Webinar 2: ProUCL A to Z

2. * Moderator Jean Balent: ProUCL can be downloaded for free from https://www.epa.gov/land-research/proucl-software

3. * Moderator Jean Balent: The sample datasets were updated on 2/7/20, they can be downloaded from https://clu-in.org/conf/tio/ProUCLAtoZ2/#tabs-3

X 60. : Wouldn't the substitution with a fixed value artificially alter the distribution of the data and decrease the variance?
Simple substitution usually leads to greater statistical bias and inaccuracy.

X 61. : The data set didn't appear to have a column for the date of the sampling event. Does the program automatically convert the dates to a corresponding event number? Does it calculate the number of days between events?
ProUCL does not convert dates to days between the events, this needs to be entered manually.

64.: Do we need to standardize them?
* Travis(privately): Standardization is an option for OLS but I wouldn't recommed it for our two non-parametric options as they do not have distributional requirements

65. : Can there be a text column with dates for presentation purposes?
* Travis(privately): You can load it in, but not use it for analysis

66. : I definitely think that how to handle 'J' values would be a valuable webinar - can deal with not only the statistical aspect, but the 'reality' aspect (asking lab for info, knowing the different limit (method DL, reporting limit, etc.).
* Travis(privately): Agreed! However that is worth its own full webinar and we already have a lot to cover on basic ProUCL functionality

68. : Cannot see either the data or the regression line on the screen
* Moderator Jean Balent(privately): Hi are you referencing her demo in ProUCL or the slides? If your screen is blank/white, try exiting and rejoining the webinar.

69.: Is there a similar way to evaluate the strength of the Theil Sen result?
* Travis(privately): Short anwer, yes. However due to the nature of the test it's not quite the same

70.: The decreasing trend in this dataset doesn't seem to be linear. Are other models available for testing?
* Travis(privately): Don't worry we're getting there :)

71.: Definitely a webinar by itself.
* Moderator Jean Balent(privately): Thanks for the feedback!

74.: What should you do if your data appear to exhibit a seasonal trend?

76: How do you save all the work in ProUCL?
* Travis(privately): just highlight your file on the nav panel. then click file -> save as

79. * Moderator Jean Balent: Thanks for joining us!
Seminar slides and other materials can be accessed from https://clu-in.org/conf/tio/ProUCLAtoZ2/#tabs-4
Please submit feedback after the event at https://clu-in.org/conf/tio/ProUCLAtoZ2/#tabs-5

X 87. : If J-coded values can't be used, should only values above the PQL be used?
J-coded values are treated the same as non-J-coded values. If they are non-detects they should be treated accordingly.

X 88. : Can Pro-UCL handle data that has multiple PQL values?
Please refer to ProUCL 5.1 User Guide Chapter 1.11 for detailed discussion on samples with non-detect observations

X 89.: If J-coded values can't be used, should only values above the PQL be used?

J-coded values are treated the same as non-J-coded values. If they are non-detects they should be treated accordingly.

X 91. : Will multiple PQL values alter trend results?

Trend analysis methods available in ProUCL don't account for NDs, but will take reported PQLs as they are. Please refer to recommendations presented in the webinar for treatment of NDs in trend analysis.

X 92. : Can ProUCL do Kaplan-Meier substitution for non-detects?
It does within UTL and UCL calculations, certain tests will not make this implementation directly. For specific test questions consult the ProUCL Technical Guide.

X 93. : What do you mean by attempt in "If they are greater than highest attempt, you are quite safe to reject them."?
High non-detects always create severe problems in analyzing data. They should be removed from the data set before further analysis. Rule of thumb to reject NDs:

- always reject NDs greater than the largest detect
- Use judgement to reject NDs greater than 10x the smallest detect

X 94.: I remember reading in the ProUCL documentation that substituting NDs using 1/2 DL or 1/2 RL is not recommended anymore. This slide seems to say otherwise.
For trend analysis, substituting a small number of NDs (10% to 15%) with ½ DL or ½ RL is acceptable. : Please -refer to ProUCL 5.1 User Guide Chapter 1.11 for detailed discussion on samples with non-detect observations and ProUCL 5.1 Technical Guide for information on analysis of data sets with non-detects for specific statistical methods.

X 95.: Would you say that Mann-Kendall is better suited to trend analysis in the presence of non-detects than linear regression?
It is easier to reliably implement as there are less assumptions to satisfy but not necessarily better.

