

- [7] 赵笑云,张敏,任铭威,等. 大学生健康素养现状及影响因素分析[J]. 卫生职业教育, 2020, 38(22):52-54.
- [8] 程靖,刘淑聪,郭梦娜,等. 杭州市大学生健康自我管理能力及影响因素[J]. 中国学校卫生, 2018, 39(9):1310-1313.
- [9] 高磊,程建伟,辛显营. 高职院校大学生体质与心理健康的现状[J]. 中国健康心理学杂志, 2020, 28(5):759-762.
- [10] 蒋忆遥,周驰凯,李晓,等. 生活方式对大学生 BMI 的影响研究[J]. 当代护士(下旬刊), 2020, 27(6):22-26.
- [11] 杨晓卿,杨秀丽,张莉华,等. 不良生活方式对高校大学生动态血压的影响分析[J]. 内科, 2017, 12(5):705-707.
- [12] 陆涵钰,徒文静,彭文雯,等. 高校大学生健康行为状况和中医体质的关系研究[J]. 医学信息, 2020, 33(10):81-84.
- [13] 熊霖,万飞,林海凤,等. 重庆 950 名高职高专学生中医体质调查研究[J]. 中医药导报, 2019, 25(11):35-37.
- [14] 肖全红. 大学生生活方式与中医体质类型对应分析[J]. 亚太传统医药, 2018, 14(9):88-91.
- [15] 刘雪凯. 不同中医体质大学生生活质量调查[J]. 贵州中医药大学学报, 2020, 42(1):99-102.
- [16] 钟小文,黄华聪,任小红,等. 1 034 例大学生中医体质成因分析及调养方法[J]. 中国民间医药, 2019, 28(17):8-10.
- [17] 田春颖,黄洋,孙璇,等. 大学生中医体质调查研究现状系统评价[J]. 天津中医药大学学报, 2020, 39(6):673-678.
- [18] 张娣. 江苏省 2 所高校大学生“隐性饥饿”现状及影响因素调查[J]. 中国健康教育, 2019, 35(7):619-622.

(收稿日期:2022-08-08 修回日期:2023-04-02)

• 教学探索 •

Python 在医学文献检索与统计学课程教学实践中的应用

王家武,姜 庆[△]

(重庆医科大学附属第二医院泌尿外科,重庆 400010)

[摘要] 随着大数据与人工智能时代的到来,数据分析的重要性日益凸显,渗透在社会各行各业,已成为临床医学学生不可或缺的技术与技能。只有将临床医学教学特点和大数据时代特征有机融合,不断深化适应医学大数据的趋势,才能更好地培养适应新时代的医学人才。该文通过实际案例分析了适合应用 Python 语言的 3 种情景,包括 PubMed 网站下载文献、整理与分析基因表达数据、基因表达相关性及其结果可视化,探讨了 Python 在文献检索与统计学课程教学实践中的应用。通过分析,将 Python 语言应用于文献检索与统计学课程教学实践能够大大提高学习效率,激发医学生学习兴趣,提高教学成效。

[关键词] 医学生; 大数据; Python; 文献检索; 统计分析

DOI:10.3969/j.issn.1009-5519.2023.09.032

中图分类号:G642

文章编号:1009-5519(2023)09-1590-06

文献标识码:C

随着我国医疗科学的快速发展,人工智能与大数据时代的到来,数据分析的影响及重要性日益凸显,已渗透在社会各行各业^[1-3]。在大数据时代,医学数据具有结构多样、数据量大、更新及增长迅速及应用价值高等特点,迫切需要对现有数据进行有效挖掘并发现其潜在价值^[4]。因此,新时代的高水平医学人才不仅要有扎实过硬的基本临床技能,还必须具备较好的科研创新能力^[5-6],这就要求其能顺应大数据、人工

智能在医学专业发展的趋势,具备较好的文献检索及数据分析能力,才能适应当前医学发展的需要。当前,如何与大数据时代深度融合,在数以万计的研究中寻找与自己研究相关领域的文献,并结合临床问题,从中找到科研方向及设计思路,并对实验数据结果进行科学统计分析,提高效率,是摆在研究者面前的一大难题^[7]。

