



SAS統計軟體操作 (2)

醫學研究部 基礎醫學科 生統小組：陳韻仔 博士

授課日期：113年6月5日

Analyzing Data

Descriptive Statistics

- Organise
- Summarise
- Simplify
- Describe and present data

Inferential Statistics

- Generalise from samples to populations
- Hypothesis testing
- Make predictions

Five types of statistical analysis

Descriptive

What are the characteristics of the respondents?

Inferential

What are the characteristics of the population?

Differences

Are two or more groups the same or different?

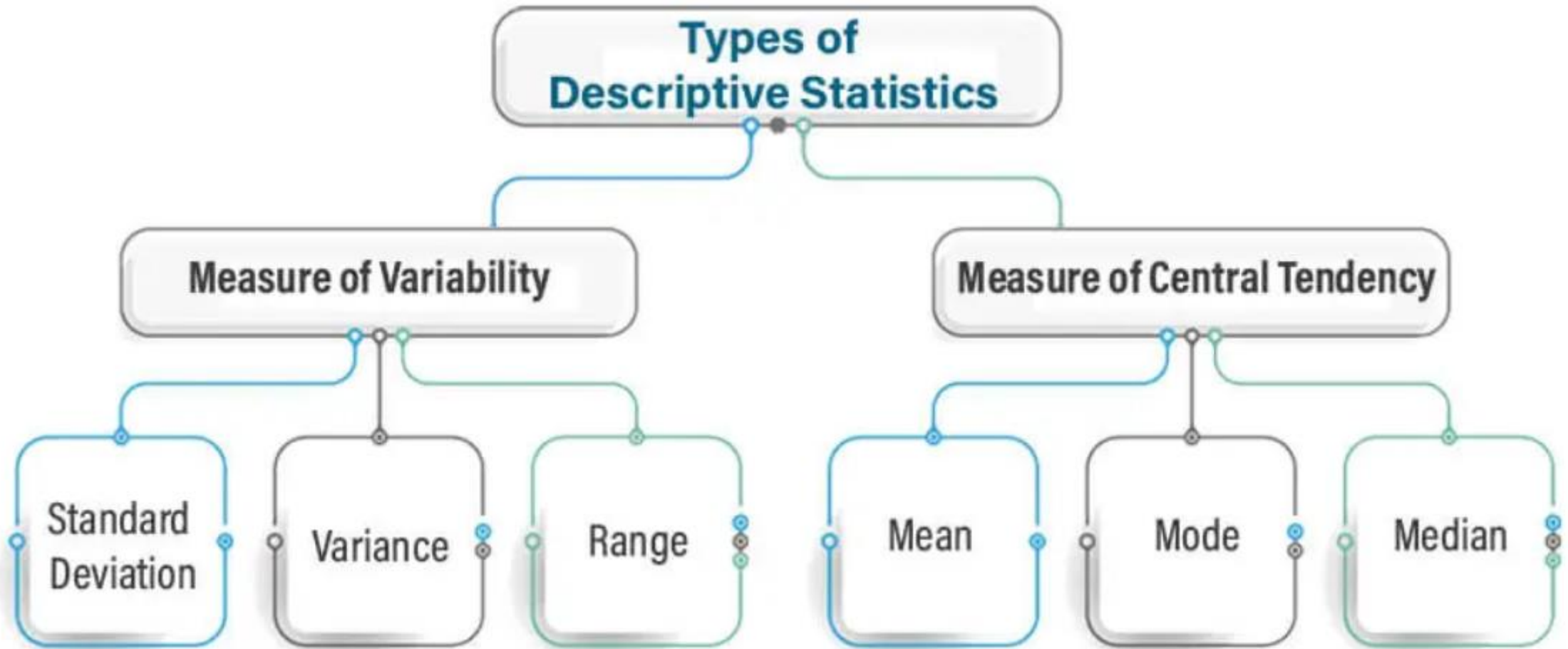
Associative

Are two or more variables related in a systematic way?

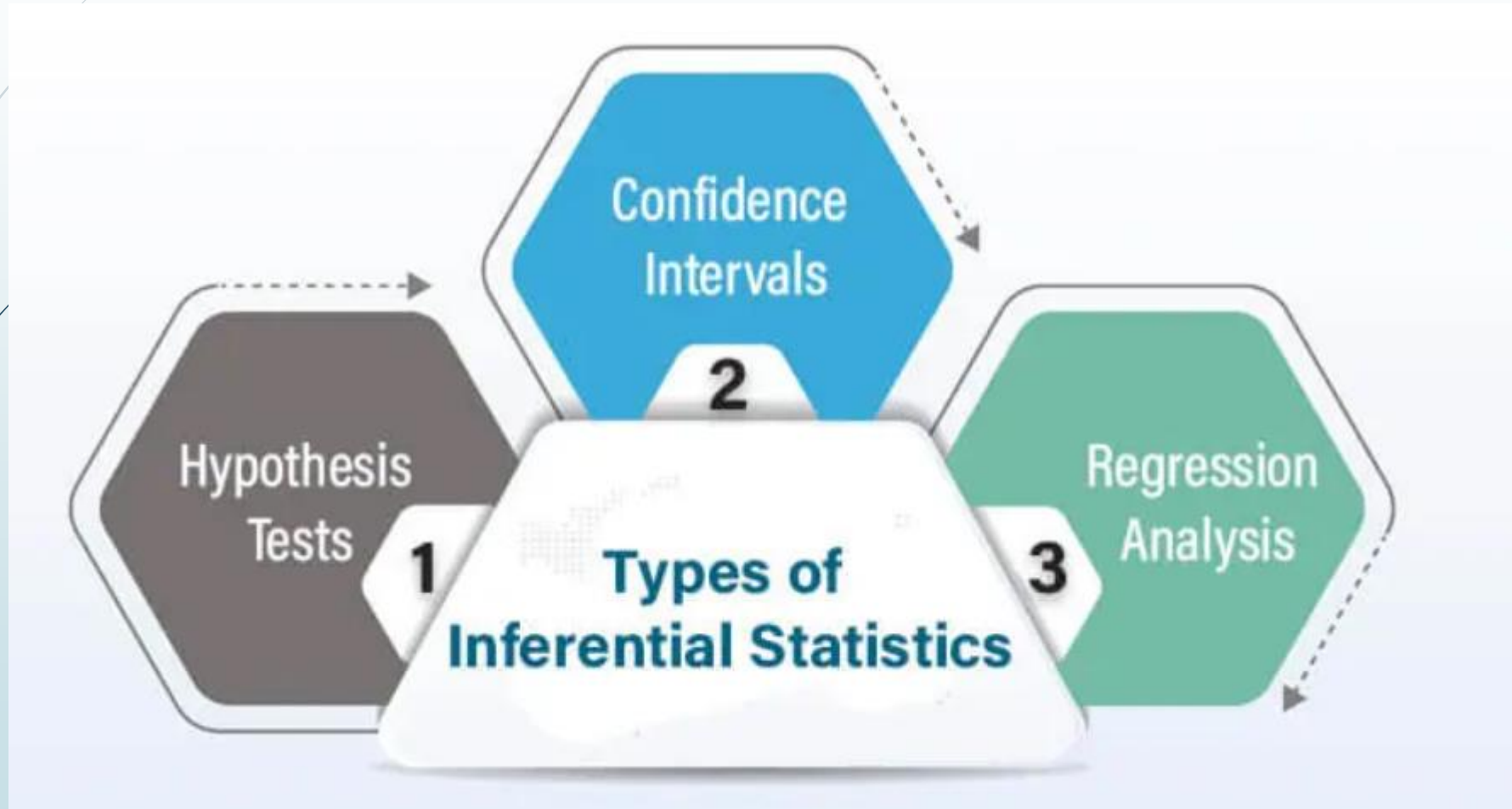
Predictive

Can we predict one variable if we know one or more other variables?

