

INTERNATIONAL STANDARD

NORME INTERNATIONALE

**Guide for the statistical analysis of ageing test data –
Part 1: Methods based on mean values of normally distributed test results**

**Guide pour l'analyse statistique de données d'essais de vieillissement –
Partie 1: Méthodes basées sur les valeurs moyennes de résultats d'essais
normalement distribués**



THIS PUBLICATION IS COPYRIGHT PROTECTED

Copyright © 2011 IEC, Geneva, Switzerland

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 IEC or IEC's member National Committee in the country of the requester.

If you have any questions about IEC copyright or have an enquiry about obtaining additional rights to this publication, please contact the address below or your local IEC member National Committee for further information.

Droits de reproduction réservés. Sauf indication contraire, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de la CEI ou du Comité national de la CEI du pays du demandeur.

Si vous avez des questions sur le copyright de la CEI ou si vous désirez obtenir des droits supplémentaires sur cette publication, utilisez les coordonnées ci-après ou contactez le Comité national de la CEI de votre pays de résidence.

IEC Central Office
3, rue de Varembe
CH-1211 Geneva 20
Switzerland
Email: inmail@iec.ch
Web: www.iec.ch

About the IEC

The International Electrotechnical Commission (IEC) is the leading global organization that prepares and publishes International Standards for all electrical, electronic and related technologies.

About IEC publications

The technical content of IEC publications is kept under constant review by the IEC. Please make sure that you have the latest edition, a corrigenda or an amendment might have been published.

- Catalogue of IEC publications: www.iec.ch/searchpub

The IEC on-line Catalogue enables you to search by a variety of criteria (reference number, text, technical committee,...). It also gives information on projects, withdrawn and replaced publications.

- IEC Just Published: www.iec.ch/online_news/justpub

Stay up to date on all new IEC publications. Just Published details twice a month all new publications released. Available on-line and also by email.

- Electropedia: www.electropedia.org

The world's leading online dictionary of electronic and electrical terms containing more than 20 000 terms and definitions in English and French, with equivalent terms in additional languages. Also known as the International Electrotechnical Vocabulary online.

- Customer Service Centre: www.iec.ch/webstore/custserv

If you wish to give us your feedback on this publication or need further assistance, please visit the Customer Service Centre FAQ or contact us:

Email: csc@iec.ch

Tel.: +41 22 919 02 11

Fax: +41 22 919 03 00

A propos de la CEI

La Commission Electrotechnique Internationale (CEI) est la première organisation mondiale qui élabore et publie des normes internationales pour tout ce qui a trait à l'électricité, à l'électronique et aux technologies apparentées.

A propos des publications CEI

Le contenu technique des publications de la CEI est constamment revu. Veuillez vous assurer que vous possédez l'édition la plus récente, un corrigendum ou amendement peut avoir été publié.

- Catalogue des publications de la CEI: www.iec.ch/searchpub/cur_fut-f.htm

Le Catalogue en-ligne de la CEI vous permet d'effectuer des recherches en utilisant différents critères (numéro de référence, texte, comité d'études,...). Il donne aussi des informations sur les projets et les publications retirées ou remplacées.

- Just Published CEI: www.iec.ch/online_news/justpub

Restez informé sur les nouvelles publications de la CEI. Just Published détaille deux fois par mois les nouvelles publications parues. Disponible en-ligne et aussi par email.

- Electropedia: www.electropedia.org

Le premier dictionnaire en ligne au monde de termes électroniques et électriques. Il contient plus de 20 000 termes et définitions en anglais et en français, ainsi que les termes équivalents dans les langues additionnelles. Egalement appelé Vocabulaire Electrotechnique International en ligne.

- Service Clients: www.iec.ch/webstore/custserv/custserv_entry-f.htm

Si vous désirez nous donner des commentaires sur cette publication ou si vous avez des questions, visitez le FAQ du Service clients ou contactez-nous:

Email: csc@iec.ch

Tél.: +41 22 919 02 11

Fax: +41 22 919 03 00

INTERNATIONAL STANDARD

NORME INTERNATIONALE

**Guide for the statistical analysis of ageing test data –
Part 1: Methods based on mean values of normally distributed test results**

**Guide pour l'analyse statistique de données d'essais de vieillissement –
Partie 1: Méthodes basées sur les valeurs moyennes de résultats d'essais
normalement distribués**

