UJME 14 (1) 2025: 1-11
UNNES JOURNAL OF MATHEMATICS EDUCATION
https://journal.unnes.ac.id/journals/ujme
ISSN: 2252-6927 (print); 2460-5840 (online)

The Statistical Thinking Process of Senior High School Students in Solving Data Centralization Problems

Shofi Farihah Muazarah[a,*], Toto Nusantara[a], Sisworo[a]

[a] *State University of Malang, Jl. Semarang 5 Malang, East Java, Indonesia*

** E-mail address: shofimuazarah@gmail.com*

ARTICLEINFO

Article history:
Received 20 January 2025
Received in revised form 10 February 2025
Accepted 18 February 2025

Keywords:
Statistical Thinking; Problem Solving; Data Centralization Problem.

Abstract

This study aims to describe the statistical thinking process of high school students in solving mathematics problems. The method used was descriptive explorative with a qualitative approach, which was conducted at Maarif Lawang High School. The research subjects consisted of 24 students in class X MIPA who had taken statistics material. After being investigated, there were 3 out of 24 students who had solved the problem correctly. Based on the statistical thinking process, it can be grouped into two, namely the complete group and the incomplete group. One person in each group was taken as a group representative. Students completed the test with one data centralization problem. Written exams and in-depth interviews were the means of data collection. The data were next subjected to analysis using data reduction techniques, data display, and the drafting and verification of conclusions. Triangulation techniques were used to assess the veracity of the data. Based on the study's findings, Subject 1 accurately solved the problem and demonstrated proficiency in all four stages of the statistical thinking process: describing data presentation, organizing and reducing data, representing data, and analyzing and interpreting data. Although Subject 2 is capable of finding the proper solution, they are lacking in three key areas of statistical thinking: explaining the presentation of data, organizing and decreasing data, and expressing data.

1. Introduction

Statistics is a branch of mathematics that includes activities related to data collection, data processing, data analysis, and drawing conclusions based on the results of data analysis (Nani, 2015). In the current era of globalization, statistics plays an important role in our daily lives, not only in terms of knowledge and theory Obrial & Lapinid (2020). This is in accordance with the statement Meletiou-Mavrotheris & Mavrotheris (2007) that "Statistics education is the focus of reformers in mathematics education as an important aspect of citizen education in a democratic society". Statistics is at the core of mathematics education reform.

In addition to the field of education, statistics also has an important role in human life to solve the problems faced. Moore (1997) argues that statistics is considered a tool for solving problems in everyday life, in the workplace, and in science. Almost all activities are related to statistics, for example, counting population, using social networks, in the market, in the laboratory, and even in government. Therefore, a basic understanding of statistics is very important and necessary to be able to solve the problems faced can be resolved correctly and accurately.

In order to develop static thinking, careful consideration needs to be given to the way people learn statistics. The thinking process is a learning activity that takes place by involving mental processes that occur in the student's brain, thus learning becomes an activity that is always associated with the thinking process (Darma, 2013). Martadiputra & Suryadi (2012) defines statistical thinking as the ability to recognize and understand the entire statistical process and apply this understanding to everyday problems by making judgments, analysis and generalizations involving the statistical thinking process. Several stages of the statistical thinking process can be used as a reference to determine whether someone is able or

To cite this article:
Muazarah, S. F., Nusantara, T., & Sisworo (2025). The Statistical Thinking Process of Senior High School Students in Solving Data Centralization Problems. *Unnes Journal of Mathematics Education*, *14*(1), 1-11. https://doi.org/10.15294/ujme.v14i1.20106

accustomed to thinking statistically. These stages are: *describing* data *displays*, *organizing* and *reducing* data, *representing* data, and *analyzing and interpreting* data (Jones et al., 2000; Mooney, 2002; Zahner & Corter, 2010; Martadiputra & Suryadi, 2012).

Research on statistical thinking in solving problems has been done before. One of them is research belonging to Naili et al. (2021), which observes about students' statistical thinking when solving descriptive statistics problems in terms of personality type. Individuals who tended towards introversion scored higher on all measures of statistical thinking, including those that measured their ability to describe data presentation, organise and reduce data, represent data, and analyse and understand data. Contrarily, the extroverted individual has failed to demonstrate all signs of statistical thinking, particularly when it comes to understanding and analysing data.

Based on this description, the importance of statistical thinking in learning statistics, therefore it is necessary to have statistical thinking skills at all levels of education in Indonesia. By knowing students' statistical thinking skills, it is hoped that each student will be able to understand and interpret statistical messages in various contexts and enable students to draw conclusions or make decisions accurately based on information or data (Masjudin et al., 2020). Therefore, researchers need to analyze how the statistical thinking process of high school students in solving data centralization poblems?

