In [1]: `#Ho Wing Wong`

In [144]:
```
import statsmodels.api as sm
import numpy as np
import matplotlib.pyplot as plt
import scipy as sci
from scipy import stats
import pandas as pd
```

In [145]:
```
#14.4
#a
x1 = 2.5
x2 = 32
x3 = 0
x4 = 1
x5 = 16
x6 = 1
x7 = 3
x8 = 0
x9 = 0

y = 3.6 - 0.01*x1 + 0.01*x2 - 0.07*x3 + 0.12*x4 + 0.02*x5 + -0.4 * x6 -
0.01*x7 - 0.04*x8 - 0.02*x9
print('a, y = ', y)
x4 = 0
y = 3.6 - 0.01*x1 + 0.01*x2 - 0.07*x3 + 0.12*x4 + 0.02*x5 + -0.4 * x6 -
0.01*x7 - 0.04*x8 - 0.02*x9
print('b, y = ', y)
x4 = 1
x3 = 1
y = 3.6 - 0.01*x1 + 0.01*x2 - 0.07*x3 + 0.12*x4 + 0.02*x5 + -0.4 * x6 -
0.01*x7 - 0.04*x8 - 0.02*x9
print('c, y = ', y)
```

```
a, y =  3.9050000000000002
b, y =  3.785
c, y =  3.835000000000001
```

In [146]:
```
#14.6
x1 = 10
x2 = 0.5
x3 = 50
x4 = 100
y = 1.52 + 0.02*x1 - 1.4*x2 + 0.02*x3 - 0.0006*x4
print('a, y = ', y)
x1 = 20
x2 = 0.5
x3 = 10
x4 = 30
y = 1.52 + 0.02*x1 - 1.4*x2 + 0.02*x3 - 0.0006*x4
print('b, y = ', y)
```

```
a, y =  1.96
b, y =  1.402
```

In [147]:
```
#14.18
print('H0: B1 = B2 = B3 = B4 = B5 = B6 = 0')
print('Ha: B1 or B2 or B3 or B4 or B5 or B6 != 0')
R2 = 0.83
k = 6
n = 37
df1 = k
df2 = 30
f = (R2 / df1) /((1-R2)/df2)
pvalue = 1-stats.f.cdf(f, df1, df2)
print('df1 = ', df1)
print('df2 = ', df2)
print('F = ', f)
print('P-value = ', pvalue)
print('Because P < 0.01, we reject H0')
```

```
H0: B1 = B2 = B3 = B4 = B5 = B6 = 0
Ha: B1 or B2 or B3 or B4 or B5 or B6 != 0
df1 =  6
df2 =  30
F =  24.411764705882348
P-value =  2.75130362937e-10
Because P < 0.01, we reject H0
```

```
In [148]: #14.24
          pl =np.array([7, 6, 5, 4, 3, 3, 4, 6, 7, 7, 6, 5, 7, 6, 7, 5, 3, 6, 3])
          ps =np.array([20, 20, 20, 20, 20, 40, 40, 40, 40, 80, 60, 80, 100, 100,
          120, 80, 80, 100, 120])
          ct =[1.1, 1.20, 1.23, 1.4, 1.5, 1.4, 1.36, 1.3, 1.28, 1.4, 1.38, 1.4, 1.
          43, 1.43, 1.7, 1.5, 1.4, 1.5, 1.9]
          X = [pl, ps]
          X = np.array(X).T
          X = sm.add_constant(X)
          y = ct

          ols = sm.OLS(y, X).fit()
          ols.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=19
  "anyway, n=%i" % int(n))

Out[148]:

OLS Regression Results

Dep. Variable:	y	R-squared:	0.750
Model:	OLS	Adj. R-squared:	0.719
Method:	Least Squares	F-statistic:	24.02
Date:	Sat, 13 May 2017	Prob (F-statistic):	1.52e-05
Time:	17:42:57	Log-Likelihood:	19.785
No. Observations:	19	AIC:	-33.57
Df Residuals:	16	BIC:	-30.74
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	1.4396	0.083	17.293	0.000	1.263 1.616
x1	-0.0523	0.015	-3.583	0.002	-0.083 -0.021
x2	0.0040	0.001	6.409	0.000	0.003 0.005

Omnibus:	0.520	Durbin-Watson:	1.836
Prob(Omnibus):	0.771	Jarque-Bera (JB):	0.010
Skew:	-0.038	Prob(JB):	0.995
Kurtosis:	3.082	Cond. No.	281.

```
In [149]: print('a, Catch time = 1.439- 0.0523pl + 0.004 ps')
          x1 = 6
          x2 = 50
          y = 1.439 - 0.0523*x1 + 0.004*x2
          print('b, y = ', y)
          print('c, ')
          print('H0: B1 = B2 = 0')
          print('Ha: B1 or B2 !=0')
          print('F = ', 24.02)
          print('P = 0.000')
          print('Because P < 0.05, we reject H0')
          print('d, ')
          print('x = ', pl/ps)
          print('e, ')
          x = pl/ps
          x = sm.add_constant(x)
          ols = sm.OLS(ct, x).fit()
          ols.summary()
```

```
a, Catch time = 1.439- 0.0523pl + 0.004 ps
b, y =  1.3252
c,
H0: B1 = B2 = 0
Ha: B1 or B2 !=0
F =  24.02
P = 0.000
Because P < 0.05, we reject H0
d,
x = [ 0.35        0.3         0.25        0.2         0.15        0.07
5      0.1
  0.15        0.175       0.0875      0.1         0.0625      0.07
   0.06
  0.05833333  0.0625      0.0375      0.06        0.025      ]
e,

/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=19
  "anyway, n=%i" % int(n))
```

Out[149]:

OLS Regression Results

Dep. Variable:	y	R-squared:	0.543
Model:	OLS	Adj. R-squared:	0.516
Method:	Least Squares	F-statistic:	20.22
Date:	Sat, 13 May 2017	Prob (F-statistic):	0.000318
Time:	17:43:02	Log-Likelihood:	14.053
No. Observations:	19	AIC:	-24.11
Df Residuals:	17	BIC:	-22.22
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	1.5865	0.048	33.031	0.000	1.485 1.688
x1	-1.4044	0.312	-4.496	0.000	-2.063 -0.745

Omnibus:	13.636	Durbin-Watson:	1.289
Prob(Omnibus):	0.001	Jarque-Bera (JB):	11.374
Skew:	1.583	Prob(JB):	0.00339
Kurtosis:	5.084	Cond. No.	11.3

```
In [150]:  print('Catch time = 1.59 - 1.41(length/speed)')
           print('Part d would recommended, because R2 is smaller')
```

```
Catch time = 1.59 - 1.41(length/speed)
Part d would recommended, because R2 is smaller
```

In [151]:
```python
#14.25
volume = [125, 135, 175, 285, 330, 90, 120, 520, 330, 570, 340, 175,
240, 240, 360, 310, 635, 1250, 650, 305, 315, 305, 245, 200, 1205, 2330,
 730]
maxwid = [2.5, 2.9, 2.15, 2.9, 3.2, 2, 1.6, 4.8, 5.9, 5.8, 2.9, 2.45, 2.
6, 2.6, 2.7, 3.1, 5.1, 10.2, 3.5, 2.7, 3, 2.7, 2.5, 2.4,4.4, 7.5, 4.25 ]
minwid = [1.8, 2.7, 2, 2.6, 3.15, 1.8, 1.5, 3.8, 5, 4.75, 2.8, 2.1, 2.2,
 2.6, 2.6, 2.9, 5.1, 10.2, 3.5, 1.2, 1.7, 1.75, 1.7, 1.2, 1.2, 7.5,
4.25]
elan = [1.5, 1.07, 1.98, 1.79, 1.25, 2.17, 3.19, 1.09, 0.29, 0.59, 1.88,
 1.98, 1.94, 2.5, 2.41, 1.77, 0.85, 0.84, 2.36, 3.06, 1.62, 3.3, 2.8, 2.
83, 3.17, 1.3, 1.62]
material = [1, 1, 1, 1, 2, 1, 1, 2, 2, 2, 3, 2, 1, 1, 1, 1, 4, 4, 1, 1,
1, 1, 1, 1, 2, 3, 3]
X = np.array([maxwid, minwid, elan]).T
X = sm.add_constant(X)
y = volume
X = sm.add_constant(X)
ols = sm.OLS(y, X).fit()
ols.summary()
```

Out[151]:

OLS Regression Results

Dep. Variable:	y	R-squared:	0.676
Model:	OLS	Adj. R-squared:	0.634
Method:	Least Squares	F-statistic:	16.03
Date:	Sat, 13 May 2017	Prob (F-statistic):	7.65e-06
Time:	17:43:06	Log-Likelihood:	-188.95
No. Observations:	27	AIC:	385.9
Df Residuals:	23	BIC:	391.1
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	-859.1598	272.938	-3.148	0.005	-1423.774 -294.545
x1	225.8059	85.763	2.633	0.015	48.391 403.221
x2	23.7245	85.657	0.277	0.784	-153.471 200.920
x3	225.2382	90.651	2.485	0.021	37.713 412.764

Omnibus:	24.675	Durbin-Watson:	1.536
Prob(Omnibus):	0.000	Jarque-Bera (JB):	61.885
Skew:	1.619	Prob(JB):	3.65e-14
Kurtosis:	9.672	Cond. No.	29.7

```
In [152]: print('Volume = -859.2 + 225.80max + 23.72min + 225.23elan')
          x1 = 5.1
          x2 = 5.1
          x3 = 0.85
          y = -859.2 + 225.8*x1 + 23.72*x2 + 225.23*x3
          print('d, y = ', y)
          print('e, 95% interval of Max = (48.391, 403.221)')
          print('f, Becasue P = 0.015 < 0.05, we reject B = 0')
          print('g, ', material)
          X = np.array([maxwid, minwid, elan, material]).T
          X = sm.add_constant(X)
          y = volume
          X = sm.add_constant(X)
          ols = sm.OLS(y, X).fit()
          print('h, ')
          ols.summary()
```

```
Volume = -859.2 + 225.80max + 23.72min + 225.23elan
d, y =  604.7974999999999
e, 95% interval of Max = (48.391, 403.221)
f, Becasue P = 0.015 < 0.05, we reject B = 0
g,  [1, 1, 1, 1, 2, 1, 1, 2, 2, 2, 3, 2, 1, 1, 1, 1, 4, 4, 1, 1, 1, 1,
  1, 1, 2, 3, 3]
h,
```

Out[152]:

OLS Regression Results

Dep. Variable:	y	R-squared:	0.684
Model:	OLS	Adj. R-squared:	0.626
Method:	Least Squares	F-statistic:	11.89
Date:	Sat, 13 May 2017	Prob (F-statistic):	2.70e-05
Time:	17:43:06	Log-Likelihood:	-188.65
No. Observations:	27	AIC:	387.3
Df Residuals:	22	BIC:	393.8
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
const	-894.6377	280.483	-3.190	0.004	-1476.325 -312.951
x1	214.2699	88.234	2.428	0.024	31.283 397.257
x2	10.6740	88.551	0.121	0.905	-172.969 194.317
x3	228.8165	91.793	2.493	0.021	38.449 419.184
x4	66.6769	94.332	0.707	0.487	-128.955 262.309

Omnibus:	27.658	Durbin-Watson:	1.615
Prob(Omnibus):	0.000	Jarque-Bera (JB):	77.119
Skew:	1.815	Prob(JB):	1.79e-17
Kurtosis:	10.441	Cond. No.	31.6

In [153]:

```
print('i, R2 increase')
print('j, 90% intevial = (-128.955, 262.309)')
print('k, we cannot reject B(Material) = 0 because P = 0.707 > 0.1')
print('l, No, we are not able to run regression if it is not numeric')
print('m, we would need 4')
```

```
i, R2 increase
j, 90% intevial = (-128.955, 262.309)
k, we cannot reject B(Material) = 0 because P = 0.707 > 0.1
l, No, we are not able to run regression if it is not numeric
m, we would need 4
```