COURSE MODULE IN

IT 318-Quantitative Methods

Bachelor of Science in Information Technology

1st Semester S/Y 2020 – 2021

I. Cover Sheet

A. MODULE NUMBER : **2. 7**

B. MODULE TITLE : **Introduction to Data Analysis: Data Preparation and Presentation**

C. DURATION : **Two (2) weeks**

D. MODULE PREREQUISITE : **Module 1.6**

E. INSTRUCTOR : **ICHELLE F. BALUIS**

II. Module Objectives (Desired Learning Outcomes)

1. Organize raw data for analysis ;
2. Encode data in the statistical software editor (MS Excel, SPSS, PSPP);
3. Present data in tabular and graphical format.

III. Vocabulary List

- **Data preparation:** involves logging in the data ; entering the data into the computer; checking the data for accuracy.
- **Data Presentation:** organizition of data in a manner which makes it easy for the statistician to analyse and consequently interpret the result.
- **PSPP:** is a GNU-based free software application for analysis of sampled data, intended as a free alternative for IBM SPSS Statistics.
- **SPSS:** a user-friendly program that facilitates data management and statistical analyses.

IV. Learning Plan

ACTIVITY	DURATION (HH:MM)	REMARKS*
Do Pre-Test	00:15	
Study Objective 1,2 ,3, Information Sheet 1: Introduction to Data Analysis	03:00	
Do Objective 1,2, 3 and Learning Task M2.1	02:30	
Do Post-Test	00:15	

*Check if accomplished

V. *Pre-test*

This pre-test will try to assess your knowledge prior to the discussion of the topic.

Instructions are provided for each part of the test.

Item		
Part 1 True or False: Please evaluate the following statement if it expresses a correct thought or not. Please put a check on the column that corresponds to your answer.	**True**	**False**
1. **Data preparation involves checking data for accuracy.**		
2. **Descriptive statistics tests specific research hypothesis.**		
3. **As a researcher, it is very vital that you screen the collected data for accuracy.**		
4. **Google Forms has a feature of automatically organizing the collected responses.**		
5. **Response data set produced by Google Forms cannot be exported to excel spread sheet file.**		
6. **Variable tab in SPSS allows you to encode questionnaire responses.**		
7. **In SPSS, if the variables will be used as predictor the Role should be set to Input.**		
8. **SPSS can count missing data for a variable.**		
9. **Frequency of responses can be counted in MS Excel by using the formula or function COUNT.**		
10. **Same with MS Excel, SPSS can count frequency of responses by allowing end user to create formula.**		

VI. *Information Sheet*

Information Sheet 1: Introduction to Data Analysis

What is Data Analysis?

Data analysis involves four major steps:

1. **Data Preparation:** involves logging in the data ; entering the data into the computer; checking the data for accuracy.
2. **Data Presentation:** once data have been collected, they must be organized and presented in a manner which makes it easy for the statistician to analyse and consequently interpret the result.
3. **Descriptive Statistics:** describe the basic features of the data in a study. It provides meaningful summaries about the sample so that potential patterns may emerge from the data. We will discuss this further in the succeeding modules.

4. **Inferential Statistics:** inferential statistical analysis tests specific research hypotheses. We will discuss this further also in the succeeding modules.

In this module, we will first cover the first two major steps in data analysis, the data preparation and data presentation.

Data Preparation

In your research project, you might have data coming from several different sources at different times, such as:

- Survey returns;
- Coded interview data;
- Pre-test and post-test result;
- Observational data.

Having any of these mentioned data, you need to set up a procedure for checking the accuracy of collected data, *logging the data* on and keeping track of it until you are ready to do a comprehensive data analysis.

Let us consider first the checking for data accuracy. As a researcher, it is very vital that you screen the *collected data for accuracy*. Doing this right away allows you to go back to the sample to clarify any problems or errors.

The following questions should be asked as part of this initial data screening:

- Are the responses legible / readable?
- Are all important questions answered?
- Are the responses complete?
- Is all relevant contextual data included?

Ensuring the accuracy of the collected data helps ensure the overall quality of subsequent analysis.

Once checked for accuracy, you can now log the data for you to keep track them. There is no single way how researcher log and keep track of incoming data. In most cases, researchers use standard computerized spread sheet (MS-Excel) or database (MS-Access) program. Some are also using standard statistical program (SPSS, PSPP, SAS,Minitab etc.) by running simple descriptive analysis to get report on data status.

****In this course we will be using Microsoft Excel, SPSS (Statistical Package for Social Sciences), and PSPP. You can open the following link to download SPSS and PSPP.*

PSPP: https://sourceforge.net/projects/pspp4windows/

SPSS: https://spss-64bits.en.softonic.com/download

Let's Try

Please recall the sample electronic questionnaire from this link: https://docs.google.com/forms/d/1StFn6PYw_w3BCk0aO4PM2VoKpb0sqzATq4MNsFbOODQ/edit?usp=sharing

Or the paper-based questionnaire provided as attachment in Module 1.6

Suppose you conducted the survey already and the responses are now ready (the sample electronic questionnaire has six responses already, you can resend them and collect at least 30 responses).

How are you going to prepare the collected data for analysis? We can make use either of the following ways:

Via Google Forms

One of the good features of Google Forms is that it automatically organize the collected responses in spread sheet so no need to encode manually the responses which will really save time for the researcher. It also automatically generates charts/graphs that show the trend/pattern of the collected data. The following steps will guide you on how to view the organized data made by Google Forms for you:

1. Open the electronic questionnaire via this link: https://docs.google.com/forms/d/1StFn6PYw_w3BCk0aO4PM2VoKpb0sqzATq4MNsFbOODQ/edit?usp=sharing
2. Open Responses tab (this will show the trend/pattern of the collected responses in a form of graph/chart.
3. Retrieve the collected responses organized in the spread sheet by clicking the icon (view responses in sheet). Automatically it will generate a spread sheet that contains the collected responses.

   **you may also download the spread sheet file via File menu , then Download.

Manual Inputting of Responses to MS-Excel

In case you run a paper-based questionnaire, the most common way of logging the collected data is through MS-Excel. The following steps will guide you on how to organize the collected questionnaire responses to MS-Excel.

1. Open MS-Excel. Select a sheet and name it (ex. questionnaire_responses).
2. In the selected sheet, add column headings. The name of the column is suggested to be named according to variable naming conventions and

must be related to its intended purpose. The following spread sheet structure is suggested for the collected data:

Respondent Number	QuestionnaireMainIndicator1	QuestionnaireMainIndicator2	QuestionnaireMainIndicator3	QuestionnaireMainIndicatorN

Example (based on the sample questionnaire):

RespondentNum	PositionandDesignation	ExistingSystemPerformance_Storage					ExistingSystemPerformance_Organization				ExistingSystemPerformance_Security					SystemFeatures
		ST1	ST2	ST3	ST4	ST5	O1	O2	O3	O4	SE1	SE2	SE3	SE4	SE5	

***

ST1…ST5 (STORAGE indicators 1 to 5)
O1…O5 (ORGANIZATION indicators 1 to 4)
SE1…SE5 (SECURITY indicators 1 to 5)

3. Once you are done structuring your spread sheet you can now start inputting the responses.

RespondentNum	PositionandDesignation	ExistingSystemPerformance_Storage					ExistingSystemPerformance_Organization				ExistingSystemPerformance_Security					SystemFeatures
		ST1	ST2	ST3	ST4	ST5	O1	O2	O3	O4	SE1	SE2	SE3	SE4	SE5	
1	**2**	**1**	**2**	**2**	**2**	**2**	**3**	**2**	**1**	**2**	**2**	**3**	**1**	**2**	**3**	**1,2**
2	**1**	**1**	**1**	**1**	**2**	**2**	**1**	**1**	**2**	**1**	**2**	**1**	**1**	**1**	**1**	**2**
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
30	**1**	**1**	**2**	**2**	**2**	**2**	**2**	**2**	**1**	**1**	**1**	**2**	**3**	**3**	**1**	**4,5**

What to input?:

RespondentNum: you can number the questionnaire with responses from 1 to N (total number of respondents); the number of that questionnaire will serve also as the respondent number; this is one way of keeping the confidentiality of collected data wherein each respondent will be represented only with a number)

PositionandDesignation: you can assign code to position or designation so that there is no need to code the text "Record Officer" and "Record Section Staff"; you can assign code as follows: **1** for Record Officer and **2** for Record Section Staff; since it is a nominal data, the numbers are just labels; once code are assigned already, you can now input the PostionandDesignation responses using the code only.

ExistingSystemPerformance_Storage: the actual scale response should be inputted;

ExistingSystemPerformance_Organization: the actual scale response should be inputted

ExistingSystemPerformance_Security: the actual scale response should be inputted

SystemFeatures: you can assign code to system features so that there is no need to code the text "Accessible;Usable;Maintainable;Reliable;Secure;Others"; you can assign code as follows: **1** for Accessible; **2** for Usable; **3** for Maintainable; **4** for Reliable; **5** for Secure; **6** for Others ; since it is a nominal data, the numbers are just labels; once code are assigned already, you can now input the SystemFeatures responses using the code only.

4. Continue inputting the responses up to the last respondent. Check for erroneous inputs. Save any changes made to the spread sheet.

   **Please see sample excel file: https://drive.google.com/file/d/10XMscPFeMfCxgEzXu_9vDY3VcBZsCdWp/view?usp=sharing

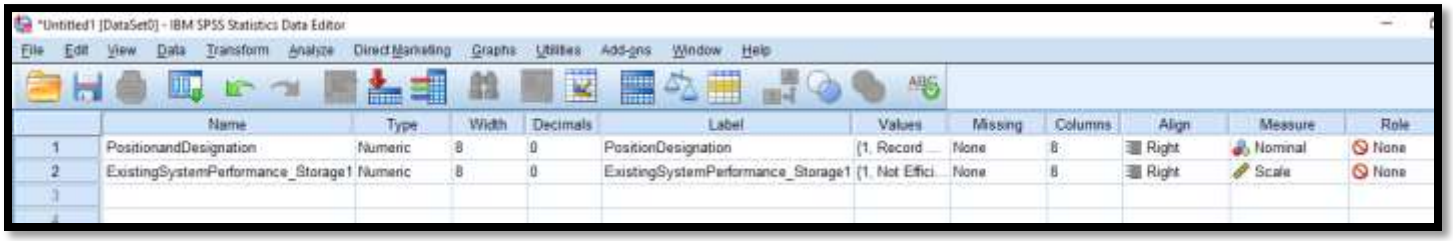
Manual Inputting of Responses to SPSS

SPSS is a user-friendly program that facilitates data management and statistical analyses. However the end-user should subscribe to a paid license to fully take advantage of its features.

1. Open IBM SPSS Statistics.
2. In the Data Editor View, set up the following:

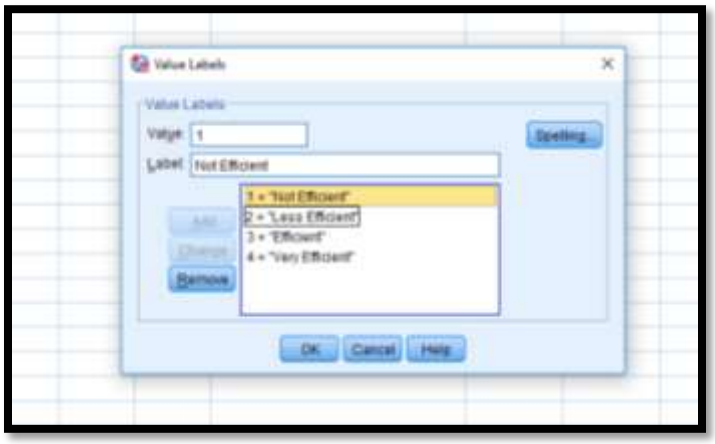
   Variables (via Variable View)
   Variables being measured in the questionnaire should be clearly identified first.

   a. Click **Variable View Tab**.
   b. Set-up the properties of the variables by inputting in the field the following details:



Please take note of the following properties:

Name: researcher made variable name; this is sensitive, illegal characters, white spaces are not allowed;
Type: specify the type of variable, if it is numeric, string etc.
Width: The number of digits displayed for numerical values or the length of a string variable.
Decimals: The number of digits to display after a decimal point for values of that variable
Label: A brief but descriptive definition or display name for the variable.
Values: For coded categorical variables, the value label(s) that should be associated with each category abbreviation. Value labels are useful primarily for categorical (i.e., nominal or ordinal) variables, especially if they have been recorded as codes (e.g., 1, 2, 3). It is strongly suggested that you give each value a label so that you (and anyone looking at your data or results) understands what each value represents. Please see example below.

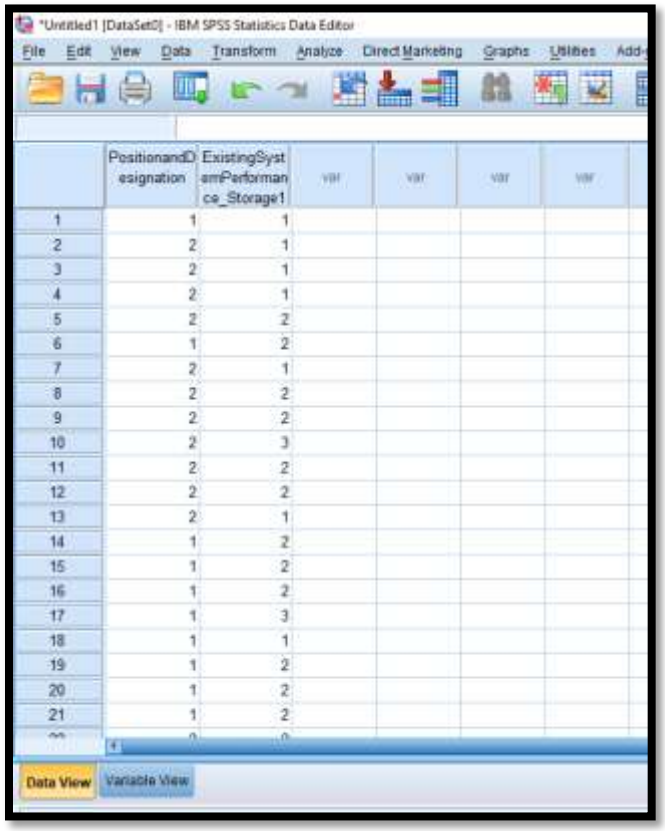


Missing: The user-defined values that indicate data are missing for a variable
Columns: The width of each column in the Data View spreadsheet.
Align: The alignment of content in the cells of the SPSS Data View spreadsheet. Options include left-justified, right-justified, or center-justified.
Measure: The level of measurement for the variable (e.g., nominal, ordinal, or scale).
Role: The role that a variable will play in your analyses (i.e., independent variable, dependent variable, both independent and dependent).

Input: *The variable will be used as a predictor (independent variable). This is the default assignment for variables.*
Target: *The variable will be used as an outcome (dependent variable).*
Both: *The variable will be used as both a predictor and an outcome (independent and dependent variable).*
None: *The variable has no role assignment.*
Partition: *The variable will partition the data into separate samples.*
Split: *Used with the IBM® SPSS® Modeler (not IBM® SPSS® Statistics).*

Data (via Data View)

a. Click **Data View tab**. The pre-setup variable will now appear as column heading.
b. Manually input the responses into each designated cells same as what you did in MS-Excel. Please see sample screen below:



3. Continue inputting the responses up to the last respondent. Check for erroneous inputs. Save the file (ex. sample_data.sav) .

Manual Inputting of Responses to PSPP

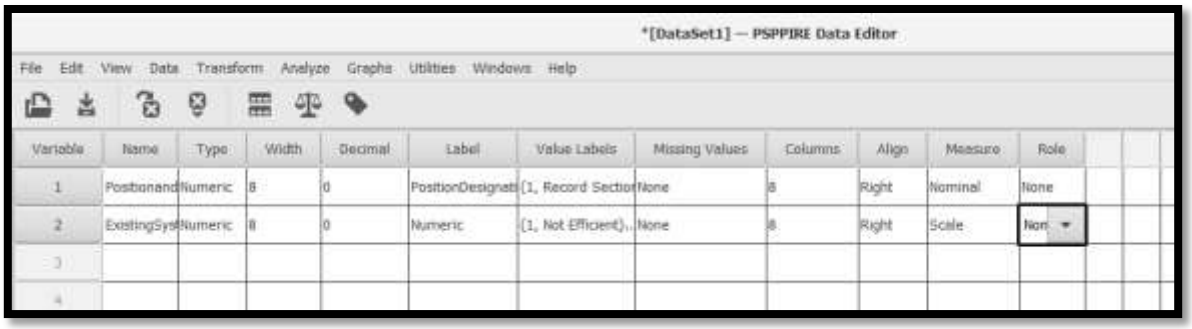
PSPP is a GNU-based free software application for analysis of sampled data, intended as a free alternative for IBM SPSS Statistics. The features and procedures for

data analysis are almost similar with SPSS. You can follow almost the same procedure as SPSS when inputting data.

1. Open PSPP.
2. In the Data Editor View, set up the following:

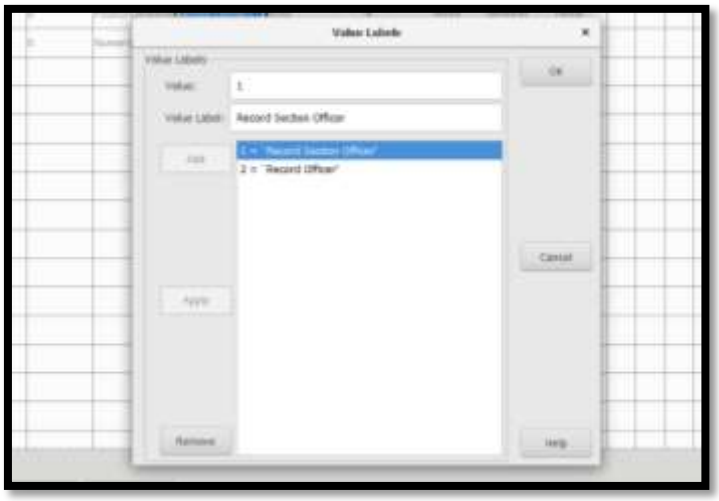
Variables (via Variable View)
Variables being measured in the questionnaire should be clearly identified first.

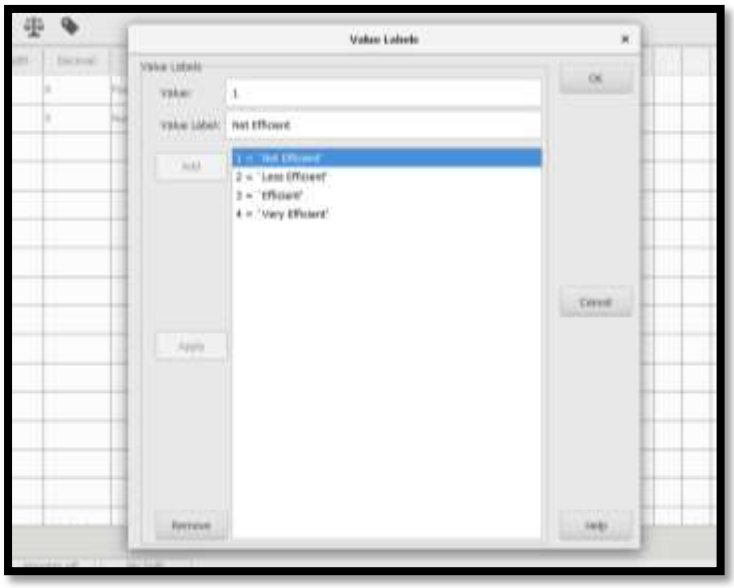
a. Click **Variable View Tab**.
b. Set-up the properties of the variables by inputting in the field the following details:



Please take note of the following properties:

Name: researcher made variable name; this is sensitive, illegal characters, white spaces are not allowed;
Type: specify the type of variable, if it is numeric, string etc.
Width: The number of digits displayed for numerical values or the length of a string variable.
Decimals: The number of digits to display after a decimal point for values of that variable
Label: A brief but descriptive definition or display name for the variable.
Values: For coded categorical variables, the value label(s) that should be associated with each category abbreviation. Value labels are useful primarily for categorical (i.e., nominal or ordinal) variables, especially if they have been recorded as codes (e.g., 1, 2, 3). It is strongly suggested that you give each value a label so that you (and anyone looking at your data or results) understands what each value represents. Please see example below.





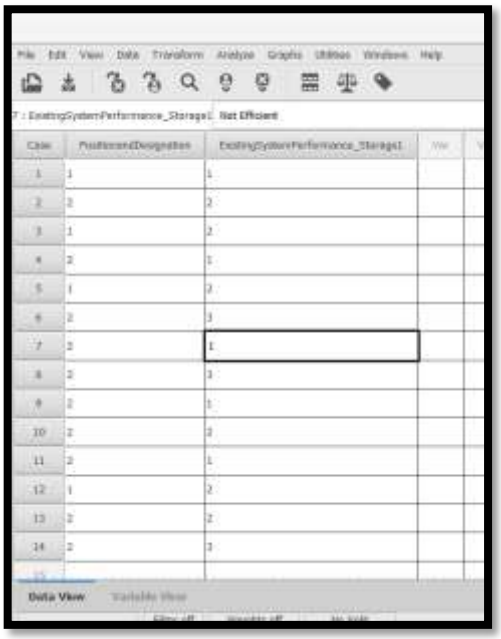
Missing: The user-defined values that indicate data are missing for a variable
Columns: The width of each column in the Data View spreadsheet.
Align: The alignment of content in the cells of the SPSS Data View spreadsheet. Options include left-justified, right-justified, or center-justified.
Measure: The level of measurement for the variable (e.g., nominal, ordinal, or scale).
Role: The role that a variable will play in your analyses (i.e., independent variable, dependent variable, both independent and dependent).

Input: *The variable will be used as a predictor (independent variable). This is the default assignment for variables.*
Target: *The variable will be used as an outcome (dependent variable).*

Both: *The variable will be used as both a predictor and an outcome (independent and dependent variable).*
None: *The variable has no role assignment.*
Partition: *The variable will partition the data into separate samples.*
Split: *Used with the IBM® SPSS® Modeler (not IBM® SPSS® Statistics).*

Data (via Data View)

a. Click **Data View tab**. The pre-setup variable will now appear as column heading.
b. Manually input the responses into each designated cells same as what you did in MS-Excel and SPSS. Please see sample screen below:



3. Continue inputting the responses up to the last respondent. Check for erroneous inputs. Save the file (ex. sample_data.sav) .

Data Presentation

In general, the methods of presenting data are *textual, tabular and graphical*. In textual method, data are presented in paragraph form wherein through sentences, the behaviour of the data is being described. The tabular method uses tables as a means of data presentation. Lastly, the graphical method gives a visual representation of the data set.

In this section, we will be focusing on tabular and graphical presentation. Application software like MS-Excel, SPSS and PSPP will be used for this purpose.

Tabular Presentation

When table is used to present data it should consider different parts. Table number and heading are very necessary on top of the table. The important data presented in the table as reflected in the heading must be in important parts, that is the *box head.* Further, the *body* of the table shall contain the specific data that pertains to the box head. See sample table format below:

<Table #. Table Title>

***shaded cells is the **box head**; non-shaded cells is the **body**.

Example :

Table 1. Frequency and Percentage of Respondents

Respondent's Designation	Frequency	Percentage
Record Officer	19	63.33%
Record Section Staff	11	36.67%
Total	30	100.00%

- *Tables via MS-Excel*
  Since MS-Excel allows us to organize the data in columns and in rows, presenting data in tabular form is not difficult.

1. Select the spread sheet where collated data were encoded. Look for appropriate part where you can present your data in tabular format.
2. Key in the Table number and Table Title.
3. Key in the box heading. See example below:

Table 1. Frequency and Percentage of Respondents

Position /Designation	Frequency	Percentage
Record Officer		
Record Section Staff		
Total		

4. Using excel formula compute for the frequency, percentage and total.
   Sample formula:
   Frequency (record officer)
   =COUNTIF($C$2:$C$31,B35)
   Percentage(record officer)

=C35/$C$37
Total
=SUM(C35:C36)
See sample computed frequency, percentage and total below:

Table 1. Frequency and Percentage of Respondents

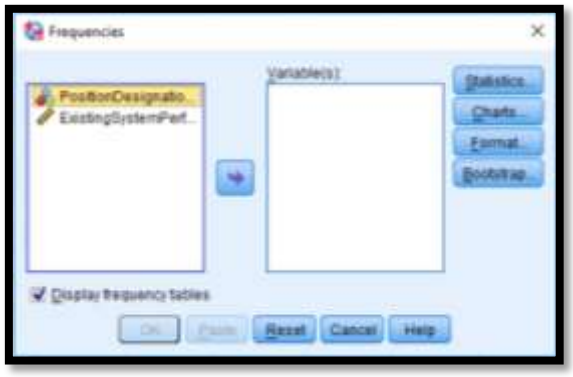
Position /Designation	Frequency	Percentage
Record Officer	19	63.33%
Record Section Staff	11	36.67%
Total	30	100.00%

5. It is recommended to apply borders enclosing the box heading and table body.
6. Save your work.

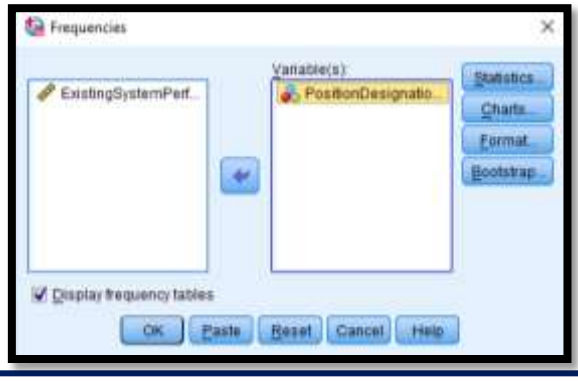
- *Tables via SPSS*

Unlike in excel, tables in SPSS can be generated automatically. Please be guided by the following steps:

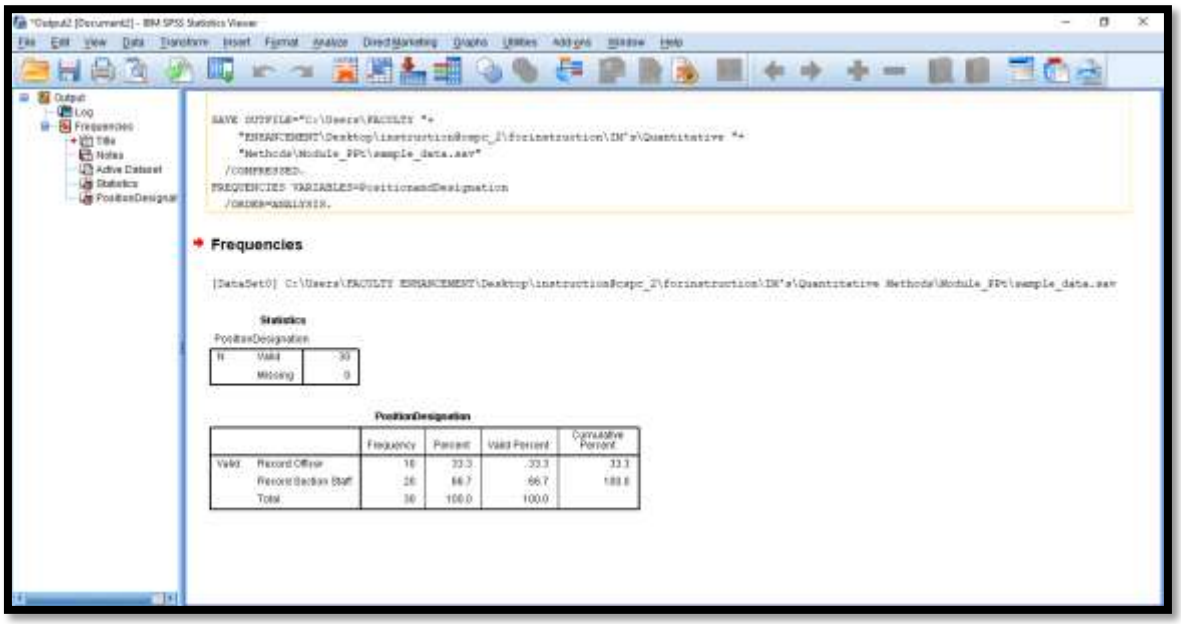
1. In the **Data View** tab, check if all the necessary data are already encoded completely.
2. Select **Analyze> Descriptive Statistics> Frequencies.**
3. In the Frequencies dialog box, select the variables that you want to create frequency table.



4. Check the box display frequency table then click OK.



5. In the output viewer, frequency table will now appear.

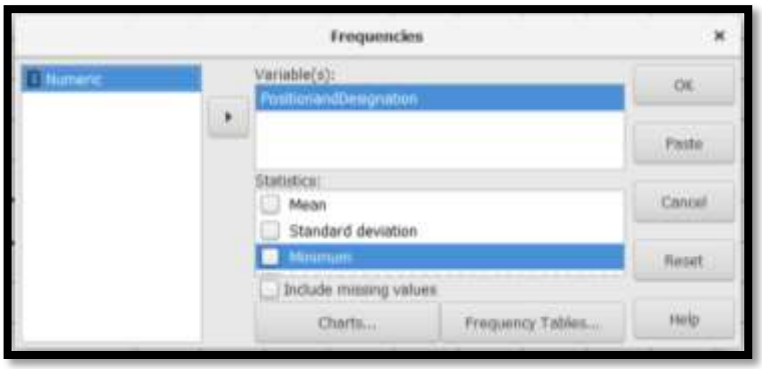


6. Save the output via File> Save As(sample file name: frequencytable.spv)

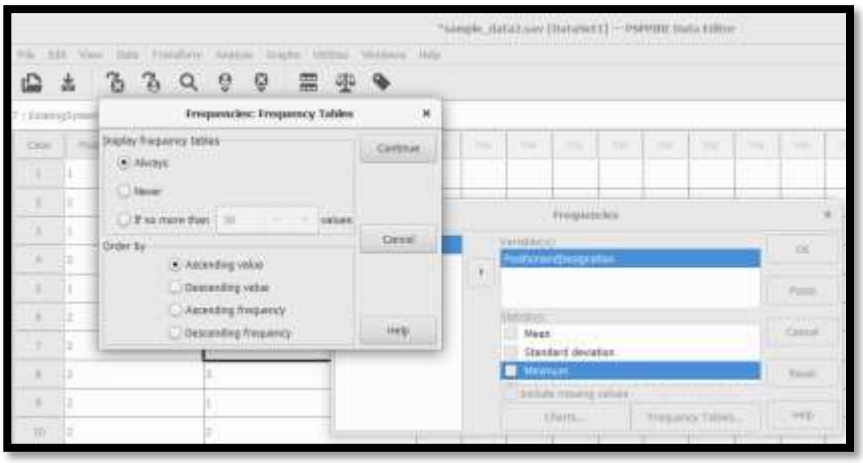
- *Tables via PSPP*

  The same in SPSS, tables in PSPP can be generated automatically also. Please be guided by the following steps:

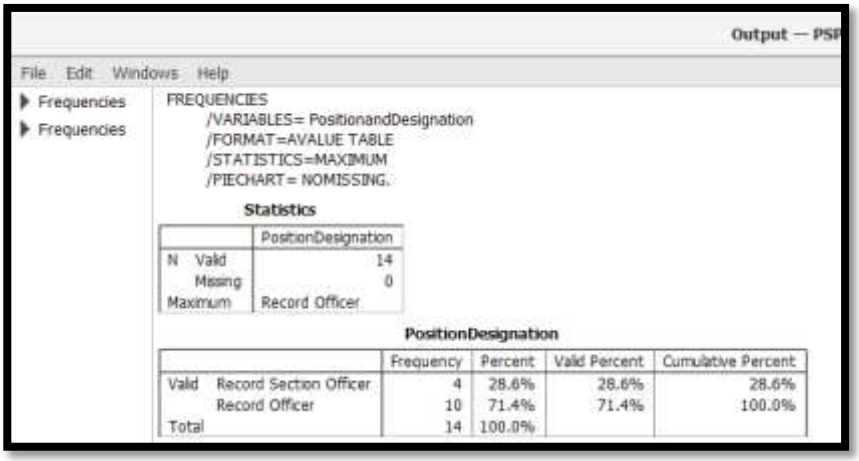
  1. In the **Data View** tab, check if all the necessary data are already encoded completely.
  2. Select **Analyze> Descriptive Statistics> Frequencies.**
  3. In the Frequencies dialog box, select the variables that you want to create frequency table.



4. Click Frequency Tables, keep the default settings then click Continue then OK.



5. In the output viewer, frequency table will now appear.



6. Save the output via File> Export(sample file name: frequencytable.spv)

Graphical Presentation

Graphs are simply pictures of numerical data. These are use in books, newspaper, magazines and reports. These are widely used because they can present clear pictures of numerical data.

Graphical presentation has the following advantages:

1. Graphs in able readers to easily grasp essential facts that numerical data intends to convey.
2. Graphs can easily attract attention and are more readily understood.
3. Graphs can simplify concepts that would otherwise have been expressed in so many words.

Some of common kinds of graphs are the following:

- Circle Graph or pie chart
- Histogram
- Bar graph
- Line graph
- Box plot
- Stem and leaf plot

When constructing a graph it is important to take note of the following considerations:

1. Title of graphs must be clearly stated and labeled as Figure.
2. All axes should be scaled and labelled with variables they represent.
3. Sources and footnotes should be provided when necessary.

Graphical Presentation with MS Excel

Graphs can be automatically generated in MS Excel. Please be guided with the following steps:

1. Open MS Excel.
2. Select the spread sheet that contains the frequency table.
3. Highlight the cells of the frequency table that you intend to create a graph. See sample below:

Table 1. Frequency and Percentage of Respondents

Position /Designation	Frequency	Percentage
Record Officer	19	63.33%
Record Section Staff	11	36.67%
Total	30	100.00%

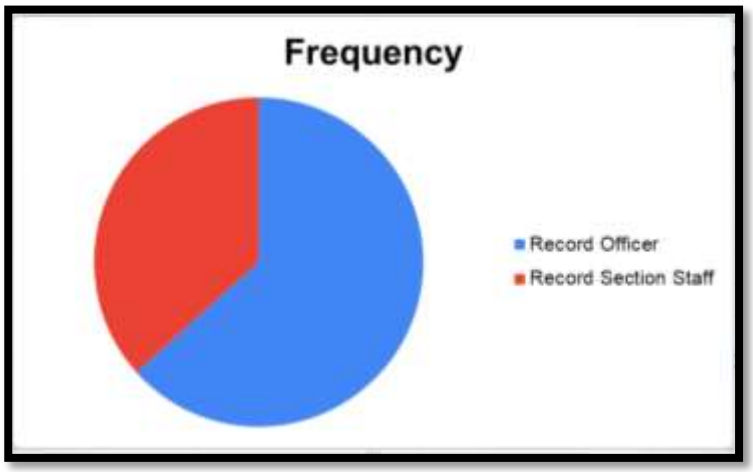
Or

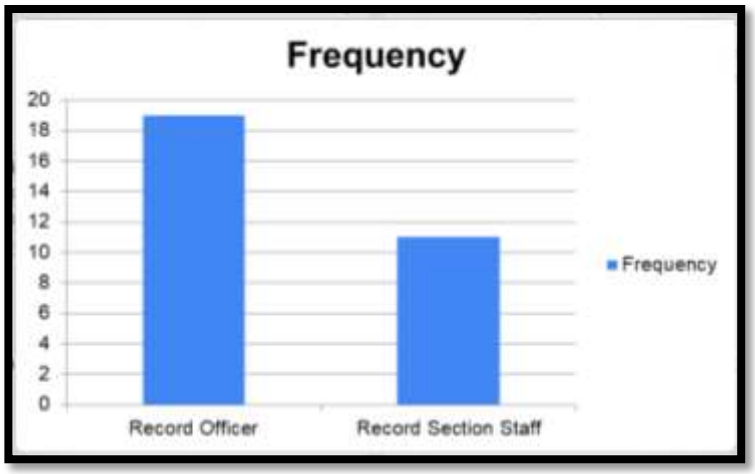
Table 1. Frequency and Percentage of Respondents

Position /Designation	Frequency	Percentage
Record Officer	19	63.33%
Record Section Staff	11	36.67%
Total	30	100.00%

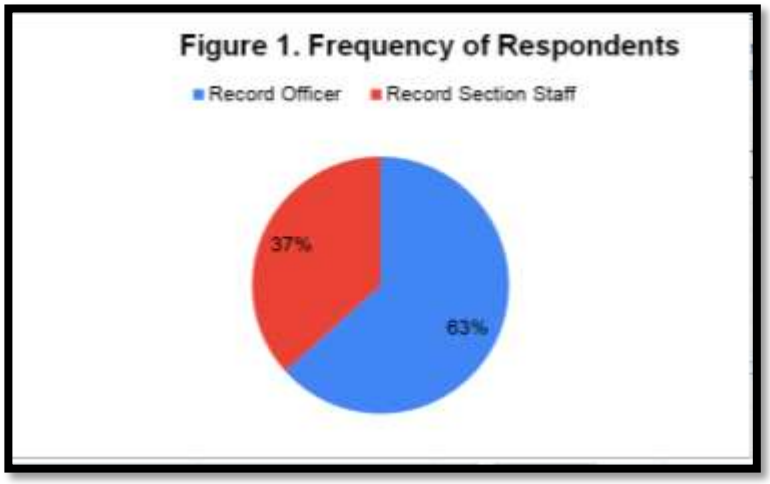
* Press ctrl key to highlight multiple non-adjacent column

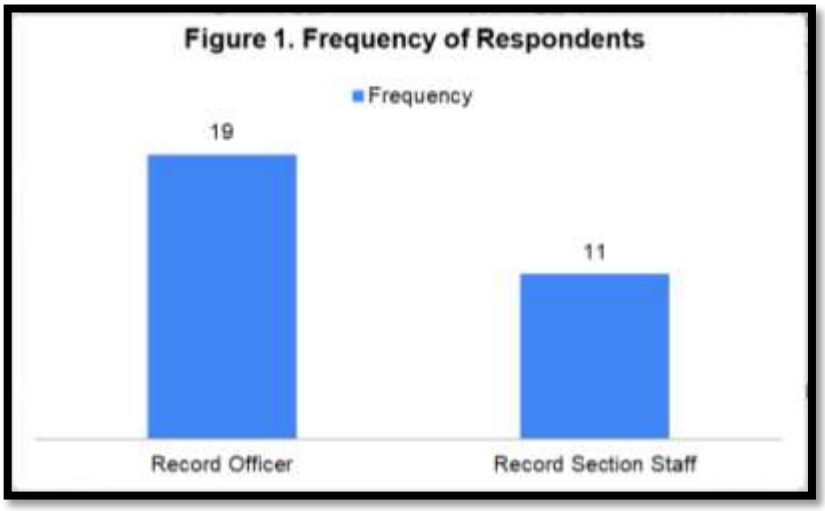
4. Click Insert tab. In the Chart group select the chart you want to use (ex. pie graph, bar graph etc.)Automatically excel will generate a chart for your highlighted data. See sample below:





5. You can edit some details of the chart such as the title and labels by customizing the chart layout. See sample below:



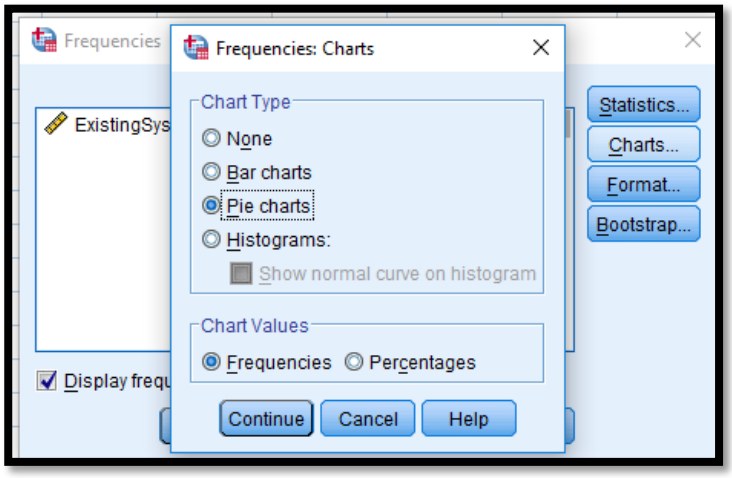


6. Save your work.

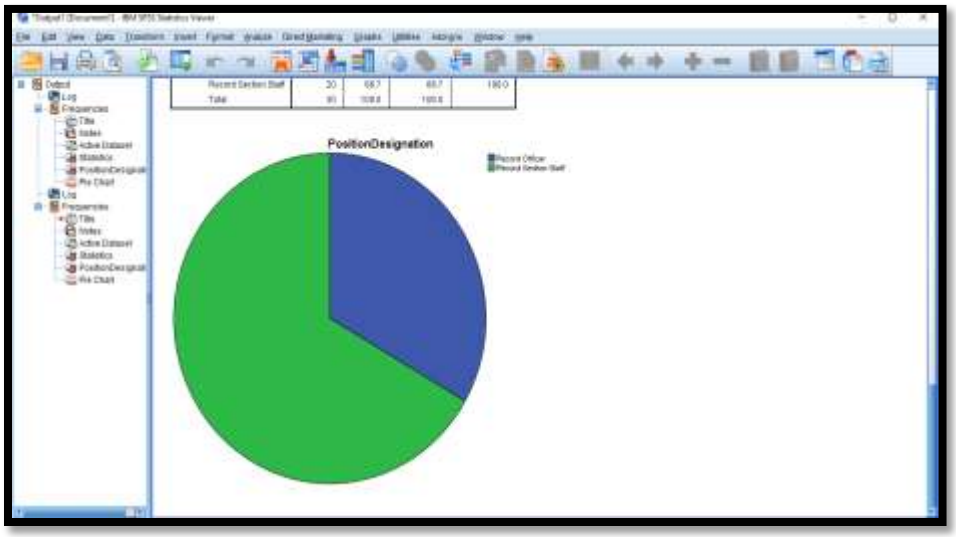
Graphical presentation with SPSS

Like in MS Excel, SPSS automatically generate also graphs or chart. Please be guided of the following steps:

1. Click Data view tab. Make sure that all the data needed for graphical presentation are already encoded.
2. Select Analyze> Discriptive Statistics> Frequencies.
3. In the Frequencies dialog box select the variables that you want to present graphically.
4. Click Charts then select the chart you want to use. See sample below:



5. Click continue then OK. The output viewer will now show the created graph. See sample below:

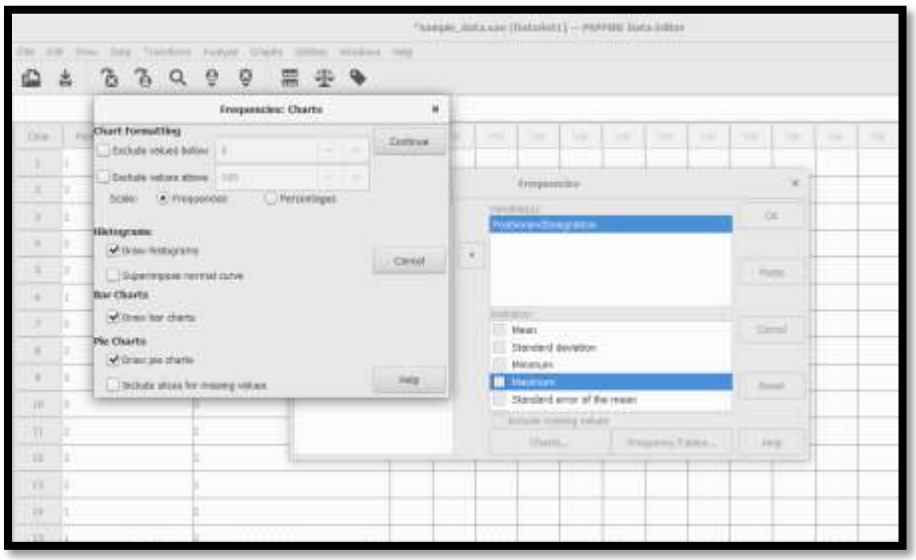


6. Save your work.

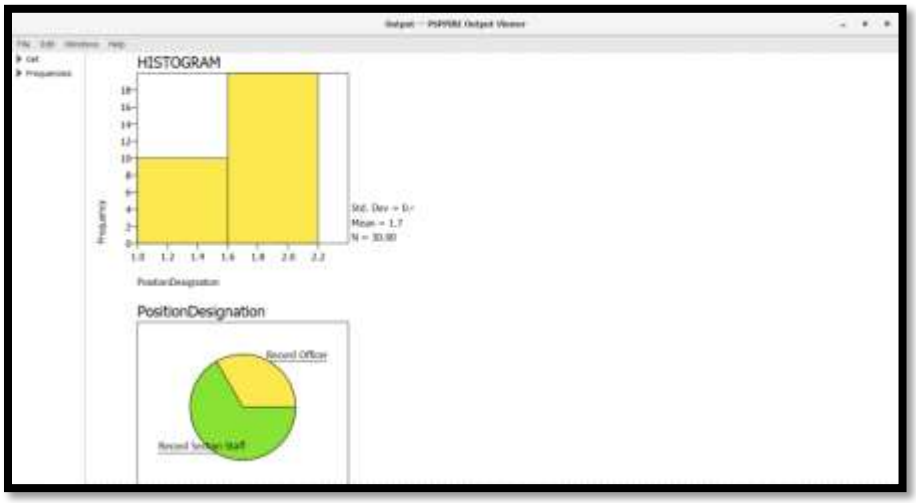
Graphical Presentation with PSPP

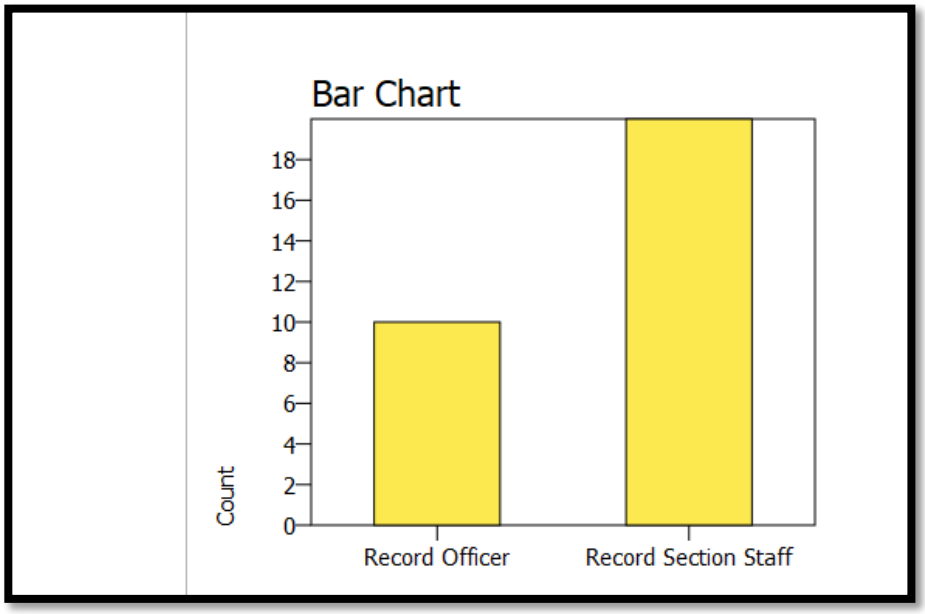
Graphs in PSPP is generated almost the same with SPSS. Please be guided with the following steps:

1. Click Data view tab. Make sure that all the data needed for graphical presentation are already encoded.
2. Select Analyze> Discriptive Statistics> Frequencies.
3. In the Frequencies dialog box select the variables that you want to present graphically.
4. Click Charts then select the chart you want to use. In PSPP you can select multiple charts or graphs at the same time. See sample below:



5. Click continue then OK. The output viewer will now show the created graph. See sample below:





6. Save your work.

****I hope you have enjoyed reading the discussion. How did you find the things you learned about data preparation and data presentation? Can you now organize data out of the survey questionnaire responses collected? If you are now ready, you may proceed answering LEARNING TASK M2.1*

References:

Daniel Muijs.2010. Doing Quantitative Research in Education with SPSS. Retrieved from https://dl.uswr.ac.ir/bitstream/Hannan/132194/1/076194382X.Sage.Publications.Ltd.Doing.Quantitative.Research.in.Education.with.SPSS.May.2004.pdf

William Trochim et.al. 2016. Research Methods: The Essential Knowledge Base. Cengage Learning

Beth M. Schwartz et.al. *2015. An Easy Guide to Research Design and SPSS. SAGE Publications, inc.*

VII. *Assignment/Activities*

LEARNING TASK M2.1

Objective: *This learning task intends to achieve objective 1, 2 and 3 of this module.*

Instruction: *Please carefully read the instruction below:*

1. Form a group of 3 to 4 members(preferably the same group with your prelim project).
2. Access the sample questionnaire response data set from this URL: https://drive.google.com/file/d/10XMscPFeMfCxgEzXu_9vDY3VcBZsCdWp/view?usp=sharing
3. With the given data set create the following:
   a. Frequency table
   b. Graph or chart(either pie or bar graph)
      For the following variables:
      ExistingSystemPerformance_Storage_ST1
      ExistingSystemPerformance_Storage_ST2
      ExistingSystemPerformance_Storage_ST3
      ExistingSystemPerformance_Storage_ST4
      ExistingSystemPerformance_Storage_ST5

      ExistingSystemPerformance_Organization_O1
      ExistingSystemPerformance_Organization_O2
      ExistingSystemPerformance_Organization_O3
      ExistingSystemPerformance_Organization_O4

ExistingSystemPerformance_Security_SE1
ExistingSystemPerformance_Security_SE2
ExistingSystemPerformance_Security_SE3
ExistingSystemPerformance_Security_SE4
ExistingSystemPerformance_Security_SE5

4. Accomplish step 3 by using either MS Excel, SPSS or PSPP whichever is preferable and available as your statistical tool.

Submission Method: Submission of the learning task will be either of the following depending on your level of technology.

- High Level: Output file containing the frequency table and graph or chart will be submitted via CSPC LeOns
- Middle Level: : Output file containing the frequency table and graph or chart will be submitted via CSPC LeOns or printed output that will be collected on designated place and time.
- Low Level: Printed output that will be collected on designated place and time.

Evaluation Criteria:
Please refer to the following score distribution.
1. *Frequency table(5 points for each variable)*
2. *Graph or chart(5 points for each variable)*

VIII. *Post-test*

This post-test will try to assess your knowledge after the discussion of the topic. Instructions are provided for each part of the test.

Item		
Part 1 True or False: Please evaluate the following statement if it expresses a correct thought or not. Please put a check on the column that corresponds to your answer.	**True**	**False**
1. **Data preparation involves checking data for accuracy.**		
2. **Descriptive statistics tests specific research hypothesis.**		
3. **As a researcher, it is very vital that you screen the collected data for accuracy.**		
4. **Google Forms has a feature of automatically organizing the collected responses.**		
5. **Response data set produced by Google Forms cannot be exported to excel spread sheet file.**		

6. **Variable tab in SPSS allows you to encode questionnaire responses.**		
7. **In SPSS, if the variables will be used as predictor the Role should be set to Input.**		
8. **SPSS can count missing data for a variable.**		
9. **Frequency of responses can be counted in MS Excel by using the formula or function COUNT.**		
10. **Same with MS Excel, SPSS can count frequency of responses by allowing end user to create formula.**		

****please answer this post-test so that you will be able to determine if there is a significant difference on your score compared to your pre-test.*

Submission of your post-test will be either of the following based on your level of technology:

High Level: submission via CSPC LeOns
Middle Level: submission via CSPC LeOnS
Low Level: Written/printed output that will be collected on designated place and time.

IX. Answer Sheet (For The Faculty Only)