

重庆理工大学

数学建模

A 题 重庆各区县人口与经济时空演化建模分析

参赛队员信息

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 学号 | 学院 | 电话 |
| 汤荣杰 | 11703990806 | 计算机科学与工程学院 | 15823502380 |
| 谢锦涛 | 11703990827 | 计算机科学与工程学院 | 18315179327 |
| 黄凤翔 | 11703990304 | 计算机科学与工程学院 | 18383091831 |

重庆各区县人口与经济时空演化建模分析

摘要

重庆直辖以来，人口与经济都发生了深刻变化，而人口对经济的影响又是一个牵动全局，事关长远的重大战略性问题。本文则依据获取的数据，通过处理和分析，对人口和经济进行建模分析，并对重庆各区县的未来发展作出评价与预测。

针对问题一，收集数据并进行描述性统计分析和数据可视化。我们通过在重庆统计年鉴网站上获取重庆近20年的各区县的人口与经济的数据，通过查阅资料和利用Python查看数据的相关性以地区生产总值和三大产业（农业，工业和服务业）总值作为经济的主要指标，而人口分为了农村人口和非农村人口。然后通过折线图，柱状图等统计图形对人口和经济主要指标进行可视化，了解数据的基本信息，如基本趋势，连续性等。

针对问题二，从时间角度分析人口和经济的走势。本文利用折线图对人口分别和选取的经济主要指标（地区生产总值和三大产业总值）进行了走势分析，总结出人口因素与地区生产总值和三大产业总值都存在正相关性，但和地区生产总值的正相关性更强。

针对问题三，从空间角度分析人口与经济的聚集性与空间分布。本文利用Python中pyecharts的Map模块对人口密度和地区生产总值进行了地图上的空间分布可视化，通过对比分析两者各自的分布以及之间的联系。发现人口密度和地区生产总值的分布都存在空间性和层次性，并且两者的分布趋势相近。此外，通过地图分析出，区县的占地面积和地区生产总值有着负相关性。

针对问题四，对人口和经济进行建模并分析，通过分析将其抽象成了一个多元回归问题。通过问题一，问题二，问题三的分析，最终选择地区生产总值作为经济的指标，也作为模型的目标变量，从分析中我们定义出地区生产总值的五个影响因素，即人口总数，地区占地面积，人口密度，农村人口数，非农村人口数，通过最小二乘法拟合出四个影响因素的系数，分别为：0.0356，-0.0326,0.1335,-0.0431,1.6566，其偏置项即截距为1.5006，通过数据进行验证得到系数是合理的。

针对问题五，对各区县未来发展做出评价和预测。通过分析和建模得到的结果，本文对各区县就其发展的优势和缺点进行评价，同时通过建模的结果做出一定的预测。

最后，给出模型的优缺点以及评价推广。

【**关键词**】：描述性数据分析，数据可视化，Python，多元回归

一、问题重述

**1.1 问题背景**

一直以来，人口与经济的关系一直都是国家关注的重点，充分，定量地的研究人口发展和各种因素对经济带来的影响效益，能够为城市或者国家发展提供数据和方法支持，使得能够及时掌握人口信息，在一定程度上控制人口产生的影响，从而促进城市或者国家经济又好又快的发展。

重庆直辖以来，人口与经济都发生了深刻变化，人口与经济有着密切关系，而重庆作为国家各个方面的枢纽城市，存在着一定的人口流动变化，因而研究重庆人口与经济的发展也变得极为重要。

**1.2 问题提出**

在以上所述的背景下，就重庆各区县人口与经济方面的统计数据进行建模分析。

1、收集到近20年以上重庆各区县人口与经济相关的主要指标数据，并进行描述性统计分析及数据可视化。

2、从时间角度分析人口与经济的走势。

3、从空间角度分析人口与经济的聚集性与空间分布。

4、建立人口与经济的相互关系的数学模型并进行分析。

5、对各区县的未来发展作出评价与预测。

二、问题分析

**2.1 问题1的分析**

问题1要求的是让我们收集到近20年以上重庆各区县的人口数据和经济相关的主要指标数据。收集到数据后，再进行一些信息的描述性统计以及数据的可视化。对于数据，我们可以从互联网上公开的数据网站获取。

在分析数据之前，通过查阅文献，我们了解到地区生产总值（地区GDP）以及农业（第一产业），工业（第二产业）和服务业（第三产业）三大产业的生产总值对经济的代表性是比较大的，所以我们选择了地区生产总值（地区GDP）和三大产业的生产总值这四项作为经济的主要指标。

选择好经济的主要指标后，我们则通过数据的描述性统计和数据可视化的方法，包括柱状图，折线图，散点图等统计图形分析人口，地区GDP以及三大产业的数据情况，包括数据的集中性，离散性以及数据的分布趋势等信息。

**2.2 问题2的分析**

问题2考虑到在时间角度下，人口和经济的相互关系。对于这点，同样也可以通过借助基本的统计图形，以时间为横坐标轴，然后以人口和经济的主要指标为纵坐标，展示出随着时间的推进，人口和经济指标两者的变化情况。通过分析人口和经济指标的变化趋势，我们可以判断两者之间的变化是否存在一定的关系。

通过所选择的经济的主要指标，我们以时间为自变量，分析地区GDP以及农业，工业和服务业三大产业的生产总值在自变量之下的变化情况，从而总结人口和四个经济指标在时间下的走势情况

**2.3 问题3的分析**

问题3是要求从空间角度方面，分析人口和经济的聚集性信息以及空间分布情况。对于这个问题，则可以在重庆区域地图上通过颜色深浅展示人口以及经济指标的数值情况，以分析人口和经济指标的分布情况。

通过Python语言库中pyecharts的Map模块可以实现地图的绘制，绘制地图后，首先可以单独分析两者的空间分布情况，然后则结合两者的共同或不同的空间分布情况，分析两者之间的空间聚集关系。

**2.4 问题4的分析**

问题4详细到了模型的建立问题，要求建立人口与经济之间相互关系的数学模型。

在我们选择的四个主要经济指标中，地区GDP和人口的关系是最大的，所以我们选择它作为模型的目标变量。要分析人口与经济的相互关系，则要找到人口中影响地区GDP的因子。

我们通过前面的描述性统计分析，以及从时间和空间两个角度对经济的讨论中，将经济的影响因子定量化，并定义出影响地区GDP的影响因子，完善建立的数学模型。然后则根据已有的数据对模型进行拟合求解出影响因子的相关系数，最后根据得到的因子系数对模型的合理性进行分析，对结果进行验证。

