ICS 67.240
B 04

中华人民共和国国家标准

GB/T 12315—2008/ISO 8587:2006
代替 GB/T 12315—1990

感官分析　方法学　排序法

Sensory analysis—Methodology—Ranking

(ISO 8587:2006,IDT)

2008-06-25 发布　　2008-12-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会　发布

数码防伪

目　次

前言 …… Ⅰ
1　范围 …… 1
2　规范性引用文件 …… 1
3　术语和定义 …… 2
4　方法提要 …… 2
5　检验的一般条件 …… 2
6　评价员 …… 2
6.1　条件 …… 2
6.2　评价员人数 …… 2
6.3　检验前的统一认识 …… 3
7　检验步骤 …… 3
7.1　样品提供 …… 3
7.2　参比样 …… 3
7.3　检验技术 …… 3
7.4　回答表 …… 3
8　结果的表述与解释 …… 3
8.1　排序结果与秩和计算 …… 3
8.2　统计分析和解释 …… 4
9　检验报告 …… 11
附录A（规范性附录）　检验条件的确定 …… 12
附录B（资料性附录）　应用实例——完全区组设计 …… 13
附录C（资料性附录）　应用实例——平衡不完全区组设计 …… 15
附录D（资料性附录）　检验回答表格式样 …… 17
参考文献 …… 18

表1　排序结果与秩和计算 …… 4
表2　Spearman相关系数的临界值 …… 4
表3　完全区组设计中Page检验的临界值 …… 6
表4　Friedman检验的临界值(0.05和0.01水平) …… 7
表5　χ^2分布临界值 …… 8
表6　符号检验的临界值(双侧) …… 10

前　言

本标准等同采用国际标准 ISO 8587:2006《感官分析　方法学　排序法》(Sensory analysis—Methodology—Ranking)。

为便于使用,本标准做了下列编辑性修改:

a) "本国际标准"一词改为"本标准"。

b) 用小数点"."代替作为小数点的逗号",";

c) 删除国际标准的前言。

d) 对标准中出现的公式进行了编号,以表述更明确。

e) 将 8.2.5 中的注,改为附注 3),以更突出重点、表达更清晰。

f) 对附录 A 中表 A.1 的第一列适当进行了拆分合并处理,以使检验目的更加明确。对附录 B 和附录 C 根据检验分析的步骤,划分出条,进行了编号,并根据每条的内容列出了标题,以表达更清晰,便于理解。

本标准代替 GB/T 12315—1990《感官分析方法　排序法》。本标准与 GB/T 12315—1990 相比,重大的技术内容变化主要有:

——对采标的国际标准原文处理不同,旧版标准参照采用 ISO 8587:1988,新版标准等同采用 ISO 8587:2006;

——增加了采用 Spearman 相关系数评判评价员个人表现和采用 Page 检验评判小组表现的技术内容;

——区分了不同试验设计进行结果表达与解释的差异,如完全区组设计和平衡不完全区组设计中,计算 LSD 最小显著差数值的不同公式;

——明确了不同检验目的所需的检验条件,如评价员水平、评价员人数以及可采用的统计方法等。

本标准的附录 A 为规范性附录,附录 B、附录 C、附录 D 为资料性附录。

本标准由中国标准化研究院提出并归口。

本标准主要起草单位:中国标准化研究院、北京工商大学、中国人民解放军总后勤部军需装备研究所、今麦郎食品有限公司。

本标准主要起草人:赵镭、刘文、汪厚银、宋焕禄、钱平、侯国友。

本标准所代替标准的历次版本发布情况为:

——GB/T 12315—1990。

感官分析　方法学　排序法

1　范围

本标准规定了将一系列被检样品按其某种特性或整体印象的顺序进行排列的感官分析方法。

本标准适用于评价样品间的差异，如样品某一种或多种[1]感官特性的强度，或者评价人员对样品的整体印象。该方法可用于辨别样品间是否存在差异，但不能确定样品间差异的程度。

本标准规定的方法适用于以下情况：

a）　评价员评估：包括培训评价员以及测定评价员个人或小组的感官阈值；

b）　产品评估：

1）　在描述性分析或偏爱检验前，对样品初步筛选；

2）　在描述性分析和偏爱检验时，确定由于原料、加工、包装、贮藏以及被检样品稀释顺序的不同，对产品一个或多个感官指标强度水平的影响；

3）　在偏爱检验时，确定偏好顺序。

2　规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准，然而，鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 10221　感官分析　术语（GB/T 10221—1998，idt ISO 5492:1992）

