

NOTES ON USING COPILOT FOR TEACHING STATISTICS IN CZECH

Nikola Kaspříková

Abstract

With the recent development of generative artificial intelligence tools based on large language models and their increasing availability, there is an opportunity to utilize these tools in statistics and data analysis education. The paper focuses on exploring and evaluating some of the current possibilities of using Bing chat with Copilot for solving tasks specified by users in the Czech language.

A couple of basic use cases in data science and statistics are explored in the paper. Namely the explanation of the p-value, which is still one of the basic concepts in the introductory statistics courses, then a problem-solving use case involving the calculation of the mean time before failure and finally a use case which is based on a request to generate the code in R statistical software.

Considering the results obtained in the case studies discussed in this paper and supposing that the tools with generative AI are going to be further developed and fine-tuned, the output is rather promising and may be supposed to provide valuable support for the students learning statistics and data science.

Key words: statistics education, Copilot, chatbot

JEL Code: C80, C88

Introduction

With the recent development of generative artificial intelligence tools based on large language models and their increasing availability, there is an opportunity to utilize these tools in statistics and data analysis education. As reflected in (Campos Trinidad et al., 2021), there is a high demand for data science learning support as well as an increased demand for online assistance systems in education which may provide the requested support even when not restricted to particular schedule.

The assistance systems for data science have been developed recently using the deep learning models and advanced natural language processing techniques. The domain specific systems,

focused on the data analysis, include for example the chatbot described in (Carlander-Reuterfelt et al., 2020). Ellis and Slade in (Ellis and Slade, 2023) discuss the usage of a more general tool, namely the ChatGPT system to support teaching statistics and data science. The authors also mention the recent discussions and concerns about the risks of using such tools associated with the potential to enhance plagiarism and undermine independent thinking of the students. For a detailed discussion of such problems see also (Gallent Torres et al., 2023), (Sun and Hoelscher, 2023) or (Chan, 2023). The paper (Xing, 2024) provides a recent evaluation of the possibilities ChatGPT provides in supporting teaching statistics, considering both high responsiveness of the tool and the concerns about the correctness of the answers.

The literature considering the application of generative AI tools for support in education in Czech language is rarer and includes the paper focused on mathematics education (Víta, 2023). This paper addresses the application of Bing.com chat with Copilot to provide support students of statistics and data science in Czech language. Bing.com chat with Copilot is rather general and is highly available. A couple of basic use cases in data science and statistics are explored in the paper. Namely the explanation of the p-value, which is still one of the basic concepts in the introductory statistics courses, then a problem-solving use case involving the calculation of the mean time before failure and finally a use case which is based on a request to generate the code in R statistical software (R Core Team, 2024) to load the dataset and build a regression model.

1 Concept explanation

Many requests by the students aim at obtaining the definition or explanation of some concepts used in statistics. The tools based on large language models may provide a better user experience compared with the output provided by traditional search engines. The first request tested in this study is a request about what is p-value (in Czech “co znamená p-hodnota?”) and the response generated by Copilot is shown in Figure 1.

The response generated by the model uses websites such as wikiskripta.eu, the educational website which has been developed within project of the First Faculty of Medicine of the Charles University and is economically supported by EU and national development projects. The response in English, using automated translation in Copilot is as follows: “The p-value is a statistical indicator that indicates how likely it is that the achieved (or even more significant) result (e.g., the difference observed between two groups) could have occurred by chance.

The p-value is a number between 0 and 1, with a smaller value indicating a greater likelihood of refuting the so-called null hypothesis (i.e., the assertion that there is no difference between the groups). A (agreed) value of 0.05 (5%) is usually used as the boundary, and if the p-value is less than 0.05, the scientific result is considered statistically significant, and the null hypothesis is rejected.

Therefore, the p-value, in simple terms, means what is the probability that the test criterion would reach its value, or values even more indicative against, if the null hypothesis really held. The smaller the p-value, the less likely a result (assuming the validity of the null hypothesis) was achieved.”

Fig. 1: Response to request to explain p-value

P-hodnota je statistický ukazatel, který udává, jak pravděpodobné je, že by dosažený (nebo ještě výraznější) výsledek (např. rozdíl pozorovaný mezi dvěma skupinami) mohl vzniknout náhodou ¹.

P-hodnota je číslo mezi 0 a 1, přičemž menší hodnota znamená větší pravděpodobnost vyvrácení tzv. nulové hypotézy (tj. tvrzení, že mezi skupinami není žádný rozdíl) ¹. Jako hranice se obvykle se používá (dohodnutá) hodnota 0,05 (5%), přičemž, pokud je p-hodnota menší než 0,05, vědecký výsledek se považuje za statisticky významný a nulová hypotéza se zamítá ¹.

P-hodnota tedy znamená, zjednodušeně řečeno, jaká je pravděpodobnost, že by testovací kritérium dosáhlo své hodnoty, případně hodnot ještě více svědčících proti, pokud by opravdu platila nulová hypotéza ². Čím menší p-hodnota, tím nepravděpodobnějšího výsledku (za předpokladu platnosti nulové hypotézy) bylo dosaženo ².

