

# MRI and Alzheimers

111007050 傅景裕

112029024 陳暉凱

112029044 蔡碩恩



# OUTLINE

1. 研究背景

2. 問題描述與研究目的

3. 資料說明

4. 資料前處理

5. 資料視覺化分析

6. 資料實驗設計

7. 研究議題和預期結果

8. 參考文獻

9. 工作分配

報告人：傅景裕

# 研究背景

# 阿茲海默症

## 病症描述

- 進行性神經退行性疾病
- 早期：短期記憶缺失
- 中後期：語言、判斷能力慢慢喪失，獨立生活困難

## 病理變化

- $\beta$  - 類澱粉蛋白斑塊
- Tau蛋白纏結的累積



- 神經元損傷死亡、腦部萎縮
- 海馬迴（記憶）、大腦皮質

# MRI

## 結構細節高

- 海馬迴體積
- 皮質厚度
- 全腦體積

## 特徵模型

- 資料探勘、機器學習
- 分類（判別）
- 預測（失智機率）

## 非侵入式、無輻射

- 縱向掃描
- 腦部萎縮速率
- 失智進程
- 病症腦區

# 問題描述與研究目的



# 問題描述與研究目的

- T1 加權 MRI 影像量化資料
- 臨床功能量表分數（如 MMSE、CDR）
- 基本人口統計資料（年齡、性別等）
  - 患者 VS 健康組
  - 橫斷面 VS 縱向研究
- 研究目的：
  - 建立模型協助早期辨識（分類）
  - 預測未來病症發展（回歸）

報告人：傅景裕

# 資料說明



# 橫斷面研究 - 概覽

- 受試者總數：416人
- 資料數：436筆
  - 每位受試者一次訪視算一筆
  - 每次訪視皆提供 3 至 4 次 MRI 影像
  - 再追加「可靠度子資料集」20筆：評估影像測量穩定性
- 年齡：18 至 96 歲、性別：男女皆有、慣用手：皆為右手
- 在 60 歲以上，100 名為極輕度至中度阿茲海默症（其餘皆健康）

# 橫斷面研究 - 資料示意圖

報告人：傅景裕

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	M/F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	Delay
2	OAS1_0001_MR1	F	R	74	2	3	29	0	1344	0.743	1.306	N/A
3	OAS1_0002_MR1	F	R	55	4	1	29	0	1147	0.81	1.531	N/A
4	OAS1_0003_MR1	F	R	73	4	3	27	0.5	1454	0.708	1.207	N/A
5	OAS1_0004_MR1	M	R	28					1588	0.803	1.105	N/A
6	OAS1_0005_MR1	M	R	18					1737	0.848	1.01	N/A
7	OAS1_0006_MR1	F	R	24					1131	0.862	1.551	N/A
8	OAS1_0007_MR1	M	R	21					1516	0.83	1.157	N/A
9	OAS1_0009_MR1	F	R	20					1505	0.843	1.166	N/A
10	OAS1_0010_MR1	M	R	74	5	2	30	0	1636	0.689	1.073	N/A
11	OAS1_0011_MR1	F	R	52	3	2	30	0	1321	0.827	1.329	N/A
12	OAS1_0012_MR1	M	R	30					1574	0.842	1.115	N/A
13	OAS1_0013_MR1	F	R	81	5	2	30	0	1664	0.679	1.055	N/A
14	OAS1_0014_MR1	F	R	19					1525	0.856	1.151	N/A
15	OAS1_0015_MR1	M	R	76	2		28	0.5	1738	0.719	1.01	N/A
16	OAS1_0016_MR1	M	R	82	2	4	27	0.5	1477	0.739	1.188	N/A

# 橫斷面研究 - 類別變項

報告人：傅景裕

變項名稱	型態	中文解釋	補充	範圍	統計摘要
ID	名目	訪視編號	每筆資料對應一次訪視 (每次包含 3-4 張 MRI)	436 唯一值	-
M/F	名目	性別	男、女	M、F	F: 約 64% M: 36%
Hand	名目	慣用手	控制慣用手 降低大腦發展差異	R	全部右手
SES	次序	社經地位	綜合教育與職業等級計算 1=最高、5=最低	1-5	眾數: 3 低社經1、2略多於 高社經 (4、5)

# 橫斷面研究 - 數值變項

報告人：傅景裕

變項名稱	型態	中文解釋	補充	範圍	統計摘要
Age	連續	年齡	-	18-96	平均：77 中位數：78 SD=8.43 分布較對稱平均
Educ	連續	受教育年數	年數越高 教育程度越高	6-23	平均：14.62 SD=2.93 分布較對稱平均
Delay	連續	重新掃描 間隔天數	評估影像測量穩定性 僅 20 名年輕（二十幾歲） 受試者有資料	1-89 限九十天內	平均：20.55 正偏，大部分間隔短

# 橫斷面研究 - 數值變項

報告人：傅景裕

變項名稱	型態	中文解釋	補充	範圍	統計摘要
MMSE	等距	簡易心智 狀態量表	檢查記憶、注意力、語言、 時間定向等認知能力 分數越高，認知越好 小於 24 視為異常	7-30	平均：27.41 中位數：29 明顯負偏 多正常者
CDR	等距	臨床失智症 評分	用於判斷失智症的嚴重程度 0 = 正常、0.5 = 極輕度 1 = 輕度、2 = 中度、3 = 重度	0, 0.5, 1, 2	平均數：0.31 中位數：0 正偏態 多正常者

# 橫斷面研究 - 數值變項

報告人：傅景裕

變項名稱	型態	中文解釋	補充	範圍	統計摘要
eTIV	連續	估計顱內容積	頭顱內可容納大腦的空間大小 數字越大，表示頭顱空間越大	1120–1991	平均：1488.3 SD=145.5 正偏 頭顱容量大者多
ASF	連續	影像縮放係數	校正頭顱大小個體差 頭越大，係數越大	0.881–1.563	平均：1.199 SD=0.129 正偏 頭顱較大者多
nWBV	連續	標準化全腦體積	數字越低 腦占顱內體積佔比越低 腦萎縮越嚴重	0.644–0.893	平均：0.792 SD=0.060 負偏態 多數萎縮少、健康

# 縱向研究 - 概覽

- 受試者總數：150人、資料數：373筆
  - 每次訪視算一筆，每位受試者 2 次以上，每次間隔至少一年
  - 每次訪視皆提供 3 至 4 次 MRI 影像
- 年齡：60 至 96 歲、性別：男女皆有、慣用手：皆為右手
- 失智症狀態分組
  - 72 名為非失智者Nondemented（持續）
  - 64 名為失智者，包含極輕度至中度阿茲海默症Demented
  - 14 名首次訪視非失智，後續轉為失智Converted



# 縱向研究 - 資料示意圖

報告人：傅景裕

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Subject ID	MRI ID	Group	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
2	OAS2_0001	OAS2_0001_MR1	Nondemented	1	0	M	R	87	14	2	27	0	1987	0.696	0.883
3	OAS2_0001	OAS2_0001_MR2	Nondemented	2	457	M	R	88	14	2	30	0	2004	0.681	0.876
4	OAS2_0002	OAS2_0002_MR1	Demented	1	0	M	R	75	12		23	0.5	1678	0.736	1.046
5	OAS2_0002	OAS2_0002_MR2	Demented	2	560	M	R	76	12		28	0.5	1738	0.713	1.01
6	OAS2_0002	OAS2_0002_MR3	Demented	3	1895	M	R	80	12		22	0.5	1698	0.701	1.034
7	OAS2_0004	OAS2_0004_MR1	Nondemented	1	0	F	R	88	18	3	28	0	1215	0.71	1.444
8	OAS2_0004	OAS2_0004_MR2	Nondemented	2	538	F	R	90	18	3	27	0	1200	0.718	1.462
9	OAS2_0005	OAS2_0005_MR1	Nondemented	1	0	M	R	80	12	4	28	0	1689	0.712	1.039
10	OAS2_0005	OAS2_0005_MR2	Nondemented	2	1010	M	R	83	12	4	29	0.5	1701	0.711	1.032
11	OAS2_0005	OAS2_0005_MR3	Nondemented	3	1603	M	R	85	12	4	30	0	1699	0.705	1.033
12	OAS2_0007	OAS2_0007_MR1	Demented	1	0	M	R	71	16		28	0.5	1357	0.748	1.293
13	OAS2_0007	OAS2_0007_MR3	Demented	3	518	M	R	73	16		27	1	1365	0.727	1.286
14	OAS2_0007	OAS2_0007_MR4	Demented	4	1281	M	R	75	16		27	1	1372	0.71	1.279
15	OAS2_0008	OAS2_0008_MR1	Nondemented	1	0	F	R	93	14	2	30	0	1272	0.698	1.38
16	OAS2_0008	OAS2_0008_MR2	Nondemented	2	742	F	R	95	14	2	29	0	1257	0.703	1.396

# 縱向研究 - 數值變項

變項名稱	型態	中文解釋	補充	範圍	統計摘要
Subject ID	名目	受試者編號	同個受試者不同訪視次數 受試者編號相同	1-150 唯一值	-
MRI ID	名目	訪視編號	每個受試者至少兩次訪視 每次訪視會有一個不同MRI ID	1-373 唯一值	-
Visit	名目	訪視次序	訪視第幾次 每位至少 2 次訪視	1-5	眾數為二
M/F	名目	性別	男、女	M/F	F: 60%、M: 40%

# 縱向研究 - 數值變項

報告人：傅景裕

變項名稱	型態	中文解釋	補充	範圍	統計摘要
Hand	名目	慣用手	控制左右腦差異	R	全部右手
Group	名目	組別	Nondemented：一直正常 Demented：首次訪視便確診 Converted：正常後變確診	Nondemented Demented Converted	Nondemented: 74 Demented: 64 Converted: 14
SES	次序	社經地位	1＝最高；5＝最低	1-5	眾數：3 低略多於高

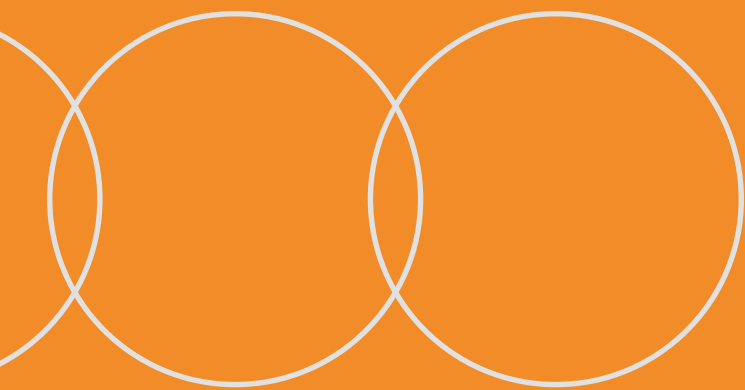
# 縱向研究 - 數值變項

變項名稱	型態	中文解釋	補充	範圍	統計摘要
MMSE	等距	簡單心智 狀態量表	≥24 視為正常	0–30	平均：約26 中位數：約28 明顯負偏（健康多）
CDR	等距	臨床失智症 評分	0＝正常、0.5＝極輕度 1＝輕度、2＝中度	0、0.5、1、2	平均：約 0.45 中位數：0 正偏（健康者多）
Delay	連續	訪視間隔 天數	訪視相隔至少一年	300–1500	平均：約 450–550 正偏 大部分間隔較短

