Ithaca 2018 Abstract Submission

Title
Wine, Women and Men: Large Sample Results

I want to submit an abstract for:
Conference Presentation

Corresponding Author
Jeffrey Bodington

E-Mail
jcb@bodingtonandcompany.com

Affiliation
Bodington & Company

Keywords
Wine tasting, women, men, bias, awards, statistics

Research Question
Does the gender balance of wine judges at a wine competition result in biased awards?

Methods
Statistical analysis of a scores assigned to 260 wines by 21 female and 24 male judges.

Results
No significant difference in the distributions of scores. No bias in awards.

Abstract
Bodington (2017) compared the scores assigned by women and men to the same wines in 23 tastings. Using four different tests, he concluded that indications that the genders prefer different wines are difficult to detect. An issue with that finding, and most other evaluations of wine tasting results, is that sample sizes are small. To address that issue, this abstract presents results for the 2016 Wines of Portugal Challenge in which panels that included at least three female judges and at least three male judges assigned scores to 260 wines. The resulting sample contains 1,736 wine-score observations, 833 assigned by female judges and 903 assigned by male judges.

Corbin (2006) reviewed over 40 published evaluations of taste tests that differentiated between female and male tasters. She found gender-related differences in taste preferences for sweet, salt, sour and bitter flavors. Corbin also found gender-related differences in preferences for fruit and fruit juice flavors. While her results did not include any comparisons of women's and men's preferences for wine, others have reported wine-related differences concerning all of sight, taste and smell. See Bodington (2017) for a review of that literature. Despite the differences between women and men enumerated above, the awards made in dozens of international, national, state fair, county fair, magazine and other wine competitions are determined by pooling the scores assigned by female and male judges. If women and men do have different wine taste preferences, then a different gender balance may lead to different awards. The gender balance of judges alone may bias results. This abstract presents a large-sample test of the potential for that bias.

The Wines of Portugal Challenge is an annual competition among wines produced by over 1,000 vintners in over 30 of the country's wine growing regions. In 2016, each of 1,328 wines were assessed and scored by panels of five to seven judges. The wines were tasted blind and each judge assigned a score between zero and 100 to each wine

that he or she assessed. See further explanation of the Portugal Challenge in Bodington & Malfeito-Ferreira (2017). Of 151 judges, 40 were women and 111 were men. Panels that included at least three women and three men assessed and scored a total of 260 wines. Those panels were composed of 21 female who assigned 833 scores and 24 male judges who assigned 903 scores. That sample of 1,736 wine-score observations is the subject of this article.

The distribution of the difference between the mean of female judges' scores and the mean of male judges' scores on the same wines appears in Figure 1. The distribution of that difference has a mean of -0.8 score points, a standard deviation (SD) of 4.2 points, and the shape is nearly symmetric with a skewness of -0.2. Although that –0.8 difference in means is small at less than one point, due to the large sample, it is marginally significant. A two-sample t-test of the null hypothesis that the difference is zero has a p-value of 0.09.

Figure 1
Difference in the Means of Scores Assigned by Female and Male Judges to the Same Wines
(260 wines scored by at least three female and three male judges)

Is the distribution in Figure 1 a mixture of component distributions that may reveal differences between female and male judges' scores? The Portugal Challenge categorized wines as white, red, rosé, white sparkling, red sparkling, rosé sparkling, sweet and late harvest. Statistics that describe the distributions of differences in gender means for each of those categories appear in Table 1. Still subject to the requirement that at least three men and three women have assigned scores to a wine, there are too few observations except for white and red to be useful. The results for white and red indicate that men assigned higher scores than did women and that men also assigned even higher scores to red than white wine. However, both of those differences are small and two sample t-tests show that they are not statistically significant.

Table 1
Su…
Category Number
of wines Mean SD Skew

White 95 -0.4 4.7 -0.3
Red 149 -1.0 4.0 -0.1
Rosé 4 -2.9 2.9 -1.1
White sparkling 0 - - -
Red sparkling 2 0.7 1.3 0.0
Rosé sparkling 10 -1.7 3.0 -0.6
Sweet 0 - - -
Late harvest 0 - - -

Total 260 -0.8 4.2 -0.2

What is the cause of the variation in Figure 1 and the SDs in Table 1? Bodington (2017), Corbin (2006, p. 2) and Jackson (2009, p. 87) conclude that non-gender-related differences between individuals are the primary reason for the variance in scores. That view is tested here by calculating the distribution of results for 1,000 cases in which men's and women's scores were shuffled randomly amongst each other. The number of wines, the wines, the numbers of male and female judges, the numbers of scores assigned by women, the number of scores assigned by men and the actual wine-score observations were not changed. Only the gender-assignment of each wine-score observation was random. The resulting distribution appears in Figure 2 below. The observed results also appeared in Figure 1, and the random distribution of results appears as the heavy sold line in Figure 2.