Další informace ¹ [wikiskripta.eu](#) ² [cs.wikipedia.org](#) ³ [cs.wikipedia.org](#) ⁴ [bing.com](#) [+1 další](#)

Source: Bing.com chat with Copilot

Fig. 2: Response to additional request to explain p-value

Samozřejmě, rád vám poskytnu podrobnější vysvětlení p-hodnoty.

P-hodnota je statistický ukazatel, který se používá v hypoteticko-deduktivním přístupu k testování statistických hypotéz. Je to pravděpodobnost, že bychom pozorovali data tak extrémní, nebo ještě extrémnější, pokud by platila nulová hypotéza.

Nulová hypotéza je předpoklad, že mezi sledovanými proměnnými neexistuje žádný vztah nebo rozdíl. Například, pokud testujeme účinnost léku, nulová hypotéza by mohla tvrdit, že lék nemá žádný účinek.

Pokud je p-hodnota menší než předem stanovená hladina významnosti (často 0,05), pak zamítáme nulovou hypotézu. To znamená, že data poskytují dostatečný důkaz pro to, aby bylo možné prohlásit, že existuje statisticky významný rozdíl nebo vztah.

Je důležité si uvědomit, že p-hodnota neříká nic o velikosti nebo praktické významnosti rozdílu nebo vztahu. Také neříká nic o pravděpodobnosti nulové hypotézy. Je to pouze míra statistické významnosti.

Doufám, že toto vysvětlení je užitečné! Pokud máte další otázky, neváhejte se zeptat.

Source: Bing.com chat with Copilot

If the user is not satisfied with the answer, it is possible to ask for more information or a better explanation. The answer to the request “try to improve the answer” (in Czech “zkus odpověď vylepšit”) is shown in Figure 2, in answer in English is “The p-value is a statistical indicator used in the hypothetico-deductive approach to testing statistical hypotheses. It is the probability that we would observe data as extreme, or even more extreme, if the null hypothesis were true.

The null hypothesis is the assumption that there is no relationship or difference between the variables being observed. For example, if we are testing the effectiveness of a drug, the null hypothesis might assert that the drug has no effect.

If the p-value is less than a predetermined significance level (often 0.05), then we reject the null hypothesis. This means that the data provide sufficient evidence to declare that there is a statistically significant difference or relationship.

It is important to note that the p-value says nothing about the size or practical significance of the difference or relationship. It also says nothing about the probability of the null hypothesis. It is merely a measure of statistical significance.”

The results obtained may be considered more valuable than the results obtained using traditional search engines, even though the time to generate the response is higher compared to the time needed to handle the request by the traditional search engines.

2 Problem solving

The calculation of an expected value of a random variable is a common problem to be solved in probability and statistics courses. The solution often requires basic knowledge of calculus.

Fig. 3: Response to request to calculate mean time before failure

Střední hodnota (také známá jako očekávaná hodnota) náhodné veličiny je definována jako

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

kde

$$f(x)$$

je hustota pravděpodobnosti náhodné veličiny

$$X$$

. V tomto případě je

$$X$$

doba bezporuchového provozu stroje a

$$f(t) = 0.2e^{-0.2t}$$

pro

$$t > 0$$

.

Takže střední hodnota doby bezporuchového provozu stroje je

$$E[T] = \int_0^{\infty} t \cdot 0.2e^{-0.2t} dt$$

Tento integrál lze vyřešit pomocí metody per partes (integrace per partes), kde

$$u = t$$

a

$$dv = 0.2e^{-0.2t} dt$$

. Po provedení integrace dostaneme

$$E[T] = 5$$

Takže střední doba bezporuchového provozu stroje je **5 jednotek času**.

Source: Bing.com chat with Copilot

The request to calculate mean time before failure for a machine with specified probability density function of the failure-free operating time (in Czech and using TeX notation for mathematical expressions “Doba bezporuchového provozu stroje, je náhodná veličina určená hustotou pravděpodobnosti definovanou jako $f(t) = \begin{cases} 0.2 e^{-0.2t} & \text{pro } t > 0 \\ 0 & \text{pro } t \leq 0 \end{cases}$. Spočti střední hodnotu doby bezporuchového provozu stroje.”) resulted in the output shown in Figure 3.

The response states a general formula for calculating the expected value first, then it calculates the expected value using the particular probability density function provided in the

request by the user. Finally, the conclusion is provided stating that the mean time before failure is 5 units of time. If asked to provide more details about how the integration was performed, Copilot makes the calculation more verbose and detailed. So, it can be concluded that the performance of the tool in this use case is good and may be quite helpful even for the students struggling with calculus.

Another task tested was the request to generate a contingency table from a two columns dataset and analyze it. The tool is able to produce the contingency table and provide general recommendations about chi square tests and in is able to generate R or python (Python software foundation, 2024) code which loads the particular user data and performs the hypothesis test. It does not seem to be able to produce the chi square test output.

