

R Notebook

Code ▾

Q3 題目

請根據上課所提到的 **kid_iq dataset**，建構一迴歸模型探索變數的非線性關係

(a)請探討 mom.iq 與依變數的 U 型關係，寫下迴歸的模型，並利用上課的 R 套件執行迴歸分析

(b)請說明你會如何檢驗 U 型關係的步驟，並搭配 R 套件驗證 U 型關係是否成立

(c)如若要探討媽媽高中學歷是否會影響 mom.iq 對於依變數的關係，請說明你會如何修改上述的迴歸模型，以及媽媽高中學歷會如何影響上述的 U 型關係

安裝套件以及載入資料

Hide

```
# 載入套件
library(tidyverse)
library(ggplot2)
```

Hide

```
# 載入 `kid_iq` 資料集
kid_iq <- read.csv("kid_iq.csv")
head(kid_iq)
```

	kid.score <int>	mom.hs <int>	mom.iq <dbl>	mom.work <int>	mom.age <int>
1	65	1	121.11753	4	27
2	98	1	89.36188	4	25
3	85	1	115.44316	4	27
4	83	1	99.44964	3	25
5	115	1	92.74571	4	27
6	98	0	107.90184	1	18

6 rows

建立二次迴歸模型

建構迴歸模型探討 mom.iq 與依變數 kid_score 的 U 型關係，建立二次迴歸模型：

$$y = a + b_1 \cdot mom.iq + b_2 \cdot mom.iq^2$$

Hide

```
# 建構二次項的迴歸模型
model_u <- lm(kid.score ~ mom.iq + I(mom.iq^2), data = kid_iq)
summary(model_u)
```

```
Call:
lm(formula = kid.score ~ mom.iq + I(mom.iq^2), data = kid_iq)

Residuals:
    Min       1Q   Median       3Q      Max
-54.824 -11.640   2.883  11.372  50.813

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -99.033675   37.301385  -2.655 0.008226 **
mom.iq        3.076800    0.730291   4.213 3.07e-05 ***
I(mom.iq^2)   -0.011917    0.003517  -3.389 0.000767 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.05 on 431 degrees of freedom
Multiple R-squared:  0.2217,    Adjusted R-squared:  0.2181
F-statistic: 61.38 on 2 and 431 DF,  p-value: < 2.2e-16
```

驗證 U 型關係是否成立

分析 Coefficients

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -99.033675   37.301385  -2.655 0.008226 **
mom.iq        3.076800    0.730291   4.213 3.07e-05 ***
I(mom.iq^2)   -0.011917    0.003517  -3.389 0.000767 ***
```

分析：

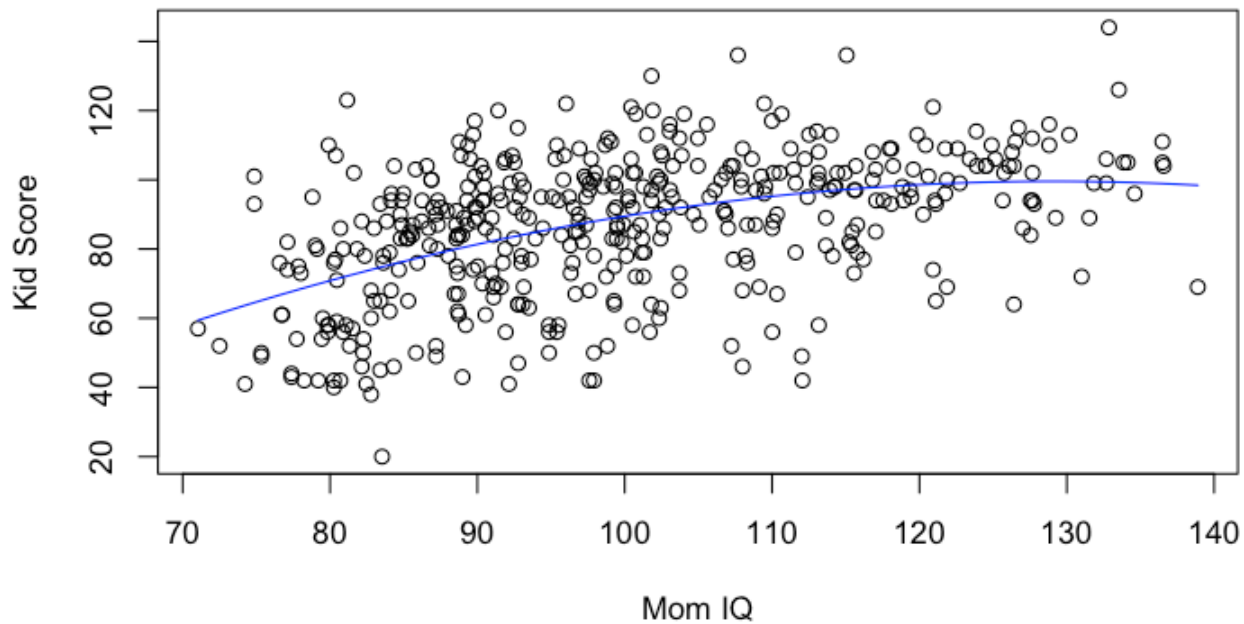
1. t-value 中顯示 mom.iq² 與 mom.iq 具有統計意義顯著性。
2. 二次項係數 (mom.iq²) 為負，顯示存在倒 U 關係
3. mom.iq 對 kid.score 有顯著的線性和非線性影響

視覺化

[Hide](#)

```
# 繪製 mom.iq 與 kid.score 的散點圖與迴歸曲線
plot(kid_iq$mom.iq, kid_iq$kid.score, main = "U-shaped Relationship",
     xlab = "Mom IQ", ylab = "Kid Score")
curve(predict(model_u, newdata = data.frame(mom.iq = x)), add = TRUE, col = "blue")
```

U-shaped Relationship



一次微分視覺化

[Hide](#)

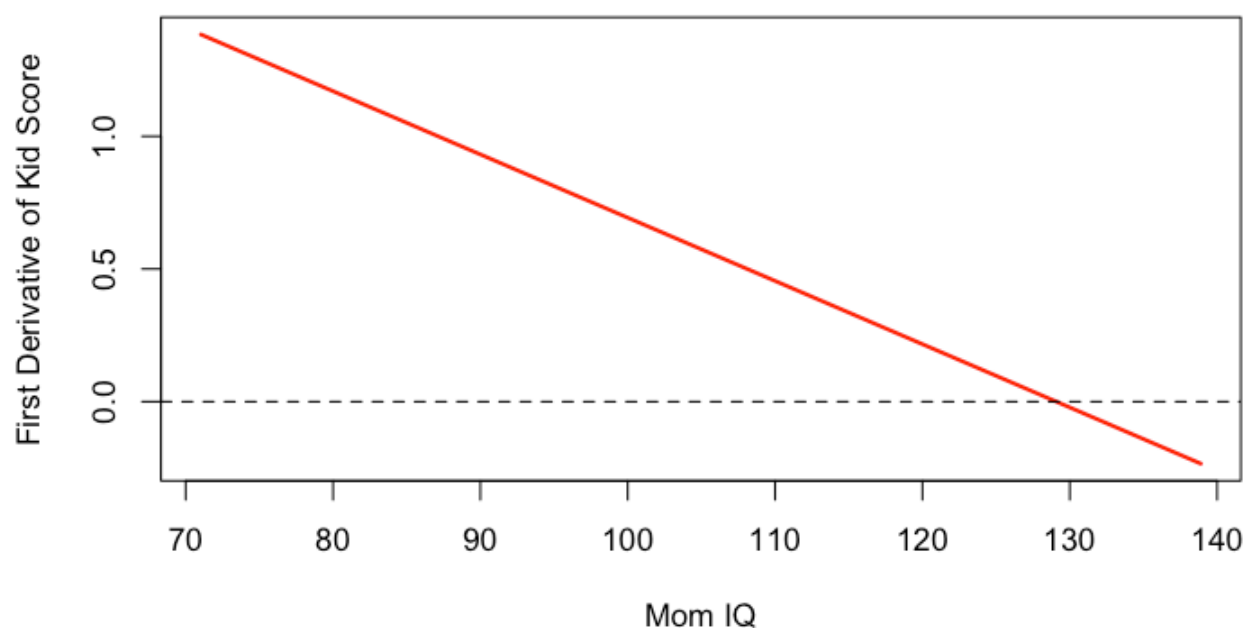
```
# 提取模型係數
b1 <- coef(model_u)["mom.iq"]
b2 <- coef(model_u)["I(mom.iq^2)"]

