limit_of_detection

Tony Ni

9/2/2020

Why the Limit of Detection (LOD) Value is Not an Appropriate Specification for Automotive Emissions Analyzers

- Very specific study on checking the reliability of analyzers (for vehicle emissions) at low concentrations.

- Generally, you want to have a low LOD value

- The general form of the equation used to determine LOD is of the form:

$$LOD = k * s_{zero}$$

  where "k is the constant for defining LOD" and s_{zero} is the standard deviation of the zero/blank"

- 2 or 3 SDs is often what is used, refer to picture (I think it's neat)

- Limit of quantification (LOQ) and limit of detection (LOD) are often concepts which are mixed up

- LOQ is the "minimum value that users of analytical instrumentation can report a value"

A Model of Measurement Precision at Low Concentrations

- Article focusing on the precision of measurements at low concentrations, seeks to explore how effective the total variance model performs

- The method detection limit (MDL) is a "method's ability to determine an analyte in a sample matrix, regardless of its source of origin" - it is estimated from data

- Background noise always exists, even when there is no analyte present

- EPA definition of method detection limit is as follows: "method detection limit (MDL) is the minimum concentration of a substance that can be measured and reported with 99

- In the total variance model – we can't measure the actual analyte concentration. We have to estimate it from the measured concentrations. These measurements innately contain random errors. Palleton (look for source? 1985) suggests that at low concentrations, total error consists of: background noise b_i and analytical error a_i – both of which are random, independent, and normally distributed with mean 0 (refer to diagram in paper, nice figure/picture)

Statistical methods for assays with limits of detection: Serum bile acid as a differentiator between patients with normal colons, adenomas, and colorectal cancer

- THIS SOURCE LOOKS REALLY GOOD. BACKSOURCE FROM HERE AND FIND MORE FROM THIS PAPER'S CITATIONS

- There are many different ways that researchers are dealing with detection limits, which are ubiquitous in the scientific realm. Substitution, nonparametric methods, and maximum likelihood methods are all ways to combat this problem.

- Substitution is the worst way, nonparametric ways do better, maximum likelihood methods are the best

A Study of the Precision of Lead Measurements at Concentrations Near the Method Limit of Detection

- Study involving generating collection of measurements on samples near the limit of detection. Gave 5 labs samples of lead with low concentrations and asked them to measure them, forbade them from recording values as "less than MDL" and they had to record a number.

- Method Detection Limit (MDL) is often misunderstood – it isn't involved with bias, it only deals with precision.

- Throwing away/Discarding values that are below of the LOD gets rid of tons of useful information

- The MDL is a statistical concepet more so than a chemical concept. It is a statistic which is estimated from the given data.

- The way that chemists/researchers report low concentrations are varied and not standardized. Some may report: "ported. They may report the datum to the data analyst as (1) trace, (2) the letters ND (not detected), (3) the numerical value of MDL itself, (4) a "less than" value, that is, the numerical value of the MDL preceded by a "<" sign, (5) zero, (6) some value between zero and the MDL, for example, one-half the MDL, (7) the actual measured concentration even if it is below the MDL (that is, whether the value is positive or negative), (8) the actual measured value followed by the MDL in parenthesis, or (9) the actual measured value with a statement of its precision (for example, 2 ± 4 Mg/L, where the $\pm$ value indicates the pre cision of the estimate). The last three methods are the best."

- Paper emphasizes not throwing away ANY numerical values, but doesn't reallyl talk about how to perform statistical analysis with these values...

Real-time detection of intentional chemical contamination in the distributional system

- "80

- "Miller and Miller (2000) defined the LOD as being equal to blank + 3(SD)":

$$Limit of Detection = \gamma_B\ + 3\sigma_B$$

- Note: in their case, the blank is 0 because they are measuring the different between normal/control/baseline conditions against the contaminant

- Recommends using artificial neural networks with large datasets (to help detect contamination)

- The rest of the article focuses on their experiment to try to detect contamination in real-time.

A Distribution-Based Multiple Imputation Method for Handling Bivariate Pesticide Data with Values below the Limit of Detection

- Study exploring different options to handle LOD laboratory data – specifically with regards to multiple imputation methods for left-censored data. They concluded that "the distribution-based MI method" worked well for bivariate data where the values were < LOD.

- LOD entries still contain information that a lot of people don't realize – specifically information that the values is between 0 and the LOD.

