AIHA Virtual Seminar, Exposure Assessment Strategies

Statistical Methods Review

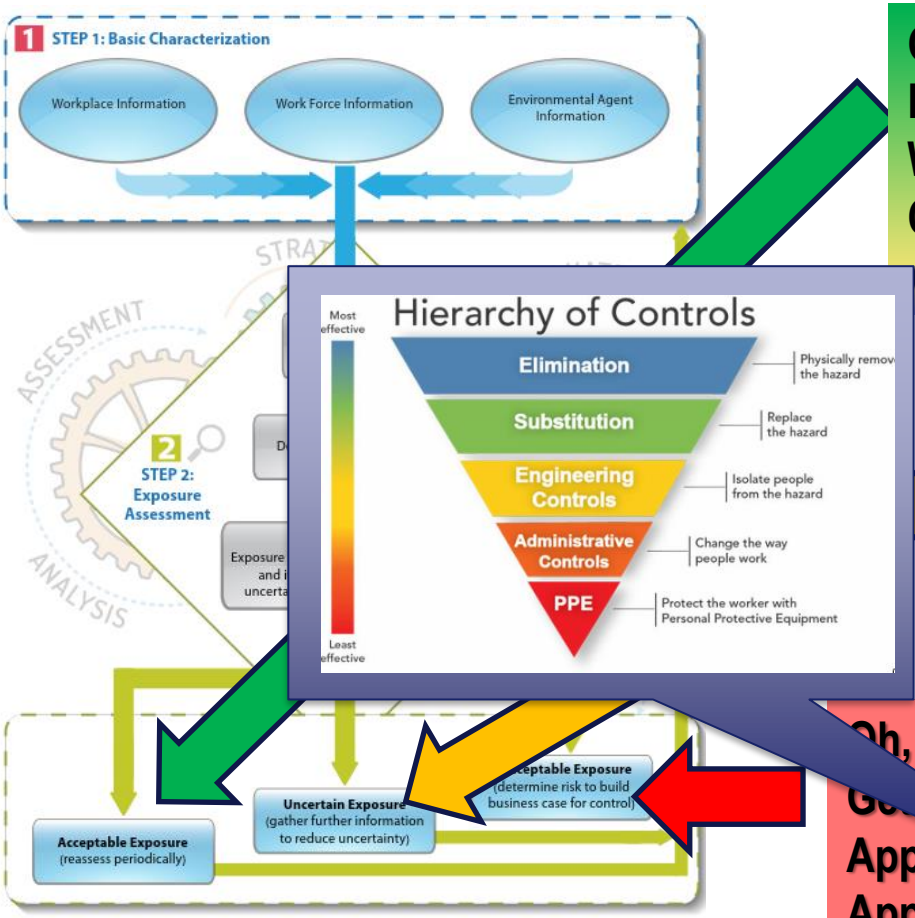
Thomas W. Armstrong, CIH, PhD, FAIHA

TWA8HR Occupational Hygiene Consulting, LLC
TWA8HR@GMAIL.COM

CONTENTS OVERVIEW

- Survey design implications and decision rules impact
- Rules of thumb for low n data sets
- Descriptive, inferential, and Bayesian approaches
- Introduction to several tools for data analysis
  - IH Stat – descriptive and inferential statistics
  - Two Bayesian tools - descriptive and inferential statistics with Bayesian methods
    - IHDA – Student version
    - EXPOSTATS

WHAT DECISIONS MIGHT YOU MAKE WITH SURVEY DATA?



Great! But, is it a TRUE finding?
False NEGATIVES can be *DANGEROUS!*
What should you do?
Confirm with more data?
Apply statistics?

Hmm! A common finding!
What do you do? Get more data?
Apply statistics? Control?

Oh, NO! Is it REALLY true? What do we do?
Get more data?
Apply Statistics?
Apply interim or final controls?

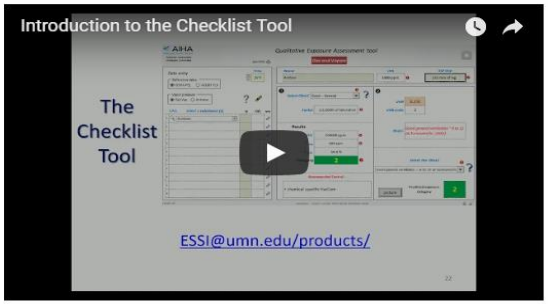
QUALITATIVE (I.E., PROFESSIONAL JUDGEMENT DECISIONS)

- Most exposure assessment decisions are with n=0
- Without specific training and use of “rules of thumb” almost 2/3 of judgements with n=0 may be WRONG and usually LOW.
- With data and statistical tools, the decisions may improve
- We will not cover the “rules of thumb” for n=0 here since we are addressing STATISTICS and n=1 or more

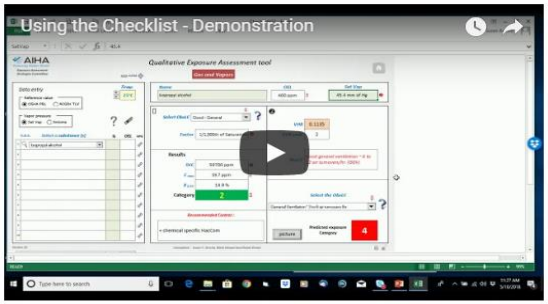
Check at **http://essi.umn.edu/** tab RESOURCES, then PRODUCTS for

More about an IH Qualitative Exposure Assessment Tool – Checklist

INTRODUCTION TO THE CHECKLIST TOOL



USING THE CHECKLIST – DEMONSTRATION



FIRST, A BIT ABOUT DATA QUALITY FOR STATISTICS

- Do you actually have relevant data?
- Are the data from a stationary population?
- Are the samples from a mostly random or stratified mostly random survey strategy?
- How good is your Similar Exposure Group (SEG) formation process?
  - SWAG from an office?
  - Walkthrough survey?
  - Walkthrough And discussions with workers?
  - If reasonably good, statistical methods we will discuss can help show if it is reasonably an SEG

