บทที่ ๔

การรวบรวมและวิเคราะห์ข้อมูล

ในการสำรวจด้วยตัวอย่าง ค่าที่ได้เป็นเพียงค่าประมาณ ถ้าสามารถควบคุมความคลาดเคลื่อนให้มีค่าน้อยสุด ค่าที่ได้จะเป็นค่าที่ถือว่าใกล้ความจริงมาก

๔.๑ วิธีเปรียบเทียบตัวประมาณค่า

การที่จะทราบว่าตัวประมาณค่าตัวใดให้ความเชื่อถือได้ หรือมีประสิทธิภาพเพียงไรนั้นอาจพิจารณาได้ดังนี้

ก. ประสิทธิภาพสัมพัทธ์ เป็นการเปรียบเทียบค่าความแปรปรวนของตัวประมาณค่าต่าง ๆ เพื่อจะดูว่าตัวประมาณค่าตัวใดดีกว่า หรือให้ความเชื่อถือได้มากกว่ากัน {เช่น} เปรียบเทียบตัวประมาณค่า A และ B ของพารามิเตอร์เดียวกัน

$$\text{ประสิทธิภาพสัมพัทธ์ระหว่าง A และ B} = \frac{V(B)}{V(A)} \times 100$$

ถ้าค่าของประสิทธิภาพสัมพัทธ์ที่คำนวณได้เกิน ๑๐๐ % แสดงว่าตัวประมาณค่า A ดีกว่า B ถ้าค่าที่ได้ต่ำกว่า ๑๐๐ % แสดงว่าตัวประมาณค่า B ดีกว่า A

ข. สัมประสิทธิ์ของความแปรปรวน (จะใช้สัญลักษณ์ว่า C.V.) เป็นวิธีการตรวจสอบหรือประเมินขนาดของความแปรปรวนกับตัวประมาณค่าที่ได้มีค่าคลาดเคลื่อนมากหรือน้อย ในกรณีที่ตัวประมาณค่ามีหน่วยต่างกัน การเปรียบเทียบจะพิจารณาดูจากค่าสัมประสิทธิของความแปรปรวน C.V. ของค่าประมาณของผลรวมจะมีค่าเป็น

$$C.V.(\hat{T}) = \frac{\sqrt{V(\hat{T})}}{\hat{T}} \times 100$$

ถ้าค่า C.V. ที่หาได้มีค่ามากแสดงว่าตัวประมาณค่านั้นจะให้ความถูกต้องน้อยกว่าตัวประมาณค่าที่มี C.V. น้อย

๔.๒ แหล่งที่มาของข้อมูลและการเก็บข้อมูล

ข้อมูลที่จะนำมาใช้เพื่อเปรียบค่าประมาณความแปรปรวนของตัวประมาณค่าโดยวิธีตรงและวิธีลัดนั้น เป็นข้อมูลเกี่ยวกับแรงงานซึ่งได้มากจากรายงานการสำรวจแรงงานปี ๒๕๑๔ สำนักงานสถิติแห่งชาติ เพราะข้อมูลของงานสำรวจแรงงานได้มากจากวิธีการสำรวจด้วยตัวอย่างขนาดใหญ่

แผนแบบการสุ่มตัวอย่างงานสำรวจแรงงานปี ๒๕๑๔

โครงการสำรวจแรงงานนี้ได้จัดทำทั่วราชอาณาจักร ดังนั้นแผนแบบเลือกตัวอย่างจึงเป็นแบบการจัดสตรา ๓ ขั้น โดยแบ่งประเทศไทยเป็น ๔ ภาค คือ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง ภาคใต้ แต่ละภาคถือเป็นสตราตา ยกเว้นภาคกลางแยกจังหวัดพระนครและธนบุรี (สมัยนั้นยังไม่เรียกกรุงเทพมหานคร) ออกเป็นสตราตาหนึ่ง รวมเป็น ๕ สตราตา สตราตาในที่นี้คือกลุ่มของจังหวัด ดังนั้นจึงมีจังหวัดเป็นหน่วยตัวอย่างขั้นที่ ๑ เขตสำมะโน หรือชุมรุมอาคาร เป็นตัวอย่างขั้นที่ ๒ สำหรับในเขตเทศบาล และหมู่บ้านสำหรับนอกเขตเทศบาล ครัวเรือนเป็นหน่วยตัวอย่างขั้นที่ ๓ การเลือกตัวอย่างในแต่ละขั้นเป็นดังนี้

