

# SAS 統計軟體介紹二

## --paired-t test

吳淑娟

2006/11/03

生物統計學實習

# SAS 程式的運作

## SAS 程式

所有的SAS程式都是由一系列的SAS STEP組成

- DATA STEPS用來建立SAS DATA SETS
- PROCEDURE STEPS用來分析或處理SAS DATA SETS

## SAS程式的步驟(STEPS)

SAS程式包含DATA STEPS和PROC STEPS. 每一個STEP是由一個或更多的敘述(STATEMENTS)所組成

### SAS STATEMENTS

- 通常由特定的關鍵字開頭
- 用一個分號[;]作為敘述的結束
- 可在一行的任何位置開始或結束, 或可跨越多行

在DATA STEP中,需要合理的程式語法來處理資料

PROC STEP中只告訴SAS系統要作什麼,不用指出如何作

建議在程式撰寫中,DATA, PROC 和RUN寫在每行的第一格,其他敘述內縮

# SAS DATA Step

## SAS DATA SETS和變數(Variables)的命名

有很多的敘述中包含有[資料集]和[變數]的名稱在內  
所有的SAS命名規則都要

- 包含1-8個字元
- 必須以A-Z或底線[\_]開頭
- 第二個字每開始可用任何字母或數字或底線

建議:用有意義的文字命名較能瞭解資料含義

注意:所有的SAS名稱是不管字母大小寫的

## 讀取SAS程式中的資料

DATA步驟提供完整的程式語法以建立和維護SAS資料集

- SAS程式中可直接讀取條例式的資料 → CARDS

- 由外部檔案讀取資料 → INFILE

SAS程式中可直接讀取外部檔案中的條例式資料

# SAS DATA Step

## 原始資料輸入

你可能已經有不同的資料格式,SAS提供多種可用的方式,將你的資料格式轉入SAS資料集中

在DATA步驟中,SAS能讀取程式內資料和外部資料

- 使用DATA敘述準備建立SAS資料集
- 使用CARDS告知SAS,資料就在下面程式中或使用INFILE告知SAS資料在那個外部檔案中
- 使用INPUT去定義變數名稱和屬性,並定義讀取資料的方法

# SAS DATA Step

## DATA敘述

DATA敘述有二個主要功能

- 代表DATA步驟的開始
- 命名即將建立的資料集

DATA敘述的一般格式如下: DATA SASdataset(s);

```
DATA STUDENT;
```

```
INPUT NAME $ SEX $ HEIGHT WEIGHT;
```

```
CARDS;
```

```
..... data lines
```

```
RUN;
```

# SAS DATA Step

## 讀取原始資料(INPUT)

- 行輸入

假如你的資料是整齊的排列在每一行中, 那你就可以用行輸入來讀取資料  
當使用行輸入讀取資料時,

- 正確告知那個變數的資料在那幾行間
- 使用[\$]指出這個變數是文字變數

**INPUT敘述的一般格式: INPUT variable [\$] startcol-endcol;**

```
DATA STUDENT;  
  INPUT NAME $ SEX $ HEIGHT WEIGHT;  
  CARDS;  
  王小明 M 34.5 125.5  
  王淑德 F 35.0 135.0  
  李大同 M 37.6 128.5  
  孫小威 M 40.0 136.5  
  林文華 F 36.5 35.5  
;
```

# SAS DATA Step

- 行輸入時的資料值

## 用行輸入讀入文字變項時

- 變數內容左靠
- 變數的長度由INPUT敘述的行寬決定
- 文字變數的最大長度是200

## 用行輸入讀入數字變項時

- 原始資料的數值可填在欄位中的任何位置
- 可在數字中加入正負號,小數點或科學符號

# SAS DATA Step

## • 格式輸入

格式輸入是另一種讀取原始資料的方法,幾乎可以讀取任何形態的原始資料  
使用格式輸入時,

- 使用指標控制器,移動指標到要讀取資料的位置
- 取一個變數名稱
- 用一個informat(格式)去描述原始資料

格式輸入的一般格式為: **INPUT** 指標控制器 變數名稱 格式;  
指標控制器

- @n 移動指標至第n行
- +n 往後移動n行

讀取標準文數字用的格式

- w. 讀取w位欄寬的標準數字
- w.d 讀取w位欄寬的標準數字並加上d位數的小數
- \$w. 讀取w位元欄寬的標準文字,略去帶頭的空白格
- \$VHARw. 讀取w位元欄寬的標準文字,保留帶頭的空白格

注意:所有的格式必須在最後加上**PERIOD[.]**

簡例: **INPUT NAME \$8. @10 SEX \$1. +1 HEIGHT 7. +1 WEIGHT 4.;**

# SAS DATA Step

- 混合輸入  
在INPUT中同時使用行輸入和格式輸入讀取資料
- 讀取外部原始資料  
SAS可以從程式中或外部檔案中直接讀取原始資料,而後建立的SAS資料集內容是完全一樣的,從程式中讀取資料是使用**CRARDS**敘述,從外部檔案中是使用**INFILE**敘述

# SAS PROC Step

SAS系統的一些基本程式(PROCedures)

PROC步驟會從SAS的程式館藏中呼叫一程式使用,以下列出一  
些程式

## STATISTICAL

CORR

FREQ

MEANS

SUMMARY

TABULATE

UNIVARIATE

## REPORT WRITING

CHART

FORMS

PLOT

PRINT

## UTILITIES

APPEND

COMPARE

CONTENTS

FORMAT

PRINTTO

SORT

TRANSPOSE

## SCORING

RANK

STANDARD

# SAS PROC Step

## 起動SAS PROC(程式)

PROC敘述呼叫一個SAS程式，SAS程式是預先寫好的SAS程式，可作下列事項

- 讀取SAS資料集
- 建立SAS資料集
- 計算統計量
- 印出結果報告
- 處理資料

PROC的一般格式: PROC program options;

- **PROC PRINT;**
- **PROC PRINT DOUBLE NOBS DATA=MONTH;**

DATA= 選項指出要處理那個SAS資料，如果省略此選項，SAS將處理最近產生的SAS資料集

# SAS 幾個常用敘述

加入表頭行表尾行

TITLE敘述可定義在每頁表頭印出Title,最多可定義10行TITLE行

- TITLE的一般格式為: TITLE $n$  'Ttitle';

FOOTNOTE敘述可定義在每頁表尾印出FOOTNOTE,最多可定義10行FOOTNOTE行

- FOOTNOTE $n$ 的一般格式為: FOOTNOTE $n$  'Footnote';

當使用TITLE(FOOTNOTE)定義TITLE $n$ 時,那麼比 $n$ 大的TITLE行,將被抑制不印出

```
PROC PRINT DATA=STUDENT;  
TITLE '中山國民小學';  
TITLE '82年度學童身高 體重記錄';  
FOOTNOTE '製表日期:88/07/30';  
RUN;
```

TITLE 和FOOTNOTE敘述是屬於全域性的,一旦定義了TITLE,那麼往後的報表皆會印出,直到重新定義TITLE或重新起動SAS

重新定義TITLE和取消FOOTNOTE

```
PROC PRINT DATA=STUDENT;  
TITLE '中山國小 82年度學童身高 體重記錄';  
FOOTNOTE ;  
RUN;
```

# SAS 幾個常用算式

## 各種運算式 常式

- **N=0;** 數字常式
- **DEPART='IBMS'** 文字常式

## 簡單運算

- **SUM=X+Y;** 加法
- **DIFF=X-Y;** 減法
- **TWICE=X\*2;** 乘法
- **HALF=X/2;** 除法
- **CUBE=X\*\*3;** 次方
- **Neg=-X;** 改變正負號

## 混合運算

- **A=X-Y+Z** 由左至右
- **A=X+Y\*Z** 先乘除後減
- **A=X/Y/Z** 由左至右
- **A=X/(Y/Z)** 先計算()

## 函數

- **S=SQRT(X);**
- **A=ABS(X);**
- **Z=ABS(SQRT(X)-2);**

# SAS 幾個常用函數

## SAS的函數可分為下列幾類

- 算數函數
- 截斷函數
- 數學函數
- 三角及雙曲線函數
- 機率函數
- 計量函數
- 樣本統計函數
- 亂數函數
- 文字函數
- 日期及時間函數
- 特殊函數

函數的一般格式為：**FUNCTIONNAME(引數1,引數2, ...)**;

# SAS 幾個常用統計函數

常用的統計函數有

- CV 變異係數
- MAX 最大值
- MIN 最小值
- MEAN 平均值
- N 非缺失值的個數
- RANGE 值域
- STD 標準差
- SUM 總和
- VAR 變方

函數的一般格式為：**FUNCTIONNAME(引數1,引數2, ...)**;

# SAS 幾個常用統計函數

## 函數和程式的不同

- 函數和程式都可以用來計算一些統計量
- 函數是用來計算每個觀測值的統計量(橫向計算)
- 程式是用來計算一個變數的統計量(縱向計算)

# 有條件地執行SAS程式

- **IF-THEN / ELSE敘述**

使用**IF-THEN**敘述可以依據某些條件執行部份**SAS**敘述

**IF-THEN**的一般格式為：

- **IF** 條件式 **THEN** 敘述；
- **ELSE** 敘述；

比較運算

- **EQ** 或 **=** 等於
- **NE** 或 **^=** 或 **~=** 不等於
- **LT** 或 **<** 小於
- **GT** 或 **>** 大於
- **LE** 或 **<=** 小於等於
- **GE** 或 **>=** 大於等於

# DO和END敘述

## DO和END敘述

DO敘述和它的配對敘述,可以將一群敘述包裝成一個單元,在某一條件下一起執行

DO和END敘述的一般格式為

```
IF 條件式 THEN DO;  
    欲執行的敘述;  
END;
```

# PROC MEANS

MEANS程序可對數字變數計算描述性統計量

MEANS程序的一般格式為：

- **PROC MEANS options;**

一般的選項有**DATA=**資料集, **MAXDEC=n**指出報表中的小數點位數

統計選項有：**CV MAX MIN MEAN NMISS PRT RANGE  
STD SUM T VAR**等

在MEANS程序中使用的敘述有

- **CLASS** 類別變數；依類別分開統計, 可為數字或文字
- **VAR** 分析變數；指出欲統計的變數名稱