2. Methods

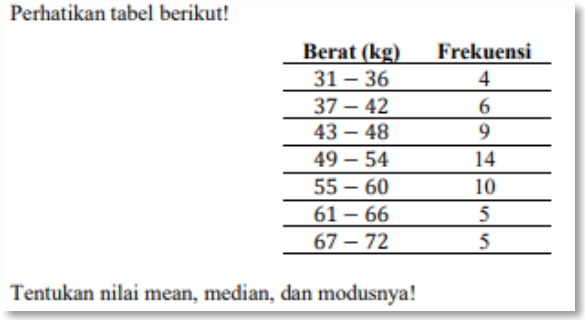
This type of research was descriptive and explorative with a qualitative approach. The main objective was to describe the statistical thinking process of high school students in solving mathematics problems. This research was conducted at Maarif Lawang High School. The subjects of this study were students in grade X MIPA in the 2023/2024 school year who had taken statistics material. The subject selection process began with giving a test on data centralization questions and was strengthened by verbal explanations through interviews. There are 3 out of 24 students have solved the problem correctly, found 21 other students are still wrong when doing calculations so that the final results obtained do not match the correct answer.

Based on the stages of the statistical thinking process, it can be grouped into two groups, namely complete and incomplete groups. The complete group is students who solve the problem correctly and fulfill the 4 stages of the statistical thinking process. Likewise, the incomplete group is students who solve the problem correctly but have not fulfilled the 4 stages of the statistical thinking process. Then one student in each group was taken as a group representative and analyzed the statistical thinking process further. The results of the grouping of subject work based on the stages of statistical thinking can be seen in the following Table 1.

Table 1. Grouping of Subjects Based on Statistical Thinking Stage

Statistical Thinking	Subject 1	Subject 2
Describe data presentation (DD)	✔	✔
Organizing and reducing data (OD)	✔	✔
Representing data (RD)	✔	✔
Analyzing and interpreting data (AI)	✔	–

The researcher acted as the main research instrument by collecting data, analyzing, interpreting, and reporting the results of the study. The tools used in this research were data centralization test questions and interview guidelines. The Statistical Thinking Test Instrument could be seen in the following figure:


Perhatikan tabel berikut!

Berat (kg)	Frekuensi
31 – 36	4
37 – 42	6
43 – 48	9
49 – 54	14
55 – 60	10
61 – 66	5
67 – 72	5

Tentukan nilai mean, median, dan modusnya!

Figure 1. Data Centralization Problem Test

The data analysis technique referred to the point of view of Miles & Huberman (1994) that described the interactive and ongoing nature of the actions involved in qualitative data processing. Data analysis encompassed a wide range of tasks, such as data reduction, data display, conclusion formulation, and verification.

Statistical thinking criteria, including data presentation description, data organisation and reduction, data representation, and data analysis and interpretation, were used to examine the data (Mooney, 2002). Technical triangulation, which comprises comparing written and interview test results, as well as analysing data from the same source using multiple approaches, ensures the validity of the data utilised in this study (Sugiyono, 2017).

3. Results & Discussions

Students of class X MIPA solve the problem procedurally. In more detail, the researcher will present the results of students' answers in solving mathematics problems based on students' statistical thinking skills. The following description aims to see the statistical thinking process of each student.

Result and Discussion of Statistical Thinking Process of Subject 1 (S1)

The first analysis is to analyze the statistical thinking process in subject 1 (S1). The results or answers to data centralization questions can be seen at Figure 2.

