

ON THE ADEQUACY OF THE  $-2 \log \lambda$   
APPROXIMATION IN MULTIVARIATE ANALYSIS

by

Charles A. Bruce, Jr.

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

MASTER OF SCIENCE

in

STATISTICS

APPROVED:

APPROVED:

---

Director of Graduate Studies

---

Head of Department

---

Dean of Applied Science and  
Business Administration

---

Supervisor

September, 1958  
Blacksburg, Virginia

## TABLE OF CONTENTS

	Page
I. INTRODUCTION	3
II. CHAPTER I, MULTIVARIATE ANALYSIS OF VARIANCE	6
III. CHAPTER II, TESTS OF INDEPENDENCE	21
IV. SUMMARY	31
V. ACKNOWLEDGEMENTS	32
VI. BIBLIOGRAPHY	33
VII. VITA	36

## I. INTRODUCTION

One of the most widely used techniques for the identification of test statistics is the likelihood-ratio method. Quite usually, it leads to optimum or near-optimum tests in univariate analysis. Wilks (16, 17) introduced this method into multivariate analysis and, in this area, the likelihood-ratio statistics are the most widely used ones. Roy (11) was able to show that they are but special cases of a more extensive class of test statistics in multivariate analysis, each member of which shares the same asymptotic properties. One member, the likelihood-ratio statistic, seems to lead to the easiest test statistics. For the construction of simultaneous confidence bounds on parametric functions, it is not very useful (2), and another member of the general class of "union-intersection" statistics, viz, the largest and/or smallest root of certain matrices, seems to be superior for this purpose. In the univariate case (and also in the quasi-univariate situation such as Hotelling's  $T^2$  (6) ), both approaches lead to identical or equivalent test statistics.

A further advantage of the likelihood-ratio statistics is the fact that, in the central case, their distributions can be readily approximated by a short series of  $\chi^2$  or Beta distributions (3, 4, 13, 15, 16). There are, however, several cases where even the central distribution is difficult to obtain; examples occur in Factor Analysis (12), the study of dependence patterns (2, 10, 14, 18) and other situations (see, e. g. (1, 9) ). In such cases the asymptotic

distribution of  $-2 \log \lambda$  is used to advantage. While this approximation is usually quite satisfactory in the univariate analysis, a comparison of results in multivariate analysis between this approximation and the exact technique sometimes seems to indicate rather poor agreement. The use of slight variations of the approximation (3), however, usually leads to quite satisfactory results. In many cases (e. g., the ones cited above) we have no method, at this time, to find such improvements, and are thus compelled to retain the first approximation.

The present report is to serve as a demonstration study. We selected two important applications of multivariate analysis for which exact distributions are known: The multivariate analysis of variance (or "analysis of dispersion" as some authors prefer to call it), and the test of independence of a multivariate normal distribution. We made the following (naturally arbitrary) definition of an "adequately large sample size". If, for an exact probability value of .05, the  $-2 \log \lambda$  approximation produced a value of .045, we considered this agreement as "adequately close" for the .05 level, and noted the sample size for this case. For the comparison of two means in univariate analysis, this size is 57. To obtain the same quality of agreement in a comparison of five mean vectors (four degrees of freedom) in a five-variate study, we need a sample size of 441, (see entry 4, 5 in Table I).

If, for an exact probability value of .01, the  $-2 \log \lambda$  approximation produced a value of .0095, we considered this agreement as "adequately close" for the .01 level, and noted the sample size for this case. For the comparison of two means in univariate analysis,

this size is 185. To obtain the same quality of agreement in a comparison of five mean vectors (four degrees of freedom) in a five-variate study, we need a sample size of 1225 (see entry 4, 5 in Table II).

The tables and charts which show the minimum sample size required for this quality of agreement are intended to serve as a guide for the applied statistician who is confronted with the task of testing a multivariate hypothesis in a case where the exact distribution of the test statistic is unknown or extremely hard to evaluate. He will be able to see whether the sample size at his disposal justifies the use of the  $-2 \log \lambda$  approximation. For certain areas of application, these sample sizes are probably unattainable. In other areas, however, (e. g. psychology, education) it is frequently not too difficult to obtain sample sizes of the required magnitude, and use of the  $-2 \log \lambda$  approximation seems quite justified. Where, as in the case under study, the exact distributions are easily obtained or well tabulated (5), they should be used in preference to the asymptotic technique.

## II. CHAPTER I

## MULTIVARIATE ANALYSIS OF VARIANCE

The likelihood-ratio technique of testing hypotheses in the multivariate extension of analysis of variance leads to the following test statistic (16)

$$(1.1) \quad \mathcal{L} = \frac{|W|}{|W + B|}$$

where  $W$  is a symmetric matrix of sums-of-squares and cross-products corresponding to the "within" sums-of-squares in univariate analysis, and  $B$  is the analogous matrix corresponding to the "between" sums-of-squares. Under the null hypothesis (e.g., in single classification,  $\mu_1 = \mu_2 = \dots = \mu_k$ , where  $\mu_i$  is the vector of expectations in the  $i$ 'th group) the distribution of

$$(1.2) \quad V = -m \ln \mathcal{L}$$

can be approximated (see e.g. (1, 9) ) by

$$(1.3) \quad V : \chi^2(pq \text{ d.f.}) + \frac{Y}{m^2} \left[ \chi^2(pq + 4 \text{ d.f.}) - \chi^2(pq \text{ d.f.}) \right]$$

where

$$m = n - \frac{p+q+1}{2}$$

$$Y = \frac{pq}{48} (p^2 + q^2 - 5)$$

$p$  = number of variates

$q$  = degrees of freedom due to the hypothesis  
(e.g.,  $k-1$  for a  $k$  group, single classification analysis)

$n$  = total degrees of freedom (usually equals total sample size minus one).

It can be shown from the multivariate normal likelihood function,  $\log L$ , and from  $\log L(\hat{\omega})$ , and  $\log L(\hat{\Omega})$ ,

$$(1.4) \quad \frac{2}{N} \log \lambda = \log \frac{|W|}{|W+B|} = \log \mathcal{L}$$

or

$$(1.5) \quad \lambda^{\frac{2}{N}} = \mathcal{L} = \frac{|W|}{|W+B|}$$

In  $\omega$ , we have  $\frac{p(p+1)}{2}$  variances and covariances and  $p$  means. In  $\Omega$ , we have  $\frac{p(p+1)}{2}$  variances and covariances, and  $kp$  means, ( $k$  equals number of groups). Hence:

$$-2 \log \lambda \approx \chi^2(\text{no. of parameters in } \Omega \text{ minus no. of parameters in } \omega)$$

$$-2 \log \lambda \approx \chi^2(kp-p \text{ d.f.})$$

$$(1.6) \quad -2 \log \lambda \approx \chi^2(pq \text{ d.f.})$$

comparing (1.4) and (1.2) we see that

$$(1.7) \quad -2 \log \lambda = -\frac{N}{m} V,$$

where  $N$  is the total sample size. Our objective is to find  $N$  such that the value of  $V$  obtained by inserting the .05 values for the  $\chi^2$  expressions in (1.3) equals the value of  $V$  obtained from (1.6) and (1.7) obtained by inserting the .045 value for the  $\chi^2$  expression in (1.6). We will then say that the sample size is adequately large for using at the .05 level, the approximate procedure. For, it should be noted that for this  $N$ , the  $-2 \log \lambda$  approach will produce a probability of .045 if the true probability is .05. The same procedure will be followed for the .01 level, where we consider the approximation as adequate if the  $-2 \log \lambda$  approach yields a result of .0095 if the exact probability equals .01, and the sample size,  $N$ , needed for this case will be noted.

As a result of comparisons of various methods of obtaining these critical values of .045 and .0095 of  $\chi^2$  with  $pq$  degrees of freedom, Pearson's Tables of the Incomplete Gamma Function (8) were utilized. Quadratic interpolation in the above mentioned tables proved satisfactory to five significant figures.



## Example 1.1

Object: To obtain a satisfactory method for computing the desired  $\chi^2$  values.

Solution: For an exact value of  $\chi^2$  with which to compare the computed  $\chi^2$  's, we will take a tabular value of  $\chi^2$  from Biometrika Tables for Statisticians (7) with certain probability, say .05, and degrees of freedom, say 50;  $\chi^2(.05, 50 \text{ d.f.}) = 67.5048$ .

## Method 1

Computation by inverse interpolation of Table 7 of the Biometrika Tables. For 50 d.f.  $\chi^2 = 68$  corresponds to .04596 and  $\chi^2 = 66$  corresponds to .06418. Inverse interpolation leads to  $\chi^2 = 67.540$  (corresponding to .05).

## Method 2

Computation by use of the cubic formula:

$$\chi^2 = (\text{d.f.}) \left[ 1 - \frac{2}{9(\text{d.f.})} + X \sqrt{\frac{2}{9(\text{d.f.})}} \right]^3$$

where d.f. means degrees of freedom; in our case, 50; and X is the standardized normal deviate corresponding to .05, then

$$\chi^2 = 67.500.$$

## Method 3

Computation by use of the incomplete gamma tables with quadratic interpolation,

$$\chi^2 = 67.506.$$

This kind of comparison (as in Example 1.1), which was made for several specific values, led us to the adoption of Method 3 for the evaluation of the .045 and .0095 levels of  $\chi^2$ .

For every  $(p,q)$ , where  $p$  equals number of variates to be measured on each individual and  $q$  equals the number of degrees of freedom of variation due to hypothesis, we thus determined the minimum sample size for the test of the multivariate analysis of variance by the  $-2 \log \lambda$  approach.

## Example 1.2

Object: To obtain an approximate minimum sample size, we shall make use of only the first approximation of the distribution function of  $V$  (1.3), and after we have found the approximate sample size  $N$ , we will test it by the second approximation (and the third if necessary) to determine the necessary change in  $V$ , if any, so as to adjust  $N$ .

Solution: To determine an approximate  $N$ , we will solve the equation

$$\frac{N}{m}(V) = \chi^2(\text{Cr. value})$$

From (1.3),  $V : \chi^2(pq \text{ d.f.})$

$$m = \left(n - \frac{p+q+1}{2}\right) = \left(N - 1 - \frac{p+q+1}{2}\right)$$

Therefore,

$$\frac{N}{\left(N - \frac{p+q+3}{2}\right)} \left[ \chi^2(\alpha, pq \text{ d.f.}) \right] = \chi^2(\text{Cr. Value})$$

To illustrate, we shall take  $(p,q) = (4, 6)$  at the .05 level:

$$\chi^2(.05, 24 \text{ d.f.}) = 36.415$$

$$\chi^2(.045, 24 \text{ d.f.}) = 36.878$$

Our equation is for  $(p,q) = (4,6)$

$$\frac{N}{(N - 6.5)} (36.415) = 36.878$$

Solving the above for  $N$  we find,  $N = 518$ .

Now we wish to utilize the second approximation of  $V$  (1.3) to determine the necessary change in our approximate value of  $N$ ,

Recalling relation (1.3)

$$V : \chi^2(.05, 24) + \frac{Y}{m^2} \left[ \chi^2(.05, pq+4) - \chi^2(.05, pq) \right]$$

$$Y = \frac{pq}{48} (p^2 + q^2 - 5) = 23.50$$

$$m^2 = \left( N - \frac{(p+q+3)}{2} \right)^2 = (511.5)^2 = 261,632.25$$

$$\chi^2(.05, 28) = 41.337$$

$$\chi^2(.05, 24) = 36.415$$

$$V : 36.415 + .000 = 36.415$$

By substituting these values, we have

$$\frac{518}{511.50}(36.415) = 36.878$$

Therefore, we find our result for  $(p,q) = (4,6)$

**RESULT:**       $(4,6)$                $N = 518$

We intended utilizing Hotelling's  $T^2$  for the special case of  $(p, 1)$  in which we have  $p$  variates as previously, but only two groups, giving one degree of freedom for variation due to hypothesis. However, since available tables of the  $F$  distribution and the Tables of the Incomplete Beta Distributions proved inadequate when the degrees of freedom exceeded 30, we utilized the same method as before for those values.

Description of Tables I and II. Tables I and II contain the minimum sample sizes which justify the  $-2 \log \lambda$  approximation in multivariate analysis of variance (in the sense described above). Table I contains the minimum sample sizes for  $(p,q)$ , where  $p$  = number of variates and  $q$  = degrees of freedom of variation due to the hypothesis, at the .05 level of significance for  $(p,q)$  from (1,1) to (10, 10). Table II contains the minimum sample sizes at the .01 level for  $(p,q)$  from (1,1) to (10, 10).

Instruction for use: To find the minimum sample size necessary for testing a multivariate hypothesis by  $-2 \log \lambda$  approximation at the  $\alpha$  level of significance for a set of  $p$  variates and  $(q+1)$  groups, first find the table for  $\alpha$ , next find the value,  $p$ , in the left-hand column, and read across (directly right) to the column headed by the value  $q$ . At this point is the minimum sample size necessary for  $(p,q)$  at the  $\alpha$  level of significance.

Example: Find the minimum sample size necessary for testing, by the  $-2 \log \lambda$  criterion, the multivariate hypothesis for 7 variates and 5 groups at the .05 level of significance. In this case,  $p = 7$ , and  $q = (k-1) = 4$ ; selecting Table I ( $\alpha = .05$ ), we read down the left-hand column (headed No. of Variables) to the value 7, and directly right to the column headed 4, and we find the size  $N = 595$ . Therefore we should have a sample of about 600 in order to justify the  $-2 \log \lambda$  approximation in a study of this size.

Note: For a certain  $(p, q)$  in proximity to those listed, Charts I and II may be helpful for obtaining an approximate value of  $N$ , which may be read from extensions of the curves shown. However, the reader should exert caution in such a case since investigation beyond the extent of the curves is lacking.

Table I

## Multivariate Analysis of Variance

Minimum Sample Sizes for (p, q) by  $-2 \log \lambda$  Criterion  
 (p equals no. of variates, q equals d. f. due to hypothesis)

$\alpha = .05$

p \ q	1	2	3	4	5	6	7	8	9	10
1	57	88	120	153	188	221	262	302	343	385
2	88	134	177	226	275	326	379	435	493	550
3	120	177	237	297	358	422	487	558	629	701
4	153	226	297	369	441	518	595	675	761	847
5	188	275	358	441	527	614	704	798	894	991
6	221	326	422	518	614	713	814	918	1025	1135
7	262	379	487	595	704	814	927	1045	1160	1280
8	302	435	558	675	798	918	1045	1170	1295	1430
9	343	493	629	761	894	1025	1160	1295	1435	1585
10	385	550	701	847	991	1135	1280	1430	1585	1740



