

日期: 2024/04/16(二) 上機考: Open Book
授課教師: 吳漢銘 (政治大學統計學系副教授)

請仔細閱讀每一個注意事項 (禁止討論)

1. 考試期間 (第一階段: 04/16 (二) 13:10-15:50, 於上課教室。)
 - (a) 可參考課本、上課講義 (包含電子檔) 及其它資料, 但不能與別人討論。
 - (b) 可使用計算機、自己的筆記型電腦、平板電腦和手機, 但不能使用通訊 APP。
 - (c) 全程可上網查詢, 但不能用通訊軟體 (例如: FB/LINE) 討論, 也不可抄襲網路上之程式碼, 不可使用 chatGPT 之類的工具。
 - (d) 有問題者, 請舉手發問。勿與同學交談。不按照規定作答者, 酌量扣分。
 - (e) 不可使用它人之隨身碟。『作弊』或『疑似作弊』, 往後各項考試不予評分。
2. 下載題目卷, 上傳答題檔案:
 - (a) 於課程網站下載題目卷, 請使用樣版檔作答:「學號-Consulting-MidtermExam.tex 學號-Consulting-MidtermExam.R」, 「學號」改成自己學號。
 - (b) 上傳答題檔案: 於教師網站首頁登入 [作業考試上傳區], 帳號: consulting112。密碼: xxx (上課教室號碼)。
 - (c) 請上傳「學號-Consulting-MidtermExam.tex、學號-Consulting-MidtermExam.pdf、學號-Consulting-MidtermExam.R」。(資料夾: 「MidtermExam」)。
3. 答題檔案原則:
 - (a) 作答時請隨時存檔, 檔案毀損、變空白或消失, 請自行負責。
 - (b) 若上傳檔案格式錯誤, 內容亂碼, 空檔等等問題。請自行負責。
 - (c) 依公平性原則, 檔案不得上傳兩次以上。若上傳兩次 (含) 以上 (以第一次上傳檔案計分)、格式不合等等酌量扣分。
 - (d) 如果上傳網站出現「空白頁」, 請將滑鼠移至「網址列」後, 按「Enter」即可。若再不行, 請換 (Edge/Firefox/Chrome)。
4. 完成考試: 上傳完畢, 請通知教師確認。
5. 第二階段: 至 04/21 (日) 24:00 止, Take Home。
 - (a) 同一份考題, 自行找時間作答再上傳。(註: 不是只做第一階段沒做的題目, 要全做。)
 - (b) Open Book, 可使用網路查尋。不可問人, 不可使用 chatGPT 等 AI 工具, 一定要親自做。
 - (c) 最後得分 = 第一階段 * 0.6 + 第二階段 * 0.4。(註: 各階段分數若超過 100, 以 100 計再計算最後得分)。(總分: 150 分)

我已經仔細閱讀上述各注意事項, 若有違背, 會自行負責。

日期: 2024/04/16(二) 上機考: Open Book
授課教師: 吳漢銘 (政治大學統計學系副教授)

1 L^AT_EX

1.1 以 L^AT_EX 打出一模一樣的文字及數學式 (10 分)

Student's t -distribution has the probability density function (PDF) given by

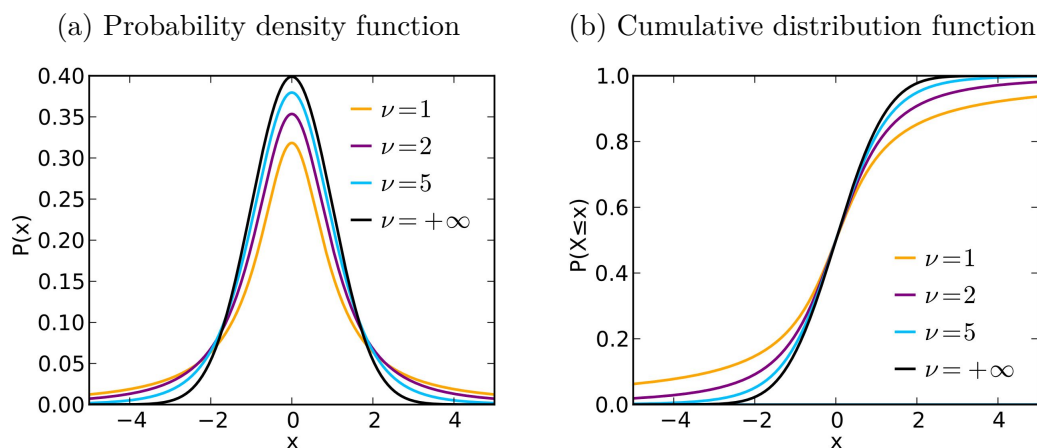
$$p(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < x < \infty,$$

