



## 我国区域传染病状况分析<sup>1</sup>

---

学校：湖南农业大学

院系：理学院信息与计算机科学系

专业：统计学

作者：刘影 刘文妮 康艺龄

指导老师：王访

---

<sup>1</sup> 注:该论文获得由中国统计教育学会举办的“2015 年（第四届）全国大学生统计建模大赛”大数据统计建模类本科生组二等奖。

# 我国区域传染病状况分析

**摘 要：** 传染病是世界范围内造成死亡人数最多的疾病，也是我国死亡率最高的三类疾病之一( 心脑血管疾病、恶性肿瘤和传染性疾病 )，每年全世界约有 1640 万人死于传染性疾病。医学技术的进展使传染病的预防与控制取得了重大进展。在我国，传染性疾病的年死亡率为 43.2/100000。

本文首先利用 SPSS 21.0 软件以发病率、死亡率及病死率为指标对全国各省份进行聚类，依据每种指标分别将全国城市归为三类：高发病率、中发病率、低发病率；高死亡率、中死亡率、低死亡率；高病死率、中病死率、低死亡率，通过表格以及图像方式清晰明了地给出具体分类情况，再从宏观政策等角度挖掘产生这种现状的可能原因。

而后，利用 SPSS 21.0 软件分别对 2004-2013 年我国城镇和农村传染病死亡率时序数据进行三次方程拟合，并预测出 2014-2015 年城镇和农村传染病死亡率数据，同理，分别对男性、女性传染病死亡率时序数据进行拟合并进行预测得出：城镇居民传染病死亡率略高于农村居民传染病死亡率、男性传染病死亡率略高于女性传染病死亡率，且两者间差距正在不断缩小。

最后，从东部、中部和西部地区中分别选取三个典型省份作为代表构建 2004-2013 年 9 个地区的面板数据，通过协整检验等方法构建相应的个体固定效应模型，从而研究各省份在 2004 年至 2013 年 10 年间传染病的现状，通过模型检验以及真实值与预测值的比较分析，发现模型拟合优度较高且具有较好的解释性，从而分析我国东部、中部和西部的传染病蔓延情况，以便改善各地区传染病蔓延从而预防传染病的快速传播或突然爆发。

**关键词：** 传染病 聚类分析 回归拟合 面板数据建模

## 一、 研究背景

### （一） 问题描述

2015 年伊始，各大传染病频繁出现于各类新闻报道：香港流感卷土重来，美国麻疹疫情远超去年同期记录，而来自遥远西非大陆的埃博拉继续出现在 WHO 的新闻发言中。

根据国家卫生计生委通报显示，2015 年 1 月，全国（不含港澳台，下同）共报告法定传染病 497479 例，死亡 1271 人；2 月，共报告法定传染病 338939 例，死亡 855 人；3 月，共报告法定传染病 484896 例，死亡 1454 人；4 月，共报告法定传染病 577783 例，死亡 1218 人。

法定传染病背后庞大的死亡病例数据，说明了传染病防护和预测的重要地位。由于传染病的发生和蔓延具有很强的地域性和时序性，所以理当用精准有力的研究方法对各地区的传染病发病情况进行深入研究与探讨。

### （二） 国内外研究现状

国内外学者对传染病相关因素、趋势模型的构建和发病率的时序分析及预测进行了大量研究，并有相关学者对传染病的预测模型进行分析。

樊雯婧<sup>[1]</sup>等用 ARIMA 模型预测了疟疾的发病趋势，且得到了平均误差为 2.57% 的外推预测结果。同时，杨小兵<sup>[2]</sup>等构建时间序列分析 ARIMA 乘积季节性模型预测流行性腮腺炎发趋势。张娟等<sup>[3]</sup>利用 Heesterbeek 和梅斯推导出的个别接触饱和和接触率的流行病模型的表达式，对传染病饱和和接触率的 SEIR 模型进行了研究。杨志军<sup>[4]</sup>使用马尔科夫决策方法建立了确定慢性传染病治疗监测最优策略的数学模型，同时也建立了确定慢性传染病最佳治疗监测时间的数学模型，还建立了对慢性传染病治疗检测技术进行经济学分析的数学模型。李甜甜<sup>[5]</sup>通过分析微分方程的动力稳定性，判断传染病的传播机理，从而有效地阻止了传染病的传播。李冰<sup>[6]</sup>通过建立疾病传播依赖染病者的饱和和发生率，并带有垂直感染的脉冲接种 SIR 传染病模型，利用频闪映射，得到模型的无病周期解，并利用脉冲微分不等式，证明无病周期解的全局稳定性，从而研究具有脉冲接种的多类传染病模型，分析模型的动力学行为。Zhihang Peng<sup>[7]</sup>等人首先应用聚类分析的方法建立在未知的情况下的传染病分类标准。其次，使用加权马尔科夫链的方法对未来传染病情况做出了预测。最后通过马尔可夫链蒙特卡罗方法做出最优的长期利益，从而分析传染病的特点。

综上所述，国内传染病的研究主要使用的 ARIMA 模型和普通的线性回归模

型。本文采用面板模型进行分析，更具有时效性以及全面性。

## 二、 研究思路及方法

### （一） 研究思路

本文通过对全国传染病发展趋势及现状的整体统计研究，以发病率、病死率和死亡率三个指标作为统计指标，大体上揭示我国各省、城乡以及男女间传染病分布的空间特征，从而有效地认识到我国传染病分布的内在规律及其动态演变过程，为我国传染病分布趋势的预测及防治提供合理的参考。

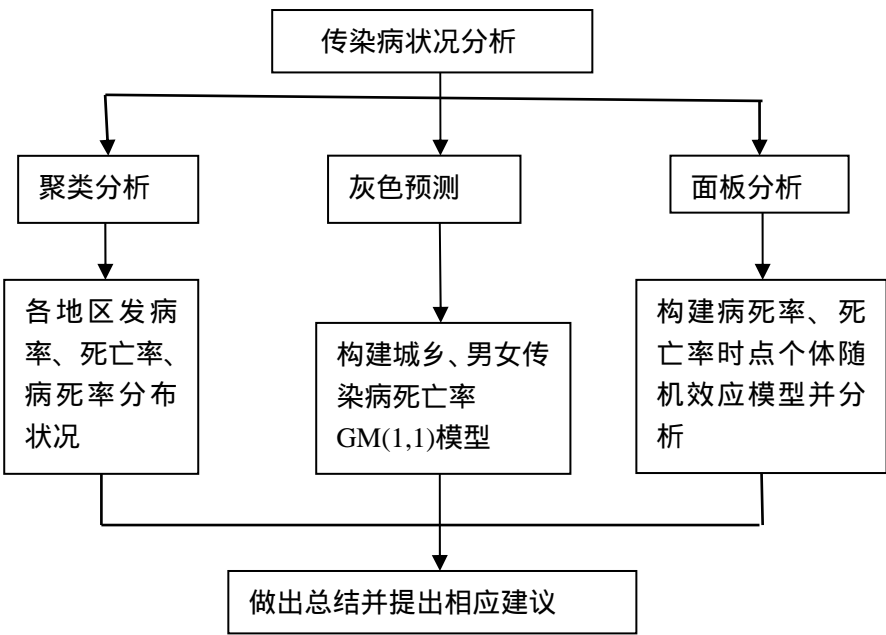


图 1 研究思路流程图

### （二） 研究方法

1. 本文利用 K 均值聚类分析，对各地区进行聚类，得出在特定个别年份各省份传染病的蔓延程度强弱。

2. 本文利用灰色预测原理，构建 2004-2011 年城乡及男女传染病死亡率灰色预测模型并预测出 2012-2013 年的相应死亡率，同时将预测值与真实值进行比较，以判断模型精度，随后对 2014-2015 年死亡率进行预测。

3. 本文利用格兰杰因果检验、面板数据分析方法，根据 10 年 31 个地区的面板数据，建立了各个地区的个体时点随机效应模型。

### 三、数据来源及指标确定

文中所用数据均来自 2003 年至 2013 年中国卫生统计年鉴。文中所用的指标分别是发病率（F）、死亡率（S）以及病死率（B）：

发病率（十万分之一）是指一定期间内某病新病例的发生频率<sup>[8]</sup>。

$$\text{发病率} = \frac{\text{观察期间内发现的新病例数}}{\text{同期平均人口数}}$$

死亡率（十万分之一）是指在一定期间一定人群中，死于某病的频率，是测量人群死亡危险最常用的指标，用于衡量一定期间某地区人群死亡危险性大小的一项指标<sup>[8]</sup>。

$$\text{死亡率} = \frac{\text{某时期内（因某病）死亡总数}}{\text{同期平均人口数}}$$

病死率（百分之一）是指一定时期内（通常为 1 年）患某病的全部患者中因该病死亡者的比例<sup>[8]</sup>。

$$\text{病死率} = \frac{\text{某时期内因某病死亡人数}}{\text{同期患某病的患者数}}$$

### 四、聚类分析

#### （一）方法介绍

聚类分析是指将个体或对象分类，使得同一类中的对象之间的相似性比其他类的对象的相似性更强。通过聚类分析，可以将研究目的、专业知识和数据特征相结合，合理地把数据分成若干个类别，聚类分析的实质就是按照距离的远近将数据分为若干个类别，使得类别内部的差异尽可能小，类别间差异尽可能大。K-means 算法是很典型的基于距离的聚类算法，采用欧式距离作为相似度测度，它是求对应某一初始聚类中心向量的最优分类，使得评价指标求得最小。即认为两个对象的距离越近，其相似度就越大。该算法认为簇是由距离靠近的对象组成的，因此把得到紧凑且独立的簇作为最终目标<sup>[9]</sup>。

