问题 2 临床证明，男胎孕妇的 BMI 是影响胎儿 Y 染色体浓度的最早达标时间（即浓度达到或超过 4%的最早时间）的主要因素。试对男胎孕妇的 BMI 进行合理分组，给出每组的 BMI 区间和最佳 NIPT

时点，使得孕妇可能的潜在风险最小，并分析检测误差对结果的影响。

问题2：

第二问侧重于探究BMI对胎儿Y 染色体浓度最早达标时间的影响而非关注同一个人BMI随时间的变化，又因为单个样本BMI随时间变化的幅度非常小且个体差异远大于BMI随时间的变化量，所以在这一问中我们将使用个体所有记录中BMI的平均值来进行统计和分类。

为了更合理地划分BMI区间，我们先依据Y染色体浓度达标情况对孕妇进行了分组并计算了每组孕妇的BMI平均值，以此来为划分提供参考。

我们先对原始数据进行了筛选和分组。我们依据孕妇是否存在Y染色体浓度达标情况将原数据大致分成了三组：第一组为从第一次检测开始Y染色体浓度就达标（always\_can\_test）的情况，结果保存在bmi\_Y\_always\_can\_test\_result.xlsx中，包括【孕妇代码，BMI，最早达标天数】三列数据；第二组为开始Y染色体浓度不达标，后来Y染色体浓度达标的情况（middle），结果保存在bmi\_Y\_middle\_result.xlsx中，包括【孕妇代码，BMI，预测达标天数】三列数据；第三组为Y染色体浓度一直不达标（cannot\_test）的情况，结果保存在bmi\_Y\_cannot\_test\_result.xlsx中，包括【孕妇代码，BMI，最晚达标天数】三列数据。其中最早达标天数和最晚达标天数直接取自表中，而预测达标天数使用**插值法**得到（详见附录代码？？？line）。

如果某编号样本某次Y染色体浓度大于等于4%，后面又下降至低于4%，则该样本被判断为无效样本并被筛去。

分类过后总样本量为230人，第一组样本量为186人（占总比80.87%），第二组样本量为37人（占总比16.09%），第三组样本量为7人（占总比3.04%）。、bmi\_Y\_middle\_result.xlsx和bmi\_Y\_cannot\_test\_result.xlsx文件中。

为了提炼三种情况(always\_can\_test；middle；cannot\_test)孕妇的BMI特点，我们分别对三组所有样本的BMI和达标天数求了平均值，结果如下：

|  |  |  |
| --- | --- | --- |
|  | BMI | 对应天数 |
| always\_can\_test | 31.96548919 | 88.51075269 |
| middle | 33.37592436 | 112.1758142 |
| cannot\_test | 35.21942627 | 139.7142857 |

从统计数据不难看出：BMI小的个体Y染色体达标的时间有更早的趋势，下面我们就依据三组的BMI平均值将BMI分为5个范围：<30,30-32,32-34,34-36,>36。

接着我们进行每组的最佳 NIPT 时点的计算。