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**	**: 2.11**
B.	**MODULE TITLE**	**: Inferential Statistics: Introductory Concepts and Comparison of Means**
C.	**DURATION**	**: Two (2) weeks**
D.	**MODULE PREREQUISITE**	**: Module 2.9**
E.	**INSTRUCTOR**	**: ICHELLE F. BALUIS**

II. Module Objectives (Desired Learning Outcomes)

1. Apply the model step in hypothesis testing for comparison of means.
2. Analyze and solve problems on comparison of means involving one-sample, two dependent and independent samples.
3. Analyze data and interpret results provided by statistical software.

III. Vocabulary List

- **Inferential Statistics:** The results obtain from this statistics is used as the basis for drawing broader inferences or generalization. In other words, inferential statistics is used to try to infer from the sample data to the larger population.
- **T-test:** The most commonly used method to evaluate the difference in means between two groups. It can be used for small sample sizes (n<30) and population standard deviation is not given.
- **Z-test:** Another commonly used method to evaluate the difference between the two groups. . It can be used for larger sample sizes (n>=30) and the population standard deviation is given.
- **Dependent sample:** Two sets of data taken from the same group.
- **Independent sample:** Two sets of data taken from two different groups.
- **Level of significance:** it has the maximum value of the probability of rejecting the null hypothesis.
- **p-value:** known as the probability value. It used in hypothesis testing to come up with the appropriate decision whether to accept or reject the null hypothesis. It is the evidence against a null hypothesis. The smaller the p-value, the stronger the evidence that the null hypothesis should be rejected.

IV. Learning Plan

ACTIVITY	DURATION (HH:MM)	REMARKS*
Do Pre-Test	00:15	
Study Objective 1,2 ,3, Information Sheet 1: Introductory Concept ; Information Sheet 2: Comparison of Means	03:00	
Do Objective 1,2, 3 and Learning Task M2.3	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.

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. Inferential statistics demands higher order of critical thinking skill than that of in descriptive statistics.		
2. To determine the average age of facebook users in the Philippines, requires inferential statistics.		
3. Statistically speaking, when a statistical hypothesis is being tested, it is the null hypothesis that is being tested.		
4. The 5% level of significance implies that we are 95% confident that we have made the right decision in rejecting or accepting the null hypothesis.		
5. If data or variables are nominal or ordinal, parametric tests such t-test and z-test should be used.		
6. T-test is used if the sample size is larger than 30.		
7. Z-test is used only if the population mean is given.		
8. In determining the difference between the sample mean from the two sets of data taken from the same group, paired samples t-test should be used.		
9. If the p-value is greater than the level of significance, reject the null hypothesis.		
10. With a 1% level of significance, a t-test returns a p-value of p=0.0000, the researcher should decide to accept the null hypothesis.		

VI. Information Sheet

Information Sheet 1: Inferential Statistics (Introductory Concepts)

Let us start the discussion by recalling the definition of inferential statistics given in the previous module:

- ***Inferential Statistics*** is the statistical analyses used to reach conclusions that extend beyond the immediate data alone. The result obtain from this statistics is used as the basis for drawing broader inferences or generalization. In other words, we use inferential statistics to try to infer from the sample data to the larger population. Unlike descriptive statistics, it demands higher order of critical thinking and mathematical methods. Its main purpose is to draw conclusions or judgements regarding an entire set of observation by using only a small representative of the whole set.

  With that definition, the next question would be, when are we going to use this inferential statistics in our data analyses? Inferential statistics shall be used if the purpose of your data analyses is either of the following:

  1. ***Test significant difference among variables;***
     Example: To determine the significant difference in the level of acceptability between existing Learning Management system and the proposed Learning Management System in terms of reliability.
  2. ***Test the significant relationship among variables as well as testing the strength of relationship;***
     Example: To determine if the level of efficiency of the Learning Management System is significantly related to the level of acceptability.

     and
  3. ***Make predictions out of a given data set.***
     Example: To predict the rate of employment of BSIT graduates by year 2025.

  Before going forward to the different statistical tools for inferential statistics, let us review first your knowledge on hypotheses.
  As what you've learned from Module1.4, hypothesis can be derived directly from the statement of problem or objectives. For example the research problem or objective is:

  Is there a significant relationship between the level of efficiency and level of acceptability of the new Learning Management System?

  The corresponding hypotheses can be stated as:

  Alternative Hypothesis: *There is a significant relationship between the level of efficiency and level of acceptability of the new Learning Management System.*

Null Hypothesis: *There is no significant relationship between the level of efficiency and level of acceptability of the new Learning Management System.*

To test whether you still recall these concepts, try to state alternative and null hypothesis for the following research problem or objective:

Is there a significant difference in the level of acceptability between existing Learning Management system and the proposed Learning Management System in terms of reliability?

Please write below the alternative and null hypothesis derived from the above research problem or objective:

Alternative hypothesis: _____

Null hypothesis: _____

Note: Technically speaking, when a statistical hypothesis is being tested using inferential statistics, it is the null hypothesis that is being tested. When the null hypothesis is tested, it is either accepted or rejected. When it is rejected, the statement that is accepted is the alternative hypothesis.

