Data Science using openAI: testing their new capabilities focused on data science

Jorge Guerra Pires, Ph.D.

Funding: No specific funding was received for this work.
Potential competing interests: No potential competing interests to declare.

Abstract

Even though statistics is taught in several courses, including life sciences, their applications tend to be challenging; basics errors are inevitably done by researchers. It is not uncommon discussions on errors done during statistical analysis done on data by non-experts in statistics, in scientific papers. However, statistics is too important and crucial for science for being left to statisticians only. On this paper, we discuss the possibilities opened by openAI latest API called coder interpreter. This API is able to read CSV files, and perform statistical analysis on the dataset, having as starting point human's commands. Different from classical statistical tools, it is able to choose what method and library to use, get the results back, and give an evidence-based answer. The user need only to know what to ask, which may require minimal knowledge on statistical inference, most likely, covered on any course on statistical analysis. Several areas can benefit from this tool, we think, especially evidence-based medicine, an area where the researchers are also medical doctors. This tools is actually using open source python libraries: the python community was able to build a rich and powerful set of libraries for data science, and this new API from openAI is exploring those libraries. We believe that those tools can be used on real research, and we present a couple of examples, one of them, studying a dataset in diabetic patients. We hope to influence the usage of this tool by non-experts, as well experts, on statistics.

Jorge Guerra Pires[1,*]

Founder at IdeaCoding Lab and JovemPesquisador.com

Keywords: openAI, data science, statistical inference, python, open source, diabetes.

1. Introduction

When chatGPT appeared, started to gain momentum, and people were still trying to understand this new artificial intelligence, it was common comments on social media about their fear of losing their jobs as programmers: they feared that since chatGPT could code, they would no longer be needed (see meme on fig.1). Even today, we cannot understand fully what it can do [1][2].

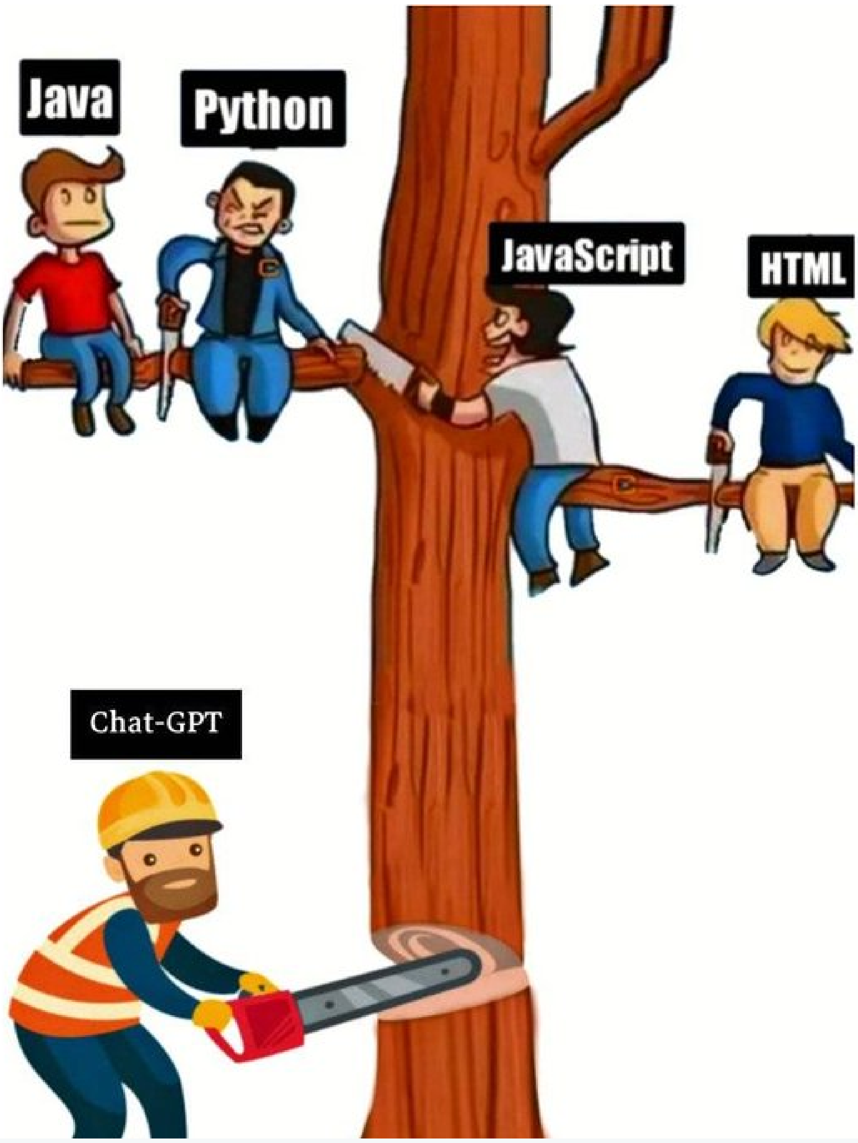

Figure 1. A very popular meme on social media depicting chatGPT as a higher computer

language.