**2.5 问题5的分析**

问题5是要求在通过分析和建模之下，对重庆各区县的未来发展做出评价和预测。

在问题1-3中我们借助数据对人口和经济之间的关系进行了详细的分析，在模型的分析之下，我们也可以看到很多问题，针对这些问题，我们对重庆各区县的发展做出一些评价。同时在问题4下进行了模型的建立，对模型的一些参数进行了求解和验证，所以我们可以通过模型对重庆各区县的未来经济等发展进预测。

1. 模型假设
2. 假设获取的数据信息准确可靠。
3. 假设在数据涉及的时间年限内，重庆人口无太大的人口迁入或者迁出。
4. 假设在数据涉及的时间年限内，除个别地区以外，重庆各区县的占地面积无太大的变化。
5. 符号说明

|  |  |
| --- | --- |
| 变量 | 说明 |
| Y | 地区生产总值 |
| Ci | 影响因子系数 |
| z1 | 人口总数 |
| z2 | 地区占地面积 |
| z3 | 人口密度 |
| z4 | 农村人口数量 |
| z5 | 非农村人口数量 |

1. 模型的建立与求解

**5.1 问题一**

针对问题1中的问题，我们通过一些统计图形来浏览数据的基本信息，同时发现数据中的一些规律或者其它信息。

首先，通过Python读取数据后，查看相关性，根据相关性我们从获取的数据中选择了用于分析的数据项信息。人口中，我们选择了地区的人口总数，自然增长率，非农村人口数信息，经济指标中选择了地区生产总值，三大产业生产总值信息。

**5.1.1 人口信息分析**

通过Python语言及对应的编译器，我们首先画出了各地区总人口随时间（1996-2017年）的变化情况。通过对总人口随时间变化的分析，我们总结出三种变化趋势。每一种变化趋势以某一个区县为例，如下图：

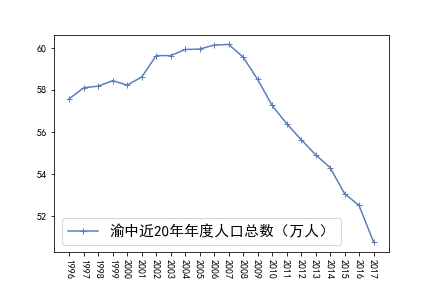


图5.1-1 渝中区近20年人口数变化图

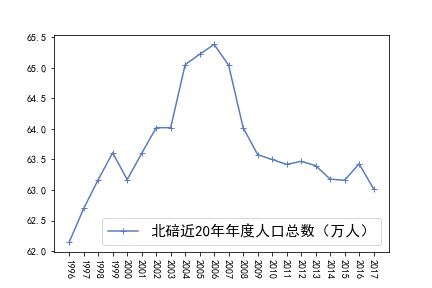


图5.1-2 北碚区近20年人口数变化图

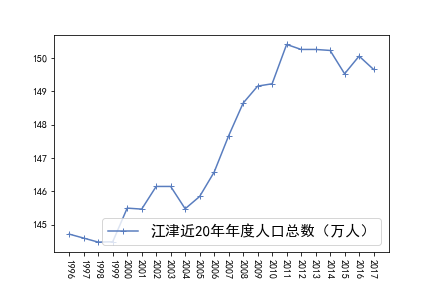


图5.1-3 江津区近20年人口数变化图

从以上三个图中可以很清晰的看出各区县人口的变化的三种情况。

第一种，下降型。如图中渝中区的情况，虽在前十年人口趋于稳定，但近20年总体的情况有下降的趋势。

第二种，稳定型。如图中北碚区的情况，人口先有一定的上升，然后又有较小的下降，总体人口情况则趋于稳定。

第三种，上升型。如图中江津区的情况，人口数量在波动中上升，最后逐渐有稳定的趋势。

当然，从得到的图来看，各个区县若是人口上升或者下降，其幅度也是偏小的，没有过大的人口变化。

运用同样的方法，画出各个区县人口自然增长率随时间（1996-2017年，有部分年限得到的数据自然增长率是缺失的）的变化的折线图，以江津区的人口自然增长率变化为例，如下图：

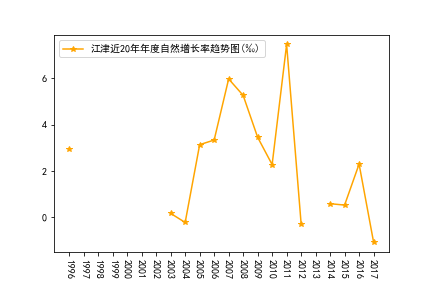


图5.1-4 江津区近20年人口自然增长率变化图

从图中可以看出，江津区的人口自然增长率处于波动状态，而且波动的范围也较小。同时，其人口自然增长率有正的情况也有负的情况，但正的年份较多，可以知道其人口大体是处于上升状态，对比前面江津的人口情况，可以看出信息是合理的。