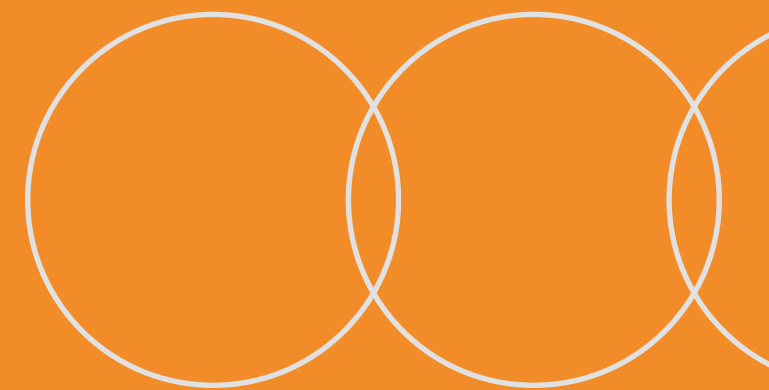
# 縱向研究 - 數值變項

報告人：傅景裕

變項名稱	型態	中文解釋	補充	範圍	統計摘要
eTIV	連續	估計顱內容積	頭顱空間大小 數字越大，頭越大	約 1100–2000	平均：約 1470 SD≈140 正偏 頭顱容量偏小者略多
ASF	連續	影像縮放係數	係數越大，頭越大	約 0.85–1.6	平均：約 1.20 SD≈0.12 正偏、頭較小者略多
nWBV	連續	標準化全腦體積	數值越低，腦萎縮越嚴重	約 0.60–0.90	平均：約 0.77 SD≈0.065 負偏 多數體積較大、健康



# 資料前處理



# Excel 重整資料 - 遺漏值處理

報告人：陳偉凱

## Cross-sectional

- ses, educ : 缺失比例過高: 缺失比例過高
- mmse, CDR : 臨床量測值

```
,Missing Count,Missing Percent (%)  
Delay,416,95.41  
SES,220,50.46  
Educ,201,46.1  
MMSE,201,46.1  
CDR,201,46.1
```

## Longitudinal

- SES、MMSE : 缺失比例小，不會造成樣本量大量流失

```
,Missing Count,Missing Percent (%)  
SES,19,5.09  
MMSE,2,0.54
```



# Excel 重整資料

報告人：陳偉凱

刪除欄位：

## 1. SES

統計證據：SES 與 EDUC 高度相關

LONGITUDINAL：R = -0.72

CROSS-SECTIONAL：R = -0.74

Longitudinal

## 2. ASF

統計證據：ASF 與 ETIV 幾乎完全共線

LONGITUDINAL：R = -0.989（超強負相關）

CROSS-SECTIONAL：R = -0.976

A	B	C	D	E	F	G	H	I	J	K
Subject ID	MRI ID	Group	Visit	M/F	Age	EDUC	MMSE	CDR	eTIV	nWBV
OAS2_000	OAS2_000	Nondemen	1	M	87	14	27	0	1987	0.696
OAS2_000	OAS2_000	Nondemen	2	M	88	14	30	0	2004	0.681
OAS2_000	OAS2_000	Demented	1	M	75	12	23	0.5	1678	0.736

## 3. HAND

與 CDR、MMSE、NWBV 完全無關

Cross-sectional

ID	M/F	Age	EDUC	MMSE	CDR	eTIV	nWBV
OAS1_000	F	74	2	29	0	1344	0.743
OAS1_000	F	55	4	29	0	1147	0.81
OAS1_000	F	73	4	27	0.5	1454	0.708
OAS1_001	M	74	5	30	0	1636	0.689

## 4. DELAY / MR DELAY

缺失值過多，且與疾病無關

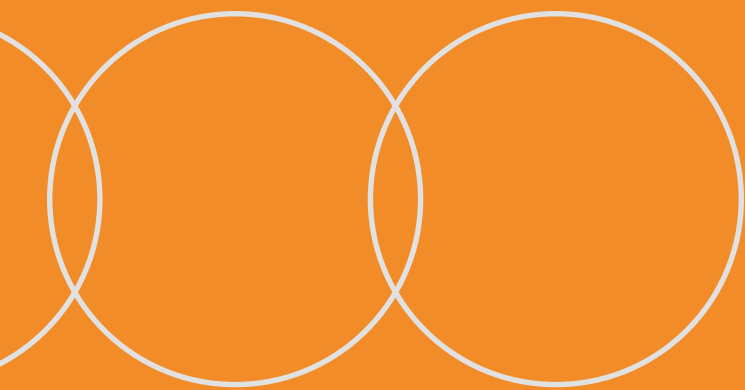
# Excel 重整資料

標準化：

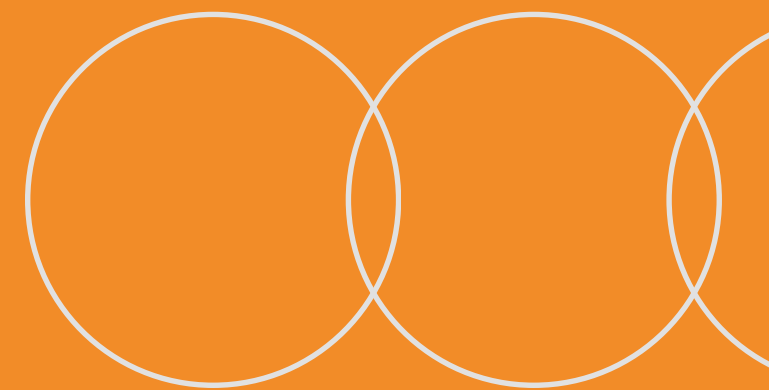
特徵尺度差異太大

報告人：陳偉凱

Age	MMSE	CDR	eTIV	nWBV
-0.3946	0.450668	0	-0.83475	0.36242
-2.87808	0.450668	0	-1.96204	2.164513
-0.52531	-0.09306	0.5	-0.20529	-0.57897
-0.3946	0.722535	0	0.836167	-1.09001
-3.27021	0.722535	0	-0.96636	2.621761
0.520373	0.722535	0	0.996392	-1.35898
-0.13318	0.178802	0.5	1.419843	-0.28311
0.651083	-0.09306	0.5	-0.07368	0.254832
-4.96944	0.178802	0	0.836167	2.245204
1.566052	0.722535	0	0.263936	-0.39069
-3.79305	0.450668	0	-0.93775	1.49209
0.389663	-1.18053	0.5	1.740293	0.954152
-1.04815	-1.18053	0.5	-0.24535	0.738977
0.651083	-0.09306	0.5	-0.39985	-0.52518
-2.48595	0.722535	0	-1.45848	2.433483
1.173923	-0.09306	1	-0.23391	0.227935
-1.57098	0.450668	0	-0.56008	0.927255
1.435342	-0.36493	1	-0.40557	-1.49347
1.566052	0.178802	0	0.807556	-1.27829
0.389663	0.450668	0	-0.95492	0.147244
-3.40092	0.450668	0	0.275381	2.729349
0.912503	0.178802	1	-0.50285	-0.92863
-0.91744	0.450668	0.5	-0.15379	1.14243



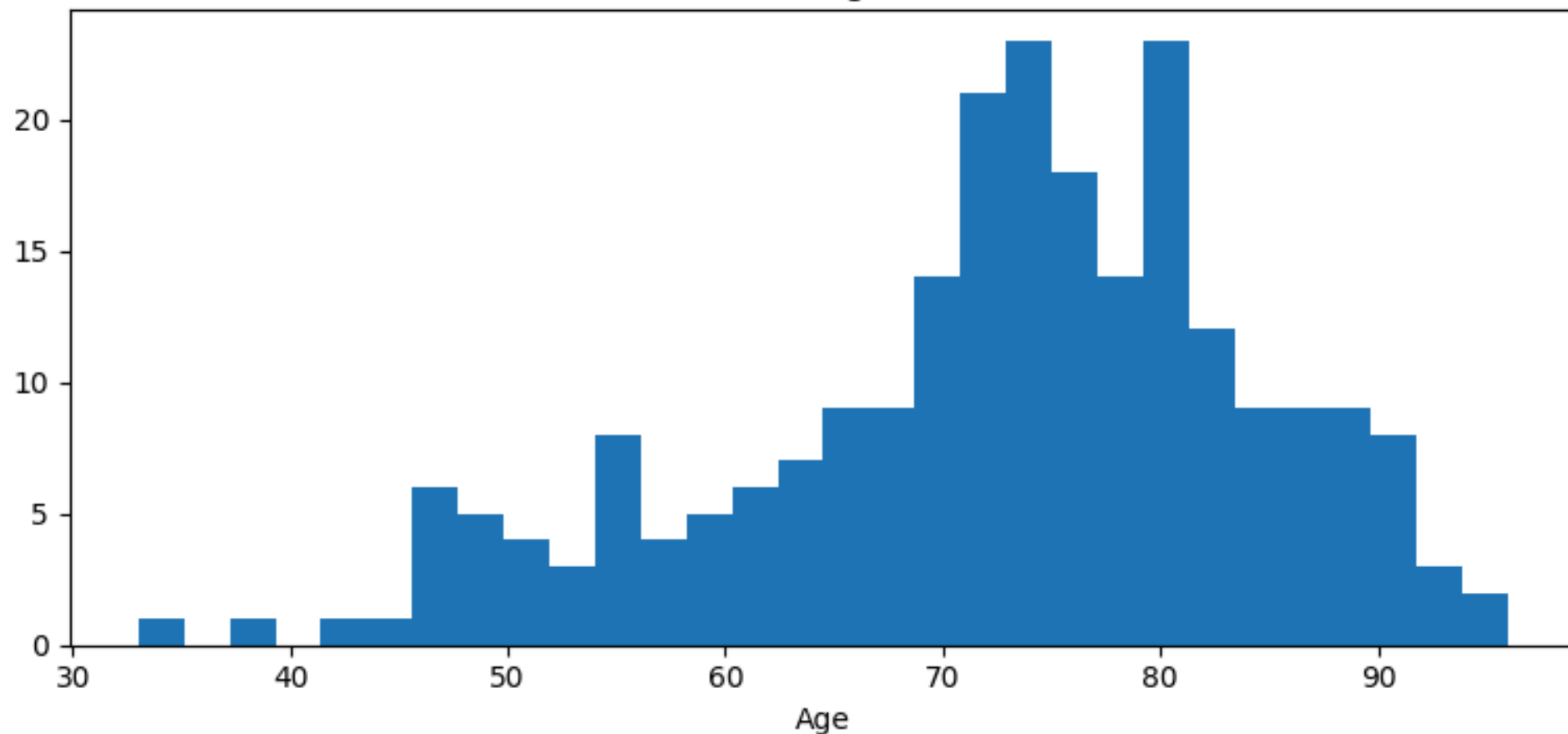
# 資料視覺化分析



# AGE-資料分析

報告人：陳曄凱

Cross-sectional Age Distribution



Cross-sectional

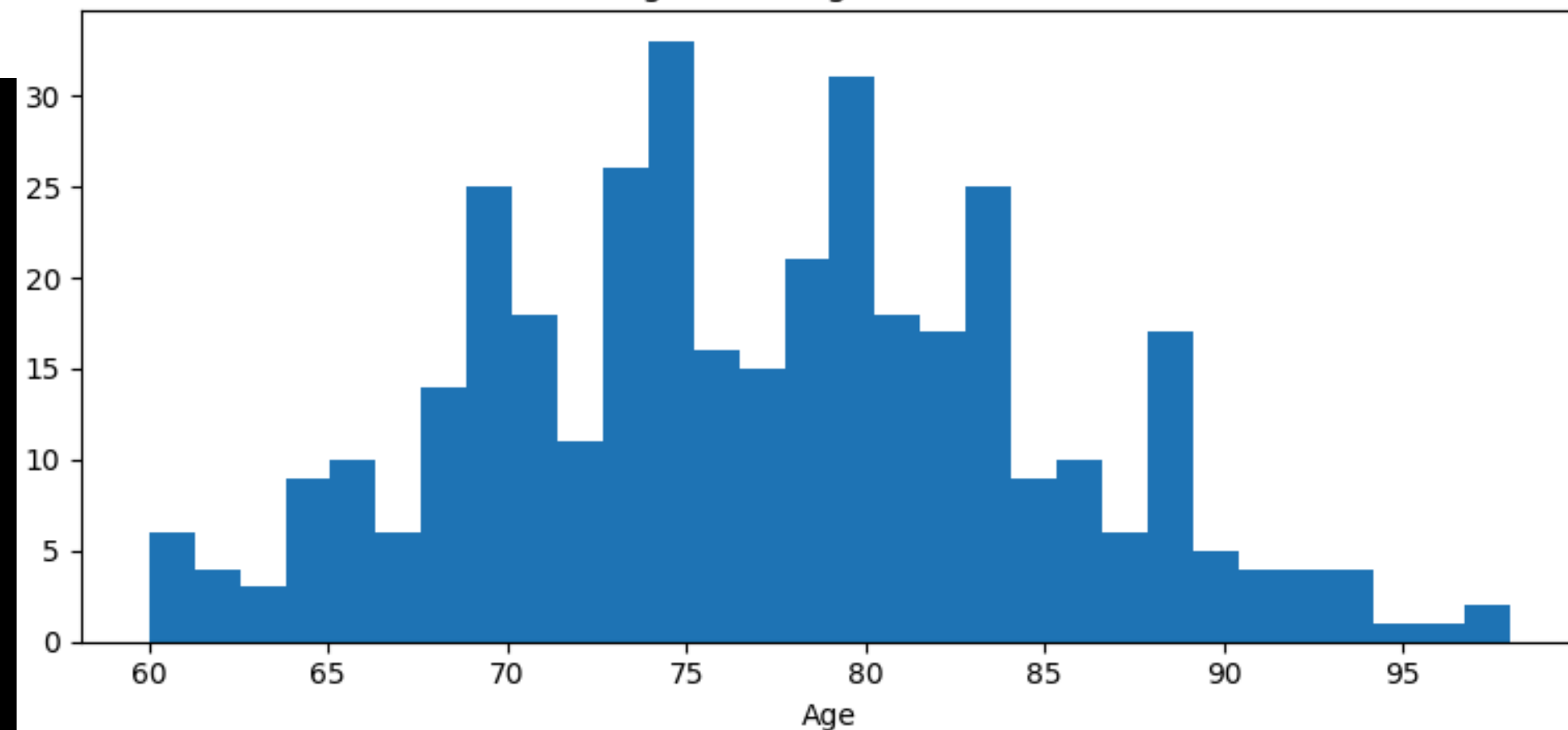
平均值 72.3

標準差 12.09

範圍 33 - 96

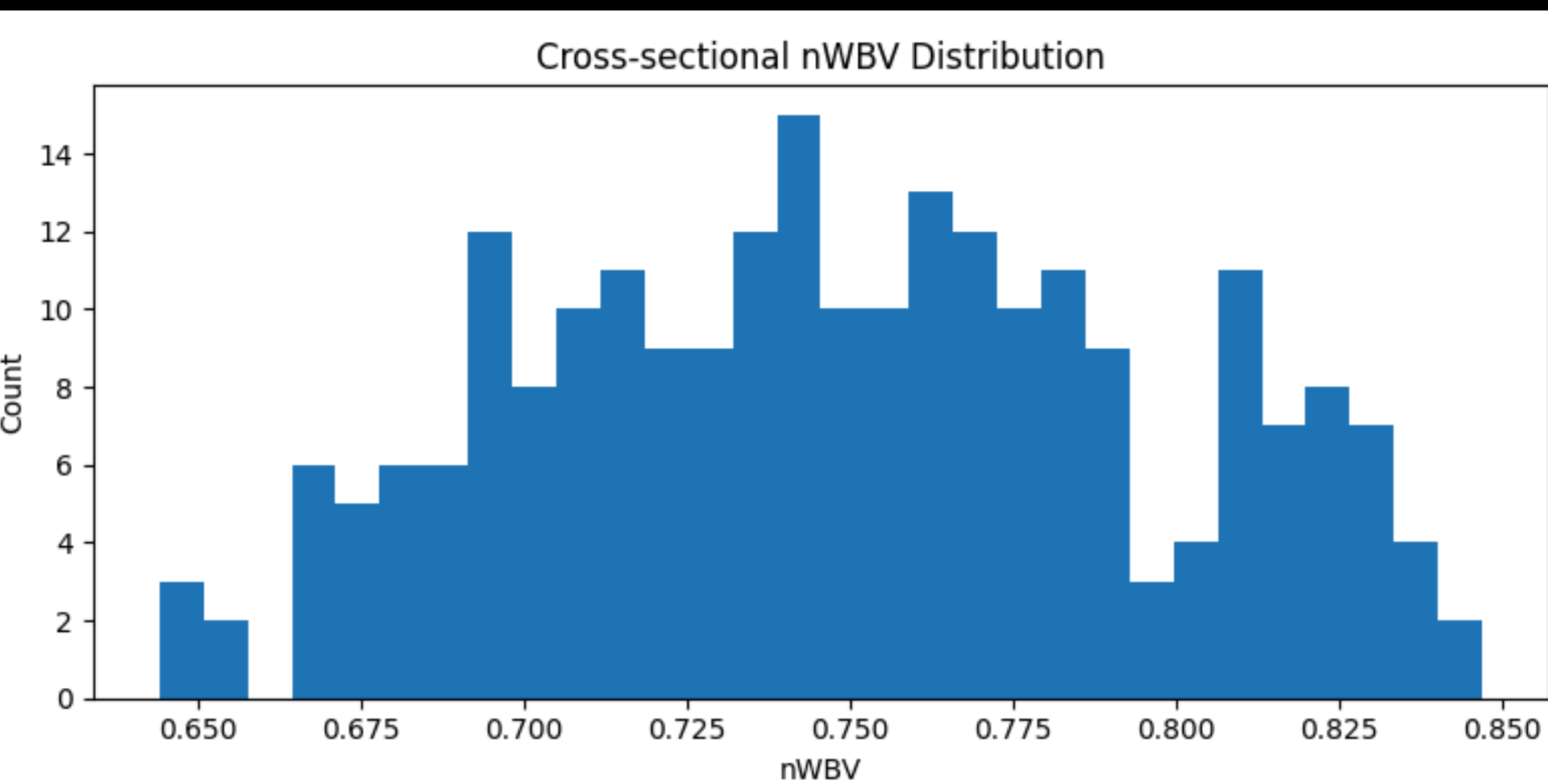
Longitudinal  
平均值 77.0  
標準差 7.66  
範圍 60 - 98

