Mitteilungen Klosterneuburg 55 (2005): 107-113

Technical Report

Evaluation of wine tasters

Fedor Malík, Štefan Varga and Dušan Slugeň

Faculty of Chemical and Food Technology, Slovak Technical University
SK-812 37 Bratislava 1, Radlinského 9
E-mail: Fedor_Malik@stuba.sk

Results of every wine competition are official evaluations of wines (using for example median or average method) and individual evaluations of the wines of each member of jury. The principle question is, whether there is a good agreement between the values of official evaluations of the wines and the values of individual evaluations of each wine taster. This paper discusses the possibility, if one could classify each taster according to the degree of agreement between his individual results and the values of the official evaluations of the wines. From the mathematical point of view, the problem is to compare two qualitative random variables. This problem can be solved on four levels. First, using the chi-square test of fit it can be decided whether there is a good agreement or a statistically significant difference between values of these two variables. Second, using the Pearson contingency coefficient the strength of the agreement (dependence) can be measured. The bigger the dependence, the better the professional expert opinion of the taster. Third, using the test of symmetry in contingency tables, it can be verified, whether the values of the first variable are significantly bigger or significantly smaller than the values of the second variable. It means that we can decide whether the wine taster significantly underestimates or significantly overestimates particular wine samples. Fourth, using the X-Y chart the strength of the correspondence of official evaluations of the wines and the values of individual evaluations of some wine tasters can be visualized. As an example for this procedure the results of such an evalution of the wine tasters of the international wine competition Vinoforum 2004 are presented.
Keywords: Wine, competitions, taster evaluation, statistical methods

Information on data elaboration acquired by the evaluation of wines is very poor in technical literature, on the other hand, however, we have lots of data at our own disposal (Malík and Varga, 1996; Varga et al., 1992a). Moreover, we have knowledge of suitable mathematical-statistical methods enabling adequate methods of data elaboration and wine valuations. Therefore we suppose that the proposed method may characterize the professionalism and quality of members of a jury at a wine competition and also the quality of the work of the whole jury.

Material and methods

Data acquired at the Vinoforum 2004, one of the international wine competitions under the patronage of the O.I.V., were processed by mathematical-statistical methods. The basis for the evaluation were the results of an evaluation of seven commissions each consisting of five members. Cut average values taken from the competition catalogue are statistically confronted with individual values, which had been willingly provided by individual evaluators for scientific evaluation purposes.
The wines were evaluated by the 100 points system of the U.I.O.E. (Union Internationale des Oenologues). There were evaluated ten qualitative properties (resp. eleven for sparkling wine) of each wine sample (visual: limpidity, color; nose: intensity, genuineness, quality; taste: intensity, genuineness, quality, persistence; overall appreciation). 713 wine samples were evaluated by 35 wine tasters, that had been assorted into seven juries (A, B, C, D, E, F). Each jury had five members (X1 to X5). The highest and the lowest score in each jury were eliminated and from the remaining three values the arithmetic average was calculated. This is the so-called 'cut average', which is a coarse estimator of expectation of obtained points of evaluated wines.
In an effort to estimate the evaluator's activity within

the jury, we investigated the strength of an agreement of two qualitative random variables - the cut average values of the wine (variable *Y*) and the ratings given by each individual wine taster (variable *X*).
The strength of the agreement of the variables *X* and *Y* was examined at four steps:
1) the test of dependence (the chi-square test of fit),
2) the Pearson contingency coefficient,
3) the test of symmetry and
4) the *X*-*Y* chart.
All tasters were evaluated anonymously and only range values are presented.
Step 1) Using the test of dependence we can decide whether there is high dependence (good agreement) or there is a low dependence (bad agreement) between the cut average values (*Y*) and the individual values given by each wine taster (*X*) for the wines.

The hypothesis
H_0 : the variables *X*, *Y* are independent
is rejected (it means that the variables *X*, *Y* are dependent) if

$$T_1 = n \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{n_{ij}^2}{n_{i.}\, n_{.j}} - n \quad \in \quad \left(\chi^2_{(r-1)(s-1),1-\alpha}\ , \infty\right).$$

The values *n*, *r*, *s*, $n_{i.}$, $n_{.j}$ are from the contingency table of the cut average and taster's values and the value $\chi^2_{(r\text{-}1)(s\text{-}1),\ 1\text{-}\alpha}$ is the quantil of the chi-square distribution (MALIK and VARGA, 1996; VARGA et al., 1992a).

Step 2) The Pearson contingency coefficient

$$C = \sqrt{\frac{r\,T_1}{(r-1)(T_1+n)}}$$

(VARGA et al., 1992b; VARGA et al., 1993a; VENABLES and RIPLEY, 1994) expresses the strength of the dependence of the variables *X*, *Y* (the quality of the agreement of the variables *X*, *Y*). Values of the Pearson coefficient are from the interval (0, 1). The nearer to 1 the value, the bigger the strength of the agreement of the values *X*, *Y*, the better is the quality of the taster.

Step 3) The test of symmetry decides whether the wine taster generally underestimates or overestimates or symmetrically evaluates particular wine samples.
The hypothesis
H_0 : the variables X, Y are symmetric
is rejected (it means that the taster underestimates or overestimates particular wine samples) if

$$T_2 \ = \ \sum\sum \frac{(n_{ij}-n_{ji})^2}{n_{ij}+n_{ji}} \quad \in \quad \left(\chi^2_{r(r-1)/2,1-\alpha}\ , \infty\right)$$

(VARGA et al., 1993b; VENABLES and RIPLEY, 1994)

Step 4) Using the *X*-*Y* chart, we can see whether there is a good or bad agreement between the cut average values (*Y*) and the individual values of a taster (*X*). Each point in the chart represents one evaluated wine. The *x*-coordinate of the point is an individual evaluation of a taster and the *y*-coordinate is the cut average evaluation of the evaluated wine. The nearer to a line the points are in the chart, the bigger is the agreement between the variables *X*, *Y*.
In order to solve the above mentioned calculations we used the statistical software S-plus. Individual values of the wine tasters and cut average values of *n*-wine samples were entered into the program, which calculates the Pearson coefficient, the result of the test of dependence, the result of the test of symmetry and the *X*-*Y* chart.

Results

The quality evaluations of 35 wine tasters of seven commissions of the international wine competition Vinoforum 2004 are given in Table 1. The results of the five best and the five worst tasters are shown in the table. In the first column of each table, the denominations of the wine tasters are coded. In the second column the values of the Pearson contingency coefficient are given, in the third column the results of the test of dependence (good agreement) of the variables *X*, *Y* are introduced and in the fourth column, the results of the test of symmetry of the variables *X*, *Y* are mentioned. The last column qualitatively refers to the ranking of the taster. The tasters of each commission evaluated approximately *n* = 100 wine samples (white and red natural still grape wines as well as sparkling wines) during the two days.
The detailed evaluation of the two best and the two worst tasters are given in Tables 2, 3, 4 and 5. Besides the Pearson coefficient, the complete result of the test of dependence, the complete result of the test of symmetry, the contingency tables of all pairs of observations and the *X*-*Y* chart of the pairs are integrated. The first number in the pair is the individual taster evalua-

Table 1:
Five best and five worst tasters of the wine competition Vinoforum 2004 and 5. Vinofest Pezinok, 8th to 9th July, 2004