欧式距离也称欧几里得距离，它描述  $p$  维空间中两点间的真实距离。任意两个点  $P = (x_1, x_2, \dots, x_p)$  与  $Q = (y_1, y_2, \dots, y_p)$  之间的欧式距离为<sup>[10]</sup>：

$$d(P, Q) = \sqrt{(x_1 - y_1)^2 + \dots + (x_p - y_p)^2} \quad (1)$$

## （二）实例分析

由图 2 所展示的 2004-2013 年传染病发病率、死亡率以及病死率变化趋势可知，10 年间传染病发病率以及病死率均呈现较平稳的发展趋势，死亡率增长也相对平稳。

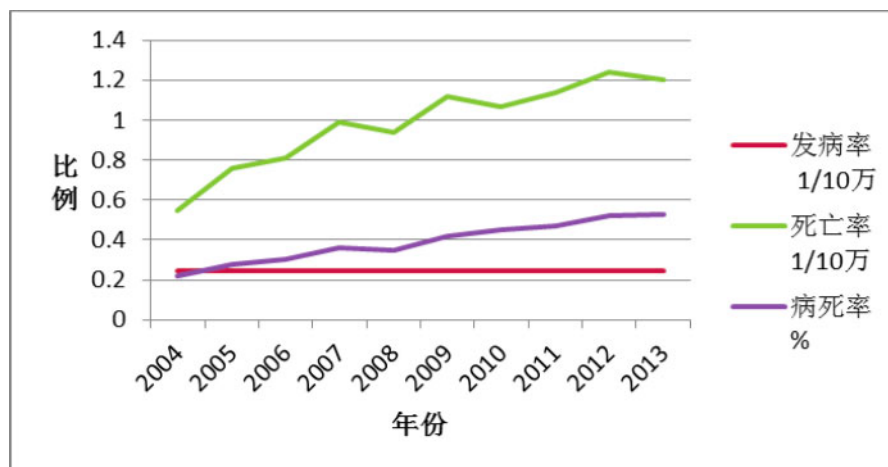
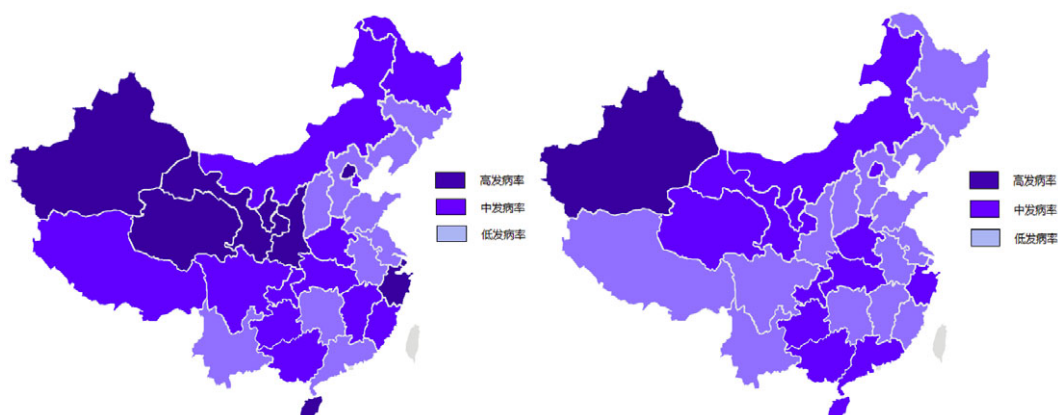


图 2 2004-2013 发病率、死亡率、病死率

故此，本文从 10 年中等距离抽取 4 年：2004 年、2007 年、2010 年和 2013 年，并用 SPSS 软件根据这四年的传染病发病率、病死率以及死亡率情况对全国各个省份采用欧式距离作为相似度测度进行 k-means 聚类，分析各地区传染病发展情况。最后，根据 SPSS 中的聚类分类情况，借用 Photoshop 绘图软件，分别用蓝、绿、红三种颜色标示出全国各省份传染病发病率、死亡率、病死率状况，并加以简要说明。

### 1. 发病率的聚类分析



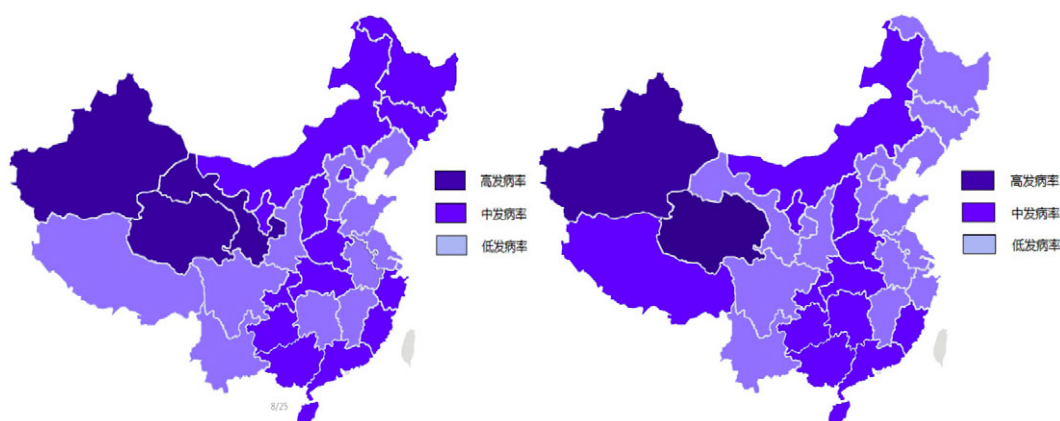
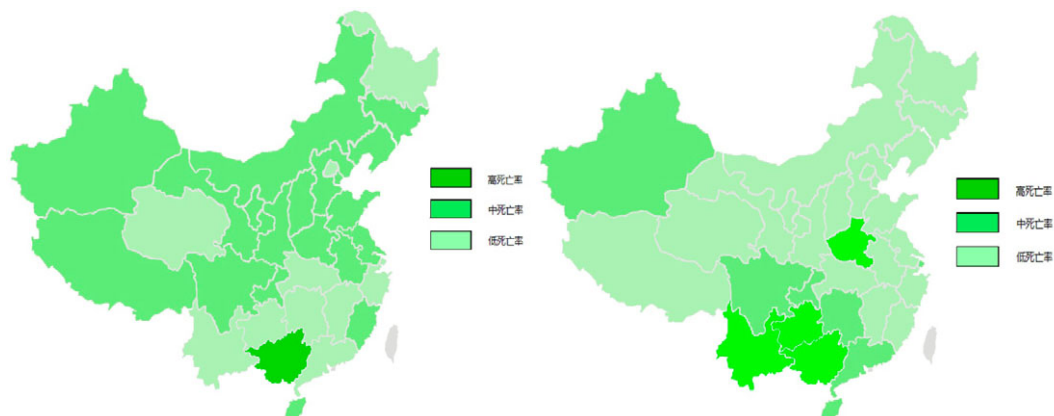


图3 2004、2007、2010、2013年全国发病率分布状况

由图3可知：全国发病率分类状态呈现出明显的区域性现象。2004年，我国多数地区均被分类为发病率高和较高两类，经分析发现，这是由于H5N1禽流感疫情的爆发所导致；2007年，我国较少地区属于发病率相对较高的簇，主要是因为该年全国经历的重大灾害性天气较少，加之中央对传染病加强重视，在第十届全国人民代表大会第31届会议颁布了《中华人民共和国国际卫生检疫法》修订的决议；2010年，我国总体传染病控制情况较好，故该年被分为较高发病率簇的省份均位于我国西北部；至2013年，相对较高发病率的簇集聚在我国西北部，中发病率类别大多于中部，而被分为发病状况相对良好类别的地区主要于东、西部。

总体而言，发病率较高的簇基本集中于我国西北部，其中新疆的死亡率在4年间一直保持高发病率状态；此外，持续被分至中发病率簇的省份主要是蒙古、河南、湖北、重庆、贵州、广西等部分北部、中部和南部地区，同时从2010年到2013年浙江、福建、广东等沿海地区从同年较低发病率簇转向中发病率簇。

## 2. 死亡率的聚类分析



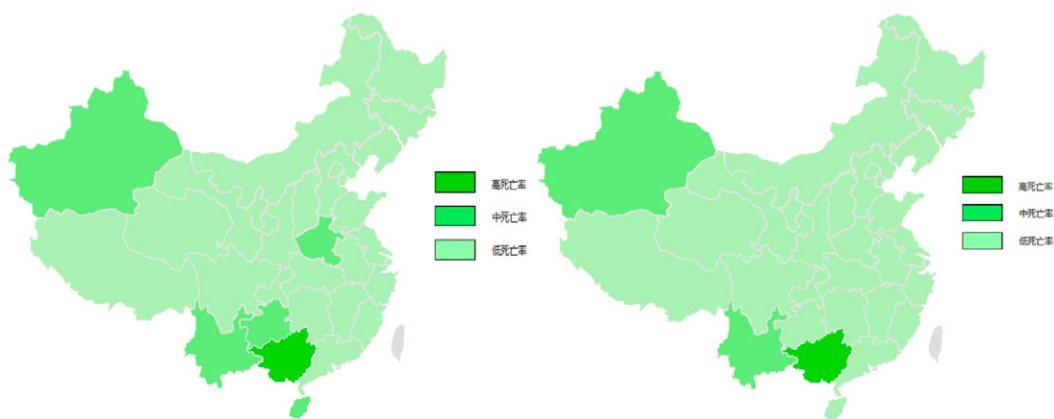
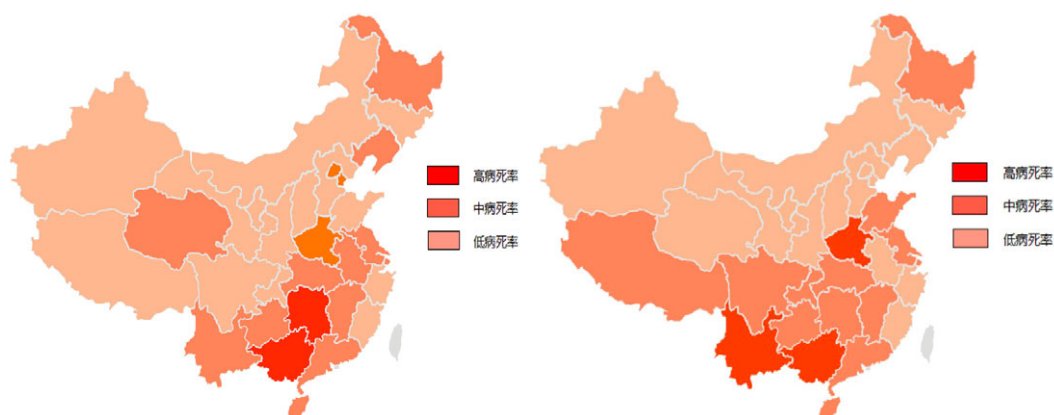


图4 2004、2007、2010、2013年全国死亡率分布状况

由图4可知：2004年，全国多数地区被分至死亡率中等水平的组中，唯有广西壮族自治区属于死亡率较高的一类，当年H5N1疫情灾害泛滥导致；2007年中部的河南和南部的云南、广西皆被归于较高死亡率簇，主要因由8月中旬，超强台风波及云南，且同年，河南省发生了自1954年以来第二次流域性大洪水导致，而归于中死亡率簇的地区集中分布于我国西部和中部，如新疆、四川、重庆、湖南等地区；2010年和2013年，全国各地区死亡状况明显较好，同年死亡率较高簇皆仅涵盖广西一个省份，而死亡率中等水平的簇主要分布于我国西北、西南和南部少部分地区，两年分类结果相较而言没有见大改变，唯有河南省从2010年的中等死亡率簇转为2013年低死亡率簇。

总体而言，全国死亡率分布状况存在区域性的现象，因此高死亡率簇中的地区基本分布在西北地区 and 西南部地区的边境省份，而中部、东部城市总被分为死亡率偏低的簇。

### 3. 病死率的聚类分析





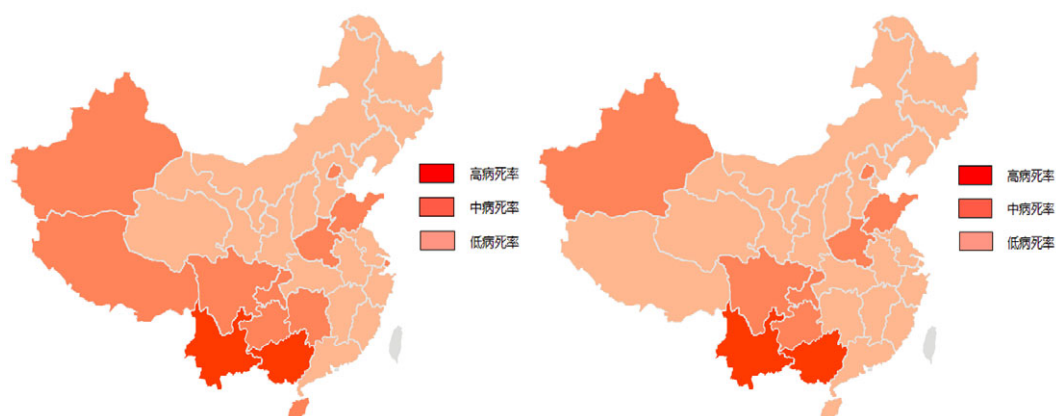


图 5 2004、2007、2010、2013 年全国病死率分布状况

由图 5 可知：我国病死率分布状况存在明显区域性变化的现象。2004 年，我国的高病死率簇地区主要集中于较南部的湖南、广西等地；中病死率簇地区明显聚集于我国东南和西北部。2007 年，西南部的广西、云南以及中部的河南等地区被归为病死率较高组中；由于区域性影响，使得归类于病死率中等水平的地区分布于周边的西南、中部和南部的多个省市。2010 年和 2013 年间，分类中病死率相对较高的地区始终集中在西南地区的云南、广西两地。由两年的病死率分布图可以比较发现，西部地区的西藏和南部地区的海南均从 2010 年中等病死率水平簇转化为了 2013 年的低病死率分类簇中。此外，广西省病死率始终高于其他地区，云南也基本属于同等水平下的高病死率簇。

## 五、灰色预测

### （一）方法介绍

#### 1. 灰色预测方法

灰色预测是指利用 GM 模型对系统行为特征的发展变化规律进行估计预，同时也可以对行为特征的异常情况发生的时刻进行估计计算，以及对在特定时区内发生事件的未来时间分布情况做出研究等等。这些工作实质上是将“随机过程”当作“灰色过程”，“随机变量”当作“灰变量”，并主要以灰色系统理论中的 GM(1,1) 模型来进行处理。灰色预测在工业、农业、商业等经济领域，以及环境、社会和军事等领域中都有广泛的应用。特别是依据目前已有的数据对未来的发展趋势做出预测分析<sup>[11]</sup>。

设已知参考数据列为  $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\}$  , 做一次累加 (AGO) 生成数列

$$\begin{aligned} X^{(1)} &= (X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)) \\ &= (X^{(1)}(1), X^{(1)}(1) + X^{(0)}(2), \dots, X^{(1)}(n-1) + X^{(0)}(n)) \end{aligned}$$

其中  $X^{(1)}(k) = \sum_{i=1}^k X^{(0)}(i)$  ( $k=1, 2, \dots, n$ )。求均值数列

$$Z^{(1)}(k) = 0.5X^{(1)}(k) + 0.5X^{(1)}(k-1), \quad k=2, 3, \dots, n$$

则  $Z^{(1)} = (Z^{(1)}(2), Z^{(1)}(3), \dots, Z^{(1)}(n))$ 。于是建立灰微分方程为

$$X^{(0)}(k) + aZ^{(1)}(k) = b, \quad k=2, 3, \dots, n$$

相应的白化微分方程为

$$\frac{dX^{(1)}}{dt} + aX^{(1)}(t) = b, \quad (2)$$

记  $u = (a, b)^T, Y = (X^{(0)}(2), X^{(0)}(3), \dots, X^{(0)}(n))^T, B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix}$  , 由最小二乘法, 求

得使  $J(\hat{u}) = (Y - B\hat{u})^T (Y - B\hat{u})$  达到最小值的  $\hat{u} = (a, b)^T = (B^T B)^{-1} B^T Y$ 。于是求解方程 (2) 得

$$X^{(1)}(k+1) = (X^{(1)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, \quad k=1, 2, \dots, n-1 \quad (3)$$

## 2. 灰色预测的步骤

### (1) 数据的检验与处理

首先, 为保证建模方法的可行性, 需要对已知数据列做必要的检验处理。设参考数据为  $X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n))$ , 计算数列的级比

$$\lambda(k) = \frac{X^{(0)}(k-1)}{X^{(0)}(k)}, \quad k=2, 3, \dots, n \quad (4)$$

如果所有的级比  $\lambda(k)$  都落在可容覆盖内  $(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$  , 则数列  $X^{(0)}$  可以作为模型 GM(1,1) 的数据进行灰色预测。否则, 需要对数列  $X^{(0)}$  做必要的变换处理, 使

其落入可容覆盖内。即取适当的常数  $c$  , 作平移变换

$$y^{(0)}(k) = X^{(0)}(k) + c, \quad k = 1, 2, \dots, n \quad (5)$$

则使数列  $y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n))$  的级比

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \in X, \quad k = 2, 3, \dots, n$$

## (2) 建立模型

按 (1) 中介绍的方法建立模型 GM(1,1) , 则可以得到预测值

$$\hat{X}^{(1)}(k+1) = (X^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, \quad k = 1, 2, \dots, n-1$$