# 定義一階導數的函數
first_derivative <- function(x) {
  b1 + 2 * b2 * x
}

# 繪製一階導數曲線（紅色線）
curve(first_derivative(x), from = min(kid_iq$mom.iq), to = max(kid_iq$mom.iq),
      main = "First Derivative of the Model",
      xlab = "Mom IQ", ylab = "First Derivative of Kid Score", col = "red", lty = 1, lwd = 2)

# 在 y 軸畫出 0 的水平線
abline(h = 0, col = "black", lty = 2)
```

First Derivative of the Model



從模型的散布圖以及一階微分的視覺化來看，可以確定該模型是倒 U 型關係。

小結

係數的顯著性以及視覺化的結果，證實了 mom.iq 與 kid.score 之間的倒 U 型關係在統計上顯著。

加入媽媽高中學歷

建立迴歸模型

將媽媽的高中學歷（mom.hs）加入模型，並與 mom.iq 和二次項交互作用，來探討高中學歷是否影響 U 型關係：

$$y = a + b1 \cdot mom.iq + b2 \cdot mom.iq^2 + b3 \cdot mom.hs + b4 \cdot (mom.iq \times mom.hs) + b5 \cdot (mom.iq^2 \times mom.hs)$$

[Hide](#)

```
# mom.iq:mom.hs 和 I(mom.iq^2):mom.hs 表示 mom.hs 的交互作用項，  
# 用來檢驗高中學歷是否改變 mom.iq 和二次項對依變數的影響。  
  
model_interaction <- lm(kid.score ~ mom.iq + I(mom.iq^2) + mom.hs + mom.iq:mom.hs + I(mom.iq^  
2):mom.hs, data = kid_iq)  
  
summary(model_interaction)
```

Call:

```
lm(formula = kid.score ~ mom.iq + I(mom.iq^2) + mom.hs + mom.iq:mom.hs +
    I(mom.iq^2):mom.hs, data = kid_iq)
```

Residuals:

Min	1Q	Median	3Q	Max
-53.466	-10.085	2.655	11.352	46.517

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.237e+02	1.000e+02	-1.237	0.217
mom.iq	3.334e+00	2.093e+00	1.593	0.112
I(mom.iq^2)	-1.221e-02	1.078e-02	-1.133	0.258
mom.hs	8.129e+01	1.093e+02	0.743	0.458
mom.iq:mom.hs	-1.253e+00	2.259e+00	-0.554	0.580
I(mom.iq^2):mom.hs	4.623e-03	1.151e-02	0.402	0.688

Residual standard error: 17.91 on 428 degrees of freedom

Multiple R-squared: 0.2387, Adjusted R-squared: 0.2298

F-statistic: 26.84 on 5 and 428 DF, p-value: < 2.2e-16

分析：

1. 所有係數均未達到統計顯著性。
2. “I(mom.iq^2)” 與 “I(mom.iq^2):mom.hs” 觀察：
 1. “I(mom.iq^2)” 為 -1.221e-02 (<0)
 2. “I(mom.iq^2):mom.hs” 顯示加入 “mom.hs” 項目後係數為 4.623e-03 (>0)
3. “mom.hs” 項目對於 “mom.iq” 跟 “kid.score” 之間的關係有調節效果，但統計上不顯著。

因此我們查看一下兩種變數的相關係數：

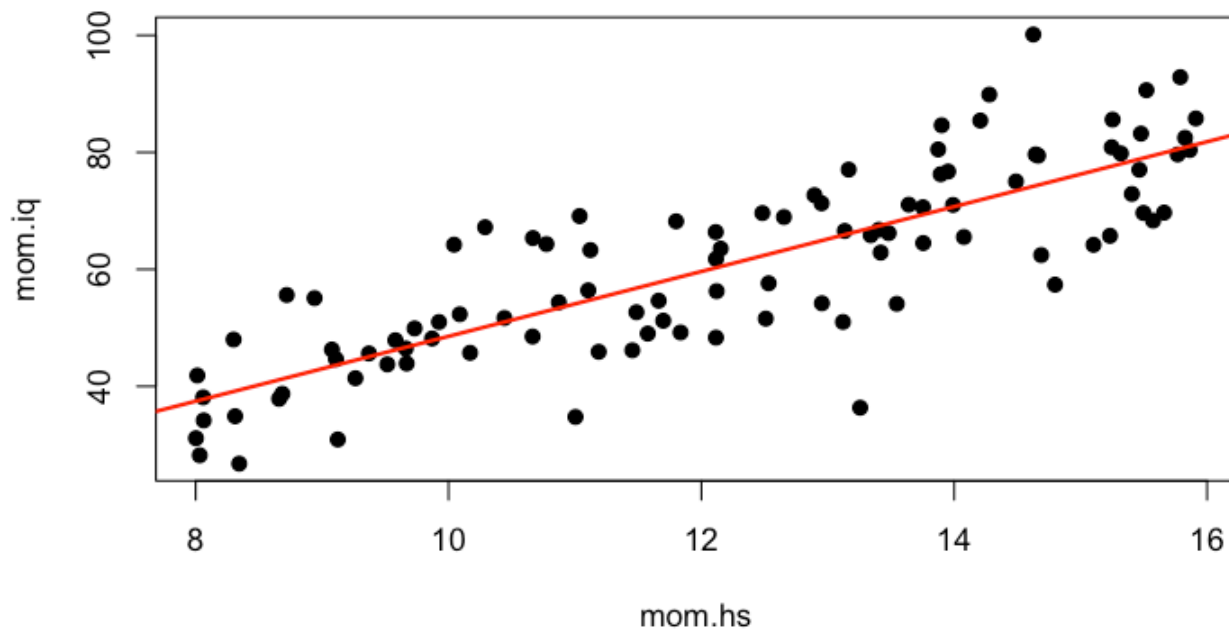
Hide

```
# 生成樣本數據
set.seed(42)
mom.hs <- runif(100, 8, 16)
mom.iq <- mom.hs * 5 + rnorm(100, mean = 0, sd = 10)

# 繪製散佈圖
plot(mom.hs, mom.iq, main = "Scatter Plot of mom.hs vs mom.iq",
     xlab = "mom.hs", ylab = "mom.iq", pch = 19)

# 添加趨勢線
model <- lm(mom.iq ~ mom.hs)
abline(model, col = "red", lwd = 2)
```

Scatter Plot of mom.hs vs mom.iq



計算兩個變數的相關係數：

[Hide](#)

```
# 計算相關係數
correlation <- cor(mom.hs, mom.iq)
print(paste("Correlation: ", round(correlation, 3)))
```