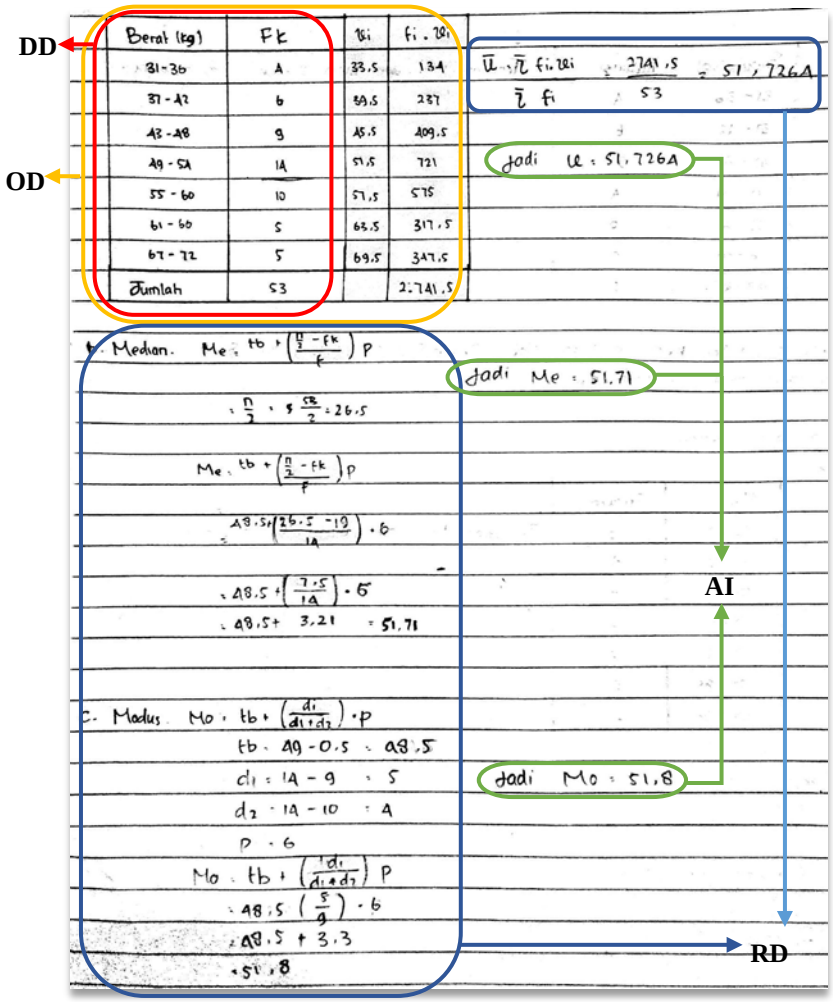


Figure 2. S1 Answer in Solving Data Centralization Problem

S1's statistical thinking process, starting from the stage of describing data presentation (DD) by being able to understand the problem by reading, the subject describes the problem information according to the

problem and the subject reveals starting from the data information in the problem in the form of a table display and what is asked in the problem. This means that S1 students have the ability to describe data presentation. The results of the work can be seen in Figure 2 with code DD. This is in accordance with the snippet of the researcher's interview (P) with S1 in the DD stage as follows.

P : "What did you do after getting this problem?"
S1 : "Preceded by reading and understanding first what is meant in the problem, and rewriting the known table"
P : "What do you know from this problem?"
S1 : "From the information in the problem, it is known that the lowest weight is 31 kg and the highest weight is 72 kg. Then, what is asked in the question is the mean, median, and mode values of the data"
P : "Then what are you doing?"
S1 : "Apart from writing the table as in the question, there is information in the question that shows that the number of frequencies is 53 people.
P : "In the table, can you explain why it says that the score range for each class is 6 numbers?"
S1 : "Yes, because the difference between the highest and lowest weight is $72 - 31 = 41$. Then I chose the number of classes as 7, so the length of the class is $41/7 = 5.86$ and rounded to6 . From there, the length of each class is made as many as 6 numbers, such as the first class with a weight of $31 - 36$, meaning that the weight between 31 to 36 is included in the first class. And don't forget to fill in the frequency, which is how many people weigh in each class."

At the stage of organizing and reducing data (OD), S1 performs several stages, the first is to find and determine the middle value of each class and written in the column (x_i). S1 looks for and determines the middle value of each class by sorting first. For example, the 1st class is $31 - 36$ where the numbers consist of $31, 32, 33, 34, 35, 36$. The median / middle value of the 1st class is between 33 and 34 , because there are 2 numbers, then S1 does the sum of the two numbers and divided by two. So that the middle value of the 1st class is $\frac{33+34}{2} = 33.5$. Likewise, in the next class, S1 determines the middle value in the same way for each class. The second stage S1 calculates the multiplication between the frequency of the *i-th* class and the middle value of the *i-th* class ($f_i \cdot x_i$) . For example, in the 1st class with a weight of $31 - 36$ kg there is a frequency (f_i) of 4 and the middle value (x_i) obtained is 33.5 , then the result of the multiplication is $f_i \cdot x_i = 4 \cdot 33.5 = 134$, as well as in the next class. The results of the work can be seen in Figure 2 OD code, and this is in accordance with the excerpts of the researcher's interview (P) with S1 in the OD stage as follows.

P : "After that, what did you do?"
S1 : "First find the middle value of the *i-th* class and then calculate the frequency of the *i-th* class multiplied by the middle value of the *i-th* class($f_i \cdot x_i$) , and do not forget to write it in the table"
P : "Then, what are you doing?"
S1 : "Because I knew it was group data, and what was asked was the mean, median, and mode, so I worked according to the formula"

At the stage of representing data (RD), S1 performs the calculation process of mean, median, and mode according to each formula for group data. In the process of calculating the mean, S1 calculates the formula $\bar{X} = \frac{\sum f_i \cdot x_i}{\sum f_i}$. By first finding the sum of the calculation of the frequency of the *i-th* class multiplied by the middle value of the *i-th* class ($\sum f_i \cdot x_i$) and the sum of the frequencies of each class ($\sum f_i$), S1 found the mean $\bar{X} = 51.7264$. In the median calculation process, S1 calculates the formula $Me = tb + \left(\frac{\frac{n}{2} - Fk}{f}\right) \cdot p$. First, S1 looks for the lower edge of the mode class (tb), the number of frequencies (n), the amount of data before the median class (*Fk)*, the amount of data in the median interval class (*f)*, and the length of each class (p) which will be used in the median calculation process. So that the final result of S1 in the median calculation process is $Me = 51.71$.

