Padua 2017 Abstract Submission

I want to submit an abstract for:
Conference Presentation

Corresponding Author
Dom Cicchetti

E-Mail
dom.cicchetti@yale.edu

Affiliation
Yale University

Keywords
Reliability; Accuracy; Wine Tasters; Binary Judgments

Research Question
What is the correlation between the reliability of wine tasters' binary judgments and the extent of disagreement on positive and negative cases; and how do statistical and enological significance relate?

Methods
A recently developed method, enothetics, is applied; it allows for a level of control that is usually not possible in the prototypical enological research design.

Results
There were near perfect correlations between the reliability of wine Tasters' judgments and the level of disagreement on positive and negative decisions; statistical and enological significance do not always correspond.

Abstract
The purpose of this presentation is to demonstrate the association between the reliability and validity of wine Tasters' binary judgments, such as: Is a wine filtered or not? oaked or not? is the varietal gamay or pinot noir? Cabernet sauvignon or cabernet franc? (Cicchetti, 2016). Conceptually, the methodological umbrella that will be utilized was developed earlier by the author in the context of resolving binary variable paradoxes whereby acceptably high levels of inter-rater agreement (e.g., > 80%) are associated with low levels of chance-corrected Cohen 1960 Kappa values of 0.40 or less)-Cicchetti (1988); and Cicchetti (1994).
In this enological context, the method allows one to investigate how the reliability and accuracy of wine Tasters' binary judgments vary across a meaningful range of possibilities. Here two hypothetical grape varietals are Grenache (a positive (+) judgment) and Syrah (not Grenache, or a negative (-) judgment. The distinct advantage of the method is that it allows for a level of scientific control that would usually not be possible in the typical enological study. In this hypothetical investigation, the following variables will be controlled: the levels of overall wine Taster agreement at three acceptable levels: 70%; 80%; and 90%; ranges of meaningful patterns of agreement at each of these three levels. For 70% these are: (35++)-(35- -); (40++)-(30- -); (45++)-(25- -); (50++)-(20- -); (55++)-(15- -); (60++)-(10- -); (65++)-(5- -); and (70++)-(0- -). For each case the errors in judgment were set/controlled at 5% Grenache (+)-Syrah(-) and the remaining 5% Syrah (-)-Grenache (+) errors in judgment. For 80%, the patterns of Taster agreement were as follows: (40++)-(40- -); (45++)-(35- -); (50++)-(30- -); (55++)-(25- -); (60++)-(20- -); (65++)-(15- -); (70++)-(10- -); (75++)-(5- -); and (80++)-(0- -). The positive-negative and negative-positive errors were set equally as 10% for each of the two hypothetical conditions. And, finally, for 90% overall Taster agreement, the range of agreement patterns were distributed as the following: (45++)-(45- -); (50++)-(40- -); (55++)-(35- -); (60++)-(30- -); (65++)-(25- -); (70++)-(20- -); (75++)-(15- -); (80++)-(10- -); (85++)-(5- -); and (90++)-(0- -). The 10% error rate also distributed equally between positive-negative and negative-positive cases.
Taster reliability was assessed applying the aforementioned Cohen 1960 statistic of choice, as recommended by the author in an upcoming publication that will appear in the Journal of Nervous & Mental Disease in January (2017). The resulting levels of chance-corrected agreement were evaluated for enological significance according to the criteria of Cicchetti (1994) and earlier by Cicchetti & Sparrow (1981), whereby: <0.40=Poor; 0.40-0.59=Fair;

Accuracy was measured by assessing the statistical significance of the Z values of Kappa whereby, in the standard manner, for a 2 tailed test: <1.96=not statistically significant (NS); 1.96 represents a probability p level of 0.05; 2.25 represents a p of 0.025; 2.58 gives a p of 0.01; 3 represents a p value of 0.003; a Z of 4 reveals a p of <0.0005; and one of 5 is synonymous with a p of <0.0001.

RESULTS:

For 70% overall Taster agreement, the correlation between the reliability of their binary judgments and the size of the discrepancy between their positive and negative decisions was -0.98; in 3 of the 8 patterns of 70% agreement (37.5%), the results were statistically significant but not enologically significant.

For 80% overall Taster agreement, the correlation between the reliability of their binary judgments and the size of the discrepancy between their positive and negative decisions was -0.99; in 1 of 9 patterns of 80% overall agreement (11%), the result was statistically significant but not enologically significant, and finally:

For 90% overall Taster agreement, the correlation between the reliability of binary judgments and the discrepancies between positive and negative decisions was a perfect -1.00; and, for this high level of overall agreement, the 10 patterns of 90% agreement showed a perfect relationship between statistical and enological significance; specifically, the 9 cases that were statistically significant were also enologically significant and the remaining case was neither statistically nor enologically significant.

DISCUSSION:

There are two major findings deriving from this enothetic investigation. First, there is a near perfect inverse correlation between the reliability and accuracy of wine Taster judgments of binary variables such as : whether a wine is filtered or not; is oaked or not; or is one of two varietals, such as pinot noir or gamay; cabernet sauvignon or cabernet franc, etc. The second finding is that enothetic and statistical significance are not always in concert one with the other. While it was consistently true that Taster agreement levels that were enothetically significant were also statistically significant, the reverse was not always true. More specifically whether the overall level of Taster agreement was 70%, 80% or 90%, there were a number of instances in which agreement levels were statistically significant while not reaching levels of enothetic significance. It should be noted here that these results are consonant with a strong correspondence between the size of Cohen's 1960 Kappa for a given binary judgment and the (Sensitivity added to the Specificity)-1 of that same judgment (this discovered by Kraemer, 1982) as based upon the original conceptualization of Youden (1950); and finally, the relationship improved when the Predicted Positive and Predicted Negative diagnostic components were added to the equation, as demonstrated by the Dom Index, due to Cicchetti (2011).

Total Count: 874 Words