

基于组合模型预测短期甘肃省医护人才需求*

徐瑞璞^{1,2} 钱国宏² 路杰³ 王晓辉^{1△}

【摘要】目的 分析甘肃省 2011–2021 年执业(助理)医师和注册护士(以下简称医护)现状,预测 2022–2025 年医护人员数量,判断未来医护发展情况,评估其是否达到国家规划既定目标,为医护数量和质量发展方向提供建议。**方法** 利用最小二乘法 and 灰色预测法并联选择最优方法,与 BP 神经网络法串联,形成用于短中期预测的组合模型。**结果** 经测算,2022–2025 年甘肃省执业(助理)医师数将达到 78325 人、78120 人、77469 人、76507 人;2022–2025 年甘肃省注册护士数将达到 97216 人、96990 人、96265 人、95240 人;2025 年甘肃省医护数将符合并分别超出国家标准的 5% 和 10%。经分析,2021 年年底,甘肃省 14 个市(州)医护数达到全国平均水平,分别占 50% 和 57%,86 个区(县)医护数达到全国平均水平,分别占 34% 和 37%。**结论** 甘肃省基层医疗服务队伍数量、质量和结构方面仍需提升。最小二乘法和 BP 神经网络组合模型预测卫生人才数量可行且较稳健。

【关键词】 人才需求 灰色关联分析法 最小二乘法 BP 神经网络模型

【中图分类号】 R192 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.02.030

随着甘肃省人口的增长、健康需求的提升以及医疗技术的进步,卫生人力资源的供需矛盾日益凸显。为了进一步做好卫生人力资源规划,需对未来卫生人力资源的需求数、供给数和供需关系进行预测,比较卫生人力资源需求数和供给数是否平衡,计算与国家规划中相应时间节点目标的差距。

1987 年以来,国内外采取不同方法对卫生人力资源进行研究^[1]。灰色模型应用广泛、拟合度高,但是对波动较大的数据拟合较差^[2];时间序列模型无需大规模调查,却只考虑到时间因素没有考虑其他指标^[3];健康需求法和卫生需求法所需资料少、计算简单,然而很少用于具体领域或地区的预测^[4–5];组合模型可以克服单一模型缺陷、减少误差、提高精确度^[6–8]。近年来,研究多使用灰色模型、时间序列模型等经典模型与 BP 神经网络模型组合,预测各类卫生

人力资源数^[9–13]。

本研究拟采用灰色预测法和最小二乘法中拟合程度较高的方法,与 BP 神经网络模型组合,预测与医护数相关的影响因素数量和未来医师和护士的数量,与国家规划中要求的相应时间节点目标做比较。同时,计算医护数与全国平均水平以及国家规划相应时间节点目标数相比较的缺口,结合近十年医护的发展趋势分析,为卫生人力资源规划提供建议。

资料与方法

1. 数据来源

从《甘肃省卫生健康统计年鉴》中提取 2011–2021 年甘肃省各市(州)、区(县)执业(助理)医师、注册护士和常住人口数等,见表 1。

表 1 2011–2021 年甘肃省卫生资源数据

年份 (年)	门急诊人次 (万次)	卫生总费用 (亿元)	诊疗人次 (万次)	医院入院人数 (万人)	人口数 (万人)	人口出生率 (‰)	人口死亡率 (‰)	医疗卫生机构 床位数(万张)	执业(助理) 医师数(人)	注册护士数 (人)
2011	9689.13	393.60	10578.28	233.56	2564.19	12.08	6.03	10.11	41121	33713
2012	10835.86	444.72	11781.67	288.68	2577.55	12.11	6.05	11.20	43302	37212
2013	11192.67	518.21	12066.74	249.35	2582.18	12.16	6.08	11.61	43442	40668
2014	11493.90	569.75	12330.11	271.87	2590.78	12.21	6.11	12.24	47791	45196
2015	11736.32	654.07	12569.80	283.08	2599.55	12.36	6.15	12.70	49663	47832
2016	12317.40	761.20	13150.60	322.15	2609.95	12.18	6.18	13.66	53195	50653
2017	12643.45	812.70	13527.40	348.47	2625.71	12.54	6.52	14.73	56398	58682
2018	12332.93	905.32	13208.10	386.00	2637.26	11.07	6.65	16.27	59370	63807
2019	12017.66	935.01	12732.60	416.44	2647.43	10.60	6.75	17.26	62891	79531
2020	10958.16	1015.31	11591.10	362.14	2501.98	10.55	7.91	17.79	66396	85264
2021	10771.50	1110.71	11458.10	361.90	2490.02	9.68	8.26	18.32	70673	91723

