

健保資料庫分析應用

National Health Insurance Research Database

講師：王鵬傑

台灣析數資訊 專案顧問



Business Intelligence Analysis

The Best Solution
Provider



台灣析數



Agenda

- Prevalence of attention-deficit/hyperactivity disorder in patients with pediatric allergic disorders 簡介
 - ✓ 摘要說明
 - ✓ 研究方法
 - ✓ 統計報表
- 資料庫處理 - 以SPSS為例
 - ✓ Input data
 - ✓ Merge Dataset- Add Cases & Variables
 - ✓ Date / Time
 - ✓ Find Case
- Paper報表操作分析

Prevalence of attention-deficit/hyperactivity disorder in patients with pediatric allergic disorders簡介



Journal of Microbiology, Immunology and Infection (2012) 45, 237–242



Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.e-jmji.com



ORIGINAL ARTICLE

Prevalence of attention-deficit/hyperactivity disorder in patients with pediatric allergic disorders: A nationwide, population-based study

Ching-Shan Shyu^{a,c}, Heng-Kuei Lin^a, Ching-Heng Lin^{b,**}, Lin-Shien Fu^{a,d,*}

^a Division of Immunology and Nephrology, Department of Pediatrics, Taichung Veterans General Hospital, Taichung City, Taiwan

^b Department of Medical Research, Taichung Veterans General Hospital, Taichung City, Taiwan

^c Department of Pediatrics, Chutung Veterans Hospital, Jhudong Township, Hsinchu County, Taiwan

^d Institute of Technology, National Chi-Nan University, Nan-to County, Taiwan

摘要說明 1/3

背景：

- 過敏性疾病的特徵性症狀主要包括支氣管哮喘（BA）、過敏性鼻炎（AR）和過敏性皮膚炎（AD），所有這些都可能擾亂睡眠，導致白天注意力不集中、煩躁、過動，這也是人們關注注意力障礙（ADHD）的要素。
- 這項研究的目的是研究兒童族群於過敏性疾病與注意力障礙之盛行情況。



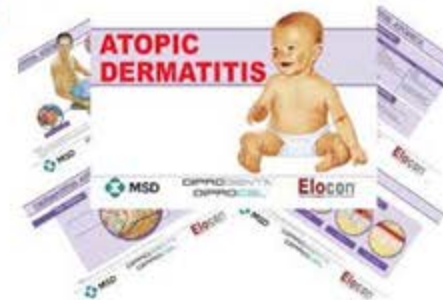
Copyright© AsiaAnalytics Taiwan 2013. All Rights Reserved.

BRONCHIAL ASTHMA EMERGING THERAPEUTIC STRATEGIES

Edited by Elizabeth Sapoz



Allergic Rhinitis



摘要說明 2/3

方法：

- 數據資料共有226550為18歲以下的兒童患者，資料來源為台灣的全民健康保險研究資料庫。
- 計算各種過敏性疾病的人口統計變量，包括過動症情況數據資料。
- 使用**Logistic迴歸**分析ADHD的危險因子。



摘要說明 3/3

結果：

- 過敏性疾​​病和ADHD在18歲以下孩童的患病率分別為21.5%和0.6%。
- 有過敏性疾病的兒童患者就盛行情形與勝算比而言，在年齡、性別、城市、都市化的程度均有明顯的比率會發展為ADHD。
- 除AD外，BA和AR，被確定為過動症的危險因素。
共病過敏性疾​​病，除BA+AD外，對於AR+AD，AR+BA和AR+BA+AD，也被確認是增加過動症的風險。



研究方法 1/2

2005年承保抽樣歸人檔：

- 由中央健康保險署所提供的2005年承保資料檔以「身分證字號加上生日加上性別」歸人，選取在保者22,717,053人之資料為抽樣母群體。

(定義「在保者」為「2006年以前出生，2005年1月1日至2006年1月1日任何一日曾在保者」，並剔除年齡非0-120歲者)

- 隨機抽出的100萬人樣本，利用身分證字號(已加密)每4萬人為一組共25組，與健保資料庫串聯，擷取1996-2006年該100萬人在全民健保研究資料庫中所有就醫資料。

(所串接的就醫資料包含：門診處方及治療明細檔(CD)、門診處方醫令明細檔(OO)、住院醫療費用清單明細檔(DD)、住院醫療費用醫令明細檔(DO)、特約藥局處方及調劑明細檔(GD)、特約藥局處方調劑醫令明細檔(GO)，以及原始承保資料)

研究方法 2/2

研究對象：

- 篩選出年齡為0~17歲族群的觀察值資料。
- 識別出過敏性疾病的患者。

(BA codes 493.xx, AR codes 477.xx, AD codes 691.xx)

- 識別出ADHD的患者。
- (ADHD codes 314.xx)



統計報表 1/3

Table 1 Demographic characteristics of patients with allergic disorders and general population

variable	Total (n = 226550) n (%)	General population (n = 178093) n (%)	with Allergic disease (n = 48457) n (%)	p-value for χ^2 test
Age(years old)				
<6	74973 (33.1)	53277 (29.9)	21696 (44.8)	<0.001
6-11	81529 (36.0)	64322 (36.1)	17207 (35.5)	
12-17	70048 (30.9)	60494 (34.0)	9554 (19.7)	
Gender				
Female	108314 (47.8)	87750 (49.3)	20564 (42.4)	<0.001
Male	118236 (52.2)	90343 (50.7)	27893 (57.6)	
Area				
North	111616 (49.3)	86101 (48.4)	25515 (52.7)	<0.001
Middle	44600 (19.7)	35316 (19.8)	9284 (19.2)	
South	64907 (28.7)	52081 (29.2)	12826 (26.5)	
East	5427 (2.4)	4595 (2.6)	832 (1.7)	
Urbanization				
Urban	131663 (58.1)	102507 (57.6)	29156 (60.2)	<0.001
Suburban	32372 (14.3)	25613 (14.4)	6759 (14.0)	
Rural	62515 (27.6)	49973 (28.0)	12542 (25.9)	
ADHD				
No	225243 (99.4)	177214 (99.5)	48029 (99.1)	<0.001
Yes	1307 (0.6)	879 (0.5)	428 (0.9)	

統計報表 2/3

Table 2 Prevalence of ADHD in children with allergic disorders and the general population

variable	General population (n = 178093)	with Allergic disorders (n = 48457)	Odds ratio	95% CI	p-value for χ^2 test
	n (%)	n (%)			
Age(years old)					
<6	219 (0.4)	123 (0.6)	1.38	1.11 ~ 1.72	0.005
6-11	519 (0.8)	257 (1.5)	1.86	1.60 ~ 2.17	<0.001
12-17	141 (0.2)	48 (0.5)	2.16	1.56 ~ 3.00	<0.001
Gender					
Female	182 (0.2)	69 (0.4)	1.62	1.23 ~ 2.14	<0.001
Male	697 (0.8)	359 (1.3)	1.68	1.48 ~ 1.91	<0.001
Area					
North	532 (0.6)	278 (1.1)	1.77	1.53 ~ 2.05	<0.001
Middle	105 (0.3)	50 (0.5)	1.82	1.30 ~ 2.54	<0.001
South	220 (0.4)	91 (0.7)	1.68	1.32 ~ 2.15	<0.001
East	22 (0.5)	9 (1.1)	2.27	1.04 ~ 4.95	0.044
Urbanization					
Urban	619 (0.6)	309 (1.1)	1.76	1.54 ~ 2.02	<0.001
Suburban	87 (0.3)	42 (0.6)	1.83	1.27 ~ 2.65	0.002
Rural	173 (0.4)	77 (0.6)	1.78	1.36 ~ 2.33	<0.001

統計報表 2/3

Table 3 Logistic regression model of factors with prevalence of ADHD in pediatric allergic patients

variable	Odds ratio	95% CI		p-value
		Lower	Upper	
Age(years old)				
<6	1.00	—	—	
6-11	2.10	1.85	2.39	<0.001
12-17	0.62	0.51	0.74	<0.001
Gender				
Female	1.00	—	—	
Male	3.76	3.28	4.32	<0.001
Area				
East	1.00	—	—	
North	1.12	0.78	1.61	0.528
Middle	0.61	0.41	0.89	0.011
South	0.81	0.56	1.17	0.265
Urbanization				
Rural	1.00	—	—	
Urban	1.58	1.37	1.82	<0.001
Suburban	0.99	0.80	1.23	0.930
Allergic condition				
no allergic disorders	1.00	—	—	
AR or BA or AD	1.56	1.38	1.75	<0.001
AR	1.71	1.48	1.98	<0.001
AD	0.73	0.48	1.09	0.124
BA	1.38	1.02	1.86	0.039
AR and AD	2.14	1.37	3.35	0.001
AR and BA	1.63	1.28	2.06	<0.001
AD and BA	0.72	0.18	2.89	0.639
AR and BA and AD	2.02	1.19	3.46	0.010

R-Square = 0.0666.

