การสร้างประชากรไทยสังเคราะห์ในระดับบุคคลและระดับครัวเรือน

นางสาวณัฐพร วัฒนสุทธิ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

GENERATING SYNTHETIC THAI POPULATION AT INDIVIDUAL AND HOUSEHOLD LEVELS

Miss Natthaporn Watthanasutthi

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2015

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การสร้างประชากรไทยสังเคราะห์ในระดับบุคคลและระดับครัวเรือน
โดย	นางสาวณัฐพร วัฒนสุทธิ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.วีระ เหมืองสิน

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.วีระ เหมืองสิน)

.....กรรมการ
(รองศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา)

.....กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.วรเศรษฐ สุวรรณิก)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ณัฐพร วัฒนสุทธิ : การสร้างประชากรไทยสังเคราะห์ในระดับบุคคลและระดับครัวเรือน (GENERATING SYNTHETIC THAI POPULATION AT INDIVIDUAL AND HOUSEHOLD LEVELS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.วีระ เหมืองสิน, 167 หน้า.

ในงานวิจัยนี้ผู้วิจัยได้นำเสนอกระบวนการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนที่เหมาะสมกับข้อมูลของประเทศไทยที่มีข้อจำกัด คือ ข้อมูลประชากรมีการเปิดเผยเพียงข้อมูลที่มาจากการสำมะโนประชากร, ข้อมูลที่เปิดเผยถูกเก็บมาจากหลายหน่วยงานซึ่งมีความละเอียดในการเก็บข้อมูลที่แตกต่างกัน และข้อมูลระดับครัวเรือนมีการเปิดเผยเพียงบางคุณลักษณะของข้อมูล เพื่อให้สามารถสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนได้อย่างน่าเชื่อถือและมีความใกล้เคียงกับประชากรจริงมากที่สุด และผู้วิจัยได้ประยุกต์ใช้กระบวนการนี้ในการสร้างประชากรสังเคราะห์ของจังหวัดพิษณุโลก เพื่อประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบ ผลลัพธ์แสดงให้เห็นว่ากระบวนการสร้างประชากรสังเคราะห์นี้มีประสิทธิภาพในการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนให้มีความใกล้เคียงกับประชากรจริง ซึ่งประชากรสังเคราะห์ที่ได้สามารถนำมาศึกษา วิเคราะห์และวางแผนนโยบายต่างๆ เกี่ยวกับประชากรได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2558

5670522721 : MAJOR COMPUTER ENGINEERING

KEYWORDS: SYNTHETIC POPULATION / MICROSIMULATION / SYNTHETIC RECONSTRUCTION

NATTHAPORN WATTHANASUTTHI: GENERATING SYNTHETIC THAI POPULATION AT INDIVIDUAL AND HOUSEHOLD LEVELS. ADVISOR: ASST. PROF. VEERA MUANGSIN, Ph.D., 167 pp.

Synthetic population represents the population in area of interest. We proposed a synthetic population generator at individual and household levels for Thai population. This generator solves the limitation of population data in Thailand: only aggregate census data are available, aggregate census data collected from several government agencies, and lack of data at household level. Moreover, this method takes many attributes of population data to identify various synthetic populations. We applied this method within Thailand's Phitsanulok Province as a case study. This case study analyzes the accuracy of a synthetic population. The results show that the synthetic population at individual and household levels is reliably close to the real population. Therefore, this synthetic population generator can generate synthetic population at individual and household levels effectively. The synthesis population can be used in the study, analysis, and planning policies about the population.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department: Computer Engineering Student's Signature .....
Field of Study: Computer Engineering Advisor's Signature .....
Academic Year: 2015

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาอย่างสูงจากผู้ช่วยศาสตราจารย์ ดร.วีระ เหมืองสิน อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่กรุณาสละเวลาช่วยเหลือในการให้ความรู้ คำปรึกษา คำแนะนำ รวมทั้งให้แนวคิดที่เป็นประโยชน์ในการทำวิจัย และช่วยชี้แนะการแก้ไขปัญหาต่างๆ ที่เกิดขึ้นระหว่างการทำวิจัย ขอกราบขอบพระคุณอาจารย์มา ณ โอกาสนี้ และขอขอบพระคุณคณาจารย์ทุกท่านที่แนะนำสั่งสอนและให้ความรู้แก่ข้าพเจ้าตลอดระยะเวลาการศึกษา

ขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่านเป็นอย่างสูง ผู้ช่วยศาสตราจารย์ ดร.ณัฐวุฒิ หนูไพโรจน์ ประธานคณะกรรมการสอบวิทยานิพนธ์, รองศาสตราจารย์ ดร.ทวิตีย์ เสนีวงศ์ ณ อยุธยา กรรมการสอบวิทยานิพนธ์ และรองศาสตราจารย์ ดร.วรเศรษฐ สุวรรณิก กรรมการสอบวิทยานิพนธ์ ที่ให้ความกรุณาตรวจสอบและให้ข้อเสนอแนะอันเป็นประโยชน์ในการปรับปรุงแก้ไข ทำให้วิทยานิพนธ์นี้มีความถูกต้องและสมบูรณ์มากยิ่งขึ้น

ขอขอบคุณทุนอุดหนุนการศึกษาระดับบัณฑิตศึกษาจากบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย เพื่อเฉลิมฉลองวโรกาสที่พระบาทสมเด็จพระเจ้าอยู่หัวภูมิพลอดุลยเดชทรงเจริญพระชนมายุครบ 72 พรรษา สำหรับการสนับสนุนค่าใช้จ่ายในระหว่างการศึกษาและการทำวิจัย

ขอกราบขอบพระคุณนายสุทัศน์ วัฒนสุทธิ, นางพวงทิพย์ วัฒนสุทธิ และนางธนวรรณ ปรีพูล บิดามารดาและพี่สาวของผู้วิจัยที่ให้การสนับสนุน และให้กำลังใจแก่ผู้วิจัยเสมอมา

สุดท้ายนี้ ผู้วิจัยขอขอบคุณทุกท่านที่มีส่วนช่วยเหลือจนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จเรียบร้อยลงได้ด้วยดีทุกประการ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญภาพ ..... ฐ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ 6

1.3 ขอบเขตของงานวิจัย 6

1.4 ประโยชน์ที่คาดว่าจะได้รับ 7

1.5 ขั้นตอนการดำเนินงานวิจัย 7

1.6 งานวิจัยที่ได้รับการตีพิมพ์ 8

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 9

2.1 ข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ 9

2.1.1 คุณลักษณะของข้อมูล 9

2.1.2 ประเภทของข้อมูล 10

2.1.2.1 จำแนกตามระดับของการวัด 10

2.1.2.2 จำแนกตามลักษณะของข้อมูล 11

2.1.3 การแจกแจงของข้อมูล (Probability Distribution) 12

2.1.3.1 การแจกแจงแบบปกติ (Normal Distribution) 12

2.1.3.2 การแจกแจงแบบปัวซอง (Poisson Distribution) 13

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.1.4 ชนิดของความสัมพันธ์ 13
2.2 กระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ 14
2.2.1 Combinatorial Optimization.....15
2.2.2 Synthetic Reconstruction 18
2.3 การประเมินความถูกต้องของแบบจำลอง (Model Validation).....20
2.4 เอกสารและงานวิจัยที่เกี่ยวข้อง.....21
2.4.1 ข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ 21
2.4.2 กระบวนการสร้างประชากรสังเคราะห์ 23
บทที่ 3 แนวคิดและวิธีการดำเนินงานวิจัย.....33
3.1 การเลือกและการจัดการข้อมูล.....34
3.1.1 ข้อมูลประชากรของไทย 34
3.1.2 การเลือกข้อมูล.....34
3.1.3 การจัดการข้อมูลให้เหมาะสม 39
3.2. การออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับบุคคล 43
3.3 การออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือน.....49
3.3.1 การสร้างหัวหน้าครัวเรือน.....54
3.3.2 การสร้างคู่สมรส 56
3.3.3 การสร้างบุตร.....57
3.4 การประเมินกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ 61
3.4.1 ประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย 62
3.4.2 ประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์.....62
3.4.3 ประเมินโดยใช้ตารางที่ถูกซ่อน.....63
3.5 การออกแบบกระบวนการจำลองสำหรับข้อมูลอื่นๆ 65

หน้า

บทที่ 4 การทดลองและผลการทดลอง 71
4.1 การประยุกต์ใช้กระบวนการสร้างประชากรสังเคราะห์กับข้อมูลของจังหวัดพิษณุโลก.....71
4.2 การออกแบบและพัฒนาโปรแกรมการสร้างประชากรสังเคราะห์ 74
4.2.1 เครื่องมือที่ใช้ในการพัฒนา 74
4.2.2 การออกแบบโปรแกรมสำหรับสร้างประชากรสังเคราะห์ 75
4.3 ผลของการประเมินกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ 76
4.3.1 ผลของการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย 76
4.3.2 ผลของการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์.....83
4.3.3 ผลของการประเมินโดยใช้ตารางที่ถูกซ่อน 91
บทที่ 5 บทสรุปและแนวทางในการพัฒนาต่อ.....95
5.1 บทสรุป 95
5.2 ข้อเสนอแนะและแนวทางในการพัฒนาต่อ 97
รายการอ้างอิง 98
ภาคผนวก..... 101
ประวัติผู้เขียนวิทยานิพนธ์ 167

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 ตัวอย่างรายงานสถิติจำนวนประชากรจำแนกตามเพศ ปี 2553.....1

ตารางที่ 2 ตัวอย่างรายงานสถิติจำนวนประชากรจำแนกตามอายุ ปี 2553.....2

ตารางที่ 3 ตัวอย่างจำนวนนักเรียนในระดับชั้นอนุบาลของจังหวัดพิษณุโลก ปี 2553.....2

ตารางที่ 4 ตัวอย่างของข้อมูลรายบุคคลซึ่งมีหลายคุณลักษณะ (disaggregated data).....3

ตารางที่ 5 ความน่าจะเป็นของประชากรในแต่ละกลุ่มอายุ (P(j)).....10

ตารางที่ 6 ความน่าจะเป็นที่ประชากรสถานภาพสมรสต่างๆในแต่ละกลุ่มอายุ (P(h | j)).....10

ตารางที่ 7 ความน่าจะเป็นของประชากรซึ่งทราบว่ามีอายุตามกลุ่มอายุ จะมีสถานภาพสมรสต่างๆ (P(h ∩ j)).....14

ตารางที่ 8 ความแตกต่างของกระบวนการสร้างประชากรสังเคราะห์.....20

ตารางที่ 9 ตัวอย่างการปรับปรุงการแจกแจง.....22

ตารางที่ 10 รายละเอียดของข้อมูลประชากรของปริมณฑลปักกิ่ง.....25

ตารางที่ 11 คุณลักษณะข้อมูลของระดับบุคคล.....27

ตารางที่ 12 คุณลักษณะข้อมูลของระดับครัวเรือน.....27

ตารางที่ 13 รายละเอียดของข้อมูลที่มีการเปิดเผยของประเทศไทย.....35

ตารางที่ 14 ตัวอย่างตารางจำนวนประชากรอายุตั้งแต่ 13-24 ปี จำแนกตามสถานภาพสมรส กลุ่มอายุ และเพศ.....36

ตารางที่ 15 ความละเอียดและลักษณะของข้อมูลจำแนกตามคุณลักษณะ.....38

ตารางที่ 16 จำนวนคนและจำนวนครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง.....39

ตารางที่ 17 การคำนวณจำนวนคนและจำนวนครัวเรือนในแต่ละขนาดครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง.....40

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 18 การปรับปรุงข้อมูลจำนวนคนและจำนวนครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง 41

ตารางที่ 19 การปรับปรุงข้อมูลของการคำนวณจำนวนคนและจำนวนครัวเรือนในแต่ละขนาดครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง 42

ตารางที่ 20 ลำดับการสร้างประชากรสังเคราะห์.....47

ตารางที่ 21 ลำดับของกระบวนการสร้างประชากรสังเคราะห์ระดับบุคคล.....47

ตารางที่ 22 ลำดับของกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือน.....51

ตารางที่ 23 ตัวอย่างการสร้างหัวหน้าครัวเรือน 55

ตารางที่ 24 ตัวอย่างการสร้างคู่สมรส 56

ตารางที่ 25 ตัวอย่างการเลือกเด็กโดยจำแนกตามการอยู่อาศัย.....58

ตารางที่ 26 ตัวอย่างการเลือกบุตรตามการอยู่อาศัย 58

ตารางที่ 27 ตัวอย่างการคำนวณจำนวนครัวเรือน.....61

ตารางที่ 28 ลำดับการสร้างข้อมูลในกระบวนการสร้างแบบจำลอง.....67

ตารางที่ 29 ลำดับการสร้างข้อมูลในกระบวนการสร้างแบบจำลอง.....70

ตารางที่ 30 ค่าของแต่ละคุณลักษณะของข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับบุคคลของจังหวัดพิษณุโลก.....72

ตารางที่ 31 ค่าของแต่ละคุณลักษณะของข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับครัวเรือนของจังหวัดพิษณุโลก.....73

ตารางที่ 32 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับบุคคล.....77

ตารางที่ 33 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับครัวเรือน.....78

ตารางที่ 34 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับบุคคล.....84

ตารางที่ 35 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับครัวเรือน.....85

ตารางที่ 36 ความคลาดเคลื่อนสัมบูรณ์ของตารางที่ถูกซ่อนของคุณลักษณะสถานภาพสมรสของเพศชาย 92

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 37 ความคลาดเคลื่อนสัมบูรณ์ของตารางที่ถูกซ่อนของคุณลักษณะสถานภาพสมรส ของเพศหญิง 92

สารบัญภาพ

หน้า

ภาพที่ 1 ลำดับการสร้างของคุณลักษณะข้อมูล 10
ภาพที่ 2 ลักษณะกราฟของการแจกแจงแบบปกติ 12
ภาพที่ 3 ลักษณะกราฟของการแจกแจงแบบปัวซอง 13
ภาพที่ 4 ตัวอย่าง Combinatorial Optimization 16
ภาพที่ 5 ตัวอย่าง synthetic reconstruction approach 19
ภาพที่ 6 ตัวอย่างกระบวนการสร้างประชากรอย่างง่ายสำหรับคุณลักษณะการมีชีวิตอยู่หรือเสียชีวิต 24
ภาพที่ 7 ความสัมพันธ์ของคุณลักษณะของข้อมูลประชากร 25
ภาพที่ 8 กระบวนการสร้างประชากรสังเคราะห์ 26
ภาพที่ 9 กระบวนการสร้างโครงสร้างของครัวเรือน 28
ภาพที่ 10 กระบวนการสร้างประชากรสังเคราะห์ในระดับครัวเรือน 29
ภาพที่ 11 กระบวนการที่ใช้ในการสร้างเมืองจำลอง 31
ภาพที่ 12 กระบวนการสร้างประชากรสังเคราะห์ 34
ภาพที่ 13 ขั้นตอนการทำงานของการสร้างประชากรสังเคราะห์ระดับบุคคล 44
ภาพที่ 14 ตัวอย่างความสัมพันธ์ของข้อมูลในรูปแบบตาราง 45
ภาพที่ 15 ความสัมพันธ์ของคุณลักษณะของข้อมูล 46
ภาพที่ 16 คุณลักษณะข้อมูลประชากรรายบุคคล 48
ภาพที่ 17 ขั้นตอนการทำงานของการสร้างประชากรสังเคราะห์ระดับครัวเรือน 50
ภาพที่ 18 คุณลักษณะข้อมูลครัวเรือน 51
ภาพที่ 19 กระบวนการสร้างโครงสร้างของครัวเรือน 52
ภาพที่ 20 กระบวนการสร้างคุณลักษณะของบุคคลในครัวเรือน 53

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาพที่ 21 กระบวนการสำหรับเลือกบุคคลเข้าในครัวเรือน 54
ภาพที่ 22 กระบวนการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย 62
ภาพที่ 23 กระบวนการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์.....63
ภาพที่ 24 ตัวอย่างการสร้างตารางที่ถูกซ่อน (hidden table).....64
ภาพที่ 25 กระบวนการการประเมินโดยใช้ตารางที่ถูกซ่อน (hidden table).....64
ภาพที่ 26 ตัวอย่างตารางของข้อมูลและความสัมพันธ์ระหว่างตาราง.....66
ภาพที่ 27 ความสัมพันธ์ของแต่ละคุณลักษณะข้อมูล 66
ภาพที่ 28 ข้อมูลที่ใช้ในการสร้างแบบจำลองของมหาวิทยาลัย 68
ภาพที่ 29 ความสัมพันธ์ของตารางข้อมูล.....69
ภาพที่ 30 ความสัมพันธ์ของตัวอย่างข้อมูล.....69
ภาพที่ 31 สถานที่ตั้งและอำเภอต่างๆ ของจังหวัดพิษณุโลก 72
ภาพที่ 32 ตัวอย่างข้อมูลแบบจำลองระดับครัวเรือน.....74
ภาพที่ 33 ขั้นตอนการออกแบบโปรแกรมการสร้างประชากรสังเคราะห์.....75
ภาพที่ 34 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลเขตเทศบาล.....79
ภาพที่ 35 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรส 80
ภาพที่ 36 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุ 82
ภาพที่ 37 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก 83
ภาพที่ 38 ความคลาดเคลื่อนที่เกิดขึ้นในการสร้างประชากรครั้งที่ 2.....86
ภาพที่ 39 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลเขตเทศบาล.....87
ภาพที่ 40 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรส 88
ภาพที่ 41 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุ 90
ภาพที่ 42 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก 91

ภาพที่ 43 กราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศชาย 93
ภาพที่ 44 กราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศหญิง 94

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ประชากรสังเคราะห์ (synthetic population) เป็นการสร้างฐานข้อมูลประชากรที่อยู่ในรูปแบบบุคคล (individual) และครัวเรือน (household) ในพื้นที่ที่สนใจ ซึ่งถูกสร้างขึ้นโดยการใส่คุณลักษณะ (attribute) ต่างๆ ที่จะเป็นตัวกำหนดรายละเอียดของประชากร สำหรับประชากรรายบุคคลจะมีคุณลักษณะที่เป็นตัวกำหนดรายละเอียด ดังนี้ เพศ อายุ สถานที่อยู่อาศัย การศึกษา สถานภาพสมรส สำหรับคุณลักษณะที่เป็นตัวกำหนดรายละเอียดของครัวเรือน เช่น อายุและเพศของหัวหน้าครัวเรือน ขนาดของครัวเรือน

ประชากรสังเคราะห์จะถูกสร้างขึ้นจากชุดข้อมูลของประชากรเรียกว่า population microdata [1, 2] ซึ่งเป็นข้อมูลที่มาจากการจดทะเบียนราษฎร (Population registration), การสำมะโนประชากร (census data) และการสำรวจ (survey) ที่ประกอบด้วยคุณลักษณะต่างๆ ที่เกี่ยวข้องกับตัวบุคคลและครัวเรือน เช่น อายุ, เพศ, การศึกษา, สถานภาพสมรส, สถานที่อยู่อาศัย, สถานะของบุคคลในครัวเรือนและขนาดของครัวเรือน ซึ่งข้อมูลเหล่านี้จะสามารถแบ่งได้เป็น 2 ประเภท คือ ข้อมูลรวม (aggregate data) และข้อมูลรายบุคคล (disaggregated data)

ข้อมูลรวม คือข้อมูลของประชากรทั้งหมดที่มีการแจกแจงให้อยู่ในรูปแบบของตารางสถิติสำหรับคุณลักษณะต่างๆ ของประชากร ตัวอย่างเช่น ตารางข้อมูลเพศ, ตารางข้อมูลอายุ และตารางข้อมูลการศึกษาที่แสดงในตารางที่ 1, 2 และ 3 ตามลำดับ ซึ่งเป็นส่วนหนึ่งของตารางที่ ก.1, ก.2 และ ก.3 ในภาคผนวกตามลำดับ โดยในแต่ละตารางจะระบุคุณลักษณะของข้อมูลประชากรเพียงหนึ่งอย่างเท่านั้นและข้อมูลที่เป็นข้อมูลรวมนั้นจะได้มาจากข้อมูลที่ได้จากทะเบียนราษฎรและการจัดทำสำมะโนประชากร

ตารางที่ 1 ตัวอย่างรายงานสถิติจำนวนประชากรจำแนกตามเพศ ปี 2553

อำเภอ	ชาย (คน)	หญิง (คน)	รวม (คน)	หลังคาเรือน (หลัง)
อำเภอเมืองพิษณุโลก	95,417	99,211	194,628	78,989
อำเภอนครไทย	33,293	32,931	66,224	20,331
อำเภอชาติตระการ	17,585	16,984	34,569	10,822
อำเภอบางระกำ	37,043	37,864	74,907	22,683

ตารางที่ 1 ตัวอย่างรายงานสถิติจำนวนประชากรจำแนกตามเพศ ปี 2553 (ต่อ)

อำเภอ	ชาย (คน)	หญิง (คน)	รวม (คน)	หลังคาเรือน (หลัง)
อำเภอบางกระทุ่ม	16,436	17,262	33,698	9,984
อำเภอพรหมพิราม	40,849	42,132	82,981	26,406
อำเภอวัดโบสถ์	14,546	14,575	29,121	9,548
อำเภอวังทอง	57,264	57,394	114,658	36,648
อำเภอเนินมะปราง	19,598	19,386	38,984	12,643

ตารางที่ 2 ตัวอย่างรายงานสถิติจำนวนประชากรจำแนกตามอายุ ปี 2553

อายุ	ชาย (คน)	หญิง (คน)	รวม (คน)	อายุ	ชาย (คน)	หญิง (คน)	รวม (คน)
1 ปี	4,754	4,410	9,164	**2 ปี**	4,733	4,553	9,286
3 ปี	4,901	4,724	9,625	**4 ปี**	4,965	4,680	9,645
5 ปี	5,170	4,687	9,857	**6 ปี**	5,148	4,652	9,800
7 ปี	4,740	4,616	9,356	**8 ปี**	4,819	4,633	9,452
9 ปี	4,769	4,647	9,416	**10 ปี**	5,151	4,836	9,987

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3 ตัวอย่างจำนวนนักเรียนในระดับชั้นอนุบาลของจังหวัดพิษณุโลก ปี 2553

ระดับชั้น	เพศ	จำนวน (คน)
อนุบาล 1	ชาย	4,843
	หญิง	4,349
อนุบาล 2	ชาย	4,615
	หญิง	4,321

ข้อมูลรายบุคคล คือ ข้อมูลของแต่ละบุคคลที่จะแสดงข้อมูลของคุณลักษณะต่างๆ อย่างครบถ้วน เช่น บุคคลหนึ่งเป็นเพศหญิง อายุ 15 ปี มีระดับการศึกษามัธยมศึกษาปีที่ 3 ดังแสดงตัวอย่างในตารางที่ 4 ซึ่ง disaggregated data นี้จะได้มาจากการสำรวจข้อมูลประชากร หรือเรียกว่า survey microdata [2]

ตารางที่ 4 ตัวอย่างของข้อมูลรายบุคคลซึ่งมีหลายคุณลักษณะ (disaggregated data)

เพศ	หญิง
อายุ	15
ระดับการศึกษา	มัธยมศึกษาปีที่ 3

ชุดข้อมูลประชากรจะถูกเปิดเผยและสามารถเข้าถึงข้อมูลได้แตกต่างกันในแต่ละประเทศ ซึ่งแบ่งได้ 3 ระดับ [2] ดังนี้

1. มีการเปิดเผยของชุดข้อมูลอย่างสมบูรณ์ ซึ่งประกอบไปด้วยชุดข้อมูล census data และ survey microdata ดังตัวอย่างของข้อมูลในประเทศสวีเดนซึ่งสามารถเข้าถึงชุดข้อมูล survey microdata ได้อย่างอิสระ

2. มีการเปิดเผยของชุดข้อมูล census data อย่างเพียงพอแต่มีการเปิดเผยของ survey microdata เพียงบางส่วน เช่น ข้อมูลของประเทศอังกฤษ

3. มีการเปิดเผยเพียงชุดข้อมูล census data เท่านั้นแต่ไม่มีการเปิดเผยของชุดข้อมูล survey microdata เลย ซึ่งตัวอย่างของการเปิดเผยชุดข้อมูลแบบนี้มักจะอยู่ในพื้นที่ของประเทศที่กำลังพัฒนาดังเช่น ประเทศจีนและประเทศไทย เป็นต้น

สำหรับกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ในปัจจุบันจะมีการเลือกใช้กระบวนการที่แตกต่างกันไป โดยเลือกให้เหมาะสมกับข้อมูลที่มีการเปิดเผยไว้ในแต่ละประเทศและรวมไปถึงข้อจำกัดของตัวข้อมูล ไม่ว่าจะเป็นข้อมูลที่สามารถเปิดเผยได้ซึ่งมาจากหลายแหล่งข้อมูลที่มีความไม่สอดคล้องกันและมีไม่เพียงพอหรือมีระดับความละเอียดของข้อมูลที่แตกต่างกัน ซึ่งจะมีผลต่อความถูกต้องแม่นยำในการสร้างฐานข้อมูลประชากร จากการรวบรวมข้อมูลที่ได้จากงานวิจัยที่ผ่านมาพบว่ามี 2 วิธีหลัก คือ Synthetic Reconstruction และ Combinatorial Optimization

Synthetic Reconstruction เป็นการสร้างประชากรสังเคราะห์ที่มีข้อมูลตั้งต้นคือ aggregate data ในรูปแบบของตารางสถิติเพื่อบันทึกคุณลักษณะต่างๆ ของประชากรโดยการค่อยๆ เพิ่มคุณลักษณะต่างๆ เข้าไปทีละอย่างซึ่งอาศัยตารางข้อมูลสถิติประชากรและหลักการของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) เพื่อทำให้ได้รายละเอียดของข้อมูลประชากรรายบุคคลที่สนใจ [3]

Combinatorial Optimization เป็นการสร้างประชากรสังเคราะห์ที่มีข้อมูลตั้งต้นคือข้อมูลกลุ่มตัวอย่างหรือ survey microdata จำนวนหนึ่งและข้อมูลทางสถิติของประชากร โดยจะสุ่มเลือก

ประชากรจากกลุ่มตัวอย่างมาสร้างประชากรในพื้นที่ที่สนใจที่แบ่งออกเป็นโซน (zone) แล้วเปรียบเทียบกับข้อมูลทางสถิติที่มีอยู่ เพื่อให้มีความใกล้เคียงกับข้อมูลทางสถิติตั้งต้นมากที่สุด หากยังมีความคลาดเคลื่อนมากก็จะทำการสุ่มเพื่อเปลี่ยนประชากรไปเรื่อยๆ จนใกล้เคียงที่สุด [3, 4]

ประชากรสังเคราะห์ที่ถูกสร้างขึ้นจะเป็นส่วนหนึ่งในการจำลองระดับไมโคร (microsimulation) เนื่องจากการจำลองระดับไมโครจะเป็นการสร้างแบบจำลองของพฤติกรรมและการมีปฏิสัมพันธ์ของหน่วยขนาดเล็ก (micro unit) ซึ่งในที่นี้คือ ประชากรสังเคราะห์ระดับบุคคลและระดับครัวเรือน การจำลองระดับไมโครที่ใช้ประชากรสังเคราะห์นี้สามารถสะท้อนสิ่งที่เกิดขึ้นหรือการเปลี่ยนแปลงต่างๆ เพื่อประเมินผลกระทบของการเปลี่ยนแปลงและสามารถบอกถึงความน่าจะเป็นไปได้ในการคาดคะเนเป้าหมายของนโยบายต่างๆ [1] จึงกลายเป็นเครื่องมือที่ได้รับการยอมรับในการประเมินผล [5] ในหลายด้าน ไม่ว่าจะเป็นการวิเคราะห์เพื่อวางแผนเกี่ยวกับนโยบายด้านเศรษฐกิจและสังคม เช่น การดูแลเด็กและผู้สูงอายุ, การวางแผนตลาดแรงงาน, การกำหนดอัตราการเสียภาษี, สวัสดิการและค่าจ้างพนักงาน, เป็นต้น นโยบายด้านสาธารณสุข เช่น การขยายบริการรักษาพยาบาล, การป้องกันและควบคุมโรค เป็นต้น นโยบายด้านการศึกษา เช่น การจัดสรรครูให้เพียงพอต่อนักเรียน การขยายหรือลดขนาดของโรงเรียนให้เหมาะสมกับเด็กที่อยู่ในวัยเรียน เป็นต้น ด้านอาชญากรรม สามารถวิเคราะห์ได้ว่าคุณลักษณะใดที่มีผลต่อการเกิดอาชญากรรม เช่น อายุ, การศึกษา, รายได้ เพื่อที่จะวางแผนป้องกันและลดการเกิดอาชญากรรมในแต่ละพื้นที่ได้ รวมไปถึงด้านการจราจร ซึ่งสามารถที่จะวางแผนการเดินทางของคนแต่ละประเภทซึ่งมีรูปแบบการเดินทางที่แตกต่างกัน เช่น นักเรียนและบุคคลวัยทำงาน เป็นต้น การจัดสรรการบริการของรถสาธารณะให้ครอบคลุมกับจำนวนประชากรและให้เหมาะสมกับช่วงเวลาเพื่อรองรับคนที่จะใช้บริการได้อย่างมีประสิทธิภาพ

จากงานวิจัยที่ผ่านมา ได้มีการใช้ประโยชน์จากการวิเคราะห์ประชากรสังเคราะห์อย่างกว้างขวางในหลายด้าน ไม่ว่าจะเป็นด้านเศรษฐกิจและสังคม เช่น SMILE model (Simulation Model for the Irish Local Economy model) ซึ่งเป็นแบบของการจำลองระดับไมโครในเชิงพื้นที่ที่มีการเปลี่ยนแปลงตลอดเวลาโดยมีจุดมุ่งหมายเพื่อวิเคราะห์ผลกระทบของการเปลี่ยนแปลงทางนโยบายและการพัฒนาในด้านเศรษฐกิจบนพื้นที่ชนบทของประเทศไอร์แลนด์ [6], การจำลองเพื่อประเมินการวิเคราะห์ข้อมูลเกี่ยวกับผลกระทบในด้านการกระจายของนโยบายที่แตกต่างกัน [7], การจำลองเพื่อวิเคราะห์ระบบภาษีและผลประโยชน์ในพื้นที่ต่างๆ ของนโยบายรัฐบาล [8] ด้านศาสนาก็มีการจำลองเพื่อวิเคราะห์รูปแบบของการเปลี่ยนแปลงทางศาสนาของประชากรในสังคมสเปน [9] สำหรับด้านสาธารณสุขก็ยังมีการสร้างแบบจำลองเพื่อใช้วางแผนเกี่ยวกับนโยบายทางด้านการดูแลสุขภาพของประชากร เช่น การพัฒนารูปแบบการจำลองระดับไมโครในประเทศนิวซีแลนด์ปี 2002

เพื่อสร้างระบบประวัติการรักษาสุขภาพสำหรับบุคคลโดยผลลัพธ์ที่ได้เป็นจำนวนของผู้มาใช้บริการในแต่ละปี, การกระจายของเงื่อนไขของโรคที่ประชากรมารับบริการด้านการรักษาและกิจกรรมการรักษาพยาบาลของแพทย์เพื่อใช้วางแผนในปี 2021 [10] และการจำลองการกระจายของโรคต่างๆ [11] นอกจากนี้ยังได้มีการนำไปประยุกต์ใช้กับเรื่องของการเดินทาง การจราจรและเรื่องที่เกี่ยวข้องกับสัตว์ต่างๆ เช่น การสร้างแบบจำลองตามรูปแบบของการกระจายตัวของการเดินทางมาพักผ่อนในสถานที่ท่องเที่ยวที่มีลักษณะเป็นป่าโดยมีความสัมพันธ์เกี่ยวข้องกับพื้นฐานของผู้เข้าชมและลักษณะของป่าซึ่งได้มีกรณีศึกษาในแคว้น Wallonia ของประเทศเบลเยี่ยม [12], การสร้างแบบจำลองของวัฏจักรชีวิต, การกระจายตัวในสถานที่ต่างๆ และการควบคุมแมลงวันเซ็ทซี (Tsetse Fly) ซึ่งเป็นแมลงวันขนาดใหญ่จากแอฟริกาที่ดูดเลือดของสัตว์มีกระดูกสันหลังเป็นอาหารและเป็นพาหะนำโรคของโรคเหงาหลับที่มีอาการร้ายแรงถึงขั้นเสียชีวิตได้ โดยงานวิจัยนี้ได้ใช้กรณีศึกษาที่เมือง Nguruman ประเทศเคนย่า [13], การสร้างแบบจำลองประชากรเสือและการเปลี่ยนแปลงอาณาเขตซึ่งมีการจำลองประชากรของเสือทั้งทางด้านเพศ อายุ การให้กำเนิด การตายและอาณาเขตที่อยู่อาศัยของเสือเพื่อประเมินภัยคุกคามที่เกิดขึ้นและวางแผนในด้านการอนุรักษ์ประชากรเสือ [14]

แม้ว่าการประยุกต์ใช้ประชากรสังเคราะห์จะมีประโยชน์อย่างกว้างขวาง แต่หากประชากรสังเคราะห์ที่ถูกสร้างขึ้นนั้นมีความคลาดเคลื่อนจากประชากรจริง ก็ทำให้การวิเคราะห์ประชากรสังเคราะห์นั้นคลาดเคลื่อนไปด้วย ซึ่งการสร้างประชากรสังเคราะห์ให้มีความถูกต้องแม่นยำ ใกล้เคียงกับประชากรจริงนั้น ขึ้นอยู่กับข้อมูลที่นำมาใช้และกระบวนการสร้างประชากรสังเคราะห์ที่ต้องออกแบบให้เหมาะสมกับข้อมูลที่นำมาใช้ด้วย

ดังนั้นในงานวิจัยนี้ทางผู้วิจัยจึงนำเสนอกระบวนการสร้างประชากรสังเคราะห์ (synthetic population generator) ทั้งในระดับบุคคลและระดับครัวเรือนที่เหมาะสมกับข้อมูลของประเทศไทยเพื่อที่จะสามารถสร้างประชากรสังเคราะห์ของประเทศไทยได้อย่างถูกต้อง และใกล้เคียงกับประชากรจริง ซึ่งข้อมูลของประเทศไทยมีข้อจำกัดคือ

1. มีการเปิดเผยเพียงข้อมูลที่เป็นข้อมูลรวม (aggregate data) ที่มาจากการสำมะโนประชากรเท่านั้น แต่ในส่วนของข้อมูลรายบุคคล (disaggregated data) ที่มาจากการสำรวจยังเป็นข้อมูลที่ไม่สามารถเข้าถึงได้หรืออาจเข้าถึงได้ในบางพื้นที่แต่มีจำนวนจำกัด คือ มีข้อมูลเป็นสัดส่วนที่น้อยมากเมื่อเทียบกับประชากรรวมทั้งหมดในพื้นที่แห่งนั้น จึงไม่สามารถใช้ข้อมูลที่เป็น disaggregated data ในการศึกษาคุณลักษณะของประชากรในพื้นที่นั้นๆ ได้ ทำให้ไม่สามารถใช้กระบวนการ Combinatorial Optimization ได้

2. ข้อมูลระดับครัวเรือนยังมีการเก็บและเปิดเผยข้อมูลน้อย ไม่เพียงพอต่อการวิเคราะห์ข้อมูล
3. ข้อมูลที่ถูกเปิดเผยนั้นถูกเก็บมาด้วยความละเอียดที่แตกต่างกันจากหลายหน่วยงาน ทำให้ข้อมูลมีความไม่สอดคล้องกันของตัวข้อมูล เช่น จำนวนประชากรจำแนกตามเพศของหน่วยงานหนึ่ง ไม่เท่ากับอีกหน่วยงานหนึ่ง

สำหรับกระบวนการสร้างประชากรสังเคราะห์ในงานวิจัยนี้ได้ออกแบบเพื่อแก้ไขข้อจำกัดของข้อมูลดังกล่าว โดยกระบวนการสร้างประชากรสังเคราะห์ที่ออกแบบนี้จะเป็นส่วนหนึ่งของกระบวนการ Synthetic Reconstruction โดยการสร้างประชากรสังเคราะห์ระดับครัวเรือนถือเป็นความท้าทายอย่างหนึ่งสำหรับการสร้างความสัมพันธ์ของบุคคลเข้าในครัวเรือนเดียวกัน เพื่อให้ประชากรสังเคราะห์ระดับครัวเรือนมีความแม่นยำใกล้เคียงกับประชากรที่มีอยู่จริง สำหรับประเทศไทยที่มีรูปแบบของครอบครัวที่ซับซ้อนอีกทั้งยังเป็นพื้นที่มีการเปิดเผยของข้อมูลอย่างจำกัด ดังนั้นกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบนี้จะใช้ข้อมูลที่ได้จากการจัดทำสำมะโนประชากรซึ่งเป็นข้อมูลรวม (aggregate data) เพื่อให้ได้ผลลัพธ์คือข้อมูลรายบุคคล (disaggregated data) โดยประชากรสังเคราะห์ที่ได้นี้มีความละเอียดถึงระดับตำบลซึ่งมีความเหมาะสมที่จะนำไปใช้เป็นข้อมูลตัวแทนของคุณลักษณะของประชากรในพื้นที่ต่างๆ ที่ต้องการศึกษาได้ โดยจะมีการประยุกต์ใช้กับกรณีศึกษาในการสร้างประชากรสังเคราะห์ของจังหวัดพิษณุโลกเพื่อเปรียบเทียบระหว่างประชากรสังเคราะห์ที่สร้างขึ้นใหม่นี้กับฐานข้อมูลประชากรที่มีอยู่เดิมเพื่อประเมินความถูกต้องและความน่าเชื่อถือของกระบวนการสร้างประชากรสังเคราะห์ในการสร้างประชากรสังเคราะห์ระดับบุคคลและครัวเรือน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 วัตถุประสงค์

เพื่อพัฒนากระบวนการสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือนจากข้อมูลสถิติที่มีหลายระดับความละเอียดให้มีความใกล้เคียงกับประชากรจริง

1.3 ขอบเขตของงานวิจัย

- งานวิจัยนี้ใช้ข้อมูลเกี่ยวกับประชากรที่มีการเปิดเผยของประเทศไทยในรูปแบบตารางข้อมูลเท่านั้น
- งานวิจัยนี้ใช้ข้อมูลสถิติของจังหวัดพิษณุโลกในปี 2553 และข้อมูลบางส่วนเป็นข้อมูลสถิติรวมของทั้งประเทศในกรณีที่ข้อมูลสถิตินี้ไม่มีอยู่ในข้อมูลสถิติของจังหวัดพิษณุโลก

- งานวิจัยนี้จะทำการศึกษาความสัมพันธ์ของข้อมูลสถิติเพื่อออกแบบกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์
- กระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ที่สร้างขึ้นเพื่อสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือน
- งานวิจัยนี้ดำเนินงานภายใต้โปรแกรม MySQL บนคอมพิวเตอร์เครื่องเดียว

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับในงานวิจัยนี้ ได้แก่

1. ได้กระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์จากการศึกษาความสัมพันธ์ของข้อมูลที่เปิดเผยของประเทศไทยอย่างเหมาะสม เพื่อเป็นแนวทางในการสร้างประชากรของทั้งประเทศ
2. ได้ประชากรสังเคราะห์ทั้งระดับบุคคลและระดับครัวเรือนที่อยู่ในพื้นที่ที่สนใจคือ จังหวัดพิษณุโลก โดยสามารถนำไปวิเคราะห์และวางแผนนโยบายต่างๆ เกี่ยวกับประชากรได้อย่างถูกต้องเหมาะสม
3. เป็นแนวทางสำหรับการศึกษาความสัมพันธ์ของข้อมูลอื่นๆ ที่ไม่ใช่ข้อมูลประชากรเพื่อประยุกต์ใช้ในการออกแบบกระบวนการในการจำลองในแบบอื่นๆ ต่อไปในอนาคต

1.5 ขั้นตอนการดำเนินงานวิจัย

ขั้นตอนการดำเนินงานวิจัยที่ได้วางแผนไว้มีดังนี้

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือน
2. ศึกษาวิธีที่ใช้ในการสร้างประชากรสังเคราะห์และข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์แบบต่างๆ
3. สำรวจข้อมูลที่มีการเปิดเผยจากหน่วยงานต่างๆ ของประเทศไทย กลั่นกรองข้อมูลที่จะนำมาใช้ พร้อมทั้งเลือกจังหวัดที่มีข้อมูลที่เหมาะสมเพื่อใช้เป็นกรณีศึกษา
4. ออกแบบกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ที่เหมาะสมกับข้อมูลของประเทศไทย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5. นำกระบวนการที่ออกแบบได้มาประยุกต์ใช้กับข้อมูลของจังหวัดพิษณุโลก โดยสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือน
6. ประเมินความถูกต้องของประชากรสังเคราะห์ที่สร้างขึ้นกับข้อมูลทางสถิติทั้งในระดับบุคคลและระดับครัวเรือน
7. ปรับปรุงกระบวนการสร้างประชากรสังเคราะห์
8. สรุปผลการทดลองและตีพิมพ์ผลการทำวิจัย
9. เรียบเรียงและจัดทำวิทยานิพนธ์

1.6 งานวิจัยที่ได้รับการตีพิมพ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้นำเสนอในการประชุมวิชาการ ดังนี้

Natthaporn Watthanasutthi and Veera Muangsin, *"Generating Synthetic Population at Individual and Household Levels with Aggregate Data",* The 13th International Joint Conference on Computer Science and Software Engineering (JCSSE2016), Khon Kaen, Thailand, July 13-15, 2016.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์

การเลือกข้อมูลที่จะนำมาใช้ในการสร้างประชากรสังเคราะห์นั้นมีความสำคัญต่อกระบวนการสร้างประชากรเป็นอย่างมาก เนื่องจากข้อมูลที่เลือกมานี้จะเป็นข้อมูลต้นแบบในการสร้างประชากรสังเคราะห์ ซึ่งประชากรสังเคราะห์ที่สร้างขึ้นได้นี้จะมีความใกล้เคียงกับประชากรที่มีอยู่จริงหรือไม่ก็ขึ้นอยู่กับว่าข้อมูลที่ได้เลือกมานั้นมีความน่าเชื่อถือเพียงใดและยังนำไปสู่การเลือกกระบวนการที่ใช้ในการสร้างประชากรด้วย ซึ่งในแต่ละประเทศจะมีการเก็บและเปิดเผยข้อมูลประชากรที่แตกต่างกันไป ไม่ว่าจะเป็นด้านชนิดของข้อมูลคือ ข้อมูล survey microdata และข้อมูล census data, ด้านคุณลักษณะของข้อมูลตัวอย่างเช่นอายุ เพศและการศึกษา, ด้านความละเอียดของข้อมูลตัวอย่างเช่น ประเทศ จังหวัด อำเภอ ตำบล เป็นต้น

2.1.1 คุณลักษณะของข้อมูล

ข้อมูลประชากรที่ได้จากการจัดทำสำมะโนประชากรหรือ census data ซึ่งเป็นข้อมูลแบบ aggregate data จะเป็นข้อมูลรวมของประชากรทั้งประเทศที่อยู่ในรูปแบบตารางสถิติซึ่งไม่ได้จำแนกเป็นข้อมูลรายบุคคล โดยที่คุณลักษณะของข้อมูลนั้นจะเรียกว่า attribute ซึ่งในที่นี้คือข้อมูลเพศ อายุ การศึกษาและสถานภาพสมรส เป็นต้น ข้อมูลเหล่านี้มักจะมาในรูปแบบของตัวเลขหรือรูปแบบการกระจายของค่าความน่าจะเป็น (probability distribution) และมักจะมีความสัมพันธ์ที่ขึ้นอยู่กับคุณลักษณะต่างๆ ของข้อมูลในแต่ละตารางด้วย เช่น ความสัมพันธ์ของสถานภาพสมรสได้มาจากการรวบรวมข้อมูลจากตารางของอายุและตารางของสถานภาพสมรส ส่วนความสัมพันธ์ระหว่างรายได้และการศึกษาได้มาจากการรวบรวมข้อมูลจากตารางของรายได้และการศึกษา จึงทำให้ลำดับการสร้างของแต่ละคุณลักษณะของข้อมูลควรได้รับการพิจารณาด้วย กล่าวคือจะทำการสร้างคุณลักษณะที่เป็นอิสระและไม่ขึ้นอยู่กับคุณลักษณะอื่นก่อนเป็นอันดับแรกและในอันดับต่อมาจึงค่อยสร้างคุณลักษณะที่สัมพันธ์กับคุณลักษณะที่ได้สร้างไปแล้วก่อนหน้านั้น ตัวอย่างเช่น ตารางข้อมูลอายุที่บ่งบอกถึงจำนวนคนแต่ละอายุดังแสดงในตารางที่ 5 และตารางข้อมูลที่บ่งบอกสถานภาพสมรสที่จำแนกตามอายุดังแสดงในตารางที่ 6 จะมีคุณลักษณะที่ต้องสร้างคืออายุและสถานภาพสมรส จะเห็นได้ว่ามีความสัมพันธ์กันระหว่าง 2 ตารางดังกล่าวจึงจำเป็นต้องสร้างคุณลักษณะของอายุก่อนเป็นอันดับแรกเนื่องจากตารางข้อมูลอายุไม่ขึ้นอยู่กับคุณลักษณะอื่นๆ ลำดับต่อมาจึงสร้างคุณลักษณะของสถานภาพสมรสเนื่องจากตารางข้อมูลสถานภาพสมรสมี

ความสัมพันธ์กับคุณลักษณะข้อมูลอายุ [15] โดยลำดับการสร้างของคุณลักษณะข้อมูลดังแสดงในภาพที่ 1

ตารางที่ 5 ความน่าจะเป็นของประชากรในแต่ละกลุ่มอายุ (P(j))

อายุ (ปี)	**จำนวน (%)**
18 – 30	40
31 – 60	60

ตารางที่ 6 ความน่าจะเป็นที่ประชากรสถานภาพสมรสต่างๆในแต่ละกลุ่มอายุ (P(h | j))

อายุ (ปี)	**แต่งงานแล้ว (%)**	**ยังไม่ได้แต่งงาน (%)**
18 – 30	60	40
31 – 60	80	20

ภาพที่ *1* ลำดับการสร้างของคุณลักษณะข้อมูล

2.1.2 ประเภทของข้อมูล

เนื่องจากข้อมูลที่เกี่ยวกับประชากรมีหลายคุณลักษณะจึงมีการแบ่งประเภทของข้อมูลได้หลายวิธีตามเกณฑ์ในการจำแนก เช่น

2.1.2.1 จำแนกตามระดับของการวัด

ข้อมูลจำแนกตามระดับของการวัด แบ่งได้ 4 ชนิด คือ

1. *ข้อมูลระดับนามบัญญัติ (Nominal scale)* เป็นการแบ่งกลุ่มข้อมูลเพื่อสะดวกต่อการวิเคราะห์ โดยการแบ่งกลุ่มข้อมูลนั้นจะถือว่าแต่ละกลุ่มจะมีความเสมอภาคกันหรือเท่าเทียมกัน

ค่าที่กำหนดให้แต่ละกลุ่มจะไม่มีความหมาย ไม่สามารถนำมาจัดลำดับข้อมูลได้และไม่สามารถนำมาใช้ในการคำนวณได้ เช่น เพศ อาชีพ ศาสนา สถานภาพสมรสและระดับการศึกษา เป็นต้น

2. *ข้อมูลระดับอันดับ (Ordinal scale)* เป็นการรวบรวมข้อมูลที่สามารถแสดงผลเปรียบเทียบข้อมูลในแต่ละกลุ่มได้ สามารถบอกได้ว่าข้อมูลที่อยู่ในแต่ละกลุ่มนั้นมีความแตกต่างกันโดยสามารถพิจารณาจากการจัดลำดับข้อมูลด้วย นั่นคือสามารถบอกได้ว่ากลุ่มใดดีกว่ากลุ่มอื่นๆ หรือกลุ่มใดที่มากกว่าหรือน้อยกว่ากลุ่มอื่นๆ แต่ไม่สามารถบอกปริมาณความมากกว่าหรือน้อยกว่าเป็นจำนวนเท่าใดและค่าที่กำหนดให้แต่ละกลุ่มไม่สามารถนำมาใช้ในการคำนวณได้ เช่น ลำดับของลูกหรือทัศนคติว่ามีความคิดเห็นอย่างไร เช่น เห็นด้วยหรือไม่เห็นด้วย เป็นต้น

3. *ข้อมูลระดับช่วงชั้น (Interval scale)* เป็นการรวบรวมข้อมูลที่แบ่งสิ่งที่ศึกษาออกเป็นระดับหรือเป็นช่วงๆ โดยแต่ละช่วงมีขนาดหรือระยะห่างเท่ากัน ทำให้สามารถบอกระยะห่างของช่วงได้ อีกทั้งสามารถบอกได้ว่าข้อมูลนั้นมากกว่าหรือน้อยกว่ากันเท่าไร จึงทำให้ข้อมูลมีความแตกต่างกันในเชิงปริมาณ เช่น วันที่ อุณหภูมิ เป็นต้น

4. *ข้อมูลระดับอัตราส่วน (Ratio scale)* เป็นการรวบรวมข้อมูลที่ละเอียดและสมบูรณ์ที่สุด สามารถบอกความแตกต่างในเชิงปริมาณได้ โดยแบ่งสิ่งที่ต้องการศึกษาออกเป็นช่วงๆ เหมือนมาตรวัดอันตรภาคที่แต่ละช่วงมีระยะห่างเท่ากัน และค่า 0 ของข้อมูลชนิดนี้เป็นค่า 0 แท้ซึ่งหมายถึงไม่มีอะไรเลยหรือมีจุดที่เริ่มต้นที่แท้จริงและสามารถนำตัวเลขในข้อมูลชนิดนี้มา บวก ลบ คูณ หารได้ เช่น อายุ รายได้ เป็นต้น

2.1.2.2 จำแนกตามลักษณะของข้อมูล

ข้อมูลจำแนกตามลักษณะของข้อมูล แบ่งได้เป็น 2 ประเภท คือ

1. *ข้อมูลเชิงคุณภาพ (Qualitative data)* เป็นข้อมูลที่ไม่ได้อยู่ในรูปของตัวเลขโดยตรงแต่จะแสดงลักษณะของข้อมูลที่แตกต่างกัน ซึ่งจะมีข้อมูลระดับนามบัญญัติและข้อมูลระดับอันดับรวมอยู่ด้วย

2. *ข้อมูลเชิงปริมาณ (Quantitative or Variable data)* เป็นข้อมูลที่อยู่ในรูปของตัวเลขสามารถแสดงความแตกต่างในเรื่องของปริมาณหรือขนาดได้ ซึ่งจะมีข้อมูลระดับช่วงชั้นและข้อมูลระดับอัตราส่วนรวมอยู่ด้วย แบ่งได้เป็น 2 ประเภท คือ

2.1 ข้อมูลแบบไม่ต่อเนื่อง (Discrete data) หมายถึง ข้อมูลที่มีค่าเป็นเลขจำนวนเต็มที่มีความหมายและที่ค่าทศนิยมจะไม่มีความหมาย เช่น จำนวนสิ่งของ จำนวนคน เป็นต้น

2.2 ข้อมูลแบบต่อเนื่อง (Continuous data) หมายถึง ข้อมูลที่อยู่ในรูปตัวเลขที่มีค่าได้ทุกค่าในช่วงที่กำหนดและทศนิยมมีความหมายด้วย เช่น รายได้ น้ำหนัก เป็นต้น

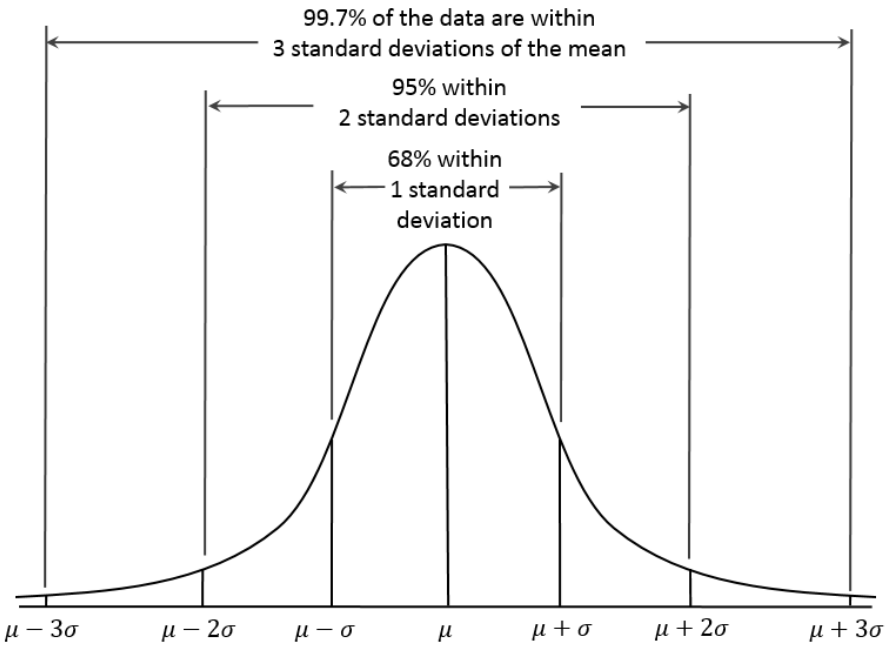
2.1.3 การแจกแจงของข้อมูล (Probability Distribution)

2.1.3.1 การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงแบบปกติหรือเรียกอีกอย่างหนึ่งว่าการแจกแจงแบบเกาส์หรือการแจกแจงแบบเกาส์เซียน (Gaussian distribution) เป็นการแจกแจงความถี่ของข้อมูลที่เป็นค่าที่วัดได้เป็นปกติ ได้แก่ อายุ คะแนน ความสูง น้ำหนัก เป็นต้น เนื่องจากข้อมูลเหล่านี้จะมีค่าเฉลี่ยอยู่ระดับกลางๆ และมีจำนวนไม่มากที่มีค่าสูงหรือมีค่าต่ำกว่าปกติ ลักษณะกราฟของข้อมูลจะเป็นรูประฆังคว่ำดังที่แสดงในภาพที่ 2 โดยมีฟังก์ชันการแจกแจงของความน่าจะเป็น ดังนี้

$$f(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad ; \quad -\infty < X < \infty$$

โดยที่ μ คือค่าเฉลี่ยของการแจกแจง, **σ** คือส่วนเบี่ยงเบนมาตรฐานของการแจกแจง, **π** คือค่าคงที่เท่ากับ 3.1416 และ e คือค่าคงที่เท่ากับ 2.7128



ภาพที่ 2 ลักษณะกราฟของการแจกแจงแบบปกติ

2.1.3.2 การแจกแจงแบบปัวซอง (Poisson Distribution)

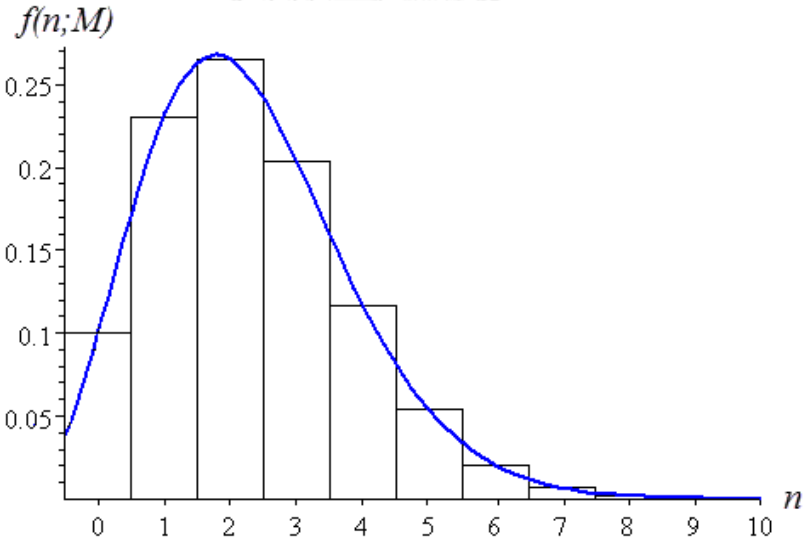
การแจกแจงแบบปัวซองนั้นจะอธิบายถึงจำนวนที่เป็นไปได้ของวัตถุที่จะสามารถพบในปริมาณบางอย่างหรือจำนวนของเหตุการณ์ที่เกิดขึ้นที่จะสามารถสังเกตได้ในช่วงเวลาใดเวลาหนึ่ง [16] ลักษณะกราฟของข้อมูลจะแสดงในภาพที่ 3 โดยมีฟังก์ชันการแจกแจงความน่าจะเป็น [17] คือ

$$f(n;M) = \frac{M^n e^{-M}}{n!}$$

โดยที่ n = จำนวนของโอกาสการเกิด (โดยที่ n = 0 -> infinity)

M = จำนวนประชากรเฉลี่ยประชากรเฉลี่ย

f (n; M) คือ ความน่าจะเป็นของ n เมื่อมีค่าเฉลี่ย = M



ภาพที่ *3* ลักษณะกราฟของการแจกแจงแบบปัวซอง

2.1.4 ชนิดของความสัมพันธ์

คุณลักษณะของข้อมูลจะมีความสัมพันธ์ของข้อมูล [15] โดยใช้ความสัมพันธ์ของเงื่อนไขของความน่าจะเป็น (conditional probability relationship) หรือเรียกว่าความน่าจะเป็นร่วม (joint probability) นั้นจะมีความเกี่ยวข้องกับคุณลักษณะที่มีความสัมพันธ์กัน โดย P(h | j) คือความน่าจะเป็นของประชากรที่ทราบว่ามีคุณลักษณะ j ซึ่งจะมีคุณลักษณะ h, P(h ∩ j) คือ ความ

น่าจะเป็นของประชากรที่จะมีคุณลักษณะ h และ j เทียบกับประชากรทั้งหมด และ P(j) คือความน่าจะเป็นของประชากรที่มีคุณลักษณะ j ดังนั้นความน่าจะเป็นของประชากรที่ทราบว่ามีคุณลักษณะ j ซึ่งจะมีคุณลักษณะ h สามารถคำนวณได้โดยใช้สมการ

$$P(h \mid j) = \frac{P(h \cap j)}{P(j)}$$

ตัวอย่างเช่น คุณลักษณะ j คือ อายุ ดังแสดงในตารางที่ 5 ซึ่งสามารถแบ่งได้เป็น 2 กลุ่ม คือ กลุ่มอายุ 18-30 ปีและกลุ่มอายุ 31-60 ปีโดยมีความน่าจะเป็นคือ 40 % และ 60 % ตามลำดับ และทราบความน่าจะเป็นที่จะมีสถานภาพสมรสในแต่ละช่วงอายุดังแสดงในตารางที่ 6 จะเห็นได้ว่าในประชากรกลุ่มอายุ 18-30 ปีมี 60% ที่แต่งงานแล้วและอีก 40% ยังไม่ได้แต่งงาน ส่วนประชากรในกลุ่มอายุ 31-60 ปีจะมี 80% ที่แต่งงานแล้วและอีก 20% ยังไม่ได้แต่งงาน ดังนั้นในกลุ่มอายุ 18-30 ปีจะมี 40% * 60% = 24 % ของประชากรทั้งหมดที่แต่งงานแล้วและมี 40% * 40% = 16 % ของประชากรทั้งหมดที่ยังไม่ได้แต่งงาน สำหรับประชากรในกลุ่มอายุ 31-60 ปีจะมีประชากรที่แต่งงานแล้ว 60% * 80% = 12 % ของประชากรทั้งหมดและคนที่ยังไม่ได้แต่งงานมี 60% * 20% = 12 % ของประชากรทั้งหมดดังแสดงในตารางที่ 7

ตารางที่ 7 ความน่าจะเป็นของประชากรซึ่งทราบว่ามีอายุตามกลุ่มอายุ จะมีสถานภาพสมรสต่างๆ (P(h ∩ j))

อายุ (ปี)	แต่งงานแล้ว (%)	ยังไม่ได้แต่งงาน (%)
18 – 30	40 *60 = 24	40 * 40 = 16
31 – 60	60 *80 = 48	60 * 20 = 12

2.2 กระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์

การเลือกกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์มีความเกี่ยวข้องกับข้อมูลที่จะนำมาใช้ เนื่องจากกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ต้องออกแบบให้เหมาะสมกับข้อมูลเพื่อที่จะได้ข้อมูลของประชากรสังเคราะห์ที่มีความถูกต้องแม่นยำและมีความน่าเชื่อถือ สำหรับกระบวนการสร้างประชากรสังเคราะห์ที่เป็นที่ยอมรับและถูกนำไปใช้อย่างแพร่หลาย สามารถแบ่งได้เป็น 2 แบบ [18] คือ

2.2.1 Combinatorial Optimization

Combinatorial Optimization หรือเรียกอีกอย่างหนึ่งว่า Reweighting เป็นกระบวนการสร้างประชากรสังเคราะห์ที่ใช้กันอย่างแพร่หลาย ที่มีข้อมูลตั้งต้น (input data) เป็นข้อมูลของประชากรกลุ่มตัวอย่าง (disaggregated data) ที่ถูกสุ่มคัดเลือกมาจำนวนหนึ่งโดยจะมีคุณสมบัติต่างๆ อย่างครบถ้วนในประชากรแต่ละราย ซึ่งเป็นข้อมูลที่ได้จากการสำรวจประชากร (survey) และและข้อมูลสำมะโนประชากร (aggregate data) ซึ่งเป็นข้อมูลที่อยู่ในรูปแบบตารางสถิติ โดยกระบวนการพื้นฐานของ Combinatorial Optimization คือ การสุ่มเลือกประชากรจากกลุ่มตัวอย่างมาสร้างประชากรในพื้นที่ที่สนใจโดยแบ่งออกเป็นโซน (zone) แล้วจึงเปรียบเทียบกับข้อมูลทางสถิติที่มีอยู่เพื่อให้มีความใกล้เคียงกับข้อมูลทางสถิติตั้งต้นมากที่สุด หากยังไม่ใกล้เคียงจะทำการสุ่มเปลี่ยนประชากรไปเรื่อยๆ จนใกล้เคียงที่สุด [1]

สำหรับตัวอย่างการสร้างประชากรสังเคราะห์อย่างง่ายด้วยวิธี Combinatorial Optimization จะแสดงในภาพที่ 4 ซึ่งมีขั้นตอนดังนี้

1. จะใช้ข้อมูล 2 แบบ คือ ข้อมูล survey microdata ซึ่งเป็นข้อมูลกลุ่มตัวอย่างที่ได้จากการสำรวจ และข้อมูลแบบ census tabulation ซึ่งเป็นข้อมูลในรูปแบบตารางสถิติ ซึ่งในตัวอย่างนี้จะสนใจข้อมูลใน 2 คุณลักษณะ คือ ขนาดของครัวเรือน และอายุของผู้อยู่อาศัย ซึ่งแบ่งเป็นประชากรผู้ใหญ่และประชากรเด็ก
2. ขั้นตอนนี้จะสุ่มเลือกครัวเรือนจาก survey microdata มา 2 ครัวเรือน ซึ่งตัวอย่างนี้ได้สุ่มเลือกครัวเรือน a และ e โดยนำมาทำเป็นข้อมูลต้นแบบเพื่อนำไปใช้ในการสร้างประชากรสังเคราะห์
3. ในขั้นตอนนี้จะคำนวณค่าความแตกต่าง (absolute different) ระหว่างครัวเรือนที่สุ่มเลือกมาในขั้นตอนที่ 2 ซึ่งคือ ครัวเรือน a และ e ในที่นี้จะแทนด้วย Estimated frequency กับข้อมูล census tabulation ที่ได้รับมาจากขั้นตอนแรก ซึ่งจะแทนด้วย Observed frequency โดยจะคำนวณหาค่าความแตกต่างใน 2 คุณลักษณะ คือ ขนาดของครัวเรือนและอายุของผู้อยู่อาศัย

Step 1: Obtain sample survey microdata and small area constraints

Survey microdata

Household	Characteristics		
	size	adults	children
(a)	2	2	0
(b)	2	1	1
(c)	4	2	2
(d)	1	1	0
(e)	3	2	1

Known small area constraints [Published small area census tabulations]

1. Household size (persons per household)

Household size	Frequency
1	1
2	0
3	0
4	1
5+	0
Total	**2**

2. Age of occupants

Type of person	Frequency
adult	3
child	2

Step 2: Randomly select *two* households from survey sample [(a) & (e)] to act as an initial small-area microdata estimate

Step 3: Tabulate selected households and calculate (absolute) difference from known small-area constraints

Household size	Estimated Frequency (i)	Observed Frequency (ii)	Absolute difference | (i)-(ii) |
1	0	1	1
2	1	0	1
3	1	0	1
4	0	1	1
5+	0	0	0
		Sub-total:	*4*

Age	Estimated Frequency (i)	Observed Frequency (ii)	Absolute difference | (i)-(ii) |
adult	4	3	1
child	1	2	1
		Sub-total:	*2*

Total absolute difference = *4 + 2* = **6**

Step 4: Randomly select one of selected households (a or e). Replace with another household selected at random from the survey sample, provided this leads to a reduced total absolute difference

Households selected: (d) & (e) [Household (a) replaced]

Tabulate selection and calculate (absolute) difference from known constraints

Household size	Estimated Frequency (i)	Observed Frequency (ii)	Absolute difference | (i)-(ii) |
1	1	1	0
2	0	0	0
3	1	0	1
4	0	1	1
5+	0	0	0
		Sub-total:	*2*

Age	Estimated Frequency (i)	Observed Frequency (ii)	Absolute difference | (i)-(ii) |
adult	3	3	0
child	1	2	1
		Sub-total:	*1*

Total absolute difference = *2 + 1* = **3**

Step 5: Repeat step 4 until no further reduction in total absolute difference is possible:

Result: Final selected households: (c) & (d)

Household size	Estimated Frequency (i)	Observed Frequency (ii)	Absolute difference | (i)-(ii) |
1	1	1	0
2	0	0	0
3	0	0	0
4	1	1	0
5+	0	0	0
		Sub-total:	*0*

Age	Estimated Frequency (i)	Observed Frequency (ii)	Absolute difference | (i)-(ii) |
adult	3	3	0
child	2	2	0
		Sub-total:	*0*

Total absolute difference = *0 + 0* = **0**

ที่มา : [2]

ภาพที่ *4* ตัวอย่าง Combinatorial Optimization

3.1 สำหรับคุณลักษณะขนาดของครัวเรือน จากครัวเรือนที่สุ่มมาคือ a และ e คือ 3 คนและ 2 คนตามลำดับ ทำให้ Estimated frequency ของขนาดครัวเรือน 2 คนและ 3 คนมีความถี่อย่างละ 1 สำหรับข้อมูลสำมะโนประชากร มีขนาดครัวเรือนคือ 1 คนและ 4 คนทำให้ Observed frequency ของขนาดครัวเรือน 1 คนและ 4 คนมีความถี่อย่างละ 1 เมื่อหาความแตกต่างของคุณลักษณะครัวเรือนจะได้ผลรวม (Sub-total) เท่ากับ 4

3.2 สำหรับคุณลักษณะอายุผู้อยู่อาศัย จากครัวเรือนที่สุ่มมาคือ a และ e คือ มีผู้อยู่อาศัยที่เป็นผู้ใหญ่ คือ 2 คนทั้งคู่ สำหรับผู้อาศัยที่เป็นเด็ก คือ 0 คนและ 1 คนตามลำดับ ทำให้ผู้อยู่อาศัยที่เป็นผู้ใหญ่รวมเป็น 4 คนและผู้อาศัยที่เป็นเด็กรวมเป็น 1 คน ทำให้ Estimated frequency ของอายุผู้อยู่อาศัยที่เป็นผู้ใหญ่คือ 4 คนและอายุผู้อยู่อาศัยที่เป็นเด็กคือ 1 สำหรับข้อมูลสำมะโนประชากร Observed frequency ที่ได้คือผู้อยู่อาศัยที่เป็นผู้ใหญ่ คือ 3 และผู้อาศัยที่เป็นเด็ก คือ 2 คน จากนั้นหาความแตกต่างของคุณลักษณะอายุของผู้อยู่อาศัยจะได้ผลรวม (Sub-total) เท่ากับ 2

3.3 รวมค่าความแตกต่างของทั้ง 2 คุณลักษณะ คือ 4 + 2 = 6

4. สุ่มเลือกครัวเรือนมา 1 ครัวเรือนจากข้อมูลแบบ survey microdata มาแทนครัวเรือนที่สุ่มเลือกมาในขั้นตอนที่ 2 (a หรือ e) 1 ครัวเรือน เพื่อให้ผลรวมค่าความแตกต่างมีค่าลดลง ซึ่งหากค่านี้ยิ่งน้อยแสดงว่าประชากรสังเคราะห์ที่สร้างขึ้นได้นั้นมีความใกล้เคียงกับข้อมูลทางสถิติมาก ในตัวอย่างนี้จะสุ่มเปลี่ยนครัวเรือน a ด้วยครัวเรือน d

4.1 สำหรับคุณลักษณะขนาดของครัวเรือน จากครัวเรือนที่สุ่มมาคือ d และ e คือ 1 คนและ 2 คนตามลำดับ ทำให้ Estimated frequency ของขนาดครัวเรือน 1 คนและ 3 คนมีความถี่อย่างละ 1 สำหรับข้อมูลสำมะโนประชากร มีขนาดครัวเรือนคือ 1 คนและ 4 คนทำให้ Observed frequency ของขนาดครัวเรือน 1 คนและ 4 คนมีความถี่อย่างละ 1 เมื่อหาความแตกต่างของคุณลักษณะครัวเรือนจะได้ผลรวม (Sub-total) เท่ากับ 2

4.2 สำหรับคุณลักษณะอายุผู้อยู่อาศัย จากครัวเรือนที่สุ่มมาคือ d และ e คือ มีผู้อยู่อาศัยที่เป็นผู้ใหญ่ คือ 1 คนและ 2 คนตามลำดับ สำหรับผู้อาศัยที่เป็นเด็ก คือ 0 คนและ 1 คนตามลำดับ ทำให้ผู้อยู่อาศัยที่เป็นผู้ใหญ่รวมเป็น 3 คนและผู้อาศัยที่เป็นเด็กรวมเป็น 1 คน ทำให้ Estimated frequency ของอายุผู้อยู่อาศัยที่เป็นผู้ใหญ่คือ 3 คนและอายุผู้อยู่อาศัยที่เป็นเด็กคือ 1 สำหรับข้อมูลสำมะโนประชากร Observed frequency ที่ได้คือผู้อยู่อาศัยที่เป็นผู้ใหญ่ คือ 3 และผู้อาศัยที่เป็นเด็ก คือ 2 คน จากนั้นหาความแตกต่างของคุณลักษณะอายุของผู้อยู่อาศัยจะได้ผลรวม (Sub-total) เท่ากับ 1

4.3 รวมค่าความแตกต่างของทั้ง 2 คุณลักษณะ คือ 2 + 1 = 3

5. ทำขั้นตอนที่ 4 ซ้ำจนกว่าจะได้ค่าความแตกต่างที่มีค่าน้อยที่สุดที่จะเป็นไป เพื่อให้ได้ประชากรสังเคราะห์ที่ใกล้เคียงกับความเป็นจริงมากที่สุด ซึ่งผลสุดท้ายของตัวอย่างนี้คือการสุ่มเลือกครัวเรือน c แทนครัวเรือน d จาก survey microdata เป็นผลทำให้ค่าความแตกต่างมีค่าน้อยสุดคือ 0

2.2.2 Synthetic Reconstruction

Synthetic Reconstruction เป็นกระบวนการพื้นฐานที่ใช้ในการสร้างประชากรสังเคราะห์ที่มีข้อมูลตั้งต้น (input data) เป็น aggregate data ซึ่งก็คือ census data ในรูปแบบของตารางข้อมูลสถิติประชากรสำหรับคุณลักษณะต่างๆ เช่น อายุ เพศและ สถานภาพสมรส เป็นต้น [19] ซึ่งกระบวนการพื้นฐานของ Synthetic Reconstruction คือการสร้างประชากรทั้งหมดขึ้นมาก่อนแล้วค่อยๆ เพิ่มคุณลักษณะต่างๆ ที่มีเข้าไปทีละอย่างโดยอาศัยข้อมูลจากตารางสถิติประชากรและหลักการของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) [2]

สำหรับตัวอย่างของกระบวนการสร้างประชากรสังเคราะห์อย่างง่ายด้วยวิธี Synthetic Reconstruction จะแสดงในภาพที่ 5 โดยในขั้นตอนแรก ประชากรจะถูกประเมินคุณลักษณะของอายุ เพศ และสถานภาพสมรสเพื่อพิจารณาสถานะของการจ้างงานที่ขึ้นอยู่กับคุณลักษณะดังกล่าว ดังเช่นตัวอย่างนี้นั้นในขั้นตอนแรกจะพบว่าประชากรคนแรกคือผู้ชาย อายุ 18 ที่มีสถานภาพสมรสแล้ว ดังนั้นประชากรคนแรกจะถูกพิจารณาสถานะของการจ้างงานที่เหมาะกับค่าของคุณลักษณะเหล่านี้คือ มีการจ้างงาน 0.4% ว่างงาน 0.3% และไม่ได้ทำงาน 0.3% ต่อมาในขั้นตอนที่ 2-4 สถานะของการจ้างงานของบุคคลจะถูกนำมาคำนวณ โดยในขั้นตอนที่ 2 นั้นข้อมูลความน่าจะเป็นของบุคคลที่มีสถานะของการจ้างงานในแต่ละแบบที่เป็นไปได้จะถูกระบุ ซึ่งในตัวอย่างนี้ความน่าจะเป็นจะมี

เงื่อนไขที่ขึ้นอยู่กับกลุ่มของอายุและสถานภาพสมรส ดังนั้นความแตกต่างของแต่ละบุคคลในสามคนนี้จะถูกแสดงดังรูป โดยที่ความน่าจะเป็นจะถูกแปลงให้อยู่ในรูปชุดของความน่าจะเป็นแบบสะสม ‘bins’ และในขั้นตอนที่ 3 และ 4 นั้นการสุ่มสร้างข้อมูลจะถูกสร้างขึ้นตามความน่าจะเป็นแบบสะสม ‘bin’ โดยตัวเลขที่ได้จากการสุ่มจะอยู่ในการกำหนดสถานะของการจ้างงานของบุคคลนั้นที่ถูกคำนวณขึ้น จากตัวอย่างพบว่าประชากรคนแรกมีค่าที่ได้จากการสุ่มคือ 0.281 ซึ่งตกอยู่ในความน่าจะเป็นแบบสะสมอันแรกสุดของ ‘bin’ (0.0-0.4) ดังนั้น จึงได้สถานะมีการจ้างงานแล้ว

	Person		
Steps	1st	2nd	Last
1. Age, sex and marital status of person[a]	Age: **18** Sex: **Male** M: **Married**	Age: 34 Sex: Male M: Married	Age: 87 Sex: Male M: Married
2. Probability of employment status given age, sex and marital status *E: employed* *U: unemployed* *I: inactive*	*Prob* *Cum Prob Bin* E: 0.4 (0.0-0.4) U: 0.3 (>0.4-0.7) I: 0.3 (>0.7-1.0)	*Prob* *Cum Prob Bin* E: 0.6 (0.0-0.6) U: 0.3 (>0.6-0.9) I: 0.1 (>0.9-1.0)	*Prob* *Cum Prob Bin* E: 0.0 (0.0-0.0) U: 0.0 (0.0-0.0) I: 1.0 (>0.0-1.0)
3. Random number *(computer generated)*	0.281	0.709	0.481
4. Employment status assigned on basis of random sampling	Employed	Unemployed	Inactive
5. Next person *(repeat until all persons have been assigned an employment status)*	*Move on to next person*	*Move on to next person*	*End*

ที่มา : [19]

ภาพที่ *5* ตัวอย่าง synthetic reconstruction approach

เนื่องจากกระบวนการสร้างประชากรสังเคราะห์ทั้ง 2 วิธีมีความแตกต่างกัน จึงสามารถสรุปความแตกต่างได้ [20] ดังแสดงในตารางที่ 8

ตารางที่ 8 ความแตกต่างของกระบวนการสร้างประชากรสังเคราะห์

Combinatorial Optimization	Synthetic Reconstruction
- เป็นกระบวนการที่ทำซ้ำ : ที่มีความเหมาะสมระหว่างข้อมูลจริงและตัวอย่างที่ถูกเลือกจากการสำรวจ ซึ่งควรจะได้รับการจากลดค่าของความคลาดเคลื่อน	- ขึ้นอยู่กับขั้นตอนที่มีความต่อเนื่องตามกระบวนการของขั้นตอน : โดยที่คุณลักษณะของแต่ละตัวอย่างจะถูกประมาณโดยการสุ่มที่ใช้ความน่าจะเป็นที่มีเงื่อนไข
- ลำดับการสร้างไม่มีความจำเป็น เนื่องจากมีการทำซ้ำของกระบวนการหลายครั้ง	- ลำดับการสร้างมีความจำเป็นในกระบวนการนี้ ซึ่งแต่ละค่าจะถูกสร้างโดยการกำหนดลำดับ
- มีความซับซ้อนในมุมมองของทฤษฎี และใช้เวลาน้อยกว่า	- ค่อนข้างซับซ้อน และใช้เวลานาน
- สามารถช่วยให้เลือกตารางของข้อมูลให้ตรงกับรูปแบบการวิจัยและ/หรือความต้องการของผู้ใช้	- ผลกระทบของความไม่สอดคล้องระหว่างตารางข้อมูลอาจจะมีนัยสำคัญกับกระบวนการนี้ เนื่องจากเกิดความไม่ตรงกันของผลรวมหรือผลรวมย่อยของตาราง

ที่มา : [20]

2.3 การประเมินความถูกต้องของแบบจำลอง (Model Validation)

การประเมินความถูกต้องของกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ที่ถูกสร้างขึ้นนั้นเป็นการประเมินของข้อมูลที่ได้สร้างขึ้นกับข้อมูลต้นแบบว่ามีความคลาดเคลื่อนกันมากน้อยเพียงใด เพื่อที่จะนำมาใช้วัดประสิทธิภาพของกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบว่ามีความถูกต้องและน่าเชื่อถือเพียงใด ตลอดจนสังเกตความคลาดเคลื่อนของข้อมูลที่ถูกสร้างขึ้น โดยหากมีความคลาดเคลื่อนน้อยก็แสดงว่ากระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบนั้นมีความถูกต้องแม่นยำสูงและมีความเหมาะสมที่จะนำมาใช้ในการสร้างประชากรได้อย่างแม่นยำ โดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (mean absolute error) [2, 5, 6, 17] ในการประเมินความถูกต้องของแบบจำลอง โดยค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์เป็นการหาผลต่างของข้อมูลที่เป็น

ข้อมูลต้นแบบกับข้อมูลที่ถูกสร้างขึ้น ซึ่งสามารถคำนวณค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ได้ในแต่ละเซลล์ของตารางได้ดังนี้

$$\textit{Mean Absolute Error} = \frac{\sum|T' - T|}{n}$$

โดยที่ T’ คือ เปอร์เซ็นต์ของคุณลักษณะของตารางข้อมูลที่ถูกสร้างขึ้น

T คือ เปอร์เซ็นต์ของคุณลักษณะของตารางข้อมูลต้นแบบ

n คือ จำนวนเซลล์ของตาราง

2.4 เอกสารและงานวิจัยที่เกี่ยวข้อง

2.4.1 ข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์

เนื่องจากข้อมูลที่จะนำมาใช้ในการสร้างประชากรสังเคราะห์นั้นมีความสำคัญต่อกระบวนการสร้างประชากร จึงได้มีงานวิจัยที่ให้ความสำคัญเกี่ยวกับการจัดการข้อมูลก่อนที่จะนำไปใช้สร้างประชากรสังเคราะห์ ดังนี้

งานวิจัย [21] ได้ให้ความหมายและกำหนดระดับคุณภาพของข้อมูลโดยจะขึ้นอยู่กับความสามารถในการใช้แหล่งข้อมูลนั้นในการสร้างประชากรสังเคราะห์ ซึ่งระดับคะแนนนี้จะสะท้อนให้เห็นถึงการคาดคะเนความยากลำบากในการสร้างประชากรสังเคราะห์ให้มีความน่าเชื่อถือโดยแบ่งได้ดังนี้

- พื้นที่ที่ได้รับคะแนน 0 คือพื้นที่ที่มีข้อมูลที่ไม่สามารถนำมาใช้สำหรับการสร้างประชากรสังเคราะห์ได้
- พื้นที่ที่ได้รับคะแนน 1 คือพื้นที่ที่มีคุณภาพของข้อมูลต่ำกว่าค่าเฉลี่ยซึ่งข้อมูลของประเทศเหล่านี้มักมาจากการไม่มีการจัดทำสำมะโนประชากรหรืออาจมีการจัดทำสำมะโนประชากรแต่ไม่บ่อยและยังหมายรวมไปถึงพื้นที่ที่มีความขัดแย้งด้วย
- พื้นที่ที่ได้รับคะแนน 2 นั้นเป็นพื้นที่ที่มีคุณภาพของข้อมูลเป็นคุณภาพเฉลี่ยและข้อมูลทั่วไปที่สามารถนำไปใช้ได้โดยมักพบในประเทศที่กำลังพัฒนา
- พื้นที่ที่ได้รับคะแนน 3 คือพื้นที่ที่มีการจัดทำสำมะโนประชากรเป็นประจำทุกปีด้วยความถูกต้องและแม่นยำ

ซึ่งในงานวิจัยนี้ออกแบบกำหนดระดับคะแนนนี้กับข้อมูลใน 6 ด้าน คือ การเมือง การทหาร เศรษฐกิจ สังคม สารสนเทศและด้านโครงสร้างพื้นฐานประชากร ทำให้สามารถออกแบบงานวิจัยได้ตามคุณภาพของข้อมูล ด้วยเหตุนี้เองทางผู้จัดทำจะนำการกำหนดระดับคุณภาพของข้อมูลนี้ไปประกอบเป็นส่วนหนึ่งในศึกษาความละเอียดของแต่ละคุณลักษณะที่ใช้ในการสร้างประชากรสังเคราะห์ที่จะถูกเก็บมาจากหลายแหล่งข้อมูลซึ่งมีผลทำให้ข้อมูลจากแต่ละแหล่งมีความละเอียดและความน่าเชื่อถือแตกต่างกัน อีกทั้งยังสามารถบอกความถูกต้องได้จากความสอดคล้องกันของข้อมูลในการสร้างประชากรสังเคราะห์ของแต่ละคุณภาพของข้อมูลได้อีกด้วย

งานวิจัย [17] นำเสนอการคาดคะเนขนาดของครัวเรือนโดยใช้การแจกแจงแบบปัวซอง (Poisson distribution) จากพื้นฐานปัญหาที่พบว่าในข้อมูลที่สามารถเปิดเผยได้ซึ่งคือ census data นั้นมีความไม่สอดคล้องกัน เช่น ข้อมูลจำนวนประชากรและข้อมูลขนาดครัวเรือน โดยเมื่อนำขนาดครัวเรือนมาคูณกับจำนวนครัวเรือนในแต่ละขนาดทำให้ได้จำนวนประชากรทั้งหมด ซึ่งพบว่ามีค่าไม่เท่ากับจำนวนประชากรที่ได้เปิดเผยไว้ ทำให้การนำข้อมูลประเภทนี้ไปใช้มีความน่าเชื่อถือน้อยลง งานวิจัยนี้จึงได้คาดคะเนขนาดของครัวเรือนใหม่เพื่อทำให้จำนวนประชากรที่ได้จากการหาค่าจากงานวิจัยมีค่าใกล้เคียงกับจำนวนของประชากรจริงที่ได้รับข้อมูลจากข้อมูลที่มีการเปิดเผยไว้มากที่สุด โดยใช้การแจกแจงแบบปัวซองดังกล่าวไว้ในหัวข้อที่ 2.1.3.2 ซึ่งจะมีการปรับปรุงค่าเฉลี่ย (mean) กำหนดให้ M = Avg.HH Size -1 โดยที่ Avg.HH Size คือค่าเฉลี่ยของขนาดครัวเรือน ดังแสดงตัวอย่างในตารางที่ 9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 9 ตัวอย่างการปรับปรุงการแจกแจง

Avg HH Size	Modified Mean (M = Avg HH Size - 1)	Size 1 → n = 0	Size 2 → n = 1	Size 3 → n= 2	etc...
2.0	1.0	f(0,1) = 37%	f(1,1) = 37%	f(2,1)=18%	...
3.5	2.5	f(0,2.5) = 8%	f(1,2.5) = 21%	...	...
4.0	3.0	f(0,3.0) = 5%	...	...	...