Python 语言是一种适用范围广泛、跨平台的高级

[△] 通信作者, E-mail:jq001002@sina.com。

计算机程序设计语言,是人工智能与数据分析领域的第一语言,由于其功能强大、编写简单、支持中文编辑,且其拥有丰富的内置库,直接调用即可实现技术先进、功能强大的程序,因此,Python 已广泛应用于医学研究、数据分析、图像处理等领域^[8]。数据的收集、处理及分析、数据结果的可视化及数据建模均可在 Python 中找到相应的库来处理。学习应用 Python 进行数据收集及数据分析,主要就是学习 Python 第三方库的使用,例如 Requests 库是一个简洁的能够简单地处理 HTTP 请求的第三方库,支持非常丰富的链接访问功能;Matplotlib 库具有丰富的绘图功能,能够把数据可视化,将其更好地呈现出来;Numpy 库和 Pandas 库可以高效地处理数据、生成统计数据、聚合数据等,它们为 Python 提供了数据快速收集、分析及可视化等重要功能。目前,国内越来越多的医学院校都将 Python 程序设计课程设置为必修课^[9-11]。本文通过实际案例探讨 Python 语言在医学文献检索及统计学课程实践教学中的应用。

1 应用 Python 的情景及示例

将 Python 应用于医学文献检索及统计学课程实践教学,打破了传统的计算机搜索及统计软件的枯燥学习,可以较好地激发学生学习课程的积极性,提高学习效率及教学效果。下面利用一些案例展示出 Py-

thon 在医学文献检索及统计学课程实践相关问题中的高效性及便利性。

1.1 案例 1 应用 Python 语言爬虫技术在 PubMed 网站批量下载“renal cancer”与“TCGA”相关的文献。下载文献具体算法流程如下:(1)导入“requests”、“lxml”及“re”模块。(2)打开 PubMed 网站,搜索关键词:“renal cancer”与“TCGA”,得到搜索结果地址“https://pubmed.ncbi.nlm.nih.gov/?term=renal%20cancer%20and%20tcga&page=1”,然后查看并分析网站源代码。(3)使用“requests”获取网页数据:data,再使用“xpath”获取搜索结果总页数:total_pages。代码算法见表 1。(4)使用“xpath”获取文献在线网址:address。(5)判断可下载文献,并获取其 pmid 储存于列表 list3 及在线网址 address[1]储存于列表 list1。算法见表 1。(6)根据(5)中的 pmid 拼接得到文献所在网址 url_2。代码见表 1。(7)根据(6)中的 url_2,使用“requests”获取网页数据:data_2,再使用“xpath”获取文献下载地址 PDF_download。代码见表 1。(8)根据(7)得到的下载地址 PDF_download 下载文献 article。代码见表 1。输出结果见表 2。(9)将下载文献 article 命名为 pmid.pdf,并根据保存路径 path 保存于本地文件夹。代码见表 1。

表 1 案例 1 下载文献算法流程相关代码

流程	相关代码
(3)	pages=data.xpath('//label[@class="of-total-pages"]/text()') total_pages=re.search('\d+',pages[1])
(5)	if address! =[] and address[1]! = '#':list3.append(pmid) list1.append(address[1])
(6)	url_2='https://pubmed.ncbi.nlm.nih.gov/'+pmid+'/'
(7)	download=data_2.xpath('//ul[@class="pmc-sidebar__formats"]//li[@class="pdf-link other_item"]//@href') PDF_download='https://www.ncbi.nlm.nih.gov/'+download
(8)	article=requests.get(url=PDF_download,headers=headers).content
(9)	path='本地文件夹'+pmid+'.pdf' with open(path,'wb') as study: study.write(article)

表 2 可下载文献的 pmid 及下载地址

pmid	PDF_download
33402128	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7786469/pdf/12885_2020_Article_7726.pdf
34384473	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8359575/pdf/13046_2021_Article_2026.pdf
31200666	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6567474/pdf/12885_2019_Article_5768.pdf

续表 2 可下载文献的 pmid 及下载地址

pmid	PDF_download
31493764	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6756904/pdf/aging-11-102233.pdf
30877265	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6461179/pdf/aging-11-101856.pdf
32234847	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7174079/pdf/jitc-2020-000552.pdf
32396851	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7335756/pdf/main.pdf
32003756	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7053637/pdf/aging-12-102715.pdf