where ν is the number of degrees of freedom and Γ is the gamma function. The corresponding cumulative distribution function (CDF) is

$$F(x) = \frac{1}{2} + x \Gamma\left(\frac{\nu+1}{2}\right) \times \frac{{}_2H_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)},$$

where ${}_2H_1(\cdot; \cdot; \cdot)$ is the hypergeometric function.

1.2 以 R 程式實作上述 PDF 和 CDF 公式 (不得直接用 dt, pt 指令), 並畫出以下兩圖 (儘量一模一樣) (20 分)

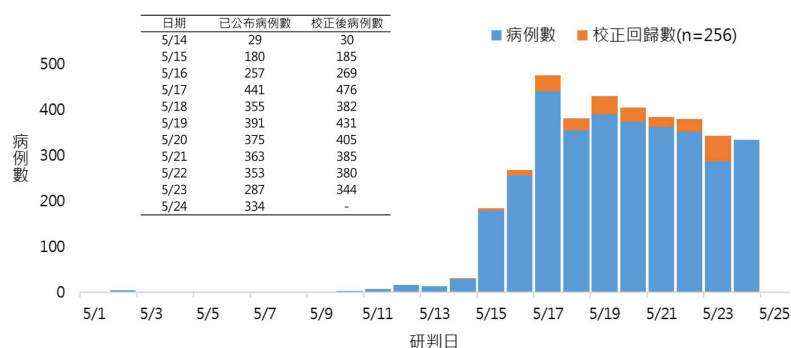


Student's t distribution.

2 ggplot2

下圖是 2021 年台灣中央流行疫情指揮中心發佈的「5/1-5/23 COVID-19 本土確定病例校正回歸情形」長條圖，相關資訊可由衛生福利部疾病管制署<https://www.cdc.gov.tw>查詢得知。

5/1-5/23 COVID-19 本土確定病例校正回歸情形



- **回歸病例定義：**個案採檢通報後，因檢體運送、實驗室量能、無法即時登打檢驗結果等因素，未於2日內完成檢驗結果報告之個案。
- **研判基準：**確診個案之採檢日至確診日(陽性檢驗結果通知)超過2日以上。

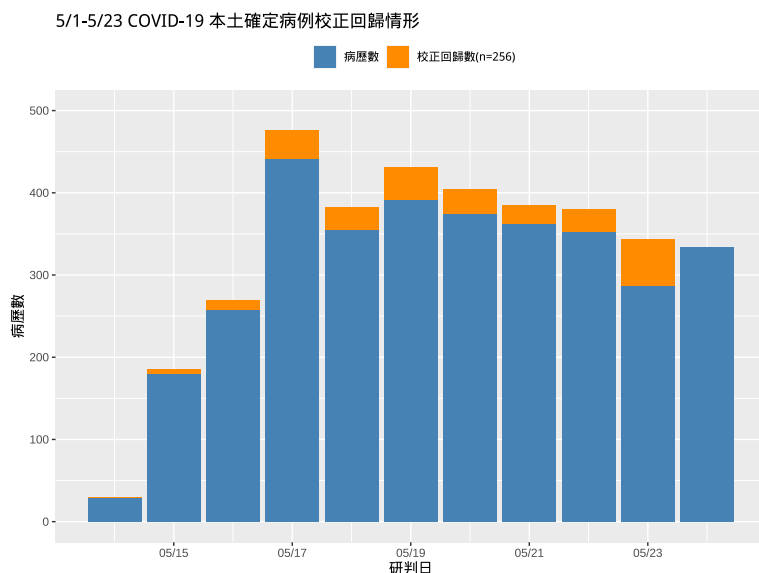
中央流行疫情指揮中心

2021/05/24 更新版

2.1 tidy data (10 分)