研究方法

选择灰色关联分析法找到与医护数强相关影响因素,用最小二乘法和灰色预测法对医护数逐一预测,选用两种方法中预测数据与实际数据更相似的一种方法

* 基金项目:甘肃省卫生健康行业科研项目(GSWSKY2020–27)

1. 兰州大学公共卫生学院(730000)

2. 甘肃省卫生健康委员会人才服务中心

3. 甘肃省卫生健康委员会统计信息中心

△通信作者:王晓辉, E-mail: wangxiaohui@lzu.edu.cn

预测 2022-2025 年影响因素。最后将 2011-2021 年影响因素和医护数放入 BP 神经网络模型中训练出误差较小的模型,带入 2022-2025 年影响因素预测模型来预测医护人员数。

(1)灰色关联分析法:影响医护人数的因素是复杂的。根据沈钰^[14]和张志明^[15]的研究综合考虑选用经济、社会、人口指标,结合《卫生健康统计年鉴》选择卫生设施、卫生经费、医疗服务、人民健康水平、人口指标作为一级指标,医疗卫生机构床位数、卫生总费用、医疗诊疗人次数、门急诊人次数、医院出院人数、人口出生率、人口死亡率、人口自然增长率、人口数作为二级指标,一般高于 70%为重要因素,处于 50%~70%之间的为比较重要因素,其余的为不重要因素^[16]。经过灰色关联分析,选择相关性大于 70%的影响因素作为医护数的重要影响因素,见表 2。

表 2 医护人才预测指标体系

一级指标	二级指标
卫生设施	医疗卫生机构床位数
卫生经费	卫生总费用
医疗服务	医疗诊疗人次数 门急诊人次数 医院出院人数
人民健康水平	人口出生率 人口死亡率 人口自然增长率
人口指标	人口数

(2)最小二乘法:通过数据组最小化误差平方和,找到与其规律最匹配的函数^[17]。函数表达式为:

$$erf(x) = \frac{2}{\pi} \int_0^x e^{-t^2} dt$$

灰色预测法:根据各因素间有不确定因素且已知部分信息来预测数据^[18]。函数表达式为:

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

通过 Pearson 相关系数衡量两个数组拟合程度大小。相关分析可以研究几组数据集之间有没有关系以及关系是否密切等情况^[19-20]。

(3)BP 神经网络:正向传播从输入端开始,沿着箭头指向乘以相关权重系数加和,得到的结果作为输入用激活函数再次进行归一化运算,将得到结果作为此节点输出和下一节点的输入,计算到最终输出;反向传播在得到最终输出后计算与实际输出的误差,多次迭代对权重进行梯度下降法的调节,每次调节后均进行测试,循环运行到测试结果满意则模型建成,直接应用到预测中。模型共有 3 层:输入层、输出层节点为影响因素个数和预测数个数,隐含层一般模型中都选一

层,既可满足需要,又大大降低了运算的时间复杂度,节点数为 $\sqrt{\text{输入层节点数} + \text{输出层节点数} + a}$ ($1 < a < 10$),通过对 a 的调试结果分析隐含层节点数为 13。其他参数根据文献和超参数寻优调试后,设置学习速率为 0.01、训练目标最小误差为 0.00001、训练次数为 1000 次。隐藏层、输出层归一化选用激活函数 logsig 和 purelin,训练函数选用能最快收敛的 trainlm 函数^[21-23]。

研究结果

1.甘肃省医护现状分析

由历史数据可知,甘肃省医护数在 2011-2021 年期间呈逐年上升趋势,经过与全国平均水平的对比,执业(助理)医师逐年缩小差距,注册护士自 2018 年开始极速接近全国平均水平,2020 年后超过全国平均水平,数据见表 3。

表 3 2011-2021 年甘肃省与全国平均医护水平对比(人)

年份 (年)	每千人口执业(助理)医师			每千人口注册护士		
	甘肃	全国	需求	甘肃	全国	需求
2011	1.60	1.82	5641.22	1.32	1.66	8718.25
2012	1.68	1.94	6701.63	1.44	1.85	10567.96
2013	1.74	2.04	7746.54	1.59	2.04	11619.81
2014	1.84	2.12	7254.18	1.74	2.20	11917.59
2015	1.91	2.22	8058.61	1.84	2.37	13777.62
2016	2.04	2.31	7046.87	1.94	2.54	15659.70
2017	2.15	2.44	7614.56	2.23	2.74	13391.12
2018	2.25	2.59	8966.68	2.42	2.94	13713.75
2019	2.38	2.77	10324.98	3.00	3.18	4765.37
2020	2.65	2.90	6254.95	3.41	3.34	-1751.39
2021	2.84	3.04	4980.04	3.68	3.56	-2988.02

