

STATA®



S F I

昊青股份有限公司
SCIENTIFIC FORMOSA, INC.

台中榮總2023/01/09

STATA 入門體驗

昊青公司 蕭鎮宇

02-25050525

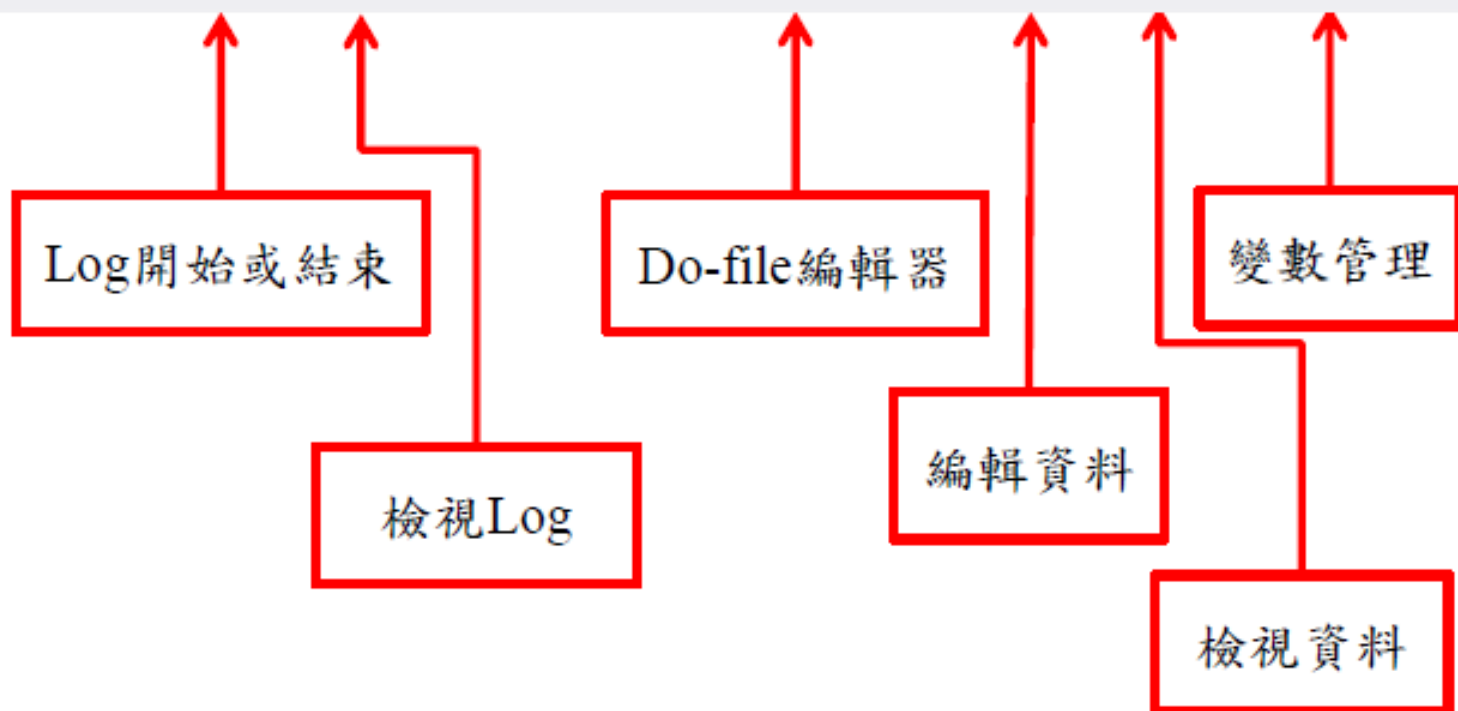
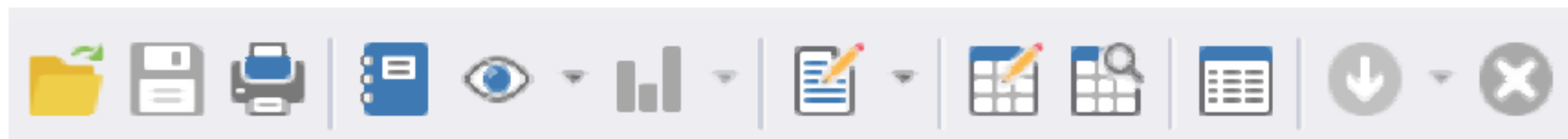
課程大綱

- 操作介面簡介
- 匯入資料
- 敘述統計
- 檢視資料 & 繪圖
- t檢定

STATA簡介

- 目前版本為**STATA17**
- 以**CLI**介面為主的統計軟體
- 常見應用於經濟、社會科學、管理、公衛、醫學等領域
- 能夠獨立完成資料整理、模型估計、圖表報表輸出
- **Do-file**易於相互交換，可與研究同伴互相交流資源；多人共同研究更容易溝通、確認研究想法

工具列



資料類型(data type)

