

腦中風資料的統計分析與探討

陳奕造 0687203

施奕如

楊欣儒

李長穎

May 22, 2021

Contents

1	前言：背景與動機	3
1.1	背景	3
1.2	動機	3
2	資料探討	4
2.1	資料來源	4
2.2	資料基本型態	4
2.3	單變數分析: 類別型	5
2.3.1	gender	5
2.3.2	Hypertension	5
2.3.3	Heart Disease	6
2.3.4	Ever married	6
2.3.5	Residence type	6
2.3.6	Stroke	7
2.3.7	Smoking status	7
2.3.8	Work type	7
2.4	單變數分析: 數值型	8
2.4.1	age	8
2.4.2	Average glucose level	8
2.4.3	BMI	9
3	影響中風的因素	10
3.1	Analysis of Contingency Table	10
3.1.1	中風與性別	10
3.1.2	中風與居住環境	10
3.1.3	中風與婚姻	11
3.1.4	中風與壓力	12
3.1.5	中風與心臟病	12

<i>CONTENTS</i>	2
3.1.6 中風與抽煙習慣	13
3.1.7 中風與年齡	14
3.1.8 中風與血糖	14
3.1.9 中風與BMI	14
4 Clustering	16
4.1 The table of correlation	16

Chapter 1

前言：背景與動機

1.1 背景

隨著年齡增長，罹患腦中風、缺血性心臟病與癌症的風險也越高，亦為1990年以來WHO統計已開發國家的前三大死因^[1]，2016年全球的中風病患高達八千萬人，其中近一千四百萬人為新增病例，同年有550萬人死於中風，為全球第二大死因^[2]，其中75%來自、低收入國家^[3]，過去十年來(4/6止，衛服部公布至2019，每年國人十大死因)一直列在臺灣國人十大死因的前五名^[4;5]。

中風、或是腦中風(stroke, cerebrovascular accident)，在衛生福利部上，我們又可以看到另一個稱呼「腦血管疾病」。當腦部血管受到阻塞或破裂，腦部細胞欠缺血液的運輸而導致缺氧，細胞進而損傷或死亡，便稱之為腦中風^[6]。罹患中風後，可能需要面對許多併發症，包含肌肉能力喪失，部分癱瘓、吞嚥困難，中樞神經系統大腦上則有記憶喪失、思覺異常、癲癇等等^[7;8]，其中又有1/3的患者在癒後仍會伴隨這些後遺症^[9]。

除了高死亡率，國家亦需要付出大量的醫療成本，對中風患者進行治療與病後照護，近數十年，腦中風對於全球各國均是不可忽視的問題。

1.2 動機

腦中風已被證實與血壓、糖尿病(肥胖)、年齡密切相關^[10]，而後兩者又可能導致高血壓，因此光探討糖尿病與肥胖，可能就已經足以涵蓋血壓所造成的影響力。故，本次報告除了想要再次證實已知的腦中風與其他病史已知的關係，我們想要瞭解，當不考慮糖尿病史與年齡時，生活環境對血壓與腦中風所造成的關係。

Chapter 2

資料探討

2.1 資料來源

我們選擇了一筆2018年釋出的網路資料

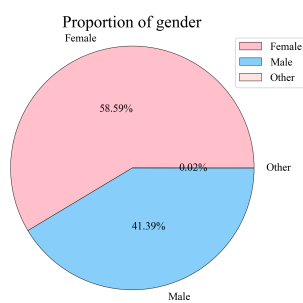
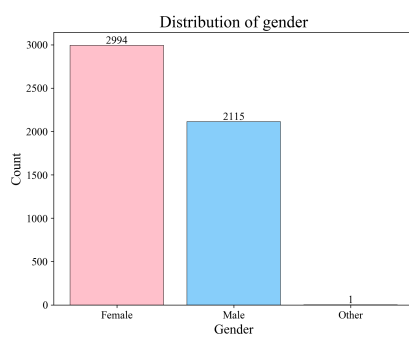
- McKinsey Analytics: Online Hackathon on Healthcare

2.2 資料基本型態

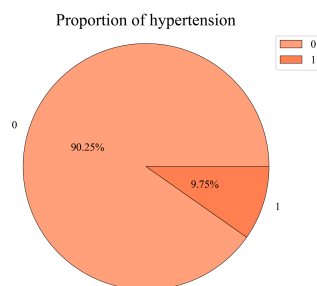
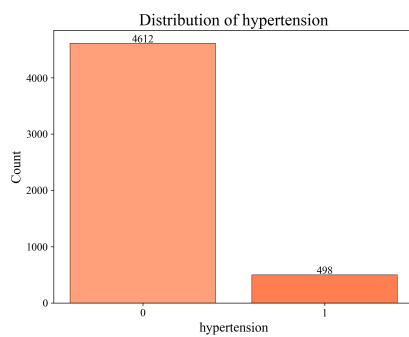
- 資料總數: 5110個人
- 類別型: gender, hypertension, heart_disease, ever_married, work_type, Residence_type, smoking_status, stroke
- 數值型: age, avg_glucose_level, bmi

