HILGARDIA

A Journal of Agricultural Science Published by the California Agricultural Experiment Station

VOLUME 30 MAY, 1961 NUMBER 19

SMALL PANEL SENSORY EVALUATIONS OF WINES BY SCORING

C. S. OUGH and G. A. BAKER

UNIVERSITY OF CALIFORNIA · BERKELEY, CALIFORNIA

Several conclusions can be made from the data on the extensive wine tastings reported here.

The over-all distributions of wine scores are normal with some exceptions. The score card in present use gives more lower scores for some tasters than for others. Some tasters show distinct nonconformity when compared with most of the tasters. Training periods necessary for consistent results are not long.

Wines should be compared within region-type groups.

Good tasters pinpoint differences better than a mixture of good and not-so-good tasters. The present methods of wine scoring are an effective means of wine-quality evaluation with respect to comparison of variety of parent grape within region-types.

H I L G A R D I A

A Journal of Agricultural Science Published by the California Agricultural Experiment Station

VOL. 30 | MAY, 1961 | No. 19

SMALL PANEL SENSORY EVALUATIONS OF WINES BY SCORING[1]

C. S. OUGH[2] and G. A. BAKER[3]

INTRODUCTION

THE DETECTION of the differences between wines by scoring paired, triangular, and duo-trio sensory techniques has been explained and illustrated by Amerine, Roessler, and Filipello (1959).[4] The evaluation of wines by scoring gives much more information to the investigator but the analyses are more complicated and are based on a mathematical probabilistic model that may require some justification in a given case. In particular it is usually assumed that variances of the tasters are equal. Also, it makes a great difference whether the analysis is considered from the standpoint of the actual tasters involved or whether the tasters are regarded as a random sample from some hypothetical population of tasters. These points are excellently explained by Scheffé (1959).

The purpose of this paper is to make a comprehensive report on the results of a small panel of not more than eight tasters on the following number of wines by years, region, and type.

Region I—Oakville (Napa Valley)	**1958**	**1959**
Red Dry (RD)	64	33
White Dry (WD)	109	37

Region IV—Davis (Sacramento Valley)	**1958**	**1959**
Red Dry (RD)	36	37
Red Sweet (RS)	30	14
White Dry (WD)	14	45
White Sweet (WS)....	29	14

The tastings were carried out for two years. Differences in the wines within a region and type are due to variety distinctions and time of picking. There were two tastings—one before and one after clarification.

The study is to continue over several years to make an assessment of the effects of season, time of picking, and varietal differences.

The considerations that were of prime importance include the over-all distributions of each taster's scores; defects in the scoring system; variances

[1] Submitted for publication October 6, 1960.

[2] Associate Specialist, Department of Viticulture and Enology, Davis.

[3] Professor of Mathematics and Statistician in the Experiment Station, Davis.

[4] See "Literature Cited" for citations referred to in the text by author and date.

of the tasters; ratings of the tasters; effects of regions, wines, tasters, years, time-of-tasting; incompatability of tasters; and the improvement of the scoring system.

METHODS AND MATERIALS

Wines. The wines were produced in 1958 and 1959 at the University winery from grapes grown in the Oakville and Davis vineyards of the University of California. Wines were made by standard fermentation techniques. Grapes for white dry wines and white sweet wines were machine crushed, stemmed, and pressed immediately. For red dry and red sweet wines the grapes were machine crushed, stemmed, and allowed to ferment on the skins for an average of three days for the RD (red dry wines) and until fortification, (about two days) for the RS (red dessert) wines. Sweet wines were fortified to 19.5 per cent alcohol and had a resulting reducing sugar of about 10 to 12 per cent.

Regions. California is divided into five grape-growing regions (I–V, inclusive) (Amerine and Winkler, 1944), the division based on the summation of degree-days above 50°F for the period April to October. Region I is the coolest region and V the warmest. The Oakville vineyard (in Napa Valley) is in region I and the Davis vineyard is in region IV. These experiments involved these two regions only.

Tasters. Four members (I, II, III, VI) of the panel have been tasting wines for 15 or more years and are considered expert tasters. Two members (IV, V) of the panel have been associated with wine for 10 years but had done little critical tasting. One member (VII) had no experience or association with wines or with other critical taste testing. Another member (VIII) had several years' experience with critical taste testing of other products.

Tastings. The first tasting of the wines was done in November–December, approximately 50 days after the wines had finished fermenting. Only normal clarification (settling-out of the yeast) had taken place at this time. The second tasting was made in March–April after the wines had been clarified by several rackings and a filtration. Normal cellar care of the wines was maintained during the four months between the tastings. This care involved aeration, addition of sulfur dioxide, and fining with bentonite (rarely) if it was necessary. The tasting was conducted in individual tasting booths. The wines were presented cold in random order in standard 9-ounce wine glasses. A different order of presentation was assigned to each taster. If it was practical, only dry wines or sweet wines were given in one day's tasting. If both were given, the assignment order always presented the dry wine before the sweet wine. If red and white wines were tasted on the same day, the white wines were presented first. The reasons for these procedures are discussed by Amerine *et al.* (1959). Tasters were presented with 6 or 12 samples each day. The following day one of the previous day's samples was repeated (daily duplicate). At the conclusion of the first tasting three series of wines were retasted: wines tasted the first two days, wines tasted one day in the intermediate portion of the tasting, and wines from the last day's tasting. This same procedure was followed in the second tasting. These repeated tastings are called "end repeats."

SCORE CARD

Name or No.					
Appearance.....2					
Color.....2					
Aroma and Bouquet.....4					
Acescent.....2					
Total Acid.....2					
Sugar.....1					
Body.....1					
Flavor.....2					
Astringency.....2					
General Quality.....2					
Total.....					

Fig. 1. Score card now in use.

17 to 20, wines must have some outstanding characteristic and no marked defect; 13 to 16, standard wines with neither an outstanding character or defect; 9 to 12, wines of commercial acceptability but with a noticeable defect; 5 to 8, wines of below commercial acceptability; 1 to 4, completely spoiled wines.

Scoring. The tasters were given a score sheet (fig. 1) commonly used in this laboratory and were asked to score the wines on a 1 to 20 point basis (1 to 4, spoiled wine; 5 to 8, wines below commercial acceptance; 9 to 12, wines commercially acceptable but with a noticeable defect; 13 to 16, standard wines; and 17 to 20, wines with no marked defects and some superior characteristics).

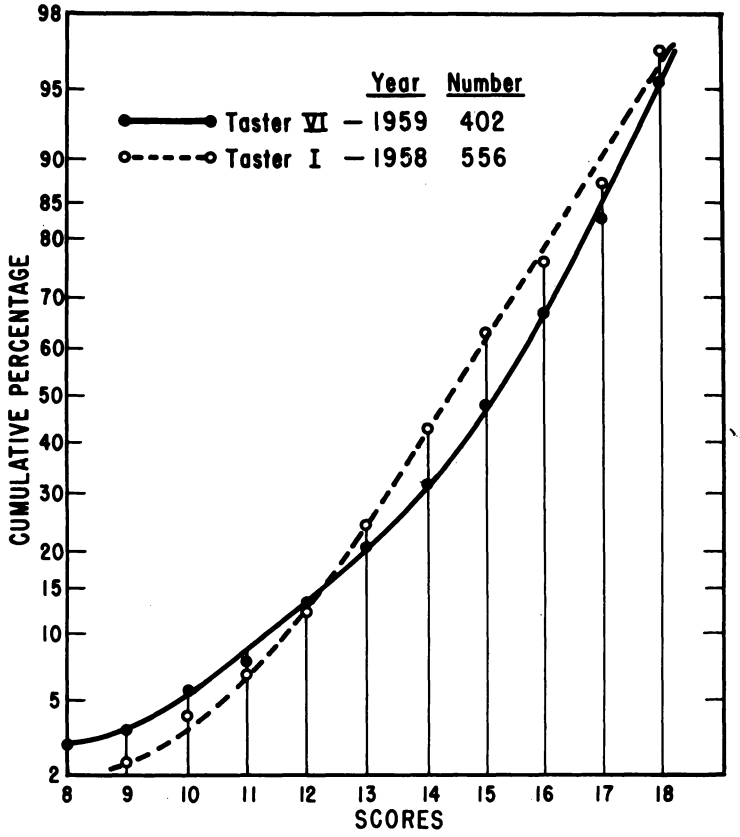

Fig. 2. Plot of score distributions for tasters I in 1958 and VI in 1959.

Analysis of Variance. Standard techniques for the fixed effects model (Model I) were used as presented by Scheffé (1959) for the analysis of variance. Testing for homogeneity of variances was done by the method of Bartlett (1945) to determine if the tasters' error terms (from daily duplicate tastings) were significantly different. To determine significant differences in mean scores the method of Duncan (1955) was applied.

Analysis of Distribution. The wines made are expected to have distribution as far as quality is concerned. The wines came from a large number of grape varieties from two regions and were picked at several different times through-

out the whole season. Production techniques are extremely uniform and should not have any effect on the distributions. The intuitively reasonable premise is that the wines are normally distributed. The distributions for tasters, region-types, and total scores were plotted on normal-probability paper (the cumulative percentage of the scores against the scores). For a normal distribution a straight line should result. The plots for most of the

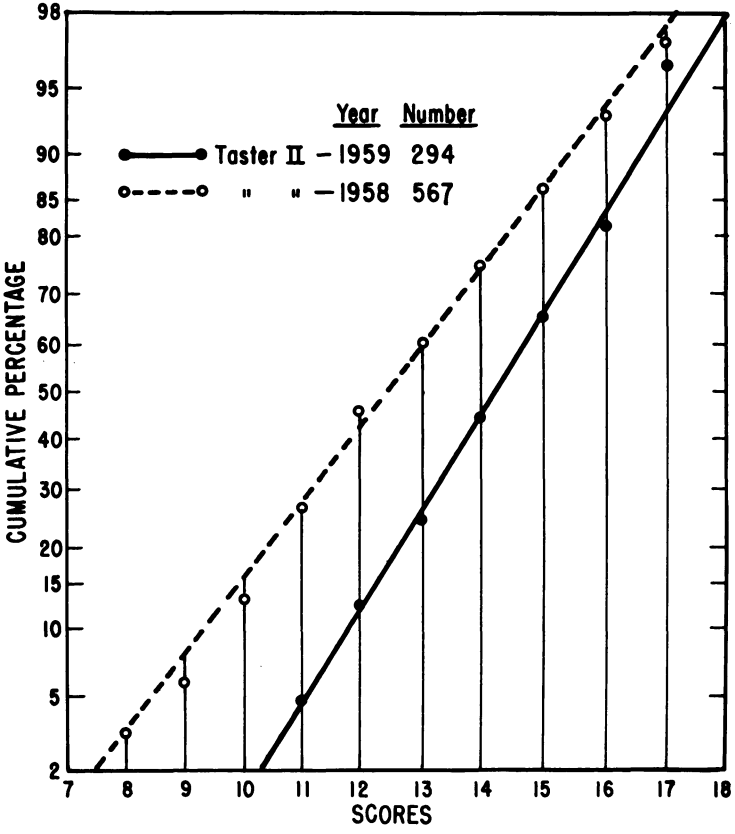

Fig. 3. Score distribution for taster II in 1958 and 1959.

tasters and situations considered here were approximately straight. In other experiments where fewer wines are selected more rigorously, the distributions of scores may be far from normal.

Evaluation of Tasters. The error variance and normality of scoring are important but do not entirely standardize the value of a taster's results. The range the taster uses must be considered. It was decided that the range for 96 per cent of the scores divided by the standard deviation ($\sqrt{\text{error variance}}$) of the taster is the best characterization of a taster. This ratio could be called the taster's "effective range" and represents the range that the taster could use and have a standard deviation of 1 unit.

RESULTS AND DISCUSSION

Distribution. To determine the distributions, the cumulative scores for each taster (figures 2, 3, 4, 5, 6), for the region-types (figures 7, 8, 9), and for the total scores (fig. 10) for each year were plotted on normal-probability paper.

Tasters II (fig. 3) and III (fig. 4) show exceptionally good normal distribution curves, tasters V and VII (fig. 6) show fairly normal distribution although both tasters show a marked reluctance to use certain scores. Taster

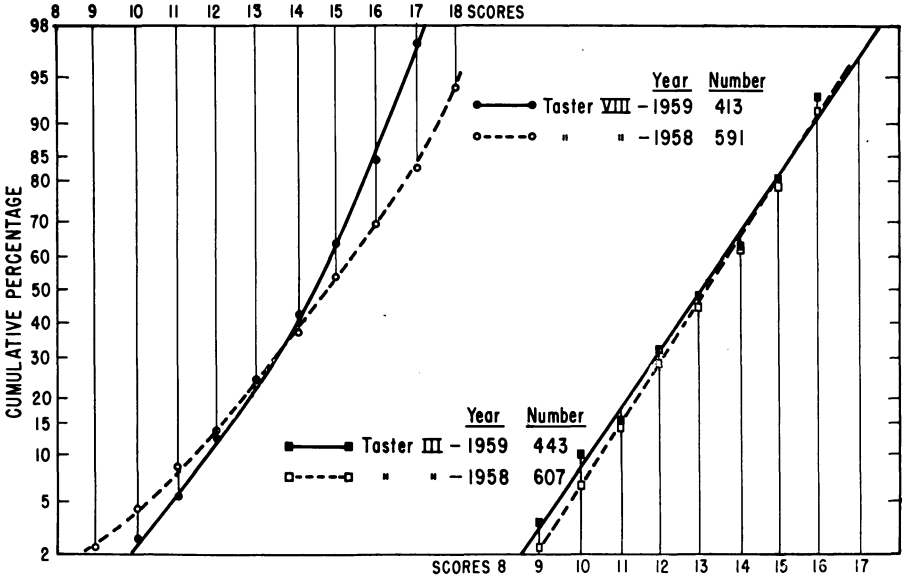

Fig. 4. Score distribution for tasters III and VIII in 1958 and 1959.

VII probably shows the best improvement in distribution from first year to the next while taster VIII (fig. 4) shows a decided progression away from a normal distribution in the second year, as does taster V to a lesser extent. Taster I (fig. 2) has a normal distribution over about 80 per cent of the range. Taster VI (fig. 2) shows a very marked deviation from a normal distribution. Taster IV (fig. 5) shows a very large range with nearly normal distribution. The general tendency for most tasters is to assign more low scores than expected.

The distributions for the region-type scores (figures 7, 8, 9) show near normal distributions in general—the worst being RS IV (fig. 9). In this case the lower 20 per cent of the assigned scores were given more often than would be expected for a normal distribution. This tendency of giving too low scores to poorer wines is also seen in the other region-type score-distributions but not to so great an extent. The most noticeable change in the shape of the cumulative distributions curve is noted on the WD IV between the 1958 to 1959 curves. The range used by the tasters was shortened from 13 points to 9 points.

The total distribution of scores (fig. 10) is fairly normal. The tendency to assign too many low scores is again observed here. This could probably be averted for these particular data by the use of a different type scoring system. A large portion of the scores available to be used are in general not used frequently, especially the scores between 1 and 10. The wines which rate in this category occur less than 10 per cent of the time. Consequently

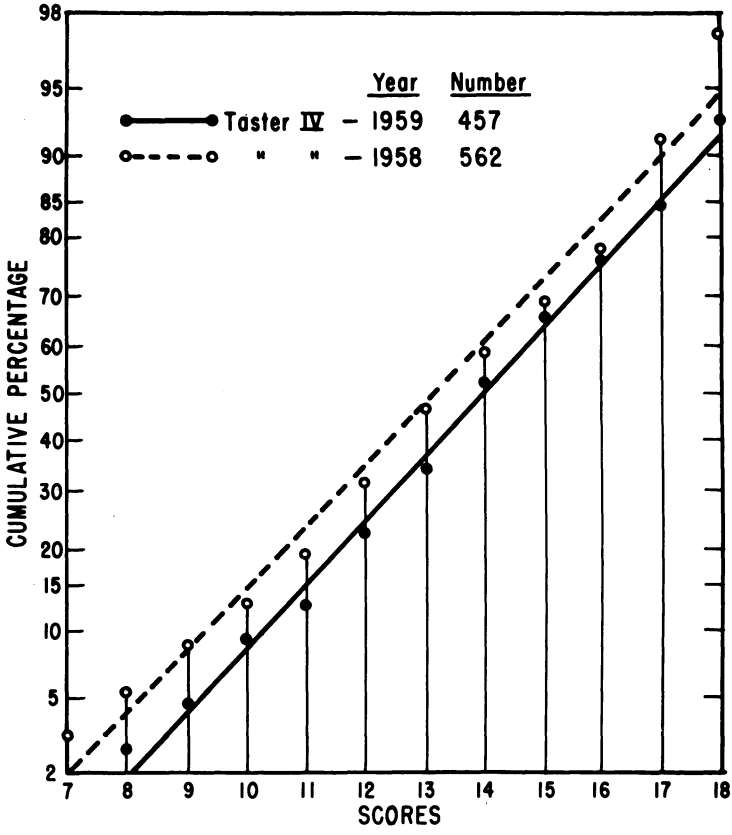

Fig. 5. Score distribution for taster IV in 1958 and 1959.

90 per cent of the scores assigned are between 10 and 20. The infrequent real need to use this lower portion of the score range, and the tendency of the tasters to mark the scores lower than is probably justified in an effort to use the whole range of 1 to 20 are the apparent reasons for these deviations from normal distribution in the lower 10 or 20 per cent of the scores. This could be rectified by the use of a score sheet with the distribution of scores more symmetrical about the mean score. It may be true that as the tasters become more experienced the tendency to downgrade the below-average wines lessens as is demonstrated by the two most experienced tasters, II and III. However, this latter conjecture is made with some reservations as tasters I and VI have had experience in tasting for more than 15 years and

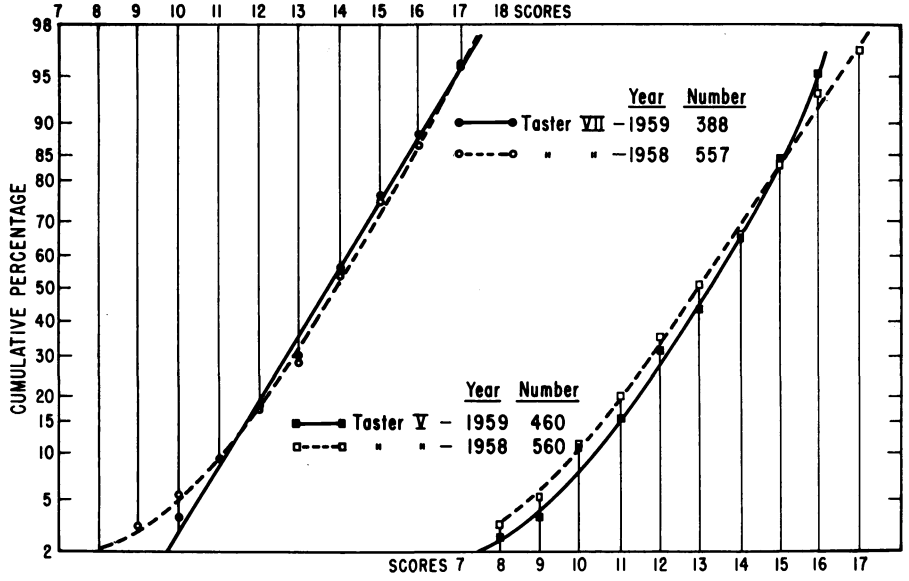

Fig. 6. Score distribution for tasters V and VII in 1958 and 1959.

