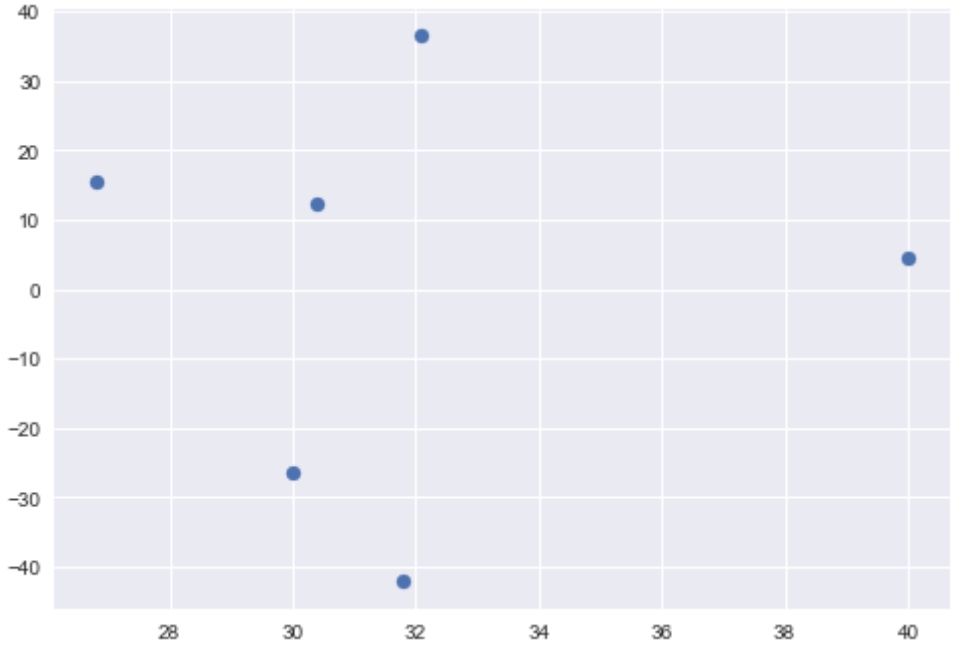
In [ ]:
```
#Assignment 2- Ho Wing Wong
```

In [71]:
```
from scipy import stats
import matplotlib.pyplot as plt
import numpy as np
import scipy as sci
import seaborn as sns
```

In [72]:
```
#5.38
Pollution = np.array([30, 31.8, 32.1, 26.8, 30.4, 40])
COMC = np.array([915, 891, 968, 972, 952, 899])
SixResiduals = COMC - (Pollution*(-4.691) + 1082.2)
print('a, Six Residuals = ', SixResiduals)
stats.linregress(Pollution, COMC)
print('b, CorrelationCofficient = ', sci.corrcoef(Pollution, COMC)[0,1])
print('   r = -5.8 indicates the linear relationship between Pollution a
nd COMC is moderately strong and negative')
plt.scatter(Pollution, SixResiduals)
plt.show()
print('c, All are positive except the first two data')
```

```
a, Six Residuals =  [-26.47   -42.0262  36.3811  15.5188  12.4064   4.4
4  ]
b, CorrelationCofficient =  -0.581289611565
   r = -5.8 indicates the linear relationship between Pollution and COM
C is moderately strong and negative
```




```
c, All are positive except the first two data
```

In [73]:
```
#5.46

print('a, The Predicted maximum salmonid size = ', 89.1 + 0.729*375)
print('   Residual = ', 165 - (89.1 + 0.729*375))
print('b, r-square = 96.3%, 96% of variation in y can be explained')
print('c, the value of se = ', sci.sqrt(((1-(0.96*2520))/9)))
```

```
a, The Predicted maximum salmonid size =  362.475
   Residual =  -197.47500000000002
b, r-square = 96.3%, 96% of variation in y can be explained
c, the value of se =  16.3917323334j
```

In [74]:
```
#13.2a
print('a, A drop of 10 in. = ', -0.12+0.095*10)
print('   A drop of 15 in. = ', -0.12+0.095*15)
```

```
a, A drop of 10 in. =  0.83
   A drop of 15 in. =  1.3050000000000002
```

