INTERNATIONAL STANDARD

ISO 8595

First edition
1989-12-15

Interpretation of statistical data — Estimation of a median

Interprétation des données statistiques — Estimation d'une médiane

Reference number
ISO 8595 : 1989 (E)

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 approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 8595 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods.*

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland
Printed in Switzerland

Interpretation of statistical data — Estimation of a median

1 Scope

This International Standard describes procedures for establishing the point estimate and the confidence intervals of a median of a probability distribution of a population, based on a random sample of size n from the population. These procedures give a method of distribution-free estimation and can be applied also to estimate quartiles and/or percentiles.

NOTE — The median is the second quartile. Similar procedures can be used for other percentiles but these are not described in this International Standard.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were 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 editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 2602 : 1980, *Statistical interpretation of test results — Estimation of the mean — Confidence interval.*

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

ISO 3534-2 : —[1], *Statistics — Vocabulary and symbols — Part 2: Statistical quality control.*

ISO 3534-3 : 1985, *Statistics — Vocabulary and symbols — Part 3: Design of experiments.*

3 Definitions

For the purposes of this International Standard, the definitions given in ISO 3534-1, ISO 3534-2 and ISO 3534-3, together with the following, apply.

3.1 kth order statistic of a sample of size n: The value $x_{(k)}$ of the kth element in a sample when the elements are arranged in the non-decreasing order of their values, i.e. when they are numbered such that

$$x_{(1)} \leqslant x_{(2)} \leqslant \dots \leqslant x_{(n)}$$

3.2 median of a continuous probability distribution, M [or $x_{(0,5)}$]: A quantity M such that

$$F(M) = 1/2$$

where $F(x)$ is a distribution function. In this International Standard, it is called the population median.

4 Conditions of application

The method described in this International Standard is valid for any continuous population, provided that the sample is drawn at random.

NOTE — If the distribution of the population can be assumed to be approximately normal, the median is equal to the mean and the confidence limits should be calculated in accordance with ISO 2602.

Under these conditions, the number k of sample elements that have value at most (at least) equal to the population percentile is the realization of a random variable K which can be described by a binomial distribution:

$$P\,(K=k\,;\,n,\,p) = \binom{n}{k}\,p^k\,(1-p)^{n-k}$$

In the case of the median, with $p = 0{,}5$, the binomial distribution is

$$P\,(K=k\,;\,n,1/2) = \binom{n}{k}\Big/\,2^n$$

5 Point estimation

A point estimate of the population median is given by the sample median. The sample median is obtained by numbering

1) To be published.

the sample elements in non-decreasing order of their values and taking the value of

— the $[(n + 1)/2]$th order statistic, if n is odd, or

— the arithmetic mean of the $[n/2]$ and $[n/2 + 1]$th order statistic, if n is even.

NOTE — This estimator is in general biased, but an estimator that is unbiased for any population does not exist.

6 Confidence interval

6.1 General

A *two-sided confidence interval* for the population median is a closed interval of the form

$$[T_1, T_2]$$

where

$$T_1 < T_2;$$

T_1 and T_2 are called the *lower* and *upper confidence limits*, respectively.

A *one-sided confidence interval* may be

$$[T_1, \infty) \text{ or } (-\infty, T_2]$$

T_1 and T_2 are called the *lower* and *upper confidence limits*, respectively.

The practical meaning of a confidence interval is that the experimenter claims that the unknown M lies within the interval, while admitting a small nominal probability that his conclusion might be false. The probability that the interval thus calculated covers the population median is called the confidence level.

6.2 Accepted method

The lower and upper limits of a two-sided confidence interval of confidence level $1 - \alpha/2$ are given by the pair of order statistics $[x_{(k)}, x_{(n-k+1)}]$ where the integer k is determined in such a way that

$$\sum_{i=0}^{k-1} \binom{n}{i} < 2^{n-1}\alpha/2$$

and

$$\sum_{i=0}^{k} \binom{n}{i} > 2^{n-1}\alpha/2$$

In the one-sided case, $\alpha/2$ shall be replaced by α.

The solutions to these inequalities for different values of n are given in table 1.

For $5 < n < 30$ and for the usual values of $1 - \alpha$, both for one-sided and two-sided confidence intervals, the values of k are given such that the lower confidence limit will be

$$T_1 = x_{(k)}$$

and the upper confidence limit will be

$$T_2 = x_{(n-k+1)}$$

where $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, are the ordered observations in the sample.