Furthermore, in the process of calculating the mode S1 performs calculations with the formula $Mo = tb + \left(\frac{d_1}{d_1 + d_2}\right) \cdot p$. Before doing the calculation, S1 first looks for the values to be used, such as the value of

the lower edge of the mode class (tb), the frequency value of the mode class to the frequency of the class before the mode (d_1), the frequency value of the mode class to the frequency of the class after the mode (d_2), and the class interval (p). After finding these values, S1 did the calculation, and found the mode result $Mo = 51.8$. The results of the work can be seen in Figure 2 RD code, and this is in accordance with the excerpts of the researcher's interview (P) with S1 in the RD stage as follows.

P : "What formula did you write?"
S1 : "Mean, median, and mode formulas for group data"
P : "Explain the process of calculating the median in group data"
S1 : "The median calculation process is that the lower edge value (tb) is summed with the result of the number of frequencies (n) divided by two and then subtracting the amount of data before the median class (Fk) and then divided by the amount of data in the median interval class (f) and the result is multiplied by the length of each class (p)."
P : "What about the mode formula for group data?"
S1 : "The mode formula is the lower edge of the mode class (tb) plus the result of the frequency of the mode class to the frequency of the class before the mode (d_1) divided by the sum of the frequency of the mode class to the frequency of the class before the mode and the frequency of the mode class to the frequency of the class after the mode ($d_1 - d_2$) then multiplied by the class interval (p)."

At the stage of analyzing and interpreting data (AI), S1 draws conclusions from the results that have been obtained, by writing "So, $\bar{X} = 51.7264$ ", then "So, $Me = 51.71$ ", and "So, $Mo = 51.8$ ". In addition, S1 has also checked the answers first before submitting them, so S1 already knows that the results in the calculation process are correct or not. The results of the work can be seen in Figure 2 AI code, and this is in accordance with the excerpts of the researcher's interview (P) with S1 in the AI stage as follows.

P : "Did you double check your answers?"
S1 : "I checked my answers before submitting them"
P : "Did you make the conclusion correctly?"
S1 : "I have listed the conclusions obtained, namely the mean value of 51.72, for the median value is 51.71, while the mode value is 51.8"

Based on the analysis of the answers and also the results of the interview, it can be concluded that S1 already understands the concept and can find the solution of the problem given. In general, S1's thought process is presented in Figure 3.

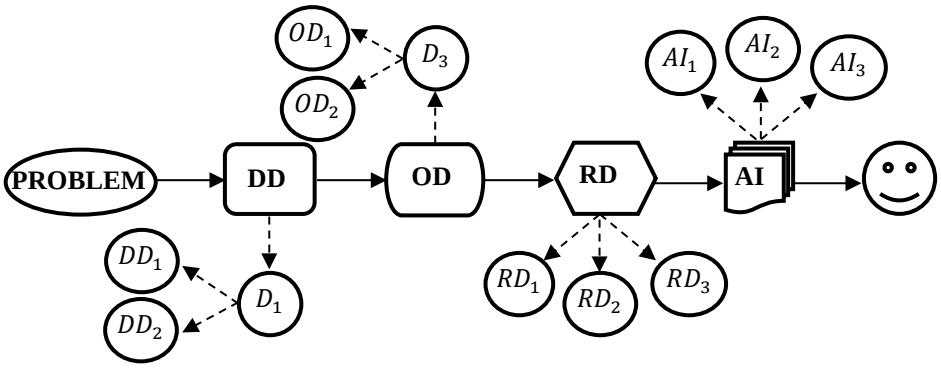


Figure 3. S1 Statistical Thinking Process

For clarity, a description of the coding for S1's thought process when solving problems based on the stages of statistical thinking will be presented in Table 2.

Table 2. Description of Figure 3. S1 Statistical Thinking Process

Code	Description
DD	Describe the data presentation. Characterized by mentioning known data information and written in the form of a table (D_1). The table consists of 2 columns, the first is the weight column (kg) with the range of values for each class predetermined (DD_1). The second column is the frequency or the amount of data for each class (DD_2).
OD	Organizing and reducing data. Marked by answering (D_3), namely determining the middle value of the *i-th* class (OD_1) and calculating the multiplication between the frequency of the *i-th* class and the middle value of the *i-th* class (OD_2).
RD	Representing data. This activity is characterized by the process of calculating the mean (RD_1), median (RD_2), and mode (RD_3) according to the respective formulas on group data.
AI	Analyzing and interpreting data. Marked by checking back and mentioning the conclusions obtained that the mean value is 51.72 (AI_1), the median value is 51.71 (AI_2), while the mode value is 51.8 (AI_3).