2.3 單變數分析：類別型

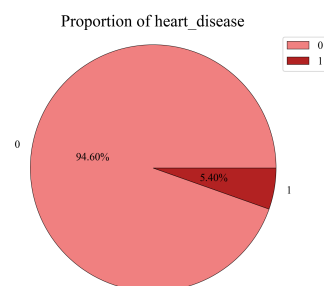
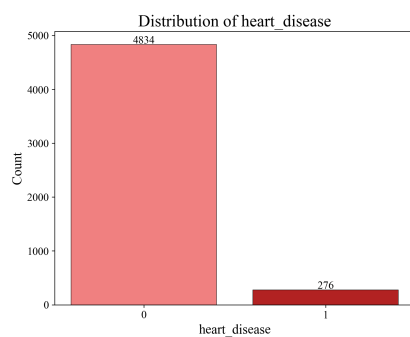
2.3.1 gender



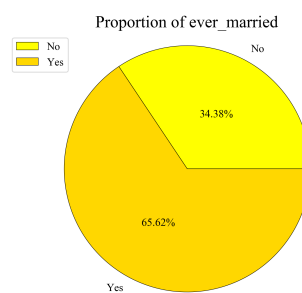
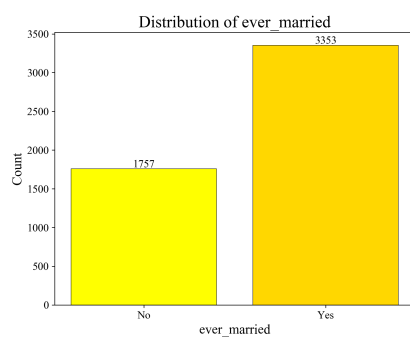
2.3.2 Hypertension



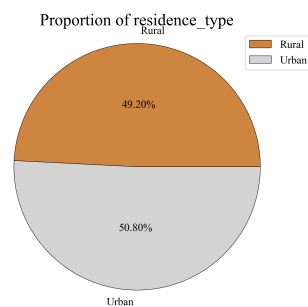
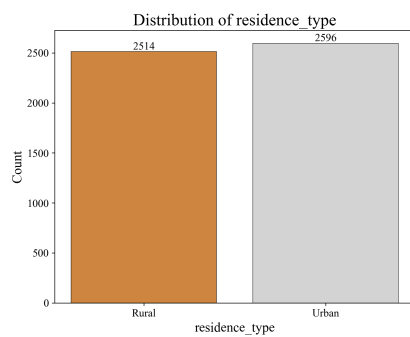
2.3.3 Heart Disease



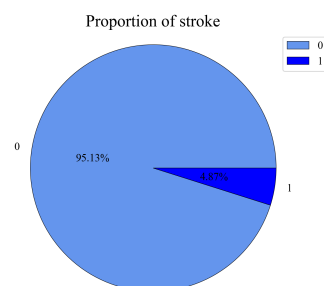
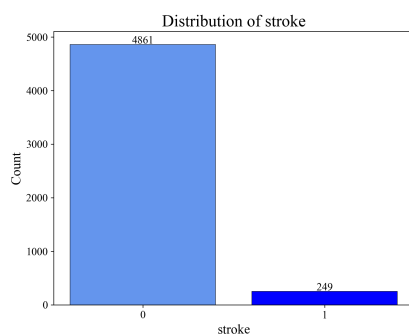
2.3.4 Ever married



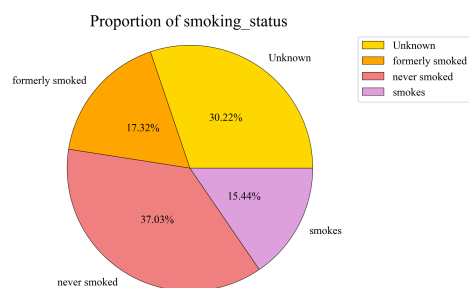
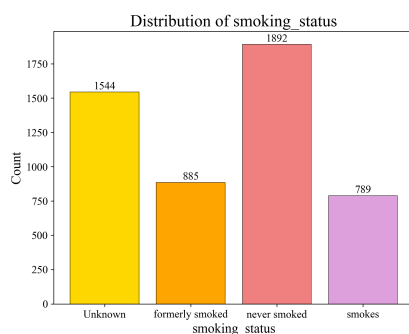
2.3.5 Residence type



