Task1: Oscar student problem ¶

Hypothesis:

Null(Ho)= The class average score be 71 and above

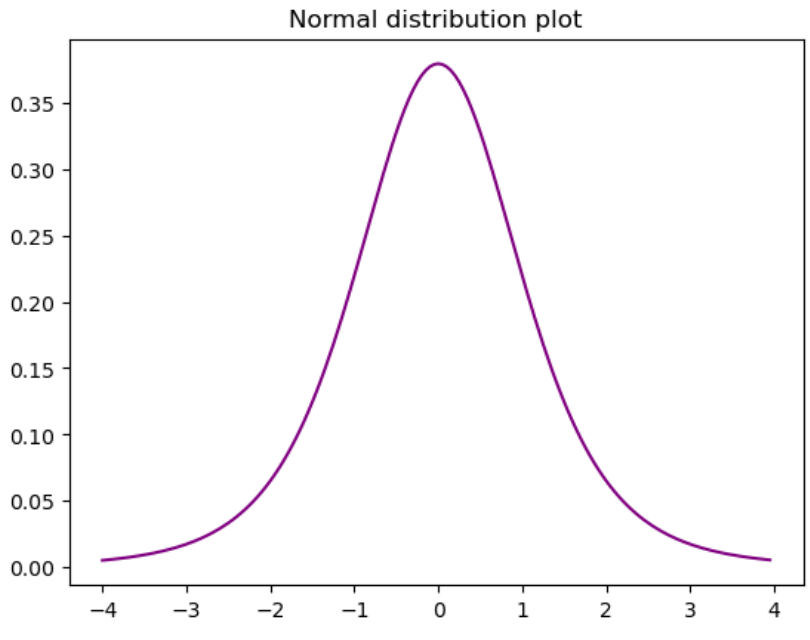
Alternate(H1)= The class average might be lower

We use T statistics since number of samples less than 30 and also we don't know population standard deviation

In [1]:
```
from scipy.stats import t,f,f_oneway
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
scores=[61,93,75,67,84,95]
mean_s=np.mean(scores)
alpha=0.1
dof=6-1
s_samp=np.std(scores)

t_crit=t.ppf(1-alpha,dof)

t_stats=(mean_s-71)/(s_samp/(dof)**0.5)

p_val=t.sf(t_stats,dof)

print("The test critical value is = ",t_crit)
print("The test statistics for given sample = ",t_stats)
print("The p value for this sample = ",p_val)


nr=np.arange(-4,4,.05)
plt.plot(nr,t.pdf(nr,5),color="purple")
plt.title("Normal distribution plot")
plt.show()
```

```
The test critical value is =  1.4758840487820273
The test statistics for given sample =  1.4430501569317138
The p value for this sample =  0.10429845495147679
```



In [3]:
```
if t_stats > t_crit and p_val<alpha :
    print("Reject Null Hypothesis")
else :
    print("Accept Null Hypothesis")
```

```
Accept Null Hypothesis
```

Thus from our findings and hypothesis test we can see that we are getting acceptance of null hypothesis so there is quite a chance that oscar will get an mark of 71 with 90% confidence level and Ho is True

