redesigning
Bounce back as
Persistent and
Sustainable
University

Course Packet 04

LM01-NGEC 0423

Learning Module

Mathematics in the Modern World

Course Packet 04

Mathematics as a Tool

Knowledge Area Code	:	**MATH**
Course Code	:	**NGEC 0423**
Learning Module Code	:	**LM-MATH0423**
Course Packet Code	:	**LM-MATH0423-04**

Course Packet 04

Mathematics as a Tool

Introduction

In scientific research, we develop theories and then conduct experiments to test them. Since we cannot achieve perfect control in any experiments, some chance variations always results. For example, even the most precise measurement systems has some limits to its accuracy, and the laboratory mice breed for biological research to be as similar as possible as if not completely identical. Also, there is a limit to the number of times observations can be repeated, so our measurements are always only a sample of reality.

To generalize from our experimental sample, we use statistical techniques to test how big a part chance plays in the outcome of anexperiment. We begin by assuming that our experimental results reflect in the random variations caused by assorted factors beyond our control. This assumption is called the *null hypothesis*. If our experiment is successfuland our theory is true, we will be able to reject the null hypothesis by showing that chance variation is not a reasonable explanation in our results.
.

Objectives

- State a null hypothesis and alternative and establish a critical reasoning for theories about P.
- Use a variety of statistical tools to process and manage numerical data
- Apply the steps in hypothesis testing
- Judge the probability of accepting a theory when it is false or reject a theory when it is true.

Learning Management System

You will be requested to join the Google Classroom. A link will be sent in your g-mail account.

Duration

- *Topic 01:Hypothesis Testing* = 9 hours
  (7 hours self-directed learning with practical exercises and 2 hours assessment)

Delivery Mode

Online learning (synchronous and asynchronous)

Assessment with Rubrics

Problem sets in activities, assignment, assessment

Requirement with Rubrics

Solving problem sets.

Readings

For further understanding of the lesson refer to the link https://www.khanacademy.org/

HYPOTHESIS

A hypothesis an approximate explanation that relates to the set of facts that can be tested by certain further investigations.

Types of Hypothesis

Null Hypothesis

Generally, null hypothesis is denoted by H_o. It states the exact opposite of what an investigator or an experimenter predicts or expects. It basically defines the statement which states that there is no exact or actual relationship between the variables.

Alternative Hypothesis

Alternative hypothesis is denoted by H_a. It makes a statement that suggests or advises a potential result or an outcome that an investigator or the researcher may expect. It has been categorized into two categories: directional alternative hypothesis and non-directional alternative hypothesis.

Illustrative examples:

1) Question: Does the data suggest that, on the average, people are able to lose more weight on a low carbohydrate diet than on a low-fat diet?

Null Hypothesis: There is no difference in the mean amount of weight loss when comparing a low carbohydrate diet with a low-fat diet
(pop. mean weight loss on a low carbohydrate diet = pop. mean weight loss on a low-fat diet).

$$\mu_1 = \mu_2$$

Alternative Hypothesis: There is a difference in the mean amount of weight loss when comparing a low carbohydrate diet with a low-fat diet
(pop. mean weight loss on a low carbohydrate diet ≠ pop. mean weight loss on a low fat diet).

$$\mu_1 \neq \mu_2$$

2) Question: Does the data suggest that females are more likely than males to eat vegetarian meals on a regular basis?

Null Hypothesis: There is no sex effect regarding those who eat vegetarian meals on a regular basis
(pop. percent of females who eat vegetarian meals on a regular basis = pop. percent of males who eat vegetarian meals on a regular basis or in symbol,

$$p_{females} = p_{males}$$

Alternative Hypothesis: Females are more likely than males to eat vegetarian meals on a regular basis
(pop. percent of females who eat vegetarian meals on a regular basis > pop. percent of males who eat vegetarian meals on a regular basis or in symbol,

$$p_{\text{females}} > p_{\text{males}}$$

3) Question: A financial analyst believes there might be a positive association between the change in a stock's price and the amount of the stock purchased by non-management employees the previous day (stock trading by management being under "insider-trading" regulatory restrictions).

