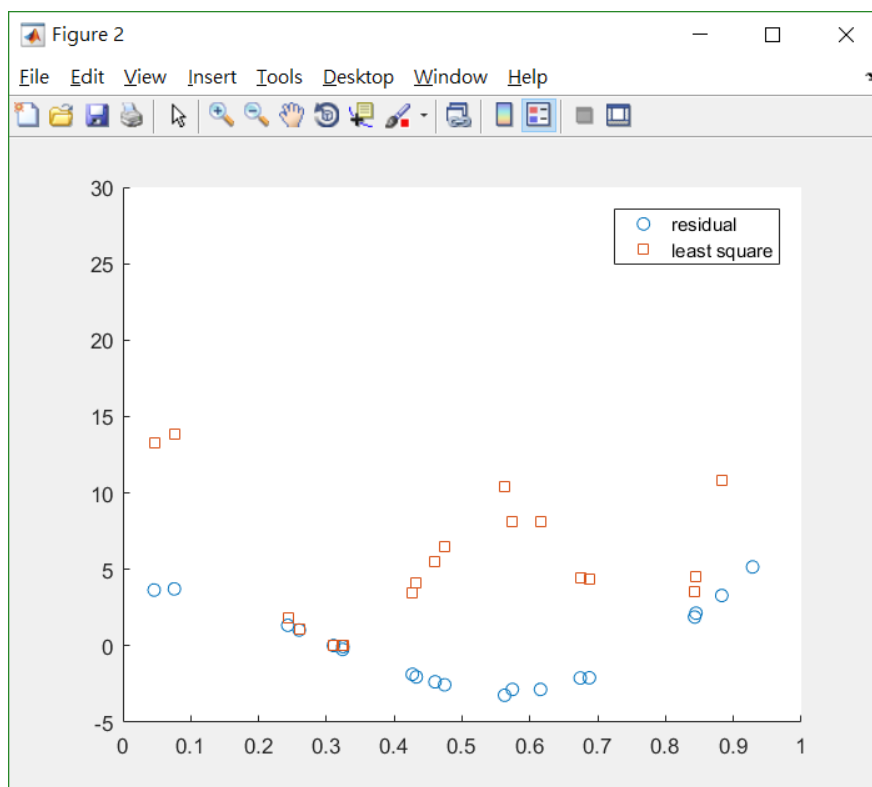
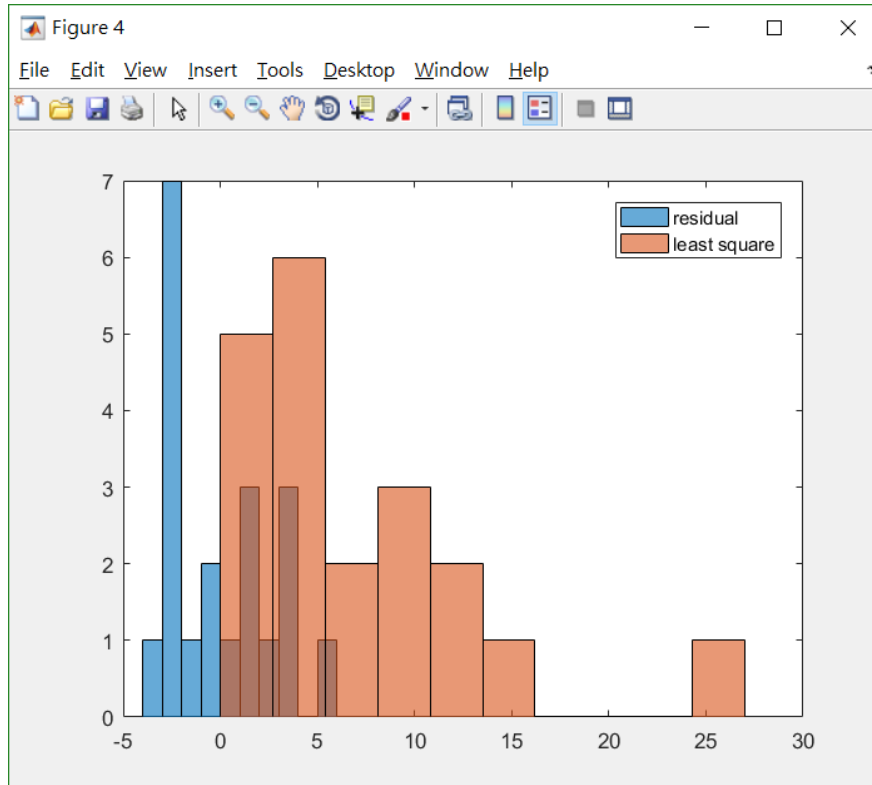
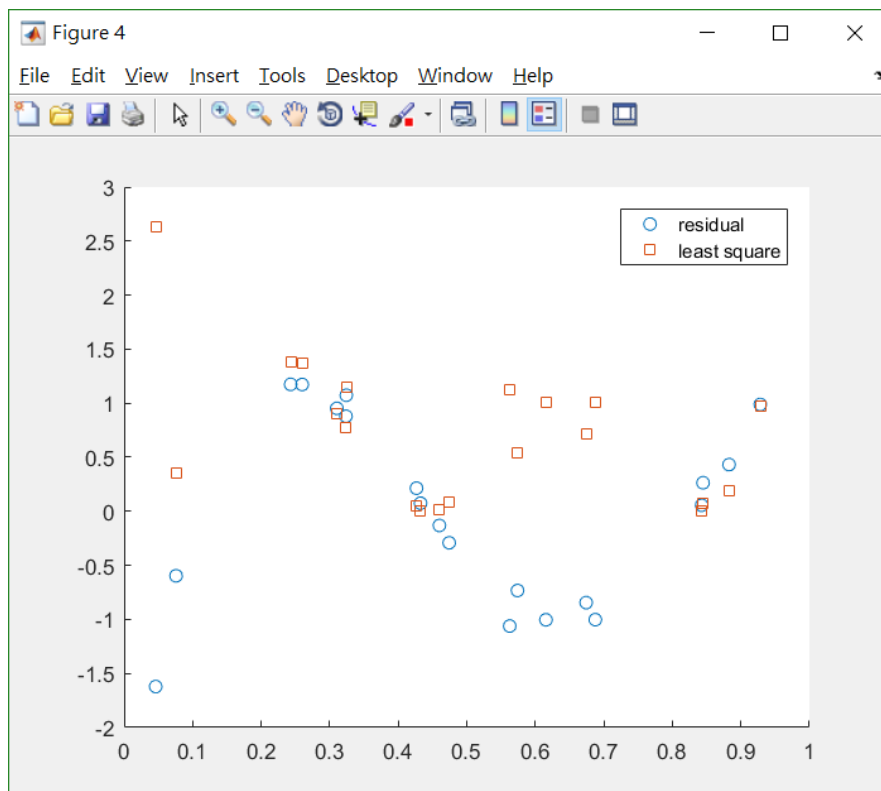
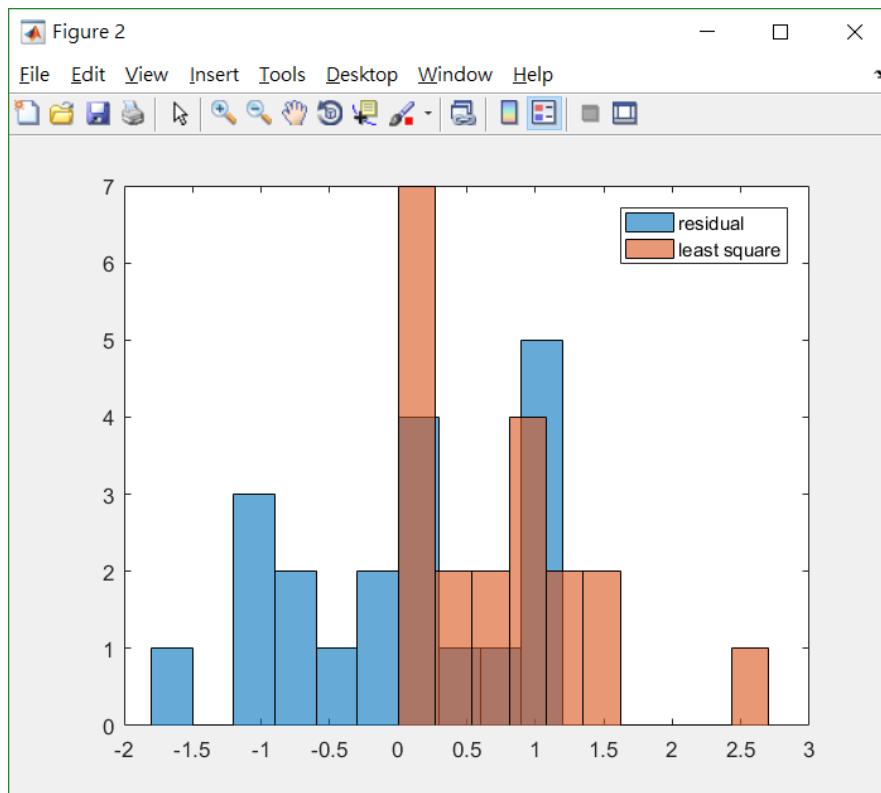


