

สร้างขึ้นมาจากลักษณะของฟังก์ชันที่เกิดขึ้นจากการประกอบกัน (Combination) ของรหัสจากทุก field ใน record¹ ซึ่งฟังก์ชันดังกล่าวจะเป็นเซตของรหัสที่ขัดแย้งกับความเป็นจริง เรียกว่า เงื่อนไขขัดแย้ง (fail edit) แล้วนำฟังก์ชันดังกล่าวไปใช้เป็นหลักในการบรรณาธิกรณั่ถ้ารหัสของ record ใดสอดคล้องกับเงื่อนไขขัดแย้งให้ถือว่า record นั้นผิดพลาดจำเป็นต้องสร้างข้อมูลขึ้นทดแทน (imputed data) ให้แก่รายการข้อมูล (field) ที่ผิดพลาดนั้นต่อไป

ให้ f คือ Combination ของรหัสคำตอบของ field ต่าง ๆ ที่ไม่อาจเป็นไปได้ Unacceptable เรียก f ว่า edit restriction นั่นคือ

$$f(A_1^0, A_2^0, \dots, A_n^0)$$

โดยที่ $A_i^0 \subseteq A$ และ f คือฟังก์ชันที่เชื่อมโยงอนุเซตเหล่านี้ด้วย set intersection และ set union operation¹.

จากฟังก์ชัน f ถ้า record a ใดสอดคล้องกับ f กล่าวคือ $a \in f(A_1^0, A_2^0, \dots, A_n^0)$ ก็แสดงว่า record a ผิดหรือมีข้อบกพร่อง (fail)

โดยอาศัย Distributive law ซ้ำ ๆ กันกับ $f(A_1^0, A_2^0, \dots, A_n^0)$ จะได้ Union of Intersection ดังนี้

$$f(A_1^0, A_2^0, \dots, A_n^0) = (A_{i_1}^1 \cap A_{i_2}^1 \cap \dots \cap A_{i_m}^1) \cup (A_{j_1}^2 \cap A_{j_2}^2 \dots \cap A_{j_{m_2}}^2) \cap \dots \cap (A_{k_1}^r \cap A_{k_2}^r \cap \dots \cap A_{k_{m_r}}^r)^2$$

และถ้า a เป็นสมาชิกหนึ่ง (หรือมากกว่า 1) ของเซตที่อยู่ในวงเล็บคือ $a \in C \in f$ แล้วเราจะถือว่า record a ผิดปกติหรือมีความบกพร่อง

¹. อาจต้องตัดบาง field ที่อยู่นอกความสนใจหรือไม่จำเป็นทิ้งไป

². ขอให้สังเกตดูว่า f เป็นอนุเซตของ A โดย f เป็นเฉพาะเซตของรหัสคำตอบที่ไม่อาจเป็นไปได้

³. ดูตัวอย่างเซต A และ A_i^0 ในข้อ ข. เลือกเฉพาะเซต (สมาชิกของ A และ A_i^0) ที่มีลักษณะเป็นไปได้ (Unacceptable) และตัวอย่างท้าย Lemma

ด้วยเหตุนี้เราจึงสามารถแจกแจง $\text{acf}(A_1^0, A_2^0, \dots, A_n^0)$ ออกมาเป็นเงื่อนไข เรียกว่า series of edit ได้ดังนี้คือ

$$a \in \bigcap_{i \in s} A_i^*$$

เมื่อ s คือเซตของเงื่อนไขการบรรณาธิกรณ (index set on the fields of questionnaire) และ A_i^* คืออนุเซตของ code space ในรูป $a \in A_i^0$

เราเรียก $\bigcap_{i \in s} A_i^* = F$ หรือ Normal form of edit¹ เพื่อความเข้าใจ ลองพิจารณาตัวอย่างต่อไปนี้

สมมุติรายการข้อมูล (field) ในแบบสำรวจ (record) ประกอบด้วยรายการที่เราสนใจ 4 รายการ คือ อายุ ระดับการศึกษา ความสัมพันธ์กับหัวหน้าครอบครัว และสถานภาพสมรส

จากรายการข้อมูลทั้ง 4 สามารถสร้างเงื่อนไขการบรรณาธิกรณได้ดังนี้

“ถ้าผู้ตอบอายุไม่เกิน 15 ปี หรือกำลังศึกษาอยู่ในระดับมัธยมศึกษา ย่อมแสดงว่าผู้นั้นจะต้องไม่ใช่หัวหน้าครอบครัว และต้องเป็นโสด” เงื่อนไขการบรรณาธิกรณนี้สามารถจัดรูปย่อจนเข้าสู่ Normal form เป็นลำดับขั้นตอนตามหลักตรรกวิทยาได้ดังนี้²

1. ถ้า ((อายุไม่เกิน 15 ปี) หรือ (กำลังศึกษาอยู่ในระดับมัธยมศึกษา)) => ((ไม่ใช่หัวหน้าครอบครัว) และ (เป็นโสด))
2. ((อายุไม่เกิน 15 ปี) หรือ (กำลังศึกษาอยู่ในระดับมัธยมศึกษา)) และไม่ใช่ ((ไม่ใช่หัวหน้าครอบครัว) และ (เป็นโสด)) = F

1. F = Failure ดูตัวอย่างท้าย Lemma

2. 1. if $(A \cup B) - (\sim C \cap D)$
 2. $(A \cup B) \cap \sim (\sim C \cap D)$
 3. $(A \cup B) \cap (C \cup D)$
 4. $A \cap C$
 5. $A \cap \sim D$
 6. $B \cap C$
 7. $B \cap D$

3. ((อายุไม่เกิน 15 ปี) หรือ (กำลังศึกษาอยู่ในระดับมัธยมศึกษา)) และ ((เป็นหัวหน้าครอบครัว) หรือ (ไม่โสด)) = F
4. (อายุไม่เกิน 15 ปี) และ (เป็นหัวหน้าครอบครัว) = F
5. (อายุไม่เกิน 15 ปี) และ (ไม่โสด) = F
6. (กำลังศึกษาอยู่ในระดับมัธยมศึกษา) และ (เป็นหัวหน้าครอบครัว) = F
7. (กำลังศึกษาอยู่ในระดับมัธยมศึกษา) และ (ไม่โสด) = F

สมการที่ 4 ถึง 7 เป็นสมการที่สอดคล้องกับเงื่อนไขการบรรณาธิการ ในสมการที่ 1 ซึ่งทั้ง 4 สมการนี้จัดอยู่ในรูป Normal form $\bigcup_{i \in S} A_i^* = F$ แล้ว

จะเห็นได้ว่า record ใดที่สอดคล้องกับ ข้อใดข้อหนึ่งหรือหลายข้อ คือ $a \in \bigcup_{i \in S} A_i^*$ แสดงว่า record นั้นมีการข้อมูลผิดพลาดต้องมีการแก้ไขหรือสร้างข้อมูลทดแทนให้แก่รายการที่ตรวจพบว่าเป็นตัวการหรือเหตุแห่งความผิดพลาด ซึ่งอาจเป็นเพียง field ใด field หนึ่งหรือหลาย field ก็ได้ ปัญหาที่ก็จะตรวจหาตัวการนั้นได้อย่างไร มีวิธีการดำเนินการแก้ไขในลักษณะใด

วิธีปฏิบัติสำหรับปัญหาข้างต้นก็คือ การพยายามสร้างเงื่อนไขตรวจสอบที่เป็น Normal form of edit ขึ้นมา แล้วนำข้อมูลจาก record a ใดๆ ไปเปรียบเทียบ ถ้า record a ไม่คล้อยตามเงื่อนไขแสดงว่า record a ถูกต้อง แต่ถ้า record a สอดคล้องกับเงื่อนไขทุกเงื่อนไขหรือมากกว่า 1 เงื่อนไข แสดงว่า record a มีความบกพร่องหรือผิดพลาด หน้าที่ของผู้ตรวจสอบก็คือเสาะหารายการข้อมูลที่ผิดพลาดต่อไป เราอาจสงสัยว่า field ที่ i ผิด แต่ก็อาจเป็นไปได้ที่ field อื่นผิด เพราะอาจมีข้อขัดแย้งบางประการที่แฝงอยู่ทำให้ไม่อาจตัดสินใจได้ กรณีเช่นนี้ผู้ตรวจสอบต้องตั้งข้อสังเกตไว้ก่อนว่า เงื่อนไขที่มีอยู่อาจยังไม่ครอบคลุมหรือสมบูรณ์เพียงพอ จำเป็นต้องพยายามสร้างเงื่อนไขให้ถูกต้องสมบูรณ์เสียก่อน เงื่อนไขที่สมบูรณ์เรียกว่า Complete Set of Edit

ก่อนที่จะกล่าวถึงหลักทางทฤษฎีสำหรับสร้าง Complete Set of Edit ลองพิจารณาตัวอย่างต่อไปนี้

สมมุติแบบสอบถามประกอบด้วยรายการข้อมูล 3 รายการดังนี้

รายการข้อมูล (field)	รหัสคำตอบที่เป็นไปได้
----------------------	-----------------------

อายุ	0-14, 15 ⁺
สถานภาพสมรส	โสด, สมรส, หย่า, หม้าย, แยกกันอยู่

ความสัมพันธ์กับหัวหน้าครอบครัว หัวหน้า, คู่สมรส, อื่น ๆ

และสมมุติว่าเรามีเงื่อนไขตรวจสอบ (Initial edit) อยู่ 2 เงื่อนไข

$$e_1 = (\text{อายุไม่เกิน 15 ปี}) \text{ และ } (\text{สมรสแล้ว}) = F$$

$$e_2 = (\text{ยังไม่เคยสมรส}) \text{ และ } (\text{เป็นคู่สมรส}) = F$$

โดยกำหนดให้ สมรส = {สมรส, หย่า, หม้าย, แยกกันอยู่} และยังไม่สมรส = {โสด, หย่า, หม้าย, แยกกันอยู่} และสมมุติว่าจากแบบสอบถาม ผู้ตอบได้ตอบว่า (อายุไม่เกิน 15 ปี, สมรส, เป็นคู่สมรส)

จากการตรวจสอบรหัสคำตอบของผู้ตอบรายนี้พบว่าคำตอบสอดคล้องกับ e_1 แต่ไม่สอดคล้องกับ e_2 และเมื่อพิจารณาคำตอบว่า รายการข้อมูล (field) สถานภาพสมรส ปรากฏอยู่ใน e_1 และ e_2 แม้จะต่างรหัสกันอยู่ก็ตาม ด้วยเหตุนี้เราจึงคาดว่ารายการ (field) สถานภาพสมรสน่าจะเป็นเหตุแห่งความผิดปกติของ record a และเพ่งความสนใจไปที่รายการนี้ วิธีปฏิบัติ คือ ลองแปรรหัสของสถานภาพสมรสของ record a ไปในทุกระหัสที่เป็นไปได้ (all possible code) ของรายการ “สมรสแล้ว” โดยถือว่า “อายุ” และ “ความสัมพันธ์กับหัวหน้าครอบครัว” คงที่ ซึ่งปรากฏดังนี้¹

(อายุ 0-14 ปี, สมรส, เป็นคู่สมรส), (อายุ 0-14 ปี, หย่า, เป็นคู่สมรส), (อายุ 0-14 ปี, หม้าย, เป็นคู่สมรส), (อายุ 0-14 ปี, แยกกันอยู่, เป็นคู่สมรส)

จากการพิจารณาคำตอบต่าง ๆ ของ record a ข้างต้นจะพบว่า ไม่ว่า ทดลองแปรค่ารหัสของสถานภาพสมรสไปในลักษณะ (any code) ความผิดปกติในคำตอบก็ยังคงปรากฏอยู่เช่นเดิม ด้วยข้อนำสงสัยนี้จึงชี้ให้เห็นว่า ความขัดแย้งจริง ๆ อาจมิใช่รายการ “สถานภาพสมรส” คงจะเป็นความขัดแย้งที่แอบแฝงอยู่ในระหว่าง รายการ “อายุ” และ “ความสัมพันธ์กับหัวหน้าครอบครัว” ซึ่งจะนำไปสู่การสร้างเงื่อนไขเพิ่มเติมขึ้นมาใหม่ โดยอาศัยความขัดแย้งดังกล่าว เงื่อนไขเพิ่มเติมนี้เรียกว่า Logical Impied edit เป็นเงื่อนไขที่เกิดขึ้นจาก Combination ของเงื่อนไขที่มีอยู่เดิม (Initial edit) ซึ่งถ้าเราสามารถเสาะหาเงื่อนไขเพิ่มเติมได้ครบถ้วน จะ

¹ ในรหัสของเซท “เคยสมรส”

ทำให้ Complete set of Edit หมายความว่า แม้จะได้พยายามตรวจสอบซ้ำ ๆ กันอีกหลาย ๆ ครั้งแล้วก็ตาม ก็ยังคงไม่อาจสร้างเงื่อนไขใดเพิ่มเติมขึ้นมาได้อีกแล้ว ให้ถือว่าเซตของเงื่อนไข (ทั้งเงื่อนไขเดิมและเงื่อนไขเพิ่มเติมที่เกิดขึ้นใหม่) เป็นเซตของเงื่อนไขที่สมบูรณ์ (Complete Set) ฟังก์ชันเป็นเกณฑ์การบรรณาธิกรณต่อไป

จากวิธีการดังกล่าวสามารถสรุปวิธีสร้างเงื่อนไขเพิ่มเติม (ถือเอาความสงสัยที่ว่า อาจมีข้อขัดแย้งแอบแฝงอยู่ใน field ที่คงที่) โดยอาศัยเงื่อนไขที่มีอยู่เดิม ดังนี้

1. ถ้า (อายุไม่เกิน 14 ปี) → (เป็นโสด)
2. ถ้า (ยังไม่สมรส) → (ไม่เป็นคู่สมรส)

และโดยปกติแล้วเราทราบจากหลักตรรกวิทยาเบื้องต้นพบว่า ถ้า (เป็นโสด) (ยังไม่สมรส) เราจึงนำเงื่อนไข 1 และ 2 มาผสมกันเพื่อสร้างเงื่อนไขเพิ่มเติมโดยอาศัย Transitive Law คือ

ถ้า (อายุไม่เกิน 14 ปี) → (เป็นโสด) → (ยังไม่สมรส) → (ไม่เป็นคู่สมรส)
 ดังนั้น จากข้อสงสัยในความขัดแย้งที่แอบแฝงอยู่ระหว่าง อายุ และความสัมพันธ์

กับหัวหน้าครอบครัวจึงพบความจริงว่า

ถ้า (อายุไม่เกิน 14 ปี) → (ไม่เป็นคู่สมรส)

ดังนั้น เงื่อนไขตรวจสอบที่สมบูรณ์ (Complete Set of Edit)

- e_1 : (อายุไม่เกิน 14 ปี) และ (สมรสแล้ว) : เงื่อนไขเดิม
- e_2 : (ยังไม่สมรส) และ (เป็นคู่สมรส) : เงื่อนไขเดิม
- e_3 : (อายุไม่เกิน 14 ปี) และ (เป็นคู่สมรส)

วิธีสร้างเงื่อนไขเพิ่มเติม (logical implied edit) ตามวิธีนี้เรียกว่า Heuristic Approach เป็นวิธีที่สามารถใช้ประโยชน์ได้ดีแต่อาจทำให้ไม่อาจสร้างเงื่อนไขเพิ่มเติมได้ครบถ้วนและไม่ได้ Complete Set of Edit ได้ตามที่ต้องการ วิธีหนึ่งที่สามารถสนองความต้องการนี้ได้ครบถ้วนก็คือการใช้ Lemma ต่อไปนี้

Lemma ถ้า $e_r ; r \in s$ เป็นเงื่อนไขการบรรณาธิกรณ (edit) โดยที่ s คือดัชนี แสดงจำนวนเงื่อนไขในเซตของเงื่อนไขการบรรณาธิกรณ (index set) กล่าวคือ

$$e_r = \bigcap_{j=1}^n a_j^r \text{ for all } r \in s \text{ แล้ว}$$

สำหรับ i ใด ๆ ที่ $1 \leq i \leq n$ โดยที่ n คือจำนวนรายการข้อมูล (field) เราสามารถสร้าง implied edit e^* ขึ้นมาได้โดย

$$e^* = \bigcap_{i=1}^n A_i^* = F$$

เมื่อ A_i^* ไม่เป็นเซตว่างเปล่า (Empty set) โดยที่

$$A_j^* = \bigcap_{r \in S} A_j^r, j = 1, 2, \dots, n; j \neq i$$

เมื่อ A_j^r คือ field ที่ j ของเงื่อนไขที่ r ; $j \neq i$

$$A_i^* = \bigcup_{r \in S} A_i^r \text{ เมื่อ } A_i^r \text{ คือ field ที่ } i \text{ ของเงื่อนไขที่ } r$$

หมายเหตุ 1. e_r (ess) หมายความว่า e_r อาจเป็นอนุเซตของเงื่อนไขการบรรณาธิกรณ เช่น ถ้าเรามีเงื่อนไขการบรรณาธิกรณ 10 เงื่อนไข e_r อาจหมายถึงเพียง 2, 3, 4 เงื่อนไขก็ได้

2. implied edit ที่สามารถใช้ร่วมเป็นเงื่อนไขสำหรับสร้าง implied edit อื่น ๆ ต่อไปก็ได้ เงื่อนไขที่สร้างเพิ่มเติมขึ้นมาโดยอาศัยเฉพาะ implied edit เรียกว่า Contributing edit

ขอยกตัวอย่างเพื่อให้เข้าใจ Lemma ข้างต้น โดยอาศัยเงื่อนไขจากตัวอย่างที่ผ่านมาดังนี้

คือ $e_1 = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{สถานภาพสมรส} = \text{สมรสแล้ว})$

$e_2 = (\text{สถานภาพสมรส} = \text{ยังไม่สมรส}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{ครัว} = \text{เป็นคู่สมรส})$

จาก e_1 และ e_2 ถ้าเราเพิ่มรายการ สถานภาพสมรสเข้า e_1 และรายการอายุเข้า e_2 เพื่อให้ครบทุกรายการข้อมูล จะพบว่า สถานภาพสมรสและอายุใช้รหัสใดก็ได้

ดังนั้น $e_1 = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{รหัสใด ๆ}) \cap (\text{สถานภาพสมรส} = \text{สมรสแล้ว})$

$e_2 = (\text{อายุ} = \text{รหัสใด ๆ}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{คู่สมรส} = \text{ยังไม่สมรส}) \cap (\text{สถานภาพสมรส} = \text{ยังไม่สมรส})$

1. รหัสใด ๆ - ทุกรหัสอาจใช้รหัสใดก็ได้เพราะ e_1 ก็ยังคงเป็น fail edit อยู่เช่นเดิม

จะเห็นว่าแต่ละเงื่อนไขประกอบด้วยรายการข้อมูล (field) 3 รายการ นั่นคือ $n = 3$

$$\text{ดังนั้นจาก } e_r = \prod_{j=1}^n A_j^r$$

$$\Rightarrow e_1 = \prod_{i=1}^3 A_i^1 \quad e_2 = \prod_{j=1}^3 A_j^2$$

ให้สถานภาพสมรสทำหน้าที่ในฐานะ field ที่ i (เรียก field ที่ i ; $1 < i < n$ ว่า Generating field) ในที่นี้จะเห็นว่าเราจัดให้สถานภาพสมรสไว้ในลำดับ 3 แสดงว่า $i=3$

$$\text{ดังนั้น จาก } A_i^* = \underset{\text{res}}{U} A_i^r$$

$$\Rightarrow A_3^* = \underset{\text{res}}{U} A_3^r = A_3^1 \underset{\text{res}}{U} A_3^2$$

= (สถานภาพสมรส = สมรสแล้ว) \cup (สถานภาพสมรสแล้ว = ยังไม่สมรส)

= (สถานภาพสมรส = ทุกะหัด = ะหัดใด ๆ)

$$\text{ดังนั้นสำหรับ field อื่นคือ } A_j^* = \underset{\text{res}}{\cap} A_j^r \quad j = 1, 2, \dots, n, j \neq i$$

$$\Rightarrow A_j^* = \underset{\text{res}}{U} A_j^r \quad j = 1, 2, ; j \neq 3$$

$$\Rightarrow A_1^* = \underset{\text{res}}{\cap} A_1^r = A_1^1 \underset{\text{res}}{\cap} A_1^2$$

= (อายุ = -14 ปี) \cap (อายุ = ะหัดใด ๆ = ทุกะหัด)

= (อายุ = 0-14 ปี)

$$A_2^* = \underset{\text{res}}{\cap} A_2^r = A_2^1 \underset{\text{res}}{\cap} A_2^2$$

A_2^* = (ความสัมพันธ์กับหัวหน้าครอบครัว = ะหัดใด ๆ) \cap (ความสัมพันธ์กับหัวหน้าครอบครัว = คู่สมรส)

= (ความสัมพันธ์กับหัวหน้าครอบครัว = คู่สมรส)

1. ะหัดใด ๆ = ทุกะหัดอาจใช้ระหัดก็ได้เพราะ e_i ก็จะต้องเป็น fail edit อยู่เช่นเดิม

ดังนั้น เราสามารถหา implied edit ได้โดยที่ implied edit คือ

$$\begin{aligned} e^* &= \bigcap_{j=1}^n A_j^* \\ &= A_1^* \cap A_2^* \cap A_3^* = F \\ &= (\text{อายุ} = 0-14 \text{ ปี}) \cup (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{คู่สมรส}) \\ &\quad \cup (\text{สถานภาพสมรส} = \text{ระหัสใด ๆ}) \end{aligned}$$

หรือ $e^* = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{เป็นคู่สมรส})$ ไม่จำเป็นต้องมีรายงานสถานภาพสมรส เพราะไม่ว่าจะใช้ระหัสใด ๆ e^* ก็คง fail edit อยู่เช่นเดิม

ขอให้สังเกตว่า e^* ที่ได้จาก Lemma นี้ก็คือเงื่อนไขเดียวกับ e , ซึ่งได้จากวิธี

Heuristic Approach

สำหรับการสร้างเงื่อนไขเพิ่มเติมอื่น ๆ (รวมทั้งการนำเงื่อนไขเพิ่มเติมอื่น ๆ ที่สร้างขึ้นร่วมเป็นเงื่อนไขเดิม (initial edit) กระทำได้โดยง่ายเพียงแต่เปลี่ยนค่าของ i ไปเรื่อย ๆ จาก $i=1, 2, \dots, n$ (ในที่นี้ $i = 1, 2, 3$) หรือนัยหนึ่งเปลี่ยน (Generating field) ไปเรื่อย ๆ ก็จะได้เงื่อนไขเพิ่มเติมตามที่ต้องการแต่สิ่งที่พึงระวังก็คือ บางครั้งเงื่อนไขเพิ่มเติม (implied edit) ที่ได้อาจไม่เหมาะสมหรือไม่จำเป็น กรณีเช่นนี้ให้ตัดเงื่อนไขดังกล่าวทิ้งเหลือไว้เฉพาะเงื่อนไขที่จำเป็นจริง ๆ (essentially new implied edit)

ค.3 การสร้างข้อมูลทดแทน

การสร้างข้อมูลทดแทนคือการเปลี่ยนค่าของรายการข้อมูลใน record ที่ตรวจพบว่าบกพร่องไปสู่ค่าอื่นที่เป็นไปได้ของรายการเดียวกัน เช่นตรวจพบว่ารายการเพศ (ที่ผู้ตอบตอบว่าเพศชาย) ผิดพลาดข้อมูลทดแทนของรายการนี้ก็คือเพศหญิง ปัญหาในทางปฏิบัติก็คือจะทราบได้อย่างไรว่ารายการข้อมูล (field) ใดของ record ผิด? และพึงแก้ไขสร้างข้อมูลทดแทนในลักษณะใดจึงจะสะดวกรวดเร็ว?

