

Developing the UNPAD SAS (Universitas Padjajaran Statistical Analysis Series) Software

Ratna Jatnika

Universitas Padjadjaran Bandung
ratna@unpad.ac.id

Mustofa Haffas

Universitas Padjadjaran Bandung
haffas@unpad.ac.id

Hendriati Agustiani

Universitas Padjadjaran Bandung
hendriati.agustiani@unpad.ac.id

Abstract— Student in Faculty of Psychology think that Statistics is very difficult for them, because Statistics is viewed as a hard science than Psychology which is viewed as a soft science. Various attempts have been made to improve student attitudes toward statistics, so that student have more positive attitudes. One of the efforts is to transform the curriculum of Statistics in Faculty of Psychology Universitas Padjadjaran by adding SPSS (Statistical Packages for Social Sciences) practicum courses since 2009.

There are a variety of data analysis contained in SPSS can be used for data processing. However, there are still some statistical data analysis used by students of the Faculty of Psychology that is not available in SPSS. The aimed of this research is to develop software namely Universitas Padjadjaran Statistical Analysis Series, which is statistical data analysis software that consist analysis that does not exist in SPSS or other data analysis software. In this preliminary research, modules are developed only for Database Management and Descriptive Statistics.

The software development will be carried out by (SDLC = Software Development Life Cycle). SDLC is a series of step or phase that presents a model for development and

lifecycle management software or applications.

The resulting software is tested on 144 students in Psychology Faculty in Universitas Padjadjaran. The trial results showed that the software is most appropriate and "user friendly" software.

Keywords— Software, Statistical Data Analysis

I. INTRODUCTION

Student in Faculty of Psychology and other social sciences think that Statistics is very difficult for them, because Statistics is viewed as a hard science than Psychology which is viewed as a soft science. Therefore, teachers of Statistics courses often have to work hard to improve their instructional methods.

Various attempts have been made to improve student attitudes toward Statistics, so that student have more positive attitudes. One of the efforts is to transform the curriculum of Statistics in Faculty of Psychology Universitas Padjadjaran by adding SPSS (Statistical Packages for Social Sciences) practicum courses since 2009. The addition of this course is done by hoping that students can more easily calculate various statistical analysis so that Statistics will be perceived as easy to understand. In other words, the addition of this course is expected to improve

students' positive attitude towards Statistics. The results of research conducted by Jatnika&Abidin (2013) and Jatnika (2015) indicate that there is a significant positive attitude increasing in the cognitive aspects after learning Statistics by using practice method with SPSS compared with before using practice method. Jatnika&Abidin (2015) continued their research and found out that positive feelings about Statistics need to be developed in order for students to have low anxiety, and low fears and concerns about Statistics. Students should also feel that Statistics is easy, so that they are not being afraid and worried about Statistics, also have a more positive attitude towards Statistics. Therefore, students should grow their interest in Statistics in order for not having anxiety, fear and worry on Statistics, also to have a positive attitude towards Statistics.

The results of this study support the results of the previous research, and as a justification that the addition of practical subjects SPSS is a way that can be done to grow student interest on Statistics so that students feel that Statistics is easy, and consequently the students do not feel anxious, fearful and worried about Statistics.

SPSS (Statistical Package for the Social Sciences) is a software package used for statistical analysis. The current SPSS version (2014) is officially named IBM SPSS Statistics. This product can be used for survey authoring and deployment (IBM SPSS Data Collection), data mining (IBM SPSS Modeler), text analytics, and collaboration and deployment (batch and automated scoring services). The software name stands for Statistical Package for the Social Sciences (SPSS), reflecting the original market, although the software is now popular in other fields as well, including the health sciences and marketing. SPSS is a widely used program for statistical analysis in social science. It is also used by market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners, and others (KDnuggets, 2013). The original SPSS manual has been des-

cribed as one of "sociology's most influential books" for allowing ordinary researchers to do their own statistical analysis. In addition to statistical analysis, data management (case selection, file reshaping, creating derived data) and data documentation (a metadata dictionary is stored in the data file) are features of the base software (Wellman, B, 1998).