而且  $\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k), \quad k = 1, 2, \dots, n-1$ 。

## (3) 检验预测值

残差检验：令残差为  $\varepsilon(k)$  , 计算

$$\varepsilon(k) = \frac{X^{(0)}(k) - \hat{X}^{(0)}(k)}{X^{(0)}(k)}, \quad k = 1, 2, \dots, n \quad (6)$$

如果  $\varepsilon(k) < 0.2$  , 则可认为达到一般要求；如果  $\varepsilon(k) < 0.1$  , 则认为达到较高的要求。

级比偏差值检验：首先由参考数据  $X^{(0)}(k-1)$  ,  $X^{(0)}(k)$  计算出级比  $\lambda(k)$  , 再用发展系数  $a$  求出相应的级比偏差

$$\rho(k) = 1 - \left( \frac{1-0.5a}{1+0.5a} \right) \lambda(k) \quad (7)$$

如果  $\rho(k) < 0.2$  , 则可认为达到一般要求；如果  $\rho(k) < 0.1$  , 则认为达到较高的要求。

## (二) 测算分析

### 1. 男性传染病死亡率灰色预测模型

#### 第一步：级比检验

为了保证模型的可行性，首先对数据进行必要的检验，本文选取 2004-2012 年男性传染病死亡率数据构建灰色预测模型，用 2013 年男性传染病死亡率数据

对模型进行精度检验。原始序列如下：

$$\begin{aligned} x^{(0)} &= (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5), x^{(0)}(6), x^{(0)}(7), x^{(0)}(8)) \\ &= (4.29, 4.395, 5.055, 6.4, 6.105, 6.29, 5.545, 5.085) \end{aligned}$$

(1) 求级比  $\lambda(k)$

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}$$

$$\begin{aligned} \lambda &= (\lambda(2), \lambda(3), \lambda(4), \lambda(5), \lambda(6), \lambda(7), \lambda(8)) \\ &= (0.7961, 0.8694, 0.7898, 1.0483, 0.9705, 1.1343, 1.0904, 0.7566, 0.6614) \end{aligned}$$

(2) 级比判断

由于所有的  $\lambda(k) \in [0.6614, 1.1343]$ ,  $k = 2, 3, \dots, 8$  不满足区间要求

$$(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}}) = (e^{-\frac{2}{9}}, e^{\frac{2}{10}}) = (0.800737, 1.221403)$$

因此需要对原始序列做必要的变换处理，即取适当的常数  $c$ ，做平移变换，使数据落在区间范围内。

(3) 数据变换

根据已有条件求常数  $c$ ，本文取常数  $c=10.71$ 。

$$\begin{aligned} y^{(0)} &= (y^{(0)}(1), y^{(0)}(2), y^{(0)}(3), y^{(0)}(4), y^{(0)}(5), y^{(0)}(6), y^{(0)}(7), y^{(0)}(8)) \\ &= (15, 15.105, 15.765, 17.11, 16.815, 17, 16.255, 15.795) \end{aligned}$$

此时级比值  $\lambda(k)$  如下：

$$\begin{aligned} \lambda &= (\lambda(2), \lambda(3), \lambda(4), \lambda(5), \lambda(6), \lambda(7), \lambda(8)) \\ &= (0.9930, 0.9581, 0.9213, 1.0175, 0.9891, 1.0458, 1.0291) \end{aligned}$$

由于所有的  $\lambda(k) \in [0.9213, 1.0458]$ ,  $k = 2, 3, \dots, 8$  满足区间要求，故可以用  $y^{(0)}$  作满意的 GM (1,1) 模型。

**第二步：GM (1,1) 建模**

(1) 对数据  $y^{(0)}$  作一次累加，即

$$y^{(1)} = (15, 30.105, 45.87, 62.98, 79.795, 96.795, 113.05, 128.845, 146.275, 167.145)$$

(2) 构造数据矩阵 B 及数据向量 Y

$$B = \begin{bmatrix} -22.5525 & 1 \\ -37.9875 & 1 \\ -54.425 & 1 \\ -71.3875 & 1 \\ -88.295 & 1 \\ -104.9225 & 1 \\ -120.9475 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} 15.105 \\ 15.765 \\ 17.11 \\ 16.815 \\ 17 \\ 16.255 \\ 15.795 \end{bmatrix}$$

(3) 计算  $\hat{\mu}$

$$\hat{\mu} = (a, b)^T = (B^T B)^{-1} B^T Y = \begin{pmatrix} -0.0062 \\ 15.8169 \end{pmatrix}$$

于是得到  $a = -0.0062$  ,  $b = 15.8169$ 。

(4) 建立模型

$$\frac{dy^{(1)}}{dt} - 0.0062y^{(1)} = 15.8169 \quad (8)$$

求解得

$$y^{(1)}(k+1) = (y^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} = 2547.138e^{0.0062k} - 2532.14 \quad (9)$$

(5) 求生成数列值  $y^{(1)}(k+1)$  及模型还原值  $y^{(0)}(k+1)$

令  $k = 1, 2, 3, 4, 5, 6, 7$  , 由上面的时间响应函数可算得  $y^{(1)}(1)$  , 其中取

$$y^{(1)}(1) = y^{(0)}(1) = 15$$

求得

$$\begin{aligned} y^{(1)} &= (y^{(1)}(1), y^{(1)}(2), y^{(1)}(3), y^{(1)}(4), y^{(1)}(5), y^{(1)}(6), y^{(1)}(7), y^{(1)}(8),) \\ &= (15, 30.9604, 47.0208, 63.1819, 79.4442, 95.8085, 112.2752, 128.8452) \end{aligned}$$

由  $y^{(0)}(k) = y^{(1)}(k) - y^{(1)}(k-1)$  , 取  $k = 2, 3, 4, 5, 6, 7, 8$  , 得

$$\begin{aligned} y^{(0)} &= (y^{(0)}(1), y^{(0)}(2), y^{(0)}(3), y^{(0)}(4), y^{(0)}(5), y^{(0)}(6), y^{(0)}(7), y^{(0)}(8),) \\ &= (15, 15.9604, 16.0604, 16.1611, 16.2623, 16.3642, 16.4668, 16.5699) \end{aligned}$$

### 第三步：模型检验

模型的各种检验指标值的计算过程见下表：

表 1 男性传染病死亡率模型的各种检验指标值

序号	年份	原始变换值	模型拟合值	绝对误差	相对误差	级比偏差
1	2004	15	15	0	0	
2	2005	15.105	15.9604	0.85542	0.05663	0.00072
3	2006	15.765	16.0604	0.29543	0.01873	0.03586
4	2007	17.11	16.1611	0.94893	0.05546	0.07283
5	2008	16.815	16.2623	0.55266	0.03286	-0.02391
6	2009	17	16.3642	0.63576	0.03739	0.00468
7	2010	16.255	16.4668	0.21177	0.01302	-0.05238
8	2011	15.795	16.5699	0.77495	0.04906	-0.03557

## 2. 女性传染病死亡率灰色预测模型

与上例类似,本文中发现女性原始值级比也不满足全部在区间范围内,故对其进行序列变换,加常数  $c$  为 2.5。进行如男性灰色预测建模步骤,得模型如下:

$$y^{(1)}(k+1) = (y^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} = 1559.709e^{0.00345k} - 1554.71 \quad (10)$$

模型的各种检验指标值的计算过程见下表:

表 2 女性传染病死亡率模型的各种检验指标值

序号	年份	原始变换值	模型拟合值	绝对误差	相对误差	级比偏差
1	2004	5	5	0	0	
2	2005	4.805	5.395129	0.590129	0.122816	-0.04418
3	2006	5.225	5.413791	0.188791	0.036132	0.07720
4	2007	6.245	5.432518	0.812482	0.130101	0.16043
5	2008	5.795	5.45131	0.34369	0.059308	-0.08138
6	2009	5.56	5.470166	0.089834	0.016157	-0.04587
7	2010	5.48	5.489088	0.009088	0.001658	-0.01810
8	2011	5.05	5.508075	0.458075	0.090708	-0.08890

## 3. 城市传染病死亡率灰色预测模型

与上例类似,本文中发现女性原始值级比也不满足全部在区间范围内,故对其进行序列变换,加常数  $c$  为 5。进行如男性灰色预测建模步骤,得模型如下:

$$y^{(1)}(k+1) = (y^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} = -1123.92e^{-0.008566k} + 1132.684 \quad (11)$$

模型的各种检验指标值的计算过程见下表:

表 3 城市传染病死亡率模型的各种检验指标值

序号	年份	原始变换值	模型拟合值	绝对误差	相对误差	级比偏差
1	2004	8.76	8.76	0	0	
2	2005	8.61	9.586228	0.97622	0.1134	0.025507
3	2006	9.24	9.504465	0.26446	0.0286	-0.06402
4	2007	10.82	9.423399	1.39660	0.1291	-0.16101
5	2008	9.73	9.343024	0.38697	0.0398	0.108409
6	2009	9.42	9.263335	0.15666	0.0166	0.040118
7	2010	9.44	9.184326	0.25567	0.0271	0.006424
8	2011	8.15	9.10599	0.95599	0.1173	0.144016

