

所属类别	2024 年第三届全国大学生数学分析实践赛	参赛编号
本科生组		2408293

基于多元统计与时间序列分析的中国医疗卫生资源动态研究及未来预测

摘 要

本研究聚焦于中国医疗卫生机构数量的时间趋势、资源配置差异、未来预测及宏观预测，旨在为政策制定和资源优化提供科学依据。通过综合应用**多元线性回归**、**因子分析**、**时间序列分析**及**灰色预测模型**等方法，深入剖析了医疗卫生资源的动态变化和影响因素。

针对问题一，即医疗卫生机构数量的增长趋势，本研究采用**年均增长率**、趋势分析与可视化模型，揭示了 1990 年至 2023 年间中国各地区的增长**异质性**。其中，**河南省**以**7.1616%**的增长率位居**榜首**，而**上海市**则以**-0.5985%**的增长率居于**末位**。这一发现为资源的精准配置指明了方向。

针对问题二，研究通过多元线性回归和因子分析，深入探讨了医疗卫生资源配置差异及其背后的影响因素。构建的**经济医疗丰裕因子**与**就业市场压力因子**，清晰地展示了经济与医疗资源丰裕度以及就业市场状况对医疗卫生机构数量的显著影响。

$$Y_p = 3.372e+04 + 1.693e+04 * Z_1 - 4515.9143 * Z_2$$

结果表明，经济水平与医疗资源的丰裕度**正向促进**了医疗卫生机构数量的增长，而就业市场压力则呈现**负相关**。

针对问题三，采用 **ARIMA 时间序列模型**预测未来五年各地区医疗卫生机构数量。模型通过**平稳性检测**、**差分处理**、**自相关**和**偏自相关函数**分析建立，预测结果为资源规划提供了依据。

针对问题四，利用灰色预测模型，对全国医疗卫生机构数量的未来十年宏观趋势进行了预测。通过残差检验与级比偏差检验，模型展现出了高度的可靠性和适用性，**平均相对残差**仅为**0.050577**，**平均级比偏差**为**0.024317**，证实了模型的预测精度。

本研究**创新性地**结合**多种分析方法**，深度解析了医疗卫生资源变化的内在原因，为政策制定提供了数据驱动的策略建议。强调了**多元化投入机制**对促进医疗卫生事业**可持续发展**的关键作用，为政府决策提供了有力的数据支持。

关键词：ARIMA 时间序列预测；灰色预测模型；因子分析；可持续发展；多元化投入机制；医疗卫生资源。

一、问题重述

1.1 问题背景

随着中国经济和社会的快速发展，医疗卫生事业作为国家综合实力的重要组成部分，其发展水平直接关系到国民健康福祉和社会稳定。近年来，中国政府高度重视医疗卫生领域的改革和发展，致力于提升医疗卫生服务质量，满足人民群众日益增长的健康需求。然而，尽管取得了显著进展，我国医疗卫生资源仍然面临一系列挑战。

首先，医疗资源分布不均是我国医疗卫生领域长期存在的问题之一。优质医疗资源大多集中于大城市和经济发达地区，而农村和边远地区则面临医疗资源短缺的问题。这种不平衡导致了“看病难、看病贵”的现象，特别是在一些偏远地区，居民获得高质量医疗服务的机会极为有限。

其次，随着人口老龄化进程的加快，慢性病患者群体不断扩大，这对医疗卫生系统提出了更高的要求。一方面，需要增加医疗资源投入以应对老年人口增加带来的健康需求；另一方面，也需要提高医疗服务效率和服务质量，以更好地管理慢性病。

此外，医疗技术的进步和医疗模式的变革也在不断改变着医疗卫生资源的需求和配置方式。数字化医疗、远程医疗服务等新兴技术的应用，为缓解医疗资源分布不均提供了新的解决方案。但同时，这也要求医疗卫生机构能够适应新技术的发展，进行相应的硬件升级和人员培训。

综上所述，虽然中国医疗卫生事业在过去几十年取得了长足的进步，但面对不断变化的社会经济环境和人口结构特征，仍需通过科学合理的资源配置，加强医疗卫生体系的建设，以满足人民群众日益增长的健康需求。本研究旨在通过分析 1990 年至 2023 年间我国各地区医疗卫生机构数量的变化趋势，探讨医疗卫生资源配置的现状和问题，并对未来几年的医疗卫生资源需求进行预测，为相关政策制定提供参考依据。

1.2 问题提出

1.2.1 问题 1

医疗卫生机构数量的时间趋势分析：通过对各地区医疗卫生机构数量的年均增长率进行计算，分析不同地区医疗卫生机构数量的变化趋势，确定哪些地区的增长最快，哪些地区增长较为缓慢。

1.2.2 问题 2

医疗卫生资源配置差异及其影响因素分析：基于探索性分析的结果，结合相关数据，讨论各地区医疗卫生资源的配置差异及其背后可能的影响因素，揭示中国各地区在医疗卫生机构数量上的变化趋势和区域差异，并探讨其背后的可能原因和政策意义。

1.2.3 问题 3

医疗卫生机构数量的未来预测：利用历史数据预测各地区未来 5 年（2024 年至 2028 年）的医疗卫生机构数量，评估预测的准确性，并讨论模型预测结果的可靠性和可能的不确定性因素，以及这些因素如何影响未来的医疗卫生资源配置。

1.2.4 问题 4

医疗卫生机构数量的宏观预测及其政策启示：在综合考虑当前医疗卫生资源配置、人口结构变化、经济发展水平和政策导向等因素的基础上，分析影响医疗卫生机构数量的关键因素，并对未来 10 年（2024 年至 2033 年）的医疗卫生机构数量进行宏观预测。讨论预测结果对医疗卫生政策制定、资源分配和体系建设的启示，以及可能面临的挑战 and 应对策略。

二、问题分析

2.1 问题 1 医疗卫生机构数量的时间趋势分析

在问题 1 中，我们将对各地区医疗卫生机构数量的时间趋势进行深入分析。首先，通过绘制医疗卫生机构数量随时间变化的折线图，直观展现各地区医疗卫生机构数量的增长态势。随后，我们计算各地区的年均增长率，并利用箱型图展示这些增长率的分布情况，以此识别增长较快和较慢的地区。通过这一系列分析，我们可以揭示医疗卫生资源在不同地区的动态变化趋势。

2.2 问题 2 医疗卫生资源配置差异及其影响因素分析

问题 2 要求我们基于探索性分析的结果，讨论各地区医疗卫生资源的配置差异及其可能的影响因素。我们将采用统计分析方法，结合相关数据，探讨各地区医疗卫生机构数量的变化趋势与区域差异。同时，我们还将分析影响医疗卫生资源配置的潜在因素，如经济发展水平、人口密度、政府政策等，并探讨这些因素如何导致了资源配置的不均衡现象。

2.3 问题 3 医疗卫生机构数量的未来预测

在问题 3 中，我们的目标是预测未来五年内各地区医疗卫生机构的数量。为此，我们将首先将历史数据划分为训练集和测试集，利用训练集数据构建预测模型。我们将比较几种常用的时间序列预测模型（如 ARIMA、LSTM 等）的效果，并选择最优模型进行预测。此外，我们还需要评估模型的预测精度，并讨论预测结果的可靠性和不确定性因素，特别是那些可能影响医疗卫生资源配置的因素。

2.4 问题 4 医疗卫生机构数量的宏观预测及其政策启示

问题 4 要求我们在综合考虑当前医疗卫生资源配置、人口结构变化、经济发展水平和政策导向等因素的基础上，对未来十年医疗卫生机构数量进行宏观预测。我们将通过分析这些关键因素与医疗卫生机构数量之间的关系，建立相应的预测模型。在此基础上，我们将讨论预测结果对医疗卫生政策制定、资源分配和体系建设的启示，同时指出可能面临的挑战和提出相应的应对策略。

三、模型假设

- ✧ 经济增长率在短期内保持稳定，对医疗卫生机构数量的影响相对固定。
- ✧ 人口数量及其结构在短期内变化不大，特别是在预测的五年内。
- ✧ 假设在未来几年内，没有重大的自然灾害或公共卫生事件发生。
- ✧ 不同地区医疗卫生机构数量的增长率保持相对稳定，受区域经济和社会发展水平的影响较大。

四、符号说明

符号	含义
N_t	第 t 年全国医疗卫生机构总数
$N_{r,t}$	第 t 年地区 r 的医疗卫生机构总数
$\Delta N_{r,t}$	地区 r 在第 t 年到第 $t+1$ 年医疗卫生机构数量的增量
$G_{r,t}$	地区 r 在第 t 年的年均增长率
P_t	第 t 年全国总人口数
$P_{r,t}$	第 t 年地区 r 的人口数
E_t	第 t 年全国人均 GDP
$E_{r,t}$	第 t 年地区 r 的人均 GDP
A	所有地区集合
T	研究的时间范围
α	模型中的常数项
β	模型中的系数，表示变量的影响程度
ε	模型中的误差项，代表未被模型捕捉的因素
$\hat{N}_{r,t}$	地区 r 在第 t 年医疗卫生机构数量的预测值

符号	含义
$\bar{N}_{r,t}$	地区 r 在第 t 年医疗卫生机构数量的平滑值
R^2	模型的决定系数，表示模型解释变异的比例
MSE	均方误差，用于评估预测值与实际值之间的偏差

五、模型的建立与求解

5.1 问题 1 医疗卫生机构数量的时间趋势分析

问题一的解题流程如下图所示：

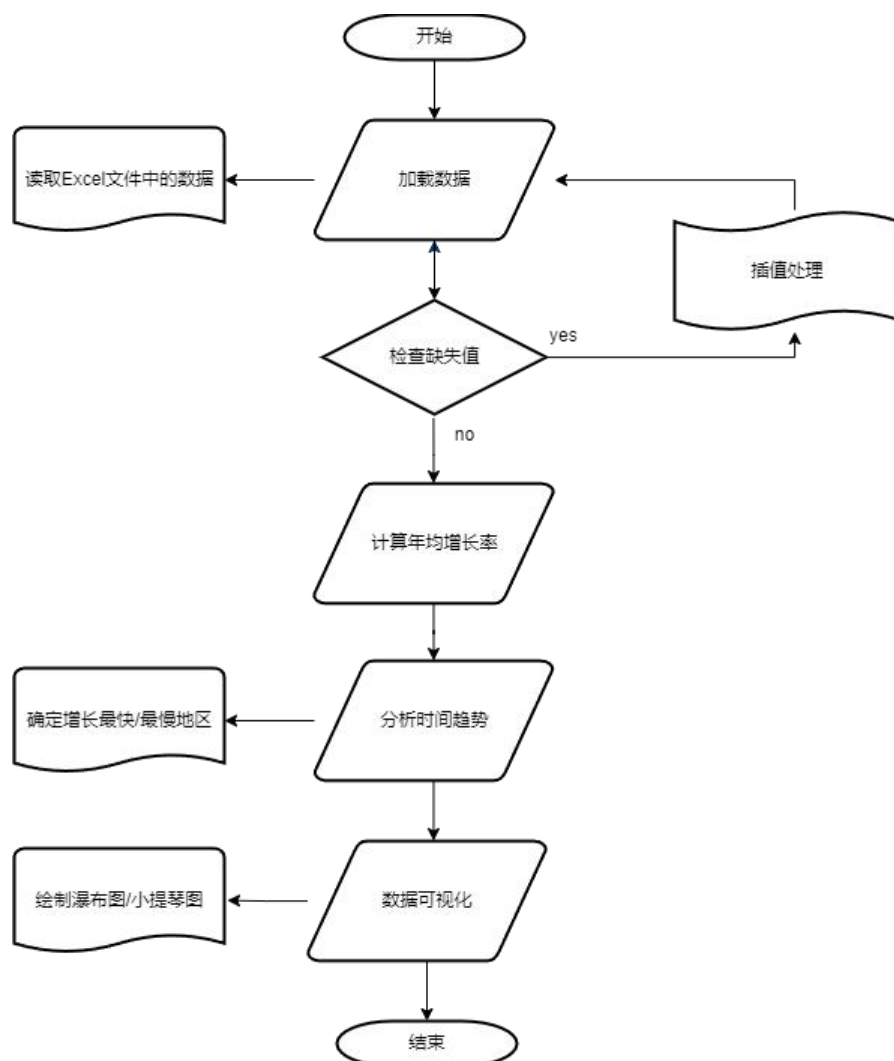


图 1 问题一的解题流程图

本研究的数据来源于中国知网中国经济社会大数据研究平台，具体数据见附件 1。这些数据记录了 1990 年至 2023 年间中国各地区医疗卫生机构的数量。

5.1.1 数据的清理与整理

本研究的数据来源于中国知网中国经济社会大数据研究平台，具体数据见附件 1。这些数据记录了 1990 年至 2023 年间中国各地区医疗卫生机构的数量。

通过对附件 1 的初步分析，发现存在部分缺失值。考虑到数据是各地区医疗卫生机构数，且为时间序列数据，我们对出重庆市之外的缺失值使用线性插补方法填补。由于重庆市的缺失值出现在序列的开始，我们采取了另一种插补策略，即使用重庆市后四年的医疗卫生机构数量平均值 3942.75 来进行填补。

5.1.2 模型的建立

5.1.2.1 年均增长率计算模型

$$G_t = \left(\frac{N_t}{N_0} \right)^{\frac{1}{T}} - 1 \quad \text{全国在第 } t \text{ 年的年均增长率}$$
$$G_{r,t} = \left(\frac{N_{r,t}}{N_0} \right)^{\frac{1}{T}} - 1 \quad \text{地区 } r \text{ 在第 } t \text{ 年的年均增长率}$$

5.1.2.2 增长趋势分析模型

利用年均增长率，我们可以分析各地区医疗卫生机构数量的增长趋势。通过比较不同地区的年均增长率，可以确定增长最快的地区和增长较慢的地区。

5.1.2.3 可视化模型

为了直观地展示不同地区的年均增长率，我们使用柱状图和箱线图进行可视化分析。柱状图可以清晰地显示每个地区的年均增长率，而箱线图则可以揭示增长率分布的中位数、四分位数等统计特征，帮助我们了解增长率的分布情况。

还采用了时间序列瀑布图和小提琴图。时间序列瀑布图有助于展现医疗卫生机构数量随时间的变化趋势，特别是对于理解增长的具体过程及其波动性非常有帮助。而小提琴图则可以提供比箱线图更丰富的信息，它不仅展示了中位数和四分位数，还能展示数据的密度分布，帮助我们了解不同地区增长率分布的差异性。

5.1.3 模型的求解

5.1.3.1 年均增长率计算模型

通过年均增长率的计算公式编写 Python 代码，代码思路的伪代码如下表所示：

表 1 计算各地区医疗卫生机构数量的年均增长率

伪代码：计算各地区医疗卫生机构数量的年均增长率

```
// 1. 加载数据
LoadData("附件 1 1990~2023 年各地区医疗卫生机构数(个).xlsx") // 读取数据

// 2. 数据预处理
SetIndex("年份") // 将"年份"列设置为 DataFrame 的索引
Regions = GetColumns() // 获取所有地区名作为列名

// 3. 数据插值
InterpolateMissingValues(LinearInterpolation()) // 对缺失值进行线性插值

// 4. 计算年均增长率
For Each Region In Regions:
    StartYear = GetMinIndex() // 获取数据开始年份
    EndYear = GetMaxIndex() // 获取数据结束年份
    NYears = EndYear - StartYear // 计算年份数量
    InitialValue = GetValueAt(StartYear, Region) // 获取初始年份的数量
    FinalValue = GetValueAt(EndYear, Region) // 获取最终年份的数量
    GrowthRate = CalculateGrowthRate(InitialValue, FinalValue, NYears) // 计算年均增长率
    StoreGrowthRate(Region, GrowthRate) // 存储年均增长率

// 5. 可视化结果
InitializePlot(14, 8) // 初始化绘图区域
BarChart(Regions, GrowthRates) // 绘制柱状图
Title("各地区医疗卫生机构数年均增长率")
XLabel("地区")
YLabel("年均增长率 (%)")
RotateXTicks(90) // 旋转 X 轴标签
ShowPlot() // 显示图表

// 6. 输出结果
For Each Region In Regions:
    PrintGrowthRate(Region, GrowthRates[Region]) // 输出每个地区的年均增长率
```


