

## Improving Statistical Analysis in Supporting Big Data Era for SMA Negeri 1 Krembung Teachers

Nur Silviah Rahmi<sup>1\*</sup>, Ni Wayan Surya Wardani<sup>2</sup>, Loekito Adi Soehono<sup>3</sup>

<sup>1,2,3</sup>Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Malang, Indonesia

### Abstract

*Classroom Action Research (CAR) is a research type conducted within the classroom environment to enhance the learning process. Nevertheless, many school teachers encounter obstacles primarily due to insufficient knowledge about research methodology and statistical analysis, which support CAR. To maximize data potential in education, SMA Negeri 1 Krembung recognizes the importance of enhancing teachers' statistical and data science-based analysis skills. To achieve this, the Department of Statistics at Universitas Brawijaya collaborates in organizing community service through training to empower the educational community at SMA Negeri 1 Krembung. To assess the community service program's effectiveness, pre-test and post-test analyses were carried out using the t-test. By applying the Wilcoxon Signed Rank Test hypothesis, a significant difference was identified in statistical and data science-based analysis skills before and after the community service program implementation. The ability to analyze based on statistics and data science has improved, evident from the post-test mean rank value surpassing the pre-test mean rank value (11.08 > 6.33). Consequently, it can be concluded that repeating training activities is essential to enhance statistical and data science-based analysis skills in the big data era and to provide a deeper understanding of Classroom Action Research.*

**Keywords:** CAR, Training, Different Test, Paired Sample T-test, Wilcoxon Signed Rank Test

### INTRODUCTION

In the ever-evolving digital era, information technology has changed the way things are done interact, work, and learn. Data is one of the most valuable assets of this era. Wali et al., (2023) affirmed that the digital era known as the "big data era" has a significant impact on various sectors, including education. The use of technology in the learning process has produced abundant data, ranging from student development to interactions in the classroom. However, large data sets will not provide benefit if they are not analyzed appropriately. The use of appropriate data can help improve the quality of learning and provide more effective education. Therefore, improving statistical and data science-based analysis skills in teachers is a critical step in supporting the era of big data in the world of education.

Statistics is a procedure for collecting, presenting, analyzing, and interpreting data (Walpole, 1993). In general, statistics are applied to numerical data in the form of measurement and enumeration results, or in the form of categorical data grouped based on certain criteria (Sutedja et al., 2020). Statistics has

an inseparable role in the implementation of research. Due to its important relevance, this course has become an essential component in the compulsory curriculum for students from various disciplines and different levels of education (Yanto, 2020).

In the research process, from collecting, processing, presenting, to interpreting data, knowledge in statistics is required. This includes research that related to actions in the classroom, which are generally carried out by teachers to improve student learning outcomes. According to Sekaran & Bougie (2013), research is a structured and organized investigation, based on data, critical, objective, and scientific, which aims to investigate certain problems and find relevant solutions or answers.

Classroom Action Research (CAR) is a type of research conducted in the classroom environment, using concrete actions to improve the quality of the learning process in order to achieve better results than before. The conduct of classroom action research requires a relatively long time because it involves the implementation of actions and

variables designed to achieve the expected goals. Classroom Action Research is realized in the form of implementing special actions in the classroom with the aim of improving the teaching-learning process in order to produce better learning progress than before (Arikunto, 2012). Classroom action research can be used as an effort to implement various programs in the school environment, by observing indicators of success of learning processes and outcomes for students, and observing the achievement of goals from various programs implemented by the school. In essence, classroom action research is carried out to improve the overall quality of learning practices in real-world situations.

Implementation of Classroom Action Research can take place effectively if the planning and implementation applies six fundamental principles (Quintania et al., 2023). The first and the most important principle is that the main task of a teacher in a school is to teach students. The use of data collection methods should not burden teachers with excessive time so as not to interfere with the course of the teaching and learning process. The methodological approach used must have an adequate level of reliability, thus enabling teachers to identify and formulate hypotheses with sufficient confidence. It is recommended to develop strategies that are relevant to each class situation and obtain data that can support the explanation of the hypotheses proposed. Furthermore, the research problem raised by the teacher must be something that really affects and unsettles him. The consistency of teachers in holding ethical values related to their profession is absolute. Understanding of problems is not only limited to the context of a particular class or subject, but also seen from a broader perspective. Teachers have a strategic role in the development of professional competencies, so teachers are expected to have a deep understanding not only of students, but also of classes, curriculum, learning mechanisms, and learning outcomes as an effort to improve student learning processes and outcomes. Teacher competency can be obtained through professional education and understanding computer-assisted instruction (Fitria et al., 2019). Classroom Action Research is the teacher's obligation in understanding all these things. However, many teachers in schools face challenges in this regard, especially due to the lack of knowledge of research methodology and statistical analysis which are essential supports in all stages of

the Classroom action research process, and a lack of mastery of technological tools.

In community development, training is provided as an effort to realize one of the missions of the Department of Statistics, Brawijaya University, namely socializing statistics through service and collaboration with the community, government and private institutions. Through intensive training, will definitely be able to produce something that has important benefits and values for self and others. The quality of training must also be emphasized because without education or training that has good weight and quality in society, efforts to increase overall productivity will be difficult to fully realize (Quintania et al., 2023). The quality of skills plays a big role in preparing skilled human resources. The quality of these skills can be obtained through creating an interesting learning environment, so that it can provide motivation and enthusiasm, as well as encourage us to make the best use of our abilities.

In the process of implementing CAR data will be produced that need to be analyzed, interpreted and concluded to improve learning so that statistics are needed to process this data. To obtain informative conclusions, data is not only analyzed using descriptive statistics, but also using inferential statistics (Afandi et al., 2022). . To optimize the potential of data in education, SMA Negeri 1 Krembung realizes the importance of improving statistics-based analysis skills and data science for teachers. To realize this goal, the Statistics Department of Brawijaya University participated in organizing training aimed at empowering the educational community at SMA Negeri 1 Krembung. This activity is a component of community service efforts to provide a concise understanding of the basics of statistics in the context of Classroom Action Research (CAR). According to Pramswari (2018), CAR will be difficult to carry out if teachers have difficulty inputting data, analyzing, and presenting data using IT-based devices, especially software. In addition, this activity also aims to introduce the use of statistical application programs (SPSS) that can assist in processing and presenting research data more effectively and accurately.

Based on the explanation above, the objectives of this community service program include: a) improving the quality of learning so that teachers will be able to identify individual needs of students and develop more effective learning methods based on data analysis, b) better assessment and

evaluation so that with better data analysis capabilities, educational institutions can assess and evaluate student performance more objectively and accurately, c) development of a curriculum that is more relevant to the times and student needs, and d) evidence-based decision making, so that teachers will be more confident in making decisions.

## MATERIAL AND METHOD

The method of implementing community service used in this program is an interactive, participatory, and practical method. The participants of the statistics and data science training in the community service program were all teachers of SMA Negeri 1 Krembung with a total of 59 teachers. Some of the methods that will be used include: a) Material Presentation: Participants will receive material presentation from statistics and data science experts from the Department of Statistics, Brawijaya University. The material will be delivered in easy-to-understand language and illustrated with relevant examples from the world of education, b) Case Studies: Participants will be given real case studies related to the world of education. Participants will be asked to analyze the data and find the right solution based on the results of the analysis. Before giving the material, a pre-test was carried out with at least 5 multiple-choice questions containing three discussion materials. After the pre-test is carried out, the material is given with a duration of 120 minutes x 2 so that the total learning is 240 minutes or 4 hours. To evaluate the success level of participants' absorption during the training, a post-test is carried out using questions as in pre-test questions.

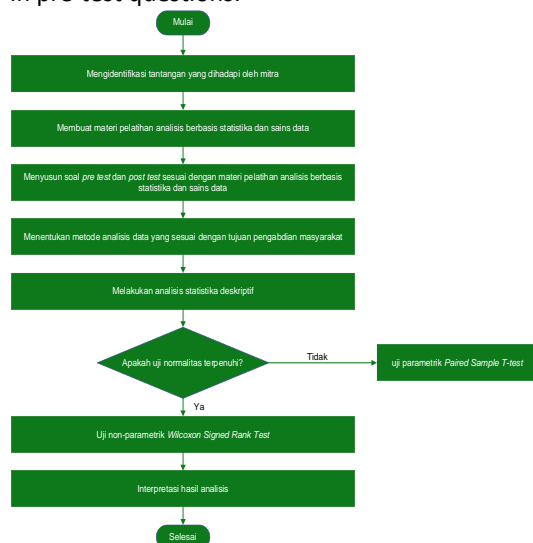


Figure 1. Research Flow Diagram

The community service program begins by identifying the challenges faced by community service partners. At this initial stage, there is a lack of understanding of research methodology and statistical analysis, which have an important role in all stages of classroom action research. In addition, there are still a number of teachers in schools who are not familiar yet with statistical software which can basically facilitate the process of processing statistical data generated from research activities. Therefore, the community service team wants to provide training on: a) the concept of big data including characteristics, potentials and challenges in processing large and complex data; b) knowledge of statistical analysis techniques such as data centering measurement and the basics of statistical inference; c) an understanding of data science concepts including data processing and data exploration; d) the introduction of data analysis software, in this case using SPSS software so that SMA Negeri 1 Krembung teachers can process and analyze data more efficiently; e) apply data analysis to educational contexts including student performance evaluation; f) improvement of learning methods, and curriculum development; g) understanding and data analysis skills that are useful in the decision-making process. Santoso (2010) stated that training is a learning process that emphasizes practice rather than theory carried out by a person or group using a various learning approach and aims to improve abilities in one or several types of certain skills.

After carrying out community service, the next step is to carry out statistical analysis to evaluate the effectiveness of the program. The efficiency of community service activities was assessed using a difference test, which aims to identify whether there are differences in statistical and data science-based analysis skills of teachers at SMA Negeri 1 Krembung before and after participating in the community service program. The data used are the pre-test and post-test scores of community service participants, namely teachers of SMA Negeri 1 Krembung. The condition that must be met by the difference test is that the group of data to be analyzed must have a normal distribution. The data normality test in this study was carried out using a statistical test, shapiro-wilk. The shapiro-wilk test is an effective and valid normality test method for small samples. According to Mohd Razali & Bee Wah (2011) the shapiro-wilk test is limited to a sample size of less than 50. The normality test hypothesis is as follows:

H0: Data is normally distributed

H1: Data is not normally distributed

with  $H_0$  accepted criteria if  $p$  value  $> \alpha$ , so it can be said that the data is normally distributed. Hypotheses related to difference tests can be grouped into two categories, namely parametric statistical tests and non-parametric statistical tests. Parametric statistical tests are a form of hypothesis testing that tests mean differences in populations, while non-parametric statistical tests do not depend on certain parameter assumptions regarding the population being tested. If in testing the data is normally distributed, then use the parametric Paired Sample T-test. However, if the data distribution does not meet the normal pattern, then the non-parametric Wilcoxon Signed Rank test is used to test the hypothesis. The hypotheses tested in this context are:

$H_0$ : There is no difference in statistical and data science-based analysis skills before and after statistical and data science-based analysis training for SMA Negeri 1 Krembung teachers.

$H_1$ : There are differences in statistical and data science-based analysis skills before and after statistical and data science-based analysis training for SMA Negeri 1 Krembung teachers.

## RESULT AND DISCUSSION

The process of implementing community service begins with applying for a permit and carrying out initial observations which are carried out within one week. This activity plan was received positively by the school, because it was considered an innovative step in providing training on statistics-based analysis and data science that had never been done before. The principal of SMA Negeri 1 Krembung assessed that training with a focus on this topic has high urgency. This is due to the importance of teachers' understanding of how to analyze data in an educational context, especially in Classroom Action Research (CAR) in the era of big data. This era is characterized by the amount of data generated on a large scale, which exceeds the capacity of conventional methods in processing. In addition, the training also aims to introduce the use of data analysis software, so that teachers at SMA Negeri 1 Krembung can manage and analyze data more efficiently.

The implementation of the community service program began with an opening ceremony led by Mr. Drs. Suswanto, M.M., as Head of SMAN 1 Krembung. In his speech, he emphasized the important role of teacher in carrying out Classroom Action Research (CAR). CAR activities are attractive to all teachers in various subjects because they can measure students' ability to understand the

material and the effectiveness of the learning methods used. After a speech by the Head of SMAN 1 Krembung, a pre-test was given to the participants. This pre-test aims to assess the extent of the abilities of SMA Negeri 1 Krembung teachers in statistics and data science-based analysis.



Figure 2. Opening by the Principal Of SMA Negeri 1 Krembung

After the pre-test activity, community service was entered the core of the activity, namely training on analytical skills based on statistics and data science guided by Nur Silviyah Rahmi, S.Si., M.Stat and consisting of Prof. Ni Wayan Surya Wardhani, MS. and Prof. Loekito Adi Soehono, M.Agr. The training session consisted of a presentation on the definition of Classroom Action Research (CAR), a Difference Test regarding paired t-test, a correlation test on learning methods vs learning styles and regression analysis on the influence of learning methods on grades. In between material presentation activities, Nur Silviyah Rahmi, S.Si., M.Stat was carried out direct practice of using data analysis software, namely SPSS software. The enthusiasm of SMA Negeri 1 Krembung teachers was quite large when training on statistics and data science-based analysis was carried out. This is reflected in the large number of questions asked by the teachers during the activity.



Figure 3. Delivery of data analysis training material based on statistics and data science

The total number of participants in community service activities at SMA Negeri 1 Krembung was 59 teachers. Of the 59 teachers, there were 34 teachers who could not be used as respondents to measure statistical and data science-based analysis skills in SMA Negeri 1 Krembung teachers. This is

because among the 34 respondents there were respondents who did not fill in personal data, did not take the pre test or post test so that the respondents' answers did not qualify as 2 paired samples in this study. The results of descriptive statistics are presented in Table 1.

**Table 1.** Descriptive statistics results

Descriptive Statistics						
	N	Min	Max	Mean	Std. Dev	Variance
Pre Test	25	0.00	100	62.286	21.381	457.143
Post Test	25	42.86	100	72.571	12.989	168.707
Valid N	25					

Based on Table 1, descriptive statistics can be seen regarding the value response of respondents when answering pre-test and post-test questions. Pre-test is carried out before the training activity takes place. The respondents provided answers to 7 questions designed to measure skills in statistical and data science-based analysis. Through Table 1, it can be observed that the average pre-test score of 25 respondents was 62.286. This illustrates that the maximum number of questions answered correctly by respondents was 4 questions. The pre-test results also showed that the lowest score obtained by respondents was 0, while the highest score reached 100. This fact reveals that there are participants who have a very limited understanding of statistics and data science-based analysis. Based on the analysis of each question, it appears that 60% of respondents gave wrong answers when asked to identify the statistical analysis that used when trying to assess the difference between two populations.

Meanwhile, the post test is carried out after the completion of the training stage. The post test was carried out by presenting the same 7 questions to the same respondents, namely the teachers of SMA Negeri 1 Krembung. The purpose of this post test is to measure the ability of respondents after participating in training that includes CAR, difference test, correlation test, and regression

analysis. Based on the results of descriptive statistics, it can be seen that the average value of the post test is 72.57. This shows that most respondents managed to answer 5 questions correctly. The results of the post test also showed that the lowest score achieved by respondents was 42.86, while the highest score reached 100. These results indicate that some participants experienced a significant increase in understanding, and overall post-test scores were higher than pre-test scores. Analysis of each component of the question showed that 68% of respondents answered incorrectly when asked questions about the scale of the data.

Normality Test is a test carried out with the aim of assessing the data distribution pattern in a certain set of data or variables, with the aim of identifying whether the data distribution pattern follows a normal distribution or not (Tiro, 1999). The normality test in this study was conducted using the Shapiro Wilk test. The normality testing hypothesis is as follows:

H0: Data is normally distributed

H1: Data is not normally distributed

With H0 accepted criteria if p value >  $\alpha$ , so it can be said that the data is normally distributed. The results of normality testing using the Shapiro Wilk test are presented in Table 2.

**Table 2.** Normality Test Result

	Test of Normality					
	Kolmogorof-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistics	df	Sig	Statistics	df	Sig
Pre Test	0.245	25	0.000	0.900	25	0.018
Post Test	0.225	25	0.002	0.904	25	0.022

a. Lilliefors Significance Correction

In this study, the degree of significance or alpha value used was 0.05. Based on Table 2, it can be seen that the p-value of the pre-test variable is  $0.018 < \alpha$  and the p-value of the post-test variable is  $0.022 < \alpha$ , then both variables are rejected  $H_0$ . This can be interpreted that the two groups of data are not normally distributed.

The hypothesis test used in this study is a difference test. The different tests are used to assess the effect of a particular treatment on the same sample in two different observation periods, namely before and after the treatment is carried out (Esomar, 2021). The observations made in this research were training activities to improve statistical and data science-based analytical skills for SMA Negeri 1 Krembung teachers. The statistical hypotheses used are as follows:

$H_0: \mu_1 = \mu_2$  (There is no difference in statistical and data science-based analysis skills before and after statistics and data science-based analysis training for SMA Negeri 1 Krembung teachers)

$H_1: \mu_1 \neq \mu_2$  (There are differences in statistical and data science-based analysis skills before and after statistics and data science-based analysis training for SMA Negeri 1 Krembung teachers)

According to research conducted by Putri et al., (2023), there are a number of assumptions that need to be considered in the application of the Wilcoxon Signed Rank Test. Among these assumptions are (a) The variables used have ordinal or interval/ratio data scales, but the data distribution does not follow a normal pattern; (b) The independent variable consists of two paired categories. This means that the data observation units are similar; (c) The shape and distribution of data between the two paired groups is symmetrical, this is necessary to qualify for the normal distribution in the difference test. The first and second assumptions of the Wilcoxon Signed Rank Test have been fulfilled, because the variables used are ratio data scaled, not normally distributed, and the independent variables come from two paired categories. The results of the non-parametric Wilcoxon Signed Rank Test are presented in Table 3.

**Table 3. Non-Parametric Rank Hypothesis Test Result**

		Ranks		
		N	Mean Rank	Sum of Rank
PostTest-	Negative Rank	6 <sup>a</sup>	6.33	38.00
PreTest	Positive Rank	12 <sup>b</sup>	11.08	133.00
	Ties	7 <sup>c</sup>		
	Total	25		

a. PostTest < PreTest  
b. PostTest > PreTest

Based on Table 3, it can be seen that the mean value in negative ranks is 6.33 while the mean value in positive ranks is 11.08. This can be interpreted that the sample in the post-test group value is different from the pre-test group value. Based on these results, it can be concluded that statistics and data science-based analytical skills in the sample group after training have increased when compared to the sample group before training. In addition, the ties value is zero, which means that there is no similarity between the post-test value and the pre-test group value.

**Table 4. Wilcoxon Signed Rank Test Result**

Test Statistics <sup>a</sup>		PostTest-PreTest
Z		-2.081 <sup>b</sup>
Asymp Sig (2-tailed)		0.37
a. Wilcoxon Signed Rank Test		
b. Based on negative ranks		

Based on the results of the Wilcoxon Signed Rank Test calculation, p-value of 0.037 is smaller than alpha ( $\alpha$ ) 0.05, so  $H_0$  is rejected. It can be concluded that there are significant differences between the pre-test and post-test groups. The difference lies in the higher level of statistical and data science-based analytical skills in the post-test group than the pre-test group. Thus, there was an increase in ability after training on statistics and data science-based analysis for teachers of SMA Negeri 1 Krembung.

The results of the different tests indicate that training on statistics and data science-based analysis for teachers of SMA Negeri 1 Krembung must continuous to be carried out because it is proven to significantly increase the knowledge and abilities of community service participants. In addition, it is hoped that the holding of this can improve decision-making skills that are more precise and evidence-based in facing educational challenges.

From the results of the message and impressions of the activity, it shows that the teachers of SMA Negeri 1 Krembung as participants in this training activity said that they really need further and in-depth training in the use of statistics and data science in Classroom Action Research (CAR). In addition, participants hope that training activities based on statistics and data science can be held regularly. So it can be concluded that this training activity needs to be carried out again in order to improve.

## IMPACT OF ACTIVITIES

The results of community service activities show that participants hope that statistical and data science-based analysis training activities can be held regularly, this is done with the aim of improving capabilities in statistics and data science-based analysis in the era of big data, and providing participant with a deep understanding of classroom action research.

## CONCLUSION

Based on descriptive statistics, the average pre-test score is 62.28, which means that 62.28% of participants understand Class Action Research (CAR), difference tests, correlation tests, and regression analysis. After the training, the participants' statistical and data science-based analysis skills increased to 72.57%. Apart from that, the results of the pre test and post test analysis are supported by the results of different tests which show the mean rank value of the post test > pre test (11.08 > 6.33). This value shows that there is a significant difference in statistical and data science-based analysis skills in the group after training, which has increased when compared to the sample group before training. In addition, based on the results of messages and impressions of community service activities, it can be concluded that this training activity needs to be carried out again with the aim of improving statistical and data science-based analysis skills in the big data era, as well as providing a deeper understanding of Class Action Research (ACR) to the participants, namely SMA Negeri 1 Krembung teachers.