#### 4. 农村传染病死亡率灰色预测模型

我们发现农村原始值级比满足要求全部在区间范围内，因此直接进行计算，进行如男性灰色预测模型的步骤，可得模型如下：

$$x^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} = 78.71523e^{0.04546k} - 75.6452 \quad (12)$$

模型的各种检验指标值的计算过程见下表：

表 4 农村传染病死亡率模型的各种检验指标值

序号	年份	原始变换值	模型拟合值	绝对误差	相对误差	级比偏差
1	2004	3.07	3.07	0	0	
2	2005	3.18	3.661299	0.481299	0.15135	-0.01032
3	2006	3.59	3.831598	0.241598	0.06729	0.072998
4	2007	4.35	4.009818	0.340182	0.07820	0.136319
5	2008	4.72	4.196327	0.523673	0.110947	0.035515
6	2009	4.98	4.391512	0.588488	0.118170	0.008116
7	2010	4.13	4.595775	0.465775	0.112778	-0.26191
8	2011	4.53	4.80954	0.27954	0.061708	0.045887

模型预测结果（模型变换后的结果）与 2012 年、2013 年的真实值，绝对误差，相对误差见下表 5：

表 5 各模型传染病死亡率预测结果

		预测值	真实值	绝对误差	相对误差
2012 年	男性	16.67378	17.43	0.756216	0.0433
	女性	5.527127	5.09	0.437127	0.0858
	城市	9.028323	9.17	0.141676	0.0155
	农村	5.033247	5.69	0.656753	0.1154
2013 年	男性	16.77826	20.87	4.091738	0.1961
	女性	5.546246	6.615	1.068754	0.1615
	城市	8.951318	11.93	2.978681	0.2497
	农村	5.267359	7.94	2.672641	0.3366

由表 5 可以发现，2013 年的绝对误差与 2012 年的相比明显变大，该年死亡率的急速增加的主要原因是该年我国发生了多起重大事件。例如：3 月，我国首次发现人感染 H7N9 禽流感病例，到 2013 年 5 月 1 日，上海、安徽、江苏、浙江、北京、河南、山东、江西、湖南、福建等 10 省(市)共报告确诊病例 127 例；4 月 20 日四川省雅安市芦山县（北纬 30.3 度，东经 103.0 度）发生 7.0 级地震，震源深度 13 千米，地震造成 196 人遇难，21 人失踪，13484 人受伤，200 余万人受灾；10 月 7 日受台风“菲特”影响，浙江余姚遭遇新中国成立以来最严重水灾，其受灾人口为 832870 人。

随后，又对 2014-2015 年传染病死亡率进行预测，得出预测结果如表 6。单就城乡而言，农村居民传染病死亡率明显高于城市居民传染病死亡率，男性传染病死亡率显著高于女性传染病死亡率，且在未来两年里传染病死亡率不断增长。

表 6 2014、2015 年各模型死亡率预测结果（单位：十万分之一）

	城市	农村	男性	女性
2014 年	3.87497	5.512361	6.285162	3.065431
2015 年	3.799273	5.768759	6.98683	3.084682

本节对 2004 年至 2011 年城市、农村、男性、女性传染病死亡率数据进行灰色预测建模和拟合，发现相对误差较小，且级比偏差效果较好，又以 2012 年和 2013 年的拟合与真实值进行了检验，除 2013 年发生重大事件为特殊数据，2012 年数据拟合结果较好，于是进行了 2014 年与 2015 年两年的传染病死亡率预测，同样发现传染病死亡率农村高于城市、男性高于女性。

## 六、 面板模型构建

本文采用面板数据对我国传染病发病率、病死率及死亡率进行实证分析。选取 2004-2013 共 10 年间全国 31 个地区传染病发病率(F)、死亡率(S) 以及病死率(B)数据构建面板数据模型，本节使用 Eviews 6.0 软件进行分析。

### （一） 单位根检验

在建立面板计量经济学模型前，首先要检验变量的平稳性，否则会产生伪回归现象，即谬误回归现象。因此，在进行计量分析前，首先要对分析的变量进行单位根检验。在单位根检验中，一般是先检验既含截距又含趋势项的模型，再检验只含有截距项的模型，最后检验既不含截距又不含趋势项的模型。在这三个模型中只有检验的结果都不能拒绝原假设时，才认为该序列是非平稳的，否则序列是平稳的。为保证检验结果的稳健性，本文采用 LLC、ADF-Fisher 和 PP-Fisher 单位根检验三种检验方法来判断变量发病率（F）、死亡率（S）和病死率（B）是否具有单位根，检验结果见表 7。

表 7 变量单位根检验

变量	检验类型	LLC	p 值	Fisher-ADF	p 值	Fisher-PP	p 值
F	I&T	-8.24929	0.0000	92.1299	0.0078	148.689	0.0000
	I	-8.02311	0.0000	121.594	0.0000	128.855	0.0000
	N	-5.19533	0.0000	115.531	0.0000	139.460	0.0000
S	I&T	-18.4508	0.0000	126.940	0.0000	172.209	0.0000
	I	-5.38255	0.0000	115.591	0.0000	146.550	0.0000
	N	0.20361	0.5807	39.3947	0.9889	26.7227	1.0000
B	I&T	-7.07998	0.0000	75.8017	0.1119	110.409	0.0002
	I	-2.36689	0.0090	74.1699	0.1384	112.183	0.0001
	N	0.99377	0.8398	29.8062	0.9998	26.2588	1.0000

注：T&I 表示含有趋势项和截距项，I 表示含有截距项，N 表示既不含截距又不含趋势项。



由表 7 可以看出，发病率（F）、死亡率（S）和病死率（B）三变量均为平稳序列，可以直接建立回归模型，而不会产生伪回归现象。

## （二）Granger 因果检验

Granger 因果检验的逻辑是：过去或现在可能预测（解释）将来，但是将来却不一定能预测过去或现在。如果事件 A 在事件 B 之后发生，显然 A 不能解释 B，但若 A 在 B 之前发生，这并不一定意味着 A 能解释 B<sup>[12-13]</sup>。正如本文中的发病率、死亡率与病死率之间，发病才能导致病死或死亡，但并不表明发病率是病死率或死亡率的原因。故此，本文用 Granger 因果检验判定发病率（F）、死亡率（S）以及病死率（B）之间是否存在因果关系。本节采用 Eviews6.0 进行分析。

Granger 因果检验假定有关变量  $y$  或  $x$  变量的预测信息全部包含在这两个变量的时间序列之中，检验要求估计以下回归：

$$y_t = \sum_{i=1}^m \alpha_i x_{t-i} + \sum_{j=1}^m \beta_j y_{t-j} + \mu_{1t}$$

$$x_t = \sum_{i=1}^m \lambda_i x_{t-i} + \sum_{j=1}^m \delta_j y_{t-j} + \mu_{2t}$$

其中干扰项  $\mu_{1t}$  和  $\mu_{2t}$  假定为不相关的。

上式第一个方程检验变量  $x$  是否是引起  $y$  变化的原因，虚拟假设为  $x$  不是  $y$  变化的原因，即  $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ ，检验统计量：

$$F = \frac{(RSS_R - RSS_{UR}) / m}{RSS_{UR} / (n - k)}$$

它遵循自由度为  $m$  和  $(n - k)$  的  $F$  分布。 $m$  是变量  $x$  的滞后阶数， $k$  是指无约束回归中为之参数个数。

第二个方程同上，但是检验  $y$  是否是  $x$  变化的原因，虚拟假设为  $y$  不是  $x$  变化的原因，即  $\delta_1 = \delta_2 = \cdots = \delta_m = 0$ 。

根据上述检验原理，对传染病的发病率、病死率以及死亡率两两之间进行 Granger 因果检验。又因为 Granger 因果检验对于滞后期长度的选择比较敏感，不同的滞后期可能会得到完全不同的检验结果。鉴于发病率（F）、死亡率（S）和病死率（B）三变量均为平稳序列，可以直接建立两两之间的 VAR 模型并根据 AIC 和 SC 准则确定其滞后阶数后再进行因果检验。因果检验的结果如下：

表8 Granger因果检验结果 (B与S)

滞后阶数	原假设	F 统计量值	P 值
1	S不是导致B变化的原因	5.05683	0.0252
	B不是导致S变化的原因	13.5795	0.0003
2	S不是导致B变化的原因	10.4070	4.E-05
	B不是导致S变化的原因	14.5735	9.E-07
3	S不是导致B变化的原因	9.95385	3.E-06
	B不是导致S变化的原因	11.5238	4.E-07
4	S不是导致B变化的原因	8.93106	8.E-07
	B不是导致S变化的原因	9.88094	2.E-07

表9 Granger因果检验结果 (F与B)

滞后阶数	原假设	F 统计量值	P 值
1	F不是导致B变化的原因	0.99095	0.3203
	B不是导致F变化的原因	0.78592	0.3760
2	F不是导致B变化的原因	0.83677	0.4341
	B不是导致F变化的原因	1.01315	0.3643
3	F不是导致B变化的原因	0.66042	0.5770
	B不是导致F变化的原因	0.96913	0.4076
4	F不是导致B变化的原因	0.63319	0.6392
	B不是导致F变化的原因	0.89090	0.4697

