

# 艾滋病疗法的评价及疗效的预测模型

## 摘 要

本文建立了艾滋病疗法的评价及疗效的预测模型，从总体上就美国艾滋病医疗试验机构 ACTG 公布的两组数据，对不同的疗法作出了优劣的评判，运用最小二乘原理结合所给数据中的 CD4 和 HIV 值，建立数据模拟模型，对三种不同程度的病人继续治疗的效果作出了预测，或者确定最佳治疗终止时间。在兼顾治疗费用的条件下，对不同疗法给出了较好的评判。最后，对模型作出分析、评价和改进。

对于问题一我们建立了模型 数据拟合模型。先根据 355 位病人的 1665 组检测数据中 CD4 和 HIV 的浓度，预测是继续服用此药物对病人进行治疗还是确定最佳终止时间提前终止治疗，我们分析认为影响这个问题的最主要的因素是病人的初始病情，为了衡量病人的疗效，根据艾滋病病理，由 CD4 浓度与 HIV 浓度的比值定义了健康指标。利用最小二乘法原理，根据不同病期平均健康指标的拟合函数确定不同病期的病人的最佳治疗终止时间或者是预测继续治疗的效果。按此模型得出结果：早期爱滋病患者继续接受治疗，病人健康会好转但不会痊愈，中期和晚期患者最佳治疗终止时间分别在第 32.8 和第 46.2081 周。

对于问题二，在评价 4 种疗法的优劣时建立了模型 统计概率模型，以 CD4 浓度的变化为标准，增加则表示疗法有效，否则疗法无效，比较 4 种疗法的有效治疗率大小，得知从优到劣的次序依次为：疗法 4、疗法 2、疗法 3、疗法 1。因为附件 2 给出了病人的年龄，对于最优疗法 4，改进了模型 我们建立了模型 数据拟合模型。增加年龄段因素，分成不同病期内不同年龄段的六类病人，仍按最小二乘原理确定各类病人的最佳治疗终止时间或者是预测继续治疗的效果。得出结果：对于最优疗法 4，青壮年早、中、晚期患者的最佳治疗终止时间分别在第 22.6592，15.7154，23.622 周；中老年早、中、晚期患者的最佳治疗终止时间分别在第 10.0294，4.3657，10.063 周。

对于问题三，综合考虑疗效、费用，我们建立模型 统计实验模型，我们固定各疗法的有效人数，来算出各疗法的人数所花费的费用，费用最小的疗法就是效果最好的疗法，我们得到在综合考虑情况下疗法 1 效果最好。疗法 1 与问题二中的疗法 4 有明显的不同，最佳停药时间明显比疗法 4 出现的早。

**关键词：**艾滋病 CD4 HIV 最小二乘法 数据拟合 Matlab

## § 1 问题的重述

### 一、背景知识

#### 1、艾滋病概况

艾滋病是当前人类社会最严重的瘟疫之一，从 1981 年发现以来的 20 多年间，它已经吞噬了近 3000 万人的生命。

#### 2、艾滋病病理

艾滋病的医学全名为“获得性免疫缺损综合症”，英文简称 AIDS，它是由艾滋病毒（医学全名为“人体免疫缺损病毒”，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。

#### 3、艾滋病治疗

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，至少要有有效地降低 CD4 减少的速度，以提高人体免疫能力。

迄今为止人类还没有找到能根治 AIDS 的疗法，目前的一些 AIDS 疗法不仅对人体有副作用，而且成本也很高。许多国家和医疗组织都在积极试验、寻找更好的 AIDS 疗法。

### 二、具体试验数据

现在得到了美国艾滋病医疗试验机构 ACTG 公布的两组数据。

#### 1、ACTG320 数据（见附件 1）

ACTG320 是同时服用 zidovudine（齐多夫定），lamivudine（拉美夫定）和 indinavir（茚地那韦）3 种药物的 300 多名病人每隔几周测试的 CD4 和 HIV 的浓度（每毫升血液里的数量）。

#### 2、193A 数据（见附件 2）

193A 是将 1300 多名病人随机地分为 4 组，每组按下述 4 种疗法中的一种服药，大约每隔 8 周测试的 CD4 浓度（这组数据缺 HIV 浓度，它的测试成本很高）。4 种疗法的日用药分别为：600mg zidovudine 或 400mg didanosine（去羟基苷），这两种药按月轮换使用；600 mg zidovudine 加 2.25 mg zalcitabine（扎西他滨）；600 mg zidovudine 加 400 mg didanosine；600 mg zidovudine 加 400 mg didanosine，再加 400 mg nevirapine（奈韦拉平）。

### 三、要解决的问题

1、**问题一**：利用附件 1 的数据，预测继续治疗的效果，或者确定最佳治疗终止时间（继续治疗指在测试终止后继续服药，如果认为继续服药效果不好，则可选择提前终止治疗）。

2、**问题二**：利用附件 2 的数据，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、**问题三**：艾滋病药品的主要供给商对不发达国家提供的药品价格如下：600mg zidovudine 1.60 美元，400mg didanosine 0.85 美元，2.25 mg zalcitabine 1.85 美元，400 mg nevirapine 1.20 美元。如果病人需要考虑 4 种疗法的费用，对问题二中的评价和预测（或者提前终止）有什么改变。

## § 2 问题的分析

艾滋病疗法的评价及疗效问题是一类大样本多数据的统计分析与预测类问题。对本问题本处理要分两个步骤进行：第一，对艾滋病的治疗方法作出合理的评价；第二，对治疗的效果给出准确的预测。由于病情治疗本问题是

要合理准确处理好本问题，关键必须弄清问题的相关知识并对问题作出深入的分析。

### 一、相关知识的介绍

艾滋病是当前人类社会最严重的瘟疫之一，它是由艾滋病毒 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。

因此，首先我们要明确点，即 CD4 数量越多或 HIV 浓度越少，人类免疫系统就越强，反之，则越弱。为此我们必须弄清以下两个关系：

#### 1、CD4 数量与病症

CD4 数量多少为好，为此我们查阅了有关资料<sup>[1]</sup>，结果表明，无症状HIV感染期：CD4 绝对数 $\geq 500$  个/ $\text{mm}^3$ ；有症状感染期：CD4 绝对数 200-499/ $\text{mm}^3$ ；AIDS 期：CD4 绝对数 $< 200/\text{mm}^3$ 。

#### 2、HIV 浓度与治疗

HIV 是一种逆转录病毒，它主要存在于感染者和病人的体液及多种器官中，它可通过含 HIV 的体液交换或器官移植而传播。HIV 的浓度指每毫升血液里的数量。显然，健康人的身体内没有 HIV，对于感染者和病人，HIV 的浓度越低说明病症越轻微，浓度越高则病症越严重。

### 二、对问题的具体分析

#### 1、对问题一的分析：

问题要求利用附件 1 的数据，预测继续治疗的效果，或者确定最佳治疗终止时间。数据 ACTG320 是使用同时服用 3 种药物的治疗方法得到的 300 多名病人每隔几周测试的 CD4 和 HIV 的浓度。把附件 1 的扩展名 txt 改成 xlt（即电子表），运用 excel<sup>[2]</sup> 可统计出结果为 355 位病人的 1665 组检测数据，其中病人、时间、CD4 数量和 HIV 的浓度是要考虑四个量。以下是处理数据的原则。

#### 剔除缺少 HIV 的数据

运用 excel 统计，有 100 组数据缺少 HIV 数值，仅占总量的 6%，为了便于研究，首先把这些组数据剔除掉。

#### 利用差值法把数据组时间统一到第 0、4、8、24、40 周

将 1665 组检测值对时间作散点图（见图 1），从图 1 可以发现：图形大致呈折线形变化，且在第 0、4、8、24、40 周附近呈水平，为了简化对问题的处理，可在总体上只考虑这五周 CD4 的大小和 HIV 的数值（以下简称为 CD4 和 HIV）来进行总体拟合，对恰巧不全是这几周检测的，我们可按下面四条原则行处理：

**按相邻原则由左右两测点推测中间点。**我们假设病人的病情在两个测试中间段的变化是平稳的，具体办法：如已知第 0、3、8、23、40 周，可由第 3 周与第 8 周的 CD4 和 HIV 分别按直线求出第 4 周的 CD4 和 HIV，同理可由第 23 周与第 40 周的 CD4 和 HIV 分别按直线求出第 24 周的 CD4 和 HIV。

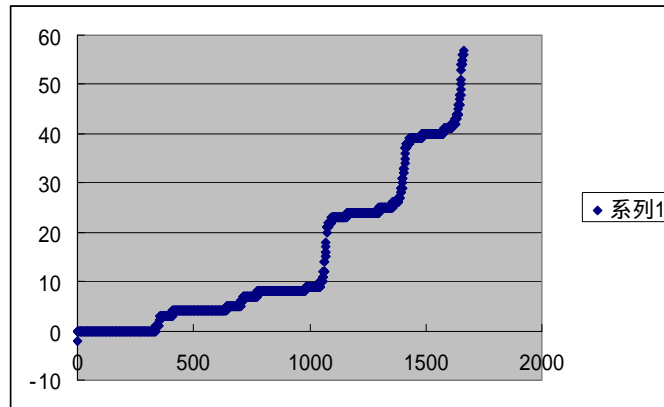


图 1 1665 组检测值点对时间散点图

**按就近原则由已知最近两点推测端点。**具体办法：如已知第 1、4、8、24、38 周，可由第 1 周与第 4 周的 CD4 和 HIV 分别按直线求出第 0 周的 CD4 和 HIV，同理可由第 24 周与第 38 周的 CD4 和 HIV 分别按直线求出第 40 周的 CD4 和 HIV。

**按多余点舍弃原则。**具体办法：如已知第 0、4、8、24、41、50 周，可由第 24 周与第 41 周的 CD4 和 HIV 分别按直线求出第 40 周的 CD4 和 HIV，对多余的第 50 周可舍弃。

**按缺少点不补原则。**具体办法：如已知第 0、4、8、24 周，对缺少第 40 周 CD4 和 HIV，不再补充，在求总体均值时要注意到总量的多少。

**第 0 周情况可作为病情分类的依据。**由于第 0 周是病人治疗前的检测结果，这既可作为病情分类的依据，也可作为治疗结果的参照。当然，在对总体拟合是也是不可缺少的点。

## 2、对问题二的分析：

问题要求利用附件 2 的数据，仅以 CD4 为标准来评价 4 种疗法的优劣，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。附件 2 涉及到五个量：病人、疗法、年龄、时间、 $\text{Log}(\text{CD4 数}+1)$ 。