Result and Discussion of Statistical Thinking Process of Subject 2 (S2)

The second analysis is to analyze the statistical thinking process in subject 2 (S2). The results or answers to data centralization questions can be seen at Figure 4.

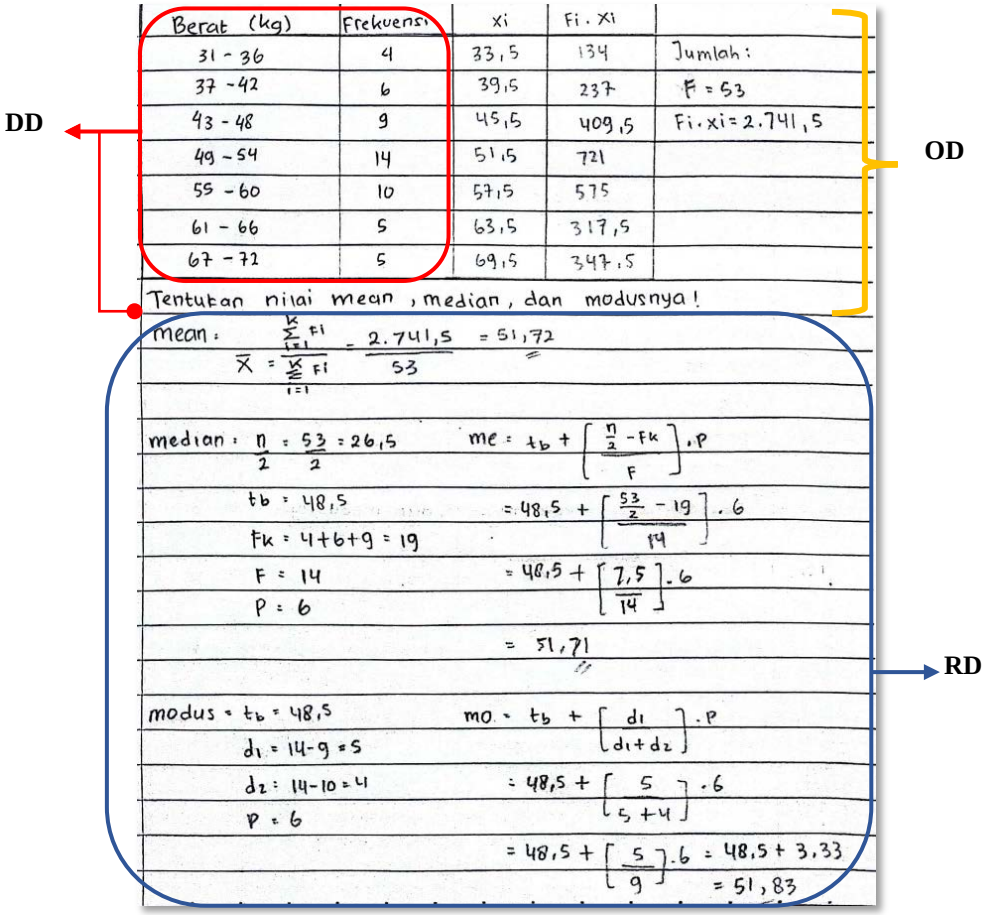


Figure 4. S2 Answer in Solving Data Centralization Problem

S2's statistical thinking process when solving the problem starts from the stage of describing data presentation (DD) by being able to present data in the form of tables, understanding the problem by reading, the subject describes the problem information according to the problem and the subject reveals starting from the data information in the problem in the form of a table display. Besides that, S2 also wrote what

was asked in the question. The results of the work can be seen in Figure 4 DD code. This is in accordance with the snippet of the researcher's interview (P) with S2 in the DD stage as follows.

P : "What did you do after getting this problem?"
S2 : "What I do after getting the problem is to read and understand first what is meant in the problem, and rewrite the known and questioned tables"
P : "From the problem presented, what information did you find?"
S2 : "The amount of data in the problem table is 53 patients. The lowest weight is 31 kg and the highest weight is 72 kg. For the question is what is the mean, median, and mode of the data"
P : "Are there any steps that you know to solve the problem? If so, what are the steps?"
S2 : "There is sis, first we write back the known data in tabular form, then determine the highest weight and lowest weight, and don't forget to write back what is asked in the question. Then do the calculation process with each formula"
P : "Let me ask you, what do you know from the table in the question about the value range?"
S2 : "The value range in the table is used when it is known that the data is group data. many classes and the range of each class. The value range is known by calculating the difference between the lowest and highest body weight, which is 41, and in the problem you want to take many classes of 7, then the value range of each class can be found from $41/7 = 5.86$ and rounded up to 6. Which makes each class with a range of 6 numbers. For example, at a body weight of $31 - 36$, because there are 4 patients between 31 and 36, the frequency written is 4, and so on until the body weight of $67 - 72$."

