



HAL
open science

Analysing teachers' knowledge about sampling using TinkerPlots 2.0

Maria Niedja Martins, Carlos Monteiro, Carolina Carvalho

► **To cite this version:**

Maria Niedja Martins, Carlos Monteiro, Carolina Carvalho. Analysing teachers' knowledge about sampling using TinkerPlots 2.0. CERME 9 - Ninth Congress of the European Society for Research in Mathematics Education, Charles University in Prague, Faculty of Education; ERME, Feb 2015, Prague, Czech Republic. pp.715-721. hal-01287103

HAL Id: hal-01287103

<https://hal.science/hal-01287103>

Submitted on 11 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysing teachers' knowledge about sampling using TinkerPlots 2.0

Maria Niedja Martins¹, Carlos Monteiro² and Carolina Carvalho¹

1 University of Lisbon, Lisboa, Portugal, marianiedjamartins@campus.ul.pt, cfcarvalho@ie.ul.pt

2 Universidade Federal de Pernambuco, Recife, Brazil, cefmonteiro@gmail.com

The conceptualization of sampling is crucial to understand statistical data. However, the teaching about sampling is not generally emphasised in school curriculum. This study investigated how teachers understand about size and representativeness of samples using TinkerPlots 2.0 software. The study was comprised of two sessions. A semi-structured interview and a familiarization about the basic use of TinkerPlots were developed during the first session, and participants engaged in three tasks on sampling using TinkerPlots during the second session. As a result, the teachers began to consider aspects of the variation of data to determine when representative samples were involved in TinkerPlots. The ability to select samples and analyse them seemed to contribute to improve their understanding about sample size and representativeness.

Keywords: Statistics education, primary school teachers, TinkerPlots.

INTRODUCTION

The recognition of the influence of statistical data in the current society demanded the inclusion of this topic in national curriculum of many countries (Monteiro & Ainley, 2004). Several studies suggested that statistics education can provide bases to students develop abilities to argue and counter-argue information, understand the generation of statistical data and make informed decisions based on their analysis (Gal, 2002). Therefore, statistical knowledge is essential for critical reflective and participatory citizenship (Carvalho & Solomon, 2012).

An important knowledge that enables citizens to understand critically statistical data is related to conceptualization of sample and sampling. Bolfarine and Bussab (2005) conceptualize sample as any subset of

a given population, and sampling as a technique of selection of such subsets. Innabi (2006) argues that in order to analyse the representativeness of a sample is necessary to know whether the sample is large enough and has the variety present in the population. It is recommended to increase the sample to ensure the variety of the population can be better visualized. However, sample sizes from a homogeneous population tend to be smaller, because such samples will have less variability.

To understand the conceptualization of sampling is crucial consider how the data were chosen, what methods are employed for the selection of these cases, what features and prioritized variables, so we can understand other contexts in which the information can be applied (Saldanha & Thompson, 2002; 2007). Therefore, the understanding about sampling seems to be essential as curriculum school content, therefore it is very relevant for teachers who teach statistics.

Although, the teaching about samples and sampling is fundamental to base the practices of statistics, it needs to be more emphasised in school curriculum (Watson, 2004). Recently, several studies investigated the conceptualization of sample and sampling among students from different levels. However, it is also important to investigate such situations among teachers who are going to approach such curriculum content (Martins, Monteiro, & Queiroz, 2013).

Several studies investigated the developing of understanding about sample and sampling using computer based tasks. For example, Manor, Ben-Zvi, & Aridor (2013) conducted a study that engaged students in designed instructional activities using computer modelling and simulations of drawing many samples. According to those authors, the research tasks ena-

bled the students to think about sampling as a process when analyses are associated with samples.

Baker, Derry and Konold (2006) involved young students in two experiments about center and variation. In one of situations they used *TinkerPlots* (Konold & Miller, 2011) to develop a task in which students can get engaged in an inferential game. According to these authors "the inferential approach acknowledges that students with their teachers have to take part in the social practice of reasoning (p. 2)". When the students were comparing two distributions, they should realize that they needed of certain concepts to reach a conclusion on the distributions were different or not. The students could come to the conclusion that the concept of average was important to identify these differences. Therefore, the uses of certain concepts involved in a game of give and ask for explanations.

These studies suggest that it seems to be important the selection of several samples. The *TinkerPlots* offers the possibility to explore the relationships between data and chance (Konold & Kazak, 2008), since it is possible to perform simulations of samples and populations.

Delmas and colleagues (1999) used a computer environment in which the students could simulate several samples of different sizes and visualize the distribution of the values of a statistic. The results suggested that students indicated that larger samples should produce a statistical distribution similar to their population of origin. According to those authors it is possible that the students had the intuition that the average is a point within the population, and that gather more averages it will have a distribution very similar to the population.

In this paper we discuss some aspects of a study that investigated teachers' knowledge about sample size and representativeness. The aim of this study was to explore possible computer based tasks which can help teachers to understand those important aspects about the sample and sampling.

METHODOLOGY

This was a qualitative exploratory study that followed an interpretative approach. The research was conducted in a rural public school located of a municipality of Metropolitan Region of Recife (RMR), Brazil.

The choice for this school was based on a survey conducted by GPEME - Research Group on Mathematics and Statistics Education (Carvalho & Monteiro, 2012), which identified 85 public schools in the RMR which had computer labs, and investigated how those labs were used.

There were two research sessions to collect the data. The first session was comprised of an individual semi-structured interview in order to have information about teaching experiences, as well as to identify their levels of understanding about the concept of sampling. The interview questions were based on a sample questionnaire used in the studies of Watson, Collis and Moritz (1995), Watson and Moritz (2000) and Watson (2004). These studies developed tasks associated with questions about sample, representativeness of small and large samples, sampling, and media news about sample surveys with inadequate statistical basis.

At the first research session, we also develop a familiarization with the *TinkerPlots2.0*. The researcher presented different functions of *TinkerPlots* to the teachers, including those to handle the database and produce graphs. This familiarization was carried out because the participants did not know about the software, and it was expected that they had certain autonomy to use the *TinkerPlots* during other research sessions.

The second session was comprised of three tasks using *TinkerPlots*. These tasks were about representativeness, size and type of sample. Therefore, in this paper due lack of space, we report examples taken only from analysis of task 1 and 2.

The study was developed with four female teachers. Due to lack of space in this paper, we report aspects of research data from one participant. For this report, her name was changed. Suzy was 30 years old, and she had 5 years of experience as a teacher. She uses the computer every day to search contents related to her teaching activities and to access emails. Suzy has university degree in Education. However, she said that never had any specific learning on sampling, and she did not know about *TinkerPlots* or other educational software for teaching Statistics. In this paper we do not discuss the data collected from the semi-structured interview.

Task 1

The aim of this task was to know if the participants understood that increasing a sample, could have better accuracy of inferences about the population, since the variability of the population would be better visualized.

Figure 1 shows a copy of screen with 625 cases (fish) of a *TinkerPlots* database called Fish Population. Each case had a numerical code, and information about the type and size of fish.

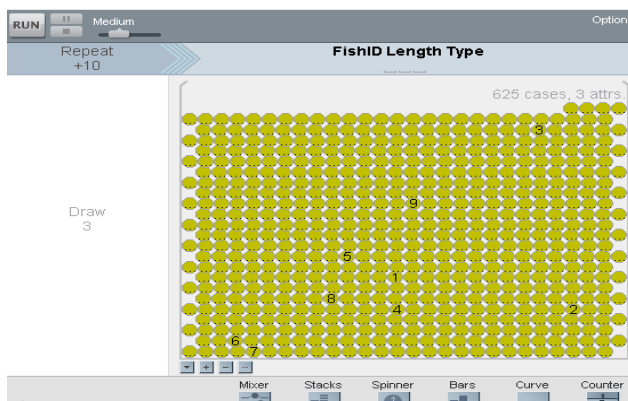


Figure 1: TinkerPlots Screen with 625 cases of Fish Population database

We asked the participant to read the following situation, which was based on the *TinkerPlots* resources:

A certain fish farmer bought some genetically modified fish of a company with the promise that they would grow more than the non GM fish. In order to check whether GM fish grow more, the fish farmer joined GM fish with other fish that he used to have in a tank in which totalized 625 fish. After the total growth time of fish, the fish farmer gradually withdrew each fish from tank, and measured each one. From the data analysis in

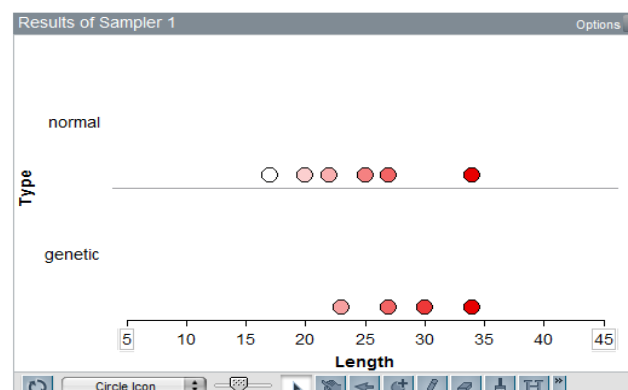


Figure 2: TinkerPlots Screen with fish population database, n = 10

TinkerPlots, indicate which type of fish had greater length. Did the fish farmer make good deal?

During participants' analyses of task 1, we took initially samples from 10 cases, and then it was increased based on their indications. The participants should infer which population had bigger fish interpreting a graph similar to the Figure 2.

For each new inclusion of cases in the sample, we asked the participant to informally rate her confidence level in a scale from 0 to 10. Therefore, rate 10 should be teacher's maximum confidence. This procedure aimed to make explicit their understanding about changes on their own analyses (Prodromou, 2011).

Task 2

The second task was based two TinkerPlots databases: *MysteryMixer1* and *MyysteryMixer2*, which were comprised of only one variable that is *number*. The Figure 3 presents a database used in this task.

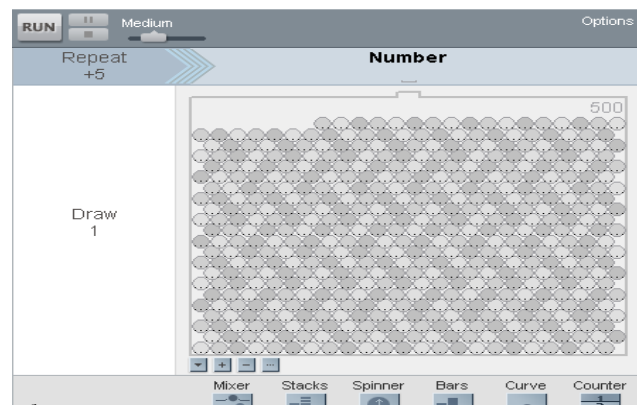


Figure 3: TinkerPlots Screen with *MysteryMixer1* database N = 500

The database had 500 cases disposed in the simulator ranging from 0 to 100. This second task aimed to identify whether the teachers could reach a conclusion on a small sample. Therefore, the participants should identify clusters of samples and infer them to the population, using the smallest possible sample. To ensure this, we engaged the participants in a fictional situation about costs of sample survey:

You have a limited amount of money to conduct survey on numbers. Each selection of five cases of this survey you should pay R\$1,00. Your task is to identify a range in which all numerical values are repeated. You need to spend the least amount

of money possible, but you need to be quite sure about your answer.

In each task, the first author, acted as researcher asking questions to make more explicit the teachers' considerations about the data, and assisting them in the selection and manipulation of *TinkerPlots* tools.

The *Camtasia Studio 7.1* software was used to record on video the participants' speeches, their gestures and manipulations developed in the computer screen while solved Tasks 1 and 2. The transcriptions of audio records generated protocols which were base to the data analysis.

RESULTS

The analyses of participants' response suggested aspects of their understandings about the relationship between size and representativeness of samples.

Task 1

During the development of task 1, when we increased sample size the participants informally rated their level of confidence about their inferences. Table 1 shows the Suzy's rates during this task.

Suzy gave a low confidence rate in their conclusions about small samples at the beginning of this task. The following extracts exemplify their arguments, when interpreting a graph similar to Figure 4.

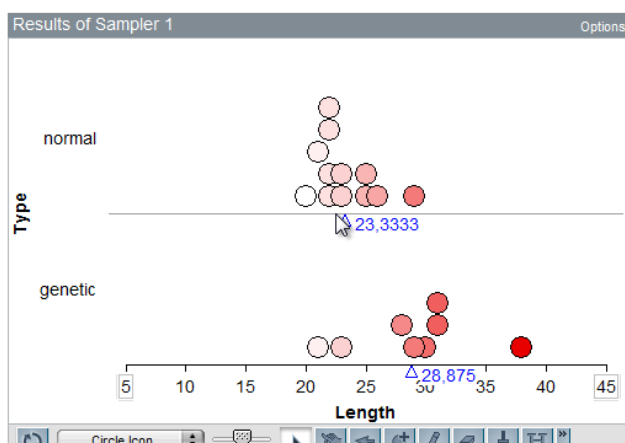


Figure 4: TinkerPlots Screen with fish population database, n = 20

Suzy: For me, this amount is not significant. It's because... well, the first time we had (a sample with) just over 2% (of the population) and now we have just over 3% (of the population). I think 3% is not significant value to buy something to put in a bowl and make a test. If so .. I would find significant 6% ... 10% ... just great! But to do a test ... to say ... (3%) I think very little.

Researcher: Right. But then, you see ... looking over here, can you observe who is showing a larger size?

Suzy: The GM.

Researcher: But, you are saying that perhaps this may not be significant for the rest?

Suzy: Exactly!

This fragment of Suzy's speech suggests that the teacher relates sample size to population size, and expressing that she considered the sample too small to make an inference. Suzy was unsure to make an inference, although that she identified a trend of genetically modified fish were larger.

Another extract from Suzy's protocol indicates that she was analysing the sample, and questioning the data variation in the samples, because she did not know the exact amount of fish for each population, since task 1 does not give this information.

Suzy: Does it [TinkerPlots] say the amount that it puts [in the sample]?

Researcher: No. It does not say the amount of one and another... whether it has more GM fish or normal ones. But, do you are absolutely sure that these here [GM] will continue growing?!

Suzy: It's because, look... 12 and 8 [amount of fish for each type]. We do not know the amount per type of fish that he put here [in the population].

Researcher: What does that mean?

Suzy: That these results may change here because these data can be very different from there [population].

Teacher	S1	Confidence	S2	Confidence	S3	Confidence	S4	Confidence
Suzy	10	0	20	0	100	8	150	10

Table 1: Informal levels of confidence about the increasing of sample sizes

Suzy seemed to be concerned about possible errors due to small samples. We can infer that she was concerned about the variation, because only taking cases at random from the population would not ensure that the values of sample were identical to those of the population of origin.

In addition, Suzy developed the strategy of seeking patterns by analyzing the distribution of data and trends in the samples. This was reflected in her assigned increasing levels of confidence to her inferences, since Suzy could confirm in each sample a tendency on genetically modified fish to be larger. The need to use the concept of average also emerged, according to view in Figure 5. Suzy's response seemed to be influenced by the average value in the different samples:

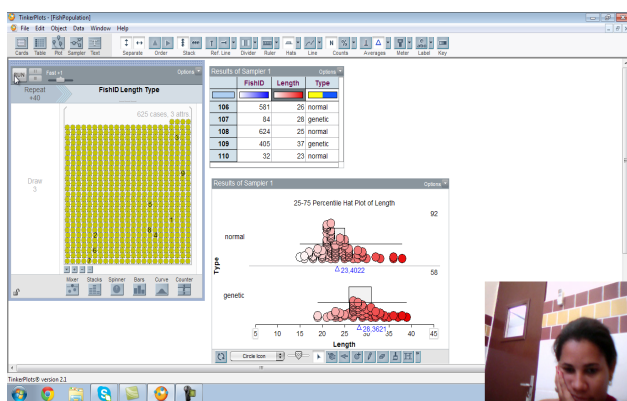


Figure 5: TinkerPlots Screen with fish population database, n = 150

Researcher: The average showed something to you?

