

CLRA – 2005 WINTER AGM

Quantifying Data Uncertainty in a Phase II ESA

by

Tai T. Wong & John G. Agar

O'Connor Associates Environmental Inc.

O'CONNOR ASSOCIATES Z

Feb. 2005

Data Uncertainties

◆ Possible Sources of Environmental Data Uncertainty:

- Environmental variability
- Measurement bias, precision & accuracy
- Random sampling errors
- Gross errors & mistakes

To Quantify Data Uncertainty

◆ Confidence limits (C.L.s)

- Example: UCL = 120 mg/kg

◆ Confidence intervals (C.I.s)

- Example: 100 ± 20 mg/kg or
 100 mg/kg $\pm 20\%$

What is Confidence Limit?

- ◆ A statistical parameter describing the probable range of the mean (usually) of a population
 - UCL, LCL & CI
 - 95% UCL => 5% chance that the true mean may be greater!

Who Use C.L.s?

- ◆ **Analytical Laboratories – CAEAL
Uncertainty Protocol**
- ◆ **Proposed GVRD Sewer Use Bylaw**
- ◆ **EPA Superfund & RCRA Sites**
- ◆ **AENV**

Example of a UCL

- ◆ If the 95% UCL of the mean metal concentration in a soil is 120 mg/kg
- ◆ We are 95% certain that the TRUE MEAN concentration does not exceed 120 mg/kg

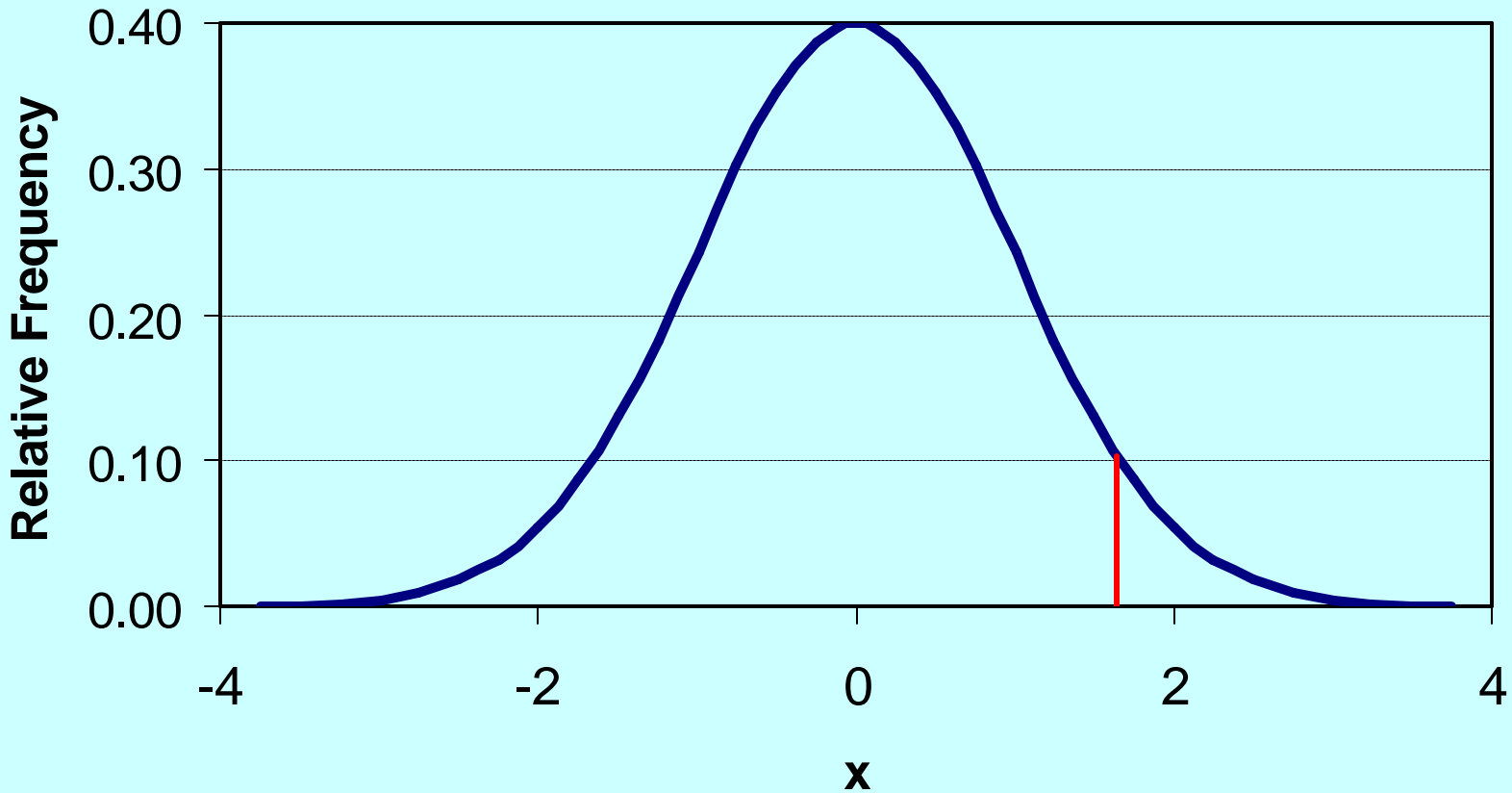
Characteristics of a CL

- ◆ **Depends on the probability distribution of the population**
 - Parametric or Non-Parametric
- ◆ **Depends on sample size**
- ◆ **Can be 1-tailed or 2-tailed**

