In [1]:

```
import numpy as np
import scipy.stats as stats
import statistics as stat
```

Problem Statement 1:

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

1. $H0: \mu = 25, H1: \mu \neq 25$ Correct. Hypothesis is either to be or not to be. Mean is 25 or it is not.
2. $H0: \sigma > 10, H1: \sigma = 10$ Not Correct.Hypothesis is either to be or not to be. Standard Deviation here defines Null hypothesis as values greater than 10.
3. $H0: x = 50, H1: x \neq 50$ Correct.Hypothesis is either to be or not to be.
4. $H0: p = 0.1, H1: p = 0.5$ Correct.
5. $H0: s = 30, H1: s > 30$ Correct.Hypothesis is either to be or not to be.Here sample Standard Deviation is defined correctly.

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 [2]:

```
μ = 52
σ = 4.50
n = 100
s_mean = 52.80
α = 0.05
print("H0: μ = 52 (The average cost of the book is same)\nH1: μ ╘ 52(The average price
print(f"Z at {α/2}: +-1.96")
z = (s_mean - μ)/(σ/(np.sqrt(n)))
print(f"Here z {z} is less than 1.96. So we accept Null hypothesis.\nTherefor The avera
```

```
H0: μ = 52 (The average cost of the book is same)
H1: μ ╘ 52(The average price of the book is changed.)
Z at 0.025: +-1.96
Here z 1.7777777777777715 is less than 1.96. So we accept Null hypothesis.
Therefor The average cost of the book is same
```

Problem Statement 3:

A certain chemical pollutant in the Genesee River has been constant for several years with mean μ = 34 ppm (parts per million) and standard deviation σ = 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 [3]:

```
μ = 34
σ = 8
α = 0.01
s_mean = 32.5
n = 50
print("H0: μ = 34 (The pollutant level is same )\nH1: μ ╘ 34(The pollutant level is not
print(f"Z at {α/2}: -2.57")
z = (s_mean - μ)/(σ/(np.sqrt(n)))
print(f"Here z {z} is less than 1.96. So we accept Null hypothesis.\nTherefor The pollu
```

```
H0: μ = 34 (The pollutant level is same )
H1: μ ╘ 34(The pollutant level is not the same.)
Z at 0.005: -2.57
Here z -1.3258252147247767 is less than 1.96. So we accept Null hypothesis.
Therefor The pollutant level is same
```

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 [4]:

```
x = [1008, 812, 1117, 1323, 1308, 1415, 831, 1021, 1287, 851, 930, 730, 699, 872, 913,
s_mean = np.mean(x)
std = np.std(x)
α = 0.5
n = 22
df = n-1
μ = 1135
c_i = 1-(α/2)
print("H0: μ = 34 (The pollutant level is same )\nH1: μ ╘ 34(The pollutant level is not
print(f"t at {α/2} and {df}: 0.686")
t = (s_mean - μ)/(std/(np.sqrt(n)))
print("C.I:",c_i)
print(f"Here t {t} is not in range(+-0.686). So we reject Null hypothesis.\nTherefore T
```

```
H0: μ = 34 (The pollutant level is same )
H1: μ ╘ 34(The pollutant level is not the same.)
t at 0.25 and 21: 0.686
C.I: 0.75
Here t -2.070747228595759 is not in range(+-0.686). So we reject Null hypoth
esis.
Therefore The pollutant level is not same
```

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 [5]:

```
μ =48432
n=400
x_mean = 48574
σ=2000
α = 0.05
z= (x_mean-μ)/(σ/np.sqrt(400))
z_α_2 = -1.96# value of z alpha/2 : 0.025
print(f"Value of Z statistics:{z}\nSince -1.96< Z <1.96.\nAccept Null Hypothesis")
```

```
Value of Z statistics:1.42
Since -1.96< Z <1.96.
Accept Null Hypothesis
```

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 [6]:

```
μ = 32.28
n = 19
s_mean = 31.67
s = 1.29
α = 0.05
t = round((s_mean-μ)/(s/np.sqrt(n)),1)
t_α_2 = 2.101 # t alpha/2(0.025) at df:18 with CP 0.975
print(f"Value of t statistics : {t}\nSince it must satisfy -2.1 < t < 2.1\nWe reject Hy
```

```
Value of t statistics : -2.1
Since it must satisfy -2.1 < t < 2.1
We reject Hypothesis.
```

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			

1) 48.5 < s_mean < 51.5

In [7]:

```
# Calculate Beta at Mu1 = 52
n1 = 10
σ =2.5
μ =52
z_11 = (48.5 - μ)/(σ/np.sqrt(n1))
z_12 = (51.5 - μ)/(σ/np.sqrt(n1))
P11 = 0
P12 = 0.2643
print(f"Z statistics value at {z_11} is {P11}")
print(f"Z statistics value at {z_12} is {P12}")
ß11 = P12 - P11
print(f"ß at μ = 52 is : {ß11}")
```

```
Z statistics value at -4.427188724235731 is 0
Z statistics value at -0.6324555320336759 is 0.2643
ß at μ = 52 is : 0.2643
```

In [8]:

```
# Calculate Beta at μ = 50.5
n1 = 10
σ =2.5
μ = 50.5
z_11 = (48.5 - μ)/(σ/np.sqrt(n1))
z_12 = (51.5 - μ)/(σ/np.sqrt(n1))
P13 = 0.005
P14 = 0.89617
print(f"Z statistics value at {z_11} is {P11}")
print(f"Z statistics value at {z_12} is {P12}")
ß12 = P13 + (1-P14)
print(f"ß at μ = {μ} is : {ß12}")
```

```
Z statistics value at -2.5298221281347035 is 0
Z statistics value at 1.2649110640673518 is 0.2643
ß at μ = 50.5 is : 0.10882999999999998
```

2) 48 < sapmle mean < 52

In [9]:

```
# Calculate Beta at μ = 48
n1 = 10
σ =2.5
μ = 52
z_21 = (48.0 - μ)/(σ/np.sqrt(n1))
z_22 = (52.0 - μ)/(σ/np.sqrt(n1))
P21 = 0.0
P22 = 0.00
print(f"Z statistics value at {z_21} is {P21}")
print(f"Z statistics value at {z_22} is {P22}")
ß21 = P22 - P21
print(f"ß at μ = {μ} is : {ß21}")
```

```
Z statistics value at -5.059644256269407 is 0.0
Z statistics value at 0.0 is 0.0
ß at μ = 52 is : 0.0
```

In [10]:

```
# Calculate Beta at μ = 50.5
n1 = 10
σ = 2.5
μ = 50.5
z_11 = (48.0 - μ)/(σ/np.sqrt(n1))
z_12 = ( 52.0 - μ)/(σ/np.sqrt(n1))
P13 = 0.00079
P14 = 0.97062
print(f"Z statistics value at {z_11} is {P13}")
print(f"Z statistics value at {z_12} is {P14}")
ß12 = P13 + (1-P14)
print(f"ß at μ = {μ} is : {ß12}")
```

```
Z statistics value at -3.1622776601683795 is 0.00079
Z statistics value at 1.8973665961010278 is 0.97062
ß at μ = 50.5 is : 0.030169999999999996
```

3) 48.81 < z < 51.9

In [11]:

```
# Calculate Beta at μ = 52
n1 = 16
σ = 2.5
μ = 52
z_21 = ( 48.81 - μ)/(σ/np.sqrt(n1))
z_22 = (51.9 - μ)/(σ/np.sqrt(n1))
P21 = 0
P22 = 00.43640
print(f"Z statistics value at {z_21} is {P21}")
print(f"Z statistics value at {z_22} is {P22}")
ß21 = P22 - P21
print(f"ß at μ = {μ} is : {ß21}")
```

```
Z statistics value at -5.103999999999965 is 0
Z statistics value at -0.16000000000000228 is 0.4364
ß at μ = 52 is : 0.4364
```

In [12]:

```
# Calculate Beta at μ = 50.5
n1 = 16
σ = 2.5
μ = 50.5
z_11 = (48.81  - μ)/(σ/np.sqrt(n1))
z_12 = (51.9 - μ)/(σ/np.sqrt(n1))
P13 = 0.0034
P14 = 0.98713
print(f"Z statistics value at {z_11} is {P13}")
print(f"Z statistics value at {z_12} is {P14}")
ß12 = P13 + (1-P14)
print(f"ß at μ = {μ} is : {ß12}")
```

```
Z statistics value at -2.703999999999996 is 0.0034
Z statistics value at 2.2399999999999975 is 0.98713
ß at μ = 50.5 is : 0.01627000000000005
```

4) 48.42 < z < 51.58

In [13]:

```
# Calculate Beta at Mu1 = 52
n1 = 16
σ = 2.5
μ = 52
z_11 = (48.42 - μ)/(σ/np.sqrt(n1))
z_12 = (51.58 - μ)/(σ/np.sqrt(n1))
P11 = 0
P12 = 0.25143
print(f"Z statistics value at {z_11} is {P11}")
print(f"Z statistics value at {z_12} is {P12}")
ß11 = P12 - P11
print(f"ß at μ = 52 is : {ß11}")
```

```
Z statistics value at -5.727999999999997 is 0
Z statistics value at -0.6720000000000027 is 0.25143
ß at μ = 52 is : 0.25143
```

In [14]:

```
# Calculate Beta at μ = 50.5
n1 = 16
σ = 2.5
μ = 50.5
z_11 = (48.42 - μ)/(σ/np.sqrt(n1))
z_12 = (51.58 - μ)/(σ/np.sqrt(n1))
P13 = 0.00045
P14 = 0.95728
print(f"Z statistics value at {z_11} is {P13}")
print(f"Z statistics value at {z_12} is {P14}")
ß12 = P13 + (1-P14)
print(f"ß at μ = {μ} is : {ß12}")
```

```
Z statistics value at -3.327999999999997 is 0.00045
Z statistics value at 1.7279999999999973 is 0.95728
ß at μ = 50.5 is : 0.04316999999999998
```

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 [15]:

```
n= 16
μ = 10
s_mean = 12
s = 1.5
t = (s_mean-μ)/(s/np.sqrt(n))
print(f"T statistics value : {round(t,1)}")
```

```
T statistics value : 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 [16]:

```
n= 16
α = 0.01
df = n-1
print("The t-score below which we can expect 99% of sample means will fall if samples c
```

```
The t-score below which we can expect 99% of sample means will fall if sampl
es of size 16 are taken from a normally distributed population is 2.6
```

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 (−t0.05 <t<t0.10).

In [17]:

```
n = 25
α = 0.05
s_mean = 60
s = 4
range_ = s_mean + (stats.t.ppf(1-α,df = 24))*(s/np.sqrt(n))
_range = s_mean - (stats.t.ppf(1-α,df = 24))*(s/np.sqrt(n))
print("Range of t-Score:",_range,range_)
```

```
Range of t-Score: 58.63129433607246 61.36870566392754
```

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
α = 0.05
z = (x1-x2)/np.sqrt((s1**2/n1)+(s2**2/n2))
z_α_2 = 1.96 # value at z 0.025
print(f"Since value range is -{z_α_2} < Z < {z_α_2}\nValue of Z-statistics is {round(z,
```

```
Since value range is -1.96 < Z < 1.96
Value of Z-statistics is -7.9
So we reject the Null Hypothesis
```

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
z = (x1-x2)/np.sqrt((s1**2/n1)+(s2**2/n2))
z_α_2 = 1.96 # value at z 0.025
print(f"Since value range is -{z_α_2} < Z < {z_α_2}\nValue of Z-statistics is {round(z,
```

```
Since value range is -1.96 < Z < 1.96
Value of Z-statistics is 5.0
So we reject the Null Hypothesis
```

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
df = (n1+n2) - 2
α = 0.05
print(f"t at df {df} and α/2 0.025 is: {stats.t.ppf(1-α,df=21)}")
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"Since t: {t_score}. So we accept the Null hypothesis.")
```

```
t at df 21 and α/2 0.025 is: 1.7207429028118775
Since t: 0.3931089218182991. So we accept the Null hypothesis.
```

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]:

```
# Population 1: Before reduction
n1 = 15
x1 = 6598
s1 = 844
# Population 2: After reduction
n2 = 12
x2 = 6870
s2 = 669
# Using T test
nume = abs(x1-x2)
den1 = np.sqrt((((n1-1)*pow(s1,2))+((n2-1)*pow(s2,2)))/(n1+n2+-2))
den2 = np.sqrt((1/n1)+(2/n2))
t = nume/den1*den2
df = (n1+n2) - 2
α = 0.05
print(f"t at df {df} and α/2 0.025 is: {stats.t.ppf(1-α,df=df)}")
print(f"Since t statistics value is : {t}\nAnd range is +-1.70\nWe accept the Null Hypo
```

```
t at df 25 and α/2 0.025 is: 1.7081407612518986
Since t statistics value is : 0.17021361344589897
And range is +-1.70
We accept the Null Hypothesis.
```

