

基于因子分析和逻辑回归的母婴关系的量化与分类模型

摘要

母亲的身心健康会对婴儿成长产生深远影响。积极的母亲情绪与心理状态和良好的母婴关系有助于建立安全的情感连接，促进婴儿情感调节和培养婴儿的社交能力。因此关心母亲的身心健康和母婴关系和对产后母亲的身心健康的量化与干预是实现优生优育的关键。

问题一需要研究母亲的**身体指标**和**心理指标**对婴儿的**行为特征**和**睡眠质量**的影响。为此本文先对题目提供的年龄、妊娠时间和心理评估的数据进行正向化，将二元指标婚姻情况、分娩方式表示成判别函数，从而将原始指标赋分，再利用 **Spearman 相关性检验**得到各指标间的相关系数。结论发现各指标间相关性不显著。

问题二需要本文在问题一的基础上量化各指标与婴儿的行为特征间的关系，并评价第 391-410 号婴儿的行为特征。为此本文对问题一中已经进行赋分对指标进行因子分析，抽象出**母亲心理状态因子** F_1 、**母亲身体状态因子** F_2 、**婴儿发育情况因子** F_3 三个指标。将三个因子与婴儿行为特征 y_1 进行多重线性回归得到 $y_1 = 0.5569F_1 + 0.2220F_2 + 0.2211F_3$ ，经检验该因子残差矩阵方差较小。评价结果在论文中给出。

问题三需要在第二问的基础上对**母亲心理状态因子**进行最小代价优化。题目提供的条件揭示了优化函数的线性性质，因此采用线性规划的方法求得目标函数最小值。然而由于治疗一种疾病需要一个起始代价，因此在线性规划的外层使用枚举的方式考察各种治疗组合的最小代价。最终得到将第 238 号婴儿的行为特征转变为“中等型”的最低价格为 2750 元，转变为“安静型”的最低价格为 5647 元。

问题四与问题二相似，目的是量化各指标与婴儿的综合睡眠质量间的关系，并将数据集中的婴儿睡眠质量进行分类。因此在使用因子多重线性回归前，先使用 **Q 型聚类分析**得出睡眠质量的临界得分，然后对回归结果进行分类，评价结果在论文中给出。

问题五与问题三相似，目的是在问题三对基础上对婴儿睡眠质量进行优化。因此此问由单目标线性规划问题转变为**双目标线性规划问题**。本文在第三问的改良基础上增加约束条件构建型的规划模型。由于第 238 号婴儿为**极早产儿**，因此需要付出较高价格，最少为 24515 元。

本文构建的优点在于合理的预处理、简洁高效和易用等。然而，评价指标不足、文献提供的评价方式欠缺等影响了本文评价的客观性，指示了本文进一步改进的方向。

关键字： Spearman 相关性检验 因子分析 多重线性回归 线性规划 聚类分析

目录

一、问题重述	4
1.1 问题背景	4
1.2 问题提出	4
二、模型的假设	5
三、符号说明	5
四、问题 1 的模型建立与求解	5
4.1 对问题 1 的分析	5
4.2 数据预处理	6
4.3 基于因子分析的评价模型建立	7
4.4 问题 1 的模型求解	9
五、问题 2 的模型建立与求解	10
5.1 对问题 2 的分析	10
5.2 基于问题 1 结果建立逻辑回归模型	10
5.3 问题 2 的模型求解	11
六、问题 3 的模型建立与求解	12
6.1 对问题 3 的分析	12
6.2 问题 3 的模型求建立	12
6.3 问题 3 的模型求解	14
七、问题 4 的模型建立与求解	14
7.1 对问题 4 的分析	14
7.2 基于 Q 型聚类分析的模型建立	14
7.3 问题 4 的模型求解	15
八、问题 5 的模型建立与求解	17
8.1 对问题 5 的分析	17
8.2 问题 5 的模型建立	17
8.3 问题 5 的模型求解	17

九、模型评价	18
9.1 模型的优点	18
9.2 模型的缺点	18
9.3 模型的改进	18
参考文献	19
附录 A 相关代码	20
1.1 问题求解	24
1.2 结果呈现	31

一、问题重述

1.1 问题背景

母亲在婴儿生命中扮演着至关重要的角色，不仅为婴儿提供营养物质和身体保护，还为婴儿提供情感支持和安全感。然而，母亲心理健康状态的不良状况，如抑郁、焦虑、压力等，可能会对婴儿的认知、情感、社会行为等方面产生负面影响。特别是，压力过大的母亲可能会对婴儿的生理和心理发展产生负面影响，例如影响其睡眠质量。

研究母亲身心健康对婴儿成长的影响对于促进婴儿的健康发展至关重要，是一项重要的课题。通过深入研究这一领域，我们可以更好地理解母亲的角色，并为提供适当的支持和干预措施提供指导，进而对婴儿的成长带来积极影响。通过数学建模，我们可以量化不同类型的母亲身心健康状况对婴儿发展的潜在影响程度，并为相关领域的研究和实践提供科学依据。

1.2 问题提出

围绕母亲的身心健康状况和婴儿的成长情况进行分析并建立数学模型，本文依次解决如下问题：

问题 1：根据附件中的数据研究母亲的身体指标和心理指标对婴儿的行为特征和睡眠质量是否有影响。

问题 2：基于婴儿行为问卷，我们将婴儿的行为特征分为安静型、中等型、矛盾型三种类型。根据附件数据表中给出的前 390 组数据建立婴儿的行为特征与母亲的身体指标与心理指标的关系模型，并依据此模型对表中最后 20 组（编号 391-410 号）行为特征信息被删除的婴儿的行为特征类型进行判断。

问题 3：对母亲焦虑的干预不仅有助于提高母亲的心理健康水平，还可以改善母婴交互质量，进而促进婴儿的认知、情感和社交发展。CBTS、EPDS、HADS 的治疗费用相对于患病程度的变化率均与治疗费用呈正比。根据调研结果给出的两个患病得分下对应的 CBTS、EPDS、HADS 治疗费用，建立模型并分析能够使编号为 238、行为特征为矛盾型的婴儿的行为特征从矛盾型变为中等型或安静型分别需要采取何种治疗策略使得治疗费用最少。

问题 4：婴儿的睡眠质量可以通过整晚睡眠时间、睡醒次数和入睡方式三个指标划分为优、良、中、差四类。根据附件中的数据对婴儿的睡眠质量进行综合评判，并根据婴儿的睡眠质量的综合评判结果建立婴儿综合睡眠质量与母亲的身体指标、心理指标的关联模型，最后依据此模型对表中最后 20 组（编号 391-410 号）婴儿的综合睡眠质量进行预测。

问题 5：在问题 3 的基础上，即在使得编号为 238 的婴儿的行为特征从矛盾型变为中等型或安静型的前提下，分析是否需要调整治疗策略使睡眠质量评级为优。

二、模型的假设

- 假设一：统计结果中的所有孕妇均为第一次分娩且均为一孩。
- 假设二：不考虑除统计结果中列举的因素以外其他因素 (如医护人员态度、个体人格类型、经济状况等) 对母亲身心健康状况的影响。
- 假设三：不考虑孕妇的个体差异，即将统计结果中指标一样的孕妇的计算结果相同。
- 假设四：统计结果中的心理量表和对婴儿行为特征的分类具有较高的效度和信度

三、符号说明

符号	意义
x_1	母亲年龄
x_2	婚姻状况
x_3	教育程度
x_4	妊娠时间
x_5	分娩方式
x_6/S_1	CBTS
x_7/S_2	EPDS
x_8/S_3	HADS
F_1	母亲产后心理状况
F_2	母亲产后身体状况
F_3	婴儿生长发育情况
y_1	婴儿的行为特征
y_2	婴儿的睡眠质量

四、问题 1 的模型建立与求解

4.1 对问题 1 的分析

问题 1 是一个评价题，本题需要我们根据附件中所给的 390 组完整数据（编号 1-390）来评价母亲的身体指标（包括年龄、婚姻状况、教育程度、妊娠时间、分娩方式）和心理指标（包括 CBTS、EPDS、HADS）对婴儿的行为特征和睡眠质量是否有影响。我们可以通过因子分析研究这八个原始变量之间的内部依赖关系，探求观测数据“婴儿的行为特征和睡眠质量”中的基本结构，并判断能否用少数几个假想变量（即公共因子）

来表示其数据结构。如果模型求解过程中能够选择两个公共因子：母亲的身体指标和心理指标，且满足它们对样本方差的累积贡献率能达到 75%，则可以表明母亲的身体指标和心理指标对婴儿的行为特征和睡眠质量有影响。

4.2 数据预处理

(一) 筛掉不满足原则的样本

通过观察附件表中的数据，我们可以看到编号 390-410 号的部分数据存在缺失；编号 181 号中婴儿整晚睡眠时间为 99 时 99 分 99 秒，属于异常值；编号 43、95、134、196、231、301、306、308、355 号的婚状况为 3 或 6，根据附表中的补充说明知，婚姻状态只用 1 和 2 表示，所以这 9 组数据也属于异常值。因此，为便于后续的数据分析和建模。我们对这 30 组数据进行清洗以提高数据的质量和可用性。

(二) 指标正向化

对于生育年龄这一指标，查阅文献可知，女性的最佳生育年龄为 22 岁到 32 岁左右，在小于 18 岁和大于 45 岁时不适合生育。因此不区分处于最佳生育年龄的女性的年龄对生育的影响，而使用 logistic 曲线拟合低龄产妇与超龄产妇的年龄对生育的影响。根据拟合结果，可以用下面的分段函数对母亲年龄这一指标进行量化。

$$scoreAge = \begin{cases} 0.2 + \frac{0.8}{1+e^{\frac{x-20}{2}}} & x < 22 \\ 1 & 22 \leq x < 29 \\ 1 - \frac{0.8}{1+e^{\frac{x-40}{5}}} & x \geq 29 \end{cases}$$

对于妊娠时间这一指标，查阅文献可知，最佳妊娠时间为 37 周到 43 周左右，妊娠时间小于 37 周的婴儿和大于 43 的婴儿在医学上分别被称之为早产儿和过期产儿，均会对母婴产生一定的不良影响。因此不区分处于最佳妊娠时间对生育的影响，而使用 logistic 曲线拟合妊娠时间不足和妊娠时间过久对生育的影响。根据拟合结果，可以用下面的分段函数对对妊娠时间这一指标进行量化。

$$scoreTime = \begin{cases} 0.4 + \frac{0.6}{1+e^{-0.7(x-31)}} & x < 37 \\ 1 & 37 \leq x < 43 \\ 1 - \frac{1}{1+e^{-2(x-45)}} & x \geq 43 \end{cases}$$

对婚姻状况和妊娠方式这两个指标采取判别函数进行重新量化

$$scoreMarr = \begin{cases} 1 & \text{已婚} \\ -1 & \text{未婚} \end{cases}$$

$$scoreMethod = \begin{cases} 1 & \text{自然分娩} \\ -1 & \text{剖宫产} \end{cases}$$

对反映心理指标的这三个量表 CBTS、EPDS、HADS 的分值进行重新量化

$$scoreForm = 30 - x \quad x \text{ 为量表分值}$$

对婴儿的行为特征这个指标进行量化

$$scoreMethod = \begin{cases} 1 & \text{安静型} \\ 0 & \text{中等型} \\ -1 & \text{矛盾型} \end{cases}$$

(三) 对婴儿的睡眠质量进行量化

与婴儿睡眠质量相关的指标有三个：整晚睡眠时间、睡醒次数和入睡方式。由于入睡方式这一指标相对于其他两个指标而言非常独立、离散；所以在量化的时候与另外两个连续性较强的变量做了一个区分。

通过查阅资料，可以了解到一般可以用睡眠效率 η 这一标准来衡量夜晚睡眠质量。

首先，我们可以通过整夜睡眠时间和睡醒次数可以得到婴儿平均一次睡眠时间 T

$$T = \frac{\text{整晚睡眠时间}}{\text{睡醒次数} + 1} = \frac{T_0}{A + 1}$$

然后通过查阅文献可知，不同年龄的婴儿入睡潜伏期 $t = \begin{cases} 42min & 1 \text{ 个月} \\ 37min & 2 \text{ 个月} \\ 24min & 3 \text{ 个月} \end{cases}$

最后可以得到睡眠效率表达式 η

$$\eta = 1 - \frac{t \times A}{T_0}$$

接下来，分析入睡方式。对附件中数据进行分析可知，哄睡法睡醒中位为 2，环境营造法睡醒中位为 0，其余三种方式为 1，而且通过查阅资料可知，使用环境哄睡法和恶定时法会使婴儿有更好的自主性。所以，我们对不同的入睡方式进行赋值：

环境营造法($w = 2$) > 定时法($w = 1.5$) > 安抚奶嘴法($w = 1$) = 抚摸($w = 1$) > 哄睡法($w = 0$)

综上，结合三个指标，最终用下面的表达式来量化婴儿睡眠质量

$$K = T \cdot \eta + w$$

4.3 基于因子分析的评价模型建立

因子分析是基于统计学和概率论的基础上进行的。其主要思想是通过对观测变量之间的协方差矩阵进行分解，找到能够解释变量之间共同性的潜在因子（即公共因子）。

(一) 假设潜在因子存在, 即假设观测变量“婴儿的行为特征和睡眠质量”的变异可以归因于少数几个公共因子, 而不是每个变量都有独立的影响因素。

通过分析附件可知,影响“婴儿的行为特征和睡眠质量”的原始变量有 8 个,母亲年龄、婚姻状况、教育程度、妊娠时间、分娩方式和 CBTS、EPDS、HADS,依次记作 $x_i(1, 2, \dots, 8)$,这 8 个原始变量可能相关,也可能独立,将 x_i 进行标准化后,可建立因子分析模型如下

其中 $F_j(j = 1, 2, \dots, m)$ 为出现在每个变量中的公共因子，含义需要根据具体问题进行解释； $U_i(i = 1, 2, \dots, m)$ 是仅与变量 x_i 有关的特殊因子；系数 $a_{ij}(i = 1, 2, \dots, j = 1, 2, \dots, m)$ 为因子载荷，反映第 i 个变量 x_i 依赖第 j 个因子的份量或比重， $A = (a_{ij})$ 为载荷矩阵。

通过计算相关系数矩阵 R 的特征值 $\lambda_1、\lambda_2 \dots \lambda_8$ 及其对应的特征向量 $\mu_1、\mu_2 \dots \mu_8$, 进一步通过 I 下面公式计算得到初等载荷矩阵, 最后计算各个公共因子的贡献率, 通常因子的累积方差贡献率达到 60% 才符合提取公共因子的要求, 选择贡献率较高的 m 个因子作为主因子。

（四）进行因子旋转，即对提取得到的 m ($m < p$) 个主因子因子进行旋转操作，以使得每个因子与尽可能多的观测变量相关。常见的旋转方法有正交旋转（如 **Varimax** 旋转）和斜交旋转（如 **Oblique** 旋转）。通过因子旋转得到矩阵 $B = A_m \cdot T$ (其中 A_m 为 A 的前 m 列, T 为正交矩阵)。构造的因子模型转化为

（五）对因子进行解释，根据因子载荷矩阵，解释每个因子与每个观测变量之间的关系。其中载荷矩阵中的元素表示每个因子对应的观测变量的权重。

4.4 问题 1 的模型求解

(一) 构造因子分析

从附件表中，我们可以找到 8 个涉及母亲身体指标和母亲心理指标的原始变量，母亲年龄 (x_1)、婚姻状况 (x_2)、教育程度 (x_3)、妊娠时间 (x_4)、分娩方式 (x_5)、CBTS(x_6)、EPDS(x_7)、HADS(x_8)。目标变量为 y_1 婴儿的行为特征和 y_2 婴儿的睡眠质量。根据题意，我们希望找到一些潜在的公共因子（如 F_1 母亲产后心理状况、 F_2 母亲产后身体状况、 F_3 婴儿生长发育情况）并能够给予合适的解释。

