

A Method for the Statistical Selection of a Panel of Rum Tasters¹

Calixta S. Torres, Juan L. Aguiar and Eleanor F. Gotay²

ABSTRACT

Employees of an agricultural research unit were evaluated for selection as rum tasters. Prospects were classified and ranked considering their relative consistency in four organoleptic tests of four rum samples and their evaluation relative to those of an experienced rum taster. Statistical techniques used were variance analysis of Latin squares for the scores of the evaluation of 10 rum attributes and the calculation of a rum evaluation index for each taster using a linear discriminant function.

INTRODUCTION

Final decision on the relative quality of an alcoholic beverage has been based historically on organoleptic tests. As tasting sensitivity may vary, the possibility of evaluating the quality of beverages on the basis of chemical analysis and physical properties has been under consideration in recent years. Analytical researchers in the area of alcoholic beverages have tried to develop instrumental alternatives to the sensitivity of the tasters in evaluation of these products. Gas chromatographic analysis is the most promising tool at present for the separation of the components of alcoholic beverages. There is no objective method presently, however, for evaluating the quality of a rum based exclusively on its chemical composition or physical characteristics. Rum quality evaluation, still depends, therefore, on the judgement of tasters.

An objective procedure is thus needed to evaluate the constancy and precision of the tasting abilities of potential rum tasters.

The Rum Pilot Plant of the Agricultural Experiment Station found it necessary to select a panel of rum tasters to complement the chemical and physical evaluation of aged rums. The method adopted for the selection of the panelists described below was designed considering the difficulties in finding available tasters for routine work. This method was developed after reviewing reports of other workers on organoleptic analysis, particularly those concerned with strong alcoholic beverages.

¹ Manuscript submitted to Editorial Board January 24, 1974.

² Assistant Chemists, and Assistant Statistician, respectively, Agricultural Experiment Station, Mayagüez Campus, University of Puerto Rico, Rio Piedras, P.R.

The published works considered were those of Byer (1), Cabezudo (2), Kramer (4), Meilgaard et al. (5), Rodríguez Benítez (7), Savory and Saranin (8), and the Standard Methods of the American Society for Testing and Materials (9). In particular, Dr. A. Kramer, statistical consultant of this Station on quality appraisal, contributed his counsel to the development of the method.

MATERIALS AND METHODS

Employees of the Agricultural Experiment Station were considered for inclusion in the rum tasting panel. Candidates interested in qualifying for an organoleptic panel had to be in good health, over 18 years old, available at all reasonable times during working hours, and not prejudiced against alcoholic beverages. Candidates were not required to be regular rum drinkers or even consume alcoholic beverages regularly to be eligible for selection.

During a meeting with prospective panelists, the panel leader delivered a brief discussion on the importance of the studies on rum ageing, the tasting mechanisms involved in the panel's work, the importance of obtaining individual, careful and thorough opinion, and the importance of being available through the entire selection period, and if selected, to be available for future rum tasting sessions. Prospects were asked to complete a questionnaire which requested specific information about age, country of origin, occupation, drinking and smoking habits, and health status.

Demonstrations were conducted on the appraisal of the body and flavor of rum. The prospects were informed that the commercial Puerto Rican rum before them was the reference sample to be used in all future tasting sessions for the evaluation of the tasters. They were instructed to swirl the sample to appreciate the body of the rum, and to note the rate at which it flowed down the walls of the glass. At the same time, they were instructed to sniff the aroma intensity and to indicate their preference and to note off-aroma, if any. They then were instructed to place the entire sample in the mouth, rolling it over the tongue and swishing it inside their cheeks, and finally to swallow it while feeling its heaviness.

The scoring sheet (fig. 1) was discussed, and definitions of rum attributes and score numbers explained. Each of the tasters rated the reference sample; then they discussed the evaluation of the sample in detail, pooling the recollection, descriptions and judgements of the panel members (flavor profile method).

A tasting mobile unit was employed to conduct the tasting sessions. The panel-wagon with a movable bar was brought to the main Station

parking lot daily. Two bar tables with movable partitions to provide 10 individual booths were set in a quiet place with adequate sunlight under the shade of a tree. Tasting sessions were open daily from 11:00 A.M. to 12:00 noon for 2 months.

The comparison method was employed, in which the tasters were presented first with the reference sample, then with known rums, one at a time, to be compared to the reference sample. A set of 14 rum attributes and general preference items were considered and evaluated using the form presented in figure 1. Each sample was presented at a time interval of not less than 5 min followed by a short mouth wash.

Date and Hour		Sample Code Number				Taster's Name			Taster's Code Number							
Score	Color	Aroma		Off-odors	Body				Taste			Effect on mouth and throat tissues		General Preference	Comments	
		I	P		I	P	I	P	I	P	I	P				
+2																
+1																
0																
-1																
-2																

I/P
P = Preference

+2 = more intense +1 = slightly more intense 0 = equal to the reference -1 = slightly less intense -2 = much less intense	+2 = much more pleasant +1 = slightly more pleasant 0 = equal to the reference -1 = slightly less pleasant -2 = much less pleasant
---	--

FIG. 1.—Score sheet sample.

A light gold, commercial Puerto Rican rum, chosen among the most accepted and most consumed in the local market, was employed as a reference sample. Tasters were informed only that the reference sample was a representative commercial Puerto Rican rum. The reference rum and all other rum samples were given code numbers to mask their identities.

Three other gold commercial Puerto Rican rums were selected for the daily tasting evaluation of the prospective panelists: a light, a medium-light, and a heavy rum. These were selected as different from the reference sample but not obviously different in all attributes. When in doubt, brands of different distilleries were selected.

The taster received each sample in a brandy glass. The samples were half-diluted to avoid intoxication or desensitization of the taste buds. They consisted of 5 ml of rum and 5 ml of demineralized water prepared a week before the evaluation.

A Latin square design was employed to assign the order in which the

samples were to be presented to the tasters on different tasting sessions. Each panelist was to evaluate each sample four times in a different order following the reference sample. Evaluation of tasting ability of the tasters was to be determined partly by his ability to judge the samples in the same way, regardless of the order of presentation or session. When all panelists completed the four sheets, booth partitions were removed so that panelists could discuss the tasting session and the samples, mainly for the benefit of the taster's training.

Thirty-five prospective panelists, 18 men and 17 women, all Puerto Rican employees of the Agricultural Experiment Station, attended the tasting sessions regularly and finished all required tests for their evaluation.

Several panelists reported discomfort during the course of the tasting sessions regarding the outdoor setting for the evaluation of the rums because of variations in sunlight; sometimes too bright and very hot; other times, considerably darker.

As some tasters found it difficult to evaluate the acid, bitter, and sweet qualities of the samples, these attributes were eliminated from the final statistical analysis of the data.

Prospective panelists also claimed five rum samples per person was too much for a single tasting session. They suggested that judging sessions should include no more than four samples (the reference and three unknowns) to be considered by each taster.

The reference sample and the four unknown samples were also presented to an experienced rum taster for his evaluation following the same comparison method. These data were handled as a reference evaluation of the taster.

The hypothesis that the rum judging ability of a taster is higher: 1) the less his appraisal of the rum characteristics varies; 2) the better it coincides with (a) the appraisal of the expert taster and (b) the mean appraisal of all the tasters; and 3) the least it is affected by (a) the order in which a sample is presented or (b) the session in which the sample is tasted; suggests the possibility of estimating an index for the judging ability of the taster as a function of these qualities. The discriminant function technique offered such a possibility.

Several rum attributes and tasters were discarded because of incomplete data (missing scores). Of the original 35 tasters and 14 rum attributes, only 20 tasters and 10 rum attributes were considered for statistical analysis. An analysis of variance was made for each taster and rum attribute for a total of 210 Latin squares, to determine the F values of the samples, order of presentation of the sample for evaluation, and of the tasting session. Mean variance ratio values, F values, were thus determined for each taster. Tasters were ranked for three variables:

sample, sample order and tasting session on the basis of F values determined from the corresponding Latin square statistical analyses. Deviation from the scores of the expert taster and from the mean score of all tasters was also determined for each taster.

The procedure consisted of evaluation of the parameters of the linear equation

$$Y = M + AX_a + BX_b + CX_c + DX_d + EX_e$$

where,

Y = evaluation index of a taster

M = regression constant

X_a = rank of F value of the taster in judging the sample

X_b = rank of F value of the taster in disregarding the order in which he judges the sample

X_c = rank of F value of individual in disregarding the tasting session

X_d = deviation of the score of the taster from the mean of all tasters

X_e = deviation of the score of the taster from the score of the expert taster

A, B, C, D and E = regression parameters

The X values were based on measurement of the taster's judgement of the different rum attributes and on the evaluation of the expert taster of the same attributes.

The parameters of the linear equation were estimated following the procedure described by Moscoso and Capó (6). By using this procedure, the fitted equation maximizes the differences between the evaluation indices of the tasters since the equation is a discriminant function. Thus, the estimated values of the parameters (constants) are the relative weights of the various qualities considered for the estimation of the rum taster's ability.

The ability index of a new taster may be calculated for classification by use of this same linear equation. The X values would be determined from the data of organoleptic tests given him following the procedure previously described. The new taster's ability index, calculated in this way, may be statistically compared with the known indices of the tasting panel members.

RESULTS AND DISCUSSION

Table 1 presents the composition of the rum samples. Total acidity ranged from 7.5 to 21 mg/100 ml at 40° P. The tannin content, directly related to rum bitterness, fluctuated from 0.9 to 11 mg/100 ml at 40° P. Sugar content varied from 45 to 440 mg/100 ml at 40° P. These three attributes were not detected independently by the tasters. Many tasters

commented on these rum characteristics but did not score them on the sheets.

Table 2 presents the previously mentioned *F*-value ranks, or *X* values of each taster, the estimated equation parameters, the value of the regression constant and the estimated rum evaluation indices of the tasters.

The characteristics of the members of the tasting panel selected by use of this procedure appear in table 3. The panel included six men and four women.

TABLE 1.—*Composition of reference and rum samples*

Rum composition	Reference			
	R-619	R-630	R-611	R-614
Ethyl alcohol (° P)	79.7	79.4	79.4	84.8
Color (% transmission)	68.1	57.1	57.0	21.0
pH	5.0	4.4	4.2	4.5
	Mg/100 ml at 40° P			
Total sugars	45	97	258	440
Total acidity	7.5	8.0	9.0	21
Dry extract	92	120	271	555
Furfural	.01	.01	.01	.03
Tannins	.9	3.1	6.0	11
Acetaldehyde	2.1	1.1	1.3	5.5
Methyl acetate	.2	.2	.4	1.0
Ethyl acetate	4.4	4.5	6.8	5.0
Acetal	1.4	.8	1.1	3.3
Propyl alcohol	5.9	7.7	3.6	.6
Isobutyl alcohol	1.5	1.3	2.1	.2
n-Butyl alcohol	.1	.0	.0	.1
Isoamyl alcohol	5.4	1.6	8.4	.5
Total fusel oil	12.9	10.6	14.1	1.4
Total esters	4.6	4.7	7.2	6.0

Young persons predominated among the selected tasters, represented by one member under 25 years of age, five between the ages of 26 and 35, three between the ages of 36 and 45, and one older than 46 years. Employees between the ages of 26–35 years formed 31% of prospects but appeared in 50% of the composition of the final panel.

All but one of the members of the panel were scientists. Among the 35 original prospects, 54% were scientists; 23% were administrative and secretarial employees; and 23% were manual workers.

All selected panelists were rum drinkers occasionally, except for one who preferred beer. In the original group of prospective panelists, 83%

TABLE 2.—Factors (F-value ranks) used for the evaluation of the discriminant linear function for each taster: $Y = M + AX_a + BX_b + CX_c + DX_d + EX_e$

Taster	Maximum sample discrimination X_a	Minimum influence of the tasting session X_b	Minimum influence of the sample position X_c	Minimum difference from the means of all tasters X_d	Minimum difference from the expert taster X_e	Taster's ability index Y
1	10.50	11.050	13.60	0.2939	-0.0409	0.04377
2	7.20	11.035	10.90	.2939	-.8039	.03460
3	9.40	7.550	7.00	-.4689	.6102	.03004
4	12.65	3.900	10.30	.9452	-.0996	.02951
5	13.70	10.400	9.85	.0120	-.0179	.02657
6	14.75	10.650	6.35	.3169	.2306	.02446
7	15.25	9.850	11.00	.5655	-.7029	.01113
8	16.50	7.900	10.95	-.3680	-.2389	.00360
9	13.60	8.700	10.35	.0959	-.3972	-.00140
10	7.15	10.400	9.75	-.0623	-.5716	-.00425
11	9.65	13.800	12.55	-.2366	-.8002	-.00466
12	8.95	10.750	12.60	-.4653	-.2330	-.00469
13	5.35	14.050	11.95	.1019	-.2810	-.00568
14	7.20	16.500	14.55	.0539	-.8009	-.00767
15	12.85	10.300	12.00	-.4659	-.1156	-.00930
16	9.00	14.250	13.15	.2192	-.2600	-.01312
17	10.20	11.100	12.40	.1409	-.1939	-.01823
18	14.75	12.050	11.75	-.5150	-.8499	-.01823
19	14.05	17.050	10.90	.3072	-.0276	-.01955
20	8.10	10.200	11.20	-.7410	-1.0759	-.03581
	A	B	C	D	E	
Regression coefficient	0.000596	0.000422	0.000406	-0.057983	-0.004598	
Regression constant = M	= 0.17846					

