DANIEL F. MERRIAM, Editor

FORTRAN IV PROGRAM FOR SAMPLE NORMALITY TESTS

By

D. A. PRESTON
Shell Development Company

in cooperation with the
American Association of Petroleum Geologists
Tulsa, Oklahoma

COMPUTER CONTRIBUTION 41
State Geological Survey
The University of Kansas, Lawrence
1970

Editor's Remarks

This computer program, "FORTRAN IV program for sample normality tests", by D.A. Preston starts our fifth year of the series. We are pleased that the publications have been so well received and are proving of use to practicing geologists the world over. About 80,000 copies of COMPUTER CONTRIBUTIONS now have been distributed to scientists in more than 40 countries!

This COMPUTER CONTRIBUTION should be of value to geologists interested in distributions of geological populations. It will have special meaning to petroleum geologists seeking to improve their predictions of finding new oil and gas fields. Examples are given from Kansas and Texas by the author in presenting results of his many years of research on the subject. I am pleased to note that the manuscript was completed by the author during his tenure as a visiting industrial scientist with the Geological Survey in 1969.

The program will be made available on magnetic tape for a limited time for $15.00 (US). An extra $10.00 charge is made if the punched cards are required.

I am pleased to welcome Dr. Frederic P. Agterberg of the Geological Survey of Canada (Ottawa) and Mr. John H. Hefner of the Humble Oil and Refining Company (Houston) as new associate editors and Mr. Paul J. Wolfe, Director of the Computation Center at The University of Kansas, as a technical editor. They will help maintain the high editorial standards of the past four years. I am most pleased to acknowledge the help of retiring board members, Dr. John R. Dempsey and Dr. R.G. Hetherington, who over the past several years gave unselfishly of their time in refereeing manuscripts. Their efforts are most appreciated.

We look forward to the new decade with all its promise and hope that geological accomplishments will equal or surpass those of the 1960's. The soaring 70's should witness a review of old techniques and ideas, a development and refinement of known methods, and a search for new untried ways and concepts.

Some recent Computer Contributions

---

* Active Member, ° Associate Member, + Junior Member, American Association of Petroleum Geologists

FORTRAN IV PROGRAM FOR SAMPLE NORMALITY TESTS

by

D.A. Preston

INTRODUCTION

The improvement of prediction is among the most important goals of any scientific investigation. Yet the results of an investigation will be subject to considerable uncertainty if the supporting data are drawn from a population whose characteristics are incompletely understood. This limitation is particularly common for geologic investigations where populations are, in general, only partially accessible so that their exact nature cannot be determined. Useful numerical approximations of their probability distributions may be derived, however, if the data are drawn in such a manner that they constitute a representative sample of the parent population (see Griffiths, 1967; Griffiths and Ondrick, 1968). Statistical analysis of such data then may improve the predictability of further sampling by suggesting an appropriate a priori probability model for the parent population from which predictive inferences can be drawn. Moreover, once an appropriate model is established for a population, the quality of any sample drawn from that population may be determined easily.

A distinction must be preserved for possible purposes for which samples are statistically analyzed. On one hand, if samples are from a population whose probabilistic nature is known, statistical analysis yields a measure of sample error created by departure from randomness, operator error, and the like. The main problem is identifying the sources of error. On the other hand, if samples are analyzed in order to derive a population distribution model, the sample must be constructed painstakingly to minimize sample error. Poorly constructed samples from populations of unknown distribution will yield completely ambiguous statistical results.

THE NORMAL A PRIORI PROBABILITY MODEL

Two generally accepted categories of a priori probability models exist. One characterizes the distribution of discrete variables based on counting or similar enumeration, and the other characterizes the distribution of continuous variables based on measurements. In the second category the normal or lognormal model seems to be most frequently applied to natural phenomena. Perhaps this is because of the mathematical basis of multivariate normality; namely, that the sum of an assemblage of random variables is distributed normally irrespective of the distribution of each contributing variable. If a natural phenomenon is assumed to be the result of several random events, by the same reasoning that phenomenon should be normally distributed. This is true, of course, only to the degree that the contributing events tend to be random with respect to each other. In the extremely complex physical systems with which the geologist deals a seemingly randomness may describe the net effect.

The normal distribution with its two parameters is not the simplest of the probability models. It is, nonetheless, relatively easy for the nonstatistician to understand because its two parameters describe (1) the average data value of the sample (mean), and (2) the variability or spread of the data around the average (standard deviation). Moreover, the separation of the two parameters makes the model particularly tractable mathematically.

In practice, to test for population normality or lognormality, the parent population is assumed to be normally or lognormally distributed. This is the so-called null hypothesis. The data (or their logarithms) then are analyzed and their statistical behavior tested against the null hypothesis of population normality. Lognormal distributions can be handled by standard normality tests because the logarithms of data elements drawn from a lognormal population are normally distributed. The hypothesis then is either accepted or rejected on the basis of closeness of fit to the normal model at some predetermined level (usually 95%). Certain risks are incurred by this procedure. An improper model for the population may be accepted, or a proper model may be rejected. The tests used in the program SNORT tend to minimize the former risk because acceptance of an improper model is usually the most undesirable of the two outcomes. Testing for sample quality from a population of known distribution is, of course, more straightforward because in this situation the answer sought is contingent only on the test results.

STATISTICAL DESCRIPTION

The principal statistical tests used in the program SNORT are the chi-square and the Kolmogorov-Smirnov parametric tests for normality. The nonparametric two sample Kolmogorov-Smirnov test is made if two samples are run. In addition, the program computes the mean, standard deviation, coefficients of skewness and kurtosis, and ratio of range of the data to the standard deviation.

Chi-square test

The chi-square test is a measure of the disparity existing between observed data values in a sample and those expected from whatever a priori probability model the sample is being tested against. The test value is calculated by

$$\chi^2 = \sum_{i=1}^{n} \frac{(\text{obs}_i - \text{exp}_i)^2}{\text{exp}_i}, \qquad (1)$$

where n = number of class intervals into which the data are divided, obs_i = frequency of observed data in class i, and exp_i = frequency of data in class i expected from the model. The chi-square values of each sample follow a distribution given by

$$Y = Y_o \, (\chi^2)^{\frac{1}{2}(\nu - 2)} e^{-\frac{1}{2}\chi^2}, \qquad (2)$$

where ν = degrees of freedom and Y_o is a constant dependent on ν so that the area under the function curve is kept at 1.0.

If the expected frequency in each class interval is at least equal to 5 then equation (2) is a close approximation to the sampling distribution of χ^2. Consequently most statisticians, when samples are small, will pool class intervals until the expected frequency in each pooled interval is at least 5. Other statisticians argue that the improved approximation is more than offset by loss of resolution in the tail portions of the curve. They recommend pooling only when an expected frequency otherwise would be close to or less than 1.0. Their argument is that excessive pooling is likely to reduce the chance of legitimately rejecting the a priori model. Rather than arbitrate this argument for the user, pooling for the chi-square statistic in the program SNORT is done, if necessary, at levels of expected values of both 5.0 and 1.5.

The class intervals are set before pooling at 0.3 standard deviations. The error of grouping is kept to less than 0.1 of the standard error, as recommended by Fisher (1946). This is slightly coarser than 0.25 standard deviation intervals, at which point the information loss due to grouping is less than 1.0 percent. Degrees of freedom in a chi-square test are based on the number of class intervals used after pooling and on whether the model parameter(s) are known or assumed. The program SNORT deducts two degrees of freedom for an assumed mean and standard deviation since they are derived from the sample itself.

Kolmogorov-Smirnov Test

This test is based on the sample distribution function, $F_n(x)$, where

$$F_n(x) = 1/n \quad (\text{number of observations} \leq x) \; . \qquad (3)$$

The function is expressed as the cumulative relative frequency of the sample. The choice of class interval is less critical than in the chi-square test, so 0.3 standard deviation intervals are used for convenience. The sample statistic used is

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)| \; , \qquad (4)$$

where D_n = the maximum absolute difference between the cumulative relative frequency curves of the observed and expected frequencies. The critical value for acceptance at the 95 percent significance level is calculated by

$$D_n = 1.36/\sqrt{n} \; , \qquad (5)$$

where n = sample size. The null hypothesis for the model is rejected if this critical value is exceeded. Valuable information about the sample distribution is given by observing where on the curve the critical value is exceeded. Accordingly, the program SNORT plots the derived values for both expected and observed frequencies in each interval to provide a direct visual comparison of the curves.

Kolmogorov-Smirnov Two Sample Test

This test compares the relative cumulative frequency curves of two samples. It determines whether or not the samples have been drawn from populations having the same frequency distribution irrespective of that distribution. The test statistic is

$$D_{mn} = \sup_{-\infty < x < \infty} | F_n(x) - F_m(x) | \; . \qquad (6)$$

The critical value for acceptance at the 95 percent significance level is calculated by

$$D_{mn} < \left| 1.36 \sqrt{(m + n) / mn} \right| , \qquad (7)$$

where $F_n(x)$ and $F_m(x)$ are the observed relative cumulative frequencies of samples n and m respectively, and D_{mn} is the maximum absolute difference between the two functions. The program SNORT makes this calculation, compares it to the maximum deviate between intervals, and prints out a statement of acceptance or rejection. This procedure is inaccurate for sample sizes less than 20, where tables should be used (Lindgren, 1962).

Skewness, Kurtosis, and Sample Range Statistics

Departure from normality in a sample is reflected by the asymmetry (skewness) and peakedness (kurtosis) of the frequency curve it describes relative to the normal curve. Measures of these effects are given by the third and fourth moments about the mean. Any moment about the mean is expressed as

$$M_r = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^r}{n} \; , \qquad (8)$$

where M_r = the r^{th} moment, x_i = i^{th} sample element, $\bar{x}$ = the sample mean, and n = sample size.

In dimensionless form the moment coefficient of skewness = $M_3 / \bar{S}^3 = \sqrt{\beta_1}$; the moment coefficient of kurtosis = $M_4 / \bar{S}^4 = \beta_2$, where $\bar{S} = \sqrt{M_2}$. For the standardized normal curve, $\sqrt{\beta_1} = 0.0$ and $\beta_2 = 3.0$.

When $\sqrt{\beta_1} > 0.0$ the distribution curve has a longer tail to the right of the central maximum indicating bias toward high data values. For $\sqrt{\beta_1} < 0.0$ the converse is true. When $\beta_2 > 3.0$ the curve tends toward strong centering about the mean with long tails (leptokurtosis) indicating less than expected variation of the data values. When $\beta_2 < 3.0$ the converse is true. Plates I and II, which graph the acceptable deviations of $\sqrt{\beta_1}$ and β_2 at the 95 percent significance level, are based on Pearson type curves that approximate the distribution of $\sqrt{\beta_1}$ and β_2 (Pearson and Hartley, 1966).

Another useful test for departure from normality is the ratio of the range of the data to its standard deviation. This test is especially sensitive for detecting maverick data values. A thorough discussion of this statistic is given in Pearson and Hartley's tables (1966). Plate III is a graph of the acceptance limit of this statistic at the 95 percent significance level.

POLYNOMIAL APPROXIMATION TO NORMAL CURVE

The use of this function in the program SNORT allows the expected value within any class interval of the normal to be generated from the specific sample being tested. Consequently, except for the chi-square test, no manual table look-up or subsequent hand calculation is required. The function is expressed

$$P(x) = 1. - Z(X)\,(0.4361836t - 0.1201676t^2 + 0.9372980t^3) + e(x) \quad , \qquad (9)$$

where $t = 1./(1. + 0.33267x)$, $Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$,

and $|e(x)|$ = error of approximation $< 10^{-5}$. The value of x is arbitrarily chosen as the class interval expressed in standard deviations from the mean ($0.3\,\overline{S}$ here). P(x) is the area under the normal curve from x to the mean. Expected values for sample size N are derived by

$$exp_i = (P(x_i) - P(x_{i-1}))\ N \ . \qquad (10)$$

PROGRAM DESCRIPTION

Input

The program SNORT (Sample NORmality Tests) is written in FORTRAN IV at a sufficiently generalized level to be compatible with or easily adapted to most computers. The program accepts one or two samples input as decks of punched cards with a limit of 5000 cards per deck, one datum to a card in fixed format. It will test the sample(s) for either normality or lognormality at the user's option. When two samples are run an additional test, the Kolmogorov-Smirnov two sample test, is run also. It is recommended, though not necessary, that sample size be kept in excess of 20. The input format may be changed to fit any special need of the user by changing format statement 1000. The sample size limit may be increased by changing the arguments of variables SMPL1 and SMPL2 in the first dimension statement.

Program Input Cards

CARD 1 An option card
Cols.
5 An integer 1 or 2 punched in this column specifies one or two samples.
10 An integer 1 or 2 punched in this column specifies a test for normality or lognormality respectively.

DATA CARDS 20 to 5000 allowed
Cols.
1-10 Sample one data elements either right justified or with decimal punched.

CARD 2
Cols.
75-80 Punch FINISH to signify end of data set.

