



“Evaluation of wine tasters”

By: F. Malík, S. Varga, and S. Slugen

In: *Mitteilungen Klosterneuburg*, 55: 107-113. 2005

The goal of this study was to find out whether there is a reasonable agreement between the final scores given in official wine judgments and the scores given by the individual tasters participating in that judging.

- The authors used as an example the results of the international wine competition Vinoforum 2004, sponsored by the O.I.V (International Wine Organization). This particular judging consisted of 7 panels - comprised of 5 judges each- tasting approximately 100 wines in 2 days. For each wine, the judges evaluated the following 10 quality properties using a scale between 0 and 100: limpidity, color, aroma intensity, aroma genuineness, aroma quality, taste intensity, taste genuineness, taste quality, persistence, and overall appreciation (these descriptors might seem bizarre, or unusually subjective, from a sensory evaluation point of view, but I’m afraid this is “real life” at wine contests around the world). The organizers then eliminated the highest and the lowest scores from each panel, and gave the wine an overall score equal to the arithmetic average of the remaining three values (this is called a “cut average”).

- The authors investigated the agreement between the “cut average” wine scores (X variable) and the individual judge scores (Y variable) using 4 statistical methods:

- 1) *Test of dependence*: through this test, the authors can see whether there exists a significant difference between X and Y for each wine.
- 2) *Test of contingency*: this measures the strength of the agreement (dependence) between X and Y. As the dependence value increases, so does the value of the professional experts’ opinions.
- 3) *Test of symmetry*: this measures whether the individual scores are smaller or larger than the wine score. That is, whether a particular judge underestimates, overestimates, or is symmetrical with the overall opinion, X.
- 4) *X-Y chart*: finally, by using a specific chart to visualize both the official evaluations and the individual ones, the authors are able to monitor the performance of each wine taster. (Don’t worry if, like me, you don’t understand these tests; you don’t need to do so to understand the final results.)

- **Results.** 1) In general, all 35 tasters in the seven panels evaluated the assigned wines in good accordance with the official cut average of the wines. 2) However, only 16 tasters evaluated the wines symmetrically, that is, 19 tasters underestimated or overestimated the average scores given the wines. 3) Finally, using the above tests, the authors were able to rank judges for their performance (in the original paper the authors present score charts for the two best and the two worst performers).

The authors present a method that, by monitoring taster performance, can help official wine contests by improving the selection of objective professional tasters. The authors admit that these tests should be applied in an anonymous manner if the reputation of many a judge is to be preserved!

Author: Bibiana Guerra, Editor: Kay Bogart. This summary series funded by J. Lohr Vineyards & Wines.