评价 4 种疗法的优劣时，首先按不同的治疗法将数据分成四组，然后对每组的 CD4 分别按问题一的方法进行处理，统计出各组 CD4 值增加的个数，据此运用统计概率模型可以求出四种治疗法的有效治疗率，由有效治疗率大小即可以判别四种疗法的优劣。对较优的疗法，然后仿照问题一依据病人治疗前的 CD4 值将病人按病情分成早期、中期和晚期三类，再依据年龄将病人分成青壮年和中老年，这样交叉可得六类病人，最后仿照问题一求出各类数据拟合函数曲线图，据此可以预测继续治疗的效果，及确定最佳治疗终止时间。

## 3、对问题三的分析：

近年来，由于不发达国家因为种种原因，感染艾滋病人数很多，而经济又相对落后，治疗费用是必须考虑的一大重要因素，为此，对问题二的模型的要进行改进，要把几种药品价格加入模型进行分析与处理。对个体而言，由于不同的疗法的费用是固定的，我们要从总体上进行考虑，在疗效相同的条件下，对每种疗法进行实验设计，然后可求得每种费用在相同时间内的花费多少，从而可以判四种疗法的优劣。对最佳停药时间，应该为有费用限制出现在无费用限制之后。

### § 3 模型的假设

- 1、假设病人的病情在两个测试周中间时段的变化是平稳的；
- 2、为了方便起见，规定早期、中期、晚期编排序号为 1、2、3；
- 3、规定第 0、4、8、24、40 周编排序号为第 0、1、2、3、4 次；
- 4、每月按 28 天计算即每个月 4 周；
- 5、对仅有初始检测值的视为未进行治疗，在研究时可剔除；
- 6、所有数据均为原始数据，来源真实可靠。

### § 4 名词解释与符号说明

#### 一、名词解释

- 1、AIDS：艾滋病的英文简称，其医学全名为“获得性免疫缺损综合症”；
- 2、HIV：艾滋病毒的英文简称，其医学全名为“人体免疫缺损病毒”；
- 3、CD4：是一组产生免疫功能的淋巴细胞的辅助细胞；
- 4、健康指标：各时刻测试出的 CD4 浓度与 HIV 浓度的比值；
- 5、病情好转速度：单位时间（周）内一个病人健康指标的变化量；
- 6、可决系数：是指 1 减误差平方和与残差平方和的差，用来检验曲线拟合程度的量。

#### 二、符号说明

序号	符号	符号说明
1	$k$	表示病人的综合健康指标
2	$c$	表示病人综合检测的 CD4 浓度
3	$h$	表示病人综合检测的 HIV 浓度
4	$k_{ij}$	表示第 $i$ 个病人在第 $j$ 次检测时的健康指标
5	$t_{ij}$	表示第 $i$ 个病人进行第 $j$ 次检测的周数
6	$v_{ij}$	表示第 $i$ 个病人在第 $j$ 次检测时的病情好转速度
7	$k'_{ij}$	表示 $t$ 期第 $i$ 个病人在第 $j$ 次的健康指标
8	$c_i$	表示用第 $i$ 种疗法第一次测试病人的 CD4 浓度
9	$c'_i$	表示用第 $i$ 种疗法最后一次测试病人的 CD4 浓度
10	$n_i$	表示第 $i$ 种疗法有效个数
11	$N_i$	表示第 $i$ 种疗法总个数
12	$s_i$	表示第 $i$ 种疗法为一个人在这段时间内的全花费

## § 5 模型的建立与求解

从所要解决的问题和对问题所做的假设出发，我们对问题一建立了模型，对问题二建立了模型和模型，对问题三建立了模型。

### 1、模型 数据拟合模型

本模型从 CD4 数量越多则病人的免疫力越强，HIV 的数值越小则病人的病情越轻出发，建立了健康指标模型，给出每个病人每个时段的健康指标，依据病人治疗前的健康指标将病人按病情分成早期、中期和晚期三类，然后利用最小二乘法原理分别求出各类数据拟合函数曲线图，据此可以预测继续治疗的效果，及确定最佳治疗终止时间。

### 2、模型 统计概率模型模型

本模型首先对四类治疗法分成组，然后对每组的 CD4 分别按问题一的方法进行处理，统计出各组 CD4 值增加的个数，据此运用统计概率模型可以求出四种治疗法的有效治疗率，由此可以判别四种疗法优劣。

### 3、模型 数据拟合模型

本模型针对模型 求出的较优的疗法，然后仿照模型 依据病人治疗前的 CD4 值将病人按病情分成早期、中期和晚期三类，再依据年龄将病人分成青壮年和中老年，这样交叉可得六类病人，最后仿照模型 求出各类数据拟合函数曲线图，据此可以预测继续治疗的效果，及确定最佳治疗终止时间。

### 4、模型 统计实验模型

对个体而言，由于不同的疗法的费用是固定的，我们要从总体上进行考虑，为此，本模型是从总体上在疗效相同的条件下，对每种疗法设计进行实验设计，然后可求得每种费用在相同时间内的花费多少，从而可以判四种疗法的优劣。对最佳停药时间，应该为有费用限制出现在无费用限制之后。

## 一、问题一的分析与求解

### 1、对问题的分析

由于利用附件 1 的数据，预测继续治疗的效果，或者确定最佳治疗终止时间（继续治疗指在测试终止后继续服药，如果认为继续服药效果不好，则可选择提前终止治疗）。

#### 相关量的定义

##### 健康指标

为了更好地反映出同时服用 *zidovudine*、*lamivudine* 和 *indinavir* 这三种药物的治疗效果，由问题的分析我们知道，CD4 数量越多，病人的免疫力越强，HIV 的数值越小则病人的病情越轻，为此兼顾这两个方面，下面我们定义一个新的指标函数。

**定义 1** 在同一时刻，病人血液中的 CD4 浓度与 HIV 浓度的比值反映一个人的健康状况，我们称为健康指标，记作  $k$ 。

由定义，我们可以给出第  $i$  个病人在第  $j$  次检测时的健康指标  $k_{ij}$  为

$$k_{ij} = \frac{c_{ij}}{h_{ij}}$$

其中  $c_{ij}$ 、 $h_{ij}$  分别表示第  $i$  个病人在第  $j$  次检测时的 CD4 与 HIV 浓度。因为在研究问题过程中单位对结果无影响，我们对健康指标的单位可以不予考虑。显然，一个病人的健康指标值越大表示这个人的病情越轻，反之病情越重。

##### 病情好转速度

为了确定最佳治疗终止时间，下面我们由健康指标与时间二者的相对改变量

给出以下定义。

**定义 2** 我们称单位时间（周）内一个病人健康指标的变化量为其病情好转速度，记作  $v$ 。

由定义，我们可以给出第  $i$  个病人在第  $j$  次检测时的病情好转速度  $v_{ij}$  为

$$v_{ij} = \frac{k_{ij+1} - k_{ij}}{t_{ij+1} - t_{ij}}$$

其中  $k_{ij}$  表示第  $i$  个病人在第  $j$  次检测时的健康指标， $t_{ij}$  表示第  $i$  个病人进行第  $j$  次检测的周数。

若  $v_{ij}$  为正值，表示艾滋病人服用这三种药物后病况好转，反之，若为负值，表示艾滋病人服用药物后病情趋于恶劣，即此时再服用这三种药物，对病人将产生负作用。综合考虑每个病人各时期的病情好转速度，我们判断每个病人是继续服药还是提前终止服药并确定最佳治疗终止时间。

## 2、模型 数据拟合模型

### 模型的准备

#### 对病期的划分

由问题的分析，可由初始检测结果对病情进行分类，CD4 绝对数  $\geq 500$  个/  $\mu\text{l}$  时为无症状 HIV 感染期，我们记为早期；CD4 绝对数为 200-499/ $\mu\text{l}$  时为有症状感染期，我们记为中期；CD4 绝对数  $< 200/\mu\text{l}$  时为 AIDS 期，我们记为晚期。考虑到附件 1 中 CD4 浓度已乘以 0.2 个/ $\mu\text{l}$ ，为保持一致，将上述 CD4 的数目乘以 0.2，可以得出与附件 1 中的数据单位一致的艾滋病病期划分标准（见表 1）。

表 1 艾滋病病期划分标准（一）

病期	晚期	中期	早期
修正 CD4 值范围	0~40	40~100	100~

#### 各病期人数的分布

利用 EXCEL 软件对数据进行自动筛选，得到最初开始测试 CD4 即还未开始服用这三种药物时 336 名患者在各病期的分布情况（见表 2）。

表 2 治疗前艾滋病人在各病期所占的比例

患病程度	频数 $m$	比例 $r$
晚期	113	33.63%
中期	94	27.98%
早期	129	38.39%

#### 统一求出第 0、4、8、24、40 周的 CD4 和 HIV 浓度

我们利用问题分析中的具体处理办法，利用差值法把时间统一到第 0、4、8、24、40 周，求出每个病人在第 0、4、8、24、40 周的测试的 CD4 和 HIV 浓度（处理结果见附录 3）。

#### 各期平均健康指标

要预测病人继续治疗的效果，或者确定最佳治疗终止时间，总体健康指标是解决问题的关键，在统计学上总体健康指标往往用样本的平均指标来估计，因此下面我们先求平均健康指标。

### [模型的建立]

结合表 2 中三个病期的病人人数，可建立以下平均健康指标模型

$$\overline{k_j^t} = \frac{\sum_{i=1}^{m_t} k_{ij}^t}{m_t}, \quad t=1,2,3 \quad j=0,1,2,3,4$$

其中  $t=1,2,3$  分别表示早、中、晚期， $k_{ij}^t$  表示  $t$  期第  $i$  个病人在第  $j$  次的健康指标。

### [模型的求解]

利用 EXCEL 软件从附录 3 的数据中分别筛选出  $m_t (t=1,2,3)$  个各病期病人的五个时段的健康指标值，然后对上述模型用 Matlab 软件编程求解（见附录 4）。

### [模型的结果]

求得的各个病期病人五个检测时间的平均健康指标结果见表 3。

表 3 各期各检测时间病人的平均健康指标

检测时间 病期	0	4	8	24	40
早 期	33.72679	77.15918	96.89483	119.2863	<b>126.7081</b>
中 期	14.18602	40.76365	55.20248	76.42377	<b>85.50784</b>
晚 期	<b>12.84206</b>	<b>18.5493</b>	<b>28.63948</b>	<b>41.67688</b>	<b>52.35392</b>

### 模型的建立与求解

#### 1) 数据的拟合原理：最小二乘法原理<sup>[3]</sup>

根据表 3 中已有的各病期的 5 个检测时刻的健康指标数据  $(s, \overline{k_j^t})$ ，为了方便起见，取为  $(x_j, y_i) (i=1,2,3; j=0,1,2,3,4)$ ，其中  $x_j$  依次为 0、4、8、24、40 五个点。在以平面直角坐标系中我们很容易作出各自的散点图，通过观察各散点图中散点的变化趋势，我们可构造含有一个或若干个未知参数  $\theta$  的拟合函数  $y_i(x_j, \theta)$ ，拟合函数  $y_i(x_j, \theta)$  可以为直线、抛物线、指数函数或其它函数，这取决于经验或软件。对拟合的优劣，可根据最小二乘法原则，拟合函数  $y_i(x_j, \theta)$  使得函数在点  $x_j (j=0,1,2,3,4)$  处的函数值与观测数据偏差的平方和达到最小，即求满足如下条件的函数  $y_i(x_j, \theta)$  使得

$$\min \sum_{j=1}^5 (y_i(x_j, \theta) - y_i)^2 = \sum_{i=1}^n (y_i(x_j, \hat{\theta}) - y_i)^2$$

