

第二届数学中国数学建模网络挑战赛

地址：内蒙古数学会

电话：0471-4343756

邮编：010021

网址：www.tzmcm.cn

Email：ceo@madio.cn

第二届“数学中国杯”数学建模网络挑战赛

承 诺 书

我们仔细阅读了第二届“数学中国杯”数学建模网络挑战赛的竞赛规则。

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。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛报名号为：1010

参赛队员（签名）：

队员 1：刘静

队员 2：张欢

队员 3：王桢

参赛队教练员（签名）： 教练组

参赛队伍组别：大学组

第二届数学中国数学建模网络挑战赛

地址：内蒙古数学会

电话：0471-4343756

邮编：010021

网址：www.tzmcm.cn

Email：ceo@madio.cn

第二届“数学中国杯”数学建模网络挑战赛

编号专用页

参赛队伍的参赛号码：（请各个参赛队提前填写好）：

1010

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

第二届数学中国数学建模网络挑战赛

地址：内蒙古数学会

电话：0471-4343756

网址：www.tzmcm.cnEmail：ceo@madio.cn

邮编：010021

2009 年 第二届“数学中国杯” 数学建模网络挑战赛

题 目 B 题 流感疫苗问题关 键 词 贝叶斯预测模型 概率分析 地域相干分析 主成份 熵值法

摘 要：

流感是一种广泛流行于世界范围内的潜在的致命性疾病，使用疫苗预防流感是抗击这种疾病的有效措施，因此对流感疫苗的研究是非常重要的。

在问题一中，本文首先从静态分析的角度，建立加权统计模型，获得了一种投放方案；然后基于动态预测的思想，建立贝叶斯预测模型，通过对三年数据正向和反向研究，预测出下年流感流行季节的可能流行病毒的，从而确定投放方案。最后建立概率评价模型，在概率的背景下，以疫苗普及率和匹配率为基本元素，构造出评价标准，可以对现行方案和本文新提出的方案进行比较评价。

在问题二中，本文借助病毒专家研制疫苗的原理建立了我们的筛选模型，首先确定流感严重的国家，然后再确定不同亚型的病毒较严重的地点，进而缩小了病毒专家的研究范围，并且通过对以往的相关数据进行思考与分析，验证了此方法的正确性。

在问题三中，我们先后利用了主成份分析模型，熵值法模型对题中所给的指标进行筛选，并利用剩余指标对国家进行评价，得到国家的综合排名，再利用前文中对各国流感状况的评价，综合之后即可得到每个国家综合经济医疗水平与流感情况的排名，进而确定优先级。

最后，在模型的拓展中，建立了地域相干分析模型，通过定量线性模拟，确定了一种投放方案。

参赛队号 1010所选题目 流感疫苗

参赛密码 _____ (由组委会填写)

第二届数学中国数学建模网络挑战赛

地址：内蒙古数学会

电话：0471-4343756

邮编：010021

网址：www.tzmcm.cnEmail：ceo@madio.cn

英文摘要

Influenza is a potentially deadly disease that widely spread all over the world. Using vaccines is an efficient way to battle against the disease. In consequence, it's of great importance to do research in the development of influenza virus vaccines.

In the first question, we used the way of static analyzing and built the model of weighted statically to get an delivery plan; and then in consideration of dynamic prediction, we built the Beyes predicting model and looked into the data of recent 3 years to do both forward and backward research, which showed us the probable subtypes of virus during the influenza epidemic periods in the following year, and helped us to determine the delivery plan. At last, a probabilistic evaluation model was built under the background of probability. A evaluation standard was also built on the basis of popularization rate and matching rate, which can be used to comparative evaluate the scheme used now and also the new one we've just put up.

In the second question, we built a selecting model by virtue of the theories used by virus researchers in vaccine development. Initially, we made sure of the countries suffering severe influenza epidemic, then we confirmed the different places under fearful virus condition. Furthermore, we reduced the range researched by the virus experts, and proved the accuracy of this method after consideration and analysis of the previous data.

In qustion three, we made use of principal component analyzing modle and entropy value modle to choose between the target index. To get a comprehensive ranking of the countries, we also evaluated the standards of different countries through the analysis of remaining index. Further more, we took advantage of the assessment of influenza spreading in different countries mentioned earlier in this assay to get the comprehensive ranking of the integrative level of economic, medical treatment and influenza spreading of each country. On top of this, we obtained the priority levels.

Ultimately, to expand the model, we set up the region relation analyzing model, with fixed amount linear imitation, we established another delivery proposal.

一、问题的背景

流行性感冒简称流感，流感是一种广泛流行于世界范围内的疾病，20 世纪的 4 次甲型流感世界大流行，中国近半个世纪内（1953 年至今）流感流行共计发生大中小规模的流感流行 17 次，其中 2 次为大流行，每次流感大流行都会造成多人死亡和巨大损失。世界卫生组织大力推荐将疫苗作为一种有效的预防措施来抗击这种潜在的致命性疾病。如果疫苗毒株和流行的病毒类型相匹配，那么大约有 50% ~ 80% 的疫苗接种者能够抵抗流感的侵袭。即使疫苗不能完全抵御流感的侵袭，它也可以降低流感发病的严重程度以及严重并发症的发生率。但流感疫苗所能产生的抗体是短效的，所以每年流感流行季节到来前，都需要重新接种疫苗。每年冬天是流感的流行季节，在流行季节到来前 1~2 个月接种疫苗，能达到较为良好的防护效果。

流感病毒分为 A, B, C（又称甲、乙、丙）三型。其中 A 型病毒容易发生变异，依其两种主要抗原(HA, NA) 的不同，区分为不同的亚型。现发现一共有 15 种 HA(H1~H15) 及 9 种 NA(N1~N9) 亚型。同一亚型的病毒之间也可能有一些区别。B 型病毒的变异比较缓慢，C 型病毒甚少对人类造成威胁。

由于流感病毒种类多样，变异迅速，所以每年使用的疫苗成分并不相同。世界卫生组织的专家通过对全球疫情的监控来收集数据，在每年 2 月份预测新的流行季节中流感流行情况(对南半球的疫苗成分推荐在每年的 9 月份确定)，并确定毒株品种作为新年度北半球流感疫苗的推荐成分。需要给药品制造商留出半年左右时间以生产和投放市场。现在的惯用推荐方案是三联装疫苗，也就是每份疫苗中有三种经过灭活或裂解处理（使之对人体失去致病性，但又能使人体的免疫系统识别它，产生针对它的抵抗力）的毒株，分别为两个 A 型和一个 B 型。一般来说推荐的疫苗成分是 A 型的 H1N1 和 H3N2 亚型，以及一个 B 型毒株。但每次建议的具体毒株并不相同，例如 2008~2009 年针对北半球冬季的疫苗组成是：一个 A/布里斯班/59/2007(H1N1) 类似株，一个 A/布里斯班/10/2007(H3N2)类似株和一个 B/佛罗里达/4/2006 类似株。

二、问题的提出与重述

问题 1：建立合适的模型，将北半球和南半球分别划分为稍小的区域，设计使用不同的双联疫苗针对不同区域来进行投放的投放方案，并设计评估标准来评估其效果，使之能与现行方案进行对比评价。

问题 2：建立合理的模型，在监测到的流感流行记录中，筛选出对下一年威胁较大的病毒可能出现的地区。

问题 3：假设联合国有部分经费可以援助不同国家，用以建设流感监测网络实验室和哨点医院。请依据数据，建立合理的模型，评估需要援助的国家或地区的优先级。

三、模型的假设与符号说明

3.1 模型的假设

- 1、如果疫苗毒株1和流行的病毒类型相匹配，那么大约有50%-80% 的疫苗接种者能够抵抗流感的侵袭。即使疫苗不能完全抵御流感的侵袭，它也可以降低流感发病的严重程度以及严重并发症的发生率；
- 2、流感疫苗所能产生的抗体是短效的，所以每年流感流行季节到来前，都需要重新接种疫苗；
- 3、假设人体对病毒的免疫完全依靠疫苗；
- 4、假设题目附件中所给的数据全部准确无误；
- 5、假设感染不同亚型的病毒，如果级别相同，感染人数相同；

3.2 符号说明

符号	符号意义
A_{ij}	第 i 年 j 月出现典型 A 亚型的概率
B_{ij}	第 i 年 j 月出现典型 B 亚型的概率
$H1_{ij}$	第 i 年 j 月出现典型 H1N1 亚型的概率
$H3_{ij}$	第 i 年 j 月出现典型 H3N2 亚型的概率
s	疫苗的成本

剩余变量将在使用时在文中列举出来，在此不做叙述。

四、问题的分析

本题是一个已知以前病毒影响的数据研究如何确定下一年的疫苗组成的相关问题。按照是否考虑亚型分为两类，即问题一与问题二，下面我们将分开讨论这个问题。

4.1 问题一的分析

问题一要求我们将北半球和南半球分别划分为稍小的区域，并使用不同的疫苗针对不同区域来进行投放双联疫苗。由于我们得到的相关材料 and 数据都是以国家为单位的，我们考虑先分析各个国家所需要的双联疫苗的种类，再将情况近似的国家划到同一个区域中，为其投放相同的疫苗，由此设计投放方案。

在本文中还要求我们寻找评估标准进行不同投放方案所产生效果的评价，由于疫苗的目的是预防和控制疫情，因而疫苗的有效性是一个重要的评估标准。另外，由于疫苗生产的成本直接关系需要接种疫苗的人或者国家是否有能力承担疫苗的费用，因此，疫苗的成本也将是一个评估标准。综合多个评估标准，才能对新方案与现有方案做出合理的对比评价。

4.2 问题二的分析

此问题要求我们在考虑选择具体的毒株的情况下，筛选出对下一年威胁较大的病毒在哪出现，目的是缩小病毒专家的研究范围，我们首先找到了病毒专家在确定疫苗时所

使用的方法，根据其原理，我们将我们的筛选模型分为两个步骤：

- (1) 确定流感较为严重的国家；
- (2) 确定不同病毒的研究范围。

在第二问中，我们假设了流感严重国家的具体体现，根据这个指标，我们进行了第一步的筛选，得到了一些流感较为严重的国家，接下来我们开始为每种病毒筛选尽可能小的研究范围，我们假设相邻或相近国家的流感病原体类似，其类似株相同的可能性很大。由此我们就确定了缩小范围后的各种亚型病毒的研究范围。

4.3 问题三的分析

仔细分析问题三，我们认为由于政府的资金有限，就不可能每个国家都资助，因此我们认为在考虑这个问题时，我们应该综合考虑两方面的情况：

- (1) 每个国家的经济及医疗水平情况；
- (2) 每个国家流感的严重程度；

最后每个两个方面的情况就可以得到每个国家的优先级。

五、问题一模型建立与求解

问题一要求我们将北半球和南半球分别划分为稍小的区域，并使用不同的疫苗针对不同区域来进行投放双联疫苗。由于我们得到的相关材料和数据都是以国家为单位的，我们考虑先分析各个国家所需要的双联疫苗的种类，再将情况近似的国家划到同一个区域中，由此设计投放方案。由于疫苗的目的是预防和控制疫情，因而疫苗的有效性是一个重要的评估标准。另外，由于疫苗生产的成本直接关系需要接种疫苗的人或者国家是否有能力承担疫苗的费用，因此，疫苗的成本也将是一个评估标准。综合多个评估标准，才能对新方案与现有方案做出合理的对比评价。

5.1 疫苗投放方案的设计

首先我们需要确定向各个国家投放的疫苗的种类。一方面，一个地区下年可能流行的流感病毒的亚型与当年的情况密切相关（这部分可详细说明，利用SPSS分析数据），我们可以通过统计当年的流感病毒亚型的情况来决定下年该地区投放的疫苗的种类。另一方面，我们希望所投放的疫苗刚好和下年流行的流感病毒的亚型相匹配，因而我们可以通过预测下年的流感病毒情况确定疫苗。针对这两种思路，我们分别建立了加权统计模型和贝叶斯预测模型来确定各国投放的疫苗的种类，进而对南北半球分别划分出稍小的区域进行疫苗投放，得到投放方案。

5.1.1 加权统计模型

从附件所给的数据（参考文献）中，可以看出，流感病毒的流行程度有四个标准，分别为零星活动（Sporadic activity）、当地活动（Local activity）、区域性活动（Regional outbreaks）和广泛传播活动（Widespread outbreaks）。

