De La Salle University- Manila
Gokongwei College of Engineering

PROLOGI
Programming Logic and Design

Project Proposal

Hippopotamus (Hypothesis Testing Calculator)

Marcus Alden Taraya
Tayshaun Pierce Alabado
Win Clarence Sy

Project Description

We have noticed that in Google, there is a lack for a program that performs the hypothesis testing. We have made it upon with a goal to develop a program that can perform hypothesis testing for various statistical scenarios. These scenarios include the following:

Single Population

- Case A: Single Population Where Sigma is Known or Sample is Large
- Case B: Sigma is Unknown or Sample is Small
- Test for Proportions: Single Population
- Variance For Single Population"

Two Population

- Case C: Two Mean Big Sample Size or Population Standard Deviation is Given
- Case D: Two Mean Small Sample Size and Population Sample Standard Deviation assumed Equal
- Case E: Two Mean Small Sample Size and Population Sample Standard Deviation not Equal
- Case F: Paired Observation
- Test for Proportions: Two Population
- Variance For Two Populations

The program should take in user inputs for the statistical data, the level of significance, and the type of test. The program will use the given data to get the test scores and critical values then sum up by either rejecting or accepting the null hypothesis.

Key Features

- User-friendly interface for inputting statistical data and hypotheses
- Calculation of test statistic and p-value
- Selection of appropriate statistical test based on user input (e.g., t-test, z-test, chi-squared test, etc.)
- Interpretation of results based on user-specified level of significance (e.g., reject or fail to reject null hypothesis)

IPO

Input	Process	Output
Case A x, u_0, o, n, alpha **Case B** x, u_0, s, n, alpha	**Case A** $z = \frac{\bar{x} - \mu_0}{\left(\sigma/\sqrt{n}\right)}$ **Case B** $t = \frac{\bar{x} - \mu_0}{\left(s/\sqrt{n}\right)} \quad df = n - 1$	Test Statistic P-Value **(if any)** Critical Value Degree of Freedom **(if any)**

Single Population Test Proportions

x, n, p_0, alpha

Case C

x_1, s_1, n_1, x_2, s_2, n_2, d_0, alpha

Case D

x_1, s_1, n_1, x_2, s_2, n_2, d_0, alpha

Case E

x_1, s_1, n_1, x_2, s_2, n_2, d_0, alpha

Case F

d, d_0, s, n, alpha

Two Populations Test Proportions

x_1, n_1, x_2, n_2, alpha

One Population Variances

n, s, o_0, alpha

Two Population Variances

n_1, s_1, n_2, s_2, alpha

Single Population Test Proportion

$$\hat{p} = \frac{x}{n} \quad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Case C

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Case D

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$df = n_1 + n_2 - 2$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Case E

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Case F

$$t = \frac{\bar{d} - d_0}{\left(s_d / \sqrt{n}\right)}$$

$$df = n - 1$$

Test Proportions for Two Population

$$\hat{p}_1 = \frac{x_1}{n_1} \text{ and } \hat{p}_2 = \frac{x_2}{n_2} \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Sample Proportions **(if any)**

Pooled Sample Proportion **(if any)**

Pooled Standard Deviation **(if any)**

Difference **(if any)**

Difference Standard Deviation **(if any)**

Whether to Reject or Not Reject the Null Hypothesis

	$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ **One Population Variances** $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ $v = n - 1.$ **Two Population Variances** $f = \frac{s_1^2}{s_2^2}$ $v_1 = n_1 - 1$ and $v_2 = n_2 - 1.$	

Methodology

This project will utilize Python language since it has powerful libraries that we could use that would help us in this project. **Numpy** and **Math** would be used for calculations. **Scipy** would be used to get the critical value for the problem. **Colorma** would be used for coloring the program so that it would be easier for the user to see the important values. We will be utilizing several functions such as if-else statements. For the formula, we would be getting it from the course FNDSTAT we have taken from Term 1 AY2022-2023. We will be using modular functions to simplify the code and avoiding code duplication would make troubleshooting easier. We would be separating each case with its own functions so that it would be more user-friendly.

Flowchart

The Google Drive folder link below contains the flowchart for this project since when imported into the word file, it would become very blurry and unreadable.

https://drive.google.com/drive/folders/10UjIdV-SWk1_epcxUHT7WwWsRHHYuHc1?usp=share_link

Schedule of Activities

Tasks	Description	Date	Person-In-Charge
Planning	In this part, we will decide on the project that we will make. We would plan on the division of tasks to make it more efficient and faster to complete the project. We will find the resources we need that would be utilized in this project.	March 6-8, 2023	Everyone in the Group will be assigned to do a portion of each task.
Structuring and IPO	In this part, we will be planning on the structuring of the code by laying out the IPO so that we will know what variables to use.	March 9-12, 2023	
Finalizing Project Proposal	We would be finalizing the project proposal and uploading it into the GitHub repository.	March 12-20, 2023	
Coding	In this part, we will be coding the program for the project.	March 20-23, 2023	
Making Hierarchy Chart	We will be making the hierarchy chart to explain how the chart would work.	March 24, 2023	
Documentation	In this part, we will be documenting the code and the project. This includes finishing all parts of the project paper.	March 24-April 4, 2023	
Reviewing Documentation and GitHub repository	This is to make sure that there are little to no errors that can be found.	April 5, 2023	
Project Presentation and Demo	We would be making the project presentation and demo which would be posted to YouTube and the	April 5, 2023	

	link would be added to README.md in the GitHub repository		
Demonstration	This would be the part where we demonstrate how the program works.	April 17, 2023	

References

- Mathematics and Statistics Department. (2022). FNDSTAT Hypothesis Testing Formula Sheet For Variances. *De La Salle University*. https://drive.google.com/drive/folders/1Y1mhqy8yeHZB9YC7WBPkQCNMH8Icvorh?usp=sharing
- Mathematics and Statistics Department. (2022). FNDSTAT Hypothesis Testing Formula Sheet for Two Populations. *De La Salle University*. https://drive.google.com/drive/folders/1Y1mhqy8yeHZB9YC7WBPkQCNMH8Icvorh?usp=sharing
- Mathematics and Statistics Department. (2022). FNDSTAT Hypothesis Testing Formula Sheet for Single Populations. *De La Salle University*. https://drive.google.com/drive/folders/1Y1mhqy8yeHZB9YC7WBPkQCNMH8Icvorh?usp=sharing