



2. 資料輸入與基本整理

鄒慶士 (Ching-Shih Tsou)
台北商業技術學院資訊與決策科學所
E-mail : cstsou@mail.ntcb.edu.tw



統計是研究資料的學問

- 如何輸入資料？
- 如何整理(manipulate or organize)資料？
- 如何進行彙總計算，產生資訊。
- 產生圖形，表達意涵。(後續主題)



向量資料輸入 concatenate函數

- 一份八頁文稿初版與再版各頁的錯誤字數
 - `typos.draft1 = c(2,3,0,3,1,0,0,1)`
 - `typos.draft2 = c(0,3,0,3,1,0,0,1)`
- 也可以用類似試算表的界面輸入
 - `data.entry("typos.draft1")` #資料物件須先存在
- 修改資料
 - `fix(typos.draft1)`



向量取值[] (同矩陣與陣列[,][,,]...)

- 取單一值
- 取連續數個值
- 間隔取數個值
- 察看哪些元素是遺缺值
 - `is.na()`



善用函數

- 再版各頁最差有多少錯誤？
 - max()
- 再版文稿有多少頁？
 - length()
- 那幾頁錯誤最多？
 - ==
 - which()



善用函數(續)

- 整份文稿共有多少錯誤？
 - sum()
- 有錯誤的頁數共幾頁？
 - sum(typos.draft2 > 0)
- 初版與再版各頁錯誤相差多少？
 - typos.draft1 - typos.draft2
- 再增加兩頁的錯誤值？
 - typos.draft2 <- c(typos.draft2,0,1)，或
 - typos.draft2[9:10] <- c(0,1)



邏輯運算子

- < (小於)
- > (大於)
- <= (小於等於)
- >= (大於等於)
- == (等於)
- != (不等於)
- & 或 && (AND)
- | 或 || (OR)
- ! (NOT)

© Vince Tsou, IDS, NTCB

100年度教育部補助技專校院建
立特色典範計畫

7



基本敘述統計函數

- 德州海灘每年擱淺的鯨魚數
 - whale=scan()
 - 74 122 235 111 292 111 211 133 156 79
- 平均數與變異數
 - mean(), var()
- 標準差呢？
 - std <- function(x) { sqrt(var(x)) }
 - sd()

© Vince Tsou, IDS, NTCB

100年度教育部補助技專校院建
立特色典範計畫

8



Summary – 向量資料的存取

- 有多少元素？
- 第 i 個元素？
- 取出所有但除了第 i 個元素。
- 取出前 k 個元素。
- 取出後 k 個元素。
- 取出特定幾個不連續的元素。
- 取出大於某值的元素。
- 取出大於某值或小於某值的元素。
- 最大元素的指標為何？



Answers- 資料的存取

- `length(x)`
- `x[2]`
- `x[-2]` #排除法
- `x[1:5]`
- `x[(length(x)-5):length(x)]`
- `x[c(1,3,5)]` #向量取值與concatenate函數併用
- `x[x>3]` #善用邏輯索引
- `x[x<-1 | x>2]` #善用邏輯索引
- `which(x == max(x))`

