การประมาณค่าเฉลี่ยจากการสำรวจตัวอย่าง ที่ปรับตามแผนแบบการชักตัวอย่างด้วย PROC SURVEYMEANS

Population Mean Estimation from Sample Survey Data with Respect to the Sample Designs by PROC SURVEYMEANS

กมล บุษบา

ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี 12121

kamon@mathstat.sci.tu.ac.th

1. บทนำ

โดยทั่วไปแล้วนักวิจัยจะทำการศึกษาลักษณะที่สนใจของประชากรโดยใช้วิธีการสำรวจด้วยตัวอย่าง และเนื่องจากลักษณะที่สนใจศึกษาในประชากรมีความผันแปรหรือแตกต่างกัน ดังนั้นนักวิจัยจำเป็นต้องใช้แผนแบบการชักตัวอย่างที่เหมาะสมในกระบวนการเลือกตัวอย่างจากประชากรหนึ่งๆ เช่น ใช้แผนแบบการชักตัวอย่างแบบมีชั้นภูมิ (stratification) ใช้แผนแบบการชักตัวอย่างแบบแบ่งกลุ่ม (clustering) ใช้แผนแบบการชักตัวอย่างที่กำหนดน้ำหนักในการชักตัวอย่างที่ต่างกัน เป็นต้น เพื่อที่จะทำการอ้างอิงเกี่ยวกับประชากรได้อย่างไม่บิดเบือนหรือมีความถูกต้องตรงตามความเป็นจริงโดยอาศัยเฉพาะข้อมูลจากตัวอย่างที่ได้มา การทำการอ้างอิงที่ถูกต้องในเชิงสถิตินั้นนักวิจัยจะต้องทำการวิเคราะห์ข้อมูลให้สอดคล้องกับแผนแบบการชักตัวอย่างที่ใช้ด้วย [1]

การคำนวณสถิติพรรณนาหรือหาค่าเฉลี่ยของข้อมูลที่สำรวจมานั้น โดยปกตินักวิจัยมักใช้โปรแกรม SPSS โดยใช้เมนู Analyze ตามด้วยเมนู Descriptive Statistics หรือใช้โปรแกรม SAS โดยใช้ PROC MEANS เป็นต้น อย่างไรก็ตามคำสั่งเหล่านี้จะทำการคำนวณสถิติต่างๆ ภายใต้ข้อตกลงเบื้องต้นว่าตัวอย่างที่ได้นั้นมาจากประชากรที่มีขนาดอนันต์ (infinite population) หรือใช้วิธีการชักตัวอย่างสุ่มแบบง่าย (simple random sampling หรือ SRS) ซึ่งโดยทั่วๆ ไป คำสั่งเหล่านี้ให้ค่าประมาณความแปรปรวนของตัวประมาณที่ไม่ถูกต้องหรือให้ค่าต่ำกว่าความเป็นจริงหากนำไปใช้กับตัวอย่างชุดที่ได้มาจากแผนแบบการชักตัวอย่างอื่นๆ ที่ไม่ใช่แผนแบบการชักตัวอย่างสุ่มแบบง่าย [1 และ 2] ปัญหาดังกล่าวในปัจจุบันสามารถแก้ไขได้โดยใช้คำสั่งที่มีการพัฒนาขึ้นมาใหม่ในโปรแกรมสำเร็จรูป SAS ตั้งแต่รุ่น 7.0 ขึ้นไป (รุ่นล่าสุดคือรุ่น 9.1.3) เช่น PROC SURVEYSELECT ใช้สำหรับการชักตัวอย่างที่ใช้ความน่าจะเป็นเป็นพื้นฐานในการเลือกตัวอย่าง PROC SURVEYFREQ ใช้สำหรับสร้างตารางแจกแจงความถี่และตารางไขว้จากข้อมูลที่ได้จากการสำรวจ ประมาณร้อยละต่างๆ ค่าคลาดเคลื่อนมาตรฐานของการประมาณ รวมทั้งการทดสอบความสัมพันธ์ระหว่างตัวแปรในตารางไขว้ PROC SURVEYMEANS ใช้สำหรับการประมาณทั้งแบบจุดและแบบช่วงของค่ายอดรวม ค่าเฉลี่ยและค่าสัดส่วนของประชากรรวมทั้งค่าคลาดเคลื่อนมาตรฐานของการประมาณด้วย PROC SURVEYREG ใช้สำหรับการวิเคราะห์ถดถอยโดยปรับให้เหมาะสมกับข้อมูลที่ได้จากการสำรวจ เช่น คำนวณสัมประสิทธิ์การถดถอยและการทดสอบสมมุติฐานเกี่ยวกับพารามิเตอร์ต่างๆ ในตัวแบบ และ ยังมี PROC SURVEYLOGISTIC ที่ใช้สำหรับการวิเคราะห์ถดถอยโลจิสติกเพื่อศึกษาความสัมพันธ์ระหว่าง

ผลตอบสนองที่มีค่าไม่ต่อเนื่องและชุดของตัวแปรอธิบาย โดยปรับให้ใช้ได้กับข้อมูลที่ได้จากการสำรวจ [3] ในที่นี้จะยกตัวอย่างเฉพาะการใช้ PROC SURVEYMEANS ประมาณค่าเฉลี่ยเท่านั้น เนื่องจากเป็นการคำนวณสถิติขั้นพื้นฐานที่ต้องใช้ในการสำรวจเกือบทุกครั้ง

2. PROC SURVEYMEANS

PROC SURVEYMEANS ในโปรแกรม SAS ใช้สำหรับการประมาณค่าเฉลี่ยและยอดรวมของประชากรที่สำรวจโดยใช้ข้อมูลจากตัวอย่าง นอกจากนี้ยังคำนวณค่าประมาณความแปรปรวน ขอบเขตความเชื่อมั่นและสถิติพรรณนาอื่นๆ โดยในการคำนวณค่าประมาณเหล่านี้ จะมีการปรับให้สอดคล้องกับแผนแบบการชักตัวอย่างแบบต่างๆ ที่ใช้สำหรับการเลือกตัวอย่างด้วย PROC SURVEYMEANS จะใช้วิธีกระจายอนุกรมเทเลอร์ (Taylor expansion) ในการประมาณความคลาดเคลื่อนจากการเลือกตัวอย่างของตัวประมาณที่สอดคล้องกับแผนแบบการเลือกตัวอย่างนั้นๆ โดยถ้าประชากรถูกมีการแบ่งเป็นกลุ่มหรือมีการใช้หน่วยตัวอย่างขั้นต้น (primary sampling units) จะทำการประมาณความแปรปรวนโดยใช้ความผันแปรระหว่างหน่วยตัวอย่างขั้นต้น แต่ถ้ามีการแบ่งประชากรเป็นชั้นภูมิจะทำการรวม (pool) ค่าประมาณความแปรปรวนของแต่ละชั้นภูมิเพื่อคำนวณค่าประมาณของความแปรปรวนรวมทั้งหมด [4]

สำหรับผลลัพธ์ที่ได้ PROC SURVEYMEANS ใช้ระบบจัดส่งผลลัพธ์ (Output Delivery System (ODS)) เพื่อให้ผู้ใช้สามารถวางผลลัพธ์ในชุดข้อมูลผลลัพธ์ ซึ่งต่างจากโปรแกรม SAS รุ่นก่อนๆ ที่ต้องใช้คำสั่ง OUTPUT ในการสร้างแฟ้มที่เก็บผลลัพธ์

3. การชักตัวอย่างสุ่มแบบง่าย

ตัวอย่างต่อไปนี้เป็นการใช้ PROC SURVEYMEANS เพื่อประมาณค่าเฉลี่ยและสัดส่วนประชากรจากข้อมูลที่ได้จากการสำรวจตัวอย่าง ประชากรที่ศึกษาได้แก่นักเรียนชั้นมัธยมศึกษาตอนต้น (เกรด 7-9) จำนวน 4,000 คน ถ้าผู้วิจัยต้องการศึกษาว่านักเรียนกลุ่มนี้โดยเฉลี่ยแล้วใช้เงินซื้อไอศกรีมสัปดาห์ละเท่าไร และร้อยละของนักเรียนที่ใช้เงินซื้อไอสกรีมอย่างน้อย 10 เหรียญต่อสัปดาห์มีจำนวนเท่าไร [4]

ถ้าทำการศึกษาทำโดยการเลือกตัวอย่างนักเรียนจำนวน 40 คนจากประชากรของนักเรียนทั้งหมดโดยใช้วิธีการชักตัวอย่างสุ่มแบบง่าย (SRS) ซึ่งโดยวิธีการนี้นักเรียนทุกๆ คนมีความน่าจะเป็นในการถูกเลือกเป็นตัวอย่างเท่าๆ กันและไม่มีนักเรียนคนใดถูกเลือกมากกว่าหนึ่งครั้ง จากนั้นจะถามนักเรียนแต่ละคนที่ตกเป็นตัวอย่างเกี่ยวกับจำนวนเงินโดยเฉลี่ยที่ใช้ซื้อไอสกรีมต่อสัปดาห์ สมมุติว่าทำการป้อนข้อมูลของนักเรียนที่ตกเป็นตัวอย่างไว้ในส่วนหนึ่งของโปรแกรม การเขียนโปรแกรม SAS เพื่อทำการประมาณค่าเฉลี่ยของค่าไอศกรีมต่อสัปดาห์และสัดส่วนของผู้ที่ซื้อไอสกรีมอย่างน้อย 10 เหรียญต่อสัปดาห์ของประชากรนักเรียนทั้งหมดจำนวน 4,000 คน ทำได้ในภาพที่ 1

```
DATA IceCream;
  INPUT Grade Spending @@;
   IF (Spending < 10) THEN Group='less';
   ELSE  Group='more';
DATALINES;
7 7 7 7  8 12  9 10  7 1  7 10  7 3   8 20  8 19  7 2
7 2  9 15  8 16  7 6  7 6  7 6  9 15  8 17  8 14  9 8
9 8  9 7  7 3   7 12  7 4  9 14  8 18  9 9  7 2  7 1
7 4  7 11  9 8  8 10  8 13  7 2  9 6  9 11  7 2  7 9
;
PROC SURVEYMEANS DATA=IceCream
TOTAL=4000;
  VAR  Spending Group;
RUN;
```

ภาพที่ 1 คำสั่ง SAS ที่ใช้ประมาณค่าเฉลี่ยและสัดส่วนสำหรับแผนแบบการชักตัวอย่างสุ่มแบบง่ายจากประชากรที่มีขนาดจำกัด

ในที่นี้ตัวแปร Grade ใช้แทนชั้นของนักเรียน (เกรด 7 หรือ 8 หรือ 9) ส่วนตัวแปร Spending ใช้แทนค่าใช้จ่ายในการซื้อไอสกรีมโดยเฉลี่ยต่อสัปดาห์ (หน่วยเป็นเหรียญ) ส่วนตัวแปร Group เป็นตัวแปรที่สร้างขึ้นใหม่โดยมีค่าเป็น "less" ถ้านักเรียนใช้เงินน้อยกว่า 10 เหรียญต่อสัปดาห์ในการซื้อไอสกรีม และมีค่าเป็น "more" ถ้านักเรียนใช้เงินตั้งแต่ 10 เหรียญต่อสัปดาห์ขึ้นไปในการซื้อไอสกรีม

ให้สังเกตว่า คำสั่ง TOTAL=4000 ใน PROC SURVEYMEANS เป็นการระบุจำนวนประชากรทั้งหมด ซึ่งโปรแกรมจะใช้ค่านี้เพื่อทำการปรับค่าประมาณความแปรปรวนอันเนื่องมาจากการชักตัวอย่างจากประชากรที่มีขนาดจำกัด (finite population) ส่วนคำสั่ง VAR ใช้สำหรับระบุตัวแปรที่ต้องการวิเคราะห์ โดยถ้าตัวแปรเป็นตัวแปรเชิงปริมาณ (เช่น ตัวแปรSpending) คำสั่งนี้จะทำการคำนวณค่าเฉลี่ย แต่ถ้าตัวแปรเป็นตัวแปรที่แบ่งเป็นพวกหรือตัวแปรเชิงกลุ่ม (เช่น ตัวแปร Group) คำสั่งนี้จะทำการคำนวณสัดส่วนในแต่ละกลุ่ม ผลลัพธ์ที่ได้แสดงในภาพที่ 2

The SURVEYMEANS Procedure

Statistics

Variable	N	Mean	Std Error of Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
Spending	40	8.750000	0.845139	7.040545	10.459455
Group=less	23	0.575000	0.078761	0.415690	0.734310
Group=more	17	0.425000	0.078761	0.265690	0.584310

ภาพที่ 2 การวิเคราะห์ค่าใช้จ่ายในการซื้อไอสกรีม แผนแบบการชักตัวอย่างสุ่มแบบง่ายจากประชากรจำกัด