Taster	Pearson coefficient	Test of dependence ($\alpha = 0.05$)	Test of symmetry ($\alpha = 0.05$)	Rank
*G*5	$C = 0.838417$	$T_1 = 112.62 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 8.042 \notin (12.6, \infty)$ *X*, *Y* are symmetric	1.
*F*1	$C = 0.836233$	$T_1 = 118.01 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 11.364 \notin (12.6, \infty)$ *X*, *Y* are symmetric	2.
*G*2	$C = 0.831381$	$T_1 = 108.72 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 12.942 \in (12.6, \infty)$ *X* is smaller than *Y*	3.
*G*4	$C = 0.814398$	$T_1 = 99.968 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 10.511 \notin (12.6, \infty)$ *X*, *Y* are symmetric	4.
*C*5	$C = 0.814394$	$T_1 = 97.987 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 8.476 \notin (12.6, \infty)$ *X*, *Y* are symmetric	5.
...	...	...	...	...
*C*3	$C = 0.658275$	$T_1 = 47.665 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 10.200 \notin (12.6, \infty)$ *X*, *Y* are symmetric	31.
*A*2	$C = 0.657576$	$T_1 = 47.516 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 24.887 \in (12.6, \infty)$ *X* is bigger than *Y*	32.
*A*4	$C = 0.650806$	$T_1 = 46.089 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 8.470 \notin (12.6, \infty)$ *X*, *Y* are symmetric	33.
*E*5	$C = 0.639659$	$T_1 = 47.816 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 15.067 \in (12.6, \infty)$ *X* is smaller than *Y*	34.
*C*2	$C = 0.622441$	$T_1 = 40.550 \in (16.9, \infty)$ *X*, *Y* are dependent	$T_2 = 6.125 \notin (12.6, \infty)$ *X*, *Y* are symmetric	35.

Table 2:
Individual evaluation of the taster *G*5 in the commission *G* (ranking 1)
X - values of the taster
Y - values of the cut average

X ... *Y*	0-80	80-84	84-88	88-100	$n_{i.}$	Pearson contingency coefficient $C = 0.838417$
0-80	33	11	2	0	46	Test of independence
80-84	6	14	1	0	21	H_0: *X*, *Y* are independent
84-88	1	6	8	1	16	$T_1 = 112.624 \in (16.9 , \infty)$
88-100	0	0	5	13	18	Conclusion: *X*, *Y* are dependent
n_j	40	31	16	14	101	Test of symmetry H_0: *X*, *Y* are symmetric $T_2 = 8.042 \notin (12.6 , \infty)$ Conclusion: *X*, *Y* are symmetric

tion of the evaluated wine and the second number in the pair is the cut average value of the wine.
The value of the Pearson coefficient grades the quality of the individual opinion of the wine taster in relation to the overall results. The closer the coefficient is to the value 1, the more valuable in quality is the professional wine taster's opinion. Using the test of dependence and the test of symmetry with a significance level of $\alpha = 0.05$, the degree of agreement and symmetry of each taster's values and the the cut average values is determined with a probability of 0.95.

Table 3:
Individual evaluation of the taster *F*1 in the commission *F* (ranking 2)
X - values of the taster
Y - values of the cut average

X ... *Y*	0-80	80-84	84-88	88-100	$n_{i.}$	Pearson contingency coefficient $C = 0.836233$
0-80	16	10	4	0	30	Test of independence
80-84	1	17	6	0	24	H_0: *X*, *Y* are independent
84-88	0	6	18	0	24	$T_1 = 118.009 \in (16.9 , \infty)$
88-100	3	0	7	19	29	Conclusion: *X*, *Y* are dependent
						Test of symmetry
n_j	20	33	35	19	107	H_0: *X*, *Y* are symmetric
						$T_2 = 11.364 \notin (12.6 , \infty)$
						Conclusion: *X*, *Y* are symmetric

Table 4:
Individual evaluation of the taster *E*5 in the commission *E* (ranking 34)
X - values of the taster
Y - values of the cut average