Let's take a look

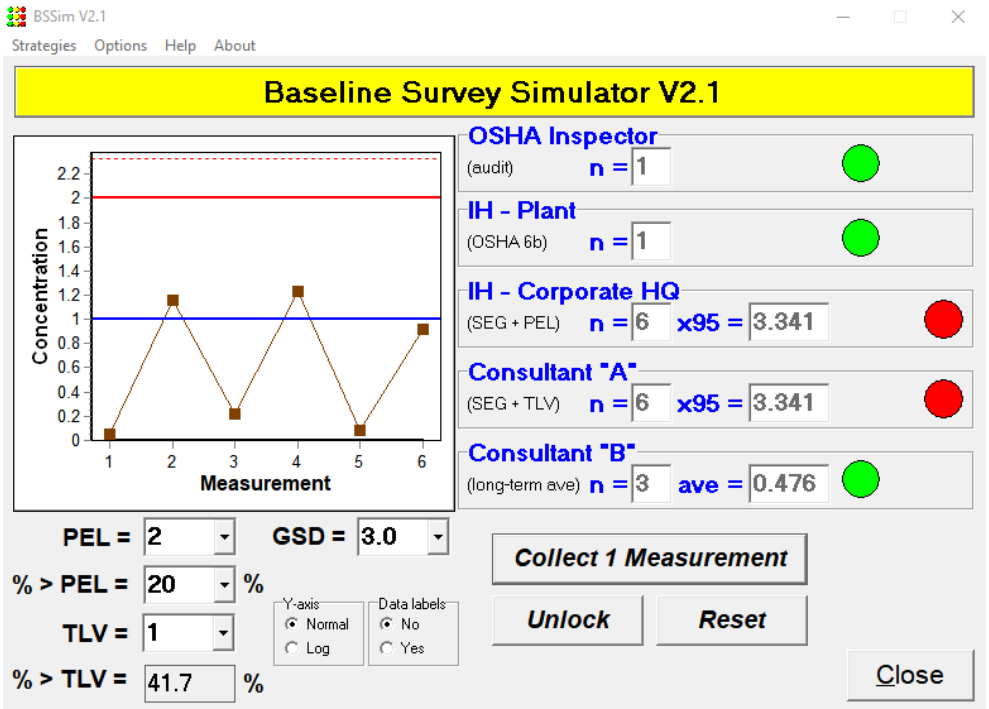
Adapted from BSSim by Paul Hewett (explained in more detain in the software:https://www.easinc.co/

Example Survey Strategies and Decision

- An OSHA inspector, who considers acceptable any single measurement that is equal to or less than the PEL [Can be one sample from one day]
- A plant IH who recommends the least complicated strategy permitted by the OSHA 6b (single substance) regulations. [If the initial C is < Action Limit (AL), then the work environment appears acceptable, If AL <= C <= PEL, then collect additional measurements until 2 consecutive measurements are < AL, or any C > PEL]
- A corporate IH who follows that the AIHA exposure assessment model with the OSHA PEL with a minimum of six samples [Estimate the SEG upper percentile (e.g., 95th percentile)]
- Consultant A)uses the AIHA model, but with the lower TLV® with a minimum of six samples
- Consultant B believes that nearly all 8-hour TWA OELs can be interpreted as the upper limit for a worker's long-term, lifetime average exposure. Single overexposures are unimportant and with a minimum of three samples

HOW MIGHT A SURVEY STRATEGY IMPACT OUR DECISIONS?

Illustration of a "What if" approach Using *Baseline Survey Simulator*
Thanks to Paul Hewett, https://www.easinc.co
I set the "conditions" of the PEL and the 20% over the PEL. 20%!!



What if … we survey a number of times, say 10 for the indicated number of samples (n)

- What might we conclude for n=1?
- OSHA OK for 9 out of 10 runs at n=1!
- Plant IH Also OK for 9 out of 10 at n=1!

And then, for n=6, ONE RUN!

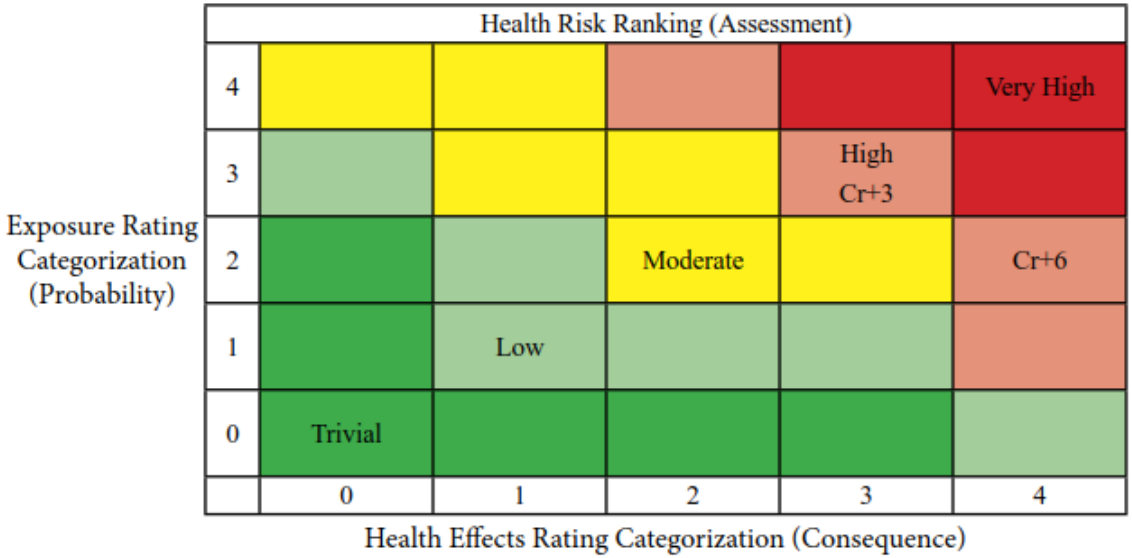
- OSHA OK, Plant IH OK
- Corporate IH PROBLEM!
- Consultant A PROBELEM!
- Consultant B OK!

If you try this, due to random samples, your results may be strikingly different. At n=1 you MIGHT get a > PEL sample. In 9 out of ten tries I did not.

Run Baseline Strategy Simulator

AIHA EXPOSURE AND RISK MATRIX

SEG Exposure Risk Rating	95th Percentile Relative to OEL
0	<1% of OEL (95th percentile less than 0.01 x OEL)
1	<10% of OEL (95th percentile less than 0.1 x OEL)
2	10-50% of OEL (95th percentile between 0.1 x OEL and 0.5 x OEL)
3	50-100% of OEL (95th percentile between 0.5 x OEL and 1.0 x OEL)
4	>100% of OEL (95th percentile greater than 1.0 x OEL)



How good is the OEL?

Dated OSHA PEL?

ACGIH TLV? Under review?

Other?

Well documented?

How "severe" is the adverse effect?

Quick Rule of Thumb Decision Rules

(Assuming log-normally distributed DATA)

1. $\hat{X}_{95\%} = GM \times GSD^{1.645}$

2. Low n, ANY value > OEL = problem

3. Estimate the median value (about the middle of the range) and use a multiplier
   2X if low spread,
   4X if moderate spread,
   6X if large spread

Est'd GSD	Multiplier
1.5	1.95
2.0	3.13
2.5	4.51
3.0	6.09

Variability and Multiplier

Low 2

Medium 4

High 6

What Can We Decide? **OEL = 2.** These are from a randomly generated distribution with parameters set by me for this exercise

- Survey 1, value 0.22
- Survey 2, values 0.65, 0.92
- Survey 3, values 0.87, 0.23, 1.75
- Survey 4, values 0.12, 1.31, 0.34, 1.54
- Survey 5, values 0.07, 0.83, 0.14, 0.48, 0.10
- Survey 6, values 2.93, 0.23, 0.18, 0.74, 0.44, 1.98

SEG Exposure Risk Rating	95th Percentile Relative to OEL
0	<1% of OEL (95th percentile less than 0.01 x OEL)
1	<10% of OEL (95th percentile less than 0.1 x OEL)
2	10-50% of OEL (95th percentile between 0.1 x OEL and 0.5 x OEL)
3	50-100% of OEL (95th percentile between 0.5 x OEL and 1.0 x OEL)
4	>100% of OEL (95th percentile greater than 1.0 x OEL)

Variability & Multiplier

Low 2, Medium 4, **High 6**

I specified the distribution as: "True" AM = 1.8 with GM 1.0, GSD 3. For n=100, UCL95% on AM = 2.3, LTL on AM = 1.5, 95th %ile of the distribution is approximately 3.6. The OEL Exceedance Fraction is about 17%.