Longitudinal Age Distribution



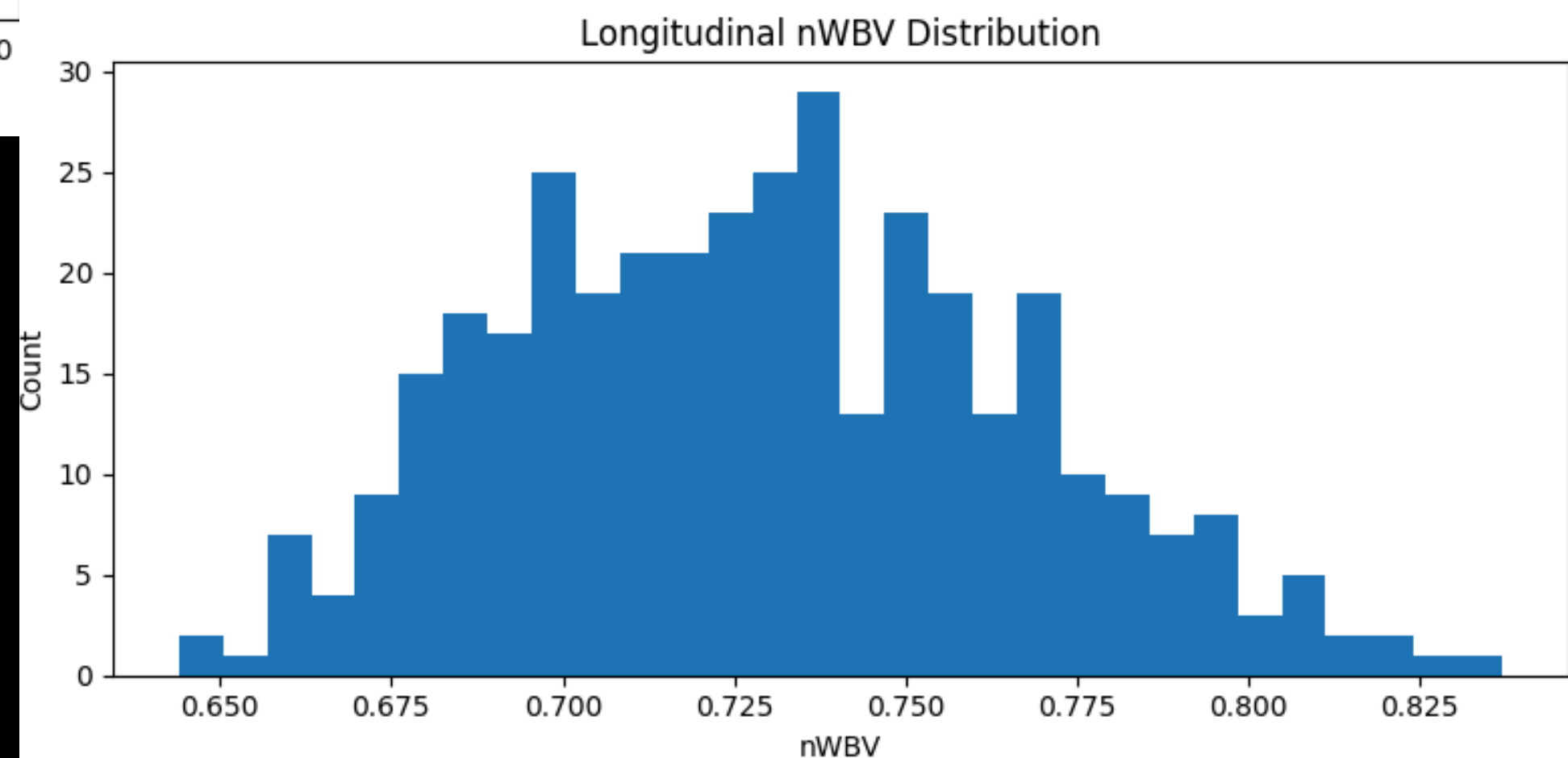
# nWBV-資料分析

報告人：陳曄凱



Cross-sectional  
平均值 0.7491  
標準差 0.0479  
平均腦體積略高，但分布更廣

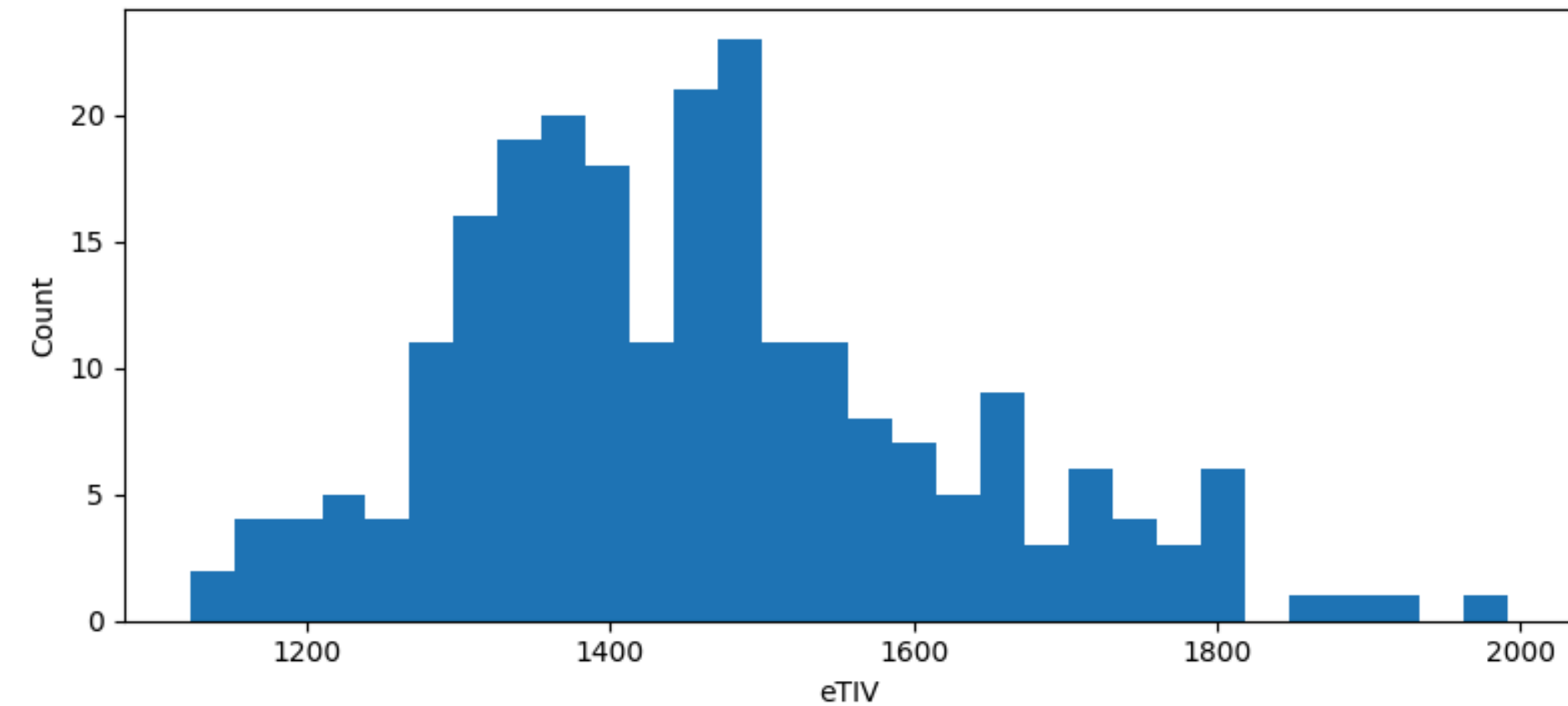
Longitudinal  
平均值 0.7295  
標準差 0.0372  
萎縮者更多偏向「追蹤長期退化的族群」。



