```
Chapter 3 - Regression Models

Part 2 - Multiple Linear Regression
```

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

from pylab import rcParams

from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import scale
```

In [2]:
```python
%matplotlib inline
rcParams['figure.figsize'] = 5,r
```

In [3]:
```python
import seaborn as sb
sb.set_style("whitegrid")
from collections import Counter
```

```
(Multiple)linear regression on the enrollment data
```

In [4]:
```python
address = 'C:/Users/danal/Desktop/Ex_Files_Python_Data_Science_EssT_Pt2/Exercise

enroll = pd.read_csv(address)
enroll.columns = ['year', 'roll', 'unem', 'hgrad','inc']
enroll.head()
```
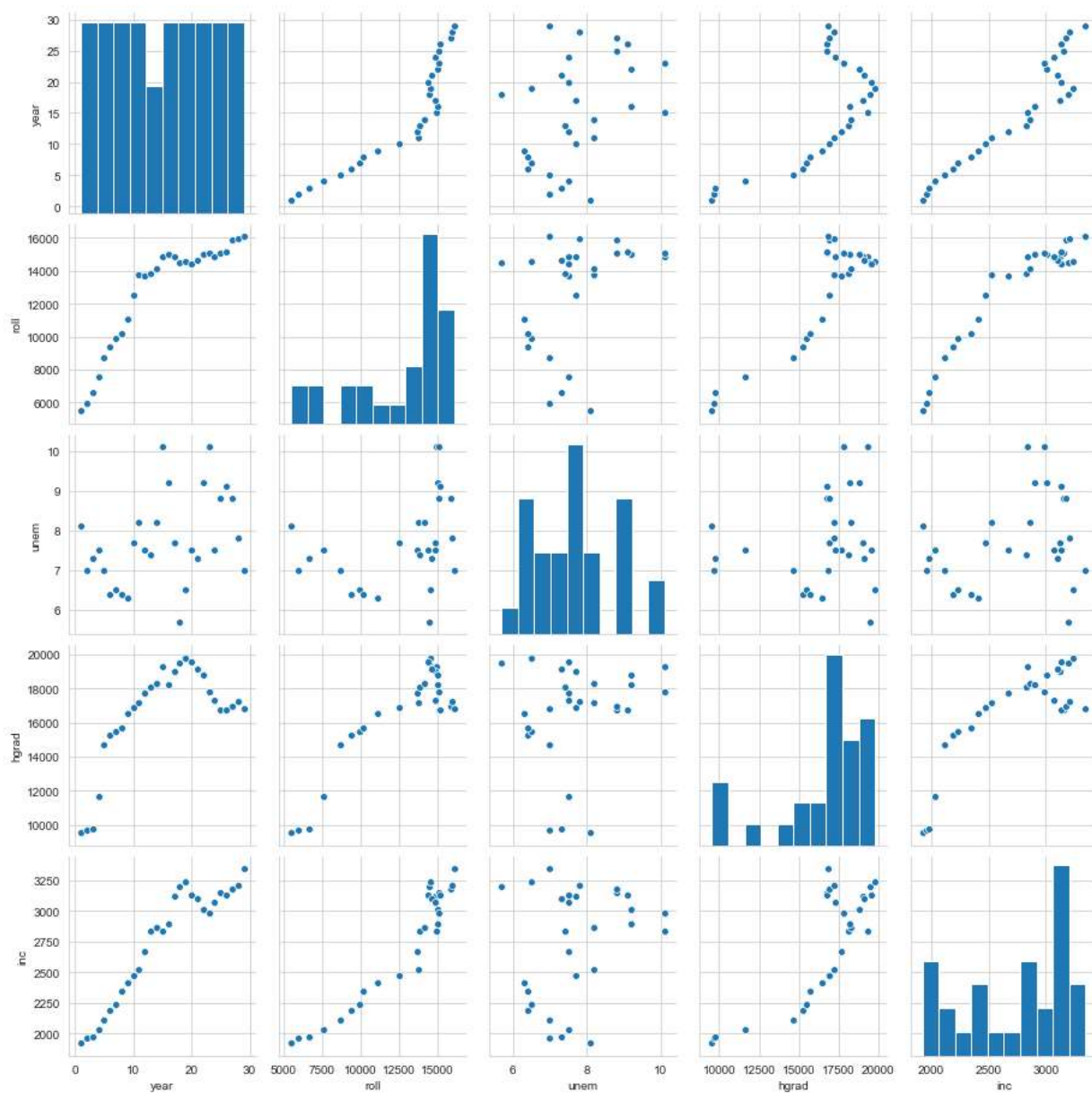
Out[4]:

	year	roll	unem	hgrad	inc
0	1	5501	8.1	9552	1923
1	2	5945	7.0	9680	1961
2	3	6629	7.3	9731	1979
3	4	7556	7.5	11666	2030
4	5	8716	7.0	14675	2112

In [5]: `sb.pairplot(enroll)`

Out[5]: `<seaborn.axisgrid.PairGrid at 0x1bf016da4c8>`

```
In [6]: print(enroll.corr())
```

```
              year      roll      unem      hgrad       inc
year      1.000000  0.900934  0.378305  0.670300  0.944287
roll      0.900934  1.000000  0.391344  0.890294  0.949876
unem      0.378305  0.391344  1.000000  0.177376  0.282310
hgrad     0.670300  0.890294  0.177376  1.000000  0.820089
inc       0.944287  0.949876  0.282310  0.820089  1.000000
```

```
In [7]: enroll_data = enroll[['unem', 'hgrad']].values

        enroll_target = enroll[['roll']].values

        enroll_data_names = ['unem', 'hgrad']

        x, y = scale(enroll_data), enroll_target
```

```
Checking for missing values
```

```
In [8]: missing_values = x==np.NAN
        x[missing_values == True]
```

```
Out[8]: array([], dtype=float64)
```

```
In [9]: LinReg = LinearRegression(normalize=True)

        LinReg.fit(x,y)

        print(LinReg.score(x,y))
```

```
0.8488812666133723
```

```
In [ ]:
```