Null Hypothesis: The correlation between the daily stock price change and the daily stock purchases by non-management employees.

$$P = 0$$

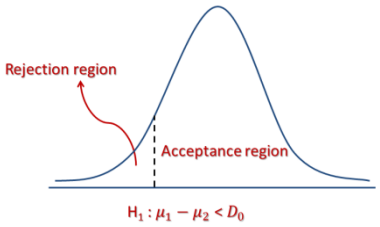
Alternative Hypothesis: The correlation between the daily stock price change and the daily stock purchases by non-management employees

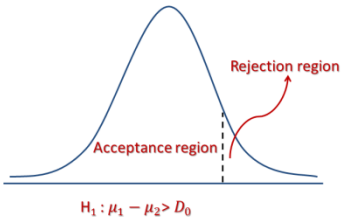
$$P > 0$$

Types of Test

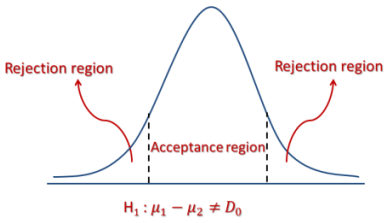
The **one-tailed test** refers to a test of null hypothesis, in which the alternative hypothesis is articulated directionally. Here, the critical region lies only on one tail. However, if the alternative hypothesis is not exhibited directionally, then it is known as the **two-tailed test** of the null hypothesis., wherein the critical region is one both the tails.

One-Tailed Test





Two-Tailed Test



Steps in Hypothesis Testing

1. Formulate a null hypothesis and an alternative hypothesis

2. Set the level of significance.

3. Identify the type of statistical test as either one-tailed or two-tailed test.

4. Determine the tabular value for the test from the table.

5. Calculate the value of the required statistical test.

6. Decide whether to accept or to reject the null hypothesis. Accepting the null hypothesis implies rejecting an alternative hypothesis, while rejecting H_O implies accepting H_1. The statement below can be used as guidelines in making a decision:

Accept H_O if the computed value is less than the tabular value.

Reject H_O if the computed value is greater than the tabular value.

7. Conclusion is the last step in hypothesis testing. It is the part where researcher explains his decision and the implication of the result.

Illustrative example:

The result of last April 2018 Electrical Engineering Licensure Examination showed that the national passing rate is 52.14%. A random sample of 82 Electrical Engineering Graduates of Bataan Peninsula State University took the same examination showed 69.14% passing rate. Is this significantly different from the national passing percentage?

Given:

National Passing Rate = 52.14%
Local Passing Rate = 69,14%

1. Formulate a null hypothesis and an alternative hypothesis

Ho: There is no significant difference between the local and national passing percentages.

Ho: $P_s = P_\mu$

Ha: There is a significant difference between the local and national passing percentages.

Ha: $P_s \neq P_\mu$

2. Set the level of significance.

$\alpha = 0.05$

3. Identify the type of statistical test as either one-tailed or two-tailed test.

Since the alternative hypothesis is expressed non-equality(non-directional) the test is said to be a **two-tailed test**.

4. Determine the tabular value for the test from the table.

At α = 0.05, the z-value obtained from the critical value of z from the table = 1.96.

$z_{crit} = \pm 1.96$

5. Calculate the value of the required statistical test.

$$z = \frac{Ps - P\mu}{\sqrt{\frac{P\mu(1 - P\mu)}{n}}}$$

$$z = \frac{.5214 - .6914}{\sqrt{\frac{.6914(1 - .6914)}{82}}}$$

$$z = \frac{-.17}{.051295}$$

$$z = -3.31$$

6. Decide whether to accept or to reject the null hypothesis.

Since the computed z-value is greater than the critical value at 0.05, reject the null hypothesis.

$z = -3.31 > z_{crit} = \pm 1.96$; Reject Ho

(Note: Rejecting Ho implies accepting Ha.)

7. Conclusion:

There is a significant difference between the local and national passing percentages.

It implies that the two passing percentages are of not equal in value with their respective proportions.

CORRELATION

The word correlation is used in everyday life to denote some form of association. We might say that we have noticed a correlation between academic performance and study habits. In statistical terms, we use correlation to denote association between two quantitative variables. We also assume that the association is linear, that one variable increases or decreases a fixed amount for a unit increase or decrease in the other. The other technique that is often used in these circumstances is regression, which involves estimating the best straight line to summarise the association.

Correlation is a statistical tool used to determine the existence of a linear relationship between variables. It is a measure of strength of association among and between variables.