The fear defense mechanism started to work: some said it could not code complex codes, which turned out to be wrong. People started to test chatGPT as a programmer, and the results were impressive. Now, we cannot live without it; and it passed just a couple of months since it gained popularity. Its capabilities range from creating codes from scratch, using APIs, reading documentation, translating codes from one language to other, explaining codes, creating unit testing and more. Therefore, even though their fear of being replaced by AI in coding was real, it was created a new level of coding. Instead of wasting time with basic codes, we can just ask for a code, and make our ideas into codes faster, instead of suffering with bugs, we can just ask chatGPT to solve them. Thus, we are not risking losing our jobs, we are risking losing basic jobs.

We believe this is the future: repetitive works will be assigned to machines, even if they are academic works. As long as they can be described algorithmically, they will be automated by AI. We believe the future of research will change dramatically with AI [3].

What happened is that the human attention shifted from basic coding to abstraction; and fast, just a couple of months. You no longer need to know a computer language in depth to code in this language: with basics of computer programming, and an idea on how to break down your goal into smaller chucks of tasks, you can guide the AI into your final goal.

The case we are going to explore here is even more interesting: it can create the code, run the code, get the results and answer the question. Thus, what would need a human to guide the coding, now was also automated. As we are going to see, the human's task went even higher in abstraction: you provide a dataset (e..g., CVS file with your samples), and you provide a question. The algorithm provided by openAI provides the rest, and then provides the answer. This is what a data scientist would normally do; with the addition of having a "team" for helping them to interpreter the results. What would take an multidisciplinary team, you can solve in a single call to their API. It will not just analyze the dataset, but also providing a nice and interdisciplinary discussion, as we are going to see.

Another interesting observation: the analysis does not limit itself to data science, to statistical inference. One can discuss, and have access to a discussion level that would require an interdisciplinary team. One can use the dataset to make inquiries, or create inquiries based on the analysis. We have added the complete discussions as Supplementary Material for your convenience, and kept the essence on the discussion section, for discussion purpose. Our hope is that we could give the reader a feeling of what they can achieve with this AI on their side.

One interesting fact, for those that are using openAPI API as chatbots: this research we are going to explore, even though can be used as chatbot, it does not work well as chabot. It is perfect for extracting information from datasets, and reading files, but not very good at "conversations"; we have actually done some tests as a possible chatbot. Nonetheless, it can be an excellent assistant as we are going to see, once you have a dataset and want to extract knowledge from it. The amount of information and details it provides can be overwhelming for a chabot, but perfect for a data science analysis.

The paper [4] highlights the crucial role of statistics in research, planning, and decision-making in health sciences. It also discusses the errors that many non-statistician researchers were making in applying statistical methods and how such errors can affect the validity and efficiency of the research conducted. The paper concludes by reflecting on the causes that have led to this situation, the consequences to the advancement of scientific knowledge, and the solutions to this problem.

This paper is divided into fours sections, three remaining. On section 2, we lay down details on how we performed the simulations. Keep in mind that this work is a discussion-focused paper, our method section is mainly for the reader to know the basic on how we arrived to our results and conclusions. On section 3, we discuss what we have found. On section 4, we close the work. We also provide a list of references we found important to cite, and a supplementary material file, which has the discussions on full length as so the curious reader can learn more.

2. Methods

We are going to test some of the just released new capabilities of the openAI API, namely:

- **Assistant** - this is a capability that allows one to build an environment with context, including files. We are going to attach the CSV files for discussion;
- **Files** - with this new option, you can add files, such as CSV files, for adding context to the conversation. For data science, we are interested in CSV files;
- **Code Interpreter** - this new feature will code for you in Python, as so to answer your questions. It is done automatically, no need of human interference. They are running Python codes. This process replicates what a data scientist normally do: they code for testing their doubts, and add the results on their final conclusions. You do not have to know Python, it is done automatically. Except when it cannot solve the problem: it will provide you with Python codes, as so you can run on your own environment;

Those features are focused on programmers: using those features, one can build their own apps, which will have advanced artificial intelligence behaviours, powered by openAI. Nonetheless, all the tests herein were done using their playground, which is an interface that one can open for testing their assistants, including the files, instructions and more that you may eventually add to your assistant. Then, you can just call the assistant from a possible code you may create, with user interface. The advantage of using this playground is testing the assistant without having to create an interface. You just create the assistant and start testing.

