Journal of Wine Economics, Volume 3, Issue 2, Fall 2008, Pages 226–227

Letters to the Editors

On Rating Wines with Unequal Judges

In the recent article by Goldstein, et al.(2008), covering over 6000 blind wine tastings, raters were asked to score wines on a scale of "Bad", "O.K", "Good", or "Great.". The ratings were subsequently coded on a numerical scale of 1 to 4 and averaged to rate the wines. Unknown to the raters, some wines were duplicates, which were used to evaluate the raters. Thus for a flight of 10 wines with 6 raters, if one of the raters was found to be quite inconsistent on the replicated wine, that rater's scores on *all* wines would have less weight. In other words, each wine's final score would not be an equally weighted average of all six raters. The exact weighting scheme was not discussed, but it brought to my attention a problem I used to present first year statistics students.

Suppose you are measuring a physical quantity, like chlorine concentration in water. You are presented with three measurements: two from "chlorine meters" that have a precision of 10ppm, and one with a precision of 50ppm. Do you average the two measurements taken with the more precise meter and discard the third? Do you average all three? If you toss out the third measurement, you are discarding information, even though the information is not very precise.

It is well known that the mean of n measurements with an instrument having a standard deviation of σ will have a standard error of $\sigma/\sqrt{n}$. If one just averages the first two measurements described above, the standard error is $10/\sqrt{2}$. Is it possible to weight all three measurements in such a way to improve the precision beyond this?

The answer is yes, and the proof follows from theorems regarding a linear combination of random variables. Let σ_p be the standard deviation of the poorer instrument and σ_i be that of the better instrument. Then we can relate the two standard deviations by $\sigma_p = k\sigma_i$, where $k = 5$ in the example. Applying differential calculus to search for a minimum standard error yields the weighting factors. The best (least standard) error occurs when the poorer measurement has a weight of $1/k^2$. The two "good" measurements have equal weighting factors of $\frac{1}{2}(1-1/k^2)$.

In the above example where $k=5$, the weighting factor of the good measurements is slightly less than 0.5 and that of the poorer measurement is 1/25. For practical purposes, you might as well throw the third measurement out.

How would this apply to the Goldstein article? Since the wines were rated 1 to 4, the maximum inconsistency would be a 3 point spread. Estimating a rater's standard deviation by

range[1], a poor rater might have his score weighted by 1/9. If raters were to use the more common 20 point scale, weighting factors could be more extreme.

Robert T. Hodgson
Professor Emeritus, Humboldt State University
email: bob@fieldbrookwinery.com

References

Goldstein, Robin, Johan Almenberg, Anna Dreber, John W. Emerson, Alexis Herschkowitsch and Jacob Katz. (1980). Do more expensive wines taste better?. Journal of Wine Economics, 3(1), 1–9

2008 Wine Valuation Analysis

Here is recent convergence of Science and economics. Physicists from the Centre National de la Recherche Scientifique determine a dating method for the glass of a wine bottle.(1) An investigator at the meeting of American Association of Wine Economists presents evidence that famous old empty wine bottles are rapidly increasing in market value.(2) While it cannot be determined whether this temporal concordance is causal, it should be noted that testing of the bottle and invasive testing of wine are already being offered by Antique Wine Company.(3)

James J. O'Donnell, M.D.
Professor Emeritus, UC San Francisco
email: JOdonn5952@aol.com

References

(1) http://www2.cnrs.fr/en/1283.htm Paris, September 1, 2008 Vintage wine bottles authenticated by high energy ion beam. Just like works of art, wine is now being subjected to advanced testing to establish its authenticity: after measuring caesium 137 radioactivity levels to test the age of the wine, the glass in vintage wine bottles is now being tested by particle acceleration. Referred by Barbara O'Donnell.
(2) http://www.wineeconomics.org/meetings/Portland2008/details.htm Portland, Oregon, August 14–16, 2008, Second Annual Conference of American Association of Wine Economists (AAWE), as described by Professor Orley Ashenfelter, President, AAWE.
(3) http://www.antiquewine.com/press_room_article.php?press_room_id=38

[1] Maximum score less minimum score