ACTIVITIES TO LEARN FUNDAMENTAL IDEAS IN STATISTICS

Marcos Nascimento Magalhães
University of Sao Paulo, Brazil
marcos@ime.usp.br

In Brazil, basic concepts of Statistics are usually presented not only in introductory statistics courses at university but also in teacher training courses. Unfortunately, most of our teachers did not have an adequate statistics preparation during their undergraduate classes and it is necessary to recall basic ideas in workshops for teachers in service. Variability, randomness, distributions and parameter estimation are examples of concepts frequently misunderstood. In this paper we discuss activities to overcome such problems. We also discuss the student's opinion and assessments results. Simulation experiments with coins and dices, Data projects and Informal inference practices are examples of such activities. Their inclusion on regular or special course is also a way to promote collaborative and active learning process.

INTRODUCTION

In Brazil, as in any other country, Statistics has gained importance in almost all aspects of human life. It is impossible to be a full citizen without a minimum knowledge of graphs, tables and basic Statistics concepts. In Brazilian middle and high schools, Statistics topics are included in the contents of Mathematics then it becomes strategically important that Mathematics teachers have an adequate statistics preparation. In general, these teachers obtain a degree in Mathematics Education after attending one or two statistics courses, however, in several colleges and universities, these courses are not sufficient to provide them confidence to teach statistics. So, we have two facing fronts, one is to improve Statistics courses at universities and the other is to expand access to continuing education workshops in statistics for in service teachers.

For an overview of statistics presence in Brazilian basic education, one can consult Campos et al (2011). We observe, in recent years, an increasing presence of Statistics Education area in Brazilian scientific environment. In addition to special sections in the main national congresses, books and journals special issues had been published. For instance, Coutinho (2013), Lopes (2014), Samá & Porciúncula (2015), Educ. Mat. Pesq. (2016), Vydia (2016) and Lopes & Mendonça (2017) contain articles discussing theoretical issues and the use of statistics activities to teach at different school levels.

In order to improve teaching, it is important to illustrate concepts and numbers use in real context, enabling an adequate environment to enhance student learning. In Garfield & Ben-Zvi (2008), research studies were reported with several suggestions on teaching statistics and related issues. Also, the GAISE Report (GAISE, 2016) summarizes a set of suggestions, including the use of activities that allowed collaboration between students to improve their participation in the learning process.

The website *AtivEstat- Atividades de Estatística* (*Statistics Activities*) was developed at Mathematics and Statistics Institute of University of Sao Paulo, Brazil with activities to statistics teaching. It is free access (*www.ime.usp.br/ativestat*) and contains a list of related sites and a set of activities, which were created by team members or selected from several internet portals. They were chosen according to topic, accuracy concepts, simple implementation and adaptation to Brazilian contexts. In case of activities from other sites, an internet link helps access to original sources. *AtivEstat* is to be used mainly by teachers since there is no formal presentation of the statistics contents. The activities in the site are divided in five types: Classroom use, Spreadsheets, Project proposals, Films and Videos and Applets. Additionally, statistics subjects are separate in four blocks: Tables and Graphs, Summary Measures, Probability and Models and Statistical Inference. All activities have a short summary to help the teacher to choose which activity is adequate for his/her class.

In this paper, we report a few *AtivEstat* activities used to understand important Statistics basic concepts as randomness, variability and estimation. These activities were applied in regular courses and in workshops for teachers in service. We also comment on a feedback questionnaire answered by Math Education students a few months after regular Statistics basic courses.

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

DATA ANALYSIS ACTIVITIES

Mathematics Education students are required to complete two sequentially statistics basic courses in their undergraduate curriculum at the Institute of Mathematics and Statistics of the University of Sao Paulo, Brazil. The courses intend to provide statistical literacy and statistical reasoning as defined in delMas (2002).

During these courses, several activities are performed. In this paper we will mention two of the most important named here as *data analysis projects*. The objective was to use statistics tools, creating contexts so that students could try working with research questions based on a real data set. For the first course, Magalhães & Magalhães (2014) bring a discussion of the course activities and their teaching-learning bases on Vygotsky´s Social-Historical Theory (Vygotsky, 1934/1987).