The assistant cannot be accessed outside that user's account, therefore, sadly, we cannot share the assistants without building an interface, which we will not do herein.

As a conversational AI algorithm, we are using *gpt-3.5-turbo-1106*. This algorithm is a good trade-off between cost and performance. It is well-known that GPT 4 (*gpt-4-1106-preview*) has a higher capability, which we did not feel we needed it for our analysis. However, it is actually available as option, if the user wants to try out this more advanced option, which costs more.

2.1. Statistical inference

Statistical inference is a largely applied field where one tries to make sense of datasets using statistics, or any similar methodology [5].

For instance, we are going to understand whether diabetes is more likely in men or women. This is done applying a hypothesis test on a dataset that contains people with pre-diabetes. So far, it seems a very strong point of this tool: doing statistical analysis on datasets.

Applying a hypothesis test, even though well-documented on textbooks, can be tricky and error-prone[4]. Statistical analysis generally is not part of the research, it is the means by which we give consistence to our work. For instance, someone studying variations on genetic information does have statistics as major topic, statistics is how they can arrive to a consistent conclusion.

Statistical inference is composed of a set of methodologies, most of them are well-developed mathematically. For instance, one can create regression curves to understand the relations between variables, Principal Component Analysis, correlation and more. Knowing which tool to use may require years of training. Most professions that use statistics have basics of statistics. For instance, medical doctors use it largely, and they have basics on statistics; also, biologists. It is not uncommon seeing those professional talking about statistics, and making basic mistakes [4]. Even well-trained professionals can make statistical mistakes [6].

Knowing which tools to use, and how to interpret the results, can be tricky even for researchers with years on their curricula using statistics.

2.1.1. Hypothesis test

This is a commonly used tool from statistics. One must deny or accept what is called a*null hypothesis*. For instance, you may want to test whether diabetes is more prevalent in men. Then, you can test whether the proportion in a group of men is higher. There are a very well-established theories on that, and can be found on any statistics book [5]. One very famous measure on how well is your hypothesis test is p-value, generally, it has to be lower than 0.05, a scientific consensus.

2.2. Questioning

For using well this new tool from openAI, one needs to have in mind what is a typical workflow for data science, for statistical inference (fig. 2). Even though one does not have to perform the test themselves, the tool will do all, one needs to know at least what to ask. This is a requirement even when you know all the tools well. You cannot use a methodology if you have no idea what is good for, and what you achieve or not achieve by using it.

