SLAGTERIERNES FORSKNINGSINSTITUT
16. august 1957

Paper no. 17
SMAGSBEDØMMELSE
Rapport over forsøg 1-S-6

Fortroligt

Statistical examination of the performance of organoleptic panels

Time: August 1956 - July 1957

Responsible: Eva Blomquist

Introduction

The selection of new tasters in August 1956 and use of the tasters during different experiments during the winter 1956-57 introduced problems of control of the tasters' performance and planning of the taste testing. In the course of work with these problems some observations and preliminaries results have emerged. In this paper a brief discussion of the work will be given.

Literature

Hansen (1957) mentions that the analysis of variance is only a little sensitive to deviations from the mathematical assumptions underlying the analysis. One of the most essential assumptions is that the standard deviation of the results is independent of their numerical values. It is known that the standard deviation of the results of organoleptic scorings varies with the result of the performances. Howard (1956) computed the value of the standard deviation over the single scores of experimental material for butter, egg-powder, milk-powder and biscuits. In the present literature there is no discussion of the maximum variations of standard deviation for organoleptic scoring of bacon that can be tolerated before a really breach of the assumptions of the analysis of variance occurs. Harries (1956) uses, by statistical analysis of results of organoleptic scorings, for instance, by a two-factor experiment where the factors are indicated A and B, an analysis of variance of the following type:

Variation:
A
B
A x B Between tasters
A x tasters
B x tasters
A x B x tasters
Residual

The variation between the tasters is an expression of the fact that the individual tasters consistently are scoring on a different level of the scoring scale. One taster, for instance, is usually giving comparatively high scores, another is usually giving low scores and so on. The above statistical analysis makes it possible to isolate this variation in such a way that it will not influence the evaluation of the effects of the experimental factors. The two interactions A x tasters and B x tasters are an expression of the fact that the individual tasters, for instance, are estimating the effect of A differently. It seems to Harries safer to use A x tasters and B x tasters as experimental error in tests of significance for effects of A and B, instead of the residual or a combination of residual and A x B x tasters.

Howard (1956) made organoleptic experiments with samples of rolled roast. The tasters were trained by scoring a number of samples with varying flavour intensity. After the palatability scoring the placing of the tests on the scoring scale had been discussed and gradually they arrived at an agreement of the placing. A number of samples with varying flavour intensity were scored after that without subsequent discussion. The statistical analysis showed a significant difference between samples, between tasters and a significant interaction samples x tasters, all of it tested against the residual. Howard considers the significant interaction as an expression of variations in the interpretation of the steps of the scoring scale by the individual tasters. He elaborates that through a calculation of the regression between the score of the individual taster and the average score of the taste panel on the same sample. He found that the regression curves were reasonably linear and with a positive gradient. Unfortunately the numerical values of the regression and the correlation coefficients are not given. Howard's conclusion is, that he found it more appropriate to accept different but reasonably constant conception of the scoring scale than to force every member of the taste panel to use an established scale which might deviate from the individual conception of the member. Howard uses residual variation as experimental error. Hopkins (1946) uses the regression between the score of the individual taster and the average score of the taste panel on the same sample as a means of inspecting the performances of the individual tasters. The experimental materials are butter, milk powder and biscuits. The value of the correlation and regression coefficients lies between the following values:

Regression coefficient:	
Butter	0.72 - 1.16
Milk powder	0.69 - 1.10
Biscuits	0.68 - 1.14

Correlation coefficient:	
Butter	0.66 - 0.88
Milk powder	0.47 - 0.68
Biscuits	0.74 - 0.84

Besides, Howard calculates the average difference between the scorings of the individual tasters and the taste panel. A good taster is characterized by a high correlation coefficient and a high regression coefficient (a good sensitivity) and little average deviation from the scoring of the taste panel. Krum (1955) recommends, by examination of the scorings of the tasters, the calculation of the ratio between the variance of the double determinations and the variance within the double determinations. If a taster distinguishes much between the different samples and at the same time is giving rather identical scores to the double determinations the ratio F between the two variances becomes high.

The conduct of the experiment

The selection of tasters in August 1956 was performed on the basis of triangle tests of lard with different peroxide number. Since the tasters were called in from the outside, attempts were made to keep the trials as short and as easily understood as possible. In all, 73 persons were tried and 21 tasters were selected from them. The present taste panels consist of 16 persons.

...

Method for organoleptic tasting

The new tasters were divised into two taste panels of 8 persons each. The panel consisting of the Institute's trained tasters was termed TT. The work of the taste panels was organised so that all the samples from the different treatments within the experiments were tasted an equal number of times by every taster. The experiments, which have been the object of statistical computations, are listed in table 1. Significant differences are marked with +.