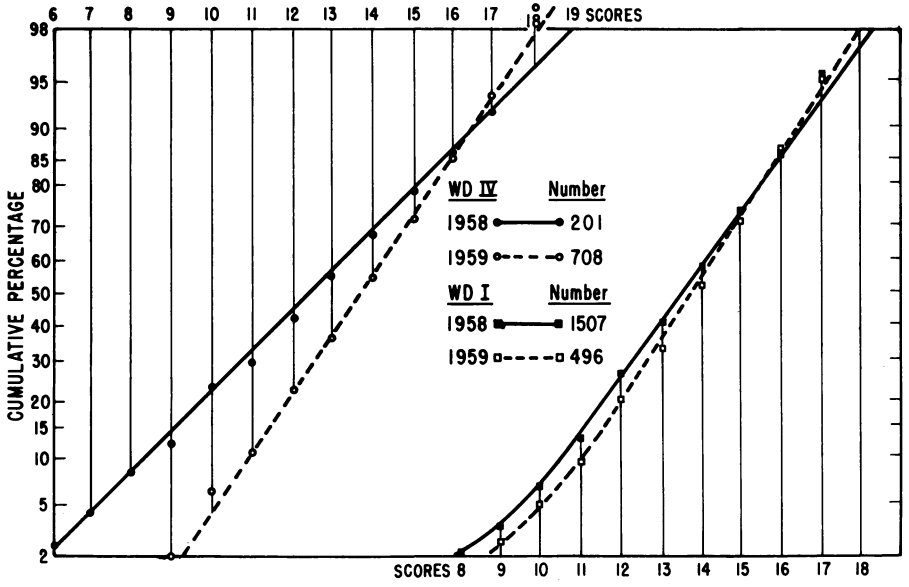

Fig. 7. Score distribution for WD IV and WD I for 1958 and 1959.

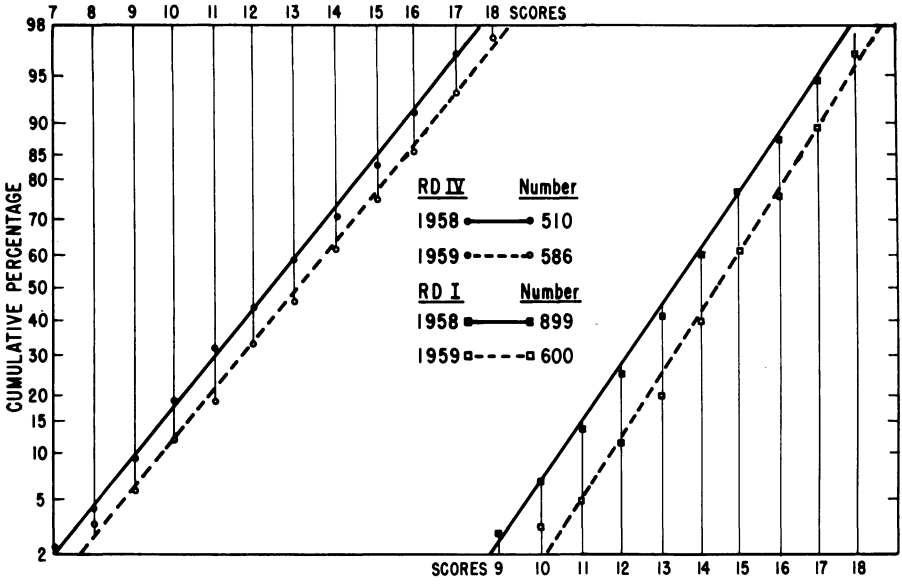

Fig. 8. Score distribution for RD IV and RD I in 1958 and 1959.

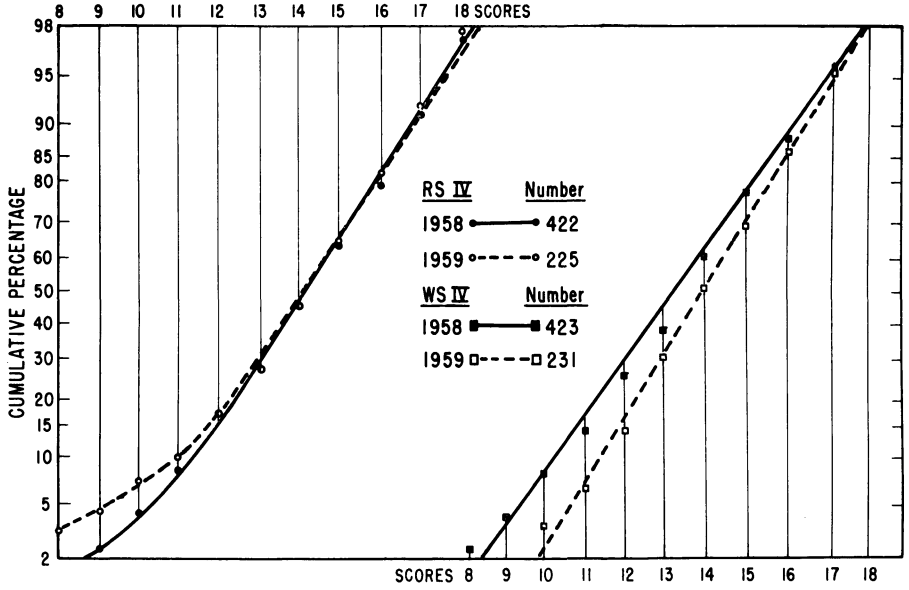

Fig. 9. Score distribution for RS IV and WS IV in 1958 and 1959.

their distributions are far from normal. In other test situations different from this one we might be considering the similarity of non-normal score-distributions.

The score card in use is really a five point system in respect to the classifications—spoiled, below commercial acceptance, acceptable with noticeable defects, standard wines, wines of some superior characteristic—with four

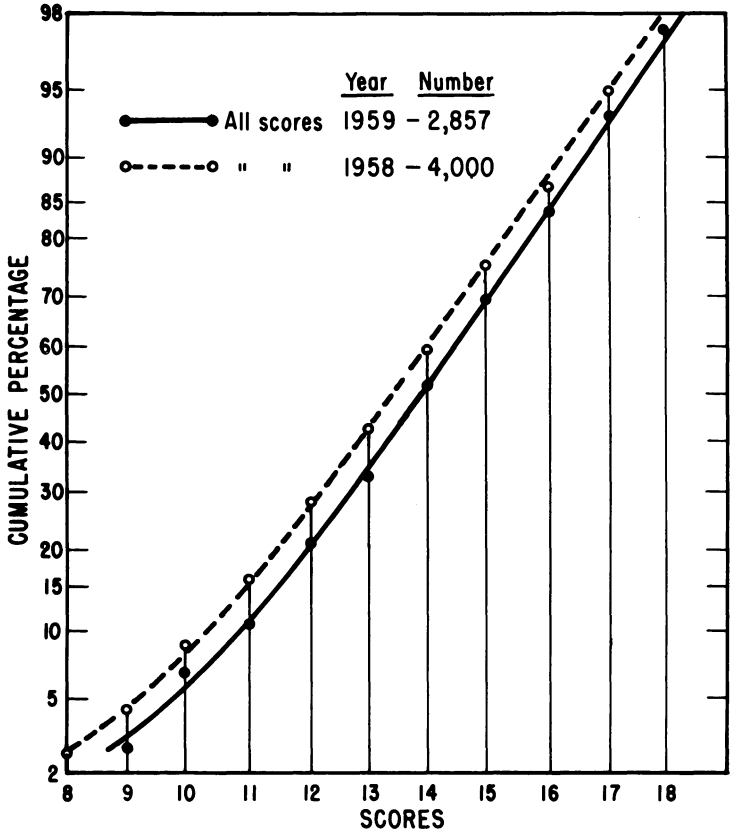

Fig. 10. Score distribution for all scores in 1958 and 1959.

points assigned to each classification. It should be stated that if the 1 to 20 score card (fig. 1) is used properly, it is extremely difficult to give excessively low (or high) values to wines and that if the directions are followed on the points assigned for each characteristic the 1 to 9 or 19 to 20 scores will occur rarely. However, there is a strong tendency for tasters not to fill in each assigned point value and add them up but to assign a final score after tasting the wine. Hence, there is more of a tendency to go oftener into the lower or higher ranges in an effort to use the complete range than would be done if the scoring of points for each characteristic was closely adhered to. To overcome this defect it might be more reasonable to assign the scoring on a 1 to 9 point basis with the following divisions—unacceptable (1) ; average quality

TABLE 1

SUMMARY OF SCORE RANGES, MEDIANS, AND MEANS FOR TWO YEARS

Distribution	Range (2–98%)[a]		Median (50%)[b]		Mean[c]	
	Years		Years		Years	
	1958	1959	1958	1959	1958	1959
	Tasters' scores					
Taster I	10.1–18.7	.....	14.4	....	14.8	....
Taster II	7.5–16.2	10.2–18.0	12.4	14.3	12.9	14.6
Taster III	9.0–17.4	8.5–17.5	13.2	13.0	13.5	13.4
Taster IV	7.0–19.1	8.0–20.0	13.2	14.0	13.7	14.4
Taster V	8.5–17.4	8.0–16.0	13.0	13.2	13.1	13.4
Taster VI	.....	8.0–19.0	....	15.2	....	15.1
Taster VII	9.7–17.6	9.7–17.6	13.7	13.7	14.2	14.0
Taster VIII	9.2–18.2	10.6–17.7	14.8	14.4	14.0	14.6
	Region-type scores					
WD I	8.7–18.3	9.5–18.0	13.5	13.7	13.9	14.2
WD IV	5.7–19.0	9.2–18.1	12.4	13.7	13.0	14.1
RD I	8.8–17.9	10.2–18.6	13.3	14.3	13.9	14.8
RD IV	7.0–17.6	7.6–18.4	12.4	13.1	12.8	13.5
WS IV	8.5–18.0	9.9–18.0	13.3	14.0	13.7	14.3
RS IV	8.7–18.2	6.8–18.3	14.1	14.1	14.5	14.6
Total scores	8.6–18.0	9.5–18.3	13.4	13.9	13.7	14.2
Total numbers	4,000	2,857	4,000	2,857	3,346	2,302

[a] Based on line of best fit if reasonable normal distribution indicated, otherwise best estimation of intersection with 2 per cent and 98 per cent lines.
[b] Based on all taste scores available and where actual plotted line goes through 50 per cent value.
[c] Based on those scores used in the analyses of variance.

but with defects (2 to 3); average quality (4 to 6); above average quality with some superior qualities (7 to 8); and superior quality (9). This gives a balanced scoring system with more points available for use in the middle ranges where most of the scoring occurs. It is realized that the degree of unacceptability of a wine has more range than 1 and likewise that of a superior quality wine. However, to get discrimination in the ranges where most interest is centered (between 2 to 8) it is felt that this type of scoring system is necessary and, further, the effective range of tasters will be shown to be no greater than 7 points (2 to 8) in later results and discussions.

