

## เทคนิคการตรวจสอบข้อมูลก่อนวิเคราะห์ด้วย โปรแกรม SPSS/PC<sup>+</sup>

ดร. ปัญญา ธีระวิทยเลิศ

การประมวลผลข้อมูลด้วยคอมพิวเตอร์ โดยใช้โปรแกรมสำเร็จรูป SPSS/PC<sup>+</sup> (Statistical Package for the Social Sciences) นั้นช่วยทำให้การวิเคราะห์ข้อมูลคำนวณค่าสถิติต่างๆ มีความรวดเร็วและแม่นยำ แต่ไม่ได้หมายความว่างานวิจัยที่ใช้คอมพิวเตอร์วิเคราะห์ค่าสถิตินี้จะมีความถูกต้องเชื่อถือได้มาก ทั้งนี้ขึ้นอยู่กับความเหมาะสมของการเลือกใช้ค่าสถิติต่างๆ ที่สอดคล้องกับระดับการวัดและชนิดของตัวแปรต่างๆ การแปลความหมายค่าสถิติที่ได้จากการคำนวณ และที่สำคัญยิ่งคือ **ความถูกต้องของข้อมูล** ซึ่งถ้าข้อมูลมีความคลาดเคลื่อนแล้วการสรุปผลการวิจัยย่อมมีความคลาดเคลื่อนด้วยอย่างแน่นอน ซึ่งการวิเคราะห์ข้อมูลด้วยคอมพิวเตอร์นี้จะมีโอกาสผิดพลาดได้จากการลงรหัส (Coding) หรือการบันทึกข้อมูล (Data Entry) ผิดพลาด จึงจำเป็นที่ผู้วิจัยจะต้องตรวจสอบข้อมูลให้มีความถูกต้องก่อนที่จะเขียนโปรแกรมคำสั่งวิเคราะห์หาค่าสถิติต่อไป

การตรวจสอบ (Editing) อาจใช้การอ่านทวนตรวจสอบด้วยตาซึ่งยังมีโอกาสผิดพลาดเช่นเดิม โดยเฉพาะเมื่อมีข้อมูลจำนวนมาก ผู้วิจัยจึงควรใช้**การตรวจสอบด้วยโปรแกรม (Machine Editing)** ด้วย เพราะจะช่วยหาตำแหน่งที่ผิดพลาดและจะได้แก้ไขข้อมูลให้มีความถูกต้องมากยิ่งขึ้น โดยการตรวจสอบนี้สามารถใช้เทคนิคการเขียนคำสั่งร่วมกันหลายๆ คำสั่งในโปรแกรม SPSS/PC<sup>+</sup>

## ลักษณะการตรวจสอบควรมีการตรวจสอบให้ครอบคลุมใน 3 ลักษณะ คือ

1. การตรวจสอบความครบถ้วนของข้อมูล (Completeness of Record)
2. การตรวจสอบความถูกต้องเฉพาะรหัสแต่ละตัวแปร (Range Check)
3. การตรวจสอบความถูกต้องของรหัสในตัวแปรหลายตัวที่มีความสัมพันธ์กัน (Relation Check)

## การตรวจสอบความครบถ้วนของข้อมูล (Completeness of Record)

ในกรณีที่ข้อมูลแต่ละ cases ใช้รหัสมากกว่า 1 record อาจจะเป็นไปได้ที่การบันทึกข้อมูลไม่ครบ เช่น บาง cases ขาดหายไปบาง record หรืออาจมีการบันทึกซ้ำทำให้จำนวน record เกินในบาง cases กรณีการผิดพลาดเช่นนี้จะทำให้ข้อมูลผิดต่อเนื่องจำนวนมาก เพราะในการทำงานของโปรแกรมคอมพิวเตอร์นั้นจะทำตามคำสั่งทุกขั้นตอน เช่นเมื่อเขียนคำสั่ง DATA LIST บอกว่าในแต่ละ cases มี 2 record แต่ข้อมูลไม่ถูกต้องคือมีมากกว่าหรือน้อยกว่า 2 record เมื่อคอมพิวเตอร์อ่านถึงข้อมูลของ cases ที่มีจำนวน record ไม่ครบถ้วน ก็จะอ่านข้อมูลของ cases ถัดไปมาแทนที่และข้อมูล cases ต่อๆ ไปก็จะสลับกันหมด ส่งผลให้ข้อมูลผิดพลาดต่อเนื่องจำนวนมาก