通过对所有区县的情况进行查看，再对比人口的变化信息，自然增长率的变化和人口总数的变化都是一致的，能够反应人口的变化趋势。

此外，我们对农村人口和非农村人口的信息也进行了可视化。我们画出了每一年各地区农村人口和非农村人口的数量信息，以下是2017的信息：

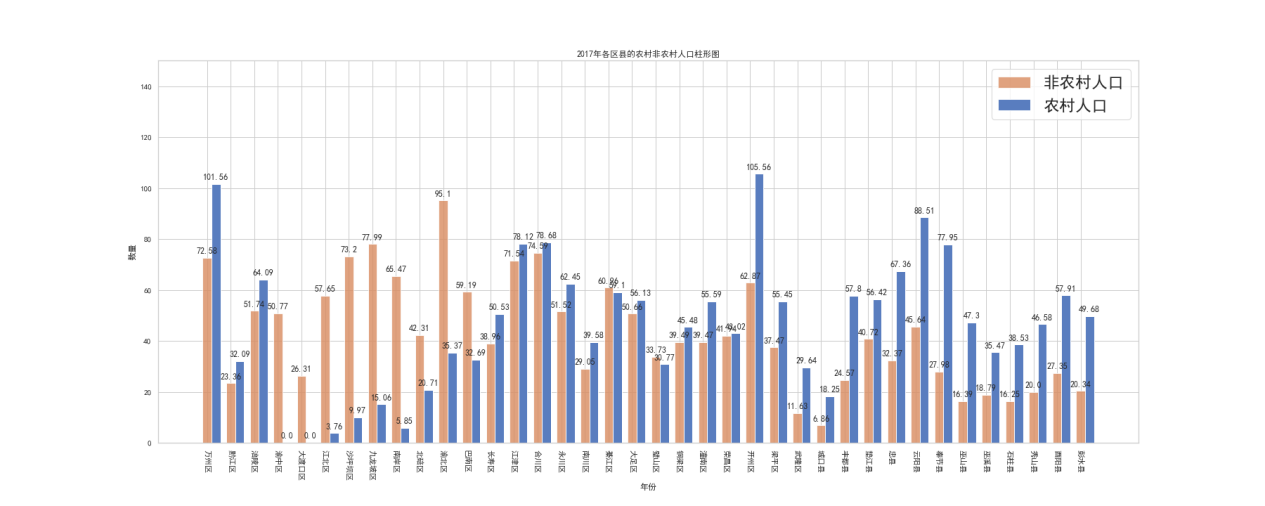


图5.1-5 2017年各区县农村与非农村人口对比图

从图中可以看见，包括渝中区，九龙坡区，江北区等几个经济发展良好的地区非农村人口与的占比是比较大的，而其它的区县农村人口都占有较大的比例。由此可知，农村人口和非农村人口与经济之间也存在着一定的相关性。