For small values of n, it may happen that there are no confidence limits.

6.3 Approximate method

For values of n not given in table 1, a solution for k can be obtained using the following formula, which yields an approximate value:

$$y = \{0{,}5(n + 1 - u\sqrt{n - 0{,}5})\},$$

k is the integer part of y, and u is a fractile of the standard normal distribution defined by

$u = u_{1-\alpha}$ in the one-sided case;

$u = u_{1-\frac{\alpha}{2}}$ in the two-sided case.

This approximation is extremely accurate for the usual values of α.

When computer programmes are used, the following formula agrees with table 1 and with the simpler approximate formula when used with values greater than those given in table 1:

$$y = \{0{,}5(n + 1 - u\sqrt{n + 0{,}5 - 0{,}25\,u^2})\}$$

k is the integer part of y.

7 Examples

7.1 Electric cords for a small appliance are flexed by a test machine until failure.

The test simulates actual use, but highly accelerated. The 24 times of failure, in hours, are given below; seven of them are censored times and are marked with an asterisk[1)]:

57,5; 77,8; 88,0; 96,9; 98,4; 100,3; 100,8; 102,1; 103,3; 103,4; 105,3; 105,4; 122,6; 139,3; 143,9; 148,0; 151,3; 161,1*; 161,2*; 161,2*; 162,4*; 162,7*; 163,1*; 176,8*.

A point estimate of the median lifetime is

$$(x_{(12)} + x_{(13)})/2 = (105{,}4 + 122{,}6)/2$$

$$= 114$$

1) When an item is removed from test without having failed, the time on test is referred to as a "censored time."

The lower one-sided confidence limit for the median with the confidence level 0,95 is obtained by reading off the value k for $n = 24$ and the confidence level 0,95 for the one-sided case, and then looking for the kth value in the list above.

We find

$k = 8$ and

$[x_{(8)}] = [102,1]$

7.2 Wilk *et al.* (1962) give data on the lifetime of 34 transistors in an accelerated life test. The lifetimes (in weeks) are given below; three of them are censored times and are marked with an asterisk:

3; 4; 5; 6; 6; 7; 8; 8; 9; 9; 9; 10; 10; 11; 11; 11; 13; 13; 13; 13; 13; 17; 17; 19; 19; 25; 29; 33; 42; 42; 52; 52*; 52*; 52*.

A point estimate of the median lifetime is

$(x_{(17)} + x_{(18)})/2 = (13 + 13)/2$

$= 13$

Since $n > 30$, the approximate method has to be used to obtain the lower one-sided confidence limit for the median with the confidence level 0,95:

$$y = 0,5\,(n + 1 - u\sqrt{n - 0,5})$$
$$= 0,5\,(34 + 1 - 1,645\sqrt{34 - 0,5})$$
$$= 12,74$$

k is the integer part of y; hence

$k = 12$

and the required confidence limit is

$[x_{(12)}] = [10]$

It can be shown that the k value thus obtained is the same as that obtained using the exact method.

For the two-sided limits:

$$y = 0,5\,(n + 1 - u\sqrt{n - 0,5})$$
$$= 0,5\,(34 + 1 - 1,960\sqrt{34 - 0,5})$$
$$= 11,83$$

which gives $k = 11$, and $n - k + 1 = 24$. Hence the desired confidence interval is

$[x_{(11)}, x_{(24)}] = [9,19]$

Table 1 — Exact values of k as a function of n

n \ k	One-sided case		Two-sided case	
	Confidence level		Confidence level	
	0,95	0,99	0,95	0,99
5	1	0	0	0
6	1	0	1	0
7	1	1	1	0
8	2	1	1	1
9	2	1	2	1
10	2	1	2	1
11	3	2	2	1
12	3	2	3	2
13	4	2	3	2
14	4	3	3	2
15	4	3	4	3
16	5	3	4	3
17	5	4	5	3
18	6	4	5	4
19	6	5	5	4
20	6	5	6	4
21	7	5	6	5
22	7	6	6	5
23	8	6	7	5
24	8	6	7	6
25	8	7	8	6
26	9	7	8	7
27	9	8	8	7
28	10	8	9	7
29	10	8	9	8
30	11	9	10	8

NOTE — The "0" indicates that a confidence interval and confidence limit(s) cannot be determined for a sample size at this confidence level.

This page intentionally left blank