CAUSAL DIAGRAMS FOR DESCRIPTIVE STATISTICS

Karsten Lübke and Matthias Gehrke
FOM University of Applied Sciences
karsten.luebke@fom.de

Without random sampling and/or random allocation, even descriptive statistics such as simple means or proportions can be quite misleading. Therefore, causal diagrams were added to existing course materials to address this topic and to illustrate the differences between random and convenience samples and between observational and experimental studies. We assessed student understanding in different courses with a pre-/post-survey. Additionally, we asked students to evaluate the helpfulness of the diagrams for their understanding. There is a statistically discernible positive effect with 280 students from more than seven different courses on pre- to post-knowledge. Also, most of the students agreed with the statement that the causal diagrams helped in their understanding.

MOTIVATION

Nowadays, data is everywhere. Statistics and data science education does not only aim to help students analyze data but also aims to help students learn from data for a problem at hand. Hernán et al. (2019) distinguish three different tasks of data science: description, prediction, and causal inference. Many introductory courses may only cover the first task, description, but as Greenland (2022) pointed out, causality might even be central for description. In her list for promoting statistical literacy, Utts (2021) also emphasizes the topics of observational studies, confounding, and causation. Statisticians are aware of the magic of randomness. As pointed out by, for example, Cobb (2007), we randomize data to protect against bias and to provide a basis for inference (p. 13). To make inferences, we draw random samples to generalize to populations and use random assignment to support conclusions about cause and effect (p. 13). The protection against bias by randomization is easy to depict with causal diagrams. Both random sampling and random assignment are erasing arrows pointing into the sampling or treatment variable in a causal diagram.

For some time now, there has been a call to include causality in statistics and data science curricula (e.g., Cummiskey et al., 2020; Greenland, 2022; Kaplan, 2018; Lübke et al., 2020; Schield, 2018). In the current study, we investigate if causal diagrams, even presented in a very informal way, could help students to draw appropriate conclusions. Therefore, we try to add to the available empirical evidence given by, for example, Ellison (2021) about classifying covariates or Reinhart et al. (2022), who explored students reasoning about correlation and causation.

METHODS

Two instructors conducted the study in seven different statistics-related courses, including introduction to quantitative research methods (in both bachelor's and master's programs). The students are majoring in business-related subjects. The voluntary, anonymous, web-based survey took place during the second lecture of the course using the same classroom response system as regularly used in classes (https://tweedback.de). The first (previous) lecture mainly covers organizational and general science topics with no statistic-specific topics such as, e.g., sample and population taught. The pre-assessment takes place at the beginning of the lecture. We provided students with only the multiple-choice question displayed in Figure 1.

> On an internet platform, 10,000 people report a positive effect of a particular shampoo on gray hair (Study A). An experiment with 100 randomly selected people finds no positive effect of the shampoo (Study B). With the information given, the result of which study is more credible?
> A: The result of study A
> B: The result of study B
> C: Both studies are equally credible

Figure 1. Multiple-choice question given to students

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org ©2022 ISI/IASE

At this point in time, we did not show the results or the correct answer (B) to the students. During the lecture that followed, topics such as measurement, random sampling, and random assignment are introduced. The introduction of random sampling and random assignment was supported by presenting causal diagrams (see Figure 2), without formally discussing graph elements such as nodes and edges. The sampling example is embedded in a fictitious study where a teacher tries to analyze the learning time of her students using a voluntary survey, i.e., a convenience sample. Both learning time and participation in the survey may, for example, depend on conscientiousness. (See the causal diagram on the left in Figure 2.) With a random sample, participation no longer depends on conscientiousness. The arrow from there pointing into the sample is erased and replaced by the researcher's study design, as shown in the causal diagram on the right in Figure 2.

