

心理测试在人才选拔中的综合评估与分类研究

摘要

心理测试作为一种重要的研究手段，广泛应用于教育、心理疾病诊断、心理咨询和人才选拔中。本文通过分析某单位 148 名求职者的心理测试数据，研究三种常用量表（症状自评量表、职业成熟度量表和大五人格量表）在测试者分类中的应用，并探讨各量表测试结果之间的关联性。此外，本文还建立了综合评价体系，综合考虑三种量表，旨在对测试者的心理状况、个人行为特征和社会适应能力进行全面画像。

对于问题 1，首先提取并汇总 148 位测试者的 391 份心理测试报告中的有价值信息，同时剔除不符合要求的报告。获得数据后，使用肘部法计算聚类数目，并使用 k-means 聚类算法将测试者分成四类。对结果进行分析，使用卡方检验方法得出结论，即症状自评量表、职业成熟度量表和大五人格量表各自分出的四组测试者之间没有显著差异。

对于问题 2，本文建立了一个较为完备的综合评价体系，能够有效地利用三种量表的信息，直观地反映测试者的心理状况、个人行为特征和社会适应能力。人才综合评价体系分为两个评价层次，包含三个维度（症状评价、职业成熟评价和人格评价），并提供十多个细分指标。基于模糊综合评价的思想，从主观和客观两方面，依据指标的重要程度赋予不同的权重。对于第一评价层，采用层次分析赋权方法；对于第二评价层，采用熵值法赋权。在此过程中，我们对数据进行了预处理，通过正向化和标准化使得综合考虑这些指标更加容易。模型建立完成后，本文依据三个维度在三维空间进行聚类分析，将测试者分为四类。最后，我们对模型进行了稳定性检验，证明该模型较原报告有更大的变化幅度，更易于进行心理健康状况分析。

对于问题 3，我们旨在分析 A 组与 B 组测试者之间的差异，两组的数字特征进行比较分析。我们采用独立样本 t 检验来比较两组样本的均值，并使用 Levene 检验来判断两组样本的方差是否相等。通过对综合评分、职业成熟度评分和 SCL90 评分三个维度的数据分析，我们发现 A 组与 B 组的测试者在这些维度上并没有显著差异。然而，在大五人格评分方面，A 组与 B 组的测试者存在显著差异。因此，A 组与 B 组的测试者在大五人格评分维度上存在差异，说明两组测试者之间确实存在差异。

对于问题 4，首先进行了敏感性分析，通过观察不同变化幅度下量表得分的变化，可以评估各量表对变化的敏感程度。如表 9 所示，SCL90 的得分变化幅度在变化幅度 5% 和 10% 时分别为 3.4% 和 13.2%，说明模型在较小变化幅度下的表现较为稳定和鲁棒。另外，通过图 9 显示的总分比较，当变化幅度较大时，得分变化幅度的误差较大，这表明该模型能够更好地反映 SCL90 对人才评分的影响。

关键字： 心理健康评估 症状自评量表 职业成熟度量表 大五人格量表 多元分析 人才选拔

一、问题重述

心理健康问题近来得到越来越多的关注，而心理测试作为评估心理健康的重要手段在各种领域中得到了广泛应用，如教育、心理疾病诊断、心理咨询和人才选拔等。心理测试利用专业的测量工具和方法，帮助了解个体的心理状况、行为特征和社会适应能力。

目前，针对不同人群和应用场景，存在多种心理测试表，其中包括症状自评量表（SCL-90）、职业成熟度量表和大五人格量表。这些量表在人才招聘领域被广泛使用。SCL-90 是全球著名的心理健康测试量表，涵盖广泛的精神病症状学内容，帮助评估个体的心理健康状况；职业成熟度量表旨在测量个体的职业成熟度；大五人格量表基于大五人格理论，评估个体的外向性、宜人性、严谨性、神经质和开放性。

由于心理测试具有间接测量的性质，其评价标准不如物理测量那样绝对和普遍，心理测试结果需要通过多个测试表的交叉验证来提高准确性。然而，交叉验证带来分析上的挑战，因为不同量表的测试项目和侧重点不同，可能导致结果的相悖。

某心理测试机构受单位委托，对该单位 148 名求职者进行心理测试。测试机构先将测试者随机分成 A、B 两组，然后对每位求职者采用症状自评量表、职业成熟度量表和大五人格量表三个表进行测试。测试数据和相关分析具体见附件。该单位希望通过测试结果分析，了解每位求职者心理健康状况，人际关系，以便对求职者有一个更加全面的了解。

请根据附件所提供的测试数据，通过数学建模完成下列问题：

问题 1 三种量表从不同角度对测试者进行分析，请分别按表对 148 名测试者进行分类。并建立模型，研究三种量表得到结果之间有无联系。

问题 2 考虑三种量表因子和测试数据，建立综合评价指标体系，对 148 名测试者进行分类。

问题 3 请建立模型，分析 A 组测试者与 B 组测试者之间有无差异。（注：问题 1，2 中不考虑分组）。

问题 4 三种量表均使用得分值总和来研究测试者，请建立模型，对这种方法的准确性加以评判。

通过上述研究，希望能够为求职者心理健康状况的全面了解提供科学依据，为人才选拔中的心理测试应用提供参考。

二、问题分析

2.1 问题一分析

对测试者分类

在解决问题一时，首先需要进行数据处理。数据收集阶段涉及汇总三张心理测试量表（SCL-90、职业成熟度量表和大五人格量表）中的有效信息，以建立完整的测试数据集。随后的数据清洗过程则专注于移除无用和不相关的数据，例如剔除测试时间小于一分钟的样本。

接下来，为了针对每种量表对测试者进行分类，选择三种量表各自具有代表性的 5、8、9 个参数进行聚类分析。而在选择使用 k-means 聚类分析前，确定聚类的数目 k 是一个很重要的议题，可以通过肘部法解决。

三种量表的联系

关于三个量表之间的联系，可以通过计算相关系数矩阵来实现，相关系数不仅可以描述变量之间的线性关系强度，还可以用作变量相似性的度量。基于相关系数矩阵，可以进一步进行变量聚类分析，以识别和描述不同量表中相关联的参数组合。

