Sample มาใช้ เพราะถ้าใช้ทุก Possible Sample งานนั้นจะมิใช่งานสำรวจด้วยกลุ่มตัวอย่าง[1] แต่ S^2_{wsy} เป็นค่าความแปรปรวนที่ต้องวิเคราะห์โดยอาศัยข้อมูลจากทุก Possible Sample (สังเกตสูตร S^2_{wsy}) ดังนั้น เมื่อเลือกตัวอย่างมาเพียง 1 ชุดจึงไม่อาจวิเคราะห์หา S^2_{wsy} แม้ข้อมูลชุดเดียวกันนี้จะใช้คำนวณหา s^2 ซึ่งเป็นค่าประมาณที่ปราศจากอคติของ S^2 ได้ก็ตาม ผลลัพธ์ก็คือเราไม่อาจหาค่าประมาณของ $V(\bar{x}_{sy})$ ได้จากโครงสร้างเดิมนี้

2. จากสูตร

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsy}$$

จะพบว่า $V(\bar{x}_{sy})$ จะมีค่าน้อยถ้าหากว่า S^2_{wsy} มีค่ามากเพราะ $V(\bar{x}_{sy})$ เสนอไว้ในรูปของผลต่างระหว่าง $\frac{N-1}{N} S^2$ กับ $\frac{k(n-1)}{N} S^2_{wsy}$ ปัญหาก็คือจะทำอย่างไรจึงจะมีผลให้ S^2_{wsy} มีค่าสูงขึ้นได้

จากความรู้เดิมเราทราบว่าถ้ากลุ่มตัวอย่างนั้นมีหน่วยตัวอย่างที่ละม้ายคล้ายคลึงกันมาก (Homogeneous) คือมีค่าใกล้เคียงกัน ความแปรปรวนจะมีค่าต่ำ และในกรณีที่ Extreme จริง ๆ คือทุกหน่วยมีค่าเท่ากัน ความแปรปรวนจะมีค่าเท่ากับ 0 เช่นกลุ่มตัวอย่างที่ประกอบด้วยหน่วยตัวอย่างที่ให้ค่าสังเกตของลักษณะที่สนใจเป็น (2,2,2,2) เราจะพบว่า $s^2 = 0$ ทั้งนี้เพราะ $x_i = \bar{x}$ ในขณะเดียวกัน ถ้าทุกหน่วยมีค่าแตกต่างกันมาก (Heterogeneous)[2] ความแปรปรวนจะมีค่าสูง ดังนั้น S^2_{wsy} จะมีค่าสูงได้ก็ต่อเมื่อค่าของหน่วยตัวอย่างมีความ

---

[1] อย่าลืมว่าถ้าหากกระจายหน่วยตัวอย่างของทุก Possible Sample ออกแล้วนำมาจัดเรียงผลลัพธ์ก็คือกลุ่มประชากรนั่นเอง ทั้งนี้เพราะใน LSS นั้นหน่วยสำรวจหนึ่ง ๆ จะตกอยู่ในกลุ่มตัวอย่างได้เพียง 1 ชุด

[2] คำว่าต่างกันหรือคล้ายคลึงกันนั้นเราต้องพูดถึงภายใต้ข้อตกลงหรือกติการ่วมอย่างใดอย่างหนึ่ง ในที่นี้ใช้ $\bar{x}$ หมายความว่าถ้าทุกหน่วยตัวอย่างมีค่าใกล้เคียงกับ $\bar{x}$ มากเราถือว่าหน่วยตัวอย่างเหล่านี้มีค่าใกล้เคียงกันมาก ถ้าต่างจาก $\bar{x}$ มากถือว่าแตกต่างกันมาก ความเข้าใจประการนี้คือพื้นฐานในการพัฒนาสูตรความแปรปรวน s^2

แตกต่างกันมาก ซึ่งเหตุการณ์เช่นนี้จะเกิดขึ้นได้เฉพาะเมื่อกลุ่มประชากรไม่มีลักษณะของ Periodic Variation ซึ่งหน่วยสำรวจในตำแหน่งเดียวกันของทุกชั้นภูมิมีค่าเท่ากันหรือใกล้เคียงกันแต่ควรมีลักษณะของ Random Order คือหน่วยที่มีค่าเท่ากันหรือใกล้เคียงกันของต่างชั้นภูมิมีการกระจายตัวไปทั่วชั้นภูมิโดยสุ่มไม่เรียงอยู่ในตำแหน่งเดียวกัน หรือประชากรควรมีลักษณะเป็น Ordered Population คือค่าสังเกตของหน่วยสำรวจในกลุ่มประชากรได้รับการจัดเรียงลำดับ อาจเป็นในลักษณะของการเรียงจากมากไปหาน้อยหรือจากน้อยไปหามากก็ได้ เช่น จัดสร้างกรอบตัวอย่างของเกษตรกรโดยเรียงรายชื่อหัวหน้าครัวเรือนไปตามลำดับของขนาดพื้นที่ถือครอง เป็นต้น[1]

ต่อไปนี้จะขอแสดงให้เห็นลักษณะการจัดเรียงของหน่วยสำรวจในแต่ละชั้นภูมิในลักษณะของ Periodic Variation ทั้งนี้ใช้ $N = 15$ และ $k = 5$ ดังนี้

ใน Periodic Variation ค่าสังเกตของหน่วยสำรวจ[2] อาจมีค่าและจัดเรียงดังนี้

1,2,3,4,5 | 1,2,3,4,5 | 1,2,3,4,5

กรณีนี้เป็นกรณีที่ extreme ที่สุดของ Periodic Variation ขอให้สังเกตว่าค่าของหน่วยสำรวจ ณ. ตำแหน่งเดียวกันในทั้งสามชั้นภูมิมีค่าเท่ากัน กลุ่มตัวอย่างที่พึงเป็นไปได้ทั้งหมดจะปรากฏดังตาราง

---

[1] จะได้กล่าวถึงธรรมชาติของกลุ่มประชากรชนิดต่าง ๆ ในตอนต่อไป

[2] ในทางปฏิบัติเราไม่อาจทราบค่าของหน่วยสำรวจในกลุ่มประชากรทั้งกลุ่มใด แต่เราอาจอาศัยดัชนีบางอย่างจากปัจจัยแวดล้อมรวมตลอดถึงประสบการณ์และการเฝ้าติดตามสังเกตเป็นเครื่องบ่งชี้ แต่ต้องแน่ใจว่าดัชนีนั้นมิได้เกี่ยวข้องกับค่าของตัวแปรในรายการข้อมูล (Field) ที่มุ่งศึกษา

กลุ่มตัวอย่างที่	ค่าของตัวแปร	$\bar{x}$	$\sum_{j}^{3} (x_{ij}-\bar{x}_i)^2$
1	1,1,1	1	0
2	2,2,2	2	0
3	3,3,3	3	0
4	4,4,4	4	0
5	5,5,5	5	0

