An Assessment of ChatGPT's Ability to Code for Different Statistical Packages

Debra R. Winberg, MSc MA; Tiange Tang, MPH; Dennis Xuan, MS
Department of Health Policy and Management, School of Public Health and Tropical Medicine, Tulane University

MSR25

Tulane SCHOOL OF PUBLIC HEALTH AND TROPICAL MEDICINE

INTRODUCTION

- ChatGPT is a **natural language processing tool** that creates human-like conversations, responds to questions, and creates written content.
- Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.
- Opportunity present to utilize ChatGPT for **innovative research methodologies in public health (1).**
- A new usage for ChatGPT is to **write code**. As generative AI gains popularity, it is essential to understand its capabilities in public health research (2).
- Researchers have identified many **use cases** where it's helpful to use ChatGPT for coding, but there has been limited testing (1-4),

AIM

This project aims to assess **the success, accuracy, and effectiveness of ChatGPT** in coding statistical programming common public health questions in Stata, R, and Visual Basic for Applications (VBA).

METHODS

Step 1. Baseline/Comparison

All researchers will create, QC, and agree upon code to answer the question in all three statistical software. An independent researcher will be brought in to QC all codes.

Step 2. GPT Prompts

Separate GPT prompts will be written to answer each of the 6 capacities we are assessing if GPT can code for. All researchers will agree upon and use the same prompts

Step 3. Running Code

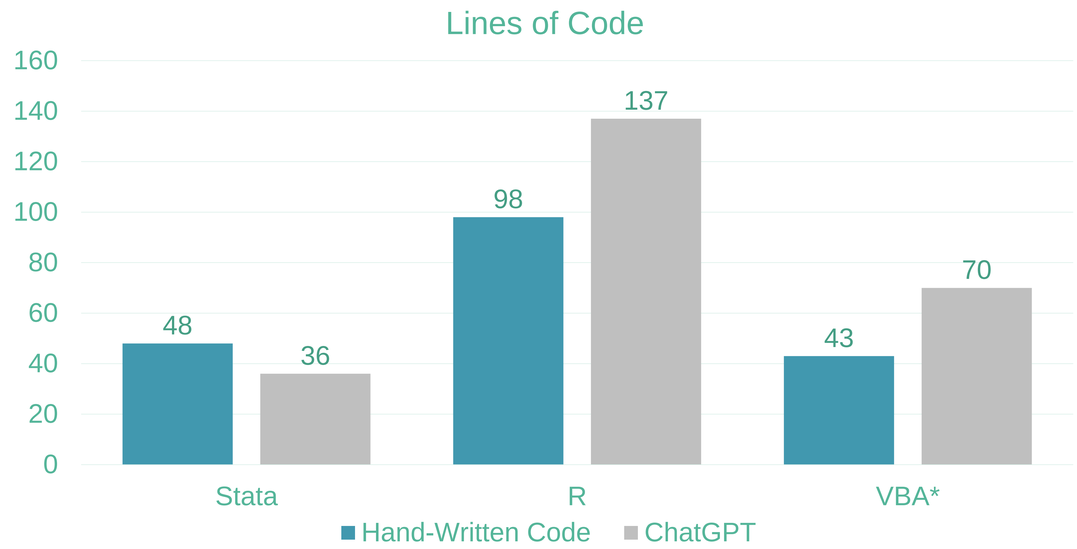
Researchers will run the prompts in GPT and test the resulting code. If there is an error message after any step, the researchers can go back and ask GPT to rewrite the code based on the specific error message

Outcomes

1. Success: Defined as a binary 'yes' or 'no' if GPT was able to create code to complete the task
2. Accuracy: If the final results are the same as the sample code created
3. Efficiency: Number of commands

RESULTS

- Results were similar between the researcher's codes and ChatGPT 4.0 codes for all software, but all three programs required changes to code



*ChatGPT prompts for VBA did not execute for all 4 steps. These step counts are for a truncated process. Handwritten code to complete the analysis process would be greater than the amount shown here

Hand-Written Code (Stata)

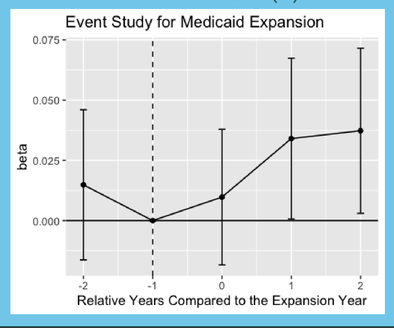
	(1) Means Sample	Treatment ~p	(2) Total Obse~s Sample	Treatment ~p	(3) control_gr~p Sample	Treatment ~p
income_per~a	44921.89		47216.72		41397.59	
population~s	578907.2		750202.6		315839.3	
transfer_p~a	8.625729		8.85214		8.278018	
employment~a	.5926644		.5953799		.588494	
percent_sex	.4939961		.4941397		.4937755	
percent_a~65	.1627252		.1609344		.1654755	
percent_bl~k	.1103659		.097172		.1306286	
percent_hi~c	.1106041		.1158157		.1026005	
charity_ca~e	7.80501		9.724199		4.857605	
charity_ca~f	.0745766		.0737214		.0758896	
bad_debt_e~e	8455204		7401093		1.01e+07	
bad_debt_p~s	1059795		1039005		1091722	
rural	.2084145		.1737649		.2616279	
teaching	.0685544		.0815824		.0485465	
N	8723		5283		3440	