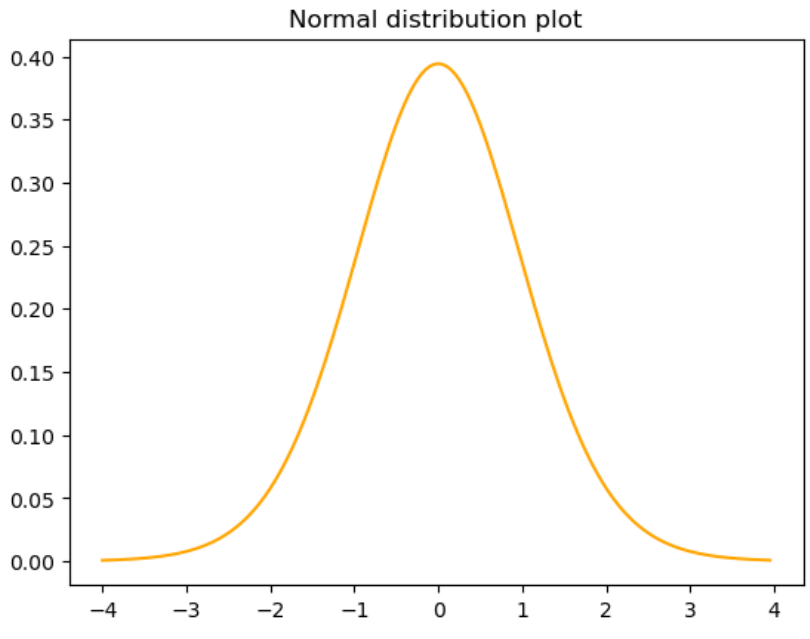
Task2: Oscar wishes to identify whether two batches can get same average of marks

Ho(Null)= The null hypothesis is considered as both means of classes are equal

H1(Alternate)= The alternate hypothesis is considered as either its higher or lower and cannot be same

In [4]:
```
b1=[9,9,9,36,45,48,51,57,69,72,96]
b2=[60,39,39,60,79,96,69,60,75,45,90]

x1=np.mean(b1)
x2=np.mean(b2)

n1=11
n2=11

v1=(np.std(b1))**2
v2=(np.std(b2))**2

alpha=0.1
dof=22-2

t_crit=t.ppf(1-alpha/2,dof)
print("The test critical value is = ",t_crit)

numerator = abs(x1-x2)
denominator = (v1*(n1-1) + v2*(n2-1))**0.5*(1/n1 +1/n2)**0.5

t_stats=numerator/denominator
print("The test statistics values is = ",t_stats)

p_val= t.sf(abs(t_stats), dof)*2
print("The p value for this samples ", p_val)

nr=np.arange(-4,4,.05)
plt.plot(nr,t.pdf(nr,21),color="orange")
plt.title("Normal distribution plot")
plt.show()
```

```
The test critical value is =  1.7247182429207857
The test statistics values is =  0.4346146276529854
The p value for this samples  0.6684953273749834
```



In [5]:
```
if (t_stats > t_crit) and (p_val<alpha) :
    print("Reject Null Hypothesis")
else :
    print("Accept Null Hypothesis")
print()

lhs=abs(x1 - x2) - 1.645*((v1/n1 + v2/n2))
rhs=(x1 - x2) + 1.645*((v1/n1 + v2/n2))
print("The difference between two means  = ",rhs+lhs)
```

```
Accept Null Hypothesis

The difference between two means  =  0.0
```

Decision made:

Since we are accepting Null hypothesis in this case we can conclude that the means for both the batches are same and do not differ