การเลือกตัวอย่างขั้นที่ ๑

เลือกจังหวัดตัวอย่างของแต่ละภาคโดยให้ความน่าจะเป็นของการสุ่มเป็นสัดส่วนกับจำนวนหลังคาเรือนในจังหวัด ใช้สถิติจำนวนหลังคาเรือนในเขตเทศบาลตามบัญชีเขตสำมะโนของหน่วยแผนที่สำนักงานสถิติแห่งชาติซึ่งจัดเตรียมสำหรับงานสำมะโนประชากรและเคหะ ๒๕๑๓ ส่วนนอกเขตเทศบาลใช้สำรวจหมู่บ้าน พ.ศ. ๒๕๑๑ พระนคร – ธนบุรี เลือกต่างหากโดยให้ความน่าจะเป็นของการสุ่มเท่ากับหนึ่งจำนวนจังหวัดตัวอย่างรวมทั้งสิ้น ๓๑ จังหวัด และจำนวนจังหวัดตัวอย่างในแต่ละภาคเป็นสัดส่วนกับจำนวนจังหวัดทั้งหมดในภาค

การเลือกตัวอย่างขั้นที่ ๒ แบ่งเป็น ๓ พวกคือ

๑. ในเขตเทศบาลพระนคร ธนบุรี เลือก ๓๕๑ ชุมรุมอาคารจากจำนวนทั้งสิ้น ๑๙๖๘ ชุมรุมอาคาร โดยให้ความน่าจะเป็นของการสุ่มเป็นสัดส่วนกับจำนวนหลังคาเรือนในชุมรุมอาคาร

๒. ในเขตเทศบาลอื่น ๆ อีก ๒๘ จังหวัด เลือกชุมรุมอาคารตัวอย่างจำนวนหนึ่งจากแต่ละเขตเทศบาลโดยวิธีเดียวกับในเขตเทศบาลพระนคร ธนบุรี จำนวนทั้งสิ้น ๓๑๑ ชุมรุมอาคารจาก ๓๖๘๔ ชุมรุมอาคาร จำนวนตัวอย่างในแต่ละจังหวัดเป็นสัดส่วนกับจำนวนชุมรุมอาคารในจังหวัดนั้น ๆ

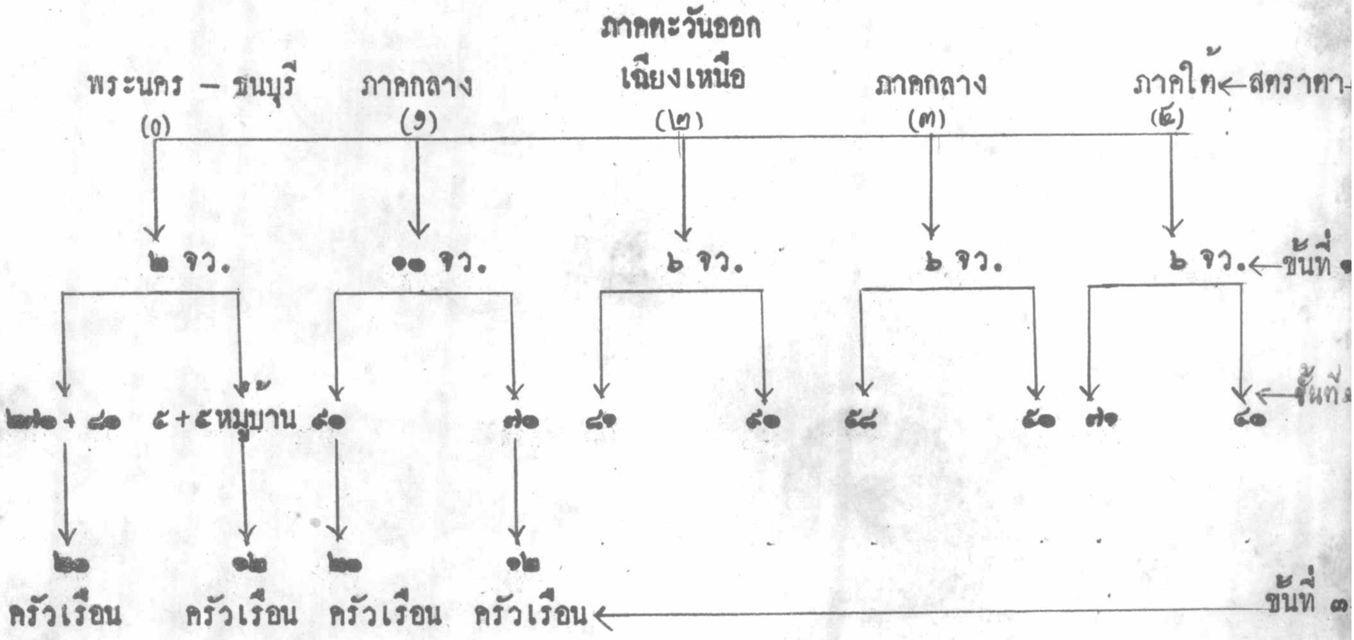
๓. <u>นอกเขตเทศบาล</u> เลือกหมู่บ้านตัวอย่างทั้งสิ้น ๒๖๑ หมู่จากรายชื่อหมู่บ้านตัวอย่างโครงการสำรวจเนื้อที่และพืชผลการเกษตร ปีการเพาะปลูก ๒๕๑๓/๑๔ โดยวิธีการสุ่มอย่างมีระบบ-หมู่บ้านตัวอย่างเดิมของการสำรวจผลิตผลเกษตร เลือกมาแต่ละอำเภอโดยให้ความน่าจะเป็นของการเลือกหมู่บ้านเป็นสัดส่วนกับจำนวนครัวเรือนเกษตร หรือครัวเรือนที่ปลูกข้าวแล้วแต่กรณีจากหมู่บ้านทั้งหมดประมาณ ๔๔,๘๙๑ หมู่บ้าน

<u>การเลือกตัวอย่างขั้นที่ ๓</u>

๑. <u>ในเขตเทศบาล</u> เลือกครัวเรือนส่วนบุคคลตัวอย่างมาชุมรุมอาคารละ ๒๕ ครัวเรือน (รวมสำรอง ๕ ครัวเรือน) อย่างมีระบบ ถ้าชุมรุมอาคารใดมีครัวเรือนพิเศษไม่เกิน ๕ ครัวเรือนก็สำรวจทั้งหมด ถ้าหากเกินกว่า ๕ จะเลือกเพียง ๕ ครัวเรือนอย่างมีระบบเช่นกัน สำหรับเขตเทศบาลพระนคร – ธนบุรี ใช้เลือกจากรายชื่อหัวหน้าครัวเรือนซึ่งนับจดใหม่ แต่ในเขตเทศบาลอื่น ๆ ไม่ได้ทำการนับจดใหม่ คงใช้แบบนับจดของงานสำมะโนประชากรและเคหะ ๒๕๑๓ (สปค.๑) เลือกครัวเรือนตัวอย่างซึ่งบางเขตแจงนับไม่ได้แยกนับจดเป็นรายชุมรุมอาคารจึงต้องเลือกทั้งเขตแจงนับซึ่งถือเสมือนว่าเป็นหนึ่งชุมรุมอาคาร

๒. <u>นอกเขตเทศบาล</u> เลือก ๑๔ ครัวเรือนตัวอย่าง (รวมสำรอง ๒ ครัวเรือน) โดยวิธีการสุ่มอย่างมีระบบจากแบบนับจดของงานสำรวจผลิตผลเกษตร ปีเพาะปลูก ๒๕๑๓/๑๔ (สคข.) และจากแบบ สปค.๑ บางส่วน ดังนั้นแผนแบบของงานสำรวจแรงงานเขียนเป็นแผนผังได้ดังข้างล่างนี้

ภาคเป็นสตราตา (๔ + ๑)



ข้อมูลที่จะนำมาใช้ในการประมาณค่าความแปรปรวนของตัวประมาณค่า เพื่อเปรียบเทียบค่าประมาณโดยวิธีตรงและวิธีลัด เป็นข้อมูลจำนวนผู้มีงานทำของพระนคร – ธนบุรี เฉพาะในเขตเทศบาล

๔.๓ <u>การประมาณค่าความแปรปรวนด้วยวิธีตรง</u>

ใช้วิธีการประมาณค่าตามแผนแบบการจัดสุ่มตัวอย่างแบบ ๒ ขั้น

ให้ T(X) เป็นค่าของจำนวนผู้มีงานทำทั้งหมดในพระนคร – ธนบุรี (สตราตาที่ ๑)

$$T(X) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} x_{ij}$$

N = จำนวนชุมรุมอาคารทั้งหมด

M_i = จำนวนครัวเรือนจากแบบนับจดของชุมรุมอาคารที่ i

X_{ij} = จำนวนผู้มีงานทำของครัวเรือนที่ j ในชุมรุมอาคารที่ i

ดังนั้นค่าประมาณของยอดรวมอาจเขียนได้เป็น

$$\hat{T}(X) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{P_i} \frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij} \qquad (1)$$

n = จำนวนชุมรุมอาคารตัวอย่างทั้งสิ้น

P_i = ความน่าจะเป็นในการสุ่มชุมรุมอาคารตัวอย่างที่ i

m_i = จำนวนครัวเรือนส่วนบุคคลตัวอย่างของชุมรุมอาคารที่ i

สูตรการหาความแปรปรวนของค่าประมาณผลรวมคือ

$$V[\hat{T}(x)] = \frac{1}{n}\left[\sum_{i=1}^{N} P_i \left(\frac{X_{i.}}{P_i} - T\right)^2 + \sum_{i=1}^{N} \frac{M_i^2}{P_i} \frac{\sigma_i^2}{m_i} \left(\frac{M_i - m_i}{M_i - 1}\right)\right]$$

แต่ในทางปฏิบัติเราไม่สามารถหาค่าความแปรปรวนตัวนี้ได้จริง ๆ เพราะจะติดค่าของ $\sigma_x^2 = \sum_{i=1}^{N} P_i \left(\frac{X_{i.}}{P_i} - T\right)^2$ ซึ่งเป็นค่าของความแปรปรวนของประชากร ดังนั้นในการคำนวณเราจะใช้ค่าประมาณค่าความแปรปรวนโดยอาศัยข้อมูลจากการสำรวจด้วยตัวอย่างคือ

$$\hat{V}(\hat{T}(x)) = \frac{1}{n(n-1)}\left[\sum_{i=1}^{n}\left\{\left(\frac{M_i}{p_i m_i}\sum_{j=1}^{m_i} x_{ij}\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{p_i m_i}\sum_{j=1}^{m_i} x_{ij}\right)\right\}^2\right]$$

$$= \frac{1}{n(n-1)}\sum_{i=1}^{n}(z_i - \bar{z})^2 \qquad (2)$$

ในเมื่อ $z_i = \dfrac{M_i}{p_i m_i}\sum_{j=1}^{m_i} x_{ij}$

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{p_i m_i}\sum_{j=1}^{m_i} x_{ij}$$

$$C.V.(\hat{T}(x)) = \frac{\sqrt{\hat{V}(\hat{T}(x))} \times 100}{\hat{T}(x)}$$

จากข้อมูลงานสำรวจแรงงาน ๒๕๐๘ และสูตรการประมาณผล (1) ได้ค่าประมาณผลรวมของกรุงเทพฯ – ธนบุรี ในเขตเทศบาลเป็น

$$\hat{T}_h(x) = \frac{1}{n_h}\sum_{i=1}^{n_h}\frac{1}{p_{hi}}\frac{M_{hi}}{m_{hi}}\sum_{j=1}^{m_{hi}} x_{hij}$$

$$= \frac{242,978,894}{350} \quad ; \quad n_h = 350$$

$$= 694,225$$

นั่นคือจำนวนคนมีงานทำในเขตเทศบาลของกรุงเทพฯ – ธนบุรี = 694,225 คน

ค่าประมาณค่าความแปรปรวนของยอดรวมตามสูตร (2) ซึ่งทำให้ง่ายขึ้น

$$\hat{V}(\hat{T}_h(x)) = \frac{1}{n_h(n_h-1)}\sum_{i=1}^{n_h}(z_i^2 - 2z_i\bar{z} + \bar{z}^2)$$

$$\hat{V}(\hat{T}_h(x)) = \frac{1}{n_h(n_h-1)}\left[\sum_{i=1}^{n_h} z_i^2 - n_h\bar{z}^2\right]$$

$$\text{ซึ่ง } \sum_{i=1}^{n_h} z_i^2 = \sum_{i=1}^{35}\left[\frac{1}{p_{hi}}\frac{M_{hi}}{m_{hi}}\sum_{j=1}^{m_{hi}} x_{hij}\right]^2$$

$$= 253,105,149,972,413$$

$$\text{และ } n_h\bar{z}^2 = 350\left[\frac{1}{n_h}\sum_{i=1}^{n_h}\frac{1}{p_{hi}}\frac{M_{hi}}{m_{hi}}\sum_{j=1}^{m_{hi}} x_{hij}\right]^2$$

$$= 168,682,214,293,250$$

ดังนั้นค่าประมาณค่าความแปรปรวนของยอดรวมเป็น

$$\hat{V}(\hat{T}_h) = \frac{1}{(350)(349)}\left[253,105,149,972,413 - 168,682\right.$$
$$\left.214,293,250,609,274,962\right]$$

มีค่าความคลาดเคลื่อนมาตรฐานเป็น

$$\hat{\sigma}_{\hat{T}_h(x)} = 24,684$$

และ $$C.V. = \frac{\hat{\sigma}_{\hat{T}(x)}}{\hat{T}_h(x)} \times 100$$

$$= \frac{24,684}{694,225} \times 100$$

$$= 3.56 \%$$

๔.๔ การประมาณค่าความแปรปรวนด้วยวิธีลัด

วิธีลัดต่าง ๆ ที่กล่าวมาในบทที่ ๓ นั้น มี ๒ วิธีคือ วิธีจัดหมู่แบบสุ่มและวิธีการสุ่มอย่างง่ายที่จะนำมาใช้กับงานสำรวจแรงงานทั้งนี้โดยพิจารณาจากความเหมาะสมในการที่จะประยุกต์เข้ากับแผนแบบการสุ่มตัวอย่างของงานสำรวจแรงงานเป็นเกณฑ์ ดังนั้นจึงขอกล่าวถึงเฉพาะ ๒ วิธีนี้เท่านั้นในการประมาณค่าความแปรปรวนของค่าประมาณยอดรวม

๔.๔.๑ วิธีการจัดหมู่แบบสุ่ม

ตามแผนแบบการสุ่มตัวอย่างของสำรวจแรงงาน ๒๕๑๔ นั้นมีจำนวนครัวเรือนตัวอย่างในเขตเทศบาลพระนคร – ธนบุรี ทั้งสิ้น ๖๙๘๐ ครัวเรือน โดยวิธีการจัดหมู่แบบสุ่มนี้แบ่งครัวเรือนทั้งหมดออกเป็น ๑๐ กลุ่มด้วยวิธีที่ ๓ ของ ๓.๒.๑ ดังนั้นในแต่ละกลุ่มจะประกอบด้วยครัวเรือนตัวอย่างจำนวน ๖๙๘ ครัวเรือน

ก. ค่าประมาณยอดรวม

ตามสูตรในข้อ ๓.๒.๑ ของบทที่ ๓ ได้ค่าประมาณยอดรวมของผู้มีงานทำในเขตเทศบาล พระนคร – ธนบุรีเป็น

$$\hat{T}_h(x) = \frac{N_h}{n_h} \sum_{g=1}^{t} x_{hg} \qquad (1)$$

x_{hg} = ผลรวมของผู้มีงานทำจากตัวอย่างในหมู่ที่ g ในสตราตา h

t = จำนวนหมู่ทั้งหมด = 10

N_h = ค่าประมาณจำนวนครัวเรือนทั้งหมดในสตราตาที่ h

n_h = จำนวนครัวเรือนตัวอย่างในสตราตาที่ h

จากข้อมูลสำรวจแรงงานของพระนคร – ธนบุรี (เฉพาะในเขต) ได้ค่าต่าง ๆ ตามสูตร (1) ดังนี้

N_h = 417,705[1] ครัวเรือน

n_h = 6,980 ครัวเรือน

$\sum_{g=1}^{t} x_{hg}$ = 14,385

$$\hat{T}_h(X) = \frac{417,705}{6,980} \times 14,385$$

$$= 860,843$$

นั่นคือจำนวนผู้มีงานทำทั้งหมดในเขตเทศบาลพระนคร – ธนบุรี มีประมาณ 860,843 คน

---

[1] ดูในภาคผนวก

ข. <u>ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม</u>

จากสูตรในข้อ ๓.๒.๑ ของบทที่ ๓ ค่าประมาณความแปรปรวนของยอดรวมตามข้อ ก. เป็น

$$\hat{V}(\hat{T}_h) = \frac{N_h^2}{n_h^2} t \sum_{g=1}^{t} \frac{(X_{hg} - \bar{X}_h')^2}{k(t-1)}$$