表10 Granger因果检验结果 (F与S)

滞后阶数	原假设	F 统计量值	P 值
1	S不是导致F变化的原因	0.61787	0.4324
	F不是导致S变化的原因	6.81629	0.0095
2	S不是导致F变化的原因	1.07911	0.3412
	F不是导致S变化的原因	3.87073	0.0219
3	S不是导致F变化的原因	1.07208	0.3612
	F不是导致S变化的原因	2.55263	0.0557
4	S不是导致F变化的原因	0.97589	0.4209
	F不是导致S变化的原因	2.05070	0.0873

根据上述Granger因果检验结果可知,B与S均不是导致F变化的原因,表明病死率和死亡率不是发病率的原因,符合实际情况;F不是导致B变化的原因,也不是导致S变化的原因,说明发病率与病死率、死亡率之间并不存在因果关系;B与S互为因果,说明病死率和死亡率之间是相互影响、相互制约的。因此,本文分别以病死率为解释变量,研究其对死亡率的影响,以死亡率为解释变量,研究其对病死率的影响。

### (三) 面板模型分类

#### 1. 混合模型

如果不同个体在时间上没有显著差异,并且同一时间上个体间也没有显著差异,那么就可以利用所有数据建立简单回归模型,即混合模型,模型形式如下:

$$y_{it} = \beta_1 + \sum_{k=2}^K \beta_k x_{kit} + \mu_{it}$$

当然也可以写成矩阵形式:  $Y = X\beta + U$

$$\text{其中, } Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}_{NT \times 1}, X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}_{NT \times K}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}_{K \times 1}, U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix}_{N \times 1}$$

该模型假定在各不同时期个体的截距和斜率都是相同的,并且用普通最小二乘估计得到的参数估计是一致有效估计。

#### 2. 固定效应模型

在面板数据回归模型中,如果模型的斜率系数不变截距随个体而变化,或模型的斜率系数不变截距随时间变化,或斜率系数不变截距随个体和时间变化,则此种模型为固定效应模型。固定效应模型一般分为个体固定效应模型、时点固定效应模型、个体时点固定效应模型这三类。

(1) 个体固定效应模型。个体固定效应模型是指模型的斜率系数不变截距随个体而变化,模型形式如下:

$$y_{it} = \delta + \lambda_i + \sum_{k=2}^K \beta_k x_{kit} + \mu_{it}$$

(2) 时点固定效应模型。时点固定效应模型是指模型的斜率系数不变截距随时间变化。时点固定效应模型形式如下:

$$y_{it} = \delta + \gamma_t + \sum_{k=2}^K \beta_k x_{kit} + \mu_{it}$$

(3) 个体时点固定效应模型。如果模型的斜率系数不变截距随个体和时间变化,则建立个体时点固定效应模型:

$$y_{it} = \delta + \lambda_i + \gamma_t + \sum_{k=2}^K \beta_k x_{kit} + \mu_{it}$$

其中,  $i = 1, 2, \dots, N-1$  表示  $N$  个个体,  $t = 1, 2, \dots, T-1$ , 表示已知的  $T$  个时点。

### 3. 随机效应模型

如果未观测到的个体差异性可以被假定与包括在模型中的变量无关, 则建立随机效应模型。模型形式如下:

$$y_{it} = \beta_1 + \sum_{k=2}^K \beta_k x_{kit} + \mu_i + \nu_t + \omega_{it},$$

其中  $\mu_i$  表示与个体相关的随机误差分量,  $\nu_t$  表示与时间相关的随机误差分量,  $\omega_{it}$  表示与个体和时间均相关的混合随机误差分量。上述模型表示的是个体时点随机效应模型, 如果模型中只存在个体随机误差分量  $\mu_i$  而不存在时间随机误差分量  $\nu_t$ , 则上述模型为个体随机效应模型; 如果模型中只存在时间随机误差分量  $\nu_t$  而不存在个体随机误差分量  $\mu_i$ , 则上述模型为时点随机效应模型。

### 4. 固定效应和随机效应模型设定检验

固定效应模型和随机效应模型各有优劣, 我们通常用Hausman检验对固定效应和随机效应模型进行选择。Hausman检验假设:

原假设  $H_0$ : 模型是随机效应模型

备择假设  $H_1$ : 模型是固定效应模型

本文用views6.0对模型设定进行检验, 检验结果如表11和表12:

表 11 固定效应模型和随机效应模型 hausman 检验

模型	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
个体效应	0.880240	1	0.3481
时点效应	0.775344	1	0.3786
时点个体效应	0.021128	1	0.8844

注: S作为被解释变量

表 12 固定效应模型和随机效应模型 hausman 检验

模型	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
个体效应	1.051222	1	0.3052
时点效应	11.141805	1	0.0008
时点个体效应	1.925666	1	0.1652

注: B作为被解释变量

由上表可知, 不论是个体效应还是时点效应, 抑或是时点个体效应, 在5%的显著性水平下, Hausman检验结果都是接受原假设, 即认为随机效应模型是合适的。由表 可知, 除去时点效应外, 不论是个体效应还是时点个体效应,

Hausman检验结果都是接受原假设，即认为随机效应模型是合适的。

#### (四) 模型选择与测度

##### 1. 模型选择

由表11、表12可知本文应建立随机效应模型，但是建立个体随机效应模型、时点随机效应模型，还是建立时点个体随机效应模型，还需对模型进行选择。模型选择各原假设及检验统计量如下：

##### (1) 混合模型

$$H_0^1: \lambda_1 = \lambda_2 = \cdots = \lambda_{N-1} = 0, \gamma_1 = \gamma_2 = \cdots = \gamma_{N-1} = 0$$

##### (2) 个体随机效应模型

$$H_0^2: \lambda_1 = \lambda_2 = \cdots = \lambda_{N-1} = 0$$

$$\text{检验统计量 } F_2 = \frac{(S_1 - S_2)/(N-1)}{S_2/(NT - N - K + 1)} \sim F(N-1, N(T-1) - K + 1)$$

其中  $S_1$  是混合模型残差平方和， $S_2$  是个体随机效应模型残差平方和，若拒绝原假设则选择个体随机效应模型。

##### (3) 时点随机效应模型

$$\text{原假设 } H_0^3: \gamma_1 = \gamma_2 = \cdots = \gamma_{T-1} = 0$$

$$\text{检验统计量 } F_3 = \frac{(S_1 - S_3)/(T-1)}{S_3/(NT - T - K + 1)} \sim F(T-1, T(N-1) - K + 1)$$

其中  $S_1$  是混合模型残差平方和， $S_3$  是时点随机效应模型残差平方和，若拒绝原假设，则选择时点随机效应模型。

##### (4) 时点个体随机效应模型

此模型检验包含两种情形，情形1：存在个体效应下，是否包含时点效应；情形2：存在时点效应下，是否包含个体效应。

存在个体效应下，是否包含时点效应：

$$\text{原假设 } H_0^4: \gamma_1 = \gamma_2 = \cdots = \gamma_{T-1} = 0, \text{ 当 } \lambda_i \neq 0, i = 1, 2, \cdots, N-1$$

$$\text{检验统计量 } F_4 = \frac{(S_2 - S_4)/(T-1)}{S_4/(NT - T - N - K + 2)} \sim F(T-1, NT - T - N - K + 2)$$

其中  $S_4$  是时点个体随机效应模型残差平方和，若拒绝原假设，则选择个体随机效应模型。存在时点效应下，是否包含个体效应：

原假设  $H_0^5: \lambda_1 = \lambda_2 = \dots = \lambda_{N-1} = 0$ , 当  $\gamma_t \neq 0, t = 1, 2, \dots, T-1$

$$\text{检验统计量 } F_5 = \frac{(S_3 - S_5) / (N - 1)}{S_5 / (NT - T - N - K + 2)} \sim F(N - 1, NT - T - N - K + 2)$$

其中  $S_5$  是时点个体随机效应模型残差平方和，若拒绝原假设，则选择时点随机效应模型。

根据上述原理，得到模型选择结果如表13。由表13可知，应选择时点个体随机效应模型。

表13 随机效应模型选择

模 型	残差平方和	F 统计量	F 统计量分布	F 分布 95%分位数
混合模型	48.29895			
个体随机效应模型	15.32788	19.9330837	F(30,279)	1.46
时点随机效应模型	47.94357	0.24625855	F(9,300)	1.911
时点个体随机效应模型 (存在个体效应情况)	15.11364	0.4252582	F(9, 270)	1.915

注：S作为被解释变量，B为解释变量

表14 随机效应模型选择

模 型	残差平方和	F 统计量	F 统计量分布	F 分布 95%分位数
混合模型	6.469259			
个体随机效应模型	2.235948	17.607651	F(17,162)	1.71
时点随机效应模型	6.092768	2.0597698	F(9,170)	1.999
时点个体随机效应模型 (存在个体效应情况)	2.118651	1.6609200	F(9, 153)	1.71

注：B 作为被解释变量

由上表可知，S 作为被解释变量时，应构建时点个体随机效应模型；B 作为被解释变量时，应构建时点个体随机效应模型。

## 2. 模型测度分析

根据上述检验结果，本文将建立病死率对死亡率的时点个体固定效应模型以及死亡率对病死率的时点个体固定效应模型。

(1) 病死率对死亡率的时点个体固定效应模型：

$$S = 0.096735 + \hat{\lambda}_i + \hat{\gamma}_t + 2.400811B + \hat{\mu}_{it} \quad (13)$$

表 15 时点个体随机效应方程回归结果

变量	系数	标准误差	T 统计量	P 值
C	0.096735	0.066088	1.463746	0.1443
B	2.400811	0.065707	36.53798	0.0000

表 16 个体随机效应部分 (Cross)

地 区	个体随机效应	地 区	个体随机效应
北 京	0.014529	湖 北	0.026836
天 津	-0.181882	湖 南	-0.280211
河 北	-0.151174	广 东	0.126590
山 西	-0.038357	广 西	0.877518
内蒙古	-0.001181	海 南	0.103571
辽 宁	-0.143005	重 庆	0.052561
吉 林	-0.109648	四 川	-0.123882
黑龙江	-0.063594	贵 州	0.272298
上 海	-0.164939	云 南	-0.583151
江 苏	-0.351087	西 藏	-0.144443
浙 江	0.011796	陕 西	-0.079945
安 徽	-0.182681	甘 肃	0.094200
福 建	-0.013476	青 海	0.190345
江 西	-0.114914	宁 夏	-0.023960
山 东	-0.415601	新 疆	1.120663
河 南	0.276226		

表 17 时点随机效应部分 (Period)

时间	时点随机效应	时间	时点随机效应
2004--C	-0.017634	2009—C	0.020666
2005--C	0.007102	2010—C	-0.016405
2006--C	0.009319	2011—C	-0.009254
2007--C	0.017789	2012—C	-0.001005
2008--C	0.008880	2013—C	-0.019458

(2) 传染病死亡率对传染病病死率的时点个体随机效应模型：

$$B = 0.037280 + \hat{\lambda}_i + \hat{\gamma}_i + 0.335402S + \hat{\mu}_{it} \quad (14)$$

表 18 时点个体随机效应方程回归结果

变量	系数	标准误差	T—统计量	P 值
C	0.037280	0.023615	1.578639	0.1154
S	0.335402	0.009315	36.00710	0.0000

表 19 个体随机效应部分 (Cross)

地区	个体随机效应	地区	个体随机效应
北 京	-0.004100	湖 北	-0.019688
天 津	0.035957	湖 南	0.130977
河 北	0.010903	广 东	-0.044216
山 西	-0.027057	广 西	-0.135437

内蒙古	-0.043520	海 南	-0.028209
辽 宁	0.024116	重 庆	0.003240
吉 林	0.006146	四 川	0.070922
黑龙江	0.016186	贵 州	-0.034583
上 海	0.054956	云 南	0.351795
江 苏	0.110678	西 藏	0.047941
浙 江	-0.036541	陕 西	-0.007457
安 徽	0.044006	甘 肃	-0.074188
福 建	-0.030715	青 海	-0.104555
江 西	0.030320	宁 夏	-0.039266
山 东	0.124020	新 疆	-0.373576
河 南	-0.059053		

表 20 时点随机效应部分 (Period)

时间	时点随机效应	时间	时点随机效应
2004--C	-0.005262	2009--C	-0.007929
2005--C	-0.013566	2010--C	0.014806
2006--C	-0.012922	2011--C	0.012806
2007--C	-0.013113	2012--C	0.011258
2008--C	-0.008520	2013—C	0.022441

最后,本文对上述两个模型残差序列进行单位根检验,检验结果显示,残差序列均为平稳序列,故认为残差符合随机序列,拟合模型效果较好。

由上述两模型结果可得出以下结论:

1. 我国传染病病死率和死亡率相互影响,且呈现正向关系。由表 15 可知,传染病病死率系数为  $2.400811 > 0$ ,说明近年来传染病病死率对传染病死亡率影响是正向的。即随着传染病病死率提高,传染病死亡率也随之提高。由表 18 可知,传染病死亡率系数为  $0.335402 > 0$ ,同样也表明传染病死亡率提高的同时,传染病病死率也在提高,并且,病死率变化对于死亡率影响大于死亡率变化对于病死率影响。

2. 我国传染病病死率和死亡率在时间和个体上存在差异。由表 16、表 19 可知,同一时间不同地区传染病病死率对传染病死亡率影响截距不同,且差异较大;由表 17、表 20 可知,同一地区不同时间传染病病死率对传染病死亡率影响不同,但差异不大。

## 七、结论与建议



## （一）结论

本文根据 k-means 聚类得到的全国各地区发病率、死亡率及病死率三指标高、中、低水平的聚类状况，并用不同的颜色画出其对应的地区图；其次，构建了城乡及男女传染病死亡率的灰色预测模型并对 2014-2015 年进行预测；最后构建了各地区死亡率和病死率的时点个体随机效应模型，从而分析我国的传染病蔓延情况，得出以下结论：

1. 本文以图像方式清晰明了地展示出具体分类情况，从聚类结果中易知，高发病率簇基本聚集于西北部，但高死亡率与高病死率均集中在南部，其原因很可能是西北地区医疗条件差，感染传染病机率较大，但由于北方地区较少发生巨大灾害，故其死亡率和病死率均相对偏低，而南方沿海地区一般较为开放，人口流动率高，且沿海地区易发生各种灾害，因此始终死亡率与病死率相对较高的分类中。

2. 传染病死亡率农村稍高于城市，且农村居民死亡率持续处于上升状态。据灰色预测模型结果，农村居民传染病死亡率将又创新高，应当大力加强农村居民传染病防治与监控，提高农村医疗水平。由于男性在生活习惯、个人卫生等方面缺乏防范意识的原因，男性传染病死亡率明显高于女性传染病死亡率。

3. 根据格兰杰因果检验，传染病发病率与病死率、死亡率之间不存在确定的因果关系，而病死率与死亡率之间存在明显的因果关系，且互为因果，表明病死率和死亡率相互影响、相互制约。

4. 根据面板模型可知，病死率与死亡率之间存在着双向的正向关系，说明传染病病死率与死亡率的增高，一定程度上也促进了死亡率与病死率的增高。病死率对死亡率的影响程度大于死亡率对病死率的影响程度。另外，传染病病死率和死亡率在时间和个体上均存在着差异，可能是由于国家医疗政策的变化或者地区发展水平之间差异所引起。

## （二）建议

### 1. 增强疾病预防意识，提高医疗水平

因新疆既常年属于高发病率簇又常被归为中等死亡率簇，故应加强新疆地区传染病预防知识的普及，进一步改善该地区环境卫生，从而减少新疆发病率和死亡率。又因广西、云南等地常被归类为高死亡率簇，故理当进一步加强广西地区的综合医疗水平，提高广西地区传染病的治疗实力，以降低该省死亡率及病死率。综上所述，应从经济和政治两方面大力支持新疆、广西、云南内各偏远地区医疗点的开设，并提出具体的优惠政策，同时可以增加医学院学生在上述地区的实习

人数,促成医疗支援的风潮;最后,由发病率的变化趋势,建议南方沿海城市当加大对城市人口的控制,增强城市居民的健身意识,培养居民常体检的习惯。

## 2. 加大农村防治力度,建议男性保持良好的个人习惯

由灰色预测对城乡、男女传染病死亡率分析比较所得,农村居民传染病死亡率要高于城镇居民传染病死亡率,且在未来的两年里仍维持着上涨趋势,由此可见,应加大农村传染病防治与监控力度,提高医疗水平,加强治疗措施。而男性居民传染病死亡率要明显高于女性居民传染病死亡率,因此建议男士在日常生活中要多注意自己的清洁卫生情况,常常保持干净整洁的状态,少抽烟不酗酒,多运动,养成良好的生活习惯。

## 3. 宣传传染病防治知识,制定监控政策

通过本文对传染病病死率与死亡率关系研究显示,传染病病死率与传染病死亡率是相互影响、相互制约的,故此,为了降低传染病死亡率,我国在提升传染病方面医疗水平的同时,也应该加强传染病防治与监控措施,通过降低传染病病死率水平,有效的控制传染病死亡率水平。加强防治措施可以通过举办宣传传染病防治知识活动或制定具体的传染病监控政策进行控制等。

## 参考文献

- [1] 樊雯婧, 陆群, 邹立巍, 等. ARIMA 模型在合肥市疟疾发病预测中的应用[J]. 安徽大学学报. 2013, 48(3): 252-256.
- [2] 杨小兵, 汪鹏, 江高峰. ARIMA 乘积季节模型在流行性腮腺炎发病率预测中的应用[J]. 公共卫生与预防医学, 2013, 24(6): 39-42.
- [3] Zhang J, Ma Z. Global dynamics of an SEIR epidemic model with saturating contact rate [J]. Mathematical Biosciences, 2003, 185(1): 15-32.
- [4] 杨志军. 我国慢性传染病预防与治疗监测的经济学分析——以慢性乙型肝炎为例[D]. 上海: 上海交通大学. 2008.
- [5] 李甜甜. 几类带有时滞的传染病模型稳定性分析[D]. 山西: 中北大学. 2014.
- [6] 李冰. 具有脉冲接种的传染病模型[D]. 哈尔滨: 哈尔滨理工大学. 2014.
- [7] Peng Z, Bao C, Zhao Y, et al. Weighted Markov chains for forecasting and analysis in Incidence of infectious diseases in jiangsu Province, China [J]. Journal of biomedical research, 2010, 24(3): 207-214.
- [8] 杨柯君编. 死亡率、发病率、病死率和患病率[J]. 上海医药, 2012, 33(20).
- [9] 张文彤, 董伟编. SPSS 统计分析高级教程[M]. 北京: 高等教育出版社, 2013. 03.
- [10] 何晓群编. 多元统计分析[M]. 北京: 中国人民大学出版社, 2013.06.