วิธีสร้างข้อมูลทดแทนมี 2 วิธีอื่น Sequential Impulation และ Joint Imputation ซึ่งจะกล่าวถึงหลักการและตัวอย่างประกอบให้เข้าใจในลำดับต่อไปนี้ ส่วนวิธีนำไปใช้ประโยชน์จริงในทางปฏิบัติจะกล่าวถึงในตอน ค.4

1. อ่าน I.P. Fellegi and D. Holt "A Systematic Approach to Automatic Edit and Impulation (Jour. of Amer. Stat. Assoc., March 1976, V.71, No, 393) p.23-29.

วิธีที่ 1 Sequential Imputation (SI)

คือวิธีการสร้างข้อมูลทดแทนให้แก่ field ที่ K ก่อน หลังจากนั้นจึงค่อย ๆ ย้อนสร้างข้อมูลทดแทนให้แก่ field อื่น ๆ คือ field ที่ K-1, K-2, ..., 3, 2, 1 ตามลำดับให้

A_k = เซตของรหัสคำตอบที่เป็นไปได้ทั้งหมด (all possible code) ของ field ที่ K

A_k^r = เซตหรืออนุเซตรหัสคำตอบของ field ที่ K ที่ปรากฏในเงื่อนไขที่ r

\overline{A}_k^r = ส่วนเติมเต็ม (Complement) ของรหัสคำตอบของ field ที่ K สำหรับเงื่อนไขที่ ra_i^o ($i = 1, 2, \dots, n$)

= รหัสคำตอบของ field ที่ i จาก record a ใด ๆ

A_i^r = เซตหรืออนุเซตรหัสคำตอบของ field ที่ i ที่ปรากฏในเงื่อนไขที่ r
 $i = 1, 2, \dots, n$

a_k^* = ข้อมูลทดแทนของ field ที่ K

สมมติว่ามีเงื่อนไขการตรวจสอบ (edit) อยู่ทั้งสิ้น M เงื่อนไข (เป็น Complete Set) ที่ field ที่ K เกี่ยวข้องอยู่ แต่ field ที่ 1, 2, ..., n, K-1 ไม่เกี่ยวข้อง ซึ่งก็คือเงื่อนไขต่อไปนี้

$$e_r = \bigcup_{i=K}^n A_i = F ; r = 1, 2, \dots, M^1$$

$$1. r = 1 \rightarrow e_1 = A_k^1 \cup A_{K+1}^1 \cup A_{K+2}^1 \cup \dots \cup A_n^1$$

$$r = 2 \rightarrow e_2 = A_k^2 \cup A_{K+1}^2 \cup A_{K+2}^2 \cup \dots \cup A_n^2$$

$$r = M \rightarrow e_M = A_k^M \cup A_{K+1}^M \cup A_{K+2}^M \cup \dots \cup A_n^M$$

ข้อสังเกต จะเห็นว่า field ที่ K จะเกี่ยวข้องอยู่กับทุกเงื่อนไข ดังนั้น A_k^r (ซึ่งหมายถึงค่าหรือรหัสหนึ่งของ field ที่ K ในเงื่อนไขที่ r) จึงมีเซตของ A_k หากเป็นเพียงค่าหนึ่งของ A_k เท่านั้น

ทั้งนี้เราจะตัดเงื่อนไขทิ้งไปได้บ้าง ถ้า $a_i^0 \in A_i^r$ สำหรับเฉพาะบาง field i $i=K+1, K+2, \dots, N$ และเหลือไว้พิจารณาเฉพาะ M' เงื่อนไขเท่านั้นที่ $a_i^0 \in A_i^r$ สำหรับทุก field i ; $i=K+1, K+2, \dots, N$; $r=1, 2, \dots, M'$ ทั้งนี้เพราะเงื่อนไขเหล่านั้นจะไม่ให้ประโยชน์อะไรหรือนัยหนึ่งไม่ว่าจะแปรค่าของ field K ไปสู่ค่า (ข้อมูลทดแทน) ใด เงื่อนไขเหล่านั้นก็ไม่เป็นเงื่อนไขที่ค่าของ record จะสอดคล้องได้

ดังนั้น เงื่อนไขที่เหลืออยู่จึงเพียงพอแล้ว และค่าใน record จะสอดคล้องหรือไม่ย่อมขึ้นอยู่กับค่า field ที่ K

ถ้าเราปรารถนาจะสร้างข้อมูลทดแทนให้แก่ field ที่ K ค่าที่ได้ (a_K^*) จะต้องเป็นสมาชิกของเซต

$$\bigcap_{r=1}^{M'} \overline{A}_K^r$$

หรือ $a_K^* \in \bigcap_{r=1}^{M'} \overline{A}_K^r$

1. $a_i^0 \in A_i^r$ สำหรับบาง field (อย่างน้อย 1 field) $i=K+1, K+2, \dots, n$ หมายความว่า เมื่อพิจารณารหัสคำตอบของ record a ใน field ที่ $K+1, K+2, \dots, n$ ของเงื่อนไขที่ $1, 2, \dots, M$ ถ้ารหัสคำตอบใน field เหล่านี้แม้เพียง field เดียวไม่ตรงกับรหัสคำตอบใน field เดียวกันของเงื่อนไขในบรรดาเงื่อนไขทั้งหมด M เงื่อนไข (เวลาพิจารณาให้ค่อยๆ พิจารณาตรวจจรวาดละเงื่อนไข) ให้ตัดเงื่อนไขนั้นทิ้ง เพราะถือว่าเป็นเงื่อนไขที่ไม่มีประโยชน์แก่การสร้างข้อมูลทดแทน จำนวนเงื่อนไขที่เหลืออยู่ M' เงื่อนไขจึงเป็นเงื่อนไขที่จำเป็นและให้ประโยชน์อย่างแท้จริง

$a_i^0 \in A_i^r$; $i = K+1, K+2, \dots, n$, $r = 1, 2, \dots, n$ หมายความว่ารหัสคำตอบใน field ที่ $K+1, K+2, \dots, n$ ของ record เป็นรหัสเดียวกัน (หรือเป็นรหัสหนึ่งในเซตของรหัส) กับรหัสคำตอบของ field เดียวกัน ที่ปรากฏอยู่ในเงื่อนไขที่ $1, 2, \dots, M$

2. ถ้า A_K^r เป็น empty set e_r จะเท่ากับ $\bigcap_{i=K+1}^n A_i^r$; $r=1, 2, \dots, M'$

A_{K+1}^r เป็น empty set e_r จะเท่ากับ $A_K^r \cap (\bigcap_{i=K+2}^n A_i^r)$ ดังนี้เรื่อยไป

สมมติว่าแบบสอบถามมีรายการข้อมูลที่ต้องแก้ไขอยู่ 5 รายการ แต่ละรายการที่รหัสคำตอบเป็นไปได้นี้

รายการข้อมูล	รหัสคำตอบ
เพศ	ชาย, หญิง
อายุ	5-14, 15-16, 17+
สถานภาพสมรส	โสด, สมรส, หย่า, แยกกันอยู่, ฆ่า
ความสัมพันธ์กับหัวหน้าครอบครัว	ภรรยา,สามี, บุตรหรือธิดา, อื่น ๆ
การศึกษา	ไม่มีการศึกษา, ประถมศึกษา, มัธยมศึกษา อุดมศึกษา

และเรามีเงื่อนไขการตรวจสอบอยู่ 5 เงื่อนไข

$$e_1 = (\text{เพศ} = \text{ชาย}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{ภรรยา}) = F$$

$$e_2 = (\text{อายุ} = 5-14 \text{ ปี}) \cap (\text{สถานภาพสมรส} = \text{สมรสแล้ว}) = F$$

$$e_3 = (\text{สถานภาพสมรส} = \text{ยังไม่สมรส}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{คู่สมรส}) = F$$

$$e_4 = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{คู่สมรส}) = F$$

$$e_5 = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{การศึกษา} = \text{อุดมศึกษา}) = F$$

สมมติจากคำตอบ (record) หนึ่งผู้ตอบบันทึกมาดังนี้

field ที่ 1 = เพศ ชาย

field ที่ 2 = อายุ 12 ปี

field ที่ 3 = สถานภาพสมรส = สมรสแล้วและอยู่ด้วยกัน

field ที่ 4 = ความสัมพันธ์กับหัวหน้าครอบครัว = ภรรยา

field ที่ 5 = การศึกษา = ประถมศึกษา

เมื่อเขียนคำตอบกับเงื่อนไขคราวละเงื่อนไขจะพบว่าคำตอบนี้สอดคล้องกับเงื่อนไข 3 เงื่อนไข e_1, e_2 และ e_4 คือ

$$e_1 = (\text{เพศ} = \text{ชาย}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{ภรรยา})$$

$$e_2 = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{สถานภาพสมรส} = \text{สมรสแล้ว})$$

$$e_4 = (\text{อายุ} = 0-14 \text{ ปี}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{คู่สมรส})$$

ลองพิจารณาเงื่อนไขทั้งสาม (e_1, e_2, e_3) จะไม่พบว่ามี field ใดเป็น field ร่วมที่ปรากฏอยู่ในทั้งสามเงื่อนไข¹ จึงจำเป็นต้องพิจารณาคราวละ 2 field² ว่า field ใดน่าจะผิดปกติ วิธีพิจารณาระทำดังนี้

(1) คูที่ e_1 ถ้าให้รายการ “เพศ” ผิดก็แสดงว่า “ความสัมพันธ์กับหัวหน้าครอบครัว” ถูกต้องเมื่อรายการ “ความสัมพันธ์กับหัวหน้าครอบครัว” ถูกต้อง ก็แสดงว่ารายการ “อายุ” (คูที่ e_2) ผิดปกติ

ข้อสรุปที่ 1 คือน่าเชื่อว่า รายการ “เพศ” และ “อายุ” ของ record ผิดปกติ

(2) คูที่ e_2 ถ้าให้รายการ “ความสัมพันธ์กับหัวหน้าครอบครัว” ผิดปกติ ก็แสดงว่ารายการ “เพศ” ถูกต้อง ขณะเดียวกัน (คูที่ e_3) รายการ “อายุ” ก็ถูกต้องด้วย เมื่อรายการ “อายุ” ถูกต้อง (คูที่ e_2) ก็แสดงว่ารายการ “สถานภาพสมรส” ผิดปกติ

ข้อสรุปที่ 2 คือน่าเชื่อว่ารายการ “ความสัมพันธ์กับหัวหน้าครอบครัว” และ “สถานภาพสมรส” ของ record ผิดปกติ

(3) จากข้อสรุปที่ 1 คือรายการ “เพศ” และ “อายุ” ผิดปกตินั้น ถ้าหากถือว่าเฉพาะรายการ “อายุ” เท่านั้นที่ผิดปกติ ก็แสดงว่ารายการ “เพศ” ถูกต้อง เมื่อรายการ “เพศ” ถูกต้อง (คูที่ e_1) ก็แสดงว่ารายการ “ความสัมพันธ์กับหัวหน้าครอบครัว” ผิดปกติ

ข้อสรุปที่ 3 คือน่าเชื่อว่ารายการ “อายุ” และ “ความสัมพันธ์กับหัวหน้าครอบครัว” ของ record ผิดปกติ

ขณะนี้เราจึงมีจุดที่น่าจะหาทางแก้ไข 3 ทางคือ

1. เพศและอายุ
2. อายุและสถานภาพสมรส
3. สถานภาพสมรสและความสัมพันธ์กับหัวหน้าครอบครัว

ผู้ตรวจแก้ไขจะเริ่มต้นทางใดก็ได้

1. ถ้ามี field ร่วมก็พอจะเชื่อได้ว่า field นั้นคงเป็นตัวการของความผิดปกติให้พยายามแก้ที่ field นั้นโดยแปลงค่ารหัสคำตอบไปเรื่อย ๆ จนพบรหัสที่ทำให้คำตอบของ record ไม่สอดคล้องกับเงื่อนไขใดเลย ซึ่งหมายความว่า record นั้นผ่านการแก้ไขและสร้างข้อมูลทดแทนได้แล้ว

2. เพื่อความสะดวกเราควรหลีกเลี่ยงการตรวจสอบคราวละหลาย field ถ้าทำได้

ในที่นี้เลือกทางที่ 1

การเลือกทางที่ 1 ก็คือการสร้างข้อมูลทดแทนให้แก่รายการ “เพศ” และ “อายุ”
แก่ record ซึ่งมีอยู่ทั้งสิ้น 5 field (ที่สนใจ)

ให้ field ที่ 1 เพศ

field ที่ 2 อายุ

field ที่ 3 สถานภาพสมรส

field ที่ 4 ความสัมพันธ์กับหัวหน้าครอบครัว

field ที่ 5 การศึกษา

เราจะเริ่มต้นสร้างข้อมูลทดแทนให้แก่รายการ “อายุ” หรือ field ที่ 2 ก่อนจะ
พบว่า $K=2$ ¹

เมื่อพิจารณาทั้ง 4 คือ e_1, e_2, e_3, e_4, e_5 จะพบว่า field ที่ $K=2$ เกี่ยวข้องอยู่ทั้ง
สิ้น $M=3$ เงื่อนไข (แต่ field $K=1$ เพศ ไม่เกี่ยวข้อง) คือ e_2, e_4 และ e_5

$e_2 =$ (อายุ 14 ปี) \cap (สถานภาพสมรส สมรสแล้ว)

$e_4 =$ (อายุ 0-14 ปี) \cap (ความสัมพันธ์กับหัวหน้าครอบครัว คู่สมรส)

$e_5 =$ (อายุ 0-14 ปี) \cap (การศึกษา - อุดมศึกษา)

พิจารณาเฉพาะ e_5 เปรียบเทียบกับ record ผู้ตอบ ๆ ว่า มีการศึกษาระดับประถม
ศึกษา (มีใช้อุดมศึกษา) หรือ $a_i^* \in A_i^*$ เมื่อ a_i^* รหัสคำตอบใน field ที่ 5 ของ record a
ในที่นี้ $a_i^* =$ (การศึกษา ประถมศึกษา) $A_i^* =$ รหัส หรือเซตของรหัสคำตอบของ field
ที่ 5 ปรากฏอยู่ในเงื่อนไขที่ 5 (ในที่นี้ $A_i^* = \{\text{การศึกษา} = \text{อุดมศึกษา}\}$) ซึ่งเมื่อเป็นเช่นนี้ก็
แสดงว่า ไม่ว่าจะแปลงค่าอายุ (ข้อมูลทดแทนเกี่ยวกับอายุ) ของ record ไปเช่นลักษณะใด
 e_5 ก็ยังมีใช่เป็นเงื่อนไขที่ record จะสอดคล้องได้ จึงตัด e_5 ออก

ดังนั้น จึงมีเงื่อนไขอยู่เพียง $M' = 2$ เงื่อนไขเท่านั้นที่รายการ “อายุ” เกี่ยวข้อง
อยู่ และเป็นเงื่อนไขที่จำเป็นจริง ๆ คือ e_2 และ e_4

ดังนั้นข้อมูลทดแทนของรายการ “อายุ” หรือรายการ $K=2$ คือ โดยที่

$$a_2^* \in \bigcap_{r=1}^{M'} \bar{A}_r^*$$

1. เปรียบเทียบกับหลักเกณฑ์ของ SI

$$\bigcap_{r=1}^{M'=2} \bar{A}_2^r = \bar{A}_2^2 \cap A_2^{-4}$$

$$A_2 = \{0-14 \text{ ปี}, 15-16 \text{ ปี}, 17^+ \text{ ปี}\} ; A_2^2 = \{0-14 \text{ ปี}\} \quad A_2^{-2} = \{15^+ \text{ ปี}\}$$

และ $A_2^4 = \{0-14 \text{ ปี}\} ; A_2^{-4} = \{15^+ \text{ ปี}\}$

ดังนั้น $\bigcap_{r=2}^{m'=2} A_2^r = \{15^+ \text{ ปี}\} \cap \{15^+ \text{ ปี}\}$
 $= \{15^+ \text{ ปี}\}$

ดังนั้นข้อมูลทดแทนสำหรับ “อายุ” ใน record คือ อายุ 15 ปีขึ้นไป

ขั้นต่อไปก็คือสร้างข้อมูลทดแทนให้แก่ field ที่ $K-1 = 2 - 1 = 1 = \text{เพศ}$
 จากการตรวจสอบเพียง e_i เท่านั้นที่เกี่ยวข้องกับเพศ (หรือ field ที่ $K-1$) คือ
 $e_i = (\text{เพศชาย}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัวและภรรยา})$

→ $M=M' = 1$

ดังนั้นข้อมูลทดแทนสำหรับรายการ “เพศ” หรือ field ที่ $K-1$ คือ a_i^* คือ

$$a_i^* \in A_i^{-1}$$

เมื่อ $A_i = \{\text{ชาย}, \text{หญิง}\}$

$$A_i^{-1} = \{\text{ชาย}\} \quad \text{ดังนั้น } A_i^{-1} = \{\text{หญิง}\}$$

ดังนั้นข้อมูลทดแทนของรายการ “เพศ” คือ a_i^* ” “เพศหญิง”

เมื่อแก้ไขและสร้างข้อมูลทดแทนให้แก่รายการ “อายุ” และ “เพศ” แล้ว record เดิมจึงได้รับการแก้ไขเป็น

(หญิง, อายุตั้งแต่ 15 ปีขึ้นไป, สมรสแล้ว เป็นภรรยา, การศึกษาระดับประถมศึกษา)

ในกรณีนี้ $a_K^* \in \bigcap_{r=1}^{M'} \bar{A}_K^r$ เมื่อ $\bigcap_{r=1}^{M'} \bar{A}_K^r$ เป็นอนุเซตของรหัสคำตอบของ field

1. อาศัย 1-1 mapping $r-1$ หมายถึง e ; $r-2$ หมายถึง e . ทั้งนี้ขึ้นอยู่กับเงื่อนไขที่เลือกอยู่

ที่ K จะเกิดปัญหาในการเลือกรหัสคำตอบ วิธีปฏิบัติสำหรับปัญหาที่สามารถปฏิบัติได้ 2 วิธีคือ

1. สุ่มรหัสคำตอบจากอนุเซตดังกล่าว และถือว่ารหัสที่สุ่มมานั้นคือค่าของข้อมูลทดแทน a_k^*

เช่น ในกรณีตัวอย่าง $a_2^* \in$ (อายุ 15-16 ปี, 17 + ปี) หรือแม้แต่ในกรณีที่รายการ “อายุ” จะจำแนกอย่างละเอียดเป็นช่วง ๆ ละ 5 ปี 10 ปี เช่น

$$a_2^* \in \{15-19, 20-24, 25-29, \dots, 35-39, 60 +\}$$

การเลือกรหัสคำตอบให้ใช้วิธีสุ่ม¹ รหัสคำตอบจากเซต {อายุ 15-16 ปี, 17 + ปี} หรือ {15-19, 20-24, 25-29, ..., 45-50, 60 +} รหัสที่สุ่มได้ถือว่าเป็นข้อมูลทดแทน a_k^*

2. ถือว่ารหัสคำตอบของ field ที่ K ใน record อื่นที่ถูกต้อง (ไม่มี field ใดสอดคล้องกับเงื่อนไข) ที่มีค่าของข้อมูล field อื่น (ที่ไม่ใช่ field ที่ K) เป็นค่าเดียวหรือใกล้เคียงกับ record a เป็นค่าของข้อมูลทดแทน โดยที่ค่าของ field ที่ K ใน record นั้นต้องเป็นสมาชิกของอนุเซต $\prod_{r=1}^{M'} \bar{A}_K^r$ และในกรณีที่มี record ให้ใช้วิธีสุ่มรหัสของ field ที่ K จาก record เหล่านั้น

วิธีที่ 2 Joint Imputation (JI)

เนื่องจากวิธี SI มีจุดอ่อนที่สำคัญอยู่ 2 ประการคือ

1. SI มิได้อาศัยข้อสนเทศอื่น ๆ ใน record มาใช้อย่างเต็มที่ทั้ง ๆ ที่บางครั้ง ข้อสนเทศเหล่านี้อาจเกี่ยวข้องหรือมีสหสัมพันธ์ต่อกัน ซึ่งกรณีนี้อาจจะละลายไปได้ถ้าค่าของข้อมูลทดแทนแบบนี้ (Consistent) กับข้อมูลในรายการอื่น ๆ ของ record เดียวกัน

2. SI คำนวณหาค่าของข้อมูลทดแทนคราวละ 1 ค่า เป็นลำดับ ๆ ต่อเนื่องกันไป เมื่อเป็นเช่นนี้ Joint pdf ระหว่างข้อมูลทดแทนกับ field อื่นจะแตกต่างกันไป (ยกเว้นกรณีที่ทุก field เป็นอิสระต่อกัน) แม้ว่า Marginal pdf จะคงที่เช่นเดิมก็ตาม

1. อาจใช้ตารางเลขสุ่ม (Random digit); Japanese dice, ฉลาก Random Generator หรืออื่น ๆ

ด้วยเหตุนี้จำเป็นต้องมีวิธีสร้างข้อมูลทดแทนที่รัดกุมขึ้นด้วยวิธี JI ซึ่งอาศัย joint pdf ของ field ทั้งนี้จำกัดการสร้างข้อมูลทดแทนเฉพาะกับ record ทั้งหลายที่มีข้อมูลในรายการอื่น (นอกเหนือจาก field ที่จะสร้างข้อมูลทดแทน) เป็นค่าเดียวกัน (หรือคล้ายคลึงกัน) กับ record ที่กำลังจะสร้างข้อมูลทดแทนให้

หลักเกณฑ์ของ JI ปรากฏดังนี้

จากการศึกษาพิจารณาทุก ๆ เงื่อนไขสมมุติมีเพียงเฉพาะ field ที่ 1 ถึง K เท่านั้นที่จะต้องสร้างข้อมูลทดแทน และมีเงื่อนไขอยู่เพียง M'' เงื่อนไขเท่านั้นที่ record สอดคล้องคือ

$$e_r = \prod_{i=1}^n A_i^r$$

โดยที่ $a_{ij} \in A_i^r$ สำหรับทุก field i ; $i = K+1, K+2, \dots, n$ และ $r = 1, 2, \dots, M''$ ¹.

$$\text{พิจารณาเซต } A_i^* = \prod_{r=1}^{M''} A_i^r \quad ; \quad i = K+1, K+2, \dots, n$$

ซึ่ง A_i^* จะไม่เป็น empty set เพราะ $a_{ij} \in A_i^r$; $r = 1, 2, \dots, M''$ ถ้าเราเลือก record ซึ่งผ่านกระบวนการบรรณาธิการแล้ว แสดงว่า record นี้จะไม่สอดคล้องกับเงื่อนไขใด และพบว่ารหัสคำตอบของ field ที่ $K+1, K+2, \dots, n$ ตกอยู่ในเซต A_i^* แล้วรหัสคำตอบของ record ดังกล่าวใน field ที่ $1, 2, \dots, K$ ของ record ที่ผิดปกติ (สอดคล้องกับเงื่อนไข)

ข้อสังเกต วิธี JI ไม่จำเป็นต้องคำนวณหาค่าที่เป็นไปได้ของ field ที่ต้องการแก้ไขเช่นเดียวกับ SI เพียงแต่เราสร้างเซตของ field อื่นขึ้นมา เซตดังกล่าวจะเป็นตัวกำหนดช่วงหรือขอบเขต (range) ของ field ที่ $K+1, K+2, \dots, n$ ซึ่งถ้าหากรหัสคำตอบของ record ที่เราถือว่าถูกต้องใน field ที่ $K+1, K+2, \dots, n$ ตกอยู่ในขอบเขตเหล่านี้เราก็ถือว่า

1. ดูคำอธิบายในเชิงอรรถที่ 2 ท้ายวิธี 1

รหัสคำตอบของ record ดังกล่าวใน field ที่ 1, 2, ..., K เป็นค่าที่เหมาะสมจะใช้เป็นข้อมูลทดแทนของ field ที่ 1, 2, ..., K ของ record ที่ผิดปกติ

ลองพิจารณาตัวอย่างต่อไปนี้เป็นตัวอย่างที่ได้แสดงให้เห็นด้วยวิธี SI

เมื่อเราเชื่อว่ารายการ “เพศ” (K=1) และอายุ (K=2) น่าจะเป็นรายการที่ผิดปกติ และควรสร้างข้อมูลทดแทนให้แก่รายการทั้งสอง ซึ่งเมื่อพิจารณาแล้วพบว่าเมื่อเงื่อนไขการบรรณาธิกรณ 4 เงื่อนไขที่รายการ “เพศ” และ “อายุ” เกี่ยวข้องอยู่คือ e_1, e_2, e_4 , และ e_5 แต่ record นี้จะไม่สอดคล้องกับ e_5 ไม่ว่าจะแปลงรหัสคำตอบของ “เพศ” และ “อายุ” ไปในลักษณะใด เพราะผู้ตอบมีรหัสคำตอบในรายการ “การศึกษา” เป็นประถมศึกษา มีไฮ่อคมศึกษา จึงตัด e_5 ทิ้ง ดังนั้นเงื่อนไขควรพิจารณาในการแก้ไขรายการ “เพศ” และอายุ จึงมีเพียง $M=3$ เงื่อนไขเท่านั้นคือ e_1, e_2 และ e_4 เมื่อ

$$e_1 = (\text{เพศ} = \text{ชาย}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{ภรรยา})$$

$$e_2 = (\text{อายุ} = 0 = 14 \text{ ปี}) \cap (\text{สถานภาพสมรส} = \text{สมรสแล้ว})$$

$$e_4 = (\text{อายุ} = 0 = 14 \text{ ปี}) \cap (\text{ความสัมพันธ์กับหัวหน้าครอบครัว} = \text{คู่สมรส})$$

ดังนั้น

จาก

$$A_i^* = \prod_{r=1}^M A_r^i ; i=K+1, K+2, \dots, n$$

$$i=K+1-2+1-3$$

$$\Rightarrow A_3^* = A_1^1 \cap A_2^1 \cap A_4^1 ; i=K+1, = 2 + 1 = 3$$

$$= (\text{สถานภาพสมรส} = \text{รหัสใด ๆ}) \cap (\text{สถานภาพสมรส} = \text{สมรสแล้ว}) \\ \cap (\text{สถานภาพสมรส} = \text{รหัสใด ๆ})$$

1. A_r^i – รหัสคำตอบของ field ที่ i ที่ปรากฏในเงื่อนไข r

2. A_r^i – รหัสคำตอบของ field ที่ 3 (field ที่ 3 คือสถานภาพสมรส) ที่ปรากฏในเงื่อนไขที่ 1 แต่ในเงื่อนไขที่ 1 ไม่ปรากฏ รายการสถานภาพสมรส ในเงื่อนไขนี้อาจเป็นรหัสใดก็ได้ จึงใช้คำว่า “รหัสใด ๆ”

= (สถานภาพสมรส = สมรสแล้ว)

$$A_4^* = A: \cap A: \cap A_4^* ; i = K+2 = 2 + 2 = 4$$

= (ความสัมพันธ์กับหัวหน้าครอบครัว = ภรรยา) \cap (ความสัมพันธ์กับหัวหน้าครอบครัว = ไร้สใด ๆ) \cap (ความสัมพันธ์กับหัวหน้าครอบครัว = คู่สมรส)¹

= (ความสัมพันธ์กับหัวหน้าครอบครัว = ภรรยา)

$$A: = A: \cap A: \cap A: ; i = K+3 = 2 + 2 = 5$$

= (การศึกษา = ไร้สใด ๆ) \cap (การศึกษา = ไร้สใด ๆ) \cap (การศึกษา = ไร้สใด ๆ)

= (การศึกษา = ไร้สใด ๆ)

ดังนั้น ถ้าพบ record ใดที่ถือว่าถูกต้องแล้ว (clean record) มีรายการ “สถานภาพสมรส” เป็นรหัสใดรหัสหนึ่งของเซต “สมรสแล้ว” มีรหัสของรายการ “ความสัมพันธ์กับหัวหน้าครอบครัว” เป็นภรรยาและมีรหัสของรายการ “การศึกษา” เป็นรหัสใดรหัสหนึ่งของเซต (ไม่มีการศึกษา, ประถมศึกษา, มัธยมศึกษา, อุดมศึกษา) เราก็จะใช้ข้อมูลในรายการ “เพศ” (field ที่ 1) และรายการ “อายุ” (field ที่ K=2) ของ record นี้เป็นข้อมูลทดแทนของรายการ “เพศ” และ “อายุ” สำหรับ record ที่ผิดปกติกคือ record ที่สอดคล้องกับเงื่อนไข) และถ้าทำการตรวจสอบติดต่อกันไป (Searching) พบว่ามี record ที่มี field ที่ 3, 4, 5 สอดคล้องกับเซต A_3^* , A_4^*

1. คู่สมรส – {สามี, ภรรยา}

2. A_i^* ต้องไม่เป็น empty set เพราะ $a_i^* \in A_i^*$; $r=1, 2, \dots, M$ ” หรือนัยหนึ่ง field ที่ i ของ record a จะต้องมีค่าตอบไม่รหัสใดรหัสหนึ่งเสมอ ถ้า A_i^* เป็น empty set จะทำให้ตีความได้ว่า ข้อมูลที่ i เป็นข้อมูลที่ไม่มีหรือไม่ต้องการคำตอบซึ่งคิดความเป็นจริง

และ A_i^* ก็ถือว่า รหัสคำตอบ field ที่ 1 และ 2 ของ record เหล่านี้ล้วนเป็นรหัสที่เหมาะสมที่จะใช้เป็นข้อมูลทดแทนให้แก่ field ที่ 1 และ 2 ของ record บกพร่องหรือผิดพลาดได้ วิธีปฏิบัติสำหรับกรณีนี้คือให้ใช้วิธีสุ่มตัวอย่างเป็นเครื่องมือช่วยตัดสินใจ กล่าวคือ สมมติมี record ที่ถูกต้องที่มี field ที่ $K+1, K+2, \dots, n$ สอดคล้องกับเซต $A_{K+1}^*, A_{K+2}^*, \dots, A_n^*$ อยู่ทั้งสิ้น X ชุด และพบว่า มี record ที่ผิดพลาดอยู่ Y ชุด การสร้างข้อมูลทดแทนใน field ที่ 1, 2, ..., K ให้แก่ record ทั้ง Y ชุดให้ดำเนินการคราวละ 1 ชุด แต่ละคราวให้สุ่มตัวอย่าง record จากชุดที่มีอยู่ X ชุดนั้นครั้งหนึ่ง ข้อมูลใน field ที่ 1, 2, ..., K ของ record ที่สุ่มได้จะใช้เป็นข้อมูลทดแทนให้แก่ field ที่ 1, 2, ..., K ของ record ที่ผิดพลาดให้ดำเนินการเช่นนี้เรื่อยไปจนกว่าจะแก้ไข record เสร็จสิ้นครบชุด Y ชุด

หมายเหตุ 1. วิธี JI สามารถนำข้อมูลสนเทศจาก field อื่นมาร่วมใช้เป็นเกณฑ์พิจารณาสร้างข้อมูลทดแทนให้แก่รายการที่สนใจได้ ถ้าหากว่าขอสสนเทศนั้นเอื้อต่อรายการที่สนใจ ทั้งนี้ขอสสนเทศดังกล่าวไม่จำเป็นต้องมีส่วนเกี่ยวข้องกับเงื่อนไขการบรรณาธิการ แต่จะต้องเกี่ยวข้องกับสัมพันธ์หรือมีสหพันธ์กับรายการที่สนใจ เช่น สถานภาพสมรสและความสัมพันธ์กับหัวหน้าครอบครัว เกี่ยวข้องสัมพันธ์กับอายุและเพศ ซึ่งถ้าไม่ทราบอายุและเพศของผู้ตอบแต่ทราบได้ว่าผู้ตอบเป็นชาย อายุอยู่ในวัยผู้ใหญ่ ขณะเดียวกัน การศึกษาก็เกี่ยวข้องกับสัมพันธ์กับอายุ ถ้าผู้ตอบตอบว่า จบชั้นอุดมศึกษา แสดงว่าผู้ตอบมีอายุอยู่ในวัยผู้ใหญ่ ถ้าตอบว่ากำลังเรียนอยู่ในระดับอุดมศึกษา แสดงว่าผู้ตอบมีอายุอยู่ระหว่าง 18-24 ปี ดังนี้ เป็นต้น

2. ถ้ามีรายการข้อมูลไม่มากนัก และแต่ละรายการมีรหัสคำตอบไม่มากเช่นกัน กรณีเช่นนี้เราอาจไม่ต้องสร้างช่วงหรือขอบเขต A_i^* ก็ได้ เราสามารถอาศัยความเกี่ยวข้องสัมพันธ์หรือรายการข้อมูลที่เอื้อต่อกัน ของรายการข้อมูลเป็นเครื่องมือสร้างข้อมูลทดแทนให้แก่รายการที่สนใจได้อย่างเพียงพอ

ค. 4 การสร้างข้อมูลทดแทนในทางปฏิบัติ

ในทางปฏิบัตินั้น เราจำเป็นจะต้องหลีกเลี่ยงความสลับซับซ้อนของวิธีการทางทฤษฎี และพยายามจัดวิธีปฏิบัติงานเสียใหม่ เพื่อให้ง่ายและเหมาะสมกับการนำเครื่องคอมพิวเตอร์เข้ามาช่วย

การนำคอมพิวเตอร์เข้ามาใช้งานบรรณารักษ์ โดยปกติจะเริ่มตั้งแต่ขั้นการสร้างเงื่อนไข การตรวจ record การสร้าง Complete Set of Edit การเสาะหารายการที่ผิดปกติ และการสร้างข้อมูลทดแทน

การนำคอมพิวเตอร์เข้ามาใช้งานบรรณารักษ์ ทำให้การปฏิบัติงานขั้นต่าง ๆ สะดวกรวดเร็วและถูกต้องแม่นยำขึ้น อย่างไรก็ตามเราจำเป็นต้องกำหนดเลขรหัสให้แก่ field ใน record และเงื่อนไขให้สอดคล้องกับ bit ในคอมพิวเตอร์ ดังนี้

ให้รหัสหมายเลข 1 แก่รหัสคำตอบที่ปรากฏในเงื่อนไขและใน record

รหัสหมายเลข 0 แก่รหัสคำตอบที่ไม่ปรากฏในเงื่อนไขและไม่ปรากฏใน record

เมื่อกำหนดหมายเลขรหัสแก่รหัสคำตอบหรือ field ต่าง ๆ เป็นไปในลักษณะนี้ ข้อมูลที่เสนออยู่ใน record และเงื่อนไขการบรรณารักษ์ จึงมีเฉพาะหมายเลข “0” และ “1” เท่านั้น การนำไปใช้ในทางปฏิบัติให้กระทำดังนี้

ค. 4.1 การตรวจสอบแบบสอบถาม (Record)

การตรวจสอบแบบสอบถามก็คือการตรวจสอบว่าแบบสอบถามที่สนใจสอดคล้องกับเงื่อนไขใดบ้างหรือว่า “ผ่าน” ทุกเงื่อนไข

การตรวจสอบ record กระทำได้โดยอาศัย Logical Edit Matrix Logical Edit Matrix คือเมตริกซ์ที่นิยามสมาชิกในแนวตั้ง (สดมภ์) และแนวนอนดังนี้

(1) สมาชิกในแนวตั้งคือ field ต่าง ๆ ของ record โดยจะเสนอไว้ทุก field แต่ละ field จะเสนอไว้ทุกรหัสคำตอบ

(2) สมาชิกในแนวนอนคือเงื่อนไขการบรรณารักษ์ทุกเงื่อนไข (Completed Set of Edit) ส่วนแถวสุดท้ายคือ record a ที่กำลังตรวจสอบ

(3) สมาชิกเมตริกซ์ค่าเท่ากับ 0 และ 1 เท่านั้นการดำเนินงานและเกณฑ์การตัดสินใจให้ปฏิบัติดังนี้

ก. จากเงื่อนไขที่อยู่ให้จัดเป็นรูปที่ประกอบไปด้วยทุก field ใน record นั่นคือ

$$e_r = \prod_{i=1}^n A_i^r ; r=1, 2, \dots, s$$
 เช่น เดิมเงื่อนไขที่ r ประกอบด้วยรหัสคำตอบจากเพียง 2 field แต่ record a มี field ที่น่าสนใจถึง 10 field อื่นในเงื่อนไขที่ r ทำให้เงื่อนไขที่ r ประกอบด้วยรหัสคำตอบจาก 10 field field ที่เพิ่มเข้ามาจะใช้รหัสคำตอบมีคำตอบเป็นรหัสใดก็ได้