ที่มา : [17]

เนื่องจากงานวิจัยของผู้จัดทำได้ประสบปัญหาด้านข้อมูลในลักษณะเดียวกันกับงานวิจัยชิ้นนี้คือจำนวนประชากรที่ได้จากการจัดทำสำมะโนประชากร (census data) มีค่าไม่เท่ากับผลรวมของขนาดของครัวเรือนคูณกับจำนวนครัวเรือนในแต่ละขนาด ซึ่งเป็นปัญหาของความไม่สอดคล้องกันในการเก็บข้อมูล ผู้วิจัยจะนำวิธีการคาดคะเนขนาดของครัวเรือนโดยใช้การแจกแจงแบบปัวซองมาใช้เพื่อปรับปรุงข้อมูลสำหรับการสร้างประชากรสังเคราะห์ระดับครัวเรือนให้มีความน่าเชื่อถือมากยิ่งขึ้น

2.4.2 กระบวนการสร้างประชากรสังเคราะห์

เนื่องจากกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ต้องออกแบบให้เหมาะสมกับข้อมูล เพื่อที่จะได้ประชากรสังเคราะห์ที่มีความถูกต้องแม่นยำและมีความน่าเชื่อถือ ซึ่งมีงานวิจัยที่ให้ความสำคัญกับการออกแบบกระบวนการในการสร้างประชากรสังเคราะห์ ดังนี้

งานวิจัย [5] ได้ทดลองสร้างประชากรสังเคราะห์ในประเทศอังกฤษโดยใช้ข้อมูลจากการจัดทำสำมะโนประชากรคือ 1991 UK Census Small Area Statistics (SAS) และข้อมูลของประชากรกลุ่มตัวอย่าง British Household Panel Survey (BHPS) โดยมีจุดมุ่งหมายเพื่อใช้ในการจำลองประชากรทั้งหมดในแต่ละพื้นที่ของประเทศอังกฤษภายในปี 2021 และเพื่อวิเคราะห์นโยบายต่างๆ ที่อาจจะเกิดขึ้นโดยใช้กระบวนการ Reweighting ซึ่งจะเริ่มต้นจากการกำหนดค่าน้ำหนัก (weight) ที่เหมาะสมให้แต่ละพื้นที่ได้และนำค่าของคุณลักษณะของข้อมูลในด้านต่างๆ มาใช้ในการปรับค่าน้ำหนักโดยการคำนวณน้ำหนักใหม่ในแต่ละรอบจนได้ค่าที่เหมาะสมที่จะใช้สร้างประชากรของทั้งประเทศ

ขณะที่งานวิจัย [22] นำเสนอกระบวนการที่ใช้ในการสร้างประชากรในระดับครัวเรือน โดยใช้ข้อมูลการสำรวจการเดินทางและตรวจสอบความถูกต้องของข้อมูลโดยการสร้างประชากรสำหรับนครนิวยอร์ก ประเทศสหรัฐอเมริกา ซึ่งคุณลักษณะของข้อมูลที่สนใจคือ ขนาดครัวเรือน รายได้ของครัวเรือน จำนวนยานพาหนะ จำนวนคนทำงานในครัวเรือน การอาศัยอยู่กับเด็ก อายุของเจ้าของบ้าน ซึ่งมีการใส่ค่าน้ำหนักในแต่ละครัวเรือนและปรับค่าน้ำหนักโดยการคำนวณน้ำหนักใหม่ในแต่ละรอบซึ่งใช้คุณลักษณะของข้อมูลดังกล่าวเพื่อให้ได้ข้อมูลที่มีค่าน้ำหนักที่ดีที่สุดในการสร้างประชากรสังเคราะห์

จากงานวิจัยที่กล่าวมานั้นพบว่าการใช้กระบวนการ Combinatorial Optimization ยังมีข้อจำกัดในด้านข้อมูลของประชากรกลุ่มตัวอย่างที่ได้จากการสำรวจอยู่ คือ หากข้อมูลที่ได้จากการสำรวจมีน้อยเกินไปจะทำให้การสร้างประชากรที่ได้นั้นเป็นประชากรที่มีลักษณะคล้ายกันมากเกินความเป็นจริง ทำให้ค่าที่จะนำมา Reweight มีความบิดเบือนไปจากความเป็นจริง ส่งผลให้ข้อมูลประชากรสังเคราะห์ที่สร้างขึ้นมีความน่าเชื่อถือน้อยและไม่เหมาะสมที่จะใช้กับประเทศไทย เนื่องจากประเทศไทยยังไม่มีการเผยแพร่ข้อมูลที่ได้จากการสำรวจอย่างเพียงพอ

งานวิจัย [6] ได้นำเสนอ SMILE model (Simulation Model for the Irish Local Economy model) ซึ่งเป็นการสร้างประชากรสังเคราะห์สำหรับวิเคราะห์ผลกระทบของการเปลี่ยนแปลงทางนโยบายและการพัฒนาเศรษฐกิจบนพื้นที่ในเขตชนบทของประเทศไอร์แลนด์ โดยสนใจในด้านการเปลี่ยนแปลงประชากร ภาวะการเจริญพันธุ์ อัตราการเสียชีวิต และการย้ายถิ่นฐาน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภายในที่พื้นที่ที่มีขนาดเล็ก เพื่อวางแผนในด้านการเปลี่ยนแปลงของประชากรในระดับย่อยของประเทศ ข้อมูลต้นแบบ (input data) ได้มาจาก Central Statistics Office (CSO) ซึ่งจะใช้การสุ่มตัวอย่างแบบ Monte Carlo เพื่อสร้างประชากรสังเคราะห์ โดยจะพิจารณาคุณลักษณะของอัตราการเสียชีวิต ซึ่งพบว่าความน่าจะเป็นของการมีชีวิตอยู่นั้นขึ้นอยู่กับเพศ อายุและสถานที่ตั้งของประเทศ จากนั้นจึงสุ่มเลือกตัวเลขเพื่อกำหนดว่าบุคคลนั้นยังมีชีวิตอยู่หรือไม่ เช่น คนแรก จะมีชีวิตอยู่เมื่อสุ่มได้ตัวเลขตั้งแต่ 0.00 – 0.80 ถ้าสุ่มได้ตัวเลขที่มากกว่า 0.80 จนถึง 1.00 จะถูกกำหนดว่าเสียชีวิตแล้ว ดังตัวอย่างที่แสดงในภาพที่ 6

ซึ่งงานวิจัยนี้เป็นกระบวนการที่เน้นการสร้างประชากรในระดับบุคคลเท่านั้น ไม่ได้ให้ความสำคัญในการสร้างประชากรในระดับครัวเรือน จึงทำให้ประชากรสังเคราะห์ที่สร้างขึ้นไม่มีความสัมพันธ์ระหว่างกันในระดับบุคคลเพื่อก่อให้เกิดประชากรสังเคราะห์ในระดับครัวเรือน

Steps	1st	2nd	...	Last
Age, sex and marital status, employment status and location (DED level) (given)	Age: 25 Sex: Male Marital status: Single Employment status: At work GeoCode: Leitrim Co., DED 001 Ballinamore	Age: 76 Sex: Female Marital status: Married Employment status: Other (e.g. retired) GeoCode: Leitrim Co., DED 002 Cloverhill	...	Age: 30 Sex: Male Marital status: Married Employment status: At work GeoCode: Leitrim Co., DED 078 Rowan
Probability (conditional on sex, age and county location) of survival	0.80	0.10	...	0.80
Random number	0.5	0.9	...	0.6
Mortality outcome	Alive	Dead	...	Alive

ที่มา : [6]

ภาพที่ 6 ตัวอย่างกระบวนการสร้างประชากรอย่างง่ายสำหรับคุณลักษณะการมีชีวิตอยู่หรือเสียชีวิต

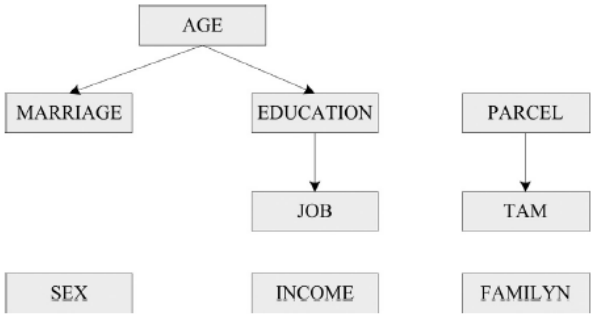
ขณะที่งานวิจัย [15] เสนอกระบวนการที่ใช้สร้างประชากรระดับบุคคลและระบุสถานที่อยู่อาศัยโดยปราศจากข้อมูลของประชากรกลุ่มตัวอย่างที่ได้จากการสำรวจของประชากรในเขตปริมณฑลของปักกิ่ง Beijing Metropolitan Area (BMA) ซึ่งงานวิจัยนี้มีการกำหนดลำดับการสร้างของคุณลักษณะของข้อมูลดังกล่าวไว้ในหัวข้อ 2.1.1 และมีการศึกษาความสัมพันธ์ของข้อมูลตามหัวข้อ 2.1.4 เพื่อสร้างประชากรสังเคราะห์ ซึ่งมีรายละเอียดของคุณลักษณะของข้อมูลดังแสดงในตารางที่ 10 และความสัมพันธ์ของคุณลักษณะของข้อมูลแสดงในภาพที่ 7

จากข้อมูลที่แสดงในตารางและภาพดังกล่าวจะเห็นได้ว่าเมื่อมีการกำหนดความสัมพันธ์ของคุณลักษณะของข้อมูลทำให้การสร้างประชากรระดับบุคคลมีความน่าเชื่อถือมากยิ่งขึ้น แต่อย่างไรก็ตาม งานวิจัยนี้ได้มุ่งเน้นไปที่การสร้างประชากรระดับบุคคลและการจัดหาสถานที่อยู่อาศัยให้กับประชากรในระดับบุคคลและไม่ได้ให้ความสำคัญกับความสัมพันธ์ของประชากรในครัวเรือน ซึ่งมีการระบุเพียงจำนวนประชากรในแต่ละครัวเรือนเท่านั้นจึงทำให้ประชากรระดับบุคคลนั้นไม่มีความสัมพันธ์กัน

ตารางที่ 10 รายละเอียดของข้อมูลประชากรของปริมณฑลปักกิ่ง

Name	Description	Type	Known information	Data source	Data type	Order
AGE	Age in years	Non-spatial attribute	Frequencies	The census	Ratio	1
SEX	Gender	Non-spatial attribute	Frequencies	The census	Nominal (male, female)	2
MARRIAGE	Marital status	Non-spatial attribute	Frequencies, RB (with AGE)	The census	Nominal (married, unmarried, divorced, remarried, widowed)	3
EDUCATION	Level of education	Non-spatial attribute	Frequencies, RB (with AGE)	The census	Ordinal (junior middle school, undergraduate, etc.)	4
JOB	Occupation	Non-spatial attribute	Frequencies, RB (with EDUCATION)	The census	Nominal	5
INCOME	Monthly income	Non-spatial attribute	Frequencies	The survey	Ratio	6
FAMILYN	Number of family members	Non-spatial attribute	Frequencies	The census	Ordinal (one person, two person, etc.)	7
PARCEL	ID of parcel at which the agent resides	Location	Frequencies	An empirical study	Nominal	8
TAM	Distance to the city center	Spatial attribute	Location of Tiananmen Square	Urban GIS	Ratio	9

ที่มา : [15]

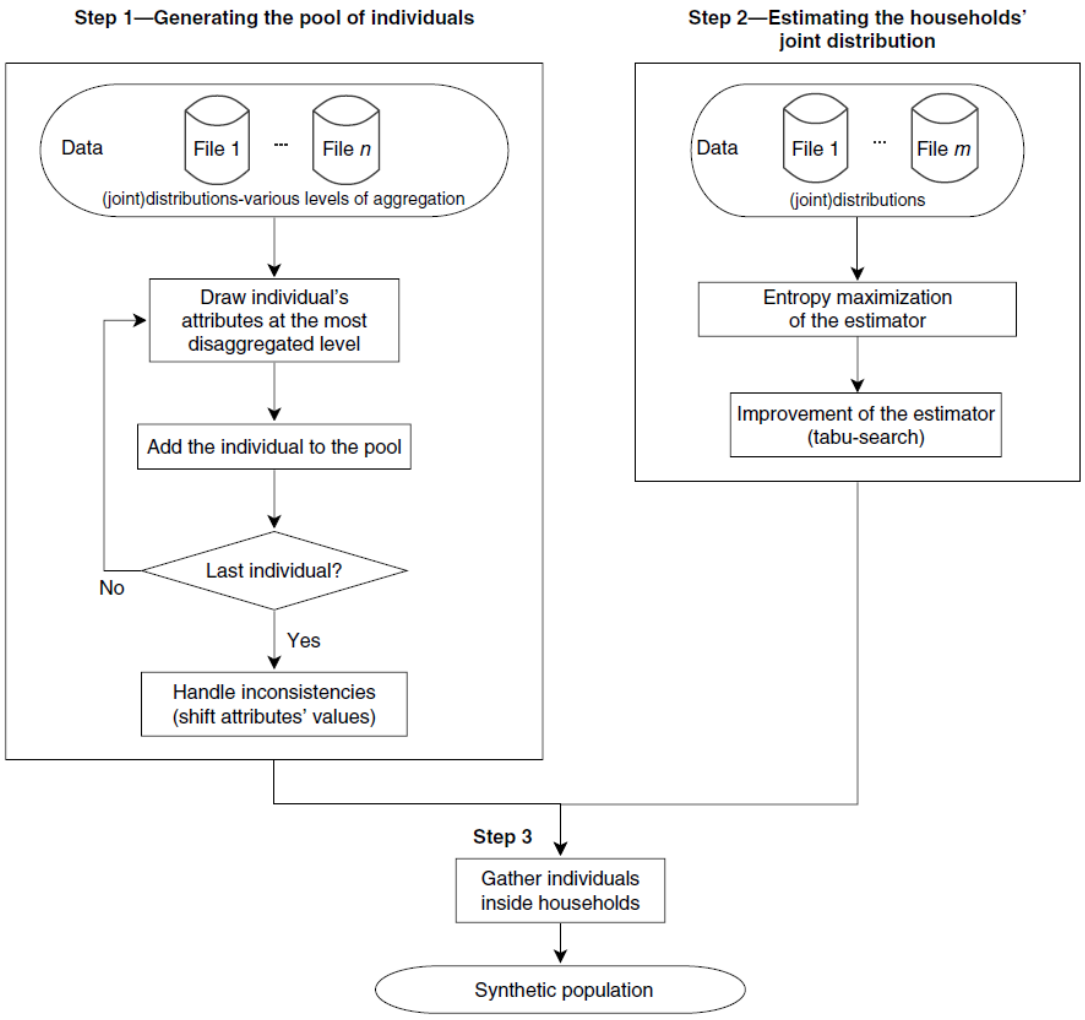


ที่มา : [15]

ภาพที่ 7 ความสัมพันธ์ของคุณลักษณะของข้อมูลประชากร

สำหรับงานวิจัย [4] ได้เสนอกระบวนการในการสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและครัวเรือนสำหรับประชากรในประเทศเบลเยียม โดยไม่ใช้ข้อมูลของประชากรกลุ่มตัวอย่างที่ได้จากการสำรวจและกระบวนการที่ถูกออกแบบขึ้นได้มีการแก้ปัญหาข้อมูลไม่สอดคล้องกันเนื่องจากถูกเก็บมาจากหลายแหล่งข้อมูลด้วย ซึ่งกระบวนการสร้างประชากรนั้นได้แบ่งออกเป็น 3 ขั้นตอนโดยในขั้นตอนแรกจะสร้างประชากรระดับบุคคลจากข้อมูลที่มีทั้งหมดและจัดการกับความไม่สอดคล้อง

กันของข้อมูล ซึ่งมีคุณลักษณะของข้อมูลที่ใช้ในงานวิจัยนี้ในระดับบุคคลดังแสดงในตารางที่ 11 ส่วนในขั้นตอนที่ 2 นั้นจะคาดคะเนการแจกแจงร่วมของครัวเรือน (Household's joint distribution) ที่มีคุณลักษณะดังตารางที่ 12 โดยมีการสร้างการแจกแจงของครัวเรือนจากข้อมูล census data และปรับปรุงให้เหมาะสม เพื่อเป็นตารางข้อมูลสำหรับใช้สร้างประชากรสังเคราะห์ในระดับครัวเรือน (contingency table) ซึ่งจะเกิดขึ้นในขั้นตอนสุดท้ายคือการสร้างประชากรในระดับครัวเรือน โดยจะใช้ตาราง contingency table เพื่อเลือกคุณลักษณะของประเภทของครัวเรือนก่อนแล้วจึงเลือกประชากรระดับบุคคลที่เก็บไว้ในขั้นตอนแรกโดยการสุ่มเพื่อจัดเข้ากลุ่มตามคุณลักษณะของประเภทของครัวเรือนดังแสดงในภาพที่ 8



ที่มา : [4]

ภาพที่ 8 กระบวนการสร้างประชากรสังเคราะห์

ตารางที่ 11 คุณลักษณะข้อมูลของระดับบุคคล

Attribute	Values
Gender	Male; female
Age class	0–5; 6–17; 19–39; 40–59; 60+
Activity	Student; active; inactive
Education level	Primary; high school; higher education; none
Driving license ownership	Yes; no

ที่มา : [4]

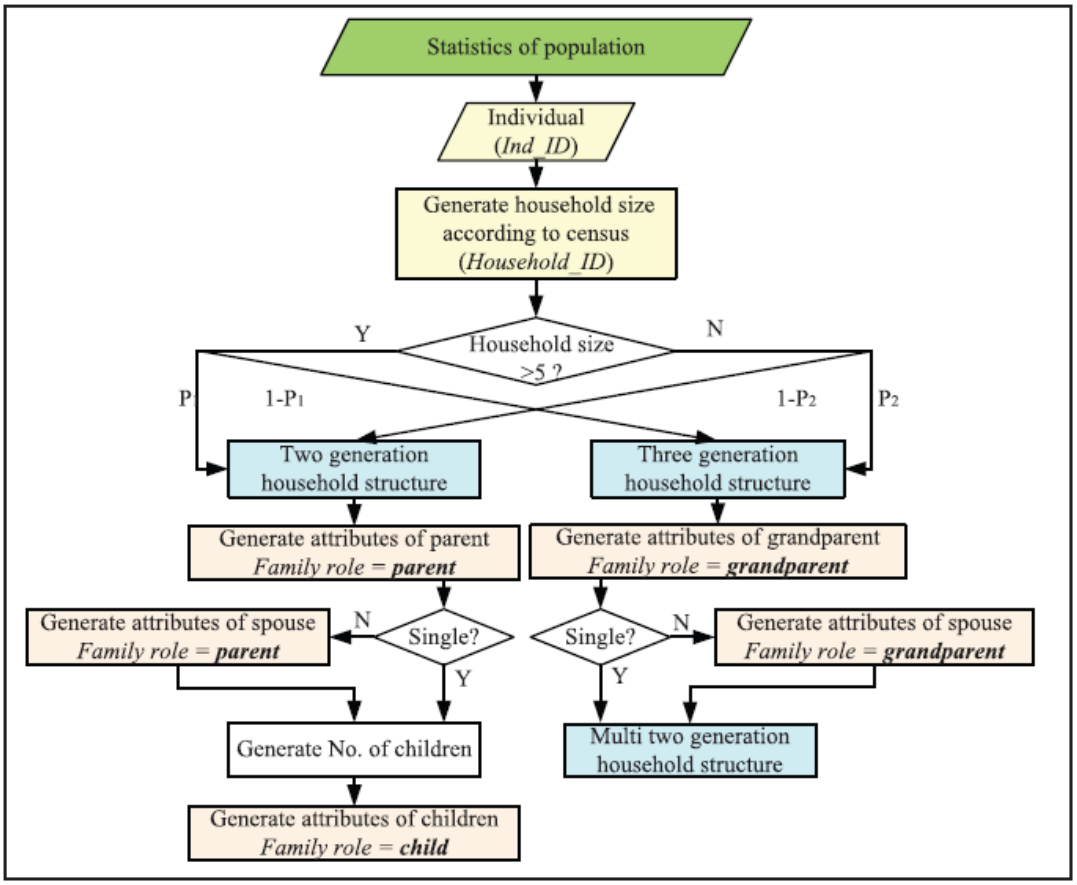
ตารางที่ 12 คุณลักษณะข้อมูลของระดับครัวเรือน

Attribute	Values
Type	Single man alone Single woman alone Single man with children (and other adults) Single woman with children (and other adults) Couple without children (and other adults) Couple with children (and other adults)
Number of children	0 to 5
Number of other adults	0 to 2 (mate not included)

ที่มา : [4]

เนื่องจากงานวิจัยนี้มีกระบวนการสร้างประชากรสังเคราะห์ที่น่าสนใจ แต่การสร้างประชากรสังเคราะห์ในระดับครัวเรือน ส่งผลให้ประชากรสังเคราะห์ระดับบุคคลมีความคลาดเคลื่อนไปด้วย อีกทั้งการใช้คุณลักษณะของข้อมูลประชากรไม่หลากหลาย ทำให้ประชากรสังเคราะห์ที่สร้างได้ไม่มีความละเอียดเพียงพอซึ่งอาจไม่ตรงกับประชากรที่มีอยู่จริงเท่าที่ควร

ขณะที่งานวิจัย [23] ได้คิดค้นกระบวนการสร้างเมืองจำลองสำหรับเมืองปักกิ่ง โดยใช้วิธี Synthetic Reconstruction จากการใช้ประโยชน์จากข้อมูลทางสถิติและสภาพทางภูมิศาสตร์ ซึ่งมีการสร้างใน 4 ด้าน คือ แบบจำลองบุคคล (individual model), แบบจำลองด้านสิ่งแวดล้อม (environment model), แบบจำลองการเดินทาง (travel model) และแบบจำลองของความสัมพันธ์ระหว่างบุคคล (social networks) สำหรับแบบจำลองบุคคลจะมีการกำหนด Family role = { ตา/ปู่, ยาย/ย่า, พ่อ, แม่, ลูก }, Social role = {ทารก, นักเรียน ,คนทำงาน, ผู้สูงอายุ, คนว่างงาน} และมีการกำหนดตัวแปร AHW คือ อายุห่างระหว่างสามี-ภรรยา, AFC คือ อายุห่างระหว่างมารดาและบุตรคนแรก และ ACC คืออายุห่างระหว่างพี่และน้อง สำหรับกระบวนการที่ใช้ในการสร้างโครงสร้างของครัวเรือนของงานวิจัยนี้แสดงในภาพที่ 9 ซึ่งอธิบายรายละเอียดได้ดังนี้



ที่มา : [23]

ภาพที่ *9* กระบวนการสร้างโครงสร้างของครัวเรือน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. สร้างประชากรระดับบุคคลจากข้อมูลทางสถิติ
2. สร้างขนาดของครัวเรือนตามข้อมูลที่ได้จากการทำสำมะโนประชากร ซึ่งแบ่งออกเป็น 2 ขั้นตอนย่อย คือ

   2.1 หากขนาดครัวเรือนน้อยกว่า 5 จะเป็นครอบครัวที่มี 2 รุ่น คือ รุ่นพ่อแม่ และรุ่นลูก

      2.1.1 กำหนด Family role เป็นพ่อหรือแม่ จากนั้นดูว่าอยู่คนเดียวหรือไม่

         2.1.1.1 หากอยู่คนเดียวให้สร้างจำนวนบุตรและใส่ข้อมูลคุณลักษณะของบุตรพร้อมทั้งกำหนด Family role = เด็ก

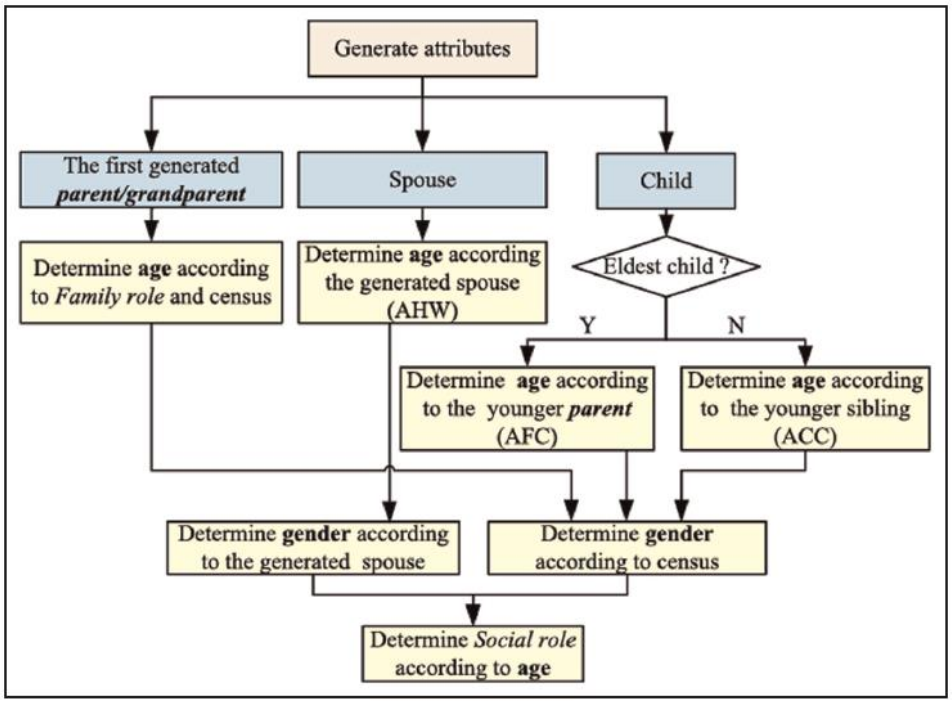
         2.1.1.2 หากไม่ได้อยู่คนเดียวให้สร้างคู่สมรสและใส่ข้อมูลคุณลักษณะของคู่สมรส จากนั้นสร้างจำนวนบุตรและใส่ข้อมูลคุณลักษณะของบุตรพร้อมทั้งกำหนด Family role = เด็ก

2.2 หากขนาดครัวเรือนมากกว่า 5 จะเป็นครอบครัวที่มี 3 รุ่น คือ รุ่นปู่ย่า/ตายาย รุ่นพ่อแม่ และรุ่นลูก

2.2.1 กำหนด Family role เป็นปู่หรือย่า/ตาหรือยาย จากนั้นดูว่าอยู่คนเดียวหรือไม่

2.2.1.1 หากอยู่คนเดียว ให้สร้างครอบครัวอีก 2 รุ่น ตามขั้นตอน 2.1.1

2.2.1.2 หากไม่ได้อยู่คนเดียว ให้สร้างคู่สมรสและใส่ใส่ข้อมูลคุณลักษณะของคู่สมรส จากนั้นสร้างครอบครัวอีก 2 รุ่น ตามขั้นตอน 2.1.1



ที่มา : [23]

ภาพที่ *10* กระบวนการสร้างประชากรสังเคราะห์ในระดับครัวเรือน

สำหรับกระบวนการสร้างประชากรในระดับครัวเรือนที่นำเสนอในงานวิจัยนี้ดังแสดงในภาพที่ 10 ซึ่งสามารถอธิบายรายละเอียดได้ดังนี้คือ

1. การสร้างประชากรครั้งแรกจะสร้างพ่อแม่หรือปู่ย่า/ตายายโดยกำหนดอายุให้ตาม Family role และจากข้อมูลที่ได้จากการจัดทำสำมะโนประชากร จากนั้นกำหนดเพศให้ตามข้อมูลสำมะโนประชากรและกำหนด Social role ตามอายุที่สร้างขึ้น

2. สร้างคู่สมรสซึ่งกำหนดอายุจาก AHW จากนั้นกำหนดเพศให้โดยให้เพศตรงข้ามกับคู่สมรสที่ได้สร้างไว้แล้วและกำหนด Social role ตามอายุที่สร้างขึ้น

3. สร้างบุตร โดยแบ่งเป็น 2 ขั้นตอนย่อย คือ

   3.1 หากเป็นบุตรคนแรกจะกำหนดอายุให้โดยดูจาก AFC จากนั้นกำหนดเพศให้ตามข้อมูลสำมะโนประชากรและกำหนด Social role ตามอายุที่สร้างขึ้น

   3.2 หากไม่ใช่บุตรคนแรก จะกำหนดอายุให้โดยดูจาก ACC จากนั้นกำหนดเพศให้ตามข้อมูลสำมะโนประชากรและกำหนด Social role ตามอายุที่สร้างขึ้น

นอกจากนี้ยังมีการสร้างแบบจำลองด้านสิ่งแวดล้อมซึ่งสามารถบอกถึงสถานที่ที่บุคคลเดินทางไป โดยศึกษาจากพฤติกรรมประจำวันของมนุษย์ เช่น การนอน การศึกษา การทำงาน การท่องเที่ยว การรับประทานอาหาร การจับจ่ายใช้สอย การพักผ่อนและการเข้ารับการรักษาตัวในโรงพยาบาล ซึ่งได้แบ่งสถานที่ที่สนใจออกเป็นบ้าน สถานศึกษา ที่ทำงาน สถานที่สำหรับจับจ่ายใช้สอย เช่น ตลาด ห้างสรรพสินค้า สถานที่ให้ความบันเทิง เช่น โรงหนัง ร้านอาหารและสถานพยาบาล โดยจะมีการระบุข้อมูลตำแหน่งของสถานที่เหล่านี้ลงในแผนที่ของเมืองด้วย ในแบบจำลองการเดินทางจะมีความสัมพันธ์ระหว่างบุคคลและสถานที่ต่างๆ ซึ่งในงานวิจัยนี้ให้ความสนใจการเดินทางไปโรงเรียนของเด็กที่ศึกษาในระดับอนุบาล ประถมศึกษา มัธยมศึกษาและมหาวิทยาลัยรวมถึงการเดินทางไปที่ทำงานของประชากรที่อยู่ในวัยทำงาน และสำหรับแบบจำลองของความสัมพันธ์ระหว่างบุคคลจะให้ความสำคัญของความสัมพันธ์ระหว่างครอบครัว เพื่อนร่วมห้องเรียน เพื่อนร่วมงานและเพื่อนบ้าน ซึ่งกระบวนการสร้างเมืองจำลองของงานวิจัยนี้ดังแสดงในภาพที่ 11 และสามารถอธิบายรายละเอียดได้ดังนี้

1. สร้างประชากรโดยใช้แบบจำลองของบุคคลด้วยข้อมูลการทำสำมะโนประชากรและข้อมูลของอายุห่างระหว่างสามี-ภรรยา (AHW), อายุห่างระหว่างมารดาและบุตร (AFC) และอายุห่างระหว่างพี่น้อง (ACC)

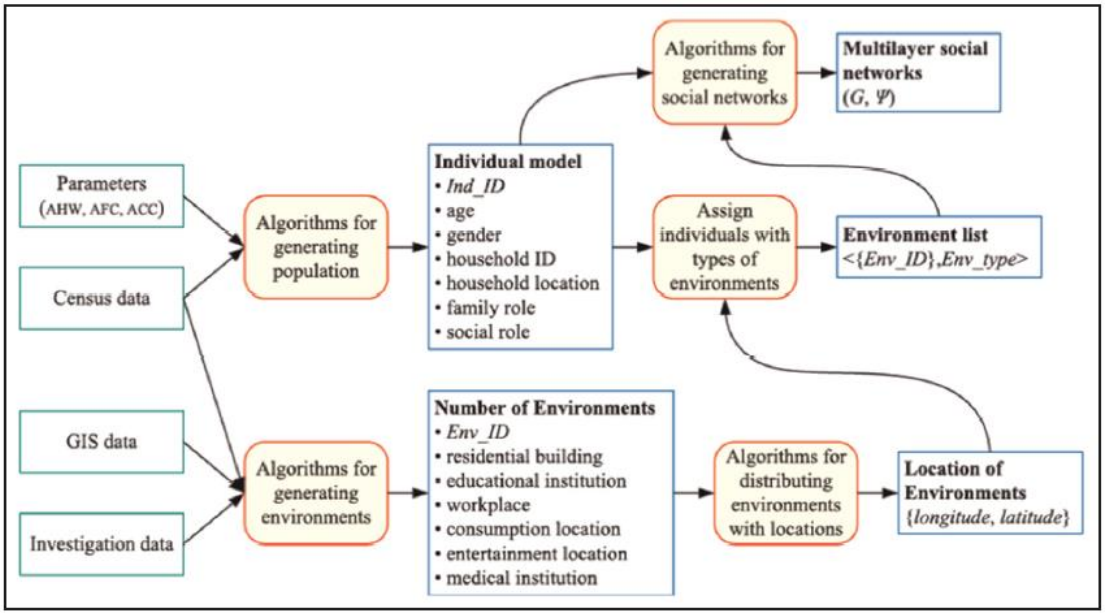
2. สร้างสถานที่ต่างๆ โดยใช้แบบจำลองด้านสิ่งแวดล้อมซึ่งได้จากข้อมูลของระบบสารสนเทศภูมิศาสตร์และข้อมูลสำหรับการตรวจสอบ จากนั้นระบุตำแหน่งบนแผนที่ให้แต่ละสถานที่ด้วยละติจูดและลองจิจูด

3. กำหนดบุคคลที่จะไปสถานที่ต่างๆ จากข้อมูลในขั้นตอนที่ 1 และ 2 เช่น นักเรียนไปโรงเรียน หรือประชากรวัยทำงานไปที่ทำงาน เป็นต้น

Chulalongkorn University

4. สร้างความสัมพันธ์ระหว่างบุคคล จากข้อมูลในขั้นตอนที่1 และ 3 เช่น นักเรียนมีความสัมพันธ์แบบเพื่อนร่วมห้อง ครูมีความสัมพันธ์แบบเพื่อนร่วมงาน เป็นต้น

อย่างไรก็ตามในงานวิจัยนี้ยังมีข้อจำกัดคือประชากรสังเคราะห์ที่ถูกสร้างขึ้นนั้นยังเป็นประชากรที่มีความหลากหลายของครัวเรือนน้อย และเป็นครัวเรือนที่ไม่มีความซับซ้อน กล่าวคือแต่ละครอบครัวที่ถูกสร้างขึ้นมาเป็นครอบครัวที่สมบูรณ์แบบ คือประกอบด้วยพ่อ แม่ ลูก หรือประกอบด้วยปู่ ย่า หรือ ตา ยาย และ พ่อ แม่ ลูก เท่านั้น ทำให้ประชากรที่ถูกสร้างขึ้นมีรูปแบบของการอาศัยอยู่ในครัวเรือนคล้ายๆ กันในทุกครัวเรือน ประชากรสังเคราะห์ที่ถูกสร้างขึ้นจึงไม่ค่อยมีความหลากหลาย



ที่มา : [23]

ภาพที่ 11 กระบวนการที่ใช้ในการสร้างเมืองจำลอง

จากงานวิจัยที่ได้กล่าวมาข้างต้น จะเห็นได้ว่าการสร้างประชากรสังเคราะห์ให้มีความเหมาะสมในแต่ละประเทศนั้นมีกระบวนการสร้างประชากรสังเคราะห์ที่แตกต่างกันออกไปตามข้อมูลที่สามารถรวบรวมได้และข้อมูลที่เปิดเผยได้ในแต่ละประเทศ สำหรับประเทศไทยมีการจัดเก็บและเปิดเผยเพียงข้อมูลทางสถิติและข้อมูลที่ได้จากการทำสำมะโนประชากรหรือเรียกว่า census data ซึ่งถือเป็นข้อมูลแบบ aggregate data เพียงอย่างเดียวเท่านั้น จึงทำให้กระบวนการสร้างประชากรสังเคราะห์ที่เหมาะสมจะอยู่ในรูปแบบของ Synthetic Reconstruction แต่จำเป็นต้องมีการ

ออกแบบกระบวนการสร้างประชากรสังเคราะห์ใหม่ให้มีความเหมาะสม สอดคล้องกับข้อมูลและโครงสร้างครอบครัวของประเทศไทยรวมถึงการเลือกข้อมูลที่มีคุณภาพดีมาใช้ในการสร้างประชากรสังเคราะห์เพื่อให้ประชากรสังเคราะห์ที่ได้นั้นมีความถูกต้องแม่นยำและน่าเชื่อถือและใกล้เคียงกับประชากรจริงมากที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3
แนวคิดและวิธีการดำเนินงานวิจัย

ในบทนี้จะกล่าวถึงแนวคิดของการแก้ปัญหา พร้อมทั้งรายละเอียดของการดำเนินงานวิจัยเพื่อพัฒนากระบวนการสร้างประชากรสังเคราะห์ให้มีความเหมาะสมกับข้อมูลที่มีการเปิดเผยของประเทศไทย เพื่อให้ข้อมูลของประชากรสังเคราะห์ที่สร้างขึ้นได้นั้นมีความถูกต้องแม่นยำและมีค่าใกล้เคียงกับข้อมูลของประชากรจริงมากที่สุด โดยในงานวิจัยนี้มีปัญหาที่จำเป็นต้องพิจารณาใน 2 ด้าน คือ

1. ปัญหาด้านข้อจำกัดของข้อมูลที่มีการเปิดเผย

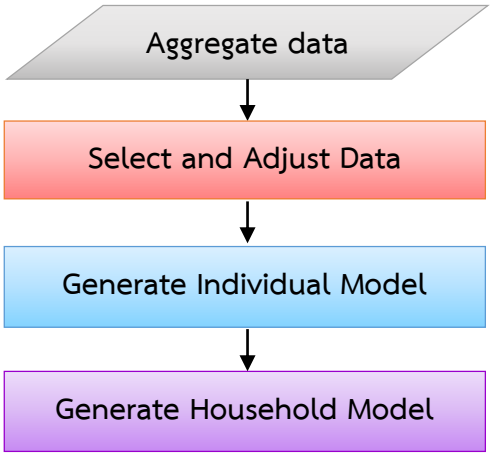
ประเทศไทยมีข้อมูล aggregate data ซึ่งอยู่ในรูปแบบ census data เท่านั้นแต่ยังไม่มีข้อมูล disaggregated data ที่อยู่ในรูปแบบ survey microdata ที่มีการเปิดเผย ทำให้กระบวนการสร้างประชากรสังเคราะห์ที่มีอยู่นั้นไม่เหมาะสมที่จะสร้างประชากรสังเคราะห์ของประเทศไทย

2. ปัญหาความสอดคล้องของข้อมูล

เนื่องจากข้อมูล census data ถูกรวบรวมมาจากหลายหน่วยงาน ซึ่งในแต่ละหน่วยงานมีการเก็บรวบรวมข้อมูลที่มีความละเอียดแตกต่างกันจึงทำให้ข้อมูลประชากรที่ถูกเปิดเผยนั้นมีความไม่สอดคล้องกันส่งผลให้ประชากรสังเคราะห์ที่ได้มีความน่าเชื่อถือน้อย

จากปัญหาเหล่านี้ทำให้การสร้างประชากรสังเคราะห์มีความซับซ้อนและท้าทายเป็นอย่างยิ่ง มีหลายงานวิจัยก่อนหน้านี้ได้พยายามพัฒนางานวิจัยเพื่อแก้ไขปัญหาดังกล่าว [4, 15, 23] แต่ก็ยังพบว่าประชากรสังเคราะห์ที่ได้นั้นยังมีความไม่หลากหลายของข้อมูลทั้งในด้านคุณลักษณะของข้อมูลประชากรและความสัมพันธ์ของประชากรสังเคราะห์ภายในครัวเรือน

จากเหตุผลดังกล่าวข้างต้น งานวิจัยนี้จึงได้ออกแบบกระบวนการสร้างประชากรสังเคราะห์ให้เหมาะสมกับข้อมูลที่มี และได้ออกแบบให้มีการสร้างประชากรสังเคราะห์ให้มีความหลากหลายคล้ายคลึงกับประชากรที่มีอยู่จริง ซึ่งกระบวนการสร้างประชากรสังเคราะห์จะแสดงในภาพที่ 12



ภาพที่ *12* กระบวนการสร้างประชากรสังเคราะห์

3.1 การเลือกและการจัดการข้อมูล

3.1.1 ข้อมูลประชากรของไทย

สำนักงานสถิติแห่งชาติจะมีการจัดทำสำมะโนประชากรและเคหะทุก 10 ปีตามข้อเสนอแนะขององค์การสหประชาชาติ ในปีค.ศ.ที่ลงท้ายด้วย 0 เช่น ค.ศ. 2000 (พ.ศ.2543), ค.ศ. 2010 (พ.ศ.2553) ซึ่งครั้งล่าสุดคือปี 2553 โดยกำหนดให้วันที่ 1 กันยายน พ.ศ.2553 เป็นวันสำมะโนเพื่อเป็นเวลาอ้างอิง โดยได้เลือกทำสำมะโนทดลองที่จังหวัดพิษณุโลกในปี 2551 เนื่องจากเป็นจังหวัดที่มีโครงสร้างของประชากรทุกรูปแบบ สามารถเป็นตัวแทนของทั้งประเทศได้ โดยเป็นจังหวัดขนาดกลางซึ่งมีประชากรอยู่ระหว่าง 5 แสนถึง 1 ล้านคนและเป็นพื้นที่ที่มีความหลากหลายโดยมีทั้งที่เป็นชุมชนปกติ ชุมชนเมืองที่มีความเจริญและมีตึกสูง เช่น คอนโดมิเนียม อพาร์ทเมนท์ มีชุมชนที่มีความเข้มแข็ง มีพื้นที่ที่มีคนต่างชาติอาศัยอยู่มากตามเขตชายแดน ดังนั้นในการทำสำมะโนทดลองครั้งนี้จึงได้เลือก 3 อำเภอคือ อำเภอเมืองพิษณุโลก อำเภอพรหมพิรามและอำเภอนครไทยเพื่อเป็นการทดลองอย่างครบวงจร เริ่มตั้งแต่ร่วมการวิเคราะห์ออกแบบระบบ แบบสอบถาม การออกสนามการใช้ ICT ในการประมวลผลจนถึงขั้นตอนการวิเคราะห์ข้อมูลและนำเสนอผลที่ได้ ซึ่งสำนักงานสถิติแห่งชาติจะได้รวบรวมและประมวลผลการดำเนินงานเพื่อปรับกระบวนการทำงานสำหรับการทำสำมะโนประชากรทั้งหมดของประเทศไทยในปี 2553 ต่อไป [24]

3.1.2 การเลือกข้อมูล

เนื่องจากงานวิจัยนี้ได้ออกแบบกระบวนการสร้างประชากรสังเคราะห์เพื่อให้มีความเหมาะสมกับข้อมูลที่มีการเผยแพร่ของประเทศไทย ดังนั้นข้อมูลที่ถูกนำมาใช้จึงมีความสำคัญมากต่อ

การออกแบบเพื่อให้ประชากรสังเคราะห์ที่สร้างขึ้นนั้นมีความถูกต้องแม่นยำมากที่สุด โดยผู้วิจัยมีความสนใจในการสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือน ข้อมูลประชากรที่จะนำมาใช้จึงจำเป็นต้องมีความละเอียดและน่าเชื่อถือและควรจะมีคุณลักษณะของข้อมูลที่เพียงพอเพื่อให้ประชากรสังเคราะห์ที่สร้างได้มีการระบุตัวตนที่มีความหลากหลาย จากการศึกษาข้อมูลที่มีการเปิดเผยของประเทศไทยพบว่ามีข้อมูลที่น่าเชื่อถือเป็นข้อมูล aggregate data ได้แก่ ข้อมูล census data ซึ่งอยู่ในรูปแบบของตารางข้อมูลสถิติ โดยจะเลือกใช้ข้อมูลของปี 2553 เป็นหลัก เนื่องจากเป็นข้อมูลครั้งล่าสุดที่ได้จากการจัดทำสำมะโนประชากร โดยมีแหล่งที่มาจากสำนักงานสถิติแห่งชาติ [25, 26], สำนักบริหารการทะเบียน กรมการปกครอง กระทรวงมหาดไทย [27] และกระทรวงศึกษาธิการ [28] โดยรายละเอียดของข้อมูลที่ได้มีการเปิดเผยนั้นเป็นข้อมูลที่มีความเกี่ยวข้องกับประชากรในระดับบุคคลและระดับครัวเรือนดังแสดงในตารางที่ 13 และตัวอย่างของตารางรายละเอียดข้อมูล เช่น ตารางจำนวนประชากรและจำนวนหลังคาเรือนของจังหวัดพิษณุโลก จังหวัดพิษณุโลก, ตารางจำนวนประชากรจำแนกตามอายุของจังหวัดพิษณุโลก และตารางจำนวนประชากรอายุตั้งแต่ 13 ปีขึ้นไป จำแนกตามสถานภาพสมรส ดังแสดงในตารางที่ 1, 2 และ 14 ตามลำดับ ซึ่งจะเห็นได้ว่าข้อมูลเป็นแบบตาราง 2 มิติ กล่าวคือ ข้อมูลนั้นจะมีความสัมพันธ์กับข้อมูลอีกหนึ่งกลุ่ม เช่น ข้อมูลอายุมีความสัมพันธ์กับข้อมูลเพศ ส่วนข้อมูลสถานภาพสมรสก็จะมีความสัมพันธ์กับข้อมูลอายุและเพศ เป็นต้น

ตารางที่ 13 รายละเอียดของข้อมูลที่มีการเปิดเผยของประเทศไทย

หมายเหตุ ตารางข้อมูลตั้งแต่ตารางที่ 1-12 ที่รวบรวมไว้ในตารางที่ 13 แสดงไว้ในภาคผนวก

ตารางที่	รายละเอียด
1	รายงานสถิติจำนวนประชากรและบ้าน พ.ศ. 2553
2	รายงานสถิติจำนวนประชากรจำแนกตามอายุ พ.ศ. 2553
3	จำนวนนักเรียนทุกชั้นปี รายจังหวัด ปีการศึกษา 2553
4	จำนวนนักเรียนนักศึกษา รายสถานศึกษา ปีการศึกษา 2553
5	จำนวนประชากรอายุ 13 ปีขึ้นไป จำแนกตามสถานภาพสมรส กลุ่มอายุ เพศ และเขตการปกครอง ภาคเหนือ พ.ศ. 2553
6	ครัวเรือนส่วนบุคคลจำแนกตามกลุ่มอายุของหัวหน้าครัวเรือน เพศ และเขตการปกครอง พ.ศ.2553

ตารางที่ 13 รายละเอียดของข้อมูลที่มีการเปิดเผยของประเทศไทย (ต่อ)

หมายเหตุ ตารางข้อมูลตั้งแต่ตารางที่ 1-12 ที่รวบรวมไว้ในตารางที่ 13 แสดงไว้ในภาคผนวก

ตารางที่	รายละเอียด
7	จำนวนและร้อยละของหญิงสมรสอายุ 15 - 49 ปี จำแนกตามความแตกต่างของอายุคู่สมรส กลุ่มอายุ ระดับการศึกษา สูงสุดที่สำเร็จ เขตการปกครอง และภาค ทั่วราชอาณาจักร พ.ศ. 2552
8	หญิงเคยสมรสอายุ 13 ปีขึ้นไป จำแนกตามจำนวนบุตรที่มีชีวิตอยู่ จำนวนบุตรที่มีชีวิตอยู่โดยเฉลี่ย ภาคเหนือ พ.ศ. 2553
9	จำนวนและร้อยละของหญิงเคยสมรสอายุ 15 - 49 ปีที่มีบุตร จำแนกตามอายุเมื่อคลอดบุตรคนแรก กลุ่มอายุ ระดับการศึกษาสูงสุดที่สำเร็จ เขตการปกครอง และภาค ทั่วราชอาณาจักร พ.ศ. 2552
10	จำนวนและร้อยละของเด็กอายุ 0 - 17 ปี จำแนกตามการอยู่อาศัยกับพ่อ แม่ กลุ่มอายุ เพศ เขตการปกครอง และภาค พ.ศ. 2551
11	ช่วงอายุห่างระหว่างการมีบุตร จำแนกตามประเทศ พ.ศ. 2530
12	ครัวเรือนส่วนบุคคล จำแนกตามขนาดของครัวเรือน ประเภทของที่อยู่อาศัย และเขตการปกครอง พ.ศ. 2553

ตารางที่ 14 ตัวอย่างตารางจำนวนประชากรอายุตั้งแต่ 13-24 ปี จำแนกตามสถานภาพสมรส กลุ่มอายุ และเพศ

เพศ	กลุ่มอายุ (ปี)	สถานภาพสมรส						
		โสด (คน)	เคยสมรส					ไม่ทราบสถานภาพสมรส(คน)
			สมรส (คน)	หม้าย (คน)	หย่า (คน)	แยกกันอยู่ (คน)	ไม่ทราบ (คน)	
หญิง	13 - 14	371,960	3,018	93	38	101	1,712	4,233
หญิง	15 - 19	767,888	69,153	912	750	3,864	8,163	2,877
หญิง	20 - 24	453,377	208,053	2,690	3,167	11,773	11,122	3,002
ชาย	13 - 14	190,843	1,079	44	20	46	689	2,548
ชาย	15 - 19	407,327	19,612	242	173	693	3,316	1,818
ชาย	20 - 24	267,096	76,098	749	912	3,456	5,488	2,143

ซึ่งสามารถแบ่งคุณลักษณะของข้อมูลที่จะใช้ทั้งในระดับบุคคลและระดับครัวเรือนได้ดังนี้

1. คุณลักษณะของข้อมูลในระดับบุคคล

- เพศ (Gender)
- อายุ (Age)
- อำเภอที่อยู่อาศัย (Location)
- เขตการปกครองที่อยู่อาศัย (Area)
- การศึกษา (Education)
- จำนวนนักเรียนแยกรายสถานศึกษา (School Code)
- สถานภาพสมรส (Marital Status)
- จำนวนบุตร (Number of Children)
- อายุเมื่อมีบุตรคนแรก (Age of First Childbearing)

2. คุณลักษณะของข้อมูลในระดับครัวเรือน

- จำนวนครัวเรือน (Number of Household)
- เพศและอายุของหัวหน้าครัวเรือน (Gender and Age of Household)
- ขนาดครัวเรือน (Size of Household)
- อายุห่างระหว่างคู่สมรส (Age Difference of Spouse)
- การอยู่อาศัยของเด็กอายุ 0-17 ปี (Youth Aged 0 - 17 Years according to Living Arrangements)
- อายุห่างระหว่างบุตรแต่ละคน (Birth intervals)

จากการเก็บรวมรวมข้อมูลในแต่ละคุณลักษณะที่จะใช้นั้นพบว่า มีความแตกต่างกันในการรวบรวมข้อมูลของแต่ละหน่วยงานทำให้ข้อมูลที่ได้มีความไม่สอดคล้องกัน เช่น หน่วยงานแรกเก็บข้อมูลจำนวนคนในระดับอำเภอ แต่หน่วยงานที่สองเก็บข้อมูลจำนวนคนในระดับจังหวัด ทำให้จำนวนคนที่ถูกเปิดเผยของทั้งสองหน่วยงานมีค่าไม่เท่ากัน และแต่ละคุณลักษณะของข้อมูลมีการเปิดเผยข้อมูลอย่างจำกัด เช่น ในระดับอำเภอมีการเปิดเผยข้อมูลที่ครบถ้วน แต่ในระดับตำบลนั้น มีการ

เปิดเผยข้อมูลที่ไม่ครบถ้วน ทำให้ไม่สามารถนำข้อมูลที่มีความละเอียดมากสุดมาใช้ได้ ดังนั้นผู้วิจัยจึงศึกษาความละเอียดของข้อมูลในแต่ละคุณลักษณะที่จะนำมาใช้ เพื่อให้ข้อมูลที่จะนำไปใช้สร้างประชากรสังเคราะห์เป็นข้อมูลที่มีคุณภาพดี ซึ่งจะส่งผลให้ประชากรสังเคราะห์ที่สร้างขึ้นนั้นมีความน่าเชื่อถือ และเหมาะสมแก่การนำไปใช้วิเคราะห์เพื่อหาสิ่งที่สนใจได้อย่างมีความคลาดเคลื่อนน้อยที่สุด ซึ่งระดับความละเอียดของข้อมูลจะแบ่งเป็น ความละเอียดระดับประเทศ, ระดับภาค, ระดับจังหวัด, ระดับอำเภอ และระดับตำบล ซึ่งมีความละเอียดน้อยสุดไปจนถึงความละเอียดมากสุดตามลำดับ แต่เนื่องจากข้อมูลที่มีความละเอียดในระดับตำบลนั้นมีการรวบรวมข้อมูลเพียงบางพื้นที่เท่านั้นจึงไม่สามารถนำมาใช้ในกระบวนการสังเคราะห์ประชากรเพื่อให้ประชากรสังเคราะห์ที่ได้มีความน่าเชื่อถือเพียงพอ จึงจำเป็นต้องใช้ข้อมูลที่มีความละเอียดมากที่สุดที่มีการเก็บรวบรวมข้อมูลไว้ครบถ้วนซึ่งคือ ความละเอียดระดับอำเภอ โดยคุณลักษณะของข้อมูลที่จะใช้ในงานวิจัยนี้มีระดับความละเอียดและลักษณะของข้อมูลที่จะใช้ในการสังเคราะห์ประชากรดังแสดงในตารางที่ 15

ตารางที่ 15 ความละเอียดและลักษณะของข้อมูลจำแนกตามคุณลักษณะ

คุณลักษณะของข้อมูล	ระดับความละเอียด	ลักษณะของข้อมูล
เพศ	อำเภอ	จำนวนคน
อายุ	อำเภอ	จำนวนคน
อำเภอที่อยู่อาศัย	อำเภอ	จำนวนคน
เขตการปกครองที่อยู่อาศัย	อำเภอ	จำนวนคน
การศึกษา	จังหวัด	จำนวนคน
จำนวนนักเรียนแยกรายสถานศึกษา	อำเภอ	จำนวนคน
สถานภาพสมรส	ภาคเหนือ	จำนวนคน
จำนวนบุตร	ภาคเหนือ	จำนวนคน
อายุเมื่อมีบุตรคนแรก	ภาคเหนือ	ร้อยละ
จำนวนครัวเรือน	อำเภอ	จำนวนครัวเรือน
เพศและอายุของหัวหน้าครัวเรือน	จังหวัด	จำนวนครัวเรือน
ขนาดครัวเรือน	จังหวัด	ร้อยละ
อายุห่างระหว่างคู่สมรส	ประเทศ	ร้อยละ
การอยู่อาศัยของเด็กอายุ 0-17 ปี	ภาคเหนือ	ร้อยละ
อายุห่างระหว่างบุตรแต่ละคน	ประเทศ	ร้อยละ

3.1.3 การจัดการข้อมูลให้เหมาะสม

เนื่องจากข้อมูลที่จะนำมาใช้ในการสร้างประชากรสังเคราะห์ระดับครัวเรือนยังมีความไม่สอดคล้องกันจึงไม่เหมาะที่จะนำข้อมูลนั้นมาใช้โดยที่ยังไม่มีการแก้ไข ซึ่งสิ่งที่เห็นได้ชัดเจนคือ ข้อมูลจำนวนครัวเรือนและข้อมูลขนาดของครัวเรือน โดยพบว่าจากความสัมพันธ์ของข้อมูลทั้งสองนี้เมื่อนำข้อมูลขนาดของครัวเรือนมาคูณกับจำนวนครัวเรือนในแต่ละขนาดแล้ว จะได้จำนวนประชากรที่คำนวณได้ ซึ่งควรจะเท่ากับจำนวนประชากรของข้อมูลที่มีการเปิดเผย แต่ในความเป็นจริงพบว่าจำนวนประชากรที่คำนวณได้มีค่าไม่เท่ากับจำนวนประชากรของข้อมูลที่มีการเปิดเผย และยังพบว่าตารางข้อมูลครัวเรือนส่วนบุคคลจำแนกตามขนาดของครัวเรือนและเขตการปกครองนั้นอยู่ในรูปข้อมูลรวมของทุกอำเภอโดยแบ่งเป็นในเขตเทศบาลกับนอกเขตเทศบาลเท่านั้น หากนำข้อมูลนี้ไปใช้โดยไม่มีการแก้ไขจะทำให้ประชากรสังเคราะห์ที่ได้มีความน่าเชื่อถือน้อยเนื่องจากจะทำให้แต่ละอำเภอมีข้อมูลที่แบ่งตามขนาดของครัวเรือนเป็นลักษณะเดียวกันทั้งหมดซึ่งไม่ตรงตามลักษณะการอยู่อาศัยของประชากรจริงดังแสดงตัวอย่างในตารางที่ 16, 17 ซึ่งแสดงให้เห็นว่าหากใช้ข้อมูลสถิติที่ได้เปิดเผยโดยที่ยังไม่มีการแก้ไขนั้นจะพบว่าข้อมูลที่มีการเปิดเผยของจำนวนประชากรนอกเขตเทศบาลและในเขตเทศบาลของอำเภอเมืองพิษณุโลกมีจำนวน 192,775 คน และ 82,275 คน ตามลำดับ ดังแสดงในตารางที่ 16 แต่จำนวนประชากรที่คำนวณได้จากจำนวนครัวเรือนในแต่ละขนาดครัวเรือนคูณกับขนาดครัวเรือนของเขตการปกครองนอกเขตเทศบาลและในเขตเทศบาล คือ 233,965 คน และ 101,218 คน ตามลำดับ ดังแสดงในตารางที่ 17 ซึ่งจะพบว่ามีความไม่สอดคล้องกัน

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 16 จำนวนคนและจำนวนครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง

อ.เมืองพิษณุโลก	**นอกเขตเทศบาล**	**ในเขตเทศบาล**
จำนวนคน	192775	82275
จำนวนครัวเรือน	78989	35741

ตารางที่ 17 การคำนวณจำนวนคนและจำนวนครัวเรือนในแต่ละขนาดครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง

ขนาดครัวเรือน	ร้อยละ		จำนวนครัวเรือนแต่ละขนาดครัวเรือน = จำนวนครัวเรือน x ร้อยละของแต่ละขนาดครัวเรือน		จำนวนคน = จำนวนครัวเรือนที่คำนวณได้ x ขนาดครัวเรือน	
	นอกเขตเทศบาล	ในเขตเทศบาล	นอกเขตเทศบาล	ในเขตเทศบาล	นอกเขตเทศบาล	ในเขตเทศบาล
1	20	23.1	78989 x 20/100 = 15797.8	35741 x 23.1/100 = 8256.171	15797.8 x 1 = 15797.8	8256.171 x 1 = 8256.171
2	22.9	25.1	78989 x 22.9/100 = 18088.48	35742 x 25.1/100 = 8970.991	18088.481 x 2 = 36176.962	8970.991 x 2 = 17941.982
3	23.9	21.7	78989 x 23.9/100 = 18878.37	35743 x 21.7/100 = 7755.797	18878.371 x 3 = 56635.113	7755.797 x 3 = 23267.391
4	17.9	16.4	78989 x 17.9/100 = 14139.03	35744 x 16.4/100 = 5861.524	14139.031 x 4 = 56556.124	5861.524 x 4 = 23446.096
5	8.6	7.4	78989 x8.6/100 = 6793.054	35745 x 7.4/100 = 2644.834	6793.054 x 5 = 33965.27	2644.834 x 5 = 13224.17
6	4.4	3.9	78989 x 4.4/100 = 3475.516	35746 x 3.9/100 = 1393.899	3475.516 x 6 = 20853.096	1393.899 x 6 = 8363.394
7	1.3	1.3	78989 x 1.3/100 = 1026.857	35747 x 1.3/100 = 464.633	1026.857 x 7 = 7187.999	464.633 x 7 = 3252.431
8	0.6	0.5	78989 x 0.6/100 = 473.934	35748 x 0.5/100 = 178.705	473.934 x 8 = 3791.472	178.705 x 8 = 1429.64
9	0.2	0.3	78989 x 0.2/100 = 157.978	35749 x 0.3/100 = 107.223	157.978 x 9 = 1421.802	107.223 x 9 = 965.007
10	0.2	0.3	78989 x 0.2/100 = 157.978	35750 x 0.3/100 = 107.223	157.978 x 10 = 1579.78	107.223 x 10 = 1072.23
รวม	100	100	78989	35741	233965.418	101218.512

จึงจำเป็นต้องปรับปรุงข้อมูลร้อยละของขนาดครัวเรือนใหม่ในทุกๆ อำเภอขึ้นอยู่กับจำนวนประชากรและจำนวนครัวเรือนในแต่ละอำเภอเพื่อให้มีความเหมาะสม โดยใช้การแจกแจงของปัวซองซึ่งได้อธิบายไว้ในหัวข้อ 2.1.3.2 กำหนดให้ข้อมูลจำนวนคน = IND, จำนวนครัวเรือน = HH ซึ่งมีขั้นตอนดังนี้

1. หาขนาดของครัวเรือนเฉลี่ย (Average Household Size)

$$Avg.HH\ Size = IND/HH$$

2. หาร้อยละของขนาดครัวเรือน โดยใช้การแจกแจงแบบปัวซอง ตามสูตร

$$f(n;M) = \frac{M^n e^{-M}}{n!}$$

โดยที่ M คือ ค่าเฉลี่ยของประชากรที่ปรับปรุง (Modified mean) ซึ่งมีค่าเท่ากับค่าขนาดของครัวเรือนเฉลี่ย - 1 (Avg.HH Size - 1) และ n คือ ขนาดของครัวเรือน – 1 (Household Size - 1)

สำหรับตัวอย่างการปรับปรุงข้อมูลของข้อมูลของขนาดครัวเรือนโดยใช้การแจกแจงของปัวซองดังแสดงในตารางที่ 18, 19 จะแสดงการปรับปรุงข้อมูลดังกล่าวของอำเภอเมืองพิษณุโลกนอกเขตเทศบาลโดยคำนวณ Avg.HH Size และ M ดังแสดงในตารางที่ 18 ได้นำการแจกแจงของปัวซองดังที่อธิบายไว้ข้างต้นมาคำนวณร้อยละในแต่ละขนาดครัวเรือนใหม่โดยแสดงร้อยละที่ปรับปรุงใหม่ในตารางที่ 19 เมื่อคำนวณจำนวนประชากรนอกเขตเทศบาลอำเภอเมืองพิษณุโลกมีจำนวน 192,772 คนและจำนวนประชากรในเขตเทศบาลอำเภอเมืองพิษณุโลกมีจำนวน 82,274 คน ดังนั้นจึงมีความจำเป็นที่จะต้องใช้การแจกแจงของปัวซองเพื่อปรับปรุงข้อมูลที่อยู่ในรูปร้อยละของขนาดครัวเรือนในทุกๆ อำเภอและเขตการปกครองเพื่อให้ประชากรสังเคราะห์ที่สร้างขึ้นมีความใกล้เคียงกับข้อมูลจริงมากที่สุด

ตารางที่ 18 การปรับปรุงข้อมูลจำนวนคนและจำนวนครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง

อ.เมืองพิษณุโลก	**นอกเขตเทศบาล**	**ในเขตเทศบาล**
จำนวนคน	192775	82275
จำนวนครัวเรือน	78989	35741
Avg HH Size = จำนวนคน/ จำนวนครัวเรือน	2.440529694	2.30197812
Modified Mean = Avg HH Size - 1	1.4405	1.30197812

ตารางที่ 19 การปรับปรุงข้อมูลของการคำนวณจำนวนคนและจำนวนครัวเรือนในแต่ละขนาดครัวเรือนของอำเภอเมืองพิษณุโลกจำแนกตามเขตการปกครอง

ขนาดครัวเรือน (HH. size)	x = HH. Size - 1	ร้อยละที่ปรับปรุงใหม่		จำนวนครัวเรือนแต่ละขนาดครัวเรือน = จำนวนครัวเรือน x ร้อยละของแต่ละขนาดครัวเรือน		จำนวนคน = จำนวนครัวเรือนที่คำนวณได้ x ขนาดครัวเรือน	
		นอกเขตเทศบาล	ในเขตเทศบาล	นอกเขตเทศบาล	ในเขตเทศบาล	นอกเขตเทศบาล	ในเขตเทศบาล
1	0	23.680	27.199	78989 x 23.680/100 = 18704.78	35741 x 27.199/100 = 9721.31	18704.78 x 1 = 18704.78	9721.31 x 1 = 9721.31
2	1	34.112	35.413	78989 x 34.112/100 = 26944.79	35742 x 35.413/100 = 12656.93	26944.79 x 2 = 53889.57	12656.93 x 2 = 25313.87
3	2	24.570	23.053	78989 x 24.570/100 = 19407.38	35743 x 23.053/100 = 8239.52	19407.38 x 3 = 58222.15	8239.52 x 3 = 24718.57
4	3	11.798	10.005	78989 x11.798/100 = 9318.97	35744 x 10.005/100 = 3575.89	9318.97 x 4 = 37275.88	3575.89 x 4 = 14303.57
5	4	4.249	3.257	78989 x4.249/100 = 3356.06	35745 x 3.257/100 = 1163.93	3356.06 x 5 = 16780.32	1163.93 x 5 = 5819.67
6	5	1.224	0.848	78989 x 1.224/100 = 966.90	35746 x 0.848/100 = 303.08	966.90 x 6 = 5801.41	303.08 x 6 = 1818.50
7	6	0.294	0.184	78989 x 0.294/100 = 232.14	35747 x 0.184/100 = 65.77	232.14 x 7 = 1624.99	65.77 x 7 = 460.38
8	7	0.060	0.034	78989 x 0.060/100 = 47.77	35748 x 0.034/100 = 12.23	47.77 x 8 = 382.18	12.23 x 8 = 97.86
9	8	0.011	0.006	78989 x 0.011/100 = 8.60	35749 x 0.006/100 = 1.99	8.60 x 9 = 77.42	1.99 x 9 = 17.92
10	9	0.002	0.001	78989 x 0.002/100 = 1.38	35750 x 0.001/100 = 0.29	1.38 x 10 = 13.77	0.29 x 10 = 2.88
รวม		100	100	78988.77	35740.96	192772.46	82274.53

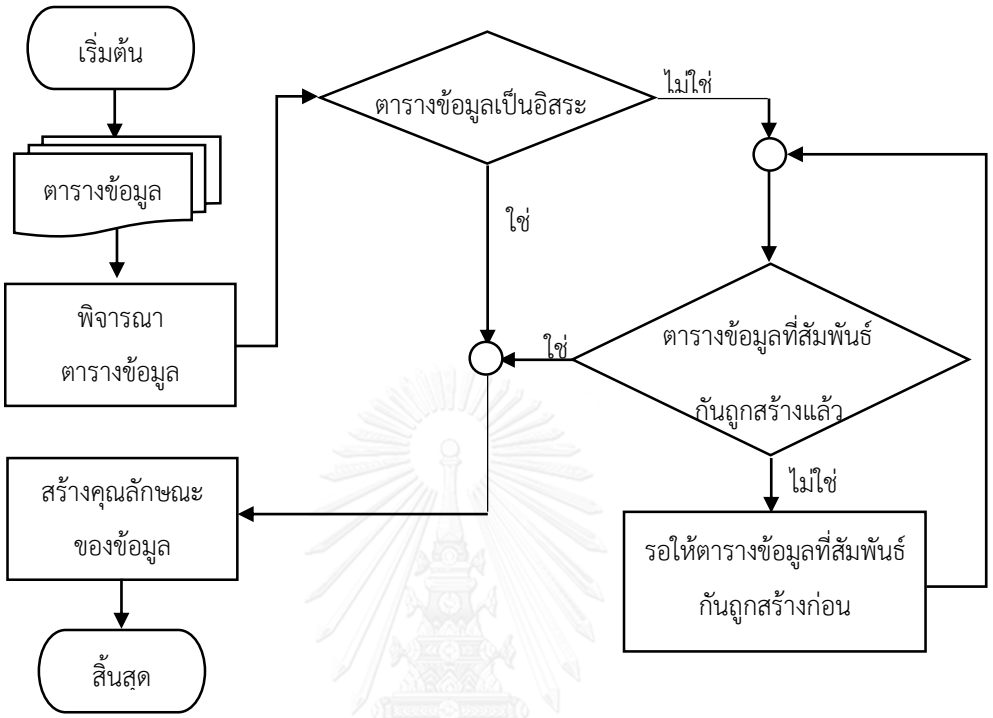
3.2. การออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับบุคคล

เนื่องจากข้อมูลต้นแบบที่ได้ศึกษามานั้นส่วนใหญ่เป็นการเก็บข้อมูลที่อยู่ในรูปแบบของตาราง 2 มิติคือแถว (Row) ของตารางจะแทนเรคคอร์ด (Record) และคอลัมน์ (Column) จะเป็นขอบเขตของข้อมูล (Field) หรือเรียกว่าแอททริบิวต์ (Attribute) เช่น ตารางจำนวนประชากรและจำนวนหลังคาเรือนของจังหวัดพิษณุโลก, ตารางจำนวนประชากรจำแนกตามอายุของจังหวัดพิษณุโลก และตารางจำนวนประชากรอายุตั้งแต่ 13-29 ปี จำแนกตามสถานภาพสมรส กลุ่มอายุ และเพศ ดังแสดงในตารางที่ 1, 2 และ 14 ตามลำดับ ซึ่งคุณลักษณะของข้อมูลในบางตารางมีความสัมพันธ์กับคุณลักษณะของตารางอื่น แต่บางคุณลักษณะก็ไม่ได้มีความสัมพันธ์กับคุณลักษณะใดเลย ทำให้การออกแบบกระบวนการสร้างประชากรสังเคราะห์ในระดับบุคคลจะต้องศึกษาหาความสัมพันธ์ของข้อมูลในแต่ละคุณลักษณะของข้อมูล เพื่อกำหนดลำดับการสร้างของแต่ละคุณลักษณะดังที่กล่าว เพื่อใส่ค่าคุณลักษณะของบุคคลทีละคุณลักษณะ ส่งผลให้ประชากรระดับบุคคลที่ถูกสร้างนั้นมีค่าของคุณลักษณะที่ถูกต้องและครบถ้วน

สำหรับขั้นตอนการทำงานของการสร้างประชากรสังเคราะห์ระดับบุคคลจะแสดงในภาพที่ 13 ซึ่งเป็นการใส่ค่าทีละคุณลักษณะจนครบถ้วน เริ่มต้นจากการนำตารางข้อมูลทั้งหมดที่อยู่ในรูปแบบตารางมาพิจารณา เพื่อแยกประเภทของตารางข้อมูลว่าตารางข้อมูลใดเป็นตารางข้อมูลที่เป็นอิสระไม่ขึ้นอยู่กับตารางข้อมูลอื่น (1^{st} level) และตารางข้อมูลใดบ้างเป็นตารางข้อมูลที่มีความสัมพันธ์กับข้อมูลอื่น (2^{nd} level) โดยมีเงื่อนไขในการนำข้อมูลในตารางไปสร้างคุณลักษณะของข้อมูลประชากรสังเคราะห์รายบุคคล ดังนี้

- ตารางข้อมูลเป็นตารางข้อมูลที่เป็นอิสระ(1^{st} level) สามารถนำข้อมูลในตารางนี้ไปสร้างคุณลักษณะของข้อมูลประชากรสังเคราะห์รายบุคคลได้เลย
- ตารางข้อมูลไม่ใช่ตารางข้อมูลที่เป็นอิสระ(2^{nd} level) จะต้องดูว่าตารางข้อมูลที่มีความสัมพันธ์กับตารางนี้ถูกนำไปสร้างคุณลักษณะของข้อมูลประชากรสังเคราะห์รายบุคคลแล้วหรือยัง
    - ถ้าตารางข้อมูลที่มีความสัมพันธ์กับตารางนี้ถูกนำไปสร้างคุณลักษณะของข้อมูลประชากรรายบุคคลแล้ว ก็สามารถนำข้อมูลของตารางนี้ไปสร้างคุณลักษณะของข้อมูลประชากรสังเคราะห์รายบุคคลต่อได้
    - ถ้าตารางข้อมูลที่มีความสัมพันธ์กับตารางนี้ยังไม่ถูกนำไปสร้างคุณลักษณะของข้อมูลประชากรรายบุคคล ต้องรอจนกว่าตารางข้อมูลที่มีความสัมพันธ์กับตารางนี้ถูก

นำไปสร้างคุณลักษณะของข้อมูลประชากรรายบุคคลก่อน จึงจะนำข้อมูลในตารางนี้ไปสร้างคุณลักษณะของข้อมูลประชากรสังเคราะห์รายบุคคลได้



ภาพที่ 13 ขั้นตอนการทำงานของการสร้างประชากรสังเคราะห์ระดับบุคคล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

โดยความสัมพันธ์ของคุณลักษณะข้อมูลสามารถยกตัวอย่างได้คือ ตารางอายุถูกสร้างจากคุณลักษณะเพศดังนั้นอายุมีความสัมพันธ์กับเพศ ตารางการศึกษาและตารางสถานภาพสมรสมีความสัมพันธ์กับอายุ ซึ่งมีขั้นตอนในการพิจารณาลำดับการสร้าง ดังนี้

1. วิเคราะห์ข้อมูลที่อยู่ในรูปแบบตาราง 2 มิติทั้งหมดเพื่อหาความสัมพันธ์ที่เชื่อมโยงกันระหว่างแถวและคอลัมน์ของตารางข้อมูลทั้งหมด โดยการใช้ข้อมูลแถวหรือข้อมูลคอลัมน์ที่เหมือนกันทั้งสองตารางเป็นตัวเชื่อมโยงความสัมพันธ์ของข้อมูล ตัวอย่างดังแสดงในภาพที่ 14 ซึ่งมีข้อมูล 3 ตาราง คือ ตารางข้อมูลจำนวนประชากรจำแนกตามเพศ, ตารางข้อมูลจำนวนประชากรจำแนกตามอำเภอและตารางข้อมูลจำนวนประชากรจำแนกตามเขตการปกครอง โดยมีการเชื่อมโยงความสัมพันธ์ของทั้งสามตารางที่ใช้ข้อมูลเพศ (gender) และข้อมูลอำเภอ (location) เป็นตัวเชื่อมโยงความสัมพันธ์ของทั้ง 3 ตาราง

ตารางจำนวนประชากรจำแนกตามเพศ

Gender	male	female
ร้อยละ	49.039	50.961

ตารางจำนวนประชากรจำแนกตามอำเภอ

Gender	อ.เมืองพิษณุโลก	อ.นครไทย	อ.ชาติตระการ	อ.บางระกำ
male	15.717	5.027	2.403	5.402
female	17.201	4.971	2.338	5.568

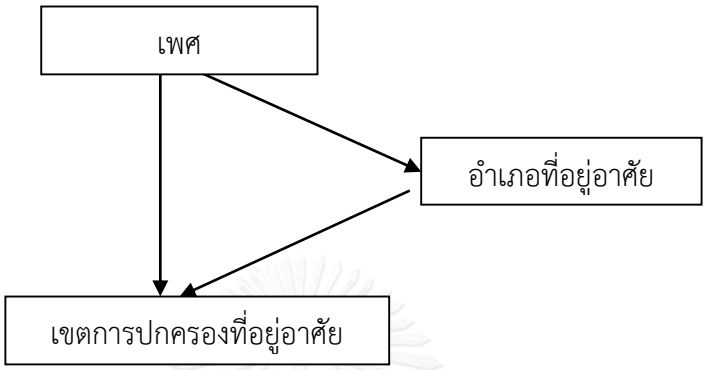
ตารางจำนวนประชากรจำแนกตามเขตการปกครอง

Gender	Location	นอกเขตเทศบาล	ในเขตเทศบาล
male	อ.เมืองพิษณุโลก	11.280	4.437
male	อ.นครไทย	3.900	1.127
male	อ.ชาติตระการ	2.076	0.327
male	อ.บางระกำ	4.313	1.089
female	อ.เมืองพิษณุโลก	11.791	5.410
female	อ.นครไทย	3.857	1.115
female	อ.ชาติตระการ	2.005	0.333
female	อ.บางระกำ	4.414	1.154

ภาพที่ *14* ตัวอย่างความสัมพันธ์ของข้อมูลในรูปแบบตาราง

2. เมื่อได้ความสัมพันธ์ที่เชื่อมโยงกันในแต่ละตารางจากที่แสดงในภาพที่ 14 จะได้ว่าค่าของตารางข้อมูลจำนวนประชากรที่จำแนกตามอำเภอจะมีความสัมพันธ์กับเพศ (gender) และตารางข้อมูลจำนวนประชากรจำแนกตามเขตการปกครองจะมีความสัมพันธ์ร่วมระหว่างค่าของข้อมูลอำเภอ (location) และค่าของข้อมูลเพศ (gender) ซึ่งจากข้อมูลที่มีความสัมพันธ์ดังกล่าวทำให้ต้องมีการสร้างข้อมูลอย่างเป็นลำดับขั้นตอนเพื่อให้การสร้างข้อมูลนั้นมีความถูกต้องและสอดคล้องกับข้อมูลในรูปแบบตารางที่มี โดยที่ลำดับในการ

สร้างข้อมูลนั้นจะเกี่ยวข้องกับคุณลักษณะของข้อมูลที่เป็นอิสระ ไม่ขึ้นอยู่กับคุณลักษณะของข้อมูลอื่น จะเป็นคุณลักษณะที่สามารถใช้สร้างข้อมูลได้เป็นอันดับแรกและคุณลักษณะของข้อมูลที่ขึ้นอยู่กับคุณลักษณะของข้อมูลอื่น จะเป็นคุณลักษณะที่ใช้สร้างข้อมูลหลังจากที่ข้อมูลอื่นได้ถูกสร้างขึ้นแล้ว จากตัวอย่างจะแสดงในภาพที่ 15



ภาพที่ 15 ความสัมพันธ์ของคุณลักษณะของข้อมูล

3. จากความสัมพันธ์ของแต่ละคุณลักษณะของข้อมูลที่ได้ในขั้นตอนที่ 2 ในลำดับต่อไปจะกำหนดลำดับขั้นตอนของการสร้างแบบจำลองของข้อมูลนี้ โดยสร้างคุณลักษณะของข้อมูลที่เป็นอิสระก่อนเป็นลำดับแรก หลังจากนั้นจึงสร้างคุณลักษณะของข้อมูลที่ขึ้นอยู่กับคุณลักษณะของข้อมูลก่อนหน้านั้นเป็นลำดับต่อไป ซึ่งจากความสัมพันธ์ของแต่ละคุณลักษณะของข้อมูลดังแสดงในภาพที่ 15 ทำให้มีการสร้างประชากรสังเคราะห์โดยมีลำดับการสร้างของแต่ละคุณลักษณะหรือแอททริบิวต์ตามตารางที่ 20 ซึ่งข้อมูลเพศจะสามารถใส่ให้กับประชากรสังเคราะห์ได้ในอันดับแรกเนื่องจากข้อมูลเพศเป็นข้อมูลที่อิสระไม่ได้ขึ้นอยู่กับข้อมูลอื่นๆ สำหรับข้อมูลอำเภอที่อยู่อาศัยจะสามารถใส่ให้กับประชากรสังเคราะห์ได้ก็ต่อเมื่อประชากรสังเคราะห์นั้นมีข้อมูลเพศอยู่แล้วเช่นเดียวกับข้อมูลเขตการปกครองที่อยู่อาศัยที่จะสามารถใส่ค่าได้ก็ต่อเมื่อประชากรสังเคราะห์นั้นมีข้อมูลเพศและอำเภอที่อยู่อาศัยมาแล้ว ซึ่งข้อมูลอำเภอที่อยู่อาศัยและเขตการปกครองที่อยู่อาศัยนี้จะเรียกว่าเป็นข้อมูลที่ไม่อิสระหมายถึงเป็นข้อมูลที่ขึ้นอยู่กับข้อมูลเพศและอำเภอที่อยู่อาศัย แต่หากตารางข้อมูลมีความสัมพันธ์กันหลายแบบ จะเลือกข้อมูลที่ถูกอ้างอิงมากที่สุดและมีความละเอียดสูงที่สุดให้ถูกสร้างเป็นลำดับแรก

ดังนั้นจากความสัมพันธ์ของข้อมูลตารางในแต่ละคุณลักษณะ ทำให้ได้ลำดับการสร้างของกระบวนการสร้างประชากรระดับบุคคล ดังแสดงในตารางที่ 21

ตารางที่ 20 ลำดับการสร้างประชากรสังเคราะห์

คุณลักษณะ	ลำดับ
เพศ	1
อำเภอที่อยู่อาศัย	2
เขตการปกครองที่อยู่อาศัย	3

ตารางที่ 21 ลำดับของกระบวนการสร้างประชากรสังเคราะห์ระดับบุคคล

คุณลักษณะของข้อมูล	ลำดับการสร้าง
เพศ	1
อำเภอที่อยู่อาศัย	2
เขตการปกครองที่อยู่อาศัย	3
อายุ	4
การศึกษา	5
จำนวนนักเรียนแยกรายสถานศึกษา	6
สถานภาพสมรส	7
จำนวนบุตร	8
อายุเมื่อมีบุตรคนแรก	9

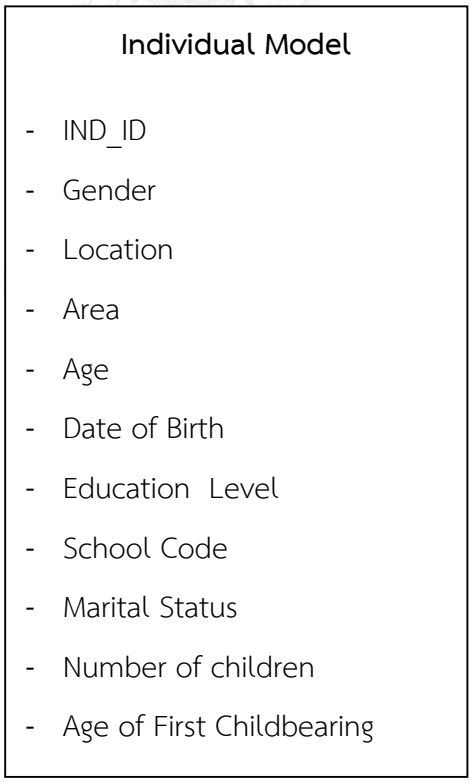
จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สำหรับกระบวนการสร้างประชากรสังเคราะห์ระดับบุคคล จะใช้ความสัมพันธ์ของเงื่อนไขความน่าจะเป็นดังที่อธิบายในหัวข้อ 2.1.4 โดย $P(h \mid j)$ คือความน่าจะเป็นของประชากรที่ทราบว่ามีคุณลักษณะ j ซึ่งจะมีคุณลักษณะ h, $P(h \cap j)$ คือ ความน่าจะเป็นที่ประชากรจะมีคุณสมบัติ h และ j เทียบกับประชากรทั้งหมด และ $P(j)$ คือความน่าจะเป็นของประชากรที่มีคุณสมบัติ j ดังนั้นความน่าจะเป็นของประชากรที่ทราบว่ามีคุณลักษณะ j จะมีคุณลักษณะ h ซึ่งสามารถคำนวณได้โดยใช้สมการ

$$P(h \mid j) = \frac{P(h \cap j)}{P(j)}$$

ตัวอย่างเช่น หากต้องการทราบความน่าจะเป็นของประชากรที่อายุ 20 ปีและแต่งงานแล้ว $P(h \cap j)$ จะได้จากการนำความน่าจะเป็นของประชากรที่อายุ 20 $P(j)$ มาคูณกันกับความน่าจะเป็นของประชากรที่มีสถานภาพสมรสในช่วงอายุ 20 ปี $P(h \mid j)$ เป็นต้น โดยคุณลักษณะข้อมูลประชากรรายบุคคลดังแสดงในภาพที่ 16 สามารถอธิบายรายละเอียดได้ดังนี้

IND_ID	:	เลขประจำตัวบุคคล
Gender	:	เพศ
Location	:	อำเภอที่อยู่อาศัย
Area	:	เขตการปกครองที่อยู่อาศัย
Age	:	อายุ
Date of Birth	:	วันเดือนปีที่เกิด
Education Level	:	ระดับการศึกษาของนักเรียน
School Code	:	รหัสโรงเรียนของนักเรียน
Marital Status	:	สถานภาพสมรส
Number of children	:	จำนวนบุตร
Age of First Childbearing	:	อายุเมื่อมีบุตรคนแรก



ภาพที่ 16 คุณลักษณะข้อมูลประชากรรายบุคคล

3.3 การออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือน

การออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือนเป็นการสร้างความสัมพันธ์ระหว่างบุคคลเพื่อเลือกประชากรที่มีความสัมพันธ์กันเข้าอยู่ในครัวเรือนเดียวกัน แต่เนื่องจากข้อมูลที่จะใช้ในการสร้างประชากรสังเคราะห์ในระดับครัวเรือนมีอยู่อย่างจำกัดและมีความละเอียดต่ำทำให้ต้องวางแผนการออกแบบอย่างละเอียดรอบคอบ รวมไปถึงต้องใช้การวิเคราะห์ความเป็นอยู่ของสังคมไทย เพื่อให้ได้กระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือนที่มีความถูกต้องแม่นยำน่าเชื่อถือและมีความใกล้เคียงกับสังคมของประเทศไทยมากที่สุด สำหรับขั้นตอนการทำงานของการสร้างประชากรสังเคราะห์ระดับครัวเรือนจะแสดงในภาพที่ 17 โดยมีลำดับขั้นตอนดังนี้

1. เริ่มต้นจากการสร้างกลุ่ม ซึ่งจะพิจารณาว่าสมาชิกนี้เป็นเจ้าของกลุ่มหรือไม่

    1.1 ถ้าเป็นเจ้าของกลุ่ม จะสร้าง ID ของกลุ่มให้ และหยิบเข้ากลุ่ม

    1.2 ถ้าไม่ใช่เจ้าของกลุ่ม จะพิจารณาต่อว่า เป็นสมาชิกที่มีความสัมพันธ์กับเจ้าของกลุ่มหรือไม่

        1.2.1 ถ้าเป็นสมาชิกที่มีความสัมพันธ์กับเจ้าของกลุ่ม จะหยิบเข้ากลุ่ม

        1.2.2 ถ้าไม่ใช่สมาชิกที่มีความสัมพันธ์กับเจ้าของกลุ่ม จะเลือกสมาชิกใหม่มาพิจารณาขั้นตอนที่ 1.2 ใหม่

2. นับจำนวนสมาชิกในกลุ่มที่มีอยู่

3. กำหนดขนาดของกลุ่ม โดยจะพิจารณาขนาดของกลุ่มว่ามีจำนวนมากกว่าจำนวนสมาชิกในกลุ่มที่มีอยู่หรือไม่

    3.1 ถ้าขนาดของกลุ่มมีจำนวนมากกว่าสมาชิกในกลุ่มที่มีอยู่ จะใส่ขนาดของกลุ่มนั้นให้

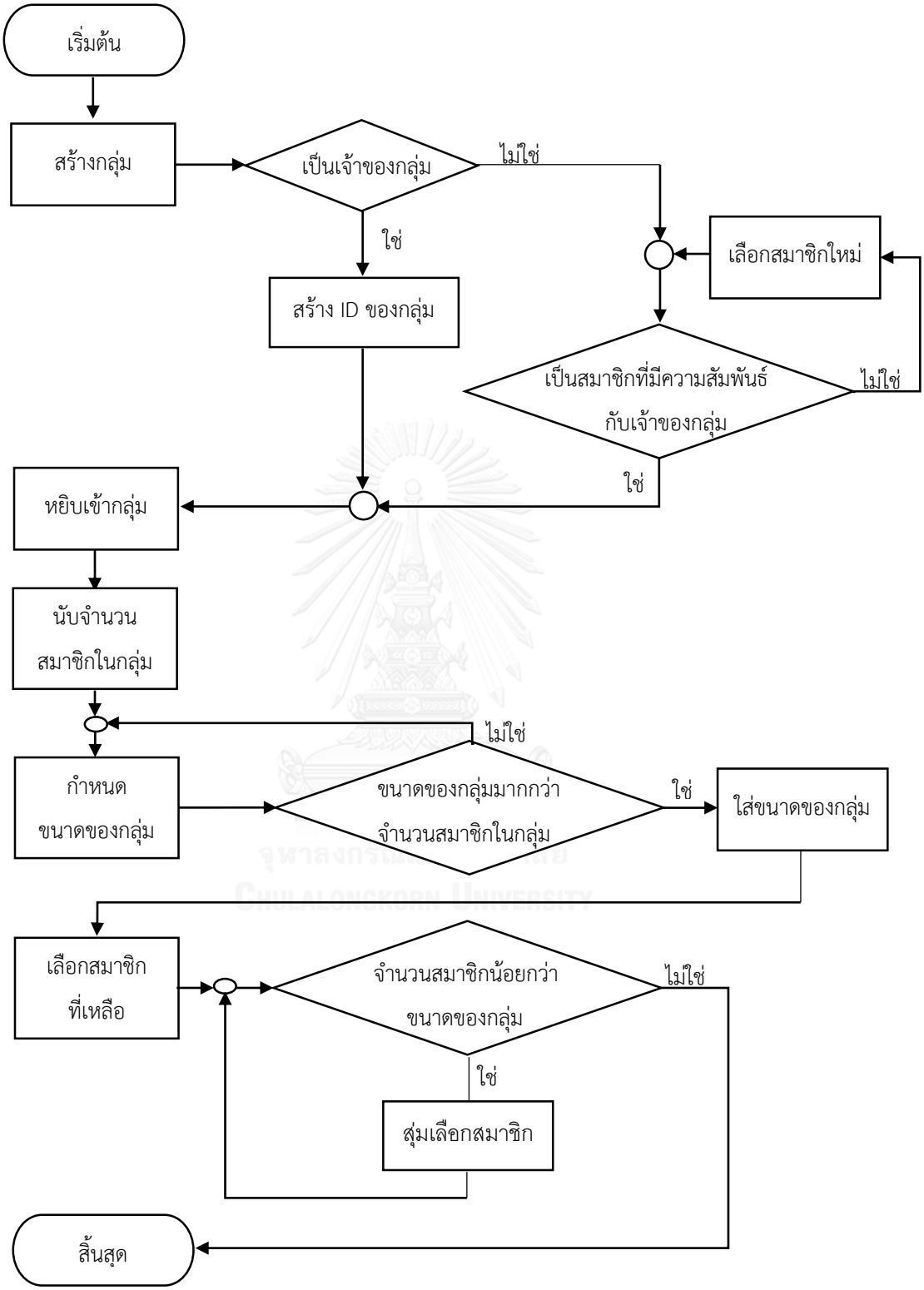
    3.2 ถ้าขนาดของกลุ่มไม่ได้มีจำนวนมากกว่าสมาชิกในกลุ่มที่มีอยู่ จะหาขนาดของกลุ่มใหม่ แล้วทำตามขั้นตอน 3 ใหม่

4. หลังจากใส่ขนาดของกลุ่มแล้ว จะเลือกสมาชิกที่เหลือเข้ากลุ่มให้ครบตามขนาดของกลุ่ม

    4.1 ถ้าจำนวนสมาชิกยังน้อยกว่าขนาดของกลุ่ม จะสุ่มเลือกสมาชิกให้เข้ามาอยู่ในกลุ่มจนครบตามขนาดของกลุ่ม

    4.2 ถ้าจำนวนสมาชิกไม่น้อยกว่าขนาดของกลุ่ม แสดงว่าจำนวนสมาชิกครบตามขนาดของกลุ่มแล้ว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

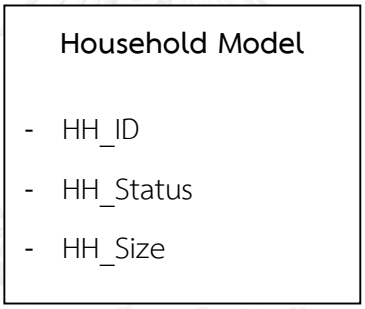


ภาพที่ 17 ขั้นตอนการทำงานของการสร้างประชากรสังเคราะห์ระดับครัวเรือน

จากคุณลักษณะของข้อมูลที่จะใช้ในขั้นตอนที่ 3.1.2 เมื่อนำมาศึกษาหาความสัมพันธ์ของข้อมูล เพื่อกำหนดลำดับการสร้างของแต่ละคุณลักษณะจะได้ลำดับดังแสดงในตารางที่ 22

ตารางที่ 22 ลำดับของกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือน

คุณลักษณะของข้อมูล	**ลำดับการสร้าง**
จำนวนครัวเรือน	1
เพศและอายุของหัวหน้าครัวเรือน	2
อายุห่างระหว่างคู่สมรส	3
การอยู่อาศัยของเด็กอายุ 0-17 ปี	4
อายุห่างระหว่างบุตรแต่ละคน	5
ขนาดครัวเรือน	6



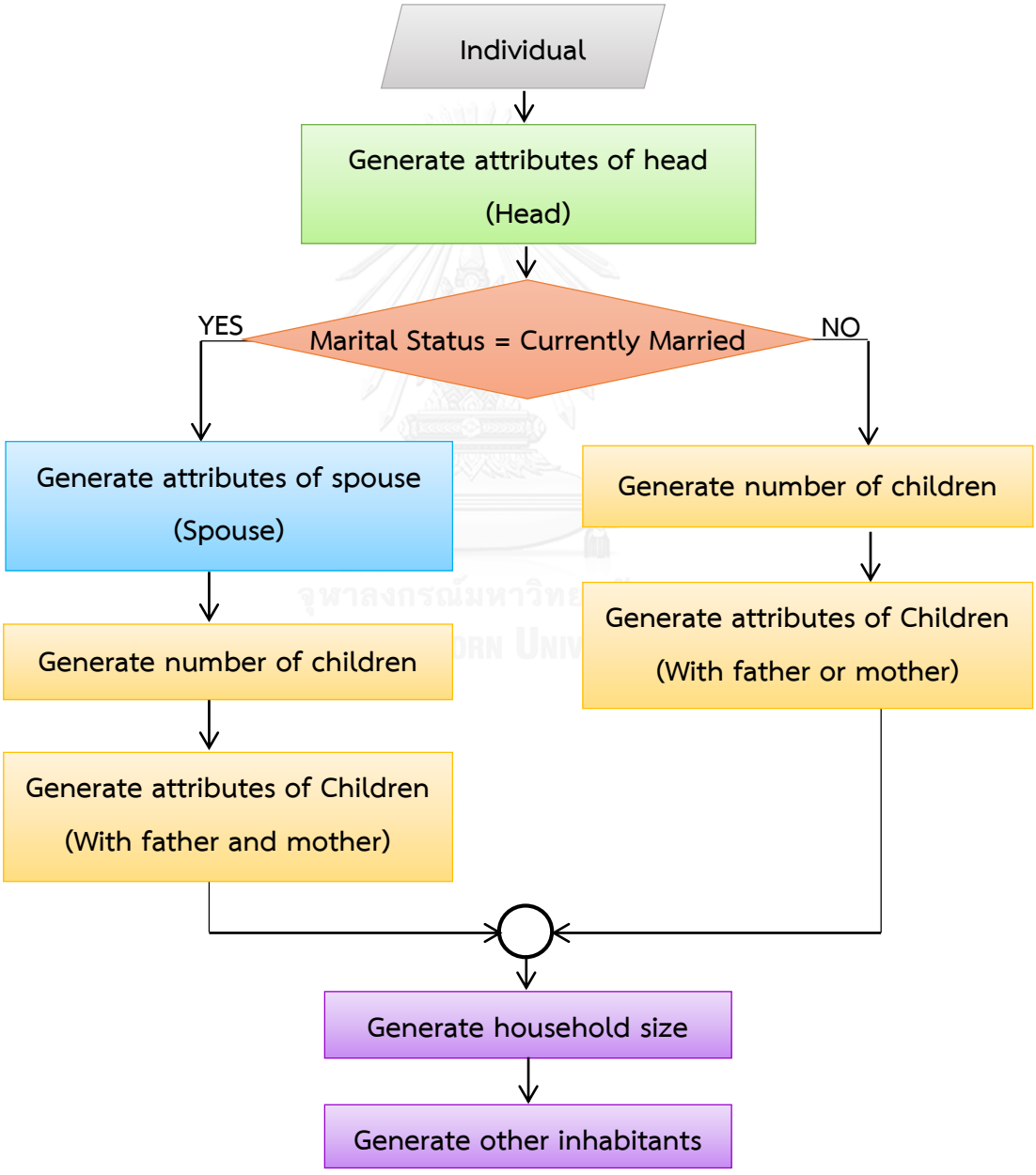
ภาพที่ 18 คุณลักษณะข้อมูลครัวเรือน

สำหรับคุณลักษณะข้อมูลครัวเรือนที่แสดงในภาพที่ 18 สามารถอธิบายรายละเอียดได้ดังนี้

HH_ID : เลขประจำครัวเรือน
HH_Status : สถานะในครัวเรือน (หัวหน้าครัวเรือน, คู่สมรส, บุตร และผู้อยู่อาศัยคนอื่น)
HH_Size : ขนาดของครัวเรือน

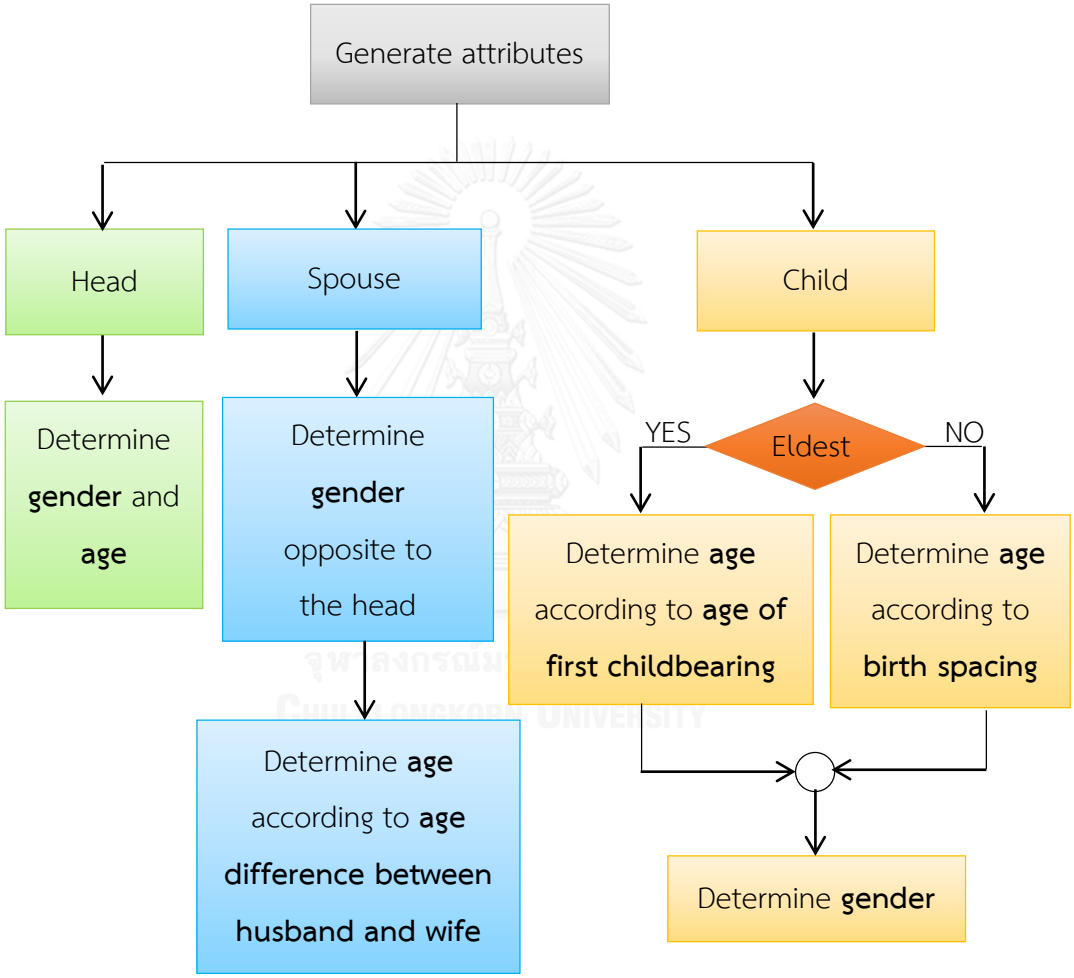
งานวิจัยนี้จะมีแนวคิดในการออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือน โดยสร้างความสัมพันธ์ของคนที่อยู่ด้วยกันในครอบครัวก่อน คือ พ่อแม่ลูก เรียกว่าหน่วยย่อยของครอบครัว (family unit) แล้วจึงนำหน่วยย่อยของครอบครัวนี้เข้าไปอยู่ในครัวเรือน (household)

ทำให้งานวิจัยนี้จะกำหนดขนาดของครัวเรือน (household size) หลังจากได้สร้างหน่วยย่อยของครอบครัวแล้ว ซึ่งการสร้างหน่วยย่อยของครอบครัวก่อนนั้น มีข้อดีคือ ประชากรสังเคราะห์ระดับครัวเรือนจะมีความหลากหลายและสอดคล้องกับความซับซ้อนของครัวเรือนในสังคมของประเทศไทย เช่น ครัวเรือนที่ถูกสร้างขึ้นจะมีการอยู่อาศัยของพ่อแม่ลูก หรือ พ่อและลูก หรือแม่และลูกก็ได้ ซึ่งแตกต่างจากงานวิจัยอื่นที่จะกำหนดขนาดของครัวเรือนก่อนแล้วจึงค่อยเลือกคนที่จะอยู่ในครัวเรือนนี้ภายหลัง ซึ่งครัวเรือนที่ถูกสร้างขึ้นจะมีการอยู่อาศัยของพ่อแม่ลูก และ ปู่ย่า/ตายาย และพ่อแม่ลูก เท่านั้น



ภาพที่ *19* กระบวนการสร้างโครงสร้างของครัวเรือน

ดังนั้นกระบวนการสร้างประชากรสังเคราะห์ระดับครัวเรือนนั้นจะแบ่งขั้นตอนออกเป็นการสร้างหัวหน้าครัวเรือน การสร้างคู่สมรส การสร้างบุตร และการสร้างผู้อยู่อาศัยอื่นๆ มีกระบวนการสร้างโครงสร้างของครัวเรือน และการสร้างคุณลักษณะของบุคคลในครัวเรือนดังแสดงในภาพที่ 19 และ 20 ตามลำดับ โดยหัวหน้าครัวเรือนจะมีเลขประจำครัวเรือนเป็นตัวระบุครัวเรือน และบุคคลอื่นที่มีความสัมพันธ์กับหัวหน้าครัวเรือนนี้ ก็จะอาศัยอยู่ในครัวเรือนเดียวกัน โดยจะมีเลขประจำครัวเรือนเดียวกันกับหัวหน้าครัวเรือน

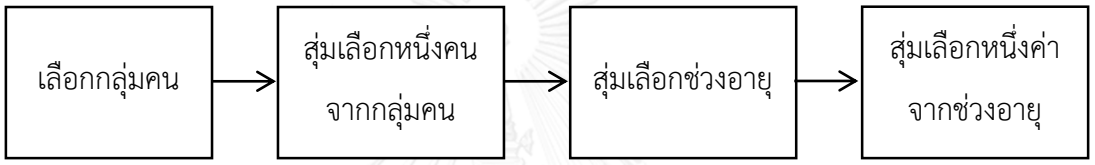


ภาพที่ 20 กระบวนการสร้างคุณลักษณะของบุคคลในครัวเรือน

เนื่องจากคุณลักษณะของข้อมูลที่เปิดเผยนั้นจะอยู่ในรูปแบบการแจกแจงของร้อยละความน่าจะเป็นที่แบ่งตามช่วงอายุ เช่น หัวหน้าครัวเรือนอายุ 21-30 ปี, 31-40 ปี, 41-50 ปี มี 25%, 35% และ

40% ตามลำดับ ทำให้กระบวนการสำหรับเลือกบุคคลเข้าในครัวเรือนจะมีลักษณะการสร้างที่คล้ายกัน ดังที่แสดงในภาพที่ 21 อธิบายได้ดังนี้

1. เลือกกลุ่มคนที่มีโอกาสที่จะถูกกำหนดค่าด้วยคุณลักษณะของข้อมูลที่ต้องการจะใส่ค่าให้
2. สุ่มเลือกหนึ่งคนจากกลุ่มที่เลือกไว้ในขั้นตอนแรก แล้วสุ่มตัวเลขเพื่อดูความน่าจะเป็นในการเลือกช่วงอายุที่จะใส่คุณลักษณะของข้อมูล
3. เมื่อได้ช่วงอายุแล้ว จะทำการสุ่มเลือกหนึ่งค่าจากช่วงอายุที่ได้ เพื่อกำหนดค่าให้กับคนที่ถูกเลือก
4. ทำขั้นตอนที่ 2-3 ซ้ำ จนครบทุกคน



ภาพที่ 21 กระบวนการสำหรับเลือกบุคคลเข้าในครัวเรือน

3.3.1 การสร้างหัวหน้าครัวเรือน

การสร้างหัวหน้าครัวเรือนจะถูกสร้างเป็นลำดับแรกในระดับครัวเรือนโดยใช้ตารางข้อมูลสถิติของจำนวนครัวเรือนจำแนกตามช่วงอายุ ซึ่งตารางข้อมูลนี้จะมีความสัมพันธ์กันกับคุณลักษณะข้อมูลของเขตการปกครอง, อายุ และเพศซึ่งแสดงเป็นตารางข้อมูลในรูปแบบร้อยละของแต่ละช่วงอายุและเพศตามเขตการปกครองดังแสดงตัวอย่างในตารางที่ ก.6 เพื่อต้องการเลือกอายุของหัวหน้าครัวเรือน ประกอบกับตารางข้อมูลของจำนวนครัวเรือนแยกตามอำเภอและเขตการปกครองซึ่งแสดงข้อมูลเป็นจำนวนคนดังแสดงตัวอย่างตามตารางที่ ก.1 เพื่อเลือกหัวหน้าครัวเรือนให้ครบตามจำนวน ซึ่งพบว่ามีข้อมูลเหล่านี้มีความไม่สอดคล้องกัน เนื่องจากมีการเก็บข้อมูลที่มีความละเอียดไม่เท่ากัน และยังเป็นความละเอียดต่ำ สำหรับการสร้างหัวหน้าครัวเรือนจะใช้ Monte Carlo โดยเป็นการสุ่มความน่าจะเป็นจากเปอร์เซ็นต์สะสม (cumulative percent) ที่ได้จากผลรวมของตารางข้อมูลในรูปแบบร้อยละของเขตการปกครอง, อายุ และเพศโดยอ้างอิงจำนวนครัวเรือนจากตารางข้อมูลของจำนวนครัวเรือนแยกตามอำเภอและเขตการปกครอง โดยแสดงตัวอย่างของการสร้างหัวหน้าครัวเรือนดังตารางที่ 23 คือ สมมติข้อมูลตารางสถิติแสดงความน่าจะเป็นของอายุหัวหน้าครัวเรือนที่มีช่วงอายุ 20-25 ปี (Dif1) มี 35%, ช่วงอายุ 26-30 ปี (Dif2) มี 30% และช่วง

อายุ 31-35 ปี (Dif3) มี 35% และจำนวนครัวเรือนของพื้นที่ location1 (CountLocation1) มี x คนจะเริ่มจากการหาหัวหน้าครัวเรือนคนที่ 1 โดยสุ่มตัวเลขแบบ Monte Carlo ได้ 35 ซึ่งหมายถึงอยู่ในช่วงอายุ 20-25 ปีจากนั้นสุ่มเลือกอายุหัวหน้าครัวเรือนจากช่วงอายุ 20-25 ปีได้เป็นอายุ 23 ปี จากนั้นเลือกประชากรที่อายุ 23 ปีจาก location1 เพื่อเป็นหัวหน้าครัวเรือนนี้ จากนั้นจะลดจำนวนครัวเรือนที่เหลือไปเรื่อยทีละหนึ่ง (x-1) จนครบทั้ง location1

ตารางที่ 23 ตัวอย่างการสร้างหัวหน้าครัวเรือน

	คนที่ถูกเลือก		
ขั้นตอน	**คนที่ 1**	**คนที่ 2**	**คนสุดท้าย**
1.ความน่าจะเป็นของอายุหัวหน้าครัวเรือน Dif1 : 20-25 = 35% Dif2 : 26-30 = 30% Dif3 : 31-35 = 35%	Cum Prob Bin Dif1 : (0-35) Dif2 : (>35-65) Dif3 : (>65-100)	Cum Prob Bin Dif1 : (0-35) Dif2 : (>35-65) Dif3 : (>65-100)	Cum Prob Bin Dif1 : (0-35) Dif2 : (>35-65) Dif3 : (>65-100)
2.สุ่มตัวเลขโดยใช้โปรแกรม	35	60	87
3.เลือกช่วงอายุ	20-25	26-30	31-35
4.เลือกอายุหัวหน้าครัวเรือนโดยใช้โปรแกรม	23	30	33
5.ลดจำนวนครัวเรือนที่ถูกเลือกในอำเภอนั้นลงหนึ่งครัวเรือน CountLocation1 : x	x-1	x-1	x-1
6.เลือกหัวหน้าครัวเรือนคนถัดไป โดยทำซ้ำตั้งแต่ขั้นตอน 1	เลือกหัวหน้าครัวเรือนคนถัดไป	เลือกหัวหน้าครัวเรือนคนถัดไป	จบการทำงาน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.3.2 การสร้างคู่สมรส

การสร้างคู่สมรสจะพิจารณาอายุของคู่สมรสตามตารางข้อมูลอายุห่างระหว่างสามี-ภรรยาดังตารางที่ ก.7 เนื่องจากข้อมูลนี้เป็นการอ้างอิงเพศหญิงที่สมรสและจะมีลักษณะเป็นช่วงอายุ ดังนั้นการเลือกคู่สมรสในขั้นตอนนี้จะเลือกคู่สมรสให้ผู้หญิงที่มีสถานภาพสมรสเป็นสมรสเท่านั้น โดยคู่สมรสนี้จะต้องอยู่ในอำเภอและเขตการปกครองเดียวกัน ซึ่งจะใช้วิธี Monte Carlo เหมือนกับขั้นตอนการสร้างหัวหน้าครัวเรือน โดยเริ่มจากสุ่มเลือกผู้หญิงทีละคนแล้วพิจารณาสถานภาพสมรสว่าเป็นสมรสหรือไม่ ถ้าเป็นสมรสจะทำการเลือกคู่สมรสให้ตามช่วงอายุที่สุ่มได้ ดังตัวอย่างในตารางที่ 24 สมมติข้อมูลตารางสถิติของผู้หญิงช่วงอายุ 30-40 ปีมีความน่าจะเป็นของคู่สมรสที่อายุอ่อนกว่า 1-2 ปี (Dif1) คือ 30%, ความน่าจะเป็นของคู่สมรสที่เท่ากัน (Dif2) คือ 40% และความน่าจะเป็นของคู่สมรสที่อายุแก่กว่า 1-2 ปี (Dif3) คือ 30% จากนั้นสุ่มเลือกผู้หญิงทีละคน เช่น คนที่ 1 มีสถานภาพสมรสเป็นสมรสจะเลือกคู่สมรสให้โดยการสุ่มตัวเลขได้ 26 ซึ่งจะได้คู่สมรสอายุน้อยกว่าอยู่ในช่วงอายุ 1-2 ปี (Dif1) คืออายุ 28-29 ปีแล้วจึงมาเลือกอายุจากช่วงอายุ 28-29 ปี จะได้อายุ 29 ปี ซึ่งเป็นเพศชายเป็นคู่สมรส แล้วเลือกผู้หญิงคนต่อไป ผู้หญิงคนที่ 2 ซึ่งมีสถานภาพสมรสเป็นโสดจึงไม่เลือกคู่สมรสให้และทำต่อไปเรื่อยๆ จนครบ

ตารางที่ 24 ตัวอย่างการสร้างคู่สมรส

	คนที่ถูกเลือก		
ขั้นตอน	**คนที่ 1**	**คนที่ 2**	**คนสุดท้าย**
1.เพศ, อายุ, สถานภาพสมรสของบุคคล	เพศ : หญิง อายุ : 30 สถานภาพสมรส : สมรส	เพศ : หญิง อายุ :34 สถานภาพสมรส : โสด	เพศ : หญิง อายุ : 40 สถานภาพสมรส : สมรส
2.ความน่าจะเป็นของช่วงอายุคู่สมรส Dif1 : อ่อนกว่า 1-2 ปี = 30% Dif2 : เท่ากัน = 40% Dif3 : แก่กว่า 1-2 ปี = 30%	Cum Prob Bin Dif1 : (0-30) Dif2 : (>30-70) Dif3 : (>70-100)	Cum Prob Bin Dif1 : (0-30) Dif2 : (>30-70) Dif3 : (>70-100)	Cum Prob Bin Dif1 : (0-30) Dif2 : (>30-70) Dif3 : (>70-100)
3.สุ่มตัวเลขโดยใช้โปรแกรม	26	-	73

ตารางที่ 24 ตัวอย่างการสร้างคู่สมรส (ต่อ)

	คนที่ถูกเลือก		
ขั้นตอน	**คนที่ 1**	**คนที่ 2**	**คนสุดท้าย**
4.เลือกช่วงอายุที่ห่างกันของคู่สมรส	อ่อนกว่า 1-2 ปี : 28-29	-	แก่กว่า 1-2 ปี : 41-42
5.เลือกอายุคู่สมรสโดยใช้โปรแกรม	29	ไม่มีคู่สมรส (เนื่องจากสถานภาพสมรสเป็นโสด)	41
6.เลือกคนต่อไปทำซ้ำตั้งแต่ขั้นตอนที่ 1	เลือกผู้หญิงคนต่อไป	เลือกผู้หญิงคนต่อไป	จบการทำงาน

3.3.3 การสร้างบุตร

การสร้างจำนวนบุตรให้แต่ละครัวเรือนเป็นขั้นตอนที่มีความซับซ้อนโดยใช้ข้อมูลดังนี้คือ ตารางข้อมูลจำนวนบุตรดังแสดงข้อมูลตามตารางที่ ก.8, ตารางข้อมูลอายุเมื่อมีบุตรคนแรกดังแสดงข้อมูลตามตารางที่ ก.9, ตารางข้อมูลการอยู่อาศัยของเด็กอายุ 0-17 ปีดังแสดงข้อมูลตามตารางที่ ก.10 และอายุห่างระหว่างบุตรแต่ละคนตามตารางที่ ก.11 [29] ซึ่งตารางข้อมูลจำนวนบุตรจะถูกกำหนดให้กับผู้หญิงที่มีสถานภาพเคยสมรสคือสมรส หม้าย หย่าร้างแยกกันอยู่และไม่ทราบ และตารางข้อมูลอายุเมื่อมีบุตรคนแรกจะถูกกำหนดให้กับผู้หญิงที่มีสถานภาพเคยสมรสมีบุตรแล้วในขั้นตอนการออกแบบกระบวนการสร้างประชากรสังเคราะห์ระดับบุคคล (หัวข้อ 3.2) จากขั้นตอนนั้นทำให้สามารถรู้อายุของบุตรคนแรกได้ นั่นคือ อายุของมารดาลบด้วยอายุเมื่อมีบุตรคนแรก

สำหรับการเลือกบุตรที่อยู่อาศัยในครัวเรือนจะใช้ตารางข้อมูลการอยู่อาศัยของเด็กอายุ 0-17 ปีและอายุห่างระหว่างบุตรแต่ละคน เพื่อคัดเลือกเด็กที่อายุ 0-17 ปีว่าอยู่อาศัยกับบิดาหรือมารดาเท่านั้นหรืออยู่ด้วยกันกับทั้งบิดาและมารดาหรือว่าไม่ได้อยู่กับทั้งบิดาและมารดา ซึ่งจะคัดเลือกเด็กที่อยู่ในอำเภอและเขตการปกครองเดียวกันกับบิดามารดา โดยการคัดเลือกเด็กนั้นจะทำโดยการเอาจำนวนเด็กที่อายุ 0-17 ปีทั้งหมดมาคูณด้วยร้อยละของการอยู่อาศัยของเด็ก ตัวอย่างดังแสดงในตารางที่ 25 โดยกำหนดให้จำนวนเด็กที่อายุ 0-17 ปีมีทั้งหมดคือ 100,000 คน จึงทำให้ได้เด็กอายุ 0- 17 ปีที่อยู่อาศัยกับบิดาและมารดามี 57,200 คน อาศัยอยู่กับมารดา 22,800 คน อาศัยอยู่กับบิดา 16,800 คนและไม่ได้อยู่กับทั้งบิดาและมารดา 3,200 คน

ตารางที่ 25 ตัวอย่างการเลือกเด็กโดยจำแนกตามการอยู่อาศัย

รูปแบบ	**อาศัยอยู่กับบิดาและมารดา (WithFather AndMother)**	**อาศัยอยู่กับมารดา (WithMother)**	**อาศัยอยู่กับบิดา (WithFather)**	**ไม่ได้อยู่กับบิดาและมารดา (WithoutFather AndMother)**
ร้อยละ	57.2	22.8	16.8	3.2
จำนวน (คน)	100000*57.2/100 = 572000	100000*22.8/100 = 22800	100000*16.8/100 = 16800	100000*3.2/100 = 3200

จากนั้นจะคัดเลือกบุตรใส่ให้ครบตามจำนวนบุตรที่แต่ละคนมีโดยจะใช้ข้อมูลอายุห่างระหว่างบุตรแต่ละคนดังที่แสดงในตารางที่ ก.11 ซึ่งข้อมูลที่ได้มาจะเป็นร้อยละอยู่ในรูปแบบช่วงของอายุห่าง คือ 1 ปี, 2 ปี, 3-4 ปี และ 5 ปีขึ้นไป โดยในที่นี้จะกำหนดให้เป็น 5-7 ปี สำหรับตารางข้อมูลการอยู่อาศัยของเด็ก 0-17 ปีจะมีความสัมพันธ์กับสถานภาพสมรส กล่าวคือบุตรที่อาศัยอยู่กับบิดาและมารดาจะเลือกมารดาที่มีสถานภาพสมรสเป็นสมรส ส่วนบุตรที่อยู่กับมารดาจะเลือกมารดาที่มีสถานภาพสมรสเป็นเคยสมรสแต่ไม่ใช่สมรสคือ หม้าย หย่า แยกกันอยู่และไม่ทราบ สำหรับบุตรที่อยู่กับบิดาจะอิงกับมารดาที่มีสถานภาพสมรสเป็นเคยสมรสเช่นเดียวกันเนื่องจากหากบิดาและมารดามีการแยกทางกันบุตรอาจจะถูกแยกให้อยู่กับบิดาหรือมารดาก็ได้โดยจะคัดเลือกเด็กที่อยู่ในอำเภอและเขตการปกครองเดียวกันกับบิดาและมารดา

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 26 ตัวอย่างการเลือกบุตรตามการอยู่อาศัย

	คนที่ถูกเลือก		
ขั้นตอน	**คนที่ 1**	**คนที่ 2**	**คนสุดท้าย**
1.เพศ, อายุ, สถานภาพสมรสของบุคคล, จำนวนบุตร	เพศ : หญิง อายุ : 35 สถานภาพสมรส : สมรส จำนวนบุตร : 2 อายุเมื่อมีบุตรคนแรก : 25	เพศ : หญิง อายุ : 45 สถานภาพสมรส : หย่า จำนวนบุตร : 2 อายุเมื่อมีบุตรคนแรก: 28	เพศ : หญิง อายุ : 50 สถานภาพสมรส : หย่า จำนวนบุตร : 2 อายุเมื่อมีบุตรคนแรก: 30

ตารางที่ 26 ตัวอย่างการเลือกบุตรตามการอยู่อาศัย (ต่อ)

	คนที่ถูกเลือก		
ขั้นตอน	**คนที่ 1**	**คนที่ 2**	**คนสุดท้าย**
	ดังนั้นลูกคนแรกอายุ : 35-25 = 10 ปี	ดังนั้นลูกคนแรกอายุ : 45-28 = 17 ปี	ดังนั้นลูกคนแรกอายุ : 50-30 = 20 ปี
2.ความน่าจะเป็นของอายุห่างระหว่างบุตร Dif1 : 1 ปี = 26.4% Dif2 : 2 ปี = 26.2% Dif3 : 3-4 ปี = 31.8% Dif4 : 5-7 ปี = 15.6%	Cum Prob Bin Dif1: (0-26.4) Dif2: (>26.4-52.6) Dif3: (>52.6-84.4) Dif4 : (>84.4-100)	Cum Prob Bin Dif1: (0-26.4) Dif2: (>26.4-52.6) Dif3: (>52.6-84.4) Dif4 : (>84.4-100)	Cum Prob Bin Dif1: (0-26.4) Dif2: (>26.4-52.6) Dif3: (>52.6-84.4) Dif4 : (>84.4-100)
3.สุ่มตัวเลขโดยใช้โปรแกรม	53	85	91
4.เลือกช่วงอายุห่างระหว่างบุตร	3-4 ปี	5-7 ปี	5-7 ปี
5.เลือกอายุห่างระหว่างบุตรโดยใช้โปรแกรม	3 ปี	6 ปี	7 ปี
6.เลือกอายุบุตร	10-3 = 7 ปี	17-6 = 10 ปี	20-7 = 13 ปี
7.ลดจำนวนบุตรตามการอยู่อาศัยลงตามจำนวนบุตร WithFatherAndMother : x WithMother : y WithFather : z	x-2	y-2	กรณีที่บุตรอาศัยอยู่กับมารดาครบแล้ว (y=0) บุตรคนนี้จะถูกเลือกให้อยู่กับบิดาแทน z-2
8.ทำซ้ำตั้งแต่ขั้นตอนที่1 จนได้ครบตามจำนวนบุตรแล้วเลือกคนต่อไป	เลือกคนต่อไป	เลือกคนต่อไป	จบการทำงาน

สำหรับตัวอย่างแสดงในตารางที่ 26 สมมติข้อมูลตารางสถิติแสดงจำนวนบุตรที่อยู่กับบิดาและมารดา (WithFatherAndMother) มี x คน จำนวนบุตรที่อยู่กับมารดา (WithMother) มี y คน จำนวนบุตรที่อยู่กับบิดา (WithFather) มี z คน และข้อมูลตารางสถิติแสดงความน่าจะเป็นของ

อายุห่างระหว่างบุตร 1 ปี (Dif1) มี 26.4%, 2 ปี (Dif2) มี 26.2%, 3-4 ปี (Dif3) มี 31.8% และ 5-7 ปี (Dif4) มี 15.6% โดยเริ่มจากสุ่มเลือกผู้หญิงที่มีจำนวนบุตรที่ไม่ใช่ 0 คน เช่น คนที่ 1 เพศหญิง สถานภาพเป็นสมรสจำนวนบุตร 2 คน อายุเมื่อมีบุตรคนแรกคือ 25 ปีดังนั้นจะมีบุตรคนแรกอายุ 35 ลบกับ 25 ปีเท่ากับ 10 ปี มาสุ่มตัวเลขเพื่อหาความน่าจะเป็นของอายุห่างระหว่างบุตรได้เลข 53 ได้บุตรที่มีอายุห่าง 3-4 ปี (Dif3) ซึ่งสุ่มเลือกอายุห่างของบุตรคนถัดมาได้ 3 ปี ทำให้ได้บุตรคนที่ 2 มีอายุ 10 ลบ 3 เท่ากับ 7 ปี จะลดจำนวนบุตรที่อยู่กับบิดาและมารดา (WithFatherAndMother) ลง x ลบจำนวนบุตรเท่ากับ x-2 คน หากมีจำนวนบุตรมากกว่า 2 คนก็จะเลือกจำนวนบุตรจนครบโดยสุ่มตัวเลขเพื่อหาความน่าจะเป็นของบุตรคนถัดไปทำตามขั้นตอนที่ 3 ในตารางไปจนครบตามจำนวนบุตร จากนั้นเลือกบุคคลในลำดับถัดไปที่มีจำนวนบุตรที่ไม่ใช่ 0 คน ดังตัวอย่างคนที่ถูกเลือกคนที่ 2 เพศหญิงมีสถานภาพสมรสเป็นหย่า เมื่อเลือกอายุของบุตรได้แล้วจะลดจำนวนบุตรที่อยู่กับมารดา (WithMother) คือ y-2 คน ส่วนคนสุดท้ายเป็นเพศหญิงมีสถานภาพสมรสเป็นหย่าเมื่อเลือกอายุของบุตรได้แล้วแต่จำนวนบุตรที่อยู่กับมารดาครบแล้วคือ y = 0 บุตรคนที่ถูกเลือกนี้จะอยู่กับบิดาแทนและจะลดจำนวนบุตรที่อยู่กับบิดา (WithFather) คือ z-2 คน

3.3.4 การสร้างผู้อยู่อาศัยคนอื่นๆ

สำหรับการเลือกผู้อยู่อาศัยคนอื่นๆ ให้ครบตามขนาดของครัวเรือนจะใช้ตารางข้อมูลขนาดของครัวเรือนที่ได้รับการปรับปรุงข้อมูลแล้วจากในหัวข้อ 3.1.3 (ตารางต้นฉบับคือ ตารางที่ ก. 12) ซึ่งข้อมูลขนาดครัวเรือนนี้จะแตกต่างกันไปในแต่ละอำเภอและเขตเทศบาลเนื่องจากมีการปรับปรุงโดยการอิงจำนวนประชากรและจำนวนครัวเรือนในพื้นที่นั้น ซึ่งขนาดของครัวเรือนจะมีตั้งแต่ 1 ไปจนถึงมากกว่า 10 คน ซึ่งจากข้อมูลที่มีจะคำนวณจำนวนครัวเรือนในแต่ละขนาดครัวเรือนจากจำนวนครัวเรือนทั้งหมดคูณกับจำนวนในแต่ละขนาดครัวเรือน ตัวอย่างดังแสดงในตารางที่ 27 สมมติให้จำนวนครัวเรือนทั้งหมดในพื้นที่นี้คือ 250,000 ครัวเรือน

เมื่อได้จำนวนครัวเรือนในแต่ละขนาดครัวเรือนมาแล้วจากนั้นจะเลือกหัวหน้าครัวเรือนที่ถูกสร้างไว้แล้วในหัวข้อ 3.3.1 และดูว่ามีคนที่อาศัยอยู่กับหัวหน้าครัวเรือนนี้กี่คนนั่นคือจำนวนขั้นต่ำของขนาดครัวเรือนที่หัวหน้าครัวเรือนนี้จะอยู่ได้ แล้วจึงสุ่มขนาดของครัวเรือน จากนั้นสุ่มผู้ที่อยู่อาศัยให้ครบตามจำนวนของขนาดครัวเรือนโดยมีเงื่อนไขคือต้องอยู่ในอำเภอและเขตการปกครองเดียวกัน

ตารางที่ 27 ตัวอย่างการคำนวณจำนวนครัวเรือน

ขนาดครัวเรือน (จำนวนคน)	จำนวนครัวเรือน (ร้อยละ)	จำนวนครัวเรือนที่คำนวณได้
1	23.68023	250000 * 23.68023/100 = 59200.58
2	34.11207	250000 * 34.11207/100 = 85280.18
3	24.56973	250000 * 24.56973/100 = 61424.33
4	11.79781	250000 * 11.79781/100 = 29494.53
5	4.248773	250000 *4.248773/100 = 10621.93
6	1.224097	250000 * 1.224097/100 = 3060.243
7	0.293891	250000 * 0.293891/100 = 734.7275
8	0.06048	250000 * 0.06048/100 = 151.2
9	0.01089	250000 * 0.01089/100 = 27.225
มากกว่า 10	0.001743	250000 * 0.001743/100 = 4.3575

3.4 การประเมินกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์

การประเมินกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์ที่ออกแบบนั้นเป็นการประเมินว่ากระบวนการนี้มีความถูกต้องและมีความเหมาะสมที่จะนำไปใช้มากน้อยเพียงใด ซึ่งจะมีการประเมินทั้งระดับบุคคลและระดับครัวเรือน หากผลที่ได้มีความคาดเคลื่อนน้อยแสดงว่าข้อมูลประชากรสังเคราะห์ที่ได้สร้างขึ้นมีค่าใกล้เคียงกับประชากรจริงมาก โดยคำนวณค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error) ได้ในแต่ละเซลล์ของตาราง ดังนี้

$$\textit{Mean Absolute Error} = \frac{\sum|T' - T|}{n}$$

โดยที่ T’ คือ เปอร์เซ็นต์ของคุณลักษณะของตารางข้อมูลแบบจำลอง

T คือ เปอร์เซ็นต์ของคุณลักษณะของตารางข้อมูลต้นแบบ

n คือ จำนวนเซลล์ของตาราง

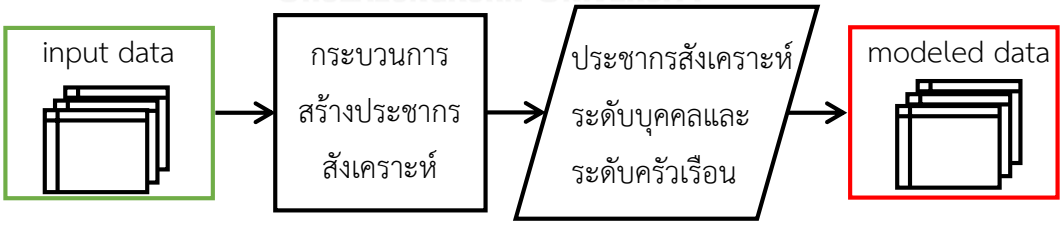
ซึ่งจะมีการประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ 3 วิธี คือ

- ประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย
- ประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์
- ประเมินโดยใช้ตารางที่ถูกซ่อน (hidden table)

3.4.1 ประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย

การประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ที่ถูกออกแบบในกรณีนี้จะใช้ข้อมูลต้นแบบ (input data) คือตารางข้อมูลที่ได้จากการสำมะโนประชากรทั้งหมดดังแสดงตามตารางที่ 13 และข้อมูลแบบจำลอง (modeled data) คือข้อมูลประชากรสังเคราะห์ที่ถูกสร้างจากกระบวนการสร้างประชากรสังเคราะห์ในขั้นตอนที่ 3.2 - 3.3 โดยนำประชากรสังเคราะห์ที่ถูกสร้างขึ้นทั้งในระดับบุคคลและระดับครัวเรือนมาเปลี่ยนกลับเป็นข้อมูลในรูปจำนวนหรือข้อมูลในรูปร้อยละที่อยู่ในรูปแบบตารางโดยมีรายละเอียดของคุณลักษณะต่างๆ เหมือนกับข้อมูลต้นแบบที่นำมาใช้ในการสร้างประชากรสังเคราะห์ ซึ่งกระบวนการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผยได้แสดงดังภาพที่ 22 จากนั้นประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ในทุกๆ คุณลักษณะของข้อมูล โดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (mean absolute error) ซึ่งการประเมินวิธีนี้จะเป็นการประเมินความแตกต่างเชิงสถิติของประชากรที่สร้างขึ้นกับประชากรตามข้อมูลสถิติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

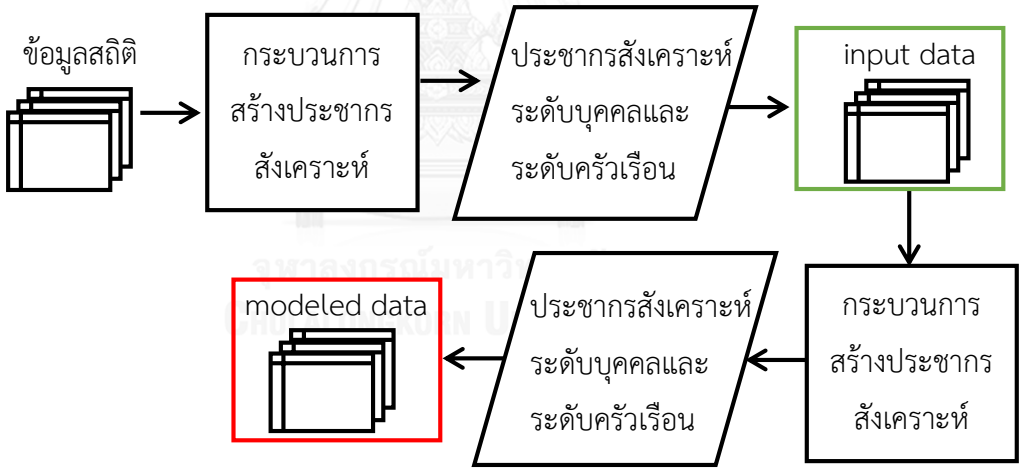


ภาพที่ 22 กระบวนการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย

3.4.2 ประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์

เนื่องจากข้อมูลที่ได้มีการเปิดเผยของประเทศไทยนั้นมีความไม่สอดคล้องกันของข้อมูลและมีการเก็บข้อมูลที่ไม่หลากหลายทำให้ข้อมูลเหล่านี้มีความคลาดเคลื่อนมากและอาจส่งผลทำให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นนั้นมีความคลาดเคลื่อนตามไปด้วย ด้วยเหตุนี้จึงทำให้ต้องมีการ

ประเมินความถูกต้องของกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์โดยใช้ข้อมูลประชากรสังเคราะห์ที่ได้ถูกสร้างขึ้นแล้วเป็นข้อมูลต้นแบบ โดยจะถือว่าข้อมูลประชากรสังเคราะห์ที่ถูกสร้างขึ้นนั้นเป็นข้อมูลที่มาจากการสำรวจอย่างสมบูรณ์และมีข้อมูลครบถ้วนในทุกคุณลักษณะของข้อมูลตามตารางที่ 13 ซึ่งในที่นี้คือผลของการสร้างประชากรสังเคราะห์ชุดแรกตามหัวข้อที่ 3.2 - 3.3 โดยจะนำประชากรสังเคราะห์เหล่านั้นมาเปลี่ยนกลับให้เป็นข้อมูลที่อยู่ในรูปแบบตารางตามคุณลักษณะของข้อมูลในตารางที่ 13 เพื่อใช้เป็นข้อมูลต้นแบบ (input data) เพื่อสร้างประชากรสังเคราะห์ชุดที่ 2 ตามหัวข้อที่ 3.2 และ 3.3 อีกครั้งจึงจะได้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นในระดับบุคคลและครัวเรือนจากการใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์ออกมาเป็นข้อมูลที่ถูกสร้างขึ้น เพื่อใช้ในการประเมินความถูกต้องและความน่าเชื่อถือของกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบ จากนั้นจะมีการนำประชากรสังเคราะห์ที่ถูกสร้างขึ้นในชุดที่ 2 ซึ่งเป็นข้อมูลแบบจำลอง (modeled data) ไปเปลี่ยนกลับให้อยู่ในข้อมูลที่เป็นรูปแบบตารางตามตารางที่ 13 ซึ่งกระบวนการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์ได้แสดงดังภาพที่ 23 จากนั้นนำมาค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (mean absolute error)



ภาพที่ 23 กระบวนการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์

3.4.3 ประเมินโดยใช้ตารางที่ถูกซ่อน

การประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ในกรณีนี้จะประเมินประชากรสังเคราะห์ที่ถูกสร้างขึ้นในแต่ละชุดว่ามีความใกล้เคียงกันมากน้อยเพียงใด โดยจะเลือกตารางคุณลักษณะของข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์มา 1 ตาราง เพื่อมาทำการเปลี่ยนแปลงขอบเขตของข้อมูลในตารางใหม่ แต่ยังคงจำนวนหรือร้อยละโดยรวมของคุณลักษณะของ

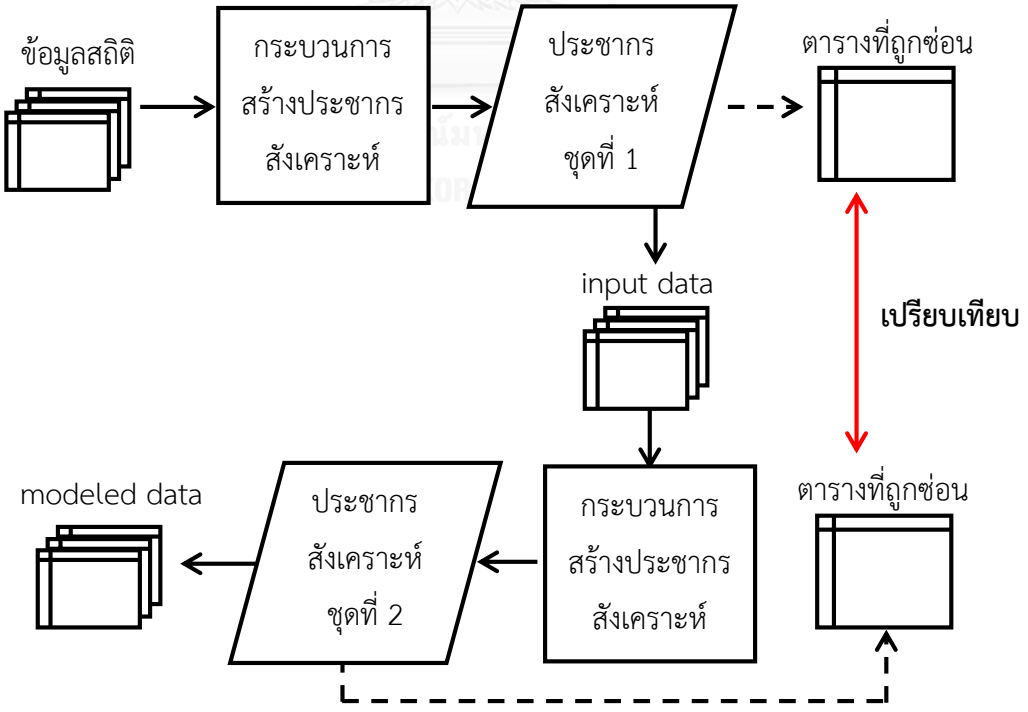
ข้อมูลในตารางนี้ไว้อย่างเดิม ซึ่งจะกำหนดให้เป็นตารางที่ถูกซ่อน (hidden table) ตัวอย่างเช่น ตารางของคุณลักษณะข้อมูลสถานภาพสมรส ซึ่งแบ่งตามช่วงอายุ ช่วงละ 10 ปี สามารถเลือกตารางข้อมูลนี้มาทำเป็นตารางที่ถูกซ่อนได้ โดยเปลี่ยนแปลงขอบเขตของข้อมูลในตารางใหม่ที่แตกต่างไปจากข้อมูลต้นแบบ เช่นแบ่งเป็นช่วงอายุละ 5 ปีแทน เป็นตารางที่ถูกซ่อนดังที่แสดงตัวอย่างในภาพที่ 24

อายุ(ปี)	โสด	สมรส
20-29	70	30
30-39	60	40
40-49	50	50

เปลี่ยนแปลงขอบเขตข้อมูล

อายุ(ปี)	โสด	สมรส
20-24	40	10
25-29	30	20
30-34	35	15
35-39	25	25
40-44	25	20
45-49	25	30

ภาพที่ 24 ตัวอย่างการสร้างตารางที่ถูกซ่อน (hidden table)



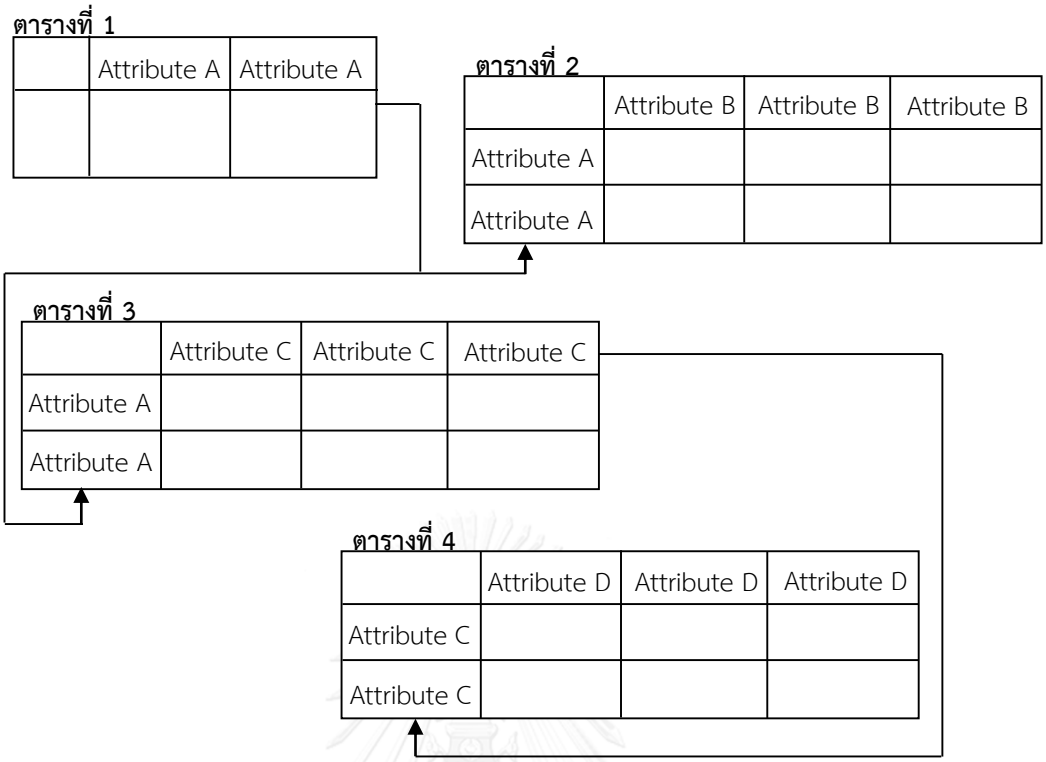
ภาพที่ 25 กระบวนการการประเมินโดยใช้ตารางที่ถูกซ่อน (hidden table)

กระบวนการการประเมินโดยใช้ตารางที่ถูกซ่อนที่แสดงในภาพที่ 25 มีวิธีการประเมิน คือ หลังจากที่ได้สร้างประชากรสังเคราะห์ชุดที่ 1 และชุดที่ 2 แล้ว ก็จะแปลงประชากรสังเคราะห์ที่ถูกสร้างขึ้นนี้มาเปลี่ยนกลับให้อยู่ในรูปแบบของตารางที่ถูกซ่อนทั้ง 2 ชุด มาเปรียบเทียบความคลาดเคลื่อนของข้อมูล เพื่อวิเคราะห์ความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ด้วยค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ (mean absolute error)

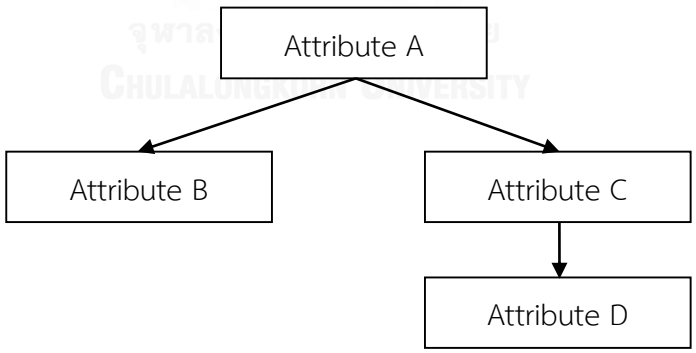
3.5 การออกแบบกระบวนการจำลองสำหรับข้อมูลอื่นๆ

นอกจากประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนที่ได้ถูกสร้างขึ้นจากการออกแบบกระบวนการสร้างประชากรสังเคราะห์นี้ยังสามารถประยุกต์กระบวนการนี้เพื่อสร้างแบบจำลองอื่นๆ ได้อีกด้วยหากมีข้อมูลที่เพียงพอ ไม่ว่าจะเป็นข้อมูลบริษัท, ข้อมูลสถานศึกษา, ข้อมูลสินค้า เป็นต้น ซึ่งส่วนใหญ่ข้อมูลเหล่านี้จะเป็นตารางข้อมูล 2 มิติที่อยู่ในรูปแบบแถวหรือคอลัมน์อยู่แล้ว โดยการสร้างแบบจำลองโดยใช้ข้อมูลอื่นๆ สามารถทำเป็นขั้นตอนได้ดังนี้

1. วิเคราะห์ข้อมูลที่อยู่ในรูปแบบตาราง 2 มิติทั้งหมดเพื่อหาความสัมพันธ์ที่เชื่อมโยงกันระหว่างแถวและคอลัมน์ของตารางข้อมูลทั้งหมด เช่นตัวอย่างในภาพที่ 26 ซึ่งมีตารางข้อมูลอยู่ 4 ตาราง หลังจากทำการวิเคราะห์ความสัมพันธ์ของแถวและคอลัมน์ของตารางทั้ง 4 แล้วพบว่ามีการเชื่อมโยงกันของความสัมพันธ์แต่ละตารางคือ Attribute B ของตารางที่ 2 นั้นจะมีความสัมพันธ์กับ Attribute A ของตารางที่ 1, Attribute C ของตารางที่ 3 จะมีความสัมพันธ์กับ Attribute A ของตารางที่ 1 และ Attribute D ของตารางที่ 4 จะมีความสัมพันธ์กับ Attribute C ของตารางที่ 3
2. เมื่อได้ความสัมพันธ์ที่เชื่อมโยงกันในแต่ละตารางแล้วก็จะสามารถวิเคราะห์ได้ว่ามีคุณลักษณะของข้อมูลอะไรบ้างที่เป็นอิสระและคุณลักษณะของข้อมูลอะไรบ้างที่ขึ้นอยู่กับคุณลักษณะอื่น เพื่อเป็นตัวกำหนดลำดับขั้นตอนของการสร้างแบบจำลองของข้อมูลนั้นๆ โดยจะสร้างคุณลักษณะของข้อมูลที่เป็นอิสระก่อนแล้วจึงสร้างคุณลักษณะของข้อมูลที่ขึ้นอยู่กับคุณลักษณะอื่นๆ ต่อไป จากตัวอย่างดังแสดงในภาพที่ 26 จะเห็นว่าคุณลักษณะของ Attribute A จะเป็นอิสระไม่ขึ้นอยู่กับคุณลักษณะอื่นๆ แต่คุณลักษณะของ Attribute B และ Attribute C เป็นคุณลักษณะที่ขึ้นอยู่กับคุณลักษณะของ Attribute A ส่วนคุณลักษณะของ Attribute D เป็นคุณลักษณะที่ขึ้นอยู่กับ Attribute C จึงได้ความสัมพันธ์ของแต่ละคุณลักษณะของข้อมูลดังแสดงในภาพที่ 27



ภาพที่ 26 ตัวอย่างตารางของข้อมูลและความสัมพันธ์ระหว่างตาราง



ภาพที่ 27 ความสัมพันธ์ของแต่ละคุณลักษณะข้อมูล

3. จากความสัมพันธ์ของแต่ละคุณลักษณะของข้อมูลที่ได้ในขั้นตอนที่ 2 ในลำดับต่อไปจะกำหนดลำดับขั้นตอนของการสร้างแบบจำลองของข้อมูลนี้ โดยสร้างคุณลักษณะของข้อมูลที่เป็นอิสระก่อนเป็นลำดับแรก หลังจากนั้นจึงสร้างคุณลักษณะของข้อมูลที่ขึ้นอยู่กับ

คุณลักษณะของข้อมูลก่อนหน้านั้นเป็นลำดับต่อไป ซึ่งจากความสัมพันธ์ของแต่ละคุณลักษณะของข้อมูลดังแสดงในภาพที่ 27 จะต้องสร้างข้อมูลของ Attribute A ก่อนเป็นอันดับแรก ต่อมาจึงสร้างข้อมูล Attribute B และ C โดยสามารถสร้างข้อมูลได้พร้อมกัน แล้วจึงสร้างข้อมูล Attribute D เป็นลำดับสุดท้าย เนื่องจาก Attribute D ขึ้นอยู่กับ Attribute C และ Attribute C ขึ้นอยู่กับ Attribute A ซึ่งลำดับขั้นตอนการสร้างดังแสดงในตารางที่ 28 หากข้อมูลมีความสัมพันธ์ที่หลากหลายแบบ ให้ข้อมูลที่ถูกอ้างอิงมากที่สุดและมีความละเอียดมากที่สุดมาสร้างก่อนเป็นอันดับแรก

ตารางที่ 28 ลำดับการสร้างข้อมูลในกระบวนการสร้างแบบจำลอง

คุณลักษณะข้อมูล	**ลำดับการสร้าง**
Attribute A	1
Attribute B	2
Attribute C	2
Attribute D	3

การสร้างแบบจำลองสำหรับข้อมูลอื่นๆ นั้นสามารถสร้างคุณลักษณะของข้อมูลต่างๆ ได้มากเท่าที่มีข้อมูลตามที่ต้องการ นั่นคือไม่เพียงแต่จะสร้างคุณลักษณะข้อมูลเพียง Attribute A, B, C และ D เท่านั้นแต่ยังสามารถสร้างคุณลักษณะอื่นๆ เพิ่มได้อีกโดยหาความสัมพันธ์ของตารางและความสัมพันธ์ของคุณลักษณะของข้อมูลใน 3 ขั้นตอนดังกล่าว ก็จะสามารถสร้างแบบจำลองได้โดยที่มีคุณลักษณะของข้อมูลที่หลากหลายมากยิ่งขึ้น สำหรับปัญหาที่มีผลต่อความถูกต้องของการสร้างแบบจำลองต่างๆ นั้นจะมีความเกี่ยวข้องกับข้อมูลที่จะใช้ในการสร้างแบบจำลอง กล่าวคือหากข้อมูลที่ใช้มีจำนวนคุณลักษณะที่เพียงพอและมีการเชื่อมโยงความสัมพันธ์ในทุกคุณลักษณะของตารางจะทำให้แบบจำลองที่สร้างขึ้นมีความหลากหลาย และมีความถูกต้อง ใกล้เคียงกับของจริง แต่หากข้อมูลที่นำมาใช้นั้นมีคุณลักษณะที่จะใช้ในการเชื่อมโยงความสัมพันธ์ของตารางที่หายไปก็จะส่งผลให้แบบจำลองที่สร้างขึ้นนั้นมีความคลาดเคลื่อนไปจากความเป็นจริงด้วย สำหรับตัวอย่างของการสร้างแบบจำลองอื่นๆ นั้นจะทำตาม 3 ขั้นตอนดังที่ได้อธิบาย เช่น ต้องการสร้างแบบจำลองของบุคคลที่อยู่ในมหาวิทยาลัยซึ่งมีข้อมูลในรูปแบบของตาราง 2 มิติอยู่ทั้งหมด 4 ตาราง ดังแสดงในภาพที่ 28 ซึ่งจะสร้างแบบจำลองตามขั้นตอนได้ดังนี้

ตารางแสดงจำนวนบุคคลในมหาวิทยาลัยจำแนกตามประเภท

ประเภท	นิสิต	อาจารย์
จำนวน (คน)	5000	100

ตารางแสดงจำนวนอาจารย์จำแนกตามคณะ

คณะ	คณะวิทยาศาสตร์	คณะวิศวกรรมศาสตร์
อาจารย์ (คน)	45	55

ตารางแสดงจำนวนนิสิตจำแนกตามคณะ

คณะ	คณะวิทยาศาสตร์	คณะวิศวกรรมศาสตร์
จำนวน (คน)	2200	2800

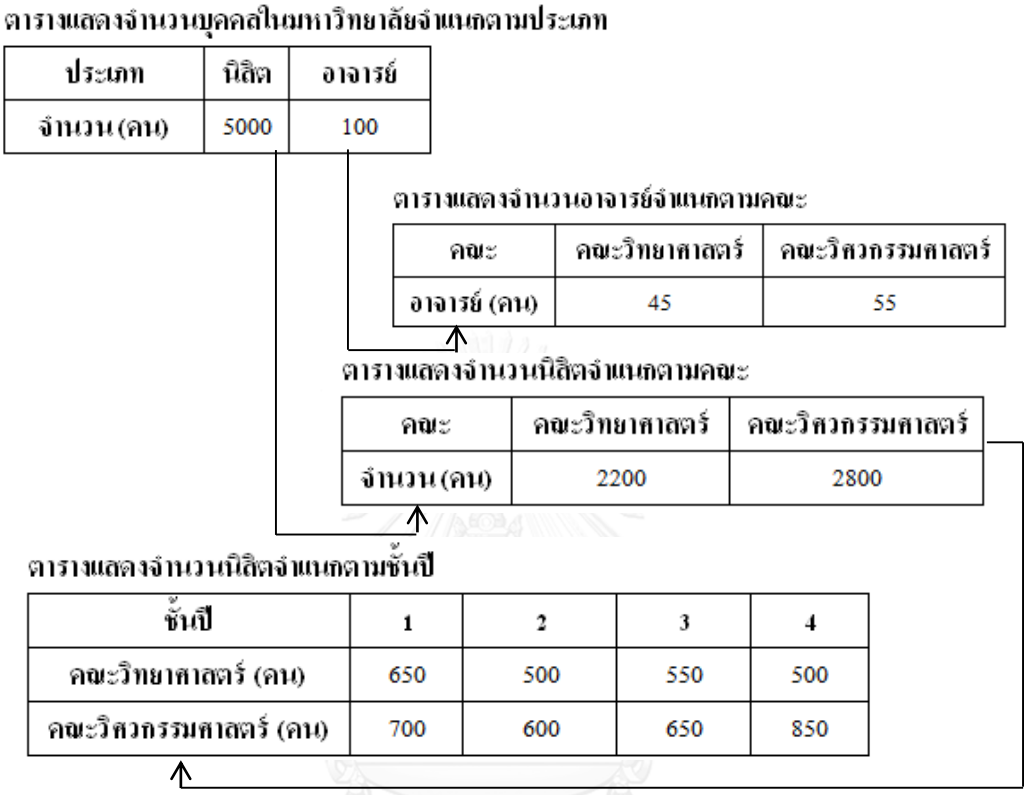
ตารางแสดงจำนวนนิสิตจำแนกตามชั้นปี

ชั้นปี	1	2	3	4
คณะวิทยาศาสตร์ (คน)	650	500	550	500
คณะวิศวกรรมศาสตร์ (คน)	700	600	650	850

ภาพที่ 28 ข้อมูลที่ใช้ในการสร้างแบบจำลองของมหาวิทยาลัย

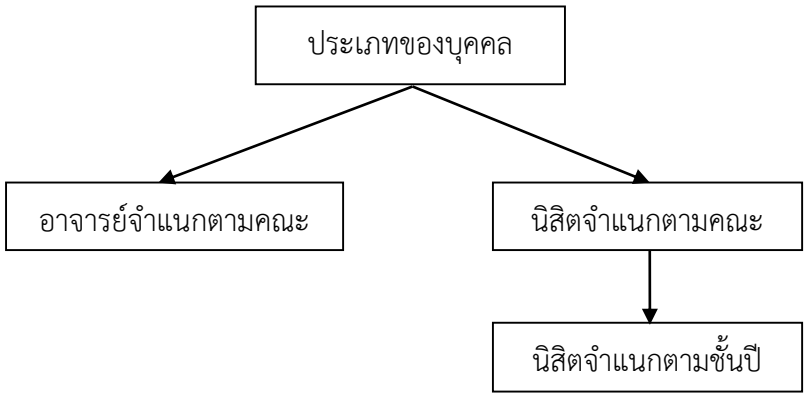
1. ตารางของข้อมูลและหาความสัมพันธ์ของการเชื่อมโยงระหว่างตารางต่างๆ โดยจะเห็นได้ว่าข้อมูลจำนวนอาจารย์ดังแสดงในตารางจำนวนบุคคลในมหาวิทยาลัยจำแนกตามประเภทนั้นจะมีความสัมพันธ์กับตารางจำนวนอาจารย์จำแนกตามคณะ, จำนวนนิสิตดังแสดงในตารางจำนวนบุคคลในมหาวิทยาลัยจะมีความสัมพันธ์กับตารางจำนวนนิสิตจำแนกตามคณะและจำนวนนิสิตดังแสดงในตารางจำนวนนิสิตจำแนกตามคณะก็จะมีความสัมพันธ์กับตารางจำนวนนิสิตจำแนกตามชั้นปี ดังแสดงในภาพที่ 29
2. วิเคราะห์คุณลักษณะของข้อมูลว่ามีคุณลักษณะของข้อมูลใดบ้างที่เป็นอิสระและข้อมูลใดที่เป็นคุณลักษณะที่ขึ้นกับข้อมูลอื่น ซึ่งจากภาพที่ 29 จะทำให้ได้ความสัมพันธ์ของข้อมูลคือจำนวนนิสิตและจำนวนอาจารย์ดังแสดงในตารางจำนวนบุคคลในมหาวิทยาลัยจำแนกตามประเภทเป็นข้อมูลอิสระ แต่จำนวนอาจารย์แต่ละคณะจะขึ้นอยู่กับข้อมูลจำนวนอาจารย์ในตารางจำนวนบุคคลในมหาวิทยาลัยจำแนกตามประเภท, จำนวนนิสิตจำแนกตามคณะก็จะขึ้นอยู่กับข้อมูลจำนวนนิสิตในตารางจำนวนบุคคลในมหาวิทยาลัยจำแนกตามประเภทและจำนวนนิสิตจำแนกตามชั้นปีก็จะขึ้นอยู่กับตารางข้อมูลจำนวนนิสิตจำแนกตามคณะ

เป็นต้น ข้อมูลเหล่านี้จึงเป็นตัวอย่างของข้อมูลคุณลักษณะที่ขึ้นกับข้อมูลอื่น ดังแสดงในภาพที่ 30



ภาพที่ *29* ความสัมพันธ์ของตารางข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ *30* ความสัมพันธ์ของตัวอย่างข้อมูล

3. กำหนดลำดับขั้นตอนของการใส่ข้อมูลคุณลักษณะเพื่อสร้างแบบจำลองของบุคคลที่อยู่ในมหาวิทยาลัย จากความสัมพันธ์ของข้อมูลดังแสดงในภาพที่ 30 นั้นจะทำให้ได้ลำดับขั้นตอนของการสร้างข้อมูลที่แสดงในตารางที่ 29

ตารางที่ 29 ลำดับการสร้างข้อมูลในกระบวนการสร้างแบบจำลอง

คุณลักษณะข้อมูล	**ลำดับการสร้าง**
ประเภทของบุคคล	1
อาจารย์จำแนกตามคณะ	2
นิสิตจำแนกตามคณะ	2
นิสิตจำแนกตามชั้นปี	3

สำหรับการสร้างแบบจำลองนั้นสามารถทำได้หลายระดับโดยจะเรียกว่าระดับตัวแทนและระดับกลุ่ม ซึ่งข้อมูลในระดับตัวแทนก็คือการใส่ข้อมูลให้แต่ละตัวแทนตามที่ได้กล่าวข้างต้นทั้ง 3 ขั้นตอน และข้อมูลระดับกลุ่มจะเป็นการรวมข้อมูลในระดับตัวแทนที่มีความสัมพันธ์กันไว้ให้อยู่ในกลุ่มเดียวกัน เช่น ครัวเรือน คณะ แผนก เป็นต้น โดยที่แต่ละครัวเรือนจะมีการรวมกลุ่มกันของพ่อ แม่ และลูก, แต่ละคณะในมหาวิทยาลัยก็จะมีการรวมกลุ่มกันของนิสิต อาจารย์และเจ้าหน้าที่, แต่ละแผนกก็จะมีการรวมกลุ่มกันของหัวหน้า เลขานุการและคณะกรรมการ เป็นต้น

ดังนั้นจึงเห็นได้ว่ากระบวนการสร้างแบบจำลองที่ผู้จัดทำได้ออกแบบนั้นสามารถนำไปประยุกต์ใช้กับข้อมูลอื่นๆ ได้ เพื่อสร้างแบบจำลองอื่นๆ ที่สนใจ อีกทั้งยังสามารถสร้างแบบจำลองได้หลายระดับของข้อมูลตามที่ต้องการแม้ว่าข้อมูลเหล่านั้นจะเป็นเพียงข้อมูลในรูปแบบตาราง 2 มิติแต่หากข้อมูลเหล่านั้นมีความสัมพันธ์ที่เชื่อมโยงกันก็สามารถที่จะสร้างแบบจำลองออกมาได้ไม่ว่าจะเป็นข้อมูลที่เกี่ยวกับรถยนต์ สินค้าต่างๆหรือการทำงานของบุคคลในองค์กร เป็นต้น ซึ่งแบบจำลองเหล่านี้จะทำให้สามารถศึกษาข้อมูลในส่วนต่างๆ ได้อย่างง่ายดายและเป็นประโยชน์ต่อการศึกษาปรับปรุง เปลี่ยนแปลง รวมทั้งการวางแผนหรือเกี่ยวข้องกับการวางนโยบายในด้านต่างๆ ต่อไปในอนาคต

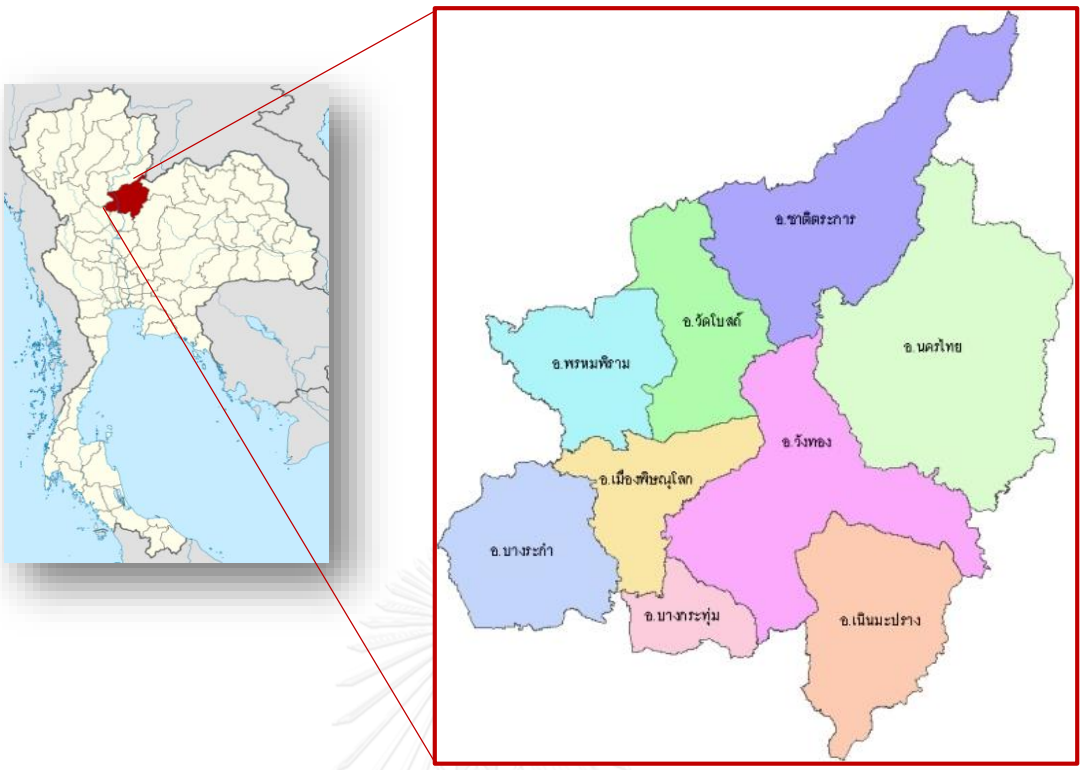
บทที่ 4
การทดลองและผลการทดลอง

ในบทนี้จะอธิบายถึงการประยุกต์ใช้กระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบที่ได้นำเสนอในหัวข้อ 3.2 - 3.3 มาใช้กับกรณีศึกษา ซึ่งคือข้อมูลของจังหวัดพิษณุโลก เพื่อนำข้อมูลประชากรสังเคราะห์ของจังหวัดพิษณุโลกที่ได้มาประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบที่นำเสนอในหัวข้อ 3.4 ดังนั้นในบทนี้ผู้วิจัยจะอธิบายรายละเอียดซึ่งประกอบไปด้วย

- การประยุกต์ใช้กระบวนการสร้างประชากรสังเคราะห์กับข้อมูลของจังหวัดพิษณุโลก
- การออกแบบและพัฒนาโปรแกรมการสร้างประชากรสังเคราะห์
- ผลของการประเมินกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์
    - ผลของการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย
    - ผลของการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์
    - ผลของการประเมินโดยใช้ตารางที่ถูกซ่อน (hidden table)

4.1 การประยุกต์ใช้กระบวนการสร้างประชากรสังเคราะห์กับข้อมูลของจังหวัดพิษณุโลก

จากกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบทั้งในระดับบุคคลและครัวเรือน ในหัวข้อที่ 3.2 - 3.3 นี้จะนำมาประยุกต์ใช้ในการสร้างประชากรสังเคราะห์ของจังหวัดพิษณุโลก เนื่องจากสำนักงานสถิติแห่งชาติได้เลือกทำสำมะโนทดลองที่จังหวัดพิษณุโลก เพราะเป็นจังหวัดที่มีโครงสร้างของประชากรทุกรูปแบบ สามารถเป็นตัวแทนของทั้งประเทศได้ ซึ่งเป็นจังหวัดที่อยู่ทางภาคเหนือและมีอำเภอต่างๆ ดังที่แสดงในภาพที่ 31 ดังนั้นผู้วิจัยจึงสนใจเลือกจังหวัดพิษณุโลกเป็นกรณีศึกษา ซึ่งข้อมูลที่ใช้จะเป็นข้อมูลของปี 2553 เพื่อนำมาสร้างประชากรสังเคราะห์ทั้งระดับบุคคลและระดับครัวเรือนของจังหวัดพิษณุโลก ซึ่งมีจำนวนประชากร 835,555 คน จำนวนครัวเรือน 296,807 ครัวเรือน ประกอบด้วย 18 เขตเทศบาลจาก 9 อำเภอ ซึ่งค่าของแต่ละคุณลักษณะของข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับบุคคลและครัวเรือนจะแสดงในตารางที่ 30 และตารางที่ 31 ตามลำดับ



ภาพที่ 31 สถานที่ตั้งและอำเภอต่างๆ ของจังหวัดพิษณุโลก

ตารางที่ 30 ค่าของแต่ละคุณลักษณะของข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับบุคคลของจังหวัดพิษณุโลก

คุณลักษณะของข้อมูล	ข้อมูล
จำนวนคน	835,555 คน
เพศ	ชาย, หญิง
อำเภอที่อยู่อาศัย	9 อำเภอ
เขตการปกครองที่อยู่อาศัย	2 เขตเทศบาล
อายุ	น้อยกว่า 1 ปี – มากกว่า 100 ปี
การศึกษา	อนุบาล 1 – มัธยมศึกษาปีที่ 6
จำนวนนักเรียนแยกรายสถานศึกษา	481 โรงเรียน
สถานภาพสมรส	โสด, สมรส, หม้าย, หย่า, แยกกันอยู่, ไม่ทราบ, ไม่ทราบสถานภาพสมรส
จำนวนบุตร	1 คน – มากกว่า 10 คน, ไม่ทราบ
อายุเมื่อมีบุตรคนแรก	13 ปี – 39 ปี

ตารางที่ 31 ค่าของแต่ละคุณลักษณะของข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับครัวเรือนของจังหวัดพิษณุโลก

คุณลักษณะของข้อมูล	**ข้อมูล**
จำนวนครัวเรือน	296,807 ครัวเรือน
เพศและอายุของหัวหน้าครัวเรือน	ชายอายุ 12 ปี – มากกว่า 100 ปี, หญิงอายุ 12 ปี – มากกว่า 100 ปี
อายุห่างระหว่างคู่สมรส	อายุสามีอ่อนกว่า 7 ปี – อายุสามีแก่กว่า 15 ปี
การอยู่อาศัยของเด็กอายุ 0-17 ปี	อยู่กับพ่อและแม่, อยู่กับพ่อเท่านั้น, อยู่กับแม่เท่านั้น, ไม่อยู่กับพ่อและแม่
อายุห่างระหว่างบุตรแต่ละคน	1 ปี – 7 ปี
ขนาดครัวเรือน	1 คน – มากกว่า 10 คน

จากการประยุกต์ใช้กระบวนการสร้างประชากรสังเคราะห์กับข้อมูลของจังหวัดพิษณุโลก ทำให้ได้ฐานข้อมูลประชากรรายบุคคล 835,555 คน และฐานข้อมูลระดับครัวเรือน 296,807 ครัวเรือน ซึ่งจะเป็นการเลือกประชากรรายบุคคลที่มีความสัมพันธ์กันเข้าอยู่ในครัวเรือนเดียวกันอย่างหลากหลาย ดังตัวอย่างข้อมูลแบบจำลองระดับครัวเรือนในภาพที่ 32 สำหรับตัวอย่างครัวเรือนที่มีขนาดครัวเรือน 1 คนในภาพที่ 32(ก) มีหัวหน้าครัวเรือนเป็นเพศชาย อาศัยอยู่ใน อ.เมืองพิษณุโลก นอกเขตเทศบาล อายุ 14 ปี เรียนอยู่ชั้นมัธยมศึกษาปีที่ 3 ที่โรงเรียนพิษณุโลกพิทยาคม (รหัสโรงเรียน 1065360459), ภาพที่ 32(ข) แสดงครัวเรือนที่มีขนาดครัวเรือน 3 คน ซึ่งมีหัวหน้าครัวเรือนเป็นเพศหญิง อาศัยอยู่ใน อ.นครไทย นอกเขตเทศบาล อายุ 36 ปี มีลูก 1 คนเป็นเพศชายอายุ 13 ปี เรียนอยู่ชั้นมัธยมศึกษาปีที่ 2 ที่โรงเรียนบ้านหนองหิน (รหัสโรงเรียน 1065360160) และมีผู้อาศัยด้วยอีก 1 คนเป็นเพศหญิง อายุ 50 ปี และภาพที่ 32(ค) แสดงครัวเรือนที่มีขนาดครัวเรือน 7 คน ซึ่งมีหัวหน้าครัวเรือนเป็นเพศชาย อาศัยอยู่ใน อ.เมืองพิษณุโลก นอกเขตเทศบาล อายุ 28 ปี มีคู่สมรสเป็นเพศหญิง อายุ 25 ปี มีบุตรด้วยกัน 5 คน เป็นเพศชาย 2 คน คือ อายุ 1 ปี และ 4 ปี และเป็นเพศหญิง 3 คน คือ อายุน้อยกว่า 1 ปี, 2 ปี และ 3 ปี

ID	Gender	Location	Area	Age	BDay	Byear	Edu_Level	SchoolCode	Marital_Status	Number_children	Age_At_First_Birth	HH_ID	HH_Status	HH_Size
16259	male	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	14	82	1996	ม.3	1065360459	โสด	0	0	123	Head	1

(ก) ขนาดครัวเรือน 1 คน

ID	Gender	Location	Area	Age	BDay	Byear	Edu_Level	SchoolCode	Marital_Status	Number_children	Age_At_First_Birth	HH_ID	HH_Status	HH_Size
136974	male	อ.นครไทย	นอกเขต เทศบาล	13	81	1997	ม.2	1065360160	โสด	0	0	232577	Child	0
569933	female	อ.นครไทย	นอกเขต เทศบาล	36	243	1974		0	หย่า	1	23	232577	Head	3
578059	female	อ.นครไทย	นอกเขต เทศบาล	50	203	1960		0	สมรส	4	0	232577	Other	0

(ข) ขนาดครัวเรือน 3 คน

ID	Gender	Location	Area	Age	BDay	Byear	Edu_Level	SchoolCode	Marital_Status	Number_children	Age_At_First_Birth	HH_ID	HH_Status	HH_Size
1593	male	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	1	130	2009		0		0	0	192246	Child	0
4110	male	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	4	294	2006		0		0	0	192246	Child	0
41816	male	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	28	340	1982		0	สมรส	0	0	192246	Head	0
410513	female	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	น้อย กว่า1	23	2010		0		0	0	192246	Child	0
412171	female	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	2	129	2008		0		0	0	192246	Child	0
413371	female	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	3	211	2007		0		0	0	192246	Child	0
444320	female	อ.เมือง พิษณุโลก	นอกเขต เทศบาล	25	218	1985		0	สมรส	5	21	192246	Spouse	7

(ค) ขนาดครัวเรือน 7 คน

ภาพที่ 32 ตัวอย่างข้อมูลแบบจำลองระดับครัวเรือน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2 การออกแบบและพัฒนาโปรแกรมการสร้างประชากรสังเคราะห์

4.2.1 เครื่องมือที่ใช้ในการพัฒนา

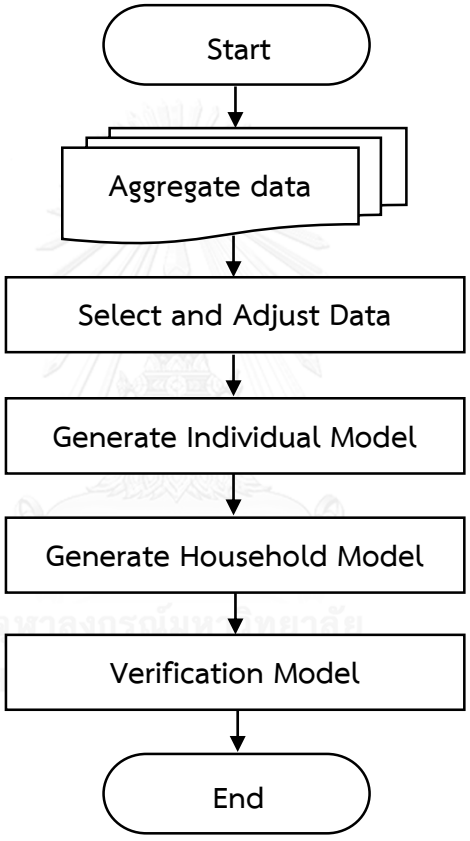
การพัฒนาโปรแกรมในงานวิจัยนี้ เป็นการดำเนินงานภายใต้ MySQL บนคอมพิวเตอร์เครื่องเดียว โดยเครื่องมือที่ใช้ในการพัฒนามีดังนี้

- โปรแกรม My SQL เป็นโปรแกรมระบบจัดการฐานข้อมูล สำหรับเก็บฐานข้อมูลในรูปแบบของตาราง และแต่ละตารางที่เก็บข้อมูลยังสามารถเชื่อมโยงเข้าหากันทำให้สามารถรวมหรือจัดกลุ่มข้อมูลได้ตามต้องการ
- ภาษา PHP เป็นภาษาสคริปต์ที่ทำงานฝั่งเครื่องบริการ (Server-Side Script) สำหรับใช้ติดต่อและจัดการฐานข้อมูลในงานวิจัยนี้
- NetBeans เป็นเครื่องมือที่ใช้สำหรับพัฒนาโปรแกรมด้วยภาษา PHP

- Microsoft Office Excel โปรแกรมสำหรับทำข้อมูลต้นแบบ (input data) ในการพัฒนาโปรแกรม

4.2.2 การออกแบบโปรแกรมสำหรับสร้างประชากรสังเคราะห์

สำหรับการสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือนในงานวิจัยนี้จะใช้ข้อมูลประชากรของจังหวัดพิษณุโลก ปี 2553 ซึ่งจะมีขั้นตอนดังที่แสดงในภาพที่ 33 สามารถอธิบายได้ดังนี้



ภาพที่ 33 ขั้นตอนการออกแบบโปรแกรมการสร้างประชากรสังเคราะห์

1. เลือกข้อมูลที่จะใช้เป็นข้อมูลต้นแบบที่เกี่ยวข้องกับคุณลักษณะของบุคคลและครัวเรือนของการสร้างประชากรสังเคราะห์จากเว็บไซต์ที่มีการเปิดเผยข้อมูล โดยจะเลือกข้อมูลที่มีระดับความละเอียดสูงสุดที่มีการเปิดเผยข้อมูลอย่างครบถ้วน

2. พิจารณาคุณลักษณะของข้อมูลว่ามีความสอดคล้องกันหรือไม่ หากไม่มีความสอดคล้องกันของข้อมูล เช่น จำนวนไม่เท่ากัน ก็จะทำการปรับปรุงข้อมูล

3. สร้างประชากรสังเคราะห์ในระดับบุคคลจากข้อมูลที่เลือกไว้ตามลำดับการสร้างที่ได้ออกแบบไว้ในหัวข้อ 3.2 โดยจะเริ่มจากการสร้างประชากรทั้งหมดตามจำนวนประชากรของข้อมูลที่มี ซึ่งประชากรนี้มีคุณลักษณะที่กำหนดรายละเอียดของบุคคลเพียงเลขประจำตัวบุคคล (IND_ID) เท่านั้น แล้วจะจึงใส่ค่าคุณลักษณะที่จะระบุตัวบุคคลทีละคุณลักษณะตามลำดับการสร้างของประชากรรายบุคคลจนครบทุกคุณลักษณะ

4. สร้างประชากรสังเคราะห์ในระดับครัวเรือนตามลำดับการสร้างที่ได้ออกแบบไว้ในหัวข้อ 3.3 โดยจะทำการสร้างความสัมพันธ์ระหว่างบุคคลที่มีความเกี่ยวข้องกัน เริ่มจากหัวหน้าครัวเรือน ซึ่งจะถูกระบุเลขประจำครัวเรือน (HH_ID) หลังจากนั้นจะหาคู่สมรส บุตรและผู้อยู่อาศัยคนอื่น โดยบุคคลที่มีความสัมพันธ์กับหัวหน้าครัวเรือนนี้จะถูกระบุเลขประจำครัวเรือนเดียวกันกับหัวหน้าครัวเรือน

5. ประเมินความถูกต้องของการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนเพื่อประเมินความถูกต้องของกระบวนการที่ได้ออกแบบ โดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ทั้ง 3 วิธีดังที่อธิบายในหัวข้อ 3.4

4.3 ผลของการประเมินกระบวนการที่ใช้ในการสร้างประชากรสังเคราะห์

4.3.1 ผลของการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลสถิติที่มีการเปิดเผย

การประเมินวิธีนี้จะเป็นการประเมินความแตกต่างเชิงสถิติของประชากรที่สร้างขึ้นกับประชากรตามข้อมูลสถิติ โดยทดลองใช้ข้อมูลต้นแบบคือ ข้อมูลตารางที่ได้จากการสำมะโนประชากรมาสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือน จึงได้ข้อมูลแบบจำลองประชากรสังเคราะห์เพื่อประเมินความถูกต้องของกระบวนการโดยการหาค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ ซึ่งได้ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนแสดงในตารางที่ 32 และตารางที่ 33 ตามลำดับ โดยค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของคุณลักษณะของข้อมูลที่มีค่าน้อย แสดงว่า ประชากรสังเคราะห์ที่สร้างในคุณลักษณะนั้นมีความใกล้เคียงกับประชากรจริงมาก จากตารางที่ 32 จะได้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการสร้างประชากรสังเคราะห์ในระดับบุคคล คือ 0.003% ซึ่งพบว่าประชากรสังเคราะห์ในระดับบุคคลมีความคลาดเคลื่อนจากข้อมูลต้นแบบเพียง 2 คุณลักษณะเท่านั้น คือ สถานภาพสมรสและจำนวนบุตร ซึ่งเป็นผลมาจากคุณลักษณะของข้อมูลที่มีความละเอียดต่ำ คือ เป็นความละเอียดที่อยู่ในระดับภาค ซึ่งคุณลักษณะอื่นในระดับบุคคล ได้แก่ จำนวนคน, เพศ, อำเภอ

ที่อยู่อาศัย, เขตการปกครองที่อยู่อาศัย, อายุ, การศึกษา, จำนวนนักเรียนแยกรายสถานศึกษา และอายุเมื่อมีบุตรคนแรก สามารถนำมาสร้างประชากรสังเคราะห์ในระดับบุคคลได้อย่างแม่นยำ ไม่มีความคลาดเคลื่อนเลย แสดงว่าประชากรในคุณลักษณะดังกล่าวมีความเหมือนประชากรจริงมากที่สุด ซึ่งจากค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ แสดงให้เห็นว่า ประชากรสังเคราะห์ในระดับบุคคลจะมีความคลาดเคลื่อนไป 3 คน จากการสร้างประชากรสังเคราะห์ทั้งหมด 100,000 คน

ตารางที่ 32 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับบุคคล

คุณลักษณะของข้อมูล	**ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์**	**ค่าต่ำสุดของความคลาดเคลื่อนสัมบูรณ์**	**ค่าสูงสุดของความคลาดเคลื่อนสัมบูรณ์**
จำนวนคน	0.000	0.000	0.000
เพศ	0.000	0.000	0.000
อำเภอที่อยู่อาศัย	0.000	0.000	0.000
เขตการปกครองที่อยู่อาศัย	0.000	0.000	0.000
อายุ	0.000	0.000	0.000
การศึกษา	0.000	0.000	0.000
จำนวนนักเรียนแยกรายสถานศึกษา	0.000	0.000	0.000
สถานภาพสมรส	0.015	0.000	0.700
จำนวนบุตร	0.015	0.000	0.534
อายุเมื่อมีบุตรคนแรก	0.000	0.000	0.000

จากตารางที่ 33 จะได้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการสร้างประชากรสังเคราะห์ในระดับครัวเรือนคือ 0.506% ซึ่งพบว่า ประชากรสังเคราะห์ในระดับครัวเรือนมีความคลาดเคลื่อนใน 3 คุณลักษณะ คือ เพศและอายุของหัวหน้าครัวเรือน, อายุห่างระหว่างคู่สมรส และขนาดครัวเรือน ซึ่งเป็นข้อมูลที่มีความละเอียดต่ำ คือ ระดับจังหวัดและระดับประเทศ และเนื่องจากข้อมูลทั้ง 3 คุณลักษณะในระดับครัวเรือนนี้เป็นข้อมูลที่ไม่สอดคล้องกับข้อมูลในระดับบุคคล ซึ่งเป็นผลมาจากการเก็บข้อมูลของหน่วยงานที่แตกต่างกัน ส่งผลให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นมีความคลาดเคลื่อนด้วย ซึ่งคุณลักษณะอื่นในระดับครัวเรือน ได้แก่ จำนวนครัวเรือน, การอยู่อาศัยของเด็ก

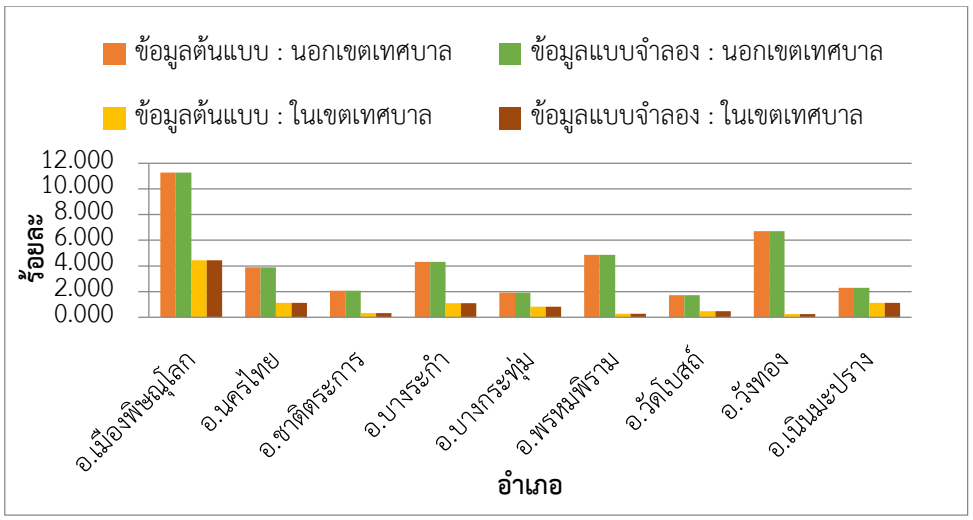
อายุ 0-17 ปี และอายุห่างระหว่างบุตรแต่ละคน สามารถนำมาสร้างประชากรสังเคราะห์ในระดับครัวเรือนได้อย่างแม่นยำ ไม่มีความคลาดเคลื่อนเลย แสดงว่าจำนวนครัวเรือนและการเลือกผู้อยู่อาศัยที่เป็นเด็กเข้าอาศัยในครัวเรือนมีความเหมือนกับสังคมจริงมากที่สุด ซึ่งจากค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ แสดงให้เห็นว่า ประชากรสังเคราะห์ในระดับครัวเรือนจะมีความคลาดเคลื่อนไป 5 ครัวเรือน จากการสร้างประชากรสังเคราะห์ระดับครัวเรือนทั้งหมด 1,000 ครัวเรือน

ตารางที่ 33 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับครัวเรือน

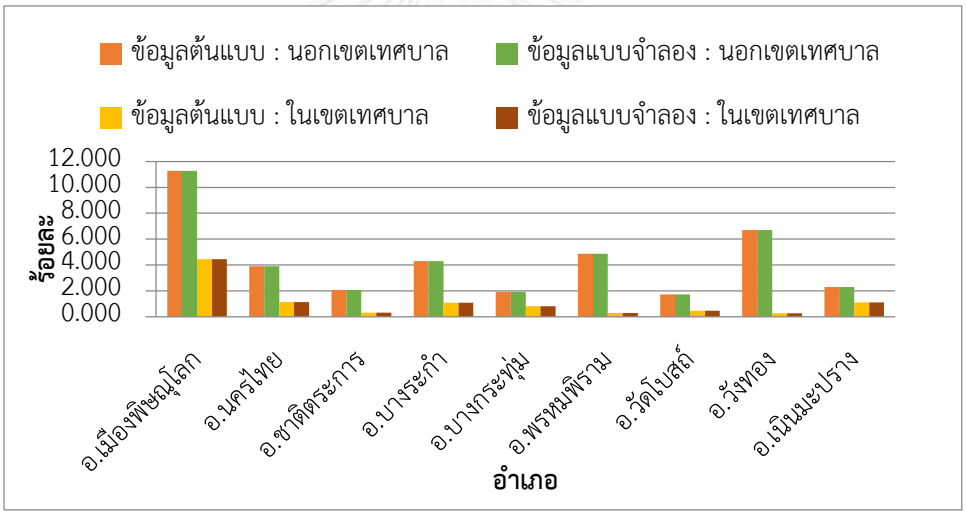
คุณลักษณะของข้อมูล	ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์	ค่าต่ำสุดของความคลาดเคลื่อนสัมบูรณ์	ค่าสูงสุดของความคลาดเคลื่อนสัมบูรณ์
จำนวนครัวเรือน	0.000	0.000	0.000
เพศและอายุของหัวหน้าครัวเรือน	1.511	0.020	4.020
อายุห่างระหว่างคู่สมรส	1.376	0.000	10.770
การอยู่อาศัยของเด็กอายุ 0-17 ปี	0.000	0.000	0.000
อายุห่างระหว่างบุตรแต่ละคน	0.000	0.000	0.000
ขนาดครัวเรือน	0.151	0.000	3.421

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สำหรับตัวอย่างของกราฟการเปรียบเทียบข้อมูลต้นแบบและข้อมูลแบบจำลองประชากรสังเคราะห์ที่ถูกสร้างขึ้นในระดับบุคคลของคุณลักษณะของข้อมูลเขตเทศบาลและสถานภาพสมรส จะแสดงในภาพที่ 34 และภาพที่ 35 ตามลำดับ ซึ่งจากภาพที่ 34 จะเห็นว่าคุณลักษณะของข้อมูลเขตเทศบาลทั้งเพศชายและเพศหญิงจะไม่มีความคลาดเคลื่อนเลย นั่นคือ ประชากรสังเคราะห์ที่สร้างขึ้นมีความเหมือนกับข้อมูลต้นแบบ แต่ภาพที่ 35 จะเห็นว่า คุณลักษณะของข้อมูลสถานภาพสมรสยังมีความคลาดเคลื่อนอยู่บ้างในสถานภาพสมรสที่เป็นโสดของเพศหญิง โดยจะมีความคลาดเคลื่อนอยู่ที่ช่วงอายุ 85 ปี - มากกว่า 100 ปี และสถานภาพสมรสที่เป็นสมรสของเพศหญิง ก็จะมีความคลาดเคลื่อนอยู่ที่ช่วงอายุ 85 ปี - มากกว่า 100 ปีเช่นเดียวกันซึ่งเป็นผลมาจากข้อมูลที่มีการเก็บข้อมูลหลายระดับความละเอียดและหลายหน่วยงาน ทำให้ข้อมูลของคุณลักษณะอายุ และคุณลักษณะสถานภาพสมรสมีความไม่สอดคล้องกัน



(ก) เพศชาย

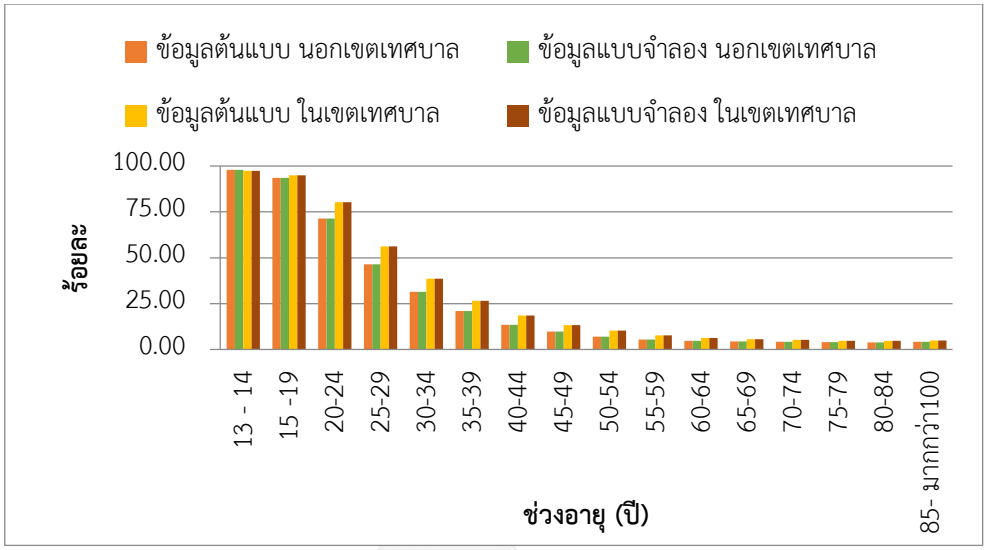


(ข) เพศหญิง

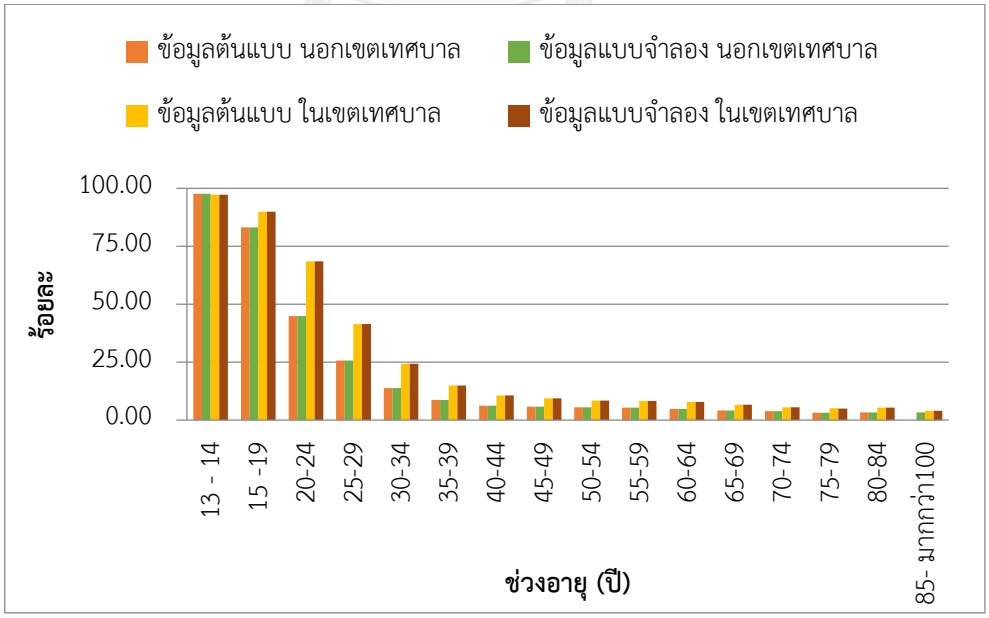
ภาพที่ 34 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลเขตเทศบาล

สำหรับตัวอย่างของกราฟการเปรียบเทียบข้อมูลต้นแบบและข้อมูลแบบจำลองประชากรสังเคราะห์ที่ถูกสร้างขึ้นในระดับครัวเรือนของคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุและอายุเมื่อมีบุตรคนแรก จะแสดงในภาพที่ 36 และภาพที่ 37 ตามลำดับ ซึ่งจากภาพที่ 36 จะเห็นว่าคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุจะมีความคลาดเคลื่อนตั้งแต่ช่วงอายุ 30 ปีขึ้นไปทั้งนอกเขตเทศบาลและในเขตเทศบาล เนื่องจากข้อมูลนี้ถูกเปิดเผยจากคนละหน่วยงานกับข้อมูลจำนวนคนในระดับบุคคล ซึ่งข้อมูลที่ได้มีความไม่สอดคล้องกันตั้งแต่จำนวนคนใน

จังหวัด ทำให้จำนวนคนในแต่ละช่วงอายุที่จะถูกเลือกให้เป็นหัวหน้าครัวเรือนมีไม่พอในการเลือกให้เป็นหัวหน้าครัวเรือน ส่งผลให้การเลือกหัวหน้าครัวเรือนตามช่วงอายุและเพศมีความคลาดเคลื่อนไปด้วย สำหรับภาพที่ 37 จะแสดงให้เห็นว่า คุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก จะไม่มีความคลาดเคลื่อนเลย นั่นคือ ประชากรสังเคราะห์ที่สร้างขึ้นมีความเหมือนกับข้อมูลต้นแบบทุกประการ โดยกราฟการเปรียบเทียบข้อมูลต้นแบบและข้อมูลแบบจำลองประชากรสังเคราะห์ที่ถูกสร้างขึ้นในแต่ละคุณลักษณะจะแสดงในภาคผนวก ข ดังภาพที่ ข.1 – ข.15

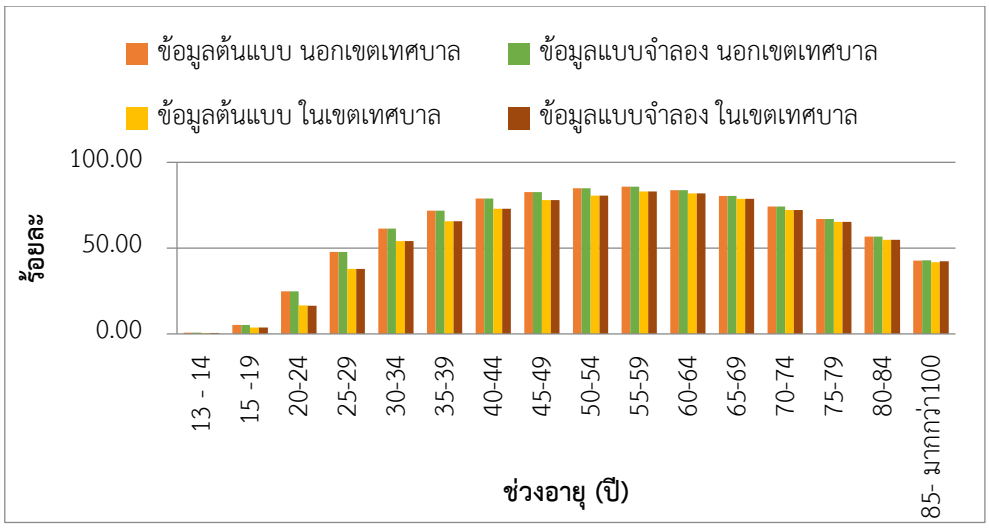


(ก) เพศชาย – โสด

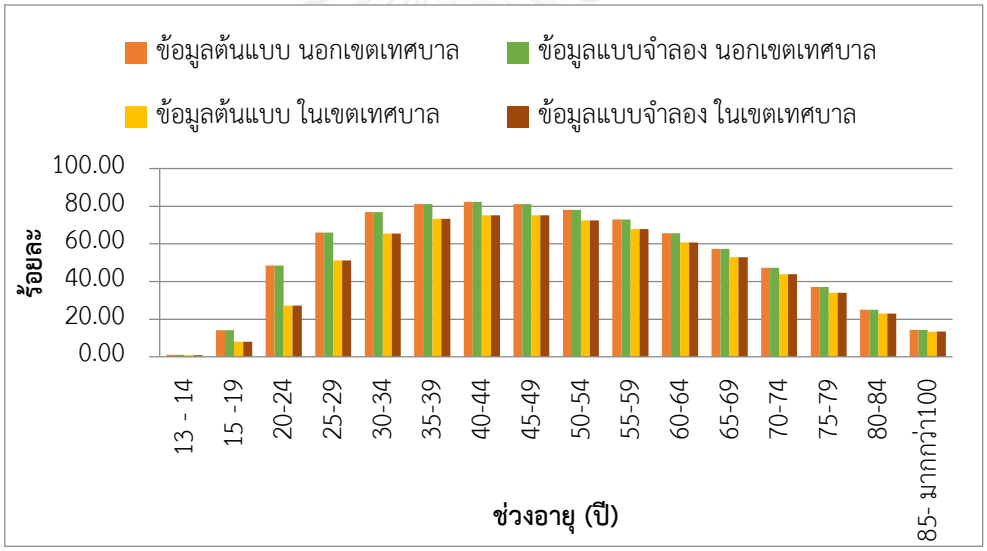


(ข) เพศหญิง – โสด

ภาพที่ 35 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรส

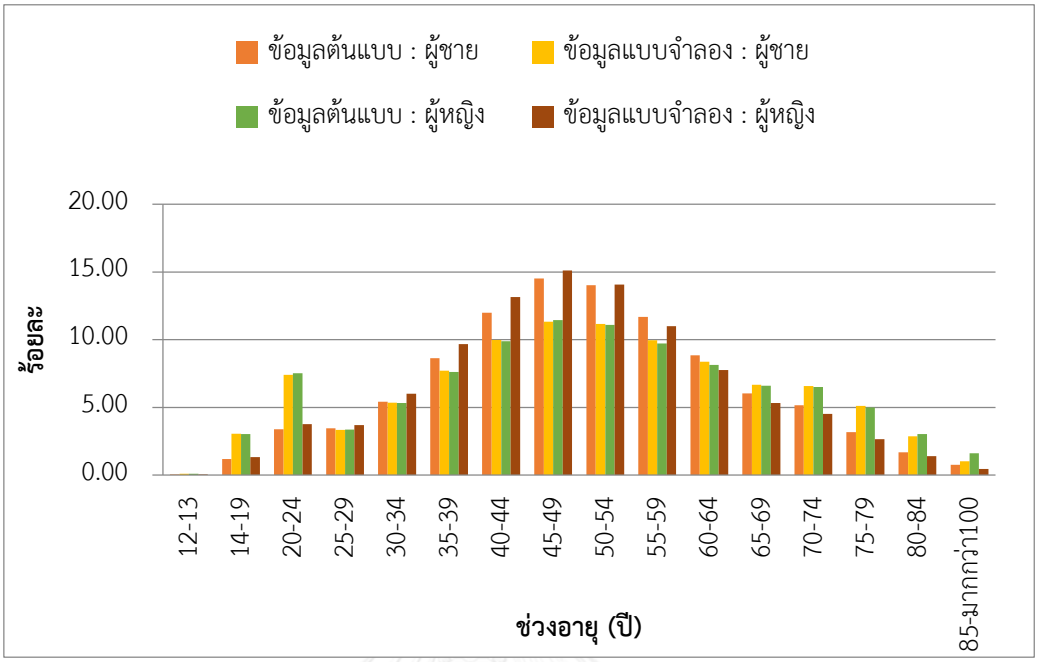


(ค) เพศชาย – สมรส

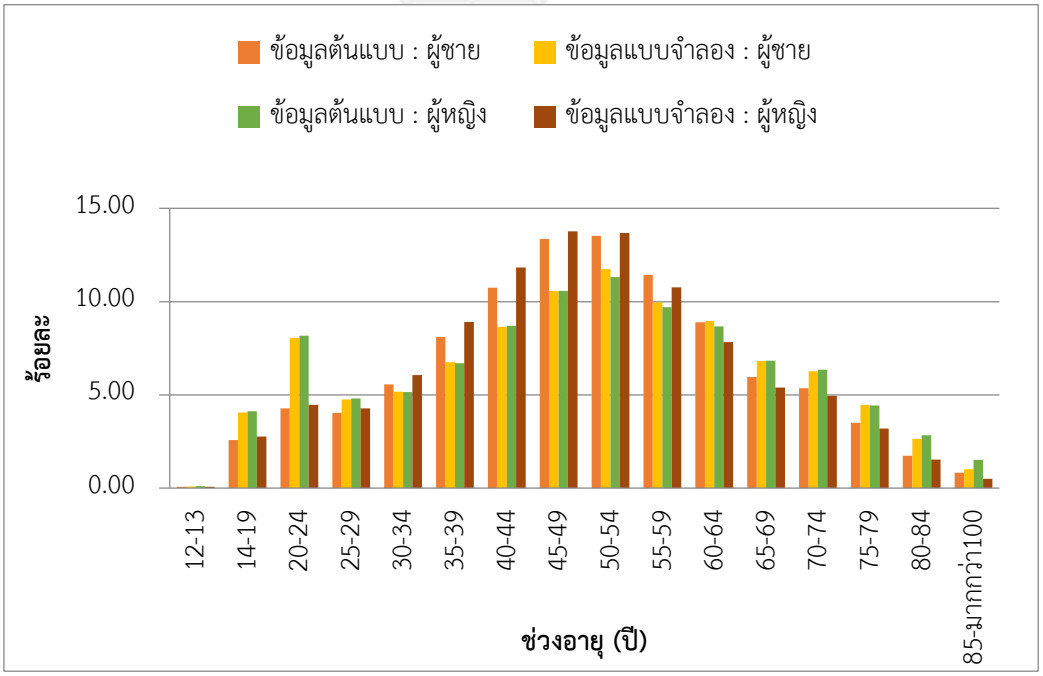


(ง) เพศหญิง – สมรส

ภาพที่ 35 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรส (ต่อ)

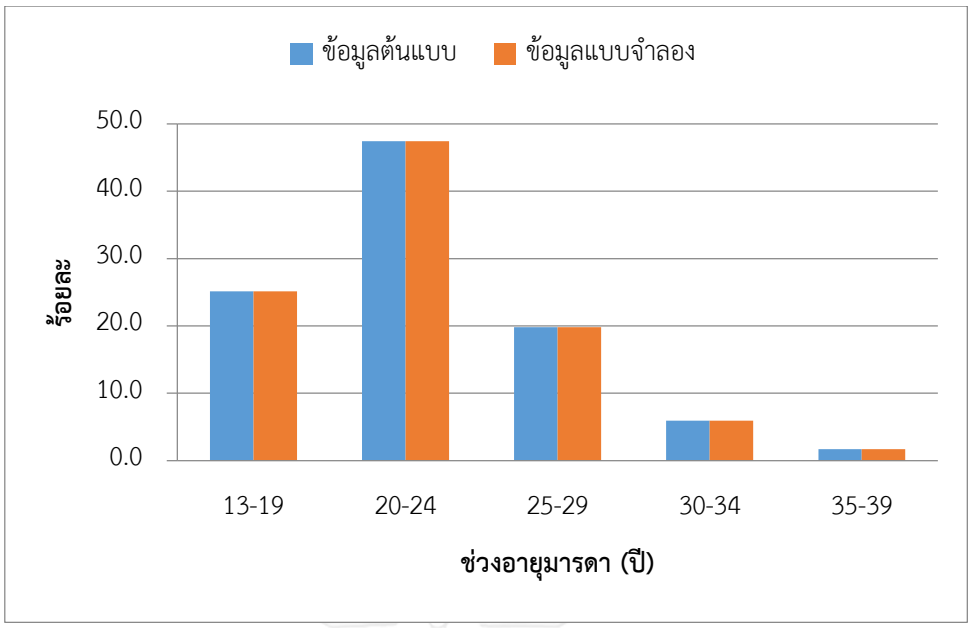


(ก) นอกเขตเทศบาล



(ข) ในเขตเทศบาล

ภาพที่ 36 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุ



ภาพที่ 37 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก

4.3.2 ผลของการประเมินโดยใช้ข้อมูลต้นแบบเป็นข้อมูลประชากรสังเคราะห์

การประเมินวิธีนี้จะเป็นการประเมินความถูกต้องและความน่าเชื่อถือของกระบวนการสร้างประชากรสังเคราะห์ที่ได้ออกแบบ โดยทดลองใช้ข้อมูลต้นแบบคือ ข้อมูลแบบจำลองประชากรสังเคราะห์ชุดที่ 1 มาสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือน จึงได้ข้อมูลแบบจำลองประชากรสังเคราะห์ชุดที่ 2 เพื่อประเมินความถูกต้องของกระบวนการโดยการหาค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ ซึ่งได้ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะที่ใช้ในการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนแสดงในตารางที่ 34 และตารางที่ 35 ตามลำดับ โดยค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนคือ 0.0003% และ 0.3698% ตามลำดับ ซึ่งมีความคลาดเคลื่อนน้อยลงเมื่อเทียบกับการประเมินในวิธีแรก จากตารางที่ 34 จะได้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการสร้างประชากรสังเคราะห์ในระดับบุคคล คือ 0.0003% ซึ่งพบว่า ประชากรสังเคราะห์ในระดับบุคคลมีความคลาดเคลื่อนจากข้อมูลต้นแบบเพียงคุณลักษณะเดียว คือ สถานภาพสมรส ซึ่งคุณลักษณะอื่นในระดับบุคคล ได้แก่ จำนวนคน, เพศ, อำเภอที่อยู่อาศัย, เขตการปกครองที่อยู่อาศัย, อายุ, การศึกษา , จำนวนนักเรียนแยกรายสถานศึกษา, จำนวนบุตร และอายุเมื่อมีบุตรคนแรก สามารถนำมาสร้างประชากรสังเคราะห์ในระดับบุคคลได้อย่างแม่นยำ ไม่มีความคลาดเคลื่อนเลย แสดงว่าประชากรในคุณลักษณะดังกล่าวมีความเหมือนประชากรจริงมากที่สุด ซึ่งจากค่าเฉลี่ยของความคลาดเคลื่อน

สัมบูรณ์ แสดงให้เห็นว่า ประชากรสังเคราะห์ในระดับบุคคลจะมีความคลาดเคลื่อนไป 3 คน จากการสร้างประชากรสังเคราะห์ทั้งหมด 1,000,000 คน

ตารางที่ 34 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับบุคคล

คุณลักษณะของข้อมูล	**ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์**	**ค่าต่ำสุดของความคลาดเคลื่อนสัมบูรณ์**	**ค่าสูงสุดของความคลาดเคลื่อนสัมบูรณ์**
จำนวนคน	0.000	0.000	0.000
เพศ	0.000	0.000	0.000
อำเภอที่อยู่อาศัย	0.000	0.000	0.000
เขตการปกครองที่อยู่อาศัย	0.000	0.000	0.000
อายุ	0.000	0.000	0.000
การศึกษา	0.000	0.000	0.000
จำนวนนักเรียนแยกรายสถานศึกษา	0.000	0.000	0.000
สถานภาพสมรส	0.003	0.000	0.631
จำนวนบุตร	0.000	0.000	0.000
อายุเมื่อมีบุตรคนแรก	0.000	0.000	0.000

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

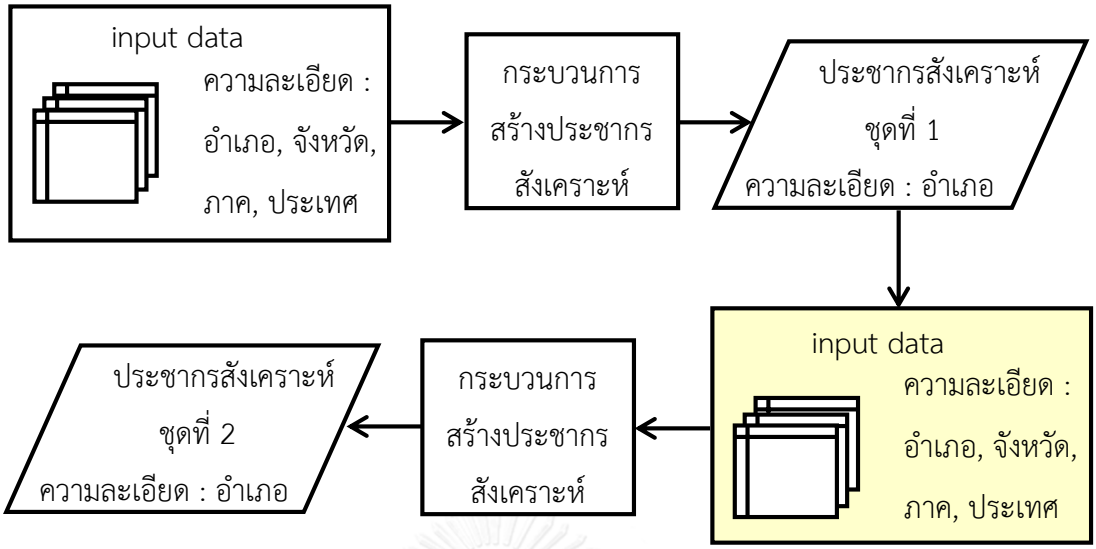
จากตารางที่ 35 จะได้ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการสร้างประชากรสังเคราะห์ในระดับครัวเรือนคือ 0.3698% โดยความคลาดเคลื่อนเหลือเพียง 2 คุณลักษณะเท่านั้น คือ เพศและอายุของหัวหน้าครัวเรือน, อายุห่างระหว่างคู่สมรส ซึ่งคุณลักษณะอื่นในระดับครัวเรือน ได้แก่ จำนวนครัวเรือน, การอยู่อาศัยของเด็กอายุ 0-17 ปี, อายุห่างระหว่างบุตรแต่ละคน และขนาดครัวเรือนสามารถนำมาสร้างประชากรสังเคราะห์ในระดับครัวเรือนได้อย่างแม่นยำ ไม่มีความคลาดเคลื่อนเลย แสดงว่าจำนวนครัวเรือน, การเลือกผู้อยู่อาศัยที่เป็นเด็กเข้าอาศัยในครัวเรือนและขนาดครัวเรือนมีความเหมือนกับสังคมจริงมากที่สุด ซึ่งจากค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ แสดงให้เห็นว่าประชากรสังเคราะห์ในระดับครัวเรือนจะมีความคลาดเคลื่อนไป 4 ครัวเรือน จากการสร้างประชากรสังเคราะห์ระดับครัวเรือนทั้งหมด 1,000 ครัวเรือน

ตารางที่ 35 ความคลาดเคลื่อนสัมบูรณ์ของแต่ละคุณลักษณะในระดับครัวเรือน

คุณลักษณะของข้อมูล	ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์	ค่าต่ำสุดของความคลาดเคลื่อนสัมบูรณ์	ค่าสูงสุดของความคลาดเคลื่อนสัมบูรณ์
จำนวนครัวเรือน	0.000	0.000	0.000
เพศและอายุของหัวหน้าครัวเรือน	1.319	0.030	4.300
อายุห่างระหว่างคู่สมรส	0.900	0.000	2.640
การอยู่อาศัยของเด็กอายุ 0-17 ปี	0.000	0.000	0.000
อายุห่างระหว่างบุตรแต่ละคน	0.000	0.000	0.000
ขนาดครัวเรือน	0.000	0.000	0.000

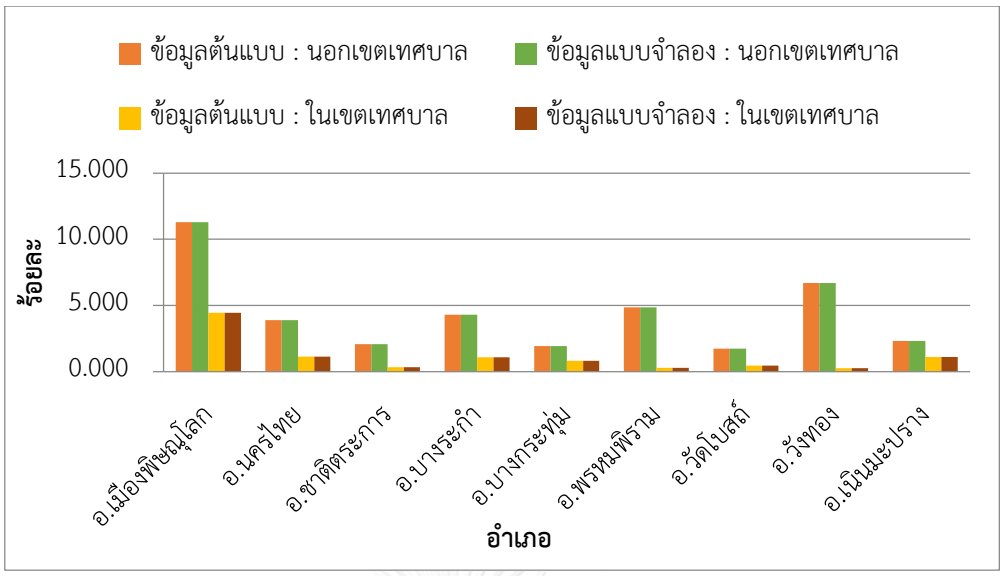
สำหรับเหตุผลที่การสร้างประชากรสังเคราะห์ในครั้งนี้ยังมีความคลาดเคลื่อนอยู่ในบางคุณลักษณะนั้นสามารถอธิบายได้ดังภาพที่ 38 คือการนำประชากรสังเคราะห์ที่ถูกสร้างขึ้นในครั้งแรกซึ่งมีความละเอียดของการระบุข้อมูลประชากรอยู่ที่ระดับอำเภอมาแปลงกลับให้เป็นข้อมูลในรูปแบบตารางเพื่อนำไปใช้เป็นข้อมูลต้นแบบสำหรับนำไปสร้างประชากรสังเคราะห์ในครั้งที่ 2 จะเป็นตารางที่มีความละเอียดที่ต่ำกว่า นั่นคือความละเอียดอยู่ที่ระดับจังหวัด ระดับภาคและระดับประเทศ ทำให้มีการเปลี่ยนแปลงความละเอียดของข้อมูลส่งผลให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นครั้งที่ 2 ก็ยังคงถูกสร้างจากข้อมูลของตารางที่มีความละเอียดต่ำ อีกทั้งยังมีความคลาดเคลื่อนจากการสุ่มของโปรแกรมด้วย จึงทำให้ยังมีความคลาดเคลื่อนอยู่ในบางคุณลักษณะ

สำหรับตัวอย่างของกราฟการเปรียบเทียบข้อมูลต้นแบบและข้อมูลแบบจำลองประชากรสังเคราะห์ที่ถูกสร้างขึ้นในระดับบุคคลของคุณลักษณะของข้อมูลเขตเทศบาลและสถานภาพสมรส จะแสดงในภาพที่ 39 และภาพที่ 40 ตามลำดับ ซึ่งจากภาพที่ 39 จะเห็นว่าคุณลักษณะของข้อมูลเขตเทศบาลทั้งเพศชายและเพศหญิงจะไม่มีความคลาดเคลื่อนเลย นั่นคือ ประชากรสังเคราะห์ที่สร้างขึ้นมีความเหมือนกับข้อมูลต้นแบบ แต่ภาพที่ 40 จะเห็นว่า คุณลักษณะของข้อมูลสถานภาพสมรสยังมีความคลาดเคลื่อนอยู่เพียงเล็กน้อยในสถานภาพสมรสที่เป็นโสดและสมรสของเพศหญิง โดยจะมีความคลาดเคลื่อนอยู่ที่ช่วงอายุ 85 ปี - มากกว่า 100 ปี

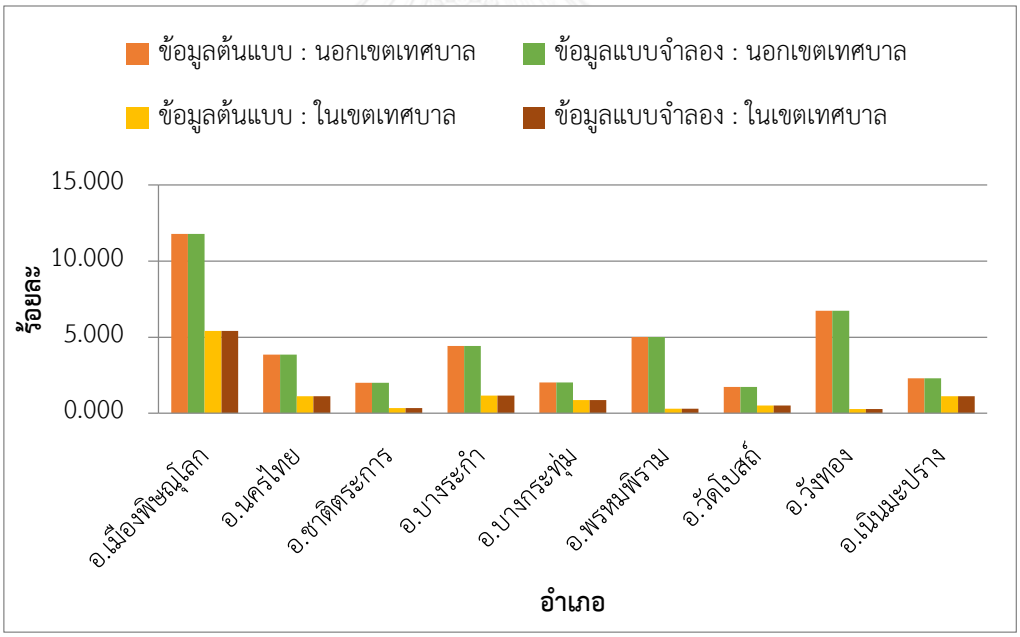


ภาพที่ 38 ความคลาดเคลื่อนที่เกิดขึ้นในการสร้างประชากรครั้งที่ *2*

สำหรับตัวอย่างของกราฟการเปรียบเทียบข้อมูลต้นแบบและข้อมูลแบบจำลองประชากรสังเคราะห์ที่ถูกสร้างขึ้นในระดับครัวเรือนของคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุและอายุเมื่อมีบุตรคนแรก จะแสดงในภาพที่ 41 และภาพที่ 42 ตามลำดับ ซึ่งจากภาพที่ 41 จะเห็นว่าคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุจะมีความคลาดเคลื่อนตั้งแต่ช่วงอายุ 30 ปีขึ้นไปทั้งนอกเขตเทศบาลและในเขตเทศบาล ซึ่งเป็นผลมาจากการแปลงความละเอียดของการระบุข้อมูลประชากรอยู่ที่ระดับอำเภอมาแปลงกลับให้เป็นข้อมูลในรูปแบบตารางเพื่อนำไปใช้เป็นข้อมูลต้นแบบสำหรับนำไปสร้างประชากรสังเคราะห์ในครั้งที่ 2 และยังมาจากความคลาดเคลื่อนของการสุ่มเลือกของโปรแกรมด้วย แต่ความคลาดเคลื่อนในการประเมินวิธีนี้จะน้อยกว่าการประเมินวิธีแรก สำหรับภาพที่ 42 แสดงให้เห็นว่า คุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก จะไม่มีความคลาดเคลื่อนเลย นั่นคือ ประชากรสังเคราะห์ที่สร้างขึ้นมีความเหมือนกับข้อมูลต้นแบบทุกประการ

