SLOVENSKI STANDARD
SIST ISO 3494:1996

01-september-1996

Statistical interpretation of data - Power of tests relating to means and variances

Statistical interpretation of data -- Power of tests relating to means and variances

Interprétation statistique des données -- Efficacité des tests portant sur des moyennes et des variances

Ta slovenski standard je istoveten z: ISO 3494:1976

ICS:

03.120.30	Uporaba statističnih metod	Application of statistical methods

SIST ISO 3494:1996 **en**

INTERNATIONAL STANDARD ISO 3494

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

Statistical interpretation of data — Power of tests relating to means and variances

Interprétation statistique des données — Efficacité des tests portant sur des moyennes et des variances

First edition — 1976-12-01

UDC 519.28

Ref. No. ISO 3494-1976 (E)

Descriptors : statistics, data, mean, variance (statistics), tests, statistical tests, efficiency, effectivness.

Price based on 44 pages

ISO 3494-1976 (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 3494 was drawn up by Technical Committee ISO/TC 69, *Applications of statistical methods*, and was circulated to the Member Bodies in March 1975.

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

Australia
Austria
Belgium
Brazil
Bulgaria
Canada
Czechoslovakia
France
Germany
Hungary
India
Israel
Mexico
Netherlands
New Zealand
Poland
Portugal
Romania
South Africa, Rep. of
Switzerland
Turkey
United Kingdom
U.S.A.
U.S.S.R.
Yugoslavia

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

Japan
Sweden

Printed in Switzerland

CONTENTS

Page

SECTION ONE : COMPARISON TESTS

SECTION TWO : SETS OF CURVES

This page intentionally left blank

Statistical interpretation of data – Power of tests relating to means and variances

SECTION ONE : COMPARISON TESTS

GENERAL REMARKS

1) This International Standard follows on from ISO 2854, *Statistical interpretation of data – Techniques of estimation and tests relating to means and variances.*

The conditions of application of this International Standard are as stated in the "General remarks" in ISO 2854. It will be recalled that the tests used are valid if the distribution of the observed variable is assumed to be normal in each population (see comments on paragraph 3 of the "General remarks" in ISO 2854). ISO 2854 is concerned only with the type I risk (or significance level). This International Standard puts forward notions of the type II risk and of power of the test.

2) It will also be recalled that the type I risk is the probability of rejecting the null hypothesis (tested hypothesis) if this hypothesis is true (case of two-sided tests), or the maximum value of this probability (case of one-sided tests). The non-rejection of the null hypothesis produces, in practice, acceptance of the hypothesis, yet non-rejection does not mean that the hypothesis is true.

Accordingly, the type II risk, designated by β, is the probability of not rejecting the null hypothesis when it is false. The complement of the probability of committing the error of the second kind $(1 - \beta)$ is the "power" of the test (see "Historical note" following these general remarks).

3) Whereas the value of the type I risk is chosen by the consumers according to the consequences that could arise from that risk (either of the values $\alpha = 0{,}05$ or $\alpha = 0{,}01$ is commonly employed), the type II risk is dependent on the true hypothesis (the null hypothesis H_0 being false), i.e. the alternative hypothesis to the null hypothesis. In the comparison of a population mean with a given value m_0, for example, a specific alternative corresponds to a value of the population mean of $m \neq m_0$ being a deviation $m - m_0 \neq 0$. As a general rule, in tests of comparison of means and variances, the alternatives are defined by the values that might be assumed by a parameter.

4) The operating characteristic curve of a test is the curve which shows the value β of the type II risk as a function of the parameter defining the alternative. β is also dependent on the value chosen for the type I risk, on size(s) of sample(s) and on the nature of the test (two-sided or one-sided).

In the tests of comparison of means, β also depends on the standard deviation of the population(s). Where this is unknown, the risk β cannot be known exactly.

5) The operating characteristic curves allow the following problems to be solved.

a) **problem 1** : For a given alternative and given size of sample, determine the probability β of not rejecting the null hypothesis (type II risk).

b) **problem 2** : For a given alternative and a given value of β determine the size of sample to be selected.

Although a single series of curve sets allows both problems to be solved, two series of sets will be presented, in order to facilitate practical applications :

– **sets 1.1 to 14.1**, giving the risk β as a function of the alternative, for $\alpha = 0{,}05$ or $\alpha = 0{,}01$ and for different values of the size(s) of sample.

– **sets 1.2 to 14.2**, giving the size(s) of sample to be selected as a function of the alternative, for $\alpha = 0{,}05$ or $\alpha = 0{,}01$ and for different values of the risk β.

6) Attention is drawn to the practical significance of interpreting statistics by means of tests of hypotheses and curves. When testing a hypothesis such as $m = m_0$ (or $m_1 = m_2$), it is generally desired to know whether it can be concluded with little risk of mistake, that m does not differ too greatly from m_0 (or m_1 does not differ too greatly from m_2). Moreover, the choice of the value $\alpha = 0{,}05$ or $\alpha = 0{,}01$ for the type I risk associated with the test has a degree of arbitrariness. Therefore, it may be useful to examine what the result of the test would be with values close to m_0 (or value of the difference $D = m_1 - m_2$ close to 0), possibly using both values of the type I risk $\alpha = 0{,}05$ and $\alpha = 0{,}01$ and, in these circumstances, to evaluate by means of the operating characteristic curves the risk β associated with different alternatives.

7) The sets of curves which are given in section two of this International Standard are described and discussed in six clauses which correspond to the tables in ISO 2854.

The detailed correspondence between the different sets, the problems which they allow to be solved, the clauses of this International Standard and the tables of ISO 2854, appear at the top of the group of sets.

HISTORICAL NOTE

The concepts "type I risk" and "type II risk" were introduced by J. Neyman and E. S. Pearson in an article which appeared in 1928. Subsequently, these authors considered that the complement of the probability of committing the error of the second kind – which they called "power" of the test, in its aptitude to reveal as significant a specified alternative to the null hypothesis (tested hypothesis) – was in general an easier concept for the users to understand. It is this "power", or the probability of revealing a given deviation from the null hypothesis, which they designated by the symbol β.

It is moreover not necessary to introduce the term "power". One can more simply speak of the probability that a statistical test applied to a sample, at a significance level α, reveals that a parameter λ of the population differs (when such is truly the case) by at least a given quantity from the specified value λ_0, or, in relation to it, in a ratio at least equal to a given number.

The change in notation was probably introduced in the United States by users of industrial applications of statistics, in order that the "consumer's risk", when designated by β, might be taken into consideration at the same time as the "producer's risk α".

The symbol β was adopted for the type II risk in ISO 3534, *Statistics – Vocabulary and symbols*, and it has therefore been adopted with the same significance in this International Standard. However, as this symbol is used, and will continue no doubt to be used, with both meanings in statistical literature, it is advisable to find out, in each case of use, the meaning which is effectively attributed to it.

1 COMPARISON OF A MEAN WITH A GIVEN VALUE (VARIANCE KNOWN)

See table A of ISO 2854.

1.1 Notations

n = sample size

m = population mean

m_0 = given value

σ = standard deviation for the population

1.2 Tested hypotheses

For a two-sided test, the null hypothesis is $m = m_0$, the alternative hypothesis corresponding to $m \neq m_0$.

For a one-sided test, the null hypothesis is

a) either $m \leqslant m_0$, the alternative hypothesis corresponding to $m > m_0$;

b) or $m \geqslant m_0$, the alternative hypothesis corresponding to $m < m_0$.

1.3 Problem 1 : n being given, determine the risk β

For the different values of m, the alternative is defined by the parameter λ $(0 < \lambda < \infty)$, with

a) $\lambda = \dfrac{\sqrt{n}\,|m - m_0|}{\sigma}$ (two-sided test) alternatives $m \neq m_0$

b) $\lambda = \dfrac{\sqrt{n}\,(m - m_0)}{\sigma}$ (one-sided test $m \leqslant m_0$) alternatives $m > m_0$

c) $\lambda = \dfrac{-\sqrt{n}\,(m - m_0)}{\sigma}$ (one-sided test $m \geqslant m_0$) alternatives $m < m_0$

According to the case, the set to be consulted is

- 1.1 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.1 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.1 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.1 (one-sided test) type I risk $\alpha = 0{,}01$

β is the ordinate of the point of the abscissa λ on the curve $\nu = \infty$ of the suitable test.

1.4 Problem 2 : β being given, determine the size n

For the different values of m, the alternative is defined by the parameter λ $(0 < \lambda < \infty)$, with

a) $\lambda = \dfrac{|m - m_0|}{\sigma}$ (two-sided test) alternatives $m \neq m_0$

b) $\lambda = \dfrac{m - m_0}{\sigma}$ (one-sided test $m \leqslant m_0$) alternatives $m > m_0$

c) $\lambda = -\dfrac{m - m_0}{\sigma}$ (one-sided test $m \geqslant m_0$) alternatives $m < m_0$

According to the case, the set to be consulted is

- 1.2 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.2 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.2 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.2 (one-sided test) type I risk $\alpha = 0{,}01$

n is the ordinate of the point on the abscissa λ on the straight line (broken line) which corresponds to the given value β.

1.5 Example

A producer of cotton yarn guarantees, for each of the batches he delivers, a mean breaking load (expressed in newtons) at least equal to $m_0 = 2{,}30$. The consumer only agrees to accept the batches after having verified on elements of yarn of a given length, taken from different bobbins, that the one-sided test, as described in ISO 2854, does not lead to a rejection of the hypothesis $m \geqslant m_0 = 2{,}30$, the value chosen for the type I risk being $\alpha = 0{,}05$ (α is therefore here the "producer's risk").

The consumer knows from experience that the mean of the different batches may vary, but the dispersion of the breaking loads within any one batch is practically constant with a standard deviation $\sigma = 0{,}33$.

1.5.1 The consumer envisages selecting $n = 10$ bobbins per batch, and wishes to know the probability that he will not reject the hypothesis $m \geqslant 2{,}30$ (hence to accept the batch) where in fact the mean breaking load would be $m = 2{,}10$.

The set to be consulted is set 3.1. The value of the parameter λ for $m = 2{,}10$ is

$$\lambda = \frac{-\sqrt{n}\,(m - m_0)}{\sigma} = \frac{\sqrt{10}\,(2{,}30 - 2{,}10)}{0{,}33} = 1{,}92$$

The straight line $\nu = \infty$ gives for $100\,\beta$ the value 36 : $\beta = 0{,}36$ or 36 %.

1.5.2 This value being considered by the consumer as much too high, he decides to select a sample of sufficient size for the risk β to be reduced to 0,10 (or 10 %) if $m = 2{,}10$.

The set to be consulted is set 3.2. The value of the parameter λ for $m = 2{,}10$ is

$$\lambda = -\frac{m - m_0}{\sigma} = \frac{2{,}30 - 2{,}10}{0{,}33} = 0{,}61$$

The value of n, read on the straight lines (broken) $\beta = 0{,}10$ is $n = 22$.

2 COMPARISON OF A MEAN WITH A GIVEN VALUE (VARIANCE UNKNOWN)

See table A′ of ISO 2854.

IMPORTANT NOTE

The type II risk β depends on the true value σ of the standard deviation for the population, which is unknown. Hence, β can only be known approximately, and this provided that an order of magnitude of σ is available. In the absence of any valid previous information, one will take for σ the estimation s obtained from the sample.

It is strongly recommended that the influence on the values read from the operating characteristic curve of an error made for the standard deviation σ should be considered. The inaccuracy can be very great where σ has been estimated from a sample of small size : allowance for this situation can be made by placing s within the confidence limits for σ calculated by the method in table F of ISO 2854.

2.1 Notations

n = sample size

m = population mean

m_0 = given value

σ = standard deviation for the population (which will be replaced by an approximate value)

$\nu = n - 1$

2.2 Tested hypotheses

For a two-sided test, the null hypothesis is $m = m_0$, the alternative hypotheses corresponding to $m \neq m_0$.

For a one-sided test, the null hypothesis is

a) either $m \leqslant m_0$, the alternative hypotheses corresponding to $m > m_0$;

b) or $m \geqslant m_0$, the alternative hypotheses corresponding to $m < m_0$.

2.3 Problem 1 : n being given, determine the risk β

For the different values of m, the alternative is defined by the parameter λ ($0 < \lambda < \infty$), with

a) $\lambda = \dfrac{\sqrt{n}\,|m - m_0|}{\sigma}$ (two-sided test) alternatives $m \neq m_0$

b) $\lambda = \dfrac{\sqrt{n}\,(m - m_0)}{\sigma}$ (one-sided test $m \leqslant m_0$) alternatives $m > m_0$

c) $\lambda = -\dfrac{\sqrt{n}\,(m - m_0)}{\sigma}$ (one-sided test $m \geqslant m_0$) alternatives $m < m_0$

According to the case, the set to be consulted is

- 1.1 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.1 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.1 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.1 (one-sided test) type I risk $\alpha = 0{,}01$

β is the ordinate of the point on the abscissa λ on the curve $\nu = n - 1$ of the suitable set.

2.4 Problem 2 : β being given, determine the size n

For the different values of m, the alternative is defined by the parameter λ ($0 < \lambda < \infty$), with

a) $\lambda = \dfrac{|m - m_0|}{\sigma}$ (two-sided test) alternatives $m \neq m_0$

b) $\lambda = \dfrac{m - m_0}{\sigma}$ (one-sided test $m \leqslant m_0$) alternatives $m > m_0$

c) $\lambda = -\dfrac{m - m_0}{\sigma}$ (one-sided test $m \geqslant m_0$) alternatives $m < m_0$

According to the case, the set to be consulted is

- 1.2 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.2 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.2 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.2 (one-sided test) type I risk $\alpha = 0{,}01$

n is the ordinate of the point of the abscissa λ on the curve which corresponds to the given value β.

2.5 Example

The example is the same as in 1.5, but the consumer does not know the exact value of the standard deviation of the breaking loads. He knows, however, from experience, that this is almost certainly within the limits

$$\sigma_I = 0{,}30 \quad \sigma_S = 0{,}45$$

2.5.1 The consumer envisages selecting $n = 10$ bobbins per batch, and wishes to know the probability that he will not reject the hypothesis $m \geqslant 2{,}30$ (hence to accept the batch), while in fact the mean breaking load would be $m = 2{,}10$.[1)]

The set to be consulted is set 3.1. The values of the parameter λ which correspond to the extreme values of σ are

$$\lambda_I = \frac{\sqrt{10}\,(2{,}30 - 2{,}10)}{0{,}30} = 2{,}1$$

$$\lambda_S = \frac{\sqrt{10}\,(2{,}30 - 2{,}10)}{0{,}45} = 1{,}4$$

1) That is, the probability that when using the Student test with the significance level $\alpha = 0{,}05$, the value $m = 2{,}10$ is not revealed to be significantly lower than $m_0 = 2{,}30$.

The corresponding values of 100 β read (by interpolation) $\nu = 9$ are 40 and 64; i.e.

$$\beta_I = 0{,}40 \text{ (or 40 \%)}$$

$$\beta_S = 0{,}64 \text{ (or 64 \%)}$$

2.5.2 The consumer wishes, in the most unfavourable hypothesis ($\sigma = \sigma_S = 0{,}45$), β not to exceed 0,10 (or 10 %) if $m = 2{,}10$.

The set to be consulted is set 3.2, curve $\beta = 0{,}10$, with

$$\lambda = \frac{2{,}30 - 2{,}10}{0{,}45} = 0{,}44$$

For $\beta = 0{,}10$ and $\lambda = 0{,}44$, one finds n in the order of 45.

If, after inspection of several batches, it is found that the standard deviation is stable, σ can be estimated with greater precision : the sample size to be taken from the following batches can probably be reduced, with the guarantees of the producer and the consumer being maintained.

3 COMPARISON OF TWO MEANS (VARIANCES KNOWN)

See table C of ISO 2854.

3.1 Notations

	Population No. 1	Population No. 2
Sample size	n_1	n_2
Mean	m_1	m_2
Variance	σ_1^2	σ_2^2
Standard deviation of the difference of the mean of the two samples	$\sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	

3.2 Tested hypotheses

For a two-sided test, the null hypothesis is $m_1 = m_2$, the alternative hypotheses corresponding to $m_1 \neq m_2$.

For a one-sided test, the null hypothesis is

a) either $m_1 \leqslant m_2$, the alternative hypotheses corresponding to $m_1 > m_2$;

b) or $m_1 \geqslant m_2$, the alternative hypotheses corresponding to $m_1 < m_2$.

3.3 Problem 1 : n_1 and n_2 being given, determine the risk β

For the different values of the difference $m_1 - m_2$, the alternative is defined by the parameter λ ($0 < \lambda < \infty$), with

a) $\lambda = \frac{|m_1 - m_2|}{\sigma_d}$ (two-sided test) alternatives $m_1 \neq m_2$

b) $\lambda = \frac{m_1 - m_2}{\sigma_d}$ (one-sided test $m_1 \leqslant m_2$) alternatives $m_1 > m_2$

c) $\lambda = \frac{m_2 - m_1}{\sigma_d}$ (one-sided test $m_1 \geqslant m_2$) alternatives $m_1 < m_2$

According to the case, the set to be consulted is

- 1.1 (two-sided test) type I risk $\alpha = 0,05$
- 2.1 (two-sided test) type I risk $\alpha = 0,01$
- 3.1 (one-sided test) type I risk $\alpha = 0,05$
- 4.1 (one-sided test) type I risk $\alpha = 0,01$

β is the ordinate of the point on the abscissa λ on the curve $\nu = \infty$ of the suitable set.

When the total size of the two samples is fixed, $n_1 + n_2 = 2n$, the best efficiency (β minimum) is obtained with :

$$\frac{n_1}{\sigma_1} = \frac{n_2}{\sigma_2}$$

hence

$$n_1 = 2n \frac{\sigma_1}{\sigma_1 + \sigma_2}$$

$$n_2 = 2n \frac{\sigma_2}{\sigma_1 + \sigma_2}$$

$$\lambda = \sqrt{2n} \frac{|m_1 - m_2|}{\sigma_1 + \sigma_2}$$

$$\left(\lambda = \sqrt{\frac{n}{2}} \frac{|m_1 - m_2|}{\sigma}, \text{ if } \sigma_1 = \sigma_2 = \sigma\right)$$

3.4 Problem 2 : β being given, determine the sizes n_1 and n_2

Using, according to the case, set 1.1, 2.1, 3.1 or 4.1, the curve $\nu = \infty$ allows the problem to be solved in the general case. The ordinate β corresponds to the point on the abscissa λ of this curve, and any pair (n_1, n_2) is suitable on the condition that

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \left(\frac{m_1 - m_2}{\lambda}\right)^2$$

The most economical sample ($n_1 + n_2$ minimum) is such that

$$\frac{n_1}{\sigma_1} = \frac{n_2}{\sigma_2}$$

hence

$$n_1 = \sigma_1 (\sigma_1 + \sigma_2) \left(\frac{\lambda}{m_1 - m_2}\right)^2$$

$$n_2 = \sigma_2 (\sigma_1 + \sigma_2) \left(\frac{\lambda}{m_1 - m_2}\right)^2$$

$$\left(n_1 = n_2 = 2 \left(\frac{\lambda\sigma}{m_1 - m_2}\right)^2, \text{ if } \sigma_1 = \sigma_2 = \sigma\right)$$

In the particular case where $\sigma_1 = \sigma_2 = \sigma$, $n_1 = n_2 = n$, it is more suitable to define, for the different values of the difference $m_1 - m_2$, the alternative by the parameter λ ($0 < \lambda < \infty$), with

a) $\lambda = \frac{|m_1 - m_2|}{\sigma\sqrt{2}}$ (two-sided test) alternatives $m_1 \neq m_2$

b) $\lambda = \frac{m_1 - m_2}{\sigma\sqrt{2}}$ (one-sided test $m_1 \leqslant m_2$) alternatives $m_1 > m_2$

c) $\lambda = \frac{m_2 - m_1}{\sigma\sqrt{2}}$ (one-sided test $m_1 \geqslant m_2$) alternatives $m_1 < m_2$

and to use according to the case, one of the following sets :

- 1.2 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.2 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.2 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.2 (one-sided test) type I risk $\alpha = 0{,}01$

n is the ordinate of the point on the abscissa λ on the straight line (broken line) which corresponds to the given value β.

3.5 Example

A producer of cotton yarn has modified his process, but according to his declaration, the mean breaking load remains the same ($m_1 = m_2$), m_1 corresponding to the old process and m_2 to the new.

The consumer is prepared to adopt the new process, but wishes to verify the declaration of the producer, by carrying out on elements of yarn of a given length taken from different bobbins, the two-sided test of the hypothesis $m_1 = m_2$ as described in ISO 2854, with for the value of the type I risk $\alpha = 0{,}05$ (α is therefore here the "producer's risk").

The consumer knows, from experience, that for all the productions of this producer, the dispersion of the breaking load is practically constant and characterized by a standard deviation $\sigma = 0{,}33$.

3.5.1 The consumer envisages selecting 10 bobbins from a batch of each of the two processes, and wishes to know the probability that he will not reject the hypothesis $m_1 = m_2$ (hence to accept the batch of the new process) while in fact $|m_1 - m_2|$ would be equal to 0,30.

The set to be consulted is set 1.1, with

$$\lambda = \frac{|m_1 - m_2|}{\sigma_d}$$

$$|m_1 - m_2| = 0{,}30$$

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2}{n}}\,\sigma = \sqrt{\frac{2}{10}} \times 0{,}33 = 0{,}147\,6$$

$$\lambda = \frac{0{,}30}{0{,}147\,6} = 2{,}03$$

The curve $\nu = \infty$ gives for $100\,\beta$ the value 47 : $\beta = 0{,}47$ or 47 %.

3.5.2 This value being considered by the consumer as much too high, he decides to select samples of a sufficiently large size for the risk β to be reduced to 0,10 (or 10 %) when $m_1 - m_2 = 0{,}30$.

The set to be consulted is set 1.2, with

$$\lambda = \frac{|m_1 - m_2|}{\sigma\sqrt{2}} = \frac{0{,}30}{0{,}33\sqrt{2}} = 0{,}64$$

The value of n, read on the straight line (broken line) $\beta = 0{,}10$, is $n = 26$.

4 COMPARISON OF TWO MEANS (VARIANCES UNKNOWN BUT MAY BE ASSUMED EQUAL)

See table C′ of ISO 2854.

IMPORTANT NOTE

The type II risk β depends on the true value σ of the standard deviation of the two populations, which is unknown. Hence β can only be known approximately, and this provided that an order of magnitude of σ is available. In the absence of any valid previous information, one will take for σ the estimation s obtained from samples.

It is strongly recommended that the influence on the values read on the curves of an error made for the standard deviation σ should be considered. The inaccuracy can be very great where σ has been estimated from samples of small size; allowance for this situation can be made by placing s within the confidence limits for σ, calculated by the method in table F of ISO 2854.

4.1 Notations

	Population No. 1	Population No. 2
Sample size	n_1	n_2
Mean	m_1	m_2
Variance (which will be replaced by an approximate value)	σ^2	σ^2
Number of degrees of freedom	$\nu = n_1 + n_2 - 2$ $[2(n-1)$, if $n_1 = n_2 = n]$	
Standard deviation of the difference between the means of the two samples	$\sigma_d = \sqrt{\frac{n_1 + n_2}{n_1 n_2}}\,\sigma$ $\left(\sqrt{\frac{2}{n}}\,\sigma\right.$, if $\left.n_1 = n_2 = n\right)$	

4.2 Tested hypotheses

For a two-sided test, the null hypothesis is $m_1 = m_2$, the alternative hypotheses corresponding to $m_1 \neq m_2$.

For a one-sided test, the null hypothesis is

a) either $m_1 \leqslant m_2$, the alternative hypotheses corresponding to $m_1 > m_2$;

b) or $m_1 \geqslant m_2$, the alternative hypotheses corresponding to $m_1 < m_2$.

4.3 Problem 1 : n_1 and n_2 being given, determine the risk β

For the different values of the difference $m_1 - m_2$, the alternative is defined by the parameter λ $(0 < \lambda < \infty)$, with

a) $\lambda = \dfrac{|m_1 - m_2|}{\sigma_d}$ (two-sided test) alternatives $m_1 \neq m_2$

b) $\lambda = \dfrac{m_1 - m_2}{\sigma_d}$ (one-sided test $m_1 \leqslant m_2$) alternatives $m_1 > m_2$

c) $\lambda = \dfrac{m_2 - m_1}{\sigma_d}$ (one-sided test $m_1 \geqslant m_2$) alternatives $m_1 < m_2$

According to the case, the set to be consulted is

- 1.1 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.1 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.1 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.1 (one-sided test) type I risk $\alpha = 0{,}01$

β is the ordinate of the point on the abscissa λ on the curve $\lambda = n_1 + n_2 - 2$ of the suitable set.

When only the total size of the two samples is fixed, $n_1 + n_2 = 2n$, it is of interest to take $n_1 = n_2 = n$ (β minimum). One then has :

$$\lambda = \sqrt{\frac{n}{2}}\,\frac{|m_1 - m_2|}{\sigma}$$

4.4 Problem 2 : β being given, determine the common size n of the samples

For the different values of the difference $m_1 - m_2$, the alternative is defined by the parameter λ $(0 < \lambda < \infty)$, with

a) $\lambda = \dfrac{|m_1 - m_2|}{\sigma\sqrt{2}}$ (two-sided test) alternatives $m_1 \neq m_2$

b) $\lambda = \dfrac{m_1 - m_2}{\sigma\sqrt{2}}$ (one-sided test $m_1 \leqslant m_2$) alternatives $m_1 > m_2$

c) $\lambda = \dfrac{m_2 - m_1}{\sigma\sqrt{2}}$ (one-sided test $m_1 \geqslant m_2$) alternatives $m_1 < m_2$

According to the case, the set to be consulted is

- 1.2 (two-sided test) type I risk $\alpha = 0{,}05$
- 2.2 (two-sided test) type I risk $\alpha = 0{,}01$
- 3.2 (one-sided test) type I risk $\alpha = 0{,}05$
- 4.2 (one-sided test) type I risk $\alpha = 0{,}01$

n is the ordinate of the point on the abscissa λ on the curve which corresponds to the given value β.

4.5 Example

The example is the same as in 3.5, but the consumer does not know the exact value of the standard deviation of the breaking loads. He only knows that there is a great likelihood that it will be the same for the two batches ($\sigma_1 = \sigma_2$).

4.5.1 The consumer envisages selecting 10 bobbins from a batch of each of the two processes, and wishes to know the probability that he will not reject the hypothesis $m_1 = m_2$ (hence to accept the batch of the new process), while in fact $|m_1 - m_2|$ would be equal to 0,30.[1)]

The measurements carried out on the two samples give the following results :

a) First batch : $\bar{x}_1 = 2{,}176 \qquad \Sigma (x_1 - \bar{x}_1)^2 = 1{,}256\ 3$

b) Second batch : $\bar{x}_2 = 2{,}520 \qquad \Sigma (x_2 - \bar{x}_2)^2 = 1{,}389\ 7$

The small difference between the two sums of the squares is perfectly compatible with the hypothesis made above that : $\sigma_1^2 = \sigma_2^2$ (see table G of ISO 2854).

The estimation of the common variance σ^2 for the two batches is

$$s^2 = \frac{1{,}256\ 3 + 1{,}389\ 7}{10 + 10 - 2} = \frac{2{,}646\ 0}{18} = 0{,}147\ 0$$

The upper limit of σ^2, at the confidence level $1 - \alpha = 0{,}95$, is (see table F of ISO 2854)

$$\sigma_S^2 = \frac{2{,}646\ 0}{\chi^2_{0,05}\ (18)} = \frac{2{,}646\ 0}{9{,}39} = 0{,}281\ 8$$

It is therefore not very probable that σ will be greater than

$$\sigma_S = \sqrt{0{,}281\ 8} = 0{,}53$$

The set to be consulted is set 1.1, with

$$\lambda_S = \sqrt{\frac{n}{2}}\ \frac{|m_1 - m_2|}{\sigma_S} = \sqrt{\frac{10}{2}} \times \frac{0{,}30}{0{,}53} = 1{,}27$$

For $\nu = 18$, one finds (by interpolation) that the corresponding value of $100\,\beta$ is close to 80 : the upper limit of the type II risk is about 0,80 (or 80 %).

4.5.2 The consumer wishes, in the most unfavourable hypothesis ($\sigma = \sigma_S = 0{,}53$), β not to exceed 0,20 (or 20 %) when $|m_1 - m_2| = 0{,}30$.

The set to be consulted is set 1.2, curve $\beta = 0{,}20$, with

$$\lambda = \frac{|m_1 - m_2|}{\sigma_S \sqrt{2}} = \frac{0{,}30}{0{,}53\sqrt{2}} = 0{,}4$$

For $\beta = 0{,}20$ and $\lambda = 0{,}4$, one finds $n = 49$.

The $2 \times 50 = 100$ measurements will permit a quite accurate estimation of σ, on the basis of which set 1.1 will give an approximate value of the type II risk associated with the alternative $m_1 - m_2 = 0{,}30$.

1) That is, the probability that when using the Student test with the significance level $\alpha = 0{,}05$, a difference $|m_1 - m_2| = 0{,}30$ will not be revealed.