

Confirmed

AUG 1993

British Standard Guide to

Statistical interpretation of data

Part 7. Tests for departure from normality

Guide pour l'interprétation statistique des données
Partie 7. Ecart de normalité — Tests

Leitfaden für die statistische Auswertung von Daten
Teil 7. Prüfung auf Normalitätsabweichungen

Contents

	Page		Page
Foreword		Inside front cover	
Committees responsible		Back cover	
Guide			
0. Introduction	1	2. Depth of sapwood; series arranged according to the non-decreasing values of 50 independent observations	4
1. Scope and field of application	1	3. Series of 35 independent observations suspected of being affected by a variation in the dispersion of measurements	5
2. General	1	4. Distribution of the independent results of 1000 resistance tests	6
3. Examination of the hypothesis of normality by means of a graph	2	5. Annual amount of rainfall collected at a weather station; ordered series of 44 values considered to be independent	7
4. Directional tests	4	6. Test for asymmetry, $\sqrt{b_1}$ fractiles of the order $P = 0.95$ and 0.99	8
4.1 General	4	7. Test for kurtosis, b_2 fractiles of the order $P = 0.01, 0.05, 0.95$ and 0.99	8
4.2 Directional test for asymmetry in $\sqrt{b_1}$: ($8 \leq n \leq 5000$)	4	8. Joint test using $\sqrt{b_1}$ and b_2 (multidirectional)	9
4.3 Directional test for kurtosis using b_2 : ($8 \leq n \leq 1000$)	5	9. Shapiro-Wilk test. Coefficients $a_k(W)$ necessary to calculate the statistic W	11
5. Joint test using $\sqrt{b_1}$ and b_2 (multidirectional): ($20 \leq n \leq 1000$)	5	10. Shapiro-Wilk test. Fractiles of the statistic W of the order $P = 0.01$ and 0.05	12
5.1 General	5	11. D'Agostino test. Fractiles of the statistic Y of the order $P = 0.005, 0.025, 0.975$ and 0.995	13
5.2 Joint test using $ \sqrt{b_1} $ and b_2	5		
6. Omnibus tests	6	Figures	
6.1 General	6	1. Normal probability graph paper	2
6.2 Shapiro-Wilk test: ($3 \leq n \leq 50$)	6	2. Graph of a series of observations on normal probability graph paper	3
6.3 D'Agostino test: ($50 \leq n \leq 1000$)	7		
Tables			
1. Results $x_{(k)}$ of a series of 15 independent wear tests on rotary equipment and corresponding values of $\log_{10} [10x_{(k)}]$	3		

Foreword

This British Standard, having been prepared under the direction of the Quality, Management and Statistics Standards Committee, is based on international discussions in which the UK has taken an active part.

Other Parts of this standard are as follows.

- Part 1 Routine analysis of quantitative data
- Part 2 Estimation of the mean-confidence interval (ISO 2602)
- Part 3 Determination of a statistical tolerance interval (ISO 3207)
- Part 4 Techniques of estimation and tests relating to means and variances (ISO 2854)
- Part 5 Power of tests relating to means and variances (ISO 3494)
- Part 6 Comparison of two means in the case of paired observations (ISO 3301)

A study of normality tests was made in order to propose a selection of tests for verification of the normality hypothesis, a hypothesis underlying the application of many statistical methods standardized within the framework of ISO/TC 69.

ISO/DIS 5479 was drawn up by Subcommittee ISO/TC 69/SC 2, Interpretation of statistical results, and submitted in 1978 to the members of the Subcommittee for postal ballot. This document was approved by a wide majority. The present standard is issued as a national standard based on a revised version of ISO/DIS 5479, incorporating editorial changes arising from comments received in the postal ballot.

Tables 6 to 11 inclusive have been reproduced by permission of the Biometrika Trustees.

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

British Standard Guide to

Statistical interpretation of data

Part 7. Tests for departure from normality

0. Introduction

Many of the statistical methods recommended in standards drawn up by the British Standards Institution, such as those described in BS 2846 and BS 6002, are based on the assumption that the variable(s) to which these methods apply are independently distributed according to the *normal distribution*.

1. Scope and field of application

1.1 This British Standard provides a selection of tests, any one of which may be used to decide whether or not the hypothesis of normality is acceptable, assuming that the observations are independent.

1.2 These tests are not intended to be used on every sample of data before it is analysed by methods based on the normal distribution; indeed this routine use of tests for normality, when the hypothesis is true, will falsify the exact consequences of this hypothesis since it will no longer refer to all possible samples but only those that pass the tests. In some cases, it can be that this hypothesis should not be challenged, either because there are theoretical reasons, e.g. physical ones, which justify it or because it has been found to be acceptable on the basis of previously accumulated information. Furthermore, in the case of *robust* methods, i.e. where the results are only altered very slightly when the *real* probability distribution of the observations is not normal, a normality test is not very useful. 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 Student test. However, where the method is not robust and there are doubts as to the normality of the distribution of the observations, the use of a normality test may be useful or even necessary.

1.3 Normality tests are also a means of checking the hypothesis of normality for other statistical purposes; for example, checking the capability where a generating process is intended to produce elements that are or appear to be randomly distributed in accordance with the normal distribution.

1.4 Finally, it should be emphasized that the *power* of these tests (i.e. for a given situation, the probability that the normality hypothesis will be rejected when it is wrong) increases as the number of observations increases. For example, a deviation from the normal distribution which would become apparent when using a normality test on a large sample might not be detected by the same test if there were fewer observations. Consequently, the relative nature of the conclusions of these tests should always be borne in mind, although it should be remembered

that these conclusions may be challenged by any other information subsequently obtained.

1.5 When carrying out a normality test, it should always be ensured that the independence of the observations cannot be questioned. It is in fact only under these conditions that a conclusion can be drawn to reject the hypothesis of normality when there is a significant test result.

NOTE. The title of the publication referred to in this standard is given on the inside back cover.

2. General

2.1 The simplest way of testing the normality of the distribution of a series of observations is by plotting a graph from these observations in a system of co-ordinate axes where the distribution function of the normal distribution is represented by a straight line.

This method, which is described in clause 3, means that it is possible to 'see' immediately whether the distribution observed is close to the normal distribution or not.

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 normality tests, by means of which it is often possible, in the case of rejection of the null hypothesis, to show the type of alternative that might be available.

2.2 A normality test is a special sort of hypothesis test. It consists of the calculation of the function T of the observations, which is called the *statistic of the test*. The null hypothesis of a normal distribution is then accepted or rejected depending on whether or not the value of T lies within a set of values near to the ideal value that corresponds to the normal distribution.

2.3 The set of values of T that leads to the rejection of the null hypothesis is the *critical region* of the test. When this hypothesis is true, the probability α of obtaining a value for T within this critical region is the *significance level* of the test. Thus 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 statistic of the test.

2.4 The *power* of the test is the probability of rejecting the null hypothesis when it is incorrect. A high power guarantees a low probability of accepting the null hypothesis erroneously (*error of the second kind*).

The power of a test varies according to which alternative hypothesis is true when the null hypothesis is false.

2.5 With normality tests, a distinction is made between two categories of tests. When the alternative hypothesis is specified, i.e. when a distribution is envisaged whose asymmetry or kurtosis is different from that of the normal distribution, the test is a directional test. However, when the form of deviation 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 chosen so that the power of the test is a maximum. In an omnibus test, the limitation is to divide the critical region in such a way that the probability of an error of the first kind is equally distributed between values of the statistic that are the most distant from the ideal value.

2.6 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.

2.7 The statistical tables 6 to 11 allow the tests described for the most usual levels to be performed; i.e. $\alpha = 0.05$ and $\alpha = 0.01$. Note that a test may result in the rejection of the null hypothesis at the 0.05 level and in the acceptance of this same hypothesis at the 0.01 level. A comparison of these two conclusions may sometimes be useful to stress the strength or weakness of the conclusion and therefore justify any need for further information.

3. Examination of the hypothesis of normality by means of a graph

3.1 If u_P is the fractile of order P of the standard normal variable U , then U_P is a function of P , $\xi(P)$, for example, and

$$P = P(U < U_P) = \phi(U_P) \\ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi(P)} \exp(-t^2/2) dt \quad (1)$$

The graph is plotted using 'normal probability graph paper'. On this paper, one of the axes (usually the horizontal) is linearly scaled in u_P and marked with the corresponding values of P . The other axis is linearly scaled for the values of x . The distribution of the variable x is then represented by the straight line of equation (2) in 3.4.

Figure 1 gives an example of normal probability graph paper. On the horizontal axis the values of $P = \phi(u)$ are given as percentages, while the vertical axis has an arbitrary linear scale.

3.2 The graphical test consists of arranging the observations $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ in non-decreasing order, and then plotting $x_{(k)}$ against

$$P_k = 100 k / (n + 1)$$

on normal probability paper.

If the points thus obtained form a more or less straight line, a normal distribution may be assumed.

However, if there is a systematic departure from a straight line, the plot often suggests the type of alternative hypothesis to be taken into consideration. In particular, if the curve of the observations is such that the large values tend to be well above the straight line defined by the other

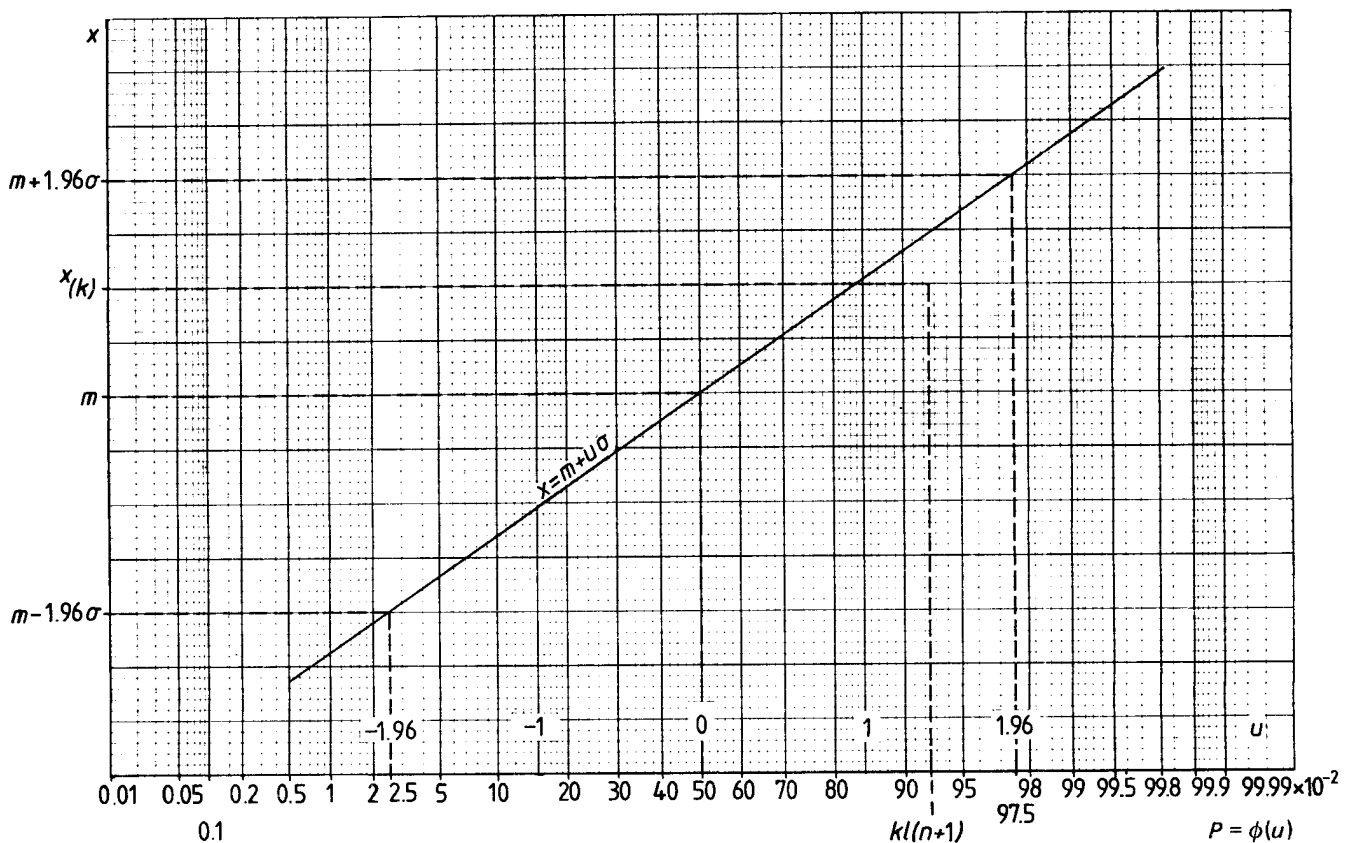


Figure 1. Normal probability graph paper

values, a transformation such as $y = \log x$ or $y = \sqrt{x}$ will generally lead to a plot that conforms more to a normal distribution.

It should be kept in mind that such a plot is in no way a normality test 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.

3.3 An example of use of normal probability graph paper is as follows.

Table 1 shows the ordinate values $x_{(k)}$ in non-decreasing order of the result of a series of 15 independent wear tests on rotary equipment.

By associating the k th value $x_{(k)}$ with the probability $P = k/16$, the series of white points shown in figure 2 is obtained. It is immediately seen from this graph that these points do not form a straight line.

However, if $\log_{10} [10 x_{(k)}]$ is substituted for $x_{(k)}$, the new graph leads to the series of black points which this time lie acceptably close to a straight line.

The hypothesis of a normal distribution of the logarithm of the observations therefore seems adequate.

Table 1. Results $x_{(k)}$ of a series of 15 independent wear tests on rotary equipment and corresponding values of $\log_{10} [10 x_{(k)}]$

k	$x_{(k)}$	$\log_{10} [10 x_{(k)}]$
1	0.200	0.301
2	0.330	0.519
3	0.445	0.648
4	0.490	0.690
5	0.780	0.892
6	0.920	0.964
7	0.950	0.978
8	0.970	0.987
9	1.040	1.017
10	1.710	1.233
11	2.220	1.346
12	2.275	1.357
13	3.650	1.562
14	7.000	1.845
15	8.800	1.944

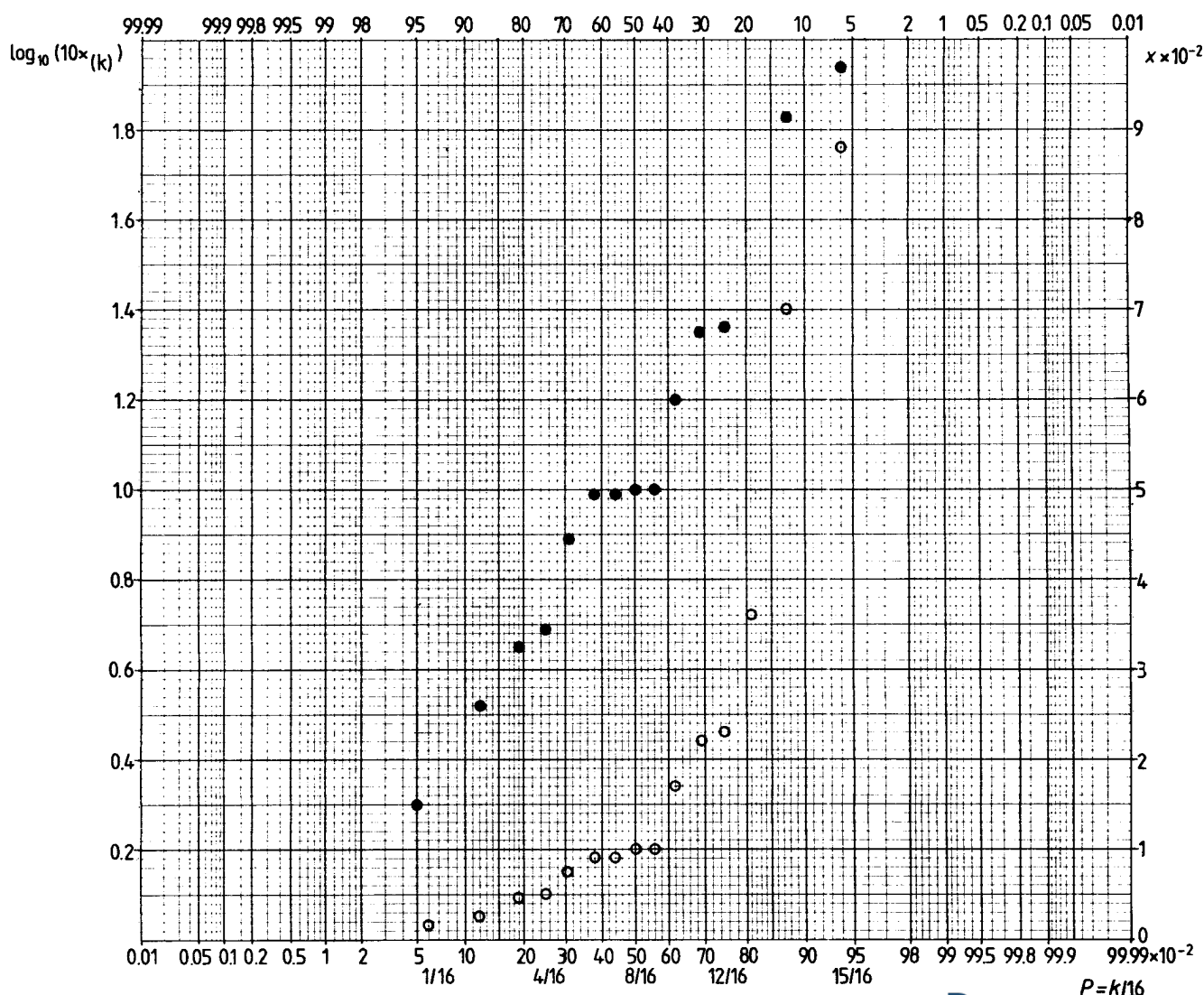


Figure 2. Graph of a series of observations on normal probability graph paper