Measuring panel performance in thawed bonito (*Thunnus alalunga*). Statistical application with R

Miriam Uranga*, Irene Gartzia, Maruxa García-Quiroga

AZTI-Tecnalia, Unidad de Investigación Alimentaria, Parque Tecnológico de Bizkaia, Astondo Bidea, Ed 609, 48160, Derio (Bizkaia), Spain

* Contact author: muranga@azti.com / magarcia@azti.com

INTRODUCTION

White tuna (*Thunnus alalunga*) belongs to the family of the Escombridae and genus *Thunnus,* commonly identified as "Bonito del Norte" or "tuna". European countries that bulk export frozen tuna are Spain and France. Spain is the largest exporter of frozen albacore in Europe since 2009 providing 90 per cent of total exports (Murua, 2010).

The aim of this study was to evaluate the quality of the assessors of a panel trained in sensory analysis of thawed Bonito del Norte (bonito), both raw and cooked. The ability to detect differences is an essential characteristic to select an assessor. Other important aspects are repeatability, concordance among panellist and consonance of assessors in using attributes (Dijksterhuis, 1995). All these factors affect panel performance (Carbonell et al, 2007). Using the results in these tasting sessions, the efficacy of the panel was studied by examining the concordance between tasters' criteria and the consistency of the team, the use of the scale and the precision of their assessments. Two-way analysis of variance (ANOVA) is usually applied to evaluate panel performance. This assessment by itself can turn out to be weak or not solid enough if there is significant effect in any of the studied variables. Thus, the proposed work consists on completing the ANOVA analysis with other statistical studies.

MATERIAL AND METHODS

To carry out panel performance 1 month frozen samples of whole bonito were used throughout 3 different sessions. On each session 3 replicates of the same thawed sample were evaluated by the panel (consisting of 10 assessors,) and all samples were coded with three random digits. The tasting sheet used was composed of 13 descriptors, 5 to evaluate raw bonito and 8 to evaluate cooked bonito.

Then, statistical analysis was carried out with R 2.14.0 programme (R Development Core Team), using the packages Commander and Agricolae. The results obtained in the two-way ANOVA (assessor vs. session) with interaction were taken as starting point (table 1).

To study the concordance, a two-way analysis of variance (ANOVA) with interaction (assessor vs. session) and Pearson Correlation Test (correlation coefficient between the mean of assessors and the mean of the panel) were applied.

The consistency of the panel (homogeneity among assessors in the evaluation of the samples) was estimated using assessors´ individual input to the Interaction Sum of Squares (SSI) of the two-way ANOVA.

The precision of the panel in the evaluation of each attribute has been calculated from each assessor´s contribution to the Error Sum of Squares (SSE) of the two-way ANOVA.

Finally, the use of the scale was tested by measuring the difference between maximum and minimum values of each attribute.

In the case of consistency, precision and use of the scale, Simple Ranking Test (Friedman analysis) was applied and individual differences between rank sums on Simple Ranking Tests were determined by LSD Fisher Test. The contribution to the Interaction Sum of Squares (SSI) and the Error Sum of Squares (SSE) was not assessed quantitatively since small numerical values in much attributes could be concealed by big numerical values in just a few of them. Under those circumstances, it is convenient to convert each assessor´s numerical contribution into ranks in order to eliminate the scale of the effect (Fig. 1).

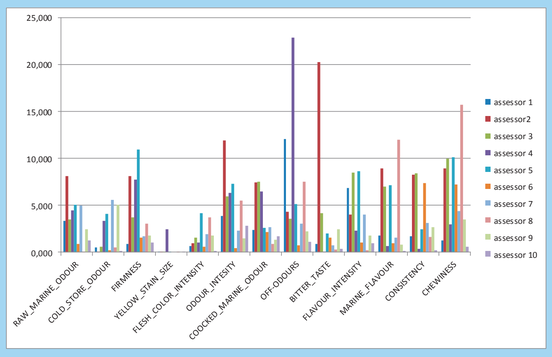

Figure 1. Individual contributions to the Sum of Squares have to be converted into ranks to eliminate the scale of the effect

RESULTS

The following results were obtained:

Concordance

ANOVA test showed that there were significant differences between the scores of the assessors (table 1). However, correlations between the scores of all assessors and the mean of the group were high or very high after performing Pearson´s correlation test (table 2).