运行 Python 代码可以得到如下表的结果：

表 2 全国及各省的年均增长率

全国及各省的年均增长率							
中国	北京市	天津市	河北省	山西省	内蒙古	辽宁省	吉林省
0.1698%	2.8495%	1.7821%	6.8006%	5.6956%	4.9831%	4.4472%	5.4441%
黑龙江	上海市	江苏省	浙江省	安徽省	福建省	江西省	山东省
2.6822%	-0.5985%	3.3770%	4.2415%	4.5007%	5.6542%	6.1303%	6.4924%
河南省	湖北省	湖南省	广东省	广西	海南省	重庆市	四川省
7.1616%	4.0233%	5.1693%	6.0728%	5.5942%	2.0947%	5.5433%	4.9907%
贵州省	云南省	西藏	陕西省	甘肃省	青海省	宁夏	新疆
4.6072%	4.5280%	5.6963%	5.2729%	5.6503%	5.4186%	4.4307%	4.4174%

5.1.3.2 增长趋势分析模型

全国年均增长率：0.1698%，这表明在全国范围内，医疗卫生机构的数量增长相对缓慢。

最高增长率：河南省以 7.1616% 的年均增长率居首，显示出该省医疗卫生机构数量增长较快。

最低增长率：上海市的年均增长率最低，为 -0.5985%，意味着在所考察的时间段内，上海市医疗卫生机构的数量有所减少。

平均增长率：大多数省份的年均增长率集中在 4% 到 6% 之间，这表明整体上医疗卫生机构数量呈现稳定增长的趋势。

北京市（2.8495%）和天津市（1.7821%）的增长率低于全国平均水平，可能与这两个直辖市已经较高的医疗资源基数有关。

内蒙古（4.9831%）、黑龙江省（2.6822%）和海南省（2.0947%）的增长率相对较低，这可能与其地理环境或人口密度等因素有关。

河南省（7.1616%）的增长率远高于其他省份，可能是因为河南省作为人口大省，对医疗卫生服务的需求增加更快。

东部沿海省份（如江苏、浙江、山东等）的增长率普遍较高，这些地区经济较为发达，对医疗卫生服务的需求较大。

西部省份（如西藏、青海、宁夏等）的增长率也相对较高，这可能反映了政府对西部地区医疗卫生资源投入的增加。

中部省份（如河南、湖北、湖南等）的增长率较高，这与这些省份的人口规模和经济增长速度有关。

综上所述，全国各省份医疗卫生机构的年均增长率存在一定的差异，但总体上呈现出稳定增长的趋势。不同地区的增长率差异反映了各省份经济发展水平、人口结构和政策导向的不同。对于增长率较低或出现负增长的地区，需要进一步分析原因并采取相应措施以促进医疗卫生资源的均衡发展。

5.1.4.3 可视化模型

利用全国及各省的年均增长率，通过编写 JavaScript 代码生成一个基于中国地图的可视化图表，展示各地区医疗卫生机构数的年均增长率，如下图所示：

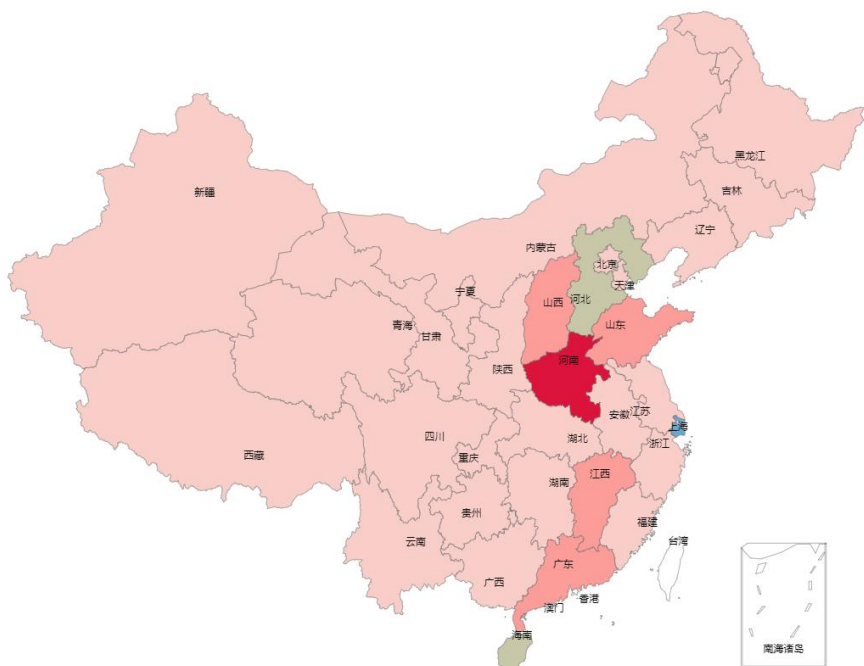
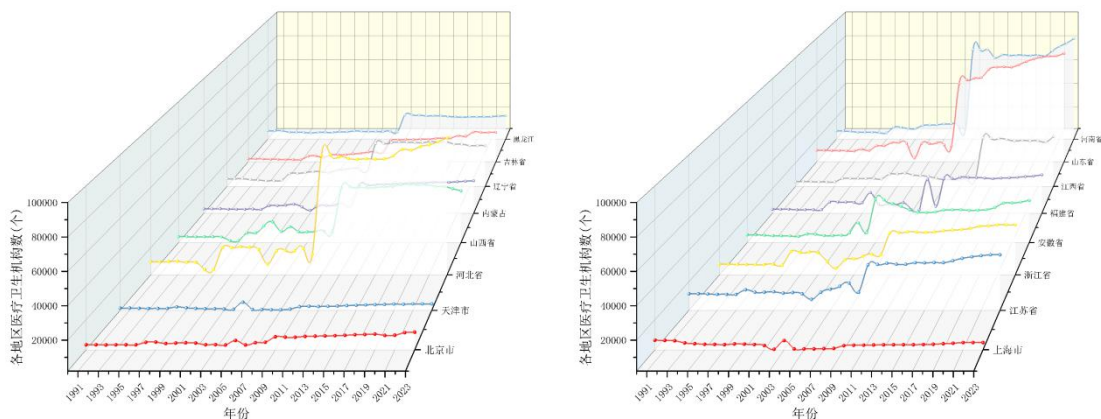


图 2 各地区医疗卫生机构数的年均增长率的中国地图热力图

绘制出各省市的 3D 瀑布时序图，如下所示：



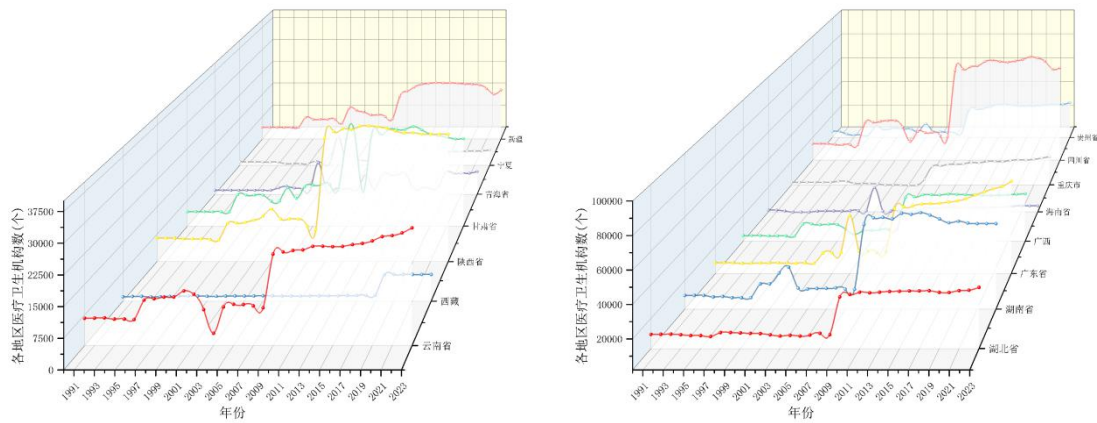


图 3 全国各省市医疗卫生机构数量 3D 瀑布时间序列图

采用 3D 瀑布时间序列图的形式展示了 1990 年至 2023 年间全国各省市医疗卫生机构数量的变化趋势。通过这种图形化的方法，我们可以直观地观察到每个时间点上医疗卫生机构数量的增减变化，并对不同地区之间的差异进行对比分析。

总体趋势：从 1990 年到 2023 年，全国医疗卫生机构的数量整体呈上升趋势，但增长速度并非线性，存在波动。在 1996 年和 2003 年左右出现了一个较为显著的下降，之后又逐渐恢复增长。

地区差异：不同地区的医疗卫生机构数量变化趋势存在显著差异。北京市和上海市的医疗卫生机构数量始终处于较高水平，并且增长速度相对稳定。西部地区如西藏、青海等地的医疗卫生机构数量增长相对较慢，但在最近几年有所加快。中部和东部地区如河南、山东等省份的医疗卫生机构数量增长迅速，反映了这些地区在医疗卫生领域的快速发展。

增长高峰期：1990 年代初期和中期，医疗卫生机构数量增长速度较快，这可能是由于国家在医疗卫生领域的投资增加。2000 年代初，增长速度放缓，可能与当时宏观经济环境和政策调整有关。2010 年之后，随着国家对医疗卫生事业的重视程度提高，医疗卫生机构数量的增长速度再次加快。

增长平缓期：某些地区在特定时间段内的增长速度较为平缓，这可能是由于地区人口增长放缓、医疗资源配置趋于饱和等原因造成的。2003 年前后多个地区的医疗卫生机构数量出现了短暂的下降，可能与当时的公共卫生事件有关。

周期性波动：个别地区的医疗卫生机构数量随时间呈现出一定的周期性波动，这可能与当地经济状况、政府政策变动等因素密切相关。

通过对 3D 瀑布时间序列图的分析，可以得出以下结论：

从 1990 年到 2023 年，全国医疗卫生机构数量整体呈上升趋势，但增长速度经历了波动。

地区之间存在明显的差异，反映出医疗卫生资源分配的不平衡。

需要采取更精准的策略来优化医疗卫生资源的配置，特别是在医疗卫生机构增长较慢的地区加强政策支持和资金投入，以促进医疗卫生事业的均衡发展。

5.2 问题 2 医疗卫生资源配置差异及其影响因素分析

问题 2 的解题流程如下图所示：

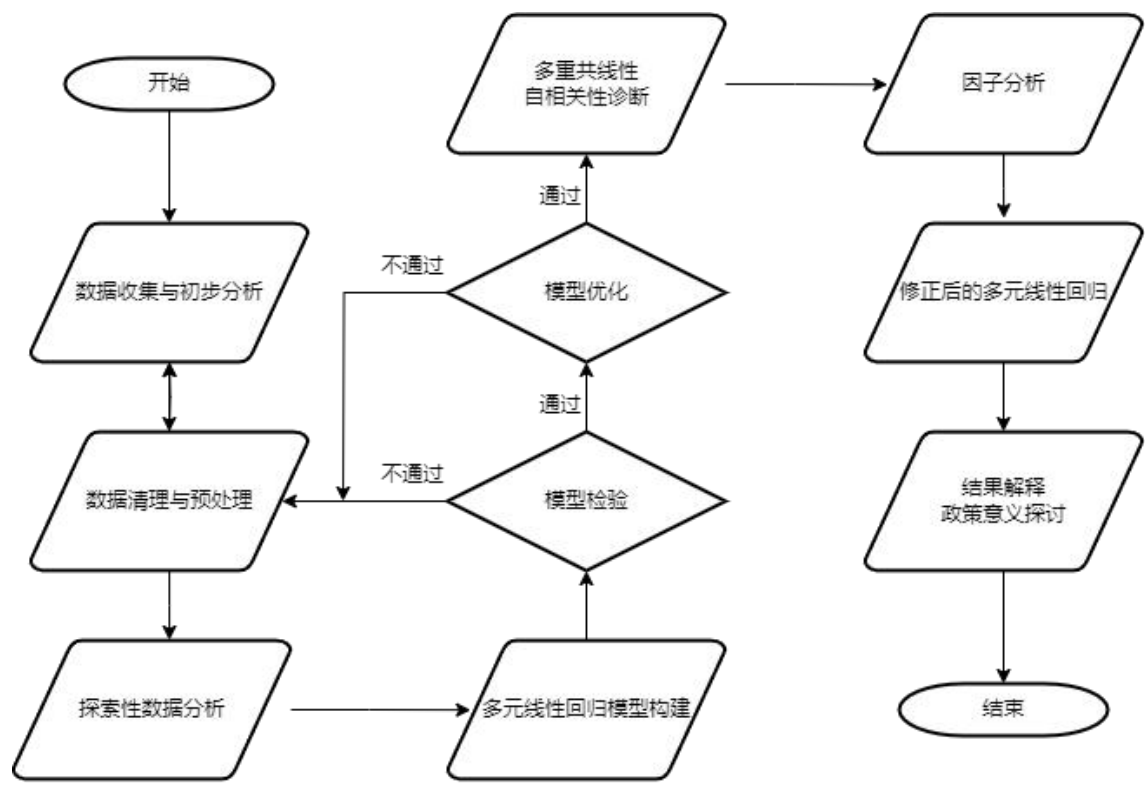


图 4 问题 2 的的解题流程图

5.2.1 数据的清理与整理

由于数据量过多，我们主要对问题一中所得到的年增长率具有代表性的省份进行分析。年增长率过高、过低，往往是由于国家政策的影响和对应省份在特定时期面临着特定的事件，所以，我们去除年增长率最高和最低两位数，对个省份再取平均值。得到的平均年增长率如下：

平均年增长率 4.7790%

根据平均年增长率，得到的代表性省份如下：

表 3 代表性省份年增长率排名表

省份	年增长率	排名
湖南省	5.1693%	15
四川省	4.9907%	16
内蒙古	4.9831%	17
贵州省	4.6072%	18
云南省	4.5280%	19
安徽省	4.5007%	20

为了深入探讨中国各地区医疗卫生资源的配置差异及其背后的影响因素，本研究不仅考虑了医疗卫生资源的数量和质量，还综合分析了经济社会因素对医疗卫生资源配置的影响。通过利用 中国经济社会大数据研究平台 (<https://data.cnki.net/>) 中的数据分析模块，获取了如表 3 所示的数据集。

表 4 新增数据集

数据集名称	数据集含义
医院数(个)	反映医疗卫生机构的数量
医院床位数(张)	反映医疗卫生机构的服务能力
医生数(人)	反映医疗卫生服务的专业人力资源
医院人员数(人)	反映医疗卫生服务领域的总人力资源
卫生人员数(人)	反映医疗卫生服务领域的总人力资源
常住人口(万人)	人口规模直接影响医疗卫生资源的需求量
职工工资总额(万元)	间接反映地区经济状况
职工平均工资(元)	反映地区的经济发展水平
城镇登记失业率(%)	反映地区的经济健康状况
城镇登记失业人员数(万人)	反映地区的经济健康状况
城镇居民消费水平(元)	反映城镇居民的生活水平
居民消费水平(元)	反映整个地区的生活水平
劳动者报酬(亿元)	反映劳动力市场的状况
农村居民消费水平(元)	反映农村地区的经济状况和生活水平
人均 GDP(元)	反映地区经济发展的水平
GDP(亿元)	反映地区的经济总量
人口自然增长率(%)	影响医疗卫生资源的需求

我们对数据集进行了系统性的符号化处理，旨在提高模型构建的效率与准确性。表 4 中，每个数据集被赋予了独特的符号标识。

表 5 数据集说明

符号	数据集	符号	数据集
X_1	职工平均工资(元)	X_{10}	农村居民消费水平(元)
X_2	职工工资总额(万元)	X_{11}	劳动者报酬(亿元)
X_3	医院数(个)	X_{12}	居民消费水平(元)
X_4	医院人员数(人)	X_{13}	城镇居民消费水平(元)
X_5	医院床位数(张)	X_{14}	城镇登记失业人员数(万人)

符号	数据集	符号	数据集
X_6	医生数(人)	X_{15}	城镇登记失业率(%)
X_7	卫生人员数(人)	X_{16}	常住人口(万人)
X_8	人口自然增长率(%)	X_{17}	GDP(亿元)
X_9	人均 GDP(元)	Y	各地区医疗卫生机构数(个)

数据清理的过程中，主要经过以下步骤：

- 1) 数据审查；
- 2) 缺失值处理；
- 3) 异常值检测；
- 4) 标准化处理；

经过以上步骤，发现原始数据集存在缺失值，为了保证数据集的有效性和真实性，我们对缺失值的部分，采用线性插补、均值填充的方法来完成。

线性插补、均值填充的伪代码如下表所示：

表 6 线性插补伪代码

伪代码：线性插补
<pre> # 定义线性插值填充数据的函数 function linearInterpolation(data, columnsToProcess): for each column in columnsToProcess: # 为每一列独立执行线性插值 data[column] = interpolate(data[column], 'linear') return data # 主程序 data = loadDataset('原数据.xlsx') // 加载数据 data = replaceSpecialValues(data, '--', NaN) // 将特殊值替换为 NaN data = linearInterpolation(data, columnsToProcess) // 应用线性插值 saveProcessedData(data, '线性插补后的数据.xlsx') // 保存处理后的数据 </pre>

表 7 均值填充伪代码

伪代码：均值填充

```
# 定义均值填充数据的函数
function meanImputation(data, columnsToProcess):
    for each column in columnsToProcess:
        meanValue = calculateMean(data[column]) // 计算列的均值
        data[column] = replaceNaNWith(data[column], meanValue) // 用均值替换 NaN
    return data

# 主程序
data = loadDataset('原数据.xlsx') // 加载数据
data = replaceSpecialValues(data, '--', NaN) // 将特殊值替换为 NaN
data = meanImputation(data, columnsToProcess) // 应用均值填充
saveProcessedData(data, '均值填充后的数据.xlsx') // 保存处理后的数据
```

编写 Python 代码，实现对数据的线性插补、均值填充和标准化处理（以中国数据为例），代码运行结果如下表所示：

表 8 标准化后的数据

年份	职工平均工资 (元)	职工工资总额 (万元)	医院数 (个)	医院人员数 (人)	医院床位数 (张)	医生数 (人)	卫生人员数 (人)	人口自然增长率(%)	人均GDP (元)	农村居民消费水平 (元)	劳动者报酬 (亿)	居民消费水平 (元)	城镇居民消费水平 (元)	城镇登记失业人员数 (万)	城镇登记失业率(%)	常住人口(万)	GDP (亿)	各地区医疗卫生机构数(个)
1990	-0.99	-1.03	-1.06	-0.79	-0.93	0.17	-0.93	2.21	-1.04	-0.93	0.75	-1.04	-1.21	-1.71	-1.83	-2.03	-1.01	1012690
1991	-0.99	-1.02	-1.03	-0.79	-0.90	0.18	-0.88	1.82	-1.03	-0.93	0.75	-1.03	-1.19	-1.84	-2.15	-1.84	-1.00	1003769
1992	-0.97	-1.01	-0.99	-0.79	-0.88	0.20	-0.83	1.44	-1.01	-0.92	0.75	-1.02	-1.16	-1.79	-2.15	-1.68	-0.99	1001310
1993	-0.96	-0.99	-0.92	-0.79	-0.85	0.22	-0.78	1.40	-0.99	-0.90	0.75	-0.99	-1.10	-1.56	-1.67	-1.51	-0.97	1000531
1994	-0.92	-0.95	-0.90	-0.79	-0.83	0.25	-0.75	1.33	-0.95	-0.86	0.75	-0.95	-1.02	-1.32	-1.34	-1.35	-0.93	1005271
1995	-0.90	-0.91	-0.89	-0.79	-0.83	0.28	-0.72	1.15	-0.91	-0.81	0.75	-0.89	-0.93	-1.14	-1.18	-1.19	-0.90	994409
1996	-0.89	-0.89	-0.87	-0.79	-0.82	0.29	-0.71	1.12	-0.88	-0.76	0.75	-0.85	-0.88	-1.00	-1.02	-1.04	-0.87	1078131
1997	-0.87	-0.88	-0.85	-0.79	-0.81	0.33	-0.68	1.02	-0.86	-0.74	0.58	-0.83	-0.86	-0.90	-0.86	-0.89	-0.85	1048657
1998	-0.84	-0.89	-0.84	-0.79	-0.80	0.34	-0.67	0.76	-0.85	-0.74	0.69	-0.82	-0.83	-0.92	-0.86	-0.75	-0.84	1042885
1999	-0.82	-0.87	-0.75	-0.79	-0.79	0.37	-0.65	0.34	-0.83	-0.74	0.80	-0.79	-0.80	-0.90	-0.86	-0.62	-0.83	1017673
2000	-0.79	-0.86	-0.80	-0.79	-0.78	0.39	-0.65	0.34	-0.81	-0.71	0.91	-0.76	-0.75	-0.82	-0.86	-0.51	-0.80	1034229
2001	-0.75	-0.83	-0.82	-0.79	-0.79	0.41	-0.66	0.16	-0.78	-0.70	1.08	-0.73	-0.72	-0.46	-0.06	-0.40	-0.77	1029314
2002	-0.71	-0.80	-0.60	-0.79	-0.75	0.23	-0.79	0.02	-0.75	-0.67	1.26	-0.70	-0.69	-0.08	0.58	-0.30	-0.75	1005004
2003	-0.67	-0.77	-0.61	-0.79	-0.73	0.24	-0.90	-0.10	-0.71	-0.65	1.50	-0.67	-0.66	0.04	1.07	-0.20	-0.71	806243
2004	-0.61	-0.72	-0.52	-0.76	-0.68	0.27	-0.86	-0.13	-0.64	-0.61	1.74	-0.62	-0.60	0.16	0.91	-0.11	-0.64	849140
2005	-0.54	-0.65	-0.48	-0.73	-0.64	0.29	-0.82	-0.13	-0.57	-0.57	1.99	-0.56	-0.52	0.21	0.91	-0.01	-0.58	882206
2006	-0.48	-0.57	-0.41	-0.66	-0.58	-1.07	-0.73	-0.30	-0.49	-0.52	0.50	-0.50	-0.45	0.24	0.74	0.07	-0.49	918097
2007	-0.37	-0.46	-0.33	-0.54	-0.53	-1.07	-0.63	-0.33	-0.35	-0.44	-0.99	-0.39	-0.31	0.17	0.58	0.16	-0.36	912263
2008	-0.25	-0.34	-0.34	-0.46	-0.42	-1.07	-0.52	-0.35	-0.22	-0.37	-0.97	-0.28	-0.19	0.40	0.91	0.24	-0.24	891480
2009	-0.16	-0.19	-0.27	-0.33	-0.30	-1.07	-0.33	-0.41	-0.14	-0.31	-0.96	-0.21	-0.11	0.55	1.07	0.32	-0.16	916571
2010	-0.04	-0.04	-0.18	-0.20	-0.17	-1.07	-0.17	-0.43	0.03	-0.23	-0.95	-0.08	0.05	0.50	0.74	0.40	0.00	936927
2011	0.12	0.25	-0.04	-0.05	-0.01	-1.07	-0.02	-0.43	0.23	-0.05	-0.95	0.13	0.27	0.56	0.74	0.50	0.20	954389
2012	0.26	0.49	0.12	0.16	0.22	-1.07	0.16	-0.39	0.36	0.07	-0.95	0.27	0.41	0.53	0.74	0.62	0.33	950297

年份	职工平均工资 (元)	职工工资总额 (万元)	医院数 (个)	医院人员数 (人)	医院床位数 (张)	医生数 (人)	卫生人员数 (人)	人口自然增长率 (%)	人均GDP (元)	农村居民消费水平 (元)	劳动者报酬 (亿)	居民消费水平 (元)	城镇居民消费水平 (元)	城镇登记失业人员数 (万)	城镇登记失业率 (%)	常住人口 (万)	GDP (亿)	各地区医疗卫生机构数 (个)
2013	0.39	0.99	0.33	0.38	0.42	-1.07	0.41	-0.40	0.49	0.21	-0.95	0.42	0.55	0.57	0.66	0.72	0.47	974398
2014	0.50	1.21	0.48	0.58	0.62	-1.07	0.57	-0.32	0.62	0.37	-0.95	0.58	0.70	0.68	0.73	0.83	0.60	981432
2015	0.69	1.41	0.71	0.77	0.80	-1.07	0.74	-0.38	0.73	0.55	-0.95	0.74	0.84	0.74	0.66	0.92	0.71	983528
2016	0.85	1.41	0.92	0.98	0.98	-1.07	0.91	-0.14	0.88	0.75	-0.95	0.93	1.01	0.81	0.62	1.03	0.86	983394
2017	1.04	1.41	1.18	1.21	1.19	-1.07	1.13	-0.29	1.09	1.01	-0.95	1.15	1.19	0.77	0.42	1.12	1.08	986649
2018	1.28	1.41	1.44	1.41	1.39	-1.07	1.33	-0.70	1.31	1.32	-0.95	1.37	1.37	0.77	0.26	1.19	1.30	997433
2019	1.52	1.41	1.62	1.62	1.57	1.62	1.56	-0.83	1.47	1.56	-0.95	1.59	1.57	0.65	-0.03	1.25	1.49	1007579
2020	1.72	1.41	1.76	1.79	1.70	1.80	1.76	-1.35	1.54	1.67	-0.95	1.59	1.50	1.56	0.97	1.27	1.55	1023000
2021	1.98	1.41	1.98	1.97	1.84	1.94	1.95	-1.65	1.85	2.10	-0.95	1.95	1.82	1.05	0.52	1.28	1.89	1031000
2022	2.09	1.41	1.98	1.97	1.97	2.03	2.10	-1.91	2.02	2.26	-0.95	2.01	1.85	1.74	0.52	1.27	2.05	1033000
2023	2.09	1.41	1.98	1.97	2.13	2.03	2.10	-2.15	2.16	2.26	-0.95	2.01	1.85	1.74	0.52	1.27	2.18	1071000

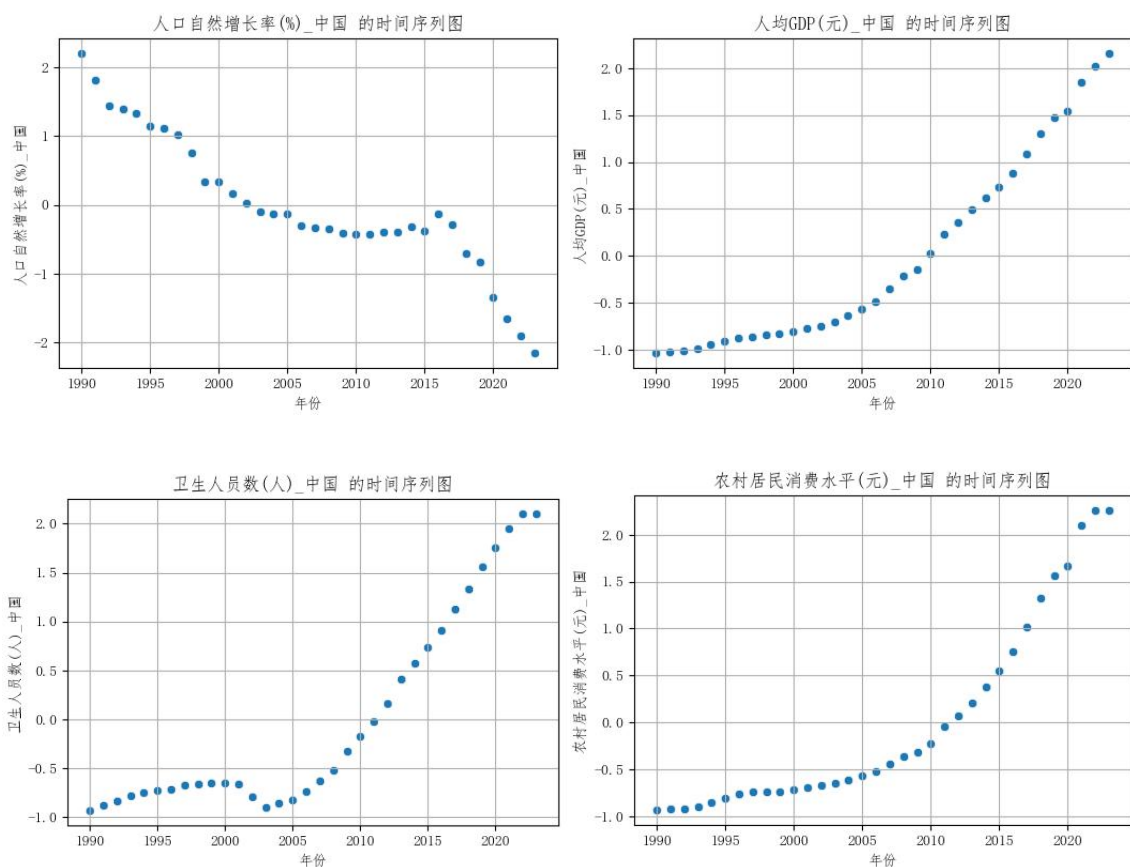


图 5 部分变量的时间序列图

通过上面这些图表，不仅能够观察到各项指标随时间的变化趋势，还能识别出数据集中的异常值和长期趋势。通过线性插补、均值填充（填充缺失值）和标准化（消除量纲影响）处理后，发现，数据集中并不存在十分明显的异常值。

基于上述内容，本研究将构建一个多元线性回归模型来研究中国各地区在医疗卫生机构数量上的变化趋势和区域差异，并探讨其背后的可能原因和政策意义。

5.2.2 模型的建立

5.2.2.1 多元线性回归

本研究将构建一个多元线性回归模型来探讨不同自变量如何影响各地区医疗卫生机构的数量。这些自变量包括职工平均工资、职工工资总额、医院人员数、医院床位数、医生数、卫生人员数、人口自然增长率、人均 *GDP*、农村居民消费水平、劳动者报酬、居民消费水平、城镇居民消费水平、城镇登记失业人员数、城镇登记失业率、常住人口以及 *GDP*。因变量为各地区医疗卫生机构的数量。

1). 模型建立的基本原理和基本思想：多元线性回归是一种常用的统计技术，用于分析多个自变量对单一因变量的影响。在数学模型中，因变量（ Y ）被假定为各自变量（ X_1, X_2, \dots, X_n ）的线性组合。

假设模型如下：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

其中：（ Y ）是因变量（各地区医疗卫生机构的数量）、（ X_1, X_2, \dots, X_n ）是自变量，包括职工平均工资、职工工资总额、医院人员数、医院床位数、医生数、卫生人员数、人口自然增长率、人均 GDP 、农村居民消费水平、劳动者报酬、居民消费水平、城镇居民消费水平、城镇登记失业人员数、城镇登记失业率、常住人口和 GDP ；

2). 构建和改进模型的详细过程：模型检验，采用统计检验（ F 检验、 t 检验）来审查模型的有效性和各系数的显著性。同时，评估模型的拟合优度（如 R^2 ）。

3). 模型优化：变量选择，采用向后或向前选择方法，通过 AIC （赤池信息量准则）、 BIC （贝叶斯信息准则）等标准确定模型中哪些变量应被包含；多重共线性问题，使用方差膨胀因子（ VIF ）来检测并处理可能的多重共线性问题。

5.2.2.2 因子分析

因子分析是一种降维、简化数据的技术。它通过研究众多变量之间的内部依赖关系，探求观测数据中的基本结构，并用少数几个“抽象”的变量来表示其基本的数据结构。这几个抽象的变量被称作“因子”，能反映原来众多变量的主要信息。原始的变量是可观测的显在变量，而因子一般是不可观测的潜在变量。因子分析中的公共因子是不可直接观测但又客观存在的共同影响因素；每一个变量都可以表示成公共因子的线性函数与特殊因子之和，它的数学模型可表示为：

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ a_{21} & \cdots & a_{2p} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pm} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

即 $X = AF + \varepsilon$ 。其中 $X = (x_1, x_2, \dots, x_p)^T$ 是可观测的 p 维随机向量，每个分量代表一个指标或者向量。 $F = (F_1, F_2, \dots, F_m)^T$ 中的 F_1, F_2, \dots, F_m 为 $m(m \leq p)$ 个公因子变量，是各个原观测变量的表达式中都出现的因子，是相互独立的不可观测的理论变量。矩阵 A 称为因子载荷矩阵， a_{ij} 称为因子载荷，表示第 i 个原有变量和第 j 个公共因子变量的相关系数， a_{ij} 越大说明公共因子 F_j 和原有变量 X 的相关性越强。 ε 为特殊因子，表示原有变量不能被公共因子变量所解释的部分，相当于多元线性回归分析中的残差部分。

因子分析利用了降维的思想，由研究原始变量相关矩阵内部的依赖关系出发，根据相关性的把原始变量分组，使得同组内的变量之间相关性高，而不同组的变量间的相关性较低。每组变量代表一个基本结构，并用一个不可观测的综合变量表示，这个基本结构就称为公共因子。抓住这些主要的因子就可以帮助我们对复杂的问题进行分析和解释。

5.2.3 模型的求解

最小二乘法：通过最小化残差平方和（RSS）来进行参数估计。

计算残差平方和：（ $RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ ），其中（ \hat{Y}_i ）是预测值。

求解正规方程：（ $X^T X \hat{\beta} = X^T Y$ ）其中（ X ）是设计矩阵，（ Y ）是因变量向量，（ $\hat{\beta}$ ）是参数估计向量。解出正规方程： $\hat{\beta} = (X^T X)^{-1} X^T Y$

5.2.3.1 多重线性诊断

在进行多元线性回归分析之前，首先使用热力图对数据间的相关性进行了初步探索，以确定哪些变量之间存在显著的相关关系。这一步骤对于后续模型构建和变量选择是十分关键的，因为它帮助识别出可能的多重共线性问题，以及选出与目标变量有显著相关性的预测因子。

编写 Python 代码，实现对自变量之间的多重共线性初步分析，如下图“变量间的相关系数矩阵热力图”所示：

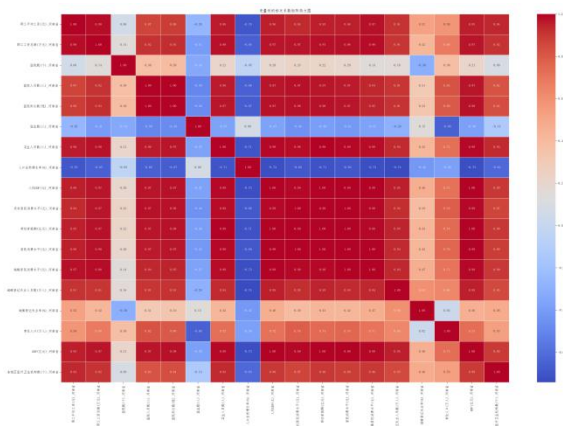


图 6 变量间的相关系数矩阵热力图_河南省

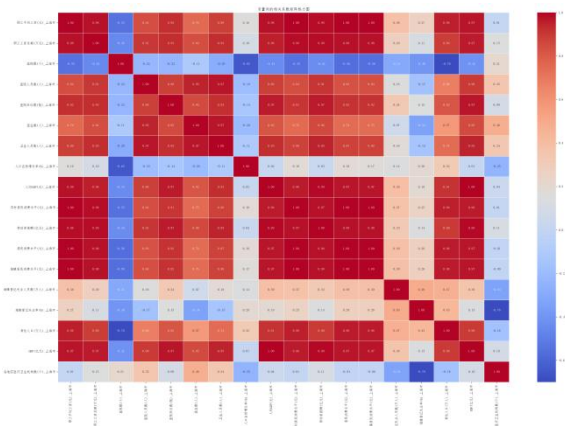


图 7 变量间的相关系数矩阵热力图_上海市

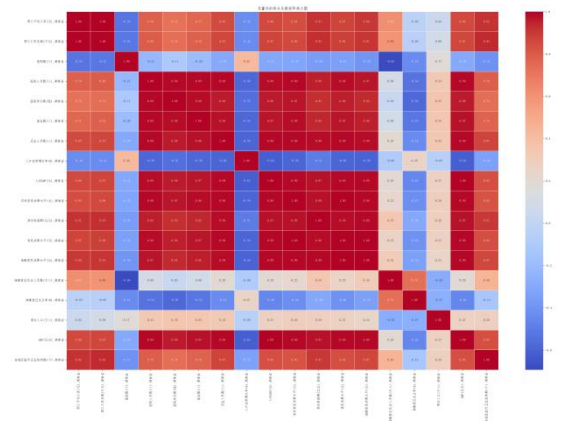


图 8 变量间的相关系数矩阵热力图_四川省

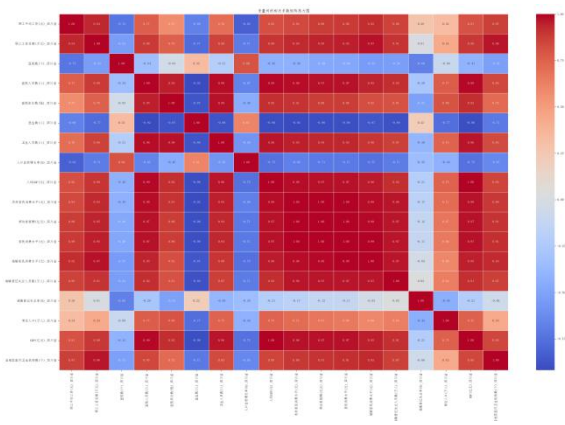


图 9 变量间的相关系数矩阵热力图_湖南省

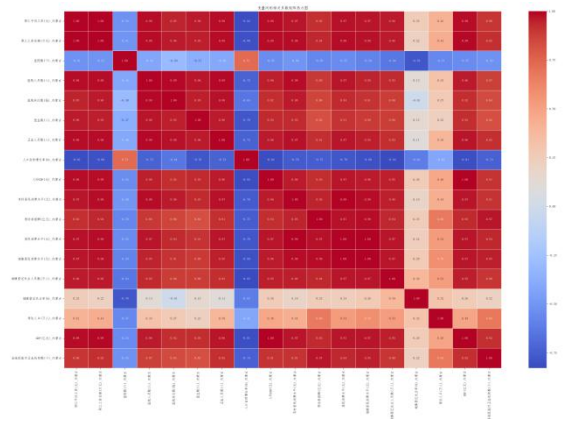


图 10 变量间的相关系数矩阵热力图_内蒙古

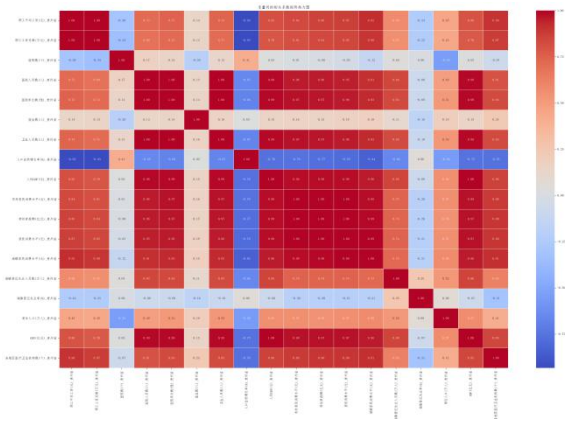


图 11 变量间的相关系数矩阵热力图_贵州省

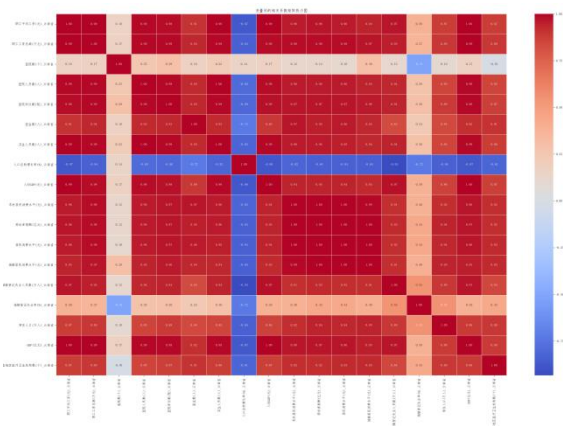


图 12 变量间的相关系数矩阵热力图_云南省

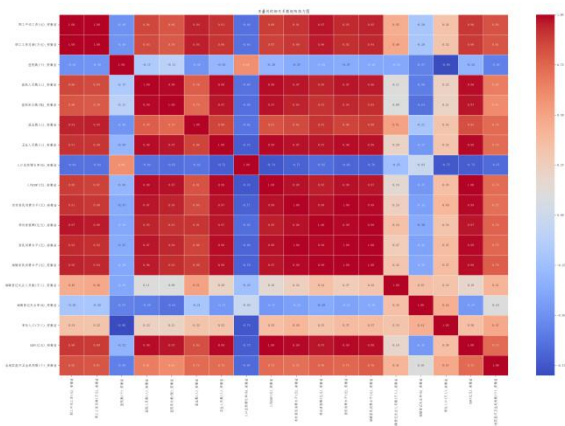


图 13 变量间的相关系数矩阵热力图_安徽省

通过上述的热力图，可以发现，变量之间存在严重的多重共线性问题。为了验证，我们进行方差膨胀因子（Variance Inflation Factor，VIF）检验，以诊断模型中的多重共线性问题。VIF 值大于 10 表示变量存在较严重的共线性。通过 VIF 检验，可以有效地评估并处理模型中的多重共线性问题（以湖南省为例），结果如下所示。

表 9 方差膨胀因子（Variance Inflation Factor，VIF）检验表

系数 ^a							
模型	未标准化系数		标准化系数	t	显著性	共线性统计	
	B	标准错误	Beta			容差	VIF
(常量)	-37978.411	83667.691		-.454	.655		
职工平均工资(元)_湖南省	.876	3.496	.441	.250	.805	.000	5205.902
职工工资总额(万元)_湖南省	-.002	.007	-.368	-.209	.836	.000	5191.789
医院数(个)_湖南省	-.688	1.160	-.039	-.593	.561	.136	7.348
医院人员数(人)_湖南省	-.309	.283	-1.216	-1.093	.289	.000	2081.026
医院床位数(张)_湖南省	.039	.068	.176	.574	.573	.006	157.652
医生数(人)_湖南省	-.252	.124	-.477	-2.037	.057	.011	91.968
卫生人员数(人)_湖南省	.513	.073	3.360	6.991	.000	.003	388.246
人口自然增长率(%)_湖南省	1108.987	576.506	.177	1.924	.070	.070	14.205
农村居民消费水平(元)_湖南省	-6.173	4.065	-1.458	-1.518	.146	.001	1550.036
劳动者报酬(亿元)_湖南省	1.851	2.104	.530	.880	.391	.002	610.065
城镇居民消费水平(元)_湖南省	.338	1.577	.162	.214	.833	.001	959.497
城镇登记失业人员数(万人)_湖南省	-3.963	506.846	-.002	-.008	.994	.013	78.436
城镇登记失业率(%)_湖南省	-171.844	5418.860	-.005	-.032	.975	.022	44.595
常住人口(万人)_湖南省	-2.297	13.444	-.009	-.171	.866	.198	5.060
GDP(亿元)_湖南省	-.268	.678	-.191	-.396	.697	.003	390.347

根据检验表可以初步得到方程如下：

$$\begin{aligned}
 Y = & -37978.411 + 0.876X_1 - 0.002X_2 - 0.688X_3 - 0.309X_4 + 0.039X_5 \\
 & - 0.252X_6 + 0.513X_7 + 1108.987X_8 - 6.173X_9 + 1.851X_{10} \\
 & + 0.338X_{11} - 3.963X_{12} - 171.844X_{13} - 2.297X_{14} - 0.268X_{15}
 \end{aligned}$$

VIF 是用来量化回归模型中解释变量之间多重共线性的指标。具体来说，VIF 衡量了某个特定独立变量与其他独立变量共线性的强度和程度。如果 VIF 值较大（通常大于 10 或某些学者推荐的阈值 5），则表明相应的解释变量可能与其他一个或多个解释变量高度线性相关，从而影响该变量在模型中的稳定性与解释力。

VIF 的计算基于以下公式： $[VIF_i = \frac{1}{1-R_i^2}]$ 其中， (R_i^2) 是第 i 个变量作为因变量，对所有其他自变量进行线性回归所得到的决定系数。

职工平均工资(元)、职工工资总额(万元)等变量的 VIF 值非常高，远大于通常的阈值 10，这表明这些变量与模型中的其他变量存在严重的多重共线性问题。

相对地，医院数(个)的 VIF 值为 7.348，虽然超过了 5，但低于 10，这表明它们具有一定程度的共线性，但不像前述变量那么严重。

VIF 值高的处理策略：移除变量：可以考虑从模型中移除 VIF 值过高的变量，尤其是那些对模型影响不大或者可由其他变量替代的。

总的来说，VIF 值的分析对于识别和处理多重共线性至关重要，可以提高回归模型的准确性和解释能力。

5.2.3.2 自相关性诊断

自相关性是：残差之间的相关性，表示了残差序列是否包含了时间的依赖性或者未建模的因素。残差之间的相关性进行如下假设：

H_0 : 残差 ε_i 之间不存在自相关

H_1 : 残差 ε_i 之间存在自相关

根据以下数据：

指标	值
Durbin-Watson 统计	0.013

理想情况下，此值接近 2 时表示残差序列之间无自相关。值接近 0 表示有强正自相关。Durbin-Watson 计算公式如下：

$$d = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2}$$

在模型中，Durbin-Watson 统计量为 0.013，接近 0，表明在模型的残差中存在明显的正自相关。这会影响模型的可靠性和预测的准确性。

通过上述分析方法，发现原始数据具有严重的共线性、存在正自相关性，所以，对多元线性回归模型进行改进，改进方法如下：

移除那些 VIF 值非常高的变量，保留 VIF 值较高和较低的变量，再通过因子分析，来减弱变量之间的共线性，自相关性。

修改 Python 代码，得到的结果如下表所示：

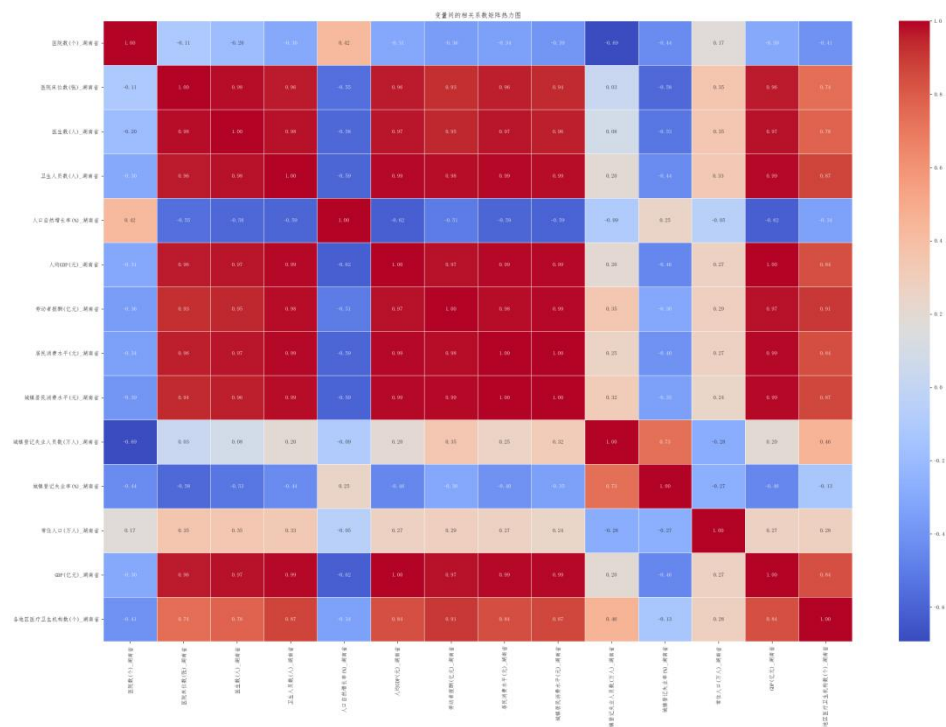


图 14 修改后的变量间相关系数矩阵热力图

剔除具有十分严重的共线性的变量，相关系数矩阵热力图中深色区域大量减少，进一步说明了剔除后的数据，不具有严重的共线性，再进行因子分析，并探讨其背后的可能原因和政策意义。

