From Bayes Through von Wieser's Marginal Utility to Jacob Cohen's Effect Sizes: A Guide to Understanding the Clinical and Statistical Significance of the Results of Oenologic Research Findings

Dom Cicchetti, Ph.D.
Department of Biometry
Yale University School of Medicine
New Haven, CT 06512

The purpose of this presentation is to trace the history of the problem of distinguishing between statistical and clinical significance of research findings, with specific application to published oenologic research. As such, I shall focus upon the historicity of the phenomenon, that all began with the raging battles between two groups of biostatisticians, the Frequentists, most notably represented by the writings of R.A. Fisher, and the Bayesian biostatisticians, whose work began with a seminal essay by Thomas Bayes (1764) that was published posthumously, and was championed and popularized by the writings of Jerzy Neyman and Egon Pearson (1928; 1933). In the meantime, the Economist Friedrich von Wieser (1851-1926) presented his key economic concept of the now famous "marginal utility." And, finally, Jacob Cohen, following the trend and implication of Bayes' reasoning, developed the concepts of Effect Size (ES) and Power Analysis that guide the design of biobehavioral, and biomedical research to determine minimal sample sizes that will produce results that have both statistical and clinical or practical meaningfulness. The conceptual relationships between the Bayesian, von Wiener, Neyman/Pearson and Cohen approaches to clinical significance will be elucidated.

The most well known Frequentist was Ronald Fisher whose contributions include, among others: the Analysis of Variance, (ANOVA), Discriminant Function Analysis, and the eponymously named Fisher exact Probability Test. The Frequentist position was that the Holy Grail of research findings is the establishment that they have occurred beyond chance, at a minimal probability, or *p* value of 0.05 (e.g., Fisher, 1925).

The Bayesian position and "Ich bin ein Bayesian," as well, is that a research finding must not only occur beyond chance expectancy, but that it must also have clinical or practical meaning, as well. The rationale behind this reasoning is that with a large enough sample size, even the most trivial of results will be statistically significant, be it a difference between group averages, a chi-square value, an F ratio deriving from the ANOVA, a correlation coefficient, a multiple r value, a reliability coefficient, or any other test statistic (Neyman & Pearson, 1928; 1933).

While clinical significance or effect size criteria for statistical test results have been published for both parametric (Borenstein, 1998; Cohen, 1988) and non-parametric statistical tests (e.g., Leach, 1979), I shall focus on a number of statistics as they apply to recently published oenologic research. The latter will include, among others: the relationship between objective and subjective sensory characteristics as determinants of

wine prices (Lecocq and Visser, 2006); and the perceived effects of terroir on the quality of wine under blind and sighted conditions (Priilaid, 2007). In each case, criteria will be applied to enable the differentiation between statistical and clinical significance. The broader implication of this approach for future oenologic research studies will also be discussed.