Environmental Health Perspectives
Vol. 53, pp. 193–194, 1983

Commentary

Answers to Some Questions Raised by the Presentation on the Graphical Analysis of the Interrelationships Between Waterborne Asbestos, Digestive System Cancer and Population Density

by Michael E. Tarter*

Two interesting questions were raised following my presentation (*1*), and I am pleased to have the opportunity to clarify these issues. The first question dealt with the data in Figure 1. The questioner asked about the interpretation of the axes as they appeared on the graph. In answer to his question it is important to note that all graph coordinate axes pass through the x and y sample means of the displayed data. For example, the mean tract number of the frequency diagram shown in Figure 1 is 216, while the mean number of individuals per square mile is 5270. Thus, the two lines shown in Figure 1 are $X \equiv 216$ and $Y \equiv 5270$ (the symbol "$\equiv$" is the usual mathematical notation for the phrase "is identical to").

The program that we use, GRAFSTAT, is designed to perform certain scaling and positioning operations automatically. Hence, it applies an initial pass through the data to compute the range of the (X,Y) observations (from which the position of the graph's boundaries are calculated). Simultaneously, it calculates the pair $(\bar{X},\bar{Y})$ (from which the axes are positioned), and the standard deviation units. For example, the very small number "2" which appears slightly below the midpoint on the Y-axis of Figure 1, is 2 y standard deviation units from the mean point $(\bar{X},\bar{Y})$.

For normally distributed data, as the sample size increases one would of course expect that the number x of all measurements which lie between plus and minus 1.95 standard deviation units of the mean would approach the value x = 5%. The fact that the data depicted in Figure 1 extends more than five standard deviation above, but less than one standard deviation unit below, the point $(\bar{X},\bar{Y})$ suggests a high degree of skewness for the population density variate. On the other hand, the fairly even spread of data on either side of the mean point in the $\pm\bar{X}$ direction suggests a fairly uniform distribution of the tract number variate.

In the second question I was asked how I concluded that the data in Figure 23 appeared to be composed of two population subgroups of the analysis displayed in Figure 25. The questioner thought that they appeared to be an artifact caused by the sparse data in this region and by the smoothing function employed.

This is a question about smoothing and bifurcation. The method used to smooth the displays shown in Figure 25 is based on the modifications of the Doetsch-Fourier transform technique for spectral decomposition as modified by Tarter and Silvers (*2*). As I stated in this paper, "the choice of λ within Λ of (2.6) will not affect the variances of any component." Parenthetically, it should be mentioned that despite the fact that the 27 references listed at the back of this paper suggest considerable interest in the λ method, there has been in the last 8 years no contention to the above assertion. This assertion implies that one can separate or smooth distributional components in one direction with absolutely no effect on the components in any other direction.

*Department of Biomedical and Environmental Health Sciences, University of California at Berkeley, Berkeley, CA 94720.

It should be mentioned that the actual λ value chosen was selected for the purpose of smoothing as opposed to separating components. One would, therefore, expect that if it had any spurious effect (which it does not), the λ method would tend to hide rather than emphasize distributional components. As discussed in one of my earlier works (3), benchmark data confirmation procedures have been used to create a series of test patterns. When the λ smoothing procedures were applied in conjunction with these test patterns, no spurious curve bifurcation was uncovered.

The views and policies presented by the author in this commentary do not necessarily reflect those of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

1. Tarter, M., Cooper, R. C., and Freeman, W. R. Agraphical analysis of the interrelationships among waterborne asbestos, digestive system cancer and population density. Environ. Health Perspect. 53: 79–90 (1983).
2. Tarter, M., and Silvers, A. Implementation and applications of bivariate Gaussian mixture decomposition. J. Am. Statist. Assoc. 70: 47–55 (1975).
3. Tarter, M. Pattern recognition in the context of an asbestos cancer threshold study. In Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface (W. F. Eddy, Ed.), Springer-Verlag, New York 1981, pp. 105–110.