

研究教育訓練課程

研究計畫統計應用

李其融

中山醫學大學護理系

大綱

- 資料結構
- 資料描述
- 統計方法
 - ▣ 變異數分析
 - ▣ 迴歸分析
 - ▣ 邏輯斯迴歸分析
- 總結

資料結構

以問卷資料為例

資料結構

基本資料

第一部份、人口學資料

1. 妳的出生年次是？ 民國____年____月
 2. 妳的宗教信仰為： 無 基督教 天主教 道教 佛教
 其他 _____
 3. 妳的籍貫是：（以父親為主）
 閩南人 外省人 客家人 原住民 其他
 4. 父、母親是否均健在：
 父、母親均健在
 父親亡故
 母親亡故
 4. 父、母親的最高教育程度：（請分別在父親及母親欄下之內打√，）。
- 父 母
- (1) 未完成國民教育（國小或國中肄業）
 (2) 國中畢業
 (3) 高中畢業、職校畢業、五專肄業
 (4) 大學、大專畢業
 (5) 研究所畢(肄)業（碩士、博士）
 (6) 其他（請註明）_____

問卷題項

第二部分、與異性朋友相處情況（關係權力）

下列題目是描述妳與最近一位與妳發生性行為的異性朋友相處的情況，請妳先回想妳們的情形，再依據與下列陳述符合的程度，以「極不符合」、「小部分符合」、「一半符合」、「大部分符合」及「完全符合」的方式，於每題的右側空格內打「√」。

題目	極不符合	小部分符合	一半符合	大部分符合	完全符合
1. 我不希望他做的事，我能說服他不要做。					
2. 在我與他決定重要事情之前，我會讓他知道我的意見。					
3. 當我想要完成某件事時，我能使他幫助我完成。					
4. 我無法圓滿地處理與他之間的衝突。					
5. 有些事即使我反對，他仍是我行我素。					
6. 在我與他的關係中讓我有受限或受困的感覺。					
7. 在與他的相處中，我仍能按照自己的意願做我想做的事。					
8. 當我認為他的要求不合理時，我會拒絕他。					
9. 在他面前我顯得不如他。					

5等級量表

氣喘兒童母親生活品質問卷

這個部分是詢問您對於自己的生活品質、健康、以及其他生活領域的感覺。請您回答所有問題。如果您對某一問題的回答不確定，請選出五個答案中最適合的一個，通常是您最早想到的那個答案。

我們的問題所關心的是您**最近兩星期內**的生活情形，請您用自己的標準、希望、愉快、以及所關注點來回答問題。請參考下面的例題：

例題一：整體來說，您滿意自己的健康嗎？

極不滿意 不滿意 中等程度滿意 滿意 極滿意

請選出最適合您在最近兩星期內對自己健康的滿意程度，如果您**不滿意**自己的健康，就在「不滿意」前的內打「**✓**」。請仔細閱讀每個題目，並評估您自己的感覺，然後就每一個題目選出最合適您的答案。謝謝您的協助！

1. 整體來說，您如何評價您的生活品質？
極不好 不好 中等程度好 好 極好
2. 整體來說，您滿意自己的健康嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
3. 您覺得身體疼痛會妨礙您處理需要做的事情？
完全沒有妨礙 有一點妨礙 中等程度妨礙 很妨礙 極妨礙
4. 您需要靠醫療的幫助應付日常生活嗎？
完全沒有需要 有一點需要 中等程度需要 很需要 極需要

10等級量表


第一部份：自我效能問卷

這個部分是想了解您在**最近兩星期內**孩子照顧經驗中，自己的信心程度。請您以數字 1 至 10 來圈選，數字愈小表示信心愈小，數字愈大表示信心愈大；若圈選 1 代表完全沒有信心，10 表示非常有信心。如果您對某一問題的回答不確定，請選出答案中最適合的一個，通常是您最早想到的那個答案。

