**2006高教社杯全国大学生数学建模竞赛**

**承 诺 书**

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则.

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们参赛选择的题号是（从A/B/C/D中选择一项填写）： B

我们的参赛报名号为（如果赛区设置报名号的话）：

所属学校（请填写完整的全名）： 江苏科技大学

参赛队员 (打印并签名) ：1. 刘硕

2. 令狐岩松

3. 余骁

指导教师或指导教师组负责人 (打印并签名)： 王承毅

日期： 2006 年9 月15 日

赛区评阅编号（由赛区组委会评阅前进行编号）：

**2006高教社杯全国大学生数学建模竞赛**

**编 号 专 用 页**

赛区评阅编号（由赛区组委会评阅前进行编号）：

赛区评阅记录（可供赛区评阅时使用）：

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 评  阅  人 |  |  |  |  |  |  |  |  |  |  |
| 评  分 |  |  |  |  |  |  |  |  |  |  |
| 备  注 |  |  |  |  |  |  |  |  |  |  |

全国统一编号（由赛区组委会送交全国前编号）：

全国评阅编号（由全国组委会评阅前进行编号）：

**艾滋病疗法的评价及疗效的预测**

**关键词：Pearson相关性 多项式逼近拟合 灰色预测 效益值模型 非线性规划模型 加权Topsis**

**摘要：**

本文通过对数据进行分析和利用，研究了数据之中的信息，得出了不同疗法的治疗特点。在拟合了大量数据后，通过不同方法、不同角度对四种疗法进行了预测和评价。在得到价格信息后针对数据和治疗特点设计出有效的模型和相匹配的求解算法，为广大需求不同的病人提供治疗建议。

**针对问题一**：通过对数据的宏观分析，确定解决方向。首先对数据进行**Pearson相关性**检测，发现CD4与HIV之间**在99%的置信水平拒绝原假设，CD4与HIV具有很强的相关性，且成中等程度负相关**。接着对数据进一步进行使用插值-拟合的方法，斟酌后最终选取**多项式逼近拟合**，得出CD4和HIV的期望拟合表达式和期望走势图，并由此得知在该疗法下**CD4增长基本趋于平稳，而HIV的在下降的同时会出现两次反弹。**在对300多名病人进行统计分析后划分为五个状态，最终预测不能继续治疗需要提前终止治疗，共有273位，占比**77%**，出现了第一次反弹和继续治疗状况的病人是可以在测试终止后继续服药，共有81位，占比**23%，**并且使用该疗法的病人**推荐在12周后停止使用**。

**针对问题二**：

**针对问题三**：在前置分析后衡量了加上消费因素的评判标准，设立了衡量效益的指标。在对该指标的表达式推导和转化后得出**效益值模型**，之后配合问题将模型再次转化为**非线性规划模型**。为求解模型针对模型特点讨论求解策略，设计出相匹配的遍历求最优解的算法并完成实现。使用算法通过计算得出**在取得最大效益下四种疗法的最佳停止时间、费用、效用、增长率积分**，并画出了不同疗法时间与效益关系的t-Q图，和费用和效益关系的C-Q图进行分析，点出四种疗法在整体上的优缺点。最后采取**加权Topsis**对得出的数据进行评价和打分，得出病人在考虑费用的情况下**最佳疗法选择应为：3>4>2>1**。

综合模型可以得出结论，该模型给出了多个因素的相关性，为治疗过程进行科学的解释，并且使用该模型的**病人可以根据不同疗法、不同经济水平的组合来给出在0~50周的范围内的期望治疗走势和最佳疗法选择建议**。

**一、问题重述**

艾滋病是当前人类社会最严重的瘟疫之一，是由艾滋病病毒(英文简称HIV)引起的，这种病毒破坏人的免疫系统，使人体丧失抵抗各种疾病的能力，从而严重危害人的生命。人类免疫系统的CD4细胞在抵御HIV的入侵中起着重要作用，当CD4被HIV感染而裂解时，其数量会急剧减少，HIV将迅速增加，导致AIDS发作。