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

COMMISSION
ELECTROTECHNIQUE
INTERNATIONALE

PRICE CODE
CODE PRIX



ICS 29.035.01

ISBN 978-2-88912-834-1

CONTENTS

FOREWORD.....	3
INTRODUCTION.....	5
1 Scope.....	6
2 Normative references	6
3 Terms, definitions and symbols	6
3.1 Terms and definitions	6
3.2 Symbols	8
4 Calculation procedures.....	9
4.1 General considerations.....	9
4.2 Single sub-group – Difference of mean and specified value.....	9
4.2.1 General	9
4.2.2 Complete data sub-group	9
4.2.3 Censored data sub-group	10
4.3 Two subgroups – Difference of means.....	10
4.3.1 General	10
4.3.2 Both sub-groups complete	10
4.3.3 One or both subgroups censored.....	11
4.4 Two or more subgroups – Analysis of variance.....	11
4.5 Three or more subgroups – Regression analysis	13
4.5.1 Regression analysis – General considerations.....	13
4.5.2 Calculations.....	14
4.5.3 Test equality of subgroup variances	15
4.5.4 Test significance of deviations from linearity.....	16
4.5.5 Estimate and confidence limit of y	16
4.5.6 Estimate and confidence limit of x	16
Annex A (informative) Statistical background	18
Annex B (informative) Statistical tables.....	22
Bibliography.....	35
Table B.1 – Coefficients for censored data calculations	23
Table B.2 – Fractiles of the F -distribution, $F_{0,95}$	30
Table B.3 – Fractiles of the F -distribution, $F_{0,995}$	32
Table B.4 – Fractiles of the t -distribution, $t_{0,95}$	34
Table B.5 – Fractiles of the χ^2 -distribution.....	34

INTERNATIONAL ELECTROTECHNICAL COMMISSION

**GUIDE FOR THE STATISTICAL ANALYSIS
OF AGEING TEST DATA –****Part 1: Methods based on mean values
of normally distributed test results**

FOREWORD

- 1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as “IEC Publication(s)”). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
- 3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
- 4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
- 5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
- 6) All users should ensure that they have the latest edition of this publication.
- 7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
- 8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
- 9) Attention is drawn to the possibility that some of the elements of this IEC Publication may be the subject of patent rights. IEC shall not be held responsible for identifying any or all such patent rights.

International Standard IEC 60493-1 has been prepared by IEC technical committee 112: Evaluation and qualification of electrical insulating materials and systems.

This second edition cancels and replaces the first edition, published in 1974, and constitutes a technical revision.

The main changes with respect to the first edition are that, besides a complete editorial revision, censored data sub-group are considered.

The text of this standard is based on the following documents:

CDV	Report on voting
112/172/CDV	112/192/RVC

Full information on the voting for the approval of this standard can be found in the report on voting indicated in the above table.

This publication has been drafted in accordance with the ISO/IEC Directives, Part 2.

A list of all the parts in the IEC 60493 series, published under the general title *Guide for the statistical analysis of ageing test data*, can be found on the IEC website.

The committee has decided that the contents of this publication will remain unchanged until the stability date indicated on the IEC web site under "<http://webstore.iec.ch>" in the data related to the specific publication. At this date, the publication will be

- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

INTRODUCTION

Procedures for estimating ageing properties are described in specific test procedures, or are covered by the general documents on test procedures for ageing tests with a specific environmental stress (e.g. temperature, radiation, partial discharges).

In many cases, a certain property is determined as a function of time at different ageing stresses, and a time to failure based on a chosen end-point criterion is found at each ageing stress. A plot of time to failure versus ageing stress may be used to obtain an estimate of the time to failure for similar specimens exposed to a specified stress, or to obtain an estimate of the value of stress which will cause failure in a specified time.

The physical and chemical laws governing the ageing phenomena may often lead to the assumption that a linear relationship exists between the property examined and the ageing time at fixed ageing stresses, or between certain mathematical functions of property and ageing time, e.g. square root or logarithm. Also, there may be a linear relationship between time to failure and ageing stress, or mathematical functions of these variables.

The methods described in this part of IEC 60493 apply to such cases of linear relationship. The methods are illustrated by the example of thermal ageing wherein the case of a simple chemical process it may be assumed that the degradation obeys the Arrhenius law, i.e. the logarithm of time to failure is a linear function of the reciprocal thermodynamic temperature. Numerical examples demonstrating the use of the methods in this case are given in IEC 60216-3 [1]¹.

The calculation processes specified in this standard are based on the assumption that the data under examination are normally distributed. No test for normality of the data is specified, since the available tests are unreliable for small sample groups of data. However, the methods have been used for a considerable time without undesirable results and with no check on the normality of the data distributions.

¹ Figures in square brackets refer to the bibliography.

GUIDE FOR THE STATISTICAL ANALYSIS OF AGEING TEST DATA –

Part 1: Methods based on mean values of normally distributed test results

1 Scope

This part of IEC 60493 gives statistical methods which may be applied to the analysis and evaluation of the results of ageing tests.

It covers numerical methods based on mean values of normally distributed test results.

These methods are only valid under specific assumptions regarding the mathematical and physical laws obeyed by the test data. Statistical tests for the validity of some of these assumptions are also given.

This standard deals with data from both complete test sets and censored test sets.

This standard provides data treatment based on the concept of "data sub-group" as defined in Clause 3. The validity of the coefficients used in the calculation processes to derive statistical parameters of the data groups are described in [1].

2 Normative references

None.

3 Terms, definitions and symbols

3.1 Terms and definitions

For the purposes of this document, the following terms, definitions and symbols apply.

3.1.1

ordered data

set of data arranged in sequence so that in the appropriate direction through the sequence each member is greater than or equal to its predecessor

Note 1 to entry: "Ascending order" in this standard implies that the data is ordered in this way, the first being the smallest.

3.1.2

order-statistic

each individual value in a set of ordered data is referred to as an "order-statistic" identified by its numerical position in the sequence

3.1.3

incomplete data

ordered data, where the values above and/or below defined points are not known

3.1.4

censored data

incomplete data, where the number of unknown values is known

Note 1 to entry: If the censoring is begun above/below a specified numerical value, the censoring is Type I. If it is begun above/below a specified order-statistic, it is Type II. This standard is concerned only with Type II.

3.1.5

truncated data

incomplete data where the number of unknown values is not known

Note 1 to entry: This report is not concerned with truncated data.

3.1.6

Saw coefficient

one of the coefficients developed by J.G. Saw for calculating the primary statistical functions of a single sub-group

Note 1 to entry: There are four coefficients used in this standard. Saw originally defined five, the fifth being intended for estimating the variance of the variance estimate (see [7]).

3.1.7

degrees of freedom

number of data values minus the number of parameter values

3.1.8

variance of a data group

sum of the squares of the deviations of the data from a reference level

Note 1 to entry: The reference level may be defined by one or more parameters, for example a mean value (one parameter) or a line (two parameters, slope and intercept), divided by the number of degrees of freedom.

3.1.9

central second moment of a data group

sum of the squares of the differences between the data values and the value of the group mean, divided by the number of data in the group

3.1.10

covariance of data groups

for two groups of data with equal numbers of elements where each element in one group corresponds to one in the other, the sum of the products of the deviations of the corresponding members from their group means, divided by the number of degrees of freedom

3.1.11

regression analysis

process of deducing the best-fit line expressing the relation of corresponding members of two data groups by minimizing the sum of squares of deviations of members of one of the groups from the line

Note 1 to entry: The parameters are referred to as the regression coefficients.

3.1.12

correlation coefficient

number expressing the completeness of the relation between members of two data groups, equal to the covariance divided by the square root of the product of the variances of the groups

Note 1 to entry: The value of its square is between 0 (no correlation) and 1 (complete correlation).

3.1.13

data sub-group

single set of data which may be used with other sub-groups to form a compound group

