

DIAS 用户手册

| | |
|---------------------------------------|----|
| 1 介绍 | 1 |
| 1.1 数据 | 1 |
| 1.2 说明 | 3 |
| 2 数据描述和检验 | 5 |
| 2.1 数据的个性（描述性统计） | 5 |
| 2.2 数据撞衫（共线性分析） | 5 |
| 2.3 数据的颜值（正太分布检验） | 6 |
| 2.4 数据的身材（方差齐性检验） | 7 |
| 2.5 数据的随机性（游程检验） | 8 |
| 3 问卷分析 | 11 |
| 3.1 用户需求分类（KANO 模型） | 11 |
| 3.2 品牌口碑（NPS 净推荐值） | 13 |
| 3.3 资源组合（Turf 组合模型） | 15 |
| 3.4 传话筒，放大器和大传话筒（中介，调节和调节中介） | 16 |
| 3.5 定价（价格敏感度测试模型） | 22 |
| 3.6 问卷可靠性（信度和重测信度） | 23 |
| 3.7 问卷有效性（内容效度与效度） | 26 |
| 3.8 亲疏远近（多维尺度分析） | 30 |
| 3.9 问卷多选题 | 31 |
| 3.10 评分可靠性（组内评分者信度） | 32 |
| 4 相关性分析 | 35 |
| 4.1 骡子还是马（Cochran's Q 检验） | 35 |
| 4.2 两人观点一致性（Kappa 一致性检验） | 36 |
| 4.3 多人观点一致性（Kendall 协调系数） | 36 |
| 4.4 齐步走和一起走（Pearson 和 Spearman 相关性分析） | 38 |
| 4.5 偏相关分析 | 40 |

| | |
|-------------------------|-----|
| 4.6 群体相关 (典型相关分析) | 42 |
| 5 必然还是巧合 (差异性分析) | 44 |
| 5.1 卡方检验 | 44 |
| 5.2 平均水平不是平均数 (t 检验) | 45 |
| 5.3 方差分析 | 50 |
| 5.4 中位数比较 (Wilcoxon 检验) | 54 |
| 5.5 Friedman 检验 | 57 |
| 5.6 预期和实际 (卡方拟合优度检验) | 57 |
| 5.7 干扰控制 (协方差分析) | 59 |
| 5.8 事后多重比较 | 60 |
| 6 设计方案选择与综合评价 | 62 |
| 6.1 专家背靠背 (德尔菲法) | 62 |
| 6.2 评价指标筛选 (主观数据) | 64 |
| 6.3 评价指标筛选 (客观数据) | 67 |
| 6.4 综合评价 (排名) | 72 |
| 6.5 综合评价 (诊断) | 79 |
| 6.6 关键因素 (极差分析) | 84 |
| 6.7 预测未来 | 85 |
| 6.8 DEMATEL 分析 | 88 |
| 7 创造一个新理论 | 91 |
| 7.1 线性回归 | 91 |
| 7.2 非线性回归 | 96 |
| 7.3 分类回归 | 98 |
| 8 找到一群好朋友 (聚类) | 103 |
| 9 软件报错 | 107 |

1 介绍

1.1 数据

1. 不同的数据

Excel 文件中的数据（除了第一行或特殊情况）都必须是数字，不能是文字或字母。

- 连续数据：例如，身高(175cm、175.5cm、175.55cm)，温度(25℃、25.3℃、25.31℃)
- 有序分类数据：有数值大小的区分，例如，满意度评分(1星、2星、3星、4星、5星)，成绩等级(不及格、及格、良好、优秀)，注意转换为数字：1, 2, 3, 4。
- 无序分类数据：没有数值大小的区分，例如，性别(1男, 0女)，职业(1教师, 2工程师, 3医生)，注意转换为数字：1, 2, 3。
- 线性关系：数据在散点图中近似是一条直线。
- 非线性关系：比如年龄和收入，30岁前收入随年龄增长，30岁后平稳。
- 单调关系：A变大时，B要么一直变大、要么一直变小，不需要是严格的直线。

2. 爱干净的数据

数据格式不能错，否则程序会报错，分析失败。

- 1) 文件的位置在电脑桌面上。
- 2) 文件没有密码保护，没有损坏。
- 3) 数据中没有缺失值、极端值、中文标点、特殊符号/, \, *, ?等。
- 4) 数据中没有空格，例如 有空格，没有空格的应为 。
- 5) 数据的格式不能是“文本”(Excel 格子的左上角有小三角)，如下图所示。选中所有数据，单击警示符号，点击下拉菜单中的“转换为数据”即可。

| 1 | 减肥方式 | 胆固醇水平 | 权重(样本数) |
|---|------|-------|---------|
| 2 | 1 | 1 | 2 |
| 3 | 1 | 2 | 10 |
| 4 | 2 | 1 | 7 |
| 5 | 2 | 2 | 4 |
| 6 | 2 | 1 | 6 |

- 6) 严格遵守文中对数据的要求。
- 7) 如果找不到数据的格式问题，而程序依旧报错，可以使用“数据清洗”功能对数据进行清洗。

3. 负数 → 正数

DIAS 默认只处理正数，因此，若存在负数，则需要“平移数据”(正数也需要一起平移)。平移后的数据将全部变为正数，且保持了原始数据的逻辑关系，可用于后续分析。

| A |
|--------|
| 1 原数据 |
| 2 -2.5 |
| 3 -1.8 |
| 4 0.5 |
| 5 1.2 |
| 6 -0.3 |
| 7 2 |

找数据中的最小值，公式为“=MIN(A:A)”

计算平移量（取最小值绝对值 + 0.0001）

| A | B | C | A | B | C | A | B | C |
|--------|---|-----------|--------|---|------|--------|---|--------|
| 1 原数据 | | =MIN(A:A) | 1 原数据 | | -2.5 | 1 原数据 | | -2.5 |
| 2 -2.5 | | | 2 -2.5 | | | 2 -2.5 | | 2.5001 |
| 3 -1.8 | | | 3 -1.8 | | | 3 -1.8 | | |
| 4 0.5 | | | 4 0.5 | | | 4 0.5 | | |
| 5 1.2 | | | 5 1.2 | | | 5 1.2 | | |
| 6 -0.3 | | | 6 -0.3 | | | 6 -0.3 | | |
| 7 2 | | | 7 2 | | | 7 2 | | |
| 8 | | | 8 | | | 8 | | |

平移单个数据，公式为“=A2+\$C\$2”，批量填充即可

| A | B | C | A | B | C | A | B | C |
|--------|------------|--------|--------|--------|--------|--------|--------|--------|
| 1 原数据 | | -2.5 | 1 原数据 | | -2.5 | 1 原数据 | | -2.5 |
| 2 -2.5 | =A2+\$C\$2 | 2.5001 | 2 -2.5 | 0.0001 | 2.5001 | 2 -2.5 | 0.0001 | 2.5001 |
| 3 -1.8 | | | 3 -1.8 | | | 3 -1.8 | | |
| 4 0.5 | | | 4 0.5 | | | 4 0.5 | | |
| 5 1.2 | | | 5 1.2 | | | 5 1.2 | | |
| 6 -0.3 | | | 6 -0.3 | | | 6 -0.3 | | |
| 7 2 | | | 7 2 | | | 7 2 | | |
| 8 | | | 8 | | | 8 | | |

4. 负向数据 → 正向数据

正向数据：数值越大，代表指标越好。

- 产品设计中的「用户满意度（1-10 分）」：得分 8 分比 5 分好。
- 交互设计中的「操作效率（次 / 分钟）」：每分钟完成 10 次操作比 5 次好。

负向数据：数值越小，代表指标越好。（注意“负向数据”不是“负数”）

- 产品使用中的「操作错误率（%）」：3%的错误率比 8%好。
- 交互设计中的「响应延迟（秒）」：0.5 秒比 2 秒好。

DIAS 默认只处理正向数据，因此，若数据中存在负向数据，我们需要把负向数据转换为正向数据。这里推荐并演示**“极值反转法”（最常用）**。如需要另外两种方法“标准化反转法”和“倒数法”，请自行学习。

初始负向数据

| A |
|-----------|
| 1 智能手表误触率 |
| 2 3.2 |
| 3 5.8 |
| 4 2.5 |
| 5 8.1 |
| 6 4.3 |
| 7 6.5 |
| 8 9.2 |
| 9 3 |
| 10 5.1 |

找数据中的最大值，公式为“=MAX(A:A)”，找数据中的最小值，公式为

“=MIN(A:A)”。

这里的“A”指的是表格中的 A 列。注意，必须使用英文输入法。

计算单元格 A2 的正向值：“=\$C\$1+\$C\$2-A2”

填充所有单元格，双击这个“点”，或向下拖动即可。

5. 注意！

在 DIAS 中输入信息时，需使用英文输入法，不能输入空格，特殊符号等。注意“逗号”也要用英文输入法。

1.2 说明

所有统计算法均源自 Python 科学计算库（如 Scikit-learn、NumPy、scipy），基于经典统计理论，广泛应用于学术研究与企业研发。

本手册中所有的数据均为“虚拟数据”，仅为“演示 DIAS 操作”和“辅助理解 DIAS”所用，所有的案例猜想，分析结果和结论均不可信。

分析结果的内容较多，手册中仅展示了一部分，案例的“结论”是借助 AI 工具对“分析结果”的解读。以“价格敏感度测试”为例，向 AI 提问：

中午好，chi

W 分析结果.docx

Word · 57KB · 约 0.1 万字

详细总结这篇文档内容 →

用通俗易懂的话，说说文档讲了什么 →

对文档内容进行润色 →

这个分析结果我看不懂



深度思考

取消



AI 编程

图像生成

AI 搜索

AI 阅读

数据分析

帮我写作

更多

2 数据描述和检验

2.1 数据的个性（描述性统计）

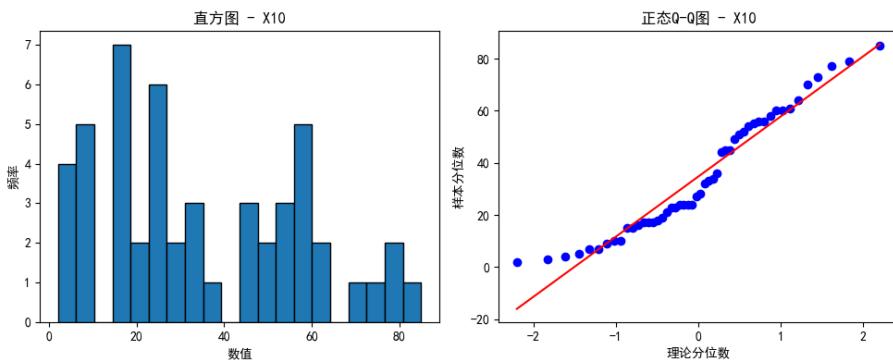
计算每一列数据的平均值，中位数，标准差，最小值，最大值，极差，第一四分位数，第三四分位数，四分位距，峰度，偏度，众数。

示例数据

| | A | B | C | D | E | F | G | H | I | J | K |
|---|----|----|----|----|----|----|----|----|----|-----|---|
| 1 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | Y |
| 2 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 2 |
| 3 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 3 |
| 4 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 56 | 1 |
| 5 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 4 | 1 |
| 6 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 5 | 0 |

分析结果

| 列名 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | Y | 解释说明 |
|-----|------|------|------|------|------|------|------|------|-------|-------|------|--|
| 平均值 | 0.46 | 0.60 | 0.30 | 0.30 | 0.18 | 0.40 | 0.46 | 0.40 | 0.68 | 34.76 | 0.48 | 平均值是数据集中所有值的平均值，反映数据的集中趋势。 |
| 中位数 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 27.5 | 0 | | 中位数是数据集按升序排列时位于中间的值，它不受极端值的影响，能更好地反映数据的中间水平。 |
| 标准差 | 0.5 | 0.50 | 0.50 | 0.50 | 0.39 | 0.50 | 0.50 | 0.47 | 23.07 | 0.5 | | 标准差衡量数据相对于平均值的离散程度，标准差越大，数据越离散。 |
| 最小值 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | | 最小值是数据集中的最小值。 |
| 最大值 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 85 | 1 | | 最大值是数据集中的最大值。 |
| 极差 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 83 | 1 | | 极差是最大值与最小值之差，反映数据的范围。 |
| | | | | | | | | | | | | ... |
| 众数 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 24 | 0 | 众数是数据集中出现频率最高的值。 |



2.2 数据撞衫（共线性分析）

“共线性”指两个或多个自变量（不涉及因变量）非常像。例如，分析“面粉用量”，“水的用量”和“面团总重量”对“面包软硬度”的影响时，“面团总重量”和“面粉用量 + 水的用量”非常像，这会严重影响结论的可靠性。

很多分析方法都要求数据没有共线性，共线性分析为它们提供“前提诊断”。

若自变量存在严重的共线性，则需要对自变量的数据进行处理，通常是删除、合并或替换变量。例如：“身高”和“体重”的 VIF=15（严重共线性），可删除其中一个变量，或用“体重/身高²”替代它们。也可以通过增加样本量或修改模型的方法解决问题。

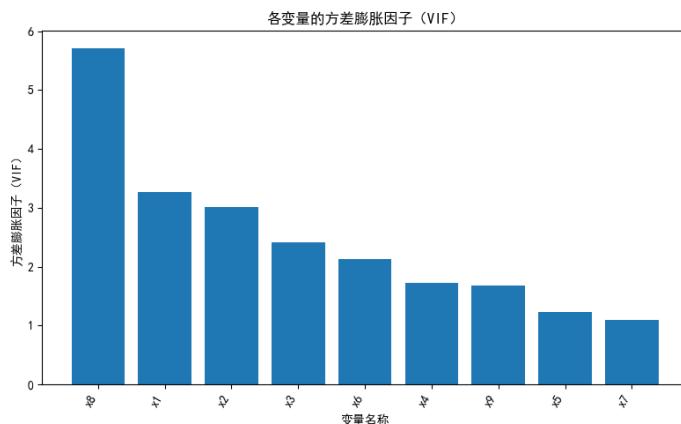
示例数据

程序默认 Excel 中的最后一列为因变量，其余列作为自变量，程序仅分析自变量。

| | A | B | C | D | E | F | G | H | I | J |
|---|----|----|----|----|----|----|----|----|----|-----|
| 1 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 |
| 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 2 |
| 3 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 3 |
| 4 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 56 |
| 5 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 4 |
| 6 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 5 |

分析结果

| 变量名 | 方差膨胀因子 | 结果解释 |
|-----|--------|--|
| X1 | 3.268 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X2 | 3.016 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X3 | 2.408 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X4 | 1.725 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X5 | 1.236 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X6 | 2.135 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X7 | 1.089 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |
| X8 | 5.719 | 方差膨胀因子 (VIF) 在 5 到 10 之间，表明该变量与其他变量之间可能存在一定的共线性。 |
| X9 | 1.682 | 方差膨胀因子 (VIF) 小于 5，表明该变量与其他变量之间不存在严重的共线性。 |



2.3 数据的颜值（正太分布检验）

“正态分布”是自然界和社会科学中最常见的分布（如身高、体重、测量误差等）。正态分布检验是用于判断数据是否近似服从正态分布的统计方法，许多统计方法的有效性依赖“数据服从正态分布”的假设。

DIAS 提供了多种正态性检验的工具。

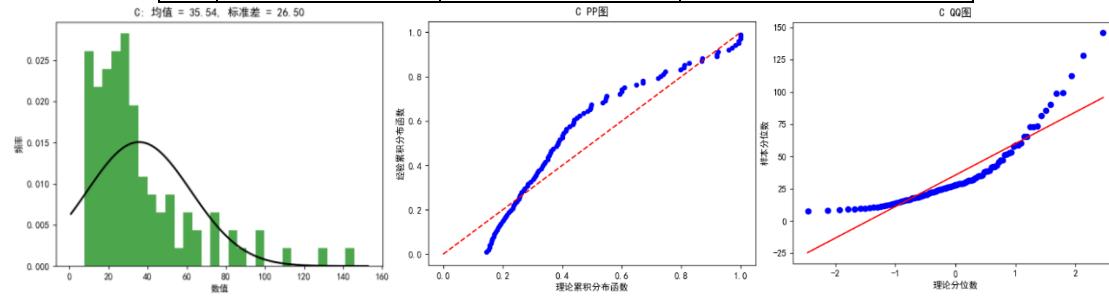
| 检验工具 | 样本量 | 优势场景 |
|--------------------------------|------------------------|-----------------|
| Shapiro-Wilk 检验（最常用） | 小样本 ($n < 50$) | 最常用的“万金油”，适合小样本 |
| Kolmogorov-Smirnov 检验 (K-S 检验) | 大样本 ($n \geq 50$) | 快速判断 |
| Lilliefors 检验 | 大样本 ($n \geq 100$) | 更贴合实际 |
| Anderson-Darling 检验 | 任意样本量 | 怀疑数据有极端值 |
| D'Agostino's K-squared 检验 | 大样本 ($n \geq 50$) | 怀疑数据存在“不对称” |
| Jarque-Bera 检验 | 极大样本 ($n \geq 2000$) | 怀疑数据存在“不对称” |

示例数据

| | A | B | C |
|---|-------|-------|------|
| 1 | A | B | C |
| 2 | 52.36 | 172.1 | 18.2 |
| 3 | 48.71 | 168.5 | 25.7 |
| 4 | 55.92 | 175.3 | 12.3 |
| 5 | 41.83 | 163.8 | 31.5 |
| 6 | 50.15 | 170.2 | 22.8 |

分析结果

| 列名 | Shapiro - Wilk 统计量 | P 值 | 结果解释 |
|----|--------------------|-----------------------|---------------|
| A | 0.9837294838899768 | 0.2562965514783715 | 样本可能服从正态分布。 |
| B | 0.9825051093743474 | 0.20731938448986387 | 样本可能服从正态分布。 |
| C | 0.8164466090525863 | 8.184974704046511e-10 | 样本不太可能服从正态分布。 |



2.4 数据的身材（方差齐性检验）

“方差齐”指多列数据有相同的方差（数据离散程度一致）。例如，比较“男生身高”和“女生身高”的均值时，需先检验两者的身高方差是否相等；分析 3 种教学方法的效果时，需检验 3 组学生成绩的方差是否一致。

方差齐性检验为其他分析提供“前提诊断”，很多的分析方法都要求数据的方差齐。

DIAS 内置了两种方差齐性检验工具，“Bartlett 检验”和“Levene 检验”。

Bartlett 检验通过计算各样本方差的“一致性程度”（即“方差齐”假设是否成立），判断总体方差是否存在显著差异。

Bartlett test 和 Levene test 的功能是一样的，区别如下。

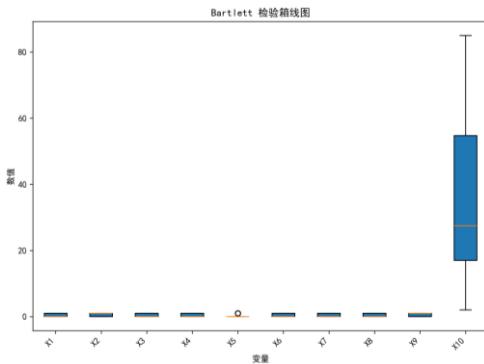
- 数据通过正态分布检验，用 Bartlett 检验。
- 数据未通过正态分布检验，数据量少或存在极端值，用 Levene 检验。

示例数据

| | A | B | C | D | E | F | G | H | I | J |
|---|----|----|----|----|----|----|----|----|----|-----|
| 1 | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 |
| 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 2 |
| 3 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 3 |
| 4 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 56 |
| 5 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 4 |
| 6 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 5 |

分析结果

| 列名 | Bartlett 统计量 | Bartlett P 值 | 结果解释 |
|-----|-------------------|--------------|-------------|
| 全面的 | 2296.282104870466 | 0.0 | 各样本方差不具有齐性。 |



2.5 数据的随机性（游程检验）

“游程”指的是数据中连续出现的、相同的元素序列。游程检验的用途是检验一组数据的排列是否随机，或比较两组样本的分布是否一致，需合并成一组数据。

- 用途一：判断一组数据的顺序是否是随机的，而非存在某种固定模式（如周期性、趋势性、聚类性）。

方法：先将数据按某种规则（如“正面 / 反面”，“合格 / 不合格”）分为两类（记为 0 和 1），再统计连续的 0 或连续的 1 构成的游程数量。
例如，抛硬币结果：正、正、反、反、反、正 → 分类为 0、0、1、1、1、0 → 游程数 = 3 (0 的游程、1 的游程、0 的游程)；
随机排列的数据中，游程数量会处于一个“合理区间”。若游程过少，说明数据存在“扎堆儿”（如连续多个 0 后连续多个 1）；若游程过多，说明数据存在“交替过频”（如 0、1、0、1 频繁切换，可能有人为干预，非随机）；若游程数量在“合理区间”内，则证明数据是随机的。
应用场景：检验生产线上的产品质量是否随机（如“合格 / 不合格”的排列是否有规律，避免因设备故障导致连续不合格）。
- 用途二：比较两组数据的分布形状、位置是否一致。
方法：合并两组数据后按大小排序，再统计合并后的序列的游程数量。

例如，两组数据合并排序后是：0、0、1、1、0、1 → 游程数=4（0的游程、1的游程、0的游程、1的游程）。

若两组数据的分布形状、位置一致，混合后0和1会“随机分布”，游程数会处于合理范围；若不一致，会出现“0集中在两端、1集中在中间”或反之，导致游程数量过少。

典型应用场景：检验两种教学方法的效果是否一致（两组学生成绩的分布是否相同）；检验两种工艺生产的产品性能是否一致（两组产品指标的分布是否相同）。

示例数据

这个数据不能直接分析，需要处理。

检验两种教学方法的效果是否一致（两组学生成绩的分布是否相同），样本量最少5。

| A | B |
|----|----|
| 85 | 98 |
| 93 | 76 |
| 87 | 86 |
| 99 | 81 |
| 91 | 97 |
| 77 | 91 |

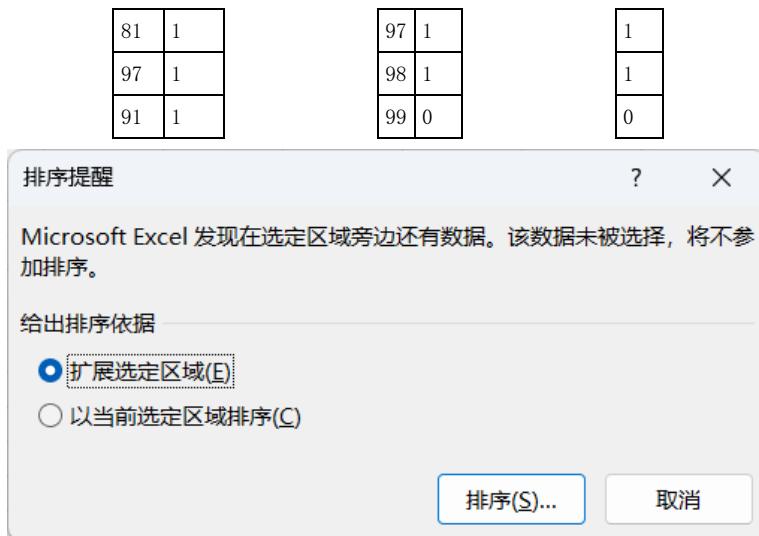
数据处理过程（Excel）

增加标签列

| A | Tag 0 | B | Tag 1 |
|----|-------|----|-------|
| 85 | 0 | 98 | 1 |
| 93 | 0 | 76 | 1 |
| 87 | 0 | 86 | 1 |
| 99 | 0 | 81 | 1 |
| 91 | 0 | 97 | 1 |
| 77 | 0 | 91 | 1 |

合并数据。按数值，升序排列C组数据（扩展选定区域）。仅保留标签列，用于游程数量的统计。

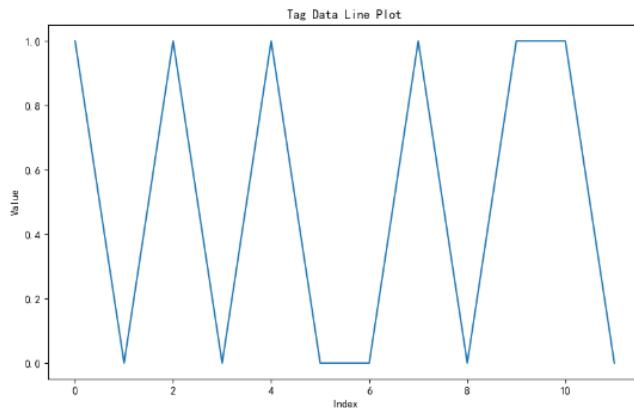
| C | Tag |
|----|-----|
| 85 | 0 |
| 93 | 0 |
| 87 | 0 |
| 99 | 0 |
| 91 | 0 |
| 77 | 0 |
| 98 | 1 |
| 76 | 1 |
| 86 | 1 |
| 91 | 1 |
| 93 | 0 |



分析结果

| 列名 | 运行次数 | Z 统计量 | P 值 |
|-----|------|--------------------|---------------------|
| Tag | 10 | 1.5138251770487459 | 0.13007018523428468 |

- 游程数：数据中相同符号的连续片段数，游程过多或过少都可能表明数据并非随机。
- Z 统计量：用于衡量游程数与预期游程数之间的差异，Z 统计量的绝对值越大，游程数与预期游程数之间的差异越显著。
- P 值：当 p 值小于显著性水平（通常为 0.05）时，拒绝零假设，表明数据并非随机；否则，接受零假设，表明数据是随机的。用于判断数据是否随机。



3 问卷分析

3.1 用户需求分类 (KANO 模型)

“用户需求”分类与优先级排序的工具，可以区分“不同的需求”对“用户满意度”的影响程度，从而分配资源和制定产品策略，避免“盲目满足需求”。

KANO 模型将用户需求分为 6 类

| 需求类型 | 核心定义 |
|-----------|--|
| 基本型需求 (M) | 有则无所谓，无则用户非常不满 (产品的“底线功能”，如手机的“通话功能”)。 |
| 期望型需求 (O) | 有则满意，无则用户不满 (用户明确期待的功能，如手机的“电池续航时长”)。 |
| 魅力型需求 (A) | 有则非常满意，无则用户无所谓 (用户未预期的惊喜功能，如早期手机的“面部识别解锁”)。 |
| 无差异需求 (I) | 无论有无，用户都无所谓 (对用户价值极低的功能，如“手机包装盒上的多余 logo”)。 |
| 反向型需求 (R) | 有则不满，无则用户满意 (用户反感的功能，如“强制弹出的广告”)。 |
| 可疑型需求 (Q) | 无效数据 (如“有此功能用户非常满意，无此功能用户也非常不满”)。 |

编写调查问卷，每一个问题需要从正反两方面共设计两道题目。

| 问题类型 | 问题描述 | 选项 (用户感受) (5 级李克特量表) |
|---------------|---|--|
| 正向问题 (需求满足时) | 假如 [产品 / 服务] 具备 “[具体需求]”，您的感受是？ 例如，“如果产品提供‘一键备份数据’功能，您的感受是？” | 1. 很满意 2. 满意 3. 无所谓 4. 不满意 5. 很不满意 |
| 反向问题 (需求不满足时) | 假如 [产品 / 服务] 不具备 “[具体需求]”，您的感受是？ 例如，“如果产品不提供‘一键备份数据’功能，您的感受是？” | 1. 很满意 2. 满意 3. 无所谓 4. 不满意 5. 很不满意 |

另外，可收集基础信息（如年龄、性别、使用频率、用户类型等），用于后续分析（如“新用户 vs 老用户对需求的差异反馈”）。

建议每个核心用户群体的样本量 ≥ 50 份（样本量过小会导致结果随机性强）；若用户群体细分（如“新用户”“付费用户”），每个细分群体需单独满足样本量的要求。

常见问题

- 问题：问卷回收后，“Q”占比过高怎么办？
原因：需求描述模糊，或用户未认真阅读问题；
- 问题：不同用户群体对同一需求的回答差异极大，如何决策？
解决：优先满足核心用户群体的需求类型。例如，若产品核心用户是“学生”，则按学生的反馈定义需求类型，而非“职场人士”。
- 问题：需求类型随时间变化，如何保持分析有效性？
解决：定期（如每季度 / 每半年）重新开展 KANO 调研，更新需求类型。例如，“手机人脸识别”曾是“A（魅力型）”，现在已变为“O（期望型）”。

示例数据

分析过程与需注意的问题

列名是第一行的内容，而非 A, B, C

| | A | B | C | D | E | F | G | H | I |
|---|----|----|----|-----|----|---------|---------|----------|----------|
| 1 | 序号 | 性别 | 年龄 | 年收入 | 学历 | 投影功能(正) | 投影功能(负) | 左右手模式(正) | 左右手模式(负) |
| 2 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 2 | 1 |
| 3 | 2 | 1 | 4 | 1 | 2 | 4 | 1 | 5 | 2 |
| 4 | 3 | 2 | 3 | 4 | 2 | 5 | 4 | 5 | 3 |
| 5 | 4 | 1 | 4 | 1 | 3 | 2 | 2 | 5 | 2 |

请输入正向问题的列名, 用逗号分隔
投影功能(正),左右手模式(正),超级快充

OK Cancel

请输入负向问题的列名, 用逗号分隔
投影功能(负),左右手模式(负),超级快充

OK Cancel

输入的正向问题的列名：投影功能(正),左右手模式(正),超级快充(正),取消 SIM 卡(正),3D 投影(正),照片搜索(正),自动美颜(正),防盗加锁(正),遥控器(正),暖手宝(正),望远镜(正),显微镜(正)

输入的负向问题的列名：投影功能(负),左右手模式(负),超级快充(负),取消 SIM 卡(负),3D 投影(负),照片搜索(负),自动美颜(负),防盗加锁(负),遥控器(负),暖手宝(负),望远镜(负),显微镜(负)

检查输入的正向 / 负向列的数量是否相等，逗号必须是英文输入。

分析结果

KANO 模型分析结果

| 问题 | KANO 分类 | 较好系数 | 较差系数 |
|-------------|-----------|------|-------|
| 投影功能(正) | 兴奋型需求 (A) | 0.58 | -0.09 |
| 左右手模式(正) | 兴奋型需求 (A) | 0.75 | -0.03 |
| 超级快充(正) | 兴奋型需求 (A) | 0.7 | -0.01 |
| 取消 SIM 卡(正) | 期望型需求 (O) | 0.52 | -0.13 |
| 3D 投影(正) | 可疑结果 (Q) | 0.41 | -0.1 |
| 照片搜索(正) | 期望型需求 (O) | 0.6 | -0.13 |
| 自动美颜(正) | 期望型需求 (O) | 0.52 | -0.19 |
| 防盗加锁(正) | 兴奋型需求 (A) | 0.46 | -0.17 |
| 遥控器(正) | 可疑结果 (Q) | 0.35 | -0.21 |
| 暖手宝(正) | 期望型需求 (O) | 0.4 | -0.21 |
| 望远镜(正) | 可疑结果 (Q) | 0.41 | -0.14 |
| 显微镜(正) | 可疑结果 (Q) | 0.46 | -0.09 |

基本型需求 (M): 用户期望产品具备的基本需求。缺少这些功能会导致用户不满意。

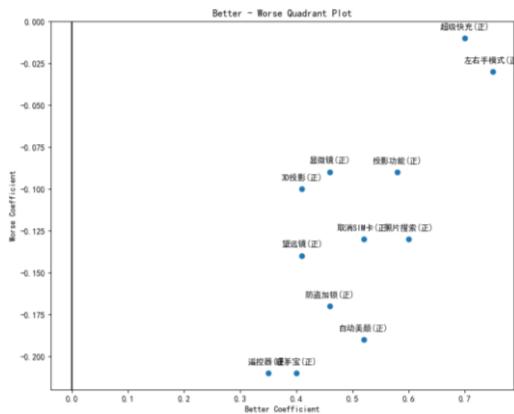
期望型需求 (O): 用户满意度随着满足程度线性增长的期望需求。

兴奋型需求(A): 用户未曾期望的激动人心的需求。满足这些需求可以提升用户满意度。

无差异型需求 (I): 用户不太在意是否满足的无关紧要的需求。

反向型需求 (R): 满足这些需求会导致用户不满意的反向需求。

可疑结果 (Q): 答案相互矛盾，结果不可靠。



3.2 品牌口碑 (NPS 净推荐值)

衡量客户忠诚度和口碑。研究表明，NPS 与企业增长高度相关：推荐者越多，口碑传播带来的新客户越多；贬损者越少，客户流失和负面影响越小。通过跟踪 NPS 变化，可预判市场份额和复购率等业务指标的趋势。

仅通过一个问题“您有多大可能向朋友或同事推荐我们的产品 / 服务？”（评分范围 0-10 分），将客户分为三类：

- 推荐者：9-10 分 → 占比 = (推荐者人数 ÷ 总样本数) × 100%
- 中立者：7-8 分
- 贬损者：0-6 分 → 占比 = (贬损者人数 ÷ 总样本数) × 100%

NPS 计算公式：NPS = 推荐者占比 - 贬损者占比（范围：-100~100 分，分数越高说明客户忠诚度越高）

确保调研样本量足够（至少 100+），且覆盖不同类型客户。

示例数据

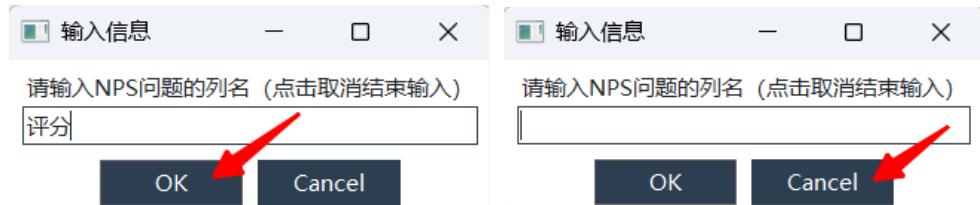
| | A | B |
|---|----|----|
| 1 | 用户 | 评分 |
| 2 | 1 | 9 |
| 3 | 2 | 5 |
| 4 | 3 | 4 |
| 5 | 4 | 6 |
| 6 | 5 | 6 |

分析过程

程序持续提示输入列名，是因为代码中设计了循环输入逻辑，目的是同时分析多个列。每次输入并点击“OK”后，会继续弹出对话框，当我们输入完所有需要分析的列名后，在弹出的输入框中点击“取消”按钮，即可结束输入并开始分析。

若只需分析 1 列，输入列名后点击“OK”，然后点击“Cancel”；若需分析 3 列，输入 3 次列名后点击“OK”，然后点击“Cancel”。

例如，数据 5 仅有一列“评分”数据，因此操作步骤如下



分析结果

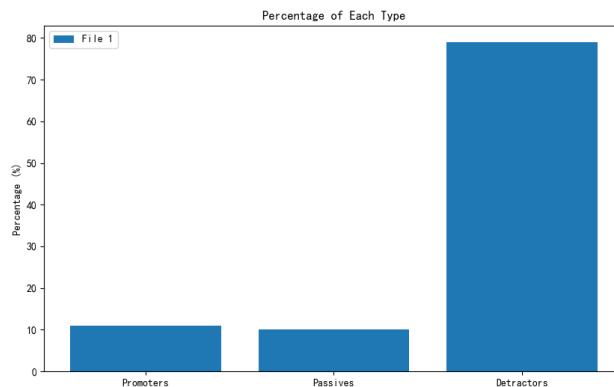
| 指标 | 数值 |
|-------------|-------|
| 评分_推荐者数量 | 11.0 |
| 评分_推荐者比例 | 11.0 |
| 评分_被动者数量 | 10.0 |
| 评分_被动者比例 | 10.0 |
| 评分_贬损者数量 | 79.0 |
| 评分_贬损者比例 | 79.0 |
| 评分_NPS 净推荐值 | -68.0 |

解释说明

- 推荐者：给出 9 – 10 分的客户，是产品或服务的忠实拥护者，会积极推荐给他人。
- 被动者：给出 7 – 8 分的客户，对产品或服务基本满意，但不会主动推荐。
- 贬损者：给出 0 – 6 分的客户，对产品或服务不满意，可能会向他人抱怨。
- NPS 净推荐值：NPS = 推荐者比例 - 贬损者比例，反映了客户对产品或服务的整体态度。

结果解读

- 推荐者：应关注推荐者的需求，提供更好的服务，鼓励他们继续推荐。
- 被动者：可以通过改进产品或服务，将被动者转化为推荐者。
- 贬损者：及时了解贬损者的不满原因，采取措施改进，避免负面影响扩大。
- NPS 净推荐值：NPS 值越高，说明客户对产品或服务越满意，忠诚度越高。



结论

NPS 净推荐值为 -68，说明大部分客户（79%）对产品 / 服务不满意，仅有少数（11%）

会主动推荐，客户忠诚度严重不足，需优先解决贬损者的不满问题。

3.3 资源组合（Turf 组合模型）

评估“资源组合”和“组合效率”，可以优化资源分配、筛选最佳组合方案，适用于从多个组合中选择“可以覆盖到最多人群”的场景。

比如咖啡机厂商有 10 种咖啡机颜色备选方案，厂商觉得颜色太多导致成本过高，消费者也有选择困难。因此，需要从 10 种颜色中找出其中 3 种，以覆盖最多的群体。

注意！咖啡机厂商并不是提供用户最喜欢的前三种，而是哪三种颜色组合在一起，可以覆盖到最多的人群，才是 Turf 组合模型解决的问题。

问卷设计

一共 10 道题，每一道题有两个选项，如下。

你对**颜色喜欢吗？ 喜欢 不喜欢

也可以只设计一道题，和 10 个选项，如下。

你喜欢下面的哪些颜色？

颜色 1 颜色 2 颜色 3 … 颜色 10

将“喜欢”的选项计为 1，“不喜欢”的选项计为 0，得到 10 列数据。

| 用户 | 颜色 1 | 颜色 2 | 颜色 3 | 颜色 4 | 颜色 5 | 颜色 6 | 颜色 7 | 颜色 8 | 颜色 9 | 颜色 10 |
|------|------|------|------|------|------|------|------|------|------|-------|
| 用户 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 用户 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 用户 3 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 用户 4 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 用户 5 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

示例数据（上图）

分析结果

| 统计量 | 统计量值 |
|-----------|---|
| 最优组合（前 3） | ['颜色 4, 颜色 5, 颜色 10', '颜色 4, 颜色 5, 颜色 6', '颜色 2, 颜色 4, 颜色 5'] |
| 各列样本量 | {'颜色 1': 99, '颜色 2': 99, '颜色 3': 99, '颜色 4': 99, '颜色 5': 99, '颜色 6': 99, '颜色 7': 99, '颜色 8': 99, '颜色 9': 99, '颜色 10': 99} |
| 各列均值 | {'颜色 1': (0.46), '颜色 2': (0.47), '颜色 3': (0.46), '颜色 4': (0.61), '颜色 5': (0.57), '颜色 6': (0.48), '颜色 7': (0.47), '颜色 8': (0.46), '颜色 9': (0.38), '颜色 10': (0.49)} |

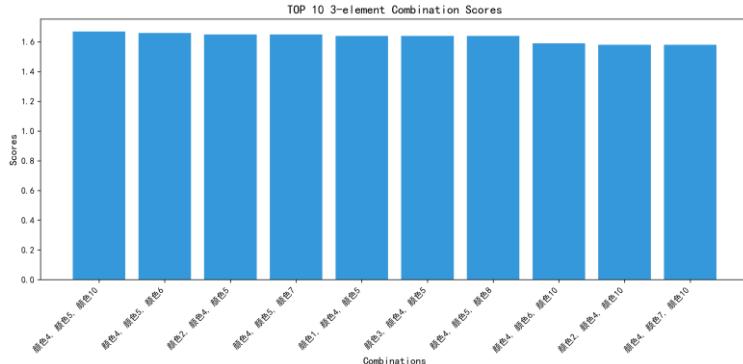
- 解释说明

Turf 组合得分：Turf 组合模型中每个组合的得分，反映了该组合的吸引力。

最优组合：得分最高的组合，代表了最有吸引力的产品或服务组合。

样本量：每个样本中的观测值数量。

- 均值：样本数据的平均值。
- 结果解读
 - Turf 组合得分：Turf 组合得分越高，说明该组合越受用户欢迎。
 - 最优组合：最优组合是最能满足用户需求的组合，可作为产品或服务的推荐组合。
 - 样本量：样本量的大小会影响统计检验的稳定性，较大的样本量通常能提供更可靠的结果。
 - 均值：均值反映了数据的平均水平，可用于比较不同变量的集中趋势。



结论

在所有的3色组合中，包含“颜色4 + 颜色5”的组合吸引力最强，其中“颜色4 + 颜色5 + 颜色10”是最优组合，其次是“颜色4 + 颜色5 + 颜色6”和“颜色2 + 颜色4 + 颜色5”；单个颜色中，颜色4和颜色5最受欢迎。

3.4 传话筒，放大器和大传话筒（中介，调节和调节中介）

这三种方法非常相似，但用途不同。

| 变量角色 | 核心要求 |
|----------|--|
| 因变量 (Y) | 必须是连续数据（如成绩、收入、满意度评分，不能是分类变量如“性别、是否购买”） |
| 自变量 (X) | 连续 / 分类均可（分类变量需做虚拟编码，如“教学方法 A=0, B=1”） |
| 中介变量 (M) | 必须是连续数据（需传递 X 对 Y 的影响，分类变量无法传递“中间效应”） |
| 调节变量 (W) | 连续 / 分类均可（分类变量更常见，如“性别、学历”；连续变量如“年龄、收入”） |

中介作用

一个东西 (X) 到底是“绕了什么弯”影响另一个东西 (Y) 的，也就是把“X 影响 Y”的过程拆开，看看中间是不是藏着“传话筒”这个必经之路(传话筒就是 M, 中介变量)。

例如：某团队设计了一款健康饮食类 APP，发现“界面视觉吸引力 (X) → 用户持续使用意愿 (Y)”正相关。中介作用发现，“界面视觉吸引力 (X) → 用户信任感 (M1) → 操作流畅度 (M2) → 用户持续使用意愿 (Y) ”。

调节作用

研究 X 对 Y 的影响时，是否会受到调节变量 Z 的干扰，也就是“X 通过 M 影响 Y 的过程，在什么条件下更有效，Z 的干扰有可能改变 X 对 Y 的影响。

例如，运动 (X) 会对减肥 (Y) 产生影响，这种影响关系是否受到饮食控制 (Z) 的干扰，也就是说，控制饮食与不控制饮食对减肥的影响是否不一样。

调节中介作用

它把影响关系拆得更细，分析了“谁在什么情况下，通过什么方式影响了谁”。

例如：为什么同样是“多刷题 (X)”，有的学生“期末分数 (Y)”涨很多，有的涨很少？调节中介能找到背后的“关键干扰因素”（调节变量，比如“学习天赋”或“老师指导频率”）。

- “天赋高的学生”：刷题对“知识点掌握 (M1)”的提升特别大，而且“知识点掌握”对分数的拉动也更强，相当于“刷题→M1→Y”这条路径被“天赋”放大了，所以分数涨得多。
- “天赋低的学生”：刷题对“知识点掌握 (M1)”的提升弱，“知识点掌握”对分数的拉动也弱，路径被“天赋”削弱了，所以分数涨得少。
- 还能发现：“老师指导频率”只影响“刷题→解题速度 (M2)”的路径，对“刷题→知识点掌握 (M1)”没影响。

三种方法的异同点

| 维度 | 中介作用 | 调节作用 | 调节中介作用 |
|--------|------------------------|-------------------------|--|
| 核心关系 | X→M→Y (X 通过 M 影响 Y) | X 对 Y 的影响，由 W 决定强度 / 方向 | X→M→Y 的链条中，某一步 (X→M 或 M→Y) 受 W 调节 |
| 第三变量角色 | M 是“中间桥梁”（必须有它，X 才到 Y） | W 是“放大器”（不搭桥，只控制效果） | W 是“中介链条上的开关”（先搭桥，再控制效果） |
| 大白话举例 | 努力→多做题→考高分（多做题是中介） | 努力对分数的影响，看智商（智商是调节） | 努力→多做题→考高分，但“努力→多做题”的效果，看家长监督（监督是调节中介里的 W） |

中介作用 示例数据

| | A | B | C | D | E | F |
|---|------|------|------|------|------|------|
| 1 | X1 | X2 | X3 | M1 | M2 | Y |
| 2 | 4.00 | 4.00 | 4.00 | 3.93 | 4.00 | 3.27 |
| 3 | 4.25 | 3.25 | 4.00 | 3.80 | 3.00 | 3.91 |
| 4 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 |
| 5 | 4.25 | 4.00 | 4.00 | 4.00 | 5.00 | 4.00 |
| 6 | 4.50 | 4.75 | 5.00 | 4.20 | 3.22 | 2.82 |

分析过程

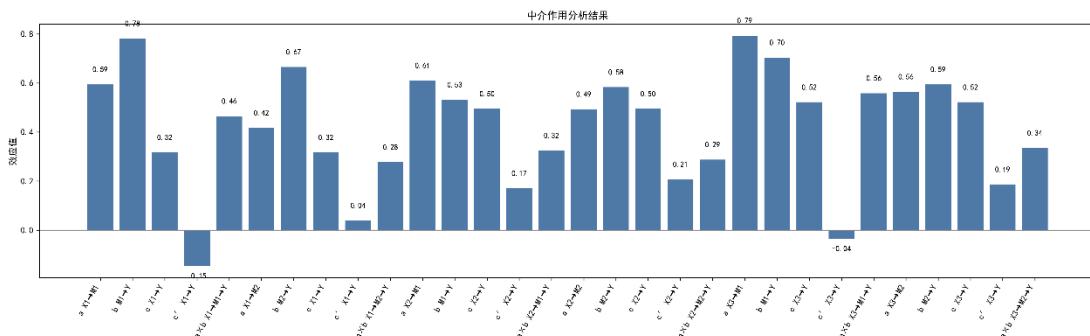
程序支持“单个自变量→单个中介变量→单个因变量”的简单中介模型，也支持“多个自变量→多个中介变量→多个因变量”的复杂中介模型。输入多个变量时，用英文逗号分隔，不能有空格。 X1,X2,X3 M1,M2 Y

| | | |
|---------------------------------|-------------------------------|----------------|
| 请输入自变量的列名 (多个用逗号分隔) X1,X2,X3 | 请输入中介变量的列名 (多个用逗号分隔) M1,M2 | 请输入因变量的列名 Y |
| OK | OK | OK |
| Cancel | Cancel | Cancel |

分析结果

| 统计类型 | 相关变量 | 效应值 | p 值 |
|---------------------|------|----------------------|------------------------|
| 自变量对因变量的总效应 | X1 | -0.15620390877237986 | 0.08849998789375227 |
| 自变量对因变量的总效应 | X2 | 0.35250160492429683 | 0.00023759826499568024 |
| 自变量对因变量的总效应 | X3 | 0.3575095252839462 | 0.0008903160593527126 |
| 自变量对中介变量[M1]的效应 | X1 | 0.10016557134340576 | 0.03291195450339405 |
| 自变量对中介变量[M1]的效应 | X2 | 0.12360705056019206 | 0.010923013196915105 |
| 自变量对中介变量[M1]的效应 | X3 | 0.6232352827560815 | 9.13307709401186e-24 |
| 自变量对中介变量[M2]的效应 | X1 | 0.017653382398746476 | 0.8488652804853278 |
| 自变量对中介变量[M2]的效应 | X2 | 0.23083891568348414 | 0.016499581536466496 |
| 自变量对中介变量[M2]的效应 | X3 | 0.369977752226337 | 0.0006991805173444058 |
| 中介变量对因变量的效应 (控制自变量) | M1 | 0.36322729648095703 | 0.002324142739579818 |
| 中介变量对因变量的效应 (控制自变量) | M2 | 0.5147087316313013 | 1.6311743287939021e-15 |

| | |
|--------------------|----------------------|
| 中介效应 [X1 → M1 → Y] | 0.0363828696795357 |
| 中介效应 [X1 → M2 → Y] | 0.009086350063461137 |
| 中介效应 [X2 → M1 → Y] | 0.04489745480096353 |
| 中介效应 [X2 → M2 → Y] | 0.11881480550259102 |
| 中介效应 [X3 → M1 → Y] | 0.22637606682703632 |
| 中介效应 [X3 → M2 → Y] | 0.19043077958021776 |
| 样本量 | 200.0 |



结论

X2 和 X3 对因变量的总效应显著，且主要通过 M1 和 M2 产生中介作用（其中 X3 通过 M1 的中介效应最强，X2 通过 M2 的中介效应强于 M1），而 X1 对因变量的总效应不显著，仅通过 M1 产生微弱中介作用。

调节作用 示例数据

仅支持单个自变量，因变量和调节变量

| | A | B | C |
|---|------|------|------|
| 1 | X | Z | Y |
| 2 | 4.00 | 4.00 | 3.27 |
| 3 | 3.25 | 4.25 | 3.91 |
| 4 | 3.00 | 3.00 | 3.00 |
| 5 | 4.00 | 4.25 | 4.00 |
| 6 | 4.75 | 4.50 | 2.87 |

操作过程



分析结果

| 统计量 | 统计量值 | p 值 |
|-----------|--------------------|-----------------------|
| 自变量主效应 | 0.5325322294459093 | 6.303797813897003e-11 |
| 调节变量主效应 | 0.1969672442373264 | 0.026416822231542857 |
| 交互项（调节效应） | 0.299807945018553 | 2.196824195042921e-08 |
| 样本量 | 200.0 | |

• 解释说明

自变量对因变量的主效应：不考虑调节变量时，自变量对因变量的影响。

调节变量对因变量的主效应：不考虑自变量时，调节变量对因变量的影响。

调节效应：调节变量对自变量和因变量关系的影响。

样本量：参与分析的样本数量。

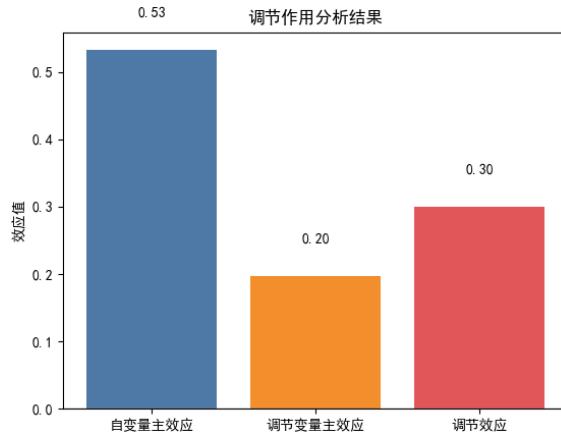
• 结果解读

自变量对因变量的主效应：主效应显著表示自变量对因变量有直接影响。

调节变量对因变量的主效应：主效应显著表示调节变量对因变量有直接影响。

调节效应：调节效应显著表示调节变量改变了自变量和因变量之间的关系。

样本量：样本量的大小会影响统计结果的可靠性，较大的样本量通常能提供更可靠的结果。

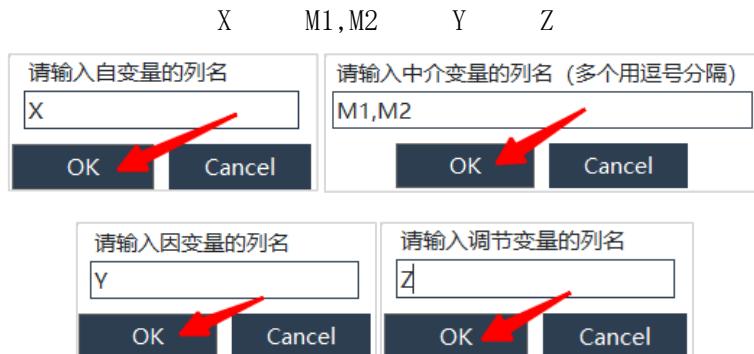


调节中介作用示例数据

X 为自变量，M1, M2 为中介变量，Y 为因变量，Z 为调节变量

| | A | B | C | D | E |
|---|------|---|-----|------|------|
| 1 | X | Z | M1 | M2 | Y |
| 2 | | 4 | 4 | 3.93 | 4 |
| 3 | 4.25 | | 4 | 3.8 | 3 |
| 4 | | 3 | 3 | 3 | 3 |
| 5 | 4.25 | | 4 | 4 | 5 |
| 6 | 4.5 | 5 | 4.2 | 3.33 | 3.89 |

操作过程



分析结果

| 统计量 | 统计量值 | p 值 |
|------------------------------|---------------------|------------------------|
| 自变量对因变量的总效应 | 0.35223523782519284 | 1.940046845255113e-15 |
| 自变量对中介变量 [M1] 的效应 | 0.5503369780959803 | 3.757263766982248e-53 |
| 调节变量对自变量 - 中介变量 [M1] 关系的调节效应 | 0.0601776064666196 | 0.014835169488449904 |
| 中介变量 [M1] 对因变量的效应 (控制自变量) | 0.6770999044420314 | 5.196710581074462e-36 |
| 调节变量对中介变量 [M1] - 因变量关系的调节效应 | 0.231391533592475 | 1.8678676616744122e-07 |
| 中介变量 [M1] 的中介效应 | 0.37263311527970455 | |
| 自变量对中介变量 [M2] 的效应 | 0.41959310980364317 | 1.4142482761811178e-18 |
| 调节变量对自变量 - 中介变量 [M2] 关系的调节效应 | 0.2699284564223313 | 1.456405850377411e-08 |

| | | |
|-----------------------------|----------------------|-----------------------|
| 中介变量 [M2] 对因变量的效应 (控制自变量) | 0. 5836509397674428 | 1. 90853075387428e-59 |
| 调节变量对中介变量 [M2] - 因变量关系的调节效应 | 0. 03626453624600978 | 0. 40490951609045367 |
| 中介变量 [M2] 的中介效应 | 0. 24489591285684015 | |
| 样本量 | 540. 0 | |

自变量对因变量的总效应：自变量对因变量的总效应，总效应显著，表明自变量对因变量有直接影响。

自变量对中介变量的效应：自变量对中介变量的效应，效应显著，表明自变量可以影响中介变量。

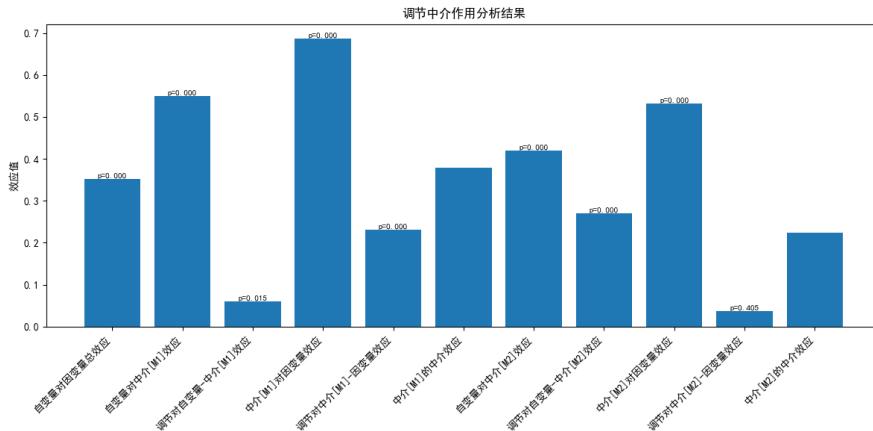
中介变量对因变量的效应(控制自变量)：控制自变量时中介变量对因变量的影响，效应显著，表明中介变量在控制自变量后仍然对因变量有影响。

调节变量对自变量-中介变量关系的调节效应：调节变量对自变量与中介变量之间关系的调节作用，调节效应显著，表明调节变量影响自变量与中介变量之间的关系。

调节变量对中介变量-因变量关系的调节效应：调节变量对自变量与中介变量之间关系的调节作用，调节效应显著，表明调节变量影响中介变量与因变量之间的关系。

中介效应：自变量通过中介变量对因变量的间接影响，中介效应显著，表明自变量通过中介变量间接影响因变量。

样本量：参与分析的样本数量，样本量影响统计结果的可靠性。样本量越大，通常结果越可靠。



结论

- 基础关系成立：X 对 Y 有显著正向总效应，且能通过 M1 (核心) 和 M2 (次要) 间接影响 Y (两条中介路径都成立)；
- 调节变量很关键：调节变量会增强 3 条路径：X→M1、X→M2、M1→Y (比如天赋越高，刷题对知识点 / 解题速度的提升越强，知识点对分数的提升也越强)；
调节变量对 1 条路径无效：M2→Y (不管天赋高低，解题速度对分数的影响都一样)；
- M1 比 M2 更重要：M1 的中介效应 (0.373) 比 M2 (0.245) 强，且 X 对 M1 的影响、调节变量对 M1 路径的调节作用都更突出。

3.5 定价（价格敏感度测试模型）

分析潜在用户对于不同价格的接受程度，从而制定合适的产品价格。

问卷设计

| 用户对某品牌文具盒不同价格时的态度 | | | | | | | |
|-------------------|------|------|------|------|------|------|------|
| 问题/价格 | 10 元 | 20 元 | 30 元 | 40 元 | 50 元 | 60 元 | 70 元 |
| 请选择你觉得‘很划算’的价格 | | | | | | | |
| 请选择你觉得‘便宜’的价格 | | | | | | | |
| 请选择你觉得‘有点贵’的价格 | | | | | | | |
| 请选择你觉得‘太贵’的价格 | | | | | | | |

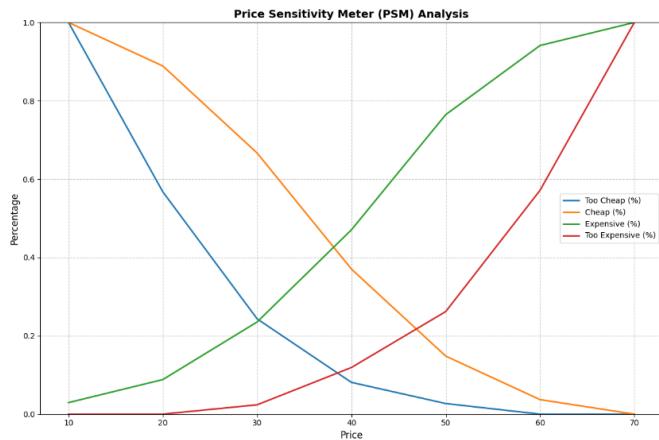
示例数据

| Price | TooCheap | Cheap | Expensive | TooExpensive |
|-------|----------|-------|-----------|--------------|
| 10 | 80 | 15 | 5 | 0 |
| 20 | 60 | 30 | 10 | 0 |
| 30 | 30 | 40 | 25 | 5 |
| 40 | 10 | 30 | 40 | 20 |
| 50 | 5 | 15 | 50 | 30 |
| 60 | 0 | 5 | 30 | 65 |
| 70 | 0 | 0 | 10 | 90 |

- 列名必须为：TooCheap、Cheap、Expensive、TooExpensive（区分大小写）
- 第一列应为价格点
- 值表示对应价格点下选择该选项的人数

分析结果

| Metric | Value |
|-----------------------------|-------|
| 无差异点 (Indifference Point) | 38.11 |
| 最优价格点 (Optimal Price Point) | 38.52 |
| 价格下限 (Lower Bound) | 10.00 |



结论

- 无差异点 (38.11 元): 在这个价格上, 觉得“便宜 / 太便宜”的人, 和觉得“贵 / 太贵了”的人数量刚好相等, 用户对价格的接受度达到平衡。
- 最优价格点 (38.52 元): 这是最推荐的定价。在这个价格下, 既不会因为太便宜损失利润, 也不会因为太贵导致用户流失, 能最大化销量和利润的平衡。
- 价格下限 (10.00 元): 低于 10 元, 大部分人 (80%) 会觉得“太便宜”, 可能会怀疑产品质量, 反响影响购买, 所以 10 元是不能再低的底价。
- 价格上限 (N/A): 表格里没给出具体数值, 但结合前面的数据能推断, 超过 60 元后, 65% 的人觉得太贵; 70 元时 90% 的人觉得太贵, 所以实际价格上限大概在 50-60 元之间, 超过就会大量流失用户。

3.6 问卷可靠性（信度和重测信度）

“信度”检验“被调研人的回答结果”是否可靠, 即“被调研人”有没有真实作答。它也是衡量“内部一致性”的方法(内部一致性即“问卷题目是否都在测量同一个核心概念”, 比如测“幸福感”的题目是否都围绕“幸福感”展开, 而非混入“压力”相关题目)。需要注意的是, 信度分析仅针对量表数据(如“问卷设计”所示)。

“重测信度”检测问卷的稳定性, 也就是“同一批人在不同时间做同一份测试, 结果会不会差太多”。比如, 让 100 个学生今天做一次“数学能力测试卷”, 1 个月后再让这 100 人做同一份卷(确保他们没刻意复习), 如果两次分数的相关性很高, 说明这份试卷的“重测信度高”(结果稳定, 不受时间影响)。

“信度”和“重测信度”相似, 区别在于,

- 如果我们想知道“问卷里的 10 个题目是不是都在测‘工作满意度’”(也就是, 题目 1 “我喜欢我的工作”和题目 5 “我对工作很满意”的结果是否一致), 这时候用“信度”(只需让员工填一次问卷即可)。
- 如果我们想知道“这份问卷的结果是否稳定”(比如员工这个月填的满意度, 和 3 个月后填的是否差不多, 排除“情绪波动”等短期因素影响), 这时候用“重测信度”(需要同一批员工填两次问卷)。

问卷设计

| 问题 A 用户对“APP 易用性”量表 | | 计分：1 = 完全不同意，5 = 完全同意 |
|-----------------------|----------------------------|-----------------------|
| A1 | 我能快速熟悉该 APP 的功能布局。 | |
| A2 | 我能轻松找到自己需要的功能（如搜索、设置）。 | |
| A3 | 该 APP 的操作步骤简单，不需要反复尝试。 | |
| A4 | 使用该 APP 时，我很少遇到卡顿或操作失误的问题。 | |

信度示例数据

注意！DIAS 对数据进行了标准化操作，输出的 Excel 文件中的结果是分析标准化后的数据而得到。

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | A1 | A2 | A3 | A4 | B1 | B2 | B3 | B4 | C1 | C2 | C3 | D1 | D2 | D3 |
| 2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 5 | 4 | 4 | 3 |
| 3 | 5 | 5 | 4 | 4 | 5 | 5 | 5 | 4 | 5 | 5 | 4 | 5 | 5 | 5 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 2 | 4 | 4 | 4 | 4 | 4 | 3 |
| 5 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

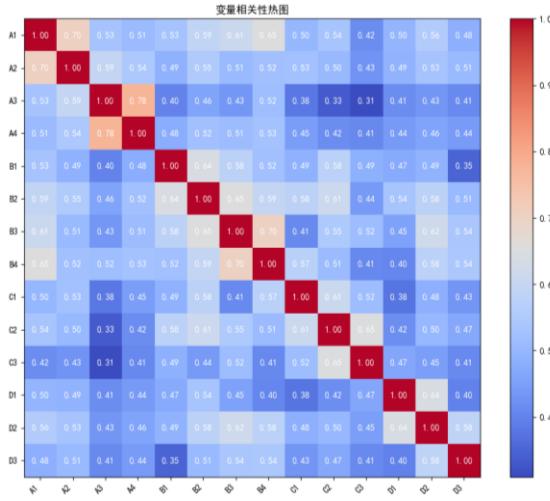
分析结果

| 统计量 | 统计量值 |
|---------------------|---|
| Cronbach's Alpha 系数 | 0.9351 |
| 样本量 | {'A1': 200, 'A2': 200, 'A3': 200, 'A4': 200, 'B1': 200, 'B2': 200, 'B3': 200, 'B4': 200, 'C1': 200, 'C2': 200, 'C3': 200, 'D1': 200, 'D2': 200, 'D3': 200} |
| 均值 | {'A1': -4.973799150320702e-16, 'A2': 2.9087843245179103e-16, 'A3': 1.9984014443252818e-16, 'A4': 2.042810365310288e-16, 'B1': 2.3980817331903383e-16, 'B2': 5.329070518200751e-16, 'B3': -2.6645352591003756e-16, 'B4': -3.552713678800501e-17, 'C1': 0.0, 'C2': -4.529709940470639e-16, 'C3': 2.486899575160351e-16, 'D1': 4.263256414560601e-16, 'D2': -1.7763568394002506e-16, 'D3': 1.7763568394002505e-17} |

Cronbach's Alpha 系数：Cronbach's Alpha 系数用于衡量量表内部一致性信度。其值范围为 0 到 1，越接近 1，信度越高。

样本量：每个样本的观测值个数，样本量影响统计检验的稳定性。样本量越大，通常结果越可靠。

均值：样本数据的平均值，均值反映数据的平均水平，可以用来比较不同变量的集中趋势。



结论

Cronbach's α 系数 0.9351 (信度极高), 样本量 200 (充足可靠), 各题项均值分布合理, 内部一致性极强, 量表完全可靠, 可以用于后续的数据分析。

重测信度 示例数据

它默认把数据分成两半, 例如一共有八列, 默认前四列是前一半, 后四列是后一半, 因此在收集数据时要确保列数为偶数。

| | A | B | C | D | E | F | G | H |
|---|------|------|------|------|------|------|------|------|
| 1 | 测试 1 | 测试 2 | 测试 3 | 测试 4 | 重测 1 | 重测 2 | 重测 3 | 重测 4 |
| 2 | 85 | 90 | 78 | 82 | 83 | 88 | 79 | 80 |
| 3 | 76 | 82 | 88 | 79 | 75 | 80 | 86 | 77 |
| 4 | 92 | 88 | 95 | 90 | 90 | 86 | 93 | 89 |
| 5 | 65 | 70 | 62 | 68 | 63 | 69 | 60 | 66 |
| 6 | 88 | 92 | 85 | 90 | 86 | 90 | 83 | 88 |

分析结果

| 变量名 | ICC (2,1) | 95% 置信区间 |
|------|-----------|--------------|
| 测试 1 | 0.9819 | [[0.06 1.]] |
| 测试 2 | 0.9816 | [[0.06 1.]] |
| 测试 3 | 0.9841 | [[0.4 1.]] |
| 测试 4 | 0.9804 | [[0.05 1.]] |

| 统计量 | 统计量值 |
|-------------|-------|
| 平均 ICC(2,1) | 0.982 |

- 解释说明

ICC (2,1) 重测信度系数: 用同一种测验在不同时间对同一组被试进行两次测量, 两次测量结果的相关系数。

样本量: 每个样本中的观测值数量。

均值: 样本数据的平均值。

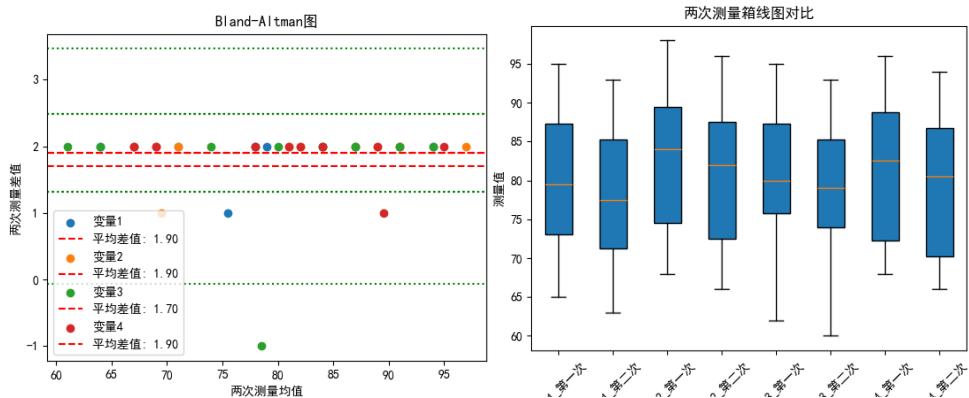
- 结果解读

ICC (2,1) 重测信度系数: 重测信度系数越接近 1, 表示测验结果越稳定; 越接近 0, 表示测验结果的稳定性越差。

样本量: 样本量的大小会影响统计检验的稳定性, 较大的样本量通常能提供更可

靠的结果。

均值：均值反映了数据的平均水平，可用于比较不同变量的集中趋势。



结论

- 平均 $ICC(2,1)=0.982$: 这是重测信度的核心指标，取值 0-1。0.982 已经非常接近 1，说明两次测量（比如“测试 1”和“重测 1”、“测试 2”和“重测 2”）的结果几乎一致，数据稳定性极强。
- 平均样本量 = 10: 每次测量用了 10 个样本，样本量虽不算大，但结合极高的 ICC 值，结果依然可信。

3.7 问卷有效性（内容效度与效度）

检验调查问卷的有效性。“效度”又称“结构效度”，它和“内容效度”的用途是一样的。区别在于，

| 对比 | 内容效度 | 效度 |
|------|---|--|
| 评估方法 | 依赖专家判断，邀请领域专家对每个题目打分 | 依赖统计分析 |
| 适用场景 | 问卷设计的初期阶段，测量目标是具体、可明确界定的内容（如“小学三年级数学知识点”“员工岗位职责”） | 问卷设计的中后期阶段，测量目标是抽象、多维度的概念（如“幸福感”“领导力”“焦虑”等无法直接观测的变量） |
| 局限性 | 可能受专家经验、视角影响，客观性较弱。 | 若样本量小或统计方法不当，可能得出错误结论；且需要明确的理论假设支撑。 |

问卷设计

小明设计了一份问卷，一共 20 道题。他希望通过专家打分的方式来判断问卷的有效性，避免个人主观臆断的情况。小明找到 8 位专家，分别给每一道题打分。

您认为第一题的有效性如何？（1 分表示较好，0 分表示较差）

您认为第二题的有效性如何？（1 分表示较好，0 分表示较差）

...

内容效度 示例数据

它更适合分析二分类的数据 (1, 0), 也兼容四分类数据。

二分类数据的意思是选项只有两个, 四分类数据的意思是选项只有四个。

| | A | B | C | D | E | F | G | H | I |
|---|------|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 测量项 | 专家1 | 专家2 | 专家3 | 专家4 | 专家5 | 专家6 | 专家7 | 专家8 |
| 2 | 测量项1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3 | 测量项2 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| 4 | 测量项3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | 测量项4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6 | 测量项5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | 测量项6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

如果是其他分类的数据, 可以先转化为二分类数据, 如下表, 为四分类数据 (1, 2, 3, 4), 可以将 3 和 4 转换为 1, 将 1, 2 转换为 0。

| | A | B | C | D | E | F | G | H | I |
|---|------|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 测量项 | 专家1 | 专家2 | 专家3 | 专家4 | 专家5 | 专家6 | 专家7 | 专家8 |
| 2 | 测量项1 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 3 |
| 3 | 测量项2 | 4 | 4 | 3 | 4 | 4 | 3 | 4 | 3 |
| 4 | 测量项3 | 3 | 3 | 4 | 4 | 3 | 3 | 4 | 4 |
| 5 | 测量项4 | 4 | 3 | 4 | 3 | 4 | 3 | 3 | 3 |
| 6 | 测量项5 | 4 | 3 | 4 | 3 | 3 | 3 | 4 | 3 |

分析结果

| 统计量 | 统计量值 |
|----------------------|--|
| 平均内容效度比 (CVR) | 0.882 |
| 量表内容效度指数 (S-CVI/UA) | 0.000 |
| 量表内容效度指数 (S-CVI/Ave) | 0.000 |
| 专家数量 | 19 |
| CVR 临界值 | 0.210 |
| I-CVI 临界值 | 0.8 |
| 样本量 | 专家 1: 19, 专家 2: 19, 专家 3: 19, 专家 4: 19, 专家 5: 19, 专家 6: 19, 专家 7: 19, 专家 8: 19 |
| 均值 | 专家 1: 0.947, 专家 2: 0.947, 专家 3: 0.947, 专家 4: 0.947, 专家 5: 0.947, 专家 6: 0.895, 专家 7: 0.947, 专家 8: 0.947 |
| 标准差 | 专家 1: 0.229, 专家 2: 0.229, 专家 3: 0.229, 专家 4: 0.229, 专家 5: 0.229, 专家 6: 0.315, 专家 7: 0.229, 专家 8: 0.229 |
| 中位数 | 专家 1: 1.000, 专家 2: 1.000, 专家 3: 1.000, 专家 4: 1.000, 专家 5: 1.000, 专家 6: 1.000, 专家 7: 1.000, 专家 8: 1.000 |
| 偏度 | 专家 1: -4.359, 专家 2: -4.359, 专家 3: -4.359, 专家 4: -4.359, 专家 5: -4.359, 专家 6: -2.798, 专家 7: -4.359, 专家 8: -4.359 |
| 峰度 | 专家 1: 19.000, 专家 2: 19.000, 专家 3: 19.000, 专家 4: 19.000, 专家 5: 19.000, 专家 6: 6.509, 专家 7: 19.000, 专家 8: 19.000 |
| 变异系数 | 专家 1: 0.242, 专家 2: 0.242, 专家 3: 0.242, 专家 4: 0.242, 专家 5: 0.242, 专家 6: |

| | |
|--------------------------|--|
| | 0.352, 专家 7: 0.242, 专家 8: 0.242 |
| p 值 ($p < 0.05$ 表示统计显著) | 专家 1: 0.000, 专家 2: 0.000, 专家 3: 0.000, 专家 4: 0.000, 专家 5: 0.000, 专家 6: 0.000, 专家 7: 0.000, 专家 8: 0.000 |

- 统计量解释说明

平均内容效度比 (CVR)：平均内容效度比用于衡量测量工具中各个题项与测量内容的相关性，取值范围在 -1 到 1 之间，越接近 1 表示相关性越强。

样本量：每个样本中的观测值数量。

均值：样本数据的平均值。

标准差：样本数据的离散程度。

中位数：样本数据的中间值。

偏度：样本数据分布的偏斜程度。

峰度：样本数据分布的峰态程度。

- 统计结果解读

平均内容效度比 (CVR)：平均内容效度比越接近 1，说明测量工具的内容与所测量的概念或领域相关性越强，内容效度越高。

样本量：样本量的大小会影响统计检验的稳定性，较大的样本量通常能提供更可靠的结果。

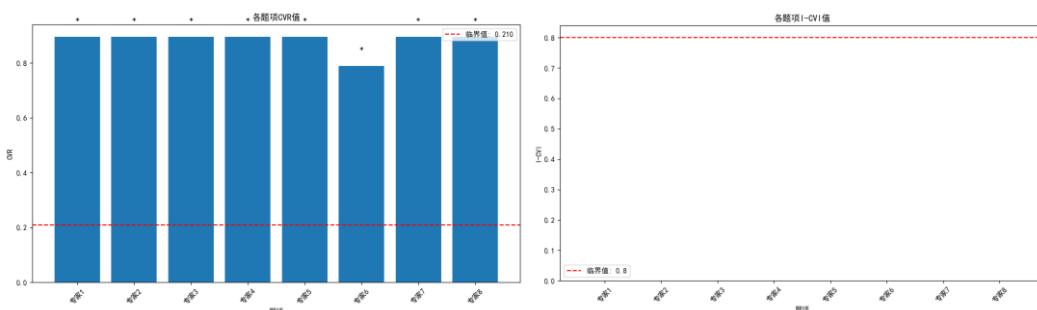
均值：均值反映了数据的平均水平，可用于比较不同变量的集中趋势。

标准差：标准差越大，说明数据的离散程度越大。

中位数：中位数不受极端值的影响，能更好地反映数据的中间水平。

偏度：偏度为正表示数据右偏，偏度为负表示数据左偏。

峰度：峰度大于 3 表示数据分布比正态分布更尖峭，峰度小于 3 表示数据分布比正态分布更平坦。



结论

- 内容效度极高：CVR=0.882，说明问卷的题目完全贴合测量目标；
- 专家评价高度一致：所有专家的平均分都在 0.8 以上，标准差小、中位数为 1.0，说明专家们对问卷的认可度几乎没有分歧；
- 数据结果稳定：偏度、峰度显示评价集中在高分段，没有异常值或极端分歧。

