

從頭開始

認識世代研究資料的分佈及 資料庫整理

醫學研究部 生物統計小組

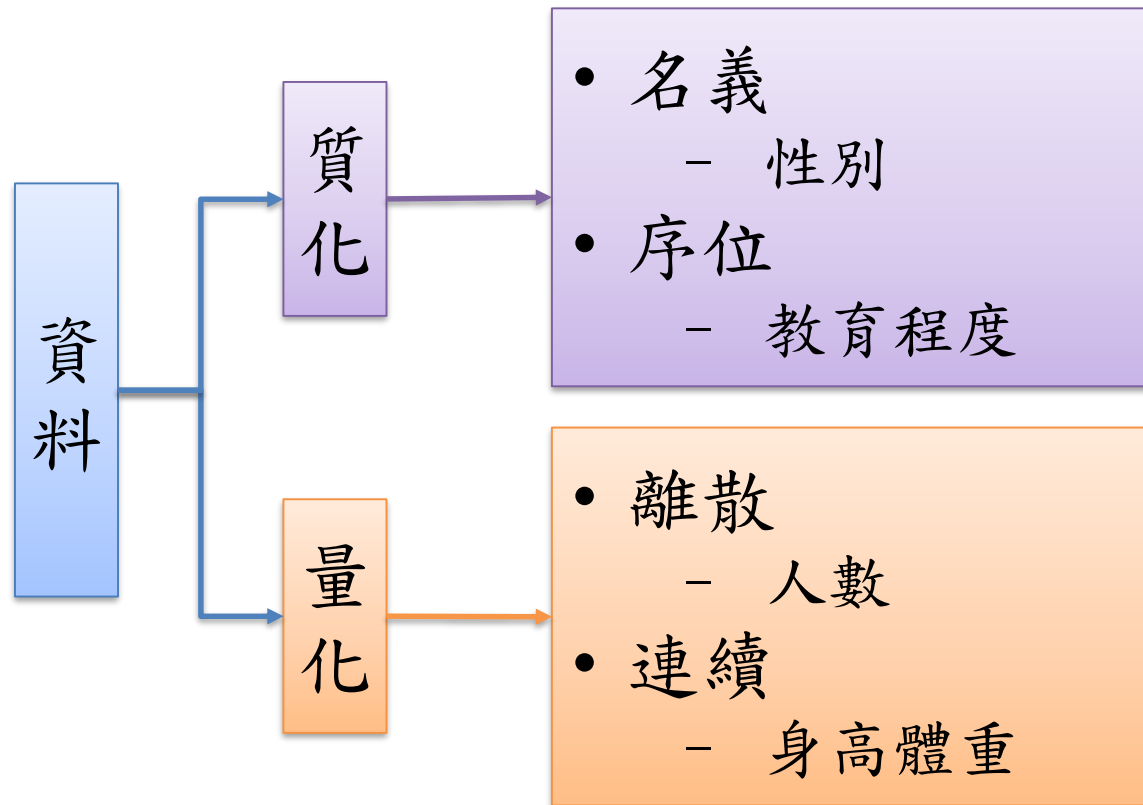
徐倩儀

2022/11/2

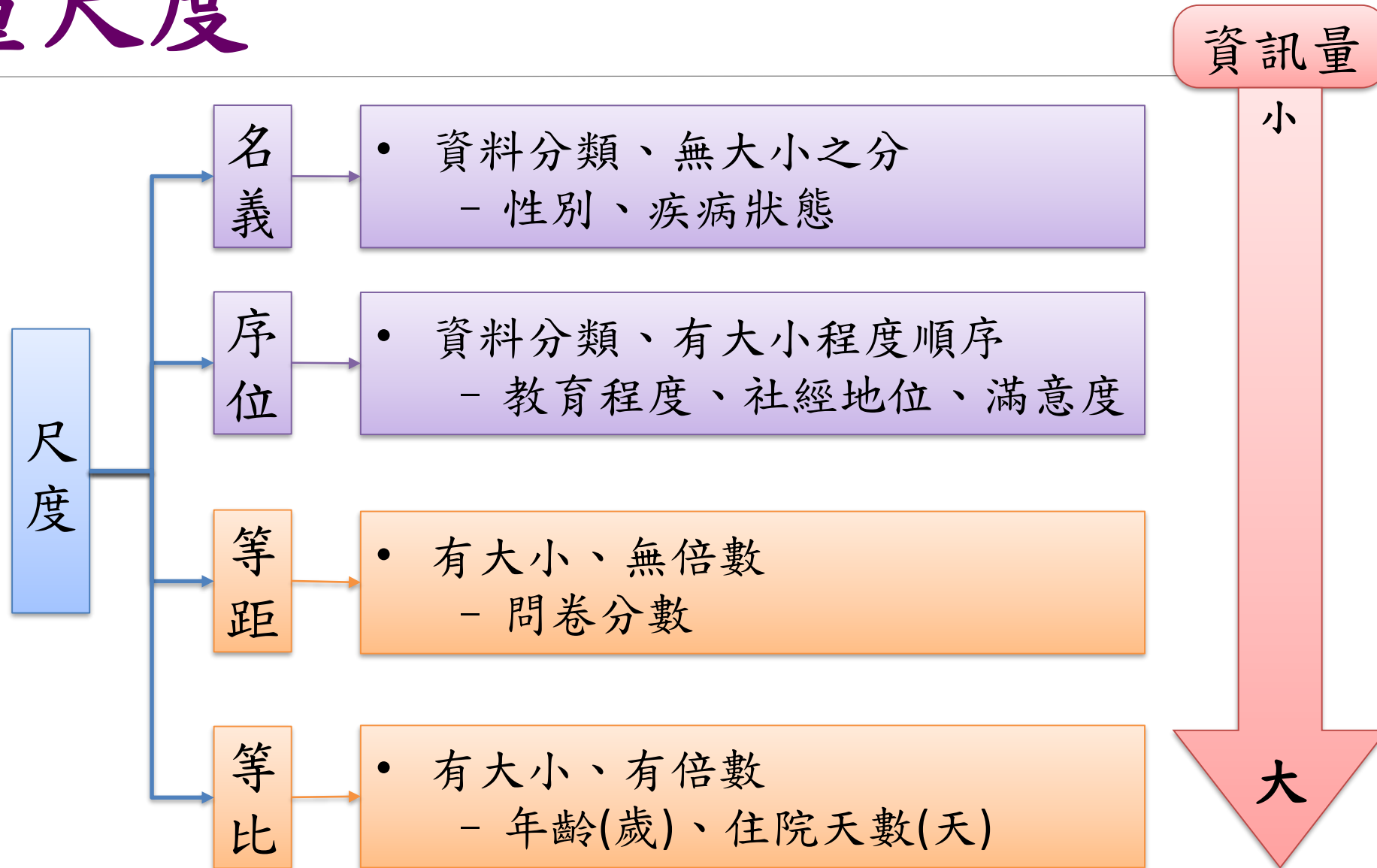
內容大綱

- 認識世代研究Medical Data的分佈
- 並學習如何整理成可用統計軟體分析的資料庫類型
- 如何探討資料的分佈

資料分類



測量尺度



測量尺度

- 研究者問卷設計(收集資料)

- 請問您今年幾歲? _____歲

- 請問您今年幾歲?

- 20歲以下 21-30歲 31-40歲 41-50歲

- 51-60歲 61-70歲 71歲以上



年齡的測量尺度?

1) 等比尺度

2) 序位尺度

測量尺度 (SPSS設定)

	Columns	Align	Measure
1		Right	Nominal
2		Right	Scale
3		Right	Ordinal
4		Right	Nominal
5		Right	Scale
6			
7			
8			
9			
10			
11			
12			
13			

Data View Variable View

IBM SPSS Statistics Processor is ready

	欄	對齊	測量
1		靠右	名義
2		靠右	比例
3		靠右	序數
4		靠右	名義
5		靠右	比例
6			
7			
8			
9			
10			
11			
12			
13			