在信息爆炸的大数据时代,医学生需要具备通过简单的计算机编写程序下载互联网上特定目标数据的能力。本研究中,运行上述代码仅下载第一页上的文献,共下载文献 8 篇,用时约 4 min 37 s。事实上 PubMed 上搜索“renal cancer”与“TCGA”相关的文献有 1 320 篇,共 132 页,应用 Python 则只需运行几行代码,即可自动下载所有文献。而传统方法通过 PubMed、万方、知网等数据库进行检索、批量下载相关文献,需要反复选择、点击,操作繁复,且耗时较多,大大降低了学习工作效率。Python 语言编写程序可以通过 Requests 库、Beautiful Soup 库及正则表达式进行网络数据的获取、解析,则可快速识别并下载文献,大大节约人力、时间成本,提高下载效率。此外,医学数据大多为 csv、tsv 等格式,且数量庞大,通过 Python 语言编程也可以准确、有效地从互联网上爬取

并直接读取数据进行统计分析。

1.2 案例 2 应用 python 提取并整理分析 612 例肾癌患者基因表达数据(共有 612 个文件夹包含 612 个文本文件,每个文件分别包含 1 例患者的 60 660 个基因表达相关信息)。数据提取整理分析具体算法流程如下:(1)导入“pandas”“os”模块。(2)使用“pandas”读取文件名数据,储存于 data。(3)根据(2)中文件名依次读取 614 例肾癌患者的基因表达数据“data1”、基因名“ID = data1[‘gene_name’]”及基因表达信息“df1 = data1[‘fpkm_unstranded’]”。(4)采用 for 循环读取 614 例肾癌患者的基因表达数据并储存于列表 list1。算法代码见表 3。(5)将 list1 中的数据整合储存于 result。输出结果见表 4。(6)提取处重复数据“repeat”并删除。代码见表 3。(7)使用“describe ()”函数进行统计分析。输出结果见表 5。

表 3 数据提取整理分析算法流程相关代码

流程	相关代码
(4)	<pre>for i in data[‘filename’]: data1 = pd.read_csv(path2, header=1, sep=‘\s+’) df1 = data1[‘fpkm_unstranded’] list1.append(df1) result = pd.concat(list1, axis=1, ignore_index=True, join=‘inner’)</pre>
(6)	<pre>repeat = result.duplicated(subset=‘gene_name’) result.dropna(axis=0, how=‘any’, inplace=True)。</pre>

表 4 整理数据结果

基因	0	1	2	...	610	611	612
TSPAN	13.636 6	13.333 8	40.139 9	...	11.304 8	7.730 3	9.506 0
TNMD	0.505 1	0.119 6	0.371 8	...	0.073 0	0.219 0	0.020 6
DPM1	23.236 0	18.109 1	22.333 2	...	23.017 1	21.943 3	19.911 3
SCYL3	2.516 5	1.639 1	1.116 2	...	2.024 0	1.846 1	2.540 8
Clorf112	0.874 2	0.602 6	0.498 6	...	0.767 3	0.711 1	0.675 6
...
AC008763.4	0.000 0	0.000 0	0.000 0	...	0.000 0	0.003 7	0.015 3

续表 4 整理数据结果

基因	0	1	2	...	610	611	612
AL592295.6	6.490 7	5.341 8	3.948 0	...	5.830 7	5.999 5	4.885 3
AC006486.3	0.000 0	0.000 0	0.000 0	...	0.000 0	0.000 0	0.000 0
AL391628.1	0.014 9	0.012 7	0.003 5	...	0.008 4	0.024 4	0.031 6
AP006621.6	0.136 6	0.121 2	0.079 2	...	0.240 6	0.306 0	0.244 6

[60 660 rows×612 columns]

表 5 612 例肾癌患者基因表达数据统计分析结果

统计指标	TSPAN6	TNMD	DPM1	...	AC006486.3	AL391628.1	AP006621.6
count	614.000 000	614.000 000	614.000 000	...	614.000 000	614.000 000	614.000 000
mean	13.894 250	1.203 302	21.876 383	...	98.796 417	98.813 480	99.003 112
std	8.140 072	5.466 893	6.422 332	...	2 448.078 726	2 448.118 460	2 448.151 222
min	1.594 600	0.000 000	1.650 400	...	0.000 00	0.000 00	0.000 000
25%	8.852 400	0.079 500	18.315 225	...	0.000 00	0.006 10	0.095 450
50%	11.645 700	0.199 150	21.1050 50	...	0.000 00	0.010 60	0.159 250
75%	16.070 300	0.538 775	24.163 700	...	0.000 00	0.016 80	0.251 325
max	57.7541 00	73.037 000	63.824 600	...	60 661.000 00	60 662.000 00	60 663.000 000

