

统计分析视域下我国医疗卫生事业发展分析

杨金梅, 黄雅仕, 曾绮茵, 王枝宁*

韩山师范学院 数学与统计学院, 广东 潮州 521041

DOI:10.61369/ASDS.12192

摘 要 : 基于全国31个地区 1990~2023年医疗卫生机构数量及相关数据, 从时间、空间两个维度对医疗卫生机构数量进行研究并分析其影响因素。剔除与医疗卫生机构数的 Pearson 相关系数值小于0.6的指标, 运用逐步回归方法构建多元线性回归模型, 并基于该模型进而构建 ARIMA 模型, 对 2024至2028 年全国及各地区医疗卫生机构数进行预测。基于预测结果, 给出了对医疗卫生政策制定、资源分配和体系建设的启示, 及可能面临的挑战和应对策略。

关 键 词 : Pearson相关系数; 逐步回归分析; LM-ARIMA 模型; 医疗卫生发展

Analysis of The Development of China's Medical And Health from The Perspective of Statistical Analysis

Yang Jinmei, Huang Yashi, Zeng Qiyin, Wang Zhining*

School of Mathematics and Statistics, Hanshan Normal University, Chaozhou, Guangdong 521041

Abstract: Based on the data of the number of medical and health institutions and related data in 31 regions across the country from 1990 to 2023, we first conducts research on the number of medical and health institutions from both temporal and spatial dimensions and analyzes its influencing factors. Then the indicators with a Pearson correlation coefficient value less than 0.6 with the number of medical and health institutions are excluded. Finally, the stepwise regression method is used to construct a multiple linear regression model, and based on this model, an ARIMA model is further constructed to predict the number of medical and health institutions nationwide and in each region from 2024 to 2028. Based on the prediction results, insights into the formulation of medical and health policies, resource allocation, and system construction are provided, as well as the possible challenges and corresponding strategies.

Keywords: Pearson correlation coefficient; stepwise regression analysis; LM-ARIMA model; development of medical and health services

引言

医疗, 始终是社会各界广泛关注的热点话题。它深度融入人民生活的各个角落, 从日常的健康体检, 到患病时的诊断与治疗, 桩桩件件都紧密关乎百姓的生活质量与幸福指数, 同时也是衡量国家发展水平和综合实力的关键维度之一。近些年来, 随着我国经济水平稳步提升、综合国力日益增强, 医疗卫生事业取得了长足进步, 先进的医疗技术层出不穷, 使得更多患者得以重获健康。然而, 当前我国医疗资源紧张的状况依然不容小觑。城乡之间以及不同地区之间, 医疗发展呈现出极为显著的不均衡态势^[1]。在发达地区的大城市里, 汇聚了大批顶尖的医学专家, 配备了众多前沿的医疗设备; 而偏远地区的医院, 却深陷设备陈旧、人才匮乏的困境。如此一来, 偏远地区群众看病时常常面临路途遥远、挂号困难等难题, “看病难”问题依旧十分突出。因此, 全面、深入地了解我国各地区医疗卫生发展水平, 是合理配置医疗资源、完善医疗制度、健全医疗体系的关键所在, 唯有如此, 才能让医疗服务更加公平、高效地惠及全体人民。

一、数据来源及数据预处理

通过国家统计局官网 (<https://www.stats.gov.cn/>) 收集了 1990~2023年全国31个地区共1088个医疗卫生机构个数信息和

1993~2023年影响全国医疗机构卫生发展的17个重要因素共558个信息 (鉴于2024年全国及各地区医疗卫生机构数量数据目前暂未公布, 本文仅采集截止到2023年的数据)。原始数据的示例见表1和表2。

基金项目: 2024年度广东省本科高校教学质量与教学改革工程项目-校企联合实验室: 数据科学创新创业实验室 (粤教高函〔2024〕30号), 韩山师范学院教学质量与教学改革工程项目 (项目编号: E24112、E24025), 韩山师范学院科研项目 (项目编号: XYN202303), 广东省教育厅普通高校青年创新人才项目 (项目编号: 2023KQNCX041)。

作者简介: 杨金梅、黄雅仕、曾绮茵均系韩山师范学院统计学专业2022级在读本科生;

通讯作者: 王枝宁, 系韩山师范学院数学与统计学院讲师。

表1 1990 ~ 2023年各地区医疗卫生机构个数

年份	中国	北京市	天津市	河北省	山西省	内蒙古	辽宁省	吉林省
1990	1012690	4953	3392	10586	6108	5161	7863	4407
1991	1003769	4970	3397	10647	6121	5172	7736	4369
1992	1001310	4868	3287	10715	5935	5253	7921	4189
1993	1000531	4962	3182	10958	6085	4932	7132	4006
1994	1005271	4958	3182	10274	5996	4918	6849	3920
1995	994409	4955	3182	10266	5922	4915	6239	3891
1996	1078131	6470	4171	5402	3289	5037	6333	3720
1997	1042885	6577	3571	5392	3286	4863	6749	3663
1998	1042885	5722	3190	20071	8592	7517	11710	6321

表2 1990 ~ 2023年影响全国医疗卫生发展的重要因素

年份	卫生技术人员数	卫生人员数	医药制造业企业数	医院床位数	医院数
1993	167748	237879	95	114222	1344
1994	133049	168560	63	79606	1153
1995	250161	321587	209	192455	1400
1996	148569	192218	262	97405	846
1997	191248	247849	109	141016	1325
1998	143670	181016	202	91464	411
1999	255458	471662	491	227556	1721
2000	267128	352530	395	155477	1410
2001	227490	288595	262	153425	1467

对于31个地区医疗卫生机构个数信息，通过观察发现存在缺失值，运用 SPSSPRO 采用 K 近邻插补法对缺失数据进行填充，其计算公式见公式 (4)。根据填充处理后的数据进行标准化处理，其计算公式见公式 (5)，运用 Python 从时间上分析 2014—2023 年全国医疗卫生机构数和从空间上分析各地区医疗卫生机构年均增长率，其计算公式见公式 (6)。对于影响全国医疗机构卫生发展的 17 个重要因素信息，通过观察发现 65 岁及以上人口数和镇数出现缺失值，同理采取 K 近邻插补法对缺失数据进行填充，因此可通过预测这些相关影响因素未来五年的数据，进而预测全国未来五年（即 2024—2028 年）的医疗卫生机构个数。

二、研究方法

（一）预备知识

1. 线性回归模型

线性回归是一种用于建立自变量和因变量之间线性关系的统计模型^[2]。其基本形式为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

其中， y 是因变量， x_i 是自变量， β_i 是回归系数， ε 是误差项。它通过最小化残差平方和（即观察值与预测值之间差值的平方和）来确定回归系数，从而找到最佳的线性拟合关系。

2. ARIMA 模型

ARIMA (p,d,q) 模型^[3]由自回归 (AR)、差分 (I) 和移动平均 (MA) 三部分组成。

(1) 自回归部分 (AR)：p 自回归阶数，它考虑时间序列的过去值对当前值的影响，其形式为：

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2)$$

(2) 差分部分 (I)：由于 AR 和 MA 模型通常要求时间序列是平稳的，而实际的时间序列往往不平稳，差分操作通过计算相邻数据点之间的差异，消除序列中的趋势或周期性成分，使其平稳，d 为差分阶数。

(3) 移动平均部分 (MA)：q 表示移动平均阶数，它考虑误差项的过去值对当前值的影响，其形式为：

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

（二）具体步骤

本文通过 SPSSPRO 和 Python 软件，基于全国 31 个地区 1990 至 2023 年医疗机构卫生数量及其相关影响因素，实现医疗卫生发展现状的分析及预测，具体步骤如下。

Step1：通过国家统计局平台收集数据；

Step2：处理收集的数据；通过 SPSSPRO 软件采用 K 近邻算法^[4]填充缺失值，该公式用于计算两个样本点之间的欧氏距离。在 K 近邻算法中，通过计算未知样本与已知样本之间的距离，找出距离最近的 K 个邻居样本，根据这些邻居样本的特征来预测未知样本的属性。计算公式如下：

$$d_{1,2} = \text{dist}(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (4)$$

并进行数据标准化处理，运用 Z-score 标准化方法，其中是 x 原始数据点，是数据的均值，是数据的标准差。其目的是将数据转换为均值是 0，标准差是 1 的标准正态分布，消除不同特征之间量纲的影响。计算公式如下：

$$Z = \frac{x - \mu}{\sigma} \quad (5)$$

Step3: 通过 KNeighborsClassifier 函数绘制全国医疗机构卫生生数柱形图, 分析变化趋势;