艾滋病治疗的目的，是尽量减少人体内HIV的数量，同时产生更多的CD4，至少要有效地降低CD4减少的速度，以提高人体免疫能力。现在得到了美国艾滋病医疗试验机构ACTG公布的同时服用3种药物疗法和另外一组4种疗法的相关数据。

**要求：**

1）预测继续治疗的效果，或者确定最佳治疗终止时间。

2）评价4种疗法的优劣（仅以CD4为标准），并对较优的疗法预测继续治疗的效果，或者确定最佳治疗终止时间。

3）如果病人需要考虑4种疗法的费用，找出对2）中的评价和预测（或者提前终止）有什么改变。

**二、模型假设与约定**

1）假设影响病人治疗的只有费用和疗效两个因素，不考虑其它因素；

2）假设采集的数据科学合理，有代表性，不受地域等因素影响或者说这种影响所造成的误差可以忽略不计；

3）假设数据处理时排除掉极个别不完全数据，不影响用于计算的数据的合理性；

4）附件二中，同一年龄段病人的身体素质视为在同一水平线。

**三、符号说明及名词定义**

**四、问题分析**

**五、问题一模型建立与求解**

**5.1 ACTG320数据的宏观分析**

分析附件一可以发现，病人接受同一疗法，可以排除不同疗法的差异。数据有CD4和HIV测试时间和对应的数量多少，由于要尽量减少人体内HIV的数量，同时产生更多的CD4，因此我们分析CD4与HIV的相关性以及对CD4和HIV进行插值-拟合中多项式逼近拟合，进而预测继续治疗的效果，或者确定最佳治疗终止时间。

**5.2 模型建立**

**5.2.1 相关系数模型**

相关系数（Correlation Coefficient）：度量变量间相关关系的一类指标的统称。就参数统计而言，常用的是皮尔逊相关系数（Pearson）：即协方差与两变量标准差之间的比值，是没有量纲的，标准化的协方差。

协方差（Covariance）：两个变量与其均值离差乘积的平均数，是相互关系的一种度量。

总体协方差：

样本协方差：

根据样本相关系数的计算公式有：



**5.2.2 模型求解**

本文通过MATLAB软件，对数据进行处理得到CD4与HIV的相关性，具体相关性如下表：

表5-1：CD4与HIV的相关性

|  |  |  |  |
| --- | --- | --- | --- |
| 相关性 | 属性分析 | CD4 | HIV |
| CD4 | Pearson相关性  显著性（双侧） | 1  无 | -0.4019  0.0000 |
| HIV | Pearson相关性  显著性（双侧） | -0.4019  0.0000 | 1  无 |

结果发现，CD4与HIV的P值为0.0000<0.0500，说明在99%的置信水平拒绝原假设，CD4与HIV具有很强的相关性。而CD4与HIV的相关性系数为-0.4019，说明CD4与HIV呈中等程度的负相关，随着CD4浓度的增加HIV的浓度会减少。

**5.3 运用插值-拟合求解最佳治疗终止时间**

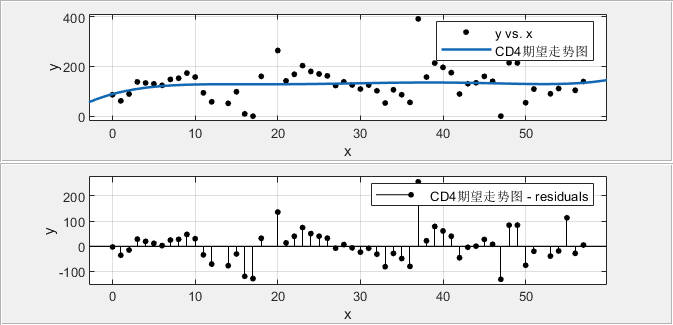
由于附件一的病人采用的疗法是一致的，所以病人对于疗法的医疗效果在总体上是大致相同的。因此我们不采用病人ID为自变量，而以周数week为自变量研究CD4浓度与HIV浓度随着周数的增加而变化的情况，并且从总体与个体两方面研究治疗效果。