#### 2) 对早中晚各期数据的拟合

##### 对早期数据的非线性拟合

对早期的五组数据  $(0, 33.72679)(4, 77.15918)(8, 96.89483)(24, 119.2863)(40, 126.7081)$ ，运用 Mathematica 软件<sup>[4]</sup>画出的散点图（见图 2）。

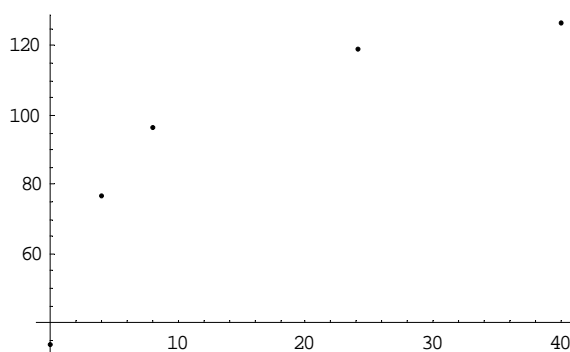


图 2 早期平均健康指标对五个时间散点图



从散点图可以看出，分布五个点呈指数形状，为此构造拟合函数

$$y_1 = ae^{\frac{c}{x+b}}$$

其中  $a$ 、 $b$ 、 $c$  为未知参数。

结合最小二乘法原理，利用MATLAB<sup>[5]</sup>数学应用软件计算出关于参数 $a$ 、 $b$ 和 $c$ 的最优估计，其中 $\hat{a}=138.72$ 、 $\hat{b}=2.78$ 、 $\hat{c}=-3.94$ （MATLAB程序见附录程序nihe.m）得到在早期病人的平均健康指标拟合曲线为：

$$y_1 = 138.72 e^{\frac{-3.94}{x+2.78}}$$

并可计算出可决系数（可决系数是判别曲线拟合效果的一个指标）：

$$R^2 = 1 - \frac{\sum_{i=1}^5 (y_i - \hat{y}_i)^2}{\sum_{i=1}^5 (y_i - \bar{y})^2} = 0.9998$$

因为可决系数接近于 1，认为得到的拟合曲线效果很好。

作出拟合曲线图（见图 3）。

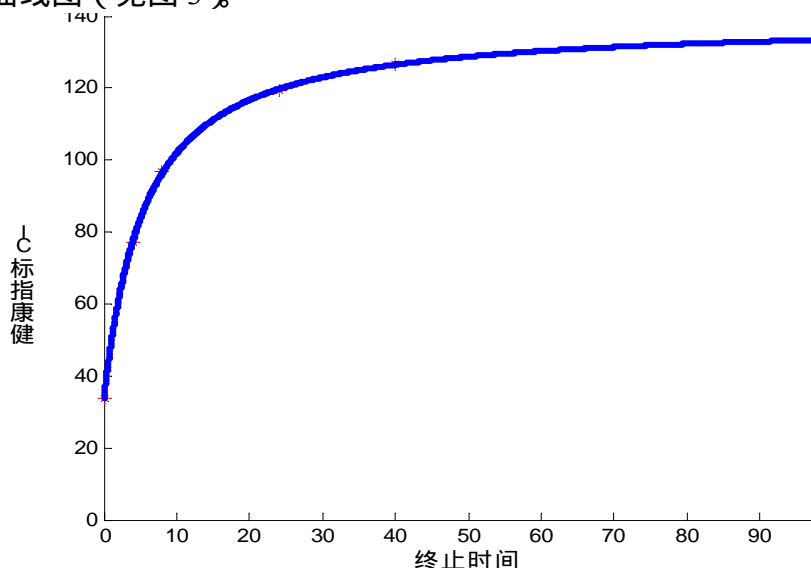


图 3 早期病人平均健康指标拟合图

**结论：**通过对早期艾滋病患者的平均健康指标拟合图进行分析，早期艾滋病患者应该继续同时服用 *zidovudine*, *lamivudine*, *indinavir* 这 3 种药物，接受治疗。从拟合图中可以直观地看出病人接受治疗的时间越长，病人的健康状况就会有所好转，也就是说同时服用这三种药物可以抑制艾滋病毒的增长，但是随着早期病人服药时间的延长，其好转速度会逐渐趋于平缓，由拟合函数可知早期患者的平均健康指标不会超过 138.72，然而我们可以利用这种药物疗法来控制早期艾滋病患者的病情恶化。

### 对中期数据的线性拟合

对中期的五组数据 (0,14.18602)( 4,40.76365)(8,55.20248)(24,76.42377), (40,85.50784), 运用 Mathematica 软件画出的散点图（见图 4）。

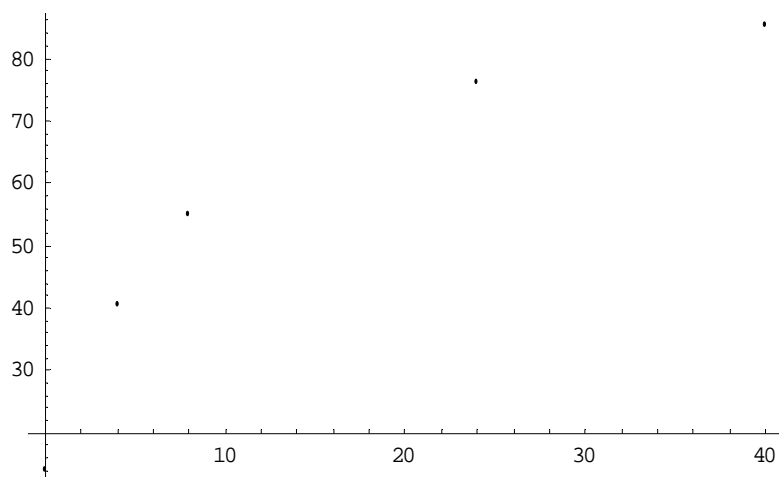


图 4 中期平均健康指标对五个时间散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_2 = -0.0614x^2 + 4.0272x + 21.9211$$

做出中期拟合曲线图见图 5，并可得到拟合曲线误差平方和为 0.957：

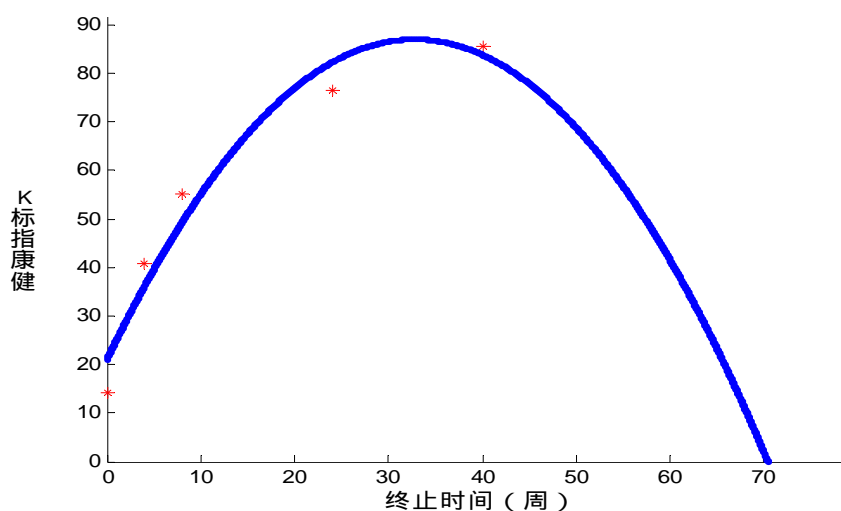


图 5 中期病人平均健康指标拟合图

因为中期艾滋病患者的健康指标拟合曲线为二次曲线，我们可以认为中期艾滋病患者在此二从此曲线的拐点处终止治疗，因为在此点之前病人的健康指标呈上升趋势，病人的健康在好转，而之后病人的健康状况又开始恶化，造成这种情况发生是因为中期患者体内 CD4 免疫细胞数目在 200 至 500 个/vl 之间(不包括 500)，随着病人服药时间的增长，病人体内开始对此药物产生抗体，而且还可能是因为当这 3 种药物同时服用量达到一定程度，会对身体内其他有益细胞（如 CD4 细胞）造成损伤，就会导致病人病情又开始恶化，因此在该点处对患者终止治疗为最佳时间。

求出拟合函数的导数为  $y_2' = -0.1226x + 4.0184$

令  $y_2' = 0$ ，即可求得在  $x=32.8$  时取得拐点；

所以中期艾滋病患者应该在第 32.8 周时终止治疗，此时经过异端时间的药物治疗，病人达到最好健康状态  $k=86.2545$ ，我们可以求出患者在终止治疗前的健康平均好转速度为 1.98466/周。

### 对晚期数据的非线性拟合

对中期的五组数据  $(0, 12.84206)$   $(4, 18.5493)$   $(8, 28.63948)$   $(20, 41.67688)$   $(40, 52.35392)$ ，运用 Mathematica 软件画出的散点图（见图 6）。

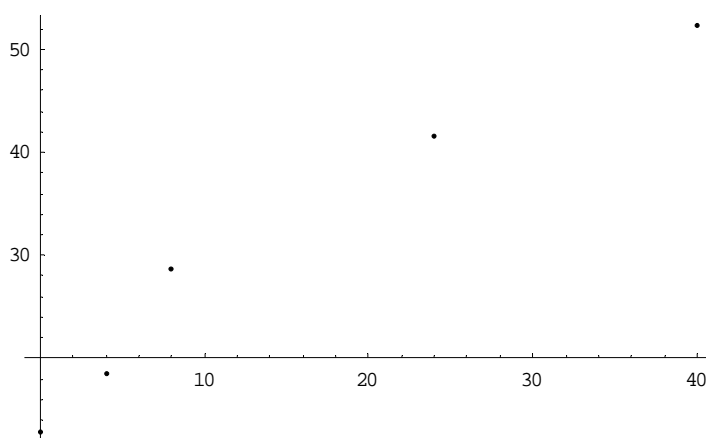


图 6 晚期平均健康指标对五个时间散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_3 = -0.0185x^2 + 1.7097x + 13$$

作出晚期拟合曲线图见图 7，并可得到拟合曲线误差平方和为 0.9867，拟合效果较好。

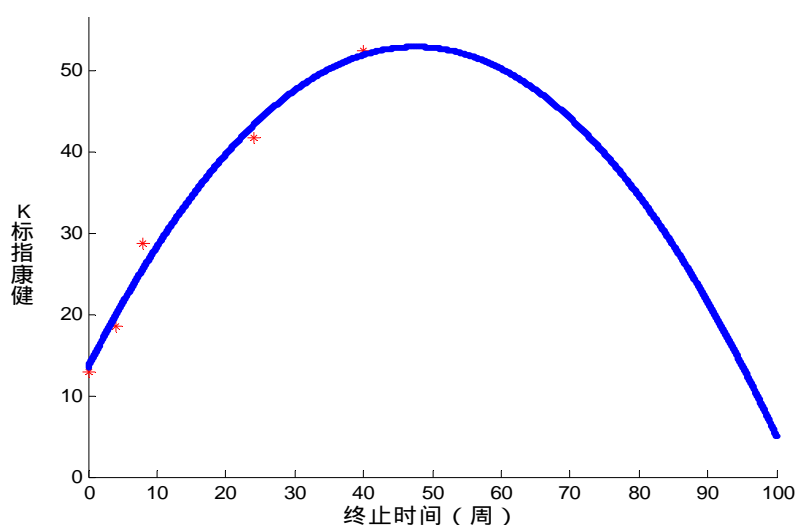


图 7 晚期病人平均健康指标拟合图

同理，我们求出晚期艾滋病患者拟合函数的导数为：

$$y_3' = -0.037x + 1.7097$$

令  $y_3' = 0$ ，即可求得在  $x=46.2081$  时取得拐点；

**结论：**我们认为晚期艾滋病患者应该在第 46.2081 周时终止治疗。

分析原因，对于晚期艾滋病患者，身体健康好转的空间已不是大，由拐点处的最大 CI 值为 52.33816，晚期艾滋病患者在终止治疗前的健康平均好转速度为 0.851326/周，远低于早期和中期艾滋病患者在终止治疗前的健康平均好转速度，恢复健康是不可能的了，此时仍对他们进行药物治疗，尽管药效甚微，但是毕竟可以在一定时间内抑制病毒的快速繁殖，可以尽可能地延长他们的生命，当他们的生命靠药物已经无法再延长时即

终止对他们的治疗。

### 3) 早中晚各期最佳治疗终止时间的比较

将早、中、晚期分别对应的最佳治疗终止时间列表如下：

表 4 各期最佳治疗终止时间

病期	早期	中期	晚期
最佳治疗终止时间	继续治疗某一时间*	32.8	46.2081

\*注：早期艾滋病患者服药后会维持较好的药效，在无特殊情况发生时最佳治疗终止时间即是该病人的自然死亡时间。

通过对比，这种疗法对早期艾滋病患者治疗效果最好，晚期次之，中期最差。

## 二、问题二的分析与求解

### 1、对问题的分析

问题要求利用附件 2 的数据，仅以 CD4 为标准来评价 4 种疗法的优劣，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。附件 2 涉及到五个量：病人、疗法、年龄、时间、Log(CD4 数+1)。

评价 4 种疗法的优劣时，首先我们按疗法把病人分开成四组，然后对每组的 CD4 分别按问题一的方法进行处理，再对各种疗法进行排序；其次我们对每组再按年龄分成若干组进行讨论，再细考虑各种疗法对不同年龄组的疗效的好坏，以期得到较优的结果。

### 2、模型 统计概率模型模型

#### 模型的分析

根据对文献资料的查阅，要判断一种疗法治疗效果的好坏，主要是通过以下三个方面进行评估：病毒学指标（即 HIV 浓度指标）、免疫学指标（即 CD4 浓度指标）和临床症状，其中病毒学的改变是最重要的指标。但是因为 HIV 浓度测试的成本很高，而且在本题中未给出临床症状的指标，所以在模型中我们仅从免疫学指标即 CD4 浓度方面来评价这四种疗法的优劣。

#### 模型的准备

首先我们利用 EXCEL 统计软件从附件 2 给出的数据中分别将按照疗法 1、疗法 2、疗法 3、和疗法 4 的病入的 CD4 与对应的测试时刻时间整理出来（各疗法对应的数据见附录 EXCEL 表格“问题 2”），为了评价 4 种疗法的优劣，我们以各疗法的治疗效果为标准，并从每个病人最后一次测试时的 CD4 浓度与第一次测试时的 CD4 浓度的变化考虑治疗效果，用  $\Delta c_i$  表示第  $i$  种疗法的治疗效果：

$$\Delta c_i = c'_i - c_i$$

其中  $c_i$  与  $c'_i$  分别表示用第  $i$  种疗法第一次和最后一次测试病人的 CD4 浓度。

若  $\Delta c_i > 0$ ，表示用第  $i$  种疗法治疗对病人有效，而且  $\Delta c_i$  值越大疗效越好；

若  $\Delta c_i < 0$ ，则表示此时用第  $i$  种疗法治疗艾滋病将对病人产生负作用。

#### 模型的建立

运用电子表 EXCEL，我们可以准确地统计出每种疗法中治疗有效与无效的病例数，具体结果见表 5。

表 5 四种疗法有效无效治疗个数分布表

治疗方案	疗法 1	疗法 2	疗法 3	疗法 4
有效个数	88	124	123	331
无效个数	197	168	174	282
总个数	285	292	297	613

令  $n_i$ 、 $N_i$  分别表示第  $i$  种疗法有效个数和总个数，则对各种疗法的有效治疗率也即总的疗效可建立如下统计概率模型：

$$p_i = \frac{n_i}{N_i}, \quad i = 1, 2, 3, 4$$

### 模型的结果

利用表 5 中的数据和概率模型，运用 MATHEMATICA 软件很容易求各种疗法对应的有效治疗率（见表 6）。

表 6 各疗法对应的有效治疗率

治疗方案	疗法 1	疗法 2	疗法 3	疗法 4
有效治疗率 $p_i$	0.31	0.42	0.41	0.54

结论：

由有效治疗率的大小可知， $p_4=0.54$  最大，也即疗法 4 的治疗方案最好。

## 3、模型 数据拟合模型

### 模型的准备

#### 对病期的划分

问题一中病期的划分为：早期 CD4 值范围为 100 以上，中期 CD4 值范围为 40~100，晚期 CD4 值范围为 0~40。由于附录 2 中的 CD4 的浓度是经过  $\text{Log}(\text{CD4}+1)$  处理所得的值，通过观察可以发现其范围在 0~6.2971。为了便于对病期的划分，我们要对问题一中病期的划分的 CD4 作相应的处理。由  $\text{Log}41 = 3.7136$ ， $\text{Log} 101 = 4.6151$ ，可相应得到具体的病期划分标准（见表 7）。

表 7 艾滋病病期划分标准（二）

病期	晚期	中期	早期
Log(CD4+1)值的范围	0~3.7136	3.7136~4.6151	4.6151~

#### 各病期人数的分布

考虑到样本数据量太大，为了避免异常点的影响，首先我们利用 DPS 软件对 CD4 数据进行异常点剔除，这样就排除了异常值对模型结果造成的误差。

按照表 7 给出的艾滋病病期划分标准，对剔除异常值的数据利用 EXCEL 软件进行自动筛选，得到最初开始测试 CD4 即还未开始服用这三种药物时 336 名患者在各病期的分布情况（见表 8）。

表 8 治疗前艾滋病人在各病期所占的比例

患病程度	年龄段	频数 $m$	比例 $r$
晚期	青壮年	161	52.27273
	中老年	85	27.5974
中期	青壮年	34	11.03896
	中老年	19	6.168831
早期	青壮年	5	1.623377
	中老年	4	1.298701

#### 统一求出第 0、8、16、24、32、40 周的 CD4 浓度

与模型 中处理数据的方法相同，我们利用差值法把时间统一到第 0、8、16、24、

32、40 周，求出每个病人在第 0、8、16、24、32、40 周的测试的 CD4 浓度（处理结果见附录 5）。

**各期不同年龄段平均 CD4 浓度**

因为题中只有 CD4 浓度的数据，现在我们就根据这组数据来预测病人继续治疗的效果，或者确定最佳治疗终止时间，在统计学上我们可以用样本的平均指标来估计总体 CD4 浓度，因此下面我们先求平均健康指标。

[模型的建立]

结合表 8 中三个病期的病人数，可建立以下平均 CD4 浓度模型，早期两个段在各检测时间的 CD4 浓度的平均值：

$$\overline{c_j^t} = \frac{\sum_{i=1}^{n_t} c_{ij}^t}{n_t} \quad t = 1,2,3 \quad j = 0,1,2,3,4$$

其中  $t = 1,2,3$  分别表示早、中、晚期， $c_{ij}^t$  表示  $t$  期第  $i$  个病人在第  $j$  次的 CD4 浓度。

[模型的求解]

利用 EXCEL 软件从附录 5 的数据中分别筛选出  $n_t(t = 1,2,3)$  个各病期不同年龄段病人的五个时段健康指标，再由上述模型用 Matlab 编程求解（见附录 6）。

[模型的结果]

各个病期不同年龄段病人五个检测时间的平均 CD4 浓度，结果见表 9。

表 9 不同阶段在不同检测时间的 CD4 的均值

CD4 均值		时间					
阶段		0	8	16	24	32	40
早期	青壮年	3.4393	16.628	30.288	55.728	9.9985	
	中老年	73.2481	254	286.9982	84.5816	61.2814	
中期	青壮年	58.1172	73.8901	91.8192	47.9226	81.7453	76.4979
	中老年	51.2374	82.1982	55.1307	71.3364	62.7776	75.7305
晚期	青壮年	15.7639	34.7782	325.887	386.915	47.4015	
	中老年	18.8295	35.2495	40.6169	32.4917	31.5210	29.8178

**模型的建立与求解**

**1) 数据的拟合原理：最小二乘法原理**

本模型的数据拟合原理与思路同模型 完全一致，即是最小二乘法原理。

**2) 对早中晚各期数据的拟合**

**对早期中的青壮年的 CD4 数据的线性拟合**

对早期的五组数据 (0，3.43932)(8，16.62806645)(16，30.28821184)(24，55.7280564) (32，9.998506749)，运用 Mathematica 软件画出的散点图（见图 8）。

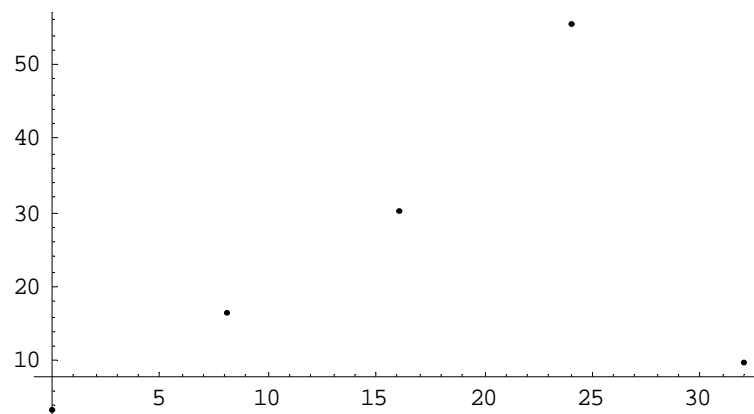


图 8 早期青壮年 CD4 浓度散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_1 = -0.0117x^3 + 0.4413x^2 - 1.9773x + 4.7859$$