Problem 1 (Regression)

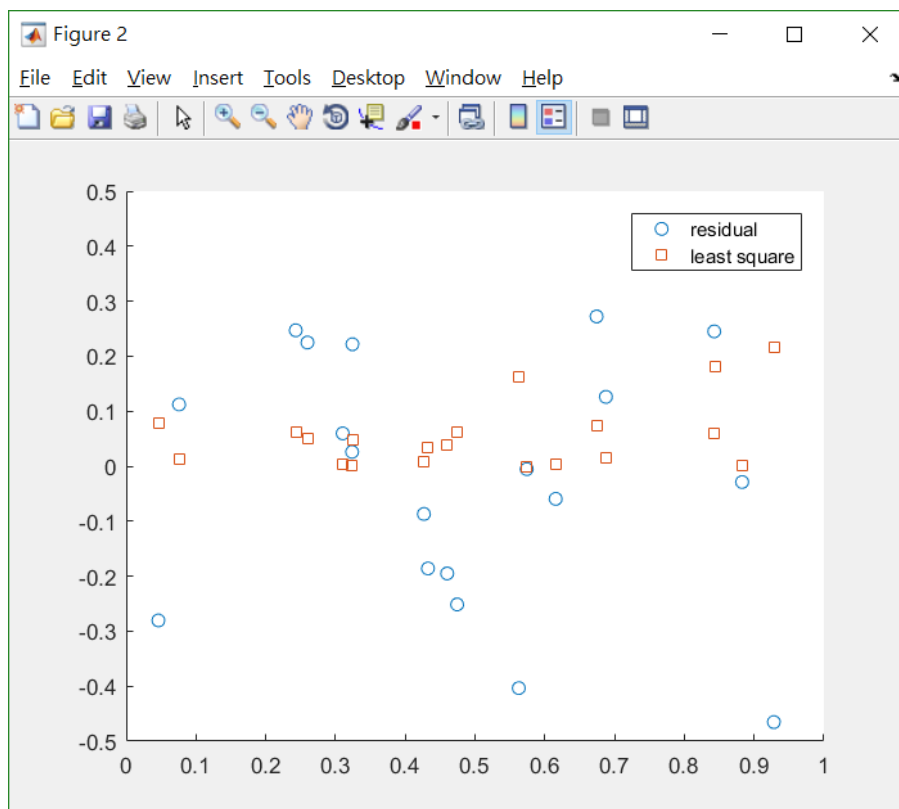
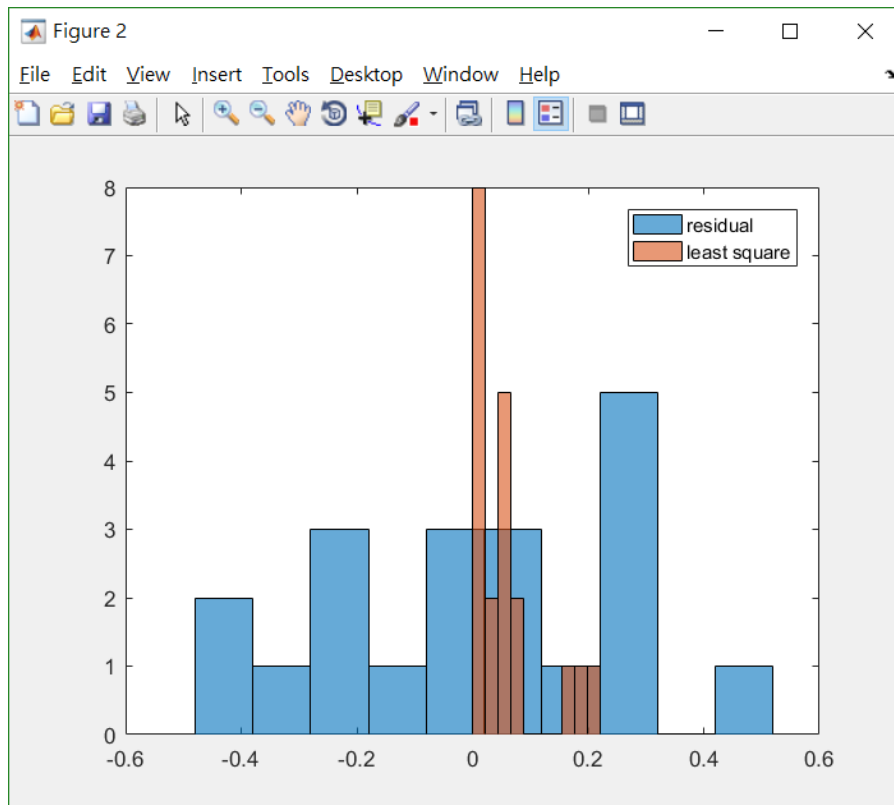
(a) 1st order \rightarrow training error = 6.5420