Figure 2

Observed and Expectation of Randomly-Assigned Gender
Difference in the Means of Scores Assigned by Female and Male Judges to the Same Wines
(260 wines scored by at least three female and three male judges)

If actual-gender results are no different than random-assignment-gender results then gender doesn't explain any difference; variation must be due to non-gender-related idiosyncratic and random factors. That is the strong implication of Figure 2. Visually, there is little difference between the observed results and the random results. The correlation between the two distributions is greater than 0.9. A chi square test of the null hypothesis that the two distributions are the same has a p-value that is also greater than 0.9. These large-sample findings are consistent with the small-sample findings in Bodington (2017). If accept-a-false-null-hypothesis Type II error is made by combining women's and men's scores and ranks when judging wine competitions, this article has shown that such error is small.

Although the potential for Type II error in wine tasting results is small, it cannot be dismissed. These results imply that the most cautious approach, particularly for wines with strong or unusual flavors, is to test tasting results for differences in gender-specific preferences. If the p-values are low, report gender-specific results and the possibility that women and men have different preferences.

References

Bodington, J. (2017). Wine, Women, Men and Type II Error. Journal of Wine Economics, 12(2), 161-172.
Bodington, J. and Malfeito-Ferreira, M. (2017). The 2016 Wines of Portugal Challenge: General implications of more than 8400 wine-score observations. Journal of Wine Research. Published online 23 Oct 2017.
Corbin, C. (2006). Sex Differences in Taste Preferences in Humans. Literature review for Psychology 451 under D. Pittman, Wofford College, Spartanburg, SC.

Jackson, R.S., (2009). Wine Tasting: A Professional Handbook, Second Edition. Elsevier, Academic Press, Amsterdam, 519 pgs.

File Upload (PDF only)

- Bodington-Wine-Women-Men-Large-Sample-Results-1-25-17.pdf

Abstract

Wine, Women and Men: Large Sample Results

Jeff Bodington

JEL Classifications: A10, C10, C00, C12, D12

Keywords: Wine tasting, women, men, bias, awards, statistics

Bodington (2017) compared the scores assigned by women and men to the same wines in 23 tastings. Using four different tests, he concluded that indications that the genders prefer different wines are difficult to detect. An issue with that finding, and most other evaluations of wine tasting results, is that sample sizes are small. To address that issue, this abstract presents results for the 2016 Wines of Portugal Challenge in which panels that included at least three female judges and at least three male judges assigned scores to 260 wines. The resulting sample contains 1,736 wine-score observations, 833 assigned by female judges and 903 assigned by male judges.