依據上圖之表格資料，製造出一個便於使用 ggplot2 繪圖的資料框 (data.frame)(命名 covid19_data_tidy)，並印出。

2.2 以 ggplot2 套件畫出下圖 (10 分)



3 EDA

有一統計諮詢案例，委託人給了兩份量表 (CBRS.csv, MBRS.csv)，內容是針對一群學生做調查 (五點李克特氏量表))，學生有分兩組。今想了解兩組學生在各變項下的平均得分是否有顯著差異。

3.1 使用 psych 套件計算基本統計量 (5 分)

Descriptive statistics

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
問項C1	1	33	4.03	1.16	4	4.26	1.48	1	5	4	-1.46	1.50	0.20
問項C2	2	33	2.73	1.42	3	2.67	1.48	1	5	4	0.09	-1.40	0.25
問項C3	3	33	4.00	1.20	4	4.19	1.48	1	5	4	-1.05	0.14	0.21

Descriptive statistics by group

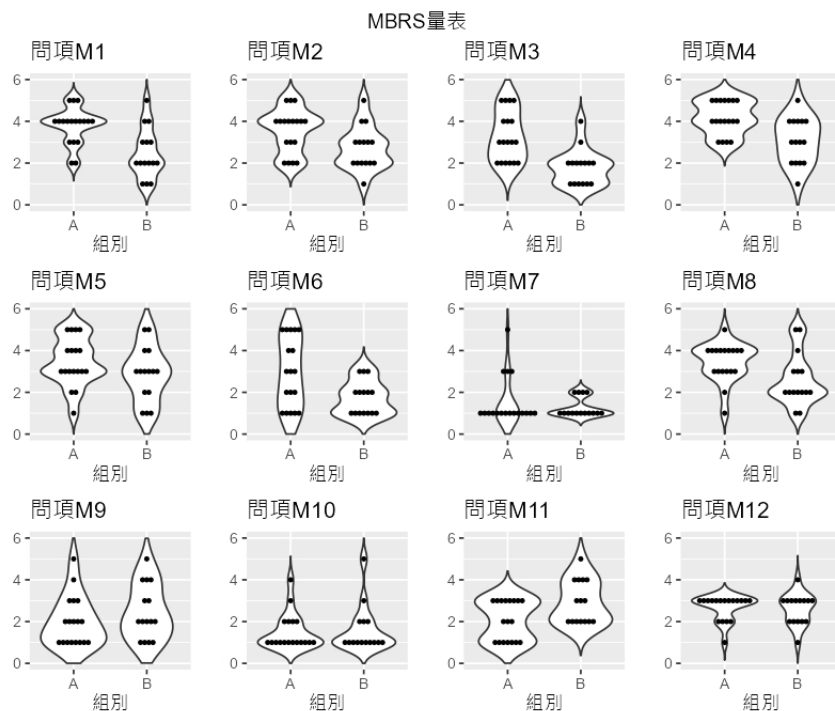
group: A

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
問項C1	1	18	4.50	0.62	5.0	4.56	0.00	3	5	2	-0.70	-0.67	0.15
問項C2	2	18	3.11	1.45	3.5	3.12	1.48	1	5	4	-0.29	-1.44	0.34
問項C3	3	18	4.44	0.86	5.0	4.50	0.00	3	5	2	-0.90	-1.07	0.20

group: B

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
問項C1	1	15	3.47	1.41	4	3.54	1.48	1	5	4	-0.80	-0.82	0.36
問項C2	2	15	2.27	1.28	2	2.15	1.48	1	5	4	0.49	-0.97	0.33
問項C3	3	15	3.47	1.36	4	3.54	1.48	1	5	4	-0.66	-1.02	0.35

3.2 以 ggplot2 套件畫出下圖 (15 分)



4 Heatmap

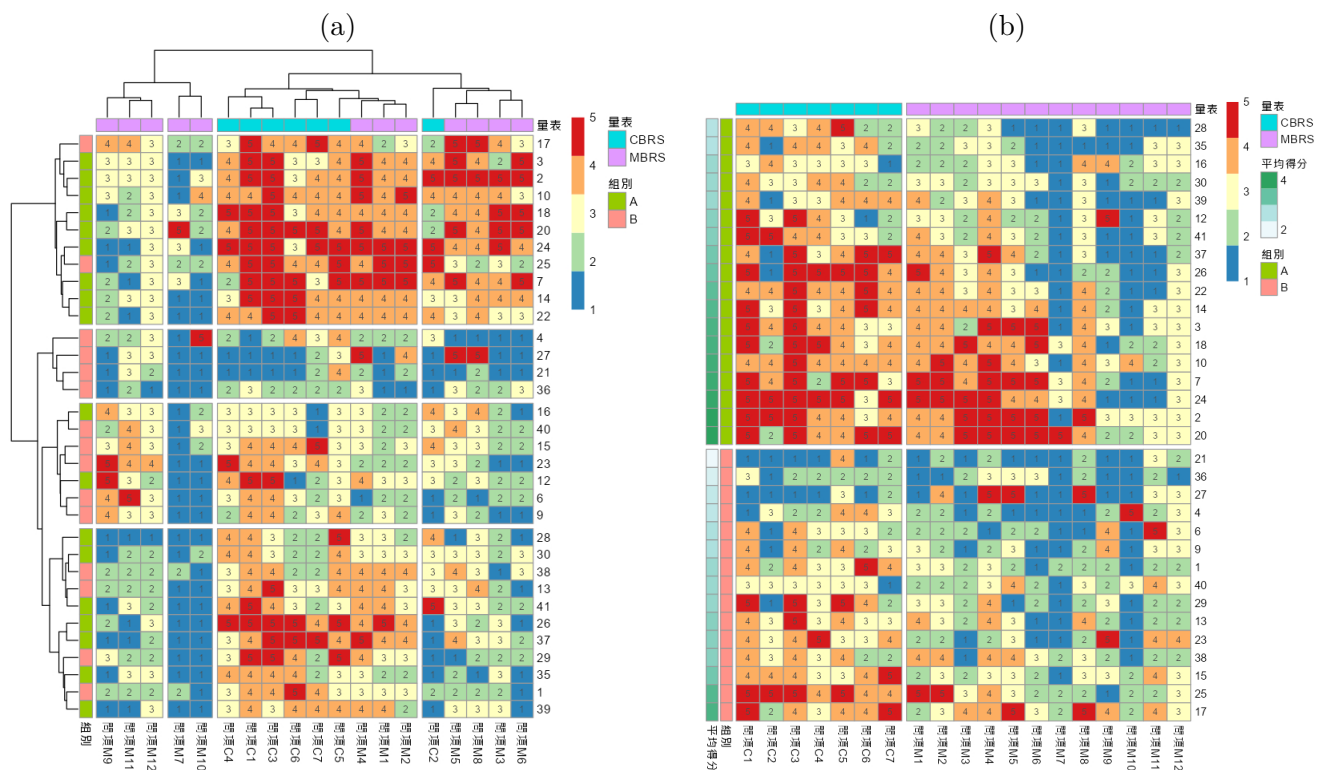
資料同上。

4.1 以 pheatmap 套件畫出下圖 (a)(10 分)

(提示: clustering_method = "ward.D2")

4.2 以 pheatmap 套件畫出下圖 (b)(20 分)