截至 2021 年,甘肃省执业(助理)医师在 14 个市(州)中,达到甘肃省和全国平均水平的分别占 57%和 50%,86 个区(县)中达到甘肃省和全国平均水平的分别占 37%和 34%;注册护士在 14 个市(州)中,达到甘肃省和全国平均水平的分别占 50%和 57%,86 个区(县)中达到甘肃省和全国平均水平的分别占 35%和 37%,见表 4。

表 4 2021 年 14 市(州)、86 区(县)医护达到平均水平数量(个)

类别	市(州)		区(县)	
	甘肃省	全国	甘肃省	全国
执业(助理)医师	8	7	32	29
注册护士	7	8	30	32

2.各模型预测分析结果

选取 2011-2021 年甘肃省医疗卫生机构床位数,卫生总费用,医疗诊疗人次数,门急诊人次数,医院出院人数,出生率,死亡率,人口数 8 个与医护数相关性系数高于 70%的指标作为影响因素,见表 5。

表 5 医护数与 9 个指标的灰色关联分析 (%)

类别	医疗卫生机构床位数	卫生总费用	医疗诊疗人次	门急诊人次	医院出院人数	人口出生率	人口死亡率	人口自然增长率	人口数
医生	94	77	80	81	87	72	84	60	77
护士	87	89	73	74	86	68	76	60	71

执业医师(助理)数与最小二乘法法和灰色预测法 Pearson 相关系数值为 0.998 和 0.998,并且存在显著性,注册护士数与最小二乘法法和灰色预测法 Pearson 相关系数值为 0.994 和 0.990,并且存在显著性,见表 6。

表 6 医护数与两种预测法 Pearson 相关性

方法	参数	执业(助理)医师数	注册护士数
最小二乘法	相关系数	0.998	0.994
	P 值	<0.001	<0.001
灰色预测法	相关系数	0.998	0.990
	P 值	<0.001	<0.001

最终影响因素选用最小二乘法预测影响因素,见表 7。

表 7 基于最小二乘法影响因素预测数

年份(年)	门急诊人次(万次)	卫生总费用(亿元)	诊疗人次(万次)	医院入院人数(万人)	人口数(万人)	人口出生率(‰)	人口死亡率(‰)	医疗卫生机构床位数(万张)
2022	9897.2	1163.3	10496.0	394.0639	2460.60	8.84	8.81	19.84
2023	8919.7	1233.8	9456.8	399.4675	2405.30	7.91	9.51	20.95
2024	7776.5	1304.2	8245.5	403.2821	2342.00	6.88	10.28	22.11
2025	6467.7	1374.4	6862.6	405.5079	2270.70	5.74	11.13	23.31

表 8 2022-2025 年基于 BP 神经网络模型预测医护数(人)

年份(年)	执业(助理)医师	注册护士	每千人口执业(助理)医师	每千人口注册护士
2022	78325	97216	3.18	3.95
2023	78120	96990	3.25	4.03
2024	77469	96265	3.31	4.11
2025	76507	95240	3.37	4.19

通过预测 2025 年甘肃省人口将达到 2270.70 万人,每千人口医护数分别达到 3.37 人、4.19 人,与《“十四五”卫生健康人才发展规划》提出的 3.2 人和 3.8 人相比,符合规划要求,并分别超出 5%和 10%。

讨论

本研究以甘肃省医师数和护士数的多种影响因素作为前提条件,运用最小二乘法法和灰色模型对影响因素进行预测,比较预测结果与原始数据,最终选择拟合度较高的最小二乘法预测影响因素,然后将预测出的影响因素代入神经网络模型预测医师数和护士数。最小二乘法拟合度更高的原因可能是:模型通过最小化误差的平方和,使每个预测值都更接近真实值^[24]。组合模型应用难度较高、专业性较强,同时,经过各种模型组合精确度会提高,还是会将各种模型的误差叠加导致精确度降低,是否还有其他更好的模型或方法用于预测,未来仍需继续探索。