**5.1.2 经济指标信息分析**

通过Python语言及其对应画图的模块，我们画出了每一年下各地区的地区生产总值情况的柱状图，以下列出2001年和2017年的总体情况：

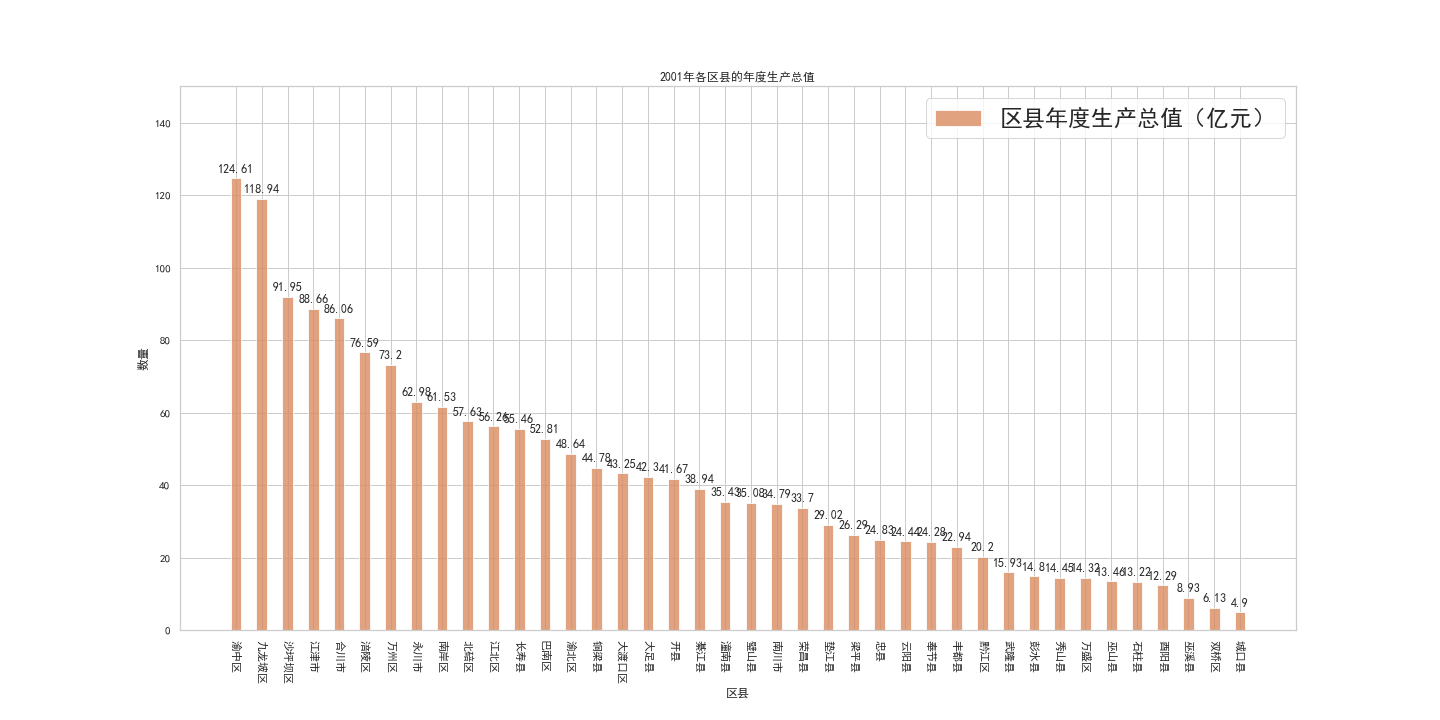


图5.1-6 2001年各区县地区生产总值

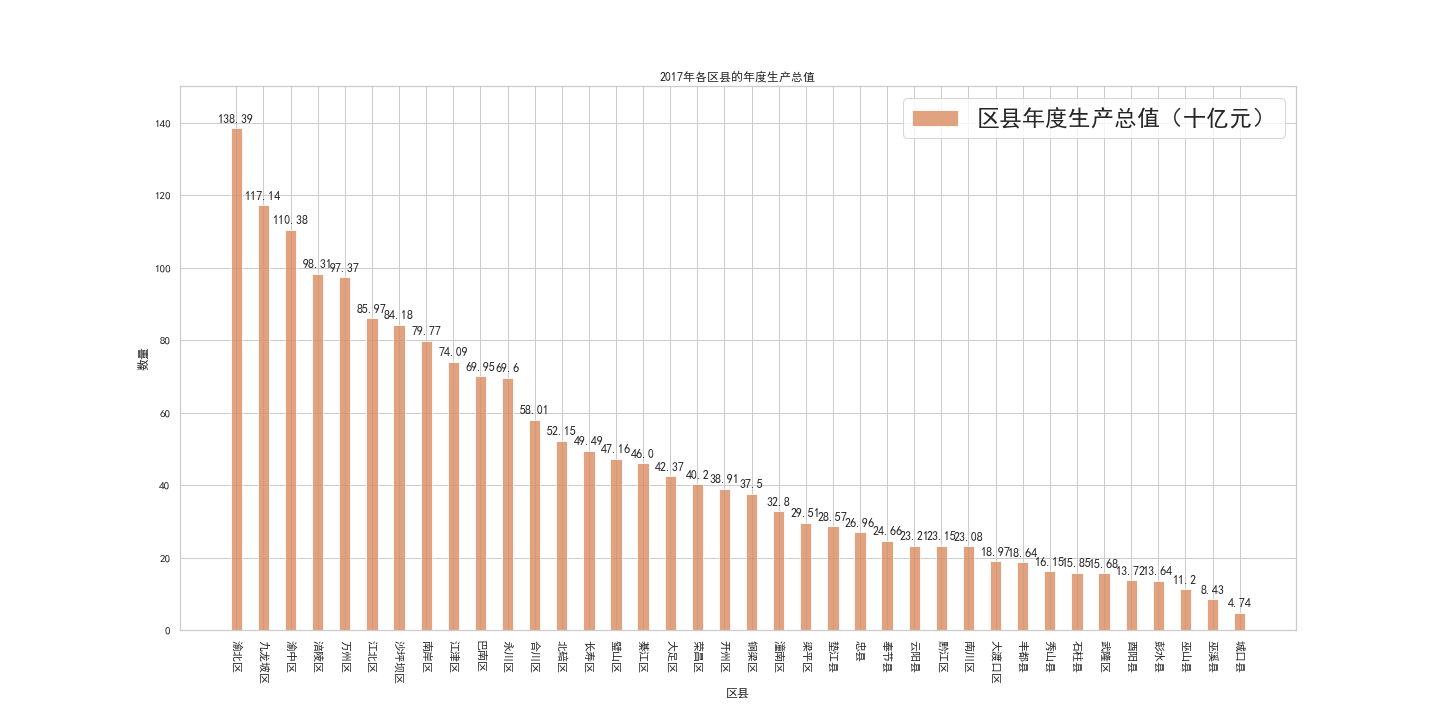


图5.1-7 2017年各区县地区生产总值

在以上的图中，是我们通过以地区生产总值对地区排序过后得到的2001年和2017年各区县的柱状图。

从图中我们可以看出在前列的几乎是渝中区，渝北区，九龙坡区等几个地区，结合它们的地理位置也可以发现，这几个地区比较靠近，而且都属于重庆经济较发达的区县。由此，我们可以将其人口信息联系起来，联系起来之后，我们发现，其地区生产总值和人口数等有着一定的正相关性。

**5.2 问题二**

问题2中，需要从时间角度分析人口与经济的走势，这个可以通过将人口和经济指标放在同一坐标下，画出其折线图，从而根据它们折线的走势分析它们之间的相关性。

通过Python语言我们画出了各个地区在时间年限下，人口总数和地区生产总值虽时间的变化情况的折线图，以巴南区和北碚区为例，如下图：

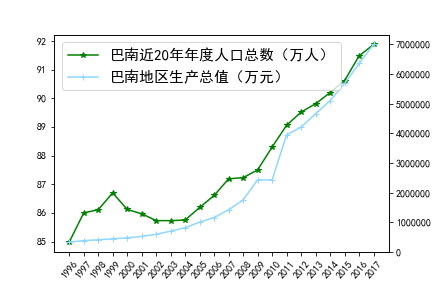


图5.2-1 巴南区近20年人口数与地区生产总值对比图

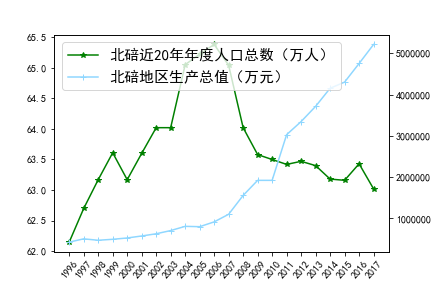


图5.2-2 北碚区近20年人口数与地区生产总值对比图

从巴南区的情况来看，人口总数和地区生产总值的走势是一样的，在同一时间下，人口总数在上升的时候，地区生产总值也有跟随上升的趋势。说明两者之前是存在一定的相关性。而在北碚区的情况图中，当将纵坐标扩大后发现，两者也是有正相关性的，但存在一些较小的误差。

从其它区的情况来看，大部分都是与巴南区的情况是相同的，而只有小部分是与北碚区的情况相同的，由此可见，认为人口数和地区生产总值之间存在正相关性是合理的。对于误差，可以通过建模的过程将其缩小。

另外，我们对三大产业则进行了占比的计算，即其在地区总产值中所占的比列，再对其进行可视化，得到各地区从1996年到2017年三大产业的占比变化情况，以下以巴南区的信息为例，同时列出巴南区的人口情况以进行对比分析：

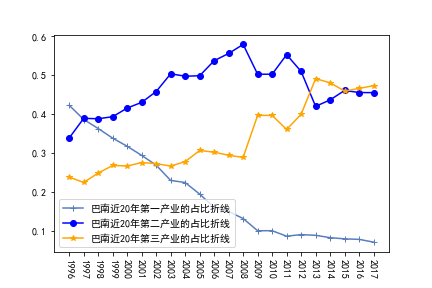


图5.2-3 巴南区近20年三大产业占比

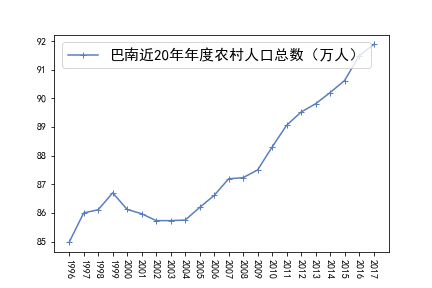


图5.2-4 巴南区近20年人口数变化情况图

从图中可以看出，巴南区第一产业即农业的比重在逐年下降，而第二和第三产业即工业和服务业在逐年缓慢上升且有趋于平稳的趋势，可见巴南区的发展方向由最初的三产业均衡发展转向于着重发展工业和服务业。而巴南区的人口也是处于上升趋势的，和第二第三产业的趋势是大体相近，由此也可知两者之间有着正相关性。

同样，其它地区也有着重发展的产业，但根据得到的产业占比图发现，经济发达的几个地区都更着重发展工业和服务业，而经济发展较为落后的地区则都会再一定程度上重视对农业的发展。并且人口数的变化和第二和第三产业的变化是相近的。

