In [1]:
```
import pandas as pd
import numpy as np
import scipy.stats as stats
import statsmodels.api as sm
import statsmodels.stats.api as sms
import statsmodels.formula.api as smf
```

Exercise 4.3 (20%)

mean 5%, median 5%, mode 5%, describe 5%

In [2]:
```
df = pd.read_excel('Xr04-03.xlsx')

print("Mean =",   df["Miles"].mean(),"miles")
print("Median =", df["Miles"].median(),"miles")
print("Mode =",   df["Miles"].mode(),"miles")
```

```
Mean = 8.541666666666666 miles
Median = 6.9 miles
Mode = 0      1.6
1      2.8
2      3.4
3      5.3
4      5.5
5      6.6
6      7.2
7      8.3
8      8.7
9     12.5
10    18.6
11    22.0
dtype: float64 miles
```

可以說沒有眾數，也可以說是每個數都是眾數。

Exercise 4.11 (25%)

rate of return 5%, mean 5%, median 5%, geometric mean 5%, explain 5%

In [3]:
```
data = [12, 10, 14 ,15, 22, 30, 25]

ROR = np.array([])
for i in range(1,len(data)):
    ror = (data[i]-data[i-1])/data[i-1]
    print("Rate of return of year ", i," = ", ror )
    ROR = np.append(ROR, ror)
```

```
Rate of return of year  1  =  -0.16666666666666666
Rate of return of year  2  =  0.4
Rate of return of year  3  =  0.07142857142857142
Rate of return of year  4  =  0.4666666666666667
Rate of return of year  5  =  0.36363636363636365
Rate of return of year  6  =  -0.16666666666666666
```

In [4]:
```
#geom 可直接用最後一期價格除以最初成本價格開6次方根計算
geom = (data[-1]/data[0])**(1/(len(data)-1)) -1

print("Mean =",   np.mean(ROR))
print("Median =", np.median(ROR))
print("Geometric mean =", geom)
```

```
Mean = 0.1613997113997114
Median = 0.21753246753246752
Geometric mean = 0.13012494323529933
```

Exercise 4.17 (15%)

mean 5%, median 5%, describe 5%

假設資料單位為公里/小時(km/hr)

In [5]:
```
df = pd.read_excel('Xr04-17.xlsx')
print("Mean =",   df["Speeds"].mean(),"km/hr")
print("Median =", df["Speeds"].median(),"km/hr")
```

```
Mean = 32.90833333333333 km/hr
Median = 32.0 km/hr
```

Exercise 4.31 (10%)

largest 5%, smallest 5%

可用全距、畫圖等方式描述，略。

Exercise 4.41 (15%)

varience 5%, standard deviation 5%, describe 5%

假設資料單位為公尺(m)

In [6]:
```
df = pd.read_excel('Xr04-41.xlsx')

print("var of Punter 1 =", df["Punter 1"].var(),"m2")
print("var of Punter 2 =", df["Punter 2"].var(),"m2")
print("var of Punter 3 =", df["Punter 3"].var(),"m2")
print("\n")
print("std of Punter 1 =", df["Punter 1"].std(),"m")
print("std of Punter 2 =", df["Punter 2"].std(),"m")
print("std of Punter 3 =", df["Punter 3"].std(),"m")
```

```
var of Punter 1 = 40.21591836734694 m2
var of Punter 2 = 14.806530612244902 m2
var of Punter 3 = 3.633061224489794 m2


std of Punter 1 = 6.341602192454753 m
std of Punter 2 = 3.847925494632777 m
std of Punter 3 = 1.9060590821088925 m
```

Exercise 4.47 (15%)

mean 5%, standard deviation 5%, describe 5%

假設資料單位為dollar

In [7]:
```
df = pd.read_excel('Xr04-47.xlsx')

print("mean of Property tax =", df["Property tax"].mean(),"dollars")
print("std of Property tax =", df["Property tax"].std(),"dollars")
```

```
mean of Property tax = 1937.3162857142854 dollars
std of Property tax = 949.9909560675101 dollars
```

Describe: 大多答案包含利用Chebyshev定理推論數值區間、利用positive skewed的前提推斷mean, median, mode大小等。