

111-2商管程式設計期末專案第20組  
在什麼樣的情況下會影響人們的事實查核能力

組員：李維 B08302330、蘇則諺 B09602027、陳開 B09602049、李心妤 B09602042、陳芊尋 B10102104

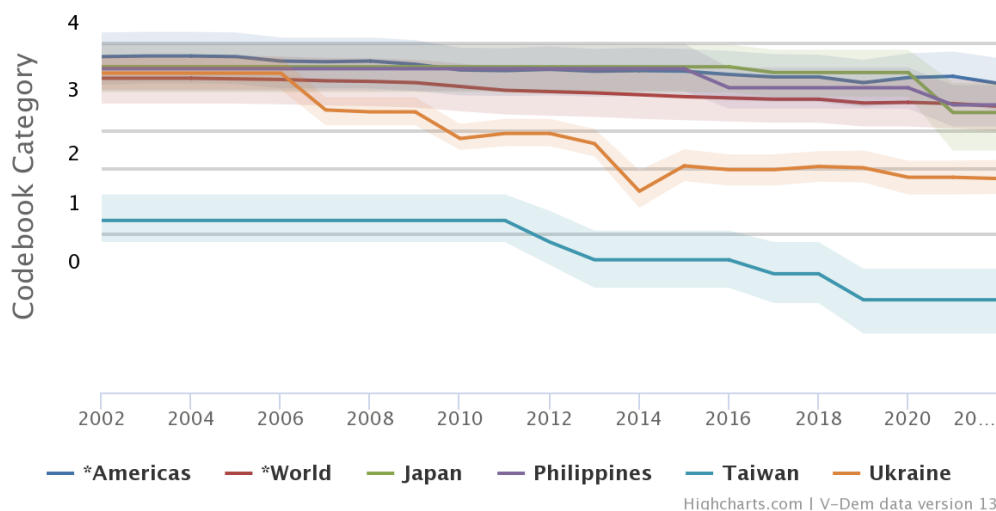
## 一. 動機 & 資料說明

近期台灣的國際情勢呈現緊張的態勢，不論是台灣周邊地區的地緣政治因素、區域安全形勢以及台海局勢的變化都對台灣的國際情勢產生了重大影響。這種緊張局勢可能對台灣的經濟、外交關係以及國家安全帶來種種挑戰，因此台灣政府需要積極應對這些挑戰，並與國際社會保持緊密的聯繫。

另一方面，台灣成為全球假消息入侵的熱點地區，其傳播量居世界之冠。這些假消息源於境外，經過精心策劃和傳播，對台灣社會造成了深遠的影響。這些假消息跨越政治、社會、經濟等各個領域，且具有極大的誘惑力和影響力，容易引起民眾的恐慌和不穩定情緒。因此，台灣政府需要加強假消息的防範和打擊，同時提高公眾對假消息的識別能力。

假消息的入侵對台灣的選舉和政治過程也產生了顯著的影響。假消息的傳播可能扭曲選民對候選人和政黨的看法，影響選舉結果的公正性和可信度。此外，假消息還可能操縱民眾對特定議題的觀點和行為，對政府的政策制定和執行產生重大影響。為了確保選舉的公正性和政治過程的透明度，政府和相關機構應該加強對假消息的監測和打擊，同時提高公眾的媒體素養和批判思維能力。

此外，透過V-Dem Institute的調查顯示，相比於烏克蘭、日本與菲律賓等其他地區，台灣民眾受到境外假訊息的影響較為嚴重（縱軸數字越小表示受假訊息影響較嚴重）。因此，為了研究在什麼樣的情況下會影響人們的事實查核能力，本專案透過中研院SRDA進行的研究調查進行資料分析，並將資料結果視覺化。



圖一、不同地區的事實查核能力

## 二. 研究方法

本次研究希望對「對生活中假消息的敏感程度如何影響事實查核能力」及「對生活中信息的猜疑程度如何影響事實查核能力」進行分析。我們首先將資料進行前處理後，建立以下兩種模型（皆採用OLS Regression Model）。最後，針對模型結果作圖並分析其趨勢。以下為本次研究所採用的模型及其所考慮的變數：

- 模型一：對生活中假消息的敏感程度如何影響事實查核能力

$$FKINDT_i = \delta_i + \beta_1 FK I_i + \beta_2 FK II_i + \beta_3 FK III_i + \beta_4 X_i + \dots + \varepsilon_i$$

	問題
FKI	請問您最近一年內有沒有收到過假消息？
FKII	請問您在日常生活當中，常不常出現假消息？
FKIII	請問您覺得假消息對社會的影響嚴不嚴重。

- 模型二：對生活中信息的猜疑程度如何影響事實查核能力

$$FKINDT_i = \delta_i + \beta_1 PRES_i + \beta_2 POLIT_i + \beta_3 CITIZ_i + \beta_4 ENMY + \dots + \varepsilon_i$$

	問題
PRES	請問您認為媒體工作者是否經常在製造假消息？
POLIT	請問您認為政治人物是否經常在製造假消息？
CITIZ	請問您認為一般民眾是否經常在製造假消息？
ENMY	請問您認為境外勢力是否經常在製造假消息？
NGO	請問您認為社運團體是否經常在製造假消息？
MAG	請問您認為報紙、雜誌是否經常在散播假消息？
TV	請問您認為電視、廣播是否經常在散播假消息？
INTNT	請問您認為網路是否經常在散播假消息？
MOB	請問您認為手機是否經常在散播假消息？
NMR	請問您認為社區鄰里是否經常在散播假消息？

### 三. 使用函數庫介紹

在本次作業中NumPy & Pandas於資料清理、scikit-learn於變數encoding、Preadstat於讀檔、Statsmodels於線性回歸及Plotly於繪圖，而下面將逐一介紹各函數庫。

#### 1. NumPy (資料清理)

NumPy (Numerical Python) 是Python中用於科學計算的核心函式庫之一。支援高階大規模的多維陣列與矩陣運算。NumPy主要功能包括：

- 快速、高效的數組操作，例如索引、切片、重塑和排序。
- 數學函式庫，包括線性代數、傅立葉變換和隨機數生成。
- 數組的廣播功能，使得不同形狀的數組可以進行元素級別的運算。

#### 2. Pandas (資料清理)

為python數據分析模組，用於數據分析和處理，並提供高效能、簡易使用的資料格式(Data Frame)供人們可以快速操作及分析資料。Pandas提供了三種數據結構：

Series、DataFrame、Panel，其中Series與DataFrame為主要，Pandas主要功能包括：

- 數據清理和預處理，包括缺失值處理、數據轉換和數據合併。
- 數據索引和選擇，可以根據標籤或位置進行數據的選取和切片。
- 數據聚合和分組，用於計算統計指標和執行群組操作。
- 數據的輸入和輸出，支持多種數據格式（如CSV、Excel和SQL）的讀寫操作。

#### 3. Scikit-learn (變數encoding)

Scikit-learn又為SKlearn，Python機器學習函式庫，提供了各種機器學習算法和工具，主要功能分為六大塊：

- Classification (分類)：確認檢測的目標屬於哪一種類別，可用於如垃圾郵件偵測、圖像識別
- Regression (回歸)：預測檢測資料的連續值屬性。
- Clustering (分群)：自動將資料分類成不同群集。可用於如客戶分類等
- Dimensionality reduction (降維)：檢索所需考慮的隨機變數數量
- Model selection (模型選擇)：比較、驗證各參數與模型
- Preprocess (前處理)：特徵提取與正規化

#### 4. Preadstat

是一個Python函式庫，用於讀取和處理不同統計軟體（如SPSS和SAS）生成的數據文件。它可以將這些文件轉換為Pandas中的DataFrame，以便進行進一步的數據分析和處理。本次作業中使用該函數將SPSS的sav檔轉成csv檔。

#### 5. Statsmodels

用於進行統計建模和推斷分析。它提供了各種統計模型和統計測試的實現，包括線性回歸、時間序列分析和方差分析等。Statsmodels還提供了對模型進行評估和診斷的工具，以及統計推斷和模型預測的功能。

#### 6. Plotly

交互式的繪圖函式庫，支持各種圖表類型，包括散點圖、線圖、長條圖和熱圖等。Plotly還提供了互動式功能（此次使用該函數而非matplotlib原因），可以在圖表上添加滑塊、按鈕和標籤等元素，用戶可以進行交互式的數據探索和分析。

#### 四. 程式介紹

##### 1. Tidying：讀取檔案並加以處理，除去不完整的資料以供分析。

資料讀取：由於在中研院SRDA資料庫所下載之資料為.sav檔，一般用於處理資料部分的numpy、pandas模組皆順利讀取，固本次使用pyreadstat套件讀取sav檔並轉存成csv檔，方便日後的檔案處理。轉檔後再次讀取csv檔案成dataframe格式方便處理。

```
tidying.py X  statis.py  vdem.png  fakenews2.png

tidying.py > ...
1  import numpy as np
2  import pandas as pd
3  import pyreadstat
4  from sklearn.preprocessing import LabelEncoder
5
6  df_1, meta = pyreadstat.read_sav("fakedata.sav", encoding="big5", \
7  .....:                          apply_value_formats=True)
8
9  df_1.to_csv("fakedata.csv", index=False, encoding="utf-8")
10
11 df_2 = pd.read_csv("fakedata.csv", encoding="utf-8")
12
```

答案轉換：本次下載的問卷題目總共可以分為三類

- 主要變數：民眾的假消息事實查核能力相關問題所得到的答案，答案分為"有"、"沒有"兩種，轉換成0以及1

```
tidying.py X  statis.py  vdem.png  fakenews2.png  models.tex  ques.pdf

tidying.py > varlist12
321 varlist12 = ["fncheck_friend", "fncheck_book", "fncheck_expert", "fncheck_factcheck", "fnrespond1",
322 .....:       "fnrespond_2", "fnrespond_3", "fnrespond_4", "fc", "fc_usage", "tfc"]
323
324 for i in varlist12:
325     df_2["fakenews_indicator"] += df_2[i]
326
327 varlist13 = ["partyhate", "@2024partyhate", "@2024presidenthate"]
328
329 df_2 = df_2.drop(varlist13, axis=1)
330
331 df_2.to_csv("fakedata_tidy.csv", index=False, encoding="utf-8")
```

- 一般連續變數：其餘問題且答案為連續者，例如："從未"、"偶而"、"經常"、"每天"，分別編為0,1,2,3

```
tidying.py X  statis.py  vdem.png  fakenews2.png

tidying.py > ...
164 def responsibility(col):
165     if col == "一點也沒責任":
166         col = 0
167     elif col == "沒有責任":
168         col = 1
169     elif col == "有責任":
170         col = 2
171     elif col == "非常有責任":
172         col = 3
173     else:
174         col = None
175     return col
176
177 varlist7 = ["fnresponsibility_src", "fnresponsibility_press", "fnresponsibility_official",
178 .....:       "fnresponsibility_platform", "fnresponsibility_citizen"]
179
180 for i in varlist7:
181     df_2[i] = df_2[i].apply(responsibility)
182
```

- 一般不連續變數：基本資料類別，包含城市、性別、職業等等，因為彼此之間並沒有有一定的關係，故使用LabelEncoder隨機生成數字並轉換

```

tidying.py X stat.py vdem.png fakenews2.png
tidying.py > varlist9
249 varlist9 = ["city", "sex", "age", "edu", "job", "job_other", "monthincome_self", "monthincome_family",
250           "father", "father_other", "mother", "mother_other", "partner", "partner_other", "carlis",
251           "motorlis", "CDC", "vaccines"]
252
253 label = LabelEncoder()
254
255 for i in varlist9:
256     df_2[i] = label.fit_transform(df_2[i])
257

```

- 存檔：再次寫出為csv檔以利後續作業

## 2. Statis

由於多數問題並沒有同樣的標準差，會造成分析的誤差，故使用robust standard errors來校正再依照下方結果來進行兩次的模型分析，並產出分析結果，結果會呈現在接下來的分析結果中。

```

tidying.py stat.py vdem.png fakenews2.png
stat.py > ...
134
135 y1 = df["fakenews_indicator"]
136
137 varlist1 = ["fakenews1", "fakenews2", "fakenews3", "city", "sex", "age", "edu", "job", "job_other",
138           "incomelevel", "classlevel", "monthincome_self", "monthincome_family", "father",
139           "father_other", "mother", "mother_other", "partner", "partner_other", "carlis", "motorlis"]
140
141
142 x1 = df[varlist1]
143
144 model1 = sm.OLS(y1, x1)
145
146 result1 = model1.fit(cov_type="HC3")

```

```

tidying.py stat.py vdem.png fakenews2.png
stat.py > ...
145
146 result1 = model1.fit(cov_type="HC3")
147
148 with open("models.tex", "w") as models:
149     models.write(result1.summary().as_text())
150     models.close()
151
152 jitter = 0.1
153 df["jittered_fakenews2"] = df["fakenews2"] + np.random.uniform(-jitter, jitter, size=len(df))
154 df["jittered_fnsce_mobile"] = df["fnsce_mobile"] + np.random.uniform(-jitter, jitter, size=len(df))
155 df["jittered_fnsce_nbr"] = df["fnsce_nbr"] + np.random.uniform(-jitter, jitter, size=len(df))
156 df["jittered_fndt_pln"] = df["fndt_pln"] + np.random.uniform(-jitter, jitter, size=len(df))
157 df["jittered_fakenews_indicator"] = df["fakenews_indicator"] + np.random.uniform(-jitter, jitter, size=len(df))
158

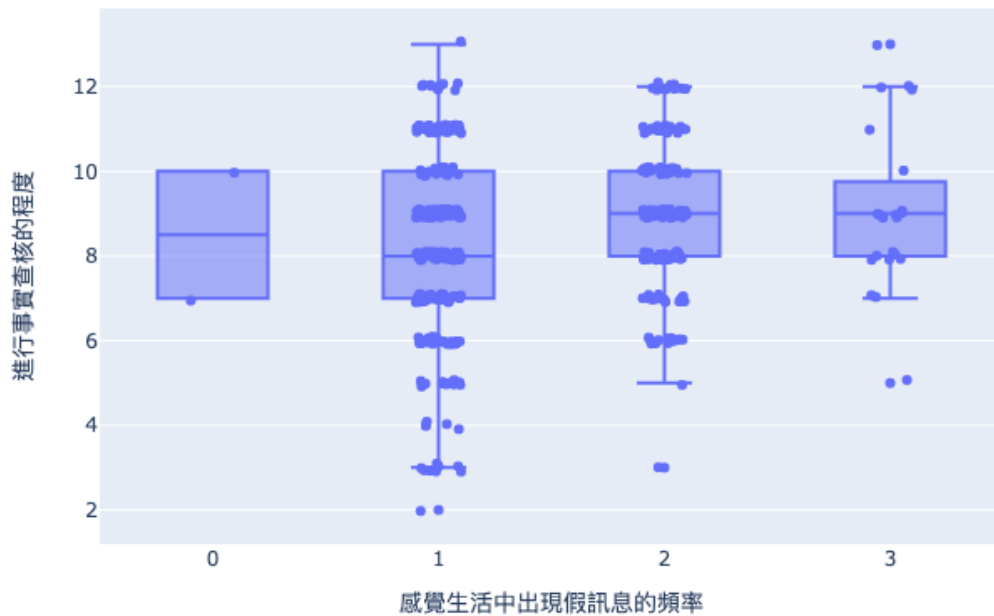
```

## 五. 分析結果

經迴歸分析發現模型一中的Fakenews2（請問您在日常生活當中，常不常出現假消息？）與模型二中的fnsce\_internet（請問您認為以下媒體、網路、手機或社區，是否經常散播假消息？- 網路）與fnsce\_nbr（請問您認為以下媒體、網路、手機或社區，是否經常散播假消息？- 社區鄰里）為顯著結果，以下將不同模型結果以視覺化展示。

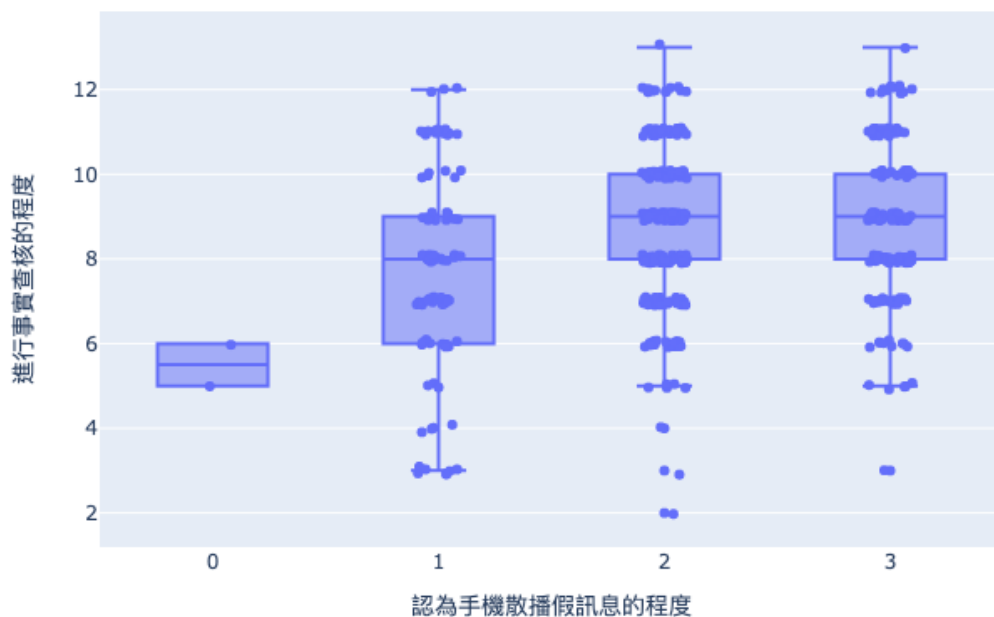
表一、模型一與模型二的OLS Regression Results

	model_1	model_2
Fakenews1	0.1519 (0.780)	
Fakenews2	0.5682*** (0.000)	
Fakenews3	-0.0518 (0.795)	
fnsce_press		-0.2308 (0.211)
fnsce_politician		0.0107 (0.957)
fnsce_citizen		0.2967 (0.067)
fnsce_enemy		-0.0796 (0.544)
fnsce_ngo		0.0994 (0.560)
fnsce_mag		0.0314 (0.853)
fnsce_tv		-0.0393 (0.833)
fnsce_internet		0.1429 (0.513)
fnsce_mobile		0.3808* (0.026)
fnsce_nbr		0.4382** (0.010)
R2	0.952	0.955
F	469.7	369.3
AIC	1922	1980
BIC	2080	2096



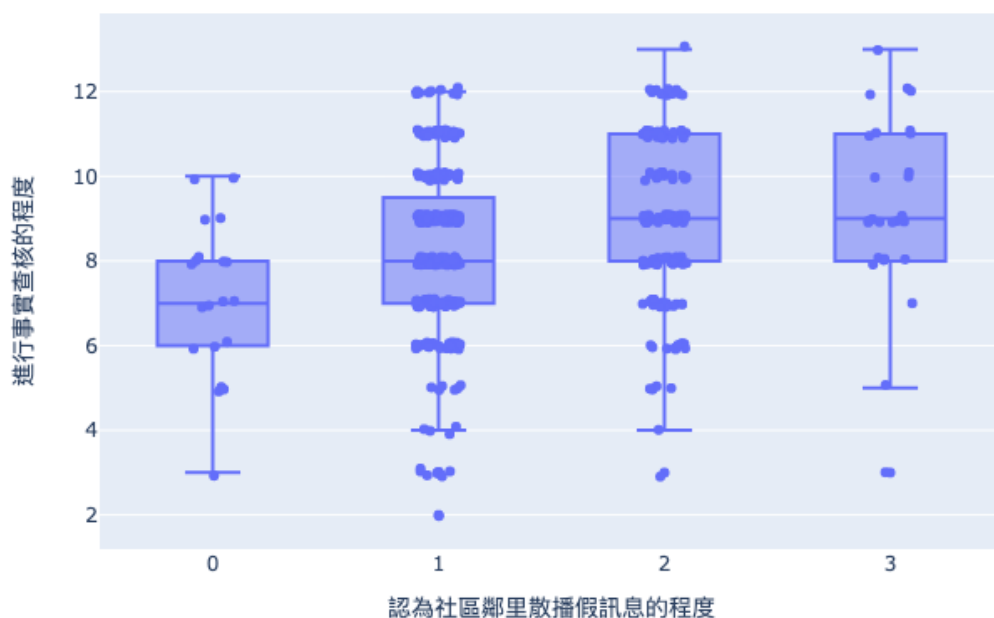
圖二、生活中出現假訊息之頻率對於進行事實查核的能力的箱型圖

根據圖二可以看出，生活中出現假消息的頻率越高，針對事實進行查核的程度亦會越高。



圖三、認為手機散撥假消息的程度對於進行事實查核的能力的箱型圖

根據圖三可以看出，認為手機散撥假消息的程度越高，針對事實進行查核的程度亦會越高。



圖四、認為社區鄰里散撥假消息的程度對於進行事實查核的能力的箱型圖

根據圖四可以看出，認為社區鄰里散撥假消息的程度越高，針對事實進行查核的程度亦會越高。

## 六. 心得感想與未來展望

近年來，我們可以看到台灣人民對假訊息的危機意識逐漸增強。儘管如此，我們仍然需要進一步加強民眾查核不顯著假消息來源的能力。製造假消息的行為涵蓋了各種角色，包括媒體工作者、政治人物、一般民眾、境外勢力和社運團體。而這些假消息則透過報章雜誌、電視及廣播、網路、手機以及社區鄰里等途徑散播開來。

面對這樣的挑戰，我們應該持續加強教育民眾辨識假消息的能力。學校可以在課程中加入媒體素養的教育，教導學生辨識真假訊息的方法和技巧。同時，政府應該提供更多的資訊來源和公開透明的政府資訊，以便民眾能夠自主地查證資訊的真實性。此外，媒體和社交平台也扮演著重要的角色。媒體應該強調事實查證和客觀報導的原則，並避免散播未經證實的消息。社交平台應該加強假消息的監控和移除機制，同時提供使用者更多工具來報告和辨識假消息。

## 七. 參考資料

張佑宗(2022)。臺灣社群媒體中的假新聞、同溫層與極化現象：一種人機整合創新的研究途徑(E10875)【原始數據】取自中央研究院人文社會科學研究中心調查研究專題中心學術調查研究資料庫。