(二) 计算相关系数矩阵

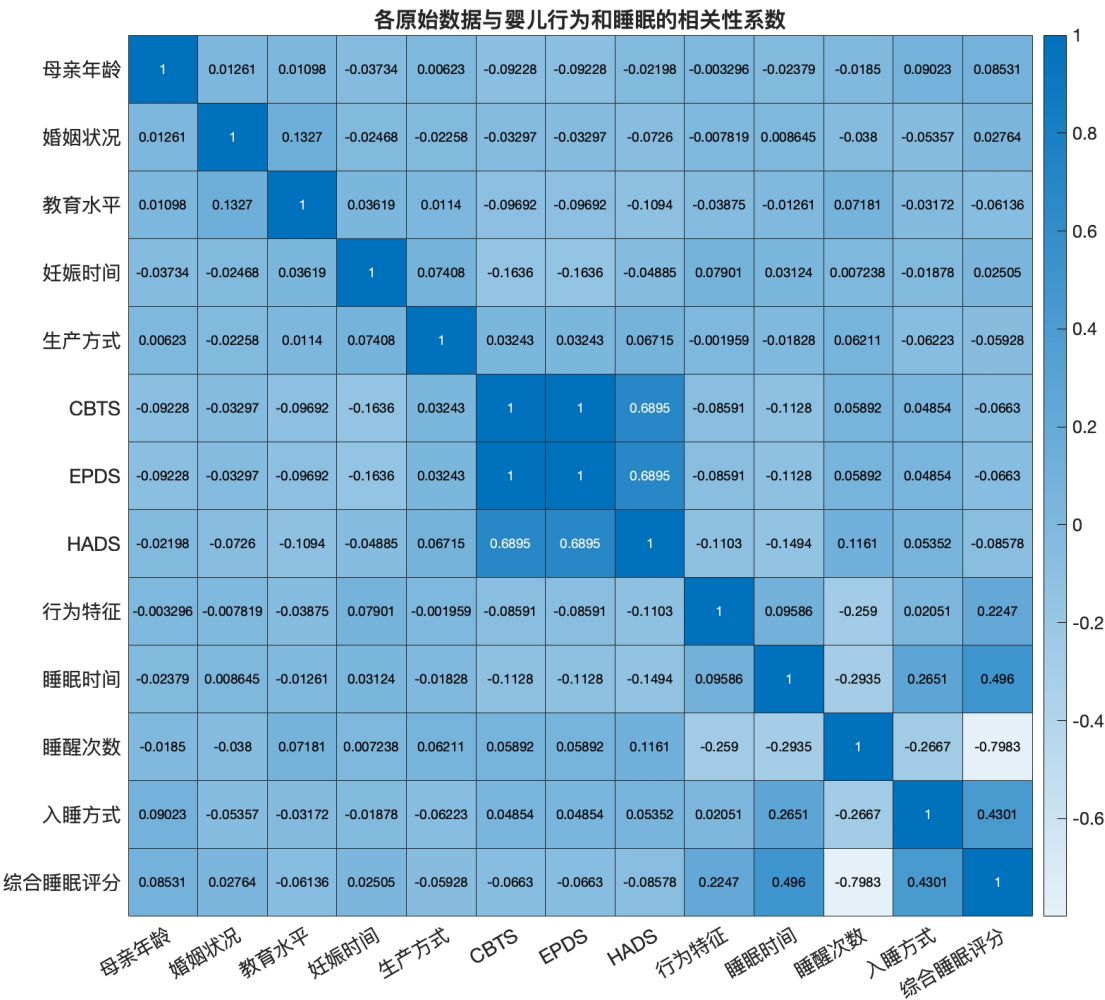


图 1 相关性系数

(三) 计算载荷矩阵并计算贡献率

在本题中，我们选取三个主因子，利用 MATLAB 程序计算得到旋转后的因子贡献及贡献率，见下表

变量	旋转因子载荷估计			共同度	特殊方差
	主因子1	主因子2	主因子3		
1	-0.0400	-0.7434	0.1289	0.0851	0.0057
2	0.0015	-0.0079	0.0566	0.6012	0.0243
3	0.1085	0.0805	0.9887	0.5703	0.0220
4	-0.0644	0.6355	0.0166	0.5396	0.0273
5	0.0043	0.0108	0.0041	0.4831	0.0281
6	0.0108	0.0391	-0.1189	0.9441	0.0028
7	0.7371	0.0566	0.1534	0.9441	0.0028
8	0.3144	0.0297	-0.0169	0.6998	0.0059
累计贡献率	0.4206	0.1675	0.1669		

图 2 因子分析表

通过观察表中数据，我们可以看出，变量 7、8（即 EPDS、HADS）在主因子 1 上有大载荷，而在 F_2 上在载荷较小或可忽略，理论上变量 6（CBTS）在 F_1 上有大载荷，但是由于 CBTS 和 EPDS 得相关性系数为 1，具有很强得相关性，二者会发生重合，所以结果较小。而变量 1 和 4（母亲年龄和妊娠状况）在 F_2 上有大载荷，而在 F_1 上在载荷较小或可忽略。因此我们有理由称 F_2 为母亲心理指标因子，在 F_1 为母亲身体指标因子。

对三个公共因子进行 Spearman 相关性检验，得到其相关系数矩阵

$$\mathbf{R} = \begin{pmatrix} 1 & -0.092053499 & -0.1136678 \\ -0.092053499 & 1 & -0.224717598 \\ -0.1136678 & 0.224717598 & 1 \end{pmatrix}$$

由此可以看出，经过因子旋转后三个公共因子的独立性强，能较好反应原指标对婴儿行为和婴儿睡眠质量的影响。

五、问题 2 的模型建立与求解

5.1 对问题 2 的分析

问题二是一个分类问题。本题需要先建立建立婴儿的行为特征与母亲的体身体指标与心理指标的关系模型，再依据模型对第 391-410 号数据进行分类。因此首先要使用合适的方法建立关系模型。经尝试 SVR（支持向量机回归）的回归效果较差，而一个人的各项身体、心理指标往往相互依赖、相互影响。因此，因此采用因子分析的方法对婴儿行为特征逻辑回归，求解回归函数。之后将回归系数带入第 391-410 号数据，得到这些样本的婴儿行为特征回归的分数。而在问题一的预处理中已经将三个类别的婴儿行为特征赋分为 1,0,-1, 对这些分数按照四舍五入取整得到分类结果。

5.2 基于问题 1 结果建立逻辑回归模型

（一）在问题 1 的基础上进行因子得分计算，即根据因子分析的结果，计算每个样本在每个因子上的得分。这些因子得分将作为逻辑回归的自变量。

(二) 建立逻辑回归模型就是将因子得分作为自变量, 根据婴儿行为特征的分类情况, 建立逻辑回归模型, 从而确定目标变量和自变量之间的关系。

$$Y = W \cdot F$$

其中 $W = (w_1, w_2, w_3)$, 表示这三个主因子在目标变量 $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ 中的权重。

(三) 对模型评估与优化, 使用合适的评价指标对逻辑回归模型进行评估, 根据评估结果, 可以对模型进行优化, 如调整自变量选择、正则化参数等。

(四) 根据最终确定的逻辑回归模型求解回归函数参数, 即可求解得到回归函数。可以使用最大似然估计等方法进行参数估计。

5.3 问题 2 的模型求解

在问题 1 的基础上, 我们得到因子旋转后的得分

表 2 旋转因子得分表

指标	常数	母亲心理状况因子	母亲身体状况因子	婴儿发育状况因子
婴儿行为	0.191116852	0.067186562	0.027829108	0.024029404
整夜睡眠时长	0.001567433	0.133750179	0.030041742	0.016198332
睡醒次数	1.463872995	0.143794162	0.046820535	0.110282611
入睡方式	3.046342453	0.066228111	0.111028566	0.028711301
综合睡眠指标	6.863520431	0.352821615	0.225271576	0.197885942

通过问题 1 中的载荷矩阵 A, 我们利用,ATLAB 将三个因子与婴儿行为特征 y_1 进行多重线性回归得到求解得到三个公共因子的权重分别为 $w_1 = (0.5569, 0.2220, 0.2211)$ 因此, 婴儿的行为特征指标 $y_1 = 0.5569F_1 + 0.2220F_2 + 0.2211F_3$ 。

对因子残差进行检验得到残差矩阵 R, 观察可以看到其方差较小, 共同度也相对较高, 模型较为合理。

方差	0.0057	0.0243	0.0220	0.0273	0.0281	0.0028	0.0028	0.0059
共同度	0.0851	0.6012	0.5703	0.5396	0.4831	0.9441	0.9441	0.6998

利用上述回归方程对最后 20 组 (编号 391-410 号) 婴儿的行为特征的预测结果如下:

编号	评分	行为特征
391	-1.581326401	矛盾型
392	-1.266486035	矛盾型
393	-0.469179441	中等型
394	0.983773326	安静型
395	0.268656657	中等型
396	0.004306753	中等型
397	-0.41876254	中等型
398	0.195895298	中等型
399	-0.499161101	中等型
400	0.431751682	中等型
401	-2.693111692	矛盾型
402	0.008282044	中等型
403	0.43524039	中等型
404	0.893321414	安静型
405	0.754973033	安静型
406	0.966602091	安静型
407	-0.45435461	中等型
408	0.978450367	安静型
409	1.441213753	安静型
410	0.019915011	中等型

