

感官分析评价小组及成员表现评估技术动态分析

史波林¹, 赵 镭^{1,*}, 汪厚银¹, 支瑞聪¹, 奂 畅², 云战友³, 苏玉芳³

(1. 中国标准化研究院食品与农业标准化研究所, 北京 100191; 2. 上海大学计算机工程与科学学院, 上海 200444;
3. 内蒙古伊利实业集团股份有限公司技术中心, 内蒙古 呼和浩特 010110)

摘 要: 作为感官评价仪器的“评价员/评价小组”是获得可靠感官分析数据的关键, 其性能表现评估技术是有效管理该仪器的重要手段。本文将信度与效度作为该“仪器”的评估指标, 发现21世纪以后该类研究进入高峰期, 其超过85%的研究成果发表于*Food Quality and Preference*与*Journal of Sensory Studies*这两个感官研究类权威杂志中, 其中欧美国家在此技术中占主导地位。透过技术内容发现, 基于定量描述能力的评价小组及成员性能表现评估技术研究频繁并趋于成熟, 技术手段重点采用单参数或多参数方差分析、多元统计方法(主成分分析、广义普罗克分析等), 涌现了以PanelCheck、Compusense Five等为代表的评估软件, 同时国际标准化组织和美国材料与试验协会机构也进入相关标准研制阶段。而有关差别与排序的评估技术研究相对缺乏。本文提出性能评估所用样品性质及数量、性能评估所用数据表现形式、性能评估感官实验设计要求、性能评估期望值及置信区间、性能评估所用感官分析方法选择、性能评估的数学统计方法选择等关键要素将成为该领域未来努力的方向, 并有助于形成统一的、系统的感官分析评价小组及成员表现评估技术规范。

关键词: 感官分析; 评价小组; 评价员; 性能评估

Overview of Assessment Techniques for Sensory Panel and Panelist Performance

SHI Bo-lin¹, ZHAO Lei^{1,*}, WANG Hou-yin¹, ZHI Rui-cong¹, HUAN Chang², YUN Zhan-you³, SU Yu-fang³

(1. Food and Agriculture Standardization Institute, China National Institute of Standardization, Beijing 100191, China;
2. Department of Computer Engineering and Science, Shanghai University, Shanghai 200444, China;
3. Technology Center, Inner Mongolia Yili Industrial Group Co. Ltd., Hohhot 010110, China)

Abstract: In sensory evaluation, panels and panelists are used as instruments that are the key tool to obtain reliable sensory data. Assessment techniques for their performance can provide an important approach for managing them effectively. In this paper, the validity and reliability are used as the main indexes for performance assessment, which have reached a plateau of research since the beginning of the 21st century. More than 85% of the research papers have been reported in the top journals in sensory science, such as *Food Quality and Preference* and *Journal of Sensory Studies*. Especially, the majority of the research reports come from the USA and Europe and have tended to be frequent and mature in assessing the performance of descriptive sensory panels-panelists. The various methods include univariate or multivariate variance analysis, and multivariate statistical methods such as principal component analysis (PCA), generalized procrustes analysis (GPA), etc. At the same time, softwares such as PanelCheck, Compusense Five, and so on have been developed, while the International Organization for Standardization and the American Society for Testing and Materials are also formulating the relevant standards on quantitative descriptive analysis. By contrast, relatively insufficient research has been done on performance assessment in ranking and discrimination sensory panels-panelists. Six key research contents are proposed for future study. They include the number and property of sample, the form of data presentation, the design of sensory experiment, expected value and confidence interval, and the selection of sensory analysis methods and statistical methods. This paper will be helpful to establish the unified and systemic framework for assessing performance in panels and panelists.

Key words: sensory evaluation; panel; panelist; performance assessment

中图分类号: TS207.3

文献标志码: A

文章编号: 1002-6630(2014)08-0029-07

doi:10.7506/spkx1002-6630-201408006

收稿日期: 2014-03-24

基金项目: 中国标准化研究院院长基金项目(562013Y-3079)

作者简介: 史波林(1981—), 男, 副研究员, 博士, 研究方向为食品感官评价与智能感官分析。E-mail: shibl@cnis.gov.cn

*通信作者: 赵镭(1968—), 女, 副研究员, 博士, 研究方向为食品感官分析标准化。E-mail: zhaolei@cnis.gov.cn

感官分析是用于唤起、测量、分析和解释产品通过视觉、嗅觉、触觉、味觉和听觉所引起反应的一种科学方法^[1]。其原始测量数据直接来源于感官分析仪器“评价小组及其评价员”的评价结果。经过筛选和培训的“仪器”具有更好的灵敏度、更稳定的评价,表现出更好的重复性和再现性,体现出更小的变异性^[2]。然而,人不是模式化的、相对封闭的物理性机器,而是主动的、开放的、发展的、变化的生物体系^[3],易受生理因素(如感官适应、感觉增强、协同或抑制等)、心理因素(如期望误差、刺激误差、时序误差、光环效应等)和环境因素(如评价间的温度、湿度、噪声、光线等)等影响^[4],使得对食品感官性状鉴别有争议时往往难以下结论,从而影响感官评价结果的准确性、客观性和一致性。