做出早期中的青壮年拟合曲线图（见图 9），并可得到拟合曲线可决系数为 0.926：

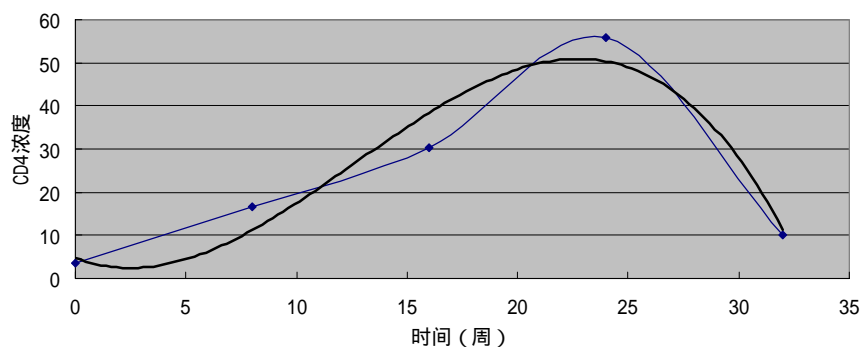


图 9 早期青壮年线性拟合

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间；令  $y_1' = 0$ ，求得  $x = 22.6592$ ，即可认为疗法 4 对青壮年早期爱滋病患者在第 22.6592 周终止治疗效果最好。

### 对早期中的中老年的 CD4 数据的线性拟合

对早期的五组数据 (0, 73.24812) (8, 254) (16, 286.9982) (24, 84.58164) (32, 61.28141)，运用 Mathematica 软件画出的散点图（见图 10）。

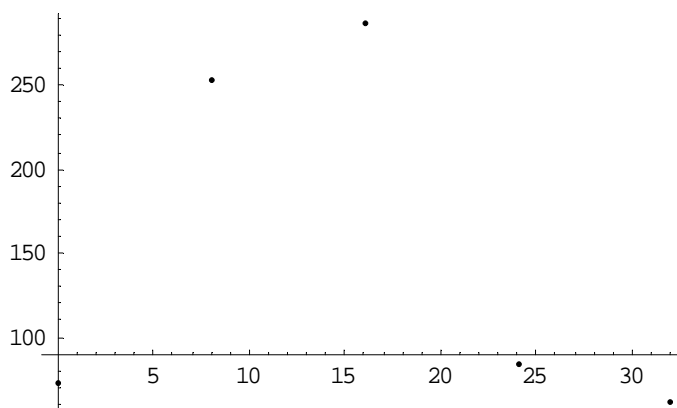


图 10 早期中老年 CD4 浓度散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_2 = 0.0523x^3 - 3.2719x^2 + 49.848x + 66.074$$

做出早期中的青壮年拟合曲线图(见图 11),并可得到拟合曲线误差平方和为 0.924：

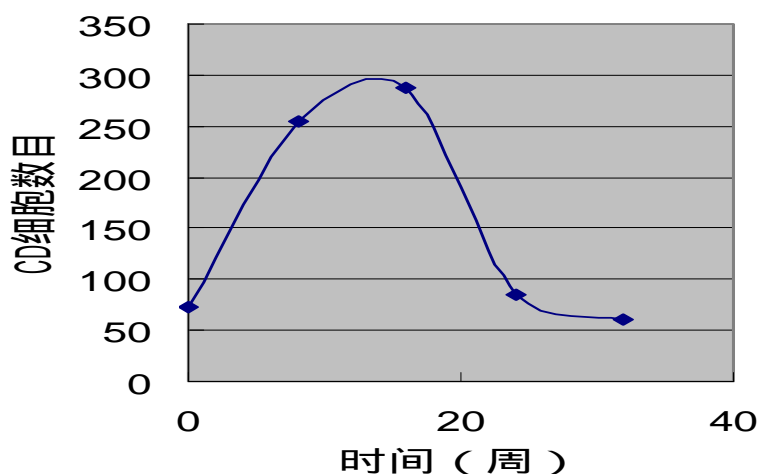


图 11 早期中老年线性拟合图

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间；令  $y_2' = 0$ ，求得  $x=10.0294$ ，即可认为疗法 4 对中老年早期爱滋病患者在第 10.0294 周终止治疗效果最好。

#### 对中期的青壮年的 CD4 数据的线性拟合

对中期的数据 (0, 58.11715624)、(8, 73.89005605)、(16, 91.81916437)、(24, 47.92268259)、(32, 81.74532234)、(40, 76.49786546)，运用 Mathematica 软件画出的散点图（见图 12）。

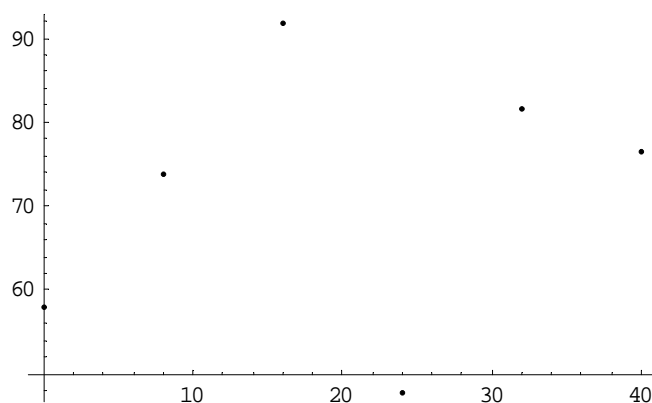


图 12 中期青壮年 CD4 浓度散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_3 = \begin{cases} -0.0208x^3 + 0.5167x^2 - 0.8291x + 58.117 & 0 < x < 25 \\ -0.3052x^2 + 21.321x - 287.97 & x \geq 25 \end{cases}$$

做出中期的青壮年拟合曲线图见图 13，并可得到拟合曲线误差平方和为 1：

观察图形，波动性很大，故我们分段处理，在第一个阶段，有最大值，当达到这一个点时就认为开始停药，即  $x=16.26$



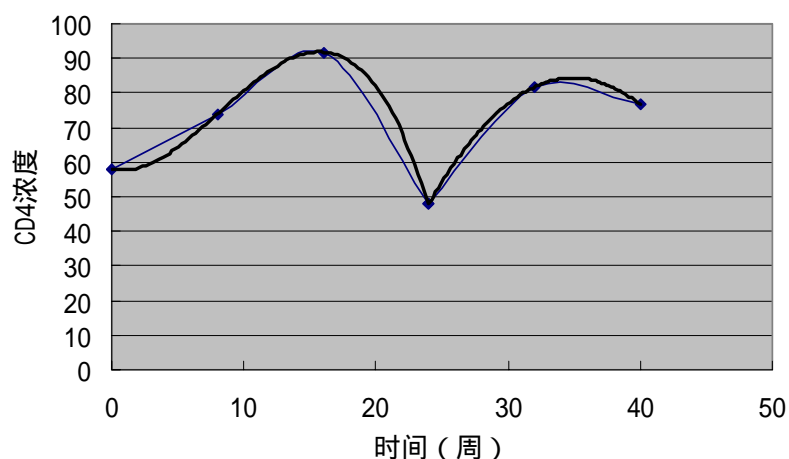


图 13 中期青壮年 CD4 浓度散点图

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间；  
令  $y'_3 = 0$ ，求得  $x=15.7154$ ，即可认为疗法 4 对青壮年中期艾滋病患者在第 15.7154 周终止治疗效果最好。

#### 对中期中的中老年人的 CD4 数据的线性拟合

对中期的六组数据  $(0, 51.23744)$   $(8, 82.19824)$   $(16, 55.13072)$   $(24, 71.33639)$   $(32, 81.74532234)$   $(40, 75.73053)$  运用 Mathematica 软件画出的散点图（见图 14）。

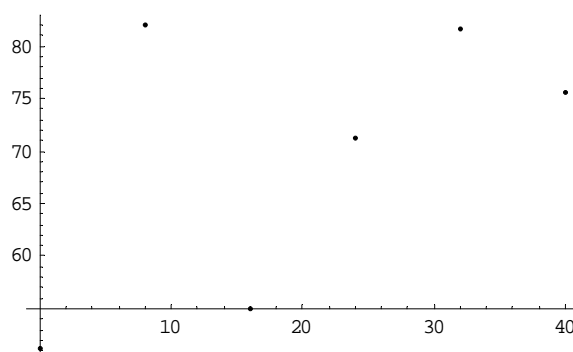


图 14 中期中老年 CD4 浓度散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_4 = 0.00007x^5 - 0.0075x^4 + 0.2772x^3 - 4.3042x^2 + 24.101x + 51.237$$

做出早期中的青壮年拟合曲线图见图 15，并可得到拟合曲线误差平方和为 1：

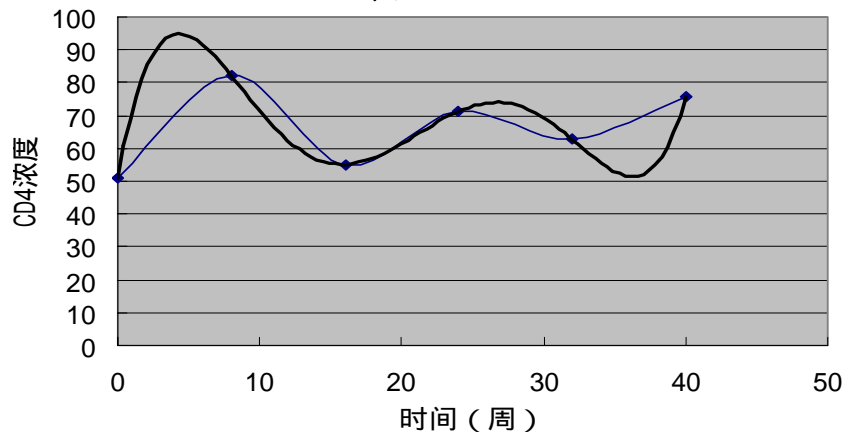


图 15 中期中老年 CD4 浓度散点图

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间；令  $y_4' = 0$ ，求得  $x=4.3657$ ，即可认为疗法 4 对中老年中期艾滋病患者在第 4.3657 周终止治疗效果最好。

#### 对晚期中的青壮年的 CD4 数据的线性拟合

对晚期的五组数据 (0, 15.76391735) (8, 34.778213) (16, 325.8866405) (24, 386.9153717) (32, 31.52096), (40, 47.40151308) 运用 Mathematica 软件画出的散点图 (见图 16)。