**5.3.1 总体治疗效果**

由于附件一给的数据点太少以及测试时间的周期是不均匀的，即无法得到完整的时间序列，因此这里采用微分方程的拟合思想，选取插值-拟合中的多项式逼近拟合（Polynomial）方法来预测未来的治疗效果。所以为了进一步研究CD4随着周数的增加其浓度的期望走势并且经过筛选发现，多项式逼近拟合中选取五次方多项式函数进行拟合最为符合要求，得到拟合函数：



其中，，，，，，，以及RMSE（均方根误差）值为64.6354，但较比其他拟合函数的均方根误差是最低的。具体期望走势图如下图：



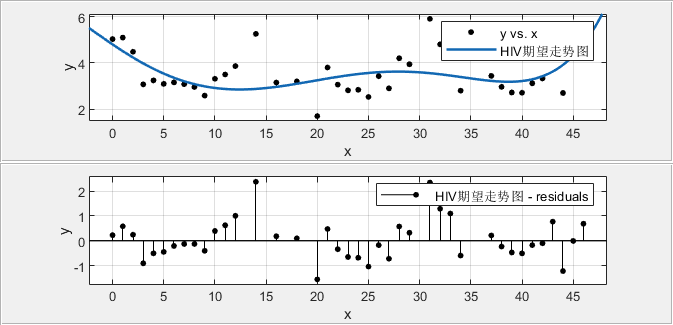
图：CD4浓度期望走势图

从图中可以发现，CD4浓度在0~10周是具有较大的上升趋势，然后在第10周之后大致平稳保持在130的水平线上。说明了该疗法在前10周具有一定的医疗效果。

为了进一步研究HIV随着周数的增加其浓度的期望走势，这里继续选取插值-拟合中的 Polynomial 方法来预测。结果发现选取五次方多项式函数进行拟合最为符合要求，得到拟合函数：



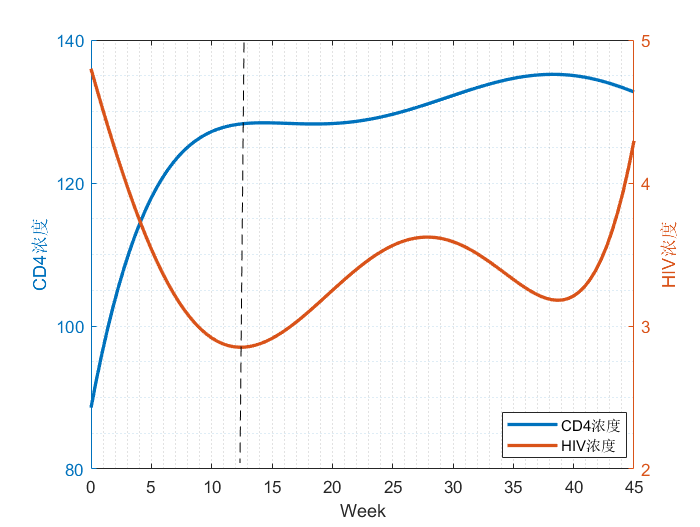
其中，，，，，，，以及RMSE（均方根误差）值为0.8655，但较比其他拟合函数的均方根误差是最低的。具体期望走势图如下图：



图：HIV浓度期望走势图

从图中可以发现，HIV浓度在0~12周时具有较大的下降趋势，在第12周时出现了较小的第一次反弹一直持续到第28周。在28~38周时具有较小的下缓态势，而在第38周时出现了较大的第二次反弹。