3.2 Symbols

Symbol	Definition
a, b	Regression coefficient
e	Sample (point) estimate of e
e_1	Lower confidence limit of e
e_2	Upper confidence limit of e
f	Number of degrees of freedom
$f(x)$	Probability density
$f_1(t), f_3(t)$	Arbitrary function of time
$f_2(\theta)$	Arbitrary function of stress
$f_4(p)$	Arbitrary function of property
F	Fisher-distributed stochastic variable
$F(x)$	Cumulative probability distribution
i	Order No. of partial sample
j	Order No. of specimen in partial sample No. i
k	Number of partial samples in total sample
m	Order No. of specimen
n	Number of observations in sample
n_i	Number of specimens in partial sample No. i
N	Total number of specimens
p	Arbitrary property of test specimens
$P(X \leq x)$	Probability that $X \leq x$
s^2	Sample variance
s_1^2	Variance within sets
s_2^2	Variance about regression line
s_{11}^2	Partial sample variance
t	Student-distributed stochastic variable
u	Standardized normal (Gaussian) distributed stochastic variable
x	Independent variable (for example $1/\theta$)
x_i	Partial sample value of x
$\bar{x}$	Sample mean
$\bar{x}$	Weighted mean of x
$\tilde{x}$	Sample median
X	Stochastic variable, specified value of x
y	Dependent stochastic variable (for example $\log v$)
y_{ij}	Individual specimen value of y
$\bar{y}_i$	Partial sample mean of y
$\bar{y}$	Total sample mean of y
Y	Specified value of y
α	Significance level
ε	Arbitrary population parameter
Θ	Thermodynamic temperature/Kelvin
θ	Ageing stress (for example temperature)

Symbol	Definition
$\bar{x}$	Mean value of X
$\tilde{x}$	Median value of X
σ	Standard deviation of X
σ^2	Variance of X
χ^2	Chi-square-distributed test variable
$1 - \alpha$	Confidence level

4 Calculation procedures

4.1 General considerations

For these calculations:

- n is the number of values known in subgroup;
- m is the total number in subgroup (= n for complete data group);
- α , β , μ and ε are the “Saw” coefficients for these values of m and n .

For an uncensored subgroup, the values of the “Saw” coefficients are as follows:

$$\alpha = 1/(n-1) \quad (1)$$

$$\beta = -1/(n(n-1)) \quad (2)$$

$$\mu = 1 - 1/n \quad (3)$$

$$\varepsilon = 1 \quad (4)$$

If convenient, these coefficients may be used to calculate the primary statistical functions (mean and standard deviation) of complete data groups, using the formulae of 4.2.3 (in place of the usual formulae as in 4.2.2). “Saw” coefficients are given in Table B.1.

4.2 Single sub-group – Difference of mean and specified value

4.2.1 General

The purpose of the procedure is to test the significance of the difference between the sub-group mean and a specified numerical value.

4.2.2 Complete data sub-group

Calculate sub-group mean
$$\bar{y} = \sum_{i=1}^n y_i / n \quad (5)$$

Calculate sub-group variance
$$\sigma^2 = \frac{\left(\sum_{i=1}^n y_i^2 - n \bar{y}^2 \right)}{(n-1)} \quad (6)$$

Calculate t
$$t = \bar{y} / \sqrt{\sigma^2 / n} \quad (7)$$

Compare the value of t with the tabulated t values.

4.2.3 Censored data sub-group

Calculate sub-group mean
$$\bar{y} = (1 - \mu)y_n + \mu \sum_{j=1}^{n-1} \frac{y_j}{(n-1)} \quad (8)$$

Calculate sub-group variance
$$\sigma^2 = \alpha \sum_{j=1}^{n-1} (y_n - y_j)^2 + \beta \left[\sum_{j=1}^{n-1} (y_n - y_j) \right]^2 \quad (9)$$

Calculate adjustment for t
$$a = \frac{(1 - n/m)}{(6,2 + n/6,4 - (m - n)/10,7)} \quad (10)$$

Calculate t
$$t = \bar{y} / \sqrt{\varepsilon \sigma^2 / n} \quad (11)$$

Calculate t_a
$$1/t_a = 1/t + a \quad (12)$$

Compare the value of t_a with the tabulated t values.

4.3 Two subgroups – Difference of means

4.3.1 General

The purpose of this procedure is to test the significance of the difference between the sub-group means.

For these calculations:

- n_i is the number of values known in subgroup i ;
- m_i is the total number of values in subgroup i ;
- $\alpha_i \beta_i \mu_i$ and ε_i are the “Saw” coefficients for these values of m and n .

For a complete sub-group, $\varepsilon_i = 1$.

4.3.2 Both sub-groups complete

Calculate sub-group means
$$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i \quad (13)$$

Calculate sub-group variances
$$\sigma_i^2 = \frac{\left(\sum_{j=1}^{n_i} y_{ij}^2 - n \bar{y}_i^2 \right)}{(n_i - 1)} \quad (14)$$

Calculate the group value of ε

$$e = \left(\frac{\varepsilon_1}{n_1} + \frac{\varepsilon_2}{n_2} \right) \quad (15)$$

Calculate the merged variance σ^2

$$\sigma^2 = \frac{((n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2)}{(n_1 + n_2 - 2)} \quad (16)$$

Calculate t

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{e\sigma^2}} \quad (17)$$

Determine probability by reference to tabulated values of t .

4.3.3 One or both subgroups censored

Calculate sub-group variances

$$\sigma_i^2 = \alpha_i \sum_{j=1}^{n_i-1} (y_{in_i} - y_{ij})^2 + \beta_i \left[\sum_{j=1}^{n_i-1} (y_{in_i} - y_{ij}) \right]^2 \quad (18)$$

Calculate sub-group means

$$\bar{y}_i = (1 - \mu_i)y_{in_i} + \mu_i \sum_{j=1}^{n_i-1} \frac{y_{ij}}{(n_i - 1)} \quad (19)$$

Calculate the group value of ε

$$e = \left(\frac{\varepsilon_1}{n_1} + \frac{\varepsilon_2}{n_2} \right) \quad (20)$$

Calculate the merged variance σ^2

$$\sigma^2 = \frac{((n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2)}{(n_1 + n_2 - 2)} \quad (21)$$

Calculate t

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{e\sigma^2}} \quad (22)$$

Calculate adjustment, a

$$a = \frac{p}{(n_1 + n_2)^2} \left[\frac{n_1}{m_1} - \frac{n_2}{m_2} \right] \left[\left(\frac{(n_1 + n_2)}{20} \right)^2 + 2 \right] \quad (23)$$

where p is the smaller of n_1 and n_2 .

Apply adjustment

$$t_a = \frac{1}{\left(\frac{1}{t} + a \right)} \quad (24)$$

Determine probability by reference to tabulated values of t .

4.4 Two or more subgroups – Analysis of variance

Individual sub-groups may be complete or censored.

For these calculations:

- n_i is the number of values known in subgroup i ;
 m_i is the total number in subgroup i ;
 α_i, β_i, μ_i and ε_i are the “Saw” coefficients for these values of m and n ;
 k is the number of subgroups;
 c is the intermediate value for χ^2 calculation;
 A is the adjustment factor for χ^2 calculation.

Calculate the total number of values
$$M = \sum_{i=1}^k m_i \quad (25)$$

Calculate the total number of values known
$$N = \sum_{i=1}^k n_i \quad (26)$$

Calculate subgroup means:

$$\bar{y}_i = (1 - \mu_i) y_{in_i} + \mu_i \sum_{j=1}^{n_i-1} \frac{y_{ij}}{(n_i - 1)} \quad (\text{Censored data})$$

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \quad (\text{Complete data subgroup}) \quad (27)$$

Calculate group general mean
$$\bar{\bar{y}} = \frac{\sum_{i=1}^k n_i \bar{y}_i}{N} \quad (28)$$

Calculate sub-group variances:

$$s_i^2 = \alpha_i \sum_{j=1}^{n_i-1} (y_{in_i} - y_{ij})^2 + \beta_i \left[\sum_{j=1}^{n_i-1} (y_{in_i} - y_{ij}) \right]^2$$

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij}^2 - n_i \bar{y}_i^2)}{(n_i - 1)} \quad (\text{Complete data subgroup}) \quad (29)$$

Calculate mean variance factor
$$\varepsilon = \frac{\sum_{i=1}^k \varepsilon_i}{k} \quad (30)$$

Calculate variance of means

$$s_N^2 = \frac{\left[\sum_{i=1}^k n_i \bar{y}_i^2 - N \bar{\bar{y}}^2 \right]}{(k-1)} \quad (31)$$

Calculate residual variance

$$s_D^2 = \frac{\varepsilon \left[\sum_{i=1}^k s_i^2 (n_i - 1) \right]}{(N - k)} \quad (32)$$

Test equality of subgroup variances:

Calculate c

$$c = 1 + \frac{\left[\sum_{i=1}^k \frac{1}{(n_i - 1)} - \frac{1}{(N - k)} \right]}{3(k-1)} \quad (33)$$

Calculate adjustment factor

$$A = 1 + \frac{\left(1 - \frac{N}{M} \right) \times \left(1 - \frac{12}{M} \right)}{2} \quad (34)$$

Calculate χ^2

$$\chi^2 = \frac{A}{c} \left[(N - k) \ln \left(\frac{s_D^2}{\varepsilon} \right) - \sum_{i=1}^k (n_i - 1) \ln(s_i^2) \right] \quad (35)$$

Test equality of residual variance and variance of subgroup means.

Calculate F

$$F = \frac{s_N^2}{s_D^2} \quad (36)$$

Degrees of freedom for F $N - k$ (denominator), $k - 1$ (numerator)

Calculate significance of general mean:

Calculate t

$$t = \bar{\bar{y}} \sqrt{\frac{N}{s_T^2}} \quad (37)$$

Adjust t for censoring

$$1/t_a = 1/t + a \quad (38)$$

Determine probability by reference to tabulated values of t with $N-1$ degrees of freedom.

4.5 Three or more subgroups – Regression analysis

4.5.1 Regression analysis – General considerations

These data differ from those of (4.4) in that the y -values in each subgroup are associated with a value of another variable, referred to in this section as x_i . The objective of the analysis is to determine whether there is a linear relationship between x and y and, if so, its parameters and properties.

The parameters and properties in question are as follows:

- slope(b) and intercept (a) of regression line;
- equality of variance of subgroups (χ^2);
- linearity of regression (F);
- confidence intervals of regression estimates.

For these calculations:

- n_i is the number of values known in subgroup i ;
- m_i is the total number in subgroup i ;
- α_i, β_i, μ_i and ε_i are the “Saw” coefficients for these values of m and n ;
- k is the number of subgroups;
- c is the intermediate value for χ^2 calculation;
- A is the adjustment factor for χ^2 calculation;
- b and a are the slope and intercept of the regression line;
- $t_{p,n-1}$ is the tabulated value of t for probability p and $n-1$ degrees of freedom.

Sub-groups may be either complete or censored. Values of y_{ij} are the actual values of variable j in subgroup i .

4.5.2 Calculations

Calculate the total number of values $M = \sum_{i=1}^k m_i$ (39)

Calculate the total number of values known $N = \sum_{i=1}^k n_i$ (40)

Calculate subgroup means:

$$\bar{y}_i = (1 - \mu_i) y_{in_i} + \mu_i \sum_{j=1}^{n_i-1} \frac{y_{ij}}{(n_i - 1)} \quad (\text{Censored data})$$

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \quad (\text{Complete data subgroup})$$
(41)

Calculate sub-group variances:

$$s_i^2 = \alpha_i \sum_{j=1}^{n_i-1} (y_{in_i} - y_{ij})^2 + \beta_i \left[\sum_{j=1}^{n_i-1} (y_{in_i} - y_{ij}) \right]^2 \quad (\text{Censored data})$$

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij}^2 - n_i \bar{y}_i^2)}{(n_i - 1)} \quad (\text{Complete data subgroup})$$
(42)

Calculate x -mean

$$\bar{x} = \frac{\sum_{i=1}^k n_i x_i}{N} \quad (43)$$

Calculate group general mean

$$\bar{y} = \frac{\sum_{i=1}^k n_i \bar{y}_i}{N} \quad (44)$$

Calculate mean variance factor

$$\varepsilon = \frac{\sum_{i=1}^k \varepsilon_i}{k} \quad (45)$$

Calculate residual variance

$$s_D^2 = \frac{\varepsilon \left[\sum_{i=1}^k s_i^2 (n_i - 1) \right]}{(N - k)} \quad (46)$$

Calculate y -sum of squares

$$SSy = \sum_{i=1}^k n_i \bar{y}_i^2 - N \bar{y}^2 \quad (47)$$

Calculate x -sum of squares

$$SSx = \sum_{i=1}^k n_i x_i^2 - N \bar{x}^2 \quad (48)$$

Calculate xy -sum of products

$$SPxy = \sum_{i=1}^k n_i x_i \bar{y}_i - N \bar{x} \bar{y} \quad (49)$$

Calculate slope b

$$b = \frac{SPxy}{SSx} \quad (50)$$

Calculate intercept a

$$a = \bar{y} - b \bar{x} \quad (51)$$

Calculate r^2

$$r^2 = \left(\frac{SPxy}{SSx} \right) \left(\frac{SPxy}{SSy} \right) \quad (52)$$

(r is the correlation coefficient)

Calculate non-linearity variance

$$s_N^2 = \frac{(1 - r^2) SSy}{(k - 2)} \quad (53)$$

4.5.3 Test equality of subgroup variances

Calculate c

$$c = 1 + \frac{\left[\sum_{i=1}^k \frac{1}{(n_i - 1)} - \frac{1}{(N - k)} \right]}{3(k - 1)} \quad (54)$$

Calculate adjustment factor

$$A = 1 + \frac{\left(1 - \frac{N}{M}\right) \times \left(1 - \frac{12}{M}\right)}{2} \quad (55)$$

Calculate χ^2

$$\chi^2 = \frac{A}{c} \left[(N - k) \ln \left(\frac{s_D^2}{\varepsilon} \right) - \sum_{i=1}^k (n_i - 1) \ln(s_i^2) \right] \quad (56)$$

4.5.4 Test significance of deviations from linearity

Calculate F

$$F = \frac{s_N^2}{s_D^2} \quad (57)$$

Degrees of freedom for F $N - k$ (denominator), $k - 2$ (numerator)

4.5.5 Estimate and confidence limit of y

Central second moment of x

$$\mu_{2(x)} = \frac{SSx}{N} \quad (58)$$

Non regression variance

$$s_T^2 = \frac{((N - k)s_D^2 + (k - 2)s_N^2)}{(N - 2)} \quad (59)$$

Adjust for extrapolation

$$s_C^2 = \frac{s_T^2}{N} \left(1 + \frac{(X - \bar{x})^2}{\mu_{2(x)}} \right) \quad (60)$$

Correction to t for censoring

$$a = \frac{\left(1 - \frac{N}{M}\right)}{\left(6,2 + \frac{N}{6,4} - \frac{(M - N)}{10,7}\right)}, \quad (61)$$

Correction applied

$$t_c = \left(\frac{1}{t_{p, N-2}} - a \right) \quad (62)$$

Estimate of y

$$\hat{y} = a + bX \quad (63)$$

Confidence limit of estimate of y

$$\hat{y}_c = \hat{y} + t_c \sqrt{s_c^2} \quad (64)$$

4.5.6 Estimate and confidence limit of x

Estimate

$$\hat{x} = \frac{(y - a)}{b} \quad (65)$$

For simplicity, calculate several temporary variables:

$$b_r = b - \frac{t_c^2 s_T^2}{Nb \mu_{2,(x)}} \quad (66)$$

$$s_r^2 = \frac{s_T^2}{N} \left(\frac{b_r}{b} + \frac{(\hat{x} - \bar{x})^2}{\mu_{2,(x)}} \right) \quad (67)$$

Confidence limit of estimate of x

$$\hat{x}_c = \bar{x} + \frac{(y - \bar{y})}{b_r} + \frac{t_c \sqrt{s_r^2}}{b_r} \quad (68)$$

Annex A (informative)

Statistical background

A.1 Statistical distributions and parameters

The distribution of a stochastic variable X is described by the distribution function:

$$F(x) = P(X \leq x) \quad (\text{A.1})$$

where $P(X \leq x)$ is the probability that the value of X is $\leq x$. Here $F(x)$ goes from 0 to 1 and is a never-decreasing function of x . If $F(x)$ is a continuous function of x , then the probability density is determined as:

$$f(x) = \frac{dF(x)}{dx} \quad (\text{A.2})$$

The distribution may be characterized by parameters, of which the most important are:

- the mean value:

$$\xi = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{A.3})$$

- the variance:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \xi)^2 f(x) dx \quad (\text{A.4})$$

The square root of the variance is termed the standard deviation σ .

A.2 Estimates of parameters

From a sample of n stochastic independent specimens from a population, estimates of the parameters of the population (see Clause A.1) may be derived.

An estimate of the mean value of the population (Formula (A.3)) is calculated as the average of the individual sample values:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{A.5})$$

where

x_i represents the individual sample values ($i = 1, 2, \dots, n$).

An estimate of the variance of the population (Formula (A.4)) is the sample variance.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)} \quad (\text{A.6})$$

where $n - 1 = f$ is called the number of degrees of freedom of s^2 .

A.3 Distribution types

A.3.1 General

The following distribution types are relevant to this application, the t , F , and χ^2 distributions being the distributions of secondary functions derived from the mean and variance parameter estimates of normally distributed data.

A.3.2 The normal distribution

The calculation processes specified in this standard are based on the assumption that the data under examination are normally distributed. No test for normality of the data is specified, since the available tests are unreliable for small sample groups of data. However, the methods have been used for a considerable time without undesirable results and with no check on the normality of the data distributions.

The normal (Gaussian) distribution is defined by:

$$f(x) = \frac{\exp\left\{-\frac{(x - \xi)^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}} \quad (\text{A.7})$$

and is completely characterized by its mean value ξ and variance σ^2 .

The standardized normal distribution:

$$f(u) = \frac{\exp\left\{-\frac{u^2}{2}\right\}}{\sqrt{2\pi}} \quad (\text{A.8})$$

where

$$u = \frac{x - \xi}{\sigma} \quad (\text{A.9})$$

and the corresponding distribution function $F(u)$ have been tabulated and computer routines for their calculation are available (see [1]).

The above use of F should not be confused with the F distribution below.

The mean value $\bar{x}$ of a sample of n specimens from a normal distribution is itself a normally distributed stochastic variable with mean value $\xi_{\bar{x}} = \xi$ and variance $\sigma_{\bar{x}}^2 = \sigma^2/n$ and the corresponding standardized variable is:

$$u = \frac{x - \xi}{\frac{\sigma}{\sqrt{n}}} \quad (\text{A.10})$$

A.3.3 The t distribution

If the true variance of the normal distribution σ^2 is not known, the sample estimate s^2 from Formula (A.6) may be substituted and the standardized sample mean value becomes:

$$u = \frac{x - \xi}{\frac{s}{\sqrt{n}}} \quad (\text{A.11})$$

The distribution of this variable is called the t distribution (or Student's t) and depends on the parameter $f = n - 1$ (the number of degrees of freedom for s^2). The t distribution has been tabulated for different values of f . It is derived from the "Incomplete Beta function".

A.3.4 The F distribution

To test if two sample variances, determined from two different samples, may reasonably be considered to be estimates of the same theoretical variance (population parameter), the following test variable is calculated:

$$F = \frac{s_1^2}{s_2^2} \quad (\text{A.12})$$

The distribution of this variable is called the F distribution (or Fisher) and depends on the parameters $f_1 = n_1 - 1$ and $f_2 = n_2 - 1$ (the number of degrees of freedom for s_1^2 and s_2^2). The F distribution has been tabulated for different values of f_1 and f_2 . It is derived from the "Incomplete Beta function".

A.3.5 The χ^2 distribution

To test if several sample variances, each determined from a different sample, may reasonably be considered to be estimates of the same theoretical variance, the following test variable is calculated (Bartlett's χ^2).

$$\chi^2 = \frac{2,3 \left(f \lg s^2 - \sum_{i=1}^k f_i \lg s_i^2 \right)}{c} \quad (\text{A.13})$$

where

$$c = 1 + \frac{\left[\sum_{i=1}^k \frac{1}{f_i} \right] - \frac{1}{f}}{3(k-1)} \quad (\text{A.14})$$

k is the number of variances, s_i^2 the individual sample variance ($i = 1, 2, \dots k$) with f_i degrees of freedom, and $s^2 = \frac{\sum f_i s_i^2}{\sum f_i}$ is a pooled variance with $f = \sum f_i$ degrees of freedom. The test hypothesis is that all k variances s_i^2 are estimates of the same theoretical variance σ^2 .

The calculated value χ^2 is compared with the tabulated value $\chi^2 (1 - \alpha, k - 1)$ which is a function of $k - 1$, the number of degrees of freedom for χ^2 and of α , the significance level. If $\chi^2 > \chi^2 (1 - \alpha, k - 1)$, the hypothesis is rejected on significance level α .

The distribution of this variable is called the Bartlett's χ^2 distribution and depends on the parameter $f = k - 1$. The χ^2 distribution has been tabulated for different values of f . It is derived from the "Incomplete Gamma function".

Bartlett's test is an approximate test, but a good approximation if the number of degrees of freedom f_i of all the individual sample variances s_i^2 is greater than 2.

If the hypothesis is accepted, s^2 is taken as a pooled estimate of the common variance with f degrees of freedom.