AN EMPIRICAL EVALUATION OF MULTIVARIATE SEQUENTIAL PROCEDURES FOR TESTING MEANS

by

Robert Houston Appleby

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

May 1960

Blacksburg, Virginia

TABLE OF CONTENTS

I. INTRODUCTION

Jackson (1959) extended Wald's (1947) sequential <u>probability ratio test</u> for testing hypotheses about means in the univariate case to that of mean vectors in a multivariate situation, both when the variance-covariance matrix is assumed known (Sequential χ^2-Test) and when the variance-covariance matrix is unknown (Sequential T^2-Test). It was shown that the sequential procedures terminate with probability one, under practically rather general conditions, and that the risks of accepting H_1 when H_0 is true and of accepting H_0 when H_1 is true are α and β respectively.

The purpose of this Monto Carlo evaluation (similar to a univariate study by K. J. Arnold, 1951) of the Sequential χ^2 and T^2 Tests is to determine empirically the Average Sample Number (ASN) values, to compare them with those obtained by Jackson using Bhate's Conjecture and to compare empirically the α and β errors to those desired.

1.1 Univariate Sequential Analysis

1.11 Probability Ratio Test.

The development of sequential testing procedures has had many contributors. The most notable of these is Abraham Wald, primarily through his book Sequential Analysis (1947).

For testing the hypotheses:

$$H_0: \quad \mu-\mu_0 = 0$$

$$H_1: \quad \mu-\mu_0 = \lambda$$

Wald sets up a probability ratio test, which is formed by taking the ratio of the likelihood function $L(x;\ \theta_1)$, given that H_1 is true, to the likelihood function $L(x;\ \theta_0)$, given that H_0 is true. The probability ratio, based on a sample with n observations, is denoted by

$$P_{1n}/P_{0n} = L(x;\ \theta_1)/L(x;\ \theta_0).$$

After n observations the ratio is evaluated as follows:

a) $P_{1n}/P_{0n} < \frac{\beta}{1-\alpha}$, accept H_0;

b) $P_{1n}/P_{0n} > \frac{1-\beta}{\alpha}$, reject H_0;

c) $\frac{\beta}{1-\alpha} < P_{1n}/P_{0n} < \frac{1-\beta}{\alpha}$, continue sampling;

where α is the probability of rejecting H_0 when H_0 is true and β the probability of accepting H_0 when H_1 is true. This procedure is continued until a decision to accept or reject is made. Proofs are given that this test procedure will terminate with probability one.

1.12 Tables for Sequential t-Test.

The procedure for known variance is contained in elementary texts. However, for unknown variance (equal under both hypotheses) the techniques are somewhat more complicated. To aid in using Wald's test in that case Tables to Facilitate Sequential t-Tests were published by the National Bureau of Standards in 1951.

To use these tables a sample of size n is drawn and

$$z_n = [\sum_{i=1}^{n}(x_i-\mu_0)]^2/\sum_{i=1}^{n}(x_i-\mu_0)^2$$

is computed and compared with the upper limit (L_A) and the lower limit (L_B) given in the table for desired values of the α and β risks and the sample size n. The upper limit in the tables was found by solving for z in:

$$\ln[(1-\beta)/\alpha] = \ln\ {}_1F_1\ (\frac{n}{2}, \frac{1}{2}, \frac{z\lambda^2}{2}) - \frac{n\lambda^2}{2}$$

and the lower limit found by solving for z:

$$\ln[\beta/(1-\alpha)] = \ln\ {}_1F_1\ (\frac{n}{2}, \frac{1}{2}, \frac{z\lambda^2}{2}) - \frac{n\lambda^2}{2}$$

where ${}_1F_1$ is a confluent hypergeometric function.

1.13 Monte Carlo Study for p = 1.

As part of the introduction to the above tables K. J. Arnold made a "Monte Carlo" study of the efficiency of two of Wald's sequential tests. A comparison of

$$(1.11) \quad P_{1n}/P_{0n} = e^{-\frac{1}{2}n\lambda^2} \; {}_1F_1\left(\frac{n-1}{2}, \frac{1}{2}, \frac{z\lambda^2}{2}\right)$$

and

$$(1.12) \quad P_{1n}/P_{0n} = e^{-\frac{1}{2}n\lambda^2} \; {}_1F_1\left(\frac{n}{2}, \frac{1}{2}, \frac{z\lambda^2}{2}\right)$$

was made by drawing samples from a table of random normal numbers and using the table to make a decision as follows:

a) $z_n < L_B$, accept H_0;

b) $z_n > L_A$, reject H_0;

c) $L_B < z_n < L_A$, take another observation.

Arnold found little difference between the ASN of the two tests but the test given by equation (1.12) tended to reject more samples under both H_0 and H_1. This tendency makes α closer but β farther away from the desired .05. The α and β errors for equation (1.12) are .044 and .034 respectively. Equation (1.12) is the same as (1.23) with $p = 1$.

Arnold's empirical ASN sampling from 500 samples was compared by Jackson to Bhate's conjectured value and found to be "of the right order of magnitude" (Jackson, 1959).

1.2 <u>Multivariate Sequential Analysis</u>

1.21 Results of Jackson's Work.

Jackson (1959) extended the univariate test to the multivariate (p-variables) situation for testing the hypotheses:

$$H_0: \quad (\underline{\mu - \mu_0})' \, \Sigma^{-1} (\underline{\mu - \mu_0}) = 0$$

$$H_1: \quad (\underline{\mu - \mu_0})' \, \Sigma^{-1} (\underline{\mu - \mu_0}) = \lambda^2 \quad \text{where } (\underline{\mu - \mu_0})$$

is a column vector of means and Σ is the variance-covariance matrix of the variables under consideration. For this set of hypotheses, the probability ratio is formed and evaluated in the same manner as for the univariate situation.

When sampling from a multivariate normal population with unknown means but Σ assumed known, henceforth to be called Case I, the probability ratio for a sample of n observations is the ratio of the Non-Central χ^2-distribution with p degrees of freedom and non-centrality parameter $n\lambda^2$, to the central χ^2-distribution with p degrees of freedom. This is given by

(1.21) $$P_{1n}/P_{0n} = e^{-\frac{1}{2}n\lambda^2} \; {}_0F_1 \left(\frac{p}{2}; \; n\lambda^2\chi_n^2/4\right), \text{ where}$$

$$ {}_0F_1 (a,x) = \sum_{n=0}^{\infty} \frac{x^n}{n! \; a(a+1)\dots(a+n-1)}$$

and

$$\chi_n^2 = n(\underline{\bar{x} - \mu_0})' \, \Sigma^{-1} \, (\underline{\bar{x} - \mu_0}).$$

For a sequential procedure based upon the above population, but with Σ assumed unknown and estimated from the sample, henceforth to be called Case II, the probability ratio is the ratio of the Non-Central T^2-distribution with non-centrality parameter $n\lambda^2$ to the central T^2-distribution. This ratio is given by

$$(1.22) \qquad P_{1n}/P_{0n} = e^{-\frac{1}{2}n\lambda^2} \; {}_1F_1 \left[\frac{n}{2}, \frac{p}{2}; \frac{n\lambda^2 T_n^2}{2(n-1+T_n^2)}\right],$$

where

$${}_1F_1 (a,b;x) = \sum_{n=0}^{\infty} \frac{a(a+1)\ldots(a+n-1)\; x^n}{b(b+1)\ldots(b+n-1)n!}$$

and

$$T_n^2 = n(\underline{\bar{x}-\mu_0})' \; S^{-1}(\underline{\bar{x}-\mu_0}),$$

and S a matrix that is a sample estimate of Σ.

