

所属类别	2024 年“华数杯”全国大学生数学建模竞赛	参赛编号
本科组		CM2400947

全国大学生数学建模竞赛论文模板

摘要

母亲的身心健康对婴儿的早期成长和发展具有重要影响。已有研究表明，母亲的身体状况以及心理健康（如压力、抑郁、焦虑等）不仅关系到婴儿的生理健康，还会影响婴儿的认知、情感、社会行为等多方面成长。婴儿的行为特征和睡眠质量常常作为衡量婴儿发展状况的重要指标。基于这方面的专业背景，本文对相关影响机制进行了研究与建模。

对于问题一，本文通过分析 390 名婴儿及其母亲的身体和心理相关指标数据，研究了母亲因素对婴儿行为特征和睡眠质量的影响关系。

对于问题二，基于婴儿行为问卷，将婴儿行为特征划分为安静型、中等型和矛盾型，建立了行为特征与母亲身体、心理指标之间的数学关系模型，并利用该模型预测了 20 组缺失婴儿行为类型。

对于问题三，建立了 CBTS、EPDS、HADS 评分与治疗费用的关联模型，针对编号 238 的矛盾型婴儿，分析通过最小治疗费用使其行为特征由矛盾型转变为中等型和安静型的策略及调整路径。

对于问题四，选取婴儿睡眠的多项指标，综合评判睡眠质量，并建立母亲各项指标与婴儿综合睡眠质量的关联模型，进而预测缺失数据婴儿的睡眠类别。

最后，结合模型结果，对如何提升 238 号婴儿的睡眠质量等级为优提出了具体的治疗方案和优化建议。

关键字： 母亲心理健康 婴儿成长 行为特征 睡眠质量 数学建模

一、问题重述

1.1 问题背景

问题背景

1.2 问题要求

问题 1

问题 2

问题 3

问题 4

二、问题分析

2.1 问题一分析

对于问题一，

2.2 问题二分析

对于问题二，

2.3 问题三分析

对于问题三，

2.4 问题四分析

对于问题四，

三、模型假设

为简化问题，本文做出以下假设：

- 假设 1
- 假设 2
- 假设 3

四、符号说明

符号	说明	单位
m	质量	kg
V	体积	m^3

五、问题一的模型的建立和求解

5.1 数据预处理与特征工程

高质量的数据是构建有效模型的基础。针对原始数据集，我们进行了以下关键的数据预处理和特征工程步骤，旨在确保数据的规范性、一致性和可用性，为后续的模型构建奠定坚实基础：

5.1.1 睡眠时间格式转换

原始数据中婴儿“整晚睡眠时间”以“HH:MM”字符串格式记录，且存在“99:99”等异常值。我们将其统一转换为浮点数表示的小时数，并对异常值进行了妥善处理，通常将其视为缺失值或根据业务逻辑进行填充。

$$\text{时间}_{\text{小时}} = \text{小时} + \frac{\text{分钟}}{60}$$

此转换确保了睡眠时间数据的数值化和可计算性，使其能够被回归模型有效利用。

5.1.2 分类变量编码与映射

数据集中包含婚姻状况、教育程度、分娩方式、婴儿性别、入睡方式和婴儿行为特征等分类变量。为使其能够被回归模型处理，我们进行了以下操作：

- **数值映射为标签：**将原始的数值编码（如 1、2、3 等）映射为更具可读性的类别标签（如“未婚”、“已婚”；“哄睡法”、“自主入睡”）。
- **独热编码 (One-Hot Encoding)：**对于无序的分类变量，如入睡方式、教育程度等，我们采用了独热编码将其转换为二元（0/1）指示变量。这有效避免了将分类变量误解释为有序数值，并消除了因任意数值赋值而引入的潜在偏差。例如，教育程度的“研究生”类别将被转换为一个独立的二元特征。

