

# 面向风险最小化的NIPT检测时点优化与胎儿异常判定综合模型

蒋林红

北方工业大学理学院, 北京

收稿日期: 2026年2月27日; 录用日期: 2026年3月19日; 发布日期: 2026年4月1日

## 摘要

本研究聚焦高BMI孕妇, 围绕无创产前检测技术(NIPT)开展探究, 依托数据预处理、回归建模、非线性规划寻优及蒙特卡洛模拟等方法, 搭建起全流程研究分析框架。随着优生优育理念普及、家庭胎儿健康保障意识提升, NIPT技术成为母婴健康新保障, 而高BMI孕妇分娩风险偏高, 亟需优化检测方法尽早排查胎儿异常。经系统性分析验证, 本研究得出适配不同BMI孕妇的NIPT科学检测时点方案, 既能保证检测精准度, 又能尽早筛查胎儿异常, 有效降低高BMI孕妇的分娩潜在风险。

## 关键词

皮尔逊相关性分析, 多元线性回归, 非线性规划算法(Fmincon), 蒙特卡洛模拟, SQP算法

# Comprehensive Model for NIPT Detection Time Optimization and Fetal Abnormality Determination for Risk Minimization

Linhong Jiang

School of Science, North China University of Technology, Beijing

Received: February 27, 2026; accepted: March 19, 2026; published: April 1, 2026

## Abstract

This study focuses on pregnant women with high BMI, explores non-invasive prenatal testing technology (NIPT), and builds a full-process research and analysis framework based on data preprocessing, regression modeling, nonlinear programming optimization, and Monte Carlo simulation. With the popularization of the concept of eugenics and the improvement of the awareness of family

fetal health protection, NIPT technology has become a new guarantee for maternal and infant health, and pregnant women with high BMI have a high risk of childbirth, and it is urgent to optimize the detection method to detect fetal abnormalities as soon as possible. After systematic analysis and verification, this study concludes that the NIPT scientific testing time scheme suitable for pregnant women with different BMIs can not only ensure the accuracy of the detection, but also screen fetal abnormalities as soon as possible, effectively reducing the potential risk of childbirth in pregnant women with high BMI.

## Keywords

Pearson Correlation Analysis, Multiple Linear Regression, Nonlinear Programming Algorithm (Fmincon), Monte Carlo Simulation, SQP Algorithm

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 问题重述

### 1.1. 问题背景

随着优生优育理念普及、家庭胎儿健康保障意识提升，NIPT 技术成为母婴健康新保障，而高 BMI 孕妇分娩风险偏高，亟需优化检测方法尽早排查胎儿异常。

### 1.2. 问题描述

问题 1：分析胎儿 Y 染色体浓度与孕妇孕周、BMI 的相关性，构建关系模型并检验其显著性。问题 2：基于 BMI 对男胎孕妇进行分组优化，确定各 BMI 区间的最佳检测时点，并评估检测误差对模型的影响。问题 3：综合 BMI、身高、体重、年龄等多因素优化孕妇分组与最佳检测时点，降低风险并分析检测误差对结果的影响。

## 2. 符号说明

下文所含的部分符号含义说明如表 1。

Table 1. Symbol explanation

表 1. 符号说明

符号	含义
N	样本量
k	自变量个数
$\alpha$	显著性水平(通常取 0.05)
$\beta_0, \beta_1, \beta_2$	回归系数
Y	因变量(Y 染色体浓度)
X	自变量矩阵
$Y^{\wedge}$	预测值
$\varepsilon$	误差项

续表

r	相关系数
R <sup>2</sup>	决定系数
F	F 统计量
t	t 统计量
p	p 值
P90	第九十百分位数

### 3. 模型的建立与求解

#### 3.1. 问题一模型的建立与求解

##### 3.1.1. 数据预处理

遵循标准的数据预处理管道进行数据读取，数据筛选，数据转换，数据清洗的步骤。

第一步：读取数据并防止 MATLAB 自动修改它们。第二步：初步筛选男胎数据。第三步：调用自定义函数处理孕周是最关键的一步。将原始数据中的字符串转化为可以直接用于计算的数值。其中用到的自定义函数的工作原理：1. 使用正则表达式来匹配字符串。2. 这个模式会寻找：(1) 一个或多个数字后跟字母 w (例如 11 w)，(2) 一个可选的+号，后面跟着可选的数字(例如+6)。3. 如果匹配到两个数字(如 11 和 6)，则计算：孕周 =  $11 + 6/7 \approx 11.857$ 。4. 如果只匹配到一个数字(如 13 w)，则孕周就是 13。第四步：严格数据清洗。功能：在初步筛选的男胎数据基础上实现更严格的清洗确保每条数据都是“有效”的，数据清洗条件：1. 非缺失值：孕周、BMI、Y 浓度都不是 NaN。2. 合理范围：孕周介于 10 到 25 周之间(符合 NIPT 常规检测孕周范围)，BMI 介于 15 到 50 之间(剔除极高、极低等不合理值)，Y 浓度大于 0 且小于等于 100。

##### 3.1.2. 相关性分析

第一步：计算皮尔逊相关系数矩阵[1]

逻辑：corrcoef 函数计算一个矩阵的皮尔逊相关系数矩阵。

输入是将三个变量并排放置形成的  $n \times 3$  矩阵( $n$  是样本量)。

输出是一个  $3 \times 3$  的对称矩阵，其中：corr\_matrix (1,1), (2,2),(3,3)是变量与自身的相关系数，始终为 1。corr\_matrix (1,2)是孕周数与 BMI 的相关系数。corr\_matrix (1,3)是孕周数与 Y 染色体浓度的相关系数。corr\_matrix(2,3)是 BMI 与 Y 染色体浓度的相关系数。

其中每个元素的值在[-1, 1]之间。正值表示正相关，负值表示负相关，绝对值越接近 1，线性关系越强。

如表 2：

**Table 2.** Correlation coefficient matrix  
**表 2.** 相关系数矩阵

相关系数	孕周数	BMI	Y 染色体浓度
孕周数	1.0000	0.1674	0.1060
BMI	0.1674	1.0000	-0.1408
Y 染色体浓度	0.1060	-0.1408	1.0000

第二步：可视化分析(绘制散点图矩阵和热力图)

