Problem 1 :

In this problem will test the engagement rate hypothesizes provided by the teacher for both knowledgeable and non-knowledgable students. We will use hypothesis testing in order to challenge the given null hypothesizes of the teacher.

In [2]:
```
import numpy as np
import matplotlib.pyplot
import scipy.stats as stats
from scipy.stats import ttest_1samp
from scipy.stats import norm
```

Q1:

```
Null hypothesis: The mean engagement of knowledgeable students who is = 0.75

Alternative hypothesis H(1): The mean engagement of knowledgeable    students is != 0.75

Therefore, we will be using hypothesis testing for a sample of size  larger than 30.
```

In [3]:
```
know_file = open('engagement_1.txt')
data1 = know_file.readlines()
data1 = [float(x) for x in data1]

nonknow_file = open('engagement_0.txt')
data0 = nonknow_file.readlines()
data0 = [float(x) for x in data0]
```

For data1 (knowledgable student engagment data):

Q2:

In [4]:
```
null_hyp1 = 0.75 #Null Hypothesis

n = len(data1) # data size
print("Knowledgeable engaged student data size: {}\n".format(n))


mean1 = np.mean(data1) # data mean
print("Knowledgeable engaged student sample Mean: {}\n".format(mean1))

std1 = np.std(data1, ddof=1) # sample std
print("Knowledgeable engaged student data Stander Deviation: {}\n".format(std1))


SE1 = std1/(n ** (0.5))
print("Knowledgeable engaged student data Stander Error: {}\n".format(SE1))


z1 = (mean1 - null_hyp1)/(std1/(n ** (0.5))) #stander-score
print("Knowledgeable engaged student data Stander Score: {}\n".format(z1))

p1 = 2 * stats.norm.cdf(-abs(z1))
# or (same result)
p1 = ttest_1samp(data1, null_hyp1)[1]
print("Knowledgeable engaged student p-value: {}\n".format(p1))
```

```
Knowledgeable engaged student data size: 937

Knowledgeable engaged student sample Mean: 0.7430304110448239

Knowledgeable engaged student data Stander Deviation: 0.12712605795317614

Knowledgeable engaged student data Stander Error: 0.004153027288269652

Knowledgeable engaged student data Stander Score: -1.6781948375012814

Knowledgeable engaged student p-value: 0.09364288349711912
```

In [6]:
```
if p1 < 0.1:    # alpha value is 0.01
    print("Therefore, we reject the null hypothesis for 90% confidence interval")
    print("and p-value is not significant\n")

else:
    print("Therefore, we accept the null hypothesis for 90% confidence interval")
    print("and p-value is significant\n")

if p1 < 0.05:    # alpha value is 0.05
    print("Also, we reject the null hypothesis for 95% confidence interval")
    print("and p-value is not significant\n")
else:
    print("Also, we accept the null hypothesis for 95% confidence interval")
    print("and p-value is significant\n")

if p1 < 0.01:    # alpha value is 0.01
    print("Therefore, we reject the null hypothesis for 99% confidence interval")
    print("and p-value is not significant\n")
else:
    print("Also, we accept the null hypothesis for 99% confidence interval")
    print("and p-value is significant\n")
```

```
Therefore, we reject the null hypothesis for 90% confidence interval
and p-value is not significant

Also, we accept the null hypothesis for 95% confidence interval
and p-value is significant

Also, we accept the null hypothesis for 99% confidence interval
and p-value is significant
```

Q3:

```
For 95% confidance intervale The largest stander error, and the      corresponding mimimum sampl
e size, are as follows:
```

In [35]:
```
alpha1 = 0.05 # significant level of 0.05
stdr_alpha1 = norm.ppf(alpha1) # stander score at the given level
SE_alpha1 = (mean1 - null_hyp1) / stdr_alpha1
print("At significant at a level of 0.05 \nThe largest standart error is:",SE_alpha1)
size_alpha1 = (std1 / SE_alpha1) ** 2
print("The minimum sample size is",int(size_alpha1))
```

```
At significant at a level of 0.05
The largest standart error is: 0.004237209220916103
The minimum sample size is 900
```

Q4:

```
This question requires us to compare two datasets. We will work on   comparing the statical resu
lts of the null hypothesis H(0) which     should be the same. On the other hand, the alternative
 hypothesis    H(1) will have different statical results.

Therefore, in this part we will use hypothesis testing since the     sample size is larger than
 30, as follows:
```

In [37]:
```
null_hyp0 = 0.75 #Null Hypothesis

n0 = len(data0) # data size
print("Non Knowledgeable engaged student data size: {}\n".format(n0))


mean0 = np.mean(data0) # data mean
print("Non Knowledgeable engaged student sample Mean: {}\n".format(mean0))

std0 = np.std(data0, ddof=1) # sample std
print("Non Knowledgeable engaged student data Stander Deviation: {}\n".format(std0))


SE0 = std0/(n0 ** (0.5))
print("Non Knowledgeable engaged student data Stander Error: {}\n".format(SE0))


z0 = (mean0 - null_hyp0)/(std0/(n0 ** (0.5))) #stander-score
print("Non Knowledgeable engaged student data Stander Score: {}\n".format(z0))

p0 = 2 * stats.norm.cdf(-abs(z0))
# or (same result)
p0 = ttest_1samp(data0, null_hyp0)[1]
print("Non Knowledgeable engaged student p-value: {}\n".format(p0))
```

```
Non Knowledgeable engaged student data size: 1977

Non Knowledgeable engaged student sample Mean: 0.6399545077035914

Non Knowledgeable engaged student data Stander Deviation: 0.2541527210309046

Non Knowledgeable engaged student data Stander Error: 0.005715989588773277

Non Knowledgeable engaged student data Stander Score: -19.252220562568542

Non Knowledgeable engaged student p-value: 7.73562183240184e-76
```

In [38]:
```
if p0 < 0.1:    # alpha value is 0.01
    print("Therefore, we reject the null hypothesis for 90% confidence interval")
    print("and p-value is not significant\n")

else:
    print("Therefore, we accept the null hypothesis for 90% confidence interval")
    print("and p-value is significant\n")

if p0 < 0.05:    # alpha value is 0.05
    print("Also, we reject the null hypothesis for 95% confidence interval")
    print("and p-value is not significant\n")
else:
    print("Also, we accept the null hypothesis for 95% confidence interval")
    print("and p-value is significant\n")

if p0 < 0.01:    # alpha value is 0.01
    print("Therefore, we reject the null hypothesis for 99% confidence interval")
    print("and p-value is not significant\n")
else:
    print("Also, we accept the null hypothesis for 99% confidence interval")
    print("and p-value is significant\n")
```

```
Therefore, we reject the null hypothesis for 90% confidence interval
and p-value is not significant

Also, we reject the null hypothesis for 95% confidence interval
and p-value is not significant

Therefore, we reject the null hypothesis for 99% confidence interval
and p-value is not significant
```

As the result, showed for each statical result the results are different. Then, the Alternative Hypothesis H(1) takes place.

Q5:

```
Now we will combine the data and work to find the needed statistical results through hypothesis
 testing, as follows:
```

In [40]:
```
null_hyp0 = 0.75 #Null Hypothesis

data = data1 + data0
n2 = len(data) # data size
print("Non Knowledgeable engaged student data size: {}\n".format(n2))


mean2 = np.mean(data) # data mean
print("Non Knowledgeable engaged student sample Mean: {}\n".format(mean2))

std2 = np.std(data, ddof=1) # sample std
print("Non Knowledgeable engaged student data Stander Deviation: {}\n".format(std2))


SE2 = std2/(n2 ** (0.5))
print("Non Knowledgeable engaged student data Stander Error: {}\n".format(SE2))


z2 = (mean2 - null_hyp0)/(std2/(n2 ** (0.5))) #stander-score
print("Non Knowledgeable engaged student data Stander Score: {}\n".format(z2))

p2 = 2 * stats.norm.cdf(-abs(z2))
# or (same result)
p2 = ttest_1samp(data, null_hyp0)[1]
print("Non Knowledgeable engaged student p-value: {}\n".format(p2))
```

```
Non Knowledgeable engaged student data size: 2914

Non Knowledgeable engaged student sample Mean: 0.673098681152711

Non Knowledgeable engaged student data Stander Deviation: 0.22655636936811666

Non Knowledgeable engaged student data Stander Error: 0.004196927844172565

Non Knowledgeable engaged student data Stander Score: -18.323240642334728

Non Knowledgeable engaged student p-value: 4.572608981610918e-71
```

Thus, the p-value is significant for all levels: 0.1, 0.05, 0.01

In [43]:
```
know_file.close()
nonknow_file.close()
```