STATISTICS - 3 - Assignment

By Prakash Ghosh

In [1]:
```
# import the libraries
import pandas as pd
import numpy as np
import scipy as sci
import matplotlib.pyplot as plt
import scipy.stats as stat
import math
from scipy.stats import binom
```

Problem Statement 1:

Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher thinks that a diet high in raw cornstarch will have a positive effect on blood glucose levels. A sample of 36 patients who have tried the raw cornstarch diet have a mean glucose level of 108. Test the hypothesis that the raw cornstarch had an effect or not.

- **Hypothesis are**

    1. Null Hypothesis: Raw cornstarch diet does have an affect
    2. Alternative Hypothesis: Raw cornstarch diet does not have any affect

Assumption: Significance level α is 5% i.e. 0.05

In [8]:
```
# Given
μ =100    # population mean of Blood glucose levels for obese patients
σ= 15     # standard deviation of Blood glucose levels for obese patients (population)
N = 36    # No of Samples who have tried the raw cornstarch diet
X= 108    # sample mean (who have tried the raw cornstarch diet)
```

In [9]:
```
print('\nCalculate Z Score Using Formula: (X - μ) / (σ/math.sqrt(N)')
Z = (X - μ) / (σ/math.sqrt(N))
print('\t Z-Score value is :',Z)


print('\nProbability of having mean less than 108:\n\t\t p = stats.norm.cdf(Z)')
p = stat.norm.cdf(Z) # cdf function takes Z- score , returns standard normal probality
print('\t i.e.\t p =',round(p,4))

print('\nThe probability of having mean more than 108:',round(1-p,4))
print('i.e. The probability of having mean more than 108 is lesser than significan Significance level 0.05')

print('\nSo, We can reject the Null Hypothesis')
print('i.e. Raw cornstarch diet does not have an affect')
```

```
Calculate Z Score Using Formula: (X - μ) / (σ/math.sqrt(N)
         Z-Score value is : 3.2

Probability of having mean less than 108:
                 p = stats.norm.cdf(Z)
         i.e.    p = 0.9993

The probability of having mean more than 108: 0.0007
i.e. The probability of having mean more than 108 is lesser than significan Significance level 0.05

So, We can reject the Null Hypothesis
i.e. Raw cornstarch diet does not have an affect
```

Problem Statement 2:

In one state, 52% of the voters are Republicans, and 48% are Democrats. In a second state, 47% of the voters are Republicans, and 53% are Democrats. Suppose a simple random sample of 100 voters are surveyed from each state.
What is the probability that the survey will show a greater percentage of Republican voters in the second state than in the first state?

In [10]:
```
# Given
p_state1_republican =52/100   # Republican voters in the first state 52%
p_state1_democract = 48/100   # Democrats voters in the first state 48%
n_state1 = 100                # No. of samples from first state=100

p_state2_republican =47/100   # Republican voters in the second state 47%
p_state2_democract = 53/100   # Democrats voters in the second state 53%
n_state2 = 100                # No. of samples from second state=100
```

In [13]:
```
# Calculate probability that the survey will show a greater percentage of
# Republican voters in the second state than in the first state

# Standard deviation
σ=  math.sqrt(((p_state1_republican*(1- p_state1_republican))/n_state1) + \
              ((p_state2_republican*(1- p_state2_republican))/n_state2))

print('Standard deviation:\t',  round(σ,5))

# Mean Difference
mean_difference = p_state2_republican  - p_state1_republican
print('Mean Difference:\t',  round(mean_difference,5))

# Z Score
# Z = (mean difference/Std Deviation)

Z = mean_difference/σ
print('Z Score:\t\t',  round(Z,5))

print('\nProbability of having greater Republican voters in the second state:\n\t\t p = stats.norm.cdf(Z)')
p = stat.norm.cdf(Z) # cdf function takes Z- score , returns standard normal probality
print('\t i.e.\t p =',round(p,4))

print('\ni.e. The probability that the survey will show a greater percentage of Republican voters \n' \
      '\t in the second state than in the first state is', round(p,4))
```

```
Standard deviation:      0.07062
Mean Difference:         -0.05
Z Score:                 -0.70803

Probability of having greater Republican voters in the second state:
                 p = stats.norm.cdf(Z)
         i.e.    p = 0.2395

i.e. The probability that the survey will show a greater percentage of Republican voters
         in the second state than in the first state is 0.2395
```

Problem Statement 3:

You take the SAT and score 1100. The mean score for the SAT is 1026 and the standard deviation is 209. How well did you score on the test compared to the average test taker?

In [15]:
```
# Given
X = 1100     # My SAT Score. i.e. Sample value of SAT score
σ = 209      # Standard deviation of SAT score
μ = 1026     # Mean SAT score
N = 1        # No of Samples - only my score considered as sample
```

In [17]:
```
print('\nZ Score Using Formula: (X - μ) / σ/math.sqrt(N)')
Z = (X - μ) / (σ/math.sqrt(N))
print('\t Z-Score value is :', Z)

print('\nProbability of having my score more than averge:\n\t\t p = stats.norm.cdf(Z)')
p = stat.norm.cdf(Z) # cdf function takes Z- score , returns standard normal probality
print('\t i.e.\t p =',round(p,4))

print('\ni.e. Probability of having my score more than averge: ', round(p*100,2),'%')
```

```
Z Score Using Formula: (X - μ) / σ/math.sqrt(N)
         Z-Score value is : 0.35406698564593303

Probability of having my score more than averge:
                  p = stats.norm.cdf(Z)
         i.e.     p = 0.6384

i.e. Probability of having my score more than averge:  63.84 %
```