基于ARIMA模型对“五城竞争”进行探究

**摘 要**

**针对问题一**，要求分析宁波、青岛、无锡、长沙、郑州这五座城市的 GDP 发展状况，预测未来五年的 GDP 变化。为此,本文采用ARIMA模型（自回归积分滑动平均模型）来预测上述五城未来五年的GDP变化。先建立ARIMA模型公式需要自回归（AR），差分（I），和滑动平均（MA）这三个部分构成公式（4），将表1的数据带入，得到结果。其结果就是述五城未来五年的GDP变化。

**针对问题二**，要求通过宁波、青岛、无锡、长沙、郑州这五座城市的总人口、城镇居民人均可支配收入、全社会用电量等社会指标，分析影响城市 GDP 的主要因素。本文通过访问相关网站,获取到了每座城市的各项社会指标数据。然后本文采用Pearson相关系数（皮尔逊相关系数）来分析上述六种因素的数据与GDP的相关性。先是建立Pearson相关系数即公式（5）,将各个城市的数据带入计算,得到各个城市与生产总值相关性最大的因素。

**针对问题三**，要求结合问题一、问题二的结果，预测宁波、青岛、无锡、长沙、郑州这五座城市未来五年的 GDP。基于在问题一、问题二中找到的数据和预测上述五座城市未来五年的GDP，为此，本文采用问题一中的ARIMA模型（自回归积分滑动平均模型）预测未来五年的GDP变化，并绘制成每个城市的实际GPD和预测GPD折线变化图，再根据问题二的Pearson相关系数和采用多元线性回归模型来对预测未来五年的GDP升降进行分析。

**针对问题四**，要求比较分析宁波、青岛、无锡、长沙、郑州这五座城市的经济发展历史、现状、未来，设计综合发展指数并对五城排序。基于问题一、问题二、问题三的结果，绘制出上述五座城市未来五年的GDP和历年GDP的折线图，来综合比较五城的GDP发展,并且对它们进行排序，得到其结果如下：第一名青岛、第二名宁波、第三名长沙、第四名无锡、第五名郑州。

**针对问题五**，要求绍兴，结合 GDP 发展状况，与上述五城相比，分析绍兴有哪些优势与不足。已知上述五城的各个因素的数据，本文将在绍兴官方网站找到相关数据，并了解绍兴各项指标和近几年的政策还有地理位置，分析绍兴和上述五城的区别，得到其优缺点和绍兴市的实际GDP与预测GDP折线变化图，评估绍兴在经济发展中的综合竞争力。

**关键词**：ARIMA模型、Pearson相关系数、多元线性回归模型、社会指标

1. 问题重述
   * 1. 问题背景

面对新冠疫情，我们国家或多或少都投入了大量的人力和物力，因此，如何在新冠疫情后的今天，抓住经济增长恢复的关键期，重新回到以往的经济繁荣，已经逐渐成为了一个时代的热门话题。而宁波，青岛，无锡，长沙，郑州五座城市也正好处于讨论的中心。也许是因为它们的经济总量都很接近，或者是它们在自己的省份里都占据着一定的地位。因此，“五城竞争”的话题也由此提出。而如何完成“五城竞争”，却是一个难题，对此，我们提出了以下几个问题来探讨。

* + 1. 需要解决的问题

**问题一**：分析宁波、青岛、无锡、长沙、郑州这五座城市的 GDP 发展状况，预测未来五年的 GDP 变化。

**问题二**：通过宁波、青岛、无锡、长沙、郑州这五座城市的总人口、城镇居民人均可支配收入、全社会用电量、农民居民人均可支配收入、货运量、社会消费品零售总额，分析影响城市 GDP 的主要因素。

**问题三**：结合问题一、问题二的结果，预测宁波、青岛、无锡、长沙、郑州这五座城市未来五年的 GDP

**问题四**：比较分析宁波、青岛、无锡、长沙、郑州这五座城市的经济发展历史、现状、未来，设计综合发展指数并对五城排序。

**问题五**：绍兴，作为[长江三角洲中心区](https://baike.baidu.com/item/%E9%95%BF%E6%B1%9F%E4%B8%89%E8%A7%92%E6%B4%B2%E4%B8%AD%E5%BF%83%E5%8C%BA/56977773) [2](https://baike.baidu.com/item/%E9%95%BF%E6%B1%9F%E4%B8%89%E8%A7%92%E6%B4%B2%E4%B8%AD%E5%BF%83%E5%8C%BA/56977773)7 城之一，经济比较发达，2021 年 GDP 达到了 6.8 千亿元，排名全国第 36。请结合 GDP 发展状况，与上述五城相比，分析绍兴有哪些优势与不足。

1. 模型假设

1.假设这六座城市在政策相同的情况下进行发展。

2.假设这六座城市的GDP发展不受除题目外其他因素影响。

1. 符号说明

|  |  |
| --- | --- |
| **符号** | **符号定义** |
|  | 常数项 |
|  | 自回归系数 |
|  | 白噪声 |
|  | 滑动平均系数 |

1. 问题分析

随着中国经济的不断发展，越来越多的城市成为了经济发展的主要力量。其中，宁波、青岛、长沙、郑州和无锡等城市已经成为具有重要地位的区域性城市。由于这些城市各自具有独特的经济发展状况，因此本文将对宁波、青岛、长沙、郑州和无锡五座城市的经济发展状况进行比较和分析。

**针对问题一**，要求分析宁波、青岛、无锡、长沙、郑州的 GDP 发展状况和预测未来五年的GDP的变化。因为GDP的发展状况是随着时间的变化而变化，所以就可以通过借助时间序列模型当中的ARIMA模型来对上述五座城市的未来GDP进行预测分析,得到它们的GDP的变化值。

**针对问题二**，要求分析宁波、青岛、无锡、长沙、郑州的 GDP 发展的主要影响因素是哪些。本文将基于总人口、城镇居民人均可支配收入、全社会用电量、农民居民人均可支配收入、货运量、社会消费品零售总额这六种因素进行查找数据和通过Pearson相关系数来对GDP和上述六种因素来绘制Pearson相关系数矩阵的热图。

**针对问题三**,将结合问题一与问题二计算出的数据结果,基于宁波、青岛、无锡、长沙、郑州过去的历史GDP数据,使用ARIMA模型,绘制出每座城市的生产总值(GDP)数据预测折线图。

**针对问题四**,本文将考虑宁波、青岛、无锡、长沙、郑州五座城市的历史GDP发展情况，ARIMA模型所绘制出的未来GDP预测数据折线图，综合比较五城的GDP发展,并且对它们进行排序。

**针对问题五**,本文需要对绍兴的相关经济数据进行收集,并且与上述的五座城市进行经济数据的比较,包括但不限于(城市GDP、GDP增长率、产业结构等)。然后,需要对绍兴的经济状况进行详细的分析，了解最近几年绍兴的经济增长趋势，分析绍兴的主导产业和新兴产业，了解其经济结构的特点和优势。同时,还需要对绍兴的地理位置以及相关政策进行分析,得出绍兴的优势和劣势，最后与上述五城进行综合比较，评估绍兴在经济发展中的综合竞争力。

1. 数据处理
   * 1. 针对问题一的数据处理

本文将通过题目中给的政府网址找到关于历年生产总值的数据表，再分别整理出各城的关于历年生产总值的数据表。然后用python算法去找到年份和生产总值单独生成一份数据表,如表 1所示：

表 1长沙GDP表

|  |  |
| --- | --- |
| 年份 | GDP（亿元） |
| 2000 | 720.85 |
| 2005 | 1589.41 |
| 2010 | 4440.32 |
| 2015 | 8502.6 |
| 2020 | 12142.5 |

可以观察到长沙在2010、2015、2018、2019、2020的GDP的数据一直呈持续上升的态势。

* + 1. 针对问题二的数据处理

基于该问题要找各城的GDP的数据与总人口、城镇居民人均可支配收入、全社会用电量、农民居民人均可支配收入、货运量、社会消费品零售总额这六种因素中哪些相关。本文将通过题目中给的政府网址找到关于历年上述六种因素的数据，分别为上述五城整理出相关数据表，如下表所示：

表 2 宁波市各项社会指标数据表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 2010 | 2015 | 2018 | 2019 | 2020 |
| 生产总值（万元） | 144238 | 227270 | 306661 | 329729 | 339034 |
| 全社会用电量（万千瓦时） | 12577 | 16029 | 21240 | 22135 | 22736 |
| 社会消费品零售总额（万元） | 44320 | 83567 | 108613 | 116947 | 115800 |
| 货运量（万吨） | 83.71 | 115.30 | 168.37 | 187.42 | 196.44 |
| 总人口（万人） | 574.08 | 586.57 | 602.96 | 608.47 | 613.66 |
| 城缜居民人均可支配收入（元） | 30166 | 47852 | 60134 | 64886 | 68008 |
| 农村居民人均可支配收入（元） | 14261 | 26469 | 33633 | 36632 | 39132 |

可以观察到宁波在2010、2015、2018、2019、2020的各因素数据和GDP的数据都呈上升的态势。

1. 模型的建立与求解
   * 1. 针对问题一的模型建立和求解
        1. ARIMA模型

通过查阅相关资料可以得知,GDP的发展状况是随着时间的变化而变化的，而ARIMA模型（自回归积分滑动平均模型）[1]是一种用于时间序列分析和预测的统计模型，因此可以通过建立ARIMA模型，通过对历史的城市GDP数据进行分析,来对未来的城市GDP进行预测。

一般来说,ARIMA模型可以分为自回归（AR），差分（I），和滑动平均（MA）三个部分。下面是关于它们的相关介绍。

自回归部分(AR)表示序列当前值与其前若干时刻值的线性组合。自回归模型的阶数用表示；积分部分（I）表示对序列进行的差分操作次数，使得序列平稳。积分部分的阶数用表示；滑动平均部分(MA)表示序列当前值与其前若干时刻的随机误差项的线性组合。滑动平均模型的阶数用表示。

为了去建立自回归部分(AR)，需要设时间序列的当前值为，常数项为，自回归系数为，白噪声（即随机误差项）为,表示当前时间，公式表示：

（1）

建立积分部分（I），进行差分操作，需要差分运算符和滞后算子，即还有差分次数，公式表示为：

（2）

建立滑动平均部分(MA)，需要常数项，滑动平均系数还有白噪声（即随机误差项），公式表示：

（3）

结合（1）（2）（3）的相关公式就可以得出ARIMA模型公式,如下所示:

（4）

* + - 1. 各城市GDP发展预测

本文通过借助ARIMA模型（附录2）,对宁波、青岛、无锡、长沙、郑州的历史GDP数据进行分析,然后利用相关的算法,计算出每座城市的未来GDP预测情况，再通过python当中的matplotlib库（附录1）绘制出以下五张地区生产总值（GDP）及未来5年预测折线图,如图 1所示：

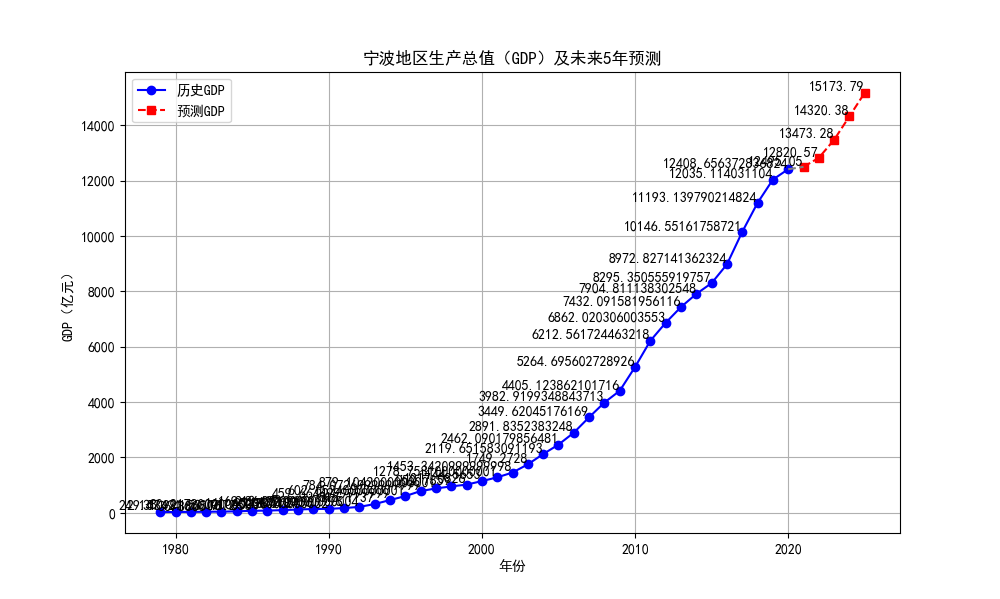


图 1宁波地区生产总值（GDP）及未来5年预测

图1描述了宁波历年GDP和预测GDP的值，X轴为年份，Y轴为GDP。可以观察出宁波在1980年到2025年GDP都在持续增加。到2025年GDP到目前最高值15173.79亿元。

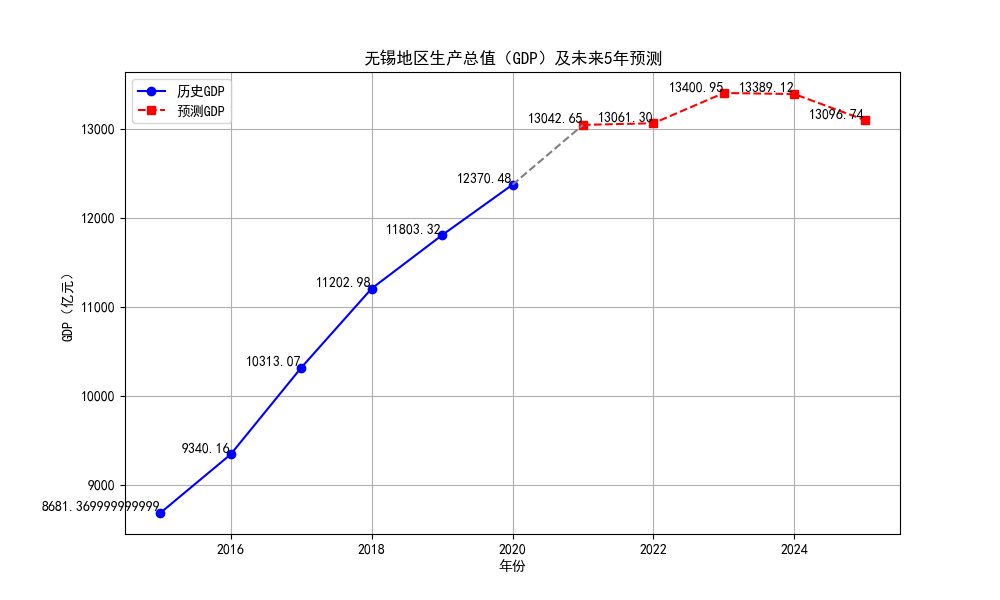


图 2无锡地区生产总值（GDP）及未来5年预测

图2描述了无锡历年GDP和预测GDP的值，X轴为年份，Y轴为GDP。可以观察出无锡在2015年到2020年的GDP都在持续增长。而2020年到2021年，无锡的GDP增长势头变缓。在2022年，GDP仅仅增长19亿元。2023年，预测GDP达到最高值为13400亿元，后续GDP开始降低，2025年预测GDP将比2023低204亿元。

