In [ ]:

Out of 500 students, 410 students pass the exam. Test the hypothesis that more than 80% of students pass the exam. Level of significance = 5%

Name:Rishabh Singhvi
Sap: 60009210206
Branch: Data Science(K3)

```
#Question 1
from scipy.stats import norm p_p =e.8 sample_success = 410 sample_size -- see from statsmodels.stats.proportion import proportions_ztest stat, p_value = nobs = sample_size,value =_print( ' z_stat = ' , stat, ' p_value z cri —— norm.ppf(ø.95) print( 'z_cri - ' ,z_cri)
```

z stat 1.16405e4929492914 p_value e. 12220177493249235 z cri = 1 .6448536269514722

In [ ]:

/usr/10ca1/1ib/python3.7/dist-packages/statsmode1s/t001s/_testing.py:19: Future Warning: pandas. util. testing is deprecated. Use the functions in the public API at pandas. testing instead .

import pandas. util. testing as tm

```
SE = np. sqrt(p_p*(l-
p_p)/sample_size)

print( ' z_stat ' , z) z cri =
norm.ppf(ø.95) norm .
sf(z)
```

Question I :

Question 2: A random sample of 35 individuals selected from a certain population showed that 6 of them are left-handed. Find 95% and 99% confidence limits for the proportions of left-handed person in the population.

```
#Question 2 import
numpy as np s_P =
6/35
n =35
SE = np.sqrt(s_p*(1-s_p)/n)
z = norm.ppf(0.975)
L = s_p - z*SE
U = s_p + z*SE
CI = (L,U)
print( '95% Confidence Interval for population means z =
norm.ppf(e.995)
L = s_p -   - z*SE                                          ' ,CI)
U = s_p +   + z*SE
CI = (L,U)
print( • 99% Confidence Interval for population means       ' ,CI)
```

95% Confidence Interval for population means = (e. 04656934249155212, e. 2962878 0036559e77)

99% Confidence Interval for population means = (e.eø7335729288939441, e. 335521 4135682e34)

Question 3: The manufacturer of patent medicine claimed that it was 90% effective in relieving allergy for a period of 8 hours. In a sample of 200 people who had the allergy, the medicine provided relief for 160 people. Determine whether manufacturer's claim is legitimate.

In [ ]:

```
#Question 3
#Ho : P = 0.9 , Ha - 0.9
sample_size      =      2Øe
sample_success = 160
 p_p = 0.9
 stat,p_value = proportions_ztest(count=sample_success,
 z_cri = norm.ppf(0.95)
 print('z_stat = ',stat,'p_value = ',p_value)          nobs = sample_size, value =
print( •z_cri = ' ,z_cri)
```

z stat = -3.535533905932737 p_value = o. eøe4e6952e1744495973 z cri = 1 .6448536269514722

Question 4: From one class of 500 students, 410 students passed. From other class of 400 students, 379 students passed. Test the hypothesis that both samples have same proportions. Level of significance = 0.025

In [

```
alpha = o. 025 success — sample stat,p_value =      nobs = sample, alternative = 'twoif
         = np.array([410,379])                          p_value > alpha:
        = np.array([500,400])                             print("Test is not
        value = proportions_ztest(count=success,
reject null hypothesis" ) else :
  print("Test is significant. Reject null hypothesis")
print( 'z_stat      stat, 'p_value __ ',p_value)
```

Test is significant. Reject null hypothesis z stat = -5.78e2476568050825
p_value - 7.459074e60078635e-09

Question 5: Two groups, A and B, each consists of 100 people who have a disease. A serum is given to group A but not group B; otherwise both the groups are treated identically. It's found that in group A and group B, 75 and 65 people respectively, recover from the disease. Test the hypothesis that serum helps to cure the disease using a level of significance (a) 0.01 (b) 0.1.

In [ ]:

```
alpha = o.el success sample = stat,p_value = nobs = sample, alternative = ' larg if p_value > alpha:
         = np.array([75,65])                              print("Test is not
          np.array([100,100])
        alue = proportions_ztest(count=success,
reject                                                    null hypothesis" )
else :
  print( "Test is significant. Reject null hypothesis" )
print( z_stat = ' , stat, 'p_value = ',p_value)
```

Test is not significant. Fail to reject null hypothesis z stat = 1.
543e334996209185 p_value = o. e61411324e5069633

In [J :

```
alpha  =  0.1  success  =                                                    '
sample = np.array([100,100])                                                 larg
stat,p_value = proportions_ztest(
if p_value > alpha:
  print("Test is not significant.
np.array([75,65])
                                        count=success, nobs = sample, alternative =
                  is not significant. Fail to reject null hypothesis" )
else:
  print( "Test is significant. Reject null hypothesis"
print( 'z_stat ' , stat, 'p_value __ ' ,p_value)
```

Test is significant. Reject null hypothesis z stat 1. 543e334996209185
p_value = o. e61411324e5Ø69633

Question 6: Check the hypothesis that blood pressure before treatment has a mean value of 156 for the given dataset.

```
import pandas as pd df = pd. ead_csv( . .
/content/blood_pressure . csv" df . head()
```

out[16]:

	patient	sex	agegrp	bp_before	bp_after
0	1	Male	30-45	143	153
	2	Male	30-45	163	170
2	3	Male	30-45	153	168
	3	4	30-45	153	142
Male			30-45	146	141
	4	5			
Male					

In [J :

```
df . describe()
```

out[17]:

	patient	bp_before	bp_after
count	120.000000	120.000000	120.000000
mean	60.500000	156.450000	151.358333
std	34.785054	11.389845	14.177622
min	1 .oooooo	138.000000	125.000000
25%	30.750000	147.000000	140.750000
50%	60.500000	154.500000	149.500000
75%	90.250000	164.000000	161.000000
max	120.000000	185.000000	185.000000

In [ ]:

```
alpha = 0.05 from statsmodels . stats import weightstats as stests
z_stats,p_val = stests.ztest(df[ ' bp_before ' = None,value = 156, alternative
= print( 'Z_stat: ' ,z_stats, 'p value-' , p_val) print( ' ' , norm . ppf(Ø.975)) if
p_val > alpha:
                    is not significant. Fail to reject null hypothesis" else
:
  print("Test is significant. Reject null hypothesis"
```

z_stat: 0.432798073526164 p value= 0.6651614730255063 z
critical- 1.959963984540054
Test is not significant. Fail to reject null hypothesis

Question 7: Can we say that the systolic blood pressure measured on arm has mean value of 125 mmHg from the given dataset?

In[ ]:

```
df = pd .   _   ' /content/systolic blood pressure.csv'
df. head ( )
```

Out[24] :

	armsys	fingsys
o	140	154
1	110	112
2	138	156
3	124	152
4	142	142

In[ ]:df. describe ( )

out[26] :

	armsys	fingsys
count	200 000000	200.000000
mean	128 520000	132.815000
std	23.287575	25.648195
nun	79 000000	60.000000
25%	111 500000	118.000000
50%	125 000000	130.000000
75%	140.000000	146.500000
max	220 000000	228.000000

```
't
alpha = 05 z_stats , p_val   stests  'armsys'        _ None,value  125, alternative
print( • Z_stat : '     •p value=• ,p_val) print( ' z_critical- ' ,norm. ppf(O.975)) if
p_val alpha:
   print("Test is not significant.Fai1 to reject null hypothesis" else:
   print( "Test is significant. Reject null hypothesis"
```

In C]:

z_stat: 2.1376342148925 p value- ø.e32546442541153964 z
critical= 1.95996398454ee54

Test is significant. Reject null hypothesis

Question 8: A sample of 50 pieces of certain type of string was tested. The mean breaking strength turned out to be 14.5 pounds. Test whether the sample is from a batch of string having mean breaking strength of 15.6 pounds and standard deviation of 2.2 pounds.

localhost:8888/notebooks/Downloads/SDS

S_Expt_6

]:

x bar = 14 5

mean = 15.6
sigma - —
2.2 n = 5e
SE = sigma/np.sqrt(n)

```
z  =(x_bar-mean)/SE
print('Z_stat=',z)
print( ' Z_critical=' norm . ppf(O.Ø25))
p_value = 2*norm.cdf(z) print( '
p_value- - ' sp_val) if p_val > alpha:
   print("Test is not significant.Fai1 to reject null hypothesis") is
else:
   print("Test significant. Reject null hypothesis" )
```

Z stat— -3.5355339059327364

Z critical— -1.9599639845400545

alue=0.032546442541153964

Test is significant. Reject null hypothesis

Question 9: Check the hypothesis that mean value of systolic blood pressure measured on arm and finger are the same.

In [ ]:df . describe()

Out[29] :

	armsys	fingsys
count	200.000000	200.000000
mean	128.520000	132.815000
std	23.287575	25.648195
min	79.000000	60.000000
25%	111.500000	118.000000
50%	125.000000	130.000000
75%	140.000000	146 500000
max	220.000000	228.000000

```
alphae .05 z_stats, p_val = stests.ztest(df['armsys'],x2 =  'fingsys '                    =          e,
],valueprint( 'Z_stat: ' z_stats,_'p value- - ' sp_val) print( ' z_critical-  alternati
' , norm . ppf(0.025)) if p_val > alpha:
  print("Test is not significant.Fai1 to reject null hypothesis"
else:
  print("Test is significant. Reject null hypothesis"
```

In [ ]:

Z stat: -1.7533233440871472 p value- e. 07954652069053099 z
critical= -1.9599639845400545
Test is not significant. Fail to reject null hypothesis

