

© PracticalStats.com  <

## Nondetects And Data Analysis: Storing Nondetects in Databases

Dennis R. Helsel, Ph.D  
PracticalStats

1

© PracticalStats.com  <

## Goals for Storing Nondetects in Databases

1. Censored and detected observations can both clearly be identified and distinguished
2. Left censored (nondetect) data can be distinguished from right-censored (greater-than) values
3. Data are easily incorporated and used by statistical software
4. Censoring by detection versus quantitation limits can be distinguished and recorded

2

2

© PracticalStats.com  <

## Storing Nondetects in Databases

Two methods used in statistics software

1. Indicator column
2. Interval Endpoints

3

3

© PracticalStats.com  <

## Indicator Column Format

		<u>Conc</u>	<u>ConcCens</u>
<1	---->	1	1
<5	---->	5	1
10	---->	10	0

- First column are the concentrations plus RLs
- Second column is the indicator. 1 = censored value
- In R, second column will become a logical (TRUE/FALSE) variable with 1 = TRUE. If this isn't done when reading in the data, use the as.logical function:

```
> TF <- as.logical(ConcCens)
```

4

4

© PracticalStats.com  <

## Indicator Column Format

Text can be used as the indicator

		<u>Conc</u>	<u>Remark</u>
<1	---->	1	E, J, < etc.
<5	---->	5	<
10	---->	10	Detect, “ “, etc.

Either text or numbers could be used to differentiate DL from QL, but the interval endpoints format is more appropriate for that

5

5

© PracticalStats.com  <

## Interval Endpoints Format

1<sup>st</sup> column is the lower limit, 2<sup>nd</sup> column is the upper limit

		<u>Start</u>	<u>End</u>
<1	---->	0	1
<5	---->	0	5
10	---->	10	10

- Detects have same value in both columns.
- Used by most commercial software for parametric (Maximum likelihood) censored methods

6

6



## Interval Endpoints Format

This is the most flexible format. Can easily differentiate between a DL and a QL. Easy to explain to general public.

For example, for a DL=1 and QL=3:

		<u>Start</u>	<u>End</u>
<1	---->	0	1
2J	---->	1	3



## To 'translate' from indicator to interval format

End column is already there -- the Data column. For nondetects it contains the DL, the max possible

Start = Data column\*(1 - Indicator)

where the indicator is a 1 for nondetects, 0 for detects

Resulting End column has 0 for nondetects, data value for detects

© PracticalStats.com  <

## The RStudio Demo

The Three Steps to Start Using RStudio

1. Start up RStudio
2. Set the working directory  
*Session > Set working directory > Choose directory*
3. Load the course packages  
*Code > Source file* and select Loadlibs.R

- **OR** check the boxes next to all 15 packages required.

Then we will:

1. Read in data sets using several different methods
2. Convert 1/0s to TRUE/FALSE

9

9

© PracticalStats.com  <

## Reading Data in R

Data sets that come with R packages:

- > data (ShePyrene)
- > attach(ShePyrene)

Use  button in RStudio  
Can read text files, .csv or .txt and excel files .xls or .xlsx

**Command Line** ↖ in the working directory, or full path name

- > Pyr = read.table(filename, header=TRUE)
- > Pyr = read.table(file = file.choose())

10

10