For data analysis projects, groups of 3 to 6 students are created. During first course, students develop *Project 1* using a data set previously collected, since sampling method is not a subject of this course. However, in the second course, for *Project 2*, they collect their own data. The projects last several weeks and include a written report, an oral presentation, and a class discussion.

In Project 1 emphasis is to work with frequency tables, graphs as histogram and box plot, and quantile analysis. We wish to reinforce the fact that simple statistical tools can be a powerful instrument of analysis. In general, it is used spreadsheets to produce tables and graphs and few groups want to use specialized statistics software. During Project 1, students produce different frequency tables, according the sub-population considered and it serves to illustrate variability.

Project 2 is similar to Project 1, adding sampling and inference issues. Groups prepare a survey to obtain information from personal (or e-mail) interviews. It is important students have the experience of getting your own data, which is sometimes a difficult task. During group presentations, the whole class discusses on sampling difficulties, randomness and inference conclusions.

In order to evaluate the contributions of 2016 data analysis projects, a questionnaire was answered in a voluntary basis by 16 students, in March/2017. It had several items, but we will only comment on the three items referred to data analysis projects. In these items, we present a possible benefit of these activities and ask students to choose one option among: *Fully agree*, *Partially agree*, *Disagree*, *No opinion*. The results are shown in Figure 1.

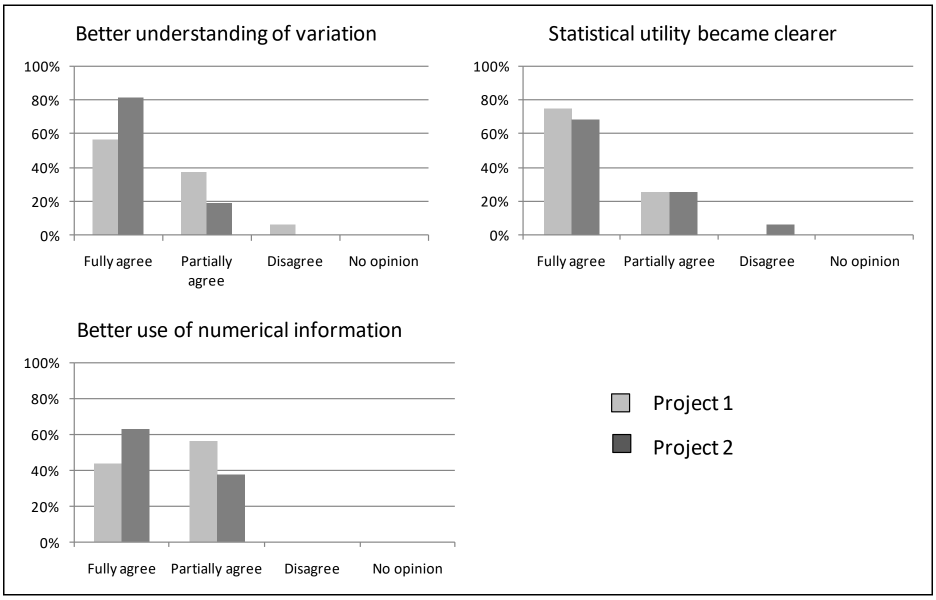

Figure 1. Students' opinion about data analysis projects

Results reveal that students think they have improved knowledge on variability. The answer *Fully agree* has the highest percentage in both projects with better score in Project 2, reinforcing the importance of students have collected their own data. With respect to usefulness of Statistics, students seem to gain insight on that according to Figure 1. Large majority of them marked the option *Fully agree*, indicating they feel they have developed better understanding of statistics utility after the data analysis projects. In another item, students' responses indicated they improved regarding the use of numerical information.

SHORT-TERM ACTIVITIES

In order to improve statistics conceptual understanding of Mathematics teachers in service, workshops are an important alternative. In general, they last 2 or 3 hours including a statistics review and 2 or 3 practical activities which can be applied later in classes of basic schools. In the sequence we presented four activities used in such meetings which take about 30 minutes each.

Fair dice?

Informally, a *fair dice* is an equilibrate dice with equal probability for each of its six faces. That is, any face occurs with probability 1/6. Participants are grouped in pairs and receive a dice which will be thrown six times. The dices can be of different sizes and colors or even, some of them, deliberately unbalanced. The activity objective is to discuss the dice behavior based on these tossing.