效度 示例数据

分类数据或连续数据均可。

| | A | B | C | D |
|---|------|------|------|------|
| 1 | 题项 1 | 题项 2 | 题项 3 | 题项 4 |
| 2 | 4 | 5 | 4 | 1 |
| 3 | 5 | 4 | 5 | 2 |
| 4 | 3 | 4 | 4 | 2 |
| 5 | 5 | 5 | 5 | 1 |
| 6 | 2 | 2 | 2 | 5 |

分析结果

| 统计量 | 统计量值 | | | | |
|-------------------|---|--|--|--|--|
| KMO 检验值 | 0.7434807907593524 | | | | |
| Bartlett 球形检验 p 值 | 2.1584363295269473e-05 | | | | |
| 因子载荷矩阵 | 0 1 2 3 题项 1 0.9781189449228391 0.08854963999415362 0.17447731521674942 0.0 题项 2 0.9237221245579419 -0.287485639000527 0.05612341304368745 0.0 题项 3 0.8945129805663484 0.3029875818730911 -0.06262738049919098 0.0 题项 4 -0.8830309355441274 0.10427897419705112 0.18853359979846462 0.0 | | | | |
| 样本量 | 题项 1: 10 题项 2: 10 题项 3: 10 题项 4: 10 | | | | |
| 均值 | 题项 1: 3.4 题项 2: 3.8 题项 3: 3.9 题项 4: 2.7 | | | | |

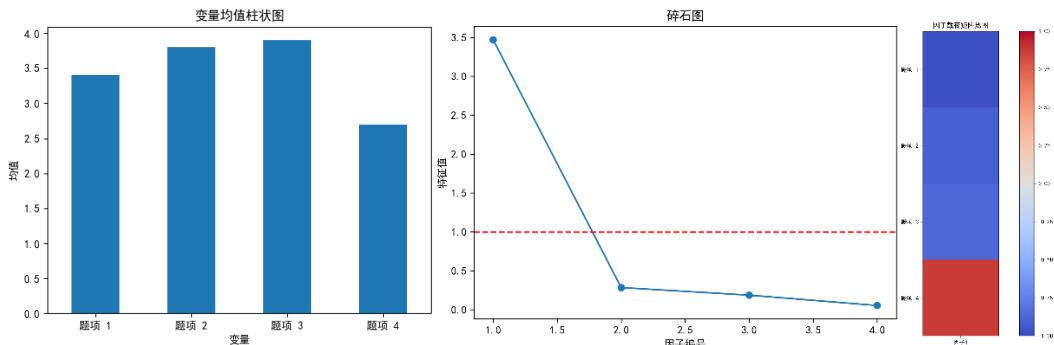
KMO 检验值：衡量数据是否适合进行因子分析。取值范围为 0 到 1，越接近 1，越适合。

Bartlett 球形检验 p 值：用于检验变量之间是否存在相关性，p 值小于 0.05，表示变量之间存在相关性，适合进行因子分析。

因子载荷矩阵：反映各变量与各因子之间的相关性，绝对值越大表明该变量与相应因子的相关性越强。

样本量：每个样本的观测值个数，影响统计检验的稳定性，样本量越大通常结果越可靠。

均值：样本数据的平均值，反映数据的平均水平，用来比较同变量的集中趋势。



结论

- KMO=0.743、Bartlett p<0.05 → 4 个题目数据适合做因子分析；
- 4 个题目都指向同一个核心因子 → 说明设计的题目“目标一致”，没有跑偏；
- 题目 4 和前 3 个题目“方向相反”（载荷为负、均值低）→ 要么是“反向题”（设计合理），要么是题目设计有歧义（需要检查题目表述，比如前 3 个问“我喜欢 XX”，题目 4 问“我不喜欢 XX”，就是合理的反向题）。

3.8 亲疏远近（多维尺度分析）

把复杂的“相似 / 不同”关系，变成 1 张能看懂的地图。

比如，对比 8 款手机（华为、苹果、小米、OPPO 等），但要考虑的维度太多：价格、处理器性能、电池容量、摄像头像素和重量一共 5 个维度，脑子根本记不住谁和谁更像。

示例数据

| | A | B | C | D | E | F |
|---|------------------|-------|----------|-----------|----------|-------|
| 1 | 手机型号 | 价格（元） | 处理器性能（分） | 电池容量（mAh） | 摄像头像素（分） | 重量（g） |
| 2 | 苹果 iPhone 15 Pro | 9999 | 9.5 | 3561 | 9 | 206 |
| 3 | 华为 Pura 70 Ultra | 8999 | 9.2 | 5000 | 9.5 | 225 |
| 4 | 小米 14 Pro | 4999 | 9 | 5000 | 9 | 220 |
| 5 | vivo X100 Pro+ | 5999 | 9 | 5400 | 9.3 | 221 |
| 6 | OPPO Find X7 Pro | 5400 | 9 | 5000 | 9 | 216 |

分析结果

| 统计量 | 统计量值 |
|--------|---|
| MDS 坐标 | [[331.9041867449383, 3804.576935470109], [-694.054845486214, 2412.651379071943], [457.430261538917, -1401.9141304257514], [-244.8600451529959, -583.6095192704736], [294.9805233414133, -929.6000029575004], [-1359.20165966331, 4303.471499557848], [1488.064177234199, -4014.8908637876807], [-274.2625985569471, -3590.6852976584964]] |

- 统计量解释

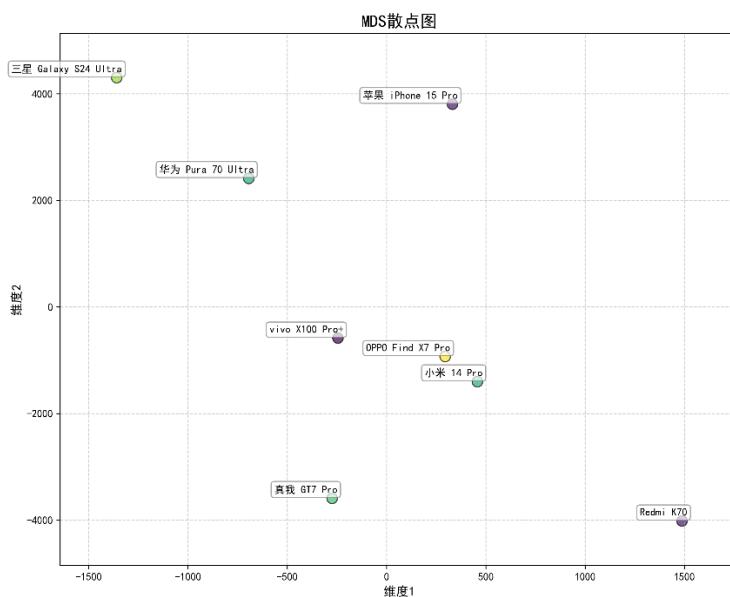
MDS 坐标：经过多维尺度分析后得到的低维坐标

MDS 散点图：展示样本在低维空间中的分布情况

- 统计量解读

MDS 坐标：可用于观察样本在低维空间中的相对位置关系

MDS 散点图：直观展示样本之间的相似性，距离近的样本更相似



结论

- 第一类：OPPO Find X7 Pro 和小米 14 Pro，二者距离极近，在多维特征上具有很高的相似性，可能在性能、配置等多个方面较为接近。
- 第二类：苹果 iPhone 15 Pro 和华为 Pura 70 Ultra，二者距离相对较近，在特征上有较多的共同之处。
- 而三星 Galaxy S24 Ultra、真我 GT7 Pro、Redmi K70 则相对独立，各自在特征上有较为鲜明的特点，与其他机型区分度较高。

3.9 问卷多选题

计算每个选项的总的选择次数。(例如：选项 A 被选择了多少次)

示例数据

| | A | B | C | D | E |
|----|-----|-------|-------|-------|-----|
| 1 | 题目1 | 题目2 | 题目3 | 题目4 | 题目5 |
| 2 | A,B | A,C | A,B,D | A,D | B,C |
| 3 | B,C | B,D | A,C | B,C | A,D |
| 4 | A,D | A,B,D | B,C | A,C,D | A,B |
| 5 | B,D | C,D | A,D | A,B | B,D |
| 总计 | 53 | 56 | 49 | 52 | 48 |

分析结果

| 统计量 | 统计量值 |
|--------|--------------------------------------|
| 选项选择次数 | {'A': 53, 'B': 56, 'C': 49, 'D': 52} |
| 样本量 | 20 |

解释说明

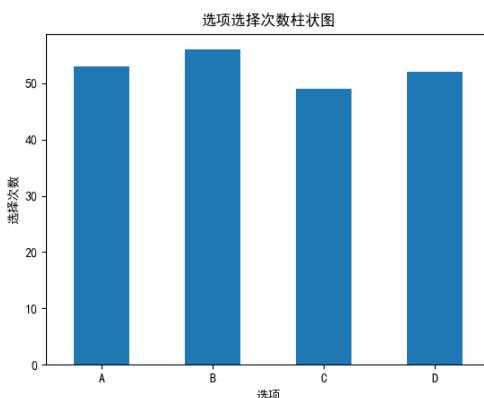
选项选择次数：每个选项被选择的总次数。

样本量：参与问卷的总人数。

结果解读

选项选择次数：选择次数越多，说明该选项在问卷中越受欢迎。

样本量：样本量的大小



3.10 评分可靠性（组内评分者信度）

验证评分数据的可靠性和有效性，确保数据质量，支持结论的可信度等，原理是评估同一组内不同评分者对相同对象的评分一致性。

例如，在课堂教学评估中，若 3 名评委对同一堂课的评分一致性低（如分别打 2 分、4 分、5 分），则评估结果难以用于判断教学质量；而高一致性的评分（如 4 分、4 分、5 分）则更具说服力。

示例数据

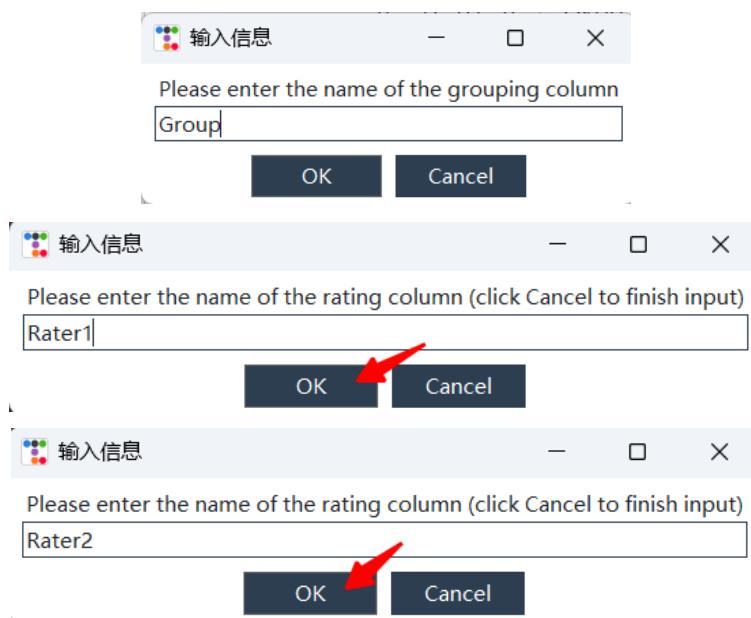
数据需满足正态分布和方差齐，不满足时也可以分析，但是结果中会有“NAN”。

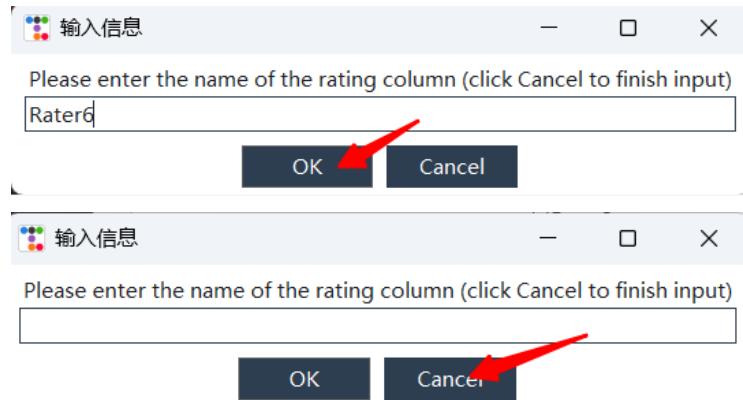
程序对评分数据的量表等级没有严格限制，只要数据是数值型评分（如 1-5 分、1-7 分、0-10 分等李克特量表，或其他数值评分形式）均可处理。程序对分组数量没有固定上限，每组的样本量需 ≥ 2 ，否则该组会被自动跳过。

| | A | B | C | D | E | F | G |
|---|-------|--------|--------|--------|--------|--------|--------|
| 1 | Group | Rater1 | Rater2 | Rater3 | Rater4 | Rater5 | Rater6 |
| 2 | 1 | 4 | 2 | 1 | 1 | 2 | 2 |
| 3 | 1 | 4 | 3 | 3 | 1 | 3 | 2 |
| 4 | 1 | 5 | 2 | 2 | 3 | 2 | 3 |
| 5 | 1 | 4 | 3 | 3 | 2 | 3 | 2 |
| 6 | 1 | 4 | 4 | 5 | 4 | 1 | 4 |

操作过程

Group Rater1 Rater2 Rater3 Rater4 Rater5 Rater6





分析结果

| 指标 | 数值 |
|--------------------|---------|
| 1_rwg 值 | -0.7500 |
| 2_rwg 值 | -0.2500 |
| Kruskal-Wallis_H 值 | 2.8783 |
| Kruskal-Wallis_p 值 | 0.0898 |
| MSB | NaN |
| MSW | NaN |
| Rwg 值标准差 SD | 0.2500 |
| P25 | -0.6250 |
| 中位数 | -0.5000 |
| P75 | -0.3750 |
| ICC1 | NaN |
| ICC2 | NaN |

- 解释说明

rwg 值：组内评分者信度 rwg 用于评估组内成员评分的一致性，值越接近 1 表示一致性越高。

Rwg 值标准差 SD：Rwg 值的标准差，反映了 Rwg 值的离散程度。

P25：Rwg 值的第 25 百分位数。

...

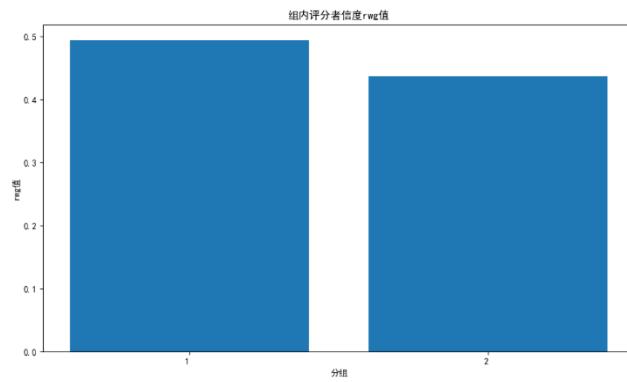
- 结果解读

rwg 值：rwg 值越接近 1，说明组内成员的评分越一致；值越低，说明组内成员的评分差异越大。

Rwg 值标准差 SD：标准差越大，说明 Rwg 值的离散程度越大。

P25：第 25 百分位数较低表示有 25% 的 Rwg 值低于该值。

...



结论

- 所有组（尤其是第 2 组）的评分者一致性都很差，组内评分者打分标准不统一，关联性弱，无法形成可靠的评分结果，可信度很低。
- 第 1 组和第 2 组的整体评分结果有显著差异（不是巧合），但因为两组内部一致性都差，这种“组间差异”也不能说明“两组评分标准有本质区别”，更可能是组内评分混乱导致的。

4 相关性分析

4.1 骡子还是马 (Cochran's Q 检验)

同一批人 / 东西，在 3 种及以上不同条件下，针对“成或不成”“行或不行”这种二选一的问题，其结果是否有明显差别。例如，

- 老师想知道“刷题”“背笔记”“看错题”这 3 种复习方法，对同一批 15 个学生的“及格 / 不及格”有没有影响。
- 同一批面包，3 种储存方式，看“会不会发霉”。
- 同一批用户，3 种 APP 界面，看“能不能找到功能”。
- 同一批病人，3 种止咳药，看“能不能止咳嗽”。

示例数据

- 至少包含 3 列数据，“用户 ID”不包含在内。
- 数据为二分类数据（0 和 1，或可转换为 0 和 1），代表“无效”和“有效”。

| | A | B | C | |
|---|--------|--------|--------|---|
| 1 | Plan A | Plan B | Plan C | |
| 2 | | 1 | 0 | 1 |
| 3 | | 0 | 0 | 1 |
| 4 | | 1 | 1 | 1 |
| 5 | | 1 | 0 | 0 |
| 6 | | 0 | 1 | 1 |

分析结果

| 统计量 | 统计量值 | p 值 |
|----------------|--|--------|
| Cochran's Q 检验 | 0.1818 | 0.9131 |
| 样本量 | {'Plan A': 15, 'Plan B': 15, 'Plan C': 15} | |
| 阳性比例 | {'Plan A': '0.6000', 'Plan B': '0.5333', 'Plan C': '0.6000'} | |

Cochran's Q 检验：用于比较三个或三个以上相关二元变量样本的分布是否显著不同。

样本量：每个样本的观测值个数，影响统计检验的功效，样本量越大，结果越准确。

阳性比例：每个样本中阳性结果的比例，反映样本中阳性结果的相对数量，可用于比较不同样本的阳性情况。

p 值：当 p 值小于显著性水平（通常为 0.05）时，表明样本间存在显著差异；否则，表明样本间无显著差异。

结论

- 方案无差异：p 值 0.9131 → 用户对 A、B、C 三个方案的阳性评价（比如愿意购买）没有显著差异；
- 比例很接近：A 和 C 的阳性比例 60%，B 是 53.33% → 整体评价都偏“积极”，且差距很小，三个方案的用户接受度在同一水平；
- 样本待优化：15 个样本 → 结论对“这 15 人”有效，若想推广到更多用户，建议增加样本量（比如 50 人），让结论更有普适性。

4.2 两人观点一致性 (Kappa 一致性检验)

两个人 / 方法，对同一批“东西”做判断时，他们的“看法”到底有多一致，而且会排除“瞎猜碰巧一致”的情况。例如，

- 两个医生看同一批 CT 片，判断“有没有肺癌”。
- 两位 UI 设计师对“APP 界面风格分类”的一致性验证。
- 用户和设计专家对口红包装的“吸引力”做“高 / 中 / 低”三级分类，判断“用户感知”与“专家判断”是否一致。

示例数据

数据必须为两列

| | A | B |
|---|-----|-----|
| 1 | A医生 | B医生 |
| 2 | | 1 |
| 3 | | 2 |
| 4 | | 3 |
| 5 | | 1 |
| 6 | | 2 |

分析结果

| 统计量 | 统计量值 | p 值 |
|------------|--------|--------------------------|
| 样本量 | 9 | |
| 未加权 Kappa | 0.5000 | 95% CI: (0.1000, 0.8333) |
| 线性加权 Kappa | 0.5385 | 95% CI: (0.0526, 0.8889) |
| 二次加权 Kappa | 0.5714 | 95% CI: (0.0000, 0.9412) |

| 混淆矩阵 | | | |
|------|---|---|---|
| | 1 | 2 | 3 |
| 1 | 1 | 0 | 0 |
| 2 | 1 | 3 | 1 |
| 3 | 1 | 0 | 2 |

- 未加权 Kappa：不考虑分类错误严重程度的一致性指标
- 线性加权 Kappa：考虑分类错误线性严重程度的一致性指标
- 二次加权 Kappa：考虑分类错误二次严重程度的一致性指标
- 95%置信区间：用于估计统计量真实值的范围（95%置信水平）

结论

根据 Kappa 值 0.4-0.6 的范围，两个评估者的分类结果是“中度一致”，不算特别准，但也不是瞎判，处于中等水平。

4.3 多人观点一致性 (Kendall 协调系数)

Kendall 协调系数的用途是判断“多个评分者（比如评委、专家）对同一批项目（比如选手、作品）的排名 / 评分是否一致”，简单说就是“一群人对同一组事物的评价是否有共识？”，避免因“评分者看法混乱”导致评价结果不可信。例如，

- 专家对产品的评审一致性检验。
- 用户对服务的满意度评价一致性。

示例数据

| 评委 | 选手 1 | 选手 2 | 选手 3 | 选手 4 | 选手 5 | 选手 6 | 选手 7 | 选手 8 | 选手 9 | 选手 10 |
|----|------|------|------|------|------|------|------|------|------|-------|
| 1 | 9 | 2 | 4 | 10 | 7 | 6 | 8 | 5 | 3 | 1 |
| 2 | 10 | 1 | 3 | 8 | 7 | 5 | 9 | 6 | 4 | 2 |
| 3 | 8 | 4 | 2 | 10 | 9 | 7 | 5 | 6 | 3 | 1 |
| 4 | 9 | 1 | 2 | 10 | 6 | 7 | 4 | 8 | 5 | 3 |

分析结果

| 统计量 | 统计量值 | p 值 |
|------------------|---------|----------|
| Kendall 协和系数 (W) | 0.853 | 0.000332 |
| 评价者数量 | 4 | |
| 项目数量 | 10 | |
| 总样本量 | 40 | |
| 中位数 (按项目) | | |
| - 选手 1 | 9.0000 | |
| - 选手 2 | 1.5000 | |
| - 选手 3 | 2.5000 | |
| - 选手 4 | 10.0000 | |
| - 选手 5 | 7.0000 | |
| - 选手 6 | 6.5000 | |
| - 选手 7 | 6.5000 | |
| - 选手 8 | 6.0000 | |
| - 选手 9 | 3.5000 | |
| - 选手 10 | 1.5000 | |

• 统计量解释说明

Kendall 协和系数：用于衡量多个评价者对多个项目的排序一致性。

评价者数量：参与评价的人数。

项目数量：被评价的项目总数。

总样本量：评价数据的总观测数（评价者数量 × 项目数量）。

中位数 (按项目)：每个项目评分的中位数。

• 统计量结果解读

统计量：Kendall 协和系数的值，范围从 0 到 1，越接近 1 表示一致性越高。

p 值：p 值小于显著性水平（通常为 0.05）时，拒绝原假设，认为评价者之间存在显著一致性；否则，接受原假设，认为评价者之间无显著一致性。

评价者数量：参与评价的人数，影响一致性分析的可靠性。

项目数量：被评价的项目总数，数量越多结果越可靠。

总样本量：总观测数 = 评价者数量 × 项目数量，样本量越大统计检验功效越高。

中位数 (按项目)：每个项目评分的中间值，反映该项目的整体评价水平。

结论

- 评委共识极高：W=0.853、p=0.000332，4 位评委对 10 位选手的评价几乎完全一

致，打分公平可信，没有“各执己见”的混乱情况。

- 选手排名清晰：选手 4 评价最高（10.0），选手 2、10 最低（1.5），中间梯队明确，可以直接用这个结果做最终排名（比如颁奖、晋级）。
- 分析条件合格：4 位评委、10 位选手、40 个总观测值，样本规模合理，结果稳定性有保障，不用怀疑“数据太少导致结论不准”。

4.4 齐步走和一起走（Pearson 和 Spearman 相关性分析）

他们的用途相似，都是用来分析变量之间的相关性，有没有“同方向变化”的趋势。

Pearson 相关性分析

两个人是不是“同步走直线”，比如 A 走 1 步，B 也走 1 步；A 走 2 步，B 也走 2 步，严格按“直线节奏”同步。

Spearman 相关性分析

两个人是不是“同步排先后”，比如 A 排第 1，B 也排第 1；A 排第 2，B 也排第 2，不管他们每步走多远，只要“排名顺序一致”就行。

区分

“背单词数量”和“英语考试分数”的关系

- 背 10 个单词，分数涨 5 分；背 20 个单词，分数涨 10 分；背 30 个单词，分数涨 15 分，呈直线同步，这时候用（Pearson），它能算出“每多背 10 个单词，分数平均涨 5 分”的线性规律；
- 背 10 个单词，分数从 50→60（涨 10 分）；背 20 个单词，分数从 60→68（涨 8 分）；背 30 个单词，分数从 68→75（涨 7 分），不是直线，但“背得越多，分数越高”（排名一直同步），这时候用（Spearman），它不管每步涨多少，只认“背单词数量排名第 1，分数排名也第 1”的规律。

对数据的要求

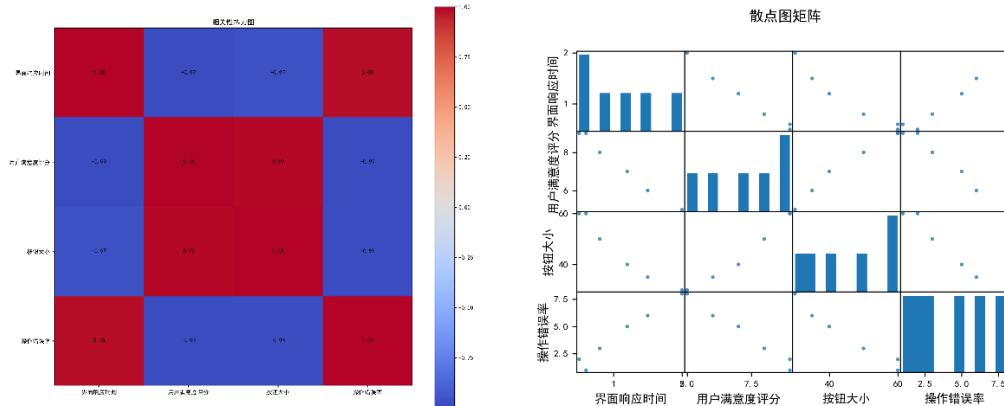
- 数据是“连续数据”，“近似正态分布”，“样本量 >100”，“线性关系”，“无异常值”，“希望精准预测”用 Pearson；
- 数据是“连续或有序分类数据”，“样本量 <100”，“单调关系”，用 Spearman；

Pearson 相关性分析 示例数据

| | A | B | C | D |
|---|--------|---------|------|-------|
| 1 | 界面响应时间 | 用户满意度评分 | 按钮大小 | 操作错误率 |
| 2 | | 0.5 | 9 | 60 |
| 3 | | 1.2 | 7 | 40 |
| 4 | | 0.8 | 8 | 50 |
| 5 | | 2 | 5 | 30 |
| 6 | | 0.6 | 9 | 60 |

分析结果

| 统计量 | 相关系数 | p 值 |
|--|---------------------|-----------------------|
| Pearson 相关系数(界面响应时间(秒)vs 用户满意度评分(1-10)) | -0.9930562493060069 | 7.215611174163112e-05 |
| Pearson 相关系数 (界面响应时间 (秒) vs 按钮大小 (像素)) | -0.969438733277124 | 0.0013867145611640007 |
| Pearson 相关系数 (界面响应时间 (秒) vs 操作错误率 (%)) | 0.980413098030004 | 0.0005717128677394607 |
| Pearson 相关系数 (用户满意度评分(1-10)vs 界面响应时间(秒)) | -0.9930562493060069 | 7.215611174163112e-05 |
| Pearson 相关系数 (用户满意度评分 (1-10) vs 按钮大小 (像素)) | 0.9877421312096074 | 0.0002244621157576952 |
| Pearson 相关系数 (用户满意度评分 (1-10) vs 操作错误率 (%)) | -0.9899011598235936 | 0.0001524648863129213 |
| Pearson 相关系数 (按钮大小 (像素) vs 界面响应时间 (秒)) | -0.969438733277124 | 0.0013867145611640007 |
| Pearson 相关系数 (按钮大小 (像素) vs 用户满意度评分 (1-10)) | 0.9877421312096074 | 0.0002244621157576952 |
| Pearson 相关系数 (按钮大小 (像素) vs 操作错误率 (%)) | -0.9807240381900842 | 0.0005537629413027729 |
| Pearson 相关系数 (操作错误率 (%) vs 界面响应时间 (秒)) | 0.980413098030004 | 0.0005717128677394607 |
| Pearson 相关系数 (操作错误率 (%) vs 用户满意度评分 (1-10)) | -0.9899011598235936 | 0.000152464886312928 |
| Pearson 相关系数 (操作错误率 (%) vs 按钮大小 (像素)) | -0.9807240381900842 | 0.0005537629413027602 |



结论

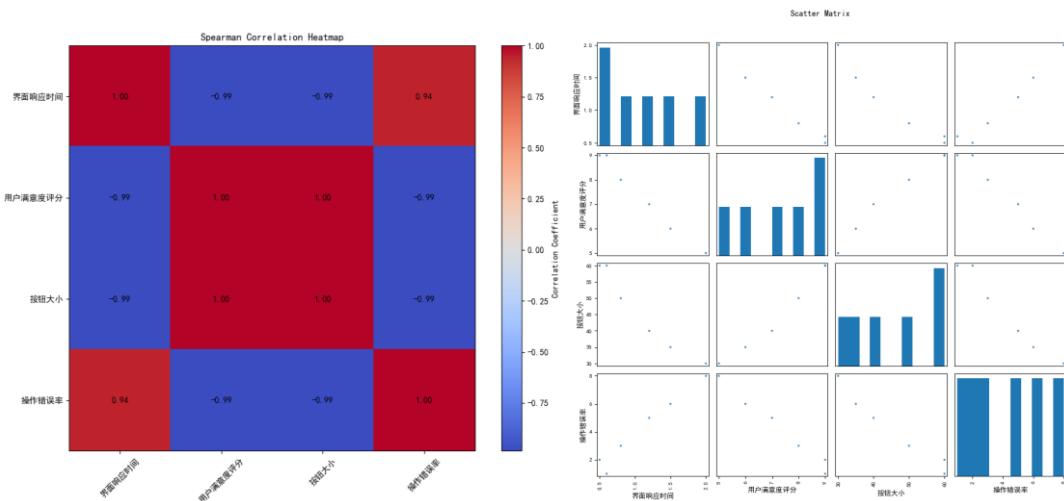
- 优先优化响应时间：这是“核心杠杆”，只要把响应时间降下来（比如从 3 秒降到 1 秒），满意度会涨、错误率会跌、按钮体验也会间接变好；
- 按钮尺寸别太小：保证用户容易点中，能直接减少错误率、提升满意度；
- 结论 100% 可信：所有强关联的 p 值都远小于 0.05，说明这些规律是真实的，不是偶然，按这两点优化，用户体验一定能提升。

Spearman 相关性分析 示例数据

| | A | B | C | D |
|---|--------|---------|------|-------|
| 1 | 界面响应时间 | 用户满意度评分 | 按钮大小 | 操作错误率 |
| 2 | | 0.5 | 9 | 60 |
| 3 | | 1.2 | 7 | 40 |
| 4 | | 0.8 | 8 | 50 |
| 5 | | 2 | 5 | 30 |
| 6 | | 0.6 | 9 | 60 |
| | | | | 1 |

分析结果

| 统计量 | 相关系数 | p 值 |
|-----------------------------------|---------|--------|
| Spearman 相关系数 (界面响应时间 vs 用户满意度评分) | -0.9856 | 0.0 |
| Spearman 相关系数 (界面响应时间 vs 按钮大小) | -0.9856 | 0.0 |
| Spearman 相关系数 (界面响应时间 vs 操作错误率) | 0.9429 | 0.0003 |
| Spearman 相关系数 (用户满意度评分 vs 按钮大小) | 1.0 | 0.0003 |
| Spearman 相关系数 (用户满意度评分 vs 操作错误率) | -0.9856 | 0.0 |
| Spearman 相关系数 (按钮大小 vs 操作错误率) | -0.9856 | 0.0003 |



4.5 偏相关分析

它解决了这个问题：“当排除其他变量的干扰后，两个变量之间的关联还存在吗？”。比如“界面响应时间和用户满意度有关系，会不会是因为‘按钮尺寸’在中间起作用？排除按钮尺寸的影响后，它们还有关系吗？”

偏相关和普通相关(Pearson 和 Spearman)的区别在于是否考虑其他因素的干扰。例如，研究“体重”和“跑步速度”的关系时。

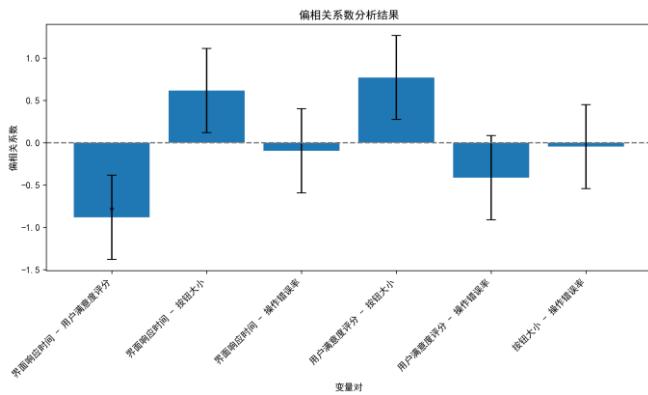
- 普通相关：直接计算两者的关联，可能发现“体重越重，跑步速度越慢”。但是，这里可能有干扰，比如“年龄”，年龄大的人体重可能更高，同时跑步速度本身也会随年龄下降（其实不是体重直接影响速度，而是年龄同时影响了两者）。
- 偏相关：排除“年龄”的干扰后，再看体重和速度的关系。可能发现：年龄相同的人里，体重和速度的关联很弱（甚至消失），这说明之前的负相关是被“年龄”干扰的结果，体重本身对速度的直接影响不大。

示例数据

| | A | B | C | D |
|---|--------|---------|------|-------|
| 1 | 界面响应时间 | 用户满意度评分 | 按钮大小 | 操作错误率 |
| 2 | 0.5 | 9 | 60 | 2 |
| 3 | 1.2 | 7 | 40 | 5 |
| 4 | 0.8 | 8 | 50 | 3 |
| 5 | 2 | 5 | 30 | 8 |
| 6 | 0.6 | 0 | 60 | 1 |

分析结果

| 变量组 | 偏相关系数 | P 值 | 自由度 |
|------------------|---------------------|-----------------------|-----|
| 界面响应时间 - 用户满意度评分 | 0.9561542086630456 | 0.0028415343812360815 | 2 |
| 界面响应时间 - 按钮大小 | 0.6951413356361906 | 0.12524160784036412 | 2 |
| 界面响应时间 - 操作错误率 | -0.5283302429121873 | 0.28124179768559604 | 2 |
| 用户满意度评分 - 按钮大小 | -0.6943650748294136 | 0.12584396829511318 | 2 |
| 用户满意度评分 - 操作错误率 | 0.5602845671729852 | 0.24751507767213607 | 2 |
| 按钮大小 - 操作错误率 | 0.31014813152709997 | 0.5496946660740781 | 2 |



结论

- 核心结论：在 4 个指标中，只有“界面响应时间”对“用户满意度”有直接、真实的影响（排除其他干扰后关联依然极强）；其他指标（按钮大小、操作错误率）和满意度 / 响应时间的相关，都是“被干扰的假关联”。
- 行动指导 1：优化用户满意度，优先抓“界面响应时间”，这是最直接的“杠杆因素”，哪怕按钮大小、错误率暂时不优化，只要把响应时间降下来（比如从 3 秒降到 1 秒），满意度一定会显著提升。
- 行动指导 2：不用过度纠结按钮大小、错误率，它们对满意度的影响，本质上是通过“响应时间”间接实现的（响应时间短的界面，自然按钮设计更合理、错误率更低），抓好响应时间，这些问题会跟着改善。

4.6 群体相关（典型相关分析）

典型相关分析的用途是，分析“两堆指标”（不是“两个指标”）整体上有没有关联、关联有多强，比如“语文、数学、英语成绩”（第一堆指标）和“每天睡眠时间、运动时间、玩手机时间”（第二堆指标）是不是有关系，还能找出这两堆指标里“最关键的核心指标”。

示例数据

智能手环的用户体验由“客观设计特征”和“主观感知评价”共同决定，现在研究两类特征之间的关联。

选取智能手环的研究对象，覆盖18-45岁，不同品牌和价格段的用户，发放问卷，问题仅包括“主观感知评价”类问题，“客观设计特征”的数据通过查询获得。

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|---------|---------|---------|---------|
| 1 | X1屏幕尺寸 | X2表带材质 | X3设备重量 | X4触控速度 | X5界面层级 | X6入口深度 | X7色彩数量 | X8图标一致 | X9信息密度 | Y1操作流畅度 | Y2学习成本 | Y3视觉吸引力 | Y4风格契合度 | Y5功能满足度 | Y6佩戴舒适度 |
| 2 | 1.69 | 2 | 28 | 120 | 2 | 2 | 2 | 4 | 28 | 4 | 4 | 5 | 4 | 4 | 3 |
| 3 | 1.28 | 1 | 22 | 150 | 1 | 1 | 1 | 3 | 22 | 3 | 3 | 2 | 2 | 3 | 4 |
| 4 | 1.47 | 2 | 35 | 200 | 2 | 2 | 2 | 2 | 34 | 2 | 2 | 2 | 2 | 2 | 2 |
| 5 | 1.1 | 1 | 20 | 100 | 1 | 1 | 1 | 4 | 18 | 5 | 5 | 3 | 3 | 5 | 5 |
| 6 | 1.55 | 3 | 42 | 250 | 3 | 3 | 3 | 1 | 42 | 2 | 2 | 1 | 1 | 1 | 2 |