การตรวจสอบว่าข้อมูลแต่ละ case มีจำนวน record ถูกต้องครบถ้วนหรือไม่นั้นสามารถใช้เทคนิคง่ายๆ ตรวจสอบโดยการเขียนคำสั่ง COMPUTE เพื่อสร้างตัวแปรใหม่ที่เกิดจากผลรวมของตัวแปรที่เป็นเลขประจำแถว (record number) ที่เป็นค่าคงที่มารวมกัน แล้วใช้คำสั่ง PROCESS IF เพื่อคัดเลือกข้อมูลที่ตัวแปรที่สร้างใหม่มีค่านอกเหนือจากค่าที่ควรเป็น และตามด้วยคำสั่ง LIST ให้แสดงว่า record ใดที่มีค่าตัวแปรที่สร้างใหม่เป็นค่าผิดปกติ ด้วยหลักการนี้สามารถค้นหาข้อมูล cases ดังกล่าวที่ผิดพลาดที่มีจำนวน record ขาดหายไปหรือเกินมา เพื่อทำการแก้ไขเพิ่มเติมให้ถูกต้องแล้วจึงทดสอบโดยโปรแกรมเดิมอีกครั้ง ถ้าพบว่ามี cases อื่นผิดอีก ก็ต้องแก้ไขต่อไปอีกจนกว่าจะพบว่าค่าตัวแปรที่สร้างใหม่นี้มีค่าเท่ากับทุกค่า อาจทดสอบได้โดยการใช้คำสั่ง FREQUENCIES แทนการใช้คำสั่ง PROCESS IF เมื่อปรากฏค่าความถี่ของค่าตัวแปรใหม่นี้เท่ากับจำนวนหน่วยตัวอย่าง แสดงว่าจำนวน record ในแต่ละ cases ถูกต้อง ข้อมูลมีความครบถ้วน และทำการตรวจสอบลักษณะอื่นต่อไป

## การตรวจสอบความถูกต้องเฉพาะรหัสแต่ละตัวแปร (Range Check)

ความผิดพลาดนี้เกิดจากการลงรหัสผิดหรือลงรหัสถูกแต่การบันทึกข้อมูลผิด โดยสามารถตรวจสอบด้วยโปรแกรมได้ ถ้ารหัสที่ผิดพลาดนั้นอยู่นอกขอบข่ายของรหัสที่เป็นไปได้ (possible code) เช่น ตัวแปร เพศ คำตอบที่เป็นไปได้มี 2 คำตอบและได้กำหนดรหัสคือ 1 กับ 2 ถ้าการลงรหัสหรือการบันทึกข้อมูลผิดเป็นรหัสอื่น คือ 3 4 หรืออื่นๆ สามารถตรวจสอบโดยโปรแกรมว่า cases ที่ผิดพลาดที่รหัสของตัวแปรเพศนี้ เป็นรหัสนอกเหนือจากเลข 1 และ 2 และ cases ไหนที่ผิดพลาดให้แสดงมาด้วยว่า cases ที่เท่าใดที่มีความผิดพลาดบ้าง เพื่อจะได้กลับไปเรียกข้อมูลใน cases นั้นๆ มาแก้ไขให้ถูกต้อง

เทคนิคง่ายๆ โดยใช้คำสั่ง PROCESS IF หรือ SELECT IF เพื่อให้คัดเลือกตัวแปรที่มีค่านอกขอบเขตของคำตอบที่เป็นไปได้ และตามด้วยคำสั่ง LIST เพื่อให้แสดงออกมาว่า cases ไหนที่มีค่าผิดพลาด คือมีรหัสนอกเหนือจากที่ควรจะเป็น และทำการแก้ไขโดยตรวจสอบจากข้อมูลดิบ

## การตรวจสอบความถูกต้องของรหัสในตัวแปรหลายตัวที่มีความสัมพันธ์กัน (Relation Check)

ในกรณีที่ข้อความบางข้อความมีความสัมพันธ์เกี่ยวเนื่องกัน คำตอบข้อหนึ่งๆ ต้องสัมพันธ์กันด้วยจึงจะเชื่อถือได้ เช่น คำถามข้อหนึ่งคือสถานภาพสมรส และคำถามอีกข้อหนึ่งคือจำนวนบุตร ซึ่งในแบบสอบถามอาจกำหนดรหัสไว้ว่าถ้าสถานภาพสมรสเป็นโสด ให้รหัส 1 และคำถามจำนวนบุตรไม่ถามให้ข้ามและใส่รหัสไม่เข้าชาย คือ 9 ถ้าการบันทึกข้อมูลผิดคือข้อแรกรหัส 1 เป็นโสด ส่วนข้อสองใส่รหัส 7 หมายถึงมีบุตร 7 คน แสดงว่าข้อมูลมีการลงรหัสหรือบันทึกผิด

เทคนิคการหาที่ผิดพลาดของข้อมูลตัวแปรที่มีความสัมพันธ์กันนี้ ใช้คำสั่งเงื่อนไข คือ คำสั่ง IF โดยใช้คำสั่งนี้ค้นหาว่า cases ไหนบ้างที่ข้อแรกมีรหัสเป็น 1 แต่ข้อ 2 รหัสไม่ใช่ 9 แล้วตามด้วยคำสั่ง LIST ให้แสดงออกมาว่า cases ไหนที่ผิดพลาดเช่นนี้ แล้วจึงกลับไปตรวจสอบและแก้ไขข้อมูลใน cases ที่ผิดพลาด

**ตัวอย่างแสดงการตรวจสอบข้อมูลด้วยโปรแกรม SPSS/PC<sup>+</sup>**

ในงานวิจัยเรื่อง การรับรู้ข่าวสารการเกษตรของเกษตรกรในภาคเหนือ โดยผู้เข้ารับการฝึกอบรมหลักสูตรนักวิจัยทางสังคมศาสตร์ รุ่นที่ 42 ของสำนักงานคณะกรรมการวิจัยแห่งชาติ แบบสอบถามมีจำนวนข้อคำถาม 38 ข้อ ใช้รหัส 2 record ในแต่ละ cases ลักษณะของ coding book เป็นดังนี้

ข้อที่	ตัวแปร	รหัสที่เป็นไปได้	เลขที่คอลัมน์	หมายเหตุ
1)	เลขที่แบบสอบถาม (ID)	1 - 1004	1 - 4	ถ้าข้อ 12 ตอบ 2 ข้อ 13 ไม่เข้าข่ายที่จะต้องถาม
	record number (RN1)	1	5	
	จังหวัด (PROV)	1 นครสวรรค์ 2 แพร่	6	
2)	พื้นที่ทำการเกษตร (AREA)	1 ในเขตชลประทาน 2 นอกเขตชลประทาน	7	
3)	เพศ (SEX)	1 ชาย	8	
		2 หญิง		
4)	อายุ (AGE)	15 - 60 ปี	9 - 10	
-	---	---	--	
-	---	---	--	
-	---	---	--	
12)	การอ่านข่าวสารเกษตรจากหนังสือพิมพ์ (B12)	1 อ่าน	23	
		2 ไม่อ่าน		
13)	ความบ่อยในการอ่านข่าว (B13)	1 ทุกวัน	24	
		2 เกือบทุกวัน		
		3 นานๆ ครั้ง		
		9 ไม่เข้าข่าย		
38.5	เลขที่แบบสอบถาม (ID)	1 - 1004	1 - 4	
	record number (RN2)	2	5	
	การทราบสถานการณ์ความต้องการข่าวโพตในปี 2536	1 ทราบ 2 ไม่ทราบ	6	
--	---	---	--	
39	ความคิดเห็นต่อข่าวสารเกษตรที่เป็นประโยชน์	1 การเพาะปลูก	19	
		2 การเลี้ยงสัตว์		
		3 บัณฑิตการผลิต		
		4 บัณฑิตสนับสนุนการผลิต		
		5 การตลาด		
		0 ไม่ฟัง/ไม่ทราบ/ไม่ตอบ		

