Statistics_Anscombe_Quartet_Stock

September 29, 2021

1 Anscombe's Quartet Stock Data

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

import warnings
warnings.filterwarnings("ignore")

# yfinance is used to fetch data
import yfinance as yf
yf.pdr_override()
```

[2]:
```
# input
symbol = 'AMD'
start = '2019-12-01'
end = '2020-01-01'

# Read data
df = yf.download(symbol,start,end)

# View Columns
df.head()
```

```
[**********100%**********]  1 of 1 completed
```

[2]:

Date	Open	High	Low	Close	Adj Close	Volume
2019-12-02	39.320000	39.410000	38.439999	38.730000	38.730000	35710200
2019-12-03	37.340000	38.930000	37.150002	38.900002	38.900002	51556400
2019-12-04	39.380001	39.820000	39.130001	39.689999	39.689999	44299400
2019-12-05	39.950001	40.220001	39.549999	39.619999	39.619999	35574000
2019-12-06	40.099998	40.189999	39.560001	39.630001	39.630001	31101400

[3]:
```
df = df.astype('float64')
df.head()
```

[3]:

Date	Open	High	Low	Close	Adj Close	Volume
2019-12-02	39.320000	39.410000	38.439999	38.730000	38.730000	35710200.0
2019-12-03	37.340000	38.930000	37.150002	38.900002	38.900002	51556400.0
2019-12-04	39.380001	39.820000	39.130001	39.689999	39.689999	44299400.0
2019-12-05	39.950001	40.220001	39.549999	39.619999	39.619999	35574000.0
2019-12-06	40.099998	40.189999	39.560001	39.630001	39.630001	31101400.0

[4]:
```
df = df[['Open', 'High', 'Low', 'Adj Close']]
df.head()
```

[4]:

Date	Open	High	Low	Adj Close
2019-12-02	39.320000	39.410000	38.439999	38.730000
2019-12-03	37.340000	38.930000	37.150002	38.900002
2019-12-04	39.380001	39.820000	39.130001	39.689999
2019-12-05	39.950001	40.220001	39.549999	39.619999
2019-12-06	40.099998	40.189999	39.560001	39.630001

[5]:
```
df.shape
```

[5]: (21, 4)

[6]:
```
for i in df.values:
  print(np.array(i))
```

```
[39.31999969 39.40999985 38.43999863 38.72999954]
[37.34000015 38.93000031 37.15000153 38.90000153]
[39.38000107 39.81999969 39.13000107 39.68999863]
[39.95000076 40.22000122 39.54999924 39.61999893]
[40.09999847 40.18999863 39.56000137 39.63000107]
[39.45999908 39.84000015 38.90999985 38.93000031]
[39.20000076 39.72999954 38.81999969 39.43999863]
[39.43999863 39.61000061 39.04000092 39.47000122]
[39.40000153 42.61999893 39.25       42.59000015]
[42.34999847 42.95000076 41.06000137 41.15000153]
[41.72999954 42.97999954 41.72999954 42.34999847]
[42.50999832 43.11000061 42.15000153 42.77000046]
[42.79000092 43.06000137 42.24000168 42.29999924]
[42.63000107 43.34000015 42.59999847 42.83000183]
[43.43999863 44.25999832 43.25999832 44.15000153]
[44.58000183 45.63000107 44.38999939 45.45999908]
[46.09999847 46.61000061 45.77000046 46.54000092]
[46.99000168 47.31000137 45.65999985 46.63000107]
[46.84999847 46.88000107 45.97999954 46.18000031]
[46.13999939 46.16999817 44.65999985 45.52000046]
[45.06999969 46.09000015 44.91999817 45.86000061]
```

[7]:
```
quartets = np.array([df['Open'], df['High'], df['Low'], df['Adj Close']])
```

[8]:
```
quartets
```

[8]:
```
array([[39.31999969, 37.34000015, 39.38000107, 39.95000076, 40.09999847,
        39.45999908, 39.20000076, 39.43999863, 39.40000153, 42.34999847,
        41.72999954, 42.50999832, 42.79000092, 42.63000107, 43.43999863,
        44.58000183, 46.09999847, 46.99000168, 46.84999847, 46.13999939,
        45.06999969],
       [39.40999985, 38.93000031, 39.81999969, 40.22000122, 40.18999863,
        39.84000015, 39.72999954, 39.61000061, 42.61999893, 42.95000076,
        42.97999954, 43.11000061, 43.06000137, 43.34000015, 44.25999832,
        45.63000107, 46.61000061, 47.31000137, 46.88000107, 46.16999817,
        46.09000015],
       [38.43999863, 37.15000153, 39.13000107, 39.54999924, 39.56000137,
        38.90999985, 38.81999969, 39.04000092, 39.25      , 41.06000137,
        41.72999954, 42.15000153, 42.24000168, 42.59999847, 43.25999832,
        44.38999939, 45.77000046, 45.65999985, 45.97999954, 44.65999985,
        44.91999817],
       [38.72999954, 38.90000153, 39.68999863, 39.61999893, 39.63000107,
        38.93000031, 39.43999863, 39.47000122, 42.59000015, 41.15000153,
        42.34999847, 42.77000046, 42.29999924, 42.83000183, 44.15000153,
        45.45999908, 46.54000092, 46.63000107, 46.18000031, 45.52000046,
        45.86000061]])
```

[9]:
```
quartets[0]
```

[9]:
```
array([39.31999969, 37.34000015, 39.38000107, 39.95000076, 40.09999847,
       39.45999908, 39.20000076, 39.43999863, 39.40000153, 42.34999847,
       41.72999954, 42.50999832, 42.79000092, 42.63000107, 43.43999863,
       44.58000183, 46.09999847, 46.99000168, 46.84999847, 46.13999939,
       45.06999969])
```

[10]:
```
quartets.shape[0]
```

[10]:
```
4
```

[11]:
```
for quartet in range(quartets.shape[0]):
    x = np.array(quartet)
    print(x)
```

```
0
1
2
3
```

[12]:
```
for names in df.columns:
    print(names)
```

```
Open
High
Low
Adj Close
```

[13]:
```
for name in df.columns: print(name)
```

```
Open
High
Low
Adj Close
```

[14]:
```
for name in df.columns:
    print("Next")
    print("Adj Close vs ", name)
```

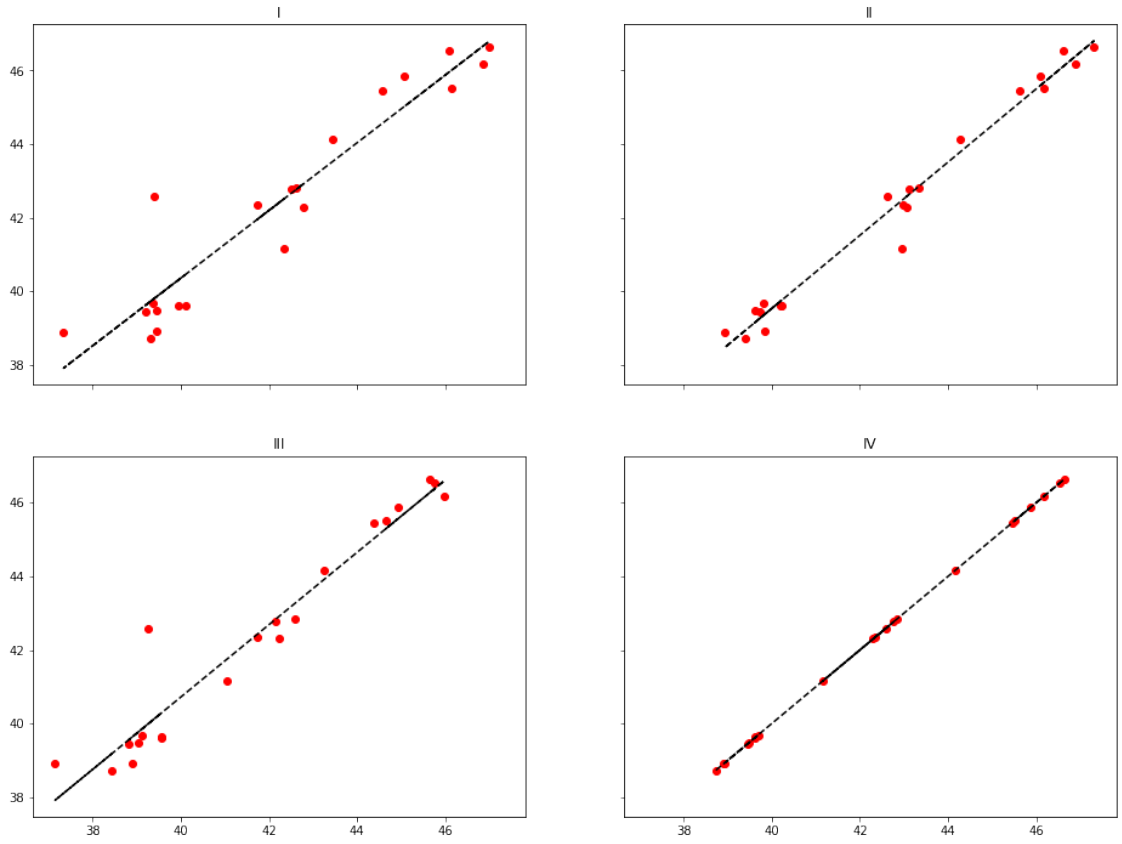
```
Next
Adj Close vs  Open
Next
Adj Close vs  High
Next
Adj Close vs  Low
Next
Adj Close vs  Adj Close
```

[15]:
```
roman = ['I', 'II', 'III', 'IV']
```

[16]:
```
%matplotlib inline
fig = plt.figure(figsize=(16,12))
fig.suptitle("Anscombe's Quartets", fontsize=14)
axes = fig.subplots(2, 2, sharex= True, sharey = True)
n = len(df.index)

for name, quartet in zip(df.columns, range(quartets.shape[0])):
    x = quartets[quartet]
    y = np.array(df['Adj Close'])
    coef = np.polyfit(x, y, 1)
    reg_line = np.poly1d(coef)
    ax = axes[quartet // 2, quartet % 2]
    ax.plot(x, y, 'ro', x, reg_line(x), '--k')
    ax.set_title(roman[quartet])

    print("Quartet:", roman[quartet])
    print("Adj Close vs", name)
```

```
    print("Mean X:", x.mean())
    print("Variance X:", x.var())
    print("Mean Y:", y.mean())
    print("Variance Y:", y.var())
    print("Pearson's correlation coef.:", round(np.corrcoef(x, y)[0][1], 2))
    print('-'*40)

plt.show()
```

```
Quartet: I
Adj Close vs Open
Mean X: 42.13190460205078
Variance X: 8.22944332000805
Mean Y: 42.32095264253162
Variance Y: 7.766971667924468
Pearson's correlation coef.: 0.95
----------------------------------------
Quartet: II
Adj Close vs High
Mean X: 42.7980953398205
Variance X: 7.692720974234099
Mean Y: 42.32095264253162
Variance Y: 7.766971667924468
Pearson's correlation coef.: 0.99
----------------------------------------
Quartet: III
Adj Close vs Low
Mean X: 41.631904783703035
Variance X: 7.489308791873262
Mean Y: 42.32095264253162
Variance Y: 7.766971667924468
Pearson's correlation coef.: 0.97
----------------------------------------
Quartet: IV
Adj Close vs Adj Close
Mean X: 42.32095264253162
Variance X: 7.766971667924468
Mean Y: 42.32095264253162
Variance Y: 7.766971667924468
Pearson's correlation coef.: 1.0
----------------------------------------
```


Anscombe's Quartets
I
II
III
IV
46
44
42
40
38
38
40
42
44
46