这里我们的思路是直接对于某个地区的病毒感染情况进行统计，寻找出对每个国家影响较大的两种亚型的病毒。但是在统计的时候，由于流感病毒的流行程度不同，所以直接进行相加之后得到的结果是存在误差的，为了避免结果的不准确性，这里我们采用加权统计模型。

所谓加权统计就是指在计算的时候为每个变量赋予一定的权值然后进行相加求和，完成统计。而这里需要我们进行赋权的就是流感流行的四个标准：

零星活动：孤立案件或实验室确诊病例在一个有限地区；

当地活动：活动的余地的价值观与伊犁实验室确诊病例在有限的区域；

区域活动：实验证实爆发流感在一个或多个地区的人口，包括低于50%的人口总数；

广泛活动：实验证实爆发流感在一个或多个地区的人口，包括不低于50%的人口总数；

从这些定义中，我们可以发现一个问题，这些标准的定义是建立在发病地区面积的基础上，零星活动的面积是最小的，而当地活动的面积是较大的，区域活动的面积又稍微大一点，而广泛活动的面积是最大的，利用我国的政府划分标准，我们希望能将这些区域具体的量化以下，因此，我们考虑为这些面积寻找对应的单位：零星活动——县，当地活动——市，区域活动——省，广泛活动——全国一半以上的省。根据相关资料的查询，我们了解到全国有22个省、5个自治区、4个直辖市、334个市、1 735个县、48 000多个乡。如果我们假设零星活动为1，那么当地活动为5.2，区域活动为78，广泛活动为857。由此我们就得到了零星活动、当地活动、区域活动与广泛活动的权重。

得到了权重，接下来要做的就是依权重进行统计，在统计时，我们确定的统计目标是某个国家感染的总人数，根据上面加权得到的权重来看，这四个标准的影响相差很大，也就是说如果某个国家某种亚型的病毒出现了广泛活动，其影响人数比几个月的零星活动加起来也要影响人数更多，区域活动较零星活动也是这种情况。如果是相同级别的，出现次数越多感染人数越多。

根据上面的分析，研究题中所给的数据，我们可以得到最后的统计结果，根据统计结果，我们就可以得到每个国家应该的投放的疫苗，结果如下表所示：

表5.1 各个国家最终的投放结果

Country, area or territory	投放 疫苗 类型	Country, area or territory	投放 疫苗 类型	Country, area or territory	投放 疫苗 类型	Country, area or territory	投放 疫苗 类型
Cameroon	H1,B	Guatemala	H3,H 1	Nepal	B,H1	Italy	B,H3
Côte d'Ivoire	B,H1	Guiana	B,H1	Oman	B	Kazakhstan	B,H1
Egypt	A,B	Honduras	B	Philippines	B,H1	Kyrgyzstan	H3
France, Réunion	A,H1	Mauritius	H1	Qatar	B,H3	Latvia	H3
Ghana	H1,B	Mexico	B,H1	Republic of Korea	H3,H 1	Lithuania	A
Kenya	B,A	Panama	A,B	Singapore	H3,H 1	Luxembourg	H3,H 1
Madagascar	B,H1	Paraguay	H1	Sri Lanka	A,B	Montenegro	H1
Mauritius	B,H1	Peru	B	Thailand	B,H1	Netherlands	B,H3
Morocco	H1,B	Suriname	H1	Viet Nam	B,H3	Norway	H1,H 3
Senegal	H1,B	Trinidad and Tobago	H3	Albania	B,H1	Poland	A,H3
Seychelles	H1	United Kingdom, Montserrat	B,H1	Austria	A,H3	Portugal	H3,H 1
South Africa	B,H1	Uruguay	H1,B	Belarus	B,A	Romania	H3,H 1

Tunisia	H3	Venezuela	B	Belgium	B,H3	Russian Federation	H3,H1
Uganda	A	Afghanistan	无	Bulgaria	H1	Serbia	B,H1
Argentina	A,B	Bangladesh	B,H3	Croatia	B,H1	Slovakia	H3
Bolivia	无	Cambodia	B,H3	Czech Republic	B,H3	Slovenia	H3,H1
Brazil	B,H1	China	B,H1	Denmark	H3,H1	Spain	H3,H1
Canada	B,H1	China(Province ofTaiwan)	B,H3	Estonia	B,H3	Sweden	B,H3
Chile	A,H1	China,Hong KongSAR	B,H1	Finland	B,H3	Switzerland	B,H1
Colombia	H1,B	India	B,H3	France	H3,H1	Turkey	B,H3
Costa Rica	B,H1	Indonesia	B,H1	Georgia	H1	United Kingdom	H3,H1
Dominica		Iran	B,H3	Germany	B,H1	Ukraine	B,H1
Ecuador	B,H3	Japan	B,H1	Greece	B,H1	Australia	B,H1
El Salvador	B,H1	Jordan	B,H1	Hungary	B,H1	New Zealand	B,H3
FranceFrench ,Guiana	H3	Kuwait	B,H3	Iceland	B,H3	United States of America, Guam	B,H3
France, Guadeloupe	A,H1	Malaysia		Ireland	B,H1	Antarctica	B
France, Martinique	B,H3	Mongolia	B,H3	Israel	B,H1	France, New Caledonia	H1

在得到上表的过程中，有些国家只能确定一种亚型，而另外的亚型影响范围是相同的，在这里我们就任意取了一种作为投放类型，导致模型出现了误差。

5.1.2 贝叶斯预测模型

由于疫苗的目的就是预防疾病的传播，所以与流感季节流行的流感病毒的亚型相匹配的疫苗的防疫效果是最好的。我们希望通过某个国家当年流感病毒流行情况确定下年流行季节流行的流感病毒的亚型，进而得到疫苗投放方案。这是一种预测。

如果下年流行季节流行的病毒的情况是已知的，那么我们可以近似地得到下年流行季节流行的是某种亚型病毒的前提下，当年某月流行某种亚型病毒的概率，这是一个条件概率，不妨记为 $P(M|N)$ 。事实上，由于我们的已知的当年的数据是非常丰富的，因而通过统计当年某月流行某种亚型病毒的频率就可以近似的认为为其概率，这一点是由概率论中的大数定律保证的。

近似地，我们可以研究历史年份数据。比如说，2008年12月的流行情况，对于2007

年2月到2008年1月的人来说，是未知的下年流行的病毒的情况，而对于我们来说，这个数据是已知的。因而我们可以据历史月数据计算条件概率 $P(M|N)$ ，作为一种经验的先验概率。

我们要预测下年的病毒流行情况，一般而言是据当年各月病毒流行情况来预测下年病毒流行季节的流行情况，其实质是求条件概率 $P(N|M)$ ，这是一个后验概率。在概率论中有一个计算后验概率的公式——贝叶斯公式，应用它可以解决我们的为题。据此建立贝叶斯预测模型。

5.1.2.1 模型的建立

以北半球为例。事件 A_{ij} , B_{ij} , $H1_{ij}$, $H3_{ij}$ 分别表示第 i 年 j 月出现典型A亚型、B亚型、H1N1亚型和H3N2亚型的概率。其中，从2007年2月至2008年1月为第一年 ($i=1$)，从2008年2月至2009年1月为第二年 ($i=2$)，从2009年2月至2010年1月为第三年 ($i=3$)。 j 的取值范围为 $1 \sim 12$ ，表示第 i ($i=1,2,3$) 年的第 j 月。一般地，事件 X_{ij} 表示第 i 年 j 月出现 X 亚型 ($X \in \{A_{ij}, B_{ij}, H1_{ij}, H3_{ij}\}$)，事件 Y_{ij} 表示第 i 年 j 月出现 Y 亚型 ($Y \in \{A_{ij}, B_{ij}, H1_{ij}, H3_{ij}\}$)。

由贝叶斯公式，我们得到据已知数据预测未来情况的贝叶斯预测模型：

$$P(Y_{(i+1),m} | X_{ij}) = \frac{P(X_{ij} | Y_{(i+1),m}) P(Y_{(i+1),m})}{P(X_{ij})} = \frac{P(X_{ij} | Y_{(i+1),m}) P(Y_{(i+1),m})}{\sum_{k=1}^4 P(X_{ij} | Y_{(i+1),m}^{(k)}) P(Y_{(i+1),m}^{(k)})} \quad (5.1)$$

$(m=10,11,12)$

其中， $Y_{ij}^{(k)}$ 在 $k=1,2,3,4$ 时分别表示 $A_{ij}, B_{ij}, H1_{ij}, H3_{ij}$ 。

在(5.1)中， X_{ij} 表示在已知的第 i 年第 j 月出现 X 亚型的流感疫情， $Y_{(i+1),m}$ 表示在未来的第 $i+1$ 年 m 月出现 Y 亚型的流感疫情。我们希望通过(5.1)得到第 i 年第 j 月出现 X 亚型的流感疫情的前提下，第 $i+1$ 年 m 月出现 Y 亚型的流感疫情的概率。

通过5.1.2.1的分析，我们可以近似地认为如果不考虑已发生月份的流感疫情，某个月出现某种亚型流感病毒的概率可以近似地认为是一段时期以来该种亚型出现的频率，即

$$P(X_{ij}) = f_X \quad (5.2)$$

我们通过统计2007年2月至2009年1月的数据可以很容易地得到这组数据。

要通过(5.1)来预测，还需要计算 $P(X_{ij} | Y_{(i+1),m}^{(k)})$ 的值。我们可以通过研究2007年2月至2008年1月的数据与2008年12月、2009年1月、2009年2月的数据，得到经验上的 $P(X_{ij} | Y_{(i+1),m}^{(k)})$ 的值。由于我们统计的数据量比较大，因此得到的结果是可信的。事实上，贝叶斯公式在应用过程中，总是需要一些由经验或者常识得到的先验概率。由于我

们所要预测的是一个季节的情况，为了简化问题，我们可以不考虑 m 的取值，将整个冬天看作一个整体。

到此，我们可以依据(5.1)，得到如果某月份出现亚型 X ，下年的冬天出现 Y 亚型的病毒的概率。

5.1.2.2 模型的讨论

由于我们的已知数据是相对比较丰富的，所以仅以某一个月发生某种亚型去判断下年相关月份将出现的病毒亚型没有充分地利用已知数据。事实上，我们并不是想得到下年相关月份出现某种亚型的概率，而是想通过比较概率的大小来确定研制疫苗使用的亚型。因此，我们可以类比“独立重复试验中试验成功的概率”的计算方法，设置一个比较标准：

$$M(i+1, m) = 1 - \prod_{j=1}^{12} (1 - T(X_{ij}) \cdot P(Y_{(i+1), m} | X_{ij})) \quad (5.3)$$

其中，

$$T(X_{ij}) = \begin{cases} 0, & \text{如果 } X_{ij} \text{ 不发生} \\ 1, & \text{如果 } X_{ij} \text{ 发生} \end{cases}$$

由于我们要使用双联装疫苗，我们只需取 $M(i+1, m)$ 比较大的两种亚型即可。

5.1.2.3 模型的求解

1、求各亚型的病毒在某个月出现的概率。

将2007年2月至2009年1月各月份的统计到的各国的各亚型病毒发生次数统计如表

5.2

表5.2 2007年2月至2009年1月各月份各亚型病毒爆发情况

A	3	7	8	5	6	4	1	4	5	9	9	12	8
B	38	32	23	13	9	5	4	18	24	29	38	35	49
H1	33	26	18	12	8	5	4	16	24	30	38	37	54
H3	45	39	28	19	8	5	4	13	19	17	15	18	30
A	6	7	4	7	6	3	0	2	3	1	2	1	
B	49	46	30	23	19	7	2	21	21	24	26	32	
H1	40	27	25	15	12	5	2	11	15	25	30	31	
H3	25	24	17	14	11	3	1	13	21	26	31	41	

将各月份数据再相加，得到2007年2月至2009年1月各亚型病毒发生地总次数和相应的频率，如表5.3所示

表5.3 2007年2月至2009年1月各亚型病毒发生地总次数和相应的频率

	A	B	H1	H3	总和
总次数	123	617	543	487	1770
频率	0.069492	0.348588	0.30678	0.275141	1

据(5.2)，从表5.3中我们可以得到各个亚型的病毒发生地概率为：

$$P(A) = 0.069492, P(B) = 0.348588, P(H1) = 0.30678, P(H3) = 0.275141$$

2、求先验概率

统计2008年和2009年间的冬天出现各亚型流感时，2007年2月至2008年1月各月份的出现的亚型个数及频率如表5-4-1，5-4-2，5-4-2，5-4-4所示：