Various statistical data analysis is contained in the SPSS menu that can be used by students and researchers of Psychology and other social sciences to process their research data. However, there are still many statistical data analyzes used by students and researchers which are not available in SPSS, such as Theta, Eta, and Jaspén correlation, as well as further tests for Kruskal Wallis and Q Cochran methods as well as minimum sample size calculations, etc. The importance of these methods in Psychological and Social research, attracted researchers to develop software to calculate the statistical data analysis.

SPSS software for single user price ranges from 1,170 USD to 7,820 USD or equivalent about 16 million rupiah up to 100 million rupiah. The high price of SPSS is causing pirated piracy against existing software. Therefore, this condition is increasingly encouraging the interest of researchers to develop statistical data analysis software which is cheap, precise, and comprehensive for researchers in the field of Psychology and other social sciences.

Jatnika&Haffas (2014) has started to develop UNPAD SAS (Universitas Padjajaran Statistical Analysis Series) which is software to analyze statistic data not found in SPSS. The UNPAD SAS that has been developed contains two analyzes, namely: the calculation of the minimum sample size and the pair of comparison from Thurstone. The calculation of the minimum sample size performed recently incorporates two methods of calculating the minimum sample size. The making of UNPAD SAS is done by using C++ programming language and method (SDLC = Software Development Life Cycle). SDLC is a step or phase sequence that presents a model for development management

and the life cycle of an application or software. This method consists of stages: Planning, Analysis, Design, Implementation, Testing, Maintenance. In 2015, Jatnika&Haffas continued to manufacture UNPAD SAS by adding Eta, Theta, and Jaspren correlation calculations which are widely used in Psychology and Social field but not yet in SPSS. In 2016, Jatnika&Haffas continued to manufacture UNPAD SAS for caculating the minimum sample size, resulting in three methods for calculating the minimal sample size, which are mean estimates, proportion estimates and correlation/regression estimates.

Initially, UNPAD SAS will be developed in several series of analysis, in which each series consists of several stand-alone modules. Patterns developed like this will allow each module to run separately, but the developed series will not be integrated. As a result, the use of UNPAD SAS is inefficient for analyzing diverse data. For that purpose UNPAD SAS needs to be developed in order to run through one door or one system, where each series of analysis will be viewed as incremental analysis modules. Therefore it is necessary to create a core system to unify the series of analysis.

In addition to making the core system, UNPAD SAS also needs to be developed by adding database management that has the ability to import export files, edit files and data management. UNPAD SAS also needs to be added with various series of statistical analyzes that many other Psychology and Social Science researchers use, but not included yet in SPSS. The series of analysis is quite a lot and will take a long time to develop it. With the core system, UNPAD SAS is also open for development by other parties, enabling the development of series of analysis to take place more quickly. At the end of this study will be obtained statistical data analysis software named UNPAD SAS, which already has a core system with complete database management and has been equipped with various series of statistical analysis that is: Descriptive Statistics, Minimum Sample Size

Calculation, Correlation Analysis Eta, Theta, Jaspren) and Non Parametric Analysis (including further tests of Kruskal Wallis and Q Cochran). Therefore the development plan of UNPAD SAS is:

1. The first year:
 - a. Modules Management and Database Management
 - b. Descriptive Statistic
2. The second year:
 - a. Minimum Sample Size
 - b. Correlation Analysis
 - c. Non Parametric Statistics

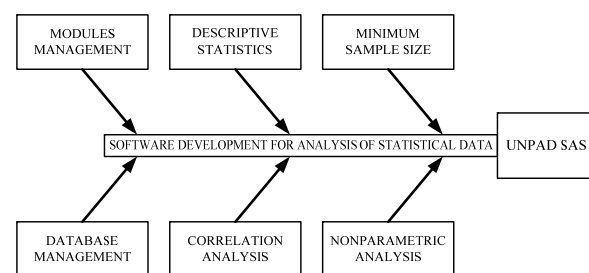


Figure 1

Fishbone Diagram for the Research Flow

Database Management which is the core system of UNPAD SAS, consists of:

1. Management of table structure:
 - Create, edit, insert, and delete variables
2. Data management table:
 - Addition, editing and deletion of table data
3. Standard data manipulation:
 - a. Sorting
 - b. Select
 - c. Replace
4. Import and export data:
 - a. Copy paste
 - b. Writing and reading from and to cvs files (excel)
5. ModulesManagement