In addition, in hypothesis testing the researcher shall consider the following model steps:

1. ***Null (H_o) and alternative (H_a) hypotheses:*** *state first the null and alternative hypotheses based from the objective.*
2. ***Level of Significance:*** *it is an important input to hypothesis testing, however, modern textbooks in statistics did not discuss thoroughly as to how the level of significance be chosen for sound statistical inference. Despite of this, the customary level of significance used is either is 10%, 5% or 1%. A 10% level of significance implies that we are 90% confident that we have made the right decision. A 5% level of significance implies that we are 95% confident that we have made the right decision. Similarly, a 1% level of significance implies that we are 99% confident that we have made the right decision.*
3. ***Statistical Test:*** *decide what statistical test to use. The next section will discuss what are the factors to consider when choosing the appropriate statistical test.*
4. ***Computation:*** *this can be done manually or through the use of statistical software like PSPP or SPSS*
5. ***Conclusion:*** *The interpretation of result. Whether to accept or reject the null hypothesis.*

Before we will go further into data analysis, let us consider first the factors necessary to consider in the choice of appropriate statistical tools:

1. ***Level of Measurement used in measuring the variables;***
    a. *If data or variables are **either interval or ratio**, parametric tests are used.*
        i. *To test the significant difference between variables, either **t-test, z-test or f-test** is used. In this module, discussion includes only on how to perform t-test and z-test.*

            ***T-test:** this is used if the sample size is less than 30 **(n<30)** and the population standard deviation is not given, only the sample standard deviation. There are two types of t-test:*

                ***(1) one sample t-test:** used to determine the significance of the difference between sample mean from a population mean (standard mean)*
                ***(2) two-sample t-test:** this can be either (1) **dependent sample t-test** or (2) **independent sample t-test**. Dependent sample t-test means two sets of data were taken from the same group while independent sample t-test means two sets of data were taken from two distinct different groups.*

            ***Z-test:** this is used if the sample size is more than or equal to 30 **(n>=30)** and the population standard deviation is given.*

        ii. *To test the dependency or relationship among variables , **Pearson'r** is used. This will be discussed further in the succeeding module.*

    b. *If data or variables are **categorical (nominal or ordinal)**, non-parametric tests are used.*
        i. *To test the significant difference between variables, either **Mann-Whitney U Test, Wilcoxon Matched Pairs, or Kruskal-Wallis Test** is used. In this module, discussion includes only on how to perform t-test and z-test.*
        ii. *To test the dependency or relationship among variables , **Spearman's r, Chi-square, Fischer's Exact or Kolmogorov-Smirnov test** is used. Spearmans'r and Chi-square will be discussed in the succeeding modules.*

2. ***Objectives of the study;***
    If the objective is to determine the significant difference between variables, the above mentioned parametric and non-parametric tests for testing significant difference between variables shall be used. On the other hand, if the objective is to determine the significant relationship between variables, the above mentioned parametric and non-parametric tests for testing significant relationship between variables shall be used.

and

3. ***Design used in the study***
   If the design is more on experimental, inferential statistics for data analysis is recommended. On the other hand, if the design is descriptive, then descriptive statistics for data analysis is suggested.

Information Sheet 2: Inferential Statistics (Comparison of Means)

T-Test (One Sample Analysis)

Example: *Suppose the researcher wishes to determine if there is a significant difference on the weight of 25 Grade IV pupils from that of the health record. If the mean weight of the sample is 31.88 kg with standard deviation of 1.76kg while that of the record is 34 kg, can the researcher claim that the sample is significantly lighter than those of in the record? Use 5% level of significance. Apply the steps in hypothesis testing. The following table shows the weight of the 25 pupils in the sample.*

Weight of Grade IV Pupils in Kilograms (n=25)				
30	30	30	30	30
29	29	29	29	29
35	35	35	35	35
33.5	33.5	33.5	33.5	33.5
29.5	29.5	29.5	29.5	29.5

Step 1: Null and alternative hypothesis

H_o: *The sample is not significantly lighter than those of in the health record*

H_a: *The sample is significantly lighter than those of in the health record.*

Step2: Level of Significance

Set 5% as the level of significance (a=.05)

Step3: Statistical Test

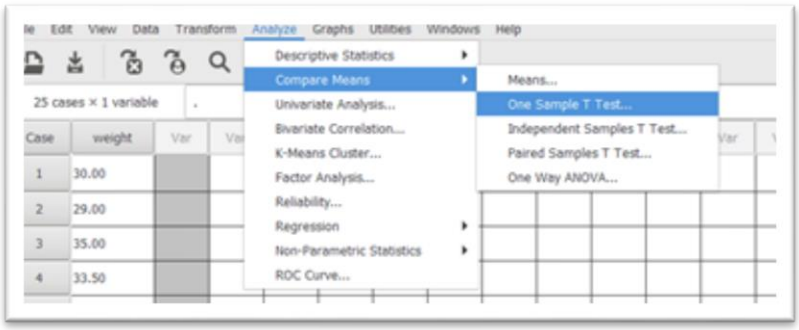
SD (standard deviation) is derived from the sample, and the sample is less than 30 therefore ***t-test*** *is the appropriate statistical test to be used. Since the problem requires determining the difference between the sample mean and population mean out of a one-sample case,* ***one-sample t-test should be used****.*

Step 4: Computation *(shown are the steps on how to do one-sample t-test using PSPP)*

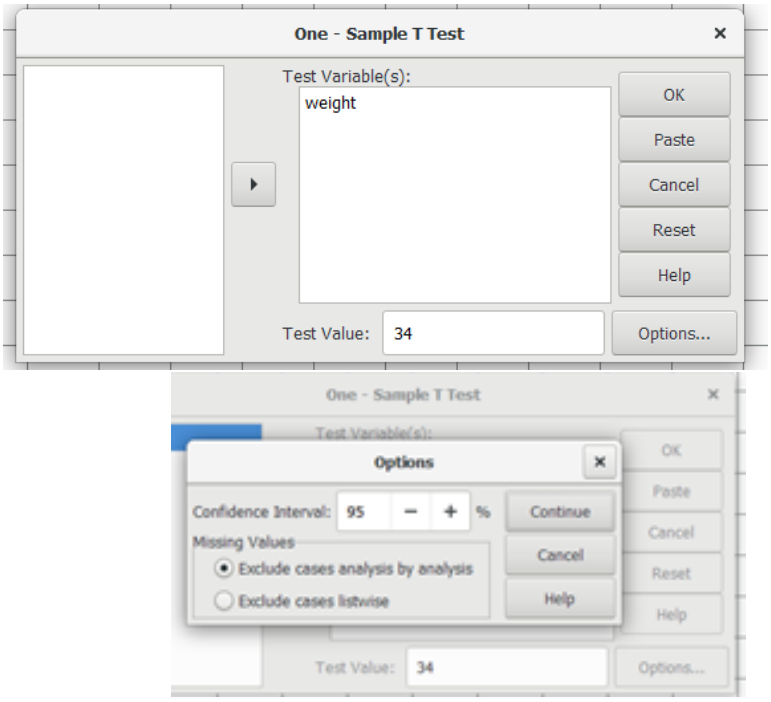
1. *In the Variable View, set up the variable. You can name the variable as Weight (review your knowledge on setting up variable in PSPP, as discussed in the previous module)*

Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1	weight	Numeric	8	2		None	None	8	Right	Scale	Input

2. *Go to Data View, input the data (weights of 25 pupils).*
3. *Click Analyze>Compare Means>One Sample T-Test.*



4. *Click weight variable and add it to Test Variable box. For Test Value box, input the population mean, as given in the problem, it is 34kg. Click Options then set Confidence Interval to 95%. Click OK.*



5. *Output Screen will display the following result.*

```
T-TEST /TESTVAL=34
      /VARIABLES= weight  /MISSING=ANALYSIS
      /CRITERIA=CI(0.95).
```

One-Sample Statistics

	N	Mean	Std. Deviation	S.E. Mean
weight	25	31.88	1.76	.35

One-Sample Test

	Test Value = 34					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
weight	-6.03	24	.000	-2.12	-2.85	-1.39

Step 5: Conclusion

To be able to come up for a conclusion, all you have to do is to take a look at the value of the highlighted column below based on the one-sample T-test result. The highlighted value is also known as the p-value.

One-Sample Test

	Test Value = 34					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
weight	-6.03	24	.000	-2.12	-2.85	-1.39

Rule: *If the p-value is less than or equal to level of significance, reject the null hypothesis, otherwise, accept the null hypothesis.*

Conclusion/Interpretation: *The above result shows that p=.000, it is less than a=.05, therefore,* ***null hypothesis should be rejected.*** *It can be said that the researcher can claim that the sample is significantly lighter than those of in the record.*

T-Test (Two Sample Analysis-Paired Sample T-Test)

Example: *Faculty members of XYZ University handling research subjects underwent a training workshop on utilization of statistical software for data analyses. Before the training proper, the 23 participants were required to take a 25 item pre-test. After the training, the same 25 item tests were given to the same participants as post-test. Having the given data below, can you determine if there is a significant difference in the pre-test and post-test results of the participants? Set the level of significance to 5%. Apply the hypothesis testing steps.*

Participant	*Pre-Test*	*Post-Test*
1	*10*	*16*
2	*5*	*17*
3	*8*	*16*
4	*13*	*20*
5	*17*	*24*
6	*5*	*15*
7	*8*	*18*
8	*13*	*22*
9	*15*	*24*
10	*19*	*23*
11	*5*	*17*
12	*11*	*21*
13	*12*	*23*
14	*18*	*19*
15	*7*	*15*
16	*13*	*18*
17	*10*	*18*
18	*11*	*19*
19	*20*	*25*
20	*15*	*20*
21	*18*	*24*
22	*13*	*25*
23	*10*	*19*

Step 1: Null and alternative hypothesis

H_o: *There is no significant difference in the pre-test and post-test result of the participants on the conducted training workshop on statistical software utilization.*

H_a: *There is a significant difference in the pre-test and post-test result of the participants on the conducted training workshop on statistical software utilization.*

Step2: Level of Significance

Set 5% as the level of significance (a=.05)

Step3: Statistical Test

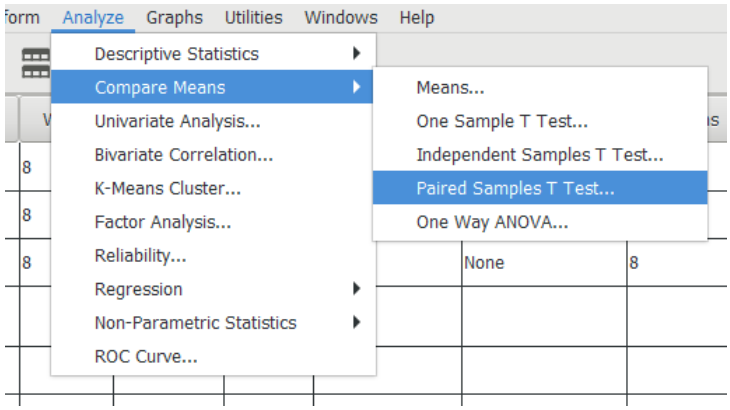
SD (standard deviation) can be derived from the sample, and the sample is less than 30 therefore ***t-test*** *is the appropriate statistical test to be used. Since the problem requires determining the difference between the sample mean from two sets of data taken from the same group,* ***two-sample t-test for dependent samples should be used****. In PSPP it is known as* ***Paired Samples T Test.***

Step 4: Computation *(shown are the steps on how to do* ***Paired Samples T Test*** *using PSPP)*

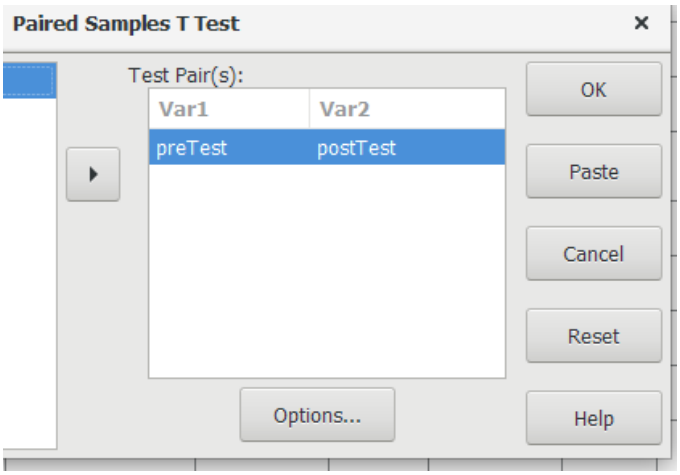
1. *In the Variable View, set up the variable. You can name the variable as preTest and postTest(review your knowledge on setting up variable in PSPP, as discussed in the previous module)*