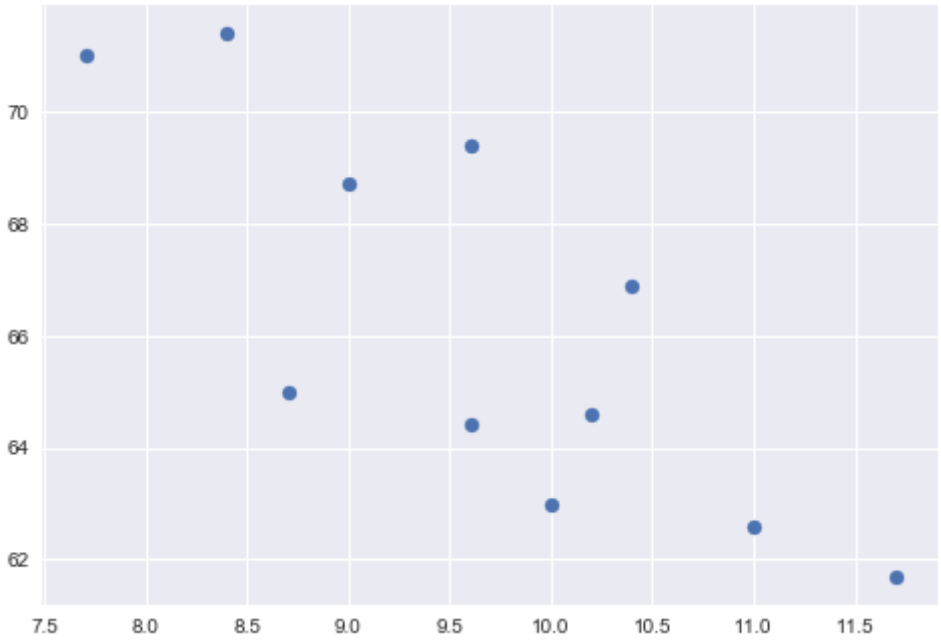
In [75]:
```
#13.4
print('a, Increase in 1 unit of housepower, the fuel efficiency decrease
s 0.15 times')
print('b, 1. When horsepower = 100, the predicted value of fuel efficien
cy is 29')
print('   2. Then Horsepower = 100, the estimated mean value of fuel eff
iciency is 29')
print('c, The fuel efficiency where horsepower (x = 300) = ', 44-0.150*3
00)
print('   Higher hoursepower means using more fuel, therefore fuel effic
iency is lower')
print('d, r-square shows the variation in fuel efficiency in y can be ex
plained')
```

```
a, Increase in 1 unit of housepower, the fuel efficiency decreases 0.15
times
b, 1. When horsepower = 100, the predicted value of fuel efficiency is
 29
   2. Then Horsepower = 100, the estimated mean value of fuel efficienc
y is 29
c, The fuel efficiency where horsepower (x = 300) =  -1.0
   Higher hoursepower means using more fuel, therefore fuel efficiency
 is lower
d, r-square shows the variation in fuel efficiency in y can be explaine
d
```

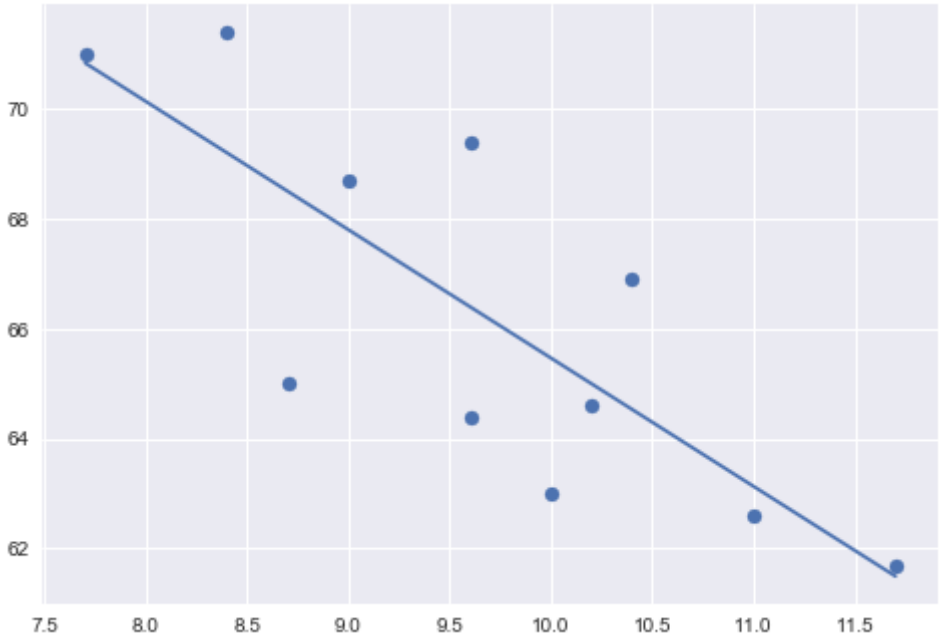
In [76]:
```
#13.6
print('a, y = a + Bx is tht true population regression line')
print('   yhat = a + bx is the estimated regression line')
print('b, B is the slope of population regression line, and b is the poi
nt estimate of B')
```

```
a, y = a + Bx is tht true population regression line
   yhat = a + bx is the estimated regression line
b, B is the slope of population regression line, and b is the point est
imate of B
```

