



The understanding of students towards statistics for research

Risnina Wafiqoh, Universitas Muhammadiyah Bangka Belitung, risnina.wafiqoh@stkipmbb.ac.id , ORCID: 0000-0001-8571-4659

Marina Zahara, Universitas Muhammadiyah Bangka Belitung, marinazahara91@gmail.com , ORCID: 0000-0002-3077-9583

Abstract. This study aims to determine the factors of causing students understanding towards statistics for research to be in the sufficient category. Data collection techniques used were tests and interviews. Based on the research, it was concluded that the causes of statistical understanding level of students were in the sufficient category: (1) Understanding based on their own wrong thoughts and logic; (2) Own thinking caused it does not remember the concept; (3) forgetting the statistical material which has been studied; (4) Don't know the formula; (5) Does not understand data analysis procedures; (6) the distance between statistics and thesis is far-off, so that what has been learned starts to be forgotten; (7) In statistical learning, the material used for research is not emphasized, only focused on giving the formula; (8) Do not understand the language of books that have been studied; (9) Don't want to ask people who are more understanding.

Keywords: Student Understanding, Statistics, Statistics for Research

Received: 22.11.2020

Accepted: 15.12.2020

Published: 11.01.2021

INTRODUCTION

Research is a mandatory requirement which must be done by final year students in every part of the higher education field. In conducting research, students are expected to be able to understand the use of statistics for research. Statistics has been widely used in daily life and even almost all fields using statistical methods in carrying out its activities (Supardi, 2013). Statistics is a branch of mathematics. Then if we want to improve the quality of statistical education, it means the same like we make improvements to the quality of mathematics education (Ulpah, 2009). Statistics is the science or art related to the procedures (methods) of data collection, data analysis and interpretation of the analysis results to obtain information for drawing conclusions and making decision (Noname, 2013). While research is a scientific technique to obtain data with specific purposes and functions (Sugiyono, 2011). It is very clear that statistics are needed in research, especially quantitative research.

Research is a mandatory requirement for final year students, and statistics is a science used to collect data, analyze data and interpret the results of research data analysis. Therefore, students are expected to be able to understand the use of statistics for research well. One of statisticians, a founding member of the ISF (International Statistical Forum), revealed that in the ISF founding event, it was stated that countries in Southeast Asia are countries which often practice rubber statistics, statistics are not in accordance with the demands of global community relations (Djauhari, 2007). Indonesia is one of the countries in Southeast Asia, this means that the practice of statistics in Indonesia has long been in the spotlight of the global community. In addition to these problems, the other problems in the field are writing thesis which students do not understand the use of statistics for research, and students do not know how to test, interpret and conclude the results of analysis, incompatibility between analytical tools and analyzed by students. Students are only able to use SPSS while the concept from the test equipment used in SPSS is not understood by students. Considering the importance of statistics for research, and looking at the problems above, this journal will discuss about "the understanding of students towards Statistics for research". The goal is: Knowing the factors that influence the students 'understanding of statistics for research are in the sufficient category and to find out students' misconceptions when using statistics for research. To determine the level of student understanding towards the use of statistics for research is looking at the test result given. The test in the form of a description given by the researcher to students is adjusted to the indicators of statistical ability [6], namely: (a) Understanding the concepts and scope of the data; (b) Grouping data; (c) Presenting data; (d) Calculating data; (e) Analyzing data. The test questions given are adjusted to the indicators of statistical ability to be able to measure understanding of the use of statistics for research.

METHODS

This research is classified as qualitative research, with descriptive methods. Qualitative research is a research procedure which produces descriptive data from people and observed behaviors (Margono, 2010). Understanding about the use of statistics for the final year research of STKIP Muhammadiyah Bangka Belitung students who were taking thesis courses was being the main focus of this research. This study involved the final year students of STKIP Muhammadiyah Bangka Belitung who were taking thesis subjects totaling 17 people to be tested using the test instruments which have been prepared. From the research subjects, 3 people were selected to be interviewed, consisting of 1 student with low ability, 1 student with medium ability and 1 student with high ability, which was able to represent all study populations and the ability category of students was based on written test results.

