

# Homework 3

711378912 蔡宜諠

2025-04-02

## 作業題目：台灣人民對「民主」概念理解的研究

本研究旨在探討台灣人民對「民主」概念的理解，特別是「程序民主」與「理想民主」之間的比較。問卷設計包含基本資料與核心問題兩個部分，目的是分析不同背景的受訪者對於民主的認知與態度。

### 研究性質說明

本研究目的在於探索台灣民眾對民主概念（特別是程序民主與理想民主）的理解差異與認知傾向，並分析不同背景變項（如年齡、性別、政治立場）對民主態度的影響。由於本研究未設立明確可檢定的假設，而是透過問卷資料進行資料整理與歸納，因此屬於 **探索性研究（Exploratory Research）**。

### 問卷設計

問卷的設計分為以下兩個部分：

- **基本資料**：包括性別、出生年份、婚姻狀況、經濟狀況與壓力程度。
- **核心問題**：涉及台灣式民主的現階段認知、民主特質的優先性、改革需求、政治議題的關注程度、政治立場與資訊來源。

### 資料收集方式

- 使用 Google 表單進行資料收集。
- 採匿名填寫，不蒐集個人身份資訊。
- 問卷結構分為 **基本資料** 與 **核心問題** 兩部分，共 10 題。
- 填答時間約為 3~5 分鐘。

### 變數介紹與分類

#### 基本資料

題目編號	問題說明	分類方式
S1	性別	男性、女性、其他
S2	出生年份	西元年份 (範例：1990)
S3-1	婚姻狀況	單身 (未婚、離婚)、已婚 (含喪偶)
S3-2	是否有小孩	有、無
S4	經濟狀況與壓力程度	五個等級 (無壓力、輕微、中度、偏高、嚴重)

# 核心問題

題目編號	問題說明	分類方式
Q1	台灣式民主階段認知：五個階段（程序民主 → 達成理想民主）	二元分類：0 = 程序民主、初步理想、穩定發展；1 = 理想落實、達成理想民主
Q2	民主最重要特質	公平、自由、負責任、平等、參與（五類別）
Q3	對民主制度的改革需求	非常同意、同意、普通、不同意、非常不同意（五類別）
Q4	對政治議題的關注程度	極度關注、關注、普通、不太關心、完全不關心
Q5	政治立場	偏綠、偏藍、偏白、無偏向、其他（五類別）
Q6	資訊來源	新聞、社群、學校、朋友家人、其他（五類別）

## 資料處理與說明

### 前處理

匯入資料，並修改column name

```
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

df <- read.csv("D:/NTPU_class/categorical_analysis/code/hw3_data_Big5.csv", header = TRUE,
fileEncoding = "Big5")
names(df) <- c(
  "timestamp",
  "S1", "S2", "S3", "S4",
  "Q1", "Q2", "Q3", "Q4", "Q5", "Q6"
)
head(df, 3)
```

```

##                                     timestamp   S1   S2           S3
## 1 2025/03/30 10:42:52 下午 GMT+8 男性 1995 單身·無小孩
## 2 2025/03/30 10:43:25 下午 GMT+8 男性 1982 已婚·有小孩
## 3 2025/03/30 10:45:06 下午 GMT+8 女性 1970 單身·無小孩
##                                         S4
## 1 輕微：經濟尚稱穩定，偶爾會擔心金錢問題
## 2 輕微：經濟尚稱穩定，偶爾會擔心金錢問題
## 3 輕微：經濟尚稱穩定，偶爾會擔心金錢問題
##                                     Q1           Q2           Q3
## 1          1. 程序民主：僅擁有程序民主。 公民參與決策的機會 同意
## 2          1. 程序民主：僅擁有程序民主。 法律面前人人平等 非常同意
## 3 2. 初步理想：開始理想民主，但隨時會消滅。 公民參與決策的機會 普通
##                                     Q4
## 1          2. 關注：我會追蹤時事新聞、和親友討論政治，投票前會認真研究候選人
## 2 1. 極度關注：政治對我很重要，我會主動討論、參與相關活動，關心政治發展是生活一部分
## 3          3. 普通：有重大事件會留意，但平常不會特別關心或談論
##                                     Q5           Q6
## 1      偏藍          新聞媒體
## 2      偏白          新聞媒體；社群平台
## 3 無偏向以上任何色彩 新聞媒體；社群平台；學校或職場

```

## timestamp

這裡將該變數分為「中午前」與「中午後」，以及「上班時間」與「非上班時間」，希望藉由這兩種時間切分方式，觀察不同時間背景下的心理狀態是否會影響對民主的理解。例如，上班時間可能處於理性、專注的決策模式，而非上班時間則可能較情緒化或放鬆，進而影響對民主性質的判斷。

```

# 移除時區字串，轉為 POSIXct 時間
df$timestamp_clean <- sub(" GMT\+\+8", "", df$timestamp)
df$timestamp_clean <- as.POSIXct(df$timestamp_clean, format = "%Y/%m/%d %I:%M:%S %p")

# 抓出小時數(24小時制)
df$hour <- as.numeric(format(df$timestamp_clean, "%H"))

# 1. 中午前 vs 中午後分類
df$time_of_day_noon <- ifelse(df$hour < 12, "before_noon", "after_noon")
df$time_of_day_noon <- factor(df$time_of_day_noon, levels = c("before_noon", "after_noon"))

# 2. 上班時間 vs 非上班時間分類(9:00~18:00 為上班時間)
df$time_of_day_work <- ifelse(df$hour >= 9 & df$hour < 18, "work_time", "non_work_time")
df$time_of_day_work <- factor(df$time_of_day_work, levels = c("work_time", "non_work_time"))

# 確認結果
table(df$time_of_day_noon) # 中午前/中午後分布

```

```

##
## before_noon after_noon
##        44       41

```

```
table(df$time_of_day_work) # 上班時間/非上班時間分布
```

```
##  
##      work_time non_work_time  
##          17           68
```

```
head(df, 3)
```

```
##                      timestamp   S1   S2       S3  
## 1 2025/03/30 10:42:52 下午 GMT+8 男性 1995 單身·無小孩  
## 2 2025/03/30 10:43:25 下午 GMT+8 男性 1982 已婚·有小孩  
## 3 2025/03/30 10:45:06 下午 GMT+8 女性 1970 單身·無小孩  
##  
##      S4  
## 1 輕微：經濟尚稱穩定，偶爾會擔心金錢問題  
## 2 輕微：經濟尚稱穩定，偶爾會擔心金錢問題  
## 3 輕微：經濟尚稱穩定，偶爾會擔心金錢問題  
##  
##      Q1                  Q2      Q3  
## 1 1. 程序民主：僅擁有程序民主。 公民參與決策的機會 同意  
## 2 1. 程序民主：僅擁有程序民主。 法律面前人人平等 非常同意  
## 3 2. 初步理想：開始理想民主，但隨時會消滅。 公民參與決策的機會 普通  
##  
##      Q4  
## 1 2. 關注：我會追蹤時事新聞、和親友討論政治，投票前會認真研究候選人  
## 2 1. 極度關注：政治對我很重要，我會主動討論、參與相關活動，關心政治發展是生活一部分  
## 3 3. 普通：有重大事件會留意，但平常不會特別關心或談論  
##  
##      Q5                  Q6 timestamp_clean hour  
## 1 偏藍                新聞媒體 2025-03-30 22:42:52 22  
## 2 偏白                新聞媒體;社群平台 2025-03-30 22:43:25 22  
## 3 無偏向以上任何色彩 新聞媒體;社群平台;學校或職場 2025-03-30 22:45:06 22  
## time_of_day_noon time_of_day_work  
## 1     after_noon    non_work_time  
## 2     after_noon    non_work_time  
## 3     after_noon    non_work_time
```

## S1. 性別

問卷中原設計的性別選項包括「男性」、「女性」與「其他」，以反映性別多元。但在實際資料中，選擇「其他」者僅有一人，樣本數過少，無法支持後續有效的統計分析與群體比較。因此，本研究將性別變數重新分類為二元類別：「男性」與「女性」，並將「其他」類別排除，以提升分析的穩定性與解釋力。

```
# 檢查性別分布  
table(df$S1)
```

```
##  
##      女性      男性 非二元性別  
##          37           47           1
```

```
# 僅保留男性與女性，將其他設為 NA  
df$gender <- ifelse(df$S1 %in% c("男性", "女性"), df$S1, NA)  
df$gender <- factor(df$gender, levels = c("男性", "女性"))  
  
# 檢查轉換後的結果  
table(df$gender, useNA = "ifany")
```

```
##  
## 男性 女性 <NA>  
## 47 37 1
```

```
# 資料清除 (移除 NA)  
df <- df[!is.na(df$gender), ]  
  
# 確認刪除後的資料筆數  
table(df$gender)
```

```
##  
## 男性 女性  
## 47 37
```

## S2. 年齡

為了方便後續分析，我將出生年份先轉換為實際年齡，並同時將年齡依照範圍切分成不同群組。這樣做可以讓我在需要時使用「連續變數」（如平均年齡）或「類別變數」（不同年齡層比較）進行不同的分析方法，更有彈性也更清楚。

```
# 計算年齡 (以 2025 年為基準)  
df$age <- 2025 - as.numeric(df$S2)  
# 檢查結果  
summary(df$age)
```

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.  
## 27.00   31.75  35.00  38.96   42.00  68.00
```

由於本次資料的受訪者年齡集中在 30 至 40 歲之間，且未涵蓋 24 歲以下的青年族群，65 歲以上者也僅極少數，因此本研究根據實際年齡分布情形，將年齡分為兩組：「青壯年」（25–44 歲）與「壯年與高齡」（45 歲以上）。此分類方式仍參照政府勞動統計的架構，但針對實際資料進行簡化，以提升分析的實用性與解釋力。

```

# 計算年齡 (如果還沒建 age 欄位)
df$age <- 2025 - as.numeric(df$S2)

# 簡化為「青壯年」與「壯年~高齡」兩組
df$age_group_simple <- cut(df$age,
                             breaks = c(24, 44, Inf),
                             labels = c("25-44 青壯年", "45+ 壯年與高齡"),
                             right = TRUE)

# 檢查結果
table(df$age_group_simple)

```

```

## 
##   25-44 青壯年 45+ 壯年與高齡
##       66          18

```

### S3. 婚姻狀況與是否有小孩

問卷原設計將婚姻狀態與育兒狀況合併為單一問題，例如「已婚・有小孩」、「單身・無小孩」等。為便於分析不同因素對民主觀念的獨立影響，本研究將其拆分為兩個變數：

- 婚姻狀況 (**marital\_status**)：區分為「已婚」與「未婚」
- 是否有小孩 (**have\_children**)：區分為「有」與「無」

```

# 拆解婚姻與小孩狀況
df$marital_status <- ifelse(grepl("已婚", df$S3), "已婚", "未婚")
df$have_children <- ifelse(grepl("有小孩", df$S3), "有", "無")

# 轉為 factor (如需)
df$marital_status <- factor(df$marital_status, levels = c("未婚", "已婚"))
df$have_children <- factor(df$have_children, levels = c("無", "有"))