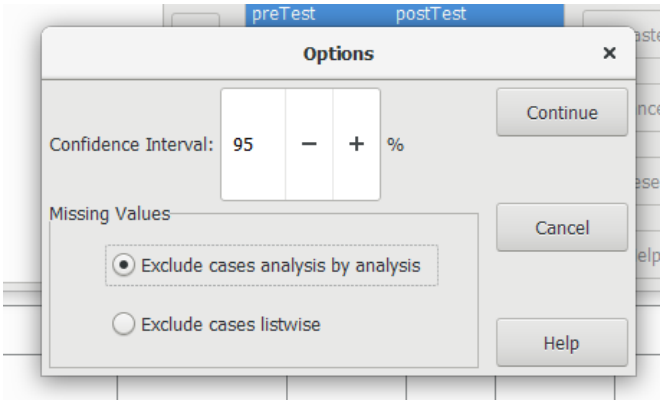
Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
2	preTest	Numeric	8	2		None	None	8	Right	Scale	Input
3	postTest	Numeric	8	2		None	None	8	Right	Scale	Input

2. *Go to Data View, input the data (pre-test and post-test scores of 23 participants).*
3. *Click Analyze>Compare Means>Paired Samples T-Test.*



4. *Click preTest variable and add it to Test Pairs box under Var1 column. Click postTest variable and add it to Test Pairs box under Var2 column. Click Options then set Confidence Interval to 95%. Click OK.*





5. *Output Screen will display the following result.*

```
T-TEST
     PAIRS =  preTest WITH  postTest (PAIRED)
     /MISSING=ANALYSIS
     /CRITERIA=CI(0.95).
```

Paired Sample Statistics

		N	Mean	Std. Deviation	S.E. Mean
Pair 1	preTest	23	12.00	4.52	.94
	postTest	23	19.91	3.27	.68

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	preTest & postTest	23	.786	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	S.E. Mean	Lower	Upper			
Pair 1	preTest - postTest	-7.91	2.81	.59	-9.13	-6.70	-13.50	22	.000

Step 5: Conclusion

To be able to come up for a conclusion, all you have to do is to take a look at the value of the highlighted column below based on the paired samples T-test result. The highlighted value is also known as the p-value.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	S.E. Mean	Lower	Upper			
Pair 1	preTest - postTest	-7.91	2.81	.59	-9.13	-6.70	-13.50	22	.000

Rule: *If the p-value is less than or equal to level of significance, reject the null hypothesis, otherwise, accept the null hypothesis.*

Conclusion/Interpretation: *The above result shows that p=.000, it is less than a=.05, therefore,* ***null hypothesis should be rejected.*** *It can be said that there is a significant difference in the pre-test and post-test result of the participants on the conducted training workshop on statistical software utilization.*

T-Test (Two Sample Analysis-Independent Samples T-Test)

Example: *Two teaching methods were used by the researcher in comparing the performance of the two groups of students in Chemistry. One group were exposed to simulation tools in Chemistry (Method A) and another group were not exposed to simulation tool (Method B). The results of the 100 item performance examination are shown below. Does the two groups' performance significantly differ? Set the level of significance to 5%. Apply the hypothesis testing steps.*

Examination Performance of Students under Method A	*Examination Performance of Students under Method B*
90	98
93	96
94	88
95	97
96	83
93	89
94	97
94	96
91	88
90	95

Step 1: Null and alternative hypothesis

H_o: *There is no significant difference in the examination performance of students who are exposed to simulation tool in Chemistry and those who are not.*

H_a: *There is a significant difference in the examination performance of students who are exposed to simulation tool in Chemistry and those who are not.*

Step2: Level of Significance

Set 5% as the level of significance (a=.05)

Step3: Statistical Test

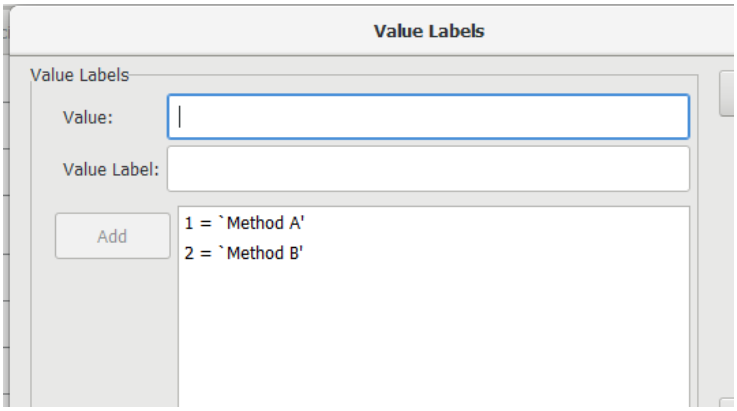
*SD (standard deviation) can be derived from the sample, and the sample is less than 30 therefore **t-test** is the appropriate statistical test to be used. Since the problem requires determining the difference between the sample mean from two sets of data taken from different groups, **two-sample t-test for independent samples should be used**. In PSPP it is known as **Independent Samples T Test.***

***Step 4: Computation** (shown are the steps on how to do **Independent Samples T Test** using PSPP)*

1. *In the Variable View, set up the variable. In Independent Samples T-Test, we have to set up the **Test Variable** and the **Grouping Variable**. Based from the problem, the Test Variable is the exam performance of students and the Groupings Variable is the method. You can name the variable as examPerformance and method (review your knowledge on setting up variable in PSPP, as discussed in the previous module)*

Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
4	method	Numeric	8	0		{1, Method A}...	None	8	Right	Nominal	Input
5	examPerformance	Numeric	8	2		None	None	8	Right	Scale	Input

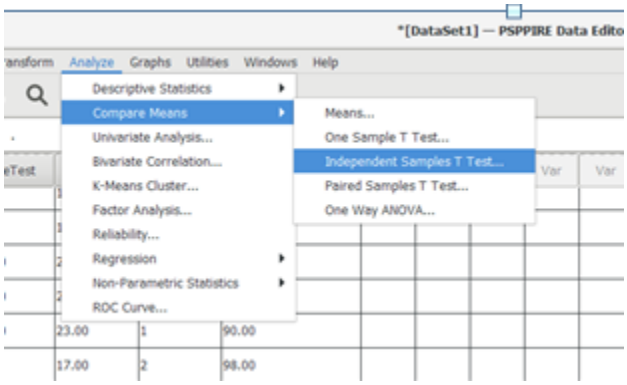
Set up also the value labels for method variable.