图 3 行为特征预测结果

六、问题 3 的模型建立与求解

6.1 对问题 3 的分析

问题三是一个优化问题。本题需要在第二问得到回归系数的基础上对求得治疗心理障碍的最小优化。题目给出了每一种疾病的治疗价格，并揭示了治疗价格与疾病变化率线性相关，因此采用线性规划的方法求得目标函数最小值。然而由于治疗一种疾病需要一个起始代价，因此在线性规划的外层使用枚举的方式考察各种治疗组合的最小代价。最后提取第 238 号样本的得分数据，带入线性优化模型得出结论。

6.2 问题 3 的模型求建立

将 CBTS、EPDS、HADS 的患病程度分别定义为 S_1 、 S_2 和 S_3 ，已知 $F_1 = C_1S_1 + C_2S_2 + C_3S_3$ 。而由题可知 CBTS、EPDS、HADS 的治疗费用相对于患病程度的变化率均与治疗费用呈正比，分别记作 k_1 、 k_2 和 k_3 。其中，系数矩阵 A 已知，则每位母亲需要发生改变的心理状况指标 ΔH 已知；系数矩阵 C 已知，则 ΔH 确定时最优调整策略和最少治疗费用已知。

我们首先设定启动治疗这三种疾病有初始花费，记作 k_1 、 k_2 和 k_3 。由于题目中涉及到的疾病只有三种，选择结合枚举的线性规划模型来解决此问题。

第一步，枚举疾病治疗方案

$$\text{sigmoid}(i) = \begin{cases} 0 & \text{表示对第 } i \text{ 种疾病不予以治疗} \\ 1 & \text{表示对第 } i \text{ 种疾病予以治疗} \end{cases}$$

矩阵 Ω 治疗方案的全部集合，是 sigmoid^3 的幂集合，其范 $\|\Omega\|$ 为 7。

第二步，构建线性规划模型

$$\text{由问题 2 可知 } y = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} \cdot \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = W \cdot F,$$

$$\text{其中 } F = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{pmatrix} \cdot \begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix} = D \cdot S$$

所以 y 的表达式可化简为 $y = W \cdot D \cdot S$ ，其中 $W \cdot D$ 为常数， S 为自变量， y 为因变量。

(一) 找出约束条件和边界条件

已知使婴儿的行为特征从矛盾型变为中等型或安静型的 Δy 分别为 1, 2，可找到约束条件为

$$\Delta S = \frac{\Delta y}{W \cdot D} = \begin{pmatrix} C_1 & C_2 & C_3 \end{pmatrix} \cdot \begin{pmatrix} \Delta S_1 \\ \Delta S_2 \\ \Delta S_3 \end{pmatrix}$$

而矩阵 $|S + \Delta S_I| \leq 1$ 为边界条件。

(二) 建立带约束条件的单目标优化模型

$$\begin{aligned} \min \quad & \text{Cost} = \begin{pmatrix} \text{sigmoid}(1) & \text{sigmoid}(2) & \text{sigmoid}(3) \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \text{linprog}(k', \Omega, \Delta S) \\ \text{s.t.} \quad & \begin{cases} \Delta S = \begin{pmatrix} C_1 & C_2 & C_3 \end{pmatrix} \cdot \begin{pmatrix} \Delta S_1 \\ \Delta S_2 \\ \Delta S_3 \end{pmatrix} \\ \|S_i + \Delta S_i\| \leq 1 \end{cases} \end{aligned}$$

6.3 问题 3 的模型求解

当 $\Delta y = 1$, 即婴儿的行为特征从矛盾型变为中等型时的治疗方案 sigmoid^3 为 (1,0,0), 可知治疗策略为仅治疗 CBTS, 最少花费代价 $\text{Cost}_1 = 2750$ 。

当 $\Delta y = 2$, 即婴儿的行为特征从矛盾型变为中等型时的治疗方案 sigmoid^3 为 (1,1,0), 可知治疗策略为同时治疗 CBTS 和 EBDS, 最少花费代价 $\text{Cost}_2 = 5647$ 。

七、问题 4 的模型建立与求解

7.1 对问题 4 的分析

问题四也是一个分类问题。本题需要先建立综合睡眠评价指标, 而这个指标已经在问题一中的预处理中给出。之后将评价得分分为四个层级。由于缺少客观的分类标准, 因此采取 Q 型聚类分析根据数据集数据进行分析。最后, 同问题二, 建立婴儿的睡眠与母亲的身体指标与心理指标的关系模型, 再依据模型对第 391-410 号数据进行分类。

7.2 基于 Q 型聚类分析的模型建立

Q 型聚类分析是一种基于数据的聚类方法, 其原理主要是对样本进行相似性度量后: 多次进行类与类之间的相似性度量。Q 型聚类分析的目标是将相似的样本归为一类, 使得同一类内的样本相似度较高, 不同类之间的样本相似度较低。通过 Q 型聚类分析, 可以发现数据中的潜在结构和模式, 并为后续的数据分析和决策提供支持。

(一) 首先, 需要准备待聚类的数据集。该数据集包含多个样本, 每个样本具有多个特征。数据可以是数值型、离散型或混合型。在本题中, 经问题 1 数据预处理后的 380 组数据构成样本点集, 每个样本有 3 个特征: 整晚睡眠时间、睡醒次数、入睡方式。记样本点集 $\Omega = \{w_1, w_2, \dots, w_{380}\}$ 。

(二) 然后计算样本之间的相似度或距离。常用的相似度计算方法包括欧氏距离、曼哈顿距离、余弦相似度等。根据相似度计算结果, 构建一个相似度矩阵 D , 记录了每个样本与其他样本之间的相似度或距离。

(三) 进行聚类初始化, 即将每个样本看作一个初始聚类, 形成初始的聚类集合 $H = \{G_1, G_2, \dots, G_{380}\}$ 。

(四) 通过合并相似度最高的聚类来逐步形成更大的聚类。合并的依据可以是两个聚类之间的平均相似度、最小相似度等。此步骤将不断迭代, 直到达到指定的聚类数目或满足停止准则。

在本题中, 考虑到样本点集中数据较少且大多为连续型数, 而 Ward 距离聚类的方法通过最小化合并群集后的方差增加, 追求紧密且均衡的聚类结果, 而且它在处理连续变量数据时表现良好且适用于中小规模的数据集, 所以我们采用 Ward 方法来度量类与

类之间的相似性。

若记

$$D_1 = \sum_{x_i \in G_1} (x_i - x_1)^T (x_i - \bar{x}_1), D_2 = \sum_{x_i \in G_2} (x_i - x_2)^T (x_i - \bar{x}_2) D_{12} = \sum_{x_k \in G_1 \cup G_2} (x_k - \bar{x})^T (x_k - \bar{x})$$

则用两个类之间的离差平方和来定义两个类之间的相似性

$$D(G_1, G_2) = D_{12} - D_1 - D_2$$

（五）聚类划分，根据聚类合并的过程，形成一棵聚类树。根据树的结构，可以选择将聚类划分为不同的层次或簇。划分聚类的方法可以是设定阈值、树剪枝等。在本题中，我们记 $L = (L_1 \quad L_2 \quad L_3)$ 为四个类别的临界值。

