Brief Overview of STATCAL Statistical Application Program

Dian Utami Sutiksno1*, Prana Ugiana Gio2, Rezzy Eko Caraka3, & Ansari Saleh Ahmar4

1Politeknik Negeri Ambon, Indonesia
2Universitas Sumatera Utara, Medan, Indonesia
3School of Mathematical Sciences, The National University of Malaysia, Malaysia
4Department of Statistics, Universitas Negeri Makassar, Indonesia

*dsutiksno@gmail.com

Abstract. Currently, there are many application programs to perform statistical analysis, such as SPSS, EViews, and Minitab, which are commercial software, while PSPP, JASP and PAST, which are free software. STATCAL is an user-friendly statistical application program which is developed using R programming language, in RStudio using various R packages. STATCAL is designed as simple as possible so that only need a bit of step to obtain result. Various statistical tests in STATCAL, such as normality, homogeneity, comparison of two or more means, correlation, association between categorical variables, reliability, linear regression, panel data regression, covariance-based structural equation modeling and partial least square path modeling are available. Inside STATCAL is also provided tutorial video and guidance menu to make easy for user.

Keywords: STATCAL, R, RStudio

1. Introduction

There are two alternatives to use statistical application program, namely free or commercial statistical application program. Commercial statistical application program such as SPSS, EViews and Minitab, while free statistical application program such as PSPP, JASP and PAST. STATCAL is also statistical application program which is free. STATCAL is created using R programming language in RStudio using various R packages (https://statcal.info/r-packages-in-statcal/) to perform graphical and statistical analysis. STATCAL is created by STATCAL team, on January 2017. STATCAL is an user-friendly statistical application program which is designed as simple as possible so that only need a few step to obtain result.

STATCAL can be used by online accessing http://gioprana.shinyapps.io/STATCAL or can be downloaded in https://statcal.info/download-statcal/. One of difference STATCAL with other statistical application program is tutorial video and guidance menu can be accessed directly inside STATCAL, so that make easy for user. Covariance-based structural equation modeling and partial least square path modeling are also available in STATCAL.
2. Getting Started STATCAL
Figure 1 is an initial appearance when STATCAL statistical application program is run.

![Initial Appearance in STATCAL](image)

There are nine main menus. STATCAL menu explains brief description about STATCAL. STATCAL has two areas to input data, namely for numeric and categorical data, that will be explained in subsequent section. Various graphs, such as bar, line, circle, scatter plot, and boxplot, can be formed in Graph menu. Descriptive measures, frequency and percentage distribution are available in Descriptive menu. Various statistical tests can be accessed in Statistics menu. Probability menu can generate random number based on certain probability distribution, such as normal and t distribution and then these random numbers can be presented in to bar chart or histogram. For user who is first time uses STATCAL, Tutorial Video and Guidance menu will help how to use STATCAL.

2.1 Data Preparation. First step before entering data in STATCAL is data preparation. Data preparation in this illustration uses Microsoft Excel. Data in Microsoft Excel will be copied and pasted to STATCAL later. Figure 2 is an example of simple data which contains three parts. In part I, there are two categorical variables, namely gender and blood. Part II displays value and label based on part I. For instance, value 1 in gender states male, while value 2 states female. Part III contains two numeric variables, namely height and weight. Based on Figure 2, consider that there are two colors, pink and yellow. Data in yellow region will be copied and pasted in to STATCAL later.
2.2 Step by Step to Input Categorical Data in STATCAL. In this part will be explained how to enter categorical data in STATCAL. In the previous section, data has been prepared in Microsoft Excel. Now, data in yellow region for part I and part II is copied and pasted to STATCAL (Figure 3). Figure 3 displays categorical data has been entered in STATCAL. First, consider part of Set Number of Column. In this part is asked to enter number of categorical variable. Number of categorical variable is 2, but it can be filled with 3, 4, 5 or a number greater than 2. This is because perhaps exists addition of variable later. So Set Number of Column part is filled with 20. Second, consider Set Number of Row part. The number of observation is 10. But it can be filled 11, 13, 1000 or a number greater than 10. Enter name of categorical variable in Name of Variable part. The last part is Label. In this part, enters the label for each categorical variable. For example for gender variable, value 1 states male and value 2 states female. For blood variable, value 1 states A, value 2 states B, value 3 states AB and value 4 states O.
2.3 Is Correct Your Categorical Data Input? In the previous part has been explained step by step how to enter categorical data in STATCAL. But, is correct your categorical data input? How to check it? To ensure it, move to Your Categorical Data part (Figure 4). In this part displays again categorical data but the data has been converted become label. Check, is there any mistake? In this part will be explained how to enter categorical.

