Advances in Social Science, Education and Humanities Research, volume 443
International Conference on Science and Education and Technology (ISET 2019)

The Identification of Pre-Service Mathematics Teachers' Statistical Reasoning on Descriptive Statistics

Yusfita Yusuf
STKIP Sebelas April
Sumedang, Indonesia
yusfita.penmat@stkip11april.ac.id

Hardi Suyitno
Universitas Negeri Semarang
Semarang, Indonesia

Y.L. Sukestiyarno
Universitas Negeri Semarang
Semarang, Indonesia

Isnarto
Universitas Negeri Semarang
Semarang, Indonesia

***Abstract* - Statistics is one of the disciplines used in solving problems that are always present in everyday life matter. Based on its function, statistics is divided into two types; they are descriptive statistics and inferential statistics. Descriptive statistics is the basis of statistical knowledge that students need to master, but in reality students have difficulties in learning statistics. One of the causes of these problems is the learning conducted by the teacher. Therefore, pre-service mathematics teacher need to have statistical reasoning in order to carry out a structured learning for students. The purpose of this study is to identify the statistical reasoning of pre-service mathematics teacher on Descriptive Statistics. The method used in this study is a qualitative method where the research subjects were taking basic statistics courses. Data collection was done by mean of tests, interviews and study documents. The results showed that only less than 50% of students who mastered the statistical reasoning on every aspect and on descriptive statistics. The cause of the lack of statistical reasoning on descriptive statistics is the students have low mathematical ability and the students' understanding of the basic concepts of descriptive statistics is not comprehensive.**

Keywords: statistical reasoning, descriptive statistics

I. INTRODUCTION

Generally people do not distinguish between the words statistics and statistic. Understanding statistics is usually associated with knowledge related to numbers or groups of numbers. Moore (1997) defined statistics as a tool to solve problems that always occur in everyday life, at work, and in science. Specifically, statistics is used to describe and predict phenomena using a collection of results from the measurements. For this purpose, Steel and Torrie (1995) defined statistics as nature and applied science regarding the creation, development, and the application of techniques in such a way that the uncertainty of inductive inference can be evaluated (calculated). In detail Sullivan (2008) stated that statistics is the science associated with gathering, organizing, summarizing and analyzing information to draw conclusions or to answer questions. Based on this opinion it can be concluded that statistics is a knowledge related to the ways of gathering materials or information, processing and analyzing, drawing conclusions and making reasonable decisions based on the analysis that has been carried out.

According to its function, Statistics is divided into two parts, namely: descriptive statistics and inferential statistics. The statistics that involves valid conclusions is called inferential statistics, whereas the statistics that only describes and analyzes groups of data provided without drawing conclusions about larger data groups is called descriptive statistics. Descriptive statistics includes central tendency, variability and distribution (Chan, Ismail, & Sumintono, 2016). Central tendency is a major component in conjecturing data analysis and graphical analysis in understanding distribution ideas (Garfield and Ben-Zvi, 2008). Central tendency consists of mean, median and mode. Variability is the same as dispersion and spread (Chan, Ismail, Sumintono, 2016).Variance consists of range, variance, standard deviation, and inter quartile range. According to Garfield and Ben-Zvi (2008) distribution is considered as one of the main and important ideas in statistics. Furthermore Wild (2006) classified the distribution into two; they are theoretical distribution and empirical distribution. Theoretical distribution shows that the probability model is the part of the normal distribution. The empirical distribution allows us to interpret variations in the data directly. The central tendency, the shape and distribution of data are the general characteristics of the distribution (Garfield and Ben-Zvi, 2008).

Descriptive statistics is the basis of statistical knowledge that students need to master, but in reality students have difficulty in learning statistics.

This was confirmed by Chan, Ismail and Sumintono (2013) who stated that students from elementary school level to university level faced difficulties in learning statistics. One of the causes of these problems is the learning conducted by the teacher. Teachers should use the stage of statistical reasoning in their learning (Rosidah, 2016). Although some studies show that teaching statistical reasoning can affect student learning achievement (Zuraida et al., 2012 and Tempelar, Van der Loeff & Gijlselaers, 2007). Therefore, statistical reasoning needs to be possessed by pre-service mathematics teacher (Garfield, 2002), so that it can be transmitted to students (Fennema & Franke, 1992; Heaton & Mickelson, 2002 in Karatoprak, Root and Börkan, 2015).

Statistical reasoning is defined as a way of reasoning with statistical ideas and understanding statistical information (Garfield and Gal, 1997). Lovett (2001) interpreted statistical reasoning is an action such as using statistical tools and concepts to make summaries, predictions, and draw conclusions from the data. In line with Lovett's opinion, Ben-Zvi and Garfield (2004) stated that statistical reasoning is a way of thinking using statistical information. Del Mas (2002) suggested that statistical reasoning is the ability to explain why and how an outcome is produced and why and how to draw the conclusions. Chan, S.W & Ismail Z (2014) stated that there are four key constructs of statistical reasoning assessment based on the framework of Jones et al, namely: 1) describing data; 2) organizing and reducing data; 3) representing data; 4) analyzing and interpreting data. Based on the opinions above, it can be concluded that statistical reasoning is a logical thinking process that includes describing data, organizing and reducing data, representing data, analyzing data and interpreting data, in other words, they can understand the statistical ideas and interpret them based on the conclusions obtained of the given data.

Considering the need for pre-service mathematics teacher to have statistical reasoning on descriptive statistics, therefore it is necessary to identify students' statistical reasoning. This has to be done so that learning statistical material conducted by students can be carried out systematically and structured. In addition, by knowing the reasoning abilities of students who have attended Basic Statistics lectures can be used as a benchmark in making improvements to the next statistical learning. Through this research, a thoroughly description of the statistical reasoning on descriptive statistics material of pre-service mathematics teacher candidates will be obtained.

II. RESEARCH METHODS

The method used in this study is a qualitative method. The research location is at STKIP Sebelas April Sumedang. Considering that descriptive statistics is studied in depth in the Basic Statistics course, therefore the subject of this research is the third semester students who are currently taking the course.

The main data source comes from students who took the written test. The written test is based on statistical reasoning which consists of 5 questions in the form of essays. The given reasoning problem includes the matter of aspects of data description (No. 1), data representation (No. 2a), organization and reduction (No. 2b), as well as analysis and interpretation (no. 3 and 4). The statistical reasoning instrument was developed based on the results of the research instruments of Chan, Ismail and Sumintono (2016). In addition to written tests, data collection was also carried out through interviews and documentation studies. Six studentswere interviewed who have been selected with the consideration determined by the researcher. Documentation studies were carried out during learning and statistical reasoning tests weretaken place. The leveling or grouping was done by matching the existing conditions with the characteristics of the grouping conducted by Chan, Ismail and Sumintono (2016).

Data analysis was performed using the constant comparative method. In general the data analysis process includes: data reduction, data categorization, synthesis and ended with a working hypothesis.

III. RESULTS AND DISCUSSION

Identification of students' statistical reasoning was done by first providing statistical reasoning test questions to 43 students who were taking the Basic Statistics course. Researchers have not been able to classify students at a certain level based only on the answers given at the reasoning test. Moreover, there are some students who did not answer the questions given on certain numbers. It is possible that students did not answer because of the limited test time. Therefore further tracing is carried out by conducting in-depth interviews and documentation studies. Interviews were conducted on six students. The selected students were three students with low scores and three students with medium scores.

Interviews were conducted a few times until a consistent answer was obtained. Based on the results of the interview, students claimed to experience anxiety when they were doing statistical reasoning test, it made them carelessly understanding and solving the given questions. This is consistent with the research result of Yusuf, Suyitno, Sukestiyarno and Isnarto (2019) that anxiety during exams is at the highest level.

