```
In [5]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import statsmodels.formula.api as sm
         %matplotlib inline
```

```
In [40]:  data = pd.read_csv(r"R:\Usuarios\OMAR\FCFM\7 SEMESTRE\MINERIA DE DATOS\
          TAREAS\datos.csv")
```

```
In [41]:  data.head()
```

Out[41]:

	PESO	ALTURA
0	86.78	162
1	74.11	212
2	71.73	220
3	69.88	206
4	67.25	152

```
In [46]:  lm = sm.ols(formula="PESO~ALTURA",data = data).fit()
```

```
In [47]:  lm.params
```

```
Out[47]:  Intercept    55.507502
          ALTURA        0.076392
          dtype: float64
```

Modelo lineal predictivo :

.ALTURA= 71.002495 + 1.588166* PESO

```
In [48]:  lm.pvalues
```

```
Out[48]:  Intercept    1.517694e-08
          ALTURA       5.925591e-02
          dtype: float64
```

```
In [50]:  lm.rsquared
```

Out[50]: 0.12132341067587182

```
In [51]:  lm.rsquared_adj
```
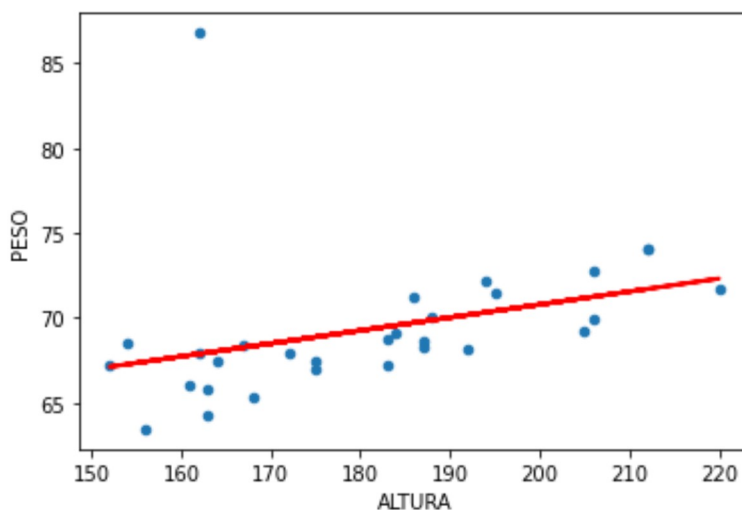
Out[51]: 0.08994210391429569

```
In [54]: peso_pred = lm.predict(pd.DataFrame(data["ALTURA"]))
         peso_pred
```

```
Out[54]: 0     67.883030
         1     71.702638
         2     72.313775
         3     71.244285
         4     67.119109
         5     69.487265
         6     68.264991
         7     68.876128
         8     67.424677
         9     69.716442
         10    69.487265
         11    67.959422
         12    67.959422
         13    68.646952
         14    70.327579
         15    68.341383
         16    67.806638
         17    68.035815
         18    69.869226
         19    69.792834
         20    67.883030
         21    70.174795
         22    69.563657
         23    71.244285
         24    68.876128
         25    67.271893
         26    69.792834
         27    71.702638
         28    70.403971
         29    71.167893
         dtype: float64
```

```
In [55]: data.plot(kind = "scatter" ,x ="ALTURA", y ="PESO")
         plt.plot(pd.DataFrame(data["ALTURA"]),altura_pred,c="red", linewidth=2)
```

```
Out[55]: [<matplotlib.lines.Line2D at 0x16074770af0>]
```

In [ ]: