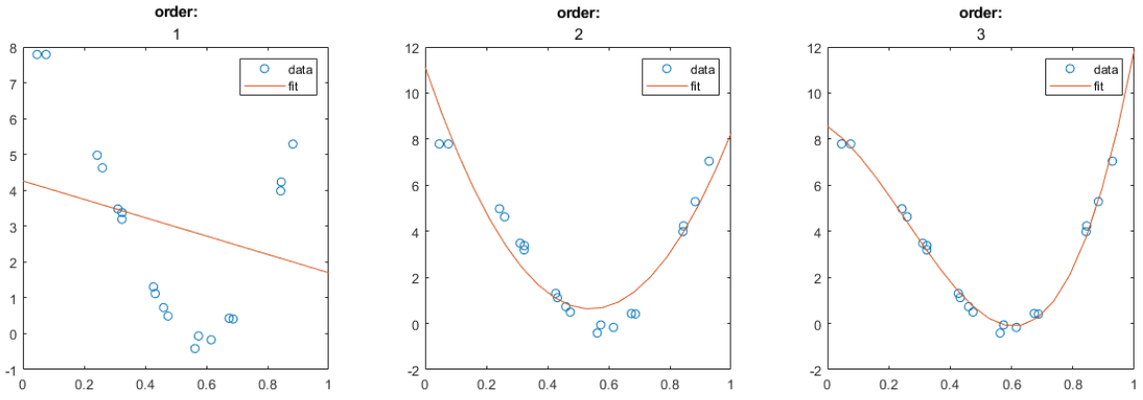
Name: 黃廷睿

ID: r11631026

Problem 1 (Regression).

可發現階數越高能越好的擬合原始資料，並且經計算殘差值的總和，非常接近 0，滿足 least squares assumptions 的條件



Order: 1

X	Y	Fit	FitError
0.4743	0.49301	3.0459	-2.5529
0.32424	3.3789	3.43	-0.051063
0.46004	0.7282	3.0824	-2.3542
0.67417	0.43161	2.5344	-2.1028
0.57406	-0.062789	2.7906	-2.8534
0.84283	3.9846	2.1028	1.8818
0.075822	7.7896	4.0657	3.7239
0.25985	4.6277	3.5947	1.0329
0.68756	0.40936	2.5002	-2.0908
0.4325	1.1195	3.1529	-2.0334
0.32373	3.1928	3.4313	-0.23852
0.046072	7.791	4.1418	3.6492
0.31035	3.4809	3.4655	0.015388
0.42662	1.3058	3.1679	-1.8621
0.61557	-0.16971	2.6844	-2.8541
0.24301	4.9794	3.6378	1.3416
0.84472	4.2343	2.098	2.1363
0.8828	5.2895	2.0005	3.289
0.92826	7.0419	1.8842	5.1577
0.56263	-0.41458	2.8199	-3.2345

Sum of error: 7.5495e-15

Order: 2

X	Y	Fit	FitError
0.4743	0.49301	0.78423	-0.29122
0.32424	3.3789	2.3051	1.0738
0.46004	0.7282	0.85901	-0.13081
0.67417	0.43161	1.2761	-0.84454
0.57406	-0.062789	0.67024	-0.73303
0.84283	3.9846	3.9284	0.056192
0.075822	7.7896	8.386	-0.59639
0.25985	4.6277	3.4547	1.1729
0.68756	0.40936	1.4118	-1.0025
0.4325	1.1195	1.045	0.074521
0.32373	3.1928	2.3131	0.8797
0.046072	7.791	9.4121	-1.6211
0.31035	3.4809	2.5278	0.95312
0.42662	1.3058	1.0917	0.21413
0.61557	-0.16971	0.83391	-1.0036
0.24301	4.9794	3.8045	1.1749
0.84472	4.2343	3.9698	0.26452
0.8828	5.2895	4.8572	0.43228
0.92826	7.0419	6.0533	0.98856
0.56263	-0.41458	0.64696	-1.0615

Sum of error: 7.5495e-15

Order: 3

X	Y	Fit	FitError
0.4743	0.49301	0.74446	-0.25145
0.32424	3.3789	3.1572	0.22174
0.46004	0.7282	0.92324	-0.19503
0.67417	0.43161	0.15936	0.27225
0.57406	-0.062789	-0.057477	-0.0053116
0.84283	3.9846	3.7395	0.24514
0.075822	7.7896	7.6773	0.11228
0.25985	4.6277	4.4028	0.22486
0.68756	0.40936	0.28338	0.12598
0.4325	1.1195	1.3054	-0.18592
0.32373	3.1928	3.1668	0.025951
0.046072	7.791	8.0715	-0.28049
0.31035	3.4809	3.4213	0.059612
0.42662	1.3058	1.3927	-0.086858
0.61557	-0.16971	-0.11029	-0.059423
0.24301	4.9794	4.7323	0.24715
0.84472	4.2343	3.8072	0.42708
0.8828	5.2895	5.3183	-0.028792
0.92826	7.0419	7.5072	-0.46528
0.56263	-0.41458	-0.011081	-0.4035

Sum of error: 7.2553e-14

Problem 2 (Variance of linear regression model)

$\because$ Ko $Y = Xw + \varepsilon$, $E[\hat{w}] = w$, $\hat{w} = (X^TX)^{-1}X^TY$, $Var(Z) = E\left[(Z - E[Z])(Z - E[Z])^T\right]$

$\Rightarrow Var(\hat{w}) = E\left[(\hat{w} - E[\hat{w}])(\hat{w} - E[\hat{w}])^T\right]$

$= E\left[(\hat{w} - E[\hat{w}])(\hat{w}^T - E[\hat{w}]^T)\right]$ ($E[\hat{w}] = w$, $E[\hat{w}]^T = w^T$)

$= E\left[\hat{w}w^T - \hat{w}w^T - w\hat{w}^T + ww^T\right]$

$= E\left[\hat{w}^T\hat{w} + ww^T - 2w\hat{w}^T\right]$

$= E\left[\hat{w}^T\hat{w}\right] + E\left[ww^T\right] - 2E\left[w\hat{w}^T\right]$

$= E\left[\hat{w}^T\hat{w}\right] + wE[\hat{w}^T] - 2wE[\hat{w}^T]$ ($E[\hat{w}^T] = w^T$)

$= E\left[\hat{w}^T\hat{w}\right] - ww^T$

$= E\left[\left((X^TX)^{-1}X^TY\right)^T \cdot \left((X^TX)^{-1}X^TY\right)\right] - ww^T$, $\hat{w} = (X^TX)^{-1}X^TY$

$= E\left[(X^TX)^{-1}X^TY \cdot (X^TX)^{-1}X^TY\right] - ww^T$

$= (X^TX)^{-1}E\left[XX^Tww^T + Xw\varepsilon + X^Tw^T\varepsilon + \varepsilon^2\right] - ww^T$, $Y = Xw + \varepsilon$,

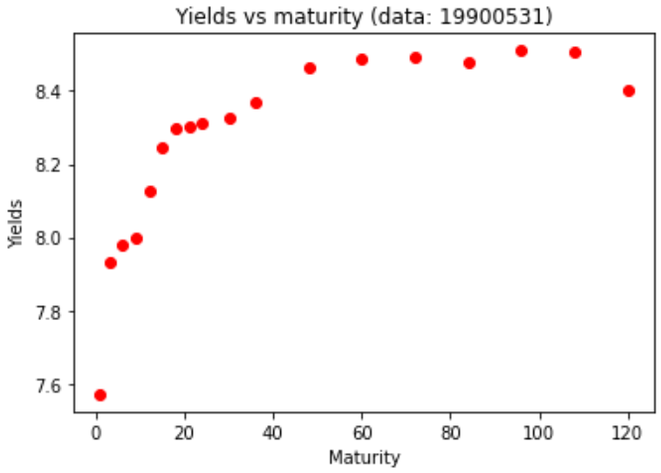
$= (X^TX)^{-1}(XX^T)(ww^T) + (X^TX)^{-1}\left[Xw + X^Tw^T E[\varepsilon] + E[\varepsilon^2] - ww^T\right]$ ($(X^TX)^{-1}(XX^T) = I$, $E[\varepsilon^2] = \sigma^2$)

$= (X^TX^{-1})\sigma^2$

#

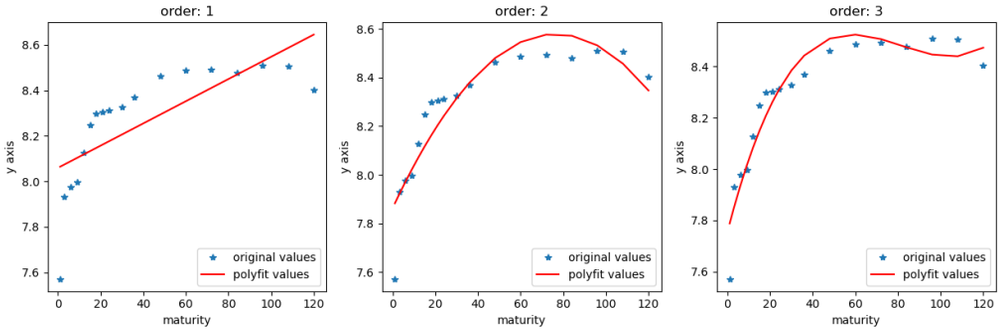
Problem 3 (Residue analysis)

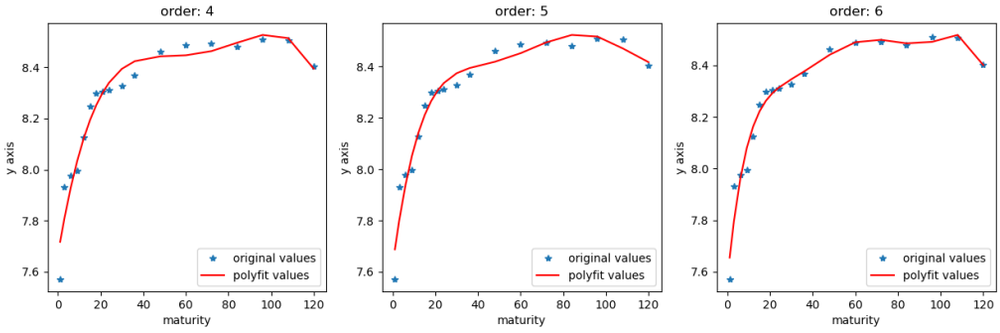
a. Generate a scatter plot of the yields vs maturity. Comment on the figures.

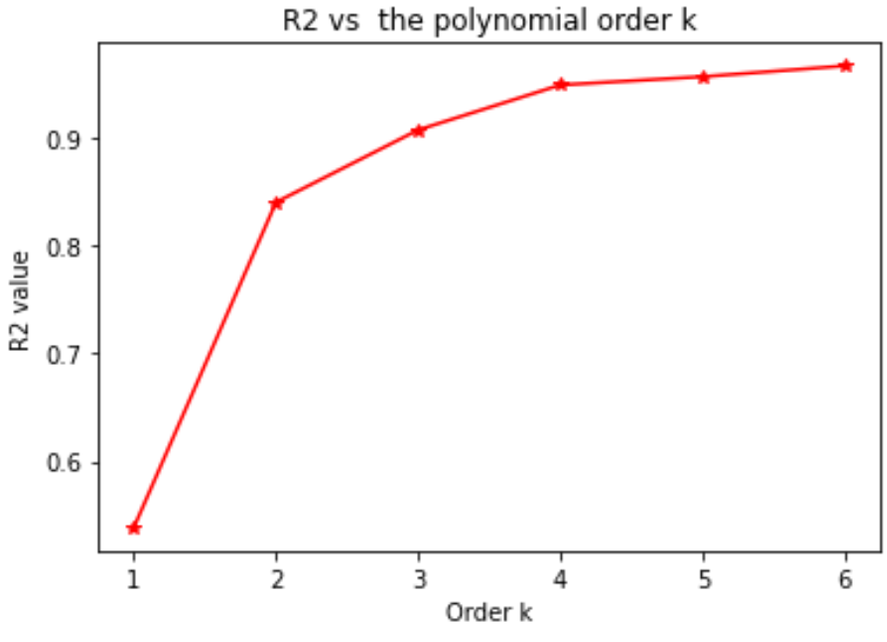


b. You want to fit a polynomial regression model to the data. Since you don't know the order you need, you fit six polynomial models, with orders from 1 to 6. Plot the $R2$ vs the polynomial order k. Comment on the result.

由下圖可以觀察不同階數的擬合情況，越高階數擬合的狀況就越佳，並且 R2 值與階數關係也滿足階數越高，R2 越高的趨勢，意即高階擬合結果較能解釋 yields 的分布狀況。

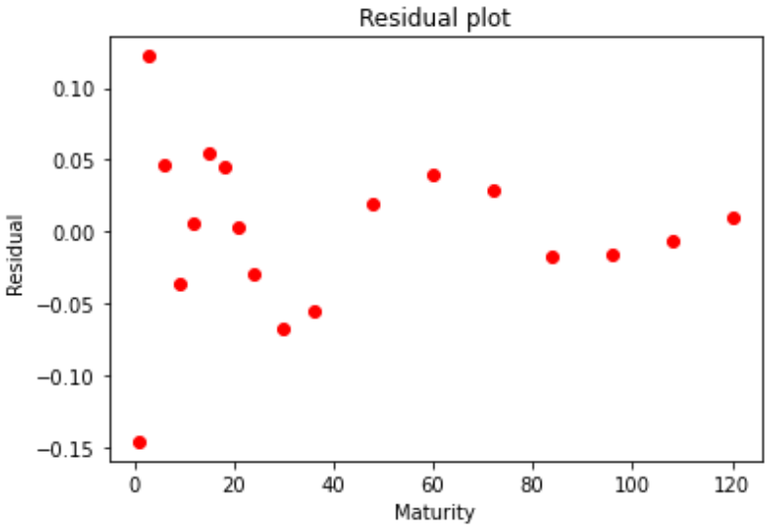

order: 1
order: 2
order: 3
y axis
maturity
original values
polyfit values


order: 4
order: 5
order: 6
y axis
maturity
original values
polyfit values

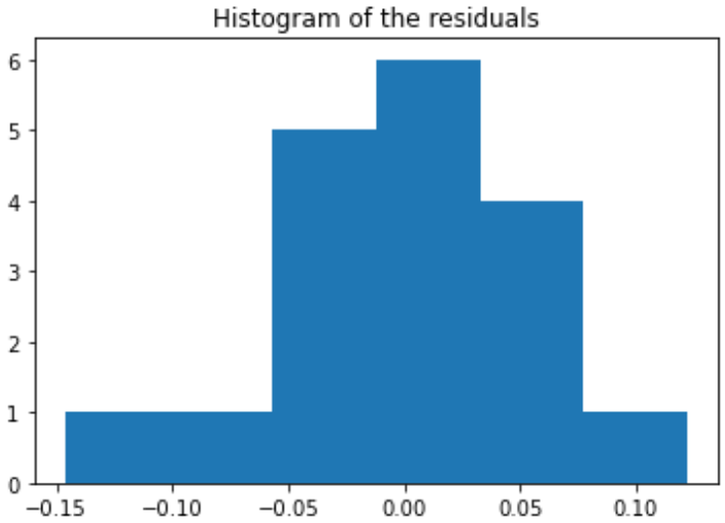

R2 vs the polynomial order k
R2 value
Order k

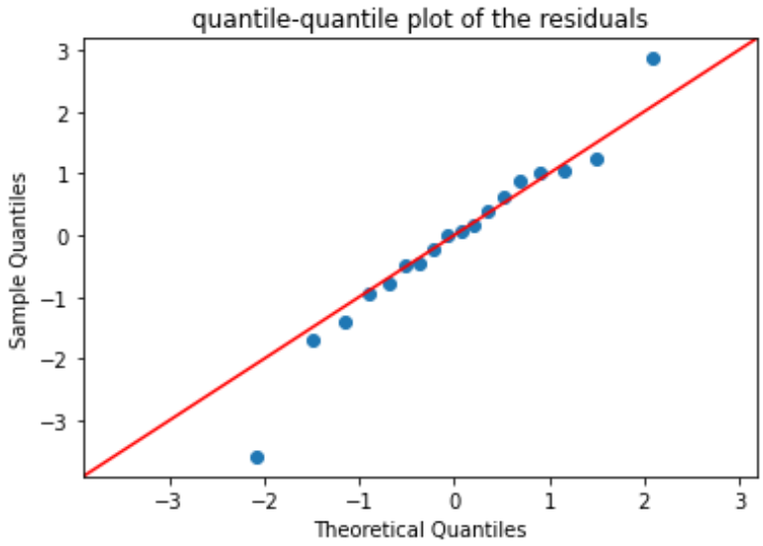
c. For the 4th-order polynomial model, draw a residual plot (vs maturity). Comment on it.

經計算 residual，可發現殘差值大致皆落在 0 上下，並且其均值大致為 0



d. Draw a histogram and a quantile-quantile (Q-Q) plot of the residuals. Comment on it.





透過 QQplot 趨勢可發現圖中點多數皆分散在對角線上，意即這筆資料大致為常態分佈的