資料視圖 變數視圖

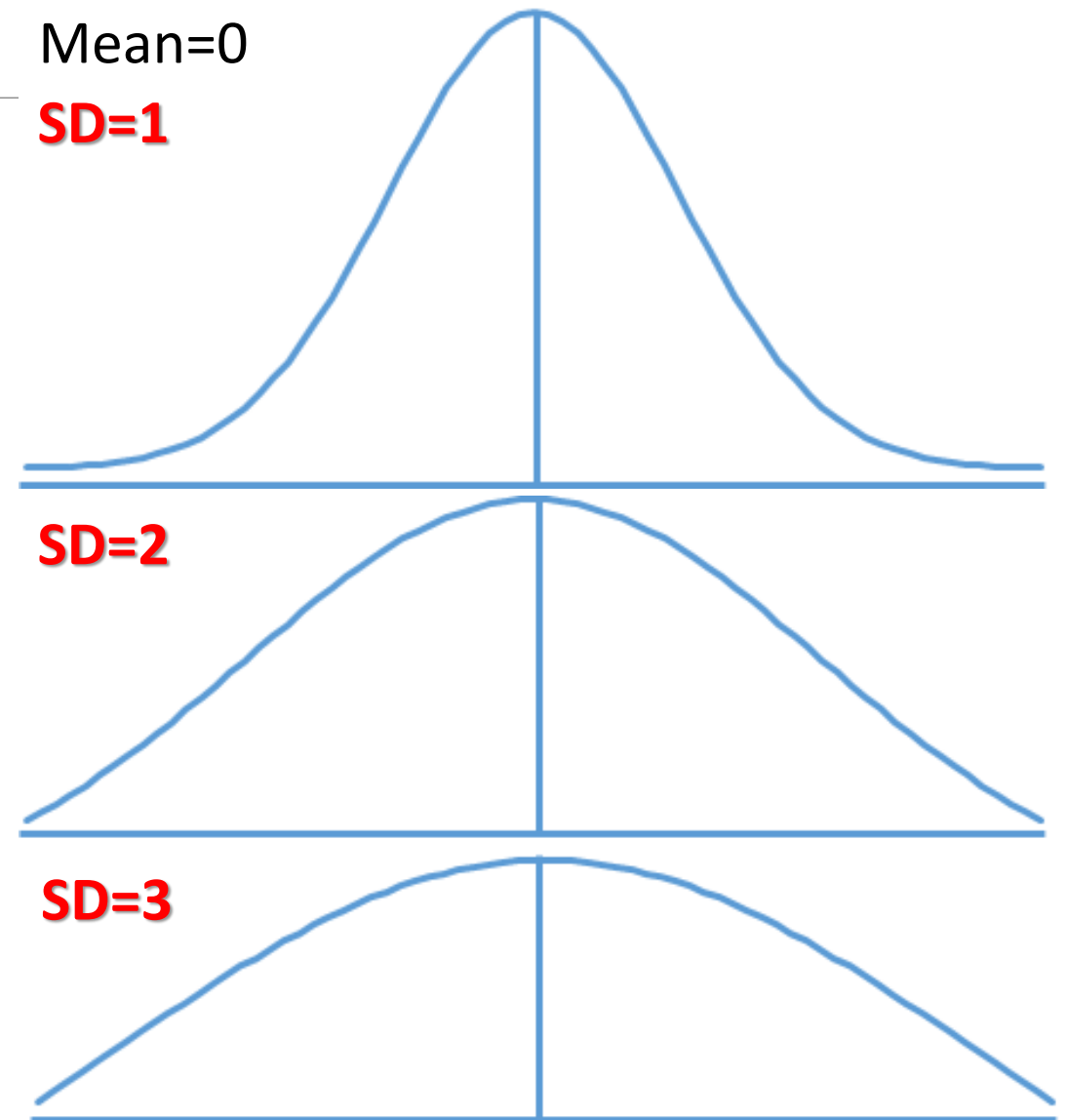
IBM SPSS Statistics 處理器已備妥

描述性統計

- 類別型(名義/序位尺度)
 - n (%)
- 連續型(等距/等比尺度)
 - Mean \pm SD
 - Median, IQR

Mean \pm SD

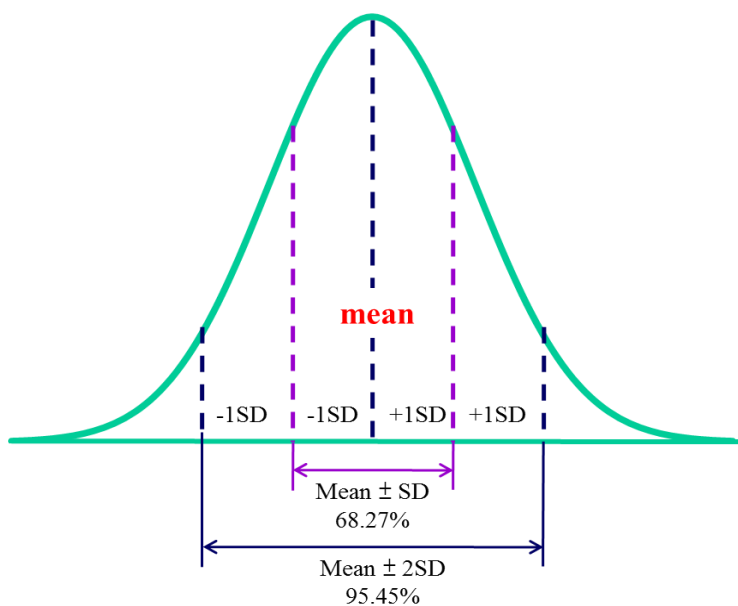
- 平均值(mean)
 - 中心點位置
- 標準差(SD)
 - 鐘型曲線形狀



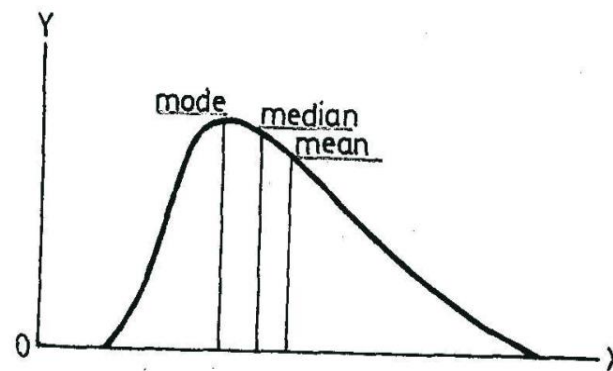
資料分佈

常態分佈

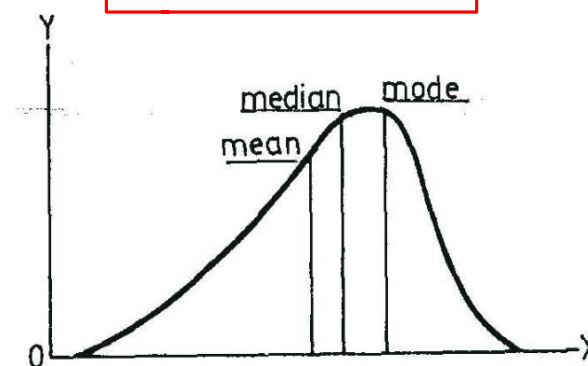
- Gaussian distribution
- 以平均值為中心的對稱曲線
 - Mean = Median = Mode



非常態分佈

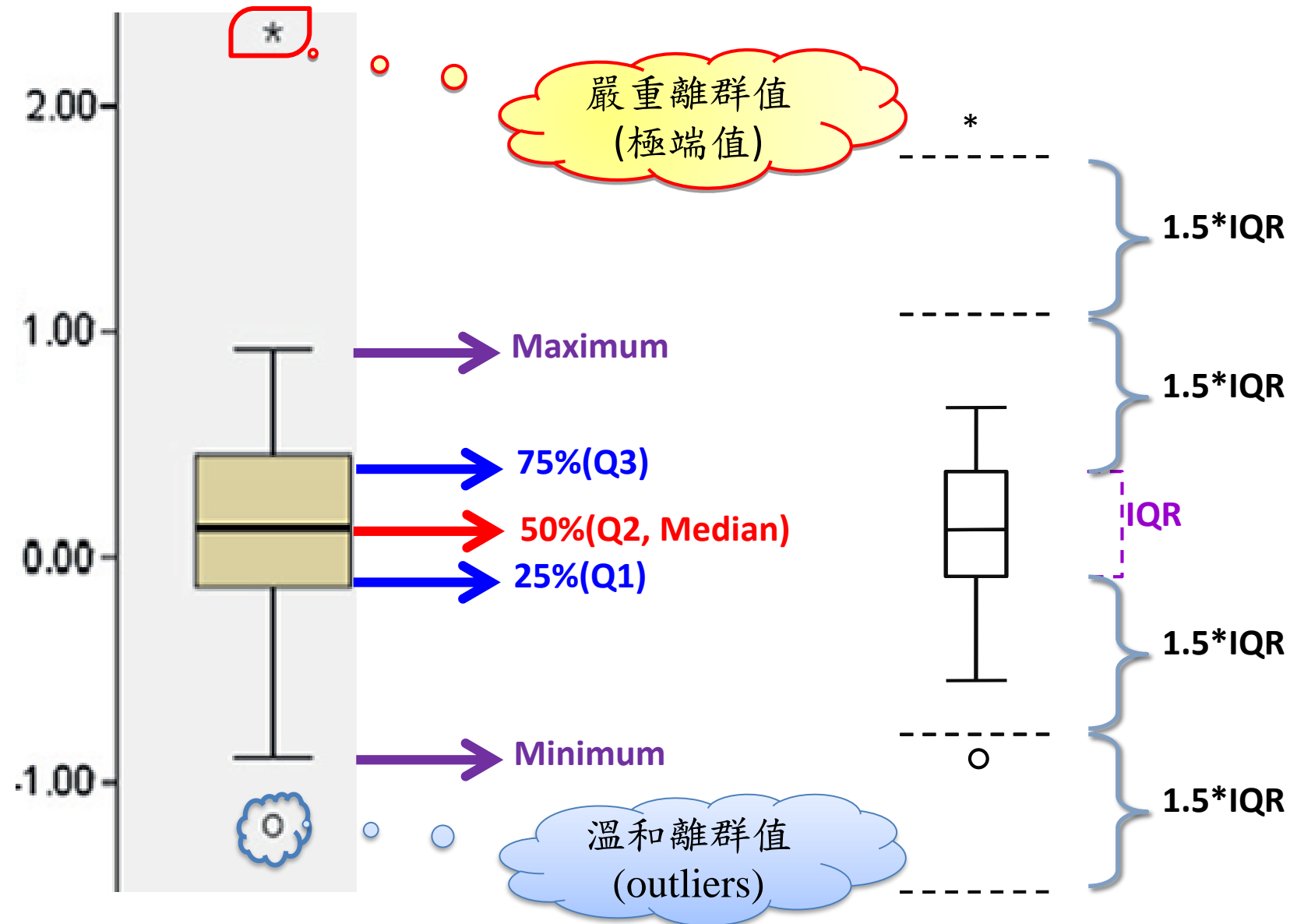


右偏分布



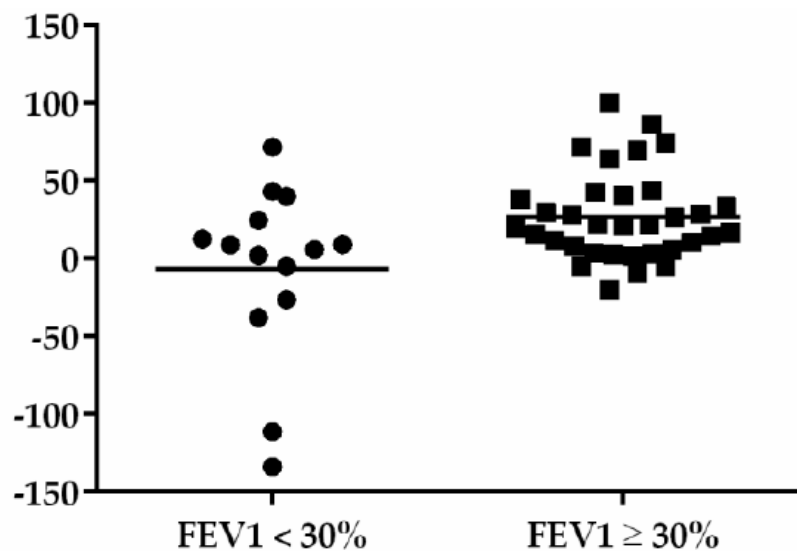
左偏分布

Box plot