[8 rows×60 660 columns]

注:count 为数量统计,表示每个基因表达量共有多少有效值;mean 表示每个基因表达量的平均值;std 表示每个基因表达量的标准差;min、max 分别表示每个基因表达量的最小、大值;25%、50%及 75%分别表示每个基因表达量的四分之一、二分之一及四分之三位数。

大多数情况下数据采集后是不能直接用于分析的,需要进行数据整理、重复值、异常值或空值等的检测与处理后才能进行数据分析。异常值是指离群点或噪声,其数值明显偏离其他观察值。缺失值及空值可以删除,而异常值则需通过统计学方法进行检测识别。上述代码用到的第三方库为 Pandas 库,是基于 Numpy 库构建的含有更高级数据结构和工具的数据分析包,有 2 个核心数据结构——Series、DataFrame,分别对应了一维的序列和二维的表结构。而“describe()”函数就是返回这 2 个核心数据结构的统计变量。其目的在于观察这一系列数据的范围、大小、波动趋势等,为后面的模型选择打下基础。Pandas 库提供了高效地操作大型数据集所需的工具。上述代码运行总耗时 6 min 43 s,极大地缩短了分析

时间。

1.3 案例 3 根据案例 1 得到的 614 例患者 TSPAN6、TNMD、DPM1、AC006486.3、AL391628.1 基因的表达水平,应用 Python 分析这 5 个基因之间的相关性。输入并分析数据,得到 TSPAN6、TNMD、DPM1、AC006486.3、AL391628.1 基因的表达相关性系数(表 6)及其可视化结果(图 1)。具体算法流程如下:(1)导入“pandas”“seaborn”“matplotlib.pyplot”及“numpy”模块。(2)读取数据并储存于 data。(3)使用“corr()”函数对‘TSPAN6’‘TNMD’‘DPM1’‘AC006486.3’‘AL391628.1’进行相关性分析。输出结果见表 6。(4)使用“sns.heatmap”进行相关性分析结果可视化,结果见图 1。

表 6 TSPAN6、TNMD、DPM1、AC006486.3、AL391628.1 基因的表达相关性系数

基因	TSPAN6	TNMD	DPM1	AC006486.3	AL391628.1
TSPAN6	1.000 000	0.133 654	0.226 462	-0.049 134	-0.049 135
TNMD	0.133 654	1.000 000	-0.001 733	0.028 073	0.028 073
DPM1	0.226 462	-0.001 733	1.000 000	-0.099 927	-0.099 930
AC006486.3	-0.049 134	0.028 073	-0.099 927	1.000 000	1.000 000
AL391628.1	-0.049 135	0.028 073	-0.099 930	1.000 000	1.000 000

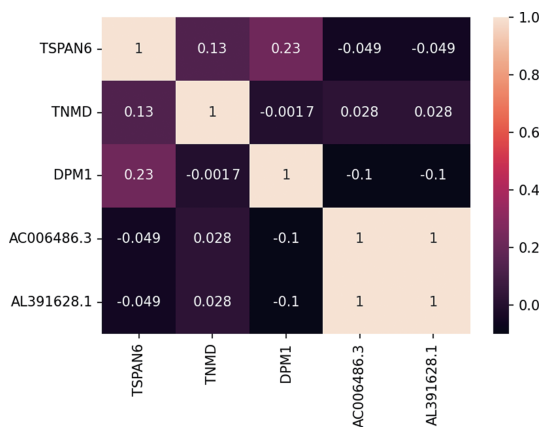


图 1 5 个基因(TSPAN6、TNMD、DPM1、AC006486.3、AL391628.1)表达相关性热图

这 5 个基因的表达量为连续型变量,需要用相关性分析来完成。Pandas 中的“corr()”函数为数据相关性分析中的常用函数,默认采用皮尔逊相关系数(Pearson correlation coefficient)表示 2 个变量之间的相关性。其取值范围为[-1,1],相关系数越接近-1 或 1,表示变量之间的负相关或正相关性越强;越接近 0,则表示相关性越弱。运行代码结果显示,TSPAN6 与 DPM1 相关性最强,相关系数为 0.226 462;TNMD 与 DPM1 相关性最弱,相关系数为-0.001 733,同时 5 个基因之间的相关情况采用热力图更加直观地显示出来,代码运行耗时 1 min 46 s。

