

Guide to

Statistical interpretation of data —

**Part 7: Tests for departure from
normality**

ICS 03.120.30

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW

BSI
Documents

AVA

Document Sharing Hub

Committees responsible for this British Standard

The preparation of this British Standard was entrusted to Technical Committee SS/2, Statistical interpretation of data, upon which the following bodies were represented:

Clay Pipe Development Association Limited
Royal Statistical Society

This British Standard, having been prepared under the direction of the Management Systems Sector Board, was published under the authority of the Standards Board and comes into effect on 15 September 1997

© BSI 10-1999

The following BSI references relate to the work on this standard:
Committee reference SS/2
Draft for comment 95/402132 DC

ISBN 0 580 28343 7

Amendments issued since publication

Amd. No.	Date	Comments

Contents

	Page
Committees responsible	Inside front cover
National foreword	ii
Foreword	iii
Text of ISO 5479	1

National foreword

This Part of BS 2846 has been prepared by Technical Committee SS/2. It is identical with ISO 5479:1997 *Statistical interpretation of data — Tests for departure from the normal distribution*, published by the International Organization for Standardization (ISO). This Part of BS 2846 supersedes BS 2846-7:1984 which is withdrawn.

The importance of correct interpretation and presentation of test results has been increasingly recognized in the analysis of data obtained from manufacturing processes based on sample determinations and prototype evaluations in industry, commerce and educational institutions. It was for this reason that a series of guides on statistical interpretation of data was prepared.

At present this British Standard consists of the following Parts:

- *Part 1: Routine analysis of quantitative data;*
- *Part 2: Estimation of the mean: confidence interval;* (Identical to ISO 2602)
- *Part 3: Determination of a statistical tolerance interval;* (Identical to ISO 3207)
- *Part 4: Techniques of estimation and tests relating to means and variances;* (Identical to ISO 2854)
- *Part 5: Power of tests relating to means and variances;* (Identical to ISO 3494)
- *Part 6: Comparison of two means in the case of paired observations;* (Identical to ISO 3301)
- *Part 7: Tests for departure from normality;* (Identical to ISO 5479)
- *Part 8: Estimation of a median;*

Part 9 is in preparation and will deal with proportions as follows:

- *Section 1: Estimation of a proportion;*
- *Section 2: Comparison of a proportion with a given value;*
- *Section 3: Comparison of two proportions;*
- *Section 4: Fractiles of the F distribution (set of tables).*

This Part of BS 2846 is the national implementation of ISO 5479. However, the latter is itself an adaptation of BS 2846-7:1984, and the present version of this Part of BS 2846 is therefore effectively a direct revision of the version it replaces. Although much of the text has been rewritten, the present version differs fundamentally only to the extent that of the omnibus tests the d'Agostino test has been replaced by the Epps-Pulley test. Extensive comparative studies have shown that the latter test is generally superior to the former.

Cross-references

The British Standards that implement international publications referred to in this document may be found in the BSI Catalogue under the section entitled "International Standards Correspondence Index".

A British Standard does not purport to include all the necessary provisions of a contract. Users of British Standards are responsible for their correct application.

Compliance with a British Standard does not of itself confer immunity from legal obligations.

Summary of pages

This document comprises a front cover, an inside front cover, pages i and ii, the ISO title page, pages ii to iv, pages 1 to 31 and a back cover.

This standard has been updated (see copyright date) and may have had amendments incorporated. This will be indicated in the amendment table on the inside front cover.

INTERNATIONAL STANDARD

**ISO
5479**

First edition
1997-05-15

Statistical interpretation of data — Tests for departure from the normal distribution

*Interprétation statistique des données — Tests pour les écarts à la
distribution normale*



Contents

	Page
Foreword	iii
Introduction	1
1 Scope	1
2 Normative references	1
3 Definitions and symbols	2
4 General	3
5 Graphical method	4
6 Directional tests	10
7 Joint test using $\sqrt{b_1}$ and b_2 (multidirectional test)	14
8 Omnibus tests	14
9 Joint test using several independent samples	19
10 Statistical tables	22
Annex A (informative) Blank normal probability graph paper	30
Annex B (informative) Bibliography	31
Figure 1 — Annotated normal probability graph paper	5
Figure 2 — Graph of a series of observations on normal probability graph paper	6
Figure 3 — Density function with kurtosis in default	8
Figure 4 — Density function with kurtosis in excess	8
Figure 5 — Density function with positive skewness	9
Figure 6 — Density function with negative skewness	9
Figure 7 — Superposition of two different density functions	10
Figure 8 — Flow chart for the computation of the test statistic T_{EP} of the Epps-Pulley test	17
Figure 9 — Joint test using $ \sqrt{b_1} $ and b_2 (multidirectional test)	24
Table 1 — Results, $x_{(k)}$ of a series of 15 rotating-bend fatigue tests and corresponding values of $\lg(10 x_{(k)})$	7
Table 2 — Depth of sapwood	12
Table 3 — Series of 50 observations suspected of being affected by a variation in the dispersion of measurements	13
Table 4 — Annual amount of rainfall collected at a weather station	16
Table 5 — Breaking strength of rayon yarn	18
Table 6 — Breaking strength of rayon yarn — Calculation of the test statistic T_{EP}	19
Table 7 — Values of W_j and G_j for 22 samples of size $n = 20$ drawn from the same population	21
Table 8 — Test for skewness, $\sqrt{b_1}$ (p -quantiles of $ \sqrt{b_1} $ for $p = 1 - \alpha = 0,95$ and $0,99$)	22
Table 9 — Test for kurtosis, b_2 (p -quantiles of b_2 for $p = \alpha = 0,01$ and $0,05$ and $p = 1 - \alpha = 0,95$ and $0,99$)	23
Table 10 — Shapiro-Wilk test coefficients a_k for calculating the test statistic W	26
Table 11 — Shapiro-Wilk test: p -quantiles of the test statistic W for $p = \alpha = 0,01$ and $0,05$	28
Table 12 — Epps-Pulley test: p -quantiles of the test statistic T_{EP} for $p = 1 - \alpha = 0,90; 0,95; 0,975$ and $0,99$	28
Table 13 — Joint test using several independent samples: Coefficients for converting W to a standardized normal variate for $n = 8(1)50$	29
Table 14 — Quantities u_p of the standardized normal distribution	29

Descriptors: Statistical analysis, statistical quality control, statistical distribution, tests, estimation, use.

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard ISO 5479 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

Annex A and Annex B of this International Standard are for information only.

Introduction

Many of the statistical methods recommended in International Standards, such as those described in ISO 2854[1], are based on the assumption that the random variable(s) to which these methods apply are independently distributed according to a normal distribution with one or both of its parameters unknown. The following question therefore arises. Is the distribution that is represented by the sample sufficiently close to the normal distribution that the methods provided by these International Standards can be used reliably?

There is no simple yes or no answer to this question which is valid in all cases. For this reason a large number of “tests of normality” have been developed, each of which is more or less sensitive to a particular feature of the distribution under consideration; e.g. asymmetry or kurtosis.

Generally the test used is designed to correspond to a predetermined *a priori* risk that the hypothesis of normality is rejected even if it is true (error of the first kind). On the other hand, the probability that this hypothesis is not rejected when it is not true (error of the second kind) cannot be determined unless the alternative hypothesis (i.e. that which is opposed to the hypothesis of normality) can be precisely defined. This is not possible in general and, furthermore, it requires computational effort. For a distinct test, this risk is particularly large if the sample size is small.

1 Scope

1.1 This International Standard gives guidance on methods and tests for use in deciding whether or not the hypothesis of a normal distribution should be rejected, assuming that the observations are independent.

1.2 Whenever there are doubts as to whether the observations are normally distributed, the use of a test for departure from the normal distribution may be useful or even necessary. In the case of robust methods, however (i.e. where the results are only altered very slightly when the real probability distribution of the observations is not a normal distribution), a test for departure from the normal distribution is not very helpful. This is the case, for example, when the mean of a single random sample of observations is to be checked against a given theoretical value using a *t*-test.

1.3 It is not strictly necessary to use such a test whenever one refers to statistical methods based on the hypothesis of normality. It is possible that there is no doubt at all as to the normal distribution of the observations, whether theoretical (e.g. physical) reasons are present which confirm the hypothesis or because this hypothesis is deemed to be acceptable according to prior information.

1.4 The tests for departure from the normal distribution selected in this International Standard are primarily intended for complete data, not grouped data. They are unsuitable for censored data.

1.5 The tests for departure from the normal distribution selected in this International Standard may be applied either to observed values or to functions of them, such as the logarithm or the square root.

1.6 Tests for departure from the normal distribution are very ineffective for samples of size less than eight. Accordingly, this International Standard is restricted to samples of eight or more.

2 Normative reference

The following standard contains provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*.

3 Definitions and symbols

3.1

Definitions

For the purposes of this International Standard, the definitions given in ISO 3534-1 apply.

3.2

Symbols

a_k	coefficient of the Shapiro-Wilk test
A	auxiliary quantity for the Epps-Pulley test
b_2	empirical kurtosis
$\sqrt{b_1}$	empirical skewness
B	auxiliary quantity for the Epps-Pulley test
E	expectation
G_j	auxiliary quantity for the joint test using several independent samples
h	number of consecutive samples
H_0	null hypothesis
H_1	alternative hypothesis
k	within the sample, arranged in non-decreasing order, the number of the observed value x
m_j	central moment of order j of the sample
n	sample size
p	probability associated with the p -quantile of a distribution
P	probability
P_k	probability associated with $X_{(k)}$
S	auxiliary quantity for the Shapiro-Wilk test
T	test statistic
T_{EP}	test statistic of the Epps-Pulley test
u_p	p -quantile of the standardized normal distribution
v_j	auxiliary quantity for the joint test using several independent samples
W	test statistic of the Shapiro-Wilk test
W_j	auxiliary quantity for the joint test using several independent samples
x	value of X
X	random variable
$x_{(j)}$	j^{th} value in the sample, arranged in non-decreasing order
$x_{(k)}$	k^{th} value in the sample, arranged in non-decreasing order
$\bar{x}$	arithmetic average
α	significance level
β	probability of an error of the second kind
β_2	kurtosis of the population
$\beta_2 - 3$	excess of the population
$\sqrt{\beta_1}$	skewness of the population
γ	auxiliary quantity for the joint test using several independent samples
$\gamma_{(n)}$	coefficient of the joint test using several independent samples

δ	auxiliary quantity for the joint test using several independent samples
$\delta_{(n)}$	coefficient of the joint test using several independent samples
ε	auxiliary quantity for the joint test using several independent samples
$\varepsilon_{(n)}$	coefficient of the joint test using several independent samples
μ	expectation
μ_2	variance of the population
μ_3	central moment of the third order of the population
μ_4	central moment of the fourth order of the population
σ	standard deviation of the population ($=\sqrt{\mu_2}$)

4 General

4.1 There are several categories of tests for departure from normality. In this International Standard, graphical methods, moment tests, regression tests and characteristic function tests are considered. Chi-squared tests are appropriate for grouped data only but, because grouping results in a loss of information, they are not considered in this International Standard.

4.2 If no additional information about the sample is available, it is recommended first to do a normal probability plot; i.e. to plot the cumulative distribution function of the observed values on normal probability graph paper consisting of a system of coordinate axes where the cumulative distribution function of the normal distribution is represented by a straight line.

This method, which is described in clause 5, allows one to “see” immediately whether the distribution observed is close to the normal distribution or not. With this additional information it can be decided whether to carry out a directional test, or to carry out either a regression test or a characteristic function test, or no test at all. In addition, although such a graphical representation cannot be considered as a rigorous test, the summary information that it provides is an essential supplement to any test for departure from the normal distribution. In the case of rejection of the null hypothesis it is often possible to envisage by this means the type of alternative that might be applicable.

4.3 A test for departure from the normal distribution is a test of the null hypothesis that the sample consists of n independent observations coming from one and the same normal distribution. It consists of the calculation of a function T of the observations, which is called the test statistic. The null hypothesis of a normal distribution is then not rejected or rejected depending on whether or not the value of T lies within a set of values near to the expected value that corresponds to the normal distribution.

4.4 The **critical region** of the test is the set of values of T that leads to the rejection of the null hypothesis. The **significance level** of the test is the probability P of obtaining a value of T within the critical region when the null hypothesis is correct. This level gives the probability of erroneously rejecting the null hypothesis (error of the first kind).

The boundary of the critical region is (or, in the case of a two-sided test, the boundaries of the critical region are) the critical value(s) of the test statistic.

4.5 The **power** of the test is the probability of rejecting the null hypothesis when it is incorrect. A high power corresponds to a low probability of not rejecting the null hypothesis erroneously (error of the second kind).

It should be emphasized that the power of a test (i.e. for a given situation, the probability that the null hypothesis of a normal distribution will be rejected if it is wrong) increases as the number of observations increases. For example, a departure from the normal distribution which would become apparent when using a test for departure from the normal distribution on a large sample might not be detected by the same test if there were fewer observations.

4.6 A distinction is made between two categories of tests for departure from the normal distribution. When the form of departure from the normal distribution is specified in the alternative hypothesis, then the test is a **directional test**. However, when the form of departure from the normal distribution is not specified in the alternative hypothesis, the test is an **omnibus test**.

In a directional test, the critical region is determined in such a way that the power of the test reaches its maximum value. In an omnibus test, it is necessary to divide the critical region in such a way that the critical region consists of those values of the test statistic which lie far away from the expected value.

If assumptions are present about the type of departure from the normal distribution, i.e. when a distribution is envisaged whose asymmetry or kurtosis is different from that of the normal distribution, a directional test should be applied, because its power is greater than the power of an omnibus test.

4.7 Note that a directional test is essentially one-sided. In the case of asymmetry, for example, it centres either on positive asymmetry or on negative asymmetry. However, when several alternatives are considered jointly, the test is multidirectional. This is the case particularly when a non-null asymmetry and a kurtosis different from that of the normal distribution are considered together.

4.8 Table 8 to Table 14 and Figure 9 allow the tests to be performed for the most usual levels of α ; i.e. $\alpha = 0,05$ and $\alpha = 0,01$. The level of significance has to be stipulated before the test is performed. Note that a test may result in the rejection of the null hypothesis at the 0,05 level and the non-rejection of this same hypothesis at the 0,01 level.

4.9 During computation of test statistics, it is necessary to use at least six significant digits. Subtotals, intermediate results and auxiliary quantities shall not be rounded to less than six significant digits.

5 Graphical method

5.1 The cumulative distribution function of the observed values is plotted on normal probability graph paper. On this paper, one of the axes (in this International Standard it is the vertical axis) is non-linearly scaled according to the area under the standardized normal distribution function and is marked with the corresponding values of the cumulative relative frequency. The other axis is linearly scaled for the ordered values of X . The cumulative distribution function of the variable X then approximates to a straight line.

Sometimes these two axes are interchanged with each other. Furthermore, if a normalizing transformation of the variable X is made, the linear scale may be replaced by a logarithmic, quadratic, reciprocal or other scale.

Figure 1 gives an example of normal probability graph paper. On the vertical axis the values of the cumulative relative frequency are given as percentages, while the horizontal axis has an arbitrary linear scale.

A sheet of blank normal probability graph paper is provided in Annex A.

If a plot on this paper gives a set of points that appears to be scattered around a straight line, this provides crude support for the assumption that the sample can reasonably be regarded as having come from a normal distribution.

However, if there is a systematic departure from the straight line, the plot often suggests the type of distribution to be taken into consideration.

The importance of this approach is that it easily provides visual information on the type of departure from the normal distribution.

If the graph indicates that the data come from a shaped distribution (e.g. if the graph of the cumulative distribution function is as shown in Figure 5 or Figure 6), a transformation of the data might result in a normal distribution.

If the graph indicates that the data do not come from a simple homogeneous distribution, but rather from a mixture of two or more homogeneous subpopulations (e.g. if the graph of the cumulative distribution function is as shown in Figure 7), it is recommended that the subpopulations be identified and the analysis on each subpopulation be continued separately.

It should be kept in mind that such a plot is in no way a test for departure from the normal distribution in the strict sense. In the case of small samples, pronounced curves may occur for normal distributions, whilst for large samples slight curves may indicate non-normal distributions.

5.2 The graphical procedure consists of arranging the observed values $[x_{(1)}, x_{(2)}, \dots, x_{(n)}]$ in non-decreasing order, and then plotting

$$P_k = (k - 3/8)/(n + 1/4) \quad \dots (1)$$

against $x_{(k)}$ on normal probability graph paper.