INTERNATIONAL STANDARD

ISO 16269-8

First edition
2004-09-15

Statistical interpretation of data —

Part 8: Determination of prediction intervals

Interprétation statistique des données —

Partie 8: Détermination des intervalles de prédiction

Reference number
ISO 16269-8:2004(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. 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.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 16269-8 was prepared by Technical Committee ISO/TC 69, *Application of statistical methods*.

ISO 16269 consists of the following parts, under the general title *Statistical interpretation of data*:

— *Part 6: Determination of statistical tolerance intervals*

— *Part 7: Median — Estimation and confidence intervals*

— *Part 8: Determination of prediction intervals*

Introduction

Prediction intervals are of value wherever it is desired or required to predict the results of a future sample of a given number of discrete items from the results of an earlier sample of items produced under identical conditions. They are of particular use to engineers who need to be able to set limits on the performance of a relatively small number of manufactured items. This is of increasing importance with the recent shift towards small-scale production in some industries.

Despite the first review article on prediction intervals and their applications being published as long ago as 1973, there is still a surprising lack of awareness of their value, perhaps due in part to the inaccessibility of the research work for the potential user, and also partly due to confusion with confidence intervals and statistical tolerance intervals. The purpose of this part of ISO 16269 is therefore twofold:

— to clarify the differences between prediction intervals, confidence intervals and statistical tolerance intervals;

— to provide procedures for some of the more useful types of prediction interval, supported by extensive, newly-computed tables.

For information on prediction intervals that are outside the scope of this part of ISO 16269, the reader is referred to the Bibliography.

Statistical interpretation of data —

Part 8: Determination of prediction intervals

1 Scope

This part of ISO 16269 specifies methods of determining prediction intervals for a single continuously distributed variable. These are ranges of values of the variable, derived from a random sample of size n, for which a prediction relating to a further randomly selected sample of size m from the same population may be made with a specified confidence.

Three different types of population are considered, namely:

a) normally distributed with unknown standard deviation;

b) normally distributed with known standard deviation;

c) continuous but of unknown form.

For each of these three types of population, two methods are presented, one for one-sided prediction intervals and one for symmetric two-sided prediction intervals. In all cases, there is a choice from among six confidence levels.

The methods presented for cases a) and b) may also be used for non-normally distributed populations that can be transformed to normality.

For cases a) and b) the tables presented in this part of ISO 16269 are restricted to prediction intervals containing *all* the further m sampled values of the variable. For case c) the tables relate to prediction intervals that contain at least $m - r$ of the next m values, where r takes values from 0 to 10 or 0 to $m - 1$, whichever range is smaller.

For normally distributed populations a procedure is also provided for calculating prediction intervals for the mean of m further observations.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Statistical quality control*

3 Terms, definitions and symbols

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1 and ISO 3534-2 and the following apply.

3.1.1
prediction interval
interval determined from a random sample from a population in such a way that one may have a specified level of confidence that no fewer than a given number of values in a further random sample of a given size from the same population will fall

NOTE In this context, the confidence level is the long-run proportion of intervals constructed in this manner that will have this property.

3.1.2
order statistics
sample values identified by their position after ranking in non-decreasing order of magnitude

NOTE The sample values in order of selection are denoted in this part of ISO 16269 by $x_1, x_2, \ldots, x_n$. After arranging in non-decreasing order, they are denoted by $x_{[1]}, x_{[2]}, \ldots, x_{[n]}$, where $x_{[1]} \leqslant x_{[2]} \leqslant \ldots \leqslant x_{[n]}$. The word "non-decreasing" is used in preference to "increasing" to include the case where two or more values are equal, at least to within measurement error. Sample values that are equal to one another are assigned distinct, contiguous integer subscripts in square brackets when represented as order statistics.

3.2 Symbols

a lower limit to the values of the variable in the population

α nominal maximum probability that more than r observations from the further random sample of size m will lie outside the prediction interval

b upper limit to the values of the variable in the population

C confidence level expressed as a percentage: $C = 100\,(1 - \alpha)$

k prediction interval factor

m size of further random sample to which the prediction applies

n size of random sample from which the prediction interval is derived

s sample standard deviation: $s = \sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \Big/ (n-1)}$

r specified maximum number of observations from the further random sample of size m that will not lie in the prediction interval

T_1 lower prediction limit

T_2 upper prediction limit

x_i ith observation in a random sample

$x_{[i]}$ ith order statistic

$\overline{x}$ sample mean: $\overline{x} = \sum_{i=1}^{n} x_i / n$

4 Prediction intervals

4.1 General

A *two-sided prediction interval* is an interval of the form (T_1, T_2), where $T_1 < T_2$; T_1 and T_2 are derived from a random sample of size n and are called the *lower* and *upper prediction limits*, respectively.

If a and b are respectively the lower and upper limits of the variable in the population, a *one-sided prediction interval* will be of the form (T_1, b) or (a, T_2).

NOTE 1 For practical purposes a is often taken to be zero for variables that cannot be negative, and b is often taken to be infinity for variables with no natural upper limit.

NOTE 2 Sometimes a population is treated as normal for the purpose of determining a prediction interval, even when it has a finite limit. This may seem incongruous, as the normal distribution ranges from minus infinity to plus infinity. However, in practice, many populations with a finite limit are closely approximated by a normal distribution.

The practical meaning of a prediction interval relating to individual sample values is that the experimenter claims that a further random sample of m values from the same population will have at most r values not lying in the interval, while admitting a small nominal probability that this assertion may be wrong. The nominal probability that an interval constructed in such a way satisfies the claim is called the confidence level.

The practical meaning of a prediction interval relating to a sample mean is that the experimenter claims that the *mean* of a further random sample of m values from the same population will lie in the interval, while admitting a small nominal probability that this assertion may be wrong. Again, the nominal probability that an interval constructed in such a way satisfies the claim is called the confidence level.

This part of ISO 16269 presents procedures applicable to a normally distributed population for $r = 0$ and procedures applicable to the mean of a further sample from a normally distributed population. It also provides procedures applicable to populations of unknown distributional form for $r = 0, 1, \ldots, 10$ or 0 to $m - 1$, whichever range is smaller. In all cases, the tables present prediction interval factors or sample sizes that provide *at least* the stated level of confidence. In general, the actual confidence level is marginally greater than the stated level.

The limits of the prediction intervals for normally distributed populations are at a distance of k times the sample standard deviation (or, where known, the population standard deviation) from the sample mean, where k is the prediction interval factor. In the case of unknown population standard deviation, the value of k becomes very large for small values of n in combination with large values of m and high levels of confidence. Use of large values of k, for example in excess of 10 or 15, should be avoided whenever possible, as the resulting prediction intervals are likely to be too wide to be of any practical use, other than to indicate that the initial sample was too small to yield any useful information about future values. Moreover, for large values of k the integrity of the resulting prediction intervals could be badly compromised by even small departures from normality. Values of k up to 250 are included in the tables primarily to show how rapidly k decreases as the initial sample size n increases.

For prediction intervals relating to the individual values in a further sample, Form A may be used to organize the calculations for a normally distributed population and Form C when the population is of unknown distributional form. Form B is provided to assist with the calculation of a prediction interval for the mean of a further sample from a normally distributed population.

Annexes A to D provide tables of prediction interval factors. Annexes E and F provide tables of sample sizes required when the population is of unknown distributional form. Annex G gives the procedure for interpolating in the tables when the required combination of n, m and confidence level is not tabulated. Annex H presents the statistical theory underlying the tables.

4.2 Comparison with other types of statistical interval

4.2.1 Choice of type of interval

In practice, it is often the case that predictions are required for a *finite* number of observations based on the results of an initial random sample. These are the circumstances under which this part of ISO 16269 is appropriate. There is sometimes confusion with other types of statistical interval. Subclauses 4.2.2 and 4.2.3 are presented in order to clarify the distinctions.

4.2.2 Comparison with a statistical tolerance interval

A prediction interval for individual sample values is an interval, derived from a random sample from a population, about which a confidence statement may be made concerning the maximum *number* of values in a further random *sample* from the population that will lie outside the interval. A statistical tolerance interval (such as that defined in ISO 16269-6) is also an interval derived from a random sample from a population for which a confidence statement may be made; however, the statement in this case relates to the maximum *proportion* of values in the *population* lying outside the interval (or, equivalently, to the *minimum* proportion of values in the population lying *inside* the interval).

NOTE 1 A statistical tolerance interval constant is the limit of a prediction interval constant as the future sample size, m, tends to infinity while the number, r, of items in the future sample falling outside the interval remains a constant fraction of m, provided $r > 0$. This is illustrated in Table 1 for a 95 % confidence level for one-sided and two-sided intervals when r/m = 0,1.

However, there is no such analogy between statistical tolerance interval constants and prediction interval constants for r = 0, the case on which this part of ISO 16269 is primarily focussed.

Table 1 — Example of prediction interval constants

r	1	2	5	10	20	50	100	1 000	Statistical tolerance interval constants for a minimum proportion of 0,9 of the population covered
m	10	20	50	100	200	500	1 000	10 000	
	Prediction interval constants								
One-sided intervals	1,887	1,846	1,767	1,718	1,686	1,663	1,655	1,647	1,646
Two-sided intervals	2,208	2,172	2,103	2,061	2,034	2,014	2,007	2,000	2,000

NOTE 2 The case r = 0 is particularly important in applications related to safety.

4.2.3 Comparison with a confidence interval for the mean

A prediction interval for a mean is an interval, derived from a random sample from a population, for which it may be asserted with a given level of confidence that the mean of a further random *sample* of specified size will lie. A confidence interval for a mean (such as that defined in ISO 2602) is also an interval derived from a random sample from a population for which a confidence statement may be made; however, the statement in this case relates to the mean of the *population*.

5 Prediction intervals for all observations in a further sample from a normally distributed population with unknown population standard deviation

5.1 One-sided intervals

A one-sided prediction interval relating to a normally distributed population with unknown population standard deviation is of the form $(\bar{x} - ks, b)$ or $(a, \bar{x} + ks)$ where the values of the sample mean $\bar{x}$ and the sample standard deviation s are determined from a random sample of size n from the population. The prediction interval factor k depends on n, on the further sample size m and on the confidence level C; values of k are presented in Annex A.

EXAMPLE The pressures in gun barrels caused by firing artillery shells of a given type are known from past experience to be closely approximated by a normal distribution. A sample of 20 rounds has a mean pressure of 562,3 MPa and a standard deviation of pressure of 8,65 MPa. A batch of 5 000 further rounds in total is to be produced under identical manufacturing conditions. What barrel pressure can one be 95 % confident will not be exceeded by any of the 5 000 shells fired under identical conditions ?

Table A.2 provides prediction interval factors at the 95 % confidence level. From Table A.2 it is found that the appropriate prediction interval factor is k = 5,251. The upper limit to a one-sided prediction interval at 95 % confidence is therefore

$$\bar{x} + ks = 562{,}3 + 5{,}251 \times 8{,}65 = 607{,}7 \text{ MPa}$$

Hence one may be 95 % confident that none of the batch of 5 000 rounds will produce a barrel pressure in excess of 607,7 MPa.

This example is also used to illustrate the use of Form A.

5.2 Symmetric two-sided intervals

A symmetric two-sided prediction interval for a normally distributed population with unknown population standard deviation is of the form $(\bar{x} - ks, \bar{x} + ks)$. The prediction interval factor k depends on n, on the further sample size m and on the confidence level C; values of k are presented in Annex B.

EXAMPLE The time to detonation of a particular type of hand grenade after the pin has been removed is known to have an approximate normal distribution. A random sample of size 30 is drawn and tested, and the times to detonation are recorded. The sample mean time is 5,140 s and the sample standard deviation is 0,241 s. A symmetric two-sided prediction interval is required for all of the next lot of 10 000 grenades at 99 % confidence.

Table B.4 provides prediction interval factors at the 99 % confidence level. Entering Table B.4 with n = 30 and m = 10 000 yields the value k = 6,059. The symmetric prediction interval is

$$(\bar{x} - ks,\ \bar{x} + ks) = (5{,}140 - 6{,}059 \times 0{,}241,\ 5{,}140 + 6{,}059 \times 0{,}241) = (3{,}68,\ 6{,}60)$$

One may therefore be 99 % confident that none of the next lot of 10 000 grenades will have a time to detonation outside the range 3,68 s to 6,60 s.

5.3 Prediction intervals for non-normally distributed populations that can be transformed to normality

For non-normally distributed populations that can be transformed to normality, first the procedures for normally distributed populations are applied to the transformed data; the prediction interval is then found by applying the inverse transformation to the resulting prediction limits.

EXAMPLE Suppose that for the data of the example in 5.2 it is known instead that times to detonation are approximately log-normally distributed, i.e. the logarithm of the time to detonation is approximately normally distributed. The sample times $x_1, x_2, \ldots, x_n$ are accordingly transformed to normality by taking their natural logarithms, namely $y_i = \ln x_i$ for i = 1, 2, …, 30. Suppose that the sample mean of the transformed data is $\bar{y}$ = 1,60 and the sample standard deviation is s_y = 0,05. The prediction interval factor for 99 % confidence that none of the next 10 000 times falls outside a two-sided interval is, of course, unchanged at k = 6,059. The symmetric prediction interval for the transformed data is

$$(\bar{y} - ks_y,\ \bar{y} + ks_y) = (1{,}60 - 6{,}059 \times 0{,}05,\ 1{,}60 + 6{,}059 \times 0{,}05) = (1{,}297,\ 1{,}903)$$

The units of measurement of y are log-seconds. The inverse transformation to convert the units back to seconds is exponentiation. The prediction interval at 99 % confidence for the time to detonation of all of the next ten thousand grenades is therefore

$$\left(\mathrm{e}^{1{,}297},\ \mathrm{e}^{1{,}903}\right) = (3{,}66,\ 6{,}71) \text{ s}$$

NOTE 1 The same result would have been obtained using logarithms to any other base, provided that the antilogarithm to the same base is used when converting back to the original units.

NOTE 2 When a two-sided prediction interval is determined in accordance with 5.2 or 6.2, its limits for normally distributed populations are symmetric about (i.e. equidistant from) the estimated median of the population. This symmetry is lost for non-normally distributed populations that are transformed to normality in accordance with 5.3 or 6.3.

5.4 Determination of a suitable initial sample size, n, for a given maximum value of the prediction interval factor, k

Sometimes the confidence level, future sample size m and approximate desired value of the prediction interval factor are given and it is required to determine the initial sample size n. Locate the table for the given confidence level and the sidedness of the prediction interval (i.e. one of the tables in Annex A for a one-sided interval or one of the tables in Annex B for two-sided intervals) and find the column for the given value of m. Look down this column until the first value of k no greater than the given maximum is found. The value of n in the leftmost column of this row of the table gives the required initial sample size.

NOTE If the entry at the bottom of this column exceeds the maximum acceptable value of k then there is no initial sample size large enough to satisfy the requirement. A reduction in the confidence level should be considered.

EXAMPLE Consider a situation in acceptance sampling in which, prior to the use of this part of ISO 16269, it has been the practice to accept lots of size 5 000 whenever $\bar{x} + 4{,}75\, s \leqslant 0{,}1$, where x is the normally distributed porosity of a sintered component and $\bar{x}$ and s are the sample mean and sample standard deviation based on a random sample of size 30 from a normally distributed population. Suppose that it has been decided to replace this acceptance criterion with one that will provide 95 % confidence that none of the items in the lot has $x > 0{,}1$. The producer says that he will be satisfied with the acceptance criterion provided the prediction interval factor is no larger than the 4,75 that he is used to, subject to the sample size requirement not being excessive.

Look down the column for m equal to 5 000 on the third page of Table A.2. It is found that k = 4,771 for sample size 40, but falls below 4,75 to k = 4,717 for a sample of size 45. The producer agrees to increase the sample size to 45 with k = 4,717 for future lots.

5.5 Determination of the confidence level corresponding to a given prediction interval

Rather than determining the prediction interval corresponding to a given confidence level, it may sometimes be required to determine, from the initial sample, the confidence level corresponding to a specified interval. This may be a one-sided interval $(\bar{x} - ks, b)$ or $(a, \bar{x} + ks)$, or a two-sided interval $(\bar{x} - ks, \bar{x} + ks)$ that is symmetrical about the sample mean.

First calculate the value of k corresponding to the desired prediction interval. The confidence level for this interval can then be found by interpolation between tabulated values, as specified in G.1.4.

6 Prediction intervals for all observations in a further sample from a normally distributed population with known population standard deviation

6.1 One-sided intervals

A one-sided prediction interval for a normally distributed population with known population standard deviation σ is of the form $(\bar{x} - k\sigma, b)$ or $(a, \bar{x} + k\sigma)$. The prediction interval factor k depends on n, on the further sample size m and on the confidence level C; values of k are presented in Annex C.

EXAMPLE 150 mm diameter vitrified clay pipes produced by a given process are known to have lengths that are normally distributed with a standard deviation of 4,49 mm. A sample of 50 pipes is found to have a mean of 1 760,60 mm. What length can one be 99 % confident that all of the next 1 000 pipes will exceed ?

Entering Table C.4 with n = 50 and m = 1 000, the appropriate prediction interval factor is found to be k = 4,306. The lower prediction limit for all of the next 1 000 lengths is therefore

$$\bar{x} - k\sigma = 1760{,}60 - 4{,}306 \times 4{,}49 = 1741$$

Hence one may be 99 % confident that none of the lengths of the next 1 000 pipes will be less than 1 741 mm.

This kind of information could be useful if the manufacturer were thinking of providing a warranty for his production. For example, here the manufacturer would be on fairly safe ground in guaranteeing a length of at least 1 740 mm.

6.2 Symmetric two-sided intervals

A symmetric two-sided prediction interval for a normally distributed population with known population standard deviation σ is of the form $(\bar{x} - k\sigma, \bar{x} + k\sigma)$. The prediction interval factor k depends on n, on the further sample size m and on the confidence level C; values of k are presented in Annex D.

EXAMPLE Suppose that for the data of the example in 6.1 it is required to calculate a two-sided prediction interval for all of the next 10 000 pipe lengths at 95 % confidence. The appropriate table for a confidence level of 95 % is Table D.2. Entering Table D.2 with n = 50 and m = 10 000, it is found that the two-sided prediction interval factor is 4,605. The prediction interval is

$$(\bar{x} - k\sigma,\ \bar{x} + k\sigma) = (1\,760{,}60 - 4{,}605 \times 4{,}49,\ 1\,760{,}60 + 4{,}605 \times 4{,}49) = (1\,739{,}9,\ 1\,781{,}3)$$

One may therefore be 95 % confident that all of the further 10 000 pipes have lengths that lie between 1 739,9 mm and 1 781,3 mm.

6.3 Prediction intervals for non-normally distributed populations that can be transformed to normality

For non-normally distributed populations that can be transformed to normality, the procedures for determining a prediction interval for known population standard deviation are similar to those for unknown population standard deviation, described in 5.3. First the procedures for normally distributed populations are applied to the transformed data; then the prediction interval is found by applying the inverse transformation to the resulting prediction limits.

EXAMPLE The fatigue life of a structural component for an aircraft is known to have an approximately lognormal distribution, i.e. the logarithm of the failure time can be assumed to be normally distributed. The standard deviation of $\log_{10}$(life) is also known from previous experience to be approximately equal to 0,11. Six samples of the component are subjected to fatigue testing and the number of loading cycles to failure recorded as follows:

— 229 200;

— 277 900;

— 332 400;

— 369 700;

— 380 800;

— 406 300.

Two further components are to be manufactured to the same specification. How many loading cycles can be applied whilst having 99,9 % confidence that none of these two components will fail?

Taking logarithms to base 10 and then averaging, the mean of $x = \log_{10}$(life) is found to be $\bar{x}$ = 5,513 86. Entering Table C.6 with n = 6 and m = 2, the appropriate one-sided prediction interval factor is found to be k = 3,554. The lower prediction limit for all of the next two components is therefore

$$\bar{x} - k\sigma = 5{,}513\,86 - 3{,}554 \times 0{,}11 = 5{,}122\,92$$

and, taking antilogarithms, $10^{5{,}122\,92}$ = 132 715.