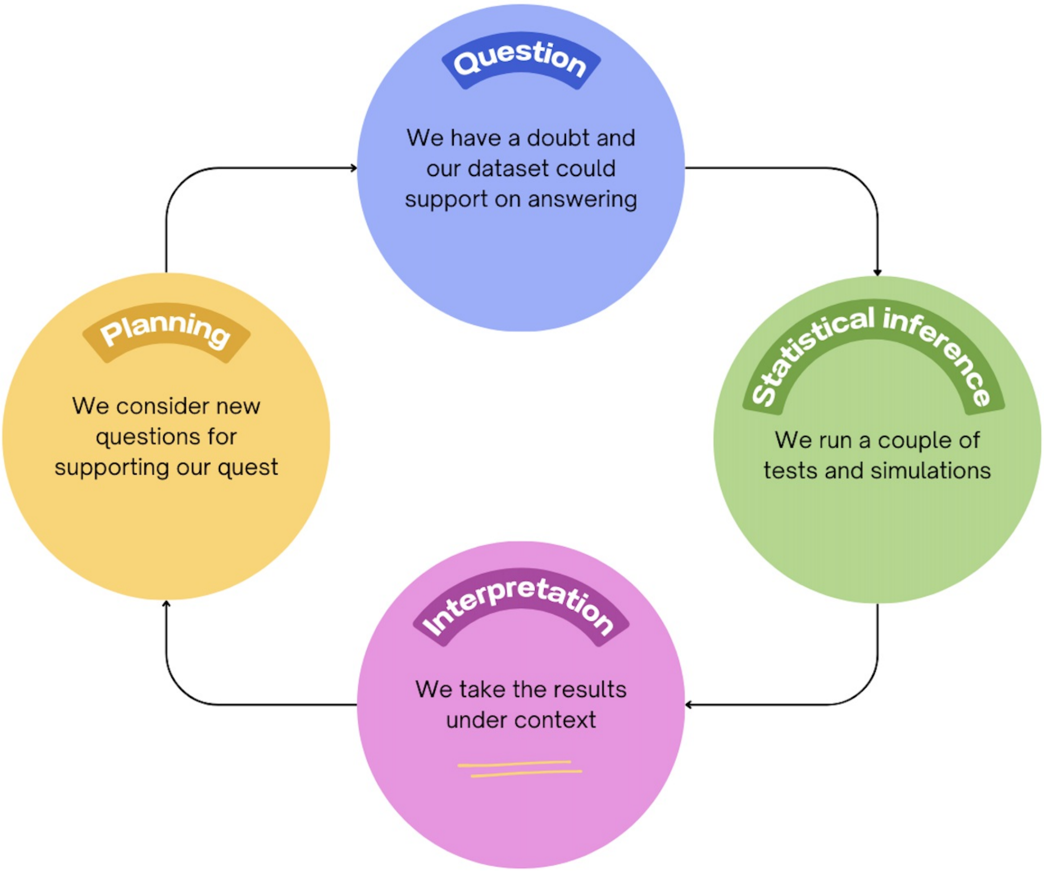

Figure 2. A typical data science workflow

Typically:

1. We have a dataset and a set of questions;
2. We need to know which tests to apply;
3. We need to know how to interpret the results;
4. We arrive to a conclusion, *evidence-based conclusion*.

The openAI tools we are using automate steps 2 and 3. Once you make a question, the tool is able to identify what is the best tests to run. Sometimes you may need to be more specific, but generally, it knows what tests to run. It also can interpret the results, and add information to the interpretation.

2.3. Configurations of the assistants

2.3.1. Understanding the relationship between apparent temperature and real temperature

Instruction given.

> *Attached is a set of observations of the Apparent Temperature (C) and the Temperature (C) in CSV, also other measurements. Your job is to analyze this dataset. and answer my questions based on this analysis.*

The dataset we are using is here in CSV. It was attached to the conversation. See part of the conversation in attachment.

Kaggle provides several datasets, with different levels of quality: we have chosen this one without too much concerns on quality since our goal herein is presenting the new openAI API for data scientist; instead of actually concerning about the datasets used.

The configurations for the other assistants were the same: just changing the dataset. For algorithm, we have used *gpt-3.5-turbo-1106*.

3. Results and Discussion

On this section, we are going to present our findings. See that our focus is on the coder interpreter, not on the datasets themselves. Therefore, we are not concerned whether what we found is novel, or whether we should make a literature review on each finding.

3.1. Understanding the relationship between apparent temperature and real temperature

The first conversation was about the relationship between real temperature and other factors that create the apparent temperature, the perceived one. When you feel a temperature, it is influenced by several factors: such as humidity and wind. We are going to explore this dataset, but using the *coder interpreter* instead of manually doing it, as we did on previous works.

It was a long and interesting conversation with their interpreter model: as I asked, it would plot graphs and bring me augmentations to either support my claims or defy them, using the dataset. I have used the playground they provided, which required no coding, for communication with their algorithm. Keep in mind that those algorithms are for APIs, that is, for programmers to build their apps on top of them. Even though they have not yet ruled out to everyone, it seems GPTs can do this and more, with no need of coding.

For example, I had the impression that when the humidity is high, that would increase the heat transference, therefore, the apparent temperature will be lower. I was wrong, Code Interpreter, using the dataset I have provided, convinced me I was forgetting that high humidity makes it harder for the movement of hot air. Whereas, I have asked about if the wind is moving, windy condition. And, now, I was proven right!

What is interesting is this ability to be used to prove hypothesis, to defy them with data. This is what a scientist in general will do, and certainly a data scientist will have to do. This is a new level of artificial intelligence: the ability to use data and knowledge to either confirm or deny an assertion. This is a scientific research being taken to higher levels, where the scientist can focus on asking questions, and providing data.

It could be particularly interesting for *evidence-based medicine*, a new field in medicine focused on data, on evidence. A medical doctor can use evidence, beyond scientific papers, without having a strong background in statistical inference. See our section about diabetes, section 3.2.