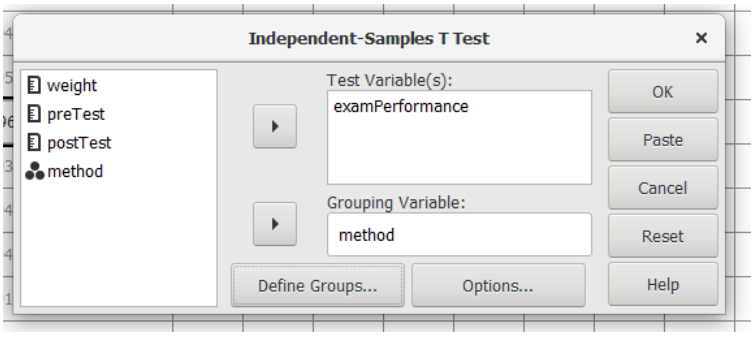
2. *Go to Data View, input the data (methods and the exam performance of 10 students for each group).Sample data view shown on the following page.*

method	examPerformance
1	90.00
1	93.00
1	94.00
1	95.00
1	96.00
1	93.00
1	94.00
1	94.00
1	91.00
1	90.00
2	98.00
2	96.00
2	88.00
2	97.00
2	83.00
2	89.00
2	97.00
2	96.00
2	88.00
2	95.00

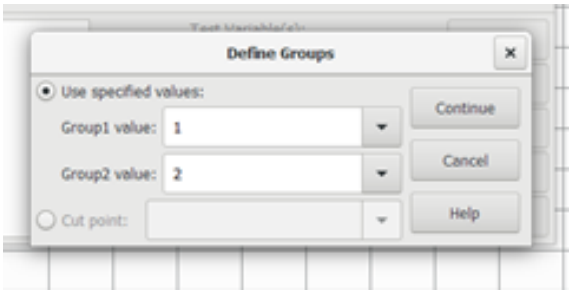
3. *Click Analyze>Compare Means>Independent Samples T-Test.*



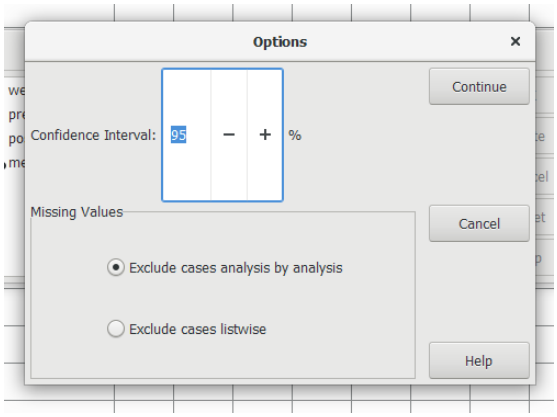
4. *Click examPerformance variable and add it to Test Variable box. Click method variable and add it to Grouping variable box.*



5. *Click Define Groupings then set it up as follows.*



6. *Click Options then set Confidence Interval to 95%. Click Continue then OK.*



7. *Output Screen will display the following result.*

```
T-TEST /VARIABLES= examPerformance
    /GROUPS=method(1,2)    /MISSING=ANALYSIS
    /CRITERIA=CI(0.95).
```

Group Statistics

	Group	N	Mean	Std. Deviation	S.E. Mean
examPerformance	Method A	10	93.00	2.05	.65
	Method B	10	92.70	5.21	1.65

Independent Samples Test

		Levene's Test for Equality of Variances		T-Test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
examPerformance	Equal variances assumed	16.24	.001	.17	18.00	.867	.30	1.77	-3.42	4.02
	Equal variances not assumed			.17	11.74	.868	.30	1.77	-3.57	4.17

Step 5: Conclusion

To be able to come up for a conclusion, all you have to do is to take a look at the value of the highlighted column below based on the independent samples T-test result. The highlighted value is also known as the p-value.

Independent Samples Test

		Levene's Test for Equality of Variances		T-Test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
examPerformance	Equal variances assumed	16.24	.001	.17	18.00	.867	.30	1.77	-3.42	4.02
	Equal variances not assumed			.17	11.74	.868	.30	1.77	-3.57	4.17

Rule: *If the p-value is less than or equal to level of significance, reject the null hypothesis, otherwise, accept the null hypothesis.*

Conclusion/Interpretation: *The above result shows that p=.867, it is greater than a=.05, therefore,* ***null hypothesis should be accepted.*** *It can be said that there is no significant difference in the examination performance of students who are exposed to simulation tool in Chemistry and those who are not.*

Z-Test (One Sample Analysis)

Example: *Data from the survey shows that the average age of users of social media website in the Philippines is 29 years old, with a standard deviation of .75 year old. A sample of 40 users of social networking site facebook was obtained with a mean of 27 years old. Is the average age of sample facebook users are different from that of Philippines population of social media users? Set the level of significance to 5%. Apply the hypothesis testing steps.*

Step 1: Null and alternative hypothesis

H_o: *There is no significant difference in the average age of sample facebook users and the average age of Philippines population of social media users.*

H_a: *There is a significant difference in the average age of sample facebook users and the average age of Philippines population of social media users.*

Step2: Level of Significance

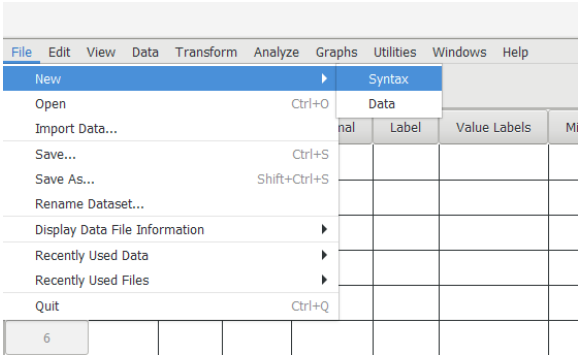
Set 5% as the level of significance (a=.05)

Step3: Statistical Test

SD (standard deviation) from the population is given, and the sample size is greater than 30 therefore ***z-test*** *is the appropriate statistical test to be used. Since the problem requires determining the difference between the sample mean from that of the population mean,* ***one-sample z-test should be used****.*

Step 4: Computation *(PSPP and SPSS doesn't provide a menu-driven steps of performing z-test computation, instead it can be done by working directly with syntax. Shown are the steps on how to do* ***z-test*** *using PSPP)*

1. *Open a new PSPP data editor window. Go to File>New>Syntax*



2. *In the PSPP Syntax Editor , copy and paste the following codes:*

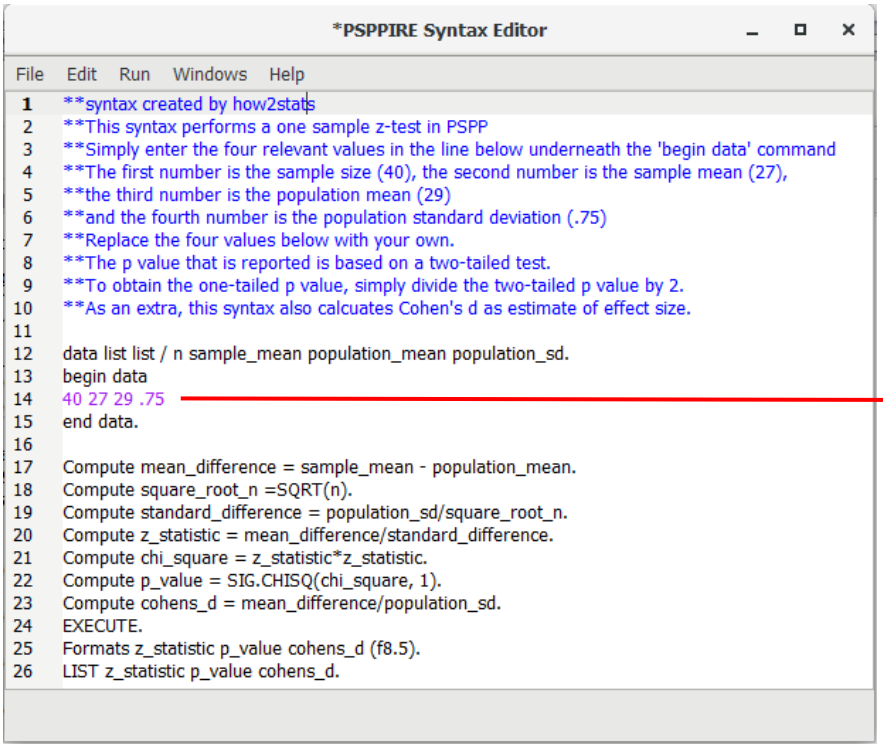
```
**syntax created by how2stats
**This syntax performs a one sample z-test in PSPP
**Simply enter the four relevant values in the line below underneath the 'begin data' command
**The first number is the sample size (40), the second number is the sample mean (27),
**the third number is the population mean (29)
**and the fourth number is the population standard deviation (.75)
**Replace the four values below with your own.
**The p value that is reported is based on a two-tailed test.
**To obtain the one-tailed p value, simply divide the two-tailed p value by 2.
**As an extra, this syntax also calcuates Cohen's d as estimate of effect size.

data list list / n sample_mean population_mean population_sd.
begin data
40 27 29 .75
end data.

Compute mean_difference = sample_mean - population_mean.
Compute square_root_n =SQRT(n).
Compute standard_difference = population_sd/square_root_n.
Compute z_statistic = mean_difference/standard_difference.
Compute chi_square = z_statistic*z_statistic.
Compute p_value = SIG.CHISQ(chi_square, 1).
Compute cohens_d = mean_difference/population_sd.
EXECUTE.
Formats z_statistic p_value cohens_d (f8.5).
LIST z_statistic p_value cohens_d.
```

****Code reference: http://www.how2stats.net/2014/03/one-sample-z-test.html*

The syntax editor should look like this:



Based from the given problem:
40 is the sample size;
27 is the sample mean;
29 is the population mean;
.75 is the population standard deviation

3. *Click Run>All. Output Screen will display the following result.*

Data List

z_statistic	p_value	cohens_d
-16.8655	.00000	-2.66667

Step 5: Conclusion

To be able to come up for a conclusion, all you have to do is to take a look at the value of the highlighted column below based on the one sample z-test result. The highlighted value is also known as the p-value.

Data List

z_statistic	p_value	cohens_d
-16.8655	.00000	-2.66667

Rule: *If the p-value is less than or equal to level of significance, reject the null hypothesis, otherwise, accept the null hypothesis.*

Conclusion/Interpretation: *The above result shows that p=.00000, it is less than a=.05, therefore,* ***null hypothesis should be rejected.*** *It can be said that there is a significant difference in the average age of sample facebook users and the average age of Philippines population of social media users.*

Z-Test (Two Sample Analysis)

Example: *Suppose the researcher wishes to determine the significant difference on the performance of students in Programming subject with those under high technology(with internet and with gadget) and those under low level technology(without internet and no gadget). The table below shows the midterm grades of students under high level and and low level technology. Use 5% level of significance.*

Midterm Grade of Students under Low Level			*Midterm Grade of Students under High Level*		
2.2	*2.5*	*2.6*	*2.1*	*1.3*	*1.5*
2.4	*2.1*	*2.1*	*2.4*	*1.4*	*2.6*
2.5	*2.0*	*2.2*	*2.6*	*1.5*	*2.7*
3.0	*2.75*	*2.5*	*2.9*	*2.1*	*2.6*
2.75	*2.6*	*2.0*	*3.0*	*2.0*	*2.75*
2.9	*2.1*	*3.0*	*3.0*	*1.9*	*1.3*
2.7	*1.3*	*2.1*	*2.9*	*1.7*	*1.5*
1.5	*1.5*	*1.4*	*2.8*	*2.2*	*1.6*
2.3	*1.8*	*1.5*	*2.1*	*2.75*	*2.1*
2.0	*1.9*	*1.8*	*2.6*	*2.6*	*2.0*
2.1	*2.0*	*1.7*	*2.4*	*2.3*	*2.3*

Step 1: Null and alternative hypothesis

H_o: *There is no significant difference in the midterm grades of Programming students under high level technology and low level technology.*

H_a: *There is no significant difference in the midterm grades of Programming students under high level technology and low level technology.*

Step2: Level of Significance

Set 5% as the level of significance (a=.05)

Step3: Statistical Test

The sample size is greater than 30 and since the sample is large, it is enough to assume that SD or variance can be derived from the population, therefore ***z-test*** *is the appropriate statistical test to be used. Since the problem requires determining the difference between two independent samples,* ***two-sample z-test should be used****.*

Step 4: Computation *(PSPP and SPSS doesn't provide a menu-driven steps of performing two-sample z-test computation, nor it can be done by working directly with syntax, but good thing it is available in MS-Excel. Shown are the steps on how to do two-sample* ***z-test*** *using MS-Excel)*

1. *Open MS-Excel. Encode the data as follows:*

	Low Level Tech	High Level Tech
2	2.2	2.1
3	2.4	2.4
4	2.5	2.6
5	3.0	2.9
6	2.8	3.0
7	2.9	3.0
8	2.7	2.9
9	1.5	2.8
10	2.3	2.1
11	2.0	2.6
12	2.1	2.4
13	2.5	1.3
14	2.1	1.4
15	2.0	1.5
16	2.8	2.1
17	2.6	2.0
18	2.1	1.9
19	1.3	1.7
20	1.5	2.2
21	1.8	2.8
22	1.9	2.6
23	2.0	2.3
24	2.6	1.5
25	2.1	2.6
26	2.2	2.7
27	2.5	2.6
28	2.0	2.8
29	3.0	1.3
30	2.1	1.5
31	1.4	1.6
32	1.5	2.1
33	1.8	2.0
34	1.7	2.3

2. *Compute the variance of the two samples (low level tech and high level tech). Select the cell where the computed variance will be placed. In the cell create a formula by using the built-in formula for computing the variance then press enter key. Shown below is the sample formula for variance computation.*

Low Level Tech	High Level Tech		Var_Low	Var_Hi
2.2	2.1		=VAR(A2:A34)	
2.4	2.4			
2.5	2.6			
3.0	2.9			
2.8	3.0			
2.9	3.0			

Low Level Tech	High Level Tech	Var_Low	Var_High
2.2	2.1	0.219863	=VAR(B2:B34)
2.4	2.4		
2.5	2.6		

Computed Variance:

Var_Low	Var_High
0.2199	0.2781

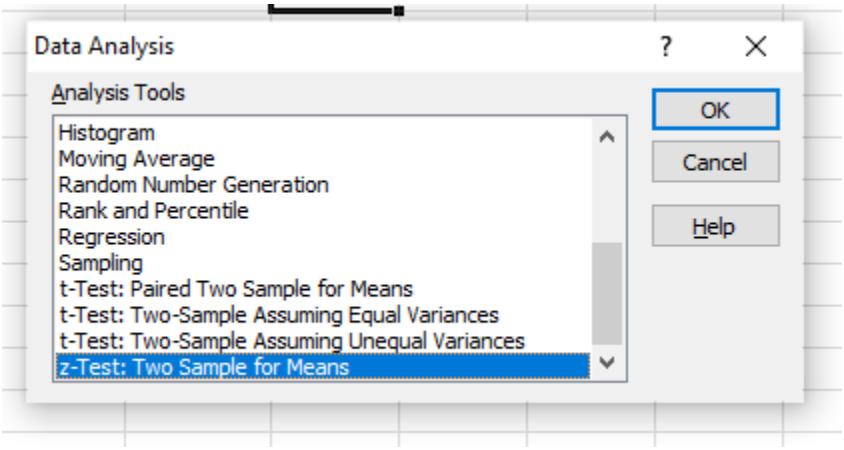
3. *Click Data Tab, then click Data Analysis.*
** If the Data Analysis command is not available, you need to load the Analysis ToolPak add-in program through the following steps:

    1. Click the **File** tab, click **Options**, and then click the **Add-Ins** category.

        If you're using Excel 2007, click the **Microsoft Office Button**  , and then click **Excel Options**
    2. In the **Manage** box, select **Excel Add-ins** and then click **Go**.

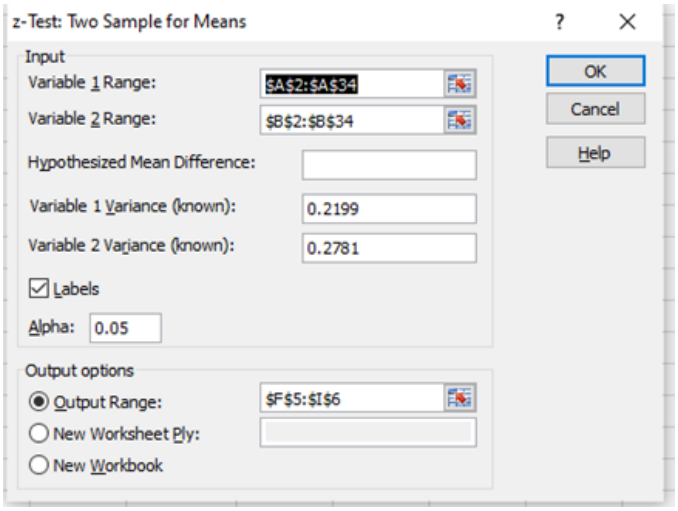
        If you're using Excel for Mac, in the file menu go to **Tools** > **Excel Add-ins.**
    3. In the **Add-Ins** box, check the **Analysis ToolPak** check box, and then click **OK**.
        - If **Analysis ToolPak** is not listed in the **Add-Ins available** box, click **Browse** to locate it.
        - If you are prompted that the Analysis ToolPak is not currently installed on your computer, click **Yes** to install it.

4. *In the Data Analysis dialog box, select* ***z-Test: Two Sample for Means.****Then click OK.*



5. *In the z-Test Two Sample for Means dialog box, set up the following details:*

Variable 1 Range: *(highlight the range of cells containing the data set for low level technology)*
Variable 2 Range: *(highlight the range of cells containing the data set for high level technology)*
Hypothesized Mean Difference*: (just leave it blank)*
Variable 1 Variance: *(input the computed variance for low level technology data set)*
Variable 2 Variance: *(input the computed variance for high level technology data set)*
Alpha: *(set to .05)*
Labels: *(check the box)*
Output Range: *(highlight the cell where you want the output will be displayed.*



Click OK once done.

6. *Output will display the following result.*

z-Test: Two Sample for Means		
	Low Level Tech	*High Level Tech*
Mean	2.175	2.23125
Known Variance	0.2199	0.2781
Observations	32	32
Hypothesized Mean D	0	
z	-0.450902709	
P(Z<=z) one-tail	0.326029835	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.652059671	
z Critical two-tail	1.959963985	

Step 5: Conclusion

To be able to come up for a conclusion, all you have to do is to take a look at the value of the highlighted row below based on the two sample z-test result. The highlighted value is also known as the p-value. The p-value for two-tail test will be the basis for conclusion since the hypothesis that is being tested is a two-tailed hypothesis (recall your knowledge on two-tailed hypothesis in Module 1.4).

z-Test: Two Sample for Means		
	Low Level Tech	*High Level Tech*
Mean	2.175	2.23125
Known Variance	0.2199	0.2781
Observations	32	32
Hypothesized Mean D	0	
z	-0.450902709	
P(Z<=z) one-tail	0.326029835	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.652059671	
z Critical two-tail	1.959963985	

Rule: *If the p-value is less than or equal to level of significance, reject the null hypothesis, otherwise, accept the null hypothesis.*

Conclusion/Interpretation: *The above result shows that p=0.652059671, it is greater than a=.05, therefore,* ***null hypothesis should be accepted.*** *It can be said that there is no significant difference in the midterm grades of Programming students under high level technology and low level technology.*

****I hope you have enjoyed reading the discussion. How did you find the things you learned about concepts of inferential statistics and comparison of means? Can you now do the steps of hypothesis testing on your own? If you are now ready, you may proceed answering LEARNING TASK M2.3*

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 M2.3

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

Instruction: *Applying the steps for hypothesis testing, solve the following problems (under Step 4, no need to show the step by step procedure, just show the result):*

1. Results on the study of the differences of grades of students in statistics who are not exposed and not exposed to the different statistical software are shown in the following table. Determine whether or not there is a significant difference in the sample mean grades of the students at 5% level of significance.

Students exposed to Statistical Software		Students who are not exposed to Statistical Software	
86	96	95	81
85	85	82	82
94	98	81	88
83	85	88	83
90	92	85	84
96	81	83	84
98	80	82	85
85	96	81	88
91	95	92	89
82	98	89	87

2. A group of 15 new masteral students took an entrance examination as requirement for admission to any masteral program. From the records, the mean rating of students taking masteral entrance examination is 88.50. Can the researcher claim that the new entrants to masteral education whose ratings are shown in the following table are significantly lower? Use 5% level of significance.

Entrance examination ratings of masteral students (n=15)		
89	91	89
88	98	93
90	81	85
85	80	86
82	87	87

3. The average zinc concentration recovered from the sample of zinc measurements in 35 different locations was found to be 2.9 grams/millilitre with a standard deviation of 0.3 gram/millilitre. Is there a reason to believe that the average water pollutant present in the river is more than that of the record which is 2.5 grams/millilitre at 0.05 level of significance.

Note: this Learning Task shall be accomplished by group (prelim project group)

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

- High Level: Output file will be submitted via CSPC LeOns
- Middle Level: : Output file 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 based on the accomplished step in hypothesis testing.

Step1: 5pts (3 pts deduction for inappropriate hypothesis statement)
Step2: 2 pts (1 pt deduction for inappropriate level of significance)
Step3: 3 pts (2 pts deduction for inappropriate statistical test)
Step4: 10 pts(8 pts deduction for incorrect result)
Step5: 5 pts (3 pts deduction for incorrect conclusion)

VIII. *Post-test*

This post-test will try to assess your knowledge after the discussion of the topic.

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. Inferential statistics demands higher order of critical thinking skill than that of in descriptive statistics.		
2. To determine the average age of facebook users in the Philippines, requires inferential statistics.		
3. Statistically speaking, when a statistical hypothesis is being tested, it is the null hypothesis that is being tested.		
4. The 5% level of significance implies that we are 95% confident that we have made the right decision in rejecting or accepting the null hypothesis.		
5. If data or variables are nominal or ordinal, parametric tests such		

t-test and z-test should be used.		
6. T-test is used if the sample size is larger than 30.		
7. Z-test is used only if the population mean is given.		
8. In determining the difference between the sample mean from the two sets of data taken from the same group, paired samples t-test should be used.		
9. If the p-value is greater than the level of significance, reject the null hypothesis.		
10. With a 1% level of significance, a t-test returns a p-value of p=0.0000, the researcher should decide to accept the null hypothesis.		

****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: online text submission via CSPC LeOns

Middle Level: online text submission via CSPC LeOnS or written/printed output that will be collected on designated place and time

Low Level: Written/printed output that will be collected on designated place and time.

IX. Answer Sheet (For The Faculty Only)