Chapter 5 - Basic Math and Statistics

Segement 7 - Transforming dataset distributions

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

import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sb

import sklearn
from sklearn import preprocessing
from sklearn.preprocessing import scale
```

In [3]:
```python
%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
```
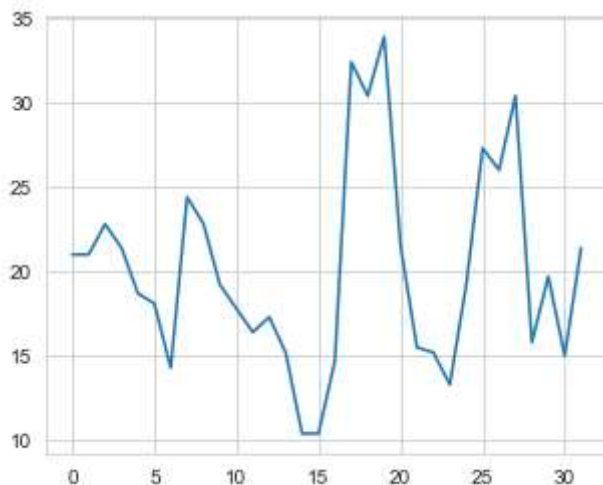
Normalizing and transforming features with MinMaxScalar() and fit_transform()

In [5]:
```python
address = 'C:/Users/danal/Desktop/ExerciseFiles/Data/mtcars.csv'

cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'v
```

In [6]:
```python
mpg = cars.mpg
plt.plot(mpg)
```

Out[6]: [<matplotlib.lines.Line2D at 0x13f39d41dc8>]
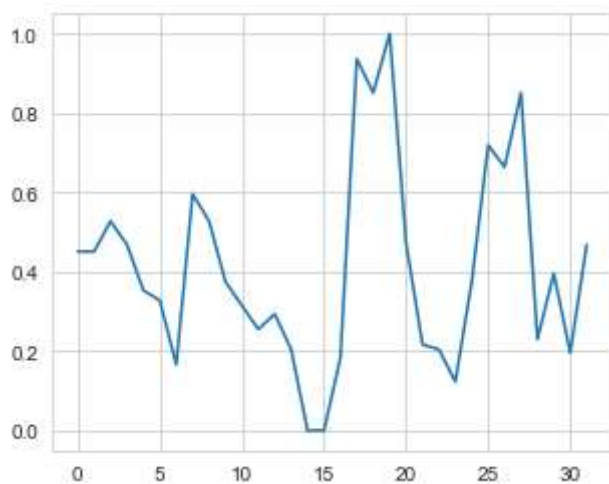
In [7]: `cars[['mpg']].describe()`

Out[7]:

	mpg
count	32.000000
mean	20.090625
std	6.026948
min	10.400000
25%	15.425000
50%	19.200000
75%	22.800000
max	33.900000

In [9]:
```
mpg_matrix = mpg.values.reshape(-1,1)

scaled = preprocessing.MinMaxScaler()

scaled_mpg = scaled.fit_transform(mpg_matrix)
plt.plot(scaled_mpg)
```
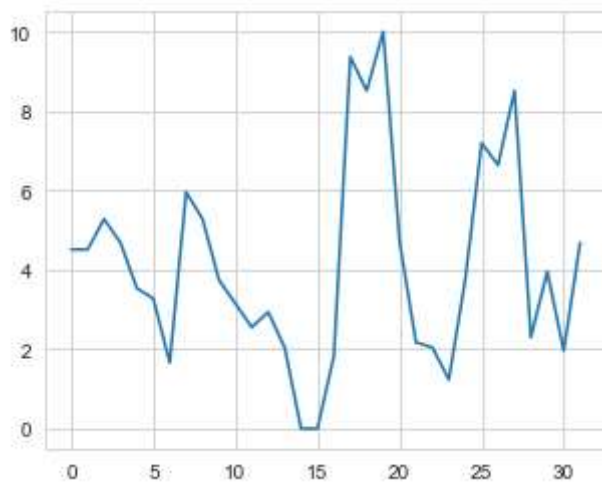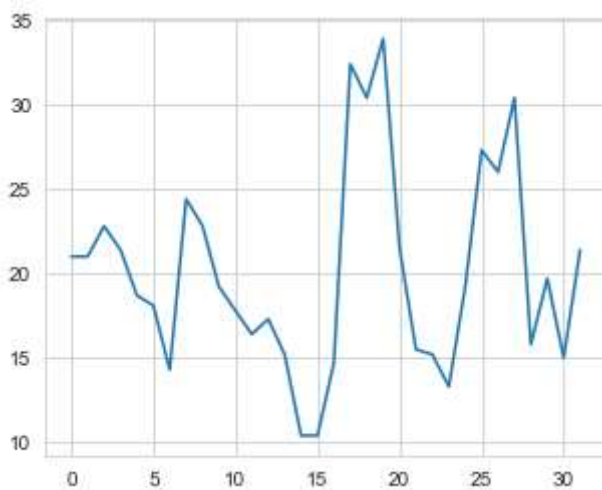
Out[9]: `[<matplotlib.lines.Line2D at 0x13f39e3b3c8>]`

In [11]:
```python
scaled = preprocessing.MinMaxScaler(feature_range=(0,10))

scaled_mpg = scaled.fit_transform(mpg_matrix)
plt.plot(scaled_mpg)
```

Out[11]: [<matplotlib.lines.Line2D at 0x13f39ead908>]

In [12]:
```python
standardized_mpg = scale(mpg, axis=0, with_mean=False, with_std=False)
plt.plot(standardized_mpg)
```

Out[12]: [<matplotlib.lines.Line2D at 0x13f39f0f988>]

In [13]:
```python
standardized_mpg = scale(mpg)
plt.plot(standardized_mpg)
```

Out[13]: [<matplotlib.lines.Line2D at 0x13f39f73a08>]