1-1-1975

The utilization of sequential analysis in psychological research.

James Michael Goett

THE UTILIZATION OF SEQUENTIAL ANALYSIS IN PSYCHOLOGICAL RESEARCH

by

James Michael Goett

A Thesis

Presented to the Graduate Committee

of Lehigh University

in Candidacy for the Degree of

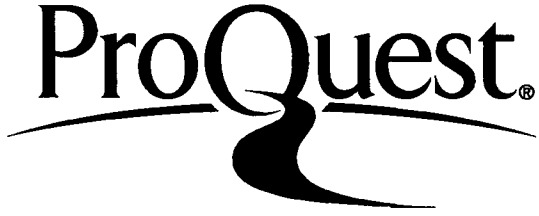
Master of Science

in

Psychology

Lehigh University

1975



THE UTILIZATION OF SEQUENTIAL ANALYSIS IN PSYCHOLOGICAL RESEARCH

by

James Michael Goett

A Thesis

Presented to the Graduate Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Psychology

Lehigh University

1975

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

9/18/75
(date)

Professor in Charge

Chairman of Department

This thesis is dedicated to my wife, Lucy, whose interest and encouragement has enabled me to complete this work.

I would like to thank Dr. Edwin Kay for his sincere interest and generous assistance in the preparation of this manuscript. I would also like to thank Drs. Roger Loeb, Arthur Brody, Martin Richter, and Lawrence Paul for their helpful comments.

Finally, a special note of thanks to my brother, Steve, for his assistance in obtaining certain documents referenced herein.

Table of Contents

A Description 2

History 4

Theory. 6

Delta 19

Implementation. 25

A Demonstration 32

Reference List and Bibliography 56

Appendices. 61

Vita. 74

List of Tables

Table 1: The Sequential Test's Probability Matrix .14

Table 2: Sensitivity of the t-Test with Paired Observations. 21

Table 3: The Probability Distributions for Both Experimental Groups Given the Incremental Learning Theory 39

Table 4: The NBS Approximation t-Test 42

Table 5: The Rushton Approximation t-Test 44

Table 6: The Sequential Sign Test. 46

Table 7: The Economy of the Sequential Tests . . . 49

Table 8: Quadratic Interpolation of the Accept-H_o Boundary 68

Table 9: Quadratic Interpolation of the Reject-H_o Boundary 70

List of Figures

Figure 1: Graphical Representation of the NBS Approximation t-Test. 51

Figure 2: Graphical Representation of the NBS Approximation t-Test with Square Root Transformation. 53

Abstract

The statistical technique of sequential analysis was presented with a historical and theoretical introduction. Its economy of sampling and ability to control the power of a statistical test were shown to be advantages for the scientific researcher despite the necessity to specify the critical size of the effect to be detected. The actual procedures were then detailed for the sequential sign test and two approximations of the sequential t-test.

To demonstrate this technique's usefulness in psychological research, a verbal learning experiment was replicated. This previous experiment's design was an attempt to control the serial position effect in paired-associate learning that may minimize any difference between the experimental and control groups. The finding of no difference was replicated.

In this paper I intend to present the technique of sequential analysis and show its application to scientific research, and in particular, psychological research. I will give a brief description and history of this technique, explain its theory (in part) and finally, show its practical use by means of a demonstration within psychology.

A Description

Sequential analysis is a procedure for testing statistical hypotheses. It is similar to the traditional statistical techniques, especially in the types of hypotheses tested and some of the theoretical concepts involved. It does, however, have some characteristics which make it superior to the tests used at present.

Briefly, a test procedure is sequential when the test is performed as each observation is taken within an experiment. The experimental manipulations may be stopped after any particular observation, depending on which of three decisions is made:

1) accept the null hypothesis;

2) reject the null hypothesis (accept the alternative hypothesis);

3) take another observation.

Barnard put it another way.

> It is at this point that sequential analysis poses the question in a more natural manner than the classical theory of testing hypotheses. In the classical approach, the question is put: Which of the two hypotheses, H or H', should we adopt, on the basis of the data R? As if we were always compelled to choose one or the other of these two alternatives. Sequential analysis, on the other hand, poses the question: Are the data R sufficient ground for adopting H, or for adopting H', or are the data insufficient?
> (Barnard, 1947, p.660)

Thus, in sequential analysis the sample size is not predetermined. Instead, the experimenter continues to collect data until a decision can be reached within predetermined limits.

The advantages of the sequential procedure are twofold. First, the experimenter not only specifies a level of significance, α, but also specifies the power of the test procedure. This specification of power is otherwise unavailable to the statistician except in special cases. Second, all things being equal, the sequential procedure when compared with the traditional test shows a savings of 50 per cent in the number of observations needed to reach a decision. The economics of time and money make this a significant reason to consider this new (?) procedure. More consideration will be given to the above ideas later, but first let

me explain "new (?)" with a bit of history.

History

The formal theory of sequential analysis was developed as part of the war effort in 1943 when the principal author, Abraham Wald, was part of the Statistical Research Group of Columbia University. It was declassified in 1945, and previous publications were reprinted in Wald's book Sequential Analysis. Research into furthering the theory has been widespread since then.

Applications of sequential analysis have not been as advanced, probably because most people are satisfied with traditional procedures. The first uses of sequential analysis were in industry where the assembly line allowed for sequential quality inspection (see Davies, 1954). The sequential procedure's economy of sampling made its use all the more advantageous.

Other early users were test constructors and administrators in education who used sequential techniques for purposes of item selection (Schmid, 1952; Anastasi, 1953) and grading (Cowden, 1946; Moonan, 1950). After disappearing for some time in this area, it resurfaced in connection with computer-assisted instruction (Olivier, 1973).

The mid-fifties showed its introduction into medical research.

> This investigation used, for the first time in clinical research, the statistical technique of sequential analysis, which enables clear-cut conclusions to be reached with, on the average, the smallest number of experimental subjects.
>
> (Kilpatrick & Oldhan, 1954, p.1391)

Its continued use (Armitage, 1954; Sainsbury & Lucas, 1959; Hajnal, 1960) was fostered by the medical ethics of not using a test drug that was not helpful and possibly harmful. The advantage of sequential analysis was highlighted here.

> The ability of the pharmaceutical chemists to synthesize compounds of potential therapeutic value in psychiatry is not matched by a corresponding ingenuity in the methods for their clinical evaluation... A need has therefore arisen to develop methods of clinical trial which are rapid and which use simple but reliable criteria of therapeutic effect...
>
> We therefore had two complementary purposes in undertaking a clinical trial of a new tranquillizer. The first was to see if the statistical method of sequential analysis could be successfully used in a psychiatric drug trial of this kind; if so, the advantages of simplicity in design and of economy in time and numbers of patients treated might be considerable.
>
> (Sainsbury & Lucas, 1959, p.737)

Sequential analysis grew even closer to psychology as fields related on all sides took up its use (in

small doses). Physiology used it in radiation studies (Kimball et al., 1957; Garb, 1961; Doubravsky et al., 1964); sociology used it in its survey techniques (Peel & Skipwort, 1970); and even psychiatry used it to determine the direction of a patient's progress (Stroebel & Glueck, 1970).

Actually, the first reference to sequential techniques in the psychological literature came early in a paper entitled "Sequential analysis in psychological research" (Fiske & Jones, 1954). The authors of this paper expressed optimism about the procedure.

> Sequential analysis is not presented adequately in the statistical tests most commonly consulted by psychologists and has rarely been used in psychological research. Yet it is a statistical method that has the important advantage that it minimizes the average number of observations required to reach a specific statistical decision...
>
> (Fiske & Jones, p.264)

However, little if any psychological research has used the techniques presented by Fiske and Jones. The only use of sequential analysis in psychology of which I am aware was a discrimination experiment involving rats (Chisum, 1965). In fact, Fiske and Jones did not even mention the sequential t-test.

Theory

Before discussing the sequential procedure, let us

look at the theory of the general test procedure presently used. The principal sources for this presentation are Wald (1947), Johnson (1961), and Wetherill (1966).

The General Test Procedure

In order to test a statistical hypothesis we sample a fixed number of observations, n, and apply a particular test procedure. This procedure is a rule which states whether the hypothesis should be accepted or rejected for each possible set of samples. This rule defines a "critical region" of rejection of the hypothesis. For example, given that we want to test the hypothesis that the mean of a population equals some fixed value μ_0, the test procedure tells us to reject if

$$|\bar{X} - \mu_0| \geq c$$

and accept if

$$|\bar{X} - \mu_0| < c$$

where $\bar{X}$ is the sample mean and c is some suitably chosen numerical constant.

The choice of a critical region should be guided by the following considerations. Once a critical region is fixed, the probability of Type I error (α) and the probability of Type II error (β) are uniquely determined.[1]

---

[1] A Type I error is rejecting a true null hypothesis. A Type II error is accepting a false null hypothesis.