資料庫處理 - 以SPSS為例



處理流程



資料讀取



資料處理



資料分析

若是有一套方便操作的
軟體會更得心應手

SPSS

Input -Excel File

✓ Reading Excel Files

有A到 N個變數扣除第一列為變數名稱，每個變數有100個觀察值。

	A	B	C	D	E	F	G	H	I	J
1	FEE_YM	APPL_TY	HOSP_ID	APPL_D	CASE_TY	SEQ_NO	FUNC_DA	ID	BIRTH_ID	ACODE
2	201209	1	EMNJMSI	20120916	JP	361931.00	20120901	19401124	JP	
3	201202	1	CMSQSYI	20120210	CM	111106.00	20120201	19120722	CM	
4	201210	1	WJGHXD	20121122	WJ	860279.00	20121001	19970318	WJ	
5	201207	1	PGUYDHI	20120826	PG	609454.00	20120701	19681113	PG	
6	201204	1	JDHPJLDI	20120503	JD	358629.00	20120401	19400711	JD	
7	201202	1	CAVGOP	20120210	CA	107804.00	20120201	19120308	CA	
8	201210	1	WXIXUU	20121122	WX	856977.00	20121001	19961102	WX	
9	201207	1	PUWNZY	20120825	PU	606152.00	20120701	19680630	PU	
10	201204	1	JRJEFCYV	20120503	JR	355327.00	20120401	19400226	JR	
11	201202	1	COXVLG	20120210	CO	104502.00	20120201	19111024	CO	
12	201210	1	WLKMQL	20121122	WL	853675.00	20121001	19960618	WL	
13	201207	1	PIYDWPU	20120825	PI	602850.00	20120701	19680214	PI	
14	201204	1	JFLUCTSI	20120503	JF	352025.00	20120401	19391012	JFLUCTSI	2504
15	201202	1	CDZLHXF	20120210	CD	101200.00	20120201	19110609	CDZLHXF	2501
16	201210	1	WANBNC	20121121	WA	850373.00	20121001	19960203	WANBNC	2508
17	201207	1	PXASSGC	20120825	PX	599548.00	20120701	19671001	PXASSGC	2506
18	201204	1	JUOJYKN	20120502	JU	348723.00	20120401	19390529	JUOJYKN	2504



設定讀取範圍

Input -Text File (分隔) 1/3

CD_Tab.dat - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

FEE_YM	APPL_TYPE	HOSP_ID	APPL_DATE
201209	1	EMNJMSLGTQJVEXXKMGVAKBROZIRXU	
201202	1	CMSQSYHYFVRYDPFMHVUFVCQRLKLW	
201210	1	WJGHXDGBWPDVTEUFBFXKGYGXIPLTL	
201207	1	PGUYDHEFNJPSJTKXWOAQCLDYUNCZ	
201204	1	JDHPJLDIEDBPZIZQQXCZVZFKPYPLN	
201202	1	CAVGOPCMVXNNPXPILHFAJJVQGCQTB	
201210	1	WXIXUUAAPRAKFMFBFQIFSMZXXHSCP	
201207	1	PUWNZYZTDKMHWBUTAZLLCQEDNLTkd	
201204	1	JRJEFCYWUEYFMRJMUJQOLTJJJEQVTS	
201202	1	COXVLGWZLYKCCGZEPSRVVXQVUXBG	
201210	1	WLKMLVDCSWZSVPXJBUEASWYXKU	
201207	1	PIYDWPUGSMIWIKEQELXGOEXDCDATI	
201204	1	JFLUCTSKJGUUYZUIYUALXICJTHBBW	
201202	1	CDZLHXRNAZGRPOJBTDCQHLHOKLDK	
201210	1	WANBNCQORTSOFDZTNMFQPLWBOFSZ	
201207	1	PXASSGOUINEMVSQMLIBZSQDRUGBN	
201204	1	JUOJYKNXZHQJLHEECGLGJWVJIZIJB	
201202	1	CRBAEOMBQBDGBWYXPOLSZAQZDJSP	
201210	1	WOPRJTEHVPDRLJPRYQCDEWQHLAD	
201207	1	PLCIPXJJOBBHAYMIUWLGJdGMNJR	
201204	1	IIQZUBILOINYYPOAGRxBVKOJXQOSF	
201202	1	CFDPAGGOFPCZVOPETBAAGENTQOUQAU	

文字匯入精靈 - 步驟 6 之 1

歡迎進入文字匯入精靈！
此精靈將協助您讀取文字檔的資料，並指定變數的相關資訊。

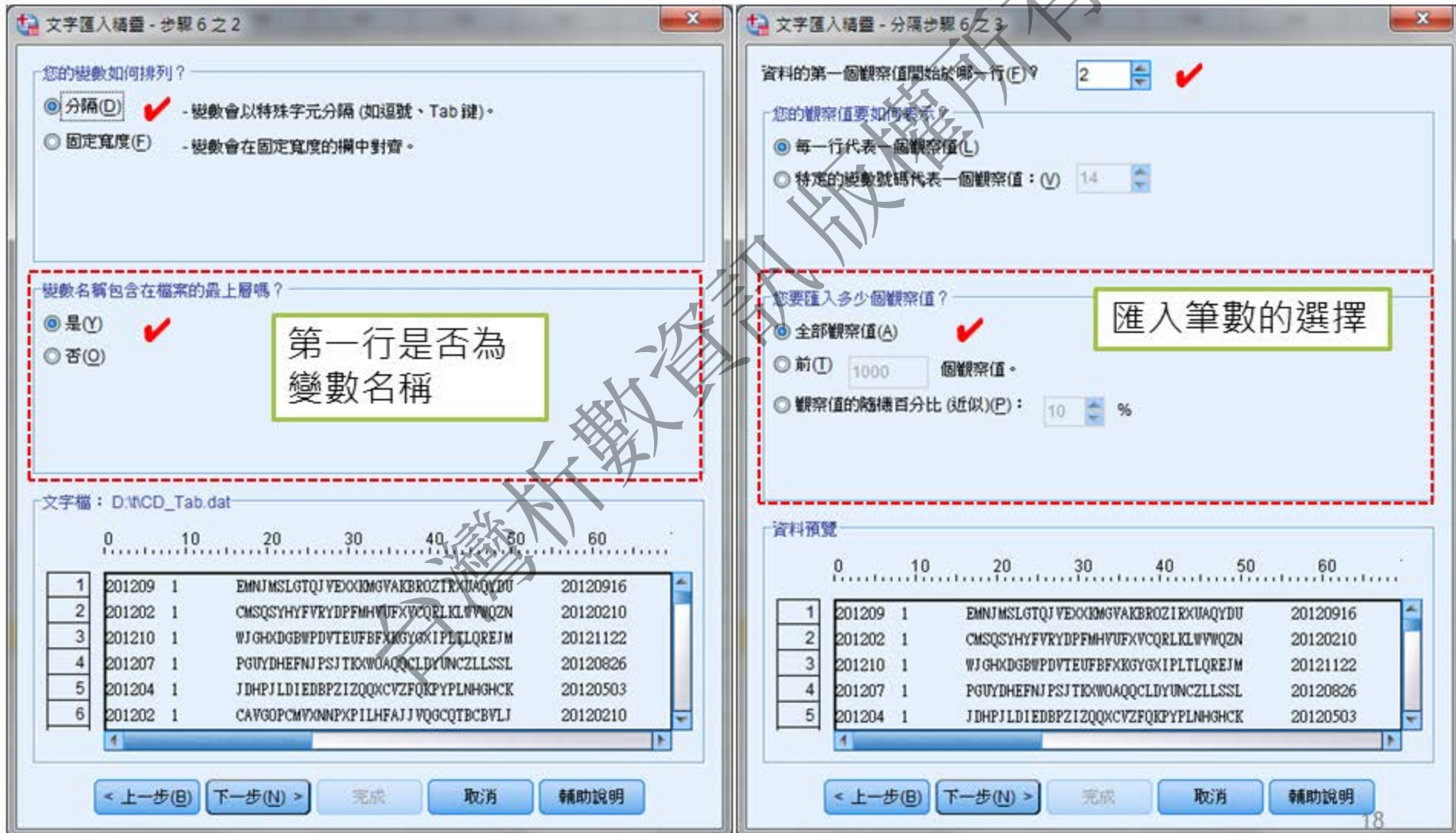
您的文字檔符合預先定義的格式嗎？

☐ 是(Y) ☒ 否(O) ✓ 瀏覽(B)...

文字檔: D:\CD_Tab.dat

	0	10	20	30	40	50	60
1	FEE_YM	APPL_TYPE	HOSP_ID	APPL_DATE	CASE_TYPE	SEQ_N	
2	201209	1	EMNJMSLGTQJVEXXKMGVAKBROZIRXUAQYDU			20120916	
3	201202	1	CMSQSYHYFVRYDPFMHVUFVCQRLKLWVWQZN			20120210	
4	201210	1	WJGHXDGBWPDVTEUFBFXKGYGXIPLTLQREJM			20121122	
5	201207	1	PGUYDHEFNJPSJTKXWOAQCLDYUNCZLLSSL			20120826	
6	201204	1	JDHPJLDIEDBPZIZQQXCZVZFKPYPLNHGCK			20120503	
7	201202	1	CAVGOPCMVXNNPXPILHFAJJVQGCQTBCLVJ			20120210	
8	201210	1	WXIXUUAAPRAKFMFBFQIFSMZXXHSCPXVJVH			20121122	
9	201207	1	PUWNZYZTDKMHWBUTAZLLCQEDNLTkdSQKEG			20120825	
10	201204	1	JRJEFCYWUEYFMRJMUJQOLTJJJEQVTSOLLOF			20120503	
11	201202	1	COXVLGWZLYKCCGZEPSRVVXQVUXBGJGAXE			20120210	

< 上一步(B) 下一步(N) > 完成 取消 輔助說明



Input -Text File (分隔) 3/3

文字匯入精靈 - 分隔步驟 6 之 4

變數間要顯示何種分隔符號？

☒ Tab(T) ☐ 空格(S)
☐ 逗點(C) ☐ 分號(E)
☐ 其他(H):

文字修飾詞為何？

☒ 無(O)
☐ 單一報價(Q)
☐ 雙重報價(D)
☐ 其他(H):

分隔符號的選擇

資料預覽

FEE_YM	APPL_TYPE	HOSP_ID	APPL_DATE	CASE_TYPE	SEQ_NO	F
201209	1	EMNJMSL...	20120916	JP	361931	2
201202	1	CMSQSYH...	20120210	CM	111106	2
201210	1	WJGHXDG...	20121122	WJ	860279	2
201207	1	PGUYDHE...	20120826	PG	609454	2
201204	1	JDHPJLDI...	20120503	ID	358629	2
201202	1	CAVGOPC...	20120210	CA	107804	2
201210	1	WXIXUUP...	20121122	WX	856977	2
201207	1	PUWNZYZ...	20120825	PU	606152	2
201204	1	JRJEFCYW...	20120503	JR	355327	2
201202	1	COXVLGW...	20120210	CO	104502	2
201210	1	WLKMLV...	20121122	WL	853675	2
201207	1	PIYDWPLJ...	20120825	PL	602850	2

< 上一步(B) 下一步(N) > 完成 取消 輔助說明

文字匯入精靈 - 步驟 6 之 5

資料預覽中所選的變數規格

為各變數命名及設定資料格式

Data format is determined by the longest value present in the first 200 records. If subsequent records have longer values, they will be truncated.

變數名稱(V):

資料格式(D):

資料預覽

FEE_YM	APPL_TYPE	HOSP_ID	APPL_DATE	CASE_TYPE	SEQ_NO	F
201209	1	EMNJMSL...	20120916	JP	361931	2
201202	1	CMSQSYH...	20120210	CM	111106	2
201210	1	WJGHXDG...	20121122	WJ	860279	2
201207	1	PGUYDHE...	20120826	PG	609454	2
201204	1	JDHPJLDI...	20120503	ID	358629	2

< 上一步(B) 下一步(N) > 完成 取消 輔助說明

Input -Text File (固定寬度) 1/3



Input -Text File (固定寬度) 2/3

文字匯入精靈 - 步驟 6 之 2

您的變數如何排列？

☐ 分隔(D) - 變數會以特殊字元分隔 (如逗號、Tab 鍵)。

☒ 固定寬度(F) - 變數會在固定寬度的欄中對齊。

變數名稱包含在檔案的最上層嗎？

☐ 是(Y)

☒ 否(N) ✓

第一行是否為變數名稱

文字檔： D:\CD_test.dat

0	10	20	30	40	50	60
1	2012091EMNJMSLTQJVEKXKMGVAKBROZIRKUAQYDU20120916JP361931201209011940					
2	2012021CMSQSYHYFVRYDPFMHVUFVCQRLKLWVWQZM20120210CM111106201202011912					
3	2012101WJGHKDBGWPDVTEUFBFKGYGXIPLTLOREJ20121122WJ860279201210011997					
4	2012071PGUYDHEFNJPSJTKKWOAQCLDYUNCZLLSSL20120826P6609454201207011968					
5	2012041JDHPJLDIEDBPZIZQXCVCZFQKPYPLNHGHC20120503JD358629201204011940					
6	2012021CAVGOPCMVXNNPXPILHFAJJVQGCQTBCBVLJ20120210CA107804201202011912					

< 上一步(B)

下一步(N) >

完成

取消

輔助說明

文字匯入精靈 - 固定寬度步驟 6 之 3

資料的第一個觀察值開始於哪一行(F)? 1 ✓

多少行代表一個觀察值(L)? 1 ✓

您要匯入多少個觀察值？

☒ 全部觀察值(A)

☐ 前(L) 1000 個觀察值。

☐ 觀察值的百分比：(P) 10 %

匯入筆數的選擇

資料預覽

0	10	20	30	40	50	60
1	2012091EMNJMSLTQJVEKXKMGVAKBROZIRKUAQYDU20120916JP361931201209011940					
2	2012021CMSQSYHYFVRYDPFMHVUFVCQRLKLWVWQZM20120210CM111106201202011912					
3	2012101WJGHKDBGWPDVTEUFBFKGYGXIPLTLOREJ20121122WJ860279201210011997					
4	2012071PGUYDHEFNJPSJTKKWOAQCLDYUNCZLLSSL20120826P6609454201207011968					
5	2012041JDHPJLDIEDBPZIZQXCVCZFQKPYPLNHGHC20120503JD358629201204011940					
6	2012021CAVGOPCMVXNNPXPILHFAJJVQGCQTBCBVLJ20120210CA107804201202011912					
7	2012101WXIXUAPMRAKFMFBFQIFSMZXXHSCFPXJVH20121122WX856977201210011996					
8	2012071PUWNYZTDRMHWBUTAZLLCQEDNLTKDSQXEG20120825PU606152201207011968					
9	2012041JRJEFYWUEYFMRJMUJQQLTJJ EQVTSOLLOF20120503JR355327201204011940					
10	2012021COKVLGWZLYKCCGEZPSRVVXQVUXBGJGAXE20120210CO104502201202011911					
11	2012101WLMQLVDCSWZSVPXJBUEASWMMYKUEAHD20121122WL853675201210011996					
12	2012071PIYDWPUGSMIWIKEQELXGOEXDCDATIZVCQC20120825PI602850201207011968					
13	2012041JFLUCTSKJGUUYZUIYUALXICJTHBBWVQQA20120503JF352025201204011939					
14	2012021CTZLHXRNAGRPDIRTICOHILHOKIKKKOKEI20120210CH101200201202011911					

< 上一步(B)

下一步(N) >

完成

取消

輔助說明

Input -Text File (固定寬度) 3/3

文字匯入精靈 - 固定寬度步驟 6 之 4

指定各變數開始的位置。第一欄是欄 0。

若要插入變數分行，請在尺規或資料區中所需的位置按一下。或者，使用方向鍵或輸入欄號移至所需位置；然後按下「插入分行」按鈕。

若要移動變數分行，請將它拖曳至新位置。

若要刪除變數分行，請選取分行或輸入分行位置。接著按下 Delete 鍵或「刪除分行」按鈕。

尺規(0... 70)

1 2012091EMNJMSL... 201209161361931201209011940112

2 2012021CMSQSYH... 20120210CM111106201202011912072

3 2012101WJGHXDG...