Descriptive Statistics consists of:

1. Frequency Module:
 - a. Measured of Central Tendency
 - b. Measured of Dispersion
 - c. Measured of Distribution
 - d. Graph (Bar, Pie)
2. Descriptive Module:
 - a. Measured of Central Tendency
 - b. Measured of Dispersion
 - c. Measured of Distribution

- d. Standardized Z
- 3. ExploreModule:
 - a. Boxplot

Based on previous explanation, the purpose of this research is to develop the most appropriate and "user friendly" UNPAD SAS (Database Management and Statistics Descriptive) to analyze statistical data effectively and efficiently. The resulting software will be applied to the teaching of Statistics courses, making it easier for students to understand Statistics and increase interest and reduce students' anxieties and concerns in studying Statistics.

II. LITERATUR REVIEW

Jatnika & Abidin (2011), has conducted a study aimed at researching the psychometric properties of the Survey of Attitudes Toward Statistics (SATS) developed by Schau (2003) which has been translated into Indonesian. This research was conducted to adapt SATS in 81 students of Faculty of Psychology UNPAD. The results of this study indicate that SATS has Goodness of Fit Index = 0.89, Root Mean Square Error of Approximation (RMSEA) = 0.017 and Chi-Square = 9.26 (P = 0.32). The Alpha Cronbach coefficients for the SATS aspects are Affect (0.844), Cognitive (0.766), Value (0.716), Difficulty (0.400), Interest (0.828), Effort (0.783).

Jatnika & Abidin (2012) study on 81 students of the Faculty of Psychology UNPAD in 3rd semester (have completed the subjects of Statistics 1, Statistics 2 and Statistics 3) also shows that basically the students understand the importance of controlling Statistics for the future in their profession later (value). The students have also made strenuous efforts to study Statistics (effort). However, the students judge that Statistics is a difficult course (difficulty)..

In 2012, Jatnika & Abidin continued their research on 67 students of the Faculty of Psychology UNPAD who have completed the courses of Statistics 1 and Statistics 2 using the Wilcoxon test analysis. The results showed that there was a significant positive attitude increase in the cognitive aspects after

statistical learning using SPSS practice method compared with before using SPSS practice method. This means that after studying Statistics using SPSS practice method, students feel that knowledge and skills in using the science of statistics to be increased.

Based on the explanation above, in 2014, Jatnika&Abidin continue their research to identify the factors that influence the mastery of Statistics materials on Unpad Faculty of Psychology students. The results of this study provide the findings as shown in the following figure:

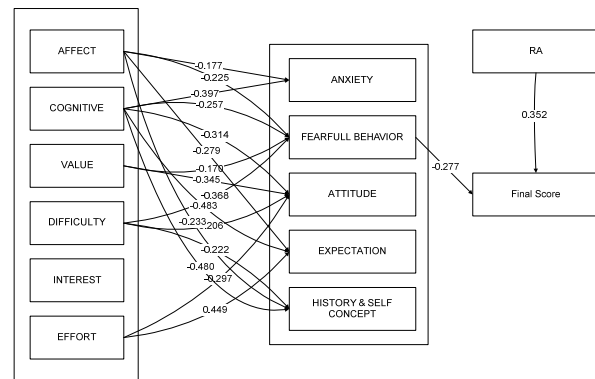


Figure2
Identification of Factors Influencing Mastery of Statistic Material to Students of Unpad Faculty of Psychology

Based on the results of this study, it can be concluded that positive feelings towards Statistics should be developed so that students have a low anxiety, as well as low fears and concern to Statistics. Students should also feel that Statistics is easy, so that students are not afraid and worried about Statistics, and have a more positive attitude towards Statistics. Therefore, students should grow their interest in Statistics in order for not having anxiety, fear and worry on Statistics, also to have a positive attitude on Statistics.

The results of this study support the results of previous research, and as a justification that the addition of SPSS practical courses will be a way that can be done to grow student interest in Statistics so that students feel that the statistics is easy, and as a result students do not feel anxious, fearful and worried about Statistics.

