Abstract: In the current era of digitalization, statistical skills are needed to interpret and understand and make good decisions about statistical data. However, many students still have the view that statistics is a difficult subject to understand. This is based on the fact that in working on the problem sufficient statistical skills are needed, starting from basic skills to application abilities. Students who have low statistical abilities become less interested in statistics lectures. One of these statistical abilities is statistical reasoning. Statistical reasoning is the student's ability to make logical conclusions from a problem by using statistical concepts in a comprehensive manner, and making decisions as well as evaluating and interpreting the information processed. Efforts to improve statistical reasoning skills require a learning innovation using an application. The role of technology in learning statistics includes using modules assisted by the R program. The research method used is quasi-experimental with a research design, namely pretest-posttest control group design. The results showed that the statistical reasoning abilities of students whose lectures used the R program assisted module were higher than regular lectures for both the low, medium and overall groups. However, the high group had the same increase in statistical reasoning abilities between the experimental and control classes. The effect of statistics lectures using the R program assisted module was high in the medium, low and overall groups. Thus, the use of the R program assisted module is effective in increasing statistical reasoning.

Keywords: R Software, Statistical Reasoning.

I. INTRODUCTION

The role of technology in the learning process is very important. The 4.0 generation technology era forces all components of society, including students, to understand and be able to operate technology because almost all fields already use technology. The inability to utilize technology will result in an inability to deal with change. Therefore, students need to use technology in the learning process. One technology that can be applied in learning is the use of applications. There are various applications that have been made, one of which is the R application.[1]

Statistics is a branch of knowledge that is needed in everyday life and is used in various fields, so humans cannot be separated from the use of statistics itself. Statistics is also a branch of mathematics that studies data collection, processing, analyzing data, and drawing conclusions based on the results of data analysis. [2] Statistics has an important role in conducting research, especially in data processing. Therefore, Statistics needs to be well mastered by students. Through Inferential Statistics lectures it is expected that students can understand the concept of Statistics including describing data, presenting and analyzing data and drawing conclusions. [3] Students are expected to be able to understand the material in the Inferential Statistics course as a whole. This student understanding can encourage an increase in statistical reasoning abilities.

Chervaney [4] defines statistical reasoning as what students can do with statistical content and use their skills in using statistical concepts for solving statistical problems. Statistical reasoning ability is the ability to understand information that exists in everyday life based on statistical data, the ability to answer problems properly based on existing data in different ways and obtain results that are not much different [5].

Statistical reasoning ability is also the ability to understand information that exists in everyday life based on statistical data, the ability to answer problems properly based on existing data in different ways and obtain results
that are not much different [6], [7], [8].

So far, many students have difficulty understanding the concept of statistical data analysis. Conceptual difficulties and misunderstandings commonly experienced by students include the tests used in data analysis, inaccuracies in calculating data analysis and the accuracy and accuracy of solutions which result in drawing conclusions [9]. On the other side, limited access to data sources in the field, either directly or indirectly, does not provide enough motivation for students to study more deeply the issues of data analysis under study. As a result >75% of students obtain unsatisfactory learning outcomes.

The fact is that there are still many students who have low statistical reasoning abilities. This can be seen from the results of the final semester exam scores that students who score less than 3.5 on a scale of 4 are 56.51%. This fact is still far from expectations. While students are expected to be able to improve their ability to use software in processing data, presenting data in graphical form, and analyzing data. Currently, there are many developing software or programs that can assist us in processing data, presenting data in graphical form, and analyzing data. Thus, it is deemed necessary in learning statistics to improve statistical reasoning abilities by using R technology applications.

With regard to R software, it is an open source programming language that is growing rapidly because R users around the world can contribute in the form of code, report bugs, and create documentation, so that R can be used to analyze data according to real-life cases [10], [11].

As for some of the uses of R software, namely: a). Data processing. data processing from simple to complex. Its use is to use a Package or want to build its own syntax from scratch to make it more flexible with various needs. Some examples are Spearman Rho Correlation, ordinal logistic regression analysis, independent t test, independent t test and Mann-Whitney. b). Variety of Statistical Computing. Various statistical computations ranging from random number generation, Bootstrapping, Experimental Design and so on. c). Machine Learning. Unsupervised Learning method, Supervised Learning, or hyperparameter tuning method, d). Text Mining, e). Neural Networks, and f). Web-based Application.

The R application has various advantages including: a). Open Source. R is one of them well-known because it is a powerful OpenSource application, even so well-known in the world of international research. One of the benefits that can be felt is that the R application can be downloaded for free. b). Rich in Package. has various Packages that are ready to assist in completing data analysis as needed. You can even create your own package. c). Capable graphics. d). Machine learning support.