三种量表得出结果之间的联系

最后，为了确定三种量表得到的结果是否存在联系，将结合前述的量表参数聚类结果，探索是否存在某种模式或者特定算法能够有效地描述量表间的联系和关联性。这一过程旨在提供对求职者心理健康状况综合评估的科学依据，并为心理测试在人才选拔过程中的应用提供理论支持。

2.2 问题二分析

为了充分利用测试者三种测试的所有指标，最大化利用所有指标的价值，建立我们自己的评价体系，我们采用基于模糊评价模型的双层人才评价体系，通过层次分析法和熵权法，将主观与客观相结合，建立全面的综合评价体系。通过正向化，标准化的方法处理评价数据，均衡的分析每个指标的权重。建立模型后，再通过聚类分析进行分类。

2.3 问题三分析

为了分析 A 组与 B 组测试者之间的差异。我们将人群分为 AB 两组，并对两组的数字特征进行比较分析。我们采用独立样本 t 检验来比较两组样本的均值，并使用 Levene 检验来判断两组样本的方差是否相等，因为 Levene 检验是一种非参数检验，不依赖于数据的分布特性。

2.4 问题四分析

根据第二问建立的人才评价模型，从三个方面入手分析人才综合评价模型与量表得分总和。第一方面，采用正向化，标准化处理方法对数据进行分析处理，消除数据不标准、量纲不统一的误差。第二方面，特殊数据的显著反应，通过对比发现，人才总和评价模型在特殊数据点处拥有更好的评分标准。第三方面，总体分析，该模型可以放大关键指标，扩大方差，使总分更易反映出人才的心理状况。

三、假设与约定

1. 假设使用的所有数据是科学有效，能够较好地反映测试者的心理状况、个人行为特征和社会适应等能力。
2. 假设测试时间小于一分钟的报告不具有参考价值，也就是不考虑测试时间小于一分钟的测试者的报告。
3. 约定只研究同时做过三个人格测试的测试者，不考虑只做过其中一项或两项测试的测试者的报告。

四、符号说明及名词解释

五、模型建立与求解

5.1 问题一求解

5.1.1 数据汇总及清洗

数据选择及汇总 在附录提供的 148 人的 391 份心理测试报告中，不是每一个人都做了三份测试，同时有的测试报告的完成时间小于 1 分钟。由于本题大部分问题旨在探讨三种测试的之间的议题，而且正常的测试需要 10-30 分钟完成 [1][2][3]，本文做出假设 2 和 3，即假设测试时间小于一分钟的报告不具有参考价值，舍弃测试时间小于一分钟的测试者的报告，并约定只研究同时做过三个人格测试的测试者，不考虑只做过其中一项或两项测试的测试者的报告。

为了建立完整的测试数据集，需要选择三种测试中有价值的测验指标。对于**症状自评量表 SCL-90**，选择躯体化、强迫症状、人际关系敏感、抑郁、焦虑、敌对、恐怖、偏执、精神病性 9 大指标的均分较为合适；对于**职业成熟度量表**，选择信息应用、职业认知、自我认知、个人调适、职业态度、价值观念、职业选择、条件评估这 8 大指标的均分比较合适；而对于**大五人格量表**，应当选取 N（神经质），E（外向性），O（开放性），A（宜人性），C（严谨性）这五大指标的得分；

表 1 符号说明

符号	意义
$d(x_i, c_i)$	数据点 x_i 到质心 c_i 的欧式距离
M	数据的维度, 指标的数量
n	样本数量, 测试者人数
x_{im}	数据点 i 在第 m 个维度的坐标
c_{im}	质心 i 在第 m 个维度的坐标
k	簇的数量
N_j	第 j 个簇中的数据点数量。
SSE	数据集的误差平方和
r_{jk}	两变量 x_j 与 x_k 的相关系数
\bar{x}_j	第 j 个变量的样本均值
χ^2	卡方统计量
E_i	第 i 类的期望频数, 即根据理论分布或假设计算出的期望值。
$R(G_1, G_2)$	表示两类变量 G_1 和 G_2 之间的距离。
H_j	第 j 个指标的熵值
w_j	第 j 个指标的权重
p_{ij}	第 i 个样本在第 j 个指标上的标准化值占该指标总和的比重
μ	在综合评价指标体系下得出的总分
t	t 统计量
s_p	合并标准差
df	自由度
W	Levene 检验的统计量

经过两轮筛选和处理, 我们最终得到了 **106** 人的各自的三大测试数据, 这样处理的数据很好地保留了测试结果中有价值的信息, 同时具有很高的研究意义。

5.1.2 k-means 聚类分析

为了实现对测试者的分类, 可以采用分散化聚类方法。**k-mean** 聚类方法简洁且快速, 适合解决这种简单的分类。数据对象间的相似度量一般是通过数据之间的相互关系来确定, 而得到距离值之后, 元素间才可以被联系起来。通常用欧式距离作为衡量数据对象间相似度的指标, 相似度与数据对象间的距离成反比, 相似度越大, 距离越小。算法需要预先指定初始聚类数目 k 各初始聚类中心, 根据数据对象与聚类中心之间的相似度, 不断更新聚类中心位置, 不断降低类簇的误差平方和 SSE , 当其不在变化时, 聚

类结束，得到最终结果。

空间中数据对象与聚类中心间的欧式距离计算公式为：

$$d(x_i, c_j) = \sqrt{\sum_{m=1}^M (x_{im} - c_{jm})^2} \quad (1)$$

整个数据集的误差平方和 SSE 计算公式为：

$$SSE = \sum_{j=1}^k \sum_{i=1}^{N_j} d(x_i, c_j)^2 \quad (2)$$

K-means 算法 归纳为 (J. MacQueen, 1967) [4]：

1. 随机选择 k 个初始质心 (centroids)。
2. 将每个点分配到离它最近的质心，形成 k 个簇 (clusters)。
3. 重新计算每个簇的质心，即将每个簇中的点的平均值作为新的质心。
4. 重复步骤 2 和 3，直到质心不再变化或达到预定的迭代次数。