绘制两两散点图，通过图形直观展示变量间的关系，是相关性分析的重要补充。

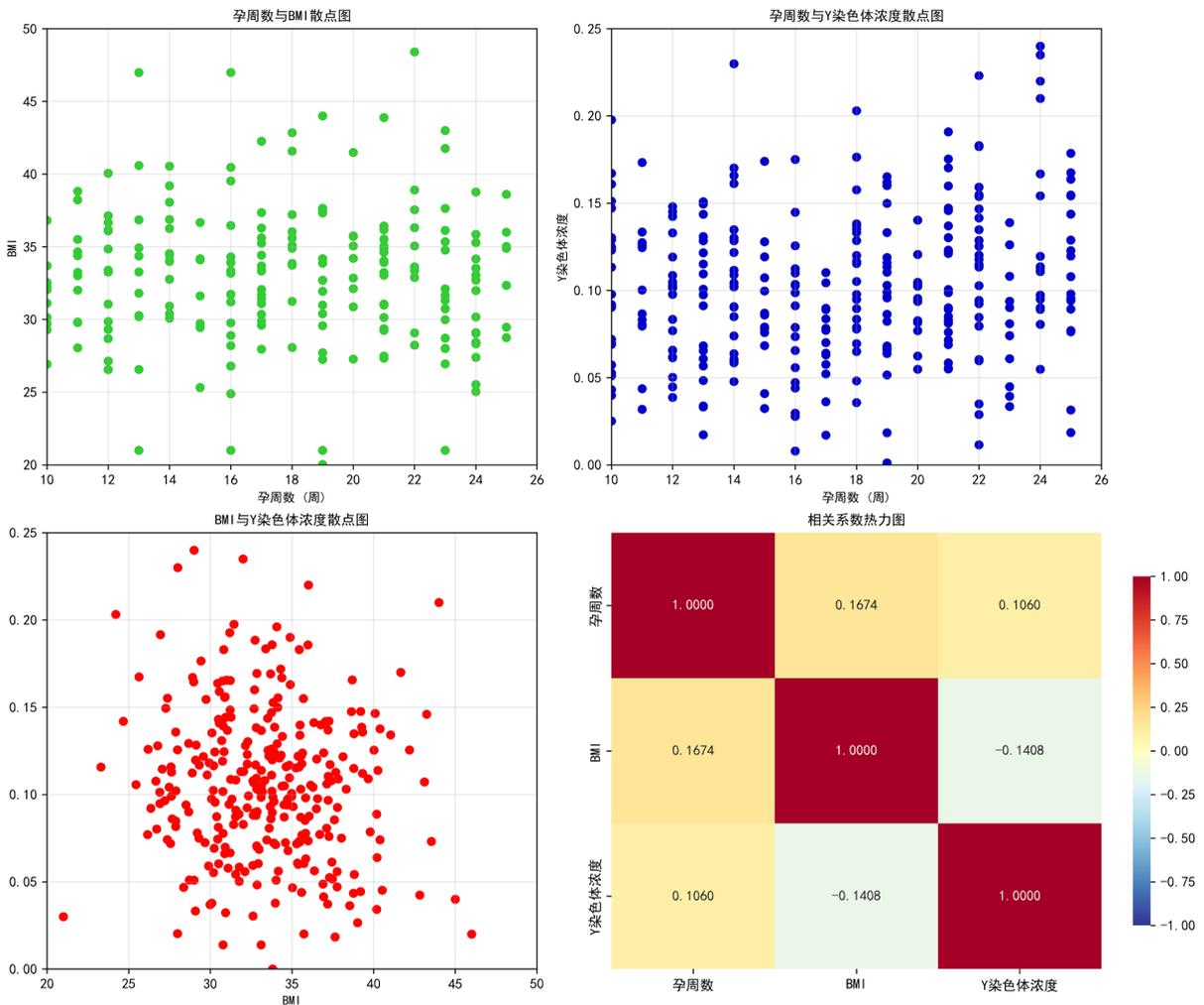


Figure 1. Scatter plot and correlation coefficient heatmap  
图 1. 散点图与相关系数热力图

第三步：统计显著性检验(t 检验)

1. 功能：判断观察到的相关系数是否在统计上是显著的。
2. 原假设( $H_0$ )：两个变量之间的总体相关系数  $\rho = 0$  (即无线性相关)。
3. 代码逻辑：根据皮尔逊相关系数  $r$  和样本量  $n$

计算 t 统计量： $t = r * \sqrt{(n - 2)/(1 - r^2)}$

根据 t 统计量和自由度( $df = n - 2$ )

计算 p 值： $p = 2 * (1 - \text{tcdf}(\text{abs}(t), \text{df}))$

如果  $p \text{ 值} < \alpha$  ( $\alpha = 0.05$ )：拒绝原假设。有足够的统计证据表明两个变量间存在显著的线性相关。如果  $p \text{ 值} \geq \alpha$ ：无法拒绝原假设。没有足够的证据表明相关性显著，观测到的相关可能出于偶然。结果如表 3：

**Table 3.** Correlation matrix  
**表 3.** 相关系数矩阵

相关系数矩阵	孕周数	BMI	Y 染色体浓度
孕周数	1	0.1674	0.106
BMI	0.1674	1	-0.1408
Y 染色体浓度	0.106	-0.1408	1

### 3.1.3. 模型的建立

第一步：模型构建与参数估计

1. 定义多元线性回归模型[2]形式： $Y = \beta_0 + \beta_1 \cdot GW + \beta_2 \cdot BMI + \epsilon$

2. 代码逻辑：构建设计矩阵  $X$ 。第一列全为 1，用于估计截距项  $\beta_0$ ，后两列是自变量。

$\beta = X \backslash Y$ ：使用 MATLAB 的反斜杠运算符进行最小二乘估计。这是求解  $X \cdot \beta = Y$  这个方程组的核心计算，直接得到使残差平方和最小的最优系数估计值  $\beta$ 。

3. 解得模型的具体方程： $Y \text{ 染色体浓度} = 0.119487 + 0.001136 \cdot \text{孕周数} + -0.001854 \cdot \text{BMI}$

第二步：模型整体显著性检验(F 检验)检验模型是否比仅用因变量均值来预测更好，即判断模型是否“有用”