**5.3 问题三**

通过导入Python语言库中pyecharts的Map模块，同时将地区生产总值数据导入进来，从而得到重庆地区各区县的地区生产总值的分布情况。此处我们列出2017年的地区生产总值分布图如下（其它年份的图则在附件中展示）：

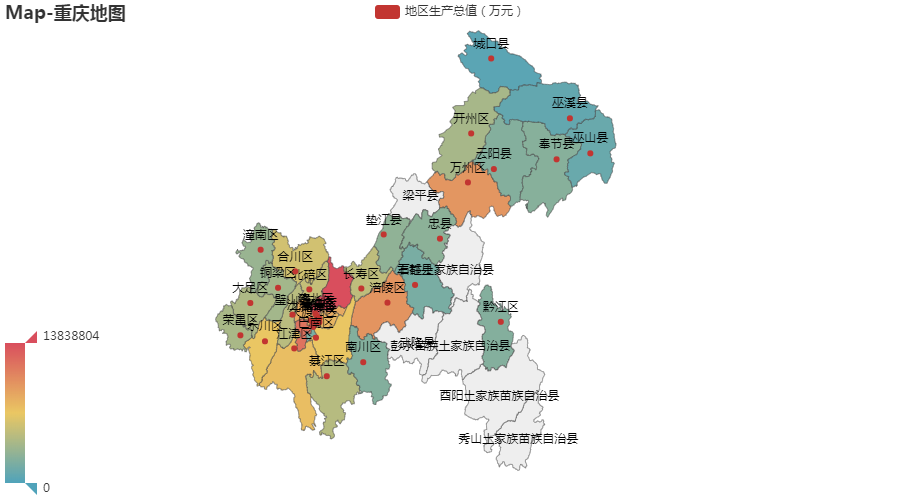


图5.3-1 重庆各区县地区生产总值空间分布图

由于包括渝中区的几个主城区面积较小，同时也为了更好的观察这几个地区的情况，我们单独把渝中区及周围几个地区的地区生产总值分布情况单独画出，如下图：

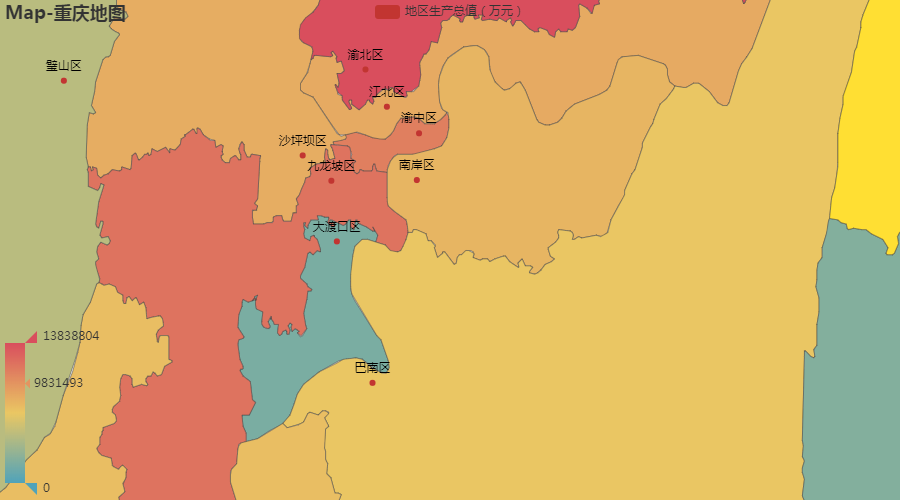


图5.3-2 重庆主城区生产总值空间分布图

从图左下角的图表我们可以知道，颜色越深，其地区生产总值的数值越大，其经济的发展则更好。同时，通过各个区县的地区生产总值的情况对比，也能分析出经济发展良好或者经济发展落后的区县，从而可以对某些区县进行针对性的分析，发现其优点或者缺点，从而制定对应的解决对策。

从两个图中可以发现，重庆地区显著地形成了以主城9区为主的高经济增长的聚集区，以及主要分布在渝东南和渝东北的低经济增长聚集区。

通过对比各地区经济以及对应的占地面积，可以总结出其地区生产总值和占地面积有负相关性。

同时，我们利用数据中的年末人口总数除以各地区的占地面积得到个地区的人口密度，然后利用Map模块画出人口密度的空间分布图，同样列出2017年各地区的人口密度图如下：