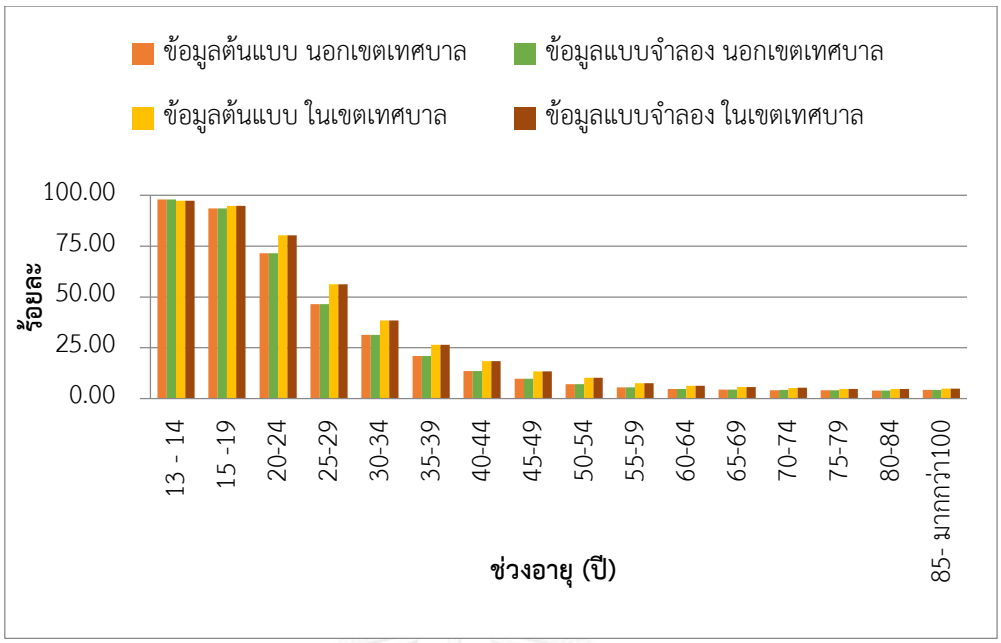


(ก) เพศชาย

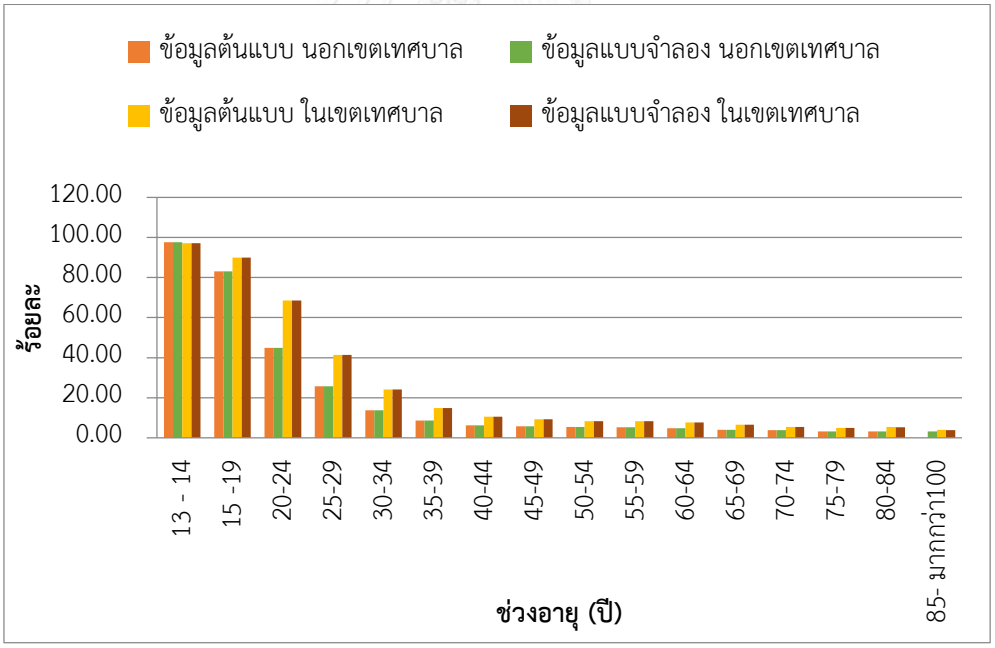


(ข) เพศหญิง

ภาพที่ 39 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลเขตเทศบาล

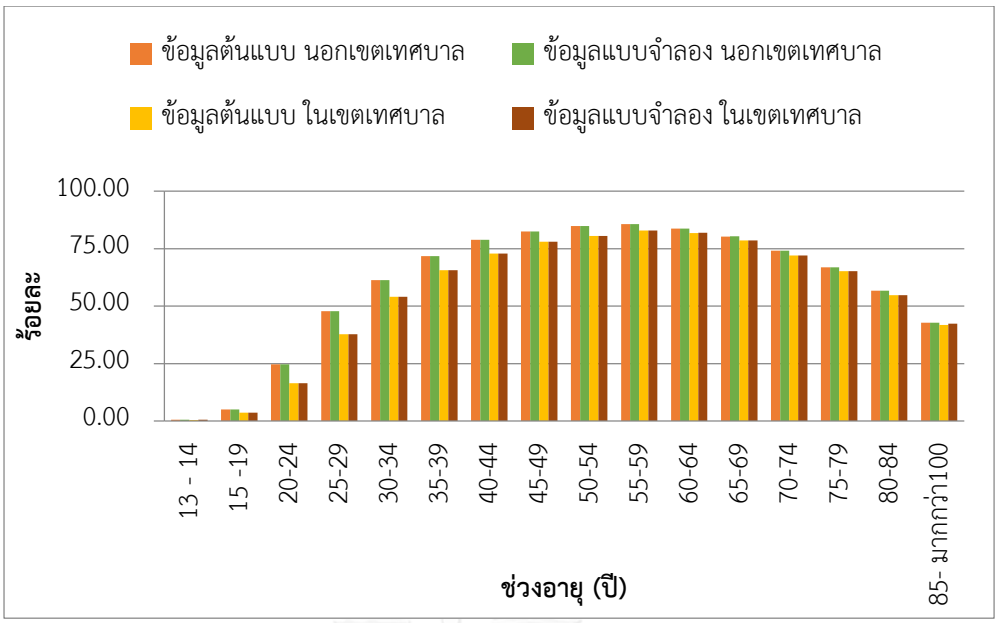


(ก) เพศชาย – โสด

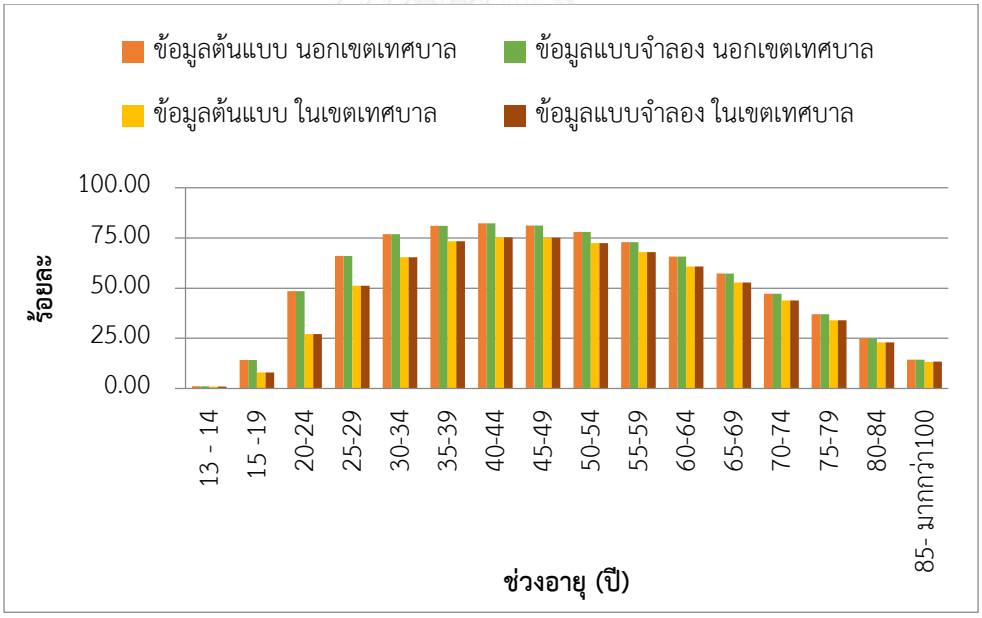


(ข) เพศหญิง – โสด

ภาพที่ 40 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรส

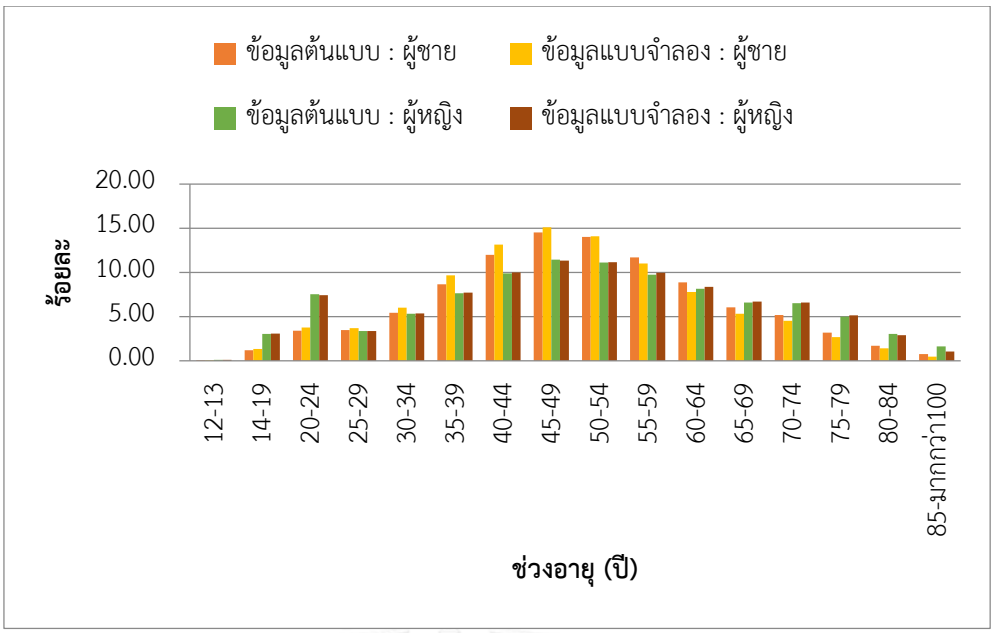


(ค) เพศชาย – สมรส

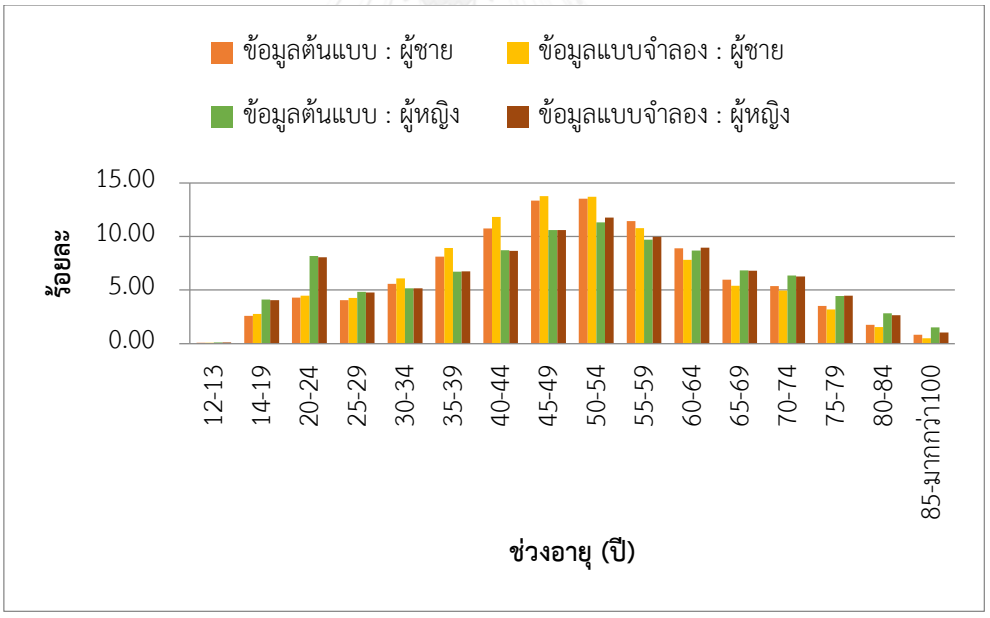


(ง) เพศหญิง – สมรส

ภาพที่ 40 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรส (ต่อ)

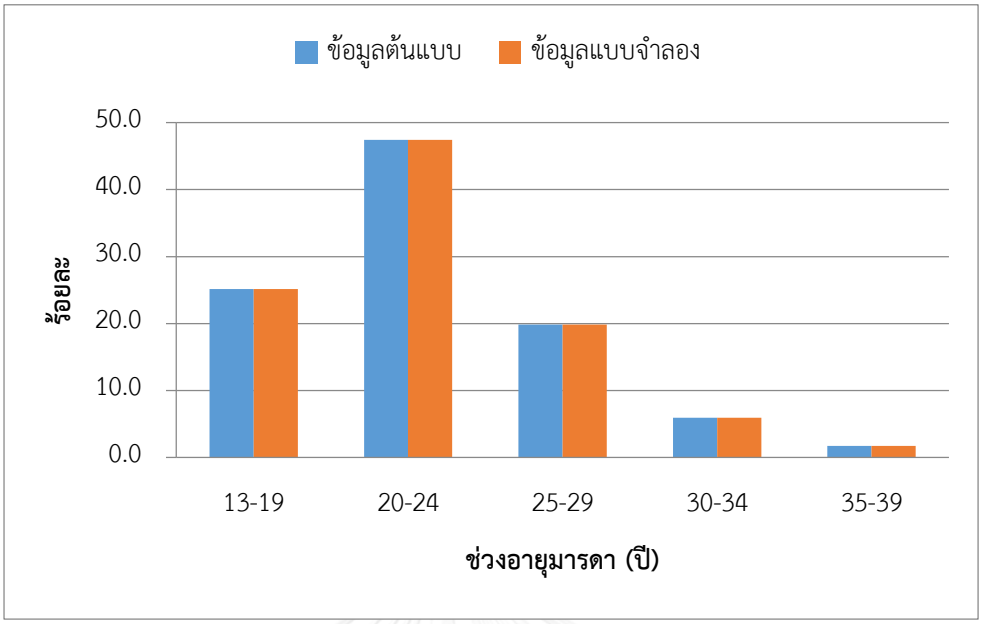


(ก) นอกเขตเทศบาล



(ข) ในเขตเทศบาล

ภาพที่ 41 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุ



ภาพที่ 42 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก

4.3.3 ผลของการประเมินโดยใช้ตารางที่ถูกซ่อน

การประเมินโดยใช้ตารางที่ถูกซ่อนในงานวิจัยนี้จะประเมินประชากรสังเคราะห์ที่ถูกสร้างขึ้นในแต่ละชุดว่ามีความใกล้เคียงกันมากน้อยเพียงใด โดยจะเลือกใช้ตารางของคุณลักษณะข้อมูลสถานภาพสมรสของเพศชายและเพศหญิง ซึ่งข้อมูลต้นแบบของคุณลักษณะนี้เป็นการแบ่งตามช่วงอายุ ช่วงละ 5 ปี นำมาเปลี่ยนแปลงขอบเขตของข้อมูลใหม่เป็นทุก 1 ปี เพื่อเป็นตารางที่ถูกซ่อนของเพศชายและเพศหญิง ใช้ในการเปรียบเทียบความคลาดเคลื่อนของประชากรสังเคราะห์ที่ถูกสร้างขึ้นระหว่างชุดที่ 1 และชุดที่ 2 โดยความคลาดเคลื่อนสัมบูรณ์ของตารางที่ถูกซ่อนของคุณลักษณะสถานภาพสมรสของเพศชายและเพศหญิงแสดงในตารางที่ 36 และตารางที่ 37 ตามลำดับ มีค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของคุณลักษณะสถานภาพสมรสของเพศชายและเพศหญิงคือ 0.126% และ 0.102% ตามลำดับ

สำหรับตัวอย่างของกราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศชายและเพศหญิง จะแสดงในภาพที่ 43 และภาพที่ 44 ตามลำดับ ซึ่งจะเห็นว่าประชากรสังเคราะห์ทั้ง 2 ชุด ในคุณลักษณะของสถานภาพสมรสที่เป็นโสดและสมรสทั้งเพศชายและเพศหญิงจะมีความคลาดเคลื่อนเล็กน้อยเพียงเล็กน้อย ซึ่งแสดงว่าประชากร

สังเคราะห์ที่ถูกสร้างขึ้นทั้ง 2 ชุดมีความใกล้เคียงกันมาก โดยกราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศชายและเพศหญิงแสดงในภาพที่ ข.16 – ข.17

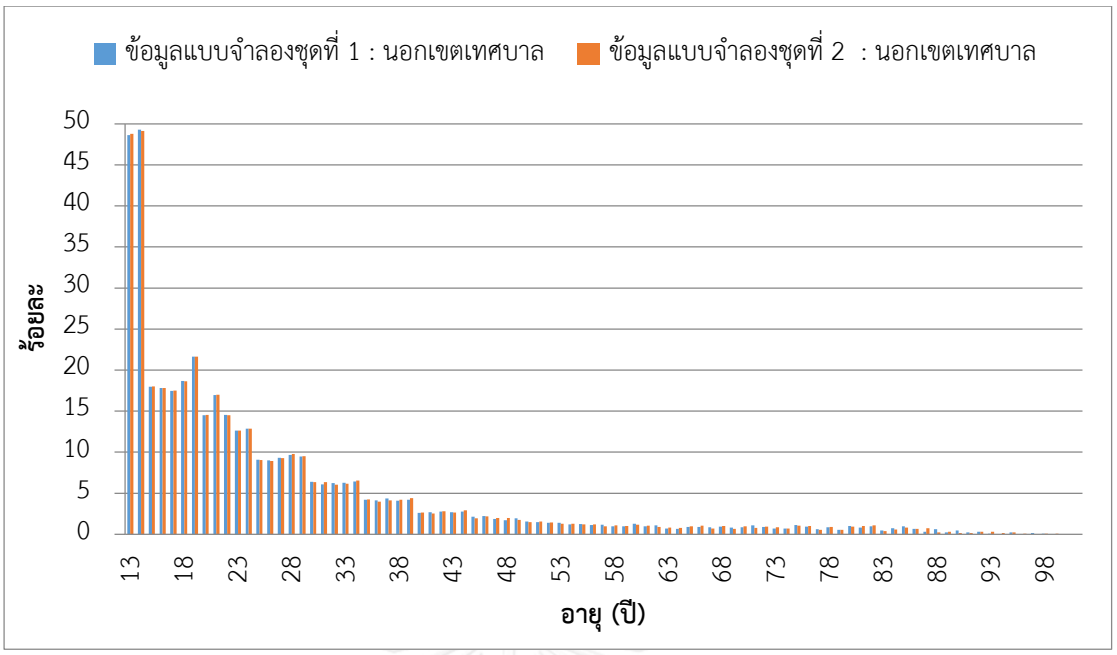
ตารางที่ 36 ความคลาดเคลื่อนสัมบูรณ์ของตารางที่ถูกซ่อนของคุณลักษณะสถานภาพสมรสของเพศชาย

สถานภาพสมรส	ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์	ค่าต่ำสุดของความคลาดเคลื่อนสัมบูรณ์	ค่าสูงสุดของความคลาดเคลื่อนสัมบูรณ์
โสด	0.152	0.000	1.099
สมรส	0.291	0.000	0.964
หม้าย	0.247	0.000	1.491
หย่า	0.048	0.000	0.260
แยกกันอยู่	0.065	0.000	0.357
เคยสมรสแต่ไม่ทราบสถานภาพสมรส	0.039	0.000	0.379
ไม่ทราบสถานภาพสมรส	0.039	0.000	0.235

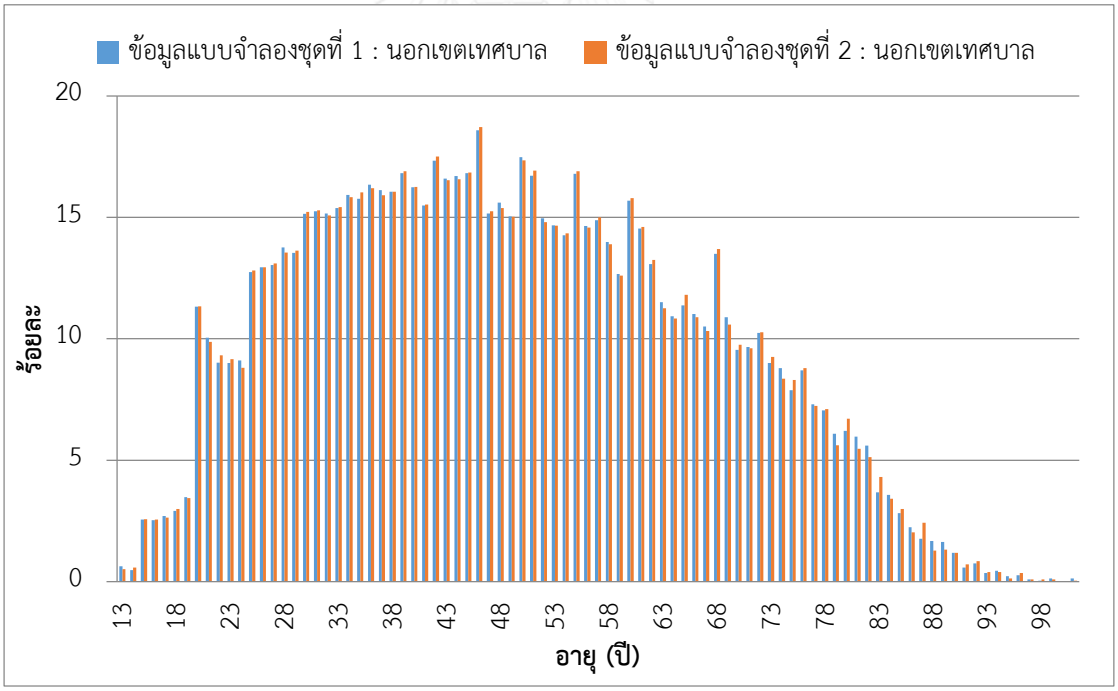
ตารางที่ 37 ความคลาดเคลื่อนสัมบูรณ์ของตารางที่ถูกซ่อนของคุณลักษณะสถานภาพสมรสของเพศหญิง

สถานภาพสมรส	ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์	ค่าต่ำสุดของความคลาดเคลื่อนสัมบูรณ์	ค่าสูงสุดของความคลาดเคลื่อนสัมบูรณ์
โสด	0.136	0.000	0.691
สมรส	0.208	0.000	2.479
หม้าย	0.173	0.000	3.099
หย่า	0.060	0.000	0.362
แยกกันอยู่	0.071	0.000	0.349
เคยสมรสแต่ไม่ทราบสถานภาพสมรส	0.047	0.000	0.209
ไม่ทราบสถานภาพสมรส	0.021	0.000	0.362

Chulalongkorn University

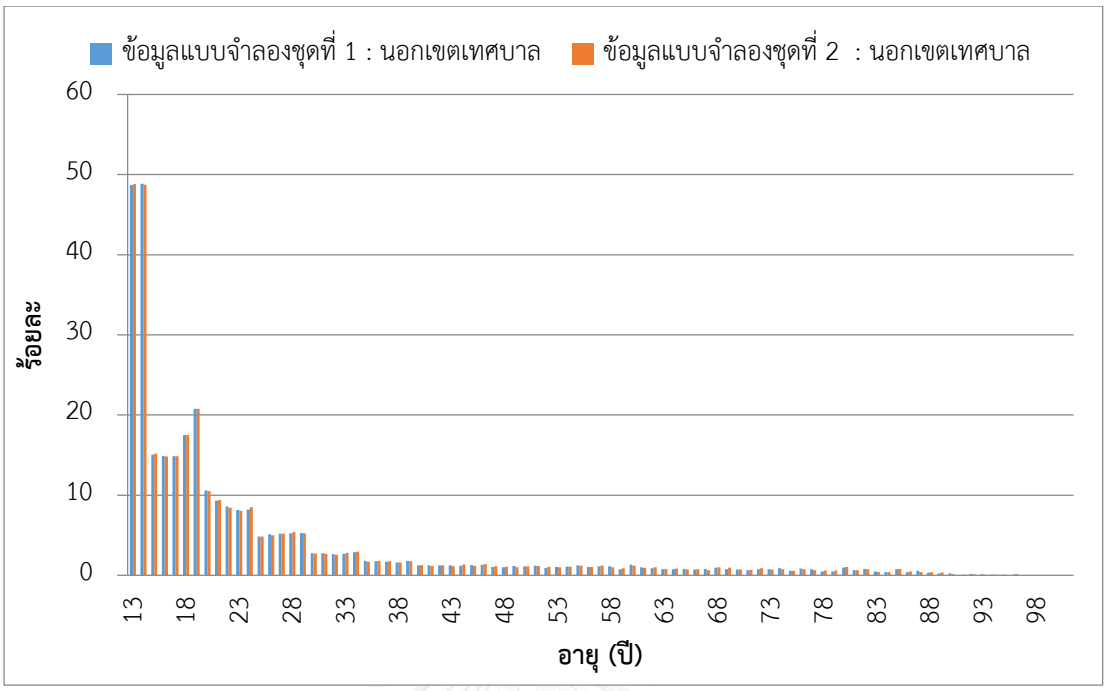


(ก) โสด – นอกเขตเทศบาล

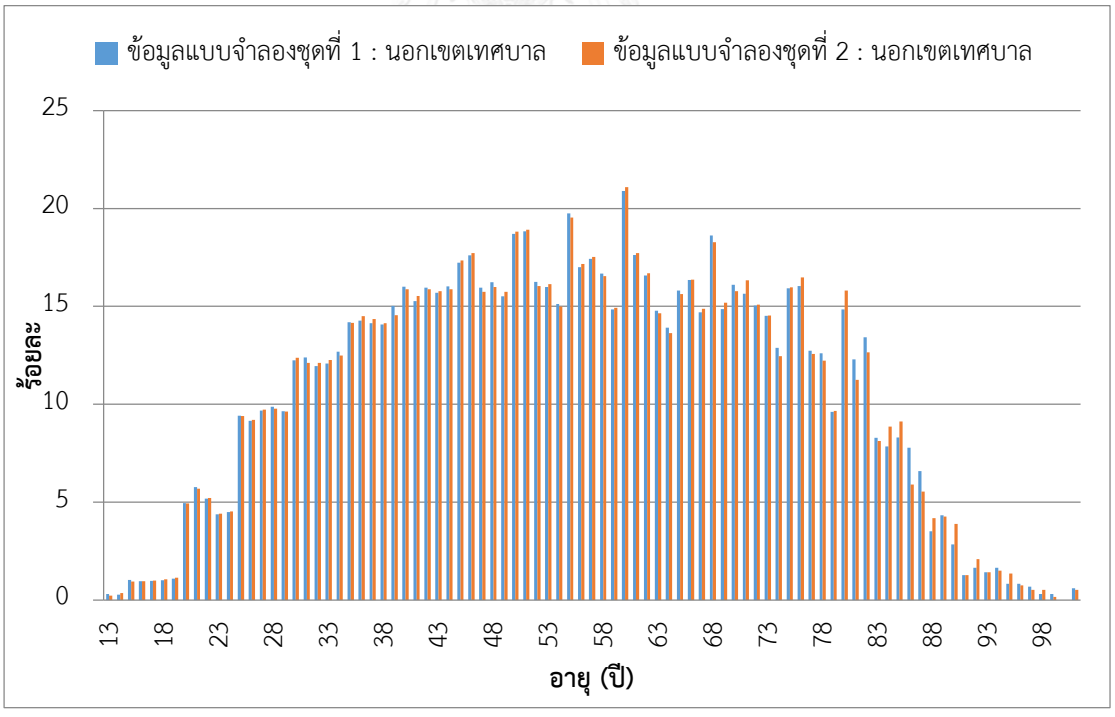


(ข) สมรส – นอกเขตเทศบาล

ภาพที่ 43 กราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศชาย



(ก) โสด – นอกเขตเทศบาล



(ข) สมรส – นอกเขตเทศบาล

ภาพที่ 44 กราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง *2* ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศหญิง

บทที่ 5
บทสรุปและแนวทางในการพัฒนาต่อ

5.1 บทสรุป

งานวิจัยนี้ได้นำเสนอกระบวนการสร้างประชากรสังเคราะห์ทั้งในระดับบุคคลและระดับครัวเรือนที่เหมาะสมกับข้อมูลของประเทศไทยซึ่งมีการเปิดเผยข้อมูลเพียงข้อมูลที่ได้จากสำมะโนประชากร อีกทั้งยังใช้คุณลักษณะต่างๆ ของข้อมูลประชากรอย่างหลากหลาย เหมาะสำหรับประเทศที่มีรายละเอียดของครัวเรือนที่ซับซ้อน เช่น จำนวนบุตรที่มีตั้งแต่ 0 คนถึงมากกว่า 10 คน, ขนาดของครัวเรือนที่มีตั้งแต่ 1 คนถึงมากกว่า 10 คน เป็นต้น ส่งผลให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นมีการระบุตัวตนของบุคคลและครัวเรือนที่มีความละเอียดใกล้เคียงกับประชากรจริงมากที่สุด แต่เนื่องจากการใช้คุณลักษณะของข้อมูลที่มากทำให้การออกแบบกระบวนการสร้างประชากรสังเคราะห์นั้นซับซ้อนไปด้วย นอกจากนี้ยังมีข้อจำกัดของการสร้างประชากรสังเคราะห์ซึ่งขึ้นอยู่กับข้อมูลที่จะนำมาใช้เป็นต้นแบบประชากรด้วย โดยจากการศึกษาและเลือกข้อมูลที่จะนำมาใช้เป็นข้อมูลต้นแบบประชากร พบว่ามีข้อจำกัดดังนี้

- ข้อมูลประชากรมีการเปิดเผยเพียงข้อมูลมาจากการสำมะโนประชากร ซึ่งอยู่ในรูปแบบ aggregate data ส่งผลให้ไม่สามารถใช้กระบวนการ Combinatorial Optimization ได้
- ข้อมูลที่เปิดเผยถูกเก็บมาจากหลายหน่วยงานซึ่งมีความละเอียดในการเก็บข้อมูลที่แตกต่างกัน ทำให้ข้อมูลเหล่านั้นไม่สอดคล้องกัน
- ข้อมูลระดับครัวเรือนมีการเปิดเผยเพียงบางคุณลักษณะของข้อมูล ทำให้การเชื่อมโยงคุณลักษณะสำหรับการสร้างประชากรสังเคราะห์ระดับครัวเรือนเป็นไปด้วยความยากลำบาก

ด้วยเหตุผลทางข้อจำกัดดังกล่าว ผู้วิจัยจึงออกแบบกระบวนการสร้างประชากรสังเคราะห์เพื่อแก้ไขข้อจำกัดดังกล่าว ซึ่งกระบวนการสร้างประชากรสังเคราะห์ที่ออกแบบนั้นเป็นส่วนหนึ่งของกระบวนการ Synthetic Reconstruction ซึ่งใช้เพียงข้อมูลที่อยู่ในรูปแบบ aggregate data และได้ใช้ข้อมูลต้นแบบที่มีความละเอียดของข้อมูลสูงที่สุด พร้อมทั้งได้ปรับปรุงข้อมูลบางส่วนเพื่อให้มีความสอดคล้องกับข้อมูลในคุณลักษณะอื่นๆ เช่น ข้อมูลขนาดของครัวเรือน เป็นต้น และได้ออกแบบกระบวนการให้รองรับกับครัวเรือนที่มีความหลากหลายของประเทศไทย

จากนั้นผู้วิจัยได้ประยุกต์ใช้กระบวนการสร้างประชากรสังเคราะห์กับข้อมูลของจังหวัดพิษณุโลกที่นำเสนอในหัวข้อที่ 4.1 และนำประชากรสังเคราะห์ที่ได้มาประเมินความถูกต้องของ

กระบวนการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนเมื่อเทียบกับข้อมูลต้นแบบที่ได้จากข้อมูลสถิติเพื่อที่จะประเมินของประชากรที่สร้างขึ้นกับประชากรตามข้อมูลสถิติดังที่แสดงในตารางที่ 32 และตารางที่ 33 พบว่า ประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนมีค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ 0.003% และ 0.506% ตามลำดับ ซึ่งความคลาดเคลื่อนของการประเมินวิธีนี้มาจากการเก็บข้อมูลที่มีความละเอียดที่แตกต่างกันและยังถูกเก็บมาจากคนละหน่วยงาน ทำให้ข้อมูลในแต่ละคุณลักษณะมีความไม่สอดคล้องกัน ส่งผลให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นมีความคลาดเคลื่อนด้วย

นอกจากนี้ผู้วิจัยได้นำประชากรสังเคราะห์ของจังหวัดพิษณุโลกที่ถูกสร้างในครั้งแรกมาแปลงกลับให้เป็นข้อมูลในรูปแบบตารางเพื่อเป็นข้อมูลต้นแบบนำไปสร้างประชากรสังเคราะห์อีกรอบ โดยจะถือว่าข้อมูลต้นแบบนี้เป็นข้อมูลที่ได้มาจากการสำรวจอย่างสมบูรณ์ เพื่อประเมินความน่าเชื่อถือของกระบวนการสร้างประชากรสังเคราะห์ที่ออกแบบ จากผลการทดลองที่แสดงในตารางที่ 34 และตารางที่ 35 พบว่า ประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนมีความคลาดเคลื่อนน้อยลงเมื่อเทียบกับการประเมินในวิธีแรก โดยประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนมีค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ 0.0003% และ 0.3698% ตามลำดับ โดยเหตุผลที่การสร้างประชากรสังเคราะห์ในครั้งนี้ยังมีความคลาดเคลื่อนอยู่ในบางคุณลักษณะนั้นมาจากการนำประชากรสังเคราะห์ที่ถูกสร้างขึ้นในครั้งแรกซึ่งมีความละเอียดของการระบุข้อมูลประชากรอยู่ที่ระดับอำเภอมาแปลงกลับให้เป็นข้อมูลในรูปแบบตารางเพื่อนำไปใช้เป็นข้อมูลต้นแบบสำหรับนำไปสร้างประชากรสังเคราะห์ในครั้งที่ 2 จะเป็นตารางที่มีความละเอียดที่ต่ำกว่า คือความละเอียดอยู่ที่ระดับจังหวัด ระดับภาคและระดับประเทศ ทำให้มีการเปลี่ยนแปลงความละเอียดของข้อมูลส่งผลให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นครั้งที่ 2 ก็ยังคงถูกสร้างจากข้อมูลของตารางที่มีความละเอียดต่ำ รวมทั้งมีความคลาดเคลื่อนมาจากการสุ่มของโปรแกรม ทำให้ยังมีความคลาดเคลื่อนอยู่ในบางคุณลักษณะ

จากการประเมินผลการทดลองทั้งสองแบบที่กล่าวมานั้น แสดงให้เห็นว่ากระบวนการสร้างประชากรในระดับบุคคลและระดับครัวเรือนที่ได้ออกแบบนั้นเป็นกระบวนการที่มีประสิทธิภาพ สามารถสร้างประชากรสังเคราะห์ที่มีความใกล้เคียงกับข้อมูลต้นแบบและประชากรจริง และผู้วิจัยยังประเมินความถูกต้องของกระบวนการสร้างประชากรสังเคราะห์โดยใช้ตารางที่ถูกซ่อน ซึ่งจากผลการทดลองที่แสดงในตารางที่ 36 และตารางที่ 37 พบว่ามีค่าเฉลี่ยความคลาดเคลื่อนของความคลาดเคลื่อนสัมบูรณ์ของตารางที่ถูกซ่อนของคุณลักษณะสถานภาพสมรสของเพศชายและเพศหญิง คือ 0.126% และ 0.102% ตามลำดับ ซึ่งแสดงว่าประชากรสังเคราะห์ที่ถูกสร้างขึ้นทั้งสองครั้งนั้นมีความใกล้เคียงกันมาก

ผลการประเมินการทดลองทั้งหมด พบว่า กระบวนการสร้างประชากรสังเคราะห์ในระดับบุคคลและระดับครัวเรือนที่ได้ออกแบบมีความเหมาะสมกับการสร้างประชากรสังเคราะห์ของประเทศไทย รวมถึงประเทศที่มีข้อมูลอย่างจำกัด เช่น ประเทศที่กำลังพัฒนา โดยสามารถระบุประชากรได้อย่างชัดเจนซึ่งมีคุณลักษณะที่หลากหลายและสามารถรองรับการสร้างครัวเรือนที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ

5.2 ข้อเสนอแนะและแนวทางในการพัฒนาต่อ

จากการทดลองพบว่าประชากรสังเคราะห์ที่ถูกสร้างขึ้นจะขึ้นอยู่กับคุณภาพของข้อมูลที่นำมาใช้ ซึ่งหากมีข้อมูลคุณลักษณะอื่นๆ เพิ่มเติม พร้อมกับข้อมูลแต่ละคุณลักษณะมีความละเอียดสูงและเป็นความละเอียดเดียวกันในทุกคุณลักษณะ จะทำให้ข้อมูลมีความสอดคล้องกัน ทำให้ประชากรสังเคราะห์ที่ได้มีความคลาดเคลื่อนน้อยลง ส่งผลให้ประชากรสังเคราะห์ที่ถูกสร้างขึ้นมีความถูกต้องใกล้เคียงกับประชากรจริงมากที่สุด นอกจากนี้ถ้ามีข้อมูลของประชากรรายปี เมื่อนำมาสร้างประชากรสังเคราะห์ ทำให้สามารถดูความเปลี่ยนแปลงของประชากรในแต่ละปีได้ ซึ่งเป็นประโยชน์ต่อการวิเคราะห์และวางแผนต่างๆ เกี่ยวกับประชากร และกระบวนการสร้างแบบจำลองที่ได้ออกแบบนี้ยังสามารถนำไปประยุกต์ใช้กับข้อมูลอื่นๆ เพื่อสร้างแบบจำลองในรูปแบบอื่นๆ ที่สนใจได้อีกด้วย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

1. Harding, A., et al., *Assessing Poverty and Inequality at a Detailed Regional Level: New Advances in Spatial Microsimulation*, in *United Nations University World Institute for Development Economics Research*. 2004.
2. Huang, Z. and P. Williamson, *A Comparison of Synthetic Reconstruction and Combinatorial Optimization Approaches to the Creation of Small-Area Microdata*, in *Working paper. Liverpool, United Kingdom: University of Liverpool, Department of Geography*. 2002.
3. Johan, B. and C. Eric, *Synthetic population: review of the different approaches*, in*. Working paper. University of Namur, Belgium*. 2012.
4. Barthelemy, J. and P.L. Toint, *Synthetic Population Generation Without a Sample*, in *Transportation Science*. 2012. p. 1-14.
5. Ballas, D., et al., *SimBritain: A Spatial Microsimulation Approach to Population Dynamics*, in *Population, Space and Place* 2005. p. 13-34.
6. Ballas, D., G.P. Clarke, and E. Wiemers, *Building a Dynamic Spatial Microsimulation Model for Ireland*, in *Population, Space and Place* 2005. p. 157-172.
7. Hancock, R., J. Mallender, and S. Pudney, *Constructing a computer model for simulating the future distribution of pensioner's income for Great Britain*. Microsimulation models for public policy analysis: new frontiers, Suntory-Toyota International Centre for Economics and Related Disciplines – LSE, ed. (eds). 1992, London. 33-66.
8. Hancock, R. and H. Sutherland, *Microsimulation models for public policy analysis: new frontiers. Suntory-Toyota International Centre for Economics and Related Disciplines – LSE*. 1992, London.
9. Pavon, J., et al., *Agent-based modelling and Simulation for the analysis of social patterns. Pattern Recognition Letters*, in *Pattern Recognition Letters*. 2008. p. 1039-1048.

10. Pearson, J., et al., *Primary Care in an Aging Society: Building and Testing a Microsimulation Model for Policy Purposes.* , in *Social Science Computer Review*. 2011. p. 21-36.
11. BP, W., et al., *Canada's Population Health Model (POHEM): a tool for performing economic evaluations of cancer control interventions*, in *European Journal of Cancer*. 2001. p. 1797–1804.
12. Li, S., et al., *Agent-based modelling of the patial pattern of leisure visitation in forests : A case study in Wallonia, south Belgium*, in *Environmental Modelling & Software* 2015. p. 111-125.
13. Lin, S., M.H. DeVisser, and J.P. Messina, *An agent-based model to simulate tsetse fly distribution and control techniques: A case study in Nguruman, Kenya*, in *Ecological Modelling* 2015. p. 80–89.
14. Carter, N., et al., *Modeling tiger population and territory dynamics using anagent-based approach*, in *Ecological Modelling*. 2015. p. 347–362.
15. Long, Y. and Z. Shen, *Disaggregation heterogeneous agent attributes and location*, in *Computer, Environment and Urban Systems*. 2013. p. 14-25.
16. Motulsky, H., *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*. 2010, New York,: Oxford University Press.
17. Jarosz, B., *Estimating Households by Household Size Using the Poisson Distribution*, in *Population Association of America*. 2013.
18. Williamson, P., M. Birkin, and P. Rees, *The estimation of population microdata by using data from small area statistics and samples of anonymised records*, in *Environment and Planning A*. 1998. p. 785-816.
19. Williamson, P., *An Evaluation of Two Synthetic Small-Area Microdata Simulation Methodologies: Synthetic Reconstruction and Combinatorial Optimisation*. Vol. 6. 2013.
20. Rahman, A., et al., *Methodological Issues in Spatial Microsimulation Modelling for Small Area Estimation*, in *International Journal of Microsimulation* 2010. p. 3-22.
21. Fernandez, S.J., et al., *Construction of Synthetic Populations with Key Attributes : Simulation Set-up while Accommodating Multiple Approaches*

within a Flexible Simulation Platform, in *Social computing: IEEE international conference on privacy, security, risk and trust, PASSAT* 2010. p. 701-706.

22. Mohammadian, A.K., M. Javanmardi, and Y. Zhang, *Synthetic household travel survey data simulation*, in *Transportation Research Part C* 2010. p. 869–878.
23. Ge, Y., et al., *Virtual city : An individual-based digital environment for human mobility and inreractive behavior*, in *Simulation: Transactions of the Society for Modeling an Simulation International*. 2014. p. 917-935.
24. สำนักงานสถิติแห่งชาติ. สื่อมวลชนสัญจรกับการทำสำมะโนประชากรและเคหะ พ.ศ. 2553. 2552 [cited 2558 16 กันยายน].
25. สำนักงานสถิติแห่งชาติ. สำมะโนประชากรและเคหะ. 2547 [cited 2557 25 กุมภาพันธ์].
26. สำนักงานสถิติจังหวัดพิษณุโลก. รายงานสถิติจากการสำมะโนของจังหวัดพิษณุโลก. [cited 2557 25 กุมภาพันธ์].
27. กรมการปกครอง, ส.ส. ระบบสถิติทางการทะเบียน. [cited 2557 25 กุมภาพันธ์].
28. สำนักงานปลัดกระทรวงศึกษาธิการ, ก.ศ. สารสนเทศเพื่อการศึกษาแห่งชาติ. [cited 2557 25 กุมภาพันธ์].
29. Rutstein, S.O., *Trends in Birth Spacing*. 2011, Demographic and Health Research Division, ICF Macro: Calverton, Maryland, USA.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก

แสดงตัวอย่างตารางข้อมูลที่ใช้ในการสร้างประชากรสังเคราะห์ของจังหวัดพิษณุโลก

ตารางที่ ก.1 แสดงรายงานสถิติจำนวนประชากรและจำนวนหลังคาเรือนของจังหวัดพิษณุโลก พ.ศ. 2553

อำเภอ	ชาย (คน)	หญิง (คน)	รวม (คน)	หลังคาเรือน (หลัง)
จังหวัดพิษณุโลก	417,293	432,399	849,692	296,807
อำเภอเมืองพิษณุโลก	95,417	99,211	194,628	78,989
อำเภอนครไทย	33,293	32,931	66,224	20,331
อำเภอชาติตระการ	17,585	16,984	34,569	10,822
อำเภอบางระกำ	37,043	37,864	74,907	22,683
อำเภอบางกระทุ่ม	16,436	17,262	33,698	9,984
อำเภอพรหมพิราม	40,849	42,132	82,981	26,406
อำเภอวัดโบสถ์	14,546	14,575	29,121	9,548
อำเภอวังทอง	57,264	57,394	114,658	36,648
อำเภอเนินมะปราง	19,598	19,386	38,984	12,643
ท้องถิ่นเทศบาลตำบลบ้านมุง	3,466	3,394	6,860	2,189
ท้องถิ่นเทศบาลตำบลบึงระมาณ	2,064	2,235	4,299	1,510
ท้องถิ่นเทศบาลตำบลพลายชุมพล	3,068	4,839	7,907	2,610
ท้องถิ่นเทศบาลตำบลพันเสา	3,029	3,054	6,083	1,836
ท้องถิ่นเทศบาลตำบลไทรย้อย	4,306	4,236	8,542	2,688
ท้องถิ่นเทศบาลตำบลบ้านแยง	4,804	4,518	9,322	3,731
ท้องถิ่นเทศบาลตำบลเนินมะปราง	1,707	1,780	3,487	1,201
ท้องถิ่นเทศบาลตำบลวังทอง	2,351	2,476	4,827	1,509
ท้องถิ่นเทศบาลตำบลวัดโบสถ์	3,921	4,351	8,272	3,363

ตารางที่ ก.1 แสดงรายงานสถิติจำนวนประชากรและจำนวนหลังคาเรือนของจังหวัดพิษณุโลก พ.ศ. 2553 (ต่อ)

อำเภอ	ชาย (คน)	หญิง (คน)	รวม (คน)	หลังคาเรือน (หลัง)
ท้องถิ่นเทศบาลตำบลวงฆ้อง	1,793	1,830	3,623	1,333
ท้องถิ่นเทศบาลตำบลพรหมพิราม	607	658	1,265	488
ท้องถิ่นเทศบาลตำบลบางกระทุ่ม	733	777	1,510	548
ท้องถิ่นเทศบาลตำบลเนินกุ่ม	6,537	6,860	13,397	3,982
ท้องถิ่นเทศบาลตำบลปลักแรด	1,798	2,008	3,806	1,156
ท้องถิ่นเทศบาลตำบลบางระกำ	2,299	2,447	4,746	1,728
ท้องถิ่นเทศบาลตำบลป่าแดง	2,756	2,796	5,552	2,050
ท้องถิ่นเทศบาลตำบลนครไทย	4,708	4,959	9,667	3,700
ท้องถิ่นเทศบาลตำบลบ้านใหม่	918	991	1,909	672
ท้องถิ่นเทศบาลนครพิษณุโลก	34,397	40,451	74,848	32,459

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.2 แสดงรายงานสถิติจำนวนประชากรจำแนกตามอายุของจังหวัดพิษณุโลก พ.ศ. 2553

อายุ	ชาย (คน)	หญิง (คน)	รวม (คน)	อายุ	ชาย (คน)	หญิง (คน)	รวม (คน)
น้อยกว่า 1 ปี	4,460	4,279	8,739	**1 ปี**	4,754	4,410	9,164
2 ปี	4,733	4,553	9,286	**3 ปี**	4,901	4,724	9,625
4 ปี	4,965	4,680	9,645	**5 ปี**	5,170	4,687	9,857
6 ปี	5,148	4,652	9,800	**7 ปี**	4,740	4,616	9,356
8 ปี	4,819	4,633	9,452	**9 ปี**	4,769	4,647	9,416
10 ปี	5,151	4,836	9,987	**11 ปี**	5,090	4,937	10,027
12 ปี	5,668	5,215	10,883	**13 ปี**	6,200	5,870	12,070
14 ปี	6,268	5,972	12,240	**15 ปี**	6,081	5,816	11,897
16 ปี	6,058	5,712	11,770	**17 ปี**	5,962	5,692	11,654
18 ปี	6,254	6,885	13,139	**19 ปี**	7,042	8,094	15,136
20 ปี	6,822	7,291	14,113	**21 ปี**	7,526	6,420	13,946
22 ปี	6,677	5,968	12,645	**23 ปี**	5,963	5,877	11,840
24 ปี	6,092	5,864	11,956	**25 ปี**	6,391	6,116	12,507
26 ปี	6,257	6,141	12,398	**27 ปี**	6,559	6,248	12,807
28 ปี	6,765	6,568	13,333	**29 ปี**	6,615	6,507	13,122
30 ปี	6,769	6,599	13,368	**31 ปี**	6,792	6,668	13,460
32 ปี	6,679	6,571	13,250	**33 ปี**	6,714	6,672	13,386
34 ปี	6,925	6,832	13,757	**35 ปี**	6,781	6,881	13,662
36 ปี	6,776	7,036	13,812	**37 ปี**	6,731	6,883	13,614
38 ปี	6,615	6,791	13,406	**39 ปี**	6,977	7,266	14,243
40 ปี	6,994	7,245	14,239	**41 ปี**	6,774	7,146	13,920
42 ปี	7,108	7,792	14,900	**43 ปี**	6,938	7,500	14,438
44 ปี	7,064	7,534	14,598	**45 ปี**	7,141	7,655	14,796
46 ปี	7,343	8,498	15,841	**47 ปี**	6,668	7,006	13,674
48 ปี	6,797	7,094	13,891	**49 ปี**	6,561	6,934	13,495
50 ปี	6,392	6,814	13,206	**51 ปี**	6,519	6,711	13,230
52 ปี	5,594	5,946	11,540	**53 ปี**	5,431	5,914	11,345
54 ปี	5,229	5,852	11,081	**55 ปี**	5,153	5,779	10,932
56 ปี	4,476	4,966	9,442	**57 ปี**	4,589	5,135	9,724
58 ปี	4,306	4,847	9,153	**59 ปี**	3,896	4,344	8,240

ตารางที่ ก.2 แสดงรายงานสถิติจำนวนประชากรจำแนกตามอายุของจังหวัดพิษณุโลก พ.ศ. 2553 (ต่อ)

อายุ	ชาย (คน)	หญิง (คน)	รวม (คน)	อายุ	ชาย (คน)	หญิง (คน)	รวม (คน)
60 ปี	3,964	4,416	8,380	**61 ปี**	3,423	3,921	7,344
62 ปี	3,215	3,573	6,788	**63 ปี**	2,810	3,168	5,978
64 ปี	2,543	2,996	5,539	**65 ปี**	2,069	2,589	4,658
66 ปี	2,218	2,523	4,741	**67 ปี**	1,996	2,355	4,351
68 ปี	2,453	3,018	5,471	**69 ปี**	2,010	2,491	4,501
70 ปี	1,951	2,382	4,333	**71 ปี**	1,964	2,387	4,351
72 ปี	1,930	2,364	4,294	**73 ปี**	1,827	2,201	4,028
74 ปี	1,619	2,155	3,774	**75 ปี**	1,336	1,797	3,133
76 ปี	1,379	1,939	3,318	**77 ปี**	1,103	1,604	2,707
78 ปี	1,112	1,537	2,649	**79 ปี**	842	1,268	2,110
80 ปี	910	1,299	2,209	**81 ปี**	708	1,095	1,803
82 ปี	747	1,119	1,866	**83 ปี**	468	846	1,314
84 ปี	447	702	1,149	**85 ปี**	374	612	986
86 ปี	291	463	754	**87 ปี**	250	440	690
88 ปี	187	326	513	**89 ปี**	157	281	438
90 ปี	137	234	371	**91 ปี**	65	148	213
92 ปี	83	146	229	**93 ปี**	60	106	166
94 ปี	62	75	137	**95 ปี**	40	55	95
96 ปี	32	64	96	**97 ปี**	23	31	54
98 ปี	14	21	35	**99 ปี**	14	14	28
100 ปี	9	10	19	**มากกว่า 100 ปี**	24	35	59

ตารางที่ ก.3 แสดงจำนวนนักเรียนทุกระดับชั้นของจังหวัดพิษณุโลก ปีการศึกษา 2553

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้แสดงตั้งแต่ระดับชั้นอนุบาล 1 จนถึงระดับชั้นมัธยมศึกษาปีที่ 6

ระดับชั้น	เพศ	จำนวน (คน)
อนุบาล 1	ชาย	4,843
	หญิง	4,349
อนุบาล 2	ชาย	4,615
	หญิง	4,321
ประถมศึกษาปีที่ 1	ชาย	4,943
	หญิง	4,481
ประถมศึกษาปีที่ 2	ชาย	4,602
	หญิง	4,540
ประถมศึกษาปีที่ 3	ชาย	4,694
	หญิง	4,531
ประถมศึกษาปีที่ 4	ชาย	4,876
	หญิง	4,583
ประถมศึกษาปีที่ 5	ชาย	4,982
	หญิง	4,564
ประถมศึกษาปีที่ 6	ชาย	5,567
	หญิง	5,087
มัธยมศึกษาปีที่ 1	ชาย	6,369
	หญิง	5,718
มัธยมศึกษาปีที่ 2	ชาย	5,930
	หญิง	5,749
มัธยมศึกษาปีที่ 3	ชาย	5,826
	หญิง	5,570
มัธยมศึกษาปีที่ 4	ชาย	2,952
	หญิง	3,684

ตารางที่ ก.3 แสดงจำนวนนักเรียนทุกระดับชั้นของจังหวัดพิษณุโลก ปีการศึกษา 2553 (ต่อ)

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้แสดงตั้งแต่ระดับชั้นอนุบาล 1 จนถึงระดับชั้นมัธยมศึกษาปีที่ 6

ระดับชั้น	เพศ	จำนวน (คน)
มัธยมศึกษาปีที่ 5	ชาย	2,282
	หญิง	3,074
มัธยมศึกษาปีที่ 6	ชาย	2,131
	หญิง	2,834

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.4 แสดงตัวอย่างข้อมูลจำนวนนักเรียนและนักศึกษาจำแนกตามสถานศึกษาของจังหวัดพิษณุโลก ปีการศึกษา 2553 โดยแสดงสถานะของสถานศึกษาเป็นรหัสสถานศึกษา

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลของสถานศึกษาขนาดเล็ก

รหัสสถานศึกษา	อำเภอ	อนุบาล 1 (คน)		อนุบาล 2 (คน)		ประถมศึกษาปีที่ 1 (คน)		ประถมศึกษาปีที่ 2 (คน)		ประถมศึกษาปีที่ 3 (คน)		ประถมศึกษาปีที่ 4 (คน)		ประถมศึกษาปีที่ 5 (คน)		ประถมศึกษาปีที่ 6 (คน)	
		ช	ญ	ช	ญ	ช	ญ	ช	ญ	ช	ญ	ช	ญ	ช	ญ	ช	ญ
1065360001	เมืองพิษณุโลก	8	3	10	7	11	9	10	9	12	7	8	8	8	5	11	9
1065360002	เมืองพิษณุโลก	3	1	7	3	7	5	5	2	0	4	3	2	5	7	2	5
1065360003	เมืองพิษณุโลก			2	1			1	1	0	1	3	0	3	0	2	1
1065360004	เมืองพิษณุโลก					4	0	2	0	3	2	3	0	5	3	3	0
1065360005	เมืองพิษณุโลก	6	2	5	6	2	8	7	4	4	5	6	6	6	4	6	4
1065360006	เมืองพิษณุโลก					1	0					0	2			1	1
1065360007	เมืองพิษณุโลก	7	5	2	3	1	2	4	6	2	2	6	4	1	5	1	5
1065360009	เมืองพิษณุโลก	12	9	16	10	11	14	15	12	17	11	11	9	17	10	15	17
1065360010	เมืองพิษณุโลก	20	19	25	18	31	18	25	15	19	18	22	19	26	17	28	28
1065360011	เมืองพิษณุโลก	13	7	6	7	12	9	7	5	5	7	7	11	12	2	16	5
1065360012	เมืองพิษณุโลก	7	5	12	2	20	22	16	17	19	12	7	11	16	18	15	14
1065360014	เมืองพิษณุโลก	10	6	12	10	3	6	4	1	4	2	7	7	8	5	6	4
1065360015	เมืองพิษณุโลก					156	156	145	146	139	149	135	146	148	158	150	178

ตารางที่ ก.5 แสดงจำนวนประชากรอายุตั้งแต่ 13 ปีขึ้นไป จำแนกตามสถานภาพสมรส กลุ่มอายุ เพศ และเขตการปกครองของภาคเหนือ พ.ศ. 2553

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามกลุ่มอายุของภาคเหนือ

กลุ่มอายุ (ปี)	รวม (คน)	สถานภาพสมรส						
		โสด (คน)	เคยสมรส					ไม่ทราบสถานภาพสมรส (คน)
			สมรส (คน)	หม้าย (คน)	หย่า (คน)	แยกกันอยู่ (คน)	ไม่ทราบ (คน)	
13 - 14	381,155	371,960	3,018	93	38	101	1,712	4,233
15 - 19	853,607	767,888	69,153	912	750	3,864	8,163	2,877
20 - 24	693,185	453,377	208,053	2,690	3,167	11,773	11,122	3,002
25 - 29	697,443	287,508	361,997	5,183	7,639	19,467	12,311	3,339
30 - 34	761,138	195,415	500,574	10,244	13,881	25,825	12,520	2,678
35 - 39	844,247	140,387	626,304	18,280	18,575	28,214	10,378	2,110
40 - 44	958,804	107,527	752,918	33,072	23,539	30,169	9,175	2,405
45 - 49	1,043,979	93,218	835,215	50,347	25,768	29,204	7,957	2,271
50 - 54	985,794	72,015	784,260	74,433	22,523	23,980	6,119	2,464
55 - 59	783,081	49,284	608,455	87,649	14,656	17,060	4,019	1,960

ตารางที่ ก.5 แสดงจำนวนประชากรอายุตั้งแต่ 13 ปีขึ้นไป จำแนกตามสถานภาพสมรส กลุ่มอายุ เพศ และเขตการปกครองของภาคเหนือ พ.ศ. 2553 (ต่อ)

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามกลุ่มอายุของภาคเหนือ

กลุ่มอายุ (ปี)	รวม (คน)	สถานภาพสมรส						
		โสด (คน)	เคยสมรส					ไม่ทราบสถานภาพสมรส (คน)
			สมรส (คน)	หม้าย (คน)	หย่า (คน)	แยกกันอยู่ (คน)	ไม่ทราบ (คน)	
60 - 64	563,715	31,334	412,743	97,429	7,997	10,086	2,749	1,378
65 - 69	391,164	19,058	262,823	95,747	4,514	6,148	1,885	989
70 - 74	342,779	15,293	201,231	116,309	2,960	4,474	1,611	901
75 - 79	257,167	10,362	127,207	112,708	1,938	2,877	1,415	661
80 - 84	146,080	6,035	55,319	81,085	1,000	1,207	911	522
85 ขึ้นไป	83,878	3,235	21,163	57,391	517	608	663	300

ตารางที่ ก.6 แสดงจำนวนครัวเรือนส่วนบุคคลจำแนกตามกลุ่มอายุของหัวหน้าครัวเรือน เพศและเขตการปกครอง พ.ศ.2553

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามกลุ่มอายุและเขตการปกครอง

กลุ่มอายุของหัวหน้าครัวเรือน (ปี)	รวม (ครัวเรือน)	เขตการปกครอง (ครัวเรือน)	
		ในเขตเทศบาล	นอกเขตเทศบาล
น้อยกว่า 14	218	61	157
14 - 19	6638	2270	4369
20 - 24	15610	4143	11467
25 - 29	11192	3103	8089
30 - 34	16568	3867	12701
35 - 39	24938	5408	19530
40 - 44	33630	7120	26510
45 - 49	40467	8781	31686
50 - 54	39682	9063	30619
55 - 59	33632	7694	25938
60 - 64	26573	6298	20275
65 - 69	19203	4506	14697
70 - 74	17417	4112	13305
75 - 79	11779	2767	9012
80 - 84	6628	1549	5079
85 ขึ้นไป	3301	780	2521

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ ก.7 แสดงจำนวนและข้อมูลร้อยละของหญิงสมรสอายุ 15 - 49 ปี จำแนกตามความแตกต่างของอายุคู่สมรส กลุ่มอายุ ระดับการศึกษาสูงสุดที่สำเร็จ เขตการปกครองและภาคทั่วราชอาณาจักร พ.ศ. 2552

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตาม**ความแตกต่าง**ของอายุคู่สมรส และกลุ่มอายุ

กลุ่มอายุ (ปี)	ความแตกต่างของอายุคู่สมรส						
	สามีอ่อนกว่า 5 ปีขึ้นไป	สามีอ่อนกว่า 3-4 ปี	สามีอ่อนกว่า 1-2 ปี	อายุเท่ากัน	สามีแก่กว่า 1-2 ปี	สามีแก่กว่า 3-4 ปี	สามีแก่กว่า 5 ปีขึ้นไป
	ร้อยละ	ร้อยละ	ร้อยละ	ร้อยละ	ร้อยละ	ร้อยละ	ร้อยละ
15-19	0.2	0.4	6.8	5.8	20.4	28.5	38.0
20-24	0.6	1.7	7.7	10.6	17.2	19.0	43.3
25-29	1.4	2.8	9.1	13.6	19.8	18.2	35.2
30-34	3.5	4.5	9.7	11.4	20.7	18.6	31.6
35-39	5.0	2.8	9.9	10.9	22.7	17.8	30.8
40-44	5.5	3.5	9.5	10.5	21.8	17.5	31.6
45-49	5.9	3.8	7.4	10.8	23.9	19.5	28.7

ตารางที่ ก.8 แสดงสถานภาพของหญิงเคยสมรสอายุ 13 ปีขึ้นไป จำแนกตามจำนวนบุตรที่มีชีวิตอยู่และจำนวนบุตรที่มีชีวิตอยู่โดยเฉลี่ยของภาคเหนือ พ.ศ. 2553

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามจำนวนบุตรที่มีชีวิตอยู่ของภาคเหนือ

หญิงเคยสมรส จำนวนบุตรที่มีชีวิตอยู่		**กลุ่มอายุของหญิง (ปี)**												
		13 - 14	**15 - 19**	**20 - 24**	**25 - 29**	**30 - 34**	**35 - 39**	**40 - 44**	**45 - 49**	**50 - 54**	**55 - 59**	**60-64**	**65 - 69**	**70 ปีขึ้นไป**
หญิงอายุ 13 ปีขึ้นไป		185,886	420,425	337,244	345,770	385,278	433,185	503,859	545,439	511,649	405,781	291,427	204,901	461,296
หญิงเคยสมรสอายุ 13 ปีขึ้นไป		3,085	58,805	150,102	234,321	316,932	386,102	464,389	506,573	477,829	379,433	274,095	194,515	441,868
จำนวนบุตร (คน)	**0**	2,843	30,760	49,824	57,193	50,684	41,778	40,746	42,436	36,898	28,198	19,154	12,565	28,663
	1	177	25,777	80,375	108,206	119,078	119,841	128,274	129,929	112,406	80,251	49,249	29,191	55,539
	2	27	1,961	17,038	58,049	117,736	171,725	220,006	239,689	210,509	143,813	81,372	43,461	66,074
	3	-	144	2,113	8,178	22,421	40,428	53,358	65,359	75,123	74,644	60,940	41,537	75,968
	4	-	16	374	1,869	4,510	7,725	15,335	20,322	27,829	34,181	34,301	30,849	69,840
	5	-	1	34	263	1,194	2,237	3,199	4,861	10,162	11,972	17,093	19,050	59,121
	6	-	-	6	46	275	851	1,318	1,595	1,877	3,355	5,471	9,575	41,492
	7	-	-	-	1	58	194	449	642	752	1,028	4,294	5,103	24,978
	8	-	-	-	-	17	35	161	215	358	519	946	1,752	11,223

ตารางที่ ก.8 แสดงสถานภาพของหญิงเคยสมรสอายุ 13 ปีขึ้นไป จำแนกตามจำนวนบุตรที่มีชีวิตอยู่และจำนวนบุตรที่มีชีวิตอยู่โดยเฉลี่ยของภาคเหนือ พ.ศ. 2553 (ต่อ)

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามจำนวนบุตรที่มีชีวิตอยู่ของภาคเหนือ

หญิงเคยสมรส จำนวนบุตรที่มีชีวิตอยู่		**กลุ่มอายุของหญิง (ปี)**												
		13 - 14	**15 - 19**	**20 - 24**	**25 - 29**	**30 - 34**	**35 - 39**	**40 - 44**	**45 - 49**	**50 - 54**	**55 - 59**	**60 - 64**	**65 - 69**	**70 ปีขึ้นไป**
จำนวนบุตร (คน)	**9**	-	-	-	-	2	14	102	92	226	162	260	537	4,950
	10	-	-	-	-	-	7	20	37	46	74	105	194	2,004
	11	-	-	-	-	-	-	4	5	15	12	20	60	529
	12	-	-	-	-	-	-	4	-	5	5	20	19	209
	13	-	-	-	-	-	-	-	-	-	-	3	8	49
	14	-	-	-	-	-	-	-	-	-	1	2	1	22
	>15	-	-	-	-	-	-	-	-	-	-	-	2	12
ไม่ทราบ		38	146	338	516	955	1,267	1,413	1,391	1,622	1,219	867	612	1,194

ตารางที่ ก.9 แสดงจำนวนและข้อมูลร้อยละของหญิงเคยสมรสอายุ 15 - 49 ปีที่มีบุตร จำแนกตามอายุเมื่อคลอดบุตรคนแรก กลุ่มอายุ ระดับการศึกษาสูงสุดที่สำเร็จ เขตการปกครองและภาคทั่วราชอาณาจักร พ.ศ. 2552

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามอายุเมื่อคลอดบุตรคนแรก และกลุ่มอายุ

กลุ่มอายุ (ปี)	อายุเมื่อคลอดบุตรคนแรก (ปี) Age at first birth (years)									
	น้อยกว่า 20		20-24		25-29		30-34		35 ปีขึ้นไป	
	จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ
15-19	222,866	100.0	-	-	-	-	-	-	-	-
20-24	541,917	45.6	645,657	54.4	-	-	-	-	-	-
25-29	383,478	22.8	843,693	50.1	456,722	27.1	-	-	-	-
30-34	407,840	19.8	868,598	42.1	600,703	29.1	187,814	9.1	-	-
35-39	504,860	20.8	1,074,652	44.2	528,215	21.7	258,442	10.6	63,333	2.6
40-44	444,497	17.9	1,122,814	45.2	587,537	23.6	230,349	9.3	100,093	4.0
45-49	464,730	20.1	1,055,979	45.6	526,370	22.7	187,911	8.1	79,132	3.4

ตารางที่ ก.10 แสดงจำนวนและข้อมูลร้อยละของเด็กอายุ 0 - 17 ปี จำแนกตามการอยู่อาศัยกับพ่อ แม่ กลุ่มอายุ เพศ เขตการปกครองและภาค พ.ศ. 2551

หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามการอยู่อาศัยกับบิดา มารดา และภาค

ภาค	จำนวนเด็กอายุ 0 - 17 ปี (คน)	อาศัยอยู่กับบิดาและมารดา		ไม่ได้อาศัยอยู่กับบิดาและมารดา		อาศัยอยู่กับมารดาเท่านั้น		อาศัยอยู่กับบิดาเท่านั้น	
		จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ	จำนวน (คน)	ร้อยละ
กรุงเทพมหานคร	1,494,411	1,053,460	70.5	177,150	11.9	197,177	13.2	66,623	4.4
กลาง	3,878,106	2,491,089	64.2	683,511	17.7	555,204	14.3	148,302	3.8
เหนือ	2,923,953	1,673,351	57.2	667,076	22.8	491,388	16.8	92,138	3.2
ตะวันออกเฉียงเหนือ	6,385,863	3,497,864	54.8	1,657,558	26.0	1,073,740	16.8	156,701	2.4
ใต้	2,739,959	2,057,477	75.1	313,748	11.4	295,448	10.8	73,286	2.7

ตารางที่ ก.11 แสดงช่วงอายุห่างระหว่างการมีบุตร จำแนกตาม พ.ศ. 2530

ประเทศ	น้อยกว่า 24 เดือน	น้อยกว่า 36 เดือน	35-59 เดือน	60 เดือนขึ้นไป
ไทย	26.4	52.6	31.8	15.6

ตารางที่ ก.12 แสดงจำนวนและข้อมูลร้อยละของครัวเรือนส่วนบุคคล จำแนกตามขนาดของครัวเรือน ประเภทของที่อยู่อาศัยและเขตการปกครองพ.ศ. 2553

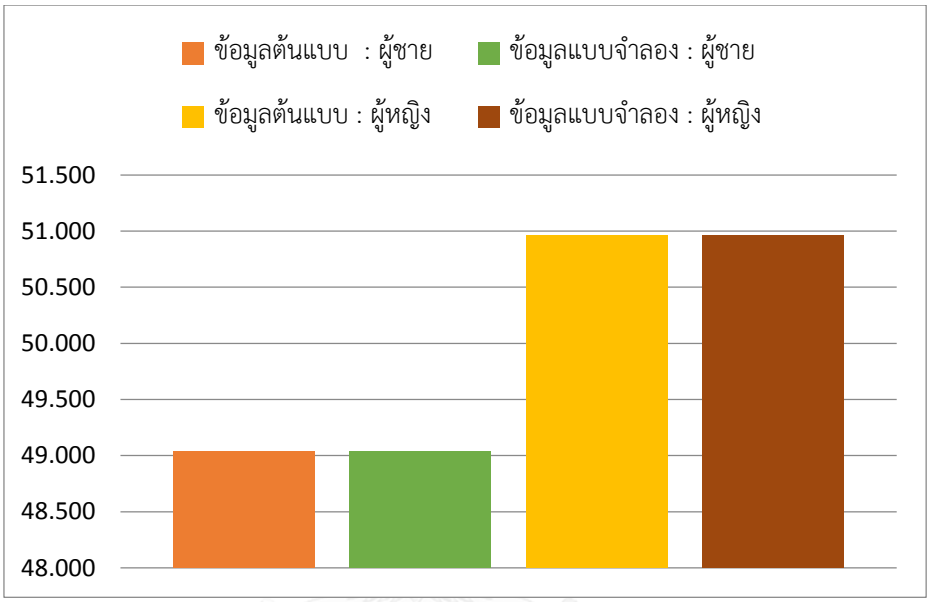
หมายเหตุ : แสดงตัวอย่างเพียงส่วนหนึ่งของข้อมูลที่มีการบันทึกไว้จริง มิใช่เป็นข้อมูลรวมทั้งหมด ในส่วนนี้เป็นตัวอย่างข้อมูลจำแนกตามขนาดของครัวเรือน

ขนาดของครัวเรือน	รวม	
	ครัวเรือน	ร้อยละ
ยอดรวม	307,476	100.0
1 คน	63,731	20.7
2 คน	71,924	23.4
3 คน	71,914	23.4
4 คน	54,077	17.6
5 คน	25,509	8.3
6 คน	13,255	4.3
7 คน	3,923	1.3
8 คน	1,721	0.6
9 คน	730	0.2
10 คนขึ้นไป	692	0.2

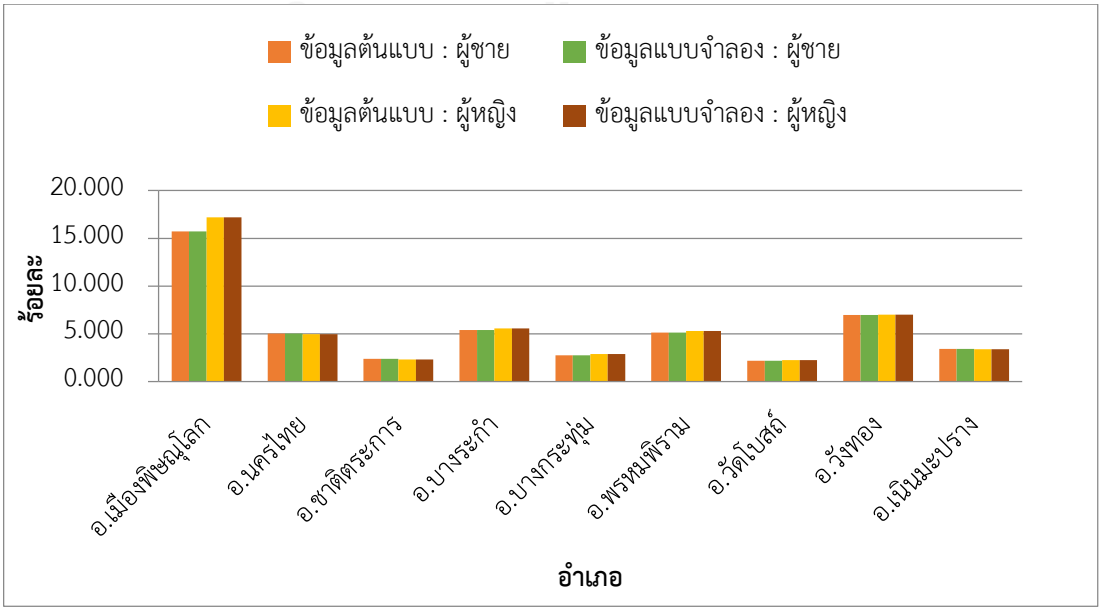
ภาคผนวก ข

แสดงตัวอย่างการเปรียบเทียบระหว่างข้อมูลต้นแบบและข้อมูลแบบจำลองของการสร้างประชากรสังเคราะห์