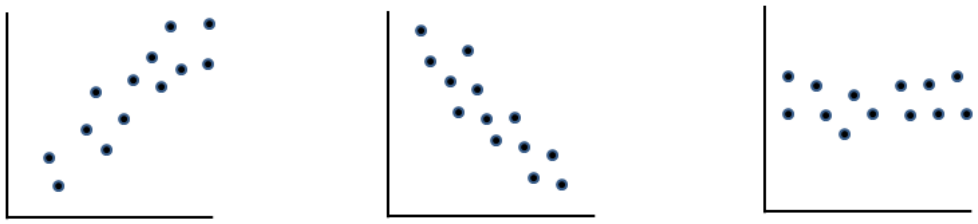
Correlation Coefficient is a numerical measure to determine whether the two or more variables are linearly related and to determine the strength of the relationship between or among the variables.

Types of Correlation

Simple correlation can be positive or negative. Positive correlation exists when one variable is directly proportional with other. Negative correlation is when variables are inversely proportional with each other.

A **correlation** of +1 means a perfect positive linear relationship
A **correlation** of 0 means no linear relationship.
A **correlation** of –1 means a perfect negative linear relationship

Scatterplot showing the different types of correlation.



Positive Correlation Negative Correlation No Correlation

Scale of Correlation:
±0.81 to ±1.00 (very strong relationship)
±0.61 to ±0.8 (strong relationship)
±0.41 to ±0.6 (moderate relationship)
±0.21 to ±0.4 (weak relationship)
± 0 to ±0.2 (weak or no relationship)

Measures of Correlation

Some statistical tools used to determine correlation coefficient are:

-Pearson's Product Moment Correlation Coefficient.

-Spearman Rank Order Correlation Coefficient (rho)

Spearman Rank Order Correlation Coefficient (rho)

-is a non-parametric measure of statistical dependence between two variables. It is used when the variables to compare are ordinal/ranked data.

Formula:

$$r_s = 1 - \frac{6\sum d^2}{(n^3 - n)}$$

where:

r_s= spearman rank correlation coefficient

$\sum d^2$ = sum of the squared difference between ranks

n = the total number of cases

English Psychologist
Charles Spearman (1863-1945)

Steps in Calculating Spearman Rho Correlation Coefficient

1. Rank the scores separately for each variable. In case of tied scores, get the average rank of the tied scores.
2. Subtract corresponding ranks for each subject/respondent.
3. Square each rank difference.
4. Get the sum of the squared rank difference.
5. Substitute the values into the formula.

Illustrative examples:

Example 1: Twelve students in Grade 9 are used as pilot sample to test the reliability of an achievement test in Mathematics.

Student	Score 1	Score 2	Ranks of Score 1	Ranks of Score 2	Difference between Ranks	Squared Rank Difference
	S_1	S_2	R_1	R_2	d	d^2
1	55	66	8	7	1	1
2	71	79	3	1	2	4
3	72	70	2	4	–2	4
4	43	50	10.5	9.5	1	1
5	35	31	12	12	0	0

6	64	72	6	3	3	9
7	57	57	7	8	–1	1
8	70	67	4	6	–2	4
9	69	69	5	5	0	0
10	48	50	9	9.5	–0.5	0.25
11	43	41	10.5	11	–0.5	0.25
12	75	75	1	2	–1	1
n=12						**$\Sigma d^2 = 25.50$**

$$r_s = 1 - \frac{6\sum d^2}{(n^3 - n)}$$

$$r_s = 1 - \frac{6(25.50)}{(12^3 - 12)}$$

$$r_s = 1 - \frac{153}{1716}$$

$r_s = 0.91$ very strong relationship

Example 2: The table shows the serum and bone magnesium levels of 14 patients. Can we conclude from these data that a relationship exist between serum magnesium and bone magnesium in the sample population.

Serum (X)	**Bone (Y)**	**Rank X**	**Rank Y**	**d**	**d^2**
3.60	672	14	14	0	0
2.85	610	13	10	3	9
2.80	621	12	12	0	0
2.70	567	11	8	3	9
2.60	570	10	9	1	1
2.55	638	8.5	13	–4.5	20.25

2.55	612	8.5	11	–2.5	6.25
2.45	552	7	7	0	0
2.25	524	6	6	0	0
1.80	400	5	5	0	0
1.45	277	4	2	2	4
1.35	294	3	3	0	0
1.40	338	2	4	–2	4
0.90	230	1	1	0	0
n=14					**$\Sigma d^2 = 51.50$**

$$r_s = 1 - \frac{6\sum d^2}{(n^3 - n)}$$

$$r_s = 1 - \frac{6(51.50)}{(14^3 - 14)}$$

$$r_s = 1 - \frac{309}{2730}$$

$r_s = 0.8868$ very strong relationship

Pearson's Product Moment Correlation Coefficient.

