

护肤品感官评估方法实验设计与数据分析

朱少君

华美康妍(苏州)生物科技有限公司, 中国·江苏 苏州 558103

摘要: 在护肤品领域, 准确了解消费者对产品的感官体验至关重要。论文聚焦于护肤品的感官评估, 深入探讨了多种常用感官评估方法的实验设计要点、适用场景、评估人员要求、操作流程、结果分析方法以及数据分析的全过程。旨在为护肤品研发、质量控制和市场调研提供科学且实用的参考。

关键词: 感官评价; 评估; 护肤品; 三角评价法; 数据分析

Experimental Design and Data Analysis of Sensory Evaluation Methods for Skin Care Products

Shaojun Zhu

Huamei Kangyan (Suzhou) Biotechnology Co., Ltd., Suzhou, Jiangsu, 558103, China

Abstract: In the field of skin care products, accurately understanding consumers' sensory experience of products is of crucial importance. This paper focuses on the sensory evaluation of skin care products and deeply discusses the key points of experimental design, applicable scenarios, requirements for evaluators, operation procedures, result analysis methods, and the entire process of data analysis of various commonly used sensory evaluation methods. It aims to provide scientific and practical guidance for skin care product research and development, quality control, and market research.

Keywords: sensory evaluation; assessment; skin care products; triangle test method; data analysis

0 前言

随着消费者对护肤品需求的不断增长和品质要求的提高, 护肤品企业愈发重视产品的感官特性。通过科学合理的感官评估实验设计和精准的数据分析, 能够深入洞察消费者的需求和偏好, 为研发出更具竞争力的护肤品提供有力支撑。

1 感官评估实验设计的基本原则

①随机性原则, 在护肤品感官评估实验中, 确保样品的选择、评估人员的分组以及样品呈现的顺序等都是随机的, 以避免系统性偏差对评估结果的影响。

②重复性原则, 通过多次重复实验, 降低偶然因素的干扰, 增强结果的可靠性和稳定性, 特别是在护肤品功效评估方面, 多次重复能更准确地反映产品的真实效果。

③对照性原则, 设置合适的对照组, 如使用安慰剂或市场上的标杆产品, 有助于清晰地凸显出被评估护肤品的独特优势或潜在问题。

④均衡性原则, 在实验条件的设计上, 如环境温度、湿度、评估时间等, 应保持均衡, 以确保不同样品在相同的条件下接受评估, 减少外部因素对评估结果的干扰。

2 感官评估实验的分类

①区别性测试, 旨在检测不同护肤品之间在质地、气味、使用感受等方面是否存在显著差异。

②描述性测试, 对护肤品的各项感官特性进行详细、

准确的描述和量化, 如肤感的油腻程度、吸收速度、保湿效果、持续时间等。

3 常用护肤品区别性测试方法

3.1 排序法

应用范围: 作为筛选护肤品配方的初步手段, 为后续更深入的研发提供方向; 在多款护肤品中挑选出最符合市场需求或特定功效要求的产品; 用于选拔和培训对护肤品质地、气味等感官特征敏感且具有良好记忆能力的评价员; 适用护肤品数量, 一般 3~6 种护肤品。

评估员要求: 对于大规模的消费者检验, 理想情况下需要 30 个以上评估员, 以获取具有广泛代表性的结果。

操作步骤: 实验前, 向评估员清晰且详细地阐述被评价的护肤品指标和准则, 确保所有评估员对评估要点有一致且准确的理解; 评估员按照事先确定的随机顺序依次检验编码的护肤品样品, 并根据指定的感官属性(如质地的细腻程度、气味的清新程度等)的感受强弱初步排列样品; 允许评估员重新检验样品, 对初步的排序进行校验和调整。