Table 1 shows a summary of the ranges (in which 96 per cent of the scores occur), the median (50 per cent value of cumulative distribution), and the mean scores for the tasters, region-types, and total scores for the two years. There is a general tendency for the mean to be slightly larger than the median; this reflects the unbalanced use of the lower range of scores. The mean score range that would be expected if a balanced normal distribution were occurring would be 13 to 15 since this would be the score assigned to standard wines. With two exceptions the medians and the means fall within

TABLE 2

TESTS FOR HOMOGENEITY OF TASTERS AND TASTINGS VARIANCES FOR DAILY DUPLICATE SAMPLES

Year		Tasters								Bartlett's test F
		I	II	III	IV	V	VI	VII	VIII	
1958	N	56	56	56	56	56	....	56	56	
	Variance	3.70	2.41	1.98	4.66	3.54	....	2.53	2.09	3.05**
	Without taster IV									1.98
1959	N	..	21	38	42	42	32	33	31	
	Variance	....	2.59	1.66	5.78	.31	2.15	2.57	2.61	3.37**
	Without taster IV									0.51
	Comparison of 1958 to 1959 Ratio Test F	....	1.07	1.19	1.24	1.53	....	1.02	1.25	

Year		Tastings	df	Variance	Ratio test F
1958	Without tasters IV or VI	1st	168	3.38	
		2d	168	2.27	1.49**
	Without tasters I, IV, VI	Total	222	2.63	
1959	Without taster IV	1st	109	2.44	
		2d	87	2.68	1.10
	Without tasters I, IV, VI	Total	138	2.62	1.00

** Significant at 1 per cent level.

this range. There was a tendency on the part of several tasters to assign higher scores to the wines in 1959 than in 1958 (II, IV, V, VII) while taster III scored the wines slightly lower. This general increase in the 1959 scores is reflected in the median and mean scores, also by the region-type scores.

Homogeneity of Variances of Tasters. As a means of evaluating the tasters' error variances, in connection with the use of their scores in an analysis of variance of all the scores, the daily duplicate samples for each taster were summarized and Bartlett's test for homogeneity applied. As is shown in table 2 in both 1958 and 1959 the taster IV error variance was significantly larger than the others. Removal of taster IV in each case showed the rest of the tasters to have homogeneous variances. Testing the variances of tasters II, III, IV, VII, and VIII showed no significant change in the error variance between the two years on the individual tasters. Pooling the error variances of tasters I, II, III, V, VII, and VIII, and checking for difference between the first and second tasting in 1958 showed a highly significant decrease in the pooled error variance for the second tasting. The same test in 1959 using tasters II, III, V, VI, VII, and VIII showed no significant difference in the pooled error variance between the first and second tastings. Using only the tasters who had tasted for both years and testing the pooled variances be-

tween years showed no difference. The significant difference found would be expected since the panel was starting for the first time and several of the tasters had very little or no experience. It is, however, unexpected that the tasters' pooled error variance did not improve significantly the second year. Several explanations are available for this lack of change: (1) the panel reached a good efficiency the first year, or (2) less intense effort was made the second year by the tasters. It is thought that a mixture of both reasons probably was the cause.

Table 3 gives a comparison of the variances of years, tasters, and tastings with the error variances. It shows the improvement of the second tasting (end repeats) in 1958 over the first tasting (end repeats) and no significant difference for the pooled end repeat variances when compared to the pooled error variance. The comparison of the pooled end repeat variances of 1958 to 1959 showed significant improvement in 1959. There was no significant difference between the 1959 first and second tasting end repeat pooled variances. Individual significant increases or decreases in end repeat comparisons by tastings, error variances, and years are seen throughout the table. Why there is an improvement here between the end repeat pooled variances of 1959 and 1958 and not in the pooled error variances of the previous table between 1959 and 1958 is explainable by the fact that the tasters have become experienced to the point, in 1959, where they give more nearly the same score to a wine after retasting it, several weeks or more later, than they could in 1958. Apparently the ability to score a wine with the same score as in a previous day's tasting is learned quite rapidly while ability to score it the same way several weeks (or more) later is learned slightly less rapidly; however, the tasters had no chance to do the latter until the end of the 1958 first tasting.

Effective Range of Tasters. Table 4 is presented in an effort to show a method of evaluation of tasters which takes into consideration the range used by the taster and the taster's error variance. Range alone or error variance alone does not tell one much about the taster's ability. It was considered that a ratio of the range in which the middle 96 per cent of the scores occur and the square root of the error variance would be an appropriate measure. This term we will call the effective range. It is the range of points which the taster can use and have a standard deviation of 1.0. The table shows that the effective range of the tasters varies from 4.4 to 7.0. In the case of taster III, with the lowest standard deviation, he does have the most effective range. This evaluation does not take into consideration the distributions of the scores.

Analysis of Variance. The fundamental model for our analyses of variance is based on the assumptions of fixed effects and independent observations of equal variance (see Scheffé, 1959). The failure of this model in respect to equal variance is recognized and discussed in some detail. No extrapolation of our results is attempted and no fine distinction with respect to significance is made. More complicated models for the analyses of variance are suggested by the data, especially mixed models, but are not considered in detail because nearly all of the troublesome interactions are insignificant or can be reduced to insignificance by omitting one aberrant taster. Further,

TABLE 3

COMPARISON OF END REPEAT TASTING VARIANCES AND COMPARISONS TO DAILY DUPLICATE ERROR VARIANCE

End repeat tasting variances and degrees of freedom

Year	Testing		Tasters									Total—Taster IV
			I	II	III	IV	V	VI	VII	VIII	Total	
1958	1st	df	33	33	33	33	33	....	33	....	198	165
		variance	2.38	4.76	2.14	5.09	3.70	....	2.45	....	3.42	3.08
	2d	df	33	33	33	33	33	33	33	33	198	165
		variance	1.86	2.72	2.44	2.85	2.39	4.14	1.79	4.78	2.24	2.12
		F	1.28	1.75	1.14	1.79*	1.55	....	1.37	....	1.53**	1.45**
	1st and 2d	df	66	66	66	66	66	....	66	....	396	330
		variance	2.12	3.74	2.29	3.97	3.05	....	2.12	....	3.42	3.08
	Daily duplicate	df	56	56	56	56	56	....	56	....	336	280
		error variance	3.70	2.41	1.98	4.66	3.54	....	2.53	....	3.14	2.83
		F	1.75*	1.55	1.16	1.17	1.16	....	1.19	....	1.09	1.09
1959	1st	df	....	....	17	29	29	27	20	28	150	121
		variance	....	....	1.47	2.72	2.01	4.04	1.40	2.95	2.49	2.43
	2d	df	....	19	28	28	28	24	23	24	155	127
		variance	....	2.86	1.02	3.66	3.09	2.75	0.65	3.02	2.40	2.12
		F	....	....	1.44	1.34	1.54	1.47	2.15*	1.02	1.04	1.15
	1st and 2d	df	.....	19	45	57	57	51	43	52	324	248
		variance	....	2.86	1.21	3.19	2.54	3.46	0.91	3.04	2.48	2.27
	Daily duplicate	df	....	21	38	42	42	32	31	33	239	170
		error variance	....	2.59	1.66	5.78	2.31	2.15	2.61	2.57	3.07	2.32
		F	....	1.10	1.37	1.82*	1.10	1.61	2.87**	1.18	1.24	1.04
1958	1st and 2d	df	....	66	66	66	66	33	66	33	396	330
		variance	....	3.74	2.29	3.97	3.05	4.14	2.12	4.78	3.27	3.13
1959	1st and 2d	df	....	19	45	57	57	51	43	52	324	267
		variance	....	2.86	1.21	3.19	2.54	3.46	0.91	3.04	2.48	2.33
		F	....	1.30	1.89*	1.24	1.21	1.20	2.32**	1.57	1.32**	1.38**

* Significant at 5 per cent level.
** Significant at 1 per cent level.

TABLE 4
RATIO OF TASTER'S RANGE TO STANDARD DEVIATION AS A MEASURE OF ABILITY

Tasters	Range used (for 96% of scores)		Standard deviation $\sqrt{\text{error variance (daily duplicate)}}$		Effective range $\frac{\text{range 96\%}}{\sqrt{\text{Error variance}}}$	
	1958	1959	1958	1959	1958	1959
I	8.6	....	1.92	....	4.5	....
II	8.7	7.8	1.55	1.61	5.6	4.8
III	8.4	9.0	1.40	1.29	6.0	7.0
IV	12.1	12.0	2.16	2.40	5.6	5.0
V	8.9	8.0	1.88	1.52	4.7	5.3
VI	...	11.0	....	1.47	...	6.8
VII	7.9	7.9	1.59	1.60	5.0	4.9
VIII	9.0	7.1	1.45	1.62	6.2	4.4
Average	.....	.....	.....	.....	5.37	5.46
II-V, VII-VIII	.....	.....	.....	.....	5.51	5.24

a complete specification of a realistic model is impossible and we would not expect such a model to modify much the importance of the various elements of structure that our analyses indicate are present in the data.

The variables tested in these tastings are: wines, tasters, tasting (first and second), region-type, color, and the various interactions. It is desirable to get the lowest correct remainder error term. This can be checked by comparing the remainder error term with the daily duplicate error variance. If the remainder error is larger by a significant amount than the daily duplicate error term it may indicate that some methodological error is involved in the tasting or that homogeneity of tasters may be lacking. Such errors can and did arise in our tasting to some degree. If the remainder term is significantly larger and suspicion exists that methodological or other errors are present, the remainder error should be used for tests of significance of the other variables.