```
[1] "Correlation: 0.824"
```

分析：

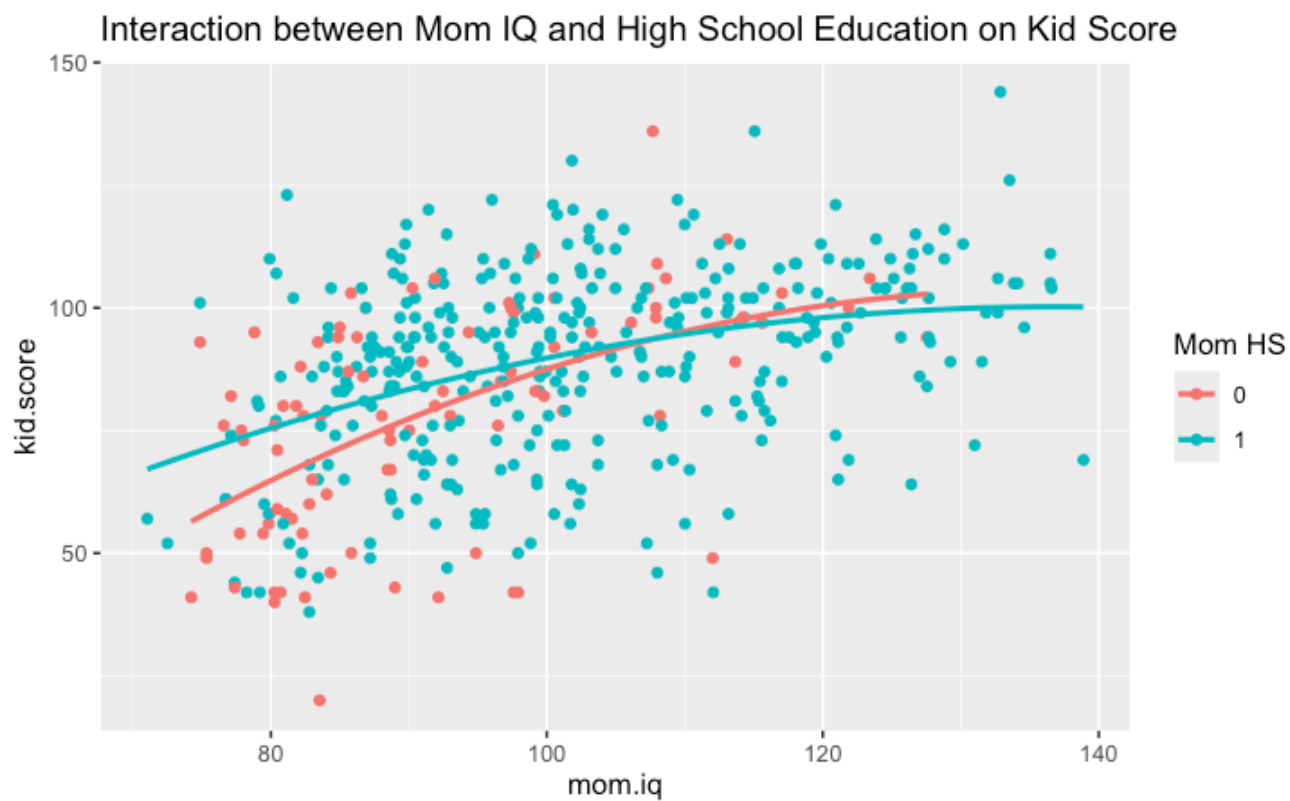
1. “mom.hs”, “mom.iq” 兩個變數之間有高度的相關性。

視覺化

為不同的 mom.hs 分組繪製二次迴歸曲線

[Hide](#)

```
ggplot(kid_iq, aes(x = mom.iq, y = kid.score, color = factor(mom.hs))) +
  geom_point() +
  stat_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE) +
  labs(color = "Mom HS") +
  ggtitle("Interaction between Mom IQ and High School Education on Kid Score")
```



Hide

```
# 為不同的 mom.hs 分組繪製二次迴歸曲線  
# formula = y ~ poly(x, 2) 指定二次多項式模型，以便捕捉可能的非線性關係
```

一次微分視覺化

$$d(y)/d(\text{mom.iq}) = b1 + 2b2 \cdot \text{mom.iq} + b4 \cdot \text{mom.hs} + 2b5 \cdot \text{mom.iq} \cdot \text{mom.hs}$$

Hide

```
# 提取模型係數
coefficients <- coef(model_interaction)
b1 <- coefficients["mom.iq"]
b2 <- coefficients["I(mom.iq^2)"]
b3 <- coefficients["mom.hs"]
b4 <- coefficients["mom.iq:mom.hs"]
b5 <- coefficients["I(mom.iq^2):mom.hs"]

# 定義一階導數的函數，針對不同的 mom.hs 狀態
first_derivative_hs0 <- function(x) {
  b1 + 2 * b2 * x
}

first_derivative_hs1 <- function(x) {
  (b1 + b4) + 2 * (b2 + b5) * x
}

# 繪製一階導數的圖，顯示 mom.hs = 0 和 mom.hs = 1 的情況
plot_range <- seq(min(kid_iq$mom.iq), max(kid_iq$mom.iq), length.out = 100)

# 繪製圖形
plot(plot_range, first_derivative_hs0(plot_range), type = "l", col = "red", lty = 1, lwd = 2,
      main = "First Derivative of the Model with Interaction",
      xlab = "Mom IQ", ylab = "First Derivative of Kid Score")

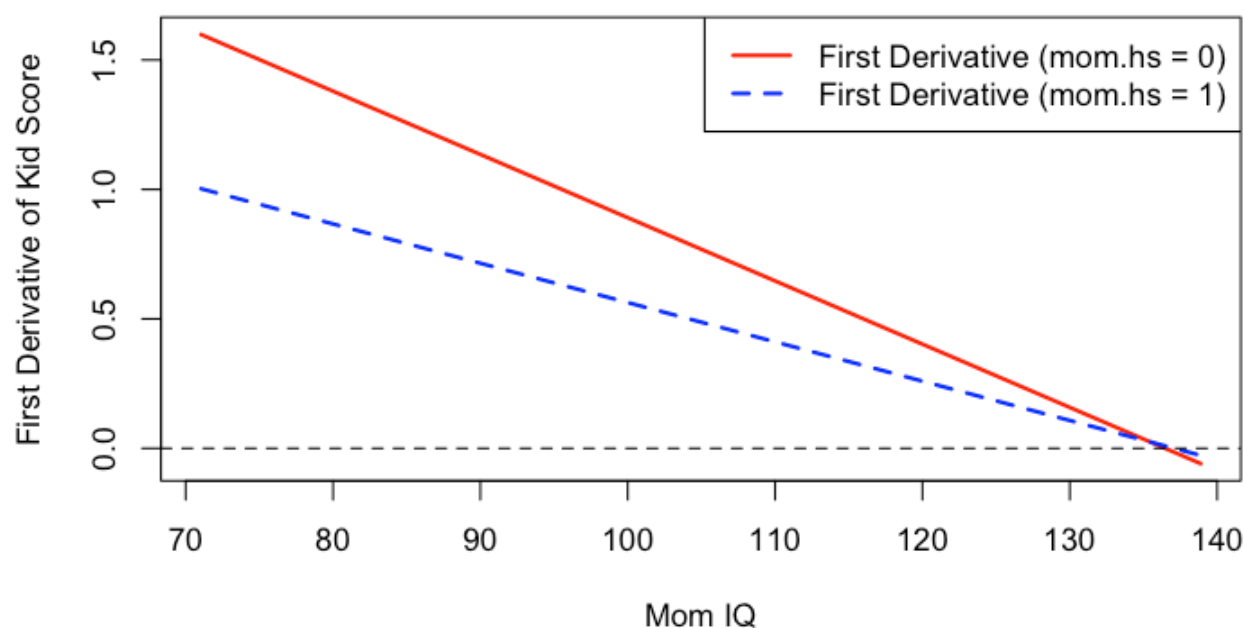
lines(plot_range, first_derivative_hs1(plot_range), col = "blue", lty = 2, lwd = 2)
```

Hide

```
# 在 y 軸畫出 0 的水平線
abline(h = 0, col = "black", lty = 2)

# 添加圖例
legend("topright", legend = c("First Derivative (mom.hs = 0)", "First Derivative (mom.hs = 1)"),
      col = c("red", "blue"), lty = c(1, 2), lwd = 2)
```


First Derivative of the Model with Interaction



分析：

1. 從頓模型的一階導數視覺化結果看來，隨著 mom.iq 變大，一階導數的數值變化是從正值到負值，顯示無論是否「媽媽學歷」，模型結果都呈現倒 U 型。
2. 而 “mom.hs” = 0 的一階導數變化比起 “mom.hs” = 1 數值變化較大，有陡峭的倒 U 形狀。（但統計上不顯著）

小結

將媽媽的高中學歷（mom.hs）加入模型後，並與 mom.iq 和二次項交互作用，雖然模型顯示 mom.hs 具有交互作用，在視覺化上具有倒 U 關係，但統計上係數均 **未達到顯著性**，因此不能說模型具有倒 U 關係。

本組進一步對 “mom.hs”, “mom.iq” 模型變數之間相關性為 0.824，兩種變數高度相關，這可能導致加入 “mom.hs” 變數可能影響模型解釋性。