In [53]:
```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings("ignore")
```

In [29]:
```
df= pd.read_excel("edited walmart.xlsx")
```

In [30]:
```
df
```

Out[30]:

	Region	Profit	Product Base Margin	Discount	Sales	Order Quantity
0	East	-111.8000	0.68	0.10	180.36	32
1	East	-342.9100	NaN	0.08	872.48	9
2	West	-193.0800	0.56	0.00	1239.06	4
3	West	247.7900	0.39	0.08	614.80	43
4	West	-1049.8500	0.69	0.07	4083.19	43
...	...	...	...	...	...	...
3846	East	5045.3025	0.40	0.08	12690.33	36
3847	West	-164.9200	0.59	0.07	178.70	45
3848	East	-144.7390	0.36	0.05	181.80	40
3849	West	-166.9600	0.70	0.06	391.12	6
3850	West	-15.0700	0.59	0.10	448.10	35

3851 rows × 6 columns

In [31]:
```
data=df.dropna()
```

In [32]:
```
data
```

Out[32]:

	Region	Profit	Product Base Margin	Discount	Sales	Order Quantity
0	East	-111.8000	0.68	0.10	180.36	32
2	West	-193.0800	0.56	0.00	1239.06	4
3	West	247.7900	0.39	0.08	614.80	43
4	West	-1049.8500	0.69	0.07	4083.19	43
5	West	26.7100	0.40	0.09	137.63	16
...	...	...	...	...	...	...
3846	East	5045.3025	0.40	0.08	12690.33	36
3847	West	-164.9200	0.59	0.07	178.70	45
3848	East	-144.7390	0.36	0.05	181.80	40
3849	West	-166.9600	0.70	0.06	391.12	6
3850	West	-15.0700	0.59	0.10	448.10	35

3822 rows × 6 columns

In [33]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3822 entries, 0 to 3850
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Region               3822 non-null   object 
 1   Profit               3822 non-null   float64
 2   Product Base Margin  3822 non-null   float64
 3   Discount             3822 non-null   float64
 4   Sales                3822 non-null   float64
 5   Order Quantity       3822 non-null   int64  
dtypes: float64(4), int64(1), object(1)
memory usage: 209.0+ KB
```

In [52]:
```
data.describe()
```

Out[52]:

	Profit	Product Base Margin	Discount	Sales	Order Quantity
count	3822.000000	3822.000000	3822.000000	3822.000000	3822.000000
mean	157.947426	0.515212	0.049908	1823.417760	25.434328
std	1235.160181	0.136424	0.031751	3618.210333	14.559096
min	-14140.701600	0.350000	0.000000	3.200000	1.000000
25%	-88.767500	0.380000	0.020000	139.107500	13.000000
50%	-3.780000	0.540000	0.050000	437.129000	25.000000
75%	144.530000	0.590000	0.080000	1764.077500	38.000000
max	12748.860000	0.850000	0.100000	45923.760000	50.000000

In [34]:
```
X =data[data['Region']=='East']
X
```

Out[34]:

	Region	Profit	Product Base Margin	Discount	Sales	Order Quantity
0	East	-111.8000	0.68	0.10	180.3600	32
10	East	-32.4800	0.36	0.05	188.7300	24
11	East	-108.8000	0.62	0.02	2357.4500	23
14	East	-11.5805	0.40	0.05	10.5800	3
15	East	-2314.7400	0.41	0.07	2119.6700	3
...	...	...	...	...	...	...
3836	East	-768.1400	0.58	0.01	4547.9000	45
3842	East	707.1660	0.58	0.10	3883.4715	42
3843	East	6670.4090	0.37	0.10	15337.5800	30
3846	East	5045.3025	0.40	0.08	12690.3300	36
3848	East	-144.7390	0.36	0.05	181.8000	40

1880 rows × 6 columns

In [35]:
```
Y =data[data['Region']=='West']
Y
```

Out[35]:

	Region	Profit	Product Base Margin	Discount	Sales	Order Quantity
2	West	-193.080	0.56	0.00	1239.06	4
3	West	247.790	0.39	0.08	614.80	43
4	West	-1049.850	0.69	0.07	4083.19	43
5	West	26.710	0.40	0.09	137.63	16
6	West	1438.490	0.59	0.05	4902.38	32
...	...	...	...	...	...	...
3844	West	-17.457	0.35	0.01	44.45	2
3845	West	488.818	0.40	0.02	2116.70	30
3847	West	-164.920	0.59	0.07	178.70	45
3849	West	-166.960	0.70	0.06	391.12	6
3850	West	-15.070	0.59	0.10	448.10	35

1942 rows × 6 columns

In [44]:
```
def hotelling_t_2(X,Y,p):
    a=X.shape ;b=Y.shape
    nx = a[0]
    ny = b[0]
    delta = np.matrix(X.mean()- Y.mean())
    Sx = np.matrix(X.cov())
    Sy = np.matrix(Y.cov())
    S_pooled = (((nx-1)*Sx + (ny-1)*Sy)/((nx+ny)-2))
    # fot inverse of pooled
    inv_Spooled = np.linalg.inv(S_pooled)
    t_squared = ((nx*ny)/(nx+ny) *delta)*(inv_Spooled*np.transpose(delta))
    statistics = t_squared * ((nx+ny-p-1)/((nx+ny-2)*p))
    from scipy.stats import f
    F = f(p , nx+ny-p-1)
    pvalue = 1- F.cdf(statistics)
    print("Hotelling_T^2" ,t_squared)
    print("Degree of Freedom" , p , "and", (nx+ny-p-1))
    print('p_value', pvalue)
```

In [45]:
```
hotelling_t_2(x.iloc[:,1:5],y.iloc[:,1:5],5)
```

```
Hotelling_T^2 [[7.04385426]]
Degree of Freedom 5 and 3816
p_value [[0.21821517]]
```

In [ ]:
```

```