PRELIMINARY RESULTS ON A ROBUST TEST FOR VARIANCES

Prepared Under Contract No. DA-36-034-ORD-1177 (RD)
(Experimental Designs for Industrial Research)

by

G. E. P. Box
S. L. Andersen

Institute of Statistics
Mimeo Series No. 73
June, 1953

ABSTRACT

In testing the equality of several variances, it has been shown that the probabilities given by standard tests, derived on the assumption that the samples are drawn from a normal parent population, are subject to large errors when the normality assumption is not satisfied. Therefore a modified form of the F test for testing equality of two variances, when the means are known has been developed. This modified test involves calculating the standard F, but looking it up with degrees of freedom δn_1 and δn_2 rather than n_1 and n_2, where

$$\delta = \frac{1}{1 + \frac{1}{2} C_2} \cdot \frac{N-1}{N} - \frac{2}{N}$$

n_1, n_2 = number of observations in samples 1 and 2 respectively

$C_2 = b_2 - 3$

By means of a sampling experiment with 12,000 values of F calculated from samples drawn from three parent populations, rectangular, normal, and double exponential, the errors in probability levels found in the standard test for the non normal cases are greatly reduced by the modified test. Furthermore, this empirical work indicates this gain in "robustness" of the test is accompanied by a loss of power of only about ten percent by the use of the modified rather than standard test, where the parent population is normal.

1. INTRODUCTION

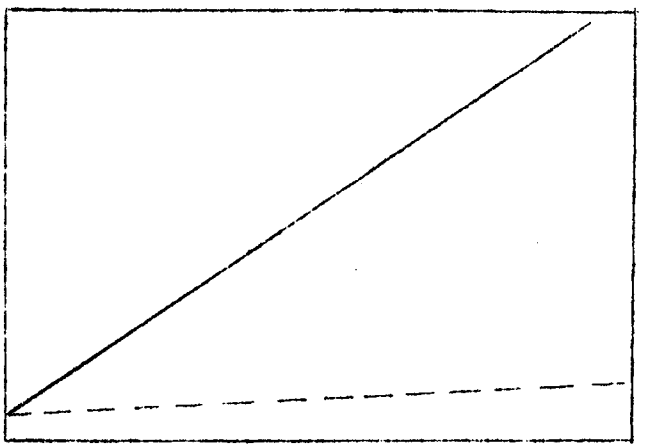
1.1 Problem

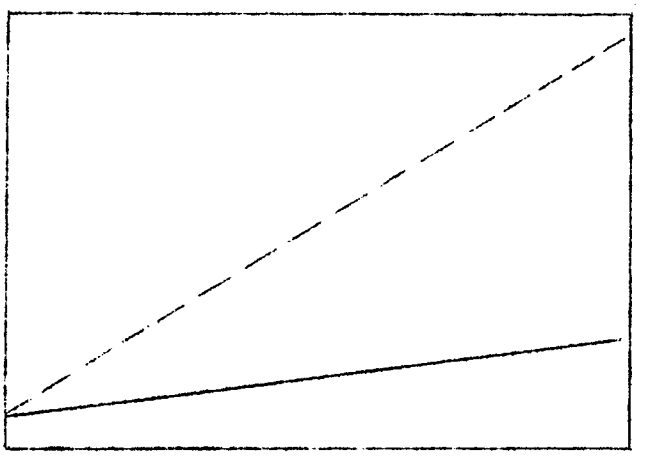
Most statistical estimation and hypothesis testing procedures are based on certain assumptions concerning the parent population. For example, the analysis of variance test for the one way classification assumes that the samples are drawn from k normal populations with equal variances; the t test for comparing two means makes a similar assumption and the F test for the equality of two variances assumes normality.

In practice assumptions such as normality of the parent distribution are never exactly true. Therefore, the justification for the use of procedures must depend on lack of sensitivity to departures from assumptions. If the tests do not possess this insensitiveness, they can be of little value to the experimenter.

It has sometimes been suggested that difficulties concerning the assumptions can be avoided by use of a preliminary test to test the assumptions underlying the main test; that this procedure is unsatisfactory may best be illustrated graphically. The two solid lines in Figure 1 describe the behavior of the preliminary test as a buffer for the main test when there exists a situation in which it is likely that departures from the assumptions will be detected, yet such departures will have scarcely any effect on the main test. If the dotted lines obtain, the situation is one in which the preliminary test has very little chance of detecting departures from the assumptions, yet such departures will seriously affect the main test. Obviously neither of these situations is tolerable.

Figure 1


Chance of Detecting Failure of the Assumptions by Preliminary Test
Degree of Departure of Parent Population from Assumptions


Amount of Interference with Main Test Caused by Departure from Assumptions
Degree of Departure of Parent Population from Assumptions

Even if a satisfactory preliminary test exists, the problem of finding an appropriate main test when the preliminary test warns that the assumptions are not satisfied still exists. It therefore seems necessary to reject the idea of preliminary tests.

The difficulty can be completely avoided by the use of non-parametric tests. This is often a good method. In other situations the loss in power and increase in computation involved in selecting a non-parametric test in favor of the standard test may be rather serious.

1.2 Robust Tests

A compromise solution to this problem would be the use of a test which, although not completely independent of the assumptions concerning the parent population, is sufficiently insensitive to departures from the assumptions. For example it has been demonstrated that the analysis of variance for the comparison of several means is relatively insensitive to normality, and is, in fact, non-parametric to order N^{-1}.

For example, consider a statistic λ, with critical value λ_α which is derived on the assumption that some nuisance parameter Δ has a value Δ_o in the parent population. In testing the null hypothesis, $\beta = \beta_o$, suppose it is true that:

$$\Pr(\lambda > \lambda_o \mid \Delta = \Delta_o, \beta = \beta_o) = \alpha \tag{1.2:0}$$

where α = the prescribed type I error.

Let

$$(1.2{:}1) \qquad \Pr\left(\lambda > \lambda_o \mid \Delta = \Delta_1,\ \beta = \beta_o\right) = \delta$$

where Δ_1 = an alternative value of the nuisance parameter Δ not equal to the assumed value Δ_o.

The statistic, λ, is said to be "robust"[3] with respect to the nuisance parameter, Δ, if the probabilities, α and δ, are very nearly equal for values of Δ different from the assumed value, Δ_o, and of a magnitude likely to be met in practice while the test is still able to detect reasonable deviations from the null hypothesis. In some cases the standard tests possess this property of robustness, while in others the standard tests clearly do not have this property. In the latter case modifications of the standard tests must be developed which are robust.

The first problem therefore is to determine classes of standard tests which are not robust in order that modified tests may be developed where they are most needed. At the same time robust standard tests may be given a clean bill of health and used by the practicing statistician without trepidation. With this purpose in mind, the robustness of two classes of tests will be investigated.

1.3 Two Classes of Tests

In an experiment in which k sets are drawn, with n_t samples per set, there are two hypotheses which are often tested:

$$H_{01}: \quad \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

(1.3:0)

$$H_{02}: \quad \mu_1 = \mu_2 = \ldots = \mu_k$$

The test for equality of two variances and the test for equality of k means both results in the F test and have both been referred to as the variance ratio test. This similarity in outward appearance is unfortunate because the two tests are intrinsically quite different.

In the general case, $k \neq 2$, it is convenient to consider these tests in the form:

(1) For variances, the Bartlett criterion

$$M_1 = N \ln s^2 - \sum n_t \ln s_t^2 \qquad (1.3:1)$$

(2) For means, the logarithm of the likelihood ratio test

$$M_2 = (\phi + k - 1) \ln (1 + X/\phi) \qquad (1.3:2)$$

where

$$X = (k - 1) F = \left(\sum n_t (\bar{y}_t - \bar{y})^2 \right) / s^2$$

By an argument of Neyman and Pearson[5] it may be shown that under the assumption of normality both of these statistics are asymptotically distributed as chi square with k - 1 degrees of freedom.

However in the general non-normal case, Box[3] has shown

M_1 is asymptotically distributed as $\left[1 + \frac{1}{2}\gamma_2\right] \chi^2_{k-1}$

M_2 is asymptotically distributed as χ^2_{k-1}

where

$\gamma_2 = \beta_2 - 3$ = normalized fourth cumulant

Here it can be seen that these two classes of tests are strikingly different. The asymptotic behavior of the test on means (analysis of variance) in the general case is no different from that in the normal case, whereas this is decidedly not true for the test on variances (F test for two variances and Bartlett test). The following tables taken from the work of Gayen (4) and Box (3) illustrate these points.