Step4: 计算各地区医疗卫生机构数年增长率^[6], 通过 Map 函数绘制地图热力图, 分析区域差异。该公式原理是假设指标在统计期内以相同增长率逐年增长, 将总体增长转化为平均年度增长水平, 其各部分分别对应增长后的数量、起始数量和时间跨度等。其计算公式如下:

$$CAGR = \left[\left(\frac{\text{末年医疗卫生机构数}}{\text{初年医疗卫生机构数}} \right)^{\frac{1}{\text{年份差}}} - 1 \right] \times 100\% \quad (6)$$

Step5: 通过 Pearson 相关系数分析影响医疗机构卫生发展的重要因素^[6], 该公式用于衡量两个变量之间的线性相关程度, 其取值范围范围是 [-1,1]。当 $r = 1$ 时, 表示两个变量完全正相关; 当 $r = -1$ 时, 表示两个变量完全负相关; 当 $r = 0$ 时, 表示两个变量之间不存在线性相关关系。值越接近 1 或 -1, 线性相关越强; 越接近 0, 线性相关性越弱。其计算公式如下:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

其中, x_i 表示影响因素的值, y_i 表示医疗卫生资源的值。

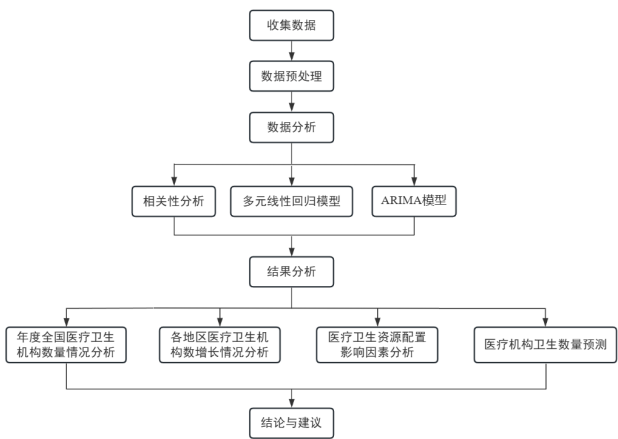
Step6: 通过 sm.OLS 函数拟合多元线性回归模型, 并通过 VIF 检验剔除值大于 10 的变量后重新拟合模型;

Step7: 对重新拟合的模型去除回归系数不显著的变量, 保留显著变量;

Step8: 通过 ARIMA 函数分别对显著变量预测其 2024 年至 2028 年的数值;

Step9: 将预测结果带入回归模型中, 求得目标变量的预测结果。

本文的整体算法框架如图 1 所示。



> 图1 整体算法框架

三、结果与分析

(一) 年度全国医疗卫生机构数量情况分析

将 2014 年至 2023 年的全国医疗卫生机构数量进行统计, 如图

2 所示。



> 图2 2014 年至 2023 年全国医疗卫生机构数柱状图 (单位: 个)

由图 2 可以看出, 从 2014 年到 2023 年, 全国医疗卫生机构的数量总体上呈现增长趋势, 展现出我国医疗卫生事业持续发展、医疗资源不断扩充的良好局面。2016 年机构数量出现短暂小幅度下降, 这可能是局部资源整合等原因导致。2019 年机构数量实现重要跨越, 突破原有规模。此后数量稳定增长, 2023 年达到新高, 这主要得益于国家政策的有力支持、社会对医疗卫生服务需求的增长以及医疗卫生体系的持续完善等。整体上, 这些变化反映出我国在医疗卫生领域持续加大投入和优化资源配置, 以更好地满足民众健康需求。

(二) 各地区医疗卫生机构数增长情况分析

各地区 1990~2023 年医疗卫生机构数的年均增长率数据, 见表 3。

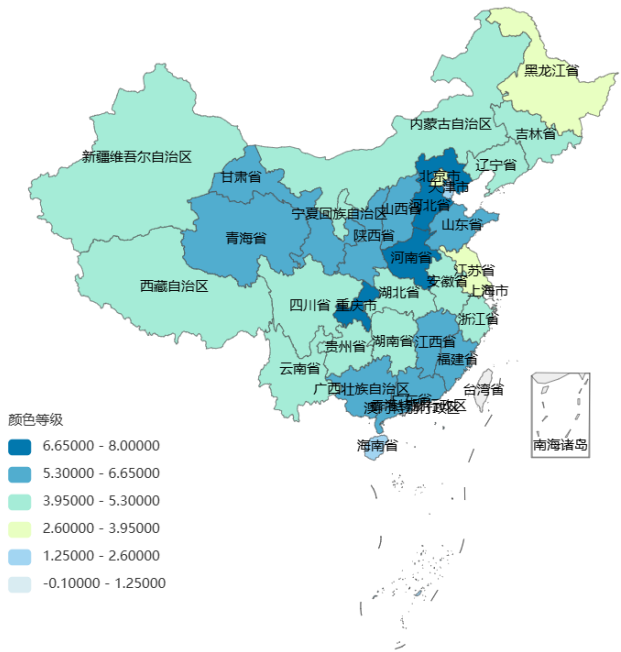
表3 各地区 1990 ~ 2023 年医疗卫生机构数量年均增长率 (单位: %)