X 96. Can ProUCL "bootstrap" the non-detects? (pick an imaginary number below the detection limit)
ProUCL currently does not do that.

97. : taking a lot of time to load data sets.

X 98. Is there any trend analysis method (may not be in ProUCL) that does not require substitution for non detects.
The trend analysis methods in ProUCL do not account for non-detects. A user needs to manually substitute their imputed ND values if they wish to use them in trend analysis. However, if the proportion of NDs in your dataset is low (<10%) utilizing trend analysis without replacing these NDs should still give a reasonable result.

X 99. How would you go about when one has to series of paired data and objective is to evaluate the ratio between paired data? If one of the two series has no detects, how would it be possible to calculate the ratios? EXAMPLE: pairs of concentration data on the two sides of a membrane to evaluate the ability

of such membrane to mitigate the movement of a given contaminant across the membrane
This is a very problem specific question. I suggest consulting a statistician for a more specific answer.

X 100.: Using some other Mann Kendall Calculation software, trends can be categorized based on the "S" statistic as increasing, probably increasing, no trend/stable, probably decreasing, and decreasing. ProUCL doesn't calculate in quite the same way so is there a recommendation regarding a simlar breakdown based on how postive or negative the "S" statisitc is?
We don't provide such recommendation as one needs to take into consideration all parameters of a situation (sample size, number of non-=detects, etc.) that have impact on p-value.

X 101.: If you did have NDs in this set, could you set these data points to appear differently on the plot? Could you have the test convert them to 1/2 DL if that's how you decide to handle them?
ProUCL unfortunately doesn't have the capability to deal with ND in trend analysis. However, it does display NDs as inverted triangles within QQ-plots.

X 102.: where did the time in days come from? I missed that.
This is user defined variable in data set.

X 103.: Are there any special steps to manage ND values, such as the first training (e.g. stats/sample sizes), before running trend analysis?
High non-detects always create severe problems in analyzing data. They should be removed from the data set before further analysis. Rule of thumb to reject NDs:

- always reject NDs greater than the largest detect
- Use judgement to reject NDs greater than 10x the smallest detect

- Reasonable substitutions for remaining NDs in trend analysis are:
  - Substitute only if there is a small number of Nds (10-15%)
  - ½ of reporting limit (RL)
  - ½ of detection limit (DL)
- For percentage of NDs between 15%-50% use statistical methods such as Kaplan Meier

X 104.: Are the rule of thumbs mentioned for non-detects widely accepted or those proposed by the speakers and/or the EPA?
Rule of thumbs for dealing with high NDs is our recommendation and we have seen other environmental statisticians using them as well.

105.: What do you recommend for minimum sample size to use the Thiel Sen test?
Minimum sample size is prescribed by applicable guidance for site location and field of application.

X 106.: How would we handle a trend analysis of 5 data points? Do either of these analyses work?
Mann-Kendal test can be used with as little as 4 data points, however the power of test is low. This means that you may not detect the trend even if it is present, because you don't have enough evidence to confirm it.

X 107.: Is there functionality to deal with autocorrelation and seasonality within ProUCL?
ProUCL does not support analysis of autocorrelated data and seasonality.

108.: How does thhe software pe4form
: if there are more than two duplicates at the same time does proUCL take the average or the median value
Theil-Sen test for trend averages repeated observations.

110.: J is estimated not non-detect.

111: Against lsrge values for dstes

X 112.: For the time axis can you enter the information as dates or does it have to be entered as days?

Time / event needs to be entered as a numeric variable.

113. : These data have more uncertainty than the unqualified data.

114.: J-coded should refer to estimated values. A J qualifier or J flag typically means an estimated value.

X 115.: Please explain what is meant by the "Power" of a statistical test?
The power of a statistical test is the probability that it correctly rejects a false null hypothesis. For example, Mann-Kendall test performed on small number of observations may not detect the trend even if it is present, because you don't have enough evidence to confirm it.

116.: Shouldn't J-coded values be considered as detect since there is a 99% chance it's there. The exact values is uncertain.

117.: Statistically, these should probably be used as is rather than a substitution as it will result in less uncertainty.

X 118.: J values are estimated levels- can be detects or NDs, right?
this is a question for the lab and data collection.

X 125.: Can ProUCL do the Seasonal Kendall?'
Currently it cannot.

X 126.: Seems like you cannot save all the files in the nav panel as a group?
Currently they need to be saved individually.