GB/T 16860　感官分析方法　质地剖面检验（GB/T 16860—1997，idt ISO 11036:1994）

GB/T 16861　感官分析　通过多元分析方法鉴定和选择用于建立感官剖面的描述词（GB/T 16861—1997，idt ISO 11035:1994）

ISO 3534-1　统计学术语　第1部分：一般统计术语（Statistics—Vocabulary and symbols—Part 1:General statistical terms and terms used in probability）

ISO 6658　感官分析方法总论（Sensory analysis—Methodology—General guidance）

ISO 8586-1　感官分析　专家的选拔、培训和管理导则（Sensory analysis—General guidance for the selection training and monitoring of assessors—Part 1:Selected assessors）

ISO 8589　感官分析　建立感官分析实验室的一般导则（Sensory analysis—General guidance for the design of test rooms）

ISO 8586-2　感官分析　选拔与培训感官分析优选评价员导则（Sensory analysis—General guidance for the selection and training of selected assessors）

注：与本标准规范性引用的国际文件有一致性对应关系的我国标准如下：

——GB/T 3358.1—1993　统计学术语　第一部分　一般统计术语（ISO 3534-1:1993，NEQ）；

——GB/T 10220—1988　感官分析方法总论（ISO 6658:1985，NEQ）；

——GB/T 16291—1996　感官分析　专家的选拔、培训和管理导则（ISO 8586-1:1994，NEQ）；

——GB/T 13868—1992　感官分析　建立感官分析实验室的一般导则（ISO 8589:1988，EQV）。

1）　在评价样品多种感官特性强度的差异时，被检的每一种感官特性都应通过不同的检验来排序，即检验时，同一样品被编上不同的编码，以不同的次序提供给同一评价员。

3 术语和定义

GB/T 10221 和 ISO 3534-1 中确立的感官分析和统计学的术语和定义适用于本标准。

4 方法提要

评价员同时接受三个或三个以上的样品，排列顺序是随机的。

注：虽然对两个样品也能进行排序，但此时通常采用成对比较法(参照 GB/T 12310)。

评价员按照规定的准则，对样品进行排序，给出每个样品的序位，即秩。秩既可按照某个属性或特性给出，也可按整体印象给出综合秩。

计算秩的和(秩和)，然后进行统计比较(检验)。

5 检验的一般条件

检验时，对样品、实验室和检验用具的具体要求，见 ISO 6658 和 ISO 8589 等相关标准。

准备被检样品时，应注意以下三个方面：

a) 被检样品的制备、编码和提供。

b) 被检样品的数量。被检样品的数量应根据被检样品的性质(如饱和敏感度效应)和所选的试验设计来确定，并根据样品所归属的产品种类或采用的评价准则进行调整。如：优选评价员或专家最多一次只能评价 15 个风味较淡的样品，而消费者最多只能评价 3 个涩味的、辛辣的或者高脂的样品。甜味的饱和度较苦味的饱和度低，甜味样品的数量可比苦味样品的数量多。

c) 被检样品的说明。

6 评价员

6.1 条件

评价员应具备的条件依检验的目的而定，见附录 A。

参加检验的所有评价员应尽可能地具有同等的资格水平，所需水平的高低由检验的目的来决定。

a) 如要开展以下三方面的工作，需要选择优选评价员或专家：

1) 培训评价员；

2) 进行描述性分析，确定由于原料、加工、包装、贮藏以及被检样品稀释顺序的不同，而造成的对产品一个或多个感官指标强度水平的影响；

3) 测试评价员个人或小组的感官阈值。

b) 只进行偏爱检验或者样品的初步筛选(即从大量的产品中挑选出部分产品作进一步更精细的感官分析)，可选择未经培训的评价员或消费者，但要求他们接受过该方法的培训；所有参加检验的评价员均应符合 ISO 6658、ISO 8586-1 和 ISO 8586-2 的要求，并应接受关于排序法和所使用描述词的专门培训。

6.2 评价员人数

评价员人数依检验目的确定，见附录 A。

进行描述性分析时，按照可接受统计风险的水平以及标准 GB/T 16861 和 GB/T 16860 的要求，确定最少需要的评价人数，宜为 12 位～15 位优选评价员。

进行偏爱检验中确定偏好顺序时，同样依据可接受风险的水平，确定最少需要的评价员人数，一般每组至少 60 位消费者类型评价人员。

进行评价员工作检查、评价员培训以及测试评价员个人或小组的感官阈值时，评价员人数可不限定。

表 1 排序结果与秩和计算

评价员	样品				秩和
	A	B	C	D	
1	1	2	3	4	10
2	4	1.5	1.5	3	10
3	1	3	3	3	10
4	1	3	4	2	10
5	3	1	2	4	10
6	2	1	3	4	10
7	2	1	4	3	10
每个样品的秩和	14	12.5	20.5	23	70
注：每行秩和等于 $0.5p(p+1)$，其中 p 为样品的数量。					

