

Problem Statement 1:

In each of the following situations, state whether it is a correctly stated hypothesis testing problem and why?

1. $H_0: \mu = 25$, $H_1: \mu \neq 25$

2. $H_0: \sigma > 10$, $H_1: \sigma = 10$

3. $H_0: x = 50$, $H_1: x \neq 50$

4. $H_0: p = 0.1$, $H_1: p = 0.5$

5. $H_0: s = 30$, $H_1: s > 30$

1. Yes, because values in both are statement about population or distribution, have equal values and has inequality in ull hypothesis.
2. No, because null hypothesis has an eqaulity claim and alternate hypothesis has inequality.
3. No, because hypothesis is stated in terms of statistics and not sample data
4. No, because values in both hypothesis is different and has equal sign.
4. No, because hypothesis are always statements about population or distribution and not about sample

Problem Statement 2:

The college bookstore tells prospective students that the average cost of its textbooks is Rs. 52 with a standard deviation of Rs. 4.50. A group of smart statistics students thinks that the average cost is higher. To test the bookstore's claim against their alternative, the students will select a random sample of size 100. Assume that the mean from their random sample is Rs. 52.80. Perform a hypothesis test at the 5% level of significance and state your decision.

```
In [1]: from scipy.stats import norm
import numpy as np
import math
import scipy.stats as stats
from scipy.stats import chi2_contingency

p_mean = 52
p_std = 4.50
n = 100
sample_mean = 52.80

SE = p_std/n**0.5
Z = (sample_mean-p_mean)/SE
print(f"Z score is:{Z}")
alpha=0.05 #test_significance
print(f"Critical region is {norm.ppf(alpha/2)}, {-norm.ppf(alpha/2)}")
```

Z score is:1.7777777777777715

Critical region is -1.9599639845400545, 1.9599639845400545

*** From above we can conclude we can accept null hypothesis since z_score<critical region**

Problem Statement 3:

A certain chemical pollutant in the Genesee River has been constant for several years with mean $\mu = 34$ ppm (parts per million) and standard deviation $\sigma = 8$ ppm. A group of factory representatives whose companies discharge liquids into the river is now claiming that they have lowered the average with improved filtration devices. A group of environmentalists will test to see if this is true at the 1% level of significance. Assume \ that their sample of size 50 gives a mean of 32.5 ppm. Perform a hypothesis test at the 1% level of significance and state your decision.

```
In [2]: p_mean = 34
p_std = 8
n = 50
sample_mean = 32.5

SE = p_std/n**0.5 #standard Error
Z = (sample_mean-p_mean)/SE
print(f"Z score is:{Z}")
alpha=0.01
print(f"Critical region is {norm.ppf(alpha/2)}, {-norm.ppf(alpha/2)}")
```

Z score is:-1.3258252147247767

Critical region is -2.575829303548901, 2.575829303548901

*** From above we can conclude we can accept null hypothesis since critical_region<Z_score**

Problem Statement 4:

Based on population figures and other general information on the U.S. population, suppose it has been estimated that, on average, a family of four in the U.S. spends about \$1135 annually on dental expenditures. Suppose further that a regional dental association wants to test to determine if this figure is accurate for their area of country. To test this, 22 families of 4 are randomly selected from the population in that area of the country and a log is kept of the family's dental expenditure for one year. The resulting data are given below. Assuming, that dental expenditure is normally distributed in the population, use the data and an alpha of 0.5 to test the dental association's hypothesis.

- 1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851, 930, 730, 699, 872, 913, 944, 954, 987, 1695, 995, 1003, 994

```
In [3]: given_data=[1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851, 930, 730, 699, 872, 913, 944, 954, 987, 1695, 995, 1003, 994]
p_mean =1135
sample_std = np.std(given_data)
n=22
sample_mean = np.sum(given_data,axis=0)/len(given_data)
SE = sample_std/n**0.5
alpha = 0.5
test_1 = (sample_mean-p_mean)/SE
print(f"t_Score is{test_1}")
print(f"Critical Region is {stats.t.ppf((alpha/2),df=21)} {stats.t.ppf(1-(alpha/2),df=21)}")
```

```
t_Score is-2.070747228595759
Critical Region is -0.6863519891164291 0.6863519891164291
```

*** We reject null hypothesis alpha=0.5 since it lies within critical region**

Problem Statement 5:

In a report prepared by the Economic Research Department of a major bank the Department manager maintains that the average annual family income on Metropolis is 48432.

What do you conclude about the validity of the report if a random sample of 400 families shows and average income of 48,574 with a standard deviation of 2000?

```
In [4]: p_mean = 48432
p_std = 2000
n =400
sample_mean =48574

SE = p_std/n**0.5
Z = (sample_mean-p_mean)/SE
alpha=0.05
print(f"Critical region is {norm.ppf(alpha/2)} {-norm.ppf(alpha/2)}")
```

```
Critical region is -1.9599639845400545 1.9599639845400545
```

***We accept null hypothesis at alpha=0.05 since z_score<critical_region**

Problem Statement 6:

Suppose that in past years the average price per square foot for warehouses in the United States has been 32.28. A national real estate investor wants to determine whether that figure has changed now. The investor hires a researcher who randomly samples 19 warehouses that are for sale across the United States and finds that the mean price per square foot is 31.67, with a standard deviation of 1.29. assume that the prices of warehouse footage are normally distributed in population. If the researcher uses a 5% level of significance, what statistical conclusion can be reached? What are the hypotheses?

```
In [5]: p_mean =32.28
n=19
sample_mean =31.67
sample_std =1.29
alpha =0.05

SE=sample_std/(n**0.5)
t=(sample_mean-p_mean)/SE
print(f"t_score is {round((t),1)}")
print(f"Critical region is {round(stats.t.ppf((alpha/2),df=18),1)} {-round(stats.t.ppf((alpha/2),df=18),1)}")

t_score is -2.1
Critical region is -2.1 2.1
```

***we can reject null hypothesis at alpha=0.05 since it lies within critical region**

Problem Statement 7:

Fill in the blank spaces in the table and draw your conclusions from it.

Acceptance region	Sample size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10			
$48 < \bar{x} < 52$	10			
$48.81 < \bar{x} < 51.9$	16			
$48.42 < \bar{x} < 51.58$	16			

i) Acceptance region $48.5 < x < 51.5$

```
In [6]: # Calculate Beta at Mu1 = 52
n1 = 10
Sig = 2.5
Mu1 = 52

p11 = (48.5 - Mu1)/(Sig/math.sqrt(n1))
print(p11)
p12 = (51.5 - Mu1)/(Sig/math.sqrt(n1))
print(p12)
# As p12 > p11, so our z score lies in between these two p12 < z < p11.
# Find probability at these z score
P11 = 0
P12 = 0.2643
# now Beta = p12 - p11
Beta11 = P12 - P11
print(f"Beta at Mu1 = 52 is : {Beta11}")
```

```
-4.427188724235731
-0.6324555320336759
Beta at Mu1 = 52 is : 0.2643
```

```
In [7]: # Calculate Beta at Mu2 = 50.5
n1 = 10
Sig = 2.5
Mu2 = 50.5

p13 = (48.5 - Mu2)/(Sig/math.sqrt(n1))
print(p13)
p14 = (51.5 - Mu2)/(Sig/math.sqrt(n1))
print(p14)

# so our z score lies in between these two p13 < z < p14.
# Find probability at these z score
P13 = 0.0057
P14 = 0.8962
# now Beta = p13 + (1 - p14)
Beta12 = P13 + (1 - P14)
print(f"Beta at Mu2 = 50.5 is : {Beta12}")
```

```
-2.5298221281347035
1.2649110640673518
Beta at Mu2 = 50.5 is : 0.1095
```