7.3 问题 4 的模型求解

利用 Matlabsh 实现上述模型，得到如下的聚类示意图

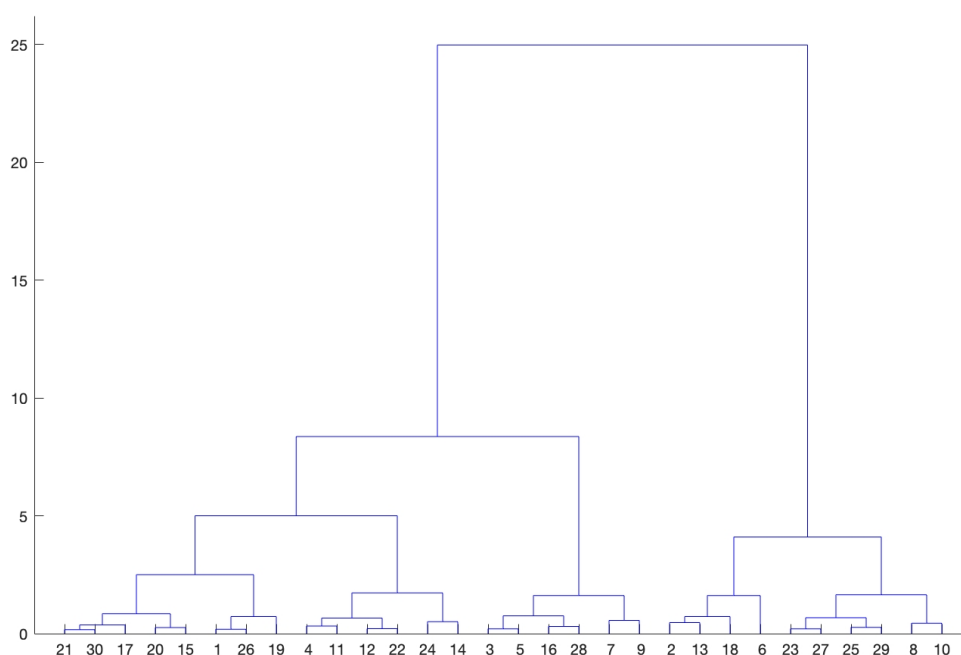


图 4 聚类示意图

根据 Q 型聚类分析模型得到 380 组数据中婴儿睡眠质量的综合评判结果如下：

睡眠质量为优	2	6	10	13	14	17	19	21	23	25	26	30	31	32	34	35	39	40	47	48	53	59	61	62	65
	70	73	74	78	79	80	83	85	86	87	90	93	95	98	99	101	111	115	118	120	124	125	127	136	138
	142	147	148	151	156	157	170	171	175	177	179	183	185	191	193	194	196	197	202	205	206	207	209	211	218
	220	228	230	235	237	239	245	260	277	284	287	289	297	298	302	305	306	307	308	309	314	316	319	322	323
	324	326	329	332	333	335	336	338	339	340	343	347	349	353	356	361	364	366	369	371	374	380			
睡眠质量为良	3	5	7	8	9	16	18	28	33	44	50	55	56	57	60	63	64	68	77	82	89	96	103	104	106
	109	110	113	114	132	134	135	139	140	141	145	152	154	158	160	164	172	176	186	188	190	195	198	201	203
	204	210	215	216	217	221	222	224	226	227	236	246	250	263	264	279	280	285	293	294	310	315	318	320	325
	327	341	345	346	348	357	359	360	368	372	373														
睡眠质量为中	1	11	12	22	24	27	29	36	37	41	43	45	51	69	71	75	76	88	92	105	107	108	112	116	117
	119	121	123	126	128	131	133	137	149	150	153	162	165	166	167	168	169	174	178	180	181	187	189	199	208
	212	214	219	232	233	240	241	242	243	247	248	249	251	252	253	255	256	257	258	259	265	268	271	272	275
	283	288	290	291	295	300	301	304	313	321	328	337	342	344	350	363	370	376	377						
睡眠质量为差	1	15	20	38	42	46	49	52	54	58	66	67	72	81	84	91	94	97	100	102	122	129	130	143	144
	146	155	159	161	163	173	182	184	192	200	213	223	225	229	231	234	238	244	254	261	262	266	267	269	270
	273	274	276	278	281	282	286	292	296	299	303	311	312	317	330	331	334	351	352	354	355	358	362	365	367
	375	378	379																						

图 5 综合睡眠质量评判结果

类似于问题 2 中建立婴儿的行为特征与母亲的身体健康指标与心理指标的关系模型，我们可以使用因子多重线性回归建立起婴儿综合睡眠质量与母亲的身体健康指标、心理指标的关联模型为

$$y_2 = w_2 F = w_2 \cdot \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix}$$

根据上述关联模型得到的最后 20 组（编号 391-410 号）婴儿的综合睡眠质量的预测结果如下

编号	评分	行为特征
391	-1.581326401	矛盾型
392	-1.266486035	矛盾型
393	-0.469179441	中等型
394	0.983773326	安静型
395	0.268656657	中等型
396	0.004306753	中等型
397	-0.41876254	中等型
398	0.195895298	中等型
399	-0.499161101	中等型
400	0.431751682	中等型
401	-2.693111692	矛盾型
402	0.008282044	中等型
403	0.43524039	中等型
404	0.893321414	安静型
405	0.754973033	安静型
406	0.966602091	安静型
407	-0.45435461	中等型
408	0.978450367	安静型
409	1.441213753	安静型
410	0.019915011	中等型

图 6 综合睡眠质量预测结果

八、问题 5 的模型建立与求解

8.1 对问题 5 的分析

问题五也是一个优化问题。本题需要在第四问得到回归系数和第三问中实现的改善婴儿行为特征的基础上对求得治疗心理障碍的最小优化。因此本题是在问题三的基础上的双目标线性规划问题。只需将问题三需要达成的目标作为另外一个约束条件，即可得到问题五的线性优化方程。

8.2 问题 5 的模型建立

(一) 给出调整方案

在问题 3 中已经得到单目标优化的可行解，我们需要做出一定的调整

$$S = S' + \Delta S' = \begin{pmatrix} S_1 + \Delta S_1 \\ S_2 + \Delta S_2 \\ S_3 + \Delta S_3 \end{pmatrix}$$

$$y_1' = WDS'$$

$$y_2' = WDS'$$

$$\Delta S' = \frac{\Delta y_2}{W \cdot D}$$

(二) 基于问题三构造的优化模型

$$\min \quad Cost = \left(\text{sigmoid}(1) \quad \text{sigmoid}(2) \quad \text{sigmoid}(3) \right) \cdot \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \text{linprog}(k', \Omega, \Delta S)$$

$$s.t. \quad \begin{cases} \Delta S = \begin{pmatrix} C_1 & C_2 & C_3 \end{pmatrix} \cdot \begin{pmatrix} \Delta S'_1 \\ \Delta S'_2 \\ \Delta S'_3 \end{pmatrix} \\ \|S'_i + \Delta S'_i\| \leq 1 \\ \Delta y_2 \geq L_1 - y_2' \\ y_{1''} = WD(S'_i + \Delta S'_i) \end{cases}$$

8.3 问题 5 的模型求解

在问题 4 的基础上我们可以计算得到第 238 号婴儿得综合睡眠质量指标与睡眠质量为优的指标相差 5.411，即 $\Delta y_2 = 5.411$ ，在此基础上我们通过上述优化模型得到使睡眠质量评级为优的最小代价为 $Cost = 24515$ 。

九、模型评价

9.1 模型的优点

(一) 本题的主要目标是评价母亲的身体因素和心理因素对以婴儿的行为特征为代表母婴关系和各类睡眠指标为代表的婴儿生长发育情况的影响。而个人的各项身体、心理指标往往相互依赖、相互影响。因此,本文采用因子分析回归方法进行评价,充分考虑了各项的内部机理,将8个指标抽象成3个根据代表性的抽象因子,较好地建立了原始指标和评价目标的对应关系。

(二) 本文在对新样本的综合睡眠质量进行分类时采用了Q型系统聚类分析,准确、细致地在原数据集的基础上建立了分类情况,得到了比较清晰的分类结果,并且能快速得到不同粒度上的多层次聚类结构。

(三) 本题的优化模型为线性模型,简单易用。本文建立的优化方程在多重线性规划确定权重的基础上,约束方程简洁,具有较高的求解效率和稳定性。在可行域内能快速找到全局最优解。

9.2 模型的缺点

(一) 综合评价的结果具有一定的相对性。尽管采用了因子分析和聚类分析的定量方法、结果也是用数值表示的,但是大部分结果只是具有相对意义,因此大多适用于在性质相同的对象之间进行比较和排序,而不适用于提出母婴关系的客观评价指标。

(二) 结果无法考虑个体差异。人的心理健康和生产过程都是非常复杂的过程,受众多因素影响,而统计结果提供的数据并不足以量化心理构建和产前产后各因素对婴儿的行为特征和睡眠情况的影响,从而减少个体差异对评价、分类结果的影响。

(三) 结果受主观评价因素影响。在综合评价中,评价指标的选择、指标权重的确定等常常需要依靠相关专家来确定。而目前相关文献在量化各个指标对母婴关系的影响上有一定分歧。因此,本文在对原始指标的赋权和综合指标的建立上具有较强的主观性。

9.3 模型的改进

(一) 可进一步考虑评价指标与评价目标间的相互影响。母亲的心理和身体状态会影响婴儿的行为特征和睡眠情况,同时母婴关系与婴儿的行为特征和睡眠情况会对母亲的心理和身体状态有反作用。查阅文献可知,产妇产后的心理障碍发展一般有三种发展模式。而统计结果中包含不同年龄的婴儿与当时的母婴状态。因此可以进一步探究母婴关系与婴儿的行为特征和睡眠情况会对母亲的心理和身体状态的反作用,从而建立时间序列模型,更准确地反映评价指标与评价目标之间的关系。

(二) 可进一步查阅相关文献,降低模型的主观性。提供本文研究使用的数据集指标较少,难以建立较客观的关系函数。因此需要进一步查阅相关文献,了解各要素间内

部机理，采用专家建议，使结论更稳定、更具有客观性。

参考文献

- [1] Wang, N. R., Ye, Y.(2016), Zhongguo dang dai er ke za zhi = Chinese journal of contemporary pediatrics, 18(4).
- [2] Johnson, K. J. et al. Parental age and risk of childhood cancer: a pooled analysis. Epidemiology20, 475-483 (2009).
- [3] 黎秋勤. 基于 EPDS 分析女性产后抑郁水平和影响因素分析 [J]. 国际检验医学杂志,2018,39(A02):249-251
- [4] 姚青芳, 夏澜, 王先伟, 江湖, 候英英, 蒲亨萍. 产后创伤后应激障碍的现状调查及影响因素分析 [J]. 护士进修杂志,2023,38(1):88-93

附录 A 相关代码

数据导入

```
% import.m
clc;clear;close all;
%% 导入标记数据
opts = spreadsheetImportOptions("NumVariables", 15);
opts.Sheet = "Sheet1";
opts.DataRange = "A2:O391";
opts.VariableNames = ["ID", "motherAge", "Marriage", "Education", "Time", "Method",
    "CBTS", "EPDS", "HADS", "Behavior", "Sex", "babyAge", "Sleeping", "Awake", "Asleep"];
opts.VariableTypes = ["double", "double", "double", "double", "double", "double",
    "double", "double", "double", "char", "double", "double", "datetime", "double",
    "double"];
cateData = readtable("2023暑期培训第一轮模拟题A题附件.xlsx", opts, "UseExcel", false);
clear opts
marriage=cateData.Marriage;
ind=find(marriage<=2);
cateData=cateData(ind,:);
clear ind marriage
tmpBehavior=cateData.Behavior;
len=length(tmpBehavior);
tmpValue=zeros(len,1);
for i=1:len
    if (tmpBehavior{i}(1)=='安')
        tmpValue(i)=1;
    end
    if (tmpBehavior{i}(1)=='中')
        tmpValue(i)=0;
    end
    if (tmpBehavior{i}(1)=='矛')
        tmpValue(i)=-1;
    end
end
clear i
cateData.Behavior=tmpValue;
clear tmpValue tmpBehavior
tmpSleeping=cateData.Sleeping;
tmpChar=char(tmpSleeping);
tmpChar=tmpChar(:,13:16);
tmpValue=zeros(len,1);
for i=1:len
    tmpValue(i)=str2double(tmpChar(i,1:2));
    if (tmpChar(i,4)=='3')
        tmpValue(i)=tmpValue(i)+0.5;
    end
end
```

```

end
end
clear i
cateData.Sleeping=tmpValue;
sleep=cateData.Sleeping;
ind=find(sleep<20);
len=length(ind);
cateData=cateData(ind,:);
clear ind sleep
clear tmpSleeping tmpChar tmpValue

%% 导入未标记数据
opts = spreadsheetImportOptions("NumVariables", 12);
opts.Sheet = "Sheet1";
opts.DataRange = "A392:L411";
opts.VariableNames = ["ID", "motherAge", "Marriage", "Education", "Time", "Method",
    "CBTS", "EPDS", "HADS", "Behavior", "Sex", "babyAge"];
opts.SelectedVariableNames=["ID", "motherAge", "Marriage", "Education", "Time", "Method",
    "CBTS", "EPDS", "HADS", "Sex", "babyAge"];
opts.VariableTypes = ["double", "double", "double", "double", "double", "double",
    "double", "double", "double", "categorical", "double","double",,];
preData = readtable("2023暑期培训第一轮模拟题A题附件.xlsx", opts, "UseExcel", false);

%% 收尾
clear opts
save rawData.mat

```

```

%Prewrite.m
clc;clear;close all;
load rawData.mat
%% 预处理
function score=CalcPregTime(time)
if ((time>=37) && (time<43))
score=1;
elseif (time<37)%26~37
score=0.4+0.6/(1 + exp(-0.7*(time-31)));
else%43-45
score=1-1/(1 + exp(-2*(time-45)));
end
end
function score=CalcMatherAge(age)
if ((age>=22) && (age<29))%最佳生育年龄为22-29,在这个年龄段不区分年龄对生育的影响
score=1;
elseif (age<22)%18-22岁
score=0.2+0.8/(1 + exp(-0.5*(age-20)));
else%30-50岁

```

```

score=1-0.8/(1 + exp(-0.2*(age-40)));
end
end
scoreAge=zeros(len,1);
scoreTime=zeros(len,1);
for i=1:len
scoreAge(i)=CalcMatherAge(cateData.motherAge(i));
scoreTime(i)=CalcPregTime(cateData.Time(i));
if (cateData.Method(i)==1)
cateData.Method(i)=1;
else
cateData.Method(i)=-1;
end
if (cateData.Sex(i)==1)
cateData.Sex(i)=-1;
else
cateData.Sex(i)=1;
end
if (cateData.Marriage(i)==1)
cateData.Marriage(i)=-1;
else
cateData.Marriage(i)=1;
end
cateData.CBTS=30-cateData.CBTS;
cateData.EPDS=30-cateData.EPDS;
cateData.HADS=30-cateData.HADS;
end
clear i
cateData.motherAge=scoreAge;
cateData.Time=scoreTime;

len2=length(preData.ID);
scoreAge=zeros(len2,1);
scoreTime=zeros(len2,1);
for i=1:len2
scoreAge(i)=CalcMatherAge(preData.motherAge(i));
scoreTime(i)=CalcPregTime(preData.Time(i));
if (preData.Method(i)==1)
preData.Method(i)=1;
else
preData.Method(i)=-1;
end
if (preData.Sex(i)==1)
preData.Sex(i)=-1;
else
preData.Sex(i)=1;
end
end

```

```

if (preData.Marriage(i)==1)
preData.Marriage(i)=-1;
else
preData.Marriage(i)=1;
end
end
clear i
preData.motherAge=scoreAge;
preData.Time=scoreTime;
clear scoreTime scoreAge len2
%% 标准化
cateData.motherAge=zscore(cateData.motherAge);
cateData.Time=zscore(cateData.Time);
cateData.Education=zscore(cateData.Education);
cateData.CBTS=zscore(cateData.CBTS);
cateData.EPDS=zscore(cateData.CBTS);
cateData.HADS=zscore(cateData.HADS);
cateData.Sleeping=zscore(cateData.Sleeping);
writetable(cateData,'scoreData.xlsx');
save scoreData.mat

```

```

% sleep.m 计算综合睡眠质量
clc;clear;close all;
load rawData.mat
quality=cell(5,2);
meanSex=zeros(5,1);
singleSleep=zeros(len,1);
effSleep=zeros(len,1);
scoreSleep=zeros(len,1);
cost_awake=[0.7,0.6167,0.4];
con_asleep=[0,1,1,1.5,2];
for i=1:5
ind=find(cateData.Asleep==i);
quality{i,1}=cateData(ind,:).Sleeping;
quality{i,2}=cateData(ind,:).Awake;
meanSex(i)=mean(cateData(ind,:).Sex);
end
clear i ind
for i=1:len
singleSleep(i)=cateData.Sleeping(i)/(cateData.Awake(i)+1);
effSleep(i)=1-cateData.Awake(i)*cost_awake(cateData.babyAge(i))/cateData.Sleeping(i);
scoreSleep(i)=singleSleep(i)*effSleep(i)+con_asleep(cateData.Asleep(i));
end
clear i
save sleepData.mat quality singleSleep effSleep scoreSleep