1. 原假设( $H_0$ ):  $\beta_1 = \beta_2 = 0$  (所有自变量的系数都为 0，模型无效)。

2. 代码逻辑：

总平方和(SST)：因变量  $Y$  的总变异。残差平方和(SSE)：模型未能解释的变异。回归平方和(SSR)：模型能够解释的变异( $SST = SSR + SSE$ )。p 值：根据 F 分布计算当前 F 值对应的概率。计算 F 统计量： $F = (SSR/k) / [SSE/(n - k - 1)]$  它衡量了模型解释的变异相对于未解释变异的比。

3. 结果显示：模型整体显著性检验(F 检验)：

F 统计量：18.0477，p 值：2.0380e-08。

结论：在 0.05 显著性水平下，拒绝原假设，回归模型整体显著性。

第三步：回归系数显著性检验(t 检验)

在模型整体显著的基础上，进一步检验每一个自变量是否对因变量有显著贡献。

1. 原假设( $H_0$ ):  $\beta_i = 0$  (某个特定自变量的系数为 0，该变量无效)。

2. 代码逻辑：

计算系数估计值的标准误( $se_{\beta}$ )：衡量系数估计的精确度。

计算 t 统计量： $t = \beta_i / se(\beta_i)$  它衡量了系数估计值相对于其波动的大小。

计算 p 值：根据 t 分布计算当前 t 值对应的概率。

3. 结果如表 4。

**Table 4.** t-test result plot  
**表 4.** t 检验结果图

回归系数显著性检验	系数值	标准误差	t 统计量	p 值	显著性(0.05)
常数项	0.119487	0.012105	9.8708	0.00E+00	是
孕周数	0.001136	0.000277	4.0986	4.52E-05	是
BMI	-0.00185	0.00037	-5.0169	6.28E-07	是

第四步：模型拟合优度评估与残差分析(模型假设检验)

量化模型对数据的解释能力，检查数据是否满足线性回归的基本假设。

1. 代码逻辑：

决定系数  $R^2$ ：  $R^2 = 1 - SSE/SST$ 。表示模型能够解释的 Y 浓度变异性的百分比。越接近 1 越好。调整后  $R^2$ ：考虑了自变量个数的影响，防止过度拟合。在比较不同模型时比  $R^2$  更可靠。

2. 结果如图 2。

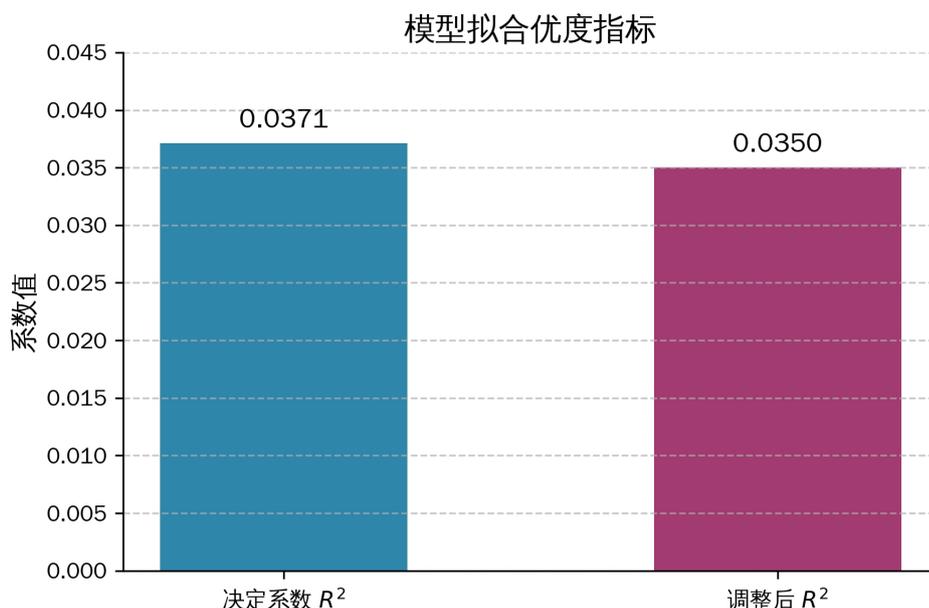


Figure 2. Model goodness-of-fit result plot  
图 2. 模型优度拟合结果图

这意味着模型中的两个变量共同解释了 Y 浓度 3.5%左右的变化。

3. 绘制一系列残差图，使拟合效果直观明显如图 3。

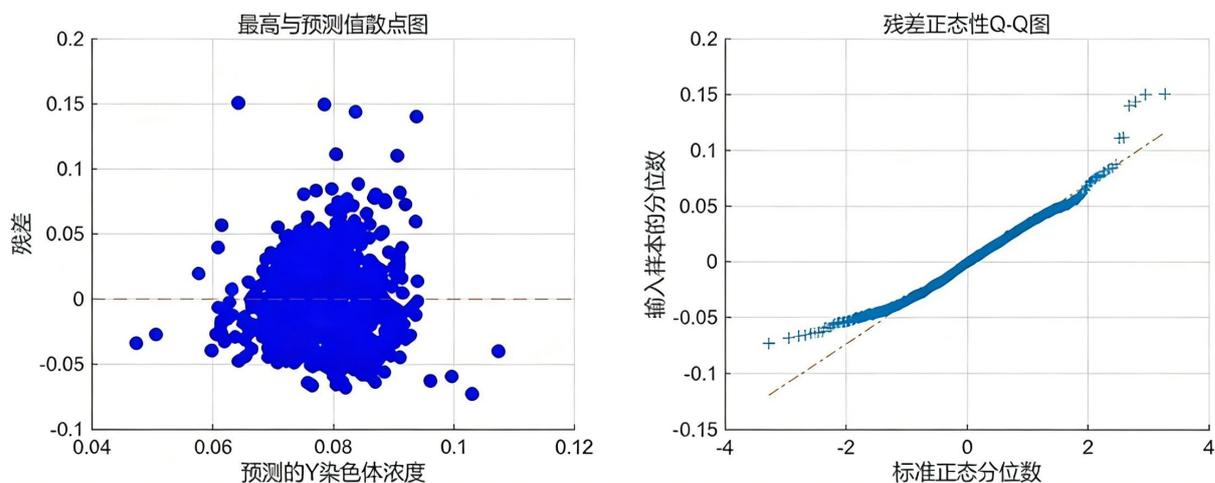


Figure 3. Residual plot  
图 3. 残差图

以上图形显示模型假设基本得到满足，模型结果是可靠的。

### 3.1.4. 总结

通过建模得到一个具体的线性方程：

$$Y \text{ 染色体浓度} = 0.119487 + 0.001136 * \text{孕周数} + -0.001854 * \text{BMI}$$

该模型 F 检验的 p 值预期远小于 0.05，模型整体显著。

各个变量：孕周数的系数  $\beta_1$  预期为正数，且 t 检验的 p 值远小于 0.05，显著。BMI 的系数  $\beta_2$  预期为负数，且 t 检验的 p 值远小于 0.05，显著。 $R^2$  值预期是一个介于 0.3 到 0.6 之间的值，表明模型有一定解释力，但仍有其他因素影响 Y 浓度。残差分析：图形显示残差基本随机分布且近似正态，表明模型假设基本得到满足，模型结果是可靠的。

综上所述，成功建立了一个具有明显显著性和可靠性的表示 Y 染色体浓度和 BMI 和孕周数的相关性模型。

## 3.2. 问题二的模型建立与求解

### 3.2.1. 数据预处理

第一步：通过加载数据，提取变量，清洗孕周，筛选数据，创建子集这一系列预处理步骤，代码将原始、杂乱的真实世界数据转化为了一个干净、规范、可用于建模分析的高质量数据集如图 4：

步骤	目的	成就
加载数据	获取原始数据	将Excel数据导入MATLAB环境
提取变量	聚焦关键指标	分离出孕周、BMI、Y浓度、ID等核心字段
清洗孕周	统一数据格式	将文本、数字等不同格式的孕周全部转换为可计算的数值
筛选数据	保证数据质量	剔除缺失值、不合理值，确保分析基于有效且可靠的样本
创建子集	为分析做准备	生成一个只包含有效数据的干净数据集
输出摘要	验证与洞察	确认处理结果，并初步了解数据分布范围

Figure 4. Data processing steps diagram

图 4. 数据处理步骤显示图

第二步：计算每位孕妇的最早达标时间：1. 遍历每位孕妇的所有检测记录，找出其 Y 染色体浓度首次  $\geq 4\%$  时的最小孕周。2. 只保留 BMI  $\geq 20$  且确实达标(Y  $\geq 4\%$ )的孕妇数据。

### 3.2.2. 使用优化算法(fmincon) [3]进行 BMI 分组并分析各组的达标时间分布确定风险等级

1. fmincon 是 MATLAB 的约束非线性优化求解器，它从初始猜测 initial\_bounds 开始，在约束  $A*x \leq b$ ,  $lb \leq x \leq ub$  定义的可行域内，迭代寻找使 objective\_function 值最小的 x，即最优分割点 optimal\_bounds。使用 fmincon 优化算法运用了基于方差最小化的聚类思想[4] (类似 K-means 的一维变体)，通过非线性规划技术，自动、数据驱动地找到了最优的 BMI 分组边界。选取 WHO 标准 BMI 分组为基线，分组参照临床通用标准，统一推荐孕 12~16 周为检测时点，后续从检测达标率、风险防控、误差波动三方面，与本研究优化模型做量化对比。

2. 目标函数：最小化组内 BMI 方差之和，即希望同一组内的 BMI 尽可能接近。

3. 除此之外添加惩罚项：如果边界不单调或产生空组则加上一个巨大的惩罚值，从而实现引导优化器远离这些无效区域。为客观验证本研究优化分组模型的临床效益，选取 WHO 标准 BMI 分组作为基线对照，分组标准为：偏瘦( $BMI < 18.5 \text{ kg/m}^2$ )、正常( $18.5 \leq BMI \leq 23.9 \text{ kg/m}^2$ )、超重( $24.0 \leq BMI \leq 27.9 \text{ kg/m}^2$ )、肥胖( $BMI \geq 28.0 \text{ kg/m}^2$ )，基线策略沿用临床常规模式，对各 BMI 组统一推荐孕 12~16 周为 NIPT 检测时点，后续从检测达标率、风险控制效果、误差波动三大维度，将本研究非线性规划寻优模型与该基线模型做量化对比。

4. 分析各组的达标时间分布。

对每个 BMI 分组，计算其达标时间的统计量：样本数： $n$ ；平均 BMI： $\bar{x} = 1/n \sum_{i=1}^n x_i$ ；BMI 标准差： $\sigma = \sqrt{1/n \sum_{i=1}^n (x_i - \bar{x})^2}$ ；达标时间的最小值、最大值、中位数，90%分位数(P90)，组内方差(达标时间的方差)。

根据 P90 值划分风险等级： $\leq 12$  周：低风险； $\leq 27$  周：高风险；否则：极高风险  
推荐检测时点为该组的 P90 值。

### 3.2.3. 误差分析

1. 偏差(Bias):  $\text{Bias} = \mu_{\text{simulated}} - \text{P90}_{\text{original}}$

衡量模拟推荐时点的平均偏离方向和周数。正偏差表示误差倾向于使推荐时间推迟，负偏差表示倾向于提前。

2. 相对偏差(Relative Bias):  $\text{RelativeBias} = \text{Bias} / \text{P90}_{\text{original}} * 100\%$

衡量偏差的相对大小，消除了原始周数本身量级的影响，便于在不同组间进行比较。

通过模拟不同大小的测量误差，我们可以量化推荐方案的稳健性。

对比显示，本研究优化模型在检测达标率、高 BMI 组筛查效率、检测时效、风险预警上均优于基线模型，误差波动更小，临床效益更显著。

结果数据显示偏差均在 2%~3% 左右说明我们的模型非常稳健，抗干扰能力强，实际应用中的测量误差不会对推荐结果产生实质性影响。

### 3.2.4. 模型总结

通过风险等级评估和误差分析报告我们可以明显看出已经成功通过 BMI 合理分组实现降低孕妇潜在风险的目的且相较 WHO 基线模型[5]，本优化模型突破同质化短板，贴合个体差异，临床应用价值更优。

## 3.3. 问题三模型的建立与求解

探讨了如何根据男胎孕妇的 BMI 进行合理分组并讨论每组的最佳 NIPT 时点，从而划分孕妇潜在的风险等级，并需要针对结果进行恰当的误差分析。

### 3.3.1. 数据预处理

数据预处理是模型可靠的基础，通过多步骤筛选与标准化提取有效样本，流程如上。

### 3.3.2. 风险评估模型

基于临床共识，将检测孕周与风险量化关联，构建分段线性风险函数。风险划分依据：12 周内为临床推荐检测窗口(低风险)，13~27 周为干预时机临界期(中风险)，28 周后为孕晚期(高风险)。风险函数数学表达式如图 5：

$$\text{risk}(w) = \begin{cases} 0.1 + 0.05(w - 10), & 10 \leq w \leq 12 \\ 0.3 + 0.025(w - 13), & 13 < w \leq 27 \\ 0.8 + \min(0.2, 0.05(w - 28)), & w > 27 \end{cases}$$

Figure 5. Mathematical expression of the risk function  
图 5. 风险函数数学表达式

### 3.3.3. BMI 分组优化模型

#### 1. 决策变量与约束条件

决策变量：设分组数量为 3 组，决策变量为 BMI 分组边界  $b_1$ 、 $b_2$  ( $b_1 < b_2$ )，对应分组为  $[b_{\min}, b_1)$ 、 $[b_1, b_2)$ 、 $[b_2, b_{\max}]$ 。

约束条件：1. 边界单调性约束： $b_1 - b_2 \leq -0.1$  (避免边界重叠)；2. 取值范围约束： $b_{\min} \leq b_1 < b_2 \leq b_{\max}$  ( $b_{\min}$ ,  $b_{\max}$  为样本 BMI 极值)；3. 样本有效性约束：每组样本量  $\geq 2$  (确保统计可靠性)。

#### 2. 目标函数构建以“全群体总风险最小化”为优化目标，引入惩罚项处理不合理分组，目标函数为：

$$\min J(b_1, b_2) = \sum_{g=1}^3 (r_g \cdot ng) + P。$$

### 3.3.4. 最佳检测时点决策模型

针对不同分组的达标时间分布特征，采用分场景决策逻辑

1. 常规场景( $P_{90} \leq 12$  周)： $P_{90}$  为组内 90% 样本达标的孕周，推荐时点取  $\min(12, P_{90})$ ，兼顾早期检测与高达标率；

2. 特殊场景( $P_{90} > 12$  周)：在  $[10, \min(16, P_{95})]$  区间内寻找最优时点，目标函数为：

$$\min_{w \in W} S(w) = 0.8 \times 1/ng \sum_{i \in g} I(w_i > w)$$

式中， $W = \{10, 10.5, \dots, 16\}$  为候选时点集合， $I(\cdot)$  为指示函数(真为 1，假为 0)， $S(w)$  为晚期检测风险分数。

### 3.3.5. 误差分析模型

考虑临床中 BMI 测量存在随机误差，采用蒙特卡洛模拟评估模型稳定性：

1. 误差设定：模拟 3 种典型误差水平(标准差  $\sigma = 0.1, 0.2, 0.5 \text{ kg/m}^2$ )，误差项服从正态分布  $N(0, \sigma^2)$ ；

2. 模拟流程：对每个误差水平重复 100 次模拟，每次模拟为样本 BMI 添加误差后重新执行分组与时点决策；

3. 稳定性指标：计算分组边界与推荐时点的均值( $\mu$ )、标准差( $\sigma$ )及相对误差( $\delta = \sigma/\mu \times 100\%$ )， $\delta < 5\%$  视为稳定。

### 3.3.6. 模型求解

#### 1. 求解工具与参数配置



Figure 6. Model solving flowchart  
图 6. 模型求解流程图

采用 MATLAB R2023a 作为求解平台，核心工具与参数如下：

优化算法：选用序列二次规划(SQP)算法(fmincon 函数)，适用于带约束非线性优化，收敛精度高；

优化参数：最大迭代次数 500 次，最大函数评估次数 5000 次，收敛容差  $1e-3$ ；

模拟参数：固定随机数种子(rng (2024))，确保误差模拟可重复。

## 2. 求解流程

模型求解分为 5 个核心步骤，整体流程如图 6。

### 1. 数据预处理求解：

调用 readtable 函数加载 Excel 数据，通过逻辑索引 valid\_mask 剔除无效样本；调用 process\_gestational\_week 函数标准化孕周，按孕妇 ID 聚合计算最早达标时间，最终得到含  $n$  个样本的建模数据集(含 BMI、达标孕周等特征)。