However, it is impossible to make both α and β arbitrarily small for some fixed value of n. So, by tradition, we choose the sample size n and a level of significance α. We then want that critical region which makes β a minimum in order to maximize the "power" of the test procedure. Neyman and Pearson (1936) have shown that a most powerful region for testing H_o against an alternative hypothesis H_1 is that region consisting of all independent samples $(x_1, \ldots, x_n)$ which satisfy the inequality

$$\frac{f(x_1,h_1)\ f(x_2,h_1)\ \ldots\ f(x_n,h_1)}{f(x_1,h_o)\ f(x_2,h_o)\ \ldots\ f(x_n,h_o)} \geq k,$$

where $f(x,h_i)$ is the underlying probability density function of the random variable x given the hypothesis H_i, and k is a constant chosen so that the region will have the required size α. (For an example, see Appendix A.)

The Sequential Test Procedure

As stated previously, the sequential test procedure differs from the traditional approach in that the sample size is not predetermined but, instead, is a random variable which depends on the outcome of the observations. The sequential analysis proceeds in a stepwise manner with one of three decisions being made at each stage

of the sampling. They are:

1) accept the hypothesis H;

2) reject the hypothesis H;

3) continue sampling.

If we take the set of all possible samples of size n and divide it into three regions such that

R_n^o, accept H;

R_n^1, reject H;

R_n, continue sampling;

the sequential test procedure is defined when these three regions are defined. Before looking at this, let us look at some of the properties of a sequential test procedure.

The operating characteristic function (OC). After a particular sequential test has been adopted, the probability that the hypothesis H will be accepted depends only on the distribution of the experiment's dependent variable x. It is assumed that the distribution is known except for a finite number of parameters λ_1 through λ_k. These can be called a parameter vector $\underline{\lambda} = (\lambda_1, \ldots, \lambda_k)$. Since the distribution of x is determined by $\underline{\lambda}$, the probability of accepting H will be a function of $\underline{\lambda}$, the operating characteristic function, $L(\underline{\lambda})$. If only one parameter is unknown, then $L(\underline{\lambda})$ is a curve which can be plotted.

The OC function describes what a sequential test procedure accomplishes. For any value of $\underline{\lambda}$, the probability of making a correct decision can be obtained from the OC function. For values of $\underline{\lambda}$ acceptable under the hypothesis $\underline{H}$, the probability of a correct decision equals $L(\underline{\lambda})$. For values of $\underline{\lambda}$ unacceptable under the hypothesis $\underline{H}$, the probability of a correct decision equals $1-L(\underline{\lambda})$. It follows that the OC function is considered more favorable the higher the value of $L(\underline{\lambda})$ for acceptable $\underline{\lambda}$ and the lower the value of $L(\underline{\lambda})$ for unacceptable $\underline{\lambda}$. However, the closer to the ideal OC function one gets, that is, the smaller the values of $\underline{\alpha}$ and $\underline{\beta}$, the larger will be the number of observations required to make a decision. To this we now turn our attention.

The average sample number function (ASN). Since the number of observations is a random variable, it would be helpful to investigate the value of the average number of samples needed to reach criterion, E(n). That value, E(n), depends on the distribution of $\underline{x}$ which, previously stated, is known except for the parameter vector $\underline{\lambda}$. Therefore, E(n) is a function of $\underline{\lambda}$. This is called the average sample number function, which is a curve that can be plotted if only one parameter is unknown. Thus, given the ASN function, we may deter-

mine how many observations will be needed (on the average) given a particular sequential test.

These properties allow us to develop guidelines for choosing a sequential test. First, we decide on the values of α and β, which is analogous to deciding what OC functions are allowable. Then, given the subset of sequential procedures which fit these required probabilities, we select that sequential test which minimizes the required number of observations (taken from the ASN functions).

Now let us turn our attention to an optimum sequential test procedure.[2]

The Sequential Probability Ratio Test (SPRT)

Abraham Wald's sequential probability ratio test is designed to choose between two simple hypotheses. It is from this procedure that the sequential t-test is eventually derived.

Given a sample $(x_1, x_2, \ldots, x_n)$ the probability of the sample given H_1 is

$$p_{1n} = f(x_1, h_1) \ldots f(x_n, h_1).$$

The probability of the sample given H_o is

$$p_{on} = f(x_1, h_o) \ldots f(x_n, h_o).$$

---

[2] Wald does not make clear why the test procedure that follows is optimum. The procedure, however, is analogous to an optimum traditional test procedure.

The SPRT is defined as follows. Two positive constants $\underline{A}$ and $\underline{B}$ ($\underline{B} < \underline{A}$) are chosen. At each stage of the experiment a likelihood ratio is computed. Continue as long as

$$B < \frac{p_{1n}}{p_{on}} < A.$$

Reject H_o (accept H_1) if

$$\frac{p_{1n}}{p_{on}} \geq A.$$

Accept H_o if

$$\frac{p_{1n}}{p_{on}} \leq B.$$

The choice of values for $\underline{A}$ and $\underline{B}$ is dictated by the values of α and β. For simplification take logs of the inequality. This allows for the addition of the new observation ratio at each stage of the experiment. Thus,

$$\log \frac{p_{1n}}{p_{on}} = \log \frac{f_{11}}{f_{01}} + \log \frac{f_{12}}{f_{02}} + \ldots.$$

Using the above criterion, the values of $\underline{A}$ and $\underline{B}$ have been found to be

$$A = \frac{(1-\beta)}{\alpha} \; ; \; B = \frac{\beta}{(1-\alpha)} .$$

To see how these values for $\underline{A}$ and $\underline{B}$ were found consider the following (for details see Wald, pp.40-42).

Table 1 shows the four possible outcomes of any sequential test. One rule for termination states

$$\frac{p_1(x_1, \ldots, x_n)}{p_o(x_1, \ldots, x_n)} \leq B$$

or

$$p_1(x_1, \ldots, x_n) \leq B\, p_o(x_1, \ldots, x_n).$$

The sum of the probabilities of obtaining sets of results $(x_1, \ldots, x_n)$ given H_o is true, summed over all sets of results which terminate in a decision in favor or H_o is equal to the probability of deciding for H_o when H_o is true. Denote this by

$$\sum_{H_o} p_o(x_1, \ldots, x_n) = 1 - \alpha .$$

Similarly,

$$\sum_{H_o} p_1(x_1, \ldots, x_n) = \beta .$$

Substituting into the above inequality, we get

$$\sum_{H_o} p_1(x_1, \ldots, x_n) \leq B \sum_{H_o} p_o(x_1, \ldots, x_n).$$

Substituting again

$$\beta \leq B\,(1 - \alpha)$$

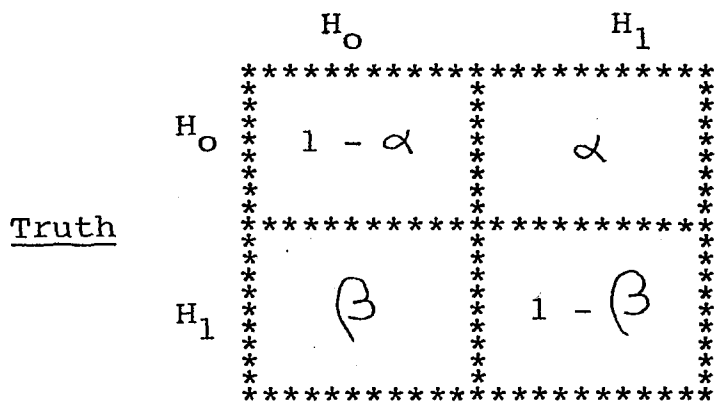
which becomes

$$\frac{\beta}{(1 - \alpha)} \leq B.$$

The same line of thinking is used to show the other inequality.

Table 1

Given any sequential test the following four outcomes may occur with the following associated probabilities.


Test Decision
H_o
H_1
Truth
H_o
$1 - \alpha$
α
H_1
β
$1 - \beta$

Since sampling is stopped when one of the limits A or B is passed, in inequalities are very nearly equalities. Given the overshoot in a sequential test, that is, termination with a value greater than A or less than B, Wald has shown that no appreciable increase in α or β will result and, at most, one of these quantities will be increased (for details see Wald, pp.44-47).

The inequalities have been derived under the assumption that the sequential procedure terminates with probability equal to one. Wald has shown this (see Wald, pp.157-158), but only after using the additional assumption of independence of observations. This last assumption is not true for many sequential tests. Therefore, it is necessary to show that a sequential procedure with dependent observations terminates with probability equal to one (for a detailed proof see David & Kruskal, 1956).

The Sequential t-Test

We now turn our attention to a problem of chief interest in this paper: testing the mean of a normal distribution given that the variance is unknown (a sequential t-test). However, before the theory of the sequential probability ratio test can be applied to this situation, two characteristics of the SPRT must be satisfied: the hypotheses tested are simple (in

contrast to the composite hypotheses found in a t-test) and all parameters except the one to be tested are known. The t-test, however, is used when the hypotheses are composite and the variance is unknown.

Wald's solution to this dilemma is to weight the simple hypotheses included in a given composite hypotheis and ascribe a prior distribution to the undefined ("nuisance") parameter. This solution is not unique but only Wald's attempt to evaluate the likelihood ratio. In fact, Wald never makes explicit why he has chosen this approach. It will be presented here as only one method of developing a sequential t-test.[3]

The possible outcomes (sample space) for the unknown parameters are divided into three regions:

w_o: H_o preferred,

w_1: H_1 preferred,

remainder: indifference.