对于本题的三种量表，分别取每个测试者作为数据点，其测试结果的指标作为数据点的 M 个维度的值，三张表的维度分别为 9, 8 和 5。

Elbow Method 选择好数据集和聚类方法，还有很重要的问题是确定合适的 k 值 (即聚类的数量)，有很多种方法可以确定最佳的 k 值，其中肘部法 (Elbow Method) 简单直观，易于理解和实现，且常用于 K-means 聚类。

1. 计算不同 k 值对应的总平方误差 (SSE, Sum of Squared Errors)。
2. 将 k 值与 SSE 绘制成图。
3. 找到图中 SSE 下降速度显著变缓的位置，即“肘部”，作为最佳 k 值。

下 (上) 方以症状自评量表为例，不同 k 值和 SSE 的关系图，可以看出当 $k=4$ 时，SSE 下降速度显著变缓，可以作为最佳 k 值。

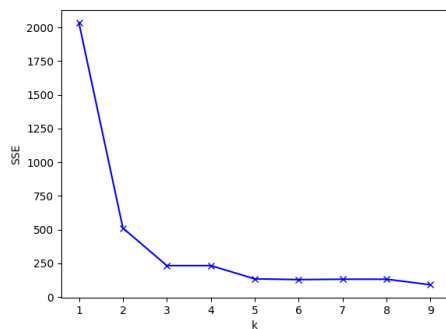
职业成熟度量表与大五人格量表得到的结果与之相似，最佳 k 值均为 4。

对于每种量表，都把人群分为了四类，聚类执行结果如图 1：

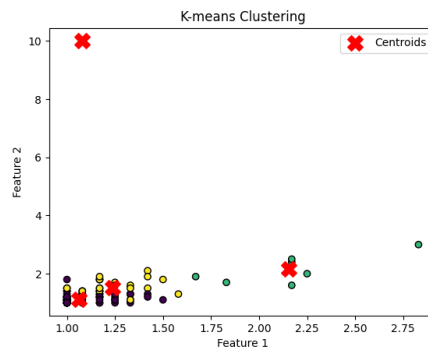
分别对职业成熟度、SCL90 和大五人格各 8、10、5 项指标进行聚类，可得到四个类别的中心点特征值，并且做出可视化雷达图如图 3：

5.1.3 三种量表得出结果之间的联系

依据三种量表分别聚类，都将测试者分为了四类，现在探究三种量表分类结果之间的联系。

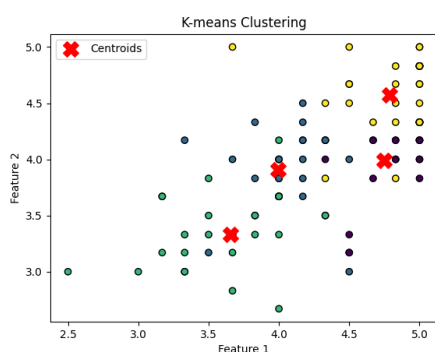


(a) 症状自评量表中 k 值和 SSE 的关系

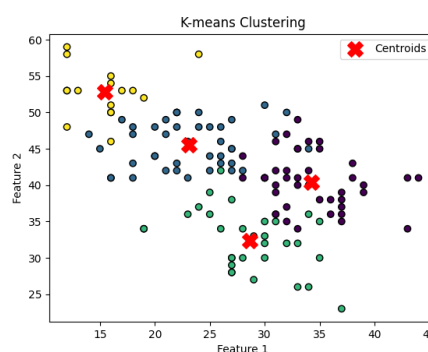


(b) 症状自评量表聚类结果

图 1 症状自评量表



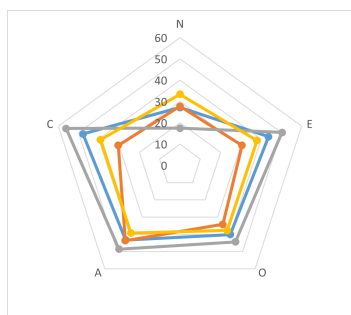
(a) 职业成熟度量表聚类结果



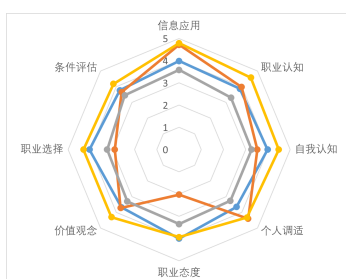
(b) 大五人格量表聚类结果

图 2 职业成熟度量表与大五人格量表聚类结果

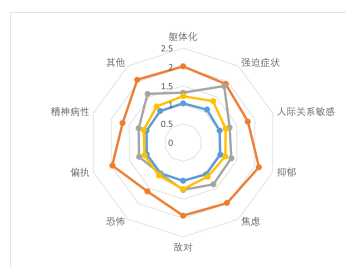
卡方检验 三种分类结果彼此之间互有重叠，但若需要证明它们之间是否存在联系，本文需要用到卡方检验。卡方检验（Chi-Square Test）[5] 是一种统计检验方法，用于判断观测数据与预期数据之间是否有显著差异。它广泛用于检验两个分类变量之间的关联性，或是检验一个变量的实际分布是否符合理论分布。卡方检验有多种形式，最常用的有独立性检验，用于检验两个分类变量是否独立，即是否存在关联性。



(a) 大五样本中心均值雷达图



(b) 职业样本中心均值雷达图



(c) 病症样本中心均值雷达图

图 3 聚类样本中心均值雷达图

卡方统计量的计算公式为：

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

其中卡方统计量 (χ^2)：反映观测值与期望值之间的差异大小。如果 χ^2 值较大，说明观测数据与预期数据之间存在显著差异。而观测频数 (O_i) 和期望频数 (E_i) 分别表示实际观测到的数据频数和基于理论分布或假设的预期数据频数。

给定数据如表 9，其中行表示职业成熟度量表分类的四组，列表示症状自评量表分类的四组，例如数据 19 表示有十九个人既属于职业成熟度量表分好的类 1，又属于症状自评量表分好的类 2。

期望频数计算：

$$E_{ij} = \frac{\text{行总计} \times \text{列总计}}{\text{总计}} \quad (4)$$

期望频数表格如表 2b：

	类 1	类 2	类 3	类 4	行总计
类 1	13	0	0	13	26
类 2	19	1	0	3	23
类 3	6	4	5	8	23
类 4	27	0	1	4	32
列总计	65	5	6	28	104

(a) 给定数据

	类 1	类 2	类 3	类 4
类 1	16.25	1.25	1.5	7
类 2	14.375	1.106	1.327	6.192
类 3	14.375	1.106	1.327	6.192
类 4	20	1.538	1.846	8.615

(b) 期望频数

表 2 数据与对应期望频数

计算每个单元格的 χ^2 值：

$$\chi_{11}^2 = \frac{(13 - 16.25)^2}{16.25} \approx 0.650$$

$$\chi_{12}^2 = \frac{(0 - 1.25)^2}{1.25} \approx 1.250$$

$$\chi_{13}^2 = \frac{(0 - 1.5)^2}{1.5} \approx 1.500$$

重复上述过程计算所有单元格的 χ^2 值，并求和：

总的 χ^2 值：

$$\chi^2 = 0.650 + 1.250 + 1.500 + 5.143 + 1.497 + 0.010 + 1.327 + \dots \approx 43.0124$$

自由度 (df) 计算:

$$\text{自由度 (df)} = (r - 1)(c - 1) = (4 - 1)(4 - 1) = 9$$

在显著性水平 $\alpha = 0.05$ 下, 自由度为 9 的卡方临界值为 16.919。

因为计算的 $\chi^2 = 43.0124$ 大于临界值 16.919, 所以拒绝零假设, 认为职业成熟度量表分类结果与 SCL 量表分类结果之间有显著关联。

通过相同的处理, 可以得出职业成熟度量表分类结果与大五人格量表分类结果的卡方值为 85.0672, SCL 量表分类结果与大五人格量表分类结果卡方值为 18.6717, 均可以拒绝原假设, 认为他们之间有显著联系。

5.2 问题二求解

三种量表分别从不同角度对测试者进行分析, 侧重各有不同。人才评价也应当如此, 多维度, 多指标地对测试者进行画像。

为了充分利用测试者三种测试的所有指标, 最大化三种量表的参考价值, 建立我们自己的评价体系, 直观地反映测试者的心理状况、个人行为特征和社会适应能力, 需要找到一种综合评价指标体系, 用来结合三个量表, 得到最终结果。

本文建立的人才综合评价体系分为两个评价层次, 包含三个维度 (症状评价、职业成熟评价和人格评价), 并提供十多个细分指标。维度的选择立足于三种量表的侧重各有不同。指标的选择及建立基于对量表参数关系的研究和变量聚类讨论。基于模糊综合评价的思想, 通过主观和客观两方面考察赋予不同的权重。对于第一评价层, 采用层次分析赋权方法; 对于第二评价层, 采用熵值法赋权。模型建立完成后, 本文依据三个维度在三维空间进行聚类分析, 将测试者分类。最后, 还需要对模型进行稳定性检验, 测试模型进行心理健康状况分析的能力。

5.2.1 三种量表的联系

在给出综合评价指标体系的方案前, 搞清楚三种量表之间的联系很重要。而探究三种量表之间的联系, 最直接的就是探究他们参数之间的联系, 这里我们可以对指标变量进行相似性度量。

记数据点 x_j 的为 $(x_{1j}, x_{2j}, \dots, x_{nj})^T$, 则可以用两变量 x_j 与 x_k 的样本相关系数作为他们的相似性度量。

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}} \quad (5)$$

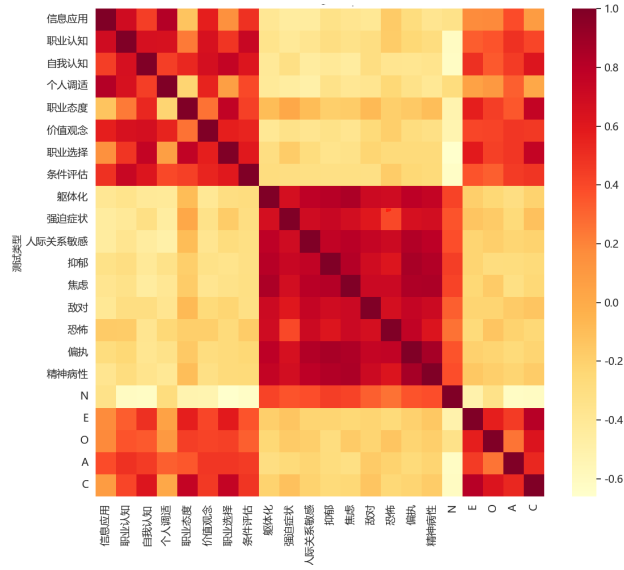


图 4 热力矩阵

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (6)$$

22 个参量分别两两计算得到相关系数矩阵及其热力矩阵如图 4 所示:

观察热力系数矩阵可以发现表内的联系明显大于表间的联系, 其中职业成熟度量表与 SCL90 表之间呈现负相关的关系, 与大五人格度量表之间呈现正相关的关系。通过阅读文献 [5] 可以得知, 长时间的抑郁症患者的 SCL90 评分也相对较高, 会影响我们的生理状况, 进一步影响其他心理状态。在职业成熟度量表上的体现为评分下降, 在大五人格量表的体现为评分偏离最佳评分, 颜色均很浅。

最大系数法 解读相关系数矩阵, 探究不同变量之间的相似性和关系, 还可利用**最大系数法**将因素进行**变量聚类**, 发现隐藏在数据中的结构, 从而帮助理解数据的内在联系并为进一步的分析和建模提供依据。在高维数据集中, 变量之间可能存在冗余。通过聚类分析, 可以将相似的变量归为一类, 从而简化数据结构, 减少维度, 提高分析效率。

在最大系数法中, 定义两类变量的距离为

$$R(G_1, G_2) = \max_{x_j \in G_1, x_k \in G_2} \{r_{jk}\}, \quad (7)$$

其中 $R(G_1, G_2)$ 表示两类变量 G_1 和 G_2 之间的距离。 $\max\{r_{jk}\}$ 表示在所有 $x_j \in G_1$ 和 $x_k \in G_2$ 中, 最大的一对 r_{jk} 的值。