综上所述，结合CD4浓度期望走势和HIV浓度期望走势，预测继续治疗的效果，具体预测结果如下图：

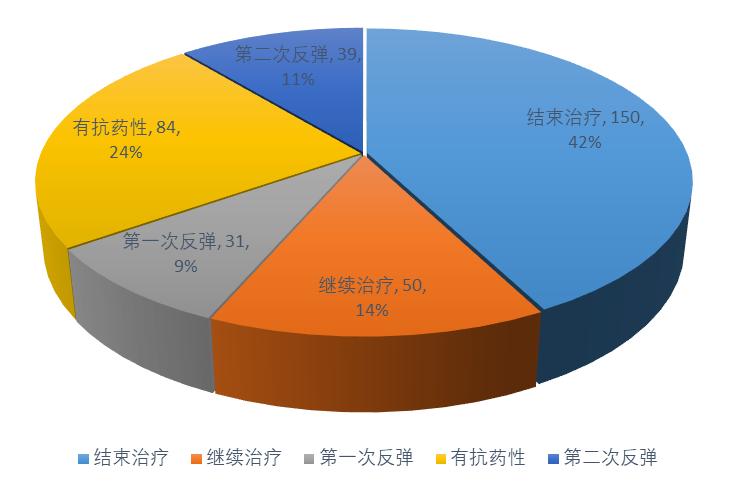


图：CD4浓度与HIV浓度的期望走势

从图中可以发现，在0~12周时CD4浓度急剧上升而HIV浓度急剧下降，在第12周之后CD4浓度大致平稳保持在130的水平线上而HIV浓度出现了两次反弹。因此在总体层面来看建议在第12周时选择提前终止治疗。

**5.3.2 个体治疗效果**

具体到每个病人的治疗效果，分析提出了病人的5种治疗状况以及对应的病人数量，如下图所示：



图：5种治疗状况以及对应数量

从图中可以发现，结束治疗的病人有150位，占比最大高达42%而占比最小低至9%是出现了第一次反弹的病人，有31位。继续治疗的病人有50位，占比14%；出现第二次反弹的病人有39位，占比11%；有抗药性的病人有84位，占比24%。

由于在CD4浓度与HIV浓度的期望走势图中HIV浓度出现了两次反弹，在28~38周时HIV浓度具有下降的趋势而在38周后HIV浓度急剧上升，因此第一次反弹是良性反弹而第二次是恶性反弹。所以出现了第二次反弹、也有抗药性和结束治疗状况的病人是不能继续治疗需要提前终止治疗，共有273位，占比77%，而出现了第一次反弹和继续治疗状况的病人是可以在测试终止后继续服药，共有81位，占比23%。

**六、问题二模型建立与求解**

**6.1 193A数据的宏观分析**

分析附件二可以发现，由于仅以CD4为标准对4种疗法的优劣进行评价，但存在病人年龄这一指标，所以这里先分析在四种疗法下年龄与CD4增长率的相关性。为了进一步评价4种疗法的优劣，本文以CD4的增长率作为评判标准并且从个体与总体两方面研究不同的疗法对CD4的增长率的影响。

**6.2 模型求解**

根据第一问的相关系数模型，得出四种疗法下年龄与CD4增长率的相关性，具体相关性如下表：

表：四种疗法下年龄与CD4增长率的相关性

|  |  |  |
| --- | --- | --- |
| 疗法类型 | 年龄与CD4增长率相关系数 | 相关系数对应的P值 |
| 疗法一 | 0.0169 | 0.7760 |
| 疗法二 | -0.0331 | 0.5775 |
| 疗法三 | 0.0050 | 0.9329 |
| 疗法四 | -0.0311 | 0.5886 |

