STUDENTS' PREFERENCES WHEN CHOOSING DATA SETS WITH DIFFERENT CHARACTERISTICS

Johannes Schulz[1], Burkhard Priemer[1] and Amy Masnick[2]
[1]Humboldt-Universität zu Berlin, Germany
[2]Hofstra University, Hempstead, NY, U.S.A.
schulzj@physik.hu-berlin.de

We report a study on how 57 undergraduate students identify and evaluate characteristics of data sets like variance, sample size, number of decimal places of the values, and repeated values. The findings suggest that the participants who are not experts in physics or statistics are able to notice obvious physical differences in data sets and find criteria to judge these data sets. They also have an intuition for the (mathematical) variance of the numbers, whether through spread or through the repetition of individual values. This means that students hold correct prior beliefs on what is important to look at when confronted with two different data sets which is a promising condition for developing future instructional curricula and exercises to improve understanding of data.

OBJECTIVES OF THE STUDY

Scientific literacy includes students' abilities in judging the quality of data (Chinn & Malhotra, 2002). To judge the quality, they must be able to identify and evaluate characteristics of data sets like number of decimal places of the values, variance, sample size, and repeated values. In this exploratory study, we wanted to examine the judgements about these characteristics from non-expert undergraduate students. The key research question is: How do students use the four characteristics of data stated above in determining which type of data they would use in a hypothetical investigation? In particular, the participants compared data sets with different numbers of decimal points, with different levels of variance, with different sample sizes, and with one or more repeated numbers, to assess which were most informative in determining which of two measurement methods would be better. Learning more about students' reasoning about these topics can aid in developing methods to teach students about measurement uncertainty and variance, a fundamental component of physics instruction.

THEORETICAL FRAMEWORK

Scientists and statisticians strive to have reliable data for precise and accurate measurement. Data that are more reliable tend to be more precise and have less variability. In addition, larger sample sizes often increase confidence that the sample represents the population well. Experimental data often display certain characteristics dependent on the method used to obtain the data. For example, some experiments generate data with precise values (e.g. with a couple of decimal places) while others do not. Sometimes the data sets have a high variance (e.g. due to the variability of what's being measured, or statistical randomness) and other times the contexts produce less variance. In some settings, only small sample sizes can be gathered (e.g., of rare phenomena), whereas in other cases there is more data available. Finally, some sets of data have a certain number of repeated values, where multiple data points are identical (e.g. when the measurements have low statistical variation), while others have fewer or no repetitions.

It is a key competence of quantitative work to evaluate data sets (see for example the different models of experimental work: Klahr & Dunbar, 1988; Millar, Le Maréchal & Tiberghien, 1999: Osborne, 2014). However, students' competences in statistical methods are often very weak. This difficulty has been demonstrated for example in the field of evaluating measurement uncertainties (Buffler, Allie, Lubben & Campbell, 2001; Lubben & Millar, 1996; Priemer & Hellwig, 2016). Thus, it is an open question how students judge the quality of data.

Due to a lack of deeper background knowledge in statistics, students may rely on their limited conceptions (e.g., knowing how to calculate a mean), use their intuition or heuristics they've heard of, or apply simplifications of common known phenomena (e.g., more measurements are better; similar to p-prims suggested by di Sessa, 1983). There is little research on what data characteristics students prefer when judging data sets, and knowledge of such beliefs can help in understanding the misconceptions students hold, as well as in how to address them.

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.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Given the strong evidence that context matters in reasoning (e.g., Kanari & Millar, 2004), here we chose to present datasets with minimal context, to focus attention on the statistical features without interference, as a first exploration of the topic.

METHOD

In order to address the research questions, we asked undergraduate students to look at different pairs of two data sets 37 times, which were framed as tests of two potential methods for measuring the density of an object. We asked students to judge which method seemed more reliable, given the data, by asking them which method they would choose to measure the density of a new object, or whether they were equally likely to choose either one.

Before the questions, the participants read: "There are a lot of methods to measure the density of an object. In the following example we want to determine which measurement method is the best to determine the density of the object just due to the measured data.

Therefore, in each question we will present you the results of two of the different measurement methods – without telling you anything about other characteristics of the different methods – and ask you which method you would prefer due to the data." Then they were given 37 items in total in the form as shown in Fig. 1.

We measured the density of an object with two different methods. The raw-scores are shown below. The methods A and B are not related to any methods A and B in any of the previous questions.

Data from method A Density in kg/m³	Data from method B Density in kg/m³
343	375
339	319
365	293
346	350
357	398
350	365

(kg: kilogram, m: meter)

If you have to measure a new object´s density, which method would you choose based on data given here?

Figure 1. Example of an item (low variance vs. high variance)

We prepared 74 pairs of data sets in total with different numbers of decimal points, with different levels of variance, with different sample sizes, and with one or more repeated numbers. In each of the 74 trials, participants saw two columns of data. The data varied along one of the dimensions noted above. All of the pairs of data sets had the same pairwise mean and were fixed on each of the four parameters above except one.

In examples where we varied the variance, we created 24 pairs of data sets (all with a sample size of six). In 12 of them the variance relative to the mean (standard deviation/mean) in one set was double that of the other (with examples from 2.5 vs. 5% of the mean, 5% vs. 10% of the mean, and 10% vs. 20% of the mean). In 8 of the data sets the variance of one data set was four times higher than in the other data set (with examples from 2.5% vs. 10% and 5% vs. 20%) and in 4 of them the variance was eight times higher (2.5% vs. 20%).

When varying the sample size, we used 18 pairs of data sets. We used the sample sizes of one, three, six, and ten values in the data sets and combined them pairwise in 3 pairs of data sets for each case (e.g., we had 3 pairs of data sets with one sample of size 1 – a single data point – and one sample of size 3; 3 pairs with a set of 3 compared to a set of 6; etc.). All possible combinations of the four levels were used.

For the number of decimal places, we used 12 pairs of data sets. We took 4 pairs where one data set had no decimal places and the other had one, 4 pairs where one had one decimal place and the other had two, and 4 pairs where one had no decimals places and the other two.

For the repeated numbers we created 20 pairs of data sets. If a column of six numbers had two identical numbers, that was considered two repeated numbers. There were four pairs for each case with no repeated number in one data set and two, three, four repeated numbers in the other data set. And 4 pairs for each case with two repeated numbers in one and three, four repeated numbers in the other data set.

The pairs of data sets were presented digitally on laptops to 57 undergraduate students from an urban university in Germany. No more information about the context was included in the task for the students except that the data sets were derived from different physics experiments. For each pair of these data sets the students had to decide which experimental setting they would prefer due to the data. They also had to state how sure they are in their decision on a Likert scale from 1 to 5. The pairs of data sets were presented in a random order for each student. The order of the data sets in each pair were chosen at random. The students had no time limit for the test, and no student took longer than 40 minutes.

Finally, after answering questions about the pairs of data sets we asked in an open question the criteria the students used to decide between the data sets. Afterwards we presented the criteria "variance in the data", "sample size", "repeated numbers", "decimals places" and "difference between the highest and lowest value" and asked students to rate the importance of these characteristics on a Likert Scale.

RESULTS

Participants responded to each of the data characteristics. When considering differing levels of variance, they more often said that the data with smaller variance was more informative (28-60%), with a higher proportion doing so when the ratio of variances was greater (Table 1). They very rarely chose the data column with more variance as the one they were most confident in (9-13%), though in many cases they more often chose the option saying both data sets were equally useful than saying they preferred the one with less variance (31-59%).

Table 1. Students' choice of the data set with the smaller variance depended on the difference of the variance in the data sets given

Variance of the two given data sets in % from the mean value		Percentage of students choosing the data set with the smaller variance
Data set 1	Data set 2	
2.5	5	28
2.5	10	41
2.5	20	60
5	10	39
5	20	57
10	20	41

On average students were more likely to say that the data set with more decimal places was more informative than the data set with fewer decimal places, particularly when comparing no decimal places to some decimal places (55%-64% of the time). When comparing 1 to 2 decimal places, they only preferred two decimal places 37% of the time but chose both 42% of the time.