At the stage of organizing and reducing data (OD), S2 did first to find the middle value of each class and listed in the column (x_i). Then, S2 calculated the multiplication between the frequency of the *i-th* class and the middle value of the *i-th* class $(f_i \cdot x_i)$, and the results of each class are listed in the column$(f_i \cdot x_i)$. The results of this work can be seen in Figure 4 OD code. In this case S2 did the same OD stage as S1 did, besides that it was also in accordance with the excerpts of the researcher's interview (P) with S2 in the OD stage as follows.

P : "After that, what did you do?"
S2 : "Find the mean value of each *i-th* class (x_i) and calculate the multiplication between the frequency of the *i-th* class and the mean value of the *i-th* class $(f_i \cdot x_i)$"
P : "What kind of calculations are you doing?"
S2 : "Because the data is group data, I continue the calculation according to the mean, median, and mode formulas for group data"

At the stage of representing data (RD), S2 calculates the mean, median, and mode according to each formula for group data. In the process of calculating the mean, S2 calculated the formula $\bar{X} = \frac{\sum f_i \cdot x_i}{\sum f_i}$. By first finding the sum of the calculation of the frequency of the *i-th* class multiplied by the middle value of the *i-th* class ($\sum f_i \cdot x_i$) and the sum of the frequencies of each class ($\sum f_i$). In the process of calculating the mean, S2 found the mean result $\bar{X} = 51.72$. In the median calculation process, S2 calculated with the formula $Me = tb + \left(\frac{\frac{n}{2} - Fk}{f}\right) \cdot p$. S2 first looks for the lower edge of the mode class (tb), the number of frequencies (n), the amount of data before the median class (Fk), the amount of data in the median interval class (f), and the length of each class (p) which will be used in the median calculation process. Then, S2 found the median result $Me = 51.71$.

In the process of calculating the mode, S2 calculates with the formula $Mo = tb + \left(\frac{d_1}{d_1 + d_2}\right) \cdot p$. Before doing the calculation, S2 first looks for the values that will be used, such as the value of the lower edge of the mode class (tb), the frequency value of the mode class to the frequency of the class before the mode (d_1), the frequency value of the mode class to the frequency of the class after the mode (d_2), and the class interval (p). After finding these values, S2 did the calculation, and found the mode result $Mo = 51.83$. In the process of calculating the mean, median, and mode S2 did not make any mistakes, so the results found were correct. The results of the work can be seen in Figure 4 RD code, and this is in accordance with the excerpts of the researcher's interview (P) with S2 in the RD stage as follows.

P : "Then, what did you do after that?"
S2 : "Write down the mean, median, and mode formulas for group data first, then do the calculation process"
P : "In the median calculation process, I would like to ask in the formula you wrote, what are *p* and *tb*?"
S2 : "For *p* is the length of each class, or the number of numbers in each class. For example, in class $31-36$, which means there are 6 body weights $31, 32, 33, 34, 35, 36$. If *Tb* is the lower edge, to find it we are told to find from the existing data, where is the median? Knowing that the number of data is 53, the median is in the 27th data. Judging from the number of frequencies, we can find that the 27th data is in the $49-54$ class. Then the *Tb* used is the lower edge of the class $49-54$ which is 48.5 "
P : "Explain the process of calculating the mean?"
S2 : "The mean formula is the sum of the middle values of the *i-th* class divided by the number of frequencies of the *i-th* class. because the sum of the middle values of the i-th class is 2741.5 and the number of frequencies of the *i-th* class is 53, the mean of the data is $2741.5/53 = 51.72$ "

At the stage of analyzing and interpreting data (AI), S2 has checked the answers first before submitting them. On the other hand, S2 did not write the conclusion of the results obtained because he forgot. This is in line with research conducted by Martadiputra & Suryadi, (2012) which states that students' statistical thinking skills with a focus on analyzing and interpreting data are still not optimal. In addition, this study is also in line with research by Masjudin et al. (2020) which states that 70.8% of class VIII students have low statistical thinking skills, especially in the process of analyzing and interpreting data in the low level category. However, during the interview S2 could express the conclusion of the results that had been obtained. This is in accordance with the excerpt of the researcher's interview (P) with S2 in the AI stage as follows.

P : "Are you sure you've done the calculations correctly?"
S2 : "Are you sure sis?"
P : "Have you checked your answer again? And have you written the conclusion correctly?"
S2 : "I didn't check my answer first because I was in a hurry to submit it, and I didn't write the conclusion because I forgot"
P : "Can you read out the conclusion of your results?"
S2 : "I can, sis. So, from the questions that have been given, the mean value is 51.72 while the mode value is 51.83 and the median value is 51.71"

Based on the analysis of the answers and also the results of the interview, it can be concluded that S2 has understood the concept and can find the solution to the problem given. In general, S2's thinking process is presented in Figure 5.