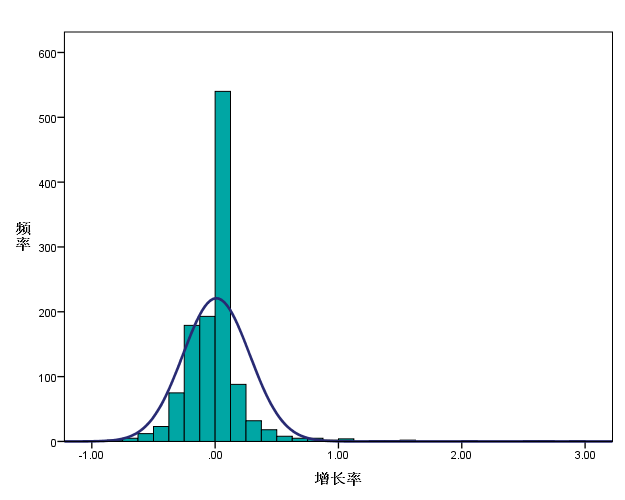
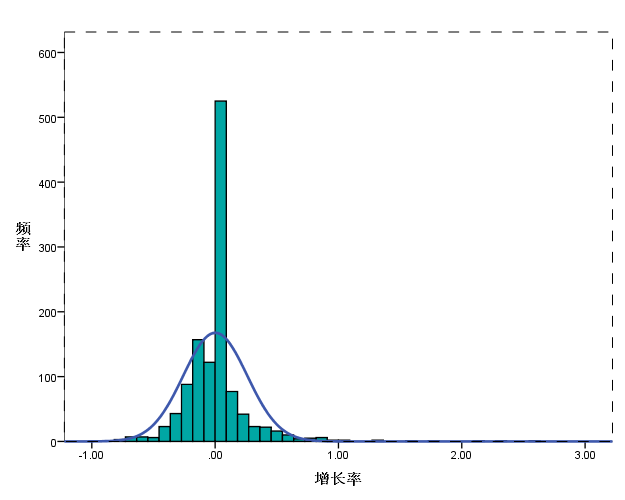
从表中可以发现，每种疗法下的年龄与CD4增长率相关系数在0.0~0.2中间，可以认为是极弱相关甚至是无相关，因此在处理CD4增长率可以剔除年龄的影响。

**6.3 数据分析处理**

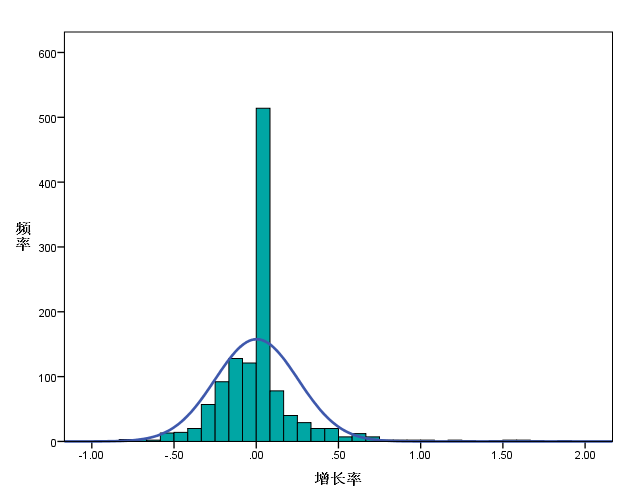
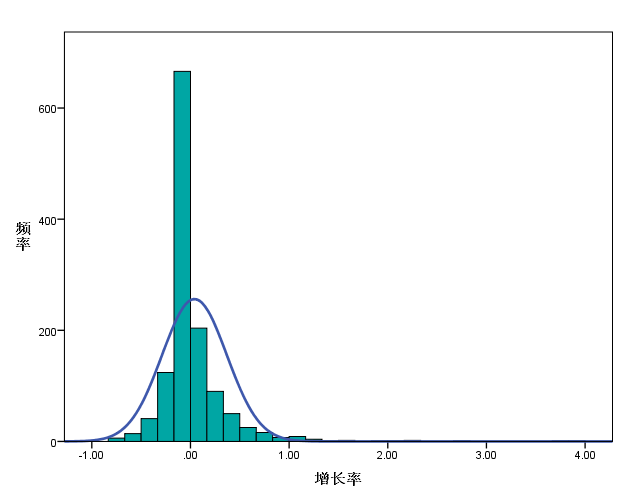
题目要求仅以CD4为标准，本文以CD4的增长率作为评判标准来评价4种疗法的优劣，并且从个体与总体两方面研究不同的疗法对CD4的增长率的影响。