5.2.3.3 因子分析

1) 因子分析适用性检验

利用 SPSS 软件对各省市的医疗卫生数据进行 KMO 和 Bartlett 球度适用性检验，结果如表 2 所示，一般认为 KMO 度量值若大于 0.5，则可以进行因子分析。且显著性 $p=0$ ，说明原有变量之间存在一定的关联性，具备进行因子分析的条件。

2) 提取公因子

对数据进行因子分析，通过主成分分析法进行主成分的提取。绘制的碎石图如下所示：

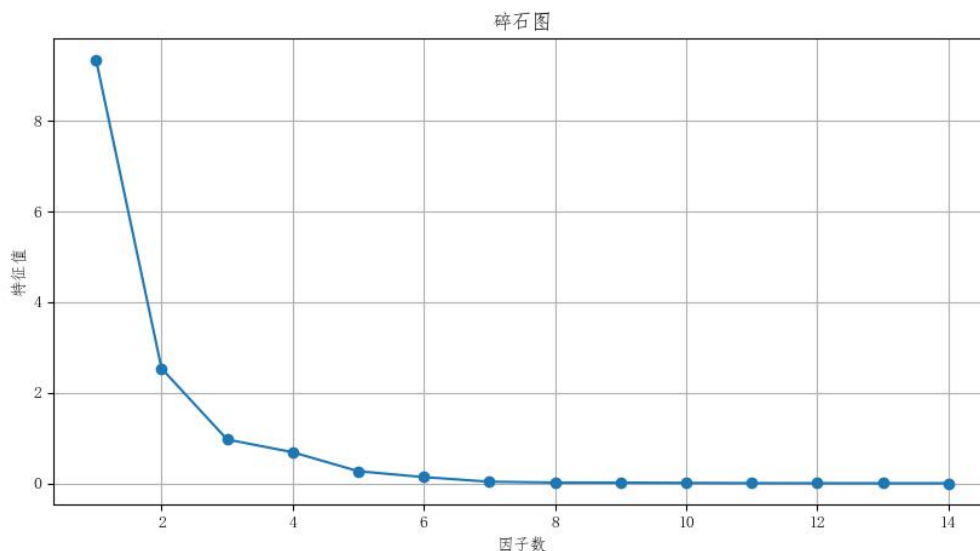


图 15 特征值的碎石图

在特征值为 1 的原则下，保留两个主因子，即 14 个变量归为 2 类。

表 10 KMO 和巴特利特检验

KMO 和巴特利特检验		
KMO 取样适切性量数		0.781
	近似卡方	1422.524
巴特利特球形度检验	自由度	91
	显著性	0.000

表 11 总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	9.347	66.766	66.766	9.280	66.286	66.286
2	2.532	18.083	84.849	2.344	16.741	83.027
3	.965	6.893	91.743			
4	.684	4.883	96.625			
5	.265	1.894	98.519			
6	.135	.965	99.484			
7	.033	.238	99.722			
8	.015	.104	99.827			
9	.013	.093	99.920			
10	.007	.053	99.972			
11	.003	.021	99.993			
12	.001	.006	99.999			
13	.000	.001	100.000			
14	4.714E-6	3.367E-5	100.000			

3) 公共因子命名

通过提取出来的 2 个公共因子，进行最大方差正交旋转，对原始因子载荷矩阵进行旋转，得到方差最大正交旋转矩阵，如下表所示。

表 12 旋转后的因子载荷矩阵

旋转后的因子载荷矩阵		
	Factor1	Factor2
医院数(个)	-0.32204	-0.67731
医院床位数(张)	0.95927	-0.22696
医生数(人)	0.97789	-0.15500
卫生人员数(人)	1.00067	-0.03166
人口自然增长率(%)	-0.57628	-0.04547
人均 GDP(元)	0.99802	-0.02727
劳动者报酬(亿元)	0.98386	0.11067
居民消费水平(元)	0.99827	0.02426
城镇居民消费水平(元)	0.99527	0.09276
城镇登记失业人员数(万人)	0.22847	0.97300
城镇登记失业率(%)	-0.41522	0.83015
常住人口(万人)	0.28672	-0.29654
GDP(亿元)	0.99833	-0.03298
各地区医疗卫生机构数(个)	0.84274	0.24097

根据旋转后的成分矩阵，可将 2 个公共因子进行命名。

第一个因子 Z1 在医院床位数、医生数、卫生人员数、人均 GDP、劳动者报酬、居民消费水平、城镇居民消费水平、GDP、各地区医疗卫生机构数指标上具有较大载荷；第二个因子 Z2 在城镇登记失业人员数、城镇登记失业率指标上具有较大载荷。

可以发现，Z1 所对应的评价指标是“经济医疗丰裕因子”，Z2 所对应的评价指标是“就业市场压力因子”分别命名为经济医疗丰裕因子、就业市场压力因子。

通过因子分析后，再对因子分析得到的得分矩阵进行多元线性回归，得到的结果如下表所示：

表 13 多元线性回归分析

	B	标准误差	自由度	显著性	EXP(B) 的 95% 置信区间	
					下限	上限
经济医疗丰裕因子	1.693e+04	2123.718	1	0.000	1.26e+04	2.13e+04
就业市场压力因子	4515.9143	2098.865	1	0.039	235.252	8796.577
常量	3.372e+04	2232.437	1	0.000	2.92e+04	3.83e+04

4) 由多因素的回归分析，建立多元线性回归，回归方程：

$$Y_p = 3.372e + 04 + 1.693e + 04 * Z_1 - 4515.9143 * Z_2$$

$$P = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

揭示了经济医疗丰裕因子对因变量具有显著且强有力的影响，而就业市场压力因子的影响虽较弱但依然显著。这些发现强调了经济与医疗资源的丰裕度以及就业市场状况对研究目标（医疗卫生机构数量）的重要作用，为政策制定和资源分配提供了数据支持。

医疗卫生资源与经济发展水平：作为主要影响因素之一，医疗卫生资源与经济发展水平的提升对增加医疗卫生机构数量有显著的正面影响。

城镇就业状况：虽然城镇就业状况因子与医疗卫生机构数量之间存在显著关系，但具体影响程度需要进一步分析才能确定。

政策意义：本研究的结果表明，提高医疗卫生资源与经济发展水平，改善城镇就业状况，是促进医疗卫生机构数量增长的重要途径。政府可以通过制定相关政策来加强医疗卫生资源的配置，提高经济发展水平，同时关注城镇就业问题，以实现医疗卫生资源的均衡分配和发展。

5.3 问题 3 医疗卫生机构数量的未来预测

问题 2 的解题流程如下图所示：

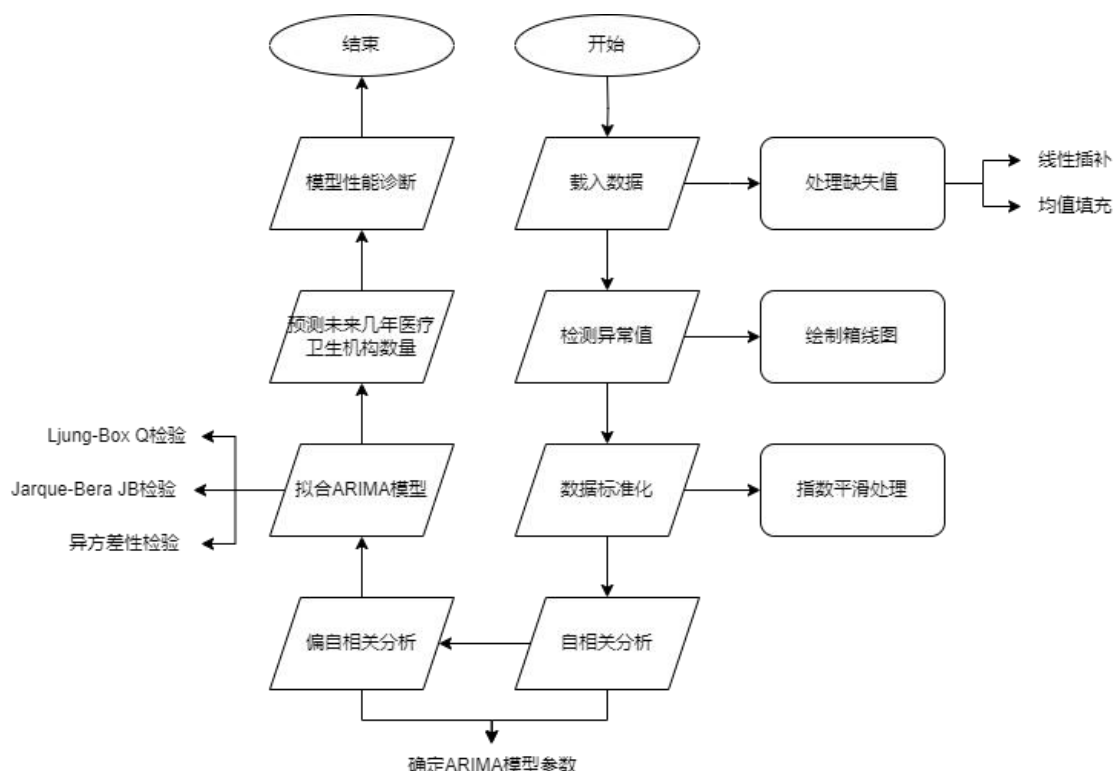


图 16 问题 3 的解题流程图

5.3.1 数据的清理与整理

- 1). 缺失值处理，在问题一、问题二中已经进行了处理（线性插补、均值填充）
- 2). 异常值检测，为了识别并处理异常值，我们绘制了各省市医疗卫生机构数量的箱线图。通过这些箱线图，我们可以直观地观察到数据的分布情况以及潜在的异常值。具体而言，箱线图上的异常值是指位于上下须触线之外的数据点。

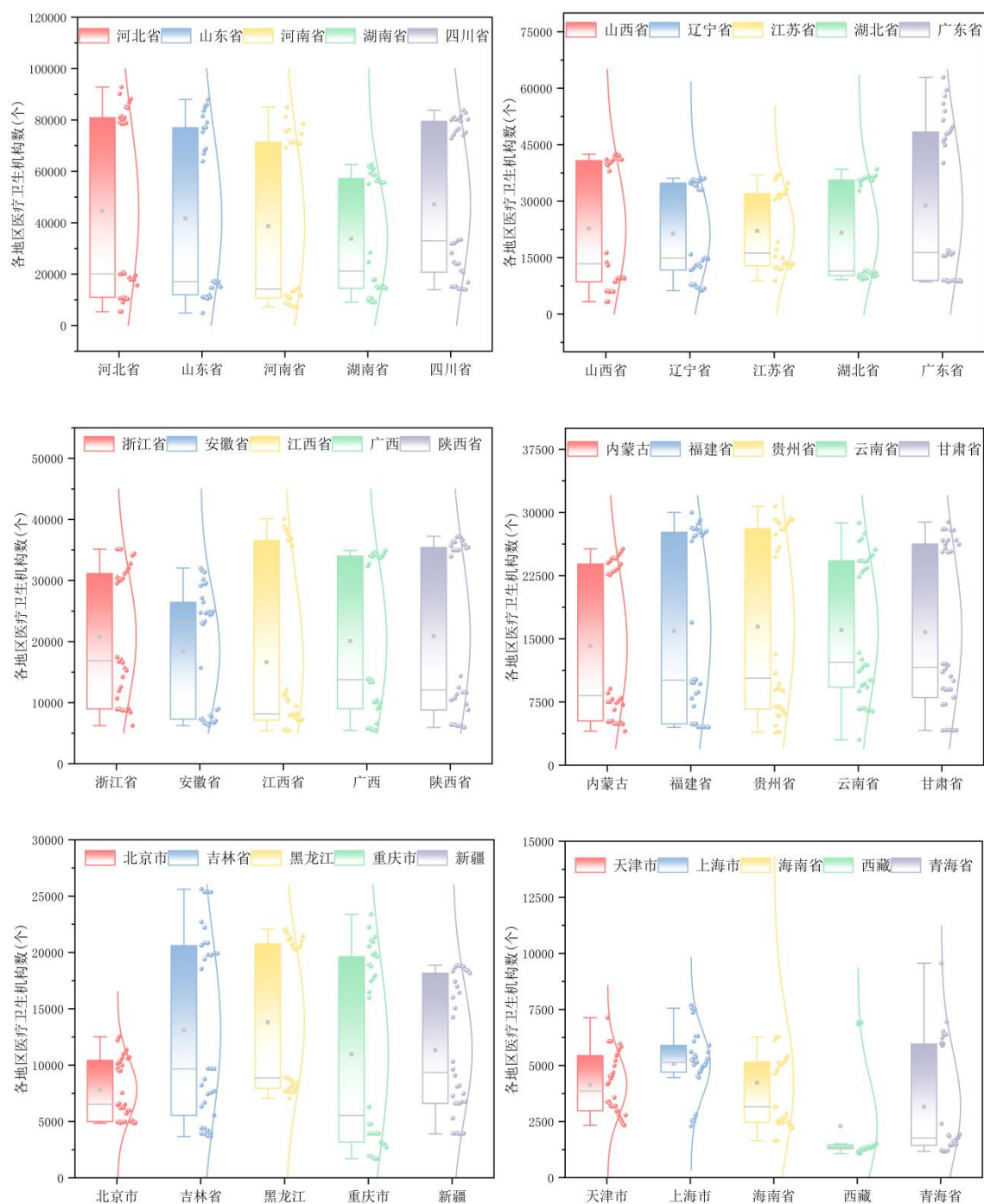


图 17 全国各地区医疗卫生机构数箱线图

通过绘制上述的箱线图，发现，海南省、西藏存在部分异常值。

处理办法：使用异常值附件 4 年内的数据的均值进行填充。

考虑到不同地区医疗卫生机构数量的量级差异，我们进行了数据标准化处理，以便于跨地区比较。为了降低数据中的随机波动，我们对数据进行了平滑处理，采用移动平均法来提高趋势的清晰度。

为了更好地识别医疗卫生机构数量随时间变化的趋势，我们对数据进行了平滑处理。采用指数平滑（Exponential Smoothing, ES）方法，通过给最近的数据点赋予更大的权重来反映最新的趋势。下图展示了原始数据（以北京市为例）与平滑后数据的对比。

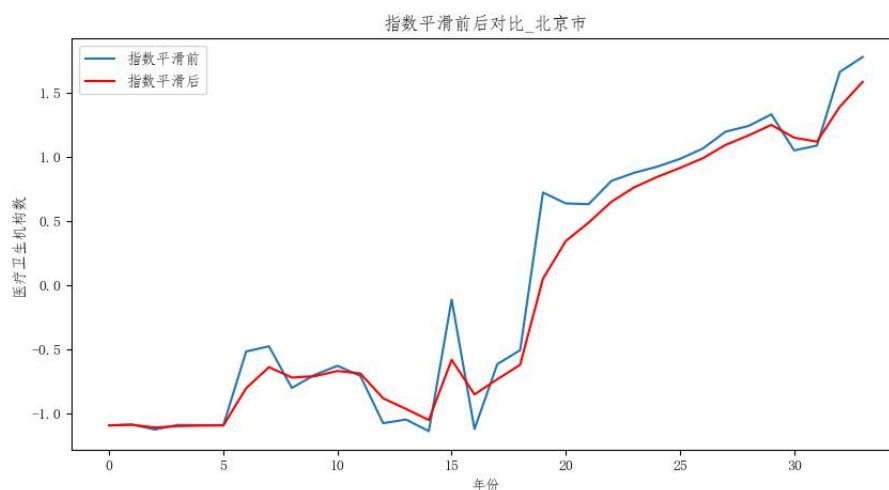


图 18 指数平滑前后对比折线图_北京市

原始数据（蓝色曲线）：表示未经处理的数据点，可以看到明显的随机波动。

平滑后的数据（红色曲线）：表示应用指数平滑后的数据点，波动减少，趋势更加明显。

通过这种方法，我们消除了数据中的短期波动，使长期趋势更加清晰可见。这有助于我们更好地理解 and 解释医疗卫生机构数量随时间的变化规律。

表 14 对时间序列数据进行指数平滑

伪代码：指数平滑数据并保存至 Excel 文件

```
// 1. 加载数据
// 假设数据存储在一个 Excel 文件中
LOAD_DATA_FROM_EXCEL("latest_data_path.xlsx", data)

// 2. 设置平滑因子 alpha
SET_SMOOTHING_FACTOR(alpha, 0.5)

// 3. 对数据框中的每列数据应用指数平滑
FOR EACH_COLUMN in data:
    // 应用指数平滑
    APPLY_EXPONENTIAL_SMOOTHING(data[each_column], alpha)

// 4. 将平滑后的数据保存到新的 Excel 文件
SAVE_SMOOTHED_DATA_TO_EXCEL(smoothed_data, "smoothed_data.xlsx")

// 输出完成信息
OUTPUT("指数平滑后的数据已保存")
```

5.3.2 模型的建立

5.3.2.1 时间序列模型

1). ADF 检验、自相关和偏自相关函数

ADF (Augmented Dickey-Fuller) 检验, ADF 检验是一种用于判断时间序列是否平稳的单位根检验方法。它的目的是检测时间序列是否存在单位根, 即序列是否表现出非平稳性特征。时间序列具有单位根, 即序列是非平稳的。ADF 检验通过构建一个包含时间序列的滞后项和差分项的回归模型来测试单位根的存在。检验的统计量是回归模型的最小二乘估计值的 t 统计量。如果 p 值小于显著性水平 (通常是 0.05), 则拒绝原假设, 认为时间序列是平稳的; 反之, 则接受原假设, 表明时间序列非平稳。

自相关函数 (ACF), ACF 表示时间序列数据在不同滞后阶数下的自相关性。ACF 图显示了不同滞后阶数下的自相关系数。如果 ACF 图中的系数迅速衰减至零, 则表示序列可能不具有显著的自相关性; 若出现明显的周期性或逐渐衰减的趋势, 则表明存在自相关。ACF 图有助于识别序列中可能存在的自相关模式, 指导模型选择。

偏自相关函数 (PACF), PACF 在排除了其他滞后阶数影响的情况下, 计算给定滞后阶数与当前值之间的关联度。PACF 图显示了去除其他滞后影响后, 特定滞后阶数下的偏自相关系数。如果在某些滞后阶数上出现显著的非零系数, 说明存在偏自相关性。PACF 图与 ACF 图结合使用, 能更准确地识别序列的自回归特征, 辅助确定 ARIMA 模型的 AR(p) 和 MA(q) 部分。

2). ARIMA 模型

自回归模型 (AR), AR(p) 模型表示当前值是前 p 个值的线性组合加上随机误差。自回归模型的表达式如下:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

其中 ϵ_t 是均值为零、方差为 σ^2 的白噪声序列。

移动平均模型 (MA), MA(q) 模型的当前值是最近 q 个随机误差的线性组合。移动平均模型的表达式如下:

$$y_t = \theta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

自回归移动平均模型 (ARMA), ARMA(p,q) 模型结合了 AR 和 MA 的特性, 同时考虑序列的自回归和移动平均特征。自回归移动平均模型的表达式如下:

$$(1 - \sum_{i=1}^p \phi_i L^i) y_t = \theta_0 + (1 + \sum_{j=1}^q \theta_j L^j) \epsilon_t$$

其中 L 是滞后算子。

自回归积分移动平均模型 (ARIMA), ARIMA(p,d,q) 模型是 ARMA 的扩展, 其中 d 表示差分次数, 用以使序列达到平稳。

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

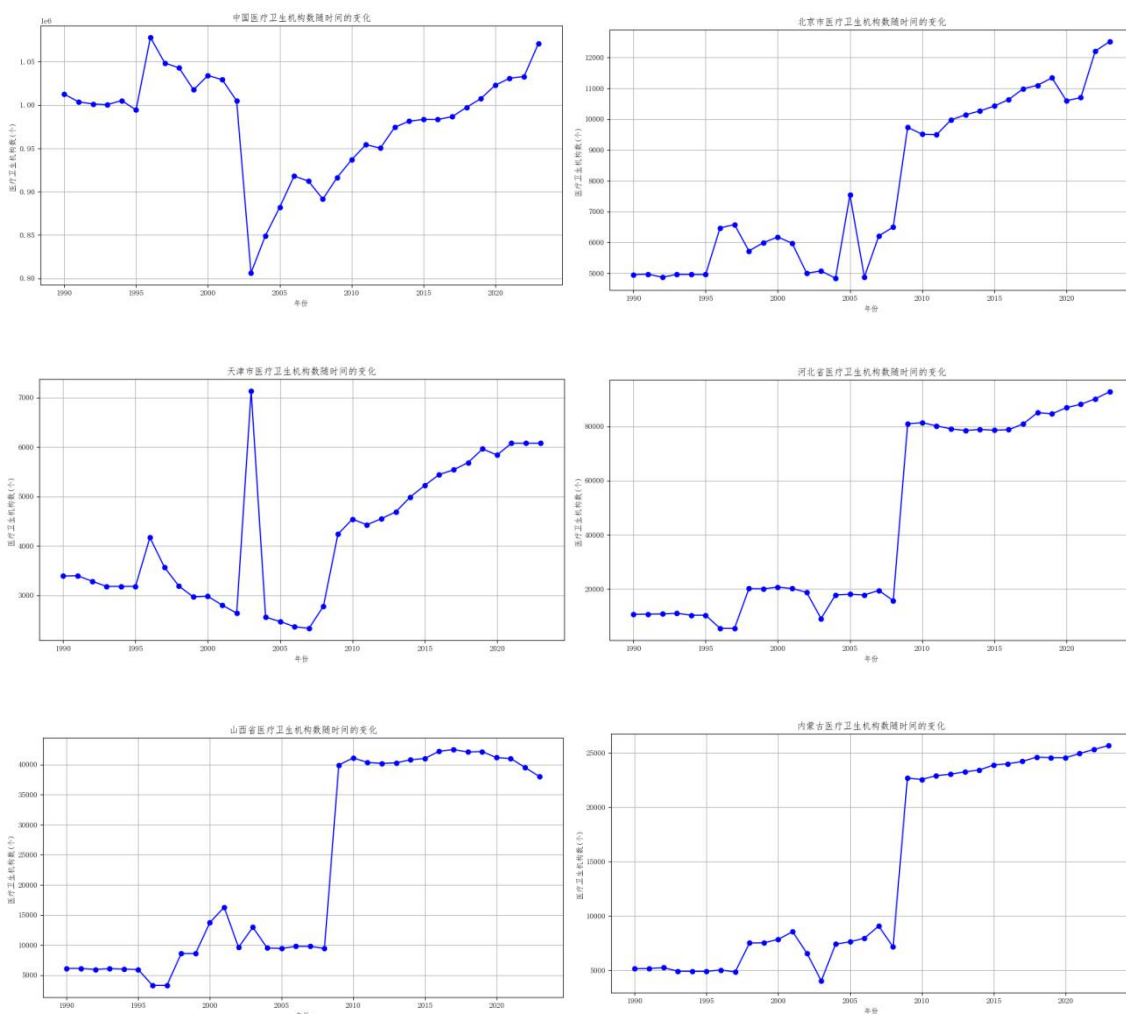
$$y_t = (1-L)^d y_t$$

平稳性, ARMA 平稳性: 平稳性主要由模型的自回归部分决定。对于 ARMA 模型, 平稳性条件要求自回归特征多项式的根都在单位圆外。

5.3.3 模型的求解

5.3.3.1 ARIMA 模型

平稳性检测: 首先需要通过绘图确认时间序列的平稳性。若序列非平稳, 需要进行差分处理直至变为平稳。使用 Python 编程, 绘制出各省市的医疗卫生机构数的时间序列图, 如下所示 (顺序参照附件一):



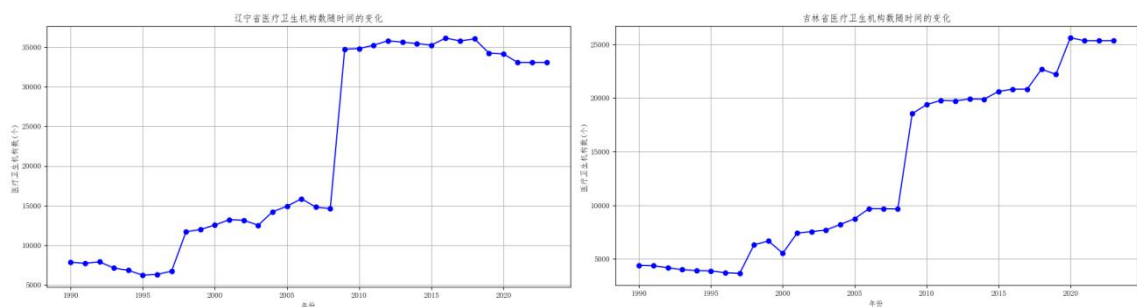


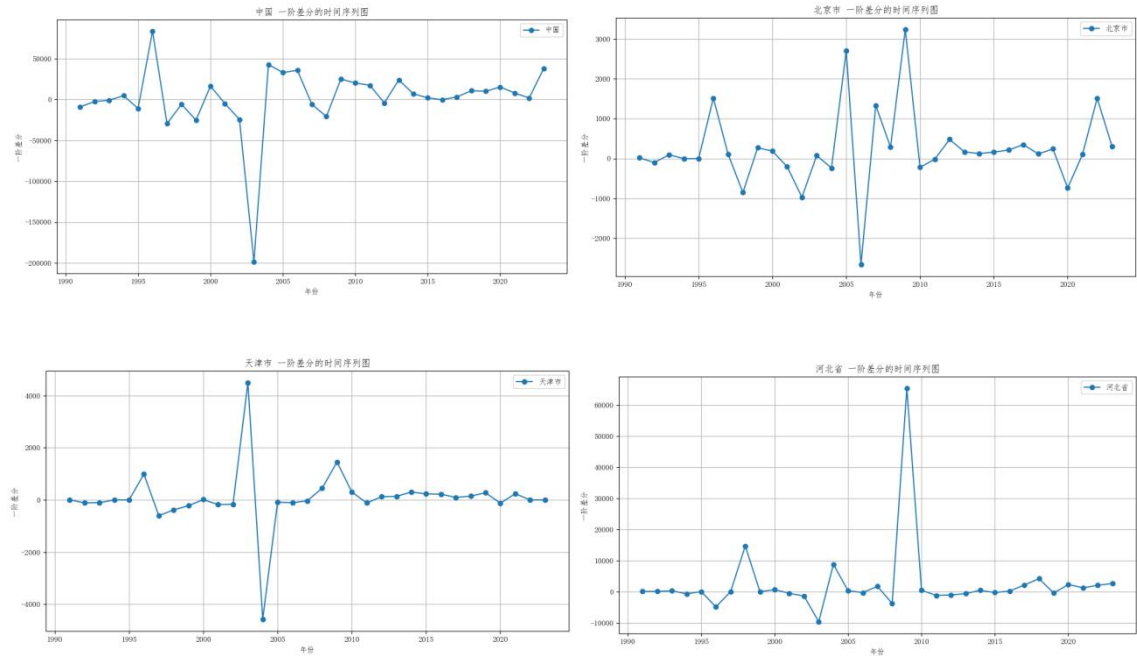
图 19 部分省市的医疗卫生机构数的时间序列图

总体趋势：从 1990 年至 2023 年，中国的医疗卫生机构数量整体呈现上升趋势。这一增长反映了中国经济和社会发展的成果，以及政府对医疗卫生事业的重视和支持。

地区差异：不同地区的增长速度存在明显差异。经济较发达的地区（如北京、上海、广东等）医疗卫生机构数量的增长较快，而一些经济相对落后的地区（如西部省份）增长速度较慢。

年均增长率：通过计算年均增长率，我们可以识别出增长最快的地区和增长较慢的地区。东部沿海城市和省份由于经济发展水平较高，医疗卫生投入较大，其医疗卫生机构数量的增长速度较快；而一些中西部地区则因经济发展相对滞后，医疗卫生机构数量的增长速度较慢。