- Lots of different approaches has already been done to handle this (they cited Helsel 2005b, 2010)

- Substitution methods are easy to implement, but are biased (common values: LOD/2, LOD/sqrt(2), LOD) but are discouraged b/c results in estimates of parameters being biased

- What this study used was distribution-based multiple imputation methods – they used MLEs to estimate distribution parameters based on all datas (< LOD and those not). They repeatedly imput the values to create multiple complete sets of data, and then analyzed each one individually
- Mathematically, they created a log-likelihood function with all the data, then derived MLEs of each parameters on multiple bootstrapped datasets. Each bootstrap data gives different estimates for the mean, sd, etc. (refer to article for math)

When Nothing Is Something : Understanding Detection Limits
- There are many different forms of detection limits, ranging from: thresholds at which something can be detected (instrument detection limit) or the limit to which a lab reports a value based regulatory requirements (reporting limit).
- Detection limits are constantly changing: as technology improves, so too does our ability to accurately measure substances.
- Reporting limits can often be used be misleading as non-statisticians may often interpret ND (non-detect) as nullity, when in fact it only means that the measurement falls below a certain limit.
- The rest of the paper focus on environmental law practices and cases in which LOD has played a role in.

Lowering the detection limit for arsenic: Implications for a future practical quantitation limit
- States can set their own regulatory standards regarding the maximum contaminant levels for chemicals. (own thoughts: maybe this is another issue...)
- Rest of the article is focused on chemical instruments in measuring things, not relevant...

Estimating Population Distributions When Some Data Are Below a Limit of Detection by Using a Reverse Kaplan-Meier Estimator
- Suggests using the reverse Kaplan-Meier (KM) estimator to estimate the distribution function and population percentiles for data where there is "left-censored data" (data point is below a certain value but known by how much)
- Again, talks about how substituting values with LOD/2 etc. is common in environmental sciences but is not recommended
- (https://publicifsv.sund.ku.dk/ tag/Teaching/share/R-tutorials/Advanced-statistics/SurvivalAnalysis.html) found this link for Kaplain-Meier tutorial in R
- After their study, they found that even though THEORETICALLY the reverse KM is for left-censored data (just like KM is for right-censored data), it is still limited in its usage since all it does it estimate a distribution function (it's really just an exploratory thing, they said)
- Proposes other ways to handle left-censored data like the Turnbull estimator (but not much work has been done in that field)

Statistical tests for latent class in censored data due to detection limit
- Tobit regression model should be used when data is from a single normal distribution, with "some" observations under the detection limit (if not normal, do transformation)
- Paper seems to focus on latent class analysis with missing data – not on methods to handle it really, not that relevant...

Week of 9/9/2020

Estimating mean exposures from censored data: Exposure to benzene in the Australian petroleum industry

- Task in the study is to enalyze exposure to benzene in a cancer study, and they found nearly all datasets had some values below LOD, with some datasets where 95

- First method they tried was replacing all values with LOD/2, they claim these method was recommended for datasets where lots of data is below LOD or when data is highly skewed with a geometric sd of 3 or more.

- Second method is replacing with LOD/sqrt(2) and is recommended to be used when few data us below LOD or when data is not highly skewed

- Third method is "Cohen's Method" where one extrapolates the left hand side of distribution based on the distribution of the uncensored data and then calculat the MLE estimate of the arithmetic mean – found to be unreliable with data with outliers, this method can ONLY be used with data where there is a single LOD

- Results of the studied showed the first two methods similar, third method gave high, unlikely results of the mean

- There is another method (Hald's Method) they discussed which isn't applicable when more than 50

- (IDEA: Do some sort of plot like they did where we check the percentage of censored data for each chemical? refer to page 4...)

-

A comparison of several methods for analyzing censored data

Methods for Handling Left-Censored Data in Quantitative Microbial Risk Assessment

- Compared 5 methods: found that in terms of performance: imputation method using MLE to estimate distribution parameters and then imputing censored data points with values from this distribution below the LOD > imputation from a uniform distribution > other 3 methods (substitution, log-normal MLE to estimate mean and SD, and kaplan-meier estimate)

- KM method is better than MLE for data where there are TONS of missing data or if data is highly skewed (distribution not assumed in KM method)

- Imputing from a uniform is useful when you don't know the distribution of the data

- MLE and KM methods were implemented using the NADA package (https://cran.r-project.org/web/packages/NADA/NA in R ('cenmle' and 'cenfit' functions) where data is labeled as censored or uncensored, for censored values, LOD is used as a placeholder, since these methods aren't imputation methods – these censored values weren't replaced. Instead, summary statistics were generated with the entire data set (including the censored data)

- The imputations methods used mostly followed the general ideas: we have to assume the entire data set follows a particular distribution. Then we use this distribution to impute in values for the censored data. The MLE imputation method uses MLE methods to estimate the parameters of a distribution to fit the dataset, then values lower than the LOD are imputed FROM this dataset for all censored values (they used the function 'fitdistcens' from the R package 'fitdistRplus'). The second uniform imputation method assumes a uniform distribution with minimum 0, maximum LOD – for all values less than the LOD, then the left-censored values are replaced with a number randomly selected from this uniform distribution.

- uses RMSE (root mean squared error) to see how close the estimated values are to the true values (lower RMSE means closer estimation to known values)