

# DeepSeek与SPSS在统计分析中的优劣比较研究

庞沙沙, 叶孟良\*

重庆医科大学公共卫生学院, 重庆

收稿日期: 2025年5月26日; 录用日期: 2025年6月18日; 发布日期: 2025年6月30日

## 摘要

目的: 比较DeepSeek与SPSS软件在数据分析方面的优劣。方法: 比较两者对同一数据集的分析结果验证两者的性能差异。结果: 与SPSS软件相比, DeepSeek进行数据预处理的准确性较差; 定量资料的统计描述误差不超过2%, 但定性资料的误差较大; Logistic回归模型预测准确率较低为67%; 因子分析的各个部分输出结果误差均较大, 且提取的公因子累计解释方差未达到60%。结论: DeepSeek进行数据处理的准确性不如传统分析软件SPSS, 但在易用性和可访问性方面具有优势。

## 关键词

DeepSeek, 统计分析能力, 比较研究

# A Comparative Study of the Advantages and Disadvantages of DeepSeek and SPSS in Statistical Analysis

Shasha Pang, Mengliang Ye\*

College of Public Health, Chongqing Medical University, Chongqing

Received: May 26<sup>th</sup>, 2025; accepted: Jun. 18<sup>th</sup>, 2025; published: Jun. 30<sup>th</sup>, 2025

## Abstract

The objective of this study is to compare the strengths and weaknesses of DeepSeek and SPSS software in data analysis. To achieve this, the analysis results of the two software on the same dataset were compared to verify their performance differences. The results show that, compared with SPSS software, DeepSeek has lower accuracy in data preprocessing. While the error in the statistical description of quantitative data does not exceed 2%, the error in qualitative data is relatively large.

\*通讯作者。

**Additionally, the prediction accuracy rate of the Logistic regression model is relatively low, at 67%. The error in each part of the factor analysis output is also relatively large, and the cumulative variance explained by the extracted common factors does not reach 60%. In conclusion, DeepSeek is less accurate than the traditional analysis software SPSS in data processing and analysis. However, it has advantages in terms of user-friendliness and accessibility.**

## Keywords

**DeepSeek, Statistical Analysis Capability, Comparative Study**

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

自 2020 年生成式人工智能(AI)发展以来,它在医学教育、培训和先进的医疗实践中发挥了重要作用。一些学者探索了使用生成式人工智能来解释数据和医学图像[1][2];还有学者采用不同国家的医疗执照考试来评估生成式人工智能是否有可能达到医生的资格要求水平[3]-[6]。此外,一些学者对生成式人工智能(如 ChatGPT)作为数据分析工具的效用进行了研究,然而研究结论不一致[7]-[10]。DeepSeek 是生成式人工智能的一种,由于它“低成本高效率”以及在一众能力相等模型中率先开源,对人工智能的资本支出以及其他行业的更广泛应用都产生了深远影响[11],成为新一代人工智能技术的代表。本研究将比较当前火热的大语言模型 DeepSeek 与传统统计分析软件 SPSS 的数据分析能力,为 DeepSeek 应用于统计分析领域提供参考。

## 2. 对象与方法

### 2.1. 研究对象

本研究采用的验证数据集由某一健康素养数据集的部分原始数据经过预处理后得到。原始数据集共 100 条观测,包含 12 个职业健康素养测试判断题的实际答案、调查时间、出生年月、性别、文化程度和平均月收入等 17 个变量。数据预处理包括以下三个步骤:① 根据调查时间、出生年月计算年龄;② 判断实际答案是否正确,每答对一题得一分,答错不得分,先计算每一题的得分再计算总分;③ 根据公式  $\text{正确率} = \text{总分}/12$ ,判断是否具备职业健康素养,正确率  $\geq 80\%$ 即为具备职业健康素养。最后的验证数据集由 12 个职业健康素养测试判断题得分、总分、是否具备职业健康素养、年龄、性别、文化程度和平均月收入等 18 个变量构成,其中年龄和总分为连续变量,性别和是否具备职业健康素养为二分类变量、文化程度为无序多分类变量,平均月收入为有序多分类变量。研究的内容为了解 DeepSeek 与 SPSS 在数据预处理、统计描述、假设检验、回归分析、高级分析的完成情况。

### 2.2. 研究方法

本研究使用 Microsoft Edge 浏览器网页访问 DeepSeek 官网开展对话。首先上传 csv 文件格式的附件,采用结构化提问公式进行提问:“提取附件中的数据表格,命名为 TAB1,” + 统计分析需求 + “以表格的方式输出分析结果”。同时,使用 SPSS 26 软件进行分析,该软件为菜单式操作。比较两种方式的分析结果是否准确和一致。

