Normality and Continuous Distributions are Myths

Numerous studies have shown that Normal (Gaussian) distributions simply do not occur in empirical research. Despite the widespread evidence against the existence of such distributions, we continue to see statements such as the following in statistics, measurement and research texts: “Data often approximates a normal “bell-shaped” curve...” Micceri (1989), in a study of 440 large-sample empirical distributions from Education and Psychology shows that even here, where some contend that given the developmental process used to produce measures, "...a bell shaped distribution is guaranteed." (Walberg, Strykowski, Rovai and Hung, 1984), only about seven percent are reasonably close to, and none are, in fact, Gaussian. From this and many other sources, we see that Gaussian-like distribution rarely, if ever, occur among empirical distributions. The following works show the non-normality of distributions across a wide variety of disciplines: Pearson, 1895; Allport, 1934; Wilson and Hilferty, 1929; Simon, 1955; Tukey and McLaughlin, 1963; Stigler, 1977; Pearson and Please, 1975; Bradley, 1977, 1982; Tapia and Thompson, 1978; Taillie, Patil & Baldessari, 1981; Tan, 1982.

Although there are several reasons why the Gaussian never occurs, two are particularly relevant for data typically used in social and behavioral science fields. First, empirical measures are always discrete and frequently multinomial, while the Gaussian is continuous. In a theoretical continuous measure, between any two obtained values (say 0.03 and 0.04) an infinite number of possible intermediate values exist. Unfortunately, the lack of precision inherent to all measurement assures that obtained distributions of value are always discrete, and except in the physical sciences, distributions rarely have a sample space exceeding 100 different values.

Second, as anyone who has looked at any social or behavioral science measurement text knows, theoretically, a measure’s values reflect a distribution of “error” about a central true score. Thus a sample of the weights of 50 people represents the distribution of error around the one “true” human weight of the target population. However, this appears an unrealistic assumption for measures involving humans. In any representative sample from a human population there will be at least two subgroups (male and female), each of which varies around a different center on almost any measure, and certainly for measures such as weight. Of course, there are almost always other subpopulations having different values on any measure involving humans. In a population of weight, perhaps the two most extreme subpopulations would be Polynesians (the physically largest people on earth) and Ituri Pygmies (among the smallest people on earth). Almost any real-world distribution, and certainly any that deal with humans, will therefore necessarily be a mixture of various subpopulation distributions.

Three additional reasons are: (1) ceiling or floor effects, (2) in the two or more group test, treatment effects which change not only the location parameter and variability but also the shape of a distribution, and, for Educational measures (3) variability in the difficulty of items within a measure.

Given the preceding, why would any reasonable individual (e.g. statistician), persist in working with continuous or Gaussian distributions despite their improbability? Geary (1947) provides one important reason:

> ...following on the brilliant work of R.A. Fisher who showed that, when universal normality could be assumed, inferences of the widest practical usefulness could be drawn from samples of any size. Prejudice in favour of normality returned in full force ... and the importance of the underlying assumptions was almost forgotten

Fuller (1987) notes another extremely important element: "Solving becomes nontrivial in a non-Gaussian environment... Therefore, this type of suggestion is considered dangerous to theory bound statisticians who prefer to remain enclosed in an Ivory Tower." Kempthorne (1979) explains further: "...'almost all' distributional problems are insoluble with a discrete sample space, notwithstanding the fact that elementary texts are replete with finite space problems that are soluble...I have often taken the view that real data are necessarily discrete and that the consequences of this are inadequately appreciated..." While statisticians can earn tenure writing about mythical distributions, those of us who are forced to work with real-world, empirical data must address the issue raised by Kempthorne (1979): "The real question is whether results from continuous approximation, coupled with the assumption that observations are possible of indefinitely fine precision, are approximately valid with observations one really can obtain."

References

Allport, F. M. (1934). The J-curve hypothesis of conforming behavior. *Journal of Social Psychology*, 5, 141-183.

Bradley, J. W. (1977). A common situation conducive to bizarre distribution shapes. *The American Statistician*, 31, 147-150.

Bradley, J. W. (1982). The insidious L-shaped distribution. *Bulletin of the Psychonomic Society*, 20, 85-88.

Fuller, Wayne A. (1987). *Measurement error models*. John Wiley & Sons: New York.

Geary, R. C. (1947). Testing for normality. Biometrika, 34, 209-242. Hampel, F. R. 1973). Robust estimation: A condensed partial survey. *Zie Wahrscheinlichkeitstheorie Verw. Geb*, 27, 87-104.

Kempthorne, O. (1978). Some aspects of statistics, sampling and randomization (pp. 11-28). In H. A. David (Ed.) *Contributions to survey sampling and applied statistics*. New York: Academic Press.

Pearson, E. S. and Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics. *Biometrika*, 62, 223-241.

Pearson, K. (1895). Contributions to the mathematical theory of evolution - II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society*, A186, 343-414.

Simon, H. A. (1955). On a class of Skew Distribution Functions. _Biometrika_, 42, 425-440.

Stigler, S. M. (1977). Do robust estimators work with real data?. _The Annals of Statistics_, 5, 1055-1098.

Taillie, C., Patil, G. P. and Baldessari, B. A. (1981). _Statistical distributions in scientific work: Volume 5 -Inferential problems and properties_. Boston: Reidel.

Tapia, R. A. and Thompson, J. R. (1978). _Nonparametric probability density estimation_. Baltimore: Johns Hopkins University Press.

Tan, W. Y. (1982). Sampling distributions and robustness of t, F and variance-ratio in two samples and ANOVA models with respect to departure from normality. _Communications in Statistics_, A11, 2485-2511.

Walberg, Herbert J., Strykowski, B.F., Rovai E., and Hung S.S. (1984). Exceptional performance. _Review of Educational Research_, 54, 87-112.

Wilson, E. B., and Hilferty, M. M. (1929). Note on C. S. Peirce's experimental discussion of the law of errors. _Proceedings of the National Academy of Science_, 15, 120-125.