Data were collected using test and interview techniques. The test was used to determine the level of understanding of the final year students of STKIP Muhammadiyah Bangka Belitung on the use of statistics for research, as well as to determine the research subjects for interviews. The type of test given was description test which consists of 4 item adjusted for statistical ability indicators. Interviews were conducted to obtain data directly from research subjects regarding the understanding of students towards the use of statistics for research, what errors that students did in understanding the use of statistics for research and the causes of errors in the use of statistics for research. The validity of the data in this study used Triangulation. Triangulation is checking data from various sources in various ways and times (Satori & Komariah, 2010). In this study the triangulation used is source triangulation and technique triangulation. Source triangulation is carried out by searching more than one informan. The informan used was several students as research subjects. The technique triangulation was carried out using two data collection techniques, namely test and interview techniques.

RESULTS

Here are the analysis results of problem solving by Subject 1. Figure 1, 2, 3 and 4 are the answers of Subject 1 which are categorized as having low ability in solving problems given by researchers:

The image shows a handwritten student answer for question number 1. The question asks for the definition of a nominal value. The student's answer is as follows:

1) Definisi nominal adalah kemampuan peserta didik pada saat pembelajaran guru harus banyak menggunakan gambar atau animasi.

2) kemampuan kebidanan adalah suatu Praktek Reserbo didik dalam Praktek Pembelajaran di kelas. contoh cara berpakaian

3) kompetensi profesional adalah nilai atau prestasi yang dapat digambarkan guru berupa cara mengajar yg baik kepada peserta didik

4) kompetensi sosial adalah cara atau hubungan sekolah dengan lingkungan masyarakat.

contoh : hubungan pengajar dengan masyarakat kat smltar (lingkungan) social saling beres sama.

Saal termasuk rasio nilai

b) 0
= 0
3 0
1 0

Saal termasuk nominal nilai dapat berubah-ubah.

c) 0%, 1%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%

Saal a termasuk ordinal

c) nilai Lina = 9,8
Fani = 8,5
Pendi = 6,0

nilai rata-rata = $\frac{9,8 + 8,5 + 6,0}{3} = 8,1$

Umur sama dengan Lina sendi Loris mendapat nilai besar. Sekelompok kelas.

Saal termasuk data nominal karena nilai memiliki nilai yang teroris dan teroris.

Errors in understanding the concept

Figure 1. The answer of Subject 1 for the question number 1

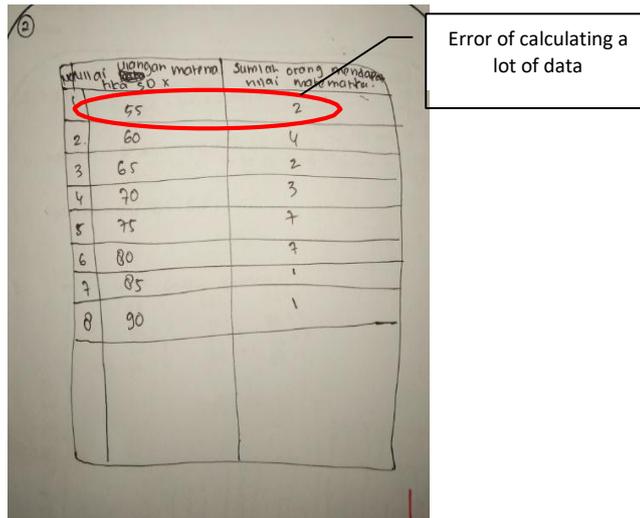


Figure 2. The answer of Subject 1 for the question number 2

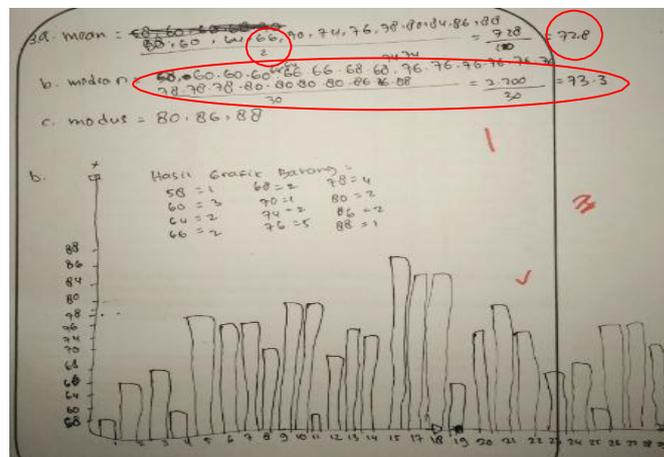


FIGURE 3. The answer of Subject 1 for the question number 3

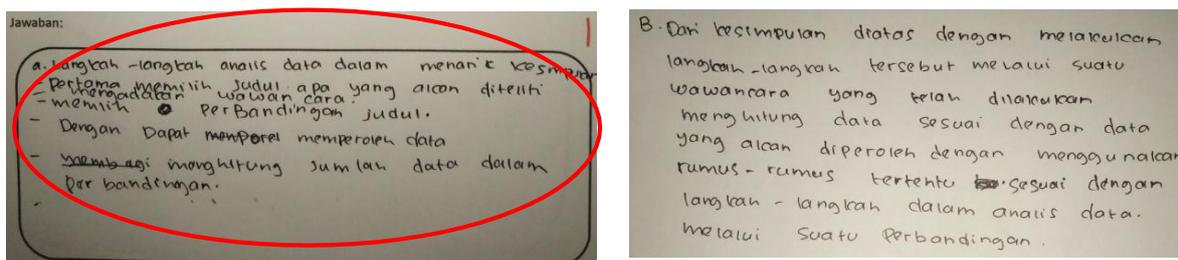


Figure 4. The answer of Subject 1 for the question number 4

Based on the picture above, in solving problem number 1, Subject 1 of research made a mistake in understanding the concept, that was in sentences b and c which should include ordinal data and ratio data, but first subject wrote both sentences including nominal data, by providing definitions that different and less precise regarding nominal data. In sentence d which should include interval data, however Subject 1 wrote d sentence including value ratio data. At the completion of problem number 2, Subject 1 made a mistake in understanding the concept and mistake of calculating a lot of data, which should make a frequency distribution table that was the value written in intervals, but Subject 1 wrote a lot of data. When calculating a lot of data, Subject I also experienced errors, 3 students got 55, but Subject 1 wrote 2 people. At the completion of problem number 3, the Subject 1 had mistakes; the first was in understanding the concept, which should have 76 mode, but Subject 1 wrote 80, 86, and 88, then the

second mistake was in writing mathematical symbols that should be "+" but the Subject 1 wrote ".", The third mistake was that there was no consistency in working on the problem, initially it was divided " 2 " and in the next step it was divided " 10 ". The last mistake was in substituting data. There should be 30 data, but Subject 1 just substituted 28 data. Whereas in solving problem number 4, the Subject 1 had errors in understanding the concept, which was supposed to write the procedure of analyzing data, but what was written was the procedure for starting the research.

Figure 5, 6, 7 and 8 were the results of Subject 2 in completing test questions:

1) a) Merupakan kalimat yang termasuk ke dalam data nominal karena berupa angka yang sesuai dengan data yang sudah ada.
 c) Merupakan kalimat yang termasuk ke dalam data interval karena kalimat tersebut mempunyai rentang / jarak.
 e) Merupakan kalimat yang termasuk ke dalam data rasio karena menggunakan simbol dan dapat dibandingkan.
 b) Merupakan kalimat yang termasuk ke dalam data ordinal karena menggunakan skala likert dan angka yang mempunyai arti seperti penskoran.

