

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): 重庆大学

参赛队员(打印并签名): 1. 曾爱凤

2. 杨源飞

3. 刘耀中

指导教师或指导教师组负责人(打印并签名): 龚劬

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

【摘要】本问题要求我们利用艾滋病服药治疗期间测试所得的数据预测继续治疗的效果并评价各种疗法的优劣。对三个问题我们都运用回归分析方法和方差分析方法来建立统计模型。

就问题一，首先对所有数据进行处理，建立了一元多项式回归模型，将多项式回归模型转化为多元线性回归模型运用 Matlab 软件求解，通过最小二乘法参数估计和统计检验，得到关于与测试时刻的比较准确的函数关系及其回归曲线，由此预测出药物治疗的效果，最终得出在药物治疗的第 25.0000 周停止用药最为合适的结论。然后提出分类的思想，即按不同病人在服药前期的 CD4 浓度大小不同将数据分成三组，对每组数据同样按前面的思路建立回归模型并求解，确定最佳治疗终止时间依次为第 25.1295 周，第 24.0000 周，第 25.0000 周。分类前后的预测结果大致相符。

问题二在增多了病人年龄因素的基础上对四种疗法的优劣进行了评价。首先，对每组治疗方案建立 4 种多元回归子模型进行定性分析，利用 Matlab 统计工具箱来处理，拟合出交互式图形，并分析剩余标准差 s ，筛选出每组治疗方案最佳模型。然后分析 CD4 浓度曲线，对 4 种疗法的优劣进行了排序，选择疗法四为最优疗法，其最佳终止服药时间为第 17.8854 周。然后用方差分析进行定量评价。定性和定量评价的结果相符。

问题三在问题二基础上增加了对治疗费用的考虑，建立多元回归模型，同样利用 Matlab 统计工具箱来处理。然后提出“疗效性价比”这一概念并建立模型，以单位费用的治疗的效果来评价 4 种疗法，在交互式图形中分别取治疗费用为 0 和 200 美元的两个时刻作为观测点，分析 4 种疗法的疗效性价比，最后得到第 1, 2, 3, 4 种疗法依次较优，即疗法四为最优方案，在预测继续治疗的效果不佳的情况下，确定最佳终止时间为第 17.8854 周，终止治疗时共花费 453.2464 美元。

关键词：回归分析 治疗效果预测 方差分析 疗效性价比

一、问题的重述

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、模型的假设

- 1、附件 1、2 数据所反映的规律基本符合 AIDS 病人药物治疗过程的病况规律，即能客观的反映现实情况。
- 2、在测试和预测期间 AIDS 病人不会死亡。
- 3、病人在服药测试初始 CD4 浓度相等。
- 4、问题三第一种疗法中第奇数个月（如第 1 个月）服用 400mg didanosine（0.85 美元/天），第偶数个月（如第 2 个月）服用 600mg zidovudine（1.60 美元/天）。
- 5、问题三治疗费用只由治疗药品费用组成。
- 6、每月按 30 天计算。

三、符号说明

符号	表示意义
C	病人的 CD4 浓度（0.2 个/ml）
h	病人的 HIV 浓度（单位不详）
c_i	第 i 类病人的 CD4 浓度（0.2 个/ml）， $i=1,2,3$
h_i	第 i 类病人的 HIV 浓度（单位不详）， $i=1,2,3$
d	测试 CD4 和 HIV 浓度的时刻（单位：周）
d_j	问题二、三中第 j 种疗法的时刻（单位：周）， $j=1,2,3,4$
y_j	对数处理后第 j 种疗法病人体内 CD4 浓度， $j=1,2,3,4$

符号	表示意义
a_j	第 j 种疗法的 AIDS 病人年龄, $j=1,2,3,4$
k_j	第 j 种疗法的拟合曲线的斜率的绝对值, $j=1,2,3,4$
f_j	第 j 种疗法的费用 (单位: 美元), $j=1,2,3,4$
s_1	停止治疗时为奇数月份时总的治疗费用
s_2	停止治疗时为偶数月份总的治疗费用
Δf_j	采用第 j 种疗法时费用增量, $j=1,2,3,4$
Δy_j	费用增量为 Δf_j 时 CD4 的浓度增量 $j=1,2,3,4$
p_j	疗效性价比, $j=1,2,3,4$

四、问题的分析、模型的建立及求解

4.1 问题一的分析及模型建立:

4.1.1 问题一的提出和数据分析

已知: 同时服用 3 种药物 (zidovudine, lamivudine, indinavir) 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)。

要求: 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)。

分析: 题目的附件 1 给出的 CD4 和 HIV 浓度是 300 多名病人每隔几周测试的, 每个病人测试 3 到 6 次不等, 加起来共有 1665 组数据。根据题目要求, 我们应当寻找出药物治疗过程中 CD4 和 HIV 浓度随治疗时间变化而变化的规律图线, 以此来判断和预测。

首先, 以测试时刻 (周) 为横轴, 分别以 CD4、HIV 的浓度为纵轴, 作出浓度关于时刻的散点图进行观测, 由于数据点太多, 测试时刻相对较集中且每个病人测试次数只有几次, 从散点图上不容易看出它们之间的规律。然后抽取一小部分病人数据作每个病人的浓度折线图, 可以看出浓度值大致变化趋势。于是, 再用回归分析法拟合出浓度和时刻的关系曲线并做统计检验。仔细观察分析数据我们发现, 不同的病人在初始时刻 CD4 浓度 (或 HIV 浓度) 有所不同, 即每个病人的病程是不一样的, 于是我们提出按服药前 (第 0 周或者第 1 周) CD4 浓度的不同将病人分为三类, 如表 1 所示:

表 1 以 CD4 浓度为标准分类

第 p_i 类病人	p_1	p_2	p_3
CD4 浓度 (0.2 个/ml)	0—75	75—150	>150

然后对每一类数据各自分析讨论,同样用回归分析法拟合出浓度和时刻的关系曲线并做统计检验。以上两种方法同时用来解决问题一(即预测继续治疗的效果,确定最佳治疗终止时间),再比较两种方法的优劣。

4.1.2 模型的建立及求解

(一) 模型一: 分类前的回归模型

首先,记病人的 CD4 浓度 (0.2 个/ml) 为 C , HIV 浓度 (单位不详) 为 h

(为了便于在同一个图像上作出 CD4 和 HIV 浓度与时刻的回归曲线,在数据时均将 HIV 浓度数据扩大了 50 倍), d 为测试 CD4 和 HIV 浓度的时刻(单位:周),由附件 1 初步分析,随着服药时间的增加,CD4 浓度先是增加,但是随着时间增加过多后 C 反而减少,而 HIV 浓度 h 先是减少然后又增加,所以不应拟合线性函数,可能是二次函数比较合适。

建立如下二元多项式回归模型:

$$c = \beta_0 + \beta_1 d + \beta_2 d^2 + \varepsilon \quad (\text{I})$$

$$h = \beta_0 + \beta_1 d + \beta_2 d^2 + \varepsilon \quad (\text{II})$$

其中 $\beta_0 \sim \beta_2$ 为回归系数, ε 是相互独立的、期望为 0、方差为 σ^2 、正态分布的随机变量,即 $\varepsilon \sim N(0, \sigma^2)$, ε 称(随机)误差。

- 1) 利用最小二乘法进行参数估计,确定系数 $\beta_0 \sim \beta_2$, 得到预测方程;
- 2) 给定显著水平 $\alpha=0.05$, 对回归系数进行相关系数检验及 F 检验,确定回归分析的效果,直到回归分析的效果好。

用 Matlab 软件命令 regress 求解,将回归系数的估计值如表 2 所示,并代入 (I)、(II) 式得到 CD4 浓度 C 和 HIV 浓度 h 的预测方程为

$$c = 97.7277 + 6.6004d - 0.1114d^2 \quad (\text{III})$$

$$h = 439.7715 - 16.2617d + 0.3289d^2 \quad (\text{IV})$$

表 2 模型一 CD4 和 HIV 浓度的计算结果

回归系数	回归系数估计值		回归系数置信区间	
	CD4 浓度	HIV 浓度	CD4 浓度	HIV 浓度
β_0	97.7277	439.7715	[89.3095 106.1458]	[429.4523 450.0907]
β_1	6.6004	-16.2617	[5.2526 7.9481]	[-17.9164 -14.6071]
β_2	-0.1114	0.3289	[-0.1452 -0.0776]	[0.2874 0.3705]
CD4 浓度	$R^2=0.1113$ $F=98.1$ $p=0$ $s^2=9990.4$			
HIV 浓度	$R^2=0.2401$ $F=247.3568$ $p=0$ $s^2=1.5001 \times 10^4$			