Consistency

In the ANOVA test there was significant effect of the interaction assessor vs. session in several cases (table 1). Nevertheless, no significant differences between assessors were found after performing Friedman analysis and LSD Fisher Test (table 3).

Use of the scale

Three groups were determined applying Friedman analysis and LSD Fisher Test. "Group a" agglutinates assessors who use the whole scale to evaluate samples, "group b" assessors who use just a small portion of the scale and "group ab" assessors showing an intermediate use of the scale (table 3).

Precision

Finally, it was observed that groups that appeared in the evaluation of the use of the scale were exactly the same as observed after analysing precision of the panel. "Group a" put together less accurate assessors, "group b" assessors who make a more precise judgement and "group ab" assessors showing an intermediate precision in their evaluations (table 3).

Use of the scale and precision were mutually related, so assessors who made a restricted use of the scale were the most accurate ones whereas those who used the whole scale turned out to be the least precise.

	Assessor F	p-value	Session F	p-value	Interaction F	p-value
RAW_MARINE_ODOUR	3,923	0,001	22,574	0,000	3,117	0,001
COLD_STORE_ODOUR	7,562	0,000	16,648	0,000	2,618	0,005
FIRMNESS	6,688	0,000	22,913	0,000	4,207	0,000
YELLOW_STAIN_SIZE	22,635	0,000	3,039	0,056	6,403	0,000
FLESH_COLOR_INTENSITY	8,008	0,000	35,752	0,000	5,563	0,000
ODOUR_INTESITY	13,449	0,000	0,108	0,898	1,131	0,353
COOCKED_MARINE_ODOUR	7,870	0,000	0,266	0,768	1,155	0,334
OFF-ODOURS	8,758	0,000	1,625	0,206	3,402	0,000
BITTER_TASTE	5,498	0,000	0,163	0,850	1,472	0,149
FLAVOUR_INTENSITY	17,221	0,000	3,408	0,040	2,156	0,020
MARINE_FLAVOUR	22,202	0,000	4,886	0,011	5,103	0,000
CONSISTENCY	10,222	0,000	4,559	0,015	3,089	0,001
CHEWINESS	5,973	0,000	1,986	0,147	1,361	0,201

Table 1. Results of the 2 way ANOVA with interaction (assessor Vs session)

Assessor	F	p-value	correlation
1	6,003	0,000	0,875
2	3,800	0,003	0,753
3	5,273	0,000	0,846
4	9,776	0,000	0,947
5	3,569	0,004	0,733
6	6,725	0,000	0,897
7	9,444	0,000	0,944
8	8,584	0,000	0,933
9	4,393	0,001	0,798
10	9,057	0,000	0,939

Table 2. Concordance study results applying Pearson´s correlation test

ASSESSOR	CONSISTENCY		USE OF THE SCALE		PRECISION	
	Rank sum	Ranking	Rank sum	Ranking	Rank sum	Ranking
1	133 a	6º	76,5 ab	5º	64 ab	4º
2	137 a	8º	90,5 a	8º	100 a	9º
3	121 a	5º	81,5 a	7º	98 a	8º
4	113 a	4º	79 ab	6º	76 ab	7º
5	136 a	7º	98 a	9º	103 a	10º
6	160 a	10º	60 ab	3º	45 ab	2º
7	111 a	3º	62 ab	4º	73 ab	6º
8	106 a	2º	79 ab	6º	66 ab	5º
9	159 a	9º	53 ab	2º	59 ab	3º
10	89 a	1º	35,5 b	1º	32 b	1º

Table 4. Results of consistency, use of the scale and precision after performing Friedman analysis and LSD Fisher Test.

CONCLUSION

The results showed that, overall, the panel is highly qualified for sensory analysis of bonito. Nevertheless, those assessors who showed significant differences were retrained in order to improve panel performance.

It can be conclude that ANOVA test alone is insufficient to analyse panel performance and it is necessary to make a deeper statistical study. This methodology is a good way to measure panel performance and to identify assessors whose performance needs to be corrected.

REFERENCES

Carbonell L., Izquierdo L., Carbonell I. (2007). Sensory analysis of Spanish mandarin juices. Selection of attributes and panel performance. Food Quality and Preference, 18, 329-341.

Dijksterhuis G. (1995). Assessing panel consonance. Food Quality and Preference, 6 (1), 7–14.

Murua H. (2010). Situación de las principales poblaciones de túnidos tropicales. Industria Conservera. Mayo-Junio 2010, 14 -26.