2nd order → training error = 0.7178



3rd order → training error = 0.05579

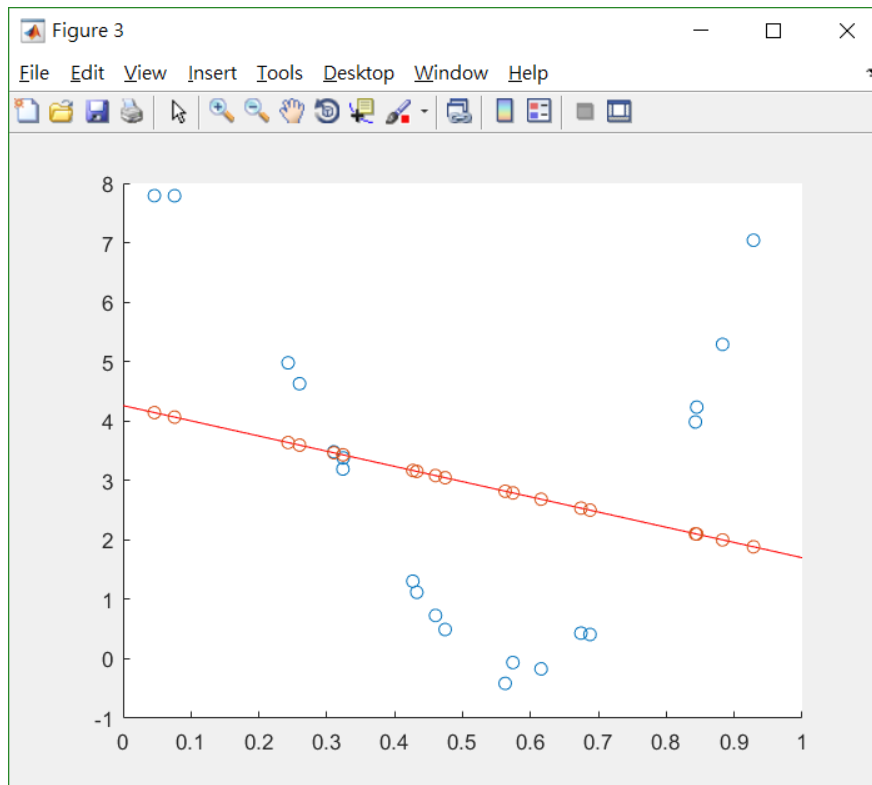


結果討論:

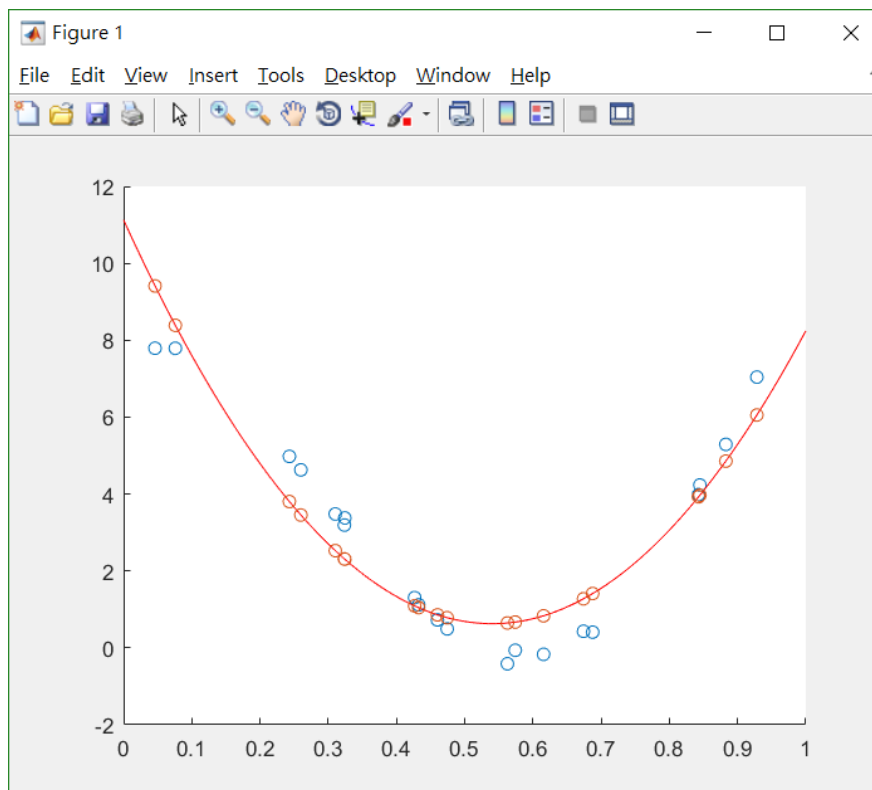
由上面三張直方圖和散佈圖，可看出 **residual** 和 **least square** 之間的關係，當 **order** 越大時，**residual** 和 **least square** 的數值範圍皆越來越小，此外，**least square** 的數值也有集中的趨勢。

(b)