We construct two weighting functions, $W_i(\theta)$ $(i=0,1)$, for the sample space such that

$$\int_{w_o} W_o(\theta)\, d\theta \quad = \quad \int_{w_1} W_1(\theta)\, d\theta \quad = \quad 1.$$

[3]For a criticism of the weight function approach see Barnard (1947). Cox (1952) transforms the data so that it no longer depends on the unknown variance and, therefore, allows for evaluation of the likelihood ratio. Both methods, however, lead to similar approximations of the likelihood ratio.

Now consider two hypotheses that the probability of a set of observations $(x_1, \ldots, x_n)$ equals

$$f(x_1,\theta)f(x_2,\theta)\ldots f(x_n,\theta)\ W_i(\theta)\ d\theta \qquad \text{for } i=0,1.$$

This, with the addition of a prior distribution for the unknown variance σ^2, will allow for the construction of a SPRT with the desired values of α and β.

The weight functions which allow α and β to take their desired values are:

$$W_o(\theta) = 1 \quad \text{when } \theta = \theta_o$$
$$= 0 \quad \text{otherwise}$$

$$W_1(\theta) = \tfrac{1}{2} \quad \text{when } \theta = \theta_o \pm \delta\sigma$$
$$= 0 \quad \text{otherwise.}$$

To allow for the fewest assumptions about the distribution of σ, a rectangular distribution is used. It is:

$$\phi(\sigma) = 1/c \quad 0 \le \sigma \le c$$
$$= 0 \quad \text{otherwise.}$$

The modified likelihood ratio with the limit $c \to \infty$ is:

$$\frac{\frac{1}{2}\left[\int_0^\infty \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\theta_o-\delta\sigma)^2\right\} d\sigma + \int_0^\infty \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\theta_o+\delta\sigma)^2\right\} d\sigma\right]}{\int_0^\infty \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\theta_o)^2\right\} d\sigma}$$

In essence, this procedure integrates out the nuisance parameter and allows the likelihood ratio to be evaluated. (For a complete derivation of the above, see Wald, pp.80-84.) It should be noted, however, that since the variance is unknown, the OC and ASN functions cannot be calculated for the sequential t-test.

Delta

One of the characteristics of the likelihood ratio for the sequential t-test is the need to specify the parameter delta (δ). This parameter is defined as μ/σ, and can be interpreted as the size of the effect to be detected. But before attempting to go further, let us look at the reason we are using such a statistical test.

The sort of experiment that is often being attempted when using the t-test is the following. Two groups of subjects are given two different treatments (for instance, an experimental group and a control group), and we wish to ascertain whether the two groups differ on the effects of the treatment of interest. Statistical tests are used in order to answer this question since random differences in subjects often lead to differences in numbers that are of no real consequence (at least, as far as the experimental treatment is concerned).

The t-Test

Traditionally, we specify the number of subjects we wish to run and the level of significance, α. We have seen previously that, although incalculable, β is minimized for that number of observations. What is often ignored is that when specifying the number of observations, the size of the effect that can be detected (in terms of the unknown variance) is also set. Table 2 shows this for various values of n.[4]

What this means in practice is that if the true size of the effect is larger than the critical value that the test can detect, the probability of detection increases over the preassigned value of β. However, if the true size is smaller than the critical value, the probability of detection decreases. This loss of power may mean that a scientifically interesting effect will be ignored.

> This brings our analysis to a very important realization: acceptance of the null hypothesis will often occur even when $\mu_1 < \mu_0$ but when the difference $\mu_0 - \mu_1$ is small. That is why in cases of statistically not significant results the null hypothesis can be considered as true only provisionally, i.e. as a working hypothesis. We therefore

[4] These can be determined by a power analysis of the t-test for fixed sample size. This particular table comes from Armitage (1954, p.267).

Table 2

Sensitivity of the t-Test with Paired Observations

A test of the mean of n differences has a probability of .95 of yielding a significant result (at the $p=.05$ level) when the true mean difference is a multiple δ of the true standard deviation of the difference.

Number of Pairs (n)	Mean diff./s.d. of diff. "δ"
10	1.29
15	1.00
20	0.85
30	0.68
50	0.52
100	0.36
200	0.26
1000	0.11

> can obtain a statistically not significant result when in fact, the alternative hypothesis is true which, however, is only a little different from the null hypothesis. This erroneous conclusion will be formed more frequently when fewer experimental objects are used in the experiment. The reverse is also true, i.e. the more objects there are in the experiment the smaller may be the difference from the null hypothesis which will be detected.
>
> (Doubravsky et al., 1964, p.97)

No strict guidelines have been given for selecting the number of observations we should take except to say "the more, the better". This has two problems, however, both stemming from ignoring the critical size of the effect that can be detected. First, as mentioned above, we may screen out a difference that is important. But as serious, we may make a t-test so sensitive that a statistically significant difference may appear that, in fact, is of no consequence to the experimenter. The treatment effect may be so small as to be scientifically unimportant and uninteresting.

The Sequential t-Test

As mentioned previously, by using a sequential analysis we may specify before the experiment not only α, but also β which allows us to safeguard against a low powered experiment. But this makes the number of observations needed for a decision a random variable whose expectation (average) is determined by the value

of δ. And here is a crucial difference between the traditional test and the sequential test. Before taking the first observation, we must specify what size effect we wish to detect. Since this has been largely ignored, there appears at first to be little basis for that decision. However, various guidelines can be used.

Approximating Delta

Considering the history of sequential analysis and its early use in industry, it is no surprise that one method for approximating delta is practical considerations. For example, managers and technicians may be able to decide what tolerance levels are required for their products on the assembly line. However, this method offers little to the scientific researcher. Theoretical considerations may show more promise. If some theory or model exists then a precise prediction may be generated. Unfortunately, this is not always the case in psychology. If previous investigations of the experiment in question have been made, then an empirical estimate may be possible, particularly if the results have been somewhat steady. And of course, a combination of the above is possible as will be shown later.

Mention should be made here that taking an "educated guess" at a value may not be as crude as it ap-

pears. It seems no worse than arbitrarily fixing the sample size in the traditional procedure. In fact, Armitage (1954) suggests "it may be useful to refer to" the table of critical values of $\underline{\delta}$ for the traditional test (Table 2, p.21). What he appears to be suggesting is that the experimenter decides how many subjects he would run given the traditional test and use that value of $\underline{\delta}$. This seems an interesting way to avoid many of the arguments that might come from those who still adhere to the traditional methods. That "educated guess" method has in fact been used (Sainsbury & Lucas, 1959).

Implementation

The likelihood ratio derived by Wald for the sequential t-test has been approximated in various ways. In this section I will briefly discuss two of these ways and show how they are used to perform a sequential t-test.

NBS Tables

Arnold (1951) in his evaluation of the likelihood ratio noticed the similarity of the likelihood ratio to the confluent hypergeometric function. Using various transformations, he showed that the original inequalities in the form

$$\frac{\beta}{(1-\alpha)} < \frac{p_{1n}}{p_{on}} < \frac{(1-\beta)}{\alpha}$$

could be written in the following way:

$$L_L < z_n < L_U$$

where

$$z_n = \frac{\left[\sum (x_i - \theta_o)\right]^2}{\sum (x_i - \theta_o)^2}$$

and the boundaries L_L and L_U are solutions to the confluent hypergeometric function given the parameters α, β, δ, and n (for details see Arnold, pp.v-vii). The National Bureau of Standards Tables prepared by Arnold list these solutions given various values of the above parameters.[5] Therefore, to perform a sequential t-test given the above approximation, the following steps are taken:[6]

1) choose values for the parameters α, β, and δ;

2) at each stage of the experiment calculate z_n;

3) find the values of L_L and L_U in the NBS tables for the given value of n; interpolate if necessary;

---

[5]Arnold points out that the likelihood ratio evaluated is slightly different from Wald's, but this difference is of no consequence for our purposes.

[6]Note that this is a two-sided t-test. No values are listed for a one-sided test in the NBS tables, and I know of no other tables that contain values for a one-sided test.

4) if $z_n \leq L_L$, sampling is terminated and the hypothesis that the mean equals θ_o is accepted with probabilities α and β;

5) if $z_n \geq L_U$, sampling is terminated and the hypothesis that the mean equals θ_o is rejected with probabilities α and β;

6) if step 4 or step 5 does not hold, continue sampling and repeat the procedure (step 2).

Rushton's Approximation

Rushton (1950) used the similarity with the confluent hypergeometric function to directly approximate the likelihood ratio. His method allows for the calculation of the likelihood ratio for any δ and n. This calculation is compared with the original boundaries

$$A = \frac{(1-\beta)}{\alpha} \quad \text{and} \quad B = \frac{\beta}{(1-\alpha)} .$$

To perform this sequential t-test the following steps are taken:[7]

1) choose values for the parameters α, β, and δ;

2) calculate

$$\ln A = \ln \frac{(1-\beta)}{\alpha}$$

(this is the upper boundary for all n);

[7] This is a one-sided test where H_o states that the mean equals zero. See Rushton (1952) for the two-sided case.

3) calculate

$$\ln B = \ln \frac{\beta}{(1 - \alpha)}$$

(this is the lower boundary for all n);

4) calculate

$$u = \Sigma x / (\Sigma x^2)^{\frac{1}{2}}$$

at each stage of the experiment;

5) calculate the first approximation to the log of the likelihood ratio, l_1:

$$l_1 = (n)^{\frac{1}{2}} \delta u - \tfrac{1}{2} n \delta^2 + \tfrac{1}{4} \delta^2 u^2 ;$$

6) when one of the boundaries is approached (relative to the opposite boundary), calculate the second approximation, l_2:

$$l_2 = (n)^{\frac{1}{2}} \delta u \left\{ 1 - \frac{1}{(4n)} \right\} - \tfrac{1}{2} n \delta^2 + \tfrac{1}{4} \delta^2 u^2$$

which can be obtained from the first by subtracting $a_1 = \left[(n)^{\frac{1}{2}} \delta u \right] / (4n)$ from the first so that

$$l_2 = l_1 - a_1 ;$$

7) if $l_2 \leq \ln B$, sampling is terminated and the hypothesis that the mean equals zero is <u>accepted</u> with probabilities $\leq \alpha$ and β;

8) if $l_2 \geq \ln A$, sampling is terminated and the hypothesis that the mean equals $\delta\sigma$ is <u>accepted</u> with probabilities $\leq \alpha$ and β;

9) for further precision in the approximation, calculate the third approximation, l_3:

$$l_3 = (n)^{\frac{1}{2}} \delta u \left\{ \left[1 - \frac{1}{(4n)} + \delta^2 u^2 \right] / (24n) \right\} - \tfrac{1}{2} n \delta^2 + \tfrac{1}{4} \delta^2 u^2$$

which can be obtained from the second by adding $a_2 = \frac{2}{3} a_1 \tfrac{1}{4} \delta^2 u^2$ to the second approximation so that

$$l_3 = l_2 + a_2;$$

10) if neither boundary is crossed, continue taking observations and repeat the procedure (step 4).

This seemingly complicated procedure is simplified by tabulating the calculations as will be shown later.

Graphing

It is possible to use a graphing procedure to accomplish the NBS-approximation t-test. For instance, the boundaries from the NBS table could be plotted on the ordinate of a graph (with n on the abcissa) before any observations are taken. Then, as each observation is taken, the resulting value of the test statistic, z, could be plotted and joined with a straight line to the previous point. As long as the line is within the "channel" formed by the boundaries, observations are continued. When one of the boundaries is crossed, the sampling is completed and the appropriate decision made depending on which region is entered.

The advantage of this method is the visual picture presented of the experiment's progress. However, pre-

cise graphing techniques should be used since the possibility of error is greater with this method.

The Sequential Sign Test

The sequential t-test is one approach to testing the hypothesis that the mean of a normal distribution with unknown variance is zero. The sequential sign test, analogous to the traditional sign test, is another alternative.[8] This test, in short, converts the difference scores into probabilities which can be tested using a sequential binomial distribution test (see Wald, chapter 5).

In considering the difference between two groups, the size of the difference between pairs is ignored, and only the sign of the difference is used. Given this way of looking at the data, the hypothesis that the mean equals zero is transformed into the hypothesis that the probability of an observation being a minus is ½, that is,

$$p(-) = \tfrac{1}{2}.$$

The hypothesis that the mean equals $\delta\sigma$ (the alternative hypothesis) is transformed into the hypothesis that the probability of a minus equals

$$p_1 = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\delta}^{\infty} e^{-\frac{1}{2}x^2}\, dx.$$

[8] Note that this is a one-sided test where the null hypothesis states that the mean equals zero. See Armitage (1947) for the two-sided case.

The test is defined by Armitage (1947) with the following procedure:

1) choose values of α, β, and δ;

2) find p_1 from a table of the normal distribution;

3) calculate the following:

$$A_1 = \frac{\log \frac{(1-\beta)}{\alpha}}{\log 2(1-p_1)}$$

$$A_2 = \frac{\log \frac{(1-\alpha)}{\beta}}{\log 2(1-p_1)}$$

$$B = \frac{\log \frac{1}{2p_1}}{\log 2(1-p_1)}$$

4) start a score at A_2 units;

5) if an observation is a "+", add 1 unit to the score;

6) if an observation is a "-", subtract $\underline{B}$ units from the score;

7) if the score reaches zero, terminate sampling and accept H_o;

8) if the score reaches A_1+A_2, terminate sampling and reject H_o;

9) if neither of the above, repeat the procedure (step 5).

This procedure has the advantage of being very easy to administer once the initial calculations are completed. However, since the size of the difference between pairs is ignored and only the sign of the difference is used, a loss of information occurs. The result is an added number of observations needed to reach a decision compared to a sequential t-test set at the same α and β levels. This is in contrast to the traditional sign test. Its loss of information usually results in a loss of power when compared to a traditional t-test set at the same α level with the same sample size.

A Demonstration

In order to demonstrate the sequential test procedure and help evaluate its usefulness, part of a previous experiment (Sunday & Kay, unpublished manuscript) was replicated except that the analysis was performed sequentially. A brief account of the experiment follows.

The Experiment: A Summary

Rock's (1957) interpretation of his data as evidence for all-or-none learning led to repeated criticism of his drop-out method for paired associate learning (e.g. Postman, 1962; Underwood, Rehula, & Keppel,

1962; Cohen & Murray, 1968). One criticism suggested by Sunday and Kay was that the serial position effect in paired associate learning (Hovland, 1938) was minimizing the differences between the experimental group (drop-out condition) and the control group (drop-none condition). They proposed to restrict the positions of the test items by only shuffling stimuli within blocks between each study and presentation trial. Hopefully, this procedure would minimize the serial position effect. Their results, however, continued to support the all-or-none interpretation.

One further criticism of the Rock experiment is that the design is possibly a low powered one and cannot produce a large-enough difference for detection. In fact, no one has attempted to predict what degree of difference is to be expected given the incremental theory. (The all-or-none theory predicts no difference.) For these reasons, the sequential precedure was seen as ideal for this design. First, the analysis could be made as high powered as necessary (without affecting the significance level, α). But even more interesting, the necessity to specify a value for δ, dictated that a prediction be made for the two opposing theories.

Method. Stimuli consisted of letters A through Z excluding I and AA through ZZ excluding II randomly

paired with the numbers 11 through 60. Each paired associate was printed on a 10 x 6 cm card.

Five study-test trials consisting of 10 stimuli each were given to each subject. The stimuli were divided into three blocks with block I and block III consisting of the first three and last three stimuli positions respectfully. Block II consisted of stimuli positions 4 through 7. Between each study-test trial, the stimuli were shuffled within their blocks.

In the drop-errors condition (DE) the errors in block II were replaced after each trial. In the drop-none condition (DN) no stimuli were replaced. The stimuli in blocks I and III were replaced after each trial for both conditions. Since subjects were not treated differently until the beginning of trial 2, the dependent variable was the difference between the number of correct responses for a subject in the DN condition and the number of correct responses for a subject in the DE condition (DN-DE) in block II on trials 2-5. Thus, subjects were paired by when they participated in the experiment. This results in an equal number of subjects in each condition.

The study trials consisted of 3 seconds for each stimulus and an interstimulus interval of 3 seconds. The test trials consisted of 5 seconds per stimulus

for recall with zero seconds in between stimuli. The intertrial interval was 40 seconds.

Subjects were undergraduates enrolled in the introductory psychology course at Lehigh University. They were alternately placed in either condition and tested individually.

Hypotheses. The null hypothesis states that there is no difference between groups (μ = DN-DE = 0). Acceptance of this hypothesis favors an all-or-none interpretation. The alternative hypothesis states that the control group (DN) will make more correct responses than the experimental group (DE), that is, μ = DN-DE = $\delta\sigma > 0$. Acceptance of this hypothesis is evidence for an incremental interpretation.

The analysis. The following sequential test procedures were used after each pair of observations:

1) a two-sided sequential t-test using the NBS approximation;

2) a one-sided sequential t-test using the Rushton approximation;

3) a one-sided sequential sign test.

All sequential procedures were set at the .05 level of significance (α = .05) and 95 per cent power (β = .05). The value of δ calculated in the next section was used in all three tests. Observations were taken until

all three sequential tests reached a decision, although each test was terminated by its own decision rule.

Calculating Delta

The following procedure for calculating delta is a combination of the empirical and theoretical approaches. Although complicated, it was considered to be the most accurate method for setting this parameter.

The data from the previous experiment were used in the empirical section of the procedure. The number of correct paired associates in the first trial was used in estimating the learning parameter c. This is the probability of a correct response given the initial presentation of that stimulus. Its value was .30337.

Now we derive a prediction of the results of this particular experiment given the learning parameter c. The prediction of the all-or-none theory is straightforward: no difference between the two experimental groups ($\delta_{dif.} = 0$). The incremental theory, however, is a more complicated matter.