When looking at data with different sample sizes, they were more likely to say they preferred more data to less (52%-64%), they preferred one data point to several (regardless of how many additional points there were). The smaller sample size was preferred by 23%-27% of the participants. With the different set sizes, participants rarely chose the option of both columns being equally informative (only 6-25% of the time). It is interesting to note here that for the different sample sizes presented the students' preferences were quite stable.

Repeated numbers led to the least clear-cut responses, with the majority of participants choosing both as their answer, except in the case of the most dramatic contrast, with one column of all unique numbers and another column with four numbers the same. Even there, both was the most common answer (45%), as compared to the column with four repeated numbers (36%) and the column with no repeated numbers (18%).

In the open question about the criteria the students used to decide, several students gave multiple answers. The most popular criteria are shown in Table 2. Ten students mentioned "gut feelings" or "guessing" as their criteria. Additionally, there were criteria which were mentioned just a few times, like the existence of outliers (3), and the pairwise comparison of the values (3). Finally, the criteria uneven numbers and smaller numbers were only mentioned once each, together with the criterion that the data sets have always similar qualities if they have the same sample size.

Students' answers to the closed question indicate that when criteria were given, the students rate them more or less equally important (Table 2). This is especially important for the repeated values since this criterion was rarely mentioned when students were asked in the open question. The considerably high standard deviations show that there is a notably range in judging the importance of the criteria.

Table 2. Students' answers to the open and closed question (scale from 1 to 5) concerning the importance of the different data set characteristics

Data set characteristic (criterion)	No. of times mentioned by students in the open question	Average rating of the students in the closed question	Standard deviation
Sample size	21	3.8	1.41
Number of decimal places	20	3.2	1.63
Variance (deviation)	17	3.5	1.62
Repeated values	7	3.4	1.63
Difference between highest and lowest value	4	3.3	1.64

DISCUSSION

These findings suggest that undergraduate students who are not experts in physics or statistics are able to notice obvious differences in data sets (such as sample size and different numbers of decimal points/precision) and have a (slight) intuition for the (mathematical) variance of the numbers, whether through spread or through the repetition of individual values. Thus, the answer to our research question is that students hold mostly correct prior beliefs about selected statistical concepts (the characteristics of the data sets), like for example, they prefer a smaller variance over a larger variance.

However, it is an open question how students made their decisions and the reasoning behind their choices. Are these decisions made by intuition or by applying known statistical concepts? Further, we do not know if they are able to quantify the difference between the two data sets (e.g., to judge if the difference really matters in a statistical sense). Both questions will be answered by additional studies using the tasks of this study and interviews.

Currently, we are trying to identify several sub-groups of the participants who consistently responded to one of the given criteria. That means that their preference between two data sets regarding one criterion is stable and might be based on a conceptual understanding (regardless of whether it is correct or not). The results will show (1) which criteria are applied consistently by which number of students, (2) how the different criteria are related to each other with respect to this consistency, and (3) how the consistency of the students answers in one criterion affects another.

Our findings indicate that students hold correct (though sometimes naïve) prior beliefs on what is important to look at when confronted with two different data sets. This is a promising condition for developing future instructional curricula and exercises to improve understanding of data.

ACKNOWLEDGEMENTS

We would like to thank Kristoph Schumann for his help with the study.

REFERENCES

Buffler, A., Allie, S., Lubben, F. & Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education, 23*(11), 1137-1156.

Chinn, C. A. & Malhotra, B. A. (2002), Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175–218. doi:10.1002/sce.10001

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner and A. Stevens (Eds.), *Mental models* (pp. 15–33). Mahwah, NJ: Lawrence Erlbaum.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748–769. doi:10.1002/tea.20020

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1–48.

Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education, 18*(8), 955-968.

Millar, R., Le Maréchal, J., & Tiberghien, A. (1999). 'Mapping' the domain: Varieties of practical work. In J. Leach & A. Paulsen (Hrsg.), *Practical work in science education. Recent research studies* (S. 33–59). Frederiksberg: Roskilde University Press; Kluwer Academic Publishers.

Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education*. Routledge.

Priemer, B. & Hellwig, J. (2016). Learning about measurement uncertainties in secondary education: A model of the subject matter. *International Journal of Science and Mathematics Education*. doi:10.1007/s10763-016-9768-0.