臨床試驗 常見統計方法



臨床試驗 常見統計方法





Europace (2014) **16**, 174–181
doi:10.1093/europace/eut293

CLINICAL RESEARCH

Atrial fibrillation

Effect of dronedarone on clinical end points in patients with atrial fibrillation and coronary heart disease: insights from the **ATHENA** trial

Ron Pisters¹, Stefan H. Hohnloser², Stuart J. Connolly³, Christian Torp-Pedersen⁴, Lisa Naditch-Brûlé⁵, Richard L. Page⁶, and Harry J.G.M. Crijns^{1*} for the ATHENA Investigators

¹Department of Cardiology, Maastricht University Medical Centre, PO Box 5800 6202 AZ, Maastricht, The Netherlands; ²Department of Cardiology, J. W. Goethe University Hospital, Frankfurt, Germany; ³McMaster University, Population Health Research Institute, Hamilton, Ontario, Canada; ⁴Institute of Health, Science and Technology, Aalborg University, Aalborg Øst, Denmark; ⁵Sanofi, Paris, France; and ⁶University of Washington, Seattle, USA

Received 15 April 2013; accepted after revision 28 August 2013; online publish-ahead-of-print 9 October 2013

Table 1 Baseline characteristics of patients with and without coronary heart disease^a

	CHD		No CHD	
	Placebo (n = 737)	Dronedarone (n = 668)	Placebo (n = 1590)	Dronedarone (n = 1633)
Mean age, years (SD)	73.5 (8.2)	73.1 (7.7)	70.8 (9.3)	70.9 (9.3)
Male gender	485 (65.8%)	419 (62.7%)	804 (50.6%)	751 (46%)
Hypertension	639 (86.7%)	593 (88.8%)	1357 (85.3%)	1406 (86.1%)
Hypercholesterolemia	436 (59.2%)	416 (62.3%)	566 (35.6%)	618 (37.8%)
Diabetes mellitus	199 (27.0%)	166 (24.9%)	264 (16.6%)	316 (19.4%)
Chronic renal failure	38 (5.2%)	40 (6.0%)	45 (2.8%)	45 (2.8%)
Congestive heart failure	287 (38.9%)	261 (39.1%)	406 (25.5%)	411 (25.2%)
NYHA class III	67 (9.1%)	58 (8.7%)	42 (2.6%)	33 (2.0%)
LVEF <35%	58/723 (8.0%)	52/658 (7.9%)	29/1558 (1.9%)	40/1605 (2.5%)
Oral anticoagulant	436 (59.2%)	414 (62.0%)	948 (59.6%)	989 (60.6%)
Low dose of aspirin (≤ 365 mg)	413 (56.0%)	390 (58.4%)	606 (38.1%)	628 (38.5%)
Beta-blocking agents ^b	559 (75.8%)	534 (79.9%)	1082 (68.1%)	1094 (67.0%)
ARB or ACE inhibitor	551 (74.8%)	495 (74.1%)	1051 (66.1%)	1119 (68.5%)
Statins ^c	453 (1.5%)	429 (64.2%)	461 (29.0%)	449 (27.5%)

ACE, angiotensin-converting enzyme; ARB, angiotensin II receptor blocker; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association.

^aData are numbers (%) unless otherwise specified.

^bNot including sotalol.

^cStatins are defined as 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors.

Mean (Std)

N (%)



Efficacy and Safety of Empagliflozin in Renal Transplant Recipients With Posttransplant Diabetes Mellitus

Diabetes Care 2019;42:1067–1074 | <https://doi.org/10.2337/dc19-0093>



*Thea Anine Strøm Halden,¹
Kine Eide Kvitne,² Karsten Midtvedt,¹
Laavanyaah Rajakumar,¹ Ida Robertsen,²
Jan Brox,³ Jens Bollerslev,^{4,5}
Anders Hartmann,¹ Anders Åsberg,^{1,2} and
Trond Jenssen^{1,4}*

Table 1—Baseline characteristics presented as median (absolute range) or number of patients (%)

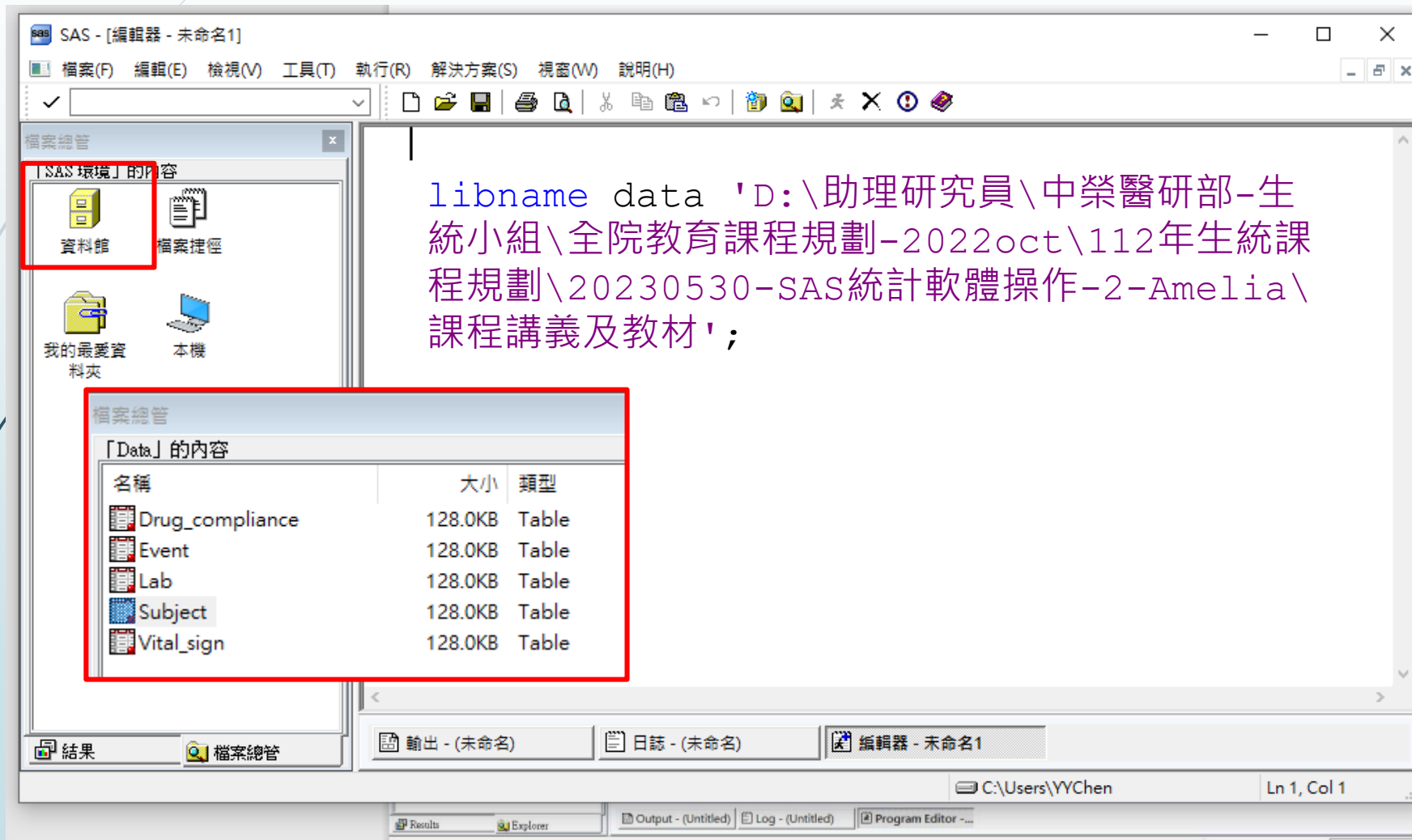
	Empagliflozin, <i>n</i> = 22	Placebo, <i>n</i> = 22
Sex (male/female), <i>n</i>	17/5	17/5
Age (years)	63 (31, 72)	59 (21, 75)
Time since transplantation (years)	3 (1, 16)	3 (1, 15)
BMI (kg/m ²)	28.8 (24.7, 39.3)	27.5 (22.4, 45.8)
WHR (cm)	1.01 (0.82, 1.25)	0.98 (0.80, 1.11)
Systolic blood pressure (mmHg)	143 (111, 176)	140 (100, 163)
Diastolic blood pressure (mmHg)	79 (63, 94)	82 (55, 94)
HbA _{1c} (%)	6.9 (6.5, 8.2)	6.8 (6.1, 7.2)
HbA _{1c} (mmol/mol)	52 (38, 83)	51 (40, 73)
FPG (mmol/L)	8.0 (5.0, 13.1)	7.3 (4.5, 12.5)
eGFR (mL/min/1.73 m ²)	66 (41, 83)	59 (44, 82)
LDL (mmol/L)	2.8 (1.2, 4.2)	2.8 (2.0, 3.8)
Triglycerides (mmol/L)	1.8 (1.1, 3.2)	2.2 (1.1, 5.6)
Smoking, <i>n</i>		
Smoker	4 (18.2)	0 (0.00)
Ex-smoker	13 (59.1)	11 (50.0)
Never smoked	5 (22.7)	11 (50.0)

Median (IQR)

N (%)

SAS 9.4 起始畫面

設定SAS永久檔位置



利用SAS進行常態性檢定資料分佈：

```
/*Normality Test*/  
data lab; set data.final;  
if visit=1; keep Group age SBP AST Creatinine; run;  
  
proc univariate data=lab normal ;  
var sbp;  
class group; run;
```

SAS 系統

UNIVARIATE 程序
變數: SBP (SBP)
Group = 2

Mean (std)

動差			
N	34	總和權重	34
平均值	130.558824	總和觀測	4439
標準差	18.6664359	變異數	348.435829
偏態	0.42712288	峰度	0.42294941
未校正平方和	591049	校正平方和	11498.3824
變異係數	14.2973377	標準誤差平均值	3.20126734

基本統計量值

位置		變異性	
平均值	130.5588	標準差	18.66644
中位數	129.5000	變異數	348.43583
眾數	119.0000	全距	86.00000
		內四分位距	21.00000

分位數 (定義 5)

層級	分位數
100% 最大值	180.0
99%	180.0
95%	164.0
90%	154.0
75% Q3	140.0
50% 中位數	129.5
25% Q1	119.0
10%	107.0
5%	101.0
1%	94.0
0% 最小值	94.0

Median (IQR)

常態性檢定

檢定	統計值		p 值	
Shapiro-Wilk	W	0.981536	Pr < W	0.8198
Kolmogorov-Smirnov	D	0.100622	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.046868	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.26081	Pr > A-Sq	>0.2500

常態性檢定 → 符合常態分佈 $P > 0.05$

Kolmogorov-Smirnov (K-S) 檢定：樣本數50個以上

Shapiro-Wilk (S-W) 檢定：樣本數50個以下

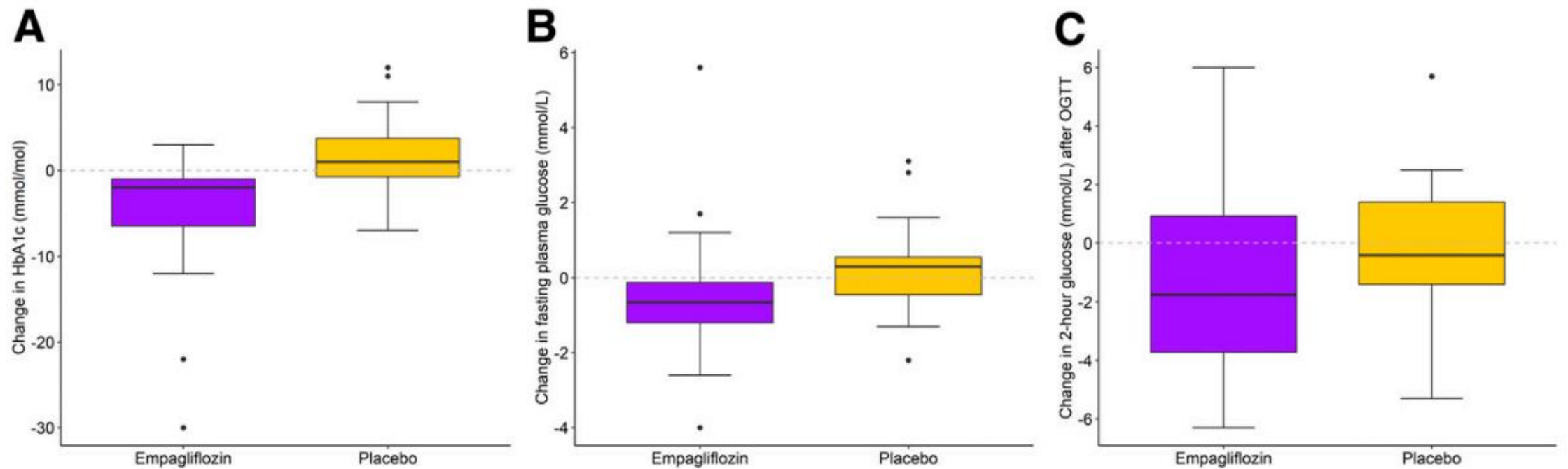
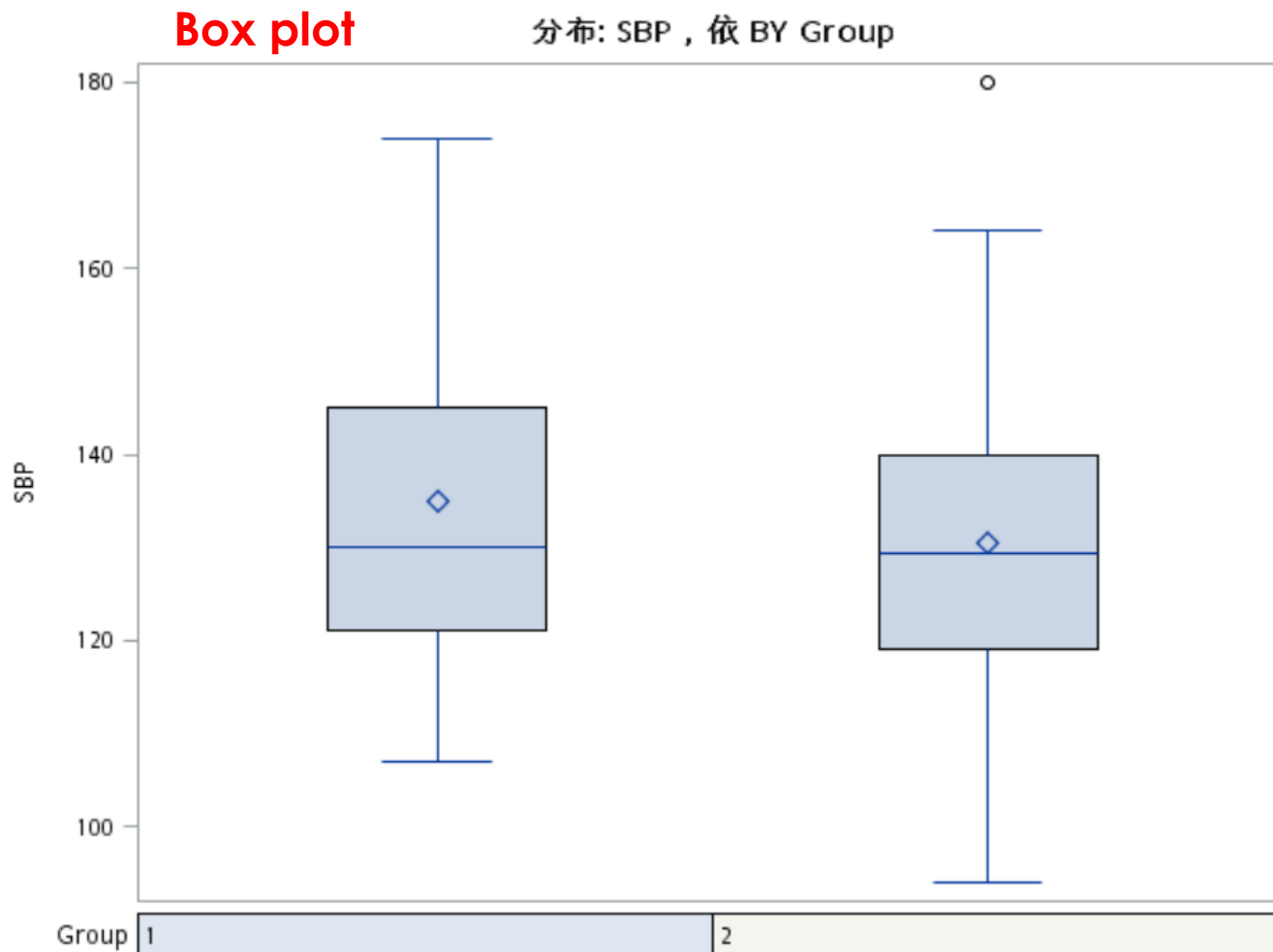


Figure 1—Median (IQR) change from baseline to week 24 in HbA_{1c} ($P = 0.018$) (A), FPG ($P = 0.27$) (B), and 2-h glucose after an OGTT ($P = 1$) (C) in the two intervention groups.

