

資料整理與檢誤經驗談

以 SPSS 程式進行邏輯檢查

蘇婉雯

邏輯檢誤相較於不合理值檢誤要來的更複雜、費事，也更需要研究者的專業知識。邏輯檢誤，主要是檢查受訪者在通篇問卷的答案是否前後一致、是否合理。藉由這道檢查，找到數據資料的可疑之處，進而修改之。例如：某人回答目前的信仰為「佛教」，但回答最常求助的神職人員為「神父」；某人回答年齡為「25 歲」，回答親生子女就讀學校卻為「XX 大學」；某人性別為「男」，應跳過生育經驗不需作答，卻回答了生育第一胎的年齡為 30 歲。像這些狀況，都需要翻查原始問卷，查看是否因鍵入錯誤而產生的不合理現象。

一般而言，邏輯檢誤所發現的錯誤答案，可能源自：

1. 問卷設計不明確，未將跳題或續答之相關敘述標明清楚，而導致訪員或受訪者答題時失誤。
2. 受訪者的失誤。如：誤解題意、記憶模糊、精神狀態不佳、蓄意隱瞞等。
3. 訪員的失誤。如：弄錯限答對象、應跳問卻沒有跳問、填錯格子、寫錯答案等。
4. 過錄員的失誤。如：誤填跳答碼、填錯過錄格、寫錯過錄碼等。
5. 輸入員的失誤。如：看錯格子或代碼、按錯輸入鍵(通常以正確鍵前後左右四個鍵最常誤觸)等。

進行「邏輯檢誤」時，覆查問卷的工作相當費神，查核者往往需要前後對照數道題目。如果能盡量減少這個階

段需要查驗的變項，將會是比較有效率的作法。因此，檢誤工作比較理想的方式是先做「不合理值檢誤」，再做「邏輯檢誤」(「不合理值的檢誤」請參考第二期內容)。不合理值檢誤能夠先確立每一個變項的數據皆落在合理範圍內；也就是說，理想上，可以去除大半的書寫及鍵入錯誤；此時，再做邏輯檢誤，可以比較專注於真正邏輯問題的查驗，較有效率。

邏輯檢誤通常需要同時考量兩題以上的答題情形，以判斷合理的邏輯；但為了簡化複雜的情形，最常見的做法是兩兩考量，如果還需要第三題合併考量，再慢慢加進來。需要邏輯檢誤的情形包羅萬象，最重要的是，檢誤者需要先對問卷結構與內涵、受訪者特性有深入瞭解之後，再依照需要設計程式檢誤。職是之故，唯有原問卷設計的研究人員最清楚該檢誤的邏輯，二手資料蒐集者及使用者是很難進行這部份的工作。在此將邏輯檢核分成三類型來說明。第一種類型是跳/續答邏輯檢核，依據問卷的跳題設計來檢驗受訪者或是訪員是否跳答正確；第二種類型是一般邏輯檢核，用來檢驗受訪者在問卷題目中前後的答案是否一致；第三種類型是複選題的邏輯檢核，用來檢查複選題選項之間的邏輯關係等。

一、跳 / 續答邏輯檢核

依據變項屬性的不同，我們分別介紹「類別變項」對「類別變項」以及「類別變項」對「連續變項」的處理方式。

TABLE1.問卷題目範例一

題號	變項名稱	變項說明	選項數值說明
5	V5	請問您的婚姻狀況	1.未婚/同居(跳答第8題) 2.已婚/分居 8.拒答 3.離婚/喪偶 9.missing
6	V6	請問您是在哪一年結婚的？民國____年	0. 不適用 98.拒答 99.missing
7	V7	請問您配偶是哪裡人？	1.本省人 8.拒答 2.外省人 9.missing 3.原住民 0.不適用 4.其他

(一)「類別變項」對「類別變項」
交叉列聯表：將交叉列聯表中前後不合理邏輯的域值挑出。

步驟一：執行交叉列聯表程式

【自撰語法】

CROSSTABS TABLES= v7 BY v5.

【點選選單】

Analyze Descriptive Statistics
Crosstabs 選擇變項(row:v5
column:v7)

TABLE 2. V5 與 V7 交叉列聯表結果

v7 請問您配偶是哪裡人？ * v5 請問您的婚姻狀況		Crosstabulation			
Count		v5 請問您的婚姻狀況			
		1 未婚/同居	2 已婚/分居	3 離婚/喪偶	4 Total
v7 請問您配偶是哪裡人？	0 不適用	43	1		45
	1 本省人	1	21	10	32
	2 外省人		12	8	21
	3 原住民		13	4	17
	4 其他		4		4
	5	1			1
	6	1			1
Total		46	51	22	121

在此例中，可以輕易地發現部分選項數值未定義，這些數值就是不合理值，應在邏輯檢查之前先排除掉。本例的跳 /

續答邏輯中，應該挑出 V5.婚姻狀況回答(2.已婚/分居)或(3.離婚/喪偶)者，其 V7.詢問配偶狀況時應續答而未續答者。

步驟二：從交叉列聯表中找出不合邏輯的值域，配合語法找出樣本編號

【自撰語法】

temporary.

select if any(v7,0) and any(v5,2,3).

list id v5 v7.

執行結果，列出下面資訊：

```
ID   V5  V7
1103 2   0
```

【點選選單】

Data Select Cases If
condition is satisfied any(v5,2,3)
and any(v7,0) 設定不合格樣本
處理方式 (filtered or deleted)

Data Sort Cases 選擇上述
filter 的新變項 將資料排序，檢查
DATA VIEW 視窗中的資料內容

FIGURE1. 點選選單篩檢條件，經排序後DATA VIEW所顯示的結果

未被 filter 的資料，即為不合邏輯條件者。

	id	u1	v1	v2	v3	v4	v5	v6	v7	
1	1113	1115	女	35	73744531	專科學校	已婚分居	55	不適用	4
2		1347	女	44	34719475	高中(高職)	未婚同居	67	本省人	1
3		1352	女	45	55866130	初中國中	已婚分居	71	本省人	1
4		1158	男	54	78314914	初中國中	已婚分居	78	本省人	1
5		1111	男	55	73415955	高中(高職)	已婚分居	75	外省人	1
6		1185	女	56	27524038	專科學校	離婚喪偶	90	本省人	2
7		1156	女	58	89232355	未入學	已婚分居	88	本省人	1
8		1182	男	68	45246811	研究所以上	已婚分居	83	外省人	1

續答與跳答是經常同時存在的兩個相對應的邏輯，以此例為例，亦應檢查 V5.婚姻狀況回答(1.未婚)者，其 V7.詢問配偶狀況時應跳答(0.不適用)而未跳答者。

在 SPSS 中，需關係運算符號時，可以使用下列三種語法：

1. any(varname, value1, value2,...)
2. varname = value1, varname ~= value2,
varname > value3, varname < value4,
varname >= value5, varname <= value6
3. varname eq value1, varname ne value2,
varname gt value3, varname lt value4,
varname ge value3, varname le value4

第一種語法可以同時選用多個數值，而省去每個數值撰寫而使程式長而不容易檢查錯誤。第二種語法與第三種語法所代表的意義是相同的。這三種語法可以依需要與習慣混合使用。就上面所提到的跳答邏輯，就可以有下面的兩種寫法：

```
temporary.
select if v5=1 and any(v7,1,2,3).
list id v5 v7.
```

或

```
temporary.
select if v5=1 and v7~=0.
list id v5 v7.
```

邏輯運算符號"and"也可以寫成"&"，"or"也可以寫成"|"。關係運算符號與

邏輯運算符號可以完全依照個人習慣來使用。

另外，利用變項間數值的對應關係，將變項排序後可以找出不相符的個案。但是當樣本數較大或檢核連續變項時，則不建議使用。接下去的檢查，將介紹語法的撰寫，視窗的點選不再另外說明。

(二)「類別變項」對「連續變項」

當樣本數較大或連續變項數值分佈較廣時，為避免執行 crosstab 的結果表格太大，建議直接以語法設定邏輯運算關係以進行檢核。因此，以下範例省略交叉列聯表結果。「連續變項」對「連續變項」亦建議以下列語法直接進行檢核。

【自撰語法】

```
crosstab v6 by v5. (交叉列表結果略)
temporary.
select if v5=1 and v6~=0.
list id v5 v6.
```

```
temporary.
select if any(v5,2,3) and v6=0.
list id v5 v6.
```

執行結果，列出下面資訊：

ID V5 V6
1147 1 67

ID V5 V6
1114 2 0

二、一般邏輯檢核

一般變項邏輯的檢核，意指檢查受訪者對問卷中具相關性的問題，是否有一致的回答。例如詢問家中是否有電腦，以及在什麼地點使用網際網路，若受訪者先回答家中沒有電腦，不應再回

答主要使用網際網路的地點在家裡。一般變項間的邏輯是較難檢驗的，通常需要瞭解整份問卷的設計以及填答方式，才能夠掌握所有變項間的邏輯關係。一般邏輯檢核的方式與跳答邏輯檢核的方式大致上相同。下列介紹以交叉列聯表檢核「類別變項」對「類別變項」；及利用新建變項檢核「連續變項」對「連續變項」兩種。

(一)「類別變項」對「類別變項」

TABLE 3. 問卷題目範例二

題號	變項名稱	變項說明	選項數值說明
9	V9	請問您家中是否有電腦設備？	1.是 2.否 8.拒答 9.missing
10	V10	請問您大部分都是在什麼地點使用網際網路？	1.家中 2.公司 3.親朋好友或同學家 4.網咖 5.不會/沒有在上網 6.其他 8.拒答 9.missing

【自撰語法】

crosstabs tables v10 by v9.

TABLE 4. V9與V10交叉列聯表結果

V10 請問您大部分都是在什麼地點使用網際網路？
V9 請問您家中是否有電腦設備？ Crosstabulation

Count		V9 請問您家中 是否有電腦設 備？			Total
		1 是	2 否		
V10 請問 您大 部分 都是 在什 麼地 點使 用網 際網 路？	1 家中	8	11	19	
	2 公司	16	19	35	
	3 親朋好友或同學家	11	4	15	
	4 網咖	9	15	24	
	5 不會/沒有在上網	11	9	20	
	6 其他	4	3	7	
	8 拒答	1		1	
	Total	60	61	121	

在此例中，V9.答家中無電腦者，v10中不應出現在家裡上網，因此需列出這些不合邏輯者受訪者，進一步查詢內容。

【自撰語法】

temporary.

select if any(v9,2) and any(v10,1).

list id v9 v10.

列出結果如下：

ID	V9	V10
1147	2	1
1183	2	1
1156	2	1
1182	2	1
1133	2	1
1134	2	1
1149	2	1
1105	2	1
1191	2	1
1203	2	1
1218	2	1

(二)「連續變項」對「連續變項」

對兩連續變項進行數學運算產生新

變項，再界定新變項的範圍，藉以挑出無理值及樣本編號。

TABLE 5. 問卷題目範例三

題號	變項名稱	變項說明	選項數值說明
2	V2	請問您是在哪一年出生的？民國____年	98.拒答 99.missing
.(略).	.(略).	.(略)	.(略)
6	V6	請問您是在哪一年結婚？民國____年	0.不適用 98.拒答 99.missing

【自撰語法】

compute v6_2=v6-v2.

temporary.

select if v6_2<=15 and v6~=0.

list id v2 v6 v6_2.

列出結果如下：

ID	V2	V6	V6_2
1182	68	83	15
1184	69	84	15
1122	75	90	15
1214	75	90	15
1169	68	80	12
1128	54	32	-22

【點選選單】新建變項

Transform Compute 定義
Target Variables 設定變數邏輯關係(v6_2=v6-v2)

Data Select Cases If condition
is satisfied 設定 v6_2<=15
設定不合格樣本處理方式
(filtered or deleted)

Data Sort Cases 選擇上述
filter 的新變項 將資料排序，檢
查 DATA VIEW 視窗中的資料內
容

三、複選題組的檢查

複選題組的檢查是進階級的資料整理工作，運用上面提過的方式，在比較複雜的邏輯中使用。常見的複選題邏輯可以簡單的分為兩類：一種是複選題組的不合理值檢查，另一種是複選題組的邏輯檢查。接下來就直接舉例介紹。

TABLE.6 問卷題目範例四

題號	變項名稱	變項說明	選項數值說明
一	A1	請問您是否聽說過中央研究院調查研究工作室「學術調查研究資料庫」？	(1) 是(續答第二題) (2) 否(跳答第二十題) (9) missing
二、請問您最初是從何處得知本室資料庫？(可複選)			
(1)	A2.1	中央研究院週報	(1) 是 (2) 否 (8) 跳答 (9) missing
(2)	A2.2	網際網路(WWW)	
(3)	A2.3	本室出版品(「調查研究」等)	
(4)	A2.4	學術論文會議上	
(5)	A2.5	學校課堂上	
(6)	A2.6	同學或朋友	
(7)	A2.7	執行國科會研究計畫	
(8)	A2.8	其他(請說明)	

《過錄說明》

本題組中，第一題中回答"1.是"者，需續答第二題；回答"2.否"者，第二題全數為"8.跳答"；未回答者，過錄為"9.missing"，第二題亦全為"9.missing"。

需回答第二題者(第一題回答"1.是"者)，有勾選之題目過錄為"1.是"，未勾選過錄為"2.否"，至少應勾選一小題。未勾選(1)-(7)題者，應勾選第(8)題。但(1)-(7)題有勾選者，仍可勾選第(8)題。全數未填答者，則應過錄為"9.missing"

(一) 複選題題組的不合理值檢查

在資料整理時應先就第二題(1)-(8)小題的資料進行不合理值檢查。光就第二題，可以整理出下列三條邏輯：

- 1.至少要有一題回答 1，不可全為 2。
(題目設計(8)與(1)-(7)互斥)
- 2.需跳答者，應全數跳答；資料遺漏者，應全數遺漏。
- 3.選項(1,2)與選項(8,9)應為互斥。(跳 / 續答邏輯)

在此特別說明，若將數值設為系統遺失值(system missing)，SPSS 邏輯語法篩選資料時，將略過系統遺失值的運算以及與其他數值的比對，因此篩選時宜取消系統遺失值的設定。以下說明如何挑出第二題的八小題全數回答 "2.否" (選項 8.與選項 1.-7.有互斥的設計)者，利用在上面介紹過的邏輯語法，將變項與條件一個個串接起來。

【自撰語法】

/*挑出第二題的八小題全數回答 "2.否" */

TEMP.

SELECT IF A2.1=2 and A2.2=2 and A2.3=2 and A2.4=2 and A2.5=2 and A2.6=2 and A2.7=2 and A2.8=2.

LIST ID A2.1 TO A2.8.

列出結果

```
ID A2.1 A2.2 A2.3 A2.4 A2.5 A2.6 A2.7 A2.8
495 2 2 2 2 2 2 2 2
```

題目數與邏輯條件越多，逐一撰寫語法隨之多而且容易遺漏。在這裡利用新建變項的方式，將複選題組轉換成單純的

幾個變項，在其後的檢查都可以利用。

【自撰語法】

/*依據邏輯第 1,2 條，過錄題組依據其邏輯轉成單一變項 a2，未符合邏輯者，a2=0*/

COMPUTE A2=0.

IF

(A2.1=1|A2.2=1|A2.3=1|A2.4=1|A2.5=1|A2.6=1|A2.7=1|A2.8=1) A2=1.

IF (A2.1=8 & A2.2=8 & A2.3=8 & A2.4=8 & A2.5=8 & A2.6=8 & A2.7=8 & A2.8=8) A2=8.

IF (A2.1=9 & A2.2=9 & A2.3=9 & A2.4=9 & A2.5=9 & A2.6=9 & A2.7=9) A2=9.

/*依據邏輯第 3 條，挑出在八題中任一位置出現 8 與 9 的數值，轉成單一變項 A2a=88 與 99*/

COMPUTE A2a=0.

IF (A2.1=8 | A2.2=8 | A2.3=8 | A2.4=8 | A2.5=8 | A2.6=8 | A2.7=8 | A2.8=8) A2a=88.

IF (A2.1=9 | A2.2=9 | A2.3=9 | A2.4=9 | A2.5=9 | A2.6=9 | A2.7=9) A2a=99.

/*挑出不符合邏輯第 1,2 條的資料*/

temp.

select if a2=0.

list id a2 a2.1 to a2.8.

(結果一)

/*挑出不符合邏輯第 3 條的資料*/

temp.

select if a2=1 and (a2a=88 | a2a=99).

list id a2 a2a a2.1 to a2.8.

