迴歸分析 第 2 次 期中考

總分 120, 時間 3:10pm-5:00pm., 請寫下重要計算過程與理由.

1 (20%) 簡答題與是非題, 若下列敘述錯誤, 請簡答錯誤理由, 並簡答正確的定義

1. (2%) 何謂統計學?
   A: 處理與分析量化資料, 發展出特定的分析方法, 將這些常見的分析資料方法歸納與收集成, 便成爲所謂的 **統計學** (**statistics**).

2. (2%) 何謂迴歸分析, 有何特點?
   迴歸分析 (**regression analysis**) 的目的在於討論一個研究資料內許多變數之間的關係或關聯性, 迴歸分析主要用於分析一個研究資料的 1 個連續型反應變數與多個解釋變數之間的關係.

3. (2%) 迴歸分析的用途?
   A: 迴歸分析通常是用數學方程式或統計模型量化變數之間的關係, 迴歸分析用途可以對資料進行摘要性描述 (data summary), 量化解釋變數對連續型反應變數的影響 (effect quantification), 提供研究者統計方程式, 可以用解釋變數的數值來預測連續型反應變數的數值 (response prediction).

4. (2%) 型一誤差表示虛無假說爲錯誤的機率.
   A: 型一誤差表示給定虛無假說爲眞, 拒絕虛無假說的機率.

5. (3%) 一個觀測到的研究樣本, 計算樣本內連續型反應變數之樣本平均值的抽樣分配, 表示母體平均值在研究樣本內的分配.
   A: 樣本平均值的抽樣分配表示從母體中抽取各種可能的樣本, 每一組樣本的樣本平均值之數值分配.

6. (3%) 一個觀測到的研究樣本, 計算樣本內連續型反應變數之樣本平均值 $\bar{x}$ 的 95% 信賴區間爲 $\text{CI} = (C_L, C_U)$, 表示此信賴區間 CI 會包含母體平均值的機率有 95%.
   A: 單一觀測樣本, 不是包含母體平均值, 就是不包含母體平均值, 95% 信賴區間爲從母體中抽取各種可能的樣本, 每一組樣本計算一個 95% 信賴區間, 所有可能的樣本的 95% 信賴區間中, 約有 95% 機會會包含母體平均值.

7. (3%) 一個觀測到的研究樣本, 計算觀測樣本所得到的 p-value, 表示此研究樣本是來自虛無假說情形的機率.
   A: p-value 表示虛無假說爲眞, 檢定統計量 (隨機變數) 比一個研究樣本觀測到的檢定統計量還要極端的機率.

8. (3%) 型二誤差表示對立假說爲錯誤的機率, 且樣本數愈大, 越容易拒絕虛無假說, 型二誤差也會隨樣本數愈大而增大.
   A: 型二誤差表示給定對立假說爲眞, 接受虛無假說的機率, 給定型一誤差, 型二誤差爲樣本數函數, 樣本數愈大檢定力愈大, 型二誤差也愈小.

2 (10%) 簡答題

Example 1: 比較 2 種運動治療對保齡球成績之影響

一位研究者想要比較 2 種運動治療對年輕且喜好保齡球運動之病患, 在肢體損傷後復建的治療效應, 進行一個前導性的預備研究, 收集了 60 位同意參與的個體, 分成 2 組 (Method), 分別進行 2 種不同運動治療. 在運動治療開始前, 紀錄基本變數包含年紀 (歲, age), 身高 (HT), 體重 (WT), 使用保齡球重量 (Ball). 然後預測 2 局保齡球賽局, 紀錄成績 (Test0), 球瓶一球全倒次數 (Strike0), 球瓶兩球全倒次數 (Spare0). 接著 2 組分別進行運動治療六週後, 再次進行 2 局保齡球賽局, 同樣紀錄成績 (Test1), 球瓶一球全倒次數 (Strike1), 球瓶兩球全倒次數 (Strike1). 部分個體資料在表 1, 資料在檔案 **ExTherBowling.csv**.

