In [171]:
```
#Exam1
#Ho Wing Wong
```

In [172]:
```
import numpy as np
import scipy as sci
import matplotlib.pyplot as plt
from scipy import stats
```

In [173]:
```
#1

n1 = 1379
n2 = 1342
P1 = 0.62
P2 = 0.50
Pc = (n1*P1+n2*P2)/(n1+n2)
z = (P1-P2)/((Pc*(1-Pc)/n1)+(Pc*(1-Pc)/n2))**(0.5)
P_value = stats.norm.sf(z)
print('''
pi1 = parents who regrad science and Math are crucial
pi2 = students of grade 6-12 thought understanding science and strong ma
th  are crucial
H0: pi1 - pi2 = 0
Ha: pi1 - pi2 !=0
''')
print('Pc = ', Pc)
print('z = ', z)
print('P value = ', P_value)
print('Because P value < 0.05, we reject the null hypothesis')
```

```
pi1 = parents who regrad science and Math are crucial
pi2 = students of grade 6-12 thought understanding science and strong m
ath  are crucial
H0: pi1 - pi2 = 0
Ha: pi1 - pi2 !=0

Pc =  0.5608158765159867
z =  6.305824800428203
P value =  1.43331580363e-10
Because P value < 0.05, we reject the null hypothesis
```

In [174]:
```
#2
n1 = 20
n2 = 25
mean1 = 87
mean2 = 73
std1 = 14.4
std2 = 17.5
std12 = ((std1**2)/n1+(std2**2)/n2)**(0.5)
t = (mean2-mean1)/std12
v1 = (std1**2)/n1
v2 = (std2**2)/n2
ddof = (v1+v2)**2/((v1**2)/(n1-1)+(v2**2)/(n2-1))
p = 1 - stats.t.sf(t, ddof)
tcri = stats.t.isf(0.01, ddof)
a = (mean2-mean1 )- (tcri*((std1**2)/n1+(std2**2)/n2)**(1/2))
b = (mean2-mean1 )+ (tcri*((std1**2)/n1+(std2**2)/n2)**(1/2))
print('(a,)')
print('ddof = ', ddof)
print('t(P(0.98), df = 43) = ', tcri)
print('98% Interval = (', a, ', ', b, ')')
print('(b,)')
print('''H0: mean2-mean1 = 0
Ha: mean2- mean1 < 0''')
print('t = ', t)
print('ddof = ', ddof)
print('P-value = ', p)
print('Because P < 0.01, we reject H0')
```

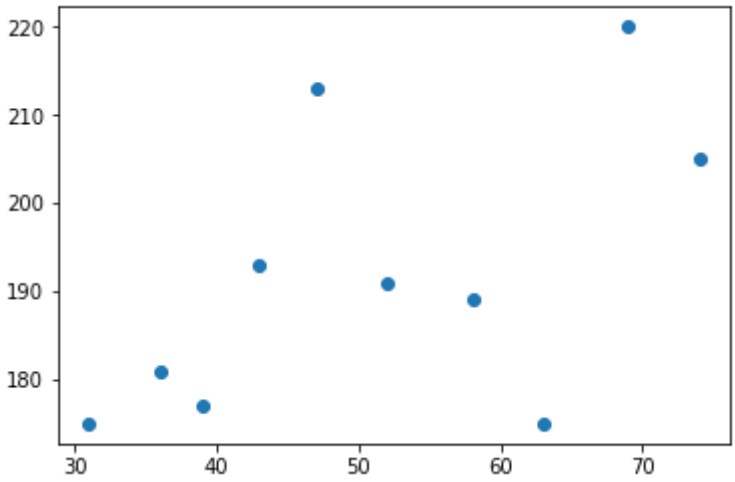
```
(a,)
ddof =  42.952379483515244
t(P(0.98), df = 43) =  2.41635342845
98% Interval = ( -25.4917867057 ,  -2.50821329425 )
(b,)
H0: mean2-mean1 = 0
Ha: mean2- mean1 < 0
t =  -2.9437500768557205
ddof =  42.952379483515244
P-value =  0.00260680402772
Because P < 0.01, we reject H0
```

In [175]:
```
#3
x = np.array([31, 39, 36, 43, 74, 52, 58, 63, 69, 47])
y = np.array([175, 177, 181, 193, 205, 191, 189, 175, 220, 213])
n = 10
Ex = sum(x)
Ey = sum(y)
Exx = sum(x**2)
Eyy = sum(y**2)
Exy = sum(x*y)
meanx = sci.mean(x)
meany = sci.mean(y)
Sxy = Exy - (Ex*Ey)/n
Sxx = Exx - (Ex**2)/n
a = (Exy - (Ex*Ey)/n)/(Exx - (Ex**2)/n)
b = meany-(Exy - (Ex*Ey)/n)/(Exx - (Ex**2)/n)*meanx
SSR = Eyy - b*Ey-a*Exy
```