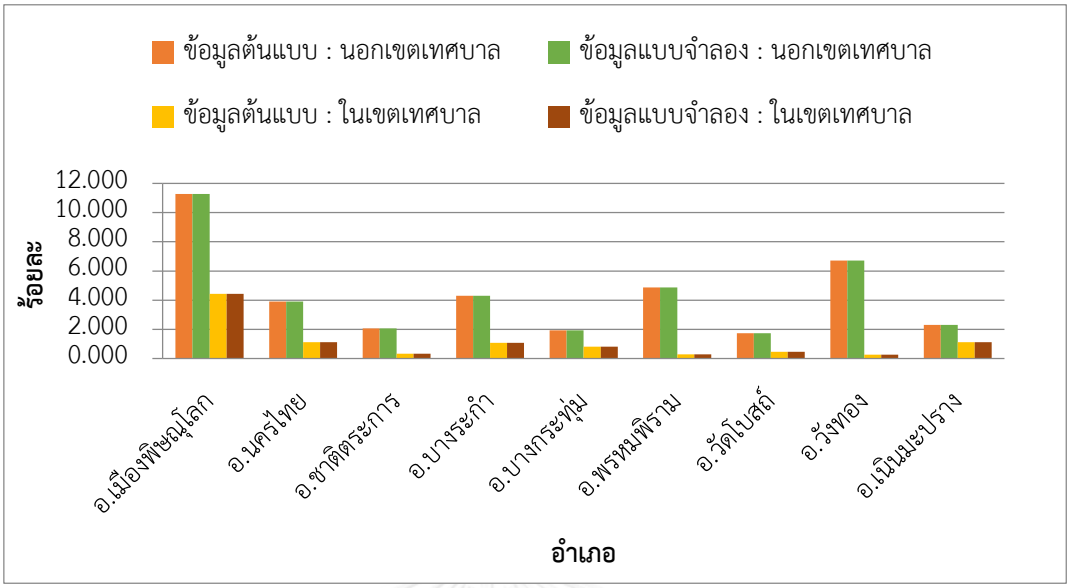
ภาพที่ ข.1 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลเพศ



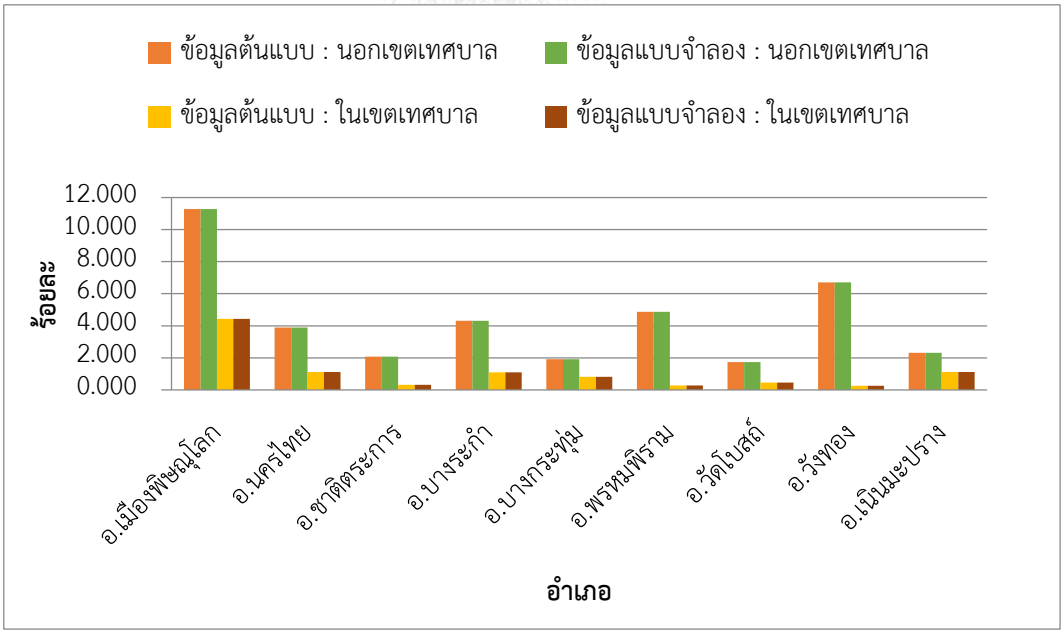
ภาพที่ ข.2 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอำเภอที่อยู่อาศัย



ภาพที่ ข.3 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลเขตเทศบาล

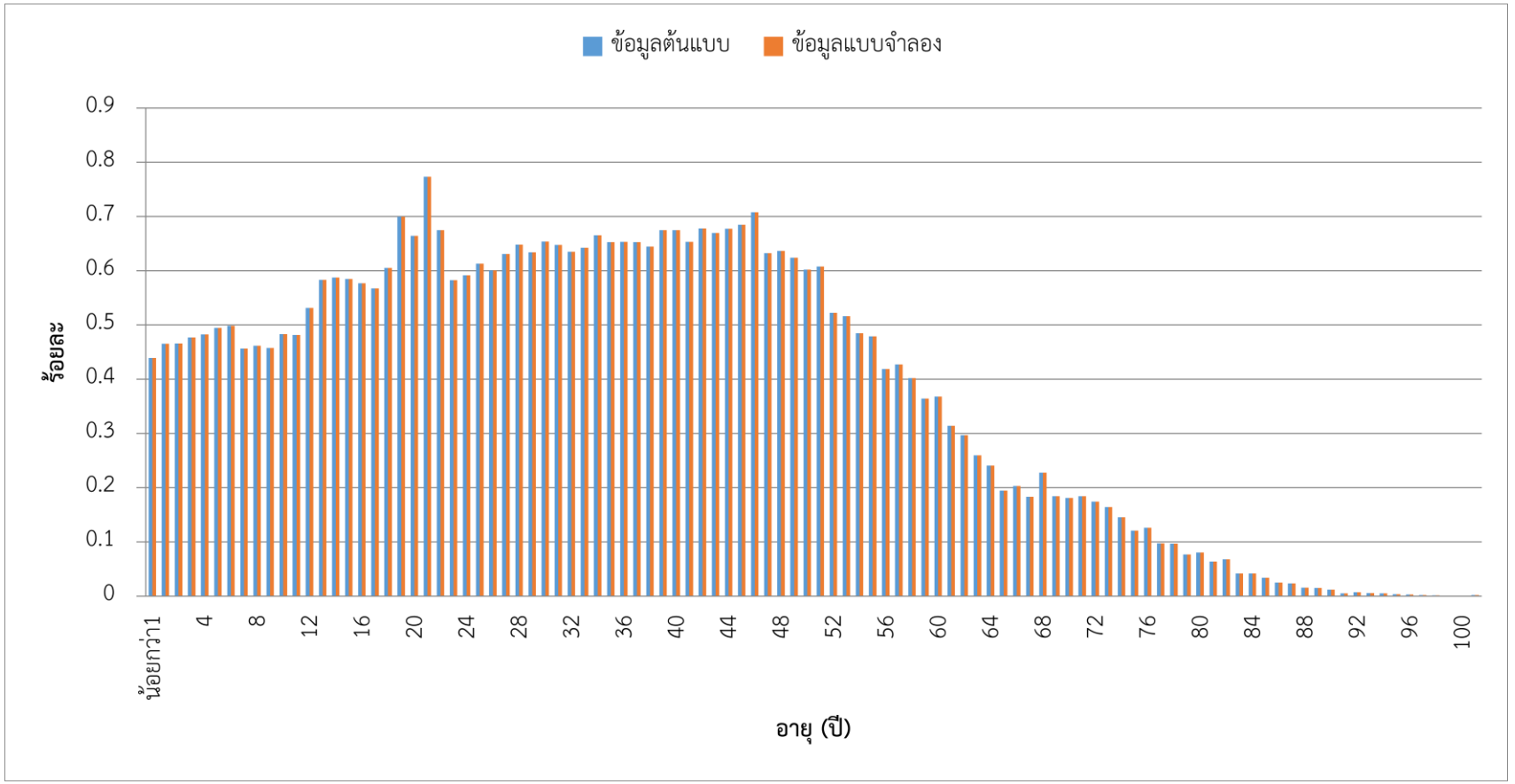


(ค) เพศชาย

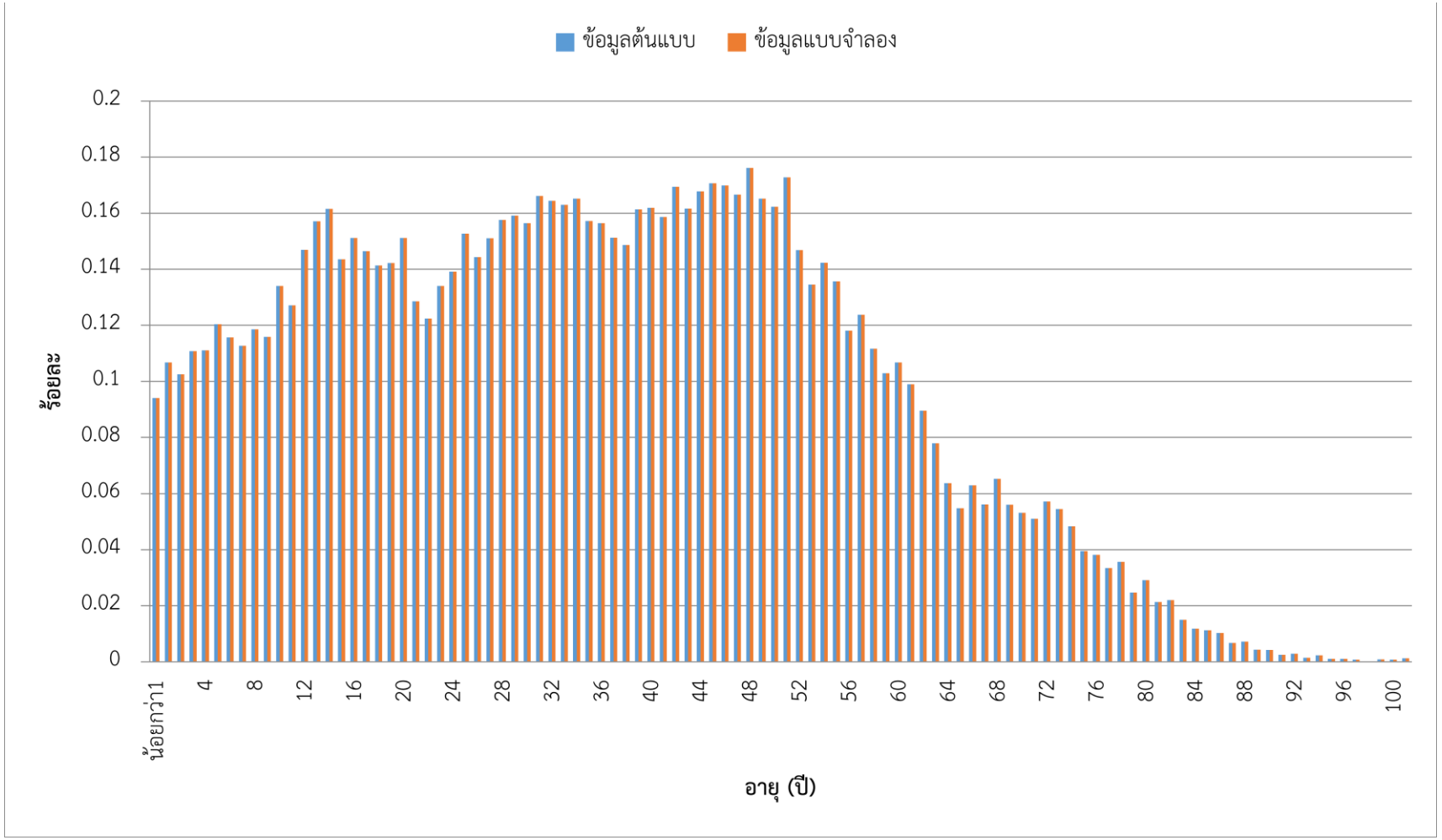


(ง) เพศหญิง

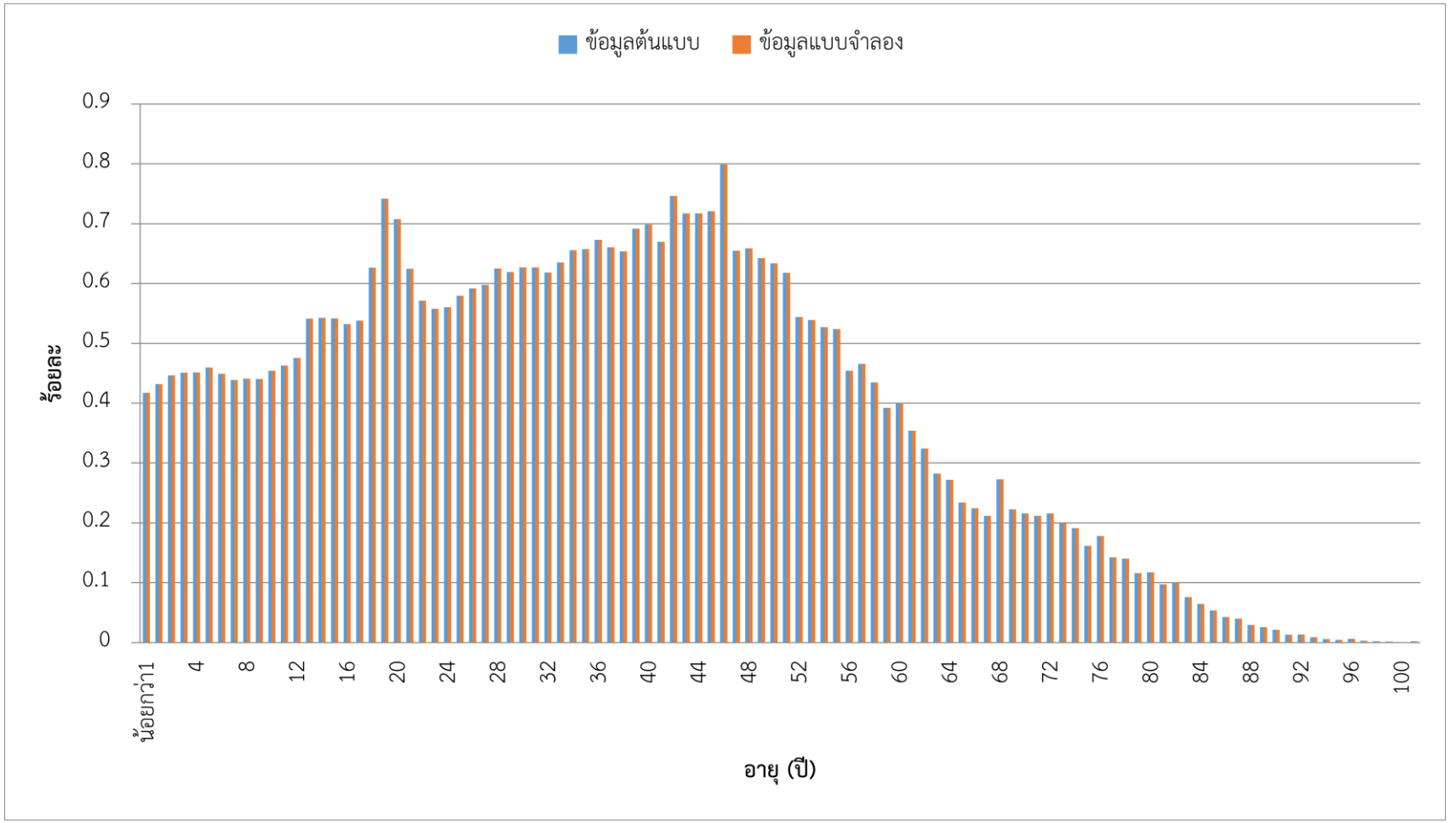
ภาพที่ ข.4 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุ



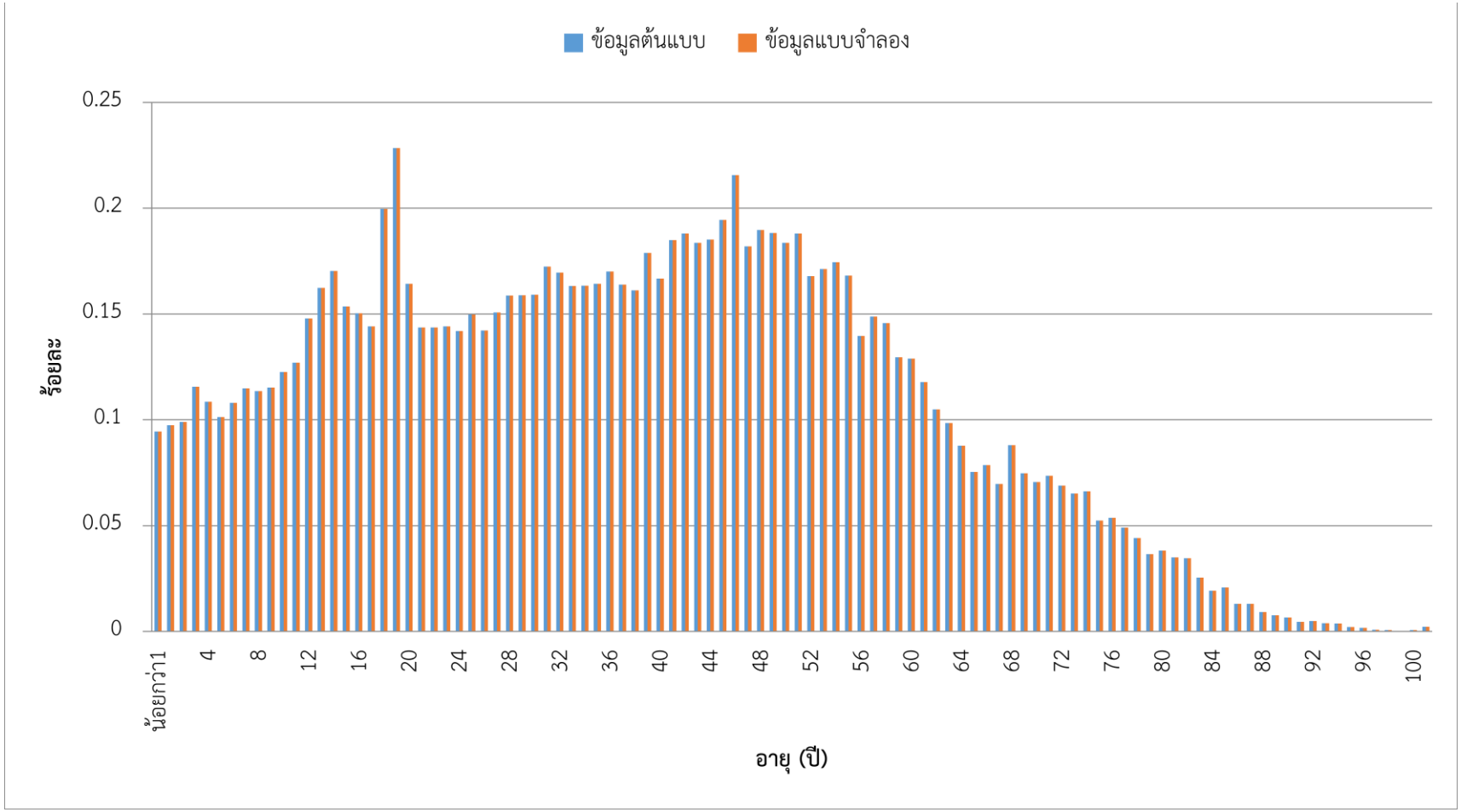
(ก) เพศชาย – นอกเขตเทศบาล



(ข) เพศชาย – ในเขตเทศบาล

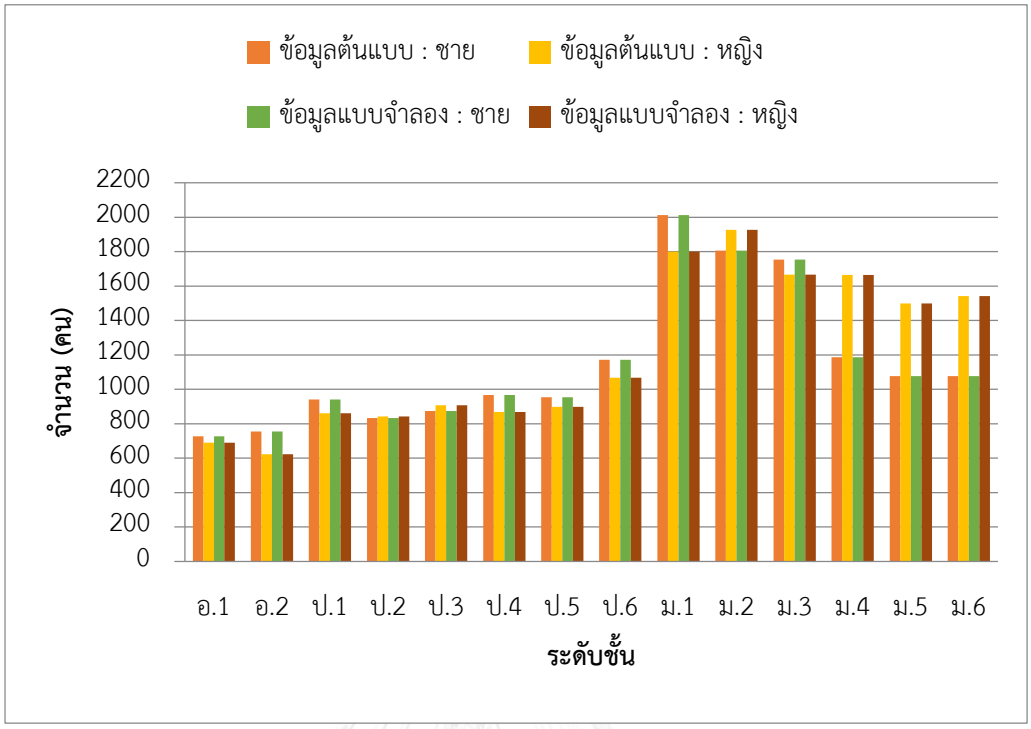


(ค) เพศหญิง – นอกเขตเทศบาล

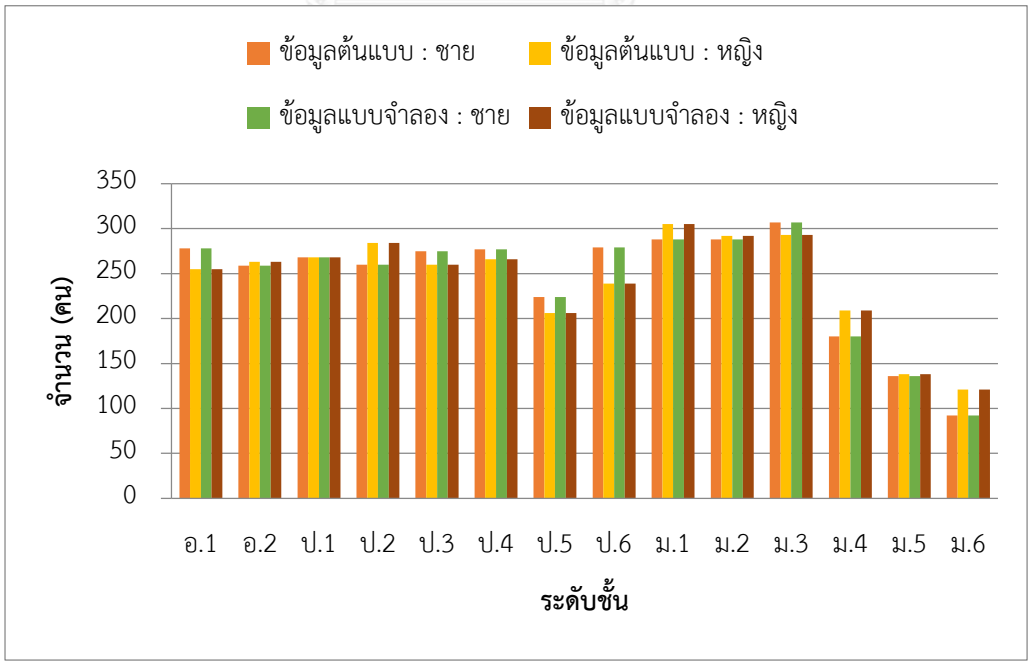


(ง) เพศหญิง – ในเขตเทศบาล

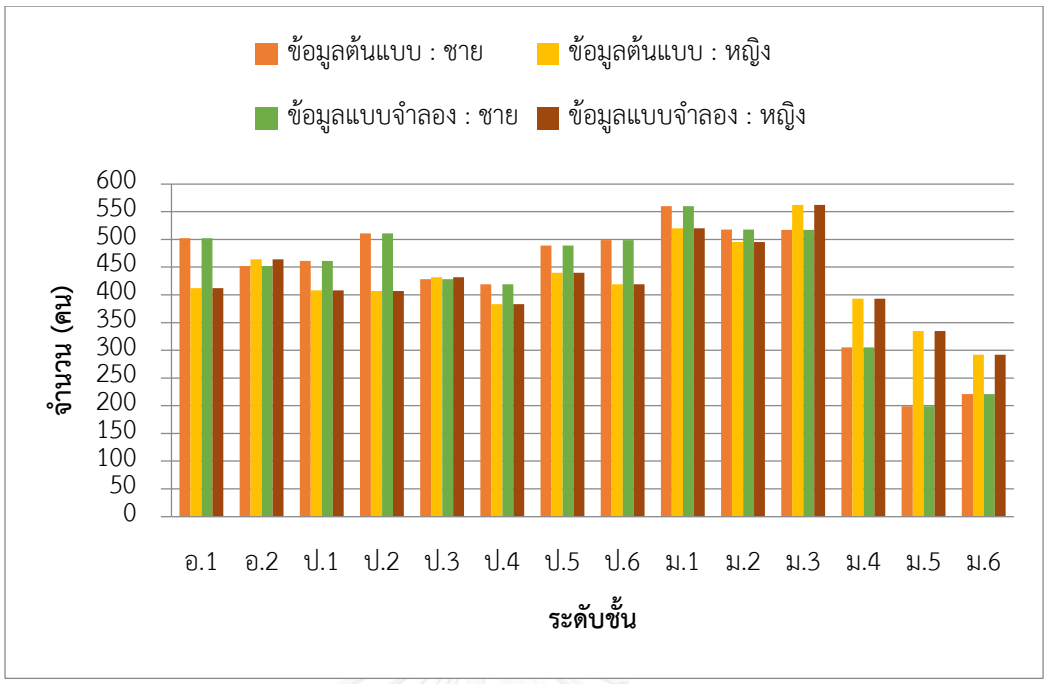
ภาพที่ ข.5 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลระดับการศึกษา



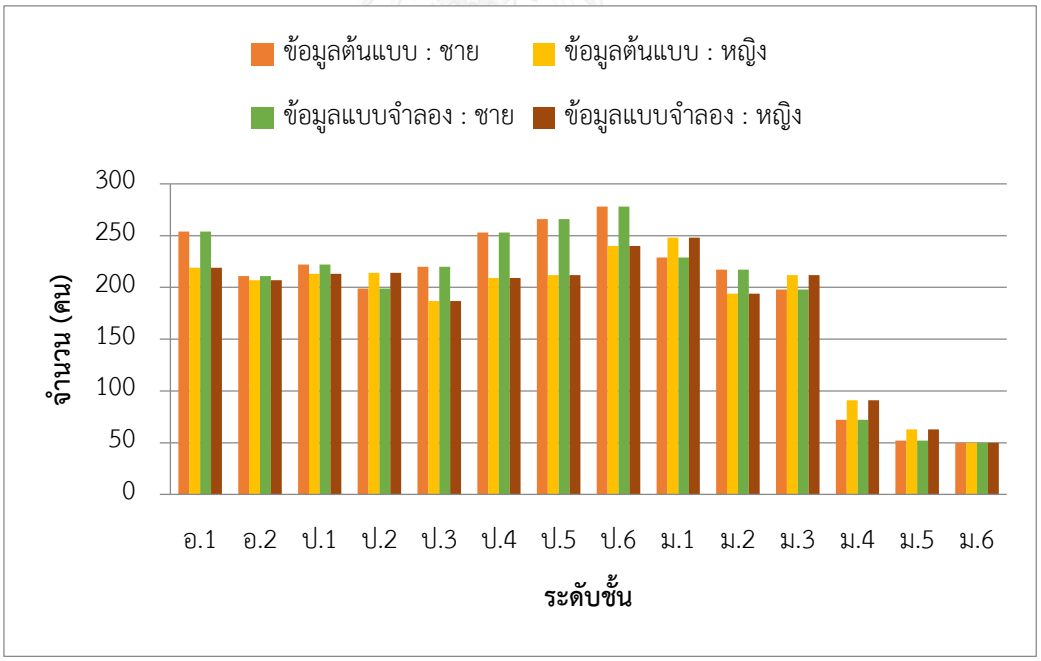
(ก) อ.เมืองพิษณุโลก



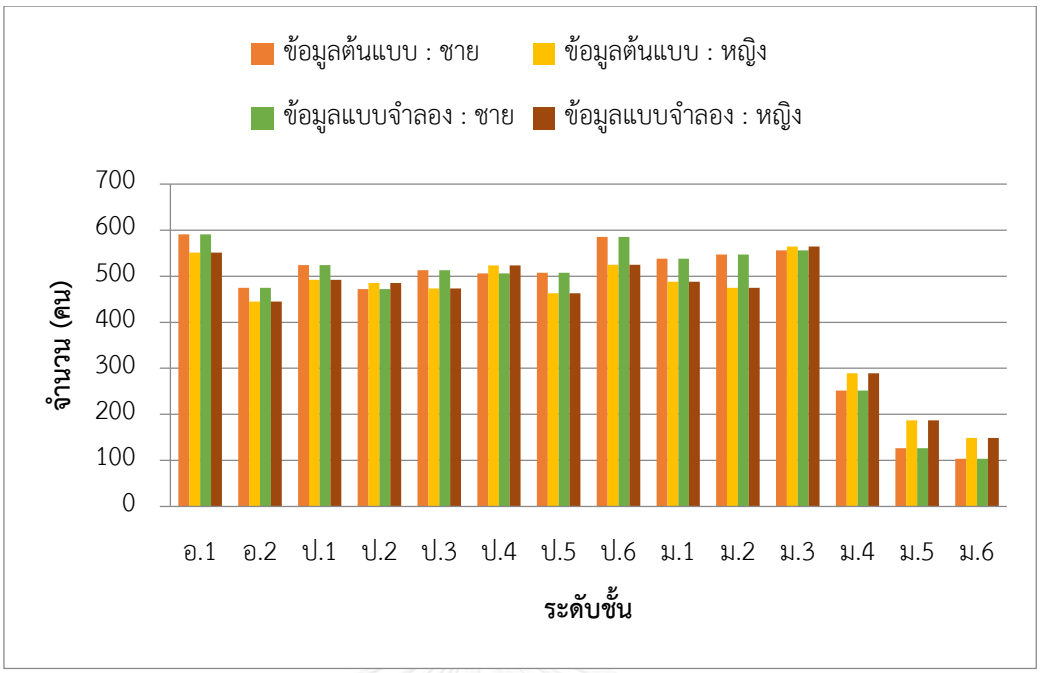
(ข) อ.ชาติตระการ



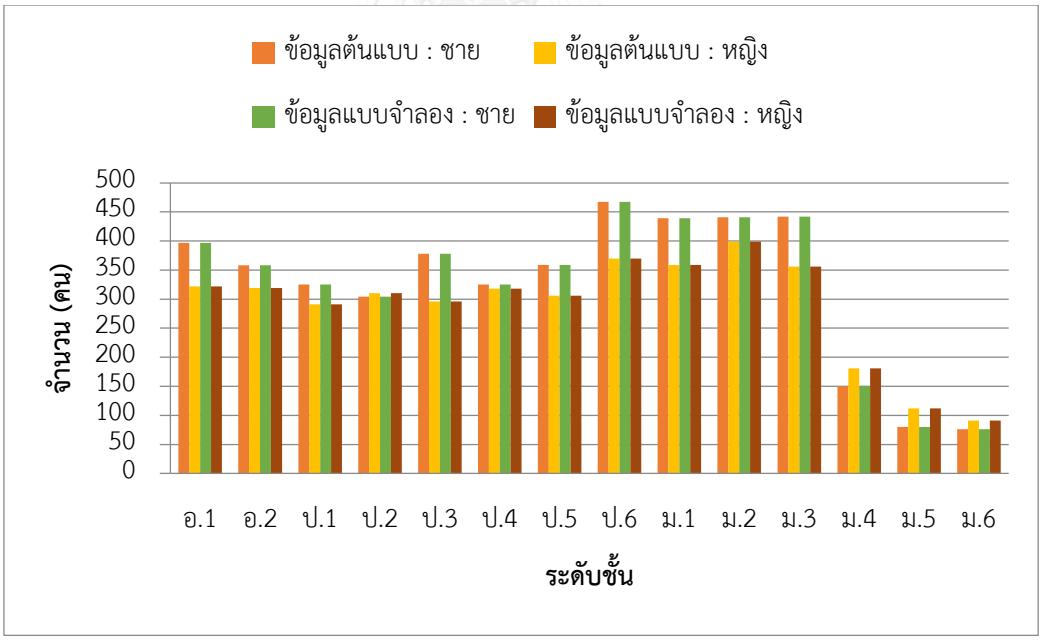
(ค) อ.นครไทย



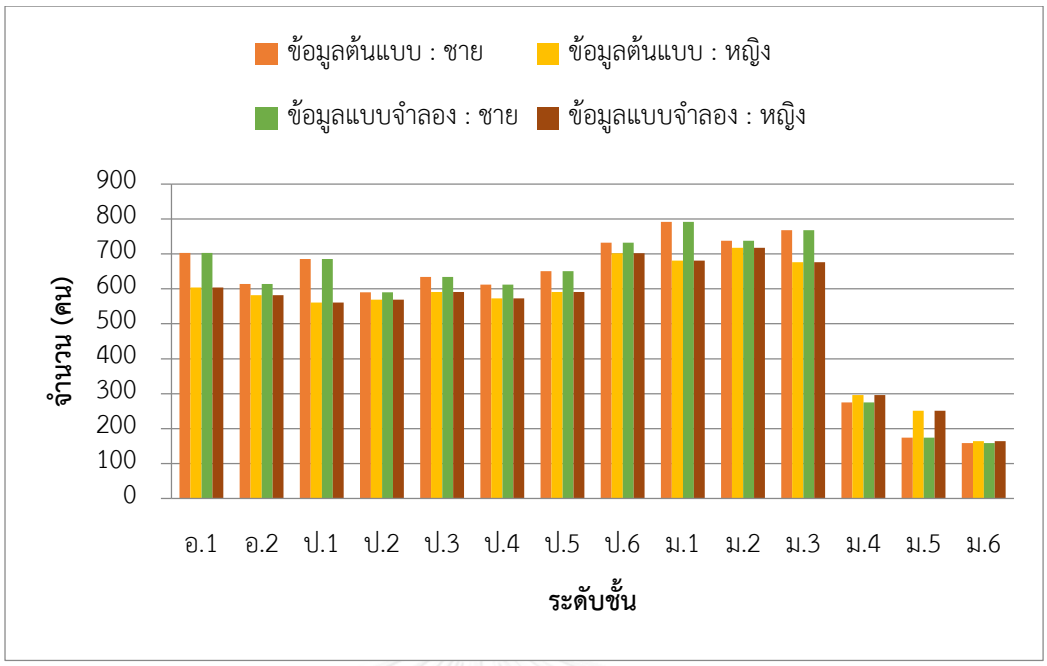
(ง) อ.บางกระทุ่ม



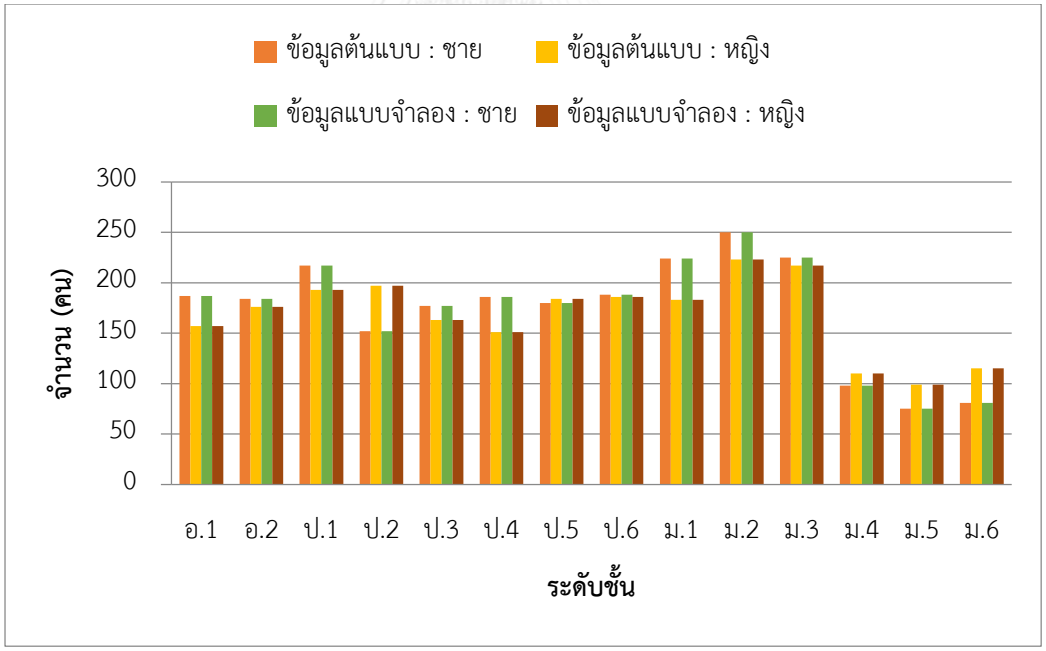
(จ) อ.บางระกำ



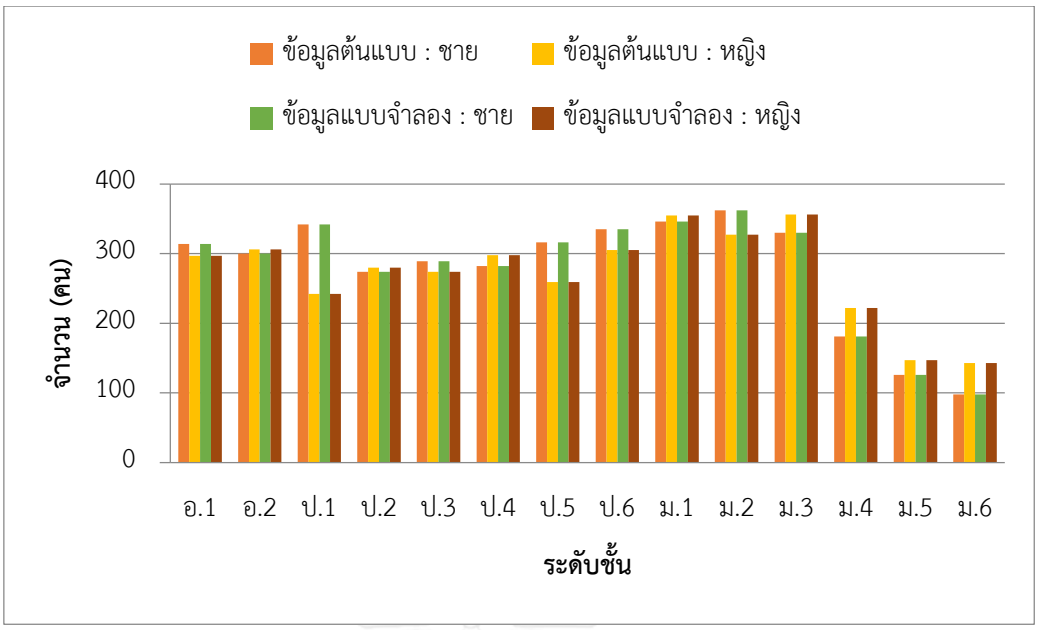
(ฉ) อ.พรหมพิราม



(ช) อ.วังทอง

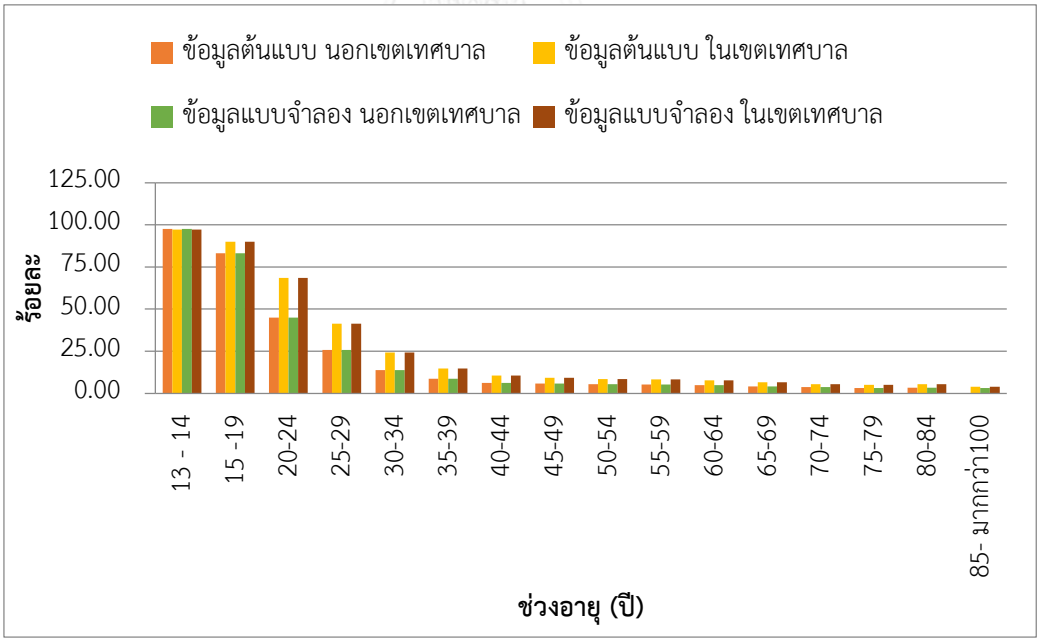


(ซ) อ.วัดโบสถ์

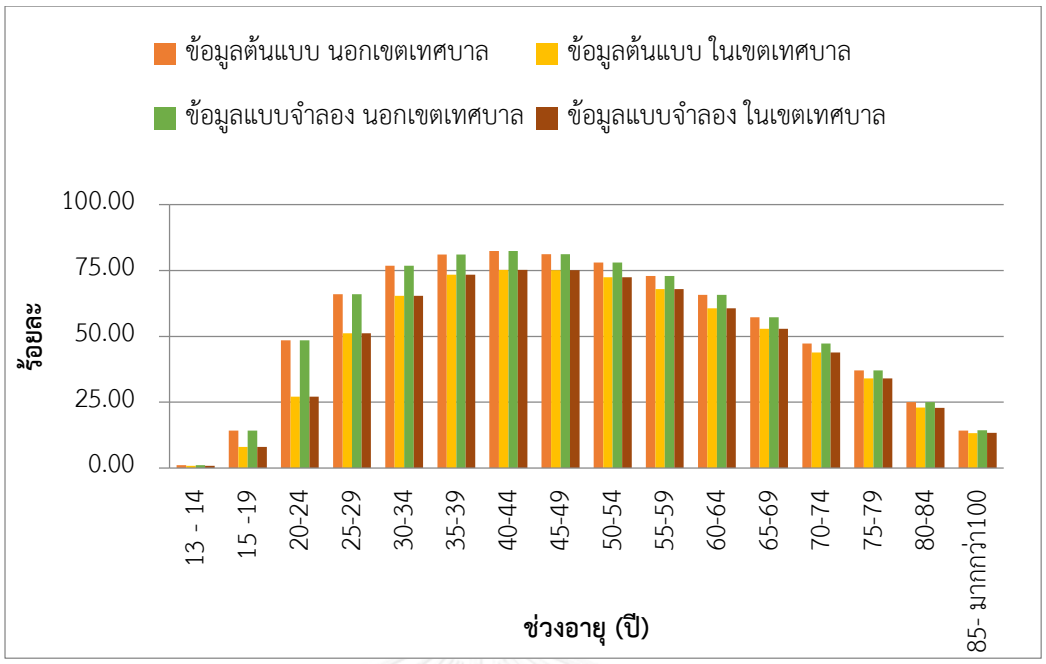


(ฌ) อ.เนินมะปราง

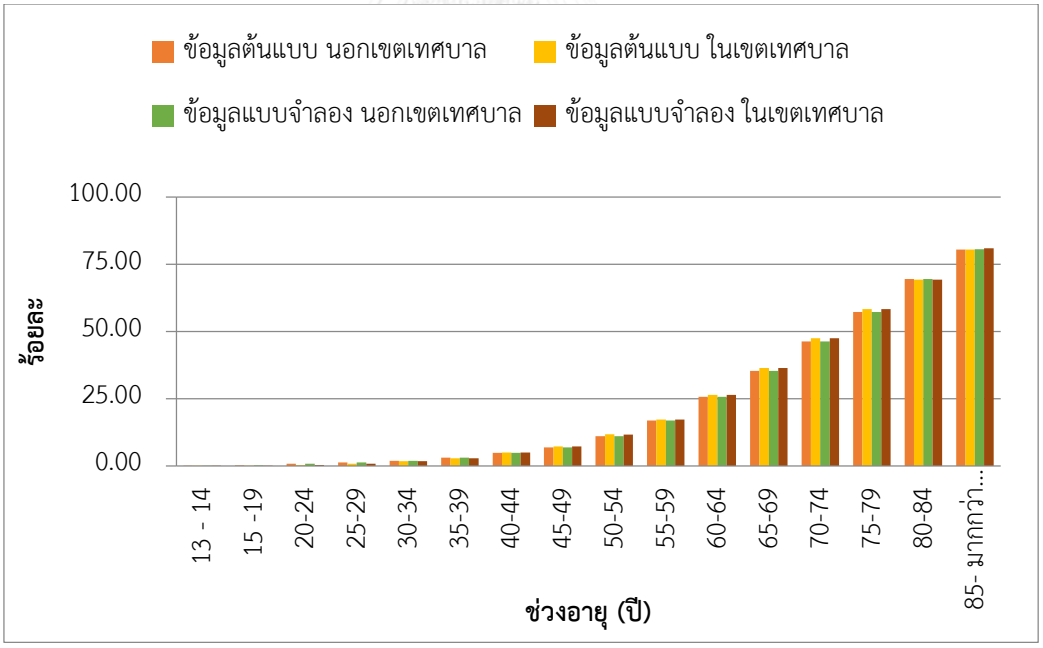
ภาพที่ ข.6 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรสของเพศหญิง



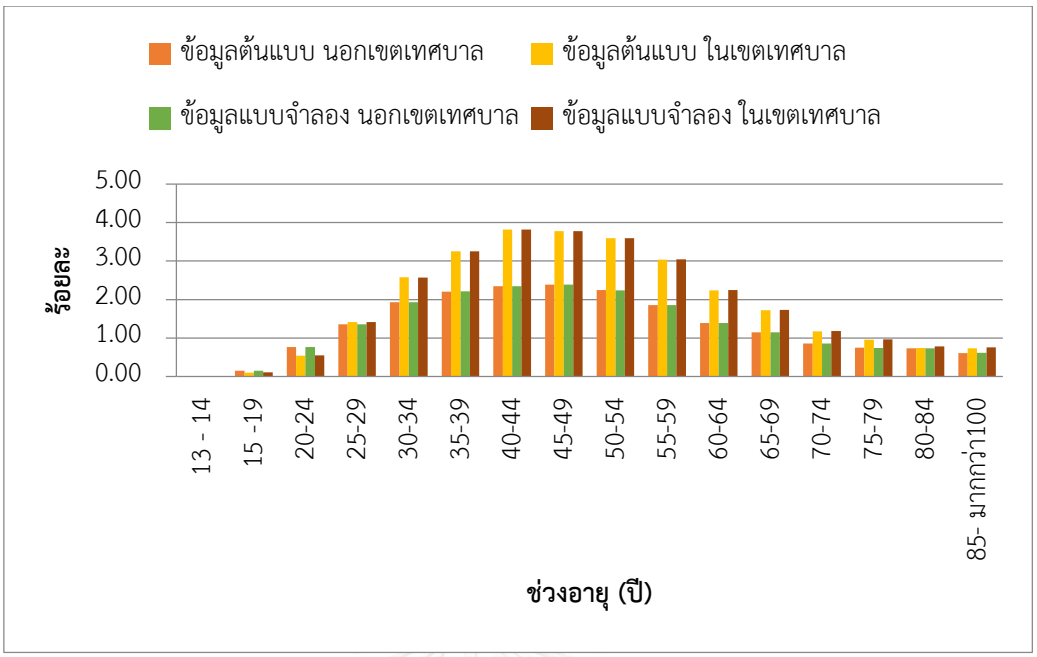
(จ) โสด


ข้อมูลต้นแบบ นอกเขตเทศบาล
ข้อมูลต้นแบบ ในเขตเทศบาล
ข้อมูลแบบจำลอง นอกเขตเทศบาล
ข้อมูลแบบจำลอง ในเขตเทศบาล
100.00
75.00
50.00
25.00
0.00
ร้อยละ
13 - 14
15 -19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85- มากกว่า100
ช่วงอายุ (ปี)

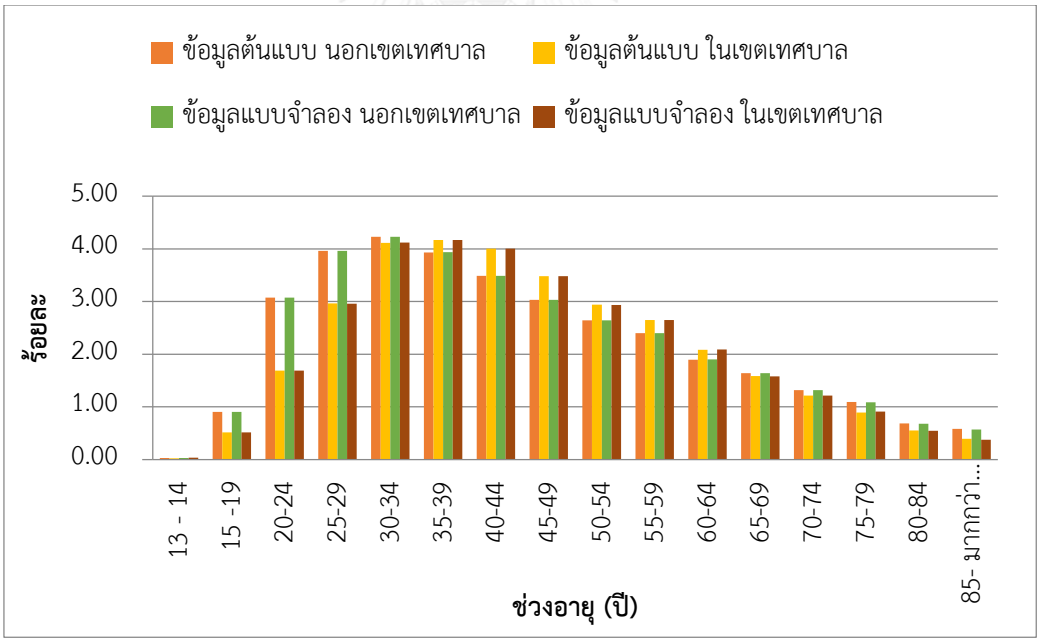
(ฉ) สมรส


ข้อมูลต้นแบบ นอกเขตเทศบาล
ข้อมูลต้นแบบ ในเขตเทศบาล
ข้อมูลแบบจำลอง นอกเขตเทศบาล
ข้อมูลแบบจำลอง ในเขตเทศบาล
100.00
75.00
50.00
25.00
0.00
ร้อยละ
13 - 14
15 -19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85- มากกว่า...
ช่วงอายุ (ปี)

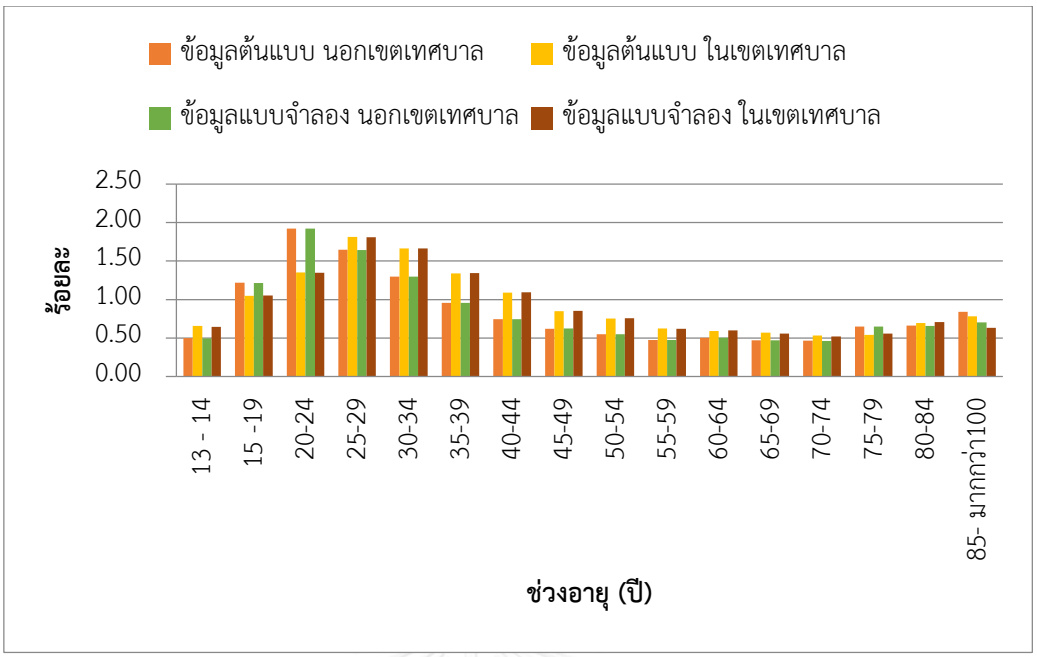
(ช) หม้าย



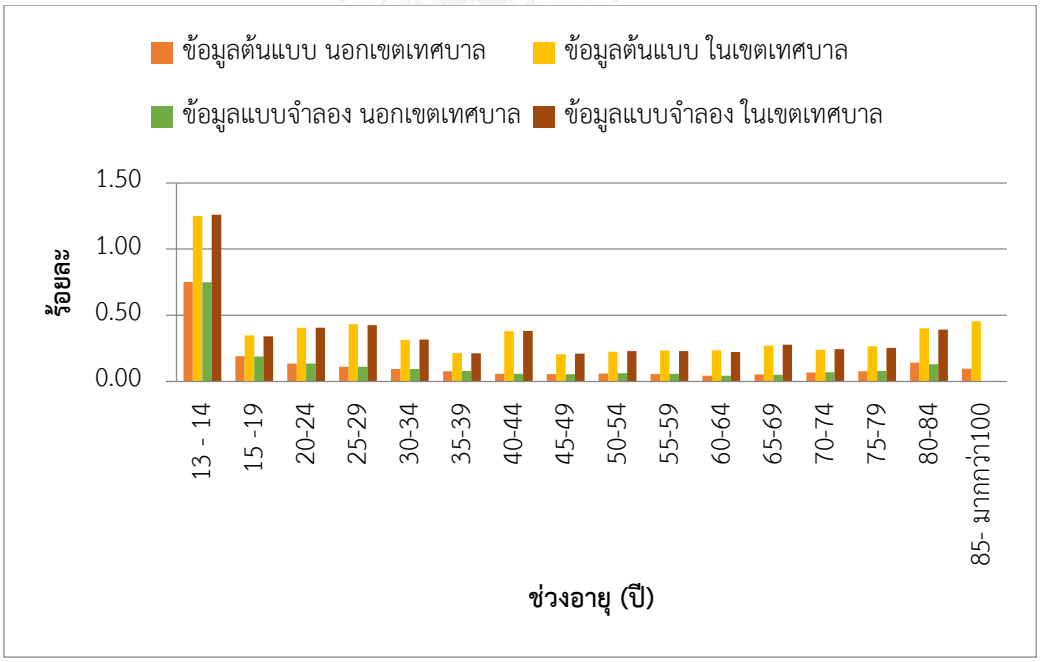
(ซ) หย่า



(ฌ) แยกกันอยู่

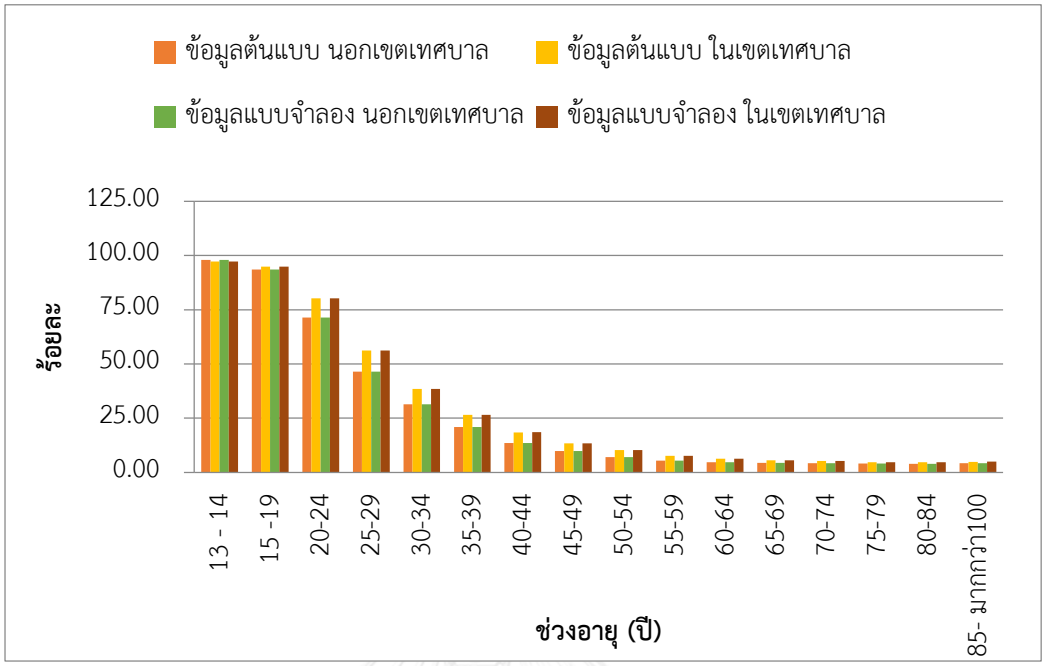


(ญ) เคยสมรสแต่ไม่ทราบสถานภาพสมรส

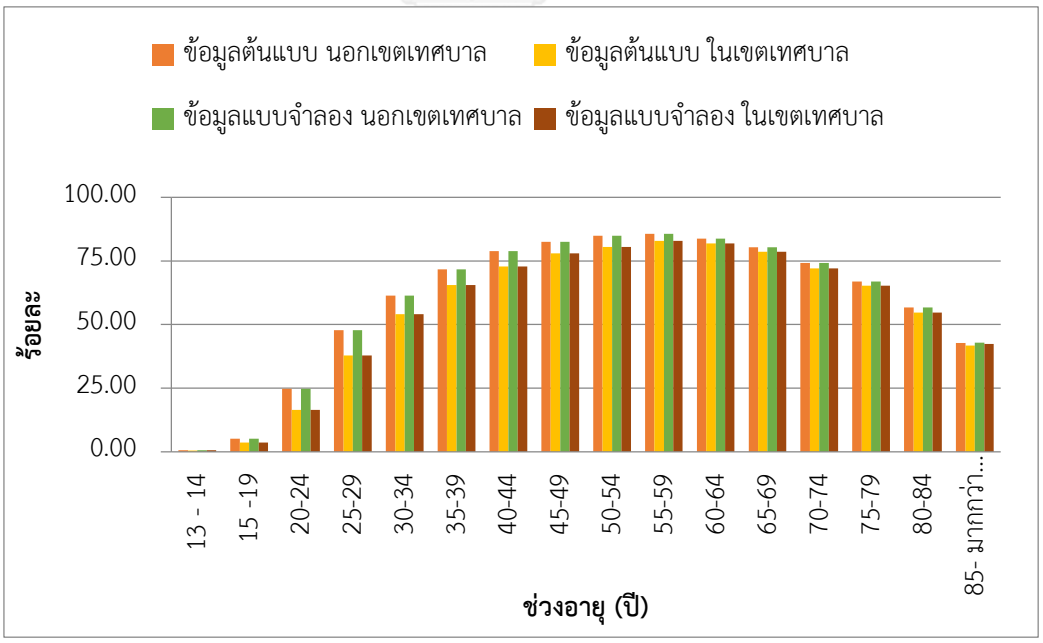


(ฎ) ไม่ทราบสถานภาพสมรส

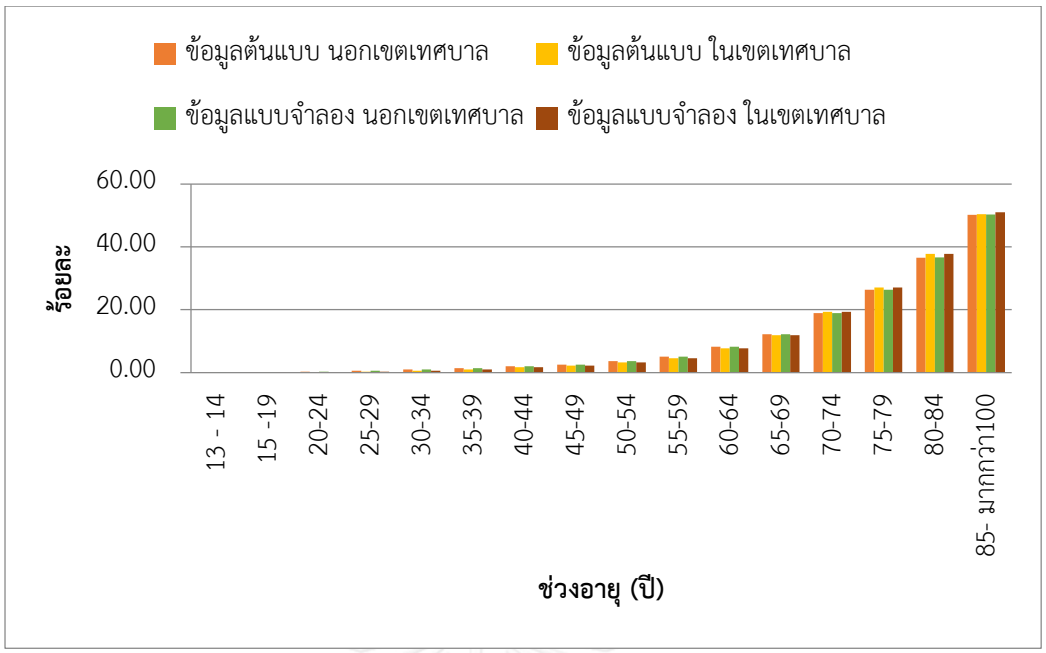
ภาพที่ ข.7 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลสถานภาพสมรสของเพศชาย



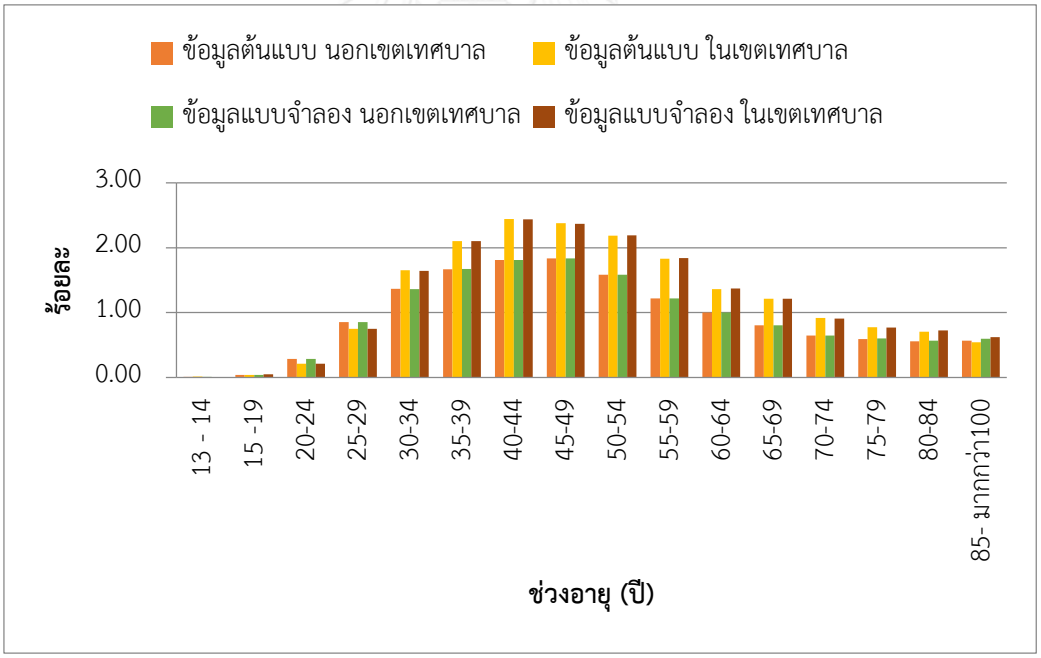
(ก) โสด



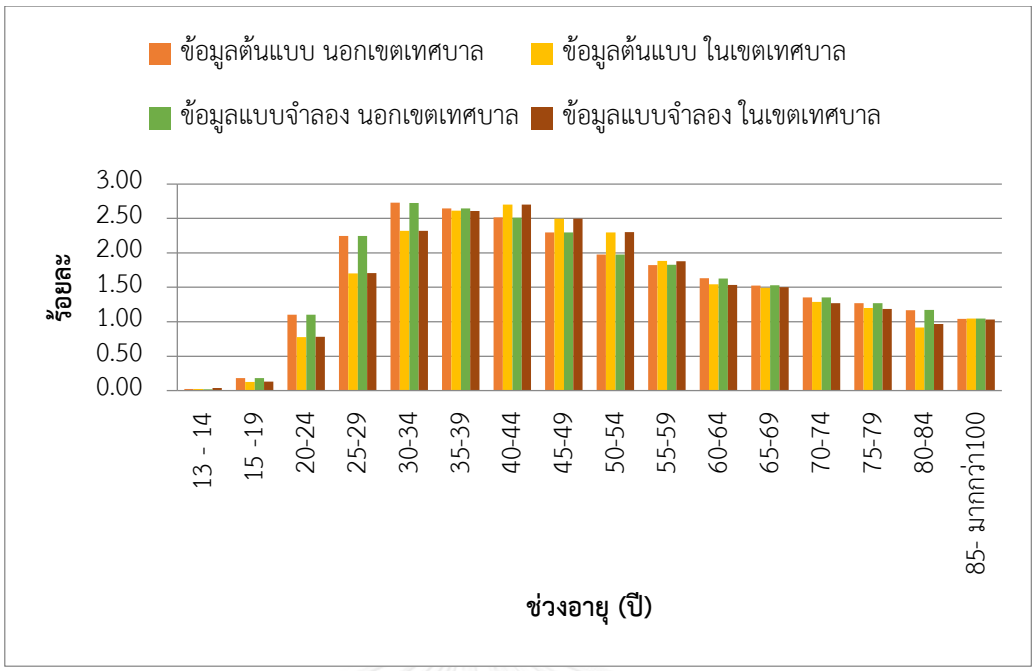
(ข) สมรส


ข้อมูลต้นแบบ นอกเขตเทศบาล
ข้อมูลต้นแบบ ในเขตเทศบาล
ข้อมูลแบบจำลอง นอกเขตเทศบาล
ข้อมูลแบบจำลอง ในเขตเทศบาล
60.00
40.00
20.00
0.00
ร้อยละ
13 - 14
15 -19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85- มากกว่า100
ช่วงอายุ (ปี)

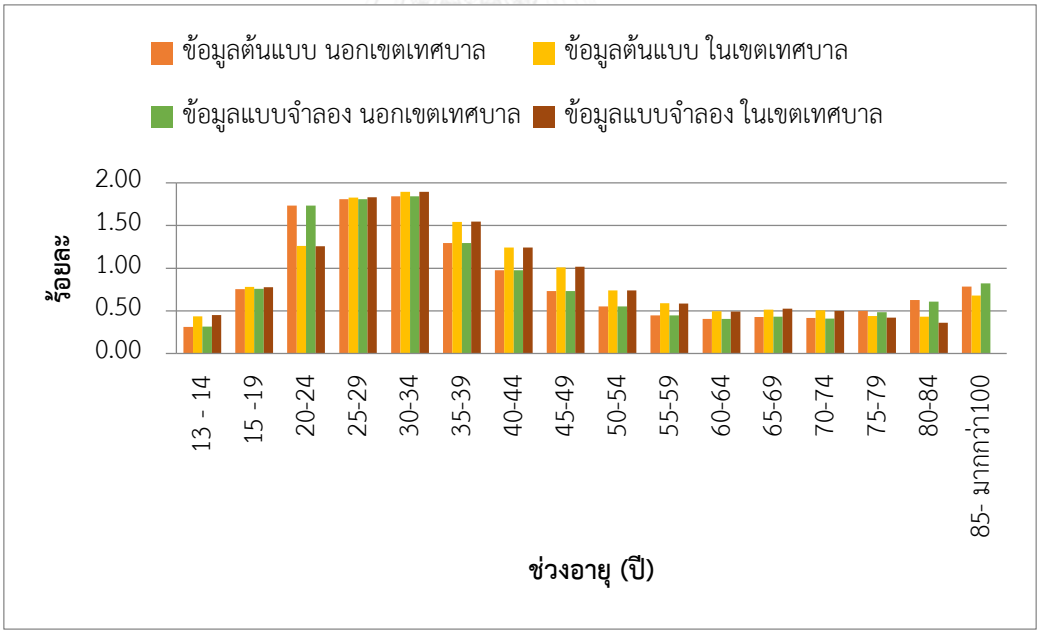
(ค) หม้าย


ข้อมูลต้นแบบ นอกเขตเทศบาล
ข้อมูลต้นแบบ ในเขตเทศบาล
ข้อมูลแบบจำลอง นอกเขตเทศบาล
ข้อมูลแบบจำลอง ในเขตเทศบาล
3.00
2.00
1.00
0.00
ร้อยละ
13 - 14
15 -19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85- มากกว่า100
ช่วงอายุ (ปี)

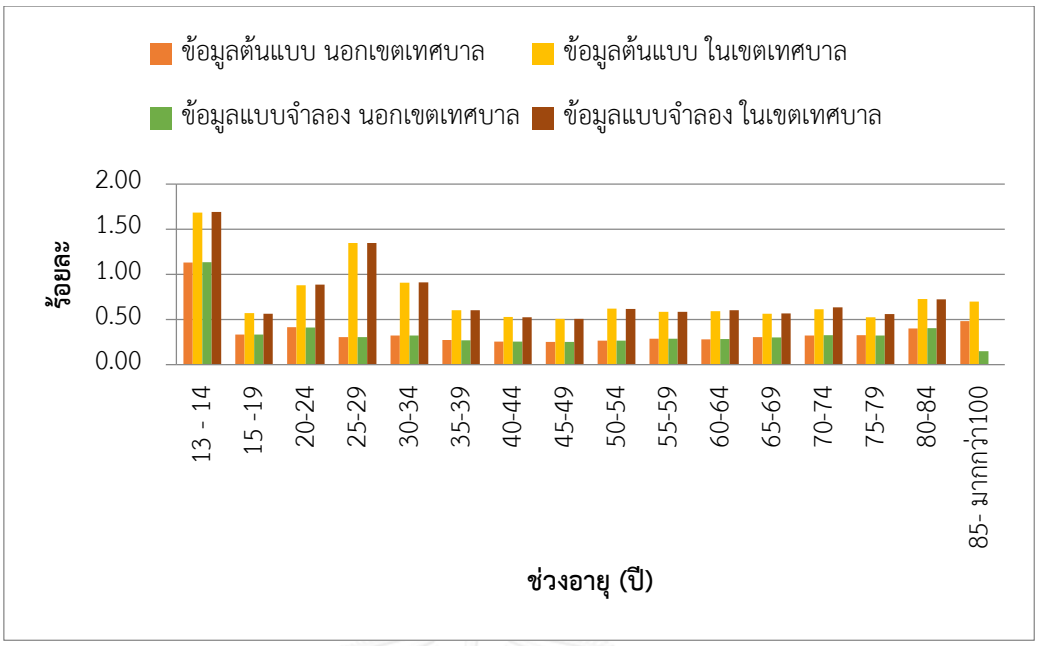
(ง) หย่า



(จ) แยกกันอยู่

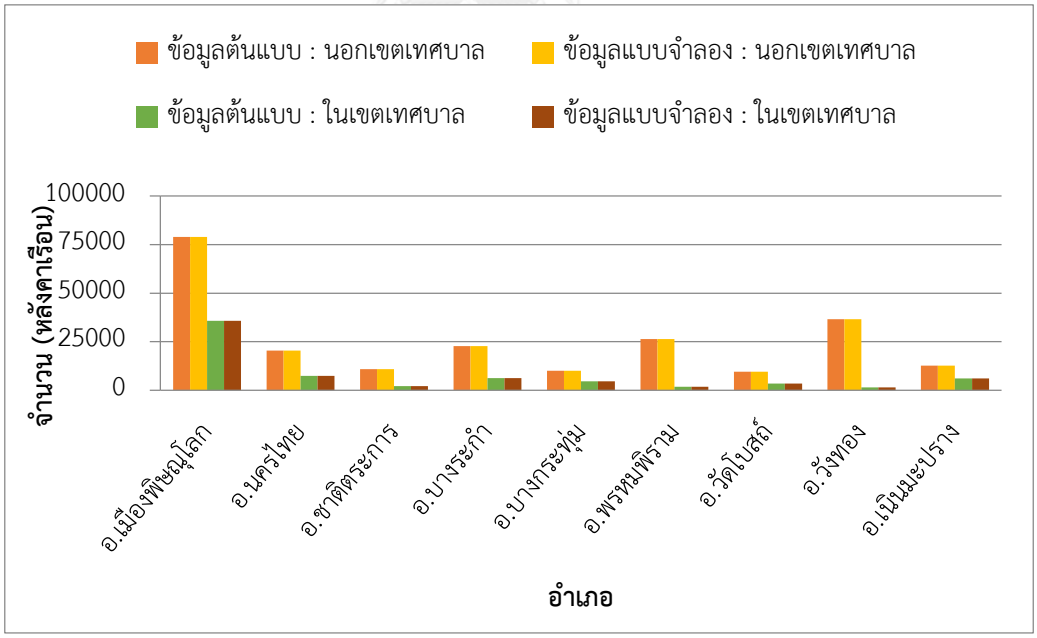


(ฉ) เคยสมรสแต่ไม่ทราบสถานภาพสมรส

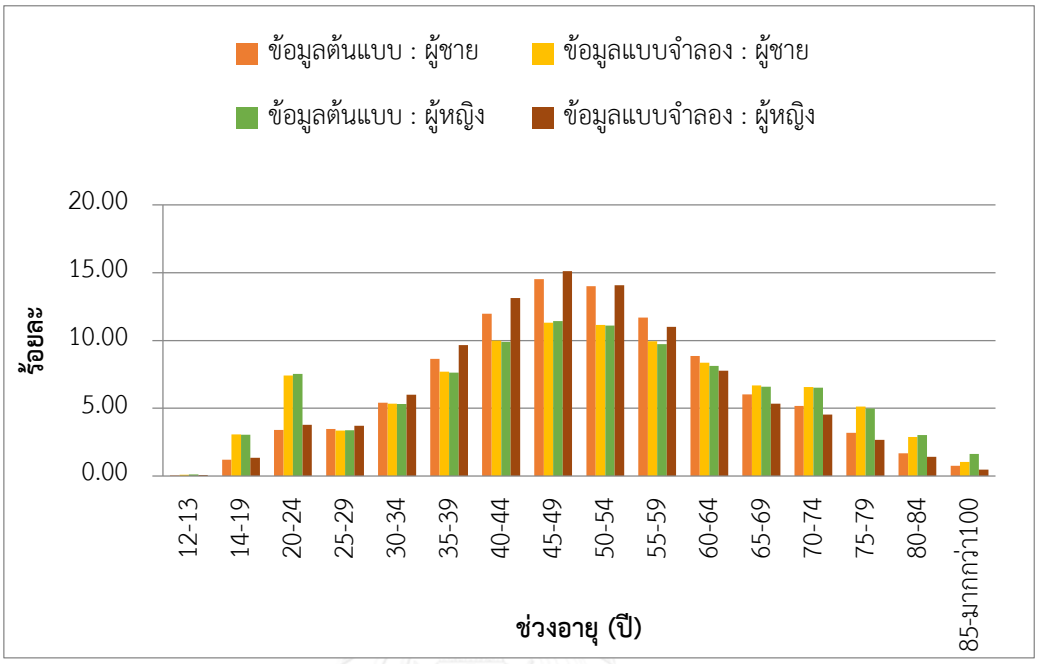


(ช) ไม่ทราบสถานภาพสมรส

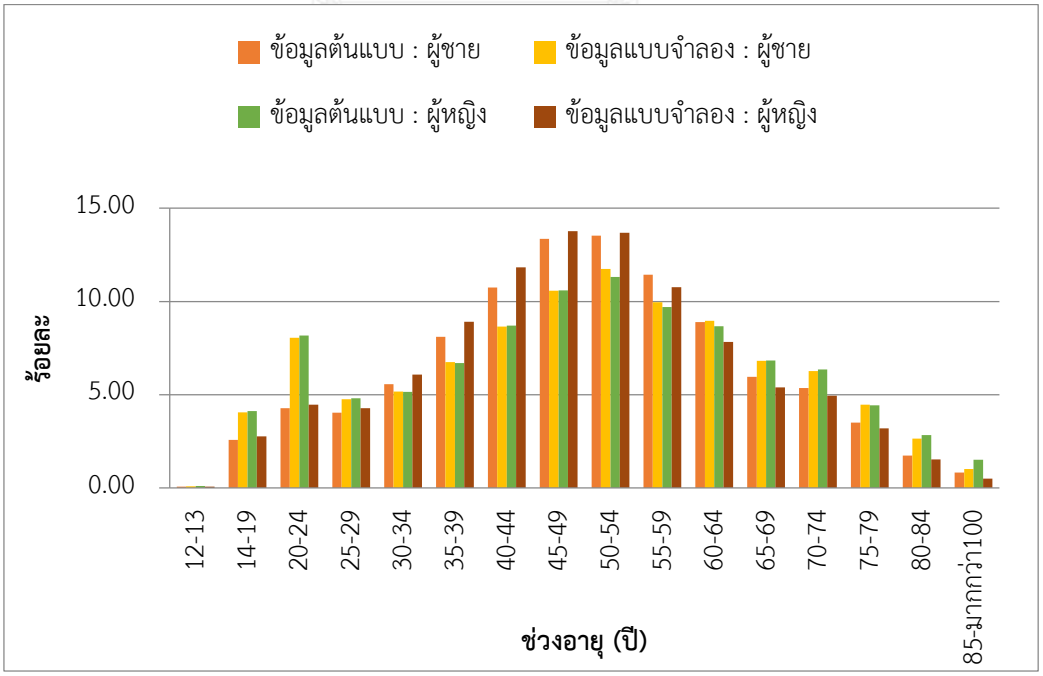
ภาพที่ ข.8 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลจำนวนครัวเรือน



ภาพที่ ข.9 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลหัวหน้าครัวเรือนจำแนกตามเพศและอายุ

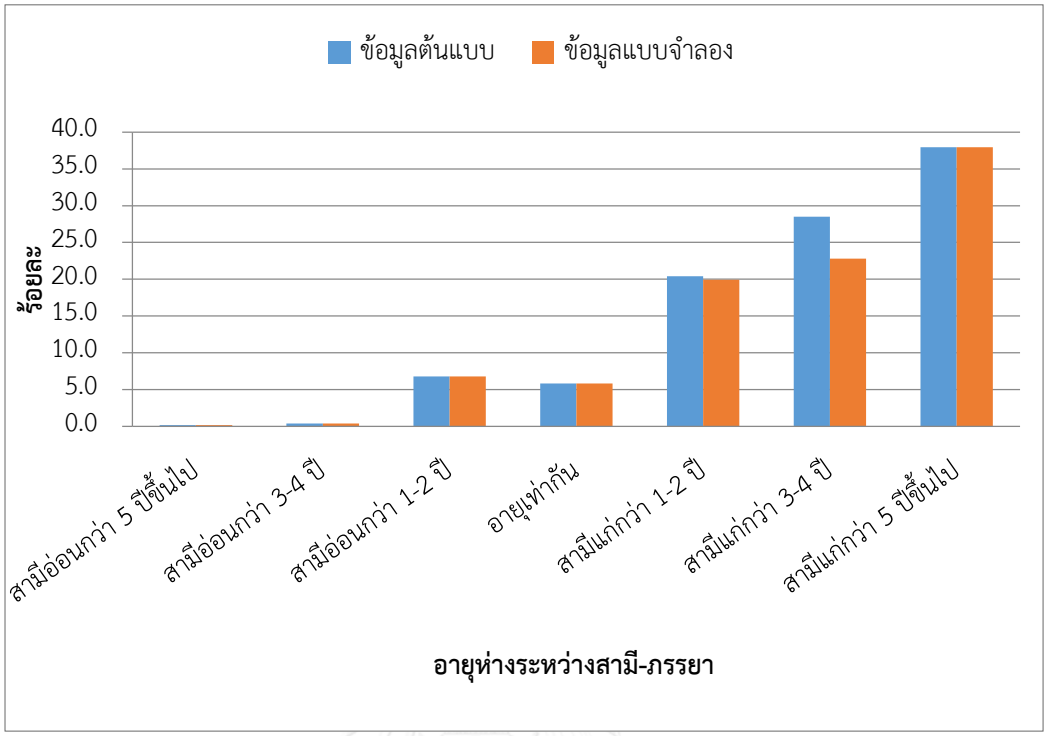


(ค) นอกเขตเทศบาล

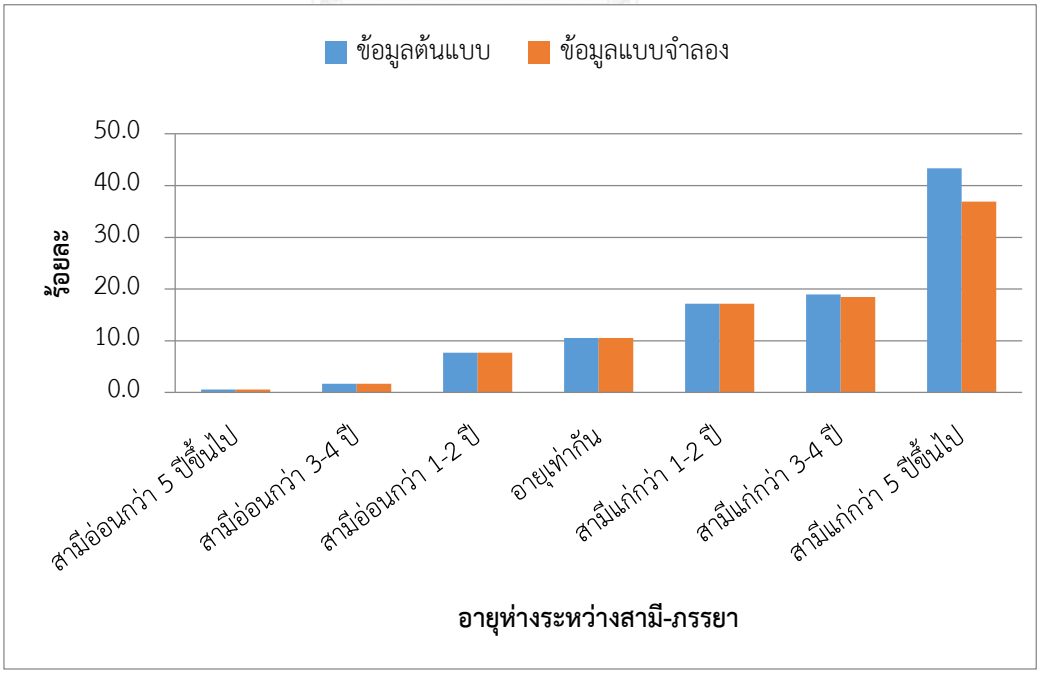


(ง) ในเขตเทศบาล

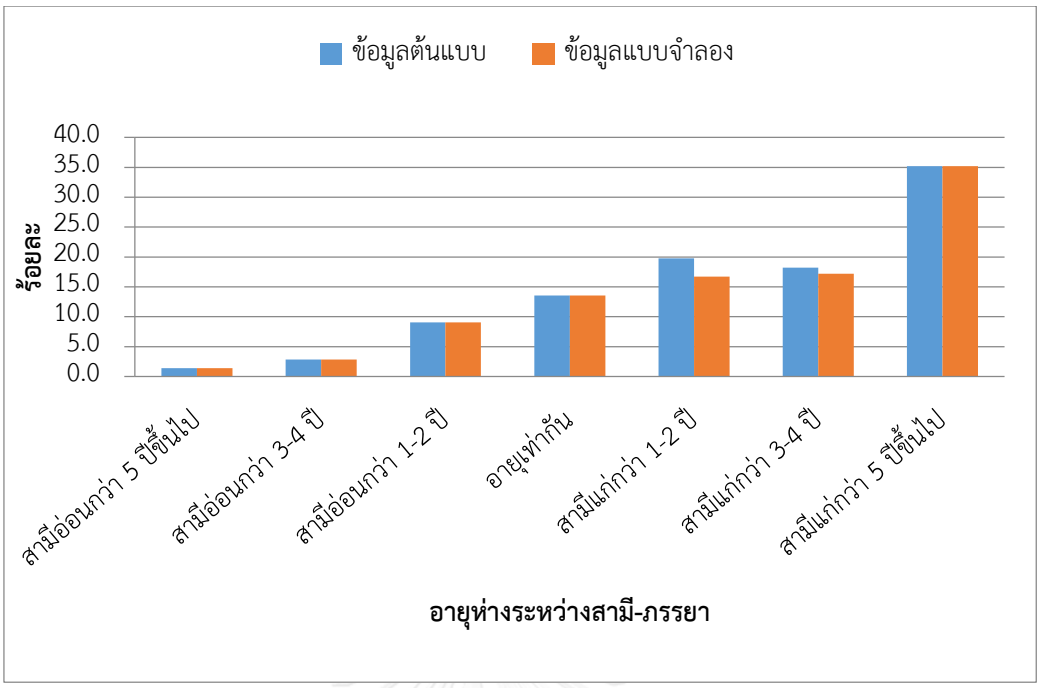
ภาพที่ ข.10 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุห่างระหว่างสามี-ภรรยา



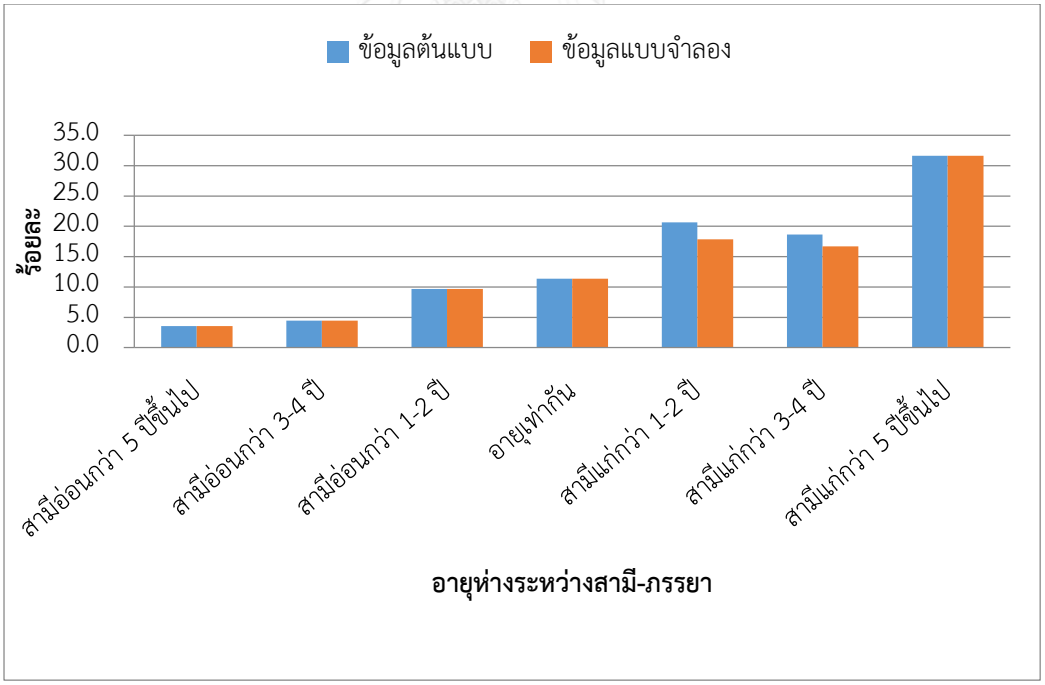
(ก) ผู้หญิงสมรส อายุ 15- 19 ปี



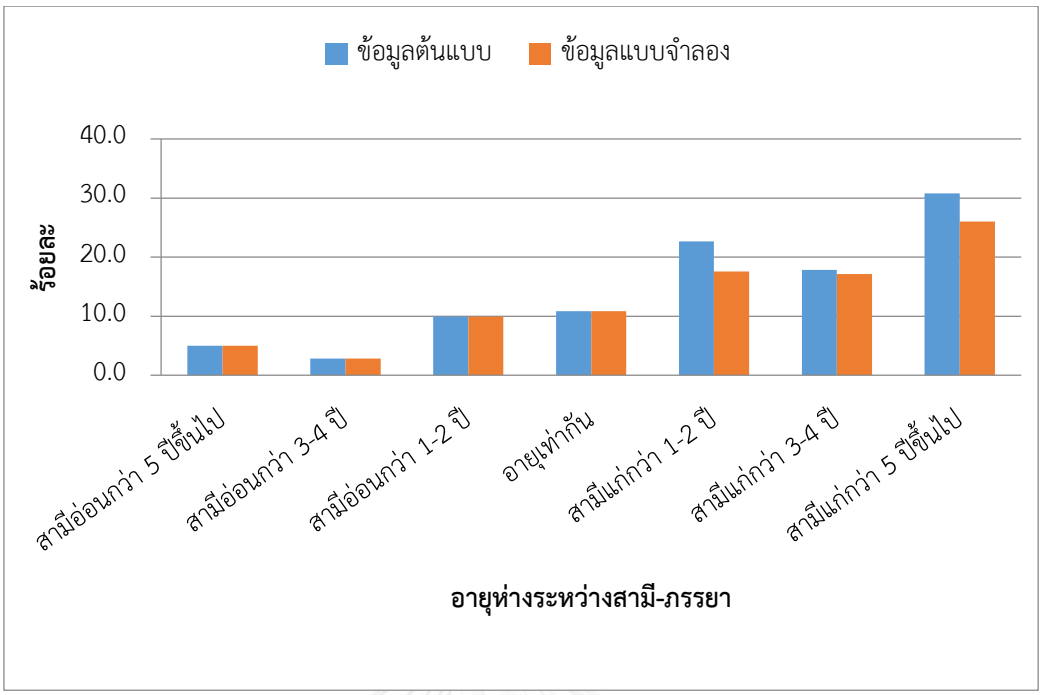
(ข) ผู้หญิงสมรส อายุ 20-24 ปี



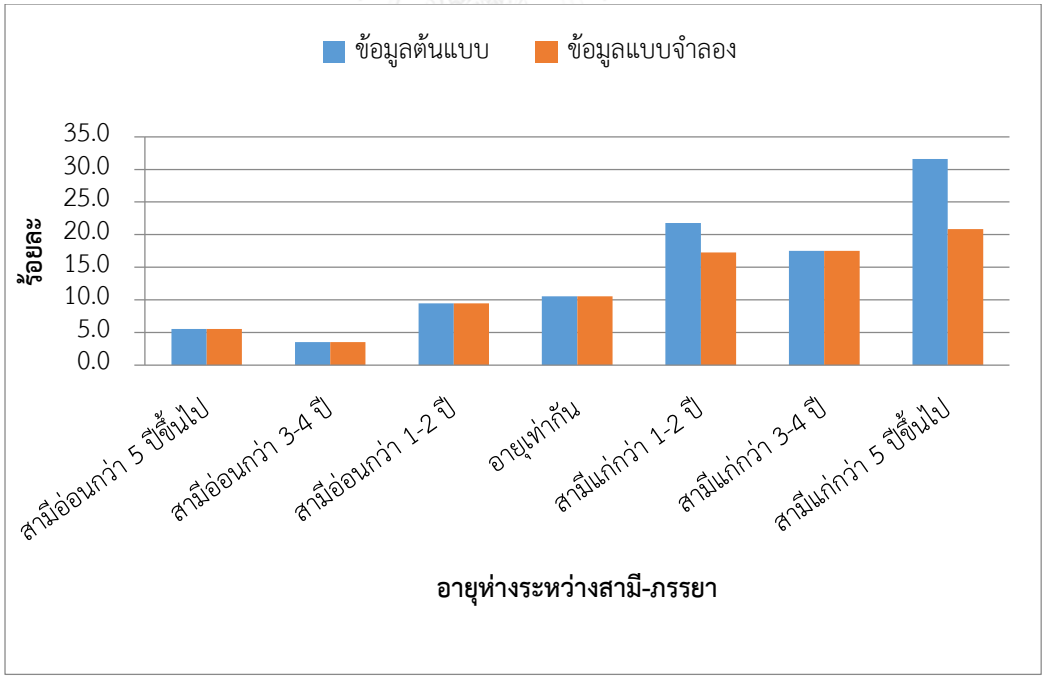
(ค) ผู้หญิงสมรส อายุ 25-29 ปี



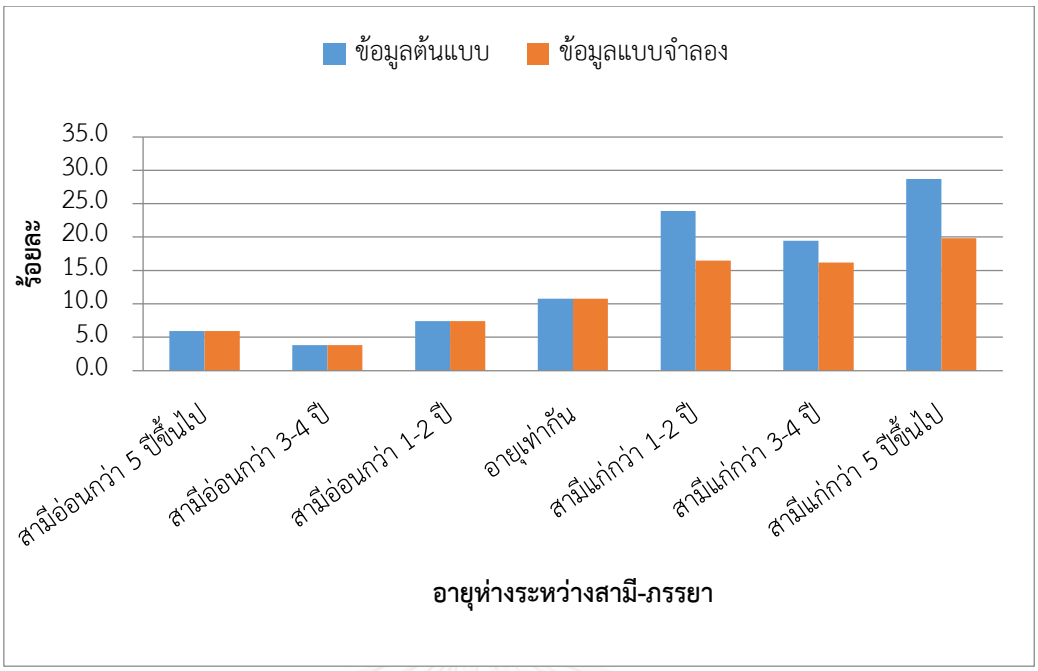
(ง) ผู้หญิงสมรส อายุ 30-34 ปี



(จ) ผู้หญิงสมรส 35-39 ปี

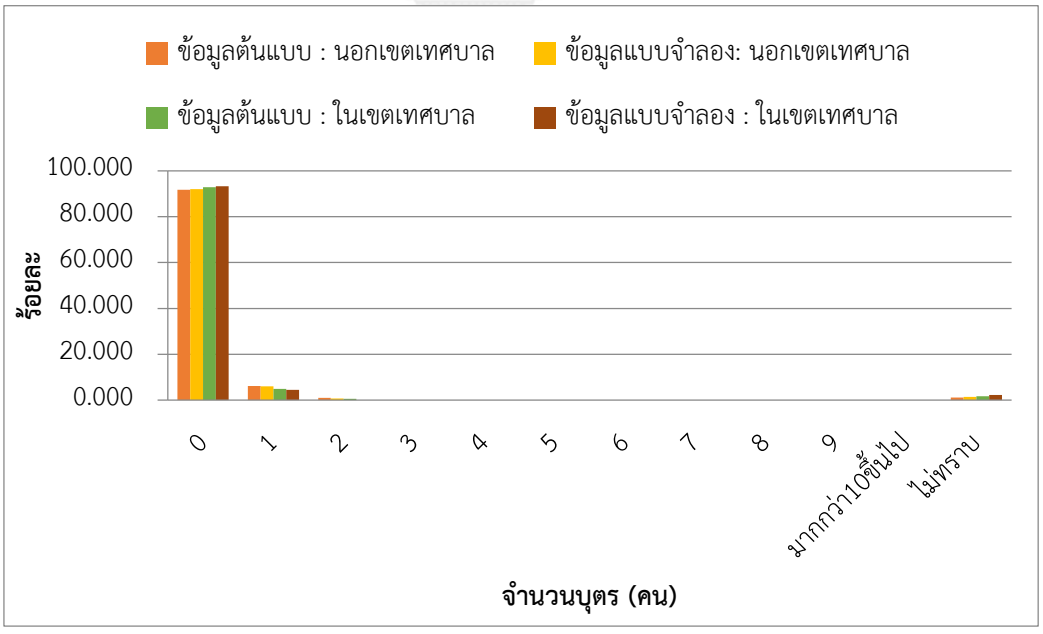


(ฉ) ผู้หญิงสมรส 40-44 ปี

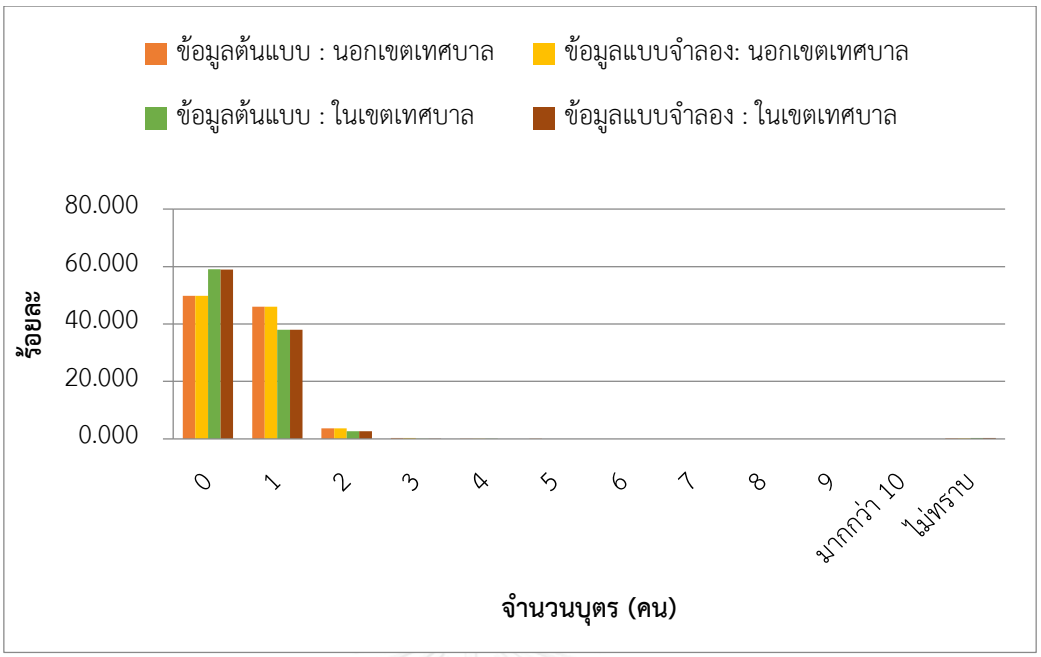


(ช) ผู้หญิงสมรส 45-49 ปี

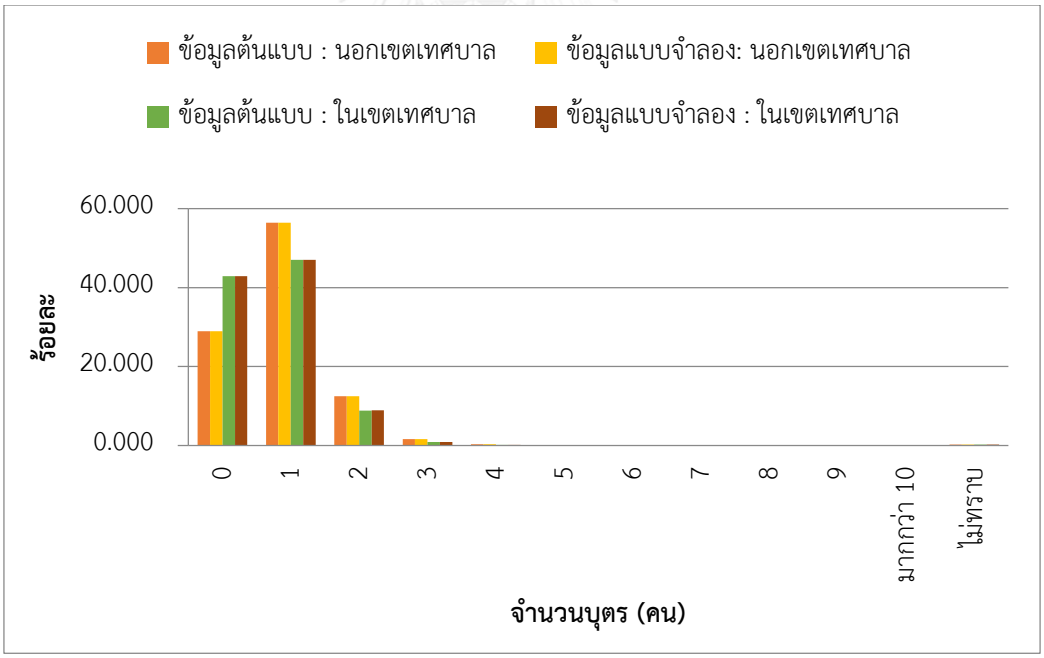
ภาพที่ ข.11 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลจำนวนบุตร



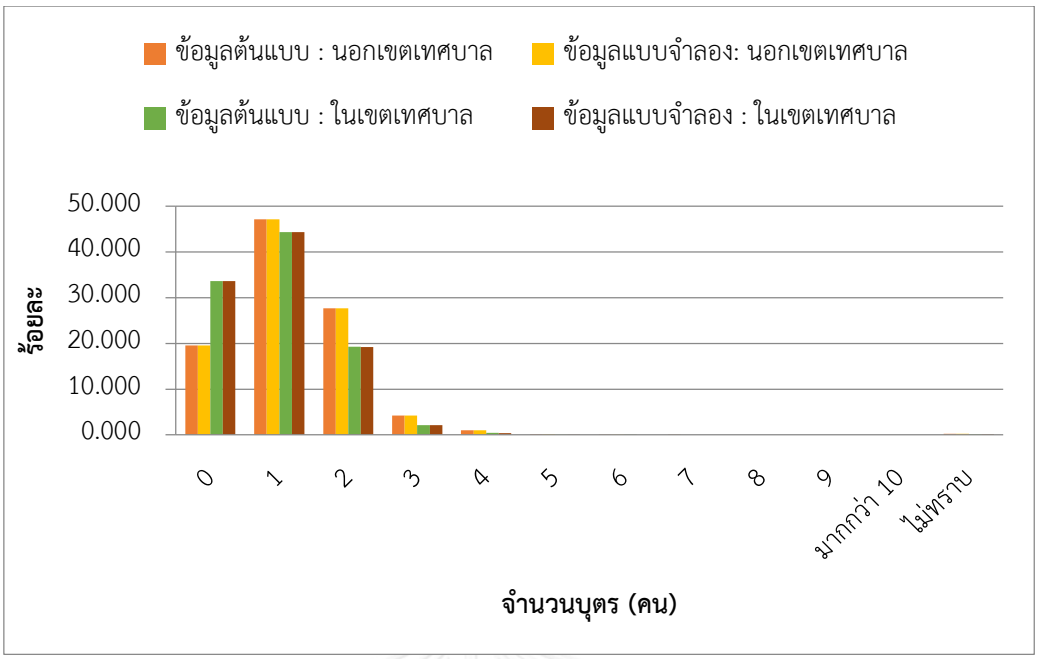
(ก) ผู้หญิงที่เคยสมรส อายุ 13-14 ปี



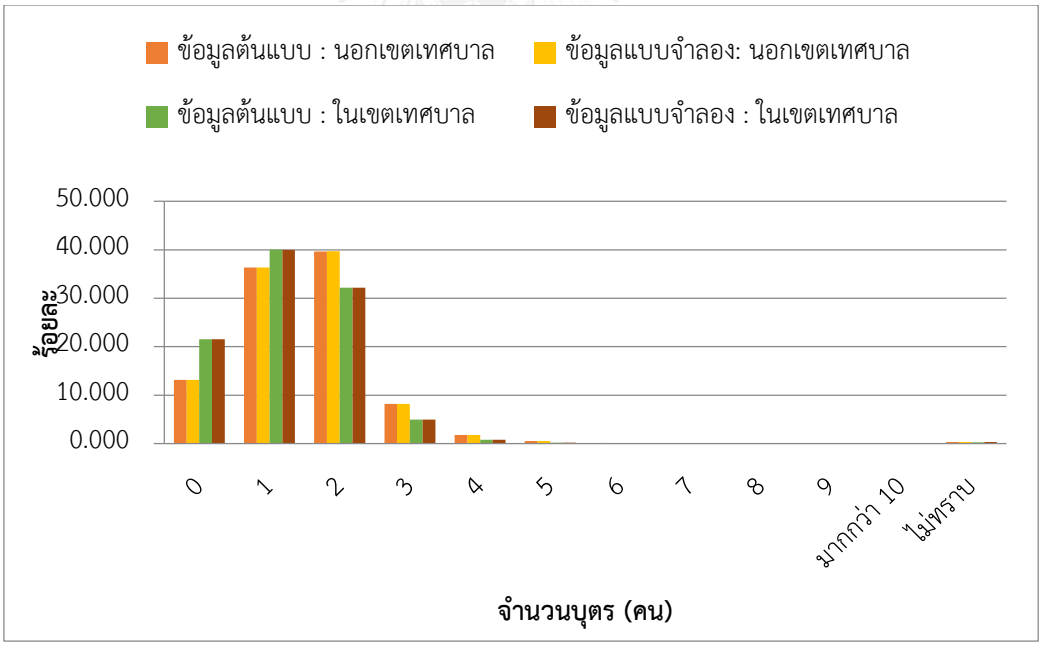
(ข) ผู้หญิงที่เคยสมรส อายุ 15-19 ปี



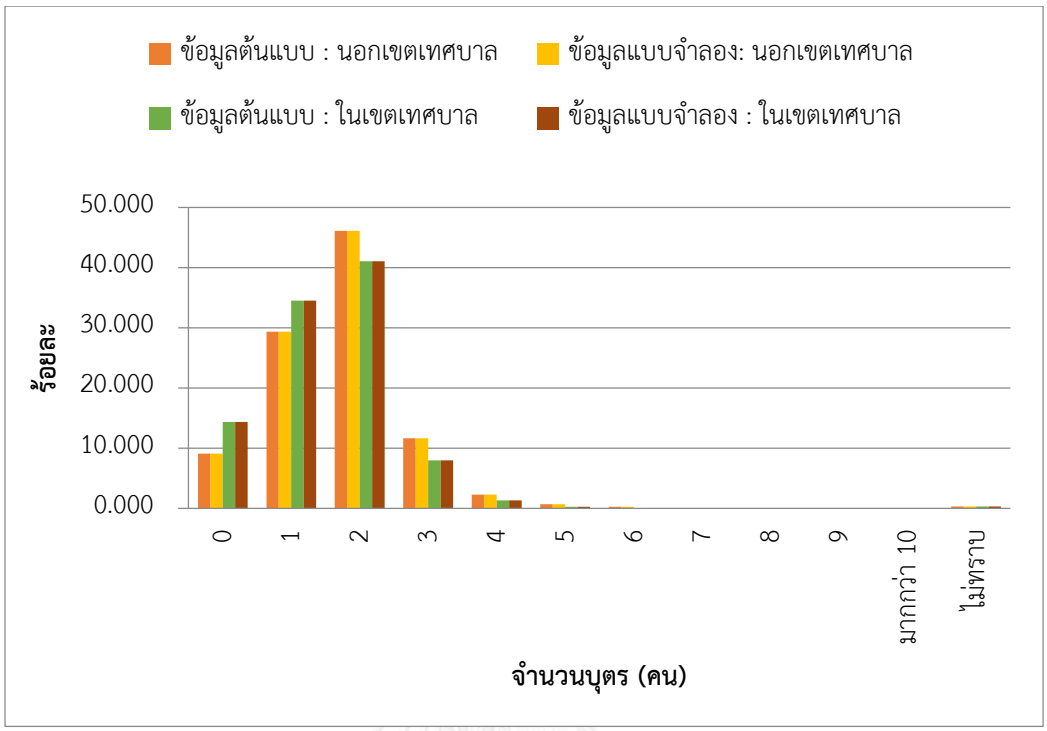
(ค) ผู้หญิงที่เคยสมรส อายุ 20-24 ปี



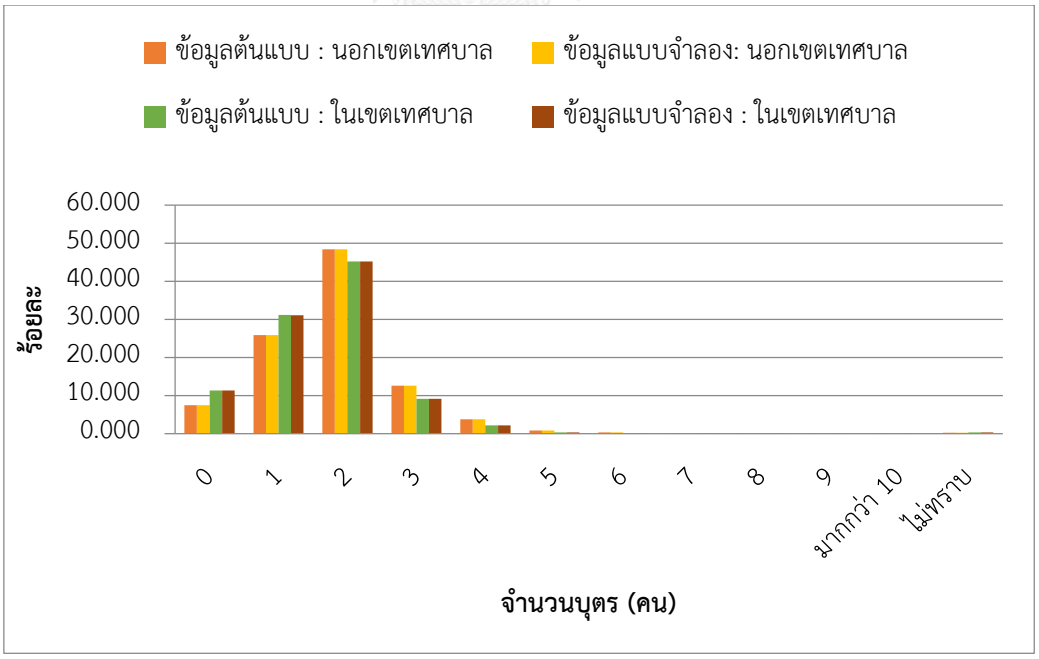
(ง) ผู้หญิงที่เคยสมรส อายุ 25-29 ปี



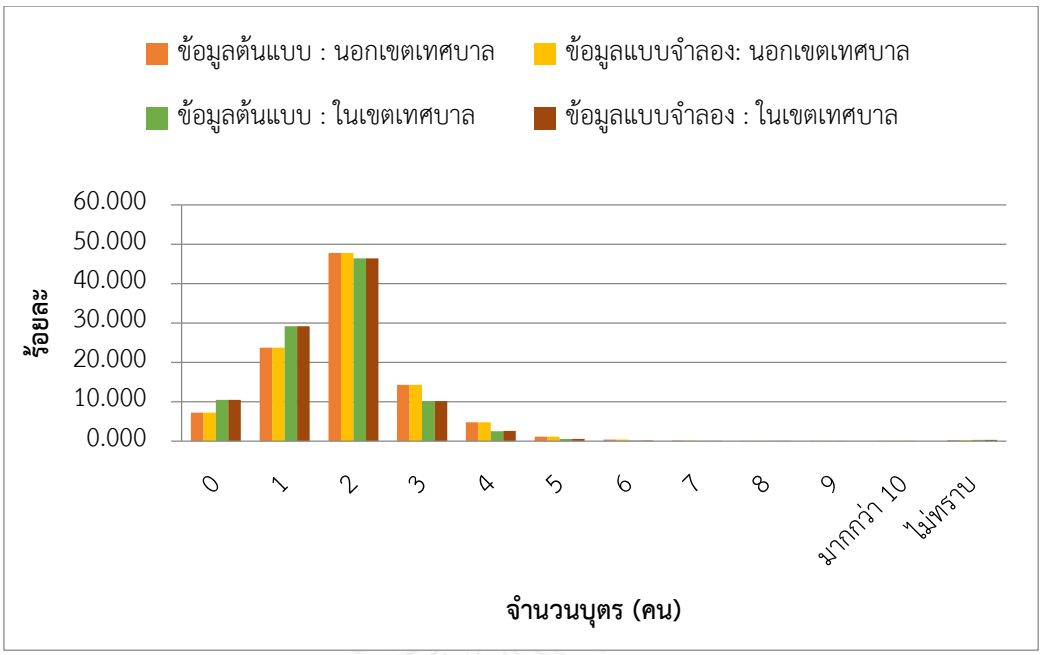
(จ) ผู้หญิงที่เคยสมรส อายุ 30-34 ปี



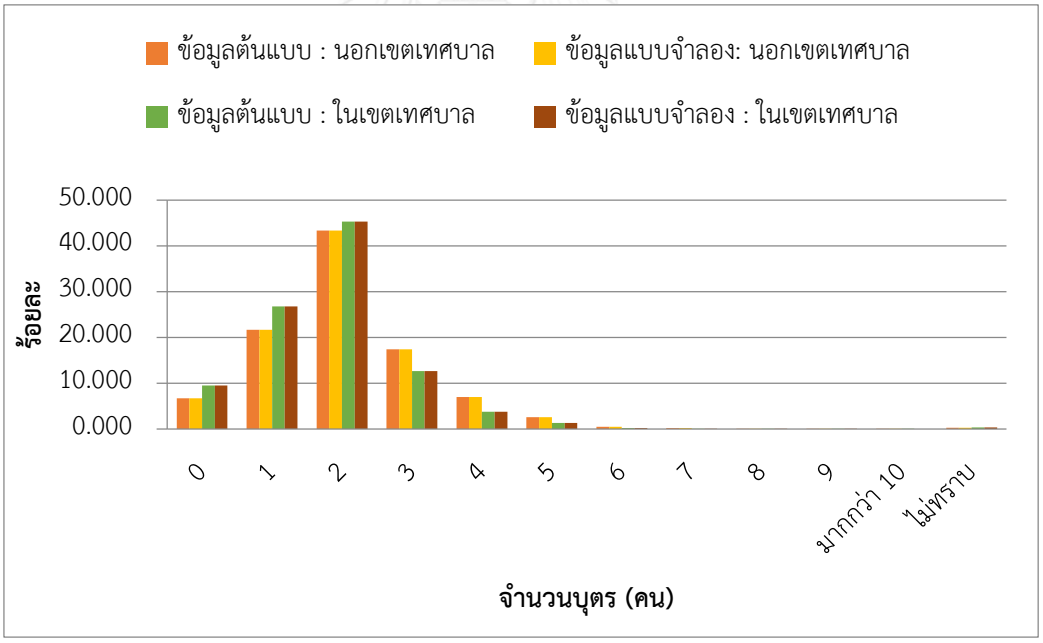
(ฉ) ผู้หญิงที่เคยสมรส อายุ 35-39 ปี



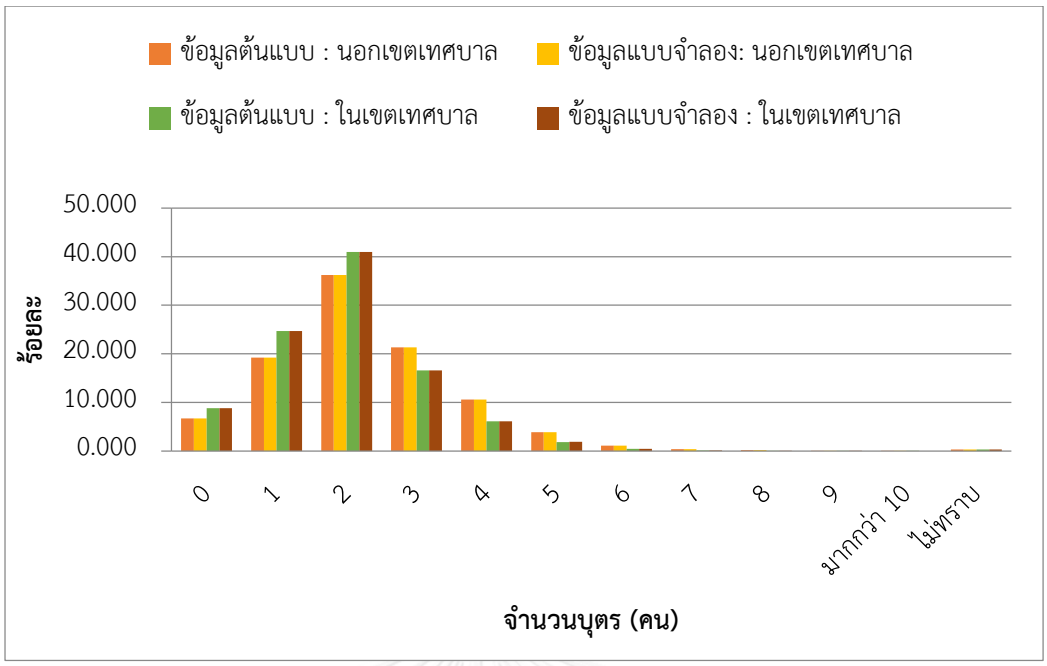
(ช) ผู้หญิงที่เคยสมรส อายุ 40-44 ปี



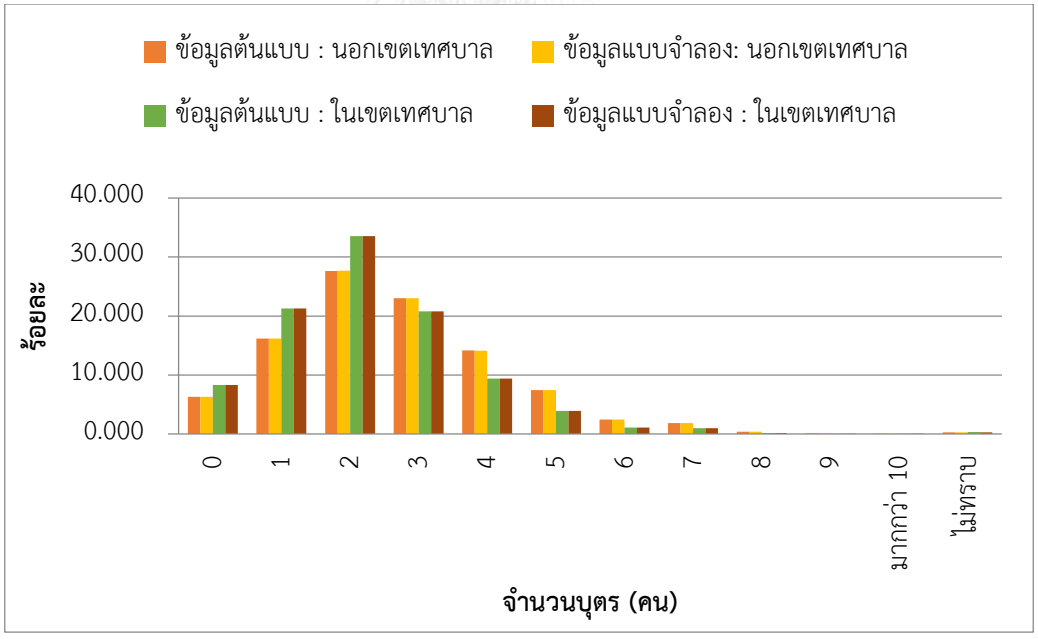
(ซ) ผู้หญิงที่เคยสมรส อายุ 45-49 ปี



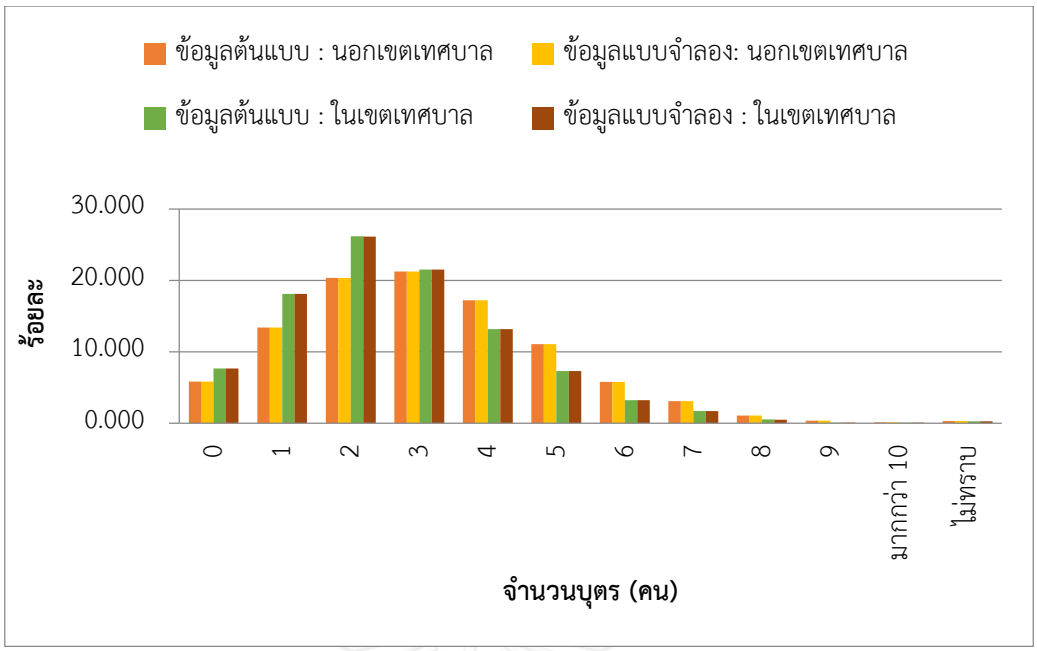
(ฌ) ผู้หญิงที่เคยสมรส อายุ 50-54 ปี



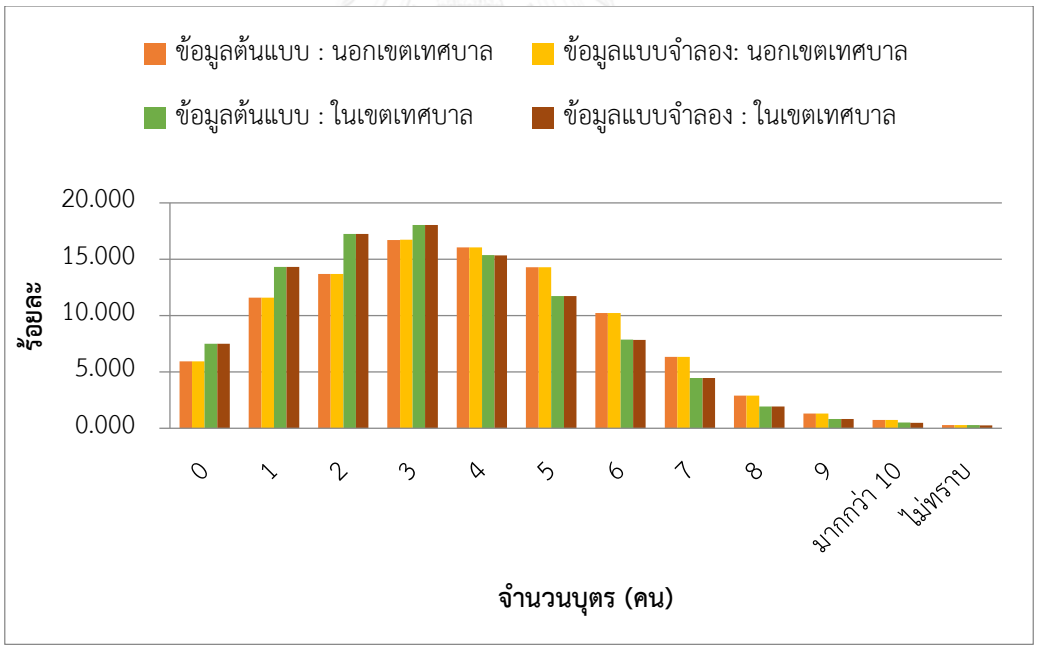
(ญ) ผู้หญิงที่เคยสมรส อายุ 55-59 ปี



(ฎ) ผู้หญิงที่เคยสมรส อายุ 60-64 ปี

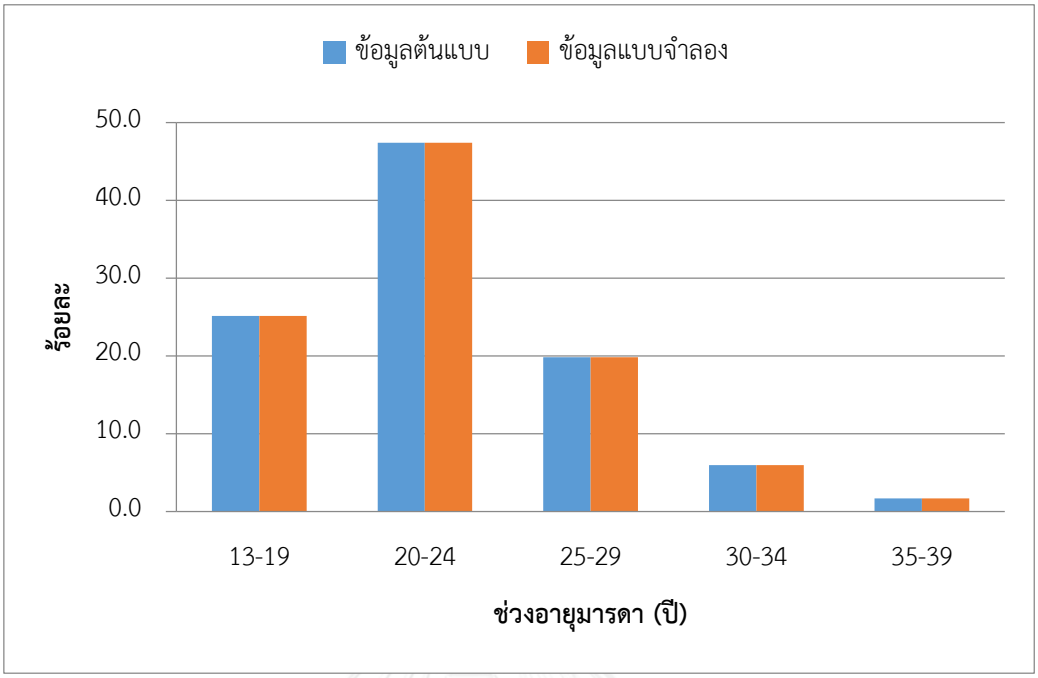


(ฎ) ผู้หญิงที่เคยสมรส อายุ 65-69 ปี

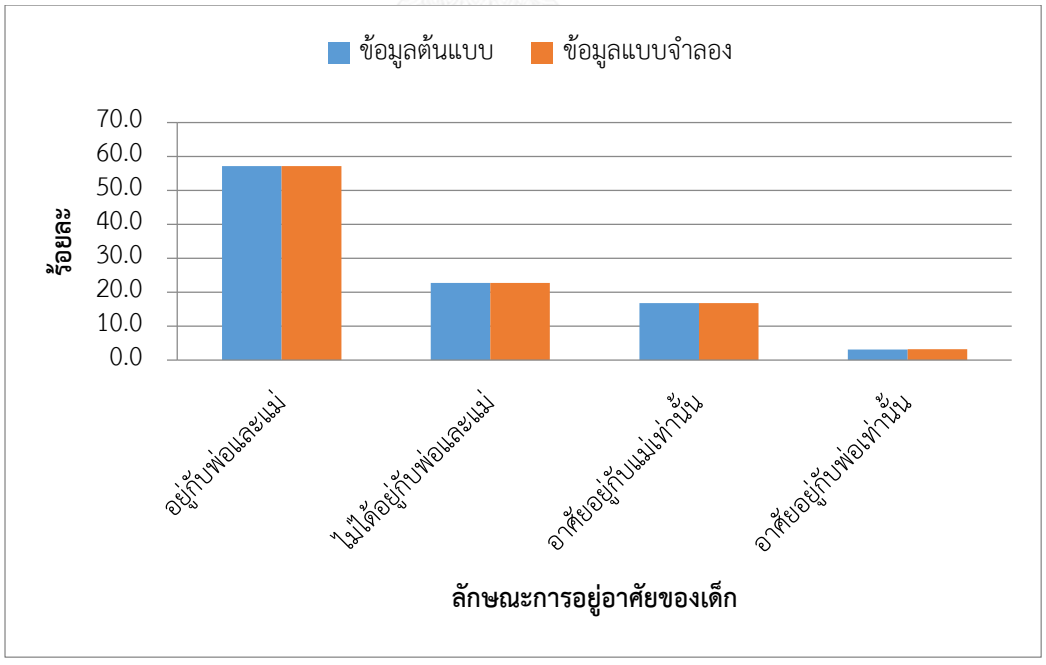


(ฐ) ผู้หญิงที่เคยสมรส อายุ 70 ปีขึ้นไป

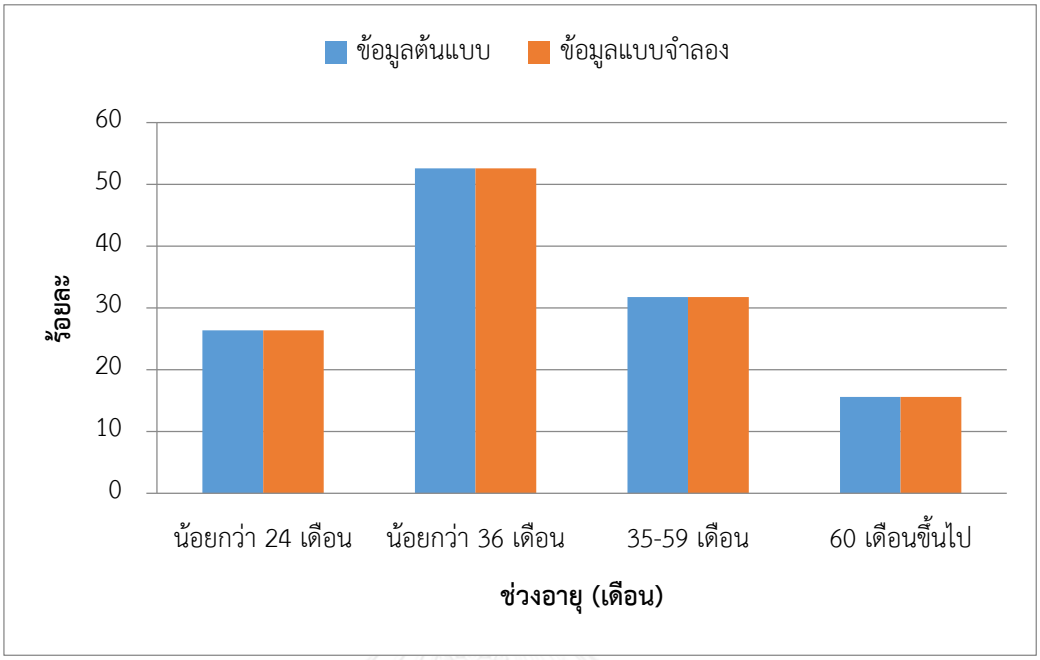
ภาพที่ ข.12 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุเมื่อมีบุตรคนแรก



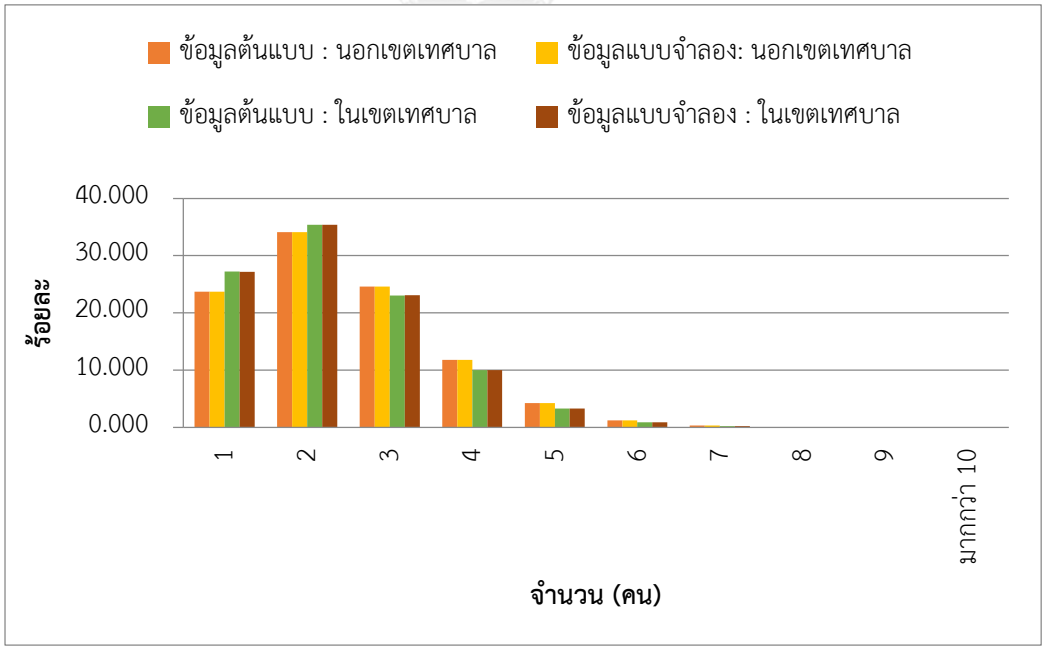
ภาพที่ ข.13 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลการอยู่อาศัยของเด็กอายุ 0-17 ปี



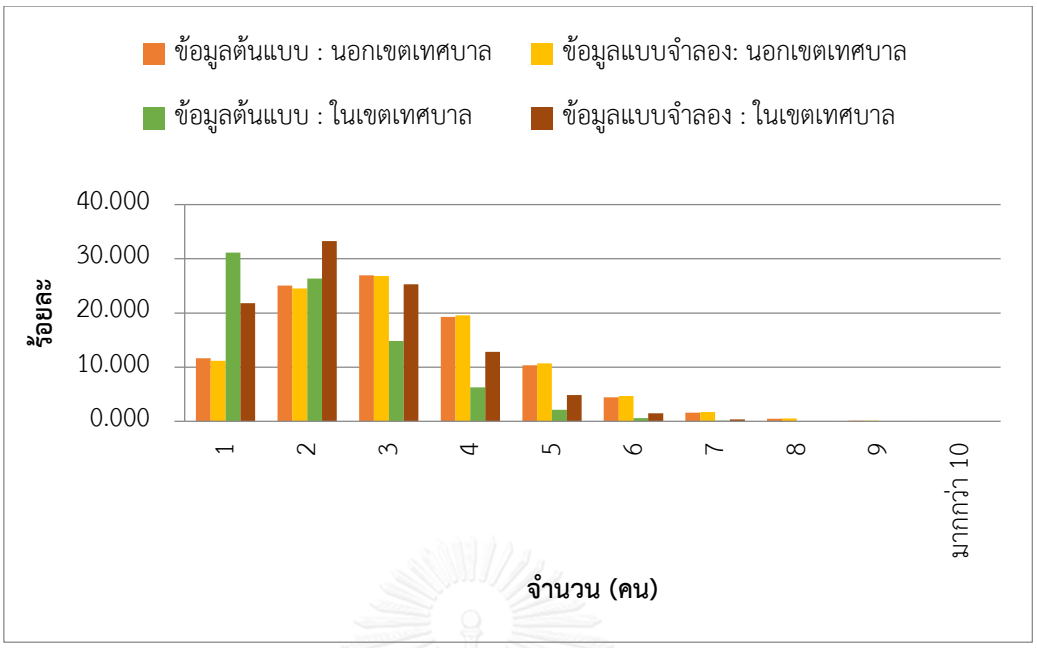
ภาพที่ ข.14 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลอายุห่างระหว่างบุตร



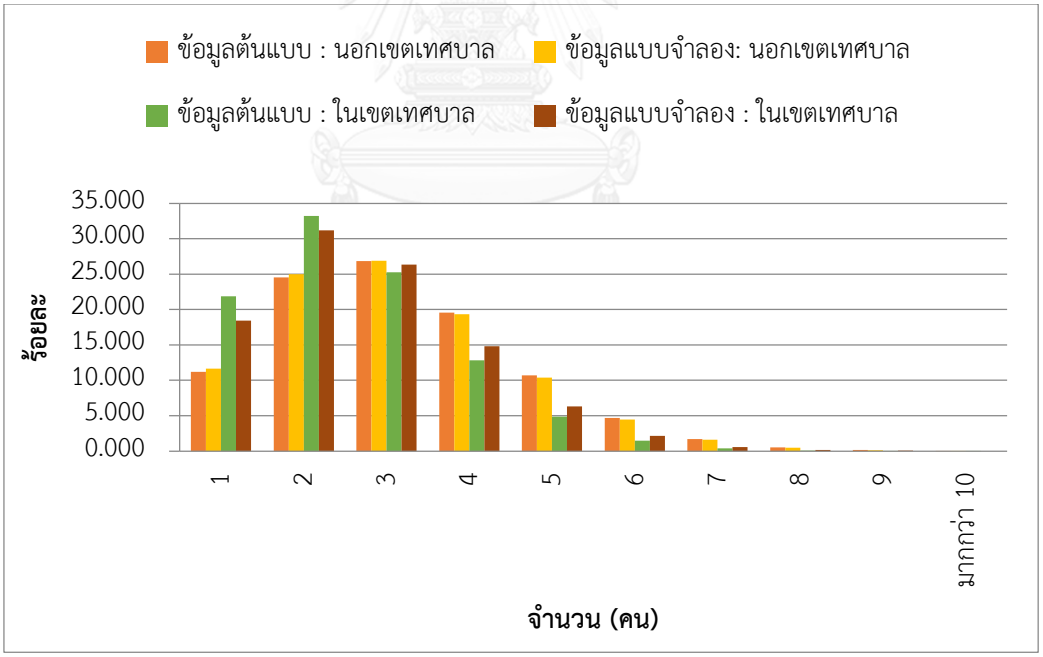
ภาพที่ ข.15 กราฟการเปรียบเทียบคุณลักษณะของข้อมูลขนาดครัวเรือน



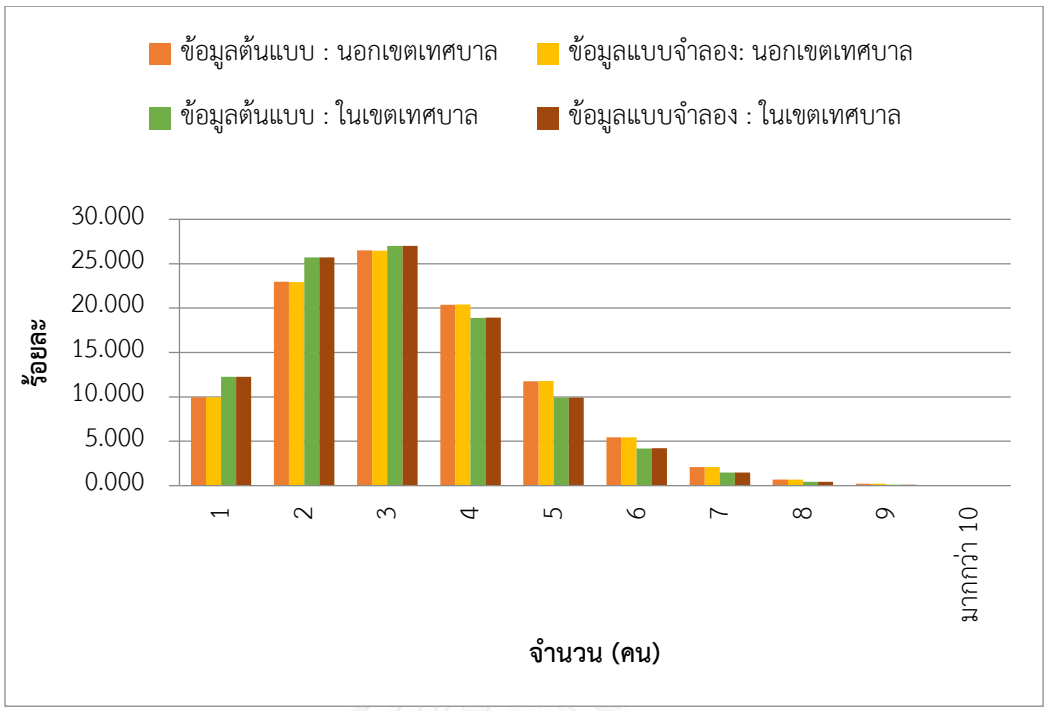
(ก) อ.เมืองพิษณุโลก



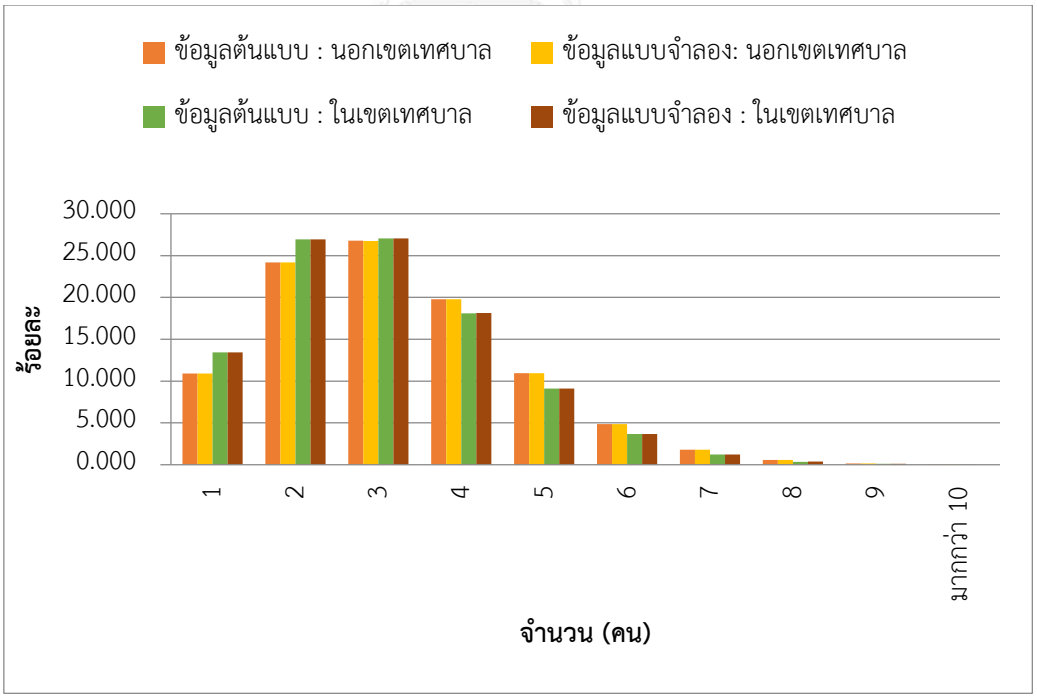
(ข) อ.ชาติตระการ



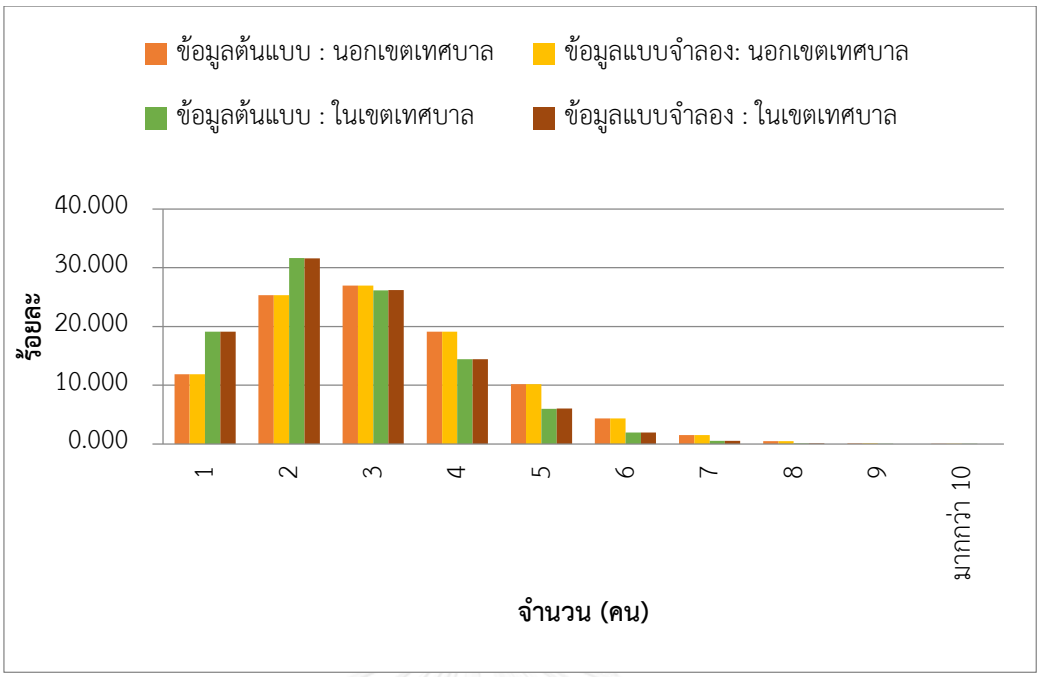
(ค) อ.นครไทย



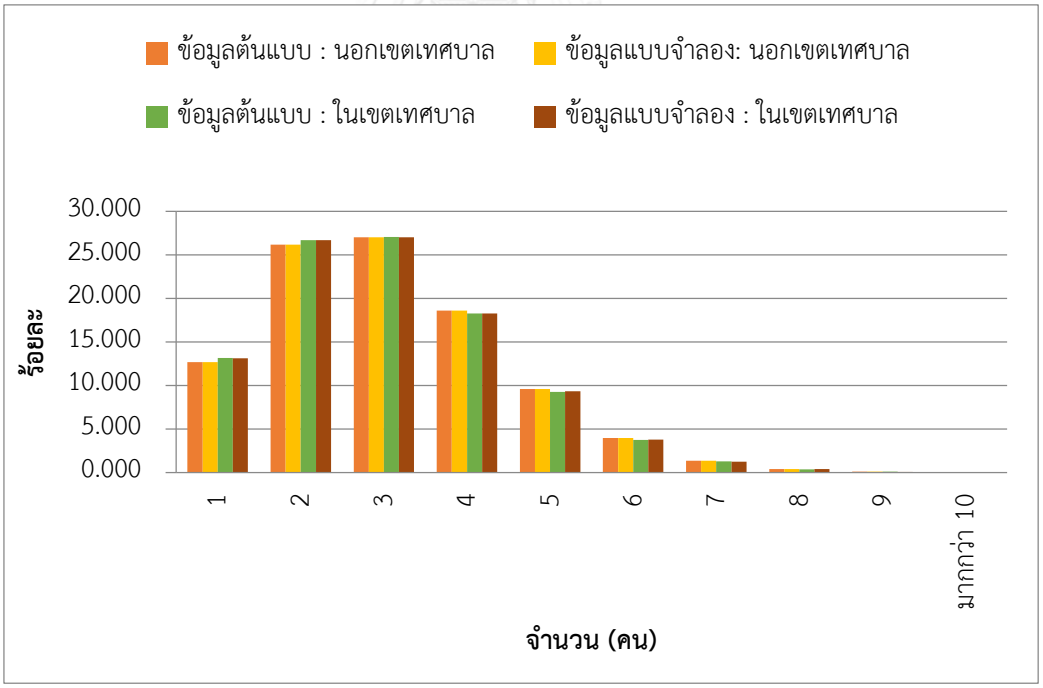
(ง) อ.บางกระทุ่ม



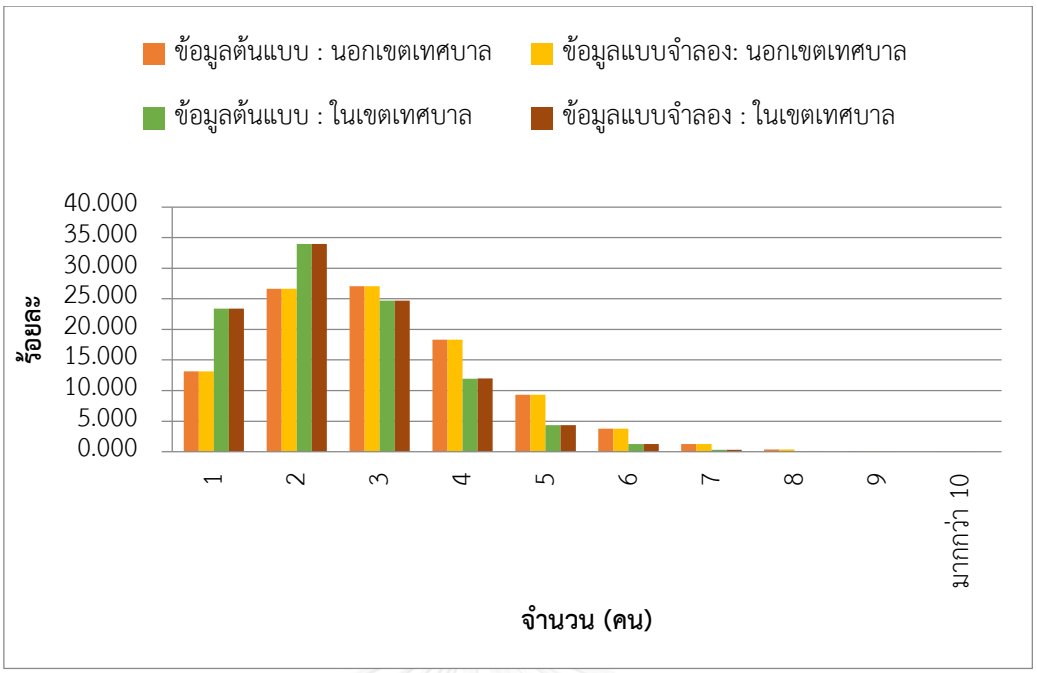
(จ) อ.บางระกำ



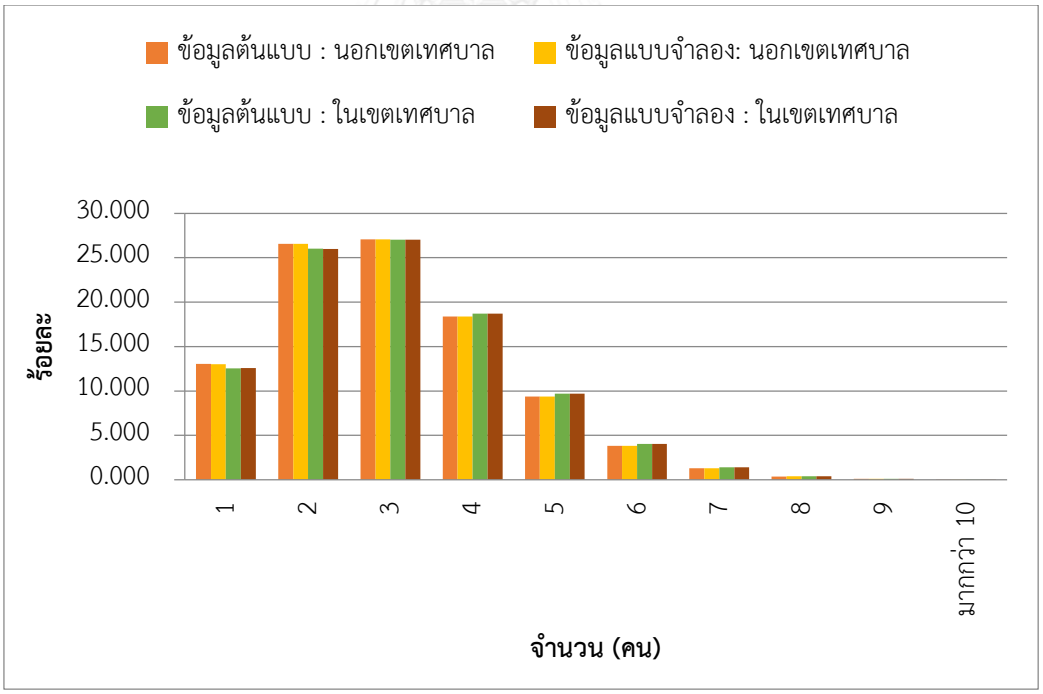
(ฉ) อ.พรหมพิราม



(ช) อ.วังทอง

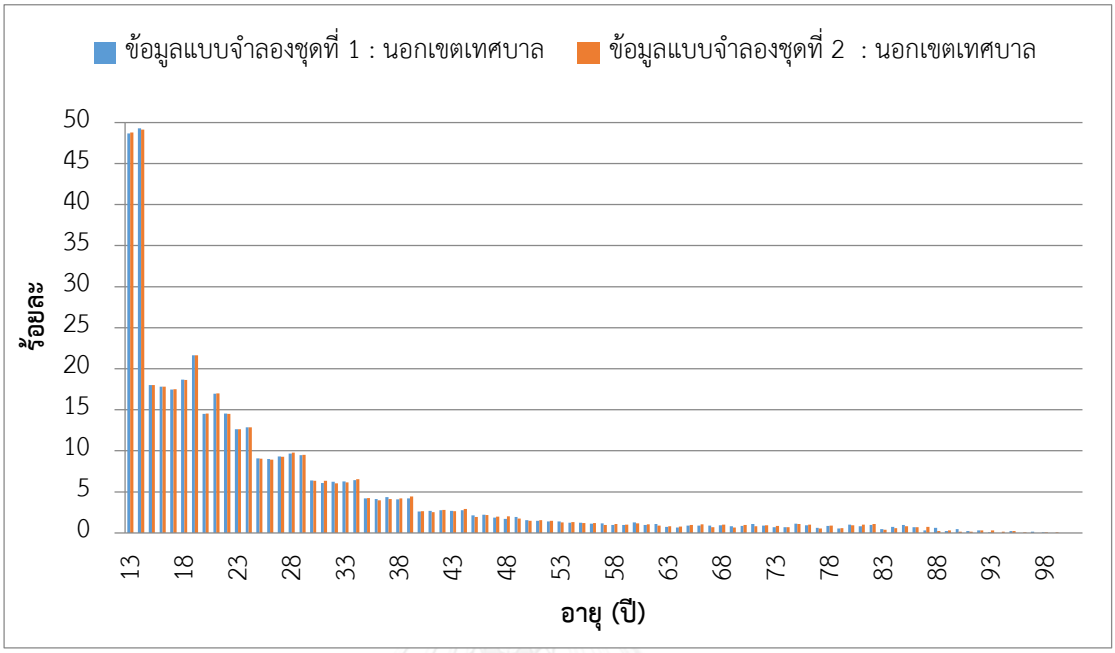


(ซ) อ.วัดโบสถ์

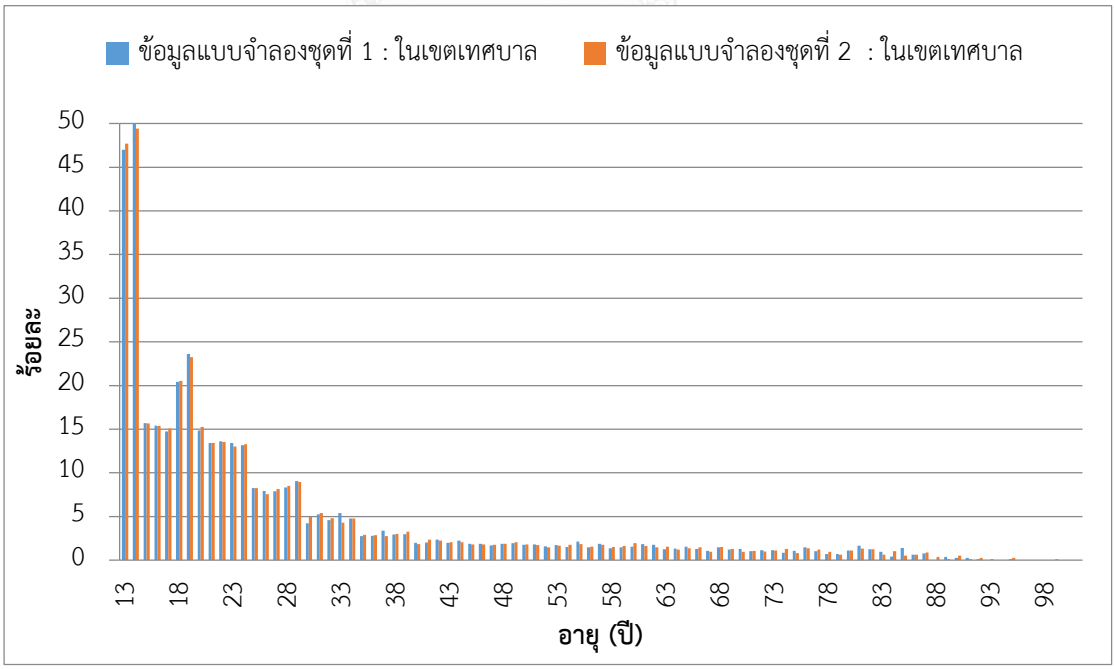


(ฌ) อ.เนินมะปราง

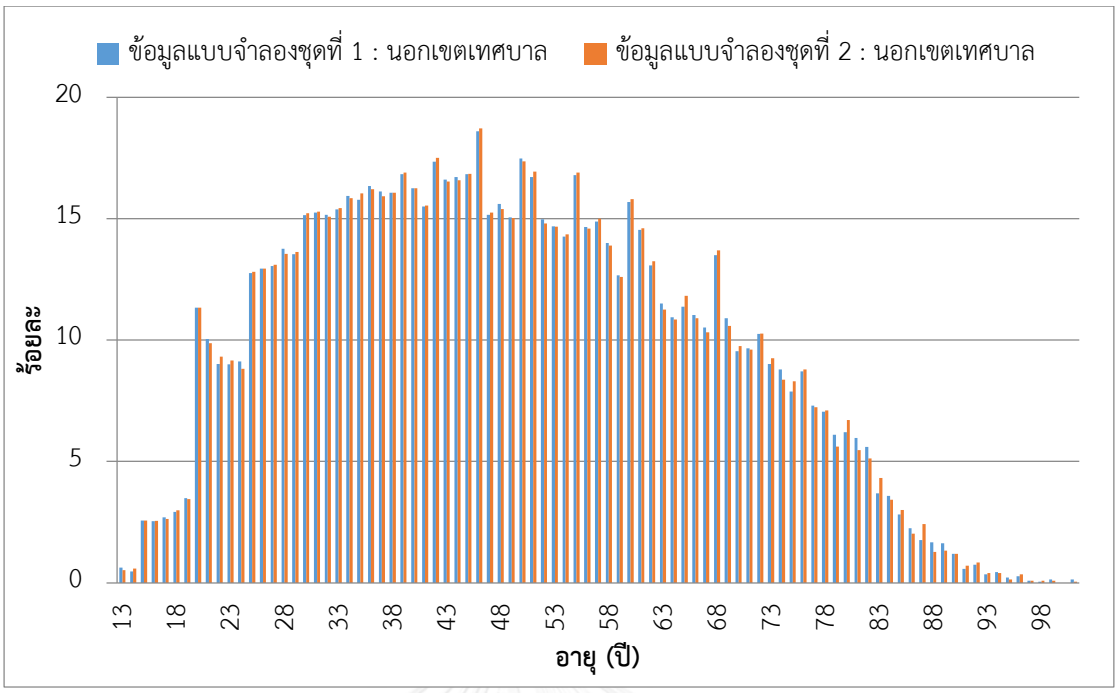
ภาพที่ ข.16 กราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศชาย



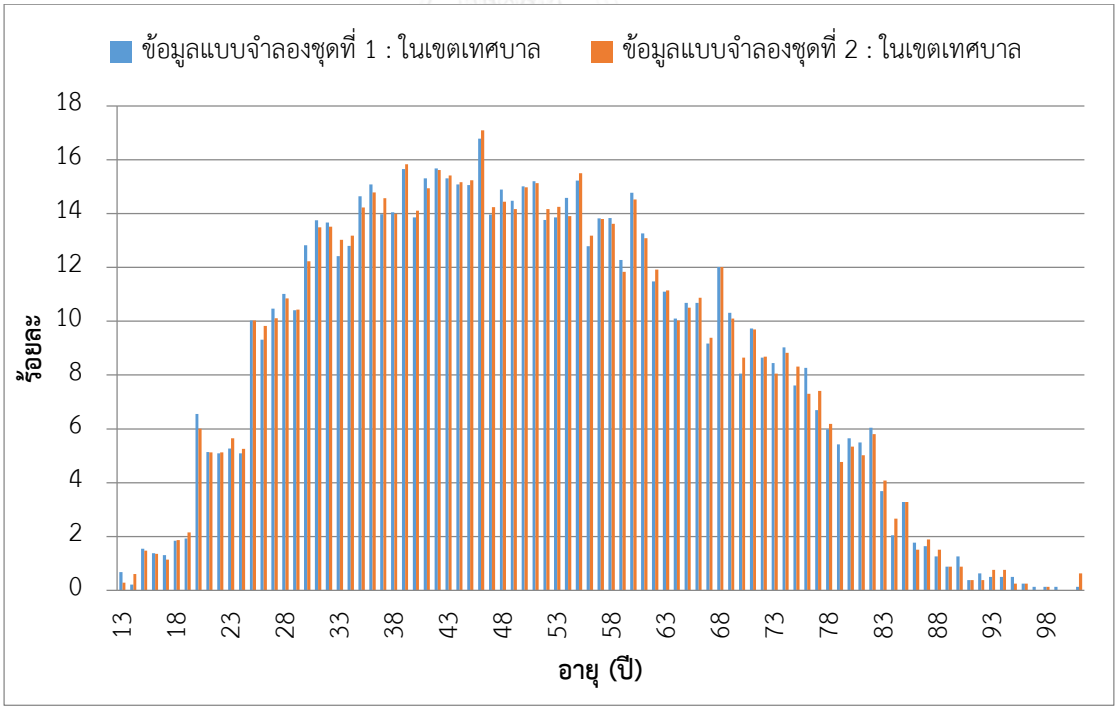
(ค) โสด – นอกเขตเทศบาล



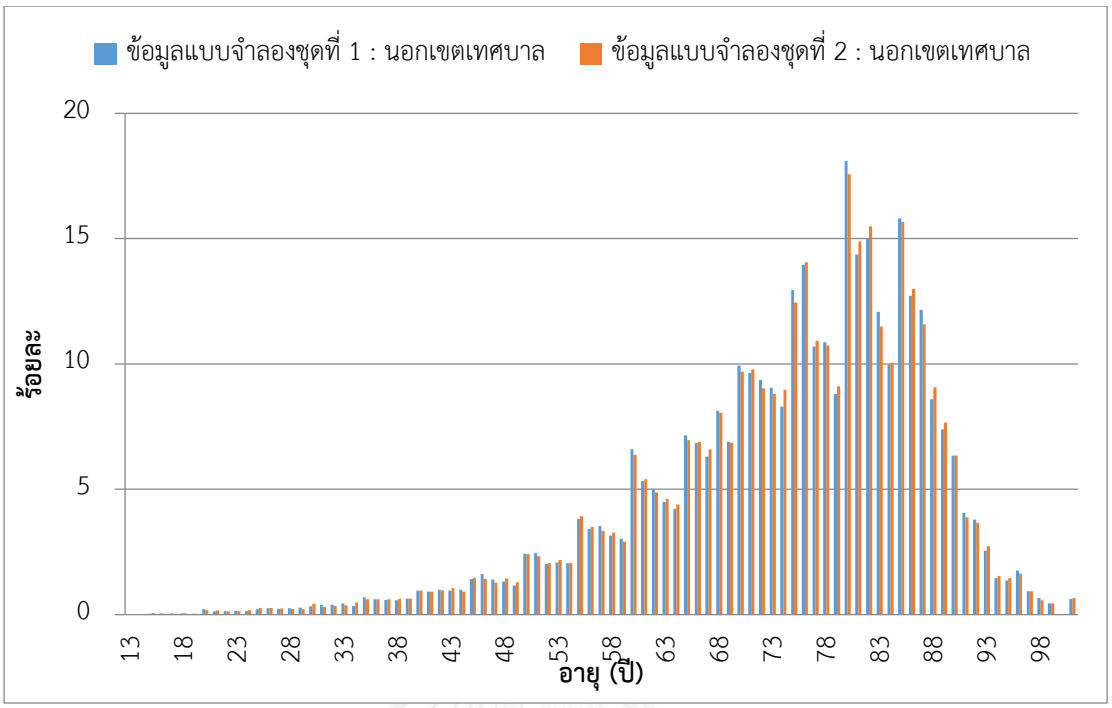
(ง) โสด – ในเขตเทศบาล



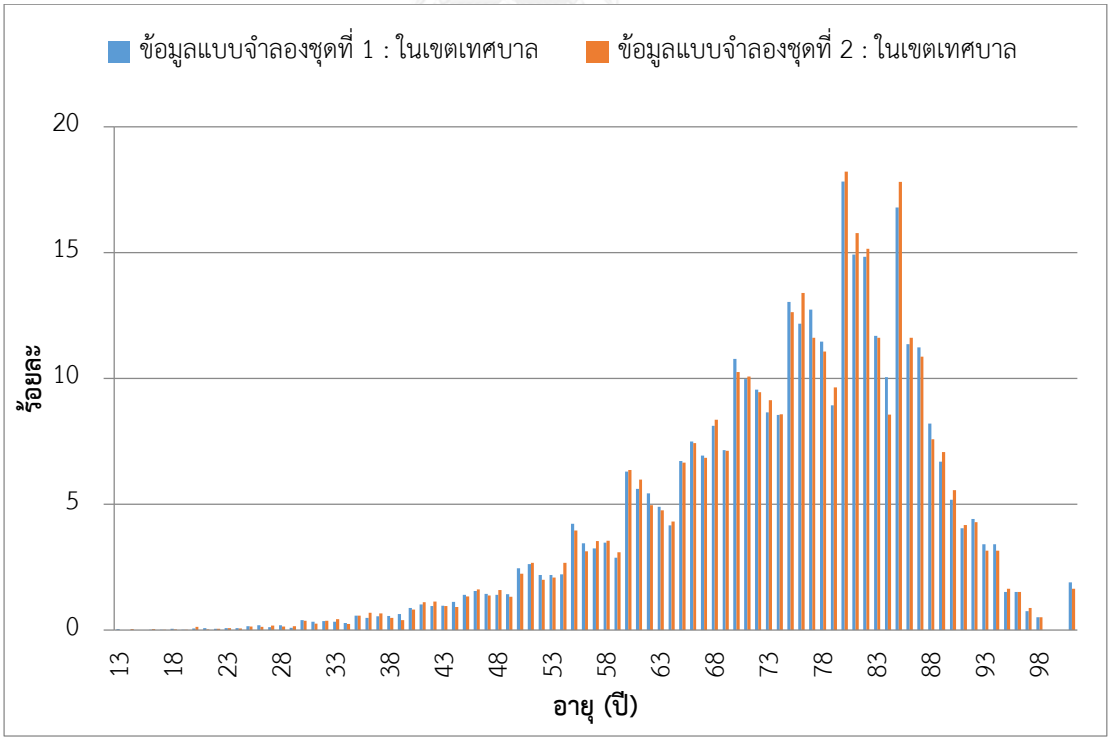
(จ) สมรส – นอกเขตเทศบาล



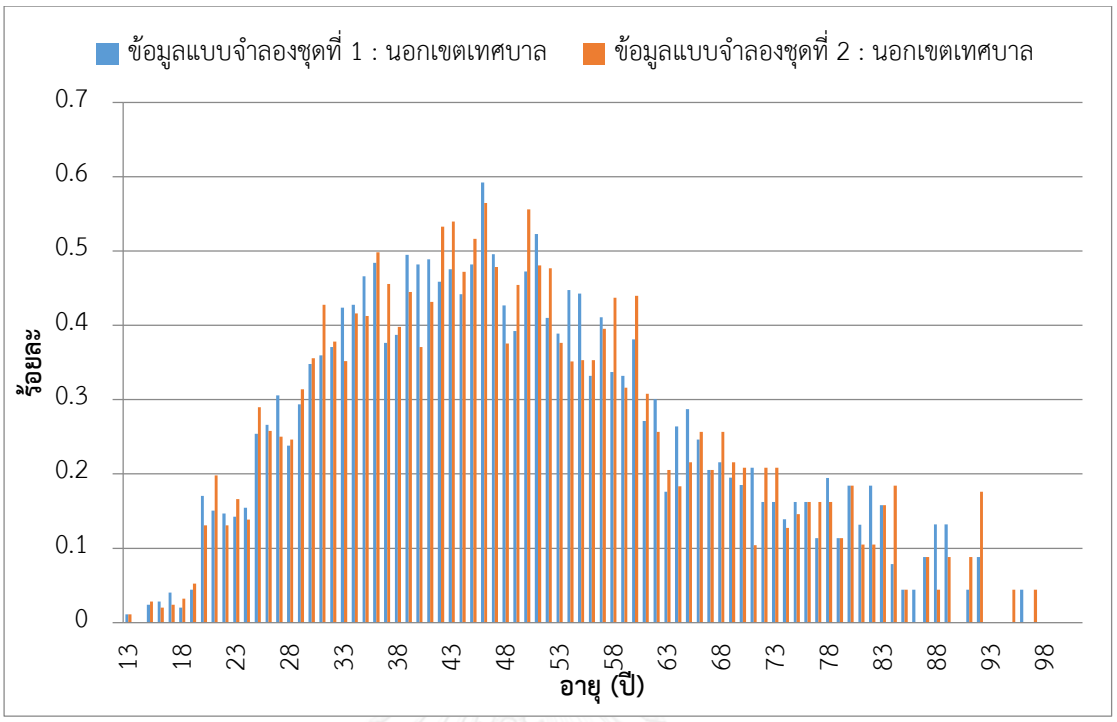
(ฉ) สมรส – ในเขตเทศบาล



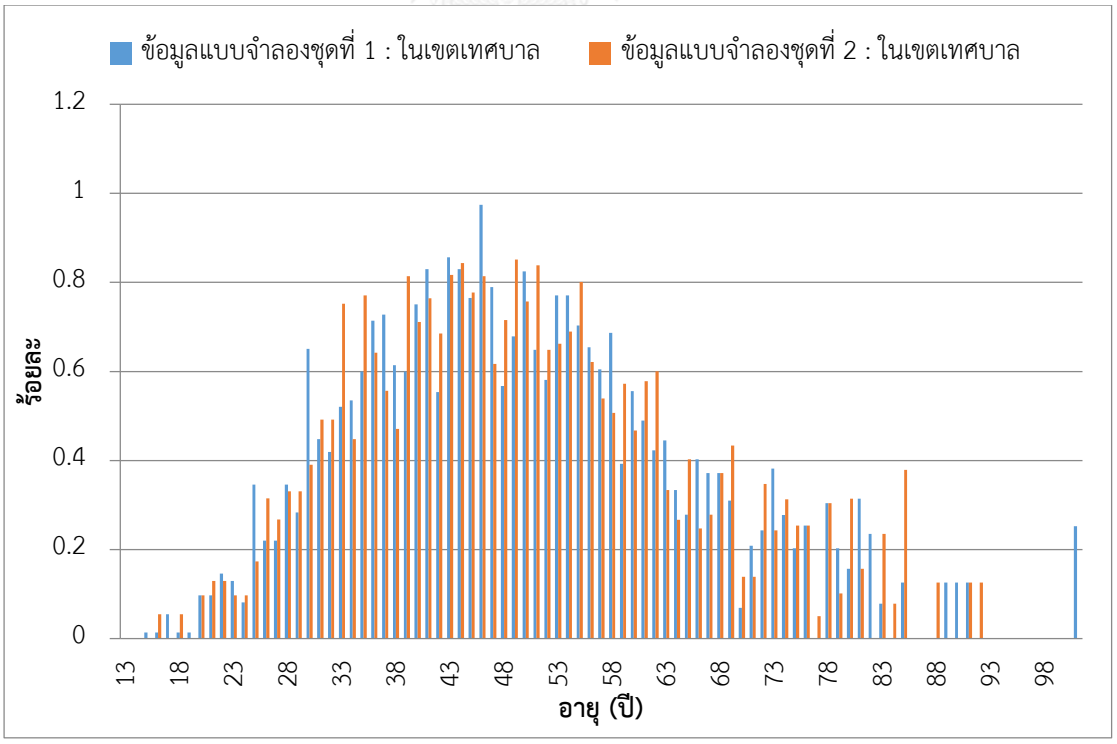
(ช) หม้าย – นอกเขตเทศบาล



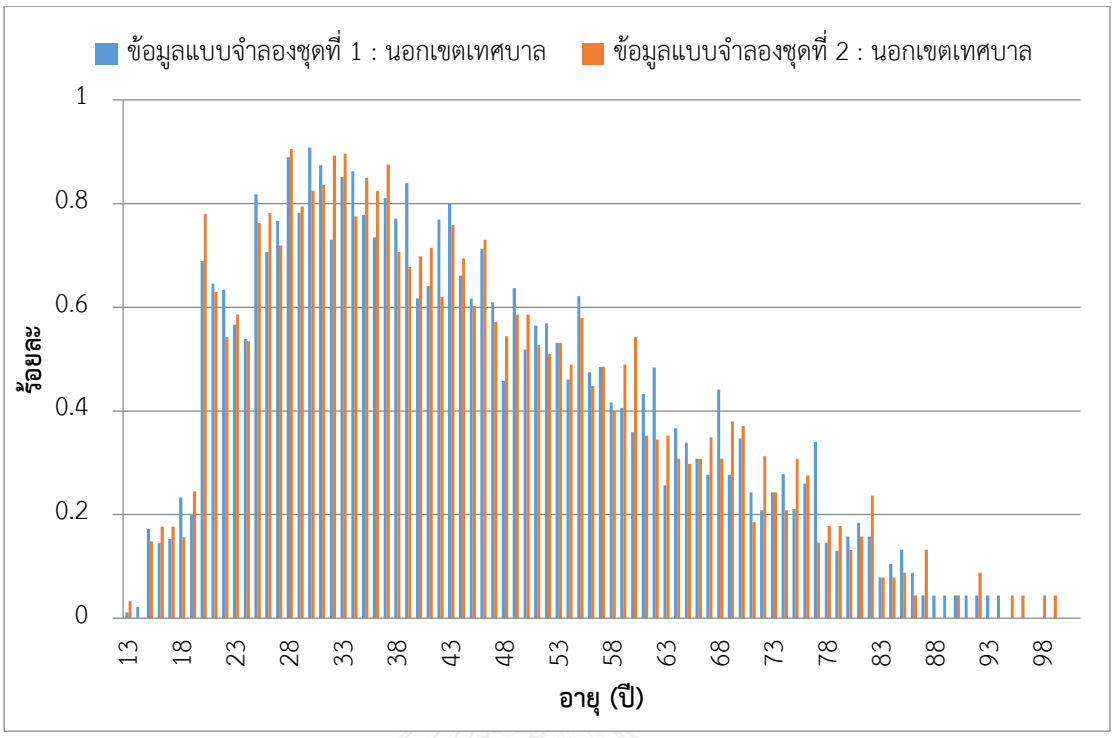
(ซ) หม้าย – ในเขตเทศบาล



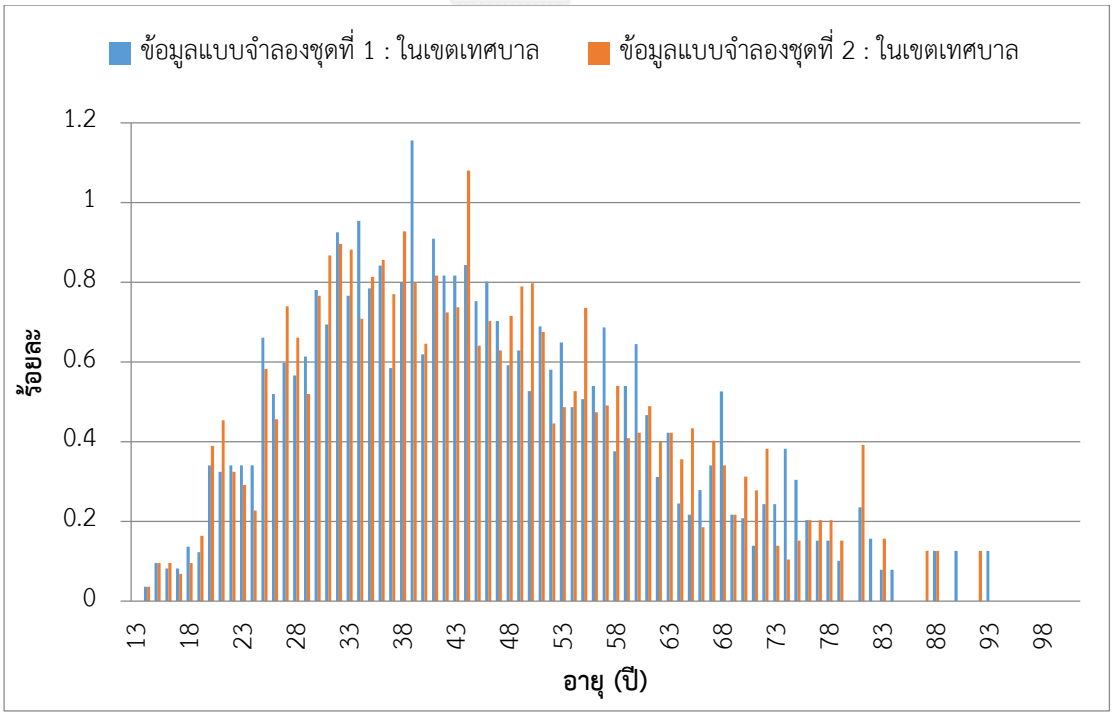
(ฌ) หย่า – นอกเขตเทศบาล



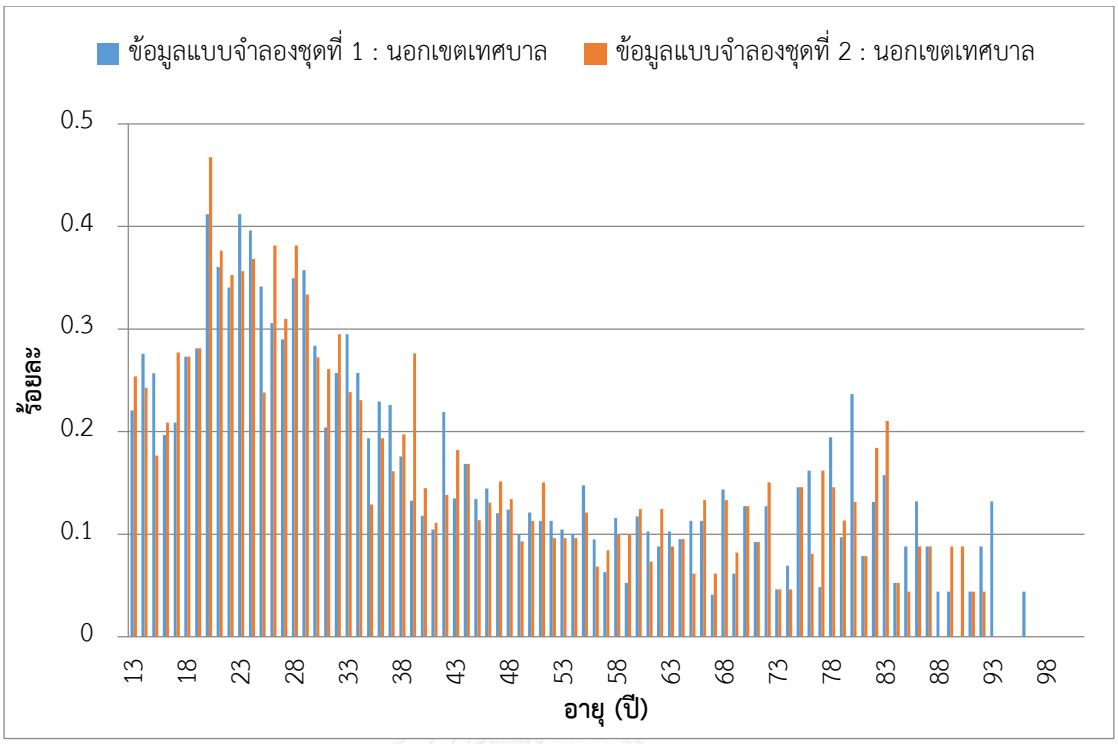
(ญ) หย่า – ในเขตเทศบาล



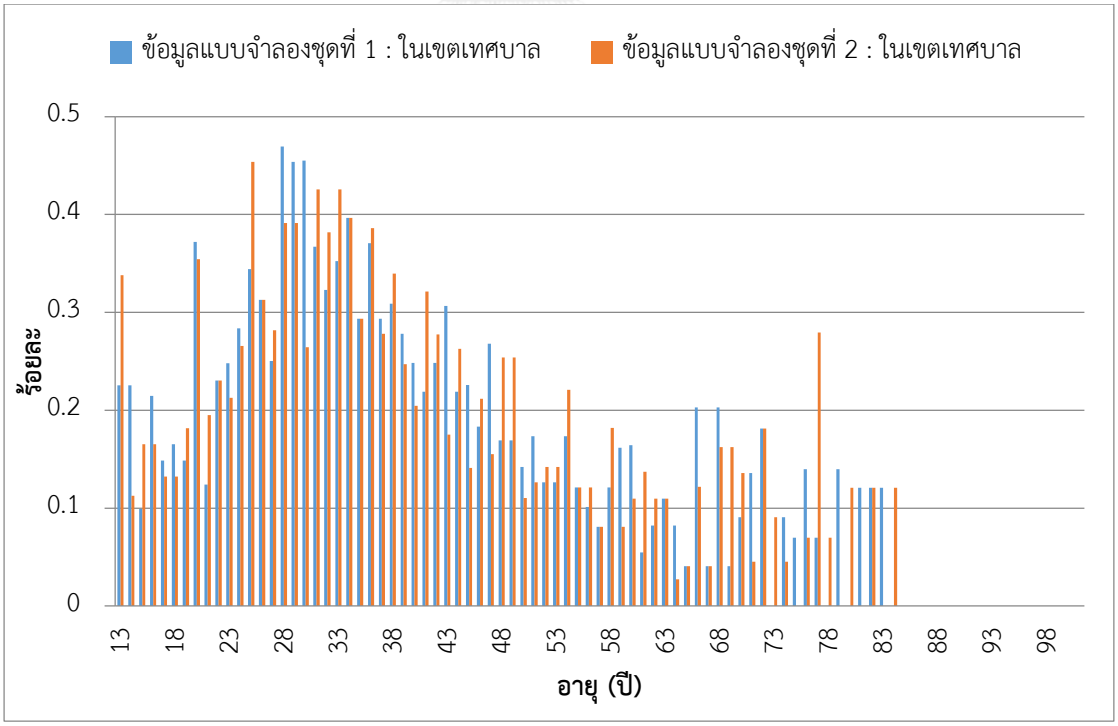
(ฎ) แยกกันอยู่ – นอกเขตเทศบาล



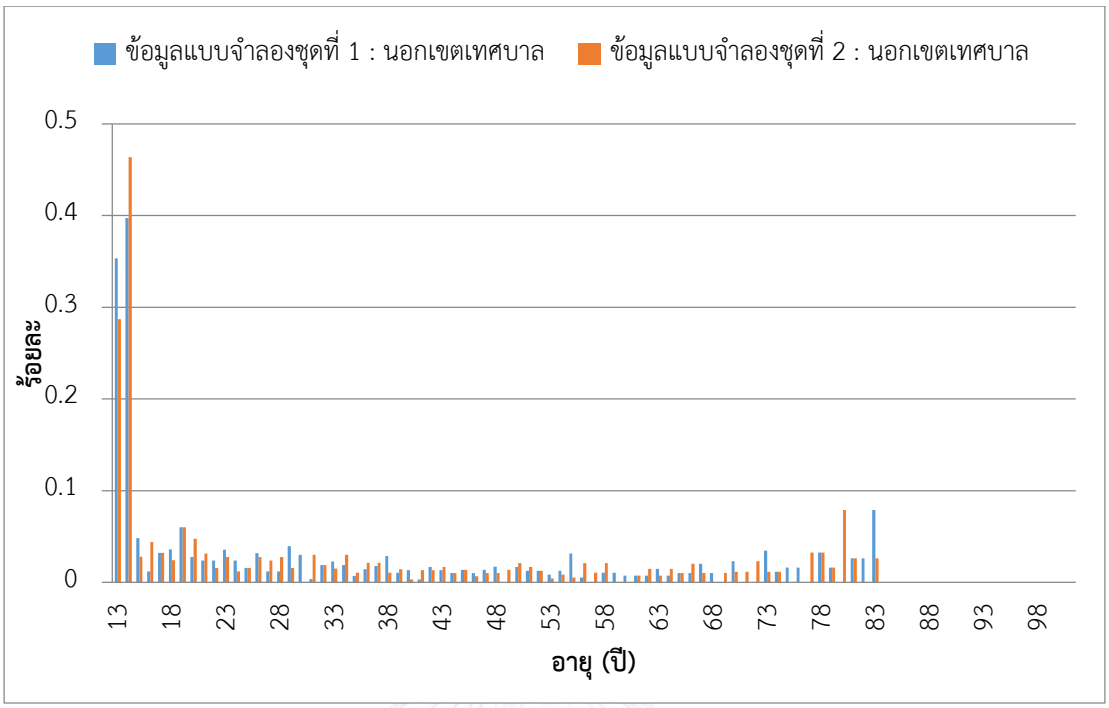
(ฎ) แยกกันอยู่ – ในเขตเทศบาล



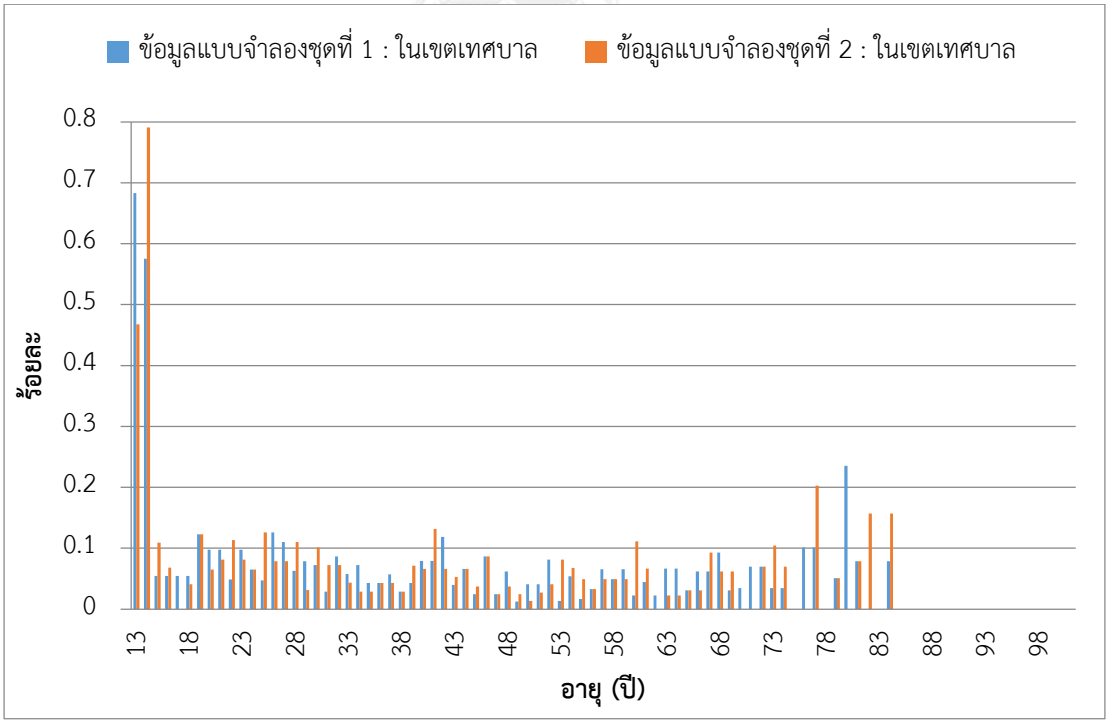
(ฐ) เคยสมรสแต่ไม่ทราบสถานภาพสมรส – นอกเขตเทศบาล



(ฑ) เคยสมรสแต่ไม่ทราบสถานภาพสมรส – ในเขตเทศบาล

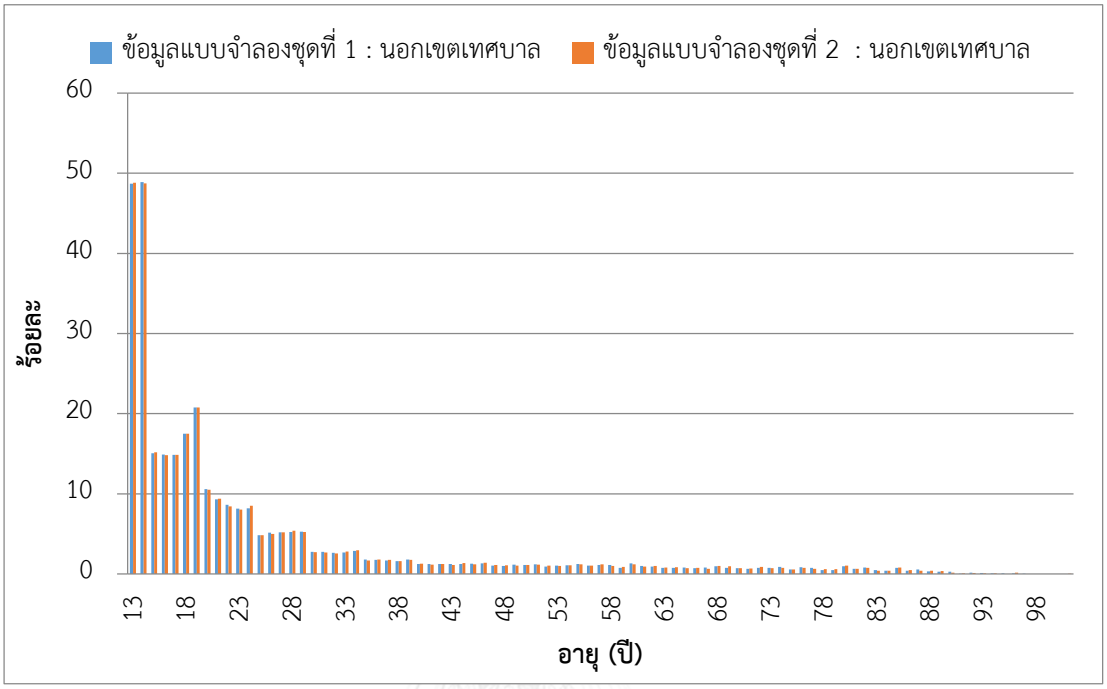


(ฒ) ไม่ทราบสถานภาพสมรส – นอกเขตเทศบาล

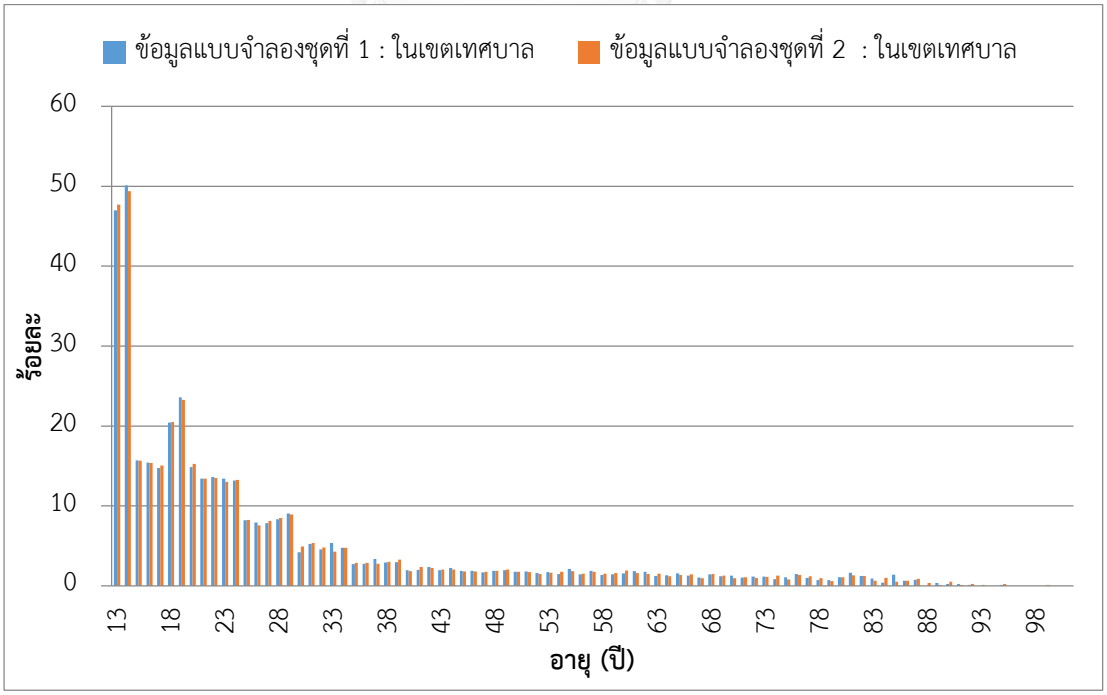


(ณ) ไม่ทราบสถานภาพสมรส – ในเขตเทศบาล

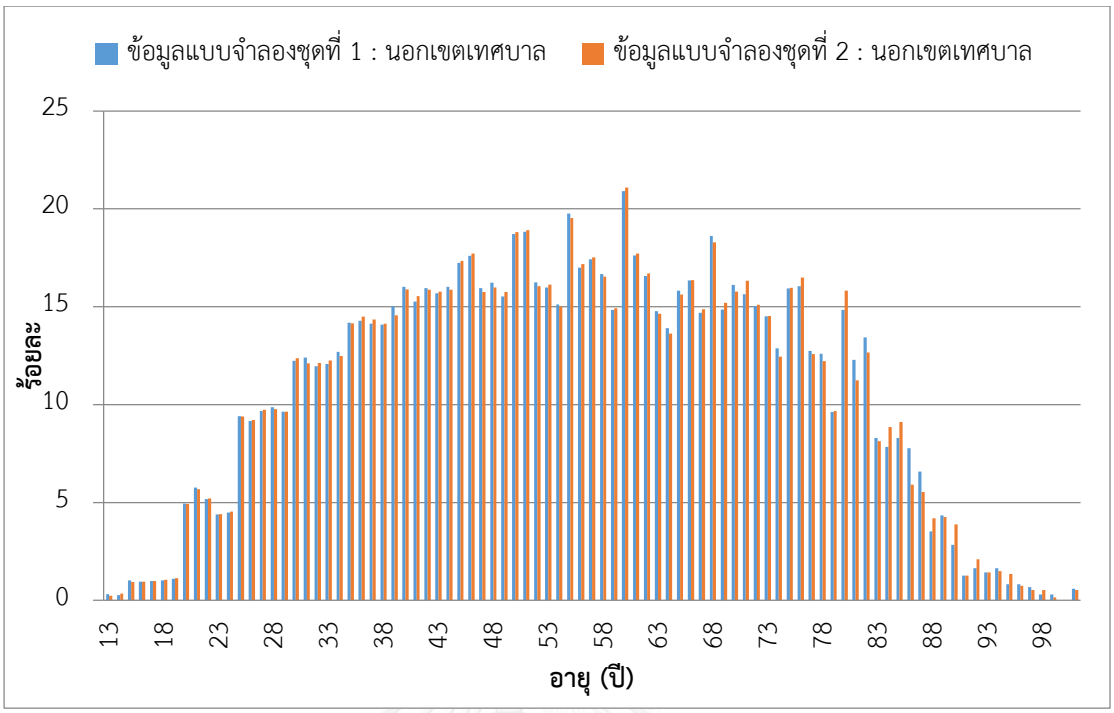
ภาพที่ ข.17 กราฟการเปรียบเทียบตารางที่ถูกซ่อนของข้อมูลแบบจำลองทั้ง 2 ชุดของคุณลักษณะข้อมูลสถานภาพสมรสของเพศหญิง



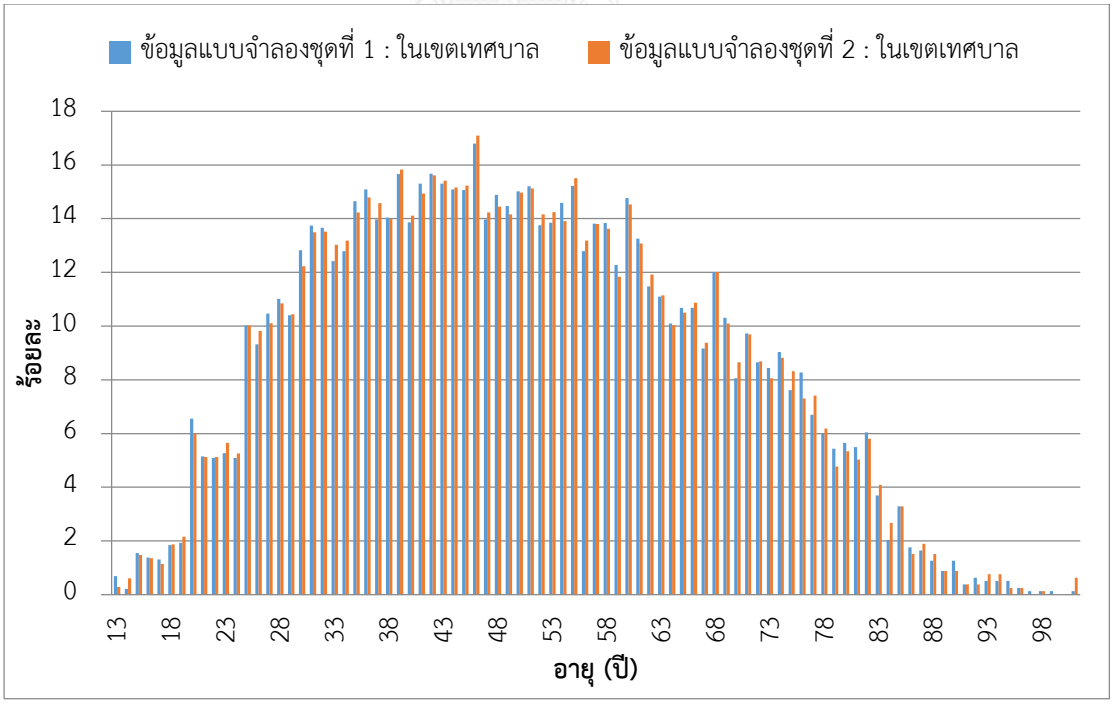
(ก) โสด – นอกเขตเทศบาล



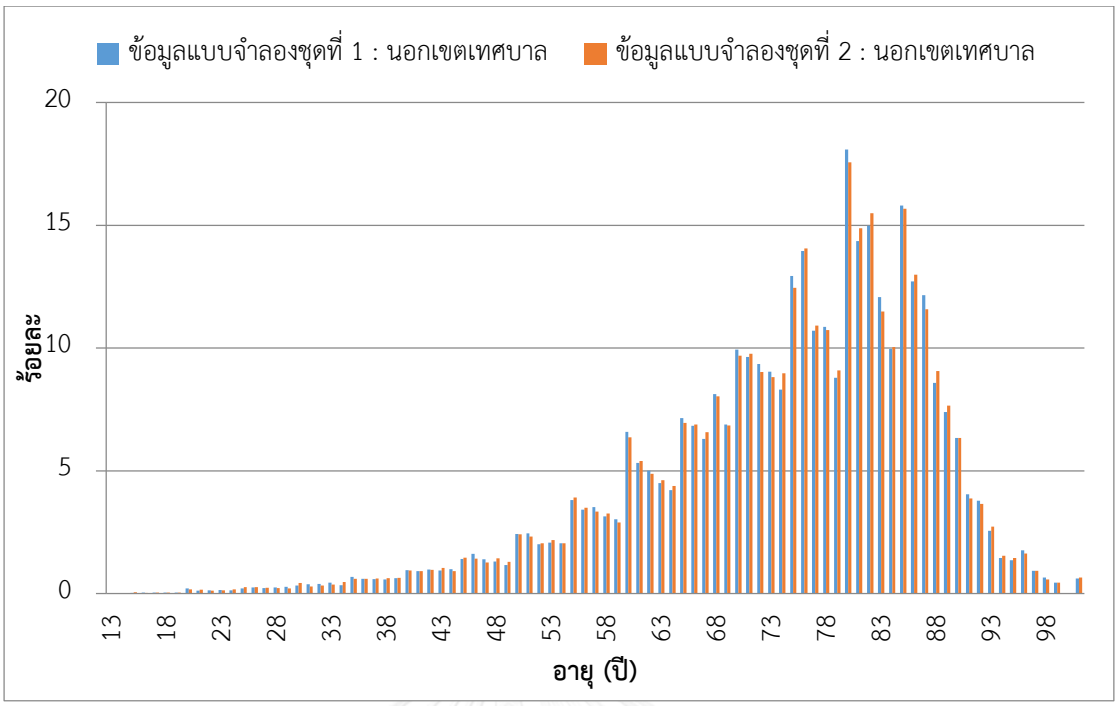
(ข) โสด – ในเขตเทศบาล



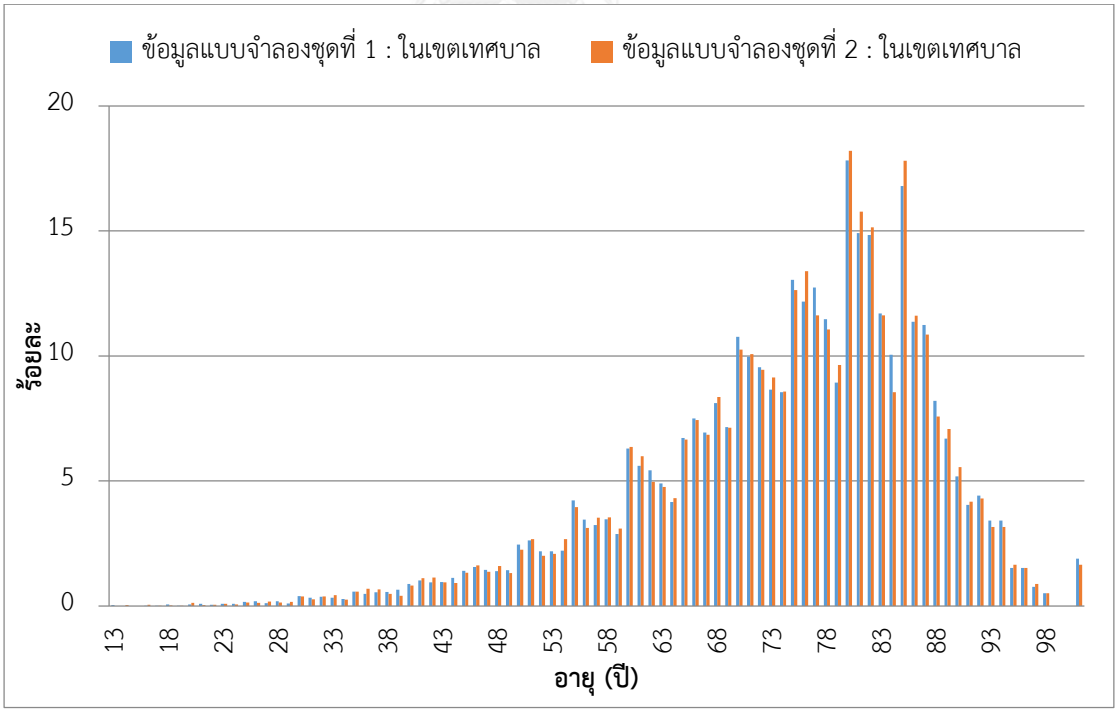
(ค) สมรส – นอกเขตเทศบาล



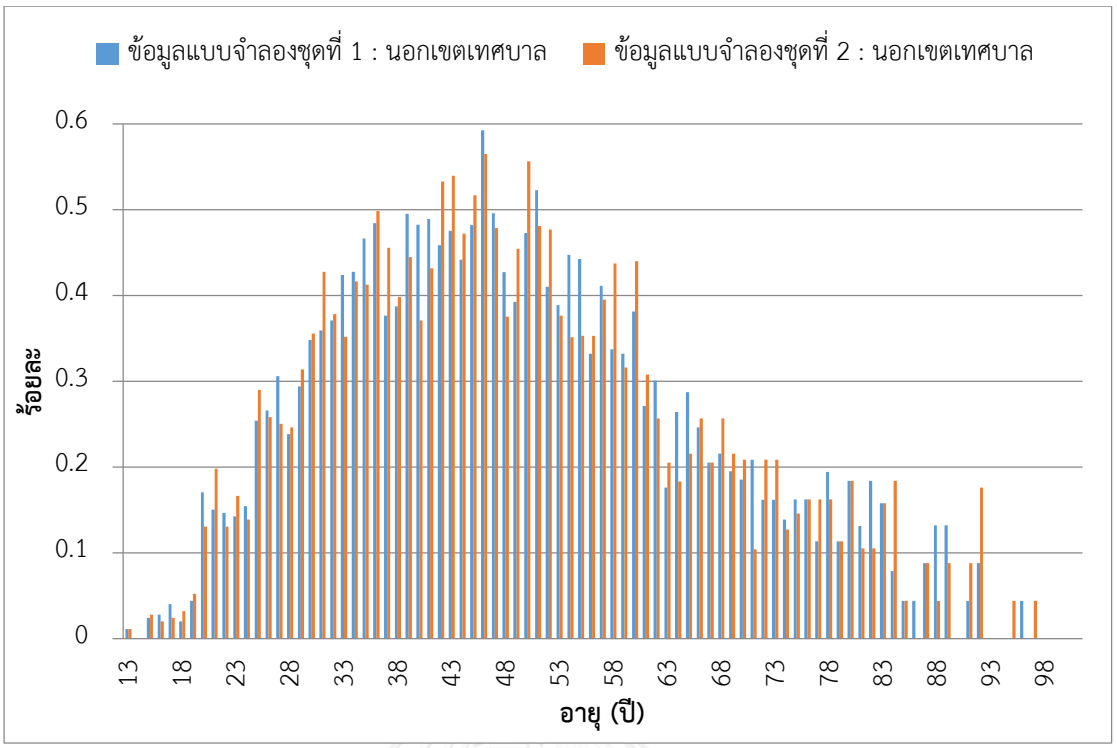
(ง) สมรส – ในเขตเทศบาล



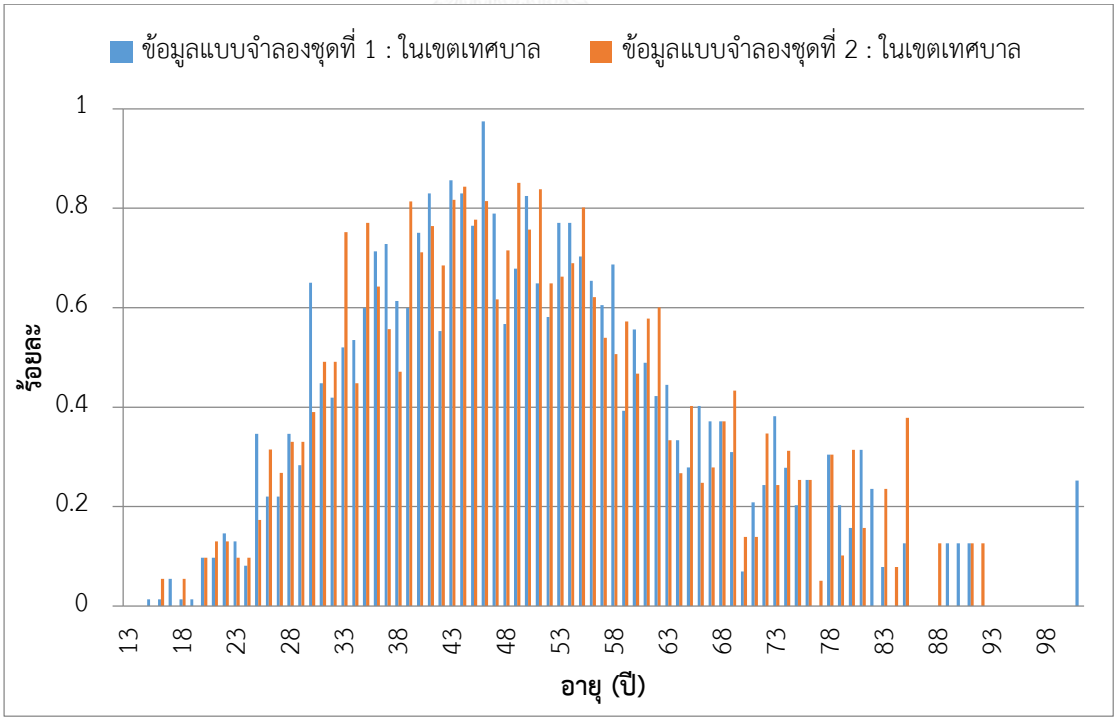
(จ) หม้าย – นอกเขตเทศบาล



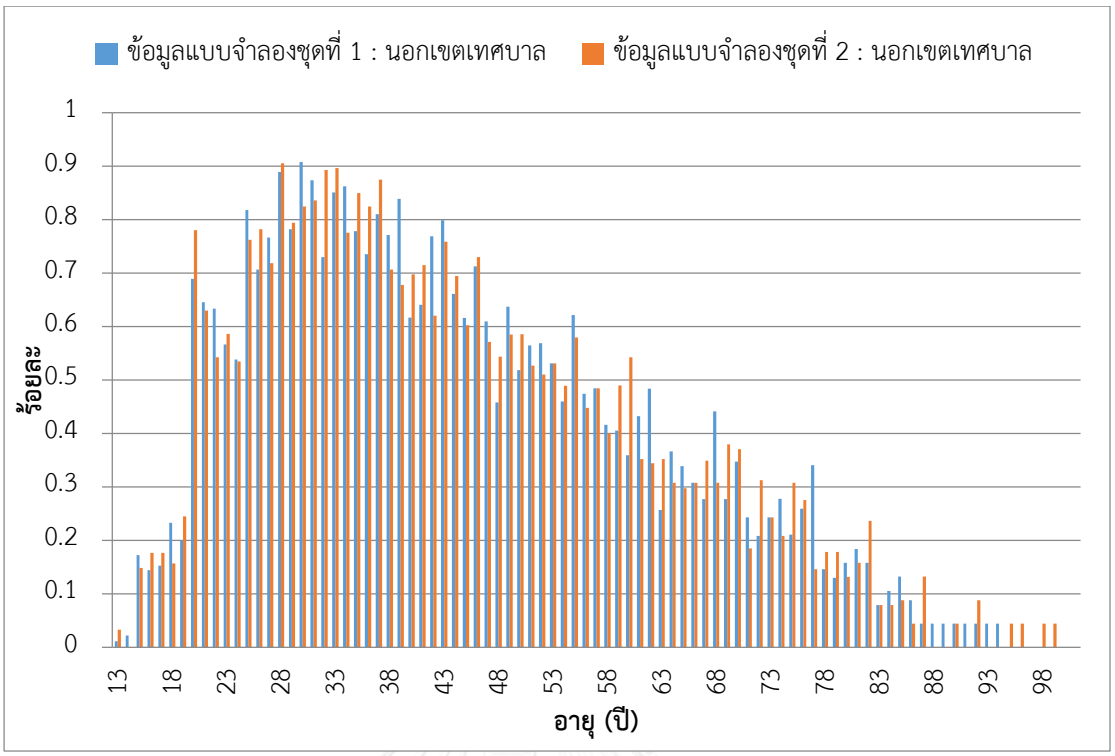
(ฉ) หม้าย – ในเขตเทศบาล



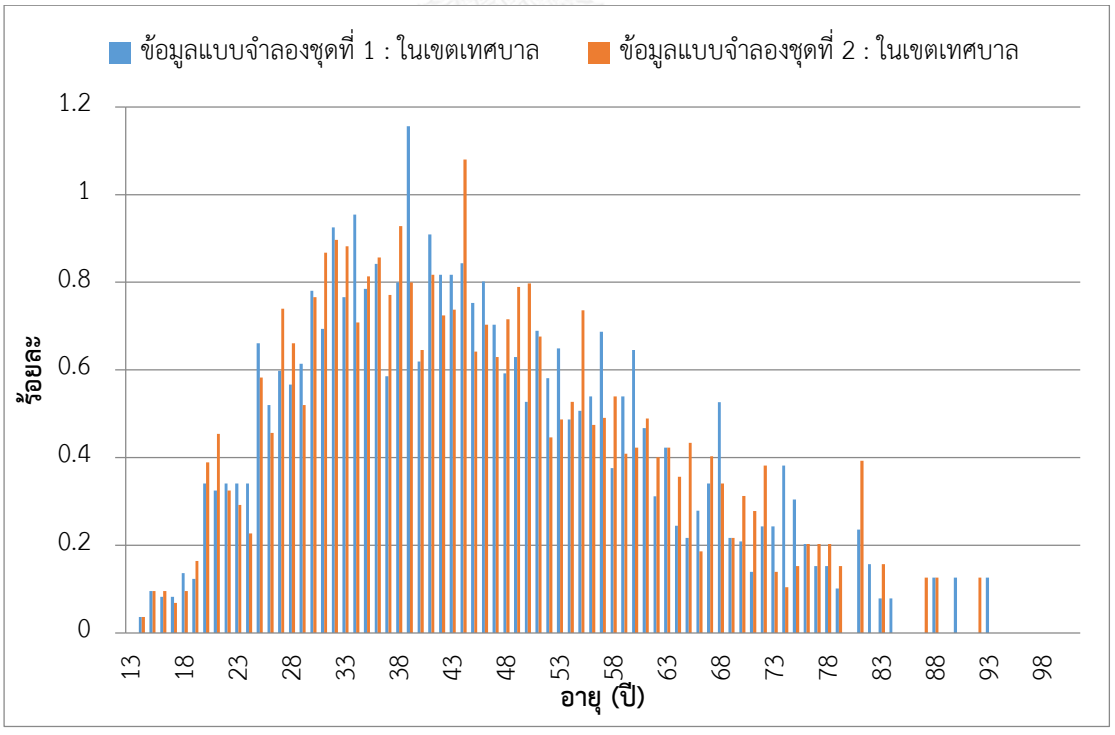
(ช) หย่า – นอกเขตเทศบาล



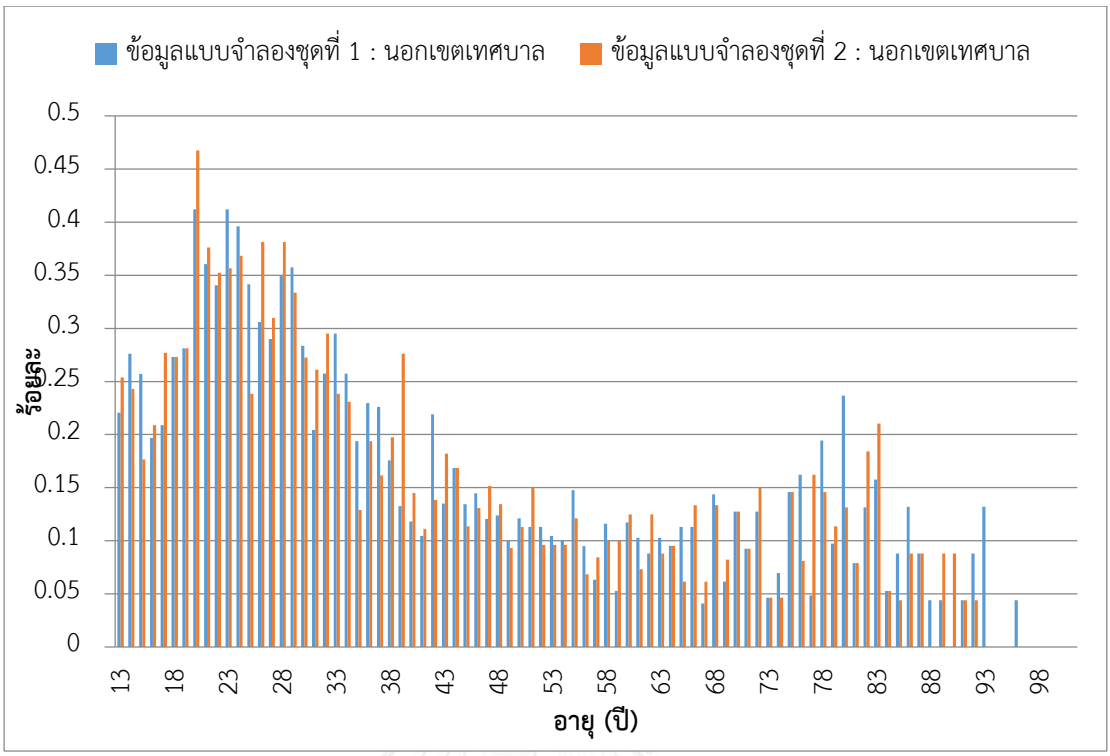
(ซ) หย่า – ในเขตเทศบาล



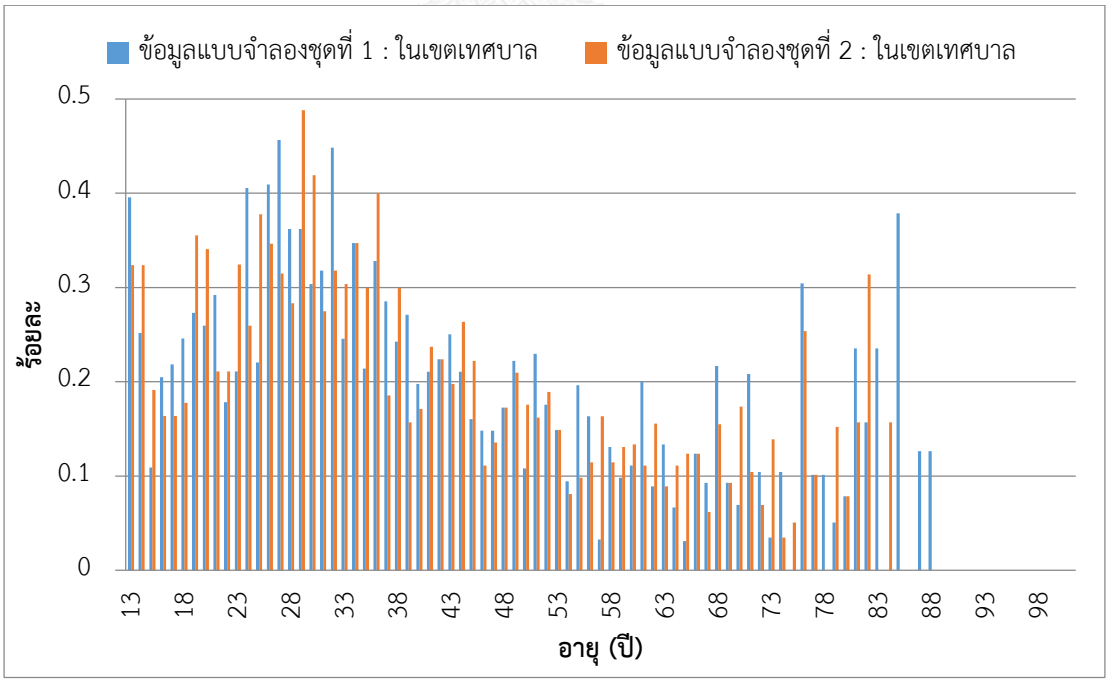
(ฌ) แยกกันอยู่ – นอกเขตเทศบาล



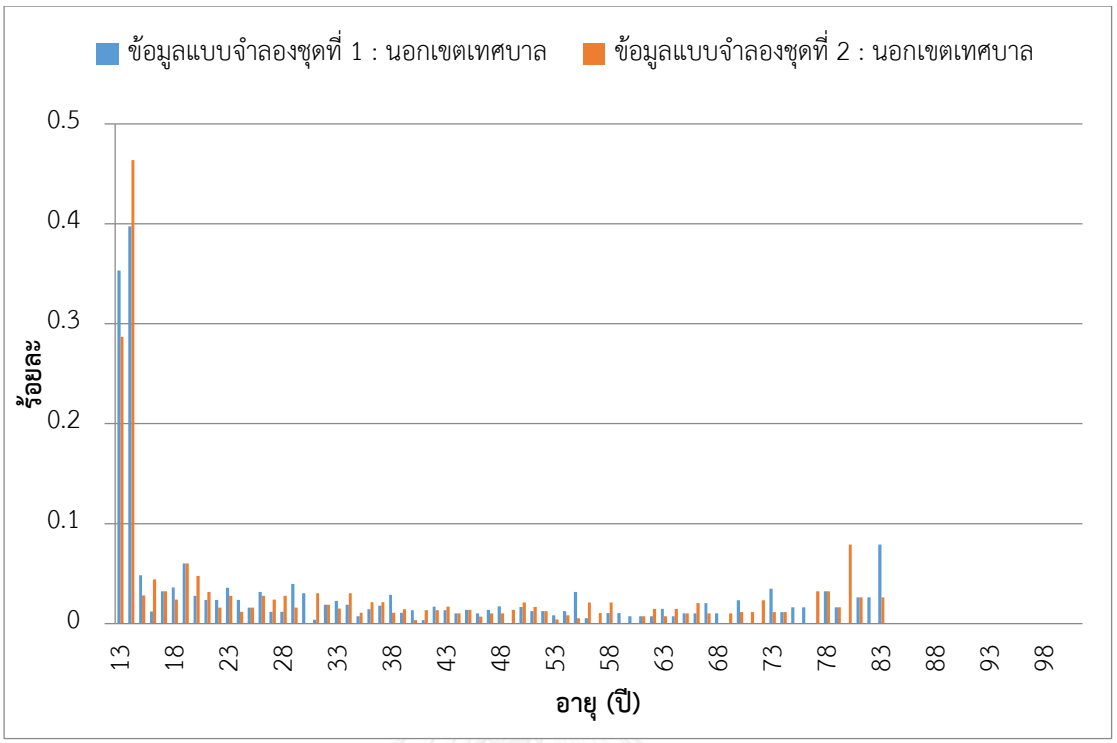
(ญ) แยกกันอยู่ – ในเขตเทศบาล



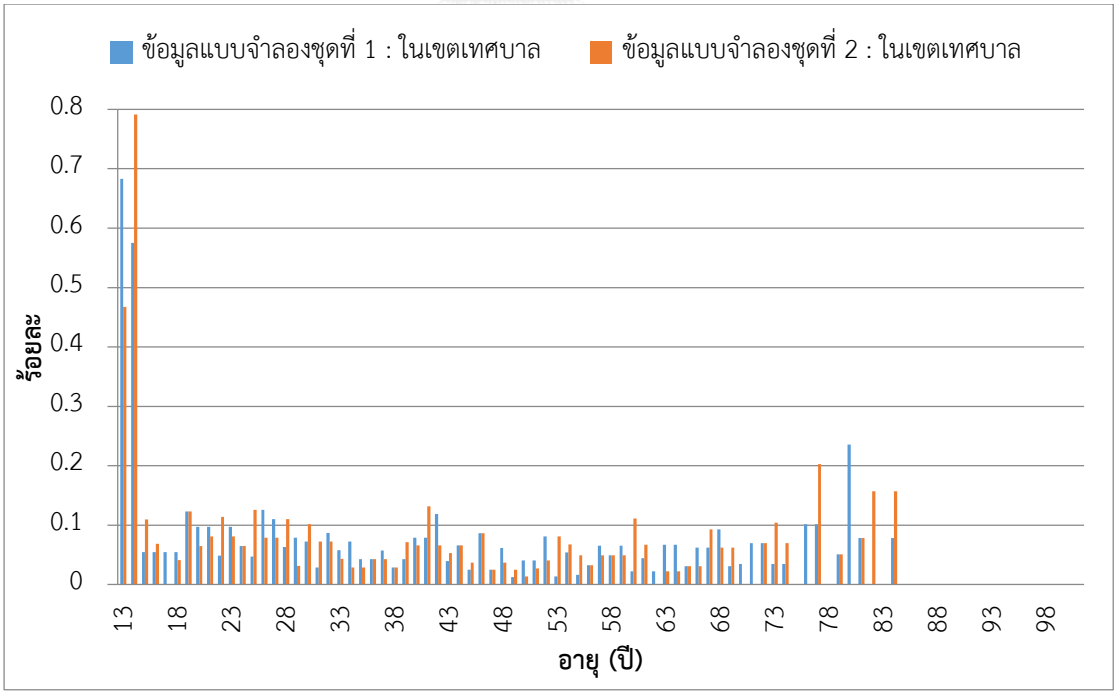
(ฎ) เคยสมรสแต่ไม่ทราบสถานภาพสมรส – นอกเขตเทศบาล



(ฎ) เคยสมรสแต่ไม่ทราบสถานภาพสมรส – ในเขตเทศบาล



(ฐ) ไม่ทราบสถานภาพสมรส – นอกเขตเทศบาล



(ฑ) ไม่ทราบสถานภาพสมรส – ในเขตเทศบาล

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวณัฐพร วัฒนสุทธิ เกิดเมื่อวันที่ 27 มกราคม พ.ศ.2534 ที่จังหวัดพิษณุโลก สำเร็จการศึกษาระดับปริญญาบัณฑิต หลักสูตรวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ (เกียรตินิยมอันดับ 2) จากคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ เมื่อปี พ.ศ. 2555 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ. 2556

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University