X ... *Y*	0-80	80-84	84-88	88-100	$n_{i.}$	Pearson contingency coefficient $C = 0.639659$
0-80	29	10	10	1	50	Test of independence
80-84	5	13	8	1	27	H_0: *X*, *Y* are independent
84-88	0	7	10	2	19	$T_1 = 47.816 \in (16.9 , \infty)$
88-100	0	2	6	4	12	Conclusion: *X*, *Y* are dependent
						Test of symmetry
n_j	34	32	34	8	108	H_0: *X*, *Y* are symmetric
						$T_2 = 15.067 \in (12.6 , \infty)$
						Conclusion: *X* is smaller than *Y*

Table 5:
Individual evaluation of the taster *C*2 in the commission *C* (ranking 35)
X - values of the taster
Y - values of the cut average

X ... *Y*	0-80	80-84	84-88	88-100	$n_{i.}$	Pearson contingency coefficient $C = 0.622441$
0-80	23	14	3	1	41	Test of independence
80-84	15	11	5	0	31	H_0: *X*, *Y* are independent
84-88	3	6	7	0	16	$T_1 = 40.550 \in (16.9 , \infty)$
88-100	1	3	3	4	11	Conclusion: *X*, *Y* are dependent
						Test of symmetry
n_j	42	34	18	5	99	H_0: *X*, *Y* are symmetric
						$T_2 = 6.125 \notin (12.6 , \infty)$
						Conclusion: *X*, *Y* are symmetric

Discussion

All 35 wine tasters in the seven commissions evaluated the presented wine samples in good accordance with the official cut average values of the wines (probability > 0.95). On the other hand, only 16 tasters evaluated the wine samples symmetrically, which means that these members did not significantly underestimate and overestimate the particular wine samples. This conclusion cannot be applied to the other 19 tasters. The va-

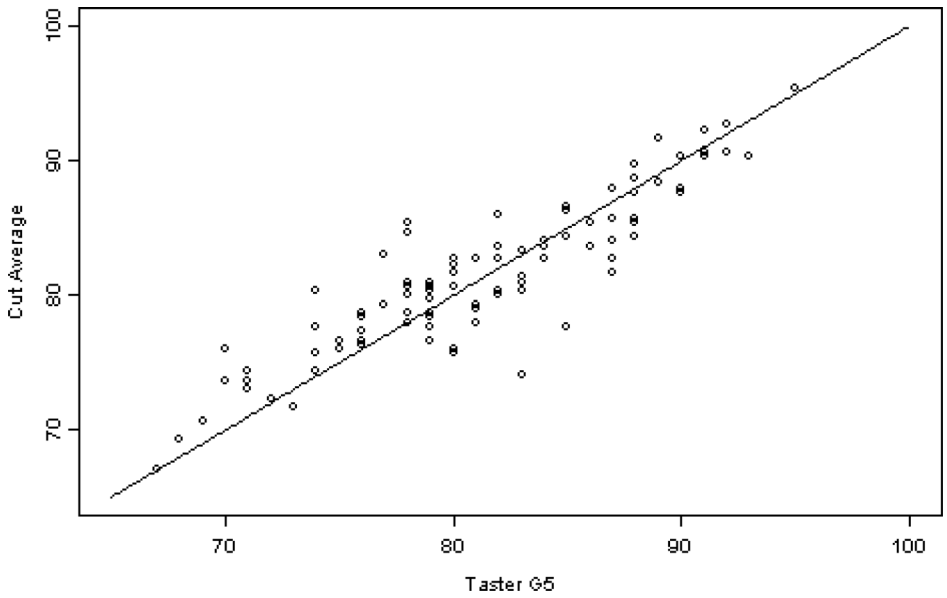

Fig. 1: Correlation between scores of taster G5 and cut average appreciation of wine samples in his jury