GPT-Written Code (Stata)

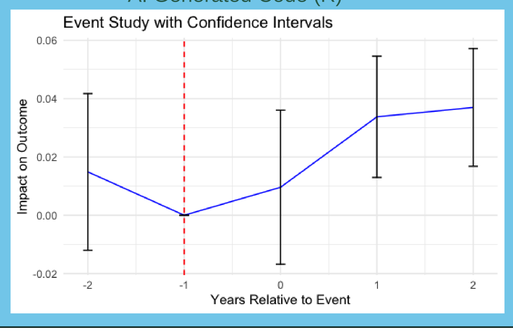
	e(count)	e(sum_w)	e(mean)	e(Var)	e(sd)	e(min)	e(max)	e(sum)
income_per~a	3440	3440	41397.59	1.00e+08	10007.39	22260	120577	1.42e+08
population~s	3440	3440	315839.3	3.61e+11	600473.8	1270	4589928	1.09e+09
transfer_p~a	3440	3440	8.278018	2.993794	1.730258	3.006127	14.93026	28476.38
employment~a	3440	3440	.588494	.0194927	.1396163	.2110687	1.075574	2024.419
percent_sex	3440	3440	.4937755	.0002391	.0154633	.4291201	.6308162	1698.588
percent_a~65	3440	3440	.1654755	.0022132	.0470447	.0722929	.3525822	569.2359
percent_bl~k	3440	3440	.1306286	.0236689	.153847	.0024643	.7278561	449.3624
percent_hi~c	3440	3440	.1026005	.0181628	.1347694	.0052258	.9183861	352.9458
charity_ca~e	3440	3440	4.857605	137.8326	11.74021	-64.6288	125.1774	16710.16
charity_ca~f	3440	3440	.0758896	.0302858	.1740282	-.72	1	261.0603
bad_debt_e~e	3440	3440	1.01e+07	4.10e+14	2.02e+07	-682521	4.16e+08	3.47e+10
bad_debt_p~s	3440	3440	1091722	1.24e+13	3515921	0	5.86e+07	3.76e+09
rural	3440	3440	.2616279	.1932349	.4395849	0	1	900
teaching	3440	3440	.0485465	.0462032	.2149492	0	1	167

The generative-AI code was successful at writing code to calculate descriptive statistics for each group, however it was not able to generate code to create a table with the appropriate columns and rows in an easy-to-read table

Hand-Written Code (R)



AI Generated Code (R)



DISCUSSION

- ChatGPT can code for common public health statistical problems and is most effective when the operator is familiar with the program language.
- All programs required researchers to make significant changes to the ChatGPT code.
- When prompting for code, operator should specify which version of the software they are coding with.
- For Stata, ChatGPT had difficulties creating tables with proper columns and was most efficient when coding for event studies.
- For R, ChatGPT could complete all 4 steps, although additional specificity was required for each prompt.
- For VBA, ChatGPT accurately coded the initial summary tables but could not complete the full data analysis and recommended using more advanced statistical software.
- ChatGPT 4.0 provided more comprehensive logic and explanations about the outputted code compared to ChatGPT 3.5.

LIMITATIONS

- This was a pilot and exploratory study with a comparatively light methodology in assessing the viability of ChatGPT in producing code.
- There was no external validation of the results and the coders for each language were unique

CONCLUSION

ChatGPT works well within a statistical platform's capabilities. Researchers need to continue learning coding languages to properly create and edit specific prompts

WORKS CITED

1. Jungwirth D, Haluza D. Artificial Intelligence and Public Health: An Exploratory Study. Int J Environ Res Public Health. 2023;20(5):4541. Published 2023 Mar 3. doi:10.3390/ijerph20054541
2. Huang Y, Wu R, He J, Xiang Y. Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. J Glob Health. 2024;14:04070. Published 2024 Mar 29. doi:10.7189/jogh.14.04070
3. Biswas SS. Role of Chat GPT in Public Health. Ann Biomed Eng. 2023;51(5):868-869. doi:10.1007/s10439-023-03172-7
4. Parray, Ateeb & Inam, Zuhrat & Ramonfaur, Diego & Haider, Shams & Mistry, Sabuj Kanti & Pandya, Apurva Kumar. (2023). ChatGPT and global public health: Applications, challenges, ethical considerations and mitigation strategies. Global Transitions. 5. 50-54. 10.1016/j.glt.2023.05.001.

CONTACT

If you are interested in this research, please email Debra Winberg at dwinberg@tulane.edu