表5-4-1 下年冬天出现A亚型时当年各月情况统计

月份		2	3	4	5	6	7	8	9	10	11	12	1
A	发生次数	0	1	0	0	0	0	0	0	0	1	2	2
	频率	0	0.2	0	0	0	0	0	0	0	0.33	0.29	0.25
B	发生次数	2	2	1	0	0	0	0	0	0	1	3	3
	频率	0.4	0.4	0.5	0	0	0	0	0	0	0.33	0.43	0.38
H1	发生次数	1	0	0	0	0	0	0	0	0	1	2	3
	频率	0.2	0	0	0	0	0	0	0	0	0.33	0.29	0.38
H3	发生次数	2	2	1	0	0	0	0	0	0	0	0	0
	频率	0.4	0.4	0.5	0	0	0	0	0	0	0	0	0
总计	发生次数	5	5	2	0	0	0	0	0	0	3	7	8
	频率	1	1	1	0	0	0	0	0	0	1	1	1

表5-4-2 下年冬天出现B亚型时当年各月情况统计

月份		2	3	4	5	6	7	8	9	10	11	12	1
A	发生次数	2	5	4	2	2	1	0	0	0	4	5	9
	频率	0.027	0.074	0.089	0.071	0.10	0.083	0	0	0	0.067	0.071	0.122
B	发生次数	22	20	16	8	5	3	2	9	14	21	25	26
	频率	0.301	0.294	0.356	0.286	0.25	0.25	0.20	0.333	0.341	0.350	0.357	0.351
H1	发生次数	20	17	10	6	6	4	4	10	14	22	27	26
	频率	0.274	0.250	0.222	0.214	0.30	0.333	0.4	0.370	0.341	0.367	0.386	0.351
H3	发生次数	29	26	15	12	7	4	4	8	13	13	13	13
	频率	0.397	0.382	0.333	0.429	0.350	0.333	0.40	0.296	0.317	0.217	0.186	0.176
总计	发生次数	73	68	45	28	20	12	10	27	41	60	70	74
	频率	1	1	1	1	1	1	1	1	1	1	1	1

表5-4-3 下年冬天出现H1亚型时当年各月情况统计

月份		2	3	4	5	6	7	8	9	10	11	12	1
A	发生次数	3	3	4	3	3	2	1	1	1	4	4	7
	频率	0.04	0.04	0.08	0.08	0.14	0.15	0.08	0.03	0.02	0.06	0.06	0.10
B	发生次数	24	23	17	10	5	3	3	11	18	21	26	23
	频率	0.31	0.34	0.33	0.27	0.23	0.23	0.25	0.33	0.35	0.34	0.37	0.34
H1	发生次数	21	16	13	8	7	4	4	12	17	24	27	25
	频率	0.27	0.24	0.25	0.22	0.32	0.31	0.33	0.36	0.33	0.39	0.38	0.37
H3	发生次数	29	26	18	16	7	4	4	9	15	13	14	12
	频率	0.38	0.38	0.35	0.43	0.32	0.31	0.33	0.27	0.29	0.21	0.20	0.18

总计	发生次数	77	68	52	37	22	13	12	33	51	62	71	67
	频率	1	1	1	1	1	1	1	1	1	1	1	1

表5-4-4 下年冬天出现H3亚型时当年各月情况统计

月份		2	3	4	5	6	7	8	9	10	11	12	1
A	发生次数	3	4	6	2	2	1	0	0	1	4	6	10
	频率	0.03	0.05	0.11	0.06	0.12	0.11	0.00	0.00	0.02	0.06	0.07	0.11
B	发生次数	25	23	17	9	4	2	1	9	16	23	30	31
	频率	0.29	0.30	0.32	0.28	0.24	0.22	0.14	0.35	0.36	0.36	0.36	0.35
H1	发生次数	24	19	12	8	5	3	3	10	16	25	34	32
	频率	0.28	0.25	0.23	0.25	0.29	0.33	0.43	0.38	0.36	0.39	0.41	0.36
H3	发生次数	34	31	18	13	6	3	3	7	11	12	13	15
	频率	0.40	0.40	0.34	0.41	0.35	0.33	0.43	0.27	0.25	0.19	0.16	0.17
总计	发生次数	86	77	53	32	17	9	7	26	44	64	83	88
	频率	1	1	1	1	1	1	1	1	1	1	1	1

表5-4-1至表5-4-4中的“频率”表示2008年和2009年间的冬天出现某种亚型的流感时，2007年2月至2008年1月各月份的出现的各个亚型的次数及频率。由5.1.2.1中的论述可知，这些“频率”可以近似地认为是先验概率 $P(X_{ij} | Y_{(i+1),m}^{(k)})$ 。

3、求后验概率

得到了 $P(A), P(B), P(H1), P(H3)$ 和先验概率 $P(X_{ij} | Y_{(i+1),m}^{(k)})$ 以后，根据5.1.2.1中的分析，由(5.1)可以计算出 $P(Y_{(i+1),m} | X_{ij})$ ，如表5.5所示。

表5.5 后验概率计算值

		A	B	H1	H3			A	B	H1	H3
February	A	0	0.307	0.384	0.309	August	A	0	0	1	0
	B	0.09	0.341	0.31	0.259		B	0	0.375	0.413	0.212
	H1	0.052	0.354	0.31	0.285		H1	0	0.388	0.284	0.328
	H3	0.071	0.355	0.296	0.278		H3	0	0.388	0.284	0.328
March	A	0.206	0.381	0.201	0.212	September	A	0	0	1	0
	B	0.088	0.324	0.328	0.26		B	0	0.37	0.326	0.304
	H1	0	0.384	0.318	0.299		H1	0	0.373	0.322	0.305
	H3	0.071	0.342	0.301	0.285		H3	0	0.396	0.321	0.284
April	A	0	0.361	0.275	0.363	October	A	0	0	0.49	0.51
	B	0.1	0.357	0.289	0.254		B	0	0.364	0.331	0.306
	H1	0	0.358	0.354	0.288		H1	0	0.37	0.318	0.311
	H3	0.099	0.331	0.303	0.267		H3	0	0.41	0.335	0.255
May	A	0	0.372	0.371	0.257	November	A	0.278	0.279	0.237	0.206
	B	0	0.383	0.319	0.298		B	0.067	0.351	0.299	0.284
	H1	0	0.356	0.316	0.328		H1	0.061	0.339	0.315	0.285

	H3	0	0.379	0.337	0.284		H3	0	0.395	0.336	0.269
June	A	0	0.32	0.384	0.297	December	A	0.242	0.304	0.211	0.243
	B	0	0.393	0.315	0.292		B	0.081	0.34	0.307	0.272
	H1	0	0.369	0.345	0.286		H1	0.052	0.35	0.304	0.294
	H3	0	0.385	0.308	0.307		H3	0	0.385	0.359	0.256
July	A	0	0.272	0.442	0.286	January	A	0.141	0.344	0.26	0.254
	B	0	0.398	0.323	0.279		B	0.074	0.349	0.3	0.276
	H1	0	0.384	0.312	0.303		H1	0.072	0.337	0.315	0.276
	H3	0	0.384	0.312	0.303		H3	0	0.376	0.337	0.288

4、求各国不同亚型对应的 $M(i+1, m)$

利用表5.5中的后验概率，代入相关数据。通过EXCEL软件进行数据处理，可以得到各国不同亚型对应的 $M(i+1, m)$ 。表5.6给出了一部分值：

表5.6 部分亚型对应的 $M(i+1, m)$ 值

	Cameroon	Côte d'Ivoire	Egypt	Ghana	Morocco
A	0.85064625	0.46273985	0.76580428	0.57883276	0.91774284
B	0.9964376	0.73858164	0.95854827	0.98315705	0.92823922
H1	0.99650523	0.74831077	1	0.98244322	0.93190871
H3	0.99704059	0.75514808	0.9549428	0.9850129	0.92781737

进而可以得到向各国投放疫苗的种类，如表5.7所示。

表5.7 各国投放疫苗种类

Africa		China	B,H3	Germany	B,H1
Cameroon	B,H3	India	B,H3	Greece	B,H3
		Japan	B,H3	Hungary	H1,H3
Egypt	B,H1	Jordan	H1,H3	Iceland	B,H1
Ghana	B,H3	Kuwait	B,H3	Ireland	B,H1
Morocco	B,H1	Malaysia	H1,H3	Israel	B,H1
Senegal	B,H3	Mongolia	A,H1	Italy	B,H1
Tunisia	H1,H3	Oman	B,H1	Kazakhstan	B,H1
America		Philippines	B,H3	Kyrgyzstan	A,B
Canada	H1,H3	Qatar	A,H1	Latvia	H1,H3
Colombia	H1,H3	Singapore	B,H1	Lithuania	A,H1
Costa Rica	H1,H3	Sri Lanka	H1,H3	Luxembourg	B,H1
El Salvador	H1,H3	Thailand	B,H3	Netherlands	B,H1
Guatemala	B,H3	Europe		Norway	B,H1
Honduras	H1,H3	Austria	B,H1	Poland	B,H1
Mexico	H1,H3	Belarus	A,H3	Portugal	B,H1
Panama	H1,H3	Belgium	B,H1	Romania	B,H1
Paraguay	H1,H3	Bulgaria	B,H3	Serbia	B,H1

Peru	H1,H3	Croatia	B,H1	Slovakia	H1,H3
Suriname	B,H3	Denmark	B,H1	Slovenia	B,H1
Venezuela	A,H1	Estonia	B,H3	Spain	H1,H3
Asia		Finland	B,H1	Sweden	B,H1
Bangladesh	B,H3	France	B,H1	Switzerland	B,H1

进而可以将划分一些区域，每个区域使用一种双联疫苗。如表5.8所示。

表5.8 疫苗投放方案

B,H3	H1,H3	B,H1	B,H1	B,H3	A,H1
Africa	Africa	Africa	Iceland	Africa	America
Cameroon	Côte d'Ivoire	Egypt	Ireland	Cameroon	Trinidad and Tobago
Ghana	Tunisia	Morocco	Israel	Ghana	Venezuela
Senegal	America	America	Italy	Senegal	Asia
America	Canada	France, Martinique	Kazakhstan	America	Mongolia
Guatemala	Colombia	United States of America	Luxembourg	Guatemala	Qatar
Suriname	Costa Rica	Asia	Netherlands	Suriname	Europe
United Kingdom, Montserrat	El Salvador	China (Province of Taiwan)	Norway	United Kingdom, Montserrat	Lithuania
Asia	France, French Guiana	China, Hong Kong SAR	Poland	Asia	Oceania
Bangladesh	France, Guadeloupe	Iran (Islamic Republic of)	Portugal	Bangladesh	
China	Honduras	Oman	Romania	China	
India	Mexico	Republic of Korea	Russian Federation	India	A,H3
Japan	Panama	Singapore	Serbia	Japan	Europe
Kuwait	Paraguay	Europe	Slovenia	Kuwait	Belarus
Philippines	Peru	Austria	Sweden	Philippines	
Thailand	Asia	Belgium	Switzerland	Thailand	
Europe	Jordan	Croatia	Turkey	Europe	
Bulgaria	Malaysia	Czech Republic	United Kingdom of Great Britain and Northern Ireland	Bulgaria	A,B
Estonia	Sri Lanka	Denmark	Ukraine	Estonia	Europe
Georgia	Europe	Finland	Oceania	Georgia	Kyrgyzstan
Greece	Hungary	France	United States of America,	Greece	

			Guam		
Oceania	Latvia	Germany		Oceania	
	Slovakia				
	Spain				
	Oceania				

5.2 概率评价模型

流感疫苗的制造成本较高，因而考虑将三联装疫苗改变成双联装，分地区投放。成本的高低直接影响着疫苗的价格，进而易感人群的购买力。因此降低疫苗成本可以使更多的人有能力使用疫苗，从而对于整体的抗击流感工作是有益的。但另一方面，双联装的疫苗比三联装的疫苗需要承担更多的疫苗与实际流行病毒不匹配的风险。因而我们希望建立一个评价模型，来评价5.1.1和5.1.2的投放方案的效果，并与现行方案进行对比评价。