结果的分析方法: 一般采用双向方差秩分析。

分析步骤: 将观测值按组从小到大顺序赋秩。每一行为一个组, 计算秩的数量。按列计算秩和, 依据公式 $\chi^2 = 12/nk(k+1) \sum R_i^2 - 3n(k+1)$ 计算统计量, 其中 n 表示评估人数, k 表示秩次, R_i 表示第 i 个样品的次序总和。

判断标准: 若 $\chi^2 >$ 临界值, 则拒绝原假设, 表明样品排名显著不同; 若 $\chi^2 <$ 临界值, 则接受原假设, 即样品

排名无显著不同。例如，假设有 10 位评估员对 4 款面霜的滋润度进行打分，然后求和后排序，排序结果如下：

对观测值进行计算秩和：

面霜 A 的秩和：

$$R1=2+2+1+3+2+2+3+2+2+2=21$$

面霜 B 的秩和：

$$R2=1+1+2+1+1+1+3+4+1+1=15$$

面霜 C 的秩和：

$$R3=4+4+3+4+4+4+2+4+3+4=36$$

面霜 D 的秩和：

$$R4=3+3+4+2+3+1+1+3+4+3=28$$

假设评估人数 $n=10$ ，秩次 $k=4$ ，代入公式计算统计量 χ^2 。

最后，根据 χ^2 分布临界值表和设定的显著性水平（如 95%），判断面霜之间的滋润度排名是否存在显著差异。

3.2 成对比较法

成对比较法的本质要求对每一个观测单元（个案）就两种护肤品做两次观测，形成配对样本，以更好地控制除研究的变化量外的其他变量，突出两款护肤品之间的差异。

应用范围：确定两种护肤品在质地、吸收速度、保湿效果等方面是否存在差异，并明确差异的方向；了解消费者对两款护肤品的偏爱倾向，为市场定位和推广策略提供依据；针对不同配方的面霜、乳液等进行主观评估；研究不同添加量的某护肤成分对产品使用效果的影响。也可用于选拔和培训评价员，评估其对护肤品特定感官属性的辨别能力。

成对比较法一般是直接比较 2 种护肤品，如竞品与自家产品的对比。对于 3 种护肤品，可以通过排列组合进行 3 次成对比较（AB、AC、BC），再对收集的数据进行综合统计处理，但当比较的护肤品数量进一步增加时，成对比较的组合数会急剧上升，导致操作难度加大。

评估员要求：理想情况下，需要 30 个以上经过半训练的评估员；实际操作中，12~18 个经过半训练的评估员通常也能满足基本要求；对于综合性的研究，如涉及大规模消费者偏爱检验的调研，则需要根据检验内容和要求配备更多的评估员。

操作步骤：以确定或随机的顺序将成对的护肤品分发给评估员；向评估员明确询问关于两款护肤品之间差别的具体问题，或者偏爱某一款的原因；严格区分差别检验和偏爱检验的问题，避免混淆。

结果的分析：当采取评分制或仪器测试出数据时一般有两种检验方法：

① t 检验法：适用于比较两组独立样本的均值是否存在显著差异。

② 符号检验法：适用于定类一定序变量的研究，对于不满足正态分布要求的定类一定距变量的研究也适用；即使对于满足分布要求的定类一定距变量，符号检验法同样可

用，只是效率相对较低。若研究变量仅为定序变量，则应优先选择符号检验法。

对差别检验结果的解释主要运用统计学的二项分布参数检验；规定不允许出现“无差别”的回答，强制评价员进行选择。

判断原理及判断标准：依据统计学比例的区间估计原理，参照相关资料制定《二项分布判断标准表》，通过比较实际投票结果与标准表中的临界值，判断两款护肤品之间是否存在显著差异。

例如，有 15 位评估员对两种乳液的保湿持久度进行成对比较评估，其中 10 人认为乳液 A 的保湿持久度优于乳液 B，5 人认为乳液 B 的保湿持久度更好。查二项分布判断标准表（95% 置信度），评估人数 $n=15$ 时，临界值为 9。由于 $10 > 9$ ，因此可以得出结论：在 95% 的置信度下，乳液 A 的保湿持久度显著优于乳液 B。