同样作为一门测量技术的感官分析,也像其他分析检测一样要求其测量结果的信度与效度^[5]。其中信度为同一被测样在相同的实验条件下应该得到相近的结果,体现结果精密性,反应结果的可信度或可靠性或稳定性。在感官评价中,信度包含重复性与再现性。重复性为同一评价员或评价小组在相同时间(包括同天、同月等)、相同环境下对同一样品重复感官评价结果的离散情况;再现性为评价小组内部不同评价员之间或不同评价小组之间对同一样品感官评价结果的离散情况。而效度是指测量工具或手段能够准确测出所需测量事物的正确程度,即测量结果与要考察内容内在特征的一致程度,体现结果的准确性或正确性,反应结果的有效性。感官评价检测中的效度为评价员或评价小组单次测量与被测内容本质的一致性,即为评价员测试结果与理论值或评价小组得出的最优估计值的离散情况,以及评价小组测试结果与理论值的离散情况。

为了保证感官检验结果的可靠性、重复性和准确性,需要对人的感官能力提出严格的要求,进而才能在可控的条件下规范化地开展感官检验活动^[6]。感官评价员及评价小组的表现评估技术是体现“仪器”可用性特点的技术保障,它能指导评价小组与评价员在投入使用前进行校准并达到要求的精确度,也能帮助评价小组与评价员在使用一定周期后进行定期检定以符合检测的规范要求,保证检测结果的有效性或正确性。该技术是实现感官评价数据价值的关键保障,是反映感官分析实验室检测能力水平的重要手段,是构成感官分析实验室能力建设与认可的主要内容。因此,感官分析实验室的评价小组及成员表现评估技术能有效管理“感官分析仪器”以保持良好的状态,达到通过该仪器检测所获得数据的可靠性,保证感官分析科研、实验和生产的需要,也有力助推感官分析技术的广泛应用。

1 理化仪器评估技术不能直接应用于感官分析仪器

针对理化检测仪器设备性能评估及其实验室间能力比对技术已相当成熟,并形成了系列规范技术体系,如《检测和校准实验室能力认可准则》^[7]、《检测和校准实验室能力认可准则在化学检测领域的应用说明》^[8]、《能力验证规则》^[9]、《检测和校准实验室能力的通用要求》^[10]等,就仪器设备校准与检定、仪器设备期间核查、实验室能力比对与验证、设备维护、设备量值溯源、测量不确定度等均有详细说明与规定。

然而这些文件中的规定与方法不适用于感官评价小组及成员的表现评估及能力比对,图1对比了基于理化的仪器测量与基于心理的感官测量的特征区别。感官评价是基于感觉差别的一种心理测量,属于实验心理学的范畴,是一种基于样品间相对差别的比较检验和测量,而不是理化检验中的绝对物理量的测量^[11]。因此,感官评价活动可分成有无差别、差别程度与差别方向这3个层次测量。具体包括不同样品之间整体品质和特定感官属性的差别检验、样品感官特性强度与消费者喜好的标度检验、产品评分和分等分级的类别检验、以及感官质量特征确定的描述性分析检验等^[12]。这些检测结果的数据特征分别为:名义数据(如:有无差异或分类)、顺序数据(如:从弱到强的排序)、数值数据(如:评分或定量描述)。由此,仪器测量与感官测量在信度与效度的评估技术方面有很大的不同。

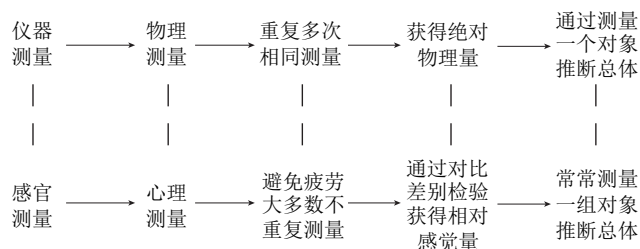


图1 仪器测量与感官测量的特征比较

Fig.1 Comparison of features between instrumental and sensory testing

2 国内外技术动态整体分析

通过Elsevier、Springer Link、Wiley-Blackwell、EBSCOhost、ACS、CUP等数据库检索到86篇(1979—2014年)有关评价员及评价小组表现评估技术的SCI收录论文。从发表论文的总量上讲,是研究很窄的一个方向。为了进一步把握在此方向上研究的风向标,按照研究时间分析热门年度,从论文来源上分析成果出处以及根据研究的区域性分析核心团队等。

早在1979年美国的Hovenden等^[13]采用8点标度感官分析牛柳品质时涉及到评价小组的重复性与差异性问题,

并认为评价小组的稳定性好于评价员个体。但在20世纪90年代对评价员及评价小组表现评估研究才真正进入萌芽阶段并逐渐展开,在此期间累计发表论文20篇(图2),并于1998年单年发表7篇相关论文而达到高峰。进入21世纪相关研究全面发展、不断深入,并且在2006年后发展更为活跃,特别是2008年单年发表论文8篇。虽然2006—2010年属于研究高峰期(发表论文24篇),但2010年后还是有研究者继续相关技术研究,并越来越系统与成熟,同时陆续研制国际标准化组织(International Organization for Standardization, ISO)与美国材料与试验协会(American Society for Testing and Materials, ASTM)相关标准。

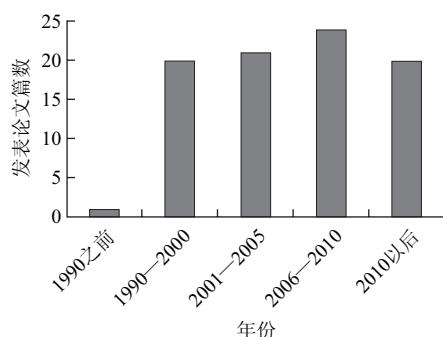


图2 发表论文年度分布图

Fig.2 Annual distribution diagram of the published papers

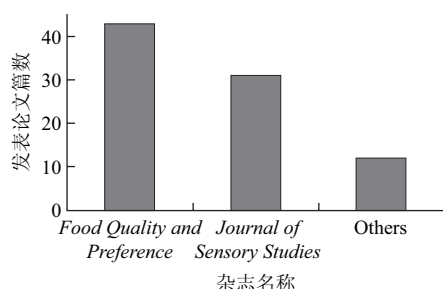


图3 发表论文杂志出处分布图

Fig.3 The journals in which the papers were published

在所检索到的87篇文章中,85%以上发表于目前展示感官科学相关研究成果最为权威的*Food Quality and Preference* (43篇)与*Journal of Sensory Studies* (31篇)这两个杂志中(图3),由此说明这两个杂志中所发表的文章代表了评价员及小组表现评估技术的发展历史、现有研究水平及未来的发展趋势,同时每个时间段前者几乎都比后者的发表文献多(除了2001—2005年)(图4)。不过其他文章也发表在如*Journal of Chemometrics*、*Food Control*、*European Food Research Technology*、*LWT-food Science and Technology*、*Journal of the Science of Food and Agriculture*、*Journal of Food Science*等知名的食品与化学计量学领域相关杂志中。

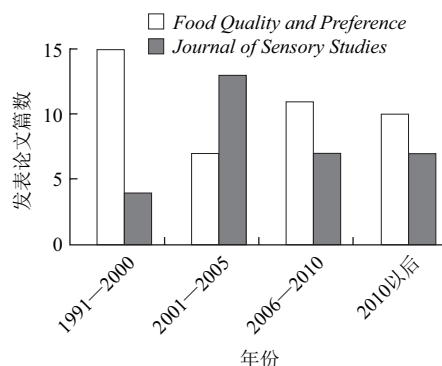
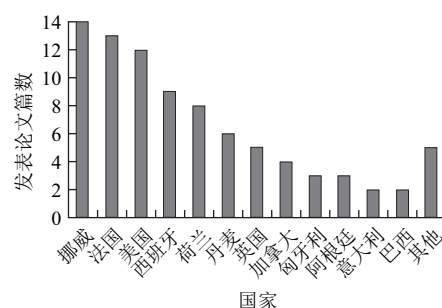
图4 在*Food Quality and Preference*与*Journal of Sensory Studies*发表论文年度分布表Fig.4 Annual distribution diagram of papers published in *Food Quality and Preference* and *Journal of Sensory Studies*

图5 发表论文第一作者所属国家分布表

Fig.5 The first author's countries of the published papers

对发表文章第一作者所在国的归类,发现欧美在该技术研究中占主导地位,因为他们是感官科学研究的发源地,也是感官分析技术应用频繁的葡萄酒、化妆品、其他食品等工业非常发达的国家。特别是挪威(14篇)、法国(13篇)、美国(12篇)、西班牙(9篇)、荷兰(8篇)、丹麦(6篇)、英国(5篇)等国家(图5)。在国际上一直致力于这方面研究的科研团队主要为来自挪威的Næs (10篇)^[14-23]与Lea (3篇)^[15,24-25]、英国的McEwan (5篇)^[24-28]、丹麦的Martens (5篇)^[19,29-32]与Brockhoff (4篇)^[33-36]、美国的Bi (4篇)^[37-40]、匈牙利的Kókai (3篇)^[41-43]等7个团队(按照发表论文参与作者统计)。当然这些研究团队也研究其他感官分析技术内容。同时发现作为研究非常活跃的法国,居然没有一个相对固定的团队长期致力于这方面的研究,因为没有一位法国学者在这方面的发表论文数大于等于2篇。