Descriptive statistics deal with the data you have

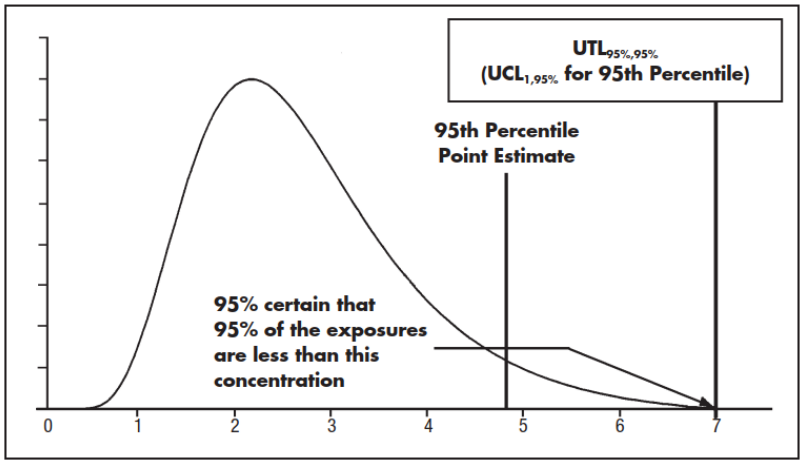
- Number of samples
- Maximum value
- Minimum value
- Percent of measurements above the OEL
- Mean value
- Standard deviation of the data
- Mean of log-transformed data
- Standard deviation of the log-transformed data
- Geometric mean
- Geometric standard deviation

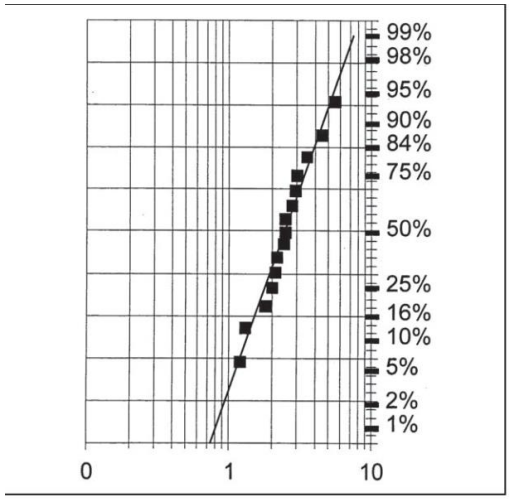
With low n, you really do not know much and cannot infer with confidence

INFERENTIAL STATISTICS….

Make predictions about the POPULATION of exposures from the SAMPLED exposures.

- Mean exposure and its confidence interval
- 95th percentile exposure and the confidence interval on that value
- Exceedance fraction, more
- Low n, wide confidence intervals





Logprobability plots
can tell you a lot about the data set

DID YOU REMEMBER TO BRING A CALCULATOR?

Let's check if you remember how to calculate the GM, GSD, exceedance fraction for a presumed lognormal data set of n=10.

Here are the numbers.

0.98, 0.87, 3.01, 2.50, 2.65,

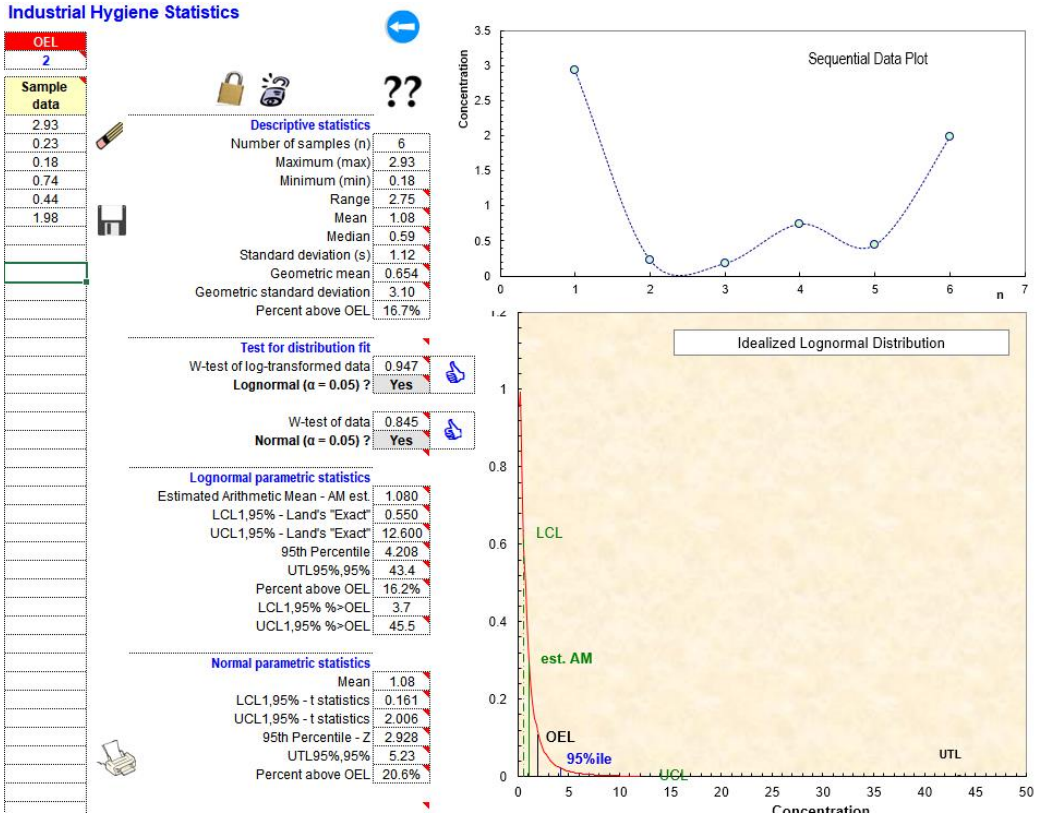
1.26, 6.76, 3.59, 0.41, 0.34

You may begin!