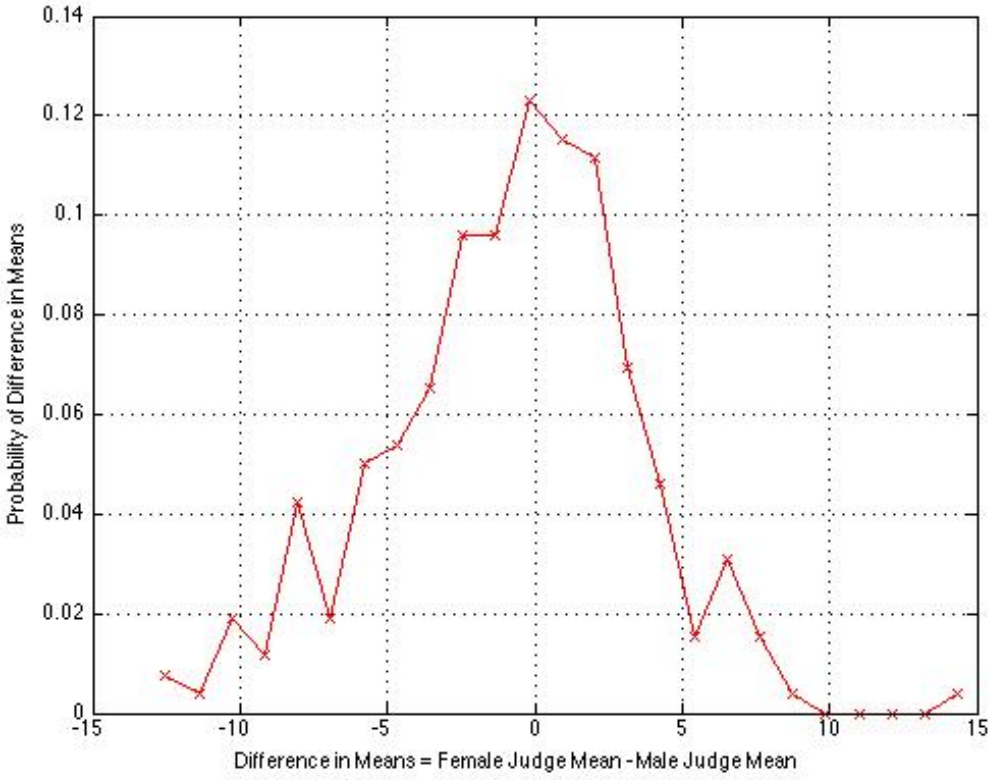
Corbin (2006) reviewed over 40 published evaluations of taste tests that differentiated between female and male tasters. She found gender-related differences in taste preferences for sweet, salt, sour and bitter flavors. Corbin also found gender-related differences in preferences for fruit and fruit juice flavors. While her results did not include any comparisons of women's and men's preferences for wine, others have reported wine-related differences concerning all of sight, taste and smell. See Bodington (2017) for a review of that literature. Despite the differences between women and men enumerated above, the awards made in dozens of international, national, state fair, county fair, magazine and other wine competitions are determined by pooling the scores assigned by female and male judges. If women and men do have different wine taste preferences, then a different gender balance may lead to different awards. The gender balance of judges alone may bias results. This abstract presents a large-sample test of the potential for that bias.

The Wines of Portugal Challenge is an annual competition among wines produced by over 1,000 vintners in over 30 of the country's wine growing regions. In 2016, each of 1,328 wines were assessed and scored by panels of five to seven judges. The wines were tasted blind and each judge assigned a score between zero and 100 to each wine that he or she assessed. See further explanation of the Portugal Challenge in Bodington & Malfeito-Ferreira (2017). Of 151 judges, 40 were women and 111 were men. Panels that included at least three women and three men assessed and scored a total of 260 wines. Those panels were composed of 21 female who assigned 833 scores and 24 male judges who assigned 903 scores. That sample of 1,736 wine-score observations is the subject of this article.

The distribution of the difference between the mean of female judges' scores and the mean of male judges' scores on the same wines appears in Figure 1. The distribution of that difference has a mean of -0.8 score points, a standard deviation (SD) of 4.2 points, and the shape is nearly symmetric with a skewness of -0.2. Although that –0.8 difference in means is small at less than one point, due to the large sample, it is marginally significant. A two-sample *t-test* of the null hypothesis that the difference is zero has a *p-value* of 0.09.

Figure 1

Difference in the Means of Scores Assigned by Female and Male Judges to the Same Wines
(260 wines scored by at least three female and three male judges)



Is the distribution in Figure 1 a mixture of component distributions that may reveal differences between female and male judges' scores? The Portugal Challenge categorized wines as white, red, rosé, white sparkling, red sparkling, rosé sparkling, sweet and late harvest. Statistics that describe the distributions of differences in gender means for each of those categories appear in Table 1. Still subject to the requirement that at least three men and three women have assigned scores to a wine, there are too few observations except for white and red to be useful. The results for white and red indicate that men assigned higher scores than did women and that men also assigned even higher scores to red than white wine. However, both of those differences are small and two sample *t-tests* show that they are not statistically significant.

Table 1

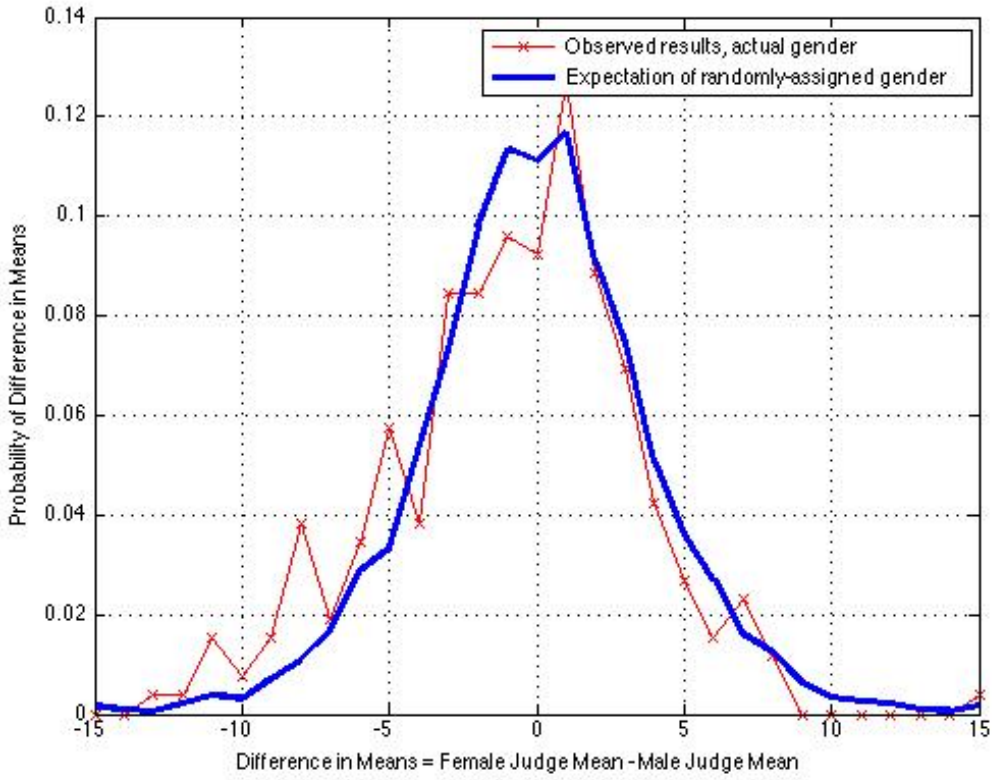
Su…

Category	Number of wines	Mean	SD	Skew
White	95	-0.4	4.7	-0.3
Red	149	-1.0	4.0	-0.1
Rosé	4	-2.9	2.9	-1.1
White sparkling	0	-	-	-
Red sparkling	2	0.7	1.3	0.0
Rosé sparkling	10	-1.7	3.0	-0.6
Sweet	0	-	-	-
Late harvest	0	-	-	-
Total	260	-0.8	4.2	-0.2

What is the cause of the variation in Figure 1 and the SDs in Table 1? Bodington (2017), Corbin (2006, p. 2) and Jackson (2009, p. 87) conclude that non-gender-related differences between individuals are the primary reason for the variance in scores. That view is tested here by calculating the distribution of results for 1,000 cases in which men's and women's scores were shuffled randomly amongst each other. The number of wines, the wines, the numbers of male and female judges, the numbers of scores assigned by women, the number of scores assigned by men and the actual wine-score observations were not changed. Only the gender-assignment of each wine-score observation was random. The resulting distribution appears in Figure 2 below. The observed results also appeared in Figure 1, and the random distribution of results appears as the heavy sold line in Figure 2.

Figure 2

Observed and Expectation of Randomly-Assigned Gender
Difference in the Means of Scores Assigned by Female and Male Judges to the Same Wines
(260 wines scored by at least three female and three male judges)



If actual-gender results are no different than random-assignment-gender results then gender doesn't explain any difference; variation must be due to non-gender-related idiosyncratic and random factors. That is the strong implication of Figure 2. Visually, there is little difference between the observed results and the random results. The correlation between the two distributions is greater than 0.9. A chi square test of the null hypothesis that the two distributions are the same has a *p-value* that is also greater than 0.9. These large-sample findings are consistent with the small-sample findings in Bodington (2017). If accept-a-false-null-hypothesis Type II error is made by combining women's and men's scores and ranks when judging wine competitions, this article has shown that such error is small.

Although the potential for Type II error in wine tasting results is small, it cannot be dismissed. These results imply that the most cautious approach, particularly for wines with strong or unusual flavors, is to test tasting results for differences in gender-specific preferences. If the *p*-values are low, report gender-specific results and the possibility that women and men have different preferences.

References

Bodington, J. (2017). Wine, Women, Men and Type II Error. *Journal of Wine Economics*, 12(2), 161-172.

Bodington, J. and Malfeito-Ferreira, M. (2017). The 2016 Wines of Portugal Challenge: General implications of more than 8400 wine-score observations. *Journal of Wine Research*. Published online 23 Oct 2017.

Corbin, C. (2006). Sex Differences in Taste Preferences in Humans. Literature review for Psychology 451 under D. Pittman, Wofford College, Spartanburg, SC.

Jackson, R.S., (2009). Wine Tasting: A Professional Handbook, Second Edition. Elsevier, Academic Press, Amsterdam, 519 pgs.