# 檢查結果
table(df$marital_status)

```

```

## 
## 未婚 已婚
##    50   34

```

```
table(df$have_children)
```

```

## 
## 無 有
## 58 26

```

## S4. 經濟狀況與壓力程度

本研究將 S4 的「經濟狀況與壓力程度」同時轉換為兩種型態：1. **有序類別變數**：保留問卷原始語意與壓力等級的順序特性，供交叉分析與比例模型使用。2. **數值變數 (1~5)**：便於進行平均值比較、視覺化與迴歸分析。

如此設計能兼顧變數語意清晰與統計實用性。

```
# 有序 factor (保留語意)
df$econ_stress <- factor(df$S4,
                           levels = c(
                               "沒有：經濟情形穩定，完全不會為金錢煩惱",
                               "輕微：經濟尚稱穩定，偶爾會擔心金錢問題",
                               "中度：收入與支出差不多，有時會擔心生活開銷",
                               "偏高：經濟狀況造成壓力，需精打細算過日子",
                               "嚴重：金錢問題已影響生活與心理健康"
                           ),
                           ordered = TRUE)

# 對應數值 (1~5)
df$econ_stress_num <- as.numeric(df$econ_stress)

# 檢查
table(df$econ_stress, df$econ_stress_num)
```

```
##                                     2   3
## 沒有：經濟情形穩定，完全不會為金錢煩惱    0   0
## 輕微：經濟尚稱穩定，偶爾會擔心金錢問題    47   0
## 中度：收入與支出差不多，有時會擔心生活開銷  0  13
## 偏高：經濟狀況造成壓力，需精打細算過日子    0   0
## 嚴重：金錢問題已影響生活與心理健康        0   0
```

## Q1. 台灣式民主階段認知 → y

Q1 為本研究的核心變數，衡量受訪者對台灣民主所處階段的認知。原始設計為五階段分類，反映從「程序民主」至「理想民主」的不同發展層次。

為利分析與詮釋，本研究將五個選項歸納為二元分類變數：

- **0**：尚未達成理想民主（程序民主、初步理想、穩定發展）
- **1**：已接近或實現理想民主（理想落實、達成理想民主）

此分類有助於釐清不同族群在民主概念理解上的傾向。

```
# 根據開頭數字自動分類為 0 或 1
df$Q1_binary <- ifelse(grepl("^\d\\.|^\d\\.", df$Q1), 1, 0)

# 檢查轉換結果
table(df$Q1, df$Q1_binary)
```

```

##                                     0  1
## 1. 程序民主：僅擁有程序民主。      12  0
## 2. 初步理想：開始理想民主，但隨時會消滅。    24  0
## 3. 穩穩定發展：理想民主開始慢慢發酵，不容易被消滅。 32  0
## 4. 理想落實：理想民主已落實在許多地方。      0 13
## 5. 達成理想民主：已完全達成理想民主。      0  3

```

## Q2. 民主最重要的特質

Q2 旨在探討受訪者心目中「民主」最重要的核心特質。問項提供五種選擇：公平、自由、負責任、平等與參與，反映受訪者對民主價值的優先排序。

此變數為名義尺度（nominal variable），無內在順序，適用於交叉分析與圖形呈現，以觀察不同背景下的價值傾向差異。

```

df$Q2_simple <- ifelse(df$Q2 == "公平的選舉制度", "選舉",
                         ifelse(df$Q2 == "言論與新聞自由", "自由",
                                ifelse(df$Q2 == "廉能負責的政府", "責任",
                                       ifelse(df$Q2 == "法律面前人人平等", "平等",
                                          ifelse(df$Q2 == "公民參與決策的機會", "參與", NA)))))

table(df$Q2_simple)

```

```

##                                     0  1
## 平等 自由 參與 責任 選舉
##    7   15   35   14   13

```

## Q3. 對民主制度改革需求 (Reform Attitude)

Q3 測量受訪者對台灣現行民主制度是否需要大幅改革的主觀態度，採用五點 Likert 量表，選項依同意程度由高至低排列：

- 非常同意
- 同意
- 普通
- 不同意
- 非常不同意

此變數具備明確的順序性，因此轉為 ordered factor，可保留等級差異性。

```

# 有序 factor (保留順序性)
df$Q3 <- factor(df$Q3,
                  levels = c("非常同意", "同意", "普通", "不同意", "非常不同意"),
                  ordered = TRUE)

# 使用 recode 對應分數 (文字轉成分數：越同意分數越高)
df$Q3_num <- recode(as.character(df$Q3),
                      "非常同意" = 5,
                      "同意" = 4,
                      "普通" = 3,
                      "不同意" = 2,
                      "非常不同意" = 1)

# 檢查結果
print(table(df$Q3, df$Q3_num, useNA = "ifany"))

```

```

##          2   3   4   5
## 非常同意 0   0   0 19
## 同意      0   0 21  0
## 普通      0 39  0  0
## 不同意     5   0   0  0
## 非常不同意 0   0   0  0

```

## Q4. 政治議題關注程度

Q4 測量受訪者平時對政治議題的關注程度，屬於 Likert 五點量表，由「極度關注」到「完全不關心」。此變數具備明確順序性，適合轉為 ordered factor 處理，以利進行群體差異分析與視覺化。同時，本研究亦建立對應數值變數 (Q4\_num)，以利模型應用與統計檢定。

```

# 1. 擷取前面等級文字 (只留下「極度關注」、「關注」...)
df$Q4_level <- sub("^\\d\\.\\s+", "", df$Q4) # 移除開頭 "1. " 等
df$Q4_level <- sub(":.*", "", df$Q4_level) # 再移除「：」後面的敘述句

# 2. 轉為有序 factor
df$Q4_level <- factor(df$Q4_level,
                        levels = c("極度關注", "關注", "普通", "不太關心", "完全不關心"),
                        ordered = TRUE)

# 3. 反向編碼為數值分數 (越關注分數越高)
df$Q4_num <- recode(df$Q4_level,
                      "極度關注" = 5,
                      "關注" = 4,
                      "普通" = 3,
                      "不太關心" = 2,
                      "完全不關心" = 1)

# 檢查結果
table(df$Q4_level, df$Q4_num)

```

```

##          1 2 3 4 5
## 極度關注 0 0 0 0 9
## 關注      0 0 0 32 0
## 普通      0 0 32 0 0
## 不太關心  0 9 0 0 0
## 完全不關心 2 0 0 0 0

```

## Q5. 政治立場傾向

Q5 為名義尺度變數，用以描述受訪者自我認定的政治立場。原始問卷選項包含：實際資料中，除前四類標準選項外，亦有少數受訪者於「其他」欄填入自由文字（如「反共的，目前偏綠」、「要是非知識份子」等）。由於這些回覆樣本極少，語意模糊且無法準確分類，故本研究在資料處理階段即予以排除，以提升分類清晰度與分析穩定性。此外，考量「偏白」受訪者數量不低，具有實質政治傾向，故將其保留為獨立類別。本研究最終將政治立場歸納為四大類：

- 偏綠
- 偏藍
- 偏白
- 中立（原為「無偏向以上任何色彩」）

此分類有助於釐清不同政治光譜對民主觀念之影響。

```

# 1. 清除非預期自由輸入的少數樣本（如反共的、要是非 等）
valid_Q5 <- c("偏綠", "偏藍", "偏白", "無偏向以上任何色彩")
df <- df[df$Q5 %in% valid_Q5, ]

# 2. 四類政治傾向分類
df$Q5_grouped4 <- case_when(
  df$Q5 == "偏綠" ~ "偏綠",
  df$Q5 == "偏藍" ~ "偏藍",
  df$Q5 == "偏白" ~ "偏白",
  df$Q5 == "無偏向以上任何色彩" ~ "中立",
  TRUE ~ NA_character_
)

# 3. 設定順序（方便圖表呈現）
df$Q5_grouped4 <- factor(df$Q5_grouped4, levels = c("偏綠", "偏藍", "偏白", "中立"))

# 4. 檢查結果
table(df$Q5_grouped4)

```

```

## 
## 偏綠 偏藍 偏白 中立
##   34     4     7    36

```

## Q6. 政治資訊來源

Q6 為複選題，受訪者可勾選一至多個常用的政治或民主資訊來源，包含：

- 新聞媒體
- 社群平台

- 學校或職場
- 家人或朋友
- 其他 (自由填答)

為利後續分析與交叉比對，本研究將此題轉換為五個二元變數 ( dummy variables )，每個欄位標示該受訪者是否勾選該資訊來源 ( 1 = 有 · 0 = 無 )。其中「其他」選項不再細分內容，統一歸為一類。

```
# 定義你要拆的類別 (需與問卷一致)
info_sources <- c("新聞媒體", "社群平台", "學校或職場", "家人或朋友")

# 把所有回答轉成 character · 再用；拆分
split_sources <- strsplit(as.character(df$Q6), ":")

# 建立 dummy 欄位 (每個來源一欄)
for (src in info_sources) {
  df[[paste0("Q6_", src)]] <- sapply(split_sources, function(x) as.integer(src %in% x))
}

# 統一處理「其他」類別：只要有不是上述類別的，就視為選了其他
df$Q6_其他 <- sapply(split_sources, function(x) {
  other_selected <- setdiff(x, info_sources)
  as.integer(length(other_selected) > 0)
})

# 確認欄位建立結果
head(df[, grep1("^\w{2}_", names(df))])
```

```
##   Q6_新聞媒體 Q6_社群平台 Q6_學校或職場 Q6_家人或朋友 Q6_其他
## 1       1         0         0         0         0
## 2       1         1         0         0         0
## 3       1         1         1         0         0
## 4       1         0         0         0         0
## 5       1         0         0         0         0
## 6       1         1         0         1         0
```



# 資料分析階段

本章節開始進行資料的探索性分析，針對核心變數（如民主觀念、政治傾向、經濟壓力、關注程度等）進行描述性統計與視覺化。此步驟有助於理解資料特性與變項之間的潛在關聯，作為後續統計檢定與模型建立的基礎。

```
# 建立新的分析用資料表，只保留需要的欄位
df_model <- df[, c(
  "Q1_binary",
  "hour", "time_of_day_noon", "time_of_day_work",
  "gender", "age", "marital_status", "have_children", "econ_stress_num",
  "Q2_simple", "Q3_num", "Q4_num", "Q5_grouped4",
  "Q6_新聞媒體", "Q6_社群平台", "Q6_學校或職場", "Q6_家人或朋友", "Q6_其他"
)]
str(df_model)
```