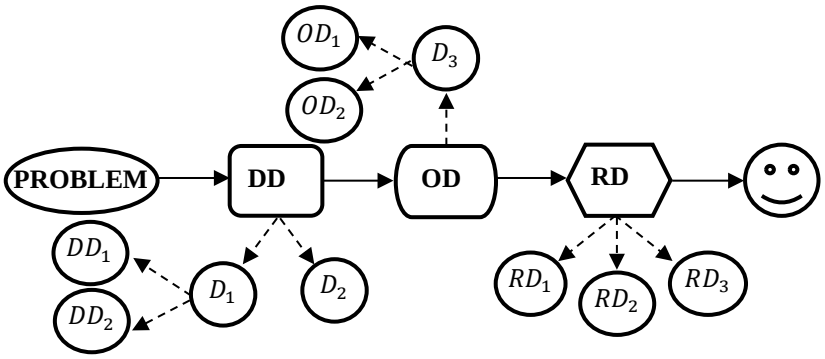


Figure 5. S2 Statistical Thinking Process

For clarity, a description of the coding for S2's thought process when solving problems based on the stages of statistical thinking will be presented in Table 3.

Table 3. Description of Figure 5. S2 Statistical Thinking Process

Code	Description
DD	Describe the data presentation. Characterized by mentioning known data information with a presentation in the form of a table (D_1). The table consists of 2 columns, the first is the weight column (kg) with the range of values for each class predetermined (DD_1), and the second column is the frequency or amount of data for each class (DD_2). In addition, S2 also mentioned what was asked (D_2).
OD	Organizing and reducing data. Marked by answering (D_3), namely determining the middle value of the *i-th* class (OD_1) and calculating the multiplication between the frequency of the *i-th* class and the middle value of the *i-th* class (OD_2).
RD	Representing data. This activity is characterized by the process of calculating the mean (RD_1), median (RD_2), and mode (RD_3) according to the respective formulas on group data.

Construction of references

Based on the research results that have been presented, there are some similarities and differences in the statistical thinking process of the two subjects. Similarities and differences in the stages of students' statistical thinking are as follows:

- Stages of Describing Data Presentation

Relatively the same ability in constructing their understanding of the data given. Furthermore, the subject is able to read the data completely and mention the components of the data completely. This is in line with the opinion of Jones et al. (2000) that students show statistical thinking in terms of understanding and describing diverse data presentations, students are able to read data and show complete awareness in the aspect of describing data presentations.

- Stages of Organizing and Reducing Data

At the stage of organizing and reducing data, both subjects showed many similarities. Both subjects fully explained the grouping and sorting of data, fully explained the general steps of data centralization measures, and fully explained the description of valid measures. In terms of determining the measures of data centralization, both subjects shared the same understanding of mean, median, and mode. In general, both subjects have fulfilled all indicators in the aspect of organizing and reducing data. This is in accordance with the results.

- Stages of Representing Data

At this stage, S1 and S2 are able to implement the plan by performing systematic calculations and steps. This is in line with research by Yuwono et al. (2018), which states that it is very important to think about the steps to be taken in solving problems to improve students' mathematical problem solving skills. In addition, research Wilujeng dalam Darmawan & Ramlah (2021), states that having accuracy and correct calculations is also very important in this stage, because the errors that occur are caused by errors in the calculation process and lack of accuracy of students in solving problems.

- Stages of Analyzing and Interpreting Data

In terms of analyzing and interpreting data, students can draw conclusions and re-check all calculations and solutions that have been obtained. This is in line with Zulfitri (2019) research, which found that the indicator of checking back was the least frequent. This indicator only appeared in S1, while S2 did not perform the checking-bank stage. This was shown in S2's answer, which only provided the final solution to the question. This indicated that the student did not check their answer by substituting it into the original equation and did not draw a conclusion. The student only reached the solution, that the indicator of checking back is the least frequent indicator. This indicator only appeared in S1, while S2 did not perform the checking back stage. This is shown in S2's answer which only writes the final solution of what is asked. This shows that students do not recheck their answers by substituting into the initial equation and also do not make conclusions. Students can only get to the solution. This causes students to find a solution but not the right solution. This is in line with Puspita (in Faturrochmah et al., 2021) said that students make mistakes by not writing conclusions or writing conclusions but wrong.

All student activities will affect the development of students' way of thinking. Therefore, students need to develop statistical thinking skills which are the focus of the development of mathematics learning.

4. Conclusion

The conclusion of the research which aims to describe the statistical thinking process of high school students in solving mathematics problems is that there are differences in the statistical thinking process of each student. In subject 1, in solving mathematical problems on statistics material, the subject carried out the stages of the statistical thinking process completely, namely the stage of describing data presentation; the stage of organizing and reducing data; the stage of representing data, as well as the stage of analyzing and interpreting data. Whereas in subject 2, when solving the problem the subject performed the stages of the statistical thinking process incompletely, namely: the stage of describing the data presentation; the stage of organizing and reducing the data; and the stage of representing the data. From the conclusions presented, there are several suggestions for further research, namely: Further research is needed on students' statistical thinking in solving problems by considering more varied review variables. In addition, based on the findings in this study, students tend to forget the fourth stage in the statistical thinking process, namely data analysis and interpretation. Therefore, teachers or readers may be able to emphasize students to familiarize with this stage.

References

Darma, A. N. (2013). Proses berpikir siswa sma dalam field independent dan field dependent. *Pedagogia*, *2*(1), 71–83.

Darmawan, S. M., & Ramlah. (2021). Analisis Kemampuan Pemecahan Masalah Matematis dalam Menyelesaikan Soal TIMSS Berdasarkan Tahapan Polya. *Jurnal Maju*, *8*(2), 283–292.

Faturrochmah, H., Sary, R. M., & Azizah, M. (2021). Kesalahan Siswa Dalam Mengerjakan Soal Materi Bangun Datar Berdasarkan Teori Nolting Pada Siswa Kelas Iv Sekolah Dasar. *Elementary School: Jurnal Pendidikan Dan Pembelajaran Ke-SD-An*, *8*(2), 310–321. https://doi.org/10.31316/esjurnal.v8i2.1404

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A Framework for Characterizing Children's Statistical Thinking. *Mathematical Thinking and Learning*, *2*(4), 269–307. https://doi.org/10.1207/s15327833mtl0204_3

Martadiputra, B. A. P., & Suryadi, D. (2012). Peningkatan Kemampuan Berpikir Statistis Mahasiswa S1 Melalui Pembelajaran MEAs yang Dimodifikasi. *Pasundan Journal of Mathematics Education : Jurnal Pendidikan Matematika*, *1*(Vol 2 No. 1), 79–89. https://doi.org/10.23969/pjme.v2i1.2466

Masjudin, M., Muzaki, A., Abidin, Z., & Ariyanti, I. A. P. (2020). Analysis of student's statistical thinking ability in understanding the statistical data. *Journal of Physics: Conference Series*, *1521*(3). https://doi.org/10.1088/1742-6596/1521/3/032063

Meletiou-Mavrotheris, M., & Mavrotheris, E. (2007). Online Communities of Practice Enhancing Statistics Instruction: The European Project EarlyStatistics. *Electronic Journal of E-Learning Volume*, *5*(2), 113–122. http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=26281833&site=ehost-live&scope=site

Miles, M. B., & Huberman, A. M. (1994). *An Expanded Sourcebook: Qualitative Data Analysis*. Sage Publications.

Mooney, E. S. (2002). A Framework for Characterizing Middle School Students' Statistical Thinking. *Mathematical Thinking and Learning*, *4*(1), 23–63. https://doi.org/10.1207/s15327833mtl0401_2

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. International Statistical Review. *International Statistical Institute*, *65*(2), 123–165.

Naili, I., Shodiqin, A., & Sudargo, S. (2021). Profil berpikir statistis siswa SMP dalam menyelesaiakan masalah statistika deskriptif ditinjau dari tipe kepribadian. *Prosiding Seminar Nasional …*.

Nani, K. La. (2015). *Kemampuan Penalaran Statistis, Komunikasi Statistis, dan Academic Help-Seeking Mahasiswa melalui Pembelajaran Berbasis Proyek Berbantuan ICT*. Universitas Pendidikan Indonesia.

Obrial, J. T., & Lapinid, M. R. C. (2020). The Use of Statistical Investigation in Assessing Students' Performance in Statistics. *Action Research and Innovation in Science Education*, *3*(2), 47–54. https://doi.org/10.51724/arise.36

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Yuwono, T., Supanggih, M., & Ferdiani, R. D. (2018). Analisis Kemampuan Pemecahan Masalah Matematika dalam Menyelesaikan Soal Cerita Berdasarkan Prosedur Polya. *Jurnal Tadris Matematika*, *1*(2), 137–144. https://doi.org/10.21274/jtm.2018.1.2.137-144

Zahner, D., & Corter, J. E. (2010). The process of probability problem solving: Use of external visual representations. *Mathematical Thinking and Learning*. https://doi.org/10.1080/10986061003654240

Zulfitri, H. (2019). Analisis Kemampuan Pemecahan Masalah Matematika Setelah Pembelajaran dengan Pendekatan MEAs pada Materi Sistem Persamaan Linier Tiga Variabel. *Jurnal Gantang*, *4*(1), 7–13. https://doi.org/10.31629/jg.v4i1.881