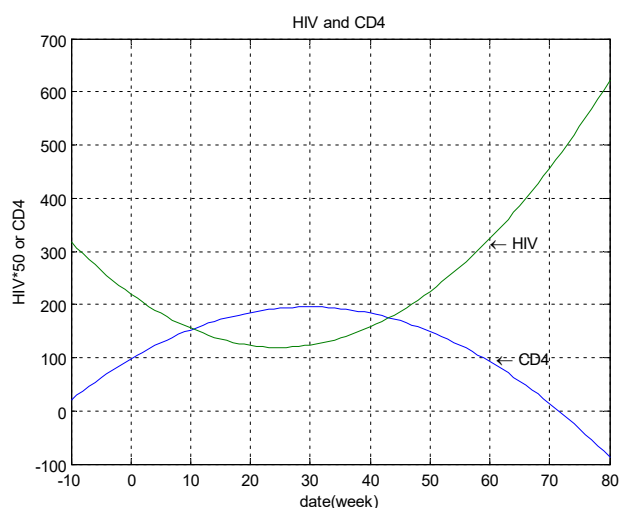


图 1 HIV 浓度 $\times 50$ 和 CD4 浓度的回归曲线

由预测方程 (III) (IV) 及图 1 上的回归曲线可以看出 C 是开口向下的抛物线而 h 是开口向上的抛物线, 这符合 AIDS 患者用药后体内 CD4 浓度和 HIV 浓度随时间变化的规律。

回归系数 $\beta_0 \sim \beta_2$ 的置信区间比较短, 而且都没有包含零点, 表明回归模型有效。

回归模型的统计检验 检验水平为 $\alpha=0.05$

1) **相关系数检验法:** 相关系数 R 是反映随机变量 X 与 Y 呈线性程度的一个度量指标, 其取值范围 $|R| \leq 1$, 当 $|R|$ 接近于 1 时, 变量 X 与 Y 呈密切线性相关, 当 $|R|$ 接近于 0 时, 变量 X 与 Y 非线性相关。本模型

CD4 浓度和 HIV 浓度的计算结果中 R^2 分别为 0.1113 和 0.2401, 都明显偏小, 可见 HIV 和 CD4 浓度与检测时刻呈非线性相关, 与我们的模型假设是相符的。

2) **F 检验法**：若 $F > F_{\alpha}(1, n-2)$ 则回归效果好。用 Matlab 求得

$F_{0.05}(1, 1567) = 3.8474$ ，由表 2 可看出， F 的值都明显大于 3.8474，

表明回归方程的效果显著。

问题结果及分析

由图 1 得到曲线中 CD4 浓度最大值所对应的时刻为 29.2317（周），HIV 浓度最小值所对应的时刻为 25.0000（周），也就是病人服药治疗期间，在第 25.0000 周体内 HIV 浓度降到最低，继续服药则 HIV 浓度又会反弹上升；同时 CD4 浓度还在上升，到了第 29.2317 周才上升到最大值，之后呈下降趋势。由于艾滋病治疗的目的，是尽量减少人体内 HIV 的数量，提高 CD4 的数量也是为了抑制 HIV，为了防止 HIV 浓度上升导致病情的恶化，在 HIV 浓度降到最低且尚未反弹时就应当停止服药。由此可以确定最佳治疗终止时间也就是 HIV 浓度降到最低的时间。

模型一得到的最佳治疗终止时间为第 25.0000 周。

（二）模型二：分类后的回归模型

将不同病人在初始时刻 CD4 浓度的不同进行分类后得出的数据分为三组，分别得到 M_1 ， M_2 ， M_3 组矩阵。对每组矩阵数据再建立关于 HIV、CD4 浓度与测试时间的二次函数：

$$c_i = \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + \varepsilon \quad (i = 1, 2, 3) \quad (V)$$

$$h_i = \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + \varepsilon \quad (i = 1, 2, 3) \quad (VI)$$

用 Matlab 软件命令 regress 求解,将回归系数的估计值代入 (V) (VI) 式得到 CD4 浓度 c_i 和 HIV 浓度 h_i 的预测方程

表 3 回归模型表

浓度类型	数据类型	预测方程	R^2	F	F_a	s^2
CD4 浓度	M_1	$c_1 = 43.5456 + 5.9858d_1 - 0.0906d_1^2$	0.1997	98.944	3.853	5351.5
	M_2	$c_2 = 129.8522 + 7.0061d_2 - 0.1241d_2^2$	0.1599	43.5	3.862	6734.1
	M_3	$c_3 = 196.26 + 6.912d_3 - 0.1132d_3^2$	0.1550	26.038	3.874	8083.6
HIV 浓度	M_1	$h_1 = 458.1105 - 15.0890d_1 + 0.3010d_1^2$	0.2029	100.91	3.853	16387
	M_2	$h_2 = 425.5908 - 17.9384d_2 + 0.3761d_2^2$	0.2963	96.229	3.862	1285.9

	M_3	$h_3 = 411.35 - 17.595d_3 + 0.3466d_3^2$	0.3670	82.319	3.874	10527
$p = 0 < 0.0001$						

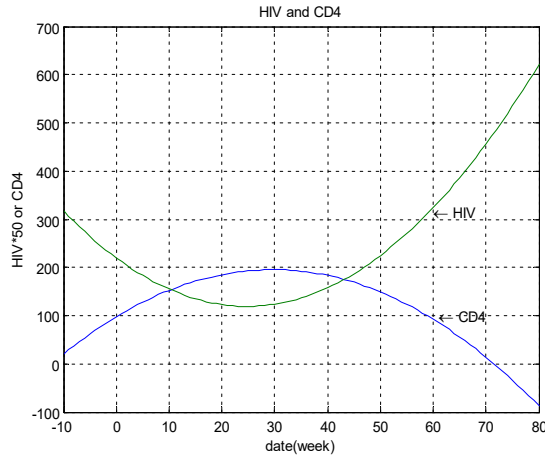


图2 第一类回归曲线

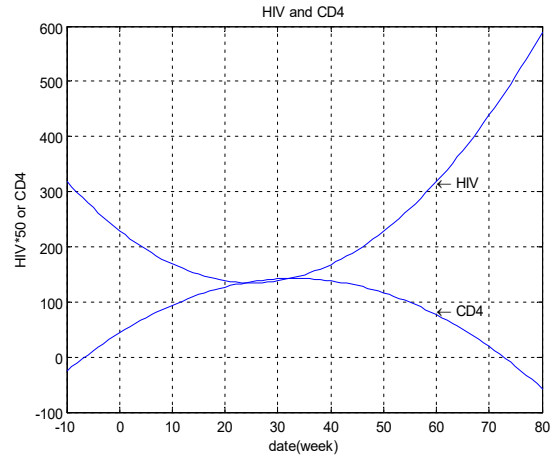


图3 第二类回归曲线

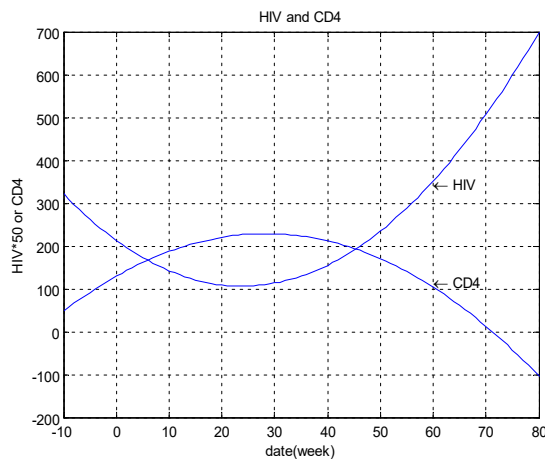


图4 第三类回归曲线

观察图2~图4，每一类CD4浓度和HIV浓度的回归曲线的变化趋势都大致相同，CD4浓度曲线都是先增大后减少，HIV浓度曲线都是先减少后上升。

问题结果及分析

从上面三个回归曲线图可以得到CD4浓度曲线最小值和HIV浓度曲线最大值所在的时刻如表4所示，如同模型一的分析，在HIV浓度降到最低且尚未反弹时就应当停止服药。由此可以判断不同类型病人都不宜在测试结束后还继续治疗，而是在服药到一定时间后终止，并且确定出最佳治疗终止时间分别为第25.1295周，第24.0000周，第25.0000周。