Table 1

Probability of Exceeding Critical Value, Nominal $\alpha = 5.00$

γ_2 \ γ_1^2 *	0	1	2
-1.5	5.36		
0	5.00	5.10	5.20
+2.0	4.52	4.62	4.72

based on analysis of variance test for equality of means with 5 groups of 5 observations each, giving F (4,20).

* γ_1 = normalized third cumulant

Table 2

Probability of Exceeding Critical Value, Nominal $\alpha = 5.00$

γ_2	Number of groups, k	
	2	20
-1	0.56	0.0004
0	5.00	5.00
2	16.6	71.8

based on Bartlett's M_1 criterion for equality of variances for large samples.

These results clearly indicate that comparative tests for variances are not robust with respect to γ_2 as compared with test on means and therefore require modified tests for situations where the assumption of normality is suspect.

It may seem at first glance that values of γ_2 such as those indicated in the above tables are either much larger than would be encountered in real life or if not extremely large, at least readily detected by observing the frequency distribution. That neither of these wishful supositions is true may be seen from a host of published data, for example the Monier-William data[4] on percentage butter fat in milk samples.

2. A ROBUST TEST FOR VARIANCES

The simplest case in which a test may be made for equality of variances is that of two groups with the mean of each group known. There is no loss in generality in assuming that the two known means are zero. The statistic most commonly used to test this hypothesis is F:

$$F = \frac{\sum_{i=1}^{n_1} X_i^2 / n_1}{\sum_{j=1}^{n_2} y_j^2 / n_2} \qquad (2.0{:}1)$$

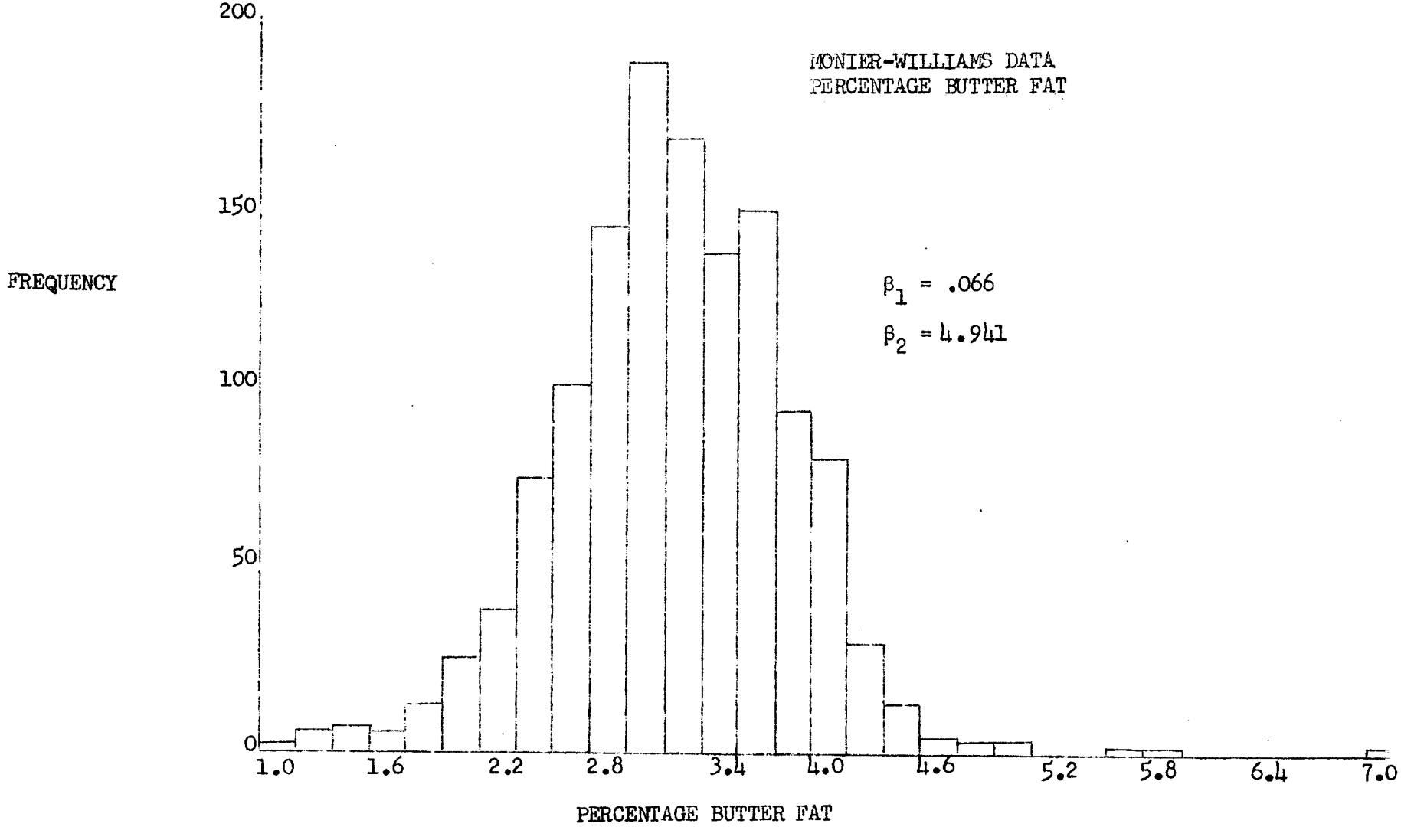
where: $X_1, \ldots, X_{n_1}$ = the n_1 observations of the first sample

$Y_1, \ldots, Y_{n_2}$ = the n_2 observations of the second sample

An equivalant statistic for testing this same hypothesis is w:

$$w = \frac{\sum_{i=1}^{n_1} X_i^2}{\sum_{i=1}^{n_1} X_i^2 + \sum_{j=1}^{n_2} y_j^2} \qquad (2.0{:}2)$$

FIGURE 2


MONIER-WILLIAMS DATA
PERCENTAGE BUTTER FAT
β_1 = .066
β_2 = 4.941
FREQUENCY
200
150
100
50
0
1.0
1.6
2.2
2.8
3.4
4.0
4.6
5.2
5.8
6.4
7.0
PERCENTAGE BUTTER FAT

Since this statistic is an exact equivilent of the more common F, and lends itself ot simpler algebraic manipulation, the theoretical development will be concerned with the distribution of w. As w is distributed as a beta function for samples drawn from a normal population, its first two moments are

$$E(w) = \frac{v_1}{v_1 + v_2} = \mu_1'(w) \qquad (2.0:3)$$

$$v(w) = \frac{2v_1v_2}{(v_1+v_2)^2(v_1+v_2+2)} = \mu_2(w) \qquad (2.0:4)$$

where: v_1, v_2 = degrees of freedom to estimate the variance in sample 1 and 2 respectively. For the case considered $v_1 = n_1$ and $v_2 = n_2$.

An approach to the problem of hypothesis testing which does not involve assumptions concerning the form of the parent distribution was given by R.A. Fisher[6] in 1937. In this method, the argument has come to be associated with the concept of randomization (i.e. the concept of arranging an experimental program in a randomized design). It should be noted, however, that, in fact, this type of test, known as the randomization test, might more pertinently be called a permutation test. Such a test has a wider validity than is commonly recognized. In fact it is only necessary to assume that the samples are drawn from the same distribution. The only restriction on the distribution is that the likelihood is unchanged when the observations are rearranged. For example independent observations from any population, satisfy this condition.

The general argument using a randomization or permutation test is the following:

1. Divide the totality of pairs of samples, S, into subsets S_1, S_2,...,S_t,... such that each subset contains the same N observations. These will be partitioned into two samples of n_1 and n_2 in many ways.
2. By the randomization test α per cent of each subset, S_t, will have H_0 rejected.
3. Since α per cent are rejected for each subset S_t, it follows that α per cent of all the samples, S, will be rejected. This result holds for any form of parent population.

The actual permuting of all the observations in a sample to obtain the actual values is far too tedious and therefore an approximation to the permutation test suggested by Welch[9] by fitting the first two moments is used. Therefore the first two moments are computed and equated to the normal moments calculated above to give the approximate criticalvalues.

Using this approach, the moments of w with respect to the permutation distribution are:

$$E(w) = n_1/N \tag{2.0:5}$$

$$V(w) = \frac{2n_1n_2}{N^2(N-1)}\left(1 + \frac{1}{2}c_2\right) \tag{2.0:6}$$

where:

$$c_2 = b_2 - 3$$

$$b_2 = \frac{N(\sum X^4 + \sum Y^4)}{(\sum X^2 + \sum Y^2)^2}$$

$$N = n_1 + n_2$$

If these moments of the permutation distribution of w are equated to the corresponding normal moments, the following relationship is found:

$$v_1 = n_1\delta$$

where:

$$(2.0:7) \qquad \delta = \frac{1}{1 + \frac{1}{2} c_2} \cdot \frac{N-1}{N} - \frac{2}{N}$$

Since there is a one to one correspondence between the w statistics here investigated and the F statistic used to test the same hypothesis, this result suggests that the appropriate statistic is:

$$(2.0:8) \qquad F(n_1\delta, n_2\delta)$$

rather than the conventional:

$$(2.0:9) \qquad F(n_1, n_2)$$

That is, the appropriate statistic is the conventional F with degrees of freedom modified by the sample fourth moment as indicated.

There are however two questions concerning the F approximation to the randomization test.

1. How good is the moment approximation to the randomization test?
2. It is well known that when the parent distribution is normal, the standard F test is uniformly most powerful. Therefore it is of interest to estimate how much power is lost by using the modified test in order to obtain the property of robustness.

In order to answer these two questions, a rather extensive empirical sampling experiment has been performed.

3. SAMPLING PROCEDURE

In order to study the power function and robustness of the standard F-test and the modified F-test, the power functions and null distributions of these statistics were investigated for the rectangular, normal, and double exponential parent distributions.

The empirical sampling procedure involved drawing 2000 samples size 20 from each of these three populations. These were paired to give 1000 values of F and 1000 values of Box's modified F, F_B, for each population. The appropriate probability associated with each F and F_B was estimated from a set of graphs prepared from Pearson's "Tables of the Incomplete Beta Function"[8]. These probabilities for the two statistics and three distributions are summarized on Charts 2.1, 2.2, and 2.3.

In addition to the calculations above, which describe the behavior of F and F_B when the null hypothesis is true in each of the three populations, the distributions of these statistics wereestimated for three alternatives to the null case:

$$H_1: \ 2\,\sigma_1^2 = \sigma_2^2$$

$$H_2: \ 4\,\sigma_1^2 = \sigma_2^2$$

$$H_3: \ 6\,\sigma_1^2 = \sigma_2^2$$

To obtain the distribution of F and F_B when each of these hypotheses was true, the values in first sample of each pair (the one used in the numerator of the F-test) were multiplied by the appropriate constant and all the calculations referred to above were repeated.

The empirical sampling was performed on IBM electric punched card machines. The random digits used in the experiment were taken from existing punched cards prepared from several published tables.

Master Cards:

First, a set of 100 master cards was prepared. Columns 1-3 contained the rectangular distribution with zero mean, with odd integer values of X from -99 to +99. Similarly the normal variates, y, and exponential variates, z, had their percentile values entered in columns 4-0 and 9-13 respectively, with values running from -99 to +99 in both cases. By means of desk calculator and direct punching the following quantites were put on the cards:

Columns.	14-17	18-21	22-25	26-32	33-39	40-46	47-54	55-62	63-70	71-72
										Sequence NO.(00-99)
Variates.	X^2	y^2	z^2	X^3	y^3	z^2	X^4	y^4	z^4	

Sampling Deck:

A sampling deck of 40,000 cards was prepared in three steps. The 40,000 blank cards first had two digit random numbers entered in columns 71-72, with numbers running from 00 to 99. These random numbers were taken from successive pairs of columns of the existing random digit cards. At the same time the cards were serially numbered by a 6 digit number such as 867006. The first three digits (000-999) indicate to which pair of samples the card belongs. The fourth digit indicates by either 0 or 1 whether the card is the first or second sample in the pair. The last two digits (00-19) indicate the item number within the sample.

On the second step the sampling deck was sorted into 100 groups according to columns 71-72, the two-digit random numbers. Having been thus sorted, all those cards containing random digit 01 were given the first percentile values of the variables, X, X^2, ..., Z^4 in the same columns as master card number one. This procedure was repreated by gang punching all the percentile values on the appropriately sorted group of cards.

Finally this sampling deck was sorted to put the samples back in original sequence according to columns 73-76, thus randomizing the samples from all the three distributions. It is to be noted, however, that samples drawn from this ordered deck will have the samples from the various distributions correlated, but this imperfection is accepted to expedite this preliminary sampling experiment.

Summary Deck:

The 40,000 cards in the sampling deck were passed through the 405 tabulator to prepare a summary punched card for each of the 2000 samples using columns 73-76 as a control group. All the columns were totaled except the columns of cubes. Columns 73-76 were reproduced on the summary cards as sequence numbers.

Statistics Deck:

From the summary deck, the following calculations were performed on the 602-A calculating punch. The computations were performed from figures obtained from successive pairs of cards from the summary deck, making 1000 calculations of each type. One such set of calculations will be indicated, using the subscript, a, for the quantities taken from the first card of each pair and b for those taken from the second card of the pair. Furthermore, only the calculations for the variable X are indicated as the identical calculations will be

required for y and Z

$$(3.0{:}1) \qquad F_i(X) = k_i \frac{\sum X_a^2}{\sum X_b^2} \qquad i = 0, 1, 2, 3 \qquad k_0 = 1$$

$$(3.0{:}2) \qquad b_{2i}(X) = 40 \frac{(k_i^2 \sum X_a^4 + \sum X_b^4)}{(k_i \sum X_a^2 + \sum X_b^2)^2}$$

$$(3.0{:}3) \qquad \delta_i(X)n = \frac{1}{1 + \frac{1}{2} C_{2i}(X)} \cdot \frac{39}{40} - \frac{1}{20}$$

where

$$C_{2i}(X) = b_{2i} - 3$$

The final form of the data from the IBM laboratory was a printed tabulation of the 1000 pairs of values $F_i(\ )$ and $\delta_i(X)n\ (\ )$, one tabulation for each value of i and each of the three variables, making 12 tabulations of 1000 pairs of values. These were printed in order of ascending values of F.

To calculate the 12,000 probability values corresponding to the statistics, F and δn, five graphs were prepared to permit rapid interpolation of the Incomplete Beta Function Table. Several graphs had to be made to permit reading three digits in all regions. Photographs of these are shown as Charts 1.1,...,1.5.

With these data charts 2.1, 2.2, and 2.3 were constructed in order to show the relative robustness of the standard and modified F tests with respect to the rectangular, normal and double exponential distributions. Charts 3.1, 3.2 and 3.3 show the power curves for these two tests for the same three parent distributions.

4. SAMPLING RESULTS

4.1 Verification of Robustness

Before analyzing the results of this experiment it is instructive to note the imperfection in the three populations due to the truncation caused by selecting only one hundred values to approximate the three continuous distributions. Comparisons of the theoretical and actual values of β_2 are shown below.

Value of β_2

	Rectangular	Normal	Double Exponential
Theoretical	1.8	3.0	6.0
Actual	1.7998	2.8340	4.7301

Actually the deviations of these populations from their theoretical counterparts is of no serious consequence, as the purpose of the experiment was to obtain populations representing three degrees of kurtosis centered about the normal.

The results indicating the robust property of the modified F test are shown on Charts 2.1, 2.2 and 2.3. These are frequency distributions for 1000 pairs of samples each showing the frequencies of the probabilities calculated from the standard and modified F statistic when sampling from each of the three parent distributions. For a sufficiently large number of samples these

distributions should be rectangular for a completely robust test. This is equivilent to saying that 5 per cent of the values of F and F_B should occur in each 5 percentile range of the distributions of these statistics.

The charts establish empirically the failure of the normal theory F test for variancesto be robust with respect to deviations of kurtosis from normal. They also establish the relative robustness of the modified F test for these same leptokurtic and platykurtic populations.

For the rectangular parent population, the standard F test shows 0.7 per cent of the values below the 5 per cent point and 0.6 per cent of the values above the 95 per cent point. The modified test corrects almost perfectly for this lack of robustness in the standard test, showing 4.5 per cent of the values below the 5 per cent point and 4.7 per cent of the values above the 95 per cent point.

If the test is behaving properly, fifty samples would be expected in each cell of the chart. The use of chi square to compare the actual frequencies with this ideal behavior illustrates the improvement with the modified test. With nineteen degrees of freedom in each case, the value of chi square is reduced from 254.16 to 21.84 by the modification of the F test; this clearly indicates that the standard F test is not behaving properly, while the observed frequencies of the modified test are compatible with the hypothesis of a robust test.

The results of sampling from all three parent populations may be summarized by the following table.

Table 3

Parent Population	Rectangular		Normal		Double Exponential	
Test	Standard	Modified	S	M	S	M
Per Cent below 5% point	0.7	4.5	3.0	4.0	10.2	3.6
Per Cent above 95% point	0.6	4.7	3.0	3.5	11.0	3.6
Chi Square	254.16	21.84	36.12	27.84	166.20	32.96
Probability of exceeding observed Chi Square	less than .001	.290	.011	0.86	less than .001	.025

The large value of chi square demonstrated by the standard test in the normal case is partially due to the fact that the value of β_2 in the actual population sampled was somewhat less than three, the correct normal value. This is seen by noting that the deviations from the expected frequencies of 50 for each cell are similar to those of the samples drawn from the rectangular parent population, the other platykurtic population.

4.2 Investigation of Power

Having established the robustness of the modified test, the question arises as to how much power is lost by the modification of the standard test, when in fact the parent population is normal. Chart 3.1 shows the power curve for the standard and modified F test for $\alpha = .05$. The power curve for the standard F is shown for both the theoretical power and empirically

determined power. These two agree closely.

From the graph it appears that there is very little loss of power due to modifying the F test. For example, the probability of detecting a variance ratio of $\sigma_1^2 = 3\sigma_2^2$ is reduced from about .75 to about .71 by using the modified test.

The power curve for the standard F test with 18 and 18 degrees of freedom, $\alpha = .05$, coincides almost perfectly with the modified F test curve. This indicates a loss of power of about 10 per cent in the sense of requiring 10 per cent more observations with the modified F than with the standard test to obtain the same power curve.

The power curves for the two tests although of less interest in the case of the non-normal distributions have been shown on Charts 3.2 and 3.3. The curves for the two tests are not directly comparable since the standard test does not have the proper intercept of .05. What can be noted is that the modified test is most powerful for the rectangular and least powerful for the double exponential. That is to say the power decreases with increasing β_2.

V. FURTHER RESEARCH

Theoretical work is now being conducted to develop and justify more general robust tests for variances. The cases under consideration are the testing of the equality k variances where the mean is known, the equality of two variances where the means are unknown, the equality of k variances where the means are unknown.

As these tests are developed, further empirical investigations similar to the one in this report will be conducted.

References

1. Lehman, E.L. (1953), The Power of Rank Tests, A.M.S. 24, pp. 23-43.

2. Gayen, A.K. (1950), The Distribution of the Variance Ratio in Random Samples of Any Size Drawn from Non-Normal Universes, Biometrika 37, p. 236.

3. Box, G.E.P. (1952), Non-Normality and Tests on Variances, Ph.D. Thesis, University College, London. To be published in Biometrika, December 1953.

4. Tocher, J.F. (1928), An Investigation of the Milk Yield of Dairy Cows, Biometrika 20 , p. 106.

5. Neyman and Pearson (1931).

6. Fisher, R.A. (1935), Design of Experiments.

7. Box, G.E.P. (1953) Unpublished.

8. Pearson, K. (1934), Tables of the Incomplete Beta-Function.

9. Welch, B.L. (1937), On the Z-test in Randomized Blocks and Latin Squares, Biometrika 29, p. 21.

Appendix - Charts

Interpolation Charts of Incomplete Beta Function.

Charts: 1.1

1.2

1.3

1.4

1.5

Probability Distributions

2.1 Rectangular Parent Population

2.2 Normal Parent Population

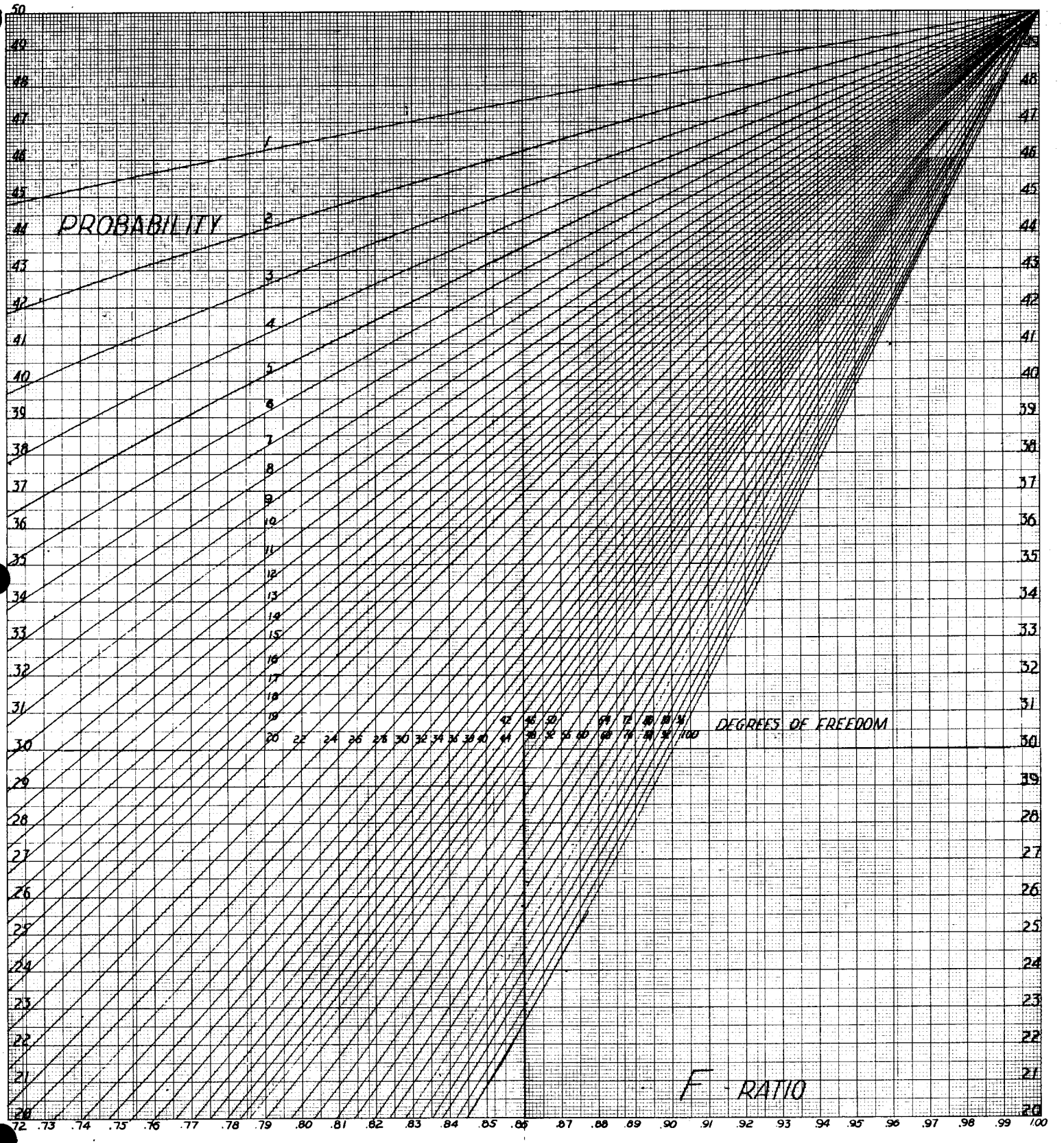
2.3 Double Exponential Parent Population

