

ICS 67.240

XX XX



中华人民共和国国家标准

GB/T XXXX—201X
代替 GB/T 12316-1990

感官分析 方法学 “A” -非 “A” 检验

Sensory analysis - Methodology - “A”-“not A” test

(ISO 8588:2017,IDT)

(征求意见稿)

201X- - 发布

201X - - 实施

中华人民共和国国家市场监督管理总局
中国国家标准化管理委员会

发布

目 次

前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 原理.....	1
5 器具.....	2
6 抽样.....	2
7 检验的一般条件.....	2
8 评价员.....	3
8.1 资格、选择及安排.....	3
8.2 评价员和评价数量.....	3
9 步骤.....	3
10 结果的表达.....	4
附录 A.....	6
附录 B.....	10
附录 C.....	11
参考文献.....	12

前 言

本标准等同采用国际标准 ISO 8588:2017 《感官分析 方法学 “A” -非 “A” 检验》（Sensory analysis - Methodology - “A” - “not A” test）。

本部分由全国感官分析标准化技术委员会（SAC/TC566）提出并归口。

本部分起草单位：

本部分主要起草人：

感官分析 方法学 “A”-非“A”检验

1 范围

本标准规定了一个用于确定两类样品之间是否存在显著感官差异的程序。该方法适用于单一或或多个感官属性差异。

“A”-非“A”检验可通过以下方式用于感官分析:

a)差异检验,特别适用于评价具有不同外观(很难获得完全相同的重复样品)或后味(很难直接比较困难)的样品;

b)识别检验,特别适用于确定评价员或评价小组能否识别与已知刺激有关的新刺激物(例如,识别新甜味剂甜味的性质);

c)知觉检验,用于确定评价员辨别刺激的能力。

由于“A”-非“A”检验本质上已包含了所有评价员对同一产品的重复评价,如果两个产品非常相似,可以互换使用(例如相似性检验),那么“A”-非“A”检验不适用。这些重复的评价违背了相似性检验在统计上有效的基本假设。其应用实例见附件 B。

2 规范性引用文件

下列文件中的部分或全部条款通过本文件的引用而成为本文件的要求。凡是注明日期的引用文件,仅注明日期的版本适用于本文件。凡是未注明日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 10221 感官分析 术语

ISO 3534-1 统计学 术语和符号 第1部分:基本统计术语和概率术语

ISO 5492 感官分析 术语

ISO 8586:2012 感官分析 选择、培训和管理优选评价员和专家感官评价员的的一般导则

ISO 8589 感官分析 建立感官分析实验室的一般导则

3 术语和定义

GB/T 10221、ISO 5492中涉及感官分析和ISO 3534-1中涉及统计学的术语和定义适用于本文件。

4 原理

分发给评价员一系列样品,其中有的是产品“A”有的是“非A”。要求评价员指出每个样品是

产品“A”还是“非A”。该检验要求评价员对产品A熟悉，其可以在接触检验样品之前，对已知的产品“A”的样品进行体验。

5 器具

器具应由检验负责人根据所分析的产品的性质、样品的数量等进行选择，且不得以任何方式影响检验结果。

如果标准化器具符合检验的需要，则应使用标准化器具。

6 抽样

参考产品或正在检验产品的感官分析抽样标准。

如果没有这样的标准，则由有关各方协商议定抽样方法。

7 检验的一般条件

7.1 应以书面形式明确定义检验目标。

7.2 在评价员不能相互交流的条件下进行每一次评价会，直至全部评价结束。

7.3 进行检验的设施应符合 ISO 8589 的规定。

7.4 评价员不能从样品提供的方式识别出样品。

例如，在品尝检验中，应避免任何温度或外观差异。

例如，使用光过滤器、柔和的灯光或不透明的样品容器来掩蔽任何不相关的颜色差异。

7.5 使用随机抽取的3位数字，按照统一的方式对包含检验样品的容器进行编码。同一组中的每个检验样品编码应不同。如果完成检验需要进行多次评价会，且不同的评价会中使用了不同的编码，那么相同的两个编码(一个用于“A”样品，另一个用于非“A”样品)可以用于同一评价会中的所有评价员。

7.6 提供的所有检验样品的产品数量或体积应相同。在品尝检验中，可指定摄入量中的数量或体积。如果未指定，应指示评价员对相同数量或体积的每个检验样品进行评价。

7.7 检验样品的温度应相同，最好为产品常规的食用温度。

7.8 应保证评价员的职业安全。评价员应被告知如何评价检验样品。例如，评价员应被告知是否需要吞下检验样品，或者是否可以自由食用。在后一种情况下，应指示评价员以同样的方式对所有检验样品进行操作。