利用SAS進行常態性檢定資料分佈：

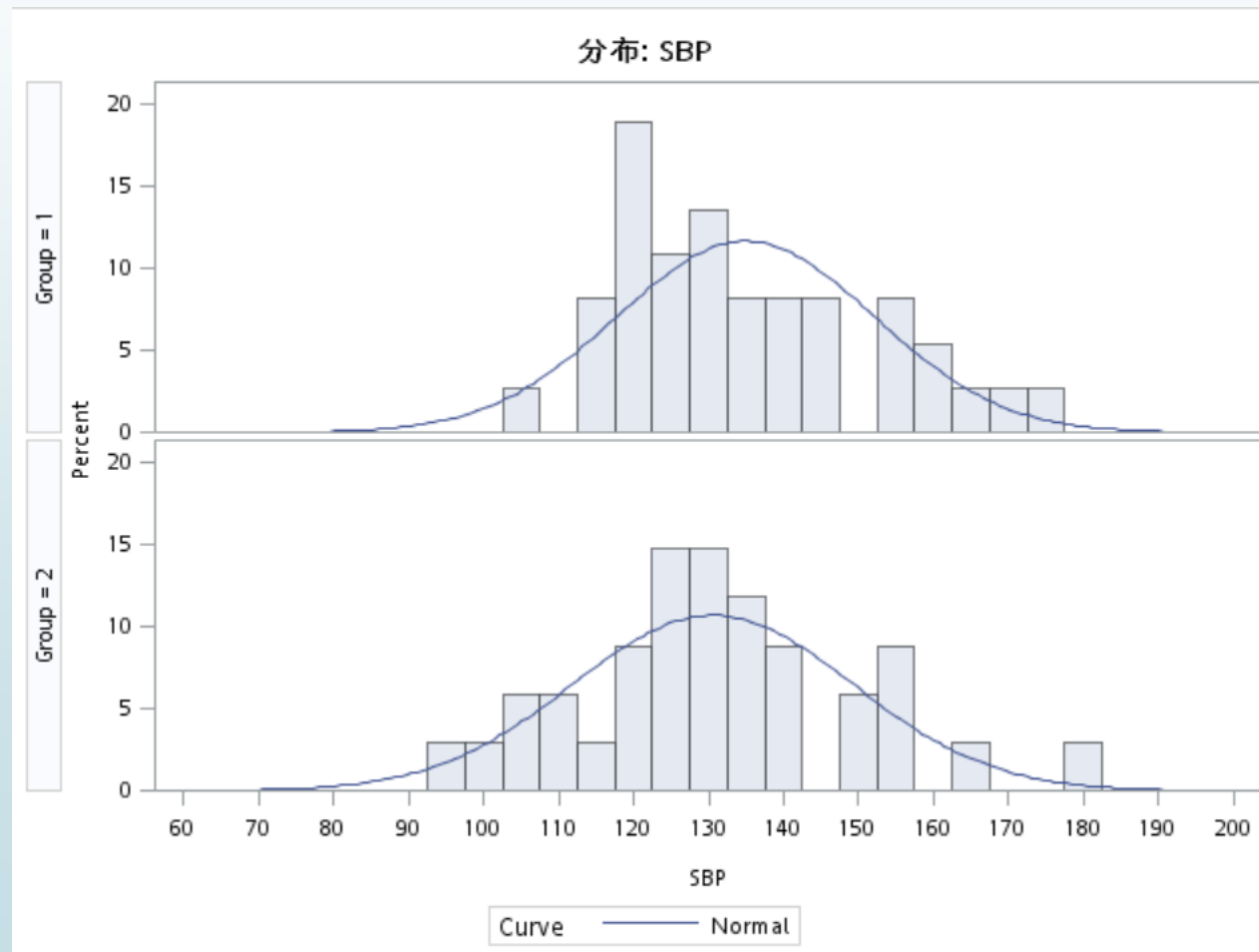
```
proc sort data=lab; by group;  
proc univariate data=lab normal plot;  
var sbp;  
by group; run;
```



利用SAS進行常態性檢定資料分佈：

```
proc univariate data=lab ;  
title histogram;  
histogram sbp / midpoint = 60 to 200 by 5 Normal;  
class group; run;
```

Histogram



Analyzing Data using SAS

長資料格式：data.final

► 利用 SAS 進行 描述性統計分析：連續型變數

```
PROC MEANS DATA=data.final
MAXDEC=2  N  MEAN  STD  MEDIAN Q1 Q3
/*MEDIAN Q1 Q3  MIN MAX  N VAR SKEWNESS KURTOSIS T PROBT CLM /*CLM 樣本信賴區間*/;
VAR  SBP DBP Creatinine  AST ALT  ;
CLASS  visit  ;
/*BY  Group  ; */ RUN;
```

Visit	觀測值數目	變數	標籤	N	平均值	標準差	中位數	下四分位數	上四分位數
1	76	SBP	SBP	76	132.80	17.30	130.00	121.50	140.00
		DBP	DBP	76	80.20	11.26	79.00	73.50	89.00
		Creatinine	Creatinine	75	1.05	0.32	0.99	0.86	1.15
		AST	AST	76	22.71	10.44	20.00	17.00	23.00
		ALT	ALT	76	24.86	20.70	20.00	15.00	25.50
2	73	SBP	SBP	73	132.77	17.84	130.00	121.00	140.00
		DBP	DBP	73	79.85	11.58	79.00	73.00	89.00
		Creatinine	Creatinine	0	.	.	.	.	.
		AST	AST	0	.	.	.	.	.
		ALT	ALT	0	.	.	.	.	.
3	71	SBP	SBP	71	128.68	15.69	127.00	120.00	138.00
		DBP	DBP	71	77.55	10.45	78.00	70.00	86.00
		Creatinine	Creatinine	71	1.03	0.29	1.00	0.87	1.18
		AST	AST	71	22.15	9.33	20.00	17.00	23.00
		ALT	ALT	71	23.28	16.33	20.00	15.00	26.00

Visit	觀測值數目	變數	標籤	N	平均值	標準差	中位數	下四分位數	上四分位數
4	71	SBP	SBP	69	125.58	16.29	125.00	114.00	135.00
		DBP	DBP	69	78.03	14.18	76.00	70.00	82.00
		Creatinine	Creatinine	0	.	.	.	.	.
		AST	AST	0	.	.	.	.	.
		ALT	ALT	0	.	.	.	.	.
5	71	SBP	SBP	64	128.08	14.53	130.00	119.50	138.50
		DBP	DBP	64	77.83	11.09	79.00	68.50	85.50
		Creatinine	Creatinine	0	.	.	.	.	.
		AST	AST	0	.	.	.	.	.
		ALT	ALT	0	.	.	.	.	.
6	71	SBP	SBP	71	128.90	15.30	128.00	119.00	138.00
		DBP	DBP	71	77.32	9.99	79.00	70.00	85.00
		Creatinine	Creatinine	68	1.03	0.28	1.00	0.89	1.12
		AST	AST	68	21.10	7.72	20.00	17.00	23.00
		ALT	ALT	68	21.54	10.77	19.00	15.00	27.00

Analyzing Data using SAS

利用 SAS 進行 推論性統計分析：Student's t test

```
/*Student's t test*/  
data lab; set data.final;  
if visit=1; keep Group SBP gender; run;
```

```
PROC TTEST DATA=lab ;  
CLASS group; VAR SBP ; RUN;
```

TTEST 程序

變數: SBP (SBP)

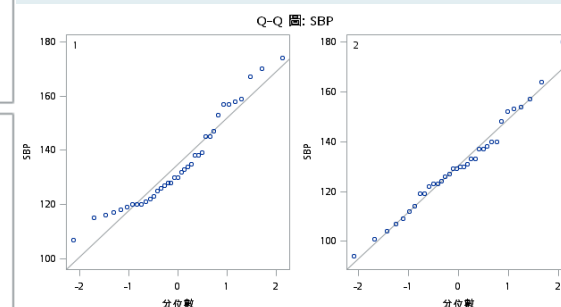
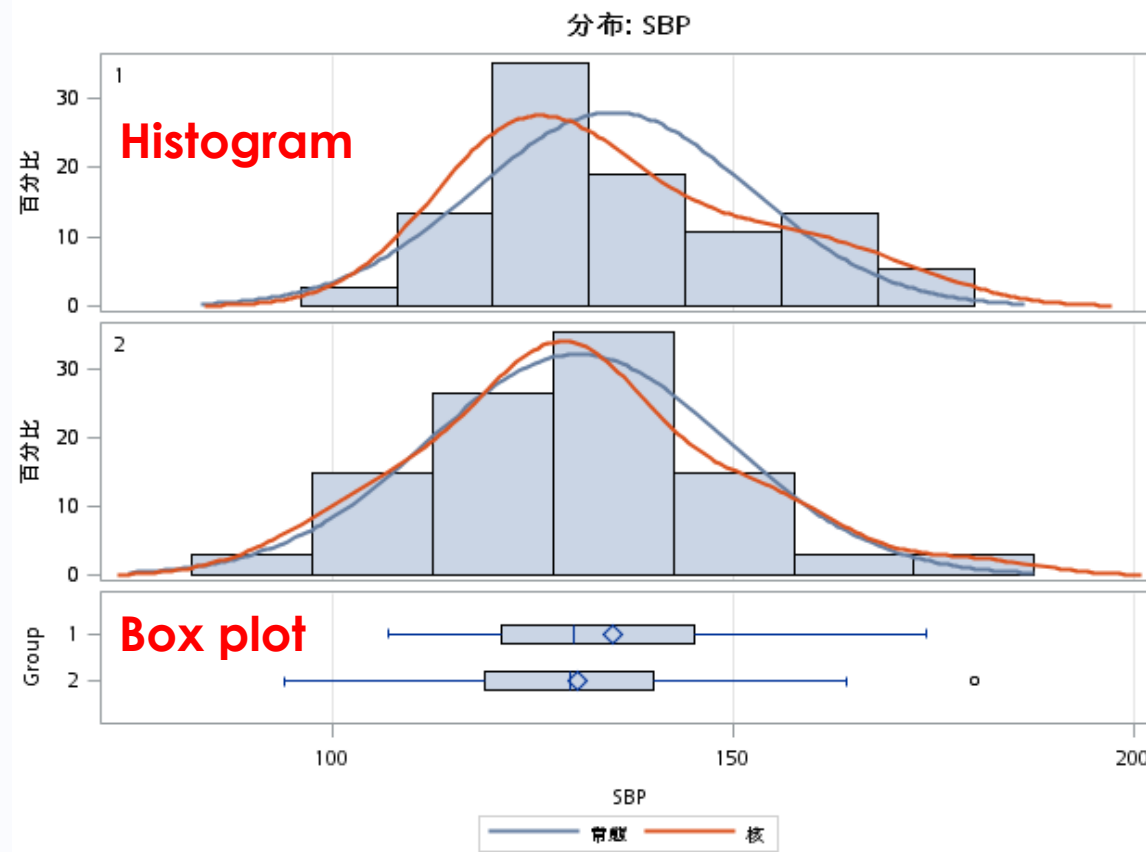
Group	N	平均值	標準差	標準誤差	最小值	最大值
1	37	134.9	17.1188	2.8143	107.0	174.0
2	34	130.6	18.6664	3.2013	94.0000	180.0
Diff (1-2)		4.3871	17.8757	4.2467		

Group	方法	平均值	95% CL 平均值	標準差	95% CL 標準差
1		134.9	129.2 140.7	17.1188	13.9212 22.2366
2		130.6	124.0 137.1	18.6664	15.0559 24.5702
Diff (1-2)	集區	4.3871	-4.0848 12.8591	17.8757	15.3269 21.4491
Diff (1-2)	Satterthwaite	4.3871	-4.1207 12.8949		

方法	變異數	自由度	t 值	Pr > t
集區	均等	69	1.03	0.3052
Satterthwaite	不均等	67.023	1.03	0.3071

變異數相等性

方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	33	36	1.19	0.6104



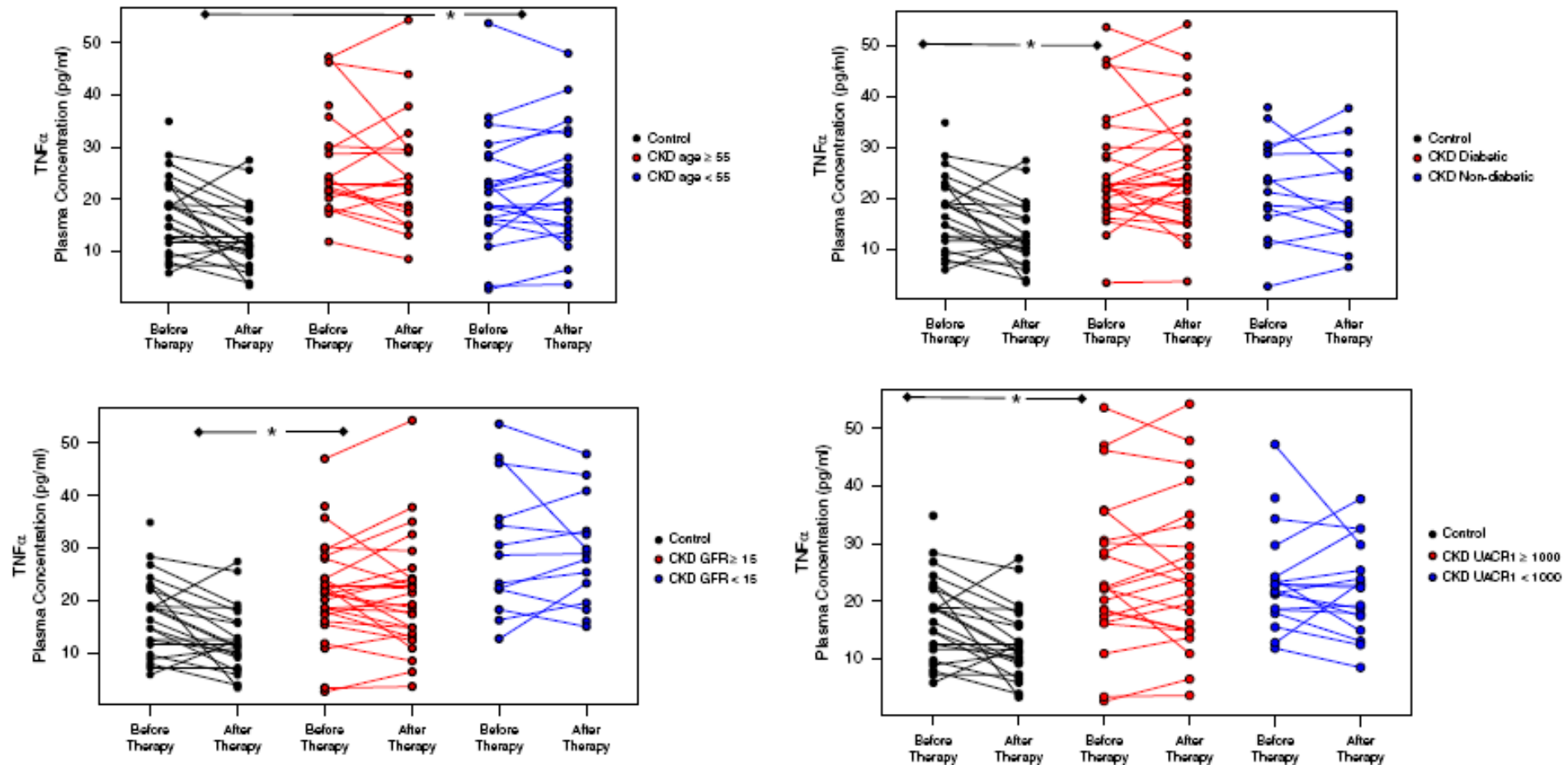


Figure 7. | Line graphs represent plasma TNF- α levels for each study participant before and after DAPT. Non-CKD controls (reference group) in each panel and participants with CKD stratified into two groups on the basis of (A) age < 55 years or ≥ 55 years, (B) presence of diabetes, (C) GFR < 15 ml/min per 1.73 m^2 or ≥ 15 ml/min per 1.73 m^2 , (D) albuminuria (UACR1) < 1000 mg/g or ≥ 1000 mg/g of creatinine. Generalized linear regression model P values with an asterisk represent significant differences when comparing mean change in TNF- α levels in a CKD subgroup with the reference group.]

Analyzing Data using SAS

➡ 利用 SAS 進行 推論性統計分析：Paired t test

```
/*Paired t test*/
```

短資料格式：data.vital_sign_m

```
PROC TTEST DATA= data.vital_sign_m ;
```

```
    PAIRED SBP1*SBP2; /*before : after 不  
同於兩組獨立樣本t檢定，此處以paired後面接著兩個變  
數，分別是前測與後側的數值，以星號連結*/
```

```
RUN;
```

SAS 系統

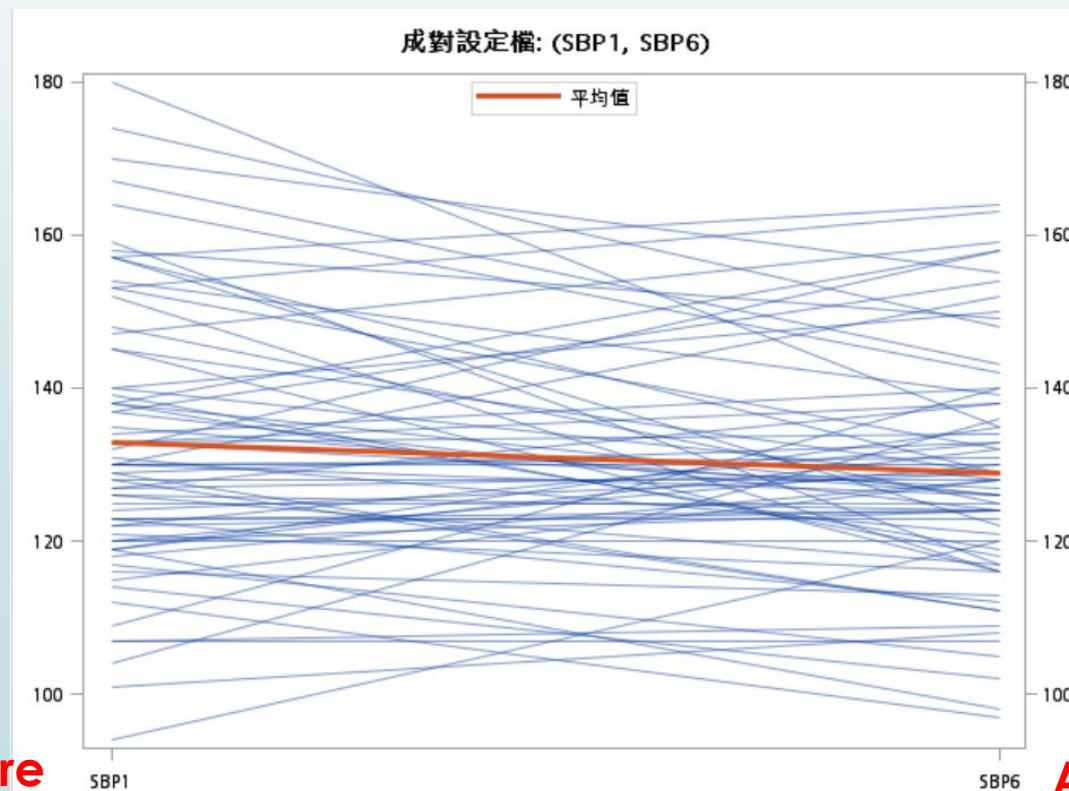
TTEST 程序

差異: SBP1 - SBP6

N	平均值	標準差	標準誤差	最小值	最大值
71	3.9437	16.7945	1.9931	-32.0000	45.0000

平均值	95% CL 平均值	標準差	95% CL 標準差
3.9437	-0.0315	7.9188	14.4145

自由度	t 值	Pr > t
70	1.98	0.0518



Before

After

Analyzing Data using SAS

► 利用 SAS 進行 推論性統計分析：

類別型變數 + Chi-square test

```
/*類別型變數 + Chi-square test*/
```

```
data lab; set data.final;
```

```
if visit=1; keep Group SBP gender; run;
```

```
PROC FREQ data=lab;
```

```
table gender*group / chisq expected exact ; run;
```

SAS 系統

FREQ 程序

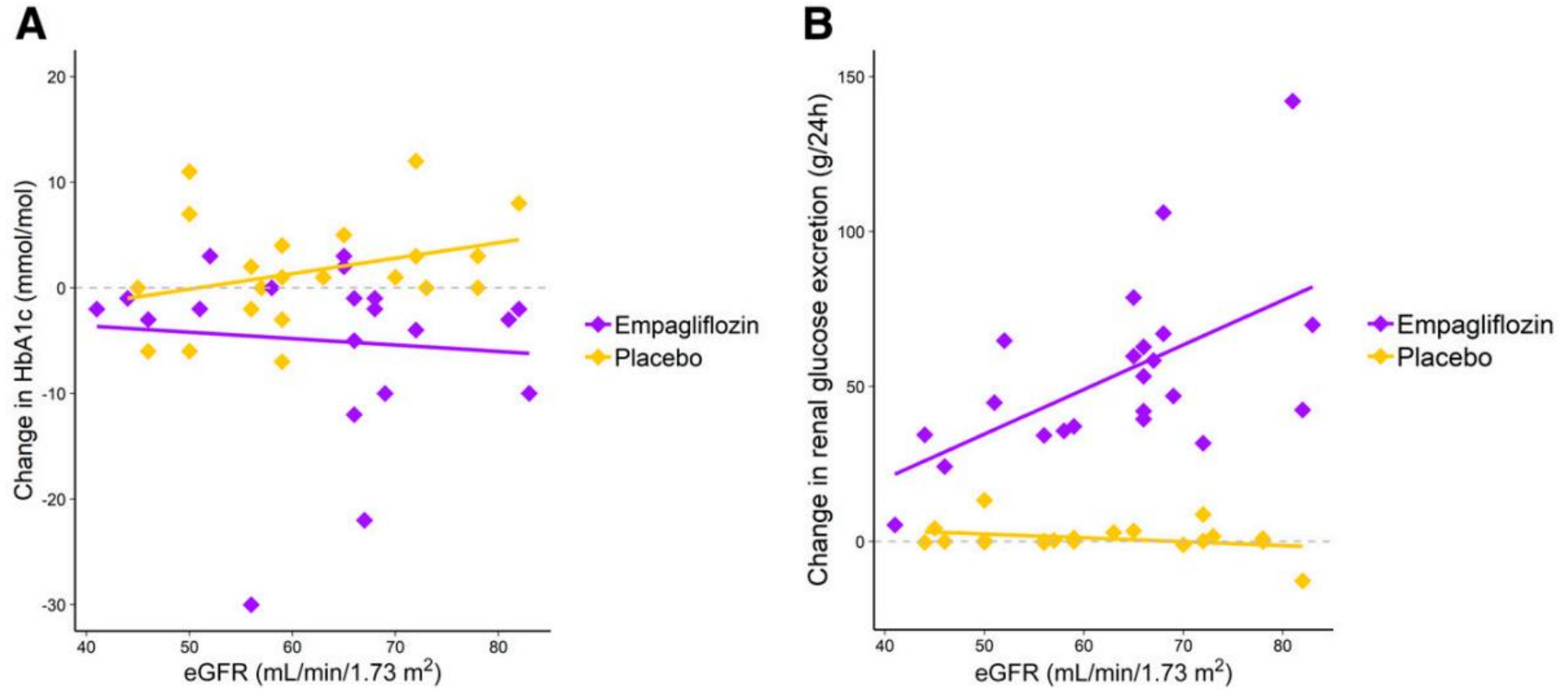
次數 預期 百分比 列百分比 欄百分比	Table of Gender by Group			
	Gender(Gender)	Group(Group)		總計
		1	2	
	0	7	8	15
		7.8169	7.1831	
		9.86	11.27	21.13
		46.67	53.33	
		18.92	23.53	
	1	30	26	56
		29.183	26.817	
		42.25	36.62	78.87
		53.57	46.43	
		81.08	76.47	
	總計	37	34	71
		52.11	47.89	100.00
次數遺漏 = 5				

表格 Group-Gender*s 的統計值

統計值	自由度	值	機率
卡方	1	0.2260	0.6345
概度比卡方	1	0.2259	0.6346
連續性調整卡方	1	0.0340	0.8537
Mantel-Haenszel 卡方	1	0.2228	0.6369
Phi 係數		-0.0564	
列聯係數		0.0563	
Cramer V		-0.0564	

Fisher 精準檢定

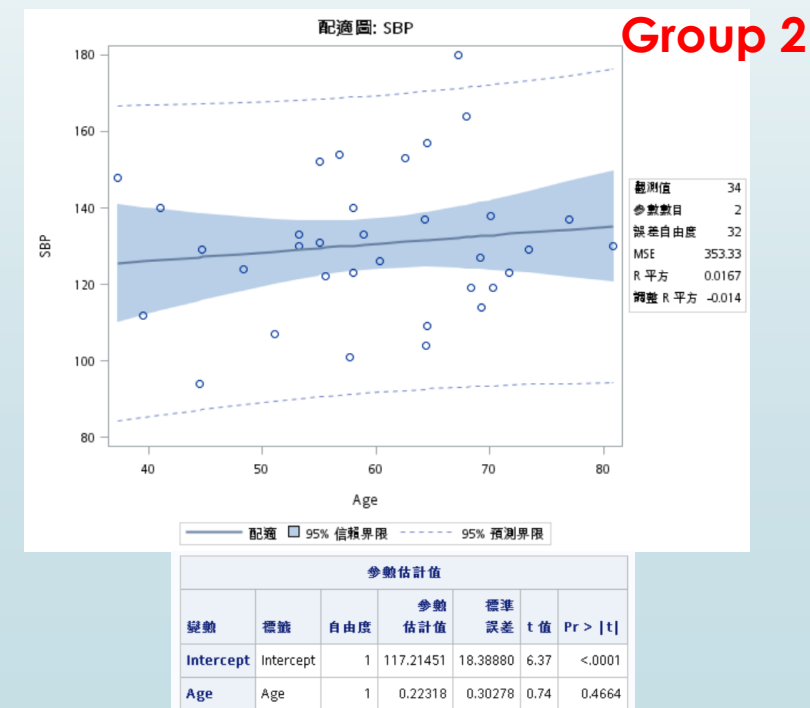
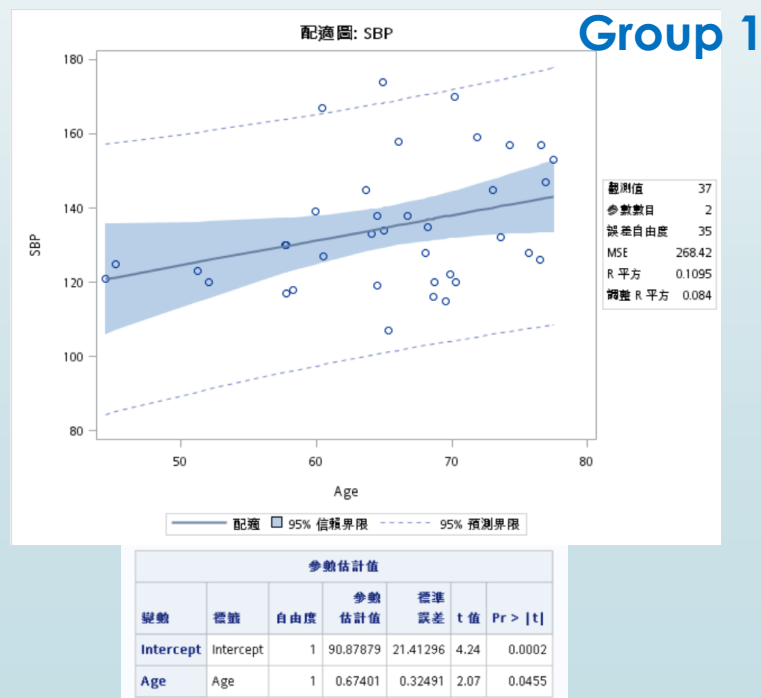
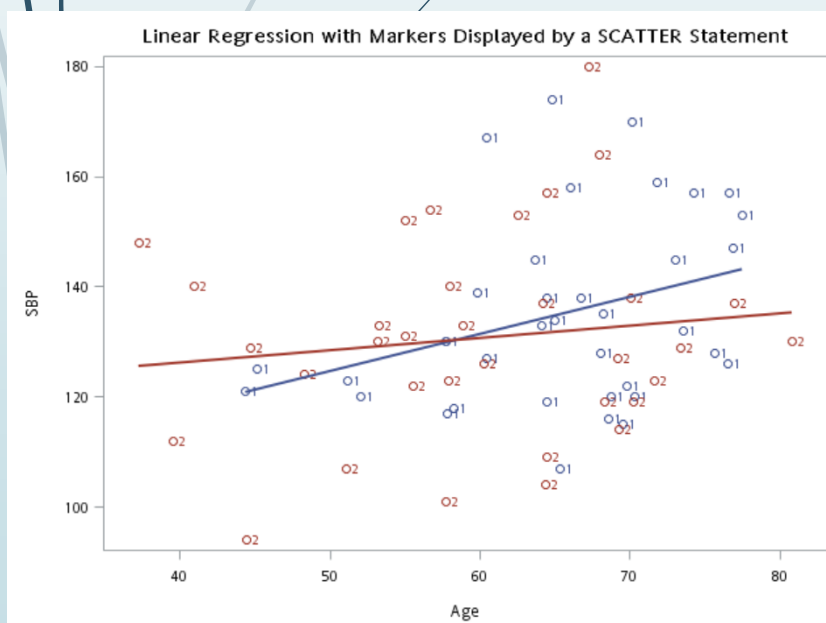
儲存格 (1,1) 次數 (F)	7
左邊 Pr <= F	0.4262
右邊 Pr >= F	0.7782
表格機率 (P)	0.2044
雙邊 Pr <= P	0.7729



Analyzing Data using SAS

► 利用 SAS 進行 推論性統計分析：
Linear regression

```
/*Linear regression*/  
proc sort data=lab; by group /*descending*/; run;  
proc sgplot data=lab noautolegend;  
  title 'Linear Regression with Markers Displayed by a SCATTER Statement';  
  scatter y=SBP x=age / group=group markerchar=group;  
  reg y=SBP x=age / group=group ; run;  
  
proc reg data =lab ; Model sbp = age ; by group; run;
```



Analyzing Data using SAS

- 利用 SAS 進行 推論性統計分析：重複測量
Generalized estimating equations (GEE)

長資料格式：data.final

```
proc sort data= data.final ; by screening_no visit; run;

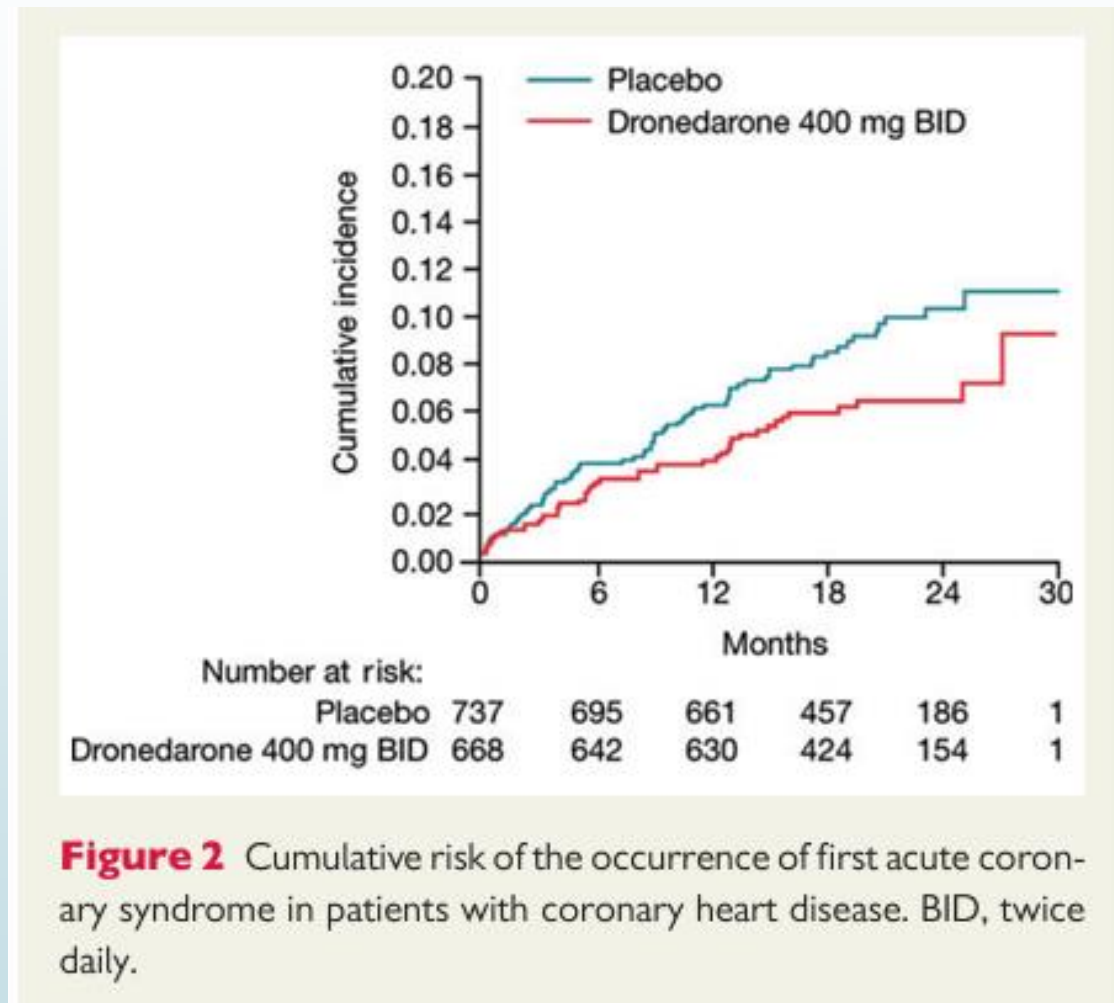
/*GEE: Repeated measure*/
proc genmod data= data.final ;
  class screening_no group (ref=first) gender (ref=first)
  visit (ref=first) / param=ref;
  model SBP = group age gender / dist = normal link = log ;
/*link=logit dist=binomial; dist = normal link = log ; dist =
poisson link = log offset = o COVB*/
repeated subject=screening_no (visit) ; run;
```