จะเห็นว่า $S^2_{wsy} = \frac{1}{k(n-1)} \sum_{i}^{5} \sum_{j}^{3} (x_{ij}-\bar{x}_i)^2 = 0$ ซึ่งมีค่าต่ำสุดอันจะส่งผลให้ $V(\bar{x}_{sy})$ มีค่าสูงสุด และถ้าค่าของหน่วยสำรวจในชั้นภูมิต่าง ๆ มีลักษณะคลาดเคลื่อนไปจาก Periodic Variation มากขึ้น จะมีผลให้ S^2_{wsy} มีค่าเพิ่มมากขึ้น และ $V(\bar{x}_{sy})$ จะค่อย ๆ มีค่าลดลง และเมื่อหน่วยสำรวจในชั้นภูมิมีลักษณะเข้าใกล้ Random Order หรือ Ordered Population- $V(\bar{x}_{sy})$ จะลดลงมาก จนในที่สุดเมื่อภายในชั้นภูมิมีลักษณะที่ Homogeneous มากที่สุดและต่างชั้นภูมิมีความแตกต่างกันมากที่สุด[1] $V(\bar{x}_{sy})$ จะมีค่าต่ำที่สุดเช่น

1,1,1,1,1 | 4,4,4,4,4 | 10,10,10,10,10

จะพบว่า S^2_{wsy} มีค่าสูงขึ้น ขอให้ดูตารางต่อไปนี้ (N = 15, k = 5)

---

[1] ความแตกต่างระหว่างชั้นภูมิ กติการ่วมที่ใช้วัดคือ $\bar{X} = \frac{1}{N}(x_1+x_2+\ldots+x_N)$ อนึ่งคำว่า "แตกต่างกันมากที่สุด" นั้น ถ้าดูด้วยตัวเลขของค่าสังเกตเรายังไม่อาจตัดสินใจได้ ควรดูด้วยดัชนีอื่น ที่นิยมใช้คือ Intraclass Correlation ρ ตัวอย่างข้างต้นแสดงให้เห็นถึงกรณีเมื่อเกิด Homogeneity ภายในชั้นภูมิและเกิด Heterogeneity ระหว่างชั้นภูมิ เครื่องมือวัดความแตกต่างระหว่างชั้นภูมิก็คือความแตกต่างระหว่าง $\bar{X}_h$ กับ $\bar{X}$ ในที่นี้ $\bar{X}_1 = 1, \bar{X}_2 = 4, \bar{X}_3 = 10, \bar{X} = 15$

ตัวอย่างที่	ค่าของตัวแปร	$\bar{x}$	$\sum_{j}^{3} (x_{ij}-\bar{x})^2$	s_i^2
1	1,4,10	5	42	21
2	1,4,10	5	42	21
3	1,4,10	5	42	21
4	1,4,10	5	42	21
5	1,4,10	5	42	21

จะเห็นว่า $S^2_{wsy} = \dfrac{1}{k(n-1)} \sum_{i}^{5} \sum_{j}^{3} (x_{ij}-\bar{x}_i)^2$

$$= \frac{1}{5(3-1)} (42+42+42+42+42) = 21$$

$$S^2 = \frac{1}{N-1} \sum_{i}^{5} \sum_{j}^{3} (x_{ij}-\bar{X})^2 = 15$$

$$V(\bar{x}_{sy}) = \frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsy}$$

$$= \frac{14}{15} \cdot 15 - \frac{5(3-1)}{15} \cdot 21 = 0$$

3. ถ้า $S^2_{wsy} > S^2$ แผนสำรวจแบบ Systematic จะมีความแม่นยำกว่าแผนสำรวจแบบ SRS

สำหรับความจริงข้อนี้สามารถแสดงให้เห็นได้ดังนี้

$V(\bar{x}_{sy}) < V(\bar{x}_{ran})$ **ก็ต่อเมื่อ** $\dfrac{N-1}{N} S^2 - \dfrac{k(n-1)}{N} S^2_{wsy} < \dfrac{N-n}{N} \dfrac{S^2}{n}$

หรือ $-\dfrac{k(n-1)}{N} S^2_{wsy} < \dfrac{N-n}{N} \dfrac{S^2}{n} - \dfrac{N-1}{N} S^2 = \dfrac{N-n-nN+n}{n} \dfrac{S^2}{N}$

หรือ $\dfrac{k(n-1)}{N} S^2_{wsy} > -\dfrac{N-nN}{n} \dfrac{S^2}{N}$

หรือ $k(n-1)S^2_{wsy} > -(k-N)S^2 = (k-nk)S^2 = k(n-1)S^2$

หรือ $S^2_{wsy} > S^2$

นั่นคือ ถ้า $S^2_{wsy} > S^2$ แล้ว $V(\bar{x}_{sy})$ จะมีค่าน้อยกว่า $V(\bar{x}_{ran})$ หรือแผนสำรวจแบบ Systematic จะให้ค่าประมาณที่มีความแม่นยำสูงกว่าค่าประมาณที่ได้จากการใช้แผนสำรวจแบบ SRS

จากข้อสังเกต 3 ประการข้างต้นชี้ให้เห็นลักษณะที่พึงยึดถือสำหรับการใช้แผนสำรวจแบบ Systematic ดังนี้

1. ควรจัดให้ภายในชั้นภูมิเดียวกันมีลักษณะเป็นเนื้อเดียวกันมากที่สุดและต่างชั้นภูมิกันมีลักษณะแตกต่างกันมากที่สุด[1] ทั้งนี้เพื่อให้ Sample Variance (s_i^2) มีค่าสูงสุด ซึ่งจะมีผลสะท้อนให้ S^2_{wsy} มีค่าสูงและ $V(\bar{x}_{sy})$ มีค่าต่ำ

2. ในทางปฏิบัติเราอาจจำเป็นต้องคล้อยตามธรรมชาติของกลุ่มประชากร ไม่อาจบังคับหรือนำมาจัดเรียงโดยนัยแห่งข้อ 1 ได้ เรื่องนี้จึงเป็นเรื่องของการศึกษาธรรมชาติของกลุ่มประชากรซึ่งจะได้กล่าวถึงต่อไป

3. ดัชนีหนึ่งที่ใช้วัดดูว่า $V(\bar{x}_{sy})$ จะมีค่ามากหรือน้อย หรือนัยหนึ่งหน่วยตัวอย่างจะมีลักษณะคล้ายคลึงกันหรือไม่เพียงใด ดัชนีดังกล่าวก็คือ ϱ เรียกว่า Intraclass Correlation Coefficient ระหว่างคู่ของหน่วยตัวอย่างในกลุ่มตัวอย่างเดียวกัน โดยที่

$$\varrho = \frac{E(x_{ij}-\bar{X})(x_{ij'}-\bar{X})}{E(x_{ij}-\bar{X})^2}$$

$$= \frac{2}{n-2} \sum_{i}^{k} \sum_{j<j'}^{n} (x_{ij}-\bar{X})(x_{ij'}-\bar{X}) \frac{1}{N-1} \frac{1}{S^2} \quad {}^{2}$$

---

[1] ทำนองเดียวกับการจัดชั้นภูมิในแผนสำรวจแบบแบ่งชั้นภูมิ

[2] เช่นในกลุ่มตัวอย่างที่ 2 ประกอบด้วย (x_{21},x_{22},x_{23}) คู่ของหน่วยตัวอย่างที่จะนำมาคำนวณหาสหสัมพันธ์ ϱ เฉพาะในกลุ่มตัวอย่างที่ 2 จึงมีได้ $\binom{n}{2} = \binom{3}{2} = 3$ คู่คือ (x_{21},x_{22}), (x_{21},x_{23}), (x_{22},x_{23})
ดังนั้น $\sum_{j<j'}^{3} (x_{2j}-\bar{X})(x_{2j'}-\bar{X})$ จึงมีค่าเท่ากับ $(x_{21}-\bar{X})(x_{22}-\bar{X})+(x_{21}-\bar{X})(x_{23}-\bar{X})+(x_{22}-\bar{X})(x_{23}-\bar{X})$

ด้วยเหตุนี้เราจึงสามารถเสนอสูตร $V(\bar{x}_{sy})$ ในลักษณะที่มี ϱ เข้าไปเกี่ยวข้องได้อีกรูปหนึ่งดังนี้

ทฤษฎี 4.4 ในการสำรวจโดยใช้แผนสำรวจแบบ Systematic ที่มี Sampling Interval ขนาด k [1] จากกลุ่มประชากรขนาด N = nk ความแปรปรวนของการประมาณค่าเฉลี่ยประชากร $\bar{x}$ คือ

$$V(\bar{x}_{sy}) = \frac{N-1}{N}\frac{S^2}{n}\{1+(n-1)\varrho\}$$

และ $V(\hat{T}_{sy}) = V(N\bar{x}_{sy}) = N^2V(\bar{x}_{sy}) = \frac{N(N-1)}{n}\frac{S^2}{n}\{1+(n-1)\varrho\}$

พิสูจน์

$$V(\bar{x}_{sy}) = E(\bar{x}_{sy}-\bar{X})^2 = E(\bar{x}_i-\bar{X})^2 = \frac{1}{k}\sum_{i}^{k}(\bar{x}_i-\bar{X})^2$$

$$= \frac{1}{k}\sum_{i}^{k}\left(\frac{1}{n}\sum_{j}^{n}x_{ij}-\bar{X}\right)^2$$

$$= \frac{1}{k}\sum_{i}^{k}\left\{\frac{1}{n}\left(\sum_{j}^{n}x_{ij}-n\bar{X}\right)\right\}^2$$

$$= \frac{1}{k}\frac{1}{n^2}\sum_{i}^{k}\left\{\sum_{j}^{n}(x_{ij}-\bar{X})\right\}^2$$

$$V(\bar{x}_{sy}) = \frac{1}{k}\frac{1}{n^2}\left\{\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{X})^2 + 2\sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\bar{X})(x_{ij'}-\bar{X})\right\} \qquad ...(1)$$

จากนิยาม $\varrho = \frac{E(x_{ij}-\bar{X})(x_{ij'}-\bar{X})}{E(x_{ij}-\bar{X})^2}$ ซึ่งแสดงสหสัมพันธ์ระหว่างหน่วยตัวอย่างคู่หนึ่ง ๆ ภายในกลุ่มตัวอย่างชุดเดียวกัน

จะพบว่า จากกลุ่มตัวอย่างขนาด n เราสามารถจัดหน่วยสำรวจออกได้เป็นคู่ ๆ ได้ทั้งสิ้น $\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$ คู่ และเนื่องจากมีกลุ่มตัวอย่างที่เป็นไปได้ทั้งสิ้นรวม k ชุด ดังนั้นเราจึงสามารถจัดหน่วยสำรวจออกเป็นคู่ ๆ ได้ทั้งสิ้น $k\binom{n}{2} = \frac{kn(n-1)}{2}$ คู่ ซึ่ง $1/\{\frac{kn(n-1)}{2}\} = \frac{2}{kn(n-1)}$ ก็คือค่าความน่าจะเป็นที่หน่วยสำรวจคู่ใด ๆ จะมีโอกาสได้รับเลือก

[1] เลือกมาหนึ่งหน่วยจากในทุก ๆ k หน่วย

ดังนั้น $E(x_{ij}-\overline{X})(x_{ij'}-\overline{X}) = \sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\overline{X})(x_{ij'}-\overline{X})\cdot\frac{2}{kn(n-1)}$

ขณะเดียวกันเราจะพบว่า

$$E(x_{ij}-\overline{X})^2 = \sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\overline{X})^2\cdot\frac{1}{N}$$

$$= \frac{N-1}{N-1}\cdot\frac{1}{N}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\overline{X})^2$$

$$= \frac{N-1}{N}S^2$$

ดังนั้น $\varrho = \frac{E(x_{ij}-\overline{X})(x_{ij'}-\overline{X})}{E(x_{ij}-\overline{X})^2} = \frac{2}{kn(n-1)}\sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\overline{X})(x_{ij'}-\overline{X})\cdot\frac{N}{N-1}\;\frac{1}{S^2}$

$$= \frac{2}{n-1}\sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\overline{X})(x_{ij'}-\overline{X})\frac{1}{N-1}\;\frac{1}{S^2} \quad \text{[1]} \qquad \ldots\ldots\ (2)$$

ดังนั้น

$$2\sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\overline{X})(x_{ij'}-\overline{X}) = (n-1)(N-1)S^2\varrho \qquad \ldots\ldots\ (3)$$

เมื่อแทนค่าสมการที่ (3) ลงในสมการที่ (1) จะพบว่า

$$V(\bar{x}_{sy}) = \frac{1}{k}\frac{1}{n^2}\left\{\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\overline{X})+(n-1)(N-1)S^2\varrho\right\}$$

$$= \frac{1}{nN}\left\{(N-1)S^2+(n-1)(N-1)S^2\varrho\right\} \quad \text{[2]}$$

$$= \frac{(N-1)S^2}{nN}\left\{1+(n-1)\varrho\right\}$$

$$V(\bar{x}_{sy}) = \frac{N-1}{N}\;\frac{S^2}{n}\left\{1+(n-1)\varrho\right\}$$

---

[1] $N = kn$

[2] $kn^2 = kn\cdot n = nN$

จากสูตร $V(\bar{x}_{sy}) = \frac{N-1}{N}\ \frac{S^2}{n}\ \{1+(n-1)\varrho\}$ เราจะพบว่า

1) ถ้า ϱ มีค่าสูงและเป็นปริมาณบวกหรือนัยหนึ่ง เมื่อ x_{ij} และ $x_{ij'}$; $i = 1,2,\ldots,k;$ $j<j'$; $j = 1,2,\ldots,n$ มีสหสัมพันธ์ในทางบวกสูงมากหรือมีความผันแปรไปในทางเดียวกันและมีค่าสูงแล้ว $V(\bar{x}_{sy})$ จะมีค่าสูง ถ้า ϱ มีค่าต่ำลง $V(\bar{x}_{sy})$ ก็จะมีค่าต่ำลง

2) ถ้า ϱ มีค่าเป็นปริมาณลบ (ติดลบ) $V(\bar{x}_{sy})$ จะมีค่าต่ำ และจะมีค่าต่ำที่สุด เมื่อ $\varrho = \frac{-1}{n-1}$ ซึ่งความจริงข้อนี้สามารถพิสูจน์ให้เห็นได้ดังนี้

$$\text{จาก } V(\bar{x}_{sy}) = \frac{N-1}{N}\ \frac{S^2}{n}\ \{1+(n-1)\varrho\}$$

ถ้า $\varrho = \frac{-1}{n-1}$ จะทำให้ $1+(n-1)\varrho = 1 - \frac{n-1}{n-1} = 0$ ซึ่งจะมีผลให้ $V(\bar{x}_{sy}) = 0$ ซึ่งเป็นค่าที่ต่ำที่สุด

แสดงว่าขีดจำกัดล่างของ ϱ ก็คือ $\frac{-1}{n-1}$

3) ถ้า $\varrho = 0$ จะมีผลให้ $V(\bar{x}_{sy}) = \frac{N-1}{N}\frac{S^2}{n}$ หรือนัยหนึ่ง เมื่อ $\varrho = 0$ แผนสำรวจแบบ Systematic จะให้ผลลัพธ์เช่นเดียวกันกับแผนสำรวจแบบ SRS นักศึกษาอาจสงสัยว่า $V(\bar{x}_{ran}) = \frac{N-n}{N}\ \frac{S^2}{n}$ มิใช่ $\frac{N-1}{N}\ \frac{S^2}{n}$ ความจริงแล้ว $V(\bar{x}_{ran})$[1] สามารถเสนอได้หลายรูปแต่รูปที่นิยมใช้ทางปฏิบัติมีเพียง 2 รูปคือ

ก. $V(\bar{x}_{ran}) = \frac{N-1}{N}\frac{S^2}{n}$ ซึ่งเป็นสูตรสำหรับกรณีของ Sampling Without Replacement

และ

ข. $V(\bar{x}_{ran}) = \frac{N-1}{N}\ \frac{S^2}{n}$ ซึ่งเป็นสูตรสำหรับกรณีของ Sampling With Replacement

การพิสูจน์สำหรับข้อ ข. นี้จะขอเว้นไว้เป็นแบบฝึกหัด ส่วนข้อ ก. ได้พิสูจน์มาแล้วในบทที่ 2

---

[1] หรือจะใช้สัญลักษณ์เป็น $V(\bar{x}_{SRS})$ ก็ได้

อย่างไรก็ตาม เมื่อพิจารณาเฉพาะในกรณีของ Sampling Without Replacement ซึ่งเป็นกรณีที่นิยมใช้กันโดยทั่วไปในทางปฏิบัติจะพบว่า แม้ว่า $\varrho = 0$ ก็มิได้ยืนยันว่าแผนสำรวจแบบ Systematic และ SRS จะให้ความแม่นยำเท่ากัน

เรื่องราวเหล่านี้เป็นเรื่องของการเปรียบเทียบประสิทธิภาพระหว่างแผนสำรวจแบบ Systematic กับแผนสำรวจแบบ SRS ซึ่งจะกล่าวถึงต่อไปในตอนต่อไป แต่ก่อนที่จะผ่านตอนนี้ไปขอยกตัวอย่างประกอบการศึกษาในส่วนของการคำนวณหา $V(\bar{x}_{sy})$ เสียก่อน

ตัวอย่าง 4.1 สมมุติว่ากลุ่มประชากรหนึ่งประกอบไปด้วยหน่วยสำรวจซึ่งมีค่าดังต่อไปนี้คือ

1,4,7,10,2,5,8,11,3,6,9,12

ก. จงเลือกตัวอย่างมา 1 หน่วยจากทุก ๆ 4 หน่วย แล้วคำนวณหาค่า ϱ และใช้ค่า ϱ คำนวณหา $V(\bar{x}_{sy})$

ข. จงคำนวณหา $V(\bar{x}_{sy})$ โดยอาศัยทฤษฎี 4.3 และนิยาม $V(\bar{x}_{sy}) = \frac{1}{k}\sum_{i}^{k}(\bar{x} - \bar{\mathbf{X}})^2$

ค. จงวาดไดอะแกรมแสดงความหมายของ ϱ

ง. จงเปรียบเทียบ $V(\bar{x}_{sy})$ กับ $V(\bar{x}_{ran})$

จ. ถ้าจัดเรียงค่าของหน่วยสำรวจเสียใหม่เป็น

1,2,3,4,5,6,7,8,9,10,11,12

จงเปรียบเทียบ $V(\bar{x}_{sy})$ กับ $V(\bar{x}_{ran})$

ฉ. ถ้าสุ่มตัวอย่างมาเพียงชุดเดียว (ใช้ข้อมูลจากข้อ จ.) จงหาค่าประมาณของ $\bar{X}$

วิธีทำ

ก. จากกลุ่มประชากรขนาด N = 12 ข้างต้นและ k = 4 จะพบว่า n = 3 เราสามารถจำแนกประชากรได้เป็น 3 ชั้นภูมิดังนี้

1,4,7,10 | 2,5,8,11 | 3,6,9,12

และกลุ่มตัวอย่างที่พึงเป็นไปได้ทั้งหมดปรากฏดังตาราง

ตัวอย่างที่	ค่าของตัวแปร	$\bar{x}_i$	$\sum_{j}^{3} (x_{ij}-\bar{x}_i)$	s^2
1	1,2,3	2	2	1
2	4,5,6	5	2	1
3	7,8,9	8	2	1
4	10,11,12	11	2	1

$$\overline{X} = \frac{1}{N}\sum_{i}^{k}\sum_{j}^{n} x_{ij} = \frac{1}{12}\{(1+2+3)+(4+5+6)+(7+8+9)+(10+11+12)\}$$

$$= 6.5$$

$$S^2 = \frac{1}{N-1}\sum_{i}^{n}\sum_{j}^{k}(x_{ij}-\overline{X})^2 = \frac{1}{11}(650 - \frac{78}{12})^2 = 13$$

$$S^2_{wsy} = \frac{1}{k(n-1)}\sum_{i}^{k}\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2 = \frac{1}{4(3-1)}(2+2+2+2) = 1$$

$$\varrho = \frac{2}{n-1}\sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\overline{X})(x_{ij'}-\overline{X})\cdot\frac{1}{N-1}\cdot\frac{1}{S^2}$$ [1]

ให้ $C_i = \sum_{j<j}^{n}(x_{ij}-\overline{X})(x_{ij}-\overline{X})$; $i=1,2,\ldots,k$

ดังนั้น

ในตัวอย่างที่ 1 จะพบว่า $C_1 = (1-6.5)(2-6.5)+(1-6.5)(3-6.5)+(2-6.5)(3-6.5) = 59.75$

ในตัวอย่างที่ 2 จะพบว่า $C_2 = (4-6.5)(5-6.5)+(4-6.5)(6-6.5)+(5-6.5)(6-6.5) = 5.75$

ในตัวอย่างที่ 3 จะพบว่า $C_3 = (7-6.5)(8-6.5)+(7-6.5)(9-6.5)+(8-6.5)(9-6.5) = 5.75$

---

[1] ค่าต่ำสุดของ ϱ คือ $\varrho = -\frac{1}{n-1} = -\frac{1}{3-1} = -0.5$

ในตัวอย่างที่ 4 จะพบว่า $C_4 = (10-6.5)(11-6.5)+(10-6.5)(12-6.5)+(11-6.5)(12-6.5)$

$$= 59.75$$

ดังนั้น $$\varrho = \frac{2}{n-1} \sum_{i}^{k} \sum_{j<j'}^{n} (x_{ij}-\overline{X})(x_{ij'}-\overline{X}) \frac{1}{N-1} \cdot \frac{1}{S^2}$$

$$= \frac{2}{3-1} (59.75+5.75+5.75+59.75) \cdot \frac{1}{12-1} \cdot \frac{1}{13}$$

$$= 0.9160838$$

จาก $$V(\bar{x}_{sy}) = \frac{N-1}{N} \frac{S^2}{n} \{1+(n-1)\varrho\}$$

$$=> V(\bar{x}_{sy}) = \frac{11}{12} \cdot \frac{13}{3} \{1+2(.9160838)\}$$

$$= 11.249998$$

ข. 1 $$V(\bar{x}_{sy}) = \frac{N-1}{N} \cdot \frac{k(n-1)}{N} \cdot S^2_{wsy}$$

$$S^2 = 13, S^2_{wsy} = \frac{1}{k(n-1)} \sum_{i}^{k} \sum_{j}^{k} (x_{ij}-\overline{X})^2 = 1$$