如果有相同秩次，取平均秩次(如，表 1 中，评价员 2 对样品 B、C 有相同秩次评价，评价员 3 对样品 B、C、D 有相同秩次评价)。

如无遗漏数据，且相同秩次能正确计算，则表中每行应有相同的秩和。将每一列的秩次相加，可得到每个样品的每列秩和。样品的每列秩和表示所有的评价员对样品排序结果的一致性。如果评价员的排序结果比较一致，则每列秩和的差异较大。反之，若评价员排序结果不一致时，每列秩和差异不大。因此，通过比较样品的秩和，可评估样品间的差异。

8.2 统计分析和解释

依据检验的目的选择统计检验方法，见附录 A。

8.2.1 个人表现判定：Spearman 相关系数

在比较两个排序结果，如：两位评价员所做出的评价结果之间或是评价员排序的结果与样品的理论排序之间的一致性时，可由式(1)计算 Spearman 相关系数，并参考表 2 列出的临界值 r_S 来判定相关性是否显著。

$$r_S = 1 - \frac{6\sum d_i^2}{p(p^2-1)} \quad \cdots\cdots(1)$$

式中：

d_i——样品 i 两个秩次的差；

p——参加排序的样品(产品)数。

表 2 Spearman 相关系数的临界值

样品数	显著性水平 α	
	$\alpha=0.05$	$\alpha=0.01$
6	0.886	—
7	0.786	0.929
8	0.738	0.881
9	0.700	0.833
10	0.648	0.794
11	0.618	0.755
12	0.587	0.727
13	0.560	0.703

PRC STANDARDS OF CHINA

表 2（续）

样品数	显著性水平 α	
	$\alpha=0.05$	$\alpha=0.01$
14	0.538	0.675
15	0.521	0.654
16	0.503	0.635
17	0.485	0.615
18	0.472	0.600
19	0.460	0.584
20	0.447	0.570
21	0.435	0.556
22	0.425	0.544
23	0.415	0.532
24	0.406	0.521
25	0.398	0.511
26	0.390	0.501
27	0.382	0.491
28	0.375	0.483
29	0.368	0.475
30	0.362	0.467

若 Spearman 相关系数接近+1，则两个排序结果非常一致；若接近 0，两个排序结果不相关；若接近−1，表明两个排序结果极不一致。此时应考虑是否存在评价员对评价指示理解错误或者将样品与要求相反的次序进行了排序。

8.2.2 小组表现判定：Page 检验

样品具有自然顺序或自然顺序已确认的情况下（例如样品成分的比例、温度、不同的贮藏时间等可测因素造成的自然顺序），该分析方法可用来判定评价小组能否对一系列已知或者预计具有某种特性排序的样品进行一致的排序。

如果 R_1、R_2、…、R_p 是以确定的排序排列的 p 个样品的理论上的秩和，那么若样品间没有差异。

a） 原假设可写成：

$H_0: R_1=R_2=\cdots=R_p$。

备择假设则是：$H_1: R_1\leqslant R_2\leqslant\cdots\leqslant R_p$，其中至少一个不等式是严格成立的。

b） 为了检验该假设，计算 Page 系数 L：

$L=R_1+2\,R_2+3\,R_3+\cdots+p\,R_p$

其中 R_1 是已知样品顺序中排序为第一的样品的秩和，依次类推，R_p 就是排序为最后的样品的秩和。

c） 得出统计结论：

表 3 给出了在完全区组设计中 L 的临界值，其临界值与样品数、评价员人数以及选择的统计学水平有关（$\alpha=0.05$ 或者 $\alpha=0.01$），当评价员的结果与理论值一致时，L 有最大值。

表 3　完全区组设计中 Page 检验的临界值

评价员人数 j	样品(或产品)数 p											
	3	4	5	6	7	8	3	4	5	6	7	8
	显著性水平 $\alpha=0.05$						显著性水平 $\alpha=0.01$					
7	91	189	338	550	835	1 204	93	193	346	563	855	1 232
8	104	214	384	625	9 501	1 371	106	220	393	640	972	1 401
9	116	240	431	701	1 065	1 537	119	246	441	717	1 088	1 569
10	128	266	477	777	1 180	1 703	131	272	487	793	1 205	1 736
11	141	292	523	852	1 295	1 868	144	298	534	869	1 321	1 905
12	153	317	570	928	1 410	2 035	156	324	584	946	1 437	2 072
13	165	343*	615*	1 003*	1 525*	2 201*	169	350*	628*	1 022*	1 553*	2 240*
14	178	368*	661*	1 078*	1 639*	2 367*	181	376*	674*	1 098*	1 668*	2 407*
15	190	394*	707*	1 153*	1 754*	2 532*	194	402*	721*	1 174*	1 784*	2 574*
16	202	420*	754*	1 228*	1 868*	2 697*	206	427*	767*	1 249*	1 899*	2 740*
17	215	445*	800*	1 303*	1 982*	2 862*	218	453*	814*	1 325*	2 014*	2 907*
18	227	471*	846*	1 378*	2 097*	3 028*	231	479*	860*	1 401*	2 130*	3 073*
19	239	496*	891*	1 453*	2 217*	3 193*	243	505*	906*	1 476*	2 245*	3 240*
20	251	522*	937*	1 528*	2 325*	3 358*	256	531*	953*	1 552*	2 360*	3 406*

注：标"*"的值是通过正态分布近似计算得到的临界值。

比较 L 与表 3 中的临界值：

如果 $L<L_{\alpha}$，产品间没有显著性差异。

如果 $L\geqslant L_{\alpha}$，则产品的秩和间存在显著性差异：拒绝原假设而接受备择假设(可以得出结论：评价员做出了与预知的次序相一致的排序)。

如果评价员的人数或样品数未在表 3 中列出，按式(2)计算 L' 统计量：

$$L'=\frac{12L-3jp(p+1)^2}{p(p+1)\sqrt{j(p-1)}} \qquad \cdots\cdots(2)$$

式中：

j——评价员人数；

p——参加排序的样品数。

L' 统计量近似服从标准正态分布。

当 $L'\geqslant 1.64(\alpha=0.05)$ 或 $L'\geqslant 2.33(\alpha=0.01)$ 时，拒绝原假设而接受备择假设(见表 3)。

若试验设计为平衡不完全区组设计，则按式(3)计算 L' 统计量：

$$L'=\frac{12L-3j\cdot k(k+1)(p+1)}{\sqrt{j\cdot k(k-1)(k+1)p(p+1)}} \qquad \cdots\cdots(3)$$

式中：

j——评价员人数；

k——每个评价员排序的样品数；

p——参加排序的总样品数。

L' 统计量近似服从标准正态分布 $N(0,1)$。

同样，当 $L'\geqslant 1.64(\alpha=0.05)$ 或 $L'\geqslant 2.326(\alpha=0.01)$ 时，拒绝原假设而接受备择假设(见表 3)。

因为原假设所有理论秩和都相等，所以即便统计的结果显示差异性显著，也并不表明样品间所有的差异都能区分。只能说明至少有一对样品的差异可以在预排序中被区分。

8.2.3 产品理论顺序未知下的产品比较

Friedman 检验能最大限度地显示评价员对样品间差异的识别能力。

8.2.3.1 至少有两个产品存在显著性差异

该检验应用于 j 个评价员对相同的 p 个样品进行评价。

R_1、R_2、…、R_p 分别是 j 个评价员给出的 1 个～p 个样品的秩和。

a) 原假设可写成：

H_0：$R_1=\cdots=R_p$ 即认为样品间无显著差异。

备择假设则是：$R_1=\cdots=R_p$ 中至少有一个等式不成立。

b) 为了检验该假设，计算 F_{test} 值：

完全区组设计中按式(4)计算 F_{test} 值。

$$F_{test}=\frac{12}{jp(p+1)}(R_1^2+\cdots+R_p^2)-3j(p+1) \qquad \cdots\cdots(4)$$

式中：

R_i——第 i 个产品的秩和。

平衡不完全区组设计中按式(5)计算 F_{test} 值。

$$F_{test}=\frac{12}{j\cdot p(k+1)}(R_1^2+\cdots+R_p^2)-\frac{3r\cdot n^2(k+1)}{g} \qquad \cdots\cdots(5)$$

式中：

k——每个评价员排序的样品数；

R_i——i 产品的秩和；

r——重复次数；

n——每个样品被评价的次数；

g——每两个样品被评价的次数。

c) 得出统计结论：

如果 $F_{test}>F$，根据表 4 中评价员人数，样品(产品)数和显著性水平($\alpha=0.05$ 或 $\alpha=0.01$)，就拒绝原假设，认为产品的秩次间存在显著差异，即产品间存在显著差异。

表 4 Friedman 检验的临界值(0.05 和 0.01 水平)

