

学 院： 数学与统计学院

专 业： 数学与应用数学

年 级： 2 0 0 5 级

姓 名： 程 为 民

学 号： 222005070101092

# 艾滋病疗法的评价及疗效的预测

**摘 要** 本文通过对大量艾滋病患者用药情况调查数据的分析处理，建立了数据的统计分析及推断模型，对不同数据按照实际情况进行合理分类。根据建立的模型比较艾滋病各种不同疗法的优劣，并通过多项式拟合的方法做出药物治疗的拟合曲线以预测其治疗的最佳终止时间以及疗效。最后综合病人自身经济条件，利用图像选择出一种疗效佳、成本又低的疗法，实际可行性很高！接着还对模型的稳定性、误差作了深入的分析，评价，并且做出细致的灵敏度分析，还有对模型优缺点的分析讨论也十分客观、实际，据此做出的改进使模型的预测、评价结果更准确，更科学！

本文还适当利用图形进行比较，增强问题分析解说的力度和真实准确性！

**关键词** 多项式拟合 统计分析 分类处理

## 一 问题的提出

当前人类社会最严重的瘟疫之一（艾滋病）。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（英文简称 HIV）引起的。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。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。如何有效防治及治疗艾滋病已经成为社会关注的焦点。

首先让我们了解一下治疗机理：

人类免疫系统的 CD4 细胞担任着“门卫”的工作。在抵御 HIV 的入侵中有重要作用，当 CD4 被 HIV 感染而裂解时，CD4 数量会急剧减少，HIV 将迅速增加，导致 AIDS 发作。因此艾滋病的治疗方向：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。人类很需要寻找一个疗效佳、成本低的 AIDS 疗法。在寻找疗法的过程中，我们需要根据收集到的数据统计分析来预测继续治疗的效果及确定最佳治疗终止时间。并对目前常用的几种疗法进行比较分析、评价优劣，并得出成本与疗效兼优的疗法。

## 二 问题分析

该问题是一个通过实验数据，预测 AIDS 疗法治疗效果和寻找最佳治疗终止时间的问题。在对数据进行分析后，考虑到数据量大，我们决定采用平均值为基本数据进行拟合。

对附件一的数据进行处理：去除其中的不完全数据，计算出在各周 CD4 和 HIV 浓度的平均增量，从而得到两组数据，进而用 Matlab 软件进行多项式拟合处理，根据得到的拟合曲线进行预测和分析，继而确定最佳治疗终止时间。

对附件二的数据进行处理：有些病人的测量只有初始数据，因而看不出疗效，对于这类数据我们予以排除，这样可以避免造成较大的误差。然后再对四种疗法分别按照不同年龄段（30 岁以下，30—40，40 岁以上）进行分类统计处理，计算出各组病人在各测试周体内 CD4 的平均增量，得到 12 组数据，对其按年龄段分为三组，运用 Matlab 软件进行多项式拟合处理，根据得到的拟合曲线比较各种疗法疗效的优劣，并对较优的疗法预测继续治疗的效果，确定最佳治疗终止时间。

最后结合各种疗法的价格对这四种疗法进行综合评估。

## 三 模型基本假设及说明

1. 我们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！

2. 数据处理时，排除掉了极个别的不完全数据，由于原始数据的随机性，因此并不影响用于计算的数据的合理性。

3. 在预测药物疗效的过程中，只以 CD4 或 HIV 的浓度为标准，而不考虑其它因素的影响。

4. 在比较四种疗法，评价优劣，对附件二进行数据处理时，将第  $8n$  周及其前后相差一周左右所测得的数据统一看作是在第  $8n$  周测得的。（ $n=0、1、2、3、4$ ）

5. 附件二中，同一年龄段病人的身体素质视为在同一水平线。

6. 考虑药品价格时，不考虑由于不同地区的运输而造成的价格差异和改变。

7.治疗过程中，不考虑病人因别的原因服用其它药物是否会影响测试结果。

#### 四 模型的建立及求解

问题一：

将附表 1 中的不全数据去除，统计出在 0 周和 3 周都有数据的病人，用这些病人在第 3 周的 CD4 的平均值减去在第 0 周的值，得出 CD4 平均增量 Z3，用其在第 0 周的 HIV 平均值减去在第 3 周的平均值，得出 HIV 的平均减少量 J3。再统计出在第 0 周和第 4 周都有数据的病人，用这些病人在第 4 周的 CD4 的平均值减去其在第 0 周的平均值，得出平均增量 Z4，用其在第 0 周 HIV 的平均值减去在第 4 周的平均值，得到 HIV 平均减少量 J4…这样重复下去,便得到两组数据:Z3, Z4, Z5…Z42 和 J3, J4, J5…J42.(若某组病人数量少于 10 人，则不

记这组数据)。数据为：

时间（周）	CD4 增量平均值	HIV 增量平均值
0	0	0
3	43.24528302	1.898
4	49.74672489	1.834361233
5	46.31147541	1.929787234
7	64.98360656	1.984745763
8	68.61538462	2.114720812
9	76.37846154	2.259615385
22	83	1.8
23	110.96875	2.249180328
24	96.38405797	2.26
25	84.59322034	2.483636364
26	56.53846154	1.385
38	82.6	2.064285714
39	130.3461538	2.225531915
40	113.6276596	2.374698795
41	100.7941176	1.795454545

考虑到 CD4 增量平均值比相应的 HIV 增量大很多倍，作出图形后不便于观察两者的关系，故做出如下处理：

将每个 CD4 增量平均值分别除以其增量平均值的最大值，记为 A，（用同样的方法处理 HIV 增量平均值，记为 B），得到相应的表格：

时间（周）	A	B
0	0	0
3	0.764202	0.331772
4	0.738579	0.381651
5	0.777001	0.355296
7	0.799129	0.498546
8	0.851462	0.526409
9	0.909801	0.585966
22	0.724744	0.636766
23	0.9056	0.851339

24	0.909956	0.739447
25	1	0.648989
26	0.55765	0.433756
38	0.831155	0.633697
39	0.896078	1
40	0.956138	0.871737
41	0.722914	0.77328

再以这两组数据的代号的数字为自变量，以这两组数据为函数值进行多项式拟合。  
以下为所编程序：

```
x=[3 4 5 7 8 9 22 23 24 25 26 38 39 40 41];
y1=[0.764202 0.738579 0.777001 0.799129 0.851462 0.909801 0.724744 0.9056
0.909956 1 0.5765 0.831155 0.896078 0.956133 0.722914];
y2=[0.331772 0.381651 0.355296 0.498546 0.526409 0.585966 0.636766 0.851339
0.739447 0.648989 0.433756 0.633697 1 0.871737 0.77328];
q5=polyfit(x,y2,5),p5=polyfit(x,y1,5);
disp('5次拟合'),poly2str(q5,'x')
poly2str(p5,'x');
xx=3:0.1:47;
y25=polyval(q5,xx),y15=polyval(p5,xx);
plot(x,y1,'r-*',x,y2,'g-*',xx,y25,'k-.',xx,y15);
```