Table 5 shows a comparison of analyses of first and second tastings in 1958 by region-types. The variables considered in this case are wines and tasters and the remainder error (wine × taster interaction) should check with the duplicate error term (this is the pooled variances of the end repeats and daily duplicates). It does not in all cases. This analysis uses taster IV and, as was shown previously, his error variance is not compatible with the rest of the tasters. Taster VI was in the second tasting in these analyses. This taster's pooled error variance (end repeats and daily duplicates) for the second tasting was 4.38. This value is almost as large as taster IV's (4.61) and, while the calculations were not made, would very likely not be homogeneous with the rest of the tasters' error variances. In every instance the tasters show significant differences in mean scores and in all instances but two (first tasting DW IV and WS IV) the wine mean scores show significant differences.

TABLE 5

COMPARISON OF THE FIRST AND SECOND TASTING ANALYSIS FOR 1958 AND A COMPARISON OF THE TASTER × WINE INTERACTION VARIANCE TO DUPLICATE VARIANCE

Region-type	Source of variations	First tasting			Second tasting		
		df	ss	ms	df	ss	ms
DW I	Total	755	5,510.74	.....	871	3,668.48	....
	Wines	107	1,871.31	17.49**	108	754.73	6.99**
	Tasters	6	497.18	82.86**	7	515.16	73.59**
	Error (remainder)	642	3,142.25	4.89	756	2,398.59	3.17
	Error (duplicate)	91	.....	4.26	72	.....	2.62
DW IV	Total	83	864.04	.....	111	855.96	....
	Wines	13	128.54	9.98	13	180.96	13.92*
	Tasters	5	204.68	40.94**	7	103.39	14.77*
	Error (remainder)	65	530.82	8.17*	91	571.61	6.28
	Error (duplicate)	24	.....	4.40	56	.....	7.56
RD I	Total	447	2,438.99	.....	519	2,672.95	.....
	Wines	63	714.70	11.34**	64	836.57	13.07**
	Tasters	6	169.30	28.22**	7	397.84	56.83**
	Error (remainder)	378	1,554.99	4.11	448	1,438.54	3.21
	Error (duplicate)	35	.....	3.69	47	.....	2.26
RD IV	Total	251	1,899.08	.....	287	1,928.75	.....
	Wines	35	664.79	18.99**	35	756.13	21.60**
	Tasters	6	126.80	21.13**	7	251.22	35.89**
	Error (remainder)	210	1,107.49	5.27**	245	921.40	3.76*
	Error (duplicate)	60	.....	2.91	64	.....	2.58
WS IV	Total	173	1,160.93	.....	231	986.00	.....
	Wines	28	187.76	6.71	28	208.62	7.45**
	Tasters	5	349.34	69.87**	7	180.69	25.81**
	Error (remainder)	140	623.83	4.46*	196	596.69	3.04**
	Error (duplicate)	144	.....	3.23	176	.....	2.12
RS IV	Total	179	1,116.91	.....	239	1,222.40	.....
	Wines	29	354.24	12.22**	29	311.52	10.74**
	Tasters	5	124.31	24.86**	7	92.83	12.36**
	Error (remainder)	145	638.36	4.40	203	818.05	4.03*
	Error (duplicate)	54	.....	3.32	72	.....	2.55

* Significant at 5 per cent level.
** Significant at 1 per cent level.

Combining the tastings and running the analyses again by region-types will show the interactions of the tasters × tastings and wines × tastings. This was done on the 1958 wines without taster VI, in several cases (DW IV, WS IV, and RS IV) without taster VII and in two instances (DW I and DW IV) with and without taster IV. The 1959 combined tastings are shown in comparison using tasters III, IV, V, and without using taster IV. The pooled error variances for the daily duplicates and end repeats are also shown

TABLE 6

COMPARISON OF ANALYSES OF VARIANCE OF WD I WITH AND WITHOUT TASTER IV FOR 1958 AND 1959

Region-type	Year	Source of variation	With taster IV			Without taster IV		
			df	ss	ms	df	ss	ms
WD I	1958[a]	Total	391	2,827.3	.....	335	1,926.0	.....
		Wines	27	593.6	21.98**	27	495.3	6.44**
		Tastings	1	131.5	131.50**	1	80.0	80.00**
		Tasters[b]	6	329.3	54.88**	5	174.5	34.90**
		Wines × tastings	27	290.4	10.75**	27	157.7	5.84**
		Wines × tasters	162	801.9	4.95**	135	616.2	4.56**
		Tastings × tasters	6	163.4	27.23**	5	17.7	3.54
		Error (remainder)	162	514.5	3.18	135	384.6	2.85
		Error (duplicate)[c]	154	.....	3.53	132	.....	3.16
WD I	1959	Total	221	1,113.9	.....	147	472.8	.....
		Wines	36	261.9	7.28*	36	246.5	6.85**
		Tastings	1	214.2	214.20**	1	23.5	23.50**
		Tasters[d]	2	8.2	4.10	1	00.8	0.80
		Wines × tastings	36	104.5	2.90	36	81.4	2.26
		Wines × tasters	72	172.8	2.40	36	69.0	1.92
		Tastings × tasters	2	51.5	25.75**	1	0.3	0.30
		Error (remainder)	72	300.8	4.18	36	51.4	2.48
		Error (duplicate)[e]	122	.....	3.14	80	.....	2.00

* Significant at 5 per cent level.
** Significant at 1 per cent level.
[a] This is analysis of the first 28 DW I wines—analysis of the complete number of wines shows essentially the same results.
[b] For tasters I, II, III, IV, V, VII, VIII.
[c] For tasters III, IV, V.
[d] Daily duplicate error term of wines of this region type.
[e] Pooled daily duplicate error term of all region-type wines.

for the 1958 analyses (see tables 6 to 11). The error (duplicate) appears to agree more favorably with the remainder error in this analysis of the 1958 tasting. In only two instances are the remainder error variances significantly larger than duplicate error variances. Removal of taster IV from this 1958 analysis seemed to have relatively little effect in the two instances (DW I and DW IV). However, in the 1959 analysis, in general, the interaction variances decreased in significance markedly when taster IV was removed from the analyses. This shows that not only was taster IV more variable but also quite different in his reactions to the various wines compared to the other tasters. However, RS IV and WS IV wines were scored differently by tasters III and V between tastings, as is shown by a significant tasting × taster interaction. Also in the RS IV analysis (without taster IV) a significant wine × taster interaction appears. In all instances in these analyses the wines are shown to have significant differences. Comparison of the error (remainder) variances for the 1959 analyses by region-types to the 1959 pooled daily duplicate error variance with and without taster IV are also given in tables 6, 7, 8, 9, 10, and 11. It is shown that without taster IV the error (remainder) variances are all compatible with daily duplicate error variances. In two instances with taster IV included, the F values be-

TABLE 7

COMPARISON OF ANALYSES OF VARIANCE OF WD IV WITH AND WITHOUT TASTER IV FOR 1958 AND 1959

Region-type	Year	Source of variation	With taster IV			Without taster IV		
			df	ss	ms	df	ss	ms
WD IV	1958	Total	167	1,624.0	.....	139	1,276.8	.....
		Wines	13	268.0	20.62**	13	182.9	14.07*
		Tastings	1	127.0	127.00**	1	61.7	61.70**
		Tasters[a]	5	239.0	47.80**	4	238.9	59.72**
		Wines × tastings	13	34.0	2.62	13	87.8	6.75
		Wines × tasters	65	343.0	5.28	52	294.5	5.66
		Tastings × tasters	5	60.0	12.00	4	24.7	6.18
		Error (remainder)	65	533.0	8.20	52	386.3	7.43*
		Error (duplicate)[b]	66	.....	5.52	55	.....	4.64
WD IV	1959	Total	269	1,521.0	.....	179	793.5	.....
		Wines	44	458.1	10.41**	44	436.2	4.00**
		Tastings	1	52.5	52.50**	1	1.5	1.50
		Tasters[c]	2	94.9	4.75	1	6.0	6.00
		Wines × tastings	44	156.4	3.55*	44	104.3	2.37
		Wines × tasters	88	498.5	5.66**	44	127.8	2.90
		Tastings × tasters	2	72.6	36.30**	1	8.5	8.50
		Error (remainder)	88	188.0	2.14	44	109.2	2.48
		Error (duplicate)[d]	122	.....	3.14*	80	.....	2.00

* Significant at 5 per cent level.
** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII.
[b] Daily duplicate error term of wines of this region type.
[c] Using tasters III, IV, V.
[d] Pooled daily duplicate error term of all region-type wines.

came significant. The pooled total error variances are not significantly different from the remainder error for either analysis for 1959.

Several simple analyses were made using all the tasters and scores of both years for differences in tastings, tasters, color with and without taster IV (taster VI not included in 1958). These data (table 12) indicate that all the analyses gave significant variances except color when taster IV was included in the analyses. The use of color as a variable is questionable because the analysis is weighted heavily by a large excess of white wines in 1958. It is of note that the 1959 analyses show no significant variance on the color variable and that the tasting variance ceases to be significant with the removal of taster IV from the analysis. To further investigate the tasting differences a simple analysis was done by region-type to show where the differences in tastings were occurring. Table 13 gives this information for 1958 without taster IV and for 1959 with and without taster IV. Reference to tables 6 to 11 will give a rough comparison of 1958 with taster IV. It is shown that the removal of taster IV has a strong effect on the tasting variance, usually decreasing it. In the case of RD IV 1959, the significant variance actually represents a decrease in the mean score of the second tasting. All other significant variances indicated increases in the second tasting mean score. Thus the RD IV 1959 variance increase is comparable to vari-

TABLE 8

COMPARISON OF ANALYSES OF VARIANCE OF RD I FOR 1959 WITH AND WITHOUT TASTER IV AND FOR 1958 WITH TASTER IV

Region-type	Year	Source of variation	With taster IV			Without taster IV		
			df	ss	ms	df	ss	ms
RD I	1958	Total	895	4,506.0	....	....	....	....
		Wines	63	1,140.0	18.10**	....	....	....
		Tastings	1	16.0	16.00*	....	....	....
		Tasters[a]	6	377.0	62.83**	....	....	....
		Wines × tastings	63	282.0	4.48*	....	....	....
		Wines × tasters	378	1,453.0	3.84	....	....	....
		Tasters × tastings	6	12.0	2.00	....	....	....
		Error (remainder)	378	1,226.0	3.24	....	....	....
		Error (duplicate)[b]	76	....	2.90	....	....	....
RD I	1959	Total	197	967.7	....	131	488.5	....
		Wines	32	215.0	6.72*	32	192.5	6.02*
		Tastings	1	3.4	3.40	1	1.4	1.40
		Tasters[c]	2	109.0	54.50**	1	5.9	5.90
		Wines × tastings	32	204.6	6.39*	32	122.6	3.83
		Wines × tasters	64	186.7	2.92	32	80.1	2.50
		Tastings × wines	2	22.8	11.40*	1	0.6	0.60
		Error (remainder)	64	226.2	3.53	32	85.4	2.67
		Error (duplicate)[d]	122	....	3.14	80	....	2 00

* Significant at 5 per cent level.
** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII, VIII.
[b] See footnote on table 7.
[c] Using tasters III, IV, V.
[d] See footnote on table 7.

ance decrease in the other instances as far as the effect of taster IV is concerned.