The most promising application in our opinion of this tool is that it was proven that researchers tend to be influenced on topics that are politicized, such as gun policies. It could support those researchers in not allowing their positions influencing their analysis. As one example where chatGPT has been used to fighting human biases is for recommendation letters [7]. We have been using to gather information when writing for avoiding internal biases, for avoiding that our current opinions and expectation will interfere on the search. We believe this new tool could also help data scientists escape from internal ideologies and expectations upon the dataset. That is, *confirmation bias*. One should never overlook human biases on interpreting scientific research [8][9][10][11][12].

See that we are not saying those large language models (LLMs) have no biases, or ideologies, because they have [13][14][15]. We are saying their biases are less likely to be individual, or even, cultural. Which is a step forward a more neutral scientific endeavour. All the biases and ideology they may eventually have must be a pattern on the dataset, that is, it must be global, not something very specific from the researcher, or their local environment.

One interesting part of the conversation was when we asked for creating a neural network, for fitting the dataset using all the features as input, even though they have calculated a correlation of about 99% just between apparent temperature and temperature. The code interpreter created a neural network, doing all the procedures, including splitting the dataset into testing and training. Once it created the neural model, we asked for a prediction, which was close. We also questioned the correlation of 99%: it is a linear measure, thus, this correlation cannot be trusted totally. Indeed, it confirmed the guess, and suggested ways to mitigate this possible misinterpretation of the correlation coefficient when dealing with nonlinear relationships.

The only limitation we noticed: we wanted a model using TensorFlow, it could not create. The algorithms they are using are predefined. Also, it seems to be stochastic the behaviour: sometimes the algorithm can be useful, in other times, even keeping the same assistant with the same configuration, it cannot help as it did before. One explanation is the stochasticity present on those LLMs, which is well-known [2].

All this magic is possible because Python has a rich set of public/open source libraries largely used in data science[16][17]. What they are doing is exploring this rich data science environment created by the data science community. What is curious: this move makes more sense than any other move from openAI since originally was to be open source driven [1].

One workaround on this issue of having to be limited by the tools they have using Python is actually defining your own external functions, and they allow in addition to attach files, to attach signature to external functions. On this case, you need to run the assistant from you own environment; we are using their playground. In fact, they allow you to host your functions on their servers, with a fee for that service.

One question we did as a result of this experience is what will happen to proprietary software. They have already been losing attention since open-source data science software gained traction.

Open-source software has gained significant traction in the field of statistics in recent years. Proprietary software such as SAS, SPSS, Minitab, and Stata, which were once the dominant players in the market, have been losing market share to open-source platforms such as Python and R [18]. The open-source movement has brought about a paradigm shift in the way statistical programming is done. Open-source languages such as Python and R are freely available, constantly updated, and enjoy near-instant worldwide distribution. The most significant difference between these new products and the proprietary hegemons of analytical programming is that open-source languages' source codes are freely available for modification by any user [18].

This approach departs markedly from the traditional software development model, where the best minds from computational statistics or social science are hired, their talents are concentrated at or near corporate headquarters, and professionally developed source code is jealously guarded [18].

While both open-source and proprietary software have their costs and benefits, open-source software has several advantages over proprietary software. For instance, open-source software is freely available, constantly updated, and enjoys near-instant worldwide distribution. Additionally, open-source software has a lower organizational cost in the long run compared to proprietary software [19].

The reason coder interpreter works is thanks to the open source community: they can access those open source libraries, and make what you see happens. This is just possible because open source became very powerful, powerful enough to shadow classical proprietary software. One drawback of those open-source libraries is that they are harder to use, compared to proprietary software that generally provide a more friendly interfaces. Is there a better interface than a chatbot? I guess not. You just make questions and assertions, and you will receive an educated response. This is what code interpreter may do in the future of data science, when using open source data science software as it is doing now.

Imagine if we could have an open source function calling file, that would be something to aim for. All the data science libraries in a single chatbot, no need to read documentations and implement the codes we need to run them, and get the analysis done.

3.2. Understanding diabetes

Diabetes is a medical condition largely studied. We have created neural network models for it on our previous paper[20]. We are going to explore the same dataset. See supplementary material for a complex version of the conversation with code interpreter.

One interesting fact is that the model can understand what each columns means, as long as you add meaningful names. It can creates a description from each column. Generally, if you go to their Kaggle Notebook, this is what generally we do before modeling the dataset: we have to know better the dataset.