运行此程序所得图像如下：

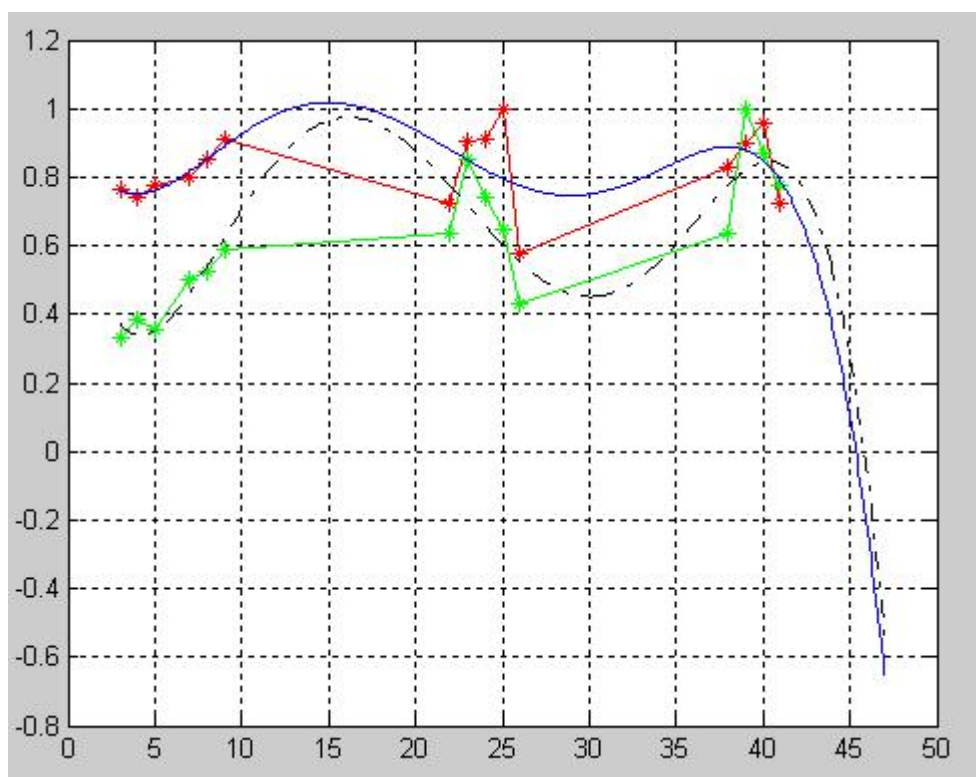


图 1-1

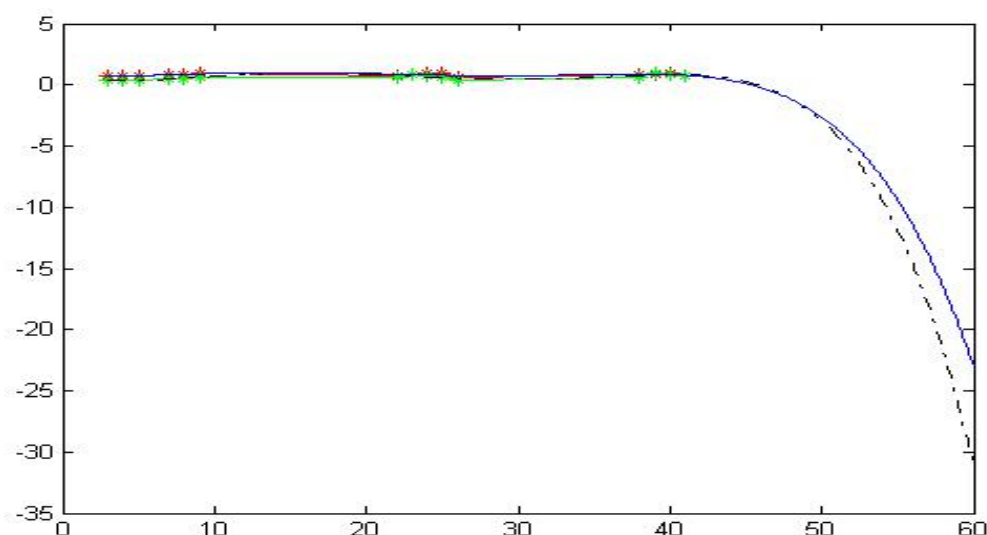


图 1-2

说明：红色实线表示 CD4 原始数据连线图象，蓝色实线表示其 5 次拟合曲线；  
绿色实线表示 HIV 原始数据连线图象，黑色虚线表示其 5 次拟合曲线。

分析：由图像 1-1 易知两条拟合曲线呈正相关，可以看出在约 40 周以后，CD4 的增加量减小，且减小速度越来越快；同时 HIV 的减少量越来越少，且减小速度越来越快。由图 1-2 看出：在 40 周以后图像不再有上升趋势。所以 40 周以后药物疗效逐渐变差甚至完全失去疗效。由此可以确定最佳治疗终止时间为 40 周左右。

**问题二：**将附表二中，数据少与三组的病人的数据删除。将余下的数据按疗法分为 4 组。再将每组中的数据按年龄段进行分组（30 岁以下，30 岁到 40 岁，40 岁以上）。对每组中的数据进行如下处理：找出在第 0 周和在第 7—9 周都有数据的病人，用这些病人在第 7—9 周 CD4 数据的平均值减去其在第 0 周的平均值，得出数据 ZZkmn(ZZ 仅为代号；k 为周数代号：0 代表第 0 周，1 代表在第 7--9 周，2 代表第 15--17 周，3 代表在第 23--25 周，4 代表在第 31--33 周，5 代表在第 38 周以上；m 为疗法代号：1 代表疗法 1，2 代表疗法 2，3 代表疗法 3，4 代表疗法 4；n 为年龄段代号：1 代表 30 岁以下，2 代表 30 岁到 40 岁，3 代表 40 岁以上的)。再找出在第 0 周和第 15—17 周都有数据的病人，在第 0 周和第 23--25 周都有数据的病人，在第 0 周和第 31—33 周都有数据的病人，第 0 周和第 38 周以上都有数据的病人，并分别进行和上边相同的处理。这样便得出 12 组数据。

