INTERNATIONAL STANDARD

ISO 29842

First edition
2011-07-15

Sensory analysis — Methodology — Balanced incomplete block designs

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

Reference number
ISO 29842:2011(E)

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

Contents

Page

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

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*.

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*

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

λ 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 λ.

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 $p*b$ and the total of evaluations per sample is then $p*r$. The total number of times each pair of samples is seen together is $p*\lambda$.

The constant values of r and λ 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 λ 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)	F
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/ν_S	MS_S/MS_E
Error	$\nu_E = t^*r - t - b + 1$	S_E	MS_E S_E/ν_E	

If the F-statistic in Table 2 exceeds the upper-α 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 diffference), 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;

ν_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 ν_E degrees of freedom.

The same value of α 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;

ν_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 ν_E degrees of freedom.

The same value of α 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*samples	$\nu_{A*S} = (p - 1)(t - 1)$	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;

ν_{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 ν_{A*S} degrees of freedom.

The same value of α 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[1)]

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

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

where t, k, r, λ and p are as defined above and R_j is the rank sum of the jth sample (Reference [8]). Tables of critical values of F_{test} are available for selected combinations of $t = 3 \ldots 6$, $k = 2 \ldots 5$, and $p = 1 \ldots 7$ (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_{test} exceeds the upper α critical value of a χ^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 χ^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 λ are as defined in Clause 4;

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

The same value of α shall be used for assessing the significance of the F_{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, $r*p$, 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.

Annex A
(informative)

Catalogue of incomplete block designs

The following BIB designs apply for $t = 3 \ldots 10$ and $k = 2 \ldots t - 1$ or $k = 2 \ldots 6$, whichever is lesser. It is possible to construct BIB designs for any values of t and k by collecting all possible combinations of t items in groups of size k. However, it is sometimes possible to construct BIB designs that have fewer blocks than are required to form all possible combinations. For any value of t and k, the catalogue presents the BIB design with the smallest number of blocks.

a) 2 of 3 BIB: All possible pairs of three samples ($t = 3$, $k = 2$, $b = 3$, $r = 2$, $\lambda = 1$).

b) 2 of 4 BIB: All possible pairs of four samples ($t = 4$, $k = 2$, $b = 6$, $r = 3$, $\lambda = 1$).

c) 3 of 4 BIB: All possible triads of four samples ($t = 4$, $k = 3$, $b = 4$, $r = 3$, $\lambda = 2$).

d) 2 of 5 BIB: All possible pairs of five samples ($t = 5$, $k = 2$, $b = 10$, $r = 4$, $\lambda = 1$).

e) 3 of 5 BIB: All possible triads of five samples ($t = 5$, $k = 3$, $b = 10$, $r = 6$, $\lambda = 3$).

f) 4 of 5 BIB: All possible four-tuples of five samples ($t = 5$, $k = 4$, $b = 5$, $r = 4$, $\lambda = 3$).

g) 2 of 6 BIB: All possible pairs of six samples ($t = 6$, $k = 2$, $b = 15$, $r = 5$, $\lambda = 1$).

h) 3 of 6 BIB: 10 Triads of six samples ($t = 6$, $k = 3$, $b = 10$, $r = 5$, $\lambda = 2$).

See Table A.1.

Table A.1 — 10 Triads of six samples

Block	Sample					
	1	2	3	4	5	6
1	×	×			×	
2	×	×				×
3	×		×	×		
4	×		×			×
5	×			×	×	
6		×	×	×		
7		×	×		×	
8		×		×		×
9			×		×	×
10				×	×	×

i) 4 of 6 BIB: All possible four-tuples of six samples ($t = 6$, $k = 4$, $b = 15$, $r = 10$, $\lambda = 6$).

j) 5 of 6 BIB: All possible five-tuples of six samples ($t = 6$, $k = 5$, $b = 6$, $r = 5$, $\lambda = 4$).

k) 2 of 7 BIB: All possible pairs of seven samples ($t = 7$, $k = 2$, $b = 21$, $r = 6$, $\lambda = 1$).

l) 3 of 7 BIB: Seven triads of seven samples ($t = 7$, $k = 3$, $b = 7$, $r = 3$, $\lambda = 1$).

See Table A.2.

Table A.2 — Seven triads of seven samples

Block	Sample						
	1	2	3	4	5	6	7
1	×	×			×		
2	×		×			×	
3	×			×			×
4		×	×				×
5		×		×		×	
6			×	×	×		
7					×	×	×

m) 4 of 7 BIB: 14 Four-tuples of seven samples (t = 7, k = 4, b = 14, r = 8, λ = 4).

See Table A.3.

Table A.3 — 14 Four-tuples of seven samples

Block	Sample						
	1	2	3	4	5	6	7
1	×	×	×		×		
2	×	×	×				×
3	×	×		×		×	
4	×	×				×	×
5	×		×	×	×		
6	×		×	×		×	
7	×			×	×		×
8	×				×	×	×
9		×	×	×			×
10		×	×		×	×	
11		×		×	×	×	
12		×		×	×		×
13			×	×		×	×
14			×		×	×	×

n) 5 of 7 BIB: All possible five-tuples of seven samples (t = 7, k = 5, b = 21, r = 15, λ = 10).

o) 6 of 7 BIB: All possible six-tuples of seven samples (t = 7, k = 6, b = 7, r = 6, λ = 5).

p) 2 of 8 BIB: All possible pairs of eight samples (t = 8, k = 2, b = 28, r = 7, λ = 1).

q) 3 of 8 BIB: All possible triads of eight samples (t = 8, k = 3, b = 56, r = 21, λ = 6).

r) 4 of 8 BIB: 14 Four-tuples of eight samples (t = 8, k = 4, b = 14, r = 7, λ = 3).

See Table A.4.

Table A.4 — 14 Four-tuples of eight samples

Block	Sample							
	1	2	3	4	5	6	7	8
1	×	×	×	×				
2	×	×			×	×		
3	×	×					×	×
4	×		×		×		×	
5	×		×			×		×
6	×			×	×			×
7	×			×		×	×	
8		×	×		×			×
9		×	×			×	×	
10		×		×	×		×	
11		×		×		×		×
12			×	×	×	×		
13			×	×			×	×
14					×	×	×	×

s) 5 of 8 BIB: All possible five-tuples of eight samples (t = 8, k = 5, b = 56, r = 35, λ = 20).

t) 6 of 8 BIB: All possible six-tuples of eight samples (t = 8, k = 6, b = 28, r = 21, λ = 15).

u) 2 of 9 BIB: All possible pairs of nine samples (t = 9, k = 2, b = 36, r = 8, λ = 1).

v) 3 of 9 BIB: 12 Triads of nine samples (t = 9, k = 3, b = 12, r = 4, λ = 1).

See Table A.5.

Table A.5 — 12 Triads of nine samples

Block	Sample								
	1	2	3	4	5	6	7	8	9
1	×	×	×						
2	×			×			×		
3	×				×				×
4	×					×		×	
5		×		×					×
6		×			×			×	
7		×				×	×		
8			×	×				×	
9			×		×		×		
10			×			×			×
11				×	×	×			
12							×	×	×

w) 4 of 9 BIB: 18 Four-tuples of nine samples (t = 9, k = 4, b = 18, r = 8, λ = 3).

See Table A.6.

Table A.6 — 18 Four-tuples of nine samples

Block	Sample								
	1	2	3	4	5	6	7	8	9
1	×	×	×	×					
2	×	×		×					×
3	×	×			×		×		
4	×		×			×		×	
5	×		×					×	×
6	×			×		×	×		
7	×				×	×			×
8	×				×		×	×	
9		×	×		×	×			
10		×	×			×	×		
11		×				×		×	×
12		×		×	×			×	
13		×					×	×	×
14			×	×	×			×	
15			×	×			×		×
16			×		×		×		×
17				×	×	×			×
18				×		×	×	×	

x) 5 of 9 BIB: 18 Five-tuples of nine samples ($t = 9$, $k = 5$, $b = 18$, $r = 10$, $\lambda = 5$).

See Table A.7.

Table A.7 — 18 Five-tuples of nine samples

Block	Sample								
	1	2	3	4	5	6	7	8	9
1	×	×	×		×				×
2	×	×	×				×	×	
3	×	×		×		×		×	
4	×	×			×	×		×	
5	×	×				×	×		×
6	×		×	×	×	×			
7	×		×	×	×		×		
8	×		×			×	×		×
9	×			×	×			×	×
10	×			×			×	×	×
11		×	×	×		×			×
12		×	×	×			×	×	
13		×	×		×			×	×
14		×		×	×	×	×		
15		×		×	×		×		×
16			×	×		×		×	×
17			×		×	×	×	×	
18					×	×	×	×	×

y) 6 of 9 BIB: 18 Six-tuples of nine samples (t = 9, k = 6, b = 18, r = 10, λ = 5).

See Table A.8.

Table A.8 — 18 Six-tuples of nine samples

Block	Sample								
	1	2	3	4	5	6	7	8	9
1	×	×		×	×		×	×	
2	×	×			×	×	×		×
3	×	×	×	×	×	×			
4	×	×	×				×	×	×
5	×	×		×		×		×	×
6	×		×	×	×			×	×
7	×		×	×		×	×		×
8	×		×		×	×	×	×	
9		×	×	×	×		×		×
10		×	×	×		×	×	×	
11		×	×		×	×		×	×
12				×	×	×	×	×	×

z) 2 of 10 BIB: All possible pairs of 10 samples (t = 10, k = 2, b = 45, r = 9, λ = 1).

aa) 3 of 10 BIB: 30 Triads of 10 samples (t = 10, k = 3, b = 30, r = 9, λ = 2).

See Table A.9.

Table A.9 — 30 Triads of 10 samples

Block	Sample									
	1	2	3	4	5	6	7	8	9	10
1	×	×	×							
2	×	×		×						
3	×		×		×					
4	×			×		×				
5	×				×		×			
6	×					×		×		
7	×						×		×	
8	×							×		×
9	×								×	×
10		×	×			×				
11		×		×						×
12		×			×			×		
13		×			×				×	
14		×				×	×			
15		×					×		×	
16		×						×		×
17			×	×			×			
18			×	×				×		
19			×		×	×				
20			×				×			×
21			×					×	×	
22			×						×	×
23				×	×				×	
24				×	×					×
25				×		×			×	
26				×			×	×		
27					×	×				×
28					×		×	×		
29						×	×			×
30						×		×	×	

bb) 4 of 10 BIB: 15 Four-tuples of 10 samples (t = 10, k = 4, b = 15, r = 6, λ = 2).

See Table A.10.

Table A.10 — 15 Four-tuples of 10 samples

Block	Sample									
	1	2	3	4	5	6	7	8	9	10
1	×	×	×	×						
2	×	×			×	×				
3	×		×				×	×		
4	×			×					×	×
5	×				×		×		×	
6	×					×		×		×
7		×	×			×			×	
8		×		×			×			×
9		×			×			×		×
10		×					×	×	×	
11			×	×	×			×		
12			×		×				×	×
13			×			×	×			×
14				×	×	×	×			
15				×		×		×	×	

cc) 5 of 10 BIB: 18 Five-tuples of 10 samples (t = 10, k = 5, b = 18, r = 9, λ = 4).

See Table A.11.

Table A.11 — 18 Five-tuples of 10 samples

Block	Sample									
	1	2	3	4	5	6	7	8	9	10
1	×	×	×	×	×					
2	×	×	×			×	×			
3	×	×		×		×			×	
4	×	×			×		×	×		
5	×		×			×		×	×	
6	×		×				×	×		×
7	×			×	×	×				×
8	×			×				×	×	×
9	×				×		×		×	×
10		×	×	×				×		×
11		×	×		×				×	×
12		×		×			×	×	×	
13		×			×	×		×		×
14		×				×	×		×	×
15			×	×	×		×		×	
16			×	×		×	×			×
17			×		×	×		×	×	
18				×	×	×	×	×		

dd) 6 of 10 BIB: 15 Six-tuples of 10 samples (t = 10, k = 6, b = 15, r = 9, λ = 5).

See Table A.12.

Table A.12 — 15 Six-tuples of 10 samples

Block	Sample									
	1	2	3	4	5	6	7	8	9	10
1	×	×	×		×		×			×
2	×	×	×					×	×	×
3	×	×		×	×			×	×	
4	×	×		×		×	×	×		
5	×	×				×	×		×	×
6	×		×	×	×	×				×
7	×		×	×		×	×		×	
8	×		×		×	×		×	×	
9	×			×	×		×	×		×
10		×	×	×	×		×		×	
11		×	×	×		×		×		×
12		×	×		×	×	×	×		
13		×		×		×	×	×		
14			×	×			×	×	×	×
15					×	×	×	×	×	×

Annex B
(informative)

Example of balanced incomplete block design with ratings data

B.1 Problem and situation

The quality control manager of a mustard factory routinely screens samples of finished product to be added to the pool of reference samples. New reference samples are needed at regular intervals, as the older samples will have changed with time and are no longer appropriate. The procedure is also used to eliminate from the pool any current reference samples that may have deteriorated.

B.2 Test design

Samples from six lots are evaluated for overall off-flavour. Evaluations are performed by 15 trained assessors using a 10 point category scale from 0, representing no off-flavour, to 9, indicating extreme off-flavour. The assessors evaluate four of the six samples. The four samples that each assessor evaluates are determined by the BIB design presented in Table B.1. Each of the 15 assessors is randomly assigned one block of four samples from the design. The order of presentation of the samples within each block is randomized.

Table B.1 — BIB design for ratings data
$t = 6$, $k = 4$, $r = 10$, $b = 15$, $\lambda = 6$

Block (assessor)	Sample					
	1	2	3	4	5	6
1	×	×	×	×		
2	×			×	×	×
3		×	×		×	×
4	×	×	×		×	
5	×	×		×		×
6			×	×	×	×
7	×	×	×			×
8	×		×	×	×	
9		×		×	×	×
10	×	×		×	×	
11	×		×		×	×
12		×	×	×		×
13	×	×			×	×
14	×		×	×		×
15		×	×	×	×	

B.3 Results

The off-flavour intensity ratings are presented in Table B.2. The data are analysed by a program capable of performing a BIB design ANOVA. The resulting ANOVA table is presented in Table B.3. The F-statistic for samples is highly significant ($p < 0{,}000\,1$), indicating that there are perceptible differences in off-flavour intensity among the samples. An LSD multiple comparison procedure is applied to the average ratings of the samples to determine which samples have significantly different off-flavour intensities. The results in Table B.4 reveal that sample 1 has significantly higher off-flavour than all of the other samples. There are no significant differences among the remaining samples.

Table B.2 — Off-flavour intensity ratings of six mustards

Assessor	Sample					
	1	**2**	**3**	**4**	**5**	**6**
1	6	1	1	2		
2	6			1	3	3
3		4	2		5	2
4	7	2	3		2	
5	3	5		1		1
6			1	1	3	2
7	7	4	4			3
8	2		1	1	1	
9		2		2	2	3
10	4	2		2	5	
11	5		3		1	1
12		3	2	1		2
13	4	2			1	1
14	5		2	2		1
15		2	4	5	3	

Table B.3 — Balanced incomplete block ANOVA table of ratings data: off-flavour intensity of six mustards

Source of variation	Sum of squares (SS)	Degrees of freedom (DF)	Mean square (MS)	F	p
Total	156,60	59			
Assessors (blocks)	38,58	14			
Samples (treatments, adjusted for blocks)	64,08	5	12,82	9,60	< 0,000 1
Error	53,42	40	1,34		

Table B.4 — Adjusted means of off-flavour intensities of six mustards

Sample	1	2	3	4	5	6
Adjusted mean	5,0 A	2,5 B	2,2 B	2,0 B	2,6 B	1,9 B
NOTE Means with no letters in common are significantly different at the 5 % level of significance (L = 1,1).						

Annex C
(informative)

Example of balanced incomplete block design with rank data

C.1 Problem and situation

A pepper sauce manufacturer wants to assess the spiciness of 15 varieties of peppers. The manufacturer wants to specify a limited number of varieties to use in its products.

C.2 Test design

Because of the high spiciness levels of the peppers, a balanced incomplete design is chosen. Each assessor evaluates $k = 3$ of the $t = 15$ samples. The basic BIB design consists of $b = 35$ blocks, in which each sample is evaluated $r = 7$ times and every pair of samples is evaluated together once ($\lambda = 1$). The basic design is repeated $p = 3$ times to obtain a total of $p*r = 21$ evaluations per sample.

A randomly selected group of 105 assessors was recruited to participate in the test. The assessors screened for their ability to distinguish different levels of spiciness and were instructed on the use of the ranking method. Because of their lack of extensive training in sensory testing, the assessors were instructed to rank the three samples as: 1 = most spicy; 2 = medium spicy; and 3 = least spicy.

C.3 Analysis of results

To make the results easier to analyse, the rank data from the study are arranged as shown in Table C.1. The rank sum for a given pepper variety is simply the sum of all the numbers in the column corresponding to that variety. The value of Friedman's test statistic, F_{test}, is computed to determine if there are any differences in spiciness among the varieties. The value of $F_{\text{test}} = 68{,}53$ exceeds the upper 5 % critical value of a χ^2 with $(t - 1) = 14$ degrees of freedom ($\chi^2_{14,\,0{,}05} = 23{,}69$), and it is concluded that there are indeed significant differences in the dataset. A 95 % LSD multiple comparison statistic for rank data is computed to determine which of the varieties are significantly different ($L = 11$). The results of the multiple comparisons are presented in Table C.2.

Table C.1 — Results obtained in BIB example with rank data: spiciness of peppers

Block (assessor)	Sample or variety [a]														
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1	2	3												
2				3			1			2					
3					1					3					2
4						3					2		1		
5							3		1					2	
6	3			2	1										
7		2						3		1					
8			3										2	1	
9						3			2						1
10							2				1	3			
101	2													3	1
102		1		3		2									
103			1					2			3				
104					1				2			3			
105							3			2			1		
Rank sum	35	45	54	43	28	37	55	42	37	50	49	50	34	42	29

[a] Responses: 1 = most spicy; 2 = medium spicy; 3 = least spicy.

Table C.2 — Summary of results and statistical analysis of the BIB rank data: spiciness of peppers

Sample or variety	Rank sum				
5	28	a			
15	29	a			
13	34	a	b		
1	35	a	b		
6	37	a	b		
9	37	a	b		
14	42		b	c	
8	42		b	c	
4	43		b	c	
2	45			c	d
11	49			c	d
10	50			c	d
12	50			c	d
3	54				d
7	55				d

NOTE Rank sums with no letters in common are significantly different at the 5 % significance level (L_{rank} = 11).

Bibliography

[1] ISO 6564, *Sensory analysis — Methodology — Flavour profile methods* [2)]

[2] ISO 6658, *Sensory analysis — Methodology — General guidance*

[3] ISO 8586-1, *Sensory analysis — General guidance for the selection, training and monitoring of assessors — Part 1: Selected assessors*

[4] ISO 8586-2, *Sensory analysis — General guidance for the selection, training and monitoring of assessors — Part 2: Expert sensory assessors*

[5] ISO 8587, *Sensory analysis — Methodology — Ranking*

[6] ISO 11036, *Sensory analysis — Methodology — Texture profile*

[7] ISO 13299, *Sensory analysis — Methodology — General guidance for establishing a sensory profile*

[8] COCHRAN, W.G., COX, G.M. *Experimental designs*, 2nd edition. New York, NY: Wiley, 1957. 611 p.

[9] GACULA, M.C. JR, SINGH, J. *Statistical methods in food and consumer research*, 2nd edition. Boston, MA: Elsevier/Academic Press, 2009. 853 p.

2) Withdrawn.

ICS 67.240

Price based on 19 pages