There is a question proposed: did you throw a fair dice? To force a decision, we require they answer only *Yes* or *No* to that question. Also, we asked them to register the results and decision in a simple form as exemplified in Figure 2. All the results are written in the blackboard and, to initiate class discussion, each group presents its decision.

Throw	1st	2nd	3rd	4th	5th	6th	Fair dice?
Face	2	3	2	4	6	3	Yes

Figure 2. Example of form in dice throwing activity

From our experience, mandatory dichotomy decision (Yes or No) incentives a polemic in the class with interesting arguments to support group decision. In general, groups that conclude they used an unfair dice, obtained results quite unbalanced. On the other hand, groups deciding for a fair dice were more tolerant to deviation from equally face frequencies. It seems they implicitly believe dices are always symmetric and equilibrate.

In fact, with only six throws there is not reliable decision and all class easily agree with that. The question that emerges is how many tossing would be enough to a safe decision? This is a great opportunity to discuss statistics objective. There are several similar statistics situations in which we will never be completely sure of the decision made, since always it will be errors involved. During the discussion, we comment on Large Numbers Laws and errors in the estimation processes whose are fundamental ideas in Statistics. The participants also had an opportunity to discuss about randomness, frequentist definition of probability, equally likely outcomes and uniform probability models. For a discussion on these issues, one can consult Chapter 3 of Batanero & Borovcnik (2016).

Less value auction

We suppose participants know mean and median definitions; nevertheless, it is quite common they confuse these two concepts in practical situations. In this activity, using dot plot graphs, we intend to reinforce the conceptual differences between these two important summary measures. The activity *Less value auction* was inspired in Dot plot summaries applet from Rossman-Chance Applets Collection (see http://www.rossmanchance.com/applets/).

It is convenient to select data sets to produce dot plot graphs according to a desired level of difficult. To present dot plot graphs in class we can use computer projection or, simply, an A3 size card. After presentation to the class, we ask students to guess mean and median values. They have a fixed amount of time to do that, say about 30 seconds. In the sequence, the true values are released and students compute their shots in the auction which are, for each measure, the absolute value of difference between guess and actual values. Auction initiates asking student´s bids. The instructor acts like in a real auction collecting successively smaller and smaller hunches until it reaches the lowest value bid among the students, who will be the winner of the auction. Figure 3 exemplifies bid calculation in an auction of mean.

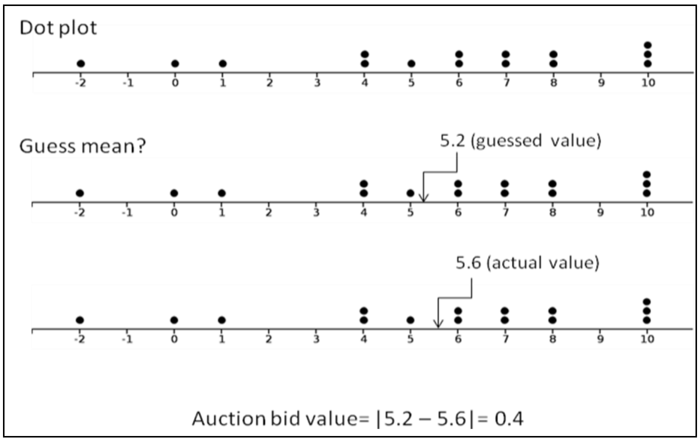

Figure 3. Example for Less value auction activity

We suggest starting with separate auctions for each measure and, then, performing a joint auction. In this case, the bid value is the addition of the two bids separately obtained. Our experience indicates that class, divided into pairs, produce better results. The preliminary discussion between each pair of students reinforces collaboration between them and helps to enrich class discussion. With this activity we can study outliers effect and the spatial location of mean and median.

Pi value

In this activity we invite students to use Geometry and Statistics knowledge to estimate the value of the irrational number π. For information about this number one can consult Wikepedia site at https://pt.wikipedia.org/wiki/Pi#.E2.80.9CHist.C3.B3ria.E2.80.9D_do_.CF.80.