以数据代号中的 k 为自变量，数据为函数值进行拟合。（每类中的四组数据的图画在同一个图中以便于比较，令 ZZ0mn=0）

年龄在 30 以下的：

时间（周）	CD4 增量			
	疗法 1	疗法 2	疗法 3	疗法 4
0	0	0	0	0
8	-0.20502	-0.11492	0.08188	0.330174
16	-0.37448	-0.1332	-0.0254	0.469919
24	-0.37879	-0.17657	-0.29495	0.233975
32	-0.38001	-0.41425	-0.22034	0.382776
40	-0.92773	-0.34662	0.116718	-0.06997

表 2-1

多项式拟合程序如下：

```
x=[0 8 16 24 32 40];
y130=[0 -0.205019512 -0.374481818 -0.378794286 -0.380005882 -0.92773];
y230=[0 -0.114917241 -0.133197436 -0.176572414 -0.414245161 -0.346622222];
y330=[0 0.08188 -0.025397059 -0.294948387 -0.22034 0.116718182];
y430=[0 0.33017381 0.469918919 0.233975 0.382776471 -0.06997273];
p130=polyfit(x,y130,5),p230=polyfit(x,y230,2),p330=polyfit(x,y330,5),p430=p
olyfit(x,y430,5),
poly2str(p130,'x');
poly2str(p230,'x');
poly2str(p330,'x');
poly2str(p430,'x');
x7=-1:0.1:45;
y1309=polyval(p130,x7);y2309=polyval(p230,x7);y3309=polyval(p330,x7);y4309=
polyval(p430,x7);
plot(x,y130,'r-*',x7,y1309),title('130');
hold on;
plot(x,y230,'k-o',x7,y2309);
plot(x,y330,'m-*',x7,y3309);
plot(x,y430,'g-pentagram',x7,y4309);grid;
hold off
```