DATA CARDS (optional) 20 to 5000 allowed
Cols.
1-10 Sample two data elements either right justified or with decimal punched.

CARD 3
Cols.
75-80 Punch FINISH to signify end of data set.

Output

An example of printer output is shown in Figures 1-8. The sample mean, standard deviation, skewness, kurtosis, ratio of range to standard deviation, and sample size are printed out first. Next, a statement of acceptance or rejection of the null hypothesis of normality at the 95 percent significance level is made based on the stated Kolmogorov-Smirnov critical value. The observed and expected cumulative relative frequencies at $0.3\,\overline{S}$ intervals are listed next with their absolute differences so that the user can determine the goodness of fit between each interval. The listed interval values are plotted below to afford a visual comparison of the curves they describe. The user can easily ascertain details and trends of deviation between the observed (empirical) and expected frequency distributions. The expected curve is the "1" curve, the empirical is the "2" curve. Where the curves are coincident an asterisk is printed. The range of data (or their logarithms) is calculated and scaled beneath the X axis. A listing of expected and observed frequencies in $0.3\,\overline{S}$ class intervals for the chi-square test is printed out below the plot.

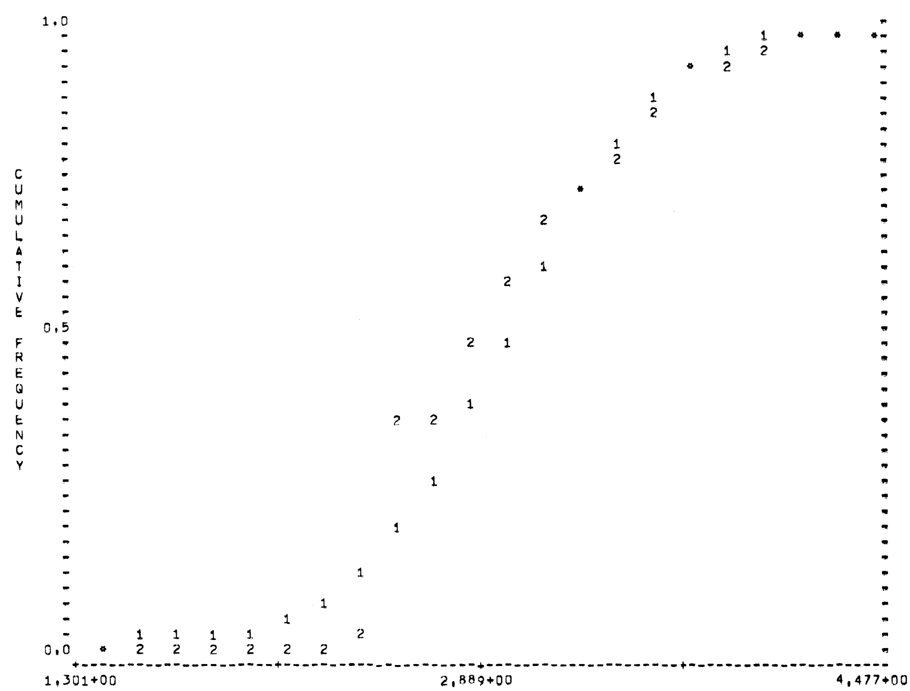
SAMPLE 1 HAS A MEAN OF 2,1431,A STANDARD DEVIATION OF ,5781--SKEWNESS IS ,9707,KURTOSIS 3,5520,

THE RATIO OF THE RANGE TO THE STD DEV = 5,49 THE SAMPLE SIZE = 264

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS REJECTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,08370 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	.00000
,00000	,00135	.00135
,00000	,00348	,00348
,00000	,00821	,00821
,00000	,01787	,01787
,00000	,03593	.03593
,00000	,06680	,06680
,00758	,11506	.10748
,37500	,18406	.19094
,37500	,27426	,10074
,48106	,38209	,09897
,57955	,50000	,07955
,69318	,61791	,07527
,73106	,72574	.00532
,79167	,81594	,02428
,86364	,88494	.02130
,92803	,93320	,00517
,94697	,96407	.01710
,96970	,98213	.01243
,97727	,99179	,01452
,98485	,99652	,01168
,99242	,99865	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE ONE
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 1-- ARE

EXPECTED	OBSERVED
,4	0
,6	0
1,2	0
2,6	0
4,8	0
8,1	0
12,7	2
18,2	97
23,8	0
28,5	28
31,1	26
31,1	30
28,5	10
23,8	16
18,2	19
12,7	17
8,1	5
4,8	6
2,6	2
1,2	2
,6	2

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 412,462 WITH 15 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 409,434 WITH 11 DEGREES OF FREEDOM

Figure 1.- Sample 1 statistics, field area, Kansas City-Lansing Groups, Central Kansas Uplift.

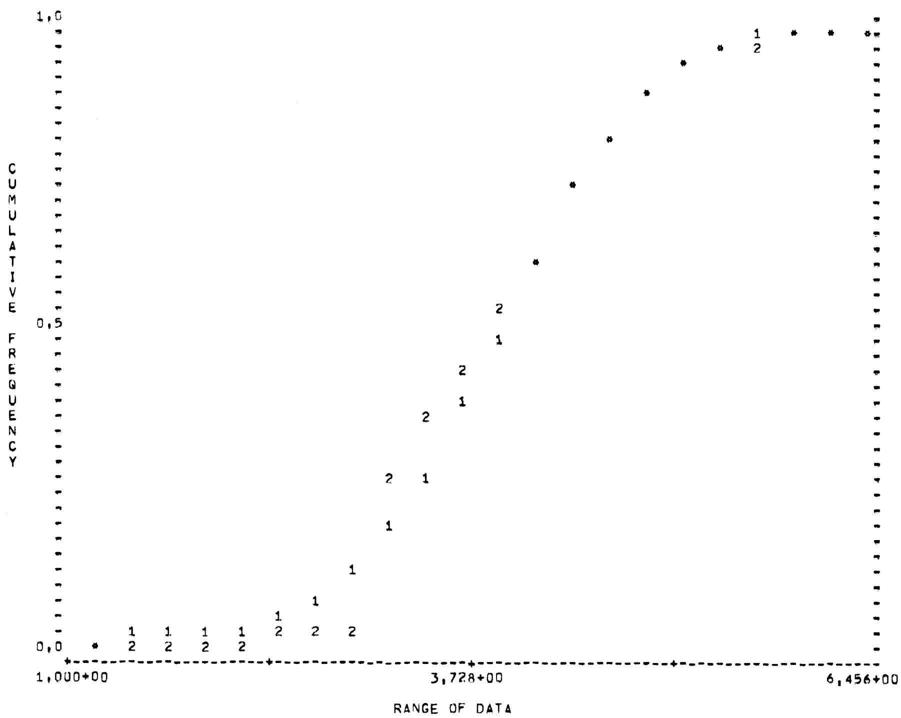
SAMPLE 2 HAS A MEAN OF 2,4472,A STANDARD DEVIATION OF ,7890--SKEWNESS IS ,9430,KURTOSIS 4,6987,

THE RATIO OF THE RANGE TO THE STD DEV = 6,92 THE SAMPLE SIZE = 250

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS REJECTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,08601 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	,00000
,00000	,00135	,00135
,00000	,00348	,00348
,00000	,00821	,00821
,00000	,01787	,01787
,00400	,03593	,03193
,00400	,06680	,06280
,01600	,11506	,09906
,26400	,18406	,07994
,37200	,27426	,09774
,44800	,38209	,06591
,54800	,50000	,04800
,62400	,61791	,00609
,72800	,72574	,00226
,82400	,81594	,00806
,87600	,88494	,00894
,92800	,93320	,00520
,96000	,96407	,00407
,96800	,98213	,01413
,98400	,99179	,00779
,99600	,99652	,00052
,99600	,99865	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE TWO
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 2-- ARE

EXPECTED	OBSERVED
,3	0
,5	0
1,2	0
2,4	0
4,5	1
7,7	0
12,1	3
17,2	62
22,6	27
27,0	19
29,5	25
29,5	19
27,0	26
22,6	24
17,2	13
12,1	13
7,7	8
4,5	2
2,4	4
1,2	3
,5	0

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 150,122 WITH 15 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 146,732 WITH 11 DEGREES OF FREEDOM

Figure 2.- Sample 2 statistics, field area, Arbuckle Group, Central Kansas Uplift.

SAMPLE 1 VS. SAMPLE 2--K-S TEST

CRITICAL VALUE OF KOLMOGOROV-SMIRNOFF TWO SAMPLE TEST FOR ACCEPTANCE AT .05 IS LESS THAN.12002 ABS DIFF

YOUR NULL HYPOTHESIS IS REJECTED

CUM FREQ	CUM FREQ	ABS.DIFF
.0000	.0000	.0036
.0076	.0040	.0036
.0076	.0040	.0036
.0076	.0040	.0036
.0076	.0040	.0036
.0076	.0160	.0084
.0076	.0160	.0084
.0076	.0160	.0084
.0076	.0160	.0084
.3750	.0160	.3590
.3750	.0160	.3590
.3750	.2640	.1110
.3750	.2640	.1110
.3750	.2640	.1110
.3750	.2720	.1030
.3750	.2720	.1030
.3750	.3720	.0030
.3750	.3720	.0030
.4811	.3720	.1091
.4811	.4120	.0691
.4811	.4120	.0691
.4811	.4160	.0651
.5000	.4480	.0520
.5000	.5120	.0120
.5795	.5160	.0635
.5795	.5360	.0435
.5795	.5480	.0315
.5833	.5680	.0153
.6326	.5840	.0486
.6326	.6080	.0246
.6364	.6240	.0124
.6932	.6440	.0492
.6970	.6560	.0410
.7083	.6880	.0203
.7083	.7280	.0197
.7083	.7600	.0517
.7197	.7760	.0563
.7462	.7880	.0418
.7500	.8120	.0620
.7576	.8320	.0744
.7727	.8400	.0673
.7803	.8600	.0797
.7917	.8640	.0723
.7992	.8800	.0808
.8144	.9000	.0856
.8182	.9120	.0938
.8333	.9200	.0867
.8523	.9280	.0757
.8636	.9320	.0684
.8712	.9440	.0728
.8902	.9480	.0578
.8977	.9560	.0583
.9053	.9600	.0547
.9280	.9640	.0360
.9318	.9640	.0322
.9318	.9680	.0362
.9356	.9680	.0324
.9432	.9760	.0328
.9470	.9760	.0290
.9470	.9800	.0330
.9583	.9840	.0257
.9659	.9840	.0181
.9659	.9840	.0181
.9659	.9880	.0221
.9735	.9920	.0185
.9735	.9960	.0225
.9773	.9960	.0187
.9773	.9960	.0187
.9773	.9960	.0187
.9773	.9960	.0187
.9811	.9960	.0149
.9811	.9960	.0149
.9811	.9960	.0149
.9811	.9960	.0149
.9848	.9960	.0112
.9848	.9960	.0112
.9848	.9960	.0112
.9848	.9960	.0112
.9848	.9960	.0112
.9886	.9960	.0074
.9924	.9960	.0036
.9924	.9960	.0036
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002
.9962	.9960	.0002

Kolmogorov-Smirnov two sample test, from samples shown in Figures 1 and 2.