3.3 均匀分布的拟合优度检验

适用条件：涉及 3 种或更多的护肤品，主要是针对单项选择性问题，如了解个人对多种护肤品功效（如美白、抗皱、保湿）的偏好情况。

适用范围：① 护肤品功效评估，如对不同功效面霜的喜好程度。② 护肤品质地评估，如对不同质地乳液（清爽型、滋润型、浓稠型）的偏好。③ 护肤品香型评估，如对多种香型护肤品的接受程度。

统计分析步骤：首先计算理论选择频数 f_e ， $f_e = \text{总频数} / \text{护肤品种数}$ 。接着计算 A 值， $A = (f_0 - f_e) / f_e$ ，其中 f_0 为实际观察频数。然后计算 χ^2 值， $\chi^2 = \sum A_i^2$ ，其中 $i=1, 2, \dots, n$ 。最后求临界值 χ^2_{α} ， $\chi^2_{\alpha} = \text{CHIINV}(\alpha, n-1)$ ，其中 α 为置信度值， n 为护肤品个数。

分析判断不同护肤品数目下（ n 一般为 2、3、4、5、6、7）， χ^2 临界值表见表 1。

表 1 χ^2 临界值表

90% 置信度	95% 置信度
2.705544	3.841459
4.60517	5.991465
6.251388	7.814728
7.77944	9.487729
9.236357	11.0705
10.64464	12.59159

将计算得到的 χ^2 值与 χ^2_{α} 值进行比较：如果 $\chi^2 < \chi^2_{\alpha}$ ，则消费者对这些护肤品的选择服从均匀分布，即消费者对这些护肤品没有显著偏好。如果 $\chi^2 > \chi^2_{\alpha}$ ，则消费者对这些护肤品的选择不服从均匀分布，即消费者对这些护肤品存在显著的偏好。

例如，评估人们对 5 种不同功效面膜的偏好情况，共收集了 500 份有效问卷。统计结果显示，每种面膜被选择的人数分别为 80、120、150、90、60。首先，计算理论选择频

数 $fe=500/5=100$, 然后计算每个面膜的 A 值, 最后计算 χ^2 值。假设计算得到的 χ^2 值为 15, 查临界值表, 5 种面膜 ($n=5$) 在 95% 置信度下的临界值为 9.487729。由于 $15 > 9.487729$, 因此可以得出结论: 消费者对这 5 种面膜存在显著的偏好差异。

3.4 Duo-tri 法

试验设计要点: 设定一款护肤品为参照样品, 另外两款为测试样品; 两个测试样品中, 有一个与参照样品完全相同; 要求评估员从测试样品中选择出与参照样品相同的那个。

适用范围: 一般应用于护肤品仿制品的评估, 在需要的评价员数量上较三角测试法少; 也可用于部分护肤品老配方调整前后的使用感评估; 也可用于当竞品和自家护肤品在除评价指标外的其他方面(如包装、颜色)难以达到完全一致时的评估。

评估步骤: 工作人员精心准备参照样品和测试样品, 清晰地告知评价员测试样品中有一个与参照样品完全相同; 评价员通过感官仔细辨别, 选择出他们认为与参照样品相同的护肤品; 强制要求评价员必须从两个测试样品中选择其一, 同时, 要求评价员评价不一样的那个样品是更好还是更差, 并阐述原因。

统计分析: 分别统计两个测试样品被选择的票数。查阅 Duo-tri 设计判断标准表, 判断两个测试样品之间是否存在显著差异。汇总评价员对描述性问题的回答, 进一步明确差异的方向, 即判断是更好还是更差。