The incremental model used in this derivation has the following assumptions. If a stimulus was correct on trial n, the probability of it being correct on trial n+1 is one. This assumption was made on the basis of informal observation in previous experiments of a

similar nature (for example see Rock, 1957, p.190). If a stimulus was incorrect on trial n, the probability of it being correct on trial n+1 is

$$p(n) + c(1 - p(n))$$

where p(n) is the probability of a correct response on trial n. This model is based on the single-operator linear model of Bush and Mosteller (see Atkinson, Bower, & Crothers, 1965). In the present case, however, the learning parameter c was calculated with a different procedure. Bush and Mosteller estimated c from the observed mean total errors per subject-item. The present experiment used the total number of correct responses on the first trial only from a previous experiment because the latter information is unavailable due to the nature of the procedure. In theory this is the same approximation.

The above model is used to predict the performance of a subject on trials 2-5 of block II in the DN condition. The probability of a correct response on each trial assuming an effect due to practice may be expressed mathematically

$$p(n) = c \sum_{i=1}^{n} (1 - c)^{i-1}.$$

In the DE condition any stimulus that the subject is given on trial n+1 that was shown on trial n must

have been correct on trial n since all errors in block II had been replaced. Therefore, the probability of the repeated stimulus being correct again is one (as assumed previously). All replacement stimuli have the probability of being correct equal to the learning parameter c since these stimuli are being seen for the first time.

The expected difference between the two conditions given the incremental model may be quantified by constructing a probability distribution for each condition and comparing their parameters. These distributions are of the random variable given by the number of correct responses in block II for trials 2-5, which can vary from zero to 16. An example of these calculations is given in Appendix B. The distributions are shown in Table 3.

The expectation, E(x), and the variance, σ^2, of each distribution can be determined in a straightforward manner. Finally, delta is calculated from that information by the following:

$$\underline{\delta} = \frac{\mu_{DN} - \mu_{DE}}{(\sigma_{DN}^2 + \sigma_{DE}^2)^{\frac{1}{2}}}$$

which is from the original definition of delta, $\underline{\delta} = \mu_{dif}/\sigma_{dif}$. The resulting value of $\underline{\delta}$ equals .86003 (see Appendix C for calculations).

Table 3

The probability of n correct responses on block II for trials 2-5 given the incremental theory with c = .30337.

Drop Errors p(n)	n	Drop None p(n)
.00073	0	.00000
.00126	1	.00000
.00264	2	.00000
.00521	3	.00000
.01525	4	.00001
.01986	5	.00003
.03122	6	.00013
.04741	7	.00053
.08448	8	.00193
.08179	9	.00614
.10139	10	.01699
.12053	11	.04196
.15384	12	.08894
.09352	13	.15857
.09038	14	.23301
.08029	15	.25960
.07018	16	.19197

Results and Discussion

The results of the NBS approximation t-test are shown in Table 4. Sampling was terminated on the 16 th trial (32 nd subject); the decision was to accept the null hypothesis.

The results of the Rushton approximation t-test are shown in Table 5. Sampling was terminated on the 6 th trial (12 th subject); the decision was to accept the null hypothesis.

The results of the sequential sign test are shown in Table 6. Sampling was terminated on the 8 th trial (16 th subject); the decision was to accept the null hypothesis.

The evidence sides with the all-or-none learning theory since there is no apparent effect due to practice. One possible source of error, however, is that the effect of practice may be quite small and the calculation of delta was in error (i.e. it may be smaller). In that case, the present tests would not be sensitive enough to detect the difference. (If the true delta is much smaller, then the probability of missing that effect would be much larger than .05.) However, until some other procedure is developed to magnify the hypothesized difference, the all-or-none interpretation will stand.

Table 4

The NBS approximation t-test on the difference, $\underline{x}$, between a pair of subjects with one in each experimental condition using the NBS boundaries L_L and L_U.

pair #	x	Σx	$(\Sigma x)^2$	x^2	Σx^2	L_L	$(\Sigma x)^2/\Sigma x^2$	L_U
1	0	0	0	0	0	-	-	-
2	+1	+1	1	1	1	-	1.0000	-
3	-2	-1	1	4	5	-	.2000	-
4	-1	-2	4	1	6	-	.6667	-
5	+4	+2	4	16	22	-	.1818	-
6	-9	-7	49	81	103	-	.4757	5.5066
7	-5	-12	144	25	128	-	1.1250	5.4542
8	-9	-21	441	81	209	-	2.1100	5.4471
9	-2	-23	529	4	213	-	2.4836	5.4728
10	+2	-21	441	4	217	.2527	2.0323	5.5220
11	+10	-11	121	100	317	.3780	.3817	5.5882
12	-3	-14	196	9	326	.5070	.6012	5.6690
13	-3	-17	289	9	335	.6384	.8627	5.7602
14	-6	-23	529	36	371	.7728	1.4259	5.8604
15	+4	-19	361	16	387	.9101	.9328	5.9672
16	+7	-12	144	49	436	1.0487	.3303 *	6.0791

* accept H_o with $\alpha = .05$ and $\beta = .05$.

See Appendix D for the interpolation of the boundary values. Boundaries not listed in this table are not found in the NBS tables (Arnold, 1951).

Table 5

The Rushton approximation t-test on the difference, x, between a pair of subjects with one in each experimental condition.

n	x	Σx	Σx^2	$(\Sigma x^2)^{\frac{1}{2}}$	u	$\frac{1}{4}\delta^2 u^2$	$(n)^{\frac{1}{2}}\delta u$	$\frac{1}{2}n\delta^2$	1_1
1	0	0	0	-	-	-	-	-	-
2	+1	+1	1	1.0000	+1.0000	.1849	+1.2162	.7396	+.6615
3	-2	-1	5	2.2361	-.4472	.0370	-.6661	1.1094	-1.7385
4	-1	-2	6	2.4495	-.8165	.1233	-1.4044	1.4792	-2.7603
5	+4	+2	22	4.6904	+.4264	.0336	+.8200	1.8490	-.9954
6	-9	-7	103	10.1489	-.6897	.0880	-1.4529	2.2188	-3.5837 *

*accept H_o with α = .05 and β = .05.

Note: 1_1 = col. 6 + col. 7 + col. 8

See Appendix E for calculation of boundaries.

Table 6

The sequential sign test on the difference, x, between a pair of subjects with one in each experimental condition.

pair #	x	sign	score
start	-	-	6.1811
1	0	+	7.1811
2	+1	+	8.1811
3	-2	-	6.2033
4	-1	-	4.2255
5	+4	+	5.2255
6	-9	-	3.2477
7	-5	-	1.2699
8	-9	-	-0.7079 *

* accept H_o with α = .05 and β = .05.

See Appendix F for calculation of sign test constants.

A few comments about the analysis are worth special mention. The number of observations needed to detect the difference predicted ($\underline{\delta}$ = .86) for the traditional t-test would be approximately 20 pairs of subjects (see Table 2, p.21). All three sequential tests showed an average savings of 50 per cent (see Table 7). However, in this particular experiment, only a one-tailed test would normally be used since the direction of the predicted outcome is well known. Considering this, the economy of the sequential tests is even greater.

Figure 1 shows the graphical representation of the two-sided NBS approximation t-test. This visual picture of the progress of an experiment serves to further highlight details that the tabulation method may miss. In this case, note how close to a decision the test procedure came on observations 11 and 15.

To further magnify these details, I took the square root of the boundary values. This seperated the rejection region into two parts: a rejection of the null hypothesis with a positive difference (DN-DE > 0) and a rejection of the null hypothesis with a negative difference (DN-DE < 0). I then plotted the test statistic $u = (z)^{\frac{1}{2}}$ using the sign of Σx. This graph is shown in Figure 2.

As mentioned earlier, rejection in the <u>positive</u>

Table 7

The economy of the sequential tests when compared with the traditional t-test.

Test	# observations	savings	tail
traditional	20	-	-
NBS	16	20%	two
Rushton	6	70%	one
sign	8	60%	one

Figure 1

The graphical representation of the NBS approximation t-test.