ii) $48 < x < 51$

```
In [8]: # Calculate Beta at Mu1 = 52
n2 = 10
Sig = 2.5
Mu1 = 52

p21 = (48.0 - Mu1)/(Sig/math.sqrt(n2))
print(p21)
p22 = (51.0 - Mu1)/(Sig/math.sqrt(n2))
print(p22)

# As p22 > p21, so our z score lies in between these two p22 < z < p21.
# Find probability at these z score
P21 = 0
P22 = 0.1038
# now Beta = p22 - p21
Beta21 = P22 - P21
print(f"Beta at Mu2 = 52 is : {Beta21}")
```

-5.059644256269407
-1.2649110640673518
Beta at Mu2 = 52 is : 0.1038

```
In [9]: # Calculate Beta at Mu2 = 50.5
n2 = 10
Sig = 2.5
Mu2 = 50.5

p23 = (48 - Mu2)/(Sig/math.sqrt(n2))
print(p23)
p24 = (51 - Mu2)/(Sig/math.sqrt(n2))
print(p24)

# so our z score lies in between these two p13 < z < p14.
# Find probability at these z score
P23 = 0.0008
P24 = 0.7357
# now Beta = p13 +(1 - p14)
Beta22 = P23 +(1- P24)
print(f"Beta at Mu2 = 50.5 is : {Beta22}")
```

-3.1622776601683795
0.6324555320336759
Beta at Mu2 = 50.5 is : 0.2651

iii) $48.81 < x < 51.9$

```
In [10]: # Calculate Beta at Mu1 = 52
n3 = 16
Sig = 2.5
Mu1 = 52

p31 = (48.81 - Mu1)/(Sig/math.sqrt(n3))
print(p31)
p32 = (51.9 - Mu1)/(Sig/math.sqrt(n3))
print(p32)

# so our z score lies in between these two p31 < z < p32.
# Find probability at these z score
P31 = 0.4364
P32 = 0
# now Beta = p31 - p32
Beta31 = P31 - P32
print(f"Beta at Mu2 = 52 is : {Beta31}")
```

```
-5.1039999999999965
-0.160000000000000228
Beta at Mu2 = 52 is : 0.4364
```

```
In [11]: # Calculate Beta at Mu2 = 50.5
n3 = 16
Sig = 2.5
Mu2 = 50.5

p33 = (48.81 - Mu2)/(Sig/math.sqrt(n3))
print(p33)
p34 = (51.9 - Mu2)/(Sig/math.sqrt(n3))
print(p34)

# so our z score lies in between these two p33 < z < 1- p34.
# Find probability at these z score
P33 = 0.0032
P34 = 0.9875
# now Beta = P33 + 1- P34
Beta32 = P33 + 1 - P34
print(f"Beta at Mu2 = 50.5 is : {Beta32}")
```

```
-2.7039999999999996
2.2399999999999975
Beta at Mu2 = 50.5 is : 0.015700000000000047
```

iv) $48.42 < x < 51.58$

```
In [12]: # Calculate Beta at Mu1 = 52
n4 = 16
Sig = 2.5
Mu1 = 52

p41 = (48.42 - Mu1)/(Sig/math.sqrt(n4))
print(p41)
p42 = (51.58 - Mu1)/(Sig/math.sqrt(n4))
print(p42)

# As P42 > P41, so our z score lies in between these two p42 < z < p41.
# Find probability at these z score
P41 = 0.0
P42 = 0.2514
# now Beta = p31 - p32
Beta41 = P42 - P41
print(f"Beta at Mu2 = 52 is : {Beta41}")
```

