



EE-495

Special Topics In Computer Engineering

Case Study No. 1

Hypothesis Testing For Coronavirus Using Python

Haneen Alamoudi - 1708436

Submission Date: 11 March 2021 (WEEK 8)

Contents

Introduction.....	3
Output	3
Code.....	4
Conclusion	5

Introduction

In this case study, the implementation of a two-sample Z test for coronavirus dataset was done, and the formula is shown below. John Hopkin's repository coronavirus dataset has been studied, the dataset includes information about state, country, last update, confirmed cases, deaths cases, recovered cases latitude and longitude. The case study concentrates on the feature of temperature and humidity for latitude and longitude. In the dataset, a temperature below 24 considered to be a cold climate and above 24 considered as hot climate

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

Main Hypothesis for testing and current perception: Warm climate is more resistant to the corona outbreak.

Null Hypothesis: Temperature doesn't affect COV-19 Outbreak

Alternate Hypothesis: Temperature does affect COV-19 Outbreak

Output

The screenshot shows the Spyder Python 3.8 IDE interface. The left pane displays the code file `temp.py` containing the implementation of a Two-Sample Z-test. The right pane shows the variable explorer with the following data:

Name	Type	Size	Value
corona	DataFrame	(206, 12)	Column names: Province/State, Country/Region, Last Update, Confirmed...
corona_t	DataFrame	(206, 2)	Column names: Confirmed, Temp_Cat
d1	Series	(31,)	Series object of pandas.core.series module
d2	Series	(175,)	Series object of pandas.core.series module
Hypothesis_Status	str	1	Do Not Reject Null hypothesis : Not Significant
m1	float	1	26.548387096774192
m2	float	1	672.9085714285715
n1	int	1	31
n2	int	1	175
p	float64	1	0.10205422251419871
p_val	float64	1	0.102054
sd1	float	1	45.93969867095882
sd2	float	1	5228.628797929274
z	float64	1	-1.6349753148440753
z_score	float64	1	-1.63497531

The screenshot shows the Spyder IDE interface with the following details:

- File Menu:** File, Edit, Search, Source, Run, Debug, Consoles, Projects, Tools, View, Help.
- Toolbar:** Includes icons for file operations like Open, Save, Run, Stop, and Help.
- Path:** C:\Users\hano0\OneDrive\university\LEVEL-10\Special topics in CE\case_study.py
- Code Editor:** Displays the Python script `temp.py`. The code performs statistical hypothesis testing on COVID-19 data. It includes imports for pandas and numpy, reads a CSV file, applies a lambda function to categorize temperature, and defines a TwoSampZ function for a two-sample Z-test. The script then filters the data for confirmed cases and prints the p-value and hypothesis status.
- Help Panel:** Shows the "Usage" panel with instructions on how to get help for objects.
- Console:** Shows the output of the script's execution, including the p-value (0.102054) and the hypothesis status ("Do Not Reject Null Hypothesis : Not Significant").
- Bottom Status Bar:** LSP Python: ready, conda: base (Python 3.8.5), Line 47, Col 1, UTF-8, CRLF, RW, Mem 78%.

Code

```
# -*- coding: utf-8 -*-
"""
Created on Thu Mar 11 08:06:26 2021

@author: Haneen Alamoudi
ID: 1708436
"""

import pandas as pd
import numpy as np

corona = pd.read_csv('Corona_Updated.csv')

corona['Temp_Cat'] = corona['Temprature'].apply(lambda x : 0 if x < 24 else 1)

corona_t = corona[['Confirmed', 'Temp_Cat']]
def TwoSampZ(X1, X2, sigma1, sigma2, N1, N2):
    from numpy import sqrt, abs, round
    from scipy.stats import norm
    ovr_sigma = sqrt(sigma1**2/N1 + sigma2**2/N2)
    z = (X1 - X2)/ovr_sigma
    pval = 2*(1 - norm.cdf(abs(z)))
    return z, pval

d1 = corona_t[(corona_t['Temp_Cat'] == 1)]['Confirmed']

d2 = corona_t[(corona_t['Temp_Cat']==0)]['Confirmed']

m1, m2 = d1.mean(), d2.mean()
sd1, sd2 = d1.std(), d2.std()
n1,n2 = d1.shape[0], d2.shape[0]

z,p = TwoSampZ(m1, m2, sd1, sd2, n1, n2)

z_score = np.round(z,8)
p_val = np.round(p,6)

if(p_val < 0.05):
    Hypothesis_Status = 'Reject Null Hypothesis : Significant'
else:
    Hypothesis_Status = 'Do Not Reject Null Hypothesis : Not Significant'

print(p_val)
print(Hypothesis_Status)
```

```

d2 = corona_t[(corona_t['Temp_Cat']==0)]['Confirmed']

m1, m2 = d1.mean(), d2.mean()

sd1, sd2 = d1.std(), d2.std()
n1,n2 = d1.shape[0], d2.shape[0]

z,p = TwoSampZ(m1, m2, sd1, sd2, n1, n2)

z_score = np.round(z,8)
p_val = np.round(p,6)

if(p_val < 0.05):
    Hypothesis_Status = 'Reject Null Hypothesis : Significant'
else:
    Hypothesis_Status = 'Do Not Reject Null Hypothesis : Not Significant'

print(p_val)
print(Hypothesis_Status)

```

Conclusion

Based on the calculation done using the Python code, $z = -1.634975$ and $p = 0.102054$. since P-value is greater than 0.05, we cannot reject the null hypothesis because it is not significant. And conclude that temperature does not effect COVID-19.