7.9 在评价会过程中，应避免提供关于产品标识、预期疗效或个体表现的信息直至检验完成。

8 评价员

8.1 资格、选择及安排

所有评价员均应具备根据 ISO 8586:2012 的检验目标选择的同等资质。根据检验目标的不同，评价员可以是完全的准评价员或训练有素的评价员。但是，在同一检验中，所有的评价员应具备同等的资质。

例如，如果正在进行的检验是因为怀疑非“A”产品可能会出现特定的污点，那么有对污点高度敏感历史的评价员可能会被选中。经验以及对产品的熟悉可能会提高评价员的表现，因而可能会增加发现显著差异的可能性。随着时间的推移，对评价人员的表现进行监控，对于增强的灵敏度可能是有用的。

所有评价员应熟悉“A”-“非A”检验的原理（评价的形式、任务和程序）。

8.2 评价员和评价数量

将要使用的评价员数量取决于检验的目标和所需的显著性水平。检验中会使用10至50名熟悉“A”产品的评价员。

每个评价员进行的重复评价的次数应根据产品的疲劳程度来确定。在“A”-非“A”检验中进行的总评价次数通常介于20到100次之间。

9 步骤

为了确保对产品“A”的熟悉，评价员在对检验样品进行评价之前可以对已知的样品“A”进行体验。根据检验目标的不同，评价员在对检验样品进行评价之前也可以对已知的非“A”样品进行体验。例如，如果研究人员认为一个或多个非“A”产品可能会呈现特定的水果味，那么评价员可能就会被安排对呈现超阈值强度水果属性的样品进行体验。一旦测试样本的评估开始，评审员就不能访问任何已知的样本。此外，在提供给评价员的系列样品中，评价员不知道“A”和非“A”样品的各自数量。

多个非“A”产品可以在同一检验中进行评价。应限制一个检验中的非“A”产品数量以避免感官疲劳。

应以随机的顺序提供“A”和非“A”样品，且对于每一位评价员，顺序应不同。提供给每个评价员的“A”样品的数目和“非A”样品的数目应相同（“A”样品的数目和“非 A”样品的数目不必相同），见A.2；类似地，如果对多个非“A”产品进行检验，则每个非“A”产品的数目不必相同，见 A.3。

根据样品的性质，以及为了避免感官适应的某些干扰效应，任何两个连续的样品提供时间间隔应相同。

回答表格式样见附件C

10 结果的表达

检验中的每一个非“A”产品单独进行分析。通过对检验中的每一个非“A”产品的分析，分析员得出一个三行三列的表（见表1）。

表1.观测数量

评价员的响应	提供的样品为“A”	提供的样品为非“A”	合计
评价员将样品标识为“A”	n_{11}	n_{12}	$n_{1.}$
评价员将样品标识为非“A”	n_{21}	n_{22}	$n_{2.}$
合计	$n_{.1}$	$n_{.2}$	$n_{..}$

其中

n_{11} 和 n_{22} 分别为所有评价员的正确的“A”和非“A”响应数；

n_{21} 和 n_{12} 分别为所有评价员的错误的“A”和非“A”响应数；

$n_{1.}$ 和 $n_{2.}$ 分别为第1行和第2行中响应的总和；

$n_{.1}$ 和 $n_{.2}$ 分别为第1列和第2列中响应的总和；

$n_{..}$ 是总响应数。

分析“A”-非“A”检验中获得的数据有两种方法。在第一种方法中，对结果的解释通过两步处理法获得。

a) 如果“A”样品被标识为非“A”样品的次数比例 ($n_{21}/n_{.1}$) 大于非“A”样品被标识为非“A”样品的次数比例 ($n_{12}/n_{.2}$)，停止解释并得出没有足够的证据表明产品之间存在显著的差异的结论。

b) 否则，应通过公式1计算 χ^2 检验统计量 T ，并将它与自由度为1的 χ^2 分布的 2α 临界值进行比较。如果检验统计量的值超过临界值，则得出样品存在显著差异的结论。如果未超过，则应通过公式1计算与检验统计量相关的p值，并将它与检验中选定的显著性水平 2α 进行比较。如果p值 $<2\alpha$ ，则得出样品存在显著差异的结论。

2α 行动标准的采用是因为 χ^2 检验本质上是双边的，其不能对“A”样品收到的太多或太少的非“A”响应进行区分。上述两步处理法的步骤a排除了由于“A”样品收到太多非“A”响应而声明其显著性的可能。因为误差的一半是由此不相关的选项造成，所以当在第二步中的 χ^2 检验使用 2α 时，两步检验处理法的I型真实误差为 α 。公式1为：

$$T = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

其中 n_{ij} 是表1第*i*行和*j*列的观测计数， E_{ij} 为第*i*行和*j*列的预期计数，其通过 $E_{ij} = (n_{i.} \times n_{.j})/n_{..}$ 计算得出（共四个单元格）。与公式1中检验统计量 T 相关的p值可以通过使用电子表格函

数（例如 CHISQ.DIST.RT函数）计算得出。p值将显示在包含“=CHISQ.DIST.RT(T, 1)”的单元格中，其中 T 是公式1中的检验统计量值。

在第二种方法中，对结果的解释通过一步处理法获得。

a) 计算公式2中的统计量 T_1 ，并将其与标准正态分布的上临界值 α 进行比较。如果检验统计量的值超过临界值，则得出样品显著不同的结论。如果未超过，则应通过公式2计算与检验统计量相关的p值，并将它与检验中选定的显著性水平 α 进行比较。如果 $p \text{ 值} < \alpha$ ，则得出样品显著不同的结论。公式2为：

$$T_1 = \frac{\sqrt{n_{..}}(n_{11}n_{22} - n_{12}n_{21})}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}} \quad (2)$$

与公式2中检验统计量 T_1 相关的p值可以通过使用电子表格函数（例如 NORM.S.DIST函数）计算得出。p值将显示在包含“=1 - NORM.S.DIST(T1, TRUE)”的单元格中，其中 T_1 是公式2中的检验统计量值。

请注意，对于样本量小于常规感官检验样本量的检验，其“A”-非“A”检验得到的数据也可以用Fisher精确检验来分析。

例子见附件A。

附录A

(规范性附录)

“A”-非“A”检验应用示例

A.1 例 1

识别蔗糖的甜味（“A”刺激）与某种甜味剂（“非 A”刺激）产生的甜味。

将两种物质置于水溶液中，使其甜味强度等于40 g/1蔗糖溶液的甜味强度。

评价员数量:20。

提供给每个评价员的样品数量：5个“A”和5个非“A”。

研究者选择在显著性水平 $\alpha = 0.05$ 处进行检验。

结果（所有评价员）：见表A.1。

表A.1 例1观测值

评价员的响应	提供的样品为“A”	提供的样品为非“A”	合计
评价员将样品标识为“A”	60	35	95
评价员将样品标识为非“A”	40	65	105
合计	100	100	200

使用第一种方法分析“A”-非“A”检验数据，“A”样品的非“A”响应比例（40%）小于非“A”样品的非“A”响应比例（65%），因此，分析可以进入第二步。公式1中的 T 统计量通过公式A.1计算得到：

$$T = \frac{(60-47,5)^2}{47,5} + \frac{(35-47,5)^2}{47,5} + \frac{(40-52,5)^2}{52,5} + \frac{(65-52,5)^2}{52,5} = 12,53 \quad (\text{A.1})$$

与检验统计量 T 相关的 p 值为 $P = 0.0004$ ，其小于研究者选择的显著性水平 $2\alpha = 0.10$ ，因此，可以得出结论，在95%置信水平下甜味剂之间存在显著的差异。

使用第二种方法分析“A”-非“A”检验数据，公式2中的 T_1 统计量通过公式A.2计算得到：

$$T_1 = \frac{\sqrt{200}[(60)(65)-(35)(40)]}{\sqrt{(95)(105)(100)(100)}} = 3,54 \quad (\text{A.2})$$

与检验统计量 T_1 相关的 p 值为 $P = 0.0002$ ，其小于研究者选择的显著性水平 $\alpha = 0.05$ ，因此，可以得出结论，在95%置信水平下甜味剂之间存在显著的差异。

A.2 例 2

与例1相同，但每个评价员评价4个“A”及6个非“A”产品。见表A.2。

表A.2 例2观测值

评价员的响应	提供的样品为“A”	提供的样品为非“A”	合计
评价员将样品标识为“A”	50	55	105
评价员将样品标识为非“A”	30	65	95
合计	80	120	200