通过时间序列图分析，发现序列非平稳需要继续对数据进行处理，使用一阶逐期差分处理，处理结果，如下图所示。



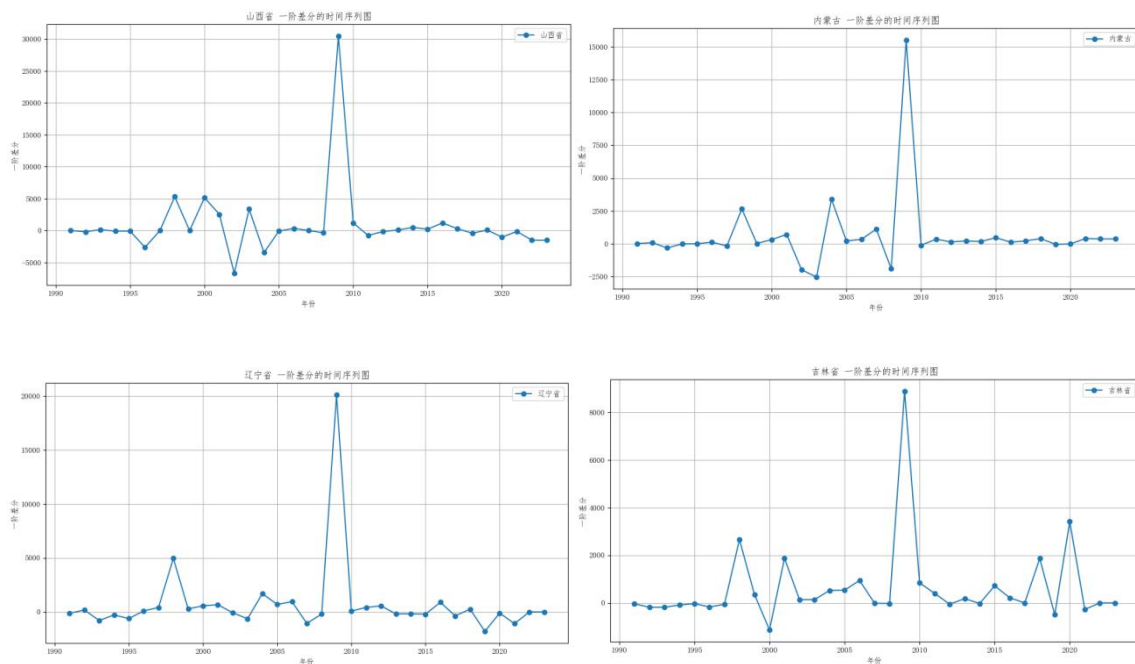
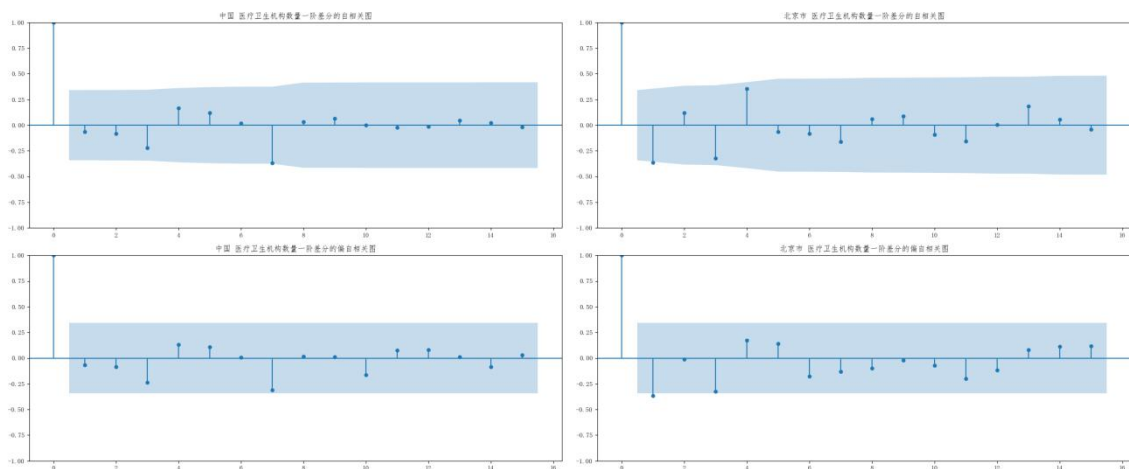
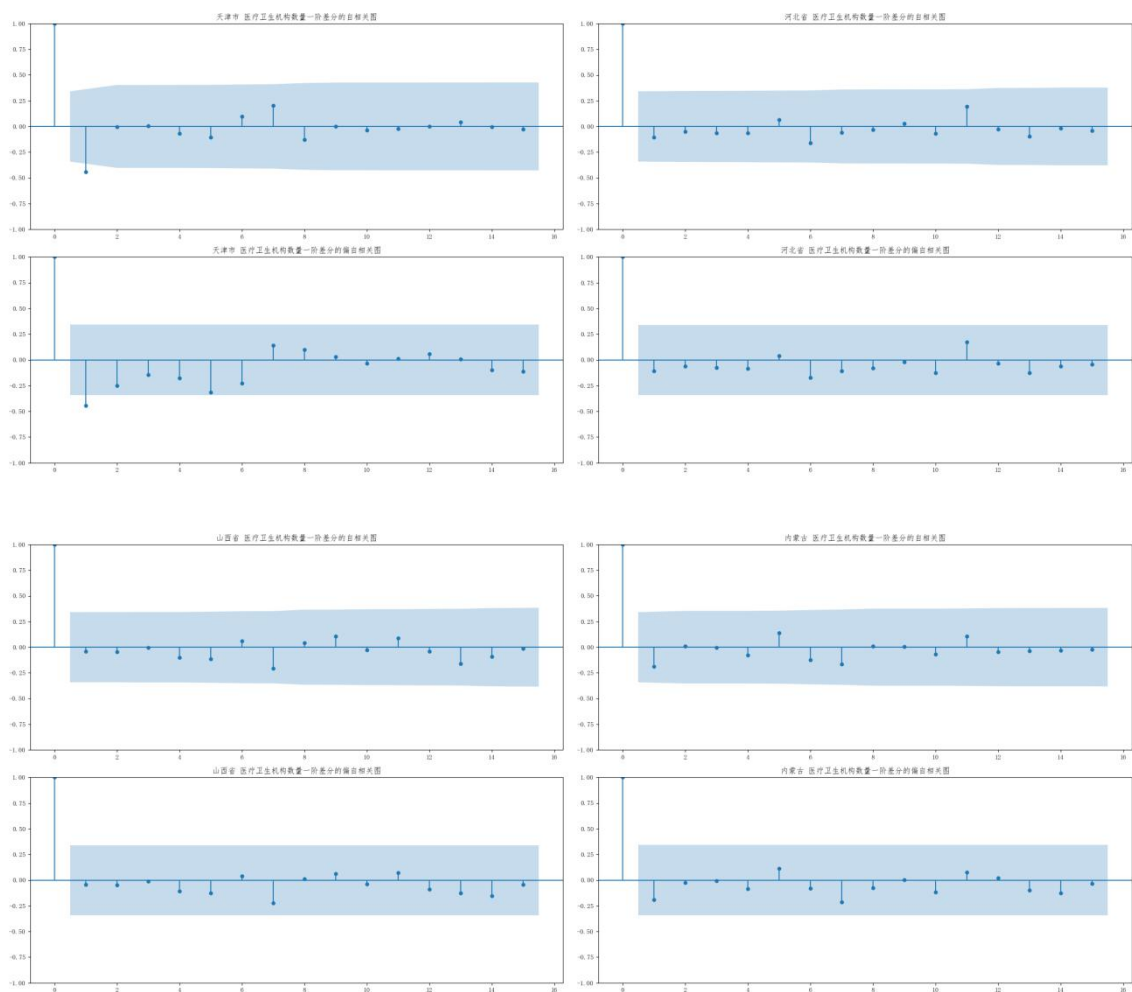


图 20 部分省市的医疗卫生机构数的一阶差分序列图

从一阶差分时间序列图可以看出，大部分地区的医疗卫生机构数呈现出波动增长的趋势，但增长速度并不稳定。其中，经济发达地区（如北京、上海等）表现出较高的增长速度，而一些经济欠发达地区（如山西省）的增长则相对缓慢。值得注意的是，在 2003 年和 2020 年，许多地区的医疗卫生机构数出现了一次性大幅增长，这可能与 SARS 和 COVID-19 疫情相关。在实施了医疗改革政策之后，部分地区的医疗卫生机构数增长加速，表明政策发挥了积极作用。总体而言，医疗卫生机构数的一阶差分时间序列图揭示了中国各地区医疗卫生资源发展的不均衡性，以及政策、经济和社会事件对其产生的影响。”

经过一阶逐期差分结果，可见，一阶逐期差分处理之后的各省市医疗卫生机构数的一阶差分序列图大部分趋于平稳，接着绘制自相关函数（ACF）和偏自相关函数（PACF）以确定 ARIMA 模型的参数。





通过上述流程，可以确定，各个 ARIMA 模型的参数，如下表所示：

表 15 各地区未来 5 年的医疗卫生机构数量

全国及各省市	ARIMA 模型参数	2024 年预测结果	2025 年预测结果	2026 年预测结果	2027 年预测结果	2028 年预测结果
中国	ARIMA(2,0,0)	1051232	1045153	1036407	1027606	1019820
北京市	ARIMA(0,1,0)	12221	12434	12648	12862	13076
天津市	ARIMA(0,1,0)	6041	6041	6041	6041	6041
河北省	ARIMA(1,1,0)	91553	91936	92096	92163	92191
山西省	ARIMA(1,1,0)	38668	38429	38323	38276	38255
内蒙古	ARIMA(1,1,0)	25868	26432	27008	27588	28169
辽宁省	ARIMA(1,1,0)	33167	33136	33123	33118	33116
吉林省	ARIMA(1,1,0)	25613	26163	26739	27324	27911
黑龙江	ARIMA(1,1,0)	21193	21225	21235	21238	21239
上海市	ARIMA(1,0,0)	6191	6161	6133	6107	6083
江苏省	ARIMA(0,1,0)	37383	38119	38854	39590	40326
浙江省	ARIMA(1,1,0)	35478	36106	36763	37429	38096
安徽省	ARIMA(0,1,0)	31052	31749	32447	33144	33842
福建省	ARIMA(1,1,0)	30149	30903	31670	32432	33196
江西省	ARIMA(1,1,0)	39035	39423	39596	39673	39707
山东省	ARIMA(1,1,0)	88498	90561	92691	94846	97011
河南省	ARIMA(1,1,0)	83232	83897	84195	84330	84390
湖北省	ARIMA(1,1,0)	37874	38036	38101	38127	38137
湖南省	ARIMA(1,1,0)	55754	55721	55706	55699	55695
广东省	ARIMA(0,1,0)	61884	63440	64996	66552	68107
广西	ARIMA(1,1,0)	34731	34809	34846	34863	34872
海南省	ARIMA(0,1,0)	6208	6208	6208	6208	6208
重庆市	ARIMA(1,1,0)	23109	23511	23784	23969	24094
四川省	ARIMA(1,1,0)	75865	75636	75540	75500	75483
贵州省	ARIMA(1,1,0)	30189	30300	30340	30355	30360
云南省	ARIMA(0,1,1)	28515	29070	29625	30181	30736

全国及各省市	ARIMA 模型参数	2024 年预测结果	2025 年预测结果	2026 年预测结果	2027 年预测结果	2028 年预测结果
西藏	ARIMA(1,1,0)	6866	6887	6898	6903	6906
陕西省	ARIMA(1,1,0)	34993	34985	34982	34980	34979
甘肃省	ARIMA(0,1,0)	26207	26856	27505	28154	28804
青海省	ARIMA(0,1,0)	6539	6539	6539	6539	6539
宁夏	ARIMA(0,1,0)	4714	4714	4714	4714	4714
新疆	ARIMA(1,1,0)	16404	16413	16417	16419	16420

使用 ARIMA 时间序列预测模型，预测后，得到如下模型性能诊断表：

表 16 ARIMA 模型诊断表_北京市

ARIMA 模型诊断表			
Ljung-Box (L1) (Q)	0.37	Jarque-Bera (JB)	14.57
Prob(Q)	0.55	Prob(JB)	0.00
Heteroskedasticity (H)	0.75	Skew	1.11
Prob(H) (two-sided)	0.65	Kurtosis	5.38

Ljung-Box (L1) (Q)检验用于检查模型的残差是否独立。这里，Ljung-Box 统计量为 0.37，其对应的 P 值（Prob(Q)）为 0.55。P 值大于常见阈值（如 0.05），表明无法拒绝残差独立的原假设，即残差间无显著自相关性，这通常被视为模型诊断中一个积极的信号。

Jarque-Bera (JB)检验用于检验模型残差是否符合正态分布。JB 统计量为 14.57，其 P 值（Prob(JB)）为 0.00，即小于 0.05，这表明残差的分布显著偏离正态分布。在实际应用中，正态分布的残差是许多统计模型的假设之一，因此，该结果可能暗示模型需要调整或改进，比如通过数据转换或考虑其他模型。

Heteroskedasticity (H)检验用于检验模型残差的方差是否随时间变化，即检查残差的异方差性。H 统计量为 0.75，P 值（Prob(H)）为 0.65，大于 0.05，表明没有足够的证据拒绝残差方差恒定的原假设，即残差不存在显著的异方差性，这是另一个正面的诊断结果。

Skew 和 Kurtosis: Skew（偏度）为 1.11，表示残差的分布向右偏斜（正偏斜），即分布的右侧尾部比左侧长；Kurtosis（峰度）为 5.38，高于正态分布的峰度值 3，表明模型残差的分布比正态分布更尖峰，尾部更重。这两个统计量与 JB 检验相呼应，共同表明残差的分布形态与正态分布有显著差异。

综上所述，以北京市为例的 ARIMA 时间序列预测模型，效果良好。

5.4 问题 4 医疗卫生机构数量的宏观预测及其政策启示

灰色预测模型特别适合医疗卫生机构数量的预测任务，尤其是在数据有限的情况下，能够提供有价值的预测结果，帮助进行资源的合理配置和规划。

问题 4 的解题流程如下图所示：

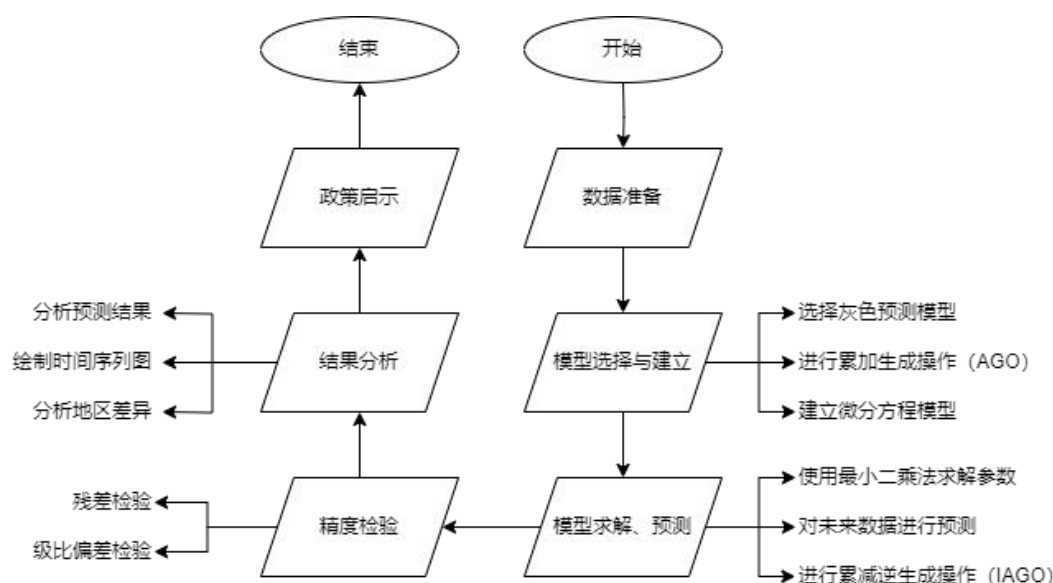


图 21 问题 4 的解题流程图

5.4.1 数据的清理与整理

前三问已经进行了相关数据的清理与整理，在问题 4 中，就不再进行。

5.4.2 模型的建立

灰色系统理论，由邓聚龙教授于 1982 年创立，其核心在于处理部分信息已知、部分信息未知的系统问题，即所谓的“灰色”信息。这一理论自提出以来，不仅在国内，而且在国际学术界引起了广泛关注，其应用领域从社会经济扩展到科技、教育、环境等多个维度，成为分析与决策的重要工具。

灰色预测模型（Gray Forecast Model）作为该理论的核心应用之一，其独特优势在于对数据量要求不高，即使在数据序列短、信息不充分的情况下，也能构建出具有高度规律性的模型。这一特性，使得灰色预测模型在处理现实世界中普遍存在的不确定性和复杂性问题时，展现出显著的优越性。

该模型的构建过程包含数据生成处理与模型建立两大步骤。首先，对原始数据进行生成处理，通过累加生成操作（AGO）增强数据序列的平滑性和规律性，为模型构建奠定基础。随后，基于生成数据，利用最小二乘法等方法建立微分方程模型，该模型能够捕捉数据变化的趋势和规律。最后，通过累减逆生成操作（IAGO）将模型预测值还原至原始数据尺度，得到最终的预测结果。

灰色预测模型的这一系列步骤，不仅简化了建模过程，降低了对大量历史数据的依赖，而且通过数学方法有效增强了预测的准确性和可靠性。因此，该模型在诸多领域，特别是对历史数据有限或信息不完全的系统分析中，展现出了独特的价值和广泛的应用前景。

5.4.2.1 灰色预测模型

1). 设 $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n),\}$ 是最初的非负数据列, 对其进行累加处理, 可以得到新的数据列 $x^{(1)}$ 。

用指数曲线的表达式来逼近序列 $x^{(1)}$ 相应可以构建一阶常微分方程来求解拟合指数曲线的函数表达式:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

想求出 $x^{(1)}$ 的表达式, 就需要解出常微分方程, 所以要先知道参数 a 和 u 。

则微分方程变为:

$$x^{(0)}(t) + ax^{(1)}(t) = u$$

上式为常见的一元线性方程, 为了消除数据随机性, 定义:

$$z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n),\};$$

其中:

$$z^{(1)}(m) = \delta x^{(1)}(m) + (1 - \delta)x^{(1)}(m-1), m = 2, 3, \dots, n;$$

且 $\delta = 0.5$, 即前后时刻的均值。

则微分方程改为:

$$x^{(0)}(t) = -az^{(1)}(t) + u;$$

2). 我们已知 $x^{(0)}(t)$, $z^{(1)}(t)$ 的数据, 结合线性规划得到知识, 可以利用线性规划或者用最小二乘法求解参数。

用矩阵表示微分方程 $Y = X\beta + \varepsilon$ 。

用最小二乘法求解, 最小二乘法即使因变量的观察值与估计值之间的离差平方和达到最小来求解, 即 $\min Q = \sum (x^{(0)} - \hat{x}^{(0)})^2$ 。

解得 \hat{a} 和 \hat{u} ，带入微分方程，并求解，可以得到：

$$\hat{x}^{(1)}(m+1) = \left[x^{(0)}(1) - \frac{b}{\hat{a}} \right] e^{-\hat{a}m} + \frac{b}{\hat{a}}, m = 1, 2, \dots, n-1;$$

