

誰是時間管理大師

R 程式設計期末報告

第 14 組

指導老師：李百靈 老師

學生：數學四 407190023 吳依潔

數學四 407190049 趙方于

數學四 407190163 陳愷晴

目錄

第一章 緒論	3
1-1 研究動機	3
1-2 研究目的	3
第二章 問卷設計的概念與架構	4
2-1 概念	4
2-2 問卷內容	4
2-3 抽樣方法	5
2-4 資料結構	5
2-5 資料篩選及轉換	6
第三章 分析結果與結論	7
3-1 敘述統計	7
3-2 假設檢定	11
3-3 迴歸分析	13
3-4 ANOVA	22
3-5 卡方獨立性檢定	25
3-6 對整體的結論	26
第四章 附錄	27
4-1 問卷網址	27
4-2 問卷網址所有步驟的 R 程式碼	27

第一章 緒論

1-1 研究動機

我們想以教授的角度去思考：在學期結束時，班上學生學習成績不如預期，想知道是什麼影響成績，是否會因為學院不同、課餘時間的分配、讀書的習慣、睡眠的時間、上課專心度等影響學期成績。

1-2 研究目的

1. 學生學習成績為什麼會不如預期，時間分配是否會影響成績
2. 比較商、理學院的總平均差異
3. 學生該如何有效安排時間
4. 教授是否要根據性別或是年級做出不同的教學方式
5. 利用什麼讀書習慣和花多少讀書時間能達到最高效率

第二章 問卷設計的概念與架構

2-1 概念

因為想要討論課後娛樂時間、讀書方法和時間分配是否會影響成績，所以問卷內容包括成績、GPA、課後的一些時間分配的問題。

2-2 問卷內容

1. 基本資料

- (1) 性別:男/女
- (2) 學院:商學院/理學院
- (3) 年級:大一/大二/大三/大四/大五/大六/碩一/碩二

2. 成績

- (1) 110 年度第 1 學期學分數
- (2) 110 年度第 1 學期學業平均
- (3) 學業總平均
- (4) GPA
- (5) 讀書習慣:紙本講義 (課本、老師自編講義、手寫筆記…) / 電子講義 (平版、電腦…) / 紙本講義電子講義並用
- (6) 是否有被當的科目:是/否

3. 時間

- (1) 每天睡眠時間
- (2) 一週讀書時間
- (3) 一週使用手機時間 (玩遊戲、滑社群…)
- (4) 一週娛樂時間 (看電影、出去玩、逛街…)
- (5) 這題請選指定答案男生請選 1 女生請選 0:0/1

4. 自我約束力

- (1) 一週參加社團時間
- (2) 一週運動時間
- (3) 一週打工時間
- (4) 幾點上床睡覺 是否常常熬夜
- (5) 是否通勤:是/否

- (6) 作業繳交時間：提早 3 天以上交/當天在時間內交/壓線到最後一刻交/直接忘記 或 不交
- (7) 這題請選指定答案男生請選 1 女生請選 0:0/1
- (8) 上課專注度 0/1/2/3/4/5/6/7/8/9/10

2-3 抽樣方法

研究的母體為淡江大學理學院和商學院的學生，並設計 Google 表單蒐集想要的資訊，利用便利抽樣的方式找數學系上的學生填寫，簡單隨機抽樣的方式放在網路平台給淡江大學的學生填寫。

2-4 資料結構

一開始利用 Google 表單蒐集的資料是 Excel 檔(圖 A 為前 10 筆的 EXCEL 資料)，陣列型資料結構，將檔案讀入 R 後，變成資料框架，而每個變數名稱和一開始的型態如下

1. 基本資料
 - 性別：文字型
 - 學院：文字型
 - 年級：文字型
2. 成績
 - 第一學期學分數：數值型
 - 第一學期平均：數值型
 - 總平均：數值型
 - GPA：數值型
 - 讀書習慣：文字型
 - 是否被當：文字型
3. 時間
 - 睡覺時間：數值型
 - 讀書時間：數值型
 - 滑手機時間：數值型
 - 娛樂時間：數值型
 - 社團時間：數值型

- 運動時間：數值型
- 打工時間：數值型

4. 自我約束力

- 幾點睡覺：文字型
- 是否通勤：文字型
- 交作業時間：文字型
- 上課專注力：數值型

※接著利用附錄的 R 程式碼將資料框架內的變數型態做轉換，並且刪除無效的資料。

1	時間戳記	1. 性別	2. 學院別	3. 年級	4. 110年度第1學期學分	5. 110年度第1學期學分	6. 學業總平均	7. GPA	8. 讀書習慣	9. 是否有被當的科目	10. 每天睡眠時數	11. 一週讀書時數	12. 一週使用手機時間	13. 一週娛樂時間 (看 #這題)
2	4/14/2022 11:30:28	女	理學院	大四	13	95	86.46	3.758	紙本講義+電子講義	否	8	45	50	18
3	4/14/2022 11:35:47	女	理學院	大四	19	91.37	80.47	3.303	紙本講義+電子講義	否	8	18	14	10
4	4/14/2022 12:25:27	女	商學院	大四	11	82.45	75.57	2.848	紙本講義+電子講義	是	5	0.5	24	5.5
5	4/14/2022 15:29:03	男	理學院	大四	30	63.03	68.68	2.269	紙本講義 (課本、是		6	0	15	10
6	4/14/2022 16:40:24	男	理學院	碩一	11	92.91	92.91	4	電子講義 (平版、否		5	5	20	0
7	4/14/2022 18:42:36	男	理學院	大二	12	54.25	50.6	2.4	電子講義 (平版、是		8	3	38.5	4
8	4/14/2022 19:10:18	男	理學院	大三	22	69.73	61.78	1.712	紙本講義+電子講義	否	8	14	30	25
9	4/14/2022 19:15:06	女	理學院	大一	18	75.35	75.35	2.882	紙本講義 (課本、否		6.5	7.5	12	4
10	4/14/2022 19:16:35	男	理學院	大四	14	80.65	83.74	2.875	電子講義 (平版、是		6.5	3	50	2
1	12. 一週使用手機時間	13. 一週娛樂時間 (看 #這題請選指定答案)	14. 一週參加社團時間	15. 一週運動時間	16. 一週打工時間	17. 幾點上床睡覺 是否	18. 是否通勤	19. 作業繳交時間#這題請填指定數字	20. 上課專注度					
2	50	18	0	0	0.5	0	11:30:00 PM	通勤 (住家裡 需要搭車 當天在時間內交	0	8				
3	14	10	0	4	4	0	12:00:00 AM	通勤 (住家裡 需要搭車 提早3天以上交	0	9				
4	24	5.5	0	0.5	0	0	01:30:00 AM	通勤 (住家裡 需要搭車 當天在時間內交	0	7				
5	15	10	1	0	7.5	0	02:00:00 AM	通勤 (住家裡 需要搭車 提早3天以上交	1	1				
6	20	0	1	0	3	21	01:30:00 AM	通勤 (住家裡 需要搭車 提早3天以上交	1	8				
7	38.5	4	1	0	6	0	01:00:00 PM	住宿 (學校附近) 壓線到最後一刻交	1	5				
8	30	25	1	8	10	4	01:00:00 AM	住宿 (學校附近) 當天在時間內交	1	5				
9	12	4	0	3	8	0	01:00:00 AM	住宿 (學校附近) 當天在時間內交	0	8				
10	50	2	1	0	12	16	01:00:00 AM	住宿 (學校附近) 當天在時間內交	1	8				

圖 A

2-5 資料篩選及轉換

1. 資料篩選掉的部分

- (1) 填答時間、防亂答 (皆問性別) 填的不一樣的
- (2) 重複率高的內容 (第一學期學分數、第一學期平均、總平均和 GPA)，因為填答的時間非常接近，絕大部分差不到一分鐘。
- (3) 總平均的離群值 (在下一章會詳細解釋刪除的原因)