使用第一种方法分析“A”-非“A”检验数据，“A”样品的非“A”响应比例（38%）小于非“A”样品的非“A”响应比例（54%），因此，分析可以进入第二步。公式1中的*T*统计量通过公式A.3计算得到：

$$T = \frac{(50-42)^2}{42} + \frac{(55-63)^2}{63} + \frac{(30-38)^2}{38} + \frac{(65-57)^2}{57} = 5,35 \quad (\text{A. 3})$$

与检验统计量*T*相关的*p*值为*P* = 0.0208，其小于研究者选择的显著性水平 $2\alpha = 0.10$ ，因此，可以得出结论，在95%置信水平下甜味剂之间存在显著的差异。

使用第二种方法分析“A”-非“A”检验数据，公式2中的*T*₁统计量通过公式A.4计算得到：

$$T_1 = \frac{\sqrt{200}[(50)(65) - (55)(30)]}{\sqrt{(105)(95)(80)(120)}} = 2,31 \quad (\text{A. 4})$$

与检验统计量*T*₁相关的*p*值为*P* = 0.0104，其小于研究者选择的显著性水平 $\alpha = 0.05$ ，因此，可以得出结论，在95%置信水平下甜味剂之间存在显著的差异。

A.3 例 3

与例1相同，但检验中对三种不同的非“A”产品（非“A1”，非“A2”和非“A3”）进行了评价。每名评价员评价5个“A”样品、3个非“A1”样品、3个非“A2”样品和2个非“A3”样品。每一个非“A”产品单独进行分析。见表A.3, A.4 和 A.5。

表A.3 例3非“A1”产品观测值

评价员的响应	提供的样品为“A”	提供的样品为非“A1”	合计
评价员将样品标识为“A”	60	32	92
评价员将样品标识为非“A”	40	28	68
合计	100	60	160

使用第一种方法分析“A”-非“A”检验数据，“A”样品的非“A”响应比例（40%）小于非“A1”样品的非“A”响应比例（47%），因此，分析可以进入第二步。公式1中的*T*统计量通过公式A.5计算得到：

$$T = \frac{(60-57,5)^2}{57,5} + \frac{(32-34,5)^2}{34,5} + \frac{(40-42,5)^2}{42,5} + \frac{(28-25,5)^2}{25,5} = 0,68 \quad (\text{A. 5})$$

与检验统计量 T 相关的 p 值为 $P = 0.4089$ ，其大于研究者选择的显著性水平 $2\alpha = 0.10$ ，因此，可以得出结论，在95%置信水平下，“A”和非“A1”产品之间存在显著的差异。

使用第二种方法分析“A”-非“A”检验数据，公式2中的 T 统计量通过公式A.6计算得到：

$$T_1 = \frac{\sqrt{160}[(60)(28) - (32)(40)]}{\sqrt{(92)(68)(100)(60)}} = 0,83 \quad (\text{A. 6})$$

与检验统计量 T 相关的 p 值为 $P = 0.2044$ ，其大于研究者选择的显著性水平 $\alpha = 0.05$ ，因此，可以得出结论，在95%置信水平下，“A”和非“A1”产品之间存在显著的差异。

表A.4 例3非“A2”产品观测值

评价员的响应	提供的样品为“A”	提供的样品为非“A2”	合计
评价员将样品标识为“A”	60	38	98
评价员将样品标识为非“A”	40	22	62
合计	100	60	160

使用第一种方法分析“A”-非“A”检验数据，“A”样品的非“A”响应比例（40%）大于非“A2”样品的非“A”响应比例（37%），则停止分析并得出没有足够的证据表明“A”和非“A2”产品之间存在显著的差异的结论。

使用第二种方法分析“A”-非“A”检验数据，公式2中的 T 统计量通过公式A.7计算得到：

$$T_1 = \frac{\sqrt{160}[(60)(22) - (38)(40)]}{\sqrt{(98)(62)(100)(60)}} = -0,42 \quad (\text{A. 7})$$

与检验统计量 T 相关的 p 值为 $P = 0.6624$ ，其大于研究者选择的显著性水平 $\alpha = 0.05$ ，因此，可以得出结论，在95%置信水平下，“A”和非“A2”产品之间存在显著的差异。

表A.5 例3非“A3”产品观测值

评价员的响应	提供的样品为“A”	提供的样品为非“A3”	合计
评价员将样品标识为“A”	60	12	72
评价员将样品标识为非“A”	40	28	68
合计	100	40	140