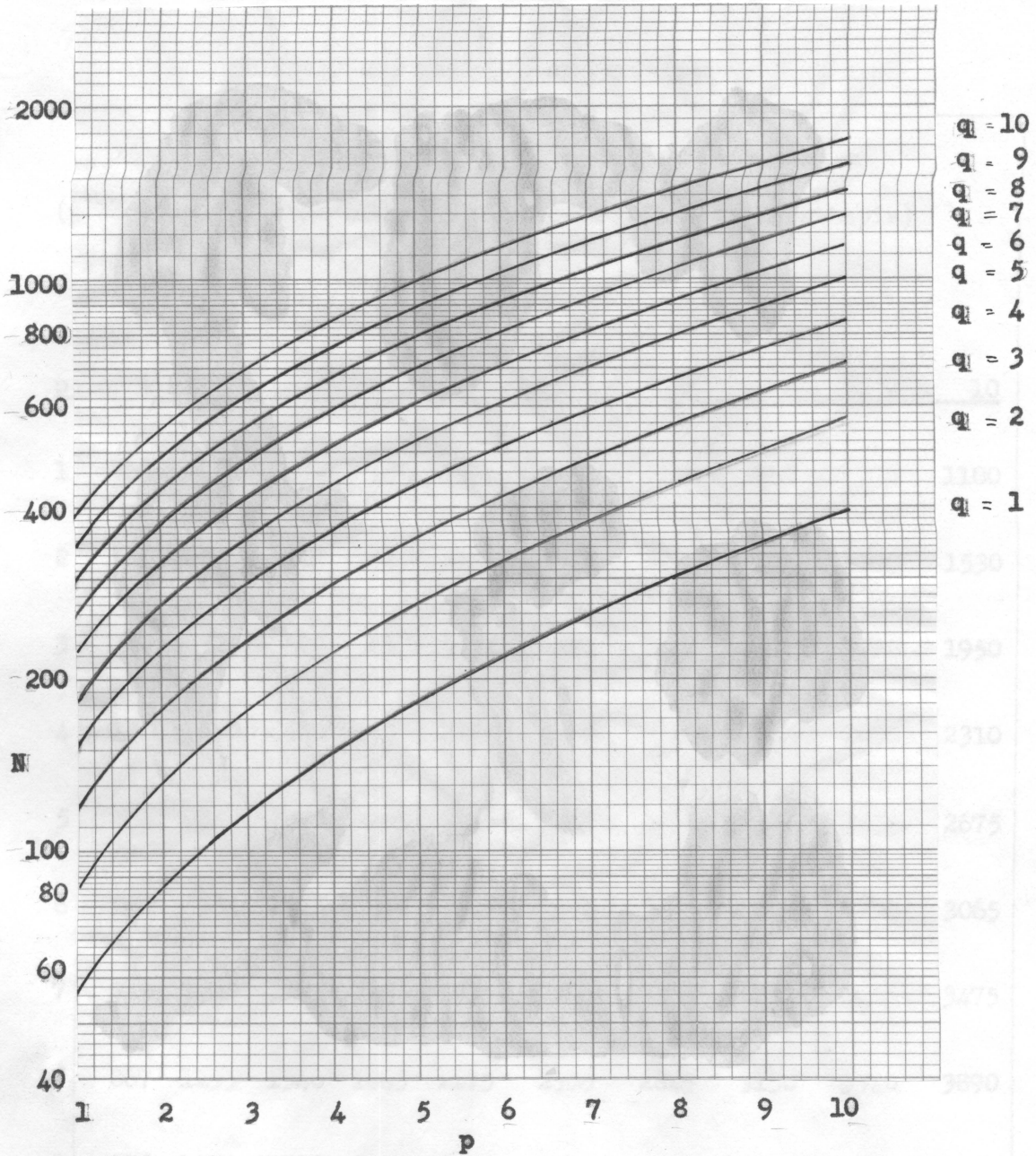


Chart I

Multivariate Analysis of Variance

Minimum Sample Sizes for  $(p, q)$  at .05 level

( $p$  = Number of Variates,  $q$  = d.f. due to hypothesis)

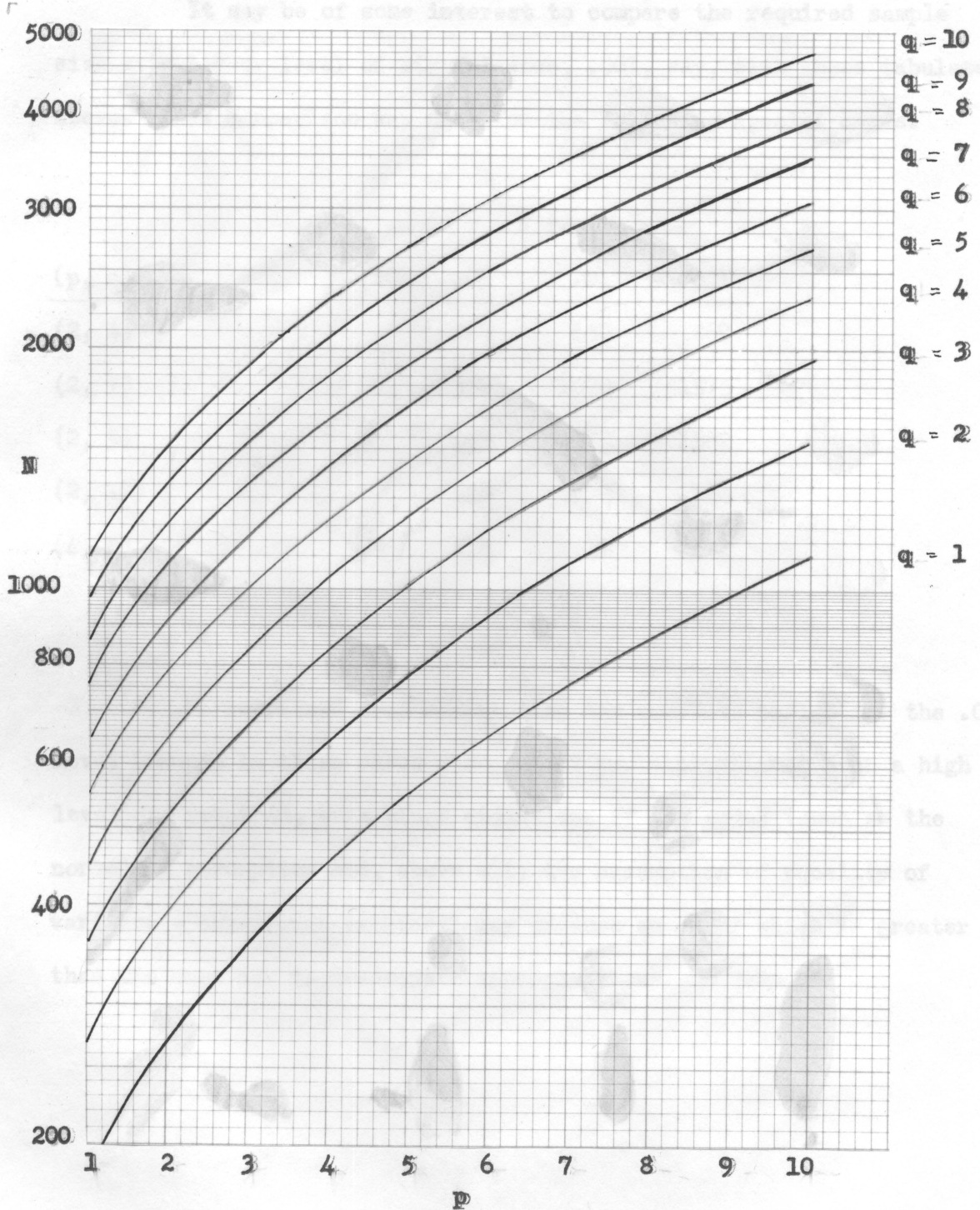
Table II

## Multivariate Analysis of Variance

Minimum Sample Sizes for (p, q) by  $-2 \log \lambda$  Criterion  
 (p equals no. of variates, q equals d. f. due to hypothesis)

$\alpha = .01$

p \ q	1	2	3	4	5	6	7	8	9	10
1	183	270	364	450	556	652	759	867	978	1100
2	270	394	521	650	784	924	1070	1235	1370	1530
3	364	521	677	840	1020	1175	1355	1540	1730	1950
4	450	650	840	1045	1225	1430	1655	1865	2090	2310
5	556	784	1020	1225	1455	1705	1915	2175	2455	2675
6	652	924	1175	1430	1705	1960	2225	2505	2795	3065
7	759	1070	1355	1655	1915	2225	2535	2845	3145	3475
8	867	1235	1540	1865	2175	2505	2845	3150	3510	3890
9	978	1370	1730	2090	2455	2795	3145	3510	3895	4305
10	1100	1530	1945	2310	2675	3065	3475	3890	4305	4630



**Chart II**  
**Multivariate Analysis of Variance**  
**Minimum Sample Sizes for (p, q) at .01 level**  
 (p = number of variates, q = d.f. of variation due to hypothesis)

It may be of some interest to compare the required sample size at a higher level of significance, .001, say, with those tabulated above. This comparison was made for the following special cases.

<u>(p, q)</u>	N for $\alpha =$		
	<u>.05</u>	<u>.01</u>	<u>.001</u>
(2, 4)	226	650	869
(2, 7)	379	1070	1575
(2, 9)	493	1370	1882
(2, 10)	550	1530	2186
(4, 6)	518	1430	2471

We refrained from calculating the complete tables for the .001 level because in those cases where exact probabilities at such a high level are required, even minor violations of the model (such as the normality assumption and, above all, the assumption of equality of variance - covariance matrices) may produce an error which is greater than the ones due to inadequate approximations.

## III. CHAPTER 11

## TESTS OF INDEPENDENCE

In this chapter, we wish to determine the minimum sample sizes for tests of overall independence in a set of  $p$  variates which justify the  $-2 \log \lambda$  approximation.

Denoting the sample correlation matrix by  $R$ , it can be shown that

$$(2.1) \quad -2 \log \lambda = -N \log |R| = -N \log V$$

where  $V = |R|$

The exact distribution of  $-m \log V$  (see, e. g., (1)) can be approximated by

$$(2.2) \quad -m \log V : \chi^2\left(\frac{p(p-1)}{2} \text{ d.f.}\right) + \frac{\gamma_2}{m^2} \left[ \chi^2\left(\frac{p(p-1)}{2} + 4\right) - \chi^2\left(\frac{p(p-1)}{2}\right) \right]$$

where

$$m = N - \frac{2p+11}{6}$$

$$\gamma_2 = \frac{p(p-1)}{288} (2p^2 - 2p - 13)$$

The  $-2 \log \lambda$  approximation yields

$$(2.3) \quad -N \log V : \chi^2\left(\frac{p(p-1)}{2} \text{ d.f.}\right)$$

As in Chapter I, we substitute the .05 values for  $\chi^2$  in (2.2) and determine that value of  $N$  which produces  $\chi^2$ -value corresponding to .045 in (2.3). Analogously, we obtain that  $N$  which, for an exact value of .01, produces an approximation of .0095 in (2.3).

For values of  $p$ , where  $p$  = number of variates, we shall determine the minimum sample sizes for testing the overall independence in a set of  $p$  variates in order to justify the  $-2 \log \lambda$  approximation.

As in Chapter I, the values of  $\chi^2$  for .045 and .0095 probability are computed by use of the incomplete gamma tables (8), wherever possible. For degrees of freedom exceeding 100, the cubic formula of Example 1.1, Method 3, was used.

## Example 2

To obtain an approximate sample size, we shall make use of only the first approximation of the distribution function of  $-m \log V$ , and after we have found the approximate minimum sample size  $N$ , we will test it by the second approximation to determine the change in  $-m \log V$ , if any, so as to adjust  $N$  if necessary.

To determine an approximate  $N$ , we will solve the equation

$$\frac{N}{m} (-m \log V) = \chi^2(\text{Cr. Value})$$

recalling from (2.2)

$$-m \log V : \chi^2\left(\frac{p(p-1)}{2} \text{ d.f.}\right), \text{ and}$$

$$m = N - \frac{2p+11}{6}$$

Therefore

$$\frac{N}{N - \frac{2p+11}{6}} \left[ \chi^2\left(\alpha, \frac{p(p-1)}{2} \text{ d.f.}\right) \right] = \chi^2(\text{Cr. Value})$$

To illustrate, let us take  $p = 5$  at the .01 level.

$$\chi^2(.01, \frac{5(4)}{2} \text{ d.f.}) = \chi^2(.01, 10 \text{ d.f.}) = 23.209$$

$$\chi^2(.0095, 10 \text{ d.f.}) = 23.358$$

Our equation for determining approximate N for  $p = 5$  is

$$\frac{N}{(N - 21/6)} [ (23.209) ] = 23.358$$

and solving the above for N, we find  $N = 549$ .

Now we wish to utilize the second approximation of  $-m \log V$  to determine the necessary change in N, if any:

$$\gamma_2 = \frac{p(p-1)}{288} [ (2p^2 - 2p - 13) ] = 1.875$$

$$m^2 = \left( \frac{N-21}{6} \right)^2 = (545.5)^2 = 297,570.25$$

$$\chi^2(.01, 10 \text{ d.f.}) = 23.209$$

$$\chi^2(.01, 14 \text{ d.f.}) = 29.141$$

$$-m \log V : 23.209 + 0.000 = 23.209$$

On substituting these values we have

$$\frac{549}{545.5} [ (23.209) ] = 23.358$$

Therefore we find our result for  $p = 5$ ,

**RESULT:**  $p = 5, \quad N = 549$



Description of Tables III and IV. Tables III and IV enable us to find the minimum sample size necessary for the  $-2 \log \lambda$  approximation in testing a set of  $p$  variates for overall independence at the .05 (Table III) and the .01 (Table IV) levels of significance, ranging for  $p$  (number of variates) from  $p = 2$  to  $p = 100$ .

Instruction for use. To find the minimum sample size necessary for testing overall independence in a set of  $p$  variates by the  $-2 \log \lambda$  approximation at a desired significance level  $\alpha$ , first find the appropriate table for  $\alpha$ , next find the value of  $p$  in the left-hand column, and read across from this value to the minimum sample size  $N$  directly right of  $p$ .

Example. Find the minimum sample size necessary for testing, by the  $-2 \log \lambda$  approximation, the independence in a set of 20 variates at a level of significance of .01. Selecting Table IV ( $\alpha = .01$ ), we read down the column headed "Number of Variates" to 20, and directly right of 20, we find  $N = 4665$ . Therefore we should have a sample of about 4500 in order to justify the use of the  $-2 \log \lambda$  approximation in a test of independence among 20 variates, at the .01 level.

Notes: For values of  $p$ , which are not included in the Tables III and IV, approximate minimum sample sizes at the .05 and .01 levels may be read from the Charts III and IV which follow Tables III and IV respectively.

Table III

Test of Overall Independence by  $-2 \log \lambda$  Criterion

Minimum Sample Size ( N ) for a Set of p Variates

$$\alpha = .05$$

<u>Number of Variates</u>	<u>Minimum Sample Size</u>
2	57
3	97
4	141
5	194
6	250
7	331
8	383
9	460
10	543
12	727
14	975
16	1160
18	1445
20	1735
25	2580
30	3585
35	4760
40	6100
50	9270
75	20,085
100	37,465