Task3: Oscar wishes to identify the same as task 2 for task 3 with 4 batches

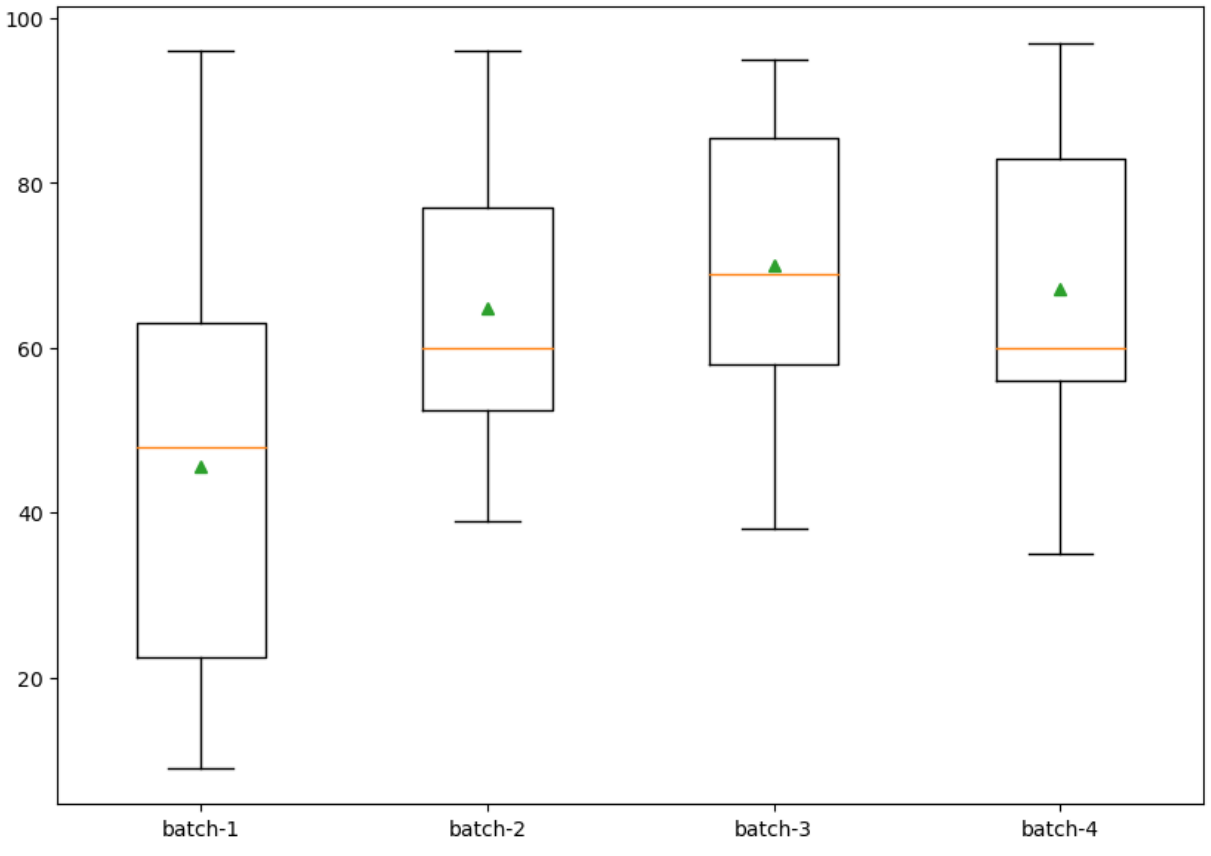
Hypothesis :

Null(Ho): The means of all the samples remain the same

H1(Alternate): The alternate hypothesis is to claim the means are not equal to each other

In [6]:
```
b3=[38,61,62,89,86,50,85,55,80,69,95]
b4=[45,55,57,88,78,60,35,97,58,88,78]
dfc = 4-1
dfe = 44-4

f_critical =f.ppf(1-0.01,dfc,dfe) # 99% confidence level
print('F-critical for the 4 samples =', f_critical)

f_statistics, pvalue= f_oneway(b1,b2,b3,b4)
print('F-statistics for the 4 samples =', f_statistics)
print("p value for 4 samples = ",pvalue)

fig = plt.figure(figsize =(10,7))
data=[b1,b2,b3,b4]
plt.boxplot(data, showmeans= True)
plt.xticks([1,2,3,4], ['batch-1','batch-2','batch-3','batch-4'])
plt.show()


```

```
F-critical for the 4 samples = 4.312569212492142
F-statistics for the 4 samples = 2.8403108468764224
p value for 4 samples =  0.04991227386422173
```



In [7]:
```
if f_statistics > f_critical and pvalue< 0.01 :
    print("Reject Null Hypothesis")
else :
    print("Accept Null Hypothesis")
```

```
Accept Null Hypothesis
```

Decision :

From above hypothesis null we can see that its acceptable so we can state that the averages of all the batches are same