

```
In [1]: import pandas as pd
```

```
In [3]: df=pd.read_csv("f:/dataset/regression/salary_1_variable.csv")
X=df.iloc[:, :-1].values
y=df.iloc[:, -1].values
```

```
In [4]: from sklearn.linear_model import LinearRegression
```

```
In [5]: model=LinearRegression() #fit_intercept=True
model.fit(X,y)
print(model.coef_) #m(w)
print(model.intercept_) #c(b)
```

```
[9449.96232146]
25792.20019866871
```

```
In [6]: model=LinearRegression(fit_intercept=False)
model.fit(X,y)
print(model.coef_) #m(w)
print(model.intercept_) #c(b)
```

```
[13254.93845442]
0.0
```

```
In [7]: model=LinearRegression() #fit_intercept=True
model.fit(X,y)
print(model.coef_) #m(w)
print(model.intercept_) #c(b)
```

```
[9449.96232146]
25792.20019866871
```

```
In [12]: pred=model.predict(X)
pred
```

```
Out[12]: array([ 36187.15875227,  38077.15121656,  39967.14368085,  44692.12484158,
        46582.11730587,  53197.09093089,  54142.08716303,  56032.07962732,
        56032.07962732,  60757.06078805,  62647.05325234,  63592.04948449,
        63592.04948449,  64537.04571663,  68317.03064522,  72097.0155738 ,
        73987.00803809,  75877.00050238,  81546.97789525,  82491.9741274 ,
        90051.94398456,  92886.932681 , 100446.90253816, 103281.8912346 ,
       108006.87239533, 110841.86109176, 115566.84225249, 116511.83848464,
       123126.81210966, 125016.80457395])
```

```
In [10]: y
```

```
Out[10]: array([ 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], dtype=int64)
```

```
In [13]: y-pred
```

```
Out[13]: array([ 3155.84124773,  8127.84878344, -2236.14368085, -1167.12484158,
       -6691.11730587,  3444.90906911,  6007.91283697, -1587.07962732,
        8412.92037268, -3568.06078805,  570.94674766, -7798.04948449,
       -6635.04948449, -7456.04571663, -7206.03064522, -4159.0155738 ,
       -7958.00803809,  7210.99949762, -183.97789525, 11448.0258726 ,
        1686.05601544,  5386.067319 ,  855.09746184, 10530.1087654 ,
       1424.12760467, -5259.86109176,  1402.15774751, -3876.83848464,
       -735.81210966, -3144.80457395])
```

```
In [14]: (y-pred)**2
```

```
Out[14]: array([9.95933398e+06, 6.60619258e+07, 5.00033856e+06, 1.36218040e+06,
                4.47710508e+07, 1.18673985e+07, 3.60950167e+07, 2.51882174e+06,
                7.07772292e+07, 1.27310578e+07, 3.25980189e+05, 6.08095758e+07,
                4.40238817e+07, 5.55926177e+07, 5.19268777e+07, 1.72974105e+07,
                6.33298919e+07, 5.19985138e+07, 3.38478659e+04, 1.31057296e+08,
                2.84278489e+06, 2.90097212e+07, 7.31191669e+05, 1.10883191e+08,
                2.02813943e+06, 2.76661387e+07, 1.96604635e+06, 1.50298766e+07,
                5.41419461e+05, 9.88979581e+06])
```

```
In [15]: sum((y-pred)**2)
```

```
Out[15]: 938128551.668429
```

```
In [16]: sum((y-pred)**2)/len(y)
```

```
Out[16]: 31270951.722280968
```

```
In [17]: model.coef_
```

```
Out[17]: array([9449.96232146])
```

```
In [18]: model.intercept_
```

```
Out[18]: 25792.20019866871
```

```
In [19]: from sklearn.metrics import mean_squared_error
```

```
In [20]: mean_squared_error(y,pred)
```

```
Out[20]: 31270951.722280968
```

```
In [ ]:
```