INTERNATIONAL STANDARD 3301

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION • МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ • ORGANISATION INTERNATIONALE DE NORMALISATION

Statistical interpretation of data — Comparison of two means in the case of paired observations

Interprétation statistique des données — Comparaison de deux moyennes dans le cas d'observations appariées

First edition — 1975-08-15

UDC 519.28

Ref. No. ISO 3301-1975 (E)

Descriptors : statistical analysis, mean, comparison.

Price based on 6 pages

ISO 3301-1975 (E)

FOREWORD

ISO (the International Organization for Standardization) is a worldwide federation of national standards institutes (ISO Member Bodies). The work of developing International Standards is carried out through ISO Technical Committees. Every Member Body interested in a subject for which a Technical Committee has been set up 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.

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.

International Standard ISO 3301 was drawn up by Technical Committee ISO/TC 69, *Applications of statistical methods*, and circulated to the Member Bodies in March 1974.

It has been approved by the Member Bodies of the following countries :

Austria	India	South Africa, Rep. of
Belgium	Israel	Spain
Brazil	Italy	Switzerland
Czechoslovakia	Netherlands	Turkey
France	Poland	United Kingdom
Germany	Portugal	U.S.S.R.
Hungary	Romania	Yugoslavia

The Member Bodies of the following countries expressed disapproval of the document on technical grounds :

Sweden
U.S.A.

Statistical interpretation of data – Comparison of two means in the case of paired observations

0 INTRODUCTION

The method specified in this International Standard, known as the method of paired observations, is a special case of the method described in table A′ of ISO 2854, *Statistical interpretation of data – Techniques of estimation and tests relating to means and variances.*[1)]

This special case is mentioned in section two of ISO 2854 immediately after the numerical illustration of table A′, and a complete example of applications of the method of paired comparisons has been given in annex A of that International Standard. The importance and wide applicability of the method justify a separate International Standard being devoted to it.

1 SCOPE

This International Standard specifies a method for comparing the mean of a population of differences between paired observations with zero or any other preassigned value.

2 DEFINITION

paired observations : Two observations x_i and y_i of a certain property or characteristic are said to be paired if they are made :

– on the same element i from a population of elements but under different conditions (for example, comparison of results of two methods of analysis on the same product);

– on two distinct elements considered similar in all respects except for the systematic difference which is the subject of the test (for example, comparison of the yield from adjacent plots sown with two distinct varieties of seed).

However, it should be noted that in the second case the efficiency of the test depends on the validity of the hypothesis that there is no other systematic difference between the individuals in the same pair other than the systematic difference under test.

3 FIELD OF APPLICATION

The method may be applied to establish a difference between two treatments. In this case, the observations x_i are carried out after the first treatment and y_i after the second treatment. The two series of results of the observations are not independent because each result x_i of the first series (first treatment) is associated with a result y_i of the second series (second treatment). The term "treatment" should be understood in a wide sense. The two treatments to be compared may, for instance, be two test methods, two measuring instruments or two laboratories, in order to detect a possible systematic error. Two treatments carried out successively on the same experimental material might interact and the value obtained might depend on the order. Good experimental design should enable such biases to be eliminated. Alternatively, only one treatment may be applied and its effect may be compared to the absence of treatment; the purpose of this comparison is then to establish the effect of that treatment.

4 CONDITIONS FOR APPLICATION

The method can be applied validly if the following two conditions are satisfied :

– the series of differences $d_i = x_i - y_i$ can be considered as a series of independent random items;

– the distribution of the differences $d_i = x_i - y_i$ between the paired observations is supposed to be normal or approximately normal.

If the distribution of these differences deviates from the normal, the technique described remains valid, provided the sample size is sufficiently large; the greater the deviation from normality, the larger the sample size required. Even in extreme cases, however, a sample size of 100 will be sufficient for most practical applications.

1) At present at the stage of draft.

5 FORMAL PRESENTATION OF CALCULATIONS

Problem studied . . . .

Experimental conditions . . . .

Statistical data

Sample size :

$n =$

Sum of the observed values :

$\Sigma x_i =$ $\Sigma y_i =$

Sum of the differences :

$\Sigma d_i =$

Sum of the squares of the differences :

$\Sigma d_i^2 =$

Given value :

$d_0 =$

Degrees of freedom :

$\nu = n - 1 =$

Chosen significance level :

$\alpha =$

Calculations

$$\bar{d} = \frac{1}{n}(\Sigma x_i - \Sigma y_i)$$

$$= \frac{1}{n}\Sigma d_i =$$

$$s_d^2 = \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{1}{n}(\Sigma d_i)^2\right] =$$

$$\sigma_d^* = \sqrt{s_d^2} =$$

$$A_1 = [t_{1-\alpha}(\nu)/\sqrt{n}]\ \sigma_d^* =$$

$$A_2 = [t_{1-\alpha/2}(\nu)/\sqrt{n}]\ \sigma_d^* =$$

Results

Two-sided case :

The hypothesis that the population mean of the difference is equal to d_0 (null hypothesis) is rejected if :

$$|\bar{d} - d_0| > A_2$$

One-sided cases :

a) The hypothesis that the population mean of the differences is at least equal to d_0 (null hypothesis) is rejected if :

$$\bar{d} < d_0 - A_1$$

b) The hypothesis that the population mean of the differences is at most equal to d_0 (null hypothesis) is rejected if :

$$\bar{d} > d_0 + A_1$$

NOTE $t_{1-\alpha}(\nu)$ is the fractile of order $1-\alpha$ of Student's variate t with ν degrees of freedom. The values of $t_{1-\alpha}(\nu)/\sqrt{n}$ are given in table 1.

TABLE 1 — **Values of the ratio** $t_{1-\alpha}(\nu)/\sqrt{n}$ **for** $\nu = n - 1$

$\nu = n - 1$	Two-sided case		One-sided case	
	$\frac{t_{0,975}}{\sqrt{n}}$	$\frac{t_{0,995}}{\sqrt{n}}$	$\frac{t_{0,95}}{\sqrt{n}}$	$\frac{t_{0,99}}{\sqrt{n}}$
1	8,985	45,013	4,465	22,501
2	2,434	5,730	1,686	4,021
3	1,591	2,920	1,177	2,270
4	1,242	2,059	0,953	1,676
5	1,049	1,646	0,823	1,374
6	0,925	1,401	0,734	1,188
7	0,836	1,237	0,670	1,060
8	0,769	1,118	0,620	0,966
9	0,715	1,028	0,580	0,892
10	0,672	0,956	0,546	0,833
11	0,635	0,897	0,518	0,785
12	0,604	0,847	0,494	0,744
13	0,577	0,805	0,473	0,708
14	0,554	0,769	0,455	0,678
15	0,533	0,737	0,438	0,651
16	0,514	0,708	0,423	0,626
17	0,497	0,683	0,410	0,605
18	0,482	0,660	0,398	0,586
19	0,468	0,640	0,387	0,568
20	0,455	0,621	0,376	0,552
21	0,443	0,604	0,367	0,537
22	0,432	0,588	0,358	0,523
23	0,422	0,573	0,350	0,510
24	0,413	0,559	0,342	0,498
25	0,404	0,547	0,335	0,487
26	0,396	0,535	0,328	0,477
27	0,388	0,524	0,322	0,467
28	0,380	0,513	0,316	0,458
29	0,373	0,503	0,310	0,449
30	0,367	0,494	0,305	0,441
40	0,316	0,422	0,263	0,378
50	0,281	0,375	0,235	0,337
60	0,256	0,341	0,214	0,306
70	0,237	0,314	0,198	0,283
80	0,221	0,293	0,185	0,264
90	0,208	0,276	0,174	0,248
100	0,197	0,261	0,165	0,235
200	0,139	0,183	0,117	0,165
500	0,088	0,116	0,074	0,104
∞	0	0	0	0

EXAMPLE : The data tabled below were collected during an investigation designed to determine whether the average rate of shaft-wear caused by various bearing metals in an internal combustion engine differed between metals.

TABLE 2 – **Shaft-wear after a given working time in 0.000 01 in**

Shaft i	Wear with copper-lead x_i	Wear with white metal y_i	Difference $d_i = x_i - y_i$
1	3.5	1.5	2.0
2	2.0	1.3	0.7
3	4.7	4.5	0.2
4	2.8	2.5	0.3
5	6.5	4.5	2.0
6	2.2	1.7	0.5
7	2.5	1.8	0.7
8	5.8	3.3	2.5
9	4.2	2.3	1.9
Total	34.2	23.4	10.8

Technical characteristics . . . .

Statistical data	**Calculations**
Sample size : $n = 9$	$\bar{d} = \frac{1}{9}(34.2 - 23.4) = 1.2$
Sum of the observed values : $\Sigma x_i = 34.2$ $\Sigma y_i = 23.4$	$s_d^2 = \frac{1}{8}\left(19.22 - \frac{10.8^2}{9}\right) = 0.782\,5$
Sum of the differences : $\Sigma d_i = 10.8$	$\sigma_d^* = \sqrt{0.782\,5} = 0.884\,6$
Sum of the squares of the differences : $\Sigma d_i^2 = 19.22$	$t_{0.995}/\sqrt{9} = 1.118$
Given value : $d_0 = 0$	$A_2 = 1.118 \times 0.884\,6 = 0.99$
Degrees of freedom : $\nu = 8$	
Chosen significance level : $\alpha = 0.01$	

Result

Comparison of the population mean with the given value 0 :

Two-sided case :

$$|\bar{d} - d_0| = 1.2 > 0.99$$

The hypothesis of the equality of the rate of shaft-wear by the two metals is rejected at the 1 % level.

6 ERRORS OF THE SECOND KIND

The probability of rejecting the null hypothesis when it is true is at most equal to the significance level α. Rejecting the null hypothesis when it is true is called an error of the first kind, and the choice of α therefore limits the risk of such an error.

On the other hand, it is possible to commit an error of the second kind, that is, accepting the null hypothesis when it is false. The probability $1-\beta$ of rejecting the null hypothesis when it is false is called the power of the test; the probability of an error of the second kind is therefore β.

For a given sample n and error of the first kind, these probabilities depend not only on the true mean D of the observed differences $d_i = x_i - y_i$ for which one can postulate different alternative hypotheses but also on the standard deviation σ_d of these differences. This standard deviation is in general unknown and if n is small the sample will provide only a poor estimator.

The result is that it is impossible to set an upper limit to the probability of an error of the second kind.

Nevertheless, in the following graphs the relation is shown between the power of the test, $1-\beta$, and the actual population mean divided by the corresponding standard deviation, D/σ_d for one-sided tests of the hypothesis H_0 : $D \leqslant 0$, and for various values of n and for the significance levels 0,05 and 0,01 respectively.

From these graphs the following conclusions may be drawn :

1) The power of the test is uniquely determined by the true mean of the differences, measured in units of their standard deviation, by the significance level α and the sample size.

2) The power function is a strictly increasing function of the true mean difference.

It is also strictly increasing with the sample size and the significance level α, provided $D > 0$ and α different from 0 and from 1.

3) With a significance level of 0,05 and a sample size of 50, a power of at least 0,95 is already obtained when the true mean difference exceeds one-half of the standard deviation of the differences. For $n = 20$, this power is obtained for $D/\sigma_d = 0,78$ or more.

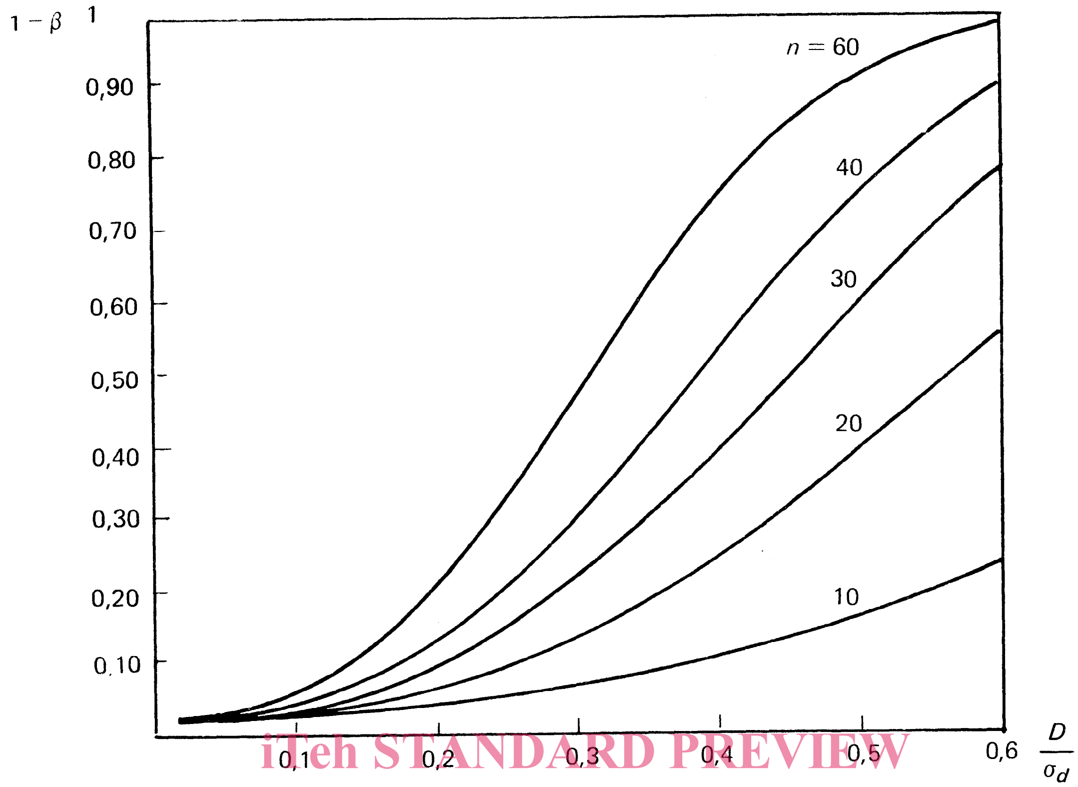

FIGURE 1 — Power of Student's one-sample test (one-sided), $\alpha = 0,01$

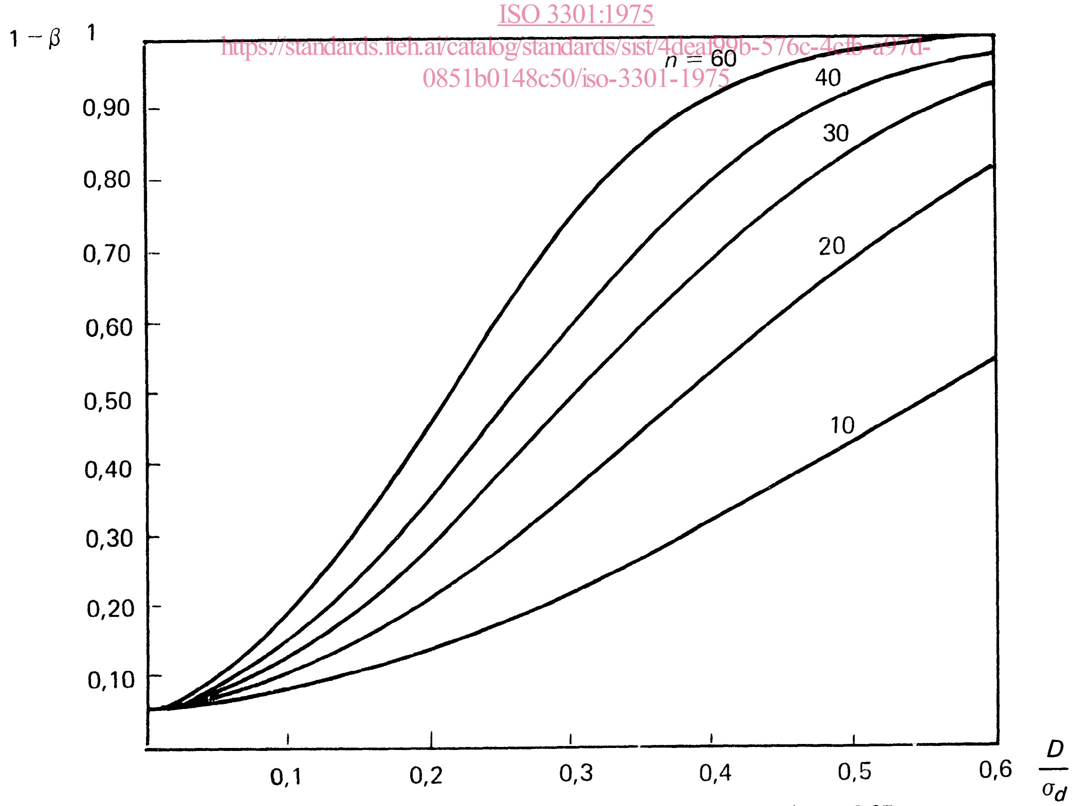

FIGURE 2 — Power of Student's one-sample test (one-sided), $\alpha = 0,05$

NOTE — The graphs are based on the work of D.B. OWEN, *Handbook of statistical tables,* Addison Wesley.