2. 轉換：幾點睡覺原始填答為準確的時間 (例如 23:00)，以

22:00~00:00、00:01~04:00、04:01~07:00 及 07:01~21:59 依序分為早
睡、熬夜、通宵與其他

第三章 分析結果與結論

3-1 敘述統計

1. 性別

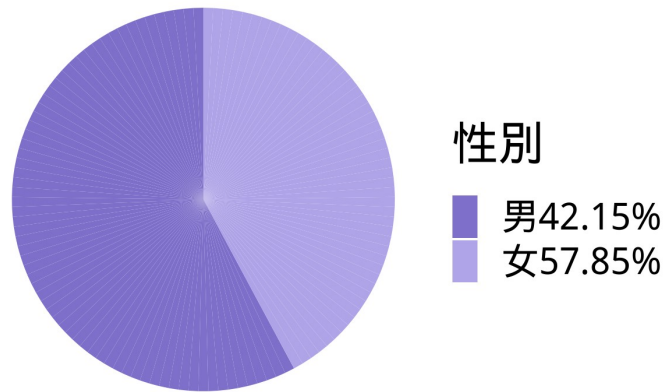


圖 B

由圖 B 得知資料中的男女比大約為 1:1.5，蒐集到的資料女生比較多，和淡江大學理學院加商學院的男女比有一點落差，實際的男女比大約是 1:1。

2. 學院

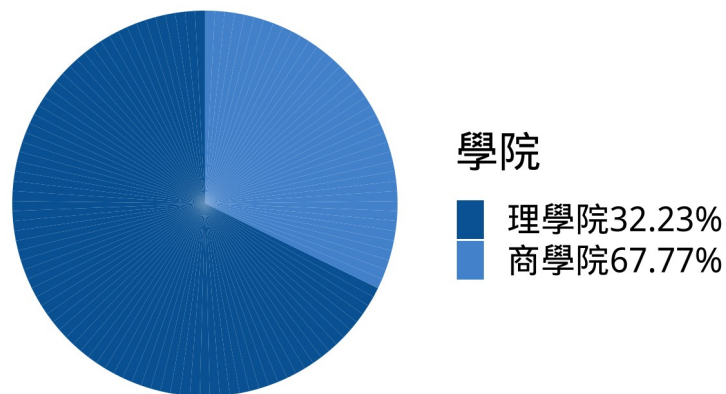


圖 C

由圖 C 得知資料中的學院比例大約為 1:2，蒐集到的資料商學院的人數相對較多，但和淡江大學理學院、商學院的人數比有一點落差，實際比例大約是 1:7.5，沒有和實際上的比例相同。

3. 年級

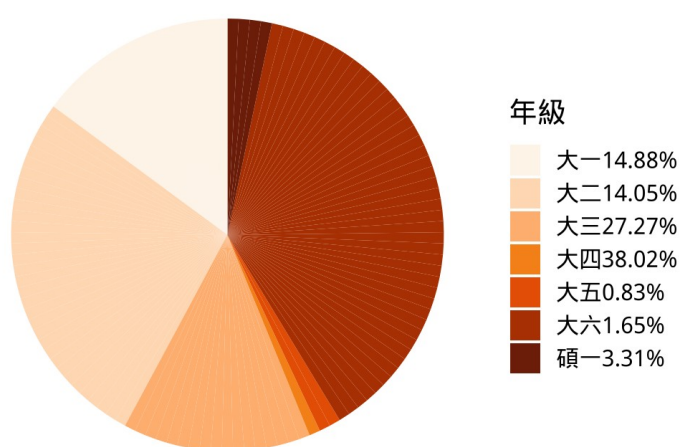


圖 D

由圖 D 得知蒐集來的樣本接近一半是大四的學生，第二多的是大三，而大一和大二人數差不多，大五以上較少。

大一	大二	大三	大四	大五	大六	碩一	碩二
18	17	33	46	1	2	4	0

表(一)

4. 總平均（此報告主要使用的數值型變數）

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	std
16.30	70.00	78.00	75.61	82.30	92.91	10.92

表（二）

總平均直方圖

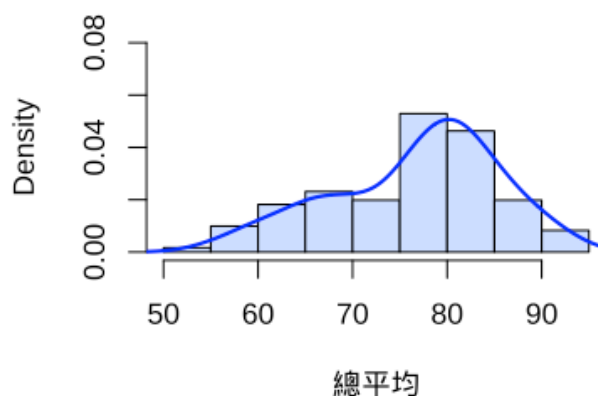


圖 E

由圖 E 得知總平均多數為 75 到 85 分，且分佈有些微的左偏

GPA和總平均的散佈圖



圖 F

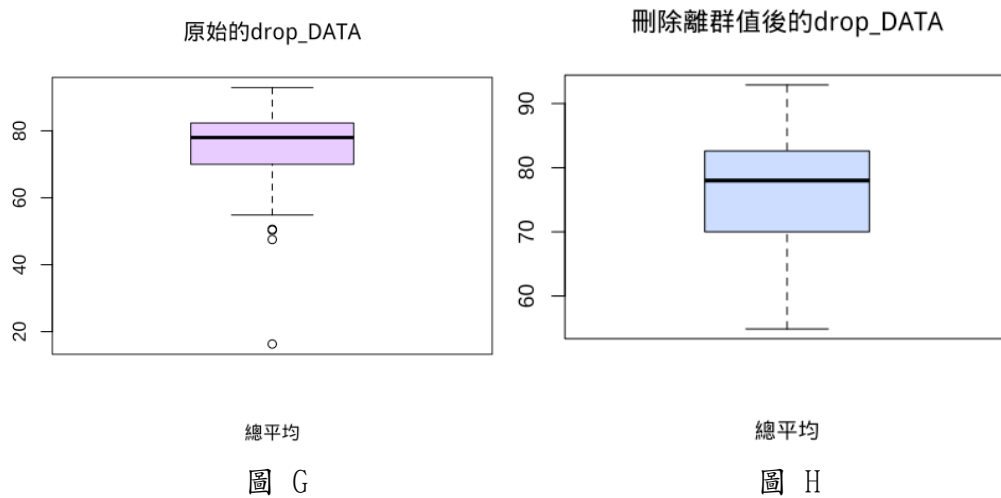
由圖 F 得知總平均和 GPA 呈正向關係，整體的相關係數為 0.730695，理學院的相關係數為 0.7665512，商學院的相關係數為 0.6950785，代表兩者高度相關；理學院的分佈較商學院分散，商學院分數高於 70 分的人佔大多數且集中在 80 分上下。

先篩過重複性太高的資料統計表

分數組距 學院	11~20	41~50	51~60	61~70	71~80	81~90	91~100
理學院	0	1	6	14	8	12	2
商學院	1	0	0	8	33	37	4

表(三)

離群值對迴歸分析影響很大，所以使用箱形圖法[$Q1-1.5*(Q3-Q1)$, $Q3+1.5*(Q3-Q1)$] 來刪除離群值以降低誤差。



刪除離群值後的統計表

分數組距 學院	51~60	61~70	71~80	81~90	91~100
理學院	4	14	8	12	2
商學院	0	8	33	37	4

表(四)

總平均以 10 分為單位類別化，比較容易看得出成績的人數分佈情形。
由表(四)可知，商學院的總平均大多數都是 71 到 90 分，而理學院可能是因為難度較高，所以有一點分散，不過比較多人是分佈在 61 到 90 分。

報告中所有分析方法的 α 皆設定為0.05

3-2 假設檢定

利用假設檢定來了解淡江大學理學院和商學院的總平均是否有差異

i. 利用 Shapiro-Wilk 常態性檢定理學院和商學院的總平均是否符合常態分配

(1) 理學院：

Shapiro-Wilk normality test	
data: S_DATA\$總平均	
W = 0.96256	p-value = 0.2169

表 (五)