JUST KIDDING!
We have great software tools to do the heavy lifting!
Now we will show a few of the tools and their utility!
Give thanks to the developers…

HERE IS IH STAT, WITH THE SURVEY 6, N=6 DATA

Isn't this slick?! This is the n=6 data



LET'S RUN IH STAT

Let's take a look at the n=5 data values
0.07, 0.83, 0.14, 0.48, 0.10

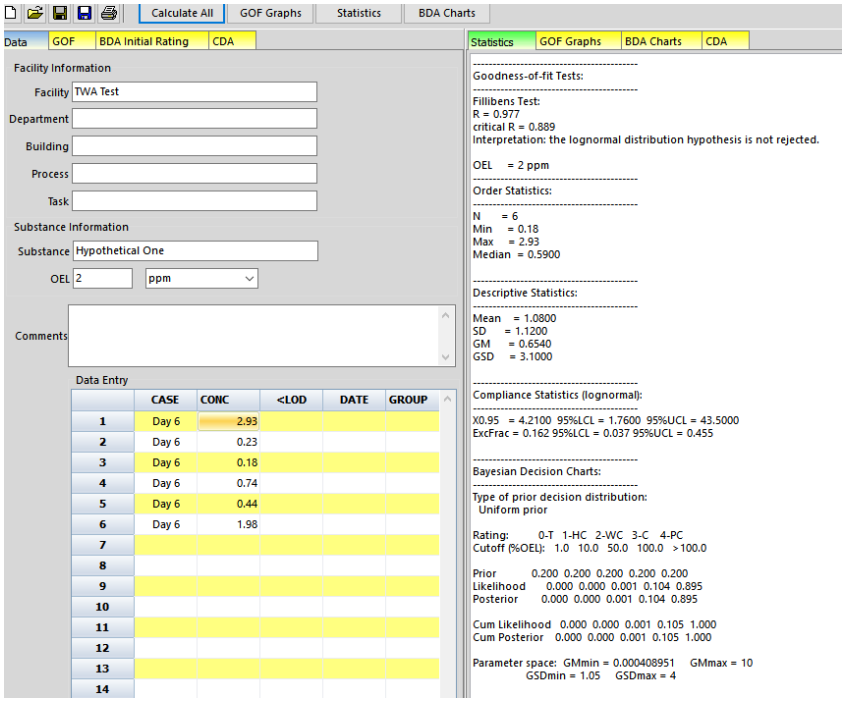
Run IH Stat

HERE ARE THE N=6 DATA

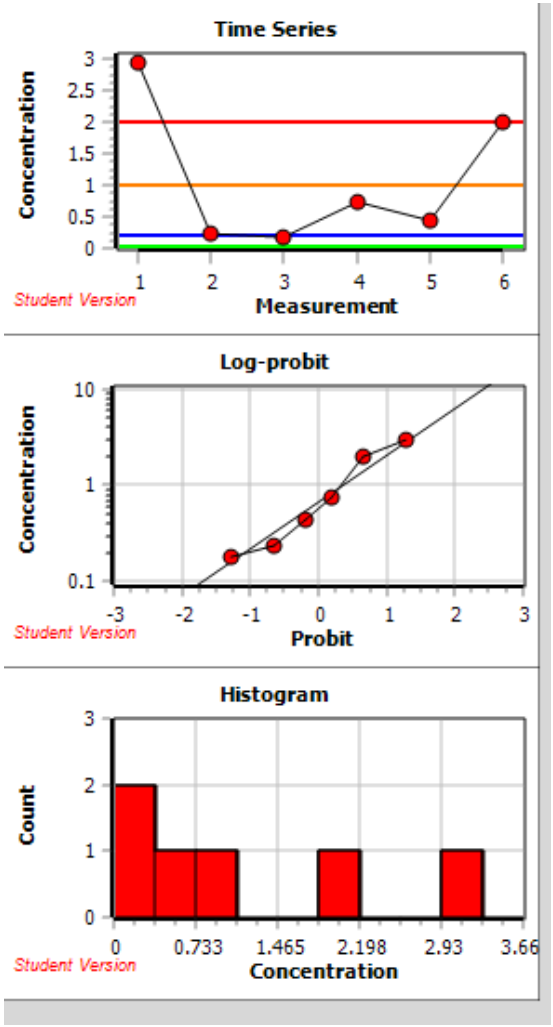
…. with a Bayesian Tool, IHDA Student Version

(Thanks to Paul Hewett, https://www.easinc.co/ihda-student/)

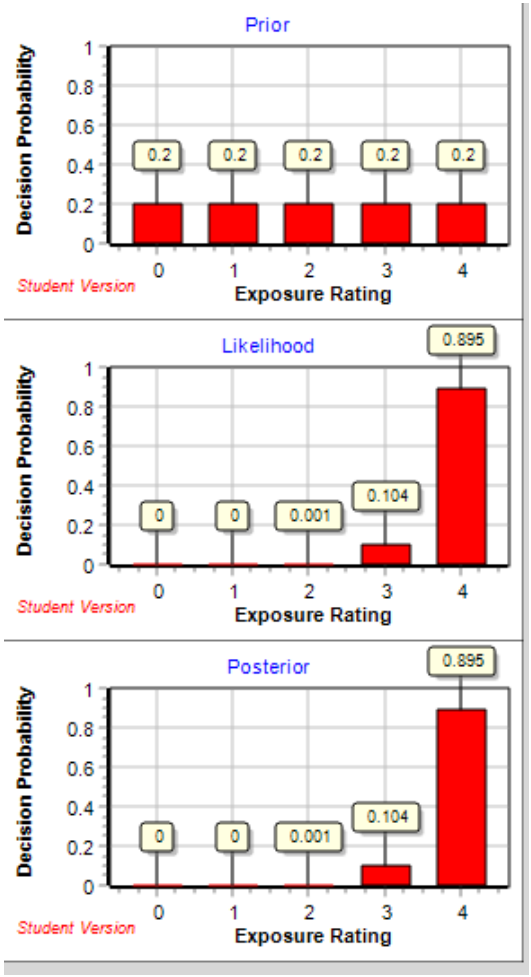
The design is set to give CATEGORICAL probabilities