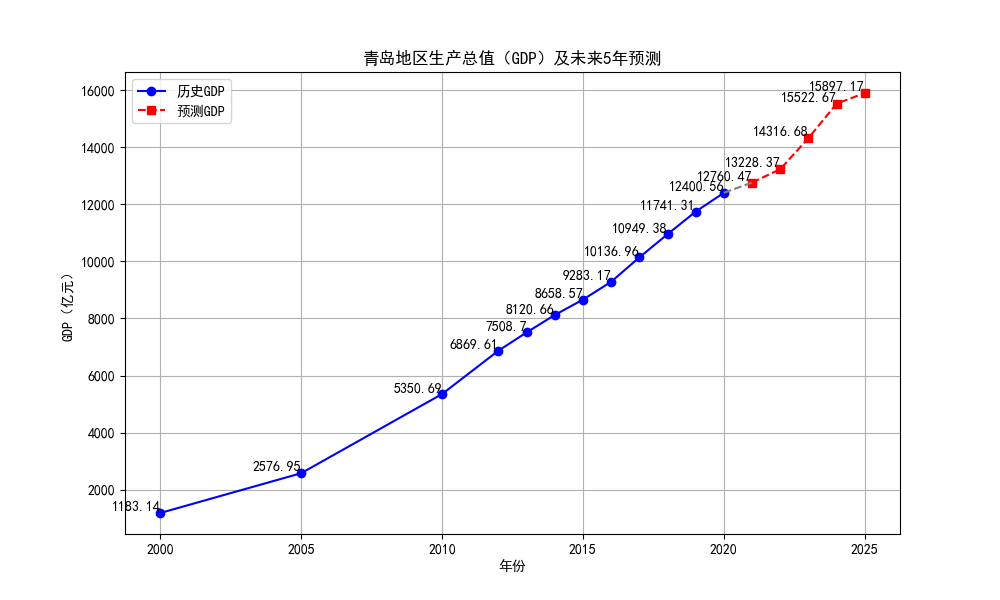


图 3青岛地区生产总值（GDP）及未来5年预测

图3描述了青岛历年GDP和预测GDP的值，X轴为年份，Y轴为GDP。可以观察出青岛的GDP在2000年到2020年在持续增长，从2020年到2022年和2024年到2025，青岛GDP在缓慢增进，整体趋势呈现上升。在2025年，GDP达到现在所预测的最高值15897，17亿元。

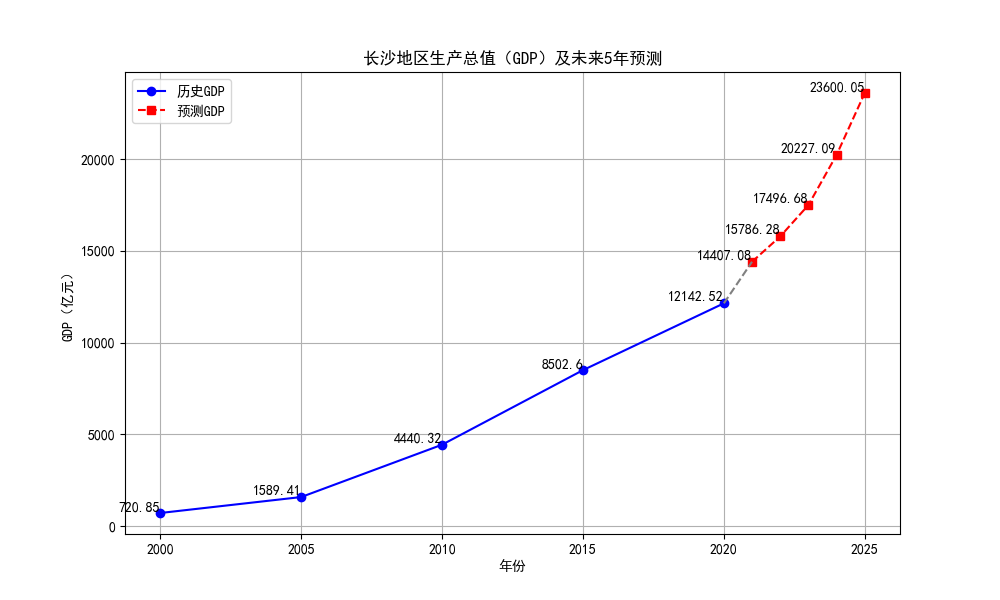


图 4长沙地区生产总值（GDP）及未来5年预测

图4描述了长沙历年GDP和预测GDP的值，X轴为年份，Y轴为GDP。可以清晰地看到长沙地区GDP的增长趋势。从2000年的720.85亿元开始，长沙地区的GDP呈现稳步增长态势。到了2020年，GDP已经增长到接近12142.52亿元的水平。而在未来五年（预测至2025年），GDP预计将继续增长，最终可能达到或超过23600.05亿元。整体趋势上来说，长沙地区的GDP增长趋势呈现出明显的上升态势。

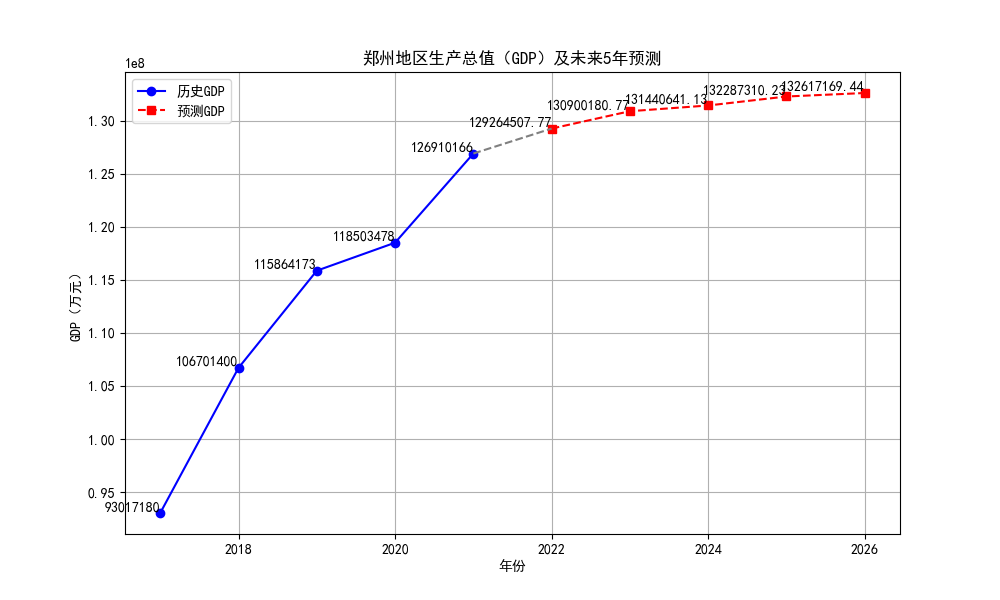


图 5郑州地区生产总值（GDP）及未来5年预测

图5描述了郑州历年GDP和预测GDP的值，X轴为年份，Y轴为GDP。可以观察出郑州地区GDP的增长趋势。从2017年到2021年，郑州GDP呈现出平稳增长的趋势，整体保持上升的态势。对于2022年到2026年，GDP预计将继续保持增长，但增长幅度相当较小。

* + 1. 针对问题二的模型建立和求解
       1. Pearson相关系数矩阵

基于问题二是要求找到影响城市 GDP 的主要因素，而Pearson相关系数（皮尔逊相关系数）[2]是测量两个变量之间线性关系强度和方向的统计量，其值介于 -1 和 1 之间：值1表示两个变量完全正相关，值-1表示两个变量完全负相关，值0表示两个变量之间没有线性相关。Pearson相关系数正好符合该题解法。

建立Pearson相关系数，需要两个变量的观测值和、两个变量的均值和、X和Y之间的Pearson相关系数、观测值的数量，公式表示：

（5）

* + - 1. 问题二Pearson相关系数的求解

为了去分析影响城市GDP的相关因素,本文首先对上述五座城市的总人口、城镇居民人均可支配收入、全社会用电量、固定资产投资总额、货运量、社会消费品零售总额等数据进行了收集,然后借助Pearson相关系数矩阵算法[3]（附录3）,绘制出每座城市的Person相关矩阵系数图,计算出每一年当中各城市影响GDP因素的指标得分以及总得分,如图 6所示:

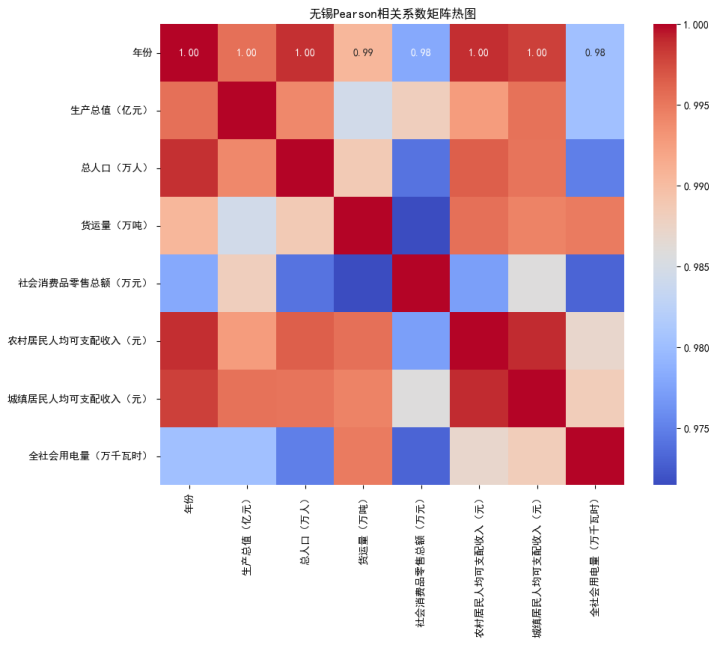


图 6无锡Pearson相关系数矩阵热图

图 6所展示的就是无锡2015年到2020年这6年间各可能影响城市GDP的因素的无锡Pearson相关系数矩阵热图,通过观察可以发现,大部分颜色较深的块位于对角线的两侧，这表示许多变量之间存在较强的正相关性。特别是“生产总值”与“总人口”这两个因素之间的相关性达到了0.995，接近1，表示两者之间存在极强的正相关性,这可能是因为人口增长通常伴随着生产总值的增加。只有当一个城市的人口数量得到增加,这个城市的生产总值才有可能得到一个比较大的增长趋势。不仅如此,“货运量”与“社会消费品零售总额”之间也有较高的相关性，这可能是因为货运量的增加反映了商品流通的活跃，从而促进了社会消费品零售总额的增长。总的来说,无锡地区的多个经济和社会指标之间存在较强的正相关性，这反映了无锡地区经济和社会发展的协调性,同时也反映出了无锡的强大社会经济发展实力。

为了去探索其他四座城市各影响城市GDP的主要因素,本文需要去绘制和无锡Pearson相关系数矩阵热图的其他城市Pearson相关系数矩阵热力图。如图 7所示:

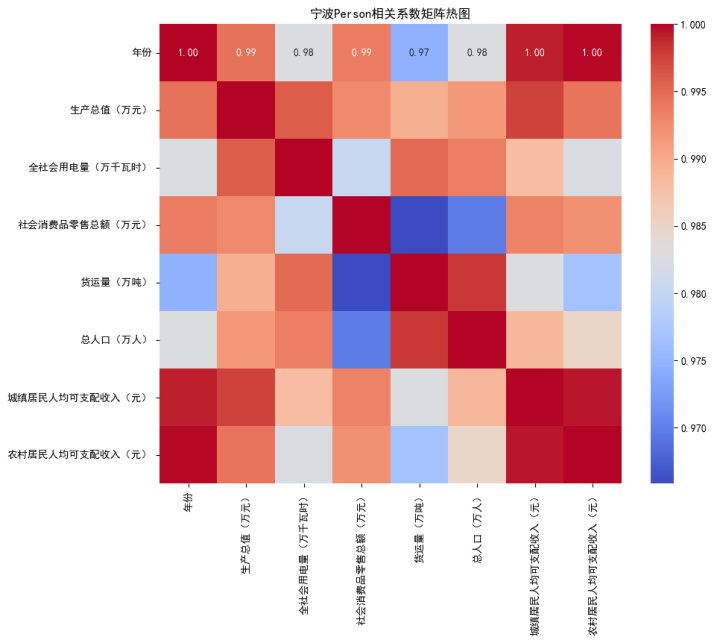


图 7宁波Pearson相关系数矩阵热图

根据上图的宁波Pearson系数相关矩阵热图可以看出,宁波的各个社会指标都与宁波的生产总值有着相关性,其中“全社会用电量（万千瓦时）”与“生产总值（万元）”的相关系数为0.990,“货运量（万吨）”与“生产总值（万元）”的相关系数为0.980,“总人口（万人）”与“生产总值（万元）”的相关系数为0.985。因此看以看出对宁波来说,对生产总值影响最大的不是城市人口数量,而是全社会的用电量,由此可以进行推断,影响宁波的生产总值的最大因素可能是全社会的用电量。

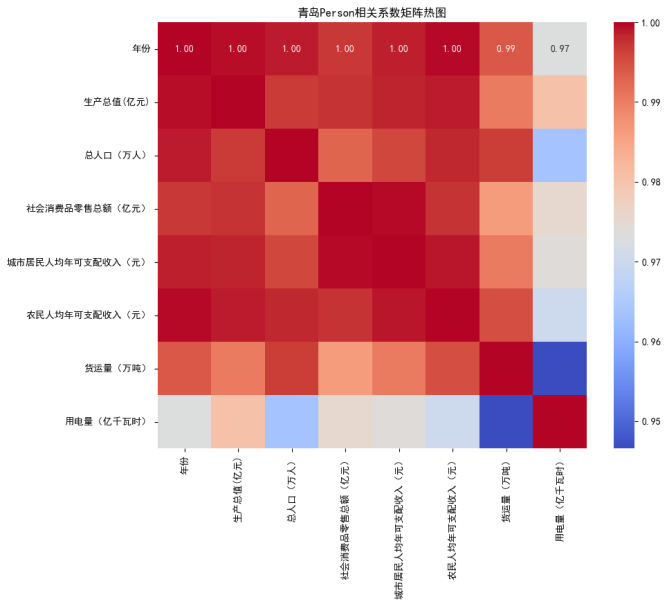


图 8青岛Pearson相关系数矩阵热图

观察上面青岛Pearson系数相关矩阵热图可以得知,青岛的“生产总值（亿元）”与它的“总人口（万人）”之间的相关性为0.99，这远远大于了其他社会指标与生产总值的相关系数,所以这就表示它们两个之间存在很强的正相关性。这通常就意味着人口增长与经济增长之间呈正相关关系,也是在众多社会指标当中,影响生产总值最大的一个因素。

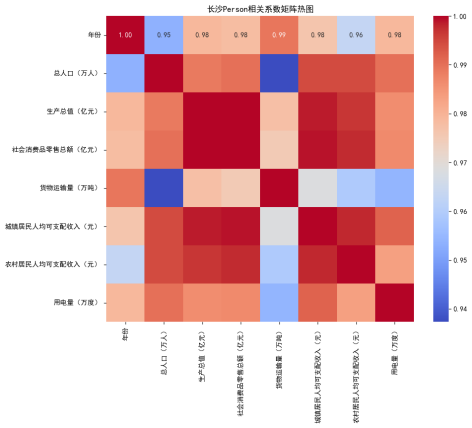
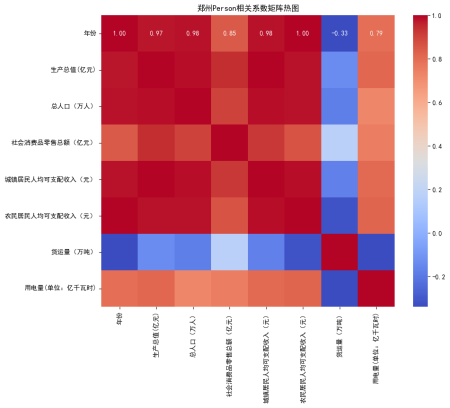
 

