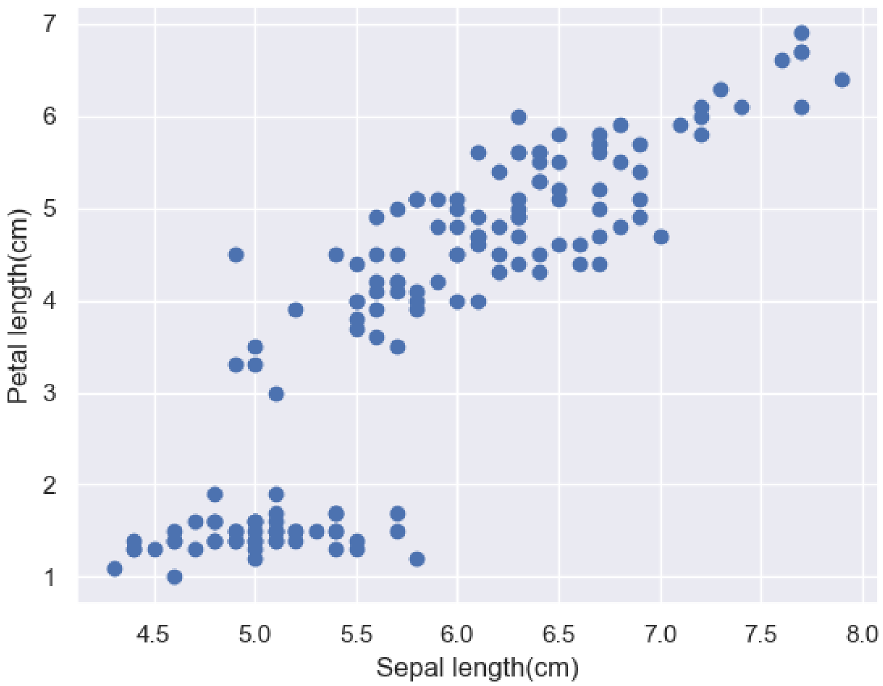
Term Report

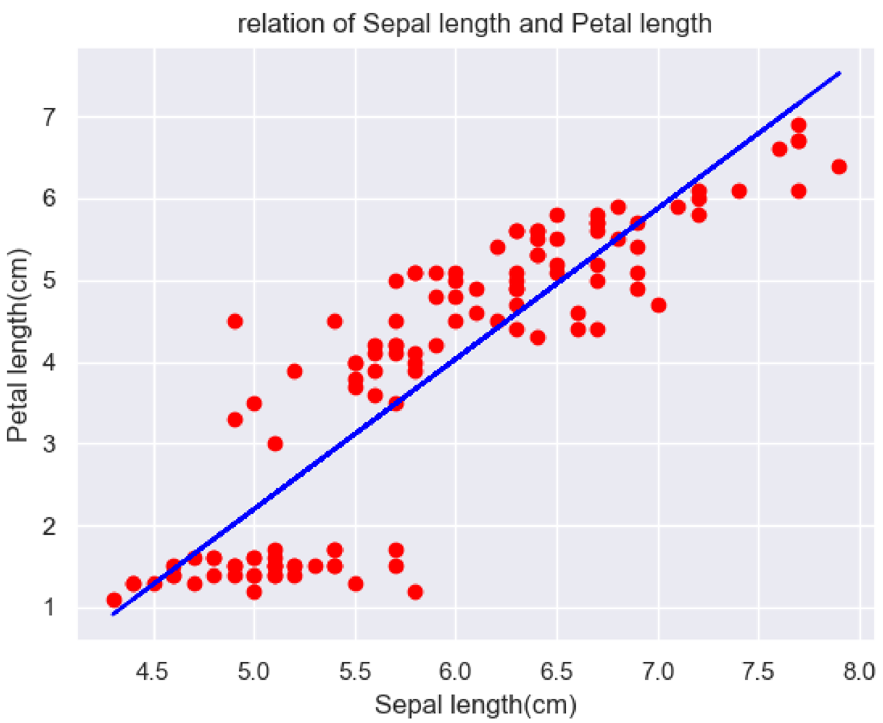
1106022 陳柏嘉

1.



資量呈現正相關，能夠使用線性回歸進行分析！

分析結果:



2.

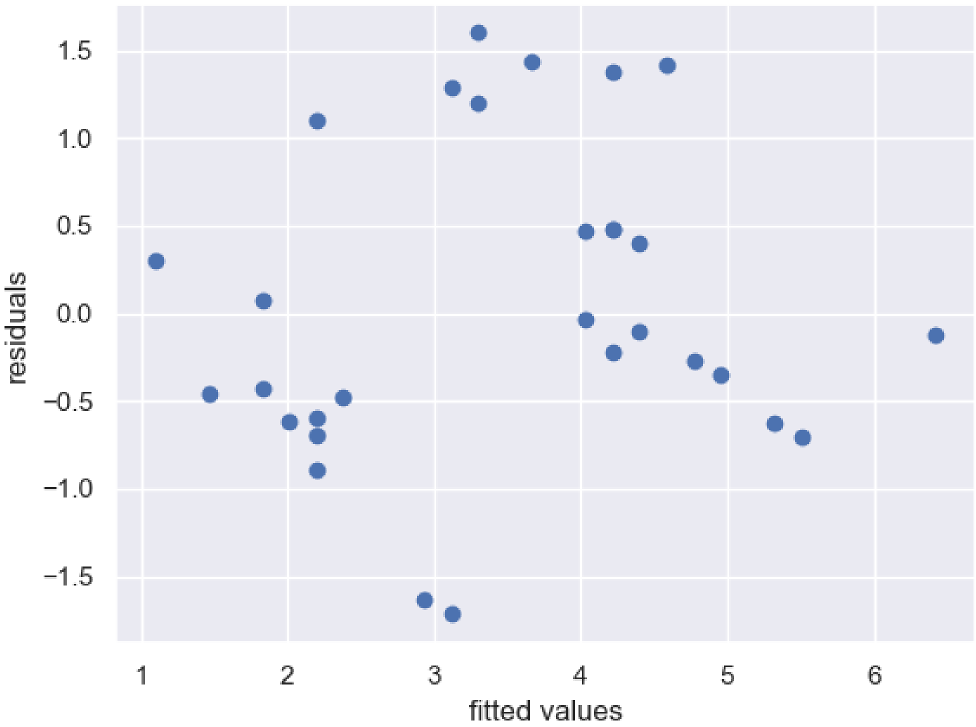
```
Interception :  -6.9936345301227485
Coeficient :  [1.83808438]
Score:  0.7214093715123957
Accuracy: 72.14093715123957%
Durbin-Watson statistic:  2.147908354176179
```