- [11] 徐国祥 编.统计预测与决策[M]. 上海: 上海财经大学出版社, 2010: 206-220.
- [12] 陈雄兵, 张宗成. 再议 Granger 因果检验[J]. 数量经济技术经济研究, 2008, (1): 154-160.
- [13] 李永立, 吴冲. 基于多变量的 Granger 因果检验方法[J]. 数理统计与管理, 2014, 33(1): 50-58.
- [14] 易丹辉编.数据分析与 Eviews 应用[M]. 北京: 中国人民大学出版社, 2008: 294-314.
- [15] 巴蒂 H. 巴尔塔基著. 面板数据计量经济分析[M]. 北京: 机械工业出版社, 2010: 244-248.
- [16] 朱祥淑, 舒昌惠. 传染病预防与管理工作中存在的问题及解决对策[J].深圳中西医结合杂志, 2015, 25(1): 177-178.

## 附录：

### 附录 1：2004 年各地区聚类结果

地区	发病率	死亡率	病死率	发病率聚类结果	死亡率聚类结果	病死率聚类结果
北京	408.03	0.61	0.15	1	1	3
天津	261.26	0.28	0.11	2	3	3
河北	192.1	0.13	0.07	3	3	3
山西	192.16	0.11	0.05	3	3	3
内蒙古	256.02	0.24	0.1	2	3	3
辽宁	206.31	0.43	0.21	3	3	1
吉林	220.48	0.3	0.14	3	3	3
黑龙江	275.56	0.87	0.32	2	1	1
上海	357.76	0.99	0.28	1	1	1
江苏	202.46	0.51	0.25	3	3	1
浙江	407.33	0.57	0.14	1	1	3
安徽	180.16	0.42	0.24	3	3	1
福建	268.16	0.26	0.1	2	3	3
江西	258.56	0.6	0.23	2	1	1
山东	128.87	0.21	0.16	3	3	3
河南	239.33	0.45	0.19	2	3	3
湖北	254.21	0.86	0.34	2	1	1
湖南	130.24	0.98	0.76	3	1	2
广东	205.06	0.74	0.36	3	1	1
广西	300.3	1.87	0.62	2	2	2
海南	404.54	0.32	0.08	1	3	3
重庆	265.91	0.37	0.14	2	3	3
四川	318.87	0.34	0.11	2	3	3
贵州	322.91	1.12	0.35	2	1	1
云南	216.29	0.67	0.31	3	1	1
西藏	239.05	0.19	0.08	2	3	3
陕西	388.29	0.44	0.11	1	3	3
甘肃	444.16	0.32	0.07	1	3	3
青海	384.59	0.89	0.23	1	1	1
宁夏	516.4	0.17	0.03	1	3	3
新疆	435.31	0.47	0.11	1	3	3

## 附录 2：2007 年各地区聚类结果

地区	发病率	死亡率	病死率	发病率聚类结果	死亡率聚类结果	病死率聚类结果
北京	421.02	0.76	0.18	1	3	3
天津	294.77	0.45	0.15	3	3	3
河北	206.5	0.37	0.18	3	3	3
山西	283.91	0.35	0.12	3	3	3
内蒙古	341.66	0.53	0.16	1	3	3
辽宁	222.41	0.54	0.24	3	3	3
吉林	208.09	0.35	0.17	3	3	3
黑龙江	254.68	0.85	0.33	3	3	2
上海	220.67	0.99	0.45	3	1	2
江苏	172.38	0.59	0.34	3	3	2
浙江	348.02	0.56	0.16	1	3	3
安徽	241.52	0.59	0.24	3	3	3
福建	299.24	0.57	0.19	3	3	3
江西	232.64	0.82	0.35	3	3	2
山东	124.17	0.42	0.34	3	3	2
河南	354.29	2.36	0.67	1	2	1
湖北	310.33	0.9	0.29	1	3	2
湖南	212.47	1.18	0.55	3	1	2
广东	325.26	1.07	0.33	1	1	2
广西	329.2	2.62	0.79	1	2	1
海南	322.02	1.41	0.44	1	1	2
重庆	333.77	1.26	0.38	1	1	2
四川	291.87	1.21	0.41	3	1	2
贵州	342.89	1.95	0.57	1	2	2
云南	215.58	1.86	0.87	3	2	1
西藏	197.7	0.68	0.34	3	3	2
陕西	254.2	0.34	0.13	3	3	3
甘肃	487.63	0.62	0.13	1	3	3
青海	463.63	0.4	0.09	1	3	3
宁夏	405.51	0.31	0.08	1	3	3
新疆	657.53	1.65	0.25	2	1	3

### 附录 3：2010 年各地区聚类结果

地区	发病率	死亡率	病死率	发病率聚类 结果	死亡率聚类 结果	病死率聚类 结果
北京	268.99	1.35	0.5	1	3	1
天津	173.54	0.52	0.3	2	3	3
河北	199.99	0.32	0.2	2	3	3
山西	264.94	0.5	0.2	1	3	3
内蒙古	324.57	0.36	0.1	1	3	3
辽宁	207.08	0.44	0.2	2	3	3
吉林	245.63	0.51	0.2	1	3	3
黑龙江	239.48	0.73	0.3	1	3	3
上海	185.62	0.81	0.4	2	3	1
江苏	138.19	0.44	0.3	2	3	3
浙江	296.48	0.58	0.2	1	3	3
安徽	176.16	0.54	0.3	2	3	3
福建	292.65	0.52	0.2	1	3	3
江西	210.36	0.64	0.3	2	3	3
山东	102.48	0.33	0.3	2	3	3
河南	289.49	1.88	0.7	1	1	1
湖北	278.59	0.82	0.3	1	3	3
湖南	212.91	1.16	0.5	2	3	1
广东	327.53	1.21	0.4	1	3	3
广西	331.41	4.24	1.3	1	2	2
海南	261.51	1.69	0.7	1	1	1
重庆	240.82	1.44	0.6	1	3	1
四川	225.42	1.44	0.6	2	3	1
贵州	277.47	1.9	0.7	1	1	1
云南	178.81	3.06	1.7	2	1	2
西藏	187.85	1.03	0.6	2	3	1
陕西	212.07	0.45	0.2	2	3	3
甘肃	402.98	0.38	0.1	3	3	3
青海	456.13	0.5	0.1	3	3	3
宁夏	239.7	0.35	0.2	1	3	3
新疆	539.3	2.58	0.5	3	1	1

#### 附录 4：2013 年各地区聚类结果

地区	发病率	死亡率	病死率	发病率聚类结果	死亡率聚类结果	病死率聚类结果
北京	155.87	1.13	0.72	2	3	1
天津	143.07	0.44	0.31	2	3	3
河北	181.6	0.27	0.15	2	3	3
山西	296.86	0.49	0.17	1	3	3
内蒙古	295.05	0.34	0.12	1	3	3
辽宁	209.54	0.65	0.31	2	3	3
吉林	194.96	0.55	0.28	2	3	3
黑龙江	216.1	0.84	0.39	2	3	3
上海	173.78	0.63	0.36	2	3	3
江苏	116.74	0.47	0.4	2	3	3
浙江	192.46	0.47	0.25	2	3	3
安徽	198.32	0.55	0.28	2	3	3
福建	280.57	0.42	0.15	1	3	3
江西	206.37	0.61	0.29	2	3	3
山东	124.05	0.29	0.24	2	3	3
河南	242.13	1.51	0.62	1	3	1
湖北	254.81	0.73	0.29	1	3	3
湖南	254.89	1.26	0.49	1	3	3
广东	317.12	1.01	0.32	1	3	3
广西	289.47	6.67	2.3	1	1	2
海南	313.16	1.14	0.36	1	3	3
重庆	254.55	1.89	0.74	1	3	1
四川	198.29	2	1.01	2	3	1
贵州	261.95	1.96	0.75	1	3	1
云南	206.43	4.07	1.97	2	2	2
西藏	237.57	0.75	0.31	1	3	3
陕西	207.73	0.47	0.23	2	3	3
甘肃	208.99	0.45	0.22	2	3	3
青海	480.59	0.63	0.13	3	3	3
宁夏	253.33	0.46	0.18	1	3	3
新疆	618.99	3.88	0.63	3	2	1