In [112]:
```
#Ho Wing Wong
import statsmodels.api as sm
from scipy import stats
import matplotlib.pyplot as plt
import numpy as np
```

In [113]:
```
#1
bc4000 = [131, 138, 125, 129, 131, 135, 132, 134, 138]
bc1850 = [129, 134, 136, 137, 137, 129, 136, 138, 134]
ad150 = [128, 138, 136, 139, 141, 142, 137, 145, 137]
F, P = stats.f_oneway(bc4000, bc1850, ad150)
print('a, ')
print('H0: 4000B.C. mean - 1850 B.C. mean - 150 A.D. mean = 0')
print('Ha: 4000B.C. mean - 1850 B.C. mean - 150 A.D. mean != 0')
print('F = ', F)
print('P = ', P)
print('Because P < .05, we reject null hypothesis. There are changes in
 head shape over time')
print('b, ')
print('I think 4000 B.C. and 150 A.D. are not equal, because thay have t
o longest time range')
print('H0: mean of 4000 B.C. = mean of 150 A.D')
print('H0: mean of 4000 B.C. != mean of 150 A.D')
F, P = stats.f_oneway(bc4000, ad150)
print('F = ', F)
print('P = ', P)
print('Because P < .05, we reject null hypothesis. The means are not equ
al')
```

```
a,
H0: 4000B.C. mean - 1850 B.C. mean - 150 A.D. mean = 0
Ha: 4000B.C. mean - 1850 B.C. mean - 150 A.D. mean != 0
F =  4.16989247312
P =  0.0279054717857
Because P < .05, we reject null hypothesis. There are changes in head s
hape over time
b,
I think 4000 B.C. and 150 A.D. are not equal, because thay have to long
est time range
H0: mean of 4000 B.C. = mean of 150 A.D
H0: mean of 4000 B.C. != mean of 150 A.D
F =  6.87757909216
P =  0.0184754538327
Because P < .05, we reject null hypothesis. The means are not equal
```

In [114]:
```
#2
year = list(range(1,11))
precip = np.array([32.4, 41.8, 36.5, 37.5, 31.6, 40.3, 33.3, 21.6, 24.7,
 39.4])
avetemp = np.array([49.8, 46.9, 48.5, 48.2, 46.5, 49.3, 51.8, 49.3,
47.1, 49.9])
acres = np.array([13850, 13150, 8550, 12900, 13550, 12050, 10150, 10700,
 12250, 12400])
corn = np.array([1731, 1578, 744, 1445, 1707, 1627, 1320, 899, 1446, 156
2])
X = np.array([precip, avetemp, acres]).T
X = sm.add_constant(X)
model = sm.OLS(corn, X).fit()
model.summary()
```

```
/Library/Frameworks/Python.framework/Versions/3.5/lib/python3.5/site-pa
ckages/scipy/stats/stats.py:1327: UserWarning: kurtosistest only valid
 for n>=20 ... continuing anyway, n=10
  "anyway, n=%i" % int(n))
```

Out[114]:

OLS Regression Results

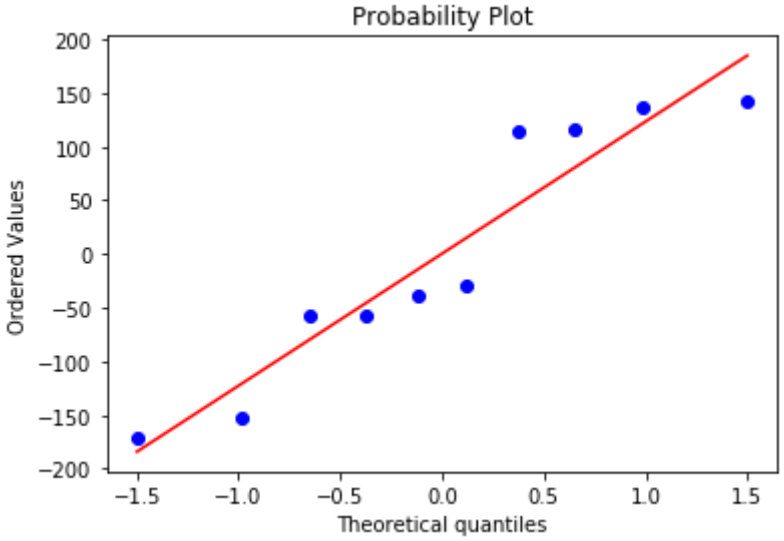
Dep. Variable:	y	**R-squared:**	0.873
Model:	OLS	**Adj. R-squared:**	0.810
Method:	Least Squares	**F-statistic:**	13.81
Date:	Mon, 15 May 2017	**Prob (F-statistic):**	0.00421
Time:	17:47:17	**Log-Likelihood:**	-61.441
No. Observations:	10	**AIC:**	130.9
Df Residuals:	6	**BIC:**	132.1
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
const	-3130.4749	1759.505	-1.779	0.126	-7435.828 1174.878
x1	10.0378	7.412	1.354	0.224	-8.099 28.175
x2	40.0501	32.492	1.233	0.264	-39.456 119.556
x3	0.1877	0.032	5.871	0.001	0.109 0.266

Omnibus:	2.231	**Durbin-Watson:**	2.412
Prob(Omnibus):	0.328	**Jarque-Bera (JB):**	0.860
Skew:	-0.039	**Prob(JB):**	0.651
Kurtosis:	1.566	**Cond. No.**	4.61e+05

In [115]:
```
pvalue = 1-stats.f.cdf(13.81, 3,6)
x1 = 24.7
x2 = 47.1
x3 = 12250
y = -3130.4749 + 10.0378*x1 + 40.0501*x2 + 0.1877*x3
print('a, Corn Production = -3130.4749 + 10.0378* Precipitation + 40.050
1 * Average temp + 0.1877 * Acres harvested')
print('b, ')
print('H0: B1=B2=B3=0')
print('Ha: B1 or B2 or B3 != 0')
print('F = 13.81')
print('P-value = 0.00421')
print('Becuase P = 0.00421 < 0.05, we reject H0')
print('c, y(year9) = ', y)
print('d, 87.3%')
print('e, Yes, the residuals looks like it come from a normal distributi
on')
cornhat = -3130.4749 + 10.0378*precip + 40.0501*avetemp + 0.1877* acres
stats.probplot(corn-cornhat, plot = plt)
plt.show()
print('f, ')
X = np.array([avetemp]).T
X = sm.add_constant(X)
model = sm.OLS(corn, X).fit()
model.summary()
```

```
a, Corn Production = -3130.4749 + 10.0378* Precipitation + 40.0501 * Av
erage temp + 0.1877 * Acres harvested
b,
H0: B1=B2=B3=0
Ha: B1 or B2 or B3 != 0
F = 13.81
P-value = 0.00421
Becuase P = 0.00421 < 0.05, we reject H0
c, y(year9) =  1303.1434700000004
d, 87.3%
e, Yes, the residuals looks like it come from a normal distribution
```




```
f,
```

```
/Library/Frameworks/Python.framework/Versions/3.5/lib/python3.5/site-pa
ckages/scipy/stats/stats.py:1327: UserWarning: kurtosistest only valid
 for n>=20 ... continuing anyway, n=10
  "anyway, n=%i" % int(n))
```

Out[115]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.028
Model:	OLS	**Adj. R-squared:**	-0.094
Method:	Least Squares	**F-statistic:**	0.2283
Date:	Mon, 15 May 2017	**Prob (F-statistic):**	0.646
Time:	17:47:18	**Log-Likelihood:**	-71.637
No. Observations:	10	**AIC:**	147.3
Df Residuals:	8	**BIC:**	147.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
const	3070.3371	3484.895	0.881	0.404	-4965.846 1.11e+04
x1	-34.1563	71.478	-0.478	0.646	-198.986 130.673