1. 我會注意保護孩子預防他(她)氣喘發作
0 1 2 3 4 5 6 7 8 9 10
| | | | | | | | | | |
2. 我在孩子感冒時會處理他(她)的症狀
0 1 2 3 4 5 6 7 8 9 10
| | | | | | | | | | |
3. 我會適當調整孩子在感冒時的用藥方法
0 1 2 3 4 5 6 7 8 9 10
| | | | | | | | | | |
4. 我會調整全家人的飲食內容減少氣喘孩子的發作機會
0 1 2 3 4 5 6 7 8 9 10
| | | | | | | | | | |
5. 我會安排孩子接受緩和性的運動如游泳、爬山、體操等
0 1 2 3 4 5 6 7 8 9 10
| | | | | | | | | | |
6. 我會交代孩子需劇烈運動前先用 15 分鐘暖身運動
0 1 2 3 4 5 6 7 8 9 10
| | | | | | | | | | |

量化資料

- 性別 (名目資料) :
男=1、女=0
- 教育程度 (順序資料) :
國中以下=0、高中高職=1、大學大專=2、研究所以上=3
- 出生年、月、日
- 身高、體重、BMI、膽固醇含量、ALT...
- 量表分數 (間斷資料? 連續資料?)

- 
- 編號
 - 變數
 - 遺失值：
通常記為“.”
 - **CODEBOOK**

資料結構

變數

編號	性別	出生年	出生月	出生日	教育程度	問題1	問題2	問題...
0001	1	1975	12	.	3	0	1	...
0002	0	1977	2	12	3	.	2	...
0003	1	1982	6	8	2	2	1	...
:								
0101	1	1963	8	3	0	1	3	...
:								
0201	0	1978	4	21	.	0	4	...
:								
1001	1	1969	5	29	1	1	2	...
:								

觀測個體

資料描述

次數分配、集中趨勢、分散程度與關係

次數分配、集中趨勢、分散程度

□ 敘述性統計

□ 間斷資料

- 習慣以百分比或次數表達

Ex1. 氣喘照護研究中，蒐集229個有氣喘孩童的家庭，男童家庭佔69%(158)，女童家庭佔31%(71)

Ex2. 在高中職青少年避孕行為調查的2325份回收問卷裡，有404位青少年於近三個月內有發生性行為，並且每次性行為都有避孕佔37.1%(150)，每次都沒有避孕佔11.9%(48)，其餘佔51.0%(206)

□ 連續資料

- 平均數、變異數(標準差)、中位數、百分位數

關係(relationship)

- 單一變數
 - ▣ 男女比重、性行為避孕頻率、平均年齡、平均膽固醇含量、ALT中位數(百分位數)
 - ▣ 量表分數：次數分配、分數的平均值與標準差
- 關係
 - ▣ 至少兩個變數
 - ▣ 並非所有的兩兩變數之間都能測量皮爾森相關(Pearson's correlation)

變數類型 (Y)	變數類型 (X)	
	連續	間斷
連續		
間斷		

- Y : response, dependent variable
- X : covariate, independent variable

連續變數(Y) vs. 間斷變數(X)