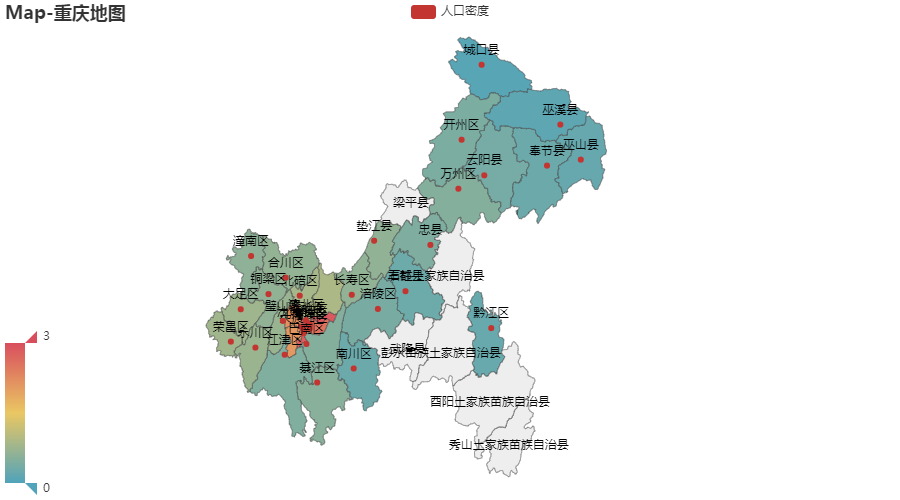


图5.3-3 重庆各区县人口密度空间分布图

对主城区的人口密度分布同样单独画出，如下图：

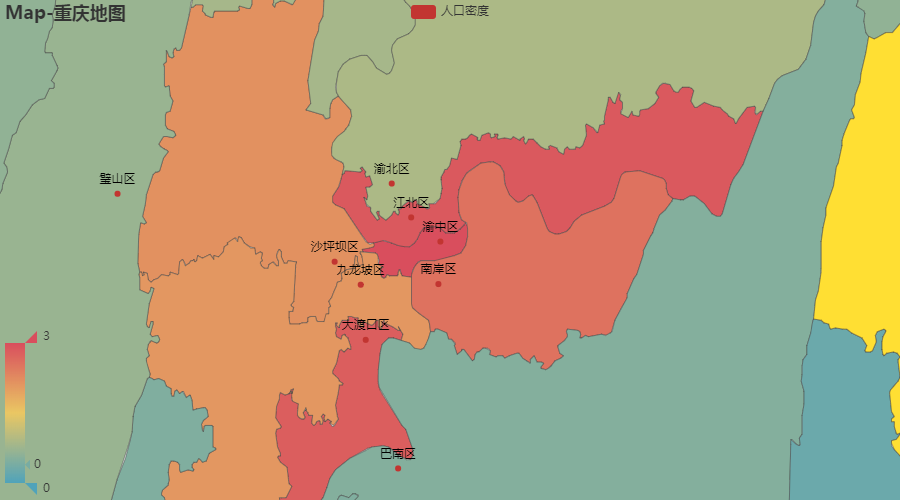


图5.3-4 重庆主城区人口密度空间分布图

这样，通过各区县的地区生产总值和人口密度的空间分布图，我们可以从以下三个方面去分析人口与经济的聚集性与空间分布。

1. 空间性。从上面地区生产总值的总体分布图中，可以看到其分布是具有空间性的。即某个地方的地区生产总值偏高，也就是经济聚集程度越高时，其周围的几个地区的地区生产总值也会偏高。这也体现了经济空间性聚集的特点：当某个地区的经济聚集程度越高时，它会对周围的地区形成一个辐射现象，并且辐射能力随其聚集程度的变高而增强。同时，在这些经济聚集程度较高的地方，它的人口密度也是较大的，由此可见，人口密度也具有一定的空间性，而且可以看出人口密度与经济之间存在着一定的正相关关系。
2. 经济累积性。经济累积性则从时间上进行阐述，通过对比每一年的地区生产总值分布情况，我们发现，地区生产总值较高的区县，比如渝中区及其周围几个城市，其在连续几年内的地区生产总值都是偏高的，而且有一定的增长。即可见，经济聚集程度的高度聚集，也会给该区域带来一定的经济增长。由经济的增长我们可以联想到人口密度，这些因经济的聚集而带来经济增长的地区，它们的人口密度也是偏大的。
3. 层次性。从上面的人口密度和地区生产总值空间分布图，可以很清晰明确的看出两者都是具有层次分布的。主城区比如渝中区及其周围地区，其人口密度较大，地区生产总值也较大，以其作为中心向四周扩大，人口密度和地区生产总值在一定程度上都有减小。由此可见，其经济都聚集在了像渝中区这样的主城区中，而四周城区的经济聚集程度都较低。将人口和经济指标对比分析时，发现它们的变化有着相同的趋势，都是向四周递减，而且中间主城区和四周城区形成了梯度上的差异。

**5.4 问题四**

**5.4.1 数据的预处理**

我们所获取的数据中，有部分数据项的值是缺失的，同时，各地区同样的一项是数据，有的单位可能不一样。经过统计和检查，我们发现数据缺失的值是比较少的，通过数据整体的分布情况，发现这些少数的缺失值有无的情况下,整体的分布几乎不受影响，所以我们选择直接剔除掉。对于单位不一样的，则通过单位换算将其转化为同样的单位。

**5.4.2 经济指标影响变量**

在问题1和问题2的分析中，我们总结出了人口总数与地区生产总值之间存在着较大的正相关性，人口总数也能在一定程度上反应经济的增长情况，即对经济有着一定的影响。同时，农村人口与非农村人口也对经济有一定的影响。

在问题3的分析中，我们可以得到一些启发，即人口密度和经济的相关性是比较大的。从人口密度和经济指标的空间分布老看，它们都有着同样的分布趋势，都具有空间性和层次性，因此我们选择人口密度作为经济指标的影响变量之一。

另外从问题3的空间分布中，还可以看出的是，几个经济聚集区的占地面积都是较小的，而其它较为落后的地区的占地面积也相对较大。因此，有理由认为地区的占地面积对经济是有影响的。

因此，通过分析，将问题转化为一个多元回归问题。

**5.4.3 问题4模型的建立**

记Y为地区生产总值，则：

Y=a+C1z1+C2z2+C3z3+C4z4+C5z5 （1）

其中，a为偏置项即常数项，z1为人口总数，z2为地区占地面积，z3为人口密度，z4为农村人口，z5为非农村人口，Ci为影响变量的系数。

此外，几个影响变量的单位以及数量级都是不同的，如果直接用于计算，则会因为这些不同而求解错误，所以我们使用标准化的方法解决量纲和数量级的问题。

记x为样本数据的影响变量，μ为样本数据的数学期望，σ2为样本数据的方差，则标准化方法为：

CodeCogsEqn(1)

对模型中涉及的影响变量都进行标准化处理，得到处理后的影响变量。

**5.4.4 问题4模型的求解**

根据模型的要求，对得到的数据进行选取，找到对应的变量，包括1996-2017年各个地区的数据。通过使用Python编写程序，通过最小二乘法拟合（1）中的系数，结果如下表所示：

表5.4-1 系数计算结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| a | C1 | C2 | C3 | C4 | C5 |
| 1.5006 | 0.0356 | -0.0326 | 0.1335 | -0.0431 | 1.6566 |

通过抽取数据进行检验，对抽取的数据用模型进行计算，再将结果与已有数据进行比较，得出系数计算结果是可靠的。

所以，带入（1）中得到：

Y=1.5006+0.0356z1-0.0326z2+0.1335z3-0.04314z4+1.6566z5

从结果来看，得到的结果符合我们在前3个问题中的探究，即人口总数，地区占地面积，人口密度和非农业人口是和地区生产总值是正相关的，而农村人口和地区生产总值是负相关的。

**5.5 问题五**

重庆各个区县的发展不论是从时间角度分析，还是从空间角度分析，都呈现出其很强的特点：1. 在时间上经济呈现稳步上升的状态；2. 在空间上经济则是呈现出很强的聚集性。

纵观重庆市各区县20年的统计数据，上述两种状态基本一致。对每个区县每年的地区生产总值用折线图或柱状图进行表示，可以明显的看到逐年稳步上升的趋势。按照各区县所处的地理位置，将全市 38 个区县的地区生产总值表示在地图上，可以发现各区域之间的经济差异较大，发展状况不平衡。主城 9 区与渝东北和渝东南地区差距比较明显。

**5.5.1 各区县发展的优势**

1.区县经济增长快速。从经济总量看，全市 38 个区县中，有 7 个区县地区生产总值超过 500 亿元，最大的为渝北、九龙坡、渝中。最小的三个县分别为城口、巫溪、巫山，均不到100亿元。从增长速度看，璧山、沙坪坝、涪陵增速位居全市前三；除了处于转型升级中的九龙坡和大渡口以外，全市其余 36 个区县均保持了两位数以上的经济增长，呈现出较好的增长势头，并且这个势头在一段时间内都会一直保持下去。

2.区域性增长极初步形成。近年来，万州、涪陵、黔江、永川、江津、合川等六大区域性中心城市充分利用区位和政策优势，不断完善城市基础设施和综合服务功能，加快推进城镇化进程，很好地发挥了承接主城、联动周边的作用，初步形成区域性增长极。

3.除了主城少数区以外，重庆市各区县总体处于工业化初期或中期起步阶段，大多属于承接沿海产业转移，产业分散化特点明显，区域内自身培育的品牌企业比较少，真正有影响力、有竞争力的产业集群还比较欠缺。在市级层面除了信息产业和汽摩以外，各区县还缺乏带动力强的支柱产业。江津、璧山等新兴工业大区虽然崛起较快，但大而不强的特征日益突出。两翼区县工业仍以传统的初级产品加工业为主，产业链相对较短，附加值低。

未来各区县需要继续发展优势，保持良好的势头，从而实现经济上的更近一步发展。

**5.5.2 各区县发展的缺点**

相邻区县产业趋同，没有发展好属于自己的产业。比如，“一圈”区县中，涉及摩托车产业的有 9 个区县，涉及汽车产业的有 7 个区县；随着全市电子信息产业的兴起，“一圈”内几乎每个区县都把电子信息作为主导产业。第三产业仍以传统的商贸、餐饮、运输服务业为主，现代服务业比重低，尤其是生产性服务业发展滞后，制约了大项目的引进。农业结构调整规划不够科学，产业化程度不高，科技含量比较低，名牌培育相对滞后。各区县对外开放层次不高，区县间的经济合作处于起始阶段，区县合作等实质性举措不多。

在未来，若若区县能着重关注上述缺点，在发展上必定能起到很好的效果，实现更进一步的发展和提升，突破历史，成就新高。

**5.5.3 预测**

在没有较大政策干预的情况下，重庆市都市区在未来十年的人口发展呈持续增长趋势，人口年均增长率接近1%，每年平均增长人数约 7.5 万人。城市发展新区在未来十年的人口发展呈持续增长趋势，人口平均年增长速度高于都市区，人口年均增长率在 1%-2%之间，每年平均增长人数约 15 万人。渝东北生态涵养发展区在未来十年的人口发展呈缓慢减少趋势，人口年均增长率接近-1%，每年人口减少量约 2.6 万人。渝东南生态保护发展区在未来十年的人口发展呈减少趋势，人口年均增长率约为-4.3%，平均每年人口减少约 11 万人。总体而言，重庆市都市区、城市发展新区是吸引人口在此集聚的人口流入区，常住人口数量呈增长的发展趋势，其中城市发展新区人口增长率会高于都市区人口增长率。渝东北生态涵养发展区、渝东南生态保护发展区是主要的人口流出区，常住人口数量呈负增长的发展趋势。这样的人口发展趋势，既有利于重庆市集中发展城市经济，又有利于涵养生态、保护生态环境。

对于类似重庆这样的地区，加大资金和人力的投资力度是根本之策，但是不能盲目增加投资，更不能以牺牲结构优化、产业升级、生态环境等为代价。科学合理的选择是，以加大投资促进总量增加，以总量增加带动结构调整，以结构调整推动经济可持续发展。

从分析中可以看出，重庆市人口数量多，增长速度快，但过快的速度，会影响资金积累，人口增长速度超过经济增长速度会导致部份经济发展成果被增长的人口规模抵消。虽然一定数量和质量的劳动力是促进经济发展必不可少的条件，但是如果一个地区的人口数量过多、增长过快，会使消费需求增加，积累减少，从而引起劳动力过剩、就业困难。紧接着就会导致一些社会问题产生，如：交通拥挤、住房紧张、教育和卫生设施不足等，这些都会阻碍经济的增长。

如果人口增长过快，地区生产资料的供给赶不上人口增长的需求，只得向外地购买生产资料，这样就降低了经济社会发展的资金，不利于投资。这种由于人口增长过快而形成的资金分散化”，会阻碍经济的良性循环和持续、稳定的发展。因此，人口增长的速度应与经济增长的速度协调一致，以使劳动力资源的供需一致，从而实现充分就业。

1. 模型的评价与推广

**6.1 模型优点**

本文的模型是在大量的分析之下，包括数据的描述性统计分析与数据可视化，人口与经济的聚集性和空间分布分析以及时间变化下的分析等，对数据的分析较为全面，从而总结出的经济的影响因子是合理的，则构建的模型的可靠性也更高。

除此之外，模型还具有较好的泛化能力，对更加复杂的数据不敏感，模型能较好的处理这些复杂情况。同时，模型的计算时间快，对于大量的数据下能节省更多的时间。

**6.2 模型缺点**

当假设下的几个条件有较大变化时，比如人口较多迁移或者土地面积有较大变化，模型的预测可能会不稳定，求得的结果也可能不准确。

**6.3 模型推广**

本文模型适合各个省市的人口和经济的分析，而不只是针对重庆的区县数据。可以按照本文使用的数据格式，构建其它省市的人口与经济数据，利用本文的模型进行分析和预测。

1. 参考文献

[1］姜启源，谢金星，叶俊.数学模型（第四版）［M］.北京：高等教育出版社，2011.

[2] 贺丹.我国人口与经济发展的关系.http://www.rmzxb.com.cn/c/2018-10-30/2203403.shtml, 访问时间：2019/5/20.

[3] 百度百科.多元回归.https://baike.baidu.com/item/%E5%A4%9A%E5%85%83%E5%9B%9E%E5%BD%92/1796460?fr=aladdin,访问时间：2019/5/19.

[4] CSDN博主goodshot.数据标准化/归一化,https://blog.csdn.net/goodshot/article/details/79488629,访问时间：2019/5/19.