而国内在这方面的研究比较落后,这跟我国感官分析技术发展水平息息相关。目前只有3篇相关文献报道。其中李华等^[44]在葡萄酒感官评价结果的统计分析方法研究中发现,品酒员间存在评价尺度、评价位置和评价方向等方面的差异,而标准化法不仅没有消除品酒员间的异质性,反而加大了品酒员间的差异;而置信区间法能有效地降低品酒员间的差异,真实地反映酒样间的客观

差异。苏玉华等^[45]利用灰色关联性分析两组评酒员的葡萄酒感官评价结果的显著性差异,通过克朗巴哈 α 系数(Cronbach's alpha coefficient)分析每组评酒员评价结果的可信度。常玉梅等^[46]在豆腐干质构感官分析及评价小组能力评估研究中,通过计算信噪比变化趋势分析了培训过程中评价小组的表现;采用Panel Check软件^[21],评估了培训后的评价小组一致性、评价员区分能力和重复性。

3 基于定量描述的评估技术趋于成熟

仪器测量的评估方法有很多可借鉴的技术。目前有关评价员及评价小组表现评估技术的研究,基本上都围绕具有量值特征的描述性分析方面。因为该感官分析评价方法所获得检测数据是定量的,能最大程度参考理化仪器的性能评估方法,其技术方法特征主要包括定性分析与定量分析方法、单变量与多变量方法、参数与非参数方法、分析型与图表型方法、传统计算与计算机软件评估法^[40]。在研究的应用方面各有侧重,包括实验室间多个评价小组结果的比对^[27,47-48]、评价小组内异常评价员分析技术^[22,35]、评价小组结果可靠性(重复性、一致性等单一评价内容或综合评价)^[2,23,49]和基于性能的评价队伍长期监测^[50-52]等。但其技术核心大多是基于方差分析的统计方法与基于多元统计方法两个方面。

3.1 基于方差分析的性能评估技术动态

在方差分析方法使用方面,具有代表性的首先是长期致力于混合评价员模型(mixed assessor model, MAM)研究的Brockhoff等^[33-36],针对定量性描述分析检验,在多因素方差分析方法的基础上增加了评价员标度使用影响因素,提出了MAM评估评价小组及成员区分能力、一致能力、重复能力和标度能力的方法。Pineau等^[53]采用了均方根误差(RMSE)、相对预测误差(REP)等性能评估参数,建立同时监测评价小组及成员表现的混合模型,由监测图直观展现评价员性能的动态表现与异常评价值。Etaio等^[51]在红酒感官质量控制中,考察了评价员对参比样和样品特征的识别力,样品间特征的区别力、重复性、再现性;对于评价小组重点评估产品品质评分与辨别、感官特性识别与区分方面的重复性和再现性指标;但其主要统计方法都为方差分析。

3.2 基于多元统计方法的性能评估技术动态

其实在多元统计方法使用方面,早在20世纪初King等^[54]就比较了聚类分析(CA)、一致性分析、主成分分析(PCA)、GRAPES法4种方法对评价小组的性能评价,研究得出互相组合使用最为理想,但研究中并未提到评价小组结果与产品特征最优估计值或期望值的一致性。Castura等^[50]提出采用4种与产品感官特

征期望值的距离检测分析描述性评价小组能力监控。Næs等^[14,19-21]也长期研究评价小组及成员评估技术,并开发了针对感官剖面数据的评价小组性能分析软件(PanelCheck),主要采用经典方差分析(ANOVA)、PCA、多因素分析(MFA)、广义普罗克分析(GPA)等方法用于实验室内外部的感官分析能力验证,并通过Profile、Tucker-1、Manhattan、MSE、 F 值等多种可视化图来展现评价性能。

3.3 针对描述性分析的性能评估软件动态分析

目前市面上有多款评价小组及评价员性能表现评估方面的软件。其中前面提到的PanelCheck软件最具有代表性、针对性和实用性。荷兰OP&P公司的Senstools软件主要采用方差分析与多元统计分析(PCA、GPA、MDPref、聚类分割法等)。而法国Biosystemes公司FIZZ不如前面两款软件全面和专业,它主要应用在质地描述分析中,通过对样品属性的统计,顺带对评价员的区别能力进行检测。加拿大Compusense公司的Compusense Five软件是在感官评价实验设计计算机管理软件的基础上升级的,其含有2-way ANOVA、Friedman分析、Tukey's HSD、Fisher's LSD、Duncan's Multiple Range检验与Crosstabulations等各种数学统计方法,专门应用于排序能力与定向描述分析能力的评价员及小组性能评估^[50]。相关的软件还有挪威Camo公司的Unscrambler^[55]、荷兰Logic8 BV公司的EyeQuestion与美国Tragon公司的Tragon QDA^[56]。同时一些数据统计软件也可以应用于此领域,比如Senpaq、SAS、SPSS、Excel(特别是带有XLStat辅助工具的)等。