Based on the results of tests, interviews and

document studies there are students who still do not know which concepts to use in working on a given problem; in the end they left the answer sheet blank. Some of them do not know which concept to use in solving problems, give answers carelessly or just copy the given questions into the answer sheet. Such conditions are in accordance with the characteristics of the Level of Reasoning Statistics (LPS) 0 developed by Yusuf (2017), it was found that students did not provide the answers and some students only copied the questions. This means that the student has no statistical reasoning ability.

In question no 1 there were students who did not complete or only completed some of the attributes on the graph but they gave relevant explanation. The student claimed that they focused on the question to elaborate of what was presented (the table) whereas they were asked to complete the graph first before explaining the answer. In addition, they also claimed to be accustomed to seeing graphics that were only given the title without any label on the axis and ordinate, even though they know what labels must be presented on a graph. Based on the comprehensive analysis of the data, it can be grouped into: (1) five students were aware of the completeness of the presented graph (Title, axis and ordinate) but did not complete it answered it partly incorrect (Verbal Level/Level 2); (2) twelve students were aware of the completeness of the presented graph (Title, axis and ordinate) correctly but did not provide a complete explanation (Procedural Level/Level 3); and (3) twenty students were aware of the completeness of the presented graph (Title, axis and ordinate) correctly and could explain and relate to the actual data or graph (Integrated Process Level/Level 5).

Problem 2a is a matter of aspects of data representation on data presentation material. In this problem there are also students who did not answer and gave short answers because they did not understand what the concepts to use in answering the questions. There are six students with this condition. There are five students who can correctly identify and provide a complete explanation (level integrated process/Level 5). Seven students are in conditions where they can provide identification but the answer is incorrect(idiosyncratic level/ Level 1), there are ten students who can provide correct identification results but cannot provide an explanation (Transitional level/ Level 3) and there are 15 students who can provide identification results but the explanation is incomplete (Procedural Level/ Level 4). Problem 1 and 2a are matters of data presentation material, only that the reasoning aspects are different. Presentation of data is an important material that needs to be mastered by pre-service mathematics teacher. At this time a lot of data presented in the form of tables or diagrams, therefore statistical reasoning on this material is needed. In addition, the data presentation material is one of the materials that often appear on the National Examination at various levels of education. But in fact, students who master this material are still under 50%. Whereas Owens and Clements (Agus et al., 2014) stated that visual portrayal provides a very important role in the preparation and method of problem solving as well as giving a very strong influence in the way of constructing reasoning.

Problem no 2b is a matter of aspects of data reduction and organization on data dissemination (variance). There are two students who answered perfectly who made a summary of the data using the data distribution correctly and can explain and relate to the actual data or graphs. There are 16 students who did not answer the questions because they did not know what concept to use (L0), there are 18 students who knew how to calculate the standard deviation but cannot interpret the formula used (Level 1), there are 4 students who knew how to calculate the standard deviation, can interpret the formula used, but were wrong in doing mathematical calculations so that the data summary is incomplete (L2), there are 3 students who knew how to calculate the standard deviation, can interpret the formula in use, and do mathematical calculations correctly but could not interpret the results (L3). Based on these data, the lack of mastery of reasoning on this material is largely due to students' mathematical abilities. This is in line with the results of research (Wilson and MacGillivray, 2006) which stated that the thinking ability and statistical reasoning is clearly connected to students' numerical abilities. Chiesi and Primi (2010) firmly stated that the mathematical knowledge acquired by students during high school has a direct and strong influence on statistical learning achievement. Therefore, Wilson and MacGillivray (2006) suggested that teachers should know the background of numeracy skills of students who are taking statistics. In addition to being weak in counting, students are also weak in interpreting results; this research shows that only 4.65% students are able to interpret the result. The observations are in line with the opinion of Leavy, Hannigan and Fitzmaurice (2013) which stated that students have difficulties when learning to interpret graphs and statistical results.