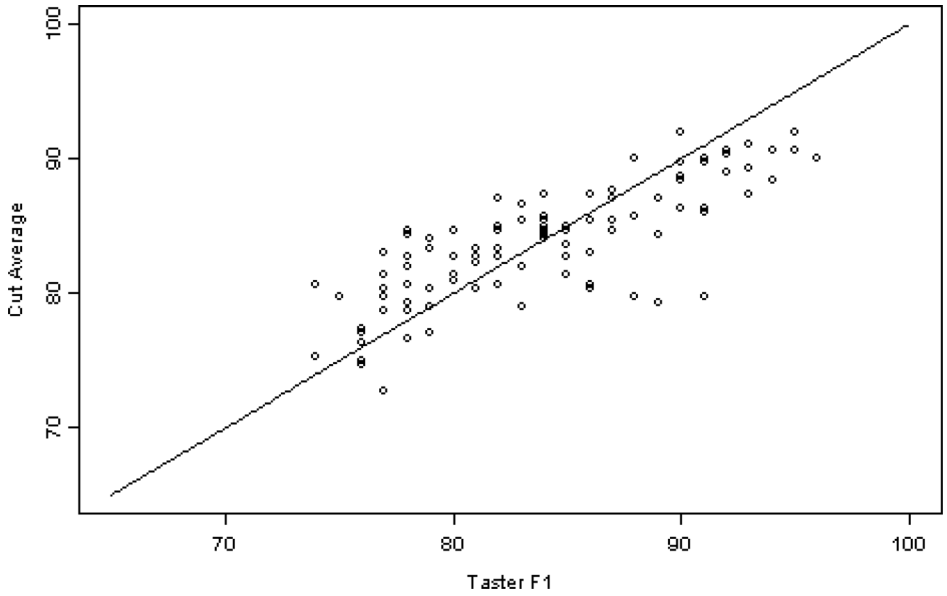

Fig. 2: Correlation between scores of taster F1 and cut average appreciation of wine samples in his jury

lues of the Pearson contingency coefficient are within the range of 0.622 (the second taster in the commission C (= C2)) to 0.838 (the fifth taster in the commission G (= G5)), which means that there were big differences in the quality of the tasters concerning the evaluation of the wine samples. This can be seen very well in the *X-Y* charts in the tables 2, 3, 4 and 5.

The presented and verified evaluation method of individual wine tasters of the international jury may be looked upon as one of the first attempts which might contribute to the improvement of the activities of such an executive body. It should be admitted that a sensory wine evaluation is influenced by many subjective factors. The choice and professional selection of wine ta-

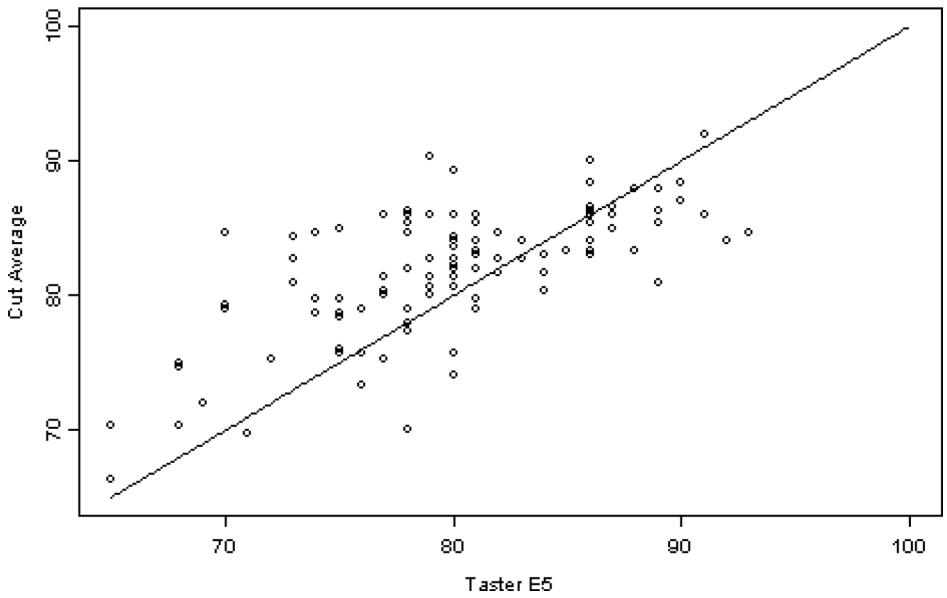

Fig. 4: Correlation between scores of taster C2 and cut average appreciation of wine samples in his jury

