SOME STATISTICAL METHODS IN TASTE TESTING AND QUALITY EVALUATION[a,b]

RALPH ALLAN BRADLEY

The Virginia Agricultural Experiment Station
Virginia Polytechnic Institute
Blacksburg, Virginia

1. INTRODUCTION

1.1 *Introductory Remarks.*

The title suggested for this paper is a general one and the discussion which follows is necessarily of a rather broad nature. There is a definite need for improved communication between the consumer of statistical methods, in this case the food technologist, the home economist and the horticulturist, on the one hand and the manufacturer of statistical methods, the mathematical statistician, on the other. The difficulty is accentuated by the general lack of mathematical training for the agricultural sciences and the sometimes aloofness of the statistician.

The mathematical statistician has tended to publish his research in concise mathematical style for an audience consisting principally of mathematical statisticians. Eventually the statistical methods, if they are good ones, become translated for use in applied problems. The time lag in some cases is considerable and should be decreased or eliminated. This can be accomplished in several ways. The ideal method would seem to involve the publication of an applied paper as a companion to the usual paper setting forth the theory of a new method. Lacking this, it is at least necessary that the research worker have some means of understanding of or reference to new statistical procedures. It is to this purpose that this paper may be of some small value.

---

[a]Prepared under a Research and Marketing Act contract, Project RM:c-629.1, with the Bureau of Agricultural Economics.

[b]This paper was presented to the Joint Symposium of The Biometric Society and the American Society of Horticultural Science at Cornell University, September 9, 1952.

1.2 *Bibliography.*

A fairly extensive classified bibliography is appended to this paper. No claim is made that it is exhaustive, which it obviously is not, nor that it affords a complete classification. The references given have been selected with a view to showing typical illustrative examples of statistical methods used in taste testing and procedures which are thought to be applicable in taste testing and quality evaluation.

Considerable assistance was derived from two available lists of references, [62] and [67]. The bibliography here is a condensation of abstracted references which were reported to the Bureau of Agricultural Economics [63, 64, 65, 66] and which received limited circulation. Many references are indicated which do not receive consideration in the body of this paper.

It is hoped that the bibliography provided may assist in bridging the gap between statistician and food technologist.

1.3 *Types of Taste Panels and Their Purposes.*

Taste panels may be classified in four types and exist for reasons which are primarily different.

(i) *Taste Panels for the Detection of Differences.*

These panels are usually small and it seems preferable to have from three to ten good judges to a larger untried panel. This sort of panel is one which is used for research purposes only. Rather intensive training of panel members is usually undertaken but in some cases it is not necessary that the members of a panel agree on their preferences or even on their judgments. It is necessary that a judge demonstrate ability to repeat his judgments.

For the development of new products, the improvement of old ones and the detection of insecticides and adulterants or effects of packaging or storage one is concerned only with the question of the existence of true differences. Unfortunately, presence or absence of differences is confounded with the presence or absence of taste acuity on the part of the judges.

(ii) *Taste Panels for Quality Control.*

Taste panels for quality control are usually panels of long standing and of more experience than the first type. Such panels are usually used for the maintenance of standards and as such are interested in the lack of differences or variability of some few specified products. Chart records may be kept both of the day-by-day performance of individual judges and of the panel comparisons as a whole. Again, these panels do not do preference testing in any way. Taste panels for quality control may be quite small but must be efficient.

(iii) *Taste Panels for Consumer Preference.*

In consumer preference testing panels are large and untrained. Usually no standards are provided and decisions are based on preferences alone. To be useful such panels must be representative of the consumer market of interest. Test procedures should be kept simple and the number of items compared should be small.

(iv) *Taste Panels for Quality Evaluation.*

Taste testing for quality evaluation is usually one phase of a more elaborate evaluation procedure. Composite quality scores consist of weighted averages of a variety of determinations. This kind of taste procedure is used in certain United States Standards for Grades. The tasting is usually done by a very small number of official graders. An attempt is made to conform to a uniform scoring system over long periods of time. Interest is in an absolute taste score and not in comparative scores for several products as is usually sufficient in the other types of panel testing.

Problems of considerable interest arise from the proper weighting of attribute measurements in quality evaluation.

Fundamental to any work with taste panels are problems in the selection of the panel and the choice of experimental designs and scoring techniques. We turn to a consideration of these problems.

II. THE SELECTION OF A TASTE PANEL

2.1 *General Discussion.*

Triangle tests, wherein a taster is required to pick the odd sample from a set of three samples, have been widely used for the selection of a taste panel. The procedure has the advantage of simplicity. In some cases each potential panel member is required to perform a predetermined number of these comparisons and the best judges of the group are selected for a panel. Concern has existed regarding the quality of the judges so obtained and the number of triangle tests which should be given. In general insufficient testing is done due to the time consumed and limitations on suitable experimental material.