-5.727999999999997
-0.67200000000000027
Beta at Mu2 = 52 is : 0.2514

```
In [13]: # Calculate Beta at Mu2 = 50.5
n4 = 16
Sig = 2.5
Mu2 = 50.5

p43 = (48.42 - Mu2)/(Sig/math.sqrt(n4))
print(p43)
p44 = (51.58 - Mu2)/(Sig/math.sqrt(n4))
print(p44)

# so our z score lies in between these two p33 < z < 1- p34.
# Find probability at these z score
P43 = 0.0035
P44 = 0.9875
# now Beta = P43 + 1- P44
Beta42 = P43 + (1 - P44)
print(f"Beta at Mu2 = 50.5 is : {Beta42}")
```

-2.7039999999999996
2.2399999999999975
Beta at Mu2 = 50.5 is : 0.015999999999999955

Problem Statement 8:

Find the t-score for a sample size of 16 taken from a population with mean 10 when the sample mean is 12 and the sample standard deviation is 1.5.


```
In [14]: #t_score = ?
n = 16
p_mean = 10
sample_mean = 12
sample_std = 1.5

SE = sample_std/(n**0.5)
t = (sample_mean-p_mean)/SE
print(f"t_score is {round((t),1)}")
```

t_score is 5.3

Problem Statement 9:

Find the t-score below which we can expect 99% of sample means will fall if samples of size 16 are taken from a normally distributed population.

```
In [15]: n= 16
alpha=(1-0.99)/2
print(f"t_score is {stats.t.ppf(1-alpha,df=15)}")
```

t_score is 2.946712883338615

Problem Statement 10:

If a random sample of size 25 drawn from a normal population gives a mean of 60 and a standard deviation of 4, find the range of t-scores where we can expect to find the middle 95% of all sample means. Compute the probability that $(-t_{0.05} < t < t_{0.10})$.

```
In [16]: n=25
std=4
mean=60
alpha=(1-0.95)/2
t_score=stats.t.ppf(1-alpha,df=24)
print(f"Range is : {mean+t_score*(std/(n**0.5))} {mean-t_score*(std/(n**0.5))}")
```

Range is : 61.651118849302414 58.348881150697586

```
In [17]: p=stats.t.cdf(0.1,df=24)-stats.t.cdf(-0.05,df=24)
print(f"probability that (-t0.05 < t < t0.10) is {p}")
```

probability that $(-t_{0.05} < t < t_{0.10})$ is 0.05914441613731247

Problem Statement 11:

Two-tailed test for difference between two population means Is there evidence to conclude that the number of people travelling from Bangalore to Chennai is different from the number of people travelling from Bangalore to Hosur in a week, given the following:

- **Population 1:---** Bangalore to Chennai

```
n1 = 1200
x1 = 452
s1 = 212
```

- **Population 2:----** Bangalore to Hosur

```
n2 = 800
x2 = 523
s2 = 185
```

```
In [18]: n1 = 1200
x1 = 452
s1 = 212
n2 = 800
x2 = 523
s2 = 185
s_1=s1**2
s_2=s2**2
alpha=0.05
se=((s_1/n1)+(s_2/n2))**0.5
z_score=(x1-x2)/se
print(f"Z_Score is {z_score}")
print(f"Critical region is {norm.ppf(alpha/2)} {-norm.ppf(alpha/2)}")
```

```
Z_Score is -7.926428526759299
Critical region is -1.9599639845400545 1.9599639845400545
```

***We reject null hypothesis since it lies within critical region at alpha=5%.So,number of people travelling from Bangalore to Chennai is different from the number of people travelling from Bangalore to Hosur in a week**

Problem Statement 12:

Is there evidence to conclude that the number of people preferring Duracell battery is different from the number of people preferring Energizer battery, given the following:

- **Population 1:----** Duracell

```
n1 = 100
x1 = 308
s1 = 84
```

- **Population 2:----** Energizer

```
n2 = 100
x2 = 254
s2 = 67
```

```
In [19]: n1 = 100
x1 = 308
s1 = 84
n2 = 100
x2 = 254
s2 = 67
s_1=s1**2
s_2=s2**2
alpha=0.05
SE=((s_1/n1)+(s_2/n2))**0.5
z_score=(x1-x2)/SE
print(f"Z_Score is {z_score}")
print(f"Critical region is {norm.ppf(alpha/2)} {-norm.ppf(alpha/2)}")
```

Z_Score is 5.025702668336442

Critical region is -1.9599639845400545 1.9599639845400545

***We reject null hypothesis since it lies within critical region at alpha=5%.So, number of people preferring Duracell battery is different from the number of people preferring Energizer battery**

Problem Statement 13:

Pooled estimate of the population variance Does the data provide sufficient evidence to conclude that average percentage increase in the price of sugar differs when it is sold at two different prices?

- **---Population 1: ----**

Price of sugar=Rs. 27.50

n1 = 14

x1 = 0.317%

s1 = 0.12%

- **-----Population 2: -----**

Price of sugar=Rs. 20.00

n2 = 9

x2 = 0.21%

s2 = 0.11%

```
In [20]: n1 = 14
x1 = 0.317
s1 = 0.12
n2 = 9
x2 = 0.21
s2 = 0.11
s_1=s1**2
s_2=s2**2
s=((n1-1)*s_1)+((n2-1)*s_2)
n=(n1+n2-2)
se=(s/n)**0.5
n_1=((1/n1)+(1/n2))**0.5
t_score=(x1-x2)/se*n_1

print(f"T_score is {t_score}")
print(f"Critical Region is {stats.t.ppf(1-0.05,df=n)}")
```

```
T_score is 0.3931089218182991
Critical Region is 1.7207429028118775
```

***We accept null hypothesis at alpha=5%.So,average price do not increase**

Problem Statement 14:

The manufacturers of compact disk players want to test whether a small price reduction is enough to increase sales of their product. Is there evidence that the small price reduction is enough to increase sales of compact disk players?

- **Population 1: Before reduction**

```
n1 = 15
x1 = Rs. 6598
s1 = Rs. 844
```

- **Population 2: After reduction**

```
n2 = 12
x2 = RS. 6870
s2 = Rs. 669
```

```
In [21]: n1 = 15
x1 = 6598
s1 = 844
n2 = 12
x2 = 6870
s2 = 669
s_1=s1**2
s_2=s2**2

s=((n1-1)*s_1)+((n2-1)*s_2)
n=(n1+n2-2)
se=(s/n)**0.5
n_1=((1/n1)+(1/n2))**0.5
t_score=(x1-x2)/se*n_1

print(f"T_score is {t_score}")
print(f"Critical Region is {stats.t.ppf(0.05,df=n)}")
```

T_score is -0.1364745051598569
Critical Region is -1.708140761251899

***We accept null hypothesis at alpha=5%.So average price remains same**

Problem Statement 15:

Comparisons of two population proportions when the hypothesized difference is zero Carry out a two-tailed test of the equality of banks' share of the car loan market in 1980 and 1995.

- Population 1: 1980**

n1 = 1000
x1 = 53
p 1 = 0.53

- Population 2: 1985**

n2 = 100
x2 = 43
p 2= 0.53

```
In [22]: n1 = 1000
x1 = 53
p1 = 0.53
n2 = 100
x2 = 43
p2 = 0.53
p = (x1+x2)/(n1+n2)

n = (1/n1)+(1/n2)
p_1 = p*(1-p)
Z = (p1-p2)/((p_1*n)**0.5)
print(f"Z_score is {Z}")
print(f"Critical region is {norm.ppf(0.05)}")
```

```
Z_score is 0.0
Critical region is -1.6448536269514729
```

***We can't reject null hypothesis at alpha=10%**

Problem Statement 16:

Carry out a one-tailed test to determine whether the population proportion of traveler's check buyers who buy at least \$2500 in checks when sweepstakes prizes are offered is at least 10% higher than the proportion of such buyers when no sweepstakes are on.

