

This is a preview of "ISO 16269-4:2010". [Click here to purchase the full version from the ANSI store.](#)

First edition
2010-10-15

Statistical interpretation of data — Part 4: Detection and treatment of outliers

Interprétation statistique des données —

Partie 4: Détection et traitement des valeurs aberrantes



Reference number
ISO 16269-4:2010(E)

© ISO 2010

This is a preview of "ISO 16269-4:2010". [Click here to purchase the full version from the ANSI store.](#)

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.



COPYRIGHT PROTECTED DOCUMENT

© ISO 2010

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

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

This is a preview of "ISO 16269-4:2010". Click here to purchase the full version from the ANSI store.

Contents

Page

Foreword	iv
Introduction.....	v
1 Scope.....	1
2 Terms and definitions	1
3 Symbols.....	10
4 Outliers in univariate data	11
4.1 General	11
4.1.1 What is an outlier?	11
4.1.2 What are the causes of outliers?	11
4.1.3 Why should outliers be detected?.....	11
4.2 Data screening.....	12
4.3 Tests for outliers	14
4.3.1 General	14
4.3.2 Sample from a normal distribution.....	14
4.3.3 Sample from an exponential distribution.....	16
4.3.4 Samples taken from some known non-normal distributions.....	18
4.3.5 Sample taken from unknown distributions.....	19
4.3.6 Cochran's test for outlying variance	21
4.4 Graphical test of outliers	22
5 Accommodating outliers in univariate data.....	23
5.1 Robust data analysis.....	23
5.2 Robust estimation of location.....	24
5.2.1 General	24
5.2.2 Trimmed mean	24
5.2.3 Biweight location estimate	25
5.3 Robust estimation of dispersion	25
5.3.1 General	25
5.3.2 Median-median absolute pair-wise deviation	25
5.3.3 Biweight scale estimate.....	26
6 Outliers in multivariate and regression data.....	26
6.1 General	26
6.2 Outliers in multivariate data	26
6.3 Outliers in linear regression.....	28
6.3.1 General	28
6.3.2 Linear regression models.....	29
6.3.3 Detecting outlying Y observations.....	31
6.3.4 Identifying outlying X observations.....	31
6.3.5 Detecting influential observations.....	32
6.3.6 A robust regression procedure.....	35
Annex A (informative) Algorithm for the GESD outliers detection procedure	36
Annex B (normative) Critical values of outliers test statistics for exponential samples	37
Annex C (normative) Factor values of the modified box plot	44
Annex D (normative) Values of the correction factors for the robust estimators of the scale parameter	47
Annex E (normative) Critical values of Cochran's test statistic	48
Annex F (informative) A structured guide to detection of outliers in univariate data	51
Bibliography.....	54

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 16269-4 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*.

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

- *Part 4: Detection and treatment of outliers*
- *Part 6: Determination of statistical tolerance intervals*
- *Part 7: Median — Estimation and confidence intervals*
- *Part 8: Determination of prediction intervals*

This is a preview of "ISO 16269-4:2010". [Click here to purchase the full version from the ANSI store.](#)

Introduction

Identification of outliers is one of the oldest problems in interpreting data. Causes of outliers include measurement error, sampling error, intentional under- or over-reporting of sampling results, incorrect recording, incorrect distributional or model assumptions of the data set, and rare observations, etc.

Outliers can distort and reduce the information contained in the data source or generating mechanism. In the manufacturing industry, the existence of outliers will undermine the effectiveness of any process/product design and quality control procedures. Possible outliers are not necessarily *bad* or *erroneous*. In some situations, an outlier may carry essential information and thus it should be identified for further study.

The study and detection of outliers from measurement processes leads to better understanding of the processes and proper data analysis that subsequently results in improved inferences.

In view of the enormous volume of literature on the topic of outliers, it is of great importance for the international community to identify and standardize a sound subset of methods used in the identification and treatment of outliers. The implementation of this part of ISO 16269 enables business and industry to recognize the data analyses conducted across member countries or organizations.

Six annexes are provided. Annex A provides an algorithm for computing the test statistic and critical values of a procedure in detecting outliers in a data set taken from a normal distribution. Annexes B, D and E provide the tables needed to implement the recommended procedures. Annex C provides the tables and statistical theory that underlie the construction of modified box plots in outlier detection. Annex F provides a structured guide and flow chart to the procedures recommended in this part of ISO 16269.