5.2.1.1 成本与疫苗使用率的关系

设疫苗的成本为 s ，由于成本与售价之间是正相关的，为了简化模型，不妨设它们之间成线性关系，因而可以设疫苗的价格 $q = ks + a$ 。

易感人群对于疫苗的价格的承受力是有一定限度的，设这个限度为 S_M 。可以认为，当疫苗的价格大于 S 时，疫苗的使用率为0。另外，如果假设易感人群都是愿意注射疫苗的，那么，当疫苗的价格足够低时，不妨设为 S_m ，疫苗使用率几乎为1。当疫苗的价格水平很低（接近 S_m ）时，提高价格对于疫苗的使用率影响不是很大，即疫苗使用率随着价格提高而降低的速度比较平缓；当疫苗的价格水平很高（接近 S_M ）时，降低价格对于疫苗的使用率的影响也不是很大，也即疫苗使用率随着价格降低而提高的速度也比较平缓。但是在疫苗的价格适中的时候，疫苗价格的提高或者降低对于疫苗的使用率的影响却比较大。（如图5.1所示）因而，疫苗的使用率与疫苗的价格之间的函数关系可以用“降岭函数”来描述。

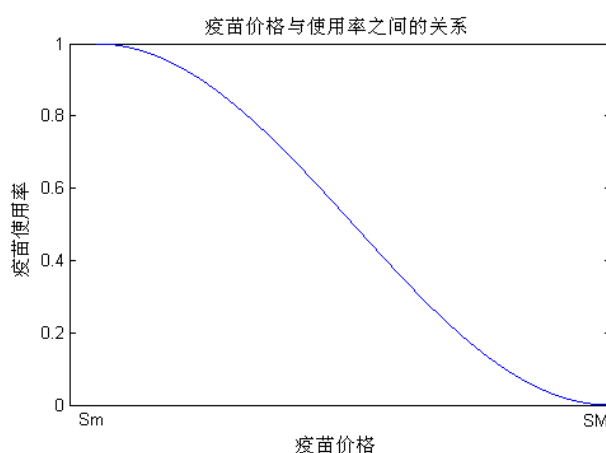


图5.1 疫苗价格与使用率之间的关系

“降岭函数”是一种隶属函数，各相邻等级之间的数值差距并不是相等的，在起始等级和结束等级其变化差距较大，在中间等级时其变化较之稳定，而且函数是对称的。这个“等级”，就相应于上述的疫苗使用率。使用“降岭函数”对于成本与疫苗使用率之间的关系进行定量描述：

$$H(q) = \begin{cases} 1 & q < S_m \\ 0.5 - 0.5 \sin \frac{\pi}{S_M - S_m} \left(q - \frac{S_M + S_m}{2} \right) & S_m \leq q \leq S_M \\ 0 & q > S_M \end{cases} \quad (5.4)$$

$$q = ks + a$$

其中 $H(q)$ 表示疫苗价格为 q 时的疫苗使用率。

5.2.1.2 疫苗与实际流行病毒不匹配的风险的度量

双联装的疫苗比三联装的疫苗需要承担更多的疫苗与实际流行病毒不匹配的风险。事实上，无论是哪一种方案，都存在这种风险。我们用疫苗与实际流行病毒不匹配的概率 p 来衡量这种风险的大小。

在实际计算中，2009年12月至2010年2月这个冬天的实际情况是未知的，所以计算2009年至2010年间的冬天的疫苗的不匹配概率是不容易实现的。考虑到方案的优劣主要是由方案本身决定的，所以我们历史月份的数据进行 p 的度量以评价方案。具体而言，就是依据2007年2月至2008年1月的数据，使用不同的方案的方法，确定2008年12月至2009年1月这个冬天所使用的疫苗。而2008年12月至2009年1月这个冬天的实际流感病毒流行情况我们是知道的，因而比较容易计算概率 p 。具体过程为：

(1) 使用某种方案，据2007年2月至2008年1月的数据，确定2008年12月至2009年1月这个冬天所使用的疫苗。

(2) 参照2008年12月至2009年1月的数据，统计(1)中的疫苗不能满足实际需要的国家的个数占总国家个数的比例。这个比例即为疫苗与实际情况不匹配的频率。

由于国家的个数比较多，根据概率论的大数定律，这个频率即可近似地作为疫苗与实际情况不匹配的概率 p 。

5.2.1.3 综合评价

由于投放方案的优劣与疫苗的使用率和疫苗的不匹配率有关，并且，疫苗的使用率越高，疫苗的不匹配率低，投放方案就越优。这样，我们可以以

$$Q = H(q)(1 - p) \quad (5.5)$$

为标准进行衡量， Q 越大，方案越优， Q 越小，方案越劣。此即为概率评价模型。

六、问题二模型建立与求解

6.1 筛选模型：

依据题中给出的参考文献，我们可以获取出病毒专家在确定具体疫苗时所采用的方法：首先寻找流感比较严重的国家，然后从这些国家感染流感病毒的人群中分离出流感病原体，对分离出的流感病原体采用血凝抑制试验等方法寻找其类似的毒株，如果不同的病原体有相同的类似毒株，这类病原体将作为同种毒株看待，相同亚型的前提下如果最后某种毒株的量最多就作为此种亚型病毒的具体毒株，即下一年的疫苗。

分析题目我们可以知道，题目要求我们筛选出对下一年威胁较大的病毒，并确定其最可能出现的地方。在此我们利用病毒专家在确定疫苗组成时利用的方法进行我们的筛选过程，认真思考这个方法不难发现，最终所获取的疫苗实际上就是今年出现概率最多的病毒，也就是说对下一年威胁较大的病毒是今年较流行的病毒，也是我们选取疫苗时应该选择的。分析参考文献中的数据我们也可以得到如下的规律：

(1) 如果此种类型的病毒在今年的威胁相对比较大，那么在下一年这个病毒的危害将会相对比较小，反之今年的威胁相对较小的病毒，那么下一年时威胁将会较大；

(2) 如果某一地区今年发生了严重的流感，那么下一年发生流感的相对可能性会较大，如果今年没有发生或者发生的是比较轻微的流感，那么下一年发生的仍是比较轻微的流感；

(3) 如果某一地区今年所有的病毒每个月都有威胁，而且大小不一，有较大有较小，那么下年可能会有较强的威胁；

针对上述规律，我们进行了如下的思考：对于规律(1)，我们认为是因为在下一年中接种了上一年比较严重的病毒组合而成的疫苗，导致其影响能力下降，而其他病毒由于没有疫苗的抵制作用影响范围得到上升；对于规律(2)，某个地区如果发生了较为严重的流感，我们可以认为它的环境因素是病毒所需的环境，因此在下一年发生流感的概率也就很大；规律(3)，因为这种病毒的危害不是很大，往往容易被忽略，随着时间的推移危害可能越来越大。

通过上面对数据的分析与思考也进一步验证了我们前面的想法，所以说我们的想法是正确的，我们应该选择今年较流行的病毒作为威胁较大的病毒，说明流感较严重的国家的流感病毒具有很强的可研究性。那么首先我们要确定的就是流感较严重的国家。

在这里我们做出了如下的模型假设：

(1) 假设区域活动和广泛活动即表示流感较严重，零星活动、当地活动表示流感不严重；

(2) 假设同种亚型的不同病毒仅仅是结构不同，但是产生的威胁没有区别，例如在不同国家***B所造成的威胁仅与这些国家的人口、面积有关；

(3) 假设有一个月份出现了威胁较大的病毒，即认为此国家的流感较严重；

(4) 假设由于流感病毒的可传染性，我们认为相近地区的病毒具有相似性，其是同一种类似株的概率也比较大；

6.1.1. 确定流感较严重的国家

利用上面的分析及假设对每个国家进行判定，即可得到流感较严重的国家，并且得到了使其流感严重的具体亚型，具体如下表所示：

表 6.1 流感较严重的国家

Country, area or territory	2008年威胁较大	2008年威胁较小
Africa		
South Africa	***H1	*H3,*B
Tunisia	***H3	*H1,*B

America		
Argentina	****A, ****B, ***H1	*H1, *A, *B
Brazil	****A, ****B, ****H1	*H1, *A, *B
Chile	***H1	*H1, *A, *B
United States of America	****A, ****B, ****H1	**H1, *H3, **B
Asia		
Japan	****H1, ****H3,	*H1, *H3, *B
Europe		
Austria	***A	*H3,
Belgium	****H3,	*H1, *B
Czech Republic	***H3	*B
Denmark	****H3	*H1
Estonia	****H3	*H1, *B
France	****H3	*H1, *B
Germany	****H3	*H1, *B
Hungary	***H3	
Latvia	***H3	*H1, *A, *B
Luxembourg	****H3	*H1
Netherlands	****H3	*H1, *B
Norway	****H3	*H1, *B
Portugal	****H3	*H1, *B
Slovenia	****H3	*H1, *B
Spain	****H3	*H1, *B
Sweden	****H3	*H1, *B
Switzerland	****H3	*B
United Kingdom	****H3	*B
Oceania		
Australia	***B, ***H3	*H1, *H3, *B
New Zealand	****B, ****H3	*H1, *B

6.1.2.不同病毒的具体研究区域

通过上面的过程，我们就筛选出了流感较为严重的国家。那么接下来的任务就转化为确定不同病毒所需的更小的研究范围。由题目我们可以知道，现在的疫苗多为三联装，所以我们需要确定三种具体毒株的影响范围，而观察上表我们发现 A 的影响最小，所以我们决定排除 A 的作用，只研究 B，H1 和 H3 三种亚型的具体毒株。

通过观察上表，我们可以发现对于 B 病毒只有阿根廷、巴西、美国、澳大利亚与新西兰这几个国家的影响情况较为严重。所以为了研制针对 B 病毒的疫苗，专家们只需在这些国家中研究即可。这样 B 的研究范围就确定了：阿根廷、巴西、美国、澳大利亚与新西兰。

B 病毒的研究范围确定后，接下来我们需要研究的就是 H1, H3 的研究范围，通过观察数据发现，受 H1 和 H3 影响的区域呈现一定的规律，受 H1 影响的区域主要集中在美洲和亚洲，而受 H3 影响的区域主要集中在欧洲地区，这样确定的国家中受 H1 影响的

区域不是很大，符合题中缩小范围的要求，然而尽管我们把 H3 的研究范围缩小至欧洲，但是国家仍然很多，区域也很大。利用假设（4），我们认为进一步缩小范围时应根据地理位置关系，通过在地图上确定这些国家的地理位置，取相邻或相近的国家作为我们的研究范围，由此就进一步减小了 H3 的研究范围，所得的最终结果如下：

表 6.2 最终确定的 H1, H3 的研究范围

	B 型病毒	H1 亚型病毒	H3 亚型病毒
研究范围	Argentina	South Africa	Belgium
	Brazil	Argentina	Denmark
	United States of America	Brazil	France
	Australia	Chile	Germany
	New Zealand	United States of America	Netherlands
		Japan	Spain
			United Kingdom

为了对模型进行检验，这里我们用 07 年的数据预测 08 年的数据，得到 08 年预测流感较严重的国家，并与 08 年各国的实际情况做对比，根据 07 年的数据预测出的 08 年流感较严重的国家如下表所示：

表 6.3 根据 07 年的数据估计 08 年流感严重的国家

Country, area or territory	2007年威胁较大	2007年威胁较小
America		
Canada	***H1	*H1,*H3,*B
France, Guadeloupe	***A	
United States of America	****A,****B,****H1	**H1,*H3,**B
Asia		
China	***B	**B, **H3
Japan	***H1,	*H1,*H3,*B
Europe		
Austria	***H1	*H3,
Belarus	***H1	*H3,*B
Belgium	****H1,	*H1, *B
Czech Republic	***H1	*B
Estonia	****H1	*H1,*B
France	****H1	*H1,*B
Hungary	***H1,***B	
Luxembourg	****H1	*H1
Portugal	****H1	*H1,*B
Slovenia	****H1	*H1,*B
Spain	****H1	*H1,*B
Switzerland	****H1	*B
United Kingdom	****H1	*B
Oceania		
New Zealand	***B	*H1,*B

而我们知道表 6.1 即是 08 年实际流感较严重的国家，对比两个表格，我们可以

发现这两个表格的吻合艺术很高，而且变化也符合我们在问题之前所做的假设，所以我们认为我们的模型是正确的。

七、问题三模型建立与求解

对题目进行思考总结分析，我们可以认为在问题一中题目要求我们做的是分区域进行二联装的投放疫苗，并将投放方案与三联装时进行比较，问题二中是希望我们在三联装的前提下为流感专家缩小确定流感疫苗的范围，也就是说确定每种亚型的流感病毒比较严重的国家，这样流感专家就可以通过提取分离这些地区的流感病毒得到下一年的疫苗，仔细分析问题三，我们认为由于政府的资金有限，就不可能每个国家都资助，因此我们认为在考虑这个问题时，我们应该综合考虑两方面的认识：