3 Producing R code

The ability to load the data set and analyze the data in statistical software is crucial for data analysis. The request to write an R code to load the text file which has the columns separated with semicolon (the prompt in Czech was “napiš kód v R, který načte textový soubor se sloupci oddělenými středníkem”) resulted in a satisfactory output, which used the readr R extension package. After a request to solve the problem without using the readr package (in Czech the request was “udělej to bez použití readr”), Copilot generated the answer shown in Figure 4.

After loading the dataset, a model may be needed. The request to generate an R code for building a regression model to predict probability of problems based on age, GMI, gender and smoking (in Czech the prompt was “napiš v R kód pro regresní model předpovídající pravděpodobnost obtize v závislosti na vek, BMI, pohlavi a koureni”) resulted in the output shown in Figure 5. The prompt not using the term probability (in Czech “pravděpodobnost”), resulted in a code producing a linear regression model.

The code generated by the tool together with the comments provides useful guidance to the user in this case, even if the user were not familiar with R syntax.

Fig. 4: Response to request generate R code to load the dataset

Jistě, zde je kód v R pro načtení textového souboru se sloupci oddělenými středníkem bez použití knihovny `readr` :

```
R
```

```
# Cesta k souboru
cesta_k_souboru <- "cesta/k/vasemu/souboru.txt"

# Načteme soubor
data <- read.table(cesta_k_souboru, sep = ";", header = TRUE)

# Zobrazíme prvních pár řádků dat
head(data)
```

Kód vygenerovaný AI. Pečlivě si je projděte a používejte. [Další informace v nejčastějších dotazech.](#)

Poznámka: Nahradte `"cesta/k/vasemu/souboru.txt"` skutečnou cestou k vašemu textovému souboru. Tento kód předpokládá, že máte nainstalovaný R. Pokud ne, můžete ho nainstalovat z oficiálních stránek R.

Source: Bing.com chat with Copilot

Fig. 5: Response to request generate R code to build a classification model

```
# Předpokládejme, že máme data uložená v dataframe s názvem 'data'
# a 'obtize' je binární proměnná označující, zda pacient má obtíže (1) nebo ne (0)

# Načteme potřebnou knihovnu
library(glm)

# Převědeme pohlaví a kouření na faktory
data$pohlavi <- as.factor(data$pohlavi)
data$koureni <- as.factor(data$koureni)

# Vytvoříme model
model <- glm(obtize ~ vek + BMI + pohlavi + koureni, family = binomial(link = 'logit'), data = data)

# Vypíšeme shrnutí modelu
summary(model)
```

Source: Bing.com chat with Copilot

Conclusion

The answers generated to the requests considered in this paper may generally be considered satisfactory and are definitely helpful to students learning basics of statistics and data science seeking support online. Regarding the speed of generating the answers, the answers generated in Czech seem to be generated slightly more slowly compared with queries in English.

Considering the results obtained in the case studies discussed in this paper and supposing that the tools with generative AI are going to be further developed and fine-tuned, the output is rather promising and may be supposed to provide valuable support for the students learning statistics and data science subjects. The application of the tools based on large language models should be included in statistics and data science courses, including the discussion of the results and sharing best practices in regular classes.

References

- Campos Trinidad, M. J., Coelho, F. B., Arauco Canchumuni, S. W., & Cavalcanti Pacheco, M. A. (2021). Data science mentoring support using neural networks. (2021). *21st International Conference on Computational Science and Its Applications (ICCSA)*. <https://doi.org/10.1109/iccsa54496.2021.00045>
- Carlander-Reuterfelt, D., Carrera, A., Iglesias, C. A., Araque, O., Sanchez Rada, J. F., & Munoz, S. (2020). Jaicob: A data science chatbot. *IEEE Access*, 8, 180672–180680. <https://doi.org/10.1109/access.2020.3024795>
- Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for CHATGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Gallent Torres, C., Zapata González, A., & Ortego Hernando, J. L. (2023). El Impacto de la Inteligencia Artificial Generativa en Educación Superior: Una mirada desde la ética y la Integridad Académica. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 29(2). <https://doi.org/10.30827/relieve.v29i2.29134>
- Chan, C. K. (2023). A comprehensive AI Policy Education Framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00408-3>
- Python software foundation (2024). Python Language Reference. [Computer software]. Retrieved from www.python.org

R Core Team. (2024). R: A language and environment for statistical computing. [Computer software]. Retrieved from www.r-project.org

Sun, G. H., & Hoelscher, S. H. (2023). The CHATGPT storm and what faculty can do. *Nurse Educator*, 48(3), 119–124. <https://doi.org/10.1097/nne.0000000000001390>

Víta, M. (2023). Srovnání chatbotů založených na velkých jazykových modelech z perspektivy základních kurzů matematiky na VŠE. *Mundus Symbolicus*, 31, 31–40.

Xing, Y. (2024). Exploring the use of chatgpt in learning and instructing statistics and data analytics. *Teaching Statistics*, 46(2), 95–104. <https://doi.org/10.1111/test.12367>

Contact

Nikola Kaspříková

Department of Mathematics, Prague University of Economics and Business

W. Churchill Sq. 1938/4, Prague 3

nikola.kasprikova@vse.cz