```
## 'data.frame': 81 obs. of 18 variables:
## $ Q1_binary      : num 0 0 0 0 0 0 0 0 0 ...
## $ hour           : num 22 22 22 22 22 22 23 23 23 ...
## $ time_of_day_noon: Factor w/ 2 levels "before_noon",...: 2 2 2 2 2 2 2 2 2 ...
## $ time_of_day_work: Factor w/ 2 levels "work_time","non_work_time": 2 2 2 2 2 2 2 2 2
2 ...
## $ gender          : Factor w/ 2 levels "男性","女性": 1 1 2 2 1 2 2 1 2 1 ...
## $ age             : num 30 43 55 41 57 54 36 38 58 35 ...
## $ marital_status  : Factor w/ 2 levels "未婚","已婚": 1 2 1 2 1 1 2 2 2 1 ...
## $ have_children   : Factor w/ 2 levels "無","有": 1 2 1 2 1 2 1 2 2 1 ...
## $ econ_stress_num : num 2 2 2 NA NA NA 2 NA NA 2 ...
## $ Q2_simple       : chr "參與" "平等" "參與" "平等" ...
## $ Q3_num          : num 4 5 3 3 5 4 4 3 3 3 ...
## $ Q4_num          : num 4 5 3 2 3 4 3 4 3 3 ...
## $ Q5_grouped4    : Factor w/ 4 levels "偏綠","偏藍",...: 2 3 4 4 4 4 2 1 2 1 ...
## $ Q6_新聞媒體    : int 1 1 1 1 1 1 0 1 1 ...
## $ Q6_社群平台    : int 0 1 1 0 0 1 1 1 0 ...
## $ Q6_學校或職場  : int 0 0 1 0 0 0 0 0 0 ...
## $ Q6_家人或朋友  : int 0 0 0 0 0 1 0 0 0 1 ...
## $ Q6_其他         : int 0 0 0 0 0 0 0 0 0 0 ...
```

```
# 確認結構與基本概況
summary(df_model)
```

```

##   Q1_binary          hour       time_of_day_noon    time_of_day_work
## Min.   :0.0000  Min.   :0.00  before_noon:42      work_time   :17
## 1st Qu.:0.0000  1st Qu.: 8.00 after_noon :39      non_work_time:64
## Median :0.0000  Median : 9.00
## Mean   :0.1852  Mean   :13.02
## 3rd Qu.:0.0000  3rd Qu.:22.00
## Max.   :1.0000  Max.   :23.00
##
##   gender        age       marital_status have_children econ_stress_num
## 男性:44  Min.   :27.00 未婚:49      無:57      Min.   :2.000
## 女性:37  1st Qu.:32.00 已婚:32      有:24      1st Qu.:2.000
##                  Median :35.00
##                  Mean   :38.84
##                  3rd Qu.:42.00
##                  Max.   :68.00
##                  NA's   :23
##
##   Q2_simple        Q3_num      Q4_num      Q5_grouped4
## Length:81  Min.   :2.000  Min.   :1.000  偏綠:34
## Class  :character  1st Qu.:3.000  1st Qu.:3.000  偏藍: 4
## Mode   :character  Median :3.000  Median :3.000  偏白: 7
##                  Mean   :3.617  Mean   :3.407  中立:36
##                  3rd Qu.:4.000  3rd Qu.:4.000
##                  Max.   :5.000  Max.   :5.000
##
##   Q6_新聞媒體     Q6_社群平台    Q6_學校或職場    Q6_家人或朋友    Q6_其他
## Min.   :0.0000  Min.   :0.000  Min.   :0.0000  Min.   :0.0000  Min.   :0
## 1st Qu.:1.0000  1st Qu.:0.000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0
## Median :1.0000  Median :1.000  Median :0.0000  Median :0.0000  Median :0
## Mean   :0.8642  Mean   :0.679  Mean   :0.1481  Mean   :0.4321  Mean   :0
## 3rd Qu.:1.0000  3rd Qu.:1.000  3rd Qu.:0.0000  3rd Qu.:1.0000  3rd Qu.:0
## Max.   :1.0000  Max.   :1.000  Max.   :1.0000  Max.   :1.0000  Max.   :0
##

```

## 數值型變數

由於 Q6 為複選題，轉換後變數過多，初步分析時資訊來源與民主概念間關聯不明顯，因此先將 Q6 的相關變數排除，專注在其他個人特徵與態度上的分析。

```

numeric_data <- df_model[, c(
  "Q1_binary",      # 應變數
  "hour",           # 填寫問卷的時間
  "age",            # 年齡
  "econ_stress_num",# 經濟壓力分數 (1-5)
  "Q3_num",         # 政治改革態度分數
  "Q4_num"          # 政治關注程度
)]
str(numeric_data)

```

```

## 'data.frame': 81 obs. of 6 variables:
## $ Q1_binary : num 0 0 0 0 0 0 0 0 0 ...
## $ hour      : num 22 22 22 22 22 22 22 23 23 23 ...
## $ age       : num 30 43 55 41 57 54 36 38 58 35 ...
## $ econ_stress_num: num 2 2 2 NA NA NA 2 NA NA 2 ...
## $ Q3_num    : num 4 5 3 3 5 4 4 3 3 3 ...
## $ Q4_num    : num 4 5 3 2 3 4 3 4 3 3 ...

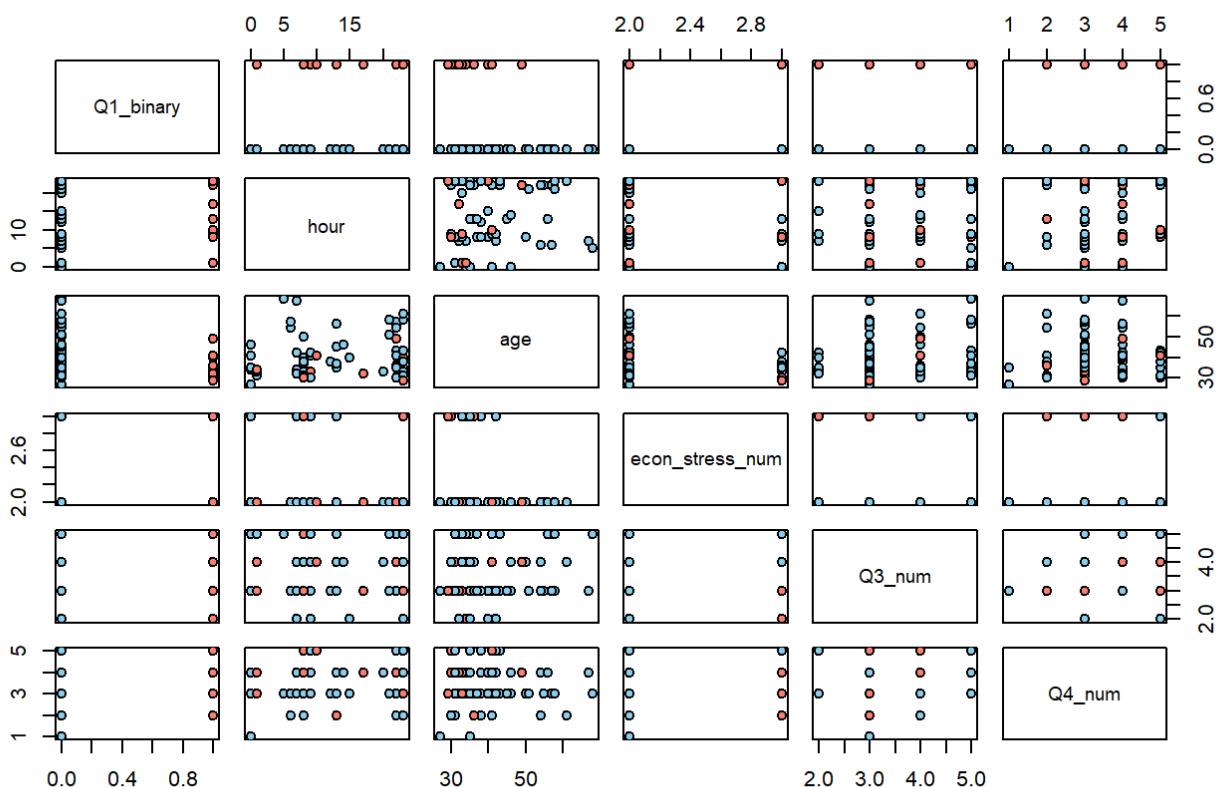
```

```

pairs(numeric_data,
      main = "數值變數之間的散佈圖",
      pch = 21,
      bg = c("skyblue", "salmon")[numeric_data$Q1_binary + 1])

```

數值變數之間的散佈圖



雖然從整體散佈圖來看，大多數變數間未展現強烈的線性關係，但個別變數如改革需求（Q3\_num）與經濟壓力（econ\_stress\_num）之間略呈正向傾向，可能隱含潛在影響力，仍值得透過後續統計模型進行驗證與探討。

```

# 數值變數清單
numeric_vars <- c("hour", "age", "econ_stress_num", "Q3_num", "Q4_num")

# 一次跑完 t-test 並整理輸出
t_results <- lapply(numeric_vars, function(var) {
  test <- t.test(df[[var]] ~ df$Q1_binary)
  data.frame(
    variable = var,
    mean_Q1_0 = mean(df[[var]][df$Q1_binary == 0], na.rm = TRUE),
    mean_Q1_1 = mean(df[[var]][df$Q1_binary == 1], na.rm = TRUE),
    t_statistic = test$statistic,
    p_value = test$p.value
  )
})
}

# 合併為單一表格
t_results_df <- do.call(rbind, t_results)

# 結果清楚顯示
print(t_results_df)

```

	variable	mean_Q1_0	mean_Q1_1	t_statistic	p_value
## t	hour	12.772727	14.133333	-0.5763283	0.570708273
## t1	age	39.863636	34.333333	2.9913707	0.004734077
## t2	econ_stress_num	2.177778	2.307692	-0.8949288	0.383504003
## t3	Q3_num	3.666667	3.400000	1.2054981	0.239329474
## t4	Q4_num	3.348485	3.666667	-1.3296713	0.196760267

只有年齡 (age) 一個變數與受訪者認為民主是否達到理想的認知有統計上的顯著關聯。因此，後續分析或模型建構時，可特別關注 age 對民主認知的影響，其他變數則可能不是主要解釋民主認知差異的因素。

## 類別型變數

```

# 需要測試的類別變數名稱
cat_vars <- c("time_of_day_noon", "time_of_day_work", "gender", "marital_status",
            "have_children", "Q2_simple", "Q5_grouped4", "age_group_simple")

# 初始化結果儲存
chi_results <- data.frame(variable = character(), chi_square = numeric(), df = integer(),
                            p_value = numeric())

# 進行卡方檢定
for (var in cat_vars) {
  tbl <- table(df[[var]], df$Q1_binary)
  test <- chisq.test(tbl)
  chi_results <- rbind(chi_results, data.frame(
    variable = var,
    chi_square = test$statistic,
    df = test$parameter,
    p_value = test$p.value
  ))
}

```

```
## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect
## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect
## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect
## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect
## Warning in chisq.test(tbl): Chi-squared approximation may be incorrect
```

```
# 顯示結果
print(chi_results)
```

```
##           variable chi_square df      p_value
## X-squared  time_of_day_noon 0.02528721 1 0.87365348
## X-squared1 time_of_day_work 0.06108163 1 0.80479447
## X-squared2          gender 0.60259242 1 0.43759061
## X-squared3   marital_status 0.69609375 1 0.40409937
## X-squared4    have_children 3.40204844 1 0.06511551
## X-squared5        Q2_simple 7.71202979 4 0.10271504
## X-squared6     Q5_grouped4 4.11187930 3 0.24963386
## X-squared7 age_group_simple 1.34024273 1 0.24699117
```

根據卡方檢定結果，包含填答時間（中午前後、上班與非上班）、性別、婚姻狀況、育兒狀況、民主最重要特質以及政治立場等類別變數，皆未與民主認知（程序民主 vs. 理想民主）有顯著的相關性 ( $p > 0.05$ )。這表示這些類別變數在本次樣本中未觀察到與民主認知的統計關聯。

## 資料分析

依據老師的題目需放入性別、年齡、以及選擇另外其中一個變數加入，因其他的都沒有顯著，因此這裡挑選是否有小孩 (have\_children) 接近統計顯著水準 ( $p=0.065$ )，且在現實社會情境下，是否育有子女可能直接或間接影響個人對民主制度的感受與需求，因此將此變數納入模型進行後續的深入分析與探討。

## 建立完整模型