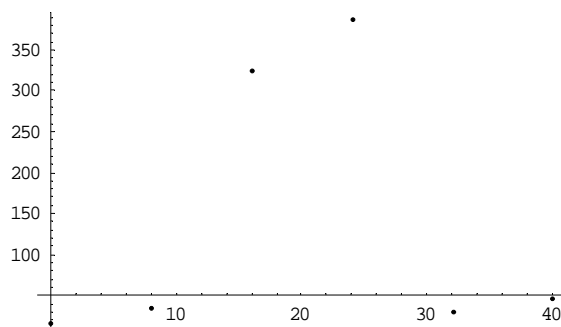


图 16 晚期青壮年 CD4 浓度散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_5 = 0.0063x^4 - 0.505x^3 + 11.811x^2 - 64.683x + 18.016$$

做出晚期中的青壮年拟合曲线图见图 17，并可得到拟合曲线误差平方和为 0.991：

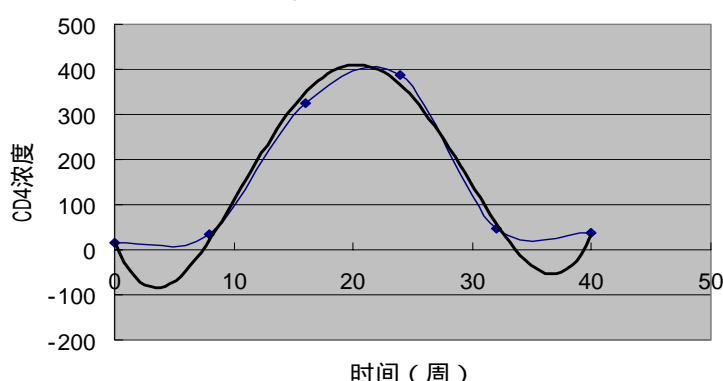


图 17 晚期青壮年线性拟合图

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间；令  $y_5' = 0$ ，求得  $x=23.622$ ，即可认为疗法 4 对青壮年晚期爱滋病患者在第 23.622 周终止治疗效果最好。

#### 对晚期中的中老年的 CD4 数据的线性拟合

对晚期的六组数据 (0, 18.82945) (8, 35.24953) (16, 40.61687) (24, 32.49173) (32, 31.52096), (40, 29.81775) 运用 Mathematica 软件画出的散点图 (见图 18)。

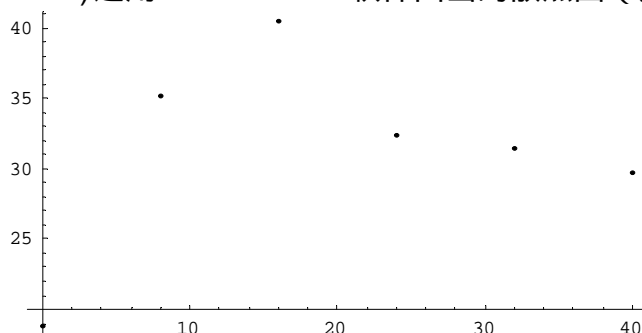


图 18 晚期中老年 CD4 浓度散点图

利用 EXCEL 软件对其用二次函数进行拟合，得到拟合函数如下：

$$y_6 = 0.0021x^3 - 0.1556x^2 + 3.2219x + 18.819$$

做出晚期中老年拟合曲线图见图 19，并可得到拟合曲线误差平方和为 0.955：

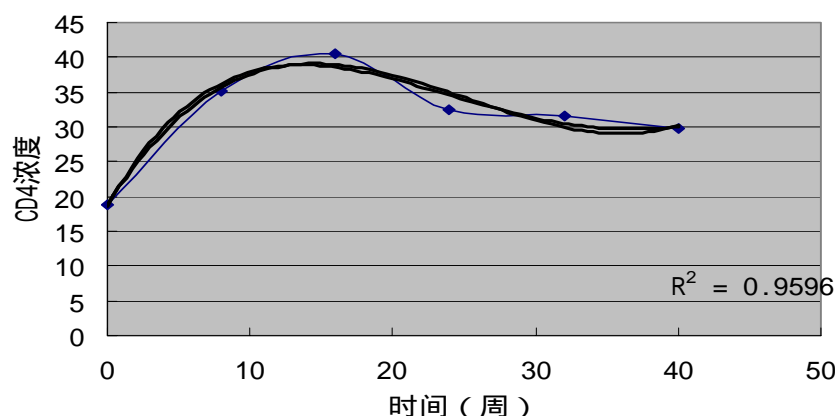


图 19 晚期中老年线性拟合图

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间；

令  $y'_6 = 0$ ，求得  $x=15.603$ ，即可认为疗法 4 对中老年晚期艾滋病患者在第 15.603 周终止治疗效果最好。

#### 模型结果：

将以上所求不同病期青壮年与中老年的最佳治疗终止时间汇总（见表 7）。

表 10 不同病期青壮年与中老年的最佳治疗终止时间

患病程度	年龄段	最佳治疗终止时间	年龄段	最佳治疗终止时间
晚期	青壮年	23.622	中老年	15.603
中期	青壮年	15.7154	中老年	4.3657
早期	青壮年	22.6592	中老年	10.0294

### 三、问题三的分析与求解

#### 1、对问题的分析

在不发达的国家和地区，人们收入普遍低下，大多数艾滋病患者的个人经济能力有限，而接受治疗需要支付高额的费用，这是患者难以承受的。尽管国家有政策对艾滋病患者进行一定帮助，但帮助往往是有限的，不是解决问题的根本。由于患者无法得到及时和有效的治疗，其病情会迅速恶化。现实生活中费用和疗效之间往往是对立的，即疗效高的疗法所花费的费用就高，而费用低的疗效相对要差得多。因此，如何能够花费少又能得到较好的疗效，这对不同经济承受力的人是值得考虑的一个有意义的事。我们在处理这个问题时，重点兼顾到费用和疗效这一对立问题的两个方面。问题三需要在这两个因素中找出一个合理的匹配关系，使得两个方面的都得到较好的兼顾。

#### 2、模型 统计实验模型<sup>[6]</sup>

##### 模型的分析

为了解决问题三，下面我们来设计一个统计实验。要对四种方法进行优劣比较，先固定一个量（比如各疗法的有效人数，由问题二记为  $n = n_1 = n_2 = n_3 = n_4$ ），由前面的统计概率模型：

$$p_i = \frac{n_i}{N_i}, \quad i = 1, 2, 3, 4$$

我们有

$$N_1 \cdot p_1 = N_2 \cdot p_2 = N_3 \cdot p_3 = N_4 \cdot p_4$$

其中  $p_i$  表示第  $i$  种疗法的疗效，然后比较  $N_i$  个人在一段时间内所花费的费用，显然，费用最小者为较优的治疗方法。这可由下面单目标模型来求解。

### 模型的建立

取  $s_i$  第  $i$  种疗法为一个人在一段时间内的全花费，对第  $i$  种疗法的  $N_i$  个人在一段时间内所花费的费用可由下面公式给出

$$s_i \times N_i, \quad i = 1, 2, 3, 4$$

则较优的治疗方法的花费为：

$$\min_{1 \leq i \leq 4} (s_i \times N_i)$$

### 模型的求解

#### 确定样本容量 $N_i$

我们任意选取  $N_i$  中一个值，如选  $N_1 = 1000$ ，又由表 6 中四种疗法对应的有效治疗率依次为  $p_1 = 0.31$ 、 $p_2 = 0.42$ 、 $p_3 = 0.41$ 、 $p_4 = 0.54$ ，把以上变量代入公式  $N_1 \cdot p_1 = N_2 \cdot p_2 = N_3 \cdot p_3 = N_4 \cdot p_4$ ，可分别求得： $N_2 = 738$ 、 $N_3 = 756$ 、 $N_4 = 574$ 。

#### 确定治疗期限

可根据具体情况进行确定。在这里为了方便疗法 1 两种药的按月交替使用，故选  $t=8$  周（二个月）时间进行比较。

由 4 种疗法的日用药分别为：600mg zidovudine 或 400mg didanosine（去羟基苷），这两种药按月轮换使用；600 mg zidovudine 加 2.25 mg zalcitabine（扎西他滨）；600 mg zidovudine 加 400 mg didanosine；600 mg zidovudine 加 400 mg didanosine，再加 400 mg nevirapine（奈韦拉平）。

又由 600mg zidovudine 1.60 美元，400mg didanosine 0.85 美元，2.25 mg zalcitabine 1.85 美元，400 mg nevirapine 1.20 美元。则

疗法 1 的费用为： $28 \times 1.60 + 28 \times 0.85 = 70.3$  (美元)；

疗法 2 的费用为： $56 \times (1.60 + 1.85) = 193.2$  (美元)；

疗法 3 的费用为： $56 \times (1.60 + 0.85) = 137.2$  (美元)；

疗法 4 的费用为： $56 \times (1.60 + 0.85 + 1.2) = 204.4$  (美元)。

则每人各疗法在  $t=8$  周内的费用如表 2：

表 11：各疗法在二个月内的费用

疗 法	疗法 1	疗法 2	疗法 3	疗法 4
费用 $s_i$ (美元)	70.3	193.2	137.2	204.4

#### 确定 $N_i$ ( $i = 1, 2, 3, 4$ ) 个人在一定时间内的总费用。

利用公式  $s_i \times N_i$  得到在二个月内的总费用（见表 12）。

表 12：第  $i$  种疗法的  $N_i$  个人在 8 周内所花费的总费用

疗 法	疗法 1	疗法 2	疗法 3	疗法 4
人数 $N_i$	1000	738	756	574
总费用 ( $N_i \times s_i$ )	70300	142580	103720	117330

从表 12 可以看出，费用由少到大依次为：疗法 1、疗法 3、疗法 4、疗法 2。综合考虑疗法 1 效果最好，疗法 2 效果最差。

#### 多次实验验证：

对疗法 1 的人数  $N_1$  取不同的值，或治疗期间  $t$  取不同的时段，同样我们可以得到的相同的结论，即疗法 1 效果仍然最好，疗法 2 效果最差。

#### 结论：

与问题二比较，疗法优劣改变了，综合考虑疗效、费用，我们得出疗法 1 的效果好。由于疗法的改变，他们的预测（或者提前结束）必然改变，再加上费用的限制，最佳停药时间一定在无费用限制的情况下出现的早。

#### 四、疗法 1 的最佳停止时间的进一步讨论

对疗法 1 的最佳治疗终止时间的分析，我们同样运用问题一中所用的线性插值法进行处理，把时间统一到第 0、8、16、24、32、40 周，求出每个病人在第 0、8、16、24、32、40 周的测试的 CD4 浓度，再借助 EXCEL 软件用模型 进行拟合，做出各期的拟合图：结果如下