Hence, taking into account that the standard deviation of $\log_{10}$(life) is only known to two significant figures, one may be 99,9 % confident that the further two components will survive for at least 130 000 loading cycles.

6.4 Determination of a suitable initial sample size, n, for a given value of k

The procedure is the same as described in 5.4 except that Annex C or D is used instead of Annex A or B.

6.5 Determination of the confidence level corresponding to a given prediction interval

The confidence level corresponding to a one-sided interval $(\bar{x} - k\sigma, b)$ or $(a, \bar{x} + k\sigma)$, or a two-sided interval $(\bar{x} - k\sigma, \bar{x} + k\sigma)$ that is symmetrical about the sample mean, may be calculated from Annexes C and D.

First calculate the value of k corresponding to the desired prediction interval. The confidence level for this interval can then be found by interpolation between tabulated values, as described in G.1.4.

7 Prediction intervals for the mean of a further sample from a normally distributed population

A simple two-stage process may be used to obtain the prediction interval factor for the mean of a further sample of m observations from the same normally distributed population, using the same tables. First find the prediction interval factor corresponding to a single future observation. Then multiply this prediction interval factor by $\sqrt{(n+m)/[m(n+1)]}$. This procedure applies to one-sided and two-sided intervals and to the cases of both known and unknown population standard deviation.

EXAMPLE Suppose that for the data of the example in 6.1 it is required to provide a lower prediction limit at 99 % confidence on the mean of the lengths of the next 1 000 pipes. From Table C.4 it is found that the prediction interval factor for an initial sample size of 50 and a single future observation is 2,350. The required prediction interval factor is therefore

$$k = 2{,}350 \times \sqrt{(n+m)/[m(n+1)]} = 2{,}350 \times \sqrt{1\,050/51\,000} = 0{,}337\,2$$

It follows that the lower prediction limit for the mean length of the next 1 000 pipes is

$$\bar{x} - k\sigma = 1\,760{,}60 - 0{,}337\,2 \times 4{,}49 = 1\,759 \text{ mm}$$

This example is used to illustrate the use of Form B.

8 Distribution-free prediction intervals

8.1 General

When the variable is continuous but the distributional form of the population is unknown, distribution-free methods should be used to produce prediction intervals. These are based on the order statistics $x_{[1]}, x_{[2]}, \ldots, x_{[n]}$. In general, one-sided distribution-free prediction intervals are of the form $(x_{[i]}, b)$ or $(a, x_{[i]})$ where $1 \leqslant i \leqslant n$, while two-sided distribution-free prediction intervals are of the form $(x_{[i]}, x_{[j]})$ where $1 \leqslant i \leqslant j \leqslant n$. This part of ISO 16269 provides procedures for one-sided intervals of the form $(x_{[1]}, b)$ or $(a, x_{[n]})$ and two-sided intervals of the form $(x_{[1]}, x_{[n]})$.

The problem with such intervals is in determining how large the initial sample size needs to be in order that one may have the required confidence that the prediction interval will contain at least $m - r$ values from the next m. Annexes E and F are provided for this purpose.

8.2 One-sided intervals

Tables E.1 to E.6 provide initial sample sizes n from which one may have confidence C that the one-sided distribution-free prediction interval $(x_{[1]}, b)$ [or alternatively $(a, x_{[n]})$] will include at least $m - r$ of a further sample of m values from the same population, for a range of values of C, m and r.

EXAMPLE A distribution-free lower prediction limit is required for the strength in bending of vitrified clay pipes, such that one may be 90 % confident that no more than 10 pipes in each further batch of 200 will have a lower strength. What initial sample size is required?

Table E.1 provides the initial sample sizes for a confidence level of 90 %. Entering this table with m = 200 and r = 10 it is found that the appropriate sample size is n = 46. A random sample of 46 pipes is drawn and tested for strength in bending. The lowest strength is found to be 6,4 kN·m. Thus one may be 90 % confident that, for pipes manufactured under identical conditions to the initial sample, no more than 10 pipes in each batch of 200 will have strength in bending below 6,4 kN·m.