图 9长沙Pearson相关系数矩阵热图 图 10郑州Pearson相关系数矩阵热图

通过比对分析上面的长沙Pearson系数相关矩阵热图和郑州Pearson系数相关矩阵矩阵热图,可以发现,长沙的各社会指标与生产总值的关系就像无锡的一样,有着最高相关系数的都是“总人口(万人)”和“生产总值（亿元）”,这就说明了影响长沙生产总值最大的也是人口这一因素。而在郑州的Pearson系数相关矩阵矩阵热图当中,可以看出影响郑州市生产总值最大的依旧是“总人口”这一社会指标,这就说明郑州市的生产总值与人口数量也有着密不可分的关系。

通过Pearson系数相关矩阵,本文得出了无锡、青岛、宁波、长沙、郑州五座城市的Pearson系数相关矩阵热力图,明确了各社会指标与城市生产总值之间的相关性,并且通过对比分析可以得出,除了宁波以外,其他几座城市的生产总值都与它们的人口有着很大的相关性,这说明一座城市的生产总值往往离不开它的人口,城市的总人口数量往往就决定着该城市的生产总值量。

* + 1. 多元线性回归模型预测未来数据变化

本文将根据任务6.2当中计算出的各个城市的社会指标数据,然后结合每个城市的历史GDP数据，利用多元线性回归模型绘制出每个城市的实际GDP与预测GDP变化折线图,对比分析得出每一座城市的GDP变化趋势以及对未来五年每个城市的GDP进行预测。

在绘制实际GDP和预测GDP的时候,需要对6.2当中的各社会指标创建一个新的字典,然后将6.2当中计算的出的数据导入到一个新的数据表当中去,通过拟合多元线性回归模型[4]（附录4）,预测未来五年的自变量数据,从而实现预测未来GDP的目的。然后就需要对预测数据和实际数据进行合并操作,绘制出实际GDP与预测GDP的折线图。具体的折线图如图 11所示:

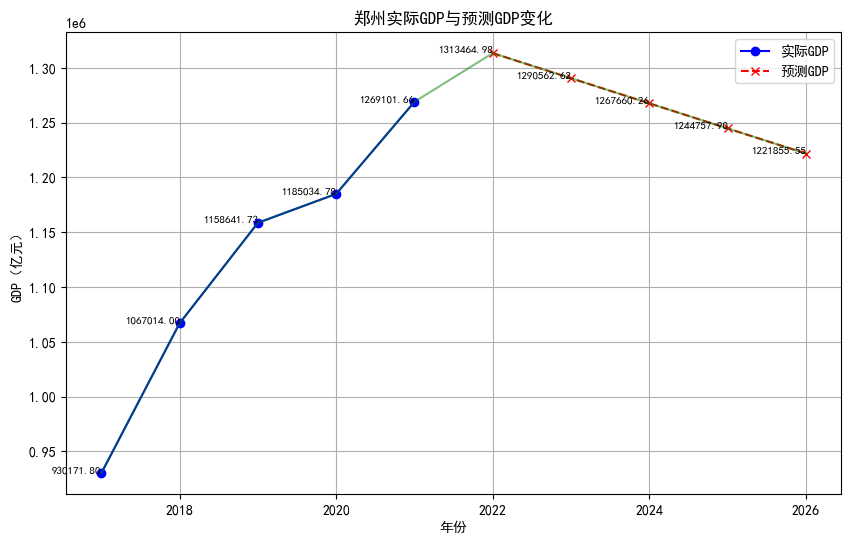


图 11郑州地区实际GDP与预测GPD折线变化图

上图的折线图就是郑州市的实际GDP与预测GDP变化折线图,通过观察可以发现,在各项社会指标的影响下,郑州市的未来五年GDP将会呈现出一个下降的趋势,这也许会和郑州市的人口流失相关,因为通过6.2.1当中的郑州市Pearson系数相关矩阵热图可以得知,郑州市的生产总值收到总人口的影响是最大的,因此在在未来五年当中,郑州市的人口流失就会导致城市GDP呈现一种下降的趋势。

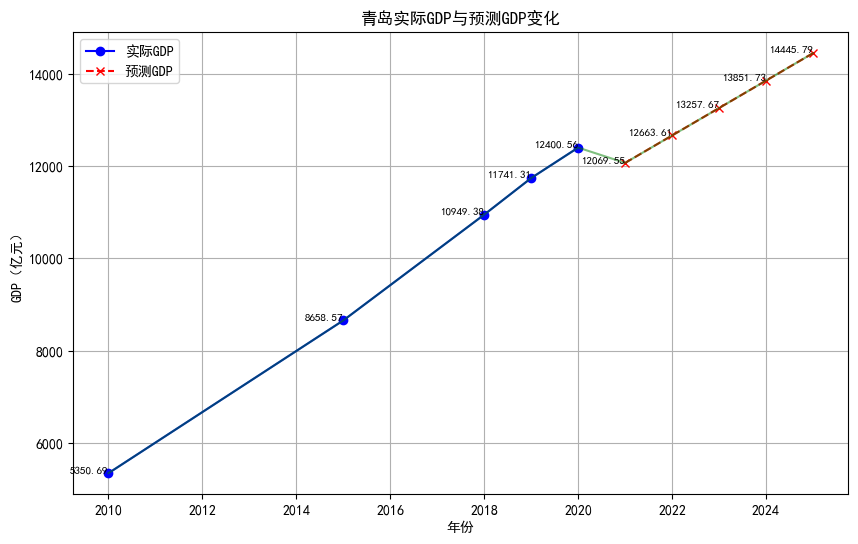


图 12青岛地区实际GPD与预测GPD折线变化图

与郑州市不同的是,青岛市的实际GDP与预测GDP变化折线图总体呈现着一种上升的趋势,无论是青岛过去几年的历史GDP数据还是未来五年的预测GDP数据,青岛市的GDP都是逐年在增加,而在6.2.2的青岛市Pearson系数相关矩阵热图当中也演示了青岛市的生产总值与青岛市的各个社会指标都有着很大的相关性,而青岛作为我们国家的新一线城市,往往会吸引大量的人口,因此在未来的五年间,青岛市的总人口肯定会呈现着一个爆炸式的增长,相对应的,青岛市的GDP也会不断地增加。

和青岛市不一样的是,宁波市的生产总值与全社会用电量的相关系数最大,因此宁波市的实际GDP与预测GDP折线变化图的增长趋势相对比较平缓(相对青岛市而言),如图 13所示:

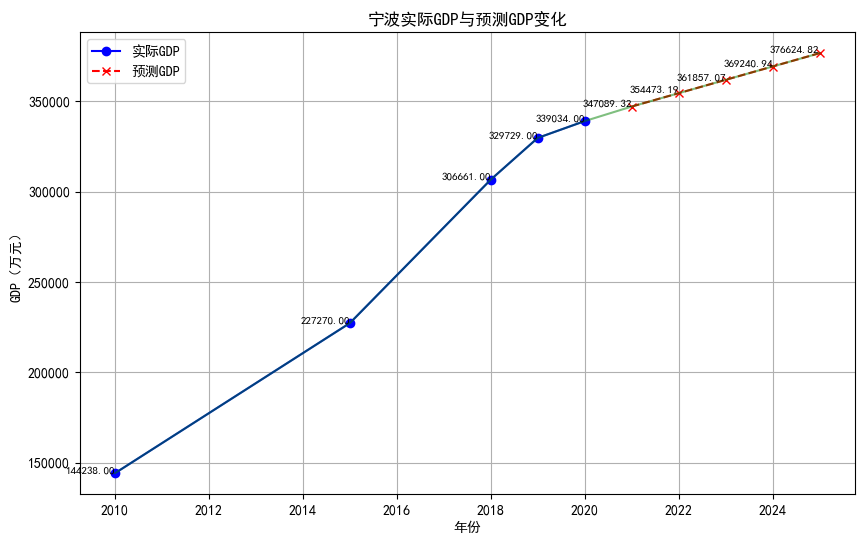


图 13宁波地区实际GPD与预测GPD折线变化图

从上方的折线图中可以看出,无论是历史的GDP还是预测的GDP,宁波的GDP都是呈现着稳步增长的趋势,可能是由于宁波市的全社会用电量与全市GDP有着密切的联系。通过查阅相关资料可以得知,宁波的人口在这五座城市当中并不算很多,但是宁波的GDP却比很多城市都要高,结合6.2.2中得出的结论,猜测宁波的全社会用电量主要是用在全市的各个工厂上,通过查阅资料可以得知宁波的第二产业在整个浙江省当中都是前几名的存在,所以说全市的社会用电量与生产总值的系数才会比总人口和生产总值的系数大。由于宁波的第二产业发展迅速,在未来的五年间依旧会不断地发展,因此宁波的GDP也会逐年地往上增加。