Question 10: A common perception about COVID-19 is that warm climate is more resistant to the corona outbreak. Verify this using hypothesis testing.

```
[ ]: corona pd . read_csv( ' / content/Corona_Updated . csv ' corona.
     head()
```

Out[33] :

	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered	Latitude	Longitude
	Hubei	Mainland China	2020-03-10T17:53:02	67760	3024	47743	30.9756	112.2707
1	NaN	Italy	2020-03-10T15:13:05	10149	631	724	43.0000	12.0000
2	NaN	Iran (Islamic Republic of)	2020-03-10T19	8042	291	2731	32.0000	53.0000
3	NaN	Republic of Korea	2020-03-10T19	7513	54	247	36.0000	128.0000
4	NaN	France	2020-03-10T18:53:02	1784	33	12	47.0000	2.0000

```
[ ]: corona . describe(
```

Out [34] :

	Confirmed	Deaths	Recovered	Latitude	Longitude	Temprature	Humidity
count	206.000000	206.000000	206.000000	206.000000	206.000000	206.000000	206.000000
mean	575.640777	20.689320	312.640777	31.184989	11.752030	12.161165	67.728155
std	4822.697784	215.794845	3332.764713	21.305149	84.576291	10.229763	21.780588
min	0.000000	0.000000	0.000000	-41.454500	-157.498300	-21.900000	6.000000
25%	3.000000	0.000000	0.000000	25.069200	-74.841325	6.100000	55.000000
50%	12.000000	0.000000	0.000000	36.030550	15.234250	11.750000	73.000000
75%	75.750000	1 .oooooo	4.000000	43.870250	101.363375	20.375000	84.000000
max	67760.000000	3024.000000	47743.000000	64.963100	174.886000	33.100000	98.000000

In [ ]:

```
corona['Temp_cat']                                                    if x<24 else 1)
corona_t = corona[[= corona( 'Temprature ' ] . apply(lambda
corona_t.head()      'Confirmed' , 'Temp_cat
corona_t[(coro
corona_t[(coro
dl = d2 — ml         'Temp_cat '    'Confirmed' ] 'Temp_cat '
dl.mean() d2.mean()   ' Confirmed ' ]
= dl.std() s2 = d2.stc
dl.shape[e]    n2
d2.shape[e]    prin
'Mean confirme       cases for warm climate : ' ,ffll) cases
print( "Mea          for not warm climate:
confirmed
```

Mean confirmed cases for warm climate 26.548387e96774192

Mean confirmed cases for not warm climate: 672.9085714285715

In[ ]:

```
#Ho: mul = mu2 , Ha : mul<mu2
SE = np.sqrt(s1**2/n1+s2**2/n2)
z = (ml-m2)/SE p_val __ norm.
cdf(z) print(z,p_val)
if(p_val<0.05):
 print("Test is is significant. Reject null hypothesis")
else :
  print("Test is not significant.Fai1 to Reject null hypothesis"
```

```
-1.634975314844e753 0.05102711125709931
```

Test is significant. Reject null hypothesis

Question I l: The mean consumption of food grains among 400 samples middle class consumers is 380 grams per day per person with a standard deviation of 120 grams. A similar sample survey of 600 working class consumers gave a mean of 410 grams with standard deviation of 80 grams. Are we justified in saying that the difference between the averages ofthe two classes is 40? Use 5% level of significance.

In [ ]:

```
#Ho : Iml -m2 —49, Ha : I (ml-m2)l       40
n1 = 400
- 380 sl -
120 n2 =
6eø
m2 = 410                                 = 410
m = np.abs(m1-m2)
s2 = 80
alpha = 0.05
SE = np.sqrt(s1**2/n1+s2**2/n2)
z = (m-40)/SE                            = (m-
4Ø)/SE  p_val  =  2*norm.cdf(z)  print(
'Z_stat:  ' z_stats,_ 'p  value-  '  ,p_val)
if(p_va1<Ø.05):
  print( "Test is significant. Reject null hypothesis") else :
    print( "Test is not significant. Fail to Reject null hypothesis")
```

Question 12: Two populations have same mean but the standard deviation of one is twice that ofthe other. Show that in samples, each of size 500, drawn under simple random conditions, the difference ofthe means will not exceed 0.3 0, where o is the smaller standard deviation.

In [ ]:

```
"
mu1 = mu2, sigma2 = 2*sigma1,n1=n2=500
z = (X1-X2)/sqrt((sigma1^2/n1)+(sigma2^2/n2))
z = (X1-X2)/sqrt((sigma^2/500)+(4sigma^2)/500)
z = (X1-X2)/(sigma/10)
(X1-X2)<=0.30
(X1-X2)/(sigma/10)<=0.3sigma/(sigam/10)
z<=3
Prob (z=3)
"""
norm . cdf(3)
```

out[48] : o. 99865e1e19683699