The author has recently learned of attempts by food technologists to develop sequential testing and selection procedures and that without regard to the systems of sequential analysis developed by Wald [85] and Rao [84]. In applying a sequential analysis control is obtained over the quality of judges selected but simple applications do not necessarily lead to the selection of the best judges available. It is thought that sequential methods will provide considerable improvements over most selection procedures now used and constitute a saving of time and material.

Lombardi [83] in a thesis developed the applications of both the Wald and Rao methods of sequential analysis to the use of triangle tests in judge selection.

2.2 *Wald's Sequential Analysis Applied.*

Let p be the true proportion of correct decisions in triangle tests if the judge could continue testing indefinitely. This may be thought of as the judge's inherent ability under the test administered. Judges having abilities less than p_0 will be ruled unacceptable for a panel and those with abilities greater than p_1 will be selected.

A graph is drawn on which the number of observations m is plotted as abscissa and the number of correct decisions d_m is specified as the ordinate. Two lines L_0 and L_1, having a common slope, divide the graph into acceptance and rejection regions. The slope and intercepts of L_0 and L_1 are given by Wald [85] (c.f. section 5.3.3) and they depend on the specification of p_0 and p_1 and parameters α and β. α is defined as the probability of selecting an unacceptable judge and β is the probability of rejecting an acceptable judge. p_0, p_1, α, β are at the disposal of the experimenter. If potential judges are in good supply, he may take α small and β large.

The common slope of L_0 and L_1 is given by

$$(1) \qquad s = \frac{\log \dfrac{1 - p_0}{1 - p_1}}{\log \dfrac{p_1}{p_0} - \log \dfrac{1 - p_1}{1 - p_0}}$$

and intercepts h_0 and h_1 respectively are

$$(2) \qquad h_0 = \frac{\log \dfrac{\beta}{1 - \alpha}}{\log \dfrac{p_1}{p_0} - \log \dfrac{1 - p_1}{1 - p_0}}$$

and

$$(3) \qquad h_1 = \frac{\log \dfrac{1 - \beta}{\alpha}}{\log \dfrac{p_1}{p_0} - \log \dfrac{1 - p_1}{1 - p_0}}$$

Before a definite decision is reached on the specification of p_0, p_1, α, β it is useful to compute the average number of tests which will be required. This depends on the ability of the judge p and the test specification. A rough plot or table can be computed by considering special values of p as shown by Wald (c.f. section 5.5). $E_p(n)$ is the average sample number for a judge of ability p. Then,

(4) for $p = 0$ (no ability)

$$E_0(n) = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{1-p_1}{1-p_0}},$$

(5) for $p = p_0$ (maximum unacceptable ability)

$$E_{p_0}(n) = \frac{(1-\alpha)\log \frac{\beta}{1-\alpha} + \alpha \log \frac{1-\beta}{\alpha}}{p_0 \log \frac{p_1}{p_0} + (1-p_0)\log \frac{1-p_1}{1-p_0}},$$

(6) for $p = p_1$ (minimum acceptable ability)

$$E_{p_1}(n) = \frac{\beta \log \frac{\beta}{1-\alpha} + (1-\beta)\log \frac{1-\beta}{\alpha}}{p_1 \log \frac{p_1}{p_0} + (1-p_1)\log \frac{1-p_1}{1-p_0}},$$

and

(7) for $p = 1$ (infallible ability)

$$E_1(n) = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_0}}.$$

One further average sample number may sometimes be computed and this case occurs when $p = s$, the slope of the control lines. Then,

$$E_s(n) = -\frac{\left(\log \frac{\beta}{1-\alpha}\right)\left(\log \frac{1-\beta}{\alpha}\right)}{\log \frac{p_1}{p_0} \log \frac{1-p_0}{1-p_1}}. \tag{8}$$

An example has been constructed in which judges with inherent abilities .33 and .60 have been simulated. The test specification used the values $p_0 = .40$, $p_1 = .65$, $\alpha = .05$ and $\beta = .05$. These values are not necessarily suitable for all judge selection and in general p_0 and p_1

are too small while β may be sometimes profitably increased. Average sample numbers are shown in Table I.

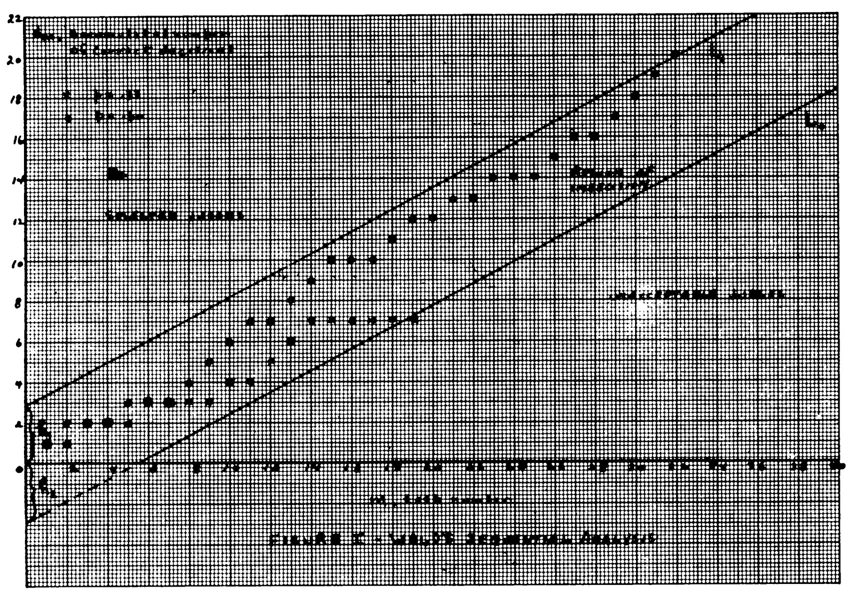
TABLE I
AVERAGE SAMPLE NUMBERS FOR THE TRIANGLE TEST

p	$E_p(n)$
0	5
0.40	21
0.65	21
1	6

Substitution in (1), (2) and (3) yields $s = .53$, $h_0 = -2.87$ and $h_1 = 2.87$. The equations of L_0 and L_1 become

(9) $$L_0 : \; d_m = .53m - 2.87 \qquad \text{and}$$
$$L_1 : \; d_m = .53m + 2.87.$$

The diagramatic procedure is shown in Figure I for two judges of



abilities .33 and .60. The analysis rejected the first of these judges but accepted the second. In so doing a judge has been obtained with ability somewhat below that desired but one who was better than those which were ruled to be unacceptable.

2.3 *Rao's Sequential Analysis.*

Rao [84] in 1950 published a method of sequentially testing a null hypothesis and this procedure may also be applied to the selection of judges for a taste panel. The theory of the procedure needs further investigation since its properties are not well known. However, when it is applied to sequences of triangle tests, apparently satisfactory results are obtained.

Lombardi in conjunction with the present author has adapted the Rao method to use with the binomial distribution. The Rao procedure differs from that of Wald in that a limit to the testing of any one potential judge may be set. We define N to be the maximum number of tests to be given to any judge. Only one limit of ability is set. Individuals having ability greater than p_1 will be accepted and those with ability less than p_1 will be rejected. α is the probability of selecting an unacceptable judge.

One central line L is used with slope p_1 and intercept

$$h = \frac{Np_1}{\alpha}\,[I_{p_1}(n_0 - 1, N - n_0 + 1) - \alpha] \tag{10}$$

where $I_{p_1}(n_0 - 1, N - n_0 + 1)$ is the incomplete beta function. n_0 is the minimum number of successes in N trials which rejects p_1 as a judge's ability in favor of the alternative $p > p_1$ at the significance level α. Approximately,

$$n_0 = p_1N + \sqrt{2.72Np_1(1 - p_1)} \qquad \text{if} \qquad \alpha = .05$$

and

$$n_0 = p_1N + \sqrt{5.43Np_1(1 - p_1)} \qquad \text{if} \qquad \alpha = .01.$$

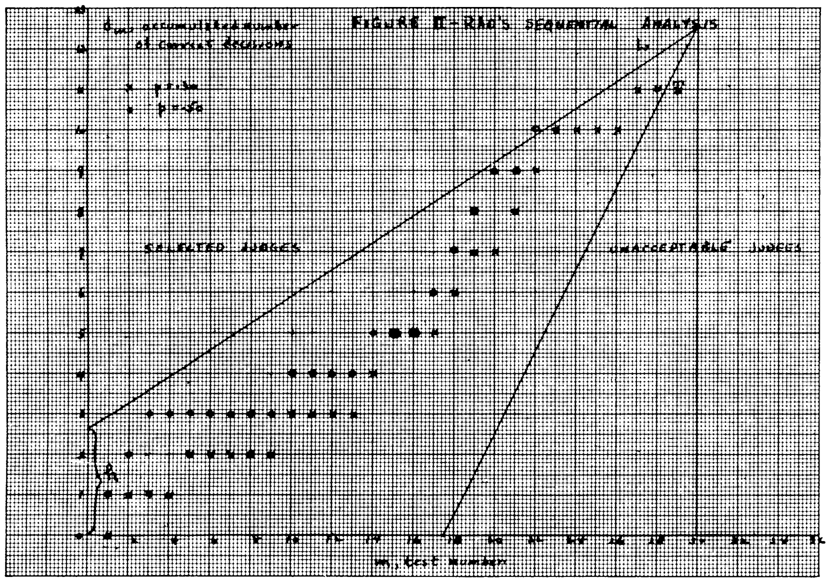
Tables of the incomplete beta function are available [70].

Let us consider an example wherein $p_1 = .33$, $N = 30$, $\alpha = .05$. Then $n_0 = 15$ and $h = 2.62$. The equation of L becomes

$$L: \; d_m = .33m + 2.62. \tag{11}$$

where d_m is the accumulated number of correct decisions and m is the number of tests.

Two judges with abilities .3 and .5 have been simulated using tables of random numbers. The procedure is illustrated in Figure II. An additional line T named the truncation line has been drawn. T is such that if the sequence of trials leads to a point below T it will be impossible to accept the judge within the specified maximum number



of tests. One should then, if this happens, stop testing at such a point. T is drawn with slope 1 through the intersection of L and the vertical line $m = N$.

Both methods of sequential selection of judges presented have features which are desirable. The computation for Rao's procedure is simpler than for that of Wald if tables of the incomplete beta function are available. In the Rao method in practice one would usually choose p_1 considerably larger than in the illustrative example.

III. THE DESIGN OF TASTE TESTING EXPERIMENTS

3.1 *Problems of Scaling and Scoring.*

In most experimentation the units of measurement are clearly defined and follow naturally from the formulation of the problem. In a majority of taste testing problems it is impossible to avoid using scores or orderings based on the purely subjective opinions of judges or panels of judges. The statistical problems peculiar to this field of experimentation arise out of the limitations and uncertainties of human behavior. Some of the difficulties introduced are non-uniformity of scoring procedures, lack of consistency on the part of an individual judge, lack of agreement between a judge and what may be regarded as a "true criterion", doubt as to the appropriateness of analyses of variance on scores or orderings, and limitations of numbers of observations due to taste fatigue.

Scoring scales in taste testing are usually set up arbitrarily. A range of values symmetrically placed about an origin are used to measure variation from poor to excellent as related to some food attribute. The choice of scoring scale has been discussed by Hopkins [78]. When a scale has been specified, there is no guarantee that all members of a panel will utilize it in the same way. Panel members may spread their scores over the complete range or may use only a segment of the scale. Too, they may record scores falling in similar size ranges but differ in the location of their average scores. To offset these difficulties standards are sometimes used although all too often only one is specified. The standards in effect have been given a prespecified score and test items are scored in comparison with the standards. The following quotation by Harrison and Elder [82] is pertinent.

> "A numerical scale for scoring often proves to be a 'rubber yardstick' unless the precaution is taken to include more than one fixed standard at predetermined points on the scale for the orientation of the taster. Repeatedly we have observed that a wide average score difference, observed between two samples when presented by themselves, will shrink when a third higher quality sample is presented along with them. This phenomenon is familiar to the psychologist who recognizes it as 'adaption'."

Alternate to using discrete scores, Baten [72, 73] has used an interesting continuous line scoring system ranging from very poor to very good. He indicated in his earlier paper that increased accuracy was obtained by the method but that it proved unpopular with the judges. Standards would again be required to obtain uniformity of scoring.

Most scoring systems lead to some doubt as to the validity of analysis of variance techniques and it is not clear how serious this may be. The basic assumptions of analysis of variance [7] are

(i) Observations are independently distributed,
(ii) Observations come from a normal population,
(iii) Error variances are homogeneous,
(iv) Treatment and environmental effects are additive.

Taste fatigue may introduce departures from (i) and (iii). A discrete scoring method without the use of standards usually leads to departures from (iii) and, with the effect of 'adaption', from (iv). The discreteness of a scale always leads to violation of (ii) although this may not be too serious.

Taste fatigue and the necessity of including at least two standards in a tasting experiment require either a reduction in the number of treatments to be compared or the use of incomplete block designs.

Difficulties in the establishment of suitable scales suggest more extensive use of ranking methods than has been made to date.

3.2 *Experimental Designs.*

Designs which increasingly control sources of variation and permit flexibility in experimental procedures are the randomized block, the latin square, the split-plot, the factorials, and the more complex incomplete block designs and lattices. It is a good rule, applicable in taste testing as well as elsewhere, to use the simplest possible design that meets the need of the experiment. However, the incomplete block designs are particularly appropriate to taste testing. The number of samples tasted in one sitting is usually limited sharply and the use of small blocks of treatments is of great assistance. It should also be noted that this eliminates the need for the judge to have longer term memory for he need only be consistent on his level of judgment within the incomplete block unit. An excellent reference to experimental designs by Cochran and Cox [19] is available.

Latinized Rectangular Lattices developed by Harshbarger and Davis [25] are of some particular interest. In most incomplete block designs the analysis separates out the effects of replications, treatments and blocks and interaction effects are not available. The latinized rectangular lattice introduces something of the advantage of the latin square. The available effects are sets, rows, interaction of rows by sets, blocks, and treatments. Then it is possible to associate judges with sets and days with rows and obtain some indication of judge behaviour from the interaction effect. This seems to constitute an advantage over other incomplete block designs. The design is available for treatment numbers which are products of two consecutive integers, $k(k - 1)$.

It is believed that the analysis of variance may be used without serious error in some taste testing experiments but that in cases of doubt ranking methods should be substituted when they are available.

3.3 *Ranking techniques*

The usual criticism of ranking methods stems from a supposed loss in efficiency. When quantitative judgments can be obtained, the magnitude of differences is obscured by the use of ranks. On the other hand, when treatment differences are very small and difficult to detect, it would appear reasonable to simplify the procedure for the judge and use a ranking technique. In many cases of the latter sort the use of a scoring scale or the continuous line may give the appearance of a precision of judgment which does not in fact exist. Again, the application of rank order methods is usually computationally simple and they may be often preferred on this ground alone. An interesting discussion on this subject was recently presented by White [58].

The modern development of rank tests has been largely limited to tests of two treatments. Such tests have been considered by Mann and Whitney [53], Wilcoxon [59, 60], Wald and Wolfowitz [56] and others. These tests could often be conveniently used to replace the t-test in sensory testing. Terry [55] has developed a test of this type which has the feature of being the most powerful rank order test in situations where a t-test would have been appropriate if quantitative measures could have been taken. His test depends on order statistics as obtained by transforming ranks using Table XX, *Scores for Ordinal (or Ranked) Data* of Fisher and Yates [46].

When ranks are employed in the analysis of variance, it has become common in taste-testing problems to transform these ranks using Table XX. Bliss [10] has supported the procedure and other references may be found in the bibliography. That the procedure is most powerful in that case considered by Terry may suggest confidence in its more general use.

Rank order methods analogous to the analysis of variance have been considered. The methodology is not complete and usually involves approximations since the computation of exact probability tables is often an exhaustive process. Kendall [52], Friedman [47] and Kendall and Babington Smith [51] have considered tests of agreement or concordance with ranked data and these methods are applicable to taste testing. Mood [54] (c.f. Chap. 16) has also contributed useful rank order methods developed from a slightly different viewpoint.

3.4 *Paired Comparisons.*

The method of paired comparisons may be regarded as a special rank order technique. It is a method long used in psychological experimentation and one that is well adapted to sensory difference testing. Only two treatments need be considered at one time and qualitative decisions alone are required. The design becomes somewhat cumbersome if many items are compared but hidden replication offsets some of that difficulty.

Two somewhat comparable methods of analysis have been presented by Thurstone [38] and by Terry with the present author [28, 37]. The mathematical formulations of the models are apparently different but may be related. Mosteller [33] has summarized Thurstone's model and listed the following underlying principles [c].

[c] Actually Mosteller lists six principles. The remaining two relate to the method of experimentation and the purpose of the analysis.

(1) There is a set of stimuli which can be located on a subjective continuum or sensation scale.

(2) Each stimulus, when presented to an individual, gives rise to a sensation in the individual.

(3) The distribution of sensations from a particular stimulus for a population of individuals is normal.

(4) It is possible for paired sensations to be correlated. The model may in a sense be summarized by writing

$$\pi_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-(S_i - S_j)}^{\infty} e^{-\frac{1}{2}y^2}\, dy \tag{12}$$

where S_i and S_j are the "true" treatment locations on the sensation continuum and π_{ij} is the probability that treatment i be rated above treatment j.

In the second method of analysis the mathematical model is formulated as follows:

(1) t treatments in an experiment using paired comparisons have true ratings $\pi_1, \cdots, \pi_t$ ($\pi_i \geq 0$).

(2) Observations on pairs of treatments are independent in probability.

(3) When treatment i is compared with treatment j, the probability π_{ij} that treatment i be rated above treatment j is $\pi_i/(\pi_i + \pi_j)$. (This further specifies the nature of the true ratings).

If we redefine

$$\pi_{ij} = \frac{1}{4} \int_{-(\log_e \pi_i - \log_e \pi_j)}^{\infty} \operatorname{sech}^2 \frac{y}{2}\, dy, \tag{13}$$

it is easy to show that then $\pi_{ij} = \pi_i/(\pi_i + \pi_j)$. Thus the substitution of the 'squared hyperbolic secant' density for the normal density of Thurstone's model yields the second method of analysis[d]. The squared hyperbolic secant density is very similar to the normal. The substitution in Thurstone's model is a sufficient condition for the application of the model developed by Terry and the author. It would not appear to be necessary. Values $\log_e \pi_i$ correspond to values S_i on a subjective continuum. Methods developed for the estimation of these sets of parameters differ.

In the second procedure estimates p_i of π_i are obtained by the method of maximum likelihood. When some one treatment is always rated above all others, that treatment obtains a rating $p = 1$ while the others have relative ratings zero. It then appears that one good

[d]The author is indebted to Professor R. A. Fisher for a suggestion on this point.

treatment obscures the differences among the others. This is in accordance with the idea of "adaption". Actually secondary ratings may be obtained for the remaining treatments by considering the subexperiment consisting of those $(t - 1)$ treatments. These secondary ratings were actually used in the evaluation of test statistics to distinguish between experimental results which had the one 'perfect' treatment but differed otherwise. This is a point which has not heretofore been exhibited.

The second test formulation has considerable flexibility and for the detection of treatment differences the results of several judges may be combined without the requirement of uniformity of ranking judgments over the judges. Fairly extensive tables are available [28] for the easy application of this method and further computation is in progress.

Two other methods of paired comparisons are of interest. Kendall and Babington Smith [31] proposed a method which is a combinatorial type test. They form a coefficient of agreement which essentially measures discrepancies from perfect agreement among judges and a coefficient of consistency for a single judge measured in terms of "circular triads".

Scheffé [36] has developed a method of paired comparisons which differs from the others in that it uses a scoring method and the analysis of variance. The method has the feature that the effect of order of presentation of paired samples to the judges is taken into account. This method seems admirably suited to consumer preference studies wherein a considerable time lag may occur between the testing of the two samples of a pair.

IV. DISCUSSION AND SUMMARY

In discussing a subject of the scope of the title of this paper certain sacrifices must be made. For completeness some topics have been included and discussed in a very superficial manner. Others may seem to have received more attention than is warranted. It was felt that the sections on sequential analysis were important to bring that phase of statistics to the attention of those interested in food and color testing by subjective appraisals. Emphasis was placed on the method of paired comparisons since some doubts and misunderstanding regarding the two models shown have arisen.

Most of the remarks have dealt with statistical methods for taste testing for differences. The subject of quality evaluation is a difficult one and one which requires considerable study. The author has been interested in the application of discriminant function techniques to the establishment of weights for the scores of various attributes in grading. This would appear possible if reliable grade determinations could be obtained independent of the present systems of weights. It is not clear how this could be managed and any practical applications of the technique would seem to involve somewhat circular arguments.

In conclusion the author would like to acknowledge the assistance of M. E. Terry, Boyd Harshbarger, Lyle L. Davis and D. B. Duncan in the studies of statistical methods for taste testing undertaken at the Virginia Agricultural Experiment Station.

V. CLASSIFIED BIBLIOGRAPHY

ANALYSIS OF VARIANCE

1. Bartlett, M. S. The use of transformations, *Biometrics* 3, 39, 1947.
2. Cochran, W. G. Problems arising in the analysis of a series of similar experiments, *Jour. of the Royal Statistical Society*, Supplement 4, 102, 1937.
3. Cochran, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied, *Biometrics* 3, 22, 1947.
4. Cox, G. M. Statistics as a tool for research, *Journal of Home Economics*, 36, 575, 1944.
5. Crump, S. L. The estimation of variance components in analysis of variance, *Biometrics* 2, 7, 1946.
6. Curtiss, J. H. On transformations used in the analysis of variance, *Annals of Math. Stat.* 14, 107, 1943.

6a. Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance, *Va. Jour. of Sci.*, 2, 171, 1951.

6b. Duncan, D. B. On the properties of the multiple comparisons test, *Va. Jour. of Sci.*, 3, 49, 1952.

7. Eisenhart, C. The assumptions underlying the analysis of variance, *Biometrics* 3, 1, 1947.
8. Snedecor, G. W. *Statistical Methods*, Iowa State College Press, Ames, 1946.
9. Tukey, J. W. Comparing means in the analysis of variance, *Biometrics* 5, 99, 1949.

CONSUMER PREFERENCES

10. Bliss, C. I., Anderson, E. D. and Marland, R. E. A technique for testing consumer preferences with special reference to the constituents of ice cream, *Storrs Agricultural Experiment Station Bulletin* 251, 1943.
11. Bogert, J. L. A method of consumer product testing, *Food Industries* 13, 47, 1941.
12. Garnatz, G. But what do the consumers say? *Food Industries* 22, 1333, 1950.
13. Nair, J. H. Mass taste panels, *Food Tech.* 3, 131, 1949.

DISCRIMINANT FUNCTION ANALYSIS

14. Baten, W. D. The use of discriminating functions in comparing judges' scores concerning potatoes. *J.A.S.A.* 40, 223, 1945.
15. Brown, G. W. Discriminant functions. *Annals of Math. Stat.* 18, 514, 1947.
16. Cochran, W. G. and Bliss, C. I. Discriminant functions with covariance. *Annals of Math. Stat.* 19, 151, 1948.

17. Johnson, P. O. The quantification of data in discriminant analysis. *J.A.S.A.* 45, 65, 1950.

EXPERIMENTAL DESIGNS

18. Bose, R. C. Partially balanced incomplete block designs with two associate classes involving only two replications. *Calcutta Statistical Association Bulletin* 3, 120, 1951.
19. Cochran, W. G. and Cox, G. M. *Experimental Designs.* John Wiley and Sons, New York, 1950.
20. Connor, W. S. On the structure of balanced incomplete block designs. *Annals of Math. Stat.* 23, 57, 1952.
21. Greenwood, M. L., Potgieter, N. and Bliss, C. I. The effect of certain pre-freezing treatments on the quality of eight varieties of cultivated high bush blueberries. *Food Research* 16, 154, 1951.
22. Hanson, H. L., Cline, L. and Lineweaver, H. Application of balanced incomplete block design to scoring of ten dried egg samples. *Food Tech.* 5, 9, 1951.
23. Harshbarger, B. Triple rectangular lattices. *Biometrics* 5, 1, 1949.
24. Harshbarger, B. Near balanced rectangular lattices. *Va. Jour. of Sci.* 2 (New Series), 13, 1951.
25. Harshbarger, B. and Davis, L. L. Latinized rectangular lattices. *Biometrics* 8, 73, 1952.
26. Hopkins, J. W. Laboratory flavor scoring—a two-factor experiment in complete and incomplete blocks. *National Research Council, Canada, Publication, Committee on Applied Mathematical Statistics* No. 7 (mimeo).
27. Shrikhandi, S. S. Designs for two-way elimination of heterogeneity. *Annals of Math. Stat.* 22, 235, 1951.

PAIRED COMPARISONS

28. Bradley, R. A. and Terry, M. E. Rank analysis of incomplete block designs I. To be published in *Biometrika*, 39, 1952.
29. Guttman, L. An approach for quantifying paired comparisons and rank order. *Annals of Math. Stat.* 17, 145, 1946.
30. Hening, J. C. Flavor evaluation procedures, *New York Agricultural Experiment Station Technical Bulletin* 284, 1948.
31. Kendall, M. G. and Babington Smith, B. On the method of paired comparisons. *Biometrika* 31, 324, 1939–40.
32. Moran, P. A. P. On the method of paired comparisons, *Biometrika* 34, 363, 1947.
33. Mosteller, F. Remarks on the method of paired comparisons I. *Psychometrika* 16, 3, 1951.
34. Mosteller, F. Remarks on the method of paired comparisons II. *Psychometrika* 16, 203, 1951.
35. Mosteller, F. Remarks on the method of paired comparisons III. *Psychometrika* 16, 207, 1951.
36. Scheffé, H. On analysis of variance for paired comparisons, *J.A.S.A.* 47, 381, 1952.
37. Terry, M. E., Bradley, R. A. and Davis, L. L. New designs and techniques for organoleptic testing. *Food Tech.* 6, 250, 1952.
38. Thurstone, L. L. Psycho-physical analysis. *Am. Jour. of Psychology* 38, 368, 1927.

39. Thurstone, L. L. *An Experiment in the Prediction of Choice.* Publication of the Psychometric Laboratory, University of Chicago, No. 68, 1951.

RANK ORDER METHODS

40. Baten, W. D. Analysis of scores from sampling tests. *Biometrics* 2, 11, 1946.
41. Baten, W. D. and Trout, G. M. A critical study of the summation difference-in-rank method of determining proficiency in judging dairy products. *Biometrics* 2, 67, 1946.
42. Charley, H. Effect of baking pan material on heat penetration during baking and on quality of cakes made with fat. *Food Research* 15, 155, 1950.
43. Dixon, W. J. The criterion for testing the hypothesis that two samples are from the same population. *Annals of Math. Stat.* 11, 199, 1940.
44. Dixon, W. J. and Mood, A. M. The statistical sign test. *J.A.S.A.* 41, 557, 1946.
45. Durbin, J. Incomplete blocks in ranking experiments. *Br. Jour. of Psychology,* Stat. Section 4, 1951.
46. Fisher, R. A. and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, Edinburgh, 1948.
47. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J.A.S.A.* 32, 675, 1937.
48. Galinat, W. C. and Everett, H. L. A technique for testing flavor in sweet corn. *Agronomy Journal* 41, 1949.
49. Greenwood, M. L. and Salerno, R. Palatability of kale in relation to cooking. *Food Research* 14, 314, 1949.
50. Hotelling, H. and Pabst, M. R. Rank correlation and tests of significance involving no assumption of normality. *Annals of Math. Stat.* 7, 29, 1936.
51. Kendall, M. G. and Babington Smith, B. The problem of *m* rankings. *Annals of Math. Stat.* 10, 275, 1939.
52. Kendall, M. G. *Rank Correlation Methods.* Charles Griffin and Co., London, 1948.
53. Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Math. Stat.* 18, 50, 1947.
54. Mood, A. M. *Introduction to the Theory of Statistics.* McGraw-Hill, New York, 1950.
55. Terry, M. E. Some Rank Order Tests Which Are Most Powerful Against Specific Parametric Alternatives. *Annals of Math. Stat.*, 23, 346, 1952.
56. Wald, A. and Wolfowitz, J. On whether two samples are from the same population, *Annals of Math. Stat.* 11, 147, 1940.
57. Walsh, J. E. Some significance tests for the median which are valid under very general conditions, *Annals of Math. Stat.* 20, 64, 1949.
58. White, C. The use of ranks in a test of significance for comparing two treatments, *Biometrics* 8, 33, 1952.
59. Wilcoxon, F. Individual comparisons by ranking methods, *Biometrics* 1, 80, 1945.
60. Wilcoxon, F. Probability tables for individual comparisons by ranking methods, *Biometrics* 3, 119, 1947.

REFERENCE MATERIAL

61. *Minutes of Conference on Taste Panel Procedures and Methods of Evaluation for Soy Bean Oil and Soy Bean Oil Products,* Northern Regional Research Laboratory, Peoria, Ill., 1949 (mimeo).

62. *Taste Panels, Bibliography.* Quartermaster Food and Container Institute, Chicago, Ill., (mimeo).
63. Bradley, R. A. and Duncan, D. B. *Bi-Annual Report No. 1 on Statistical Methods for Sensory Difference Testing,* (mimeo), 1950.
64. Bradley, R. A. and Terry, M. E. *Bi-Annual Report No. 2 on Statistical Methods for Sensory Difference Testing,* (mimeo), 1951.
65. Bradley, R. A. and Terry, M. E. *Bi-Annual Report No. 3 on Statistical Methods for Sensory Difference Testing,* (mimeo), 1951.
66. Bradley, R. A. and Terry, M. E. *Bi-Annual Report No. 4 on Statistical Methods for Sensory Difference Testing,* (mimeo), 1952.
67. Dawson, E. H. and Harris, B. L. Sensory methods for measuring differences in food quality. *Agricultural Information Bulletin No. 34,* U.S.D.A., 1951.
68. Gilford, J. P. *Psychometric Methods.* McGraw-Hill, New York, 1936.
69. Justin, C. and Adams, G. *Published and Processed Reports of Research in Foods.* Human Nutrition and Home Economics at the Land-Grant Institutions, Oct. 1950–Oct. 1951, Office of Experiment Stations, Agricultural Research Administration, U.S.D.A. (mimeo).
70. Pearson, K. *Tables of the Incomplete Beta-Function.* Cambridge University Press, Cambridge, England, 1934.
71. Scheffé, H. Statistical inference in the non-parametric case, *Annals of Math. Stat.* 14, 305, 1943.

SCORING TECHNIQUES

72. Baten, W. D. Organoleptic tests pertaining to apples and pears, *Food Research* 11, 84, 1946.
73. Baten, W. D. Reaction of age groups to organoleptic tests. *Food Tech.* 4, 277, 1950.
74. Boggs, M. M. and Ward, A. C. Scoring techniques for sulfited foods. *Food Tech.* 4, 282, 1950.
75. Cochran, W. G. The comparison of different scales of measurement for experimental results. *Annals of Math. Stat.* 14, 205, 1943.
76. Dove, W. F. The relative nature of human preference. *Jour. of Comparative Psychology* 35, 219, 1943.
77. Hening, J. C. Operations of a routine testing group in a small laboratory, *Food Tech.* 3, 162, 1949.
78. Hopkins, J. W. A procedure for quantifying subjective appraisals of odor, flavor, and textures of foodstuffs. *Biometrics* 6, 1, 1950.
79. Norton, H. W. Uses of scores in an exact test of significance in a discontinuous distribution. *Annals of Eugenics* 7, 349, 1937.
80. Ward, A. C. and Boggs, M. M. Comparison of scoring results for two and four samples of corn per taste session. *Food Tech.* 5, 219, 1951.

SEQUENTIAL ANALYSIS

81. Anscombe, F. J. Large-sample theory of sequential estimation. *Biometrika* 36, 455, 1949.
82. Harrison, S. and Elder, L. W. Some applications of statistics to laboratory taste testing. *Food Tech.* 4, 434, 1950.
83. Lombardi, G. J. *The Sequential Selection of Judges for Organoleptic Testing.* (Thesis) Virginia Polytechnic Institute, 1951.
84. Rao, C. R. Sequential tests of null hypotheses. *Sankhya* 10, 361, 1950.
85. Wald, A. *Sequential Analysis.* John Wiley and Sons, New York, 1947.