![Figure 4 Examination of Categorical Data Input](image)

2.4 Your Categorical Data Input is Wrong and Not Available Data <NA> Figure 5 is an example when categorical data input is wrong. For example observation of row 8 and column 1 (gender variable) is replaced with 4. In Your Categorical Data part will display "Error: 'from' and 'to' vectors are not the same length.". This is because there are only two labels in gender variable, namely male and female, while there are three kind of numbers 1, 2 and 4. For example observation of row 8 and column 1 (gender variable) is not filled or empty (Figure 6). Now, consider Your Categorical Data part, observation of row 8 and column 1 is filled with <NA>. 

![Image 2.4](image)
Figure 5  Example of When Categorical Data Input is Wrong

Observation of row 8 and column 1 (gender variable) is filled with 4.

Figure 6  Example of Not Available Data <NA>

Observation of row 8 and column 1 (gender variable) is not filled or empty.
2.5 Step by Step to Input Numeric Data in STATCAL. The final stage is how to input numeric data in STATCAL. This is simpler than input categorical data. Based on Figure 2, copies numeric data, namely part III for yellow region and moves to **Input Numeric Data** menu (Figure 7). Consider part of **Set Number of Column**. In this part will be filled with a value representing the number of numeric variable. The number of numeric variable is 2, but in this part can be filled with a number which is greater than 2. In this illustration is filled with 20. This is because anytime perhaps will exist addition of variable. Now consider the part of **Set Number of Row**. For the **Set Number of Row** part is filled with a value representing the number of observation. The number of observation is 10, but it can be filled with a number which is greater than 10. Based on Figure 7, Set Number of Row part is filled with 40. In part of **Name of Variable** is filled with name of numeric variable.

**Figure 7**  **Input Numeric Data**

3. Graph
In the previous section has been deeply explained step by step how to input categorical and numeric data. This part will display how to make graph in STATCAL. STATCAL provides various types of graph such as bar, line, circle, boxplot, scatter plot, and correlation. Bar menu provides 6 types of bar, line 5 types, circle 2 types, boxplot 2 types, scatter plot 2 types and correlation two types (Figure 8). But in this article only will be displayed a bit of graph.
3.1 Frequency Bar Graph for Single Categorical Variable. In this section will be explained how to make frequency bar graph for single categorical variable. First, click Graph => Bar (Figure 9). There are six menus of bar. Each menu has different characteristic such as selection of input variable. In this part chooses Bar I (Figure 10). In Bar I is asked to choose single categorical variable. For this illustration chooses gender in Choose Variable part (Figure 10). The result can be seen in Result part (Figure 11). Adjustment of bar such as title of bar graph, x-axis, y-axis, legend and color of bar can be set there as your wish.
3.2 Frequency Bar Graph for Two Categorical Variables. In the previous section has been explained how to make frequency bar graph for single categorical variable. In this part will be explained how to make frequency bar graph for two categorical variables. First, click **Graph => Bar**. Then chooses **Bar II** (Figure 12). In **Bar II** is asked to choose single categorical variable for x-axis and single categorical variable for y-axis. For this illustration chooses blood for x-axis and gender for y-axis (Figure 12). The result can be seen in **Result** part (Figure 13). Adjustment of bar such as title of bar graph, x-axis, y-axis, legend and color of bar can be set there as your wish.
4. Normality Test
Suppose that variable of height and weight will be tested normality assumption. To do this, first click Statistics => Test of Normality (Figure 14). In Figure 15 selects height and weight variable. The result is displayed in part of Result (Figure 16). In STATCAL, there are 5 results of normality test such as Kolmogorov-Smirnov (asymptotic or exact approach), Shapiro-Wilk, Jarque-Bera and Anderson-Darling. Figure 16 displays normality test results. In this Result part is also displayed boxplot for height and weight variable (Figure 17).
Figure 14 Normality Test Menu