ผลลัพธ์ที่ได้แสดงค่าประมาณสำหรับตัวแปรแต่ละตัว โดยปริยายแล้ว PROC SURVEYMEANS จะแสดงจำนวนค่าสังเกต ค่าประมาณค่าเฉลี่ยและค่าคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยนั้น ค่าขอบเขตความเชื่อมั่นร้อยละ 95 ของการประมาณค่าเฉลี่ยประชากร สำหรับสถิติอื่นๆ ถ้าผู้ใช้ต้องการก็สามารถระบุเพิ่มเติมได้ และเนื่องจากตัวแปร Group มีสองระดับคือ less กับ more ดังนั้นกระบวนคำสั่งนี้จะทำการประมาณสัดส่วนของแต่ละกลุ่ม สรุปได้ว่าโดยเฉลี่ยแล้วค่าใช้จ่ายสำหรับไอสกรีมต่อสัปดาห์ของนักเรียนชั้นมัธยมศึกษาตอนต้นของโรงเรียนนี้ประมาณ 8.75 เหรียญ ค่าคลาดเคลื่อนมาตรฐานของการประมาณนี้คือ 0.84 เหรียญ และสามารถประมาณได้ว่าค่าใช้จ่ายโดยเฉลี่ยในการซื้อไอสกรีมต่อสัปดาห์อยู่ระหว่าง 7.04 ถึง 10.46 เหรียญด้วยความเชื่อมั่นร้อยละ 95 นอกจากนี้พบว่าร้อยละ 57.5 ของนักเรียนทั้งหมดซื้อไอสกรีมโดยเฉลี่ยไม่เกินสัปดาห์ละ 10 เหรียญ ขณะที่ร้อยละ 42.5 ของนักเรียนทั้งหมดซื้อไอสกรีมโดยเฉลี่ยตั้งแต่ 10 เหรียญขึ้นไปต่อสัปดาห์ การประมาณสัดส่วนดังกล่าวแต่ละค่ามีค่าคลาดเคลื่อนมาตรฐานของการประมาณคือ ร้อยละ 7.9

4. การชักตัวอย่างสุ่มแบบมีชั้นภูมิ

ถ้าสมมุติว่านักเรียนจำนวน 40 คนในตัวอย่างก่อนหน้านี้จริงๆ แล้วเป็นตัวอย่างที่ได้จากการชักตัวอย่างสุ่มแบบมีชั้นภูมิ (stratified random sampling) ซึ่งโดยวิธีการนี้ประชากรจะถูกแบ่งออกเป็นชั้นภูมิที่ไม่ซ้ำซ้อนกันก่อน จากนั้นจึงทำการเลือกตัวอย่างจากแต่ละชั้นภูมิอย่างเป็นอิสระกัน

ถ้าประชากรนักเรียนทั้งหมด 4,000 คนถูกแบ่งตามชั้นเรียน หรือชั้นภูมิได้ดังตารางที่ 1 จากนั้นทำการชักตัวอย่างสุ่มแบบง่ายจากแต่ละชั้นภูมิและได้ตัวอย่างรวมทั้งหมด 40 คน [4]

ตารางที่ 1 จำนวนนักเรียน 4,000 คนจำแนกตามชั้นเรียน

ชั้นเรียนหรือชั้นภูมิ	จำนวนนักเรียน
เกรด 7	1,824
เกรด 8	1,025
เกรด 9	1,151
รวม	4,000

การวิเคราะห์ตัวอย่างแบบมีชั้นภูมิจากประชากรที่มีขนาดจำกัดด้วยโปรแกรม SAS จะต้องระบุขนาดของประชากรในแต่ละชั้นภูมิในกระบวนคำสั่ง PROC SURVEYMEANS ในที่นี้จะใช้ชุดข้อมูลของ SAS ชื่อ Student Totals ในการระบุชั้นภูมิและขนาดประชากรของชั้นภูมิ ดังคำสั่งในภาพที่ 3

ตัวแปรชื่อ Grade คือตัวแปรที่ระบุชั้นเรียนหรือชั้นภูมิ ส่วนตัวแปร _TOTAL_ เป็นตัวแปรที่ใช้แทนจำนวนนักเรียนทั้งหมดในแต่ละชั้นหรือขนาดของชั้นภูมิซึ่งจะต้องตั้งชื่อตัวแปรโดยใช้คำว่า _TOTAL_ เท่านั้นเนื่องจากกระบวนคำสั่ง PROC SURVEYMEANS บังคับให้ผู้ใช้เก็บขนาดของชั้นภูมิไว้ที่ตัวแปร

```
DATA StudentTotals;
 INPUT Grade _TOTAL_;
DATALINES;
7  1824
8  1025
9  1151;
```

ภาพที่ 3 คำสั่ง SAS ที่ใช้สร้างชุดข้อมูลที่ระบุค่าของชั้นภูมิและขนาดของชั้นภูมิ

ชื่อ_TOTAL_ นี้ จากนั้นกระบวนคำสั่งจะใช้ขนาดของชั้นภูมิในการปรับค่าประมาณความแปรปรวนอันเนื่องมาจากผลของการชักตัวอย่างจากประชากรที่มีขนาดจำกัด ถ้าผู้ใช้ไม่ระบุขนาดของประชากรหรืออัตราการชักตัวอย่าง (sampling rate) จะไม่มีการปรับแก้ไขเนื่องจากประชากรที่มีขนาดจำกัด (finite population correction) ในการคำนวณสถิติต่างๆ

สำหรับแผนแบบการชักตัวอย่างแบบมีชั้นภูมิ เมื่ออัตราการชักตัวอย่างในชั้นภูมิไม่เท่ากัน จำเป็นต้องมีการคำนวณน้ำหนักของการชักตัวอย่าง (sampling weight) ที่สะท้อนให้เห็นถึงอัตราการชักตัวอย่างที่แตกต่างกันไปในแต่ละชั้นภูมิเพื่อที่จะใช้ในการคำนวณค่าประมาณของค่าเฉลี่ยที่ไม่เอนเอียง ซึ่งในที่นี้น้ำหนักของการชักตัวอย่างที่เหมาะสมได้แก่ ส่วนกลับ (reciprocals) ของค่าความน่าจะเป็นของการเลือกตัวอย่าง [4] การคำนวณน้ำหนักทำได้ดังคำสั่งในภาพที่ 4

```
DATA IceCream;
SET IceCream;
 IF Grade=7 THEN Prob=20/1824;
 IF Grade=8 THEN Prob=9/1025;
 IF Grade=9 THEN Prob=11/1151;
Weight=1/Prob;
```

ภาพที่ 4 คำสั่ง SAS ที่ใช้สร้างตัวแปรที่แทนน้ำหนัก

อนึ่ง ถ้าผู้ใช้ใช้ PROC SURVEYSELECT ในการเลือกตัวอย่าง กระบวนคำสั่งดังกล่าวจะสร้างน้ำหนักในการเลือกตัวอย่างให้โดยที่ผู้ใช้ไม่ต้องคำนวณเอง จากนั้นทำการวิเคราะห์ข้อมูลจากการชักตัวอย่างแบบมีชั้นภูมิตามคำสั่งในภาพที่ 5

```
PROC SURVEYMEANS DATA=IceCream
 TOTAL=StudentTotals;
 STRATUM Grade / LIST;
 VAR Spending Group;
 WEIGHT Weight
RUN;
```

ภาพที่ 5 คำสั่ง SAS ที่ใช้ประมาณค่าเฉลี่ยและสัดส่วนสำหรับแผนแบบการชักตัวอย่างสุ่มแบบมีชั้นภูมิ

ในที่นี้คำสั่ง TOTAL= ใช้สำหรับระบุว่าชุดข้อมูลที่ชื่อ StudentTotals เป็นข้อมูลนำเข้าที่บอกรายละเอียดเกี่ยวกับขนาดของชั้นภูมิ (ต่างจากการชักตัวอย่างสุ่มแบบง่ายที่บอกขนาดรวมของประชากรทั้งหมด) ส่วนคำสั่ง STRATUM หรือ STRATA ใช้ระบุตัวแปรที่ใช้แบ่งชั้นภูมิซึ่งได้แก่ชั้นเรียนหรือตัวแปรชื่อ Grade สำหรับคำสั่งย่อย LIST ในคำสั่ง STRATUM นั้นไม่จำเป็นต้องใช้ แต่ถ้าใช้จะให้ผลลัพธ์ที่แสดงรายละเอียดของแต่ละชั้นภูมิ เช่น

ค่าของตัวแปรที่ใช้แบ่งชั้นภูมิ อัตราการชักตัวอย่างในแต่ละชั้นภูมิเป็นต้น ส่วนคำสั่ง WEIGHT บอกให้โปรแกรมทราบว่าตัวแปรชื่อ Weight เป็นตัวแปรที่ใช้เก็บค่าของน้ำหนักการเลือกตัวอย่าง ผลลัพธ์ที่ได้จากคำสั่งข้างต้นแสดงในภาพที่ 6

จากภาพที่ 5 สรุปได้ว่าโดยเฉลี่ยแล้วค่าใช้จ่ายสำหรับไอสกรีมต่อสัปดาห์ของนักเรียนชั้นมัธยมศึกษาตอนต้นของโรงเรียนนี้ประมาณ 9.14 เหรียญ ค่าคลาดเคลื่อนมาตรฐานของการประมาณ นี้คือ 0.63 เหรียญ และสามารถประมาณได้ว่าค่าใช้จ่ายโดยเฉลี่ยในการซื้อไอสกรีมต่อสัปดาห์อยู่ระหว่าง 8.06 ถึง 10.22 เหรียญ ด้วยความเชื่อมั่นร้อยละ 95 นอกจากนี้พบว่าร้อยละ 54.5 ของนักเรียนทั้งหมดซื้อไอสกรีมโดยเฉลี่ยไม่เกินสัปดาห์ละ 10 เหรียญ ขณะที่ร้อยละ 45.5 ของนักเรียนทั้งหมดซื้อไอสกรีมโดยเฉลี่ยตั้งแต่ 10 เหรียญขึ้นไปต่อสัปดาห์ การประมาณสัดส่วนดังกล่าวแต่ละค่ามีค่าคลาดเคลื่อนมาตรฐานของการประมาณคือ ร้อยละ 5.8

The SURVEYMEANS Procedure

Stratum Information

Stratum Index	Grade	Population Total	Sampling Rate	N Obs	Variable	N
1	7	1824	1.10%	20	Spending	20
					Group=less	17
					Group=more	3
2	8	1025	0.88%	9	Spending	9
					Group=less	0
					Group=more	9
3	9	1151	0.96%	11	Spending	11
					Group=less	6
					Group=more	5

Statistics

Variable	N	Mean	Std Error of Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
Spending	40	9.141298	0.531799	8.063771	10.218825
Group=less	23	0.544555	0.058424	0.426177	0.662932
Group=more	17	0.455455	0.058424	0.337068	0.573823

ภาพที่ 6 การวิเคราะห์ค่าใช้จ่ายในการซื้อไอสกรีม แผนแบบการชักตัวอย่างสุ่มแบบมีชั้นภูมิจากประชากรจำกัด

5. การใช้ PROC MEANS และโปรแกรม SPSS

ในกรณีที่สมมุติว่า ตัวอย่างเป็นตัวอย่างสุ่มจากประชากรที่มีขนาดอนันต์ หรือเป็นตัวอย่างจากแผนแบบการชักตัวอย่างสุ่มแบบง่าย [5] ทั้งๆ ที่จริงๆ แล้วตัวอย่างนั้นถูกเลือกมาจากการใช้แผนแบบการชักตัวอย่างชนิดอื่นซึ่งเป็นการสำรวจจากประชากรที่มีขนาดจำกัด และใช้กระบวนคำสั่ง MEANS หรือทำการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSS ด้วยเมนูคำสั่งที่จัดทำไว้ ผลลัพธ์ที่ได้จะแตกต่างกับการใช้ PROC SURVEYMEANS อย่างไรให้พิจารณาจากผลลัพธ์ที่ได้จากคำสั่งในภาพที่ 7

```
DATA IceCream;
 INPUT Grade Spending @@;
  IF (Spending < 10) THEN Group=1;
  ELSE Group=0;
DATALINES;
7 7 7 7 8 12 9 10 7 1 7 10 7 3 8 20 8 19 7 2
7 2 9 15 8 16 7 6 7 6 7 6 9 15 8 17 8 14 9 8
9 8 9 7 7 3 7 12 7 4 9 14 8 18 9 9 7 2 7 1
7 4 7 11 9 8 8 10 8 13 7 2 9 6 9 11 7 2 7 9
;
PROC MEANS DATA=IceCream
     N MEAN STDERR CLM;
 VAR Spending Group;
RUN;
```

ภาพที่ 7 คำสั่ง SAS ที่ใช้ประมาณค่าเฉลี่ยและสัดส่วนสำหรับตัวอย่างสุ่มจากประชากรที่มีขนาดอนันต์ด้วย PROC MEANS

ผลลัพธ์ในภาพที่ 8 เป็นผลลัพธ์จากการใช้คำสั่งในภาพที่ 7 ที่มีการแปลงค่าตัวแปร Spending ให้เป็น 1 หรือ 0 เพื่อที่จะสามารถคำนวณสัดส่วนของนักเรียนที่ซื้อไอสกรีมโดยเฉลี่ยไม่เกินสัปดาห์ละ 10 เหรียญถ้านำข้อมูลตัวอย่างนักเรียนจำนวน 40 คนไปวิเคราะห์ด้วยโปรแกรม SPPSS โดยเมนูคำสั่ง Analyze ตามด้วย Descriptive Statistics และ Explore จะได้ผลลัพธ์ดังภาพที่ 9 และ 10

6. สรุป

ตารางที่ 2 และ 3 แสดงการเปรียบเทียบค่าประมาณของตัวแปร Spending และ Group ที่ได้จากคำสั่งต่างๆ ของโปรแกรม SAS และโปรแกรม SPSS ในกรณีต่างๆ จะเห็นว่าการใช้ PROC SURVEYMEANS กรณีตัวอย่างสุ่มแบบ

The MEANS Procedure

Variable	N	Mean	Std Error of Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
Spending	40	8.750000	0.8493965	7.031934	10.468066
Group	40	0.575000	0.079158	0.414887	0.735113

ภาพที่ 8 การวิเคราะห์ค่าใช้จ่ายในการซื้อไอสกรีม ด้วย PROC MEANS กรณีตัวอย่างเป็นตัวอย่างสุ่มจากประชากรขนาดอนันต์

Descriptives

			Statistic	Std. Error
SPENDING	Mean		8.75	.85
	95% Confidence Interval for Mean	Lower Bound	7.03	
		Upper Bound	10.47	
	5% Trimmed Mean		8.58	
	Median		8.00	
	Variance		28.859	
	Std. Deviation		5.37	
	Minimum		1	
	Maximum		20	
	Range		19	
	Interquartile Range		8.75	
	Skewness		.352	.374
	Kurtosis		-.803	.733

ภาพที่ 9 การประมาณค่าเฉลี่ยของตัวแปร Spending ด้วยโปรแกรม SPSS

Descriptives

			Statistic	Std. Error
GROUP	Mean		.5750	7.916E-02
	95% Confidence Interval for Mean	Lower Bound	.4149	
		Upper Bound	.7351	
	5% Trimmed Mean		.5833	
	Median		1.0000	
	Variance		.251	
	Std. Deviation		.5006	
	Minimum		.00	
	Maximum		1.00	
	Range		1.00	
	Interquartile Range		1.0000	
	Skewness		-.315	.374
	Kurtosis		-2.003	.733

ภาพที่ 10 การประมาณสัดส่วนของตัวแปร Group ด้วยโปรแกรม SPSS

ง่ายให้ผลไม่แตกต่างจากการใช้ PROC MEANS และการใช้เมนูคำสั่งในโปรแกรม SPSS เนื่องจาก PROC MEANS และ การใช้เมนูคำสั่งในโปรแกรม SPSS สมมุติว่าข้อมูลเป็นตัวอย่างสุ่มจากประชากรที่มีขนาดอนันต์ ซึ่งสอดคล้องกับแผนแบบการชักตัวอย่างสุ่มแบบง่ายในกรณีประชากรมีขนาดจำกัด แต่ถ้าผู้วิจัยเลือกตัวอย่าง โดยใช้แผนแบบการชักตัวอย่างแบบมีชั้นภูมิแล้วใช้ PROC MEANS หรือเมนูคำสั่งในโปรแกรม SPSS ตามความคุ้นเคย ค่าประมาณที่ได้จะไม่ถูกต้องและให้ค่าที่ต่ำกว่าค่าที่ควรจะเป็น (ทั้งในกรณีค่าเฉลี่ยและสัดส่วน แม้ว่าสัดส่วนของนักเรียนที่ใช้เงินซื้อไอสกรีมน้อยกว่า10 เหรียญต่อสัปดาห์มีค่าเท่ากับ 57.5 % โดยการใช้ PROC MEANS แต่สัดส่วนนักเรียนที่ใช้เงินซื้อไอสกรีมตั้งแต่ 10 เหรียญต่อสัปดาห์ขึ้นไปจะมีค่าน้อยกว่าที่ควรจะเป็น) และมีค่าคลาดเคลื่อนมาตรฐานของการประมาณสูงกว่า ทำให้ช่วงเชื่อมั่นร้อยละ 95 ของการประมาณพารามิเตอร์ที่สนใจมีความกว้างมากกว่าช่วงเชื่อมั่นที่มีการปรับให้สอดคล้องกับแผนแบบการชักตัวอย่าง

