STATISTICAL ANALYSIS OF SENSORY FRESHNESS DATA

Peter Howgate
May 2010

INTRODUCTION

The use of trained panels and scalar systems is quite common in the measurement of freshness of fish by sensory methods, but not common in published papers on the subject to include any statistical analysis of the errors involved in using a scalar procedure or any treatment of the biases of the assessors. The statistical analysis is quite straightforward and can be performed on computer spreadsheets. Routine analysis of panel data is important in the training and selection of assessors and in monitoring the performance of individual assessors and of the panel as a whole.

Another consideration for statistical analysis of sensory data is in sampling. A quality controller has to make a decision about the use of a batch of fish depending on its freshness. He/she takes a sample, determines the mean freshness score and compares this with the limiting value given in the specification for quality. In making the decision he/she needs to know the error associated with using the sample mean as an estimate of the batch mean. There are three sources of variance contributing to the error: the variation amongst individual fish within the batch; the error of the measurement; and the variation amongst assessors within the panel (assuming more than one assessor has been used). The first two are common to all methods of analysis, sensory and non-sensory; the last, variation between analysts/instruments, is usually not significant in non-sensory procedures, but important in sensory analysis. The variances can be estimated, by analysis of variance methods, from experimental data.

ANALYSIS OF VARIANCE FOR A PANEL SESSION

Consider the results of an actual trial illustrated in Table 2. A panel of 5 assessors has scored the freshnesses of 12 fish from a batch of fish stored in ice for the same time; each assessor individually and independently has scored each fish. The overall mean score of 5.4 is typical of cod at 12-14 days in ice and a change of 1 score unit is equivalent to about 3 days in ice. The means for the fish are not the same and this is to be expected in any sample drawn from a biological population. An assessor's score for a fish is not the same as the panel mean score for the fish and biases are defined as the score given by an assessor minus the mean score for all the assessors for that fish. An objective of the analysis is to determine if there truly is a difference between the assessors in their scorings or whether the differences between them is within the experimental error. The analysis of the data here follows that described in the Appendix of Shewan *et al*. 1953.

The analysis of variance table is shown in Table 3; it is obtained from the software in the spreadsheet. The mathematical model used is

$$X_{ij} = X_m + S_i + A_j + E_{ij}$$

where X_{ij} is the score given to the ith sample by the jth assessors, X_m is the mean value of all the scores, S_i is the effect of the ith sample, A_j the effect of the jth assessor and E_{ij} is the residual error associated with the method. The S's are distributed with a mean of zero and variance, V_s, which is an estimate of the fish-to-fish variance within the batch. The E's are distributed with mean of zero and variance, V_r which is the experimental error of the method as a whole. A number of factors contribute to V_r. One is the fact that the scale points are discrete. Though freshness is a continuum only unit points on the scale are defined, though assessors may give half points if they feel so inclined. The effect of rounding to 0.5 will contribute about 0.05 to the residual variance. Of more importance is the error of allocating the sensory perceptions as a score on the freshness continuum. For example, the odours do not smell exactly as they are described on the score sheet and there might be some doubt about the assigning a sample to a position on the scale.

The A's, the differences between assessors, are distributed with a mean of zero and variance V_a. It is this variance that should be minimised in an objective sensory method. In the example, the variance ratio, 8.6, for the assessor effect is very highly significant and shows that there are systematic differences, biases, among the assessors. The bottom line of Table 2 shows the biases, but these need to examined further to determine if they are significant or whether some assessors are more erratic then others.

ANALYSIS OF BIASES

To examine the biases form a table of deviations for each assessor and sample as the assessor's score for a sample minus the panel's mean score for the sample. This is Table 4. The variance of the deviations are calculated for each assessor using the functions in the spreadsheet. The standard error of the mean deviation, the bias, is calculated from the expression:

SE bias = Sqrt(Var bias/p $+V_r/p*q$), where p=no. samples, q=no. assessors.

(The formula includes a term for the error of the measurement obtained from the residual variance; see Appendix to Shewan *et al*., 1953). Dividing the value of a bias by its standard error gives Student's for the bias which can be tested for significance. It can seen that the significance of a bias is dependent on both the size of the bias and its variance.

In the example of Table 4 assessor no. 4 has the lowest bias, which is not statistically significant the $p<0.05$ level. This assessor also has the largest value of the variance and consequently the highest value of the SE. Assessor no 2 has the lowest value of the variance - not much different from the rounding error referred to above - so this assessor's scores are almost the same as the panel means allowing for rounding to 0.5 units. The biases for assessors 1, 2, & 5 are statistically significant, and that for assessor 3 is on the borderline. The largest absolute value of bias in this example, 0.41, is not large, equivalent to about 1.2 days of storage in ice in the case of cod A manager of trained and expert sensory panels would prefer that assessors do not show any biases as this might suggest inadequate training, extensive experience of monitoring the performance of sensory panels for freshness at TorryResearch Station showed that though the bias of any one assessor can often be shown to be statistically significant in any one panel session the bias is not maintained from one session to another and the average values of biases for an individual over many assessment sessions tends to zero.

Of more importance in my view is the size of the variance of the biases. A large variance shows that the assessor is erratic, sometimes scoring markedly above the panel mean and sometimes markedly above it. This might indicate some lack of ability to discriminate amongst freshnesses, or perhaps that the assessor is not concentrating, but whatever the reason a large variation is not a desirable feature in an assessor. It's not easy to decide when the size of the variation is too large, but if an assessor consistently shows a markedly larger variation than the rest of the assessors the manager might perhaps consider withdrawing that assessor from the panel. Statistical analysis of data in this way is useful in the training and selection of assessor's. Once the initial training and selection of assessors has been accomplished and they are drafted into existing panels that performance can be monitored and compared with the existing members. My experience is that when new members are first drafted in to existing panels their variances are large compared with the existing members, but with experience the variances will decrease in size until they approach those of existing members.

The sizes of the variances included in the model shown above can be estimated from the mean squares in the analysis of variance as shown in Table 3. The data used in Table 2 show a rather high value of V_a, (the data are from an actual session, but the example has been selected to illustrate various points). The average values of V_a and V_r obtained over a long period at Torry research Station using well trained and experienced assessors were 0.03 and 0.16 respectively. The value for V_s varies from batch to batch depending on the processing and storage of the fish and on the degree of mixing of qualities before sampling. For very well handled fish stowed carefully in ice and not disturbed until sampling it can be as low as 0.07, but under normal commercial conditions at fishing ports it is more likely to be around 0.25.

APPLICATION IN SAMPLING

Having an estimate of the variances makes it possible to compare alternative sampling plans. The quality controller has only a small number of trained assessors and a large batch of fish and need to make a judgement as to the freshness of the batch as a whole. A number of sampling plans can be devised but there are basically two strategies that can be adopted in this situation:

a) a sample of m fish is taken and each is scored by all of the n assessors,

b) each of the n assessors independently selects a sample of m fish and scores each fish in the sample.

In each case there are nm scores contributing to the sample mean value but in strategy (a) m fish are selected and nm fish in strategy (b). In (a) the mean value for each fish, being based on n observations, is known with a greater degree of precision than in strategy (b) but this is of no great consequence since it is the batch mean that is to be estimated.

The standard errors of the sample means are different for the two strategies. They can be calculated for the two cases from the expressions:

for strategy a), $\text{SE mean} = [V_r/nm + V_s/m + V_a/n]^{½}$

for strategy b), $\text{SE mean} = [V_r/nm + V_s/nm + V_a/n]^{½}$

Table 4 tabulates the standard errors for some combinations of n and m for the two strategies. The variances used are those appropriate for very well trained assessors sampling a rather mixed batch of fish. In this case V_r and V_a are low and V_s high. It can be seen that for this case strategy (b) is superior to (a) in that for the same number of assessments the former produces a mean with a smaller error. For the case of a less well trained panel where V_r and V_a could be double those used in the example, and sampling a more homogeneous batch the difference between the two strategies is much less. If the quality controller has only one assessor available then the two strategies are the same. As well as giving a lower standard error for the sample mean, strategy (b) is probably more convenient in practice than strategy (a). In (a), one of the assessors has to select a fish, score it, then pass it in turn to the other assessors for scoring. In (b) each assessor works through the batch independently, and they do not even have to be operating at the same time.

REFERENCE

Shewan, J.M., MacIntosh, R.G., Tucker, C.G. & Ehrenberg, A.S.C. (1953). The development of a numerical scoring system for the sensory assessment of the spoilage of wet white fish stored in ice. Journal of the Science of Food and Agriculture, 4, 283-298

Table 1

SCORING SYSTEM FOR FRESHNESS OF ICED COD BASED ON ODOUR OF GILLS

Score	Description
10	Fresh; sharp; seaweedy
9	Less sharp seaweedy odours; shellfish
8	Little odour; neutral
7	Musty; mousy; milky; garlic; caprylic; peppery
6	Bready; malty; beery; yeasty
5	Lactic acid; sour; stale oily; leathery
4	Lower fatty acids; stale grassy; cheesy; sweaty; sweet; fruity; chloroform-like
3	Stale cabbage water; turnipy; 'sour sink'; wet matches; ammoniacal; byre-like

Table 2

SENSORY PANEL FRESHNESS SCORES BASED ON ODOUR OF RAW COD

Sample	Assessor number					Sample mean
	1	2	3	4	5	
1	6.0	5.0	6.0	6.0	6.5	5.90
2	5.5	5.0	5.5	4.5	6.5	5.40
3	5.5	5.0	5.0	4.5	5.0	5.00
4	5.5	5.0	4.5	4.5	6.0	5.10
5	5.5	5.0	4.5	5.0	6.0	5.20
6	6.0	5.5	5.5	5.5	6.0	5.70
7	6.0	5.5	5.0	6.0	6.0	5.70
8	6.0	4.5	4.0	4.0	5.0	4.70
9	6.0	5.5	5.5	5.0	6.0	5.60
10	6.0	5.0	6.0	6.0	5.0	5.60
11	6.0	5.0	5.0	5.0	5.5	5.30
12	5.5	5.0	5.0	6.0	6.5	5.60
Assessor means	5.79	5.08	5.13	5.17	5.75	5.38
Assessor biases	0.41	-0.30	-0.26	-0.22	0.37	overall mean

Table 3

ANALYSIS OF VARIANCE

Effect	DF	Sum of squares	Mean square	F	Variance estimated
assessors	4	6.058	1.515	8.6	$V_r + 12V_a$, $V_a = 0.112$
samples	11	6.883	0.626	3.6	$V_r + 5V_s$, $V_s = 0.090$
residual	44	7.74	0.176		V_r

Table 4

ANALYSIS OF BIASES

Sample	Assessor number				
	1	2	3	4	5
1	0.1	-0.9	0.1	0.1	0.6
2	0.3	-0.2	0.3	-0.7	0.3
3	0.5	0.0	0.0	-0.5	0.0
4	0.4	-0.1	-0.6	-0.6	0.9
5	0.3	-0.2	-0.7	-0.2	0.8
6	0.3	-0.2	-0.2	-0.2	0.3
7	0.3	-0.2	-0.7	0.3	0.3
8	1.3	-0.2	-0.7	-0.7	0.3
9	0.4	-0.1	-0.1	-0.6	0.4
10	0.4	-0.6	0.4	0.4	-0.6
11	0.7	-0.3	-0.3	-0.3	0.2
12	-0.1	-0.6	-0.6	0.4	0.9
Assessor bias	0.41	-0.30	-0.26	-0.22	0.37
Variance of bias	0.117	0.069	0.163	0.180	0.175
S.E. of bias	1.289	0.760	1.789	1.977	1.927
t-test bias	3.62	-3.22	-2.01	-1.62	2.77
p	0.0007	0.0024	0.0504	0.1126	0.0082

Table 5

STANDARD ERRORS OF SAMPLE MEANS FOR DIFFERENT SAMPLING PLANS

Strategy a) m fish taken from the batch, and n assessors score each fish

$$SE\ mean = [V_r/nm + V_s/m + V_a/n]^{1/2}$$

	standard error of sample mean				
no of samples, m	number of assessors, n				
	1	2	3	4	5
2	0.48	0.42	0.4	0.39	0.38
4	0.36	0.31	0.29	0.28	0.28
8	0.29	0.24	0.22	0.21	0.2
16	0.24	0.19	0.17	0.16	0.15

Strategy b) m fish taken from the batch and scored independently by each of n assessors

$$SE\ mean = [V_r/nm + V_s/nm + V_a/n]^{1/2}$$

	standard error of sample mean				
no of samples, m	number of assessors, n				
	1	2	3	4	5
2	0.48	0.34	0.28	0.24	0.22
4	0.36	0.26	0.21	0.18	0.16
8	0.29	0.20	0.16	0.14	0.13
16	0.24	0.17	0.14	0.12	0.11

For both strategies, $V_r = 0.16$, $V_s = 0.25$, $V_a = 0.03$