- **Population 1: With sweepstakes**

```
n1 = 300
x1 = 120
p = 0.40
```

- **Population 2: No sweepstakes**

```
n2 = 700
x2 = 140
p_2 = 0.20
```

```
In [23]: n1 = 300
x1 = 120
p1 = 0.40
n2 = 700
x2 = 140
p2 = 0.20
p = (x1+x2)/(n1+n2)

n = (1/n1)+(1/n2)
p_1 = p*(1-p)
Z = (p1-p2-0.1)/((p_1*n)**0.5)
print(f"Z_score is {Z}")
print(f"Critical region is {-norm.ppf(0.05)}")
```

Z_score is 3.303749523611152
Critical region is 1.6448536269514729

***We reject null hypothesis at alpha=5%**

Problem Statement 17:

A die is thrown 132 times with the following results:

Number turned up: 1, 2, 3, 4, 5, 6

Frequency: 16, 20, 25, 14, 29, 28

Is the die unbiased? Consider the degrees of freedom as $p - 1$.

```
In [24]: f_obs = [16, 20, 25, 14, 29, 28]
f_exp = [22, 22, 22, 22, 22, 22]
result = stats.chisquare(f_obs, f_exp)
print(f"Chi square value is {result[0]} and p-value is {result[1]}")
print('Dias is unbiased.')
```

Chi square value is 9.0 and p-value is 0.1090641579497725
Dias is unbiased.

Problem Statement 18: ¶

In a certain town, there are about one million eligible voters. A simple random sample of 10,000 eligible voters was chosen to study the relationship between gender and participation in the last election. The results are summarized in the following 2X2 (read two by two) contingency table:

	Men	Women
Voted	2792	3591
Not voted	1486	2131

We would want to check whether being a man or a woman (columns) is independent of having voted in the last election (rows). In other words, is "gender and voting independent"?

```
In [25]: observed_voted_men=2792
observed_voted_women=3591
observed_not_voted_men=1486
observed_not_voted_women=2131
total_voted=2792+3591
total_not_voted=1486+2131
total_men=2792+1486
total_women=3591+2131
expected_voted_men=(total_voted*total_men)/10000
expected_voted_women=(total_voted*total_women)/10000
expected_not_voted_men=(total_not_voted*total_men)/10000
expected_not_voted_women=(total_not_voted*total_women)/10000
chisquare1=((observed_voted_women-expected_voted_women)**2)/expected_voted_women
chisquare2=((observed_voted_men-expected_voted_men)**2)/expected_voted_men
chisquare3=((observed_not_voted_men-expected_not_voted_men)**2)/expected_not_voted_men
chisquare4=((observed_not_voted_women-expected_not_voted_women)**2)/expected_not_voted_women
chisquare=chisquare1+chisquare2+chisquare3+chisquare4
print(f"Chi Square value is {chisquare}")
print(f"Critical region with alpha=0.05 is 3.84")
print("We reject null hypothesis.It is not gender and voting independent")
```

Chi Square value is 6.660455899328067

Critical region with alpha=0.05 is 3.84

We reject null hypothesis.It is not gender and voting independent

Problem Statement 19:

A sample of 100 voters are asked which of four candidates they would vote for in an election. The number supporting each candidate is given below:

Higgins	Reardon	White	Charlton
41	19	24	16

Do the data suggest that all candidates are equally popular? [Chi-Square = 14.96, with 3 df, p 0.05 .]

```
In [26]: #print(np.mean([41,19,24,16]))
obs=[41,19,24,16]
exp=[25,25,25,25]
result=stats.chisquare(obs,exp)
print(f"Chi Square value is {result[0]}")
print(f"Critical region with 3df and alpha=0.05 is 7.82")
print("We reject null hypothesis. All candidates are not equally popular")
```

Chi Square value is 14.959999999999999

Critical region with 3df and alpha=0.05 is 7.82

We reject null hypothesis. All candidates are not equally popular

Problem Statement 20:

Children of three ages are asked to indicate their preference for three photographs of adults. Do the data suggest that there is a significant relationship between age and photograph preference? What is wrong with this study? [Chi-Square = 29.6, with 4 df: $p < 0.05$].

###		Photograph		
		A	B	C
Age of child	5 – 6 years	18	22	20
	7 – 8 years	2	28	40
	9 – 10 years	20	10	40

```
In [27]: obs=([[18,22,20],[2,28,40],[20,10,40]])
result=chi2_contingency(obs)
print(f"Chi Square value is {result[0]}")
print(f"Critical region with 4df and alpha=0.001 is 18.47")
print("We reject null hypothesis. There is significant relationship between age and photograph preference")
```

Chi Square value is 29.603174603174608

Critical region with 4df and alpha=0.001 is 18.47

We reject null hypothesis. There is significant relationship between age and photograph preference

Problem Statement 21:

A study of conformity using the Asch paradigm involved two conditions: one where one confederate supported the true judgement and another where no confederate gave the correct response.

	Support	No support
Conform	18	40
Not conform	32	10

Is there a significant difference between the "support" and "no support" conditions in the frequency with which individuals are likely to conform? [Chi-Square = 19.87, with 1 df: $p < 0.05$].

```
In [28]: obs=np.array([[18,40],[32,10]])
result=chi2_contingency(obs)
print(f"Chi Square value is {result[0]}")
print(f"Critical region with 1df and alpha=0.001 is 10.83")
print("We reject null hypothesis. So, there is significant difference between the 'support' and 'no support' conditions in the frequency with which individuals are likely to conform")
```

Chi Square value is 18.10344827586207

Critical region with 1df and alpha=0.001 is 10.83

We reject null hypothesis. So, there is significant difference between the 'support' and 'no support' conditions in the frequency with which individuals are likely to conform

Problem Statement 22:

We want to test whether short people differ with respect to their leadership qualities (Genghis Khan, Adolf Hitler and Napoleon were all stature-deprived, and how many midget MP's are there?) The following table shows the frequencies with which 43 short people and 52 tall people were categorized as "leaders", "followers" or as "unclassifiable". Is there a relationship between height and leadership qualities? [Chi-Square = 10.71, with 2 df: $p < 0.01$].

##	Height	
	Short	Tall
Leader	12	32
Follower	22	14
Unclassifiable	9	6

```
In [29]: obs=([[12,32],[22,14],[9,6]])
result=chi2_contingency(obs)
print(f"Chi Square value is {result[0]}")
print(f"Critical region with 2df and alpha=0.001 is 13.82")
print("We accept null hypothesis.there is no relationship between height and leadership qualities")
```

Chi Square value is 10.712198008709638

Critical region with 2df and alpha=0.001 is 13.82

We accept null hypothesis.there is no relationship between height and leadership qualities

Problem Statement 23:

Each respondent in the Current Population Survey of March 1993 was classified as employed, unemployed, or outside the labor force. The results for men in California age 35- 44 can be cross-tabulated by marital status, as follows:

	Married	Widowed, divorced or separated	Never married
Employed	679	103	114
Unemployed	63	10	20
Not in labor force	42	18	25

Men of different marital status seem to have different distributions of labor force status. Or is this just chance variation? (you may assume the table results from a simple random sample.)

```
In [30]: obs = np.array([[679,103,114], [63,10,20],[42,18,25]])
result=chi2_contingency(obs)
print(f"Chi Square value is {result[0]}")
print(f"Critical region with 4df and alpha=0.001 is 18.47")
print("We reject null hypothesis at alpha=0.001 .there is relationship between ma
```

Chi Square value is 31.61310319407798

Critical region with 4df and alpha=0.001 is 18.47

We reject null hypothesis at alpha=0.001 .there is relationship between martial status and employment status