MORE FROM IHDA_STUDENT

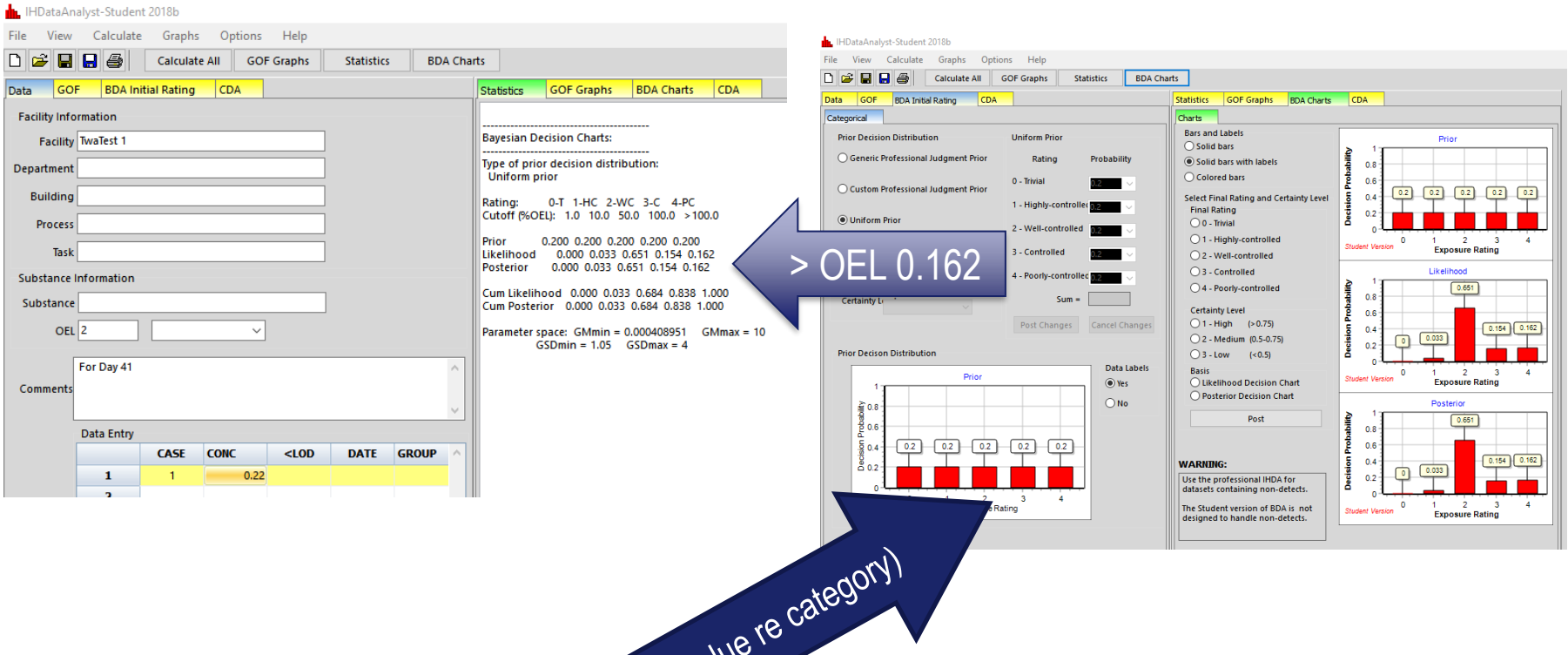


Bayesian Decisions Analysis Charts



BAYESIAN TOOL, IHDA STUDENT VERSION FOR N=1

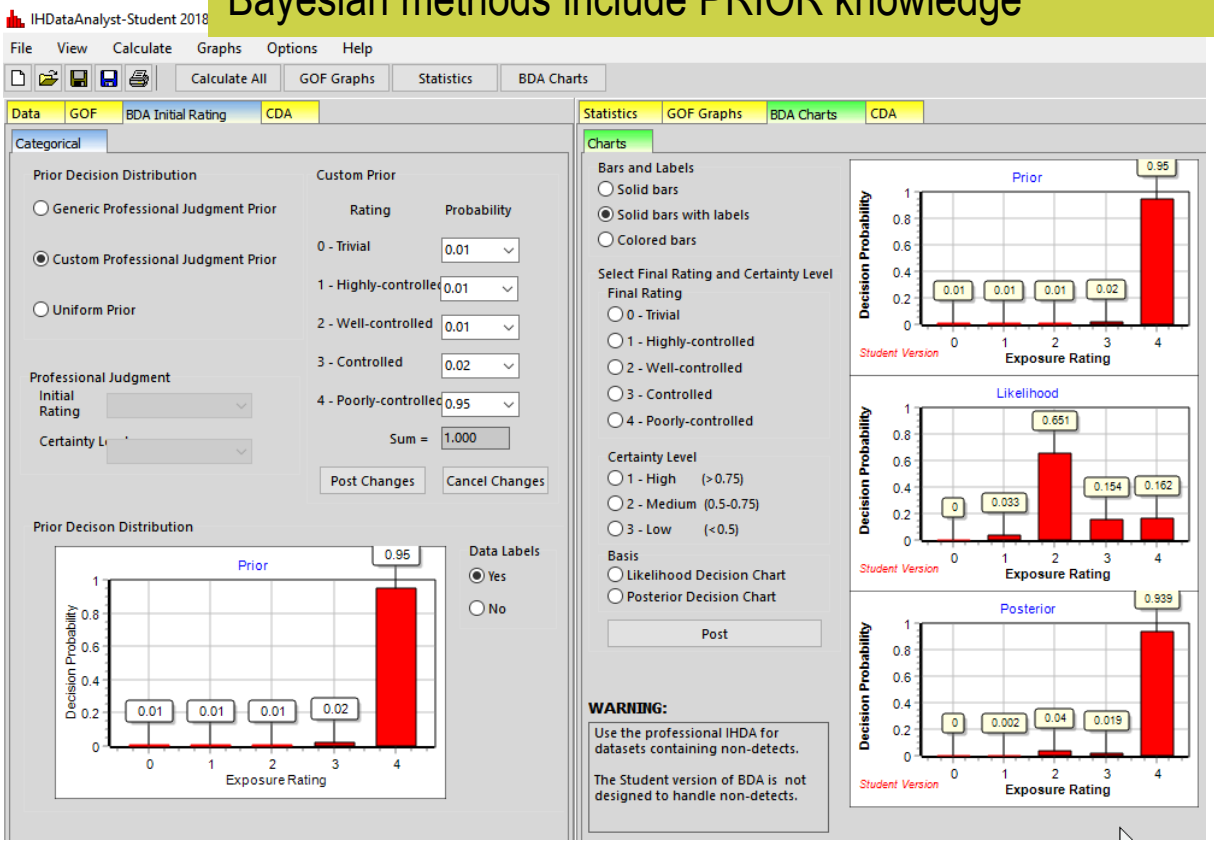
SAMPLE IS 0.22, OEL IS 2.



IN A BAYESIAN ANALYSIS, "WHAT IF…."

We say IN OUR OPINION the n=1 value is off, and is way low.
We can use a "more informed" prior.

This is a key to Bayesian power. Frequentist statistics use current data only. Bayesian methods include PRIOR knowledge



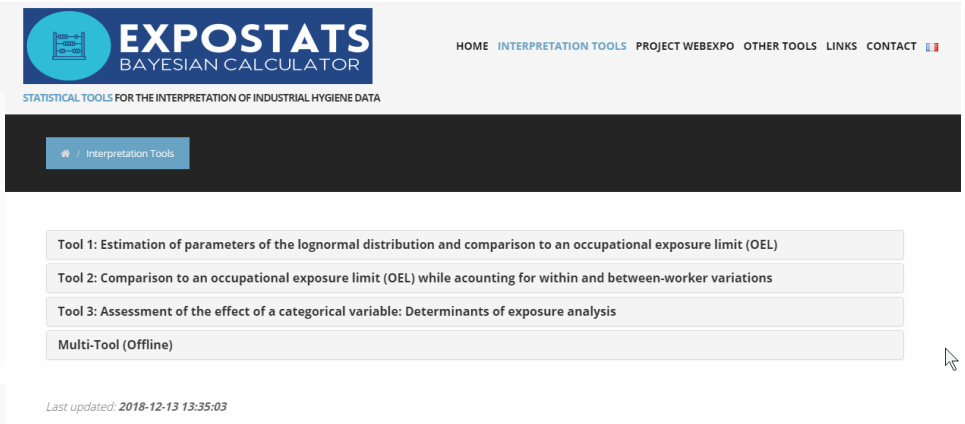
We may not always be this MATHEMATICAL in Industrial Hygiene when we use judgement and opinion with limited data.

