BS 2846 : Part 5 : 1977
ISO 3494 – 1976
UDC 31 : /519.222 : /519.724.6

Guide to Statistical interpretation of data

Part 5. Power of tests relating to means and variances

[ISO title: Statistical interpretation of data – Power of tests relating to means and variances]

Guide pour l'interprétation statistique des données
Partie 5. Efficacité des tests portant sur des moyennes et des variances

Hinweise für die statistische Auswertung von Daten
Teil 5. Schärfe von Prüfungen mit Bezug auf Durchschnittswerte und Variante

Contents

Foreword

This Part of BS 2846 is identical with ISO 3494 — 1976 'Statistical interpretation of data — Power of tests relating to means and variances'.

The correct interpretation and presentation of test results have been assuming increasing importance in the analysis of data obtained from manufacturing processes based on sample determinations and prototype evaluations in industry, commerce and educational institutions. It was for this reason that Subcommittee 2 of Technical Committee 69, 'Application of statistical methods', of the International Organization for Standardization (ISO), was charged with the task of preparing a guide to statistical methods for the interpretation of test results. As international agreement was reached on the statistical tests relevant to specific situations, it was decided to publish them as Parts of a revised BS 2846 as follows:

Statistical interpretation of data

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]

Part 5 of this standard is complementary to Part 4 which provides a series of statistical techniques applicable when the characteristic of interest is measured on a continuous scale and its distribution is Normal. In presenting these techniques Part 4 was only concerned with the error of the first kind, i.e. the error of rejecting the null hypothesis when true, or more particularly the maximum value of the probability of committing such an error that could be tolerated, the *significance level* of the test.

continued on inside back cover

British Standard Guide to

Statistical interpretation of data

Part 5. 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.