地区	年均增长率	地区	年均增长率	地区	年均增长率
广东	6.0728	安徽	4.5007	贵州	4.6072
江苏	3.2428	湖北	4.0233	天津	1.6496
浙江	4.1088	河北	6.8006	海南	1.7779
上海	-0.931	湖南	5.2309	内蒙古	4.9831
山东	6.4924	陕西	5.3031	甘肃	5.6503
重庆	6.6733	广西	5.5942	宁夏	4.4307
北京	2.8495	江西	6.1303	青海	5.4186
福建	5.6542	吉林	5.1806	新疆	4.4174
四川	4.9907	云南	4.528	西藏	5.1533
辽宁	4.5962	黑龙江	2.6822		
河南	7.1616	山西	5.6956		

由表 3 可以看出, 全国各地区的医疗卫生机构数量在增长方面呈现出显著的差异。尽管大部分地区的医疗卫生机构数量都呈现增长趋势, 但不同地区的增长速度存在明显差距。其中, 河南表现尤为突出, 年均增长率位居全国首位, 显示出该地区在医疗卫生基础设施建设方面的强劲发展势头。相比之下, 上海市的年均增长率不仅最低, 而且为负值, 可能与上海市的医疗卫生资源相对饱和、机构优化调整等因素有关。

图 3 直观地展示了不同省份医疗卫生机构数的年均增长率情况。东部沿海地区的增长速度普遍较快, 这与该地区经济发展水平较高、人口密集以及政府对医疗卫生事业的大力支持密切相关。其中河北、河南和重庆的增长速度尤为突出, 河南更是遥遥领先, 彰显当地医疗卫生快速发展和政策成效。然而, 海南和上

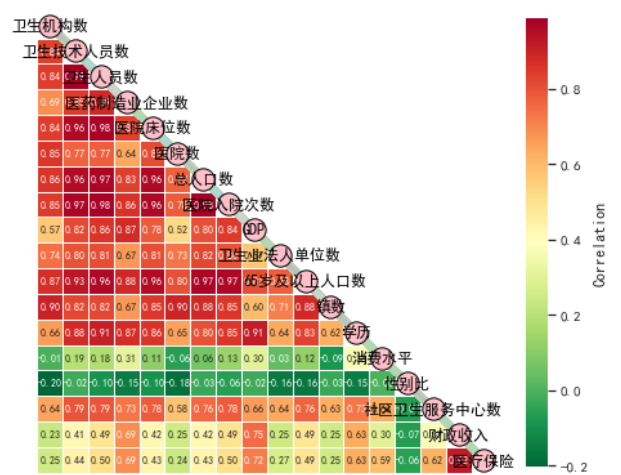
海的增长速度明显较慢，上海更是处于全国范围内最低水平。这可能与上海的医疗卫生资源已经相对完善、机构数量趋于稳定，也可能受到城市发展空间有限、政策调整等因素的影响。总体而言，图3的数据印证了表3中各地区在医疗卫生机构增长方面存在显著差异，且东部沿海地区的增长势头普遍较强，而部分地区的增长速度则相对缓慢的结论。



> 图3 1990年至2023年医疗卫生机构数年均增长率中国热力图

（三）医疗卫生资源配置影响因素分析

从医疗卫生事业中的人力资源、物力资源、财力资源出发，结合社会、人口、财政等宏观指标，共收集17个可能影响医疗卫生机构数量的指标^[7]。根据 Pearson 相关系数来绘制热力图，如图4所示。



> 图4 Pearson相关系数热力图

根据图4，筛选出与医疗卫生机构数的相关系数值大于0.6的12个指标，分别为卫生技术人员数、卫生人员数、医疗制造业企业数、医院床位数、医院数、总人口数、医院入院次数、卫生业法人单位数、65岁及以上人口、镇数、学历、社区卫生服务中心数。

（四）医疗卫生机构数量预测

根据上文筛选出来的影响医疗卫生机构数量的显著指标，建立医疗卫生机构数量与其影响指标之间的多元线性回归模型，并剔除 VIF 值大于10及回归系数不显著的指标。结果如表4所示。

表4 多元线性回归模型分析结果

	coef	P> t
const.	-3314.599	0.000
卫生人员数	0.9178	0.000
65岁及以上人口	-6.626	0.002
镇数	42.564	0.014
拟合优度	0.980	
调整后拟合优度	0.978	

数据标准化后得到的最终模型如式（8）：

$$y = -3314.599 + 09178x_1 - 6.626x_2 + 42.564x_3 \quad (8)$$

其中，是医疗卫生机构数量，是卫生人员数，是65岁及以上人口，是镇数，拟合优度为0.978，说明模型拟合效果比较好。

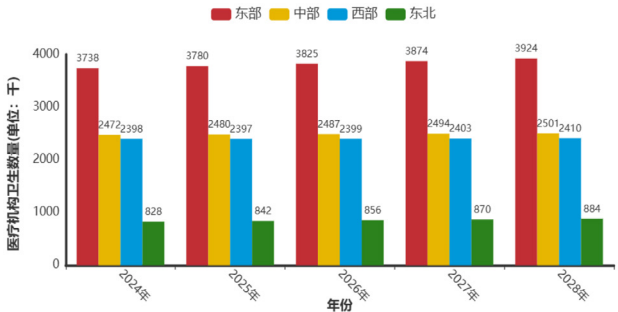
通过 ARIMA 模型分别预测全国和31个地区2024年至2028年的卫生技术人员数量以及镇数，并由多元线性回归模型可以得到2024年至2028年的医疗卫生机构数量的预测结果^[8]，见表5。

表5 2024年至2028年医疗卫生机构数量的预测结果（单位：个）

地区	2024年	2025年	2026年	2027年	2028年
中国	826340	875975	927650	981353	1037077
北京	151123	151121	151115	151111	151106
天津	104612	107901	111191	114480	117770
河北	527810	544439	561050	577661	594258
山西	228664	228643	228595	228564	228523
内蒙古	159294	159328	159422	159470	159494
辽宁	387199	387214	387227	387238	387247
吉林	190400	196504	202608	208713	214817
黑龙江	250649	258535	266408	274287	282163
上海	206597	206599	206594	206593	206589
江苏	697301	697273	697215	697208	697184
浙江	450587	450694	450801	450909	451016
安徽	425487	425631	425767	425863	425909
福建	210398	195529	183703	174298	166819
江西	265219	272634	280052	287469	294887
山东	783984	808061	832138	856215	880292
河南	643790	643679	643567	643456	643345
湖北	435131	435132	435134	435135	435137
湖南	473831	473795	473758	473723	473686
广东	559688	575805	591919	608031	624142
广西	280850	263193	248815	237109	227577
海南	45803	42259	39395	37080	35210
重庆	265566	273328	281093	288855	296618
四川	675845	675849	675852	675855	675858
贵州	213198	213163	213129	213094	213060
云南	242529	242462	242396	242329	242263
西藏	5035	4532	4158	3878	3671
陕西	260061	260069	260075	260080	260084
甘肃	145006	149381	153684	157900	162023
青海	23963	24744	25526	26307	27089
宁夏	30865	32039	33215	34395	35576
新疆	95731	98996	101359	104092	106673

从表5可以看出,2024年全国医疗机构数量为826,340家,预计到2028年将增长至1,037,077家,年均增长率为5.7%,整体增长趋势平稳,表明中国医疗机构卫生资源的供给能力正在稳步提升。然而,不同地区的增长情况存在显著差异。其中,宁夏回族自治区年均增长率3.6%,增长最快,主要得益于政府大力投入、基建与人才推进及经济支撑;天津市年均增长率3.1%,增速居第二,凭借雄厚经济实力,加大医疗投入,完善设施,引入先进技术与经验。相比之下,福建省年均增长率-5.6%,呈负增长,因产业结构调整致医疗投入减少,人口流动改变需求结构,小型机构整合或关停。此外,东部人均医疗机构卫生资源数量显著高于西部地区。以2024年为例,北京市人均医疗机构卫生资源数量处于较高水平,而西藏自治区仅为较低水平,进一步揭示了资源分配的区域不均衡问题。

图5展示了中国不同地区(东部、中部、西部、东北)在2024年至2028年期间的医疗卫生机构数量预测值。各区域呈现不同特点,其中东部地区机构数量最多,且逐年稳步增长,显示其强大的经济实力和对医疗卫生的高度重视;中部和西部地区数量相对东部较少,但也都有缓慢增长趋势,反映出在政策支持下,医疗卫生事业在不断发展;东北地区机构数量最少,虽有增长,但幅度较小,可能受经济转型等因素制约。整体上,各区域医疗卫生机构数量都在上升,体现全国医疗卫生资源在持续扩充。



> 图5 2024年至2028年医疗卫生机构数预测值（单位：个）

四、结论与建议

以全国31个地区1990~2023年医疗卫生机构数及其可能的影响因素为研究对象,使用Python软件和SPSSPRO,利用Pearson相关系数、多元线性回归模型、ARIMA等模型,从全国及各地区医疗卫生机构数量变化趋势、区域差异、影响因素等方面进行分析,预测2024年至2028年全国及各地区医疗卫生机构数量。研究结果表明,我国医疗卫生机构数量整体呈稳步增长趋势,但区域差异显著,东部地区增长较快,西部地区增长较慢。

根据分析结果,结合医疗卫生机构发展特点,提出以下建议:(1)均衡区域资源:政府加大对西部政策扶持与资金投入,通过财政转移、人才引进、基建加强等,推动优质医疗资源向中西部倾斜,缩小区域差距。(2)强化人才培养:加大中西部医疗卫生人才培养,以定向培养、人才引进、提高待遇等吸引人才,提升当地医疗水平^[9]。(3)促进城乡均衡:加强农村医疗卫生体

系建设,建乡镇卫生院、村卫生室,推动城乡医疗资源均衡,保障农村居民医疗服务。(4)推动体系改革:政府持续投入,优化资源配置,制定政策鼓励社会资本进入,促进公私医院协同,提升整体服务水平。(5)应对健康新需求:针对老龄化和慢性病,提前布局,加大老年医疗和慢性病管理投入,发展医养结合。

同时,在发展过程中可能面临以下挑战及应对策略:(1)区域发展不均:加大对西部和东北财政支持,鼓励东部资源向中西部流动,借助远程医疗等提升服务。(2)人才短缺:医学院校扩大招生,提高待遇、改善环境吸引人才到中西部和基层。(3)费用上涨:推动医疗费用控制机制,加强价格监管,推广医保支付改革降低费用^[10]。(4)体制改革难题:优化资源配置,加强医疗质量管理,促进机构协同,提升服务质量和效率。

参考文献

- [1]欧阳红兵,张支南.我国省域医疗卫生资源配置的均等化水平分析[J].中国卫生经济,2016,35(11):36-38.
- [2]陈素梅,刘梦,张雨夏,等.运用线性回归模型预测不同月龄肉用褐牛体重[J].中国牛业科学,2024,50(06):14-27.
- [3]吴喜之,刘苗.应用时间序列分析:R语言陪同(第二版)[M].机械工业出版社,2018.1(2023.6重印)
- [4]吴雅琴,轩兴栋,张玉婷,等.基于KNN算法的室内定位系统的设计与实现[J].物联网技术,2022,12(12):35-37+42.
- [5]邓丽珠,钟健文.某眼科医院2018年~2020年降眼压药物的使用分析[J].中国处方药,2022,20(03):60-62.
- [6]王金炜,单雨涵,单保海.基于皮尔逊相关系数的短道速滑竞赛数据分析[J].冰雪运动,2023,45(04):9-12+17.
- [7]李彤,宋跃晋,周仪,等.基于多元线性回归模型的广东省卫生总费用的影响因素分析[J].医学与社会,2020,33(04):66-69+83.
- [8]《我国橡胶产业发展技术经济评价》课题组,雷仲敏,邵晓燕.基于回归分析的国内天然橡胶需求预测分析[J].青岛科技大学学报(社会科学版),2012,28(03):44-49.
- [9]陆银峰.医疗人才队伍建设与可持续发展路径[J].四川劳动保障,2025,(02):113-114.
- [10]张世冉,张雯雯,程方安,等.基于CiteSpace的国内医疗卫生费用研究现状及热点分析[J].大众科技,2023,25(04):155-159.