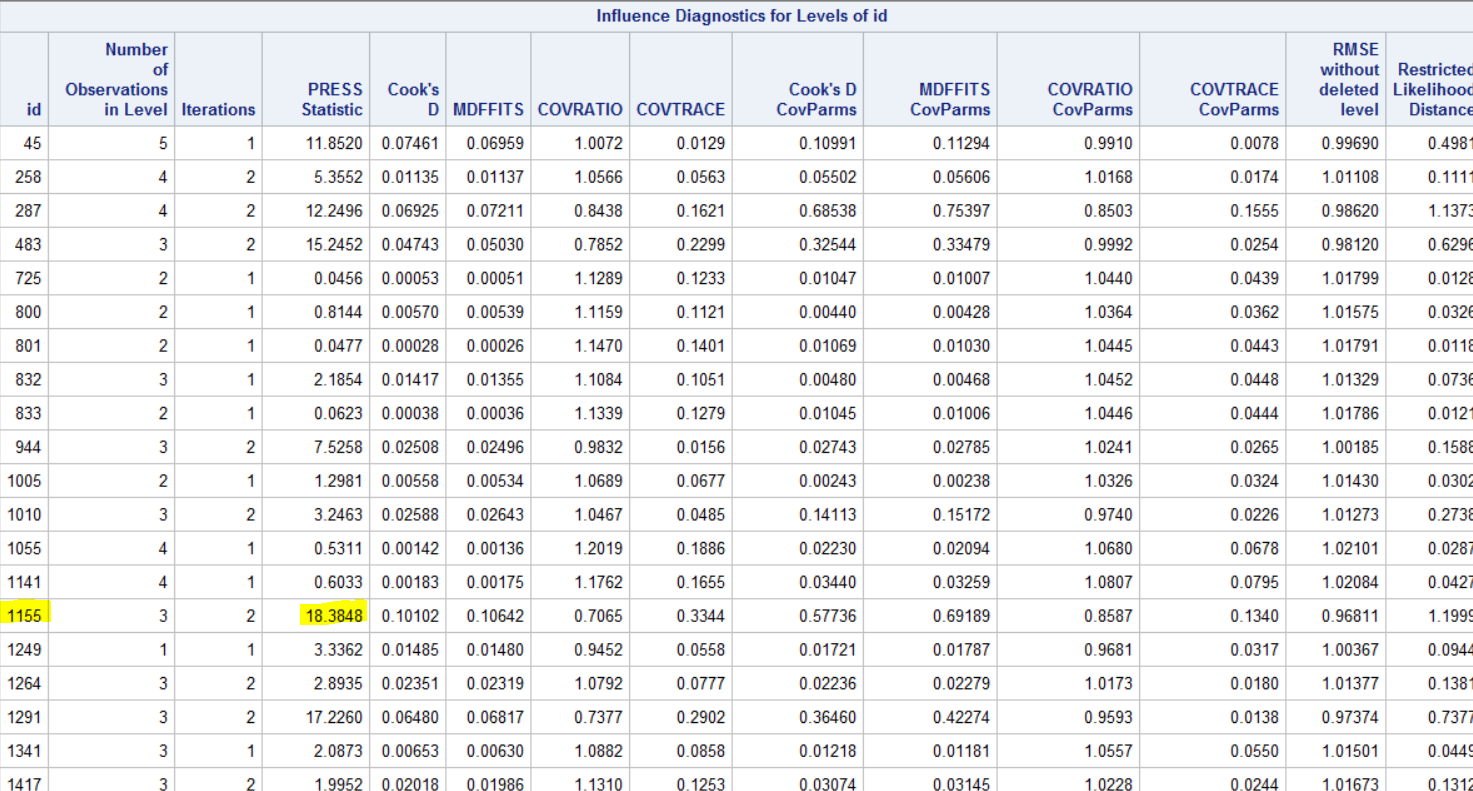
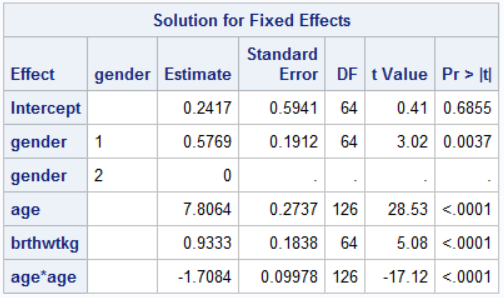
**Homework (10 points):** 課程網頁上有一檔案growth.csv，若幼兒的體重 (weight,單位：公斤) 可能受到出生時的體重 (brthwtkg,單位：公斤)、性別 (gender, 1:male; 2:female，以female作為基準組)、年紀 (age,單位：年)、年紀的平方 (age\*age) 的影響，以SAS proc mixed分析之，請配適R-side analysis --- ar(1) covariance matrix，請依本講義各步驟進行模式診斷：

1. (3 points) 請問PRESS最高的個案為？請自行計算出該個案的PRESS，並與SAS output做比對。



Id = **1155**



使用計算機計算：

0.2417 + 0.5769 \* 1 + 7.8064 \* 0.125941142 + 0.9333 \* 3.65 - 1.7084 \* ((0.125941142) ^ 2) = 5.1811947059482655022224

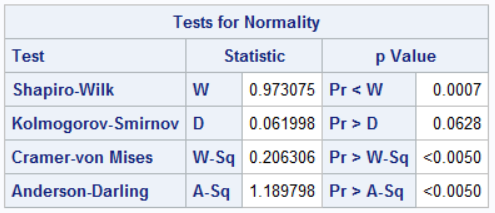
0.2417 + 0.5769 \* 1 + 7.8064 \* 1.073237538 + 0.9333 \* 3.65 - 1.7084 \* ((1.073237538) ^ 2) = 10.6354650885613202130704

0.2417 + 0.5769 \* 1 + 7.8064 \* 2.299794674 + 0.9333 \* 3.65 - 1.7084 \* ((2.299794674) ^ 2) = 13.1424396542062036940816

(5.190000057 - 5.1811947059482655022224) ^ 2 + (12.02000046 - 10.6354650885613202130704) ^ 2 + (17.20000076 - 13.1424396542062036940816) ^ 2 = 18.38

與SAS output相同。

1. (2 points) 殘差是否呈常態分布？如果違反常態，請將weight取「自然對數」(natural log) 轉換後重新配適模型。

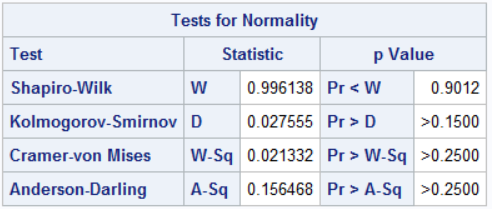


H0: 誤差呈常態分布

H1: 誤差不呈常態分布

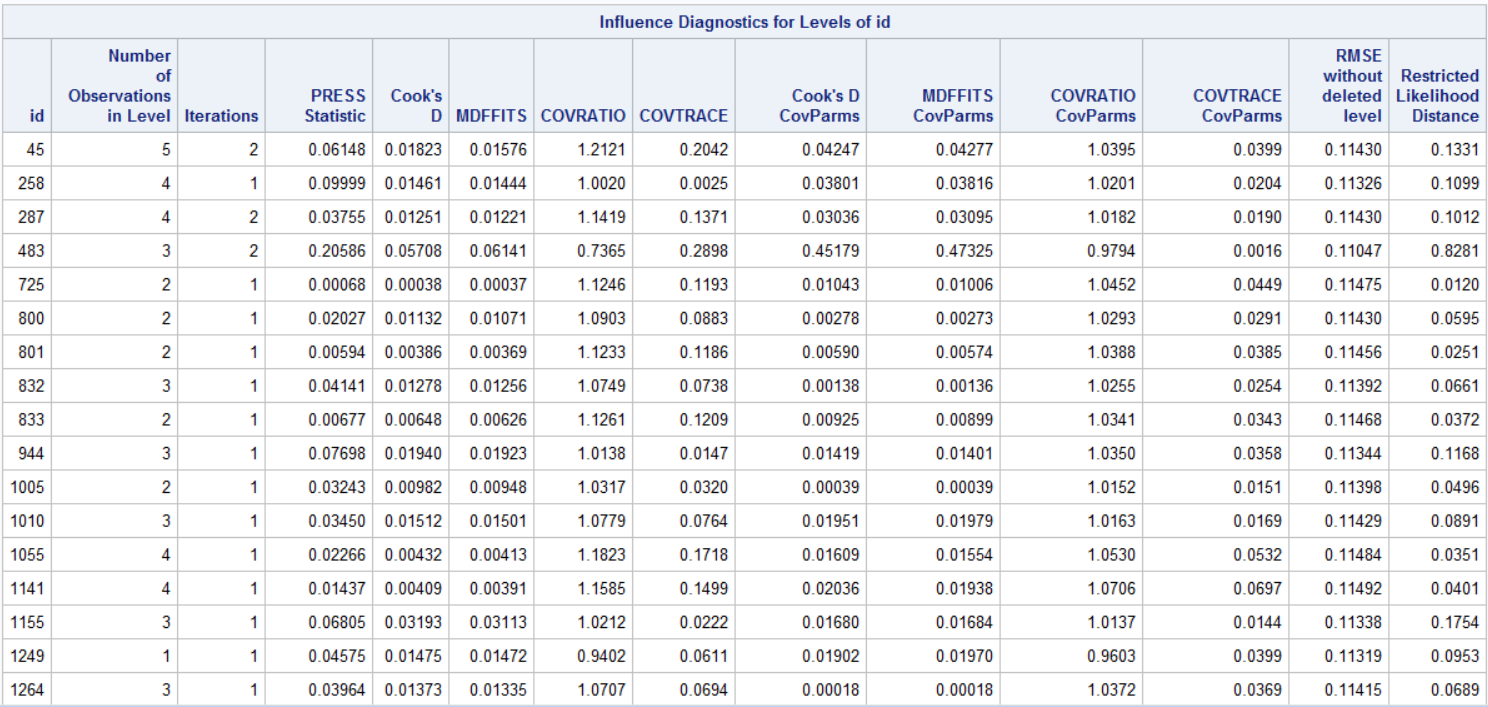
樣本數為198，落在S-W test的限制範圍,做誤差是否為常態分佈的檢定：shaprio-Wilk test 中p-value=0.0007<0.05，拒絕虛無假設，誤差不為常態分佈。因此將weight取自然對數的轉換，再重新適配模型。

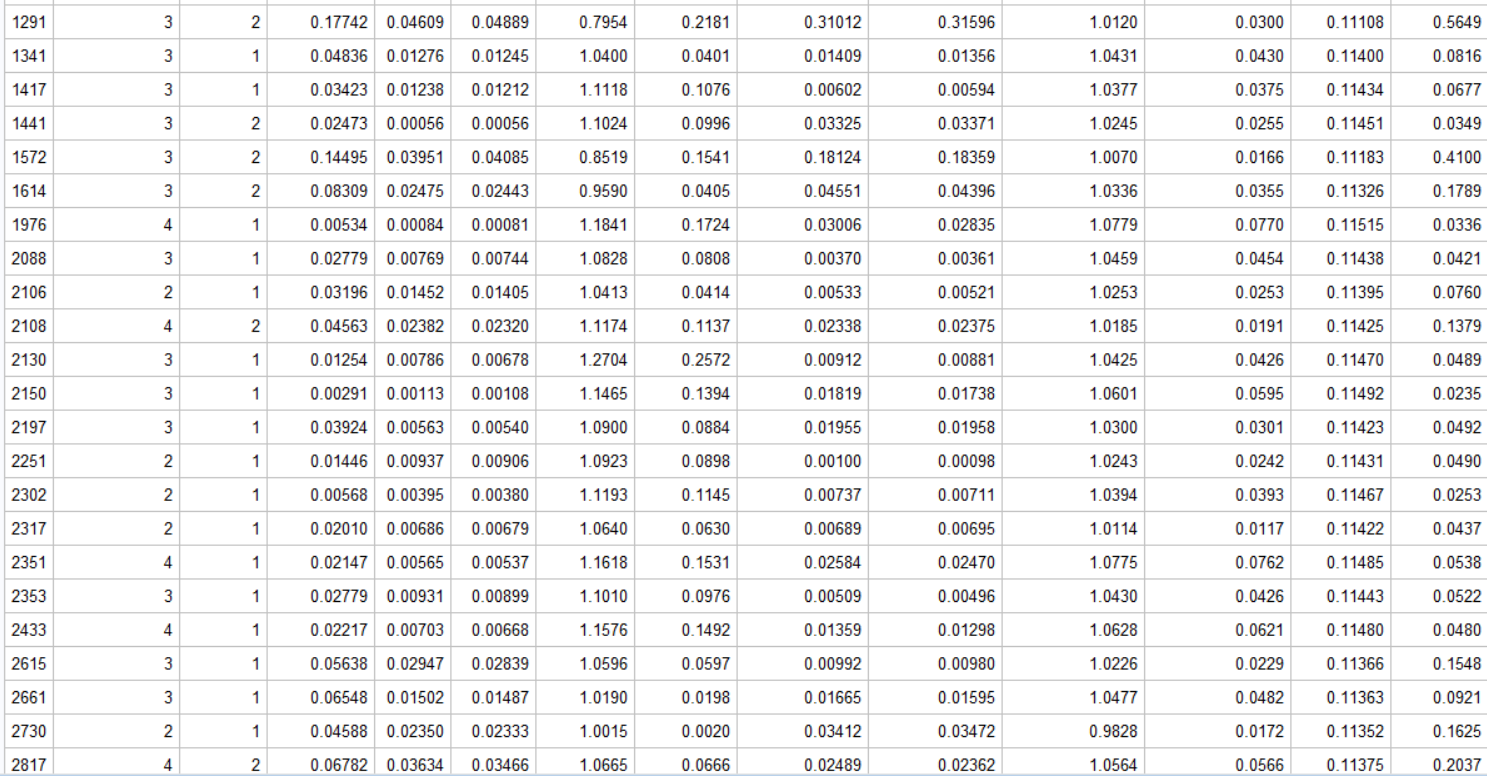
1. (1 points) 重新配適模型後殘差是否呈常態分布？

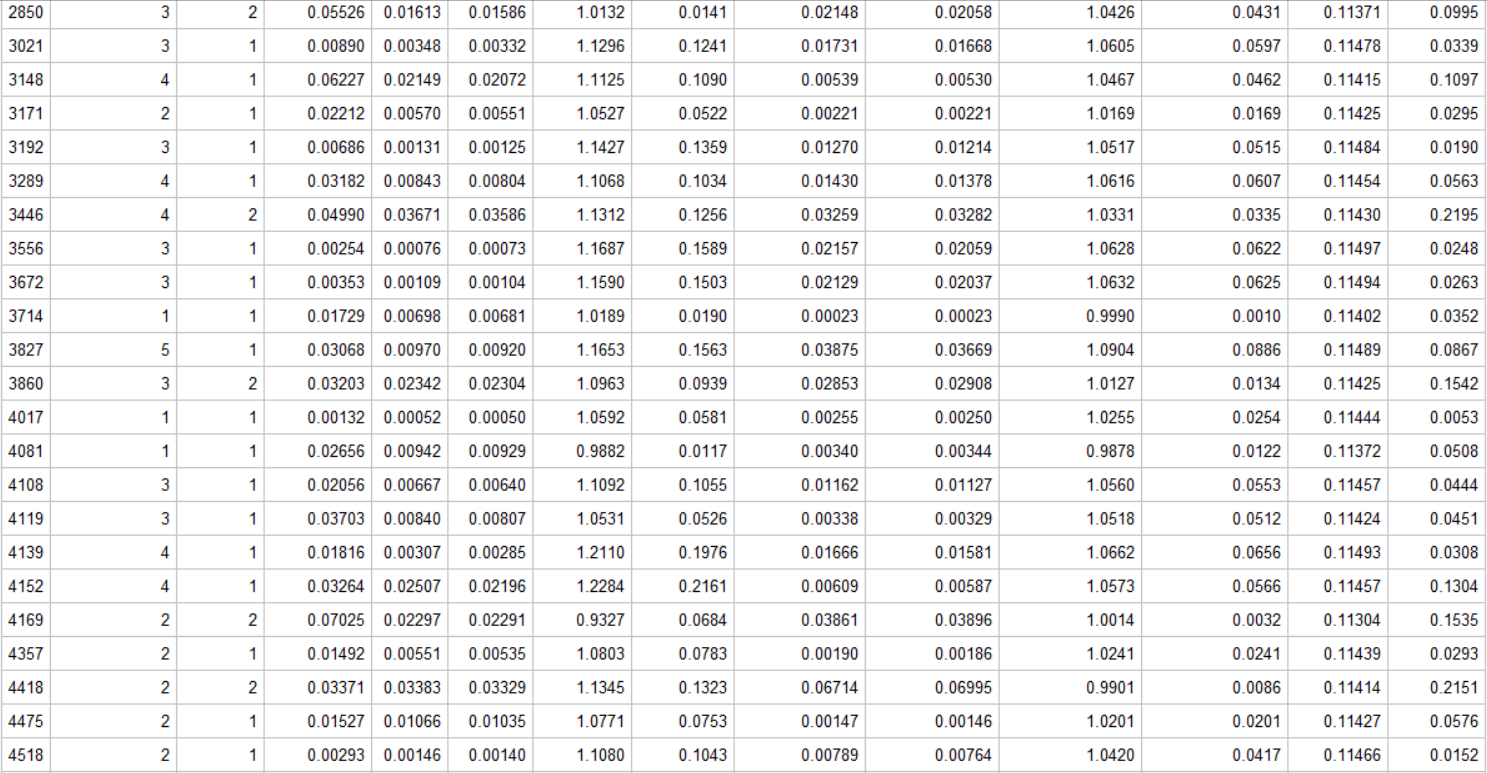


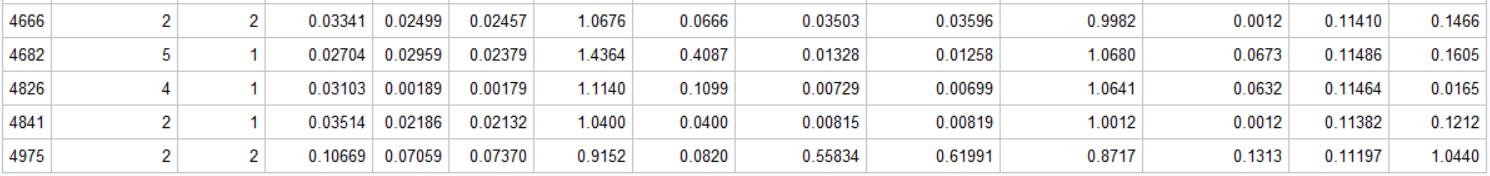
重新配適後，S-W test, K-S test p-value皆>0.05，不拒絕虛無假設，殘差呈常態分布。

1. (4 points) 請依本講義各步驟進行高影響力個案分析，是否發現任何可疑個案？





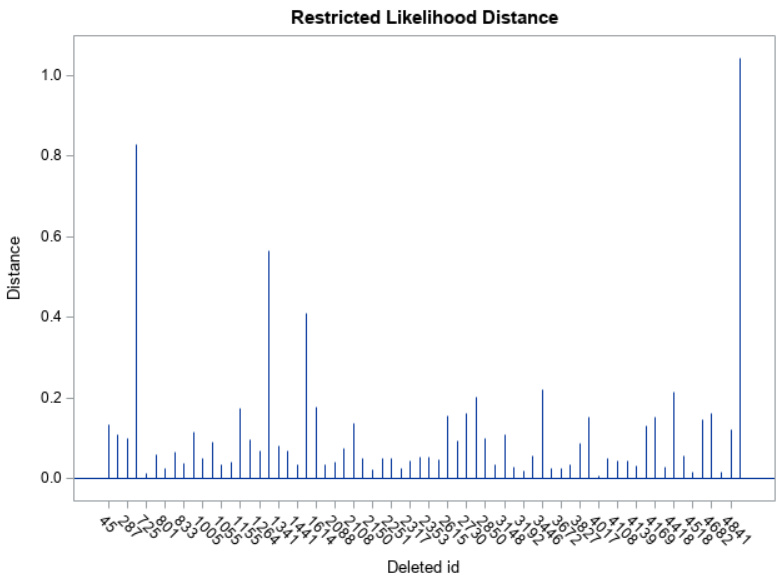




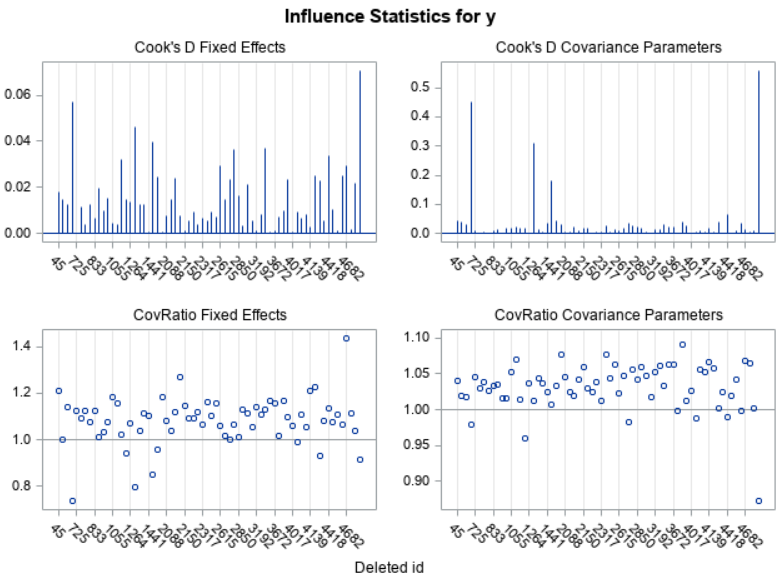
用Cook’s D來看對regression coefficient estimates哪個ID為高影響力個案，與 qf( 0.5 , 5 , 198 - 5 ) = 0.8733461比較後，並未發現有大於其值的個案。接著用Cook’s D CovParms來看對Covariance parameters哪個ID為高影響力個案，與qf( 0.5 , 2 , 198 - 2 ) = 0.6956403比較後，也沒有發現有大於其值的個案。最接近的 id 為 4975，其 Cook’s D CovParms 值為 0.55834。

觀察MDFFITS、MDFFITS CovParms 與 PRESS，若其值越大則越有可能是高影響力個案。MDFFITS中最大的有ID為4975、483、1291、1572。MDFFITS CovParms中最大的有ID為4975、483、1291、1572。PRESS中最大的有ID為483、1572、1291、4975。

接著看 CovRatio 和 CovTrace。其中 CovRatio 越接近 0 代表其id 為高影響力個案，檢查CovRatio 和 CovRatioCovParms 發現沒有 id 明顯接近 0；另外 CovTrace 越大代表其 id 為高影響力個案，檢查 CovTrace 和 CovTraceCovParms 發現沒有 id 明顯大於其他。



用Likelihood distances 同時看 fixed-effect parameters 和covariance parameters，若LDi > qchisq(0.95,7) = 14.06714，則第i個 id 為高影響力個案。可以發現上圖沒有id 超過此 threshold。但仍有四個較其他高的ID(483、1572、1291、4975)



綜合各項指標及模式診斷圖形，可疑個案為483、1572、1291、4975。

SAS code:

/\*1\*/

**proc** **mixed** data=growth cl alpha=**0.05** covtest;

class id Gender (ref="2");

model weight=gender age brthwtkg age\*age / s residual

influence(effect=id iter=**10**) outp=growthout;

repeated / type=ar(**1**) subject=id R rcorr;

**run**;

**proc** **mixed** data=growth1155 cl alpha=**0.05** covtest;

class id Gender (ref="2");

model weight=gender age brthwtkg age\*age / s residual

influence(effect=id iter=**10**) outp=growthout1155;

repeated / type=ar(**1**) subject=id R rcorr;

**run**;

/\*2\*/

**proc** **univariate** data=growthout normal;

var Resid StudentResid PearsonResid;

**run**;

/\*3\*/

/\*自然對數轉換\*/

**data** growth;

set growth;

y=log(weight);

**run**;

**proc** **mixed** data=growth cl alpha=**0.05** covtest;

class id Gender (ref="2");

model y=Gender Age brthwtkg Age\*Age / s residual influence(effect=id

iter=**10**) outp=growthout1;

repeated / type=ar(**1**) subject=id R rcorr;

**run**;

**proc** **univariate** data=growthout1 normal;

var Resid StudentResid PearsonResid;

**run**;