4 2012071PGUYDHE...

5 2012041TDHPJLD...

6 2012021CAVGOPCM...

7 2012101WXIXUAE...

8 2012071PUWNYZ1...

9 2012041RJEFYUJEYFMRJMUJQJLTJJ EQVTSOLLQF201205031355327201204011940022

10 2012091FOWJL... 201209161361931201209011940112

欄編號(C): 129

目前的變數寬度(V): 1

插入分行(S) 刪除分行(D)

< 上一步(B) 下一步(N) > 完成 取消 輔助說明

文字匯入精靈 - 步驟 6 之 5

資料預覽中所選的變數規格

i Data format is determined from the values present in the first 200 records. If a column contains multiple data types in the first 200 records, the variable type is set to string. The length (number of characters) for string variables is determined by the longest value present in the first 200 records. If subsequent records have longer values, they will be truncated.

變數名稱(V): FEE_YM

資料格式(D): 字串

編輯變數名稱與格式

資料預覽

FEE_YM	APPL_TYPE	HOSP_ID	APPL_DATE	CASE_TYPE	SEQ_NO	F
201209	1	EMNJMSL...	20120916	JP	361931	2
201202	1	CMSQSYH...	20120210	CM	111106	2
201210	1	WJGHXDG...	20121122	WJ	860279	2
201207	1	PGUYDHE...	20120826	PG	609454	2
201204	1	TDHPJLD...	20120503	ID	259620	2

< 上一步(B) 下一步(N) > 完成 取消 輔助說明

SPSS的起始欄位為0，
利用資料集譯碼簿來編輯變數寬度。

合併檔案-Add Cases

✓ 新增觀察值

VAR 1
X
Y
Z

+

VAR 1
M
N
O

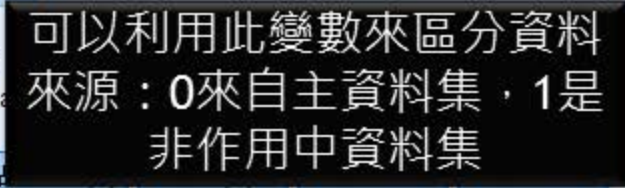


VAR 1
W
X
Y
M
N
O

變數固定，增加樣本數則為新增觀察值

Ex：合併不同年度的門診資料

新作用中資料集會 出現的變數



合併檔案-Add Variables

✓ 新增變數

ID	VAR 1
1	W
2	X
3	Y
4	Z

+

ID	VAR 2
1	10
2	20
3	30
4	40



ID	VAR 1	VAR 2
1	W	10
2	X	20
3	Y	30
4	Z	40

兩個資料集中，增加同樣關鍵變數樣本的其他變數資訊。

Ex：合併case的疾病紀錄與藥品紀錄

新增變數 1/2

新增變數來源 資料集26

被排除的變數(E):

新作用中資料集(N):

- ORDER_TYPE<(*)
- DRUG_NO>(*)
- DRUG_USE<(*)
- DRUG_FRE>(*)
- UNIT_PRICE(*)
- TOTAL_QTY(*)
- TOTAL_AMT(*)
- ORDER_SEQ_NO(*)
- CURE_ITEM_NO1<(+)
- CURE_ITEM_NO2<(+)

更名(A):

☒ 重要變數上的相符觀察值

☐ 在兩個資料集中，以重要變數的順序排序觀察值

☒ 兩檔皆提供觀察值(B)

☐ 非作用中資料集是索引表(K)

☐ 作用中資料集是索引表(A)

☐ 指明觀察值來源為變數(I): source01

(*)=作用中資料集

(+)=資料集26

關鍵變數(V):

- APPL_DATE<
- APPL_TYPE<
- CASE_TYPE<

確定 貼上之後(P) 重設(R) 取消 輔助說明

設定關鍵變數來確保合併時的位置

新增變數 2/2

主作用資料集

ID	VAR 1
1	W
2	X
3	Y
4	Z

非作用資料集

ID	VAR 2
3	10
4	20
5	30
6	40

+



兩檔皆提供觀察值

ID	VAR 1	VAR 2
1	W	-
2	X	-
3	Y	10
4	Z	20
5	-	30
6	-	40

非作用中資料集是索引表

ID	VAR 1	VAR 2
1	W	-
2	X	-
3	Y	10
4	Z	20

以主作用資料集的關鍵變數為主，串聯非作用資料集的變數資料

作用中資料集是索引表

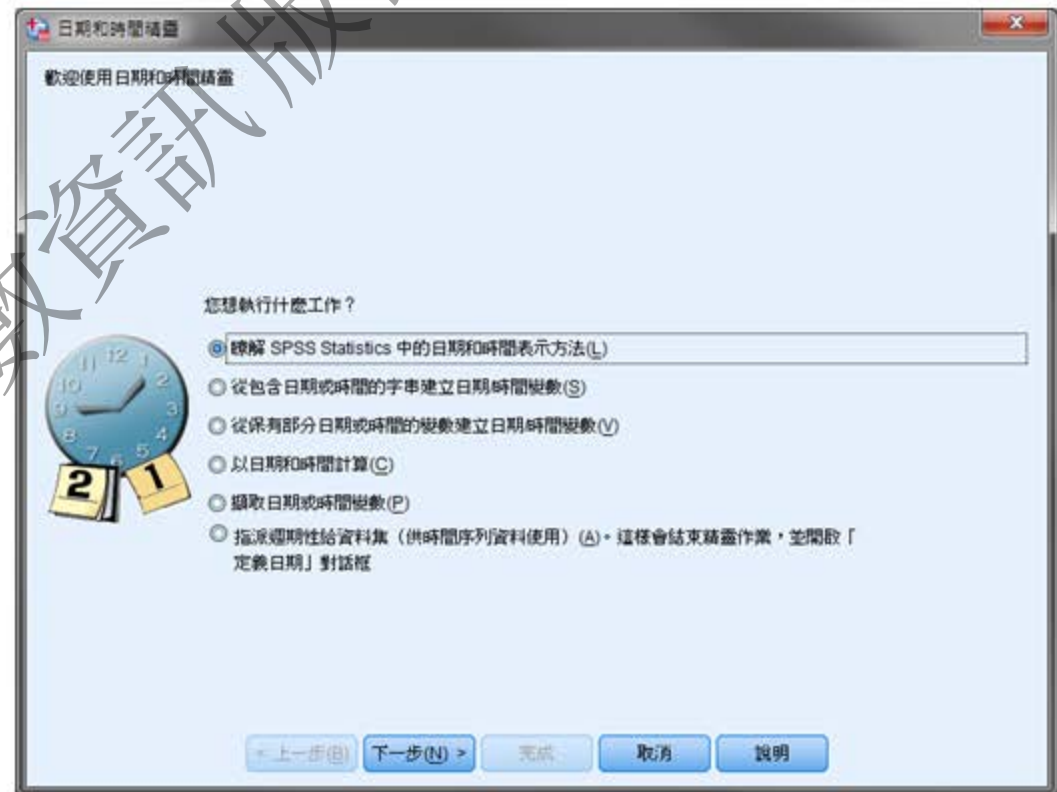
ID	VAR 1	VAR 2
3	Y	10
4	Z	20
5	-	30
6	-	40