分析过程

程序把所有的指标分为X组和Y组，在输入框中输入X组的所有表头信息，中间用英文的逗号连接，注意不能有空格。

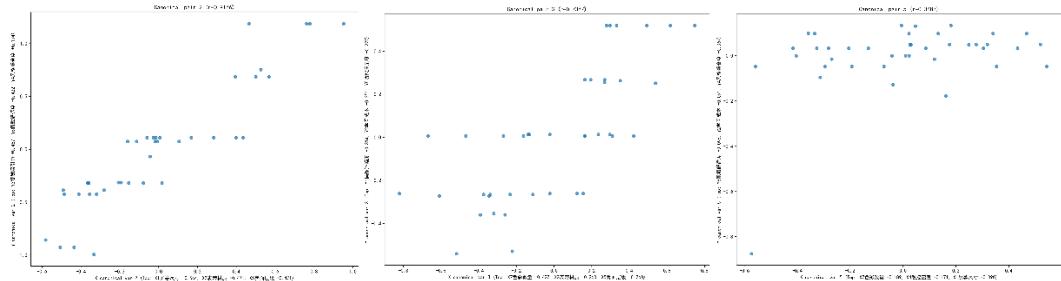
请输入X组变量的列名（以英文逗号分隔）：
例如：A,B,C
可选列：X1屏幕尺寸, X2表带材质, X3设备重量, X4触控速度, X5界面层级, X6入口深度, X7色彩数量, X8图标一致, X9信息密度, Y1操作流畅度, Y2学习成本, Y3视觉吸引力, Y4风格契合度, Y5功能满足度, Y6佩戴舒适度
未输入的列将自动作为Y组变量。
X1屏幕尺寸,X2表带材质,X3设备重量,X4触控速度,X5界面层级,X6入口深度,X7色彩数量,X8图标一致,X9信息密度

分析结果

内容过多，仅展示部分内容。

| 变量 | 加载中 | 正在加载 | 核心层 |
|---------|---------|--------|-----------|
| X4 触控速度 | -0.9717 | 0.9717 | 绝对核心 |
| X8 图标一致 | 0.9630 | 0.9630 | 绝对核心 |
| X3 设备重量 | -0.9171 | 0.9171 | 绝对核心 |
| X9 信息密度 | -0.8788 | 0.8788 | 绝对核心 |
| X7 色彩数量 | -0.8654 | 0.8654 | 绝对核心 |
| X2 表带材质 | -0.8240 | 0.8240 | 绝对核心 |
| X5 界面层级 | -0.8240 | 0.8240 | 绝对核心 |
| X6 入口深度 | -0.8240 | 0.8240 | 绝对核心 |
| X1 屏幕尺寸 | -0.3988 | 0.3988 | Secondary |

| 变量 | 加载中 | 正在加载 | 核心层 |
|----------|--------|--------|------|
| Y5 功能满足度 | 0.9637 | 0.9637 | 绝对核心 |
| Y4 风格契合度 | 0.9575 | 0.9575 | 绝对核心 |
| Y2 学习成本 | 0.9520 | 0.9520 | 绝对核心 |
| Y1 操作流畅度 | 0.9386 | 0.9386 | 绝对核心 |
| Y6 佩戴舒适度 | 0.8970 | 0.8970 | 绝对核心 |
| Y3 视觉吸引力 | 0.8834 | 0.8834 | 绝对核心 |



结论

- 优先抓“绝对核心设计”（改了必出效果）：触控速度 \leqslant 150 毫秒、图标风格统一（同色系 + 同形状）、设备重量 \leqslant 28 克、信息密度 25-30 字符 / 平方厘米、屏幕色彩 3-5 色、表带选硅胶、界面层级 \leqslant 2 层（常用功能放首页）；
- 次优化“次要设计”（辅助提升体验）：屏幕尺寸控制在 1.47-1.55 英寸（兼顾视觉和佩戴）；
- 别浪费精力在“无关设计”上：屏幕尺寸 1.47 与 1.55 英寸的微小差异、设备重量差 3-5 克、图标颜色深浅微调（用户几乎没感觉，优化成本高且无效）。

5 必然还是巧合（差异性分析）

5.1 卡方检验

卡方检验的用途是判断“两个分类变量之间的关系，是真的存在，还是纯粹巧合”。仅能分析两个变量，不支持多个变量。例如，想知道“用新教学方法”和“学生是否及格”有没有关系，新方法到底有没有用。

两个班级：

- 老方法班（50人）：及格 30人，不及格 20人。
- 新方法班（50人）：及格 40人，不及格 10人。

这时候你会疑惑：“新方法班的及格率高，会不会只是碰巧？”

卡方检验就会帮你算：

- 先假设“教学方法和及格率没关系”（理想情况）：那两个班的及格率应该差不多，理想及格人数都是 $(30+40) \div 2 = 35$ 人，不及格都是 15人；
- 再对比“实际情况”（老班 30人及格，新班 40人及格）和“理想情况”（都该 35人及格）的差异；
- 最后告诉你：“这个差异是不是巧合”，如果 p 值 < 0.05 ，就说明“新方法和高及格率不是巧合，新方法大概率真有用”。

注意!! DIAS 内置了四种卡方检验方法，并根据数据特征自动选择：“Pearson 卡方、Yates 校正卡方、似然比卡方和 Fisher 卡方”。

示例数据

| 性别 | 吸烟 | 不吸烟 |
|----|----|-----|
| 男性 | 40 | 60 |
| 女性 | 10 | 90 |

分析结果

| 检验类型 | 样本量 | 卡方值 | 自由度 | p 值 | 显著性 ($\alpha = 0.05$) | Phi 系数 |
|------------|-----|---------|-----|----------|-------------------------|--------|
| Pearson 卡方 | 200 | 24.0000 | 1 | 0.000001 | 显著 | 0.3464 |

| Row Labels | 吸烟 | 不吸烟 |
|------------|-------|-------|
| 男性 | 25.00 | 75.00 |

用于在 $2*2$ 列联表中，当 $n >= 40$ 且所有预期频率 $E >= 5$ 时，测量观测频率与预期频率之间的差异。

卡方值：卡方值越大，表示观测频率与预期频率之间的差异越大。

p 值：当 p 值小于显著性水平（通常为 0.05）时，拒绝零假设，表明变量之间存在显著关联；否则，接受零假设，表明变量之间不具有显著关联。

自由度：自由度反映数据的独立变异程度，用于计算卡方分布的临界值。

显著性 ($\alpha = 0.05$)：表示在 0.05 显著性水平下，变量之间是否存在显著关联。

结论

核心结论：性别和吸烟有显著关联 ($p=0.000001 < 0.05$)，男性吸烟比例 (40%) 远高于女性 (10%)。

差异程度：实际吸烟人数和“无关联”的理论人数差距很大（卡方值 = 24），且关联强度为中等 ($\Phi=0.3464$)。

检验前提合格：样本量 $200 \geq 40$ ，所有期望人数 ≥ 5 (男性吸烟期望 25 人)，符合 Pearson 卡方检验的要求，结果可靠。

示例数据

| 年龄段 | 品牌 A | 品牌 B | 品牌 C | 品牌 D |
|---------|------|------|------|------|
| 18-30 岁 | 15 | 10 | 3 | 2 |
| 31-50 岁 | 12 | 8 | 4 | 1 |

分析结果

| 检验类型 | 样本量 | 卡方值 | 自由度 | p 值 | 显著性 ($\alpha=0.05$) | Cramer's V |
|-------|-----|--------|-----|----------|-----------------------|------------|
| 似然比卡方 | 55 | 0.5847 | 3 | 0.899931 | 不显著 | 0.1031 |

当 R*C 列联表中的预期频率不满足使用 Pearson 卡方检验的条件时使用。

卡方值：卡方值越大，表示观测频率与预期频率之间的差异越大。

p 值：当 p 值小于显著性水平（通常为 0.05）时，拒绝零假设，表明变量之间存在显著关联；否则，接受零假设，表明无显著关联。

自由度：自由度反映数据的独立变异程度，用于计算卡方分布的临界值。

显著性 ($\alpha=0.05$)：表示在 0.05 显著性水平下，变量之间是否存在显著关联。

结论

- 核心结论：年龄段 (18-30 岁 / 31-50 岁) 和品牌偏好 (A/B/C/D) 无显著关联 ($p \approx 0.9 > 0.05$)，两个年龄段对四个品牌的喜欢程度差不多。
- 差异本质：实际数据中“18-30 岁选 A 多一点、选 C 少一点”是偶然差异，不是真实的年龄偏好差异。
- 行动指导：如果做品牌营销，不用按“18-30 岁”“31-50 岁”分年龄段设计策略，两个群体对品牌的偏好一致，用统一的营销方案（比如，主推品牌 A，因为两个年龄段选 A 的人都最多）即可。

5.2 平均水平不是平均数 (t 检验)

DIAS 中有三种不同的 t 检验工具，“单样本 t 检验，独立样本 t 检验和配对 t 检验”。它们的计算逻辑和结论的可靠性建立在“数据近似符合正态分布”的基础上，因此，必须先对数据进行正态分布检验。数据中不能有极端值。用于独立样本 t 检验的数据还要进行方差齐性检验。

单样本 t 检验

检验“一组数据的总体均值”是否等于某个预设的“理论值 / 标准值”。这个标准值的设定需要“有依据”，比如行业规范、历史数据、目标要求等。

- 例 1：检验某班学生的数学平均分（均值 85 分）是否等于全国高考数学平均分（标准值 80 分）。
- 例 2：检验某工厂生产的零件直径（均值 5.1mm）是否符合设计标准（理论值 5mm）。

独立样本 t 检验

检验“两组相互独立的数据”的总体均值是否存在显著差异（两组数据来自不同群体，无配对关系）。

配对关系：同一组对象在不同条件下（如时间前后、不同处理方法等）的两次测量结果。它关注的是每个对象内部的变化差异。

比如，我们有 10 个男生的身高、10 个女生的身高，数量一样，但这 20 个人没有任何“搭档关系”，就不算配对；我们有 10 个人，每个人的“早上身高”和“晚上身高”，虽然也是 10 对、数量一样，但因为是同一个人的前后数据，就是配对。

- 例 1：检验男生的体重均值（65kg）与女生的体重均值（55kg）是否有本质差异。
- 例 2：检验“新药组”患者的血压降低值均值（15mmHg）与“安慰剂组”（5mmHg）是否有显著差异（两组患者无关联）。

配对 t 检验

检验“两组有配对关系的数据”的总体均值差异是否显著（两组数据来自同一群体或配对群体，存在配对关系）。

- 例 1：检验同一批患者“服药前的血压”与“服药后的血压”的均值差异（每个患者有两个数据，形成配对）。
- 例 2：检验“双胞胎 A 的智商得分”与“双胞胎 B 的智商得分”的均值差异（配对群体的对应关系）。

T 检验 和 平均数 的区别

平均数只能告诉我们“两组数据的平均水平不一样”，但 t 检验能告诉我们“这个‘不一样’是真的有差异，还是纯粹因为随机波动导致的”。

| 对比维度 | 平均数（描述性统计） | t 检验（推断性统计） |
|------------|-----------------------------|---|
| 核心目的 | 描述数据的“平均水平”，告诉我们“数据长什么样” | 推断“样本差异”是否能代表“总体差异”，回答“差异是不是真的” |
| 是否考虑“随机误差” | 完全不考虑。只看表面数值差异，忽略波动和样本量 | 必须考虑。会结合“数据波动（标准差）”和“样本量”，判断差异是否超过随机误差的范围 |
| 输出结果与结论 | 只给一个“平均数”或“两组平均数的差值”，无结论性判断 | 输出“t 值”“p 值”等统计量，能明确判断“差异是否显著” |
| 适用场景 | 快速了解数据的基本情况 | 需要从样本数据推断总体结论（如“新方法是否真的比旧方法有效”“某班成绩是否真的高于全国标准”） |

假设比较“新减肥方法”和“旧减肥方法”的效果，找了 10 个人用新方法（平均减重 5.2 斤），10 个人用旧方法（平均减重 4.8 斤）。

- 平均数：只能得出“新方法平均多减 0.4 斤”，但我们不知道这 0.4 斤的差异是“新方法真的更好”，还是“碰巧这 10 个用新方法的人本身代谢就好一点”；
- t 检验：数据的波动程度（如每个人减重的差异）、样本量等，计算出“这个 0.4 斤的差异由随机误差导致的概率（p 值）”，如果 $p < 0.05$ ，就说明“0.4 斤的差异大概率不是巧合，新方法真的有效果”；如果 $p > 0.05$ ，就说明“0.4 斤的差异可能是随机的，不能证明新方法更好”。

三种方法的异同点

| 对比维度 | 单样本 t 检验 | 独立样本 t 检验 | 配对 t 检验 |
|-----------|----------|--------------|------------------|
| 样本数量 / 结构 | 1 组数据 | 2 组独立数据 | 2 组具有配对关系的数据 |
| 数据关联性 | 无 | 独立（无关联） | 配对（一一对应，如同一对象前后） |
| 正态分布要求 | 需近似正态分布 | 两组数据均需近似正态分布 | 两组数据的“差值”需近似正态分布 |
| 额外前提条件 | 无 | 需满足方差齐性 | 无 |

单样本 t 检验示例数据

数据要求近似正态分布，分析前需要做正太分布检验。

某设计团队为一款购物 APP 设计了新版结算页面，团队设定的设计标准是：用户完成结算操作的平均时间应 ≤ 25 秒（基于行业调研，25 秒是用户可接受的阈值，超过会显著降低购买意愿）。为验证新版设计是否达标，团队招募了 20 名用户进行测试，记录每位用户完成结算的时间（秒）。

| A | 结算时间（秒） |
|---|---------|
| 1 | 28 |
| 2 | 25 |
| 3 | 30 |
| 4 | 27 |
| 5 | 22 |

分析过程

这里需要输入“理论值 / 标准值”。

| |
|---------|
| 总体均值: 0 |
|---------|

分析结果

| t 统计量 | 自由度 | p 值 | 置信区间 | 样本均值 | 样本标准差 |
|-------------------|-----|----------------------------|--|------|-------------------|
| 4.882615196099746 | 19 | 0.0001033850 9212290358 | (1.542594650625655), (3.8574053493743437) | 27.7 | 2.473012226755907 |

- 解释说明

单样本 t 检验：用于检验样本均值与已知总体均值之间是否存在显著差异。

t 统计量：衡量样本均值与总体均值之间差异的统计量。

自由度：在统计计算中可以自由变动的数值的个数。

p 值：用于判断是否拒绝原假设的概率值。

均值差异的置信区间：包含总体均值差异的一个区间范围。

样本均值：样本数据的平均值。

样本标准差：样本数据的离散程度度量。

效应量 (Cohen's d)：衡量差异实际意义的指标，不受样本量影响。

- 结果解读

t 统计量：t 统计量的绝对值越大，说明样本均值与总体均值之间的差异越显著。

p 值：当 p 值小于显著性水平（通常为 0.05）时，拒绝原假设，认为样本均值与总体均值存在显著差异；否则，接受原假设。

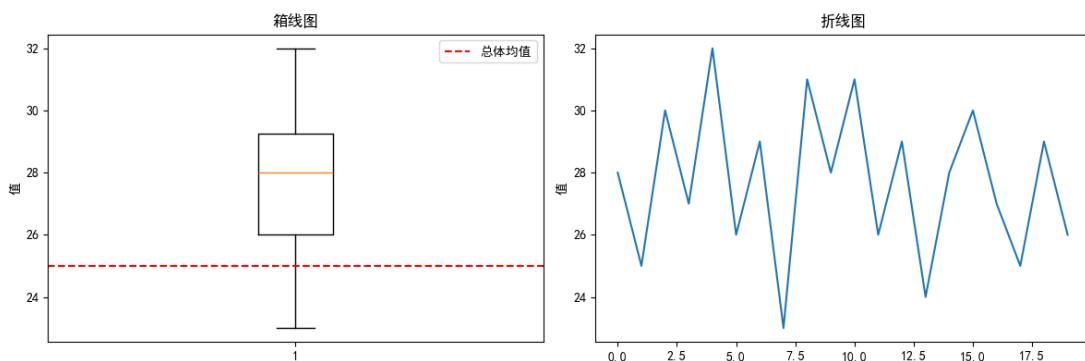
自由度：自由度影响 t 分布的形状，自由度越大，t 分布越接近正态分布。

均值差异的置信区间：如果置信区间不包含 0，则说明样本均值与总体均值存在显著差异。

样本均值：反映了样本数据的平均水平。

样本标准差：反映了样本数据的离散程度。

效应量 (Cohen's d)： $|d| < 0.2$ 为微小效应， $0.2 \leq |d| < 0.5$ 为小效应， $0.5 \leq |d| < 0.8$ 为中等效应， $|d| \geq 0.8$ 为大效应。



结论

- 结算速度严重不达标：p 值 ≈ 0 , t 统计量 = 50.09, 样本平均 27.7 秒远高于标准值，差异极其显著，必须紧急优化结算速度；
- 问题是普遍的：样本标准差仅 2.47 秒，数据波动小，说明不是“偶尔慢”，而是“每次都慢”，是系统性问题；
- 优化目标明确：真实结算平均时间在 26.54~28.86 秒之间，要达到标准（比如 10 秒），需要把结算速度至少提升 16 秒以上 ($26.54 - 10 = 16.54$ 秒)。
- 立刻排查结算流程的瓶颈（比如服务器响应、代码效率），结算速度比标准慢太多，已经是严重影响体验的问题了。

独立样本 t 检验 示例数据

数据要求方差齐，因此需要先进行方差齐性检验。

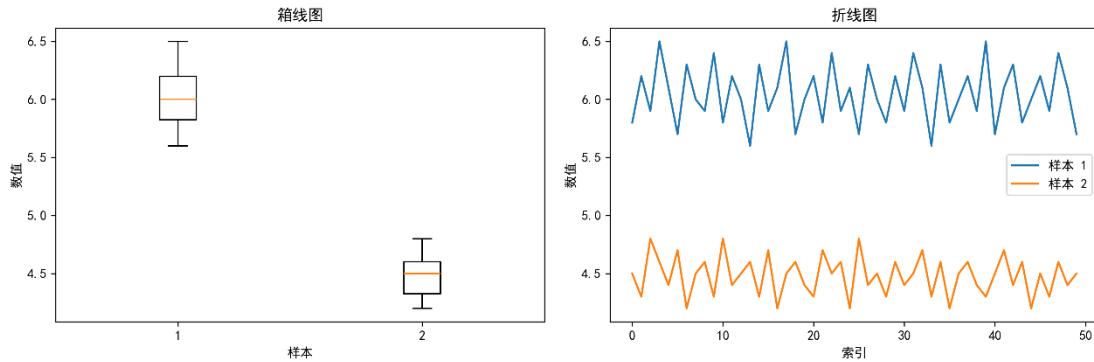
两种界面设计方案（极简风格 vs 拟物风格）的用户体验评分场景。随机招募 60 名用户（无购物 APP 使用偏好差异），随机分为两组（每组 30 人），每组用户分别使用对应风格的界面完成相同任务（如查找商品、加入购物车），任务后填写“用户体验评分量

表”（1-7 分，1 = 极差，7 = 极佳）。

| | A | B |
|---|-------------------|-------------------|
| 1 | 极简风格综合体验评分（1-7 分） | 拟物风格综合体验评分（1-7 分） |
| 2 | | 5.8 |
| 3 | | 6.2 |
| 4 | | 5.9 |
| 5 | | 6.5 |
| 6 | | 6.1 |

分析结果

| t 统计量 | 自由度 | p 值 | 置信区间 | 样本量 | 均值 | 标准差 |
|------------------------|-----|-------------------------------|-------------------------------|---|---|---|
| 24.1672494 10223423 | 58 | 5.72790 7623254 726e-32 | (1.4391165 710040168) , | {'极简风格综合 体验评分（1-7 分）': 30, '拟 物风格综合体验 评分（1-7 分）': 289959843) | {'极简风格综合体验 评分（1-7 分）': 6.066666666666666, '拟物风格综合体验评 分（1-7 分）': 4.466666666666666} | {'极简风格综合体验评分 (1-7 分)': 0.2564120247507547, '拟 物风格综合体验评分（1-7 分）': 0.2564120247507547} |



结论

- 极简风格更优：在购物 APP 商品详情页设计中，极简风格的用户体验评分（6.07）显著高于拟物风格（4.47）， p 值 ≈ 0 ，差异无可争议。
- 优势可量化：极简风格比拟物风格平均高 1.44~1.76 分（1-7 分量表），这个优势在用户体验中属于“大差距”（比如从“中等体验”到“优秀体验”的跨越）。
- 设计决策明确：后续迭代应优先选择极简风格，因为数据证明，它更符合用户对“易用性、美观度、满意度”的需求，不是设计师的主观偏好，而是有统计依据的结论。

配对 t 检验示例数据

注意，数据需要先进行“差值”的正太分布检验，也就是用第一列数据减第二列数据，得到一列新的数据，检验新的数据是否通过正太分布检验。分析时仍使用两列数据。

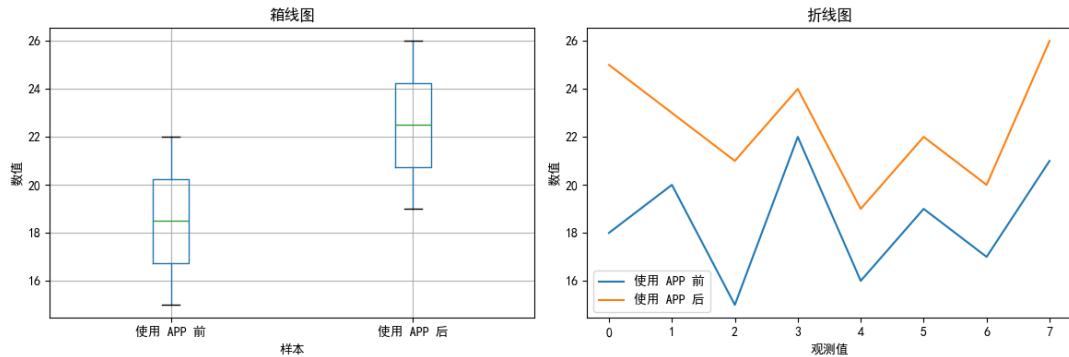
随机抽取 15 名当前初级班学员，记录 1 天内实际背会的单词数量，分析是否达到教学目标。过去 3 年初级班学员的平均“日背词量”稳定在 25 个，且这个水平能保证学员后续考试通过率达 85%，属于“经过验证的有效标准”。

| 使用 APP 前 | 使用 APP 后 |
|----------|----------|
| 18 | 25 |

| | |
|----|----|
| 20 | 23 |
| 15 | 21 |
| 22 | 24 |
| 16 | 19 |
| 19 | 22 |
| 17 | 20 |
| 21 | 26 |

分析结果

| 统计量 | t 统计量 | 自由度 | p 值 | 置信区间 |
|---------|--------------------------------------|-----|----------|--------------------|
| 配对 t 检验 | -6.381792281739726 | 7 | 0.000374 | (-5.4821, -2.5179) |
| 样本量 | {'使用 APP 前': 8, '使用 APP 后': 8} | | | |
| 均值 | {'使用 APP 前': 18.5, '使用 APP 后': 22.5} | | | |



结论

- APP 真的有效: p 值 = 0.00037 < 0.05, 使用 APP 后指标显著提高 (从 18.5 涨到 22.5), 不是偶然。
- 效果有明确范围: 使用 APP 后, 指标比使用前高 2.52~5.48 个单位, 效果稳定。
- 样本可靠: 8 个样本的结果符合统计要求, 结论可信 (如果想推广到更多人, 可后续增加样本量)。

5.3 方差分析

DIAS 中的方差分析有四种不同的工具, “单样本方差, 双样本方差, 多样本方差和多元方差”。“方差分析”和“t 检验”的用途是一样的。

单样本方差

检验“一组样本的方差”是否和“某个已知的标准值”有显著差异(本质是单样本 t 检验的扩展)。

例: 某工厂生产的零件标准长度是 10cm, 随机测了 30 个零件的长度, 想知道这批零件的平均长度是否真的等于 10cm (是否达标)。

多样本方差

检验“三组及以上样本的均值”是否有显著差异。

例：比较“节食”“运动”“节食+运动”三种减肥方法的效果，每组30人，看三种方法的平均减重是否有显著差异（比如是否有一种方法明显更好）。

多元方差

同时检验“三组及以上样本”在“多个指标”上的均值是否有显著差异。

例：比较“公立学校”“私立学校”“国际学校”的学生，在“数学成绩”“英语成绩”“创造力评分”三个指标上的均值是否有整体差异（不只是单一成绩，而是综合能力）。

方差分析和t检验的异同点

| 对比维度 | t 检验 | 方差分析 |
|------|-----------------------|--------------------------------------|
| 分组数量 | 1组或2组（独立样本/配对样本） | 1组或3组及以上（独立样本） |
| 数据前提 | 数据需满足“正态分布”，“方差齐” | 数据需满足“正态分布”，“方差齐”和“无配对关系” |
| 检验核心 | 直接计算“两组均值差的t值” | 先算“组间差异”和“组内差异”的F值，再判断整体是否有差异 |
| 结果解读 | 直接看“两组是否有差异”（如A比B高/低） | 先看“所有组是否有差异”，若有差异，需做“事后检验”，找具体是哪两组不同 |

怎么选？

若要比较“2组数据”（如A vs B、前 vs 后）：用t检验；

若要比较“3组及以上数据”（如A vs B vs C）：用方差分析；

若要同时比较“多组+多个指标”：用多元方差，t检验做不到。

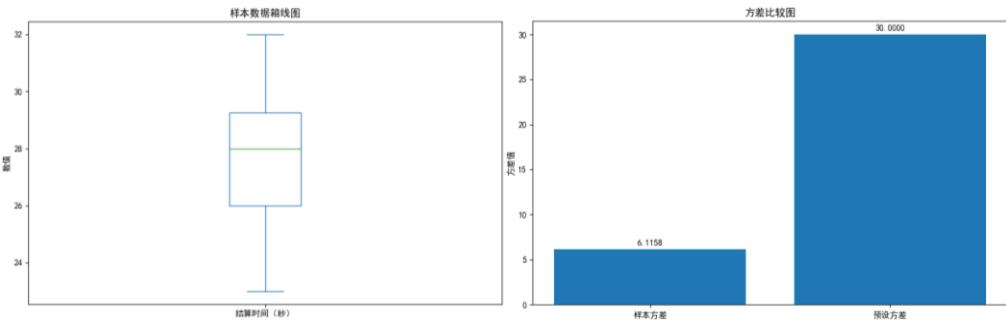
单样本方差示例数据

标准值30

| A | |
|---|---------|
| 1 | 结算时间(秒) |
| 2 | 28 |
| 3 | 25 |
| 4 | 30 |
| 5 | 27 |
| 6 | 22 |

分析结果

| 样本量 | 样本方差 | 样本标准差 | 预设总体方差 | 与预设值的差异 | 卡方统计量 | 自由度 | p值 | 效应量（方差比） |
|-----|--------|--------|---------|----------|--------|-----|----------|----------|
| 20 | 6.1158 | 2.4730 | 30.0000 | -23.8842 | 3.8733 | 19 | 0.000166 | 0.2039 |

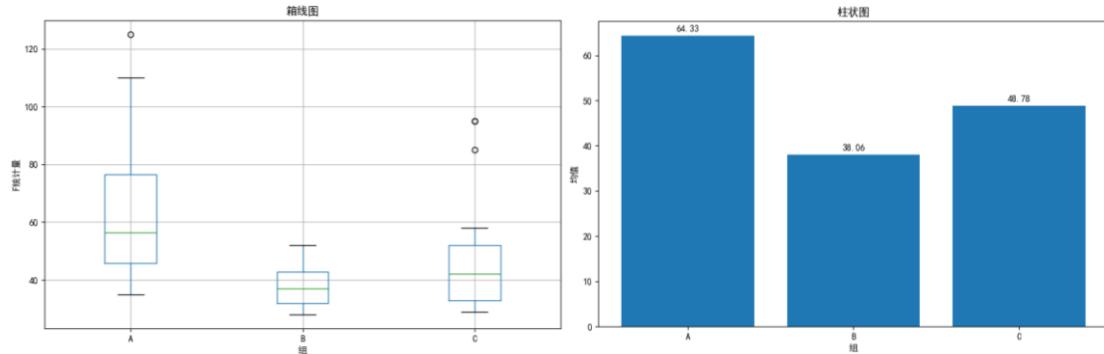


多样本方差 示例数据

| | A | B | C |
|---|-----|----|----|
| 1 | A | B | C |
| 2 | 45 | 28 | 52 |
| 3 | 62 | 35 | 95 |
| 4 | 38 | 30 | 35 |
| 5 | 125 | 42 | 52 |
| 6 | 55 | 32 | 95 |

分析结果

| F统计量 | 组间自由度 | 组内自由度 | p值 | 效应量(Eta 平方) | 样本量 | 均值 |
|--------|-------|-------|-------------|-------------|-----------------------------|------------------------------------|
| 7.9558 | 2 | 51 | 0.000983*** | 0.2378 | {'A': 18, 'B': 18, 'C': 18} | A: 64.3333, B: 38.0556, C: 48.7778 |



多元方差 示例数据

注意，数据需包含两类数据，第一类是分类数据（如 A, B, C / 1, 0 / 男, 女），第二类是连续数据，连续数据至少有两列。

| | A | B | C | D | E |
|---|------|-------|------|-------|-------|
| 1 | 设计风格 | 易用性评分 | 美学评分 | 创新性评分 | 用户满意度 |
| 2 | 极简主义 | | 4.8 | 4.5 | 3.2 |
| 3 | 极简主义 | | 4.6 | 4.7 | 3 |
| 4 | 极简主义 | | 4.7 | 4.6 | 3.1 |
| 5 | 极简主义 | | 4.9 | 4.4 | 2.9 |
| 6 | 极简主义 | | 4.5 | 4.8 | 3.3 |

分析过程

输入分类数据的列名：设计风格



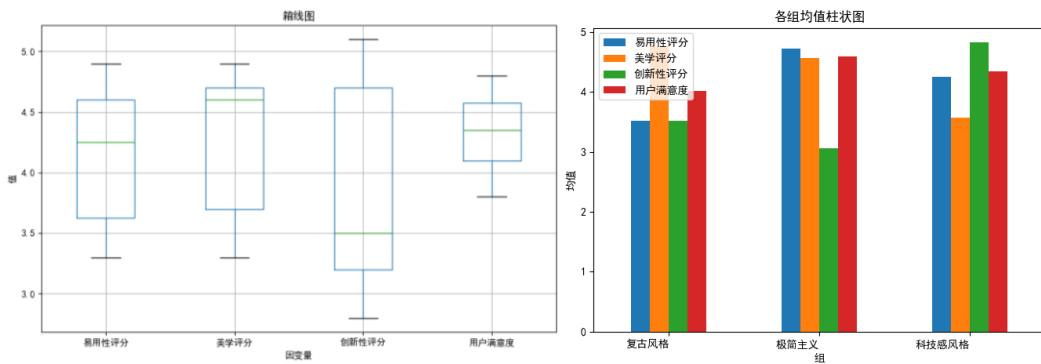
输入连续数据的列名：易用性评分，美学评分，创新性评分，用户满意度



分析结果

| F 统计量 | 组间自由度 | 组内自由度 | p 值 | Pillai's Trace | 偏 Eta 平方 | Cohen's f 效应量 |
|-----------------------|-------|-------|---------------------------|------------------------|------------------------|----------------------|
| 469.564885 9006931 | 4.0 | 25.0 | 4.032220193526 214e-23 | 0.986864639620 9966 | 0.986864639620 9966 | 8.6677783626550 4 |

| 样本量 | 均值 |
|---------------------------------------|---|
| {'复古风格': 10, '极简主义': 10, '科技感风格': 10} | {'易用性评分': {'复古风格': 3.5200000000000005, '极简主义': 4.720000000000001, '科技感风格': 4.25}, '美学评分': {'复古风格': 4.770000000000005, '极简主义': 4.57, '科技感风格': 3.570000000000003}, '创新性评分': {'复古风格': 3.520000000000005, '极简主义': 3.06, '科技感风格': 4.83}, '用户满意度': {'复古风格': 4.020000000000005, '极简主义': 4.6, '科技感风格': 4.35}} |



结论

- 三种风格综合差异显著， p 值 ≈ 0 , Eta 平方 = 0.95, 说明设计风格对用户体验影响极强，选对风格比细节优化更重要；
- 风格选择要匹配需求，若优先“易用性、整体满意度”：选极简主义（易用性 4.72, 满意度 4.60, 都是最高）；若优先“美观度、复古氛围”：选复古风格（美学 4.77,

- 行业顶尖水平); 若优先“创新性、科技感”: 选科技感风格(创新性 4.83, 远超其他两种);
- 短板优化方向明确, 极简风格: 提升创新性(当前 3.06, 可增加新交互设计); 科技感风格: 提升美学评分(当前 3.57, 可优化色彩和视觉细节); 复古风格: 提升易用性(当前 3.52, 可简化操作步骤)。

5.4 中位数比较 (Wilcoxon 检验)

DIAS 中有两种 Wilcoxon 检验工具, “单样本 Wilcoxon 检验和配对样本 Wilcoxon 检验”。它们和“t 检验”的功能相似。区别在于,

- t 检验比较均数, 且要求数据近似正太分布, 不能有极端值
- Wilcoxon 检验比较中位数, 不要求数据近似正太分布, 且不受极端值影响。

单样本 Wilcoxon 检验

检验“一组有序分类数据的总体中位数”是否等于某个预设的“理论值 / 标准值 / 目标值”。

- 例 1: 检验某款家具的“用户舒适度评分(1-10 分, 非正态分布)”的中位数是否达到行业推荐标准(如中位数 ≥ 8 分);
- 例 2: 检验某类患者的“术后恢复天数(偏态分布, 有极端值)”的中位数是否等于临床预期(如中位数 ≤ 7 天);
- 例 3: 检验某批零件的“误差值(分布未知)”的中位数是否等于设计允许的误差(如中位数 $\leq 0.1\text{mm}$)。

配对样本 Wilcoxon 检验

检验“两组配对有序分类数据的总体中位数差异”是否显著。

- 例 1: 检验同一批学生“课前测试成绩”与“课后测试成绩”的中位数是否有差异(成绩偏态, 且前后配对);
- 例 2: 检验同一组消费者“使用产品前的满意度”与“使用后的满意度”的中位数是否有提升(满意度为有序分类数据, 如 1-5 星);
- 例 3: 检验同一块土地“施用肥料 A 后的产量”与“施用肥料 B 后的产量”的中位数是否有差异(产量受地块差异影响, 需配对控制)。

单样本 Wilcoxon 检验 示例数据

某家电品牌设计了一款新型扫地机器人, 行业对“扫地机器人用户满意度”的推荐标准为“中位数 ≥ 4 分”(评分尺度: 1-5 分, 1 = 非常不满意, 5 = 非常满意)。为验证该产品是否达标, 团队招募 25 名用户试用后打分, 评分数据存在偏态(多数用户打 4-5 分, 少数打 1-2 分, 不符合正态分布), 需用单样本 Wilcoxon 检验判断“总体满意度中位数是否 ≥ 4 分”。

| A | |
|---|---------------|
| 1 | 满意度评分 (1-5 分) |
| 2 | 4 |
| 3 | 5 |
| 4 | 4 |
| 5 | 2 |
| n | |

分析过程

这里输入“理论值 / 标准值 / 目标值”，数据的理论值为 4

预设值：请输入假设的中位数（默认为0）

分析结果

| 检验统计量(W) | 样本量 | 有效样本量 | 中位数 | 预设值 | 与预设值的中位数差异 | 效应量 r | p 值 | 中位数差异的 95% 置信区间 |
|----------|-----|-------|--------|--------|------------|--------|----------|------------------|
| 46.5000 | 25 | 25 | 4.0000 | 4.0000 | 0.0000 | 9.3000 | 0.684470 | [0.0000, 1.0000] |

检验统计量 (W): W 值越大，说明样本与假设中位数的差异越明显。

p 值：用于判断样本与假设中位数之间是否存在显著差异的依据。

样本量：样本量的大小会影响统计检验的功效，较大的样本量通常能提供更准确的结果。

有效样本量：实际参与分析的样本数量，排除了缺失值。

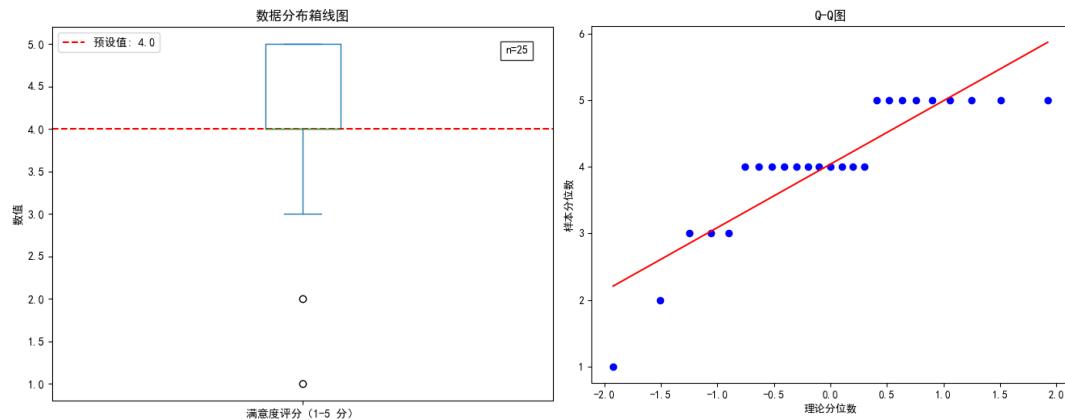
中位数：中位数反映了数据的中心位置，可用于比较样本与假设中位数的差异。

中位数差异的 95% 置信区间：如果置信区间不包含 0，说明样本与假设中位数存在显著差异。

预设值：作为比较基准的假设值，是统计检验的参考点。

与预设值的中位数差异：正值表示样本中位数大于预设值，负值表示样本中位数小于预设值，绝对值越大差异越明显。

效应量 r：通常 0.1 为小效应，0.3 为中等效应，0.5 为大效应。



结论

- 满意度刚好达标：p 值 = 0.684 > 0.05，样本中位数 = 4.0 (和目标一致)，说明 25 个用户的满意度达到了团队设定的目标，不是偶然。
- 数据很稳定：与预设值差异为 0，置信区间包含 0，说明没有“隐性不达标”（比

如看似中位数 4.0，但多数用户评 3 分），是真的符合目标。

- 后续行动建议：暂时不用调整产品，当前满意度达标，若想进一步提升，可扩大样本量（比如调查 50 人），看是否能稳定在 4.0 以上。

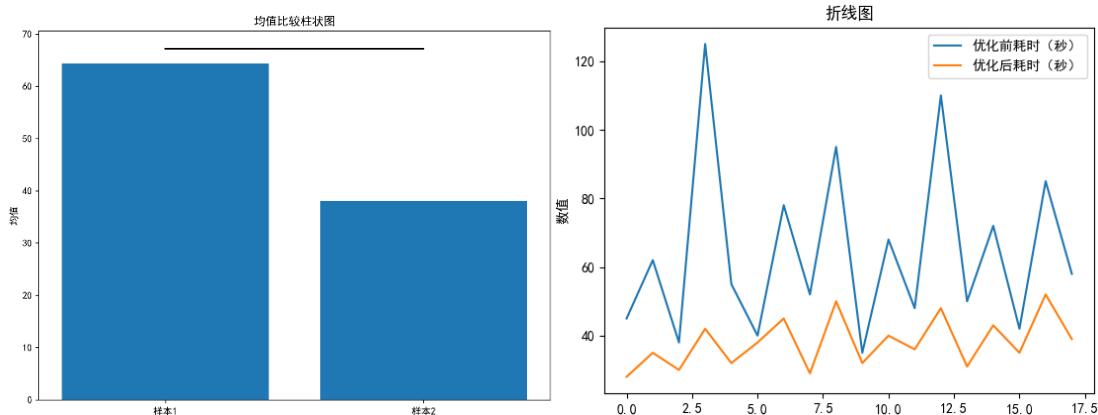
配对样本 Wilcoxon 检验示例数据

某 APP 设计团队对“首页推荐算法”进行优化，为验证优化效果，招募 18 名用户，分别记录“优化前使用 APP 时的‘找到目标商品耗时’”和“优化后相同任务的耗时”（单位：秒）。数据存在极端值（部分用户在优化前因为找不到商品导致耗时远超均值，差值非正态），需用配对样本 Wilcoxon 检验判断“优化后耗时是否显著降低”。

| | A | B |
|-----|-----------|-----------|
| 1 | 优化前耗时 (秒) | 优化后耗时 (秒) |
| 2 | 45 | 28 |
| 3 | 62 | 35 |
| 4 | 38 | 30 |
| 5 | 125 | 42 |
| ... | ... | ... |

分析结果

| 样本量 | 中位数 | W 统计量 | 显著性 | Z 统计量 | 效应量 (η) | p 值 | 均值差异的置信区间 |
|------|------|-------|-------------|-------|---|----------------------------|-------------------------------|
| 18.0 | 23.0 | 0.0 | $p < 0.001$ | - | 0.877650426007001 3.7235554064448757 | 0.000195407937 40897315 | (15.0), (38.88888888888886 |



结论

- 优化效果极显著： p 值 = 0.000195 < 0.05，优化后，用户耗时显著减少，方案有效。
- 效果可量化：优化后平均减少 23 秒，且在 95% 可信度下，减少范围是 16~36.5 秒，效果稳定且有明确边界。
- 用户反馈一致：18 个用户中几乎没人觉得优化后变慢，优化体验无争议。

5.5 Friedman 检验

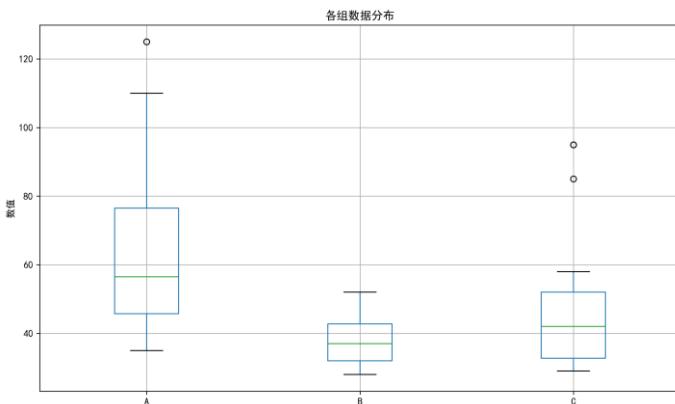
它的用途和“t 检验”，“方差分析”，“Wilcoxon 检验”相似，如果数据有三组及以上，且不满足正太分布时，可以用 Friedman 检验。

示例数据

| | A | B | C |
|---|-----|----|----|
| 1 | A | B | C |
| 2 | 45 | 28 | 52 |
| 3 | 62 | 35 | 95 |
| 4 | 38 | 30 | 35 |
| 5 | 125 | 42 | 52 |
| n | 22 | 22 | 22 |

分析结果

| 统计指标 | 数值 |
|-------------------|-----------------------------------|
| Friedman 检验统计量 | 17.0725 |
| p 值 | 0.0002 |
| 效应量 (Kendall's W) | 0.4742 |
| 效应解释 | 中等效应 |
| 样本量 | {'A': 18, 'B': 18, 'C': 18} |
| 中位数 | {'A': 56.5, 'B': 37.0, 'C': 42.0} |



结论

A、B、C 三组数据不仅中位数有明显差距 (A>C>B)，而且这种差距是统计上显著的 (不是偶然的)，A 组的整体表现显著优于 B 组和 C 组，B 组的表现最差。

5.6 预期和实际（卡方拟合优度检验）

它的用途是对比两组数据 (一组是实际观测到的结果，另一组是预期的理论结果)，看看它们是不是“差不多”，还是“差得很远”。

例如，一名包装设计师想测试自己设计的 5 种零食包装颜色 (红、黄、蓝、绿、紫) 是

否符合他预期的“吸引力分布”。

理论预期：根据前期调研，他认为这5种颜色的吸引力比例应该是：红30%、黄25%、蓝20%、绿15%、紫10%。如果总共有200个测试者，理论上每种颜色的喜欢人数应该是：60、50、40、30、20。

实际观测：让200个测试者投票选出最喜欢的颜色，实际得到的喜欢人数是：红70、黄45、蓝35、绿30、紫20。

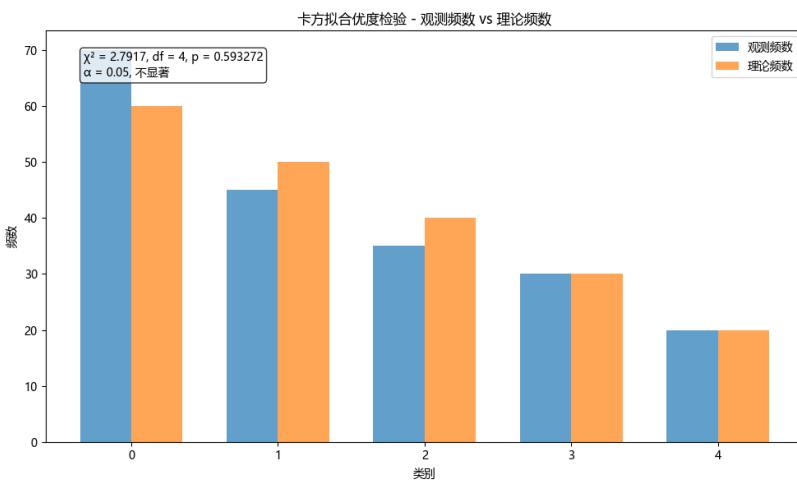
示例数据

程序默认第一列是实际数据，第二列是理论数据。

| 实际数据 | 理论数据 |
|------|------|
| 70 | 60 |
| 45 | 50 |
| 35 | 40 |
| 30 | 30 |
| 20 | 20 |

分析结果

| 统计量 | 值 |
|-------|----------|
| 卡方值 | 2.7917 |
| 自由度 | 4 |
| p 值 | 0.593272 |
| 显著性水平 | 0.05 |
| 结果 | 不显著 |



结论

p 值 = 0.593 > 0.05，预设的“理论分布猜想”是成立的，包装设计师以为的“数据分布”，真的和现实一致！

5.7 干扰控制（协方差分析）

它的用途是当我们研究 X 对于 Y 的影响时，考虑干扰项 z 是否有影响。X 必须是分类数据，Y 和 z 必须是连续数据。

例如，研究不同的 APP 界面风格（简约风 A/ 复古风 B/ 科技风 C）对用户“操作时间”的影响。“界面风格”是“X”，“操作时间”（比如完成任务的时间）是“Y”，用户的“年龄”，“手机使用经验年限”等因素可能会干扰结果，它们就是“Z”。

注意!! 排除“干扰因素”和手动删除“干扰因素”是不同的。手动删除会直接毁掉数据的完整性和准确性，而协方差分析是“控制干扰”，既保留数据又能排除干扰。

协方差分析和调节作用的区别

这两种方法的用途非常相似，区别在于，

- 调节作用：假设研究“广告投入（X，连续变量：万元）”对“产品销量（Y，连续变量：件）”的影响，同时怀疑“品牌知名度（M，分类变量：高 / 中 / 低）”会改变这种影响。
- 协方差分析：假设比较“两种减肥方法（X，分类变量：方法 A / 方法 B）”对“体重下降量（Y，连续变量：kg）”的影响，但发现“初始体重（C，连续变量：kg）”会干扰结果。

示例数据

第一列为“分组变量”（要比较的不同组别，比如“界面风格”，“设计方案”等）。

最后一列为“因变量”（结果指标，比如“操作时间”，“满意度评分”等）。

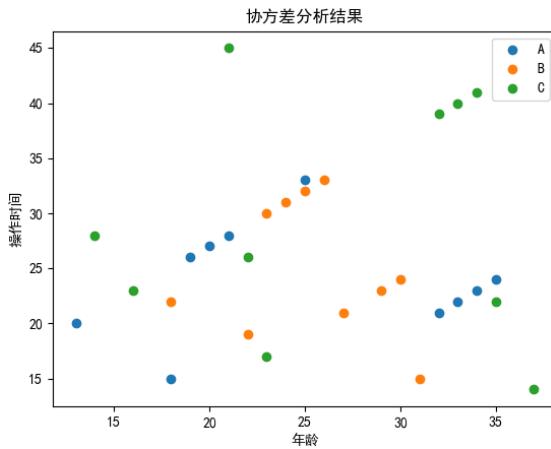
中间所有列为“协变量”（干扰因素，比如“年龄”，“使用经验”等），可以有多列。

数据需满足正太分布和方差齐。

| | A | B | C | D |
|---|------|----|------|------|
| 1 | 界面风格 | 年龄 | 使用经验 | 操作时间 |
| 2 | A | 25 | 1 | 33 |
| 3 | A | 13 | 0 | 20 |
| 4 | A | 32 | 2 | 21 |
| 5 | A | 33 | 3 | 22 |
| 6 | A | 31 | 2 | 23 |

分析结果

| 来源 | SS | DF | F 值 | 未校正 p 值 | 偏 eta 平方 |
|----------|-----------|----|--------|---------|----------|
| 界面风格 | 148.8647 | 2 | 1.1696 | 0.3269 | 0.0856 |
| 年龄 | 0.0549 | 1 | 0.0009 | 0.9768 | 0.0000 |
| 使用经验 | 52.2774 | 1 | 0.8215 | 0.3734 | 0.0318 |
| Residual | 1590.9873 | 25 | nan | nan | nan |



结论

在排除了“年龄”和“使用经验”的干扰后，3 种界面风格（A/B/C）在用户操作时间上没有显著差异，选择哪种风格都不会影响用户的操作效率。

5.8 事后多重比较

它是“多样本方差的‘后续补充’”：当多样本方差发现“多组数据整体有显著差异”后，它会进一步说明“具体哪两组之间有差异”。多样本方差只说明了“整体不一样”，但没说“具体哪两组不一样”。

例如，产品设计师想测试 3 种手机壳材质（A：硅胶，B：TPU，C：硬塑料）的“防滑性能”，每种材质测了 5 个样品，记录了“滑落时间（秒）”（时间越长，防滑越好）

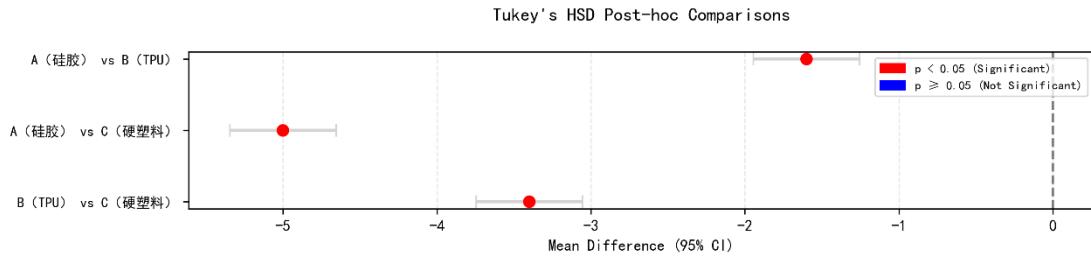
事后多重比较会帮我们判断哪种材质的防滑性能更好，为设计决策提供依据。

示例数据

| 材质分组 | 样品 1 | 样品 2 | 样品 3 | 样品 4 | 样品 5 |
|---------|------|------|------|------|------|
| A (硅胶) | 8.2 | 7.9 | 8.5 | 8.1 | 7.8 |
| B (TPU) | 6.5 | 6.3 | 6.7 | 6.4 | 6.6 |
| C (硬塑料) | 3.2 | 3 | 3.1 | 2.9 | 3.3 |

分析结果

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---------|---------|----------|-------|---------|---------|--------|
| A (硅胶) | B (TPU) | -1.6 | 0.0 | -1.9444 | -1.2556 | True |
| A (硅胶) | C (硬塑料) | -5.0 | 0.0 | -5.3444 | -4.6556 | True |
| B (TPU) | C (硬塑料) | -3.4 | 0.0 | -3.7444 | -3.0556 | True |



Tukey's HSD: 用于方差分析后检验多组均值之间的差异是否显著。

p-adj: 调整后的 p 值。当 p 值小于显著性水平（通常为 0.05）时，拒绝零假设，表明两组均值之间存在显著差异。

meandiff: 两组均值之间的差异，反映两组之间差异的大小。

lower: 差异置信区间的下限。

upper: 差异置信区间的上限。

reject: 是否拒绝零假设。True 表示拒绝，表明两组均值之间存在显著差异。

结论

- 三种材质防滑性能排序：硬塑料 > TPU > 硅胶，硬塑料防滑最好，硅胶最差；
- 差异都显著：所有两两对比的 p-adj 都 = 0.0（远小于 0.05），没有“差不多”的情况，差距都是真实存在的；
- 差距大小量化：硬塑料比 TPU 好 3.4 分，比硅胶好 5 分（硬塑料优势最明显），TPU 比硅胶好 1.6 分（差距较小，但仍显著）。
- 优先选硬塑料做手机壳材质，防滑性能最优；其次选 TPU，硅胶防滑最差。

6 设计方案选择与综合评价

6.1 专家背靠背（德尔菲法）

它是一种“让一群专家悄悄达成共识”的方法，不让专家们当面吵架，而是通过“背靠背”的多轮沟通，把分散的意见慢慢捏合到一起，最后得出一个“相对靠谱”的结论。它适合没有标准答案，需要靠经验判断的问题，步骤是“多轮匿名反馈 + 意见收敛”。

数据收集

步骤 1：定具体问题。比如：“设计一款儿童手表，哪些安全功能是必须的？”，“明年智能家具的流行趋势会有哪些？”。

步骤 2：找“懂行的人”。选 5-15 个懂行的人，比如：做过相关设计的工程师，设计师，该领域的教授，研究员，有丰富经验的从业者。

步骤 3：第一轮提问。把问题编写成问卷，发给专家，自由发挥，匿名回答，比如：“请列出你认为儿童手表必须有的 3 个安全功能”。

关键点：匿名（专家不知道别人是谁），避免“权威压制”（不敢反对老专家）。

步骤 4：分析第一轮结果。把分析结果整理成“反馈报告”，但不透露是哪个专家说的。

步骤 5：第二轮提问，把“反馈报告”发给专家，问：“看到大家的想法后，你要不要修改自己的答案？”，“对于分歧大的‘防摔’功能，你坚持原来的观点吗？为什么？”

步骤 6：分析第二轮结果。第一轮和第二轮的数据都需要。

步骤 7：重复几轮，直到意见相似。

如果第二轮后，大部分专家对“防摔、防水”的重要性打分接近（比如都在 8-9 分），就算“意见相似”了。

如果还有大分歧（比如有人说“续航最重要”，有人说“续航不重要”），就再来第三轮、第四轮，直到多数人看法一致。

步骤 8：把所有的数据放在一个表格里，最后分析一次，得出最终分析结果。

示例数据 第一轮

注意，一次只能分析一个指标。

| 轮次 | 专家 1 | 专家 2 | 专家 3 | 专家 4 | 专家 5 |
|-----|------|------|------|------|------|
| 第一轮 | 8 | 7 | 9 | 6 | 8 |

分析结果 第一轮

| | |
|---------|---------------|
| 各轮评分均值 | ['第一轮: 7.6'] |
| 各轮评分标准差 | ['第一轮: 1.02'] |
| 最终共识评分 | [(7.6)] |
| 评分收敛情况 | ['否'] |
| 数据范围 | ['6.0-9.0'] |

| | |
|---------|---|
| 收敛阈值 | ['0.45'] |
| 专家意见偏离度 | ['专家 1: 0.4', '专家 2: 0.6', '专家 3: 1.4', '专家 4: 1.6', '专家 5: 0.4'] |

结论 第一轮

第一轮专家对该设计要素的整体认可度为 7.6 分，但意见有明显分歧（标准差 1.02），未达成共识，需要进行第二轮。后续重点是解决专家 3、4 的分歧，推动意见集中。

数据 第二轮

| 轮次 | 专家 1 | 专家 2 | 专家 3 | 专家 4 | 专家 5 |
|-----|------|------|------|------|------|
| 第一轮 | 8 | 7 | 9 | 6 | 8 |
| 第二轮 | 8 | 8 | 8 | 7 | 9 |

分析结果 第二轮

| | |
|---------|---|
| 各轮评分均值 | ['第一轮: 7.6', '第二轮: 8.0'] |
| 各轮评分标准差 | ['第一轮: 1.02', '第二轮: 0.63'] |
| 最终共识评分 | [(7.85)] |
| 评分收敛情况 | ['否'] |
| 数据范围 | ['6.0-9.0'] |
| 收敛阈值 | ['0.45'] |
| 专家意见偏离度 | ['专家 1: 0.2', '专家 2: 0.3', '专家 3: 0.7', '专家 4: 1.3', '专家 5: 0.7'] |

结论 第二轮

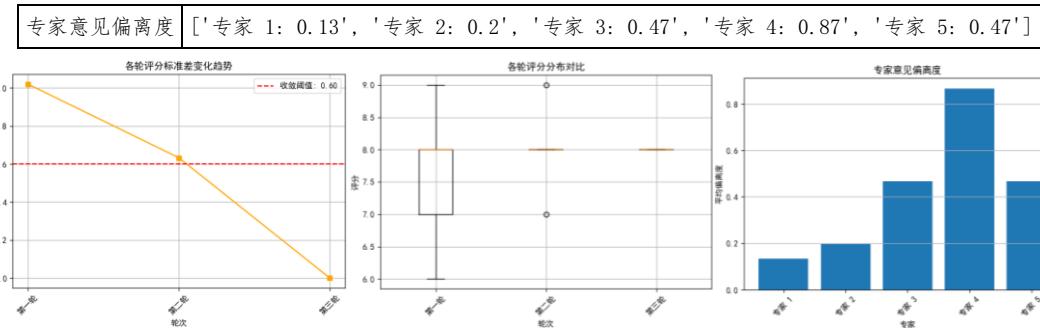
两轮之后，专家意见从“明显分歧”变成“相对集中”，整体认可度也在提升，但还没达到“完全收敛”的标准（标准差 $0.63 > 0.45$ ），需要再进行一轮咨询，就能达成最终共识。

数据 第三轮

| 轮次 | 专家 1 | 专家 2 | 专家 3 | 专家 4 | 专家 5 |
|-----|------|------|------|------|------|
| 第一轮 | 8 | 7 | 9 | 6 | 8 |
| 第二轮 | 8 | 8 | 8 | 7 | 9 |
| 第三轮 | 9 | 8 | 8 | 8 | 8 |

分析结果 第三轮

| | |
|---------|--|
| 各轮评分均值 | ['第一轮: 7.6', '第二轮: 8.0', '第三轮: 8.0'] |
| 各轮评分标准差 | ['第一轮: 1.02', '第二轮: 0.63', '第三轮: 0.0'] |
| 最终共识评分 | [(7.92)] |
| 评分收敛情况 | ['是'] |
| 数据范围 | ['6.0-9.0'] |
| 收敛阈值 | ['0.60'] |



第三轮已收敛。

结论 第三轮

专家已达成高度共识，8.02 分就是最终结论，可直接用于设计决策。

6.2 评价指标筛选（主观数据）

依据主观数据筛选指标的方法有两种，“层次分析法 AHP 与模糊层次分析法 FAHP”，它们都是用来比较多个因素的重要性的，并且它们的“数据收集和处理”的过程完全一样。它们的区别在于，

- AHP 适用于“评价标准清晰、决策者能精确判断”的场景，例如，工程方案中“成本”与“工期”的重要性对比（可基于历史数据明确判断“成本比工期重要，比值为 5”），主观判断无模糊性，或模糊性可忽略。
- FAHP 适用于“评价标准模糊、决策者难以精确判断”的场景，例如，设计方案中“美观度”与“用户体验”的重要性对比（无法明确说“比值是 3 还是 4”，只能说“美观度比用户体验略重要，大概在 3 左右”），主观评价存在明显模糊性，且无法忽略。

例如，在“智能穿戴设备设计方案评估”的场景中，我们需要考虑 8 个评价指标中，哪个更为重要，需要投入更多资源。

人机交互体验 (A), 视觉美学设计 (B), 功能实用性 (C), 材质耐久性 (D), 续航能力 (E), 价格合理性 (F), 便携性 (G), 创新技术应用 (H)

数据收集

- 选择三类核心评价者，回答这些题目
 设计领域专家（3-5 人，侧重人机交互与美学）
 技术工程师（2-3 人，擅长硬件性能评估）
 目标用户代表（3-5 人，提供实际使用反馈）
- 采用 Saaty 1-9 标度法作为判断矩阵的数值基础
 1 级：同等重要（如两款设备的便携性无显著差异）
 3 级：稍微重要（如人机交互体验略优于视觉设计）
 5 级：明显重要（如功能实用性显著高于材质耐久性）
 7 级：强烈重要（如创新技术应用远胜于价格因素）

9 级：极端重要（如核心功能缺失将导致产品失效）

偶数 2/4/6/8 用于相邻等级的过渡判断。

- 每一道题的提问都是这样的
你认为“人机交互体验”和“视觉美学设计”哪个更重要？
- 8 个指标的“不重复两两组合”数量是： $n \times (n-1)/2 = 8 \times 7/2 = 28$ ，因此共设计 28 道题，调研问卷如下表所示。

| 序号 | 左侧指标 | 右侧指标 | 左侧比右侧更重要 | 右侧比左侧更重要 |
|----|-----------|-----------|--|--|
| 1 | A. 人机交互体验 | B. 视觉美学设计 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 2 | A. 人机交互体验 | C. 功能实用性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 3 | A. 人机交互体验 | D. 材质耐久性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 4 | A. 人机交互体验 | E. 续航能力 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 5 | A. 人机交互体验 | F. 价格合理性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 6 | A. 人机交互体验 | G. 便携性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 7 | A. 人机交互体验 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 8 | B. 视觉美学设计 | C. 功能实用性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 9 | B. 视觉美学设计 | D. 材质耐久性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 10 | B. 视觉美学设计 | E. 续航能力 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 11 | B. 视觉美学设计 | F. 价格合理性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 12 | B. 视觉美学设计 | G. 便携性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 13 | B. 视觉美学设计 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 14 | C. 功能实用性 | D. 材质耐久性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 15 | C. 功能实用性 | E. 续航能力 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 16 | C. 功能实用性 | F. 价格合理性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 17 | C. 功能实用性 | G. 便携性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 18 | C. 功能实用性 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 19 | D. 材质耐久性 | E. 续航能力 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 20 | D. 材质耐久性 | F. 价格合理性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 21 | D. 材质耐久性 | G. 便携性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 22 | D. 材质耐久性 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 23 | E. 续航能力 | F. 价格合理性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 24 | E. 续航能力 | G. 便携性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 25 | E. 续航能力 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 26 | F. 价格合理性 | G. 便携性 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 27 | F. 价格合理性 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |
| 28 | G. 便携性 | H. 创新技术应用 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 | <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 |

数据处理

制作 Excel 表格时，“左侧比右侧重要”列的数字不变，“右侧比左侧重要”列的数字需改为倒数，例如 3 改为 1/3, 7 改为 1/7, 1/5 改为 5，然后把所有的数据填入上三角。为了方便理解，这里只取了八行八列数据中的前四行四列的数据。

| | A | B | C | D |
|---|---|---|-----|---|
| A | | 3 | 1/2 | 5 |
| B | | | 1/4 | 3 |
| C | | | | 7 |
| D | | | | |

每一个评价者都会有一个矩阵，我们需要把所有的矩阵合成一个“最终矩阵”。

假如一共有3个评价者Tony, Mike和Lily, 下面的表格依次为Tony, Mike和Lily的。首先需要为他们分配权重，也就是他们的话语权，权重的分配由研究者自己决定，但需要保证权重的和为1。假设Tony为0.5, Mike为0.3, Lily为0.2。

我们需要计算最终矩阵中的数值，计算方法就是每一个“数字×权重”的和，以标记黄色的位置为例，最终矩阵中的数值为： $5 \times 0.5 + 8 \times 0.3 + 1/4 \times 0.2 = 4.95$ 。我们可通过四舍五入调整数字，4.95可修正为5。

| | A | B | C | D |
|---|---|---|-----|---|
| A | | 3 | 1/2 | 5 |
| B | | | 1/4 | 3 |
| C | | | | 7 |
| D | | | | |

| | A | B | C | D |
|---|---|---|-----|---|
| A | | 7 | 3 | 8 |
| B | | | 1/4 | 3 |
| C | | | | 5 |
| D | | | | |

| | A | B | C | D |
|---|---|---|-----|-----|
| A | | 5 | 3 | 1/4 |
| B | | | 1/4 | 1/7 |
| C | | | | 1/5 |
| D | | | | |

得到“最终矩阵”后，需要补全整个矩阵，上三角和下三角中的数据对称且互为倒数，对角线填1即可，如下图所示。为了更容易理解数据的“对称性”，对称轴上的数字已经标记了红色。

| | A | B | C | D |
|---|-----|-----|-----|---|
| A | 1 | 5 | 2 | 5 |
| B | 1/5 | 1 | 1/4 | 2 |
| C | 1/2 | 4 | 1 | 5 |
| D | 1/5 | 1/2 | 1/5 | 1 |

最后，表格中的“分数”需要通过公式重新生成，并设置分数的单元格格式为分数。

(例如单元格 3B = 1/2C, 单元格 3B 和单元格 2C 的位置是对称的)

The screenshot shows the 'Format Cells' dialog box for cell B3. The formula bar at the top has the formula $=1/C2$. The 'Number' tab is selected. In the 'Fraction' section of the list, the '分母为一位数(1/4)' option is highlighted with a red arrow. The preview area on the right shows the fraction $1/5$.

示例数据 上表

程序自动区分第一行和第一列是表头而非数据，因此第一行和第一列非数字。确保所有数据依据对角线对称且互为倒数。

分析结果

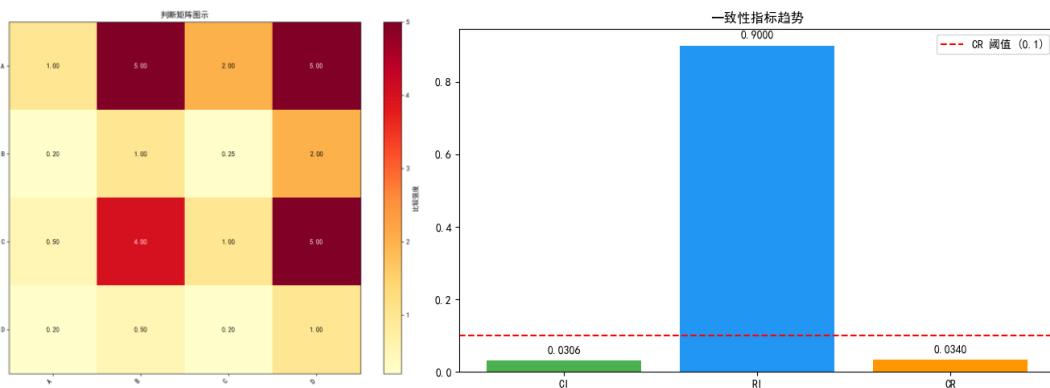
- 特征向量: 0.4946, 0.1046, 0.3308, 0.0700
- 最大特征值: 4.0917
- 一致性指标 CI: 0.0306
- 随机一致性指标 RI: 0.9
- 一致性比率 CR: 0.034
- 通过 ($CR < 0.1$) - 结果可靠

特征向量: 反映各因素相对重要性的向量

一致性指标 CI: 衡量判断矩阵一致性的指标

随机一致性指标 RI: 根据矩阵阶数确定的随机一致性指标

一致性比率 CR: CI 与 RI 的比值, 判断矩阵是否具有满意一致性



结论

8个因素的重要性排序: A > C > B > D (A 最重要, D 最不重要)。

这个排序很靠谱 ($CR=0.0581 < 0.1$, 通过检验), 可以放心用。

6.3 评价指标筛选 (客观数据)

指标筛选的方法有六种, “熵值法, CRITIC 权重法, 独立性权重法, 信息量权重法, AHP 和 FAHP”, 它们都是用来计算“权重”的, 也就是“指标筛选工具”。它们可以从一堆评价指标里, 找出“谁更重要”, 并对这些指标按照重要性排序。区别在于,

- “熵值法, CRITIC 权重法, 独立性权重法和信息量权重法”是数据驱动的(客观计算), 必须有实际数据(比如用户评分、测试结果), 数据不变, 权重就固定(客观性强)。

- “AHP 和 FAHP”是经验驱动的（主观判断），不需要实际数据（只要有人能判断“谁重要”就行），换一批人判断，权重可能变（主观性强），数据格式必须是矩阵，且必须提前为每个人分配权重。

怎么选

- 连续数据，指标互相独立 → 熵值法；
- 连续数据，指标间高度相关，保留所有指标 → CRITIC；
- 连续数据，指标间高度相关，精简指标 → 独立性权重法；
- 分类数据 → 信息量权重法。

熵值法 示例数据

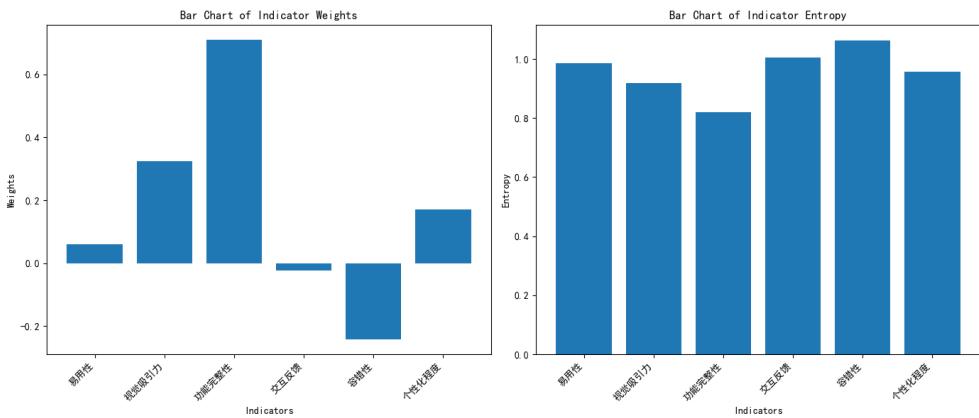
在“电商APP用户体验评价”的场景中，收集了每一个用户对六种用户体验指标的评分，现在需要判断这六种用户体验指标谁更重要。

程序默认所有指标为正向指标（即数值越大越好的指标），如果有负向指标，可以通过取倒数转换为正向指标。数据中不能有极端值。

| | A | B | C | D | E | F |
|---|-----|-------|-------|------|-----|-------|
| 1 | 易用性 | 视觉吸引力 | 功能完整性 | 交互反馈 | 容错性 | 个性化程度 |
| 2 | 8.2 | 7.5 | 9 | 6.8 | 7.2 | 5.5 |
| 3 | 7.8 | 8.1 | 8.5 | 7.3 | 6.9 | 6.2 |
| 4 | 9.1 | 6.9 | 8.8 | 7.7 | 8 | 5.8 |
| 5 | 6.5 | 8.7 | 7.2 | 8 | 6.5 | 7 |
| 6 | 0.5 | 7.0 | 0.2 | 0.5 | 7.0 | 0.2 |

分析结果

| 统计量类型 | 序号 | 指标名称 | 数值 |
|-------|----|-------|-----------|
| 指标权重 | 1 | 易用性 | 0.060505 |
| 指标权重 | 2 | 视觉吸引力 | 0.324284 |
| 指标权重 | 3 | 功能完整性 | 0.708979 |
| 指标权重 | 4 | 交互反馈 | -0.022639 |
| 指标权重 | 5 | 容错性 | -0.242361 |
| 指标权重 | 6 | 个性化程度 | 0.171231 |
| 指标熵值 | 1 | 易用性 | 0.984620 |
| 指标熵值 | 2 | 视觉吸引力 | 0.917571 |
| 指标熵值 | 3 | 功能完整性 | 0.819787 |
| 指标熵值 | 4 | 交互反馈 | 1.005755 |
| 指标熵值 | 5 | 容错性 | 1.061605 |
| 指标熵值 | 6 | 个性化程度 | 0.956475 |



结论

- 优先优化“功能完整性”：它是最关键指标（权重 0.709），设计时先确保“核心功能都有”（比如购物 APP 必须有下单、支付功能），再考虑其他。
- 其次打磨“视觉吸引力”：重要性第二（权重 0.324），功能完整后，再优化界面美观度（如配色、图标）。
- 暂时忽略“容错性”和“交互反馈”：这两个指标权重为负、熵值高，不是说它们不重要，而是当前用户评分数据太混乱，没法用它们指导设计。
- “易用性”和“个性化”先不用管：权重低，对综合体验影响小，等核心指标优化完再微调即可。

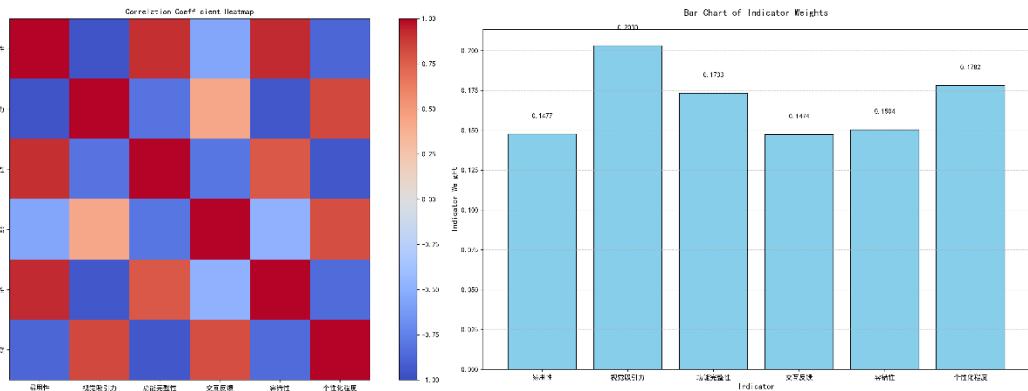
CRITIC 权重法 示例数据

程序默认所有指标为正向指标（即数值越大越好的指标），如果有负向指标，可以通过取倒数转换为正向指标。数据中不能有极端值。

分析结果

| | |
|--------|---|
| 原始数据 | 行标题: ['样本 1', '样本 2', '样本 3', '样本 4', '样本 5', '样本 6', '样本 7', '样本 8', '样本 9', '样本 10'] |
| 矩阵 | 列标题: ['易用性', '视觉吸引力', '功能完整性', '交互反馈', '容错性', '个性化程度'] 数据: [[8.2, 7.5, 9.0, 6.8, 7.2, 5.5], [7.8, 8.1, 8.5, 7.3, 6.9, 6.2], [9.1, 6.9, 8.8, 7.7, 8.0, 5.8], [6.5, 8.7, 7.2, 8.0, 6.5, 7.0], [8.5, 7.8, 9.2, 6.5, 7.6, 5.2], [7.3, 8.3, 8.0, 7.1, 7.0, 6.5], [8.8, 7.0, 8.9, 6.9, 8.2, 5.0], [6.9, 8.5, 7.5, 7.8, 6.8, 6.8], [8.0, 7.6, 8.7, 7.0, 7.5, 5.6], [7.6, 8.0, 8.3, 7.5, 7.1, 6.3]] |
| 标准差矩阵 | 指标标题: ['易用性', '视觉吸引力', '功能完整性', '交互反馈', '容错性', '个性化程度'] 数据: [0.8246884933786278, 0.600370256129036, 0.6607235091591304, 0.48350571638583323, 0.5391351098441526, 0.6756889488185784] |
| 相关系数矩阵 | 指标标题: ['易用性', '视觉吸引力', '功能完整性', '交互反馈', '容错性', '个性化程度'] 数据: [[1.0, -0.953307249265088, 0.9100697577186957, -0.5495065242229458, 0.9281364737281838, -0.8819372355645252], [-0.953307249265088, 1.0, -0.8134206151170503, 0.4233424671590225, -0.9378241290708313, 0.8337505908575934], [0.9100697577186957, -0.8134206151170502, 1.0, -0.9378241290708313, 0.8337505908575934], [0.9100697577186957, -0.8134206151170502, 1.0, -0.9378241290708313, 0.8337505908575934], [0.9100697577186957, -0.8134206151170502, 1.0, -0.9378241290708313, 0.8337505908575934], [0.9100697577186957, -0.8134206151170502, 1.0, -0.9378241290708313, 0.8337505908575934]] |

| | |
|-------|---|
| | 0.7985608257052893, 0.7741799024433437, -0.938030700325135], [-0.5495065242229458, 0.4233424671590225, -0.7985608257052893, 1.0, -0.4808032139672864, 0.8114831398951474], [0.9281364737281836, -0.9378241290708315, 0.7741799024433438, -0.4808032139672864, 1.0, -0.8546360152598682], [-0.8819372355645252, 0.8337505908575934, -0.938030700325135, 0.8114831398951474, -0.8546360152598682, 1.0]] |
| 相关系数 | 指标标题: ['易用性', '视觉吸引力', '功能完整性', '交互反馈', '容错性', '个性化程度'] 数据: [[0.0, 1.9653114624459118e-05, 0.00025641634547185553, 0.09988858420936739, 0.00010692002558841023, 0.0007354185598753838], [1.9653114624459118e-05, 0.0, 0.004204665622699856, 0.22282086819294686, 6.0630064530154766e-05, 0.002720436539493603], [0.00025641634547185553, 0.004204665622699875, 0.0, 0.005604296554216612, 0.008574752119715813, 5.984350597749122e-05], [0.09988858420936739, 0.22282086819294686, 0.005604296554216612, 0.0, 0.1595118783075405, 0.004371133311017667], [0.00010692002558841023, 6.0630064530154766e-05, 0.008574752119715813, 0.1595118783075405, 0.0, 0.001632918160269163], [0.0007354185598753838, 0.002720436539493603, 5.984350597749122e-05, 0.004371133311017667, 0.001632918160269163, 0.0]] |
| 信息量矩阵 | 指标标题: ['易用性', '视觉吸引力', '功能完整性', '交互反馈', '容错性', '个性化程度'] 数据: [4.120999948537577, 4.125024797887926, 3.875647170330663, 2.411280569149436, 2.744275219569024, 4.073978826257946] |
| 指标权重 | 指标标题: ['易用性', '视觉吸引力', '功能完整性', '交互反馈', '容错性', '个性化程度'] 数据: [0.1930101674775553, 0.19319867435862398, 0.18151888346780787, 0.11293415974248316, 0.1285302175074008, 0.1908078974461289] |