- Y：連續變數

Ex. 年齡、生活品質分數、避孕行為量表分數

X：二分類的間斷變數

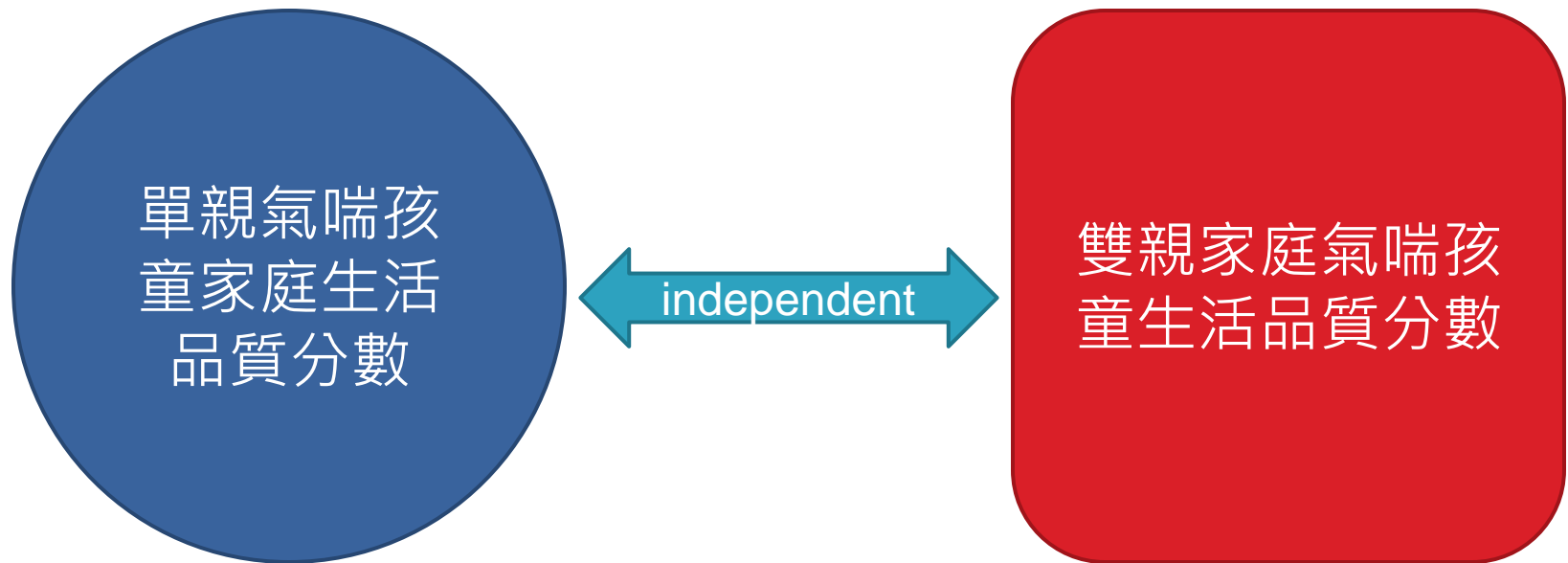
- X表示兩個不同的母體

Ex. X=男生or女生、X=單親or雙親家庭、X=近三個月發生性行為頻率高or低、X=實驗組or對照組

- 兩獨立樣本t檢定(two independent samples t-test)

Ex. 單親與雙親家庭的生活品質分數是否相同、近三個月性行為頻率高與低兩群青少年的避孕行為量表分數是否相同

- 目的在比較兩母體平均數是否相同



□ Y：連續變數

X：二分類の間斷變數

□ X表示相同觀測個體的兩次測量

Ex. X=服藥前or服藥後(同一個病患)、X=長子女or次子女(同一對父母)、X=左眼or右眼(同一個人)

➤ 成對t檢定(paired t-test)

Ex. 比較成人飯前飯後的血糖變化、比較同一家庭的長子女與次子女的e抗原抗體血清轉換年齡

➤ 目的在比較一個成對資料的差異的平均數是否為0

□ Y：連續變數

X：二分類以上的間斷變數

□ X表示兩個以上不同的母體

Ex. 教育程度(高中以下、大專、研究所以上)、家庭收入(低、中、高)、飲酒量(無、適量、過量)、居住地區(北、中、南、東)

➤ 變異數分析(ANOVA)

連續變數(Y) vs. 連續變數(X)

□ Y：連續變數

Ex. 生活品質量表分數、避孕行為量表分數

X：連續變數

Ex. 年齡、兒童氣喘量表分數、關係權力量表分數

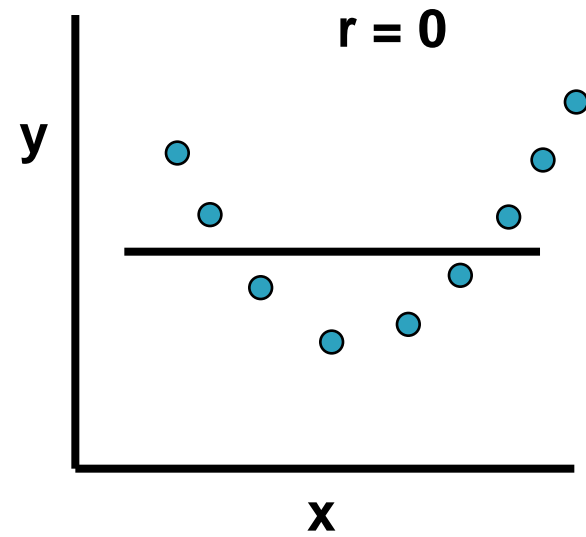
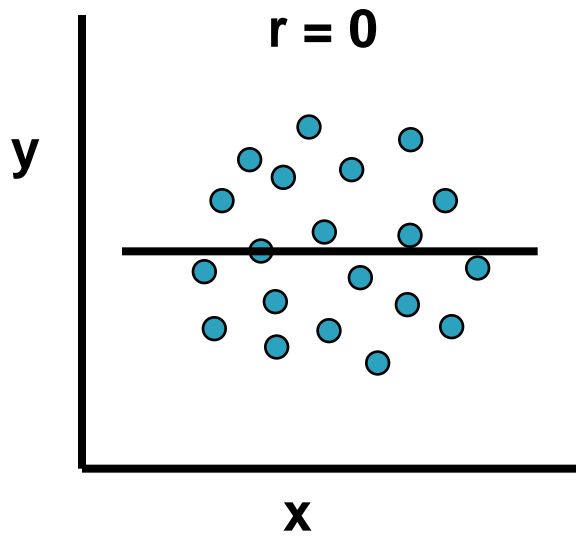
➤ 皮爾森相關(Pearson's correlation)

➤ 目的在提供Y、X兩連續變數之間的線性相關程度

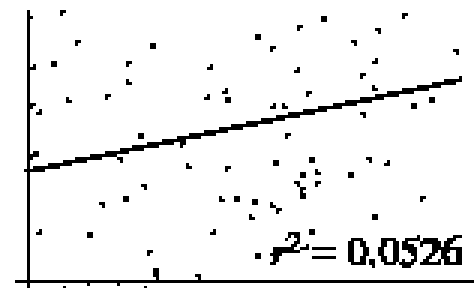
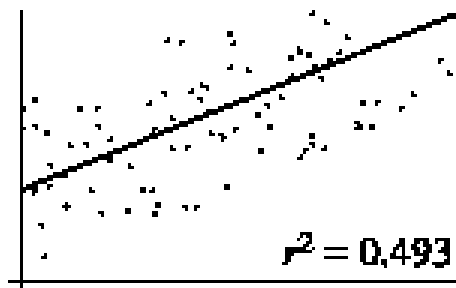
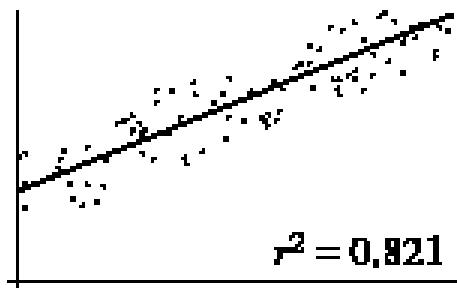
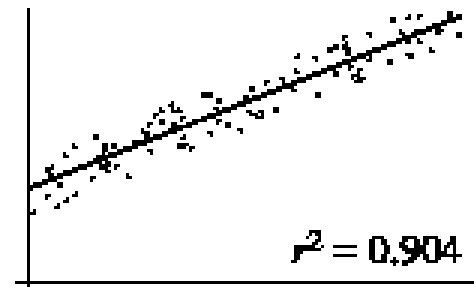
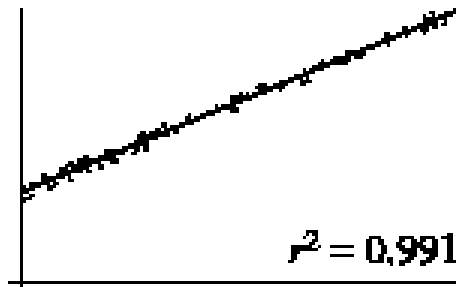
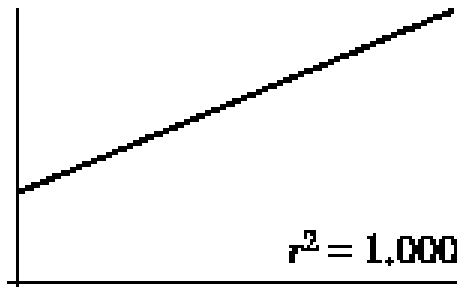
➤ $\rho(\text{rho})$ 、 r

➤ $-1 \leq \rho \leq 1$

➤ 線性相關



➤ 相關數值大小



間斷變數(Y) vs. 間斷變數(X)

- Y：間斷變數

Ex. 生活品質等級；避孕行為量表等級

- X：間斷變數

Ex. 性別、教育程度、兒童氣喘量表等級、關係權力量表等級

- 卡方檢定(Chi-square test)

Ex. 不同生活品質等級的百分比是否因為氣喘孩童的性別而有所差異、不同避孕行為量表等級的百分比是否因為青少年母親教育程度而有所差異

孩童性別	生活品質			total
	劣	普通	優	
男	37 (23.4)	82 (51.9)	39 (24.7)	158 (100.0)
女	19 (26.8)	36 (50.7)	16 (22.5)	71 (100.0)
total	56 (24.5)	118 (51.5)	55 (24.0)	229 (100.0)

避孕行為量表

母親教育程度

劣

普通

優

total

國中以下

高中高職

大學大專

研究所以上

total

□ Y、X：皆為順序資料

➤ Spearman's rank-order correlation

➤ 目的在提供Y、X兩順序變數之間的相關程度

Ex.生活品質等級與兒童氣喘等級的關係、避孕行為量表等級與關係權力等級的關係

➤ 模仿Pearson's correlation

➤ ρ_s 、 r_s

➤ $-1 \leq \rho_s \leq 1$

間斷變數(Y) vs. 連續變數(X)

□ Y：二分類的間斷變數

Ex. Y=存活or死亡

X：連續變數

Ex. X=急性生理改變與慢性健康狀況評估
(APACHE)分數

- 邏輯斯迴歸分析(Logistic regression analysis)
 - Post-ICU mortality會隨著APACHE分數變動而改變
(Chen et al. 2007. ICU discharge APACHE II scores help to predict post-ICU death)

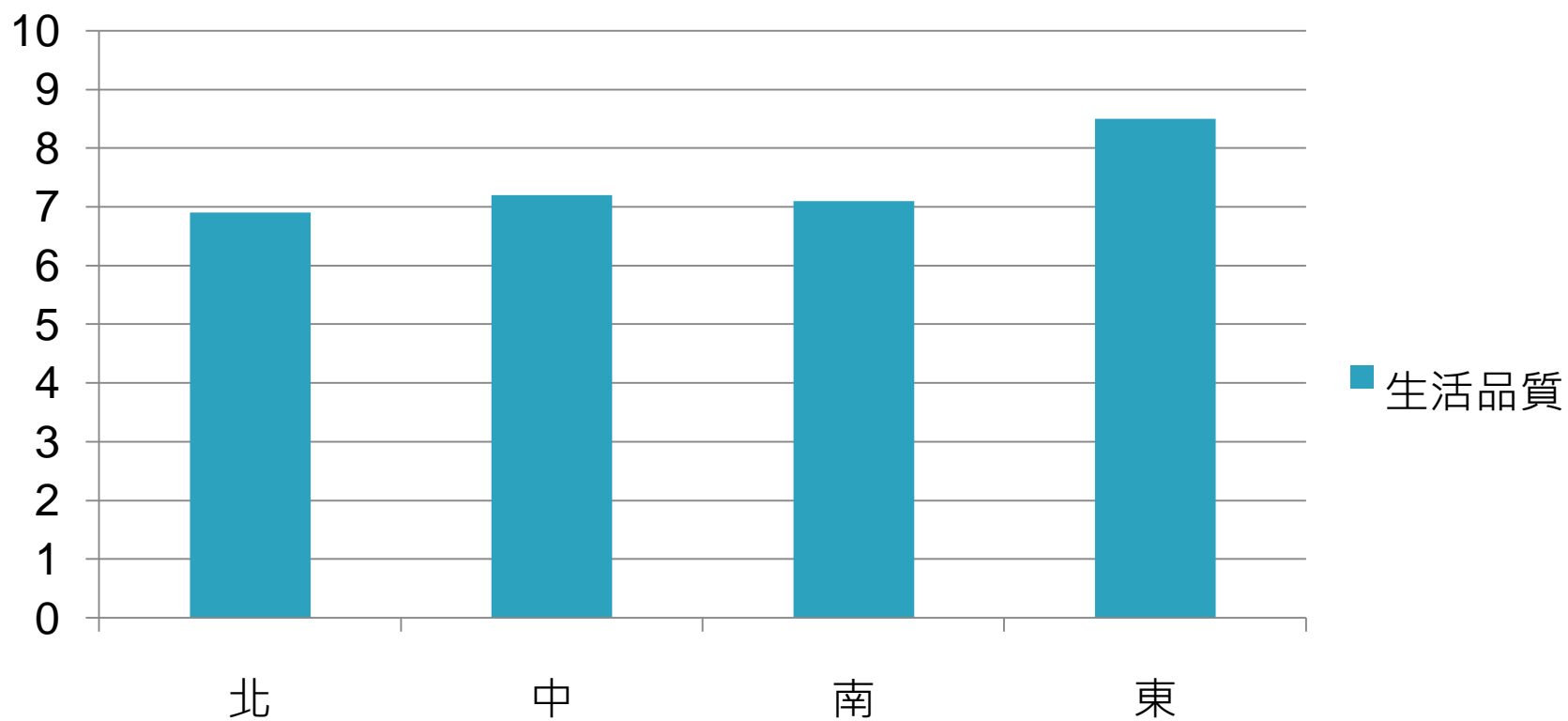
統計分析

模式建立

變異數分析(ANOVA)

- 描述連續變數(Y)與間斷變數(X)的關係
- X可以為二分類以上的間斷變數
Ex. 居住地區、婚姻狀況、教育程度、處置方式
- Q. 欲比較不同居住地區的氣喘家庭生活品質分數是否相同、欲了解青少女避孕行為分數是否因為青少女母親教育程度不同而有顯著差異...
- 變異數分析

氣喘家庭生活品質分數



北部氣
喘家庭
生活品
質分數

獨立

中部氣
喘家庭
生活品
質分數

獨立

南部氣
喘家庭
生活品
質分數

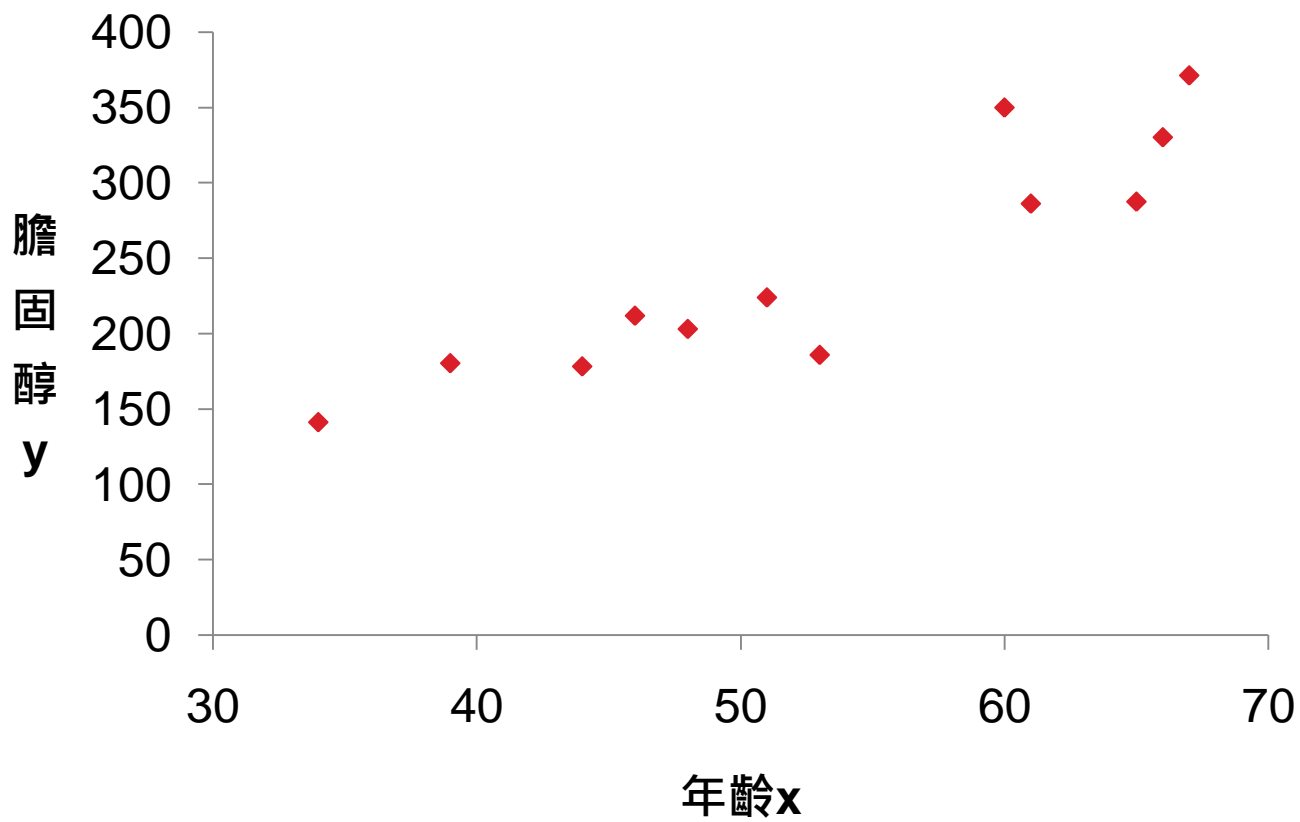
獨立

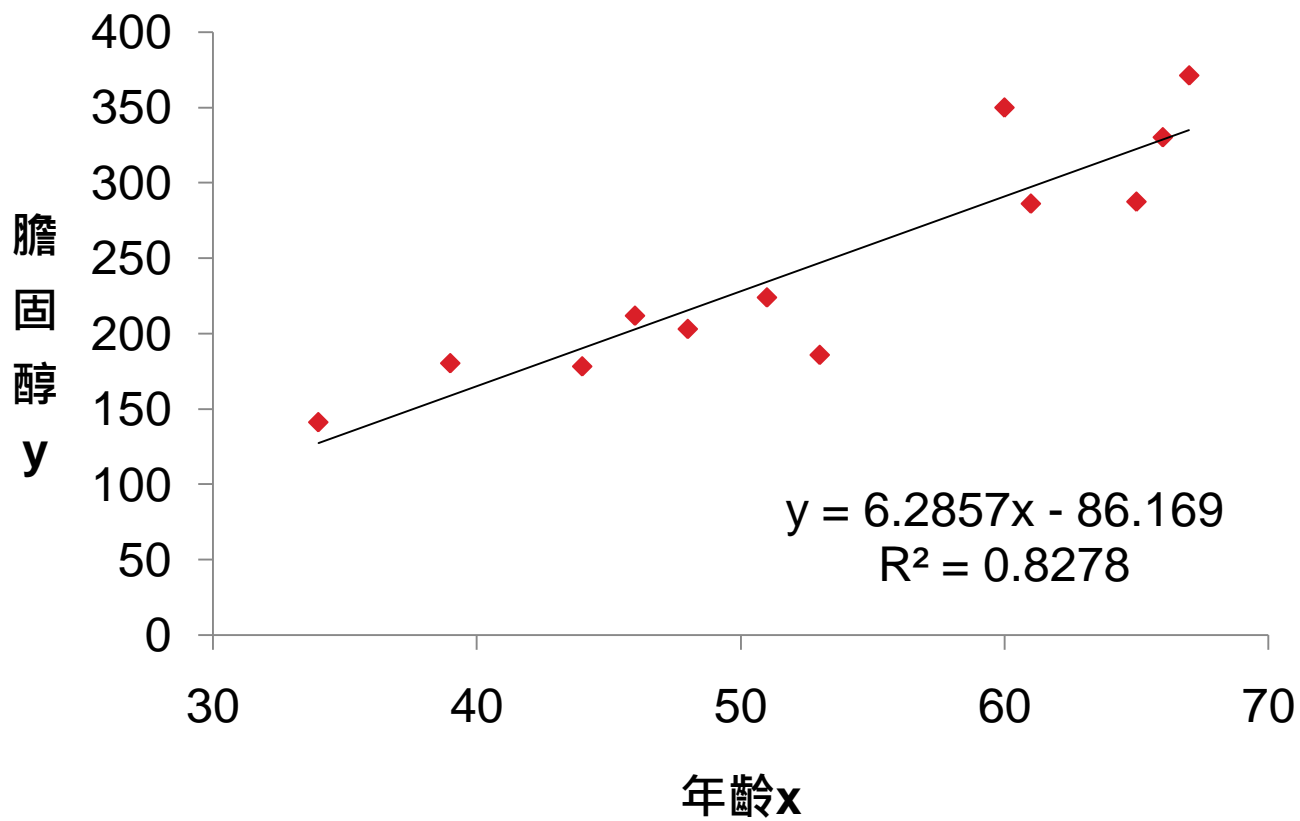
東部氣
喘家庭
生活品
質分數

- 藉著ANOVA證明居住地區(X)是否顯著影響氣喘家庭生活品質分數(Y)的因子
- 如果是→事後比較(Post-hoc comparison)
- 在發現居住地區是影響因子後，才去找哪些居住地區的分數和其他不同
- Scheffe's method (特定對比比較)
Bonferroni method (多個成對比較)
Tukey HSD
Duncan

迴歸分析(Regression Analysis)

- 描述連續變數(Y)與連續變數(X)的關係
- Ex. 避孕行為量表分數(Y)與關係權力量表分數(X)的關係、膽固醇(Y)與年齡(X)的關係
- 因？果？
- Pearson's correlation ?
- Model : $Y = \beta_0 + \beta_1 X + \varepsilon$
 - β_1 : x對Y的影響大小
 - ε : 誤差





- 藉著迴歸分析證明膽固醇與年齡有線性關係
- 年齡與膽固醇的關係？ β_1
 - ▣ 年齡每增加1歲，膽固醇平均增加6.29
- 解釋力？ R^2
 - ▣ 年齡解釋了膽固醇82.78%的變異
- 預測(針對X合理的範圍)
 - ▣ 預測50歲的膽固醇含量為228.12
- 模式診斷：殘差分析

邏輯斯迴歸(Logistic Regression)

- 描述二分類間斷變數(Y)與連續變數(X)的關係

Ex. Y=死亡or存活

- 死亡 \leftrightarrow X ? 存活 \leftrightarrow X ?

- $\pi = P(Y=\text{死亡}) \leftrightarrow X$ (or $1-\pi = P(Y=\text{存活}) \leftrightarrow X$)

Ex. Post-ICU mortality與APACHE分數的關係

- Model

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x$$

$$\Rightarrow \frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 x}$$

□ 勝算(Odd)

$$\text{odd} = \frac{\pi}{1 - \pi}$$

□ 假設 $X=1$ 為男性、 $X=0$ 為女性

$$\text{odd}_{\text{男}} = \frac{\pi_{\text{男}}}{1 - \pi_{\text{男}}} = e^{\beta_0 + \beta_1}$$

$$\text{odd}_{\text{女}} = \frac{\pi_{\text{女}}}{1 - \pi_{\text{女}}} = e^{\beta_0}$$

□ 勝算比(Odds Ratio ; OR)

$$\text{OR} = \frac{\text{odd}_{\text{男}}}{\text{odd}_{\text{女}}} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

□ Ex. APACHE分數增加1分，odd增加 e^{β_1} 倍

$$\text{OR} = \frac{\text{odd}_{x+1}}{\text{odd}_x} = \frac{e^{\beta_0 + \beta_1(x+1)}}{e^{\beta_0 + \beta_1x}} = e^{\beta_1}$$

□ 預測

$$\pi = \frac{e^{\beta_0 + \beta_1x}}{1 + e^{\beta_0 + \beta_1x}}$$

總結

為什麼要做統計分析？



- 研究主題

- ⇒ 提出假設

- ⇒ 蒐集資料

- ⇒ 分析資料

- ⇒ ...

- 統計學家不會比我更了解我的研究

- 變異數分析可以有更多的因子(factors)
(X_1 、 X_2 、...、 X_k)
- 迴歸分析可以有更多的共變數(covariates)
(X_1 、 X_2 、...、 X_p)
- 邏輯斯迴歸分析也可以有更多的共變數



- 更多的分析

隨機效果模式分析(Random effect model)、重複測量分析(Repeated measurement model)、因素分析(Factor analysis)、路徑分析(Path analysis)、結構方程式模式分析(SEM)...

- 參考書目

Polit, D. F. (1996). *Data Analysis & Statistics for Nursing Research*. Appleton & Lange