<td>s3</td>
<td>16.9</td>
<td>53.6</td>
<td>12.6</td>
<td>10</td>
<td>3.3</td>
</tr>
<tr>
<td>s4</td>
<td>21.6</td>
<td>9.5</td>
<td>106.3&lt;2</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>s5</td>
<td>&lt;2</td>
<td>45.9</td>
<td>34.5</td>
<td>77.2</td>
<td>&lt;2</td>
</tr>
</tbody>
</table>

Manganese Comparison

![Manganese Index Plot](image)
**NDs in Trend Tests**

- **Cannot use Kaplan-Meier or ROS**
  - These assume a stable population
- **If NDs are lower than any detects, can use Mann-Kendall (MK) test**
  - MK scores all possible pairs as $-1$, $0$, or $+1$
  - Pairs with two NDs are scored as ‘no change,’ ‘flat’ (0)
  - Pairs with one early ND are scored as increases (+1)
  - Pairs with one later ND are scored as decreases ($-1$)

**What About Intermixing?**

- Limit = 6 ug/L

**Same Data, Different View**

**Kendall’s tau-b**

- Nonparametric correlation measure
  - Related to Mann-Kendall test
  - Useful to assess strength of trend when data contain ‘ties’
Kendall’s tau-b (cont.)

- Intermixing means data can only be partially ranked
- Many pairs are ‘ties’ (don’t know which is bigger, smaller)
- Groups of data are ‘partially tied’, overlapping
- Must properly account for pattern of ties to judge strength (significance) of trend

Special Topics

- Dilution ‘Outliers’
- Setting useful reporting limits (RLs)
- ALL non-detects in BG

Dilution ‘Outliers’

- Large dilutions can make data essentially unusable, misleading
- Unrealistic reporting limits
- Imaginary trends, sky-high uncertainty

Vinyl Chloride ‘Trend’
Vinyl Chloride: Another View

How To Handle

- **Standard adjustment methods do not work**
- Two choices
  - **Eliminate** dilution ‘outliers’
  - **Re-classify** using a lower, reasonable RL (one with no or little dilution, if possible)

Setting Useful RLs

- Contract labs do not always report lowest achievable RLs
  - Reasons: cost; not required; avoids having to quantify ‘less certain’ observations (e.g., J-flags)
  - ‘Less thans’ blur regulatory liability

Statistical Headaches

- **RLs set too high — particularly when RL > GWPS — vastly complicate statistical decision-making**
  - Is ND benzene value of <7 ppb above or below 5 ppb MCL?
  - What about a J-flagged value of 5.2 ppb?
Fake Example: Trend or No?

<table>
<thead>
<tr>
<th>Event</th>
<th>Data</th>
<th>RL</th>
<th>All &lt; Max RL</th>
<th>Max RL</th>
<th>1/2 RL</th>
<th>1/2 Min RL</th>
<th>1/2 Max RL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>&lt;2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Q3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Q4</td>
<td>&lt;4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Q5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>p-value</td>
<td>0.014</td>
<td>0.144</td>
<td>0.065</td>
<td>0.065</td>
<td>0.134</td>
<td>0.065</td>
<td></td>
</tr>
</tbody>
</table>

Solutions

- Never treat a single J-flag value as dispositive!
  - Whether above or below GWPS
  - Regulatory decisions must be based on multiple observations
- Set RLs below the GWPS whenever possible
- Use J-flag values ‘as is’ in statistical calculations
  - Vastly improves trend, variability estimates!

100% NDs in BG: Double Quantification Rule (DQR)

- Codified in EPA guidance (Unified Guidance, 2009)
- Strategy for testing against BG when 100% of BG measurements are ND
  - Significant difference declared if two consecutive, quantified compliance values observed
  - Similar to nonparametric prediction limit with single retest

Selenium Examples
Summary

- Non-detects impact how statistical limits, trends, and confidence intervals should be computed.
- Don’t *always* need sophisticated non-detect strategy.
  - Simple substitution may give similar results or be needed due to small sample size.
- Censored models helpful for larger data sets and intermixed samples of detects and non-detects.