In making the actual test, for both cases a sample of size n is drawn; χ_n^2 or T_n^2 computed and P_{1n}/P_{0n} is evaluated; if

a) $P_{1n}/P_{0n} < \beta/(1-\alpha)$, accept H_0;

b) $P_{1n}/P_{0n} > (1-\beta)/\alpha$, reject H_0;

c) $\beta/(1-\alpha) < P_{1n}/P_{0n} < (1-\beta)/\alpha$, continue sampling.

1.22 Tables for Sequential χ^2-Test and T^2-Test.

Based on the above probability ratios, Jackson constructed tables of upper and lower limits of χ_n^2 and T_n^2 using a procedure similar to that used in establishing the National Bureau of Standards tables.

For Case I; $P_{1n}/P_{0n} = e^{-\frac{1}{2}n\lambda^2} \; {}_0F_1(\frac{p}{2}; n\lambda^2\chi_n^2/4)$

is solved for upper limit of χ_n^2 (denoted by $\overline{\chi_n^2}$) by setting $P_{1n}/P_{0n} = (1-\beta)/\alpha$, and the lower limit of χ_n^2 $(\underline{\chi_n^2})$ by setting

$P_{1n}/P_{0n} = \beta/(1-\alpha)$. From this, tables for $p = 2, 3, \ldots, 9$ and $\lambda^2 = .5, 1.0, 2.0$ with α and β equal to .05 have been computed. (More extensive, but unpublished, tables by Jackson and Freund have, for the same values of p, α and β, selected values of $\lambda^2 = .25, \ldots, 10.0$.)

To use these tables a sample of size n is drawn, $\chi_n^2 = n(\underline{\bar{x}-\mu_0})'\Sigma^{-1}(\underline{\bar{x}-\mu_0})$ is computed and compared with tabled values:

a) $\chi_n^2 < \underline{\chi_n^2}$, accept H_0;

b) $\chi_n^2 > \overline{\chi_n^2}$, reject H_0;

c) $\underline{\chi_n^2} < \chi_n^2 < \overline{\chi_n^2}$, continue sampling.

For Case II, as for Case I, the ratio

$$P_{1n}/P_{0n} = e^{-\frac{1}{2}n\lambda^2}\ {}_1F_1\left[\frac{n}{2}, \frac{p}{2}; \frac{n\lambda^2 T_n^2}{2(n-1+T_n^2)}\right] \tag{1.23}$$

is set equal to $(1-\beta)/\alpha$ and solved to yield $\overline{T_n^2}$; then it is set equal to $\beta/(1-\alpha)$ and solved to obtain $\underline{T_n^2}$. Tables have been computed for the same sets of parameters as in Case I.

To use these tables a sample of size n is drawn, $T_n^2 = n(\underline{\bar{x}-\mu_0})'S^{-1}(\underline{\bar{x}-\mu_0})$ is computed and compared with tabled values of $\underline{T_n^2}$ and $\overline{T_n^2}$:

a) $T_n^2 < \underline{T_n^2}$, accept H_0;

b) $T_n^2 > \overline{T_n^2}$, reject H_0;

c) $\underline{T_n^2} < T_n^2 < \overline{T_n^2}$, continue sampling.

1.23 Bhate's Conjecture for ASN.

No exact method of estimating the ASN when the variables are not independent exists at present. However, a method for approximating ASN, called Bhate's Conjecture, has been used for this purpose.

For Case I, and sampling from H_0 (i.e., $\lambda^2 = 0$ in the population), the ASN is found by solving for n:

$$(1.24) \qquad [(1-\beta)/\alpha]^{\alpha} \, [\beta/(1-\alpha)]^{1-\alpha} = e^{-\frac{1}{2}n\lambda_1^2} \, {}_0F_1(\tfrac{p}{2}; \tfrac{np\lambda_1^2}{4}),$$

where $n\lambda_1^2$ is the non-centrality parameter for H_1. When H_1 is assumed to be true ($\lambda^2 = \lambda_1^2$ in the population) the ASN is found by solving for n:

$$(1.25) \qquad [(1-\beta)/\alpha]^{1-\beta} \, [\beta/(1-\alpha)]^{\beta} = e^{-\frac{1}{2}n\lambda_1^2} \, {}_0F_1 \, [\tfrac{p}{2}, \tfrac{n\lambda_1^2 \, (p+n\lambda_1^2)}{4}].$$

Similarly for Case II, ASN is found assuming H_0 is true by solving for n:

$$(1.26) \qquad [\beta/(1-\alpha)]^{1-\alpha} \, [(1-\beta)/\alpha]^{\alpha} = e^{-\frac{1}{2}n\lambda_1^2} \, {}_1F_1 \, (\tfrac{n}{2}, \tfrac{p}{2}; \tfrac{p\lambda_1^2}{2}),$$

and assuming H_1 is solving for n:

$$(1.27) \qquad [\beta/(1-\alpha)]^{\beta} \, [(1-\beta)/\alpha]^{1-\beta} = e^{-\frac{1}{2}n\lambda_1^2} \, {}_1F_1\{\tfrac{n}{2}, \tfrac{p}{2}; \tfrac{n\lambda_1^2}{2}(1-e^{-\frac{1}{2}n\lambda_1^2}[\tfrac{n-p}{n}]\, {}_1F_1[\tfrac{n}{2}, \tfrac{n+2}{2}; \tfrac{n\lambda_1^2}{2}])\}.$$

Tables 3.2 and 3.6 show the results of the empirical ASN values obtained by a Monte Carlo technique as compared with the values obtained from the above conjecture.

1.3 Monte Carlo Methods

In simulation or Monte Carlo studies, a probabilistic process is simulated or generated by appropriate use of random numbers. A large number of these simulated samples can be repeated on an electronic computer and inferences can be drawn from the analysis of these repeated samples.

Such simulations are useful to 1) solve distribution problems too difficult to handle theoretically and 2) to verify approximations used in mathematical derivations. This study is of the second type. H. A. Meyer (1956) states "the efficiency of the methods ... seems unbelievable".

1.31 Method of Obtaining Random Normal Numbers.

In order to obtain results a method must be used that will give random normally distributed observations with a given mean and given variance. In Arnold's study the method used was to take a random sample from a table of random normal numbers. The method used in this study was to use a subroutine for the IBM 650 (6.5.002.1 NCS) that could be incorporated in a program to compute the statistic and test;

thereby making the program self-sustained.

This program utilizes the Von Neumann "Middle Square" method of generating random numbers with a rectangular distribution. A normalizing transformation is then used to make these numbers normal deviates N(0,1). A test is incorporated to insure normality and to exclude all numbers with absolute value greater than 4. (A more detailed description is found in the writeup of program 6.5.002.1 NCS). For sampling from a multivariate population this method generates independent variables, i.e., with a diagonal variance-covariance matrix. Linear transformations may be used to generate other multivariate normal distributions.

1.32 Validity of Generator.

A paper by N. Metropolis concerning the Von Neumann method was published in Symposium on Monte Carlo Methods, edited by H. A. Meyer (1956). In this study, for a computer that has 20 binary digits the number of random deviates generated from each starting number will be 1,048,576. (The IBM 650 has the equivalent of more than 30 binary digits and should produce an even larger amount of deviates from any starting number). For the present study the largest number of random deviates used for any one set of parameters is less than 75,000. Therefore, since a new starting number was selected for each set of parameters, there should be no duplication of samples within any set.

II. COMPUTATION PROCEDURE FOR SAMPLING STUDY

2.1 Case I

To test the hypotheses

$$H_0: \quad (\underline{\mu}-\underline{\mu_0})'\Sigma^{-1}(\underline{\mu}-\underline{\mu_0}) = 0$$

$$H_1: \quad (\underline{\mu}-\underline{\mu_0})'\Sigma^{-1}(\underline{\mu}-\underline{\mu_0}) = \lambda^2,$$

a program for the IBM 650 was written which would draw random samples from a population

$$(2.11) \qquad f(x_1, x_2, \ldots, x_p) = (2\pi)^{-\frac{1}{2}p}|\Sigma|^{-1}e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})},$$

where in this case, Σ is known, and is an identity matrix. Without loss of generality, $\underline{\mu}$ is taken to be a vector of zeros under H_0 and $\underline{\mu} = (\lambda/\sqrt{p})\underline{j}$ (a vector with all elements equal to $\lambda/\sqrt{p}$) under H_1; thus the hypotheses become

$$H_0: \quad \underline{\mu}'\Sigma^{-1}\underline{\mu} = 0$$

$$H_1: \quad \underline{\mu}'\Sigma^{-1}\underline{\mu} = \lambda^2.$$

After each sample is drawn $\chi_n^2 = n(\bar{\underline{x}})'\Sigma^{-1}(\bar{\underline{x}})$ is computed. Since Σ is an identity matrix the statistic becomes

$$\chi_n^2 = 1/n \sum_{\alpha=1}^{p} \left(\sum_{i=1}^{n} x_{\alpha i}\right)^2.$$

After each sample, beginning with the lowest values available in the corresponding tables, this value is compared to tabled values of $\underline{\chi_n^2}$ and $\overline{\chi_n^2}$, a decision being made as follows:

a) $\chi_n^2 < \underline{\chi_n^2}$, accept H_0;

b) $\chi_n^2 > \overline{\chi_n^2}$, reject H_0;

c) $\underline{\chi_n^2} < \chi_n^2 < \overline{\chi_n^2}$, continue sampling.

The sets of parameters (λ^2, p) used in this study are listed in Table 3.2. These were chosen with the intention of obtaining a cross section of the values tabled by Jackson (1959) and to get an over-all picture of the results of the testing procedures. Since there exists a limit to these tables a decision must be made to terminate the sample with no conclusion about the hypotheses when this limit is reached. These limits vary with the set of parameters involved and are indicated in the last column of Table 3.4 on page 22.

When the sample has been accepted, rejected or terminated, another sample is initiated. This continues until a "sufficient number" of samples have been tested. For Case I, a "sufficient number" was chosen to be a minimum of 500. With samples of this size a 95% confidence interval on α or β, which are parameters of a binomial, has a width of .038 if α or β are .05. For some sets of parameters computing time was shorter and thus 1000 samples were tested, which makes the confidence interval narrower by a factor of $1/\sqrt{2}$.

A flow chart of the computer program is given in Appendix B.

2.2 Case II

When testing the same hypotheses as in Case I, and taking samples from the same distribution, but with Σ assumed unknown, an estimate S of Σ must be obtained from the sample. Let

$$A = \sum_{\alpha=1}^{p} (\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = (n-1)S.$$

Then $A^{-1} = S^{-1}/(n-1)$. The statistic $T_n^2 = n(\underline{\bar{x}})'S^{-1}(\underline{\bar{x}})$ can be written $T_n^2 = (n-1)/n\ \underline{X}'A^{-1}\underline{X}$, where $X = \sum_{i=1}^{n} x_{\alpha i}$, $\alpha=1,2,\ldots,p$.

After each sample the value of T_n^2 is then compared with the tabled values of $\underline{T}_n^2$ and $\overline{T}_n^2$:

a) $T_n^2 < \underline{T}_n^2$, accept H_0;

b) $T_n^2 > \overline{T}_n^2$, reject H_0;

c) $\underline{T}_n^2 < T_n^2 < \overline{T}_n^2$, continue sampling.

For this test no decision can be reached until either $(p+1)$ or $6/\lambda^2$ (whichever is larger) samples have been drawn. Also $\overline{T}_n^2$ does not exist until about the $(p+6)$-th observation (varies with λ^2), and a decision to reject cannot be made until that sample size is reached. The lowest value of n which will allow a decision to be made for various sets of parameters will differ for both accepting and rejecting the sample. For example, when $p = 9$, $\lambda^2 = 2.0$, it is possible to accept at $n = 10$, but no rejection can be made until $n = 14$.

In the testing procedure, after the maximum of $(\lambda^2, p+1)$ samples have been drawn the estimate of Σ must be inverted to compute T_n^2. This matrix inversion is time consuming, therefore, the number of sets of parameters for which samples were drawn was reduced. The parameters chosen are listed in Table 3.5, and are the smallest and largest values of p for which $\overline{T_n^2}$ and $T_{\underline{n}}^2$ have been tabulated.

As in Case I, limits of the tabled values sometimes cause a termination of sample before a decision can be reached. Table 3.8, page 25, gives the terminating values of n.

Further restrictions on time when $p = 9$ (about 6 samples per hour can be drawn and tested) have caused a reduction in the minimum number of samples, from 500 to 100, for each value of λ^2. This sample size increases the 95% confidence interval for α and β by a factor $\sqrt{5}$.