```

1.1 问题求解

```
%CoefExam.m 计算第一问中各指标相关系数
clc;clear;close all;
load scoreData.mat
load sleepData.mat
physical=[cateData.motherAge,cateData.Marriage,cateData.Education,cateData.Time,cateData.Method];
psychosocial=[cateData.CBTS,cateData.EPDS,cateData.HADS];
basic=[physical,psychosocial];
goal=[cateData.Behavior,cateData.Sleeping,cateData.Awake,cateData.Asleep];
sleepGoal=[scoreSleep,cateData.Sleeping,cateData.Awake,cateData.Asleep];
[R,P,RL,RU] = corrcoef([basic,goal], 'Alpha',0.05);
save arrayData.mat
```

```
%m_Factor.m 构造因子分析
function [coef,weight,rate]=m_Factor(X,num)
%因子分析Factor Analysis,X为自变量矩阵,Y为因变量矩阵,num为主因子个数
Correspond=cov(X);
[rate,latent,con]=pcacov(Correspond);
%latent为Correspond的特征值, rate为各个主成分的贡献率
f1=repmat(sign(sum(rate)),size(rate,1),1);
%构造与vec1同维数的元素为±1的矩阵用于求正分量
rate=rate.*f1;
%修改特征向量的正负号, 使得每个特征向量的分量和为正, 即为最终的特征向量
f2=repmat(sqrt(latent)',size(rate,1),1);
factormatOrigin=rate.*f2; %构造全部因子的载荷矩阵
factormatMain=factormatOrigin(:,1:num);
[factormat_rotate,t]=rotatefactors(factormatMain,'Method','varimax');
%b为旋转后的矩阵,t为做变换的正交矩阵
factmat=[factormat_rotate,factormatOrigin(:,num+1:end)];
contribution=sum(factmat.^2); %计算各个因子的贡献
rate=contribution(1:num)/sum(contribution); %计算因子的贡献率
coef=Correspond\factormat_rotate; %计算得分函数的系数
weight=rate/sum(rate); %计算得分的权重
end
```

```
%m_FactorCheck.m 检验因子构造结果
function [residual,common,contribution]=m_FactorCheck(matrix,ind)
%因子分析Factor Analysis,Matrix为数据矩阵
%contribution为累计贡献率系数向量,residual为因子的残差矩阵
matrix=zscore(matrix); %数据的标准化
Correspond=corrcoef(matrix); %计算相关系数矩阵Correspond
%下面利用相关系数矩阵进行主成分分析, vec1的第一列为Correspond的第一特征向量, 即主成分的系数
[rate,latent,con]=pcacov(Correspond);
%latent为Correspond的特征值, rate为各个主成分的贡献率
```



```

f1= repmat(sign(sum(rate)),size(rate,1),1);
    %构造与vec1同维数的元素为±1的矩阵用于求正分量
rate=rate.*f1;          %修改特征向量的正负号，使得每个特征向量的分量和为正，即为最终的特征向量
f2= repmat(sqrt(latent)',size(rate,1),1);
factormat=rate.*f2;      %构造全部因子的载荷矩阵
sub=factormat(:,ind);     %计算两个因子时的载荷矩阵
tsquare=diag(Correspond-sub*sub'); %计算两个因子的特殊方差
residual=Correspond-sub*sub'-diag(tsquare); %计算两个因子时的残差矩阵
common=sum(sub.^2,2);
contribution=cumsum(con);
end

%FactorRegress.m 利用构造的因子分析模型进行回归
%% 建立回归方程
clc;clear;close all;
load arrayData.mat
[coef,weight,rate]=m_Factor(basic,3);
[residual,common]=m_FactorCheck(basic,1:3);
score=basic*coef;
Tscore=score*weight';
[ccoef,p]=corrcoef([Tscore,goal(:,1),scoreSleep]);
d=zeros(5,4);stats=zeros(5,4);
for i=1:4
[d(i,:),~,~,~,stats(i,:)]=regress(goal(:,i),[ones(len,1),score]);
end
[d(5,:),~,~,~,stats(5,:)]=regress(scoreSleep,[ones(len,1),score]);
clear i
%d=abs(d);
writematrix(d,'RegressRatio.xlsx');
writematrix([cateData.ID,score],'score.xlsx');
save regressData.mat coef d score
varResidual=var(residual)';
t1=table(residual,varResidual,common);
rate=rate';weight=weight';
t2=table(rate,weight,ccoef);
writetable(t1,'FactorRegress.xlsx','Sheet',1);
writetable(t2,'FactorRegress.xlsx','Sheet',2);

```

```

%Problem2.m 利用构造的因子分析模型进行分类得出第二问结论
%% 建立回归方程
clc;clear;close all;
load arrayData.mat
[coef,weight,rate]=m_Factor(basic,3);
[residual,common]=m_FactorCheck(basic,1:3);
score=basic*coef;

```

```

Tscore=score*weight';
[d,~,~,~,stats]=regress(goal(:,1),[ones(len,1),Tscore]);
dd=regress(goal(:,1),score);dd=normalize(dd);
save BehaviorRegress.mat d score Tscore weight dd

%% 利用回归方程预测
load scoreData.mat
len2=length(preData.ID);
physical=[preData.motherAge,preData.Marriage,preData.Education,preData.Time,preData.Method];
psychosocial=[preData.CBTS,preData.EPDS,preData.HADS];
basic=[physical,psychosocial];
Bscore=basic*coef;
BTscore=Bscore*weight';
scoreFinal=zscore([ones(len2,1),BTscore]*d);
scoreBehavior=round(scoreFinal);

for i=1:len2
if (scoreBehavior(i)<0)
cate(i)="矛盾型";
elseif (scoreBehavior(i)>0)
cate(i)="安静型";
else
cate(i)="中等型";
end
end
clear i
cate=cate';
t=table(scoreFinal,cate);
writetable(t,"problem2.xlsx");

```

```

%m_Clustering.m 利用k-means算法进行聚类
function Res=m_Clustering(matrix,num)
%系统聚类分析,Matrix为目标分析矩阵,Method为距离计算方法,Num为类的数量
%返回结果为每一类的对象归类的矩阵
matrix=zscore(matrix); %数据标准化
Distance=pdist(matrix); %求对象间的欧氏距离
z=linkage(Distance,"ward"); %按指定方法生成聚类
dendrogram(z);
Res=cell(num,1);
T=cluster(z,'maxclust',num);
for i=1:num
Target=find(T==i);
Target=reshape(Target,1,length(Target));
Res{i}=Target;
end
end

```

```

%Problem4.m 利用构造的因子分析模型和聚类结果进行分类得出第四问结论
clc;clear;close all;
load arrayData.mat
load sleepData.mat
%% 建立分类标准
Res=m_Clustering(scoreSleep,4);
aver=zeros(1,4);barr=zeros(1,4);
for i=1:4
aver(i)=mean(scoreSleep(Res{i}));
barr(i)=min(scoreSleep(Res{i}));
end
[aver,ind]=sort(aver,'descend');
Res=Res(ind);barr=barr(ind);
clear i ind
%% 建立回归方程
[coef,weight,rate]=m_Factor(basic,3);
[residual,common]=m_FactorCheck(basic,1:3);
score=basic*coef;
Tscore=score*weight';
[d,~,~,~,stats]=regress(scoreSleep,[ones(len,1),Tscore]);

%% 利用回归和聚类结果分类
load scoreData.mat
len2=length(preData.ID);
physical=[preData.motherAge,preData.Marriage,preData.Education,preData.Time,preData.Method];
psychosocial=[preData.CBTS,preData.EPDS,preData.HADS];
basic=[physical,psychosocial];
Bscore=basic*coef;
BTscore=Bscore*weight';
scoreFinal=d(1)+BTscore*d(2);
scoreFinal=1.5*(scoreFinal-min(scoreFinal)+0.5);
cate=zeros(len2,1);
name=["优","良","中","差"];
for i=1:len2
for j=1:4
if (scoreFinal(i)>barr(j))
cate(i)=j;
cateStr(i)=name(j);
break
end
end
end
clear i j
cateStr=cateStr';
res=table(cate,cateStr);
writetable(res,"problem4.xlsx");

```

```

writecell(Res,"SleepCate.xlsx");
save sleepCateData.mat aver barr

```

```

function [cost,x]=CalcProgram(select,init,Aeq,dScore)
k=zeros(1,3);
k(1)=(2812-200)/(3-0);
k(2)=(1890-500)/(2-0);
k(3)=(12500-300)/(5-0);
k=k(select);init=init(select);Aeq=Aeq(select);
A=eye(length(select));B=ones(1,length(select));
k=k./init;
x=linprog(k,A,B,Aeq,dScore)';
cost=0;
if (~isempty(x))
cost=sum(-k.*x);
else
cost=inf;
end
end
%CureCost.m 建立线性优化方法
function cost=CureCost(ind)
behavior=cateData.Behavior(ind);
cbts=cateData.CBTS(ind);
epds=cateData.EPDS(ind);
hads=cateData.HADS(ind);
s=score(ind,:);init=[cbts,epds,hads];
present=sum(weight.*s);
dScore=-0.1-present;
dPhy=dScore/weight(1);
cost=zeros(1,8);
for i=1:8
select=[];
if ((i/4)==1)
select=[select,1];
end
if (mod((i/2),2)==1)
select=[select,2];
end
if (mod(i,2)==1)
select=[select,3];
end
cost(i)=CalcProgram(select,init);
end
clear i j k
end