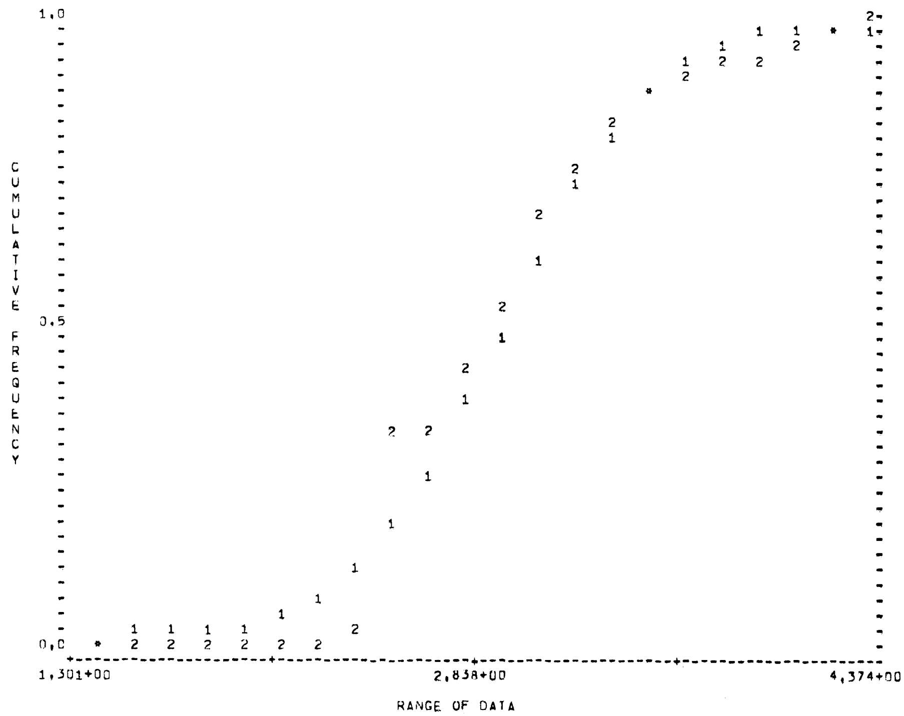
SAMPLE 1 HAS A MEAN OF 2,2911,A STANDARD DEVIATION OF ,7373--SKEWNESS IS 1,0385,KURTOSIS 3,4691,

THE RATIO OF THE RANGE TO THE STD DEV = 4,17 THE SAMPLE SIZE = 86

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS REJECTED AT THE .05 LEVEL--K-S CRITICAL VALUE IS ,14665 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	,00000
,00000	,00135	,00135
,00000	,00348	,00348
,00000	,00821	,00821
,00000	,01787	,01787
,00000	,03593	,03593
,00000	,06680	,06680
,01163	,11506	,10343
,34884	,18406	,16478
,34884	,27426	,07458
,44186	,38209	,05977
,53488	,50000	,03489
,69767	,61791	,07976
,76744	,72574	,04170
,82558	,81594	,00964
,89535	,88494	,01041
,91860	,93320	,01459
,93023	,96407	,03383
,93023	,98213	,05189
,95349	,99179	,03830
,98837	,99652	,00815
1,00000	,99865	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE ONE
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 1-- ARE

EXPECTED	OBSERVED
,1	0
,2	0
,4	0
,8	0
1,6	0
2,7	0
4,2	1
5,9	29
7,8	0
9,3	8
10,1	8
10,1	14
9,3	6
7,8	5
5,9	6
4,2	2
2,7	1
1,6	0
,8	2
,4	3
,2	1

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 113,488 WITH 12 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 53,641 WITH 7 DEGREES OF FREEDOM

Figure 3.- Sample 1 statistics, field area, Clearfork Formation, Midland Central Basin Platform.

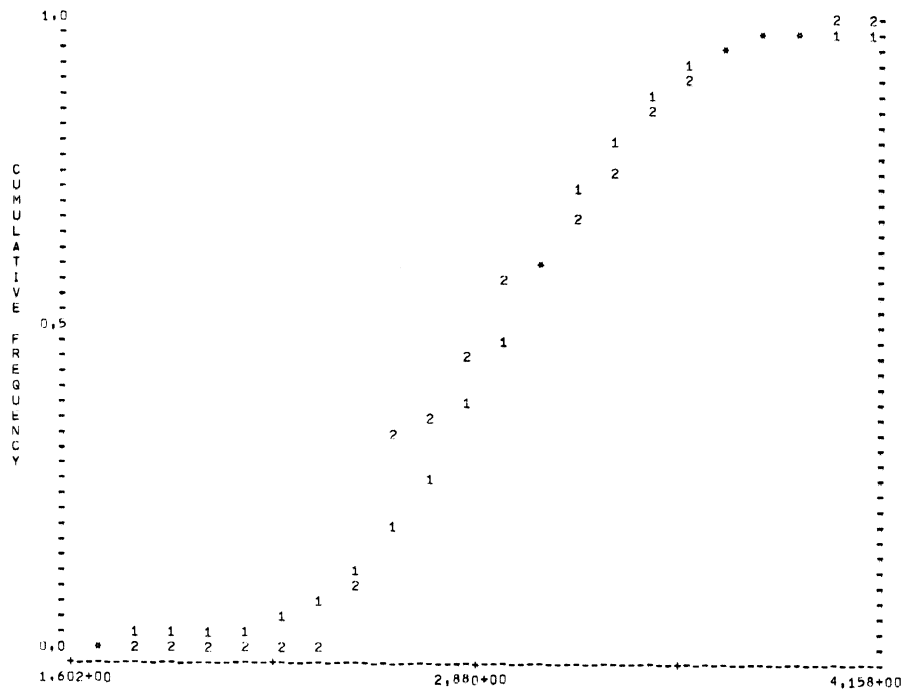
SAMPLE 2 HAS A MEAN OF 2,5205,A STANDARD DEVIATION OF ,6521--SKEWNESS IS ,4020,KURTOSIS 2,0070,

THE RATIO OF THE RANGE TO THE STD DEV = 3,92 THE SAMPLE SIZE = 105

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS REJECTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,13272 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	.00000
,00000	,00135	.00135
,00000	,00348	,00348
,00000	.00821	.00821
,00000	,01787	,01787
,00000	,03593	.03593
,00000	,06680	,06680
,08571	,11506	,02934
,33333	,18406	.14928
,35238	,27426	.07812
,46667	,38209	,08458
,58095	.50000	,08096
,60952	,61791	,00839
,69524	,72574	,03050
,75238	,81594	,06356
,85714	,88494	,02780
,90476	,93520	.02844
,97143	,96407	,00736
,99048	,98213	.00835
,99048	,99179	,00132
1,00000	,99652	,00348
1,00000	,99665	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE TWO
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 2-- ARE

EXPECTED	OBSERVED
,1	0
,2	0
,5	0
1,0	0
1,9	0
3,2	0
5,1	9
7,2	26
9,5	2
11,3	12
12,4	12
12,4	3
11,3	9
9,5	6
7.2	11
5,1	5
3,2	7
1,9	2
1,0	0
,5	1
.2	0

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 79,909 WITH 13 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 78,497 WITH 9 DEGREES OF FREEDOM

Figure 4.- Sample 2 statistics, field area, Devonian rocks, Midland Central Basin Platform.

SAMPLE 1 VS, SAMPLE 2--K-S TEST

CRITICAL VALUE OF KOLMOGOROV-SMIRNOFF TWO SAMPLE TEST FOR ACCEPTANCE AT ,05 IS LESS THAN,19779 ABS DIFF

YOUR NULL HYPOTHESIS IS REJECTED

CUM FREQ	CUM FREQ	ABS,DIFF
.0000	.0000	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.0116	.0857	.0741
.3488	.0857	.2631
.3488	.0857	.2631
.3488	.3333	.0155
.3488	.3333	.0155
.3488	.3333	.0155
.3488	.3333	.0155
.3488	.3333	.0155
.3488	.3333	.0155
.3488	.3333	.0155
.3488	.3524	.0035
.4419	.3524	.0895
.4419	.3524	.0895
.4419	.3524	.0895
.4419	.3524	.0895
.4419	.4571	.0153
.4419	.4571	.0153
.5116	.4571	.0545
.5116	.4571	.0545
.5116	.4667	.0450
.5116	.4667	.0450
.5349	.4667	.0682
.5349	.4952	.0396
.5349	.4952	.0396
.5930	.4952	.0978
.5930	.5143	.0787
.5930	.5143	.0787
.6279	.5810	.0470
.6279	.5810	.0470
.6512	.5905	.0607
.6512	.6000	.0512
.6977	.6000	.0977
.6977	.6000	.0977
.6977	.6000	.0977
.6977	.6095	.0882
.7209	.6095	.1114
.7674	.6476	.1198
.7674	.6476	.1198
.7674	.6571	.1103
.7791	.6667	.1124
.7907	.6762	.1145
.8023	.6762	.1261
.8023	.6952	.1071
.8023	.6952	.1071
.8140	.7048	.1092
.8256	.7143	.1113
.8372	.7143	.1229
.8372	.7333	.1039
.8605	.7333	.1271
.8837	.7429	.1409
.8837	.7524	.1313
.8837	.7619	.1218
.8953	.7810	.1144
.8953	.8000	.0953
.8953	.8000	.0953
.8953	.8000	.0953
.8953	.8286	.0668
.9070	.8571	.0498
.9070	.8571	.0498
.9186	.8667	.0519
.9186	.8667	.0519
.9186	.8762	.0424
.9302	.8857	.0445
.9302	.8857	.0445
.9302	.8857	.0445
.9302	.8857	.0445
.9302	.9048	.0255
.9302	.9238	.0064
.9302	.9333	.0031
.9302	.9429	.0126
.9302	.9429	.0126
.9302	.9524	.0221
.9302	.9714	.0412
.9302	.9810	.0507
.9302	.9810	.0507
.9302	.9810	.0507
.9419	.9810	.0391
.9419	.9905	.0486
.9535	.9905	.0370
.9535	.9905	.0370
.9535	.9905	.0370
.9535	.9905	.0370
.9535	.9905	.0370
.9535	.9905	.0370
.9535	.9905	.0370
.9651	.9905	.0254
.9767	.9905	.0137
.9884	.9905	.0021
.9884	.9905	.0021
.9884	.9905	.0021
.9884	.9905	.0021
.9884	.9905	.0021

Kolmogorov-Smirnov two sample test, from samples shown in Figures 3 and 4.

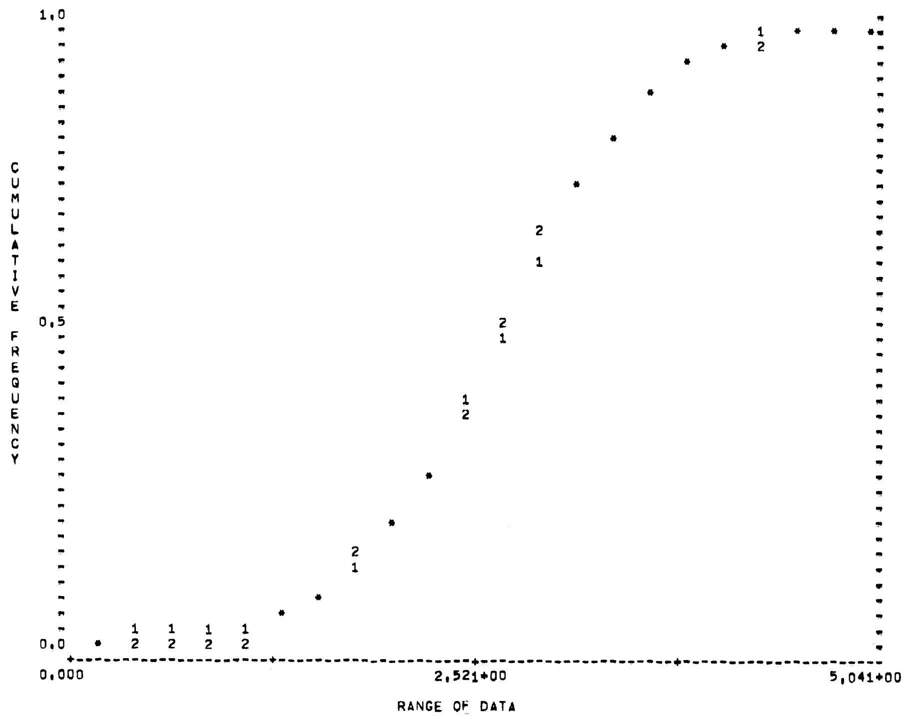
SAMPLE 1 HAS A MEAN OF 1,9704,A STANDARD DEVIATION OF ,9729--SKEWNESS IS ,1615,KURTOSIS 2,9422,

THE RATIO OF THE RANGE TO THE STD DEV = 5,18 THE SAMPLE SIZE = 264

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS ACCEPTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,08370 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	,00000
,00000	,00135	,00135
,00000	,00348	,00348
,00000	,00821	,00821
,00000	,01787	,01787
,03788	,03593	,00194
,07197	,06680	,00517
,12879	,11506	,01373
,19697	,18406	,01291
,27273	,27426	,00153
,36742	,38209	,01466
,50758	,50000	,00758
,65530	,61791	,03739
,74242	,72574	,01669
,80682	,81594	,00913
,89015	,88494	,00521
,92803	,93320	,00517
,95833	,96407	,00573
,97348	,98213	,00864
,98485	,99179	,00694
,99621	,99652	,00031
,99621	,99865	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE ONE
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 1-- ARE

EXPECTED	OBSERVED
,4	0
,6	0
1,2	0
2,6	0
4,8	10
8,1	9
12,7	15
18,2	18
23,8	20
28,5	25
31,1	37
31,1	39
28,5	23
23,8	17
18,2	22
12,7	10
8,1	8
4,8	4
2,6	3
1,2	3
,6	0

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 20,447 WITH 15 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 9,043 WITH 11 DEGREES OF FREEDOM

Figure 5. - Sample 1 statistics, ultimate volume, Kansas City-Lansing Groups, Central Kansas Uplift.