通过上述实例可以看出,Python 在文献检索下载、数据处理及数据分析中具有易于上手、直观、功能强大、运算速度快等优点,可以极大地减少学生繁杂的下载、计算过程。此外,还可让医学生了解数据分析中 Python 的基本算法及编程思路,培养其掌握程序设计语言中的一些简单常见算法,提高思维计算能力,这样不仅能提高医学生的动手能力,还能激发提高医学生的学习兴趣,为今后实际应用数据获取、处理和分析技术和学术能力打下良好基础。

2 小 结

大数据背景下,医学教学中要充分利用 Python 语言编程设计,优化教学资源,强化文献检索与数据挖掘方法的应用,掌握基本的程序设计技能,切实提高数据分析能力^[12]。同时,医学院教师要密切关注医学发展的热点和前沿,除了传统的医学教学内容之外,还要注意紧跟时代潮流,充分利用 Python 编程的优势,与传统教学相结合,培养学生提高临床应用与基础科研的能力,为国家大数据战略培养高水平医学人才^[13]。

本文探讨了将 Python 语言应用到医学文献检索

与统计学实训课程的教学中,通过 Python 对文献检索与统计学实训课程中的案例进行实际操作演示,为医学生提供对 Python 的直观认识,锻炼医学生的临床思维和科研思维能力,提高了医学生数据分析处理能力和解决实际问题的能力,实现了复杂问题简单化,对于激发学生的学习兴趣,提高教学质量都具有重要的作用。此外,医学研究常常需要收集、整理、分析大量数据,如借助 Python 语言编程,则可大大提高工作效率,一定程度上缓解了医学生因课业繁重、学制较长而面临的心理健康问题^[14-15]。

参考文献

- [1] 张晓莉. 应用大数据推动经济社会发展创新的思路[J]. 经济研究导刊,2021,17(31):1-3.
- [2] 杨为民,叶春蕾,刘子渲. 基于大数据时代的数据价值分析[J]. 科技成果管理与研究,2022,17(5):28-31.
- [3] 孙晓静. 大数据技术在统计信息化中的运用分析[J]. 财经界,2022,40(25):78-80.
- [4] 王陆,袁睿,杨萌萌,等. 大数据背景下关于重症医学研究生培养的思考[J]. 中国继续医学教育,2019,11(7):43-45.
- [5] 王晶. 医学生科研伦理线上培训模式调查分析[J]. 中国医学伦理学,2022,35(2):180-184.
- [6] 王燕,张荣,杜琰. 临床医学生科研创新能力需求的调查研究[J]. 基础医学教育,2022,24(2):135-139.
- [7] 康巍,钟武宁,韦苇,等. 大数据背景下人工智能在医学研究生个性化网络教学中的探索[J]. 中国继续医学教育,2018,10(32):41-43.
- [8] MARIANO D, MARTINS P, HELENE SANTOS L, et al. Introducing programming skills for life science students[J]. Biochem Mol Biol Educ,2019,47(3):288-295.
- [9] 魏晋,郝文延,王晓云. 医学院校 Python 程序设计课程的教学改革研究[J]. 计算机时代,2022,359(5):96-98.
- [10] 李秀敏,王希杰,刁丽娟. 面向研究型及应用型医学人才培养的 Python 教学探索[J]. 卫生职业教育,2018,36(16):74-76.
- [11] 郭洪彬,王箜杨. Python 编程在八年制医学实习生教学中的应用[J]. 中国医学教育技术,2020,34(5):613-616.

- [12] 田翔华. 医学院校开展 Python 程序设计课程教学设计与实践[J]. 电脑知识与技术, 2020, 16(19):121.
- [13] 魏晋, 郝文延, 王晓云. 医学院校 Python 程序设计课程的教学改革研究[J]. 计算机时代, 2022, 40(5):96-98.
- [14] 钱怡, 孙金海, 薛雅, 等. 八年制临床医学学生生命

质量现状调查及影响因素研究[J]. 中国卫生事业管理, 2019, 36(8):627-631.