与青岛市相同的是,长沙市和无锡市的生产总值都是和总人口数有着最大的相关系数,因此它们两座未来五年的GDP变化也都会呈现一种上升的趋势。如图 14图 15所示:

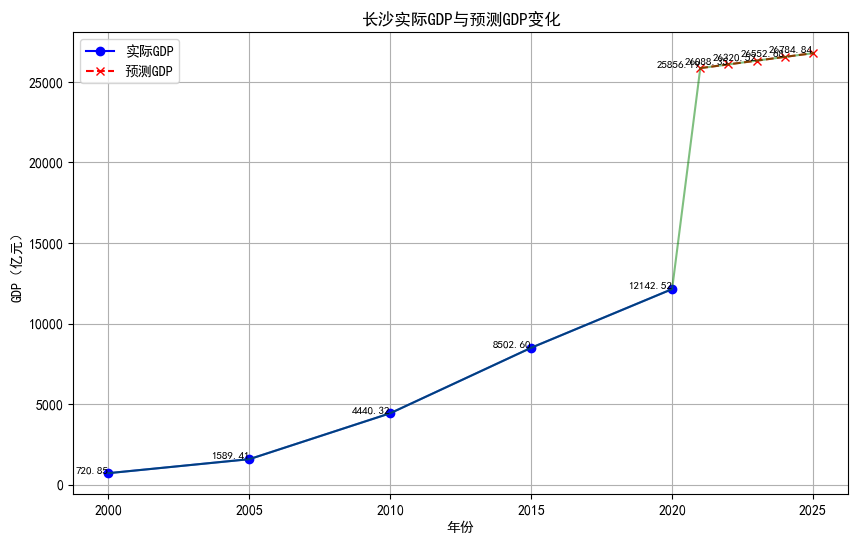


图 14长沙地区实际GPD与预测GPD折线变化图

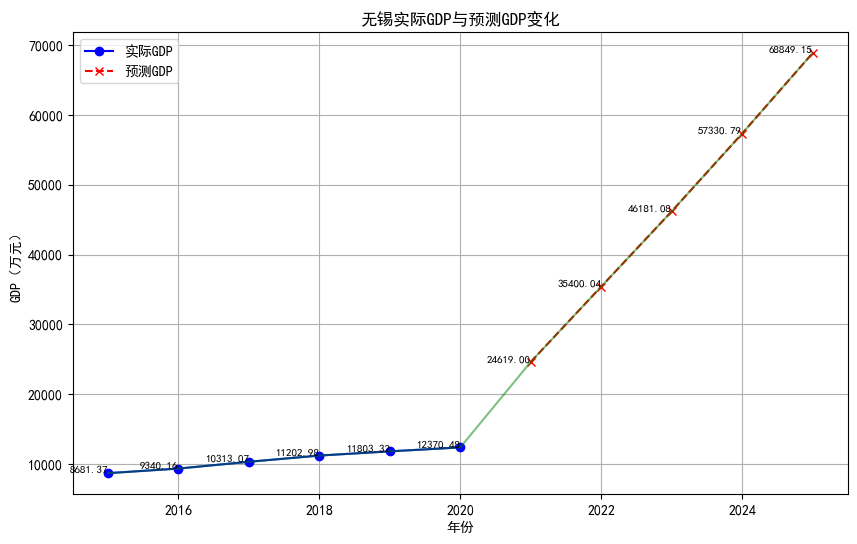


图 15无锡地区实际GPD与预测GPD折线变化图

但是和青岛不同的是,这两座城市的实际GDP变化趋势明显没有青岛市的实际GDP变化明显。尤其是无锡,在2020年之前无锡的实际GDP变化可以说是微乎其微。但是在2020年之后的预测GDP变化的折线上,无锡的GDP有了一个明显的起伏,结合相关的政策,造成无锡如此变化原因可能就是无锡市被评选为了新一线城市,从而吸引了大量的人口,结合上面任务中的无锡市Pearson系数相关矩阵热图,人口的飞速增长往往就和导致无锡市GDP的暴增。而长沙市在2020年之后的预测GDP断层增长的原因可能是长沙市政府在推动经济发展方面采取了积极有效的政策措施。例如，加大对科技创新的投入，优化营商环境，吸引更多的投资和企业入驻。这样就在吸引了人口的同时增加了城市的GDP，导致长沙市未来五年的GDP有了一个断层式的增长。

* + 1. 对五城GDP进行比较排序

通过综合分析对比宁波、青岛、长沙、无锡、郑州五座城市的经济发展历史、现在、未来状况,发现五座城市各具特色，各自在经济发展中展现了不同的风貌和潜力。

首先是青岛，这座拥有百年历史的海滨城市，近年来在科技创新和海洋经济领域取得了显著成绩。青岛正积极打造全球海洋科技创新高地，以科技创新为引领，推动海洋产业转型升级。未来，青岛将继续深化海洋科技创新，加强国际合作，提升海洋经济的国际竞争力。

长沙，作为湖南省的省会城市，近年来在制造业、科技创新和文化产业等方面取得了显著进展。长沙以智能制造为主导，加快构建现代产业体系，同时注重科技创新和人才引进，打造了一批具有国际竞争力的企业和品牌。未来，长沙将继续坚持创新驱动发展战略，推动产业高端化、智能化、绿色化发展。

无锡，这座江南水乡城市，以制造业和科技创新闻名于世。无锡拥有众多优秀的制造业企业和科研机构，形成了从原材料到成品的全产业链优势。未来，无锡将继续加强科技创新和产业升级，推动制造业向高端化、智能化、绿色化方向发展，同时注重城市环境和文化建设的协调发展。

郑州，作为中原地区的中心城市，近年来在交通物流、制造业和科技创新等领域取得了快速发展。郑州依托其独特的地理位置和交通优势，大力发展现代交通物流业，同时加强制造业和科技创新的发展，形成了具有较强竞争力的产业体系。未来，郑州将继续加强交通物流、制造业和科技创新的发展，推动城市经济持续健康发展。

宁波，这座港口城市，以开放、创新、务实的精神推动经济发展。宁波拥有完善的港口设施和便利的交通网络，为对外贸易和经济发展提供了有力支撑。同时，宁波还注重科技创新和产业升级，推动制造业向高端化、智能化、绿色化方向发展。未来，宁波将继续深化对外开放和科技创新，推动城市经济持续健康发展，并积极参与全球经济治理和合作。

总之，这五座城市在经济发展中各具特色，各自在科技创新、产业升级、对外开放等方面取得了显著成绩。但是总体来说,五城的GDP发展排序根据综合因素排序依次为青岛、宁波、长沙、无锡、郑州。

* + 1. 绍兴市GDP情况数据

与上面五座城市相比,绍兴市作为一座二线城市,肯定有着很多的不足之处。为了去比较它们之间的差距,本文也收集了绍兴市历年的城市GDP数据,绘制处理绍兴市的实际GDP与预测GDP折线变化图。如图 16所示:

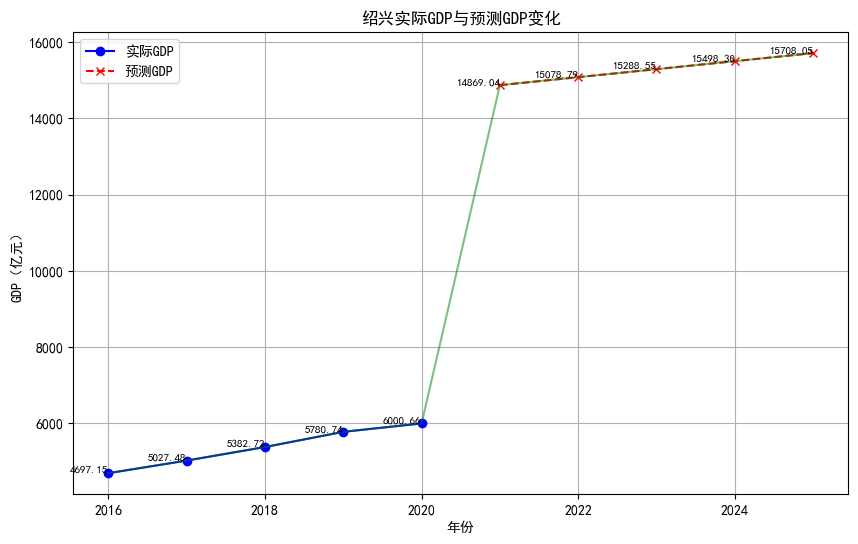


图 16绍兴市的实际GDP与预测GDP折线变化图

从绍兴市的实际GDP与预测GDP变化折线图可以看出,绍兴市在2020年之前的GDP的变化量并不大,但是在2020年之后绍兴市的GDP就得到了一个快速增长,因此需要去了解绍兴的经济结构以及近年来的相关政策[5]。通过查阅相关资料可以得知,绍兴市和宁波市一样,第二产业是经济的主导部分。其中,在工业领域特别是规模以上工业的发展势头良好，如计算机通信和其他电子设备制造业、专用设备制造业、纺织业、电气机械和器材制造业等行业增速高于全市平均水平。绍兴市的工业发展受益于大中型企业的快速增长和民营企业的支撑。结合绍兴市的相关政策,可以发现绍兴市在近几年全市推动创新驱动发展政策,数字经济核心产业发展政策,人才引进和培养政策以及出台相关的财政支持政策,这大大地吸引了外地人口来绍,结合上面几座城市的GDP发展,一座城市的生产总值往往与总人数有着很大的相关性,因此绍兴的GDP在人口数量增多的情况下也得到了一个很大的提示。

但是,与上面五座城市相比,绍兴的劣势还是比较明显的。例如绍兴市为了保护老城区的古建筑物不被破坏,限制了老城区的工业发展,导致地区产业结构分配不均匀,城市GDP难以得到一个平稳的发展。而且,绍兴市作为一个二线城市,知名度远远不如上面的几座一线城市,因此,在吸引人才的方面,绍兴市可能会占据一定的劣势。

1. 模型的优化与评价
   * 1. 模型的优点和缺点
        1. 模型的优点

ARIMA模型的优点不仅仅是其简洁性和直观性。它还允许我们捕获时间序列数据中的趋势、季节性和随机性，这是许多其他模型所不具备的。

其次，ARIMA模型具有很好的预测性能。通过调整模型的参数，我们可以精确地模拟过去的数据，并据此预测未来的趋势。这使得ARIMA模型在预测分析领域具有广泛的应用等。

再者，ARIMA模型具有很好的鲁棒性。这意味着即使数据中存在一些异常值或缺失值，ARIMA模型仍然可以给出相对准确的预测结果。

最后，ARIMA模型还具有很好的可解释性。模型的参数可以直观地解释数据的特征，如趋势的强度、季节性的周期和幅度等。这使得ARIMA模型不仅是一个预测工具，还是一个很好的数据分析工具，可以帮助我们更好地理解数据背后的规律。

* + - 1. 模型的缺点

尽管ARIMA模型具有诸多优点，但它也存在一些局限性。例如，它对于非线性关系的处理能力相对较弱，对于某些复杂的时间序列数据可能无法给出满意的预测结果。此外，ARIMA模型还需要对数据进行差分处理，这可能会损失一些原始信息。因此，在实际应用中，我们需要根据数据的具体特点选择合适的模型和方法。

* + 1. 模型的优化

由于在使用ARIMA模型对五城GDP进行预测时会出现准确度过低的可能,因此需要对ARIMA模型进行优化。一般常用的优化方法可以分为参数选择优化、差方次数优化、自动模型选择方法等。通过相关方法对模型进行优化后,可以显著提高ARIMA模型的预测精度和适用性，更好地适应不同类型的时间序列数据预测需求。

参考文献

1. 陈满丽,张慧娟,焦楠楠,等.基于MATLAB的ARIMA模型对我国GDP预测的研究[J].中国市场,2024(12):1-4.DOI:10.13939/j.cnki.zgsc.2024.12.001.
2. 张露康,王海蓉,古可言,等.基于Pearson相关性分析的交通枢纽站客运量预测模型分析研究[J].黑龙江交通科技,2023,46(03):137-139.DOI:10.16402/j.cnki.issn1008-3383.2023.03.030.
3. 熊沫,黄钦,王超,等.老年肺癌化疗患者化疗相关性味觉改变与生存质量的Pearson相关分析[J].齐鲁护理杂志,2023,29(22):13-17.
4. 王家俊.基于线性回归模型的铁路客运量预测与实证分析[J].智慧轨道交通,2024,61(03):102-105+114.
5. 赵国正,周建芳.绍兴市经济发展路径、困囿及对策研究[J].农场经济管理,2023(09):41-44.

[6]金岗. 长三角经济一体化背景下的绍兴地区经济发展对策研究[D].浙江大学,2006.

附录

附录1

import pandas as pd  
from matplotlib import pyplot as plt  
  
# 设置matplotlib以支持中文显示  
plt.rcParams['font.sans-serif'] = ['SimHei'] # 指定默认字体为黑体  
plt.rcParams['axes.unicode\_minus'] = False # 解决保存图像是负号'-'显示为方块的问题  
  
# 指定文件路径  
file\_path = r"E:\数据建模\z\图片\宁波市GDP.xls"  
  
# 使用pandas的read\_excel函数读取Excel文件  
df = pd.read\_excel(file\_path)  
  
# 显示数据框的内容（如果需要）  
# print(df)  
  
# 从第八行开始提取年份和生产总值数据  
years = df.iloc[**7**:**, 0**].tolist() # 第一列是年份数据  
gdp\_values = df.iloc[**7**:**, 1**].tolist() # 第二列是生产总值数据  
  
# 绘制年份与生产总值的折线图  
plt.figure(figsize=(**10, 6**)) # 设置图形的大小  
plt.plot(years**,** gdp\_values**,** marker='o'**,** linestyle='-'**,** color='b') # 绘制折线图  
plt.title('年份与生产总值的关系') # 添加标题  
plt.xlabel('年份') # 添加x轴标签  
plt.ylabel('生产总值（亿元）') # 添加y轴标签  
plt.xticks(rotation=**45**) # 旋转x轴标签，以便更好地显示  
plt.grid(True) # 显示网格   
plt.show() # 显示图形

附录2

import numpy as np  
import pandas as pd  
from matplotlib import pyplot as plt  
from statsmodels.tsa.arima.model import ARIMA  
  
# 设置matplotlib以支持中文显示  
plt.rcParams['font.sans-serif'] = ['SimHei'] # 指定默认字体为黑体  
plt.rcParams['axes.unicode\_minus'] = False # 解决保存图像是负号'-'显示为方块的问题  
  
# 指定文件路径  
file\_path = r"E:\数据建模\2020的题目\表\宁波市GDP.xls"  
  
# 使用pandas的read\_excel函数读取Excel文件  
df = pd.read\_excel(file\_path)  
  
# 假设第一列是年份，第二列是GDP  
years = df.iloc[**7**:**, 0**].values  
gdp = df.iloc[**7**:**, 1**].values  
  
# 数据清理：确保gdp数组是数值类型并且没有NaN值  
gdp = pd.to\_numeric(gdp**,** errors='coerce') # 将无法转换为数字的值设置为NaN  
gdp = pd.Series(gdp).dropna().values # 删除NaN值并转换为numpy数组  
  