利用本資料說明.

1. (2%) 本研究的目的與研究設計 (Study design)?
   A: 研究者想要比較 2 種運動治療對年輕且喜好保齡球運動之病患, 在肢體損傷後復建的治療效應, 研究設計爲實驗性質的研究設計.
2. (2%) 本研究的主要的反應變數 (Response variable) 與主要的解釋變數 (Explanatory variables)?
   A: 主要的反應變數爲紀錄成績 (Test1), 主要的解釋變數爲 2 組 (Method), 代表 2 種不同運動治療.
3. (2%) 本研究的樣本空間 (Sample space) 與隨機變數 (Random variable)?
   A: 樣本空間爲反應變數紀錄成績 (Test1) 的各種可能分數的集合, 隨機變數爲函數, 將各種可能分數轉換成實數, $X(\text{Test1}) = Test1$.
4. (2%) 本研究的科學假說 (Scieitific hypothesis) 與統計假說 (Statistical Hypothesis)?
   根據研究目的的科學假說爲 2 種不同運動治療的紀錄成績 (Test1) 不相等.
   統計假說中的虛無假說爲
   $$H_0 : \mu_{\text{method}_1} = \mu_{\text{method}_2}, \tag{2.1}$$
   $$H_A : \mu_{\text{method}_1} = \mu_{\text{method}_2}. \tag{2.2}$$
5. (2%) 分析本研究的合適的統計分析方法 (Statistical method) 有哪些?
   A: (Any one of) two-sample t test, two-sample wilcoxon rank sum test, permutation test ...

Table 1: 比較 2 種運動治療對保齡球成績之影響: 部份個體資料

ID	Method	Age	HT	WT	Ball	Test0	Strike0	Spare0	Test1	Strike1	Spare1
1	1	21	160	50	9	142	0	0	160	0	1
2	1	22	158	50	8	105	0	1	102	0	1
3	1	23	158	50	8	80	0	0	136	1	0
4	1	22	155	46	10	135	0	0	196	2	5
5	1	21	156	47	8	117	1	2	112	1	1
						...					

3 (40%) 複迴歸模型

Example 2: 血壓與身體質量指數

一個關於血壓與身體質量指數的調查研究, 研究者收集 32 爲自願者的資料, 測量收縮壓 (sbp, mmHg), 身體質量指數 (bmi, kg/m^2), 年紀 (age, years old) 與抽菸習慣 (smk), smk = 0, 無抽菸, smk = 1, 有抽菸習慣. 資料在檔案 **HTBMI.xls**, 資料如表 2.

考慮配適複迴歸模型, 反應變數爲 sbp, 解釋變數爲 bmi 與 age, 分析結果的參數估計與變異數分析表, 如下表 3.

1. (5%) 依據資料, 請寫下複迴歸模型, 說明上述複迴歸模型的統計假設.
   A: 複迴歸模型

$$Y_i(\text{sbp}) = \beta_0 + \beta_1 \text{bmi}_i + \beta_2 \text{age}_i + \varepsilon_i, i = 1, 2, \ldots, 32; \tag{3.1}$$

$$\text{或}\ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i, i = 1, 2, \ldots, n; \tag{3.2}$$

$$\text{或}\ \mathcal{E}(Y_i \mid \underline{\mathbf{X}}_i = \underline{\mathbf{x}}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}. \tag{3.3}$$

複迴歸模型關於誤差項的統計假設如同簡單線性迴歸模型,

$$\textbf{A1}.\quad \mathcal{E}(\varepsilon_i) = 0 \quad (\text{線性}); \tag{3.4}$$

$$\textbf{A2}.\quad \textbf{Var}(\varepsilon_i) = \sigma^2; (\text{同質性}) \tag{3.5}$$

$$\textbf{A3}.\quad \textbf{Cov}(\varepsilon_i, \varepsilon_j) = 0,\ \text{對任一}\ i \neq j\ \text{成立}; \quad (\text{獨立性}) \tag{3.6}$$

$$\textbf{A4}.\quad \varepsilon_i \sim \text{NID}(0, 1), i = 1, 2, \ldots, n. \tag{3.7}$$

(常態分配性, NID 表示爲獨立的高斯 (常態) 分配)

或是 A: 複迴歸模型

$$\underline{\mathbf{y}}_{n\times 1} = \mathbf{X}_{n\times(p+1)} \underline{\boldsymbol{\beta}}_{(p+1)\times 1} + \underline{\boldsymbol{\varepsilon}}_{n\times 1} \tag{3.8}$$

$$\text{且}\quad y_i = \underline{\mathbf{x}}_i^T \underline{\boldsymbol{\beta}} + \varepsilon_i, \quad i = 1, 2, \ldots, n. \tag{3.9}$$

複迴歸模型關於誤差項的統計假設爲

$$\textbf{C1}.\quad \mathcal{E}(\underline{\boldsymbol{\varepsilon}}) = \underline{\mathbf{0}} \quad (\text{線性}); \tag{3.10}$$

$$\textbf{C2}.\quad \textbf{Var}(\underline{\boldsymbol{\varepsilon}}) = \sigma^2 \mathbf{I}_n; (\text{同質性}) \tag{3.11}$$

$$\textbf{C3}.\quad \textbf{Cov}(\varepsilon_i, \varepsilon_j) = 0,\ \text{對任一}\ i \neq j\ \text{成立}; \quad (\text{獨立性}) \tag{3.12}$$

$$\textbf{C4}.\quad \underline{\boldsymbol{\varepsilon}} \sim \text{N}(\underline{\mathbf{0}}, \sigma^2 \mathbf{I}_n). \tag{3.13}$$

(常態分配性)

2. (5%) 依據資料, 計算參數估計表中的 “t value” 空格, 請寫其虛無假說 H_0 與對立假說 H_A.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  65.0742    12.7964   5.085  0.00002
bmi           0.9751     0.5402   1.805   0.0815
age           1.0452     0.3861   2.707   0.0113
```

   A: 給定 age (X_2) 在迴歸模型中,

$$H_0 : \beta_{\text{bmi}} = 0, \quad H_A : \beta_{\text{bmi}} \neq 0 \tag{3.14}$$

$$\text{或}\quad H_0 : \beta_1 = 0, \quad H_A : \beta_1 \neq 0 \tag{3.15}$$

3. (10%) 依據資料, 參數估計表 bmi 與 age 的迴歸係數 (“Estimate”) 估計值的意義如何解釋?
   A: 給定 age 不變, bmi 每增加 1 kg/m^2, sbp 增加 0.9751 mmHg.
   給定 bmi 不變, age 每增加 1 Y/O, sbp 增加 1.0452 mmHg.

4. (5%) 變異數分析表中的虛無假說 H_0 與對立假說 H_A?

$$
\begin{aligned}
& H_0 : \mathcal{E}(Y \mid X) = \beta_0 + \beta_1 X_1 && (3.16)\\
\text{相對}\quad & H_A : \mathcal{E}(Y \mid X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2; && (3.17)\\
\text{或}\quad & H_0 : y_i = \beta_0 + +\beta_1 X_{1i} + \varepsilon_i && (3.18)\\
\text{相對}\quad & H_A : y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \quad i = 1, 2, \ldots, n. && (3.19)\\
\text{或}\quad & H_0 : \text{給定 bmi } (X_1) \text{ 在迴歸模型中}, \beta_2 = 0 && (3.20)\\
\text{相對}\quad & H_A : \text{給定 bmi } (X_1) \text{ 在迴歸模型中}, \beta_2 \neq 0 && (3.21)\\
\text{或}\quad & H_0 : \text{給定 bmi } (X_1) \text{ 在迴歸模型中, age } (X_2) \text{ 與 } Y \text{ 沒有線性關係}; && (3.22)\\
\text{相對}\quad & H_A : \text{給定 bmi } (X_1) \text{ 在迴歸模型中, age } (X_2) \text{ 與 } Y \text{ 沒有線性關係}. && (3.23)
\end{aligned}
$$

5. (10%) 完成變異數分析表中空格 “?”.