Probability Distributions

- ◆ AKA – probability density functions (pdf's)
- ◆ A plot of values (say, concentrations) vs relative frequency
- ◆ Common PDFs used are:

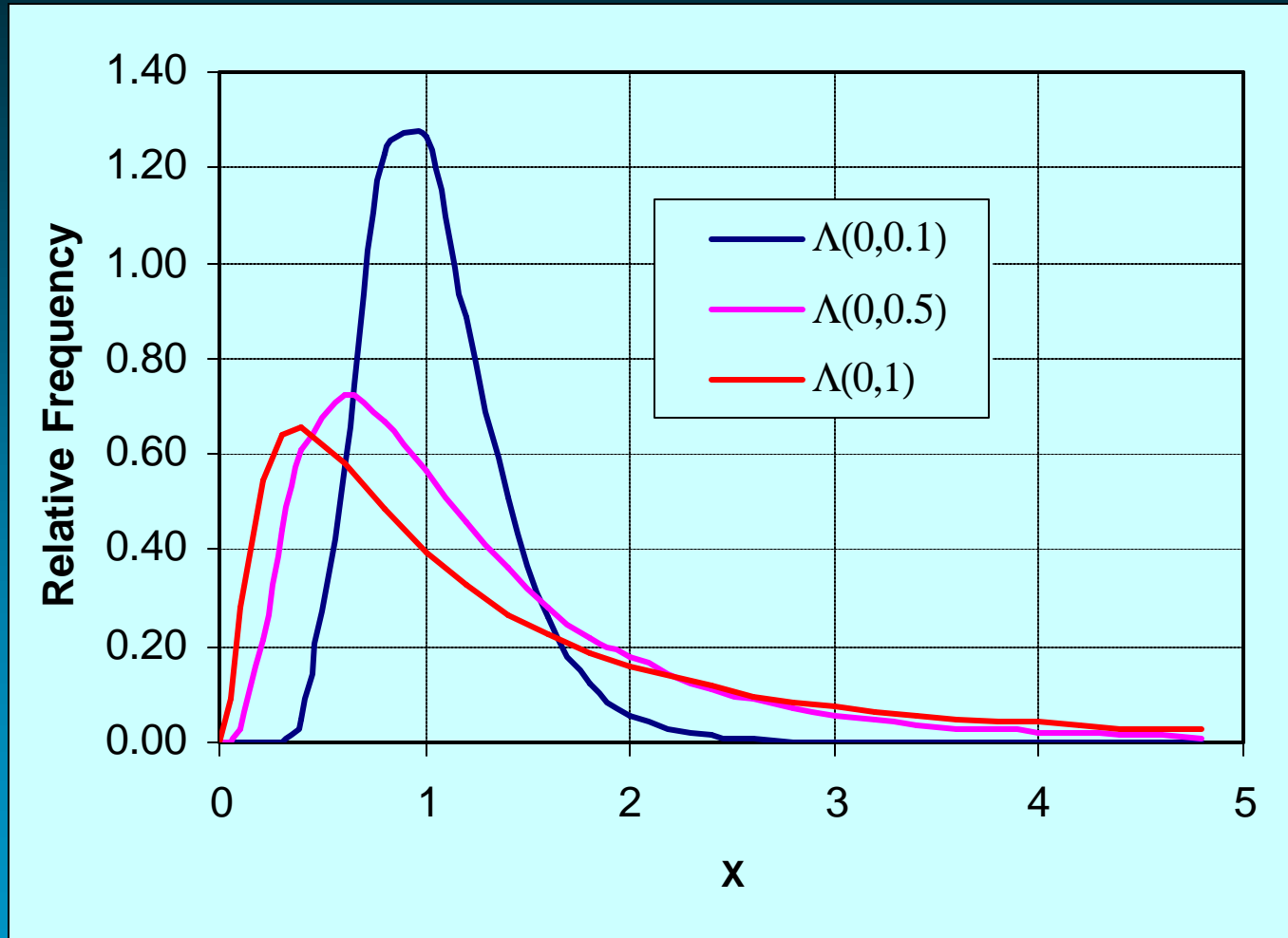
Std Normal PDF: $N(0,1)$



O'CONNOR ASSOCIATES Z

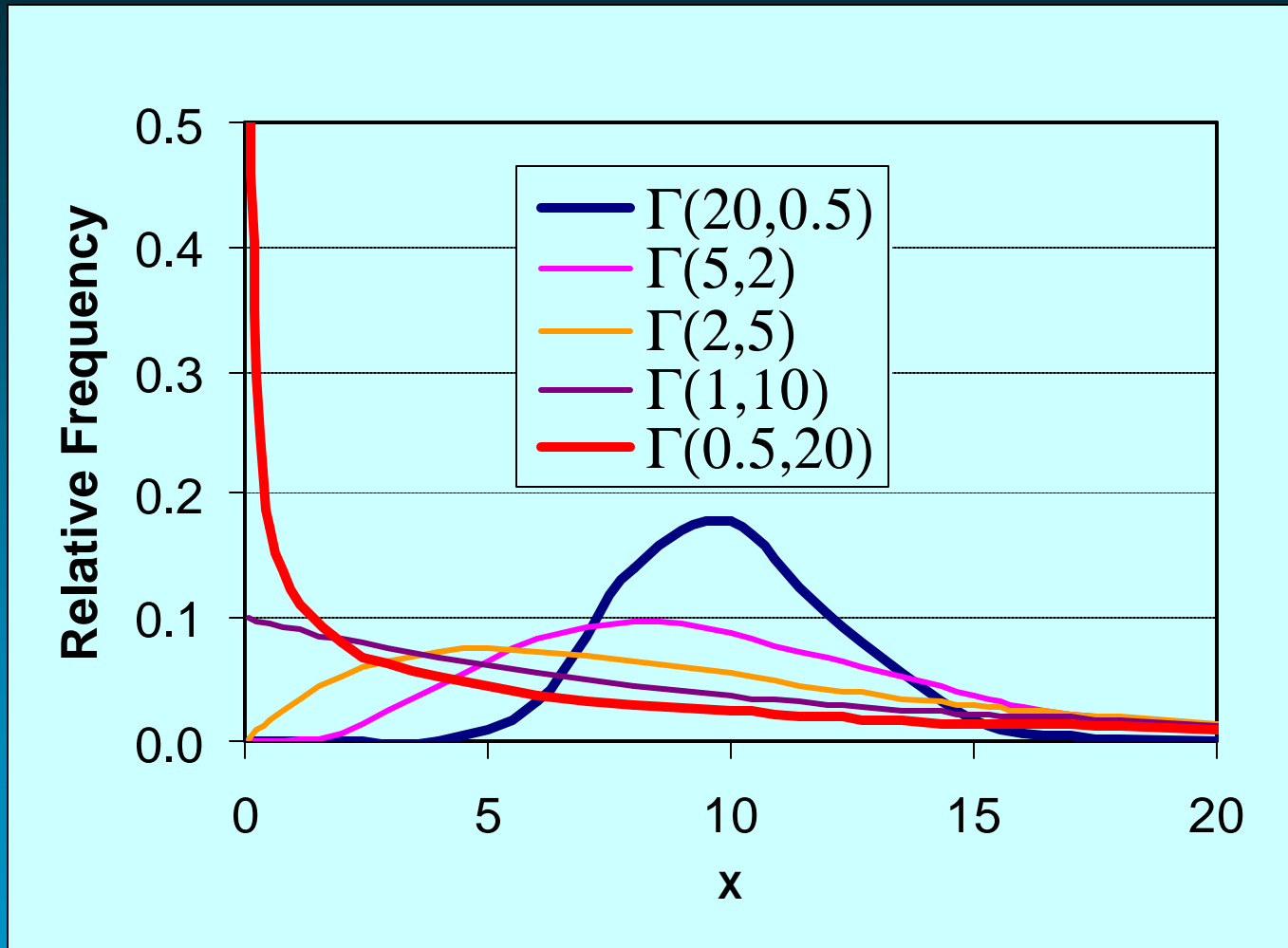
Feb. 2005

Log Normal PDF: $L(m, s)$



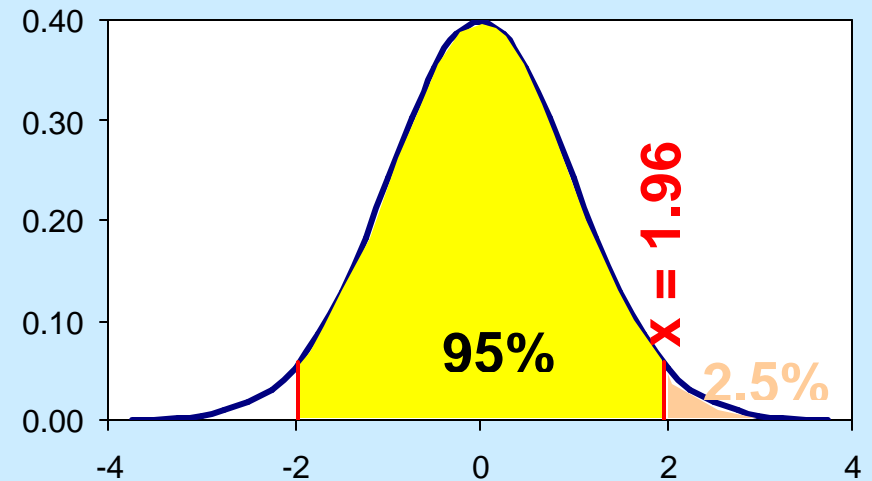
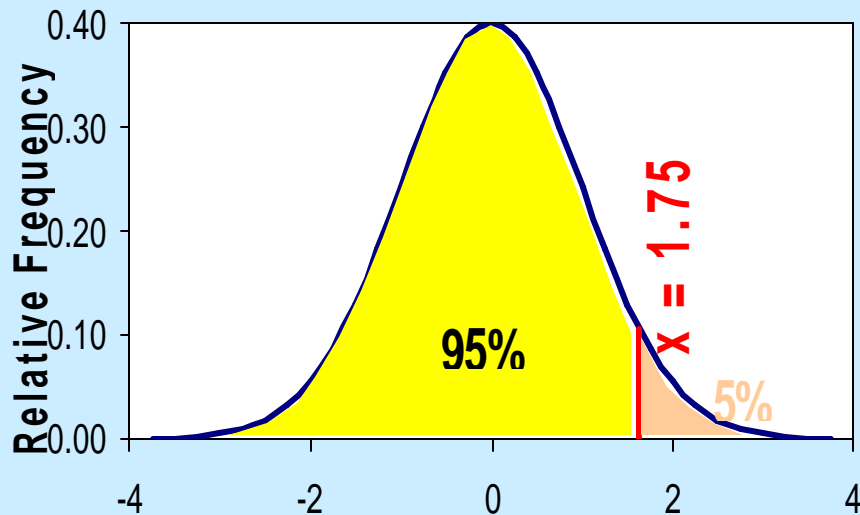
O'CONNOR ASSOCIATES Z

Gamma PDF: $G(k,q)$



O'CONNOR ASSOCIATES Z

Tails – 1 or 2?



1- tailed for exposure/contamination data

2- tailed for measurement/lab data

O'CONNOR ASSOCIATES Z

How to Calculate UCL's

◆ For populations with Normal or Lognormal pdf's:

- Standard formulae; MS Excel

◆ For populations with unknown pdf's (non-parametric):

- PRO-UCL, WQSTAT, SAS, etc.

PRO-UCL 3.0

◆ **Developed by USEPA**

◆ **Capabilities:**

- **Estimates common statistics**
- **Computes 1-tailed UCL of the mean**
 - » **Uses parametric (5) and non-parametric (10) methods**

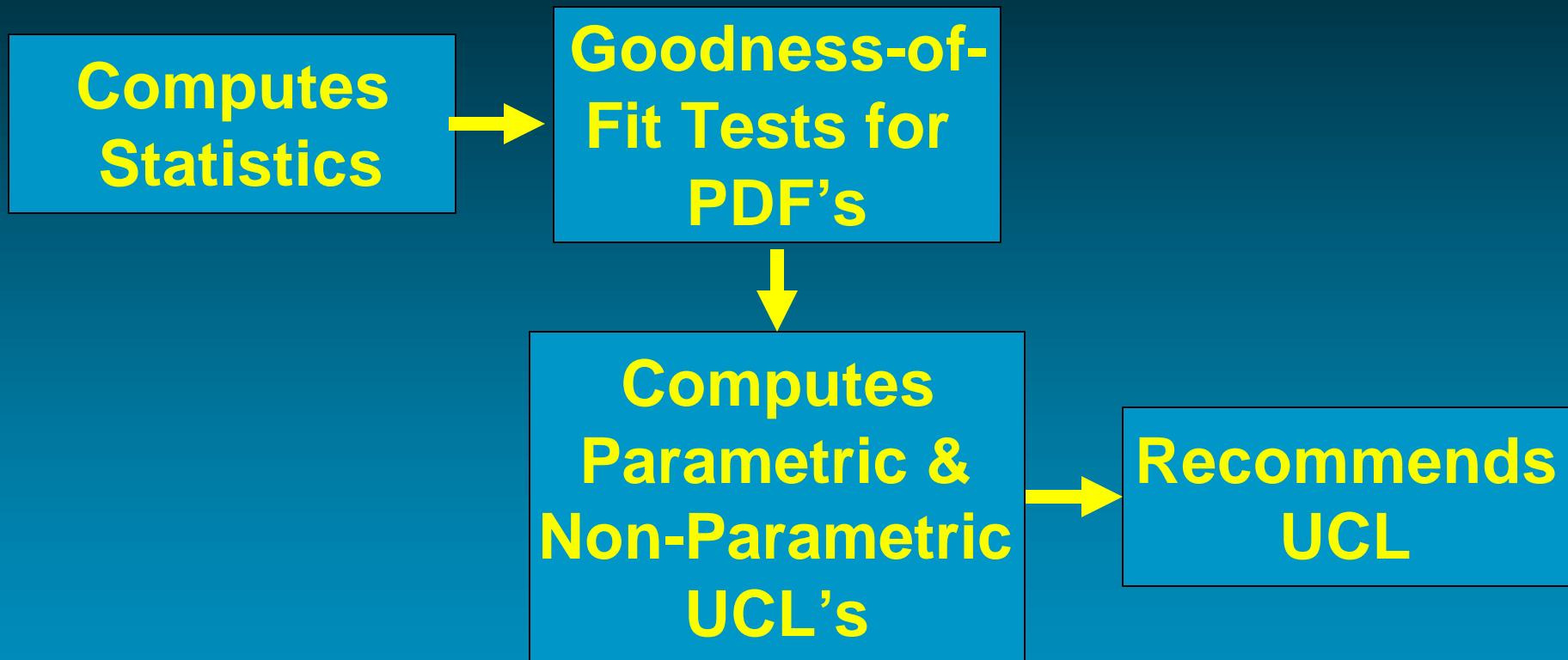
O'CONNOR ASSOCIATES Z

$$P\{X \leq (m_x + h s_x)\} \geq 1 - \frac{1}{h^2}$$

Non-Parametric Methods

- ◆ AKA distribution-independent
- ◆ Uses mean and s.d.
- ◆ Based on –
 - Central Limit Theorem (CLT)
 - Chebyshev Inequality
 - Etc.

Pro-UCL - Overview

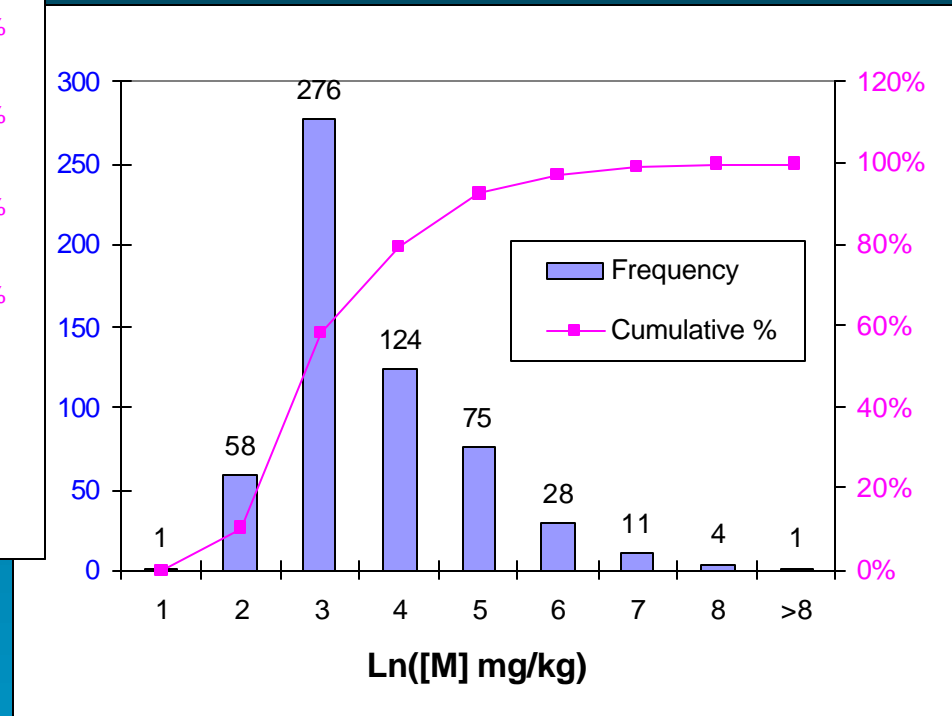
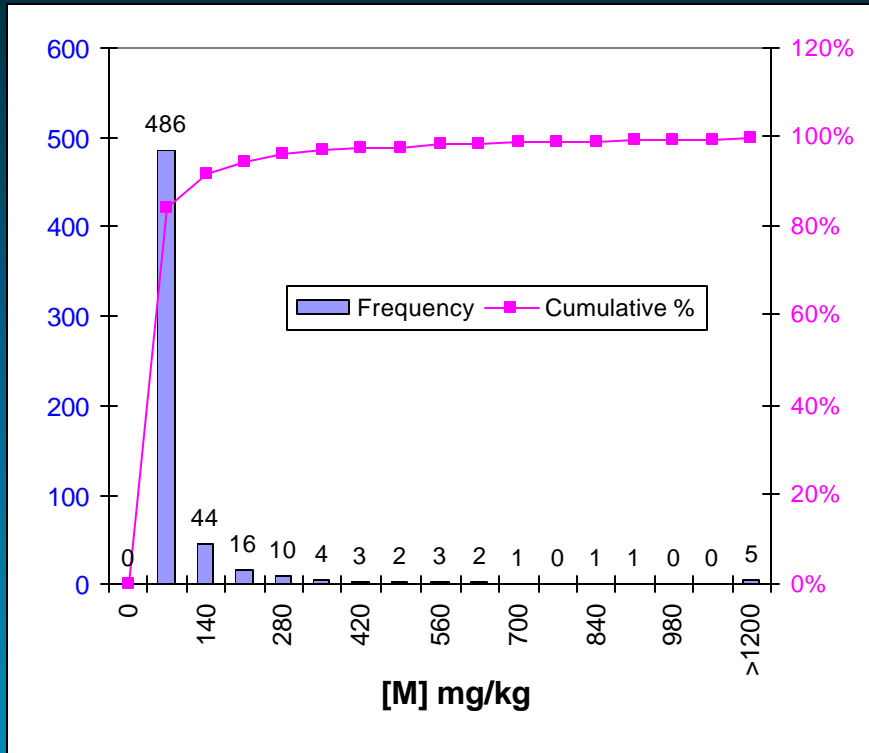


Example 1 - Statistics

- ◆ Metal M Conc. in Soil (mg/kg)
- ◆ >0.3 m to 1.5 m
- ◆ N = 578
- ◆ 328 boreholes

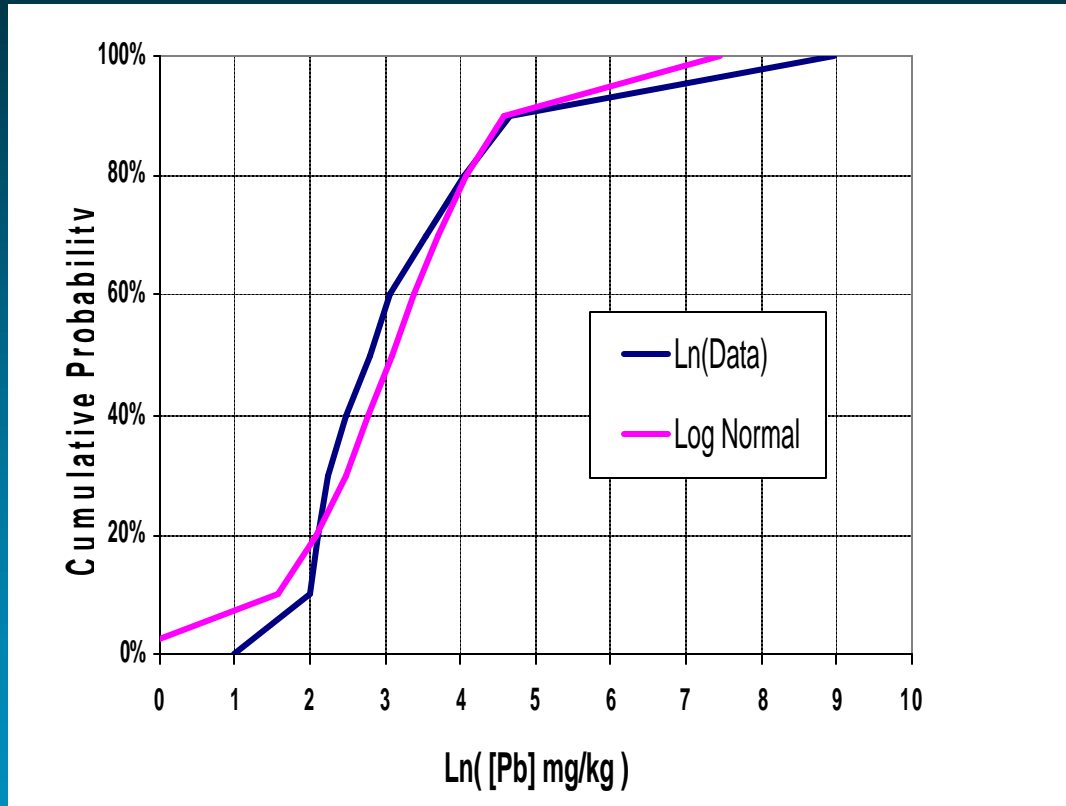
	[M]	Ln([M])
Mean	71.9	3.09
Median	16.4	2.80
S.D.	367.3	1.17
Kurtosis	352.2	1.96
Max.	7840.0	8.97
Min.	1.7	0.53

Example 1. - PDF's



O'CONNOR ASSOCIATES Z

Example 1.. OAEI Analysis



- N = 578
- ~ Lognormal

UCL	Log Normal	Normal
90%	23.4	91.6
95%	23.8	97.1

Example 1... Pro-UCL

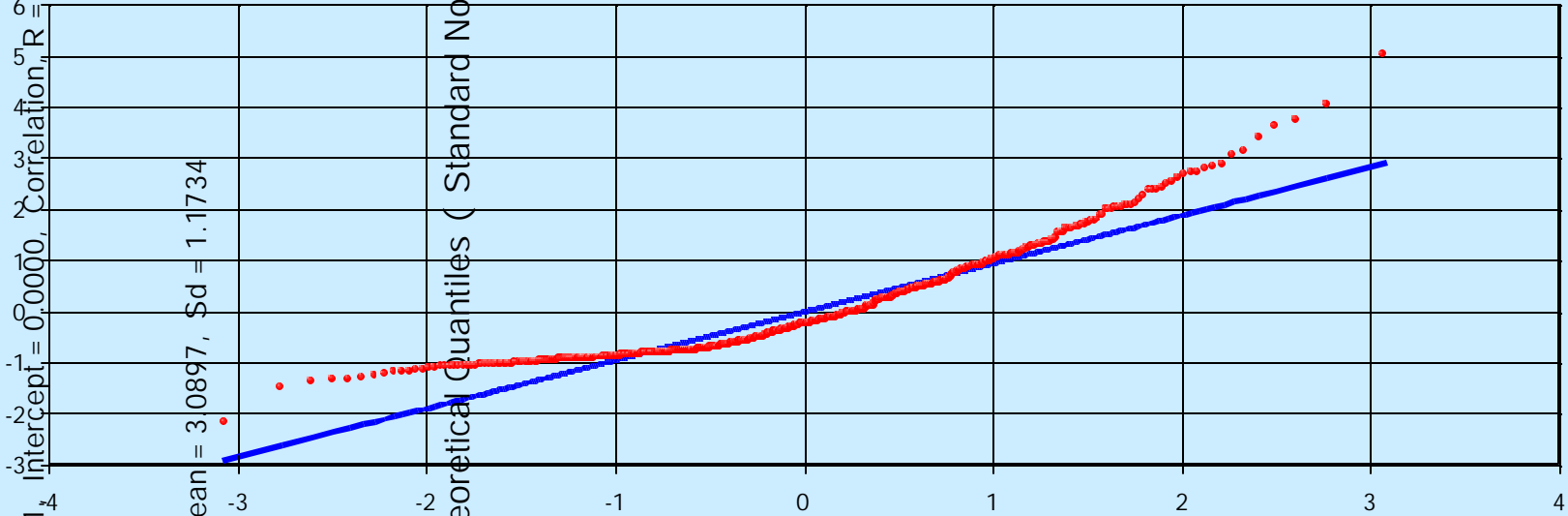
Lilliefors Statistic = 0.117, Critical Value (0.05) = 0.037, Data not Lognormal
 Standardized Ordered Observations

Slope = 0.9461, Intercept = 0.0000, Correlation, $R = 0.94364641$

N = 578, Mean = 3.0897, Sd = 1.1734

Lognormal Plot for Me

Theoretical Quantiles (Standard Normal)



O'CONNOR ASSOCIATES Z

Example 1.... Pro-UCL Results

◆ Data are non-parametric

◆ Recommended UCL's:

- 90% - No recommendation.

- 95% - 167.4 mg/kg

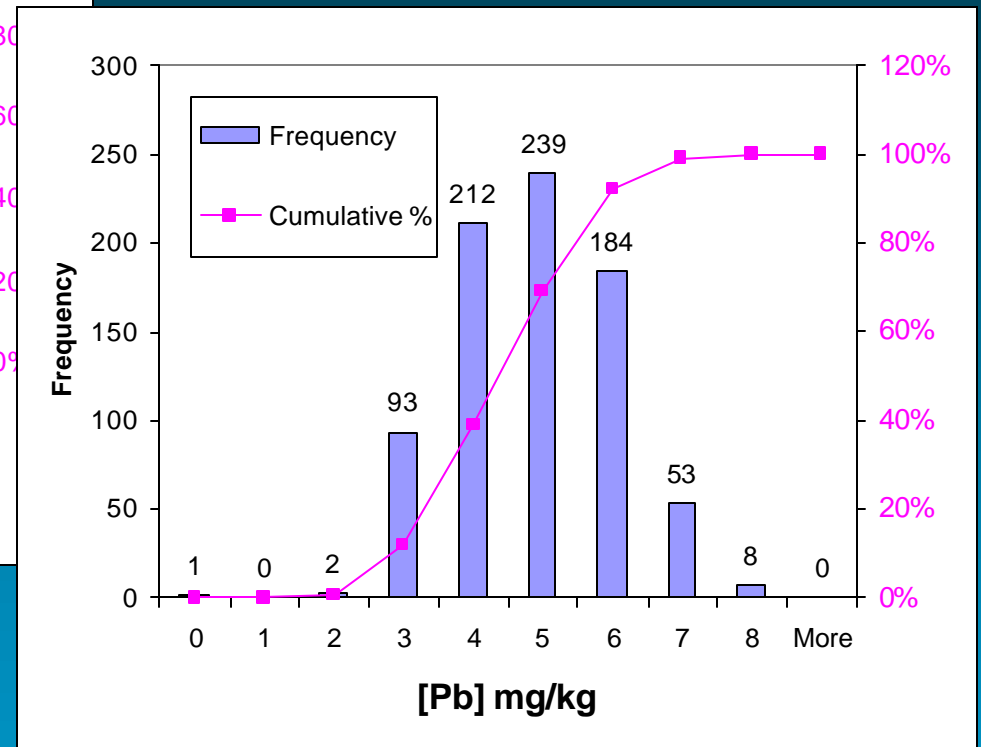
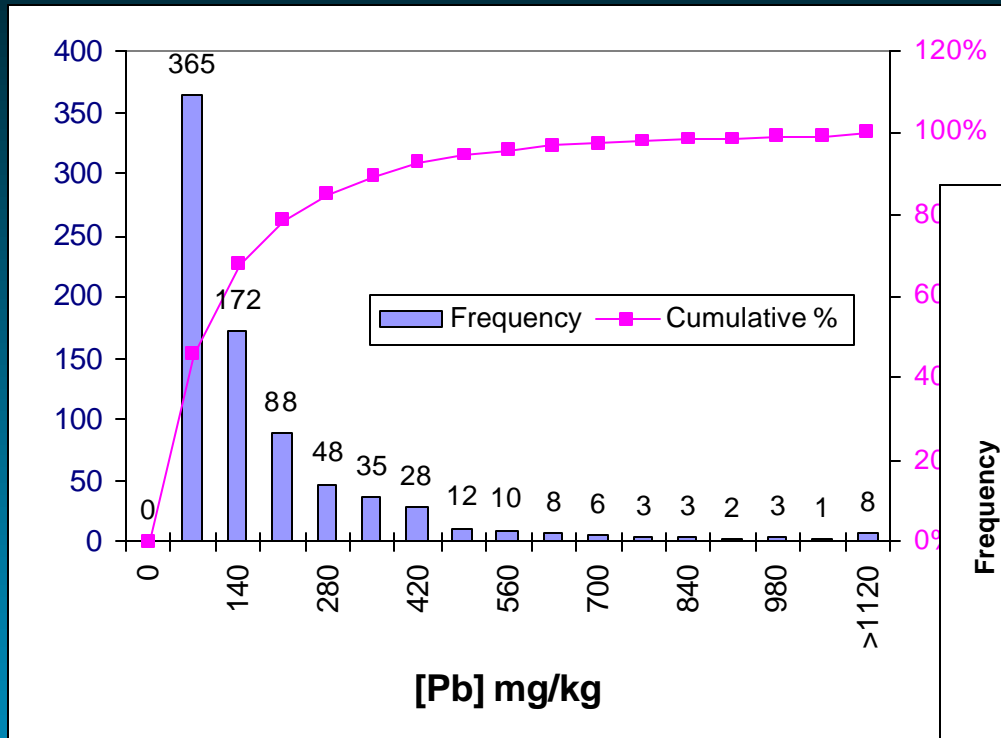
(Non parametric results between
97.1 and 194.7 mg/kg)

Example 2 - Statistics

- ◆ Soil Metal M
Conc. (mg/kg)
- ◆ 0.0 m to 0.3 m
- ◆ N = 792
- ◆ 365 unique
locations

	[M]	Ln([M])
Mean	154.4	4.39
Median	80.0	4.38
s.d.	232.8	1.14
Kurtosis	38.5	- 0.27
Max.	2810.0	7.94
Min.	0.9	- 0.11

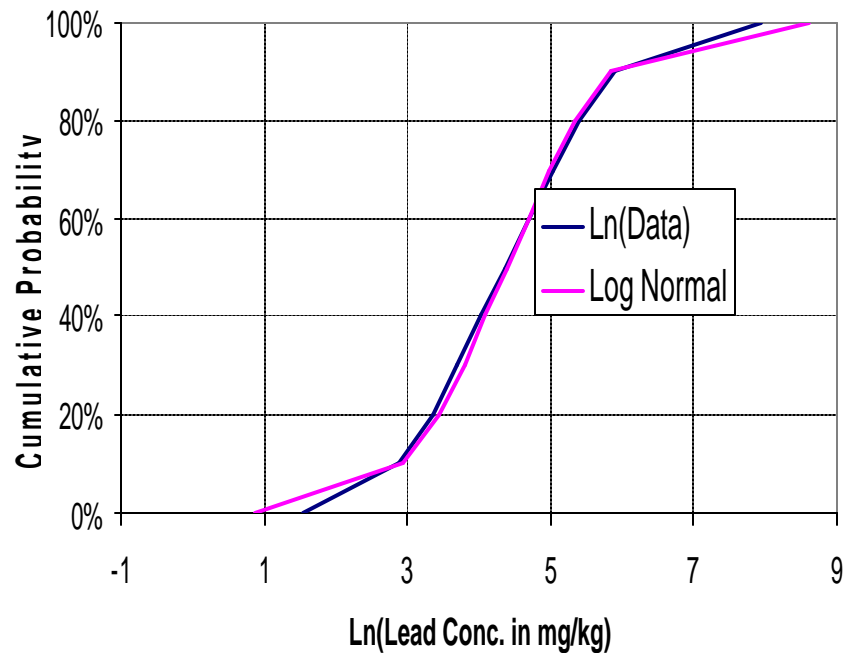
Example 2.1 - PDF's



O'CONNOR ASSOCIATES Z

Feb. 2005

Example 2.2 OAEI Analysis



- N = 792
- ~ Lognormal

UCL	Log Normal	Normal
90%	85.0	165.1
95%	86.3	168.1

Example 2.3 Pro-UCL

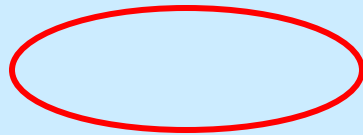
Lilliefors Statistic = 0.036, Critical Value(0.05) = 0.031, Data not Lognormal
Standardized Ordered Observations

Slope = 0.9984, Intercept = 0.0000, Correlation, R = 0.99614900

N = 792, Mean = 4.3909, Sd = 1.1360

Lognormal Q-Q Plot for Metal C

Theoretical Quantiles (Standard Normal)



O'CONNOR ASSOCIATES Z

Example 2.4 Pro-UCL Results

◆ **Data are non-parametric**

◆ **Recommended UCL's:**

- **90% - No recommendation**

- **95% - 206.4 mg/kg**

**(Non parametric results between
168.3 and 206.4 mg/kg)**

Estimating UCL – Lessons Learned

- ◆ **Need to find the underlying PDF**
 - Remove outliers
- ◆ **Consider non-parametric methods**
- ◆ **When using “canned” software**
 - Critical look at results – do they make sense

◆ **Certainly, thank YOU for your
attention!**

Sample Size

◆ For estimating the mean of exposure data:

< 10	Poor
10 – 20	Somewhat better
20 – 30	Fairly Consistent

EPA (1992)

O'CONNOR ASSOCIATES Z

Why Use Mean or Average?

- ◆ Toxicity criteria are based on lifetime average exposure
- ◆ Average concentration more appropriate for exposure over time
- ◆ NOT be used when assessing acute exposure

EPA Pub. 9285.7-08I (May 1992)

O'CONNOR ASSOCIATES Z