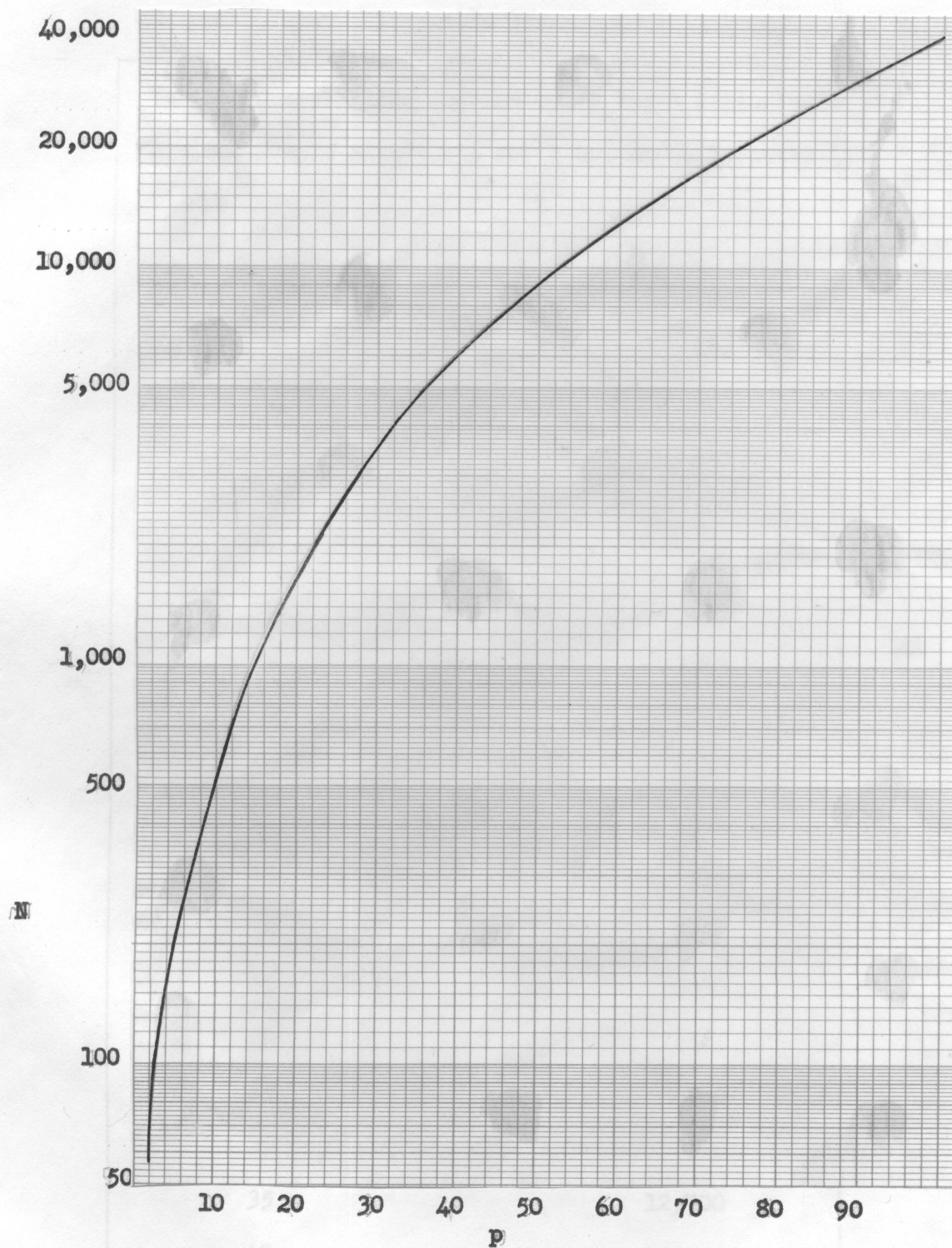


Chart III  
 Test of Overall Independence in a set of  $p$  Variates  
 (Minimum Sample Size ( $N$ ) for Number of Variates ( $p$ ) at .05 level)

Table IV

Test of Overall Independence by  $-2 \log \lambda$  Criterion  
Minimum Sample Size ( N ) for a Set of p Variates  
 $\alpha = .01$

<u>Number of Variates</u>	<u>Minimum Sample Size</u>
2	183
3	295
4	413
5	549
6	710
7	867
8	1065
9	1265
10	1490
12	1970
14	2535
16	3180
18	3890
20	4665
25	6900
30	9590
35	12,700
40	16,235
50	24,620
75	53,210
100	92,640

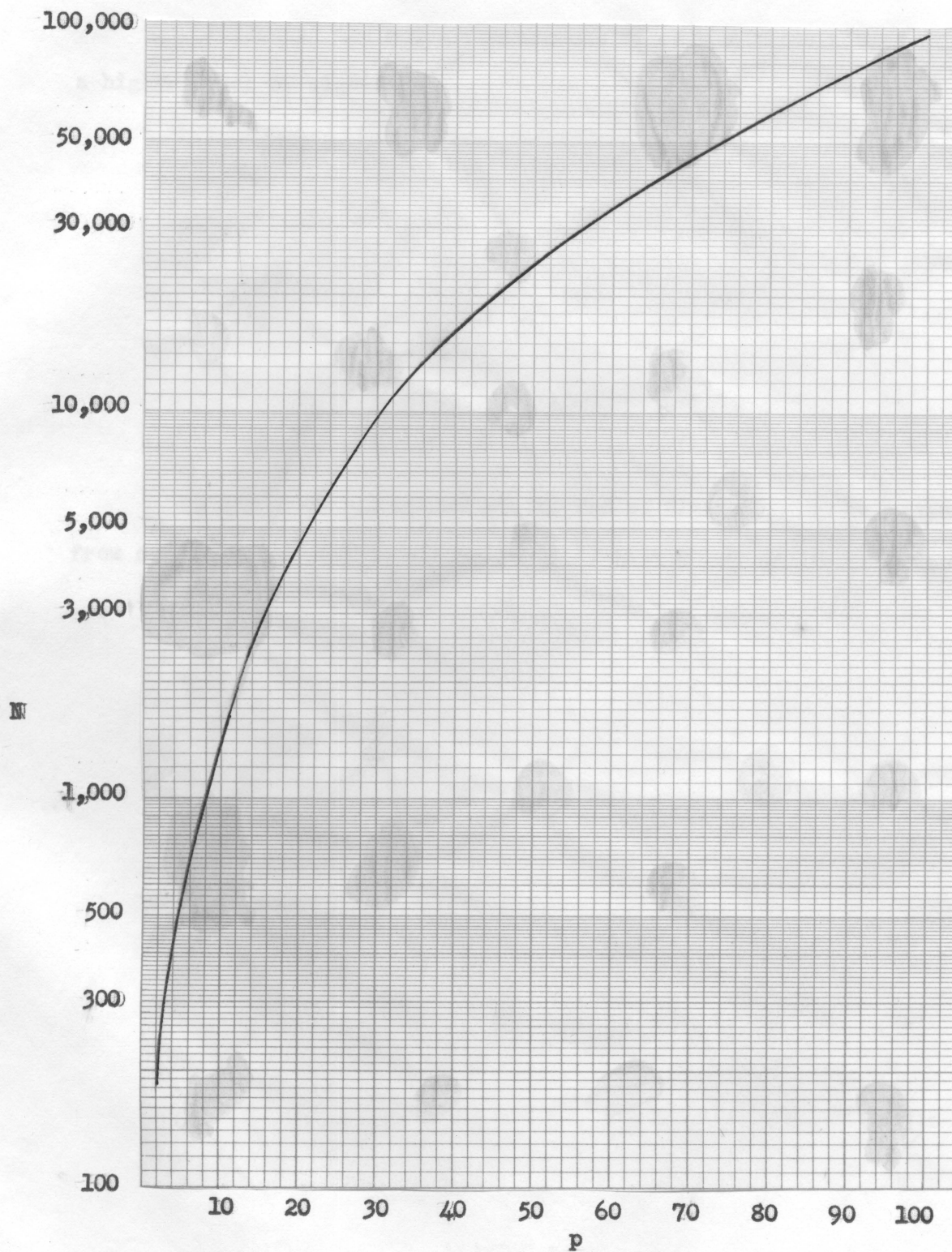


Chart IV  
Test of Overall Independence in a Set of  $p$  Variates  
Minimum Sample Size ( $N$ ) for Number of Variates ( $p$ ) at .01 level

Following are a few comparisons of sample sizes needed at a higher level of significance.

<u>p</u>	<u>.05</u>	<u>.01</u>	<u>.001</u>
3	97	295	422
10	543	1490	1975
20	1735	4665	5874

For the reason stated at the end of Chapter I, we refrained from calculating complete tables for this high level of significance.

## IV. SUMMARY

Exact distributions of statistics for the tests of hypotheses in multivariate analysis of variance and for the test of independence are compared with the asymptotic  $\chi^2$  distribution for  $-2 \log \lambda$ .

"Critical" sample sizes have been recorded which indicate the magnitude of a sample needed so that the approximate technique may produce satisfactory results for testing at the .05 and .01 significance levels.

## V. ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Professor Rolf E. Bargmann for his generous assistance and helpful advice during the preparation of this thesis. The results owe much to his many ideas and suggestions.

The author would also like to acknowledge with thanks the assistance of \_\_\_\_\_ in preparing the final copies for presentation.



## VI. BIBLIOGRAPHY

- (1) Anderson, T. W. An Introduction to Multivariate Statistical Analysis, John Wiley (1958).
- (2) Bargmann, R. E. "A Study of Independence and Dependence in Multivariate Normal Analysis", Inst. of Statistics, Univ. of North Carolina, Mimeograph Series No. 186 (1957).
- (3) Bartlett, M. S. "A Note on the Multiplying Factors for Various  $\chi^2$  Approximations", J. Roy. Stat. Soc. Series B, XVI, (1954) p. 296.
- (4) Box, G. E. P. "A General Distribution Theory for a Class of Likelihood Criteria", Biometrika XXXVI, (1949) p. 317.
- (5) Heck, D. L. "Some Uses of the Distribution of the Largest Root in Multivariate Analysis", Inst. of Statistics, Univ. of North Carolina, Mimeograph Series No. 194 (1958).
- (6) Hotelling, H. "The Generalization of Student's Ratio", Ann. Math. Stat. II (1931) p. 360.
- (7) Pearson, E. S. and Hartley, H. O. (ed.) Biometrika Tables for Statisticians Vol. I, Cambridge University Press, Fourth Edition (1956).
- (8) Pearson, Karl (ed.) Tables of the Incomplete Gamma Function, Cambridge University Press, Second Re-issue (1957).

- (9) Rao, C. R. Advanced Statistical Methods in Biometric Research, John Wiley (1952).
- (10) Roy, S. N. Some Aspects of Multivariate Analysis, John Wiley, (1957).
- (11) Roy, J. "On Some Tests of Significance in Samples from Bipolar Normal Distributions", Sankhya XIV, Part 3 (1954) p. 203.
- (12) Thurstone, L. L. Multiple Factor Analysis, University of Chicago Press (1947).
- (13) Tukey, J. W. and Wilks, S. S. "Approximation of the Distribution of the Product of Beta Variables by a Single Beta Variable", Ann. Math. Stat. XII, p. 318.
- (14) Votaw, D. F. "Testing Compound Symmetry in a Multi-variate Normal Distribution", Ann. Math. Stat. XIX (1948) p. 447.
- (15) Wald, A. and Brookner, R. J. "On the Distribution of the Wilks' Statistic for Testing the Independence of Several Groups of Variates", Ann. Math. Stat. XIII (1944) p. 358.
- (16) Wilks, S. S. "Certain Generalizations of the Analysis of Variance", Biometrika XXIV (1932) p. 471.
- (17) \_\_\_\_\_, "On the Independence of  $k$  Sets of Normally Distributed Statistical Variables", Econometrica III (1935) p. 309.

- (18) \_\_\_\_\_, "Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution", Ann. Math. Stat. Vol. XVII (1946) p. 257.

**The vita has been removed from  
the scanned document**

**ABSTRACT:**

ON THE ADEQUACY OF THE  $-2 \log \lambda$   
APPROXIMATION IN MULTIVARIATE ANALYSIS

by

Charles A. Bruce, Jr.

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

MASTER OF SCIENCE

in

STATISTICS

September, 1958

Blacksburg, Virginia

## ALSTRACT

This thesis compares the .05 and .01 points of several exact distributions used in multivariate analysis with the  $-2 \log \lambda$  approximation for the corresponding likelihood-ratio statistics.

A set of tables and graphs contains the minimum sample sizes required to justify the  $-2 \log \lambda$  approximation in the following sense: If, for an exact probability of .05 the  $-2 \log \lambda$  approximation produced a value of .045 the agreement was considered adequately close, and the sample size producing this result was recorded. At the .01 level, a value of .0095 was considered appropriate.

In the comparison of two means in univariate analysis, a sample size of 57 produces the desired agreement at the .05 level. In multivariate situations, the samples have to be considerably larger. Their sizes have been recorded for the multivariate analysis of variance, and the test of independence.