The first thing to do is to download and install it on your personal computer. You can also download the application on the official page for free. Make sure to download the application according to the operating system version of the personal computer. Next, install the application by running the R installer on the personal computer.

To see if the application has been installed properly, try using the R–version command. If no command appears, then there is no problem. However, if the command not found command appears, it is possible that the program was not installed properly and cannot be used.

Next, prepare the data to be processed and analyzed by the program. Make sure the data has been saved in the specified format so that it can be opened in the program. To change the data to be neater and add information attributes, you can change it in excel. Then save the data in the form of an excel workbook.

To start using R-Studio, open the program first. Next, select Packages and the Rcmdr command. A screen will appear with two blank boxes. On the menu, select Data, Import Data, select from Excel then select the data that was saved earlier.
To start using R-Studio, open the program first. Next, select Packages and the Rcmdr command. A screen will appear with two blank boxes. On the menu, select Data, Import Data, select from Excel then select the data that was saved earlier [12].

Based on the description above, the learning used must be able to encourage students to think and be actively involved which is complemented by practicum in the teaching and learning process. Through practicum students can apply the concepts given in a data processing with the help of computer media to solve real problems. However, to overcome students' lack of understanding of the concepts and procedures that should be known, an R-assisted lecture module was developed.

In this module, material and manual solutions are first presented, then problem solving using the R program is presented. Learning with the help of the R program is thought to improve statistical reasoning abilities. Modules and computers as learning media can empower lecturers and students because using modules and computers allows students to study teaching materials in different ways and have time that is not limited only in class.

Based on the problems described above, the formulation of the problem in this study are: (1). Is the increase in the statistical reasoning abilities of students who receive learning using the R program assisted modules higher than conventional learning in terms of: a) overall, b) initial statistical abilities? (2) is there any effect of learning with the R program assisted module on statistical reasoning abilities?

In accordance with the formulation of the problem raised above, the objectives of this study are: (1). To find out whether the increase in statistical reasoning abilities of students who receive learning using the R program assisted module is higher than conventional learning in terms of: a) overall, b) initial statistical ability (2). To find out whether there is an effect of learning with the R-assisted module on statistical reasoning abilities.

II. METHOD

The research method used in this study is a quasi-experimental method. In this quasi-experimental, the subjects were not randomly grouped, but the researchers accepted the subject's condition as it was. The research design carried out in this study was a pretest-posttest control group design [13]. The illustration of the research design is as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eksperimen</td>
<td>$O_1$</td>
<td>$X$</td>
<td>$O_2$</td>
</tr>
<tr>
<td>Control</td>
<td>$O_1$</td>
<td>-</td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

Noted:

$O_1$: Preliminary Aptitude Test

$X$: Learning using the R program assisted module

$O_2$: Final Aptitude Test

The population in this study are all fourth-semester students at the Indonesian Institute of Education for the 2021/2022 academic year. The research sample was taken by purposive sampling as many as two classes. One class is for the experimental class whose lectures use modules assisted by the R program and another class as the Control class whose lectures are conventional.

The statistical reasoning ability test is an instrument used to measure students' statistical reasoning abilities. The test materials were taken from several statistics books, where the questions were validated by experts, the validity that was assessed was face validity and content validity.
The data in this study were analyzed through several steps as follows: (1). Conduct descriptive analysis of statistical reasoning ability data and calculate the pre-test and post-test N-Gain (normalized gain), so that it is known the magnitude of the increase in students' statistical reasoning abilities before and after learning for the experimental and control classes. (2). Testing some of the assumptions needed for hypothesis testing in the assumption test analysis, namely testing for normality and homogeneity of variance. (3). Testing the entire hypothesis using the two-sample independent t-test, the Mann-Whitney U test, and the effect size test.

III. RESULT AND DISCUSSION

Data obtained from students through tests are used to determine the effect of learning carried out on increasing statistical reasoning abilities. The increase in statistical reasoning abilities achieved by students can be seen from the normalized gain data. Normalized Gains (N Gains) were obtained with the Hake formulation. The statistical hypothesis to be tested is:

\[ H_0: \mu_1 = \mu_2 \]

\[ H_1: \mu_1 > \mu_2 \]

The calculation of the difference test of the two mean N-Gains was carried out with the help of the R Program at a significance level of \( \alpha = 0.05 \). The summary is presented in Table 1.

<table>
<thead>
<tr>
<th>Reasoning Statistical Ability</th>
<th>t</th>
<th>W</th>
<th>df</th>
<th>p-Value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>-</td>
<td>1108, 0</td>
<td>65</td>
<td>0.00052</td>
<td>( H_0 ) accepted</td>
</tr>
<tr>
<td>High</td>
<td>-0,53</td>
<td></td>
<td>15</td>
<td>0.75890</td>
<td>( H_0 ) not accepted</td>
</tr>
<tr>
<td>Medium</td>
<td>-</td>
<td>835.5</td>
<td>45</td>
<td>0.00000</td>
<td>( H_0 ) accepted</td>
</tr>
<tr>
<td>Low</td>
<td>-</td>
<td>13</td>
<td>5</td>
<td>0.01889</td>
<td>( H_0 ) accepted</td>
</tr>
</tbody>
</table>

In the high group, students who have high abilities are able to adapt to any way of learning, so they are not too affected by the lectures given. This is in line with Allen [14], that students with high abilities are always suitable for any learning approach. In contrast to the low and medium groups, they are still influenced by the learning approach or method used.

In the low group of students, lectures with modules assisted by the R program can first increase their motivation towards statistics, after their interest increases, they will start to study statistics. After successfully testing the hypothesis with a certain level of significance, the effect size calculation is then carried out. The data for the high group was not calculated because the increase in statistical reasoning abilities in the experimental class was not better than the control class. Summary of calculations can be seen in Table 2.

<table>
<thead>
<tr>
<th>Reasoning Statistical Ability</th>
<th>Class</th>
<th>Mean</th>
<th>Standar Deviasi</th>
<th>d</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Eksperiment</td>
<td>0.66</td>
<td>0.07</td>
<td>5.92</td>
<td>High</td>
</tr>
<tr>
<td>Control</td>
<td>0.08</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Based on table 2, it can be seen that the Cohen effect size (d) in the medium group, low group and overall is at $0.8 \leq d \leq 2.0$ and more than 2.0. Where based on the criteria proposed by Cohen this value is included in the large effect criteria. So it can be concluded that lectures with modules assisted by the R program have a big positive influence in improving students' statistical reasoning abilities both as a whole and for medium and low groups.

Lectures using modules assisted by the R program make students interact directly with the material so that the material will be easier for them to understand. Besides that, lectures with modules assisted by the R program, the learning process can be adjusted according to the learning abilities of the students themselves, meaning that the learning process is controlled by the students. The benefits of learning media put forward that allow direct interaction between students and subject matter and the learning process can take place individually according to student learning abilities [15].

Apart from the R program-assisted module, the R program can also perform complex calculations so that students who lack calculus skills find it helpful. This makes students focus on statistical material, that is, they focus more on choosing a suitable analytical method and how the effect of using the R program assisted module has on statistical reasoning abilities in interpreting results. In addition, the R program menu is presented in Indonesian, so students are not constrained by the language factor. Where the menu in the statistics program is generally presented in English.

In doing data processing with the R program, students must do some reasoning. The reasoning done in the use of the R program is data reasoning, students must know that the data to be processed is numeric data or character (category) data. When the selected data is numerical, a number of statistical tests that can be carried out can be selected, and vice versa.

When choosing a statistical test to be carried out in the R program, it is done by selecting the one to be tested first. For example, what will be tested is the average, then a number of statistical tests will appear that can be used to test the average, such as single sample t-test, independent sample t-test, paired sample t-test, one-way ANOVA and multi-way ANOVA. Associated reasoning is needed at this stage, where they need to know the variables to be tested, how are the relationships between the variables tested, and how many of these variables are so they can determine the right statistical test. Statistical reasoning in using the R program is still needed when a statistical test has been selected and the output of the R program already exists. After selecting the statistical test to be performed, there are a number of options that must be determined. The options are the hypotheses that will be tested in statistical tests, the significance level used and the assumptions needed.

The effect of the specified options can be determined by observing the output of the R program. Students can compare the differences in the output of the R program when the specified options are different. After students carry out simulations with the R program, they are expected to know the assumptions used when they do data processing. Interpreting the output of the R program requires association reasoning, namely making conclusions by taking into account the calculation results. Analysis can be used by comparing p-value with $\alpha$ or comparing calculation results with table data, for example t-count with t-table.

The preparation of modules in this lecture is to overcome the jumps of reasoning that students have. For example, students understand the meaning of the p-value with $\alpha$, but students do not know what conditions are needed when they are going to test two independent means. With modules arranged sequentially starting from the concept, then
the manual problem-solving procedure then solving the problem using the R program, this can overcome jumps of reasoning and does not rule out student knowledge of the concepts and procedures that should be possessed.

CONCLUSION

Based on the results of the research, several things were obtained that could be concluded about the influence of the R program assisted research statistics lecture. The conclusions are as follows. (1). The statistical reasoning abilities of students who received lectures using the R program were better than students who received conventional learning both as a whole and in the medium and low groups. Whereas in the high subgroup there was no difference in the increase in statistical reasoning abilities. (2). There is an effect of learning with the R-assisted module on statistical reasoning abilities.

REFERENCES