In [77]:
```
#13.8
x = np.array([7.7, 8.4, 8.7, 9.0, 9.6, 9.6, 10.0, 10.2, 10.4, 11.0,
11.7])
y = np.array([71.0, 71.4, 65, 68.7, 64.4, 69.4, 63, 64.6, 66.9, 62.6, 6
1.7])
plt.figure(1)
plt.scatter(x, y)
plt.show()
print('a, Yes, the scatter plot shows it is appropriate')
m, b, r, p, stderr = stats.linregress(x,y)
print('b, Slope, Intercept, r-value, p, stderr = ', m, b, r, p, stderr)
print('   yhat = 88.79-2.33x')
plt.figure(2)
plt.scatter(x, y)
plt.plot(x, x*m+b)
plt.show()
print('c, -2.33 is the estimate of the average change in ski time with 1
 min increase in treadmill time')
print('d, f(x=10) = ', 88.79-2.33*10)
print('e, No, because 15 is out of the x range, as 7.7<x<11.7')
print('f, r-square = ', r**2)
print('g, se = SSR/(n-2) = 43.097/9 = ', 2.188)
```




```
a, Yes, the scatter plot shows it is appropriate
b, Slope, Intercept, r-value, p, stderr =  -2.33351021491 88.7956487132
-0.796192141109 0.00336604784251 0.591095188614
   yhat = 88.79-2.33x
```




```
c, -2.33 is the estimate of the average change in ski time with 1 min i
ncrease in treadmill time
d, f(x=10) =  65.49000000000001
e, No, because 15 is out of the x range, as 7.7<x<11.7
f, r-square =  0.633921925564
g, se = SSR/(n-2) = 43.097/9 =  2.188
```

In [78]:
```
#13.24
x = np.array([2024, 5038, 905, 3572, 1157, 327, 378, 191])
y = np.array([1.9, 3.96, 2.44, 0.88, 0.37, -0.9, 0.49, 1.01])
m, b, r, p, stderr = stats.linregress(x,y)
print('a, B = true average change in growth rate for 8 industries with 1
 % increase')
print('   H0: B = 0')
print('   Ha: B!= 0')
print('   Significance level = 0.05')
print('   Slope, Intercept, r-value, p, stderr = ', m, b, r, p, stderr)
print('   p-value = 0.0624, which greater than 0.05 significance value,
 we fail to reject null hypothesis')
print('b, 90% confidence interval when n = 6 is 1.94')
print('   b +- (t-critical)Sb = 0.000575+-1.94*0.0002518 = (0.0000865,
 0.001063)' )
```

```
a, B = true average change in growth rate for 8 industries with 1 % inc
rease
   H0: B = 0
   Ha: B!= 0
   Significance level = 0.05
   Slope, Intercept, r-value, p, stderr =  0.000575064882174 0.29171476
5186 0.68200610403 0.0624358949356 0.000251753858686
   p-value = 0.0624, which greater than 0.05 significance value, we fai
l to reject null hypothesis
b, 90% confidence interval when n = 6 is 1.94
   b +- (t-critical)Sb = 0.000575+-1.94*0.0002518 = (0.0000865, 0.00106
3)
```

In [79]:
```
#13.26
x = [78, 75, 78, 81, 84, 86, 87]
y = [850, 775, 750, 975, 915, 1015, 1030]
m, b, r, p, stderr = stats.linregress(x,y)
print('Slope, Intercept, r-value, p, stderr = ', m, b, r, p, stderr)
print('yhat = -908.7+22.26x')
print('Cranial capacity would not increase by 20, but 22.26')
```

```
Slope, Intercept, r-value, p, stderr =  22.2569444444 -907.743055556 0.
893510189224 0.00670636923285 5.00229017767
yhat = -908.7+22.26x
Cranial capacity would not increase by 20, but 22.26
```

In [ ]:
```

```