结论

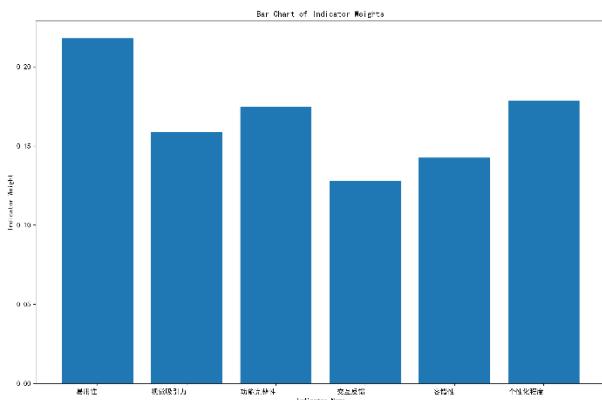
- 最重要的两个指标: 易用性、视觉吸引力 (权重最高, 各占 19.3%);
- 最不重要的指标: 交互反馈 (权重最低, 仅 11.3%);
- 需要注意的关联: 易用性好的样本, 视觉吸引力往往差 (负相关), 不用追求“两个指标都满分”;
- 区分样本的关键: 易用性最能看出样本好坏, 能快速筛掉差样本。

独立性权重法 示例数据

程序默认所有指标为正向指标 (即数值越大越好的指标), 如果有负向指标, 可以通过取倒数转换为正向指标。数据中不能有极端值。

分析结果

| | |
|--------|--|
| 原始数据矩阵 | [8.2, 7.5, 9.0, 6.8, 7.2, 5.5], [7.8, 8.1, 8.5, 7.3, 6.9, 6.2], [9.1, 6.9, 8.8, 7.7, 8.0, 5.8], [6.5, 8.7, 7.2, 8.0, 6.5, 7.0], [8.5, 7.8, 9.2, 6.5, 7.6, 5.2], [7.3, 8.3, 8.0, 7.1, 7.0, 6.5], [8.8, 7.0, 8.9, 6.9, 8.2, 5.0], [6.9, 8.5, 7.5, 7.8, 6.8, 6.8], [8.0, 7.6, 8.7, 7.0, 7.5, 5.6], [7.6, 8.0, 8.3, 7.5, 7.1, 6.3] |
| 预处理后数据 | [8.2, 7.5, 9.0, 6.8, 7.2, 5.5], [7.8, 8.1, 8.5, 7.3, 6.9, 6.2], [9.1, 6.9, 8.8, 7.7, 8.0, 5.8], [6.5, 8.7, 7.2, 8.0, 6.5, 7.0], [8.5, 7.8, 9.2, 6.5, 7.6, 5.2], [7.3, 8.3, 8.0, 7.1, 7.0, 6.5], [8.8, 7.0, 8.9, 6.9, 8.2, 5.0], [6.9, 8.5, 7.5, 7.8, 6.8, 6.8], [8.0, 7.6, 8.7, 7.0, 7.5, 5.6], [7.6, 8.0, 8.3, 7.5, 7.1, 6.3] |
| 标准差矩阵 | '0.7824', '0.5696', '0.6268', '0.4587', '0.5115', '0.6410' |
| 指标权重 | '0.2179', '0.1587', '0.1746', '0.1278', '0.1425', '0.1786' |
| 权重计算过程 | '0.7824 / 3.5899 = 0.2179', '0.5696 / 3.5899 = 0.1587', '0.6268 / 3.5899 = 0.1746', '0.4587 / 3.5899 = 0.1278', '0.5115 / 3.5899 = 0.1425', '0.6410 / 3.5899 = 0.1786' |



结论

- 优先优化“易用性”：它是最关键的指标（权重 21.79%），设计时先确保“操作简单”（比如按钮位置合理、流程短），这是提升用户体验的核心。
- 其次打磨“个性化”和“功能完整性”：两者权重都近 18%，功能要全（核心需求别漏），同时增加个性化选项（比如主题切换），能显著提升体验。
- “交互反馈”不用优先管：权重最低（12.78%），不是说它不重要，而是当前用户对它的感知差异小，等核心指标优化完，再微调即可。
- 评估设计时，不用“平均用力”：把 70% 的精力放在“易用性 + 个性化 + 功能完整性”上（三者权重加起来近 57%），效率最高。

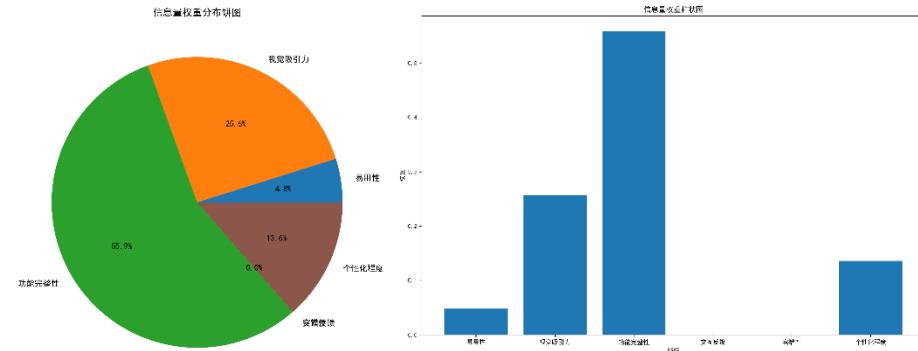
设计不用面面俱到，先把“易用性”做好，再补“个性化”和“功能”，就是提升体验的最快方式！

信息量权重法 示例数据

程序默认所有指标为正向指标（即数值越大越好的指标），如果有负向指标，可以通过取倒数转换为正向指标。数据中不能有极端值。

分析结果

| | |
|-------|--|
| 指标熵值 | [0.9843523134134794, 0.9172627126314764, 0.8195769554238457, 1.0054698923728005, 1.0612735106927342, 0.9561560711406422] |
| 指标冗余度 | [0.015647686586520604, 0.0827372873685236, 0.18042304457615432, -0.005469892372800489, -0.061273510692734234, 0.04384392885935784] |
| 信息量权重 | [0.048497108767112905, 0.2564289105882062, 0.5591878370340656, 0.0, 0.0, 0.1358861436106153] |



结论

- 优先优化“功能完整性”：它是 55.9% 的核心，先确保 APP 有用户需要的核心功能（比如购物 APP 要有下单、支付），这是体验的基础。
- 其次打磨“视觉吸引力”：25.6% 的重要性，功能全了再优化界面颜值（配色、图标），能快速提升用户好感。
- “交互反馈、容错性”暂时忽略：权重 0，当前评分数据没用，要么重新找用户评这两个指标（让评分有规律），要么先不管，等核心指标优化完再说。
- “易用性、个性化程度”不用急：权重低，对综合体验影响小，后续微调即可。

6.4 综合评价（排名）

综合评价(排名)包括“灰色关联分析法,综合指数,功效系数法,因子分析和 TOPSIS”，五种方法。它们都是常用的多指标综合评价方法，也就是，在一堆“各有优缺点”的选项里，综合来看，挑出“最靠谱的”那个，并且给所有选项排个先后顺序。比如企业选择供应商、学生选大学、项目选方案，只要是“多个选项、多个评价维度(有好有坏)”的场景，都能用它来做客观的选择或排序。

它们的核心区别是

| 方法 | 用途 | 最适合的场景 |
|---------|--|--|
| 灰色关联分析法 | 找数据列和参考列的关联程度。比如成本、重量、美观度和“销量”的关联程度。数据少、不全，模糊，有很多不确定因素。只看“谁在前、谁在后” | 1. 评价刚上市的 5 款新能源汽车（数据少，只有 10 个用户评分）。 2. 分析某新政策实施 3 个月的效果（数据不全，只有部分区域数据）。 3. 测试一款新研发的 APP 体验（样本量小，不确定因素多）。 |
| 综合指数 | 给多个对象“算总分”。指标少(3-5 个)，自己定权重。 | 1. 算“城市幸福感指数”（整合收入、教育、医疗、环境 4 个指标）； 2. 评“班级学生综合表现”（成绩 60%、纪律 20%、活动 20%）； 3. 算“店铺经营指数”（销售额 50%、客流量 30%、复购率 20%）。 |

| | | |
|--------|---|--|
| 功效系数法 | 给多个对象“算总分”。有“目标标准”（比如满分线、及格线），有极端值。 | 1. 企业给员工做绩效考核（定好“营收满分 100 万、及格 30 万”“利润满分 50 万、及格 10 万”）； 2. 评“学校教学质量”（升学率满分 90%、及格 60%，平均分满分 85 分、及格 70 分）； 3. 政府考核区域环保（PM2.5 满分 20 $\mu\text{g}/\text{m}^3$ 、及格 50 $\mu\text{g}/\text{m}^3$ ）。 |
| 因子分析 | 找出背后的“潜在因素”，简化指标。指标多（10 个以上），且有关联（比如“收入”和“消费”相关）。 | 1. 分析 1000 个用户对某产品的满意度（有 20 个指标：外观、价格、功能、售后等，提炼出“性价比”，“体验感”等）； 2. 评 50 个城市的竞争力（有 30 个指标：GDP、人口、交通、教育等，简化成“经济活力”，“基础设施”2 个核心因子）； 3. 研究消费者购物偏好（15 个指标：品牌、价格、口碑等，找出“品质导向”，“性价比导向”等潜在需求）。 |
| TOPSIS | 多个方案各有优劣，综合对比，找出最接近完美的选项。不仅看排名，还看“和最好的差多少”，“和最差的差多少”。自己定权重。 | 1. 公司从 10 个供应商里选最优（对比价格、交货速度、质量、售后 4 个指标，看谁离“最好供应商”最近）； 2. 评 8 个旅游城市的宜居度（对比气候、交通、物价、景点，找“最接近理想城市”的）； 3. 选最优的产品方案（5 个方案，对比成本、效果、风险，看谁离“最好方案”近、离“最差方案”远）。 |

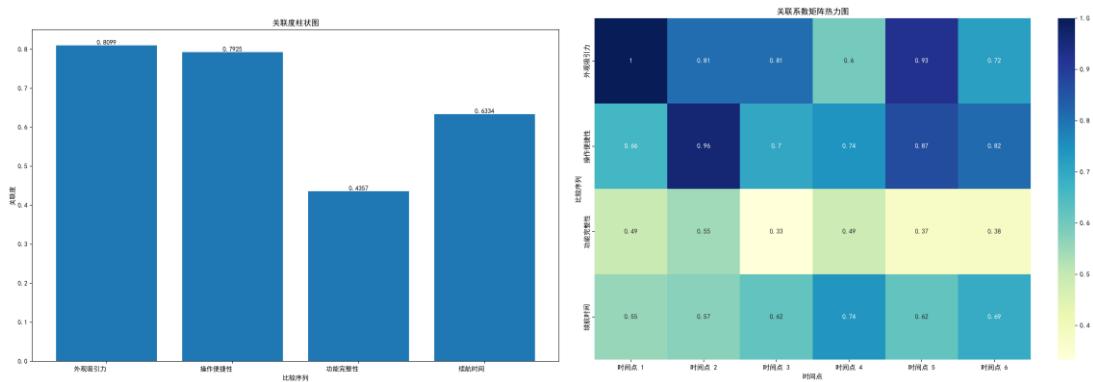
灰色关联分析法示例数据

第一列是参考序列，数据不能有 0，不能有负值

| 用户满意度 | 外观吸引力 | 操作便捷性 | 功能完整性 | 续航时间 |
|-------|-------|-------|-------|------|
| 8.5 | 9.2 | 7.8 | 12 | 7 |
| 7.9 | 8.5 | 8 | 10 | 6 |
| 9.1 | 9 | 8.5 | 15 | 8 |
| 6.8 | 7.2 | 6.5 | 8 | 5 |
| 8.3 | 8.8 | 8.2 | 13 | 7 |
| 7.5 | 8 | 7.6 | 11 | 6 |

分析结果

| | |
|---------|---|
| 参考序列 | [8.5, 7.9, 9.1, 6.8, 8.3, 7.5] |
| 比较序列 | 外观吸引力：[9.2, 8.5, 9.0, 7.2, 8.8, 8.0]，操作便捷性：[7.8, 8.0, 8.5, 6.5, 8.2, 7.6]，功能完整性：[12.0, 10.0, 15.0, 8.0, 13.0, 11.0]，续航时间：[7.0, 6.0, 8.0, 5.0, 7.0, 6.0] |
| 关联系数矩阵 | [[1.0, 0.8085106382978724, 0.8085106382978724, 0.5984251968503936, 0.9268292682926832, 0.7169811320754716], [0.6619958341223255, 0.9620253164556963, 0.7026194144838213, 0.7435792906644924, 0.8697009102730813, 0.8150134048257374], [0.4947636569487688, 0.5467625899280576, 0.3333333333333333, 0.48927038626609437, 0.3730893673993084, 0.3769373837569746], [0.554832566259324, 0.5714285714285715, 0.6212534059945504, 0.7384615384615386, 0.6221395348837208, 0.69248291571754]] |
| 关联度 | '外观吸引力：0.8099'，'操作便捷性：0.7925'，'功能完整性：0.4357'，'续航时间：0.6334' |
| 关联度排序结果 | ['1: 外观吸引力 (关联度: 0.8099)', '2: 操作便捷性 (关联度: 0.7925)', '4: 功能完整性 (关联度: 0.4357)', '3: 续航时间 (关联度: 0.6334)'] |



结论

4个因素对“用户满意度”的影响程度排序是：1. 外观吸引力(影响最大, 关联度 0.8099) > 2. 操作便捷性(影响第二, 关联度 0.7925)> 3. 续航时间(影响第三, 关联度 0.6334) > 4. 功能完整性 (影响最小, 关联度 0.4357)

想提升用户满意度，优先优化“外观”和“操作便捷性”，效果比优化“功能”和“续航”更明显。

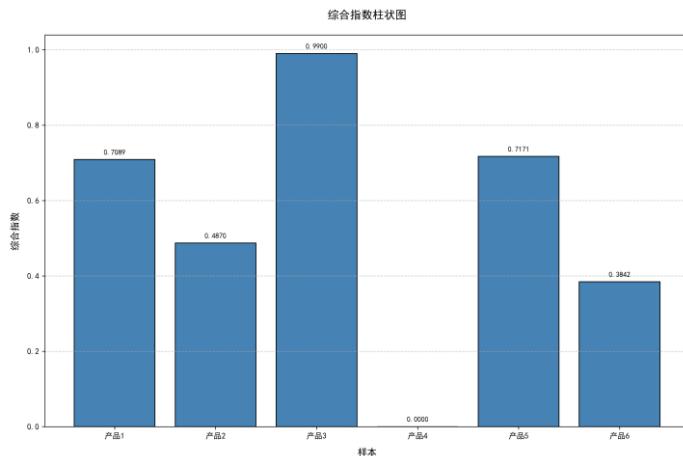
综合指数示例数据

注意，第一行和第一列不是数据，第二行是权重值。

| | 用户满意度 | 外观吸引力 | 操作便捷性 | 功能完整性 | 续航时间 |
|------|-------|-------|-------|-------|------|
| 权重 | 0.3 | 0.1 | 0.2 | 0.1 | 0.3 |
| 产品 1 | 8.5 | 9.2 | 7.8 | 12 | 7 |
| 产品 2 | 7.9 | 8.5 | 8 | 10 | 6 |
| 产品 3 | 9.1 | 9 | 8.5 | 15 | 8 |
| 产品 4 | 6.8 | 7.2 | 6.5 | 8 | 5 |
| 产品 5 | 8.3 | 8.8 | 8.2 | 13 | 7 |
| 产品 6 | 7.5 | 8 | 7.6 | 11 | 6 |

分析结果

| | |
|------|--|
| 综合指数 | 0.7088819875776398, 0.48704968944099386, 0.99, 0.0, 0.7170807453416149, 0.3841614906832298 |
| 指标权重 | 用户满意度: 0.3000, 外观吸引力: 0.1000, 操作便捷性: 0.2000, 功能完整性: 0.1000, 续航时间: 0.3000 |



结论

产品 3 最好 (0.99 分), 产品 4 最差 (0 分); 而且算分时, “用户满意度”和“续航时间”这两个维度最关键 (各占 30%), 想提升产品综合分, 优先优化这两项。

功效系数法 示例数据

第一行表头必须是“指标名称 | 实际值 | 不允许值 | 满意值 | 权重”, 满意值不能等于不允许值, 实际值不能超出[不允许值, 满意值]范围

程序默认所有指标均为正向指标, 数字越大越好。负向指标可通过“转换后实际值 = 不允许值 - 原实际值 + 满意值”转化为正向指标, 交换不允许值和满意值的位置。

如下表, 重量为负向指标

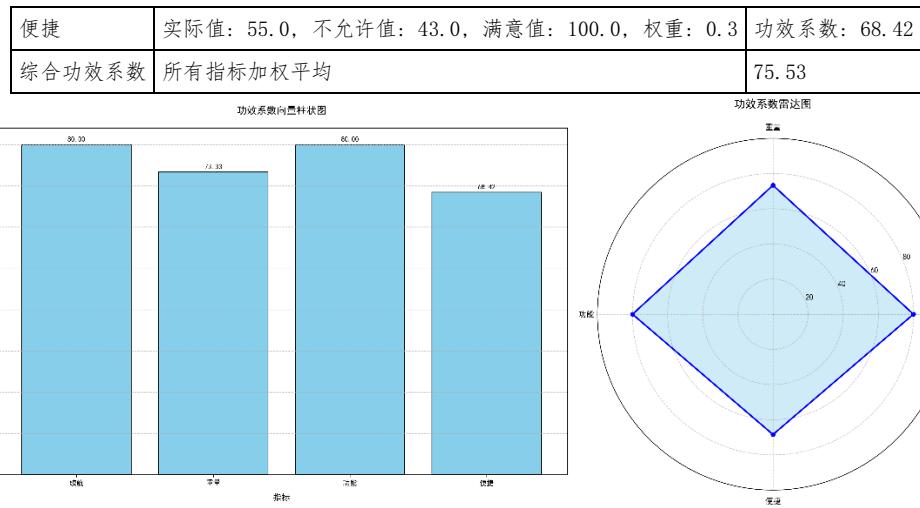
| 指标名称 | 实际值 | 不允许值 | 满意值 | 权重 |
|-----------|-----------|-----------|-----------|-------------|
| 续航 | 30 | 12 | 48 | 0.2 |
| 重量 | 50 | 60 | 30 | 0.15 |
| 功能 | 60 | 40 | 80 | 0.35 |
| 便捷 | 55 | 43 | 100 | 0.3 |

50 转换为 “ $60 - 50 + 30 = 40$ ”

| 指标名称 | 实际值 | 不允许值 | 满意值 | 权重 |
|-----------|-----------|-----------|-----------|-------------|
| 续航 | 30 | 12 | 48 | 0.2 |
| 重量 | 40 | 30 | 60 | 0.15 |
| 功能 | 60 | 40 | 80 | 0.35 |
| 便捷 | 55 | 43 | 100 | 0.3 |

分析结果

| 指标/统计量 | 详细信息 | 结果值 |
|--------|--|-------------|
| 续航 | 实际值: 30.0, 不允许值: 12.0, 满意值: 48.0, 权重: 0.2 | 功效系数: 80.00 |
| 重量 | 实际值: 40.0, 不允许值: 30.0, 满意值: 60.0, 权重: 0.15 | 功效系数: 73.33 |
| 功能 | 实际值: 60.0, 不允许值: 40.0, 满意值: 80.0, 权重: 0.35 | 功效系数: 80.00 |



结论

- 这款产品的综合表现得分为 75.53 分 (满分 100 分, 60 分是及格线, 越高越好), 4 个指标的表现排名是: 1. 续航 (80 分, 表现最好) = 功能 (80 分, 表现最好) > 2. 重量 (73.33 分, 表现中等) > 3. 便捷 (68.42 分, 表现稍弱)
- 产品的“续航”和“功能”做得不错, “便捷性”还有提升空间, 整体综合表现处于中等偏上水平。

因子分析 示例数据

数据最好符合正态分布和方差齐。

| | A | B | C | D | E | F | G | H | I | J |
|---|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|
| 1 | 设计美观度 | 色彩协调性 | 排版合理性 | 图标识别度 | 操作便捷性 | 学习难度 | 功能完整性 | 功能实用性 | 交互流畅度 | 个性化程度 |
| 2 | 4 | 5 | 4 | 5 | 3 | 4 | 4 | 5 | 5 | 4 |
| 3 | 5 | 4 | 5 | 4 | 4 | 3 | 5 | 4 | 4 | 5 |
| 4 | 3 | 3 | 3 | 4 | 5 | 2 | 3 | 3 | 5 | 3 |
| 5 | 4 | 5 | 4 | 5 | 2 | 5 | 4 | 5 | 3 | 4 |
| 6 | 5 | 5 | 5 | 5 | 4 | 3 | 5 | 5 | 4 | 5 |

操作步骤

- 因子数量, 也就是最后筛选出的最重要的指标的数量, 可根据需要修改。
- Varimax (正交旋转): 指标之间不相关, 结果更简洁, 解释更直观。
- Promax (斜交旋转): 指标之间存在相关性, 更符合实际情况。

建议因子数量: 2
特征值大于1的因子数: 2
请输入因子数量:

因子数量:

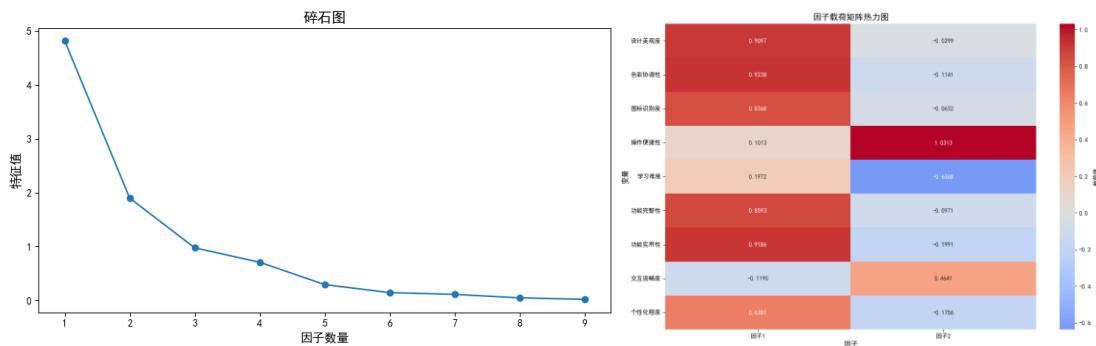
选择旋转方法:

varimax (正交旋转)
 promax (斜交旋转)

分析结果

内容较多，此处仅显示部分内容。

| 变量 | 因子 1 | 因子 2 |
|-------|---------|---------|
| 设计美观度 | 0.9097 | -0.0299 |
| 色彩协调性 | 0.9238 | -0.1141 |
| 图标识别度 | 0.8368 | -0.0632 |
| 操作便捷性 | 0.1013 | 1.0313 |
| 学习难度 | 0.1972 | -0.6368 |
| 功能完整性 | 0.8593 | -0.0971 |
| 功能实用性 | 0.9186 | -0.1991 |
| 交互流畅度 | -0.1190 | 0.4641 |
| 个性化程度 | 0.6381 | -0.1756 |



结论

- 简化评价维度：重点关注“视觉与功能综合质量”（因子 1）和“使用体验便捷性”（因子 2）两个核心维度，就能快速判断设计方案的优劣；
- 定位改进方向：如果某个设计方案“因子 1 得分低”，就重点优化外观、色彩、功能完整性；如果“因子 2 得分低”，就聚焦简化操作、降低学习难度；
- 识别无效指标：交互流畅度、个性化程度的共同度低，说明当前评价体系里这两个指标“参考价值有限”，后续可以考虑调整或替换。

数据收集 TOPSIS

某企业计划推出新一代智能手环，设计团队初步完成 5 套设计方案，但需从用户体验、技术可行性、成本控制等多维度筛选最优方案。由于设计决策涉及定量指标（如续航时间、生产成本）与定性指标（如外观吸引力、操作便捷性）的综合权衡，传统单一维度评价方法难以满足需求。

通过文献分析与专家访谈，确定 6 项核心评价指标（含效益型与成本型）：

- 效益型指标（数值越大越优）
 - 续航时间（小时）：用户对智能设备的核心需求，反映产品续航能力
 - 外观吸引力（1-10 分）：由 20 名设计专业学生评分的平均值
 - 操作便捷性（1-10 分）：基于用户测试的操作流畅度评分

- 功能丰富度（1-10分）：支持的健康监测、运动模式等功能数量评分
- 成本型指标（数值越小越优）
 - 生产成本（元/台）：单台设备的物料与组装成本
 - 重量（克）：影响佩戴舒适度

邀请5位专家（包括设计总监、产品经理、工程师）采用层次分析法AHP打分，得到指标权重（或熵值法等）。

| 指标 | 续航时间 | 外观吸引力 | 操作便捷性 | 功能丰富度 | 生产成本 | 重量 |
|-----|------|-------|-------|-------|------|------|
| 权重值 | 0.25 | 0.20 | 0.15 | 0.10 | 0.20 | 0.10 |

效益型数据（续航时间，外观吸引力，操作便捷性，功能丰富度）是通过20名设计专业学生评分后的平均值，属于多评分者的平均结果。

成本型数据（生产成本，重量）是可直接测量的物理或成本数据，无需人工打分。

TOPSIS示例数据

第一行：6项指标的权重值。不需要指标的名称。

第二行：指标类型标识（1表示效益型，0表示成本型），续航时间、外观吸引力、操作便捷性、功能丰富度为效益型（1），生产成本、重量为成本型（0）。

第三行及以后：5套设计方案的原始数据。

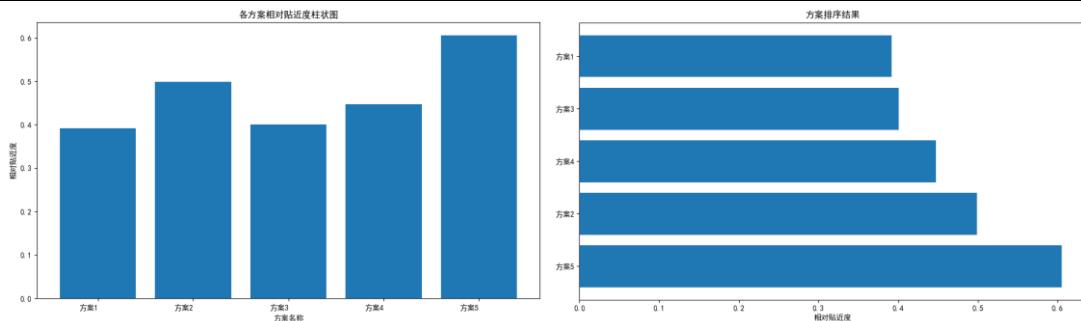
| | | | | | | |
|------|------|-----|------|-----|-----|-----|
| 权重 | 0.25 | 0.2 | 0.15 | 0.1 | 0.2 | 0.1 |
| 指标类型 | 1 | 1 | 1 | 1 | 0 | 0 |
| 方案1 | 72 | 8.5 | 7 | 9 | 180 | 35 |
| 方案2 | 96 | 7 | 8.5 | 8 | 220 | 40 |
| 方案3 | 60 | 9 | 9 | 7 | 160 | 30 |
| 方案4 | 84 | 8 | 8 | 9 | 200 | 38 |
| 方案5 | 120 | 6.5 | 7.5 | 10 | 240 | 45 |

分析结果

| | |
|-----------|---|
| 标准化决策矩阵 | 0.3625, 0.4839, 0.3898, 0.4648, 0.3985, 0.4127 0.4833, 0.3985, 0.4733, 0.4131, 0.4871, 0.4716 0.3021, 0.5124, 0.5012, 0.3615, 0.3542, 0.3537 0.4229, 0.4555, 0.4455, 0.4648, 0.4428, 0.448 0.6041, 0.3701, 0.4176, 0.5164, 0.5314, 0.5306 |
| 加权标准化决策矩阵 | 0.0906, 0.0968, 0.0585, 0.0465, 0.0797, 0.0413 0.1208, 0.0797, 0.071, 0.0413, 0.0974, 0.0472 0.0755, 0.1025, 0.0752, 0.0361, 0.0708, 0.0354 0.1057, 0.0911, 0.0668, 0.0465, 0.0886, 0.0448 0.151, 0.074, 0.0626, 0.0516, 0.1063, 0.0531 |

| 正理想解 | 负理想解 | 各方案到正理想解的距离 | 各方案到负理想解的距离 | 各方案的相对贴近度 | 方案排序结果 |
|--------|--------|-------------|-------------|-------------|-------------|
| 0.151 | 0.0755 | 方案1: 0.064 | 方案1: 0.0412 | 方案1: 0.3916 | 方案1 (排名: 5) |
| 0.1025 | 0.074 | 方案2: 0.049 | 方案2: 0.0488 | 方案2: 0.4991 | 方案2 (排名: 2) |

| | | | | | |
|--------|--------|--------------|--------------|--------------|--------------|
| 0.0752 | 0.0585 | 方案 3: 0.0771 | 方案 3: 0.0515 | 方案 3: 0.4007 | 方案 3 (排名: 4) |
| 0.0516 | 0.0361 | 方案 4: 0.0518 | 方案 4: 0.042 | 方案 4: 0.4477 | 方案 4 (排名: 3) |
| 0.0708 | 0.1063 | 方案 5: 0.0503 | 方案 5: 0.0772 | 方案 5: 0.6053 | 方案 5 (排名: 1) |
| 0.0354 | 0.0531 | | | | |



结论

方案 5 > 方案 2 > 方案 4 > 方案 3 > 方案 1。

6.5 综合评价（诊断）

综合评价（诊断）包括“耦合协调度模型、非期望 SBM 模型和障碍度模型”三种方法。

综合评价（诊断）和综合评价（排名）的异同点

- 它们都处理多指标数据，用于“综合评价”。
- 综合评价（诊断）更聚焦“系统关系”和“深层问题”，它更像“诊断工具”，看系统配合、算真实效率、找问题根源。
- 综合评价（排名）更聚焦“直接排序”和“简化指标”，它更像“排名工具”，直接比高低、排先后、简化指标。
- 简单说：想知道“谁更好”，用“综合评价（排名）”，想知道“为什么不好”“配合得怎么样”，用“综合评价（诊断）”。

耦合协调协调度模型

分析“多个系统是否配合得好”，它不仅看单个系统的好坏，更看系统之间是否“同步发展、相互促进”，比如“经济发展”和“环境保护保护”两个系统：是经济好但污染严重（配合差），还是两者经济和环保同步提升（配合好）？核心是衡量系统之间的“协同程度”，结果告诉你“系统间是否良性互动”。

非期望 SBM 模型

分析“做事的效率，还要扣掉副作用”。比如评价工厂：不仅看产量（好结果），还要扣掉污染（坏结果），算“净效率”。评价 APP：不仅看活跃用户（好结果），还要扣掉投诉量（坏结果）。核心是“在考虑代价 /副作用的前提下，评谁的效率更高”，并且给出具体的优化方案。

障碍度模型

用于评估设计方案中各因素对方案落地的阻碍程度，也就是寻找“拖后腿的关键因素”，比如某城市发展差，是“交通”，“教育”和“产业”哪个因素拖了后腿？某产品用户满意度低，是“价格”，“质量”和“售后”哪个环节最拉垮？核心是定位“短板”，告诉我们“该优先改进什么”。

三种方法的异同点

| 维度 | 耦合协调度模型 | 非期望 SBM 模型 | 障碍度模型 |
|------|-----------------|-------------------------------|---------------------------------|
| 核心关注 | 系统之间的“配合程度” | 带副作用的“效率高低” | 影响整体表现的“关键短板” |
| 典型场景 | 经济与环保、设计与体验的同步性 | 工厂生产（产量+污染）、APP 运营（用户+投诉） | 城市发展瓶颈、产品的差评原因分析 |
| 解决问题 | 回答“多个系统是否协调” | 回答“效率有多高”“如何达到最优”（提供改进的“目标值”） | 回答“为什么效率低”“先改哪项最有效”（提供改进的“优先级”） |

- 想知道“多个系统是否配合好”，用耦合协调度模型。
- 想知道“做事效率（含副作用）”，用非期望 SBM 模型。
- 想知道“该优先改什么”，用障碍度模型。

三者常结合使用：比如先看两个系统是否协调（耦合协调度），再算整体效率（非期望 SBM），最后找拖后腿的因素（障碍度），形成完整的“评价 - 优化”闭环。

耦合协调度模型示例数据

第二行数据为系统标识（0、1、2、3 等整数，相同系统的指标用相同数字标识），下表的数据中，前三个指标为“硬件设计系统”，后三个指标为“用户体验系统”。

| | A | B | C | D | E | F |
|---|--------|-------|-------|-------|-------|-------|
| 1 | 外观美学设计 | 功能完整性 | 材质耐用性 | 操作便捷性 | 使用舒适度 | 功能实用性 |
| 2 | 0 | 0 | 0 | 1 | 1 | 1 |
| 3 | 8.5 | 9.2 | 8.8 | 7.9 | 8.3 | 9 |
| 4 | 7.2 | 8 | 8.5 | 8.8 | 7.5 | 7.8 |
| 5 | 9 | 8.5 | 9.1 | 8.2 | 8 | 8.7 |
| 6 | 8.8 | 7 | 7.5 | 6.5 | 7.2 | 6.9 |

分析结果

样本 1

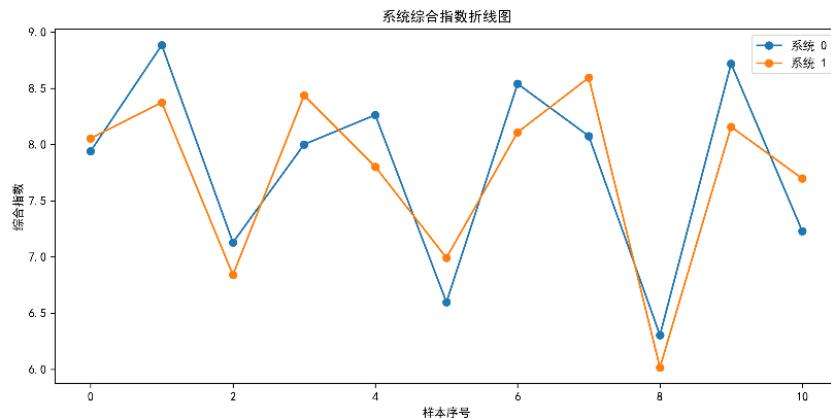
耦合度 1.0

耦合协调度 2.828 Extremely coordinated

系统 0 综合指数 7.9407

系统 1 综合指数 8.0553

...



结论

- 整体结论：11个样本的“双系统”表现都很好，既相互关联（耦合度高），又发展匹配（协调度高，全是“极度协调”），没有“偏科”的情况。
- 细节差异：
 - 表现最好的样本：样本2（耦合协调度 2.937，系统0得分最高）、样本8（系统1得分最高，协调度 2.8864）
 - 表现相对弱的样本：样本9（两个系统得分最低，协调度 2.4816，是“极度协调”里的下限）
- 图表辅助：文档最后两张图是“分布直方图”（虽然显示有点乱），核心是看数值分布，耦合度全集中在 1.0 附近，耦合协调度集中在 2.4-2.9 之间，说明所有样本的表现很集中，没有“不协调”的异常值。

11个样本的两个系统（比如硬件和体验）不仅“关系好”，而且“水平匹配”，没有出现“设计好但用着差”的问题，只是部分样本整体水平略高或略低而已。

非期望 SBM 模型 示例数据

表头需要加上“前缀”，也就是“Input_”，“Output_”和“UndesirableOutput_”

| DMU 名称 | Input_劳动力 | Input_资本 | Output_产值 | UndesirableOutput_污染 |
|--------|-----------|----------|-----------|----------------------|
| A 厂 | 100 | 200 | 300 | 50 |
| B 厂 | 150 | 300 | 400 | 50 |
| C 厂 | 80 | 200 | 200 | 70 |
| D 厂 | 160 | 250 | 300 | 100 |

参数设置

- 新手入门：**建议先尝试“投入导向 + 可变规模报酬”，这是最常用的组合，能覆盖多数实际场景，且结果解释性较强。
- 根据研究目标调整
关注“资源节约”则选“投入导向”，关注“产出提升”则选“产出导向”。
若研究对象规模差异大（如大 / 中 / 小企业），选“可变规模报酬”；若规模相近，选“不变规模报酬”。
- 投入导向 (Input Oriented)

目标：在保持产出（包括期望产出和非期望产出）不变的前提下，尽可能减少投入资源（如成本、人力等）。

适用场景：当投入资源较为稀缺或成本较高时（例如减少生产中的原材料消耗）。

- 产出导向（Output Oriented）

目标：在投入资源固定的前提下，尽可能增加期望产出（如利润、产量），同时减少非期望产出（如污染物、浪费）。

适用场景：当投入相对固定，需要最大化有效产出时（例如在固定预算下提升产品质量）。

- 不变规模报酬（Constant Returns to Scale, CRS）

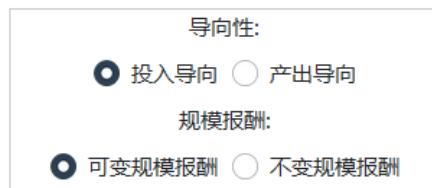
假设：决策单元的效率不受规模大小影响（投入增加一倍，产出也增加一倍）。

适用场景：适用于规模较为稳定的情况，或假设所有决策单元都在最优规模下运行（如成熟行业的企业效率对比）。

- 可变规模报酬（Variable Returns to Scale, VRS）

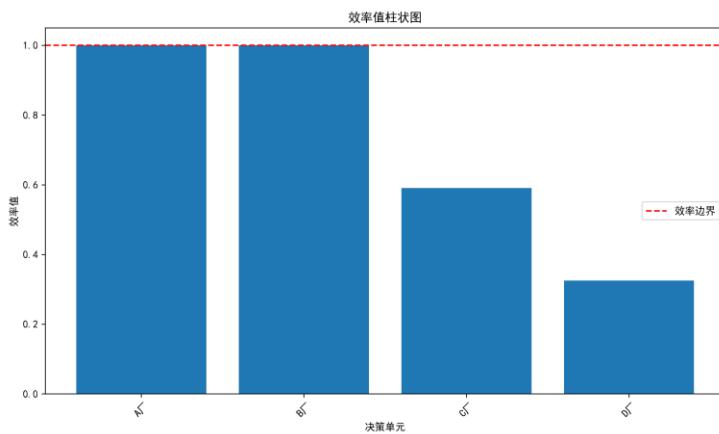
假设：效率会受规模影响（可能存在规模经济或规模不经济）。

适用场景：适用于不同规模的决策单元对比（如不同城市、不同规模企业的效率分析），能更真实反映实际运营中的规模差异。



分析结果

| 决策单元 | 效率值 |
|------|--------|
| A 厂 | 1.0 |
| B 厂 | 1.0 |
| C 厂 | 0.5908 |
| D 厂 | 0.3249 |



低效的决策单元分析

- C 厂（改进建议）

- 应减少输入: {'Input_劳动力': '0.0 (原始值: 80)', 'Input_资本': '0.0409 (原始值: 200)'}
- 应增加理想产量: {'Output_产值': '0.0 (原始值: 200)'}
- 应减少不良输出: {'UndesirableOutput_污染': '0.4418 (原始值: 70)'}
- D 厂 (改进建议)
 - 应减少输入: {'Input_劳动力': '0.0 (原始值: 160)', 'Input_资本': '0.0788 (原始值: 250)'}
 - 应增加理想产量: {'Output_产值': '0.0 (原始值: 300)'}
 - 应减少不良输出: {'UndesirableOutput_污染': '0.9 (原始值: 100)'}

结论

- A、B 厂是“榜样”，其他工厂可以参考它们的投入 / 产出比例（比如同样的劳动力和资本，A 厂能做到“少污染、高产值”）；
- C 厂和 D 厂的问题主要在“污染过多”，C 厂需要减少污染 0.4418，D 厂需要减少污染 0.9，减少污染后效率会大幅提升。

障碍度模型 示例数据

通过专家评分 (AHP, 或其他计算权重的方法) 确定了各因素的权重和当前状态值。

- 权重: 表示该因素对方案成功的重要性 (值越高越重要)。
- 状态值: 表示该因素当前的实现程度 (值越高越接近目标)。

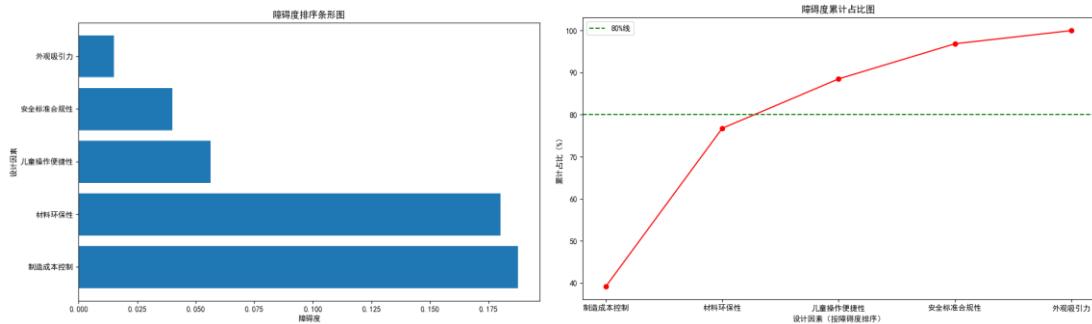
| 设计因素 | 权重 (0-1) | 状态值 (0-1) |
|---------|----------|-----------|
| 材料环保性 | 0.3 | 0.6 |
| 制造成本控制 | 0.25 | 0.4 |
| 安全标准合规性 | 0.2 | 0.8 |
| 儿童操作便捷性 | 0.15 | 0.5 |
| 外观吸引力 | 0.1 | 0.7 |

分析结果

| 因素名称 | 障碍度 | 百分比贡献 | 累计贡献 |
|---------|--------|--------|---------|
| 制造成本控制 | 0.1875 | 39.16% | 39.16% |
| 材料环保性 | 0.18 | 37.60% | 76.76% |
| 儿童操作便捷性 | 0.0562 | 11.75% | 88.51% |
| 安全标准合规性 | 0.04 | 8.36% | 96.87% |
| 外观吸引力 | 0.015 | 3.13% | 100.00% |

关键障碍因素 (累计贡献 $\geq 80\%$): 制造成本控制, 材料环保性, 儿童操作便捷性

| 因素名称 | 障碍度 |
|---------|--------|
| 材料环保性 | 0.18 |
| 制造成本控制 | 0.1875 |
| 安全标准合规性 | 0.04 |
| 儿童操作便捷性 | 0.0562 |
| 外观吸引力 | 0.015 |



障碍度：衡量设计因素对方案实施的阻碍程度（设计专用公式：权重 \times (1-状态值) \times 影响系数），障碍度越高，设计因素对方案实施的阻碍作用越大

障碍度条形图：显示各设计因素的障碍度分布，直观比较不同设计因素的障碍度

障碍度排序条形图：按障碍度从高到低显示设计因素，有助于快速识别主要障碍因素

累积比例图：显示障碍度的累积比例，可识别关键的少数障碍因素（帕累托原则）

结论

影响设计方案落地的主要障碍是“制造成本控制”和“材料环保性”，先解决这两个，方案落地会顺利很多。

6.6 关键因素（极差分析）

从一堆实验数据里，找出哪些因素对结果影响大，哪些影响小，以及每个因素选哪种状态（水平）效果最好。

比如，一个烤蛋糕的实验，自变量有“烤箱温度”，“烘烤时间”和“面粉种类”，因变量是“蛋糕评分”，极差分析可以告诉我们，

- 是温度对评分影响最大（比如 180°C 和 200°C 烤出来的评分差很多），还是时间影响更大；
- 温度选 180°C、时间选 25 分钟、面粉用低筋的，可能烤出评分最高的蛋糕。

自变量必须是分类数据，且分类不能过多（通常 2-5 个），因变量必须是连续数据，不要求满足正太分布等。

示例数据

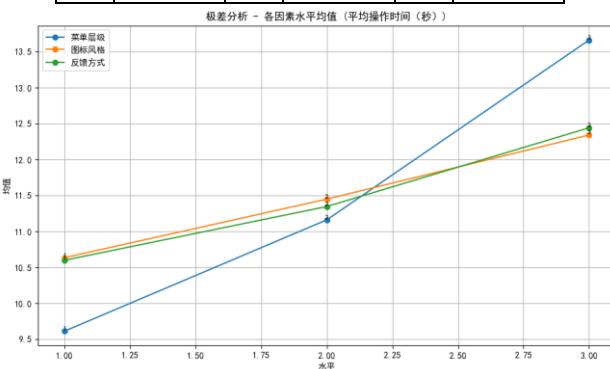
选取 3 个关键设计因素，每个因素设置 3 个水平：

- 因素 A（菜单层级）：2 层（1）、3 层（2）、4 层（3）
- 因素 B（图标风格）：线性图标（1）、拟物图标（2）、扁平化图标（3）
- 因素 C（反馈方式）：震动反馈（1）、声光反馈（2）、无反馈（3）

| | A | B | C | D |
|---|------|------|------|------------|
| 1 | 菜单层级 | 图标风格 | 反馈方式 | 平均操作时间 (秒) |
| 2 | 1 | 1 | 1 | 8.2 |
| 3 | 1 | 2 | 2 | 9.5 |
| 4 | 1 | 3 | 3 | 11.3 |
| 5 | 2 | 1 | 2 | 10.1 |
| 6 | 2 | 2 | 3 | 12.4 |

分析结果

| 等级 | 平均值 | 等级 | 平均值 | 等级 | 平均值 |
|----|-----------|----|-----------|----|-----------|
| 1 | 9.614286 | 1 | 10.635714 | 1 | 10.6 |
| 2 | 11.161538 | 2 | 11.45 | 2 | 11.346667 |
| 3 | 13.661538 | 3 | 12.341667 | 3 | 12.441667 |
| 范围 | 4.047253 | 范围 | 1.705952 | 范围 | 1.841667 |



结论

- 优先改菜单层级：把现在的 3 层 / 4 层菜单，改成 2 层（常用功能点 1-2 下就能到），能让操作时间直接减少 4 秒，用户会明显觉得“变快了”；
- 其次选震动反馈：操作后加震动提示（别用无反馈或声光反馈），让用户不用等、不用看，知道操作成功；
- 图标风格选线性的：不用花太多成本改，但选线性图标比扁平化图标能快 1.7 秒，属于“低成本提升体验”。

6.7 预测未来

“预测未来”指的是通过分析过去一段时间的数据（如每月销量、年度污染量），推测未来某一时刻的数值或状态。DIAS 提供了两种方法，“灰色预测模型和指数平滑法”。

两种方法的异同点

| 对比维度 | 灰色预测模型 | 指数平滑法 |
|------|---------------|----------------------------|
| 数据量 | 很少（3-5 个就行） | 较多（至少 10 个以上连续数据） |
| 预测逻辑 | 找少量数据里的“隐藏规律” | 给近期数据“更高权重”（想让“最近的数据影响更大”） |

| | | |
|------|----------------------|-----------------------|
| 适合场景 | 小样本、数据少的短期趋势（如新产品销量） | 连续数据的短期平稳趋势（如日常销量、库存） |
| 缺点 | 数据多了反而不准 | 遇到突发变化（如促销）会失灵 |

怎么选？

- 数据少、新业务：选灰色预测；
- 数据多、关注近期趋势：选指数平滑法；

灰色预测模型示例数据

某设计公司想优化智能手环的界面设计，但用户满意度的数据收集成本高、周期长。他们发现近几个月通过用户调研得到的满意度评分有一定变化规律，希望用少量历史数据预测未来3个月的满意度趋势，提前判断是否需要调整设计方向（比如按钮布局、数据可视化方式等），避免等到用户大量流失再改。

- 方式：每月随机抽取50名用户，用1-10分评分量表（10分最高）收集对当前手环界面的满意度，取每月的平均分。
- 数据来源：过去4个月的用户调研数据。

| |
|-----|
| 7.2 |
| 7.5 |
| 7.8 |
| 8.1 |

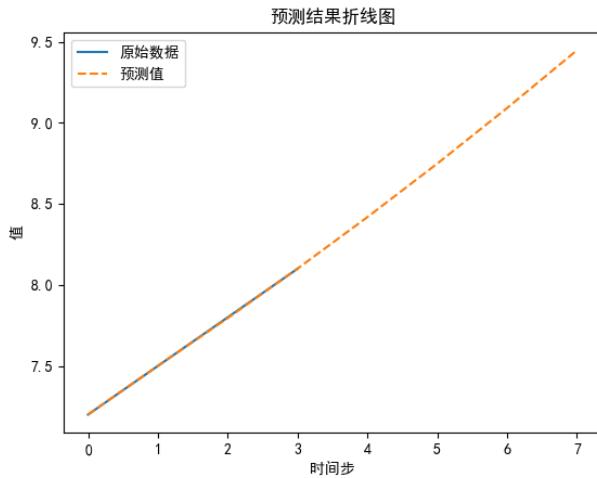
分析步骤

选择预测多少年后的未来

请输入预测步数（正整数）：

分析结果

| | |
|----------|--|
| 原始数据 | [7.2, 7.5, 7.8, 8.1] |
| 累加生成序列 | [7.2, 14.7, 22.5, 30.6] |
| 预测值 | (7.19999999999989), (7.50105568100463), (7.79514080490128), (8.100755780564867), (8.418352645419134), (8.74840115939071), (9.091389499734987), (9.447824983102066)] |
| 后验差比(C) | 0.0071 |
| 小误差概率(P) | 1.0000 |



结论

- 趋势结论：过去 4 个月满意度稳步上升 ($7.2 \rightarrow 8.1$)，预测未来 4 个月会继续涨到 9.4，整体是“持续向好”的趋势；
- 可靠性结论：模型算出来的结果很可信 ($C=0.0071$ 、 $P=1.0$)，不是随便猜的；

指数平滑法 示例数据

| |
|-----|
| 7.2 |
| 7.5 |
| 7.3 |
| 7.8 |
| 8.0 |
| 7.9 |
| 8.2 |
| 8.5 |
| 8.3 |
| 8.6 |
| 8.8 |
| 9.0 |

分析过程

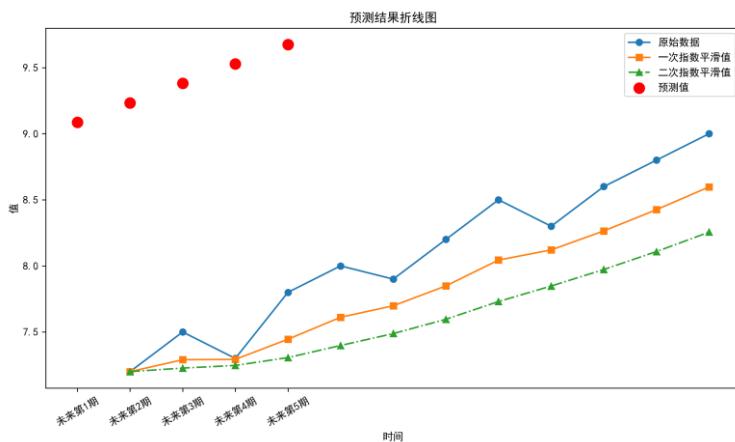
- 平滑系数的取值含义
平滑系数 α 越接近 1，预测结果对近期数据的反应越敏感（更重视新数据），适合数据波动较大（如短期趋势明显）的情况。
越接近 0，预测结果越平滑（更重视历史数据的累积影响），适合数据较平稳（如长期趋势稳定）的情况。
- 默认值为 0.3，适合多数一般性场景，若希望预测更关注近期数据，可填写 0.7，若希望结果更平滑，可填写 0.2。

| | |
|-------------|-----|
| 平滑系数 (0-1): | 0.3 |
| 预测未来时间点数量: | 5 |

分析结果

| 时间 | 一次指数平滑值 | 二次指数平滑值 |
|----|---------|---------|
| 1 | 7.2000 | 7.2000 |
| 2 | 7.2900 | 7.2270 |
| 3 | 7.2930 | 7.2468 |
| 4 | 7.4451 | 7.3063 |
| 5 | 7.6116 | 7.3979 |
| 6 | 7.6981 | 7.4879 |
| 7 | 7.8487 | 7.5962 |
| 8 | 8.0441 | 7.7305 |
| 9 | 8.1208 | 7.8476 |
| 10 | 8.2646 | 7.9727 |
| 11 | 8.4252 | 8.1085 |
| 12 | 8.5977 | 8.2552 |

| 预测时间点 | 预测值 |
|---------|--------|
| 未来第 1 期 | 9.0868 |
| 未来第 2 期 | 9.2336 |
| 未来第 3 期 | 9.3803 |
| 未来第 4 期 | 9.5271 |
| 未来第 5 期 | 9.6739 |



结论

- 趋势结论：过去 12 个月数据稳步上升，未来 5 个月会继续涨，从 9.08 涨到 9.67，涨幅稳定。

6.8 DEMATEL 分析

它可以分析因素间的影响关系（谁影响谁、影响多强），同时计算重要性。

它的数据矩阵不是对称矩阵（上三角和下三角元素互为倒数）。

数据收集

方法：让专业人士根据经验对因素间的影响程度打分。

- **第一步：明确要研究哪些要素**

影响 UI 设计满意度的关键因素：F1 = 视觉美观度、F2 = 操作便捷性、F3 = 功能完整性、F4 = 信息层级清晰度、F5 = 交互反馈及时性、F6 = 兼容性、F7 = 个性化设置、F8 = 响应速度

这些因素通常来自文献研究（查论文，看别人常用哪些因素），实际需求（比如用户反馈中频繁提到的问题）和专家经验（让设计师、产品经理列出核心要素）。

- **第二步：设计“评分表”（让专家填的表格）**

制作一个“因素影响评分表”，格式如下（以 F1 到 F8 为例）：

| 影响方\被影响方 | F1 视觉美观度 | F2 操作便捷性 | ... | F8 响应速度 |
|----------|----------|----------|-----|---------|
| F1 视觉美观度 | (填 0) | ? | ... | ? |
| F2 操作便捷性 | ? | (填 0) | ... | ? |
| ... | ... | ... | ... | ... |
| F8 响应速度 | ? | ? | ... | (填 0) |

填写规则：表格中第“行因素”对第“列因素”的影响程度，用 1-5 分打分，1 分 = 几乎无影响，2 分 = 轻微影响，3 分 = 一般影响，4 分 = 较强影响，5 分 = 极强影响，对角线位置（因素对自身的影响）必须填 0（自己不影响自己）。

- **第三步：找专家打分（收集原始数据）**

邀请 3-10 位相关专家（比如 UI 设计师、用户体验研究员、资深产品经理），让他们根据经验独立填写上面的评分表。

举个例子：专家判断“F3 功能完整性”对“F2 操作便捷性”影响很强，就在“F3 行、F2 列”填 4 分；

- **第四步：计算平均分（得到最终数据）**

如果有 3 位专家打分，同一位置的分数可能不同，需要取平均值作为最终数据。

比如“F1 对 F2 的影响”：专家 1 打 3 分，专家 2 打 4 分，专家 3 打 2 分，平均分 $= (3+4+2) \div 3 = 3$ 分，最终表格中“F1 行、F2 列”就填 3 分（对应案例中第一行第二列的“3”）。

示例数据

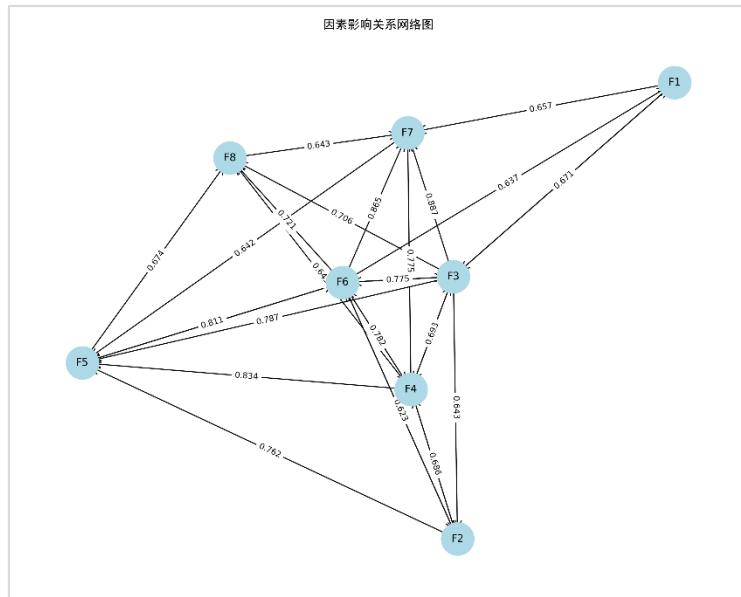
| F | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 |
|----|----|----|----|----|----|----|----|----|
| F1 | 0 | 3 | 4 | 2 | 1 | 0 | 2 | 1 |
| F2 | 2 | 0 | 1 | 3 | 4 | 2 | 1 | 0 |
| F3 | 3 | 2 | 0 | 1 | 2 | 3 | 4 | 2 |
| F4 | 1 | 3 | 2 | 0 | 3 | 4 | 2 | 1 |
| F5 | 0 | 2 | 1 | 2 | 0 | 3 | 1 | 4 |
| F6 | 2 | 1 | 3 | 4 | 2 | 0 | 3 | 2 |
| F7 | 1 | 0 | 2 | 1 | 3 | 2 | 0 | 3 |
| F8 | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 0 |

分析结果

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | 因素 | 原因度 | 中心度 |
|--|----|----|----|----|----|----|----|----|----|-----|-----|
|--|----|----|----|----|----|----|----|----|----|-----|-----|

| | | | | | | | | |
|----|--------|--------|--------|--------|--------|--------|--------|--------|
| F1 | 0.4211 | 0.5968 | 0.6799 | 0.5575 | 0.6161 | 0.5691 | 0.657 | 0.5242 |
| F2 | 0.5068 | 0.453 | 0.5436 | 0.6328 | 0.7622 | 0.6792 | 0.5997 | 0.4961 |
| F3 | 0.6714 | 0.6433 | 0.609 | 0.6281 | 0.7872 | 0.8337 | 0.8873 | 0.7056 |
| F4 | 0.56 | 0.6856 | 0.6931 | 0.5806 | 0.8341 | 0.8891 | 0.7751 | 0.6451 |
| F5 | 0.4273 | 0.5388 | 0.5255 | 0.5608 | 0.5431 | 0.711 | 0.6067 | 0.6737 |
| F6 | 0.637 | 0.6233 | 0.7748 | 0.7825 | 0.8113 | 0.7239 | 0.8652 | 0.721 |
| F7 | 0.439 | 0.4061 | 0.5415 | 0.4642 | 0.6417 | 0.615 | 0.5109 | 0.6068 |
| F8 | 0.5274 | 0.4908 | 0.4962 | 0.4168 | 0.5442 | 0.5825 | 0.6428 | 0.4239 |

| | | |
|----|---------|---------|
| F1 | 0.4317 | 8.8119 |
| F2 | 0.2356 | 9.1113 |
| F3 | 0.9021 | 10.6292 |
| F4 | 1.0394 | 10.286 |
| F5 | -0.9528 | 10.1268 |
| F6 | 0.3354 | 11.5426 |
| F7 | -1.3194 | 9.7699 |
| F8 | -0.6719 | 8.9211 |



结论

- 优先优化原因因素：先搞定 F6（兼容性）、F3（功能完整性）、F4（信息层级），这三个因素能带动其他因素变好；
- 不用单独优化结果因素：F5（交互反馈）、F7（个性化）、F8（响应速度）会随着原因因素的优化自然提升，不用额外投入；
- 视觉美观度放在最后：F1（视觉）重要性最低，等核心功能、兼容性都做好了，再调整美观度即可。

7 创造一个新理论

7.1 线性回归

线性回归的用途是分析“哪些因素影响了结果、影响有多大”，以及“未来可能会出现什么结果”，它把自变量和因变量的关系总结为一个数学公式。

DIAS 中有五种线性回归方法，“OLS，Robust，岭回归，Lasso 和 PLS”。

因变量必须为连续数据，如：收入、温度和销量等。自变量可以是连续数据，也可以是分类数据，若是分类数据，需要由 0, 1, 2 等数据代表分类数据。

核心模型形式为 $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$ ，如：收入 = $2000 + 500 \times \text{年龄} + 1000 \times \text{工作年限} + 3000 \times \text{教育水平}$

五种方法的异同点

| 方法 | 数据特征 | 典型场景举例 |
|--------|--|---|
| OLS | 数据符合正态、无异常值、无共线性 | 用年龄、工作年限预测收入 |
| Robust | 数据存在极端值，允许不符合正态分布 | 房价预测 |
| 岭回归 | 允许不符合正态分布、允许有共线性，保留所有变量，数据量较多时 (>20) | 用 GDP、GNP、居民消费指数等相关指标预测通胀率 |
| Lasso | 允许不符合正态分布，允许有共线性，自变量多，希望简化模型（剔除冗余变量） | 用 100 个基因指标预测疾病风险，筛选关键基因 |
| PLS | 允许不符合正态分布，允许有共线性，样本数量（行）远少于自变量数量（列） | 用“20 个产品的销量”去分析“1000 个设计参数的作用”，变量比要分析的对象还多。 |

数据收集

外观设计的多个维度（如形态复杂度、色彩饱和度、材质质感等）会通过用户的视觉感知与情感体验间接作用于购买意愿。

- 样本选择
选取 100 款市售智能音箱作为研究样本，覆盖不同品牌（如小米、华为、苹果等）、价格区间（200–2000 元）及设计风格（简约、科技、复古等），确保样本多样性。
- 变量定义
因变量 (Y): 用户购买意愿通过问卷调查收集，采用 7 分李克特量表 (1=“完全不愿意购买”，7=“非常愿意购买”)，取每个样本的平均得分作为量化值。
自变量 (X): 外观设计要素（选取 4 个核心维度）
 X_1 : 形态复杂度（由设计专家团队评分，1=“极简”，5=“极复杂”）；
 X_2 : 色彩饱和度（通过 PS 提取 RGB 值计算，0=“黑白灰”，1=“高饱和彩色”）；
 X_3 : 材质质感 (1=“塑料”，2=“金属”，3=“木质 / 布艺”)；
 X_4 : 界面交互面积（屏幕 / 触控区域占产品表面积的比例，单位: %）。
Y: 购买意愿数据（针对每款产品，随机抽取 20 名潜在用户进行调查，取平均值）。

OLS 示例数据

数据的第一列是“因素名称”，最后一列是“因变量”。

| | A | B | C | D | E | F |
|---|------|---------|---------|--------|----------|-------|
| 1 | 产品ID | 形态复杂度X1 | 色彩饱和度X2 | 材质质感X3 | 界面交互面积X4 | 购买意愿Y |
| 2 | 1 | 2 | 0.3 | 2 | 15 | 4.8 |
| 3 | 2 | 3 | 0.7 | 1 | 30 | 3.6 |
| 4 | 3 | 1 | 0.1 | 3 | 5 | 5.9 |
| 5 | 4 | 2 | 0.5 | 3 | 20 | 5.2 |
| 6 | 5 | 4 | 0.0 | 1 | 10 | 2.0 |

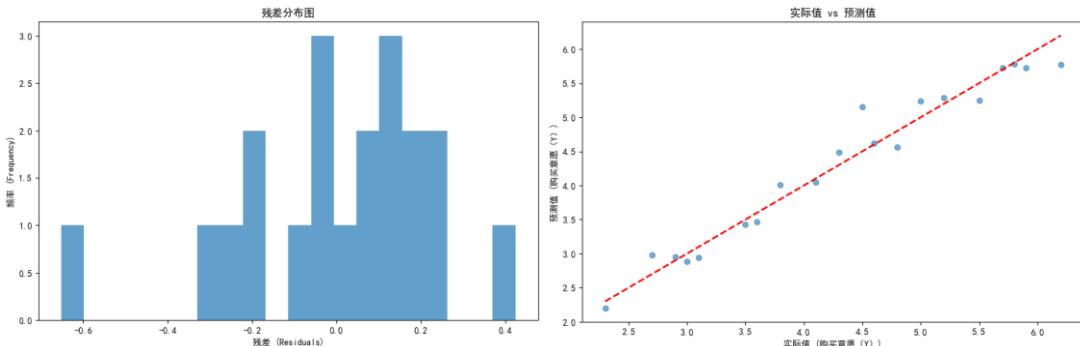
分析结果

购买意愿 Y = 4.3641 - 0.6687*形态复杂度 X1 - 0.0596*色彩饱和度 X2 + 0.6514*材质质感 X3 + 0.0166*界面交互面积 X4

| 变量名称 | 系数 | t 值 | p 值 | 95%置信区间下限 | 95%置信区间上限 |
|-----------|---------|---------|--------|-----------|-----------|
| 形态复杂度 X1 | -0.6687 | -4.0982 | 0.0009 | -1.0164 | -0.3209 |
| 色彩饱和度 X2 | -0.0596 | -0.1989 | 0.8450 | -0.6982 | 0.5790 |
| 材质质感 X3 | 0.6514 | 5.4923 | 0.0001 | 0.3986 | 0.9042 |
| 界面交互面积 X4 | 0.0166 | 0.8665 | 0.3999 | -0.0242 | 0.0574 |

| 统计量 | 值 |
|----------------|---------|
| 截距 (Intercept) | 4.3641 |
| 均方误差 (MSE) | 0.0529 |
| 决定系数 (R^2) | 0.9601 |
| 调整决定系数 | 0.9495 |
| F 统计量 | 90.3500 |
| F 统计量 p 值 | 0.0000 |

| 来源 | df | sum_sq | mean_sq | F | PR(>F) |
|-----------|---------|---------|---------|----------|--------|
| 形态复杂度 X1 | 1.0000 | 23.2445 | 23.2445 | 329.4444 | 0.0000 |
| 色彩饱和度 X2 | 1.0000 | 0.0996 | 0.0996 | 1.4117 | 0.2532 |
| 材质质感 X3 | 1.0000 | 2.1021 | 2.1021 | 29.7930 | 0.0001 |
| 界面交互面积 X4 | 1.0000 | 0.0530 | 0.0530 | 0.7508 | 0.3999 |
| Residual | 15.0000 | 1.0583 | 0.0706 | nan | nan |



- Coefficients: 回归系数，表示每个自变量对因变量的影响程度。
- Intercept: 截距，是当所有自变量为 0 时因变量的预测值。
- Mean Squared Error (MSE): 均方误差，衡量预测值与真实值之间的平均误差。
- R-squared (R^2): 决定系数，取值范围在 0 到 1 之间，越接近 1 表示模型拟合效果越好。
- Adjusted R-squared: 调整决定系数，考虑了模型中自变量的数量，对模型的拟合优度进行了调整。
- F-value: F 统计量，用于检验整个回归模型的显著性。
- t-value: t 统计量，用于检验每个自变量的显著性。

- p-value: p 值, 用于判断自变量的显著性, p 值越小, 自变量越显著。通常 p < 0.05 认为显著。

结论

优先简化形态: 别做太复杂的外观 (比如少加多余的线条、凸起), 每简化 1 级, 用户的购买意愿能提升约 0.67 分;

重点升级材质: 把塑料换成金属或木质, 每升级 1 级, “想买度”能提升约 0.65 分;

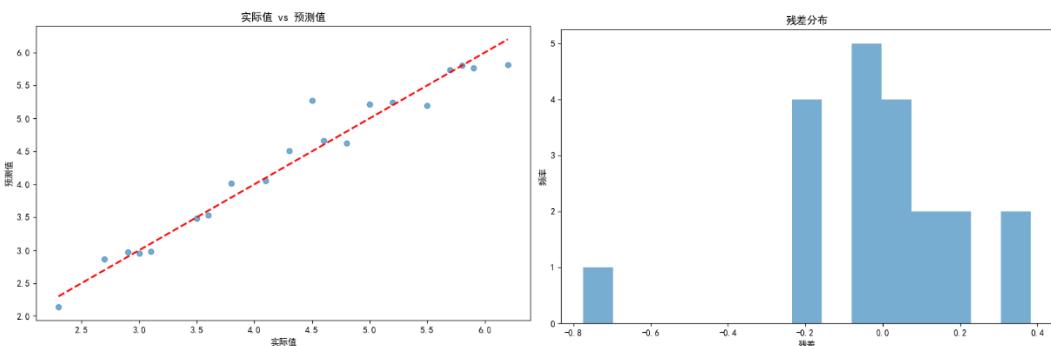
不用纠结颜色和界面大小: 不管用高饱和色还是低饱和色、界面做 15% 还是 30% 面积, 对购买意愿几乎没影响, 不用在这两项上浪费成本。

Robust 示例数据

分析结果

购买意愿 $Y = 4.7287 - 0.7265 * \text{形态复杂度 } X_1 - 0.1458 * \text{色彩饱和度 } X_2 + 0.5634 * \text{材质质感 } X_3 + 0.0175 * \text{界面交互面积 } X_4$

| 指标 | 值 | 解释 |
|-----------------------|---------|------------------------------------|
| 截距 | 4.7287 | 截距, 即当所有自变量都为 0 时, 因变量的预测值。 |
| 均方误差 (MSE) | 0.0560 | 均方误差, 测量预测值和实际值之间的平均误差。 |
| R 平方 (R^2) | 0.9579 | 判定系数, 范围从 0 到 1。值越接近 1, 表示模型拟合得越好。 |
| 调整后的 R 平方 | 0.9849 | 调整判定系数, 考虑独立变量的数量和稳健回归特性。 |
| 系数(形态复杂度 (X_1)) | -0.7265 | 回归系数, 表示各自变量对因变量的影响。 |
| t 值(形态复杂度 (X_1)) | -5.4034 | t 统计量, 用于检验每个独立变量的显著性。 |
| p 值(形态复杂度 (X_1)) | 0.0000 | p 值, 用于判断自变量的显著性。p 值越小, 自变量越显著。 |
| 系数(色彩饱和度 (X_2)) | -0.1458 | 回归系数, 表示各自变量对因变量的影响。 |
| t 值(色彩饱和度 (X_2)) | -0.5904 | t 统计量, 用于检验每个独立变量的显著性。 |
| p 值(色彩饱和度 (X_2)) | 0.5550 | p 值, 用于判断自变量的显著性。p 值越小, 自变量越显著。 |
| 系数(材质质感 (X_3)) | 0.5634 | 回归系数, 表示各自变量对因变量的影响。 |
| t 值(材质质感 (X_3)) | 5.7636 | t 统计量, 用于检验每个独立变量的显著性。 |
| p 值(材质质感 (X_3)) | 0.0000 | p 值, 用于判断自变量的显著性。p 值越小, 自变量越显著。 |
| 系数(界面交互面积 (X_4)) | 0.0175 | 回归系数, 表示各自变量对因变量的影响。 |
| t 值(界面交互面积 (X_4)) | 1.1085 | t 统计量, 用于检验每个独立变量的显著性。 |
| p 值(界面交互面积 (X_4)) | 0.2676 | p 值, 用于判断自变量的显著性。p 值越小, 自变量越显著。 |

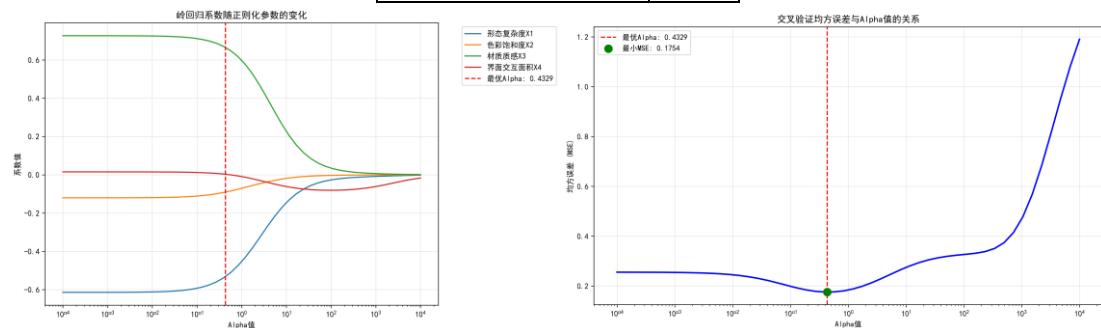


岭回归示例数据

分析结果

购买意愿 $Y = 4.2581 - 0.5318 \times \text{形态复杂度 } X_1 - 0.0905 \times \text{色彩饱和度 } X_2 + 0.6654 \times \text{材质质感 } X_3 + 0.0029 \times \text{界面交互面积 } X_4$

| 指标 | 值 |
|---------------------------|---------|
| 最优 Alpha | 0.2477 |
| 截距 | 4.4397 |
| Train MSE | 0.0610 |
| Test MSE | 0.0404 |
| Train R-squared (R^2) | 0.9575 |
| Test R-squared (R^2) | 0.9547 |
| Train Adjusted R-squared | 0.9420 |
| Test Adjusted R-squared | 1.1359 |
| F-value | 14.6999 |
| 形态复杂度 X_1 系数 | -0.6134 |
| ... | ... |