ดังนั้น $$V(\bar{x}_{sy}) = \frac{11}{12} \cdot 13 - \frac{4(3-1)}{12} \cdot 1 = 11.916666 - 0.66666666$$

$$= 11.25$$

ข. 2 $$V(\bar{x}_{sy}) = \frac{1}{k} \sum_{i}^{k} (x_i-\overline{X})^2$$

$$= \frac{1}{4} \{(2-6.5)^2+(5-6.5)^2+(8-6.5)^2+(11-6.5)^2\}$$

$$= 45/4 = 11.25$$

จะเห็นว่าเราสามารถใช้สูตรใดคำนวณหา $V(\bar{x}_{sy})$ ก็ได้เพราะให้ผลลัพธ์ตรงกัน แต่ปัญหาที่สามารถมองเห็นได้ก็คือสูตร $V(\bar{x}_{sy})$ เหล่านี้ล้วนอาศัยข้อมูลจากกลุ่มตัวอย่าง

ทุกกลุ่ม (All Possible Sample) ซึ่งเป็นสิ่งที่ไม่อาจกระทำได้ในทางปฏิบัติ ทั้งนี้เพราะในทางปฏิบัตินั้นเราจะสุ่มตัวอย่างมาใช้เพียง 1 ชุด (Single Sample of Size n) เท่านั้นและเรายังไม่มีตัวประมาณค่าที่ปราศจากอคติของ $V(\bar{x}_{sy})$ เท่าที่มีใช้อยู่มักจะเป็นตัวประมาณค่าที่มีอคติหรือพอจะใช้ได้เฉพาะในบางสถานการณ์เท่านั้น

ค. ดังที่กล่าวแล้วว่าค่าของ ϱ คือดัชนีที่ใช้วัดสหสัมพันธ์ภายในกลุ่มตัวอย่างโดยวิเคราะห์ข้อมูลจากกลุ่มตัวอย่างหนึ่ง ๆ เป็นคู่ ๆ จนครบทุก Possible Sample ดังนั้น ถ้า ϱ เป็นปริมาณบวกและมีค่าสูงก็ย่อมแสดงว่าภายในกลุ่มตัวอย่างมีความเป็นเนื้อเดียวกันมาก ซึ่งจะส่งผลสะท้อนให้ $V(\bar{x}_{sy})$ มีค่าสูง และชี้ให้เห็นว่าชั้นภูมิมีความโน้มเอียงเข้าสู่ลักษณะของ Periodic Variation ขณะเดียวกันถ้า ϱ มีค่าใกล้ 0 ก็จะชี้ให้เห็นว่ากลุ่มตัวอย่างมีความโน้มเอียงที่จะเป็น Heterogeneous จนเมื่อ ϱ มีค่าติดลบลักษณะของ Heterogeneity จะสูงมากขึ้น ซึ่งจะส่งผลสะท้อนให้ $V(\bar{x}_{sy})$ มีค่าน้อยลงตามลำดับ

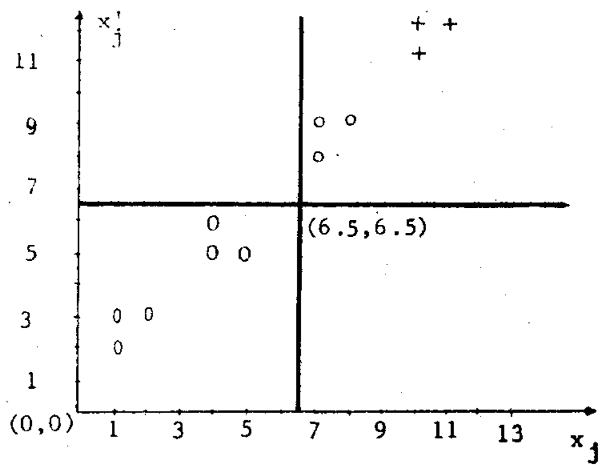
$$\text{เนื่องจาก } \varrho = \frac{2}{n-1}\sum_{i}^{k}\sum_{j<j'}^{n}(x_{ij}-\bar{X})(x_{ij'}-\bar{x})\cdot\frac{1}{N-1}\Big|\frac{1}{S^2}$$

แสดงว่าเกณฑ์ร่วมที่จะใช้เป็นหลักวัดว่าข้อมูลมีความเป็นเนื้อเดียวกันหรือไม่ก็คือ $\bar{X}$ ดังนั้นไดอะแกรมที่แสดงลักษณะการกระจายของข้อมูลคือแกน x_j และแกน $x_{j'}$ จุดกำเนิดคือ $(\bar{X},\bar{X})$ สเกลบนแกนจะเริ่มจากค่าต่ำสุดไปถึงค่าสูงสุดของ X การตีความหมายจากไดอะแกรมยึดถือหลักเกณฑ์เดียวกับ Scatterdiagram ในการวิเคราะห์สหสัมพันธ์และความถดถอย กล่าวคือถ้าข้อมูลส่วนใหญ่กระจายอยู่ในควอทแดรนท์ที่ 1 และ 3 แสดงว่า ϱ มีค่าเป็นบวก ถ้ากระจายหนาแน่นในควอทแดรนท์ที่ 2 และ 4 แสดงว่า ϱ มีค่าเป็นลบ ถ้ากระจายรอบ ๆ จุดกำเนิด แสดงว่า ϱ มีค่าใกล้ 0 ยิ่งกระจายห่างจากจุดกำเนิดมากเพียงใด ϱ จะมีค่าเพิ่มมากขึ้นเพียงนั้น

จากกลุ่มตัวอย่างในข้อ ก. เราสามารถเตรียมตารางวิเคราะห์ได้ดังนี้

ตัวอย่างที่	ค่าของตัวแปร	คู่ลำดับ $(x_{ij},x_{ij'})$	คู่ลำดับ $(x_{ij}-\bar{X}),(x_{ij'}-\bar{X})$ เมื่อ $\bar{X} = 6.5$
1	1,2,3	(1,2),(1,3),(2,3)	(−5.5, −4.5),(−5.5, −3.5),(−4.5, −3.5)
2	4,5,6	(4,5),(4,6),(5,6)	(−2.5, −1.5),(−2.5, −0.5),(−1.5, −0.5)
3	7,8,9	(7,8),(7,9),(8,9)	(0.5,1.5),(0.5,2.5),(1.5,2.5)
4	10,11,12	(10,11),(10,12),(11,12)	(3.5,4.5),(3.5,5.5),(4.5,5.5)

นำคู่ลำดับมาจัดลงไดอะแกรมได้ดังภาพ เส้นจางคือแกนเดิมที่มีจุดกำเนิด (0,0) สำหรับคู่ลำดับ $(x_{ij},x_{ij'})$ ส่วนแกนเส้นหนาที่เทียบต่อแกนรวม $\bar{X}$ มีจุดกำเนิด $(\bar{X},\bar{X})$ = 6.5,6.5)