### 2.3. 统计学方法

定量资料采用平均值、标准差、最小值、四分位数、最大值、观察值数进行统计描述;定性资料采用频数进行统计描述。采用卡方检验和 Logistic 回归分析探索两种方式完成假设检验和回归分析的能力;高级分析选择因子分析方法,对 12 个职业健康素养测试题得分采用主成分分析方法提取公因子,并采用最大方差法进行旋转。显著性水平为 0.05。

## 3. 结果

### 3.1. 数据预处理与统计描述

分别运用 DeepSeek 和 SPSS 软件对原始数据集执行预处理的三个步骤后,由于 DeepSeek 不能输出文件,也不能完整展示处理后的数据集,故采用统计描述的方法来判断两种方式进行预处理后的数据是否一致。结果发现两种方式进行预处理后的数据几乎不一致,定量资料的具体差异见表 1,定性资料的具体差异见表 2,表明 DeepSeek 在数据预处理方面的准确性较差,后续的统计分析将以 SPSS 软件处理过的数据集作为验证数据集进行。

分别用 DeepSeek 和 SPSS 软件对验证数据集进行统计描述,发现两种方式的分析结果几乎不一致,定量资料的具体差异见表 1,定性资料的具体差异见表 2,表明 DeepSeek 在对数据进行统计描述时的准确性较差,尤其是定性资料的频数统计方面。

**Table 1.** Table of statistical descriptive results of quantitative variables

**表 1.** 定量变量的统计描述结果表

变量	平均值	标准差	P0	P25	P50	P75	P100	观察值数
SPSS 对原始数据集预处理后								
年龄	39.1	9.1	23	32	39.5	46.8	58	100
总分	9.9	1.5	5	9	10	11	12	100
DeepSeek 对原始数据集预处理后								
年龄	39.3	9.6	23	32	39	46	58	100
总分	9.4	1.8	4	8	10	11	12	100
DeepSeek 对验证数据集								
年龄	38.4	9.1	23	32	38	44	58	100
总分	9.8	1.5	5	9	10	11	12	100

**Table 2.** Table of statistical descriptive results of qualitative variables

**表 2.** 定性变量的统计描述结果表

变量	SPSS	DeepSeek <sup>a</sup>	DeepSeek <sup>b</sup>	变量	SPSS	DeepSeek <sup>a</sup>	DeepSeek <sup>b</sup>
职业健康素养				性别			
具备	60	53	67	男	89	81	76
不具备	40	47	33	女	11	19	24
文化程度				平均月收入			
不识字/识字很少	3	4	4	少于 3000 元	10	10	8
小学	5	10	8	3000~4999 元	55	56	51
初中	33	36	34	5000~6999 元	19	16	17
高中/职高/中专	35	30	27	7000~8999 元	7	9	9
大专	12	12	14	9000~10,999 元	2	4	6
大学本科	12	8	13	11,000 元及以上	7	5	9

注: <sup>a</sup> DeepSeek 对原始数据集预处理后的数据进行统计描述; <sup>b</sup> DeepSeek 对验证数据集进行统计描述。

### 3.2. 假设检验

分别运用 DeepSeek 和 SPSS 软件对验证数据集进行卡方检验, 分析不同性别、文化程度、平均月收入对职业健康素养的单独影响。两种方式得出了不同的结论, DeepSeek 的结果表明性别、文化程度、平均月收入均对职业健康素养无显著性影响( $p > 0.05$ ); SPSS 软件的结果表明性别对职业健康素养无显著性影响( $p > 0.05$ ), 而文化程度、平均月收入均对职业健康素养有显著性影响( $p < 0.05$ ), 两种方式分析结果的具体差异见表 3。因此, 运用 DeepSeek 进行假设检验, 得到的结果不太可靠。

**Table 3.** Table of results of single factor analysis

**表 3.** 单因素分析结果表

变量	OHL = 0	OHL = 1	$\chi^2$	$p$	OHL = 0	OHL = 1	$\chi^2$	$p$
性别			0.56	0.454			0.15	0.695
男	24	52			35	54		
女	9	15			5	6		
文化程度			10.02	0.075			17.12	<b>0.004</b>
不识字/识字很少	2	2			2	1		
小学	4	4			5	0		
初中	15	19			16	17		
高中/职高/中专	6	21			14	21		
大专	3	11			1	11		
大学本科	3	10			2	10		
平均月收入			5.20	0.391			8.46	<b>0.004</b>
少于 3000 元	4	4			6	4		
3000~4999 元	15	36			27	28		
5000~6999 元	5	12			5	14		
7000~8999 元	4	5			1	6		
9000~10,999 元	2	4			0	2		
11,000 元及以上	3	6			1	6		

注: OHL = 0 表示不具备职业健康素养, OHL = 1 表示具备职业健康素养; 第 2~5 列为 DeepSeek 分析结果, 第 6~9 列为 SPSS 分析结果。

### 3.3. 回归分析

分别运用 DeepSeek 和 SPSS 软件对验证数据集进行 Logistic 回归分析, 分析不同性别、文化程度、平均月收入对职业健康素养的共同影响。首先对分类变量进行编码, 见表 4。虽然两种方式的分析结果完全不相同, 但得出了相同的结论: 年龄、性别、文化程度、平均月收入均对职业健康素养无显著性影响( $p > 0.05$ ), 两种方式分析结果的具体差异见表 5。两种方式建立的 Logistic 回归模型的模型摘要显示, DeepSeek 的模型预测准确率为 67%, SPSS 的模型预测准确率为 71%。因此, 与 SPSS 软件相比, DeepSeek 建立的 Logistic 回归模型预测准确率稍差。

**Table 4.** Variable assignment table

**表 4.** 变量赋值表

变量	变量取值	变量赋值
是否具备职业健康素养	否	0
	是	1
性别	男	0
	女	1

续表

文化程度	女	1
	不识字/识字很少	(0, 0, 0, 0, 0)
	小学	(1, 0, 0, 0, 0)
	初中	(0, 1, 0, 0, 0)
	高中/职高/中专	(0, 0, 1, 0, 0)
	大专	(0, 0, 0, 1, 0)
	大学本科	(0, 0, 0, 0, 1)

Table 5. Table of results of multifactor analysis

表 5. 多因素分析结果表

变量	类别	$\beta$ 系数	Wald $\chi^2$	$p$	OR	95% CI
DeepSeek						
常量		-0.892	0.300	0.584		
性别	男					
	女	-0.213	0.216	0.642	0.81	0.33~1.99
文化程度	不识字/识字很少	0.402	0.127	0.721	1.49	0.16~13.69
	小学	0.831	0.659	0.417	2.30	0.31~17.09
	初中	1.483	1.988	0.159	4.41	0.56~34.59
	高中/职高/中专	1.534	1.884	0.170	4.64	0.52~41.41
	大专	1.569	1.934	0.164	4.80	0.53~43.78
	大学本科	0.102	0.624	0.430	1.11	0.86~1.43
平均月收入		0.021	1.071	0.301	1.02	0.98~1.06
年龄						
SPSS 软件						
常量		-1.056	0.439	0.508	0.35	
性别	男					
	女	-0.483	0.449	0.503	0.62	0.15~2.54
文化程度	不识字/识字很少	4.381	0.496			
	小学	-19.972	0.000	0.999	0.00	
	初中	1.075	0.584	0.445	2.93	0.19~46.05
	高中/职高/中专	1.246	0.852	0.356	3.47	0.25~48.95
	大专	3.129	3.402	0.065	22.85	0.82~635.13
	大学本科	1.676	1.09	0.296	5.34	0.23~124.24
平均月收入		0.363	1.732	0.188	1.44	0.84~2.47
年龄		-0.015	0.285	0.594	0.98	0.93~1.04

### 3.4. 因子分析

对 12 个测试题的得分进行因子分析。DeepSeek 分析结果显示, KMO 值为 0.823, Bartlett 球形检验卡方值为 542.37,  $p < 0.001$ , 得出数据适合进行因子分析的结论; SPSS 软件的分析结果显示, KMO 值为 0.539, Bartlett 球形检验卡方值为 265.402,  $p < 0.001$ , 表明数据勉强适合进行因子分析。采用主成分分析方法初步提取公因子, DeepSeek 结果显示, 所有变量的共同度均大于 0.5, 说明提取的公因子能较好地解释原始变量, 并提取 2 个公因子, 累计解释方差为 57.13%, 尚未达到 60%, 表明未充分提取数据中的信息; SPSS 软件结果显示, 除第 10 题外的所有变量的共同度均大于 0.5, 说明提取的公因子能较好地解释原始变量, 并提取 5 个公因子, 累计解释方差为 67.84%。经过最大方差法旋转后的成分矩阵见表 6。因此, 与 SPSS 软件相比, DeepSeek 进行因子分析的能力较差。

**Table 6.** Component matrix after rotation  
**表 6.** 旋转后的成分矩阵

变量	公因子 1	公因子 2	公因子 1	公因子 2	公因子 3	公因子 4	公因子 5
第 1 题	0.742			0.786			
第 2 题	0.813			0.848			
第 3 题		0.752		0.673			
第 4 题	0.792					0.876	
第 5 题		0.712				0.731	
第 6 题	0.821						0.944
第 7 题	0.843		0.675				
第 8 题	0.819		0.702				
第 9 题		0.752			0.569		
第 10 题	0.812				0.759		
第 11 题	0.851				0.735		
第 12 题		0.761				0.486	

注：第 2 和 3 列为 DeepSeek 分析结果，第 4~8 列为 SPSS 分析结果。

## 4. 讨论

在大数据时代，选择合适的分析工具对数据进行挖掘和统计分析至关重要。本研究旨在比较新兴 AI 分析工具 DeepSeek 与传统专业统计软件 SPSS 在数据预处理和统计分析方面的性能差异。分别用两种工具对同一数据集进行系统验证，发现 DeepSeek 数据处理和统计分析的结果缺乏准确性。这可能是因为 DeepSeek 的算法实现与 SPSS 这类专业统计软件存在本质差异，SPSS 采用经过数十年验证的经典统计算法实现，每一步计算都遵循严格的数值分析原则。而 DeepSeek 可能采用了某些近似计算方法以提高处理速度，这在简单分析中可能差异不大，但在复杂模型或多步骤分析中误差将越来越大。数据预处理流程的透明度问题也是导致 DeepSeek 某些分析结果不准确的原因之一。数据预处理是统计分析的关键环节，SPSS 提供了完全透明且可自定义的预处理流程；然而，DeepSeek 的预处理操作则像“黑箱”，用户难以明确知道缺失值处理、异常值检测和数据转换编码的具体情况。在对原始数据集验证中，DeepSeek 处理后具备职业健康素养的比例为 53% 与 SPSS 用户指定处理的结果(60%)存在差异，误差超过 10%，表明 DeepSeek 的预处理结果不可靠。对同一验证数据集分析，在基础统计量计算方面，两种工具在定量资料上表现相对接近，均值、标准差等基本指标误差不超过 2%，说明 DeepSeek 在简单计算上是可靠的。然而，在定性资料的描述中，DeepSeek 报告的频数与 SPSS 结果差异较大，将影响卡方检验的结果可靠性。事实上，DeepSeek 输出的卡方检验结果确实不可靠，并未发现受教育程度、平均月收入与职业健康素养的显著影响。在 Logistic 回归分析中，DeepSeek 的模型预测准确率为 67%，低于 SPSS 的模型预测准确率(71%)。在因子分析中，DeepSeek 虽然能完整地输出因子分析的各个部分结果，但结果与 SPSS 软件的结果相差甚远，并且 DeepSeek 提取的公因子的累计解释方差未达到 60%，表明 DeepSeek 在进行高级分析(如因子分析)的可靠性较差。本研究的实证结果与一项探索生成式 AI 应用于政府统计服务的理论研究结论几乎不一致[12]，可能是由于 DeepSeek 等大模型的适用边界不同[13]。

专业统计软件 SPSS 的分析过程可控且结果可解释，用户能对每个分析步骤逐步进行验证；而 DeepSeek 的自动化处理虽然便捷，但缺乏透明度，分析过程不可控，用户不能对每个分析步骤的结果进行验证。此外，生成式 AI 复杂的工作原理使用户难以直接理解、解释模型的决策与推理过程，DeepSeek 的低可解释性最终导致用户难以证明结果的可靠性和可信度[12]。因此，建议 DeepSeek 开发者提供分析方法的详细文档和计算步骤说明，或者通过可视化技术呈现算法推理路径以解决算法透明度问题，并提

供中间结果输出, 允许用户检查预处理和模型拟合的中间步骤; 同时引入专业统计验证以确保核心算法的准确性。通过输出中间结果, 用户及时发现并解决问题; 或者开发者对中间结果分析如特征归因、梯度监控等优化模型, 可能会提高分析结果的准确性和可信度。

综上所述, 虽然 DeepSeek 在易用性和可访问性方面具有优势, 但在分析结果准确性上与传统专业软件 SPSS 仍存在明显差距。这种差距在基础统计分析中可能影响较小, 但在构建复杂模型和精确推断时可能导致实质性偏差。若未来 DeepSeek 能通过改进算法和进行专业验证不断提升分析结果的准确性, 同时保持易用性, 平衡便捷性与专业性的关系, DeepSeek 将有望成为可靠且实用的统计分析工具。

## 基金项目

重庆市教育科学规划课题 2024 年度教学改革研究专项重点课题, 新时代家校社协同的学生健康素养提升路径研究, 课题编号: K24ZG2040059。

## 参考文献

- [1] Yu, P., Xu, H., Hu, X. and Deng, C. (2023) Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. *Healthcare*, **11**, Article No. 2776. <https://doi.org/10.3390/healthcare11202776>
- [2] Busch, F., Hoffmann, L., Rueger, C., van Dijk, E.H., Kader, R., Ortiz-Prado, E., et al. (2025) Current Applications and Challenges in Large Language Models for Patient Care: A Systematic Review. *Communications Medicine*, **5**, Article No. 26. <https://doi.org/10.1038/s43856-024-00717-2>
- [3] Wang, H., Wu, W., Dou, Z., He, L. and Yang, L. (2023) Performance and Exploration of ChatGPT in Medical Examination, Records and Education in Chinese: Pave the Way for Medical AI. *International Journal of Medical Informatics*, **177**, Article ID: 105173. <https://doi.org/10.1016/j.ijmedinf.2023.105173>
- [4] Sumbal, A., Sumbal, R. and Amir, A. (2024) Can ChatGPT-3.5 Pass a Medical Exam? A Systematic Review of ChatGPT's Performance in Academic Testing. *Journal of Medical Education and Curricular Development*, **11**, 1-12. <https://doi.org/10.1177/23821205241238641>
- [5] Gilson, A., Safranek, C.W., Huang, T., Socrates, V., Chi, L., Taylor, R.A., et al. (2023) How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Medical Education*, **9**, e45312. <https://doi.org/10.2196/45312>
- [6] Lin, S., Chan, P.K., Hsu, W. and Kao, C. (2024) Exploring the Proficiency of ChatGPT-4: An Evaluation of Its Performance in the Taiwan Advanced Medical Licensing Examination. *Digital Health*, **10**, 1-11. <https://doi.org/10.1177/20552076241237678>
- [7] Huang, Y., Wu, R., He, J. and Xiang, Y. (2024) Evaluating Chatgpt-4.0's Data Analytic Proficiency in Epidemiological Studies: A Comparative Analysis with SAS, SPSS, and R. *Journal of Global Health*, **14**, Article No. 04070. <https://doi.org/10.7189/jogh.14.04070>
- [8] Ignjatović, A. and Stevanović, L. (2023) Efficacy and Limitations of ChatGPT as a Biostatistical Problem-Solving Tool in Medical Education in Serbia: A Descriptive Study. *Journal of Educational Evaluation for Health Professions*, **20**, Article No. 28. <https://doi.org/10.3352/jeehp.2023.20.28>
- [9] Jahangiri, Y. (2023) Can Chat Generative Pretraining Transformer (ChatGPT) Be Used for Statistical Analysis of Research Data? *Journal of Vascular and Interventional Radiology*, **34**, 2242-2246.e2. <https://doi.org/10.1016/j.jvir.2023.09.010>
- [10] Ordak, M. (2023) ChatGPT's Skills in Statistical Analysis Using the Example of Allergology: Do We Have Reason for Concern? *Healthcare*, **11**, Article No. 2554. <https://doi.org/10.3390/healthcare11182554>
- [11] 丁炜. DeepSeek 对生成式人工智能应用的影响与启示[J]. 银行家, 2025(3): 129-134.
- [12] 林戴忠. 生成式 AI 应用于政府统计服务的理论初探[J]. 统计科学与实践, 2023(7): 53-56.
- [13] 董昌其, 李大宇, 米加宁. 大模型嵌入政务服务: 能力边界、协同治理与发展路径——基于地方政府大规模部署 DeepSeek 的观察[J/OL]. 电子政务, 1-9. [https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbcEJxkxW7L6ZXHPCEuefTn9uDA92BLOVb3prTdtL4IZjz6Nf67BQKnSF4tPXuZOrO-HPRQnYEaWvbPflSU3aUWHpRp4\\_r9M8bXKXNykYIwMEFCAvMOUJXD3LvcSivWoNsBdInrC-BD\\_OjZg64sOFSecEdA2OOOnBX17AwI1pB4dIPO==&uniplatform=NZKPT&language=CHS, 2025-03-19](https://kns.cnki.net/kcms2/article/abstract?v=UThtwiquHbcEJxkxW7L6ZXHPCEuefTn9uDA92BLOVb3prTdtL4IZjz6Nf67BQKnSF4tPXuZOrO-HPRQnYEaWvbPflSU3aUWHpRp4_r9M8bXKXNykYIwMEFCAvMOUJXD3LvcSivWoNsBdInrC-BD_OjZg64sOFSecEdA2OOOnBX17AwI1pB4dIPO==&uniplatform=NZKPT&language=CHS, 2025-03-19)