D29 What Confidence Do We Have in Confidence Intervals?

Willem A. Schreuder, PhD, Principia Mathematica, Inc, 445 Union Boulevard, Ste 230, Lakewood, CO 80228*

The goals of this presentation are to clarify the correct interpretation of confidence intervals and illustrate how sampling bias and similar factors lead to incorrect confidence limits.

This presentation will impact the forensic science community by informing attendees that confidence intervals are frequently used to ensure compliance with criteria such as maximum contaminant limits. Attendees will learn that sample bias or inappropriate application of upper confidence limits can have significant consequences in evaluating these criteria.

Confidence intervals are widely used to express the certainty with which we can establish that a set of samples exceeds or meets criteria, such as maximum contaminant limits; however, confidence intervals and upper confidence limits are often misunderstood and misapplied.

When considering small sample sets, there may be insufficient information to discern the statistical properties of the data. As a result, many assumptions have to be made regarding important properties, such as the inherent distribution of the data. The situation is further complicated by heterogeneous sample matrices, samples taken at different times, variability introduced by sampling methods, sample preservation, analytical techniques, laboratories, and similar factors that influence data sets encountered in real-world situations.

These sources of uncertainty confound environmental forensic investigations in which investigators often have little control over the data collection efforts. They may be dealing with data sets collected for an event that occurred some time in the past and they can no longer make relevant measurements, or they may be dealing with data collected by another party and they cannot make independent measurements or influence where and how the data are collected. They are simply left with a data set they need to analyze and objectively express the confidence with which they can make various claims.

Further frustrating the establishment of confidence intervals is biased sampling. In some instances, samples are only taken when there is some indication that contaminants are present; for example, staining or when a field instrument such as a photoionization detector indicates the presence of volatile organic compounds. In these instances, forensic investigators have no negative confirmation samples showing background levels. Conversely, sample sets may be biased toward background levels if most of the samples are unrelated to a problem area or there is some other flaw in the sampling protocol, such as sampling too shallow or deep, disturbing the sample, or any number of factors that could cause the samples to be biased low.

Using synthetic data sets for which the statistical properties are known, this presentation explores how these factors contribute to the difficulty of interpreting the data and establishing confidence intervals. This presentation discusses the correct interpretation of the confidence interval, explores how the sample size influences the confidence interval, and illustrates how the distribution of the underlying data is distorted by time, sampling methods, variability in the methods of analysis, and the influence on the confidence intervals. The importance of the sampling protocols is discussed and how sampling bias translates into bias in the confidence interval is demonstrated.

Statistics, Confidence Intervals, Maximum Contaminant Limits

*Presenting Author