Training of new taste panels

The new tasters were subjected to a short training period where the results of the tastings were marked on a line ranging from very bad to very good. The results were discussed after the taste session with the Institute's trained tasters. The scoring scale from 0 to 10 for colour, saltiness, flavour and tenderness were described for the new tasters and samples accompanied by the Institute's mean scorings were given to the new tasters. Finally samples were given where the Institute's mean scorings were told to the new tasters after the taste session and the results were discussed.

The new tasters were not subjected to special experiments to examine their scorings but for most of the experiments the samples were presented in duplicate in the following way: two taste samples were made so identical as possible of the same experimental sample. A short series of different samples was at first presented to the tasters and after a short rest the duplicates were presented in random order to the samples first presented.

Results

Examinations of the taster's scorings

Samples from experiment no. 56-135 were tasted by the Institute's trained tasters. An analysis of variance of the results is listed in table 2, where x means variations exceeding the 5 % level of significance, xx exceeding the 1 % level of significance and xxx exceeding the 1 $^{o}/oo$ level of significance. All effects are tested against the residual.

The tasters apparently use different parts of the scoring scales, as there is a significant variation between tasters. The significant interactions between feeding and tasters and curing and tasters could be interpreted as a different sensibility to variations in feeding and curing of the different tasters or perhaps more generally as a different conception of the steps of the scoring scales by the different tasters. Six tasters of test panel I were used in experiment no. 56-68 together with two of the trained tasters. Variations between weeks of the experiment showed significance both to taste panel I and to the trained tasters. For taste panel I the regression and correlation coefficients of the individual scores against the mean scores of the taste panel were computed, as well as the average difference between the scorings of the individual tasters and the panel means. The coefficients were computed from 50 pairs of scorings. The experiment was one of the first where taste panel I was used. The following experiments listed in

...

table 1, nos. 56-99 to 56-116, were pig breeding experiments. As they showed no significant difference between the experimental treatments in the taste testing of flavours and only few significant differences for other testings, there was not computed any correlation or regression coefficients. The lack of significant differences between treatments is a characteristic feature of pig breeding experiments of this type, as it also was the case in the numerous former experiments of this type where the Institute's trained tasters were used.

In experiment no. 56-115, Sv. 751 and Sv. 754, and 56-117, Sv. 138, examinations were made to determine whether any of the tasters were able to discriminate between the treatments within the experiments. In experiment Sv. 751 one taster recorded a significant difference in saltiness and another taster recorded a significant difference in flavour. In experiment Sv. 754 none of the tasters recorded any differences and in experiment 56-117, Sv. 138 one taster recorded significant differences in saltiness and another taster recorded differences in flavour and tenderness.

Computation of the variance ratio between and within double samples has been used to some extent to evaluate the tasters sensitivity and consistency. It seems, however, necessary to use this method with some criticism. One of our tasters reacted extremely against one pair of samples, which gave him a comparatively high variance ratio. It is doubtful to conclude that he is a more qualified taster. Regression analysis as described below is probably a better method.

Discussion

The organoleptic tests discussed in this report are of the type, where analytical taste panels are used and where numerical scores are given to the samples. The results are treated statistically by the analysis of variance technique. A summary of literature studies and of the experimental results seems to outline the following ways in further studies concerning the use of analytical taste panels.

The analysis of variance is based on the mathematical assumption that the variance of the results is independent of the numerical value of the results. Literature reports that the variance of taste scorings augments when the quality of the samples is lowered. A similar examination of our experimental material and our tasters would be of interest. If there are real differences between samples, further training of the tasters on samples of inferior quality would probably diminish the standard deviation.

The variation between tasters is computed by the analysis of variance, and, as the variation between tasters in many of the experiments constitutes a considerable part of the total variation, the experiment plan mentioned above gives a more efficient experiment. The question concerning the use of either the interaction of treatments on tasters or the residual variation as experimental error could possibly be solved in the following way:

1. By a selection of the tasters with high correlation coefficients; the residual variation is used as experimental error and the experiments are planned with comparatively few tasters and comparatively many samples to each taster.

2. If it proves difficult to find tasters with high correlation coefficients,

...

the interaction is used as experimental error and the experiments are planned with comparatively many tasters and comparatively few samples to each taster to get the most efficient experiment.

In this connection the question of the size of the correlation coefficient arises. A discussion on this point has not yet been found in the literature.

Another question seems to be of interest: Is it possible to augment correlation and regression coefficients by training of the tasters, and in which form would it be appropriate to give such a training. It is not the purpose of the training to force the use of an established scale but to train the tasters to score consistently with respect to their own interpretation of the scale.