Problem 3 presents the aspect of data analysis and interpretation with material size of data concentration. In this problem there were 9 students who did not know the concept of central tendency at all (L0). In addition, there are still students who wereincorrect in understanding the concept of the median. The student determined the median value by using a way to determine the location of the median. Students like this are grouped at level 1 with a total of 6 people. In general, their mistakes in answering are caused by their knowledge of the central

tendency is not well integrated. Even though they know the definitions of mean, median and mode but they could not apply the three definitions in one problem. According to Chance, delMas and Garfield (Reading and Reid, 2010) students who could pick the correct definition but did not understand how the concepts are integrated, they are on the verbal level. In this study, there were 8 students with this condition. There are 3 students who were able to do analysis about the central tendencycorrectly but could not connect with the actual data (L3), there are 7 students who could conduct an analysis of the central tendencycorrectly and could connect with the actual data but did not provide appropriate explanation (L4), and there are 10 students who could determine the unknown value correctly, determine the median correctly, provide correct and complete information in accordance with the information provided in the problem. Students who mastered statistical reasoning on central tendency well are still below 25%. Students claimed that they were too focused on memorizing the definition and the procedure of counting without knowing the basic concepts of the central tendency. This also happened in a study conducted by Leavy (2010).

Problem 4 is a matter of aspects of analysis and interpretation with standard score material. In this material students are asked to compare two different data. More than 50% (22 people) students have understood how to compare two different data, where they considered the variability of the two data (L5). Only one person knew how to compare and calculate standard scores, but the student was wrong in doing calculations so that the conclusions they drew was wrong (L2). There are 2 students who knew how to compare and calculate standard scores, the results of calculations were wrong, but drew the conclusion correctly and there is one personwho knew how to compare and calculate standard scores, the results of calculations were correct, but they were wrong in making conclusions (L3). A total of 10 people did not know at all how to compare the different data (L0). There were 7 students who immediately compared the mean values of the two data without regarding to the variability (L1). They forgot that data can be compared if they have the same size (Karatoprak, Root and Börkan, 2015). Therefore, in this problem, of course, you must use the concept central tendency and variability in solving it.

IV. CONCLUSIONS

Based on the results of case studies conducted on students of the Mathematics Education Study Program STKIP Sebelas April Sumedang, the result obtained both from reasoning tests and from interviews and document studies isstudents have weak statistical reasoning on descriptive statistical material. There are less than 50% of students who mastered almost every aspect and descriptive statistics material. The causes of the lack of statistical reasoning on descriptive statistics are students' low mathematical ability and students' understanding of the basic concepts of descriptive statistics that are not comprehensive. Therefore, further research is needed to explore the factors that cause the lack of statistical reasoning of pre-service mathematics teacher so that they can be designed learning models that can improve statistical reasoning on descriptive statistics.

REFERENCES

[1] Moore, D. S. (1997). "New Pedagogy and New Content: The Case of Statistics". *International Statistics Review*, 65(2), 123-165.
[2] Steel, R. G. D. &Torrie, J. H. (1995). *Principles and Procedures of Statistics: A Biometrical Approach.* McGraw-Hill Book Company.
[3] Sulivan (2008). *Fundamental of Statistics*. Pearson; Prentice Hall
[4] Chan, Wei, Ismail, Sumintono. (2016). Assessing Statistical Reasoning In Descriptive Statistics: A Qualitative Meta Analysis. *Jurnal Teknologi (Sciences & Engineering)*,78 (6),p.29–35
[5] Garfield, J. B. & Ben-Zvi, D. 2008. *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. Springer.
[6] Wild, C. (2006). The Concept Of Distribution. *Statistics Education Research Journal*, 5(2), 10-26
[7] Chan, Ismail, Sumintono. (2013). *A* Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics. *International Conference on Innovation, Management and Technology Research,* Malaysia, Sept. 22 – 23.
[8] Rosidah. (2016). Analysis of Statistical Reasoning Process of Senior High School Students on the Size of Central Tendency (The Case Study For Student's Low Math Ability) .*International Conference On Research, Implementation And Education Of Mathematics And Science*. Yogyakarta, Indonesia, May, 16-17. p.225-232.
[9] Zuraida, Jaafar, Foo, Kien-Kheng., Rosemawati Ali. &Haslinda, Abdul Malek. (2012). Cognitive factors influencing statistical performance of diploma science students: A structural equation model approach. In Proceedings of *Langkawi Conference-ICSSBE*. p 562-566.
[10] Tempelaar, D. T., Van Der Loeff, S., And Gijselaers, W.H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior

reasoning abilities and course performance. *Statistics Education Research Journal,* 6 (2). p. 78-102

[11] Garfield, J. B. (2002). The Chalange of Develoving Statistical Reasoning. *Journal of Statistics Education*, 10(3).

[12] Karatoprak, Rabia, GülserenKaragözAkar, And BengüBörkan. (2015) Prospective Elementary And Secondary School Mathematics Teachers' Statistical Reasoning. *International Electronic Journal of Elementary Education*, 7 (2). p.107-124.

[13] Garfield,J.B. dan Gal (1997). *Teaching and Assesing Statistical Reasoning*. NCTM.

[14] Lovett, M. "*A Collaborative Convergence on Studying Reasoning Processes. A Case Study in Statistics*". In D Klahr And S. Carver (Eds). Cognition and Instruction Twenty-Five Years of.Mahwah: NJ Lawrence Erlbaum, 2001. P. 347-384.

[15] Ben-Zvi D. And Gafield, J. (2004). "*The Challenge of Developing Statistical Literacy, Reasoning, and Thinking"*. Kluwer Academic Publisher.

[16] delMas, R. (2002). Statistical Literacy, Reasoning, and Leraning: A Commentary". *Journal of Statistics Education*, 10(3).

[17] Chan dan Ismail. (2014). Developing Statistical Reasoning assessment Instrument for High School Students in Descriptive Statistics. *5th World Conference on Educational Sciences* - WCES 2013. Procedia – Socialand Behavioral Sciences 116. 4338 – 4344.

[18] Yusuf, Suyitno, Sukestiyarno, & Isnarto (2019), The Influence of Statistical Anxiety on Statistic Reasoning Of Pre service Mathematics Teachers. *Bolema*, 33 (64), p. 694 -706.

[19] Yusuf, Y. (2017). Konstruksi Penalaran Statistis Pada Statistika Penelitian. *Scholaria*, 7 (1), 60 – 69

[20] Agus, Mirian, Maria Pietronilla Penna, Maribel PeròCebollero, Joan Guârdian Olmos And ElianoPessa. (2014). The Aplication Of Graphical Representations InEstimatian Of Probabilistic Events. *Journal Of Theories and Research in Education*, 9 (1). p.235-252.

[21] Wilson, Therese & MacGillivray, H. (2006). Numeracy And Statistical Reasoning On Entering University. *International Conference on Teaching Statistics (ICOTS) 7*.Salvador, Brazilia, July, 2-7.

[22] Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' achievement. *Statistics Education Research Journal*, 9(1), 6-26.

[23] Leavy, A. M., Hannigan, A. And Fitzmaurice, O. (2013). "If You're Doubting Yourself Then, What's the Fun in That? An Exploration of Why Prospective Secondary Mathematics Teachers Perceive Statistics as Difficult," *Journal of Statistics Education,* v.21.n.3.p. 1-26.

[24] Reading, C. And J. Reid. (2006). "A Emerging Hierarchy of Reasoning of Reasoning about Distribution: From a Variation Perspective". *Statistics Education Research Journal, 5(2), 46-68.*

[25] Leavy, A. (2010). Teaching statistics at the primary level: identifying obstacles And challenges in teacher preparation from looking at teaching. *International Conference on Teaching Statistics (ICOTS) 8.* Ljubljana, Slovenia. Juli.