THE SEQUENTIAL SELECTION OF JUDGES FOR ORGANOLEPTIC TESTING

by

Gerardo J. Lombardi

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute
in candidacy for the degree of

MASTER OF SCIENCE
in
STATISTICS

APPROVED:

Director of Graduate Studies

APPROVED:

Head of Department

Dean of Applied Science

Supervisor or Major Professor

June 1951

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. R. A. Bradley for his guidance during this study so that it could be brought to a successful conclusion. Appreciation is also given to Dr. Boyd Harshbarger and Dr. M. E. Terry for their helpful suggestions during the course of this work. In addition the author expresses his thankfulness to the Bureau of Agricultural Economics, United States Department of Agriculture, for their interest in this study.

TABLE OF CONTENTS

I. INTRODUCTION

1.1 Formulation of the Problem:

Taste testing panels are being used to determine whether differences can be detected in the taste of foods due to cooking methods, age in storage, addition of preservatives, and other sources. The results of an experiment in taste testing however have little meaning unless the judges demonstrate ability to detect differences in taste when such differences do exist, and to do this with good consistency. These differences in many cases are not so pronounced that they can be detected easily.

Two tests, the duo-trio test and the triangle test, are in use for determining an individual's ability as a judge. In a duo-trio test, an individual is presented with three samples, one of which is labelled. Of the other two, one is identical to the labelled sample, the other is different. The prospective judge is asked to choose the one which he believes to be different. The probability of choosing the odd sample by chance associated with the duo-trio test is one-half. The triangle test is the same as the duo-trio test with the exception that none of the samples are labelled. The guessing probability associated with the triangle test is therefore one-third. In order to show with reasonable confidence that he is doing more than

guessing, the prospective judge is required to repeat the experiment successfully a considerable number of times.

A sequential testing plan is well suited to the selection of judges, since on the average fewer trials are required for such a procedure than for other equally reliable selection methods. The characteristic feature of a sequential plan is that the number of observations is not predetermined. The decision to terminate the experiment depends, at any stage, on the results of the observations previously made. A small number of trials is desirable in taste testing, since taste fatigue accompanies repetition, and only a few trials can be performed at one sitting. This increases the time interval for an individual to demonstrate discriminating ability, as well as the cost of preparation and performance of the experiment.

1.2 Review of Literature:

A test procedure leading to the acceptance or rejection of an hypothesis, is a rule specifying, for each sample of a given size, whether the hypothesis should be accepted or rejected. Choosing a test procedure is equivalent to determining two regions, such that if observations fall in one region, the hypothesis is accepted, and, if observations fall in the other region, the hypothesis is rejected.

i. Choice of Critical Regions: Jerzy Neyman and E. S. Pearson [2] have advanced the following criteria as a basis for choosing a critical region. In the acceptance or rejection of an hypothesis, errors of two kinds may be committed. An error of the first kind, committed with a selected probability α, occurs ~~when~~ if the hypothesis is rejected, when it is in fact true. An error of the second kind is committed with probability β, if the hypothesis is accepted when it is false. α and β are uniquely determined with the choice of a critical region. α is called the size of the critical region, and $1-\beta$, the power. From regions of fixed size α, that one should be chosen which is most powerful against a specific alternative hypothesis, that is, $1-\beta$ is the maximum of the power function and dependent on the alternative hypothesis.

ii. Sequential Tests of Statistical Hypotheses: Abraham Wald should be credited with the development of the sequential test procedure. The sequential method of testing hypotheses may be described as follows: A rule is given for making one of the following decisions at any stage of the experiment, (1) accept the hypothesis, (2) reject the hypothesis, (3) continue the experiment by taking an additional observation.

The operating characteristic function of a sequential test is closely associated with the concept of power defined

above. The probability that a sequential process will terminate with the acceptance of a null hypothesis H_o depends, according to Wald [7] (c.f. Section 2.2.1), only on the distribution of the random variable X under consideration. If this distribution is known except for the specification of a certain parameter θ, the probability of accepting H_o is a function of θ. This function is denoted by $L(\theta)$ and termed the operating characteristic function. Given that the probability is one that the test will eventually terminate, the probability of rejecting the null hypothesis is equal to $1-L(\theta)$.

Of particular interest is the average number of observations required if a particular sequential test were applied repeatedly. For any given test the number of observations depends on the distribution of X, which in turn is specialized by the value of θ. Therefore the average number of observations depends only on this specialized distribution. For any given parameter point θ, the average sample number is denoted by $E_\theta(n)$.

The operating characteristic function, and the average sample number are perhaps the most important features of a sequential test procedure. Wald has stated "The operating characteristic function describes how well the test achieves its objective of making correct decisions, and the average

sample number represents the price paid in terms of the number of observations."

1.3 Sequential Plans Particularly Applicable to the Selection of Judges:

Wald [7] (c.f. Section 5.2) has developed a test on the mean of the binomial distribution. In acceptance sampling of an industrial product it is possible to specify a percent defective p' and two values p_o and p_1, $p_o < p'$ and $p_1 > p'$, such that acceptance of the lot is regarded an error of practical importance if $p \geq p_1$ and rejection of the lot is regarded an error of practical importance if $p \leq p_o$, where p is the true proportion defective in the lot. The probability of rejecting the lot should not exceed some small preassigned value α, whenever $p \leq p_o$, and the probability of accepting the lot should not exceed some small preassigned value β, whenever $p \geq p_1$. If d_m is the number of defectives in the first m units inspected, the sequential sampling plan consists of computing at each stage

$$(1.1) \quad \log \frac{p_{1m}}{p_{om}} = d_m \log \frac{p_1}{p_o} + (m-d_m) \log \frac{1-p_1}{1-p_o} ,$$

where p_{1m} and p_{om} are respectively values of the likelihood function for sample size m when p takes the values p_1 and p_o. Inspection is continued as long as

$$(1.2)\quad \log \frac{\beta}{1-\alpha} < \log \frac{p_{1m}}{p_{om}} < \log \frac{1-\beta}{\alpha} .$$

It is terminated either the first time

$$(1.3)\quad \log \frac{p_{1m}}{p_{om}} \geq \log \frac{1-\beta}{\alpha} ,$$

in which case the lot is rejected, or the first time

$$(1.4)\quad \log \frac{p_{1m}}{p_{om}} \leq \log \frac{\beta}{1-\alpha} ,$$

in which case the lot is accepted. If wrong decisions in a taste testing experiment are associated with defectives in the sampling plan, it is seen that the above test is applicable to the selection of judges.