卫生人力资源包括卫生专业技术人员、行政人员、

3. 基于组合模型的甘肃省医护数预测

BP 神经网络模型构建一般是随机的从样本中选取 70%的数据作为训练数据,以 2011-2018 年数据作为训练集,2019-2021 年数据作为验证集,2022-2025 年数据作为预测集,预测出 2022-2025 年甘肃省医护数,结果见表 8。

平均绝对百分误差 MAPE 不超过 20%且越趋于 0 越好,相关系数 R^2 越趋于 1 越好,相对误差 RE 不大于 $\pm 0.5\%$ 满足高精度要求。BP 神经网络预测值与实际值对比误差分析:平均绝对百分误差 MAPE 为 1.30%,相关系数 R^2 为 93.34%,相对误差 RE 为 0.07,均在合理范围之内。

后勤人员等,卫生技术人员包括医师、护士、技师、药师等^[25]。卫生人力资源管理全过程包括卫生人力的规划、培训和使用三个部分,最终使卫生人力在数量、质量、结构和分布方面满足居民要求^[26]。本研究主要以医师和护士为例研究卫生人力资源数量特征,结果表明:甘肃省以医师和护士的数量在相应时间节点能按规划达到相应要求^[27],但是暂未讨论其质量、结构和分布。未来需要通过分析卫生人员的年龄、学历、职称、分布等特征,判断配置是否公平,更好地了解卫生人员的现况。同时,还应重点研究卫生人力资源的聘任、调配、培训、奖惩等,注重以需求为导向,配置和培养卫生人力资源。

参考文献

- [1] 张瑜洁,王健,王辛,等. 基于文献计量法的卫生人力预测模型研究现状分析中国医院, 2022, 26(2):43-46.
- [2] 陈彤,刘玉琢,陈倩,等. 基于灰色预测 GM(1,1)模型山东省卫生人力资源变化趋势研究. 卫生软科学, 2023, 37(3):59-62.
- [3] 丁海峰,李立清. 基于 ARIMA 模型的我国长三角地区卫生总费用时间序列预测分析研究中国医疗管理科学, 2022, 12(2):4-10.
- [4] Bayked EM, Toleha HN, Kebede SZ, et al. The impact of community-based health insurance on universal health coverage in Ethiopia: a systematic review and meta-analysis. Glob Health Action, 2023, 16(1): 2189764.
- [5] Bazyar M, Noori Hekmat S, Rafei S, et al. Supply-and-demand projections for the health workforce at a provincial level from 2015 to

- 2025 in Ilam, Iran. *Proceedings of Singapore Healthcare*, 2020, 30(1):18-27.
- [6] 朱泉同,高山. 基于组合预测模型的江苏省卫生人力资源需求预测探讨. *中国卫生统计*, 2020, 37(6):862-865.
- [7] 邵佳馨,杜敏,刘瑞华. 基于集聚度和秩和比法的广东省护理人员人力资源研究. *护理学报*, 2023, 30(4):33-37.
- [8] 薛宇,王长青. ARIMA-灰色耦合模型下区域医疗卫生资源供给预测. *中国卫生统计*, 2019, 36(4):604-608.
- [9] Bairui T, Cheng C, Lubo L, et al. Analysis of Urban Talent Demand Model Based on Neural Network. 2019 11th International Conference on Measuring Technology and Mechatronics Automation (IC-MTMA), 2019.
- [10] Li H, Wang Q, Liu J, et al. A Prediction Model of Human Resources Recruitment Demand Based on Convolutional Collaborative BP Neural Network. *Comput Intell Neurosci*, 2022;3620312.
- [11] 沈钰. 区域技术技能人才需求预测研究. 山西大学, 2020.
- [12] 锥朝君. 新医改以来甘肃省卫生人力资源发展状况及需求预测分析. 兰州大学, 2019.
- [13] 毛瑛,井朋朋,吴静娴,等. 我国卫生人力资源的组合预测模型构建及应用. *中国卫生经济*, 2015, 34(5):21-24.
- [14] 沈钰,韩永强. 技术技能人才需求预测模型及其检验:基于BP神经网络视角. *当代职业教育*, 2020,104(2):72-78.
- [15] 张志明,汪荟萃. 基于灰色BP神经网络的员工需求预测. *重庆文理学院学报(社会科学版)*, 2020, 39(2):33-42.
- [16] 孙林凯,金家善,耿俊豹. 基于修正邓氏灰色关联度的设备费用影响因素分析. *数学的实践与认识*, 2012, 42(8):140-145.
- [17] 王珍萍. 最小二乘法在商业活动中的应用及其在MATLAB的实现. *广西科技师范学院学报*, 2018, 33(1):131-133+138.
- [18] 邓聚龙. 灰色预测模型GM(1,1)的三种性质:灰色预测控制的优化结构与优化信息量问题. *华中工学院学报*, 1987(5):1-6.
- [19] Hauke J, Kossowski T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae*, 2011, 30(2):87-93.
- [20] Arndt S, Turvey C, Andreassen NC. Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *J Psychiatr Res*, 1999, 33(2):97-104.
- [21] 张怡君,左颖婷,刘近春,等. GA-BP与BP神经网络在医学研究中的应用与比较. *中国卫生统计*, 2018, 35(2):239-241+245.
- [22] 王伟,储泽楠,张修太,等. 一种改进BP神经网络的K-means算法. *电子器件*, 2020, 43(2):380-385.
- [23] 罗正亮,潘虹,赵雷,等. 基于PCA-GA-BP神经网络的水电机组状态异常判别方法. *排灌机械工程学报*, 2022, 40(4):372-377+403.
- [24] 孔刚鹏,周焯博,刘洋,等. 基于最小二乘法线性回归的火炮身管寿命预测. *兵工自动化*, 2024, 43(2):1-3+22.
- [25] 郝模,郭岩,谢阳. *中华医学百科全书·卫生事业管理学*. 北京:中国协和医科大学出版社, 2019.
- [26] 梁万年,饶克勤,王亚东. *卫生事业管理学*. 北京:人民卫生出版社, 2012.
- [27] 国家卫生健康委印发《“十四五”卫生健康人才发展规划》. *上海护理*, 2022, 22(9):5.

(责任编辑:张悦)

(上接第286页)

- [5] 王旭莹. 肿瘤标志物的临床检测方法及其应用. *中国实用医药*, 2007, 2(7):105-106.
- [6] Nguyen AH, Miller EJ, Wichman CS, et al. Diagnostic value of tumor antigens in malignant pleural effusion; a meta-analysis. *Translational Research: the Journal of Laboratory and Clinical Medicine*, 2015, 166(5):432-439.
- [7] 金宸,高志洪. CEA、CA125、CA153联合检测在恶性胸腔积液中的诊断价值. *现代医学与健康研究(电子版)*, 2021, 5(5):115-117.
- [8] 郑达礼. 胸腔积液肿瘤标志物联合检测的临床意义探讨. *临床和实验医学杂志*, 2013, 12(4):287-288.
- [9] 夏一帆,陈阳育,施焕中,等. 基于统计学习方法的肿瘤标志物联合诊断研究. *中国卫生统计*, 2021, 38(6):860-863.
- [10] 陈阳育,徐莉莉. 血清及胸腔积液中四种肿瘤标志物联合应用对良恶性肿瘤鉴别诊断价值的评估. *中华肿瘤防治杂志*, 2021, 28(3):212-222.
- [11] Cui H, Zhong W. A distribution-free test of independence based on mean variance index. *Comput Stat Data Anal*, 2019, 139, 117-133.
- [12] Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves; a nonparametric approach. *Biometrics*, 1988, 44(3):837-845.
- [13] Zou H. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 2006, 101(476):1418-1429.
- [14] Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 2010, 38(2):894-942.
- [15] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Am Stat Assoc*, 2001, 96(456):1348-1360.
- [16] 侯艳,谢宏宇,张晓凤,等. 高维组学数据的变量筛选方法及其应用. *中国卫生统计*, 2016, 33(3):521-526.
- [17] Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. *Nips*, 2017.
- [18] 林永志,梁贤明. 4种肿瘤标志物在肺癌中的诊断价值. *检验医学与临床*, 2015, 12(24):3724-3726.
- [19] 孙美琪,张薇,郝倩文,等. 肿瘤标志物对良恶性胸腔积液鉴别的研究进展. *实用肿瘤杂志*, 2020, 35(3):278-283.
- [20] 王赞,张斯为,高泉,等. 肿瘤标志物CEA、CA153、SCC、Cy211联合检测对胸腔积液鉴别诊断的意义. *肿瘤*, 2002(4):322-324.
- [21] 佟威威,佟广辉,王婧,等. Cyfra21-1、NSE、SCCA和CRP在肺癌诊断中的应用. *中国免疫学杂志*, 2015, 31(3):396-400.

(责任编辑:郭海强)