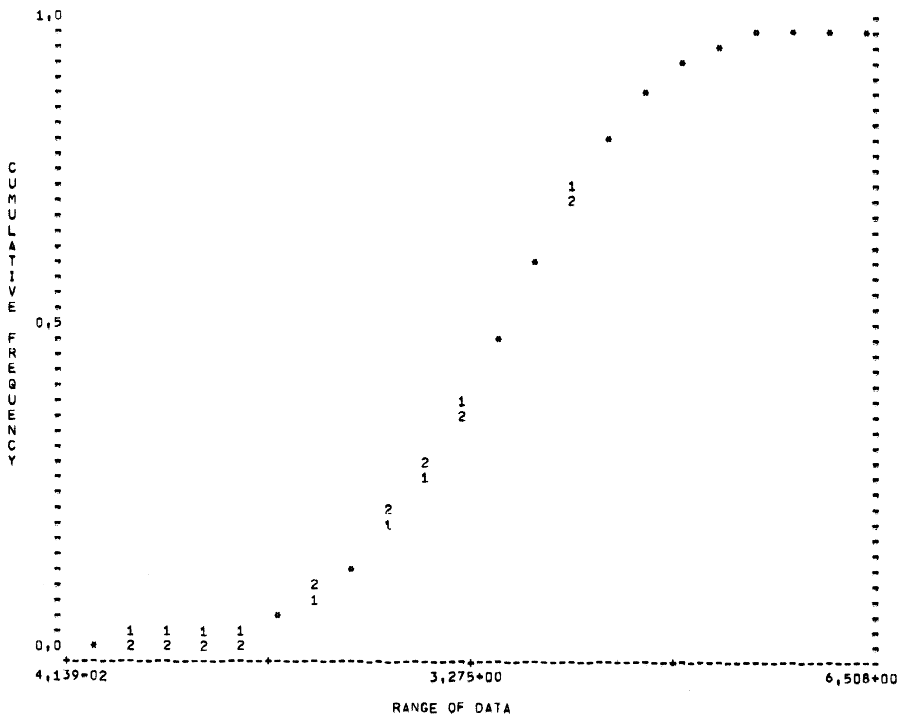
SAMPLE 2 HAS A MEAN OF 2,3855,A STANDARD DEVIATION OF 1,2049--SKEWNESS IS ,1556,KURTOSIS 2,9203,

THE RATIO OF THE RANGE TO THE STD DEV = 5,37 THE SAMPLE SIZE = 250

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS ACCEPTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,08601 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	,00000
,00000	,00135	,00135
,00000	,00348	,00348
,00000	,00821	,00821
,00000	,01787	,01787
,03600	,03593	,00007
,07600	,06680	,00920
,12000	,11506	,00494
,21200	,18406	,02794
,28000	,27426	,00574
,37200	,38209	,01009
,50000	,50000	,00000
,62000	,61791	,00209
,72400	,72574	,00174
,80800	,81594	,00794
,90000	,88494	,01506
,93600	,93320	,00280
,96000	,96407	,00407
,97600	,98213	,00613
,98400	,99179	,00779
,99600	,99652	,00052
,99600	,99865	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE TWO
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 2-- ARE

EXPECTED	OBSERVED
,3	0
,5	0
1,2	0
2,4	0
4,5	9
7,7	10
12,1	11
17,2	23
22,6	17
27,0	23
29,5	32
29,5	30
27,0	26
22,6	21
17,2	23
12,1	9
7,7	6
4,5	4
2,4	2
1,2	3
,5	0

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 18,098 WITH 15 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 8.202 WITH 11 DEGREES OF FREEDOM

Figure 6.- Sample 2 statistics, ultimate volume, Arbuckle Group, Central Kansas Uplift.

SAMPLE 1 VS, SAMPLE 2--K-S TEST

CRITICAL VALUE OF KOLMOGOROV-SMIRNOFF TWO SAMPLE TEST FOR ACCEPTANCE AT ,05 IS LESS THAN,12002 ABS DIFF

YOUR NULL HYPOTHESIS IS ACCEPTED

CUM FREQ	CUM FREQ	ABS,DIFF
,0000	,0000	,0019
,0379	,0360	,0019
,0379	,0360	,0019
,0379	,0360	,0019
,0379	,0680	,0301
,0568	,0680	,0112
,0568	,0760	,0192
,0568	,0760	,0192
,0568	,0840	,0272
,0720	,0840	,0120
,0720	,1000	,0280
,0985	,1080	,0095
,0985	,1120	,0135
,1098	,1200	,0102
,1098	,1360	,0262
,1288	,1400	,0112
,1439	,1640	,0201
,1515	,1760	,0245
,1667	,2000	,0333
,1742	,2120	,0378
,1856	,2160	,0304
,1970	,2200	,0230
,2121	,2320	,0199
,2197	,2640	,0443
,2311	,2800	,0489
,2386	,3000	,0614
,2576	,3200	,0624
,2803	,3240	,0437
,2955	,3360	,0405
,3106	,3680	,0574
,3333	,3800	,0467
,3485	,4000	,0515
,3636	,4240	,0604
,3788	,4520	,0732
,3826	,4840	,1014
,4129	,5000	,0871
,4356	,5240	,0884
,4735	,5440	,0705
,5076	,5560	,0484
,5341	,5720	,0379
,5492	,5920	,0428
,5720	,6200	,0480
,6023	,6320	,0297
,6288	,6640	,0352
,6591	,6760	,0169
,6667	,7040	,0373
,6780	,7200	,0420
,6894	,7280	,0386
,7045	,7400	,0355
,7348	,7600	,0252
,7424	,7800	,0376
,7500	,7960	,0460
,7652	,8080	,0428
,7765	,8360	,0595
,7803	,8560	,0757
,7917	,8680	,0763
,8144	,8840	,0696
,8485	,8960	,0475
,8561	,9000	,0439
,8712	,9040	,0328
,8826	,9160	,0334
,8902	,9200	,0298
,8977	,9320	,0343
,9091	,9360	,0269
,9129	,9360	,0231
,9205	,9520	,0315
,9242	,9520	,0278
,9280	,9560	,0280
,9432	,9560	,0128
,9470	,9600	,0130
,9470	,9680	,0210
,9545	,9720	,0175
,9583	,9760	,0177
,9583	,9760	,0177
,9621	,9760	,0139
,9659	,9800	,0141
,9697	,9800	,0103
,9735	,9840	,0105
,9735	,9840	,0105
,9773	,9840	,0067
,9811	,9840	,0029
,9811	,9920	,0109
,9811	,9960	,0149
,9848	,9960	,0112
,9848	,9960	,0112
,9848	,9960	,0112
,9848	,9960	,0112
,9886	,9960	,0074
,9924	,9960	,0036
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002
,9962	,9960	,0002

Kolmogorov-Smirnov two sample test, from samples shown in Figures 5 and 6.

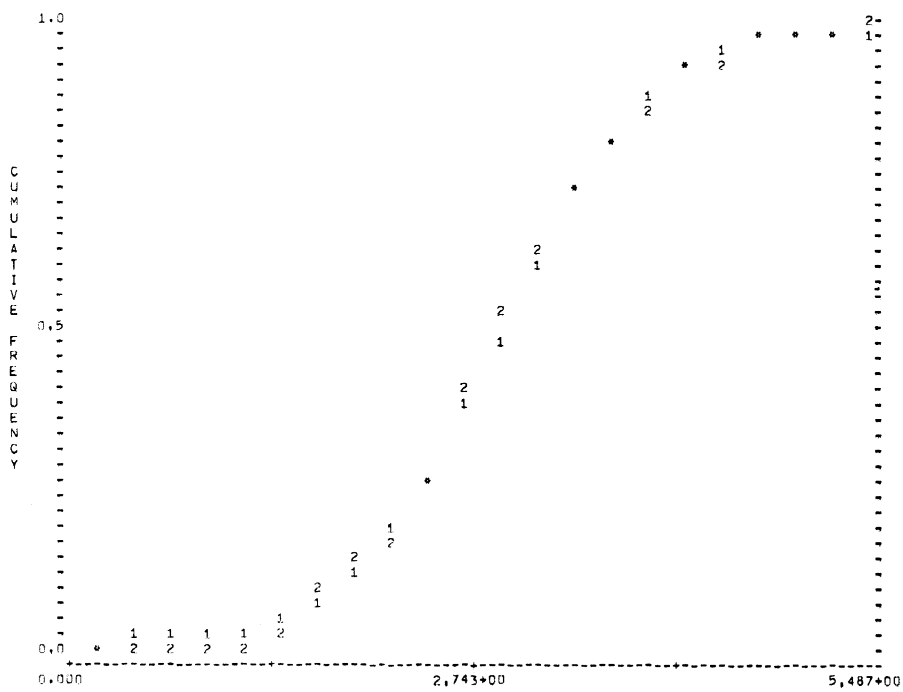
SAMPLE 1 HAS A MEAN OF 2,2451,A STANDARD DEVIATION OF 1,1557--SKEWNESS IS ,2641,KURTOSIS 2,7947,

THE RATIO OF THE RANGE TO THE STD DEV = 4,75 THE SAMPLE SIZE = 86

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS ACCEPTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,14665 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	,00000
,00000	,00135	,00135
,00000	,00348	,00348
,00000	,00821	,00821
,00000	,01787	,01787
,02326	,03593	,01268
,09302	,06680	,02622
,13953	,11506	,02448
,17442	,18406	,00964
,26744	,27426	,00682
,41860	,38209	,03652
,53488	,50000	,03489
,63953	,61791	,02162
,73256	,72574	,00682
,80233	,81594	,01362
,87209	,88494	,01285
,93023	,93320	,00297
,94186	,96407	,02221
,97674	,98213	,00538
,98837	,99179	,00342
,98837	,99652	,00815
1,00000	,99665	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE ONE
CUMULATIVE NORMAL*1* EMPIRICAL*2*



THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 1-- ARE

EXPECTED	OBSERVED
,1	0
,2	0
,4	0
,8	0
1,6	2
2,7	6
4,2	4
5,9	3
7,8	8
9,3	13
10,1	10
10,1	9
9,3	8
7,8	6
5,9	6
4,2	5
2,7	1
1,6	3
,8	1
,4	0
,2	1

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 12,100 WITH 12 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 4,369 WITH 7 DEGREES OF FREEDOM

Figure 7.- Sample 1 statistics, ultimate volume, Clearfork Formation, Midland Central Basin Uplift.

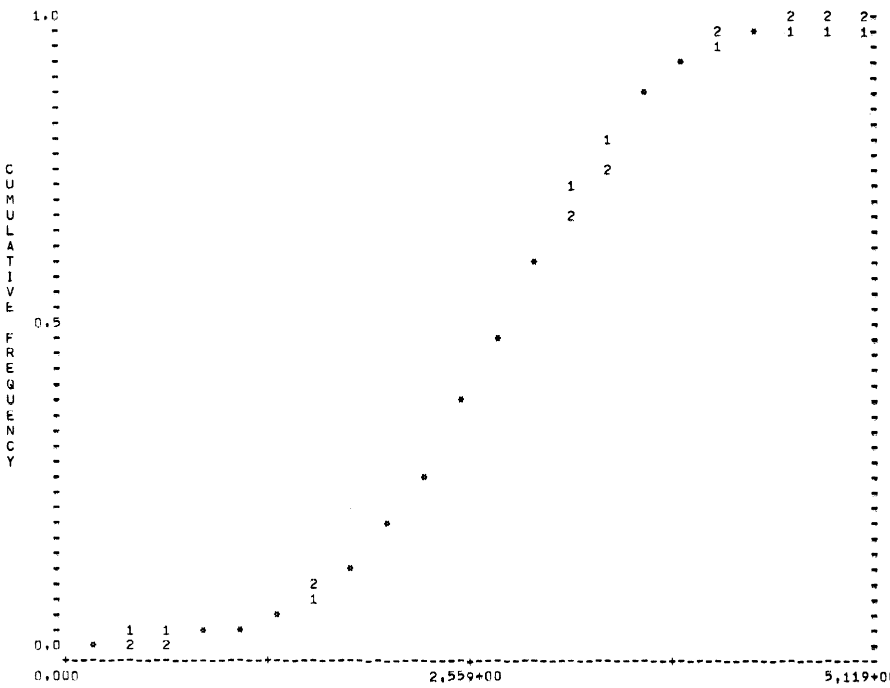
SAMPLE 2 HAS A MEAN OF 2,7554,A STANDARD DEVIATION OF 1,1203--SKEWNESS IS -,2218,KURTOSIS 2,4627,

THE RATIO OF THE RANGE TO THE STD DEV = 4,57 THE SAMPLE SIZE = 105

YOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS ACCEPTED AT THE ,05 LEVEL--K-S CRITICAL VALUE IS ,13272 ABS DIFF

CUM FREQ	EXP FREQ	ABS DIFF
,00000	,00000	,00000
,00000	,00135	,00135
,00000	,00348	,00348
,00952	,00821	,00132
,01905	,01787	,00117
,04762	,03593	,01168
.09524	,06680	,02844
,12381	,11506	,00875
,20000	,18406	,01594
,25714	,27426	,01712
,39048	,38209	,00839
,49524	,50000	,00476
,61905	,61791	,00114
,69524	,72574	,03050
,75238	,81594	,06356
,89524	,88494	,01030
,94286	,93320	,00966
,98095	,96407	,01689
,99048	,98213	,00835
1,00000	,99179	,00821
1,00000	,99652	,00348
1,00000	,99865	,00000

K-S TEST FOR (LOG)NORMALITY SAMPLE TWO
CUMULATIVE NORMAL*1* EMPIRICAL*2*



RANGE OF DATA

THE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQUARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0,3 ST DEV--SAMPLE 2-- ARE

EXPECTED	OBSERVED
,1	0
,2	0
,5	1
1,0	1
1,9	3
3,2	5
5,1	3
7,2	8
9,5	6
11,3	14
12,4	11
12,4	13
11,3	8
9,5	6
7,2	15
5,1	5
3,2	4
1,9	1
1,0	1
,5	0
,2	0

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 1,5 = 15,942 WITH 13 DEGREES OF FREEDOM

CHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF 5,0 = 14,621 WITH 9 DEGREES OF FREEDOM

Figure 8.- Sample 2 statistics, ultimate volume, Devonian rocks, Midland Central Basin Platform.