- [15] 李宁宁, 李航, 侍效春, 等. 临床医学八年制本科生多阶段心理状况调查及分析[J]. 基础医学与临床, 2020, 40(1):141-144.

(收稿日期:2022-09-03 修回日期:2023-04-18)

• 教学探索 •

邓巴数亲密朋友模式在手术室新入职护士规范化培训中的应用研究

余伟玲, 韦燕萍, 莫乙瑜, 张艺, 陈春燕, 韦宇
(广西科技大学第二附属医院, 广西柳州 545006)

[摘要] 目的 研究将邓巴数亲密朋友模式运用于手术室新入职护士规范化培训(规培)的实施效果。方法 将 2018 年 1 月至 2020 年 5 月 30 名在该院手术室规培的新入职护士作为对照组, 采用双导师制方式教学。另将 2020 年 6 月至 2022 年 6 月 30 名在该院手术室规培的新入职护士作为观察组, 选用五位导师, 建立亲密微信群, 采用多学科参与培训方式教学。经过培训后考核 2 组护士理论知识、操作技能、手术配合能力、困难情境沟通能力及与同事关系满意度。结果 观察组护士理论知识及操作技能成绩、手术配合能力、困难情境沟通能力及与同事关系满意度均高于对照组, 差异均有统计学意义($P < 0.001$)。结论 将邓巴数亲密朋友模式应用于手术室新入职护士规培, 既提高了护士的专科理论知识水平、操作技能及手术配合能力, 也增强了其困难情境沟通能力, 提升了护士与同事关系的满意度。

[关键词] 邓巴数; 亲密朋友; 微信群; 手术室; 新入职; 护士; 规范化培训; 教学模式

DOI: 10.3969/j.issn.1009-5519.2023.09.033

中图法分类号: R615; R619+.9

文章编号: 1009-5519(2023)09-1595-04

文献标识码: C

2016 年, 国家卫生健康委员会颁发了《新入职护士培训大纲(试行)》^[1], 提出开展新入职护士培训是培养合格临床护士的重要途径, 是提高临床护理质量、保证医疗安全的有力举措。新入职护士需全面掌握基础理论知识和临床操作技能, 具有良好沟通协作能力、应急处理能力及职业操守。手术室是外科治疗和抢救重症患者的科室^[2], 手术室护理专科性强, 相关知识、专业知识及专业技术均涉及外科、麻醉科及消毒供应中心专科理论。实际工作开展, 与相关科室配合沟通密不可分, 多学科培训可有效促进各科室沟通互助^[3]。传统的一对一带教^[4]及双导师带教^[5]模式, 均全程以手术室导师为主导, 并未加入外科等专科导师的教授, 不利于新入职护士系统掌握培训内容, 也无益于沟通及应急处理工作模式的形成。20 世纪 90 年代, 牛津大学学者罗宾·邓巴提出, 人类的智力水平决定社会群体规模, 150 人为稳定社交人数极限, 150 被称为“邓巴数”, 微信创始人张小龙公开表示, 微信群中很多功能都是根据“邓巴数”设置^[6]。个

体的社交网络共 150 人, 分 4 个层级, 密友 5 人, 至交 15 人, 好友 45 人, 朋友 150 人, 各层级朋友身份随着社交群而发生变化^[6]。新入职护士进入陌生且严谨的手术室大环境, 难免心生恐惧。本研究中, 全程导师主动建立起专属的规培团队, 以邓巴数亲密朋友模式建立微信群, 手术室总带教导师、全程带教导师、麻醉科总带教麻醉医师、外科总带教医师、消毒供应中心总带教导师及一名新入职护士均在微信群中。新入职护士在较强归属感的环境下, 予以多学科协作的专业知识教授, 并同时能便捷、精准地交流。研究结果提示邓巴数亲密朋友模式在手术室新入职护士规范化培训(规培)中的效果显著。

1 资料与方法

1.1 一般资料 将 2018 年 1 月至 2020 年 5 月在本院手术室进行规培的 30 名新入职护士作为对照组, 2020 年 6 月至 2022 年 6 月在手术室规培的 30 名新入职护士作为观察组。对照组护士平均年龄(22.88 ± 1.94)岁; 其中女 29 名, 男 1 名; 本科学历 5