DA Assignment (Hypothesis Test)

*- Dharun VS **20AD09**, Harish Khumar **20AD56***

Introduction

From the data-set that describes Russian and Ukrainian Equipment Losses During The 2022 Russian Invasion Of Ukraine we perform z-test on losses_total column to check if sample is from population

Statistics

Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data. In applying statistics to a scientific, industrial, or social problem, it is conventional to begin with a statistical population or a statistical model to be studied.

Hypothesis Test

A statistical ***hypothesis test*** is a method of statistical inference used to decide whether the data at the hand sufficiently supports a hypothesis. Hypothesis testing allows the researcher to determine whether the data from the sample is statistically significant. Hypothesis testing is one of the most important processes for measuring the validity and reliability of outcomes in any systematic investigation.

The ***null hypothesis*** is a typical statistical theory which suggests that no statistical relationship and significance exists in a set of given single observed variable, between two sets of observed data and measured phenomena.

*The **alternative hypothesis*** is one of the proposed proposition in the hypothesis test.

A ***z-test*** is a statistical test to determine whether two population means are different when the variances are known and the sample size is large. A z-test is a hypothesis test in which the z-statistic follows a normal distribution. A z-statistic, or z-score, is a number representing the result from the z-test.

Data-set Study

Data Description

The data-set describes Russian and Ukrainian Equipment Losses During The 2022 Russian Invasion Of Ukraine.

The data-set was created based on Oryx by scraping Ukrainian losses and Russian losses pages.

This list only includes destroyed vehicles and equipment of which photo or video-graphic evidence is available. Therefore, the amount of equipment destroyed is significantly higher than recorded. You can find numbers here 2022 Ukraine Russia War Data-set.

Data includes Equipment Losses, Equipment Models, Countries that produce Equipment, the Number of Equipment Losses, and types of Losses (abandoned, damaged, destroyed, captured, etc.).

Main Columns

equipment
model
sub_model
manufacturer
losses_total

Data-set History

2022-05-08 - data-set is updated.
2022-04-30 - data-set is created (after 66 days of the War).

Data-set Source

https://www.kaggle.com/datasets/piterfm/2022-ukraine-russia-war-equipment-losses-oryx

Research Problem

The sample from column losses_total id from the population from the data-set that describes Russian and Ukrainian Equipment Losses During The 2022 Russian Invasion Of Ukraine

Hypothesis

H0 - Sample from population

H1 - Sample not from population

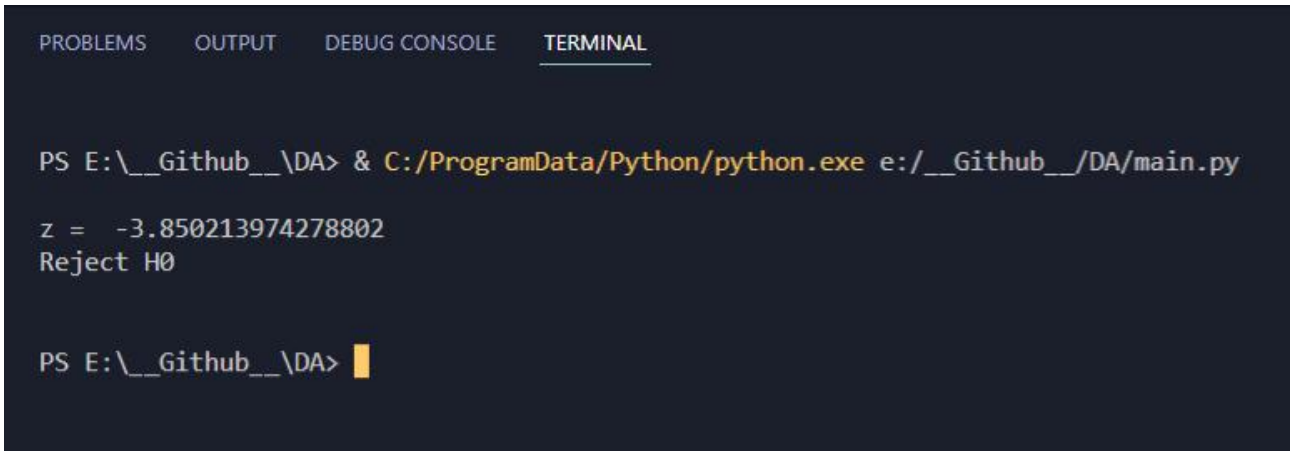
Decision Rule

Reject H0 at ***0.05*** level of significance if calculated |z| value is greater than the table z value ***1.96***

Calculation

z = **-3.850213974278802**

|z| > table z



Source repository - https://github.com/dharunvs/DA

Result

From the interpretation the sample does not belong to the population