We consider a circle inscribed in a unit square, consequently, the square area is 1 and the circle area is $\pi/4$ (see Figure 4). Supposing a random choice of a point in this square, there will be a $\pi/4$ probability of falling into the circle, which is the quotient between the circle and square areas. The objective is to estimate this probability to obtain a π estimate.

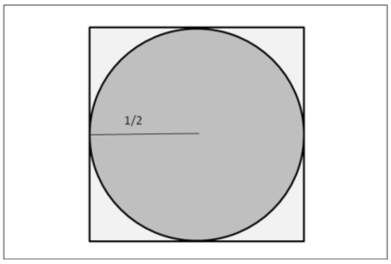

Figure 4. Square with circle inscribed

In classroom environment, we can use an A4 card with a figure as above to randomly select a point in the square. Folding the card in the edge of the square, we ask students to imitate a random choice throwing a small ball in it. Notice that the ball will be inside the square, but it can be in or out of the circle. It is convenient to obtain 20 or more points for each student or pair of students. After that, we can compute a relative frequency of points inside the circle, which will be an estimative of $\pi/4$.

An alternative procedure is to use a spreadsheet in a computer, obtaining point coordinates through a random number generator (see details in *AtivEstat* website). The manual procedure of throwing balls gives student hands on experience with randomness and it is preferential to the alternative one. However, this alternative procedure is convenient for high number of experiments, since by computer it is very easy to obtain large number of points.

Since students know π value (at least approximately), the activity allows a deeper discussion about frequency relative limits and informal inference reasoning. In this way, it improves understanding of frequentist definition of probability. Also, the different estimates obtained enrich class discussion about variability.

Sub-population comparison

In this activity, we discuss sample techniques and explore comparisons between two sub-populations. We provide class with a small data set containing a list with around 200 observations of three variables: sex, age and height. This is supposed to be a population data and it can be printed in A4 card. The list size must be large enough to avoid a visual knowledge of population characteristics. We use a data set collected from past students of regular Statistics courses.

We begin reviewing briefly different forms of collect data and commenting on quantiles as measures of location. It is propose a question that will be answered using sampling data and graphs, for instance, does the height behavior differ with respect to sex?

To develop the activity, class is divided into groups to collect 10 (or more) height data in each subpopulation of men and women. They build the central part of box plot diagram which uses 1st, 2nd (median) and 3rd quartile. After that, they will answer the proposed question and, then, the discussion begins in the classroom. It is important to mention we requested students to use a creative mode to mimic a random sample with replacement. The data collected are then ordered and the measures computed. In Figure 5, we show an example of the sequence of tables and graphs, as it would be obtained by a group of students in the activity.

Data collected:

Men	01	02	03	04	05	06	07	08	09	10
Height (mts)	1.62	1.69	1.69	1.81	1.78	1.61	1.72	1.71	1.59	1.72

Women	01	02	03	04	05	06	07	08	09	10
Height (mts)	1.60	1.67	1.56	1.62	1.58	1.72	1.62	1.63	1.55	1.75

Ordered values:

Men	01	02	03	04	05	06	07	08	09	10
Ordered	1.59	1.61	1.62	1.69	1.69	1.71	1.72	1.72	1.78	1.81

Women	01	02	03	04	05	06	07	08	09	10
Ordered	1.55	1.56	1.58	1.60	1.62	1.62	1.63	1.67	1.72	1.75

Measures:

1º quartile (3rd): Men = 1.62; Women = 1.58;

Median (mean of 5th and 6th): Men = 1.70; Women = 1.62;

3º quartile(8th): Men = 1.72; Women = 1.67;

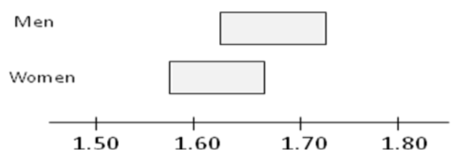

Figure 5. Example of group work in the activity

During the activity students have opportunity to reflect on the sampling process. Because they need to sample without the aid of any special apparatus, they produce their own method to collect data. In previous workshops, some groups just asked members to choose numbers and picked up the correspondent height in the data set list; others selected a number between 1 and 10 and choose their sample using systematic sampling; there were groups that threw a small ball in the card to indicate the heights to be chosen and groups in which one member closed his eyes and ran his finger on the data set table to indicate the value to be chosen. All these different sample choices offered an interesting material for debate in class, allowing students to discuss practical difficulties of performing a simple random sample with replacement in professional surveys.