以非作用資料集的關鍵變數為主，串聯主作用資料集的變數資料

Date / Time

轉換→日期和時間精靈 (也可使用計算變數中的時間函數計算)

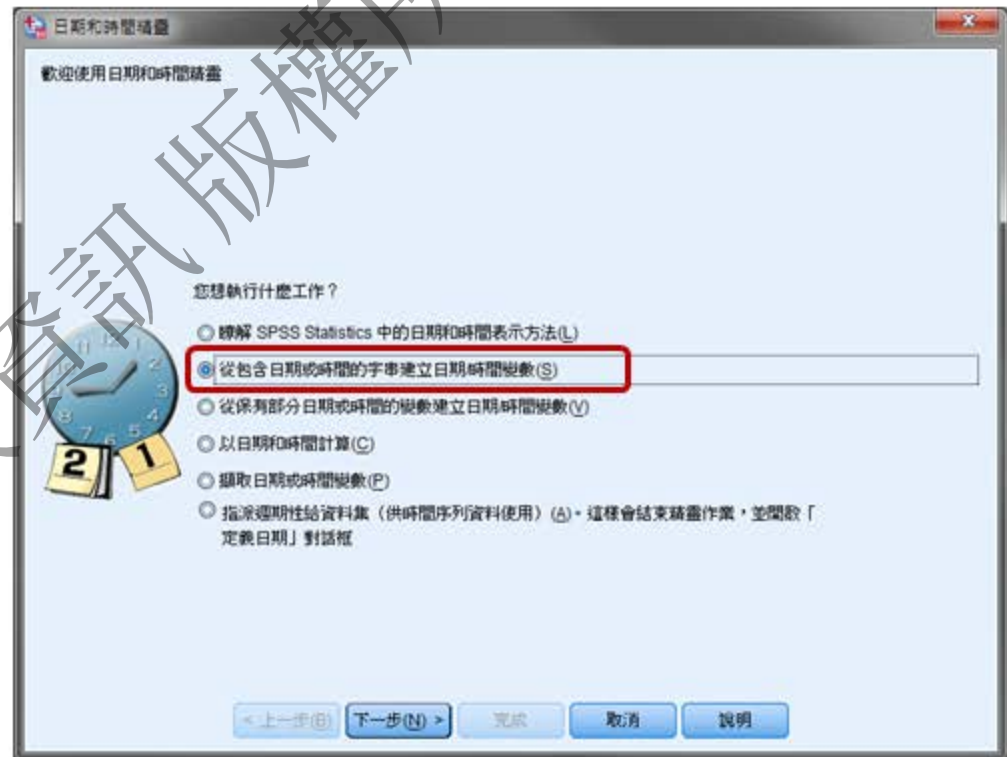
- ✓ 建立日期/時間變數
 1. 從單個完整日期字串變數
 2. 從多個部分日期字串
- ✓ 萃取部分日期/時間變數
- ✓ 計算日期時間長度



建立日期/時間變數 1/4

1. 從單個完整日期字串變數

字串	日期
APPL_DATE	APPL_DATE_Trans
20121128	2012/11/28
20120104	2012/01/04
20121111	2012/11/11
20120720	2012/07/20
20120426	2012/04/26
20120104	2012/01/04
20121111	2012/11/11
20120719	2012/07/19
20120426	2012/04/26
20120103	2012/01/03
20121110	2012/11/10
20120719	2012/07/19
20120426	2012/04/26
20120103	2012/01/03
20121110	2012/11/10
20120719	2012/07/19



建立日期/時間變數 2/4



建立日期/時間變數 3/4

2. 從多個部分日期字串

	ID	ProductNumber	Amount	Month	Day	Year
1	126	566	\$177	10	27	2012
2	862	214	\$47	7	13	2012
3	859	252	\$52	1	12	2012
4	571	236	\$43	9	15	2012
5	548	274	\$53	4	3	2012
6	569	281	\$78	6	16	2012
7	868	420	\$131	2	2	2012
8	560	555	\$157	5	23	2012
9	761	392	\$100	9	5	2012
10	443	556	\$174	9	27	2012
11	22	467	\$103	2	3	2012
12	437	558	\$158	2	11	2012
13	154	238	\$53	9	27	2012
14	936	207	\$4	10	20	2012
15	243	282	\$77	9	21	2012
16	359	323	\$84	3	24	2012
17	758	466	\$125	5	4	2012
18	425	309	\$70	1	11	2012
19	714	559	\$174	8	30	2012
20	248	118	\$10	10	28	2012
21	14	500	\$133	3	21	2012
22	929	221	\$68	8	27	2012
23	211	275	\$0	7	28	2012

透過年、月、日變數
建立日期變數，建立
完整之日期資料。

建立日期/時間變數 4/4

日期和時間變數

歡迎使用日期和時間變數。

您想執行什麼工作？

- ☐ 瞭解 SPSS Statistics 中的日期和時間表示方法 (L)
- ☐ 從包含日期和時間的字符串建立日期時間變數 (L)
- ☒ 從具有部分日期的時間的變數建立日期時間變數 (V)
- ☐ 以日期和時間計算 (C)
- ☐ 提取日期或時間變數
- ☐ 從連續性時資料集 (將時間變數列資料使用) (A) - 這裡有 (定義日期) 對話方塊

日期和時間變數 - 步驟 2 之 1

您可以包含年、月與小時等部分資訊的數值變數建立日期時間或持續時間變數。這部分的資訊將被時間變數。

在下面選取適當的變數，然後按 [下一步]

變數 (V):

- ☒ 顧客ID [ID]
- ☒ 消費產品 [ProductNumber]
- ☒ 消費金額 [Amount]

年 (Y):

- ☒ 消費日期_年 [Year]

月 (M):

- ☒ 消費日期_月 [Month]

日 (D):

- ☒ 消費日期_日 [Day]

結果變數類型: 日期

日期和時間變數 - 步驟 2 之 2

您可以為結果變數命名並指定格式。

結果變數 (R):

Date

輸出格式 (O):

- ☒ dd-mmm-yyyy
- ☐ dd-mmm-yy
- ☐ mm/dd/yyyy
- ☐ mm/dd/yy
- ☐ dd.mm.yyyy
- ☐ dd.mm.yy
- ☐ yyddd

變數標記 (L):

日期

執行:

- ☒ 立即建立變數 (C)
- ☐ 將語法貼到語法視窗 (P)

	ID	Product Number	Amount	Month	Day	Year	Date
1	126	566	\$177	10	27	2012	27-Oct-2012
2	862	214	\$47	7	13	2012	13-Jul-2012
3	859	252	\$52	1	12	2012	12-Jan-2012
4	571	236	\$43	9	15	2012	15-Sep-2012
5	548	274	\$53	4	3	2012	03-Apr-2012
6	569	281	\$75	6	15	2012	15-Jun-2012
7	868	420	\$131	2	2	2012	02-Feb-2012
8	560	555	\$157	5	23	2012	23-May-2012
9	761	392	\$100	9	5	2012	05-Sep-2012
10	443	556	\$174	9	27	2012	27-Sep-2012
11	22	467	\$103	2	3	2012	03-Feb-2012

萃取部分日期/時間變數 1/2

申報日期	申報年份
APPL_DATE_Trans	APPLY_DATE_YEAR
2012/11/28	2012
2012/01/04	2012
2012/11/11	2012
2012/07/20	2012
2012/04/26	2012
2012/01/04	2012
2012/11/11	2012
2012/07/19	2012
2012/04/26	2012
2012/01/03	2012
2012/11/10	2012
2012/07/19	2012
2012/04/26	2012
2012/01/03	2012
2012/11/10	2012
2012/07/19	2012



萃取部分日期/時間變數 2/2

日期和時間摘錄 - 步驟 1/2

將部分日期或時間作為一般數值變數。

選擇一個變數以及要摘錄的日期或時間元件。

變數(V):

- 目前日期和時間 [TIME]
- 就醫日期 [FUNC_DATE_Trans]
- 治療結束日期 [TREAT_END_...]