Figure 5. The answer of Subject 2 for the question number 1

Tabel distribusi frekuensi

Interval	x	y	$x^2 + y^2$
55 - 60			
60 - 65			
65 - 70			
70 - 75			
75 - 80			

Figure 6. The answer of Subject 2 for the question number 2

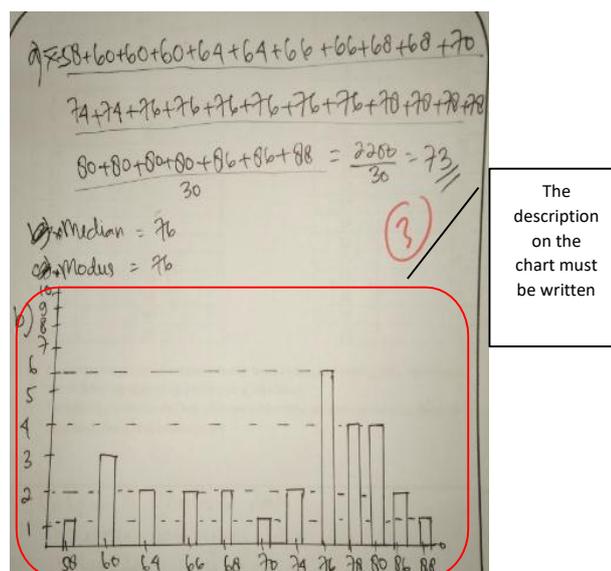


Figure 7. The answer of Subject 2 for the question number 3

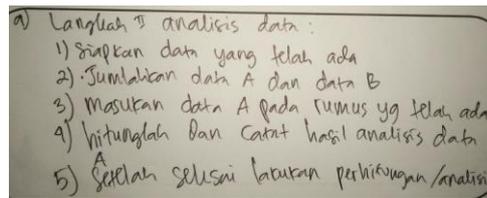


Figure 8. The answer of Subject 2 for the question number 4

Based on the picture 5, 6, 7, and 8, it can be seen that the Subject 2 solved problem number 1. Subject 2 got a mistake in understanding the concept, it was caused in sentence a that should include interval data, but Subject 2 wrote the sentence including nominal data, by providing a less definition regarding nominal data. In sentence c, it should be included ratio data. However Subject 2 wrote sentence c including interval data. In the sentence e, it should include nominal data. However Subject 2 wrote the sentence e including the ratio data. At the completion of question number 2, the Subject 2 was mistaken in understanding the concept. This can be seen what subject was written only wrote the table in the form of intervals, but without writing down the procedure for making the table. In addition, Subject 2 had a mistake in writing intervals and writing them repeatedly, such as the value of 60, 65, 70, 75, and there were two for 80. At the completion of question number 3, the subject II was mistaken in understanding the concept of the presentation of graphics, which should be written on the graph description. Whereas in solving problem number 4, the Subject 2 had errors in understanding the concept, which was supposed to write the procedure of analyzing data, but what was written was the procedure for starting the research. The subject 3 did the same kind of mistakes as Subjects 1 and 2 in understanding the concept and the mistake of using mathematical formulas.

It will discuss some examples of errors and causes of subjects in understanding educational statistics which are in the low category. The results of the interviews with subject I was:

- P : *From the questions asked, it was about nominal, ordinal, and ratio data, had you heard of it?*
 S1 : *(Nodding) Yes, I had?*
 P : *When?*
 S1 : *In statistics, when counting with Mr ******
 P : *Ohh, in statistics? Then why did you answer this? (Pointing to the answer of subject 1)*
 S1 : *I did not remember, Ms*

From the conversation, it can be seen the cause is that Subject 1 grouped the sentences into nominal / ordinal / ratio / interval data based on their own thoughts. It was caused by forgetting with concept. Below, these were the results of interviews with Subject 1. Subject 2 made a mistake for question number 1 as follows:

- S2 : *a data is nominal data because it is a number. The number you mean is 0° up to 100° C*
 P : *why nominal?*
 S2 : *Because there are numbers*
 P : *how about the others?*
 S2 : *Yes, yesterday I took it one by one, and then b I put it in ordinal. For c I put in interval (while showing the answer c). This was the ratio (while pointing to the answer e). it meant between a and d, which I thought nominal is a, because there were still 2 choices.*

From this conversation it can be seen that this error occurred due to subject II being less precise in understanding the definition of nominal data concepts. Subject 2 was not classifying the data based on data type concepts. Furthermore the results of the interview above, there were errors and other causes of Subject 2 making mistakes; these were the results of the interview:

- P : *You said that interval data had a range?*
 S2 : *Yes, Ms*
 P : *your answer was c, which one was the range?*
 S2 : *The c had a value of 98, Fani 85 rendi 60 (continue reading about the question). So the range of values was less between the value of fani and Rendi, I thought it was range.*

 P : *Which range was that?*
 S2 : *The range should be 10 to 20, but I was confused there was nothing like it, so I thought this one, because there was a gap between this and this, this and this. Actually the true ones were 5, 10, and 20. It was range.*

From this conversation it can be seen that this error occurred due to Subject 2 being less precise in understanding the definition of nominal data concepts. Subject 2 did not classify data based on data type concepts. Besides the conversation above, there were other conversations that show Subject 2 made a mistake.

- P : Had you ever met something like this before?*
S2 : Yes, in statistics, mathematics, it was made for a thesis.
P : was it important?
S2 : Important for everything
P : Important but there were people who did not understand what that was?
S2 : Learning of statistics had been a long time ago in some semester before, meeting it when thesis. For a thesis it was also not understood, wasn't it?

From these conversations, it can be seen that the cause of the error is a long statistical learning, and the distance among thesis is quite far and Subject 2 did not understand the thesis that he was working on at the moment. Besides the conversation above, there were other conversations that show Subject 2 made a mistake:

- P : In the past, were you ever taught this lesson?*
S2 : yes mom, But this separation was not too emphasized, Ms
P : For example, did not emphasize so? Why did not you ask?
S2 : Because all the data were given separately (data grouping was indicated). When learning about it, not at all met
P : ohh, there was no specific learning of grouping data?
S2 : Emm..em (Nodding)
P : So when were the statistics explained?
S2 : Just a little bit, sometimes the interval ratio appeared in the statistics, but if nominal was not focused here.

Based on the results of the interviews, Subject 2 made a mistake for question number 2 as follows:

- P : Seeing the interval you made, where is 90 from?*
S2 : ohh ... I forgot Ms, so I did not make it anymore. I forgot there was the number 90 here.

From this conversation it can be seen that this error occurred due to Subject 2 made a mistake in substituting the data. Furthermore the results of the interview above, there were errors and other causes of Subject 2 making mistakes; these were the results of the interview:

- P : Why were 55 to 60, why you did not specify 55 to 57 for example?*
S2 : Because I wanted even number Mom. So it was fitted Mom, if three to three was more, if four to four was less, but if I count five to five it was fit to the end

From this conversation, it can be seen the cause of the error is due to not understanding the concept of how to create a group distribution table and work on questions based on their own thinking and logic. Subject 2 did not make a mistake in question number 3, Mean, Mode, Median, and bar graphs that were accomplished by Subject 2 and those were all correct. However, for standard deviation question, Subjects 2 did not write the answer because they did not know the formula. Based on the results of the interviews, Subjects 2 made a mistake for question number 4 as follows:

- P : did you test the normality before the n-gain test?*
S2 : Yes, I did
P : What did it test?
S2 : Questions distributed to 3 children, and then the results of solving were calculated

From this conversation it can be seen the cause of the error due to not understanding the concept of data analysis procedures. Furthermore the results of the interview above, there were errors and other causes of Subject 2 making mistakes; these were the results of the interview:

- S2 : In my thesis I had a normality test, but I did not really understand what the normality test was and what it was for in my thesis.*
P : But did you understand the steps in the normality test?
S2 : I did not understand it, Ms.