In order to find the region-type variance and the interaction with the tasters and tastings, 10 wines were selected at random from each of the region-types and an analysis of variance (table 14) was made using all the tasters in 1958 except taster VI and all tasters except I and II in 1959. A second analysis was done on 10 wines of each region-type in 1958 using tasters I, III, V, VII, and VIII and in 1959 using tasters III, V, VI, VII, and VIII. The removal of tasters II and IV from the 1958 analysis caused no change in the significances but in 1959 the removal of taster IV showed again the taster × tasting interaction produced by this taster and also the decline to insignificance of the tasting variance when taster IV is removed from the analysis. The 1958 analysis showed a significant difference in region-type scores and a significant region-type × tasting interaction. In 1959 again a significant region-type variance is noted and in this year a region-type × taster interaction of significance was noted.

Table 15 gives a summary of the mean scores of the tasters, tastings, region-types, and years. The total mean scores for 1958 and 1959 cannot be compared because of the differing numbers of region-types tested between the two years but the rest of the mean scores may be tested using one of the

TABLE 9

COMPARISON OF ANALYSES OF VARIANCE FOR RD IV IN 1958 WITH TASTER IV AND IN 1959 WITH AND WITHOUT TASTER IV

Region-type	Year	Source of variation	With taster IV			Without taster IV		
			df	ss	ms	df	ss	ms
RD IV	1958	Total	503	3,500.0	....	....	....	....
		Wines	35	1,056.0	30.17**	....	....	....
		Tastings	1	28.0	28.00**	....	....	....
		Tasters[a]	6	332.0	55.33**	....	....	....
		Wines × tastings	35	221.0	6.31*	....	....	....
		Wines × tasters	210	1,055.0	5.02*	....	....	....
		Tastings × tasters	6	37.0	6.17	....	....	....
		Error (remainder)	210	771.0	3.67*	....	....	....
		Error (duplicate)[b]	116	....	2.74	....	....	....
RD IV	1959	Total	221	1,408.8	....	147	657.7	....
		Wines	36	410.8	11.41**	36	269.2	7.48**
		Tastings	1	4.0	4.00	1	45.4	45.40**
		Tasters[c]	2	288.5	144.25**	1	31.2	31.20**
		Wines × tastings	36	141.5	3.93**	36	120.1	1.10
		Wines × tasters	72	279.5	3.77**	36	82.3	0.76
		Tastings × tasters	2	178.4	89.20**	1	0.5	0.17
		Error (remainder)	72	114.3	1.59	36	109.0	3.03
		Error (duplicate)[d]	122	....	3.14*	80	....	2.00

* Significant at 5 per cent level.
** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII, VIII
[b] See footnote on table 7.
[c] Using tasters III, IV, V.
[d] See footnote on table 7.

TABLE 10

COMPARISON OF ANALYSES OF VARIANCE FOR WS IV IN 1958 WITH TASTER IV AND IN 1959 WITH AND WITHOUT TASTER IV

Region-type	Year	Source of variation	With taster IV			Without taster IV		
			df	ss	ms	df	ss	ms
WS IV	1958	Total	347	1,779.0	....	....	....	....
		Wines	28	214.0	7.64**	....	....	....
		Tastings	1	60.0	60.00**	....	....	....
		Tasters[a]	5	262.0	52.40**	....	....	....
		Wines × tastings	28	132.0	4.71	....	....	....
		Wines × tasters	140	559.0	3.99	....	....	....
		Tastings × tasters	5	135.0	27.00**	....	....	....
		Error (remainder)	140	417.0	2.98	....	....	....
		Error (duplicate)[b]	298	....	2.53	....	....	....
WS IV	1959	Total	83	368.8	....	55	116.0	....
		Wines	13	99.0	7.61**	13	59.2	4.55**
		Tastings	1	18.2	18.20**	1	0.5	0.50
		Tasters[c]	2	14.6	7.30*	1	0.2	0.20
		Wines × tastings	13	28.3	2.18	13	28.8	2.21
		Wines × tasters	26	129.7	4.99*	13	8.1	0.62
		Tastings × tasters	2	28.6	14.30**	1	5.1	5.10*
		Error (remainder)	26	50.4	1.94	13	14.1	1.08
		Error (duplicate)[d]	122	....	3.14**	80	....	2.00

* Significant at 5 per cent level.
** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII.
[b] See footnote on table 7.
[c] Using tasters III, IV, V.
[d] See footnote on table 7.

TABLE 11

COMPARISON OF ANALYSES OF VARIANCE FOR RS IV FOR 1958 AND 1959 WITH TASTER IV AND WITHOUT TASTER IV

Region-type	Year	Source of variation	With taster IV			Without taster IV		
			df	ss	ms	df	ss	ms
RS IV	1958	Total	359	2,054.0	. . .	....	....	....
		Wines	29	527.0	18.17**	....	....	....
		Tastings	1	1.0	1.00	....	....	....
		Tasters[a]	5	154.0	30.80**	....	....	....
		Wines × tastings	29	117.0	4.03	....	....	....
		Wines × tasters	145	775.0	5.52**	....	....	....
		Tastings × tasters	5	26.0	5.20	....	....	....
		Error (remainder)	145	454.0	3.13	....	....	....
		Error (duplicate)[b]	117	....	2.94	....	....	....
RS IV	1959	Total	83	574.3	. . .	55	365.8	. .
		Wine	13	311.0	23.92**	13	221.1	17.07**
		Tastings	1	0.0	0.00	1	13.0	13.00**
		Tasters[c]	2	15.6	7.80	1	0.1	0.10
		Wines × tastings	13	15.3	1.18	13	25.2	1.93
		Wines × tasters	26	108.7	4.18	13	75.1	5.77**
		Tastings × tasters	2	52.1	26.05**	1	13.1	13.10**
		Error (remainder)	26	71.6	2.75	13	18.2	1.40
		Error (duplicate)[d]	122	....	3.14	80	....	2.00

** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII.
[b] See footnote on table 7.
[c] Using tasters III, IV, V.
[d] See footnote on table 7.

TABLE 12

COMPARISON OF COMPLETE SCORE ANALYSIS FOR 1958 AND 1959 WITH AND WITHOUT TASTER IV

Year	Source of variation	With taster IV			Without taster IV		
		df	ss	ms	df	ss	ms
1958[a]	Total	3,787	22,695.0	.....	3,346	21,619.0	.....
	Tastings	1	43.0	43.00**	1	85.6	85.60**
	Error (remainder)	3,786	22,652.0	5.98	3,345	21,533.4	6.44
	Tasters	6	1,341.0	223.50**	5	1,399.0	226.90**
	Error (remainder)	3,781	21,354.0	5.65	3,341	20,289.0	6.07
	Color (of wine)	1	9.0	9.00	1	54.0	54.00**
	Error (remainder)	3,786	22,686.0	5.99	3,345	21,565.0	6.45
	Region-type	5	790.0	158.00**	5	997.9	199.50**
	Error (remainder)	3,782	21,905.0	5.79	3,341	20,621.0	6.17
1959[b]	Total	2,301	15,229.0	.....	1,940	12,274.0	.....
	Tastings	1	162.2	162.00**	1	15.5	15.50
	Error (remainder)	2,300	15,066.8	6.55	1,939	12,258.7	6.32
	Tasters	6	784.0	130.70**	5	764.6	152.90**
	Error (remainder)	2,295	14,445.0	6.29	1,935	11,509.6	5.95
	Color	1	0.4	0.40	1	13.7	13.70
	Error (remainder)	2,300	15,228.6	6.62	1,939	12,260.5	6.32
	Region-type	5	410.8	82.16**	5	460.0	92.00**
	Error (remainder)	2,296	14,818.0	6.45	1,935	11,814.2	6.12

** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII, VIII.
[b] Using tasters II, III, IV, V, VI, VII, VIII.

TABLE 13

COMPARISON OF ANALYSES ON COMPLETE SCORES FOR TASTINGS BY REGION-TYPE IN 1958 AND 1959 WITH AND WITHOUT TASTER IV

Region-type	Source of variation	1958[a]			1959[b]					
		Without taster IV			Without taster IV			With taster IV		
		df	ss	ms	df	ss	ms	df	ss	ms
WD I	Total	1,295	6,743.8		379	1,384.4		453	2,050.0	
	Tastings	1	5.8	5.80	1	144.0	144.00**	1	178.8	178.88**
	Error	1,294	6,738.0	5.24	378	1,240.4	3.28	452	1,871.2	4.14
WD IV	Total	164	1,512.0		492	4,812.0		583	5,680.4	
	Tastings	1	112.1	112.10**	1	40.0	40.00*	1	124.7	124.70**
	Error	163	1,399.9	8.22	491	4,722.0	9.70	582	555.7	9.55
RD I	Total	767	5,740.0		348	1,496.0		414	1,902.4	
	Tastings	1	13.1	13.10	1	1.8	1.80	1	10.9	10.90
	Error	766	5,726.9	7.48	347	1,494.2	4.31	413	1,891.5	4.58
RD IV	Total	431	2,848.7		407	2,769.7		481	2,363.2	
	Tastings	1	24.1	24.10	1	41.1	41.10*	1	11.2	11.20
	Error	430	2,824.6	6.57	406	2,728.6	6.72	480	2,352.0	4.90
WS IV	Total	336	1,693.0		154	536.3		182	812.9	
	Tastings	1	29.1	29.10*	1	1.9	1.90	1	14.4	14.40
	Error	335	1,663.9	4.97	153	534.4	3.49	181	798.5	4.41
RS IV	Total	347	2,231.6		155	816.1		183	1,009.3	
	Tastings	1	0.1	0.10	1	0.0	0.00	1	12.0	12.00
	Error	346	2,231.5	6.45	154	816.1	5.30	182.1	997.3	5.48

* Significant at 5 per cent level.
** Significant at 1 per cent level.
[a] Using tasters I, II, III, V, VII, VIII.
[b] Using tasters II, III, IV, V, VI, VII, VIII

TABLE 14

COMPARISON OF ANALYSES OF TEN WINES PICKED AT RANDOM FROM EACH REGION-TYPE GROUP FOR 1958 AND FOR 1959 WITH AND WITHOUT TASTER IV

Year	Source of variation	With taster IV			Without taster IV		
		df	ss	ms	df	ss	ms
1958[a]	Total	839	5,428.0	.....	599	3,314.9	.....
	Region-type	5	316.0	63.20**	5	232.9	46.58**
	Tasters	6	258.0	43.00**	4	182.0	45.50**
	Tastings	1	166.0	166.00**	1	66.6	66.66**
	Region-type × tasters	30	251.0	8.37	20	144.2	7.21
	Region-type × tastings	5	130.0	26.00**	5	85.7	17.14**
	Tasters × tastings	6	55.0	9.17	4	22.2	5.55
	Error	786	4,252.0	5.41	560	2,581.3	4.61
1959[b]	Total	719	3,508.0	.....	599	2,662.3	.....
	Region-type	5	131.8	26.36**	5	155.7	31.14**
	Tasters	5	155.7	31.14**	4	253.7	63.43**
	Tastings	1	71.8	71.80**	1	13.7	13.70
	Region-type × tasters	25	381.2	15.25**	20	179.1	8.95**
	Region-type × tastings	5	35.7	7.14	5	33.3	6.66
	Tasters × tastings	5	221.0	44.20**	4	25.0	6.25
	Error	673	2,510.8	3.73	560	2,001.8	3.57

** Significant at 1 per cent level.
[a] Using tasters I, II, III, IV, V, VII, VIII.
[b] Using tasters III, IV, V, VI, VII, VIII.

TABLE 15
SUMMARY OF MEAN SCORES

Region-type	Year	Tasting	Tasters								Total	Total without taster IV
			I	II	III	IV	V	VI	VII	VIII		
WD I	1958	1st	14.96	12.77	14.26	12.84	13.00	.....	14.44	13.95	13.75	13.43
		2d	14.86	12.89	14.23	13.81	13.42	.....	14.75	14.08	14.01	14.41
		Total	14.91	12.83	14.25	13.32	13.21	.....	14.60	14.02	13.88	13.92
	1959	1st	.....	13.95	13.62	11.48	13.65	14.84	14.00	14.16	13.65	14.04
		2d	.....	14.66	14.54	15.78	14.32	15.68	14.83	14.61	14.91	14.73
		Total	.....	14.30	14.08	13.63	13.98	15.34	14.33	14.37	14.24	14.36
WD IV	1958	1st	15.00	10.29	12.50	11.14	10.93	.....	12.79	11.27	11.97	12.30
		2d	15.64	13.00	13.43	14.93	12.79	.....	13.29	14.71	14.05	13.63
		Total	15.32	11.64	12.96	13.04	11.86	.....	13.04	13.20	13.02	12.97
	1959	1st	.....	14.61	13.39	13.17	12.61	14.76	12.95	14.35	13.66	13.87
		2d	.....	14.24	13.44	15.73	13.51	15.56	14.76	14.89	14.55	14.32
		Total	.....	14.42	13.41	14.45	13.06	15.19	13.62	14.58	14.08	14.01
RD I	1958	1st	14.08	12.91	12.95	14.30	13.34	.....	14.13	14.52	13.75	13.65
		2d	14.63	13.44	12.72	14.55	13.61	.....	14.45	14.66	14.01	13.92
		Total	14.35	13.17	12.84	14.42	13.48	.....	14.29	14.59	13.88	13.79
	1959	1st	.....	15.53	14.15	14.82	13.85	15.09	13.97	15.48	14.65	14.61
		2d	.....	15.25	14.06	16.03	13.52	15.62	15.52	15.25	14.97	14.76
		Total	.....	15.40	14.10	15.42	13.68	15.32	14.58	15.37	14.80	14.68
RD IV	1958	1st	13.00	11.33	12.00	13.47	12.17	.....	13.14	13.00	12.59	12.44
		2d	14.47	12.06	12.11	13.92	11.81	.....	13.28	13.75	13.06	12.91
		Total	13.74	11.69	12.06	13.67	11.98	.....	13.21	13.06	12.82	12.68
	1959	1st	.....	14.35	12.46	13.89	13.27	14.81	13.27	13.89	13.68	13.64
		2d	.....	13.30	11.24	15.30	12.27	14.42	13.03	14.22	13.37	13.01
		Total	.....	13.82	11.85	14.59	12.77	14.63	13.16	14.05	13.53	13.33
WS IV	1958	1st	15.55	12.69	14.52	11.21	14.00	.....	12.69	12.00	13.23	13.89
		2d	15.14	13.97	14.34	13.93	13.55	.....	14.69	13.93	14.23	14.34
		Total	15.34	13.33	14.43	12.57	13.78	.....	13.69	13.19	13.73	14.12
	1959	1st	.....	14.77	13.79	12.14	14.50	14.83	14.09	15.07	14.15	14.51
		2d	.....	15.56	14.57	14.57	14.07	15.79	14.71	13.92	14.71	14.74
		Total	.....	15.10	14.18	13.35	14.29	15.35	14.43	14.50	14.43	14.63
RS IV	1958	1st	16.00	13.40	13.73	14.77	14.37	.....	14.60	14.88	14.54	14.42
		2d	15.33	14.40	13.73	15.17	14.60	.....	14.13	14.53	14.56	14.44
		Total	15.67	13.90	13.73	14.97	14.48	.....	14.37	14.67	14.55	14.43
	1959	1st	.....	15.50	13.50	13.50	14.57	15.36	14.21	14.29	14.42	14.57
		2d	.....	15.66	13.50	15.43	12.64	15.72	15.57	15.20	14.72	14.58
		Total	.....	15.58	13.50	14.46	13.61	15.54	14.89	14.67	14.56	14.58

TABLE 16

MEAN WINE, TASTER, AND TASTING SCORES FOR WD IV, 1958

Wines	9112B	9021	9057	8835H	9112A	9082	8936	8762	8835B	8751A	9083B	9945	8751B	9038A
Mean scores[a]	10.25	11.58	12.50	12.67	12.92	12.92	12.92	13.00	13.17	13.42	13.75	14.08	14.08	14.67
5% level														

Tasters	II	V	III	VII	IV	I
Mean scores[a]	11.64	11.86	12.96	13.04	13.04	15.32
5% level						

Tastings	1st	2d
Mean scores[a]	12.107	13.845
5% level		

Variables $S_{\bar{x}}$ (standard error)

Wines	0.82
Tasters	0.66
Tasting	0.31

[a] Mean scores underlined by the same line indicate no significant difference in those mean scores.

TABLE 17

MEAN WINE, TASTER, AND TASTING SCORES FOR WS IV, 1959

Wines	9174	9345	9255B	9310	9255C	9307	9328A	9328B	9256	9344	9258	9257	9259	9358A
Mean scores[a]	12.50	12.75	13.00	13.25	13.50	14.00	14.25	14.75	14.75	15.00	15.00	15.50	15.50	15.50
5%														

Tasters	III	V
Mean scores[a]	14.18	14.29
5%		

Tastings	1st	2d
Mean scores[a]	14.14	14.32
5%		

Variables $S_{\bar{x}}$

Wines	0.52
Tasters	0.19
Tastings	0.19

[a] Mean scores underlined by the same line indicate no significant difference in those mean scores.

TABLE 18

COMPARISON OF MEAN SCORES OF REGION-TYPE AND TASTERS FOR 1958 AND 1959

(10 wines at random from each region-type)

	1958 (with taster IV)						1958 (without taster IV or II)					
Region-type..	WD IV	RD IV	WS IV	RD I	WD I	RS IV	RD IV	WD IV	RD I	WS IV	WD I	RS IV
Mean score[a]...	13.01	13.15	13.92	14.03	14.14	14.82	13.08	13.21	13.97	14.29	14.40	14.79
5%	($S_{\bar{x}}$ = 0.197)						($S_{\bar{x}}$ = 0.215)					
Tasters.....	II	V	III	VII	VIII	IV	I	V	III	VII	VIII	I
Mean score[a]...	13.12	13.54	13.67	13.69	13.84	14.03	15.03	13.54	13.67	13.69	13.84	15.03
5%	($S_{\bar{x}}$ = 0.212)						($S_{\bar{x}}$ = 0.196)					
	1959 (with taster IV)						1959 (without taster IV)					
Region-type..	RD IV	WD IV	RS IV	WS IV	WD I	RD I	RD IV	WD IV	RS IV	WS IV	RD I	WD I
Mean score[a]...	13.73	14.23	14.40	14.52	14.87	15.03	13.48	14.23	14.37	14.74	14.84	15.02
5%	($S_{\bar{x}}$ = 0.167)						($S_{\bar{x}}$ = 0.192)					
Tasters.....	V	III	IV	VII	VIII	VI	V	III	VII	VIII	VI	
Mean score[a]...	13.55	13.89	14.55	14.67	14.75	15.38	13.55	13.89	14.67	14.75	15.38	
5%	($S_{\bar{x}}$ = 0.176)						($S_{\bar{x}}$ = 0.170)					

[a] Mean scores underlined by the same line indicate no significant difference in those mean scores.

TABLE 19

1958 AND 1959 COMPARISON OF REGION-TYPE MEAN SCORES ON THE COMPLETE SCORES WITH ALL TASTERS INCLUDED

				1958			
Region-type.....	RD III	WD IV	WS IV	RD I	WD I	RS IV	
Mean score[a]	12.82	13.02	13.73	13.88	13.88	14.55	$S_{\bar{x}} = 0.102$
5%							

				1959			
Region-type.....	RD IV	WD IV	WD I	WS IV	RS IV	RD I	
Mean score[a]	13.53	14.08	14.24	14.43	14.56	14.80	$S_{\bar{x}} = 0.132$
5%							

[a] Mean scores underlined by the same line are not significantly different.

methods such as that of Duncan (1955). The tables to include the mean scores of all the wines would be much too extensive; however, an example of one of the smaller groups, WD IV 1958, is shown in table 16 with significant differences indicated. This example has one of the highest error terms and hence a higher standard error than would normally be encountered. Table 17 shows a similar analysis of another group of wines, WS IV 1959, using only two tasters. This error variance is one of the lowest found and gives a much better resolution of score differences than did the previous analysis shown despite the fewer number of tastings per wine in the latter (12 to 4).

Some other mean scores of interest are those of the analyses of the 10 samples from each region-type group. These are presented in table 18 and show the effects on the mean scores of region-types and tasters for the two years and changes occurring by the removal of taster IV from the analysis. It is of interest to note that the tasters who are involved both years in this analysis have mean scores of the same general order for both years. It is also evident that the dry wines produced from region I are given significantly better scores than the dry wines produced from region IV. Shown in table 19 is a similar analysis of the complete region-type mean scores for 1958 and 1959 for all the tasters. The change in the relative position of the RS IV 1959 can be attributed to the effect of taster II's higher evaluation of the RS IV wines (taster II was not included in the analysis of table 18, 1959). Otherwise the agreement is very good between the two tables. The relatively high standard error ($S_{\bar{x}}$) for the 1959 complete mean scores (table 19) is due to the fact that many of the tasters did not taste all of the wines. This lowers the effective number of replications (K_o) defined by the formula:

$$K_o = \frac{1}{n-1}\left(\sum K_i - \frac{\sum K_i^2}{\sum K_i}\right) \qquad S_{\bar{x}} = \frac{\sqrt{\text{error variance}}}{K_o}$$

n = number of tasters K_i = number of scores of each taster

TABLE 20

CLASSIFICATION OF MEAN SCORES FOR RD I 1958 WINES, USING ALL TASTERS

Classification range mean	Poor (1) 0–11.1		Below average (2) 11.10–12.20 11.65		Average (3) 12.20–14.40 13.30		Above average (4) 14.4–15.50 14.95		Superior (5) 15.50–20	
Confidence intervals	9.4–12.8		10.5–13.9		12.7–16.1		13.8–17.1			
	Wine number and score									
	Wine number	Score	Wine number	Score	Wine number	Score	Wine number	Score	Wine number	Score
	9072	10.00	8898	12.20	8748	13.85	8783	14.65	9026	15.50
			8902	12.20	8809	14.35	8848	14.65	9029	15.65
			8911	11.92	8810	14.05	8897	14.55		
			8849	11.20	8826	14.40	8913	14.65		
			9074	11.40	8895	14.20	8914	14.80		
			8910	11.35	8899	13.50	8923	14.55		
					8900	13.70	8927	14.70		
					8903	13.60	8995	14.80		
					8905	13.65	8997	14.50		
					8908	13.40	8999	14.65		
					8909	14.00	9000	14.85		
					8912	14.15	9004	14.45		
					8924	14.05	9070	14.50		
					8928	14.00	9071	14.50		
					8996	13.85	9075	14.65		
					9001	13.80	9070C	14.70		
					9011	13.80	9080B	14.70		
					9027	14.15	8926	15.20		
					9030	14.15	8998	15.15		
					9038	14.35	9005	15.05		
					9076	14.30	9006	14.95		
					8896	13.05	9008	15.15		
					8901	13.10	9079B	15.05		
					8904	12.60				
					8906	13.20				
					8907	13.00				
					8922	12.50				
					8925	12.95				
					9031	13.05				
					9073	13.30				
					8978	12.65				

Hence, in the long run, we are primarily interested in determining with a fair degree of accuracy which varieties produce the best wine from a given region. It is evident from the analysis presented that dry wines from region IV and dry wines from region I should be compared separately as the mean scores do differ significantly. Further, taster × region-type interactions do exist. Since the distribution of the region-type scores is in general normal and fairly evenly distributed on either side of the median, it is suggested that the wines of each region-type be divided into five categories each year—poor, below average, average, above average, and superior. Wines whose mean scores fall into the range assigned 16 per cent or less of the cumulative

percentage scores will be called poor, 16 to 31 per cent called below average, 31 to 69 per cent called average, 69 to 84 per cent above average, and 84 to 100 per cent superior. This information is easily derived from the cumulative percentage plots. Confidence intervals between the levels can be established by the analysis of variance results. Advantages of this type of classification are: (1) that the tabulation of the results will be reasonably independent of tasters from year to year in respect to using different tasters, and (2) as long as the mean score for all wines of the region-type does not vary too greatly from the median cumulative score, the wines will be divided into three major middle groups and two minor end groups, which is a desirable grouping for the purpose. Only truly superior or inferior wines would appear in the end groups. Other methods of grouping will be investigated. Table 20 gives a tabulation of the mean wine score for RD I 1958 and the division thereof. It is seen that there are more mean scores of the wines at higher values than at the lower values (above and below median score). This unequal division is due to a higher mean score for the region-type as compared with median score. The confidence intervals are computed from the remainder error terms of the analyses of variance for this region-type for 1958 using all tasters' scores and figured at a 5 per cent level of significance. The confidence intervals are rather large and overlapping occurs. However, tests of this type will be run for a number of years and the final results of the pooled designations should be quite meaningful.

CONCLUSIONS

Several points may be concluded from the above data on the tastings made and reported here.

1. The over-all distributions of wine scores are fairly normal for most tasters in these particular experiments but a few show decided abnormalities.
2. Region-type scores are evenly distributed over most of the range with the largest discrepancy being shown by the RS IV scores.
3. Score card used at present tends to give, by some tasters, more lower scores than for other qualified tasters.
4. Some tasters have error variances not homogeneous with the rest of the tasters and preference patterns that are distinctly different from the main group of qualified tasters. The importance of preference patterns is discussed in a series of papers by Baker and others (1953, 1954, 1958, 1960) and Mrak *et al.* (1959).
5. Training periods needed for tasters to become relatively proficient are not too long. Evidence indicates that after the first tasting, tasters do not greatly increase in ability on a short-term basis.
6. The number of points the tasters tested can use effectively varies from about 4½ to 7 with an average of about 5½.
7. Wines must be compared in region-type groups.
8. The use of a taster's scores whose error variance and/or preference pattern are not homogeneous with the other tasters will usually induce large interaction terms and in general bias the tasting results.
9. It is better in investigating differences between regions, years, varieties and types of wine to use several of the "good" taster's results and exclude

the results of a taster with a large error variance, lower effective range, or different preference pattern. (General consumer acceptance would require an extensive investigation involving many tasters.)

10. These studies indicate that the procedure now in use is an effective method of wine quality evaluation as to variety-region-type.

SUMMARY

A relatively complete breakdown of the ability of tasters, distribution of scores by tasters, region-types, and total by years are presented, as well as a rather complete analysis of variance of the taste scores. Some of the significance of taster differences, region-type differences, tasting differences, and wine score differences is given. A tentative method for the final summary of scores is presented.

ACKNOWLEDGMENTS

Much credit and appreciation are due to Professor M. A. Amerine for the far-sighted planning which laid the background for this work. We are also indebted to the panel members for giving freely of their own time to make these tastings possible.

LITERATURE CITED

AMERINE, M. A., and A. J. WINKLER
1944. Composition and quality of musts and wines of California grapes. Hilgardia **15** (6):493–675.

AMERINE, M. A., E. B. ROESSLER, and F. FILIPELLO
1959. Modern sensory methods of evaluating wine. Hilgardia **28** (18):477–567.

BAKER, G. A.
1954. Organoleptic ratings and analytical data for wines analyzed into orthogonal factors. Food Research **19**:575–80.

BAKER, G. A., and M. A. AMERINE
1953. Organoleptic ratings of wines estimated from analytical data. Food Research **18**:381–89.

BAKER, G. A., M. A. AMERINE, and E. B. ROESSLER
1954. Errors of the second kind in organoleptic difference testing. Food Research **19**:206–10.

BAKER, G. A., M. A. AMERINE, E. B. ROESSLER, and F. FILIPELLO
1960. The nonspecificity of difference in taste testing for preference. Food Research **25**:810–16.

BAKER, G. A., VERA MRAK, and M. A. AMERINE
1958. Errors of the second kind in an acid threshold test. Food Research **23**:150–54.

BARTLETT, M. S.
1947. The use of transformations. Biometrics **3**:39.

DUNCAN, D. B.
1955. Multiple range and multiple F tests. Biometrics **11**:1–42.

MRAK, VERA, M. A. AMERINE, C. S. OUGH, and G. A. BAKER
1959. Odor difference test with application to consumer preferences. Food Research **24**:574–78.

SCHEFFÉ, HENRY
1959. The analysis of variance. John Wiley & Sons, New York, XVI + 477.

4*m*-5,'61(B6863)MH

The journal *Hilgardia* is published at irregular intervals, in volumes of about 600 pages. The number of issues per volume varies.

Subscriptions are not sold. The periodical is sent as published only to libraries, or to institutions in foreign countries having publications to offer in exchange.

You may obtain a single copy of any issue free, as long as the supply lasts; please request by volume and issue number from:

Agricultural Publications
Room 207 University Hall
2200 University Avenue
Berkeley 4, California

The limit to nonresidents of California is 10 separate issues on a single order. A list of the issues still available will be sent on request.