ISSUE 41 MARCH 2022

CAMBRIDGE
√Mathematics

ESPRESSO

RESEARCH, FILTERED BY CAMBRIDGE MATHEMATICS

TALKING POINT:

WHAT DOES RESEARCH SUGGEST ABOUT TEACHING STATISTICS USING RICH DATA SETS?

IN SUMMARY

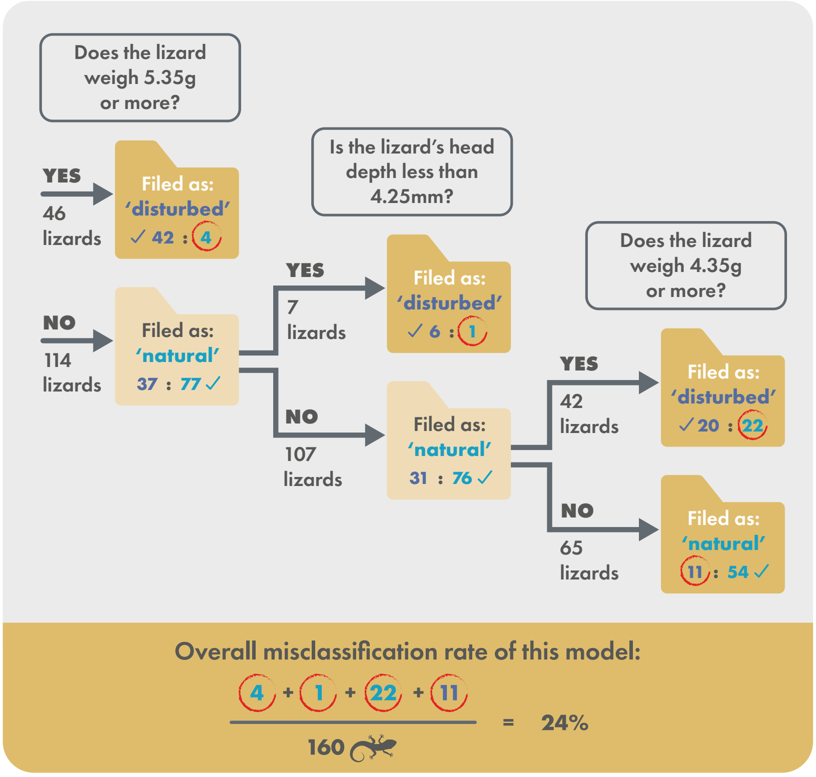
- Rich data sets – which may be messy, have many variables (multivariate), a complex structure, and a mixture of both qualitative and quantitative data – may support student engagement and create opportunities for cross-curricular work
- Approaches such as exploratory data analysis, informal inference, and data modelling are complementary to working with rich data sets
- Using rich data shifts the focus of statistics lessons away from procedural fluency towards interpretation and discussion and so may require different approaches to assessment
- When using multivariate data, it is important to use contexts with which students are familiar, access to carefully designed software tools, and well-structured activities
- When engaging with rich data a problem-solving cycle approach is recommended
- Rich, multivariate data sets should contain a range of different data types to allow the posing of interesting and engaging questions and problems

An example of a rich data classification task

160 lizards were captured from two different habitats:

disturbed (developed by humans) and **natural (no human development)**.

Are there ways of classifying the lizards that might help us predict whether a randomly chosen lizard has come from a disturbed or natural habitat?



1

The use of rich data supports a greater emphasis on statistical thinking,[2] allowing greater focus on conceptual rather than procedural understanding and creating space for more student-led lessons in line with recommendations from research.[3] Data sets that can be considered "rich" may be messy, have many variables (multivariate), a complex structure, and a mixture of both qualitative and quantitative data. Rich data reflects more closely the data students encounter in everyday contexts and can support motivation and engagement.[4] Using rich data creates opportunities for cross-curricular activity, which in turn may require cross-curricular planning to fully exploit.[5] Considering different data and variable types and engaging in multivariate thinking (making inferences about data with more than two variables) are all important from the earliest stages of mathematics education onwards.[1]

IMPLICATIONS: Rich data may support engagement and motivation in statistics learning with a greater focus on conceptual understanding

Embedding the use of rich data creates opportunities for cross-curricular work

Consideration of different data and variable types is important from the earliest stages of education

2

Early work with rich data may be supported by informal approaches such as exploratory data analysis (see *Espresso 29*) and drawing several pictures of the same data.[6] Other approaches supported by rich data include informal inference (making conjectures from data and describing a level of confidence in them)[7] and data modelling (using data to design rules that can be used to make predictions or solve problems).[8] Tasks involving multivariate data may appear more demanding than traditional univariate or bivariate data-based tasks, but carefully structured task design can support students with a wide range of prior attainment in accessing teaching and learning activities.[5] Making sense of multivariate data does not require learning advanced techniques;[9] existing curriculum objectives can be met by, for example, calculating averages for sub-groups generated from a categorical variable. Using rich data allows greater focus on developing central statistical ideas, supports statistical reasoning, and promotes classroom discourse but these elements may not be well supported by forms of assessment that focus on procedural fluency.[10] Rich data enables students to engage with a broader range of statistical activities such as classification (see infographic), and foundational concepts in data science.[1]

IMPLICATIONS: Approaches such as exploratory data analysis, informal inference, and data modelling complement work with rich data sets

It is particularly important to structure tasks carefully when working with rich data

Using rich data shifts the focus of statistics lessons away from procedural fluency towards interpretation and discussion, which may require different approaches to assessment

3

The effective exploration of rich data sets is supported by a classroom environment that includes access to technology tools, and well-designed tasks or prompts.[11] Effective technology tools for data visualisation aim to make visualisation simple and unambiguous, allow the use of colour to highlight patterns, minimise jargon and provide useable options.[12] Some tools that meet these recommendations include TinkerPlots,[13] CODAP,[14] and iNZight.[2] Web-based resources such as ProCivicStat and Gapminder also curate rich data, tasks, and visualisation resources.[9] Classroom activities related to exploration of rich data may be enhanced by engaging with a statistical problem-solving cycle that encompasses posing questions, processing data, creating representations and drawing inferences or conclusions.[15]

IMPLICATIONS: Alongside multivariate data sets, access to technology tools and carefully structured activities are important to support student exploration of data

Software specifically designed to support exploration of rich data (such as TinkerPlots, CODAP, and iNZight) may help reduce barriers to engagement

When exploring rich data, a problem-solving cycle approach is recommended

4

The effective use of rich data sets in the classroom may be supported by the use of real-life contexts in which students already have some experience. This can create opportunities for students to ask questions where their expectations are either confirmed or confounded.[16] Data should be interesting, relevant, and potentially controversial in order to stimulate interest, with sufficient contextual information to allow students to act as researchers.[17] Rich data sets may contain a range of different data types including qualitative and quantitative, discrete and continuous, cardinal, ordinal, counts, and proportions. They may also contain more modern data types such as images, sounds, videos, or words.[1]

IMPLICATIONS: Rich, multivariate data sets should contain a range of different data types to allow the posing of interesting and engaging questions and problems

Data contexts with which students are familiar may support engagement and accessibility

"Collect enough data and it's hard to know what you'll uncover"

Fry, 2019 [18(p.28)]

"Somebody, somewhere, still had to decide which information to collect, but it's easy to gather that data just in case, and decide later whether it's useful"

Harkness, 2016 [19(p.18)]

Darren Macey & Lucy Rycroft-Smith, 2022

REFERENCES

1. Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education*. National Council of Teachers of Mathematics.
2. Forbes, S., Chapman, J., Harraway, J., Stirling, D., & Wild, C. (2014). Use of data visualisation in the teaching of statistics: A New Zealand perspective. *Statistics Education Research Journal, 13*(2), 187–201.
3. Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 37–70). Springer International Publishing.
4. Gould, R., Peng, R. D., Kreuter, F., Pruim, R., Witmer, J., & Cobb, G. W. (2018). Challenge to the established curriculum: A collection of reflections. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 457–472). Springer International Publishing.
5. Ridgeway, J., Nicholson, J., & McCusker, S. (2007). Reasoning with multivariate evidence. *International Electronic Journal of Maths Education, 2*(3), 245–269.
6. Gil, E., & Gibbs, A. L. (2016). Introducing secondary school students to big data and its social impact: A study within an innovative learning environment. In J. Engel (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education*. IASE.
7. Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.
8. Ärlebäck, J. B., Blomberg, P., & Nilsson, P. (2015). An instructional design perspective on data-modelling for learning statistics and modelling. In O. Helenius, A. Engström, T. Meaney, P. Nilsson, E. Norén, J. Sayers, & M. Österholm (Eds.), *Development of mathematics teaching: Design, scale, effects: Proceedings of MADIF 9* (pp. 37–46). Swedish Society for Research in Mathematics Education.
9. Engel, J. (2017). Learning from data about society: Perspectives and experiences from ProCivicStat. In R. Biehler, L. Budde, D. Frischmeier, B. Heinemann, S. Podworny, C. Schulte, & T. Wassong (Eds.), *Paderborn symposium on data science education at school level 2017: The collected extended abstracts* (pp. 59–66). Universitätsbibliothek Paderborn.
10. Garfield, J., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics, 31*(3), *72–77.*
11. Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473–502). Springer International Publishing.
12. Krumhansl, R., Peach, C., Foster, J., Busey, A., & Baker, I. (2012). *Visualizing oceans of data: Educational interface design*. Education Development Centre, Inc.
13. Frischemeier, D. (2017). Statistisch denken und forschen lernen mit der Software TinkerPlots. Springer Spektrum.
14. Haldar, L. C., Wong, N., Heller, J. I., & Konold, C. (2018). Students making sense of multi-level data. *Technology Innovations in Statistics Education, 11*(1).
15. Watson, J., Fitzallen, N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 105–138). Springer International Publishing.
16. Rubin, A. (2021). What to consider when we consider data. *Teaching Statistics, 43*(S1), S23–S33.
17. Willet, J. B., & Singer, J. D. (1992). Providing a statistical "model": Teaching applied statistic using real-world data. In F. Gordon & S. Gordon (Eds.), *Statistics for the twenty-first century* (MMA Notes no. 26, pp. 83–98). Mathematical Association of America.
18. Fry, H. (2019). *Hello world: How to be human in the age of the machine*. Black Swan.
19. Harkness, T. (2016). *Big data: Does size matter?* Bloomsbury Sigma.

CAMBRIDGE UNIVERSITY PRESS & ASSESSMENT

UNIVERSITY OF CAMBRIDGE Faculty of Mathematics

UNIVERSITY OF CAMBRIDGE Faculty of Education