运行上述程序得到图像如下所示：

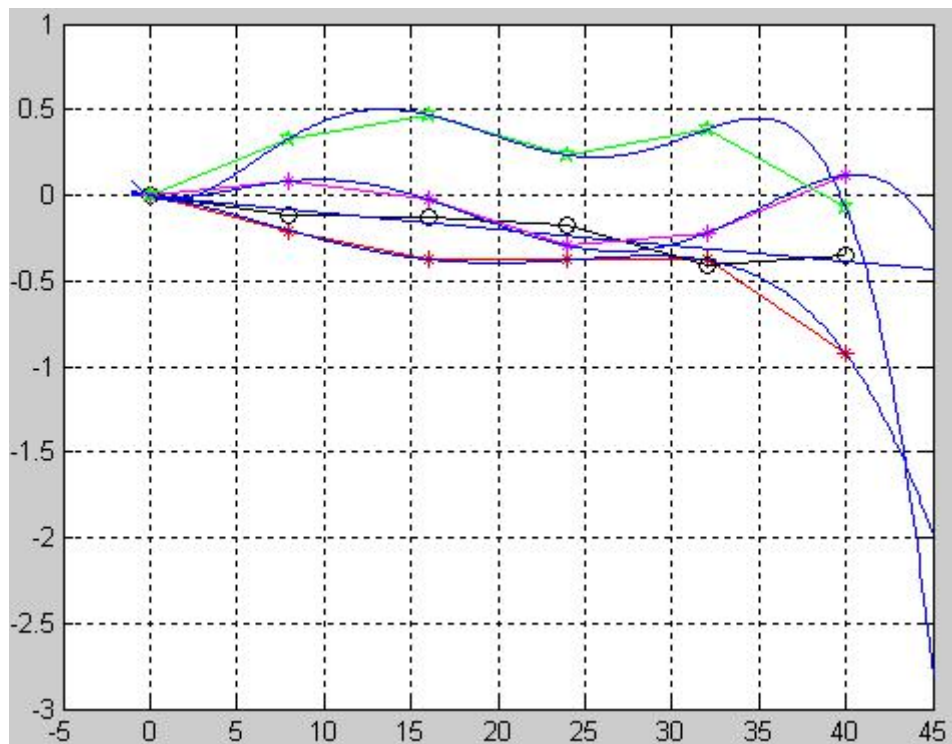


图2-1

年龄在 30—40 间的：

时间（周）	CD4 增量			
	疗法 1	疗法 2	疗法 3	疗法 4
0	0	0	0	0

8	-0.1058	0.095527	0.114986	0.34597
16	-0.20003	-0.21152	-0.05158	0.280854
24	-0.40605	-0.46709	-0.18437	-0.00523
32	-0.59008	-0.53904	-0.33374	-0.02629
40	-0.56704	-0.52004	-0.28951	-0.31184

表 2-2

多项式拟合程序如下：

```
x=[0 8 16 24 32 40];
y130=[0 -0.105802885 -0.20002844 -0.406052632 -0.590084211
-0.567043478];
y230=[0 0.095526667 -0.211523846 -0.467094737 -0.539043011
-0.520044737];
y330=[0 0.114986087 -0.051575969 -0.184373626 -0.333741414
-0.28950625];
y430=[0 0.345970175 0.280854237 -0.005233735 -0.026286139 -0.311835714];
p130=polyfit(x,y130,1),p230=polyfit(x,y230,4),p330=polyfit(x,y330,5),p430=p
olyfit(x,y430,4),
poly2str(p130,'x');
poly2str(p230,'x');
poly2str(p330,'x');
poly2str(p430,'x');
x7=-1:0.1:43;
y1309=polyval(p130,x7);y2309=polyval(p230,x7);y3309=polyval(p330,x7);y4309=
polyval(p430,x7);
plot(x,y130,'r-*',x7,y1309),title('130');
hold on;
plot(x,y230,'k-o',x7,y2309);
plot(x,y330,'m-*',x7,y3309);
plot(x,y430,'g-pentagram',x7,y4309);grid;
hold off
```

运行以上程序得到以下图像：

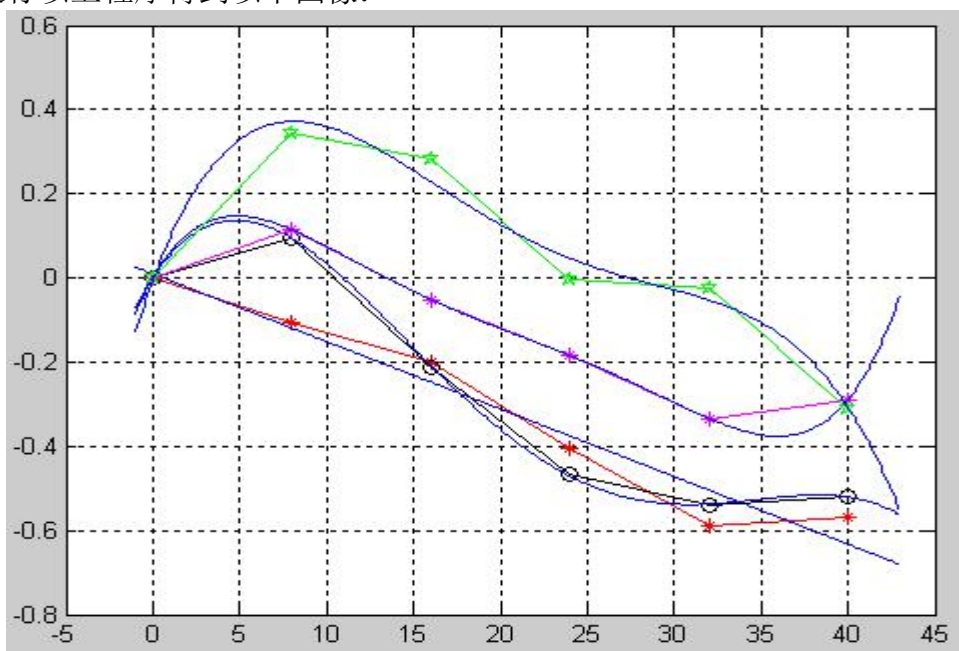


图 2-2



年龄在 40 以上的:

时间 (周)	CD4 增量			
	疗法 1	疗法 2	疗法 3	疗法 4
0	0	0	0	0
8	-0.13858	-0.05167	0.278297	0.375755
16	-0.14856	-0.03265	0.043719	0.281164
24	-0.35231	-0.39014	-0.04567	0.208459
32	-0.40341	-0.14213	-0.0928	0.001791
40	-0.71451	-0.60959	0.044394	0.004556

表 2-3

编写多项式拟合程序如下:

```
x=[0 8 16 24 32 40];
y130=[0 -0.138581928 -0.148564368 -0.352313846 -0.40341
-0.714508696];
y230=[0 -0.051674324 -0.032647436 -0.390135938 -0.14212623
-0.609588];
y330=[0 0.278297368 0.043719355 -0.045668852 -0.092798592 0.04439375];
y430=[0 0.375754762 0.281163636 0.20845942 0.001790789 0.00455625];
p130=polyfit(x,y130,4),p230=polyfit(x,y230,4),p330=polyfit(x,y330,4),p430=p
olyfit(x,y430,4),
disp('4次拟合'),poly2str(p130,'x');
poly2str(p230,'x');
poly2str(p330,'x');
poly2str(p430,'x');
x7=-1:0.1:48;
y1309=polyval(p130,x7);y2309=polyval(p230,x7);y3309=polyval(p330,x7);y4309=
polyval(p430,x7);
plot(x,y130,'r-*',x7,y1309),title('130');
hold on;
plot(x,y230,'k-o',x7,y2309);
plot(x,y330,'m-*',x7,y3309);
plot(x,y430,'g-pentagram',x7,y4309);grid;
hold off
```

运行以上程序得到图像如下:

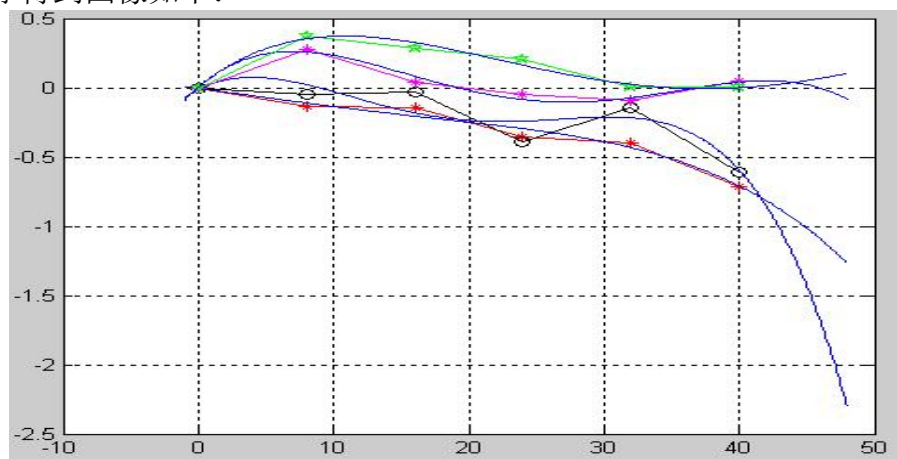


图 2-3

分析说明：在不同年龄病人的身体素质不同，因而我们先按年龄段对病人分组。共三组：

第一组：（30 岁以下）由图 2-1 看出：疗法 4 CD4 的增量明显高于其它三种疗法，且变化一直都较稳定，所以它的疗效是最好的；疗法 3 CD4 的增量次于疗法 4，但其变化也较稳定，而且其治疗最佳终止时间比其它三种疗法都长；疗法 1 和 2 的 CD4 增量一直都比较小，而且几乎没有上升趋势，治疗最佳终止时间也比较短。

还可根据图像预测出若继续采用疗法 4 进行治疗，效果不佳。因为在 36 周以后其拟合图像呈下降趋势，并一直下降下去！

第二组：（30—40）由图 2-2 看出：对于疗法 4，虽然此时 CD4 变化已不太稳定，但其增量明显高于其它三种疗法，所以疗法 4 仍是最佳疗法；疗法 3，CD4 增量仅次于疗法 4，但其变化相对稳定；疗法 1 和 2，CD4 较早就开始呈减少趋势。