表 4 曲线驻点的横纵坐标值

曲线类型	第一类		第二类		第三类	
	浓度值	时刻	浓度值	时刻	浓度值	时刻
CD4 浓度曲线 (min)	124.4511	33.0000	228.7633	28.0000	301.0769	30.0769
HIV 浓度曲线 (max)	134.5057	25.1295	106.0407	24.0000	94.1240	25.0000

(三) 模型一与模型二对所得结果优劣评价

模型一中是对整体数据开始处理，初步得到治疗效果的评价，对所有病人都适用；模型二中是逐步细化将病人分类后得到三种评价结果，工作量稍微大点，但可作为模型一的参照检验，两种评价相差不大，这在一定程度上检验了结果的正确性。两种处理方法各有优缺点，相互补充。

4.2 问题二的分析及模型建立：

4.2.1 问题二的提出和数据分析

已知：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。

四种疗法的日用药分别为：

疗法一：600mg zidovudine 或 400mg didanosine，这两种药按月轮换使用；

疗法二：600 mg zidovudine 加 2.25 mg zalcitabine；

疗法三：600 mg zidovudine 加 400 mg didanosine；

疗法四：600 mg zidovudine 加 400 mg didanosine，再加 400 mg nevirapine。

要求：以 CD4 为标准评价 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。

分析：由于采用了四种不同的用药方案，且不同年龄的病人身体免疫能力、产生 CD4 细胞的能力各有不同，这将直接影响 CD4 细胞浓度的变化趋势。由此得出影响 CD4 细胞浓度的因素有两个——病人的年龄差异和治疗方案的不同（具体表现为治疗的时间的长短），问题可归结为多元回归模型。采用多元回归来拟合出 CD4 的浓度与病人的年龄和治疗的时间的交互式图形，曲线的变化趋势表达了人体内 CD4 的浓度的变化趋势，曲线的斜率表达了 CD4 减少或增加的速率，据此评价治疗方案的优劣及确定最佳治疗时间。

4.2.2 模型的建立及求解

(一) 建立多元回归模型：

将治疗方案依次记为 $j=1,2,3,4$ ，CD4 的浓度做简化处理，处理为 $\log(CD4\ count + 1)$ ，并记做 $y_j = \log(CD4\ count + 1)$ ，将第 j 种治疗方案的病人的年龄记为 a_j ($j=1,2,3,4$)，治疗的时间记为 d_j ($j=1,2,3,4$)。

由于二元回归方程存在多种形式，为更加精确的拟合出变化曲线，我们采用多元回归模型同时建立多个子模型拟合出 CD4 的浓度分别与病人的年龄和治疗的时间的交互式曲线，然后筛选出最佳模型作为它们各自的最终模型，作出最佳

模型对应的关系曲线^[1]。

1.) purequadratic 模型：包含线性项和纯二次项

$$y_j = \beta_0 + \beta_1 a_j + \beta_2 d_j + \beta_3 a_j^2 + \beta_4 d_j^2 + \varepsilon \quad (j = 1, 2, 3, 4)$$

2.) quadratic 模型：包含线性项和完全二次项

$$y_j = \beta_0 + \beta_1 a_j + \beta_2 d_j + \beta_3 a_j d_j + \beta_4 a_j^2 + \beta_5 d_j^2 + \varepsilon \quad (j = 1, 2, 3, 4)$$

3.) interaction 模型：包含线性项和纯交互项

$$y_j = \beta_0 + \beta_1 a_j + \beta_2 d_j + \beta_3 a_j d_j + \varepsilon \quad (j = 1, 2, 3, 4)$$

4.) linear 模型：只包含线性项

$$y_j = \beta_0 + \beta_1 a_j + \beta_2 d_j + \varepsilon \quad (j = 1, 2, 3, 4)$$

(二) 模型的求解：

用 MATLAB 统计工具箱中的 rstool 命令来处理，每个治疗方案都对应可输出四个交互式画面，其中每个交互式画面都在图左面提供了两个下拉式菜单，以便我们选择不同的子模型来分析曲线。将每种治疗方案的四个模型输出的回归系数 β 和剩余标准差 s 分别列入表 5～表 8 中：

表 5 治疗方案一与 CD4 浓度的模型输出

子模型名称	β_0	β_1	β_2	β_3	β_4	β_5	s
purequadratic	2.8293	0.0041	-0.0111	-0.0000	-0.0001		0.9310
quadratic	2.8747	0.0027	-0.0137	0.0001	-0.0000	-0.0001	0.9313
interaction	2.9047	0.0018	-0.0170	0.0001			0.9307
linear	2.8665	0.0028	-0.0144				0.9303

表 6 治疗方案二与 CD4 浓度的模型输出

子模型名称	β_0	β_1	β_2	β_3	β_4	β_5	s
purequadratic	1.4525	0.0565	-0.0060	-0.0004	-0.0002		1.0321
quadratic	1.5092	0.0548	-0.0092	0.0001	-0.0004	-0.0002	1.0325
interaction	2.1844	0.0212	-0.0158	0.0001			1.0326
linear	2.1252	0.0228	-0.0120				1.0322

表 7 治疗方案三与 CD4 浓度的模型输出

子模型名称	β_0	β_1	β_2	β_3	β_4	β_5	s
purequadratic	2.2972	0.0271	0.0041	-0.0002	-0.0003		1.1215
quadratic	2.3112	0.0267	0.0031	0.0000	-0.0002	-0.0003	1.1220
interaction	2.7134	0.0073	-0.0065	0.0000			1.1220
linear	2.6930	0.0079	-0.0051				1.1215

表 8 治疗方案四与 CD4 浓度的模型输出

子模型名称	β_0	β_1	β_2	β_3	β_4	β_5	s
purequadratic	3.0105	-0.0181	0.0331	0.0004	-0.0009		1.1390
quadratic	3.0732	-0.0195	0.0285	0.0001	0.0004	-0.0009	1.1394
interaction	2.6306	0.0099	-0.0032	0.0001			1.1464
linear	2.5666	0.0116	0.0008				1.1460

筛选最佳模型：分析它们各自的剩余标准差 s ，选取 s 最小的一项，如有两项相同且最小，即取较为简单的一个方程，得到每种治疗方案的最终模型，

治疗方案一选取 linear 模型：

$$y_1 = 2.8665 + 0.0028a_1 - 0.0144d_1$$

治疗方案二选取 purequadratic 模型：

$$y_2 = 1.4525 + 0.0565a_2 - 0.0060d_2 - 0.0004a_2^2 - 0.0002d_2^2$$

治疗方案三选取 linear 模型：

$$y_3 = 2.6930 + 0.0079a_3 - 0.0051d_3$$

治疗方案四选取 purequadratic 模型：

$$y_4 = 3.0105 - 0.0181a_4 + 0.0331d_4 + 0.0004a_4^2 - 0.0009d_4^2$$

四种治疗方案按各自对应的最终模型输出的交互式画面如图 5 至图 8 下：

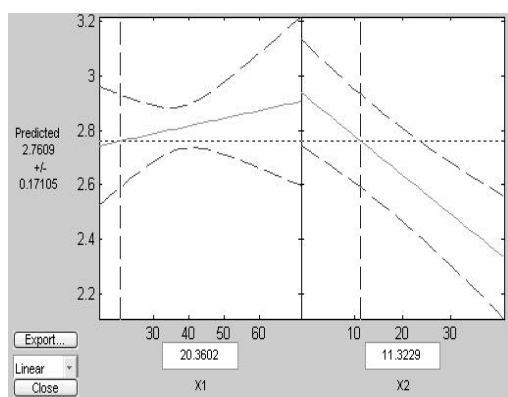


图 5 疗法一的交互式画面

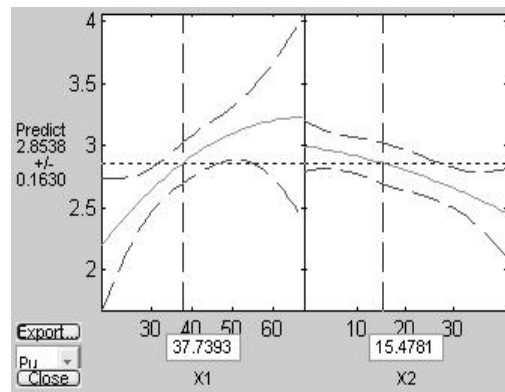


图 6 疗法二的交互式画面

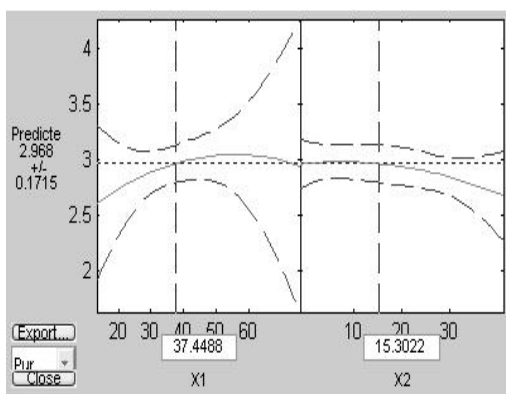


图 7 疗法三的交互式画面

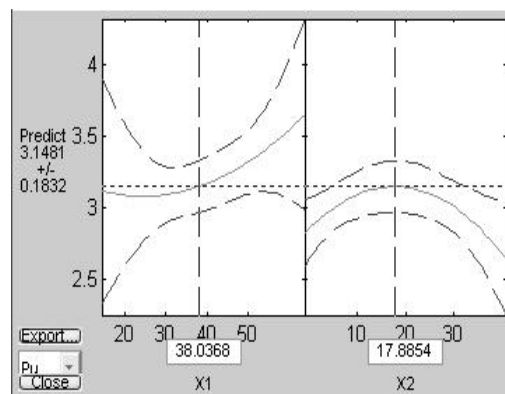


图 8 疗法四的交互式画面

注：以上四图中虚曲线为因变量的置信区间，实曲线为自变量与因变量的关系曲线。

图形分析：

图形最左边给出了 y_j 的预测值及其置信区间。

交互式画面左边一幅图形表示治疗时间 d_j 固定时 CD4 浓度 y_j 随病人年龄 a_j 变化的变化曲线。

四幅图形中左半部分曲线均呈增长的总趋势，说明在经过同一治疗时间后病人体内的 CD4 浓度会随病人年龄的增长而增加。

交互式画面右半边一幅图形表示病人年龄 a_j 固定时 CD4 浓度 y_j 随治疗时间 d_j 变化而变化的曲线。

(三) 四种疗法优劣的评价及预测结果

在图 5~图 8 中，在年龄一定的情况下，只有第四种疗法的 CD4 浓度的变化曲线随治疗时间的延迟而产生先增高后降低的趋势，其他各疗法随着治疗的进行 CD4 的浓度在持续降低，由于艾滋病治疗的目的，是尽量使人体内产生更多的 CD4，有效地降低 CD4 减少的速度，以提高人体免疫能力。所以第四种疗法最好，且最佳治疗终止时间应选在图形的最高点处，即确定 $d_4 = 17.8854$ 周为最佳治疗终止时间。

分析其他三种疗法曲线的走势，随着治疗的进行它们的 CD4 的浓度在持续降低，在此曲线斜率的绝对值表达了 CD4 降低的速率，斜率的绝对值越大，CD4 减少得越快。设 $k_j (j=1,2,3)$ 为第 j 种疗法曲线斜率的绝对值，存在以下关系：

$k_3 < k_2 < k_1$ ，说明随着治疗的进行，第三种疗法 CD4 减少的速度最慢，第二种疗法次之，第一种疗法减少的速度最快，从而确定优劣：疗法三>疗法二>疗法一。

综上可知，四种疗法的优劣依次如下（表 9）（编号由优至劣如 1 号为最好，4 号为最差）：

表 9 四种疗法优劣比较结果

优劣序号	疗法名称	最佳治疗终止时间
1.	疗法四	第 17.8854 周
2.	疗法三	
3.	疗法二	
4.	疗法一	

即选择疗法四：600 mg zidovudine 加 400 mg didanosine，再加 400 mg nevirapine 为最终方案，它的最佳终止时间为第 17.8854 周。

（四）方差分析

以上的回归分析是定性的判断四种疗法的优劣，下面我们来定量的讨论和检验四种不同的疗法对 CD4 浓度的影响，我们采用单因素方差分析方法处理。

由于附件 2 是将 1300 多名病人随机地分为 4 组，因此在此可以假设所有病人的 CD4 的初始浓度是相等的。由以上的回归模型得出四种中的前三种的 CD4 浓度几乎都是随治疗时间的增长而降低，只有第四种疗法的第 17 周附近出现了转折，因此对所有病人取具有代表性的第 14 至 19 周的 CD4 浓度作为观测值，再将所得观测值分别按对应的疗法分组，最后在 Matlab 里用命令 `anova1(X,group)` 对各疗法的优劣进行单因素方差分析并比较出优劣顺序。M 文件见附录 data203.m，结果如下^[4]：

$$P = 4.23923 \times 10^{-5}$$

并且得到两个图形界面（图 9、图 10）：

图 9 方差分析表中有 6 列：

第 1 列显示误差的来源；

第 2 列显示每一个误差来源的平方和(SS)；

第 3 列显示有误差来源相关的自由度(df)；

第 4 列显示均值平方和(MS)，它是误差来源平方和自由度的比值,即 SS/df；

第 5 列显示 F 统计量，它是均值平方和的比值；

第 6 列显示 p 值,p 值是 F 的函数(fcdf)；当 F 增加时 P 值减小。

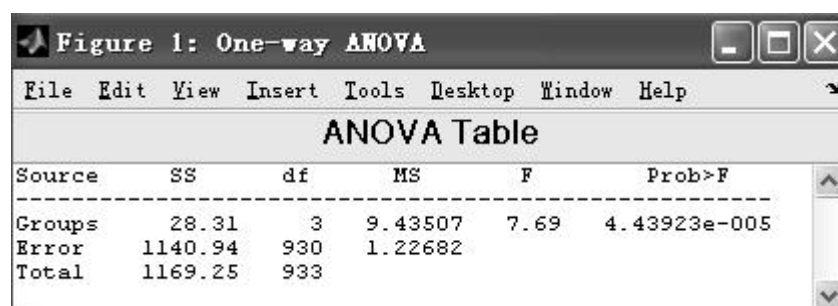


图 9 单因素方差分析表

图 10 box 图显示 X 的每组数据的箱形图。箱形图中心线上较大的差异对应于较大的 F 值和较小的 P 值。

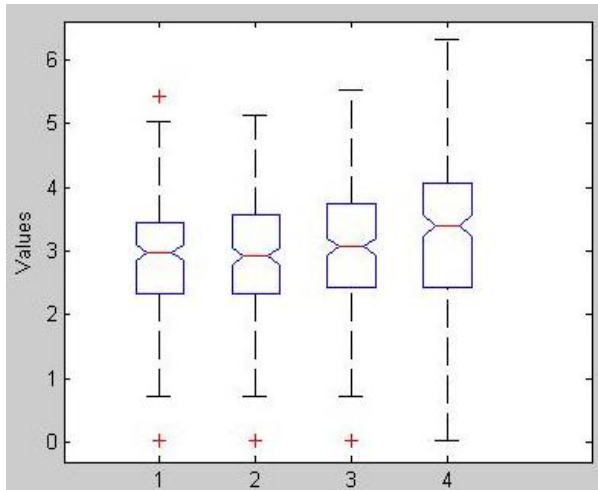


图 10

因为 $P = 4.23923 \times 10^{-5} < \alpha = 0.05$, 另一方面运用 Matlab 算得 F 值 $F_{0.05}(2, 930) = 3.0054$, 由方差分析表得 $F = 7.69 > F_{0.05}(2, 930)$, 所以认为不同的疗法对病人 CD4 浓度的提高有显著影响。

各水平的效应值如下：

$$\hat{\alpha}_1 = \bar{x}_1 - \bar{x} = 2.8112 - 2.9883 = -0.1771$$

$$\hat{\alpha}_2 = \bar{x}_2 - \bar{x} = 2.8597 - 2.9883 = -0.1286$$

$$\hat{\alpha}_3 = \bar{x}_3 - \bar{x} = 3.0147 - 2.9883 = 0.0264$$

$$\hat{\alpha}_4 = \bar{x}_4 - \bar{x} = 3.2566 - 2.9883 = 0.2683$$

计算结果表明，效应值 $\hat{\alpha}_4$ 最大。说明第四种疗法对病人 CD4 浓度的提高值最大，因此第四种疗法 $\hat{\alpha}_4$ 为最优疗法。

4.3 问题三的分析及模型建立：

4.3.1 问题三的提出和数据分析

已知： 艾滋病药品的主要供给商对不发达国家提供的各种药品价格不尽相同。具体如下：600mg zidovudine 1.60 美元/天，400mg didanosine 0.85 美元/天，2.25 mg zalcitabine 1.85 美元/天，400 mg nevirapine 1.20 美元/天。

要求： 若病人需考虑 4 种疗法的费用，对（2）中的评价和预测（或者提前终止）将产生什么改变。

分析： 在（2）中我们在讨论疗法的优劣及预测较优疗法继续治疗的效果时，只以 CD4 的浓度作为参考标准，在此应综合考虑治疗费用和 CD4 的浓度大小值，首先建立治疗费用与治疗时间的函数关系，然后采用多元回归模型来拟合出 CD4 的浓度与病人的年龄和治疗费用的交互式图形，再建立一个疗效性价比模型来进行分析。

在此提出一个疗效性价比的概念，定义：

$$\text{疗效性价比} = \text{CD4 变化的浓度} / \text{治疗费用}$$

为简化分析，我们采用比较在花费相同费用时四种疗法 CD4 浓度的变化程度，评价四种疗法的优劣并做预测。

4.3.2 模型的建立及求解

(一) 建立模型：

将第 j 种疗法的治疗费用记为 f_j ($j=1,2,3,4$)，第 j 种疗法的疗效性价比记为 p_j ($j=1,2,3,4$)，采用第 j 种疗法时费用增量为 Δf_j ($j=1,2,3,4$) 时 CD4 的浓度增量记为 Δy_j ($j=1,2,3,4$)，治疗时间记为 d_j ($j=1,2,3,4$)。在采用第一种疗法时，记停止治疗时为奇数月份时总的治疗费用为 s_1 ，记停止治疗时为偶数月份总的治疗费用为 s_2 。记：

$$w_1 = \begin{cases} 1 & \text{该月为治疗的第奇数个月份} \\ 0 & \text{该月为治疗的第偶数个月份} \end{cases}, w_2 = \begin{cases} 1 & \text{该月为治疗的第偶数个月份} \\ 0 & \text{该月为治疗的第奇数个月份} \end{cases}。$$

第一步：根据病人日用药量、药品使用方案及药品价格确定治疗费用 f_j ($j=1,2,3,4$) 与治疗时间 d_j ($j=1,2,3,4$) 的线性关系如下：

为使治疗费用尽可能少，假设采用第一种疗法时第一个月服用的是每 400mg 0.85 美元的 didanosine，第二个月服用的是每 600mg 1.60 美元的 zidovudine。

$$1.) s_1 = (7d_1 - 30 \times \left\lceil \left(\frac{7d_1}{30} \right) / 2 \right\rceil) \times 0.85 + \left\lfloor \left(\frac{7d_1}{30} \right) / 2 \right\rfloor \times 30 \times 1.60$$

$$s_2 = (7d_1 - 30 \times \left\lceil \left(\frac{7d_1}{30} \right) / 2 \right\rceil) \times 1.60 + \left\lfloor \left(\frac{7d_1}{30} \right) / 2 \right\rfloor \times 30 \times 0.85$$

$$w_1 = \left\lceil \left(\frac{7d_1}{30} \right) / 2 \right\rceil \text{ 求余}$$

$$w_2 = 1 - w_1$$

$$f_1 = s_1 \times w_1 + s_2 \times w_2$$

$$2.) f_2 = 7d_2 \times (1.60 + 1.85)$$

$$3.) f_3 = 7d_3 \times (1.60 + 0.85)$$

$$4.) f_4 = 7d_4 \times (1.60 + 0.85 + 1.20)$$

第二步：为简化计算，将 CD4 的浓度处理为 $\log(\text{CD4 count} + 1)$ ，并记做 $y_j = \log(\text{CD4 count} + 1)$ 。

我们以病人的年龄 a_j 和病人的治疗费用 f_j 作为自变量，以 CD4 的浓度 y_j 作为因变量，利用多元回归分析拟合出 CD4 的浓度 y_j 与病人的年龄 a_j 和病人的治疗费用 f_j 的交互式图形，输出每种治疗方案的四个模型输出的回归系数 β 和剩余标准差 s ，选取 s 最小的一个方程，如有两项最小，则取方程较为简单的一个，得到每种治疗方案的最终模型。（具体做法同问题二的处理）。

各种疗法的模型如下[1]:

治疗方案一选取 purequadratic 模型:

$$y_1 = 3.0105 - 0.0181a_1 + 0.0013f_1 + 0.0004a_1f_1$$

治疗方案二选取 quadratic 模型:

$$y_2 = 1.4525 + 0.565a_2 - 0.0002f_2 - 0.0004a_2f_2$$

治疗方案三选取 linear 模型:

$$y_3 = 2.8305 + 0.0042a_3 - 0.0013f_3$$

治疗方案四选取 quadratic 模型:

$$y_4 = 2.2972 + 0.0271a_4 + 0.0002f_4 - 0.0002a_4f_4$$

四种治疗方案对应的最终模型的交互式画面如下:

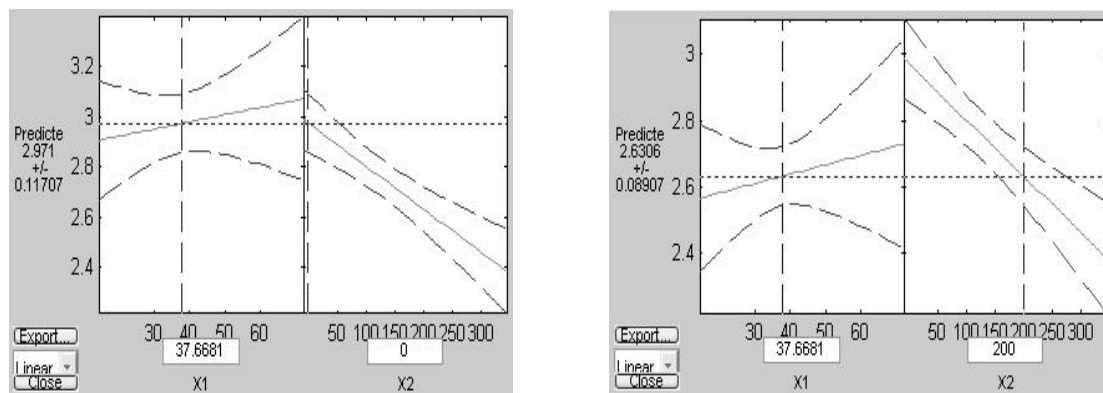


图 11 疗法一的交互式画面（左幅为费用 0 美元时，右幅为费用 200 美元时）

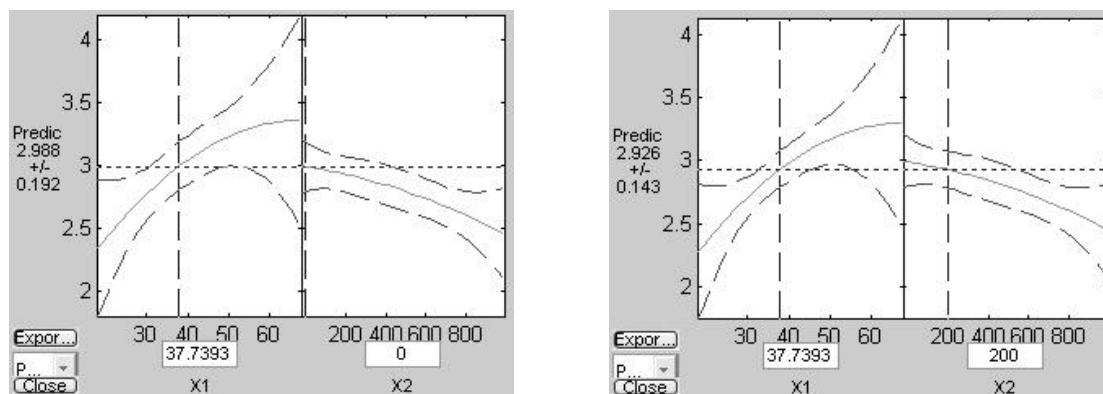


图 12 疗法二的交互式画面（左幅为费用 0 美元时，右幅为费用 200 美元时）

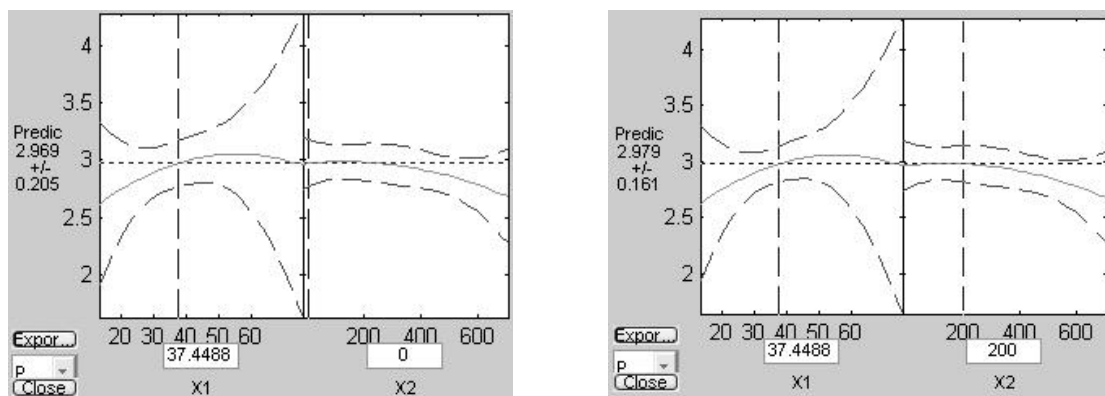


图 13 疗法三的交互式画面（左幅为费用 0 美元时，右幅为费用 200 美元时）

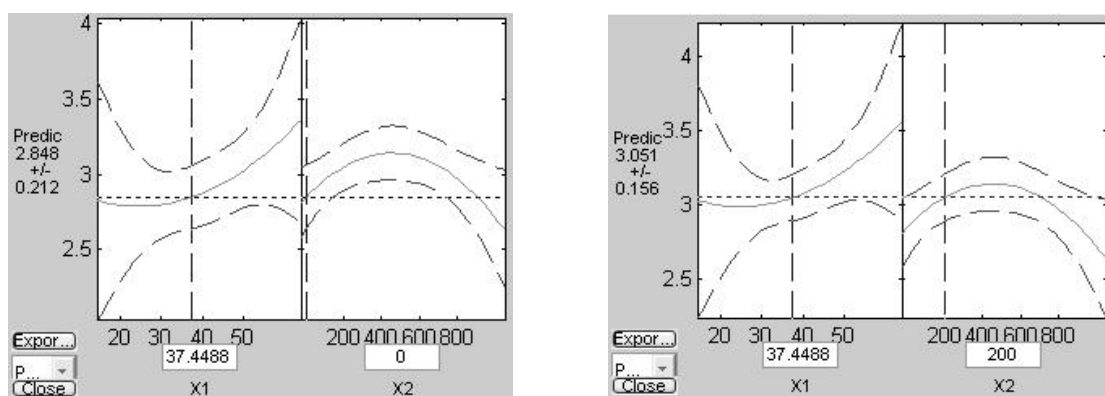


图 14 疗法四的交互式画面（左幅为费用 0 美元时，右幅为费用 200 美元时）

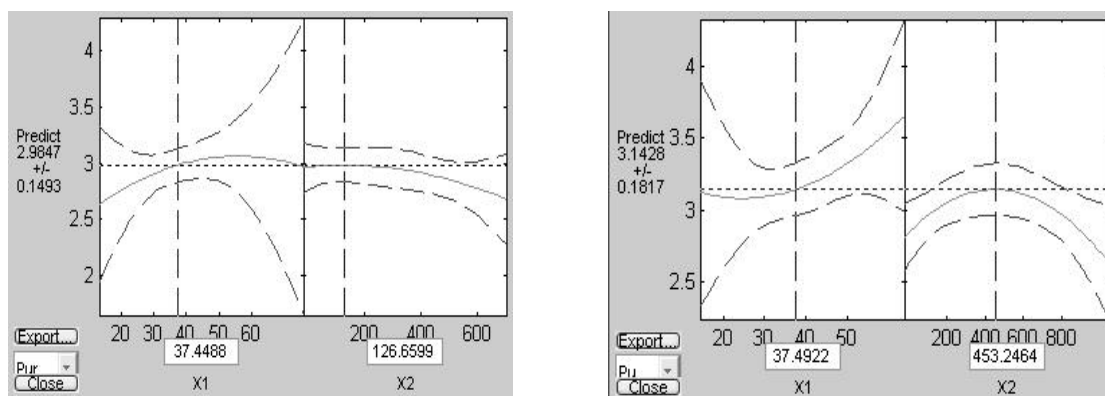


图 15 疗法三、四 CD4 浓度达最大值的交互式画面（左为方案三，右为方案四）

第三步：根据疗效性价比的定义，建立一个关于四种不同疗法的疗效性价比

模型：
$$p_j = \frac{\Delta y_j}{\Delta f_j} \quad (j=1,2,3,4)$$

花费同样多的治疗费用，疗效性价比越高，单位费用的治疗的效果越明显，则治疗效果越理想，越明显。以此作为判断四种疗法优劣的依据。

第四步：判断治疗效果

首先分析四种疗法的交互式画面的左边，当治疗费用 f_j 固定时，病人年龄 d_j 与病人体内 CD4 浓度 y_j 的变化曲线均呈增长的总趋势，说明在经过同一治疗时间后病人体内的 CD4 浓度会随病人年龄的增长而增加。

然后分析四种疗法的交互式画面的右边，前两种疗法的图形均呈下降趋势，第三、四种疗法的图形呈抛物线状，故当选择第三、四种疗法时，应在 CD4 浓度尚未达到最高或者达到最高的时刻停止治疗。第三、四种疗法 CD4 浓度达到最高值时费用分别为：126.6599 和 453.24 美元，故至少应在费用达到 453.24 美元时停止治疗。

在治疗费用在 0~453.24 美元上分析：

1. 读图分析：疗效性价比的正负；
2. 分别取治疗费用为 0 和 200 美元的两个时刻（读图 11~图 14），记录 CD4 的浓度，并计算这个区间内的疗效性价比。

表 10 取 0、200 美元分析疗效性价比

方案名称	疗效性价比的正负（读图判断）	费用为 0 时 CD4 浓度	费用为 200 美元时 CD4 浓度	疗效性价比 ($\times 10^{-4}$)
方案一	负	2.971	2.6306	-17.020
方案二	负	2.988	2.926	-3.100
方案三	先正后负	2.969	2.979	0.500
方案四	先正后负	2.848	3.015	0.835

注：疗效性价比的正负要求图形上每一点都满足，而不是只有某两点符合。

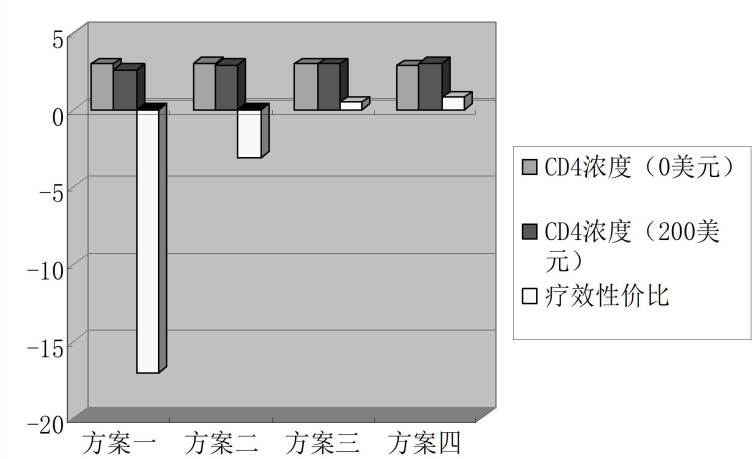


图 16 取 0、200 美元分析疗效性价比的直方图

注：图表中每种的第一、二个柱状图表示的分别为治疗费用为 0、200 美元时 CD4 的浓度（无单位），第三个柱状图表示的是此时的疗效性价比（ $\times 10^{-4}$ ）。

分析：读直方图可知，治疗方案三、四优于方案一、二。而对于前两种治疗方案，很明显方案一的疗效性价比要比方案二的小的多。

故第二种方案优于第一种方案。

在图上，疗效一和疗法二疗效性价比为负，说明采用这两种治疗方案治疗，不但没有使 CD4 的浓度升高，反倒使其降低，药物治疗起了负作用，或者带来负面影响。

以下在第三、四种疗法 CD4 浓度达到最高值时分别做疗效性价比分析。

表 11 取最值点分析疗效性价比

方案名称	费用为 0 美元时刻 CD4 浓度	费用为 126.6599 美元时刻 CD4 浓度	疗效性价比 ($\times 10^{-4}$)	费用为 453.24 美元时刻 CD4 浓度	疗效性价比 ($\times 10^{-4}$)
方案三	2.969	2.9847	1.240	/	/
方案四	2.848	/	/	3.1428	6.504

由于在两种疗法的各自的最值点处疗效性价比 $6.504 > 1.240$ ，故第四种方案优于第三种方案。

第五步：四种方案的优劣比较结果及相关信息，如表 12 所示。

表 12 方案优劣表

方案优劣顺序	方案名称	最优方案终止治疗时间（周次）	最优方案终止治疗时共花费治疗费用（美元）
1.	方案四	17.8854	453.2464
2.	方案三		
3.	方案二		
4.	方案一		

选择疗法四：600 mg zidovudine 加 400 mg didanosine，再加 400 mg nevirapine 为最终方案，最佳终止时间为第 17.8854 周，终止治疗时共花费 453.2464 美元。

五、模型的评价

5.1 模型的优点：

1. 分析附件 1 CD4 和 HIV 浓度数据的时候，先从整体数据入手，对所有数据进行统一处理，然后再细化，将数据进行分类处理，由两种方法分别预测最佳治疗终止时间。这种逐步处理方法简洁易懂，给出不同的解法，使我们的评价预测具有很强的可比性和可读性，也更严密。

2. 评价问题二 4 种疗法的优劣时，建立了回归模型并用 Matlab 统计工具箱处理，定性的作出评价，同时我们还运用方差分析作了定量评价，两种评价得出的结果是一致的。

3. 问题二三中利用 Matlab 统计工具箱命令来处理，建立了多个多元回归子模型，在交互式画面上选择变量，这样逐步筛选，使模型达到最优，提高了精确度。

4. 在问题三的解答中, 我们提出了一个“性价比”的概念, 建立“疗效性价比”模型, 使抽象问题具体化、量化, 既有效完美地结合了 CD4 的浓度和治疗费用, 又便于计算和比较。

5.2 模型的缺点:

由于所给两组数据非常庞大, 考虑到时间问题, 为了避免复杂而繁琐的工作量, 在作回归分析时, 我们没有完全对残差图中的一些异常点进行剔除处理。但是回归分析得到的结果都经过了各种统计检验, 因而总的回归效果还是很好的。

六、模型的改进和推广

1. 对问题作回归分析并建立回归模型时, 若进行残差分析, 作出残差及其置信区间分析图, 对原始数据中的异常数据剔除处理后再计算, 模型会有一些改进。

2. 我们在附件 1 中随机抽取一些病人数据分成几组, 作出 CD4 浓度 (或 HIV 浓度) 与测试时刻的关系折线图, 可以看到, 折线图大致有两种不同的走势, 一种是 CD4 浓度随着时间的增加而呈现上升的趋势, 最后浓度稳定在最高点, 另一种是 CD4 浓度随着时间的增加先是上升, 过一段时间后转而下降。这说明不同的病人进行药物治疗后大致有两种不同的效果, 一种是服药后病情都在好转, 另一种是服药后前一段时间有效果, 随后又变坏。若对问题一继续作深入讨论, 可以对附件 1 CD4 和 HIV 浓度数据作更合理的调整分类, 分类依据是药物治疗后 CD4 浓度变化趋势, 将这两类病人用**聚类分析**的方法分离出来, 得到两组相对比较类似的数据, 再分别进行回归分析处理后, 各自会预测出的治疗效果应该会不同。前一种预测应该是建议继续治疗, 即在测试终止后继续服药, 后一种预测应该是在测试期间当病人的 CD4 浓度值上升到最大时即终止服药。这种预测结论更加贴近现实。由于时间紧迫, 我们没有具体进行操作, 在此只作为模型的推广大致讨论了一下。

参考文献

[1] 姜启源, 邢文训, 谢金星, 杨顶辉 大学数学实验, 北京: 清华大学出版社, 2000

[2] 马知恩 等著, 传染病动力学的数学建模与研究, 北京: 科学出版社, 2004

[3] 傅鹏, 龚劬, 刘琼荪, 何中市编著, 数学实验, 北京: 科学出版社, 2000

[4] (美) 埃维森 (Gudmund R. Iversen) 等著; 吴喜之 等译, 统计学, 北京: 高等教育出版社, 2000

[5] 盛骤 等编, 概率论与数理统计 (第三版), 北京: 高等教育出版社, 2001

[6] 苏金明等编, MATLAB 工具箱应用

<http://sshtm.ssreader.com/html/pages/showbook.asp?ssid=11193564>,

2006 年 9 月 17 日

八、 附录

8.1 程序代码:

8.1.1 问题一的程序代码

注: 程序中 data1 为附件 1 数据的矩阵表示, 由于篇幅过大, 在此略去

```
n=size(data1);
[ra,ir]=sort(data1(:,2));
W=[ra,ir];
k=1;
for i=1:n(1)
    % if (data1(i,2)<=1)&data1(i,3)>75&data1(i,3)<=150
    if (data1(i,2)<=1)&data1(i,3)>150
        % if (data1(i,2)<=1)&data1(i,3)>=0&data1(i,3)<=75
        s(k)=i;
        k=k+1;
    end
end
s';
M=data1(s',:);
j=0;
for i=2:n(1)
    if data1(i,1)-data1(i-1,1)==1
        j=j+1;
    end
end
n2=size(M);
k=1;
for i=1:n2(1)
    for j=1:n(1)
        if data1(j,1)==M(i,1)
            ss(k)=j;
            k=k+1;
        end
    end
end
ss;
data2=(data1(ss,:));
x1=data2(:,2);
x2=x1.^2;
y=data2(:,3);

x=[x1,x2];
```

```

X=[ones(size(x1),1),x1,x2];
%stepwise(x,y);
[b1,bint1,r1,rint1,s1]=regress(y,X);%regress of cd4 after sort
xx=-10:80;
yy=b1(1)+b1(2)*xx+b1(3)*xx.^2;yyn=max(yy);
xxn=(sqrt(b1(2)^2-4*b1(3)*(b1(1)-yyn))-b1(2))/(2*b1(3));
plot(xx,yy)
max(yy)
%plot(data2(:,2),data2(:,3),'.');

hold on
x3=data2(:,4);
x4=data2(:,4).^2;
y2=data2(:,5)*50; %enlarge y2 to '50*y2' so it
%will be clearer when plotted with yy;
xs=[x3,x4];
Xs=[ones(size(x3),1),x3,x4];
%stepwise(x,y);
[b2,bint2,r2,rint2,s2]=regress(y2,Xs);%regress of hiv after sort
xx=-10:80;
yy=b2(1)+b2(2)*xx+b2(3)*xx.^2;yyn=min(yy);
xxn2=(sqrt(b2(2)^2-4*b2(3)*(b2(1)-yyn))-b2(2))/(2*b2(3));
plot(xx,yy)
min(yy)
grid
text(60,320,'\leftarrow HIV');
text(60,105,'\leftarrow CD4');
title('HIV and CD4');
xlabel('date(week)');
ylabel('HIV*50 or CD4');
hold off

xt1=data2(:,3);xt2=xt1.^2;xt3=1./(xt1+1);yt=data2(:,5);
xt=[xt1,xt2];
Xt=[ones(size(xt1),1),xt1,xt2,xt3];
%stepwise(xt,yt);
[b3,bint3,r3,rint3,s3]=regress(yt,Xt);
xx=0.0000001:300;
yyt=b3(1)+b3(2)*xx+b3(3)*xx.^2+b3(4)./(xx+1);

figure(2);

plot(xx,yyt)
hold on