C. R. Rao [5] has developed sequential tests of null hypotheses distinct from those of Wald. The above sequential tests have been considered to discriminate between two simple or composite alternative hypotheses. According to Rao, this is not strictly applicable to testing a null hypothesis when nothing is known about the alternative hypothesis. Problems arise where the main emphasis consists in testing a null hypothesis. In such problems, no single hypothesis can be offered as alternative, nor is it possible to assign any risk function.

Large differences from the null hypothesis may be detected with a small sample of observations, and even small differences may be detected with probability approaching unity with an infinitely large number of observations. It

is not possible to accept the null hypothesis before an indefinitely large sample is observed. If the sequential process is terminated only when the null hypothesis is disproved, sampling may be continued indefinitely, and this will occur with high probability when the null hypothesis is true. It is therefore desirable to set an upper limit N to the number of observations and to stop sampling if the null hypothesis can be rejected at an earlier stage. It may happen that the null hypothesis does not stand rejected even at the N^{th} stage. In such a case, it may be accepted provisionally, or the class of hypotheses from which the observed samples could have reasonably arisen may be determined.

Rao [5] states that, when the alternatives are known to be one-sided (say $\theta > \theta_o$), in standard test procedures the locally most powerful test is defined by

(1.5) $P'_N(\theta_o) \geq \mu P_N(\theta_o)$ μ is a constant dependent on α.

where $P_N(\theta)$ is the likelihood function based on N observations, and $P'_N(\theta_o)$ is its first derivitive with respect to θ evaluated at $\theta = \theta_o$. The suggested sequential test is determined by the inequality,

(1.6) $P'_n(\theta_o) \geq A(N)P_n(\theta_o)$

after n observations have been taken and where A(N) is a suitable determined constant depending on N and the level

of significance. Sampling is terminated at the smallest value of n for which the above relationship is realized. The problem is then reduced to one of finding A(N). Rao has devised an upper bound for A(N), $\frac{\tau_m}{\alpha}$, where

$$(1.7)\quad \tau_m = \int_w P'_N(\theta_o)dv,$$

where w is the region of the sample space which maximizes this integral of $P'_N(\theta_o)$ and such that

$$(1.8)\quad \int_w P_N(\theta_o)dv = \alpha.$$

dv is the element of volume in the sample space. The actual probability of rejection of H_o when true, using $\frac{\tau_m}{\alpha}$ instead of A(N), is less than α. Rao has devised plans for testing hypotheses on the parameters of a normal distribution. In a following section a plan is derived for testing hypotheses on the mean of a binomial distribution. The procedure will be directly applied to judge selection in sensory difference testing.

II. WALD'S SEQUENTIAL ANALYSIS APPLIED TO THE SELECTION OF JUDGES

2.1 Application of the Method to the Duo-Trio and Triangle Tests:

In industrial schemes, the items inspected are usually classified as defective or non-defective. For such a dichotomous population the so-called binomial distribution is generated. The value one is assigned to each defective unit and the value zero to each non-defective unit. The sequential plan for testing hypotheses on the mean of the binomial distribution then applies.

The decision to accept or reject an individual as a judge for taste testing experiments is directly analagous to an industrial sampling scheme. This decision may be based on either a series of triangle or duo-trio trials. In such a comparison each trial is associated with an inspection unit. A correct decision is taken to correspond to a non-defective unit and an incorrect decision to a defective unit. It is clear that such a testing scheme also follows the basic binomial distribution, and that the sequential plan developed by Wald again applies.

Let p be the true proportion of incorrect decisions if the judge could continue testing indefinitely. Then 1-p, the proportion of correct decisions, may be defined as the judge's inherent ability under the test administered. A

value $1-p_o$ can be specified such that individuals having abilities greater than or equal to $1-p_o$ are considered qualified judges and should be included in the taste testing panel. In a like manner, a value, $1-p_1$, can be specified such that individuals having abilities less than or equal to $1-p_1$, are not considered qualified judges and should be excluded from the taste testing panel.

In order to determine acceptance or rejection, the following procedure is utilized. A graph is drawn on which the number of observations m is plotted as the abscissa, and the number of incorrect decisions d_m as the ordinate. Two lines L_o and L_1, having a common slope, determine the acceptance and the rejection regions. Wald [7] (c.f. Section 5.3.3) gives the intercepts and slope for plotting these lines. The intercept of L_o is[1]

$$(2.1) \quad h_o = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{p_1}{p_o} - \log \frac{1-p_1}{1-p_o}},$$

and the intercept of L_1 is

$$(2.2) \quad h_1 = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_o} - \log \frac{1-p_1}{1-p_o}},$$

---

1. log is used to denote a logarithm to base 10. The formulae used are, however, independent of the logarithmic base.

and their slope is

$$(2.3)\quad s = \frac{\log \frac{1-p_o}{1-p_1}}{\log \frac{p_1}{p_o} - \log \frac{1-p_1}{1-p_o}} .$$

To use the chart, the points (m,d_m) are plotted after each trial. An additional trial is performed as long as the point (m,d_m) lies between the lines L_o and L_1. The experiment is terminated the first time the point (m,d_m) does not lie between L_o and L_1, accepting the individual as a judge if the point lies on L_o or below, and rejecting the individual if the point lies on L_1 or above.

The average sample numbers for the test can be derived quite easily for values of p equal to zero, p_o, p_1, and unity. The equations, as given by Wald [7] (c.f. Section 5.5), are as follows:

$$(2.4)\quad E_0(n) = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{1-p_1}{1-p_o}} ,$$

$$(2.5)\quad E_{p_o}(n) = \frac{(1-\alpha) \log \frac{\beta}{1-\alpha} + \alpha \log \frac{1-\beta}{\alpha}}{p_o \log \frac{p_1}{p_o} + (1-p_o) \log \frac{1-p_1}{1-p_o}} ,$$

$$(2.6)\quad E_{p_1}(n) = \frac{\beta \log \frac{\beta}{1-\alpha} + (1-\beta) \log \frac{1-\beta}{\alpha}}{p_1 \log \frac{p_1}{p_o} + (1-p_1) \log \frac{1-p_1}{1-p_o}} , \text{ and}$$

$$(2.7)\quad E_1(n) = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{p_1}{p_o}} .$$

These results are useful in the planning of a tasting series for which it is necessary to estimate the amount of experimental material required in advance, and to decide on the values of α and β that can be specified from practical considerations. In a succeeding section, the average sample number is used as a criterion for comparison of the duo-trio and triangle tests.

2.2 Illustrative Examples:

For the reader not well versed in the applications of sequential analysis, we turn to numerical examples. Let us first suppose that the duo-trio test is the basis of a sequential procedure. Reasonable ability limits may be stated as $1-p_o = 0.75$ and $1-p_1' = 0.55$. Thus,to qualify as a judge on this panel, ability of 0.75 must be displayed on duo-trio trials. Individuals with abilities 0.55 or less will be excluded. α and β are taken to be 0.05. That is, the probabilities of rejecting a satisfactory judge and of accepting an unsatisfactory one are equal as stated.

When α, β, p_o, and p_1 are determined, we may compute the equations of L_o and L_1, using the intercepts and slope formulated in equations (2.1), (2.2), and (2.3). In the

specified case above,

(2.8) L_o: $d_m = 0.34m - 3.28$,

and

(2.9) L_1: $d_m = 0.34m + 3.28$.

These equations are plotted in Figure 1, determining the test limits.

Tables of random numbers [2,6] were used to assimilate sequential trials with known values of p. For the duo-trio test, two trials with p having values 0.5 and 0.2 are recorded in Table I and plotted in Figure 1. Both trials led to correct decisions and terminated with sample numbers consistent with those computed from equations (2.4) to (2.7) and shown in Table II.

TABLE I

SEQUENTIAL TRIALS WITH A DUO-TRIO TEST

	Sample Number																								
p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
.5	1	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	1	1	1	0	1	1	
.2	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0

A similar series of sequential trials was obtained using repeated triangle tests. In this case, minimum acceptable ability is $1-p_o = 0.65$ and $1-p_1 = .4$. α and β remain at 0.05. The control limits become

(2.10) L_o: $d_m = 0.47m - 2.88$,

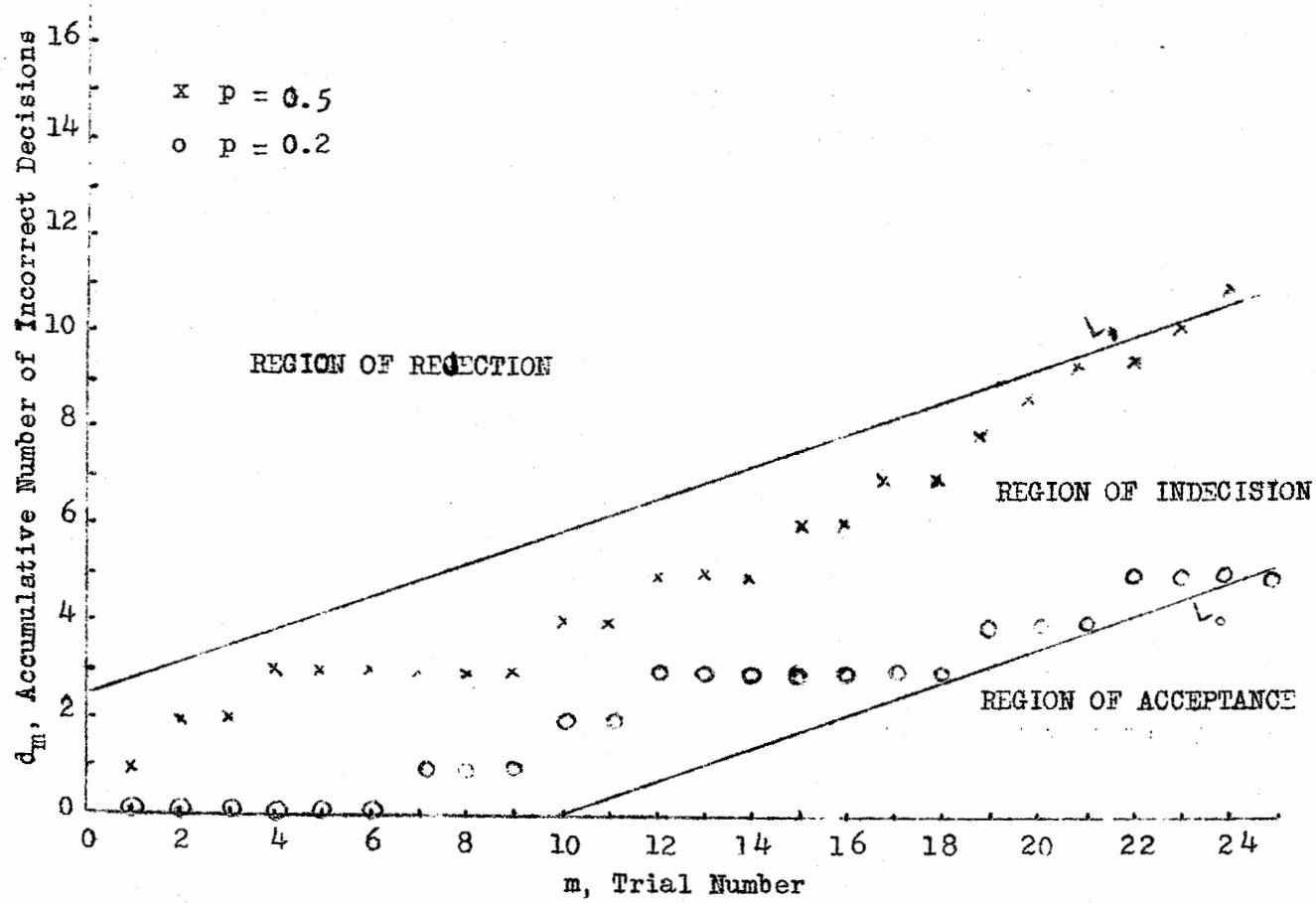

Figure 1: SEQUENTIAL TRIALS WITH THE DUO-TRIO TEST

and

$$(2.11)\quad L_1: \quad d_m = 0.47m + 2.88.$$

Two trials obtained using tables of random numbers are shown in Table III, and are plotted on Figure 2 with L_o and and L_1. Average sample numbers are shown in Table IV.

TABLE II.

AVERAGE SAMPLE NUMBERS FOR THE DUO-TRIO TEST

p	$E_p(n)$
0	2
0.25	32
0.45	27
1	5

TABLE III

SEQUENTIAL TRIALS WITH A TRIANGLE TEST

	Sample Number														
p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
.67	0	0	1	1	1	0	1	1	1	0	1	0	0	0	1
.4	0	1	0	1	0	1	1	0	0	0	0	1	0	0	0

	Sample Number														
p	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
.67	1	1	1	1	1											
.4	1	1	0	0	1	0	1	0	1	1	0	0	1	0	0	0

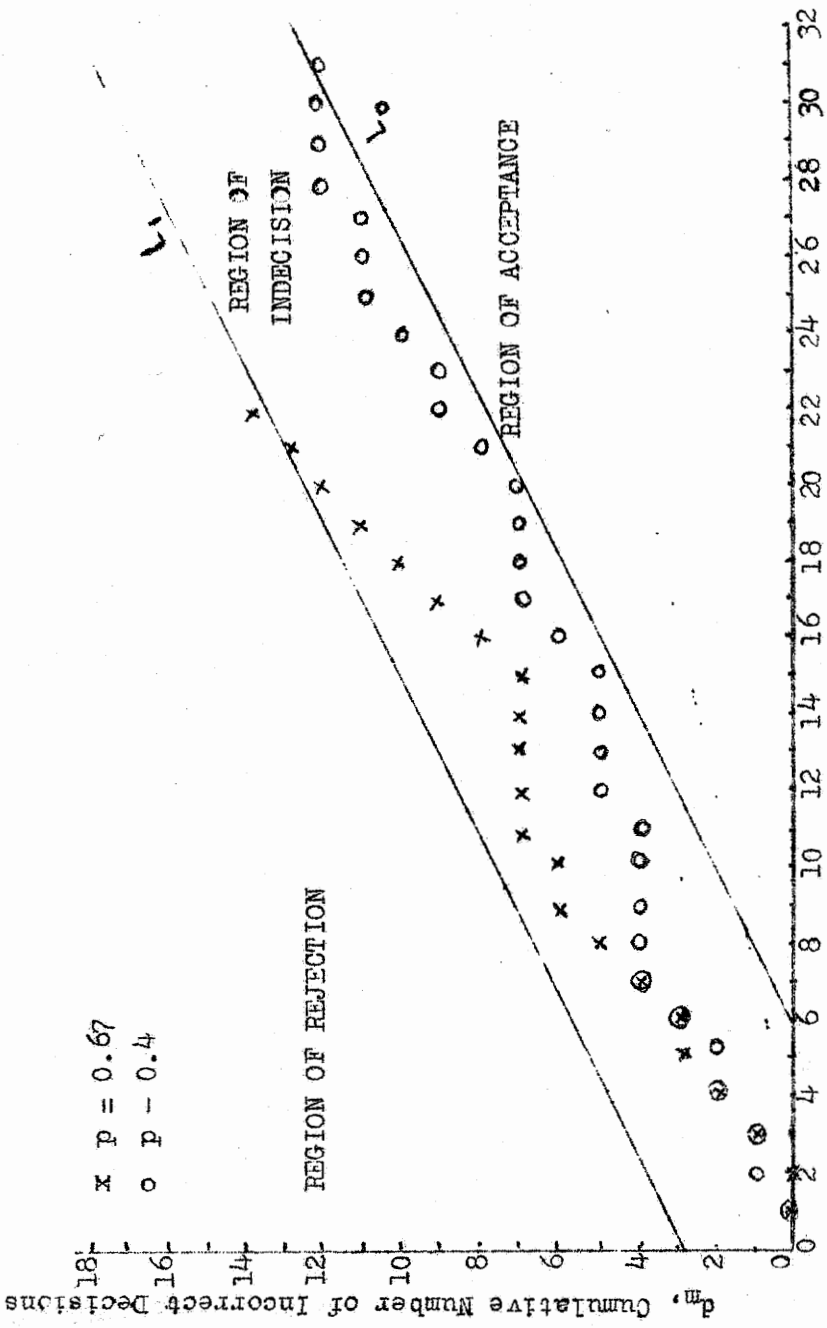

Figure 2: SEQUENTIAL TRIALS WITH THE TRIANGLE TEST.

TABLE IV

AVERAGE SAMPLE NUMBERS FOR THE TRIANGLE TEST

p	$E_p(n)$
0	6
0.35	21
0.6	23
1	5

A few general remarks are in order regarding the limiting abilities $1-p_o$ and $1-p_1$ specified. It is not suggested that the values used in the illustrations actually be used in selecting judges for an experimental panel. The experimenter may feel, and with some reason, that more stringent limits should be required. It must be noted that as p_1 approaches p_o, the average sample number required for determining whether an individual is acceptable or not, increases considerably. However, one means of reducing the average sample numbers is to increase α, accepting a greater risk of rejecting a capable judge. This may be done, for example, if large numbers of individuals are available as prospective judges so that the rejection of an able judge is not regarded as a serious error. The relationship existing between abilities in the duo-trio test and the triangle test, is not known explicitly. This problem is considered in the following section.

2.3 Abilities on Duo-Trio and Triangle Tests Compared.

The duo-trio and triangle tests could be compared as bases for sequential trials, if a correspondence between the limiting abilities $1-p_o$ and $1-p_1$ were determined. This correspondence, although apparently simple in conception, is extremely difficult to determine mathematically. The initial difficulty arises with a definition of equivalence for these two tests. The one definite feature is the knowledge that a chance proportion 1-p of correct decisions $\frac{1}{2}$ on duo-trio trials corresponds to a chance proportion of correct decisions 1/3 on triangle trials.

One criterion for obtaining equivalent values of $1-p_o$ for the two tests is to require that the probability of at least $N(1-p_{od})$ correct decisions on N duo-trio tests be equal to the probability of at least $N(1-p_{ot})$ correct decisions in N triangle tests. A similar relation may be required for p_{1d} and p_{1t}. A second reasonable, yet distinct, criterion is the requirement that the average sample numbers, evaluated at the upper control limits of ability, for the two sequential tests be equal when the corresponding lower control limits are specified at the chance proportions of correct decisions noted above.

If we use the first criterion, the probabilities in N binomial trials are generated by $\left[p+(1-p)\right]^N$. This expression may be expanded in a Taylor's series with

integral remainder [1]. Thus

$$(2.12)\quad \left[p+(1-p)\right]^N = \sum_{j=o}^{n_o-1} \binom{N}{j} p^{N-j}(1-p)^j + R_{n_o},$$

where R_{n_o} is the remainder after n_o terms.

$$(2.13)\quad R_{n_o} = \frac{N!}{(n_o-1)!(N-n_o)!} \int_o^{1-p} x^{n_o-1}(1-x)^{N-n_o}\, dx,$$

or, in the usual notation,

$$(2.14)\quad R_{n_o} = I_{1-p}(n_o, N-n_o+1).$$

This function may be evaluated by use of the incomplete beta function [4]. We note that R_{n_o} is the probability of at least n_o correct decisions and is a function of both p and N.

To compare the duo-trio and the triangle tests, we equate

$$(2.15)\quad I_{\frac{1}{2}}(n_{od}, N-n_{od}+1) = I_{1/3}(n_{ot}, N-n_{ot}+1) = \alpha.$$

On the basis

$$(2.16)\quad 1-p_{od} = \frac{n_{od}}{N}, \text{ and } 1-p_{ot} = \frac{n_{ot}}{N}.$$

Similarly, we may obtain

$$(2.17)\quad 1-p_{1d} = \frac{n_{1d}}{N}, \text{ and } 1-p_{1t} = \frac{n_{1t}}{N}.$$

These expressions can be evaluated from the tables of the incomplete beta function for specified values of N and α.

From practical consideration, it is sufficient to assume that the ratios $\frac{n_{od}}{n_{ot}}$ and $\frac{n_{1d}}{n_{1t}}$ are stable and equal to

R_N for all α. Values of n_o and R_N are given in Table V for selected values of N and α.

From the tabulated values of R_N, it is clear that R_N increases with increasing values of α and N. It is of particular interest to determine the greatest value that R_N can assume as N becomes large. In the limit, under the null hypothesis of chance match pairings in the duo-trio and triangle tests, $1-p_{od} \rightarrow 1/2$ and $1-p_{ot} \rightarrow 1/3$ in probability by Khintchine's Theorem. Thus,

$$\lim_{N \rightarrow \infty} R_N = 1.5, \quad \text{in probability} \tag{2.18}$$

Let us take an example using this method of comparing abilities on the duo-trio and triangle tests. In the previous illustrations, the control limits of ability specified for the duo-trio trials were 0.75 for $1-p_o$ and 0.55 for $1-p_1$. We will determine the corresponding control limits of ability for a series of triangle trials. As will be remembered, the average sample numbers for the sequential plan based on the duo-trio test ranged from 2 to 32. The value 1.33, consistent with these average sample numbers, is chosen for R_N from the tables. Then by dividing 0.75 and 0.55 by this value of R_N, the corresponding control limits of ability for the triangle test are 0.57 for $1-p_0$ and 0.41 for $1-p_1$. If this method is used to compute corresponding control limits for the sequential tests, the values of the

TABLE V

VALUES OF n_{od}, n_{ot} AND R_N FOR SELECTED VALUES OF N AND α

$\alpha = 0.05$

N	5	6	7	8	9	10	11	12	13
n_{od}	4.84	5.45	6.14	6.82	7.44	8.08	8.76	9.34	9.95
n_{ot}	3.95	4.40	4.92	5.41	5.88	6.31	6.79	7.22	7.67
R_N	1.22	1.23	1.25	1.26	1.27	1.28	1.29	1.29	1.30
N	14	15	16	17	18	19	20	21	22
n_{od}	10.57	11.23	11.79	12.47	12.97	13.65	14.21	14.81	15.42
n_{ot}	8.10	8.51	8.95	9.39	9.82	10.28	10.71	11.07	11.55
R_N	1.31	1.32	1.32	1.32	1.32	1.33	1.33	1.34	1.34
N	23	24	25	26	28	30	32	34	
n_{od}	15.94	16.59	17.12	17.74	18.87	19.99	21.16	22.34	
n_{ot}	11.91	12.36	12.75	13.18	13.93	14.77	15.66	16.37	
R_N	1.34	1.34	1.34	1.35	1.35	1.35	1.36	1.36	
N	36	38	40	42	44	46	48	50	
n_{od}	23.49	24.62	25.76	26.85	27.95	29.08	30.22	31.35	
n_{ot}	17.13	17.91	18.72	19.52	20.64	21.00	21.82	22.61	
R_N	1.37	1.37	1.38	1.38	1.38	1.38	1.38	1.39	
N	52	54	56	58	60	84			
n_{od}	32.47	33.58	34.69	35.79	36.88	50.05			
n_{ot}	23.37	24.19	24.88	25.46	26.43	35.33			
R_N	1.39	1.39	1.39	1.40	1.40	1.42			

$\alpha = 0.01$				$\alpha = 0.1$			
N	13	30	60	5	13	30	60
n_{od}	11.13	21.86	39.53	4.40	9.24	19.00	35.50
n_{ot}	8.90	16.66	29.00	3.53	6.99	13.77	25.00
R_N	1.25	1.31	1.36	1.26	1.32	1.38	1.42

average sample numbers for the triangle test corresponding to Table II for the duo-trio test are given in Table VI. It is noted that these average sample numbers correspond reasonably well at the extremes, but differ most seriously at the intermediate points.

TABLE VI

AVERAGE SAMPLE NUMBERS FOR THE TRIANGLE TEST BY METHOD 1

p	$E_p(n)$
0	2
0.43	54
0.59	52
1	10

If the second criterion is used for determining equivalent control limits for the two sequential tests, we substitute 0.5 and 0.667 in turn for p_1 in equation (2.5) for $E_{p_o}(n)$ and equate. Thus

$$(2.19)\quad \frac{(1-\alpha)\log\frac{\beta}{1-\alpha} + \alpha\log\frac{1-\beta}{\alpha}}{p_{od}\log\frac{0.5}{p_{od}} + (1-p_{od})\log\frac{0.5}{1-p_{od}}} = \frac{(1-\alpha)\log\frac{\beta}{1-\alpha} + \alpha\log\frac{1-\beta}{\alpha}}{p_{ot}\log\frac{0.667}{p_{ot}} + (1-p_{ot})\log\frac{0.333}{p_{ot}}}$$

The numerators cancel leaving

$$(2.20)\quad p_{od}\log\frac{0.5}{p_{od}} + (1-p_{od})\log\frac{0.5}{1-p_{od}} =$$

$$p_{ot}\log\frac{0.667}{1-p_{od}} + (1-p_{ot})\log\frac{0.333}{1-p_{od}}.$$

The upper control limit $1-p_o$ is specified for either the duo-trio or triangle test and substituted in (2.20). The corresponding upper control limit for the other test is obtained by solution of the resultant equation. The correspondence is independent of α and β. For the lower control limits, p_{od} and p_{ot} are replaced by p_{ld} or p_{lt} in (2.20) and the same procedure is followed. The solutions of the resultant equations present some difficulty, but a method is illustrated in the following example.

Let us again consider the duo-trio test for which $1-p_{od} = 0.75$ and $1-p_{ld} = 0.55$ and compute the corresponding values for the triangle test. Substitution in (2.20) for p_{od} yields, upon reduction, the equation

$$(2.21)\quad p_{ot} = \frac{\log 2 + .25\log 3 + \log(1-p_{ot})}{\log 2 - \log p_{ot} + \log(1-p_{ot})}.$$

p_{ot} is determined by assuming a first approximation which is substituted in the right hand member of (2.21) from whence a second approximation to p_{ot} is obtained. Repetition of the process, permits evaluation of closer and closer approximations to p_{ot}. Again, p_{lt} is obtained in a

similar way. For the triangle test, the control limits become

$$1-p_{ot} = 0.64 \quad \text{and} \quad 1-p_{lt} = 0.35$$

The first method yielded values 0.57 and 0.41 respectively. Average sample numbers corresponding to those of Table II and Table VI are listed in Table VII. It is noted that by this method the average sample numbers again compare reasonably well with those of Table II at the extremes, but differ considerably at the intermediate points.

TABLE VII

AVERAGE SAMPLE NUMBERS FOR THE TRIANGLE TEST BY METHOD 2

p	$E_p(n)$
0	5
0.36	15
0.65	15
1	5

Neither criterion gives average sample numbers which are consistent with those given in Table II for the intermediate region of abilities. The second criterion does, however, offer the advantage of lower sample numbers in this region and is evident from the examples. To the experimenter, this reduction in the number of trials is likely to

be a highly important consideration. It is very possible that other criteria could be defined and defended. However, on the basis of the limited information available, it is suggested that the second is preferable.

III. THE SELECTION OF JUDGES REGARDED AS A SEQUENTIAL TEST OF A NULL HYPOTHESIS

Rao [5] has developed a sequential method for testing hypotheses on the parameters of the normal distribution, in which an upper limit N is assigned to the number of observations taken. This method differs from that of Wald in that only one control limit is specified. The null hypothesis may be rejected or tentatively accepted. A similar plan, based on the binomial distribution, can be directly applied to the sequential selection of judges for tasting experiments. Rao has not considered this distribution in his work, and the following discussion of a sequential plan for testing hypotheses on the mean of the binomial distribution when the alternatives are one sided ($1-p > 1-p_0$ say) is novel. In this plan, only one limit of ability $1-p_0$ need be specified. Individuals having abilities greater than $1-p_0$ are acceptable as judges, whereas those having abilities less than $1-p_0$ are not accepted. It is an advantage to the experimenter that he is able to limit the number of trials commensurate with available funds and time.

The methematical derivation of the plan will now be discussed.

3.1 Rao's Method Applied to the Binomial Distribution:

If $1-p$ is defined as the true mean of the binomial distribution, the locally most powerful test for testing the

hypothesis $1-p = 1-p_o$ against alternatives $1-p \geq 1-p_o$ based on N observations is defined by the region,

$$(3.1)\quad P'_N(1-p_o) \geq \mu P_N(1-p_o).$$

where

$$(3.2)\quad P_N(1-p_o) = \binom{N}{j}(1-p_o)^j\, p_o^{N-j}$$ and j is the number of correct decisions in N trials. $P'_N(1-p_o)$ is the first derivative of $P_N(1-p)$ with respect to $1-p$ evaluated at $1-p_o$.

$$(3.3)\quad P'_N(1-p_o) = \frac{j-N(1-p_o)}{p_o(1-p_o)}\, P_N(1-p_o).$$

μ is a suitably determined constant depending on N and α. By analogy, a sequential plan based on m observation, $m \leq N$, is determined by the region,

$$(3.4)\quad P'_m(1-p_o) \geq A(N)P_m(1-p_o),$$

or

$$(3.5)\quad \frac{j-m(1-p_o)}{p_o(1-p_o)} \geq A(N),$$

where A(N) is a suitably determined constant depending on N and α.

In accordance with Rao's method we will determine $\frac{\tau_m}{\alpha}$, the upper limit of A(N). Upon substitution of (3.3) in (3.1) and reduction, we obtain

$$(3.6)\quad j \geq N(1-p_o) + \mu p_o(1-p_o) = n_o,$$

the number of correct decisions in N trials for significance at the specified value of α. We now solve for τ_m which was defined in equation (1.7). For a discrete variate we

replace the integral by a sum and obtain

$$(3.7)\quad \tau_m = \sum_{j\geq n_o} P_N'(1-p_o).$$

Substitution of (3.3) yields

$$(3.8)\quad \tau_m = \frac{\sum_{j\geq n_o}\left[j-N(1-p_o)\right]P_N(1-p_o)}{p_o(1-p_o)},$$

or

$$(3.9)\quad \tau_m = \frac{\sum_{j\geq n_o} jP_N(1-p_o) - N(1-p_o)\sum_{j\geq n_o} P_N(1-p_o)}{p_o(1-p_o)}.$$

But, by (1.8),

$$(3.10)\quad \sum_{j\geq n_o} P_N(1-p_o) = \alpha$$

Therefore,

$$(3.11)\quad \tau_m = \frac{\sum_{j\geq n_o} jP_N(1-p_o) - N(1-p_o)\alpha}{p_o(1-p_o)}.$$

Let us consider the sum

$$\sum_{j\geq n_o} jP_N(1-p_o).$$

This sum may be written as

$$N(1-p_o)\sum_{j\geq n_o}\binom{N-1}{j-1}(1-p_o)^{j-1}\,p_o^{N-j}.$$

But by the result (2.12) of Chapter II, it is now evident that

$$(3.12)\quad \sum_{j\geq n_o} jP_N(1-p_o) = N(1-p_o)I_{1-p_o}(n_o-1,N-n_o+1).$$

Hence,

$$(3.13)\quad \tau_m = \frac{N(1-p_o)\left[I_{1-p_o}(n_o-1,N-n_o+1) - \alpha\right]}{p_o(1-p_o)}$$

The critical region of the sequential test, are defined by (3.5), becomes on substitution of $\frac{\tau_m}{\alpha}$ for A(N),

$$(3.14)\quad j \geq m(1-p_\theta) + \frac{N(1-p_o)\left[I_{1-p_o}(n_o-1,N-n_o+1) - \alpha\right]}{\alpha}.$$

n_o, corresponding to the one in equation (2.14), and $I_{1-p_\theta}(n_o-1,N-n_o+1)$ can be evaluated from the tables of the incomplete beta function for specified values of N and d. It is noted that, when the equality holds in (3.17) the resultant equation is linear in j and m thus determining a line L with slope

$$(3.15)\quad S_r = 1-p_o$$

and intercept

$$(3.16)\quad h_r = \frac{N(1-p_o)\left[I_{1-p_o}(n_o-1,\ N-n_o+1) - \alpha\right]}{\alpha}.$$

In order to determine acceptance or rejection of a judge, the following procedure is suggested. A graph is drawn in which the number of trials m is plotted as the abscissa, and the number of correct decisions j as the ordinate. The line L determines the acceptance and rejection regions. To use the chart, the points (m,j) are plotted after each trial. An additional trial is performed as long as (m,j) lies below L and above T, a truncation line drawn

at 45 degrees to the m axis intersecting L at m equal to N. The experiment is terminated either the first time (m,j) lies on or above L, in which case the individual is accepted as a judge, or anytime (m,j) croses T. If the latter event occurs it is apparent that the experiment can not result in acceptance of the individual as a judge in N trials.

3.2 Illustrative Examples:

As a numerical example illustrating the method for the duo-trio test, let us take N = 30, $1-p_o = 0.5$, and $\alpha = 0.05$. When these values have been specified, n_o can be evaluated from the tables of the incomplete beta function.[2] The values specified above yield $n_o = 20$. Having determined n_o, $I_{1-p_o}(n_o-1, N-n_o+1)$ is evaluated by the use of the tables of the incomplete beta function and found to be 0.0680. Then the equation of the curve becomes,

(3.17) L: $j = 0.5m + 5.4$.

This equation is plotted in Figure 3 and with the limit T determines the test limits.

The tables of random numbers were again used to produce sequential trials for known values of 1-p. For the duo-trio

---

2. For $1-p_o$ equal to 0.5 and 0.333, however, values of n_o are given in Table V for selected values of N and α.

test, two trials with 1-p having values 0.5 and 0.7 are shown in Table VIII and plotted in Figure 3. The trials using p = 0.5 are terminated at m = 18, since it is clear that continuation could not result in acceptance in 30 trials.

TABLE VIII

SEQUENTIAL TRIALS WITH A DUO-TRIO TEST

	Sample Number																	
1-p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0.5	0	1	0	0	1	0	0	0	1	1	1	0	0	1	1	1	0	0
0.7	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1

As a numerical example using the triangle test, we will take $1-p_o = 0.333$, $N = 30$, and $\alpha = 0.05$. In this case,

(3.18) L: $j = 0.333m + 5.6$

Two trials using the tables of random numbers are shown in Table IX and plotted with L in Figure 4.

TABLE IX

SEQUENTIAL TRIALS WITH A TRIANGLE TEST

	Sample Number																	
1-p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0.3	1	0	1	0	0	1	0	0	1	1	0	1	0	0	0	0	1	0
0.5	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	1		

TABLE IX (Continued)

	Sample Number								
1-p	19	20	21	22	23	34	25	26	27
0.3	1	1	0	0	1	1	0	0	0

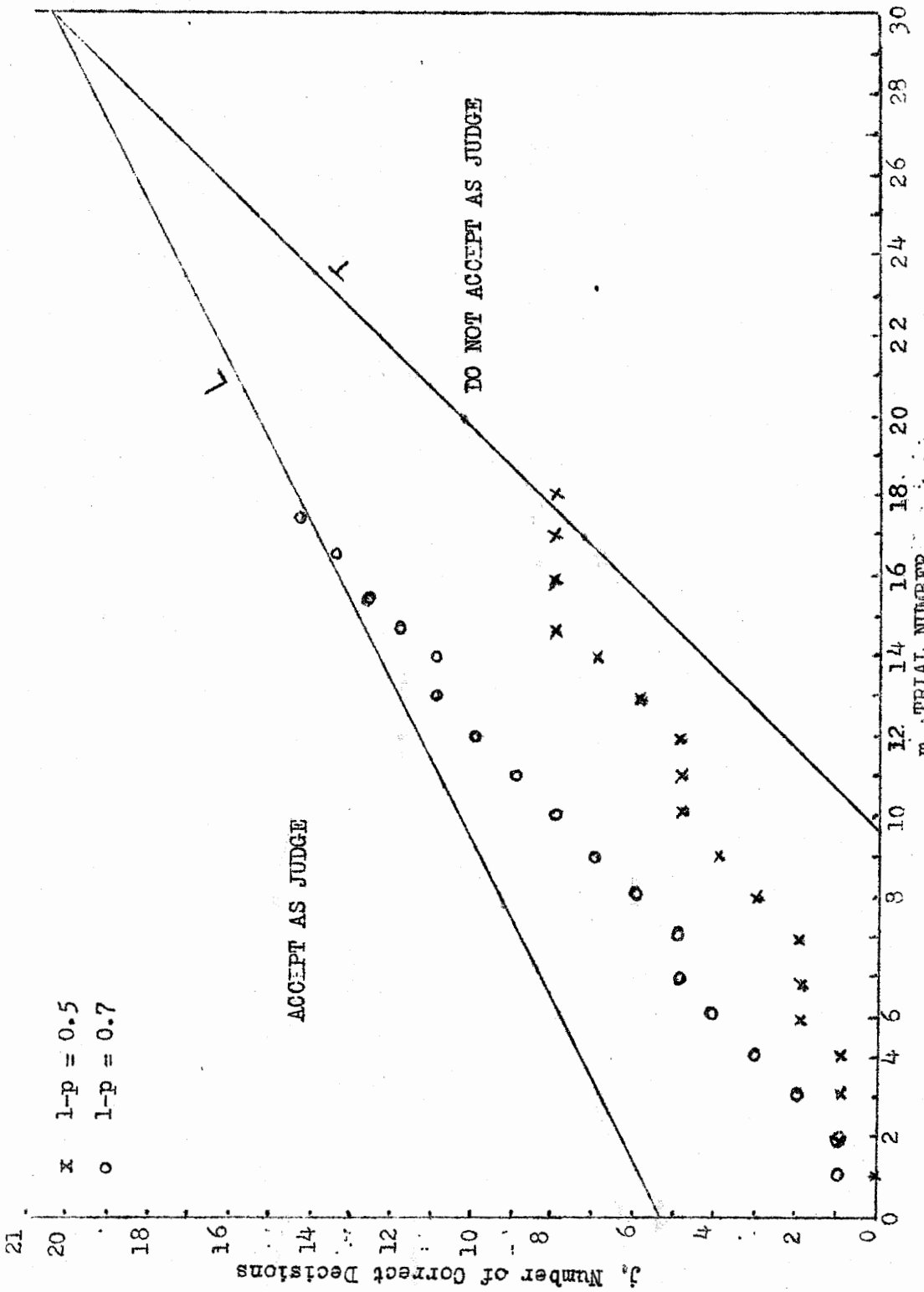

Figure 3: SEQUENTIAL TRIALS FOR DUO-TRIO TEST

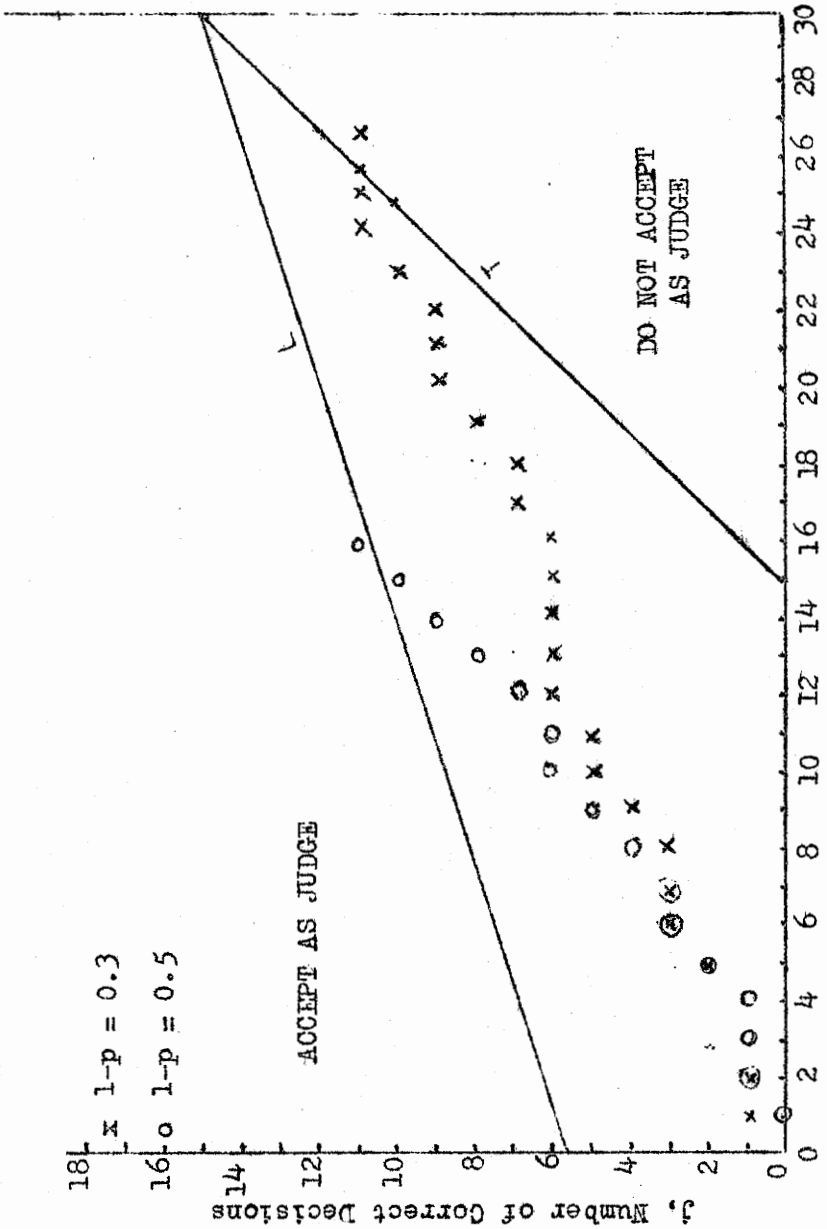

Figure 4: SEQUENTIAL TRIALS WITH THE TRIANGLE TEST

VI SUMMARY

Both the theory of sequential analysis developed by Wald and that developed by Rao may be applied to binomial distributions and may be used to select judges for taste tesing experiments. The duo-trio or triangle test may both be used as the fundamental test in either sequential process.

Wald has applied his sequential analysis to binomial distributions and it was only necessary to associate duo-trio and triangle trials with his examples on industrial sampling inspection. This was done by associating incorrect tasting decisions with defectives in industrial sampling. It has been noted that choice of α and β, the risks of type I and type II errors, may depend on the availability of potential judges and on the time and material available to the experimenter.

Rao's newer sequential process had not previously been developed for binomial type populations. The theory of this application is developed in this thesis and applied to the judge selection problem. It is shown that a single limit line is required for a graphical representation of this test and that, in addition, a truncation line may be introduced to limit the sequential process as soon as it becomes impossible to accept a judge.

The average sample number for Rao's test procedure has not been obtained either in general or for the binomial

application. Since the process is limited by an upper sample limit, computation of an average sample number function is not so important in planning experiments as it would be in using Wald's procedure. However, it is recommended that future study be made on this point. Similarly, the power of Rao's test procedure should be investigated.

Two criteria have been advanced for determining equivalence of duo-trio and triangle tests in sequential processes. The two methods do not agree and both result in some oversimplification of the situation. However, both criteria have certain features which are intuitively appealing. This problem is of some practical importance from the point of view of comparing results of different experimenters using similar sequential procedures but differing in the basic test employed. Although this problem is a somewhat paradoxical one, it would seem worthy of further study.

Apart from the developments of sequential methods for the selection of judges for taste testing experiments, this thesis has the additional objective of illustrating these methods in a manner such that food technicians may more easily recognize and understand the application to their experimental work. Several examples of both Wald's and Rao's methods are included in some detail and, although it

was not possible to conduct actual methodological experiments, the illustrations using tables of random numbers are quite analogous to experimental data.

It is expected that this thesis will be included in a bi-annual report to the Brueau of Agricultural Economics under the contract on "Statistical Methods For Sensory Difference Tests of Food Quality." It will thus obtain fairly wide-spread distribution to food technologists.

BIBLIOGRAPHY

1. Camp, B. H.: "Probability Integrals for the Point Binomial", Biometrika, Vol. 16, 1925.

2. Fisher, R. A. and Yates, Frank: Statistical Tables for Biological, Agricultural and Medical Research, Hafner Publishing Co., Inc., New York, N. Y., 1949.

3. Neyman, Jerzy and Pearson, E. S.; "On Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference", Biometrika, Vol 20A, 1928.

4. Pearson, Karl: Tables of the Incomplete Beta Function, The University Press, Cambridge, England, published by the Proprietors of Biometrika, 1934.

5. Rao, C. R.: "Sequential Tests of Null Hypotheses", Sankhyā, Vol. 10, 1950.

6. Snedecor, G. W.: Statistical Methods, The Iowa State College Press, Ames, Iowa, 1946.

7. Wald, Abraham: Sequential Analysis, John Wiley and Sons, Inc., New York, N. Y., 1947.

Gerardo J. Lombardi was born August 10, 1920, in St. Donato, Italy. His parents emigrated from Italy to the United States in November of the same year, and settled in Boston, Massachusetts. After graduation from the Huntington Preparatory School, Boston, in 1940, he attended the University of Notre Dame, receiving the B. S. degree in Chemical Engineering in 1943. Upon leaving college, he entered the U. S. Navy as a line officer, and served as executive officer of a landing ship operating on the Pacific waters. In 1946, he was discharged from the Navy and accepted a position with the engineering department of the Stauffer Chemical Company, Chauncey, New York. He left their employ in 1949 and became a condidate for the degree M. S. in Chemical Engineering at the Virginia

Polytechnic Institute, completing the requirements in December, 1950. At that time, he became a candidate for the degree M. S. in Statistics. Mr. Lombardi is married and has one child.