

# **EPA-SAP: Statistical Methods for Use of Composite Data in Acute Dietary Exposure Assessment**

May 25, 1999  
Washington, D.C.

By  
Mark E. Whalon and Patrick S. Bills  
Michigan State University  
Center for Integrated Plant Systems

Generally, I am very pleased that EPA has taken a stochastic approach to estimating acute risk from residue data. Despite my comments and problem identification below, this is the only reasonable approach given the major over estimation problems with using empirical distribution functions. It is my recommendation that the Agency move ahead in utilizing the outlined procedure of de-compositing PDP and FDA data for acute dietary risk assessment.

A general concern is exactly how will the Agency regulate around high "outlier" residue data points. Even with lognormal transformation, some of these PDP azinphos-methyl residue data points from apple and peach (see below) for instance are more than 3 standard deviations from the mean. As several authors including the Environmental Working Group have pointed out, these are the true risks for infants, children and adults. My point, what do we really know about these "outlier" residue data points? They are much higher than would be expected given a continuous lognormal distribution, and they are a key to much of the dietary exposure concern.

A 1998 Michigan study was designed to provide GLP measurements that link actual field use of pesticides on randomly sampled farms to farm gate, processing and handling residues. This study was conducted on eight commodities (apples, peaches, blueberries, tart cherries, grapes, asparagus, cucumber and potato). It was a cooperative effort by EPA Region 5, Michigan Department of Agriculture (the regional PDP) and Michigan State University. The resulting data were not nearly as variable as PDP data. The study report is available on the web ([www.cips.msu.edu](http://www.cips.msu.edu)). Samples were split and frozen. When a high residue level was detected a check routine was implemented. This routine checked all the analysis, calculations, sample chain of custody, and pesticide use and usage records. Several of the "outlier" data points were resolved as errors. If a data point was not resolvable by this routine, a second GLP residue analysis was carried out (only on a very few samples was this necessary) on the frozen split sample. At this writing, I am not sure of the exact number of re-samples conducted, but there were not more than two or three. In all of these, the resample residue was moderate or low.

From this experience, I believe that it would be worthwhile to initiate a PDP lab routine that identifies "outlier" residue values within the lab and back-check the process leading to the "outlier" data point. This process may help resolve some of the outliers observed in the data, and it would be interesting to know what proportion this process would address. A second study could use the Michigan split-sample procedure to re-extract, analyze and check the data point. GLP labs already do some variability estimating yet a third commissioned study could replicate measures from large split samples at other GLP labs

and compare the variance especially around "outlier" samples. Again, it would be very interesting to know how these compare.

The point of all of this is to characterize the risk. Is the risk from a large number of moderate residue values or from a handful of "outliers"? If it is from "outliers", what can we learn about them?

### Questions for the scientific advisory panel:

- 1- The assumption that the lognormal statistical distribution actually fits all possible PDP, FDA, registrant-field, processor, state market basket, land grant university "bridging" data sets etc. is a bit premature. What are the total data sets (n) submitted to a distribution fitting procedure so far -- 2, 3, 4, or 5? The inference or assumption deduced from these analysis is that the universe of all possible residue data sets are all represented by the lognormal statistical distribution. Without actually fitting a number of these data sets (say n = 30), we are pressed to actually infer or adopt this assumption under sound statistical inference. Note the analysis of azinphos-methyl PDP data on apples and peaches below. The lognormal transformation of the detect data improves the distribution of the data, but it still fails a normal distribution test (figures 1-3 below). What are the implications of not fitting normality?

The lognormal distribution also has additional assumptions (independence and common variance), have these assumptions actually been tested? If they are not met, what effect will these deviations have on the inferences drawn from the analysis?

- 2- The calculations for the correction of the standard deviation are (page 10) see the analysis below:

$$\sigma = S_{\text{est}} \sqrt{N}$$

When N is small (1-2 = egg plant) the standard deviation is smaller than when N is large (16-20 = small apple or peach) or very large (grapes and raisins). Sensitivity analysis should be performed on N's influence on the calculated standard deviation since this is counterintuitive to the Central Limit Theorum (where:  $\sigma(x) = S_{\text{est}} / \sqrt{N}$ ) in which the standard deviation of the mean decreases with increasing sample number. In this estimation,  $\sigma$  increases with N (number of units in a composite).

