DeepSeek v3.1

1. 资料查找阶段,使用 AI 寻找 NIPT 与 BMI 关系的相关文献

帮我寻找和无创产前检测时间和BMI的关系有关的文献

已深度思考 (用时 20 秒) >

您好!很高兴能协助您进行数学建模竞赛的文献调研。针对"无创产前检测 (NIPT)时间选择与孕妇BMI关系"这一主题,该领域确实已有较为明确的研究共识。

核心结论是:**孕妇的BMI越高,其血液中的胎儿游离DNA (fetal fraction, FF) 比例通常越低**。因此,为了达到成功检测所需的FF阈值(例如对男胎Y染色体检测的4%),高BMI孕妇可能需要等待更长的孕周(即推迟检测时间)。

以下是一些关键文献和检索策略,希望能为您的模型提供理论支持。

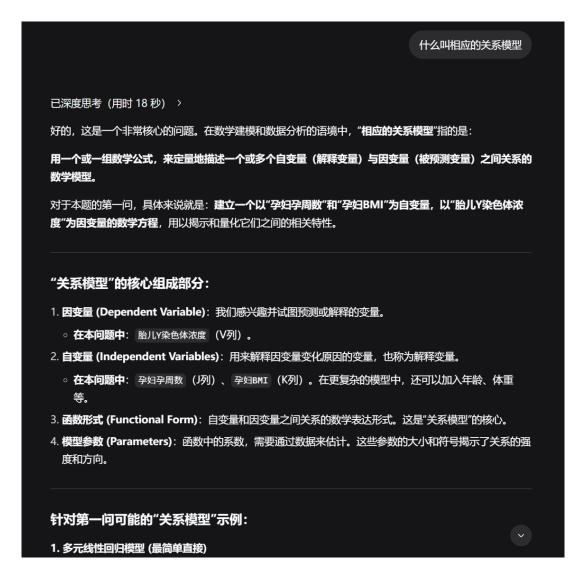
一、核心推荐文献(可直接用于支撑模型)

您可以通过【Google Scholar】、【PubMed】 (https://pubmed.ncbi.nlm.nih.gov/) 或【中国知网CNKI】等数据库,使用下述标题或DOI进行查找。

- 1. 文献标题: Analysis of fetal fraction in maternal plasma cell-free DNA and its correlation with maternal characteristics in Chinese women
 - ∘ 作者: Y. Liu, et al.
 - 。期刊: Prenatal Diagnosis
 - 发表年份: 2021年左右
 - 摘要: 这篇研究对中国人群进行了分析,明确展示了胎儿游离DNA比例与孕妇体重、孕周、BMI等指标的相关性。非常适合作为您问题1中建立关系模型的基础。文中很可能提供了具体的统计数据和相关性系数。
- 2. 文献标题: The impact of maternal plasma DNA fetal fraction on next generation sequencitests for common fetal aneuploidies

Deepseek 推荐的文献为我们寻找与 BMI 有关的因素变量提供了一定的参考,我们在文章中对其相关性进行了检验。

2. 构建模型阶段, 询问 AI 关系模型的相关解释, 了解相关信息



Deepseek 为我们解释了关系模型的含义,我们意识到构建关系模型的核心是为自变量和因变量构建合适的函数关系,以此来解释他们之间的联系。

同时他也给予了一些建模建议,我们通过分析发现并不适用本题目,最终决定不采纳。

3. 构建模型阶段, 与 AI 对话试图寻找多元线性回归的创新突破点

我应该如何在多元线性回归中做出创新

ን ዕ

已深度思考 (用时 21 秒) >

非常好的问题!在数学建模竞赛中,使用多元线性回归这类基础模型时,做出创新是脱颖而出、争取高奖的关键。单纯地跑一个回归、给出系数和p值远远不够。

以下是一些在多元线性回归框架内进行创新的策略,您可以根据数据和问题的特点选择组合使用:

1. 模型设定创新: 超越线性与独立

这是最核心、最有效的创新点。质疑"线性"和"相加"这两个基本假设。

- 引入非线性关系:
 - 多项式项: 如果散点图提示可能存在曲线关系, 可以加入平方项、立方项。
 - Y ~ BMI + BMI² + Gestational_Week
 - **创新解释**:这可以模拟"边际效应递减"或"加速效应"。例如,可能BMI在较低时对Y浓度影响不大,但超过某个阈值(如30)后,负面影响急剧增大。你可以通过求导找到这个"临界点"。
 - **分段回归(阈值效应)**: 假设临床已知BMI=35是一个重要门槛。你可以建立两个模型:
 - Y ~ Gestational_Week + BMI (对于 BMI < 35 的样本)
 - Y ~ Gestational_Week + BMI (对于 BMI >= 35 的样本)
 - **创新解释**: 比较两个模型的系数 (尤其是截距项和BMI的系数) 是否有显著差异。这能有力地证明**对于不同BMI人群,孕周与Y浓度的关系模式根本不同**,这直接引向了问题2和3的分组需求。
- 引入交互效应:
 - 添加交互项: Y ~ Gestational_Week + BMI + Gestational_Week * BMI
 - 。 **创新解释**:这检验了"孕周和BMI是否存在联合效应"。例如,交互项显著为负意味着:对于高BMI ~ 妇,每增加一周孕周所带来的Y浓度提升幅度,要小于低BMI孕妇。这是一个非常深刻且有临床意义的

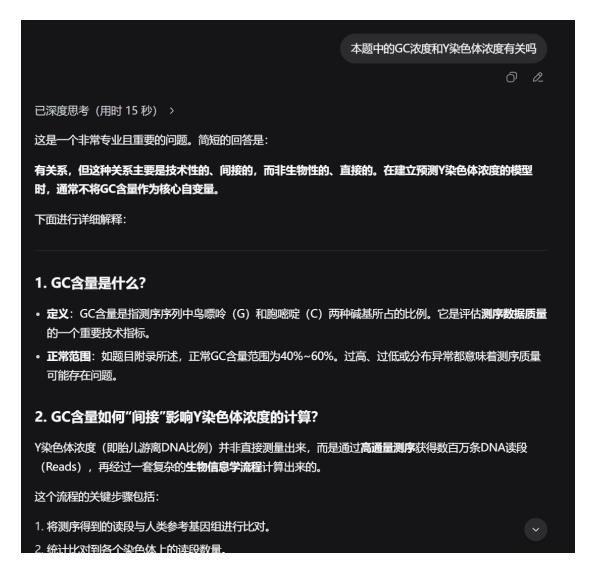
在探索建模过程中,此次 Deepseek 提到了分段,这启发了我们第二问以及第三问的建模过程,我们通过风险程度的分段构建了风险概率模型以及风险程度的量化。

4. 模型构建阶段, 询问 AI 有哪些可能与 Y 染色体浓度相关的指标

对于题中所给的指标,可能有哪些和y染色体浓度有关 已深度思考 (用时 44 秒) > 非常好,这是一个深入理解数据和问题的关键步骤。除了题目明确指出的孕周数和BMI,附件中还提供了 大量其他指标,这些都可能与Y染色体浓度 (即胎儿游离DNA比例) 直接或间接相关。 以下是对这些指标与Y染色体浓度相关性的详细分析,分为**强相关、可能相关**和**间接相关**三类: 一、 强相关指标 (建模时应优先考虑) 这些指标与Y染色体浓度的关系有明确的生物学或技术原理支持,是模型的核心自变量。 1. 孕周数 (Gestational Week) - 列J ○ 关系: 极强的正相关。随着孕周增加,胎盘增大,进入母体血液循环的胎儿细胞和游离DNA越多,因 此胎儿游离DNA比例 (ffDNA) 越高。这是最确定、最稳定的关系。 2. **孕妇BMI - 列K** ○ 关系: 显著的负相关。这是题目明确指出的重点。BMI高的孕妇,血容量更大,对胎儿游离DNA产 生"稀释效应"。同时,脂肪组织可能对游离DNA的释放、清除有影响,导致其背景浓度高而胎儿比例 相对降低。 3. **孕妇体重 (Weight)** - 列E ○ 关系:与BMI类似,负相关。体重是计算BMI的组成部分,因此它同样代表了血液稀释效应的物理基 础。注意:体重和BMI高度共线性,模型中通常只选其一,BMI是更标准化的指标。 二、 可能相关指标 (可作为模型增强项) 这些指标与V效色体浓度的关系在立起由有法法论或左左全理的生理机划。 可以作为各类结缔的 A 槽刑。 通

在思考与 Y 染色体浓度有关的因素时,我们询问了 Deepseek,期望能得到一些方向。对于 AI 的解答,我们选择了纳入考虑并检验相关性,而后继续探索新的可能的有关因素。

5. 第四问构建模型阶段, 让 AI 解释 GC 含量的意义以及与 Y 染色体浓度的关系



在解决第四问的过程中,要考虑的变量过多,且变量的具体含义我们并不 是很明白,于是求助了AI,并将其回答纳入考虑范围。

6. 第四问模型构建阶段, 使用 AI 寻找 GC 含量的相关文献



在第四问的构建过程中,我们不确定该如何理解并使用 GC 含量,可我们无 法将 AI 的回答与题目所给数据进行有效地关联起来。