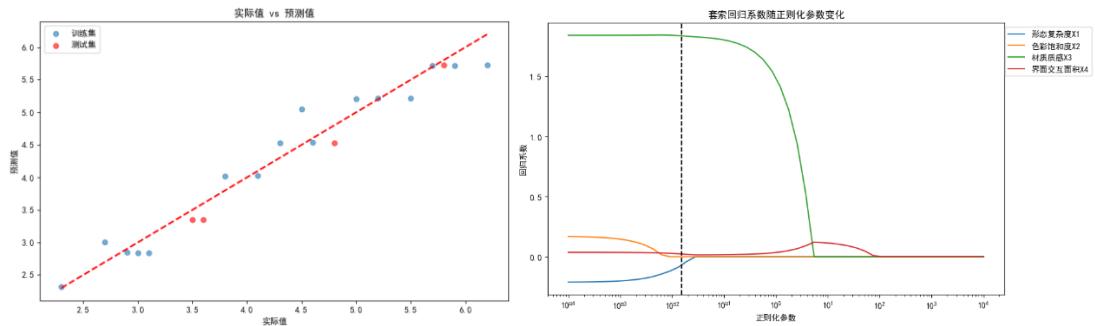


Lasso示例数据

分析结果

| 指标 | 详情 |
|--------------------|---|
| 模型信息 | 套索回归 最优 Alpha: 0.015199 |
| 回归方程 | 购买意愿 $Y = 4.2046 - 0.5208 \times \text{形态复杂度 } X_1 + 0.6762 \times \text{材质质感 } X_3 + 0.0012 \times \text{界面交互面积 } X_4$ |
| 因变量 | 购买意愿 Y |
| 截距 | 4.204612210049021 |
| 训练集 均方误差 (MSE) | 0.06176533597014681 |
| 测试集 均方误差 (MSE) | 0.0400287471777721 |
| 训练集 决定系数 (R^2) | 0.9569204282684242 |
| 测试集 决定系数 (R^2) | 0.9551184334376767 |
| 调整后决定系数 | 0.941255129456942 |
| F统计量 | 92.82965735174338 |

| | |
|------------------------|--------------------------------------|
| 交叉验证决定系数，评估模型的泛化能力。均值 | 0.8560142892374364 |
| 交叉验证决定系数，评估模型的泛化能力。标准差 | 0.06026810934139586 |
| 残差正态性检验 | Shapiro-Wilk 统计量: 0.9787 p 值: 0.9520 |
| 特征: 形态复杂度 X1 | 系数: -0.5208 标准化系数: -0.6682 |
| 特征: 色彩饱和度 X2 | 系数: -0.0000 标准化系数: -0.0185 |
| 特征: 材质质感 X3 | 系数: 0.6762 标准化系数: 0.5462 |
| 特征: 界面交互面积 X4 | 系数: 0.0012 标准化系数: -0.0000 |
| 统计检验 (仅非零系数特征) | |
| 常数项 t 值/p 值 | 8.1439 0.0000*** |
| 特征: 形态复杂度 X1 t 值/p 值 | -3.3002 0.0063** |
| 特征: 材质质感 X3 t 值/p 值 | 5.0073 0.0003*** |
| 特征: 界面交互面积 X4 t 值/p 值 | 0.7105 0.4910 |

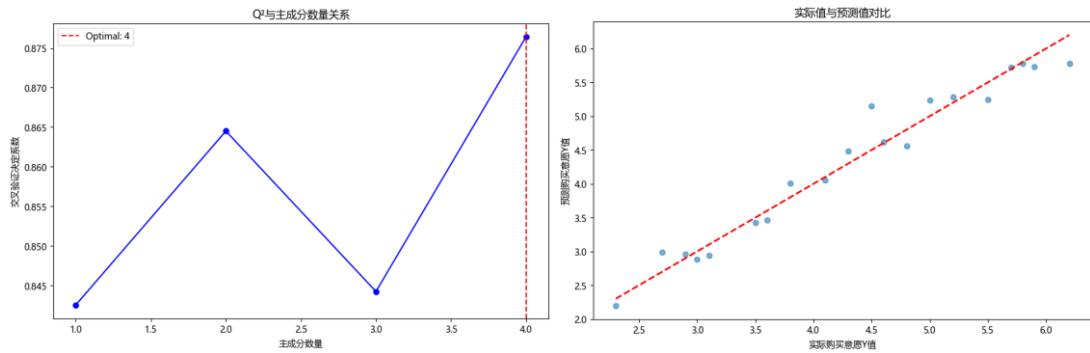


PLS 示例数据

分析结果

购买意愿 Y = 4.3250 - 0.6687 × 形态复杂度 X1 - 0.0596 × 色彩饱和度 X2 + 0.6514 × 材质质感 X3 + 0.0166 × 界面交互面积 X4

| 变量 | 系数 | 标准化系数 |
|--------------------------|---------|---------|
| 形态复杂度 (X ₁) | -0.6687 | -0.8583 |
| 色彩饱和度 (X ₂) | -0.0596 | -0.0145 |
| 材质质感 (X ₃) | 0.6514 | 0.5242 |
| 界面交互面积 (X ₄) | 0.0166 | 0.1907 |



7.2 非线性回归

线性回归和非线性回归的目的相同，对因变量和自变量的要求也相同，区别在于适用的“数据特征”不同。

- 当只有 1 个自变量 (X) 和 1 个因变量 (Y) 时：
画“X-Y 散点图”，观察点的分布是否接近一条直线。散点图呈直线趋势 → 线性回归，散点图呈“抛物线、指数增长、先增后减”的曲线 → 非线性回归。
- 当有多个自变量 (X1, X2...) 时：
分别画出每个自变量与因变量的散点图，然后选择。

DIAS 中的非线性回归包含两类，“非线性回归和多项式回归”，多项式回归属于非线性回归的特殊形式。

- 散点图呈现复杂曲线形态，如指数、对数、幂函数（钟形、指数增长 / 衰减、对数趋势），追求公式精准，使用线性回归。
- 数据量小 (< 50) 且波动大，散点图无规则图形，呈现简单弯曲形态，希望得出的函数公式更简单且易理解时，使用多项式回归。

非线性回归不要求数据符合正态分布，方差齐和无多重共线性。

非线性回归 示例数据

数据的最后一列为因变量，前面所有列均为自变量，自变量最多有 5 个。

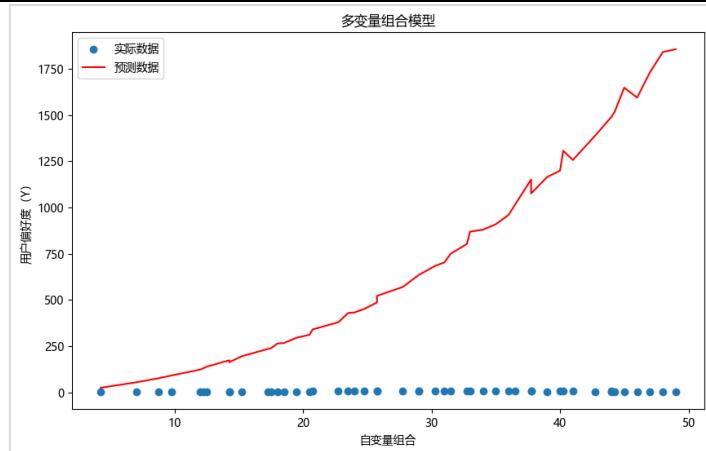
| | A | B | C | D | E |
|---|-----------|-----------|-----------|--------------|-----------|
| 1 | 线条数量 (X1) | 嵌套层数 (X2) | 色彩种类 (X3) | 复杂度综合评分 (X4) | 用户偏好度 (Y) |
| 2 | 5 | 1 | 1 | 10 | 2.8 |
| 3 | 10 | 1 | 2 | 15 | 3.1 |
| 4 | 15 | 2 | 2 | 20 | 3.5 |
| 5 | 20 | 2 | 3 | 25 | 3.9 |
| 6 | 25 | 2 | 2 | 20 | 4.2 |

分析结果

$$\begin{aligned} \text{用户偏好度 (Y)} = & ((-0.00077) * \text{线条数量 (X1)}^2 + (0.087255) * \text{线条数量 (X1)} + \\ & (0.712287) + (0.034369) * \text{嵌套层数 (X2)}^2 + (-0.1203) * \text{嵌套层数 (X2)} + (0.885245)) \\ & * (1 + \tanh(0.114482 * ((-0.00077) * \text{线条数量 (X1)}^2 + (0.087255) * \text{线条数量 (X1)} + \\ & (0.712287) + (0.034369) * \text{嵌套层数 (X2)}^2 + (-0.1203) * \text{嵌套层数 (X2)} + (0.885245)))) \end{aligned}$$

| | | |
|---------|-----------------------------|-----------|
| 多变量组合模型 | a_1 (线条数量 (X1)) | -0.000770 |
| 多变量组合模型 | b_1 (线条数量 (X1)) | 0.087255 |
| 多变量组合模型 | c_1 (线条数量 (X1)) | 0.712287 |
| 多变量组合模型 | a_2 (嵌套层数 (X2)) | 0.034369 |
| 多变量组合模型 | b_2 (嵌套层数 (X2)) | -0.120300 |
| 多变量组合模型 | c_2 (嵌套层数 (X2)) | 0.885245 |
| 多变量组合模型 | k (组合系数) | 0.114482 |
| 多变量组合模型 | MSE | 0.097050 |
| 多变量组合模型 | R-squared (R ²) | 0.908054 |

| | | |
|-------------------------|------------------------------|-------------|
| 多变量组合模型 | AIC | -102.626701 |
| 多变量组合模型 | 残差正态性检验, p 值>0.05 表示残差近似正态分布 | p=0.370649 |
| 最佳模型, 基于综合指标自动选择的最优拟合模型 | | 多变量组合模型 |



结论

关系规律：线条数量和用户偏好度是“先升后降”的曲线关系，不是简单的“越多越好”或“越少越好”；

结果可靠：69.6%的偏好度变化能被线条数量解释，关键参数都很可靠；

实用价值：可以用这个公式估算“最优线条数量”（比如求曲线顶点对应的 X1 值），帮设计决策（如界面设计、产品外观的线条数量选择）。

多项式回归 示例数据

数据的最后一列为因变量，前面所有列均为自变量，自变量的数量没有限制。

| | A | B | C | D | E |
|---|-----------|-----------|-----------|--------------|-----------|
| 1 | 线条数量 (X1) | 嵌套层数 (X2) | 色彩种类 (X3) | 复杂度综合评分 (X4) | 用户偏好度 (Y) |
| 2 | 5 | 1 | 1 | 10 | 2.8 |
| 3 | 10 | 1 | 2 | 15 | 3.1 |
| 4 | 15 | 2 | 2 | 20 | 3.5 |
| 5 | 20 | 2 | 3 | 25 | 3.9 |
| 6 | 25 | 3 | 3 | 30 | 4.3 |

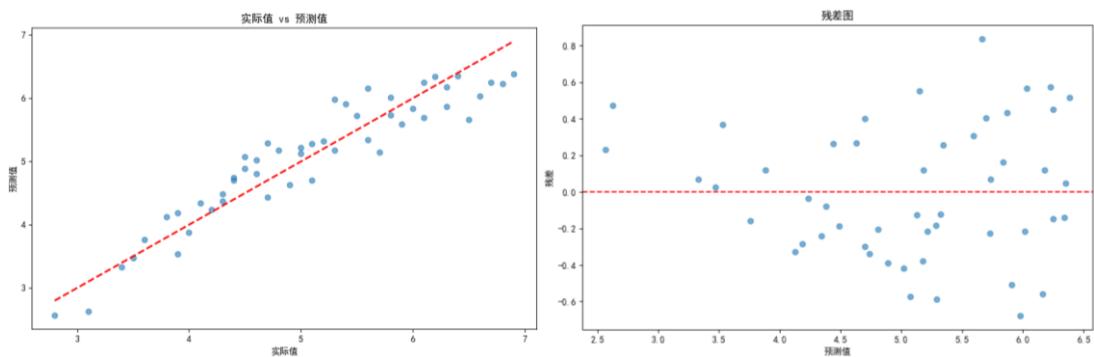
分析结果

$$y = 1.7258 + 0.0126 \times \text{线条数量 (X1)} + 0.7104 \times \text{嵌套层数 (X2)} - 0.0754 \times \text{色彩种类 (X3)} + 0.0144 \times \text{复杂度综合评分 (X4)}$$

| 指标 | 多项式回归 |
|------------------------|---------|
| 截距 | 1.7258 |
| 均方误差 (MSE) | 0.1285 |
| R 平方 (R ²) | 0.8783 |
| 调整后的 R 平方 | 0.8645 |
| F 值 | 81.1808 |
| 系数 (1) | 0.0 |

| 阶数 | 平均 MSE |
|----|------------|
| 1 | 0.2170 |
| 2 | 0.2324 |
| 3 | 28792.5009 |
| 4 | 20804.3238 |
| 5 | 68674.1816 |

| | |
|-----------------|---------|
| t 值(1) | 0.1888 |
| p 值(1) | 0.8511 |
| 系数(线条数量 (X1)) | 0.0126 |
| t 值(线条数量 (X1)) | 15.8551 |
| p 值(线条数量 (X1)) | 0.0 |
| 系数(嵌套层数 (X2)) | 0.7104 |
| t 值(嵌套层数 (X2)) | -2.5322 |
| p 值(嵌套层数 (X2)) | 0.0149 |
| 系数(色彩种类 (X3)) | -0.0754 |
| t 值(色彩种类 (X3)) | 0.2118 |
| p 值(色彩种类 (X3)) | 0.8332 |



结论

- 优先优化嵌套层数：它对结果的影响最大，适当增加嵌套层数，能明显提升结果。
- 适当关注线条数量和复杂度：这两个参数对结果有微弱的正向影响，但不用过度追求（比如不用为了多几条线强行增加设计元素）。
- 不用纠结色彩种类：它的影响不可靠，按设计需求选颜色即可。
- 模型很可靠：87.83% 的解释力 + 很小的预测误差，结果能放心用在设计决策里。

7.3 分类回归

分类回归和线性回归，非线性回归的用途是一样的，区别是，分类回归的因变量是分类数据（非连续数值），比如“用户是否购买（0/1）”，“产品评级（差 / 中 / 好）”。

DIAS 中的分类回归方法有三个，“二元 Logit 回归”，“多分类 Logit 回归”和“有序 Logit 回归”。这三种方法均不要求数据符合正太分布和方差齐，但是要求数据无多重共线性。

“有序”与“无序”

- “有序”指的是，因变量的多个类别之间没有顺序、等级或大小关系，仅代表“不同类别”，比如，出行方式：公交 / 地铁 / 自驾（三种方式只是不同选择，没有“好 / 坏”“高 / 低”的顺序）。

- “无序”指的是，因变量的多个类别之间有清晰的先后、高低或强弱顺序，但类别间的“间隔大小无法精确衡量”，比如，满意度评级：非常不满意(1) < 不满意(2) < 满意(3) < 非常满意(4)（有“从差到好”的顺序）。

“二元”与“多分类”

- “二元”指的是，因变量只有两个分类，比如，“是 / 否”，“买 / 不买”。
- “多分类”指的是，因变量有三个及更多的分类，比如，“公交 / 地铁 / 自驾”，“非常不满意 < 不满意 < 满意 < 非常满意”。

怎么选？

若因变量只有两类 → 选二元 Logit 回归；

若因变量是“多类别且无序” → 选多分类 Logit 回归；

若因变量是“多类别且有序” → 选有序 Logit 回归；

二元 Logit 回归示例数据

最后一列为因变量，且必须是二元数据（只能包含 0 和 1）

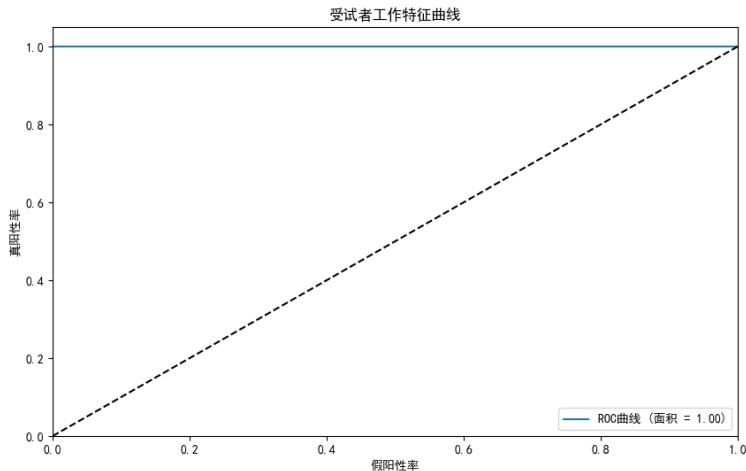
多人的数据依次向下增加行数即可。

| | A | B | C |
|---|-------|-----------|--------------------|
| 1 | 价格(元) | 用户评分(1-5) | 是否购买(1 = 是, 0 = 否) |
| 2 | 199 | 3.2 | 0 |
| 3 | 159 | 4.5 | 1 |
| 4 | 299 | 4 | 0 |
| 5 | 99 | 3.8 | 1 |
| 6 | 100 | 4.2 | 1 |

分析结果

逻辑回归($P(\text{是否购买} \ (1 = \text{是}, 0 = \text{否}) = 1) = 7.9211 - 1.1882 * \text{价格 (元)} + 63.3202 * \text{用户评分 (1-5)}$)

| | |
|----------------|-------------|
| 模型 | 二元 Logit 回归 |
| 价格(元) | -1.1882 |
| 用户评分(1-5) | 63.3202 |
| 截距 | 7.9211 |
| 准确性 | 1.0000 |
| ROC-AUC | 1.0000 |
| z 值(Intercept) | 0.0000 |
| z 值(价格(元)) | -0.0008 |
| z 值(用户评分(1-5)) | 0.0021 |
| p 值(Intercept) | 1.0000 |
| p 值(价格(元)) | 0.9994 |
| p 值(用户评分(1-5)) | 0.9983 |



结论

模型预测看似完美，但不可信：100%的准确率更可能是数据问题（如样本少、完美分离），而非模型真的能精准预测。

价格和评分的影响都不可靠： p 值远大于 0.05，不能用“价格高→购买少”，“评分高→购买多”这些结论指导决策（比如不能为了提升销量刻意降价或刷评分）。

后续建议：需要补充更多样本数据，或检查数据是否存在“完美分离”（比如是否有某个条件能完全区分购买 / 不购买），重新分析才能得到可靠结论。

多分类 Logit 回归 示例数据

最后一列为因变量。

分析用户的年龄、性别、收入等特征，如何影响他们对“智能手环界面风格”的偏好（界面风格分 3 类：Class 1 = 极简风、Class 2 = 科技风、Class 3 = 运动风）。

| | 年龄 | 性别 (1 = 男 / 0 = 女) | 月收入 (千元) | 日常通勤 (1 = 是 / 0 = 否) | 运动健身 (1 = 是 / 0 = 否) | 睡眠监测 (1 = 是 / 0 = 否) | 品牌熟悉度 (1-5 分) | 使用频率 (1 = 低 / 2 = 中 / 3 = 高) | 教育程度 (1 = 高中及以下 / 2 = 专 / 3 = 本科 / 4 = 硕士及以上) | 偏好风格 (1 = 极简风 / 2 = 科技风 / 3 = 运动风) |
|---|----|--------------------|----------|----------------------|----------------------|----------------------|---------------|------------------------------|---|------------------------------------|
| | A | B | C | D | E | F | G | H | I | J |
| 1 | 年龄 | 性别 | 月收入 | 日常通勤 | 运动健身 | 睡眠监测 | 品牌熟悉度 | 使用频率 | 教育程度 | 偏好风格 |
| 2 | 28 | 0 | 8 | 1 | 0 | 1 | 4 | 2 | 3 | 1 |
| 3 | 35 | 1 | 15 | 1 | 1 | 0 | 3 | 3 | 2 | 2 |
| 4 | 42 | 1 | 22 | 0 | 1 | 1 | 2 | 3 | 2 | 3 |
| 5 | 22 | 0 | 5 | 1 | 1 | 0 | 5 | 3 | 3 | 2 |
| c | 28 | 1 | 12 | 1 | 1 | 1 | 4 | 2 | 3 | 1 |

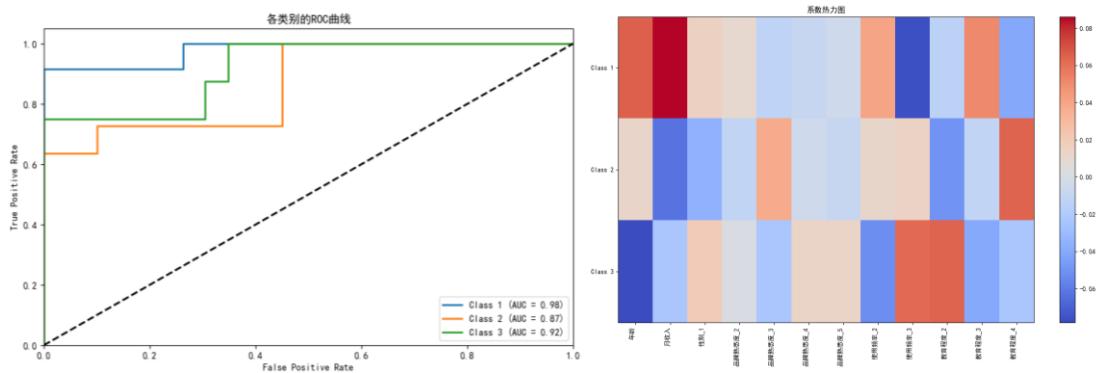
分析结果

（内容过长，仅保留回归方程和图片）

- Class 2 vs Class 1: $\ln(P(2)/P(1)) = 4.6449 - 0.0527 \times \text{年龄} - 0.1494 \times \text{月收入} - 0.0510 \times \text{性别}_1 - 0.0213 \times \text{品牌熟悉度}_2 + 0.0501 \times \text{品牌熟悉度}_3 +$

$0.0037 \times \text{品牌熟悉度}_4 - 0.0040 \times \text{品牌熟悉度}_5 - 0.0282 \times \text{使用频率}_2 +$
 $0.0916 \times \text{使用频率}_3 - 0.0352 \times \text{教育程度}_2 - 0.0632 \times \text{教育程度}_3 + 0.1042 \times$
 教育程度_4

- Class 3 vs Class 1: $\ln(P(3)/P(1)) = 6.6731 - 0.1438 \times \text{年龄} - 0.1093 \times \text{月收入} + 0.0045 \times \text{性别}_1 - 0.0088 \times \text{品牌熟悉度}_2 - 0.0103 \times \text{品牌熟悉度}_3 + 0.0219 \times \text{品牌熟悉度}_4 + 0.0177 \times \text{品牌熟悉度}_5 - 0.0929 \times \text{使用频率}_2 + 0.1393 \times \text{使用频率}_3 + 0.0781 \times \text{教育程度}_2 - 0.0913 \times \text{教育程度}_3 + 0.0170 \times \text{教育程度}_4$
- Class 3 vs Class 2: $\ln(P(3)/P(2)) = 2.0281 - 0.0911 \times \text{年龄} + 0.0401 \times \text{月收入} + 0.0554 \times \text{性别}_1 + 0.0125 \times \text{品牌熟悉度}_2 - 0.0604 \times \text{品牌熟悉度}_3 + 0.0182 \times \text{品牌熟悉度}_4 + 0.0216 \times \text{品牌熟悉度}_5 - 0.0647 \times \text{使用频率}_2 + 0.0478 \times \text{使用频率}_3 + 0.1134 \times \text{教育程度}_2 - 0.0281 \times \text{教育程度}_3 - 0.0872 \times \text{教育程度}_4$



结论

年龄越大，选科技风比极简风的概率越低，选运动风比极简风的概率低得更明显；
 常运动的人，选科技风比极简风的概率更高，选运动风比极简风的概率高很多；
 关注睡眠监测的人，选科技风比极简风的概率更低，选运动风比科技风的概率高；
 收入越高，选运动风比极简风的概率越低；
 大专学历的人，选运动风比科技风的概率高；
 针对“年轻、爱运动”的用户群体，主推科技风或运动风界面；
 针对“高收入、关注健康监测”的用户，重点优化极简风界面。

有序 Logit 回归 示例数据

| | A | B | C |
|---|------|-------|-------|
| 1 | 续航能力 | 价格合理性 | 用户满意度 |
| 2 | 48 | 6 | 4 |
| 3 | 24 | 7 | 2 |
| 4 | 72 | 5 | 5 |
| 5 | 36 | 8 | 3 |
| 6 | 60 | 7 | 4 |

分析结果

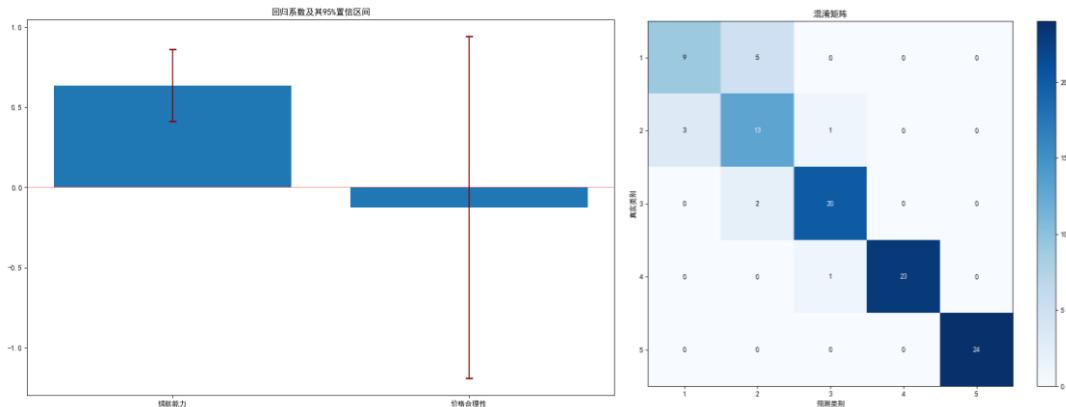
$$Z = 0.6360 \times \text{续航能力} - 0.1248 \times \text{价格合理性}$$

| 指标 | 值 |
|----------|-----------|
| 准确率 | 0.8812 |
| 加权 Kappa | 0.9215 |
| 伪 R 平方 | 0.7889 |
| 观察编号 | 101 |
| 对数似然 | -33.8710 |
| LL-Null | -160.4662 |
| LLR p 值 | 0.0000 |
| 模型收敛状态 | 融合 |
| 优化方法 | bfgs |
| 使用的迭代次数 | 1000 |

| 变量 | 系数 | z 值 | p 值 |
|-------|---------|---------|--------|
| 续航能力 | 0.6360 | 5.5516 | 0.0000 |
| 价格合理性 | -0.1248 | -0.2294 | 0.8185 |

| 变量 | 系数 | z 值 | p 值 |
|-------|---------|---------|--------|
| 临界点 1 | 13.8486 | 2.2744 | 0.0229 |
| 临界点 2 | 1.5245 | 6.7461 | 0.0000 |
| 临界点 3 | 2.0539 | 10.1933 | 0.0000 |
| 临界点 4 | 2.5271 | 13.1663 | 0.0000 |

| 预测类别: 1 | 预测类别: 2 | 预测类别: 3 | 预测类别: 4 | 预测类别: 5 |
|---------|---------|---------|---------|---------|
| 9 | 5 | 0 | 0 | 0 |
| 3 | 13 | 1 | 0 | 0 |
| 0 | 2 | 20 | 0 | 0 |
| 0 | 0 | 1 | 23 | 0 |
| 0 | 0 | 0 | 0 | 24 |



结论

关键影响因素：只有“续航能力”显著影响满意度，续航越强，满意度越高；“价格合理性”没影响；

模型准确度：预测准确率 88.12%，几乎能完美匹配真实满意度等级；

实用价值：想提升用户满意度，优先提升续航能力（比如从 48 小时升到 72 小时），不用过分纠结价格调整。

8 找到一群好朋友（聚类）

聚类就是把数据分成几堆，每堆的内部数据尽量像，堆之间尽量不像。比如：把 1000 个用户按“消费金额、购买频次”分成 3 堆（高活跃、中活跃、低活跃），方便针对性运营；或者把手机按“价格、性能”分成 4 类（入门款、性价比款、旗舰款、奢侈品款），快速归纳产品定位。

DIAS 中的聚类工具有三个，“K-Means，二阶聚类和密度聚类”，分析前需要检验数据的多重共线性，若共线性过高则无法分析。

- “K-Means 和密度聚类” 仅支持连续数据，不支持分类数据，二阶聚类都支持。
- K-Means 仅支持“圆 / 球形”数据，也就是数据在散点图中的形状大概是“圆 / 球形”。另外两种可以支持任意形状（长条形、环形等）

怎么选？

- 数据是连续型、小样本、需要快速分群 → K Means；
- 数据是混合类型（连续 + 分类）、大规模 (>1 万条) → 二阶聚类；
- 数据是连续型、有异常值、形状不规则 → 密度聚类。

K-Means 示例数据

用户对 UI 元素的偏好差异是提升产品竞争力的关键，通过眼动追踪实验与行为日志采集获取多种数据，

- 眼动数据：用户对不同 UI 元素的注视时长、注视次数、瞳孔扩张度。
- 行为数据：界面操作效率（完成任务耗时）、元素点击频率、页面停留时间。
- 主观评分：用户对 10 项 UI 元素（如“按钮圆角弧度”“图标复杂度”等）的满意度评分（1-5 分）

| 特征维度 | 说明（量化方式） |
|---------|---------------------|
| 色彩饱和度偏好 | 高饱和度元素注视时长占比 (0-1) |
| 按钮圆角偏好 | 圆角按钮点击频率 / 直角按钮点击频率 |
| 信息密度耐受度 | 高密度界面停留时间 (秒) |
| 图标复杂度偏好 | 复杂图标识别正确率 (0-100%) |
| 交互反馈敏感度 | 无反馈操作纠错次数 |
| 导航层级偏好 | 平均操作路径长度 (步骤数) |
| 视觉焦点集中度 | 页面焦点区域注视占比 (0-1) |
| 界面美学评分 | 主观满意度评分 (1-5 分) |

| | A | B | C | D | E | F | G | H |
|---|---------|--------|---------|---------|---------|--------|---------|--------|
| 1 | 色彩饱和度偏好 | 按钮圆角偏好 | 信息密度耐受度 | 图标复杂度偏好 | 交互反馈敏感度 | 导航层级偏好 | 视觉焦点集中度 | 界面美学评分 |
| 2 | 0.28 | 0.35 | 120 | 65 | 3 | 4 | 0.32 | 2.8 |
| 3 | 0.22 | 0.21 | 90 | 58 | 2 | 3 | 0.28 | 2.5 |
| 4 | 0.31 | 0.42 | 110 | 70 | 4 | 5 | 0.35 | 3 |
| 5 | 0.25 | 0.3 | 105 | 62 | 3 | 4 | 0.3 | 2.7 |
| 6 | 0.20 | 0.20 | 115 | 60 | 2 | 5 | 0.22 | 2.0 |

分析步骤

聚类数就是把用户分成的“堆数”，若无前提条件，建议勾选。

自动确定最佳聚类数

若不勾选，点击“分析文件”按钮后，会出现下面的输入框，可以输入希望的“堆数”。

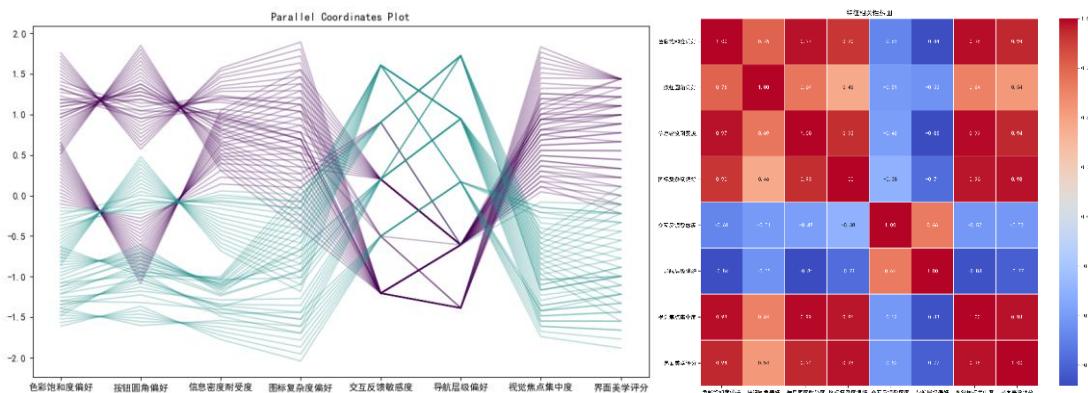
请输入聚类数量 (1-10)

选择想要重点展示的特征（即数据中的列），以便更有针对性地观察聚类结果，例如输入 "1,3,5" 表示选择第 1 列、第 3 列和第 5 列的特征进行可视化。

请选择可视化特征 (用逗号分隔列索引, 如1,2,3)
1,2,3,4,5,6,7,8

分析结果

| 簇ID | 色彩饱和度偏好 | 按钮圆角偏好 | 信息密度耐受度 | 图标复杂度偏好 | 交互反馈敏感度 | 导航层级偏好 | 视觉焦点集中度 | 界面美学评分 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|
| 0 | 0.8428 | 0.5670 | 0.8503 | 0.7724 | -0.6333 | -0.8279 | 0.8428 | 0.8218 |
| 1 | -0.9306 | -0.6260 | -0.9389 | -0.8529 | 0.6993 | 0.9141 | -0.9306 | -0.9074 |



结论

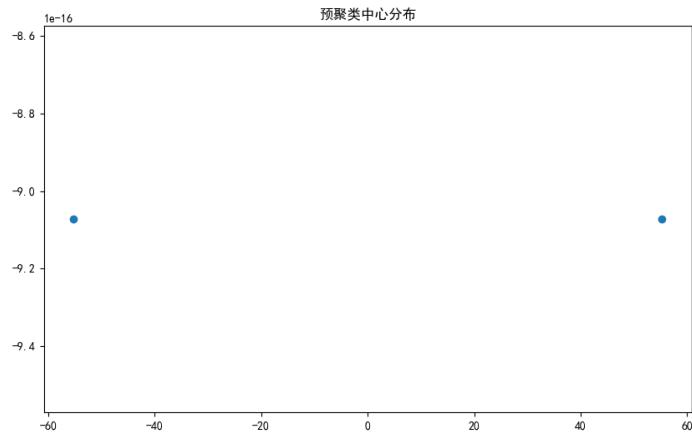
- 如果你的产品主要用户是标签 0 → 重点做“颜值”和“简化步骤”；
 - 如果是标签 1 → 重点做“效率”和“清晰反馈”；
 - 甚至可以做“个性化 UI”：让用户自己选“视觉模式”或“实用模式”，分别对应两类用户的偏好。

二阶聚类 示例数据

分析结果

| | |
|------------|--|
| | 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0 |
| 聚类数量 | 2 |
| 预聚类数量 | 2 |
| 轮廓系数 | 0.6820 |
| 卡林斯基哈拉巴什指数 | 410.8165 |
| 戴维斯-博尔丁 指数 | 0.4052 |

| 簇 ID | 样本计数 | 特征均值 | 特征标准 | 特征中值 |
|------|------|---|---|---|
| 0 | 59 | 0.6976, 0.6541, 226.8305, 84.8475, 2.1186, 1.8305, 0.6186, 4.3458 | 0.1332, 0.1953, 25.7433, 7.0730, 1.3414, 0.5570, 0.0842, 0.5130 | 0.7000, 0.7200, 229.0000, 85.0000, 1.0000, 2.0000, 0.6200, 4.5000 |
| 1 | 42 | 0.3019, 0.4012, 118.0000, 65.6429, 3.5476, 4.1190, 0.3398, 2.7929 | 0.0774, 0.1292, 25.8871, 6.6326, 1.0735, 0.6970, 0.0617, 0.4350 | 0.2950, 0.3650, 114.5000, 66.0000, 3.0000, 4.0000, 0.3400, 2.8000 |



密度聚类示例数据

分析结果

DBSCAN eps (估计值): 0.73867, 最小样本数: 16

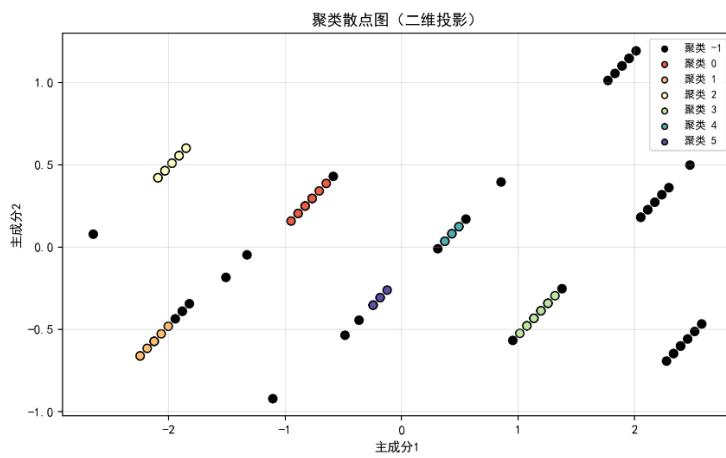
PCA 解释方差 (双样本): 89.92%

轮廓: N/A

聚类-1: 101 个样本

| 簇 | 特征均值 (截断) |
|----|-------------------------|
| -1 | 58.4000, 6.5250, 4.0000 |
| 0 | 34.9231, 8.0000, 3.0000 |
| 1 | 19.5556, 8.0000, 1.0000 |
| 2 | 25.5714, 9.0000, 2.0000 |
| 3 | 55.2308, 6.0000, 4.0000 |
| 4 | 46.0000, 7.0000, 4.0000 |

| | |
|---|-------------------------|
| 5 | 39.3333, 7.0000, 3.0000 |
|---|-------------------------|



9 软件报错

1. 大部分问题都是因为数据格式，因此首先检查数据。
使用“数据清洗”功能清洗数据。
把数据复制到一个新的 excel 文件中，重新分析。
2. An error occurred while analyzing the file: openpyxl does not support the old .xls file format, please use xlrd to read this file, or convert it to the more recent .xlsx file format.
把.xls 格式文件改为.xlsx 文件。
3. **An error occurred while analyzing the file: Singular matrix**
在 Excel 中绘制数据的散点图，查看是否有数据完全重合。
4. **An error occurred while analyzing the file: property 'result' of 'VIFHandlingDialog' object has no setter**
数据有多重共线性，需做多重共线性检验，去除或修正有多重共线性的变量。