l-plot-in-simple-linear-regression

April 7, 2024

Import packages and libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

Reading the csv file

[3]:
```
data = pd.read_csv('/content/headbrain3.csv')
data.head()
```

[3]:

	Unnamed: 0	index	Head_size	Brain_weight
0	0	0	4512	1530
1	1	1	3738	1297
2	2	2	4261	1335
3	3	3	3777	1282
4	4	4	4177	1590

Fit simple linear regression model

[4]:
```
linear_model = ols('Brain_weight ~ Head_size', data=data).fit()
```

Display the model summary

[5]:
```
print(linear_model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           Brain_weight   R-squared:                       0.516
Model:                            OLS   Adj. R-squared:                  0.511
Method:                 Least Squares   F-statistic:                     105.4
Date:                Sun, 07 Apr 2024   Prob (F-statistic):           2.85e-17
Time:                        11:26:06   Log-Likelihood:                -580.70
No. Observations:                 101   AIC:                             1165.
Df Residuals:                      99   BIC:                             1171.
Df Model:                           1
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    437.4384     87.990      4.971      0.000     262.847     612.030
Head_size      0.2360      0.023     10.268      0.000       0.190       0.282
==============================================================================
Omnibus:                        1.314   Durbin-Watson:                   2.023
Prob(Omnibus):                  0.518   Jarque-Bera (JB):                1.019
Skew:                           0.244   Prob(JB):                        0.601
Kurtosis:                       3.061   Cond. No.                     4.41e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 4.41e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

modify figure size

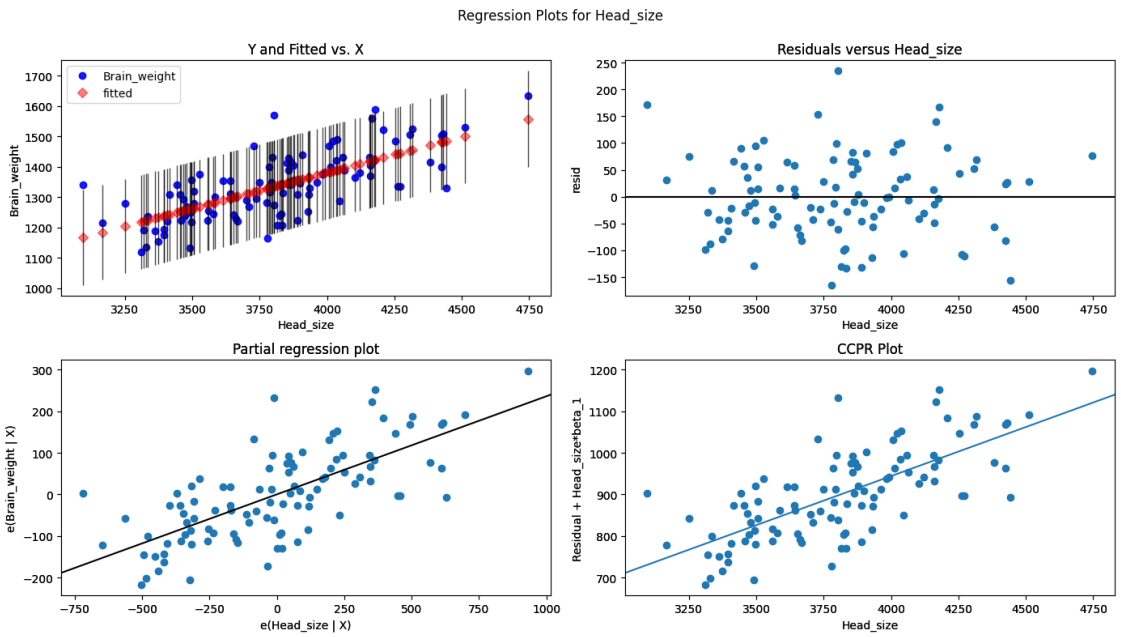
[7]:
```
fig = plt.figure(figsize=(14, 8))
```

```
<Figure size 1400x800 with 0 Axes>
```

creating regression plots

[9]:
```
fig = sm.graphics.plot_regress_exog(linear_model,
                                    'Head_size',
                                    fig=fig)
fig
```

[9]:



In conclusion we can see that the points are plotted randomly spread or scattered.points or residuals are scattered around the ‘0’ line,there is no pattern, and points are not based on the side so there is no problem of heteroscedastisity.with the predictor variable ‘Head_size’ theres no heteroscedasticity