评价员人数 j	样品(或产品)数 p									
	3	4	5	6	7	3	4	5	6	7
	显著性水平 $\alpha=0.05$					显著性水平 $\alpha=0.01$				
7	7.143	7.8	9.11	10.62	12.07	8.857	10.371	11.97	13.69	15.35
8	6.250	7.65	9.19	10.68	12.14	9.000	10.35	12.14	13.87	15.53
9	6.222	7.66	9.22	10.73	12.19	9.667	10.44	12.27	14.01	15.68
10	6.200	7.67	9.25	10.76	12.23	9.600	10.53	12.38	14.12	15.79
11	6.545	7.68	9.27	10.79	12.27	9.455	10.60	12.46	14.21	15.89
12	6.167	7.70	9.29	10.81	12.29	9.500	10.68	12.53	14.28	15.96
13	6.000	7.70	9.30	10.83	12.37	9.385	10.72	12.58	14.34	16.03
14	6.143	7.71	9.32	10.85	12.34	9.000	10.76	12.64	14.40	16.09
15	6.400	7.72	9.33	10.87	12.35	8.933	10.80	12.68	14.44	16.14

表 4（续）

评价员人数 j	样品（或产品）数 p									
	3	4	5	6	7	3	4	5	6	7
	显著性水平 $\alpha=0.05$					显著性水平 $\alpha=0.01$				
16	5.99	7.73	9.34	10.88	12.37	8.79	10.84	12.72	14.48	16.18
17	5.99	7.73	9.34	10.89	12.38	8.81	10.87	12.74	14.52	16.22
18	5.99	7.73	9.36	10.90	12.39	8.84	10.90	12.78	14.56	16.25
19	5.99	7.74	9.36	10.91	12.40	8.86	10.92	12.81	14.58	16.27
20	5.99	7.74	9.37	10.92	12.41	8.87	10.94	12.83	14.60	16.30
∞	5.99	7.81	9.49	11.07	12.59	9.21	11.34	13.28	15.09	16.81

注 1：F 可能是不连续值，此不连续性是由于 j、p 值较小而造成，故在 $\alpha=0.05$ 和 $\alpha=0.01$ 的情况下，不能得到临界值。

注 2：使用 χ^2 分布的一个近似值得到用斜体表示的值即临界值。

如果样品（产品）数或者评价员人数未列在表中，可将 F_{test} 看作自由度为 $p-1$ 的 χ^2 分布，估算出临界值。χ^2 分布的临界值参照表 5，p 为样品（产品）数。

表 5 χ^2 分布临界值

样品（或产品）数 p	χ^2 自由度（$v=p-1$）	显著性水平 α	
		$\alpha=0.05$	$\alpha=0.01$
3	2	5.99	9.21
4	3	7.81	11.34
5	4	9.49	13.28
6	5	11.07	15.09
7	6	12.59	16.81
8	7	14.07	18.47
9	8	15.51	20.09
10	9	16.92	21.67
11	10	18.31	23.21
12	11	19.67	24.72
13	12	21.03	26.22
14	13	22.36	27.69
15	14	23.68	29.14
16	15	25.00	30.58
17	16	26.30	32.00
18	17	27.59	33.41
19	18	28.87	34.80
20	19	30.14	36.19
21	20	31.4	37.6

表 5（续）

样品（或产品）数 p	χ^2 自由度（$v=p-1$）	显著性水平 α	
		$\alpha=0.05$	$\alpha=0.01$
22	21	32.7	38.9
23	22	33.9	40.3
24	23	35.2	41.6
25	24	36.4	43.0
26	25	37.7	44.3
27	26	38.9	45.6
28	27	40.1	47.0
29	28	41.3	48.3
30	29	42.6	49.6

8.2.3.2 检验哪些产品与其他产品存在显著性差异

如果 Friedman 检验的结论是产品之间存在显著性差异时，则可通过在选定的风险 α 下，计算最小显著差（LSD）来确定哪些产品与其他产品存在显著性差异（$\alpha=0.05$，或 $\alpha=0.01$）。

在考虑风险 α 水平（显著性水平，即实际不存在差异，而检验结果存在差异的概率）时，应选用以下两种方法之一：

a) 当风险水平是应用于某特定产品对时，实际风险即是 α。例如当 $\alpha=0.05$，在计算 LSD 时的 z 值为 1.96（对应于双尾正概率为 α），此时的风险称为比较风险或个别风险。

b) 当风险水平 α 应用于整个实验，则与每个产品对有关的实际风险为 $\alpha'=2\alpha/p(p-1)$。例如，当 $p=8$，$\alpha=0.05$ 时，$\alpha'=0.001\ 8$，$z=2.91$（对应于双尾正概率为 α'）。此时的风险称为实验风险或整体风险。