使用第一种方法分析“A”-非“A”检验数据，“A”样品的非“A”响应比例（40%）小于非“A3”样品的非“A”响应比例（70%），因此，分析可以进入第二步。公式1中的 T 统计量通过公式A.8计算得到：

$$T = \frac{(60-51,43)^2}{51,43} + \frac{(12-20,57)^2}{20,57} + \frac{(40-48,57)^2}{48,57} + \frac{(28-19,43)^2}{19,43} = 10,29 \quad (\text{A. 8})$$

与检验统计量 T 相关的 p 值为 $P = 0.0013$ ，其小于研究者选择的显著性水平 $2\alpha = 0.10$ ，因此，可以得出结论，在95%置信水平下，“A”和非“A3”产品之间存在显著的差异。

使用第二种方法分析“A”-非“A”检验数据，公式2中的 T 统计量通过公式A.9计算得到：

$$T_1 = \frac{\sqrt{140}[(60)(28)-(12)(40)]}{\sqrt{(72)(68)(100)(40)}} = 3,21 \quad (\text{A. 9})$$

与检验统计量 T_1 相关的 p 值为 $P = 0.0007$ ，其小于研究者选择的显著性水平 $\alpha = 0.05$ ，因此，可以得出结论，在95%置信水平下，“A”和非“A3”产品之间存在显著的差异。

注意，在例3中，“A”产品的数据在分析每一个非“A”产品时被重复使用。

附录B

(资料性附录)

 χ^2 和标准正态分布表摘录

表B.1给出了自由度为1的 χ^2 分布及与“A” - 非“A”检验相关的标准正态分布的临界值。并得出结论，如果公式1中的检验统计量 T 或公式2中的检验统计量 T 超过了表中与检验中选择的显著性水平 α 相关的临界值，则检验样品之间有显著的差异。

表B.1 临界值

显著性水平(α)	0.20 (20 %)	0.10 (10 %)	0.05 (5 %)	0.025 (2.5 %)	0.01 (1.0 %)	0.005 (0.5 %)
χ^2 临界值	0.71	1.64	2.71	3.84	5.02	6.63
标准正态临界值	0.84	1.28	1.64	1.96	2.33	2.58

附录C

(资料性附录)

“A” - 非“A”检验回答表格式样

C.1 只有样品“A”被提前出示

样品:	日期:
评价员:	
<p>1. 品尝样品“A”，并将容器交还负责人。取出编码样本。</p> <p>2. 由“A”和非“A”组成的编码样品顺序是随机的，所有非“A”样品均为同类样品。两种样品的具体数目事先不告知。</p> <p>3. 按顺序将样品一一品尝并将判断记录在下面。</p>	
样品编码	样品为
	“A” 非“A”
.....	<input type="checkbox"/> <input type="checkbox"/>
备注	
.....	
.....	

C.2 样品“A”和非“A”都被提前出示

样品:	日期:
评价员:	
<p>1. 品尝样品“A”和非“A”，并将容器交还负责人。取出编码样本。</p> <p>2. 由“A”和非“A”组成的编码样品顺序是随机的，所有非“A”样品均为同类样品。两种样品的具体数目事先不告知。</p> <p>3. 按顺序将样品一一品尝并将判断记录在下面。</p>	
样品编码	样品为
	“A” 非“A”
.....	<input type="checkbox"/> <input type="checkbox"/>
备注	
.....	
.....	

参考文献

- [1] Bi J., & Ennis D.M. The power of the A-Not A method. *J. Sens. Stud.* 2001, 16 pp. 343–359
- [2] Bi J., & Ennis D.M. Statistical methods for the A-Not A method. *J. Sens. Stud.* 2001, 16 pp. 215–237
- [3] Bi J. *Sensory discrimination tests and measurements—statistical principles, procedures and tables.* Blackwell Publishing, 2006
- [4] Brockhoff P.B., & Christensen R .H.B. 2010, Thurstonian models for sensory discrimination tests as generalized linear models. *Food Qual. Prefer.* 2010 April, 21 (3) pp. 330–338
- [5] Conover W.J. *Practical Nonparametric Statistics.* Wiley Publishing, Third Edition, 1998
- [6] Kim M.-A., Chae J.-E., van Hout D., Lee H.-S. 2012, Discriminations of the A–Not A difference test improved when “A” was familiarized using a brand image. *Food Quality and Preference*, Volume 23, Issue 1, January 2012, pp. 3-12
- [7] Lee H.-S., van Hout D., Hautus M.J. 2007, Comparison of performance in the A–Not A, 2-AFC, and same–different tests for the flavor discrimination of margarines: The effect of cognitive decision strategies. *Food Qual. Prefer.* 2007 September, 18 (6) pp. 920–928
-