A flow chart of the computer program is given in Appendix B.

III. ANALYSIS OF RESULTS

3.1 Case I

For each set of parameters (p,λ^2) an empirical ASN was determined and compared with conjectured values. The frequency of occurence of sample sizes was tabulated, and the number of rejections and acceptances recorded.

A complete breakdown is given for a typical set of parameters: p = 3 and testing the hypotheses:

$$H_0: \quad (\underline{\mu}-\underline{\mu_0})'\Sigma^{-1}(\underline{\mu}-\underline{\mu_0}) = 0$$

$$H_1: \quad (\underline{\mu}-\underline{\mu_0})'\Sigma^{-1}(\underline{\mu}-\underline{\mu_0}) = 1.0.$$

Table 3.1 gives the frequencies and sample size for testing H_0 vs. H_1 assuming $\lambda^2 = 0$ and $\lambda^2 = 1.0$ with $p = 3$. The empirical ASN is below the conjectured value when sampling from H_0, and above the conjectured value when sampling from H_1. In general, as can be seen from Table 3.2, this appears to be the pattern when comparing conjectural values with the ASN as obtained from the Monte Carlo study.

The fixed sample size for the univariate case is known to be much larger than the ASN for sequential testing. It is assumed that the ASN will be smaller for the multivariate case, therefore, the fixed sample size is included in Table 3.2.

Table 3.1

Frequency Distribution of Sample Size

($p = 3$, $\alpha = \beta = .05$)

Sample Size	$\lambda^2 = 0$ Accept	Reject	Total	$\lambda^2 = 1.0$ Accept	Reject	Total
3	0	1	1	0	17	17
4	0	1	1	0	29	29
5	0	1	1	0	30	30
6	2	2	4	0	43	43
7	39	1	40	3	43	46
8	57	1	58	1	49	50
9	53	1	54	3	38	41
10	56	2	58	3	34	37
11	57	1	58	2	29	31
12	41	0	41	0	20	20
13	39	0	39	0	21	21
14	23	2	25	0	27	27
15	17	1	18	0	20	20
16	22	1	23	0	15	15
17	17	0	17	0	10	10
18	11	0	11	0	11	11
19	11	0	11	0	6	6
20	5	0	5	0	9	9
21	7	1	8	0	8	8
22	4	0	4	0	7	7
23	7	0	7	0	4	4
24	4	0	4	0	4	4
25	2	0	2	1	3	4
26	4	0	4	0	2	2
27	2	0	2	0	2	2
28	2	0	2	0	0	0
29	0	0	0	0	2	2
30	2	0	2	0	0	0
31	1	0	1	0	0	0
32	1	0	1	0	0	0
33	0	0	0	0	0	0
34	0	0	0	0	2	2
35	1	0	1	0	0	0
41	0	0	0	1	0	1
45	0	0	0	1	0	1
Totals	487	16	503	15	485	500
Average n	12.4	9.9	12.3	14.7	10.7	10.8
Conjectured Value			13.3			8.5

Appendix A lists sample size distributions for all parameters studied.

Table 3.2 summarizes, for all parameters, the ASN and Bhate's conjectural value along with the fixed sample size. The fixed sample size was determined from the improved variance-stabilizing (square root) transformation of non-central χ^2 given by Hofer, (1960).

Table 3.2

Comparison of Conjectured, Empirical ASN

		H_0		H_1		Fixed
p	λ^2	ASN	Bhate's	ASN	Bhate's	Sample
2	.5	21.3	25	18.2	15	32
2	1.0	10.3	13	9.6	8	16
2	2.0	6.0	7	5.3	4	8
3	.5	24.0	27	21.2	17	35
3	1.0	12.4	14	10.8	9	18
3	2.0	6.5	7	5.9	5	9
5	1.0	14.5	15	12.5	11	20
5	2.0	7.6	8	6.7	6	10
5	5.0	3.3	3	3.2	3	4
9	2.0	9.1	9	8.0	7	12
9	5.0	4.0	4	3.7	3	5
9	10.0	2.3	2	2.1	2	3

For this table the conjectured values were rounded to the next largest integer, and in some cases will disagree with values given by Jackson (1959).

3.11 Alpha and Beta Errors.

The desired probability for the two types of error was stated to be .05. Table 3.3 lists the empirical probabilities that were obtained by the sampling. These probabilities are far below the desired ones in all cases, with the test becoming increasingly conservative with larger λ^2. In most cases the 95% confidence interval does not even include the desired value of .05.

Table 3.3

Actual α and β Errors

p	λ^2	α	β
2	0.5	.030	.036
2	1.0	.020	.038
2	2.0	.022	.019
3	0.5	.035	.035
3	1.0	.031	.030
3	2.0	.016	.031
5	1.0	.024	.036
5	2.0	.011	.034
5	5.0	.022	.018
9	2.0	.031	.030
9	5.0	.012	.026
9	10.0	.018	.014

3.12 On Termination.

When no decision has been reached after testing $\chi^2_{n(\lambda^2,p)}$, the sample is terminated, where $n(\lambda^2,p)$ is the termination value of n for the specified values of λ^2 and p.

Table 3.4 gives the percentage of samples that were terminated without decision for each set of parameters. As was expected the highest percentage occured when $\lambda^2 = 0.5$ and $p = 3$. Wald (1947) gives a method for testing after a sample has been terminated without decision. Compute the probability ratio and if:

a) $\beta/(1-\alpha) < P_{1n}/P_{0n} < 1$, accept H_0;

b) $1 < P_{1n}/P_{0n} < (1-\beta)/\alpha$, reject H_0.

Table 3.4

Percentage of Samples Terminated

p	λ^2	under H_0	under H_1	$n(\lambda^2,p)$
2	0.5	.006	.004	60
2	1.0	0	0	45
2	2.0	0	0	30
3	0.5	.018	.011	60
3	1.0	0	0	45
3	2.0	0	0	30
5	1.0	0	0	50
5	2.0	.002	.002	25
5	5.0	0	.006	10
9	2.0	0	0	30
9	5.0	0	.002	12
9	10.0	0	0	6

Applying this test to the samples for $p = 3$, $\lambda^2 = 0.5$, the following results were obtained. Out of 1035 samples under H_0, 19 terminated without decision, 17 samples accepted H_0;

2 rejected H_0. Under H_1, out of 6 of 543 samples, 5 rejected H_0 and 1 accepted H_0. This changed the α and β errors from .035 to .037 for both hypotheses.