In 2014, Jatnika&Haffashave started to develop alternative software to analyze sta-

tistical data named UNPAD SAS (Statistical Analysis Series). UNPAD SAS is software to analyze Statistic data not found in SPSS. The UNPAD SAS that has been developed contains two analyzes, namely: the calculation of the minimum sample size and the pair of comparison from Thurstone.

In 2015, Jatnika&Haffascontinue the development of UNPAD SAS by completing other statistical analyzes, namely: Theta, Eta, and Jaspencerrelations. Results of UNPAD SAS trials of 8 students of Faculty of Psychology UNPADshowed that:

1. UNPAD SAS (Statistical Analysis Series): Theta, Eta, and Jaspencerrelations Analysis is software that is "user friendly" to calculate Theta, Eta, and Jaspencerrelations.
2. UNPAD SAS (Statistical Analysis Series): Theta, Eta, and Jaspencerrelations Analysis can be used to calculate the correlation of Theta, Eta and Jaspencer quickly and accurately.
3. UNPAD SAS (Statistical Analysis Series): Theta, Eta, and Jaspencerrelations Analysis needs to be added with import export file facility that is easy to do, for file storage security and to be easily combined with other programs like Microsoft Office.

Based on the findings, it appears that UNPAD SAS should be developed with a better database management that has at least the ability to import export files, Edit files (Copy, Cut, Clear) and Data Management (Sort, Select).

In 2016, Jatnika&Haffas resumed the manufacturing of UNPAD SAS for the minimum sample size calculation so that three methods for calculating the minimum sample size were: mean estimation, proportion estimation and correlation/regression estimation. The results of this study indicate that:

1. UNPAD SAS (Statistical Analysis Series): Minimum Sample Size Calculation applicable to calculate minimum sample size for mean estimation, proportion estimation, correlation/regression estimation rapid and accurate

2. Advanced research needs to be done to develop the UNPAD SAS (Statistical Analysis Series): Calculation of Minimum Sample Size, especially to simplify install MySQL program, create more attractive appearance and manual creation with some alternative suggestions in case of error.

III. RESEARCH METHODOLOGY

Computer software development methods are often called Software Development Life Cycle (SDLC). SDLC is a step or phase sequence that presents a model for development management and application life cycle or software. Methodologies in the SDLC process may vary across industries and organizations, but in general are as follows:

1. **Planning** - is the planning stage of the system. This stage emphasizes the aspect of system development feasibility.
2. **Analysis** - is a purification step to formulate the necessary functions and operations in the application.
3. **Design** - is a stage to describe desirable features and detailed operations, including screen layout, business rules, process diagrams, pseudo and other documentation.
4. **Implementation** - is the stage of implementing the design from the previous stages and test.
5. **Testing** - is the stage of testing the system that has been made.
6. **Maintenance** - is the stage of system operation by users, including maintenance and feedback from users.

In a software development, generally there are two parties, the user (user) and the developer (developer). The user is the party that needs the system to be developed or the owner of the project. The developer is the party who works/builds the system or the project implementer. Other parties (third parties) may be added. It is the end-user, the party that will actually use the system.

In this study the user is the Universitas Padjadjaran, the developer is the research team, and the end user is anyone who needs

this system. However, the Universitas Padjajaran is represented by the research team in determining the requirement of the system to be developed.

In this study, the "planning" phase refers to the Goals, Benefits, and Research Urgency so that researchers view this system as feasible to develop. The "Analysis" stage will be conducted by reviewing the Statistical methods used in this study and formulating the functions and operations required to build this system. The "Design" stage will be done by breaking down the results of the analysis. The "Implementation" stage will be performed to formulate the system algorithm, coding, and experimenting the system internally. Given the team's development team is very little, then the stage of "Testing" will be done by involving students to try to use this system. The "Maintenance" stage will be done by distributing this system through Universitas Padjajaran to end users.

IV. RESULTS AND DISCUSSION

In this section will be described the stage of Testing UNPAD SAS produced that was tested on 144 students of Faculty of Psychology UNPAD. Results of student polls on the resulting program can be seen in the following table:

Table 1
Result of Polls against Software Created

NO	ITEMS	Strongly Disagree	Disagree	Agree	Strongly Agree
1	Install My SQL Server is easy to do	5.3	15.8	68.4	10.5
2	Install My SQL Connector is easy to do	5.3		84.2	10.5
3	Install Unpad SAS is easy to do			78.9	21.1
4	Setting up data base is easy to do		10.5	68.4	21.1
5	Configuring ODBC is easy to do		26.3	52.6	21.1
6	Creating data file is easy to do		21.1	63.2	15.8
7	Data file storage is easy to do		5.3	63.2	31.6
8	Editing data process is easy to do		31.6	63.2	5.3
9	The data input process using this program is easy to do		31.6	57.9	10.5
10	The displayed results are		15.8	47.4	36.8

NO	ITEMS	Strongly Disagree	Disagree	Agree	Strongly Agree
	easy to understand				
11	Display databases that appear is interesting		57.9	42.1	
12	The process of importing data from other programs (SPSS, Excel, etc.) is easy to do	10.5	10.5	63.2	15.8
13	The Database Management module is easy to understand			89.5	10.5
14	The process of data analysis using descriptive statistics on the program is easy to do			68.4	31.6
15	Display of Descriptive Statistics is interesting		47.4	47.4	5.3
16	The graphic view is interesting		63.2	31.6	5.3
17	The Descriptive Statistics Module is easy to understand			73.7	26.3

Note: all values in percent

The inputs provided to refine the program are:

1. The install process is made easier
2. The process of data input is made easier, so that when an error occurs in input data does not need to exit the application first
3. Reduce the bug so that when an error occurs it does not become an application error
4. Graphical display made more interesting

Based on the data in Table 1 above, it appears that in general software programs for Database Management and Descriptive Statistics are already software that is "user friendly". This is evident from the results of a user poll stated that 79% of users agree and strongly agree that this program making the data file is easy to do. It also appears that 94.8% of users stated that file storage is also easy to do. This is supported also by the existence of 68.4% of users stated that data input is easy to do, 68.5% said edit data is easy to do. Descriptive Statistics data analysis is also easy to do and easily understood which is indicated by 100% of users who agree and strongly agree. As for 52.7% of users stated that the Descriptive Statistics view is interesting and only 36.9% of users said the display graph is interesting.

V. CONCLUSIONS

Based on the results of the experiment can be concluded that UNPAD SAS (Database Management and Descriptive Statistics) is the right software and "user friendly" to analyze statistical data effectively and efficiently. The resulting software must be improved to make the output look more attractive. This software will be applied to the teaching of Statistics courses, making it easier for students to understand Statistics and increase interest and reduce students' anxieties and concerns in studying Statistics. This software can also reduce the rampant piracy of software that occurs.

REFERENCES

- [1] Jatnika, Ratna, "The Psychometric Properties of Student Attitudes toward Statistics", *International Meeting of Psychometric Society*, Hongkong, 2011
- [2] Jatnika, Ratna and Abidin, Fitri, The Effect of Practicum Method to Students Attitudes Toward Statistics, *International Conference on Applied Statistics*, Bandung, 2013.
- [3] Jatnika, Ratna and Abidin, Fitri, Models of Factors Affecting the Student Mastery of Statistics in Psychology Faculty Universitas Padjadjaran, *Asian Association of Indigenous and Cultural Psychology Conference*, Bandung, 2015
- [4] Jatnika, Ratna, The Effect of SPSS Course to Students Attitudes toward Statistics and Achievement in Statistics, *International Journal of Information and Education Technology*, Vol. 5 No. 11, IACSIT Press, 2015
- [5] Jatnika, Ratna and Haffas, Mustofa, Universitas Padjadjaran Statistical Analysis Series: Theta, Eta, and Jaspén Correlation, *9th Annual International Conference of Education, Research and Innovation (ICERI 2016)*, Seville Spain, 2016
- [6] "KDnuggets Annual Software Poll: Analytics/Data mining software used?" *KDnuggets*, May 2013
- [7] Schau, Candace, Students' Attitudes: The "Other" Important Outcome In Statistics Education, *2003 Joint Statistical Meetings - Section on Statistical Education*, 2003
- [8] Wellman, B. "Doing It Ourselves: The SPSS Manual as Sociology's Most Influential Recent Book." pp. 71–78 in *Required Reading: Sociology's Most Influential Books*, edited by Dan Clawson. Amherst: University of Massachusetts Press, 1998