In this activity students also reflect on variability and its measuring instruments. In general, time is limited in these workshops, so it would be convenient to focus on the main streaming of height behavior. For this reason we avoid building complete box plot. However, if time is not a problem, the complete box plot can be made with the correspondent discussion about outliers and whiskers size.

FINAL CONSIDERATIONS

The use of activities is very important to teach statistics concepts. As described here, activities allow students to work together, collaboratively involved, enabling collaborative relationships aimed at concepts development. In special, the use of data analysis projects is indispensable in regular undergraduate statistics basic courses. Students, when conducting experiments with randomness, recognize that variability is intrinsic to several situations. In this way, dealing with variability in activities helps the understanding of statistical conclusions limitations.

The short-term activities discussed here in the context of in service teachers workshops can also be applied in regular statistics courses. In *AtivEstat* site there are several activities aimed at enhancing all levels statistical teaching.

The presence of statistics in Brazilian educational scenario has increased in recent years. However, there is still a long way to go before a common citizen obtains an adequate understanding of basic statistical contents. Mathematics basic education teachers play a key role in this process. Forming a new generation of young people who would be statistically literate is a challenge that should motivate all of us to obtain better teaching alternatives.

REFERENCES

delMas. R. C. (2002). Statistical Literacy. Reasoning. and Learning. *Journal of Statistics Education 10*(3).

Batanero, C. & Borovcnik, M. (2016). *Statistics and Probability in High School*. Rotterdam: Sense Publishers.

Campos, T. M. M., Cazorla, I. M. & Kataoka, V. Y. (2011). Statistics School Curricula in Brazil. In C. Batanero, G. Burril & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*: A Joint ICMI/IASE Study,(pp. 5- 8). New York: Springer Science+Business Media.

Coutinho, C. Q. S., ed. (2013). *Discussões sobre o ensino e aprendizagem da Probabilidade e da Estatística na escola básica* [Teaching and Learning discussions on Probability and Statistics in basic school]. Campinas: Mercado de Letras (in portuguese).

Garfield, J. B. & Ben-Zvi, D. (2008). *Developing students statistical reasoning: Connecting research and teaching practice*. New York: Springer.

GAISE College Report ASA Revision Committee, (2016) *Guidelines for Assessment and Instruction in Statistics Education College Report.* http://www.amstat.org/education/gaise.

Educ. Mat. Pesq. (2016). *Educação Matemática Pesquisa v. 18, n. 3, Edição temática: Educação Estatística* (in portuguese). [Online: http://revistas.pucsp.br/index.php/emp/index]

Lopes, C. E., ed. (2014). *Os movimentos da Educação Estatística na escola básica e no ensino superior* [Statistics Education movements in basic and college levels]. Campinas: Mercado de Letras (in portuguese).

Lopes, C. E. ; Mendonça L. O. , eds. (2017). *Trilhas investigativas em Educação Estatística narradas por professores que ensinam Matemática* [Investigation paths in Statistics Education according mathematics teachers]. Campinas: Mercado de Letras (in portuguese).

Magalhães, M. N. & Magalhães, M. C. C. (2014). A critical understanding and transformation of an initial statistics course. *Statistics Education Research Journal* 3(2).

Samá, S. P.; Porciúncula, M. M. S., eds. (2015). *Educação Estatística: ações e estratégias pedagógicas no Ensino Básico e Superior* [Statistics Education: pedagogical actions and strategies for basic and college levels]. Curitiba: Editora CRV (in portuguese).

Vidya (2016). *Vidya v. 36, n. 2, Edição especial - Ensino de Probabilidade e Estatística* (in portuguese). [Online: http://periodicos.unifra.br/index.php/]

Vygotsky, L. S (1987). Thinking and Speech. In: R. W. Rieber & A. S. Carton, (Eds.) *The collected works of L. S. Vygotsky, Volume 1*. New York: Plenum Press (original work published 1934).