Bharatiya Vidya Bhavan's
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)
Academic SEM: VII **Year: 2022-23**

Experiment: Hypothesis Testing using Python

Name:	Tripathi Vinayak Ramprakash
UID:	2019110067
Class:	BE ETRX
Batch:	A
Subject:	Data Analytics Lab

Objective:
Perform statistical data analysis such as: Estimators of the main statistical measures (mean, variance, standard deviation, covariance correlation, standard error), Main distributions (Normal distribution, chi-square distribution), Hypothesis testing, pair-wise association (Pearson correlation test, t-test, ANOVA), Non-parametric test (Spearman rank)

System Requirements: CoLab, Python 3.7, SKLearn, SciKit

DataSet:
The Dataset considered in this experiment is collection of IQ and CGPA of third year college students.

Code: Please consider the ipynb file in the folder for code and output

Interpretation:
1. **Descriptive Statistics - What the Data shows?**
   →The data shows Bimodal Distribution. We can see two peaks in the histogram.
   →The data is of size 200. So it is a pretty large dataset and we will be using z-test to do the hypothesis testing.
   →The 5 point summary of the dataset shows that the max IQ for the college students is 121, median is 102, min is 83 with std deviation of 12.16 This suggests that the value of the IQ varies by 41 and the variation is consistent.
   →Same applies to the CGPA. This is also bimodal means the student either score well above 6.5 and those who can't even score this much scores very less like from 5.4 to 4.6.

2. **Building and Using Models - Can the data be used to estimate values or predict the future?**
   →We can use the data to predict the future values. This IQ and CGPA shows correlation. So we can say that assuming the same trend, If we have the CGPA then we can predict the IQ of the person.
   → Now the correlation between the IQ and CGPA is just a coincidence or then exists some relation will be checked in the next stage

3. **Hypothesis Testing - Are the descriptive statistics and models just coincidence for this sample?**

→The size of the dataset is 200 so we will use the z test
→First we did the independent test to check if the college students had higher IQ than average people. We found than this is not the case we didn't had enough evidence to prove that the IQ was greater than equal to the average IQ of the world
→Next was the two dependent test to check if the IQ and CGPA are truly correlated ot not. We Found enough evidence to prove that the Above statement is true wth a confidence of 95%.

Conclusion:

- We performed Statistical analysis usong python on real dataset
- We also did Hypothesis testing using python and followed proper steps to do the testing.
- We also learned how to define null hypothesis and alternative hypothesis.