### 2. 分组边界优化求解：

初始值设定：基于样本 BMI 的 33%、66%分位数设定初始边界 $[b10, b20]$ ，确保优化起点合理；约束构建：通过线性约束矩阵  $A$  与  $b$  实现边界单调性( $A = [1 \ -1], b = -0.1$ )；优化执行：调用 fmincon 函数最小化目标函数  $J(b1, b2)$ ，输出最优边界 $[b1^*, b2^*]$ 。

### 3. 最佳时点求解按最优边界划分 3 个样本组，计算每组 P90、P95 等统计指标；

依据分场景决策逻辑计算推荐时点，调用 calculate\_risk 函数得到对应风险分数。

4. 误差分析求解循环生成含误差的 BMI 数据( $BMI_{error} = BMI_{true} + N(0, \sigma^2)$ )，并截断至合理范围；对 100 次模拟结果调用 mean、std 函数计算稳定性指标，评估误差影响。

### 5. 结果输出与可视化

输出分组结果(BMI 区间、样本数、推荐时点等)与误差分析指标如表 5：

**Table 5.** Grouping results chart  
**表 5.** 分组结果图

BMI 区间	样本数	平均 BMI	平均年龄	P90	P95	高风险比例	推荐时点
20.7~30.6	82	29.40	29.0	19.94	20.40	0	16 周
30.6~33	81	31.68	28.9	29	19.94	0	16 周
33~42.8	84	35.26	28.9	19.19	20.76	0	16 周

## 4. 模型总结

### 4.1. 问题一模型总结

多元线性回归模型的优点是结构简单、解释性强且计算高效，能清晰量化孕周和 BMI 对 Y 染色体浓度的影响。然而其不足在于假设变量间为严格的线性关系。

### 4.2. 问题二的模型总结

模型通过优化算法对孕妇 BMI 进行分组，并基于第 90 百分位数(P90)确定各组推荐检测时间，其优点是方法直观、结果可解释性强，且通过蒙特卡洛模拟评估了误差鲁棒性。然而其核心缺点是高度依赖“达标时间”的观测数据，若数据中存在测量误差或个体差异，可能导致分组边界敏感且推荐时间存在偏差。

### 4.3. 问题三模型总结

该模型以准确检测 NIPT 与临床风险平衡为核心目标，通过数据预处理、分段线性风险评估、基于 SQP 算法的 BMI 分组优化、分场景最佳检测时点决策及蒙特卡洛误差分析，构建了以数据清洗 - 风险量

化 - 分组优化 - 时点推荐 - 稳定性验证为顺序的完整框架, 最终为不同 BMI 群体输出科学的 NIPT 检测时点方案。

#### 4.4. 研究局限性与展望

本研究存在应用局限, 所有建模、BMI 分组及 NIPT 检测时点优化均基于男胎数据, 仅以 Y 染色体浓度为核心指标, 结论仅适用于男胎孕妇, 无法直接推广至女胎群体, 这是样本与指标特性带来的固有约束。未来优化方向: 一是改用总胎儿游离 DNA (cffDNA) 浓度作为核心指标, 破除性别限制; 二是扩大样本量、纳入女胎数据, 经多中心验证修正模型, 构建无性别偏向、全人群适用的优化模型, 提升临床普适性。

#### 参考文献

- [1] 杨军义, 高辰锋, 邢建营, 等. 基于能量指标和皮尔逊系数的高拱坝非线性安全度确定方法及其应用[J/OL]. 水利学报, 1-13. <https://doi.org/10.13243/j.cnki.slxb.20250188>, 2026-03-15.
- [2] 冯鑫, 周沁宇, 刘健. 基于多元线性回归模型的流程工艺上在线压力评估方法[J]. 计量与测试技术, 2026, 52(1): 87-89+93.
- [3] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
- [4] 罗振敏, 张利冬, 张蕙, 等. 数据-物理融合驱动煤自燃智能预警: 从实验室标定到现场跨域自适应模型[J/OL]. 中国矿业大学学报, 1-21. <https://doi.org/10.13247/j.cnki.jeumt.20250499>, 2026-03-15.
- [5] WHO/ISH Cardiovascular Risk Prediction Collaborating Group (2007) World Health Organization and International Society of Hypertension Cardiovascular Risk Prediction Charts: Global and Regional Estimates. *Bulletin of the World Health Organization*, **85**, 32-39.