-is a parametric measure of statistical dependence between two variables. It is applicable when the variables to compare are continuous/scale.

Formula:

$$\boldsymbol{r} = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

where:
Σx = summation of observed data of dependent variable
Σy = summation of observed data of independent variable
Σx^2 = summation of squares of observed data of dependent variable
Σy^2 = summation of squares of observed data of independent variable
n = number of cases

English Mathematician Karl Pearson (1857–1936)

Alternative Formula:

$$r = \frac{SS_{xy}}{\sqrt{SS_x \bullet SS_y}}$$

where:
SS_{xy}= Sum of squares xy
SS_x= Sum of squares of x
SS_y= Sum of squares of y

Test of Significance of Pearson's r Correlation Coefficient

Test for statistical significance is used to address the question: what is the probability that what we think is a relationship between two variables is really just a chance occurrence? In order to test the significance of the result we must consider the level of significance and confidence level.

The level of significance is defined as the probability of rejecting a null hypothesis by the test when it is really true, which is denoted as α. That is, the probability of committing a Type I error, P (Type I error) = α.

Confidence level refers to the possibility of a parameter that lies within a specified range of values, which is denoted as c. Moreover, the confidence level is connected with the level of significance. The relationship between the level of significance and the confidence level is $c = 1 - \alpha$.

The common level of significance and the corresponding confidence level are given:

The level of significance 0.10 is related to the 90% confidence level.
The level of significance 0.05 is related to the 95% confidence level.
The level of significance 0.01 is related to the 99% confidence level.

The obtained correlation coefficient r can be tested with the test statistic:

$$t = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} \qquad or \qquad t = r\sqrt{\frac{n - 2}{1 - r^2}}$$

where:

n = sample size
r = sample correlation coefficient of the two variables
ρ = population correlation coefficient
$n - 2$ = degrees of freedom

The computed t-value will then be compared to the tabular/critical t-value which is obtained based on the degrees of freedom and alpha level.

Coefficient of Determination (R^2)

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, It is used to analyze how differences in one variable can be explained by a difference in a second variable.

Illustrative Example:

Example 1: A researcher wants to investigate the relationship between social media usage (in hours) and reading ability (words per minute). A sample of twelve students was used in the study. The data gathered are as follows:

Student	Social Media Usage (x)	Reading Ability (y)	xy	x^2	y^2
1	8	175	1400	64	30625
2	8	130	1040	64	16900
3	6	90	540	36	8100
4	10	200	2000	100	40000
5	9	120	1080	81	14400
6	11	185	2035	121	34225
7	7	140	980	49	19600
8	11	150	1650	121	22500
9	12	145	1740	144	21025
10	7	135	945	49	18225
11	5	80	400	25	6400
12	9	190	1710	81	36100
n=12	**$\Sigma x=103$**	**$\Sigma y=1740$**	**$\Sigma xy=15520$**	**$\Sigma x^2=935$**	**$\Sigma y^2=268100$**

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$r = \frac{(12)(15520) - (103)(1740)}{\sqrt{[(12)(935) - (103)^2][(12)(268100) - (1740)^2]}}$$

$$r = \frac{186240 - 179220}{\sqrt{(611)(189600)}}$$

$r = \frac{7020}{10763.16}$

$\boldsymbol{r = 0.65}$ **strong relationship**

Coefficient of Determination

$$r^2 = (0.65)^2$$

$$\boldsymbol{r^2 = 0.4225 \; or \; 42.25\%}$$

It implies that 42.25% of the variance in reading ability is explained by the variance in social media usage

Test of Significance

Level of significance $\alpha = 0.05$; $t_\alpha = 2.228$; Tabular value : $t_\alpha = 2.228$; n = 12 ; r = 0.65

$$t_{comp} = r\sqrt{\frac{n-2}{1-r^2}}$$

$$t_{comp} = 0.65\sqrt{\frac{12-2}{1-(0.65)^2}}$$

$$t_{comp} = 2.7048$$

Since $t_\alpha < t_{comp}$ we can conclude that there is a significant relationship between the compared variables.

Support Videos:
Null and Alternative Hypothesis

https://www.youtube.com/watch?v=5N7L1cGCL-w

Type of test

https://www.youtube.com/watch?v=enQDKXI2PdA

One Sample z-test

https://www.youtube.com/watch?v=BWJRsY-G8u0

One Sample t-test

https://www.youtube.com/watch?v=VPd8DOL13Iw

Independent Samples T-test

https://www.youtube.com/watch?v=jyoO4i8yUag

Pearson's Correlation

https://www.youtube.com/watch?v=2B_UW-RweSE

Hypothesis Testing Pearson's Correlation

https://www.youtube.com/watch?v=rR-jptLvhFw

Activity Sheet

Answer the following completely:

A. Do what is asked.

1. State the null hypothesis, Ho and the alternative hypothesis, Ha, for the following.

a) The mean number of years Filipinos work before retiring is 25.
b) At least 70% of the Filipinos vote in presidential elections.
c) The mean starting salary for a Bataan Peninsula State University graduates is at least P100,000 per year.
d) Thirty percent of high school seniors get drunk each month.
e) The mean number of cars a person owns in his lifetime is not more than ten.

2. The National Institute of Mental Health published an article stating that in any one-year period, approximately 12 percent of Filipino adults suffer from depression or a depressive illness. Suppose that in a survey of 100 people in a certain town, seven of them suffered from depression or a depressive illness. Conduct a hypothesis test to determine if the true proportion of people in that town suffering from depression or a depressive illness is lower than the percent in the general adult Filipino population.

a. State the null and alternative hypotheses.
   Ho :
   Ha :
b. Is this a right-tailed, left-tailed, or two-tailed test?

3. From generation to generation, the mean age when smokers first start to smoke varies. However, the standard deviation of that age remains constant of around 2.1 years. A survey of 40 smokers of this generation was done to see if the mean starting age is at least 19. The sample mean was 18.1 with a sample standard deviation of 1.3. Do the data support the claim at the 5% level? (justify your answer by hypothesis testing)

4. The mean number of sick days an employee takes per year is believed to be about ten. Members of a personnel department do not believe this figure. They randomly survey eight employees. The number of sick days they took for the past year are as follows: 12, 4, 15, 3, 11, 8, 6, 8. Let x be the number of sick days they took for the past year. Should the personnel team believe that the mean number is ten?

5. The student academic group on a college campus claims that freshman students study at least 3 hours per day, on average. One Introduction to Statistics class was skeptical. The class took a random sample of 30 freshman students and found a mean study time of 140 minutes with a standard deviation of 45 minutes. At $\alpha = 0.05$ level, is the student academic group's claim correct?

6. Ten pupils from grade six were randomly selected from a certain public school system. They were ranked according to the quality of their home environment (the scale used is (1-poor and 7-excellent) and the quality of their performance in school (1-needs improvement and 7-

outstanding. The result is shown in the table. Calculate r_s and determine whether one can conclude that the two variables are correlated.

Home Environment	3	4	5	4	6	2	3	4	5	1
Performance in School	4	5	6	4	7	3	5	6	7	3

7. A company owner wishes to find out whether there is a relationship between the age of his employees and the number of absences they take each year. The data are as follows:

Age	**No. of Days Absent**
24	2
27	4
29	3
32	3
36	5
40	6
35	6
28	1
33	0
48	8

Using the steps in hypothesis testing calculate r, and determine its significance at 0.05.

8. A researcher is interested to know whether the height of tomato plant affects the number of fruits it can produced. Fifteen tomatoes were chosen at random from the farm. And the results were recorded. Using the steps in hypothesis testing. Determine if there is a significant relationship between the two variables. (α=0.05 level of significance)

Height in (cm)	**No. of Fruits Produced**
30	34

34	40
38	46
26	30
22	32
25	28
28	26
33	28
36	36
40	56
42	58
36	34
34	28
32	24
22	30

Learner's Feedback Form

Name of Student: _____
Program : _____
Year Level : _____ Section : _____
Faculty : _____
Schedule : _____

Course Packet : Code : _____ Title : _____

How do you feel about the topic or concept presented?

- ☐ I completely get it.
- ☐ I've almost got it.
- ☐ I'm struggling.
- ☐ I'm lost.

In what particular portion of this course packet, you feel that you are struggling or lost?

_____
_____
_____

Did you raise your concern to you instructor? ☐ Yes ☐ No

If Yes, what did he/she do to help you?

_____
_____
_____

If No, state your reason?

_____
_____
_____

To further improve this course packet, what part do you think should be enhanced?

_____
_____
_____

How do you want it to be enhanced?

_____
_____
_____