```
Analysis of Variance Table

Response: sbp
          Df Sum Sq Mean Sq F value       Pr(>F)
bmi        1 3537.9  3537.9 44.5048      < 0.0001
age        1  582.6   582.6  7.3293        0.0113
Residuals 29 2305.4    79.5
```

6. (5%) 依據資料, 解釋 R^2.
A: sbp 總變異度的 64%, 可由線性迴歸模型中的 bmi 與 age 解釋.

Table 2: 血壓與身體質量指數: 資料

id	sbp	bmi	age	smk	id	sbp	bmi	age	smk
1	135	18.76	45	0	17	145	23.60	49	1
2	122	22.51	41	0	18	142	20.24	46	1
3	130	21.00	49	0	19	135	21.71	57	0
4	148	27.68	52	0	20	142	24.01	56	0
5	146	19.79	54	1	21	150	26.28	56	1
6	129	17.90	47	1	22	144	27.51	58	0
7	162	26.68	60	1	23	137	22.96	53	0
8	160	26.12	48	1	24	132	22.10	50	0
9	144	13.68	44	1	25	149	23.01	54	1
10	180	36.37	64	1	26	132	20.17	48	1

Table 3: 血壓與身體質量指數: 複迴歸模型

```
> m1<-lm(sbp~bmi+age)
> summary(m1)
sbp ~ bmi + age
            Estimate  Std. Error   t value    Pr(>|t|)
(Intercept)  65.0742     12.7964     5.085     0.00002
bmi           0.9751      0.5402      ?        0.0815
age           1.0452      0.3861     2.707     0.0113
Residual standard error: 8.916 on 29 degrees of freedom
Multiple R-squared: 0.6412,     Adjusted R-squared: 0.6165
F-statistic: 25.92 on 2 and 29 DF,  p-value: < 0.001