Storage type	Minimum	Maximum	Closest to 0 without being 0	Bytes
byte	-127	100	± 1	1
int	-32,767	32,740	± 1	2
long	-2,147,483,647	2,147,483,620	± 1	4
float	$-1.70141173319 \times 10^{38}$	$1.70141173319 \times 10^{38}$	$\pm 10^{-38}$	4
double	$-8.9884656743 \times 10^{307}$	$8.9884656743 \times 10^{307}$	$\pm 10^{-323}$	8

- 數字
- 字串(string)
- 時間
- 布林(Boolean)

String storage type	Maximum length	Bytes
str1	1	1
str2	2	2
...	.	.
...	.	.
...	.	.
str2045	2045	2045
strL	2000000000	2000000000

資料探索

- describe:檢視資料屬性
- summarize:摘要變數基本統計
- tabulate:呈現類別變數的次數分配
- browse:瀏覽(特定)資料
- label:製作資料標籤

資料整理

- generate:產生新變數
- replace:修改現有變數
- rename:重新命名變數
- drop/keep:(保留/剔除)(變數/樣本)
- egen:產生統計量欄位
- collapse:產生統計量資料表

小技巧

- `cd`:可以導向指定資料夾路徑，方便儲存呼叫檔案
- `help`:呼叫指令說明
- `search`:找尋相關指令(不知指令名稱時)
- 指令/變數名稱大多可縮寫，節省打字時間
- “=”與“==”區別: =為指定(assign)之意，若要表達相等(equal)的情況應該使用==
- `do-file`裡綠色文字為註解，不會被執行
- `do/run`:都會執行`do-file`，但`run`不會顯示執行結果(除了錯誤訊息)

統計圖

- graph bar: 長條圖
- histogram: 次數分配圖
- two way~: 同時有X,Y軸的圖，如散布圖、時間序列圖
 - ✓ scatter: 散布圖
 - ✓ lfit: 迴歸線
 - ✓ line: 折線圖
- graph save: 儲存為stata圖片
- graph export: 匯出圖片為其他格式

統計檢定

- `ttest`: 各類t檢定(檢定平均數)
- `sdtest`: 檢定變異數(卡方或F檢定)
- `oneway`: 單因子變異數分析(anova)
- `anova`: 各類anova檢定分析

Log檔

- 用以紀錄執行過程(必須搭配do，而非run)
- 開始:
- Log using *filename*, replace ///覆蓋現有檔案(如果已存在)
- Log using *filename*, append ///附加在現有檔案之後(如果已存在)
- 結束: log close
- 轉檔: translate *filename.smcl filename.txt* (**.txt, **.log, **.pdf)

```
cd "C:\業務資料\榮總 STATA"
```

```
*** 0 Import Data
```

```
sysuse auto ,clear
```

```
import delimited auto, clear
```

```
import excel using Demo , first sheet("Sheet") clear
```

```
save demo, replace
```

```
***1-1 data exploration
```

```
use demo ,clear
```

```
describe
```

```
summarize
```

```
summarize EF
```

```
summarize EF, detail
```

```
bys SEX: summarize EF
```

```
tabulate SEX
```

```
tabulate SEX age_gr
```

```
browse if CASE==412
```

```
label define gender 1 "male" 2 "female"
```

```
label value SEX gender
```

```
label dir
```

```
label list
```

```
***1-2 data preparation
```

```
use demo ,clear
```

```
generate new_EF=EF^0.5
```

```
replace age=ceil(age)
```

```
generate patient_type="young"
```

```
replace patient_type="old" if age>=31
```

```
rename EF ef
```

```
egen m_ef=mean(EF)
```

```
egen m_cad_ef=mean(EF), by(cva0)
```

```
drop if cva0==1
```

```
drop new_EF
```

```
use lbw_mod, clear
```

```
collapse (mean) lwt, by(race smoke)
```

```
export excel using table1028, replace first(var)
```

```

***2 graph
use lbw_mod, clear

graph bar, over(race)
graph bar (count) , over(race)
graph bar (mean) lwt , over(race)
graph hbar (mean) lwt , over(race)

histogram lwt
histogram lwt, width(25)
histogram lwt, bin(15)

twoway scatter bwt lwt
twoway scatter bwt lwt, by(race)
twoway scatter bwt lwt, title(Mother's Weight & Newborn's Weight) ytitle(Newborn's Weight
(gram)) xtitle(Mother's Weight (lb))
twoway scatter bwt lwt || lfit bwt lwt , title(Mother's Weight & Newborn's Weight)
ytitle(Newborn's Weight (gram)) xtitle(Mother's Weight (lb)) by(smoke)
graph save weight, replace
graph export weight.png, as(png) replace

***3-1 t-test, one-way/two-way anova
use demo ,clear
ttest EF==0.7
ttest EF, by(SEX)
sdtest EF, by(SEX)
ttest EF, by(SEX) unequal

oneway EF age_gr, t
anova EF age_gr
anova EF age_gr cva0 age_gr#cva0
table age_gr cva0 , stat(mean EF)

*** 4.log
log using test0109, replace
/*
Your Command Here
*/
log close
translate test0109.smcl test0109.txt

```