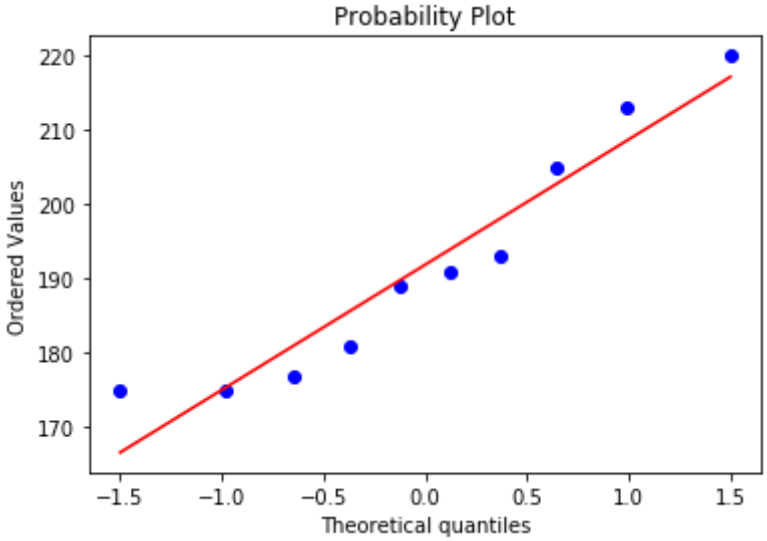
```
Se = (SSR/8)**(0.5)
Sb = Se/Sxx**(0.5)
Sab = Se*((1/10)+(((55-meanx)**2)/Sxx))**(0.5)

plt.scatter(x, y)
print('(a,)')
plt.show()
print('(b,)')
r = sci.corrcoef(age,Cho)[0,1]
print('Pearson correlation coefficient = ', r)
print('(c,)')
print(a, 'x + ',b )
print('(d,)')
print('r = 0.5569 indicates the linear relationship between Age and Chol
esteral is moderately strong and positive')
print('(e,)')
print('a(55) + b = ', a*55+b )
print('(f,)')
print('t critical value = ', stats.t.isf(0.05, 8))
print('Sa+b(55) = ', Sab)
print('90% confident Inteval = ', a*55+b-stats.t.isf(0.05, 8)*Sab, '
',a*55+b+stats.t.isf(0.05, 8)*Sab)
print('(g,)')
print('95% Confidence interval when n = 8 is ', stats.t.isf(0.025,6))
print('a +- (t-critical)Sb = 0.617324330027+-2.446*0.325479 = (',a -stat
s.t.isf(0.025,6)*Sb, ', ',a +stats.t.isf(0.025,6)*Sb, ')')
print('(h,)')
print('''H0: B = 0
Ha: B! = 0
B = 0.05''')
print('t = ', a/Sb)
print('P(t = 2.322, n = 8) = ', stats.t.sf(-t, 8))
print('Because P < B, we reject null hypothesis')
print('(i,)')
print('''r^2 is =  0.310, fitting the simple linear regression model to t
he data would result
in only about 30% of ovserved variation in y''')
print('(j,)')
stats.probplot(y, plot = plt)
plt.show()
print('(k,)')
print('''the normal probability plot is approximately linear. The plots
 would not
cause us to question the assumption of normality.''')
```

```
(a,)
```




```
(b,)
Pearson correlation coefficient =  0.556943735439
(c,)
0.617324330027 x +  160.292994303
(d,)
r = 0.5569 indicates the linear relationship between Age and Cholestera
l is moderately strong and positive
(e,)
a(55) + b =  194.245832454
(f,)
t critical value =  1.85954803752
Sa+b(55) =  4.64877000908
90% confident Inteval =  185.601221307   202.890443601
(g,)
95% Confidence interval when n = 8 is  2.44691184879
a +- (t-critical)Sb = 0.617324330027+-2.446*0.325479 = ( -0.17909409881
7 ,  1.41374275887 )
(h,)
H0: B = 0
Ha: B! = 0
B = 0.05
t =  1.89666406876
P(t = 2.322, n = 8) =  0.00930131468459
Because P < B, we reject null hypothesis
(i,)
r^2 is =  0.310, fitting the simple linear regression model to the data
would result
in only about 30% of ovserved variation in y
(j,)
```




```
(k,)
the normal probability plot is approximately linear. The plots would no
t
cause us to question the assumption of normality.
```

In [ ]: