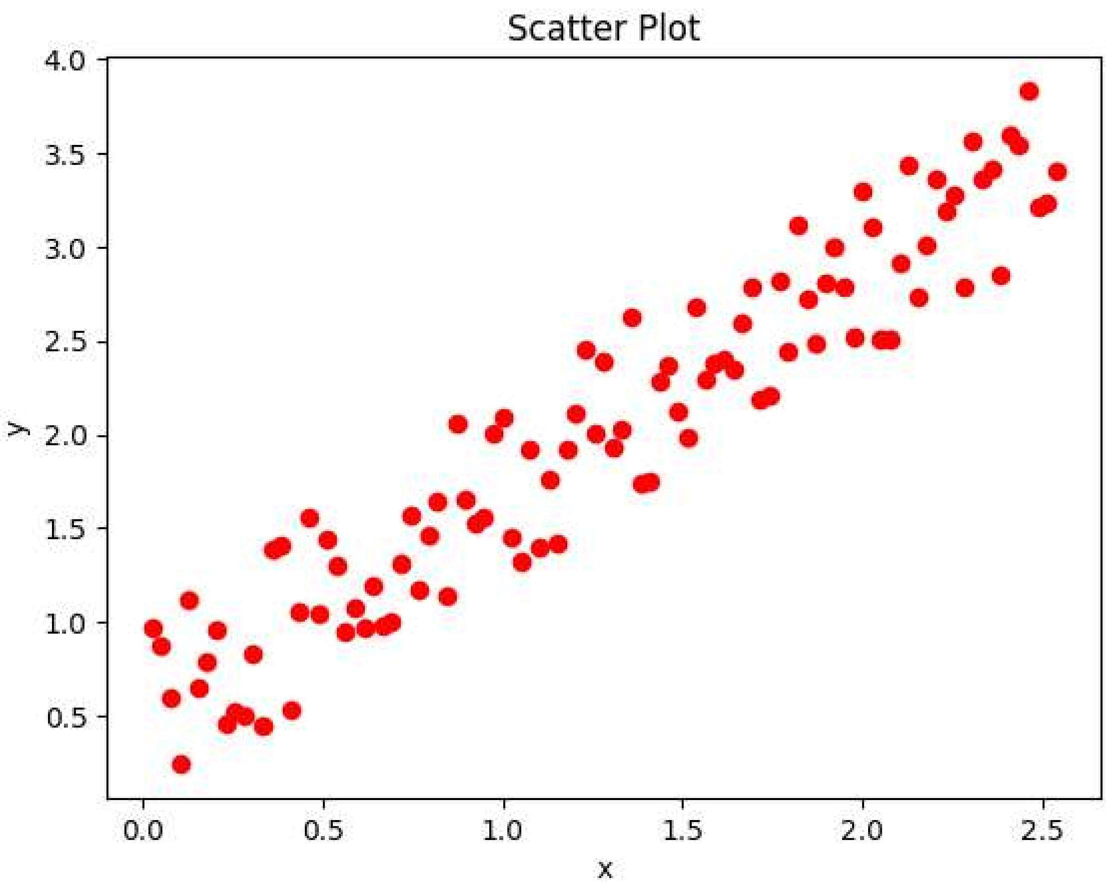
In [ ]:
```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

In [ ]:
```
arr = np.loadtxt("linear-data-set-for-regression.csv", delimiter = ',', skiprows =
```

In [ ]:
```
x = arr[:,1]
y = arr[:,0]
```

In [ ]:
```
plt.scatter(x,y,color="Red")
plt.title("Scatter Plot")
plt.ylabel("y")
plt.xlabel("x")
```

Out[ ]:
```
Text(0.5, 0, 'x')
```



In [ ]:
```
xy_bar = np.mean(x*y)
x_sq_bar = np.mean(x*x)
x_bar = np.mean(x)
y_bar = np.mean(y)
```

In [ ]:
```
a = (xy_bar - x_bar * y_bar)/(x_sq_bar - x_bar*x_bar)
b = y_bar - a * x_bar
print ("a =",a)
```

```
print("\n")
print ("b = ",b)
```

```
a = 1.1719511804459168


b =  0.5156466449319419
```

In [ ]:
```
y_cap = a*x + b
print (y_cap)
```

```
[0.54569668 0.57574671 0.60579674 0.63584677 0.6658968  0.69594683
 0.72599686 0.75604689 0.78609692 0.81614695 0.84619698 0.87624701
 0.90629704 0.93634707 0.9663971  0.99644713 1.02649716 1.05654719
 1.08659722 1.11664725 1.14669728 1.17674731 1.20679734 1.23684737
 1.2668974  1.29694743 1.32699746 1.35704749 1.38709752 1.41714755
 1.44719758 1.47724761 1.50729764 1.53734767 1.5673977  1.59744773
 1.62749777 1.65754779 1.68759783 1.71764786 1.74769789 1.77774792
 1.80779795 1.83784798 1.86789801 1.89794804 1.92799807 1.9580481
 1.98809813 2.01814816 2.04819819 2.07824822 2.10829825 2.13834828
 2.16839831 2.19844834 2.22849837 2.2585484  2.28859843 2.31864846
 2.34869849 2.37874852 2.40879855 2.43884858 2.46889861 2.49894864
 2.52899867 2.5590487  2.58909873 2.61914876 2.64919879 2.67924882
 2.70929885 2.73934888 2.76939891 2.79944895 2.82949898 2.85954901
 2.88959904 2.91964907 2.9496991  2.97974913 3.00979916 3.03984919
 3.06989922 3.09994925 3.12999928 3.16004931 3.19009934 3.22014937
 3.2501994  3.28024943 3.31029946 3.34034949 3.37039952 3.40044955
 3.43049958 3.46054961 3.49059964]
```

In [ ]:
```
e_i = y_cap - y
print(e_i)
```

```
[-0.42921961 -0.3041557   0.003199    0.38875894 -0.45047668  0.04288641
 -0.06801934 -0.2003581   0.32332505  0.29724771  0.34304666  0.03823256
  0.46084051 -0.44878168 -0.43931128  0.45906319 -0.03524993 -0.50013961
  0.04524065 -0.32896186 -0.16011529  0.2267425   0.13224475  0.26250721
  0.07319382  0.31753527  0.32328892  0.04099855 -0.18538397  0.24540027
 -0.01995145 -0.16652733  0.36884463 -0.52256392 -0.08287861  0.07124091
  0.07317348 -0.35029913 -0.40790023  0.27036007  0.4278492  -0.14640593
  0.40433763  0.08161118  0.44654862 -0.02749931 -0.18828603 -0.50092782
 -0.02190493 -0.37648907  0.11970079  0.05457808 -0.52098716  0.4007676
  0.4228406  -0.08546562 -0.13809529  0.13183617  0.301712   -0.36118439
  0.05224732 -0.003903    0.01005947  0.09495877 -0.11969188 -0.28906123
  0.34149667  0.35491056 -0.22997198  0.17782375 -0.46954154 -0.03752258
  0.21812885 -0.06414492 -0.23245067  0.01538932  0.30844589 -0.43390098
 -0.21426267  0.41648928  0.44498424  0.06310827 -0.42396298  0.30657574
  0.05931381 -0.26261984 -0.05602272 -0.11698758  0.4047754  -0.34574531
 -0.10639372 -0.13511316  0.46241066 -0.25731245 -0.16864218 -0.43045711
  0.22233108  0.22359399  0.08905178]
```

In [ ]:
```
statistic,p_value = stats.normaltest(e_i)
print("p_value is", p_value)
```

```
p_value is 0.0001567360662763278
```

Since the p_value is less than 0.05, we can conclude that this is not a normal distribution.

In [ ]:

```
y_cap_minus_y_i_sq = (y_cap - y)**2
y_cap_minus_y_bar_sq = (y_cap - y_bar)**2
y_i_minus_y_bar_sq = (y - y_bar)**2
```

In [ ]:

```
SSE = np.sum(y_cap_minus_y_bar_sq)
SSR = np.sum(y_cap_minus_y_i_sq)
SST = np.sum(y_i_minus_y_bar_sq)
R2 = SSE/SST
```

In [ ]:

```
print(f'SSR = {SSR} \n\n SSE = {SSE} \n\n SST ={SST} \n\n R2 ={R2}')
```

```
SSR = 8.09913676760857

 SSE = 73.00789919833159

 SST =81.10703596594

 R2 =0.9001426119059565
```

In [ ]:

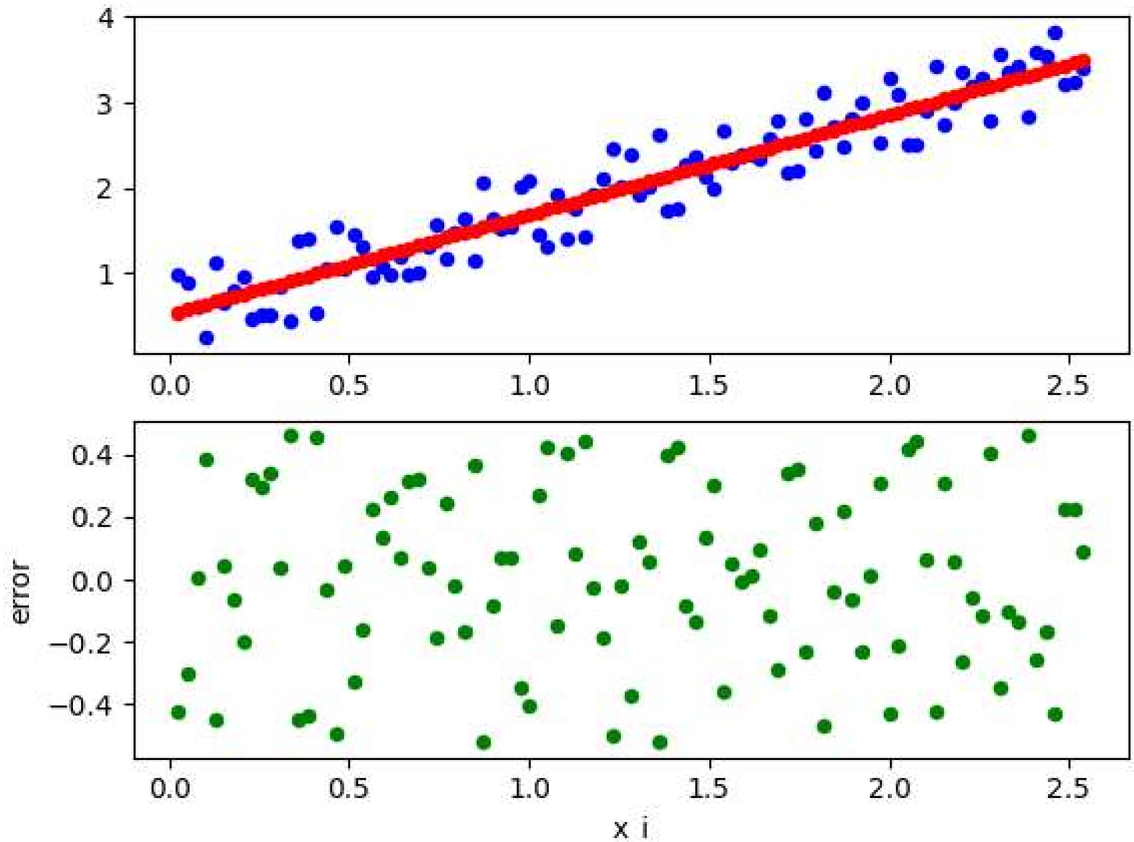
```
print(SSE + SSR)
print(SST)
```

```
81.10703596594016
81.10703596594
```

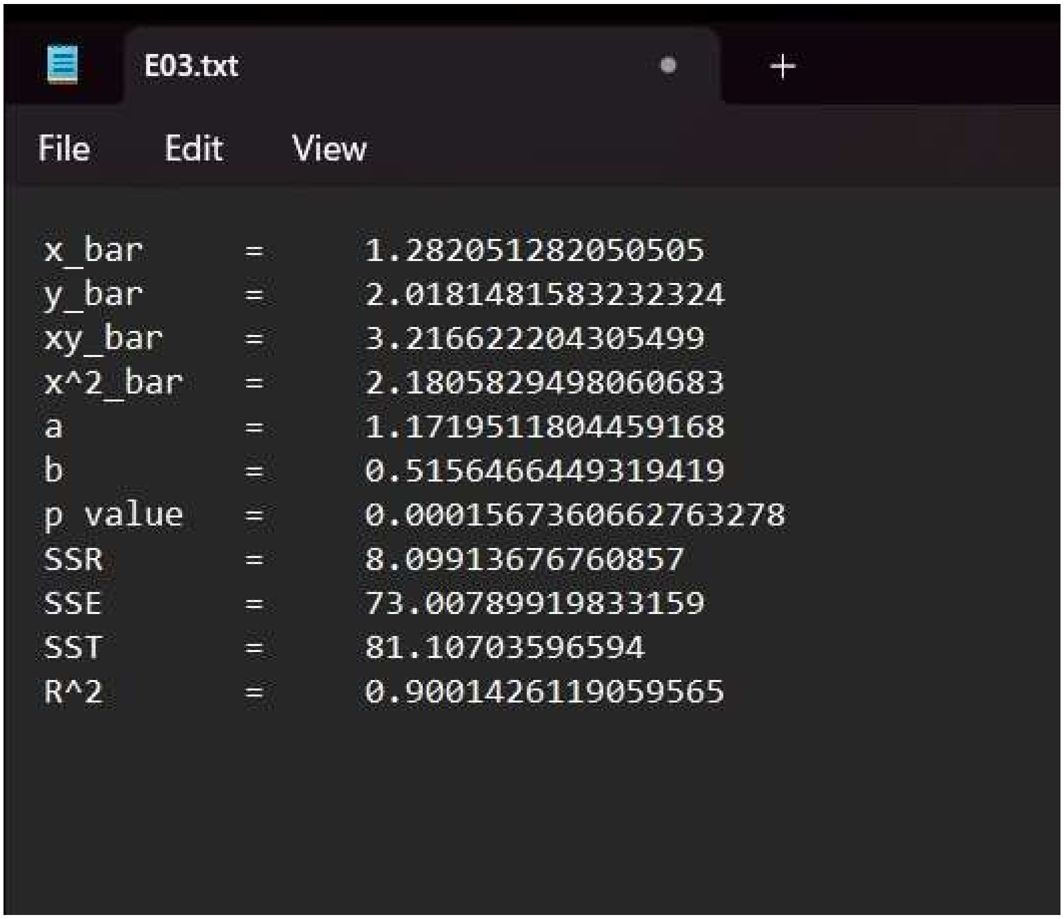
Here we can see that SST = SSE + SSR

In [ ]:

```
fig,axes=plt.subplots(2,1)
axes[0].scatter(x,y,color='Blue',s=20)
axes[0].scatter(x,y_cap,color='Red',s=20)
plt.xlabel('x_i')
plt.ylabel('error')
axes[1].scatter(x,e_i,color='Green',s=20)
plt.show()
```



In [ ]:
```
with open('eo3.txt','w') as output:
    output.writelines([
        f'x_bar = {x_bar}\n',
        f'y_bar = {y_bar}\n',
        f'xy_bar = {xy_bar}\n',
        f'x^2_bar = {x_sq_bar}\n',
        f'a = {a}\n',
        f'b = {b}\n',
        f'p value = {p_value}\n',
        f'SSR = {SSR}\n',
        f'SSE = {SSE}\n',
        f'SST = {SST}\n',
        f'R^2 = {R2}\n'])
```



E03.txt

File Edit View

```
x_bar      =     1.282051282050505
y_bar      =     2.0181481583232324
xy_bar     =     3.216622204305499
x^2_bar    =     2.1805829498060683
a          =     1.1719511804459168
b          =     0.5156466449319419
p value    =     0.00015673606627632 78
SSR        =     8.09913676760857
SSE        =     73.00789919833159
SST        =     81.10703596594
R^2        =     0.9001426119059565
```

```
In [ ]:
```