3.4 针对描述性分析的性能评估标准动态分析

相关性能评估标准的研制工作主要针对描述性分析方法。国际标准化组织食品技术委员会感官分析分技术委员会(ISO/TC34/SC12)正在研制基于方差分析的定量感官评价小组及其成员表现评估的一般导则,并于2010年形成国际标准草案(Draft International Standard, DIS)^[57],但截止到今年还未形成终稿出版。DIS稿主要是针对某一感官属性在多个样品的多次重复测试中的性能评估方法,包括:1)通过单因素方差分析评价员各自的重复性;2)通过评价员多次重复后均值与评价小组均值的标准偏差体现两者之间的接近程度;3)通过评价员标度值与评价小组平均标度值的回归模型性能来体现评价员的描述量值能力;4)通过评价员与样品双因素方差分析评价小组的重复性;5)小组正确性;6)通过三因素(样品、评价员与轮次)方差分析评价小组多轮次实验的再现性;7)通过三因素(样品、评价小组、轮次)方差分析不同评价小组间的再现性。ISO/DIS稿主要是对于各属性单独分析评价小组及成员性能评估,而对于样品多属性整体评价时的评价员与评价小组的性能

分析,标准中只提到了几种多元统计方法(PCA、DA、GPA),但未具体展开说明。

美国材料与试验协会感官分析技术委员会(ASTM/E18)也通过工作项目正在研制相关标准^[58],此标准也是针对描述性分析方法,其技术内容与ISO标准相同之处也提到了评价员及小组个体的重复性、不同评价员或不同评价小组之间的再现性、特性标度值的正确性、以及不同样品在特定属性上的区分能力(敏感性)。不同之处,ASTM标准中提到了不同评价小组或不同评价员之间对特性描述理解的一致性与对不同样品在某特性上的强度排序一致性问题,其实这两项指标同样也可以通过重复性、再现性与正确性来体现。与ISO标准最大的不同在于详细介绍了每种多元统计方法(同向缩放因子、PCA、CA、GPA)的使用范围与基本概念。

4 有关差别与排序的评估技术研究相对缺乏

感官分析方法除了描述性分析外,还有差别检验与排序法等。而目前有关这两个能力的评价小组及成员性能评估技术研究非常少。McEwan等^[24,28]于1998—2001年期间在欧洲范围开展了感官分析实验室间比对研究,并首次正式提出感官分析能力验证相关技术问题,虽然重点研究不同实验室间不同评价小组的能力比对,而未全面涉及评价员个体的性能评价,但对于该方向研究具有里程碑意义。在感官剖面能力验证中,提到如何获得最优期望值,如何获得用于评价小组判断的性能标准与置信区间。同时还采用Friedman检验、单边 t 检验分别分析了排序法与9点标度的能力验证。在2003年,McEwan等^[25]专门研究了感官排序检测的实验室比对,是目前唯一系统的介绍评价小组排序性能比对的研究。该研究重点形成了这类能力比对的技术流程与每个环节的判别标准。包括由Pearson相关系数法确定校准评价小组所建立的期望样品排序及每个样品的秩次平均值,由Friedman检验各评价小组的样品区分能力,由Conover多重比较法检验各评价小组区分样品对个数,由Kendall系数评估评价小组内部评价员之间的一致性,以及评价小组排序能力的整体性能。文章最后展望了样品选择的重要性,各性能判别标准阈值的随意性,以及设置期望排序能力置信区间的价值等问题。

2012年Sauvageot等^[48]在对9个实验室3类级别评价小组(消费者级、初选级、优选级)的三点检验结果进行比对研究中,虽然也提到了评价小组的样品间区分力、小组内部的重复性以及不同小组间的一致性等问题,但整个研究重点分析了不同性能评价小组所带来的小组评价结果对比,对于评价小组及成员差别能力评估技术研究的指导意义不是很强。Bi等^[40]首次系统比较所有目前

提到的有关评价员及小组性能评估的指标,并提出采用社会行为学测量方法中的信度与效度最为贴切;同时首次统一采用组内相关系数(ICC)建立应用于感官分析中出现的所有7类数据(评分法的连续数据、多元连续数据、分级数据、排序数据、二项式分布数据、多项选择数据、强迫选择数据)下的评价小组及成员性能评价方法,并由此确立了针对不同性质数据的评价员及小组评估统一导则,特别对差别与排序的评估技术具有重要指导意义。

5 展望

虽然在2000年以后,感官评价小组及成员表现评估技术研究频繁,但基本集中在具有数值化特征的描述性分析方法中,并且所采用数学技术方法繁多、不具统一性,也未形成固定的、系统的技术规范。针对差别区分能力或排序能力的评价小组及成员表现评估技术研究匮乏,与该两类感官分析方法的广泛应用现状极其不匹配。

由于感官分析方法中的差别检验、排序检验与量值性描述分析等分别获得不同性质检测数据(二项式分布数据、秩次数据和量值数据),因此需要分别对差别能力、排序能力、定量描述能力开展感官评价小组与成员性能表现评估技术研究,并建立对应的技术规范与准则。由此需要重点在性能评估所用样品性质及数量、性能评估所用数据表现形式、性能评估实验设计要求、性能评估期望值及置信区间、性能评估所用感官分析方法选择、性能评估的数学统计方法选择等方面加以深入,并作为未来的研究方向。

5.1 性能评估所用样品性质及数量

选择怎样的样品用于评价小组及成员性能表现的测试需要深入思考。通用型感官分析实验室应该倾向于基本味、通用香气与质地等样品;而对于具体产品生产公司,建议选择公司热销产品或类似的模拟体系物质。

感官分析是被检样品差异难度与评价成员灵敏度之间的博弈,因此在评价小组及成员性能表现评估中,样品难度的选择非常重要。若样品难度过低,所有评价小组及成员的每次评价结果一致并正确,而未真正达到性能考察的作用;若样品难度过大,几乎所有评价小组及成员的每次评价结果都不正确,性能都非常差,有时差到分辨不出哪些结果优于其他结果,这样也失去性能评估的意义。而在排序能力的评价小组及成员性能表现分析中,需要考虑几个特定性质强度不同的样品参与排序实验评估中更加合适,更能便捷的监测感官分析“仪器”性能。

5.2 性能评估所用数据表现形式

特别对于差别检验,其评价小组或评价员测试结果

为对立结果的其中一个（A或非A、正确或错误、有差异或无差异等），而这不能作为表现性能评估的基础数据，需要通过一定方法进行转化而起到科学、方便的评估目的。

在排序检验中，能否简单的根据所有评价员秩次和排序代表评价小组得出的样品顺序，或以怎样的形式代表评价小组每次的排序结果更具有科学性（总秩次已受评价员人数影响）。

5.3 性能评估感官实验设计要求

在重复性、再现性、一致性评估中，多少次的重复或再现能在保证实验成本的前提下达到统计意义，以及多长时间开展一次重复能最高效率的监测评价小组成员的表现性能。在描述性分析中，需要设置几种不同样品用于实验。每次实验需要安排几轮，每轮安排几次重复，同时重复实验之间的样品如何制备以避免评价员在非检测感官特性方面对样品的记忆，以及感觉疲劳带来的实验误差。

5.4 性能评估期望值及置信区间

为了更有效的决策单次测量与被测内容本质的一致性，以及评价小组测试结果与理论值的离散情况，如何获得评判依据、期望值或真值，以及所能接受的相应置信区间。比如在无理论结果的排序检验中，如何通过评价小组获得最优估计顺序，同时在此期间如何判断评价员异常判断结果并加以剔除。在描述性分析的实验室比对中，如何选择可靠的评价小组用于期望值获取的校准小组。差别检验中，任何样品都不是完全一致的，但这种绝对的不一致能否让评价员团队都能感知到，并又以哪个评价小组团队的检测结果代表不同样品间的差异程度。

5.5 性能评估所用感官分析方法选择

感官分析方法既是评价样品的方法，也是评估评价小组及成员的途径，关键是在已知样品性能的前提下，考察评价小组及成员的检测能力。然而针对差别能力的性能评估，由于差别检验包括成对比较、二三点、三点、A-非A、五中取二等方法，究竟哪种差别检验方法更加适合差别能力下的评价小组及成员的性能表现评估值得探究。

5.6 性能评估的数学统计方法选择

针对相对差别测量特征的感官评价，包含有无差别、差别程度与差别方向这3个层次，其不同层次感官分析方法所获得的数据特征各不相同，含有名义数据、顺序数据、数值数据，也就是分别为二项式分布数据、秩次数据、量值数据，因此针对这些不同数据特征所要进行评价小组及成员性能评估的数学统计方法也各不相同。

参考文献：

[1] STONE H, SIDEL J L. 感官评定实践[M]. 陈中, 陈志敏, 唐传核, 等. 译. 北京: 化学工业出版社, 2007: 84-90.

- [2] PINTO F S T, FOGLIATTO F S, QANNARI E M. A method for panelists' consistency assessment in sensory evaluations based on the Cronbach's alpha coefficient[J]. Food Quality and Preference, 2014, 32: 41-47.
- [3] 李宏, 刘锐萍, 张克义. 食品感官检测实验室评审方法的研究[J]. 现代检测与实验室管理, 2013(1): 39-41.
- [4] 赵镭, 刘文. 感官分析技术应用指南[M]. 北京: 中国轻工业出版社, 2011: 3.
- [5] 郑日昌, 孙大强. 心理测量与测验[M]. 北京: 中国人民大学出版社, 2008: 63; 102.
- [6] 赵镭, 刘文, 牛丽影, 等. 食品感官科学技术: 发展的机遇和挑战[J]. 中国食品学报, 2009, 9(6): 138-143.
- [7] 中国合格评定国家认可委员会. CNAS-CL10:2006 检测和校准实验室能力认可准则[S]. 北京: 中国合格评定国家认可委员会, 2006.
- [8] 中国合格评定国家认可委员会. CNAS-CL10:2012 检测和校准实验室能力认可准则在化学检测领域的应用说明[S]. 北京: 中国合格评定国家认可委员会, 2012.
- [9] 中国合格评定国家认可委员会. CNAS-RL02:2007 能力验证规则[S]. 北京: 中国合格评定国家认可委员会, 2007.
- [10] 中国合格评定国家认可中心, 国家认证认可监督管理委员会, 中国计量科学研究院, 等. GB/T 27025—2008 检测和校准实验室能力的通用要求[S]. 北京: 中国标准出版社, 2008.
- [11] 董小雷, 周广田, 崔云前. 啤酒感官品评[M]. 北京: 化学工业出版社, 2007: 4.
- [12] 徐树来, 王永华. 食品感官分析与实验[M]. 北京: 化学工业出版社, 2010: 15.
- [13] HOVENDEN J E, DUTSON T R, HOSTETLER R L, et al. Variation and repeatability of an untrained beef sensory panel[J]. Journal of Food Science, 1979, 44: 1598-1601.
- [14] NÆS T, SOLHEIM R. Detection and interpretation of variation within and between assessors in sensory profiling[J]. Journal of Sensory Studies, 1991, 6(3): 159-177.
- [15] LEA P, RODBOTTEN M, NÆS T. Measuring validity in sensory analysis[J]. Food Quality and Preference, 1995, 6(4): 321-326.
- [16] NÆS T. Detecting individual differences among assessors and difference among replicates in sensory profiling[J]. Food Quality and Preference, 1998, 9(3): 107-110.
- [17] NÆS T, LANGSRUD O. Fixed or random assessors in sensory profiling[J]. Food Quality and Preference, 1998, 9(3): 145-152.
- [18] DAHL T, NÆS T. Outlier and group detection in sensory panels using hierarchical cluster analysis with the Procrustes distance[J]. Food Quality and Preference, 2004, 15: 195-208.
- [19] TOMIC O, NILSEN A, MARTENS M, et al. Visualization of sensory profiling data for performance monitoring[J]. LWT-Food Science and Technology, 2007, 40(2): 262-269.
- [20] DAHL T, TOMIC O, WOLD J P, et al. Some new tools for visualizing multi-way sensory data[J]. Food Quality and Preference, 2008, 19(1): 103-113.
- [21] TOMIC O, LUCIANO G, NILSEN A, et al. Analysing sensory panel performance in a proficiency test using the PanelCheck software[J]. European Food Research and Technology, 2009, 230(3): 497-511.
- [22] DAHL T, NÆS T. Identifying outlying assessors in sensory profiling using fuzzy clustering and multi-block methodology[J]. Food Quality and Preference, 2009, 20: 287-294.
- [23] TOMIC O, FORDE C, DELAHUNTY C, et al. Performance indices in descriptive sensory analysis: a complimentary screening tool for assessor and panel performance[J]. Food Quality and Preference, 2013, 28: 122-133.
- [24] McEWAN J A, HUNTER E A, GEMERT L J, et al. Proficiency

- testing for sensory profile panels: measuring panel performance[J]. Food Quality and Preference, 2002, 13(3): 181-190.
- [25] McEWAN J A, HEINIO R L, HUNTER E A, et al. Proficiency testing for sensory ranking panels: measuring panel performance[J]. Food Quality and Preference, 2003, 14(3): 247-256.
- [26] RISVIKUAFORSK H, COLWILL J S, McEWAN J A, et al. Multivariate analysis of conventional profiling data: a comparison of a British and a Norwegian trained panel[J]. Journal of Sensory Studies, 1992, 7: 97-118.
- [27] HUNTER E A, McEWAN J A. Evaluation of an international ring trial for sensory profiling of hard cheese[J]. Food Quality and Preference, 1998, 9(5): 343-354.
- [28] McEWAN J A. Comparison of sensory panels: a ring trial[J]. Food Quality and Preference, 1999, 10(3): 161-171.
- [29] THYBO A K, MARTENS M. Analysis of sensory assessors in texture profiling of potatoes by multivariate modeling[J]. Food Quality and Preference, 2000, 11: 283-288.
- [30] BYRNE D V, O'SULLIVAN M G, DIJKSTERHUIS G B, et al. Sensory panel consistency during development of a vocabulary for warmed-over flavor[J]. Food Quality and Preference, 2001, 12: 171-187.
- [31] BITNES J, UELAND Ø, MØLLER P, et al. Reliability of sensory assessors: issues of retention and learning[J]. Journal of Sensory Studies, 2008, 23(6): 852-870.
- [32] BITNES J, UELAND Ø, MØLLER P, et al. Reliability of sensory assessors: issues of complexity[J]. Journal of Sensory Studies, 2009, 24(1): 25-40.
- [33] BROCKHOFF P B. Statistical testing of individual differences in sensory profiling[J]. Food Quality and Preference, 2003, 14(5): 425-434.
- [34] DEHLHOLM C, BROCKHOFF P B, BREDIE W L P. Confidence ellipses: a variation based on parametric bootstrapping applicable on multiple factor analysis results for rapid graphical evaluation[J]. Food Quality and Preference, 2012, 26: 278-280.
- [35] BAVAY C, BROCKHOFF P B, KUZNETSOVA A, et al. Consideration of sample heterogeneity and in-depth analysis of individual differences in sensory analysis[J]. Food Quality and Preference, 2014, 32: 126-131.
- [36] PELTIER C, BROCKHOFF P B, VISALLI M, et al. The MAM-CAP table: a new tool for monitoring panel performances[J]. Food Quality and Preference, 2014, 32: 24-27.
- [37] ENNIS D M, BI J. The beta-binomial model: accounting for inter-trial variation in replicated difference and preference tests[J]. Journal of Sensory Studies, 1998, 13(4): 389-412.
- [38] BI J, ENNIS D M. The power of the "A"- "not A" method[J]. Journal of Sensory Studies, 2001, 16(4): 343-359.
- [39] BI J. Agreement and reliability assessments for performance of sensory descriptive panel[J]. Journal of Sensory Studies, 2003, 18: 61-76.
- [40] BI J, KUESTEN C. Intraclass Correlation Coefficient (ICC): A framework for monitoring and assessing performance of trained sensory panels and panelists[J]. Journal of Sensory Studies, 2012, 27(5): 352-364.
- [41] HUNEK K K, HESZBERGER J, KÓKAI Z, et al. Testing panel consistency with GCAP method in food profile analysis[J]. Journal of Chemometrics, 2008, 22(3/4): 218-226.
- [42] SIPOS L, KÓVACS Z, SZOLLOSI D, et al. Comparison of novel sensory panel performance evaluation techniques with e-nose analysis integration[J]. Journal of Chemometrics, 2011, 25(5): 275-286.
- [43] LOSÓ V, GERE A, GYÖREY A, et al. Comparison of the performance of a trained and an untrained sensory panel on sweet corn varieties with the PanelCheck software[J]. Applied Studies in Agribusiness and Commerce, 2012, 6: 77-83.
- [44] 李华, 刘曙东, 王华, 等. 葡萄酒感官评价结果的统计分析方法研究[J]. 中国食品学报, 2006, 6(2): 126-131.
- [45] 苏玉华, 韦师. 葡萄酒感官评价的差异性及可信度研究[J]. 河南科技, 2013(1): 254-259.
- [46] 常玉梅, 钟芳. 豆腐干质构感官分析及评价小组能力评估[J]. 食品与生物技术学报, 2013, 32(1): 37-42.
- [47] LE S, PAGES J, HUSSON F. Comparison of the performance of a trained and an untrained sensory panel on sweet corn varieties with the PanelCheck software[J]. Food Quality and Preference, 2008(19): 179-184.
- [48] SAUVAGEOT F, HERBRETEAU V, BERGER M, et al. A comparison between nine laboratories performing triangle tests[J]. Food Quality and Preference, 2012, 24(1): 1-7.
- [49] MEYNNERS M. Panel and panelist agreement for product comparisons in studies of temporal dominance of sensations[J]. Food Quality and Preference, 2011, 22: 365-370.
- [50] CASTURA J C, FINDLAY C J, LESSCHAEVE I. Monitoring calibration of descriptive sensory panels using distance from target measurements[J]. Food Quality and Preference, 2006, 17(3/4): 282-289.
- [51] ETAIO I, ALBISU M, OJEDA M, et al. Sensory quality control for food certification: a case study on wine. Panel training and qualification, method validation and monitoring[J]. Food Control, 2010, 21(4): 542-548.
- [52] ROMANO R, VESTERGAARD J S, ZAREH M K, et al. Monitoring panel performance within and between sensory experiments by multi-way analysis: classification and multivariate analysis for complex data structures[J]. Springer Berlin Heidelberg, 2011: 335-342.
- [53] PINEAU N, CHABANET C, SCHLICH P. Modeling the evolution of the performance of a sensory panel: a mixed-model and control chart approach[J]. Journal of Sensory Studies, 2007, 22(2): 212-241.
- [54] KING M C, HALL J, CLIFF M A. A comparison of methods for evaluating the performance of a trained sensory panel[J]. Journal of Sensory Studies, 2001, 16(6): 567-581.
- [55] KERMIT M, LENGARD V. Assessing the performance of a sensory panel-panelist monitoring and tracking[J]. Journal of Chemometrics, 2005, 19: 154-161.
- [56] 张爱霞, 张卫斌, 励建荣, 等. 现代食品感官科学的国际动态与发展趋势[J]. 中国食品学报, 2008, 8(6): 177-180.
- [57] International Organization for Standardization. ISO/DIS 11132—2010 Sensory analysis-Methodology-General guidance for monitoring the performance of quantitative sensory panel[S]. United States: Information Handling Services, 2010.
- [58] American Society for Testing and Materials. ASTM/WK 8435—2010 Standard guide for measuring and tracking sensory descriptive panel and assessor performance[S]. United States: Information Handling Services, 2010.