MAYBE WE SHOULD BE!!!!

Note admitting "we do not know" in this situation of n=1, we came to a "safe" conclusion with the "uniform" prior or our highly skeptical prior

THERE IS ANOTHER BAYESIAN TOOL AVAILABLE
HTTP://WWW.EXPOSTATS.CA/SITE/EN/INDEX.HTML
WEB BASED BUT OFFLINE VERSION AVAILABLE

Home Page



Interpretation Tools Tab

Currently EXPOSATS has a preconfigured prior. It works well. A future version may have more prior specification flexibility

EXPOSTATS.CA DATA ARE FROM THE N=6 SET

Calculation parameters

Exposure limit: 2

Exposure limit multiplier: 1

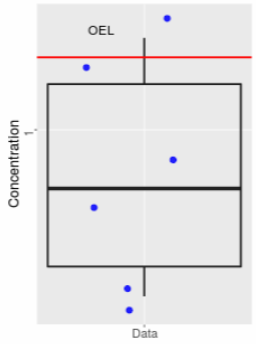
Credible interval probability: 90

Overexposure risk threshold: 5

Data:
2.93
0.23
0.18
0.74
0.44
1.98

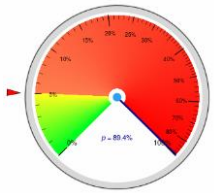
Descriptive statistics

parameter	value
n	6
Proportion censored	0
Minimum	0.18
25th percentile	0.283
Median	0.59
75th percentile	1.67
Maximum	2.93
Proportion >OEL	17 %
Arithmetic mean	1.08
Arithmetic standard deviation	1.12
Coefficient of variation	104 %
Geometric mean	0.654
Geometric standard deviation	3.1



Risk analysis based on the exceedance fraction

Exceedance threshold: 5

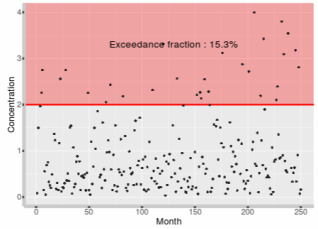


Risk decision

- Criterion defining overexposure:
  Exceedance fraction ≥ 5%
- Uncertainty management : based on the Bayesian model, the probability that this criterion is met (overexposure risk) is:
  89.4 %
- Uncertainty management : The probability of overexposure (overexposure risk) should be:
  < 5%
- As a consequence, the current situation is declared:
  Poorly controlled

Sequential plot

The sequential plot presents the estimated exposure distribution assuming 250 exposure measurements have been collected. If the measurements represent 8h TWA values, this would represent approximately a full year of exposure. The OEL is shown as a red horizontal line.



Expostats continued

Estimated parameters

Estimated parameters - Distribution

The estimate of the geometric mean and its interval of credibility (90%) are: **0.65 [0.29 - 1.5]**

The geometric standard deviation estimate and its credibility interval (90%) are: **3 [2.1 - 6.2]**

Estimated parameters - Exceeding fraction

The estimate is: **15.3%**

The credibility interval (90%) is: **[3.23 - 41]**

Estimated parameters - 95th percentile

The estimate is: **4**

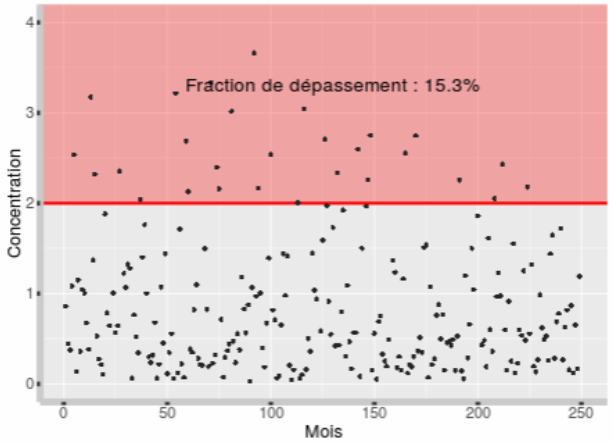
The credibility interval (90%) is: **[1.68 - 18.2]**

Estimated Parameters - Arithmetic Mean

The estimate is: **1.23**

The credibility interval (90%) is: **[0.572 - 4.92]**

Sequential graphic



The sequential graph represents the estimated distribution of exposures assuming 250 measurements were taken. If the measurements represent 8h weighted values, this is approximately a year of exposure. The ELV is in red.

EXPOSTATS.CA FOR N=1, 0.22, ESTIMATED PARAMETERS (BAYESIAN)

Calculation parameters

Exposure limit: 2

Exposure limit multiplier: 1

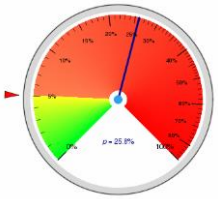
Credible interval probability: 90

Overexposure risk threshold: 5

Data: 0.22

Risk analysis based on the exceedance fraction

Exceedance threshold: 5

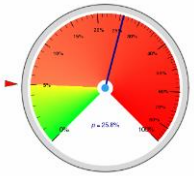


Risk decision

- Criterion defining overexposure:
  - Exceedance fraction ≥ 5%
- Uncertainty management : based on the Bayesian model, the probability that this criterion is met (overexposure risk) is:
  - 25.8 %
- Uncertainty management : The probability of overexposure (overexposure risk) should be:
  - < 5%
- As a consequence, the current situation is declared:
  - Poorly controlled

Risk analysis based on the 95th percentile

Selection of the critical percentile: 95

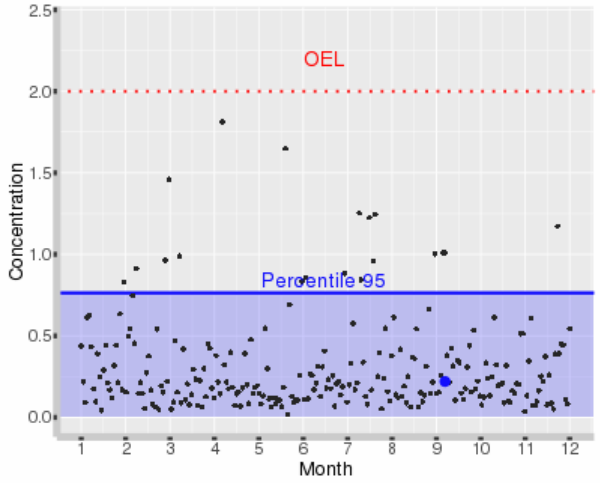


Risk decision

- Criterion defining overexposure:
  - 95th percentile ≥ OEL
- Uncertainty management : based on the Bayesian model, the probability that this criterion is met (overexposure risk) is:
  - 25.8 %
- Uncertainty management : The probability of overexposure (overexposure risk) should be:
  - < 5%
- As a consequence, the current situation is declared:
  - Poorly controlled

Sequential plot

Assuming the underlying distribution of exposure would occur over a year, the graph below shows a simple sequential plot of the approximate 240 days of exposure. The OEL is shown as a dashed red horizontal line. The point estimate of the selected percentile is shown as a blue continuous line.