Conclusions

New tasters as well as trained ones use different levels of the scoring scale and the steps of the scoring scales are interpreted differently. The organoleptic testings of samples from an experiment are therefore planned so that every taster scores samples from every treatment within the experiment an equal number af times. It is thereby possible to isolate the variation between tasters and the interaction treatments tasters, from the variation between experimental treatments. Only further examination and training of the tasters will be able to decide if the interaction treatments tasters or the residual variation is to be used as experimental error. If the interaction is to be used, an effective experimental plan must be installed.

It would be of interest to determine whether the mathematical assumption, underlying the analysis of variance are fulfilled to a satisfaction degree in the case of our experimental material and our tasters, especially with regard to the independence of the variance on the size of the results of the experiment.

Summary

An organoleptic panel was selected by triangular tests and the new tasters were trained briefly. Statistical analysis of the results showed that new as well as trained tasters use different levels of the scoring scales and the steps of the scoring scales are interpreted differently. The experiment was planned so that it was possible to compute the variation between tasters and the interaction treatments tasters. Which of them should be used as experimental error is decided by the regression between the scores of the individual tasters and the average score of the taste panel. The correlation and regression coefficients have been computed for a single experiment. A larger material perhaps combined with continued training of the tasters is desirable.

...

SLAGTERIERNES FORSKNINGSINSTITUT
16. august 1957

SMAGSBEDØMMELSE
Rapport over forsøg 1-S-6

Table 1

Experiment no.	Tasters panel I, II, TT	Experimental material	Variation between treatments				Variation between tasters	Tasters x treatments
			colour	saltiness	flavour	tenderness		
56-135 22/6	TT	Bacon	-		+	+	+	+
56-135 29/6	TT	"	-		+	+	+	+
56-135 6/7	TT	"	-		+	+	+	+
56-135 13/7	TT	"	-		+	-	+	+
56-68	I	"		+	+	+	+	
56-68	TT	"	+		+	+		
56-99	II	Ham	+	+	-	+	+	
56-115 Sv751	I	"		-	-	-	+	-
56-115 Sv754	I	"	-	-	-	-	+	-
56-116	II	"	-	+	-	-	+	-
56-117 Sv744	II	"		-	-	-	+	-
56-117 Sv138	I	Bacon		+	-	-	+	+
56-117 Sv138	I	Cutlet			-	-	+	-
2-S-3	II	Bacon	-	+	-	-	+	-
3-S-2	TT	"	-		-	-		
3-S-2	I	"			-	-	+	-

SLAGTERIERNES FORSKNINGSINSTITUT
16. august 1957

SMAGSBEDØMMELSE
Rapport over forsøg 1-S-6

Table 2

Variation	22/6 4 tasters		29/6 4 tasters		6/7 4 tasters		13/7 3 tasters	
	flavour	tenderness	flavour	tenderness	flavour	tanderness	flavour	tenderness
between feed F	xxx	xxx	xxx	xxx	xxx	x	xxx	-
" cure C	xxx	xx	xx	xxx	-	-	xx	-
" taster T	-	xxx	xx	xxx	x	xx	xxx	-
F x C	-	xx	xxx	xxx	-	-	-	-
F x T	xxx	xxx	xxx	xxx	xxx	xxx	xxx	-
C x T	xx	x	xx	x	xx	-	-	-
F x C x T	-	xx	-	xxx	-	-	-	-
residual								

Table 3

taster	Correlation coefficient			Regression coefficient			Average deviation		
	Saltiness	Flavour	Tendernes	Saltiness	Flavour	Tenderness	Saltiness	Flavour	Tenderness
A	0.54	0.37	0.44	1.08	1.05	1.47	0.72	0.49	0.17
B	0.42	0.43	0.48	1.27	1.24	2.06	-1.23	0.74	0.11
C	0.60	0.33	0.28	1.65	0.86	0.78	-0.17	-0.22	-0.23
D	0.42	0.45	0.43	0.82	0.91	0.64	0.32	-0.14	0.06
E	0.44	0.45	0.36	0.79	1.72	1.11	0.14	-0.57	-0.59
F	0.45	0.55	0.27	0.60	1.09	0.53	-0.06	-0.29	0.50

...

Bibliography

Dawson, E & Harns, B. (1951). Sensory Methods for Measuring Differences in Food Quality. Agriculture Information Bulletin 34. U.S. Department of Agriculture. Washington D. C.

Hansen, E (1957). Statistisk Forsøgsplanlægning. Forelæsninger i Dansk Ingeniørforening 1956/57.

Harries, J. M. (1957). Personal Communication.

Howard, A. (1956). Food Preservation Quarterly 16, (2), 26-30

Hopkins, J. W. (1946). Can. Jour. Res. F., 24, 203.

Krum, J. K. (1955). Food Eng., 27, (7), 74.