3.2 Case II

3.21 Tabulation for ASN.

Sample size and frequency of occurence was tabulated for each set of parameters. Table 3.5 gives a typical breakdown, when p = 9 and testing

$$H_0: \quad (\underline{\mu}-\underline{\mu}_0)' \, \Sigma^{-1}(\underline{\mu}-\underline{\mu}_0) = 0$$

$$H_1: \quad (\underline{\mu}-\underline{\mu}_0)' \, \Sigma^{-1}(\underline{\mu}-\underline{\mu}_0) = 10.0$$

Table 3.5

Frequency Distribution of sample size

($p = 9$, $\alpha = \beta = .05$)

	H_0			H_1		
Sample Size	Accept	Reject	Total	Accept	Reject	Total
10	41	0	41	1	0	1
11	23	0	23	1	0	1
12	19	1	20	0	40	40
13	8	0	8	0	35	35
14	5	0	5	0	9	9
15	1	1	2	0	11	11
16	0	0	0	0	2	2
17	0	1	1	0	0	0
18	0	0	0	0	1	1
Totals	97	3	100	2	98	100
Average n	11.1	14.7	11.2	10.5	13.0	13.0

Table 3.6 gives a comparison of values for all parameters

and the fixed sample size. Fixed sample size was determined using the improved variance-stabilizing ($\cosh^{-1}$) transformation given by Bargmann (1958).

Table 3.6

Comparison of Conjectured, Empirical ASN

		H_0		H_1		
p	λ^2	ASN	Bhate's	ASN	Bhate's	Fixed
2	0.5	22.7	26	23.1	21	35
2	1.0	12.4	14	13.9	13	21
2	2.0	7.7	8	9.1	9	14
9	2.0	16.0	15	18.5	18	36
9	6.0	11.9	11	14.0	14	18
9	10.0	11.2	10	13.0	13	16

3.22 Alpha and Beta errors

Table 3.7 lists the actual α and β errors which compare more favorably with the desired .05 than the actual errors in Case I, but are still below the desired value.

Table 3.7

Actual α and β Errors

p	λ^2	α	β
2	0.5	.034	.042
2	1.0	.047	.032
2	2.0	.042	.012
9	2.0	.030	.010
9	6.0	.020	.010
9	10.0	.030	.020

In most cases the 95% confidence interval contains the desired value .05, but in the case where p = 9, the width of the confidence interval is approximately .085. The actual values of α and β errors obtained in the study by Arnold (1951) were .044 and .034 respectively.

3.23 On Termination.

Table 3.8 lists the percentage of samples that did not reach a decision after testing $T^2_{n(\lambda^2,p)}$, where $n(\lambda^2,p)$ is the limit of n for the tabled values of λ^2 and p.

Table 3.8

Percentage of Samples Terminated

p	λ^2	under H_0	under H_1	$n(\lambda^2,p)$
2	0.5	.008	0	60
2	1.0	.002	0	45
2	2.0	0	0	30
9	2.0	0	0	50
9	6.0	0	0	30
9	10.0	0	0	20

It is interesting to note that only under H_0 are there any samples that do not reach a decision. It would seem from this that the present tables are sufficiently large for the parameters under consideration.

IV. SUMMARY

4.1 Inferences from Results

The results of the simulation study indicate that the multivariate sequential test procedures for testing means proposed by Jackson (1959) lead to satisfactory decisions. The tests are highly conservative, but the α and β errors are of the appropriate magnitude. The use of Bhate's conjecture as an estimate for the average sample number appears to give a good estimate of the size of the sample needed to obtain a decision.

Due to the time element for sampling in Case II, it was not possible to obtain as complete a coverage as was done for Case I. The data appear, however, to be following an expected pattern and the same inferences from the data should be made for Case II as for Case I.

4.2 Future Research

Outlined below are some problems to which answers may be supplied by using the programs that were developed for this study.

How good is the test if the actual λ^2 is not that specified by H_1? For example, a study could be made which would investigate the effects of testing $\lambda^2 = 0$, vs. $\lambda^2 = 1.0$

when the true λ^2 is assumed to lie between these. One may also use $\lambda^2 = 2.0$ or 3.0, but this will probably result in rejecting H_0 after very few observations. In this manner a power curve could be constructed.

The test procedure assumes that the sampling is done from a normal distribution. What effect would sampling from a non-normal distribution have on the test procedures? The robustness of the sequential test could be investigated by generating specific non-normal distributions.

Since the test procedures are highly conservative perhaps a method could be found of shortening the interval between the rejection and acceptance points of the tables thus producing α and β errors closer to those desired and consequently reducing the size of the sample needed for decision.

BIBLIOGRAPHY

Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York.

Bargmann, R. E. (1958). Some Interpretations in the Analysis of Transformed Data. Technical Report, Virginia Polytechnic Institute.

Hofer, G. F. (1960). A Variance-Stabilizing Transformation of the Non-Central χ^2-Distribution. Unpublished Master's Thesis, Virginia Polytechnic Institute.

Jackson, J. E. (1959). Multivariate Sequential Procedures for Testing Means. The Development of Statistical Methods of Experimental Designs in Quality Control and Surveillance Testing, Technical Report No. 10, Virginia Polytechnic Institute.

Meyer, H. A. (1956). Symposium on Monte Carlo Methods, John Wiley and Sons, Inc., New York.

National Bureau of Standards (1951). Tables to Facilitate Sequential t-Tests, Applied Mathematics Series No. 7.

North Carolina State College. Random Normal Deviates. IBM 650 Program 6.5.002.1 NCS.

Wald, A. (1947). Sequential Analysis, John Wiley and Sons, Inc., New York.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to whose experience with the IBM 650 made him an invaluable source of information and a great aid in writing the programs for the computer.

For his many and varied criticisms and advice, the author wishes to thank .

Thanks to the entire IBM center personnel who assisted the author in many ways.

Special thanks to and for their invaluable contributions with the typewriter.

Appendix A

Frequency Distribution of Sample Size for χ^2-Test

$p = 2$ variables

Size \ λ^2	H_0 0.5	H_0 1.0	H_0 2.0	H_1 0.5	H_1 1.0	H_1 2.0
1	0	0	0	0	0	9
2	0	1	1	0	9	108
3	0	1	32	1	31	189
4	0	0	292	7	64	172
5	0	4	245	6	97	160
6	0	25	138	19	106	99
7	0	170	93	15	114	88
8	0	129	64	16	106	56
9	0	120	39	17	65	35
10	1	126	25	30	60	33
11	1	81	32	33	64	21
12	13	66	15	27	49	13
13	43	58	10	24	47	5
14	56	49	10	23	29	5
15	32	31	2	30	23	4
16	34	24	2	24	25	1
17	41	18	0	23	19	1
18	32	21	1	16	22	2
19	29	17	1	11	21	1
20	21	14	0	23	12	1
21	20	11	1	15	6	0
22	20	7	0	14	6	0
23	20	4	0	13	7	0
24	14	7	0	9	4	0
25	10	3	0	8	4	0
26	11	4	0	11	4	0
27	7	1	0	15	2	0
28	9	1	0	5	2	0
29	10	0	0	5	0	0
30	6	1	0	8	1	0
31	6	2	0	3	1	0
32	8	0	0	4	0	0
33	6	1	0	6	0	0
34	7	0	0	5	0	0
35	6	0	0	7	0	0

p = 2 variables

(cont'd)