- Measured, randomly placed

EXPOSTATS Tool 2. Within and Between Worker Variability (data by J. Lavoue)

Calculation parameter

Exposure limit
100

Exposure limit multiplier
1

Credible interval probability
90

Overexposure risk threshold
30

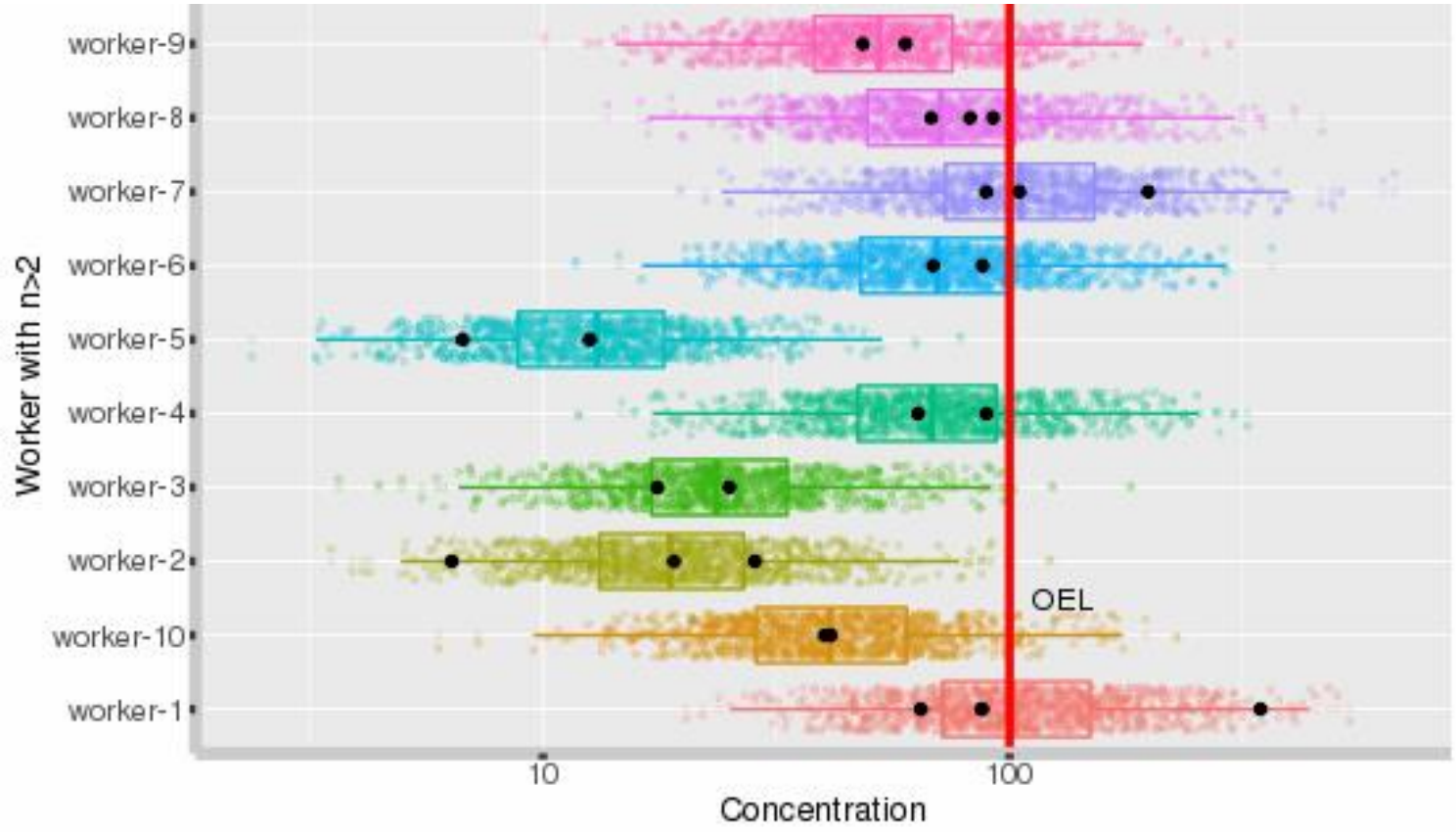
Rb ratio coverage
80

Within-worker correlation threshold
0.2

Data

40.1	worker-10
28.5	worker-2
48.5	worker-9
87	worker-1
6.73	worker-5
105	worker-7
64.6	worker-1
87.5	worker-6
6.38	worker-2
68.6	worker-6

Differences Between Workers (with at least n=3)



EXPOSTATS WORKER BY WORKER ANALYSIS

Risk analysis based on the exceedance fraction

Exceedance threshold

5

Group analysis | Between worker analysis | Individual worker results

In this panel you can obtain individual risk assessment for any worker. Workers can be selected using the dropdown menu below. Be aware that important uncertainty will be associated with these estimates, especially those based on very few measurement. The results are based on the Bayesian analysis of variance of the whole group, not restricted to individual worker data.

Select a worker

worker-1

worker-1
worker-10
worker-2
worker-3
worker-4
worker-5
worker-6

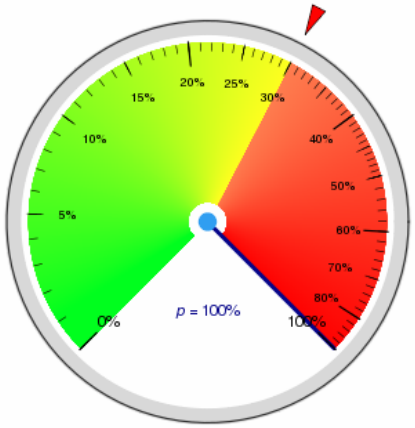
…esian model, the probability that this criterion is met (overexposure risk) is:

- Uncertainty management : The probability of overexposure (overexposure risk) should be:

  < 30%

- As a consequence, the current situation is declared:

  Poorly controlled



Do You Remember our Baseline Strategy Simulator Exercise? With n=1 or 2, or … the BAYESIAN tools may have delivered more reliable information to support the critical decisions.