由表 (五) 可知 $p\text{-value} = 0.2169 > 0.05$ 故理學院的總平均是常態分配

(2) 商學院：

Shapiro-Wilk normality test	
data: B_DATA\$總平均	
W = 0.96463	p-value = 0.02326

表 (六)

由表 (六) 可知 $p\text{-value}=0.02326$ 沒有很小，因此商學院的總平均是常態分配

ii. 利用 Z 檢定來分析淡江大學理學院和商學院的總平均是否有差異？

H_0 : 兩學院總平均相同 *v.s.* H_1 : 兩學院總平均不同

data: S_DATA\$總平均 和 B_DATA\$總平均	
z 值	-3.463
p-value	0.0005342
95%信賴區間的上界	-2.449801
95%信賴區間的下界	-8.838836
sample estimates	mean of x=72.87897
	mean of y=78.52329
	means is not equal to 0

表 (七)

- iii. 結果:由假設檢定得知，95%信賴區間不包含 0 且 $p\text{-value}=0.0005342$ 小於 0.05，拒絕虛無假設，表示有足夠證據顯示理學院和商學院的總平均有差異。

3-3 迴歸分析

1. 複迴歸:利用所有變數來看哪些變數對總平均有顯著的影響，以及預測使成績增加的方法

- I. 檢查迴歸的前提假設:

常態性:利用直方圖(圖 I)、qq 圖(圖 J)和偏度來檢驗總平均是否符合常態性，其中偏度 $-0.493 < 0$ ，稍微左偏。

三者皆顯示總平均不是常態分佈，所以需要把總平均標準化，使數據較集中、對稱。

標準化規則： $S = (\text{總平均} - \text{mean}(\text{總平均})) / \text{std}(\text{總平均})$ ，標準化後的直方圖為圖 K

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
-2.51	-0.773	0.197	0	0.667	1.856	1

表 (八)

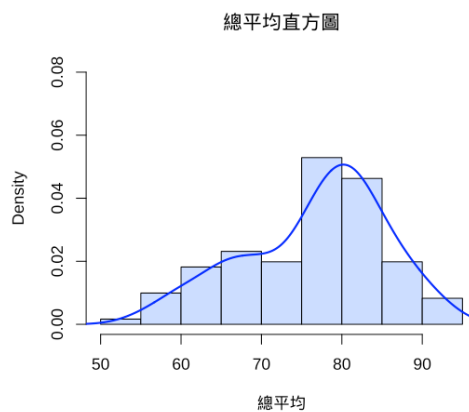


圖 I

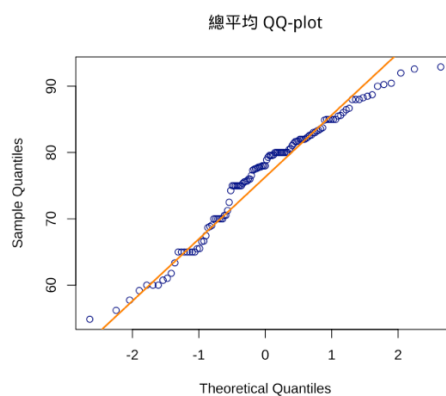


圖 J

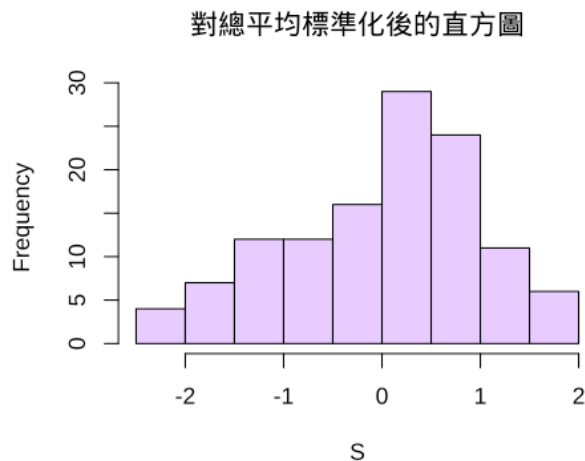


圖 K

II. 檢驗有無共線性:利用變異數膨脹因子(VIF)檢驗

模型為 $\text{lm}(\text{formula}=\text{S} \sim \text{as.factor}(\text{學院}) + \text{as.factor}(\text{年級}) + \text{第一學期學分數} + \text{第一學期平均} + \text{GPA} + \text{as.factor}(\text{讀書習慣}) + \text{讀書時間} + \text{社團時間} + \text{運動時間} + \text{打工時間} + \text{as.factor}(\text{通勤}) + \text{as.factor}(\text{sleep}) + \text{上課專注度}, \text{Data} = \text{drop_DATA})$ ，類別變數在程式裡直接用 `as.factor` 會自動轉成虛擬變數。

自由度是 1 時，GVIF 就是 VIF，不是 1 時要取 $\text{GVIF}^{1/(2 \cdot \text{Df})}$ 的平方看是否小於 5 ($(\text{GVIF}^{1/(2 \cdot \text{Df})})^2 < 5$ 等同於 $\text{VIF} < 5$)，全部都沒有大於 5，所以沒有共線性的問題。

	GVIF	Df	$\text{GVIF}^{1/(2 \cdot \text{Df})}$	$[\text{GVIF}^{1/(2 \cdot \text{Df})}]^2$
<code>as.factor(性別)</code>	1.717	1	1.310	
<code>as.factor(學院)</code>	1.969	1	1.403	
<code>as.factor(年級)</code>	6.885	6	1.174	1.378
第一學期學分數	1.536	1	1.239	
第一學期平均	3.123	1	1.767	
GPA	2.852	1	1.689	
<code>as.factor(讀書習慣)</code>	2.039	2	1.195	1.428

內容過多，只放前幾筆

III. 選定模型

使用 step 函數逐步選取法來決定模型，會自動判別向前選取、向後消去，使所有變數對總平均皆顯著為止。

結果共有 7 種模型，表（九）僅放 AIC 的值，以及選定最小 AIC 的部分結果

1	2	3	4	5	6	7
-171.97	-173.92	-175.87	-177.71	-179.34	-180.55	-181.44

表（九）

Step: AIC=-181.44

$S \sim \text{as.factor(學院)} + \text{as.factor(年級)} + \text{第一學期學分數} + \text{第一學期平均} + \text{GPA} + \text{as.factor(讀書習慣)} + \text{讀書時間} + \text{社團時間} + \text{運動時間} + \text{打工時間} + \text{as.factor(通勤)} + \text{as.factor(sleep)} + \text{上課專注度}$

$$\begin{aligned}
 Y = & -5.916817 + 0.30931D_1 - 0.165488D_{21} - 0.528992D_{23} \\
 & - 0.422192D_{24} - 0.478178D_{25} - 1.483458D_{26} \\
 & - 0.342143D_{27} - 0.022005 \times \text{第一學期學分數} \\
 & + 0.068593 \times \text{第一學期平均} + 0.509172 \times \text{GPA} \\
 & - 0.223224 \times D_{31} - 0.083149 \times D_{32} \\
 & - 0.009546 \times \text{讀書時間} - 0.028003 \times \text{社團時間} \\
 & + 0.015398 \times \text{運動時間} - 0.011428 \times \text{打工時間} \\
 & + 0.130989D_4 + 0.771407D_{51} - 0.699387D_{52} \\
 & - 0.160591D_{53} - 0.034305 \times \text{上課專注度} + \epsilon
 \end{aligned}$$

其中Y為總平均， ϵ 為隨機誤差項且 $\epsilon \sim N(0, 76.47394)$

學院	D ₁
理學院	0
商學院	1

虛擬變數表(1)

2→6→3→4→5→1→碩1

年級	D ₂₆	D ₂₃	D ₂₄	D ₂₅	D ₂₁	D ₂₇
大二	0	0	0	0	0	0
大六	1	0	0	0	0	0
大三	0	1	0	0	0	0
大四	0	0	1	0	0	0
大五	0	0	0	1	0	0
大一	0	0	0	0	1	0
碩一	0	0	0	0	0	1
碩二	0	0	0	0	0	0

虛擬變數表(2)

沒有碩二的資料，因此也是0

讀書習慣	D ₃₁	D ₃₂
電子講義	0	0
紙本講義	1	0
紙本講義+電子講義	0	1

虛擬變數表(3)

	D ₄
通勤	0
住宿	1

虛擬變數表(4)

就寢情況	D ₅₁	D ₅₂	D ₅₃
熬夜	0	0	0
其他	1	0	0
通宵	0	1	0
早睡	0	0	1

虛擬變數表(5)

IV. 殘差分析

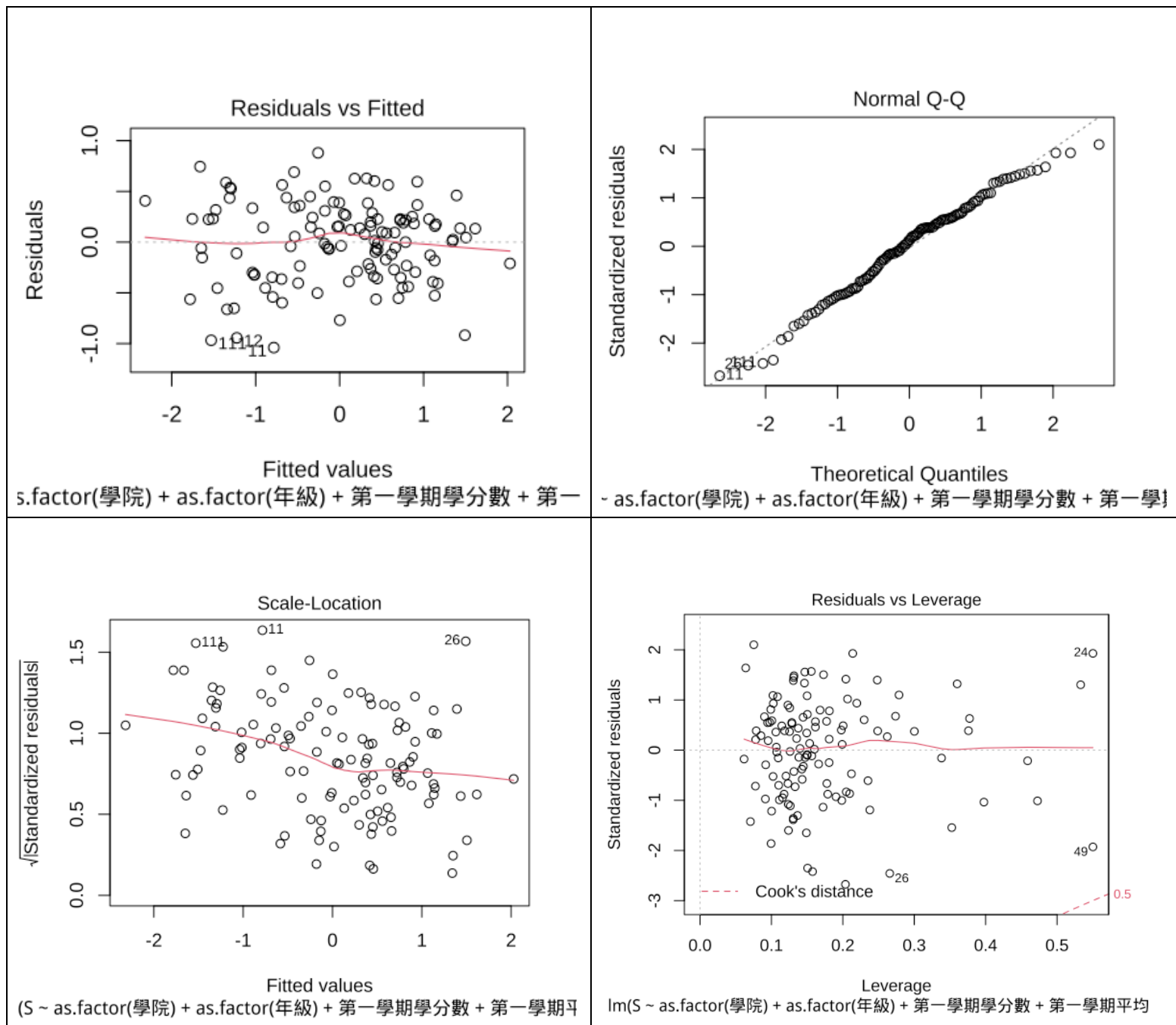


圖 L

左上：非線性（獨立）

左下：同質性（變異數一致）

右上：殘差服從常態分配

右下：無異常值

綜合以上四張圖，可知模型符合迴歸假設

V. 模型結論：

Residual standard error	0.4355 on 99 degrees of freedom
Multiple R- squared	0.8435
Adjusted R-squared	0.8104
F-statistic	25.42 on 21 and 99 DF
p-value	< 2.2e-16

表（十）

使用的自變數很多，會使自由度減少、對 R^2 高估，因此看調整後的 R^2 ，由表（十）可知 $\bar{R}^2 = 0.8104$ 和 $R^2 = 0.8435$ 差異不大，模型使用的自變數沒有太多，解釋力高。

F = 25.42, p-value 很小，代表迴歸的解釋力好。

Analysis of Variance Table					
	SS	MS	Df	F	P
Regression	101.235	4.8207	21	25.417	7.7814e-31
Error	18.777	0.1897	99		
Total	120.012	1.0001	120		

表（十一）

由模型可知：商學院的學生比理學院的總平均高一點；讀大一的人比讀大二的對總平均有一點負向影響，讀大三到大六的人比讀大二的負向影響較大，推測大三的課可能難度是比較高的，所以成績較差；在固定第一學期學分數外的變數下，前一學期每多一學分，總平均會減少 0.022005 分；固定第一學期平均外的變數下，前一學期平均每多一分，總平均會增加 0.068593 分；固定 GPA 外的變數下，GAP 每增加 1 單位，總平均會增加 0.509172 分；習慣用紙本做筆記的人以及習慣用紙本和電子講義的人相對於習慣用電子講義的人總平均比較差一點；固定讀書時間外的變數下，每增加一小時讀書，總平均會減少 0.009546 分；固定社團時間外的變數下，每增加一小時參加社團，總平均會減少 0.028003 分；固定運動時間外的變數下，每增加一小時去運動，總平均會增加 0.015398 分；固定打工時間外的變數下，每增加一小時的打工，總平均會減少 0.011428 分；住宿的人比通勤的總平均高 0.130989 分；在其他時間睡覺的人比熬夜的總平均高 0.771407 分，通宵比熬夜的總平均低 0.699387 分，早睡比熬夜的總平均低 0.160591 分；固定上課專注度外的變數下，每多專注一點，總平均會減少 0.034305 分。

• 總結：

1. 商學院的學生成績表現比理學院好
2. 就讀越高年級成績越差一些，可能是因為課程難度提升，成績會下降
3. 前一學期的學分數、社團及打工時間越多，對成績會有負向的影響
4. 前一學期平均和 GPA 會直接影響總平均，所以一定是正向關係
5. 只用電子講義比其他兩種讀書習慣的人成績較好一點，猜測是因為使用電子產品的效率比較高，做筆記時會比較有系統，節省一些印講義或是找筆記內容等等的時間
6. 上課專注度比較差一點成績未必比較差，猜測可能是跟老師的教學風格差異有關，或是覺得自己看就好，所以沒有很認真上課等等的因素
7. 住宿的人因為少了通勤時間，所以能充分使用的時間相對於通勤生多，因此成績會比較高
8. 在其他時間（7:00~22:00）睡的人比熬夜（0:00~4:00）的成績高，熬夜的人比早睡（22:00~0:00）的成績還要好，可能睡覺時間的總時數比幾點睡來得重要
9. 花一點時間運動有助於提升成績表現

2. 羅吉斯迴歸：分析性別跟有無被當是否有相關，並且利用性別來預測是否有被當。

		是否被當	
		是	否
性別	女	18	52
	男	28	23

性別和是否被當的列聯表(表十二)

I. 模型：

$$\log(\text{odds}) = \log\left(\frac{p(x)}{1-p(x)}\right) = -0.1967 + 1.2576x + \epsilon, \text{其中 } x = \begin{cases} 1, \text{女生} \\ 0, \text{男生} \end{cases}$$

II. 結論：

從表(十四)的 p-value=0.00135 顯著得知解釋變數(性別)的效應是顯著的，表示性別對於有沒有被當有影響，使用OR值 = $\exp(\text{參數估計值})$ 來作勝算比較： $\exp(\text{性別的參數估計值}) = \exp(1.2576) = 3.5169082$ ，男生被當的勝算比大約是女生的 3.52 倍

Deviance Residuals:				
Min	1Q	Median	3Q	Max
-1.648	-1.095	0.771	0.771	1.262

表(十三)

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.1967	0.2814	-0.699	0.48454
gen 女	1.2576	0.3924	3.205	0.00135 **

表(十四)

(Dispersion parameter for binomial family taken to be 1)			
Null deviance:	160.72 on 120 degrees of freedom		
Residual deviance	150.02 on 119 degrees of freedom		
AIC:	154.02		
Number of Fisher Scoring iterations	4		
	OR	2.5 %	97.5 %
(Intercept)	0.8214286	0.4691598	1.423844
gen 女	3.5169082	1.6482387	7.715954

表 (十五)

3-4 ANOVA

- I. 利用 ANOVA 分析不同年級的學生總平均是否無差異，分成一到四年級以上（大四到碩二分成一群）

	人數	中位數	標準差
大一	18	79.78	7.261
大二	17	79.55	7.965
大三	33	76	9.123
大四以上	53	78.9	9.148

表（十六）

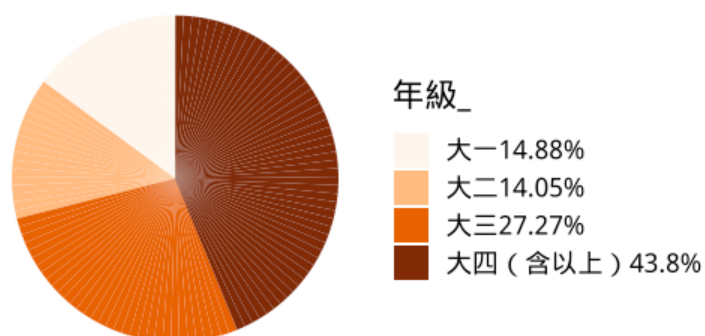


圖 M

一到四年級（含以上）的箱形圖

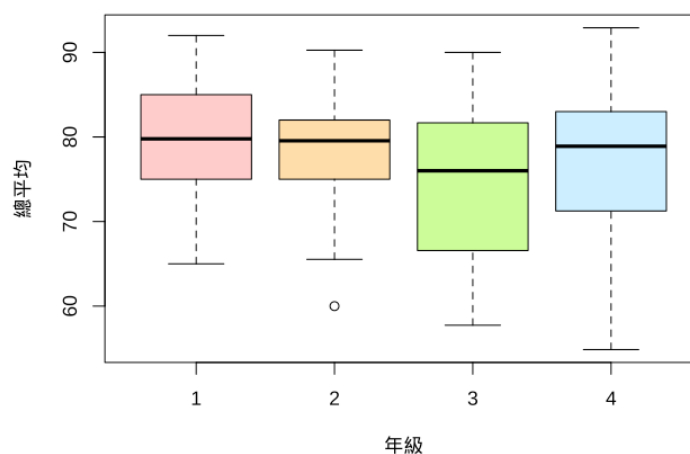


圖 N

II. H_0 : 不同年級的學生總平均無差異 v.s. H_1 : 不同年級的學生總平均有差

III. 檢驗大一到大四（含以上）四群是否互相獨立、變異數相同及是否為常態

i. 不同年級的總平均是否符合常態分配

— 大一

Shapiro-Wilk normality test	
data: 總平均[which(grade1 == 1)]	
W = 0.9594	p-value = 0.5901

表（十七）

— 大二

Shapiro-Wilk normality test	
data: 總平均[which(grade1 == 2)]	
W = 0.95195	p-value = 0.488

表（十八）

— 大三

Shapiro-Wilk normality test	
data: 總平均[which(grade1 == 3)]	
W = 0.95282	p-value = 0.1607

表（十九）

— 大四（含以上）

Shapiro-Wilk normality test	
data: 總平均[which(grade1 == 4)]	
W = 0.95455	p-value = 0.04233

表（二十）

表（十七）到表（十九）中的 p-value 皆大於 0.05，所以大一到大三的分佈為常態，而表（二十）的 p-value=0.04233 不夠顯著，因此大四（含以上）的分佈也是常態。

ii. 檢驗大一到大四（含以上）四群的變異數是否相同

Levene's Test for Homogeneity of Variance (center = median: drop_DATA)			
	Df	F value	Pr(>F)
group	3 117	0.8044	0.4939

表（二十一）

由表（二十一）的 F value =0.8044, p-value=0.4939>0.05 可知，大一到大四（含以上）四群的變異數相同

iii. 檢驗大一到大四（含以上）四群是否互相獨立

lag	Autocorrelation	D-W Statistic	p-value
1	0.001715504	1.986378	0.856
Alternative hypothesis: rho != 0			

表（二十二）

由表（二十二）的 D-W Statistic=1.986378, p-value=0.856>0.05 可知，大一到大四（含以上）四群互相獨立

由 i.~iii.三個檢驗，可確定符合 ANOVA 的定義

IV. 模型： $Y_{ij} = \mu_i + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, 76.47394), i = 1, 2, 3, 4$ （年級） $j =$

各年級的樣本數，其中 μ_i 為第 i 年級的母體平均， ε_{ij} 為第 i 年級第 j 個樣本的誤差效應

Anova Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
grade1	3	251	83.63	1.096	0.354
Residuals	117	8926	76.29		

表（二十三）

V. 結論：由表（二十三）的 F value=1.096, p-value>0.05 不顯著，不拒絕 H_0 ，所以不同年級之間的總平均沒有差異。

3-5 卡方獨立性檢定

- I. 利用獨立性檢定分析讀書習慣和讀書時間的多寡是否有關
- II. 獨立性檢定為類別分析，所以事先把讀書時間依四分位數轉成類別型，分別用第一分位數、中位數、第三分位數分成"幾乎無"，"少"，"中等"，"多"四類。

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	2	8	8.223	10	45

表(二十四)

- III. H_0 : 讀書習慣和讀書時間的多寡無關 v.s. H_1 : 讀書習慣和讀書時間的多寡有關

		study			
		幾乎無	少	中等	多
讀書習慣	紙本講義(課本、老師自編講義、手寫筆記...)	14	19	18	9
	電子講義(平版、電腦...)	5	7	3	3
	紙本講義+電子講義 並用	13	10	5	15

讀書習慣和讀書時間列聯表(表二十五)

- IV. 因表(二十五)的格子內樣本數有小於5的，要改用費雪精確獨立性檢定。

費雪精確獨立性檢定	
data	study.table
p-value	0.1128
alternative hypothesis	two.sided

表(二十六)

- V. 結論：由表(二十六)的 $p\text{-value} = 0.1128 > 0.05$ ，得知不拒絕 H_0 ，表示讀書習慣和讀書時間的多跟少無關。

3-6 對整體的結論

1. 以淡江大學理學院和商學院的學生來說，商學院學生的成績會比理學院的來得好，且不同年級的總平均差異不大，但男生比女生被當的機率較高
2. 對於成績表現
 - I. 每週花一些時間運動以及在經濟許可下，通勤時間很多，會影響到時間分配的話，住宿有助於提升成績
 - II. 應該避免修太多課、花太多時間在社團或打工上，會使成績變差
 - III. 男生被當的勝算比是女生的 3.5 倍，所以男生對於課業要多用心

第四章 附錄

4-1 問卷網址

https://docs.google.com/forms/d/1JnkZfJdg3dQrALWsXUH1LLqj7EHa0dJ5LH1eJ1PP_V4/viewform?hl=zh-tw&hl=zh-tw&edit_requested=true

4-2 問卷網址所有步驟的 R 程式碼

<程式碼>

```
#install.packages("openxlsx")
library(openxlsx)
#install.packages("readxl")
library(readxl)

#install.packages("dplyr")
library(dplyr)

#DATA=read_excel("/Users/chaofangyu/Desktop/R/期末報告/20220529
原.xlsx")
DATA=read_excel("/Users/paulawu/PAULAPAULA/R 期末報告/20220605 誰是
時間管理大師.xlsx")
DATA=DATA[-1] ;DATA
colnames(DATA)=c('性別','學院','年級','第一學期學分數','第一學期平均','總平
均','GPA',
                 '讀書習慣','是否被當','睡覺時間','讀書時間','滑手機時間','娛
樂時間',
                 '性別_1','社團時間','運動時間','打工時間','幾點睡覺','是否通
勤',
                 '交作業時間','性別_2','上課專注度')
x=c('女','男')
DATA$性別_1=x[as.numeric(DATA$性別)+1]
#y=ifelse(DATA$性別=='女',0.0,1.0)
y=ifelse(DATA$性別=='女',0,1)

#改外星人那筆
```

```

DATA[ which(DATA$性別=='外星人'),'性別']='男'

#改每天幾小時那筆
DATA[ which(DATA$滑手機時間=='每天至少 10 小時'),'滑手機時間']='10'
DATA[ which(DATA$娛樂時間=='每天 5 小時'),'娛樂時間']='5'

#改 5-6 那筆
DATA[ which(DATA$睡覺時間=='5-6'),'睡覺時間']='5.5'
DATA[ which(DATA$讀書時間=='5-6'),'讀書時間']='5.5'

#改有寫單位的
DATA[ which(DATA$睡覺時間=='9 小時'),'睡覺時間']='9'
DATA[ which(DATA$讀書時間=='15 小時'),'讀書時間']='15'
DATA[ which(DATA$滑手機時間=='85 小時'),'滑手機時間']='85'
DATA[ which(DATA$娛樂時間=='2 小時'),'娛樂時間']='2'
DATA[ which(DATA$運動時間=='3 小時'),'運動時間']='3'
DATA[ which(DATA$第一學期學分數=='16 學分'),'第一學期學分數']='16'


DATA$GPA=as.numeric(DATA$GPA)
DATA$第一學期學分數=as.numeric(DATA$第一學期學分數)
DATA$第一學期平均=as.numeric(DATA$第一學期平均)
DATA$總平均=as.numeric(DATA$總平均)


DATA$睡覺時間=as.numeric(DATA$睡覺時間)
DATA$讀書時間=as.numeric(DATA$讀書時間)
DATA$滑手機時間=as.numeric(DATA$滑手機時間)
DATA$娛樂時間=as.numeric(DATA$娛樂時間)


#無轉成 0
DATA[ which(DATA$社團時間=='無'),'社團時間']='0.0'
DATA[ which(DATA$運動時間=='無'),'運動時間']='0.0'
DATA[ which(DATA$打工時間=='無'),'打工時間']='0.0'


DATA$社團時間=as.numeric(DATA$社團時間)
DATA$運動時間=as.numeric(DATA$運動時間)
DATA$打工時間=as.numeric(DATA$打工時間)

```

```

#上床睡覺時間
DATA$幾點睡覺=sub("1899-12-31", "", DATA$幾點睡覺)

#(DATA2=>159 筆)
DATA2=DATA

###(new_DATA=>146 筆)
new_DATA=DATA2 %>%
  filter(性別==性別_1,性別_2==y,第一學期學分數<=30 ,
         !is.na(GPA) & GPA<=4 ,!is.na(睡覺時間),
         !is.na(讀書時間),!is.na(滑手機時間),
         !is.na(娛樂時間),!is.na(社團時間) ,
         !is.na(運動時間),!is.na(打工時間) ) %>%
  select(-性別_1,-性別_2) ;new_DATA

#write.xlsx(new_DATA , file="/Users/chaofangyu/Desktop/R/期末報告
/20220515 篩.xlsx")

#篩出女生 20/85/80/3 =>有 13 筆
g1=filter(new_DATA , new_DATA$性別=="女" , new_DATA$第一學期學分數
==20 ,
          new_DATA$第一學期平均==85 , new_DATA$總平均==80 ,
new_DATA$GPA==3) ;g1
G1=which(new_DATA$性別=="女" & new_DATA$第一學期學分數==20 &
new_DATA$第一學期平均==85
        & new_DATA$總平均==80 & new_DATA$GPA==3) ;G1

#篩出 85/82/3 (都女生) =>有 7 筆
g2=filter(new_DATA , new_DATA$第一學期平均==85 & new_DATA$總平均
==82
        & new_DATA$GPA==3 ) ;g2
G2=which(new_DATA$第一學期平均==85 & new_DATA$總平均==82 &
new_DATA$GPA==3) ;G2

```

```

#刪除女生 20/85/80/3 和 85/82/3 =>刪除 20 筆(剩 126 筆)
drop_DATA=new_DATA[-c(G1,G2) ,]

#write.xlsx(drop_DATA , file="/Users/chaofangyu/Desktop/R/期末報告
/20220517drop.xlsx")

#--
# plot--中文
#install.packages("showtext")
library(showtext)
showtext_auto(enable = TRUE)
font_add('新細明體')
par(family='STKaiti')

#---
# 刪掉總平均的 outlier
# 用合鬚圖找 outlier

#1.原始的 drop_DATA(126 筆)
boxplot(drop_DATA[, 6], xlab = "總平均" , col="#E8CCFF" , main="原始的
drop_DATA")
summary(drop_DATA[, 6])

#---
# 離群值的上下界
outlier_more <- 82.3 + 1.5*(82.3-70); outlier_more ##100.75
outlier_less <- 70 - 1.5*(82.3-70); outlier_less ##51.55

A = which(new_DATA$總平均<51.55); A ##位置 6 10 39 138
B = which(new_DATA$第一學期平均<40);B ##位置 39 41

#2.新的 drop_DATA:刪掉離群值 =>剩 121 筆
drop_DATA=new_DATA[-c(G1,G2, A, B) ,]

```

```

boxplot(drop_DATA[, 6], xlab = "總平均", col="#CCDDFF", main="刪除
離群值後的 drop_DATA")
summary(drop_DATA[,6])

# outlier_more1 <- 82+1.5*(82-72.94); outlier_more1
# outlier_less1 <- 70-1.5*(82-72.94); outlier_less1
# B = which(new_DATA$總平均<51.55); B
#
# drop_DATA=new_DATA[-B ,]

table(drop_DATA$性別, drop_DATA$學院)
##   商學院 理學院
##女      52      18
##男      30      21

#最後刪完的資料 (=>剩 121 筆)
write.xlsx(drop_DATA, file="/Users/paulawu/PAULAPAULA/R 期末報告
/20220605 誰是時間管理大師 drop.xlsx")

#####
# 分析 #
#####

B_DATA=filter(drop_DATA, drop_DATA$學院=="商學院")
S_DATA=filter(drop_DATA, drop_DATA$學院=="理學院")

summary(B_DATA)
summary(S_DATA)

#####
#簡化變數名稱
性別=drop_DATA$性別
# 性別=ifelse(性別=="男", 1, 0);性別
學院=drop_DATA$學院
年級=drop_DATA$年級
年級=factor(年級, levels=c("大一", "大二", "大三", "大四", "大五", "大
六", "碩一", "碩二"))

```

```

GPA=drop_DATA$GPA
第一學期學分數=drop_DATA$第一學期學分數
第一學期平均=drop_DATA$第一學期平均
總平均=drop_DATA$總平均
上課專注度=drop_DATA$上課專注度

讀書習慣=drop_DATA$讀書習慣
被當=drop_DATA$是否被當
通勤=drop_DATA$是否通勤
交作業=drop_DATA$交作業時間
幾點睡覺=drop_DATA$幾點睡覺
#上床睡覺時間
#DATA$幾點睡覺=sub("1899-12-31", "", DATA$幾點睡覺)
s1=as.numeric(substr( 幾點睡覺 , 2 , 3 )) ;s1
s2=as.numeric(substr( 幾點睡覺 , 5 , 6 )) ;s2
SS=ifelse(( s1%in%c(22,23) & s2%in%c(0:59) )|(s1==0 & s2==0) , 1 ,
  ifelse(( s1==0 & s2%in%c(1:59) )|( s1%in%c(1:4) &
s2%in%c(0:59) ) , 2 ,
  ifelse(( s1==4 & s2%in%c(1:59) )|( s1%in%c(5:7) &
s2%in%c(0:59) ) , 3 , 4))) ;SS
sleep=ifelse(SS==1 , "早睡" , ifelse(SS==2 , "熬夜" , ifelse(SS==3 ,
"通宵" , "其他")))) ;sleep

#cbind(幾點睡覺 , SS , sleep)#檢查

睡覺時間=drop_DATA$睡覺時間
讀書時間=drop_DATA$讀書時間
滑手機時間=drop_DATA$滑手機時間
娛樂時間=drop_DATA$娛樂時間
社團時間=drop_DATA$社團時間
運動時間=drop_DATA$運動時間
打工時間=drop_DATA$打工時間

aggregate(cbind(GPA , 第一學期學分數 , 第一學期平均 , 總平均) ~ 性

```



```

別 , data=drop_DATA , FUN=mean)
aggregate(cbind(GPA , 第一學期學分數 , 第一學期平均 , 總平均) ~ 學院 , data=drop_DATA , FUN=mean)
aggregate(cbind(GPA , 第一學期學分數 , 第一學期平均 , 總平均) ~ 年級 , data=drop_DATA , FUN=mean)

gap = 1*(總平均<=10)+2*(總平均>10&總平均<=20)+3*(總平均>20&總平均<=30)+4*(總平均>30&總平均<=40)+
      5*(總平均>=40&總平均<50)+6*(總平均>=50&總平均<60)+7*(總平均>=60&總平均<70)+
      8*(總平均>=70&總平均<80)+9*(總平均>=80&總平均<90)+10*(總平均>=90);gap

gap = ifelse(gap == "1", "0~10", ifelse(gap == "2", "11~20",
ifelse(gap == "3", "21~30",
      ifelse(gap == "4", "31~40", ifelse(gap == "5", "41~50",
ifelse(gap == "6", "51~60",
      ifelse(gap == "7", "61~70", ifelse(gap == "8", "71~80",
ifelse(gap == "9", "81~90", "91~100"))))))))

table(學院, gap)

#---

# 1.兩母體平均差

# 先檢定是否服從常態分配
## 在本報告中的分析方法所有  $\alpha$  皆設定為 0.05

#1.商學院
shapiro.test(B_DATA$總平均)
# p-value = 0.02326 沒有很小 算是常態
hist(B_DATA$總平均 , ylim=c(0,30) , col="#9999FF" , main="商學院的直方圖")
qqnorm(B_DATA$總平均 , col="#5500DD" , main="商學院 QQ-plot")
qqline(B_DATA$總平均, col = "#FF8800" , lwd=2)

```

```

#2.理學院
shapiro.test(S_DATA$總平均)
# p-value = 0.2003 > 0.05 是常態

#install.packages("BSDA")
library(BSDA)
z.test(x = S_DATA$總平均, y = B_DATA$總平均, alternative =
"two.sided", mu = 0,
      sigma.x = sd(B_DATA$總平均), sigma.y = sd(S_DATA$總平均),
conf.level = 0.95)

#---

# 2.Regression
set.seed(12)

#install.packages("psych")
library(psych)
pairs.panels(drop_DATA[ , c("GPA","第一學期學分數","第一學期平均","總
平均","上課專注度")])

# 先檢驗總平均是不是常態
hist(x = 總平均 , ylim=c(0,0.08) , freq = FALSE ,col="#CCDDFF" ,
main="總平均直方圖")
lines(density(總平均), col = "blue" , lwd=2)

qqnorm(總平均 , col="#000088" , main="總平均 QQ-plot")
qqline(總平均, col = "#FF8800" , lwd=2)

#install.packages("moments")
library(moments) # 偏度鋒度計算
skewness(總平均)
# -0.5036863 左偏

#---

# 標準化轉成常態分配的樣子 ( 不會改變原先的分配 所以偏度會一樣 )
S <- round(scale(總平均, mean(總平均), sd(總平均)), 3); S
hist(S , col="#E8CCFF" , main="對總平均標準化後的直方圖")

```

```

qqnorm(S , col="#66009D" , main="對總平均標準化後的 QQ-plot")
qqline(S, col = "#FF8800" , lwd=2)

skewness(S)
summary(S); sd(S)

modell_S <- lm(S ~ as.factor(性別) + as.factor(學院) + as.factor(年級) + 第一學期學分數 +
               第一學期平均 + GPA + as.factor(讀書習慣) +
as.factor(被當) +
               睡覺時間 + 讀書時間 + 滑手機時間 + 娛樂時間 + 社團時間 + 運動時間+
               打工時間 + as.factor(通勤)+ as.factor(sleep) +
as.factor(交作業) + 上課專注度, data = drop_DATA)
summary(modell_S)
anova(modell_S)

# 檢驗共線性
round(car::vif(modell_S), 3)
# df 非 1 要看  $GVIF^{1/(2*Df)}$  的平方是否小於 5

# 雙向
step(modell_S, direction = "both")

# BIC 最小的模型
modell3 <- lm(formula = S ~ as.factor(學院) + as.factor(年級) + 第一學期學分數 +
              第一學期平均 + GPA + as.factor(讀書習慣) + 讀書時間
+
              社團時間 + 運動時間 + 打工時間 + as.factor(通勤) +
as.factor(sleep) + 上課專注度, data = drop_DATA)
summary(modell3)
anova(modell3)

# 畫 ANOVA table
anova_alt = function (object, reg_collapse=TRUE,...)
{

```

```

if (length(list(object, ...)) > 1L)
  return(anova.lm(list(object, ...))
if (!inherits(object, "lm"))
  warning("calling anova.lm(<fake-lm-object>) ...")
w <- object$weights
ssr <- sum(if (is.null(w)) object$residuals^2 else w *
object$residuals^2)
mss <- sum(if (is.null(w)) object$fitted.values^2 else w *
            object$fitted.values^2)
if (ssr < 1e-10 * mss)
  warning("ANOVA F-tests on an essentially perfect fit are
unreliable")
dfr <- df.residual(object)
p <- object$rank
if (p > 0L) {
  pl <- 1L:p
  comp <- object$effects[pl]
  asgn <- object$assign[stats::qr.lm(object)$pivot][pl]
  nmeffects <- c("(Intercept)", attr(object$terms, "term.labels"))
  tlabels <- nmeffects[1 + unique(asgn)]
  ss <- c(vapply(split(comp^2, asgn), sum, 1), ssr)
  df <- c(lengths(split(asgn, asgn)), dfr)
  if(reg_collapse){
    if(attr(object$terms, "intercept")){
      collapse_p<-2:(length(ss)-1)
      ss<-c(ss[1],sum(ss[collapse_p]),ss[length(ss)])
      df<-c(df[1],sum(df[collapse_p]),df[length(df)])
      tlabels<-c(tlabels[1],"Source")
    } else {
      collapse_p<-1:(length(ss)-1)
      ss<-c(sum(ss[collapse_p]),ss[length(ss)])
      df<-c(df[1],sum(df[collapse_p]),df[length(df)])
      tlabels<-c("Regression")
    }
  }
} else {
  ss <- ssr
  df <- dfr