已收斂演算法。

GEE 模型資訊	
相關結構	獨立
主題效果	Screening_no(Visit) (433 層級)
群集數	433
含有遺漏值的群集	16
相關矩陣維度	1
最大群集大小	1
最小群集大小	0

GEE 參數估計值的分析

經驗標準誤差估計值						
參數		估計值	標準誤差	95% 信賴界限		Z Pr > Z
Intercept		4.6678	0.0425	4.5845	4.7512	109.79 <.0001
Group	2	0.0047	0.0124	-0.0196	0.0290	0.38 0.7038
Age		0.0028	0.0006	0.0017	0.0040	4.79 <.0001
Gender	1	0.0203	0.0154	-0.0099	0.0504	1.32 0.1881



Analyzing Data using SAS

► 利用 SAS 進行 推論性統計：存活分析

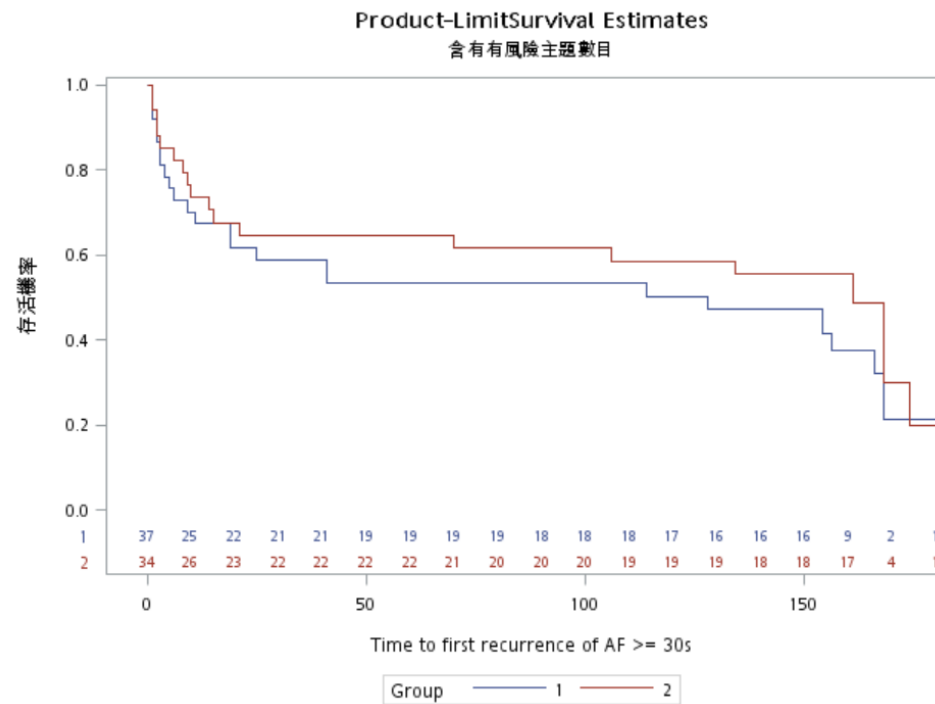
Kaplan-Meier Survival Curves

```
/*KM plots*/  
PROC LIFETEST DATA= data.itt6 METHOD=KM NOTABLE  
PLOTS=SURVIVAL (ATRISK=0 TO 180 BY 10 NOCENSOR /*CB 95%CI*/);  
/*KM/PL意思相同; (SURVIVAL) / (S) 意思相同*/  
TIME Time_first_rec* First_AF_recurrence (0);  
STRATA group ; /*diagage (65,75); age切65↓, 65-75, 75↑; */  
/* by age65; */  
RUN;
```

分層相等性的檢定

檢定	卡方	自由度	Pr > 卡方
對數排名	0.5519	1	0.4576
Wilcoxon	0.8589	1	0.3540
-2Log(LR)	0.2575	1	0.6119

Event-free survival rate



Europace (2014) **16**, 174–181
doi:10.1093/europace/eut293

Table 2 Comparison of the effect of dronedarone among patients with and without coronary heart disease

Outcome	CHD	Placebo, n/N (%)	Dronedarone, n/N (%)	HR for dronedarone (95% CI)	P value ^a
First cardiovascular hospitalization or death from any cause	Yes	350/737 (47.49)	252/668 (37.72)	0.733 (0.62–0.86)	0.535
	No	567/1590 (35.66)	482/1663 (29.52)	0.782 (0.69–0.88)	
Cardiovascular death	Yes	47/737 (6.38)	26/668 (3.89)	0.602 (0.37–0.97)	0.350
	No	47/1590 (2.96)	39/1663 (2.39)	0.814 (0.53–1.24)	
First ACS	Yes	67/737 (9.09)	42/668 (6.29)	0.671 (0.46–0.99)	0.429
	No	29/1590 (1.82)	26/1633 (1.59)	0.876 (0.52–1.49)	
First stroke, ACS or cardiovascular death	Yes	116/737 (15.74)	67/668 (10.03)	0.615 (0.46–0.83)	0.272
	No	101/1590 (6.35)	81/1633 (4.96)	0.778 (0.58–1.04)	

ACS, acute coronary syndrome; CHD, coronary heart disease; CI, confidence interval.

^aP value of interaction between CHD status and treatment based on Cox regression model.

Analyzing Data using SAS

- 利用 SAS 進行 推論性統計：存活分析
Cox proportional hazard model

```
/*Cox regression*/  
PROC PHREG DATA=data.itt6 ;  
class gender (ref=first ) group (ref="2") htn (ref=first ) dm(ref=first )  
hyperlipidemia (ref=first ) cad (ref=first ) cva (ref=first) hf (ref=first) ;  
/*by age65;*/  
MODEL Time_first_rec* First_AF_recurrence (0) =group age gender htn/ RISKLIMITS;  
RUN;
```

最大概度估計值的分析										
參數		自由度	參數 估計值	標準 誤差	卡方	Pr > ChiSq	危險 比	95% 危險比信賴界限		標籤
Group	1	1	0.26032	0.31998	0.6619	0.4159	1.297	0.693	2.429	Group 1
Age		1	-0.00476	0.01606	0.0878	0.7670	0.995	0.964	1.027	Age
Gender	1	1	-0.13546	0.35660	0.1443	0.7040	0.873	0.434	1.757	Gender 1
HTN	1	1	-0.01737	0.30505	0.0032	0.9546	0.983	0.541	1.787	HTN 1

Descriptive Statistics



Descriptive statistics is achieved with the help of tables, graphs, etc.

Inferential Statistics



Inferential statistics is achieved by probability.

Descriptive Statistics



The goal of descriptive statistics is to describe and summarize the data.

Inferential Statistics



The goal of inferential statistics is to test and make predictions.

Descriptive Statistics



The algorithm of descriptive statistics is mean, mode, and median.

Inferential Statistics



The algorithm of inferential statistics is ANOVA and regression.



SAS統計軟體基礎操作訓練 (2)



Thank you for listening