根据距离, 使用聚类算法得到变量的分类图为图 5

可看出经过变量聚类变量明显分为了九类。

以病症变量聚类结果分析来看, 躯体化、人际关系、焦虑这三大指标有很强的相似

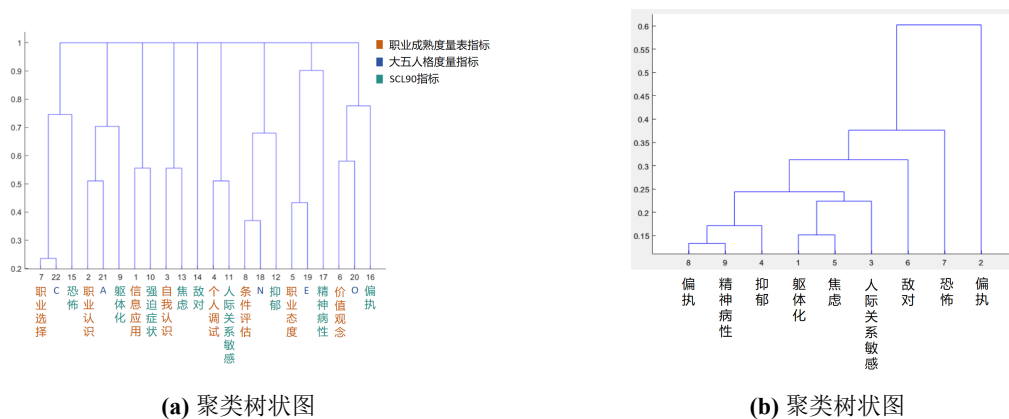


图 5 聚类树状图

性和重叠度；抑郁、偏执、精神病有很强的相似性和重叠度，分为一类；而强迫、敏感、敌对、恐怖自成一派。

这个分类结果，为后面的综合评价体系选择和构建指标奠定了基础。

5.2.2 评价体系的建立

心理测试具有间接测量的性质，其评价标准不如物理测量那样绝对和普遍，不能拿一个精确的数学化的评价模型来进行评价，所以本文基于模糊综合评价的思想，对不同的指标，依据它们不同的重要程度，就附上不同权重，加和得出来总分。

指标赋权方法 权重是评价模型中反映各评价指标相对重要程度的关键参数。指标赋权方法主要分为主观赋权、客观赋权两类。主观赋权因过度依赖经验判断而缺乏一定客观性，客观赋权因忽略指标间相关性而难以结合实际，本文模型结合主观赋权和客观赋权，采用层次分析-熵值组合赋权法对各因素对个体的评价能力进行定性-定量分析，使其权重同时反映主观经验和客观数据。

对于本题，我们建立如图的评价体系，综合考虑三种量表的影响。

人才综合评价体系分为两层，第一评价层和第二评价层。第一评价层包含三个维度，症状评价，职业成熟评价与人格评价结果。第二评价层是对三个维度展开，提供了更加细分的指标，对于部分有较高的同质性的指标，考虑综合为一个指标，减少它的维度。对于第一评价层，采用层次分析的赋权方法，对于第二评价层采用熵值法赋权。

5.2.3 数据预处理

在利用三种量表中的数据之前，我们需要对三种量表中的数据进行预处理。举例说明：对于症状自评量表 SCL-90 中，强迫症状、人际关系敏感、抑郁、焦虑、敌对、恐怖、偏执、精神病性 8 大指标都是极大型指标，指标数值越大越好。而躯体化是极小型指标，指标数值越小越好。对于职业成熟度量表，信息应用、职业认知、自我认知、个

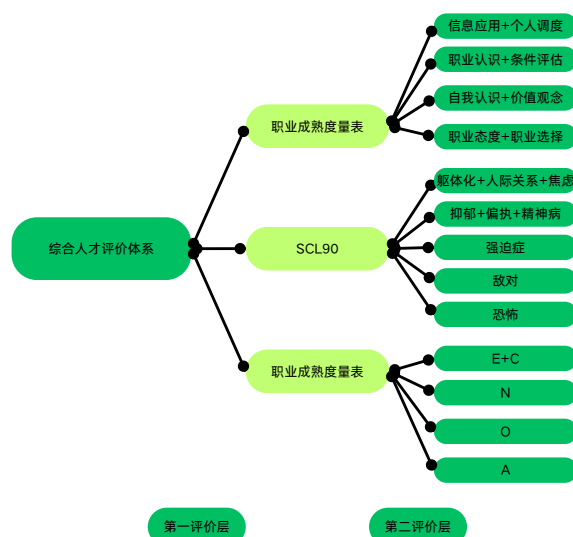


图6 人才评价体系

人调适、职业态度、价值观念、职业选择、条件评估这8大指标都是极大型指标。对于大五人格量表，N（神经质），E（外向性），O（开放性），A（宜人性），C（严谨性）这五大指标是中间型指标，标数值越接近某个值越好。

可以看出各种数据纷繁复杂大小不一，最优值也并非有明确的指标、明确的方向，那么如何综合的考虑这些指标，如何利用好这些数据呢？

正向化 有的数据是越大越好，有的数据是越小越好，有的数据是靠近某个值越好，这种不同的方向和区间让分析变得混乱，为了简化分析我们将数据进行正向化处理，都让他越大越好。

对于极大型指标、极小型指标、中间型指标这三种指标分别采用不同的正向化处理方法

极小型指标转化为极大型指标：

$$\bar{x}_i = \frac{1}{x_i} \quad (8)$$

中间型指标转化为极大型指标：

$$\bar{x}_i = 1 - \frac{|x_i - x_{best}|}{\max(|x_i - x_{best}|)} \quad (9)$$

经过这样的正向化处理后，成功将数据都变为越大越好的数值类型，方便之后的运算。

标准化 经过了正向化后，还存在一个问题就是所有的值都有他的量纲，对于每一列的数据进行标准化的方法如下：

$$z_i = 1 - \frac{x_i}{\sqrt{\sum_{i=1}^n x_i^2}} \quad (10)$$

5.2.4 熵权法

经过了正向化和标准化的修正之后，剩下的步骤就是进行评分指标的构建。熵权法构建系数利用了一定的信息论的知识，通过这种方法，可以量化各指标的重要性，使得综合评分更加客观和科学。