日期或時間(O):

申報日期 [APPL_DATE_Trans]

格式: yyyy/mm/dd

要摘錄的單位(U):

年數

TIME 是目前的日期和時間。

日期和時間摘錄 - 步驟 2/2

輸入變數: APPL_DATE_Trans

摘取的單位: 年數

結果變數(R):

輸出格式(O):

變數標籤(L):

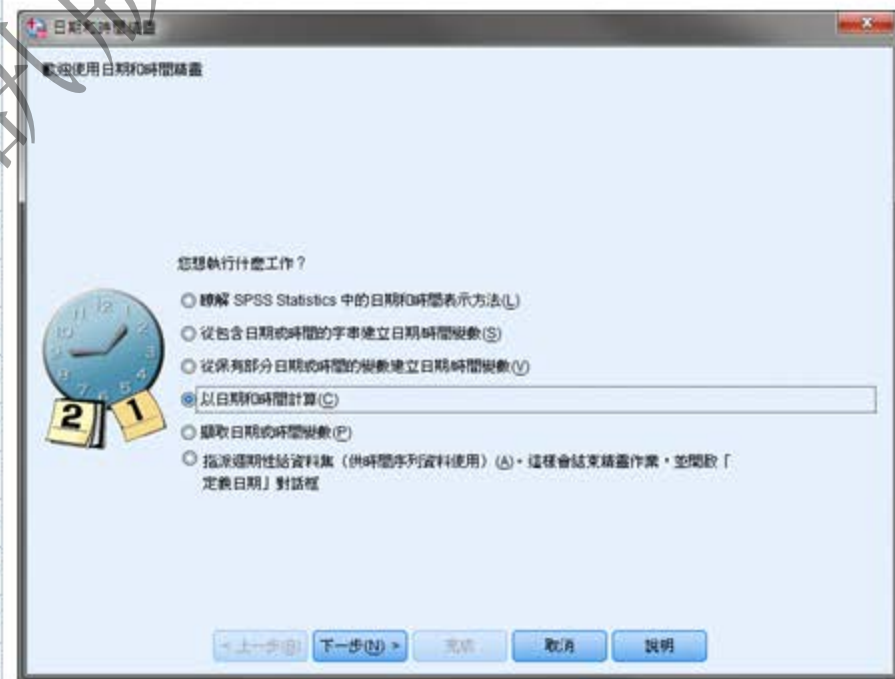
申報年份

執行

☒ 立即建立變數(C) ☐ 將語法粘到語法視窗(P)

計算日期時間長度 1/3

就醫日期	治療結束日期	
FUNC_DATE_Trans	TREAT_END_DATE_Trans	TREAT_Days
2012/11/01	2012/11/09	8
2012/01/01	2012/01/02	1
2012/09/01	2012/09/24	23
2012/06/01	2012/06/17	16
2012/04/01	2012/04/09	8
2012/01/01	2012/01/02	1
2012/09/01	2012/09/24	23
2012/06/01	2012/06/16	15
2012/04/01	2012/04/09	8
2012/01/01	2012/01/01	0
2012/09/01	2012/09/24	23
2012/06/01	2012/06/16	15
2012/04/01	2012/04/09	8
2012/01/01	2012/01/01	0
2012/09/01	2012/09/24	23
2012/06/01	2012/06/16	15
2012/04/01	2012/04/09	8



計算日期時間長度 2/3

日期和時間設置 - 步驟 1/3

計算日期

選擇下列其中一個作業並按「下一步」

- ☐ 從日期加上或減去持續時間(A) (例如, 在年紀加上一個月, 或是在日期時間變數加上時間變數)
- ☒ 計算介於兩個日期之間的時間單位數(C) (例如, 以年為單位, 計算從生日至另一個日期的年紀)
- ☐ 減去兩段持續時間(B) (例如, 工作時間 - 通勤時間)

← 上一步(B) 下一步(N) > 完成 取消 說明

日期和時間設置 - 步驟 2/3

計算介於兩個日期或日期時間變數之間的時間單位數。

結果將是整數變數。將會捨棄單位的任何小數部分。結果將是持續時間變數。僅有持續時間變數會顯示於下列變數清單。

變數(V):

- 目前日期和時間 [STL_...]
- 申報日期 [APPL_DA_...]

日期1(D):

治療結束日期 [TREAT_END_DATE_Trans]

減去日期2(U):

就醫日期 [FUNC_DATE_Trans]

單位(U):

天數

結果處理

- ☒ 截斷為整數(I)
- ☐ 捨入為整數(R)
- ☐ 保留分數部分(F)

對於月份和年份單位, 除非使用截斷功能, 否則結果是依平均單位長度為主。

TIME 是目前的日期和時間。

← 上一步(B) 下一步(N) > 完成 取消 說明

計算日期時間長度 3/3

日期和時間精靈 - 步驟 3/3

計算: TREAT_END_DATE_Trans - FUNC_DATE_Trans

結果變數(R): TREAT_Days 單位(U): 天數

變數標籤(L): 治療天數

執行

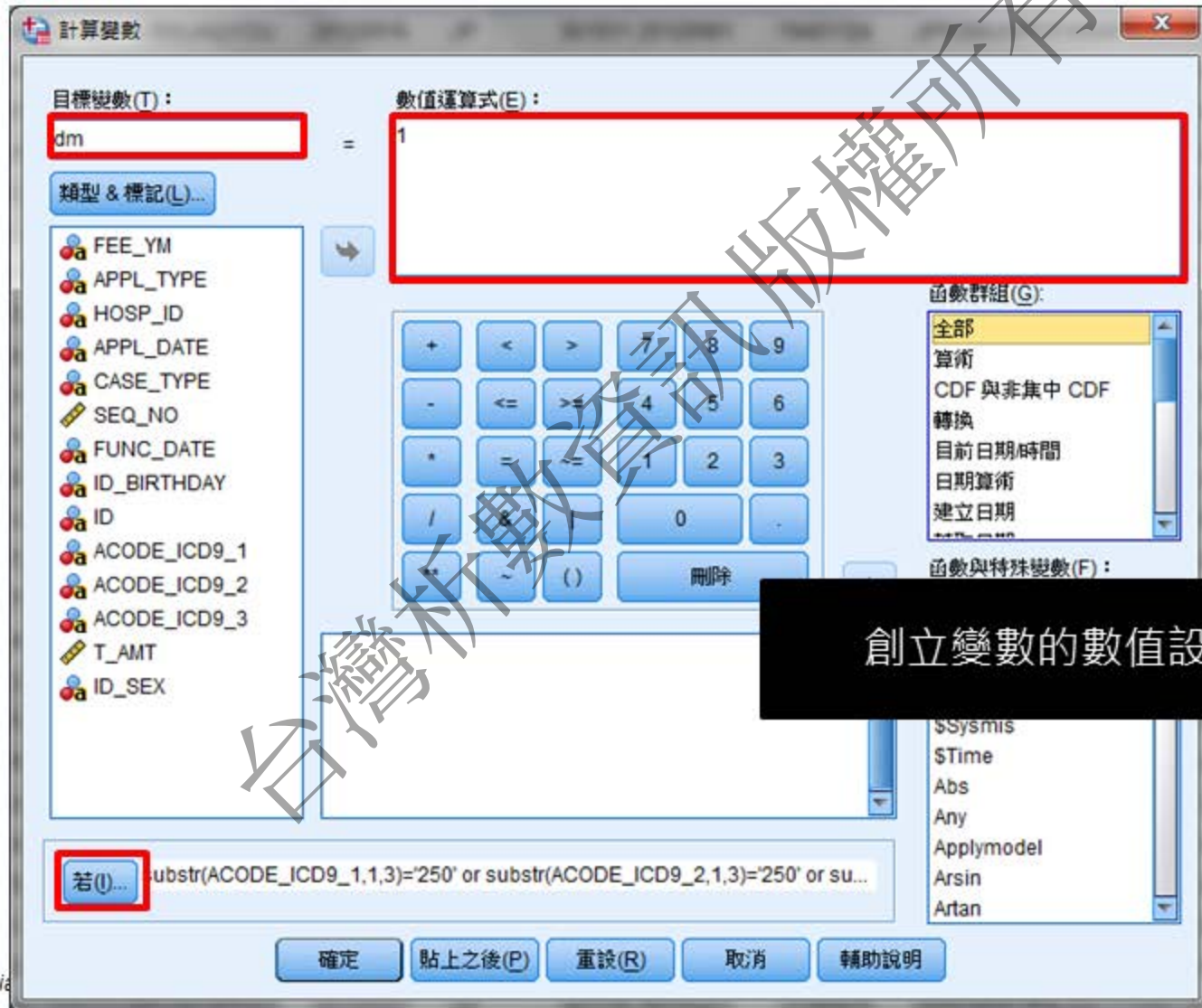
☒ 立即建立變數(C) ☐ 將語法貼到語法視窗(D)