Omnibus:	4.220	**Durbin-Watson:**	1.805
Prob(Omnibus):	0.121	**Jarque-Bera (JB):**	2.004
Skew:	-1.096	**Prob(JB):**	0.367
Kurtosis:	2.958	**Cond. No.**	1.54e+03

In [116]:
```
print('H0: There is no linear correaltion')
print('Ha: There is linear correaltion')
print('t = -0.478')
print('P = 0.646')
print('Becuase P > 0.05, we fail to reject H0')
print('g, ')
a = -34.1563+stats.t.isf(0.05, 8)*71.478
b = -34.1563-stats.t.isf(0.05, 8)*71.478
print('90% confident intervial = ', b,' ',a)
print('h, ')
print('H0: B - (-22.1)=0')
print('Ha: B - (-22.1)>0')
print('t = ', (-34.1563+22.1)/71.478)
print('P = ', stats.t.sf((-34.1563+22.1)/71.478, 8))
print('Becuase P > 0.05, we fail to reject H0')
print('i, ')
X = np.array([precip, avetemp]).T
X = sm.add_constant(X)
model = sm.OLS(corn, X).fit()
model.summary()
```

```
H0: There is no linear correaltion
Ha: There is linear correaltion
t = -0.478
P = 0.646
Becuase P > 0.05, we fail to reject H0
g,
90% confident intervial =  -167.073074626   98.7604746261
h,
H0: B - (-22.1)=0
Ha: B - (-22.1)>0
t =  -0.1686714793362993
P =  0.564879190421
Becuase P > 0.05, we fail to reject H0
i,
```

```
/Library/Frameworks/Python.framework/Versions/3.5/lib/python3.5/site-pa
ckages/scipy/stats/stats.py:1327: UserWarning: kurtosistest only valid
 for n>=20 ... continuing anyway, n=10
  "anyway, n=%i" % int(n))
```

Out[116]:

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.147
Model:	OLS	**Adj. R-squared:**	-0.097
Method:	Least Squares	**F-statistic:**	0.6015
Date:	Mon, 15 May 2017	**Prob (F-statistic):**	0.574
Time:	17:47:23	**Log-Likelihood:**	-70.985
No. Observations:	10	**AIC:**	148.0
Df Residuals:	7	**BIC:**	148.9
Df Model:	2		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[95.0% Conf. Int.]**
const	2580.5664	3525.342	0.732	0.488	-5755.543 1.09e+04
x1	17.3507	17.569	0.988	0.356	-24.193 58.894
x2	-36.1795	71.618	-0.505	0.629	-205.530 133.171

Omnibus:	7.334	**Durbin-Watson:**	1.730
Prob(Omnibus):	0.026	**Jarque-Bera (JB):**	2.861
Skew:	-1.207	**Prob(JB):**	0.239
Kurtosis:	4.021	**Cond. No.**	1.90e+03

In [117]:
```
print('H0: B1=B2=0')
print('Ha: B1 or B2!= 0')
print('F = 0.6015')
print('P-value = 0.574')
print('Becuase P = 0.574 > 0.05, we fail to reject H0, the least square
 equation is not useful')
print('j, Using only Acres harvested as the only prediction variable wou
ld not make sense. Temperature and precipitation \
do affect the corn production. Imagine an extreme cases: when precipitat
ion = 0 and average temperature = 0, \
it make sense that cron production is equal to 0. \
precipitation and average temp contributed on the model. Only using Acre
s will make the model less accurate')
```

```
H0: B1=B2=0
Ha: B1 or B2!= 0
F = 0.6015
P-value = 0.574
Becuase P = 0.574 > 0.05, we fail to reject H0, the least square equati
on is not useful
j, Using only Acres harvested as the only prediction variable would not
make sense. Temperature and precipitation do affect the corn productio
n. Imagine an extreme cases: when precipitation = 0 and average tempera
ture = 0, it make sense that cron production is equal to 0. precipitati
on and average temp contributed on the model. Only using Acres will mak
e the model less accurate
```