THERE WILL BE PROBLEMATIC DATA SETS

- Data are neither normal nor lognormal
- Data have a WILD GSD
  - Is it a mixed distribution? Should data be stratified? HOW could you check this?
  - Did a truly exceptional event occur? Should a data point be censured?
  - Is one worker different from others in the group?
- Limit of detection samples are quite prevalent
  - LOD replacement methods
  - Bayesian tools including censored data approaches
- The analytical method is very imprecise or biased
  - Use a better method next time
  - Adjust for the bias
- The samples are of mixed and significantly different duration and were not adjusted to a standard STEL or TWA duration
  - Stratify the data
  - Adjust to a standard duration
  - Can you make legitimate and suitable assumptions on unsampled time?
- Other problems

STATISTICAL METHODS NOT COVERED
(CONSIDERED ADVANCED TOPICS REQUIRING A FULL PDC)

1. Within and between worker variability (several slides prior)
2. Dealing with prevalent non-detects
3. Nonparametric methods
4. Analysis of variance for improving SEGs
5. Goodness of fit tests and importance for inferential statistics

IMHO, EXPOSTATS with Bayesian logic,

a) gives robust approaches to 1,

b) handles 2 well,

c) reduces the need for 3 (and non-parametric methods usually need a good sized data set),

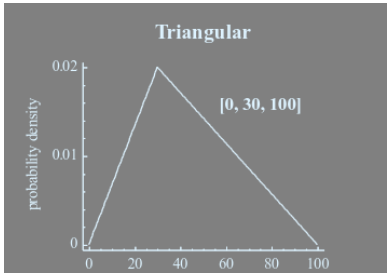
d) does well with 4 via within and between worker analysis (Tool 2) and group to group differences (Tool 3).

e) Covers 5 as does IH Stat and IHDA

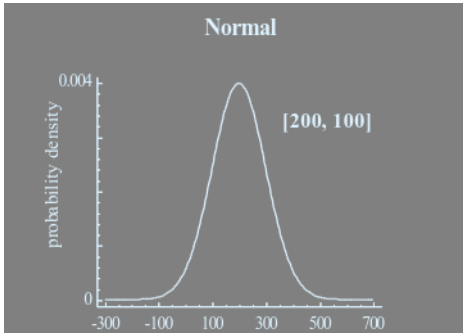
COMMONLY USEFUL PROBABILITY DISTRIBUTIONS

- Uniform (Just a range, n1 to n2)

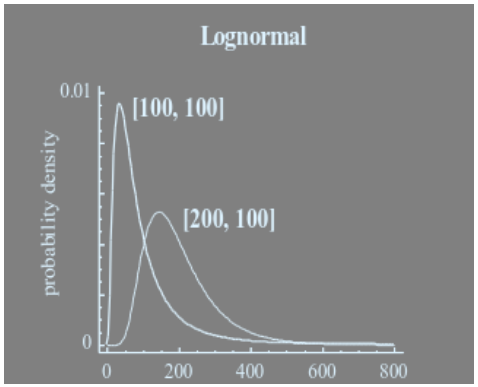
- Triangular (Min, Max, Most Likely)



- Normal (AM, SD)

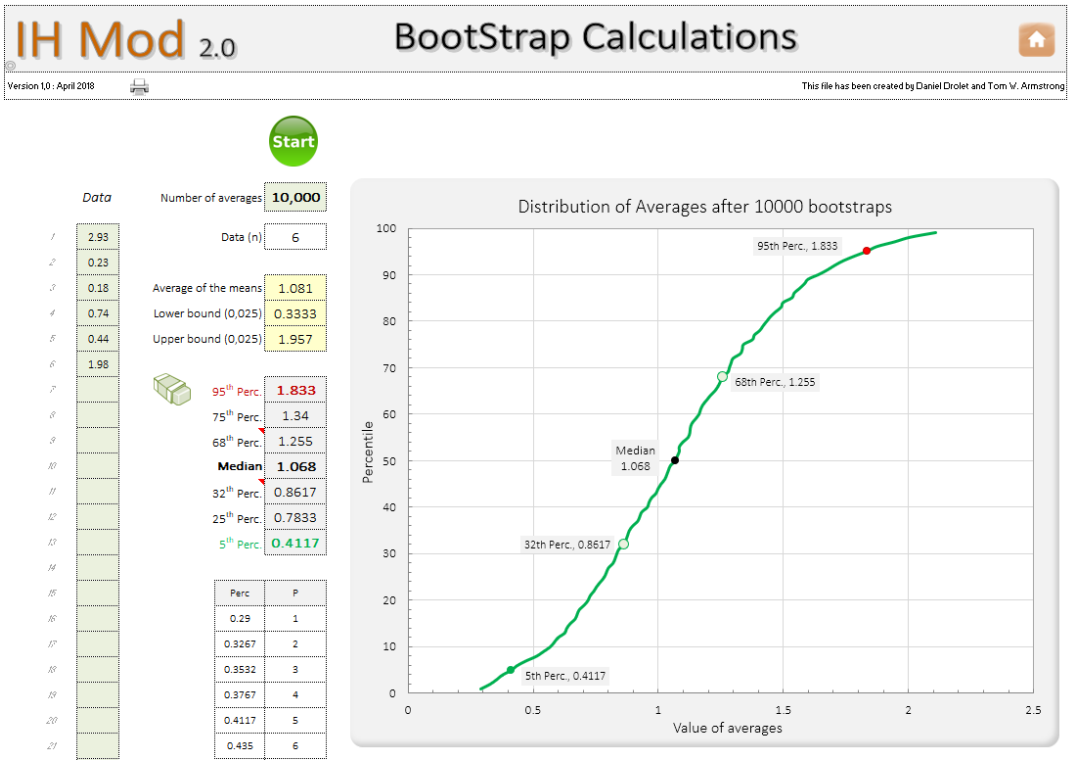


- LogNormal (AM or GM, GSD



WHAT IF YOUR DATA DO NOT FIT A DISTRIBUTION FOR INFERENTIAL STATISTICS?

TRY IH MOD SUPPORT FILE, BOOTSTRAP FOR MEAN AND ITS CI



At AIHA EASC Volunteer group Web page

https://www.aiha.org/get-involved/VolunteerGroups/Pages/Exposure-Assessment-Strategies-Committee.aspx

FOR MORE INFORMATION….

- AIHA Body of Knowledge Occupational Exposure Risk Assessment/Management
- https://www.aiha.org/membercenter/SynergistArchives/2014SynergistArchives/Pages/JudgmentDay-01.aspx
- https://www.aiha.org/publications-and-resources/TheSynergist/SynergistNow/Lists/Posts/Post.aspx?ID=22
- Fundamentals: Bayes' Theorem https://www.youtube.com/watch?v=OqmJhPQYRc8
- EXPOSTATS.CA http://www.expostats.ca/site/en/tools.html

  (Expostats: A Bayesian Toolkit to Aid the Interpretation of Occupational Exposure Measurements, Annals of Work Exposures and Health, 2018, 1–13, doi: 10.1093/annweh/wxy100, Jérôme Lavoué, et al., E-mail: jerome.lavoue@umontreal.ca)
- And, of course, A Strategy for Assessing and Managing Occupational Exposures,

  Bullock et al., AIHA Press