Size \ λ^2	H_0 0.5	H_0 1.0	H_0 2.0	H_1 0.5	H_1 1.0	H_1 2.0
36	2	0	0	4	0	0
37	5	1	0	2	0	0
38	6	0	0	3	0	0
39	2	0	0	1	0	0
40	2	0	0	5	0	0
41	2	0	0	3	0	0
42	2	0	0	1	0	0
43	0	0	0	2	0	0
44	1	0	0	3	0	0
45	3	0	0	1	0	0
46	0	0	0	0	0	0
47	1	0	0	2	0	0
48	2	0	0	2	0	0
49	2	0	0	0	0	0
50	1	0	0	1	0	0
51	6	0	0	0	0	0
52	0	0	0	1	0	0
53	0	0	0	1	0	0
54	0	0	0	0	0	0
55	1	0	0	0	0	0
56	0	0	0	0	0	0
57	0	0	0	0	0	0
58	1	0	0	0	0	0
59	0	0	0	0	0	0
60	5	0	0	3	0	0
Totals	501	1006	1005	506	1000	1005
Average n	21.3	10.3	6.0	18.2	9.6	5.3

Frequency Distribution of Sample Size for χ^2-Test

p = 3 variables

$n \backslash \lambda^2$	H_0 0.5	H_0 1.0	H_0 2.0	H_1 0.5	H_1 1.0	H_1 2.0
1	3	0	0	1	0	2
2	4	0	0	0	0	36
3	0	1	2	0	17	77
4	3	1	95	1	29	78
5	0	1	110	6	30	89
6	0	4	94	11	43	52
7	2	40	85	9	46	49
8	2	58	36	21	50	41
9	0	54	23	27	41	25
10	0	58	18	9	37	18
11	2	58	11	22	31	14
12	5	41	6	18	20	11
13	40	39	7	22	21	9
14	59	25	7	33	27	4
15	68	18	2	27	20	3
16	59	23	2	23	15	2
17	61	17	0	27	10	2
18	59	11	1	21	11	0
19	42	11	0	21	6	0
20	54	5	0	17	9	0
21	47	8	0	18	8	0
22	56	4	0	23	7	0
23	79	7	2	11	4	0
24	34	4	0	13	4	0
25	41	2	0	12	4	0
26	29	4	0	11	2	0
27	25	2	0	9	2	0
28	28	2	0	11	0	0
29	28	0	0	11	2	0
30	24	2	0	12	0	0
31	18	1	0	7	0	0
32	16	1	0	4	0	0
33	10	0	0	9	0	0
34	11	0	0	0	2	0
35	14	1	0	11	0	0
36	7	0	0	5	0	0
37	7	0	0	4	0	0

p = 3 variables

(cont'd)

$n \backslash \lambda^2$	H_0 0.5	H_0 1.0	H_0 2.0	H_1 0.5	H_1 1.0	H_1 2.0
38	17	0	0	6	0	0
39	6	0	0	1	0	0
40	8	0	0	3	0	0
41	7	1	0	6	0	0
42	9	0	0	1	0	0
43	4	0	0	1	0	0
44	2	0	0	4	0	0
45	1	1	0	2	0	0
46	3	0	0	2	0	0
47	5	0	0	4	0	0
48	2	0	0	2	0	0
49	1	0	0	4	0	0
50	4	0	0	2	0	0
51	7	0	0	2	0	0
52	0	0	0	2	0	0
53	2	0	0	2	0	0
54	1	0	0	3	0	0
55	1	0	0	1	0	0
56	3	0	0	1	0	0
57	1	0	0	0	0	0
58	1	0	0	0	0	0
59	0	0	0	0	0	0
60	20	0	0	7	0	0
Totals	1035	514	501	543	500	512
Average n	24.0	12.4	6.5	21.2	10.8	5.3

Frequency Distribution of Sample Size for χ^2-Test

p = 5 variables

	H_0			H_1		
$n \backslash \lambda^2$	2.0	5.0	5.0	1.0	3.0	5.0
1	0	0	1	0	4	27
2	0	0	117	0	14	182
3	0	1	261	1	62	128
4	0	47	41	14	87	73
5	0	91	43	22	94	48
6	0	110	18	31	57	24
7	16	67	13	33	59	9
8	30	53	5	41	62	3
9	42	39	1	42	28	3
10	60	38	0	58	24	3
11	54	25	0	37	16	0
12	45	19	0	31	10	0
13	33	10	0	19	10	0
14	33	4	0	24	7	0
15	21	7	0	18	8	0
16	33	3	0	17	6	0
17	12	5	0	25	4	0
18	14	2	0	12	2	0
19	16	1	0	11	2	0
20	15	0	0	6	1	0
21	12	1	0	6	0	0
22	6	0	0	12	0	0
23	13	1	0	13	0	0
24	10	0	0	8	0	0
25	8	1	0	7	4	0
26	6	0	0	5	0	0
27	2	0	0	1	0	0
28	3	0	0	2	0	0
29	2	0	0	4	0	0
30	1	0	0	1	0	0
31	3	0	0	1	0	0
32	1	0	0	3	0	0

p = 5 variables

(cont'd)

$n \backslash \lambda^2$	H_0 2.0	5.0	5.0	H_1 1.0	3.0	5.0
33	3	0	0	1	0	0
34	1	0	0	1	0	0
35	0	0	0	2	0	0
36	2	0	0	0	0	0
37	1	0	0	0	0	0
42	0	0	0	1	0	0
47	1	0	0	0	0	0
49	1	0	0	0	0	0
Totals	500	521	500	500	561	500
Average n	14.5	7.6	3.3	12.5	6.7	3.2

Frequency Distribution of Sample Size for χ^2-Test

p = 9 variables

Size \ λ^2	H_0 2.0	H_0 5.0	H_0 10.0	H_1 2.0	H_1 5.0	H_1 10.0
1	0	0	61	0	37	125
2	2	74	281	3	105	239
3	0	144	116	26	118	103
4	10	136	35	51	109	27
5	48	79	7	64	60	4
6	74	32	0	69	37	2
7	54	16	0	48	15	0
8	93	8	0	52	10	0
9	75	7	0	32	4	0
10	36	1	0	51	2	0
11	32	2	0	24	2	0
12	28	1	0	24	1	0
13	17	0	0	8	0	0
14	5	0	0	14	0	0
15	4	0	0	8	0	0
16	7	0	0	0	0	0
17	7	0	0	8	0	0
18	6	0	0	5	0	0
19	7	0	0	1	0	0
20	4	0	0	4	0	0
21	3	0	0	3	0	0
22	0	0	0	0	0	0
23	0	0	0	0	0	0
24	0	0	0	0	0	0
25	0	0	0	1	0	0
26	0	0	0	0	0	0
27	3	0	0	0	0	0
28	0	0	0	0	0	0
29	0	0	0	0	0	0
30	2	0	0	0	0	0
Totals	517	500	500	500	500	500
Average n	9.1	4.0	2.3	8.0	3.7	2.1