จะเห็นว่าคู่ลำดับจะตกอยู่ในควอทแดรนท์ 1 และ 3 สำหรับแกนที่ใช้กติการ่วม $\bar{X}$ เป็นตัวเปรียบเทียบความแตกต่าง และตกอยู่ในควอทแดรนท์ 1 ทั้งหมดสำหรับแกนเดิม ซึ่งแสดงว่าค่าสหสัมพันธ์ ϱ จะมีค่าบวกและมีค่าสูง เหตุที่มั่นใจว่ามีค่าบวกเพราะคู่ลำดับตกอยู่ในควอทแดรนท์ 1 และ 3 และด้วยเหตุที่คู่ลำดับตกอยู่เฉพาะในควอทแดรนท์ที่ 1 และ 3 และกระจายห่างไปจากจุดกำเนิด $(\bar{X},\bar{X})$ ค่า ϱ จึงน่าจะมีค่าสูง ทั้งนี้เพราะในควอทแดรนท์ที่ 1 และ 3 C จะมีค่าเป็นค่าบวกทั้งหมด และขอให้สังเกตว่าคู่ลำดับจะวัดได้เป็นกลุ่ม ๆ

ตามกลุ่มตัวอย่าง กลุ่มที่ห่างไกลจุดกำเนิด $(\bar{X},\bar{X})$ มากจะให้ค่า C สูง ด้วยเหตุนี้จึงทำให้เชื่อได้ว่าถ้าคู่ลำดับกระจายห่างจากจุดกำเนิดออกไปมากจะมีผลให้ ϱ มีค่าสูงและจะลดลงถ้ากระจายตัวใกล้เคียงจุดกำเนิด ϱ จะมีค่าเท่ากับ 0 ถ้าคู่ลำดับกระจายตัวรอบ ๆ ค่าเฉลี่ยอยู่ในควอทแดรนท์ 1 หรือ 3 บ้างและในควอทแดรนท์ 2 หรือ 4 บ้าง และมีผลให้ค่า C หักลบกันหมด ขณะเดียวกันถ้าคู่ลำดับกระจายตัวอยู่ในควอทแดรนท์ 2 และ 4 ค่า C จะติดลบและมีผลให้ ϱ ติดลบ

ง.
$$V(\bar{x}_{ran}) = \frac{N-n}{N}\cdot\frac{S^2}{n}$$
$$= \frac{12-3}{12}\cdot\frac{13}{3}$$
$$= 3.25$$

จะเห็นว่า $V(\bar{x}_{sy}) = 11.25$ ขณะที่ $V(\bar{x}_{ran}) = 3.25$

แสดงว่าเมื่อ ϱ มีค่าสูงแผนสำรวจแบบ Systematic จะให้ค่าประมาณที่มีความแม่นยำน้อยกว่าแผนสำรวจแบบ SRS

จ. เมื่อจัดเรียงหน่วยสำรวจเสียใหม่ดังนี้คือ

1,2,3,4 | 5,6,7,8 | 9,10,11,12

จะพบกลุ่มตัวอย่างดังตารางต่อไปนี้ อนึ่งขอให้สังเกตว่าเมื่อจัดเรียงหน่วยสำรวจเสียใหม่จะพบว่าภายในชั้นภูมิต่าง ๆ มีลักษณะเป็นเนื้อเดียวกันมากกว่าในข้อ ก. ซึ่งการกระทำดังนี้จะมีผลให้ภายในกลุ่มตัวอย่างมีลักษณะ Heterogeneity สูงอันจะมีผลให้ S^2_{wsy} มีค่าสูงเชื่อว่า $V(\bar{x}_{sy})$ จะมีค่าลดลง

ตัวอย่างที่	ค่าของตัวแปร	$\bar{x}_i$	$\sum_{j}^{n}(x_{ij}-\bar{x}_i)^2$	s^2
1	1,5,9	5	32	16
2	2,6,10	6	32	16
3	3,7,11	7	32	16
4	4,8,12	8	32	16

$S^2 = 13$, $\bar{X} = 6.5$

$$S^2_{wsy} = \frac{1}{k(n-1)} \sum_i^k \sum_j^n (x_{ij} - \bar{x}_i)^2 = \frac{1}{4(3-1)}(32+32+32+32) = 16$$

ดังนั้น
$$V(\bar{x}_{sy}) = \frac{N-1}{N}S^2 - \frac{k(n-1)}{N}S^2_{wsy}$$
$$= \frac{11}{12}.\ 13 - \frac{4(3-1)}{12}.\ 16$$
$$= 11.916666 - 10.666666$$
$$= 1.249994$$
$$= 1.25$$

จะเห็นว่า $V(\bar{x}_{sy}) < V(\bar{x}_{ran})$ หรือเมื่อ $S^2_{wsy} > S^2$ แล้ว $V(\bar{x}_{sy}) < V(\bar{x}_{ran})$

ฉ. สมมุติเลือกได้ตัวอย่างชุดที่ 3 คือ (3,7,11) จะพบว่า

$$\bar{x}_{sy} = \frac{k}{N}\sum_j^n x_{ij} = \frac{4}{12}(3+7+11) = 7$$

นั่นคือ ค่าประมาณที่ปราศจากอคติของ $\bar{X}$ คือ $\bar{x}_{sy} = 7$

ขอให้สังเกตว่าเมื่อ $nk = N$ เราสามารถใช้สูตร $\bar{x}_{sy} = \frac{k}{N}\sum_j^n x_{ij}$ หรือ $\bar{x}_{sy} = \sum_j^n \frac{x_{ij}}{n}$ ก็ได้เพราะให้ผลตรงกัน แต่ถ้าใช้ CSS สูตรการประมาณค่าของ $\bar{X}$ คือ $\bar{x}_{sy} = \frac{1}{n}\sum_j^n x_{ij}$ ซึ่งก็คือสูตรเดียวกันกับกรณีทั่วไปเมื่อ $nk = N$

4.6 การเปรียบเทียบประสิทธิภาพระหว่างแผนสำรวจ Systematic กับ SRS

การเปรียบเทียบประสิทธิภาพในที่นี้มุ่งเปรียบเทียบระหว่างแผนสำรวจแบบ Systematic กับแผนสำรวจแบบ SRS เท่านั้น อย่างไรก็ตามอาจกล่าวถึงการเปรียบเทียบแผนสำรวจนี้กับแผนสำรวจแบบแบ่งชั้นภูมิในภายหลังเมื่อได้ศึกษาเรื่องธรรมชาติของกลุ่มประชากรแล้ว เหตุที่จำเป็นต้องเปรียบเทียบประสิทธิภาพ (Relative Efficiency) ก็เพราะเราต้องทราบว่าแผนสำรวจแบบ Systematic มีคุณภาพดีเพียงใด และการจะกล่าวถึงสิ่งหนึ่งสิ่งใดว่าดีเพียงใดได้นั้น จำเป็นต้องมีกติกาหรือเกณฑ์หนึ่งใดเตรียมไว้เปรียบเทียบ ในที่นี้เกณฑ์ที่จะใช้คือแผนสำรวจแบบ SRS ผลการเปรียบเทียบจะช่วยชี้ให้เห็นว่าแผนสำรวจแบบ Systematic มีประสิทธิภาพสูงต่ำกว่าแผนสำรวจแบบ SRS หรือไม่ อย่างไรและเมื่อใด

เรานิยามรูปทั่วไปของ Relative Efficiency ไว้ดังนี้

$$\text{Relative Efficiency ของ } \hat{\Theta}_1 : \hat{\Theta}_2 = \frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)}$$

และเมื่อ $\hat{\Theta}_1$ และ $\hat{\Theta}_2$ เป็นตัวประมาณค่าที่ปราศจากอคติของ Θ ตามลำดับแล้ว

$$R.E. = \frac{V(\hat{\Theta}_1)}{V(\hat{\Theta}_2)}$$

ดังนั้น Relative efficiency ของแผนสำรวจแบบ Systematic และแบบ SRS จึงปรากฏดังนี้

$$R.E. = \frac{V(\bar{x}_{sy})}{V(\bar{x}_{ran})}$$

$$= \frac{N-1}{N} S^2\{1+(n-1)\varrho\} \;/\; (\frac{N-n}{N} \; \frac{S^2}{n})$$

$$=> \quad \frac{V(\bar{x}_{sy})}{V(\bar{x}_{ran})} = \frac{N-1}{N-n}\{1+(n-1)\varrho\}$$