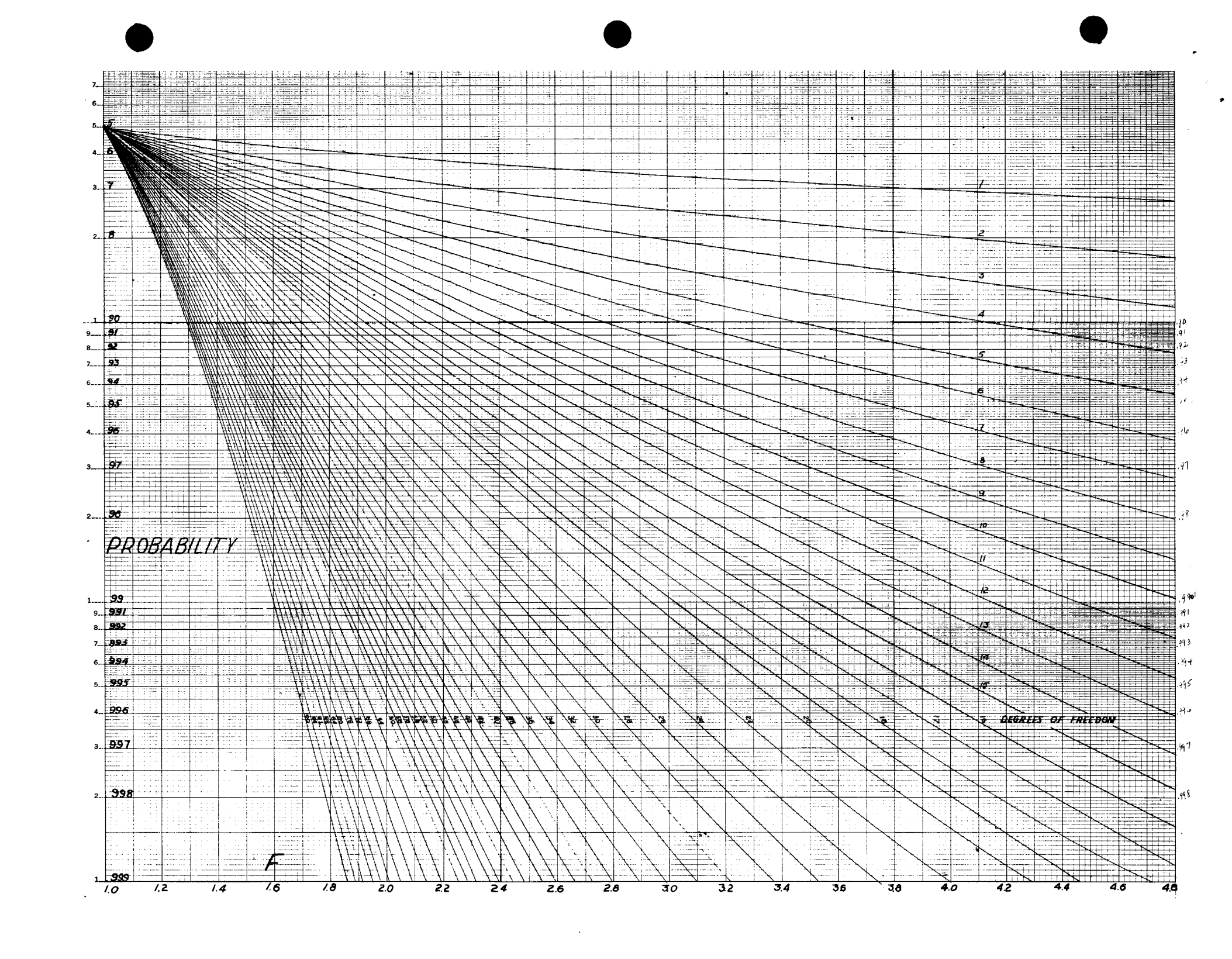
Power Curves

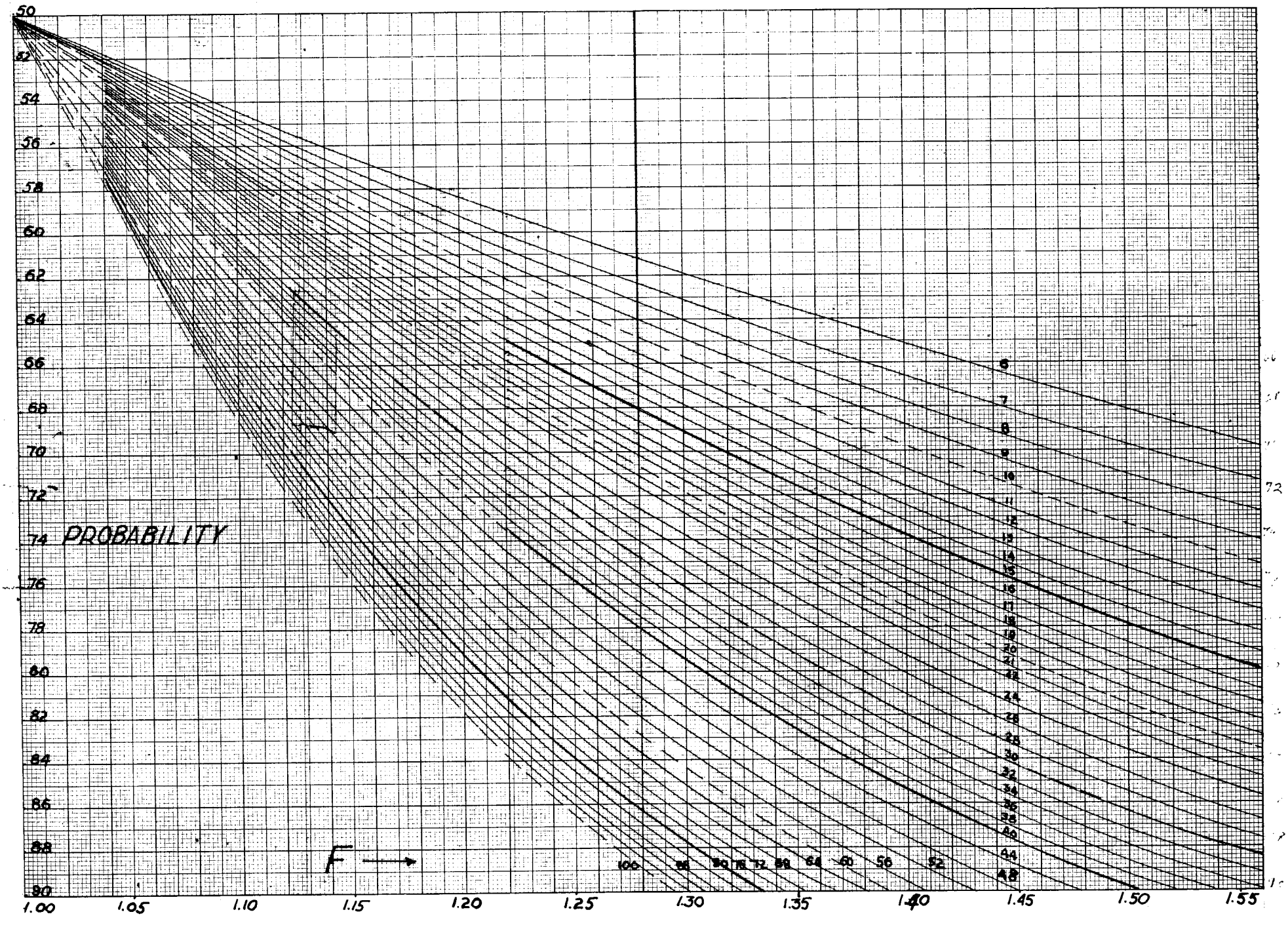
3.1 Normal Parent Population

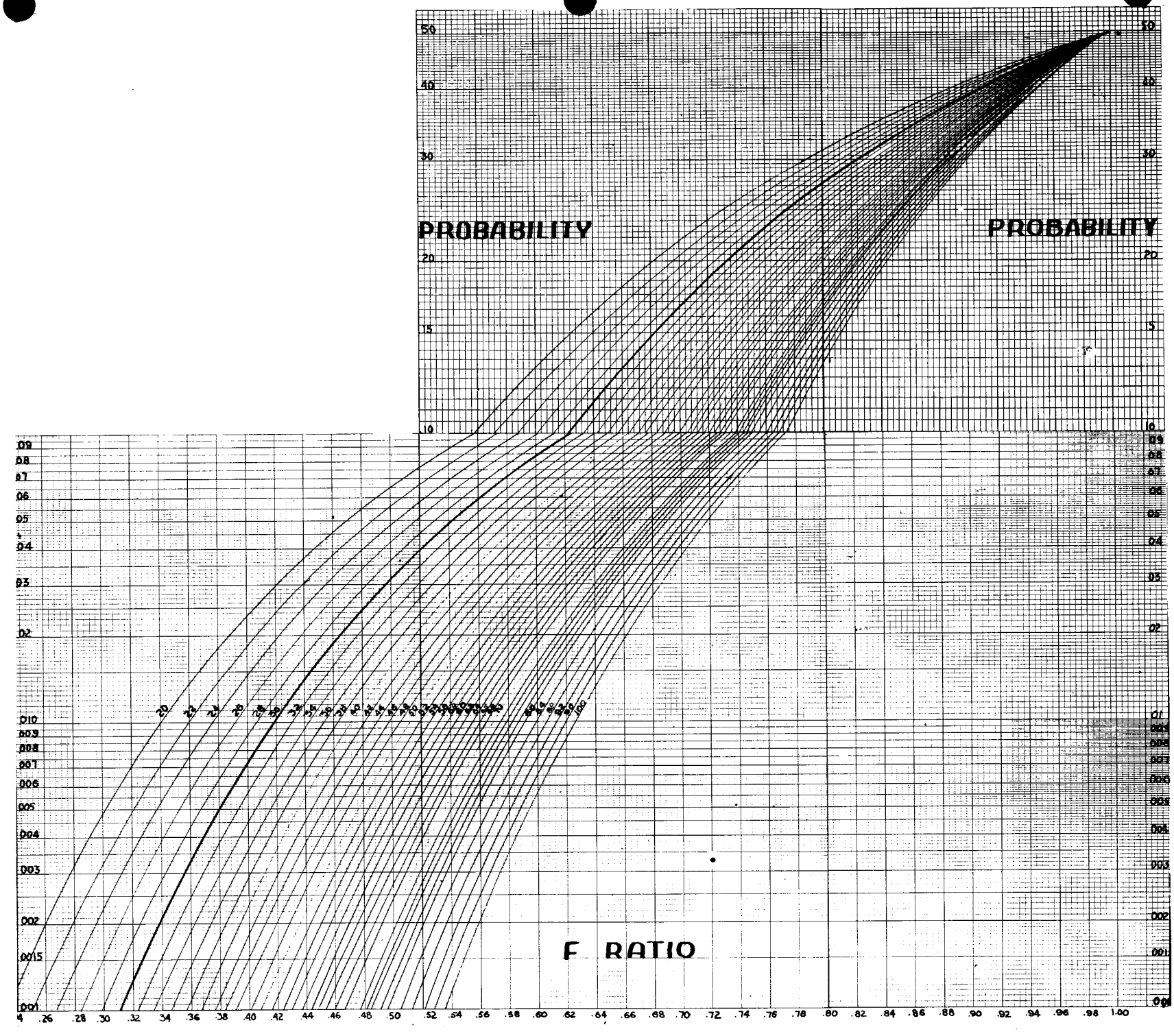
3.2 Rectangular Parent Population

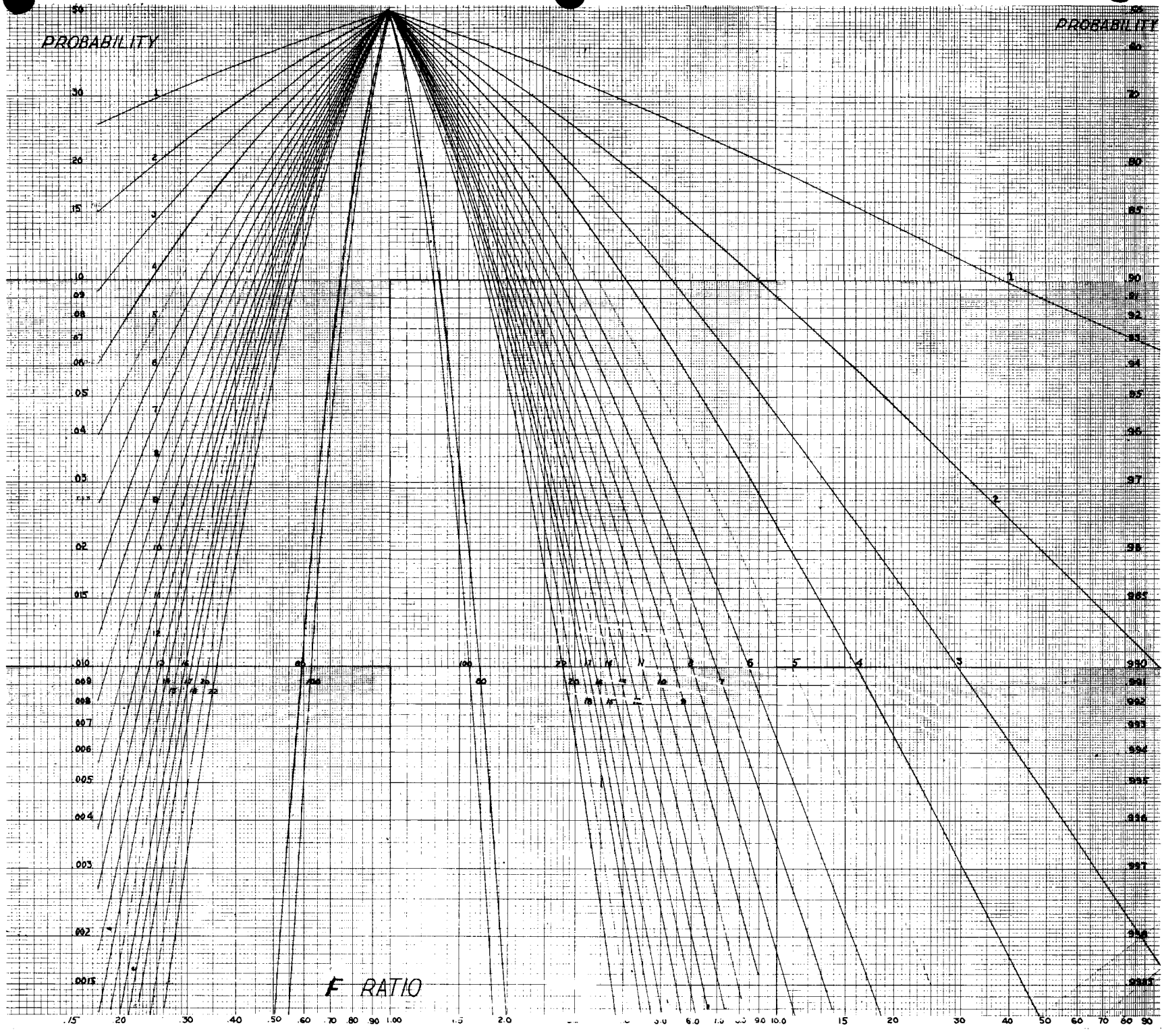
3.3 Double Exponential Parent Population

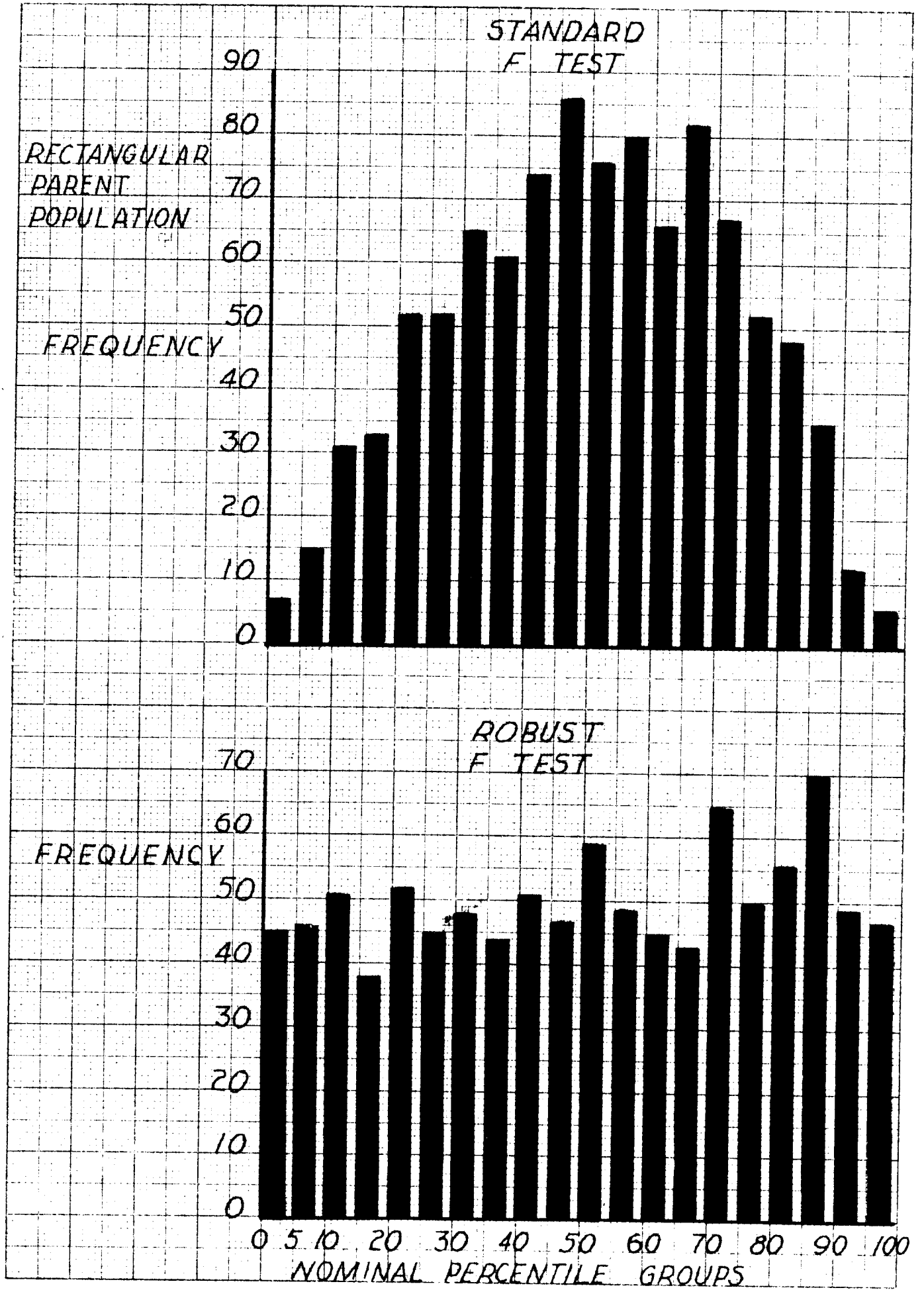

PROBABILITY
DEGREES OF FREEDOM
F - RATIO

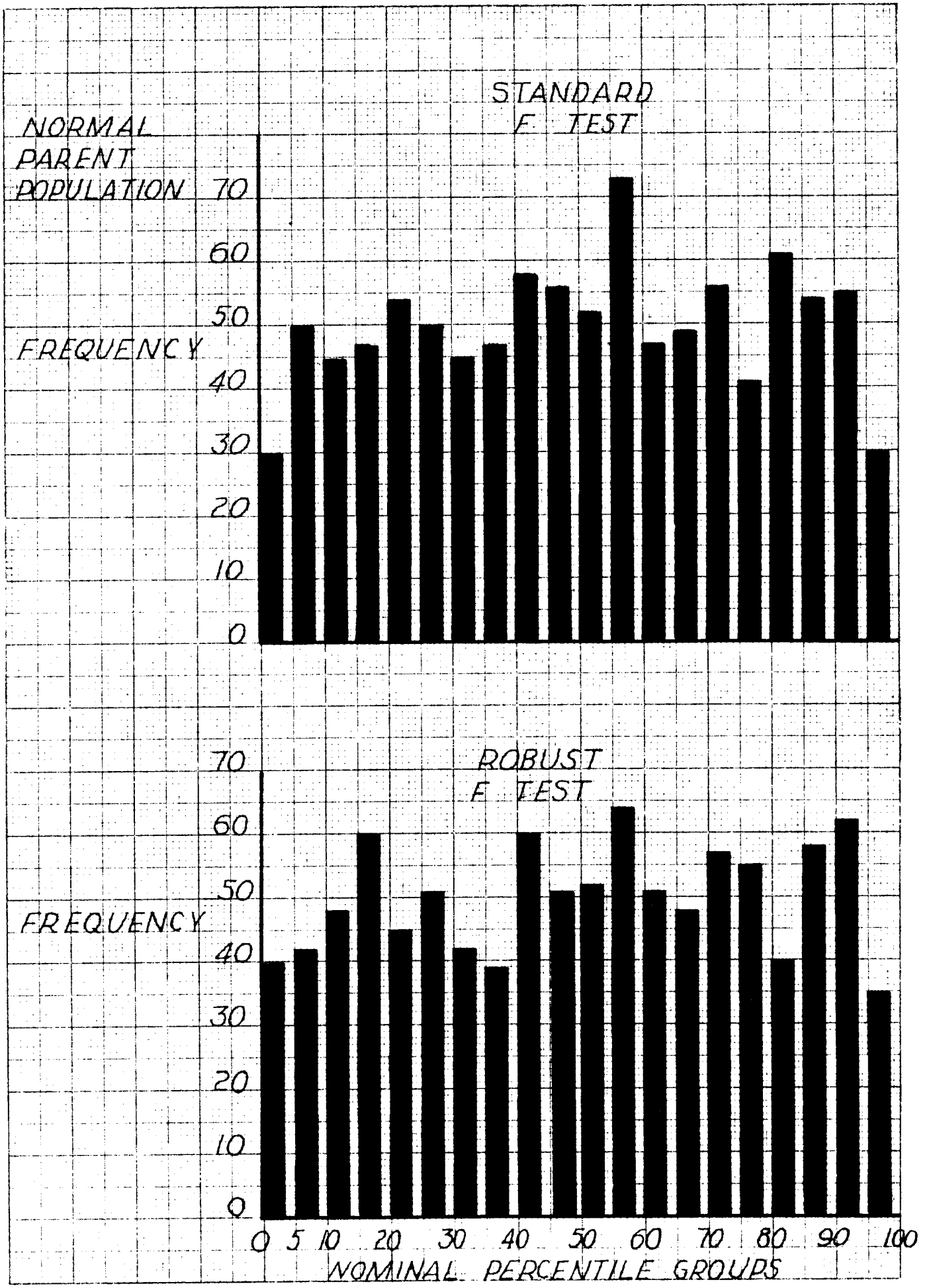

PROBABILITY
F
DEGREES OF FREEDOM
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.8
90
95
97
99
995
997
998
999

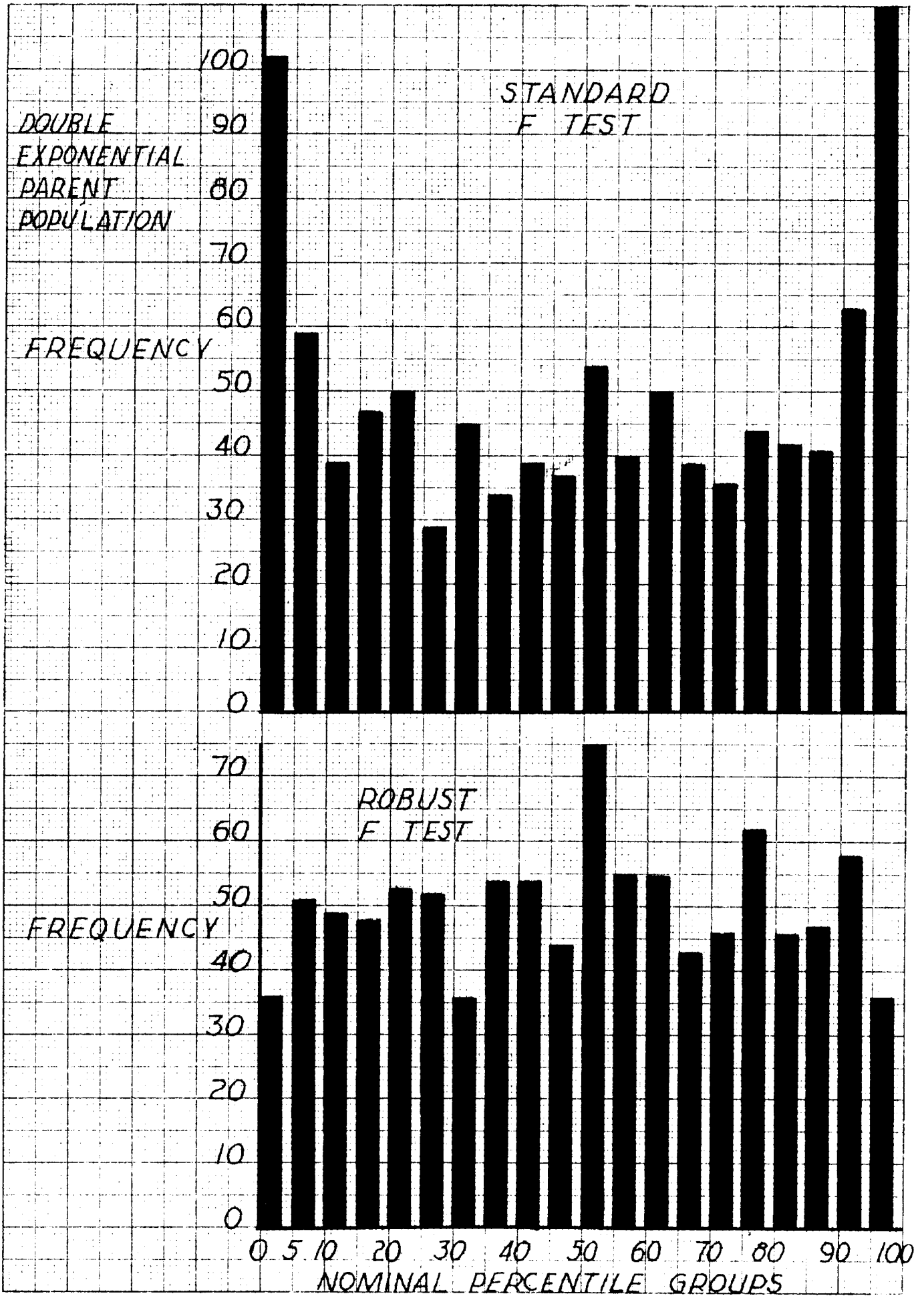

PROBABILITY
F
50
52
54
56
58
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
1.00
1.05
1.10
1.15
1.20
1.25
1.30
1.35
1.40
1.45
1.50
1.55
100
96
92
88
84
80
76
72
68
64
60
56
52
48
44
40
36
34
32
30
28
26
24
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6

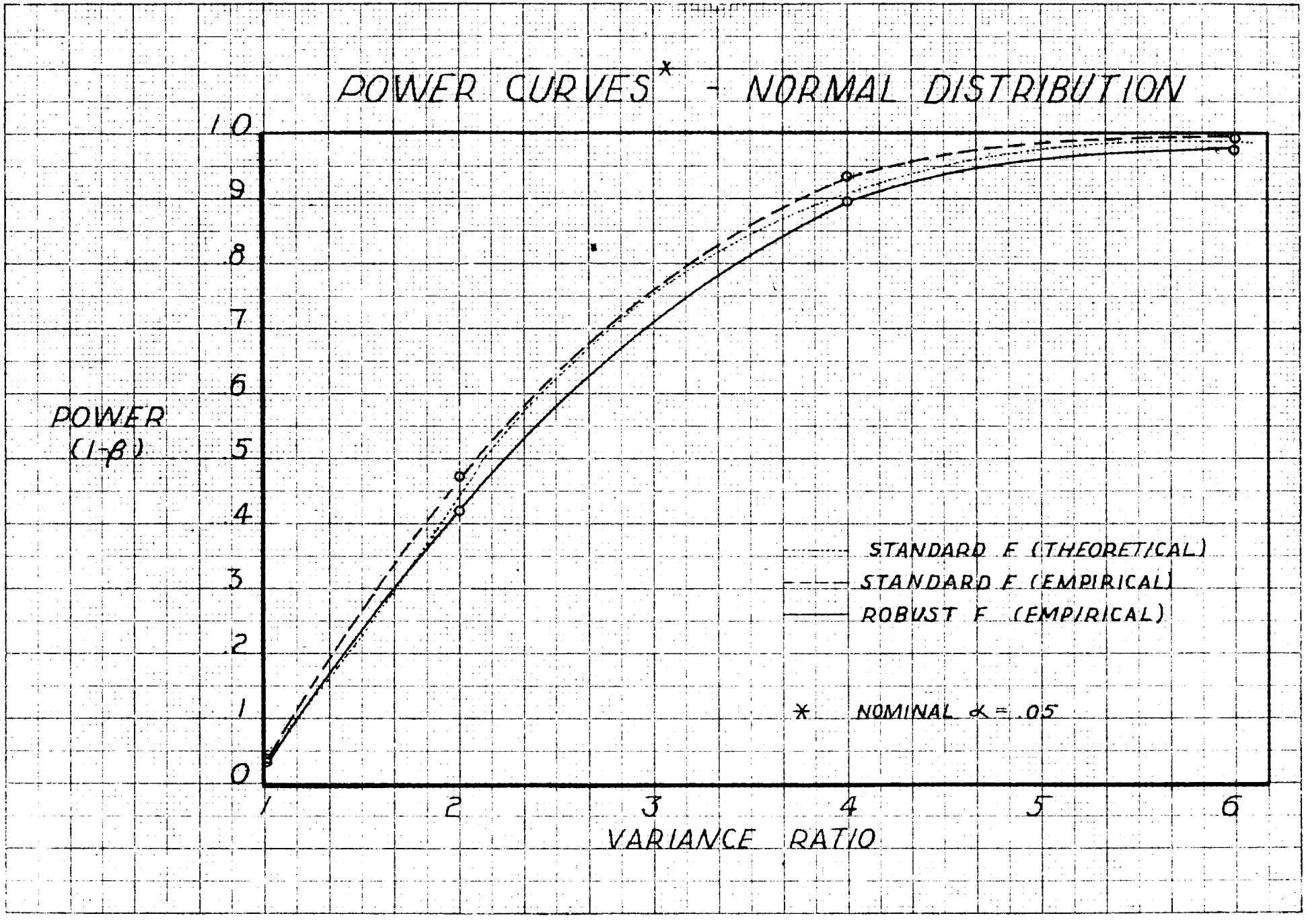

PROBABILITY
PROBABILITY
F RATIO

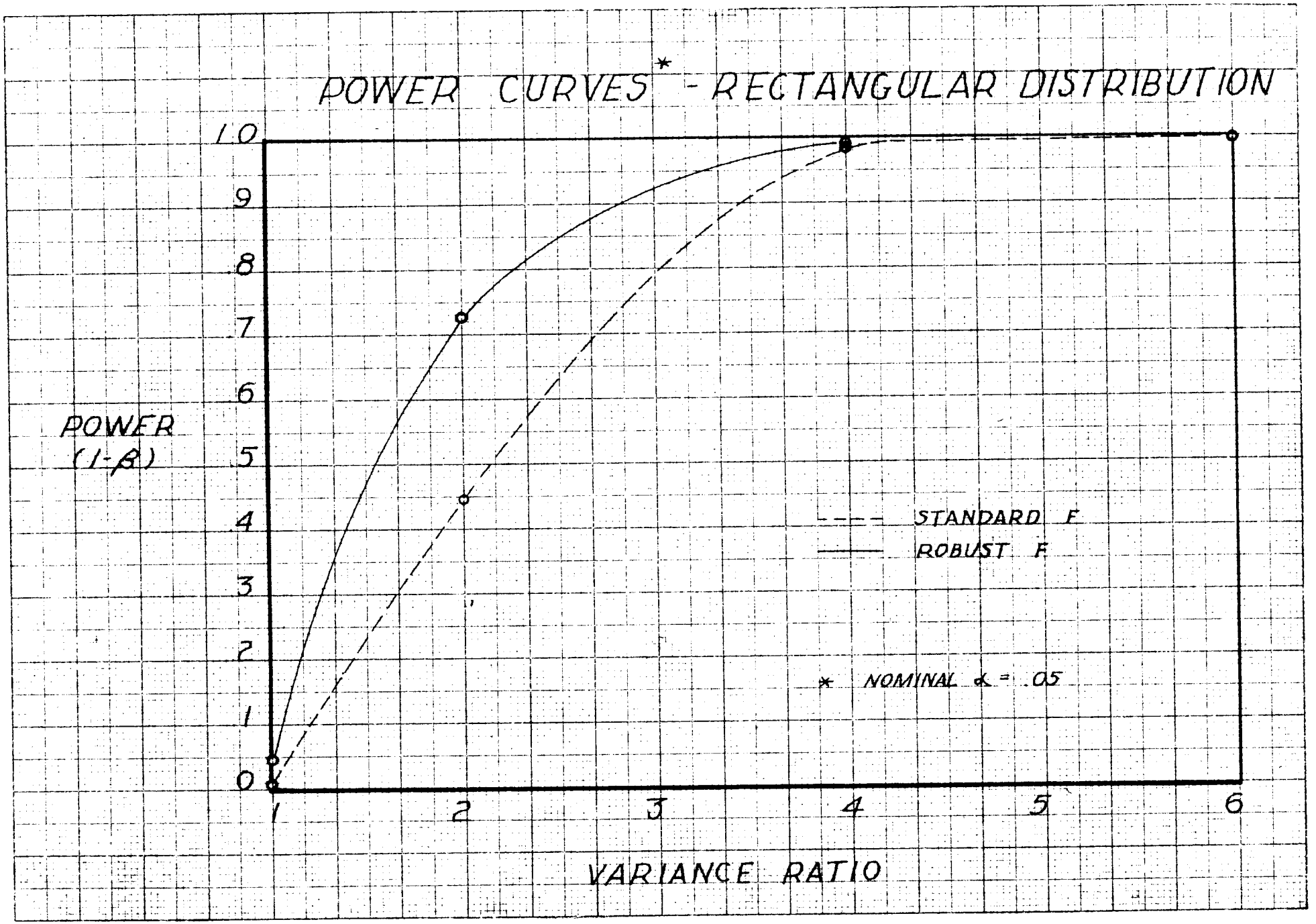

PROBABILITY
PROBABILITY
F RATIO


STANDARD
F TEST
RECTANGULAR
PARENT
POPULATION
FREQUENCY
90
80
70
60
50
40
30
20
10
0
ROBUST
F TEST
FREQUENCY
70
60
50
40
30
20
10
0
0 5 10 20 30 40 50 60 70 80 90 100
NOMINAL PERCENTILE GROUPS


NORMAL
PARENT
POPULATION
STANDARD
F TEST
FREQUENCY
70
60
50
40
30
20
10
0
ROBUST
F TEST
FREQUENCY
70
60
50
40
30
20
10
0
0 5 10 20 30 40 50 60 70 80 90 100
NOMINAL PERCENTILE GROUPS




POWER CURVES* - NORMAL DISTRIBUTION
POWER (1-β)
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
1
2
3
4
5
6
VARIANCE RATIO
STANDARD F (THEORETICAL)
STANDARD F (EMPIRICAL)
ROBUST F (EMPIRICAL)
* NOMINAL α = .05


POWER CURVES* - RECTANGULAR DISTRIBUTION
POWER (1-β)
10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
VARIANCE RATIO
STANDARD F
ROBUST F
* NOMINAL α = .05