< 上一步(B) 下一步(F) > 完成 取消 說明

治療天數

TREAT_Days
8
1
23
16
8
1
23
15
8
0
23
15
8
0
23
15
8

Find Case 1/2



創立變數的數值設定

Find Case 2/2

計算變數：觀察值選擇條件

☐ 包含全部觀察值(A)

☒ 包含滿足條件時的觀察值(I)：

`substr(ACODE_ICD9_1,1,3)="250" or substr(ACODE_ICD9_2,1,3)="250" or
substr(ACODE_ICD9_3,1,3)="250"`

函數群組(G):

全部
算術
CDF 與非集中 CDF
轉換
目前日期/時間
日期算術
建立日期
統計函數

函數與特殊變數(F)：

\$Casenum
\$Date
\$Date11
Any
Applmodel
Arsin

繼續 取消 輔助說明

設定疾病辨別條件

Paper報表操作分析



資料分析流程 1/2

資料合併

資料篩選

資料分析

NHRI 全民健康保險研究資料庫
National Health Insurance Research Database



Table 1. Demographic characteristics of patients with stroke, diabetes, and general population

Characteristic	Total (N = 1,000,000)	Stroke (N = 100,000)	Diabetes (N = 100,000)	General (N = 1,000,000)
Age (years)	65.0 ± 15.0	65.0 ± 15.0	65.0 ± 15.0	65.0 ± 15.0
Gender				
Male	500,000 (50%)	50,000 (50%)	50,000 (50%)	500,000 (50%)
Female	500,000 (50%)	50,000 (50%)	50,000 (50%)	500,000 (50%)
Ethnicity				
Chinese	950,000 (95%)	95,000 (95%)	95,000 (95%)	950,000 (95%)
Other	50,000 (5%)	5,000 (5%)	5,000 (5%)	50,000 (5%)
Education				
Elementary	100,000 (10%)	10,000 (10%)	10,000 (10%)	100,000 (10%)
High school	400,000 (40%)	40,000 (40%)	40,000 (40%)	400,000 (40%)
University	500,000 (50%)	50,000 (50%)	50,000 (50%)	500,000 (50%)
Marital status				
Married	700,000 (70%)	70,000 (70%)	70,000 (70%)	700,000 (70%)
Single	300,000 (30%)	30,000 (30%)	30,000 (30%)	300,000 (30%)

Table 2. Prevalence of stroke, diabetes, and general population

Characteristic	Stroke (N = 100,000)	Diabetes (N = 100,000)	General (N = 1,000,000)
Age (years)	65.0 ± 15.0	65.0 ± 15.0	65.0 ± 15.0
Gender			
Male	50,000 (50%)	50,000 (50%)	500,000 (50%)
Female	50,000 (50%)	50,000 (50%)	500,000 (50%)
Ethnicity			
Chinese	95,000 (95%)	95,000 (95%)	950,000 (95%)
Other	5,000 (5%)	5,000 (5%)	50,000 (5%)
Education			
Elementary	10,000 (10%)	10,000 (10%)	100,000 (10%)
High school	40,000 (40%)	40,000 (40%)	400,000 (40%)
University	50,000 (50%)	50,000 (50%)	500,000 (50%)
Marital status			
Married	70,000 (70%)	70,000 (70%)	700,000 (70%)
Single	30,000 (30%)	30,000 (30%)	300,000 (30%)

Table 3. Logistic regression results of factors which predict stroke, diabetes, and general population

Characteristic	Stroke (N = 100,000)	Diabetes (N = 100,000)	General (N = 1,000,000)
Age (years)	1.00	1.00	1.00
Gender			
Male	1.00	1.00	1.00
Female	1.00	1.00	1.00
Ethnicity			
Chinese	1.00	1.00	1.00
Other	1.00	1.00	1.00
Education			
Elementary	1.00	1.00	1.00
High school	1.00	1.00	1.00
University	1.00	1.00	1.00
Marital status			
Married	1.00	1.00	1.00
Single	1.00	1.00	1.00

資料分析流程 2/2



2005年
承保抽樣歸人檔



年齡為0~17歲
共226,550筆資料



資料篩選、分組與
合併外部都市化資料

Age(years old)
<6
6-11
12-17

Area
East
North
Middle
South

Urbanization
Rural
Urban
Suburban

ADHD
No
Yes

Allergic condition
no allergic disorders
AR or BA or AD
AR
AD
BA
AR and AD
AR and BA
AD and BA
AR and BA and AD



Table 1

Table 1 Demographic characteristics of patients with allergic disorders and general population

variable	Total (n = 226550) n (%)	General population (n = 178093) n (%)	with Allergic disease (n = 48457) n (%)	p-value for χ^2 test
Age(years old)				
<6	74973 (33.1)	53277 (29.9)	21696 (44.8)	<0.001
6-11	81529 (36.0)	64322 (36.1)	17207 (35.5)	
12-17	70048 (30.9)	60494 (34.0)	9554 (19.7)	
Gender				
Female	108314 (47.8)	87750 (49.3)	20564 (42.4)	<0.001
Male	118236 (52.2)	90343 (50.7)	27893 (57.6)	
Area				
North	111616 (49.3)	86101 (48.4)	25515 (52.7)	<0.001
Middle	44600 (19.7)	35316 (19.8)	9284 (19.2)	
South	64907 (28.7)	52081 (29.2)	12826 (26.5)	
East	5427 (2.4)	4595 (2.6)	832 (1.7)	
Urbanization				
Urban	131663 (58.1)	102507 (57.6)	29156 (60.2)	<0.001
Suburban	32372 (14.3)	25613 (14.4)	6759 (14.0)	
Rural	62515 (27.6)	49973 (28.0)	12542 (25.9)	
ADHD				
No	225243 (99.4)	177214 (99.5)	48029 (99.1)	<0.001
Yes	1307 (0.6)	879 (0.5)	428 (0.9)	

Table 2

Table 2 Prevalence of ADHD in children with allergic disorders and the general population

variable	General population (n = 178093)	with Allergic disorders (n = 48457)	Odds ratio	95% CI	p-value for χ^2 test
	n (%)	n (%)			
Age(years old)					
<6	219 (0.4)	123 (0.6)	1.38	1.11 ~ 1.72	0.005
6-11	519 (0.8)	257 (1.5)	1.86	1.60 ~ 2.17	<0.001
12-17	141 (0.2)	48 (0.5)	2.16	1.56 ~ 3.00	<0.001
Gender					
Female	182 (0.2)	69 (0.4)	1.62	1.23 ~ 2.14	<0.001
Male	697 (0.8)	359 (1.3)	1.68	1.48 ~ 1.91	<0.001
Area					
North	532 (0.6)	278 (1.1)	1.77	1.53 ~ 2.05	<0.001
Middle	105 (0.3)	50 (0.5)	1.82	1.30 ~ 2.54	<0.001
South	220 (0.4)	91 (0.7)	1.68	1.32 ~ 2.15	<0.001
East	22 (0.5)	9 (1.1)	2.27	1.04 ~ 4.95	0.044
Urbanization					
Urban	619 (0.6)	309 (1.1)	1.76	1.54 ~ 2.02	<0.001
Suburban	87 (0.3)	42 (0.6)	1.83	1.27 ~ 2.65	0.002
Rural	173 (0.4)	77 (0.6)	1.78	1.36 ~ 2.33	<0.001

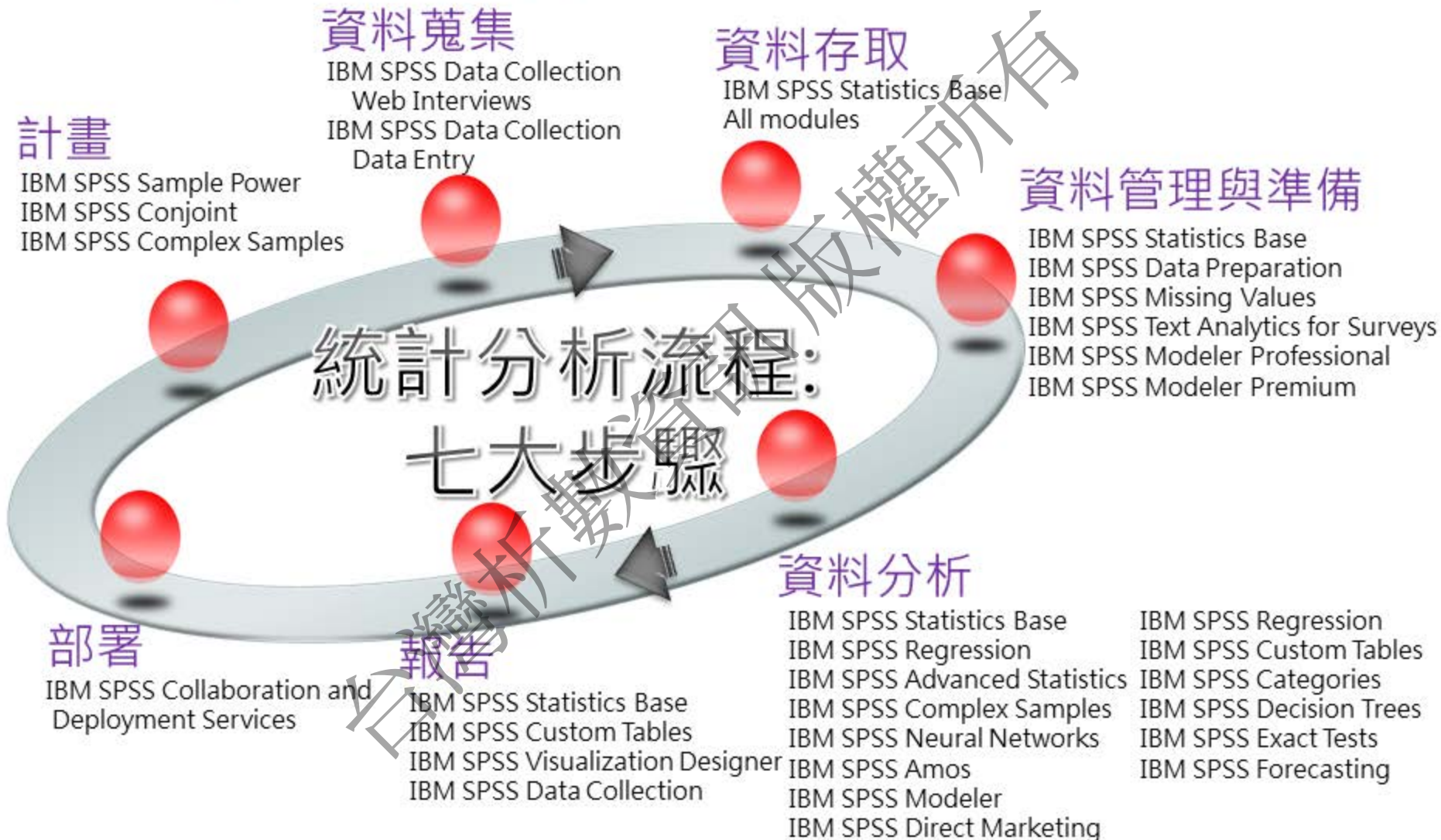
Table 3

Table 3 Logistic regression model of factors with prevalence of ADHD in pediatric allergic patients

variable	Odds ratio	95% CI		p-value
		Lower	Upper	
Age(years old)				
<6	1.00	—	—	
6-11	2.10	1.85	2.39	<0.001
12-17	0.62	0.51	0.74	<0.001
Gender				
Female	1.00	—	—	
Male	3.76	3.28	4.32	<0.001
Area				
East	1.00	—	—	
North	1.12	0.78	1.61	0.528
Middle	0.61	0.41	0.89	0.011
South	0.81	0.56	1.17	0.265
Urbanization				
Rural	1.00	—	—	
Urban	1.58	1.37	1.82	<0.001
Suburban	0.99	0.80	1.23	0.930
Allergic condition				
no allergic disorders	1.00	—	—	
AR or BA or AD	1.56	1.38	1.75	<0.001
AR	1.71	1.48	1.98	<0.001
AD	0.73	0.48	1.09	0.124
BA	1.38	1.02	1.86	0.039
AR and AD	2.14	1.37	3.35	0.001
AR and BA	1.63	1.28	2.06	<0.001
AD and BA	0.72	0.18	2.89	0.639
AR and BA and AD	2.02	1.19	3.46	0.010

R-Square = 0.0666.

SPSS與統計分析



專業顧問諮詢服務 Consultancy Services



需求評估與規劃服務
Service of Assessment
and Planning

需求評估與規劃服務

Service of Assessment and Planning

不論您的產業規模，台灣析數依據您業務上分析的需求及資料的狀況做詳細的產品需求評估，以標準的統計與數據挖掘流程為您規劃分析專案導入所需的資源及執行工作，以最適化的配置及產品的應用來達到最高的效益。



預測分析之概念驗證
Proof of Concepts for
Predictive Analysis

預測分析之概念驗證

Proof of Concepts for Predictive Analysis

為了協助您更快的評估預測分析為業務所帶來的實際效果，台灣析數基於用戶業務挑戰與現有資料的狀況，透過分析概念驗證 (Proof of Concepts, POC) 的方式，讓用戶更清楚與明顯的感覺預期效果的實現，並提出分析數據與成果報告呈報上級，了解解決方案所帶來的具體成效及投資報酬率(ROI)。



解決方案規劃與執行

Solution Planning and Execution

軟體使用引領與導正

許多用戶回覆並未能確切掌握軟體功能應用到實際的業務問題，為協助您將軟體功能發揮得淋漓盡致，讓工作更有效率，台灣析數的專業顧問將親自了解您在業務上的工作，並教導您如何正確有效的使用SPSS軟體，借助軟體功能優點為您解決手上難解的問題，領引您到正確的使用途徑。

完整解決方案規劃與執行

各產業所遭遇艱困的問題，仰賴現有商業邏輯與經驗亦無法得到預期的成果！亦可能您已理解數據解剖與預測模型能解開謎題及提供最適的決策支援，但卻沒有時間與人力來實際的深入數據的世界獲取成功的秘密。AsiaAnalytics台灣析數團隊國內外的專家顧問群，擁有各產業多年的解決方案經驗，包含銀行業、保險業、零售業、電信業、政府機關等，將為您的商業難題與挑戰，透過進階的數據分析技術與方法論來解析問題的原由，並以預測分析支援決策擬定，進行提升財務與營運之績效。

功能模組加值服務
Service of Added Module

功能模組加值服務

Service of Added Module

各產業面臨許多實務上的議題，往往是經過繁覆的流程與試誤才得到最後的結果，但重複耗時性的工作僅局限於專門的人員才能執行。因此，若能簡化資料整理及分析流程，讓分析軟體更易容使用，不論對個人或部門之工作都能有事半功倍的成效。

技術諮詢服務

為確保用戶正常的使用IBM SPSS系列軟體系統，SPSS台灣析數用戶可透過客服信箱(service@asia-analytics.com.tw)或電話聯絡專業技術支援人員，免費協助用戶處理軟體運行錯誤之問題。此外，在AsiaAnalytics台灣析數網站上『下載專區』提供SPSS各產品修正插件及相關支援手冊，台灣析數以高品質的技術服務已獲得眾多SPSS用戶的肯定。

諮詢服務聯絡方式

服務電話：02 29030885

電子郵件信箱：service@asia-analytics.com.tw

Thanks for your attention



Business Intelligence Analysis
The Best Solution
Provider



台灣析數