(1) 每个国家的经济及医疗水平情况；

(2) 每个国家流感的严重情况；

我们认为像美国、法国、英国等发达国家即使他们的流感疫情比较严重也不需要联合国的援助，因此我们首先根据各国的基本国情缩小我们所需要的研究国家的范围。

分析参考文献，我们可以从中提取出影响各国医疗水平的体系表，如下表所示：

表 7.1 医疗水平指标体系表

准则层	指标层	单位
区域和国家的图表	5岁以下年龄别低体重儿	%
	5岁以下儿童死亡率	每千例活产中5岁死亡的概率
	一岁以下儿童麻疹疫苗接种率	%
	孕产妇死亡率	每 100 000 例活产
	专业医护人员的接生	%
	避孕普及率	%
	青春期生育率	每1000 个15-19岁之间的女孩
	产前检查覆盖率	%
	未满足的计划生育需求	%
	每100 000 人中15岁（含）以上成年人艾滋病病毒感染率	
	15-24岁之间接受全面正确的艾滋病毒/艾滋病知识的男性比例	%
	15-24岁之间接受全面正确的艾滋病毒/艾滋病知识的女性比例	%
	晚期艾滋病感染者接受抗逆转录病毒疗法的覆盖率	%
	每100 000 人中疟疾造成的死亡率	
	5岁以下儿童睡在经杀虫剂处理的蚊帐里的比例	%
	五岁以下发烧儿童接受抗疟治疗的比例	%
	按照DOTS策略实现的成功治愈率	%

	安全饮用水的可及性	%
	改良后的卫生设施的可及性	%
死亡率和疾病负担	出生期望寿命	年
	出生健康期望寿命	年
	新生儿死亡率	每1000 例活产
	婴儿死亡率	每1000例活产儿中从出生到1岁时死亡的概率
	5岁以下死亡率	每1000例活产儿5岁以下死亡的概率
	成人死亡率	每1000人中在15岁到60岁之间死亡的概率
病因特异性死亡率与发病率	死亡率	%
	孕产妇死亡率	每 100 000活产儿
	病因特异性死亡率	每100 000人
	按病因分列的标准化（年龄调整）死亡率	每100 000人
	由主要病因造成的生命年损失构成	%
	5岁以下儿童死因构成	%
	发病率	
	结核病发病率	每100 000人
	结核病发病率	每年每100 000人
	≥15岁成人的艾滋病毒感染率	每100 000人
部分传染病报告病例数	霍乱	
	白喉	
	H5N1型流感	
	日本脑炎	
	麻风病	
	疟疾	
	麻疹	
	脑膜炎	
	流行性腮腺炎	
	百日咳	
	鼠疫	
	脊髓灰质炎	
	先天性风疹综合征	
	风疹	
	新生儿破伤风	
	破伤风总计	
	结核病	
	黄热病	
卫生服务覆盖率	产前检查覆盖率	%

	由专业医务人员辅助的分娩	%
	剖腹产率	%
	新生儿出生时获新生儿破伤风保护的情况	%
	1岁以下儿童的免疫覆盖率	%
	6-59个月之间的儿童接受维生素A的补充情况	%
	5岁以下儿童睡在经杀虫剂处理的蚊帐里的比例	%
	5岁以下发烧儿童接受抗疟药物治疗的比例	%
	5岁以下具有急性呼吸道感染症状儿童被送往医疗机构的比例	%
	接受口服补液疗法的5岁以下腹泻病患儿	%
	未满足的生育需要	%
	避孕普及率	%
	为预防母婴传播，感染艾滋病毒的孕妇接受抗逆转录病毒疗法的覆盖率	%
	晚期艾滋病毒感染者接受抗逆转录病毒疗法的覆盖率	%
	DOTS策略下的结核病发现率	%
	DOTS策略下的结核病治愈率	%
风险因素	改良饮用水源的可获得性	%
	改良卫生设施的可获得性	%
	低体重新生儿	%
	出生头6个月实行纯母乳喂养的婴儿	%
	5岁以下儿童年龄别发育不良的比例	%
	5岁以下儿童中年龄别低体重的比例	%
	5岁以下儿童中年龄别超重的比例	%
	≥15岁的肥胖成年人群	%
	≥15岁成人中酒精的消耗量(每年的纯酒精升量)	%
	成人现在使用烟草的流行率 (≥15 years)	%
	青少年现在使用烟草的流行率(13 - 15 years)	%
	成人在具有高风险性的性行为中使用避孕套的流行率 (15 - 49)	%
	15-24 岁之间的人群全面正确了解艾滋病毒/艾滋病知识的比例	%
	卫生人力和基础设施	
卫生人力、基础设施、基本药物	医生的数量和密度	每10 000 人
	护士和助产士的数量和密度	每10 000 人
	牙科医生的数量和密度	每10 000 人
	社区卫生工作者的数量和密度	每10 000 人
	其他卫生服务的提供者的数量和密度	每10 000人
	病床数	每10 000 人
	基本药物	

	公立和私立机构中选定的非专利药物的中位可得性	%
	公立和私立机构中对选定的非专利药物的中间消费价格比	
卫生费用	卫生总费用占GDP的比例	
	政府总体卫生投入占卫生总费用的比例	
	个人卫生支出占卫生总费用的比例	
	政府总体卫生支出占政府总支出的比例	
	外部卫生投入占卫生总费用的比例	
	社会保障性卫生费用占政府卫生投入的比例	
	自费占个人卫生费用的比例	
	个人预付计划占个人卫生费用的比例	
	按平均汇率（美元）来计算的人均卫生总费用	
	人均卫生总费用	
	按平均汇率（美元）来计算的人均政府卫生费用	
	人均政府卫生费用	
卫生不公平	专业接生方面存在的不公平	
	一岁以下儿童麻疹免疫率的不公平	
	5岁以下儿童死亡率的不公平	每1000活产儿5岁以下儿童的死亡概率
人口与社会经济统计	人口数：总计	
	人口数：年龄中位数	年
	人口数：15岁以下	%
	人口数：60岁以上	%
	人口年增长率	%
	城区人口	%
	生命登记的覆盖率：出生率和死亡率	%
	总生育率	每名妇女
	青春期生育率	每1000 名妇女
	成人识字率	%
	小学净入学率	%
	人均国民总收入	PPP int. \$
	生活在每天 1 美元以下的人口	(PPP int. \$) (%)_

在考虑这个问题时，我们首先评判各国的医疗水平，在参考文献中给出了很多可以用来评价的指标，如上表所示。通过上表我们还可以看出一个问题，那就是题中给的评价指标过多，如果直接利用这些指标进行评价，将会是一项很烦琐的工作，因此我们认为应该首先对这些指标进行筛选，获取有用指标，进而简化对各国经济水平及医疗水平的评价过程。因此，下面我们要做的就是简化评价指标。

通过整理参考文献，我们可以得到与每个国家对应的所有指标的取值，如下表所示：

表 7.2 每个国家的相应指标的取值

国家	2.5岁以下儿童死亡率(每千例活产中5岁死亡的概 率)	一岁以下儿童麻疹疫苗接种率(%)	孕产妇死亡率(每100 000例活产)	青春 期 生 育 率 (每 1000 个 15-19 岁 之 间 的 女 孩)	每 100 000 人 中 15 岁 (含) 以 上 成 年 人 艾 滋 病 毒 感 染 率	按 照 DOTS 策 略 实 现 的 成 功 治 愈 率 (%)	安全饮 用水的 可 及 性 (%)	改良后 的卫生 设施的 可 及 性 (%)
阿尔巴尼亚	15	97	92	13	0	93	97	97
阿尔及利亚	37	92	180	4	87	91	85	94
阿富汗	257	70	1800	151	0	84	22	30
阿根廷	16	99	77	62	409	63	96	91
阿拉伯利比亚民众国	18	98	97	4	0	77	0	97
阿拉伯联合酋长国	8	92	37	23	0	79	100	97
阿拉伯叙利亚共和国	17	98	130	75	0	86	89	92
阿曼	12	97	64	11	0	86	0	0
阿塞拜疆	39	97	82	44	120	60	78	80
埃及	36	97	130	27	18	87	98	66
埃塞俄比亚	119	65	720	109	1907	84	42	11
爱尔兰	4	87	1	17	161	0	0	0
爱沙尼亚	6	96	25	20	871	68	100	95
安道尔	4	94	0	10	0	75	100	100
安哥拉	158	88	1400	165	1962	18	51	50
安提瓜和巴布达	11	99	0	67	0	0	0	0
奥地利	4	79	4	12	138	71	100	100
澳大利亚	6	94	4	15	107	85	100	100
巴巴多斯	12	75	16	51	914	100	100	99
巴布亚新几内	65	58	470	70	1395	73	40	45

亚									
巴哈马	13	96	16	43	2508	75	0	100
巴基斯坦	90	80	320	20	89	88	90	58
巴拉圭	29	80	150	65	501	83	77	70
巴林	10	99	32	15	0	86	0	0
巴拿马	23	89	130	85	810	79	92	74
巴西	22	99	110	56	510	72	91	77
白俄罗斯	7	99	18	20	158	70	100	93
保加利亚	12	96	11	38	0	80	99	99
贝宁	123	61	840	114	1161	0	65	30
比利时	4	92	8	10	172	73	0	0
冰岛	3	95	4	14	212	0	100	100
波兰	7	98	8	13	59	75	0	0
波斯尼亚和黑塞哥维那	14	96	3	16	0	97	99	95
玻利维亚	57	81	290	88	132	83	86	43
伯利兹	25	96	52	90	1863	0	0	0
博茨瓦纳	40	90	380	51	22757	72	96	47
不丹	84	95	440	46	109	89	81	52
布基纳法索	191	94	700	131	1498	73	72	13
布隆迪	180	75	1100	30	1903	83	71	41
朝鲜民主主义人民共和国	55	99	370	0	0	86	100	0
赤道几内亚	150	51	680	128	3343	0	43	51
大不列颠及北爱尔兰联合王国	6	86	8	26	154	0	100	0
大韩民国	5	92	14	2	33	81	0	0
丹麦	5	89	3	6	108	77	100	100
德国	4	94	4	10	73	0	100	100

东帝汶	97	63	380	59	0	79	62	41
多哥	100	80	510	0	3184	67	59	12
多米尼加共和国	38	96	150	98	902	78	95	79
多米尼克	14	96	0	48	0	50	0	0
俄罗斯联邦	12	99	28	28	774	58	97	87
厄瓜多尔	22	99	210	100	275	74	95	84
厄立特里亚	70	95	450	85	1264	90	60	5
法国	4	87	8	8	278	0	100	0
菲律宾	28	92	230	55	14	88	93	78
斐济	18	81	210	30	0	66	47	71
芬兰	3	98	7	9	55	0	100	100
佛得角	32	74	210	92	0	79	0	0
冈比亚	109	85	690	104	741	58	86	52
刚果	125	67	740	132	3330	53	71	20
刚果民主共和国	161	79	1100	124	0	86	46	31
哥伦比亚	20	95	130	96	490	71	93	78
哥斯达黎加	11	90	30	63	295	88	98	96
格林纳达	19	98	0	53	0	0	0	97
格鲁吉亚	30	97	66	37	75	75	99	93
古巴	6	99	45	42	67	90	91	98
圭亚那	60	96	470	90	2360	68	93	81
哈萨克斯坦	32	99	140	27	102	72	96	97
海地	76	58	670	69	1823	82	58	19
荷兰	5	96	6	4	134	0	100	100
黑山	10	90	0	16	0	0	98	91
洪都拉斯	24	89	280	108	598	86	84	66
基里巴斯	63	93	0	39	0	90	65	33
吉布提	127	74	650	27	2870	78	92	67
吉尔吉斯	38	99	150	28	113	82	89	93