From this conversation it can be seen the cause of the error because they do not know the flow of analyzing data. Subject 3 had the same type of cause with Subject 1 and 2, namely understanding based on own thinking and logic that was wrong and Subject 3 did not understand data analysis procedures. However, there were other types of causes experienced by Subject 3 that were different from other research subjects such as the following conversation:

- P : Had you ever seen anything like this before? (Pointing to table)*
S3 : I had, this followed from the books which had been seen.

- P : *But (shows the formula for many classes, class length) did not you know what for this?*
 S3 : *No Ms*
 P : *Had you read the book why in applying the formula was still not quite right?*
 S3 : *I did not understand the language of the book that I read Ms*

The aforementioned interview result above reveals that the cause of Subject 3 making a mistake was caused just knowing the formula but not knowing the usefulness of the formula and Subject 3 did not understand the language of the book that has been read and studied. Besides these causes, Subject 3 had other causes, such as the following conversation:

- P : *You ever studied about it, didn't you?*
 S3 : *Yes I did Ms, during education statistics class*
 P : *So why was not it applied?*
 S3 : *When taught, I felt confused, Ms*
 P : *Why did not you ask if confused*
 S3 : *Hmmmm!!*

From the conversation above it can be seen that the cause of Subject 3 made a mistake was because the Subject 3 did not want to ask to the person who was more understanding when Subject 3 was confused.

DISCUSSION AND CONCLUSIONS

Based on the results of interviews, the cause of statistical understanding level from Subject 1 was: (1) Understanding based on own thinking and logic that was error; (2) own thinking, it is caused forgetting the concept; (3) forgetting statistical material that has been studied. The cause of statistical understanding level from Subject 2 are: (1) Understanding based on own wrong thoughts and logic; (2) Do not know the formula; (3) Not understanding data analysis procedures; (4) the distance between statistics courses and thesis is too far, so that what has been learned starts to be forgotten; (5) In statistical learning, the material used for research is not emphasized, only focused on giving the formula. The cause of statistical understanding level from Subject 3 are; (1) Do not understand the language of books that have been studied; (2) Don't want to ask people who understand better; (3) Do not know the concepts and procedures of data analysis; (4) understanding based on own wrong thoughts and logic.

Based on the results of tests and interviews, mistakes made by Subject I are; (1) Error with understanding of concepts; (2) Error of making the own formula; (3) Errors in the use of mathematical symbols. The mistakes made by the subject 2 are; (1) Error with understanding of concepts; (2) Error in knowing the usefulness of the formula. While the mistakes made by research Subjects 3 are: (1) Misconceptions about concepts; (2) Errors in the use of mathematical formulas.

REFERENCES

- Djauhari, M. A. (2007). *Statisika: Salah Satu Indikator Utama Peradaban. Jurnal Sositologi*.
 Margono, S. (2010). *Metodelogi Penelitian Pendidikan*. Jakarta: Rineka Cipta.
 Noname. (2013). *Konsep & Pengetian Statistika*. Retrieved from Statistika Ekonomi.
 Satori, D., & Komariyah, A. (2010). *Metodelogi Penelitian Kualitatif*.
 Sugiyono. (2011). *Metodelogi Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
 Supardi. (2013). *Aplikasi Statistika Dalam Penelitian*. Jakarta: Change Publication.
 Ulpah, M. (2009). *Belajar Statistika: Mengapa dan Bagaimana? Jurnal pemikiran alternatif kependidikan*, 3(14).
 Waru, M. V., Minggu, I., & Annas, S. (2015). *Deskripsi Persepsi tentang Statistika Ditinjau dari Tingkat Kemampuan Berfikir Statistis (Statistical Thinking) pada Mahasiswa Jurusan Pendidikan IPS*