还可根据图像预测出若继续采用疗法 4 进行治疗，效果不佳。因为在 8 周以后其拟合图像呈下降趋势，并一直下降下去！

第三组：（40 岁以上）由图 2-3 看出：疗法 4 CD4 的增量明显高于其它三种疗法，且变化一直都较稳定，所以它的疗效是最好的；疗法 3 CD4 的增量次于疗法 4，但其变化也较稳定，而且其治疗最佳终止时间比其它三种疗法都长；疗法 1 和 2，CD4 增量很小，较早就开始呈减少趋势。

再由图像看出第四种疗法在第 10 周以后，CD4 基本呈直线下降趋势。所以预测最佳治疗终止时间为第 10 周左右。

### 问题三

各疗法说明		价格（每日）
疗法 1	600mg zid 或 400mg did 每日 按月轮换使用	\$1.6 或\$0.85
疗法 2	600 mg zid 加 2.25 mg zal 每日	\$(1.6+1.85)
疗法 3	600 mg zid 加 400 mg did 每日	\$(1.6+0.85)
疗法 4	600 mg zid 加 400 mg did，再加 400 mg nev 每日	\$(1.6+0.85+1.2)

表 3-1

结合问题二的图像和结论：

第一组中

疗法 1 和 2 疗效不佳，不建议病人采用；

疗法 3 的最佳治疗终止时间约为第 8 周，共  $8 \times 7 = 56$  天。总费用为： $(1.6 + 0.85) \times 56 = 137.2$ （美元）；

疗法 4 的最佳治疗终止时间约为第 36 周，共  $36 \times 7 = 252$  天。总费用为： $(1.6 + 0.85 + 1.2) \times 252 = 919.8$ （美元）；

第二组中

疗法 1 疗效不佳，不建议采用；

疗法 2 的最佳治疗终止时间约为第 5 周，共  $5 \times 7 = 35$  天。总费用为  $(1.6 + 1.85) \times 35 = 120.75$ （美元）；

疗法 3 的最佳治疗终止时间约为第 8 周，共  $8 \times 7 = 56$  天。总费用为  $(1.6 + 0.85) \times 56 = 137.2$ （美元）；

疗法 4 的最佳治疗终止时间约为第 8 周，共  $8 \times 7 = 56$  天。总费用为  $(1.6 + 0.85 + 1.2) \times 56 = 204.4$ （美元）。

第三组中

疗法 1 和 2 疗效不佳，不建议病人采用；

疗法 3 的最佳治疗终止时间约为第 8 周，共  $8 \times 7 = 56$  天。总费用为： $(1.6 + 0.85) \times 56 = 137.2$ （美元）；

疗法 4 的最佳治疗终止时间约为第 8 周，共  $8 \times 7 = 56$  天。总费用为： $56 \times (1.6 + 0.85 + 1.2) = 204.4$ （美元）。

综上，若病人要综合考虑疗效和费用时，建议采用疗法 3！

## 五、稳定性分析、误差分析、灵敏度分析

本模型利用了大量的实验数据来求得平均值，但由于实验数据中各人的差异很大，利用此模型的结论来预测针对个人的疗效和选择疗法，误差会较大。所以本模型所得结论只能作为参考。如果要更好的预测针对某人的疗效来选择疗法，应该找出几十个数据和这人具有相似性的病人的数据，利用此模型的方法来预测。

## 六、优缺点的讨论

优点：本模型利用大量实验数据，给出的结论具有一般性，这会给那些从没测试过的病人以很好的建议。

按年龄分段处理，使得结果更具针对性，也更准确。

缺点：本模型所利用的平均值中，有的平均值的源数据只有十多组，这样算出的平均值很可能不具一般性，这很可能导致模型有误差。在预测最佳终止时间时，所得出的结果虽不够精确，但在实际情况中，不定因素很多，预测的结果不可能很精确，因此所给出的结果是合理的。

## 七、推广和改进

在问题 1 中，对与实验数据分布不均的问题，可以通过改变采集的数据来解决。如在大约每四周时，采集一次数据。通过移动平均来消除一些特殊值带来的负面影响，这样在数据拟合时，会得到更好的拟合曲线，使得预测的结果可靠性更佳。在对年龄分段时，缺少一个科学的标准（如：体质与年龄的关系），分段就具有主观性，这也会降低结果的可靠性。