##### 早期中的 CD4 数据的线性拟合如图 20

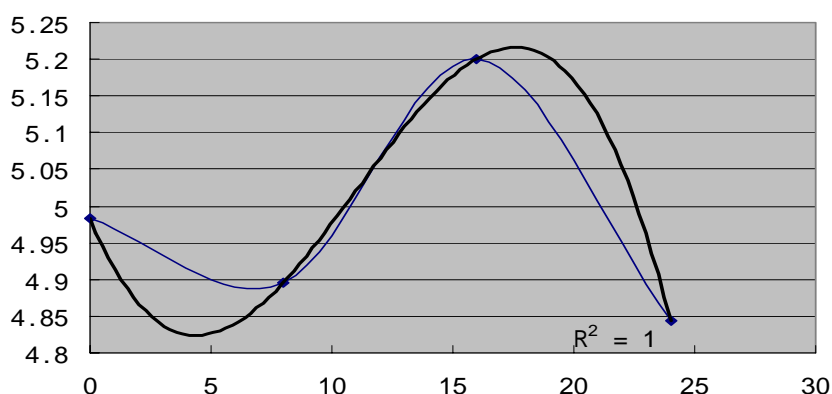


图 20 早期患者线性拟合图

根据图 20 可得到拟合函数如下：

$$y = -0.0003x^3 + 0.0113x^2 - 0.0793x + 4.9836$$

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间为：  
 $x_1 = 18.233$

##### 对中期的青壮年的 CD4 数据的线性拟合如图 21

根据图 21 可得到拟合函数如下：

$$y = -10^{-6}x^4 + 9 \times 10^{-5}x^3 - 0.0016x^2 + 0.0041x + 4.0328$$

利用 MATLAB 软件求解曲线的最大值对应的时间，即为最佳治疗终止时间为 32.45

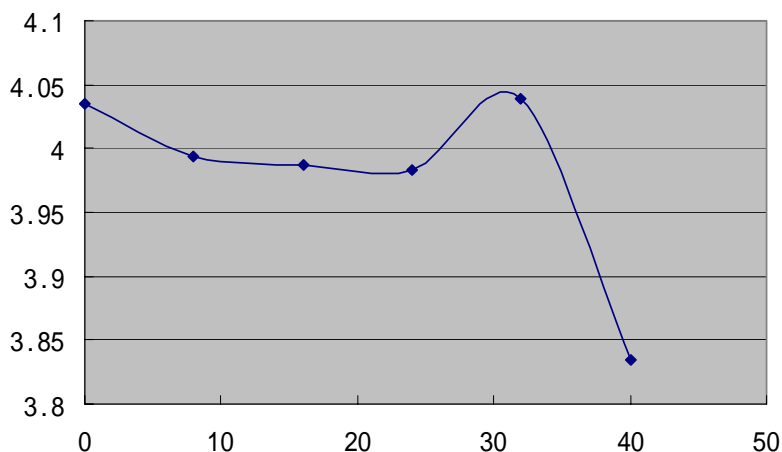


图 21 中期青壮年线性拟合图

### 对中期的中老年的 CD4 数据的线性拟合

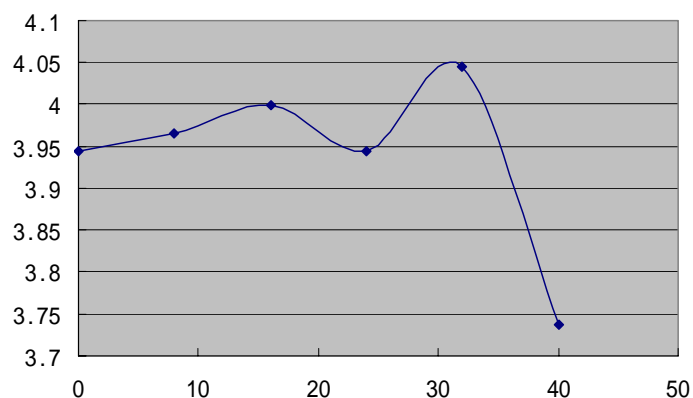


图 22 中期中老年线性拟合图

根据上图可得到拟合函数如下：

$$y = -2 \times 10^{-6} x^6 + 0.0002x^3 - 0.0035x^2 + 0.0252x + 3.939$$

利用 MATLAB 软件求解曲线的最大值对应的时间 ,即为最佳治疗终止时间为 33.58

### 对晚期中的青壮年的 CD4 数据的线性拟合

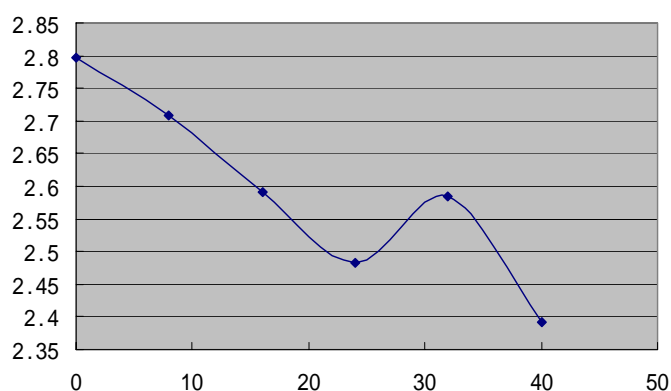


图 23 晚期青壮年线性拟合图

根据上图可得到拟合函数如下：

$$y = 0.00012x^2 - 0.013x + 2.7929$$

拟合函数在 0 至取得最大值，即用药效果始终在下降的，因此疗法 1 对于晚期青壮年的疗效不显著，即对此类患者不宜采用此疗法。

### 对晚期中的中老年的 CD4 数据的线性拟合

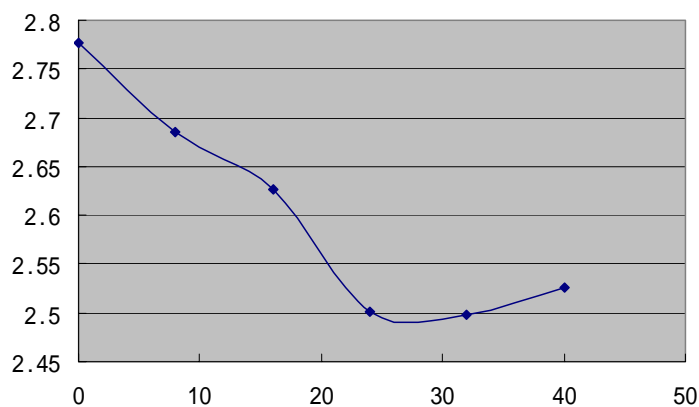


图 24 晚期中老年线性拟合图

根据上图可得到拟合函数如下：

$$y = 0.0002x^2 - 0.0162x + 2.7902$$

拟合函数在 0 至取得最大值，因此疗法 1 对于晚期青壮年的疗效不显著，即对此类患者不宜采用此疗法。

将疗法 1 与疗法 4 的最佳停药时间综合分析见表 13

表 13 疗法 1 与疗法 4 的最佳停药时间综合分析表

不同阶段的患者	疗法 1 的最佳停药时间	疗法 4 的最佳停药时间
早期患者	18.223	15.5
中期中的青壮年	32.45	15.719
中期中的中老年	33.58	4.3657
晚期中的青壮年	-	23.622
晚期中的中老年	-	15.603

注：缺省值表示用药期间效果不显著。持续下降。

从表中可得疗法 1 对早中期患者有疗效，且效果还可以；而疗法 4 对早期、中期中的青壮年、晚期的用药效果有效。建议：晚期患者在经济情况下，还是用疗法 4 中用药。

## § 6 模型的误差分析、检验和进一步讨论

### 一、误差分析

#### 1、考虑线性拟合过程

在模型 中，由于题目附件 1 中给出的是 300 多名病人每隔几周测试的 CD4 和 HIV 的浓度，每次检测的时间没有严格的周期或时间段，我们按照一维线性差值的方法，把时间统一到第 0、4、8、24、40 周，预测出病人分别在 0、4、8、24、40 周的 CD4 和 HIV 的浓度，但是实际情况中病人的终止时间和平均健康指标并不是完全严格地按照线性函数来发展的，因此在拟合预测过程中一定会产生拟合误差。

在模型 中，由于题目附件 2 中给出的是 1300 多名病人大约按照每八周的周期测试，与模型 相同，我们仍是利用一维线性差值的方法把时间统一到第 0、8、16、24、32、40 周，预测出病人分别在 0、8、16、24、32、40 周的 CD4 浓度，但是实际情况中病人的终止时间和平均 CD4 浓度并不是完全严格地按照线性函数来发展的，因此在拟合预测过程中一定会产生拟合误差。

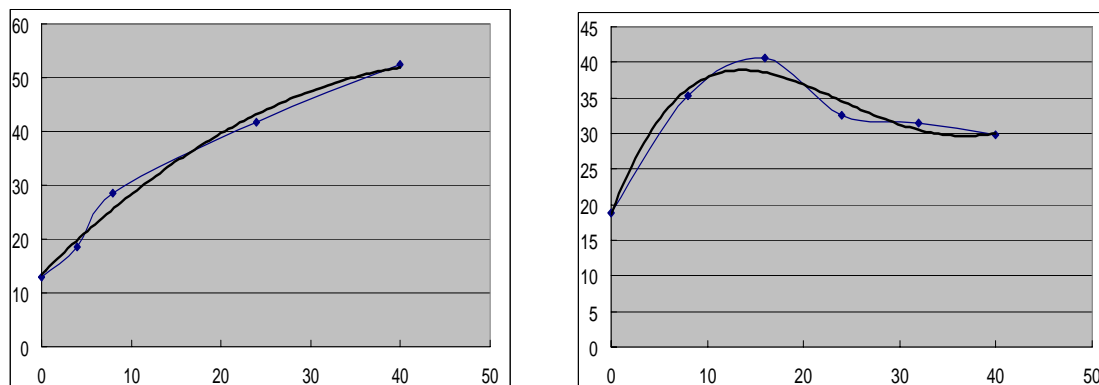


图 20 关于不同拟合下的图表

#### 2、考虑缺省值

在模型一中，原始数据中 HIV 浓度的缺省值我们可以认为是 HIV 浓度过大或过小，

由于检测技术与设施的限制,不能测试模型 中缺省值可以出来,因此此时的 HIV 浓度值缺省。我们可以通过相应时刻的 CD4 浓度的检测值判断出缺省的 HIV 浓度是过大还是过小,这样产生的误差会较小。

### 3、考虑年龄段的划分

由于将病人的年龄是按照大于 40 岁的划分为中老年,小于 40 岁的划分为青壮年,来考虑不同年龄的病人用药后的恢复状况的,但是病人用药后的恢复状况与年龄有着密切的关系,不同年龄的患者的体质也有所不同,我们只将其看成青壮年、中老年两类,不能全面的考虑年龄这个因素,这就会产生误差。