(結果二)

結果一：不只列出全為 2 的資料，也列出其他邏輯錯誤的情形

```
LINE 1: ID A2 A2.1 A2.2 A2.3 A2.4 A2.5
```

```
LINE 2: A2.6 A2.7 A2.8
```

```
ID: 379 .00 8 2 2 2 2
```

```
A2.6: 2 2 2
```

```
ID: 495 .00 2 2 2 2 2
```

```
A2.6: 2 2 2
```

```
ID: 875 .00 8 8 8 8 8
```

```
A2.6: 8 8 9
```

結果二：列出 1 與(8,9)同時存在的資料

LINE 1: ID A2 A2A A2.1 A2.2 A2.3 A2.4

LINE 2: A2.5 A2.6 A2.7 A2.8

ID: 9998 1.00 88.00 8 1 2 2

A2.5: 2 1 2 2

ID: 2829 1.00 99.00 2 9 2 2

A2.5: 2 2 1 2

SPSS 的報表中變項數超過一行時，將在報表前以 line 1, line 2 等標明每一行中變項出現的次序。由上面這兩種撰寫方式的結果比較起來，利用新建變項再撰寫邏輯語法，可以同時檢查好幾個邏輯問題，不需逐一撰寫。這樣的檢查工作，才能確切檢查複選題組中是否包含不合理值，應在進行下一個跳 / 續答的複選題組檢查前先進行。

(二)跳 / 續答的複選題組的檢查

以挑出第一題回答 "1.是"者，第二題的八小題中，不該出現"8.跳答"(跳 / 續答之後的複選題組)；第一題回答 "2.否"者，第二題的八小題中，應全數為 "8.跳答"，而不該出現"1、2、9"為例子；第一題為遺漏者，第二題的八小題中，應全數為 "9.跳答"，而不該出現"1、2、8"為例子。

【自撰語法】

temp.
select if a1=1 and a2~=8.
list id a1 a2 a2.1 to a2.8.
(結果一)

temp.
select if a1=2 and a2~=8.
list id a1 a2 a2.1 to a2.8.
(結果二)

temp.
select if a1=9 and a2~=9.
list id a1 a2 a2.1 to a2.8.
(結果三)

結果一：列出第一題選 1 而第二題不該跳答卻跳答者，以及在前一步驟就應挑出的不合理值。

LINE 1: ID A1 A2 A2.1 A2.2 A2.3 A2.4

LINE 2: A2.5 A2.6 A2.7 A2.8

ID: 2465 1 8.00 8 8 8 8

A2.5: 8 8 8 8

ID: 379 1 .00 8 2 2 2

A2.5: 2 2 2 2

ID: 495 1 .00 2 2 2 2

A2.5: 2 2 2 2

結果二：列出第一題選 2 而第二題該跳答卻未跳答者，以及在前一步驟就應挑出的不合理值。

LINE 1: ID A1 A2 A2.1 A2.2 A2.3 A2.4

LINE 2: A2.5 A2.6 A2.7 A2.8

ID: 2583 2 1.00 1 2 1 2

A2.5: 2 2 2 2

ID: 875 2 .00 8 8 8 8

A2.5: 8 8 8 9

結果三：列出第一題為遺漏值，而第二題亦應為遺漏但卻出現其他答案者。

LINE 1: ID A1 A2 A2.1 A2.2 A2.3 A2.4

LINE 2: A2.5 A2.6 A2.7 A2.8

ID: 1229 9 8.00 8 8 8 8

A2.5: 8 8 8 8

ID: 2350 9 1.00 2 2 2 2

A2.5: 2 1 2 2

上述的邏輯錯誤，在不合理值檢查時不容易被發現。因此需要進一步的複選題邏輯檢查。邏輯檢核是一個相當重要且辛苦的工作，卻常被研究者所忽視。邏輯的繁複程度，會依題目的設計而有所不同，需要心思縝密的進行。邏輯語法的撰寫，往往因人而異而多采多姿。在辛苦之中，能撰寫各種程式處理各式狀況，及時修正資料內容，確保資料品質，也是相當具有成就感的一份工作。本室連續兩期撰寫資料整理工作的經驗與學術界分享，希望藉此拋磚引玉，進而提升國內調查資料的品質。(完)