TABLE 3.—Characteristics of members of the selected tasting panel

Taster number	Sex	Age range	Occupation	Drinking habits	Beverage preference
		<i>Years</i>			
1	M	26-35	Clerk	Occasional	Beer
2	M	26-35	Scientist	"	Rum
3	F	36-45	Scientist	"	"
4	F	26-35	Scientist	"	"
5	M	36-45	Scientist	"	"
6	F	20-25	Scientist	"	"
7	M	26-35	Scientist	"	"
8	F	26-35	Scientist	"	"
9	M	36-45	Scientist	"	"
10	M	46 or more	Scientist	"	"

were weekend drinkers. Occasional rum drinkers scored better than weekend drinkers. Among all tasters, 80% preferred rum; 17%, beer; and 3% preferred whiskey.

The evaluation of prospective panelists indicated that: rum drinking frequency is not a requisite for a good rum taster; scientists showed greater ability than non-scientists to evaluate rum; and, in general, younger persons had better rum judging abilities than older people.

RESUMEN

Treinticinco empleados de la Estación Experimental fueron considerados para formar parte de un panel de catadores de ron. Los candidatos probaron y evaluaron individualmente cuatro muestras de ron en cuanto a sus cualidades de color, cuerpo, aroma, sabor, efectos sobre los tejidos bucales y de la garganta y preferencia general. Las muestras les fueron presentadas en cuatro sesiones distintas, dispuestas siempre en diferente orden. Los catadores evaluaron las muestras en contraste con un ron de referencia.

Tomando en consideración la consistencia relativa en la evaluación de cada una de las muestras en contraste con las de un catador experto, los candidatos se clasificaron y catalogaron en relación a su sensibilidad para evaluar ron.

Las técnicas estadísticas utilizadas fueron el análisis de varianza de los resultados obtenidos en un diseño experimental de cuadrados latinos de las evaluaciones de los atributos de los rones y el cálculo de un índice discriminante para cada candidato, usando para este último cálculo una función lineal discriminante en la que se determinó el peso relativo a darse a cada criterio de evaluación del catador.

Los índices de evaluación fueron tabulados en orden descendente y los 10 catadores con los mejores índices se escogieron para formar el panel de catadores de ron.

Se propone utilizar la función lineal discriminante que se evaluó en este estudio para determinar los índices de catadores sometidos a las mismas pruebas organolépticas.

LITERATURE CITED

1. Byer, A. J., Looking askance at statistical sensory testing, *Food Tech.* 1717-22, Nov. 1964.
2. Cabezudo, M. D., Método de selección previa de un grupo de futuros catadores, *Revista de Agroquímica y Tecnología de Alimentos*, 11(3): 330-7, Instituto Nacional de Ciencias y Tecnología de los Alimentos, Valencia, España, 1971.
3. Fisher, R. A., The use of multiple measurements in taxonomic problems, *Ann. Eugen.* 7: 179-88, 1936.
4. Kramer, A., The relevance of correlating objective and subjective data, *Food Tech.* 23 (7): 66-8, 1969.
5. Meilgaard, M., Elizondo, A., and Mackinney, A., Beer flavor characterization, panel testing and gas chromatography for evaluation of thresholds and interactions of rum flavor compounds, *Wallerstein Lab. Comm.* 34(114): 95-109, 1971.
6. Moscoso, C., and Capó, B. G., Characteristic (discriminant function) indices of chironja, orange and grapefruit, *J. Agr. Univ. P. R.* 57 (1): 65-77, 1973.
7. Rodríguez-Benítez, V., Conducting rum tasting tests for quality appraisal, *Rev. Col. Quím.*, 2-5, March 1948.
8. Savory, S., and Saranin, A. P., The training of tasters for the organoleptic analysis of rum, *Food Tech. Australia*, pp. 338-41, July 1969.
9. Anonymous. Standard method of test for discrimination of differences in the production of alcoholic beverages, E-339-67:1, 46:228, *Amer. Soc. Testing and Materials*, 1974.