ANALYSIS OF TEACHERS' UNDERSTANDING OF COVARIATION IN THE *VITRUVIAN MAN* CONTEXT

Verônica Yumi Kataoka[1], Claudia Borim da Silva[2] and Irene Cazorla[3]
[1]State University of Santa Cruz, Brazil
[2]Univeristy of São Judas Tadeu, Brazil
[3]State University of Santa Cruz, Brazil
veronicayumi@terra.com.br

This work aims to analyze how 24 high school in-service teachers understand covariation, within the context of Vitruvian Man. This concept was explored in an informal way during a teaching intervention mode, by applying tasks in which teachers observed the height and arm span of students. They had to describe both variables, and answer whether it was possible to say that the height and arm span measurements were equal. They were also asked to construct the scatter plot with and without drawing the linear function. Just after the construction of the scatter plot and the linear function, teachers considered that the measurements of both variables were close. We believe that at the end of the set of tasks, the understanding of covariation in this group of teachers was improved, leading us to think that proposals such as these can aid the training of math teachers for teaching this topic in schools.

TEACHING COVARIATION

It is common in our daily life that we encounter various pieces of information presented in the media through tables, graphs and statistical measurements, such as, for example, the results of election polls and market, financial indicators, as well as situations referring to the association between two or more variables (height and arm span measurements, the number of study hours and grade obtained, plant growth and the amount of water deposited).

The reading and interpretation of these information's statistics require more and more from citizens' contextual understanding and knowledge, which enable them to evaluate the data critically and reason over their conclusions, allowing us to consider that an individual is statistically literate (Gal, 2002).

We highlight the fact that in situations involving the association between two variables, if the citizen domains the concept of covariation, he can check with competence the structure and the intensity of this bivariate correlation. Therefore, we consider that it is important to approach this concept at school; however, according to Peck and Gould (2005), Cazorla (2006), and Contreras, Batanero, Diaz, and Fernandes (2011), among other researchers, there are still gaps in math teachers' training regarding conceptual and statistical aspects.

Reflecting on the need for effective action to assist teachers' training in statistics teaching, particularly covariation, we have elaborated and applied to a group of 24 high school math teachers a set of tasks to introduce this concept in an informal approach (without calculating the covariation and Pearson's correlation), using Leonardo da Vinci's Vitruvian Man figure. From the results of the teaching intervention in this article, we aim to analyze the understanding covariation of these teachers in the solution of the proposed tasks.

According to Moritz (2004), covariation can be classified into three types: logical covariation, numerical covariation and statistical covariation. Logic covariation is defined by the logic variables involved, and can be classified as true or false; numerical covariation can be expressed by an equation involving the variables that have assumed real values, while maintaining the correlation in one value when related to another; and the statistical covariation that involves the correlation between two random variables, which vary over a numerical scale. We emphasize that this research deals only with statistical covariation.

In the development of logic covariation, Moritz (2004) states that it is important to propose instruction sequences involving the translation process (process to explore the same content covariation represented by situations different from each other), for example, the representations of numerical data, graphical representations, or verbal statements, whose goal is to evaluate the reasoning skills about: the generation of speculative data (GDE) – data generated by the student based on the interpretation of a verbal sentence (for example, a graph draft to represent a verbal

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

declaration); verbal interpretation of the graph (verbal statement to describe a scatter plot); and numerical interpretation of the graph (reading and interpolating the values on the scatter plot).

According to Zeiffler (2008), it is important to understand the aspects of statistical covariation: the structure (positive or negative association) and intensity (strong or weak) of the bivariate correlation, their causal role models and the prediction of events. According to this author, the topics involving covariation can be introduced at several moments in the curriculum and can be informally developed through data production and bivariate graph construction.

Watson and Suzie (2008) recommend that to teach the bivariate context in elementary school it is interesting to use familiar situations in order to facilitate for students the inference of informal ideas, as they should be initially encouraged to investigate a particular variable, for example, the arm span of a whole group, then analyze this same variable, but separate groups, such as boys and girls, and finally to evaluate the relation between two variables, such as arm span and height. In this study, the authors used the TinkerPlots software, which assisted in checking some measurements, such as the maximum, minimum, mean, mode, median, and frequency, and also facilitated several graphical representations using these data.

Watson and Suzie (2008) point out that although the use of these software tools in TinkerPlots is very important, especially when exploring scatter diagrams, if students are not familiar with this program yet, they can perform explorations of bivariate correlation through the difference between the height and arm span to check if the result is zero or close to zero. This research can also be done by dividing one measurement by the other, and checking how close from one (1) this result is (Silva, Magina, & Silva , 2010; Watson & Suzie, 2008).

METHODOLOGY

The subjects of this study were 24 high school math teachers (54.17% males). Initially we applied a profile questionnaire, and found out that in this group of teachers the average age was 36.05 years (standard deviation of 7.99) and the average time teaching was 13.15 years (standard deviation of 6.98), 75% studied statistics during their initial training and said that they enjoyed studying this content, only two teachers said they already knew the topic covariation, but none of them taught this concept.

When elaborating the set of tasks, we used the recommendations (first explore the univariable context, work the translation processes, assess whether other measurements may indicate the bivariate correlation) for the teaching of covariation of the authors mentioned in the previous section (Moritz, 2004; Silva et al., 2010; Watson & Suzie, 2008; Zeiffler, 2008), among other researchers.

The teachers created 12 pairs and the tasks were divided into five sets, with a total workload of six hours. During the first set, the teachers observed a database (called DB1) with the measurements of height and arm span of 12 students and had to separately describe the two variables, and answer whether it was possible to say that the measurements of the height and arm span of the pupils were equal.

In the second set, the tasks were the same as in the first set, but working with a second database (DB2) with measurements of the height and arm span of 50 students. It is noted that DB2 was also used in the third and fourth sets.

In the third set, it was requested that a dot plot construction was employed using a styrofoam plate and pins of different colors for each variable (Figure 1), and that they described separately the two variables, and also they should answer whether, from the dot plot, it was possible to state that the measurements of the height and arm span of pupils were equal.

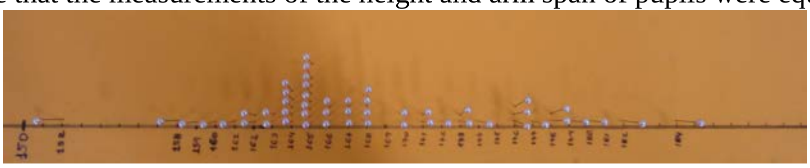


Figure 1. Dot plot height of 50 students

In the fourth set, we requested the construction of the scatter plot (also using the styrofoam plate and pins), and no theoretical line draft, and we questioned whether from the graph we could say that the measurements of the height and arm span of students were equal. Then, we asked them to draw the theoretical line (Figure 2) and the following question came up: *"Can you observe a proportional correlation between height and arm span? Justify your answer. (Show the strategy (ies) used)"*. And also answer: *"In the chart, did you use any measurement (s) to check whether the height and arm span of the students were equal? Which one (s)? Justify your answer."*

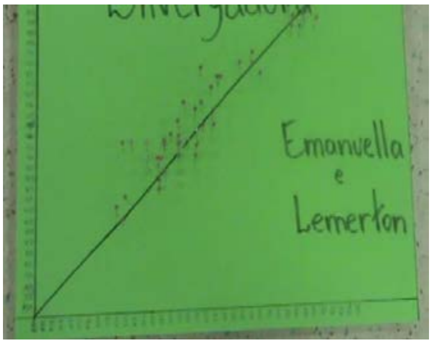


Figure 2. Scatter plot of height and arm span of 50 students

Before presenting the fifth set, we discussed the concept of covariation informally, discussing with the teachers the proposed tasks up to the fourth set. In the fifth set, the three questions proposed involved different contexts of the Vitruvian Man.

To evaluate the teachers' responses, in the tasks that are related specifically with the description of the correlation between height and arm span, we adopted four levels of knowledge about statistical covariation proposed by Moritz (2004) based on the SOLO taxonomy – Structure of Observed Learning Outcomes (Biggs & Collis, 1991), namely: Level 0 called NOT statistical, Level 1: Single Appearance, Level 2: Inadequate covariation and Level 3: Appropriate covariation.

For Moritz (2004), in the tasks involving the generation of speculative data (GDE), the responses at level 0 show a narrative context, but do not have a data set with more than one value for the variable or axis of a graph or values shown by numbers in spatial positions, without a context that indicates the variable. At level 1, the answers show a match in a single bivariate case, or a range of values for a single variable. At level 2, the correspondence is presented with inadequate variation for at least one variable: a variable having only two distinct values, or the variation is shown for each variable and in an inappropriate matching, that is, in the wrong direction. At level 3, the answers present both variables with matching fit between the variations of the values for each variable. For tasks involving verbal graphical interpretation (IGV), level 0 consists in a context without any variable or association or visual flaws. Level 1 shows a simple single point or a single variable (dependent). Level 2 refers to variables such as the correspondence is perceived comparing two or more points without generalizing or the variables are described and the correspondence is not mentioned or direction is not correct. Level 3 refers to all variables and indicates the correct direction. For tasks involving graphical and numerical interpretation (IGN), at level 0 the answers show flaws in the reading values on the axis, level 1 refers to a given value associating it to the bivariated values, without interpolating the data, level 2 makes references to values, but interpolates the points in the wrong positions, and level 3 reads the values and interpolates the points correctly.

RESULTS

We will discuss the results breaking them down by set of tasks.

1st Set of Tasks

In this 1st set, teachers used the DB1 with the measurements of the height and arm span of 12 students. We asked them to describe these two measurements, and one pair determined only the frequencies observed, while all the other pairs calculated measurements such as mode, median, average, total range, interquartile range (referring to the range of higher concentration data), maximum, and minimum, results that are similar to those found by Watson and Suzie (2008).

Then we questioned whether it was possible to say that the measurements of the height and arm span of the students were equal. All 12 pairs had their responses classified at level 0 of GDE, as they answered NO. And four pairs said that only 1/3 of the measurements were equal and the other eight pairs said that the differences between the two measurements were small, and could not be considered as equal anyway.

2nd Set of Tasks

In this 2nd set, teachers used DB2 with the measurements of height and arm span of 50 students. The description of the two measurements was made in a similar way to DB1, but with a decrease in the measurements calculated, probably due to the increase of data.

Other the task, we asked: "Is it possible to say that the measurements of the height and arm span of the students are equal? Justify your answer." All pairs continued responding NO (level 0 of GDE), explaining that many measurements were different. Only two pairs mentioned that in general the difference between the measurements was 2-5 cm .

3rd Set of Tasks

In this 3rd set, the teachers used DB2; they built the dot plot both for height and for arm span, and were asked to describe the two measurements. All pairs mentioned only the intervals where the highest concentration of data occurred. Two pairs referred to quartiles. They did neither calculate the central tendency measurement nor the dispersion. It was questioned whether the measurements could be considered equal after the construction of the dot plot. Just one pair responded YES, and their responses were classified at level 3 of the GDE. They justified this by stating that: "there is a similarity between the graphs, they tend to meet and disperse at close intervals; regarding the distribution of data, the dot plots of height and arm span preserved some similarities."

4th Set of Tasks

We asked the teachers, initially, to build the scatter plot between measurements of height and arm span. Analyzing this graph, the teachers were asked again if they could consider that these two measurements were equal. All pairs answered NO, but we note that 10 pairs justified their response, stating that to be equal, they would have to fit it in a linear function, or the results should be close to the line $y = x$; this way they were classified at level 2 of IGV. These reasons probably influenced the responses of five pairs in the second task. They had to draw the line $y = x$ in the scatter plot, and answer the same question. In their responses the teachers said that they could consider that the two measurements were equal because the points were concentrated near this line, then moving to level 3 of the IGV. Of all the pairs, two of them justified their response, considering also that the ratio between the two measurements was close to one.

We asked them to respond whether, besides the graph, they had used any measure to check whether the height and the arm span of the students were equal. One pair, besides the other two that had already responded in the previous task, said that they could verify whether the ratio of the two measurements was close to one. The other pairs answered NO, or considered measurements for this case as an inconsistent response; for example, two pairs answered: "Yes. Calculus was not done, but the estimation of the central tendency measure"; "Yes. I would use the arithmetic mean."

5th Set of Tasks – Extra Task Outside the Context of the Vitruvian Man

The first task of this 5th set was extracted from Moritz (2004), in which teachers had to consider the following situation:

> Ana and Clara developed a project to verify the study habits of six students in a particular school year. They prepared a questionnaire with two questions: How many hours do you spend studying for the math test? What was the grade that you took in the test? After the survey, Ana said: *"The longer the time spent studying, the lower the grade taken in the test."*

After that, the teachers had to answer: How do you interpret Ana's information? And they had to sketch two graphs, one that represented Ana's assertion and another that represented the statement: "Students that study longer get higher grades on the test." Both requested tasks used to assess reasoning skills about GDE.

For the first question, which involved a translation of the sentence for a verbal interpretation, three pairs' answers were classified at level 0, for example: "Ana did an analysis of data where there was a discrepancy between the results." The other nine pairs' answers were classified at level 3, relating the variables correctly and indicating the correct direction, for example: "There is an inverse relation between the hours of study and students' notes."

In the first graph construction, two pairs had their responses classified at level 0, six pairs at level 2, as they were not clear about the number of students who should be represented on the chart, although they presented correctly the direction of covariation, with explicit numbers, and axis labels indicating the correct direction (Figure 3a), and four pairs were classified at level 3, because they also had six points in the bivariate scatter plot and were clear about naming the coordinate axes (Figure 3b) .

In the second graph construction, one pair had a response rated at level 0, and all the other pairs at level 3.

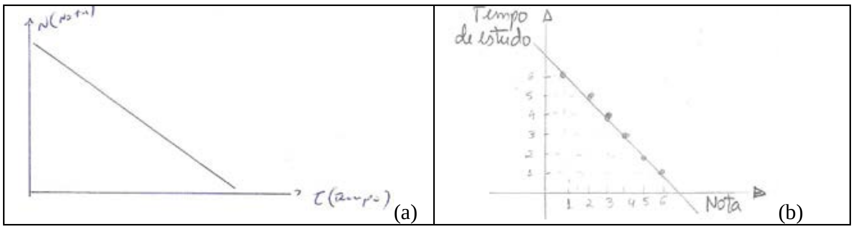
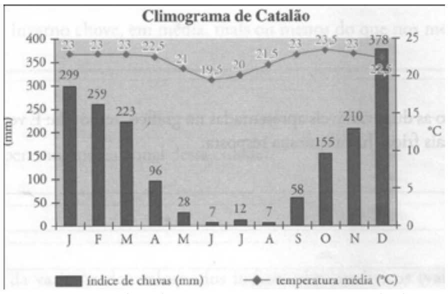


Figure 3. Graph example for two pairs, ranked at levels 2(a) and 3(b) of covariation

In the second task of this set, the question extracted from the fourth book of the third year of high school in São Paulo State (SEE-SP, 2010), teachers were supposed to watch the climograma (weather graph – Figure 4) and answer whether there was a relation between the temperature and rain rate, in addition to outlining a bivariate scatter plot that represented the temperature and rain rate.



Rain rate (mm) – Average temperature (ºC)

Figure 4. Weather graph relating the rain rate and temperature

For the first question, which involved the verbal interpretation of the graph (IVG), all the answers were classified at level 3. In the graph construction, the answers of four pairs were classified at level 0 and eight pairs at level 3.

The third task in this set was also proposed by Moritz (2004); it involved the verbal interpretation of the graph in Figure 5.

Noise level x number of people in the class

Figure 5. Noise level of six classes according to the number of people

We asked the teachers to write a sentence that could represent the graph information and that was a prediction of the noise level in a class that had 23 people. In the first question, two pairs' answers were classified at level 0 of IGV, because they had just commented on the context of the question, but made no association between the variables. The other pairs made responses at level 3. On the second question, three pairs' answers were classified at level 0 of IGN, for failing to make a correct prediction; that is, they did not provide the standard of the decreasing noise. The other pairs had their answers classified at level 3, predicting values of noise level between 60 and 65.

FINAL CONSIDERATIONS

According to the results found, we confirm that after the construction of the scatter plot with the identity function, 41.7% of the 24 teachers changed their opinion, stating that they considered the measurements of height and arm span were equal, and they realized, that way, the intensity of this bivariate relation, even if informally. Another positive result was to note that even without the identity function drawn on the scatter plot, 83.3% of the 24 teachers were able to evaluate the structure of this relation, when they said that the measurements were near that line. In the fifth set of tasks, after our teaching intervention, the majority of responses were classified at level 3, which is a level of covariation considered appropriate, noticing direct and inverse relations, and managing to make consistent oral statements with the results presented either in verbal or graphic form.

Therefore, we believe that at the end of the set of tasks, this teachers' group understanding of covariation had improved, leading us to think that proposals such as these can aid the training of math teachers for teaching this topic in schools, thereby providing a more adequate preparation of themselves and, therefore, of their students, for better reading and critical association and interpretation of the variables reported in the media.

REFERENCES

Biggs, J.B., & Collis, K.F. (1991). Multimodal learning and the quality of intelligent behavior. In H.A.H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement*. New Jersey: Lawrence Erlbaum Associates.

Cazorla, I. M. (2006). Teaching statistics in Brazil. In A. Rossman, & B. Chance, (Eds.), *Proceedings of Seventh International Conference Teaching Statistics*. International Association for Statistical Education, Salvador, Brazil [CD-ROM].

Contreras, J., Batanero, C., Diaz, C., & Fernandes, J. (2011). *Prospective teachers' common and specialized knowledge in a probability task*. Paper presented in Seventh Congress of the European Society for Research in Mathematics Education. Rzeszów, Polônia.

Gal, I. (2002). Adult's statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1-25.

Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227-255). New York: Springer.

Peck, R., & Gould, R. (2005). Preparing secondary teachers to teach statistics: A distance education model. *Proceedings of the 55th Session International Statistical Institute*, Sydney, Australia [CD-ROM].

São Paulo (2010). *Secretaria de Educação. Caderno de matemática do currículo do Estado de São Paulo*: Matemática e suas tecnologias: Ensino Fundamental e Ensino Médio. SEE/SP, 2010. 72p. [in Portguese].

Silva, C. B., Magina, S. M. P., & Silva, E.D. (2010). Homem Vitruviano. In I. M. Cazorla, & E. Santana (Eds.), *Do Tratamento da Informação ao Letramento Estatístico*. Itabuna: Via Litterarum, pp. 80-94 [in Portguese].

Watson, J., & Suzie, W. (2008). Building informal interference with TinkerPlots in a measurement context. *The Australian Association of Mathematics Teachers*, 22.

Zeiffler, A. (2008). Learning to reason about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 289-308). New York: Springer.