บันทึกข้อมูลชื่อเพิ่มข้อมูลว่า DATA จำนวนทั้งสิ้น 1004 cases case ละ 2 record ดังนี้

```
00011112602218207107011299911131299211100100000250000097A22222022222202522210909
00012111111111109092102
00021112402218201201011299922991299114600001000260000096A22012000000001521210909
000220909090909092101
00031112564218101401411299911232299214100100000250000094A22011000100008509090909
000320909220909092106
.....
.....
10041111274218301201211140511131112111600233300230000394A20011100200002309090911
100420909090909093506
```

### วิธีการตรวจสอบข้อมูล

1. การตรวจสอบความครบถ้วนของข้อมูล

เขียนคำสั่งโดยใช้ SPSS ดังนี้

```
DATA LIST FILE = 'A:DATA'
  /ID 1-4 RN1 5 PROV 6 AREA 7 SEX 8 AGE 9-10 STATUS 11 EDU 12 B12 23 B13 24
  /RN2 5 N9 TO N20 6-17 R36 18 N39 19 N40 20-21.
COMPUTE SR = RN1 + RN2.
PROCESS IF (SR NE 3).
LIST ID SR RN1 RN2.
```

ผลการทำงานของคำสั่งต่างๆ ปรากฏดังนี้

```
The raw data or transformation pass is proceeding
      1003 cases are written to the uncompressed active file.

      ID          SR  RN1  RN2
      .605        2.00  1    1

Number of cases read =      1    Number of cases listed =      1
```

**ความหมาย :** แสดงว่าเครื่องได้อ่านข้อมูลทั้งหมด 1003 cases หายไป 1 case case ที่หายไปคือ case สุดท้ายเนื่องจากจำนวน record ไม่ครบ 2 record และสาเหตุที่ record สุดท้ายไม่ครบเพราะมี case ที่ 605 ซึ่งมีจำนวน record ไม่ครบ (พิจารณาจากค่า SR ที่เท่ากับ 2 ซึ่งที่ถูกต้องทุกค่าควรเป็น 3 แสดงว่า case ที่ 605 record ที่ 2 ขาดหายไป โปรแกรมจึงไปอ่าน record ที่ 2 ของ case ที่ 606 มาเป็น record ที่ 2 แทน)

เมื่อได้เพิ่มเติม record ที่ 2 ของ case ที่ 605 แล้วจึงทำการตรวจสอบโดยโปรแกรมและหาค่าความถี่ของ SR พบว่าเท่ากับ 3 จำนวน 1004 cases แสดงว่าข้อมูลชุดนี้มีความถูกต้องครบถ้วนแล้ว จึงทำการตรวจสอบในขั้นที่ 2 ต่อไป

2. การตรวจสอบความถูกต้องเฉพาะรหัสแต่ละตัวแปร

ตัวอย่าง เช่น ตัวแปรจังหวัดซึ่งคำตอบมี 2 จังหวัด รหัสที่เป็นไปได้คือ 1 กับ 2 เขียนคำสั่งดังนี้

```
DATA LIST FILE = 'A:DATA'

  /ID 1-4 RN1 5 PROV 6 AREA 7 SEX 8 AGE 9-10 STATUS 11 EDU 12 B12 23 B13 24

  /RN2 5 N9 to N20 6-17 R36 18 N39 19 N40 20-21.

PROCESS IF (PROV > 2).

LIST ID PROV.
```

ผลการทำงานตามโปรแกรมคำสั่งปรากฏผลดังนี้

The raw data or transformation pass is proceeding

1004 cases are written to the uncompressed active file.