大多数情况下，往往选用实验风险去判定哪些产品与其他产品存在显著性差异。

在完全区组试验设计中，LSD 值由式（6）得出：

$$\mathrm{LSD}=z\sqrt{\frac{j\cdot p(p+1)}{6}} \qquad \cdots\cdots(6)$$

在平衡不完全区组试验设计中，LSD 值由式（7）得出：

$$\mathrm{LSD}=z\sqrt{\frac{r(k+1)(n\cdot k-n+g)}{6}} \qquad \cdots\cdots(7)$$

计算两两样品的秩和之差，并与 LSD 值比较。若秩和之差等于或者大于 LSD 值，则这两个样品之间存在显著性差异，即排序检验时，已区分出这两个样品之间的差异。反之，若秩和之差小于 LSD 值，则这两个样品之间不存在显著性差异，即排序检验时，未区分出这两个样品之间的差异。

8.2.4 同秩情况

若两个或多个样品同秩次，则完全区组设计中的 F 值应替换为 F'，由式（8）得出：

$$F'=\frac{F}{1-\{E/[j\cdot p(p^2-1)]\}} \qquad \cdots\cdots(8)$$

其中 E 值由式（9）得出：

令 n_1、n_2、…、n_k 为每个同秩组里秩次相同的样品数，则：

$$E=(n_1^3-n_1)+(n_2^3-n_2)+\cdots+(n_k^3-n_k) \qquad \cdots\cdots(9)$$

例如，表 1 中有两个组出现了同秩情况：

——第 2 行中 B、C 样品同秩次（评价结果来源于二号评价员），则 $n_1=2$；

——第3行中B、C和D样品同秩次(评价结果来源于三号评价员),则 $n_2=3$。

故:$E=(2^3-2)+(3^3-3)=6+24=30$

因 $j=7$,$p=4$,先计算出 F,再按式(8)计算 F':

$$F'=\frac{F}{1-\{30/[7\times4(4^2-1)]\}}=1.08F$$

然后将 F' 与表4或表5中的临界值比较,从而得出统计结论。

8.2.5 比较两个产品:符号检验

某些特殊的情况用排序法进行两个产品之间的差异比较时,可使用符号检验[3)]。

如比较两个产品A和B的差异。k_A 是产品A排序在产品B之前的评价次数。k_B 表示产品B排序在产品A之前的评价次数。k 则是 k_A 和 k_B 之中较小的那个数,即 $k=\min\{k_A,k_B\}$。而未区分出A和B差异的评价不在统计的评价次数之内。

原假设:$H_0:k_A=k_B$。

备择假设:$H_1:k_A\neq k_B$。

如果 k 小于表6中配对符号检验的临界值,则拒绝原假设而接受备择假设。表明A和B之间存在显著性差异。

表6 符号检验的临界值(双侧)

评价员人数 j	显著性水平 α		评价员人数 j	显著性水平 α	
	$\alpha=0.01$	$\alpha=0.05$		$\alpha=0.01$	$\alpha=0.05$
1			26	6	7
2			27	6	7
3			28	6	8
4			29	7	8
5			30	7	9
6		0	31	7	9
7		0	32	8	9
8	0	0	33	8	10
9	0	1	34	9	10
10	0	1	35	9	11
11	0	1	36	9	11
12	1	2	37	10	12
13	1	2	38	10	12
14	1	2	39	11	12
15	2	3	40	11	13
16	2	3	41	11	13
17	2	4	42	12	14
18	3	4	43	12	14
19	3	4	44	13	15
20	3	5	45	13	15
21	4	5	46	13	15
22	4	5	47	14	16
23	4	6	48	14	16
24	5	6	49	15	17
25	5	7	50	15	17

3) 这种情况更多采用成对比较检验(见GB/T 12310)。

表 6（续）

评价员人数 j	显著性水平 α		评价员人数 j	显著性水平 α	
	$\alpha=0.01$	$\alpha=0.05$		$\alpha=0.01$	$\alpha=0.05$
51	15	18	71	24	26
52	16	18	72	24	27
53	16	18	73	25	27
54	17	19	74	25	28
55	17	19	75	25	28
56	17	20	76	26	28
57	18	20	77	26	29
58	18	21	78	27	29
59	19	21	79	27	30
60	19	21	80	28	30
61	20	22	81	28	31
62	20	22	82	28	31
63	20	23	83	29	32
64	21	23	84	29	32
65	21	24	85	30	32
66	22	24	86	30	33
67	22	25	87	31	33
68	22	25	88	31	34
69	23	25	89	31	34
70	23	26	90	32	35

注：当 $j>90$ 时，临界值由公式 $L=(j-1)/2-k\sqrt{j+1}$ 计算，结果进行四舍五入取整。$\alpha=0.05$ 时，k 值为 0.980 0；$\alpha=0.01$ 时，k 值为 1.287 9。

9 检验报告

检验报告应包括以下内容：

a) 检验目的。

b) 样品确认所应包括的信息：

1) 样品数；

2) 是否使用参比样。

c) 采用的检验参数：

1) 评价员人数及其资格水平；

2) 检验环境；

3) 有关样品的情况说明。

d) 检验结果及其统计解释。

e) 注明根据本标准检验。

f) 如果实际实验中与本标准方法有偏差应予以说明。

g) 检验负责人的姓名。

h) 检验的日期与时间。

附 录 A
（规范性附录）
检验条件的确定

表 A.1 根据检验目的选择参数

检验目的		评价员水平	评价员人数	统计方法		
				同已知顺序比较（评价员表现评估）	产品顺序未知（产品比较）	
					两个产品	两个以上产品
评价员表现评估	个人表现评估	优选评价员或专家评价员	无限制	Spearman 检验	符号检验	Friedman 检验
	小组表现评估	优选评价员或专家评价员	无限制	Page 检验		
产品评估	描述性检验	优选评价员或专家评价员	12～15 为宜			
	偏好性检验	消费者	每组至少 60 位消费者类型的评价人员	—		

附 录 B
（资料性附录）
应用实例——完全区组设计

14 个评价员评价 5 个样品。

B.1 结果的汇集

见表 B.1。

表 B.1 评价实例

评价员	样品				
	A	B	C	D	E
1	2	4	5	3	1
2	4	5	3	1	2
3	1	4	5	3	2
4	1	2	5	3	4
5	1	5	2	3	4
6	2	3	4	5	1
7	4	5	3	1	2
8	2	3	5	4	1
9	1	3	4	5	2
10	1	2	5	3	2
11	4	5	2	3	1
12	2	4	3	5	1
13	5	3	4	2	1
14	3	5	2	4	1
秩和	33	53	52	45	27

STANDARDS PRESS OF CHINA

B.2 Friedman 检验

a) 计算统计量 F_{test}

$j=14, p=5, R_1=33, R_2=53, R_3=52, R_4=45, R_5=27$,

根据式(4),

$$F_{test}=\frac{12}{14\times5\times(5+1)}(33^2+53^2+52^2+45^2+27^2)-3\times14\times(5+1)=15.3$$

b) 作统计结论

因 F_{test}(15.31)大于表 4 中对应 $j=14, p=5, \alpha=0.05$ 的临界值 9.32,故可认为,在显著性水平小于或等于 5%时,5 个样品之间存在显著性差异。

B.3 多重比较和分组

如果两个样品秩和之差的绝对值大于最小显著差 LSD,可认为二者有显著性差异。

a) 计算最小显著差 LSD

$$LSD=1.96\times\sqrt{\frac{14\times5\times(5+1)}{6}}=16.40(\alpha=0.05)$$

b) 比较与分组

在显著性水平 0.05 下，A 和 B、A 和 C、E 和 B、E 和 C、E 和 D 的差异是显著的，它们秩和之差的绝对值分别为：

A—B：| 33—53 | =20　　E—B：| 27—53 | =26

A—C：| 33—52 | =19　　E—C：| 27—52 | =25

E—D：| 27—45 | =18

以上比较的结果表示如下：

E A　D C B

下划线的意义表示：

——未经连续的下划线连接的两个样品之间有显著性差异(在 5%的显著性水平下)；

——由连续的下划线连接的两个样品无显著性差异；

——无显著性差异的 A 和 E 排在无显著性差异的 D、C、B 前面。

因此，5 个样品可分为三组，一组包括 A 和 E，另一组包括 A 和 D，第三组包括 B、D、C。

B.4 Page 检验

根据秩和顺序，可将样品初步排序为：E≤A≤D≤C≤B，Page 检验可检验该推论。

a) 计算 L 值

$L=(1\times27)+(2\times33)+(3\times45)+(4\times52)+(5\times53)=701$

b) 作统计结论

由表 3 可知，$p=5$，$j=14$，$\alpha=0.05$ 时，Page 检验的临界值为 661。

因为 $L>661$，所以当 $\alpha=0.05$ 时，拒绝原假设，样品之间存在显著性差异。

B.5 结论

a) 基于 Friedman 检验

在 5%的显著性水平下，E 和 A 无显著性差异；D 和 C、B 无显著性差异；A 和 D 无显著性差异，但 A 和 C、B 有显著性差异，E 和 D、C、B 有显著性差异。

b) 基于 Page 检验

在 5%的显著性水平下，评价员辨别出了样品之间存在差异，并且给出的排序与预先设定的顺序一致。

附 录 C
（资料性附录）
应用实例——平衡不完全区组设计

平衡不完全区组设计中，10 个评价员每人检验 5 个样品中的 3 个。

C.1 结果的汇集

见表 C.1。

表 C.1 评价实例

评价员	样品				
	A	B	C	D	E
1	1	2	3		
2	1	2		3	
3	2	3			1
4	1		2	3	
5	2		3		1
6	1			3	2
7		1	3	2	
8		2	3		1
9		3		2	1
10			1	3	2
秩和	8	13	15	16	8