Test of Normality (Assumption)

Choose Numeric Variable (Multiple Choice)

Height
Weight

Figure 15 Selection of Numeric Variable in Normality Test Menu

Kolmogorov-Smirnov Test (Asymptotic Approach)

| Variable | Statistic of Kolmogorov-Smirnov (KS) | P-Value of KS | Conclusion |
|----------|--------------------------------------|---------------|------------|
| Height   | 0.1912                               | 0.8581        | p-value > 0.05, assumption of normality is received, at the level of significance 5% |
| Weight   | 0.1256                               | 0.9975        | p-value > 0.05, assumption of normality is received, at the level of significance 5% |

Kolmogorov-Smirnov Test (Exact Approach)

| Variable | Statistic of Kolmogorov-Smirnov (KS) | P-Value of KS | Conclusion |
|----------|--------------------------------------|---------------|------------|
| Height   | 0.1912                               | 0.7935        | p-value > 0.05, assumption of normality is received, at the level of significance 5% |
| Weight   | 0.1256                               | 0.9913        | p-value > 0.05, assumption of normality is received, at the level of significance 5% |

Shapiro-Wilk Test

| Variable | Statistic of Shapiro-Wilk (SW) | P-Value of SW | Conclusion |
|----------|--------------------------------|---------------|------------|
| Height   | 0.9093                          | 0.2762        | p-value > 0.05, assumption of normality is received, at the level of significance 5% |
| Weight   | 0.9764                          | 0.943         | p-value > 0.05, assumption of normality is received, at the level of significance 5% |
5. Tutorial Video and Guidance
For the user who is first time uses STATCAL, perhaps Tutorial Video and Guidance menu will really help to guide you step by step how to process data in STATCAL (Figure 18 and Figure 19). Various videos in Tutorial Video menu can be played in connected internet (Figure 18). Articles about STATCAL can be downloaded in Guidance menu (Figure 19).
6. Summary
Basically STATCAL is created using R programming language in RStudio using various R packages to perform graphical and statistical analysis. It means that if you install STATCAL, you will install R and various R packages. Tutorial Video and Guidance menu will guide you to learn STATCAL step by step. STATCAL will continue to be fixed and updated such as icon image, graph, and statistical tests. For the future will be attached tutorial video for each menu in STATCAL.

References
[1] https://cran.r-project.org/web/packages/shinythemes/index.html
[2] https://cran.r-project.org/web/packages/lavaan/index.html
[3] https://cran.r-project.org/web/packages/plspm/index.html
[4] https://cran.r-project.org/web/packages/ggplot2/index.html
[5] https://cran.r-project.org/web/packages/reshape2/index.html
[6] https://cran.r-project.org/web/packages/dplyr/index.html
[7] https://cran.r-project.org/web/packages/shiny/index.html
[8] https://cran.r-project.org/web/packages/plyr/index.html
[9] https://cran.r-project.org/web/packages/DT/index.html
[10] https://cran.r-project.org/web/packages/doBy/index.html
[11] https://cran.r-project.org/web/packages/GGally/index.html
[12] https://cran.r-project.org/web/packages/plotrix/index.html
[13] https://cran.r-project.org/web/packages/tsseries/index.html
[14] https://cran.r-project.org/web/packages/nortest/index.html
[15] https://cran.r-project.org/web/packages/car/index.html
[16] https://cran.r-project.org/web/packages/RVAideMemoire/index.html
[17] https://cran.r-project.org/web/packages/plm/index.html
[18] https://cran.r-project.org/web/packages/colourpicker/index.html
[19] https://cran.r-project.org/web/packages/psych/index.html
[20] https://cran.r-project.org/web/packages/DescTools/index.html
[21] https://cran.r-project.org/web/packages/semPlot/index.html