# 检查清理后的数据  
if len(gdp) != len(years):  
 years = years[-len(gdp):] # 确保年份和GDP数据长度一致  
  
# 绘制历史GDP的折线图  
plt.figure(figsize=(**10, 6**))  
plt.plot(years**,** gdp**,** label='历史GDP'**,** marker='o')  
plt.title('宁波市地区生产总值（GDP）')  
plt.xlabel('年份')  
plt.ylabel('GDP')  
plt.legend()  
plt.grid(True)  
plt.show()  
  
# 构建和拟合ARIMA模型  
model = ARIMA(gdp**,** order=(**5, 1, 0**)) # (p,d,q) 这里p=5, d=1, q=0是一个简单的选择  
model\_fit = model.fit()  
  
# 预测未来5年的GDP  
forecast\_years = np.arange(years[-**1**] + **1,** years[-**1**] + **6**)  
forecast = model\_fit.forecast(steps=**5**)  
  
# 打印预测结果  
for year**,** gdp\_pred in zip(forecast\_years**,** forecast):  
 print(f"Predicted GDP for {year}: {gdp\_pred}")  
# 绘制历史GDP和预测GDP  
# 绘制历史GDP和预测GDP  
plt.figure(figsize=(**10, 6**))  
plt.plot(years**,** gdp**,** label='历史GDP'**,** marker='o'**,** color='blue')  
plt.plot(forecast\_years**,** forecast**,** label='预测GDP'**,** linestyle='--'**,** marker='s'**,** color='red')  
  
# 连接历史GDP和预测GDP  
plt.plot([years[-**1**]**,** forecast\_years[**0**]]**,** [gdp[-**1**]**,** forecast[**0**]]**,** linestyle='--'**,** color='gray')  
  
# 在每个节点上标注值  
for year**,** gdp\_value in zip(years**,** gdp):  
 plt.text(year**,** gdp\_value**,** f'{gdp\_value}'**,** ha='right'**,** va='bottom')  
  
for year**,** gdp\_pred in zip(forecast\_years**,** forecast):  
 plt.text(year**,** gdp\_pred**,** f'{gdp\_pred:.2f}'**,** ha='right'**,** va='bottom')  
  
plt.title('宁波地区生产总值（GDP）及未来5年预测')  
plt.xlabel('年份')  
plt.ylabel('GDP（亿元）')  
plt.legend()  
plt.grid(True)  
plt.show()

附录3

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from scipy.stats import chi2\_contingency

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import classification\_report

from sklearn.utils import resample

from sklearn.ensemble import RandomForestClassifier

from sklearn.model\_selection import GridSearchCV

from pylab import mpl

mpl.rcParams["font.sans-serif"] = ["SimHei"] # 设置显示中文字体 宋体

mpl.rcParams["axes.unicode\_minus"] = False #字体更改后，会导致坐标轴中的部分字符无法正常显示，此时需要设置正常显示负号

data = pd.read\_excel(r'D:\study\begin\python\城市比赛\1\W1.xlsx', index\_col=0)

print(data.index)

print(data.columns)

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# 将数据转换为pandas DataFrame

data\_dict = {

'年份': [2015, 2016, 2017, 2018, 2019, 2020],

'GDP': [8681.37, 9340.16, 10313.07, 11202.98, 11803.32, 12370.48],

'户籍人口': [480.9007, 486.2, 493.05, 497.21, 502.83, 508.97],

'货运量': [15353, 15830, 17385, 18640, 20601, 21691],

'社会消费品零售总额': [2186.34, 2357.62, 2585.69, 2785.21, 3024.34, 2994.36],

'农村常住居民人均可支配收入': [24155, 26158, 28358, 30787, 33574, 35750],

'城镇常住居民人均可支配收入': [45129, 48628, 52659, 56989, 61915, 64714],

'用电量（亿万千瓦时）': [598.18, 600.5, 638.67, 686.67, 732.81, 750.82]

}

import pandas as pd

# 创建DataFrame

df = pd.DataFrame(data\_dict)

# 计算Pearson相关系数矩阵

pearson\_corr = df.corr()

# 计算Spearman等级相关系数矩阵

spearman\_corr = df.corr(method='spearman')

# 打印相关系数矩阵

print("Pearson相关系数矩阵:")

print(pearson\_corr)

print("\nSpearman等级相关系数矩阵:")

print(spearman\_corr)

import seaborn as sns

import matplotlib.pyplot as plt

# 绘制Pearson相关系数矩阵的热图

plt.figure(figsize=(10, 8))

sns.heatmap(pearson\_corr, annot=True, cmap='coolwarm', fmt=".2f")

plt.title('无锡Pearson相关系数矩阵热图')

plt.show()

附录4

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from pylab import mpl

# 设置显示中文字体

mpl.rcParams["font.sans-serif"] = ["SimHei"] # 设置显示中文字体 宋体

mpl.rcParams["axes.unicode\_minus"] = False # 正常显示负号

# 新数据字典

data\_dict = {

'年份': [2010, 2015, 2018, 2019, 2020],

'生产总值（万元）': [144238, 227270, 306661, 329729, 339034],

'全社会用电量（万千瓦时）': [12577, 16029, 21240, 22135, 22736],

'社会消费品零售总额（万元）': [44320, 83567, 108613, 116947, 115800],

'货运量（万吨）': [83.71, 115.30, 168.37, 187.42, 196.44],

'总人口（万人）': [574.08, 586.57, 602.96, 608.47, 613.66],

'城缜居民人均可支配收入（元）': [30166, 47852, 60134, 64886, 68008],

'农村居民人均可支配收入（元）': [14261, 26469, 33633, 36632, 39132]

}

# 创建DataFrame

df = pd.DataFrame(data\_dict)

df = df.rename(columns={'生产总值（万元）': 'GDP'})

# 准备数据

X = df[['全社会用电量（万千瓦时）', '社会消费品零售总额（万元）', '货运量（万吨）', '总人口（万人）', '城缜居民人均可支配收入（元）', '农村居民人均可支配收入（元）']] # 自变量

y = df['GDP'] # 因变量

# 拟合多元线性回归模型

model = LinearRegression()

model.fit(X, y)

# 预测未来5年自变量

future\_data\_dict = {

'年份': [2021, 2022, 2023, 2024, 2025],

'全社会用电量（万千瓦时）': [23500, 24300, 25100, 25900, 26700],

'社会消费品零售总额（万元）': [120000, 124000, 128000, 132000, 136000],

'货运量（万吨）': [205, 214, 223, 232, 241],

'总人口（万人）': [620, 627, 634, 641, 648],

'城缜居民人均可支配收入（元）': [70000, 72000, 74000, 76000, 78000],

'农村居民人均可支配收入（元）': [41000, 43000, 45000, 47000, 49000]

}

future\_df = pd.DataFrame(future\_data\_dict)

X\_future = future\_df[['全社会用电量（万千瓦时）', '社会消费品零售总额（万元）', '货运量（万吨）', '总人口（万人）', '城缜居民人均可支配收入（元）', '农村居民人均可支配收入（元）']]

# 预测未来GDP

predicted\_gdp = model.predict(X\_future)

# 结合实际数据和预测数据

future\_df['GDP'] = predicted\_gdp

combined\_df = pd.concat([df[['年份', 'GDP']], future\_df[['年份', 'GDP']]], ignore\_index=True)

# 绘制实际GDP与预测GDP的变化图

plt.figure(figsize=(10, 6))

plt.plot(df['年份'], df['GDP'], label='实际GDP', marker='o', color='blue')

plt.plot(future\_df['年份'], future\_df['GDP'], label='预测GDP', linestyle='--', marker='x', color='red')

# 将现有值和预测值连接起来

plt.plot(combined\_df['年份'], combined\_df['GDP'], linestyle='-', color='green', alpha=0.5)

# 添加数据标签

for i, row in df.iterrows():

plt.text(row['年份'], row['GDP'], f'{row["GDP"]:.2f}', ha='right', fontsize=8)

for i, row in future\_df.iterrows():

plt.text(row['年份'], row['GDP'], f'{row["GDP"]:.2f}', ha='right', fontsize=8)

plt.xlabel('年份')

plt.ylabel('GDP（万元）')

plt.title('宁波实际GDP与预测GDP变化')

plt.legend()

plt.grid(True)

plt.show()