

# 回歸分析期中報告

M122040017

吳俞憲



# 資料蒐集以及變數的選擇

從縣市重要統計指標查詢系統中,蒐集2010~2020這10年間南部地區的資料

南部地區包含:嘉義市、嘉義縣、台南市、高雄市、屏東縣

網站網址:[https://winsta.dgbas.gov.tw/DgbasWeb/ZWeb/StateFile\\_ZWeb.aspx](https://winsta.dgbas.gov.tw/DgbasWeb/ZWeb/StateFile_ZWeb.aspx)



# 資料蒐集以及變數的選擇

- 初始變數的選擇
  - 粗出生率(‰) (CBR)
  - 就業者之年齡別結構-25-44歲(%) (ER)
  - 犯罪人口率(人/十萬人) (CR)
  - 青壯年人口比率(15-64歲)(%) (YAR)
  - 低收入戶人口數占總人口比率(%) (LOWR)

# 資料蒐集以及變數的選擇

我們先以粗出生率為反應變數,其餘的變數為解釋變數作線性回歸

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.759e+01	7.922e+00	-3.483	0.001040	**
ER	1.188e-01	3.090e-02	3.845	0.000342	***
CR	-6.842e-04	6.174e-04	-1.108	0.273096	
YAR	4.145e-01	1.167e-01	3.553	0.000842	***
LOWM	-7.377e-01	2.461e-01	-2.998	0.004226	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6788 on 50 degrees of freedom

Multiple R-squared: 0.5916, Adjusted R-squared: 0.5589

F-statistic: 18.1 on 4 and 50 DF, p-value: 2.998e-09

# 資料蒐集以及變數的選擇

比較有CR跟沒有CR的模型的 $R^2$

有CR的模型

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.759e+01	7.922e+00	-3.483	0.001040	**
ER	1.188e-01	3.090e-02	3.845	0.000342	***
CR	-6.842e-04	6.174e-04	-1.108	0.273096	
YAR	4.145e-01	1.167e-01	3.553	0.000842	***
LOWM	-7.377e-01	2.461e-01	-2.998	0.004226	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6788 on 50 degrees of freedom

Multiple R-squared: 0.5916, Adjusted R-squared: 0.5589

F-statistic: 18.1 on 4 and 50 DF, p-value: 2.998e-09

沒有CR的模型

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-31.50904	7.10517	-4.435	4.93e-05	***
ER	0.11897	0.03097	3.841	0.000340	***
YAR	0.45667	0.11051	4.132	0.000134	***
LOWM	-0.75284	0.24622	-3.058	0.003549	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6803 on 51 degrees of freedom