Can we say diabetes is more likely on a specific gender?

> ***DiabetesBot***: *To test whether diabetes is more prevalent in men or women, we can conduct a hypothesis test to compare the proportion of diabetic patients between the two genders. We will use a two-sample hypothesis test for proportions to determine if there is a significant difference.*

What is interesting: I have not mentioned anything about hypothesis test, which is the correct path to follow to answer my inquiry. Also, amongst the possible hypothesis test, this is the correct one. There are several of them [5].

DiabetesBot: The p-value for the two-sample proportion hypothesis test comparing the prevalence of diabetes between men and women is approximately 1.24e-32, which is extremely low. **This indicates strong evidence against the null hypothesis (H0) that there is no difference in the proportion of diabetic patients between men and women**.

Therefore, ***we reject the null hypothesis in favor of the alternative hypothesis***(H1) that there is a difference in the prevalence of diabetes between men and women. This suggests that **the prevalence of diabetes differs significantly between genders in the given dataset**.

Which group is most likely?

> ***DiabetesBot***: *The proportion of diabetic patients is approximately 7.6% for women and 9.7% for men in the given dataset. Therefore, based on these proportions, men are more likely to have diabetes compared to women.*

See that the difference between men and women is small: it could be between the standard deviation, and be a meaningless difference. But we already know there is a difference, we just needed to find it.

But, we can confirm it.

> ***DiabetesBot***: *The 95% confidence intervals for the proportions of diabetic patients in the given dataset are approximately:*
>
> - *For women: 7.41% to 7.84%*
> - *For men: 9.47% to 10.04%*
>
> *These intervals provide a range of values within which we can be 95% confident that the true proportion of diabetic patients lies for each gender.*

That is, with a confidence of 95%, they do not cross each other, making the proportion statistically signiticant.

Even though different in numbers, it was also found by[21] that diabetes are more present in men. Therefore, this study confirms our finding.

We now want to know which feature best predict diabetes[20]. We already know from previous work.

> ***DiabetesBot**: The correlation coefficients between the features and the "diabetes" column are as follows:*
>
> 1. *Blood Glucose Level: 0.4196*
> 2. *HbA1c Level: 0.4007*
> 3. *Age: 0.2580*
> 4. *BMI (Body Mass Index): 0.2144*
> 5. *Hypertension: 0.1978*
> 6. *Heart Disease: 0.1717*
>
> *The features with the highest correlations (in absolute value) with predicting diabetes are "Blood Glucose Level" and "HbA1c Level," indicating that these two features are strongly associated with predicting diabetes in the dataset. These results suggest that blood glucose level and HbA1c level may be important indicators for predicting the presence of diabetes in the given dataset.*

Which is what we have found previously on this same dataset.

3.3. Should you buy a house in Boston?

A very nice application of data science is to gain insight on information. A very famous historical dataset tells us whether a house is over or underpriced in Boston [22][23][24].

> *"If we were the only ones in the world with access to this info, we could be the best Boston real-estate investors in 1978! Unless, somehow, **someone were able to build an even more accurate estimate**. . ." [highlight added]* [23]

Our goal is to learn as much as we can from this dataset; you can actually build neural network models using this dataset [22][23][24].

3.3.1. Would the criminality effect the house pricing?

I guess we all say yes. Nonetheless, let's see what the dataset says. We have learnt from Daniel Kahneman that we are bad at statistics. Let's see what the numbers say.

> *The correlation coefficient between the median value of owner-occupied homes (MEDV) and the per capita crime rate by town (CRIM) is approximately -0.3883.*

This is about 38% correlation. According to Kahneman[25], we can interpreter correlation as how much one variable can explain the other. That is, about 40% of the variation in the price of the house is due to crime rate. This is an expressive number for a single variable to explaining a multi-factor dependent variable. You may lose almost half the price of a house if the criminality rate increases on your neighborhood, or gain on a bought house if you manage to support on crime rate drop down. Sounds like an evil plan for gaining on house trading!

What about if we try a hypothesis test?

> ***BrokerBot****: The hypothesis test comparing the median home values between regions with crime rates above the average and regions with crime rates below the average yields the following results:*
>
> - *The t-statistic is approximately -9.70.*
> - *The p-value is approximately 1.16e-18, indicating that it is extremely small.*
>
> *Based on these results,* ***we reject the null hypothesis*** *and conclude that there is a statistically significant difference in median home values between regions with crime rates above the average and regions with crime rates below the average. Specifically, the median home values are lower in regions with crime rates above the average.*

In case you have forgotten statistical inference, like myself:

> *The p-value is the probability of obtaining a test statistic as extreme or more extreme than the one observed in the sample data, assuming that the null hypothesis is true. If the p-value is less than the significance level (usually 0.05), we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.*

> *The t-statistic is used to determine whether to support or reject the null hypothesis. It is the ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error. The best value of t-statistic is the one that maximizes the difference between the null hypothesis and the alternative hypothesis 2.*

Therefore, we can confirm that criminal rate will decrease the value of the house; of course, we knew that *a priori*. It is one of those researches to conclude the obvious!

3.3.2. Would the number of teachers per student effect the house pricing?

We know that education is something very important for the rich people. They value a lot the education of their childreen. Thus, would the number of available teachers per student influence the house pricing? I would guess so. Furthermore, that would also measure other factors. Like the number of fish on the water measures how much oxygen it has, the number of teachers per student also measures the quality of the life around. Poor places have difficulties to retain teachers, I would guess.

> *The correlation coefficient between the median value of owner-occupied homes (MEDV) and the pupil-teacher ratio by town (PTRATIO) is approximately -0.5078.*

Which means: about 50% of house pricing is explained by the ratio between teachers per student. The lower the rate, the better. Remember from Kahneman [25] that correlation measure how much one variable explain the other; it does not necessarily mean that is a causality. Correlation is not causality.

Note. This dataset is from 1978, recently, we are facing shortage of professors worldwide. With current situation, we should be careful with this conclusion.

3.4. Aristotle and the skilled questioner

Even thought code interpreter can do miracles, it seems to have a weakness: you need to know what to ask. This is similar to what Aristotle believed when teaching someone: you need to be a skilled questioner. Even though if you ask, it will do everything automatically, you need to know what to ask for getting information from the dataset you have. This is just possible if you are already minimally familiar with *statistical inference*. Thus, we believe this tool will be most efficient at the hand of data scientists minimally trained on statistical inference, and related areas that make up data science.

Aristotle believed that knowledge is not something that can be taught, but rather something that is already present within us. He believed that the role of the teacher is to help students bring this knowledge to the surface through a process of questioning and reflection. In other words, Aristotle believed that everyone has the potential to be a philosopher, but that we need to be skilled questioners in order to uncover the knowledge that is already within us. The same way, with the current tool, anyone can be a data scientist, as long as they get a minimal training in statistical inference.

This idea is closely related to Aristotle's concept of epagoge, or "induction." According to Aristotle, induction is the process of moving from particular instances to general principles. By asking questions and reflecting on our experiences, we can begin to identify patterns and make generalizations about the world around us .

In summary, Aristotle believed that knowledge is not something that can be taught, but rather something that is already present within us. By asking questions and reflecting on our experiences, we can begin to uncover this knowledge and develop a deeper understanding of the world around us.

4. Conclusion

We have explored a new tool from openAI called coder interpreter. This tool is able to find knowledge from datasets. We have shown that the tool can make basic statistical inference automatically and with guidance, even go further, like the Aristotle method: you ask to take knowledge out of a dataset. This tool can be particularly interesting to data scientists already familiar with statistical inference. We have in a problem from medicine, making the tool particularly interesting to medical doctors: once they have a basics on statistics, they can perform statisical analysis on their datasets effortless, enforcing evidence-based medicine.

References

1. [a, b] *HAO, K. The chaos inside OpenAI – Sam Altman, Elon Musk, and existential risk explained — Karen Hao. Big Think [YouTube Channel], 2023. Accessed on 2 Dec 2023. Disponível em: ⟨https://www.youtube.com/watch?v=O9sLCp2Jq74&t=3333s⟩.*
2. [a, b] *WOLFRAM, S. What Is ChatGPT Doing... and Why Does It Work? 2023. https://writings.stephenwolfram.com/2023/02/whatis-chatgpt-doing-and-why-does-it-work/.*
3. [^] *PIRES, J. G. O mercado da criatividade: Regulamentaçao da profissão de pesquisador acadêmico e científico no Brasil. 2023. Disponível em: ⟨https://www.amazon.com.br/mercado-criatividade-Regulamenta% C3%A7%C3%A3o-pesquisador-cient%C3%ADfico-ebook/ dp/B09TKRX5PW⟩.*
4. [a, b, c] *BULLARD, K. M. et al. Prevalence of diagnosed diabetes in adults by diabetes type — united states, 2016. Morbidity and Mortality Weekly Report, US Department of Health and Human Services, Centers for Disease Control and Prevention, v. 67, n. 12, p. 359, 2018.*
5. [a, b, c] *MONTGOMERY, D. C.; RUNGER, G. C. Applied Statistics and Probability for Engineers. [S.l.]: John Wiley & Sons, 2010.*
6. [^] *KAHNEMAN, D. Thinking, Fast and Slow. [S.l.]: Farrar, Straus and Giroux, 2011.*
7. [^] *LEUNG, T. I. et al. Can ai mitigate bias in writing letters of recommendation? JMIR Medical Education, JMIR Publications Inc., Toronto, Canada, v. 9, n. 1, p. e51494, 2023.*
8. [^] *MAY, J. Bias in science: natural and social. Synthese, Springer, v. 199, n. 8, p. 3345–3366, 2021.*
9. [^] *TODAY, P. Recognizing PoliticallyBiased Social Science. 2021. ⟨https://www.psychologytoday.com/us/blog/psych-unseen/202112/recognizing-politically-biased-social-science⟩.*
10. [^] *ACADEMY, E. Dealing With Bias in Academic Research. 2022. ⟨https://www.enago.com/academy/dealing-with-bias-in-academic-research/⟩.*
11. [^] *LEARNING, S. N. How bias affects scientific research. ⟨https://www.sciencenewsforstudents.org/article/how-bias-affects-scientific-research⟩.*
12. [^] *UNIVERSITY, S. Ideological Bias and Trust in Information Sources. ⟨https://web.stanford.edu/~gentzkow/research/biasmeas/biasmeas.pdf⟩.*
13. [^] *GALLEGOS, I. O. et al. Bias and fairness in large language models: A survey. arXiv preprint arXiv:2109.05252, 2021.*
14. [^] *LI, Y. et al. Generative language models exhibit social identity biases. arXiv preprint arXiv:2104.04451, 2021.*
15. [^] *SUN, H. et al. Aligning with whom? large language models have gender and racial biases in subjective nlp tasks. arXiv preprint arXiv:2104.07693, 2021.*
16. [^] *Kumar, A.; Kumar, S. Python for data analytics, scientific and technical applications. In: 2019 Amity International Conference on Artificial Intelligence (AICAI). [S.l.: s.n.], 2019. p. 1–5.*
17. [^] *Kumar, M.; Kumar, S. An overview and comparison of free python libraries for data mining and big data analysis. In: 2018 3rd International Conference on Computing Methodologies and Communication (ICCMC). [S.l.: s.n.], 2018. p. 1–*

5.

18. [a, b, c] *HEARN, E. Is proprietary software better than opensource? Built In, June 2020. Disponível em: ⟨https://builtin.com/data-science/open-source-proprietary-software⟩.*
19. [^] *CASTILLO, L. Open Source Software Statistics [Fresh Research]. 2023. Disponível em: ⟨https://blog.gitnux.com/open-source-software-statistics/⟩.*
20. [a, b] *PIRES, J. G. Machine learning in medicine using javascript: building web apps using tensorflow.js for interpreting biomedical datasets. medRxiv, Cold Spring Harbor Laboratory Press, 2023. Disponível em: ⟨https://www.medrxiv.org/content/early/2023/07/09/2023.06.21.23291717⟩.*
21. [^] *BULLARD, K. M. et al. Prevalence of diagnosed diabetes in adults by diabetes type — united states, 2016. Morbidity and Mortality Weekly Report, US Department of Health and Human Services, Centers for Disease Control and Prevention, v. 67, n. 12, p. 359, 2018.*
22. [a, b] *PIRES, J. G. Discussing a multiple regression model. 2022. Disponível em: ⟨https://medium.com/computational-thinking-how-computers-think-decide/discussing-a-multiple-regression-model-c1f45ce52211⟩.*
23. [a, b, c] *NIELSEN, S. C. E. D.; BILESCHI, S. M.*
24. [a, b] *PIRES, J. G. Computational Thinking: How computers think, decide and learn, when human limits start and computers champ, vol. 1. 2022. Disponível em: ⟨https://books.google.com/books/about/Computational_Thinking_How_computers_thi.html?id= WfV6EAAAQBAJ⟩.*
25. [a, b] *KAHNEMAN, D.; SIBONY, O.; SUNSTEIN, C. R. Noise: A Flaw in Human Judgment. [S.l.]: William Collins, 2021.*