Frequency Distribution of Sample Size for T^2-Test

p = 2 variables

$n \backslash \lambda^2$	H_0 0.5	H_0 1.0	H_0 2.0	H_1 0.5	H_1 1.0	H_1 2.0
3	0	0	2	0	0	4
4	0	0	72	0	0	4
5	0	0	88	0	0	2
6	0	4	78	0	6	57
7	0	48	68	0	4	103
8	0	73	46	0	40	100
9	0	74	46	0	65	73
10	0	53	21	0	59	39
11	0	38	16	0	66	34
12	5	38	13	54	38	23
13	27	34	18	30	32	23
14	33	23	9	29	37	11
15	35	19	9	20	21	10
16	47	21	5	27	28	8
17	28	11	4	27	20	5
18	32	14	0	19	19	3
19	30	11	3	24	20	2
20	26	7	2	19	7	0
21	28	5	3	16	8	2
22	19	11	0	18	9	0
23	15	4	1	19	7	0
24	11	1	1	14	7	1
25	17	1	0	12	6	0
26	17	6	0	20	4	0
27	9	1	0	18	3	0
28	15	3	1	8	4	0
29	12	2	0	10	3	0
30	11	2	0	9	4	0
31	6	1	0	12	2	0
32	11	0	0	11	1	0
33	14	3	0	10	0	0
34	0	0	0	10	2	0
35	3	0	0	5	0	0
36	4	0	0	9	1	0
37	5	1	0	3	0	0
38	5	0	0	5	1	0
39	3	0	0	5	0	0

p = 2 variables

(cont'd)

$n \backslash \lambda^2$	H_0 0.5	H_0 1.0	H_0 2.0	H_1 0.5	H_1 1.0	H_1 2.0
40	5	1	0	4	0	0
41	3	0	0	6	0	0
42	2	0	0	2	0	0
43	6	0	0	5	0	0
44	1	0	0	2	0	0
45	0	2	0	3	0	0
46	4	0	0	1	0	0
47	1	0	0	0	0	0
48	0	0	0	1	0	0
49	1	0	0	0	0	0
50	3	0	0	2	0	0
51	0	0	0	1	0	0
52	0	0	0	1	0	0
53	0	0	0	1	0	0
54	0	0	0	0	0	0
55	0	0	0	1	0	0
56	1	0	0	1	0	0
57	0	0	0	0	0	0
58	0	0	0	0	0	0
59	0	0	0	2	0	0
60	5	0	0	4	0	0
Totals	500	512	500	500	525	500
Average n	22.7	12.4	7.7	23.1	13.9	9.1

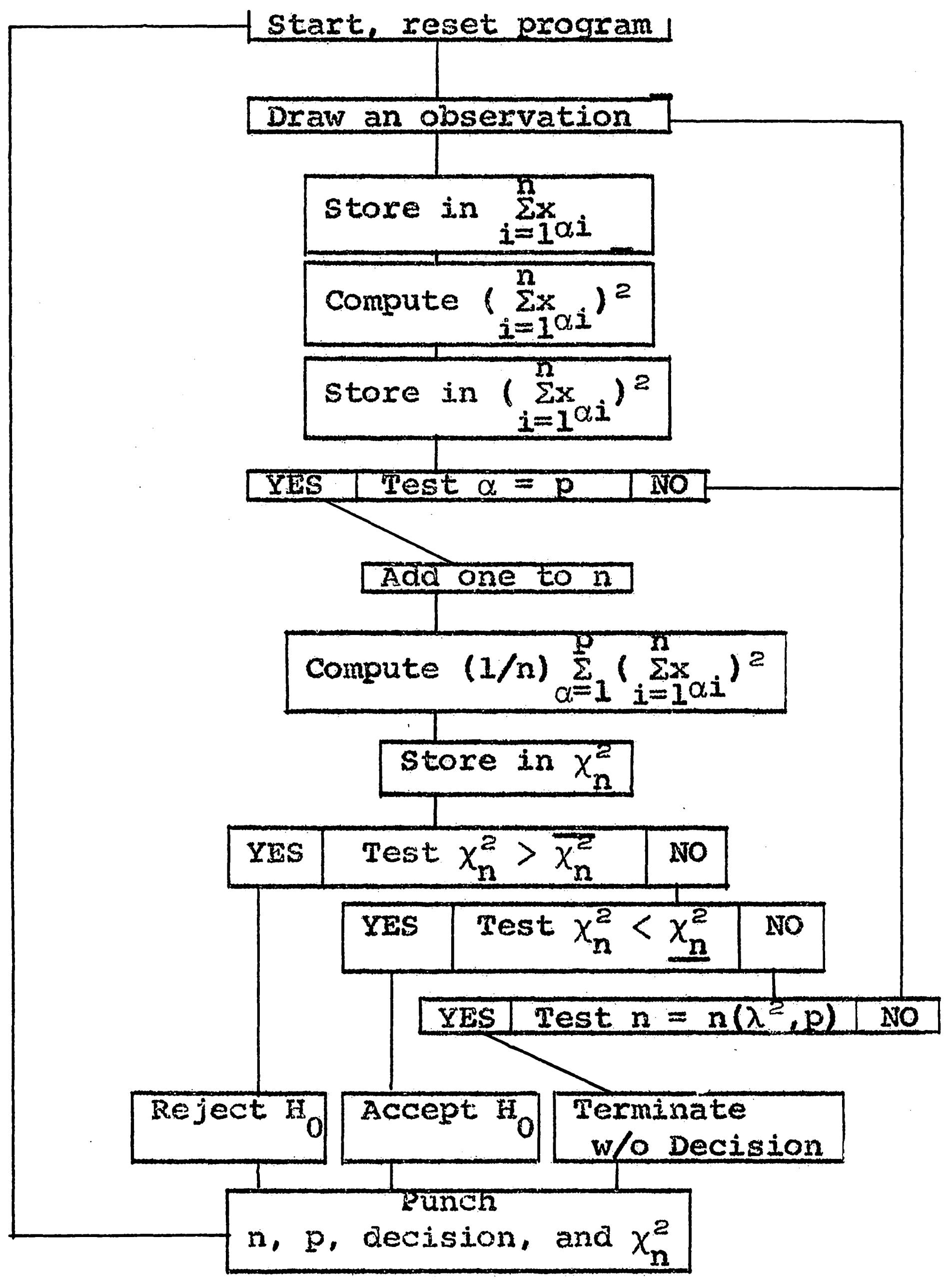
Frequency Distribution of Sample Size for T^2-Test