### 4、考虑疗效与费用

不同的患者,对待治疗费用与所产生疗效的态度是不同的,有人重视治疗费用,有人重视产生疗效,但是所建模型给人们的选择赋予了一定的权重,认为他们的选择是固定的,因此存在不可避免的误差。

### 5、考虑数据处理过程

并非所有病人都是按时去治疗的,或者在一段时期后不去治疗,这就使所得的数据并非都是完全的,对某一时刻的平均估计也就不是完全的,这也使模型与实际相偏差。

## 二、模型的检验

对于模型中年龄段的划分,标准是我们自定的,将所有患病人数分为两个年龄段,我们利用马克威分析系统中的聚类分析软件将患病人数分成两类,可以对原模型进行检验。

## 三、模型的进一步讨论：

对于模型 还可以考虑在一定的费用条件下,使用效率高的疗法,而缩短治疗时间,这样更符合实际情况.我们可以利用一些统计软件中自带的分析系统如马克威分析系统,DPS 软件等其他统计或数学应用软件对不同病期与不同年龄段的健康指标 CD4 的浓度进行拟合。

## § 7 模型的评价与推广

### 一、模型的优缺点

#### 1、优点：

利用 EXCEL 软件对数据进行处理并作出各种平面图,简便,直观、快捷;

本文建立的模型与实际紧密联系,充分考虑现实情况的不同阶段,从而使模型更贴近实际,通用性强;

运用多种数学软件进行计算,取长补短,使计算结果更加准确;

#### 2、缺点：

对附件 1 中的 HIV 病毒数目缺省值作为异常值,将所在的行删除,增加了误差;

在用插值法处理数据中,插值的点不多,拟合的曲线不能精确的说明问题。

### 二、模型的推广：

本文针对附录的数据进行插值,然后分类、拟合,从不同的患者情况、不同年龄阶段分别进行考虑,具有现实意义,可以适用于方案的选优、人才的选拔、治疗方案的确立等一些问题。

模型 又综合考虑费用的限制,确定最佳治疗方案,给艾滋病患者提供参考,有助于患者根据自己的经济情况、患病状况进行选择治疗方案。



## 参考文献

- [1]尚红,HIV检测及临床意义, [www.hiv-vct.net](http://www.hiv-vct.net), 2006-1-16 ;
- [2]于洪彦,EXCEL 统计分析与决策, 北京:高等教育出版社,2001.5 第二版 ;
- [3]张韵华,Mathematica 符号计算系统实用教程,合肥:中国科技大学出版社, 1998.9 第一版 ;
- [4]茆诗松,周纪芃,概率论与数理统计,北京:中国统计出版社,2000.7 第二版 ;
- [5]胡守信,李柏年,基于 MATLAB 的数学实验,北京:科学出版社,2004.6 第一版 ;
- [6]任建标译,数据、模型与决策, 北京:中国财政经济出版社,2004.1 第二版。

## 附录

### 附录 1

Zaoqinihe.m:

```
a=[0    33.72678841
4    77.15917835
8    96.89483412
24   119.2862735
40   126.7080751];
x1=a(:,1);y1=a(:,2);
plot(x1,y1);
k=[1:3,5:7,9:23,25:39];
y2=interp1(x1,y1,k);
s=[k;y2];
t=[s';a];
plot(t(:,1),t(:,2),'*');
[x,y,z]=solve('c=b/(ln(33.72678841)-ln(a))','4+c=b/(ln(77.15917835)-ln(a))','40+c=b/(ln(126.7080751)-ln(a))');
b01=[139.25680815982182343186662306216,-4.0468598237414255012327571993945,2.8538658680560016698512456578579];
fun1=inline('b(1)*exp(b(2)./(x+b(3)))','b','x');
[b1,r1,j1]=nlinfit(x1,y1,fun1,b01);
b1;
y=138.72*exp(-3.94./(x1+2.78));
plot(x1,y1,'*r',x1,y,'<g')
r2=1-sum(r1.^2)/sum((y1-mean(y1)).^2)
plot(x1,y1,'*',x1,y,'.')
```

### 附录 2 :

zhongqi.m:

```
a=[0    14.18601835
4    40.76365108
8    55.20248271
24   76.42376878
40   85.50783741];
x1=a(:,1);y1=a(:,2)
[x,y,z]=solve('c=14.18601835','55.20248271=64*a+8*b+c','85.50783741=1600*a+40*b+c')
b01=[-0.1045,5.963,14.186];
fun1=inline('b(1)*x.^2+b(2).*x+b(3)','b','x');
[b1,r1,j1]=nlinfit(x1,y1,fun1,b01);
b1
x=0:0.01:100;
```

```

y=-0.0614*x.^2 +4.0272* x+ 20.9211; %原数据拟和曲线
plot(x1,y1, '*r',x,y, '.g');
r21=1-sum(r1.^2)/sum((y1-mean(y1)).^2)

```

### 附录 3 :

```

wanqi.m:
a=[0      12.84206319
4      18.5492997
8      28.63948375
24     41.67688213
40     52.3539174
];
x1=a(:,1);y1=a(:,2);
plot(x1,y1)
k=[1:3,5:7,9:23,25:39];
y2=interp1(x1,y1,k);
s=[k;y2];
t=[s';a];
%plot(t(:,1),t(:,2), '*');
[x,y,z]=solve('0=ln(12.84206319*b)-ln(a-12.84206319)', '8*c=ln(28.63948375*b)-ln(a-28.63948375)', '40*c=ln(52.3539174*b)-ln(a-52.3539174)');
b01=[0.034902416791254783199873694996710,0.042966690781638260631971832622128];
fun1=inline('1./(b(1)+b(2). *exp(-x))', 'b', 'x');
[b1,r1,j1]=nlinfit(x1,y1,fun1,b01);
b1;
y=1./(0.0241+1.6867*exp(-x1));
plot(x1,y1, '*r',x1,y, '<g');
r2=1-sum(r1.^2)/sum((y1-mean(y1)).^2);

```

### 附录 4 :

Pt ID	CD4Date	CD4Count	RNADate	VLoad
23424	0 178	0 5.5		
23424	4 228	4 3.9		
23424	8 126	8 4.7		
23424	25 171	25 4		
23424	40 99	40 5		
23425	0 14	0 5.3		
23425	4 62	4 2.4		
23425	9 110	9 3.7		
23425	23 122	23 2.6		
23425	40 320			
23426	0 101	0 4.5		
23426	4 151	4 1.7		
23426	8 115	8 1.7		

23426	26	149	26	2.8
23426	46	120	46	3.4
23426	54	141		
23427	0	10	0	5.3
23427	4	11	4	2.7
23427	7	30	7	2.6
23427	24	137	24	2
23427	40	225		
23428	0	161	0	5.5
23428	3	220	3	2.5
23428	8	316	8	1.9
23428	24	645	24	1.7
23428	40	451	40	1.7
23429	0	7	0	5.5
.....				
.....				
.....				
23755	44	124	44	1.7
23756	0	112	0	5
23756	4	241	4	2.5
23756	8	205	8	2.4
23756	24	207	24	2.1
23756	39	164	39	3
23757	0	42	0	4.6
23757	4	65	4	3.2
23757	7	101	7	2.6
23757	24	71	24	3.2
23757	39	65	39	3.3
23758	0	58	0	4.3
23758	4	80	4	1.8
23758	8	141	8	1.8
23758	24	111	24	1.7
23758	40	136	40	1.7
23759	0	33	0	4.4
23759	4	65	4	2.5
23759	7	84	7	1.9
23759	23	112	23	1.7
23759	39	198	39	1.8
23760	0	59	0	5.4
23760	3	135	3	4.3
23760	7	207	7	3.1
23760	23	340	23	3.5
23760	40	210	40	4.5
23761	0	41	0	4.8

23761	4	180	4	3
23761	9	191	9	2.4
23761	24	199	24	1.7
23761	40	213	40	1.7
23762	0	17	0	5.3
23762	3	24	3	3.8
23762	7	49	7	2.9
23762	23	98	23	1.7
23762	38	115	38	1.7
23762	55	132		
23763	0	128	0	4.7
23763				
3775	9	230	9	1.7
23775	25	240	25	1.7
23775	41	233	41	1.7
23776	0	15	0	5.4
23776	4	71	4	2.6
23776	8	99	8	2
23776	24	112	24	3.3
23776	40	81	40	4.8
23777	0	197	0	3.5
23777	4	181	4	2.2
23777	8	231	8	1.9
23777	24	283	24	1.7
23777	40	350	40	1.9
23780	0	11	0	4.9
23780	7	10	7	4.7
23780	35	86		
23780	44	106	44	1.7
23990	0	63	0	4.6
23990	3	195	3	3.1
23990	7	154	7	3.4

## 附录 5

```

a=[0      12.84206319
4    18.5492997
8    28.63948375
24   41.67688213
40   52.3539174
];
x1=a(:,1);y1=a(:,2);
plot(x1,y1)
k=[1:3,5:7,9:23,25:39];
y2=interp1(x1,y1,k);

```

```

s=[k;y2];
t=[s';a];
%plot(t(:,1),t(:,2),'*');
[x,y,z]=solve('0=ln(12.84206319*b)-ln(a-12.84206319)', '8*c=ln(28.63948375*b)-ln(a-28.63948375)', '40*c=ln(52.3539174*b)-ln(a-52.3539174)');
b01=[0.034902416791254783199873694996710,0.042966690781638260631971832622128];
fun1=inline('1./(b(1)+b(2).*exp(-x))','b','x');
[b1,r1,j1]=nlinfit(x1,y1,fun1,b01);
b1;
y=1./(0.0241+1.6867*exp(-x1));
plot(x1,y1,'*r',x1,y,'<g');
r2=1-sum(r1.^2)/sum((y1-mean(y1)).^2);
a=[-5.417999938 0
100000 100000
100000 100000
100000 100000
100000 100000
12.79356216 54.86884357
100000 100000
54.86884357 0
0 9.806679651
100000 100000
100000 100000
100000 100000
0 30.83392161
100000 100000
100000 100000
3.089591104 -35.95380343
100000 100000
100000 100000
-21.23829857 -34.10309954
100000 100000
-34.10309954 -14.51330638
100000 100000
-14.51330638 36.66346558
100000 100000
100000 100000
100000 100000
6.62568748 44.64318223
100000 100000
100000 100000
.....
-23.32448322 47.28764371
100000 100000

```

```

100000 100000
100000 100000
100000 100000
100000 100000
33.8263199 -2.451634423
100000 100000
100000 100000
-29.40429377 -59.34181968
100000 100000
-59.34181968 -12.75919575
100000 100000
100000 100000
100000 100000
100000 100000
23.22832622 0
0 0
0 0
0 0
0 0
];
t=[];
for i=1:1170;
    j=1:2;
    if a(i,j)<100000
        a1=a(i,j);
        t=[t;a1];
    end
end
t

```