斯斯坦									
几内亚	150	71	910	153	1520	75	70	19
几内亚	198	76	1100	170	1692	0	57	33
比绍									
加拿大	6	94	7	14	268	57	100	100
加纳	115	95	560	74	1722	76	80	10
加蓬	91	55	520	0	5308	46	87	36
柬埔寨	91	79	540	52	755	93	65	28
捷克共和国	4	97	4	11	17	69	100	99
津巴布韦	90	66	880	101	14609	60	81	46
喀麦隆	148	74	1000	141	4580	74	70	51
卡塔尔	10	92	12	16	0	69	100	100
科摩罗	66	65	400	95	41	0	85	35
科特迪瓦	127	67	810	111	3967	73	81	24
科威特	11	99	4	14	0	78	0	0
克罗地亚	6	96	7	13	0	30	99	99
肯尼亚	121	80	560	116	0	85	57	42
库克群岛	18	98	0	47	0	0	95	100
拉脱维亚	10	97	10	17	509	73	99	78
莱索托	84	85	960	98	21548	66	78	36
老挝人民民主共和国	70	40	660	110	149	92	60	48
黎巴嫩	29	53	150	18	101	90	100	0
立陶宛	7	97	11	19	77	74	0	0
利比里亚	133	95	1200	137	1615	0	64	32
卢森堡	4	96	12	10	0	0	100	100
卢旺达	181	99	1300	40	2342	86	65	23
罗马尼亚	14	97	24	35	77	83	88	72
马达加斯加	112	81	510	154	116	78	47	12
马尔代夫	30	97	120	8	0	91	83	59
马耳他	6	79	8	17	59	100	100	0
马拉维	110	83	1100	178	11367	78	76	60
马来西亚	11	90	62	13	428	48	99	94

亚									
马里	196	68	970	190	1437	76	60	45
马绍尔群岛	54	94	0	88	0	75	0	0
毛里求斯	17	98	15	35	1348	92	100	94
毛里塔尼亚	118	67	820	88	744	41	60	24
美利坚合众国	8	93	11	41	452	64	99	100
蒙古	43	98	46	19	52	88	72	50
孟加拉国	61	88	570	127	11	92	80	36
秘鲁	20	99	240	59	382	78	84	72
密克罗尼西亚（联邦）	40	0	0	51		90	94	25
缅甸	113	81	380	0	667	84	80	82
摩尔多瓦共和国	18	96	22	25	286	62	90	79
摩洛哥	34	95	240	18	95	87	83	72
摩纳哥	5	99	0	0	0	0	0	0
莫桑比克	168	77	520	185	11761	83	42	31
墨西哥	21	96	60	82	267	80	95	81
纳米比亚	68	69	210	51	13885	76	93	35
南非	59	83	400	54	16293	74	93	59
瑙鲁	30	99	0	69	0	100	0	0
尼泊尔	55	81	830	106	389	88	89	27
尼加拉瓜	35	99	170	109	211	89	79	48
尼日尔	176	47	1800	199	757	77	42	7
尼日利亚	189	62	1100	126	2886	76	47	30
纽埃	46	99	0	28	0	0	100	100
挪威	4	92	7	9	79	93	100	0
帕劳	10	91	0	31	0	60	89	67
葡萄牙	4	95	11	17	379	87	99	99
前南斯拉夫的 马其顿	17	96	10	19	30	87	100	89

共和国									
日本	4	98	6	5	9	53	100	100
瑞典	3	96	3	6	82	63	100	100
瑞士	5	86	5	4	398	0	100	100
萨尔瓦多	24	98	170	67	743	91	84	86
萨摩亚	27	63	0	29	0	0	88	100
塞尔维亚	8	95	0	24	79	84	99	92
塞拉利昂	262	67	2100	98	1523	87	53	11
塞内加尔	114	84	980	100	886	76	77	28
塞浦路斯	4	87	10	6	0	88	100	100
塞舌尔	16	99	0	54	0	0	0	0
沙特阿拉伯	25	96	18	7	0	69	0	0
圣多美和普林西比	99	86	0	91	0	0	86	24
圣基茨和尼维斯	18	99	0	74	0	100	99	96
圣卢西亚	13	94	0	49	0	80	98	0
圣马力诺	2	92	0	1	0	0	0	0
圣文森特和格林纳丁斯	17	99	0	57	0	0	0	0
斯里兰卡	21	98	58	28	25	87	82	86
斯洛伐克	8	99	6	21	11	81	100	100
斯洛文尼亚	4	96	6	5	29	92	0	0
斯威士兰	91	91	390	111	24301	43	60	50
苏丹	109	79	450	0	1253	82	70	35
苏里南	28	85	72	63	2066	0	92	82
所罗门群岛	70	78	220	0	0	90	70	32

索马里	142	34	1400	123	495	89	29	23
塔吉克斯坦	67	85	170	27	239	84	67	92
泰国	7	96	110	46	1191	77	98	96
坦桑尼亚 联合国 共和国	116	90	950	139	5771	85	55	33
汤加	23	99	0	16	0	100	100	96
特立尼 达和多 巴哥	35	91	45	35	1240	0	94	92
突尼斯	21	98	100	6	46	91	94	85
图瓦卢	37	95	0	22	0	75	93	89
土耳其	23	96	44	51	0	91	97	88
土库曼 斯坦	50	99	130	20	14	84	0	0
瓦努阿 图	34	65	0	0	0	90	0	0
危地马 拉	39	93	290	92	691	47	96	84
委内瑞 拉（玻 利瓦尔 共和 国）	19	0	57	91		82	0	0
文莱达 鲁萨兰 国	9	97	13	26	0	84	0	0
乌干达	131	68	550	159	5155	70	64	33
乌克兰	16	98	18	30	1082	59	97	93
乌拉圭	14	96	20	63	391	87	100	100
乌兹别 克斯坦	41	99	24	26	85	81	88	96
西班牙	4	97	4	12	370	0	100	100
希腊	4	88	3	11	115	0	100	98
新加坡	3	95	14	6	113	84	0	0
新西兰	6	79	9	29	42	70	0	0
匈牙利	7	99	6	20	39	46	100	100
牙买加	31	76	170	58	1386	41	93	83
亚美尼 亚	24	92	76	25	99	69	98	91
也门	73	74	430	80	0	83	66	46
伊拉克	45	69	300	68	0	84	77	76
伊朗(伊	33		140	25	163	83		0

斯 兰 共 和 国)									
以色列	5	97	4	15	98	74	100	0
意大利	4	87	3	7	296	0	0	0
印度	72	67	450	45	290	86	89	28
印 度 尼 西 亚	31	80	420	54	161	91	80	52
约旦	20	95	62	28	0	71	98	85
越南	15	83	150	35	446	92	92	65
赞比亚	170	85	830	146	15087	85	58	52
乍得	209	23	1500	193	3102	54	48	9
智利	9	91	16	49	244	85	95	94
中 非 共 和 国	172	62	980	133	5582	0	66	31
中国	22	94	45	5	65	94	88	65

为了对指标进行筛选，这里我们建立了主成份分析模型。

7.1 主成份分析模型

主成分分析法旨在力保原始数据信息丢失最小的情况下，对高维变量空间进行降维处理，即在保证原始数据信息损失最小的前提下，经过线性变换和舍弃部分信息，以少数的综合变量取代原有的多维变量。

设原始变量为 x_1, x_2, \dots, x_p ，进行主成分分析后得到的主成分(综合变量)为 z_1, z_2, \dots, z_m ($m < p$) 它们是 x_1, x_2, \dots, x_p 的线性组合。新变量 z_1, z_2, \dots, z_m 构成的坐标系是在原坐标系经平移和正交旋转后得到的，称 z_1, z_2, \dots, z_m 空间为 m 维主超平面。在主超平面上，第一主成分 z_1 对应于数据变异(贡献率 e_1) 最大的方向。对于 z_1, z_2, \dots, z_m ，依次有 $e_1 \geq e_2 \geq \dots \geq e_m$ 。因此， z_1 是携带原始数据信息最多的一维变量，而 m 维主超平面是保留原始数据信息量最大的 m 维子空间。

主成分分析法的步骤如下：

(1) 为了排除数量级和量纲不同带来的影响，首先对原始数据进行标准化处理：

$$x'_{ij} = (x_{ij} - \bar{x}_i) / \sigma_i, (i=1, 2, \dots, p, j=1, 2, \dots, n) \quad (7.1.1)$$

式中 x_{ij} 为第 i 个指标的第 j 个样本的原始数据； \bar{x}_i 和 σ_i 分别为第 i 个指标的样本均值和标准差。

(2) 根据标准化数据表 $(x'_{ij})_{p \times n}$ ，计算相关系数矩阵 $R = (r_{ij})_{p \times p}$ ，其中

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) / \sigma_i \sigma_j \quad (7.1.2)$$

(3)计算R的特征值和特征向量。根据特征方程 $|R - \lambda I| = 0$,计算特征根 λ_i ，并使其从大到小排列, $\lambda_1 > \lambda_2 > \dots > \lambda_p$,同时可得对应的特征向量 u_1, u_2, \dots, u_p 。它们标准正交: u_1, u_2, \dots, u_p 称为主轴。

(4)计算贡献率:

$$e_i = \lambda_i / \sum_{i=1}^p \lambda_i \quad (7.1.3)$$

及累计贡献率:

$$E_m = \sum_{j=1}^m \lambda_j / \sum_{i=1}^p \lambda_i \quad (7.1.4)$$

(5)计算主成分:

$$Z_m = \sum_{j=1}^p u_{mj} x_j \quad (7.1.5)$$

这时，各主成分相互独立。

(6)综合分析。一个 m 维主超平面究竟以多大的精度来近似代替原始变量系统，才能确保尽可能多的原始数据信息?这可以通过求累计贡献率 E_m 来判断。一般取 $E_m > 85\%$ 的最小 $m(m < P)$ ，则可得主超平面的维数 m ，从而可对 m 个主成分进行综合分析。

通过题中所给的各个指标的取值，利用上述过程可以计算出每个国家的各个主成份的取值，通过求其总得分，可以对每个国家进行排名，从而得到各个国家的医疗水平比较结果。

通过上面的过程我们就得到了每个国家的医疗水平比较，下面我们要做的就变成了每个国家的流感病毒严重情况的比较，根据前面所给的数据，我们要对每个国家的流感病毒严重情况做出比较。从附件所给的数据（参考文献）中，可以看出，流感病毒的流程度有四个标准，分别为零星活动（Sporadic activity）、当地活动（Local activity）、区域性活动（Regional outbreaks）和广泛传播活动（Widespread outbreaks）。我们认为它们对应的流感疫情的严重程度分别为1级、2级、3级、4级。由于疫情严重程度随等级增加而剧增，因而，我们可以分别用1、4、16、64作为权重来量化，然后对于2007年至2009年各国的数据进行加权求和。为了便于计算，对它们进行归一化处理，即：选取“加权和”最大的项为单位1，其它国家的“加权和”与最大的“加权和”的比值作为疫情严重程度的衡量标准。如表7.2所示。

表 7.2 各个国家的流感严重情况

国家	0	1	2	3	4	1,4,16	最终	国家	0级	1级	2	3	4	1,4,1	最终
	级	级	级	级	级	,64 加						级	级	级	6,64
						权									加权
阿富汗	98	0	0	0	0	0	0	日本	36	35	15	7	4	463	0.939148

阿尔巴尼亚	97	1	0	0	0	1	0.002028	约旦	96	0	0	3	0	48	0.097363
阿尔及利亚	95	4	0	0	0	4	0.008114	哈萨克斯坦	91	5	0	2	0	37	0.075051
南极洲	98	1	0	0	0	1	0.002028	肯尼亚	69	28	2	0	0	36	0.073022
阿根廷	60	29	3	4	3	297	0.602434	科威特	95	4	0	0	0	4	0.008114
澳大利亚	50	33	9	7	0	181	0.36714	吉尔吉斯斯坦	90	6	0	1	0	22	0.044625
奥地利	83	7	2	4	1	143	0.290061	拉脱维亚	79	17	1	2	0	53	0.107505
孟加拉国	83	15	1	0	0	19	0.03854	立陶宛	97	1	1	0	0	5	0.010142
白俄罗斯	76	15	1	3	2	195	0.395538	卢森堡	81	10	1	1	4	286	0.580122
比利时	77	17	1	1	3	229	0.464503	马达加斯加	67	29	2	0	0	37	0.075051
玻利维亚	99	0	0	0	0	0	0	马来西亚	88	7	4	0	0	23	0.046653
巴西	68	17	6	3	5	409	0.829615	毛里求斯	85	13	1	0	0	17	0.034483
保加利亚	92	5	0	0	0	5	0.010142	毛里求斯	98	1	0	0	0	1	0.002028
柬埔寨	89	10	0	0	0	10	0.020284	墨西哥	59	34	3	2	0	78	0.158215
喀麦隆	82	17	0	0	0	17	0.034483	蒙古	86	7	2	2	0	47	0.095335
加拿大	48	42	5	2	2	222	0.450304	黑山	98	0	0	0	1	64	0.129817
智利	70	23	4	2	0	71	0.144016	摩洛哥	87	12	0	0	0	12	0.024341
中国	50	32	1	1	0	108	0.219067	尼泊尔	95	4	0	0	0	4	0.008114
中国台湾	64	30	3	1	0	58	0.117647	荷兰	84	6	4	1	4	294	0.596349
中国香港	24	46	2	5	0	214	0.434077	新西兰	71	14	4	5	5	430	0.872211
哥伦比亚	95	4	0	0	0	4	0.008114	挪威	51	37	1	5	3	313	0.634888

哥斯达黎加	87	11	1	0	0	15	0.030426	阿曼	93	6	0	0	0	6	0.01217
科特迪瓦	93	6	0	0	0	6	0.01217	巴拿马	84	10	2	0	2	146	0.296146
克罗地亚	84	10	1	1	2	158	0.320487	巴拉圭	98	1	0	0	0	1	0.002028
捷克共和国	80	13	3	2	0	57	0.115619	秘鲁	60	31	6	1	0	71	0.144016
丹麦	74	20	2	0	2	156	0.31643	菲律宾	63	29	6	0	0	53	0.107505
多米尼加	98	0	0	0	0	0	0	波兰	76	22	1	0	0	26	0.052738
厄瓜多尔	92	3	3	1	0	31	0.06288	葡萄牙	76	17	1	2	2	181	0.36714
埃及	76	14	2	6	0	118	0.239351	卡塔尔	95	4	0	0	0	4	0.008114
萨尔瓦多	90	9	0	0	0	9	0.018256	大韩民国	60	24	10	3	1	176	0.356998
爱沙尼亚	90	6	1	0	2	138	0.279919	罗马尼亚	85	8	3	2	0	52	0.105477
芬兰	79	13	3	2	0	57	0.115619	俄罗斯联邦	47	29	4	1	4	493	1
法国	59	33	3	2	1	141	0.286004	塞内加尔	79	19	1	0	0	23	0.046653
瓜德罗普岛	96	1	1	1	0	21	0.042596	塞尔维亚	88	10	1	0	0	14	0.028398
马提尼克岛	94	4	1	0	0	8	0.016227	塞舌尔	97	2	0	0	0	2	0.004057
新喀里多尼亚(岛)	93	4	2	0	0	12	0.024341	新加坡	51	39	8	0	0	71	0.144016
法国留尼旺	97	2	0	0	0	2	0.004057	斯洛伐克	90	5	4	0	0	21	0.042596
圭亚那	88	11	0	0	0	11	0.022312	斯洛文尼亚	78	17	2	0	2	153	0.310345
格鲁	91	7	0	0	0	7	0.0141	南非	61	31	4	2	1	143	0.29006

吉亚							99								1
德国	58	32	3	3	1	156	0.3164	西班牙	71	20	2	3	1	140	0.28397
							3								6
加纳	84	15	0	0	0	15	0.0304	斯里兰卡	77	21	0	0	0	21	0.04259
							26								6
希腊	78	16	3	0	0	28	0.0567	苏里南	98	1	0	0	0	1	0.00202
							95								8
危地马拉	86	11	2	0	0	19	0.0385	瑞典	57	38	0	0	2	166	0.33671
							4								4
洪都拉斯	88	11	0	0	0	11	0.0223	瑞士	72	20	2	0	3	220	0.44624
							12								7
匈牙利	83	10	1	4	0	78	0.1582	泰国	41	48	8	0	0	80	0.16227
							15								2
冰岛	87	11	1	0	0	15	0.0304	特立尼达和多巴哥	98	1	0	0	0	1	0.00202
							26								8
印度	74	25	0	0	0	25	0.0507	突尼斯	77	12	4	5	0	108	0.21906
							1								7
印度尼西亚	93	6	0	0	0	6	0.0121	土耳其	89	7	2	0	0	15	0.03042
							7								6
伊朗	73	23	2	0	0	31	0.0628	乌干达	81	16	2	0	0	24	0.04868
							8								2
爱尔兰	95	3	0	0	1	67	0.1359	乌克兰	83	10	2	1	1	98	0.19878
							03								3
以色列	82	11	4	2	0	59	0.1196	英国	52	40	3	1	1	132	0.26774
							75								8
意大利	71	23	3	0	1	99	0.2008	蒙特塞拉特岛	98	1	0	0	0	1	0.00202
							11								8
乌拉圭	82	15	0	0	2	143	0.2900	美利坚合众国	31	41	14	7	4	465	0.94320
							61								5
委内瑞拉	98	1	0	0	0	1	0.0020	关岛	96	3	0	0	0	3	0.00608
							28								5
越南	80	15	3	0	0	27	0.0547								
							67								

就这样，根据上文中的分析，我们就可以得到各个国家根据经济及医疗水平的排名，和依据流感严重程度的排名，综合上面每个国家的医疗水平情况及流感病毒情况，我们可以得到最终的一个国家的综合排名，进而决定最终的援助方案。这里我们在综合各国经济及医疗水平与流感疫情的排名时，我们希望通过给流感疫情和经济医疗水平，赋一定的权重之后累计相加，得到最后的总排名，由排名情况即可得到各个国家接受援助

的优先级，如下表所示：

表 7.3 接受援助国家优先级（竖排）

国家	国家	国家	国家
巴拿马	匈牙利	罗马尼亚	厄瓜多尔
南非	秘鲁	约旦	伊朗(伊斯兰共和国)
乌拉圭	新加坡	蒙古	希腊
爱沙尼亚	爱尔兰	马达加斯加	越南
突尼斯	黑山	肯尼亚	印度

仔细观察主成份分析模型的过程，我们可以发现即使通过主成份分析降维之后，得到的指标仍然是很多的，为了能够达到更大的降维效果，我们认为熵值法模型更能起到效果，因此我们决定采用熵值法模型进行求解。

7.2 熵值法模型

熵是系统无序程度的度量，可以用于度量已知数据所包含的有效信息量和确定权重，在水质评价中得到了广泛的应用。在水质模糊评价中，通过对“熵”的计算确定权重，就是根据各项监测指标值的差异程度，确定各指标的权重。当各评价对象的某项指标值相差较大时，熵值较小，说明该指标提供的有效信息量较大，其权重也应较大；反之，若某项指标值相差较小，熵值较大，说明该指标提供的信息量较小，其权重也应较小。当各被评价对象的某项指标值完全相同时，熵值达到最大，这意味着该指标无有用信息，可以从评价指标体系中去除。

使用熵权法确定权重主要有以下 3 个步骤：

(1)原始数据矩阵归一化。设 m 个评价指标、 n 个评价对象的原始数据矩阵为

$A = (a_{ij})_{m \times n}$ ，对其归一化后得到 $R = (r_{ij})_{m \times n}$ ，对小者为优的指标而言，归一化公式为：

$$r_{ij} = \frac{\max\{a_{ij}\} - a_{ij}}{\max\{a_{ij}\} - \min\{a_{ij}\}} \quad (7.2.1)$$

对大者为优的指标而言，归一化公式为：

$$r_{ij} = \frac{a_{ij} - \min\{a_{ij}\}}{\max\{a_{ij}\} - \min\{a_{ij}\}} \quad (7.2.2)$$

(2)定义熵。在有 m 个指标、 n 个被评价对象的评估问题中，第 i 个指标的熵为：

$$h_i = -k \sum_{j=1}^n f_{ij} \ln f_{ij} \quad (7.2.3)$$

式中： $f_{ij} = r_{ij} / \sum_{j=1}^n r_{ij}$ ； $k = 1 / \ln n$ ；当 $f_{ij} = 0$ 时，令 $f_{ij} \ln f_{ij} = 0$ 。

(3)定义熵权。定义了第 i 个指标的熵之后，可得到第 i 个指标的熵权：

$$w_i = \frac{1-h_i}{m-\sum_{i=1}^m h_i} (0 \leq w_i \leq 1, \sum_{i=1}^m w_i = 1) \quad (7.2.4)$$

通过 Matlab 编程求解，我们可以得到最终筛选出的评价指标。通过前面的分析及求解，我们已经利用熵值法得到了经过筛选的评价指标，接下来我们就可以利用主成份分析对各个国家的这些因素进行综合排名，再利用各个国家流感疫情的严重程度，可得到最终的援助方案，如下表所示：

表 7.4 接受援助的国家优先级(竖排)

国家	国家	国家	国家
南非	墨西哥	罗马尼亚	加纳
乌拉圭	乌干达	约旦	冰岛
巴拿马	新加坡	蒙古	厄瓜多尔
乌克兰	爱尔兰	孟加拉国	越南
泰国	黑山	危地马拉	印度

八、模型的拓展

在问题一的模型中，我们只考虑了单个国家自己的病毒传播问题，并未考虑到国家之间的相互关系，为了更符合实际的研究这个问题，我们将加入国家之间的相互影响因素进行问题一的研究。

8.1 模型的建立

首先为研究相邻国家间的影响，我们选取了欧洲部分邻近国家作为样本。为了便于讨论，我们对疫情情况按其级别赋予权重。考虑到疫情级别是以其影响范围而定，我们给疫情权重赋值为疫情级别的平方，即：

$$s_i = l_i^2, \quad (8-1)$$

其中 s_i 为疫情权重， l_i 为疫情级别。

分别就总体疫情情况和亚型情况用 SPSS 软件分析国家间相互影响的相关系数(结果见附录 n)。

从数据可以看出，疫情总体情况和亚型疫情情况在国家间相互影响的皮尔森相关系数绝大多数都在 0.600 以上，因此我们可以认为疫情有着很明显的地域相互影响性。

下面我们讨论在给定某地区疫情的情况下，该地区疫情对周边地区的影响范围。

选取某一国家 A 作为对照国，另取若干个气候条件与对照国近似(纬度接近)，但和对照国距离各不相同的若干组国家作为比较国组，研究它们与对照国组疫情相互影响的相干系数，在此基础上进一步讨论疫情与影响范围的关系，从而得出在一定概率条件下的疫情影响范围。得到数据表格如下：

表 8.1 一定概率条件下的疫情影响范围

相关系数均值	0.771	0.726	0.331	-0.047	0.737	0.689	0.243
距离(km)	8400	10500	11800	12600	8400	10500	11800
参照国的亚型疫情权重和	67	67	67	67	47	47	47

我们取相关系数在 0.600 以上为有很明显的线性相关性，由此去除第 3、4、7 组数据，于是剩下四组数据：

表 8.2 明显相关的疫情影响范围

相关系数均值	0.771	0.726	0.737	0.689
距离 (km)	8400	10500	8400	10500
参照国的亚型疫情权重和	67	67	47	47

接下来我们从这四组数据出发，讨论疫情影响范围与疫情权重和相关系数的关系。

首先讨论疫情影响范围与相关系数的关系。因为之前所讨论的是线性相关性，因此我们用 SPSS 软件对这四组数据中的距离与疫情权重和进行线性拟合。为了讨论方便，我们先不考虑疫情权重的差别，于是得到以下结果：

$$\rho_1 = 0.940 - 0.000022d \quad (8-2)$$

于是我们可以求出，若要满足 ρ 的值大于 0.900，则应当有 d 的值应小于 1818.2km。

这个数据仅是基于一组数据求出的结果，精确度较低，但由此我们可以大致划出存在较大可能的线性相关性的区域范围。

接下来我们将疫情权重也考虑进去，用 SPSS 软件对上述数据进行二元线性拟合，得到以下结果：

$$\rho_2 = 0.839 - 0.000022d + 0.002S \quad (8-3)$$

其中 S 为疫情权重和。为了讨论疫情权重和与影响范围的关系，我们现将 ρ_2 固定，于是有：

$$d = \frac{0.839 - \rho_2 + 0.002S}{0.000022} \quad (8-4)$$

于是对于给定要求的 ρ_2 范围和 S ，我们就可以得出一个相应的影响范围。

基于以上讨论，我们可以建立以下模型：

1) 对于发生过四级疫情的区域，我们认为它给周边地区带来的影响十分强烈，我们将这些地区单独取出作为一类，不妨称之为中心点区。

2) 对上述中心点区，我们利用式 (1) 中求出的 d 的结果（这里为了提高其可信度，我们可以将 d 的值缩小至 1000km；事实上考虑到权重的影响，这样做是有着较高可信度的，因为这里是对发生四级疫情的区域讨论，应该说其影响范围是较大的），以中心点区为圆心， d 为半径划出一个圆形区域，若这个圆形区域内包含了其它的中心点区，我们就将这些中心点区合为一个较大的区域，我们称这样的区域为中心区。