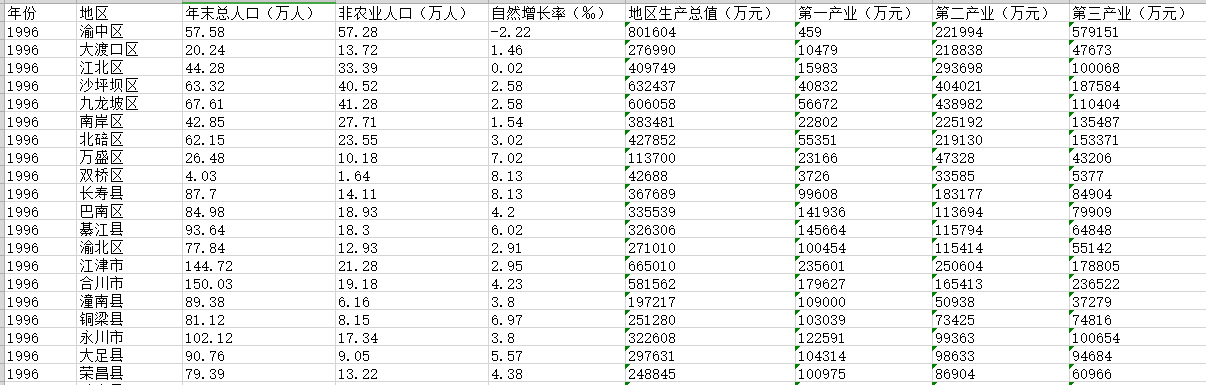
附录

1. **模型代码**

|  |
| --- |
| from sklearn.preprocessing import StandardScaler  from sklearn.metrics import mean\_absolute\_error  from sklearn import linear\_model  import pandas as pd  import numpy as np  # 数据读入  path = "E:\\kaifabu\\MathematicalModeling\\jpn\\k1.csv"  data = pd.read\_csv(path)  # 测试集和训练集构造  train\_X = data.loc[data['年份']!=2017][['年末总人口（万人）', '面积', '密度', '农业人口','非农业人口（万人）']].values  train\_y = (data.loc[data['年份']!=2017]['地区生产总值（万元）']/10000000).values  test\_X = data.loc[data['年份']==2017][['年末总人口（万人）','面积', '密度', '农业人口','非农业人口（万人）']].values  test\_y = (data.loc[data['年份']==2017]['地区生产总值（万元）']/10000000).values  # 数据标准化  scaler = StandardScaler()  train\_X = scaler.fit\_transform(train\_X)  test\_X = scaler.transform(test\_X)  # 模型训练  def train\_model(clf, train\_X, test\_X, train\_y, test\_y):  clf.fit(train\_X, train\_y)  pre = clf.predict(test\_X)  mae = mean\_absolute\_error(test\_y, pre)  print("mae: ", mae)  print("权重: ", clf.coef\_)  print("偏置： ", clf.intercept\_)  # 声明模型  clf = linear\_model.LinearRegression()  # 训练模型  train\_model(clf, train\_X, test\_Xs, train\_y, test\_y) |

1. **数据来源及信息说明**

数据来源于重庆统计年鉴中各年的人口和经济数据，由于获取的数据信息较多，我们根据需要筛选出了使用的数据，部分数据如下图：



1. **可视化部分代码（未展示的在附件中提供）**

|  |
| --- |
| import pandas as pd  import xlrd  import matplotlib.pyplot as plt  import matplotlib  import numpy as np  data = pd.read\_excel('1996-2017.xlsx')  data.head()  columns = ['年末总人口（万人）', '非农业人口（万人）', '自然增长率\n（‰）', '地区生产总值（万元）',  '第一产业（万元）', '第二产业（万元）', '第三产业（万元）']  data['年份'] = data['年份'].astype(int)  for col in columns:  data[col] = data[col].astype(float)  data['rural\_population'] = data['年末总人口（万人）'] - data['非农业人口（万人）']  data['citilization'] = data['非农业人口（万人）'] / data['rural\_population']  rate\_col = ['第一产业（万元）', '第二产业（万元）', '第三产业（万元）']  for rate in rate\_col:  data[rate+'\_rate'] = data[rate] / data['地区生产总值（万元）']  data['该年农村人口占比'] = data['rural\_population'] / data['年末总人口（万人）']  data['该年城市人口占比'] = data['非农业人口（万人）'] / data['年末总人口（万人）']  span = 10  nap = 5  size=40  font1 = {'family' : 'SimHei',  'weight' : 'normal',  'size' : size,  }  for year in range(1996, 1997):  plt.figure(figsize=(50, 25))  label\_list = data.loc[data['年份'] == year, '地区'] # 横坐标刻度显示值  num\_list1 = data.loc[data['年份'] == year, '非农业人口（万人）'] # 纵坐标值1  num\_list2 = data.loc[data['年份'] == year, 'rural\_population'] # 纵坐标值2  x = range(0, 20 \* len(num\_list1), 20)  """  绘制条形图  left:长条形中点横坐标  height:长条形高度  width:长条形宽度，默认值0.8  label:为后面设置legend准备  """  rects1 = plt.bar(x, height=num\_list1, width=span - nap / 2, alpha=0.8, color='#D98B5F', label="非农村人口（万人）")  rects2 = plt.bar([i + span - nap / 2 for i in x], height=num\_list2, width=span - nap / 2, color='#597DBF', label="农村人口（万人）")    plt.ylim(0, 150) # y轴取值范围  plt.ylabel("数量", size=size)  """  设置x轴刻度显示值  参数一：中点坐标  参数二：显示值  """  plt.xticks([index + span for index in x], label\_list, rotation=270, size=size - 10)  plt.yticks(range(0, 150, 25), [i for i in range(0, 150, 25)], size=size)  plt.xlabel("年份", size=23)  plt.title("各地区{}年份人口数量柱形图".format(year), size=size)  plt.legend(prop=font1) # 设置题注  # 编辑文本  for rect in rects1:  height = rect.get\_height()  plt.text(rect.get\_x() + rect.get\_width() / 2, height+1, str(np.round(height, 2)), ha="center", va="bottom", size=size - 10)  for rect in rects2:  height = rect.get\_height()  plt.text(rect.get\_x() + rect.get\_width() / 2, height+1, str(np.round(height, 2)), ha="center", va="bottom", size=size - 10)  plt.grid(False)  plt.show() |