```

```

tlabels <- character()
if(reg_collapse){
  collapse_p<-1:(length(ss)-1)
  ss<-c(sum(ss[collapse_p]),ss[length(ss)])
  df<-c(df[1],sum(df[collapse_p]),df[length(df)])
}
}

ms <- ss/df
f <- ms/(ssr/dfr)
P <- pf(f, df, dfr, lower.tail = FALSE)
table <- data.frame(df, ss, ms, f, P)
table <- rbind(table,
               colSums(table))
if (attr(object$terms, "intercept")){
  table$ss[nrow(table)]<- table$ss[nrow(table)] - table$ss[1]
}
table$ms[nrow(table)]<-
table$ss[nrow(table)]/table$df[nrow(table)]
table[length(P):(length(P)+1), 4:5] <- NA
dimnames(table) <- list(c(tlabels, "Error","Total"),
                       c("Df","SS", "MS", "F",
                          "P"))
if (attr(object$terms, "intercept")){
  table <- table[-1, ]
  table$MS[nrow(table)]<-
table$MS[nrow(table)]*(table$Df[nrow(table)]/(table$Df[nrow(table)]-1))
  table$Df[nrow(table)]<-table$Df[nrow(table)]-1
}
structure(table, heading = c("Analysis of Variance Table\n"),
          class = c("anova", "data.frame"))
}
anova_alt(modell3)

# 殘差診斷
plot(modell3)

```

```

#---

# 3.logistic
# 性別跟有無被當是否有相關
# 1 -> 女, 0 -> 男
gen = as.factor(性別)
fail = as.factor(被當)
fail = ordered(fail, levels = c("是", "否"))
table(gen, fail)

model = glm(fail ~ gen, family = binomial(link = logit), data =
drop_DATA)
summary(model)

exp(coef(model))
exp(cbind(OR = coef(model), confint(model)))

#---

# 4.ANOVA
# 檢定不同年級的學生總平均是否相同
grade = ifelse(年級 == "大一", 1, ifelse(年級 == "大二", 2,
ifelse(年級 == "大三", 3, 4)))
grade1 = as.factor(grade)
#grade = ordered(grade, levels = c(1,2,3,4))
shapiro.test(總平均[which(grade1==1)])
shapiro.test(總平均[which(grade1==2)])
shapiro.test(總平均[which(grade1==3)])
shapiro.test(總平均[which(grade1==4)])

# 齊一性
leveneTest(總平均, grade1, data = drop_DATA)

one.way <- aov(總平均 ~ grade1, data = drop_DATA)

# 獨立性檢定
durbinWatsonTest(one.way)

```

```

summary(one.way)

table(gap, grade)
人數 = tapply(grade, grade1, length); 人數
中位數 <- tapply(總平均, grade1, median); 中位數
標準差 <- tapply(總平均, grade1, sd); 標準差
tab = data.frame(人數, 中位數, 標準差)
tab = round(tab, 3); tab

boxplot(總平均 ~ grade, xlab="年級", col=c("#FFCCCC", "#FFDDAA",
"#CCFF99", "#CCEEFF"), main="一到四年級(含以上)的箱型圖")

#---

# 5.chi-square
# 讀書習慣和讀書時間的多寡是否有關
summary(讀書時間)
stu = 1*(讀書時間<=2) + 2*(讀書時間>2 & 讀書時間<=8) + 3*(讀書時間
>8 & 讀書時間<=10) + 4*(讀書時間>10)
study = ifelse(stu=="1", "幾乎無", ifelse(stu=="2", "少",
ifelse(stu=="3", "中等", "多"))) # 讀書時間轉成類別
study = as.factor(study)
study = ordered(study, levels = c("幾乎無", "少", "中等", "多"))
讀書習慣 = as.factor(讀書習慣)
讀書習慣 = ordered(讀書習慣, levels = c("紙本講義(課本、老師自編講
義、手寫筆記...)", "電子講義(平版、電腦...)", "紙本講義+電子講義 並
用"))
study.table = table(讀書習慣, study); study.table
#prop.table(study.table,margin=1)
#prop.table(study.table,margin=2)

chisq.test(study.table)

# 有格子樣本數小於 5，改費雪檢定
fisher.test(study.table)
#p-value = 0.5079 > 0.05

```

```

#####
# 圖形 #
#####

library(colorspace)
library(ggplot2)

# 男女比
gender = as.factor(性別)
tapply(gender, gender, length)
levels(gender)
gender1 = c(51, 70)
group = c("男", "女")
gender2 = data.frame(group, gender1); gender2
attach(gender2)
gender3 = sum(gender2$gender1); gender3
男女比 <- gender2$gender1*100/gender3 ;男女比

ggplot(drop_DATA, aes(x="", y="性別比", fill=性別)) +
  geom_bar(stat="identity", width=1)+
  coord_polar(theta = "y")+
  theme_void() +
  theme(text = element_text(family = "Heiti TC Light", size =
30))+
  scale_fill_discrete_sequential(palette = "Purples 3",
                                nmax = 6,
                                rev = FALSE,
                                order = 3:5 ,
                                labels = paste0(group, round(男女
比,2), "%"))

# 學院比

```



```

major = as.factor(學院)
tapply(major, major, length)
levels(major)
major1 = c(39, 82)
group1 = c("理學院", "商學院")
major2 = data.frame(group1, major1); major2
attach(major2)
major3 = sum(major2$major1); major3
學院比 <- major2$major1*100/major3 ;學院比

ggplot(drop_DATA, aes(x="", y="學院比", fill=學院)) +
  geom_bar(stat="identity", width=1)+
  coord_polar(theta = "y")+
  theme_void() +
  theme(text = element_text(family = "Heiti TC Light", size =
25))+
  scale_fill_discrete_sequential(palette = "Blues 3",
                                nmax = 6,
                                rev = FALSE,
                                order = 2:4,
                                labels = paste0(group1, round(學院
比,2), "%"))

# 年級比(1~4)
年級_ = ifelse(年級 == "大一", 1, ifelse(年級 == "大二", 2,
ifelse(年級 == "大三", 3, 4)))
年級_ = as.factor(年級_)

tapply(年級_, 年級_, length)
levels(年級_)
grade_1 = c(18, 17, 33, 53)
group3 = c("大一", "大二", "大三", "大四 ( 含以上 ) ")
grade_2 = data.frame(group3, grade_1); grade_2
attach(grade_2)
grade_3 = sum(grade_2$grade_1); grade_3
年級比 <- grade_2$grade_1*100/grade_3
ggplot(drop_DATA, aes(x="", y="年級比", fill=年級_ )) +

```

```

geom_bar(stat="identity", width=1)+
coord_polar(theta = "y")+
theme_void() +
theme(text = element_text(family = "Heiti TC Light" , size =
15))+
scale_fill_discrete_sequential(palette = "Oranges" ,
                                rev = TRUE,
                                labels = paste0(group3, round(年級
比,2), "%"))

#年級比
#年級=drop_DATA$年級
#年級=factor(年級 , levels=c("大一", "大二", "大三", "大四","大五","大
六","碩一","碩二"))
tapply(年級, 年級, length)
levels(年級)
grade_4 = c(18, 17, 33, 46, 1, 2, 4, 0)
group4 = c("大一", "大二", "大三", "大四","大五","大六","碩一","碩二")
grade_5 = data.frame(group4, grade_4); grade_5
attach(grade_5)
grade_6 = sum(grade_5$grade_4); grade_6
年級比 2 <- grade_5$grade_4*100/grade_6

ggplot(drop_DATA, aes(x="", y="年級比", fill=年級 )) +
  geom_bar(stat="identity", width=1)+
  coord_polar(theta = "y")+
  theme_void() +
  theme(text = element_text(family = "Heiti TC Light" , size =
15))+
  scale_fill_discrete_sequential(palette = "Oranges" ,
                                rev = TRUE,
                                labels = paste0(group4, round(年級
比 2,2), "%"))

# 成績分佈
hist(x = 總平均 , ylim=c(0,0.06) , freq = FALSE ,col="#CCDDFF" ,

```

```
main="總平均直方圖")
lines(density(總平均), col = "blue" , lwd=2)

hist(x = GPA, freq = FALSE , col="#BBFFEE" , main="GPA 直方圖")

# 散佈圖
ggplot(drop_DATA)+
  geom_point(aes(x = GPA, y = 總平均, color = 學院)) +
  theme(text = element_text(family = "Heiti TC Light"))
#scale_color_brewer(palette ="Pastel1")
```