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= 0.43

In [35]:

```
#Population 1: 1980
n1 = 1000
x1 = 53
p1 = 0.53
#Population 2: 1985
n2 = 100
x2 = 43
p2= 0.53
α_2 = 0.05/2
z = (p1-p2)/np.sqrt(((p1*(1-p1))/n1)+((p2*(1-p2))/n2))
print(f"Value of Z statistics is :{round(z,2)}")
print(f"Z at {α_2} is {-1.96}.\nSo We reject the Null Hypothesis. ")
```

```
Value of Z statistics is :0.0
Z at 0.025 is -1.96.
So We reject the Null Hypothesis.
```

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 as 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 [33]:

```
#Population 1: With sweepstakes
n1 = 300
x1 = 120
p1= 0.40
#Population 2: No sweepstakes
n2 = 700
x2 = 140
p2= 0.20
d=10/100
z = (p1-p2)-d/np.sqrt(((p1*(1-p1))/n1)+((p2*(1-p2))/n2))
#z = (p1-p2)/np.sqrt(((p1*(1-p1))/n1)+((p2*(1-p2))/n2))
print(f"Value of Z statistics is :{round(z,2)}")
print(f"Z at {α_2} is {-1.96}.\nSo We reject the Null Hypothesis. ")
```

```
Value of Z statistics is :-2.92
Z at 0.025 is -1.96.
So We reject the Null Hypothesis.
```

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 [47]:

```
p_of_each_outcome = 132/6
exp_frequency = [p_of_each_outcome,]*6
obj_Frequency = [16, 20, 25, 14, 29, 28]
result = stats.chisquare(f_obs=obj_Frequency,f_exp=exp_frequency)
print(f"ChiSquare Value is :{result[0]}\nP value is :{round(result[1],2)}.\nSo die is U
```

```
ChiSquare Value is :9.0
P value is :0.11.
So die is Unbaised.
```

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

In [78]:

```
obj_freq = [2792,3591]
exp_freq = [4278,5722]
def calChisquare(f_obs=None,f_exp=None):
    sum_ = 0
    for i in range(len(f_obs)):
        sum_ = sum_+(pow((f_obs[i]-f_exp[i]),2)/f_exp[i])
    return round(sum_,2)
result = calChisquare(f_obs=obj_freq,f_exp=exp_freq)
print(f"ChiSquare value is : {result}.\nProbability value : {3.84}\nSo we reject Null H
```

```
ChiSquare value is : 1309.81.
Probability value : 3.84
So we reject Null Hypothesis
```

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:

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

Higgins	Reardon	White	Charlton
41	19	24	16

In [79]:

```
obj_freq = [41,19,24,16]
exp_freq = [100/4]*4
result = calChisquare(f_obs=obj_freq,f_exp=exp_freq)
print(f"ChiSquare value is : {result}.\nProbability value : {7.81}\nSo We reject Null H
```

```
ChiSquare value is : 14.96.
Probability value : 7.81
So We reject Null Hypothesis
```

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 [80]:

```
hiSquare Table that the\nprobability value at df=4 and α at 0.05 is : {9.488}\nChi Square v
```

```
As Given:
We get to know frome the ChiSquare Table that the
probability value at df=4 and α at 0.05 is : 9.488
Chi Square value is 29.6
So We reject Null Hypothesis.
```

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 [88]:

```
obj_freq = [[18,40],[32,10]]
r= stats.chi2_contingency(obj_freq)
print(f"Chi Square value is {r[0]}")
print(f"Critical region with 1df and alpha=0.001 is 10.83")
print("We reject null hypoythesis.So,there is significant difference between the 'suppo
```

```
Chi Square value is 18.10344827586207
Critical region with 1df and alpha=0.001 is 10.83
We reject null hypoythesis.So,there is significant difference between the 's
upport' 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 [93]:

```
obj_freq =[[12,32],[22,14],[9,6]]
r= stats.chi2_contingency(obj_freq)
print(f"Chi Square value is {r[0]}")
print(f"Critical region with 2 df and alpha=0.01 is 13.82")
print("We reject null hypothesis.\nThere is no relationship between height and leaders
```

```
Chi Square value is 10.712198008709638
Critical region with 2 df and alpha=0.01 is 13.82
We reject 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 [98]:

```
obj_freq =[[679,103,114],[63,10,20],[42,18,25]]
r= stats.chi2_contingency(obj_freq)
print(f"Chi Square value is {r[0]}")
print(f"Critical region with 2 df and alpha=0.05 is 5.991")
print("We reject null hypothesis.\nMen of different marital status seem to have different d
```

```
Chi Square value is 31.61310319407798
Critical region with 2 df and alpha=0.05 is 5.991
We reject null hypothesis.
Men of different marital status seem to have different distributions of labo
r force status
```

In [ ]:

```

```