```

```

plot(data2(:,3),data2(:,5),'.')
hold off
xxn %date when cd4 reach max(cd4)
xxn2 %date when hiv reach min(hiv)

figure(3);

bx1=data1(:,2);bx2=bx1.^2;by=data1(:,3);
bX=[ones(size(bx1),1),bx1,bx2];
bx=[bx1,bx2];
%stepwise(bx,by)
[b4,bint4,r4,rint4,s4]=regress(by,bX);%regress of cd4 with no sort
xx=-10:80;
yy1=b4(1)+b4(2)*xx+b4(3)*xx.^2;

bx1=data1(:,4);bx2=bx1.^2;by=data1(:,5)*50;
bX=[ones(size(bx1),1),bx1,bx2];
bx=[bx1,bx2];
%stepwise(bx,by)
[b5,bint5,r5,rint5,s5]=regress(by,bX);%regress of hiv with no sort
yy2=b5(1)+b5(2)*xx+b5(3)*xx.^2;
%plot(xx,yy1,xx,yy2,bx1,by, '.')
plot(xx,yy1,xx,yy2)
grid
text(60,320,'\leftarrow HIV');
text(61,103,'\leftarrow CD4');
title('HIV and CD4');
xlabel('date(week)');
ylabel('HIV*50 or CD4');
yyn=max(yy1)
xxn3=(sqrt(b4(2)^2-4*b4(3)*(b4(1)-yyn))-b4(2))/(2*b4(3));
yyn=min(yy2)
xxn4=(sqrt(b5(2)^2-4*b5(3)*(b5(1)-yyn))-b5(2))/(2*b5(3));
xxn3
xxn4

```

8.1.2 问题二回归分析的程序代码

```

size(data201);
[ra,ir]=sort(data201(:,2));
D=[ra,ir];
for k=1:4
j(k)=0;
for i=1:5036
if D(i,1)==k

```



```

        j(k)=j(k)+1;
    end
end
end
X1=data201(D(1:j(1),2),:);

X2=data201(D(j(1)+1:j(1)+j(2),2),:);

X3=data201(D(j(1)+j(2)+1:j(1)+j(2)+j(3),2),:);

X4=data201(D(j(1)+j(2)+j(3)+1:j(1)+j(2)+j(3)+j(4),2),:);

data1=X4;
n=size(data1);
[ra,ir]=sort(data1(:,2));
W=[ra,ir];
k=1;
for i=1:n(1)
    %if (data1(i,4)<=1&data1(i,5)>3&data1(i,5)<4.5
    if (data1(i,4)<=1)&data1(i,5)<3
        %if (data1(i,4)<=1&data1(i,5)>4.5
        s(k)=i;
        k=k+1;
    end
end
s';
M=data1(s',:);
j=0;
for i=2:n(1)
    if data1(i,1)-data1(i-1,1)==1
        j=j+1;
    end
end
n2=size(M);
k=1;
for i=1:n2(1)
    for j=1:n(1)
        if data1(j,1)==M(i,1)
            ss(k)=j;
            k=k+1;
        end
    end
end
ss;

```

```

data2=(data1(ss,:));
x1=data1(:,3);
x2=x1.^2;
x3=data1(:,4);
x4=x3.^2;
x5=x1.*x3;
y=data1(:,5);
xx=[x1,x2,x3,x4,x5];
x=[x1,x3];
rstool(x,y,'purequadratic');

```

8.1.3 问题二方差分析的程序代码

```

n=size(data201);
k=1;
for i=1:n(1)
    if data201(i,4)>=22&data201(i,4)<=26
        data20(k,:)=data201(i,:);
        k=k+1;
    end
end
data20;
n2=size(data20);

%for j=1:4
    k=1;
    for i=1:n2(1)
        if data20(i,2)==1
            data21(k,:)=data20(i,:);
            k=k+1;
        end
    end
end
X1=data21;

clear data21;
k=1;
for i=1:n2(1)
    if data20(i,2)==2
        data21(k,:)=data20(i,:);
        k=k+1;
    end
end
X2=data21;

```

```

clear data21;
    k=1;
for i=1:n2(1)
    if data20(i,2)==3
        data21(k,:)=data20(i,:);
        k=k+1;
    end
end
X3=data21;

clear data21;
    k=1;
for i=1:n2(1)
    if data20(i,2)==4
        data21(k,:)=data20(i,:);
        k=k+1;
    end
end
X4=data21;

n3=size(data21)
%s(j)=
sum(data21(:,5))/n3(1)
X=[X1(:,5);X2(:,5);X3(:,5);X4(:,5)];
group=[ones(140,1);
    ones(155,1)*2;
    ones(140,1)*3;
    ones(151,1)*4];

anova1(X,group)

```

8.1.3 问题三的程序代码

```

[ra,ir]=sort(data201(:,2));
D=[ra,ir];
for k=1:4
    j(k)=0;
for i=1:5036
    if D(i,1)==k
        j(k)=j(k)+1;
    end
end
end
end

```

```

X1=data201(D(1:j(1),2),:);
x14=X1(:,4);
t=x14;    a1=0.85;    a2=1.60;
yw1=(t*7-floor(ceil(t*7/30)/2)*30)*a1+floor(ceil(t*7/30)/2)*30*a2;
yw2=(t*7-floor(ceil(t*7/30)/2)*30)*a2+floor(ceil(t*7/30)/2)*30*a1;
w1=rem(ceil(t*7/30),2);
w2=1-w1;
yw=yw1.*w1+yw2.*w2;

x11=X1(:,1);x12=X1(:,2);x13=X1(:,3);x14=X1(:,4);x15=X1(:,5);x16=yw;
XX1=[x11,x12,x13,x14,x15,x16];

X2=data201(D(j(1)+1:j(1)+j(2),2),:);
x21=X2(:,1);x22=X2(:,2);x23=X2(:,3);x24=X2(:,4);x26=X2(:,4)*(1.60+1.85)*7;
x25=X2(:,5);
XX2=[x21,x22,x23,x24,x25,x26];

X3=data201(D(j(1)+j(2)+1:j(1)+j(2)+j(3),2),:);
x31=X3(:,1);x32=X3(:,2);x33=X3(:,3);x34=X3(:,4);x36=X3(:,4)*(1.60+0.85)*7;
x35=X3(:,5);
XX3=[x31,x32,x33,x34,x35,x36];

X4=data201(D(j(1)+j(2)+j(3)+1:j(1)+j(2)+j(3)+j(4),2),:);
x41=X4(:,1);x42=X4(:,2);x43=X4(:,3);x44=X4(:,4);
x46=X4(:,4)*(1.60+0.85+1.20)*7;
x45=X4(:,5);
XX4=[x41,x42,x43,x44,x45,x46];

data1=XX4;
n=size(data1);
j=1;
ss(j)=data1(1,1);
for i=2:n(1)
    if data1(i,1)-data1(i-1,1)>=1
        j=j+1;
        ss(j)=data1(i,1);
    end
end
M=ss';
n2=size(M);

x1=data1(:,3);
x2=data1(:,6);
y=data1(:,5);

```

```
x=[x1,x2];  
rstool(x,y,'purequadratic')
```