C.2 Friedman 检验

a) 计算统计量 F_{test}

$j=14, p=5, k=3, n=6, g=3, r=1, R_1=8, R_2=13, R_3=15, R_4=16, R_5=8,$

$$F_{test}=12\times(8^2+13^2+15^2+16^2+8^2)/[1\times3\times5\times(3+1)]-[3\times1\times6^2\times(3+1)/3]=11.6$$

b) 作统计结论

因 F_{test}(11.6)大于表 4 中对应 $p=5, \alpha=0.05$ 的临界值 9.25，故可认为，在显著性水平小于或等于 5%时，5 个样品之间存在显著性差异。

C.3 利用最小显著差分组

如果两个样品秩和之差的绝对值大于最小显著差 LSD，可认为二者有显著性差异。

a) 计算最小显著差 LSD

$$LSD=1.96\times\sqrt{\frac{1\times(3+1)\times(6\times3-6+3)}{6}}=6.2(\alpha=0.05)$$

b) 比较与分组

在显著性水平 0.05 下，A 和 C、A 和 D、C 和 E、D 和 E 之间的差异是显著的，其秩和之差的绝对值分别为：

A—C：| 8−15 | =7　　C—E：| 15−8 | =7

A—D：| 8−16 | =8　　D—E：| 16−8 | =8

以上比较的结果表示如下：

A　E　B　C　D

C.4 Page 检验

根据秩和顺序，可将样品初步排序为：E≤A≤D≤C≤B，Page 检验可检验该推论。

a) 计算 L 值

$L=(1\times8)+(2\times8)+(3\times16)+(4\times15)+(5\times13)=197$

$p=5$，$k=3$，$j=10$ 时，L'的值为：

$$L'=\frac{12\times197-3\times10\times3\times4\times6}{\sqrt{10\times3\times4\times5\times6}}=2.4$$

b) 作统计结论

因为 $L'>2.33$，所以当 $\alpha=0.01$ 时，拒绝原假设，样品之间存在极显著差异。

C.5 结论

a) 基于 Friedman 检验

在 5%的显著性水平下，A、E 的秩和显著小于 C、D，而 B 与其他 4 种样品均无显著性差异。

b) 基于 Page 检验

在 1%的显著性水平下，评价员辨别出了样品之间存在差异，并且给出的排序与预先设定的顺序一致。

附 录 D
（资料性附录）
检验回答表格式样

姓名：..... 日期：..... 检验号：.....

请按从左至右顺序品尝每个样品：

请在下面表格中以甜味增加的顺序写出样品编码：

编码	最不甜			最甜

注释：

参 考 文 献

[1] ISO 5495 Sensory analysis—Methodology—Paired comparison test

[2] FRIEDMAN M. The use of ranks to avoid the assumptions of normality implicit in the analysis of variance, Journal of the American Statistical Association, 32, 1937: 675-701.

[3] PAGE E B. Ordered hypotheses for multiple treatments: a significance test for linear ranks, Journal of the American Statistical Association, 58, 1963: 216-230.

[4] XP V09-500 Analyse sensorielle—Méthodologie—Directives générales pour la réalisation d'épreuves hédoniques en laboratoire d'évaluation sensorielle ou en salle en conditions contrôlées impliquant des consommateurs

[5] COCHRAN W G, COX G M. Experimental Designs. John Wiley & Sons, Inc., Chapter 11, Balanced Incomplete Blocks, 1950: 315-346.

[6] LOTHAR S. Applied Statistics, a Handbook of Techniques, Springer Series in Statistics, Springer-Verlag, 1982.

中华人民共和国
国家标准
感官分析 方法学 排序法
GB/T 12315—2008/ISO 8587:2006
*
中国标准出版社出版发行
北京复兴门外三里河北街16号
邮政编码:100045
网址 www.spc.net.cn
电话:68523946 68517548
中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销
*
开本 880×1230 1/16 印张 1.5 字数 39千字
2008年9月第一版 2008年9月第一次印刷
*
书号:155066·1-33609 定价 20.00元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68533533

GB/T 12315-2008