Multiple R-squared: 0.5815, Adjusted R-squared: 0.5569

F-statistic: 23.62 on 3 and 51 DF, p-value: 1.006e-09



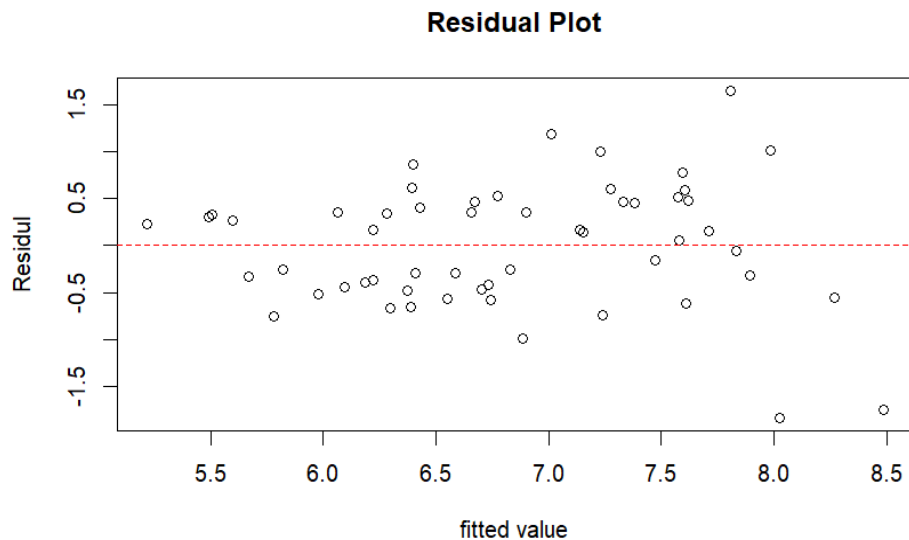
# Diagnostics

接下來我們會以沒有CR的模型來分析是否有符合線性模型的假設

- 誤差的方差是否等於常數
- 變數是否與反應變數有非線性關係
- 誤差是否為常態分佈

# 判斷誤差是否為常數

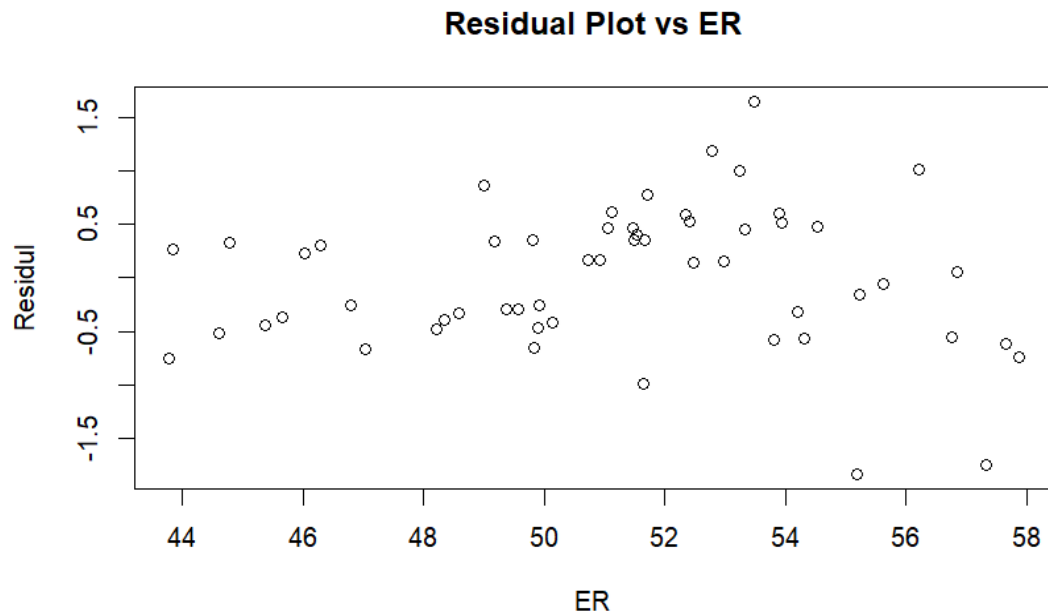
從圖中可以發現大部分的點是均勻的在一個水平帶子裡面  
所以可以推斷方差為常數



# 變數是否與反應變數有非線性關係

## 1. 就業者之年齡別結構-25-44歲(%) (ER)

從圖中可以看出沒有非線性的關係

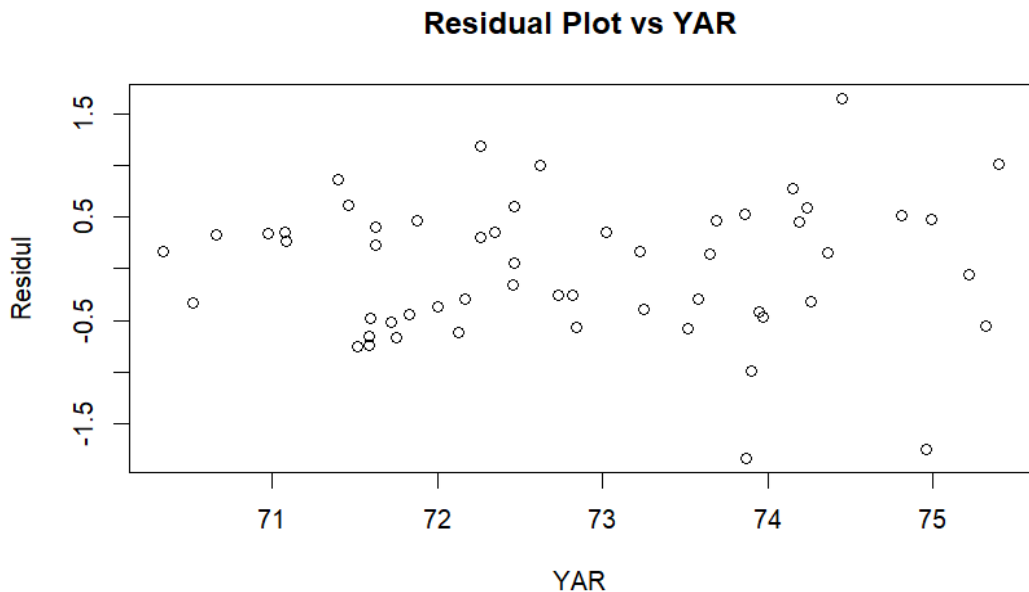




# 變數是否與反應變數有非線性關係

## 2. 青壯年人口比率(15-64歲)(%) (YAR)

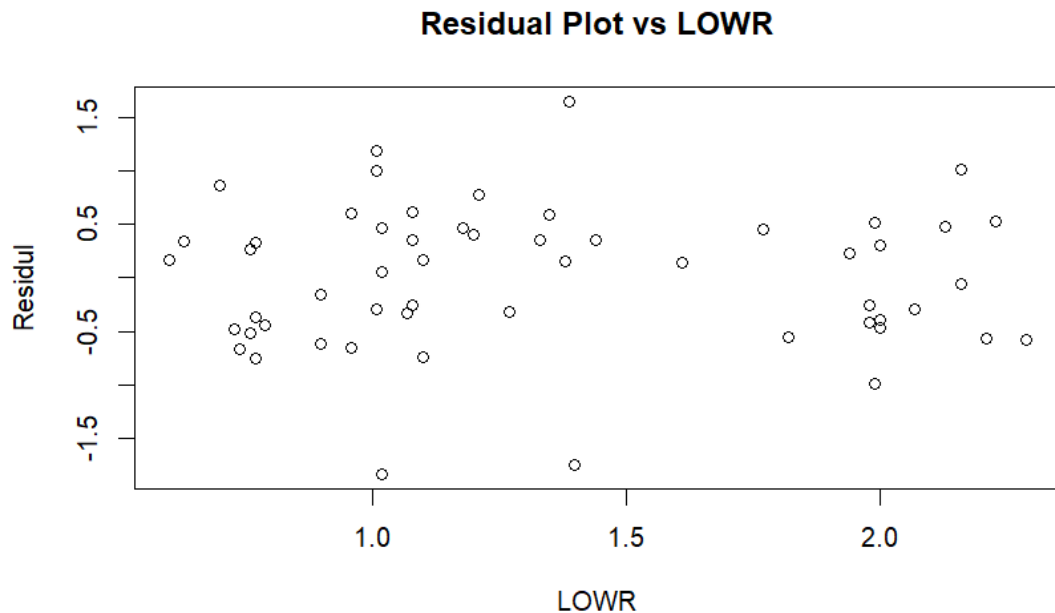
從圖中可以看出沒有非線性關係



# 變數是否與反應變數有非線性關係

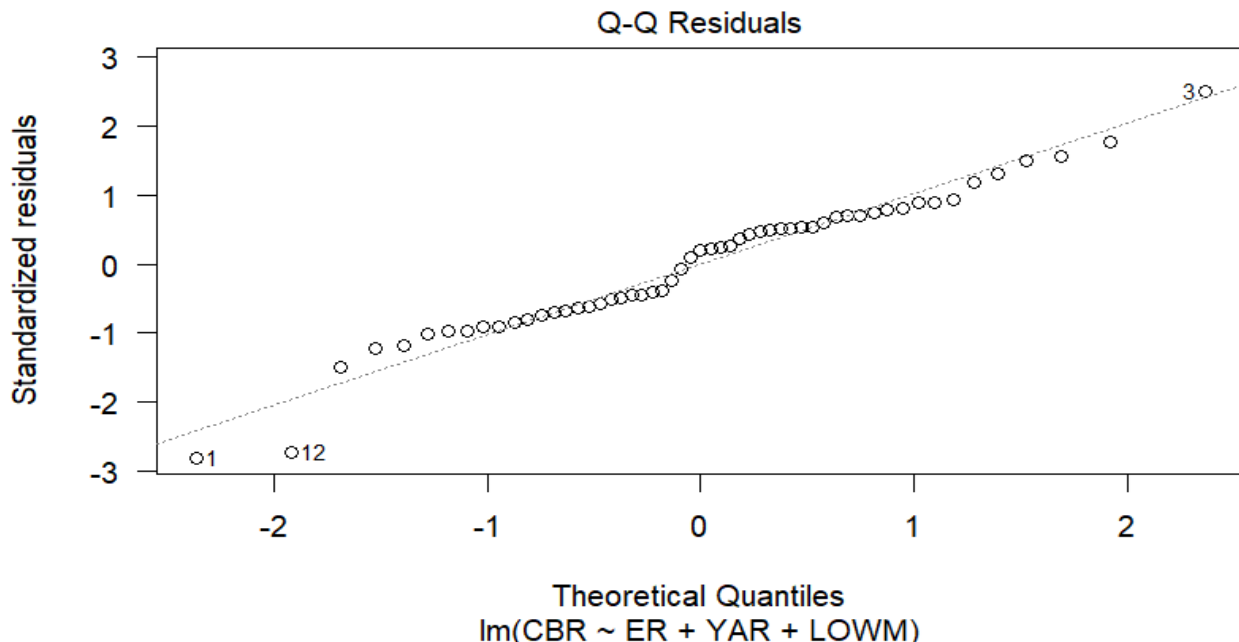
## 3. 低收入戶人口數占總人口比率(%) (LOWR)

從圖中可以看出沒有非線性關係



# 誤差是否為常態分佈

利用Q-Q Plot可以看出大部分的點都落在直線上，  
所以可以判斷出分配很接近常態分佈





# 結論

儘管模型滿足初步回歸模型的假設，然而考慮到樣本數相對較少，以及 $R^2$ 值的沒有很高，我們不能直接說這個模型是好的。因此，我們必須謹慎對待由這個模型所得出的結論，不應該對其結果過於絕對肯定。