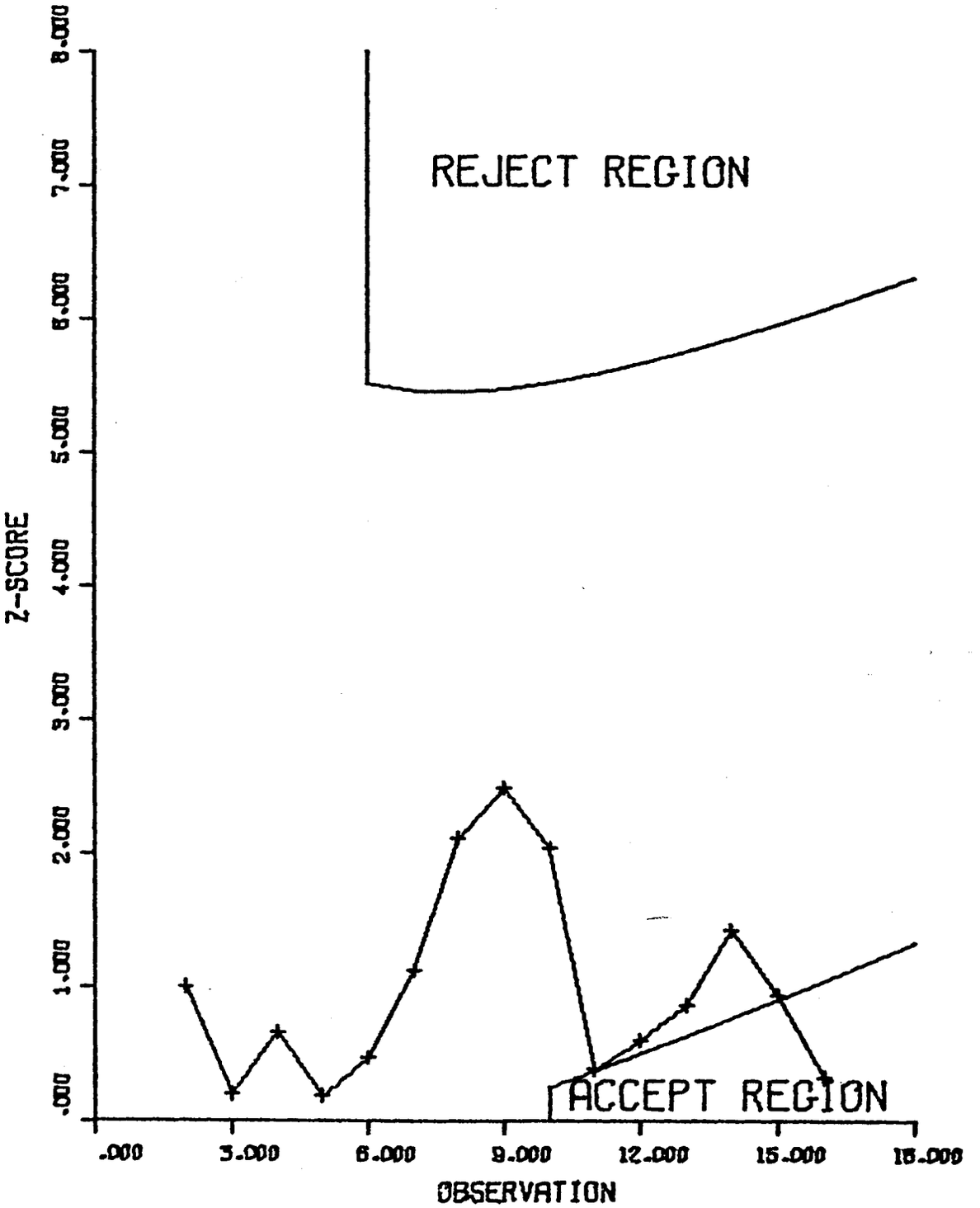
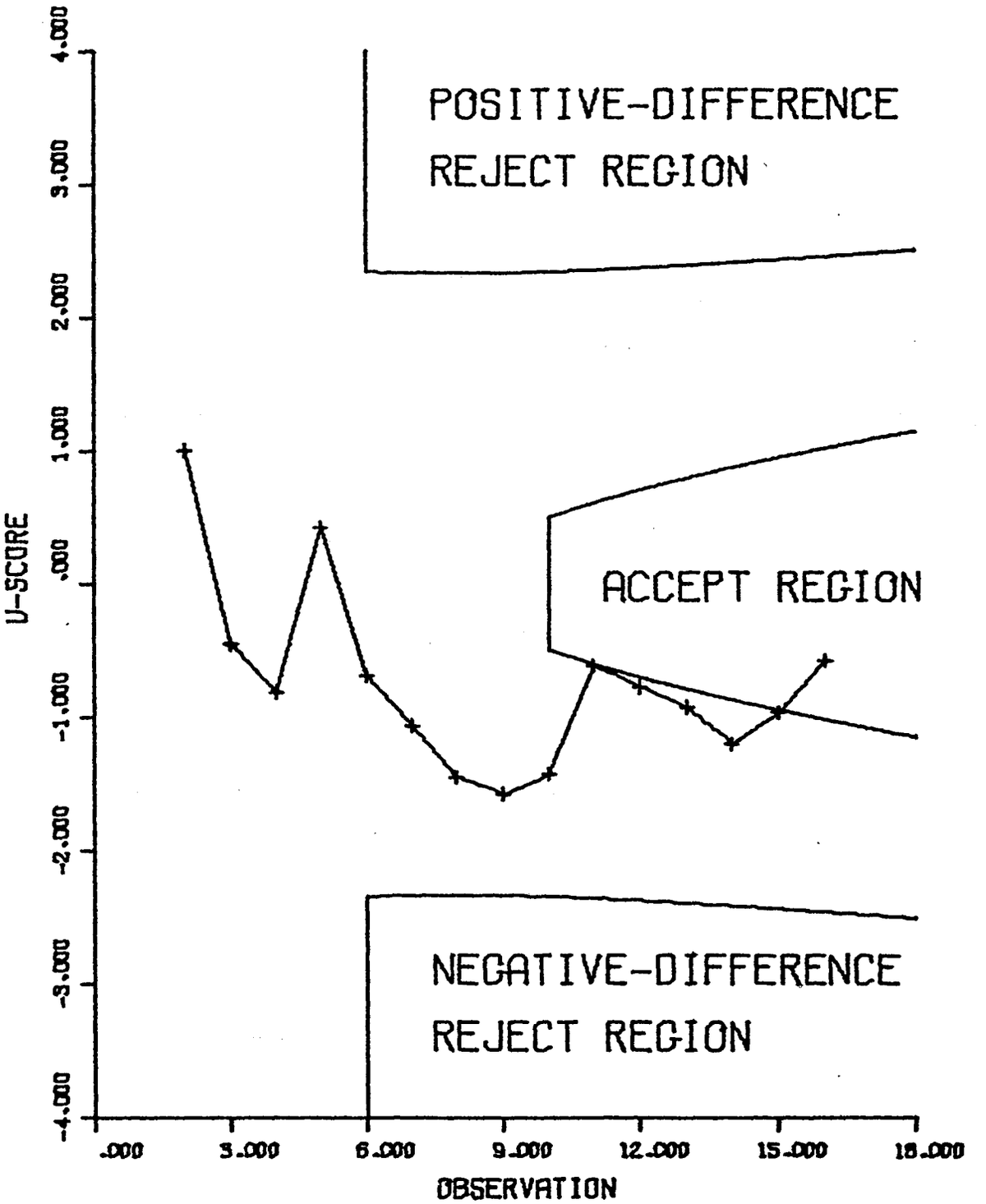

REJECT REGION
ACCEPT REGION
Z-SCORE
OBSERVATION
.000
1.000
2.000
3.000
4.000
5.000
6.000
7.000
8.000
.000
3.000
6.000
9.000
12.000
15.000
18.000

Figure 2

The graphical representation of the NBS approximation t-test after the square root transformation.


POSITIVE-DIFFERENCE
REJECT REGION
ACCEPT REGION
NEGATIVE-DIFFERENCE
REJECT REGION
U-SCORE
4.000
3.000
2.000
1.000
.000
-1.000
-2.000
-3.000
-4.000
.000
3.000
6.000
9.000
12.000
15.000
18.000
OBSERVATION

direction is evidence for the incremental theory while acceptance of the null hypothesis is evidence for the all-or-none theory. Rejection in the negative direction does not fit any proposed theory (discussed in this paper). Note in Figure 2 how the dependent variable's path enters the lower "channel". In fact, at observation 9 it is impossible to reach the upper "channel". Mathematically, it would take a value of the dependent variable (at observation 10) that is greater than 16 (its maximum). Therefore, although the sequential procedure has not terminated at observation 9, the positive-rejection boundary for termination has been eliminated. In this case, acceptance of the null hypothesis was all but certain at this observation (since rejection in the negative direction was very improbable).

Conclusions

The above demonstration clearly shows that sequential analysis is a feasible and valuable statistical tool in scientific research. It deserves an equal position with traditional statistical methods in the researcher's handbook and merits consideration in the development of experimental design. However, little is known about its compatability with psychological phenomena, and only continued use and evaluation will determine its rightful place in statistics.

Reference List

Anastasi, A. An empirical study of the applicability of sequential analysis to item selection. Educational and Psychological Measurement, 1953, 13, 3-13.

Armitage, P. Some sequential tests of Student's hypothesis. Journal of the Royal Statistical Society, Supplement, 1947, 9, 250-263.

Armitage, P. Sequential tests in prophylactic and therapeutic trials. Quarterly Journal of Medicine, 1954, 23, 255-274.

Arnold, K.J. Tables to Facilitate Sequential t-Tests National Bureau of Standards Applied Mathematics Series No. 7, 1951.

Atkinson, R.C., Bower, G.H., & Crothers, E.J. An Introduction to Mathematical Learning Theory. New York: Wiley, 1965.

Barnard, G.A. Review of A. Wald's sequential analysis. Journal of the American Statistical Association, 1947, 42, 658-664.

Chisum, G.T. Transposition as a function of number of test trials. Journal of Comparative and Physiological Psychology, 1965, 59, 419-421.

Cohen, R.L., & Murray, A. Paired-associate learning, mediation value and the drop-out method. Psychological Reports, 1968, 23, 671-674.

Cowden, D.J. An application of sequential sampling to testing students. Journal of the American Statistical Association, 1946, 41, 547.

Cox, D.R. Sequential tests for composite hypotheses. Proceedings of the Cambridge Philosophical Society, 1952, 48, 290-299.

David, H.T., & Kruskal, W.H. The WAGR sequential t-test reaches a decision with probability one. Annals of Mathematical Statistics, 1956, 27, 797-805.

Davies, O.L. The Design and Analysis of Industrial Experiments. New York: Hofner, 1954.

Doubravsky, J., Kamenicek, O., Komenda, S., & Schober, B. Sequential analysis in biological and physiological research (simultaneous study of the influence of x-ray irradiation on the gastric transit time in rat). Physiologia Bohemoslovenica, 1964, 13, 96-103.

Fiske, D.W., & Jones, L.V. Sequential analysis in psychological research. Psychological Bulletin, 1954, 51, 264-275.

Garb, J. Modification of a sequential procedure for radiation protection studies with mice. Radiation Research, 1961, 15, 616-624.

Hajnal, J. Sequential trials of analgesics in rheumatoid arthritis. In H. de Jonge (Ed.), Quantitative Methods in Pharmocology. Amsterdam: North-Holland, 1960.

Hovland, C.I. Experimental studies in role learning theory. III. Distribution of practice with varying speeds of syllable presentation. Journal of Experimental Psychology, 1938, 23, 172-190.

Johnson, N.L. Sequential analysis: a survey. Journal of the Royal Statistical Society, 1961, 124, 372-411.

Kilpatrick, GS., & Oldhan, P.D. Calcium chloride and adrenaline as bronchial dilators compared by sequential analysis. British Medical Journal, 1954, ii, 1388-1391.

Kimball, A.W., Burnett, W.T., Jr., & Doherty, D.G. Chemical protection against ionizing radiation. I. Sampling methods for screening in radiation protection studies with mice. Radiation Research, 1957, 7, 1-12.

Moonan, W.J. Some empirical aspects of the sequential analysis technique as applied to an achievement examination. Journal of Experimental Education, 1950, 18, 195-207.

Olivier, W.P. Computer-Assisted Mathematics Instruction for Community College Students. International Journal of Man-Machine Studies, 1973, 5, 385.

Peel, J., & Skipwort, G.E. Sequential sampling as a technique in sociological surveys. Applied Statistics, 1970, 19, 27.

Postman, L. Repition and paired-associate learning. American Journal of Psychology, 1962, 75, 372-389.

Rock, I. The role of repetition in associative learning. American Journal of Psychology, 1957, 70, 186-193.

Rushton, S. On a sequential t-test. Biometrika, 1950, 37, 326-333.

Rushton, S. On a two-sided sequential t-test. Biometrika, 1952, 39, 302-308.

Sainsbury, P., & Lucas, C.J. Sequential methods applied to the study of prochlorperazine. British Medical Journal, 1959, ii, 737-740.

Schmid, J., Jr. Sequential analysis of test items. Journal of Experimental Education, 1952, 20, 261-264.

Stroebel, C.F., & Glueck, B.C. Computer-derived global judgments in psychiatry. American Journal of Psychiatry, 1970, 126, 1057.

Sunday, D., & Kay, E. Serial position and paired-associate learning. Unpublished manuscript, 1974.

Underwood, B.J., Rehula, R., & Keppel, G. Item selection in paired-associate learning. American Journal of Psychology, 1962, 75, 353-371.

Wald, Abraham Sequential Analysis. New York: Wiley, 1947.

Wetherill, G.B. Sequential Methods in Statistics. New York: Wiley, 1966.

Bibliography

Barnard, G.A. The frequency justification of certain statistical tests. Biometrika, 1952, 39, 144-150.

Barraclough, E.D., & Page, E.S. Tables for Wald tests for the mean of a normal distribution. Biometrika, 1959, 46, 169-177.

Bartholomew, D.J. Hypothesis testing when sample size is treated as a random variable. Journal of the Royal Statistical Society, Series B, 1967, 29, 53.

Bhate, D.H. Approximations to the distribution of sample size for sequential tests, II. Tests of composite hypotheses. Biometrika, 1960, 47, 190-193.

Billard, L. Properties of some 2-sided sequential tests for normal distribution. Journal of the Royal Statistical Society, Series B, 1972, 34, 417.

Billard, L. Procedure for testing the mean of a normal distribution. Australian Journal of Statistics, 1973, 15, 80.

Bross, I. Sequential medical plans. Biometrics, 1952, 8, 188.

Buros, O.K. Statistical Methodology Reviews, 1941-1950. New York: Wiley, 1951.

Cox, D.R. A note on the sequential estimation of means. Proceedings of the Cambridge Philosophical Society, 1952, 48, 447-450.

Cox, D.R. Large sample sequential tests for composite hypotheses. Sankhya, 1963, 25, 5-12.

Ghosh, B.K. Sequential analysis of variance under random and mixed models. Journal of the American Statistical Association, 1967, 62, 1401.

Ghosh, J. On some properties of sequential t-tests. Bulletin of the Calcutta Statistical Association, 1960, 9, 77-86.

Gilchrist, W.G. Some sequential tests using range. Journal of the Royal Statistical Society, Series B, 1961, 23, 335-342.

Hajnal, J., Sharp, J., & Popert, A.J. A method for testing analgesics in rheumatoid arthritis using a sequential procedure. Annals of the Rheumatic Diseases, 1959, 18, 189-206.

Hoel, P.G. On a sequential test for the general linear hypothesis. Annals of Mathematical Statistics, 1955, 26, 136-139.

Ifram, A.F. On sample size and simplification of a class of sequential probability ratio tests. Annals of Mathematical Statistics, 1966, 37, 425.

Jackson, J.E. Bibliography on sequential analysis. Journal of the American Statistical Association, 1960, 55, 561-580.

Joanes, D.N. Sequential tests of composite hypotheses. Biometrika, 1972, 59, 633.

Jones, R.D. Sequential analysis of variance with unequal numbers of observations in experimental units. Dissertation Abstracts, 1968, 29 (3-B), 1203.

Kimball, A.W. Sequential sampling plans for use in psychological test work. Psychometrika, 1950, 15, 1-15.

Lechner, J.A. Optimality and OC curve for Wald SPRT. Journal of the American Statistical Association, 1964, 59, 464.

Manly, B.F.J. Choice of a Wald test on mean of a mormal population. Biometrika, 1970, 57, 91.

Nandi, H.K. Use of well-known statistics in sequential analysis. Sankhya, 1948, 8, 339-344.

Ray, W.D. Sequential analysis applied to certain experimental designs in the analysis of variance. Biometrika, 1956, 43, 388-403.

Schneiderman, M.A., & Armitage, P. Closed sequential t-tests. Biometrika, 1962, 49, 359-366.

Schwarz, G. Sequential student test. Annals of Mathematical Statistics, 1971, 42, 1003.

Siskind, V. On certain suggested formulae applied to the sequential t-test. Biometrika, 1964, 51, 97.

Weiss, L., & Wolfowit, J. Asymptotically efficient, sequential equivalent of the t-test. Journal of the Royal Statistical Society, Series B, 1972, 34, 456.

Appendix A

The following simple case will illustrate the use of the principle for choosing a critical region. It is taken from Wald (pp.18-19).

Let $\underline{x}$ be normally distributed with unknown mean and unit variance.

H_o: $\mu = \theta_o$

H_1: $\mu = \theta_1 \quad (\theta_1 \quad \theta_o)$

For testing H_o against H_1, we must determine the likelihood ratio (p.8).

$$\frac{f(x_1,h_1) \ \ldots \ f(x_n,h_1)}{f(x_1,h_o) \ \ldots \ f(x_n,h_o)} \geq k.$$

Since

$$f(x_1,h_1) \ \ldots \ f(x_n,h_1) = \frac{1}{(2\pi)^{n/2}} \ e^{-\frac{1}{2}\sum(x_i-\theta_1)^2}$$

and

$$f(x_1,h_o) \ \ldots \ f(x_n,h_o) = \frac{1}{(2\pi)^{n/2}} \ e^{-\frac{1}{2}\sum(x_i-\theta_o)^2}$$

the inequality can be written

$$\frac{e^{-\frac{1}{2}\sum(x_i-\theta_1)^2}}{e^{-\frac{1}{2}\sum(x_i-\theta_o)^2}} \geq k.$$

Taking the logarithm of both sides, we get

$$\tfrac{1}{2}\Sigma(x_i-\theta_o)^2 - \tfrac{1}{2}\Sigma(x_i-\theta_1)^2 = (\theta_1-\theta_o)\Sigma x_i + \tfrac{1}{2}n(\theta_o^2-\theta_1^2) \geq \log k.$$

Thus,

$$\Sigma x_i \geq k'$$

where

$$k' = \frac{\log k - \tfrac{1}{2}n(\theta_o^2-\theta_1^2)}{\theta_1 - \theta_o}$$

which can be written

$$\frac{\Sigma(x_i - \theta_o)}{n} \geq k''$$

where

$$k'' = \frac{k' - n\theta_o}{n} .$$

We now determine the value of k'' such that the critical region defined by the above inequality has the size $\alpha = .05$.

Under the hypothesis H_o the random variable $\Sigma(x_i - \theta_o)/n$ is normally distributed with zero mean and variance $1/n$. A table of the normal distribution tells us that

$$k'' = 1.64 / n^{\frac{1}{2}}.$$

Thus, the most powerful region of size .05 consists of all samples for which the inequality

$$\bar{X} - \theta_o \geq \frac{1.64}{n^{\frac{1}{2}}}$$

holds. This is a familiar result.

Appendix B

As an example of calculating the probability distribution for two groups given the incremental model, the situation of a subject making 14 correct responses on trials 2-5 in block II will be shown in detail. The following notation will be used. The subject's performance can be symbolized

$$x_1\ (x_2,\ x_3,\ x_4,\ x_5)$$

where x_i is the total number correct on trial i. The numbers within the parenthesis sum to the dependent variable of interest; thus, in this example, $x_2+x_3+x_4+x_5 = 14$. However, the number correct on the first trial, x_1, is shown since this value will effect the resulting probability.

The DE condition has a probability correct (given a previous error) = .30337 for any trial since a new stimulus is being tested. The DN condition has a probability correct (given a previous error) equal to

$$p(n) = c\sum(1 - c)^{i-1}$$

where c = .30337 (see p. 36). The calculated probabilities are:

Trial	p_n
1	.30337
2	.51471
3	.66193
4	.76449
5	.83594

The possible arrangements of 14 correct for the DE condition and their associated probabilities are:

$0(2,4,4,4)$: $\binom{4}{2}\binom{2}{2}\, c^4(1-c)^6 = .02466$

$1(2,4,4,4)$: $\binom{4}{1}\binom{3}{1}\binom{2}{2}\, c^4(1-c)^5 = .01667$

$2(2,4,4,4)$: $\binom{4}{2}\binom{2}{2}\, c^4(1-c)^4 = .01197$

$0(3,3,4,4)$: $\binom{4}{3}\binom{1}{1}\, c^4(1-c)^6 = .01644$

$1(3,3,4,4)$: $\binom{4}{1}\binom{3}{2}\binom{1}{1}\, c^4(1-c)^5 = .01667$

$2(3,3,4,4)$: $\binom{4}{2}\binom{2}{1}\binom{1}{1}\, c^4(1-c)^4 = .02393$

$3(3,3,4,4)$: $\binom{4}{3}\binom{1}{1}\, c^4(1-c)^3 = .01145$

.09038

The possible arrangements of 14 correct for the DN condition and their associated probabilities are:

$0(2,4,4,4)$: $6\, p_2^2\, p_3^2\, (1-p_1)^4(1-p_2)^2 = .03863$

$1(2,4,4,4)$: $12\, p_1\, p_2\, p_3^2\, (1-p_1)^3(1-p_2)^2 = .06537$

$2(2,4,4,4)$: $6\, p_1^2\, p_3^2\, (1-p_1)^2(1-p_2)^2 = .02765$

$0(3,3,4,4)$: $4\, p_2^3\, p_4\, (1-p_1)^4(1-p_2)(1-p_3) = .01611$

$1(3,3,4,4)$: $12\, p_1\, p_2^2\, p_4\, (1-p_1)^3(1-p_2)(1-p_3) = .04089$

$2(3,3,4,4)$: $12\, p_1^2\, p_2\, p_4\, (1-p_1)^2(1-p_2)(1-p_3) = .03460$

$3(3,3,4,4)$: $4\, p_1^3\, p_4\, (1-p_1)(1-p_2)(1-p_3) = .00976$

.23301

Appendix C

Given the probability distributions shown in Table 3 (p.39), the expectation and variance of each can be calculated.

Condition	Expectation	Variance
DN	14.06316	2.48599
DE	11.10773	9.32304

Delta is then calculated using the previous formula.

$$\delta = \frac{\mu_{DN} - \mu_{DE}}{(\sigma^2_{DN} + \sigma^2_{DE})^{\frac{1}{2}}}$$

$$= \frac{14.06316 - 11.10773}{(2.48599 + 9.32304)^{\frac{1}{2}}}$$

$$\delta = .86003$$

Appendix D

Since boundary values are not found in the NBS tables for $\underline{\delta} = .86$, a quadratic interpolation was performed on the given values using the Gregory-Newton formula (see Arnold, p.xi):

$$z_p = p\, z_1 + (1-p)\, z_o' - \tfrac{1}{2}\, p(1-p)(z_o - 2z_1 + z_2)$$

where $p = .6$

$z_o = z(\text{for } \underline{\delta} = .8)$

$z_1 = z(\text{for } \underline{\delta} = .9)$

$z_2 = z(\text{for } \underline{\delta} = 1.0)$

The accept-H_o boundary is shown in Table 8. The reject-H_o boundary is shown in Table 9.

Table 8

The quadratic interpolation of the accept-H_o boundary for $\underline{\delta}$ = .86.

n	z(.8)	z(.9)	z(1.)	I	II	z(.86)
8	-	0.099	0.350	-	-	-
9	-	0.234	0.519	-	-	-
10	0.086	0.371	0.692	0.2570	0.0043	0.2527
11	0.193	0.510	0.870	0.3832	0.0052	0.3780
12	0.302	0.653	1.051	0.5126	0.0056	0.5070
13	0.412	0.799	1.234	0.6442	0.0058	0.6384
14	0.525	0.948	1.421	0.7788	0.0060	0.7728
15	0.640	1.100	1.609	0.9160	0.0059	0.9101
16	0.757	1.253	1.798	1.0546	0.0059	1.0487
17	0.877	1.409	1.989	1.1962	0.0058	1.1904
18	0.999	1.566	2.181	1.3392	0.0058	1.3334

Note: $I = .6z_1 + .4z_o$; $II = .12(z_o - 2z_1 + z_2)$.

Table 9

The quadratic interpolation of the reject-H_o boundary for $\underline{\delta}$ = .86.

n	z(.8)	z(.9)	z(1.)	I	II	z(.86)
6	5.831	5.325	4.992	5.5274	0.0208	5.5066
7	5.718	5.309	5.053	5.4726	0.0184	5.4542
8	5.660	5.333	5.145	5.4638	0.0167	5.4471
9	5.641	5.386	5.258	5.4880	0.0152	5.4728
10	5.650	5.460	5.387	5.5360	0.0140	5.5220
11	5.680	5.549	5.528	5.6014	0.0132	5.5882
12	5.727	5.651	5.678	5.6814	0.0124	5.6690
13	5.787	5.762	5.835	5.7720	0.0118	5.7602
14	5.857	5.881	5.997	5.8714	0.0110	5.8604
15	5.935	6.006	6.164	5.9776	0.0104	5.9672
16	6.021	6.135	6.335	6.0894	0.0103	6.0791
17	6.112	6.269	6.509	6.2062	0.0100	6.1962
18	6.208	6.407	6.686	6.3274	0.0096	6.3178

Note: $I = .6z_1 + .4z_o$; $II = .12(z_o - 2z_1 + z_2)$.

Appendix E

The boundaries for the Rushton approximation t-test given α = .05 and β = .05 are as follows:

$$\ln A = \ln \frac{(1 - \beta)}{\alpha}$$

$$= \ln \frac{(1 - .05)}{.05}$$

$$= \ln 19.0$$

$$= 2.9444$$

$$\ln B = \ln \frac{\beta}{(1 - \alpha)}$$

$$= \ln \frac{.05}{(1 - .05)}$$

$$= \ln .0526$$

$$= -2.9444$$

Appendix F

The constants for the sequential sign test are calculated as follows.

The quantity p_1 was defined (p.30) as

$$p_1 = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{\delta}^{\infty} e^{-\frac{1}{2}x^2} dx$$

This can be found readily in a table of the normal distribution. Its value at $\delta = .86$ is .1949.

Given $\alpha = .05$, $\beta = .05$, and $p_1 = .1949$,

$$A_1 = \frac{\log \frac{(1-\beta)}{\alpha}}{\log 2(1-p_1)} = 6.1811.$$

$$A_2 = \frac{\log \frac{(1-\alpha)}{\beta}}{\log 2(1-p_1)} = 6.1811$$

$$B = \frac{\log \frac{1}{2p_1}}{\log 2(1-p_1)} = 1.9778$$

Vita

James M. Goett, son of Edward and Hazel Goett, was born on July 6, 1950 in Bronx, New York. After graduating in 1968 from Archmere Academy, Claymont, Delaware, he entered the University of Notre Dame. In 1972 he graduated with a Bachelor of Arts degree in Psychology. After a year of non-degree graduate work at Notre Dame, he accepted a teaching assistantship at Lehigh University in order to continue work towards the Ph.D.