SAMPLE 1 VS. SAMPLE 2--K-S TEST

CRITICAL VALUE OF KOLMOGOROV-SMIRNOFF TWO SAMPLE TEST FOR ACCEPTANCE AT .05 IS LESS THAN .19779 ABS DIFF

YOUR NULL HYPOTHESIS IS REJECTED

CUM FREQ	CUM FREQ	ABS.DIFF
.0000	.0000	.0137
.0233	.0095	.0137
.0233	.0095	.0137
.0233	.0095	.0137
.0233	.0095	.0137
.0465	.0190	.0275
.0465	.0190	.0275
.0465	.0190	.0275
.0930	.0190	.0740
.0930	.0286	.0645
.1047	.0286	.0761
.1047	.0381	.0666
.1279	.0381	.0898
.1279	.0476	.0803
.1279	.0476	.0803
.1395	.0571	.0824
.1512	.0571	.0940
.1512	.0762	.0750
.1512	.0762	.0750
.1512	.0762	.0750
.1628	.0952	.0676
.1744	.1048	.0697
.1744	.1048	.0697
.2093	.1048	.1045
.2209	.1143	.1066
.2326	.1143	.1183
.2558	.1238	.1320
.2558	.1333	.1225
.2791	.1333	.1457
.2907	.1429	.1478
.3140	.1524	.1616
.3256	.1714	.1542
.3605	.1810	.1795
.3837	.2000	.1837
.4302	.2095	.2207
.4302	.2095	.2207
.4651	.2190	.2461
.4651	.2190	.2461
.5000	.2476	.2524
.5233	.2476	.2756
.5349	.2762	.2587
.5465	.2952	.2513
.5465	.2952	.2513
.5698	.3048	.2650
.6279	.3143	.3136
.6279	.3619	.2660
.6395	.3905	.2491
.6512	.4000	.2512
.6628	.4190	.2437
.6860	.4381	.2480
.7093	.4381	.2712
.7209	.4762	.2447
.7209	.4952	.2257
.7442	.4952	.2489
.7558	.5048	.2511
.7558	.5143	.2415
.7558	.5429	.2130
.7558	.5429	.2130
.7558	.5619	.1939
.8140	.6095	.2044
.8256	.6381	.1875
.8372	.6381	.1991
.8372	.6476	.1896
.8605	.6667	.1938
.8605	.6762	.1843
.8721	.6762	.1959
.8837	.6952	.1885
.9070	.7048	.2022
.9186	.7048	.2138
.9302	.7238	.2064
.9302	.7429	.1874
.9302	.7524	.1779
.9419	.7524	.1895
.9419	.7619	.1800
.9419	.8095	.1323
.9419	.8190	.1228
.9419	.8381	.1038
.9419	.8571	.0847
.9419	.8952	.0466
.9651	.8952	.0699
.9651	.9048	.0604
.9767	.9238	.0529
.9767	.9333	.0434
.9767	.9333	.0434
.9767	.9333	.0434
.9767	.9429	.0339
.9767	.9429	.0339
.9767	.9429	.0339
.9767	.9429	.0339
.9884	.9619	.0265
.9884	.9619	.0265
.9884	.9810	.0074
.9884	.9810	.0074
.9884	.9810	.0074
.9884	.9810	.0074
.9884	.9905	.0021
.9884	.9905	.0021
.9884	.9905	.0021
.9884	.9905	.0021
.9884	.9905	.0021

Kolmogorov-Smirnov two sample test, from samples shown in Figures 7 and 8.

Finally, the chi-square test values for the sample at the two pooling levels, and the associated degrees of freedom are printed. Acceptance or rejection of the null hypothesis must be determined from chi-square tables.

When two samples are run, a statement of acceptance or rejection of the two sample null hypotheses, with the critical value on which the decision is based, is printed out. The cumulative relative frequencies of both samples in class intervals of range/100, and the absolute differences between the intervals is listed below the statement.

GEOLOGIC EXAMPLE

Several statistical studies of oil and gas field frequency distributions have been published (Kaufman, 1964; Drew and Griffiths, 1965; and McCrossan, 1969). Kaufman confined his study to ultimate volumes of fields, while Drew and McCrossan included both ultimate volumes and areas of fields. Kaufman's and Drew's samples apparantly included all fields within specified areas, whereas McCrossan separated fields into categories based on lithology and depositional types. The conclusions reached on the basis of their studies is that the areas and volumes of fields are acceptably approximated by a lognormal distribution. Both Drew and McCrossan, however, had problems fitting this model to part of their field area samples; Drew in the Denver Basin, and McCrossan with his reef pool areas in general. It is my view that field area distribution is of first order importance to the explorationist since it constitutes the critical parameter effecting discovery. Ultimate volumes are more within the province of those people concerned with post-discovery operations. The inconsistencies suggested by these studies and disclosed by some of my early work inspired further examination of the problem.

Sample Description

The fields in this study are categorized by geologic horizon. The horizons are carbonates in the Clearfork (lower Permian) and Devonian formations on the Midland Central Basin Platform in Texas, and the Arbuckle (Ordovician) and Kansas City-Lansing (Pennsylvanian) Groups on the Central Kansas Uplift. These horizons were chosen for analysis because: (1) they have a reasonably consistent lithology over the areas of investigation and have been thoroughly explored within these areas; (2) they contain a sufficiently large number of fields to afford a meaningful sample size; (3) field development is far enough in the past so that the field parameters of area and ultimate volume are well established; and, (4) the data are as comprehensive and reliable as available. Oil fields only are included in the sample, and associated gas volumes are ignored. All oil fields are included, even those abandoned. In fields where both horizons are produced, the pools are separated as accurately as possible. The field areas are based on the maximum acreage attained during their producing history. The field volumes are estimated ultimate volumes including secondarily recoverable oil. Fields discovered after 1965 are not included in the samples to avoid using possibly inaccurate early estimates.

Analytical Results

The statistical attributes of the four horizons considered are listed in Table 1. Almost without exception the skewness, kurtosis, and range statistics for field areas are beyond the acceptable limits for lognormal distribution. In contrast every field volume statistic is compatible with the null hypothesis. This consistent statistical difference is reflected perfectly in Table 2. The null hypothesis for lognormality at the 95 percent significance level is rejected for all field area samples by the chi-square and Kolmogorov-Smirnov tests. Again, all field volume samples are accepted.

Table 3 exhibits the outcomes of the Kolmogorov-Smirnov two sample tests between all combinations of the four samples for areas and ultimate volumes. It is interesting to note that although the field areas in the Clearfork and Kansas City-Lansing formations are rejected as being lognormally distributed they are accepted as coming from the same population distribution. Other interesting relationships are shown between field volume samples. I believe that in context with other studies, this type of analysis is helpful as a guide for investigation of underlying geologic factors that might contribute to establishing analogs for other new or more lightly explored provinces. Moreover, economic predictions may be substantially improved. The computer runs on which these tables are based comprise Figures 1-8.

Chronological Study

The exploration for oil in a newly opened horizon is a highly biased, statistically nonstationary process. The few giant fields that often contain most of the oil are easiest to find, and as a rule are discovered early in the exploration history. As exploratory holes are more densely spaced the probability of giant or large fields remaining undetected decreases as do the expected values for the remaining population. In a province with several producing horizons, all at different stages of development, this process is largely obscured. Figure 9 is included to illustrate how the output from the program SNORT was used to obtain quickly an approximation of the exploration process within the Devonian horizon. The plotted curves show the distribution of logarithms (base 10) of the field volumes in thousands of barrels. Skewness is not shown. The plot output is from another program. The field

Table 1.- Sample statistics.

	HORIZON	LITH	SAMPLE SIZE		MEAN	STANDARD DEVIATION	RANGE/ STD. DEV.	SKEWNESS	KURTOSIS
CENTRAL KANSAS UPLIFT	ARBUCKLE	DOL	250	AREA	2.4472	0.7890	6.92*	0.9430*	4.6987*
				VOLUME	2.3855	1.2049	5.37	0.1556	2.9203
	KANSAS CITY – LANSING	LS	264	AREA	2.1431	0.5781	5.49	0.9707*	3.5520*
				VOLUME	1.9704	0.9729	5.18	0.1615	2.9422
MIDLAND CENTRAL BASIN PLAT.	CLEARFORK	DOL	86	AREA	2.2911	0.7373	4.17*	1.0385*	3.4691
				VOLUME	2.2451	1.1557	4.75	0.2641	2.7947
	DEVONIAN	LS	105	AREA	2.5205	0.6521	3.92*	0.4020*	2.0070*
				VOLUME	2.7554	1.1203	4.57	−0.2218	2.4627

* NULL HYPOTHESIS REJECTED AT 95% SIGNIFICANCE

areas and volumes were arranged chronologically according to date of discovery, and then divided into 5 numerically equal subsamples. Analysis of the subsamples by SNORT showed each field volume subsample to be roughly lognormally distributed, but only the subsamples 56-58 and 58-62 were accepted by the Kolmogorov-Smirnov two sample test. None of the field area subsamples was accepted as lognormally distributed. In this example, as in all others run, the mean and standard deviation change substantially and systematically through time. Identical trends were observed for these parameters of the field area subsamples as well, but the plot routine used for Figure 9 cannot conveniently handle multiple empirical curves so no plot is shown. A listing follows, however.

Subsample	Mean	Standard Deviation
29-52	2.87	0.65
52-56	2.99	0.55
56-58	2.31	0.49
58-62	2.29	0.69
62-65	2.10	0.33

Table 2.- Chi-square and Kolmogorov-Smirnov test for lognormality.

	CLEARFORK		DEVONIAN		ARBUCKLE		KC–LANSING	
	AREA	VOLUME	AREA	VOLUME	AREA	VOLUME	AREA	VOLUME
χ^2	REJECT	ACCEPT	REJECT	ACCEPT	REJECT	ACCEPT	REJECT	ACCEPT
K–S	REJECT	ACCEPT	REJECT	ACCEPT	REJECT	ACCEPT	REJECT	ACCEPT

The parameters of the field area subsamples are based on the logarithms (base 10) of the acreages.

Table 3.- Kolmogorov-Smirnov two sample test.

	CLEARFORK	DEVONIAN	ARBUCKLE	KC–LANSING
		VOLUME		
CLEARFORK	X	REJECT	ACCEPT	ACCEPT
DEVONIAN	REJECT	X	REJECT	REJECT
ARBUCKLE	REJECT	REJECT	X	ACCEPT
KC–LANSING	ACCEPT	REJECT	REJECT	X
AREA				

Obviously, the optimum periods for exploration were 29-52 and 52-56, from the standpoint of likelihood of discovery and the economic rewards attendant on discovery. The low standard deviation of field volumes in the 52-56 period may reflect a combination of few remaining giant fields and the effect of information gleaned from the preceding period reducing the number of small fields discovered. Accidental discoveries related to deeper drilling for other primary objectives accentuate the number of small fields discovered later in the history of the horizon.

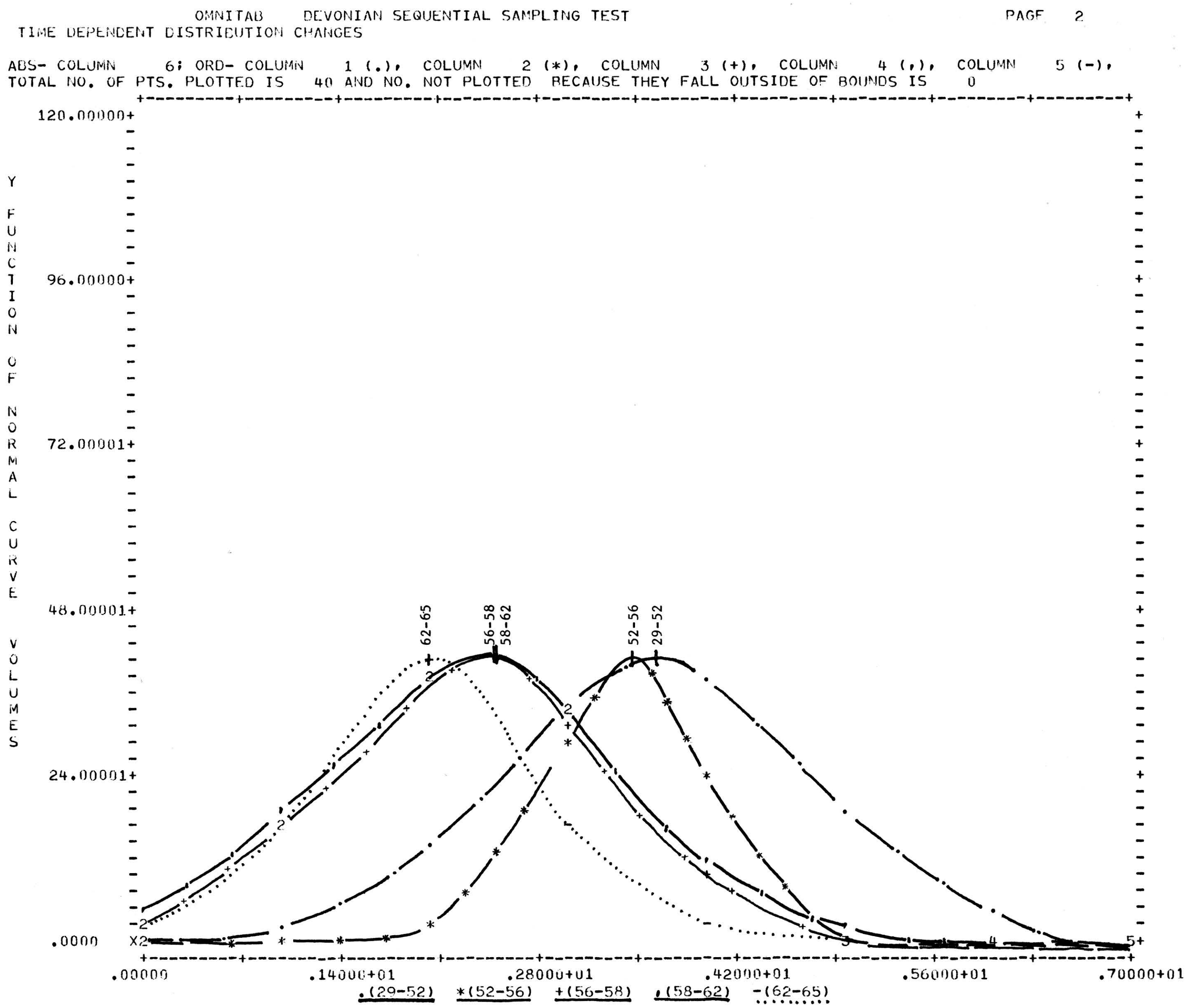


Figure 9.- Devonian field ultimate volume distribution curves by chronological subsamples (x axis plotted as $\log_{10}$ M bbls).

CONCLUSIONS

The parameter of field area, so important to the explorationist, has been accepted as lognormally distributed by most statistical measurements in other oil field studies, although some problems were recognized. The completely consistent rejection of the lognormal null hypothesis in this study is, consequently, somewhat surprising. It is important to remember that the disparity in results carries no connotation of "right" or "wrong". I believe that the explanation lies in the different structure of the samples used in this study. It is obvious that the fields within each horizon comprise a composite of several populations, physically as regards types of geologic traps, etc., and through time as is adequately demonstrated by the chronological analysis. Moreover, there is an apparent layering of populations between horizons shown by the Kolmogorov-Smirnov two sample tests. The extraction of samples with different selection schemes would be expected to lead to different outcomes. This is perfectly legitimate since it is the worker's prerogative to define his own target population depending on the purpose of the study. The target population as defined for this study is considered as well resolved by the samples which almost include the target population. Hence, it is likely that forecasts of discoveries, reserves, or economic returns made on the assumption of a stationary lognormal distribution will be subject to considerable error.

REFERENCES

Drew, L.J., and Griffiths, J.C., 1965, Size, shape, and arrangement of some oil fields in the U.S.A.: Pennsylvania State Univ., Mineral Industries, Contr. No. 64-59, p. F1-F31.

Griffiths, J.C., 1967, Scientific method in the analysis of sediments: McGraw-Hill Book Co., New York, 307 p.

Griffiths, J.C., and Ondrick, C.W., 1968, Sampling a geologic population: Kansas Geol. Survey Computer Contr. 30, 53 p.

Fisher, R.A., 1946, Statistical methods for research workers: Oliver and Boyd, London, 354 p.

Kaufman, G.M., 1963, Statistical decisions and related techniques in oil and gas exploration: Prentice-Hall, Englewood Cliffs, New Jersey, 307 p.

McCrossan, R.G., 1969, An analysis of size frequency distributions of oil and gas reserves of western Canada: Canadian Jour. Earth Sci., v. 6, no. 2, p. 201-211.

Pearson, E.S., and Hartley, H.O., 1966, Biometrika tables for statisticians: Cambridge University Press, New York, 264 p.

ADDITIONAL REFERENCES

Abramowitz, M., and Stegun, I.A., 1965, Handbook of mathematical functions: National Bur. Standards, Washington, D.C., p. 912-913.

Burington, R.S., and May, D.C., 1958, Handbook of probability and statistics: McGraw-Hill Book Co., New York, 332 p.

Lindgren, B.W., and McElrath, G.W., 1966, Introduction to probability and statistics: Macmillan, New York, 288 p.

Spiegel, M.R., 1961, Theories and problems of statistics: Schaum's Outline Series, McGraw-Hill Book Co., New York, 359 p.

Listing of program SNORT

```
C       PROGRAM SNORT
C       AUTHOR--D.A.PRESTON
C       LAST REVISION--JUNE 11,1969
C
C       THIS PROGRAM WILL TEST ONE OR TWO SAMPLE SETS FOR EITHER NORMALITY
C       OR LOGNORMALITY. A KOLMOGOROV-SMIRNOFF PARAMETRIC TEST IS MADE ON
C       EACH SAMPLE, AND THE RESULTS ARE PLOTTED. A CHI-SQUARE TEST IS
C       ALSO MADE OF EACH SAMPLE WITH POOLING OF 1.5 MINIMUM EXPECTED
C       VALUE PER INTERVAL, AS WELL AS 5.0. THE NORMAL CURVE FIT TO THE
C       DATA (OR THEIR LOGARITHMS) IS COMPUTED BY A POLYNOMIAL APPROXI-
C       MATION WHICH HAS AN AVERAGE ERROR OF 10**-6.
C
C
C       SKEWNESS, KURTOSIS, AND THE RATIO OF THE RANGE TO THE STANDARD
C       DEVIATION ARE GIVEN FOR FURTHER TESTS FOR (LOG) NORMALITY. TABLES
C       ARE PROVIDED IN COMPUTER CONTRIBUTION NO. 41 PUBLISHED BY THE
C       KANSAS GEOLOGIC SURVEY.
C
C
C       WHEN TWO SAMPLE SETS ARE RUN A KOLMOGOROV-SMIRNOFF NON-PARAMETRIC
C       TEST IS MADE TO DETERMINE IF THE SAMPLE SETS ARE FROM THE SAME
C       POPULATION.
C
C       CONFIDENCE LIMITS FOR ALL TESTS ARE AT THE .95 LEVEL.
C
C
      DIMENSION S1(101),S2(101),SDIFF(101),SMPL1(5000),SMPL2(5000),PIN(
     125),ONE(25),TWO(25),KONE(25),KTWO(25),KOBS1(25),KOBS2(25),EXPS1(25
     2),EXPS2(25),MAXD1(25),MAXD2(25)
      REAL KURT1,KURT2,MAX1,MAX2,MIN1,MIN2,MAXD1,MAXD2
      INTEGER FINISH
      DATA FINISH/6HFINISH/
      READ(5,101)IVER,IOPT
      IF(IVER.LT.1 .OR.IVER.GT.2)GO TO 50
      IF(IOPT.LT.1 .OR.IOPT.GT.2) GO TO 50
      GO TO 51
   50 WRITE(6,117)
      STOP
C READ IN DATA FROM SAMPLE ONE
   51 DO 33 IX=1,5000
      READ(5,1000)SAMPLE,LOOK
      SMPL1(IX)=SAMPLE
      IF(LOOK.EQ.FINISH)GO TO 34
   33 CONTINUE
   34 NSIZE1=IX-1
      SIZE1=NSIZE1
      MIN1=10.0**12
      MAX1=-1.0*(10.0**12)
      DO 36 I=1,NSIZE1
      IF(MAX1.GT.SMPL1(I))GO TO 35
      MAX1=SMPL1(I)
   35 IF(MIN1.LT.SMPL1(I))GO TO 36
   36 CONTINUE
      MIN1=SMPL1(I)
C CHECK IVER FOR SECOND SAMPLE
      IF(IVER.EQ.1)GO TO 524
      DO 37 IX=1,5000
```

```
      READ(5,1000)SAMPLE,LOOK
      SMPL2(IX)=SAMPLE
      IF(LOOK.EQ.FINISH)GO TO 38
   37 CONTINUE
   38 NSIZE2=IX-1
      SIZE2=NSIZE2
      MIN2=10.0**12
      MAX2=-1.0*(10.0**12)
      DO 40 I=1,NSIZE2
      IF(MAX2.GT.SMPL2(I))GO TO 39
      MAX2=SMPL2(I)
   39 IF(MIN2.LT.SMPL2(I))GO TO 40
      MIN2=SMPL2(I)
   40 CONTINUE
C CHECK IOPT FOR LOG OPTION
  524 IF(IOPT.EQ.1)GO TO 3
C CONVERT DATA TO LOGARITHMS
      MAX1=ALOG10(MAX1)
      MIN1=ALOG10(MIN1)
      DO 1 I=1,NSIZE1
    1 SMPL1(I)=ALOG10(SMPL1(I))
      IF(IVER.EQ.1)GO TO 3
      MAX2=ALOG10(MAX2)
      MIN2=ALOG10(MIN2)
      DO 2 I=1,NSIZE2
    2 SMPL2(I)=ALOG10(SMPL2(I))
C CALCULATE 1,2,3,4 MOMENTS OF SAMPLE ONE
    3 TOTAL1=0.0
      DO 4 I=1,NSIZE1
    4 TOTAL1=TOTAL1+SMPL1(I)
      AMEAN1=TOTAL1/SIZE1
      AMOM21=0.0
      AMOM31=0.0
      AMOM41=0.0
      DO 5 I=1,NSIZE1
      AMOM21=AMOM21+((SMPL1(I)-AMEAN1)**2)/SIZE1
      AMOM31=AMOM31+((SMPL1(I)-AMEAN1)**3)/SIZE1
    5 AMOM41=AMOM41+((SMPL1(I)-AMEAN1)**4)/SIZE1
      STDEV1=SQRT(AMOM21)
      SKEW1=AMOM31/STDEV1**3
      KURT1=AMOM41/STDEV1**4
      RANGE1=MAX1-MIN1
      STAT1=RANGE1/STDEV1
      CLASS1=RANGE1/100.0
      WRITE(6,100)
      WRITE(6,99)
      ILK=1
      DO 21 I=1,11
      IF(I.EQ.7)WRITE(6,124)ILK
   21 WRITE(6,120)
      WRITE(6,118)AMEAN1,STDEV1,SKEW1,KURT1
      WRITE(6,102)STAT1,NSIZE1
C CHECK IVER AND CALCULATE 1,2,3,4 MOMENTS OF SAMPLE TWO
      IF(IVER.EQ.1)GO TO 12
      TOTAL2=0.0
      DO 6 I=1,NSIZE2
    6 TOTAL2=TOTAL2+SMPL2(I)
      AMEAN2=TOTAL2/ SIZE2
      AMOM22=0.0
      AMOM32=0.0
```

```
      AMOM42=0.0
      DO 7 I=1,NSIZE2
      AMOM22=AMOM22+((SMPL2(I)-AMEAN2)**2)/SIZE2
      AMOM32=AMOM32+((SMPL2(I)-AMEAN2)**3)/SIZE2
    7 AMOM42=AMOM42+((SMPL2(I)-AMEAN2)**4)/SIZE2
      STDEV2=SQRT(AMOM22)
      SKEW2=SQRT(AMOM32/STDEV2**3)
      KURT2=AMOM42/STDEV2**4
      RANGE2=MAX2-MIN2
      STAT2=RANGE2/STDEV2
      CLASS2=RANGE2/100.0
C CALCULATE CUMULATIVE FREQUENCY FOR SAMPLES ONE AND TWO
      ARG1=MIN1
      ARG2=MIN2
      DO 8 I=2,101
      ARG1=ARG1+CLASS1
      ARG2=ARG2+CLASS2
      N=0
      M=0
      S1(1)=0.0
      S2(1)=0.0
      DO 9 J=1,NSIZE1
    9 IF(SMPL1(J).LE.ARG1)N=N+1
      DO 10 J=1,NSIZE2
   10 IF(SMPL2(J).LE.ARG2)M=M+1
      AN=N
      AM=M
      S1(I-1)=AN/SIZE1
      S2(I-1)=AM/SIZE2
    8 SDIFF(I-1)=ABS(S1(I-1)-S2(I-1))
C FIND MAXIMUM DEVIATION BETWEEN CURVES AND COMPARE WITH CRITICAL VALUE
C AS DETERMINED FOR KOLMOGOROV-SMIRNOFF TWO SAMPLE TEST(NON-PARAMETRIC)
      CRVAL5=(SQRT((SIZE1+SIZE2)/(SIZE1*SIZE2)))*1.36
      CHOICE=0.0
      DO 11 I=1,100
   11 IF(SDIFF(I).GT.CRVAL5)CHOICE=1.0
C CALCULATE POLYNOMIAL APPROXIMATION TO NORMAL CURVE
   12 Y=(-3.3)
      PIN(1)=0.0
      DO 13 I=2,22
      Y=Y+0.3
      Y1=-0.5*Y**2
      Y2=EXP(Y1)*0.39894
      Y3=ABS(Y)
      Y4=1.0/(1.0+0.33267*Y3)
      PIN(I)=((((0.937298*Y4-0.1201676)*Y4)+0.4361836)*Y4)*Y2
      IF(Y.LT.0.0)GO TO 130
      PIN(I)=1.0-PIN(I)
C CALCULATE EXPECTED VALUES IN INTERVALS OF 0.3 STANDARD DEVIATIONS
  130 EXPS1(I-1)=(PIN(I)-PIN(I-1))* SIZE1
   13 IF(IVER.EQ.2)EXPS2(I-1)=(PIN(I)-PIN(I-1))* SIZE2
      ONE(1)=0.0
      TWO(1)=0.0
      TNORM1=AMEAN1-3.3*STDEV1
      IF(IVER.EQ.2)TNORM2=AMEAN2-3.3*STDEV2
      KONE(1)=0
      KTWO(1)=0
      DO 14 I=2,22
      N=0
      M=0
```

```
      TNORM1=TNORM1+0.3*STDEV1
      IF(IVER.EQ.2)TNORM2=TNORM2+0.3*STDEV2
      DO 15 J=1,NSIZE1
      IF(SMPL1(J).LE.TNORM1)N=N+1
      AN=N
C CALCULATE CUMULATIVE FREQUENCIES AT INTERVAL BOUNDARIES FOR SAMPLE ONE
      ONE(I)=AN/SIZE1
C DETERMINE COUNT OF SAMPLE ONE DATA IN 0.3 ST DEV INTERVALS
      KONE(I)=N
      KOBS1(I-1)=KONE(I)-KONE(I-1)
C FIND MAXIMUM DEVIATION FROM NORMAL CUMULATIVE FREQUENCY FOR SAMPLE ONE
   15 MAXD1(I-1)=ABS(ONE(I-1)-PIN(I-1))
      IF(IVER.EQ.1)GO TO 14
      DO 16 J=1,NSIZE2
      IF(SMPL2(J).LE.TNORM2)M=M+1
      AM=M
C CALCULATE CUMULATIVE FREQUENCIES AT INTERVAL BOUNDARIES FOR SAMPLE TWO
      TWO(I)=AM/SIZE2
C DETERMINE COUNT OF SAMPLE TWO DATA IN 0.3 ST DEV INTERVALS
      KTWO(I)=M
      KOBS2(I-1)=KTWO(I)-KTWO(I-1)
C FIND MAXIMUM DEVIATION FROM NORMAL CUMULATIVE FREQUENCY FOR SAMPLE TWO
   16 MAXD2(I-1)=ABS(TWO(I-1)-PIN(I-1))
   14 CONTINUE
C CALCULATE CRITICAL VALUE FOR K-S TEST OF SAMPLE 1 VS NORMAL DISTRIB.
      CVNT1=1.36/SQRT(SIZE1)
      CAPUT=0.
      DO 17 I=1,22
   17 IF(MAXD1(I).GE.CVNT1)CAPUT=1.
      IF(CAPUT.EQ.0.0)WRITE(6,108)CVNT1
      IF(CAPUT.EQ.1.0)WRITE(6,109)CVNT1
      WRITE(6,111)
      WRITE(6,112)(ONE(I),PIN(I),MAXD1(I),I=1,22)
      BANNER=1.0
      CALL PLOT(ONE,PIN,MAX1,MIN1,BANNER)
      IF(IVER.EQ.1)GO TO 19
C CALCULATE CRITICAL VALUE FOR K-S TEST OF SAMPLE 2 VS NORMAL DISTRIB.
      CVNT2=1.36/SQRT(SIZE2)
      CRATER=0
      DO 18 I=1,22
   18 IF(MAXD2(I).GE.CVNT2)CRATER=1.
C OUTPUT
   19 WRITE(6,113)
      WRITE(6,114)
      WRITE(6,115)(EXPS1(I),KOBS1(I),I=1,21)
C CALCULATE THE CHI SQUARE VALUE FOR SAMPLE ONE WITH MINIMUM EXPECTED VALUES
C OF 1.5 IN EACH INTERVAL
      CULP=0.0
      KOB=0
      INDEX=0
      DO 61 I=1,21
      KOB=KOB+KOBS1(I)
      CULP=CULP+EXPS1(I)
      IF(CULP.LT.1.5)GO TO 61
      INDEX=INDEX+1
      EXPS1(INDEX)=CULP
      KOBS1(INDEX)=KOB
      CULP=0.0
      KOB=0
   61 CONTINUE
```

```
      CHI=0.0
      IDF=INDEX-3
      DO 63 I=1,INDEX
   63 CHI=CHI+((FLOAT(KOBS1(I))-EXPS1(I))**2)/EXPS1(I)
      WRITE(6,121)CHI,IDF
C CALCULATE THE CHI SQUARE VALUE FOR SAMPLE ONE WITH MINIMUM EXPECTED VALUES
C OF 5.0 IN EACH INTERVAL
      INDEC=0
      DO 65 I=1,INDEX
      KOB=KOB+KOBS1(I)
      CULP=CULP+EXPS1(I)
      IF(CULP.LT.5.0)GO TO 65
      INDEC=INDEC+1
      EXPS1(INDEC)=CULP
      KOBS1(INDEC)=KOB
      CULP=0.0
      KOB=0
   65 CONTINUE
      CHI=0.0
      IDF=INDEC-3
      DO 66 I=1,INDEC
   66 CHI=CHI+((FLOAT(KOBS1(I))-EXPS1(I))**2)/EXPS1(I)
      WRITE(6,123)CHI,IDF
      IF(IVER.EQ.1)GO TO 20
      WRITE(6,122)
      ILK=2
      DO 22 I=1,11
      IF(I.EQ.7)WRITE(6,124)ILK
   22 WRITE(6,120)
      IF(IVER.EQ.2)WRITE(6,119)AMEAN2,STDEV2,SKEW2,KURT2
      WRITE(6,102)STAT2,NSIZE2
      IF(CRATER.EQ.0.0)WRITE(6,108)CVNT2
      IF(CRATER.EQ.1.0)WRITE(6,109)CVNT2
      WRITE(6,111)
      WRITE(6,112)(TWO(I),PIN(I),MAXD2(I),I=1,22)
      BANNER=2.0
      CALL PLOT(TWO,PIN,MAX2,MIN2,BANNER)
      WRITE(6,116)
      WRITE(6,114)
      WRITE(6,115)(EXPS2(I),KOBS2(I),I=1,21)
C CALCULATE THE CHI SQUARE VALUE FOR SAMPLE TWO WITH MINIMUM EXPECTED VALUES
C OF 1.5 IN EACH INTERVAL
      CULP=0.0
      KOB=0
      INDEX2=0
      DO 62 I=1,21
      KOB=KOB+KOBS2(I)
      CULP=CULP+EXPS2(I)
      IF(CULP.LT.1.5)GO TO 62
      INDEX2=INDEX2+1
      EXPS2(INDEX2)=CULP
      KOBS2(INDEX2)=KOB
      CULP=0.0
      KOB=0
   62 CONTINUE
      CHI=0.0
      IDF=INDEX2-3
      DO 64 I=1,INDEX2
   64 CHI=CHI+((FLOAT(KOBS2(I))-EXPS2(I))**2)/EXPS2(I)
      WRITE(6,121)CHI,IDF
```

```
C CALCULATE THE CHI SQUARE VALUE FOR SAMPLE TWO WITH MINIMUM EXPECTED VALUES
C OF 5.0 IN EACH INTERVAL
      INDEX5=0
      DO 67 I=1,INDEX2
      KOB=KOB+KOBS2(I)
      CULP=CULP+EXPS2(I)
      IF(CULP.LT.5.0)GO TO 67
      INDEX5=INDEX5+1
      EXPS2(INDEX5)=CULP
      KOBS2(INDEX5)=KOB
      CULP=0.0
      KOB=0
   67 CONTINUE
      CHI=0.0
      IDF=INDEX5-3
      DO 68 I=1,INDEX5
   68 CHI=CHI+((FLOAT(KOBS2(I))-EXPS2(I))**2)/EXPS2(I)
      WRITE(6,123)CHI,IDF
      WRITE(6,125)
      WRITE(6,103)CRVAL5
      IF(CHOICE.EQ.1.0)WRITE(6,106)
      IF(CHOICE.NE.1.0)WRITE(6,107)
      WRITE(6,104)
      WRITE(6,105)(S1(I),S2(I),SDIFF(I),I=1,100)
C FORMAT STATEMENTS
   99 FORMAT(1H0,40HPUBLISHED BY THE KANSAS GEOLOGIC SURVEY-         )
  100 FORMAT(1H1,90HTHE TESTS FOR (LOG)NORMALITY IN THIS PROGRAM ARE DIS
     1CUSSED IN COMPUTER CONTRIBUTION NO. 41)
  101 FORMAT(2I5)
  102 FORMAT(1H0,40HTHE RATIO OF THE RANGE TO THE STD DEV = ,F5.2,20H  T
     1HE SAMPLE SIZE = ,I5)
  103 FORMAT(1H0,88HCRITICAL VALUE OF KOLMOGOROV-SMIRNOFF TWO SAMPLE TES
     1T FOR ACCEPTANCE AT .05 IS LESS THAN,F6.5,9H ABS DIFF)
  104 FORMAT(1H0,30H CUM FREQ  CUM FREQ  ABS.DIFF )
  105 FORMAT(2X,F6.4,4X,F6.4,4X,F6.4)
  106 FORMAT(1H0,32HYOUR NULL HYPOTHESIS IS REJECTED)
  107 FORMAT(1H0,32HYOUR NULL HYPOTHESIS IS ACCEPTED)
  108 FORMAT(1H0, 92HYOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS ACCEPTED
     1 AT THE .05 LEVEL--K-S CRITICAL VALUE IS ,F6.5,9H ABS DIFF)
  109 FORMAT(1H0, 92HYOUR NULL HYPOTHESIS FOR (LOG)NORMALITY IS REJECTED
     1 AT THE .05 LEVEL--K-S CRITICAL VALUE IS ,F6.5,9H ABS DIFF)
  111 FORMAT(1H0,20X,30H  CUM FREQ  EXP FREQ  ABS DIFF)
  112 FORMAT(20X,3F10.5)
  113 FORMAT(1H1,120HTHE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQU
     1ARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0.3 ST DEV--SAMPLE 1--
     2ARE)
  114 FORMAT(1H0,20H EXPECTED  OBSERVED )
  115 FORMAT(F10.1,I10)
  116 FORMAT(1H1,120HTHE EXPECTED AND OBSERVED FREQUENCIES FOR A CHI SQU
     1ARE TEST OF (LOG)NORMALITY IN INTERVALS OF 0.3 ST DEV--SAMPLE 2--
     2ARE)
  117 FORMAT(1H1,30HCHECK PARAMETERS IOPT AND IVER)
  118 FORMAT(1H0,23HSAMPLE 1 HAS A MEAN OF ,F10.4,25H,A STANDARD DEVIATI
     1ON OF ,F6.4,14H--SKEWNESS IS ,F6.4,10H,KURTOSIS ,F6.4,1H.)
  119 FORMAT(1H0,23HSAMPLE 2 HAS A MEAN OF ,F10.4,25H,A STANDARD DEVIATI
     1ON OF ,F6.4,14H--SKEWNESS IS ,F6.4,10H,KURTOSIS ,F6.4,1H.)
  120 FORMAT(1H0)
  121 FORMAT(1H0,58HCHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF
     11.5 = ,F8.3,6H WITH ,I2,19H DEGREES OF FREEDOM)
  122 FORMAT(1H1)
```

```
  123 FORMAT(1H0,58HCHI SQUARE TEST VALUE FOR MINIMUM EXPECTED VALUE OF
     15.0 = ,F8.3,6H WITH ,I2,19H DEGREES OF FREEDOM)
  124 FORMAT(1H ,41X,7HSAMPLE ,I1,11H STATISTICS)
  125 FORMAT(1H1,40X,31HSAMPLE 1 VS. SAMPLE 2--K-S TEST)
 1000 FORMAT(10X,F10.0,54X,A6)
   20 STOP
      END
$       FORTRAN
      SUBROUTINE PLOT(ONE,PIN,AMAX,AMIN,BANNER)
      DIMENSION INC(130),YINC(50),ONE(22),PIN(22),CURVES(4),FRAME(2),TIT
     1LE(20),SCALE(4)
      INTEGER FRAME,CURVES,TITLE,SCALE
      DATA CURVES/1H1,1H2,1H*,1H /
      DATA FRAME/1H-,1H+/
      DATA TITLE/1HC,1HU,1HM,1HU,1HL,1HA,1HT,1HI,1HV,1HE,1H ,1HF,1HR,1HE
     1,1HQ,1HU,1HE,1HN,1HC,1HY/
      DATA SCALE/1H1,1H.,1H0,1H5/
      IF(BANNER.EQ.1.0)WRITE(6,201)
      IF(BANNER.EQ.2.0)WRITE(6,204)
      WRITE(6,205)
      WRITE(6,202)
      YINC(1)=1.025
      IY=12
      IT=0
      IMARK=7
      DO 1 I=2,43
      DO 2 J=1,94
    2 INC(J)=CURVES(4)
      INC(6)=FRAME(1)
      INC(95)=FRAME(1)
      YINC(I)=YINC(I-1)-0.025
      IF(IY.EQ.I)GO TO 16
      GO TO 17
   16 IY=IY+1
      IF(IY.GT.32)GO TO 17
      IT=IT+1
      INC(1)=TITLE(IT)
   17 CONTINUE
      DO 3 J=1,22
      JI=0
      JIJ=0
      IF(PIN(J).GE.YINC(I).AND.PIN(J).LT.YINC(I-1))GO TO 4
      GO TO 5
    4 JI=J*4+6
      INC(JI)=CURVES(1)
    5 IF(ONE(J).GE.YINC(I).AND.ONE(J).LT.YINC(I-1))GO TO 6
      GO TO 3
    6 JIJ=J*4+6
      IF(JIJ.EQ.JI)GO TO 8
      INC(JIJ)=CURVES(2)
      GO TO 3
    8 INC(JIJ)=CURVES(3)
    3 CONTINUE
      IF(YINC(I).EQ.1.0)GO TO 9
      GO TO 10
    9 INC(4)=SCALE(1)
      INC(5)=SCALE(2)
      INC(6)=SCALE(3)
      GO TO 1
   10 IF(YINC(I).GT.0.499.AND.YINC(I).LT.0.5001)GO TO 11
```

```
      GO TO 12
   11 INC(4)=SCALE(3)
      INC(5)=SCALE(2)
      INC(6)=SCALE(4)
      GO TO 1
   12 IF(YINC(1).LT.(-0.001))GO TO 13
      GO TO 1
   13 INC(4)=SCALE(3)
      INC(5)=SCALE(2)
      INC(6)=SCALE(3)
    1 WRITE(6,200)(INC(M),M=1,95)
      DO 26 L=1,6
   26 INC(L)=CURVES(4)
      DO 14 L=7,94
      IF(IMARK.EQ.L)GO TO 15
      INC(L)=FRAME(1)
      GO TO 14
   15 IMARK=IMARK+22
      INC(L)=FRAME(2)
   14 CONTINUE
      WRITE(6,200)(INC(M),M=1,95)
      AMID=AMIN+(AMAX-AMIN)/2.0
      WRITE(6,203)AMIN,AMID,AMAX
      WRITE(6,206)
  200 FORMAT(1X,95A1)
  201 FORMAT(1H1,30X,39HK-S TEST FOR (LOG)NORMALITY SAMPLE ONE )
  202 FORMAT(1H0)
  203 FORMAT(1H ,1PE11.3,30X,1PE13.3,30X,1PE13.3)
  204 FORMAT(1H1,30X,39HK-S TEST FOR (LOG)NORMALITY SAMPLE TWO )
  205 FORMAT(1H ,32X,34HCUMULATIVE NORMAL*1*  EMPIRICAL*2*)
  206 FORMAT(1H0,42X,13HRANGE OF DATA)
      RETURN
      END
```