```

```

function cost=CalcProgram(select,init)
k=zeros(1,3);
k(1)=(2812-200)/(3-0);
k(2)=(1890-500)/(2-0);
k(3)=(12500-300)/(5-0);
b=zeros(1,3);
b(1)=200;b(2)=500;b(3)=300;
b=b(select);k=k(select);init=init(select);
k=k./init;
cost=sum(b);
end

```

```

%Problem3.m 利用线性规划得出优化结论
clc;clear;close all;
load behaviorRegress.mat
load scoreData.mat
load regressData.mat
coef=coef(6:8,:)';weight=weight';
b=zeros(1,3);
b(1)=200;b(2)=500;b(3)=300;
Aeq=coef*weight;Aeq=Aeq';
ind=find(cateData.ID==238);
cbts=cateData.CBTS(ind);
epds=cateData.EPDS(ind);
hads=cateData.HADS(ind);
s=score(ind,:)';init=[cbts,epds,hads];
present=sum(weight.*s);
dScore=-0.1-present;
costAny=zeros(1,7);
for i=1:7
select=[];
if (floor(i/4)==1)
select=[select,1];
end
if (floor(mod((i/2),2))==1)
select=[select,2];
end
if (floor(mod(i,2))==1)
select=[select,3];
end
[tmpCost,tmpX]=CalcProgram(select,init,Aeq,dScore);
costAny(i)=sum(b(select))+tmpCost;
if (i==1 || cost>costAny(i))
cost=costAny(i);
sol=[tmpX;select];
end

```

```

clear tmpCost tmpX
end
clear i j k
save Cure.mat costAny sol

%Problem5.m 利用线性规划和第三题结论得出第五问优化结论
clc;clear;close all;
load behaviorRegress.mat
load scoreData.mat
load regressData.mat
load sleepData.mat
load sleepCateData.mat
load cure.mat
coef=coef(6:8,:);weight=weight';
b=zeros(1,3);
b(1)=200;b(2)=500;b(3)=300;
Aeq=coef*weight;Aeq=Aeq';
ind=find(cateData.ID==238);
cbts=cateData.CBTS(ind);
epds=cateData.EPDS(ind);
hads=cateData.HADS(ind);
init=[cbts,epds,hads];
select=sol(2,:);x=sol(1,:);
init(select)=init(select)+x;
init=zscore(init);
s=score(ind,:);
present=sum(weight.*s);
dScore=-0.1-present;
dSleep=barr(1)-scoreSleep(ind);
costAny=zeros(1,7);
for i=1:7
select=[];
if (floor(i/4)==1)
select=[select,1];
end
if (floor(mod((i/2),2))==1)
select=[select,2];
end
if (floor(mod(i,2))==1)
select=[select,3];
end
costAny(i)=sum(b(select))+CalcProgram(select,init,Aeq,dSleep);
end
clear i j k
cost=min(costAny);

```

1.2 结果呈现

```
clc;clear;close all;
load arrayData.mat
labels={'母亲年龄','婚姻状况','教育水平','妊娠时间','生产方式','CBTS','EPDS','HADS','行为特征','睡眠时间','睡眠  
h = heatmap(labels,labels, R, 'FontSize',8, 'FontName','Helvetica');
h.FontSize=14;
h.Title='各原始数据与婴儿行为 and 睡眠的相关性系数';
h.CellLabelColor = 'auto';
set(gcf,'Color',[1 1 1]);
```

```
%% 数据导入与处理
clc;clear;close all;
load sleepData.mat
len1=0;len2=0;
for i=1:5
len1=max(len1,length(quality{i,1}));
len2=max(len2,length(quality{i,2}));
end
clear i
sleep=nan(len1,5);
awake=nan(len2,5);
meanSleep=zeros(5,1);
meanAwake=zeros(5,1);
for i=1:5
sleep(1:length(quality{i,1}),i)=quality{i,1};
meanSleep(i)=nanmean(quality{i,1});
awake(1:length(quality{i,2}),i)=quality{i,2};
meanAwake(i)=nanmean(quality{i,2});
end
clear i len1 len2

C1 = [0.8929,0.2929,0.2000];
C2 = [0.3772,0.7297,0.3806];
p1=[0.5 3.5 6.5 9.5 12.5];
p2=[1.4,4.4,7.4,10.4 13.4];
hold on
%% 整晚睡眠时长绘制
yyaxis left
box_1 = boxplot(sleep,'Positions',p1,'Colors','b','Widths',0.7,'Symbol','o','Notch','on');
plot(p1,meanSleep,'-o');
hYLabel1 = ylabel('整晚睡眠时间');
set(gca, 'YColor', [.1 .1 .1], 'Ylim' ,[4 13]);

%% 睡醒次数绘制
yyaxis right
```

```

box_2 = boxplot(awake,'positions',p2,'Colors','k','Widths',0.7,'Symbol','o','Notch','on');
plot(p2,meanAwake,'-o');
hYLabel2 = ylabel('睡醒次数');
set(gca, 'YColor', [.1 .1 .1], 'Ylim', [-1 11]);

%% 细节优化
%标签
hTitle = title('各类入睡方式对婴孩睡眠情况的影响','FontSize',20);
hXLabel = xlabel('入睡方式','FontSize',14);
set(gca, 'XTick', (p1+p2)/2,... % 边框
'XGrid', 'off', 'YGrid', 'on', ... % 网格
'TickDir', 'out', 'TickLength', [.01 .01], ... % 刻度
'XMinorTick', 'off', 'YMinorTick', 'off', ... % 小刻度
'XColor', [.1 .1 .1],... % 坐标轴颜色
'Xticklabel',{'哄睡法','抚摸法','安抚奶嘴法','环境营造法','定时法'}) % X轴标签
hLegend = legend([box_1(1),box_2(2)], ...
'整晚睡眠时间', '睡醒次数', ...
'Location', 'northeast','orientation','horizontal');
% 添加上、右框线
xc = get(gca,'XColor');
yc = get(gca,'YColor');
unit = get(gca,'units');
ax = axes('Units', unit,...
'Position',get(gca,'Position'),...
'XAxisLocation','top',...
'YAxisLocation','right',...
'Color','none',...
'XColor',xc,...
'YColor',yc);
set(ax, 'linewidth',1,...
'XTick', [],...
'YTick', []);

```

```

%% 数据导入与处理
clc;clear;close all;
load rawData.mat
ind1=find(cateData.Marriage==2);
ind2=find(cateData.Marriage==1);
marrT=cateData(ind1,:);
unmarrT=cateData(ind2,:);
marr=[marrT.CBTS,marrT.EPDS,marrT.HADS];
unmarr=[unmarrT.CBTS,unmarrT.EPDS,unmarrT.HADS];
meanMarr=mean(marr,1);
meanUnmarr=mean(unmarr,1);

C1 = [0.8929,0.2929,0.2000];

```



```

C2 = [0.3772,0.7297,0.3806];
p1=[0.5 3.5 6.5];
%p2=[1.4,4.4,7.4];
p2=p1;
hold on
%% 绘制
%box_1 = boxplot(marr,'Positions',p1,'Colors','b','Widths',0.7,'Symbol','o');
plot1=plot(p1,meanMarr,'-o');
%box_2 = boxplot(unmarr,'positions',p2,'Colors','k','Widths',0.7,'Symbol','o');
plot(p2,meanUnmarr,'-o');
set(gca, 'YColor', [.1 .1 .1], 'Ylim' ,[0 30]);

%% 细节优化
%标签
hTitle = title('不同婚姻情况的产妇心理量表结果','FontSize',20);
hXLabel = xlabel('量表名称','FontSize',14);
hYLabel1 = ylabel('分数');
set(gca, 'XTick', (p1+p2)/2,... % 边框
'XGrid', 'off', 'YGrid', 'on', ... % 网格
'TickDir', 'out', 'TickLength', [.01 .01], ... % 刻度
'XMinorTick', 'off', 'YMinorTick', 'off', ... % 小刻度
'XColor', [.1 .1 .1],... % 坐标轴颜色
'Xticklabel',{'CBTS','EPDS','HADS'}) % X轴标签
hLegend = legend([box_1(1),box_2(2)], ...
'已婚', '未婚', ...
'Location', 'northeast','orientation','horizontal');

```