# eTIV-資料分析

報告人：陳曄凱

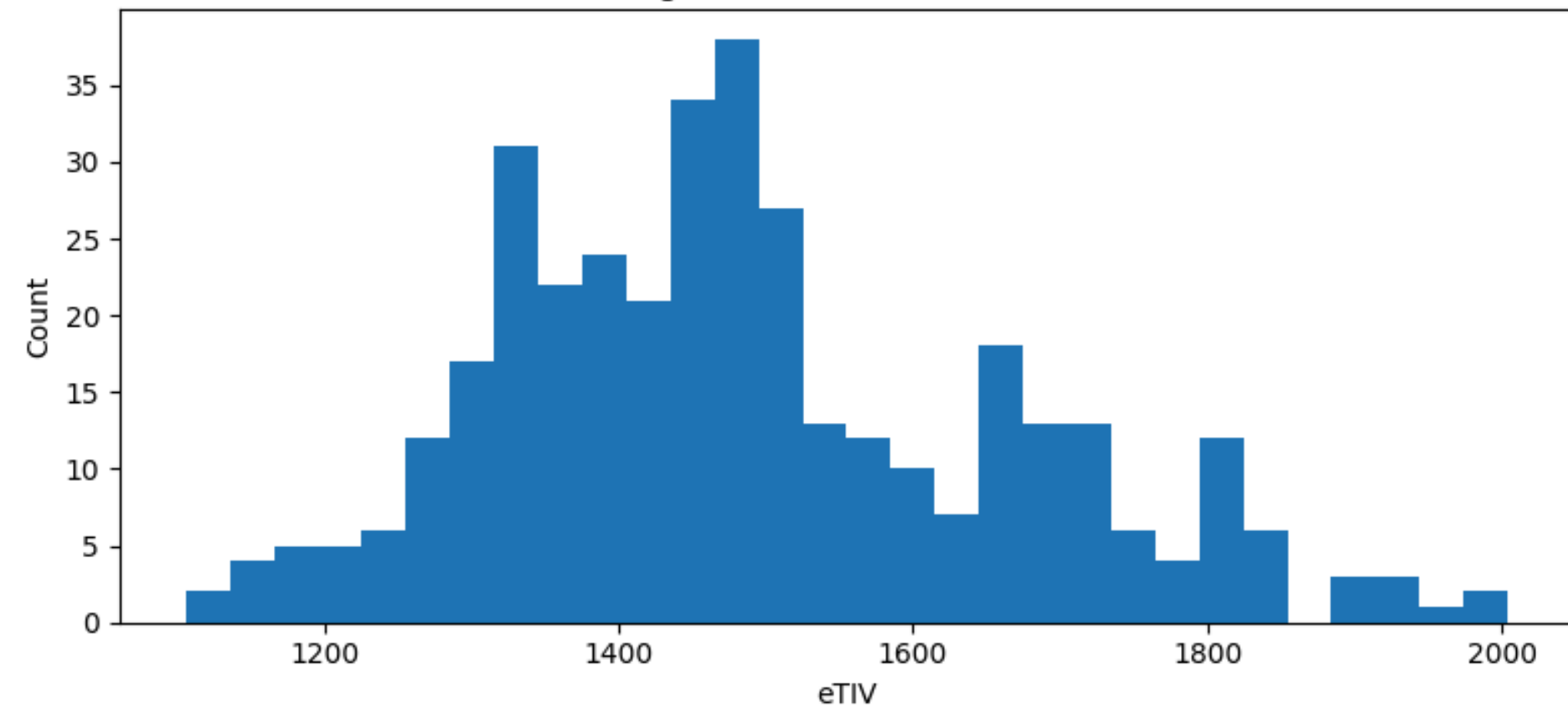
Cross-sectional eTIV Distribution



Cross-sectional  
平均值 1459.5  
標準差 160.36

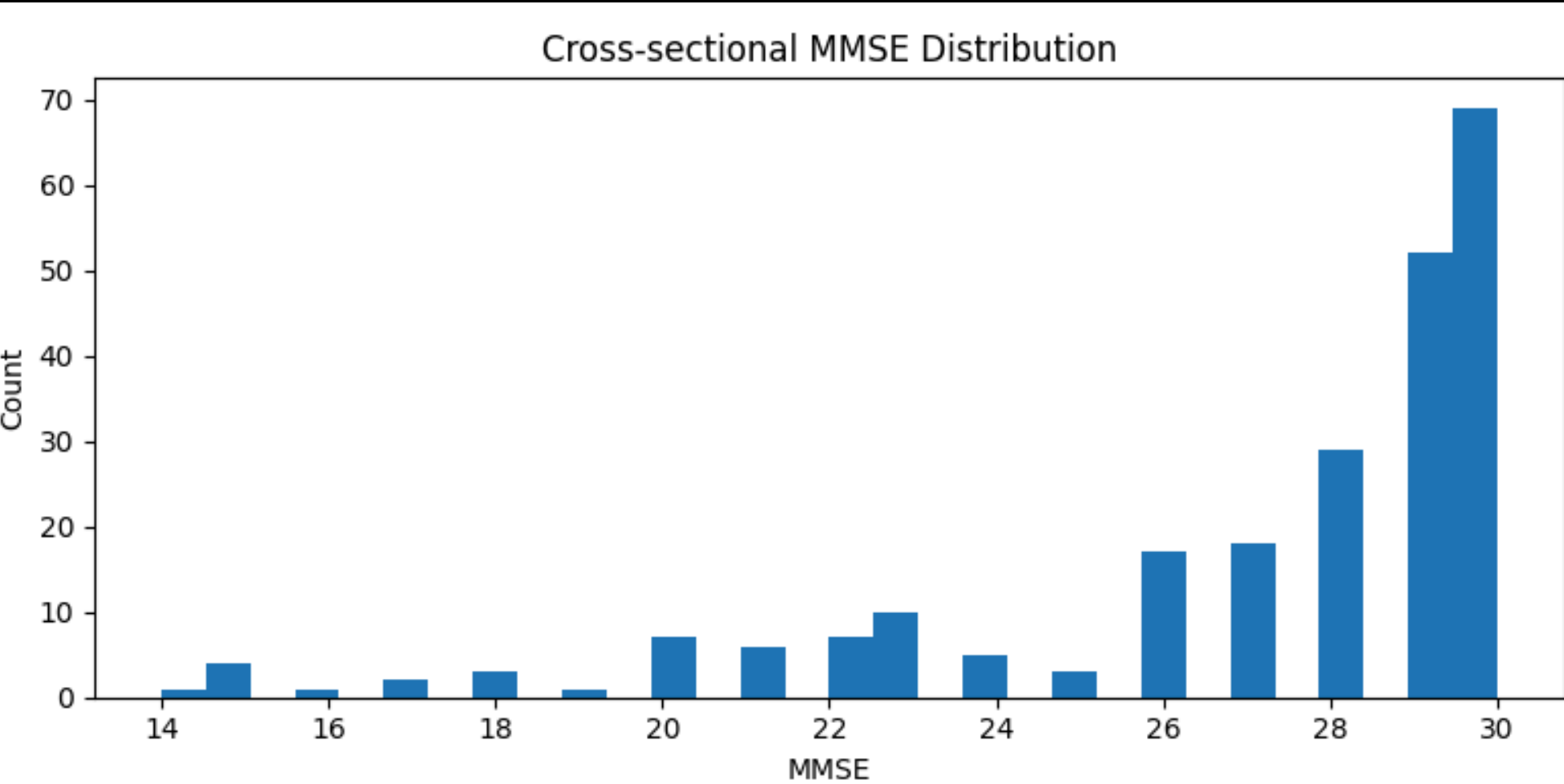
Longitudinal  
平均值 1489.9  
標準差 174.99

Longitudinal eTIV Distribution



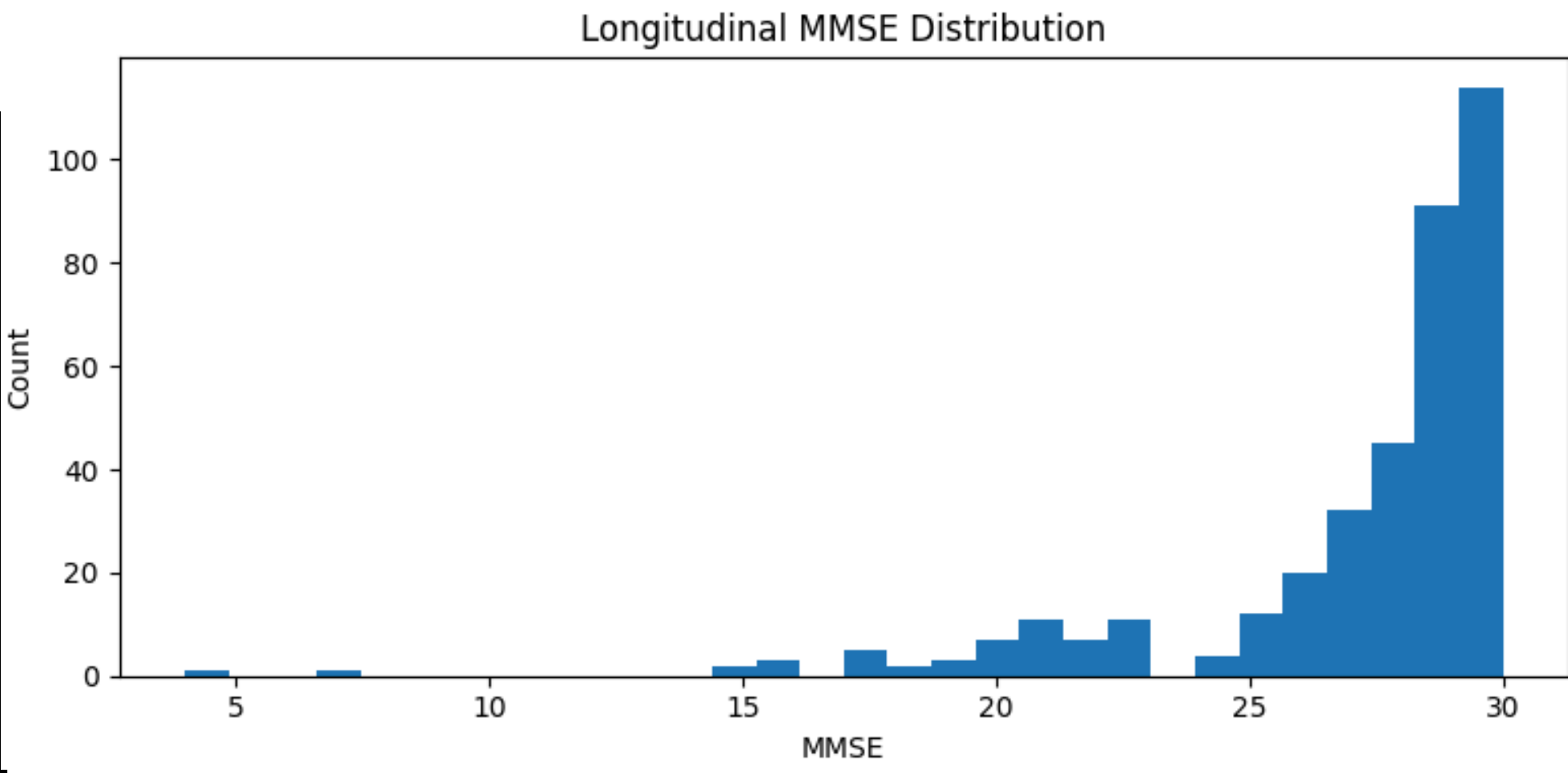
# MMSE-資料分析

報告人：陳偉凱



Cross-sectional

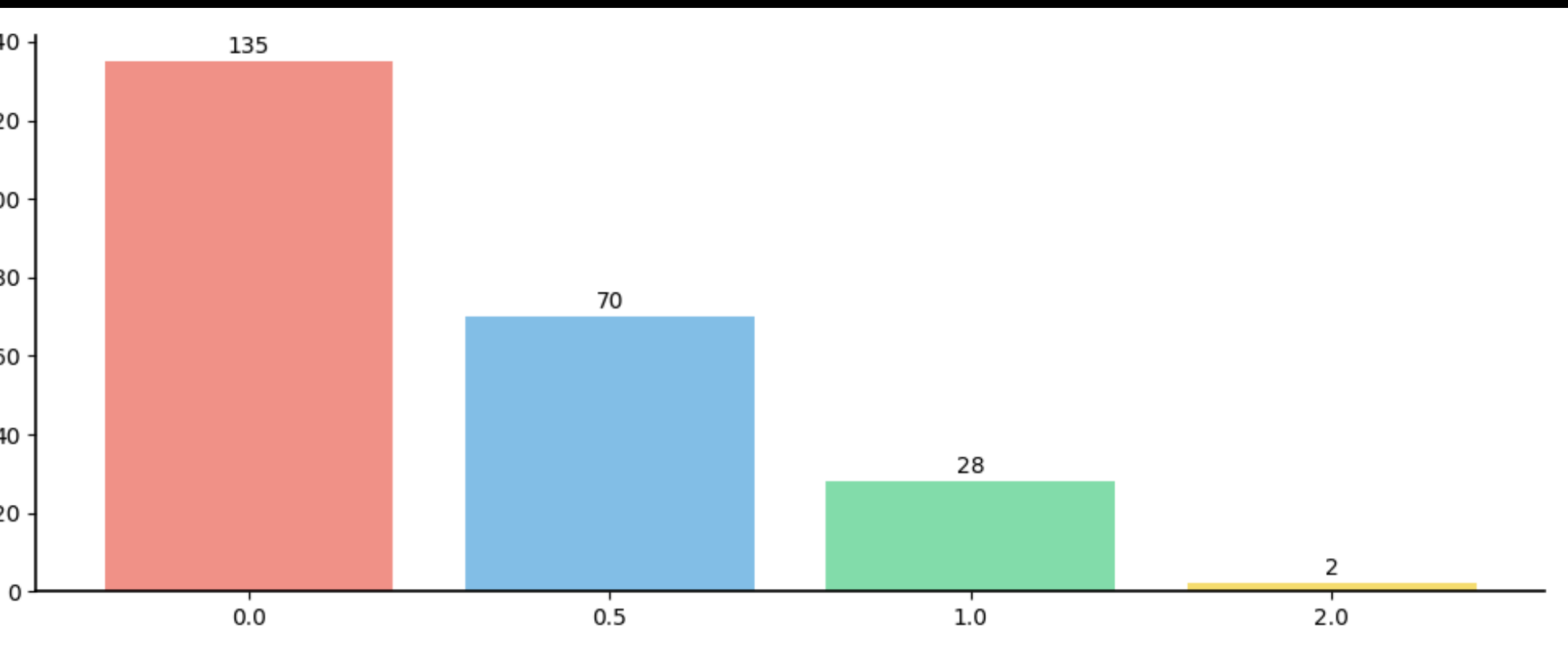
平均值	27.06
標準差	3.69
最低分	14