KANSAS GEOLOGICAL SURVEY COMPUTER PROGRAM
THE UNIVERSITY OF KANSAS, LAWRENCE

PROGRAM ABSTRACT

Title (If subroutine state in title):

SNORT - FORTRAN IV program for sample normality tests

Date:

Author, organization: D.A. Preston

Shell Development, Houston, Texas

Direct inquiries to: D.A. Preston

Name:

Address: P.O. Box 481

Houston, Texas 77001

Purpose/description: Selected tests of samples are made to determine (log) normality of parent populations of one or two samples.

Mathematical method: Polynomial approximation to normal distribution curve, chi-square test, Kolmogorov-Smirnov test (parametric and two sample).

Restrictions, range:

Computer manufacturer: GE or RAND

Model: 635 or 1108

Programming language: FORTRAN IV

Memory required: K

Approximate running time: 15 sec (1108)

Special peripheral equipment required: None

Remarks (special compilers or operating systems, required word lengths, number of successful runs, other machine versions, additional information useful for operation or modification of program)

Plate I. - Acceptance limits for $\sqrt{\beta_1}$ at 95% significance level

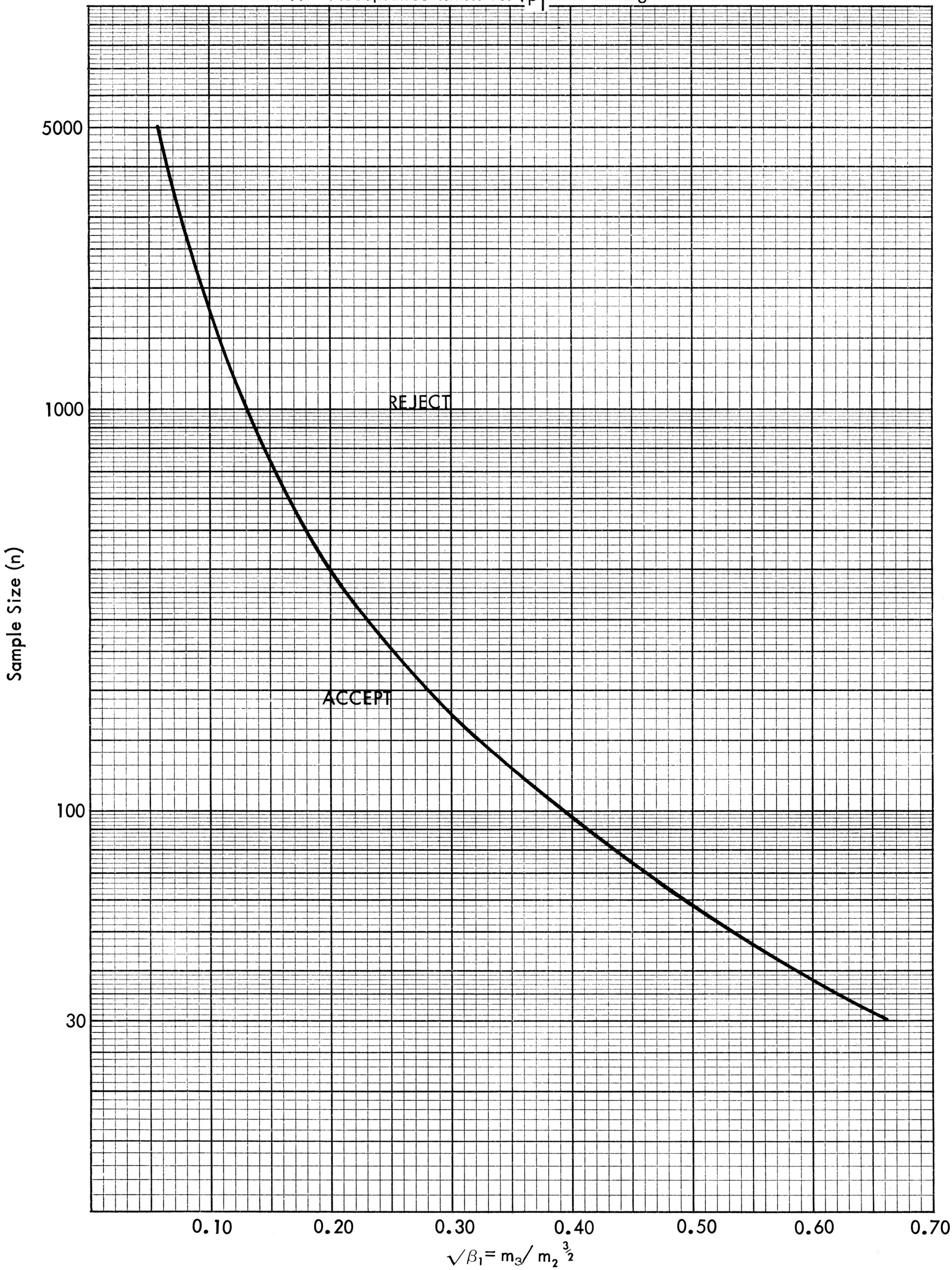


Plate II. - Acceptance limits for β_2 at 95% significance level

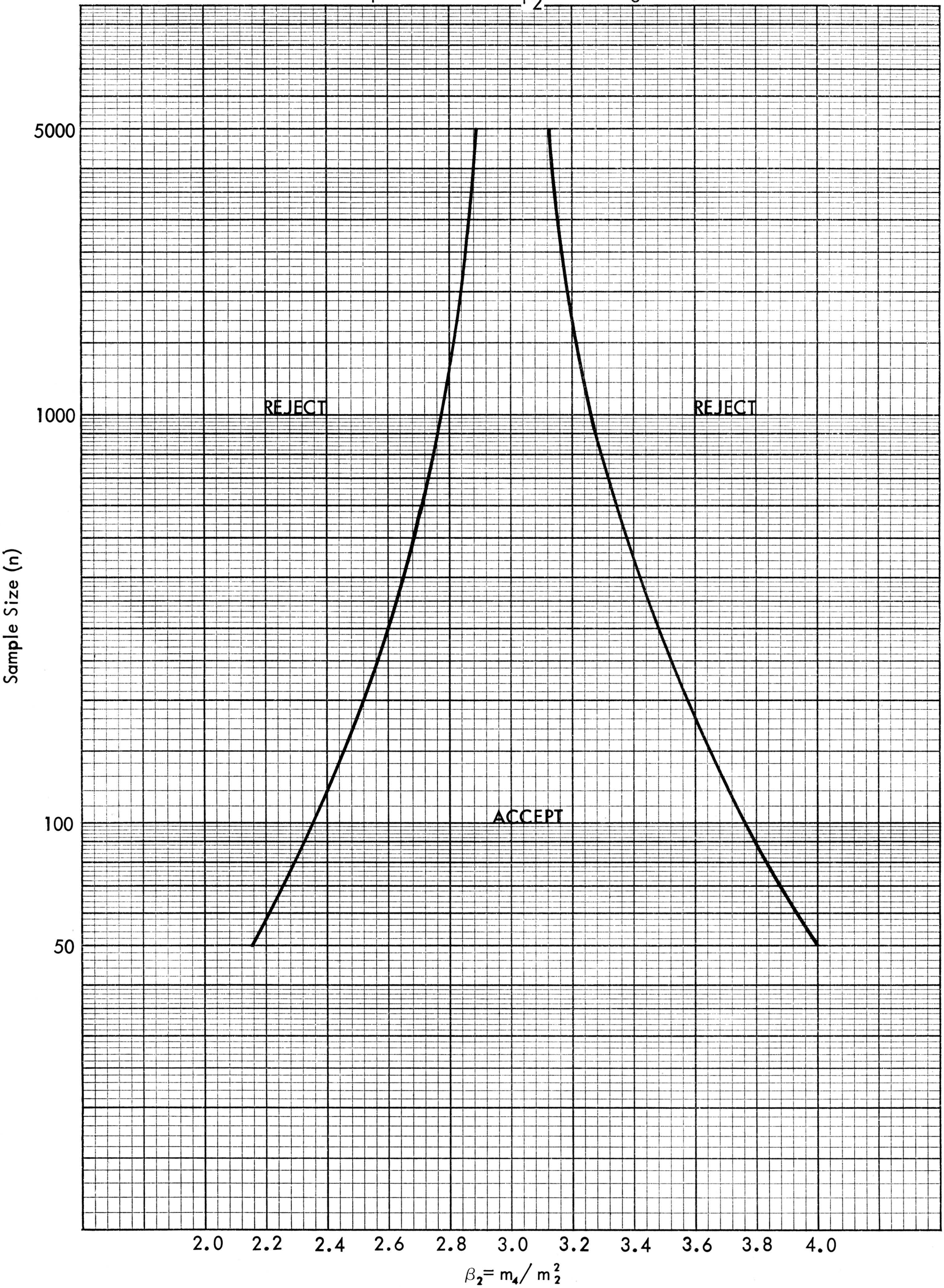


Plate III. - Acceptance limits for range/standard deviation at 95% significance level