```
log.fit <- glm(Q1_binary ~ gender + age + have_children, family = binomial, data = df)
summary(log.fit)
```

```

## 
## Call:
## glm(formula = Q1_binary ~ gender + age + have_children, family = binomial,
##      data = df)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.94265   1.74349   0.541   0.589
## gender女性         -0.15847   0.63722  -0.249   0.804
## age                 -0.05758   0.05018  -1.147   0.251
## have_children有    -1.44148   1.16042  -1.242   0.214
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 77.625 on 80 degrees of freedom
## Residual deviance: 70.095 on 77 degrees of freedom
## AIC: 78.095
##
## Number of Fisher Scoring iterations: 5

```

年齡與是否有小孩可能在實務上有潛在影響，但需更大量的樣本才能觀察到顯著性；性別在本研究中與民主理解並無明顯關聯。

```
anova(log.fit)
```

```

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Q1_binary
##
## Terms added sequentially (first to last)
##
##
##                               Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                           80    77.625
## gender            1    1.1539      79    76.471  0.28273
## age               1    4.4676      78    72.003  0.03454 *
## have_children  1    1.9084      77    70.095  0.16714
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

當依序納入性別、年齡與是否育有小孩作為預測變數時，僅年齡顯著減少模型的誤差，其卡方檢定 p 值為 0.0345，顯著低於 0.05 水準，代表年齡對於是否認為台灣已實現理想民主的預測具有統計上的意義。

```
step(log.fit)
```

```

## Start: AIC=78.09
## Q1_binary ~ gender + age + have_children
##
##          Df Deviance    AIC
## - gender      1   70.157 76.157
## - age         1   71.715 77.715
## - have_children 1   72.003 78.003
## <none>           70.095 78.095
##
## Step: AIC=76.16
## Q1_binary ~ age + have_children
##
##          Df Deviance    AIC
## - age        1   71.865 75.865
## <none>           70.157 76.157
## - have_children 1   72.283 76.283
##
## Step: AIC=75.86
## Q1_binary ~ have_children
##
##          Df Deviance    AIC
## <none>           71.865 75.865
## - have_children 1   77.625 79.625

```

```

##
## Call: glm(formula = Q1_binary ~ have_children, family = binomial, data = df)
##
## Coefficients:
## (Intercept) have_children
##             -1.122       -2.013
##
## Degrees of Freedom: 80 Total (i.e. Null); 79 Residual
## Null Deviance: 77.62
## Residual Deviance: 71.86    AIC: 75.86

```

逐步回歸結果顯示，第一階段移除 gender · AIC 從 78.09 降至 76.16；第二階段再移除 age · AIC 進一步降至 75.86 · 為最低值。雖然 AIC 有下降，但整體模型解釋力並未有明顯提升。

```

## Parsimonious Model
red.log.fit <- glm(Q1_binary ~ gender + age + have_children, family=binomial, data=df)
summary(red.log.fit)

```

```

## Call:
## glm(formula = Q1_binary ~ gender + age + have_children, family = binomial,
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.94265   1.74349   0.541   0.589
## gender女性 -0.15847   0.63722  -0.249   0.804
## age         -0.05758   0.05018  -1.147   0.251
## have_children有 -1.44148   1.16042  -1.242   0.214
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 77.625 on 80 degrees of freedom
## Residual deviance: 70.095 on 77 degrees of freedom
## AIC: 78.095
##
## Number of Fisher Scoring iterations: 5

```

在納入性別、年齡與是否育有子女的條件下，logistic 回歸結果顯示三者對預測民主認知傾向（Q1\_binary）皆未達顯著水準（ $p$  值皆  $> 0.05$ ），AIC 為 78.095，整體模型解釋力有限，顯示這些變數對於區分受訪者是否認為台灣已達理想民主影響不大。

## ROC 分析（以年齡為例）

```

library(pROC)

## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
## 
##     cov, smooth, var

# 建立只使用 age 的邏輯斯迴歸模型
log.age <- glm(Q1_binary ~ age, family = binomial, data = df)

summary(log.age) # 確認係數顯著性

```

```
##  
## Call:  
## glm(formula = Q1_binary ~ age, family = binomial, data = df)  
##  
## Coefficients:  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept)  2.0025    1.7759   1.128   0.2595  
## age        -0.0953    0.0503  -1.894   0.0582 .  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
## Null deviance: 77.625 on 80 degrees of freedom  
## Residual deviance: 72.283 on 79 degrees of freedom  
## AIC: 76.283  
##  
## Number of Fisher Scoring iterations: 5
```

```
exp(coef(log.age)) # 算 odds ratio
```

```
## (Intercept)      age  
## 7.4076089   0.9091029
```

```
confint(log.age) # 95% 信賴區間
```

```
## Waiting for profiling to be done...
```

```
##           2.5 %    97.5 %  
## (Intercept) -1.067704  6.03122993  
## age         -0.213479 -0.01232207
```

```
round(exp(confint(log.age)), 3)
```

```
## Waiting for profiling to be done...
```

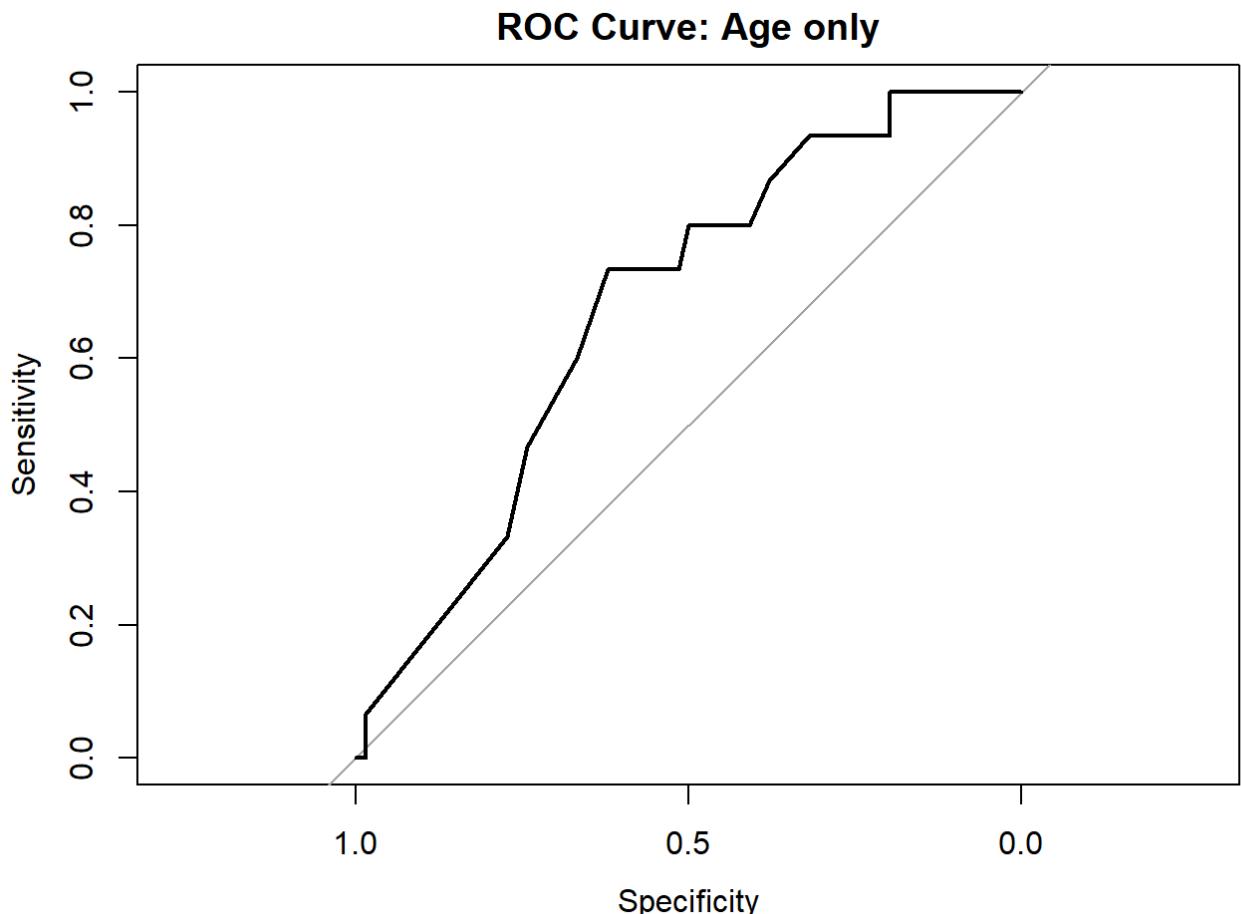
```
##           2.5 %    97.5 %  
## (Intercept) 0.344 416.227  
## age         0.808   0.988
```

```
# 預測機率  
pred_age <- predict(log.age, type = "response")  
  
# 畫 ROC 曲線  
roc_age <- roc(df$Q1_binary, pred_age)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
plot(roc_age, main = "ROC Curve: Age only")
```



```
auc(roc_age)
```

```
## Area under the curve: 0.6773
```

從 ROC 曲線觀察可知，年齡作為單一解釋變數時，對於預測受訪者是否認為台灣已接近「理想民主」的區辨能力有限。雖然曲線略高於隨機預測線（灰色對角線），但整體表現仍較為保守，顯示僅依賴年齡變數無法有效預測民主態度。年齡與民主認知呈現負相關，意即年齡愈大者，愈不傾向認為台灣已達成理想民主。這一關係也在邏輯斯迴歸與 ROC 曲線中反映出來，模型的預測機率隨年齡增加而下降。

## 小結

目前模型中，年齡（`age`）是唯一在邏輯斯迴歸中達到顯著水準的變數（ $p = 0.0345$ ），同時也是逐步篩選中最具貢獻的變數。其他變數如性別（`gender`）與是否有小孩（`have_children`）不具顯著性，也未能有效提升模型表現（AIC未明顯下降）。因此，後續分析可以僅保留「年齡」作為解釋變數進行模型簡化與詮釋。

## 結論

模型公式：

$$\log\left(\frac{\pi}{1 - \pi}\right) = 7.4076 - 0.0909 \cdot \text{age}$$

- 年齡對於是否認為台灣已實現理想民主的預測效果為負向影響，代表年齡越大，越不容易認為台灣已達理想民主。
- **Odds ratio 為 0.91**，每增加一歲，認為民主已達理想階段的機率會下降約 9%。
- 其 95% 信賴區間為 **(0.808, 0.988)**，未涵蓋 1，具統計顯著性。

綜上所述，雖然本研究中大部分變項未達顯著，但年齡在統計與實務面皆展現出預測力，顯示不同世代對民主理想的認知仍存在落差，值得未來更進一步的研究。加油！