DW statistic values 在 1.5~2.5 之間，屬於正常值。

3.

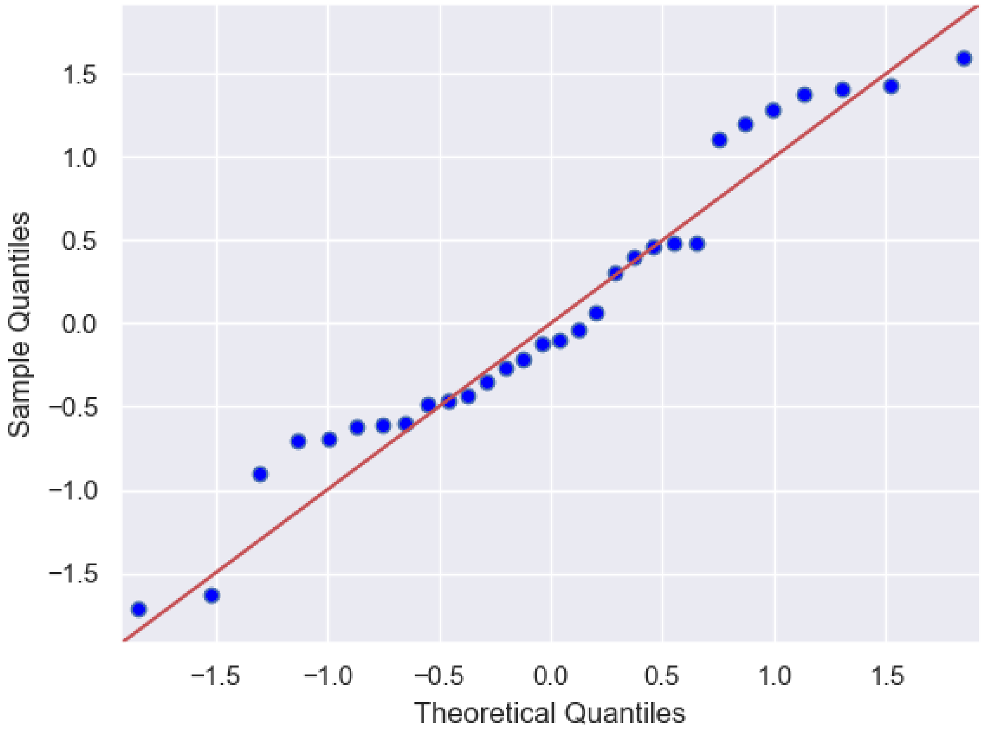
資料集只有 2 個 variables，因此沒有這個問題。

4.



殘差與擬合值圖並未呈現漏斗形狀。

The 5th assumption: **Normal Distribution of error terms**



QQ 圖呈現接近常態分佈。

Code:

```
import matplotlib.pyplot as plt
import seaborn as sns; sns.set()
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn import datasets
from statsmodels.stats.stattools import durbin_watson
import pandas as pd
from statsmodels.stats.outliers_influence import
variance_inflation_factor
import statsmodels.api as sm

df = datasets.load_iris()

X=df.data[:, 0].reshape(-1,1)    # 花萼長度
Y=df.data[:, 2]  # 花瓣長度

plt.scatter(X, Y)
plt.xlabel("Sepal length(cm)")
plt.ylabel("Petal length(cm)")
plt.show() # x, y 的散點圖

X_train, X_test, y_train, y_test = train_test_split(X,
Y,test_size=0.2,random_state=0)

model = LinearRegression() # 建立線性回歸模型
model.fit(X_train, y_train)

# 計算出截距值與係數值
w_0 = model.intercept_
w_1 = model.coef_
print('Interception : ', w_0)
print('Coeficient : ', w_1)

# 迴歸模型的準確度
score = model.score(X_test, y_test)
print('Score: ', score)
print('Accuracy: ' + str(score*100) + '%')
```

```python
# 視覺化迴歸模型與訓練集的關聯
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, model.predict(X_train), color = 'blue')
plt.title('relation of Sepal length and Petal length')
plt.xlabel("Sepal length(cm)")
plt.ylabel("Petal length(cm)")
plt.show()

y_pred = model.predict(X_test)
residuals = y_test - y_pred
plt.scatter(y_pred, residuals)
plt.xlabel("fitted values")
plt.ylabel("residuals")
plt.show()

# Durbin-Watson statistic values
print('Durbin-Watson statistic: ', durbin_watson(residuals))

# QQ 圖
fig = sm.qqplot(residuals, line='45')
plt.show()
```