## ACKNOWLEDGEMENT

The author would like to thank to Badan Penelitian dan Pengabdian Kepada Masyarakat (BPPM) Faculty of Mathematics and Natural Sciences, Brawijaya University which has funded the team of lecturers from the Department of Statistics, in carrying out community service at SMA Negeri 1 Krembung so that this service activity can be carried out. The team also expressed gratitude to the Principal of SMA Negeri 1 Krembung, teachers and staff who have supported and accepted the team of lecturers from the Department of Statistics in implementing ideas and work programs.

## REFERENCES

1. Afandi, N., Fransiska, H., Yosmar, S., Rini, D. S., Agwil, W., & Swita, B. (2022). *Pelatihan SPSS untuk Analisis Data Penelitian Tindakan Kelas*. *Jurnal Berdaya Mandiri*, 4(3), 254-263.
2. Arikunto, S. (2012). *Penelitian Tindakan Kelas*.
3. Esomar, M. J. (2021). *Analisa Dampak Covid-19 terhadap Kinerja Keuangan Perusahaan Pembiayaan di Indonesia*. *Jurnal Bisnis, Manajemen, dan Ekonomi*, 2(2), 22-29.
4. Fitria H, Kristiawan, M., & Rahmat, N. (2019). *Upaya Meningkatkan Kompetensi Guru Melalui Penelitian Tindakan Kelas*. *Abdimas Unwahas*, 4(1), 14-25.
5. Pramswari, L. P. (2018). *Penguasaan IT dan Kemampuan Meneliti Guru: Dua Hal Berkorelasi yang Diabaikan*. *Prosiding Seminar Nasional "Membangun Generasi Emas 2045 yang Berkarakter dan Melek IT" dan Pelatihan "Berpikir Suprarasional"* (p. 90). UPI Sumerdang Press.
6. Putri, V. Y., Halim, R. M., Fahira, T. N., & Janice, S. N. (2023). *Keuangan generasi z: Tingkat literasi dalam rangka peningkatan inklusi keuangan digital*. *Jurnal Inovasi Hasil Pengabdian Masyarakat (JIPEMAS)*, 6(2), 188-200. <https://doi.org/10.33474/jipemas.v6i2.19068>.
7. Quintania, M., Handayani, E. N., Anwar, R. N., Suprpto, A., Ekonomi, F., & Darma Persada, U. (2023). *Pelatiha Statistik Guna Meningkatkan Kemampuan Dalam indakan Penelitian Kelas Pada Guru-Guru SDS Keenkids*. *WASANA NYARA*, 7(1). <http://e-journal.stie-aub.ac.id>.
8. Razali, N. M., & Wah, Y. B. (2011). *Power comparisons of Shapiro-Wilk, Kolmogorof-Smirnov, Liliefors and Anderson-Darling test*. *Journal of Statistical Modeling and analytics*, 2(1).
9. Santoso, B. (2010). *Skema dan mekanisme pelatihan: panduan penyelenggaraan pelatihan*. Yayasan Terumbu Karang Indonesia.
10. Sekaran, U., & Bougie, R. (2013). *Reserch methods for business: a skill-bulikding approach*. Wiley.

11. Sutedja, A., Hanafiah, & Ahmaddien, I. (2020). *Pengantar Statistika. era Industri 4.0 dan Society 5.0). PT Sonpedia Publishing Indonesia.*
12. Tiro, M. A. (1999). *Dasar-Dasar Statistika. Makassar: UNM Makassar.*
13. Wali, M., Efitra, Sudipta, I. I., Heryani, A., Hendiyani, C., Rahman, R., . . . Sepriano. (2023). *Penerapan & Implementasi Big Data di Berbagai Sektor (Pembangunan Berkelanjutan*
14. Walpole, R. E. (1993). *Pengantar Statistika (3rd ed.). PT Gramedia Pustaka.*
15. Yanto. (2020). *Survei Penggunaan Uji Statistik dalam Penelitian Tugas akhir Mahasiswa dengan Topik Ergonomi di Prodi Teknik Industri Unika Atma Jaya Jakarta. <http://ojs.atmajaya.ac.id/index.php/metri>.*