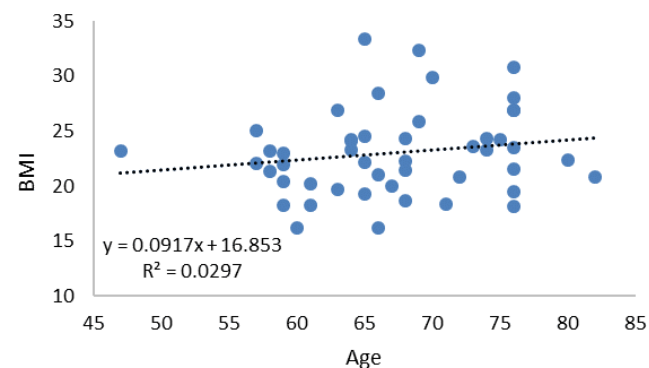
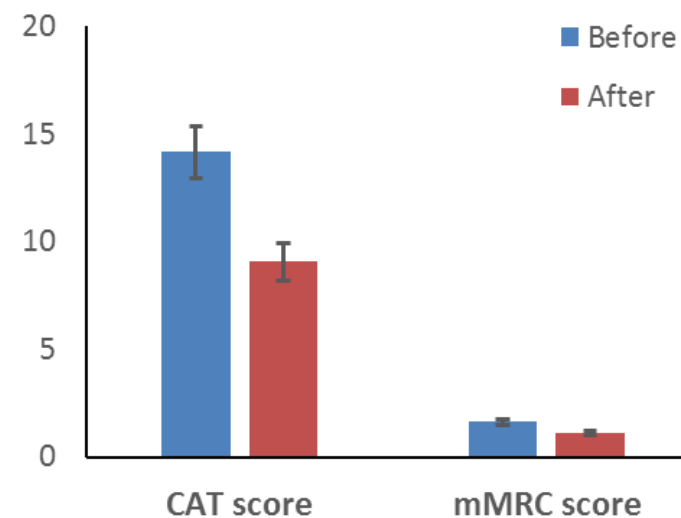
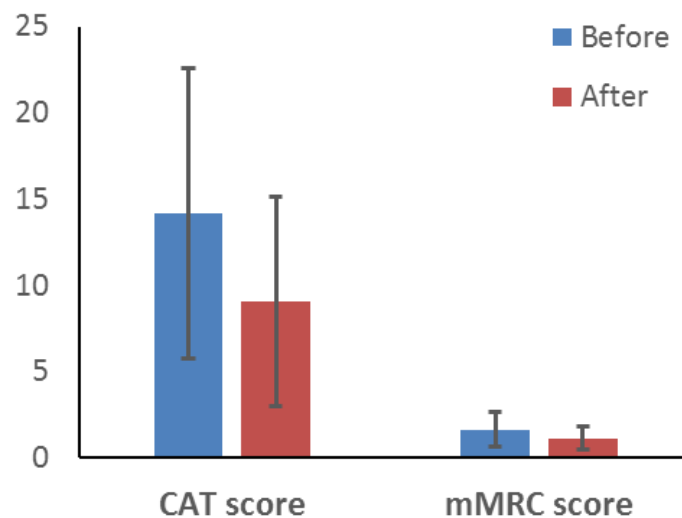


統計圖

● 資料分佈



Cheng, Y.Y., Lin, S.Y., Hsu, C.Y., & Fu, P.K. (2022). Respiratory Muscle Training Can Improve Cognition, Lung Function, and Diaphragmatic Thickness Fraction in Male and Non-Obese Patients with Chronic Obstructive Pulmonary Disease: A Prospective Study. *J Pers Med*. 2022 Mar 16;12(3):475. doi: 10.3390/jpm12030475.



Hypothesis

- Null hypothesis(H_0)：虛無假設
- Alternative hypothesis (H_1)：對立假設

		真實情形	
		The Null hypothesis(H_0) is true	The Alternative hypothesis (H_1) is true
研究結果	The Null hypothesis(H_0) is true	Accurate ($1-\alpha$)	Type II error (β)
	The Alternative hypothesis (H_1) is true	Type I error (α)	Accurate ($1-\beta$) (Power)

α level (推翻虛無假設)

- 可容忍的最大誤差
- 通常設為0.05、0.1及0.01

α level = 0.05
Power = 0.8

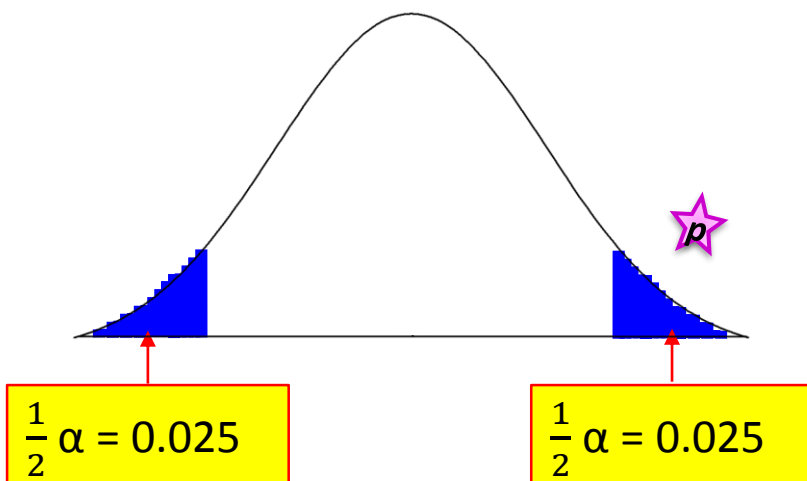
Hypothesis

α level = 0.05

雙尾檢定

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$



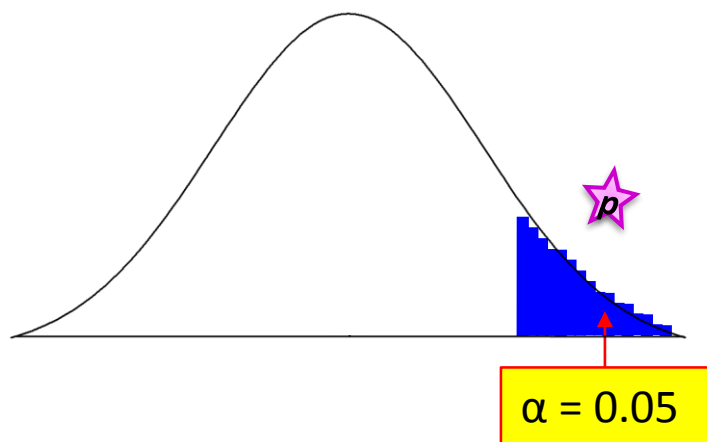
Q1. 男女體重是否相同?

Q2. 新藥是否比傳統用藥的血壓值較低?

單尾檢定

$$H_0 : \mu \leq \mu_0$$

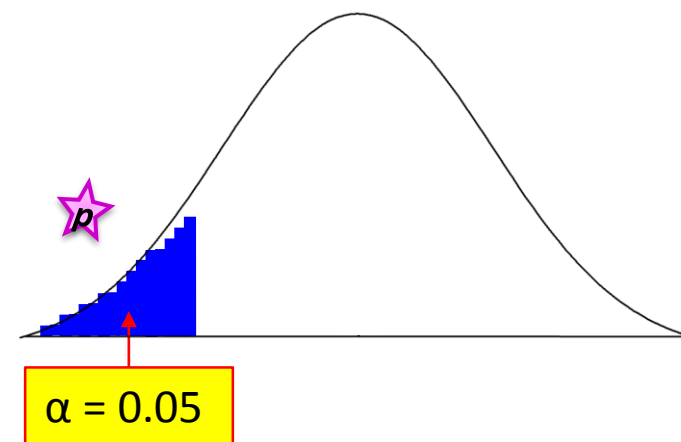
$$H_1 : \mu > \mu_0$$



單尾檢定

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$



p value

- significance level
 - 機率
 - 錯誤地接受對立假設的統計機率
 - 介於0-1之間
 - 統計顯著差異
 - 與alpha值(α level)比較
 - 接受Alternative hypothesis (H_1)
 - $p \text{ value} \leq \alpha \text{ level}$
 - 接受Null hypothesis (H_0)
 - $p \text{ value} > \alpha \text{ level}$

資料建檔-1

ID	性別	年齡	婚姻	教育	體重	身高	抽菸	喝酒	日期
001	M	57	已婚	高中	89	186	無	無	2013/4/1
002	F	46	已婚	大專以上	57	160	有	無	2012/6/28
003	F	55	已婚	國中	90	170	無	無	2013/1/21
004	M	65	已婚	大專以上	60	160	有	有	2014/6/20
005	M	60	已婚	國中	58	162	有	有	2013/9/18

- 欄
 - 變項數值
- 列
 - 變數名稱
- 譯碼簿(編碼簿 coding book)

資料建檔-2 (譯碼簿)

ID	性別	年齡	婚姻	教育	體重	身高	抽菸	喝酒	日期
001	M	57	已婚	高中	89	186	無	無	2013/4/1
002	F	46	已婚	大專以上	57	160	有	無	2012/6/28
003	F	55	已婚	國中	90	170	無	無	2013/1/21
004	M	65	已婚	大專以上	60	160	有	有	2014/6/20
005	M	60	已婚	國中	58	162	有	有	2013/9/18

ID	性別	年齡	婚姻	教育	體重	身高	抽菸	喝酒	日期
001	1	57	1	3	89	186	0	0	2013/4/1
002	0	46	1	4	57	160	1	0	2012/6/28
003	0	55	1	2	90	170	0	0	2013/1/21
004	1	65	1	4	60	160	1	1	2014/6/20
005	1	60	1	2	58	162	1	1	2013/9/18

轉碼

性別	0	F	婚姻	1	已婚	教育	1	小學	抽菸	0	無
	1	M		2	單身		2	國中		1	有
		3		離婚	3		高中	喝酒	0	無	
				4	大專以上		1		有		

文字型態
改為
數值型態

資料建檔-3 (複選題)

請問有無以下疾病？

1. 高血壓

2. 高血脂

3. 心臟病

4. 糖尿病

5. 腎炎、腎徵候群及腎性病變

6. 慢性阻塞性肺病

7. 慢性肝病及肝硬化

8. 其他_____

一個變項數值
一個概念

ID	疾病	高血壓	高血脂	心臟病	糖尿病
001		0	0	0	0
002	1,4	1	0	0	1
003	2,4	0	1	0	1
004	1,2,4	1	1	0	1
005	3	0	0	1	0
006	1	1	0	0	0

資料建檔-4 (日期格式)

ID	日期1	日期2	日期3	日期4	日期5
001	2013/4/1	2015/12/10	20130401	1020401	102/4/1
002	2012/6/28	2015/11/27	20120628	1010628	101/6/28
003	2013/1/21	2015/12/12	20130121	1020121	102/1/21
004	2014/6/20	2015/12/8	20140620	1030620	103/6/20
005	2013/9/18	2015/12/3	20130918	1020918	102/9/18

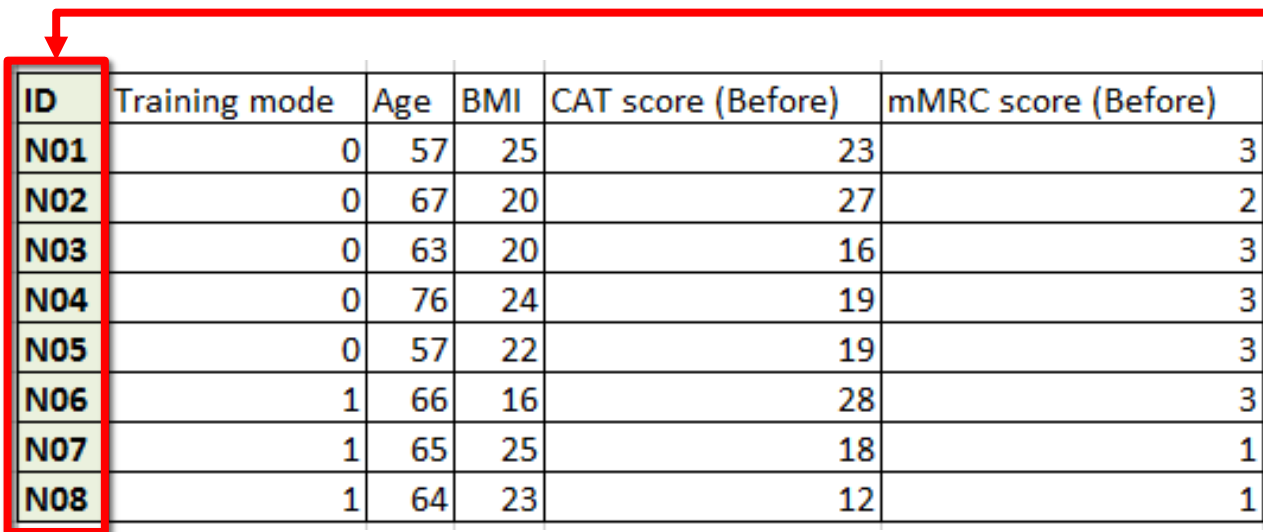
SUM X ✓ fx =C2-B2

	A	B	C	D	E	F	G
1	ID	日期1	日期2	日期3	日期4	日期5	日期2-日期1
2	001	2013/4/1	2015/12/10	20130401	1020401	102/4/1	=C2-B2
3	002	2012/6/28	2015/11/27	20120628	1010628	101/6/28	1247
4	003	2013/1/21	2015/12/12	20130121	1020121	102/1/21	1055
5	004	2014/6/20	2015/12/8	20140620	1030620	103/6/20	536
6	005	2013/9/18	2015/12/3	20130918	1020918	102/9/18	806

日期格式可直接計算
時間差(天數)

資料建檔-5

- 簡單且明確建檔
 - 不同檔案各別建立
 - 基本資料
 - 追蹤資料
 - 關鍵字(如病歷號)合併資料



ID	Training mode	Age	BMI	CAT score (Before)	mMRC score (Before)
N01	0	57	25	23	3
N02	0	67	20	27	2
N03	0	63	20	16	3
N04	0	76	24	19	3
N05	0	57	22	19	3
N06	1	66	16	28	3
N07	1	65	25	18	1
N08	1	64	23	12	1

ID	CAT score (After)	mMRC score (After)
N01	10	1
N02	10	1
N03	13	1
N04	4	1
N05	16	3
N06	13	1
N07	16	2
N08	20	2

資料檔範例-1

- 不建議建檔方式

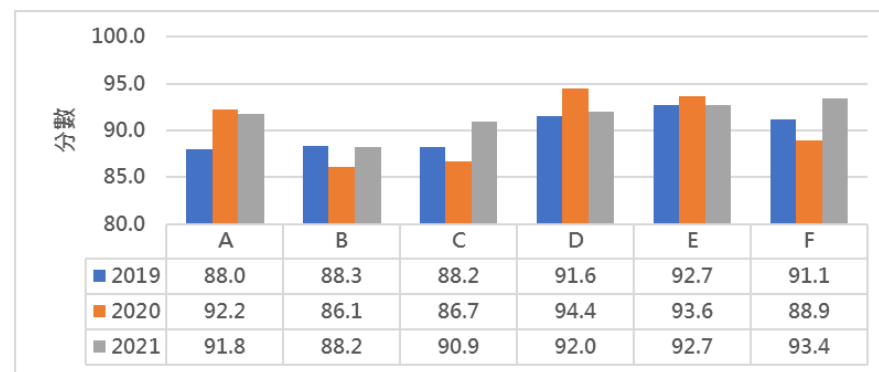
ID	Training mode	Age	BMI	Time	CAT score	mMRC score
N01	0	57	25	Before	23	3
				After	10	1
N02	0	67	20	Before	27	2
				After	10	1
N03	0	63	20	Before	16	3
				After	13	1
N04	0	76	24	Before	19	3
				After	4	1
N05	0	57	22	Before	19	3
				After	16	3
N06	1	66	16	Before	28	3
				After	13	1
N07	1	65	25	Before	18	1
				After	16	2
N08	1	64	23	Before	12	1
				After	20	2

資料檔範例-2

● 無法直接檢定的類型

— 彙整資料

年度	評分項目	Dis1						Dis2						Dis3						總平均
		優	佳	普通	待改進	差	分數	優	佳	普通	待改進	差	分數	優	佳	普通	待改進	差	分數	
2019	A	4	2	0	0	0	91.7	2	3	0	0	0	89.0	1	3	2	0	0	83.3	88.0
	B	3	3	0	0	0	90.0	5	0	0	0	0	95.0	0	4	1	1	0	80.0	88.3
	C	2	4	0	0	0	88.3	4	1	0	0	0	93.0	2	1	3	0	0	83.3	88.2
	D	4	2	0	0	0	91.7	4	1	0	0	0	93.0	3	3	0	0	0	90.0	91.6
	E	6	0	0	0	0	95.0	4	1	0	0	0	93.0	3	3	0	0	0	90.0	92.7
	F	3	3	0	0	0	90.0	5	0	0	0	0	95.0	2	4	0	0	0	88.3	91.1
2020	A	4	0	0	0	0	95.0	2	2	0	0	0	90.0	4	2	0	0	0	91.7	92.2
	B	0	2	2	0	0	80.0	2	2	0	0	0	90.0	3	2	1	0	0	88.3	86.1
	C	0	2	2	0	0	80.0	3	0	1	0	0	90.0	3	3	0	0	0	90.0	86.7
	D	4	0	0	0	0	95.0	4	0	0	0	0	95.0	5	1	0	0	0	93.3	94.4
	E	4	0	0	0	0	95.0	3	1	0	0	0	92.5	5	1	0	0	0	93.3	93.6
	F	1	2	1	0	0	85.0	2	2	0	0	0	90.0	4	2	0	0	0	91.7	88.9
2021	A	3	4	0	0	0	89.3	3	2	0	0	0	91.0	5	0	0	0	0	95.0	91.8
	B	4	3	0	0	0	90.7	3	1	1	0	0	89.0	1	3	1	0	0	85.0	88.2
	C	4	3	0	0	0	90.7	3	2	0	0	0	91.0	3	2	0	0	0	91.0	90.9
	D	5	2	0	0	0	92.1	2	3	0	0	0	89.0	5	0	0	0	0	95.0	92.0
	E	5	2	0	0	0	92.1	3	2	0	0	0	91.0	5	0	0	0	0	95.0	92.7
	F	5	2	0	0	0	92.1	4	1	0	0	0	93.0	5	0	0	0	0	95.0	93.4



資料檔範例-3

- 統計軟體無法直接計算的類型
 - 變項『數值』及『名稱』位置相反

ID	N01	N02	N03	N04	N05	N06	N07	N08
Training mode	0	0	0	0	0	1	1	1
Age	57	67	63	76	57	66	65	64
BMI	25	20	20	24	22	16	25	23
CAT score (Before)	23	27	16	19	19	28	18	12
CAT score (After)	10	10	13	4	16	13	16	20
mMRC score (Before)	3	2	3	3	3	3	1	1
mMRC score (After)	1	1	1	1	3	1	2	2

資料檔範例-4

- 需要進行清理/擷取的資料

病歷號	檢查日期	檢查項目	檢查數值
	20090115	HGB	10.4
	20090115	PLT	398
	20090115	WBC	9400
	20090115	ALB	3
	20090115	Creatinine	1.3
	20090115	GPT	11
	20090115	Glucose	110
	20090120	HGB	10
	20090121	Glucose	NEG
	20090123	Creatinine	0.8
	20090123	HGB	10.6
	20090123	LYM	9.5
	20090123	NEUT	83.3
	20090123	PLT	301
	20090123	WBC	12900
	20090204	Creatinine	0.9
	20090204	HGB	11.2
	20090204	WBC	10400
	20090205	Glucose	NEG
	20090209	HGB	10.4
	20090209	WBC	8800
	20090209	Creatinine	1

動手做做看-資料建檔

- 從以下資料試著自行輸入資料

序號0597

1. 生日：23.10.01
2. Sex： Male Female
3. BMI：24
4. Smoking： No Yes
5. Albumin (g/dL)：3.6
6. Hba1c (%)：_____
7. Proteinuria (mg/g)：0.08
8. Diuretics： No Yes
9. MRA： No Yes

序號1713

1. 生日：1945.09.05
2. Sex： Male Female
3. BMI：_____
4. Smoking： No Yes
5. Albumin (g/dL)：_____
6. Hba1c (%)：7.9
7. Proteinuria (mg/g)：_____
8. Diuretics： No Yes
9. MRA： No Yes

序號0796

1. 生日：33/11/01
2. Sex： Male Female
3. BMI：26
4. Smoking： No Yes
5. Albumin (g/dL)：3.9
6. Hba1c (%)：5.7
7. Proteinuria (mg/g)：0.13
8. Diuretics： No Yes
9. MRA： No Yes

動手做做看-統計值/統計圖

- EXECL
 - Bar chart (mean \pm SD / mean \pm SE)
 - Scatter plot

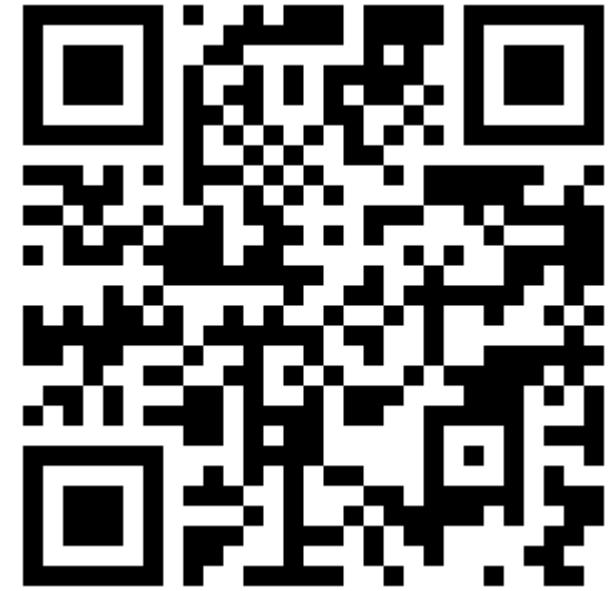
動手做做看(練習資料檔.xlsx)

ID	Age	BMI	CATO	CAT1	mMRC0	mMRC1
N01		57	25	23	10	3
N02		67	20	27	10	2
N03		63	20	16	13	3
N04		76	24	19	4	3
N05		57	22	19	16	3
N06		66	16	28	13	3
N07		65	25	18	16	1
N08		64	23	12	20	1
N09		76	22	2	3	1
N10		76	27	33	21	2
N11		69	32	20	7	3
N12		59	20	19	15	1
N13		58	21	17	9	3
N14		61	20	18	18	2
N15		71	18	34	19	4
N16		76	20	5	8	1
N17		64	24	6	2	1
N18		59	23	15	5	1
N19		72	21	9	4	1
N20		74	23	10	6	2
N21		58	23	5	3	0
N22		80	22	7	7	1
N23		68	19	15	4	2
N24		68	22	14	6	2
N25		75	24	15	7	2
N26		76	18	11	8	2
N27		47	23	21	10	0
N28		65	22	5	3	1
N29		76	31	23	10	2
N30		73	24	26	9	3
N31		65	19	24	7	3
N32		64	24	9	6	1
N33		82	21	8	27	2
N34		63	27	16	4	1
N35		76	28	11	5	1
N36		66	21	22	12	2
N37		68	22	15	11	1
N38		59	22	7	5	1
N39		74	24	2	2	1
N40		76	27	3	2	1
N41		60	16	9	10	1
N42		65	33	6	6	1
N43		70	30	18	13	1
N44		68	24	3	2	1
N45		66	28	3	1	0
N46		61	18	0	2	0
N47		59	18	14	15	1
N48		69	26	18	19	3

CATO	CAT score (Before)
CAT1	CAT score (After)
mMRC0	mMRC score (Before)
mMRC1	mMRC score (After)

問卷調查

Thank you



For your attention!!