Problem 1

7.3. Refer to Brand preference Problem 6.5.

a. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X1 and with X2, given X1.

b. Test whether X2 can be dropped from the regression model given that X1 is retained. Use the F* test statistic and level of significance .01. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

In [1]:
```
import pandas as pd, numpy as np
```

In [2]:
```
df = pd.read_csv('CH06PR05.txt', sep = '\s+', header =None, names=['Y','X1','X2'])
df.head()
```

Out[2]:

	Y	X1	X2
0	64.0	4.0	2.0
1	73.0	4.0	4.0
2	61.0	4.0	2.0
3	76.0	4.0	4.0
4	72.0	6.0	2.0

In [3]:
```
x1= df['X1']
x2= df['X2']
y= df['Y']
```

a. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X1 and with X2, given X1.

Regression of Y on X1

In [4]:
```
import statsmodels.api as sm
import statsmodels.formula.api as smf
model1 = smf.ols('y ~ x1', data=df)
results1 = model1.fit()
```

In [5]:
```
sse1 = np.sum((results1.fittedvalues - df.Y)**2)
ssr1 = np.sum((results1.fittedvalues - df.Y.mean())**2)
```

Regression of Y on X2

In [6]:
```
import statsmodels.api as sm
```

```
import statsmodels.formula.api as smf
model2 = smf.ols('y ~ x2', data=df)
results2 = model2.fit()
```

In [7]:
```
sse2 = np.sum((results2.fittedvalues - df.Y)**2)
ssr2 = np.sum((results2.fittedvalues - df.Y.mean())**2)
```

Regression of Y on X1 and X2

In [8]:
```
import statsmodels.api as sm
import statsmodels.formula.api as smf
model3 = smf.ols('y ~ x1+x2', data=df)
results3 = model3.fit()
```

In [9]:
```
sse12 = np.sum((results3.fittedvalues - df.Y)**2)
ssr12 = np.sum((results3.fittedvalues - df.Y.mean())**2)
```

In [10]:
```
ssrX1_X2 = sse2 - sse12
print('SSR(X1|X2)=',ssrX1_X2)
```

```
SSR(X1|X2)= 1566.45
```

The extra sum of squares SSR(X1|X2) is 1566.45

b. Test whether X2 can be dropped from the regression model given that X1 is retained. Use the F* test statistic and level of significance .01. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

The alternatives

- H0: β2 = 0
- Ha: β2 # 0

The decision rule

- If F* ≤ F(1−α; 1 , n-p), conclude H0
- If F* > F(1−α; 1 , n-p), conclude Ha

In [11]:
```
ssrX2_X1 = ssr12 - ssr1
print('SSR(X2|X1)=',ssrX2_X1)
```

```
SSR(X2|X1)= 306.2499999999984
```

In [12]:
```
n = len(y)
Fstar = (ssrX2_X1/1 / (sse12/(n-3)))
print('F*=',Fstar)
```

```
F*= 42.218981972428246
```

In [13]:
```
import scipy.stats as stats
f = stats.f.ppf(q=1-0.01,dfn=1,dfd=n-3)
f
```

Out[13]:
```
9.073805728515653
```

For α = 0.01, we have F = 9.07. Since F* = 42.2 > 9.07, we conclude Ha that X2 cannot be dropped from the regression model that already contains X1.

In [14]:
```
p =1- stats.chi2.cdf(x=(ssrX2_X1/2) / ((sse12/n)**2),df=1)
print('p-value=',p)
```

```
p-value= 0.03576626606100286
```

p- value = 0.03

7.12. Refer to Brand preference Problem 6.5.

Calculate R^2 Y1, R^2 Y2, R^2 12, R^2 Y1|2, R^2 Y2|1, and R^2. Explain what each coefficient measures and interpret your results.

In [15]:
```
sstoX1 = ssr1+sse1
R2_1 = ssr1/sstoX1
print('R^2 Y1=',R2_1)
```

```
R^2 Y1= 0.7963650228774783
```

In [16]:
```
sstoX2 = ssr2+sse2
R2_2 = ssr2/sstoX2
print('R^2 Y2=',R2_2)
```

```
R^2 Y2= 0.1556939501779355
```

In [17]:
```
ssto12 = ssr12+sse12
R2_12 = ssr12/ssto12
print('R^2 12=',R2_12)
```

```
R^2 12= 0.9520589730554143
```

In [18]:
```
R2_1_2 = ssrX1_X2 / sse2
print('R^2 Y1|2=',R2_1_2)
```

```
R^2 Y1|2= 0.9432184254102063
```

In [19]:
```
R2_2_1 = ssrX2_X1 / sse1
print('R^2 Y2|1=',R2_2_1)
```

```
R^2 Y2|1= 0.7645737111471687
```

In [21]:
```
R2 = 1 - (sse2/sstoX2)
print('R^2=',R2)
```

```
R^2= 0.15569395017793553
```

- R^2 Y1= 0.796
- R^2 Y2= 0.155
- R^2 12= 0.952
- R^2 Y1|2= 0.943
- R^2 Y2|1= 0.764
- R^2= 0.155

R^2 Y1|2 measures the proportionate reduction in the variation in Y remaining after X2 is included in the model that is gained by also including X1 in the model.

R^2 Y2|1 measures the proportionate reduction in the variation in Y remaining after X1 is included in the model that is gained by also including X2 in the model.