$p = 9$ variables

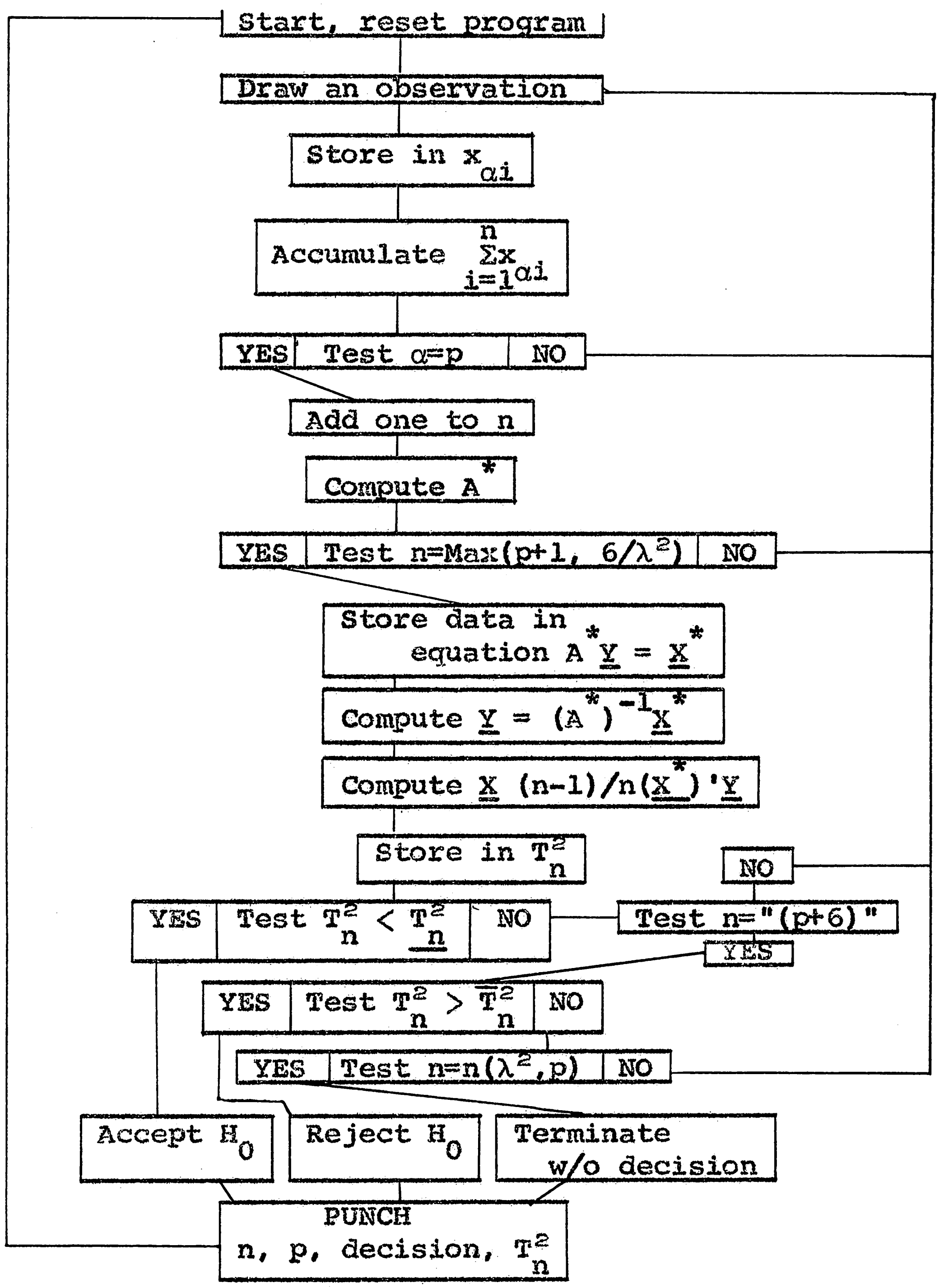
$n \diagdown \lambda^2$	H_0 2.0	H_0 6.0	H_0 10.0	H_1 2.0	H_1 6.0	H_1 10.0
10	9	35	41	1	1	1
11	10	23	23	0	0	1
12	11	13	20	0	9	40
13	9	10	8	0	36	35
14	8	5	5	6	29	9
15	5	5	2	10	12	11
16	10	3	0	15	7	2
17	5	2	1	12	2	0
18	3	2	0	17	2	1
19	5	1	0	7	0	0
20	4	0	0	10	1	0
21	5	0	0	2	0	0
22	6	1	0	7	1	0
23	6	0	0	2	0	0
24	0	0	0	4	0	0
25	0	0	0	3	0	0
26	0	0	0	1	0	0
27	1	0	0	2	0	0
28	1	0	0	1	0	0
29	1	0	0	0	0	0
30	0	0	0	0	0	0
31	1	0	0	0	0	0
Totals	100	100	100	100	100	100
Average n	16.0	11.9	11.2	18.5	14.0	13.0

APPENDIX B

Flow Chart for Sequential χ^2-Test



Flow Chart
for
Sequential T^2-Test



For explanation of symbols see following page:

Where $\underline{X}^*$ is a vector of totals $(\sum_{i=1}^{n} x_{\alpha i})$ with first element zero, and

$$A^* = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \cdots\cdots & \sum_{i=1}^{n} x_{pi} \\ \cdot & \sum_{i=1}^{n} x_{1i}^2 & \cdots\cdots & \sum_{i=1}^{n} x_{1i}x_{pi} \\ \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdots\cdots & \sum_{i=1}^{n} x_{pi}^2 \end{bmatrix}$$

is a matrix of uncorrected sums of squares and products.

The vita has been removed from
the scanned document

ABSTRACT

The purpose of this study is to make an empirical evaluation of Multivariate Sequential Procedures for Testing Means as proposed by J. E. Jackson. This was done by simulated sampling from a multivariate normal population with known means, both when the variance-covariance matrix is assumed known (Sequential χ^2-Test), and when it is assumed unknown (Sequential T^2-Test).

The results indicate the test procedure, either the Sequential χ^2-Test or the Sequential T^2-Test, lead to satisfactory decisions. The tests are conservative but the α and β errors are of an appropriate magnitude. The method (Bhate's Conjecture) used to approximate the Average Sample Number appears to give good estimates of the size of sample needed to obtain a decision.