中位數	29





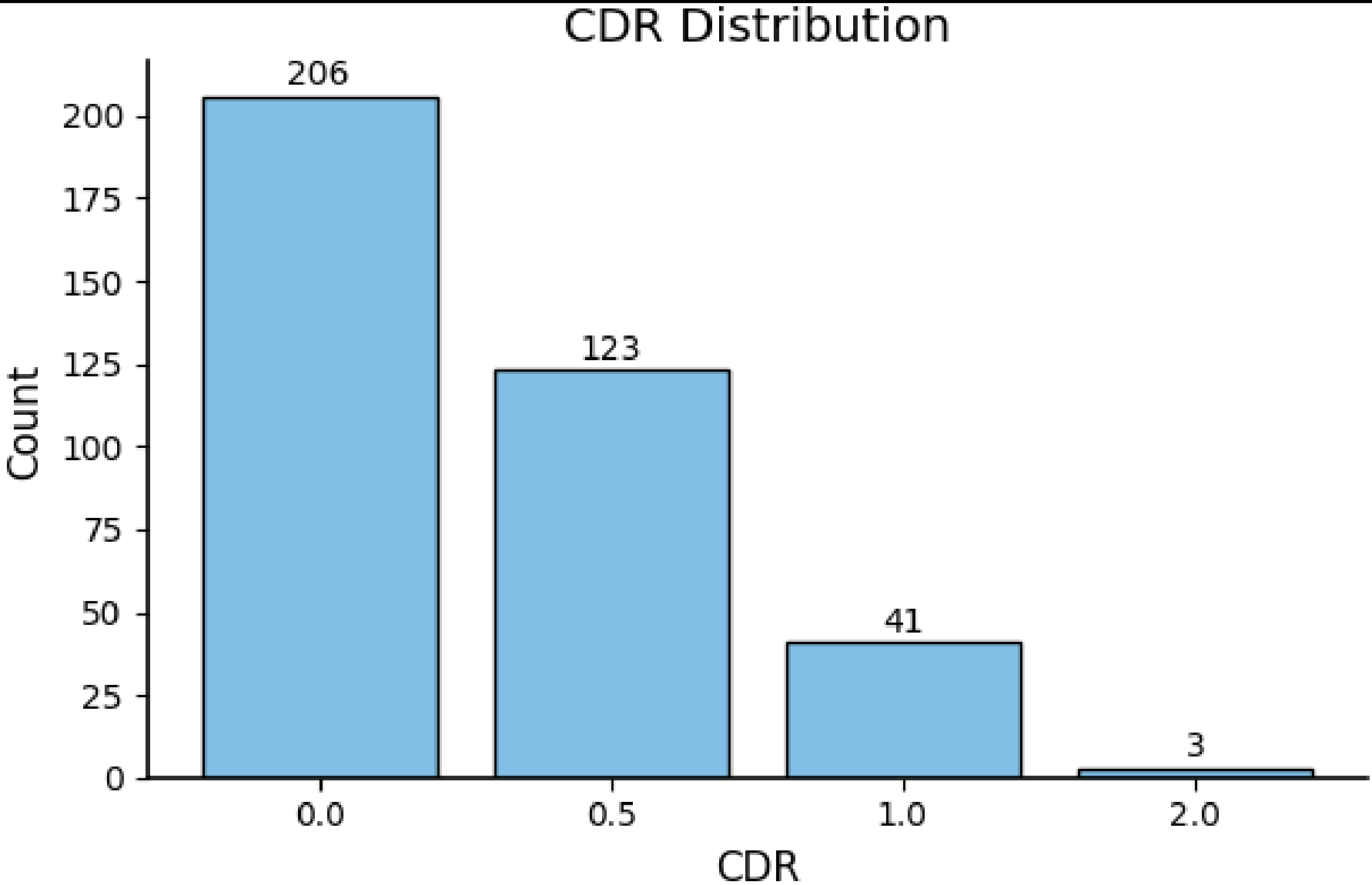
# CDR-資料分析

報告人：陳曄凱



Cross-sectional  
平均值 0.285  
中位數 0

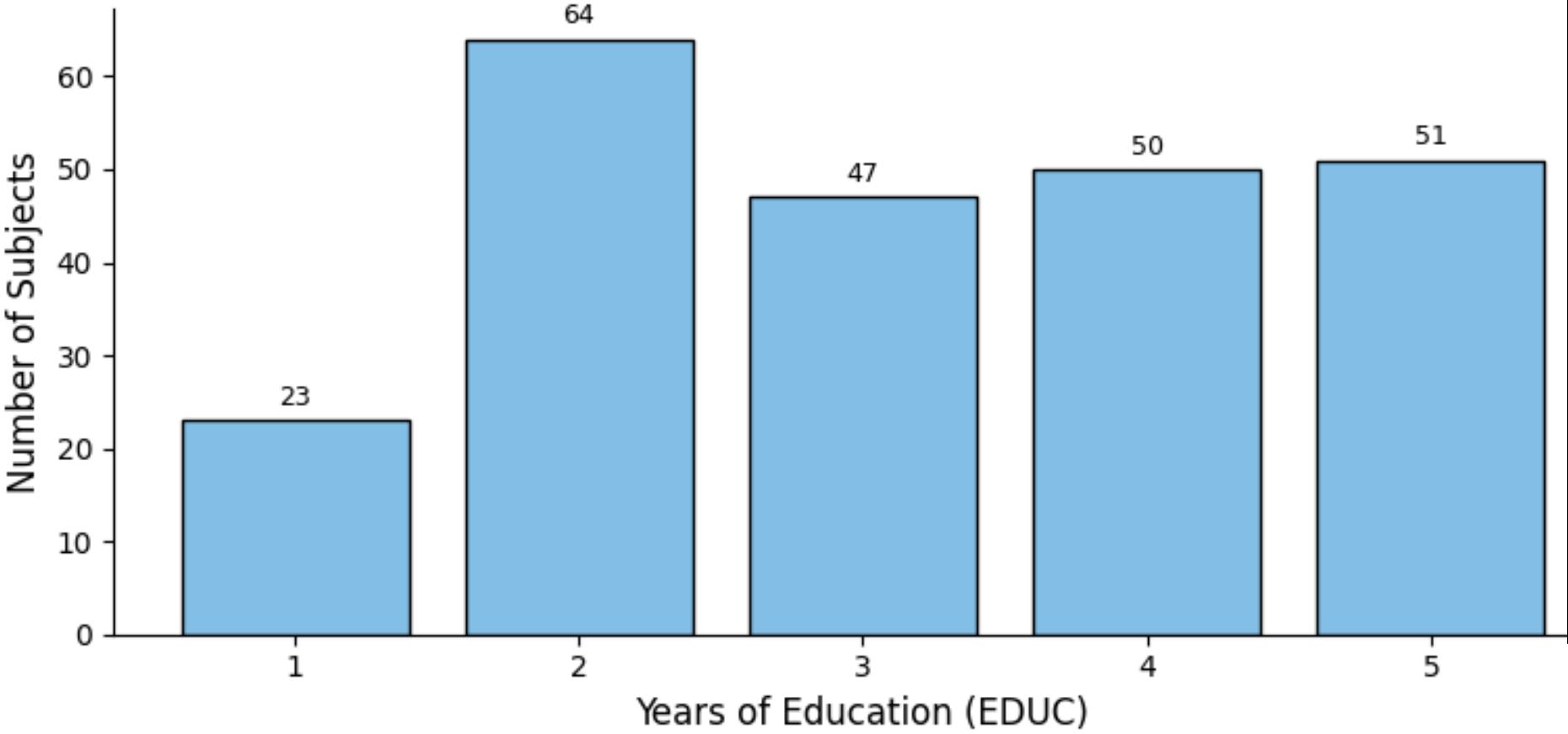
Longitudinal  
平均值 0.287  
中位數 0



# EDUC-資料分析

報告人：陳曄凱

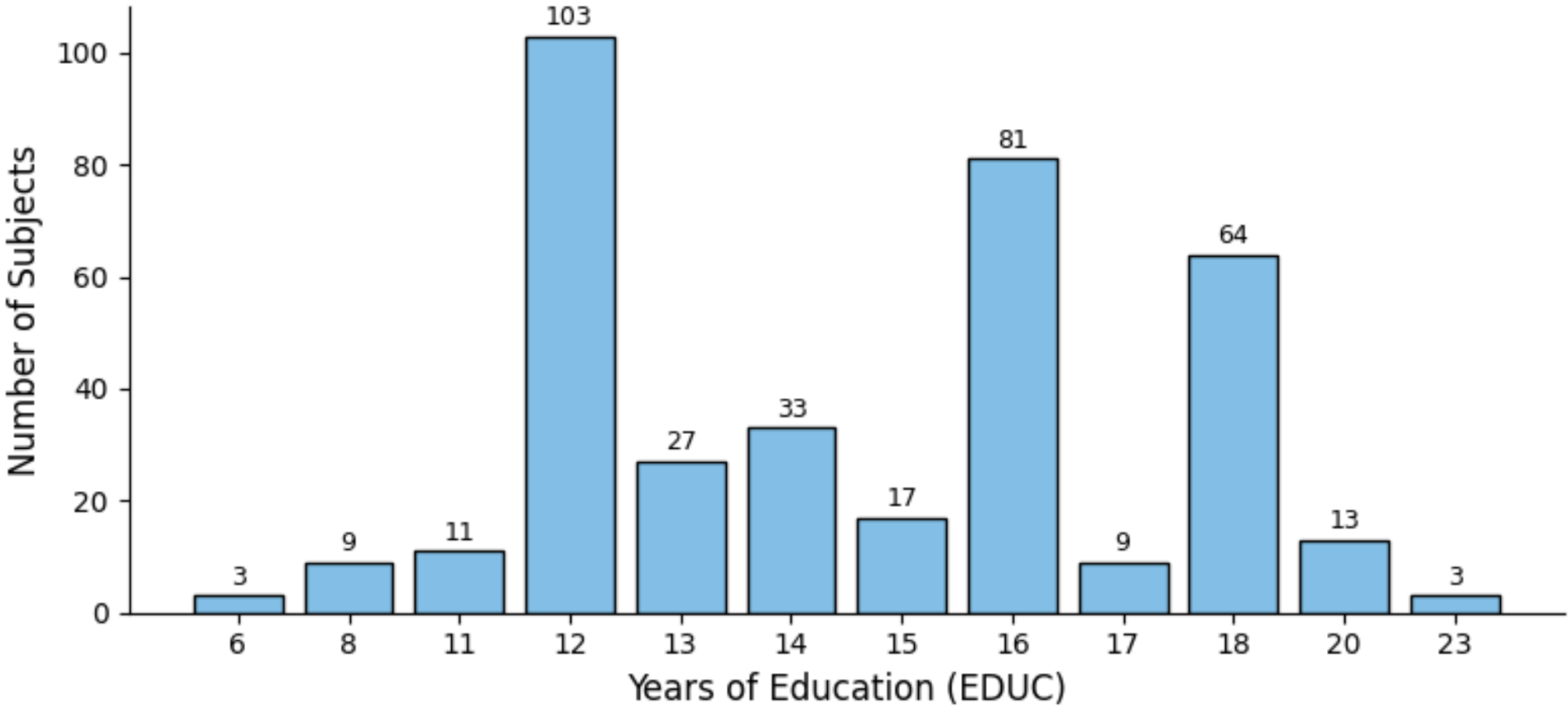
Cross-sectional EDUC Distribution



Cross-sectional  
平均值 3.18  
標準差 1.31  
範圍 1 - 5

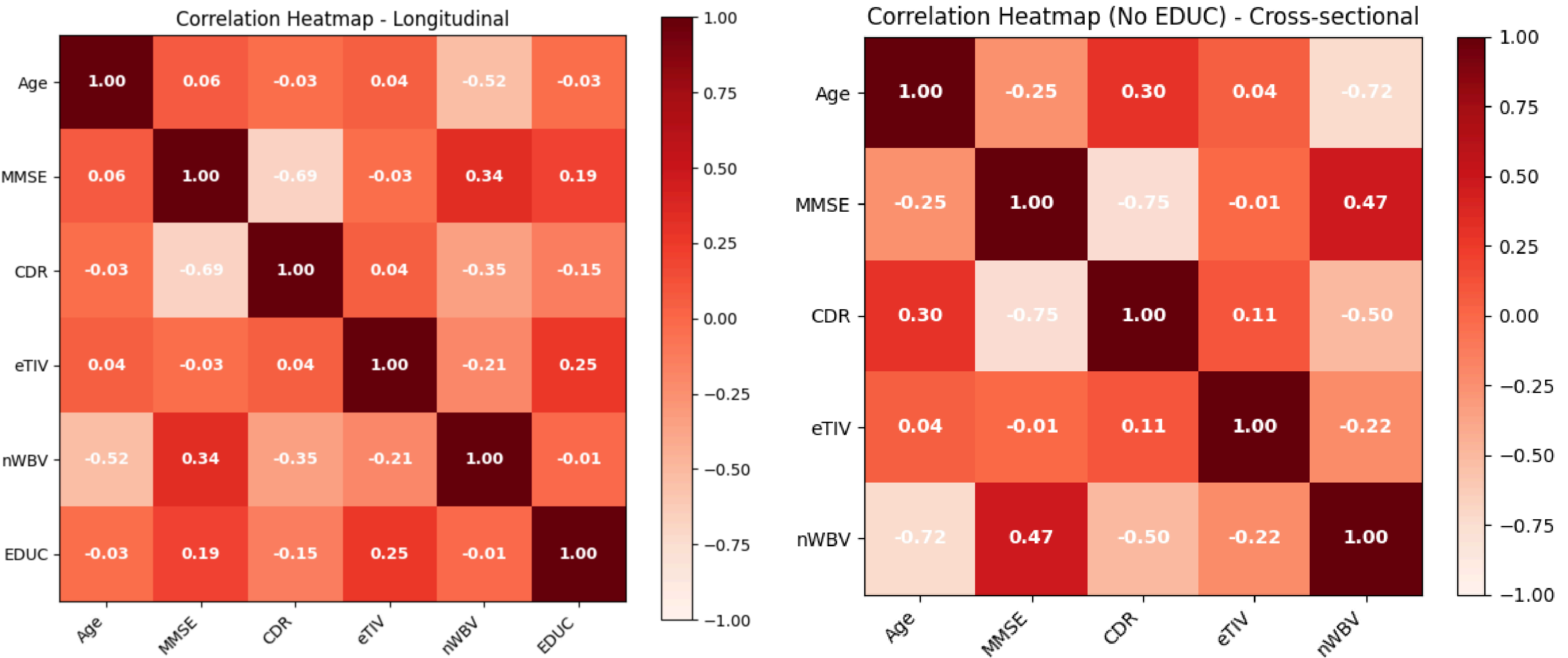
Longitudinal  
平均值 14.61  
標準差 2.87  
範圍 6 - 23

Longitudinal EDUC Distribution



# cross-sectional(有無EDUC加入)

報告人：陳曄凱



# 熱力圖解釋

報告人：陳暉凱

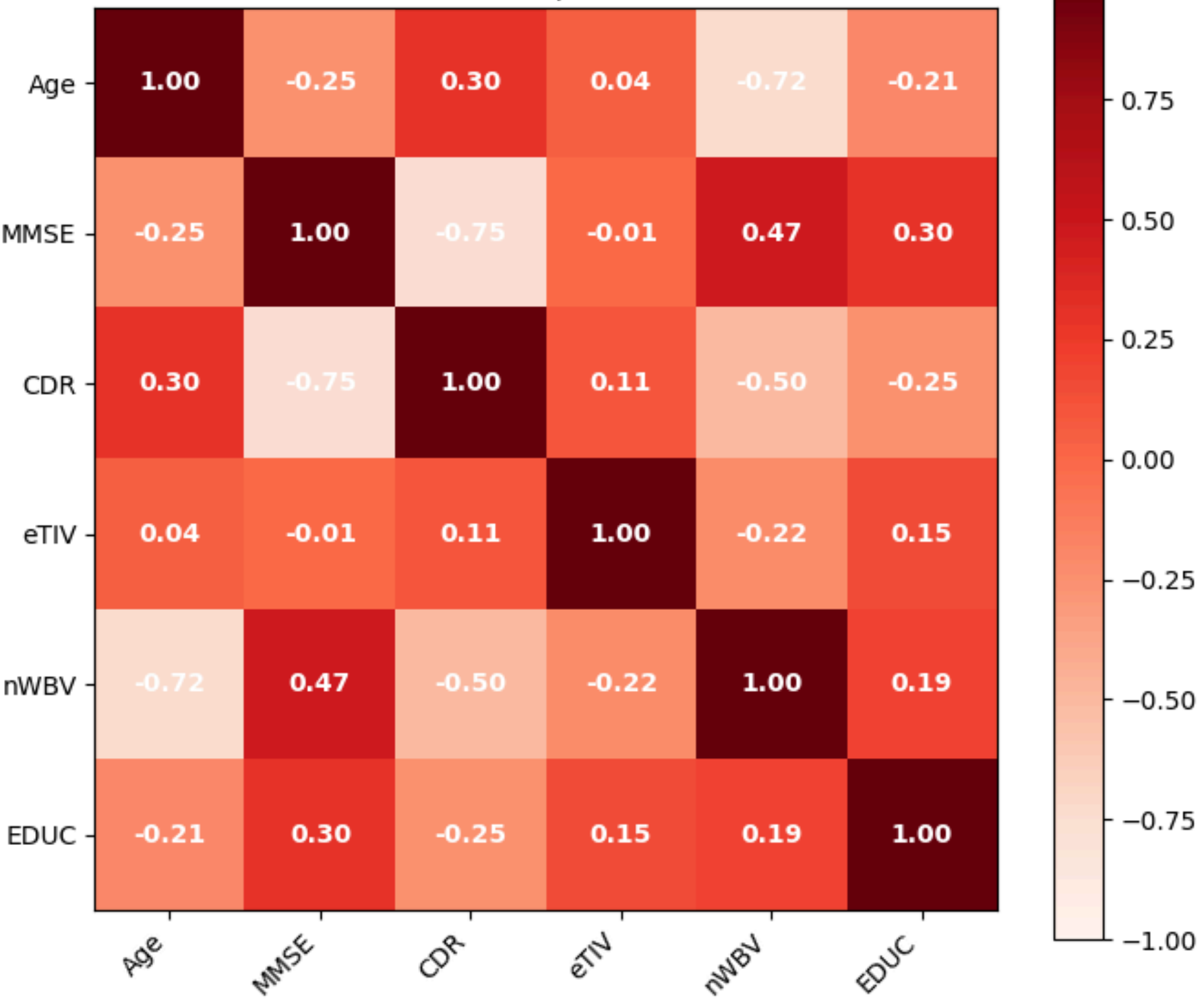
## Cross-sectional有無EDUC加入關鍵變數之差異

關鍵變數組合	無 EDUC	有 EDUC	是否改變
Age – nWBV	-0.72	-0.72	無
MMSE – CDR	-0.75	-0.75	無
CDR – nWBV	-0.50	-0.50	無
MMSE – nWBV	0.47	0.47	無
Age – MMSE	-0.25	-0.25	無

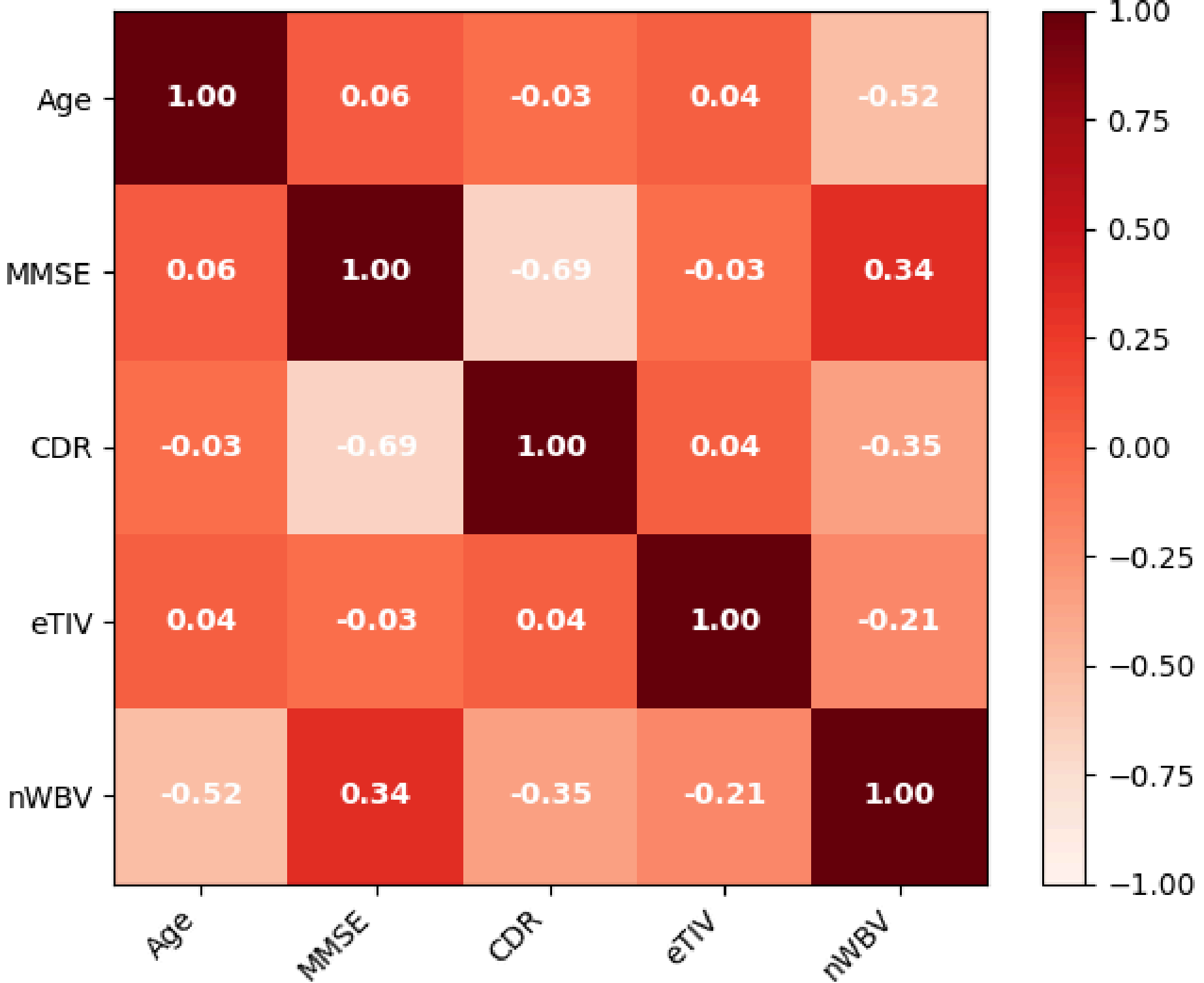
# Longitudinal (有無EDUC加入)