例如, 在一次护肤品仿制品评估中, 采用 Duo-tri 法, 共有 20 位评估员参与, 其中被认为与参照样品相同的测试样品 A 获得了 13 票, 测试样品 B 获得了 7 票, 查阅 Duo-tri 设计判断标准表(95% 置信度), 评估人数为 20 时, 临界值为 12。由于 $13 > 12$, 因此可以判断在 95% 的置信度下, 测试样品 A 与参照样品在感官上更相似, 并且存在显著差异。

3.5 三角测试法

适用范围: 适用于评估某特定指标(如吸收速度、油腻程度)有细微差别的两款护肤品, 也常用于护肤品香精的评估; 也可用于挑选和培训护肤品评价员, 以及考核评价员的能力。

评估方法要点: 提供三个编号的护肤品样品, 其中两个相同, 一个不同, 要求评价员准确挑选出那个与众不同的单个样品。

评价员要求: 通常需要 10 多位对护肤品质地、气味或使用效果等特定性能敏感的人员。

操作步骤: 工作人员向评价员提供一组三个编号的护肤品样品, 并明确告知其中两个是相同的, 要求评价员挑出单个不同的样品, 为了确保三个样品的排列次序出现的概率相等, 可运用以下 6 组组合: BAA、ABA、AAB、ABB、BAB、BBA。

结果统计与分析: 将能够正确选出不同样品的人次相加, 结合总人次和所需的置信水平, 查阅《三角测试判断标准表》, 对于在不同阶段分别进行三角测试的情况, 如护肤品香精测试, 若任何一个阶段未通过, 则判定该香精不通过。

例如, 在一次护肤品吸收速度的三角测试中, 共有 30 位评价员参与, 其中 18 人正确挑选出了不同的样品, 查阅《三角测试判断标准表》(95% 置信度), 评估人数为 30 时, 通过的临界值为 18, 由于实际正确人数等于临界值, 因此可以认为在 95% 的置信度下, 刚好能够区分这两款护肤品在吸收速度上的差异。

4 护肤品感官评估的数据分析

①表观数据分析, 对收集到的护肤品感官评估数据进行初步处理, 如计算样本的均值、中位数、求和等, 以了解数据的基本特征和集中趋势。

②统计推断, 目标是以样本数据推断总体的状况。在进行统计学检验时, 需要对数据的特性进行分析, 如正态性、方差等。对于不符合正态分布的数据, 可能需要采用非参数检验方法。

③最小样本量的确定, 因实验目的、评估指标和要求的精度不同, 最小样本量的确定方法各异。一般来说, 需要综合考虑置信水平、允许误差和总体变异程度等因素。

④统计分析软件的应用, 统计分析软件如 SPSS、SAS 等的应用可以大大提高工作效率, 增加结果的准确性和可靠性。这些软件能够处理复杂的数据分析任务, 提供丰富的统计图表和输出结果。

5 结语

护肤品的感官评估是一个复杂但至关重要的领域, 通过科学合理地选择评估方法、精心设计实验、严格筛选和培训评估员, 并结合精确的数据分析, 能够为护肤品的研发、改进和市场推广提供有力的支持。不同的评估方法和数据分析技术各有其特点和适用场景, 论文仅列举了部分方法, 研究人员和从业者应根据具体的评估目标和条件, 灵活运用这些方法, 以获取准确、可靠且有价值的评估结果, 满足消费者对护肤品品质和使用体验的不断追求。

参考文献:

- [1] Barkat S, Thomas-Danguin T, Bensafi M, et al. Odor and color of cosmetic products: correlations between subjective judgement and autonomous nervous system response[J]. *Cosmet Sci*,2003,25(6):273-283.
- [2] Poessel P, Ahrens S, Hautzinger M. Influence of cosmetics on emotional, autonomous, n-docrinological, and immune reactions[J]. *Cosmet Sci*,2005,27(6):343-349.
- [3] 殷润元.轻工产品的感官评价方法的建立[J].*现代装饰(理论)*, 2011(4):196.
- [4] International Organization for Standardization. Sensory analysis Vocabulary: ISO 5492 - 2008[S]. Geneva: ISO,2008.