- 3- See the comment above.
- 4- If there were a sufficiently large number of single serving data sets, we could test the overestimation of the estimation of residues on single servings (30 sets?). Short of this we can look at the PDP data and note the number of composite residue outliers (stem and leaf displays with calculated confidence intervals) with abnormally high (greater than two or three standard deviations from the mean) composite residue means. In my experience, when a particularly high residue value in state market basket samples is found, follow up yields a range of circumstances: lab error

(contamination, decimal point placement, miscalculation = ppt, ppm, ppb, etc.) or an actual field residue that is high (rate of application, 2x treatment of the same area, etc.). The PDP data does contain some residue values that are very high, and it is these data points that are most interesting from a risk perspective. Therefore, the most reasonable follow-up would be to develop investigation protocols that could actually determine in control tests where these high residues data points are coming from.

- 5- Estimation of N (the number of units in a single serving). Since most serving sizes are developed and standardized in the DEEM software (by age, population, ethnicity, etc.) the probabilistic estimates should already compensate for the population demographic variables. The change in N from melons (part of a fruit) to grapes (clusters to multiple clusters = 25-100's) is related to my comments in 2 above.

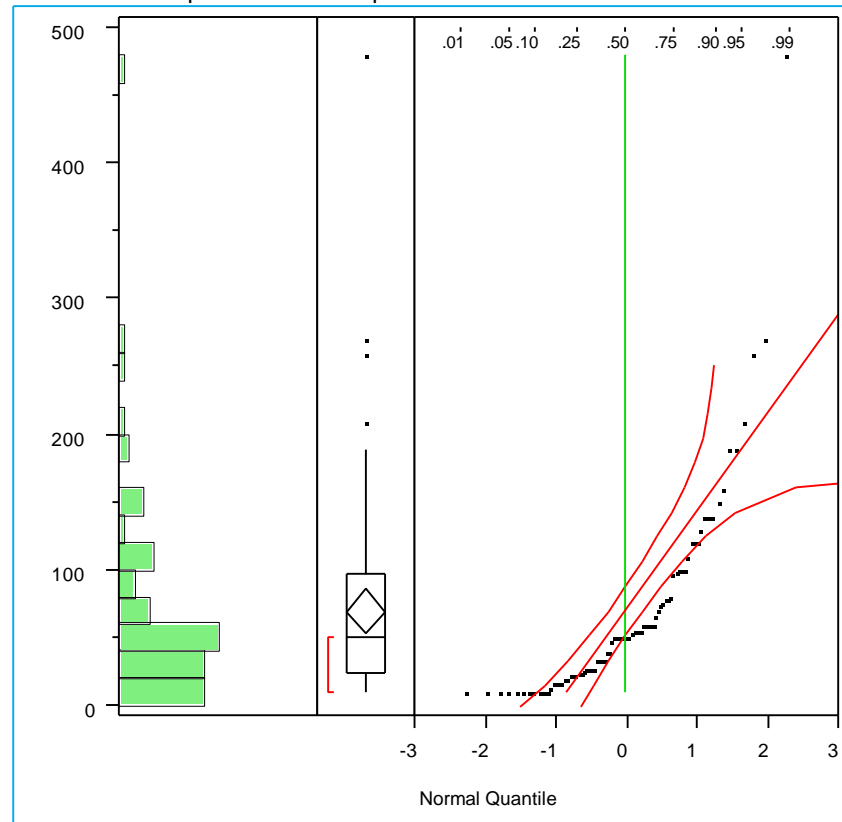
The more critical issue is the counter intuitiveness of the estimate of the standard deviation, as composite unit numbers increase the estimate of  $\sigma$ .

- 6- Is this question relating the surface area to volume ratio? In other words, are residues correlated to water weight or surface area of the fruit or vegetable? One answer is that it depends on the water solubility of the chemistry in question, the movement of the chemical across the fruit or vegetable's leaf surface or cuticle and the other uptake routes (through roots, stems, reproductive tissues, etc) together with the surface area (SA) of residue deposits. For systemic compounds, weight may be a better indicator of residue than surface area. For compounds that are not systemic, surface area probably better reflects residues. In any case, studies of the surface area to volume (V) ratio of living organisms change as follows:  $SA^2 : V^3$ . My guess is that this relationship probably does not matter that much at the extremely low LOQ and LOD that residues are measured at, especially given the sample preparation extraction protocols which tend to collect both systemic and surface residues.

If the question is asking whether or not the N will affect the sensitivity of estimation of residues see number 2 above.

**Figure 1.** Pesticide Data Program 1993 - 1996: Azinphos-Methyl /Peaches Domestic Samples-rawdata ; detects only

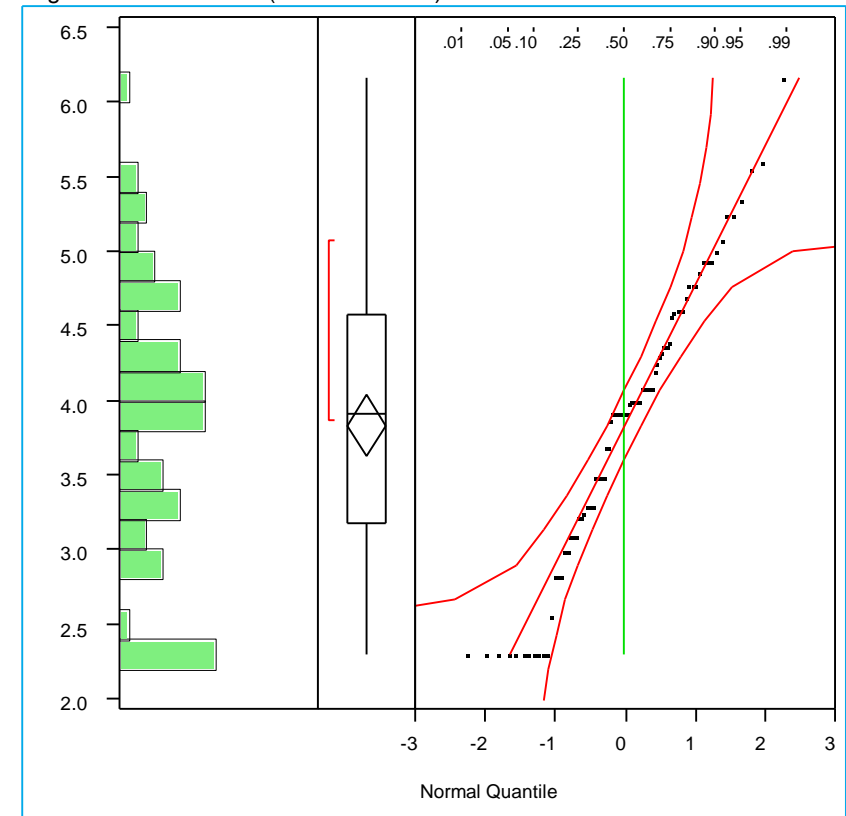
CONCEN: "Composite" PDP Sample - PPB



Mean	70.04878
Std Dev	72.84092
Std Error Mean	8.04393
Upper 95% Mean	86.05375
Lower 95% Mean	54.04381
N	82.00000

Test for Normality Shapiro-Wilk W Test	
W	Prob<W
0.737930	0.0000

Log Transformation: ln (CONCEN PPB)



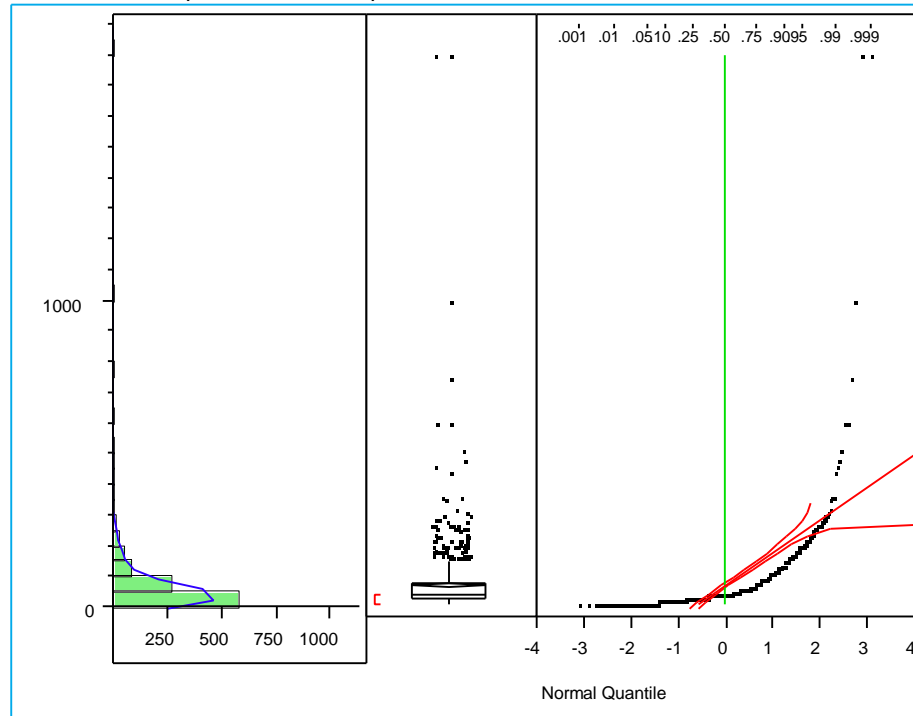
Mean	3.83396
Std Dev	0.93311
Std Error Mean	0.10304
Upper 95% Mean	4.03899
Lower 95% Mean	3.62894
N	82.00000

Test for Normality Shapiro-Wilk W Test	
W	Prob<W
0.953278	0.0145

- The " Test for Normality " tests that the distribution is normal. If the p-value reported is less than .05 (or some other alpha), then you conclude that the distribution is not normal. If you conclude from these tests that the distribution is not normal, it is useful to use the Normal Quantile command in the check border menu to help assess the lack of normality in the distribution.
- Upper 95% Mean and lower 95% Mean are 95% confidence limits about the mean. They define an interval which is very likely to contain the true population mean. If many random samples are drawn from the same population and each 95% confidence interval is determined, you expect 95% of the confidence intervals so computed to contain the true population mean. The upper and lower limits are computed as the sample mean, plus or minus a 97.25% Student's t value multiplied by the standard error of the mean.

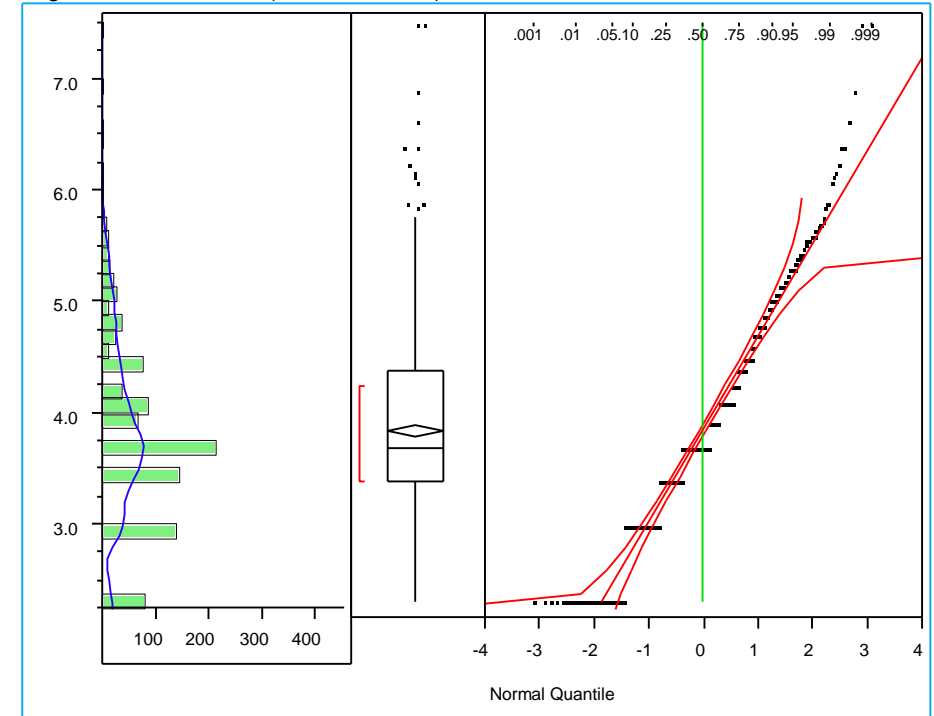
**Figure 2.** Pesticide Data Program 1993 - 1996: Azinphos-Methyl / Apples Domestic Samples-raw data ; detects only

CONCEN: "Composite" PDP Sample - PPB



Mean	70.296
Std Dev	106.959
Std Error Mean	3.306
Upper 95% Mean	76.782
Lower 95% Mean	63.810
N	1047.000
Test for Normality: Shapiro-Wilk W Test	
W	0.438479
Prob<W	0.0000