報告人：陳曄凱

Correlation Heatmap - Cross-sectional



Correlation Heatmap (No EDUC) - Longitudinal



# 熱力圖解釋

報告人：陳暉凱

## Longitudinal有無EDUC加入關鍵變數之差異

關鍵變數組合	無 EDUC	有 EDUC	是否改變
Age - nWBV	-0.52	-0.52	無
MMSE - CDR	-0.69	-0.69	無
CDR - nWBV	-0.35	-0.35	無
MMSE - nWBV	0.34	0.34	無
Age - MMSE	0.06	0.06	無

## Age（年齡）與關鍵變數的相關性

資料集	與 nWBV	與 MMSE	與 CDR
Longitudinal	-0.518	+0.056	-0.025
Cross-sectional	-0.720	-0.25	+0.30

- 1. 年紀越大 → 腦容量比例（nWBV）越低（萎縮）
- 2. Cross-sectional 更明顯，因為樣本中失智患者比例較高



## MMSE與關鍵變數的相關性

資料集	與 CDR	與 nWBV
Longitudinal	-0.686	+0.34
Cross-sectional	<b>-0.75</b>	<b>+0.47</b>

- 1. CDR 越高 → MMSE 下降明顯
- 2. nWBV 越大 → MMSE 越高
- 3. MMSE 是研究中最強的認知指標

# 熱力圖解釋

報告人：陳瑋凱

## CDR與關鍵變數相關性

資料集	與 MMSE	與 nWBV
Longitudinal	強烈負相關	-0.35
Cross-sectional	強烈負相關	-0.50

- 1. CDR 越高 → 腦體積下降 → 認知下降
- 2. 失智嚴重度的核心變項

# 熱力圖解釋

報告人：陳瑋凱

## nWBV與關鍵變數的相關性

資料集	與 Age	與 MMSE	與 CDR
Longitudinal	-0.52	+0.34	-0.35
Cross-sectional	-0.72	+0.47	-0.50

nWBV 反映三項核心

- 1. 年齡（萎縮）
- 2. 認知（越大越好）
- 3. 失智（越低越好）

→ 是模型中最具代表性的特徵

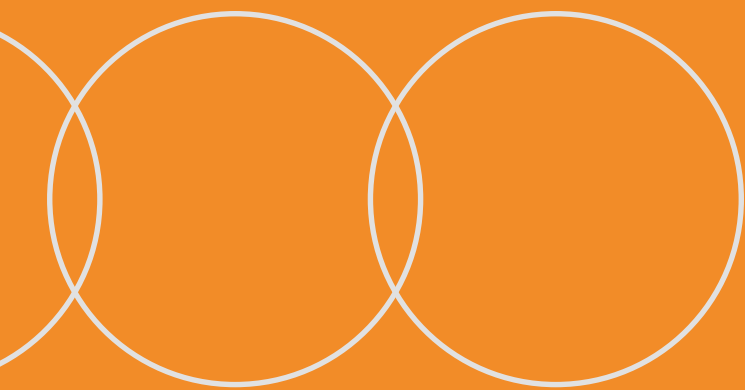
# 熱力圖解釋

報告人：陳瑋凱

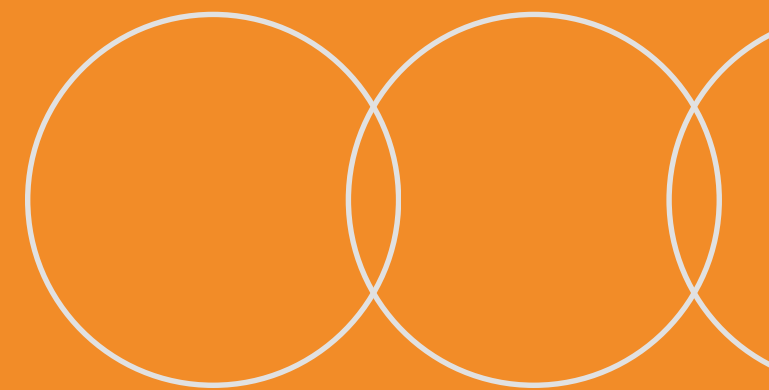
## eTIV與其他變項的相關性

資料集	與其他變項
Longitudinal	$\pm 0.20$ 以內
Cross-sectional	$\pm 0.20$ 以內

- 1. 幾乎無關係。
- 2. 主要反映 anatomical 差異（頭顱大小）
- 3. 與失智無直接關聯。



# 資料切分



# STEP1.訓練／驗證／測試資料集的分工

報告人：陳曄凱

## Longitudinal 資料集：

- 用來訓練模型
- 用來做 k-fold 交叉驗證 (GroupKFold)
- 交叉驗證只在 Longitudinal 內部進行

## Cross-sectional 資料集：

- 完整保留
- 完全不參與訓練與調參
- 最終當作 External Test (外部測試集)

## STEP2. Longitudinal 的分析方式：以 Subject ID 分組

GroupKFold (10-fold)

- 分組依據 Subject ID
- 讓同一個 Subject 的所有 visit 都一起進入同一個 fold
- 可避免模型提前「看過同一個人」

## STEP3. 標籤與特徵設定

標籤 (y\_long)

以 MMSE 建立二元分類：

$\text{MMSE} \geq 24 = 0$  (正常)

$\text{MMSE} < 24 = 1$  (異常)

→ 疑似失智 / 確診失智

特徵 (X\_long)

最穩定的三個數值特徵：

Age

nWBV

eTIV

→ 可以之後再加入變數(ex.gender)，類別會做encoding



# 用10-fold GroupKFold的原因

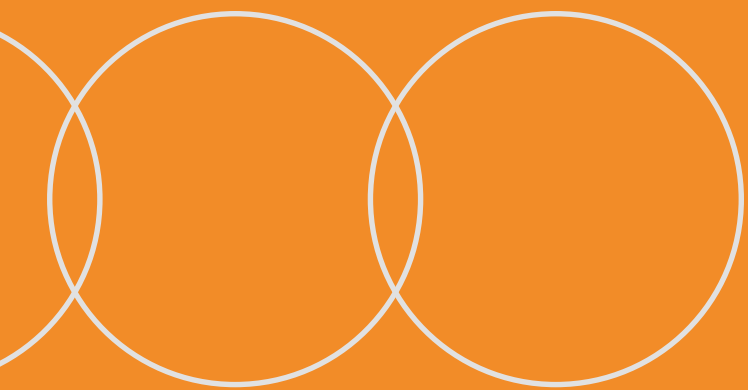
報告人：陳曄凱

## 1. Longitudinal 資料

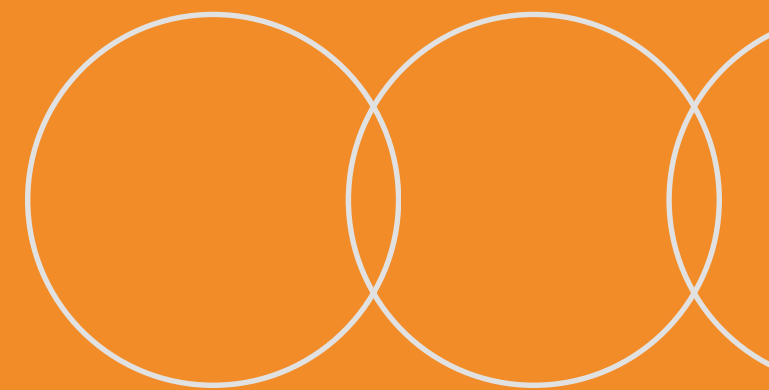
- 有重複追蹤同一個受試者
- 每 fold 大概 15 人
- 每次 135 人訓練、15 人驗證
- 因為受試者數夠多，切 10 折變異不會太大

## 2. Label ( CDR\_binary ) 分布健康

- 0 : 206
- 1 : 165
- 大約 55% vs 45% ,  
沒有嚴重不平衡  
模型比較不會因為不平衡而不準確



# 研究議題和預期結果



# 研究議題

## 假設：

- age越大，nWBV越低，MMSE分數越低
- age增加速率，與nWBV、MMSE分數下降速率高度正相關

## 分類模型：

- 使用age, nWBV作為主要輸入參數 $x$ ，搭配其他基本人口統計參數，建立有效區分正常與認知異常者 $y$ （失智症：MMSE < 24）

# 研究議題

**預測模型（回歸）：**

**使用 age、nWBV 作為主要輸入參數 X，搭配其他基本人口統計變項，建立以 MMSE 分數為輸出變數的回歸預測模型，評估 MRI 結構指標與人口學因素對認知功能表現的預測能力。**

# 參考文獻

- Daniel S., Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. 2007. “Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults.” *Journal of Cognitive Neuroscience* 19(9) (September):1498 - 1507.doi:10.1162/jocn.2007.19.9.1498.

# 參考文獻

- G. A. Ansari, Sivakani. R, S. Srisakth.(2022).Precise diagnosis of alzheimer's disease using recursive feature elimination method.Int. J. Systematic Innovation.7 (3):28-38.
- <https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers/data>
- [https://www.researchgate.net/figure/A-C-Analysis-of-ASF-eTIV-and-nWBV-for-Demented-and-Non-demented-group\\_fig5\\_359412479](https://www.researchgate.net/figure/A-C-Analysis-of-ASF-eTIV-and-nWBV-for-Demented-and-Non-demented-group_fig5_359412479)
- <https://matilda.fss.uu.nl/articles/common-data-types.html>Marcus,

# 工作分配

## 蔡碩恩

- 資料查找
- 資料分類
- 資料描述
- 資料前處理
- 視覺化分析
- 問題假設
- 製作簡報

## 傅景裕

- 資料、文獻查找
- 研究背景、目的
- 問題描述
- 資料描述、說明
- 製作簡報
- 上台報告
- 簡報總整理
- Q & A

## 陳偉凱

- 資料分類
- 資料前處理
- 視覺化分析
- 問題假設
- 製作簡報
- 資料切分
- 上台報告

# Q & A

**老師提問：橫斷面與縱向研究的區別是什麼？**

**回答：橫斷面年齡橫跨18至96歲，跨度大，此數據顯示僅六十歲以上高齡者，較容易患有失智症，故縱向研究僅針對六十歲以上較高風險者進行研究。**



# Q & A

**老師提問：橫斷面與縱向研究的 I D 各自代表什麼？沒有意義是否可以刪除？**

**回答：這些 I D 僅代表研究進行中的受試者訪視編號，沒有任何實質統計的必要，在後續資料前處理的部分會刪除**

# Q & A

**老師提問：MMSE與CRD分數與失智症間有什麼關係？**

**回答：前者代表認知能力分數，後者代表失智症嚴重分數，兩者呈現高度正相關，因此在前處理的過程，會將MMSE取代CRD，僅保留MMSE來作為判斷失智症與否的指標**