ID PROV

17 7

Number of cases read = 1 Number of cases listed = 1

**ความหมาย :** ข้อมูลทั้งสิ้น 1004 cases มีอยู่ 1 cases คือเลขที่ 17 ที่ค่าตัวแปรจังหวัดเป็น 7 แสดงว่ามีการลงรหัสหรือบันทึกข้อมูลผิดพลาดจึงควรกลับไปดูข้อมูลดิบในชุดนี้ว่าที่ถูกต้องคือจังหวัดที่ 1 หรือ 2 แล้วทำการแก้ไขและตรวจสอบตัวแปรอื่นๆ ด้วย

### 3. การตรวจสอบความถูกต้องของรหัสในตัวแปรหลายตัวที่มีความสัมพันธ์กัน

ตัวอย่างตัวแปรที่มีความสัมพันธ์กันในการลงรหัสคือข้อที่ 12 และ 13 ถ้าข้อ 12 รหัสคือ 1 ข้อ 13 รหัสควรเป็น 1 หรือ 2 หรือ 3 และถ้าข้อ 12 รหัส 2 ข้อ 13 รหัสควรเป็น 9 การใช้โปรแกรมตรวจสอบดังนี้

DATA LIST FILE = 'A:DATA'

/ID 1-4 RN1 5 PROV 6 AREA 7 SEX 8 AGE 9-10 STATUS 11 EDU 12 B12 23 B13 24

/RN2 5 N9 to N20 6-17 R36 18 N39 19 N40 20-21.

IF (B12 = 1 AND B13 > 3)B = 0.

PROCESS IF (B = 0).

LIST ID B12 B13 B.

ผลการทำงานตามโปรแกรมคำสั่งปรากฏผลดังนี้

The raw data or transformation pass is proceeding  
 1004 cases are written to the uncompressed active file.

ID	B12	B13	B
9	1	9	0.0
171	1	9	0.0
175	1	9	0.0
205	1	9	0.0
299	1	9	0.0
391	1	9	0.0
454	1	9	0.0
594	1	9	0.0

**ความหมาย :** ข้อมูลทั้งสิ้น 1004 cases มีอยู่ 8 cases ที่มีการลงรหัสหรือบันทึกข้อมูลผิดพลาดเนื่องลักษณะข้อมูลไม่สอดคล้องสัมพันธ์ ทั้งนี้อาจจะลงรหัสผิดในข้อ 12 หรือข้อ 13 ซึ่งต้องพิจารณาข้อมูลดิบใหม่และทำการแก้ไขให้ถูกต้อง

โดยสรุป อาจกล่าวได้ว่าการวิเคราะห์ข้อมูลโดยใช้โปรแกรม SPSS นั้น ถ้าไม่ตรวจสอบแก้ไขข้อมูลให้ถูกต้องก่อน โดยข้อมูลยังมีข้อผิดพลาดอยู่อาจจะส่งผลให้โปรแกรมไม่ทำงานตามคำสั่งหรืออาจทำงานตามคำสั่งแต่ได้ผลที่คลาดเคลื่อน แน่หน่อนว่าการสรุปผลการวิจัยจากข้อมูลที่ไม่ถูกต้อง งานวิจัยนั้นย่อมไม่น่าเชื่อถือทั้งทางด้านวิชาการหรือการนำไปใช้...