> anova(m1)
Response: sbp
          Df   Sum Sq   Mean Sq   F value      Pr(>F)
bmi        1   3537.9    3537.9   44.5048    < 0.0001
age        1    582.6     582.6      ?         0.0113
Residuals  ?     ?          ?
Total      ?   6425.9
```

4 (10%) 複迴歸模型與變異數分析表

Example 3: 高血壓, 年紀與壓力指數

一個醫學研究探討高血壓, 年紀與壓力指數的關係, 資料爲隨機抽樣 20 位成年人的調查結果, 變數說明如表 4, 資料如表 5.

Table 4: 血壓, 年紀與壓力指數關係變數說明

變數	說明
bp (Y)	平均動脈壓 (mean arterial blood pressure), 單位; mmHg.
age (X_1)	年紀, 單位; 年.
wt (X_2)	重量, 單位; 公斤.
bsa (X_3)	身體表面積 (body surface area), 單位; 平方公尺.
dur (X_4)	患有高血壓時間, 單位; 年.
hr (X_5)	基礎心跳速率, 單位; 次數/分.
stress (X_6)	壓力指數.

Table 5: 血壓, 年紀與壓力指數關係: 部分資料

bp	age	wt	bsa	dur	hr	stress
105	47	85.4	1.75	5.1	63	33
115	49	94.2	2.10	3.8	70	14
116	49	95.3	1.98	8.2	72	10

研究者配適多變數線性迴歸分析,

$$m1 : y_i \; = \; \beta_0 + \beta_1 \,\mathrm{age} + \beta_2 \,\mathrm{wt} + \beta_3 \,\mathrm{bsa} + \beta_4 \,\mathrm{dur} + \beta_5 \,\mathrm{hr} + \beta_6 \,\mathrm{stress} + \varepsilon_i, i = 1, 2, \ldots, 20. \quad (4.1)$$

模型變異是分析結果如表 6 中第 1 個 ANOVA table, 研究者配適另一個模型,

$$m2 : y_i \; = \; \beta_0 + \beta_1 \,\mathrm{age} + \beta_2 \,\mathrm{wt} + \beta_3 \,\mathrm{bsa} + \varepsilon_i, i = 1, 2, \ldots, 20. \quad (4.2)$$

模型變異是分析結果如表 6 中第 2 個 ANOVA table, 研究者想要比較 2 個模型, 分析 dur, hr, stress 是否同時爲重要的解釋變數, 使用變異數分析.

1. (4%) 依據資料, 請寫 表 6 中變異數分析表中的虛無假說 H_0 與對立假說 H_A,
2. (6%) 請完成變異數分析表 6 中第 3 個 ANOVA table.

A:

$$H_0 : m1 \text{ 模型成立, 模型可由 } m2 \text{ 簡化成 } m1; \quad (4.3)$$

$$H_A : m2 \text{ 模型成立, 模型無法由 } m2 \text{ 簡化成 } m1. \quad (4.4)$$

$$H_0 : y_i = \beta_0 + \beta_1 \,\mathrm{age} + \beta_2 \,\mathrm{wt} + \beta_3 \,\mathrm{bsa} + \varepsilon_i, i = 1, 2, \ldots, 20. \quad (4.5)$$

$$H_A : y_i = \beta_0 + \beta_1 \,\mathrm{age} + \beta_2 \,\mathrm{wt} + \beta_3 \,\mathrm{bsa} + \beta_4 \,\mathrm{dur} + \beta_5 \,\mathrm{hr} + \beta_6 \,\mathrm{stress} + \varepsilon_i, i = 1, 2, \ldots, 20. \quad (4.6)$$

或 H_0 : 給定 age, wt, bsa 在迴歸模型中, dur, hr, stress 對 sbp 都沒有影響力; (4.7)

H_A : 給定 age, wt, bsa 在迴歸模型中, dur, hr, stress 對 sbp 至少有一個有影響力; (4.8)

或 H_0 : 給定 age, wt, bsa 在迴歸模型中, $\beta_4 = \beta_5 = \beta_6 = 0$; (4.9)

H_A : 給定 age, wt, bsa 在迴歸模型中, β_4, β_5, β_6 至少有一個不爲 0. (4.10)

```
Model 1: bp ~ age + wt + bsa
Model 2: bp ~ age + wt + bsa + dur + hr + stress
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     16 3.05614
2     13 2.15586  3   0.90028 1.8096 0.1951
```

Table 6: 血壓, 年紀與壓力指數: 變異數分析表

```
(1)
> m1<-lm(bp~age+wt+bsa+dur+hr+stress)
Response: bp
          Df  Sum Sq Mean Sq   F value    Pr(>F)
age        1 243.266 243.266 1466.9140  < 0.001
wt         1 311.910 311.910 1880.8435  < 0.001
bsa        1   1.768   1.768   10.6599    0.006
dur        1   0.335   0.335    2.0207    0.178
hr         1   0.123   0.123    0.7421    0.404
stress     1   0.442   0.442    2.6659    0.126
Residuals 13   2.156   0.166

(2)
> m2<-lm(bp~age+wt+bsa)
Response: bp
          Df  Sum Sq Mean Sq  F value    Pr(>F)
age        1 243.266 243.266 1273.585   < 0.001
wt         1 311.910 311.910 1632.961   < 0.001
bsa        1   1.768   1.768    9.255     0.007
Residuals 16   3.056   0.191

(3)
Analysis of Variance Table
Model 1:  bp ~ age + wt + bsa
Model 2:  bp ~ age + wt + bsa + dur + hr + stress
  Res.Df       RSS   Df    Sum of Sq     F
1     ?        ?
2     ?        ?     ?         ?         ?
```

5 (10%) 複迴歸模型與解釋變數關係

(A) 例題 (Example) 3 中, 令 $X_3 = \text{age} \times \text{bsa}$, 研究者配適另一個模型,

$$m4 : y_i = \beta_0 + \beta_1 \,\text{age} + \beta_2 \,\text{wt} + \beta_9 \,\text{bsa} + \beta_3 x_3 + \varepsilon_i, i = 1, 2, \ldots, 20. \tag{5.1}$$

參數估計如表 7.

Table 7: 血壓, 年紀與壓力指數: 交互作用

```
(4)
> x3<-age*bsa
> in.fit<-lm(bp~age+wt+bsa+x3)
             Estimate Std. Error   t value   Pr(>|t|)
(Intercept)  -1.816   43.607        -0.04     0.967
age           0.457    0.899         0.50     0.619
wt            0.905    0.051        17.90   < 0.001
bsa          -1.178   21.379        -0.05     0.957
x3            0.120    0.442         0.27     0.789
```

1. (2%) 依據資料, 表 7 中參數估計 $\widehat{\beta}_3$ 如何解釋?
   A: β_3 無法單獨解釋, 或許只能說給定 wt 不變, age 與 bp 不是線性關係, 或對 bp 的影響可能存在著 age 與 bsa 的交互作用.
2. (2%) 表 7 中若 $\text{age} = X_1$ 增加 1 單位, 則 $\text{bp} = Y$ 會增加多少?
   A: 給定 wt 與 bsa 不變, age 增加 1 年, bp 增加 $\beta_1 + \beta_3 \,\text{bsa}$ mmHg, bp 的改變量大小, 不僅受到模型參數β_1, β_3, 也受到 bsa 的影響, 即受到 bp 的原始 bsa 所在的位置影響.
3. (2%) 在表 6 中的 (1) 與 (2), 都顯示 age, wt, bsa 爲顯著重要的解釋變數, 爲何表 7 中的 (4) age 與 bsa 不再是顯著重要的解釋變數?
   A: 有可能 X_3 與 age, bsa 高度相關, 或接近線性相依, 或高度共線性.

(B) 例題 (Example) 3 中, 令 $X_4 = \text{age} \times \text{age}$, 研究者配適另一個模型,

$$m5 : y_i = \beta_0 + \beta_1 \,\text{age} + \beta_2 \,X_4 + \varepsilon_i, i = 1, 2, \ldots, 20. \tag{5.2}$$

參數估計如表 8.

Table 8: 血壓, 年紀與壓力指數: 二次項

```
(5)
> x4<-age*age
> quad.fit<-lm(bp~age+x4)
              Estimate  Std. Error   t value   Pr(>|t|)
(Intercept) -154.712     254.041     -0.61      0.551
age            9.411      10.157      0.93      0.367
x4            -0.079       0.101     -0.77      0.443
```

1. (2%) 依據資料, 表 8 中參數估計 $\widehat{\beta}_2$ 如何解釋?
   β_2 無法單獨解釋, 或許只能說 age 與 bp 不是線性關係, 可能存在著 bp 與 age 的二次項關係.

2. (2%) 表 8 中若 age $= X_1$ 增加 1 單位, 則 bp $= Y$ 會增加多少?
   A: 反應變數從 bp 增加 $\beta_1 + \beta_2(2 \times \text{age} + 1)$ mmHg. 因此反應變數從 bp 改變成 bp* 的改變量大小, 不僅受到模型參數β_1, β_2 的影響, 也受到 age 的影響, 即受到 bp 的原始 age 所在的位置影響.

6 (10%) 迴歸模型與變數轉換

Example 4: 老鼠腦組織與血液中甲苯濃度

甲苯 (Toluene) 是一種無色的液體, 但具有明顯的味道, 是一種常見的有機溶劑, 常存在於油漆, 稀釋劑, 黏著劑, 指甲油與汽油. 甲苯對健康的影響以腦部最受到關切, 甲苯會引起頭痛, 恍惚與喪失記憶. 一個關於甲苯的醫學研究, 將 60 隻新生老鼠暴露於甲苯環境, 分別爲 50, 100, 500 或 1000 ppm 甲苯 (tol, 單位: ppm), 暴露時間共 3 小時, 將老鼠繼續飼養, 並在在老鼠不同年紀 (age), 測量老鼠血液中甲苯濃度 (bloodtol, 單位: ppm) 老鼠腦組織甲苯濃度 (braintol, 單位: ppm) 老鼠腦組織甲苯重量 (weight, 單位: gm), 研究者探討老鼠暴露於甲苯的量 (tol), 暴露後老鼠在不同的年紀 (age), 老鼠血液中甲苯濃度 (bloodtol), 對老鼠腦組織甲苯重量 (weight) 的影響, 資料在檔案 **TolueneRat.xls**, 部分資料如表 9.

Table 9: 老鼠腦組織與血液中甲苯濃度: 部分資料

RAT	blood	brain	tol	bwt	age
1	0.553	0.481	50	393	95
2	0.494	0.584	50	378	95
3	0.609	0.585	50	450	95
4	0.763	0.628	50	439	95
5	0.420	0.533	50	397	95

考慮建構以下幾個簡單迴歸模型如下,

$$\log_{10}(Y_i) = \lambda_0 + \lambda_1 \log_{10}(x_{1i}) + \varepsilon_i. \tag{6.1}$$

參數估計結果如表 10.

Table 10: 老鼠腦組織與血液中甲苯濃度: 變數轉換

```
> m7<-lm(log10(brain)~log10(blood))
              Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)    0.16493      0.02074     7.952    < 0.001
log10(blood)   1.01502      0.02258    44.948    < 0.001
```

1. (5%) 依據資料, 參數估計中 $\log_{10}$(blood) 的迴歸係數 $\widehat{\beta}_1$ 如何解釋?
   A: $\log_{10}$(blood) 每增加 1 單位, $\log_{10}$(brain) 1.01502 單位.

2. (5%) 若 $\text{blood}^\star = 100 \times \text{blood}$, 則 $\text{brain}^\star$ 會增加多少?

$$10^{2\times 1.01502} = 107.1618. \tag{6.2}$$

7 (10%) 迴歸理論

A: 參考答案課本 58 頁, 學生的答案要比課本 58 頁的參考答案更詳細.

8 (10%) 看圖説故事

Example 5: NIDDM 臨床試驗比較合成胰島素與一般胰島素對降低血糖的效應

一個臨床試驗針對成人第二型糖尿病 (NIDDM) 進行研究, 研究者比較 2 種不同的治療: 新的治療藥物爲 “合成胰島素 (Synthetic. Insulin, Syn)” 與 “一般胰島素 (Regular Insulin, Reg)”, 對降低血糖的效應總共 32 位病患, 隨機地分配到合成胰島素治療組與一般胰島素控制組 (Reg. Insulin), 測量基線 (baseline) 之空腹血糖 (FBS1,), 與測量治療 12 個月之後的空腹血糖 (FBS2). 研究的目地是比較 2 種不同的治療對降低血糖和總膽固醇的效應, 是否有任何差異. 表 11 資料在檔案 **CTDMInsuSimple.csv**.

令 $Y = \text{fbs2}$, $X = \text{fbs1}$, $Z = 0$, 表示合成胰島素組, $Z = 1$, 表示一般胰島素組, 研究者考慮下列 4 個統計模型, A, B, C, D,

$$(A): Y = \beta_0 + \beta_1 X + \varepsilon; \tag{8.1}$$
$$(B): Y = \beta_0 + \beta_1 Z + \varepsilon; \tag{8.2}$$
$$(C): Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 ZX + \varepsilon; \tag{8.3}$$
$$(D): Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon; \tag{8.4}$$

圖 1 中, (1), (2), (3), (4) 分別表示這 4 種統計模型.

1. 請敍述圖 1 中每一個小圖 (1), (2), (3), (4) 是 A, B, C, D 中的哪一個統計模型? 並請簡要說明你了理由.
   A:

$$(A) = 2, \quad (B) = 1, \quad (C) = 4, \quad (D) = 3. \tag{8.5}$$

Table 11: Example 5: 部分資料

合成胰島素 Syn. Insulin		一般胰島素 Reg. Insulin	
fbs1	fbs2	fbs1	fbs2
140	100	150	110
150	110	140	100
140	90	130	100

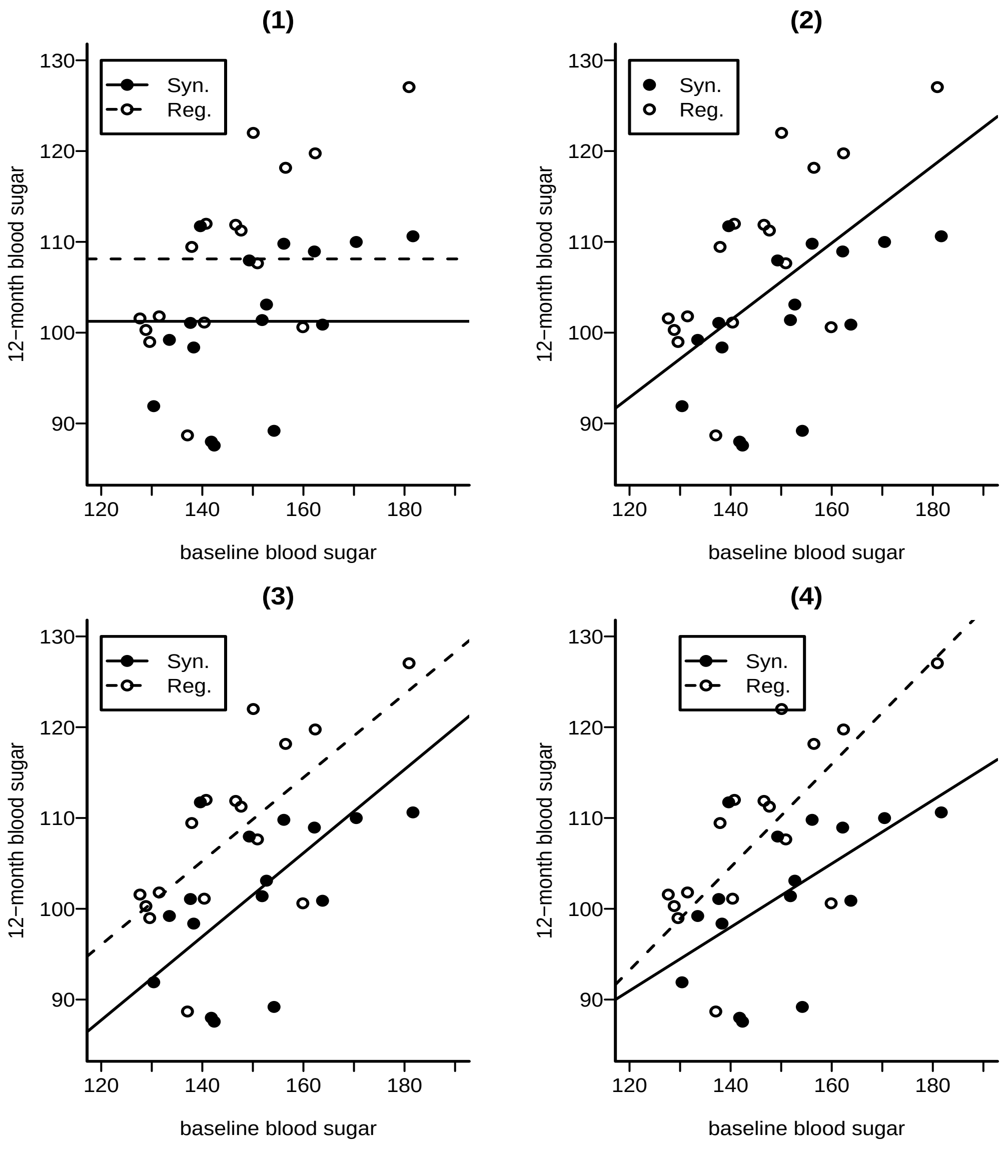

Figure 1: 看圖說故事