Not in Trade

Reprint from

Archiv für Meteorologie, Geophysik und Bioklimatologie

Serie A: Meteorologie und Geophysik, Band 16, Heft 2–3, 1967

Herausgegeben von

Doz. Dr. W. Mörikofer, Davos, und Prof. Dr. F. Steinhauser, Wien

Springer-Verlag / Wien · New York

551.501.45

(Department of Meteorology, Texas A & M University, College Station, Texas, U. S. A.)

Choosing a Test of Normality for Small Samples

By

J. J. O'Brien[1] and J. F. Griffiths

With 1 Figure

(Received March 2, 1967)

Summary. The Cornu criterion and the coefficient of skewness test are shown to be preferable to other tests when testing the hypothesis that small samples, less than 100, have been obtained from a normal population. The Cornu ratio of mean deviation to standard deviation is highly negatively correlated with the coefficient of kurtosis.

Zusammenfassung. Es wird gezeigt, daß der Cornu-Test und der Test mittels des Schiefekoeffizienten anderen Tests vorzuziehen sind, wenn man prüft, ob kleine Stichproben mit weniger als 100 Beobachtungen aus einer Normalverteilung stammen. Das von Cornu verwendete Verhältnis zwischen der mittleren Abweichung und der Streuung ist stark negativ mit dem Exzeßkoeffizienten korreliert.

Résumé. On démontre ici que le test de Cornu et celui du coefficient d'obliquité est préférable à tout autre lorsqu'il s'agit d'examiner si de petits échantillons de moins de 100 observations proviennent d'une répartition normale. Le rapport utilisé par Cornu entre l'écart moyen et l'écart quadratique moyen est corrélé d'une manière fortement négative avec le coefficient de l'excès.

1. Introduction

In many meteorological and geophysical problems there is need to study the distribution of parameters involved. This is especially so for meteorological elements and samples of organism populations. If the distribution can be identified as being not significantly different from the Gaussian or normal population then many forms

[1] Present affiliation: National Center for Atmospheric Research, Boulder, Colorado.

of analysis can be simplified. This paper discusses the problem of testing the hypothesis that a small sample belongs to the normal population and is intended for the student and for users of statistics rather than the expert.

On many occasions the sample of data available is small, less than 100. This circumstance is very common in meteorology, especially in climatology when examining annual totals of a parameter. The literature has many examples wherein either no test or non-allowable tests have been used and this short note is an attempt to clarify this situation. In addition, we shall demonstrate, using "theoretical" and practical examples, that the Cornu and skewness tests are preferable choices among the many tests available.

2. Tests of Normality

If a large series is being tested in order to compare it with a known distribution, such as the normal, then the χ^2 test is ideal. However, the test has limitations, namely:

(1) When the degree of freedom is unity, each expected frequency should be at least 5.

(2) When the degrees of freedom exceed unity, at least 80 per cent of the expected frequencies should be 5 or more (Cochran [2]).

For small samples these criteria mean that only a few classes can be used so that the χ^2 test is of limited use. When the sample size, N, is less than about 40, this test becomes subjective and inappropriate.

Other tests of normality which have been suggested by various authors are the Cornu, skewness, kurtosis, Chauvenet (e. g., Brooks and Carruthers [1]) and Kolmogorov-Smirnov test (e. g., Siegel [6]). The use of probability graph paper will not be considered here. Cornu's criterion tests the sample ratio of the mean deviation, e, to the standard deviation, $\hat{\sigma}$. The skewness and kurtosis test utilize respectively the moment coefficient of skewness, γ_1, and the moment coefficient of kurtosis, γ_2, in a Student's t test. Chauvenet's criterion tests the probability of occurrence of extreme values. The $K - S$ test is a test of goodness of fit which compares the theoretical and observed cumulative frequency distributions.

For a sample of size, N, and individual elements, x_i, we define:

$$m_r \quad \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^r, \qquad (1)$$

$$\bar{x} \quad \frac{1}{N} \sum_{i=1}^{N} x_i, \qquad (2)$$

$$\hat{\sigma}^2 = m_2, \tag{3}$$

$$\gamma_1 \quad m_3/\sigma^3, \tag{4}$$

$$\gamma_2 = m_4/\sigma^4 - 3, \tag{5}$$

$$e = \frac{1}{N} \sum_{i=1}^{N} |x_i \quad \bar{x}|. \tag{6}$$

For a normal distribution, $e/\sigma = \sqrt{2/\pi}$. GEARY [3] gives the 0.01, 0.05, and 0.10 probability limits for e/σ for samples of 11—1001. For the t test γ_1 has a standard error of approximately $\sqrt{6/N}$ and an expected value of zero. The coefficient of kurtosis, γ_2, has a standard error of approximately $\sqrt{24/N}$, but, for $N < 500$, γ_2 has an expected value which differs from zero (PEARSON [5]).

The principal difficulty which arises, however, is that each of these tests is presented in such a way that they are necessary but not sufficient for the identification of normality. For example, infinite samples drawn from all symmetrical populations will show a zero value of γ_1. Hence, the skewness test is only a test of symmetry and cannot distinguish the normal from other symmetrical populations. In the case of the Cornu test, many populations have an expected value of e/σ of approximately 0.8, e. g., uniform, $e/\sigma = 0.866$; exponential, $e/\sigma = 0.734$; etc. In other words if a random sample does not pass the Cornu test at the 99 per cent confidence level, there is only a 1 in 100 chance that the series is from a normal population (Type I error). If, on the other hand, it passes the test no definite statement can be made regarding the probability limit for a Type II error since random samples other than the normal can also pass the criterion. In our opinion, it is necessary to run several tests on small samples in order to minimize Type II errors. If a given series passes all the tests, this greatly increases the investigator's belief that the hypothesis of normality is acceptable. However, this is true only if all the tests are independent. If the tests are not independent, the results of testing are biased.

3. Results

The independence of the Cornu, skewness and kurtosis tests was examined using random samples of sizes 5(5)40 (100 samples of each size) generated by two random number subroutines on a high speed digital computer. These subroutines utilize the power-residue technique to produce pseudo-random numbers. Over 8.5 billion numbers can be obtained before repetition occurs. Random numbers were generated from a uniform population ($\mu = 50$, range $= 100$) and a normal population ($\mu = 0$, $\sigma = 25$).

The results of linear regression indicated that, while the Cornu ratio and skewness are not correlated, the Cornu ratio and kurtosis are very highly correlated (Table 1). This agrees favorably with a

Table 1. *Linear Regression of Test Parameters of 100 Series of Various Sizes*

Sample Size	Random Normal Population Correlation Coefficient between		Random Uniform Population Correlation Coefficient between	
	Cornu & γ_1	Cornu & γ_2	Cornu & γ_1	Cornu & γ_2
5	−0.079	−0.881	−0.053	−0.890
10	−0.004	−0.888	0.029	−0.843
15	0.027	−0.857	0.013	−0.779
20	0.262	−0.850	−0.116	−0.881
25	0.152	−0.800	0.003	−0.825
30	−0.139	−0.795	−0.065	−0.816
35	0.021	−0.795	−0.161	−0.887
40	−0.059	−0.850	−0.140	−0.887

theoretical analysis by Geary [3] which indicated that the correlation coefficient between the Cornu ratio and γ_2 approaches -0.78 as N tends to infinity. Hence, on the basis of this empirical evidence, we conclude that the Cornu and γ_2 tests are not linearly independent, whereas the Cornu criterion and skewness test appear to be linearly independent tests.

The next step was to examine a practical meteorological example. For this, the 50 years (1914—1963) of maximum and minimum temperatures at College Station, Texas, were grouped into 730 samples, each containing the relevant extreme value for each day, January 1, January 2, and so on. There is no physical basis to expect any of these temperature samples to be from a population with a Gaussian distribution. It may have been more pertinent to choose data (such as monthly mean temperatures) which are expected to be Gaussian. However, our purpose is to demonstrate empirically the effectiveness of the various standard tests in rejecting small samples as being from normal populations.

The resultant plot of Cornu ratio versus γ_2 for minimum temperature is shown in Fig. 1. There is a linear correlation coefficient of -0.831 for these data. The actual relationship between the variates is apparently nonlinear. Similar results were found for the maximum temperature series and the series from the random number subroutines. The curve illustrates that as the distribution becomes markedly leptokurtic (γ_2 positive), the Cornu ratio remains conservative at about 0.70 ± 0.04. As might be expected from the Figure, leptokurtic samples which failed the Cornu test generally failed the

Table 2
Tests of Normality on 730 Temperature Samples of Size 50 at 95 °/o Level

Tests	No. of Samples Failing	Tests	No. of Samples Failing
Cornu	.. 216	K-S	14
Kurtosis, γ_2.	.. 145	ANY	374
Skewness, γ_1	 243	Cornu and/or γ_1..............	355
Chauvenet..	 31	Cornu, γ_1, and/or Chauvenet ...	368

kurtosis test while for platykurtic samples, only the Cornu test indicated failures. Hence, it is possible to conclude that for small samples the Cornu test is more powerful for detecting departures from normality than the kurtosis test.

An interesting aspect of the tests on the 730 samples is illustrated in Table 2. This shows that by using the Cornu and skewness tests it is possible to reject the hypothesis of normality for the majority

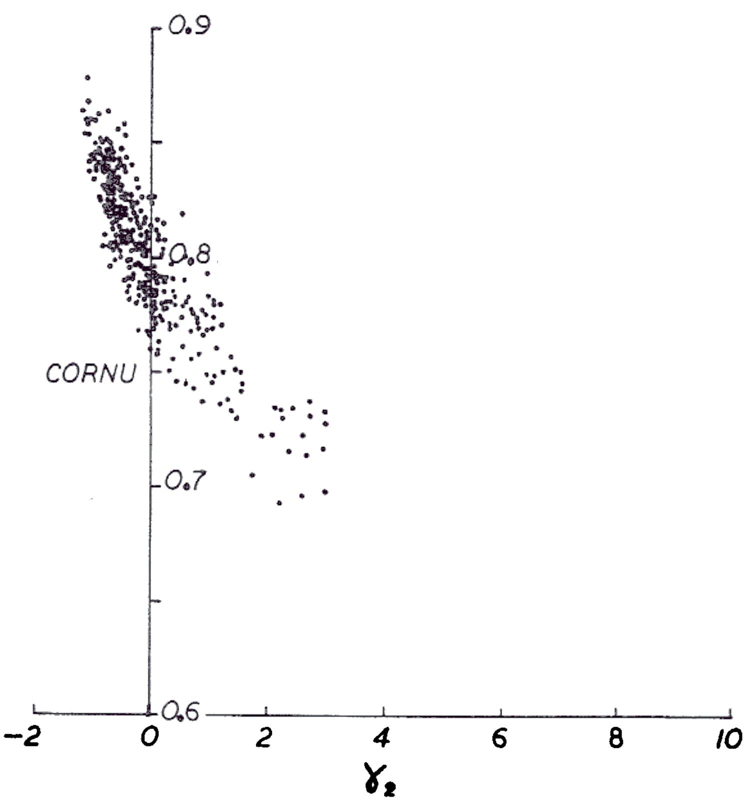

Fig. 1. Cornu criterion versus moment coefficient of kurtosis for 365 minimum temperature samples of size 50 for College Station, Texas

of samples. This capability of these two tests to be able to detect departures from the Gaussian distribution in small samples is regarded as an important and desirable attribute.

The weakness of the $K - S$ test was surprising. Only 14 samples failed at 95 % level and 47 at 90 % level. Apparently this test is similar to χ^2 in that it is a test of goodness of fit and is less sensitive to fine departures from normality.

An unexpected uniform distribution of failures from month to month arose from these investigations. Each month has a large

Table 3. *Seasonal Distribution of Non-Normal Samples*

	Jan.	Febr.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
No. of Samples Failing Any Test	25	25	38	36	21	33	35	42	43	27	26	23

proportion (30 % in May to 60 % in September) of non-normal samples (Table 3).

4. Conclusions

For small samples, less than 100, it appears to be sufficient to use only the Cornu ratio and skewness tests to detect departures from normality. The kurtosis test is related to the Cornu test. The goodness of fit tests such as χ^2 and $K - S$ tests are not as powerful and are generally inapplicable for small samples.

Acknowledgements

We wish to thank Professor R. R. Hocking of the Texas A & M University Institute of Statistics and Professor Irving Weiss of the University of Colorado for reviewing the paper and the Texas A & M University Data Processing Center for supplying the IBM 7094 computer time for this research.

References

1. Brooks, C. E. P., and N. Carruthers: Handbook of Statistical Methods in Meteorology. London, 1953.
2. Cochran, W. G.: Some Methods for Strengthening the Common χ^2 test. Biometrics **10**, 417—451 (1954).
3. Geary, R. C.: On the Correlation of B_2 and W, Section III of the Ratio of the Mean Deviation to the Standard Deviation as a Test of Normality. Biometrika **27**, 310—355 (1935).
4. Geary, R. C.: Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples. Biometrika **28**, 295—307 (1936).
5. Pearson, E. S.: A Further Development of Tests for Normality. Biometrika **22**, 239—249 (1930).
6. Siegel, S.: Nonparametric Statistics. P. 47, New York, 1956.

Authors' addresses: J. J. O'Brien, National Center for Atmospheric Research, Boulder, Colorado 80302, U. S. A. Prof. J. F. Griffiths, Department of Meteorology, Texas A & M University, College Station, Texas 77843, U. S. A.

Printed in Austria