เมื่อ $\bar{X}_h' = \sum_{g=1}^{t} X_{hg} / t$

และ k = จำนวนครัวเรือนตัวอย่างในแต่ละหมู่

จากข้อมูลการสำรวจแรงงานของเขตเทศบาลพระนคร – ธนบุรี ภายหลังการจัดหมู่แบบสุ่มแล้วสรุปได้ดังตารางข้างล่างนี้

ตารางแสดงค่าผลรวมและค่าผลต่างกำลังสอง เพื่อใช้ในการคำนวณวิธีลัด

	หมู่ที่									
	1	2	3	4	5	6	7	8	9	10
X_{hg}	1,449	1,463	1,491	1,472	1,401	1,414	1,406	1,402	1,476	1,411
$(X_{hg} - \bar{X}_{hg})^2$	110.25	600.25	2,756.25	2,910.50	1,406.25	600.25	1,056.25	1,332.25	1,406.25	756.25

เมื่อแทนค่าต่าง ๆ ลงในสูตรข้างต้นนี้จะได้ค่าประมาณความแปรปรวน ดังนี้

$$\hat{V}(\hat{T}(x)) = \left(\frac{417,715}{6,980}\right)^2 \sum_{g=1}^{10} \frac{(x_{hg} - \bar{x}'_h)^2}{698(10-1)}$$

$$= \frac{(417,715)^2 \times (12,934.75)}{6,980 \times 698 \times 9}$$

$$= 51,468,793$$

$$\sigma_{\hat{T}_h(x)} = 7,174$$

$$c.v. = \sigma_{\hat{T}_h(x)} \times 100 / \hat{T}_h(x)$$

$$= \frac{7,174 \times 100}{860,843}$$

$$= 0.83 \%$$

๔.๔.๒ วิธีการสุ่มอย่างง่าย

ตามแผนแบบการสุ่มตัวอย่างของงานสำรวจแรงงาน ๒๕๑๔ มีจำนวนครัวเรือนตัวอย่างทั้งหมด ๖๙๘๐ ครัวเรือน ซึ่งจะนำมาใช้ในการวิเคราะห์

ก. ค่าประมาณยอดรวม

ตามสูตรที่ ๑ ของ ๓.๒.๒ ได้ค่าประมาณยอดรวมเป็น

$$\hat{T}_h(x) = \frac{N_h}{n_h} \sum_{j=1}^{n_h} x_{hj}$$

x_{hj} = จำนวนผู้มีงานทำของครัวเรือนตัวอย่างที่ j ในเขตเทศบาลพระนคร – ธนบุรี

N_h = ค่าที่คาดหมายของจำนวนครัวเรือนในปี ๒๕๑๔ ในเขตเทศบาลพระนคร – ธนบุรี

= 417,705

n_h = จำนวนครัวเรือนตัวอย่างทั้งหมดในเขตเทศบาลพระนคร – ธนบุรี

= 6,980

$$\hat{T}_h(x) = \frac{417,705}{6,980} \sum_{j=1}^{6980} x_{hj}$$

$$= \frac{417,705}{6,980} \times 14,385$$

$$= 860,843$$

ข. <u>ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม</u>

จากสูตรที่ ๒ ของ ๓.๒.๒ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวมเป็น

$$\hat{V}(\hat{T}_h(x)) = \frac{1}{n_h(n_h-1)} \left[N_h^2 \sum_{j=1}^{n_h} x_{hj}^2 - n_h \hat{T}_h^2 \right]$$

$$\sum_{j=1}^{6980} x_{hj}^2 = 48,834$$

$$N_h^2 \sum_{j=1}^{n_h} x_{hj}^2 = (417,705)^2 (417,013) = 8,202,709,157,246,325$$

$$n\hat{T}_h^2 = (6980)(860,843)^2 = 5,172,533,681,130,020$$

$$\hat{V}(\hat{T}_h(x)) = \frac{1}{6,980 \times 6,979} \left[8,202,709,157,246,325 - 5,172,533,681,130,020 \right]$$

$$= \frac{1}{48,713,420} \times 3,030,175,476,116,305$$

$$= 62,204,121$$

$$\hat{\sigma}_{\hat{T}_h(x)} = 7,886.96$$

$$C.V. = \frac{\hat{\sigma}_{\hat{T}_h(x)} \times 100}{\hat{T}_h(x)}$$

$$= (7,886.96 \times 100) / 860,843$$

$$= 0.92 \%$$

๔.๕ ผลการเปรียบเทียบวิธีตรงกับวิธีลัด

จากการประมาณค่าความแปรปรวนของตัวประมาณค่าด้วยวิธีตรงและวิธีลัดซึ่งประยุกต์กับงานสำรวจแรงงาน ๒๕๑๔ สรุปผลได้ดังตารางข้างล่างนี้

วิธีการประมาณ	จำนวนคนมีงานทำ ($\hat{T}(x)$)	Var $\hat{T}(x)$	c.v ($\hat{T}(x)$)(%)
วิธีตรง	694,225	609,274,962	3.56
วิธีลัด :-			
วิธีการสุ่มอย่างง่าย	860,843	62,204,120	0.92
วิธีการจัดหมู่แบบสุ่ม	860,843	51,468,793	0.83

เมื่อพิจารณาค่าประมาณของความแปรปรวนและค่าสัมประสิทธิ์ของความแปรปรวนของค่าประมาณผู้มีงานทำทั้งหมดในเขตเทศบาลพระนคร – ธนบุรี ปรากฏว่าวิธีตรงให้ค่าสูงกว่าวิธีลัดทั้งสองวิธีซึ่งแสดงว่าค่าประมาณที่ใช้วิธีลัดมีประสิทธิภาพดีกว่าและให้ความเชื่อถือได้มากกว่าวิธีตรง แต่เป็นการขัดแย้งกับทฤษฎีทางสถิติว่าวิธีตรงจะเป็นวิธีที่ให้ค่าประมาณใกล้เคียงความจริงมากกว่าวิธีลัด ดังนั้นจึงควรทำการวิเคราะห์เพื่อหาสาเหตุต่อไป

ระหว่างวิธีลัดทั้งสองวิธีนี้ วิธีการจัดหมู่แบบสุ่มจะให้ค่าสัมประสิทธิ์ความแปรปรวนน้อยกว่า แต่เมื่อพิจารณาถึงความสะดวกในการคำนวณ ง่ายและให้ผลรวดเร็ว วิธีการสุ่มอย่างง่ายย่อมดีกว่า- -การจัดหมู่แบบสุ่ม

วิเคราะห์ผล

เนื่องจากค่าประมาณยอดรวมผู้มีงานทำโดยวิธีตรงนั้นต่ำกว่าวิธีลัดทั้งสองวิธีซึ่งเป็นการขัดกับหลักทฤษฎีดังที่กล่าวมาแล้ว ทั้งนี้อาจเนื่องมาแต่ frame ที่ใช้ในการเลือกชุมรุมอาคารและการนับจดที่ได้มานั้นไม่สอดคล้องกัน พิจารณาจากสูตรการประมาณข้างล่างนี้

$$\hat{T}_h(x) = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{1}{P_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} x_{hij}$$

พบว่าตัวที่จะทำให้กระเทือนค่าประมาณก็คือ $\frac{1}{P_{hi}} \cdot M_{hi}$ ซึ่ง $P_{hi} = \frac{M'_{hi}}{\sum_i M'_{hi}}$

หรือเขียนใหม่เป็น $K = \frac{\sum M'_{hi}}{M'_{hi}} \cdot M_{hi} = \sum M'_{hi} \cdot F$

เมื่อ F คืออัตราส่วนระหว่าง $\frac{M_{hi}}{M'_{hi}}$

M'_{hi} = เป็นจำนวนครัวเรือนในชุมรุมอาคารที่ i ที่ได้จาก frame (ข้อมูลจากสำมะโนประชากร ๒๕๑๓)

$\sum M'_{hi}$ = เป็นจำนวนครัวเรือนของชุมรุมอาคารทั้งหมดที่ได้จาก frame

M_{hi} = เป็นจำนวนครัวเรือนที่ได้จากแบบนับจดปี ๒๕๑๔

ถ้า $M'_{hi} = M_{hi}$ แล้ว $F = 1$ นั่นคือ $K = \sum M'_{hi}$

หมายถึงจำนวนครัวเรือนไม่ได้เพิ่มขึ้นเลย ซึ่งไม่น่าจะเป็นไปได้ และทำนองกลับกัน ถ้า $M'_{hi} > M_{hi}$ แล้ว $F < 1$ นั่นคือ $K < \sum M'_{hi}$ หมายถึงจำนวนครัวเรือนลดลง โดยทฤษฎีแล้วไม่ควรจะเป็น แต่ในทางปฏิบัติแล้วปรากฏว่าเป็นจริงดังตารางในหน้าถัดไป

ตารางแสดงการเปรียบขนาดชุมรุมอาคารตัวอย่างของ frame และแบบนับจด

ขนาดชุมรุมอาคารจาก Frame	ขนาดชุมรุมอาคารตัวอย่างจากแบบนับจด										
	1-50	51-100	101-150	151-200	201-250	251-300	301-350	351-400	401-450	451-500	> 500
1-50	//	卌 /	//	/							
51-100	卌 //	卌 卌 卌 卌 卌 卌 ////	卌 ////	/							
101-150	//	卌 卌 卌 卌 卌 卌 //	卌 卌 卌 卌 卌 卌 卌 卌 卌 //	卌 //	//	/	/	/			
151-200	/	卌 ///	卌 卌 卌 卌 卌 //	卌 卌 卌 卌 卌 //	//						
201-250	//	////	卌 //	卌 ///	卌 //	///	//	/	/		
250-300		/	/	卌	卌	//	/				
301-350				//	//					/	/
351-400										/	
401-450					/						
451-500			/	/	/						
> 500								/	/		

ดังข้อมูลในตารางซึ่งมีจำนวนชุมรุมอาคารตัวอย่างเป็นจำนวน ๑๒๐ ชุมรุมอาคารที่มีจำนวนครัวเรือนลดลงในขณะที่มีจำนวนชุมรุมอาคารเพียง ๔๔ ชุมรุมอาคารที่มีครัวเรือนเพิ่มขึ้น

นั่นคือสรุปแล้วชุมรุมอาคารเป็นจำนวนมากมีขนาดลดลง นั่นก็หมายความว่า จำนวนครัวเรือนทั้งหมด ($\sum M'_{hi}$) มีแนวโน้มที่จะลดลงด้วย ดังนั้นผลคูณของ $\frac{\sum_{j=1}^{m_{hi}} x_{hij}}{m_{hi}}$ กับ $\sum M'_{hi}$ ก็ลดลงด้วยซึ่งทำให้ค่าของ $\hat{T}_h(X)$ ลดลง

กลับมาพิจารณาค่าประมาณยอดรวมโดยวิธีลัดทั้งสอง จากสูตรข้างล่างนี้

$$\hat{T}_h(x) = \frac{N_h}{n_h}\sum_{g=1}^{t} x_{hg}$$

$$\hat{T}_h(x) = \frac{N_h}{n_h}\sum_{j=1}^{n_h} x_{hj}$$

ปรากฏว่าค่าประมาณยอดรวมทั้งสองนี้คำนวณขึ้นโดยไม่มีค่า P_1 เข้ามาเกี่ยวข้องด้วย ดังนั้นไม่ว่าขนาดของชุมรุมอาคารจะเปลี่ยนแปลงไปอย่างไรก็ตามก็จะไม่มีผลกระทบกระเทือนต่อค่าประมาณเหมือนดังที่เกิดขึ้นกับการใช้สูตรโดยวิธีตรง

เมื่อพิจารณาค่าของสัมประสิทธิ์ของความแปรปรวนโดยวิธีตรงนั้นมีค่าใหญ่กว่าโดยวิธีลัด ถ้าสมมติว่าค่าประมาณของวิธีตรงถูกต้องใกล้ความจริงเมื่อค่า $F > 1$ นั่นคือค่าประมาณยอดรวมสูงขึ้น ค่า C.V. อาจลดลงถ้าค่าประมาณความแปรปรวนไม่เปลี่ยนแปลง ยกเว้นแต่ค่าประมาณความแปรปรวนเปลี่ยนแปลงไปในทางเพิ่มขึ้น ค่า C.V. อาจจะสูงกว่าเดิมทั้งนี้ขึ้นกับว่าปริมาณที่เพิ่มขึ้นของค่าประมาณยอดรวมและค่าประมาณของความแปรปรวน จำนวนใดจะมากกว่ากัน ซึ่งไม่สามารถสรุปได้ในที่นี้เพราะไม่มีข้อมูลจะมาประมาณได้