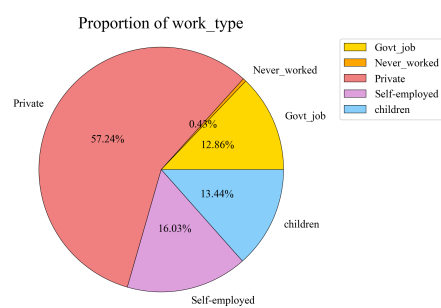
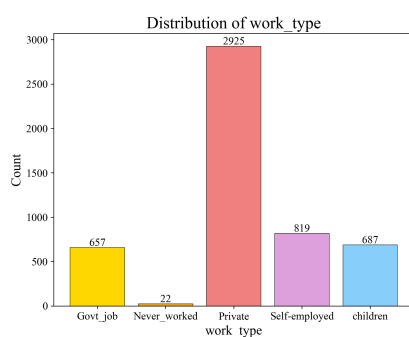
2.3.6 Stroke



2.3.7 Smoking status

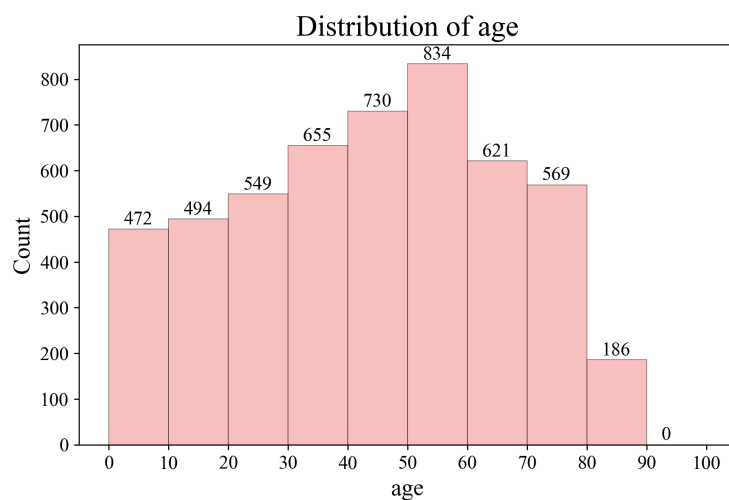


2.3.8 Work type

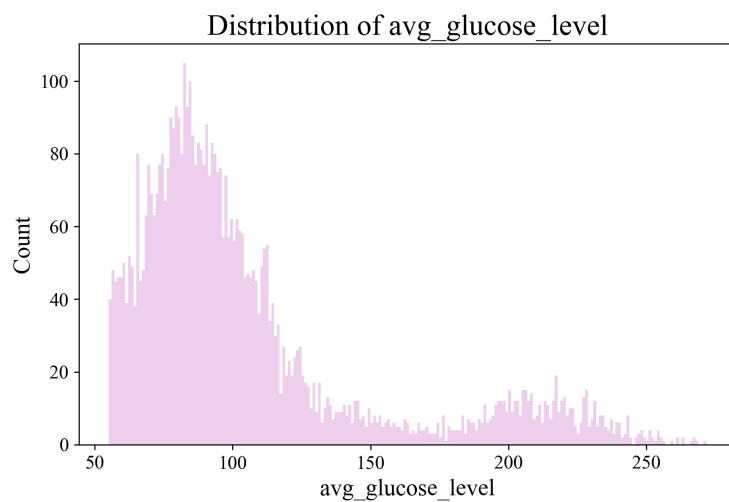


2.4 單變數分析：數值型

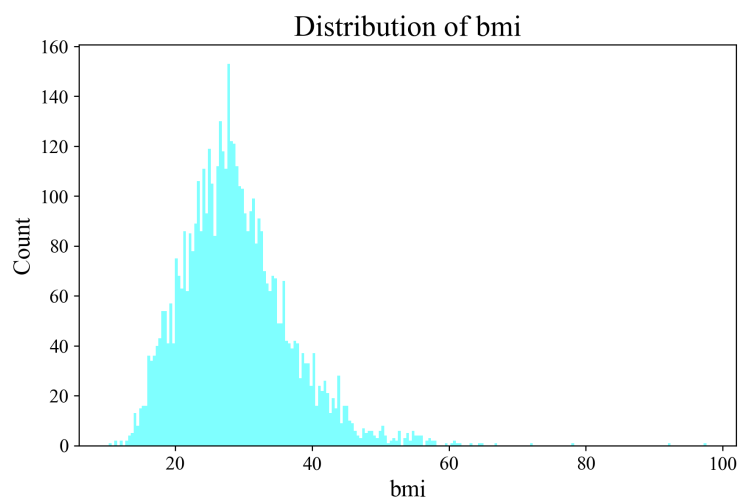
2.4.1 age



2.4.2 Average glucose level



2.4.3 BMI



Chapter 3

影響中風的因素

3.1 Analysis of Contingency Table

3.1.1 中風與性別

- 基本假設:性別與中風沒有關聯

	Stroke	No Stroke	Row Total
Male	108	2007	2115
Female	141	2853	2994
Column Total	249	4860	5109

- 註記: 有一個性別是Other，被拿掉了
- Odds ratio $\hat{\theta} = 1.089$
- 95% Confidence Interval of $\log(\hat{\theta}) = (-0.172, 0.342)$
- 結論:中風跟性別是獨立的，符合我們一開始的猜想

3.1.2 中風與居住環境

- 基本假設:只有一點點關聯，鄉下的中風比例可能比住在城市的低

	Stroke	No Stroke	Row Total
Rural	114	2400	2514
Urban	135	2461	2596
Column Total	249	4861	5110

- Odds ratio $\hat{\theta} = 0.866$
- 95% Confidence Interval of $\log(\hat{\theta}) = (-0.400, 0.112)$
- 雖然資料符合我們的猜想，鄉下中風的比例比較低，但是95%的信賴區間包含0，表示 $\log(\hat{\theta})$ 與0無顯著差異
- 結論:中風與居住環境是獨立的

3.1.3 中風與婚姻

- 基本假設:婚姻與中風沒有關聯

	Stroke	No Stroke	Row Total
Ever-Married	220	3133	3353
Never-Married	29	1728	1757
Column Total	249	4861	5110

- Odds ratio $\hat{\theta} = 4.184$
- 95% Confidence Interval of $\log(\hat{\theta}) = (1.040, 1.823)$
- 95%的信賴區間不包含0，表示 $\log(\hat{\theta})$ 與0有顯著差異。
- 結論:跟我們的基本假設相反，中風跟”有無結過婚”有關聯

Fix column比較:

- 中風患者: $\frac{\text{Ever-Married}}{\text{Never-Married}} = 7.59$
- 非中風者: $\frac{\text{Ever-Married}}{\text{Never-Married}} = 1.81$
- 中風患者中”結婚與沒結婚”的比例高於非中風者

3.1.4 中風與壓力

- 基本假設:壓力愈大的人愈容易中風

	Stroke	No Stroke	Row Total
Hypertension	66	432	498
Non-Hypertension	183	4429	4612
Column Total	249	4861	5110

- Odds ratio $\hat{\theta} = 3.698$
- 95% Confidence Interval of $\log(\hat{\theta}) = (1.009, 1.606)$
- 95%的信賴區間不包含0，表示 $\log(\hat{\theta})$ 與0有顯著差異。
- 結論:中風跟壓力有關聯

Fix column比較:

- 中風患者: $\frac{\text{Hypertension}}{\text{Non-Hypertension}} = 0.36$
- 非中風者: $\frac{\text{Hypertension}}{\text{Non-Hypertension}} = 0.10$
- 符合基本假設，壓力愈大的人愈容易中風

3.1.5 中風與心臟病

- 基本假設:中風是心血管疾病，要與心臟病相關

	Stroke	No Stroke	Row Total
Heart-Disease	47	229	276
No-Heart-Disease	202	4632	4834
Column Total	249	4861	5110

- Odds ratio $\hat{\theta} = 4.706$
- 95% Confidence Interval of $\log(\hat{\theta}) = (1.205, 1.893)$

- 95%的信賴區間不包含0，表示 $\log(\hat{\theta})$ 與0有顯著差異。
- 結論:中風跟心臟病有關聯

Fix column比較:

- 中風患者: $\frac{\text{Heart-Disease}}{\text{No-Heart-Disease}} = 0.23$
- 非中風者: $\frac{\text{Heart-Disease}}{\text{No-Heart-Disease}} = 0.05$
- 符合基本假設，中風患者中有心臟病的比例高

3.1.6 中風與抽煙習慣

- 基本假設:有抽煙的人比較容易中風

	Stroke	No Stroke	Row Total
Smoke	112	1562	1674
No smoke	90	1802	1892
Column Total	202	3364	3566

- 註記: unknown有1544個資料點，被拿掉了
- 註記: smoke包含smokes與formerly smoked
- Odds ratio $\hat{\theta} = 1.436$
- 95% Confidence Interval of $\log(\hat{\theta}) = (0.076, 0.647)$
- 95%的信賴區間不包含0，表示 $\log(\hat{\theta})$ 與0有顯著差異。
- 結論:中風跟抽煙有關聯。但相比於婚姻、壓力與心臟病，抽煙的 $\hat{\theta}$ 比較靠近1，表示抽煙相對來說不是一個重要的因子

Fix column比較:

- 中風患者: $\frac{\text{Smoke}}{\text{No smoke}} = 1.24$
- 非中風者: $\frac{\text{Smoke}}{\text{No smoke}} = 0.87$
- 符合基本假設，中風患者中有抽煙的比例高

3.1.7 中風與年齡

	Stroke	No Stroke	Row Total
<50	20	2880	2900
50-60	48	786	834
60-70	47	574	621
70-80	94	475	569
>80	40	146	186
Column Total	249	4861	5110

X-square=398.15, df=4, p-value=6.991e-85

3.1.8 中風與血糖

分類依據:

檢驗項目	未得糖尿病	糖尿病前期	糖尿病	糖尿病	嚴重糖尿病
			控制尚可	控制不良	
糖化血色素 HbA1c值 (%)	5.8未滿	5.8~6.5	6.5~7.0	7.0~8.0	8.0以上
空腹血糖(mg/dl)	80~110	110~130	130~160		160以上
飯後2小時血糖 (mg/dl)	80~140	140~180	180~220		220以上

來源：日本糖尿病協會

χ^2 test:

	Stroke	No Stroke	Row Total
<80	63	1467	1530
80-110	69	2038	2107
110-160	28	776	804
>160	89	580	669
Column Total	249	4861	5110

X-square=119.42, df=3, p-value=1.028e-25

3.1.9 中風與BMI

分類依據:

BMI指數怎麼計算？

$$\text{BMI} = \text{體重(公斤)} \div \text{身高}^2(\text{公尺}^2)$$

正常： $18.5 \leq \text{BMI} < 24$
 過重： $24 \leq \text{BMI} < 27$
 輕度肥胖： $27 \leq \text{BMI} < 30$
 中度肥胖： $30 \leq \text{BMI} < 35$
 重度肥胖： $\text{BMI} \geq 35$

 χ^2 test:

	Stroke	No Stroke	Row Total
<24	26	1292	1318
24-27	35	774	809
27-30	50	812	862
30-35	56	944	1000
>35	42	878	920
Column Total	209	4700	4909

X-square=26.56, df=4, p-value=2.441e-05

Chapter 4

Clustering

4.1 The table of correlation

	stroke	gender	age	ever_married	residence	hypertension	work_type	smoking	heart disease	blood glucose	bmi
stroke		-	+	+	-	+	+	+	+	+	+
gender	-			+	-	-			+		
age	+										
ever_married	+	+			-	+		+	+		+
residence	-	-		-		-		+	-		
hypertension	+	-		+	-			-	+		
work_type	+										
smoking_status	+			+	+	-			+		
heart disease	+	+		+	-	+		+			
blood glucose	+										
bmi	+			+							

Bibliography

- [1] 邱弘毅(2018) 腦中風之現況與流行病學特徵.
- [2] Johnson CO, et al. (2019) Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* 18(5):439–458.
- [3] Collaborators TGLRoS (2018) Global, Regional, and Country-Specific Lifetime Risks of Stroke, 1990 and 2016. *New England Journal of Medicine* 379(25):2429–2437.
- [4] 衛生福利部統計處(2018) 死因統計.
- [5] 衛福部-慢性疾病防治組(2020) 健康主題:慢性病防治-腦血管疾病.
- [6] Johnson W, Onuma O, Owolabi M, Sachdev S (2016) Stroke: a global response is needed. *Bulletin of the World Health Organization* 94(9):634–634A.
- [7] Langhorne P, et al. (2000) Medical Complications After Stroke. *Stroke* 31(6):1223–1229.
- [8] Kumar S, Selim MH, Caplan LR (2010) Medical complications after stroke. *The Lancet Neurology* 9(1):105–118.
- [9] Organization WH (2021) Stroke, Cerebrovascular Accident.
- [10] Boehme AK, Esenwa C, Elkind MS (2017) Stroke Risk Factors, Genetics, and Prevention. *Circulation Research* 120(3):472–495.