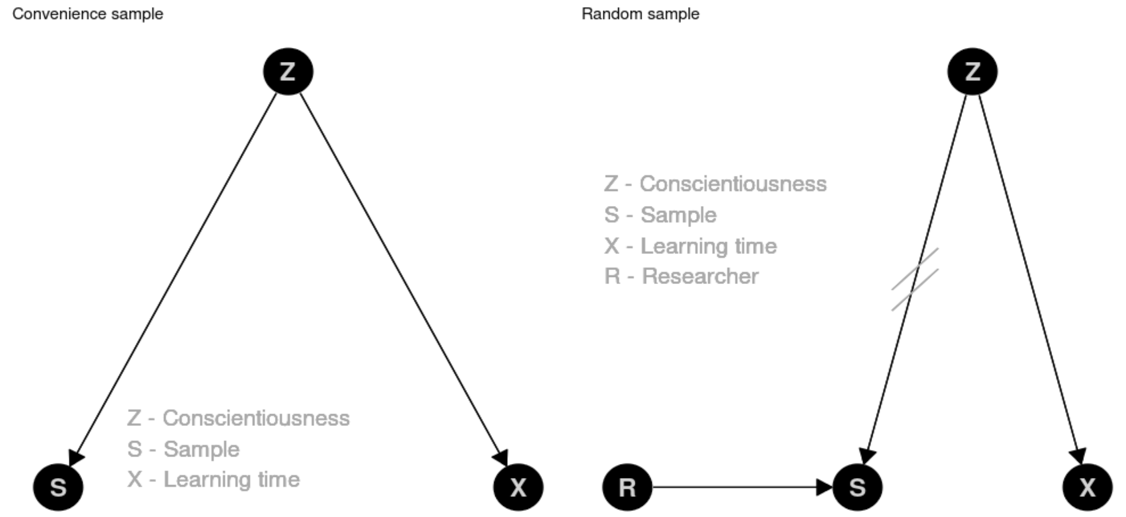

Figure 2. Causal diagram for an example on convenience sample vs. random sample

The example of random assignment within a randomized controlled trial is embedded in a fictitious study where the teacher tries to analyze the relationship between learning time and test score. Prior knowledge is one reasonable confounder here. (See the causal diagram on the left in Figure 3.) This confounder may even give rise to Simpson's paradox, i.e., observing a negative correlation between learning time and test score, whereas the true (direct) causal effect is positive. Again, randomness erases the arrow pointing into the treatment (see the causal diagram on the right in Figure 3).

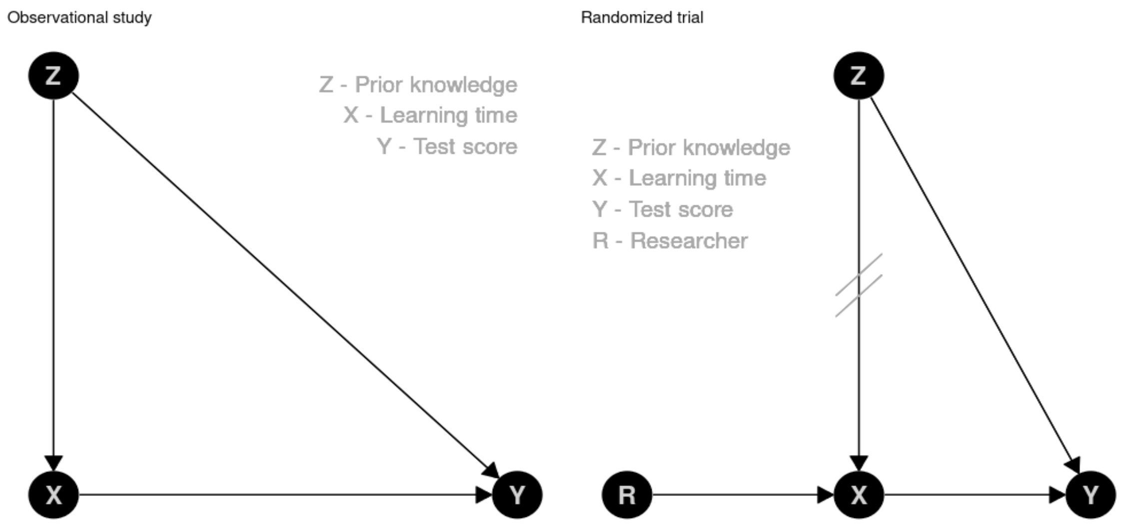

Figure 3. Causal diagram for an example on observational study vs. randomized trial

Both examples provide opportunities to discuss practical and ethical challenges in conducting these fictitious studies in reality. Students can be asked to consider the following questions. How can we achieve a random sample? Is it ethical to randomize learning time? What about non-compliance?

For post-assessment, the same question as at the beginning of the lecture is asked at the end of the lecture. Because one lecture is divided into two 90-minute segments, the post-assessment took place approximately three hours after the pre-assessment. To lower the barrier for participation, there was no attempt to link the results of the pre- and post-assessments.

We also included a short evaluation within the classroom response system to investigate students' perceptions about causal diagrams. The students were finally asked to rate their agreement on a 5-point Likert scale to the following statement: The diagrams (graphs) to describe the data generating process are helpful to understand concepts of data collection (randomized sampling and allocation).

RESULTS

Due to the COVID-19 pandemic, we conducted the survey within synchronous online lectures in the fall of 2021. Data and R code for the analysis are available from https://github.com/luebby/ICOTS-2022. A total of $n_{pre} = 282$ students took part in the pre-assessment, and $n_{post} = 280$ responded to the post-assessment. For the short evaluation question, $n_{eval} = 230$ students provided opinions about the helpfulness of the graphs. Due to the study design, it is possible that some students answered the post-assessment without answering the pre-assessment or the pre-assessment without the post-assessment, for example. Approximately $^2/_3$ of the students attending the lectures answered the survey.

In the pre-assessment only $p_{pre} = 49.6\%$ of the students provided the correct answer (B). In the post-assessment, this proportion increased to $p_{post} = 62.5\%$. The increase of 12.85 percentage points (Cohen's $h = 0.26$; Cohen, 1988) with a p-value of 0.0012 is statistically discernible (Witmer, 2019). To calculate the p-value, we permuted pre- and post-responses within each course; the one-sided p-value for $\pi_{post} - \pi_{pre} \leq 0$ is calculated based on 10,000 permutations.

We should note that these results are based on an observational study within the classes. We did not apply a randomized controlled trial nor any difference-in-difference methods, so we cannot identify a causal effect of including the diagrams versus not including the diagrams.

Figure 4 shows the result of students' evaluating the helpfulness of the causal diagrams. More than 71% of the students stated that they fully or strongly agreed that the causal diagrams were helpful to understand concepts of data collection; 10% of the students disagreed.

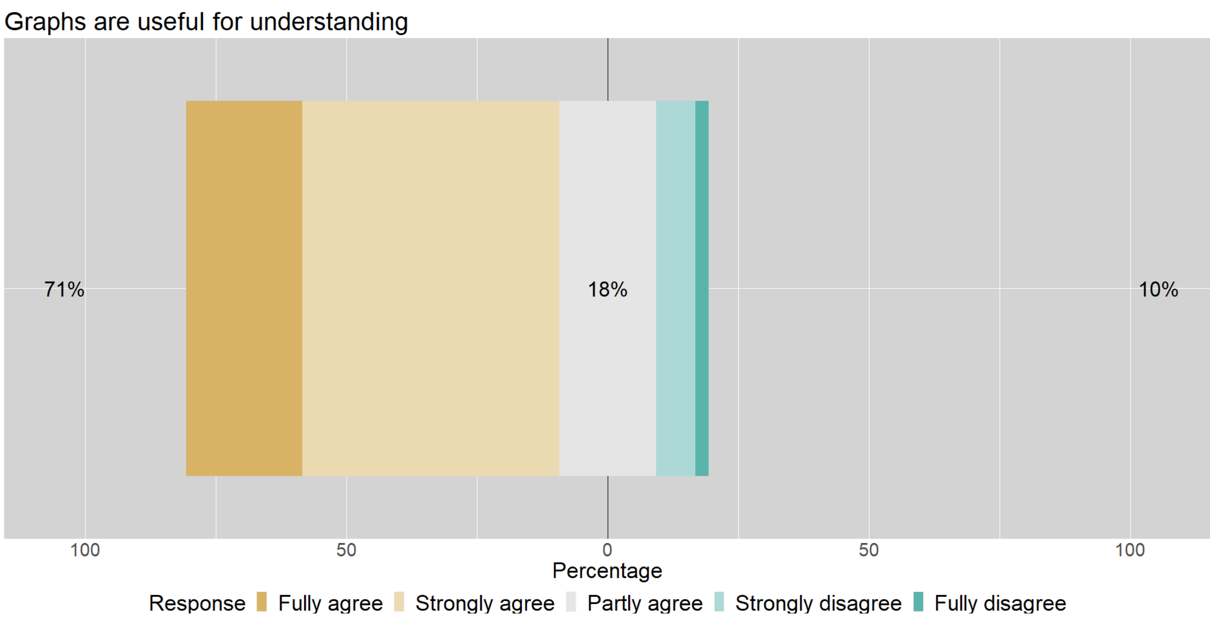

Figure 4. Evaluation of helpfulness

Our experiences as teachers in these classes are also positive. The causal diagrams of these elementary, fictitious examples provide an excellent opportunity to discuss the crucial topics of confounding and bias. In addition, we experienced lively debates about the merits of randomness in data generation and why this may be hard or even impossible to achieve.

CONCLUSION

In a world full of big data and many studies published based on the analysis of such data, we, as statistics and data science educators, face the challenge of how we can help our students to draw correct conclusions. The latter is essential even on a purely descriptive level. Causal diagrams may help prepare students not to mess with data and make trustworthy conclusions based on data. We should sensitize them as early as possible that "compensating for quality with quantity is a doomed game" (Meng, 2018). Causal diagrams may provide an easy-to-grasp language to discuss the assumptions about the data-generating process. With these diagrams, instructors can formally assess confounding and bias and illustrate the important benefits of random sampling and/or random allocation. In more advanced courses, one can discuss methods of perhaps recovering from bias and confounding in observational data.

The current study has many limitations which should be considered. For example, pre- and post-assessment took place in a single lecture, with only one question. No qualitative data was collected. Also, it is not possible to identify and estimate the causal effect of the intervention by including the diagrams. More and better designed studies are needed to investigate the effect of this teaching approach as well as student understanding and learning. For example, an open question such as “How did the causal diagram help you to come to the answer you chose?” could be added. So, we need more research on how statistics education can provide students with a conceptual framework to scrutinize the data generation process in a data-centric world.

REFERENCES

Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, *1*(1). https://doi.org/10.5070/T511000028

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.). Lawrence Erlbaum Associates.

Cummiskey, K., Adams, B., Pleuss, J., Turner, D., Clark, N. & Watts, K. (2020). Causal inference in introductory statistics courses. *Journal of Statistics Education*, *28*(1), 2–8. https://doi.org/10.1080/10691898.2020.1713936

Ellison, G. T. (2021). Might temporal logic improve the specification of directed acyclic graphs (DAGs)? *Journal of Statistics and Data Science Education*, *29*(2), 202–213. https://doi.org/10.1080/26939169.2021.1936311

Greenland, S. (2022). The causal foundations of applied probability and statistics. In H. Geffner, R. Dechter, & J. Y. Halpern (Eds.), *Probabilistic and causal inference: The works of Judea Pearl* (1st ed., pp. 605–624). Association for Computing Machinery. https://doi.org/10.1145/3501714.3501747

Hernán, M. A., Hsu J., & Healy, B. (2019). A second chance to get causal inference right: A classification of data science tasks. *CHANCE*, *32*(1), 42–49. https://doi.org/10.1080/09332480.2019.1579578

Kaplan, D. (2018). Teaching stats for data science. *The American Statistician*, *72*(1), 89–96. https://doi.org/10.1080/00031305.2017.1398107

Lübke, K., Gehrke, M., Horst, J. & Szepannek, G. (2020). Why we should teach causal inference: Examples in linear regression with simulated data. *Journal of Statistics Education*, *28*(2), 133–139. https://doi.org/10.1080/10691898.2020.1752859

Meng, X. L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, *12*(2), 685–726. https://doi.org/10.1214/18-AOAS1161SF

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wieczorek, J., Elliott, P., Burckhardt, P., & Nugent, R. (2022). Think-aloud interviews: A tool for exploring student statistical reasoning. *Journal of Statistics and Data Science Education*. Advance online publication. https://doi.org/10.1080/26939169.2022.2063209

Schield, M. (2018). Confounding and cornfield: Back to the future. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, Looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018) Kyoto, Japan*. ISI/IASE. https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_1C2.pdf?1531364185

Utts, J. (2021). Enhancing data science ethics through statistical education and practice. *International Statistical Review*, 89, 1–17. https://doi.org/10.1111/insr.12446

Witmer, J. (2019). Editorial. *Journal of Statistics Education, 27*(3), 136–137. https://doi.org/10.1080/10691898.2019.1702415