信息熵 是信息论的基本概念，描述信息源各可能事件发生的不确定性。是一种基于信息熵的多指标评价方法。熵权法利用信息论中的熵来确定各指标的权重。数据的变异程度越大，说明该指标包含的信息量越多，因此该指标的重要性也就越大。

信息熵的计算公式：

$$H_j = -\frac{1}{\ln(n)} \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (11)$$

各指标权重 w_j 的公式：

$$w_j = \frac{1 - H_j}{\sum_{j=1}^m (1 - H_j)} \quad (12)$$

生成权重系数为：

	信息应用	职业认知	自我认知	个人调适	职业态度	价值观念
W	0.0121	0.0123	0.0139	0.0115	0.0466	0.0101
	职业选择	条件评估	躯体化	强迫症状	人际关系敏感	抑郁
W	0.0212	0.0064	0.0437	0.046	0.0465	0.0221
	焦虑	敌对	恐怖	偏执	精神病性	N
W	0.0410	0.0402	0.0271	0.0374	0.0365	0.1267
	E	O	C	A		
W	0.1024	0.1037	0.0447	0.1480		

表 3 权重矩阵

5.2.5 层次分析法

第一层中使用层次分析法赋权，是主观的分析方法。分析一层指标（scl，职业成熟度量表，大五人格度量表）对人才综合评价体系的影响权重，需要对人才综合评价体系中 3 个一层指标判断矩阵，通过观察相关系数矩阵并查阅相关资料 [7][8][9] 可得：

	T1	T2	T3
T1	1	$\frac{4}{5}$	$\frac{9}{10}$
T2	$\frac{5}{4}$	1	$\frac{4}{5}$
T3	$\frac{10}{9}$	$\frac{5}{4}$	1

表 4 指标判断矩阵

计算得最优权重向量:

w	0.1556	0.6832	0.1612
---	--------	--------	--------

表 5 权重数据

总分计算公式 通过对数据的预处理，一二级权重的计算，我们可以得到总分的计算公式。

给出综合评分的公式:

$$\mu = \sum_{j=1} w_i \left(\sum_{i=1} w_{ij} \mu_i \right) \quad (13)$$

5.2.6 结果可视化展示

通过综合评价体系，我们得到从综合、职业成熟度、SCL90 和大五人格四个角度对 100 余位测试者的评分，详细评分结果见附录。同时我们将其可视化为堆叠面积图如下:

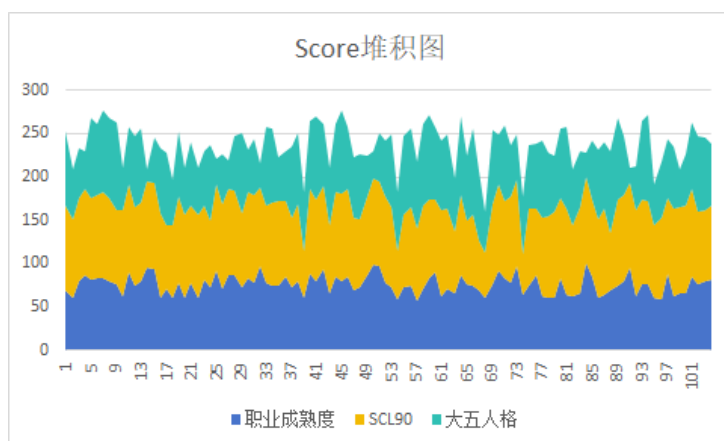


图 7 堆叠面积图

5.2.7 聚类分析

我们将处于第一分析层的职业成熟度、SCL90 和大五人格作为三项指标对 106 位测试者进行聚类，使用的 K-means 方法在之前已经讲述。通过肘部法，我们可将这些测试

者分为 4 类，其中心特征值如表 6。

	类别 1	类别 2	类别 3	类别 4
职业成熟度	77.426	71.289	92.873	66.820
SCL90	91.461	94.233	98.386	64.206
大五人格	87.676	63.007	38.201	71.619

表 6 中心特征值

样本的可视化三维展示图如图 8：

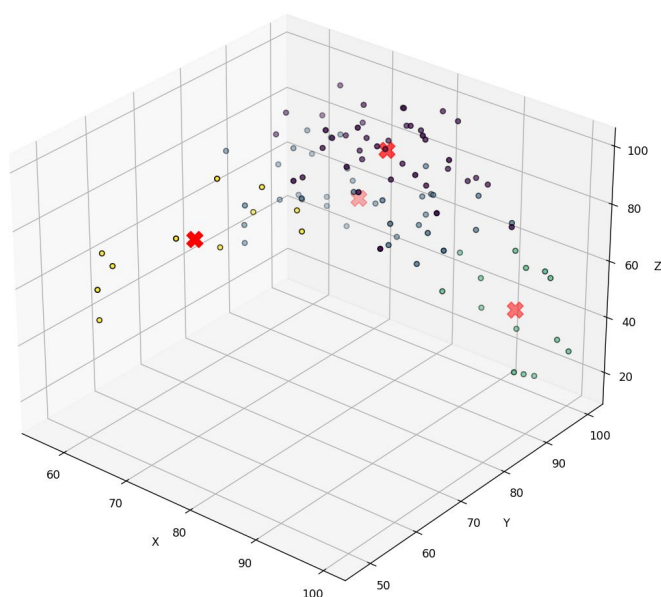


图 8 聚类分析三维图

5.2.8 稳定性检验

在实际当中，对于个体其心理因素受到心情和环境的影响，每次问卷调查的结果不可能严格的相等，实际存在的心理得分会存在一个波动范围，为检验模型的准确性，假设实际的问卷调查个体的每项指标得分存在上下 5% 的范围浮动，取样本中心作为参考，检验结果如下：