值得一提的是，婴儿行为特征作为因变量，被映射为数值（如安静型 =0，中等型 =1，矛盾型 =2），以适应多项逻辑回归模型的输入要求。

5.1.3 数值型特征标准化

对于母亲年龄、妊娠时间、CBTS、EPDS、HADS 等数值型预测变量，我们均采用了标准差标准化（Z-score normalization）处理。

$$X' = \frac{X - \mu}{\sigma}$$

其中， X 是原始特征值， μ 是特征的均值， σ 是特征的标准差。标准化后的特征均值为 0，标准差为 1。此步骤至关重要，它有助于：

- 消除不同特征因量纲和取值范围差异带来的影响，确保模型优化过程中所有特征得到同等重视。
- 加速模型收敛，尤其对于基于梯度下降的优化算法（如逻辑回归）。
- 提高模型的可解释性和稳定性，避免某些特征因数值过大而主导模型。

5.2 探索性数据分析 (EDA) 结果

在正式构建模型之前，我们进行了全面的探索性数据分析，旨在揭示数据内部的结构、变量分布及其潜在关联，为后续模型选择与构建提供数据驱动的依据。

5.2.1 变量分布特征

我们首先对关键变量的分布特性进行了细致的探究，以全面了解数据的内在结构。如图??所示，母亲的 CBTS、EPDS 和 HADS 得分均呈现右偏分布，这反映出在研究样本中，绝大多数母亲的心理健康状况处于良好范畴，然而，确实存在一部分母亲面临不同程度的心理困扰，这提示了潜在的干预需求。在婴儿睡眠质量方面，如图??所示，婴儿的整晚睡眠时间主要集中在 10 至 12 小时之间，这与该年龄段婴儿的正常生理需求基本吻合；同时，婴儿夜间睡醒的次数大多介于 0 至 2 次，但也观察到少数婴儿存在较为频繁的夜间醒来现象，这可能暗示了其睡眠模式存在某些问题。至于婴儿的行为特征，从图??的分布来看，样本中“中等型”婴儿占据了显著多数（57.7%），其次是“安静型”婴儿（30.8%），而“矛盾型”婴儿的比例最低（11.5%），这种明显的类别不平衡性在后续建模中需要予以特别关注，以避免模型偏向多数类别。

5.2.2 相关性分析

为了洞察各数值型变量之间的相互关系，我们计算了皮尔逊相关系数，并以热力图形式直观展示了其关联强度（如图??所示）。分析结果揭示了几项重要联系：首先，母亲的各项心理指标之间表现出显著的强相关性。例如，EPDS 与 HADS 的相关系数高达 0.79，CBTS 与 EPDS 为 0.78，CBTS 与 HADS 亦达到 0.69。这强烈暗示了抑郁、焦虑以及分娩后创伤应激障碍并非孤立存在，而是常常伴随出现，这一发现对制定综合性临床

干预策略具有重要的指导意义。其次，我们观察到母亲心理指标与婴儿睡眠质量之间存在负相关关系。具体而言，母亲的 CBTS、EPDS 和 HADS 得分与婴儿整晚睡眠时间呈弱到中等程度的负相关，而与睡醒次数呈正相关。这初步揭示了母亲的心理健康状况与婴儿睡眠质量之间存在紧密且负向的关联，即母亲的心理压力越大，婴儿的睡眠质量可能越受影响。

5.2.3 分类变量对连续变量的影响 (ANOVA)

为了评估分类变量对婴儿睡眠质量指标（睡醒次数和整晚睡眠时间）的潜在影响，我们运用了方差分析（ANOVA），以检验不同类别之间是否存在显著差异。分析结果分别呈现于图??和图??，明确指出“入睡方式”是影响婴儿睡眠质量的关键因素。具体来看：