当 m 取 0, 1,, 33 时，得到的 $\hat{x}^{(0)}$ 为拟合值，当 $m > 33$ 时，得到的是预测值。

5.4.2.2 精度检验

灰色预测模型的精度检验一般有两种方法，分别是残差检验和级比偏差检验。

1). 残差检验

绝对残差： $\epsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 2, 3, \dots, n$

相对残差： $\epsilon_r(k) = \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} \times 100\%, k = 2, 3, \dots, n$

平均相对残差： $\epsilon_r = \frac{1}{n-1} \sum_{k=2}^n |\epsilon_r(k)|$

如果 $\epsilon_r < 20\%$ 则认为 GM(1, 1) 对原数据的拟合达到一般要求。

如果 $\epsilon_r < 10\%$ 则认为 GM(1, 1) 对原数据的拟合效果非常不错。

2). 级比偏差检验

首先计算原始数据级比 $\sigma(k) = \frac{x^{(0)}(k)}{x^{(0)}(k-1)}, k = 2, 3, \dots, n$

再根据预测发展系数 \hat{a} 计算级比偏差和平均级比偏差：

$$\mu(k) = \left| 1 - \frac{1 - 0.5\hat{a}}{1 + 0.5\hat{a}} \frac{1}{\sigma(k)} \right|, \mu = \sum_{k=2}^n \mu(k) / (n-1)$$

如果 $\hat{\mu} < 0.2$ 则认为模型对原数据的拟合达到一般要求。

如果 $\hat{\mu} < 0.1$ 则认为模型对原数据的拟合效果非常不错。

5.4.3 模型的求解

5.4.3.1 灰色预测模型

为了直观展示 1990 年至 2023 年中国医疗卫生机构数量的变化趋势，绘制了时间序列图。

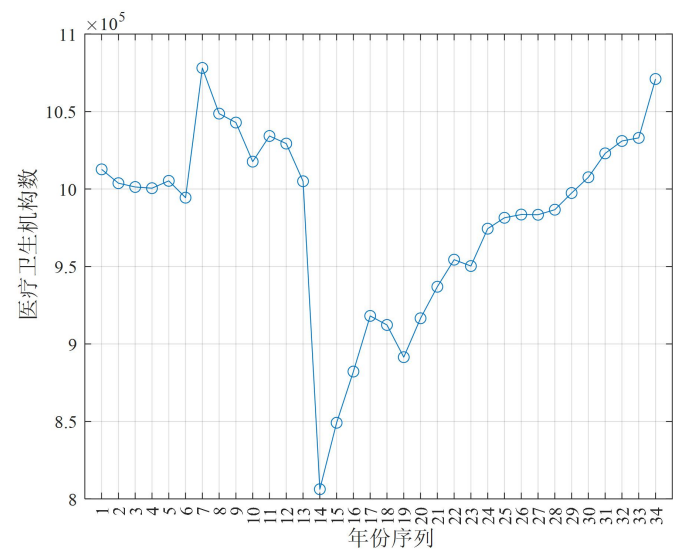


图 22 医疗卫生机构数的时间序列图_中国

从图中可以看出，全国医疗卫生机构数量总体呈现上升趋势，反映了我国医疗卫生事业的发展状况。

再绘制出原始数据的光滑度曲线。如下所示：

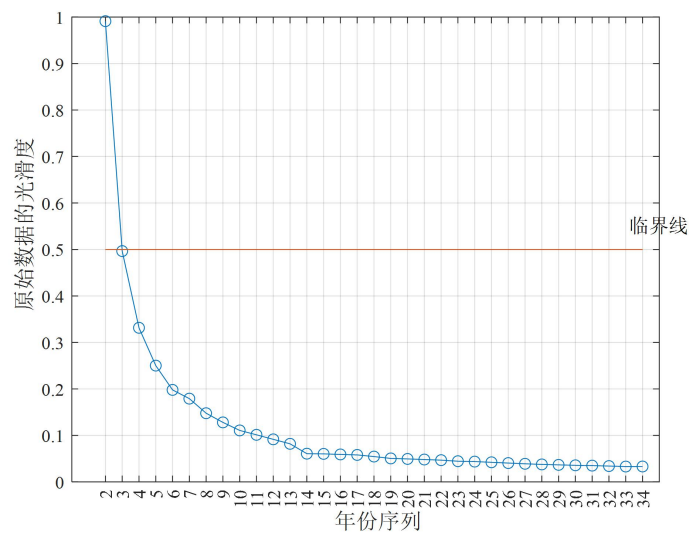


图 23 光滑度曲线

通过平滑处理原始数据，我们获得了光滑度曲线，如图 22 所示。平滑处理有助于减少噪声干扰，突出数据的基本趋势。根据计算，光滑比小于 0.5 的数据占比达到

96.9697%，这表明数据符合准指数增长规律，适合采用指数平滑法进行预测。得到的预测结果，如下表所示：

表 17 医疗卫生机构数预测_中国

年份	预测结果
2024	980193.8836
2025	981809.1454
2026	983588.8873
2027	985645.8624
2028	988379.7063
2029	990810.4171
2030	999299.7605
2031	1007256.5819
2032	1016016.8533
2033	1024305.2139

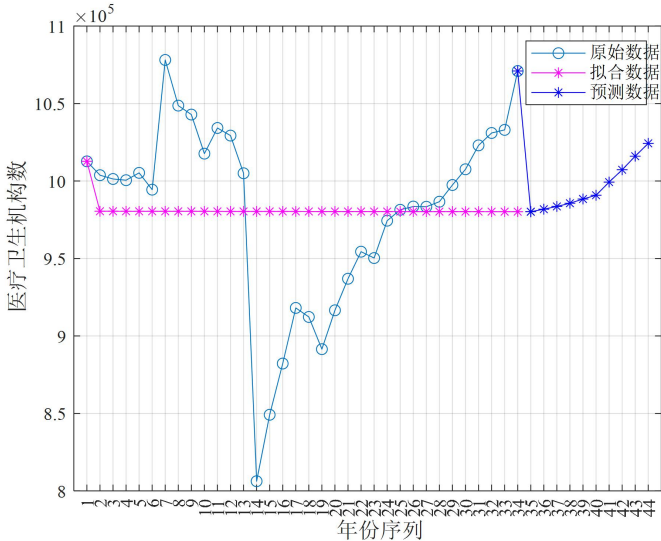


图 24 医疗卫生机构数预测序列图_中国

通过对比，可以看出模型很好地捕捉到了数据的增长趋势，并对未来十年进行了合理的预测。

精确检验
平均相对残差为 0.050577
平均级比偏差为 0.024317

残差检验的结果表明，该模型对原始数据的拟合程度较好；而级比偏差检验的结果同样显示模型对原始数据的拟合效果良好。综合两项检验结果，可以认为该预测模型具有较高的精度和可靠性。

同理，可以对全国各省市的医疗卫生机构数进行预测，预测的结果，如下：

表 18 各地区未来 10 年的医疗卫生机构数量

年份	中国	北京市	天津市	河北省	山西省	内蒙古
2024	980194	13113	6269	141806	62841	35044
2025	981809	13545	6488	151682	66657	37060
2026	983589	13991	6709	162244	70706	39191
2027	985646	14451	6931	173543	75000	41445
2028	988380	14927	7161	185628	79556	43828
2029	990810	15418	7398	198555	84387	46349
2030	999300	15926	7716	212382	89513	49014
2031	1007257	16450	8012	227172	94949	51833
2032	1016017	16991	8293	242992	100716	54814
2033	1024305	17550	8563	259913	106833	57966
年份	山西省	内蒙古	辽宁省	吉林省	黑龙江	上海市
2024	62841	35044	47698	33673	26254	4893
2025	66657	37060	50029	35704	27346	5092
2026	70706	39191	52473	37858	28483	5310
2027	75000	41445	55038	40142	29667	5464
2028	79556	43828	57727	42563	30900	5603
2029	84387	46349	60548	45130	32185	5736
2030	89513	49014	63506	47853	33523	5875
2031	94949	51833	66610	50739	34916	6016
2032	100716	54814	69864	53799	36368	6197
2033	106833	57966	73278	57045	37880	6388
年份	江苏省	浙江省	安徽省	福建省	江西省	山东省
2024	43611	43644	36962	39210	46381	136659
2025	45607	45737	38533	41491	50161	146688
2026	47695	47931	40170	43904	54247	157453
2027	49879	50231	41877	46457	58667	169009
2028	52164	52641	43657	49159	63448	181412
2029	54553	55166	45512	52018	68617	194725
2030	57052	57813	47446	55043	74208	209016
2031	59665	60586	49462	58245	80254	224355
2032	62398	63493	51564	61632	86793	240820
2033	65255	66538	53755	65216	93865	258493
年份	河南省	湖北省	湖南省	广东省	广西	海南省
2024	133153	49322	83179	82846	48270	6654
2025	142776	51977	87755	88488	50909	6842
2026	153095	54775	92582	94515	53691	7036
2027	164160	57724	97675	100952	56626	7235
2028	176024	60831	103048	107827	59721	7439
2029	188746	64105	108717	115170	62985	7649
2030	202388	67556	114697	123014	66428	7866
2031	217015	71193	121007	131391	70059	8088
2032	232699	75025	127663	140339	73889	8316
2033	249518	79064	134686	149897	77927	8551

年份	重庆市	四川省	贵州省	云南省	西藏	陕西省
2024	33035	109511	40872	33573	4769	51430
2025	35351	115290	43260	35195	5293	54258
2026	37829	121375	45786	36896	5800	57241
2027	40480	127780	48461	38679	6289	60389
2028	43317	134524	51292	40549	6830	63709
2029	46354	141624	54288	42509	7427	67212
2030	49603	149098	57458	44563	8108	70907
2031	53079	156967	60815	46717	8864	74806
2032	56800	165251	64367	48975	9692	78919
2033	60781	173972	68127	51342	10579	83258

年份	甘肃省	青海省	宁夏	新疆
2024	36795	6673	6057	23151
2025	38711	7017	6387	24171
2026	40727	7379	6735	25235
2027	42848	7760	7102	26345
2028	45079	8160	7489	27505
2029	47427	8581	7897	28716
2030	49897	9024	8328	29980
2031	52495	9489	8781	31300
2032	55229	9977	9260	32678
2033	58105	10491	9765	34117

增长趋势：中国医疗卫生机构数量在过去 30 多年里呈现稳步增长的趋势。

预测结果：预测结果显示，在未来十年内，全国医疗卫生机构数量将继续保持稳定增长，预计到 2033 年将达到约 102.4 万个。

模型可靠性：通过精确检验，我们验证了所建立的预测模型对原始数据的拟合效果较好，可以作为政策制定和资源配置的重要参考。

地区差异：不同地区的医疗卫生机构数量增长率存在差异，例如东部沿海地区增长较快，而西部地区则相对较慢。

本研究利用时间序列分析方法，成功预测了未来十年中国及其各省市医疗卫生机构数量的增长趋势。这一预测结果不仅可以帮助政府更好地规划医疗卫生资源的配置，还可以为医疗卫生领域的决策提供数据支持。未来的研究可以进一步探索影响医疗卫生机构数量增长的因素，并结合人口结构、经济发展水平等多维度数据进行综合预测，以提高预测的准确性和实用性。

5.4.3.2 政策启示

1.加强医疗卫生资源的均衡布局

依据地区差异进行资源配置，考虑到不同地区的医疗卫生机构数量增长率存在明显差异，政策制定者应当更加注重资源的均衡布局，特别是加大对西部地区和其他欠发达地区的支持力度，缩小区域间差距。优化资源配置机制，通过精准投放资源，加强对医

疗卫生机构的建设和支持，特别是那些增长较慢的地区，以提高整个国家医疗卫生服务水平的整体均衡性。

2.适应人口老龄化趋势

增加老年医疗服务设施，鉴于未来人口老龄化的趋势，需要提前规划和建设更多的老年医疗服务设施，包括但不限于老年医院、康复中心和长期护理机构。培训老年医疗专业人员，加强老年医学人才的培养，提高医护人员的专业技能和服务能力，满足老年人口的特殊医疗需求。

3.提升基层医疗卫生服务能力

加强基层医疗机构建设，扩大基层医疗卫生机构的服务范围和能力，特别是在农村和边远地区，确保基本医疗服务的普及。促进远程医疗服务发展，利用现代信息技术，如互联网医疗和远程诊疗系统，提高基层医疗服务的质量和效率。

4.推动医疗卫生信息化建设

加强数据互联互通，推动医疗卫生数据标准化和互联互通，实现医疗资源的有效共享和高效利用。提升信息化管理水平，利用大数据、人工智能等先进技术，优化医疗卫生资源的管理和调度，提高医疗服务的智能化水平。

5.强化公共卫生体系建设

加强疾病预防控制，加强公共卫生体系建设，提升疾病监测、预警和应急响应能力，预防重大传染病的发生。加大健康宣传教育力度，通过健康教育和宣传，提高公众的健康意识和自我保健能力，减少疾病发生率。

6.鼓励多元化投入机制

引入社会资本，鼓励社会资本投资医疗卫生领域，通过公私合作等方式，共同推动医疗卫生机构的发展。拓宽资金来源渠道，探索多层次、多渠道的资金支持模式，减轻政府财政负担的同时，保证医疗卫生事业的可持续发展。

六、参考文献

- [1]袁淑娟.基于时间序列模型 ARIMA 的校园供水管网暗漏检测研究[J].科学技术创新,2024,(17):94-97.
- [2]张巨峰.基于 ARIMA 模型的甘肃省碳排放预测分析[J].山西大同大学学报(自然科学版),2024,40(04):31-36.
- [3]郑德生,孙涵明,王立远,等.多元时间序列聚类算法综述[J/OL].计算机科学与探索,1-21[2024-08-21].<http://kns.cnki.net/kcms/detail/11.5602.TP.20240724.1140.004.html>.
- [4]樊冬艳,杨灿,孙海,等.基于时间序列相似性与机器学习方法的页岩气井产量预测[J].中国石油大学学报(自然科学版),2024,48(03):119-126.
- [5]盛玮,刘巧玲,刘丽燕.基于因子分析的 6 个大果沙枣新品种果实性状综合评价[J/OL].果树学报,1-15[2024-08-21].<https://doi.org/10.13925/j.cnki.gsxb.20240107>.
- [6]高琳,周剑楠,周小杰,等.基于 ReliefF 和因子分析的管道泄漏源特征识别方法[J].噪声与振动控制,2024,44(04):174-179+241.
- [7]赵爽,袁海燕,郭霆.基于弱化算子的黑龙江省人口老龄化灰色预测[J].黑龙江工程学院学报,2024,38(04):44-49.DOI:10.19352/j.cnki.issn1671-4679.2024.04.007.
- [8]吴俊妮.基于灰色预测模型的大连市快递物流业发展研究[J].黑河学院学报,2024,15(07):58-60+79.
- [9]邱林萍,宋国强,刘梦,等.我国省间基层医疗卫生机构医护资源分布特征研究[J].中国全科医学,2024,27(31):3911-3918.
- [10]林蕾,叶梅毅,李祖平,等.县级区域医学检验中心对基层医疗卫生机构检验能力与质量提升的调查分析[J].中国农村卫生,2024,16(08):36-39.DOI:10.20126/j.cnki.1674-361X.2405-065.

七、附录

支撑材料目录：

- 📁 2024年全国大学生数据分析实践赛
 - ✓ 📁 问题1
 - 📁 各省市年增长率地图热力图
 - 📁 计算年增长率
 - ✓ 📁 问题2
 - 📁 因子分析
 - 📁 多元线性回归
 - 📁 拆分数据
 - 📁 插补其他数据
 - 📁 聚合数据
 - ✓ 📁 问题3
 - 📁 差分序列
 - 📁 时间序列
 - 📁 时间序列预测
 - 📁 自相关、偏自相关图
 - ✓ 📁 问题4
 - 📁 灰色预测代码
 - 📁 灰色预测相关图像