查阅 SCL90 指标得分超越 2 时被认作心理存在轻微病态，故当变化幅度较大时得分变化幅度误差大。

通过直接对比原总分发现我们的总分将人之间分差扩大，更容易分析出心理健康状况做出评价。

	变化幅度	样本中心	得分变化幅度
职业成熟度量表	5%	76.2	3.54%
	10%		6.45%
SCL90	5%	90.5	3.4%
	10%		13.2%
大五人格度量	5%	70.5	1.7%
	10%		4.6%

表 7 变化幅度与得分变化

5.3 问题三求解

分析 A 组测试者与 B 组测试者之间有无差异，需要把人群分为两组，分析两个群的之间数字特征的差异，来判断有无明显差异。可以比较的数字特征有两类样本之间的均值，我们可以采用独立样本 t 检验；还可以判断两类样本的样本方差是否相等，用到 Levene 检验，它非参数，不依赖于数据的分布特性。

5.3.1 独立样本 t 检验

独立样本 t 检验（Independent t-test）是一种统计方法，用于比较两个独立样本的均值，以判断它们是否有显著差异。它假设这两个样本来自具有相同方差的正态分布总体，但来自不同的个体（即，样本之间是独立的）。这种检验通常用于比较两个不同组（如实验组和控制组）的平均值。

1. 提出假设：零假设（ H_0 ）：AB 两个群体的均值相等，即没有显著差异。备择假设（ H_1 ）：AB 两个群体的均值不相等，即存在显著差异。

2. 计算 t 统计量：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (14)$$

其中 \bar{X}_1 和 \bar{X}_2 分别是 AB 两个群体总分的均值。 n_1 和 n_2 分别是两个群体的样本量。 s_1 和 s_2 分别是两个群体的标准差。 s_p 是两个群体的合并标准差，计算公式为：

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (15)$$

3. 确定自由度：自由度（df）计算公式为：

$$df = n_1 + n_2 - 2$$

4. 查找临界值：根据显著性水平（如 0.05）和自由度，在 t 分布表中查找临界值。

5. 做出决策：如果计算得到的 t 值的绝对值大于临界值，则拒绝零假设，认为两个群体的均值有显著差异。如果计算得到的 t 值的绝对值小于临界值，则不能拒绝零假设，认为两个群体的均值没有显著差异。

5.3.2 Levene 检验

Levene 检验 的原假设是各组数据样本来自于方差相等的总体，备择假设则是至少有一组数据样本来自于方差不等的总体。通过计算各组数据的离差（例如，观察值与该组的中位数的绝对偏差），Levene 检验可以得出一个统计量，并基于该统计量计算出一个 p 值。如果 p 值显著小于设定的显著性水平（通常设为 0.05），则可以拒绝原假设，认为 AB 两组样本来自于方差不等的总体，即样本方差不等。

1. 计算组内离差（group deviations）：对于每一组数据，计算每个数据点与该组的中心趋势（通常是中位数）的偏差的绝对值。例如，对于第 i 组数据 X_{ij} ，其中 \bar{X}_i 是第 i 组的中心趋势（如中位数），则离差计算如下：

$$|X_{ij} - \bar{X}_i|$$

对每组数据计算上述离差。

2. 计算离差的平均值：对每组数据计算离差的平均值：

$$M_i = \frac{1}{n_i} \sum_{j=1}^{n_i} |X_{ij} - \bar{X}_i|$$

其中 n_i 是第 i 组的样本数量。

3. 总体离差平均值：计算所有组的离差平均值的平均数：

$$M = \frac{1}{k} \sum_{i=1}^k M_i$$

其中 k 是组的数量。

4. Levene 检验统计量：根据上述计算的离差平均值，可以得到 Levene 检验的统计量：

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^k n_i (M_i - M)^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (|X_{ij} - \bar{X}_i| - M_i)^2} \quad (16)$$

其中 k 表示组的数量, $k=2$. \bar{X}_i 表示第 i 组的中心趋势 (如中位数). M_i 表示第 i 组的离差平均值. M 表示所有组的离差平均值。

5.3.3 结论

	小组 (均值 \pm 标准差)		独立性 t 检验		levene 方差齐性检验
	A	B	t	p	p_levene
综合评分	76.946 \pm 14.462	79.953 \pm 13.013	-1.115	0.267	0.260
职业成熟度	10.489 \pm 78.791	73.7073 \pm 11.156	-1.622	0.108	0.178
SCL90	92.502 \pm 9.067	13.180 \pm 88.567	1.767	0.080	0.100
大五人格	67.450 \pm 20.478	73.516 \pm 17.599	2.392	0.018	0.434

表 8 统计数据表

独立样本 t 检验的使用有两个前提条件:

1. 独立性——各观察值间相互独立; 如本案例中 AB 组对象都是独立的, 不存在相互干扰。
2. 方差齐性——两样本对应的两总体方差相等。

levene 方差齐性检验 从表格得出, 不同组别 A 与 B 的测试者对于综合评分、职业成熟度评分、SCL90 评分、大五人格评分全部均不会表现出显著性 ($p>0.05$), 意味着不同组别样本数据的波动性均呈现出一致性, 通过方差齐检验, 即两样本对应的两总体方差相等。因此, 两样本对应的两总体方差相等, 可以放心使用独立样本 t 检验进行分析。

独立样本 t 检验 从表格得出, 利用独立样本 t 检验去研究试验组 A 和对照组 B 对于综合评分的差异性, 得到 $p=0.267>0.05$, 所以不拒绝原假设, 即差异无统计学意义, 尚不能认为 A 组与 B 组的测试者有差异; 对于职业成熟度评分的差异性, 得到 $p=0.108>0.05$, 所以不拒绝原假设, 尚不能认为 A 组与 B 组的测试者有差异; 对于 SCL90 评分的差异性, 得到 $p=0.080>0.05$, 所以不拒绝原假设, 尚不能认为 A 组与 B 组的测试者有差异; 对于大五人格评分的差异性, 得到 $p=0.018<0.05$, 所以拒绝原假设, 即可以认为 A 组与 B 组的测试者有差异。

综上, 虽然从综合评分、职业成熟度评分、SCL90 评分三个维度都不能认为 A 组与 B 组的测试者有差异, 不过从大五人格评分来看, A 组与 B 组的测试者仍然存在差异。所以 A 组与 B 组的测试者是有差异的。

5.4 问题四求解

5.4.1 敏感性分析

经过前三问的解决，已经将所有指标都进行了正向化处理，消除了 SCL90 极小化指标对总分所造成的影响；还通过标准化处理消除了三张表不同量纲的影响，数据标准化可以减少数值计算过程中的数值不稳定性，提高模型的表现和鲁棒性。

表 9通过观察不同变化幅度下量表得分的变化，评估各量表对变化的敏感程度。表现了衡量施加不同变化幅度后，各量表得分的具体变化情况，以评估不同量表的稳定性和响应性。由得分变化幅度 3.4% 小于 5%，可以说明模型的变化程度不高，模型的表现

	变化幅度	样本中心	得分变化幅度
SCL90	5%	90.5	3.4%
	10%		13.2%

表 9 数据表

和鲁棒性强。

5.4.2 特殊影响的显著反应

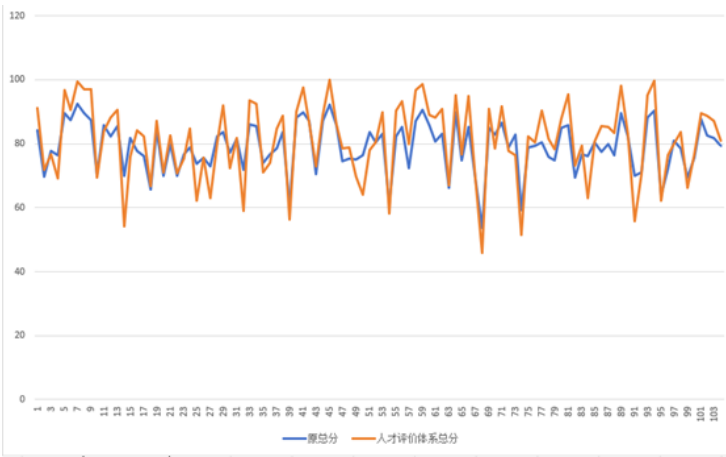


图 9 总分比较

查阅 SCL90 指标得分超越 2 时被认作心理存在轻微病态，故当变化幅度较大时得分变化幅度误差大。当变化幅度变大的时候我们有更大的理由认为该人的心理病态几率变大，得分变化幅度也应当变大。如图 9 该模型可以更好的反应 SCL90 对人才评分的影响。而 SCL90 中给出的总分则不能很好的说明这一点，92 分既可以代表一名心理健康的人也可以代表一个某项心理指标存在病态的人。

六、模型评价与推广

该综合评价体系的建立不仅能有效整合不同量表的信息，还能为个体提供全面而精确的评估结果。在症状评价方面，通过系统地分析和量化个体的心理症状表现，可以帮助精准识别问题并提供针对性的干预措施。职业成熟评价则关注个体在职场环境中的表现和适应能力，评估其专业技能、决策能力以及团队合作等方面的表现。而人格评价则深入分析个体的人格特征，包括情绪稳定性、责任感和社交能力等，为个体发展和职业选择提供重要参考。

这一综合评价体系不仅具有理论深度，还注重实际应用和推广。通过科学的探究方法和广泛的实证研究支持，确保了评估结果的客观性和可靠性，为不同领域的人才选拔、心理健康干预和职业发展提供了有力支持。随着社会和个体需求的不断变化，该评价体系也在不断优化和更新，以满足多样化、个性化的评估需求，促进社会各界的人才发展和个体成长。

七、参考文献与引用

参考文献

- [1] Derogatis, L. R. (1994). SCL-90-R: Symptom Checklist-90-R. Pearson.
- [2] Savickas, M. L. (1999). The Career Development Quarterly.
- [3] John, O. P., Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives.
- [4] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics (pp. 281-297). University of California Press.
- [5] Pearson, K. (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling."
- [6] 蔡秀算. 认知行为护理策略对老年抑郁症伴高血压患者 SCL 90 评分及血压水平的影响 [J]. 心血管病防治知识, 2023, 13(28): 55-57.
- [7] 邱林, 郑雪, 王雁飞. 积极情感消极情感量表 (PANAS) 的修订 [J]. 应用心理学, 2008, 14(03): 249-254+268.

- [8] 王若逸, 苏永强, 陈朝阳, 等, 无法忍受不确定性与抑郁倾向关系: 情绪调节困难的中介作用 [J]. 宁波大学学报 (教育科学版), 2017, 39(03): 10-14.
- [9] 汪玥, 张豹, 周晖. 中小學生正念注意觉知与心理健康: 情绪调节和积极重评的跨时间中介作用 [J]. 心理发展与教育, 2022, 38(05): 692-702. DOI: 10.16187/j.cnki.issn1001-4918.2022.05.10.

附录 A 支撑材料内容组成

文件夹	文件名	主要功能/用途
源代码	kmeans.py	进行 Kmeans 聚类分析
	feature.py	聚类后的标签汇总
	t_test.py	对数据进行独立性 t 检验
	f_test.py	对数据进行卡方检验
	Cov.py	求解相关系数矩阵
	Zhengxiang.m	将指标得分正向化
	Biaozhunhua.m	将正向化系数标准化
	Shangquan.m	利用熵权法求解第二评价层权重
	Score.m	求解个人总得分
	MDI.m	求解第一评价层权重系数
数据	disposed_data.xlsx	该表格是原始数据表格，包含所有填写三个调查问卷的人的数据
	Disposed_data_alltime.xlsx	该表格包含所有填写三个调查问卷的人的数据，剔除时间低于 1min
	Cov.xlsx	22 个指标的全相关系数矩阵
	RESULTzhengxiang.xlsx	将所有指标下的数据正向化
	Score.xlsx	人才评价模型下所有人总分
	Score_for_each.xlsx	人才评价模型下第一评价层中的每项总分
	feature.xlsx	所有测试者的聚类标签数据

表 10 支撑材料内容组成