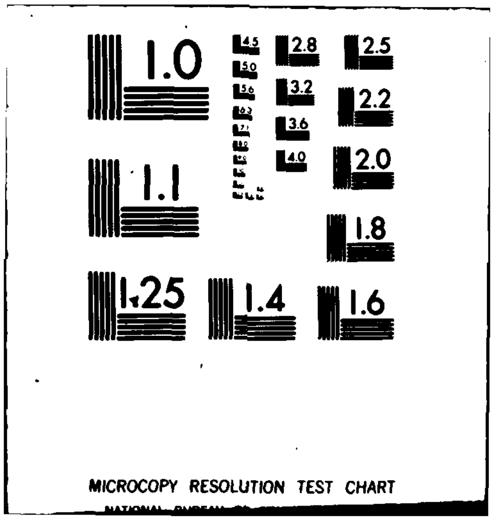
AD-A081 680   NAVY UNDERWATER SOUND LAB NEW LONDON CT   F/G 12/1
VARIANCE TEST FOR NORMALITY.(U)
DEC 69 J SKORY
UNCLASSIFIED   NUSL-TM-2211-343-69   NL

1 of 1

END
DATE
FILMED
4-80
DTIC


1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

LEVEL II (0)

(00688)

Copy No. 9

Code No. 2064.2

(14) NUSL-TM-2211-343-69

NAVY UNDERWATER SOUND LABORATORY
NEW LONDON, CONNECTICUT 06320

(6) VARIANCE TEST FOR NORMALITY

USL Problem No.
0-A-055-00-00
(17) ZR 011 01/01

by

(10) John Skory

NUSL Technical Memorandum No. 2211-343-69

(11) 30 Dec[redacted]69

INTRODUCTION

For sample sizes up to n = 50 an ordered sample $(X_1 \ldots X_n)$ may be tested for normality by a calculation involving a vector $(V_1 \ldots V_n)$ that is tabulated in (1). If V and X are column vectors the test statistic, also tabulated in (1), is

(9) Technical memo

(12) 3

$$W = \frac{(V_i^T X_i)^2}{(X_i - \bar{X})^2} \qquad i = 1 \ldots n$$

The distribution of W ranges from 0 to 1 where as W approaches 1.0 the distribution of $(X_1 \ldots X_n)$ comes closer and closer to being Gaussian.

When dealing with a very large sample, N measurements, how does one proceed to test for normality since V for samples greater than 50 is not available? A common procedure is the chi squared test for deviations from an expected distribution. A more attractive but approximate method, outlined below, is the taking of a sample of n = 50 from the cross section of the large sample of N values. These 50 values may

254200 Gm

then be tested by $(V_1 \ldots V_{50})$ from the table in (1).

METHOD

The procedure for obtaining n = 50 values from N values is:

The i^{th} ordered value of the sample $(X_1 \ldots X_{50})$ equals the $(NP_i + \frac{50-i}{49})$th ordered value of sample $(X_1 \ldots X_N)$ rounding to the nearest integral value.

The vector $P = (P_1 \ldots P_{50})$ is obtained from (2) using 50 percent ranks with the tails of the distribution adjusted outward to give a standard deviation for $(X_1 \ldots X_{50})$ close to that of the large sample $(X_1 \ldots X_N)$ and with W close to 1. The vector P is:

P = .0048,.0322,.0531,.0729,.0928,.1126,.1325,.1524,.1722,.1921,.2119,
.2318,.2517,.2715,.2914,.3113,.3311,.3510,.3709,.3907,.4106,.4305,
.4503,.4702,.4901,.5099,.5298,.5497,.5695,.5894,.6093,.6291,.6490,
.6689,.6887,.7086,.7285,.7483,.7682,.7880,.8079,.8278,.8476,.8675,
.8873,.9072,.9270,.9469,.9678,.9952

If N = 10,000 the required set $(X_1 \ldots X_{50})$ equals the (49th, 323rd, 532nd ...9952nd) ordered values of the sample $(X_1 \ldots X_{10,000})$.

The vector $(V_1 \ldots V_{50})$ = (-.3751,...-.0035,.0035,...,.3751) and W would be calculated as usual. From tables of the unit normal distribution a cross-section sample of 50 yielded a sigma of 1.004 (for 50 degrees of freedom) and a W of .9996, both sufficiently close to unity for general application.

SUMMARY

An alternate method has been outlined to test if a large sample is Gaussian in distribution. Instead of a chi squared test of fit a new statistic W is evaluated using a cross-section sample of 50 from a much larger sample of data. If the large sample is at least 100, the technique yields reliable results which may be assessed for significance against tabulated percentiles of W.

John Skory

JOHN SKORY
Math Statistician

REFERENCES

1. Shapiro, S.S., and Wilk, M.B., "An Analysis of Variance Test for Normality," Biometrika, 52, 3 and 4, p 591, (1965).

2. Failure Distribution Analyses Study, Volume 3, Tables of Plotting Positions, Dept of the Navy, CAI-NY 6052, Contract NObsr-89126, AD631527, (1964).