

---

---

**Sensory analysis — Methodology —  
Balanced incomplete block designs**

*Analyse sensorielle — Méthodologie — Plans de présentation en blocs  
incomplets équilibrés*

iTeh Standards  
(<https://standards.iteh.ai>)  
Document Preview

[ISO 29842:2011](https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011)

<https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011>



iTeh Standards  
(<https://standards.iteh.ai>)  
Document Preview

ISO 29842:2011

<https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011>



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2011

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](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

**Contents**

Page

<b>Foreword</b> .....	<b>iv</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>1</b>
<b>4 Specification of balanced incomplete block designs</b> .....	<b>1</b>
<b>5 Data analysis</b> .....	<b>3</b>
<b>5.1 General</b> .....	<b>3</b>
<b>5.2 Analysis of variance for rating data</b> .....	<b>3</b>
<b>5.3 Friedman's sum rank analysis for rank data</b> .....	<b>5</b>
<b>6 Application in sensory evaluation</b> .....	<b>6</b>
<b>Annex A (informative) Catalogue of incomplete block designs</b> .....	<b>7</b>
<b>Annex B (informative) Example of balanced incomplete block design with ratings data</b> .....	<b>15</b>
<b>Annex C (informative) Example of balanced incomplete block design with rank data</b> .....	<b>17</b>
<b>Bibliography</b> .....	<b>19</b>

iTech Standards  
(<https://standards.iteh.ai>)  
Document Preview

[ISO 29842:2011](https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011)

<https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011>

## 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 29842 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 12, *Sensory analysis*.

# iTeh Standards (<https://standards.iteh.ai>) Document Preview

[ISO 29842:2011](https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011)

<https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011>

# Sensory analysis — Methodology — Balanced incomplete block designs

## 1 Scope

This International Standard specifies a method for the application of balanced incomplete block designs to sensory descriptive and hedonic tests.

This International Standard applies when the number of test samples exceeds the number of evaluations that an assessor can perform reliably in a single session.

This International Standard also specifies the fundamental characteristics of balanced incomplete block designs and establishes guidelines for their application in sensory evaluation.

## 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 5492, *Sensory analysis — Vocabulary* [ISO 29842:2011](https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011)

<https://standards.iteh.ai/catalog/standards/iso/89c25e35-4a1e-4a50-ae2f-91e9dda1d1e6/iso-29842-2011>

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 5492, ISO 3534-1, and the following apply.

### 3.1

#### **block design**

⟨sensory analysis⟩ multi-sample serving protocol in which an assessor evaluates all or a subset of the samples in a study

### 3.2

#### **repetition**

one occurrence of an experimental design

## 4 Specification of balanced incomplete block designs

Balanced incomplete block (BIB) designs apply to sensory tests in which the total number of samples is greater than the number that can be evaluated before sensory and psychological fatigue set in. In BIB designs, each assessor evaluates only a subset of the total number of samples in a single session.

An example of a BIB design is shown in Table 1.

Table 1 — A BIB design with five samples and 10 block/assessors

Block (assessor)	Test sample				
	1	2	3	4	5
1	×	×	×	—	—
2	×	×	—	×	—
3	×	×	—	—	×
4	×	—	×	×	—
5	×	—	×	—	×
6	×	—	—	×	×
7	—	×	×	×	—
8	—	×	×	—	×
9	—	×	—	×	×
10	—	—	×	×	×

In a BIB design each assessor evaluates a subset,  $k$ , of the total number of samples,  $t$ , where  $k < t$ . The subset of samples that an assessor evaluates is selected so that, in a single repetition of the BIB design, every sample is evaluated an equal number of times and all possible pairs of two samples are evaluated by an equal number of assessors.

The notation most commonly used in a BIB design follows.

$t$  number of test samples

$k$  number of samples evaluated by an assessor in a single session ( $k < t$ )

$b$  total number of blocks (typically, assessors) in one repetition of the BIB design

$r$  number of times each test sample is evaluated in one repetition of the BIB design

$\lambda$  number of times each pair of samples is evaluated by the same assessor

$p$  number of times the basic BIB design is repeated

Notationally, each assessor evaluates  $k$  of the  $t$  samples ( $k < t$ ). The subset of  $k$  samples that an assessor evaluates is selected so that in a single repetition of the BIB design every sample is evaluated an equal number of times and all possible pairs of samples are evaluated by an equal number of assessors. The number of blocks (assessors) required to complete a single repetition of the BIB design is denoted by  $b$ . The number of times each sample is evaluated in a single repetition of the BIB design is denoted by  $r$  and the number of times every pair of two samples is evaluated together is denoted by  $\lambda$ .

The entire BIB design may need to be repeated several times in order to achieve an adequate level of precision for the study. The number of repetitions of the basic BIB design is denoted by  $p$ . The total number of blocks (typically assessors) is then  $p \cdot b$  and the total of evaluations per sample is then  $p \cdot r$ . The total number of times each pair of samples is seen together is  $p \cdot \lambda$ .

The constant values of  $r$  and  $\lambda$  for all samples in the BIB design imparts important statistical properties to data collected from the design. The constant value of  $r$  ensures that the mean values of all of the samples are estimated with equal precision. The constant value of  $\lambda$  ensures that all pair-wise comparisons between any two samples are equally sensitive.

## 5 Data analysis

### 5.1 General

Two types of data can be collected using balanced incomplete block designs. Ratings data, or scores, are obtained when assessors use a scale to report the perceived intensities of the attributes or impressions they are evaluating. Rank data are obtained when assessors order the samples from lowest to highest (or vice versa) relative to the attribute they are evaluating. Different data analysis methods are used for ratings and rank data.

### 5.2 Analysis of variance for rating data

Analysis of variance (ANOVA) is used to analyse ratings data obtained from the BIB design. The sources of variability accounted for in the ANOVA model for the BIB design are the same as those accounted for in a randomized (complete) block design. In both cases, the total variability is partitioned into the separate effects of blocks (typically assessors), treatments (typically samples) and errors. Because each assessor evaluates only a subset of the total number of test samples, more complicated formulae are required to calculate the ANOVA sum-of-squares for the BIB design than for the randomized (complete) block design. The sensory analyst shall ensure that the program used to perform the analysis is capable of handling BIB designs. In many statistical computer packages, the ANOVA procedure applies only to complete designs, i.e. studies in which every assessor evaluates all of the test samples. For incomplete designs, such as BIB designs, the general linear model (GLM) procedure or a mixed model procedure is required.

The form of the ANOVA used to analyse BIB data depends on how the design is administered.

Where the experiment is of the form of the example in Table 1, with a single repetition of the design, the ANOVA table is as shown in Table 2.

**Table 2 — ANOVA table for balanced incomplete block design (single repetition)**

Source of variation	Degrees of freedom (DF)	Sum of squares (SS)	Mean square (MS)	<i>F</i>
Total	$\nu_T = t^*r - 1$	$S_T$		
Assessors	$\nu_B = b - 1$	$S_B$		
Samples (adjusted for assessors)	$\nu_S = t - 1$	$S_S$	$MS_S = S_S/\nu_S$	$MS_S/MS_E$
Error	$\nu_E = t^*r - t - b + 1$	$S_E$	$MS_E = S_E/\nu_E$	

If the *F*-statistic in Table 2 exceeds the upper- $\alpha$  critical value of an *F* with the corresponding degrees of freedom, then the null hypothesis assumption of equivalent mean ratings is rejected. If the *F*-statistic is significant, a multiple comparison procedure, such as Fisher's LSD (least significant difference), *L*, shall be applied to determine which samples are significantly different from one another. The equation for Fisher's LSD, *L*, appropriate for a single repetition of this BIB design is:

$$L = t_{\alpha/2, \nu_E} \sqrt{\frac{2MS_E}{r} \sqrt{\frac{k(t-1)}{(k-1)t}}}$$

where

*t*, *k* and *r* are as defined in Clause 4;

$MS_E$  is the mean square for error from the ANOVA table;

$\nu_E$  is the number of degrees of freedom for error from the ANOVA table;

$t_{\alpha/2, \nu_E}$  is the upper  $\alpha/2$  critical value of Student's *t*-distribution with  $\nu_E$  degrees of freedom.

The same value of  $\alpha$  shall be used for assessing the significance of the  $F$ -statistic and in Fisher's LSD,  $L$ .

The BIB design shall be repeated  $p$  times to achieve an adequate level of precision from the study. If the total number of blocks is too large for each assessor to evaluate all of them, each of the  $p*b$  assessors shall evaluate only one block of  $k$  samples. Within each block, the order in which the  $k$  samples are evaluated shall be done at random. The ANOVA table for this design is presented in Table 3.

**Table 3 — ANOVA table for balanced incomplete block design**  
( $p$  repetitions performed by  $p*b$  assessors each evaluating a single block of  $k$  samples)

Source of variation	Degrees of freedom (DF)	Sum of squares (SS)	Mean square (MS)	$F$
Total	$\nu_T = t*p*r - 1$	$S_T$		
Blocks (assessors)	$\nu_B = p*b - 1$	$S_B$		
Samples (adjusted for assessors)	$\nu_S = t - 1$	$S_S$	$MS_S = S_S/\nu_S$	$MS_S/MS_E$
Error	$\nu_E = t*p*r - t - p*b + 1$	$S_E$	$MS_E = S_E/\nu_E$	

If the  $F$ -statistic in Table 3 exceeds the critical value of an  $F$  with the corresponding degrees of freedom, then the null hypothesis assumption of equivalent mean ratings is rejected. If the  $F$ -statistic is significant, a multiple comparison procedure, such as Fisher's LSD,  $L$ , shall be applied to determine which samples are significantly different from one another. The equation for Fisher's LSD,  $L$ , appropriate for a BIB design is:

$$L = t_{\alpha/2, \nu_E} \sqrt{\frac{2MS_E}{pr} \sqrt{\frac{k(t-1)}{(k-1)t}}}$$

where

$t, k, p$  and  $r$  are as defined in Clause 4;

$MS_E$  is the mean square for error from the ANOVA table;

$\nu_E$  is the number of degrees of freedom for error from the ANOVA table;

$t_{\alpha/2, \nu_E}$  is the upper  $\alpha/2$  critical value of Student's  $t$ -distribution with  $\nu_E$  degrees of freedom.

The same value of  $\alpha$  shall be used for assessing the significance of the  $F$ -statistic and in Fisher's LSD,  $L$ .

If each assessor evaluates all  $b$  blocks in the BIB design, then the “assessor effect” and the “assessor-by-sample” interaction can be partitioned out of the total variability (see Table 4). This approach is especially applicable when the total number of blocks in one repetition of the BIB design is small (e.g.  $b \leq 6$ ). The order in which the blocks are presented to the assessor shall be done at random. Within each block, the order in which the samples are evaluated shall be done at random. In either approach, the variability that arises from the assessors is accounted for and the interaction between assessors and samples replaces the error term that was used in Tables 2 and 3.



**Table 4 — ANOVA table for balanced incomplete block design**  
( $p$  repetitions performed by  $p$  assessors each evaluating  $b$  blocks of  $k$  samples)

Source of variation	Degrees of freedom (DF)	Sum of squares (SS)	Mean square (MS)	$F$
Total	$\nu_T = t^*p^*r - 1$	$S_T$		
Assessor	$\nu_P = p - 1$	$S_P$		
Blocks (within assessor)	$\nu_{B(P)} = p^*(b - 1)$	$S_{B(P)}$		
Samples (adjusted for assessor)	$\nu_S = t - 1$	$S_S$	$MS_S = S_S/\nu_S$	$MS_S/MS_{A^*S}$
Assessor* <i>samples</i>	$\nu_{A^*S} = (p - 1)(t - 1)$	$S_{A^*S}$	$MS_{A^*S} = S_{A^*S}/\nu_{A^*S}$	
Residual	$\nu_E = p^*(t^*r - t - b + 1)$	$S_E$	$MS_E = S_E/\nu_E$	

If the  $F$ -statistic in Table 4 exceeds the critical value of an  $F$  with the corresponding degrees of freedom, then the null hypothesis assumption of equivalent mean ratings is rejected. If the  $F$ -statistic is significant, a multiple comparison procedure, such as Fisher's LSD,  $L$ , shall be applied to determine which samples are significantly different from one another. The equation for Fisher's LSD,  $L$ , appropriate for a BIB design is:

$$L = t_{\alpha/2, \nu_{A^*S}} \sqrt{\frac{2MS_{A^*S}}{pr}} \sqrt{\frac{k(t-1)}{(k-1)t}}$$

where

$t$ ,  $k$ ,  $p$  and  $r$  are as defined in Clause 4;

$MS_{A^*S}$  is the mean square for the assessor\*sample interaction from the ANOVA table;

$\nu_{A^*S}$  is the number of degrees of freedom for the assessor\*sample interaction from the ANOVA table;

$t_{\alpha/2, \nu_{A^*S}}$  is the upper  $\alpha/2$  critical value of Student's  $t$ -distribution with  $\nu_{A^*S}$  degrees of freedom.

The same value of  $\alpha$  shall be used for assessing the significance of the  $F$ -statistic and in Fisher's LSD,  $L$ .

### 5.3 Friedman's sum rank analysis for rank data<sup>1)</sup>

A Friedman-type statistic shall be applied to rank data arising from a BIB design. Friedman's test statistic,  $F_{\text{test}}$ , is given by:

$$F_{\text{test}} = \frac{12}{p\lambda t(k+1)} \sum_{j=1}^t R_j^2 - \frac{3(k+1)pr^2}{\lambda}$$

where  $t$ ,  $k$ ,  $r$ ,  $\lambda$  and  $p$  are as defined above and  $R_j$  is the rank sum of the  $j$ th sample (Reference [8]). Tables of critical values of  $F_{\text{test}}$  are available for selected combinations of  $t = 3 \dots 6$ ,  $k = 2 \dots 5$ , and  $p = 1 \dots 7$  (Reference [9]). However, in most sensory studies, the total number of blocks exceeds the values in the tables. For these situations, the test procedure is to reject the assumption of equivalency among the samples if the value of  $F_{\text{test}}$  exceeds the upper  $\alpha$  critical value of a  $\chi^2$ -statistic with  $(t - 1)$  degrees of freedom.

1) There are several statistical methods for analysing rank data that are obtained from a BIB design. Interested readers are encouraged to review the statistical literature on the topic. Friedman's method has been chosen for detailed discussion because it is statistically powerful and computationally convenient.

If the  $\chi^2$ -statistic is significant, then a multiple comparison procedure shall be performed to determine which samples differ significantly from one another. The equation for Fisher's LSD,  $L$ , appropriate for a BIB design is:

$$L = z_{\alpha/2} \sqrt{\frac{p(k+1)(rk-r+\lambda)}{6}}$$

where

$p$ ,  $k$ ,  $r$  and  $\lambda$  are as defined in Clause 4;

$z_{\alpha/2}$  is the upper- $\alpha/2$  critical value of the standard normal distribution.

The same value of  $\alpha$  shall be used for assessing the significance of the  $F_{\text{test}}$ -statistic and in Fisher's LSD,  $L$ .

## 6 Application in sensory evaluation

The number of samples that are evaluated by an assessor in a single session,  $k$ , shall not exceed the assessor's ability to provide reliable ratings of the samples. The value of  $k$  depends on several factors including the overall intensity of the sensory characteristics of the samples, the degree of carryover effects (e.g. lingering aftertastes), and the number of attributes the assessor is rating. The sensory analyst shall limit the number of samples an assessor evaluates in a single session to avoid both sensory and psychological fatigue.

Sensory and psychological fatigue limit the number of evaluations an assessor can perform reliably. However, the sensory analyst shall make every attempt to have each assessor perform as many evaluations as possible in order to counteract the influence of context effects. Since each assessor evaluates only a subset of the total number of samples, the assessor is not exposed to the full range of variability present in the samples. This can lead the assessors to exaggerate the differences in the ratings that they assign to the products. The larger the proportion of samples the assessors evaluate in a single session, the greater the range of sensory differences they are exposed to and, therefore, the lesser the chance that context effects bias the results of the study.

The total number of evaluations of each sample,  $r \cdot p$ , shall be determined based on the level of sensitivity required of the test. The same criteria used to decide on the total number of evaluations in a complete block design shall be applied. For example, if it is standard practice to perform 12 evaluations per sample in a complete block study with a descriptive analysis panel and the number of replications in one repetition of the BIB design is  $r = 3$ , then the BIB design shall be replicated  $p = 4$  times in order to achieve the required total number of evaluations.