ผลสรุปจะแปรเปลี่ยนไปตามค่าของ ϱ ดังนี้

ค่าของ ϱ	R.E.	ความหมาย
$-\frac{1}{n-1}$	0	แผนสำรวจแบบ Systematic มีประสิทธิภาพสูงกว่าแบบ SRS
$-\frac{1}{N-1}$	1	แผนสำรวจแบบ Systematic มีประสิทธิภาพเท่าเทียมกับ SRS หรือใช้ทดแทนกันได้[1]
0	$\frac{N-1}{N-n}$	แผนสำรวจแบบ SRS มีประสิทธิภาพสูงกว่าแบบ Systematic เล็กน้อย แต่จะมี ประสิทธิภาพเท่ากันถ้าใน SRS นั้นเราสุ่มตัวอย่างมาโดยวิธีสุ่มแล้วใส่คืน[2] (with replacement sampling)
1	$\frac{nN-n}{N-n}$	แผนสำรวจแบบ SRS มีประสิทธิภาพสูงกว่า

ข้อสังเกต จากตารางจะเห็นได้ว่า $-\frac{1}{n-1} \leq \varrho \leq 1$

---

[1] ถ้า $\rho = -\frac{1}{N-1}$ แล้วจะพบว่า R.E. $= \frac{N-1}{N-n}(1-\frac{n-1}{N-1}) = \frac{N-1}{N-n} - \frac{n-1}{N-n} = 1$ แสดงว่าถ้าเราคำนวณหาค่า ϱ ได้เท่ากับ $-\frac{1}{N-1}$ เราจะใช้สูตรการประมาณค่าต่าง ๆ ในแผนสำรวจ SRS ไปประมาณค่าพารามิเตอร์ได้ แม้ว่าจะสุ่มตัวอย่าง (เพียง 1 ชุด) มาโดยวิธี Systematic แต่อย่างไรก็ตามในทางปฏิบัติเราไม่อาจคำนวณหาค่า ρ ได้แต่จำเป็นต้องคาดหมายค่า ρ เอาจากธรรมชาติของกลุ่มประชากร

[2] Des Raj หน้า 38 (ทฤษฎี 3.3) และ Taro Yamane หน้า 70 กล่าวว่าในการสุ่มตัวอย่างโดยวิธีเลือกแล้วใส่คืนจะพบว่า

$$E(\bar{x}) = \bar{X}, \; V(\bar{x}) = \sigma^2/n = \frac{N-1}{N}\frac{S^2}{n} \quad \text{และ} \quad \hat{V}(\bar{x}) = s^2/n$$

การพิสูจน์ให้ยึดถือผลและวิธีปฏิบัติและวิธีพิสูจน์ในทฤษฎี 4.2

4.7 ธรรมชาติของกลุ่มประชากร

ในการใช้แผนสำรวจแบบ Systematic นั้นปัจจัยที่มีอิทธิพลต่อคุณภาพของตัวประมาณค่าอย่างยิ่งก็คือลักษณะการเรียงค่าของหน่วยสำรวจในกลุ่มประชากรซึ่งนับเป็นข้อจำกัดที่เพิ่มภาระให้แก่นักวิจัยมากขึ้นกว่าการใช้แผนสำรวจแบบอื่น ทั้งนี้เพราะนอกจากจะต้องมีความรู้ความเข้าใจในหลักและวิธีการสำรวจแล้วนักวิจัยจำเป็นต้องมีความมั่นใจว่ากลุ่มประชากรมีธรรมชาติแบบใดอีกด้วยและโดยปกติตัวประมาณค่าของ $V(\bar{x}_{sy})$ ที่พอจะใช้ได้นั้นขึ้นอยู่กับธรรมชาติของกลุ่มประชากรแต่ละแบบ

ธรรมชาติของกลุ่มประชากรมีอยู่หลายแบบด้วยกันดังนี้

1. Periodic Variation หรือ Periodicity

กลุ่มประชากรที่เป็น Periodic Variation คือกลุ่มประชากรที่ค่าของตัวแปรสุ่มที่สนใจ ณ ตำแหน่งเดียวกันหรือ period เดียวกันในชั้นภูมิต่างกันหรือในระยะเวลา, ระยะทาง, พื้นที่ ฯลฯ ต่างกัน มีค่าใกล้เคียงกัน และกรณีที่ extreme ที่สุดก็คือมีค่าเท่ากัน ดังตัวอย่าง 1,2,3,4,5 | 1,2,3,4,5 | 1,2,3,4,5

ถ้าประชากรมีลักษณะนี้เราจะพบว่า ϱ มีค่าสูงมากเช่นในกรณีของตัวอย่างจะพบว่า $\varrho = 1$ ซึ่งสูงที่สุด สิ่งที่น่าสังเกตก็คือ ถ้าจะใช้ตัวอย่างเพียงชุดเดียวและบังเอิญเลือกตัวอย่างเริ่มต้นเป็น 1 หน่วยต่อ ๆ ไปจะเป็น 1 ผลก็คือค่าประมาณเช่น $\bar{x}_{sy}$, $\hat{T}_{sy}$, $\hat{P}_{sy}$ จะมีค่าต่ำกว่าความเป็นจริงและมีผลเสมือนเราสุ่มตัวอย่างมาใช้เพียง 1 หน่วย $(n=1)$ โดยนัยตรงกันข้าม ถ้าเลือก หน่วยเริ่มต้นเป็น 5 หน่วยต่าง ๆ ต่อไปจะเป็น 5 ผลลัพธ์ของการประมาณค่าจะสูงกว่าความเป็นจริง ทางออกอีกทางหนึ่งสำหรับกรณีเช่นนี้ก็คือให้ใช้ Centered Systematic Sampling คือแทนที่จะเลือกหน่วยเริ่มต้นด้วยเลขสุ่มระหว่าง $1-k$ ให้ใช้หน่วยที่อยู่กลางชั้นภูมิเป็นหน่วยเริ่มต้น คือใช้หน่วยที่ $\frac{k+1}{2}$ เป็นหน่วยเริ่มต้นถ้า k เป็นเลขคี่ และใช้หน่วยที่ $\frac{k+2}{2}$ หรือ $\frac{k}{2}$ เป็นหน่วยเริ่มต้น ถ้า k เป็นเลขคู่ วิธีนี้จะให้ค่าประมาณที่ใกล้เคียงกับค่าจริง ตามตัวอย่างข้างต้น $k=5$ ดังนั้นหน่วยเริ่มต้นคือหน่วยที่ $\frac{k+1}{2} = \frac{5+1}{2} = 3$ ดังนั้นกลุ่มตัวอย่างคือ $(x_3,x_8,x_{13}) = (3,3,3)$ จะเห็นว่า $\bar{x}_{sy} = 3 = \bar{X}$ ทางออกที่ดีที่สุดก็คือถ้ามีสถานการณ์ที่ชี้ชัดว่ากลุ่มประชากรมีลักษณะของ Periodic Variation ก็คือเปลี่ยนแผนสำรวจไปใช้แผนอื่น

ตัวอย่างของสถานการณ์ที่เป็น Periodic Variation ปรากฏดังนี้คือ

(1) ถ้าเราต้องการประมาณจำนวนรถยนต์ที่วิ่งผ่านด่านตรวจแห่งหนึ่งโดยจะเก็บข้อมูลติดต่อกันทุก ๆ ระยะเวลา 24 ชั่วโมง (k=24) เป็นเวลา 1 เดือน (N=720 ชั่วโมง) ในกรณีนี้เราเชื่อได้ว่าการจราจรมีลักษณะของ Periodic Variation ในช่วง 24 ชั่วโมง (ชั้นภูมิหนึ่ง ๆ มีขนาดเท่ากับ 24) เพราะใน 24 ชั่วโมงนั้นย่อมมีแบบแผน (Pattern) ของจังหวะเวลาที่การจราจรหนาแน่นและเบาบางเหมือนกันทุกวัน (24 ชั่วโมง) เช่น หนาแน่นมาก (Peak Hour) ในช่วง 7.00 น., 8.00 น., 9.00 น., 16.00 น., 17.00 น. และ 18.00 น. เวลาอื่นจะเบาบางและลดหลั่นกันลงไป ด้วยเหตุที่ ถ้าเราเลือกหน่วยเริ่มต้นได้เวลาที่การจราจรหนาแน่น เช่น หน่วยเริ่มต้นเป็น 7.00-8.00 น. เราก็จะต้องสำรวจหรือจดบันทึกจำนวนยวดยานในเวลา 7.00-8.00 น. ทุกวันจนครบ 30 วัน (n=30) ผลลัพธ์ก็คือค่าประมาณของจำนวนยวดยานที่ผ่านด่านมีค่าสูงกว่าความเป็นจริง (over estimate) โดยนัยกลับกันถ้าหน่วยเริ่มต้นคือหน่วยเวลาที่การจราจรเบาบางค่าประมาณจะต่ำกว่าความเป็นจริง

(2) การประมาณอุณหภูมิเฉลี่ยในเดือนเมษายน ถ้าเราจำเป็นต้องทำการจดบันทึกข้อมูลทุก ๆ 24 ชั่วโมง จะพบว่าอุณหภูมิของอากาศมีลักษณะเป็น Periodic Variation เพราะอุณหภูมิในเดือนหนึ่ง ๆ นั้นแต่ละชั่วโมงในแต่ละวันมักมีค่าใกล้เคียงกัน เช่นมีอุณหภูมิสูงสุดในช่วงเวลา 12.00-13.00 น. อุณหภูมิต่ำสุดในช่วงเวลา 05.00-06.00 น. เป็นต้น

(3) การบันทึกยอดขายรายสัปดาห์ของห้างสรรพสินค้าโดยจะทำการสำรวจในวันใดวันหนึ่งของสัปดาห์ (k=7) ตลอดฤดูหนาว (N=120 วัน) จะพบว่ายอดขายสินค้าของห้างสรรพสินค้ามักจะสูงมากในวันศุกร์และวันเสาร์และยอดขายจะต่ำในวันจันทร์และอังคาร ซึ่งจะมีแบบแผนทำนองนี้ทุกสัปดาห์ อันเป็นการแสดงให้เห็นลักษณะของ Periodic Variation

(4) การประมาณผลิตภาพของโรงงานอุตสาหกรรมโดยวัดจากปริมาณที่พนักงานผลิตได้ การสำรวจกระทำทุกวันเฉพาะกะกลางวันคือระหว่าง 06.00-18.00 น. โดยจะสำรวจเป็นช่วงเวลาทุก ๆ 12 ชั่วโมง (k=12) จนครบ 1 เดือน กรณีนี้จะเห็นว่าปริมาณผลผลิตจะขึ้นอยู่กับความเหน็ดเหนื่อยเมื่อยล้าของพนักงานซึ่งจะเริ่มผันแปรไปตามจำนวนชั่วโมงทำงานที่ค่อย ๆ เพิ่มจาก 1 ถึง 12 ชั่วโมง หมายความว่าพนักงานจะคึกคักเข้มแข็งในตอนเช้าและค่อย ๆ

เมื่อยล้าเหนื่อยอ่อน จนกระทั่งถึงชั่วโมงที่ 12 ของการทำงานซึ่งเมื่อยล้าที่สุด ดังนั้นจึงเห็นได้ว่าผลิตภาพของโรงงานมีลักษณะเป็น Periodic Variation ตามช่วงเวลาต่าง ๆ กัน

2. Ordered Population

คำว่า Ordered Population หมายถึงกลุ่มประชากรถูกจัดเรียงลำดับค่าตามค่าของตัวแปรจากมากไปหาน้อยหรือน้อยไปหามาก แต่ในทางปฏิบัตินั้นเราไม่อาจทราบค่าของตัวแปรได้ตราบใดที่ยังไม่มีการสำมะโน ดังนั้นเราจึงใช้ดัชนีบางประการเป็นเครื่องช่วยชี้ในการจัดเรียง เช่น จัดเรียงประชากรตามขนาดพื้นที่ที่ถือครองถ้าต้องการสำรวจภาวะเศรษฐกิจของเกษตรกร จัดเรียงห้างสรรพสินค้าตามจำนวนพนักงานหรือเงินทุนจดทะเบียนเมื่อต้องการประมาณยอดขาย จัดเรียงรายชื่อหัวหน้าครัวเรือนตามระดับรายได้เป็นต้น

Ordered Population มีผลดีที่เด่นชัดต่องานสำรวจแบบ Systematic คือทำให้กลุ่มตัวอย่างมีลักษณะของ Heterogeneity สูง ซึ่งมีผลให้ S^2_{wsy} มีค่าสูงหรือ ϱ มีค่าต่ำซึ่งจะยังผลให้ $V(\bar{x}_{sy})$ มีค่าต่ำและโดยปกติแล้ว $V(\bar{x}_{sy})$ จะมีค่าต่ำกว่า $V(\bar{x}_{ran})$ หรือนัยหนึ่ง ถ้าเราจัดกรอบตัวอย่างให้มีลักษณะเป็น Ordered Population ได้แล้วแผนสำรวจแบบ Systematic จะมีประสิทธิภาพสูงกว่าแบบแผนสำรวจแบบ SRS ตัวอย่างที่แสดงให้ดูแล้วสำหรับกรณีนี้ก็คือตัวอย่างข้อ จ. ในตัวอย่าง 4.1 ที่ผ่านมา

3. Random Order

ในหลายสถานการณ์เรามักพบว่าค่าของตัวแปรมักจะกระจัดกระจายกันอยู่ในชั้นภูมิในลักษณะเชิงสุ่ม คือไม่มีทั้งลักษณะของ Periodic Variation ขณะเดียวกันก็ไม่มีลักษณะของ Ordered Population ลองพิจารณาเปรียบเทียบจากตัวอย่างต่อไปนี้ ทั้งนี้สมมุติว่าทราบค่าของตัวแปรหรือทราบค่าดัชนีบางตัวที่พอคาดหมายค่าของตัวแปรนั้นได้ ตัวอย่างต่อไปนี้เป็นตัวเลขสมมุติ โดยสมมุติว่าทราบค่าของตัวแปรในกลุ่มประชากรขนาด $N = 12$ และเราจัดออกได้เป็น 3 ชั้นภูมิ ดังนี้คือ

วิธีที่ 1

1,4,7,10 | 2,5,8,11 | 3,6,9,12