Data management considerations for multivariate correlations of sensory and near infrared spectroscopic analyses of chilled chicken breasts

B.G. Lyon, W.R. Windham, F.E. Barton II and C.E. Lyon

USDA Agricultural Research Service, Richard B. Russell Agricultural Research Center, PO Box 5677, Athens, GA 30604-5677, USA.

Introduction

Developing models for predicting quality from spectral and sensory data requires careful pre-analysis of both data types. Whether to use all sensory attributes or only those that contribute the most to treatment differentiation is a major concern to the sensory analyst. Also, because quantitative sensory data are from more than one instrument (i.e. panelists) there is a need to compare performance of panelists to assess repeatability, reliability and/or agreement. If data are averaged over panelists, the influence of outliers and non-discriminators affects the sensory quantitative input used for modelling.[1]

Two approaches to sensory data analysis were considered in a study involving sensory, physical and spectral data: (i) a multivariate analysis of variance of all sensory data and (ii) a recently published program, GRAPES,[2] which presents, graphically, panelist and attribute performance on six criteria.

Materials and methods

The dataset used was from a study on chicken breasts that were chill-stored (5 temperatures $\times$ 3 holding periods $\times$ 4 replications) before cooking. The responses of ten trained panelists to 17 descriptive attributes were recorded on a computer-based test using 15-point line scales.[3,4] A multivariate analysis of variance (MANOVA from SAS-pc) was conducted over all panelists and by individual panelist to test panelist and attribute performance. The GRAPES program produced graphs of individual panelist and attribute statistics on scale use and data variation. Factor analysis was conducted on a full dataset of all panelists and all attributes (APA) and a modified dataset of selected panelists and/or selected attributes (SPSA).

Results

Table 1 summarizes panelist information from the MANOVA and the GRAPES analyses. MANOVA panel performance statistics of the ABC dataset (all products) and for the AB dataset (omitting storage set C) indicated significant differences ($P < 0.05$) among products for 10 of the attributes. Discriminating panelists did not always find sample differences according to the same attribute. Most significant F-values were for texture attributes.

The GRAPES program used all attributes and panelists with complete replication. P1 and P10 were omitted because of missing data. Values in Table 1 are the number of attributes for which a

From Near Infrared Spectroscopy: The Future Waves

Table 1. Summary of panelist performance statistics based on MANOVA[a] and GRAPES[b].

Panelist code	MANOVA[a]		GRAPESb	
	No. of significant attributes		Total no. of outlier values[c]	
	ABC	AB	Negative	Positive
P1	1	1	ID	ID
P2	4	4	–1	+12
P3	1	0	–2	+3
P4	8	5	–2	+7
P5	2	5	–1	0
P6	1	2	0	0
P7	8	9	–2	+5
P8	4	9	–3	+1
P9	6	5	–4	0
P10	2	2	ID	ID
ALL PAN	10	8	0	+20

[a]Multivariate analysis of variance. Values in these columns indicate the number of attributes for which treatment source of variation was significant ($P \leq 0.15$) for dataset ABC (all products) and dataset AB (omitting storage C).
[b]GRAPES procedure.
[c]Column values indicate the frequency that a panelist was a negative or positive outlier. ID = insuffcient data relative to x,y coordinates of plots of six performance variables.

panelist was considered an outlier. P2 and P4 were outliers most often by GRAPES but not MANOVA. P6 was not an outlier by GRAPES, but by MANOVA showed few differences in the samples.

GRAPES summary plots (not shown) from averaged data pointed out P2 and P7 tended to use the higher end of the scale. P2 had the highest drift. P4 had the highest mean standard deviation; P9 had the lowest standard deviation. P6 had the highest unreliability index, although data were plotted near the group mean. P6 was the least discriminating panelist. P4 contributed the most to the panelist × sample interaction.

Attributes, plotted by the same performance criteria (not shown), showed texture attributes were scored on the higher end of the scale, but also had higher mean standard deviations. Three texture terms were unreliable and had high drift while two flavor terms were unreliable but had low drift values. The same two flavor terms also showed less disagreement among the panelists and contributed the least to product discrimination.

Table 2. Factor analysis statistics of data with all panelists and all attributes (APA) and with selected panelists and selected attributes (SPSA).

	APA[a]			SPSA[b]		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Eigenvalue	16.59	12.56	2.60	24.14	12.56	2.26
Proportion	0.52	0.40	0.08	0.62	0.32	0.06
Cumulative	0.52	0.92	1.00	0.62	0.94	1.00
Attributes	Springi	Broth	Moisrl1	Broth	Springi	Moisrl1
	Cohes	Chick	Moisrl2	Chick	Cohes	Moisrl2
	Hard		Rbkdwn		Hard	Bwet
	Cohm		Bwet		Cohm	
	Chewi				Chewi	
	Fibrous				Fibrous	
	Easesw				Easesw	

[a]Non-contributing attributes in APA were bldy, cardbd, respart and met.
[b]Non-contributing attribute in SPSA was met.

Factor analysis with all panelists, all attributes (APA) and with selected panelists, selected attributes (SPSA), revealed that the two factor solutions switched flavor and texture terms (Table 2). Texture notes from APA comprised Factor 1 (52% variation). Brothy and chickeny notes comprised Factor 2 (40% variation). Omitting the presumably less discriminating panelists and attributes (SPSA) resulted in brothy and chickeny comprising Factor 1 (62% variation) and texture terms comprised Factor 2 (32% variation).

Discussion

Initially, in analyzing sensory data, it is unknown if panelists are consistent or discriminating or if all attributes are appropriate for measuring differences in the samples. Sometimes, a compensated dataset, using all panelists but eliminating non-discriminating attributes, is used. The eliminated attributes may have been determined from MANOVA F-values on the full dataset (all treatments, all panelists). This option screens the data first with all attributes and all panelists, but decisions made by the MANOVA F-value criteria could involve errors because of using "non-sensitive" panelists. Similar problems may be encountered by using all attributes, eliminating selected panelist data or using selected panelist data with selected attributes. Although GRAPES allows a more complete view of panel and attribute performance, the analyst still must use judgement in determining whether to use all data or to compensate for non-performing panelists or attributes.

The data in this study do not present extremes but represent a wide range of values for statistical modelling. However, these chill storage treatments of poultry represent real world extremes. Developing models to explain relationships presents difficulties when the differences are subtle and/or occur within a narrow range of intensity of detected attributes. Statistically valid models from instrumental measurements must still be related to sensory differences that are "noteworthy" relative to human perception.

References

1. D.S. Lundahl and M.R. McDaniel, *J. Sens. Stud.* **5,** 265 (1990).
2. P. Schlich, *J. Sens. Stud.* **9,** 157 (1994).
3. M.C. Meilgaard, G.V. Civille and B.T. Carr, "Selection and training of panel members", in *Sensory Evaluation Techniques*. CRC Press, Boca Raton, pp. 163–180 (1987).
4. C.J. Findlay, E.A. Gullett and D. Genner, *J. Sens. Stud.* **1,** 307 (1986).