- 对于婴儿的睡醒次数， F 值为 16.87， p 值远小于 0.001，表明入睡方式对睡醒次数具有极显著的影响。其中，采用自主入睡方式的婴儿，其夜间睡醒次数显著少于其他入睡方式的婴儿，这凸显了良好入睡习惯的积极作用。
- 同样，在整晚睡眠时间方面，入睡方式的影响也极为显著（ $F = 12.18, p < 0.001$ ）。自主入睡的婴儿普遍拥有更长的整晚睡眠时间，进一步印证了自主入睡对提升婴儿睡眠质量的积极作用。

值得注意的是，尽管我们考察了婚姻状况、教育程度、分娩方式和婴儿性别等其他分类变量，但它们对婴儿睡眠质量指标的影响均未达到统计显著水平（所有 $p > 0.05$ ）。这提示这些因素对婴儿睡眠的直接影响可能相对较弱，或其影响被其他更强的因素所掩盖。

5.2.4 分类变量间的关联性 (Chi-square)

我们进一步采用卡方检验来探究各个分类变量之间的关联性，以识别组别间是否存在非随机的分布模式。图??特别展示了那些具有统计显著性（ $p < 0.05$ ）的分类变量对，而图??则以热力图的形式呈现了所有分类变量对的 p 值， p 值越小意味着关联性越强。分析发现，两个主要关联：

- 婚姻状况与教育程度之间存在显著关联（ $p = 0.020$ ）。这表明在我们的样本中，已婚母亲的教育程度普遍高于未婚母亲，这可能反映了社会人口学特征的一般趋势。
- 更为重要的是，入睡方式与婴儿行为特征之间也存在显著关联（ $p = 0.014$ ）。具体而言，自主入睡的婴儿更倾向于表现出“安静型”的行为特征，这为育儿实践中推崇自主入睡提供了额外的支持证据，表明其不仅有利于睡眠，也可能塑造婴儿的整体行为模式。

除此之外，大多数分类变量之间未检测到显著关联，暗示这些因素在很大程度上是相互独立的，这简化了后续模型中的变量选择。

5.2.5 多重共线性检查

为了确保后续回归模型的稳健性与解释力，我们对所有数值型预测变量进行了多重共线性检查，并计算了方差膨胀因子（VIF）。VIF 值能够直观反映一个自变量能否被其他自变量线性解释的程度，从而判断是否存在过度相关性。如图??所示，虽然“教育程度_研究生”的 VIF 值最高（5.72）、“教育程度_高中”（4.56）和“教育程度_大学”（4.36）的 VIF 值也相对较高，但它们均未超过通常的警戒线（VIF=10）。这表明尽管教育程度的不同类别之间存在一定的共线性，但其程度仍在可接受范围内，预计不会对模型系数的估计造成严重偏差。此外，母亲的 EPDS 得分 VIF 值为 3.94，也反映了其与 CBTS、HADS 等其他心理指标之间存在一定相关性，这与我们之前的相关性分析结果一致。总体而言，绝大多数变量的 VIF 值均小于 2.5，这有力地表明本研究中的多重共线性问题得到了有效控制，不会严重影响模型的稳定性或结果的可靠性。

5.3 模型选择与理论基础

根据问题一所涉因变量的类型（连续型与多分类型），我们选择了相应的统计回归模型进行分析，以期最大化地捕捉变量间的潜在关系。

5.3.1 婴儿睡眠质量模型：多元线性回归

鉴于婴儿的“整晚睡眠时间”和“睡醒次数”均为连续数值型变量，我们选择多元线性回归 (Multiple Linear Regression) 作为其预测模型。多元线性回归旨在建立一个线性关系，用一个或多个自变量来预测因变量的取值，其优点在于模型简洁且易于解释。其一般形式为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

其中：

- Y 是因变量（如整晚睡眠时间或睡醒次数）。
- X_1, X_2, \dots, X_k 是 k 个自变量（如母亲年龄、心理指标、入睡方式等）。
- β_0 是截距项。
- $\beta_1, \beta_2, \dots, \beta_k$ 是对应自变量的回归系数，表示在其他自变量不变的情况下，该自变量每增加一个单位，因变量的平均变化量。
- ϵ 是误差项，代表模型未能解释的随机变异。

模型通过最小化残差平方和（Ordinary Least Squares, OLS）来估计回归系数。模型的评估指标包括决定系数 (R^2)、调整决定系数 (Adjusted R^2)、F 统计量及其 p 值，用于评估模型的整体拟合优度和统计显著性。

5.3.2 婴儿行为特征模型：多项逻辑回归

婴儿行为特征（“安静型”、“中等型”、“矛盾型”）是一个具有三个或更多无序类别的分类变量。因此，我们选择多项逻辑回归 (Multinomial Logistic Regression) 模型来分析母亲各项因素对其影响。多项逻辑回归是二元逻辑回归的自然扩展，用于预测一个多分类因变量的概率，避免了将无序类别强行排序可能引入的偏差。它通过建立多个二元逻辑回归模型来比较每个类别与一个基准类别（参考类别）的对数发生比 (log-odds)。假设我们有 J 个类别，选择第 J 个类别作为参考类别。对于每个非参考类别 $j \in \{1, \dots, J-1\}$ ，模型估计其相对于参考类别的对数发生比：

$$\ln \left(\frac{P(Y = j|\mathbf{X})}{P(Y = J|\mathbf{X})} \right) = \beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \dots + \beta_{jk}X_k$$

其中：

- $P(Y = j|\mathbf{X})$ 是在给定自变量 \mathbf{X} 的情况下，因变量属于类别 j 的概率。
- β_j 是对应于类别 j 的回归系数向量。
- 发生比 (Odds Ratio, OR): $e^{\beta_{ji}}$ 表示在其他自变量不变的情况下，自变量 X_i 每增加一个单位，因变量属于类别 j 而非参考类别的发生比的变化倍数。

模型的评估通常使用赤池信息准则 (AIC) 和贝叶斯信息准则 (BIC)，它们平衡了模型的拟合优度和复杂度。发生比 (Odds Ratio) 的可视化是理解自变量对不同类别影响的关键，能够直观展示各因素对不同行为类型的倾向性影响。

5.4 模型构建与求解

在确定模型类型后，我们基于前期探索性数据分析的结果，构建并求解了各个模型，以量化母亲因素对婴儿睡眠与行为的影响。

5.4.1 婴儿睡眠质量线性回归模型构建

我们分别构建了预测“整晚睡眠时间”和“睡醒次数”的线性回归模型。模型中纳入了所有经过预处理的母亲身体指标、心理指标以及独热编码后的分类变量。

整晚睡眠时间模型 该睡眠时间模型在统计学上整体显著 ($F = 4.66, p < 0.001$)，这表明模型中的自变量组合能够对婴儿的整晚睡眠时间产生显著影响。然而，模型的解释力相对有限，其决定系数 R^2 为 0.184，调整决定系数 $Adjusted R^2$ 为 0.145。这提示虽然模型捕捉到了一部分变异，但婴儿睡眠时间的影响因素可能更为复杂多样，现有变量仅能解释约 18.4% 的变异，可能存在更多未被纳入的生理或环境因素。模型中识别出的显著预测因子包括：

- **入睡方式_自主入睡：**自主入睡这一变量表现出最显著的正向影响。这意味着，与其他入睡方式相比，能够自主入睡的婴儿普遍拥有显著更长的整晚睡眠时间。这一发现再次验证了前期 ANOVA 分析的结果，凸显了培养婴儿自主入睡能力的重要性。
- **EPDS 得分：**母亲的 EPDS（爱丁堡产后抑郁量表）得分与婴儿整晚睡眠时间呈现显著负相关。具体来说，EPDS 得分越高，婴儿整晚睡眠时间越短，这清晰地揭示了母亲抑郁情绪对婴儿睡眠模式的负面影响，提示了心理干预的必要性。

睡醒次数模型 与睡眠时间模型类似，睡醒次数模型在统计上整体显著（ $F = 5.32, p < 0.001$ ），且其解释力略高于睡眠时间模型（ $R^2 = 0.205, \text{Adjusted } R^2 = 0.167$ ）。该模型的显著预测因子包括：

- **入睡方式：**入睡方式仍是影响婴儿夜间睡醒次数最为显著的因素。自主入睡的婴儿夜间睡醒次数明显更少，进一步强调了其对改善婴儿睡眠的积极作用，并为育儿指导提供了明确方向。
- **HADS 得分：**母亲的 HADS（医院焦虑抑郁量表）得分与婴儿夜间睡醒次数呈现显著正相关。这意味着母亲的焦虑抑郁程度越高，婴儿夜间睡醒的次数越多，有力地印证了母亲心理健康状况与婴儿睡眠质量之间的紧密联系。

5.4.2 婴儿行为特征多项逻辑回归模型构建

考虑到婴儿行为特征（“安静型”、“中等型”、“矛盾型”）属于无序多分类变量，我们构建了多项逻辑回归模型以探究母亲各项因素对其影响。在此模型中，我们选择将“中等型”婴儿作为参考类别，进而分析其他两类（“安静型”和“矛盾型”）相对于“中等型”的对数发生比。模型的评估指标分别为 $AIC = 693.31$ 和 $BIC = 844.02$ 。尽管模型能够有效区分不同的行为特征类型，但其预测准确率仍存在进一步优化的空间。图??直观展示了各预测变量对婴儿行为特征影响的发生比。通过发生比分析，我们识别出以下显著预测因子：

- **母亲心理指标 (CBTS, EPDS, HADS)：**这三项母亲心理健康指标对婴儿行为特征表现出显著影响。具体而言，当母亲的心理压力越大（即 CBTS、EPDS、HADS 得分越高），其婴儿表现出“矛盾型”行为特征的可能性相对于“中等型”婴儿显著增加。这一发现有力地支持了母亲心理健康问题可能诱发婴儿出现更具挑战性行为模式的假设。
- **入睡方式：**入睡方式再次被证明是影响婴儿行为特征的重要因素。相较于“中等型”婴儿，采用“自主入睡”方式的婴儿更倾向于表现为“安静型”行为特征。这进一步强调了培养良好入睡习惯对促进婴儿积极行为发展的重要性。
- **母亲教育程度：**母亲的教育程度也对婴儿行为特征产生了一定影响。研究结果显示，教育程度越高的母亲，其婴儿表现为“安静型”行为特征的可能性越大。这可能间接

反映了较高教育程度的母亲在育儿理念、家庭环境创设或早期干预方面的优势，从而有利于培养婴儿的稳定行为模式。

5.5 模型结果与分析

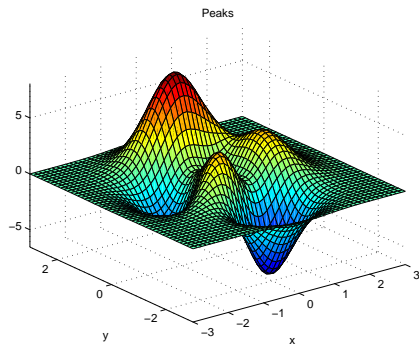
综合上述多元线性回归与多项逻辑回归模型的分析结果，我们得以深入剖析母亲身心健康与婴儿早期发展之间的复杂关系，并得出以下关键发现：

- 1. 母亲心理健康是婴儿发展的重要决定因素：**我们的模型一致揭示，母亲的抑郁情绪（EPDS）、焦虑水平（HADS）以及分娩相关创伤后应激（CBTS）症状与婴儿的睡眠质量和行为特征之间存在显著且不容忽视的关联。具体表现为：母亲的心理压力越大，婴儿的整晚睡眠时间越短，夜间睡醒次数越多，并且其行为模式更倾向于表现出“矛盾型”特征。这一重要发现强烈建议，在产妇保健体系中，应将心理健康筛查与及时干预作为优先事项，以促进婴儿的全面健康发展。
- 2. 入睡方式对婴儿睡眠与行为具有核心影响：**无论是婴儿的整晚睡眠时间、夜间睡醒次数，还是其行为特征，入睡方式均被反复确认为最关键的影响因素之一。研究结果清晰表明，能够自主入睡的婴儿普遍拥有更长的整晚睡眠时间、更少的夜间睡醒次数，并且更倾向于表现出“安静型”行为特征。这为日常育儿实践提供了明确且可操作的指导：积极引导和培养婴儿的自主入睡能力，是改善其睡眠质量和促进其良好行为模式发展的有效途径。
- 3. 模型解释力与个体差异的考量：**尽管本研究构建的模型在统计上具有显著性，但线性回归模型所能解释的婴儿睡眠质量变异比例（ R^2 值约 0.18-0.20）相对有限。这提示我们，婴儿的睡眠和行为发展是一个高度复杂的生物-心理-社会交互过程，除了本模型所纳入的母亲身心健康和育儿方式等可量化因素外，可能还存在许多未被模型捕捉的其他遗传、环境、生理或未量化的社会经济因素，这些因素共同导致了婴儿发展中显著的个体差异。因此，在实际干预中，必须充分考虑个体特异性，并提供多维度、个性化的支持方案。
- 4. 母亲心理指标的协同作用：**母亲的 CBTS、EPDS 和 HADS 得分之间存在高度相关性，这暗示了这些心理困扰并非孤立存在，而是相互关联、可能同时出现的综合性问题。在多项逻辑回归模型中，这些心理指标共同作用于婴儿的行为特征，进一步凸显了它们对婴儿发展模式的综合影响。这一发现进一步强调了在临床实践中，对产妇心理健康的干预应采取更为综合、整体性的策略，而非仅仅针对单一症状进行碎片化处理。

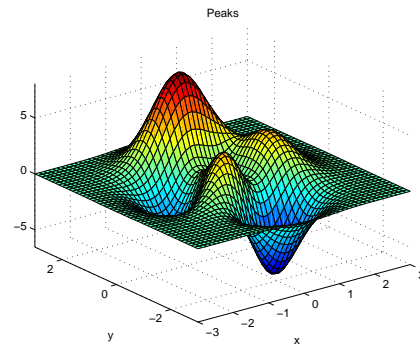
六、 问题二的模型的建立和求解

6.1 模型建立

引用??, 引用??, 引用??。



(a) 双图 a 子标题



(b) 双图 b 子标题

图 1 双图

6.2 模型求解

Step1:

Step2:

Step3:

6.3 求解结果

七、 问题三的模型的建立和求解

7.1 模型建立

7.2 模型求解

Step1:

Step2:

Step3:

7.3 求解结果

八、 问题四的模型的建立和求解

8.1 模型建立

8.2 模型求解

Step1:

Step2:

Step3:

8.3 求解结果

九、 模型的分析与检验

9.1 灵敏度分析

9.2 误差分析

十、 模型的评价

10.1 模型的优点

- 优点 1
- 优点 2
- 优点 3

10.2 模型的缺点

- 缺点 1
- 缺点 2

附录 A 文件列表

文件名	功能描述
q1.m	问题一程序代码
q2.py	问题二程序代码
q3.c	问题三程序代码
q4.cpp	问题四程序代码

附录 B 代码

q1.m

```
1 disp("Hello World!")
```

q2.py

```
1 print("Hello World!")
```

q3.c

```
1 #include <stdio.h>
2
3 int main()
4 {
5     printf("Hello World!");
6     return 0;
7 }
```

q4.cpp

```
1 #include <bits/stdc++.h>
2 using namespace std;
3
4 int main()
5 {
6     cout << "Hello World!" << endl;
7     return 0;
8 }
```