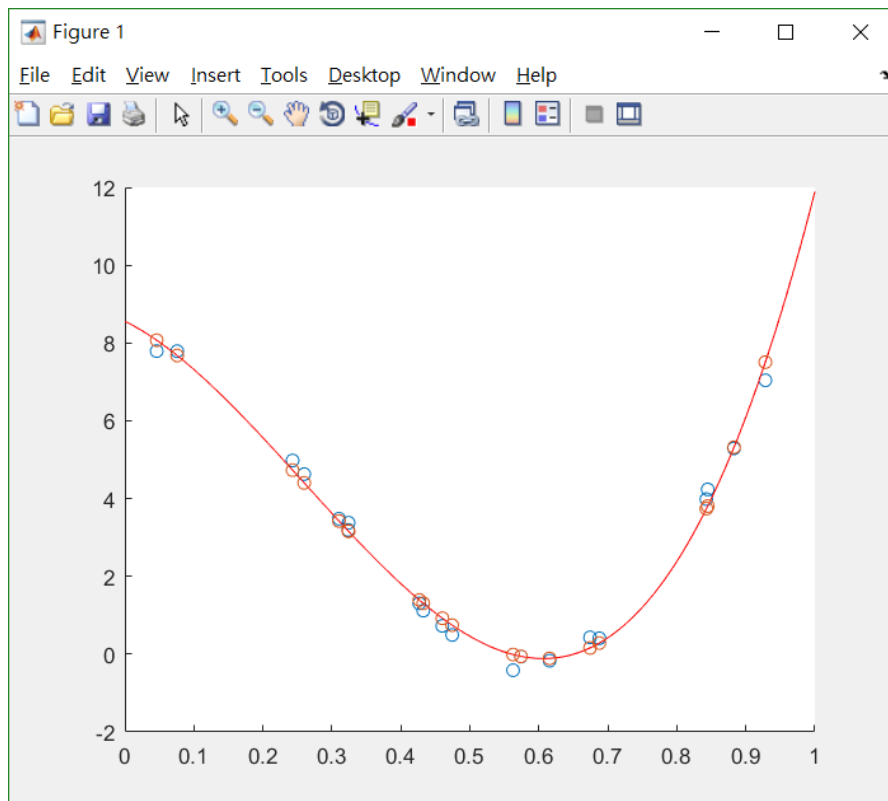
(i) 1st order



(ii) 2nd order



(iii) 3rd order

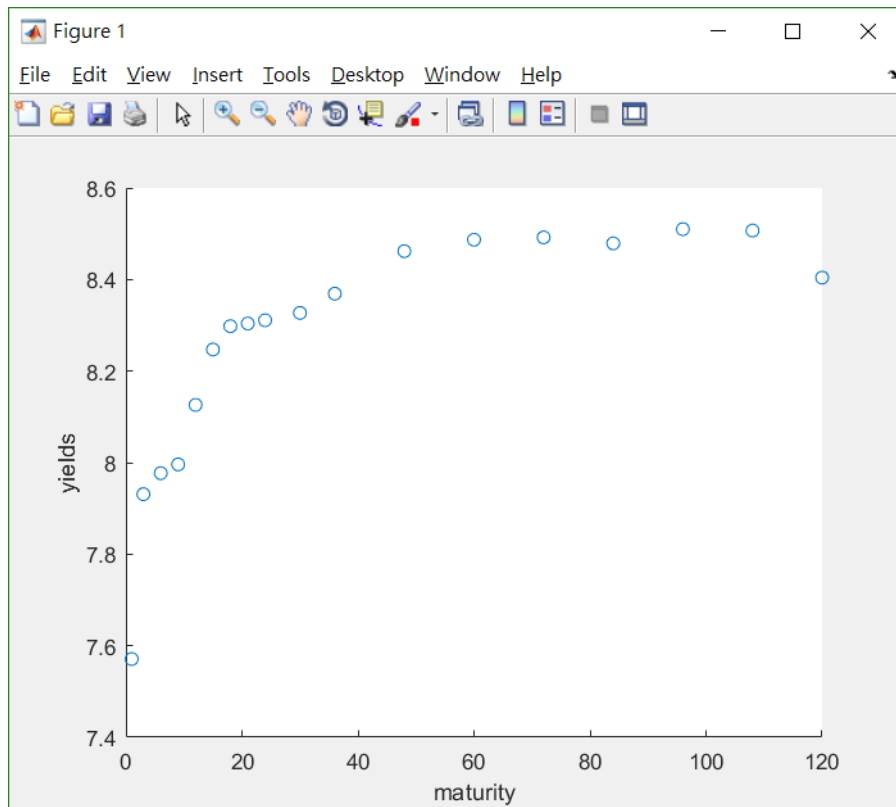


Problem 2 (Variance of linear regression model)

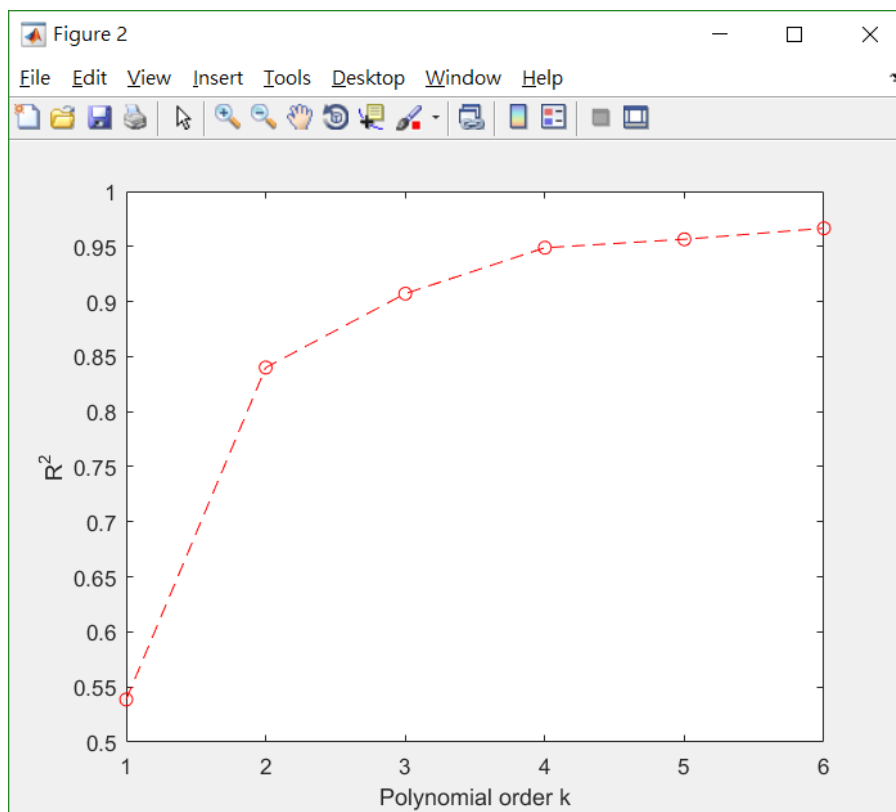
$$\begin{aligned} 2. \text{Var}(\hat{\omega}) &= E[(\hat{\omega} - E[\hat{\omega}])(\hat{\omega} - E[\hat{\omega}])^T] \\ &= E[(X'(X)')^{-1}X'\epsilon)(X'(X)')^{-1}X'\epsilon)^T] \\ &= E[(X'(X)')^{-1}X'\epsilon\epsilon'X(X')^{-1}] \\ &= E[(X'(X)')^{-1}X'\sigma^2X(X')^{-1}] \\ &= E[\sigma^2(X'(X)')^{-1}X'X(X')^{-1}] \\ &= \sigma^2(X'(X)')^{-1} \end{aligned}$$

Problem 3 (Residue analysis)

(a)



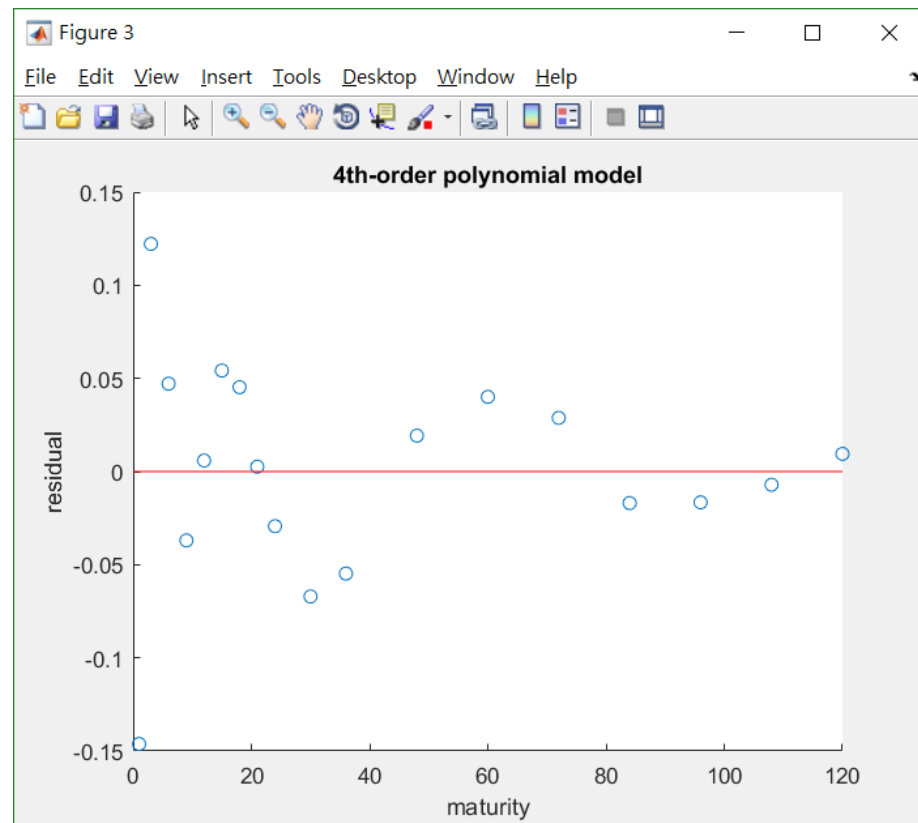
(b)



Result:

由上圖可以看出，在 polynomial order = 4 的時候，當 order 再增加， R^2 向上的幅度也並不大，因此 order 取 4 較為合適。

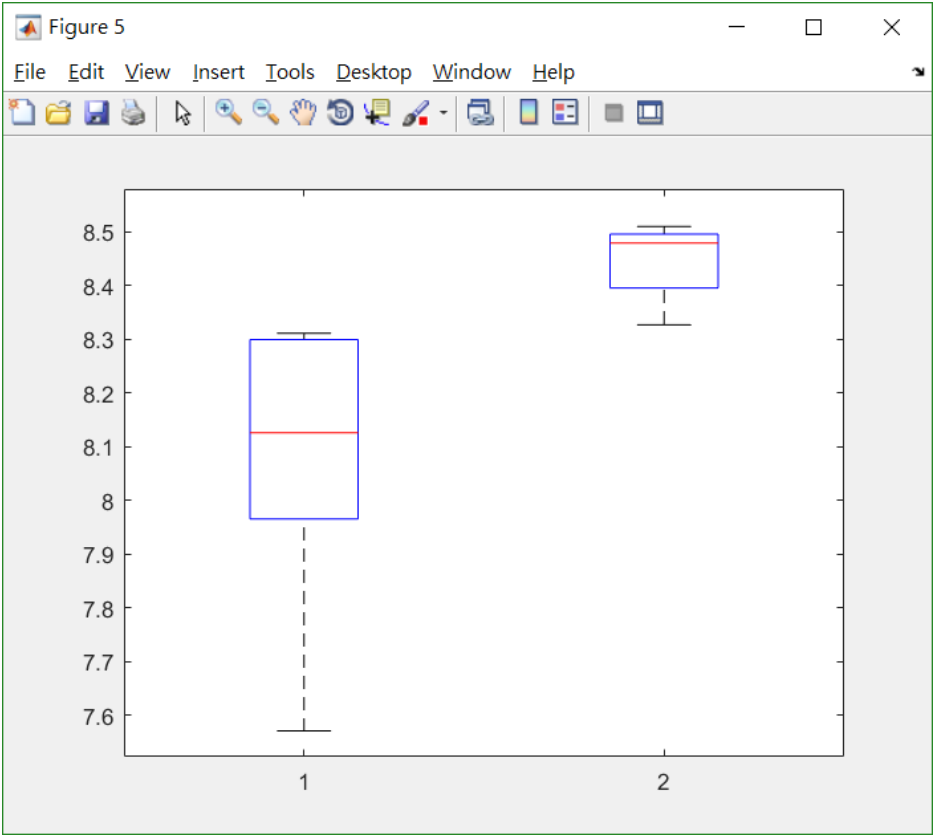
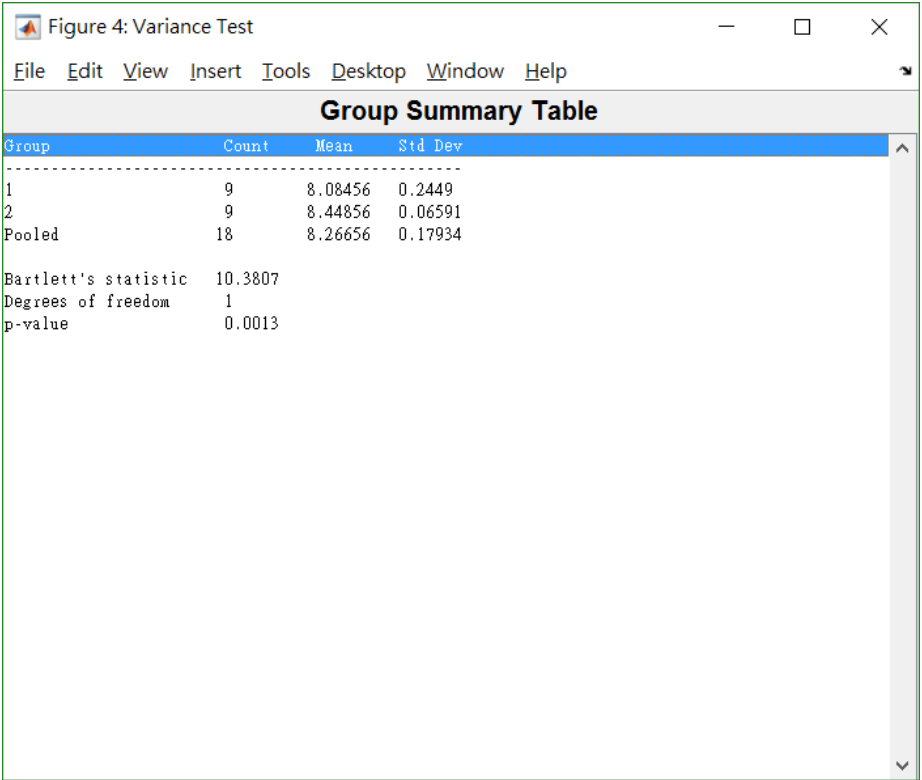
(c)



Result:

可以看出資料一開始的 residual 範圍較廣，到後面時則縮小。

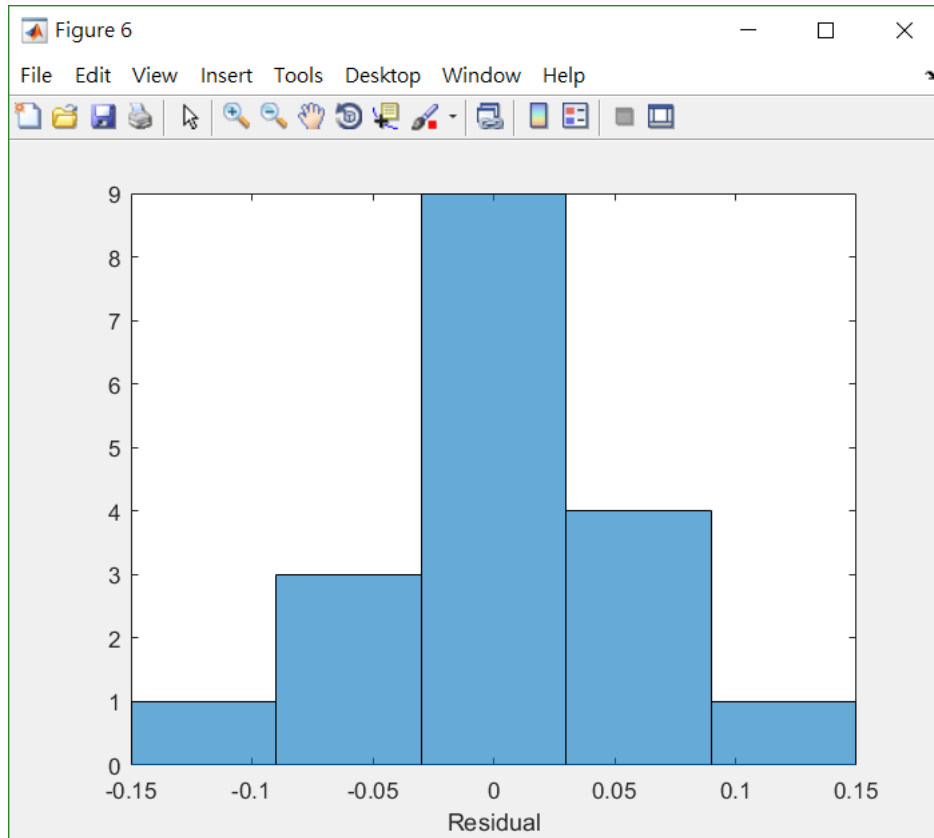
(d)

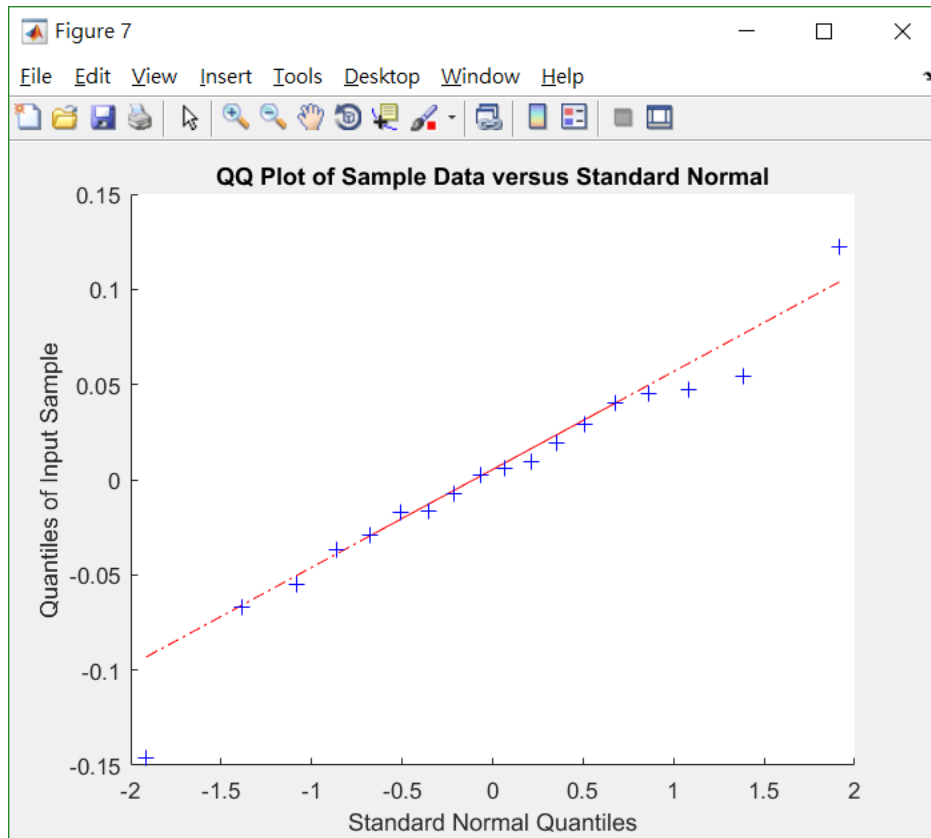


Result:

藉由 **Levene test**，我們可以來判別變異數的同質性，我便可以看到前半段的資料變異數的差距較大，而後半段的較小，而由(c)的 **residual plot**，也可看出 **residual** 的分布，於前半段分布較廣，後半段分布較窄。

(e)





Result:

由直方圖可看出 **residual** 還是較集中於 0 附近，且呈現常態分佈的趨勢。

QQ 圖可知道收集到的數據是否來自常態分佈的母體，如此樣本是來自常態分佈的母體，因此其 QQ 圖近似於一條直線。