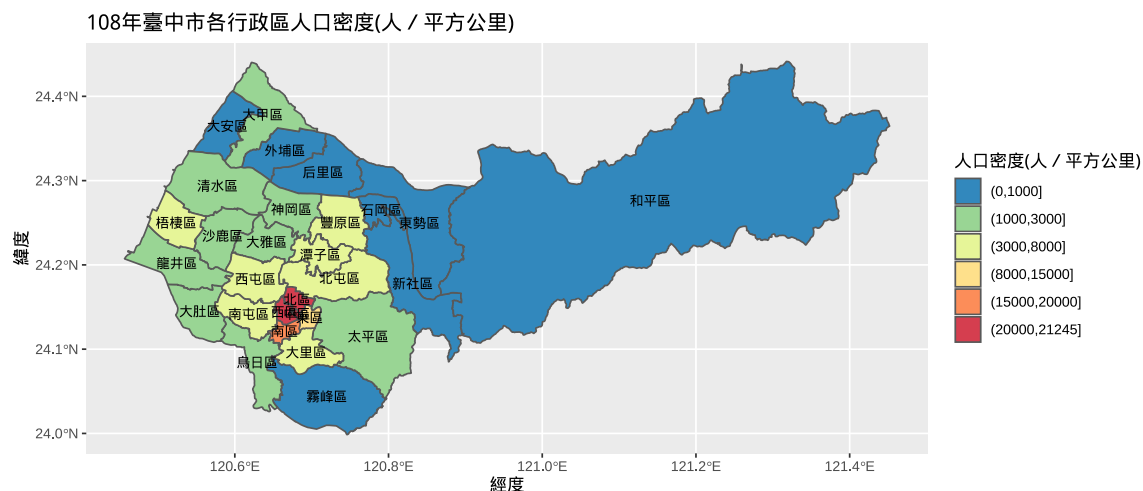
(提示: 欄位的排序是依照量表及變項，列位的排序是先依照組別、再依照組內每位學生之平均得分。)



5 Map

請畫出 108 年臺中市各行政區之「人口密度 (人 / 平方公里)」(30 分)

- 統計資料來源: 臺中市統計資料查詢平臺 (<https://govstat.taichung.gov.tw/DgbasWeb/index.aspx>)。(查詢方式為, 資料表式-統計年報 (行政區資料) ⇒ 貳、人口 ⇒ 臺中市人口、密度及性比例 ⇒ 時間: 107-108; 地區: 全選; 臺中市人口、密度及性比例。)
- 統計資料檔案名稱: taichung-pop-query.xlsx
- 台灣鄉鎮市區界線 shapefile 圖資檔: mapdata201907310953.zip
- 可參考「http://www.hmwu.idv.tw/web/R/E04-hmwu_R-Map.pdf」
- 要求: 地圖來源需使用「Shape files」, 各行政區需標上行政區名稱。
- 要求: 將「人口密度 (人 / 平方公里)」以下列六個區間分組: 「0–1000, 1001–3000, 3001–8000, 8001–15000, 15001–20000, 20001–」, 選取合適之單向連續型色階 (6 個顏色對應至前述之分組)。
- 提示: `as.numeric(gsub(",", "", "123,456"))`



6 PCA

資料來源: (UCI) Statlog (Vehicle Silhouettes),

[https://archive.ics.uci.edu/ml/datasets/Statlog+\(Vehicle+Silhouettes\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(Vehicle+Silhouettes)),

其中 `class` 為具有 4 個類別的反應變數 (Y) · {compactness, circularity, ..., hollows} 為 18 個解釋變數 (X)。資料檔: `statlog_vehicle_846x18.txt`。使用 FactoMineR 套件對此資料進行主成份分析 (PCA)。

6.1 印出 Eigenvalues/Variances, 畫出陡坡圖 (scree plot)(10 分)

6.2 畫出降維後的二維資料散佈圖, 每一觀察值需以顏色標上所屬類別 (10 分)