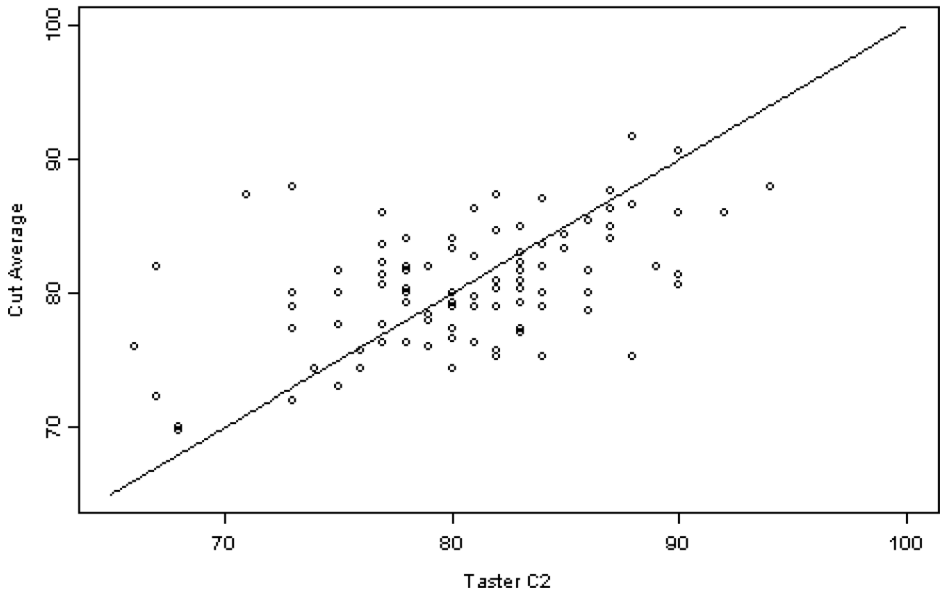

Fig. 3: Correlation between scores of taster E5 and cut average appreciation of wine samples in his jury

sters could contribute to objectify the evaluating activity of the international jury. We should not forget, however, that such activity and methods of evaluation which concern the honor and reputation of the taster, have to be done strictly anonymous.

References

MALÍK, F. and VARGA, Š. (1996): Analysis of sensorial methods for the evaluation of wine, 452 - 463. 11th Int. Oenol. Symp. - Sopron, Hungaria, 1996

VARGA, Š., MALÍK, F. and MINÁRIK, E. 1992a: Statistical treatment of data acquired by wine evaluation by the penalty system, 1st Part: Relation of median and individual values. Bull. O.I.V. 65: 36-48

VARGA, Š., MALÍK, F. and MINÁRIK, E. 1992b: Statistical treatment of data acquired by wine evaluation by the penalty system, 2nd Part: Evaluation of professionality growth of the wine tester. Bull. O.I.V. 65: 48-72

VARGA, Š., MALÍK, F. and MINÁRIK, E. 1993a: Statistical treatment of data acquired by wine evaluation by the penalty system, 3rd Part: Evaluation of specific dispositions of the wine tester. Bull. O.I.V. 66: 558-577

VARGA, Š., MALÍK, F. and MINÁRIK, E. 1993b: Statistical treatment of data acquired by wine evaluation by the penalty system, 4th Part: Evaluation of individual wine taster of the international jury. Bull. O.I.V. 66: 578-592

Venables, W.N. and Ripley, B.D. (1994): Modern applied statistics with S-plus. - New York: Springer, 1994

Received March 7, 2005