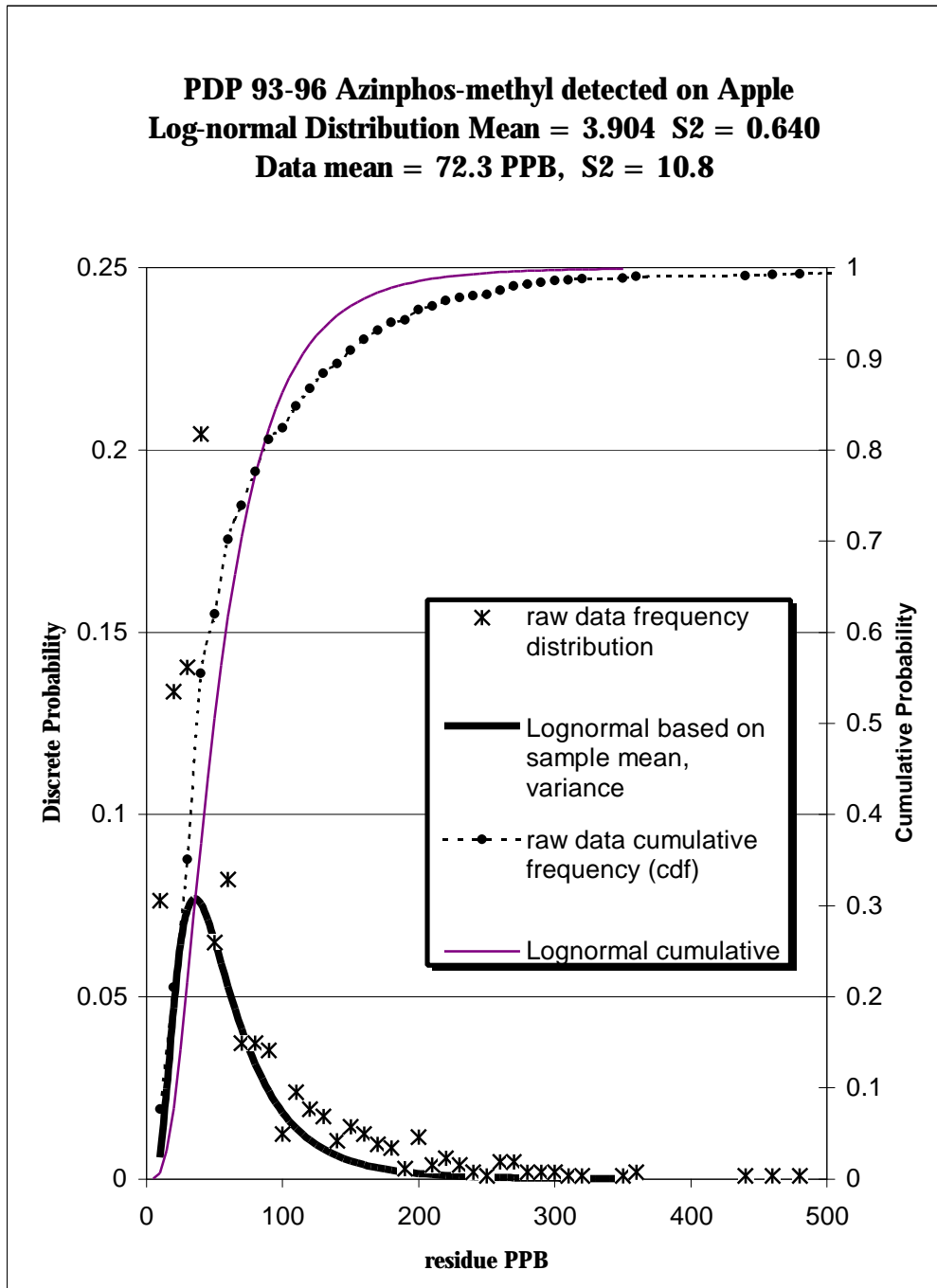
Log Transformation: ln (CONCEN PPB)



Mean	3.844
Std Dev	0.838
Std Error Mean	0.026
Upper 95% Mean	3.895
Lower 95% Mean	3.793
N	1047.000
Test for Normality: Shapiro-Wilk W Test	
W	0.953989
Prob<W	0.0000

- 
- The "Test for Normality" tests that the distribution is normal. If the p-value reported is less than .05 (or some other alpha), then you conclude that the distribution is not normal. If you conclude from these tests that the distribution is not normal, it is useful to use the Normal Quantile command in the check border menu to help assess the lack of normality in the distribution.
- Upper 95% Mean and lower 95% Mean are 95% confidence limits about the mean. They define an interval which is very likely to contain the true population mean. If many random samples are drawn from the same population and each 95% confidence interval is determined, you expect 95% of the confidence intervals so computed to contain the true population mean. The upper and lower limits are computed as the sample mean, plus or minus a 97.25% Student's t value multiplied by the standard error of the mean.

**Figure 3.** Example Lognormal Transformation of Azinphos-methyl PDP Detect data:



**Figure 4.** Sensitivity analysis of N: using the example data presented in the EPA position paper