Suzy: It shows that I'm correct! Because like this, it did not have changes. If the average had very fluctuated, there would be worrying. However, it remained constant to the extent that we have been getting more information.

Researcher: So it helped?

Suzy: It is. Now I have a 100% certainty.

Task 2

Suzy reached her answer with a smaller sample, and justify the response based on the idea of homogeneity of sample.

Researcher: Why did you find easy to say a response with lower number of cases?

Suzy: I think was because that issue of the group that I told you... because it was

concentrated in the group ... and ... don't know more.

Researcher: Concentrated in the group? What do you mean?

Suzy: So... lets I say... don't know. I thought so... to the extent that we were taking... I thought, should not have 50%, then I was dropping, 25%. And with this there, I did far less than the percentage that I thought at first. And, to the extent that I was taking (cases from the simulator), and that I was doing, the concentration kept constant. Then, I did not need to take all this data, I had focused on to analyze a bigger percentage... I believe this happens because the information is contained like that, in that group. It is not one thing mixed. I think that's it. I just do not know to explain, but I understand.

Suzy's strategy to be able to generalize the results of the sample to the population focused on the analysis of the trend of data in successive samples. She quickly realized that the curve where was concentrated most of the data remained constant even when the sample grew and relied on it to provide a final inference.

Another strategy that also seemed important for Suzy to choose a representative sample was associated with hypothetical costs of sample. The following extract from Suzy's interview exemplifies how this aspect was relevant to her analyses.

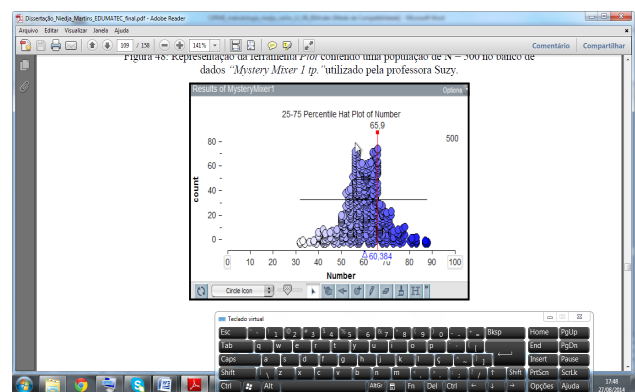


Figure 6: Graph produced by Suzy using MysteryMixer1 database

Researcher: [after show the graph with all cases according to Figure 6] Was close to what you said?

Suzy: I said 50 and 63. It was close!

Researcher: Were you pleased with your survey?
 With the amount you spent to give this approximate answer?

Suzy: Bargain, right?! [Laughing] So... Do you know one thing that I am worry about? In taking the part [sample], the bigger amount possible. And then, with the result, I can see that perhaps with smaller percentage, I could already tell you the answer. I could have been spending less because the sample was already confirming where it was [the cluster].

Suzy made inferences from 15 cases sample when she was interpreting the MysteryMixer2 database. The analyses of interview protocols suggested that this reduction was due to the fact that the teachers identify that the data of the samples were homogeneous.

CONCLUSIONS

The results of this study suggested that the teachers presented different ideas about sample when they analysed heterogeneous samples. This result corroborates the idea that the sampling involves different statistical concepts and ideas, and that inconsistencies of these notions can influence how a person perceives the representative samples.

In task 1, the analysis of variance of cases of homogeneous samples and the hypothetical cost to the sampling seemed to be the main influences on determining the appropriate sample size to make a final inference. One explanation for this result is the possibility that the teachers had to see the increasing of the samples and to compare the trends showed in *TinkerPlots* representations.

Therefore, the situations in which teachers can compare distributions may be potentially important to understand the tasks with the sampling; as seen in the study (Ben-Zvi et al., 2011) who found that the use of increasing samples can easily identify and recognize patterns representative established through comparison.

From the results of this study, further research is necessary to explore the autonomy of teachers to use software like *TinkerPlots* in order to build understandings of statistical concepts and also because teacher education in statistics software seems to have a gearing

effect on eventual student learning of statistical ideas (Pratt, Davies, & Connor, 2011). In addition, it is crucial to investigate how this knowledge constructed from their interaction with software can motivate reflective situations to explore new ways to teach statistics.

ACKNOWLEDGEMENT

This research was supported by CAPES Foundation, Ministry of Education of Brazil.

REFERENCES

- Bakker, A., Derry, J., & Konold, C. (2006). Using technology to support diagrammatic reasoning about center and variation. In A. Rossman & B. Chance (Eds.), *International Conference on Teaching of Statistics, Vol. 7* (pp. 1–6). Voorburg: International Statistical Institute.
- Ben-Zvi, D., Makar, K., Bakker, A., & Aridor, K. (2011). Children's emergent inferential reasoning about samples in an inquiry-based environment. In M. Pytlak, T., Rowland, & E. Swoboda (Eds.), *Congress of the European Society for Research in Mathematics Education, Vol. 7* (pp. 745–754). Rzeszów, Poland: University of Rzeszów.
- Carvalho, C., & Solomon, Y. (2012). Supporting statistical literacy: What do culturally relevant/realistic tasks show us about the nature of pupil engagement with statistics? *International Journal of Educational Research, 55*, 57–65.
- Carvalho, L., & Monteiro, C. (2012). Reflexões sobre implementação e uso de laboratório de informática na escola pública. *Roteiro (UNOESC), 37*, 343–360.
- delMas, R.C., Garfield, J., & Chance, B.L. (1999, March). *Exploring the role of computer simulations in developing understanding of sampling distributions*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.
- Gal, I. (2002). Adult statistical literacy: Meanings, components, responsibilities, *International Statistical Review, 70*, 1–25.
- Konold, C., & Miller, C. (2011). *TinkerPlots: Dynamic data exploration. TinkerPlots Help Version 2.0*. Emeryville, CA: Key Curriculum Press.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*, 1–37.
- Manor, H., Ben-Zvi, D., & Aridor, K. (2013). Students' emergent reasoning about uncertainty exploring sampling distributions in an "integrated approach". In J. Garfield (Ed.), *International Research Forum on Statistical Reasoning, Thinking, and Literacy: Vol. 8*. Minneapolis, MN: University of Minnesota.

- Martins, M., Monteiro, C., & Queiroz, T. (2013). Compreensões sobre amostra ao manipular dados no software TinkerPlots: um caso de uma professora polivalente. *Revista Eletrônica de Educação*, 7, 317–342.
- Monteiro, C., & Ainley, J. (2004). Exploring the Complexity of the Interpretation of media graphs. *Research in Mathematics Education*, 6, 115–128.
- Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 97–108). New York: Springer.
- Prodromou, T. (2011). Students' emerging inferential reasoning about samples and sampling. In Clark, J., Kissane, B., Musley, J., Spencer, T. & Thornton, S. (Eds.). *Annual Conference of the Mathematics Education Research Group of Australasia, Vol. 34* (pp. 640–648). Alice Springs, Australia: MERGA34.
- Saldanha, L., & Thompson, P. (2002). Conception of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257–270.
- Saldanha, L., & Thompson, P. (2007). Exploring connections between sampling distributions and Statistical inference: An analysis of students' engagement and Thinking in the context of instruction involving repeated Sampling. *International Electronic Journal of Mathematics Education*, 3, 270–297.
- Watson, J.M., & Moritz, J.B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31, 44–70.
- Watson, J.M. (2004). Developing reasoning about samples. In Ben-Zvi, D. & Garfield, J. (Org.), *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Watson, J.M., Collis, K.F., & Moritz, J.B. (1995, November). *The development of concepts associated with sampling in grades 3, 5, 7 and 9*. Paper presented of at the Annual Conference of the Australian Association for Research in Education, Hobart, Tasmania.