3) 接下来利用 (2) 式，对于某一中心区和给定的 ρ 、 S （为了更切合实际，这里 S 我们取这一中心区内国家同种病毒类型引起的疫情情况的权重和的平方根），我们可以求出一个对应的 d_i ；再以中心区的地理中点为圆心， d_i 为半径，可以划出另一更大的圆形区域，这个区域是以 ρ 为相关系数受中心区某一类型病毒线性影响的区域，我们称这个区域为该种病毒的危险区，某种病毒的危险区内就是有较大可能性需要该种病毒疫苗

的区域。通过改变 ρ 的大小，我们可以相应改变 di 的大小，也就是危险区的大小。当 di 足够大时，不同危险区间就会发生相交，不同病毒类型的危险区产生的相交区域即是需要不同种类疫苗的区域。

4) 由于危险区是圆形区域，可能难以覆盖到每个地方。对于这种情况，可以适当减小 ρ ，使 di 足够大，就可以覆盖到足够大的区域。

5) 增大 di ，可能出现的另一种情况是：某一区域被两个以上不同病毒类型危险区覆盖（这里由于需要考虑双联装，所以讨论两个以上的情况）。对于这种情况，可以建立以下子模型：设这个区域距离覆盖它的第 i 个危险区的中心区的距离为 di' ，该中心区的权重和为 ρ_i ，记 j_i 为 di' 与 ρ_i 的比值，比较 j_i 大小，取其值较小的病毒类型。

6) 还有一种情况是，在一定相关系数 ρ 的要求的前提下，某些地区仅被不到两个类型的不同病毒的危險区覆盖，但又还是有发生疫情的可能，这种情况下，为了满足双联装的要求，可以进一步划出三级疫情的危險区，若仍未达到预期覆盖要求，则继续向下划分；也可以取时间较近的优先等。

例如我们利用上述模型和WHO给出的2007年9月~2009年1月的数据分析亚欧大陆（为简单起见，这里我们不考虑其他大洲的影响），可以得到以下分析结果：

1) 发生四级疫情国家的疫情情况（中心点区）

表 7.3 发生四级疫情国家的疫情情况

国家	疫情发生时间	病毒类型	国家	疫情发生时间	病毒类型
Belgium	2008.1~2008.2	H1	Japan	2009.1	H1/H3
Metenegro	2008.1	H1	Belgium	2009.1	H3
Slovenia	2008.1	H1	Czech Republic	2008.12~2009.1	H3
Switzerland	2008.1~2	H1	Estonia	2009.1	H3
Japan	2008.2	H1	France	2009.1	H3
Republic of Korea	2008.2	B	Germany	2009.1	H3
Austria	2008.2	H1	Italy	2009.1	H3
Belarus	2008.2	H1/H3	Luxembourg	2009.1	H3
Croatia	2008.2~2008.3	H1	Netherlands	2008.12~2009.1	H3
Estonia	2008.2	H1	Portugal	2008.12~2009.1	H3
Luxembourg	2008.1~2008.3	H1/B	Portugal		
Netherlands	2008.2~2008.3	H1/B	Slovenia	2009.1	H3
Norway	2008.2	H1	Spain	2009.1	H3
Russian Federation	2008.3	H1/H3	Swiden	2009.1	H3
Ukraine	2008.2	H1	Switzerland	2009.1	H3
			UK	2008.1	H3

2)中心区划分

依据公式 (1)，取 $d=1000\text{km}$ ；考虑到俄罗斯人口主要集中在欧洲地域，将俄罗斯与欧洲的中心区划在一起，则中心区划分为两个，即 Japan 和 Japan 以外的欧洲国家。于是得到如下图中心区划分：



3) 讨论

毫无疑问在日本附近是需要选择 H1 和 H3 疫苗的，那么接下来就要讨论在欧洲区应选取什么疫苗，这也就是需要比较欧洲区三种病毒影响力的大小，利用 (2) 式不难得出 B 的影响力是最小的，那么也就是说在欧洲区附近应当选择 H1 和 H3 这两种疫苗；最后需要讨论在欧洲与日本间如何选择疫苗。事实上在欧洲区和在日本取都是 H1 和 H3 影响最大，在亚欧大陆只考虑自身影响的前提下，两中心区间也必定受 H1 和 H3 影响最大，由此可得在亚欧大陆上都是选择 H1 和 H3 的双联装疫苗。

8.2 模型分析

(1) 该模型是建立在对国家间相互影响的数据分析的基础上的，将地球看成一个二维的点集，通过影响范围的分析而划分点集的子域，达到合理预测疫苗种类的目的，具有较高的可行性。

(2) 在数据处理方面，数据分析仅是基于线性关系，这就可能影响了数据分析的准确度；此外优先分析四级疫情可能造成三级疫情被忽略，而大量的三级疫情其影响可能反而大于一起四级疫情。

(3) 对于一些较大的国家，这种模型仅有全国的统计资料可能不够，为提高准确度，需要有更精细的数据。

(4) 疫情的影响范围和疫情持续时间的长短也会有关联，这个模型没有将疫情持续时间的状况考虑进去。

(5) 在参数 S 选择上还需要经过数据的分析以更切合实际。

九、模型的推广

本文中，我们使用了加权统计模型、贝叶斯预测模型、评估模型及筛选模型。这些模型在一般的问题中都是经常遇到的，生活中我们经常要统计一个整体量，但是这个整体是由很多分量表示的，而且这些分量的重要程度又各不相同，这样就引入了加权计算。例如我们最熟悉的学分的计算就是加权平均，最后所有总分的计算也是加权以后的结果。尽管面对不同问题，权重可能会有所不同，但是原理是相通的。

再比如说贝叶斯预测模型，这是一个预测类的模型，面对其他预测类问题时，也可以用此模型很好的做出解答，比如说在地震的预测问题上。评估模型是确定一个评估标准后对两种方法或方案进行的一个评价对比，生活中的问题并不是只有唯一的正确答案，很多时候答案并不唯一，这时就需要我们挑选出比较好的方案，而怎样挑选方案呢，这就需要对这些方案进行评估，根据评估结果进行选择。

十、模型的优缺点分析

10.1 模型的优点

- (1) 加权统计模型中权重的处理很巧妙而且还很贴切，大大简化了赋权值的过程；
- (2) 使用贝叶斯预测模型确定下年疫苗种类，具有创新性；
- (3) 概率评价模型确立的评价标准很合理；
- (4) 在确定投放方案时考虑到国与国之间的作用，分析全面；
- (5) 使用的各种模型均具有可推广性；
- (6) 筛选模型利用了病毒专家在研究下一年的疫苗时所采用的方法，模型更具科学性；
- (7) 占有的数据量大，分析结果接近实际情况；

10.2 模型的缺点

- (1) 在加权统计模型的统计过程中，我们没有再具体的讨论相同严重程度情况下的两个病毒的选取，
- (2) 筛选模型太过于理论化，没有一些计算上的推导；
- (3) 贝叶斯预测模型的实际操作繁琐，本文使用 EXCEL 软件处理仍花费了大量的工夫。若能够编写出程序简化工作量会更好；

十一、参考文献

- [1] 盛骤，谢式千等，概率论与数理统计，杭州：高等教育出版社，2008 年；
- [2] 张文彤，SPSS11 统计分析教程，北京：北京希望电子出版社，2002 年；
- [3] 张兴永，Matlab 软件与数学实验，徐州：中国矿业大学出版社，2006 年；
- [4] 刘先勇等，SPSS10.0 统计分析软件与应用，北京：国防工业出版社，2002 年；
- [5] 百度知道，全国县市有多少，<http://zhidao.baidu.com/question/7375199.html>，2009 年 4 月 26 日
- [6] 青岛传媒网，市疾控中心公布今年流感疫苗种类价格，http://news.qingdaomedia.com/shownews_a.asp?articleid=159674，2009 年 4 月 26 日
- [7] 建议组成的流感病毒疫苗的使用在 2009 年-2010 年北半球流感季节
http://www.who.int/csr/disease/influenza/200902_recommendation.pdf，2009 年 4 月 25 日

[8] 建议组成的流感病毒疫苗的使用在 2009 年南半球流感季节

<http://www.who.int/csr/disease/influenza/200809Recommendation.pdf>, 2009 年 4 月 25 日

[9] 建议组成的流感病毒疫苗用于在 2008-2009 年北半球冬季流感季节

http://www.who.int/csr/disease/influenza/recommended_compositionFeb08FullReport.pdf, 2009 年 4 月 25 日

[10] 建议组成的流感病毒疫苗的使用在 2008 年南半球流感季节

<http://www.who.int/csr/disease/influenza/recommendationlong2.pdf>, 2009 年 4 月 25 日

[11] Google 地图, <http://ditu.google.cn/maps?hl=zh-CN&tab=wl>, 2009 年 4 月 25 日

附录

相关系数

		USA
France	皮尔森相关系数	.838(**)
	双尾T检验结果	.000
	样本数	18
Spain	皮尔森相关系数	.715(**)
	双尾T检验结果	.001
	样本数	18

**： 双尾显著性水平1%条件下的结果。

相关系数

		USA
Romania	皮尔森相关系数	.822(**)
	双尾T检验结果	.000
	样本数	18
Bulgaria	皮尔森相关系数	.629(**)
	双尾T检验结果	.005
	样本数	18

**： 双尾显著性水平1%条件下的结果。

*： 双尾显著性水平5%条件下的结果。

相关系数

		USA
Iran	皮尔森相关	.429

		系数
Kuwait	双尾T检验结果	.075
	样本数	18
	皮尔森相关系数	.(a)
Qatar	双尾T检验结果	.
	样本数	18
	皮尔森相关系数	.232
	双尾T检验结果	.354
	样本数	18

* 双尾显著性水平5%条件下的结果。

** 双尾显著性水平1%条件下的结果..

相关系数

		USA
Mongolia	皮尔森相关系数	-.156
	双尾T检验结果	.536
	样本数	18
China	皮尔森相关系数	-.360
	双尾T检验结果	.143
	样本数	18
Korea	皮尔森相关系数	.375
	双尾T检验结果	.125

样本数	18
-----	----

相关系数

		USA
France	皮尔森相关系数	.671(**)
	双尾T检验结果	.002
	样本数	18
SPAIN	皮尔森相关系数	.802(**)
	双尾T检验结果	.000
	样本数	18

** 双尾显著性水平1%条件下的结果

* 双尾显著性水平5%条件下的结果.

相关系数

		USA
UKRAINE	皮尔森相关系数	.767(**)
	双尾T检验结果	.000
	样本数	18
ROMANIA	皮尔森相关系数	.611(**)

双尾T检验结果	.007
样本数	18

**：双尾显著性水平1%条件下的结果

*：双尾显著性水平5%条件下的结果.

a：因至少有一个变量为常数而无法计算.

相关系数

		美国
IRAN	皮尔森相关系数	-.065
	双尾T检验结果	.798
	样本数	18
KUWAIT	皮尔森相关系数	.280
	双尾T检验结果	.261
	样本数	18
QATAR	皮尔森相关系数	.513(*)
	双尾T检验结果	.029
	样本数	18

* 双尾显著性水平5%条件下的结果.

** 双尾显著性水平1%条件下的结果

熵权法代码：

```

m=[p(1,1) p(2,1) p(3,1) p(4,1) p(5,1) p(6,1) p(7,1) p(8,1) p(9,1) p(10,1) p(11,1) p(12,1) p(13,1)
p(14,1) p(15,1) p(16,1) p(17,1) p(18,1) p(19,1) p(20,1) .....p(193,1)];%每一行的最大值
for i=1:37
    for j=1:193
        if m(i)<p(i,j)
            m(i)=p(i,j);
        end
    end
end
n=[ p(1,1) p(2,1) p(3,1) p(4,1) p(5,1) p(6,1) p(7,1) p(8,1) p(9,1) p(10,1) p(11,1) p(12,1) p(13,1)
p(14,1) p(15,1) p(16,1) p(17,1) p(18,1) p(19,1) p(20,1) .....p(193,1)];%每一行的最小值
for i=1:37
    for j=1:193
        if n(i)>p(i,j)
            n(i)=p(i,j);

```