**6.3.1 个体CD4增长率**

本文通过MATLAB软件，通过对数据进行分析处理得到各个疗法的CD4增长率分布律，具体结果如下图：



疗法一CD4增长率分布律 疗法二CD4增长率分布律

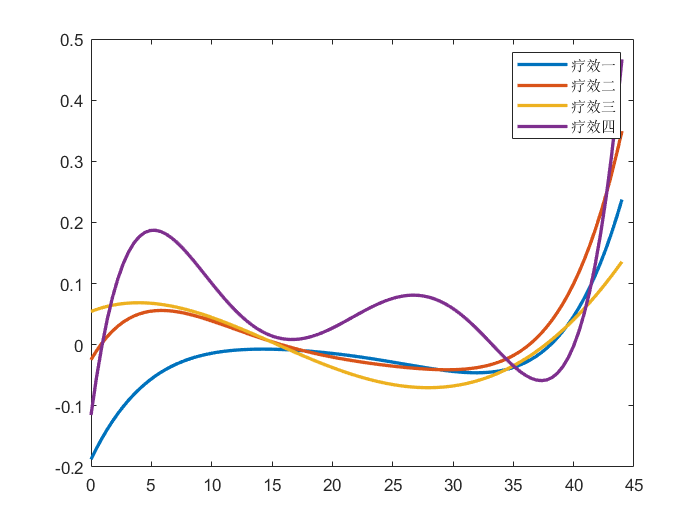


疗法三CD4增长率分布律 疗法四CD4增长率分布律

从四个图形综合分析来看，对CD4增长率的分布进行了一个正态分布拟合，结果发现只有疗法四正态曲线的高峰位于0.00偏右，说明疗法四的总体增长率较比其他三种疗法更高，效果更好。

**6.3.2 总体CD4增长率**

为了进一步理解不同疗法之间的差异性以及研究不同疗法对CD4的增长率的影响，总体的CD4增长率如下图：



图：总体CD4增长率

从图中可以发现，以周数week为横坐标，疗法一的CD4增长率大体处于负增长，仅在40周之后CD4增长率才大幅上升，因此疗法一适合长时间治疗服用，疗法耗费较高；疗法二和疗法三的增长率较为平缓但增长率较低，对病人的负荷较小，在15周之前疗法三略优于疗法二，而在15周之后均处于负增长且疗法二略优于疗法二，因此疗法二和疗法三适合在15周确定最佳治疗终止时间，但CD4增长率不高，治疗效果不佳；疗法四的总体CD4增长率明显比其他疗法高，在0~5周时CD4增长率急剧上升，而在第5周和第27周出现了两次反弹且在第32周时出现负增长，但CD4增长率总体处于正增长，治疗效果较佳。

综合所述，疗法四的治疗效果较优，而疗法二和疗法三的治疗效果其次且大致相同，疗法一的治疗效果较劣。由于在第32周时CD4增长率出现了负增长，因此建议疗法四在第32周时选择最佳治疗终止时间。

**七、问题三模型建立与求解**

**7.1 模型准备**

由于附件二病人的CD4抽取时间大体固定以八周为周期，是一个固定的时间序列并且实验数据量较少，因此采用灰色预测模型对较优的疗法进行预测继续治疗的效果。

**7.2 灰色预测模型**

**7.2.1 模型分析**

通过对有限实验数据的分析，并运用灰色理论建立预测模型，目的是在分析这些数据的基础上，实现对未知数据的预测。

**7.2.2 模型建立**

记，其中表示第次测得样品的数据。令为建模序列，表示灰导数，其中，k=1,2,3，令为的AGO序列。







令为的均值（MEAN）序列，表示白化背景值：





则得到的灰微分方程模型为：



其中：





经变化得到：



**六、模型评价与推广**

**七、参考文献**

**八、附录**