ตารางที่ 2 การเปรียบเทียบค่าประมาณของตัวแปร Spendingในกรณีต่างๆ

PROC (โปรแกรม SAS) หรือ SPSS	ค่าเฉลี่ย (ค่าคลาดเคลื่อนมาตรฐาน)	ช่วงความเชื่อมั่น ร้อยละ 95
SURVEYMEANS		
- สุ่มแบบง่าย	8.75 (0.8451)	(7.04 , 10.46)
- มีชั้นภูมิ	9.14 (0.5318)	(8.06 , 10.22)
MEANS	8.75 (0.8494)	(7.03 , 10.47)
โปรแกรม SPSS	8.75 (0.8494)	(7.03 , 10.47)

ตารางที่ 3 การเปรียบเทียบค่าประมาณของตัวแปร Group ในกรณีต่างๆ

PROC (โปรแกรม SAS) หรือ SPSS	สัดส่วน (ค่าคลาดเคลื่อนมาตรฐาน)	ช่วงความเชื่อมั่น ร้อยละ 95
SURVEYMEANS		
- สุ่มแบบง่าย	57.5 % (7.9 %)	(41.6 % , 73.4 %)
- มีชั้นภูมิ	54.5 % (5.8 %)	(42.6 % , 66.3 %)
MEANS	57.5 % (7.9 %)	(41.5 % , 73.5 %)
โปรแกรม SPSS	57.5 % (7.9 %)	(41.5 % , 73.5 %)

การเลือกใช้คำสั่งในโปรแกรมสำเร็จรูปให้เหมาะสมและถูกต้องตามข้อตกลงเบื้องต้นของการวิเคราะห์ของวิธีการทางสถิตินั้นเป็นสิ่งที่สำคัญ หากผู้ใช้โปรแกรมสำเร็จรูปทางสถิติยังคงใช้ตามความเคยชิน ไม่พัฒนาตนเองให้มีความรู้ตามความก้าวหน้าของวิธีการทางสถิติและเทคโนโลยีใหม่ๆ ผลสรุปที่ได้จากการศึกษาจากตัวอย่างจะให้ค่าที่มีความถูกต้องเชื่อถือน้อยลง ยังผลให้มีการตัดสินใจผิดเนื่องจากการเลือกใช้เทคนิคการวิเคราะห์ทางสถิติและโปรแกรมสำเร็จรูปที่ไม่ถูกต้อง ซึ่งเป็นสิ่งที่ควรระมัดระวังเป็นอย่างยิ่งหากการตัดสินใจในเรื่องนั้นๆ นำไปสู่การเปลี่ยนแปลงด้านนโยบายที่มีผลต่อสังคมและเศรษฐกิจของชาติ

7. เอกสารอ้างอิง

[1] What's New in Data Analysis : Sample Survey Design and Analysis, http://support.sas.com/rnd/app/da/new/dasurvey.html,2005. (January, 10, 2005)

[2] Berglund, P. A., Analysis of Complex Sample Survey Data Using the SURVEYMEANS and SURVEYREG Procedures and Macro Coding. http://www2.sas.com/proceedings/sugi27/p263-27.pdf (January, 10, 2005)

[3] The SURVEYLOGISTIC Procedure. http://support.sas.com/onlinedoc/913/getDoc/th/statug.hlp/surveylogistic_sect1.htm, 2003. (January, 10, 2005)

[4] The SURVEYMEANS Procedure. **http://support.sas.com/onlinedoc/913/getDoc/ th/statug.hlp/surveymeans_index.htm, 2003. (January, 10, 2005)**

[5] An A., and Watts, D. New SAS Procedures for Analysis of Sample Survey Data., http://www.ats.ucla.edu/stat/sas/library/svy_survey.pdf(January, 10, 2005)