Importing the libraries

In [11]:

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Importing the datset

In [12]:

```python
dataset = pd.read_csv('Salary_Data.csv')
x=dataset.iloc[:, :-1].values
y=dataset.iloc[:, -1].values
```

In [13]:

```python
print(x)
```

```
[[ 1.1]
 [ 1.3]
 [ 1.5]
 [ 2. ]
 [ 2.2]
 [ 2.9]
 [ 3. ]
 [ 3.2]
 [ 3.2]
 [ 3.7]
 [ 3.9]
 [ 4. ]
 [ 4. ]
 [ 4.1]
 [ 4.5]
 [ 4.9]
 [ 5.1]
 [ 5.3]
 [ 5.9]
 [ 6. ]
 [ 6.8]
 [ 7.1]
 [ 7.9]
 [ 8.2]
 [ 8.7]
 [ 9. ]
 [ 9.5]
 [ 9.6]
 [10.3]
 [10.5]]
```

In [14]:

```
print(y)
```

```
[ 39343.   46205.   37731.   43525.   39891.   56642.   60150.   54445.   64445.
  57189.   63218.   55794.   56957.   57081.   61111.   67938.   66029.   83088.
  81363.   93940.   91738.   98273.  101302.  113812.  109431.  105582.  116969.
 112635.  122391.  121872.]
```

Spliting data for testing & training

In [15]:

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=1/3, random_state = 0)
```

In [16]:

```
print(x_train)
```

```
[[ 2.9]
 [ 5.1]
 [ 3.2]
 [ 4.5]
 [ 8.2]
 [ 6.8]
 [ 1.3]
 [10.5]
 [ 3. ]
 [ 2.2]
 [ 5.9]
 [ 6. ]
 [ 3.7]
 [ 3.2]
 [ 9. ]
 [ 2. ]
 [ 1.1]
 [ 7.1]
 [ 4.9]
 [ 4. ]]
```

In [17]:

```
print(x_test)
```

```
[[ 1.5]
 [10.3]
 [ 4.1]
 [ 3.9]
 [ 9.5]
 [ 8.7]
 [ 9.6]
 [ 4. ]
 [ 5.3]
 [ 7.9]]
```

In [18]:

```python
print(y_train)
```

```
[ 56642.  66029.  64445.  61111. 113812.  91738.  46205. 121872.  60150.
  39891.  81363.  93940.  57189.  54445. 105582.  43525.  39343.  98273.
  67938.  56957.]
```

In [19]:

```python
print(y_test)
```

```
[ 37731. 122391.  57081.  63218. 116969. 109431. 112635.  55794.  83088.
 101302.]
```

Training the simple linear regression

In [20]:

```python
from sklearn.linear_model import LinearRegression

regressor= LinearRegression()
regressor.fit(x_train, y_train)
```

Out[20]:

```
LinearRegression()
```

In [21]:

```python
#predicting the test set results
y_pred=regressor.predict(x_test) #x_test=original data set, y_pred=predicted
```

In [24]:

```python
#visualising the training set tesults
plt.scatter(x_train,y_train,color='red')
plt.plot(x_train, regressor.predict(x_train), color='blue' )
plt.title('Salary vs experience (Training set)')
plt.xlabel('Years of experience')
plt.ylabel('Salary')
plt.show()
```


In [27]:

```python
#visualising the testing set tesults
plt.scatter(x_test,y_test,color='red')
plt.plot(x_test, regressor.predict(x_test), color='blue' )
plt.title('Salary vs experience (Test set)')
plt.xlabel('Years of experience')
plt.ylabel('Salary')
plt.show()
```


In [ ]: