บทที่ 13
การสุ่มตัวอย่างและการทดสอบความสุ่มของตัวอย่าง

13.1 การสุ่มตัวอย่าง (Random Sampling)

ดังได้กล่าวมาแล้วในบทแรก ๆ ว่าการเก็บรวบรวมข้อมูลนั้นเราไม่สามารถที่จะเก็บรวบรวม จากทุกหน่วยของประชากรได้ เราจึงต้องใช้วิธีการสุ่มตัวอย่างมาเพียงบางส่วนมาศึกษา เพื่อรวบรวมข้อมูลสำหรับใช้เป็นพื้นฐานในการสรุปผล หรืออ้างอิงเกี่ยวกับคุณลักษณะของประชากร ซึ่งตัวอย่างที่จะเป็นตัวแทนที่ดีของประชากรหรือไม่นั้นขึ้นอยู่กับว่าตัวอย่างนั้นจะมีลักษณะเหมือนกับประชากรมากน้อยแค่ไหน

โดยทั่ว ๆ ไปแล้วเราแบ่งตัวอย่างออกเป็น 3 ชนิด คือ

ก. ตัวอย่างชนิดสุ่ม (Random sample) เป็นตัวอย่างที่เลือกมาโดยไม่เจาะจงหรือลำเอียง

ข. ตัวอย่างชนิดเจาะจง (Purposive sample) เช่น เรียงบ้านเลขที่ไว้เป็นลำดับถ้านับได้เลขลำดับ 10, 20, 30 .... ก็เลือกบ้านที่อยู่ในลำดับนั้นมาเป็นตัวอย่างคือเลือกมา 1 บ้าน จากทุก ๆ 10 บ้าน

ค. ตัวอย่างชนิดผสม (Mixed sample) โดยใช้แบบที่ 1 และ 2 ผสมกัน เช่นต้องการสำรวจดูว่าในบรรทัดหนึ่งของหนังสือเล่มหนึ่งจะบรรจุคำได้กี่คำ ก็จะทำการเลือกหน้าในหนังสือมา 1 หน้า ในทุก ๆ 5 หน้า แล้วเลือกบรรทัดโดยสุ่มมาหน้าละ 3 บรรทัด

สำหรับวิธีการสุ่มตัวอย่างที่นิยมใช้กันมาก ซึ่งมีอยู่ 4 วิธีนั้นได้กล่าวถึงมาแล้วในตอนบทแรก ๆ จะไม่ขอกล่าวในที่นี้

เมื่อได้ตัวอย่างมาแล้วเรามักจะมีปัญหาว่า ตัวอย่างที่ได้นั้นเป็นตัวอย่างแบบสุ่มจริงหรือไม่ เราจึงต้องมีการทดสอบกันที่เรียกว่าการทดสอบความสุ่มของตัวอย่าง

13.2 <u>การทดสอบความสุ่มของตัวอย่าง</u> มีด้วยกันหลายวิธีแต่ในที่นี้จะขอกล่าวถึงวิธีที่เรียกว่า Run test

Run เป็นอนุกรมของสัญญลักษณ์ที่เหมือนกัน ซึ่งอาจจะตามหรือนำสัญญลักษณ์อื่น ๆ หรือไม่มีสัญญลักษณ์อื่นตาม หรือนำเลย ผลรวมของ Run (R) ในการจัดเรียงของสัญญลักษณ์สองชนิดหรือมากกว่าจะนำไปใช้ในการตรวจสอบการสุ่ม หรือที่เรียกว่าการทดสอบความสุ่มของตัวอย่าง และยังนำไปใช้ในการทดสอบสมมติฐาน H_o ที่ว่าสองตัวอย่างสุ่ม ซึ่งเป็นอิสระกันมาจากประชากรที่เป็นแบบเดียวกันและยังใช้ทดสอบการสุ่มที่อาศัย Runs ที่อยู่เหนือและใต้มัธยฐาน (Runs Above and Below the Median) หรือใช้ทดสอบสอบการสุ่มที่อาศัย Runs Up and Down

13.2.1 <u>ผลรวมของ Run</u> (Total Number of Runs)

ใช้ทดสอบการสุ่มของอนุกรมของสัญญลักษณ์สองชนิด ซึ่งมี 2 กรณีด้วยกันคือ

ก. กรณีที่ตัวอย่างมีขนาด n_1 และ n_2 น้อยกว่า 20 ค่าวิกฤตจะหาได้จากตารางที่

ข. กรณีที่ตัวอย่าง n_1 และ n_2 โตกว่า 20 การทดสอบจะอาศัยตัวสถิติ Z โดยที่ค่าเฉลี่ยของ r และความคลาดเคลื่อนมาตรฐานของ r คือ

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2\ (n_1+n_2-1)}}$$

และ $$Z = \frac{r-\mu_r}{\sigma_r}$$

หรือ $$Z = \frac{|r-\mu_r| - .5}{\sigma_r}$$

<u>ตัวอย่างที่ 1</u> จงทดสอบความสุ่มของ

ก. การโยนเหรียญ 1 อัน 20 ครั้ง

ตัวอย่างชุดที่ 1 : T T T T H H T T T T T T H H H H H H H H

ตัวอย่างชุดที่ 2 : H T H T H T H T H T H T H T H T H T H T

ตัวอย่างชุดที่ 3 : T T H H T H T T H T H H T T T H T H H H

จากตัวอย่างทั้ง 3 เรามองเห็นได้ชัดเจน จากตัวอย่างชุดที่ 1 ได้หัวติด ๆ กัน และก้อยติด ๆ กันมากครั้งเกินไป ส่วนตัวอย่างชุดที่ 2 ได้หัว และก้อยสลับกันไป และตัวอย่างชุดที่ 3 มีหัวและก้อยคละกันไป ดังนั้นเราจะเห็นได้ว่าตัวอย่างชุดที่ 1 และ 2 ดูจะมีความสุ่มน้อยกว่าตัวอย่างชุดที่ 3 แทนที่เราจะสังเกตด้วยตาเราอาจจะนับจากการสลับและของสัญญลักษณ์ทั้งสองว่ามีการสลับมากน้อยแค่ไหนแล้วพิจารณาความสุ่มจากจำนวนการสลับนี้ โดยให้ r = จำนวนหมู่ที่เกิดขึ้นของสัญญลักษณ์ทั้งสอง

$\therefore$ จากตัวอย่างนี้เราได้ค่า r ดังต่อไปนี้

ตัวอย่างชุดที่ 1 : $\frac{1}{\text{TTTT}}\ \frac{2}{\text{HH}}\ \frac{3}{\text{TTTTTT}}\ \frac{4}{\text{HHHHHHHH}}$

$\therefore$ ในตัวอย่างชุดที่ 1 มีค่าเท่ากับ 4

ตัวอย่างที่ 2 : $\frac{1}{\text{H}}\ \frac{2}{\text{T}}\ \frac{3}{\text{H}}\ \frac{4}{\text{T}}\ \frac{5}{\text{H}}\ \frac{6}{\text{T}}\ \frac{7}{\text{H}}\ \frac{8}{\text{T}}\ \frac{9}{\text{H}}\ \frac{10}{\text{T}}\ \frac{11}{\text{H}}\ \frac{12}{\text{T}}\ \frac{13}{\text{H}}\ \frac{14}{\text{T}}\ \frac{15}{\text{H}}\ \frac{16}{\text{T}}$

$\frac{17}{\text{H}}\ \frac{18}{\text{T}}\ \frac{19}{\text{H}}\ \frac{20}{\text{T}}$

$\therefore$ ในตัวอย่างชุดที่ 2 มีค่าเท่ากับ 20

ตัวอย่างที่ 3 : $\frac{1}{\text{TT}}\ \frac{2}{\text{HH}}\ \frac{3}{\text{T}}\ \frac{4}{\text{H}}\ \frac{5}{\text{TT}}\ \frac{6}{\text{H}}\ \frac{7}{\text{T}}\ \frac{8}{\text{HH}}\ \frac{9}{\text{TTT}}\ \frac{10}{\text{H}}\ \frac{11}{\text{T}}\ \frac{12}{\text{HHH}}$

$\therefore$ r ในตัวอย่างชุดที่ 2 มีค่าเท่ากับ 12

จากค่า r ของตัวอย่างทั้ง 3 ชุดจะเห็นได้ว่าในตัวอย่างชุดที่ 1 ค่าของ r น้อยเกินไป ส่วนตัวอย่างชุดที่ 2 ค่าของ r ก็มากเกินไปดังนั้นจึงมีผู้คิดหาขอบเขตของ r (ค่าวิกฤต ของ r) และสร้างเป็นตารางไว้ ซึ่งตารางแสดงไว้ตอนท้ายเล่ม ในตารางที่ VIII ในตารางนั้นจะมีทั้งตาราง F_1 และ F_2 โดยที่ตาราง F_1 เป็นตารางสำหรับค่าขอบเขตล่างและ F_2 สำหรับค่าของเขตบนของ r ค่าของ r นี้ขึ้นอยู่กับ n_1 และ n_2 และระดับนัยสำคัญตารางของ r ที่แสดงไว้นี้ใช้ระดับนัยสำคัญ 0.05 และ n_1, n_2 มีค่าตั้งแต่ 2 ถึง 20

จากตัวอย่างข้างต้นพิจารณาค่าของ r

จะได้ว่า $n_1 = 10$, $n_2 = 10$, $\alpha = .05$

เปิดตารางจะได้ $r = 6$ ถึง 16

ตัวอย่างชุดที่ 1 : $r = 4$ ค่าของ r เล็กเกินไปผลจึงมีนัยสำคัญแสดงว่าข้อมูลได้ไม่สุ่มพอ

ตัวอย่างชุดที่ 2 : $r = 20$ ค่าของ r ใหญ่เกินไป ผลจึงมีนัยสำคัญแสดงว่าความสุ่มยังเป็นที่น่าสงสัย

ตัวอย่างชุดที่ 3 : $r = 12$ ค่าของ r อยู่ในขอบเขตที่หามาได้จึงยอมรับว่าข้อมูลได้จากวิธีสุ่ม

ในกรณีที่ตัวอย่างมีขนาด n_1 และ n_2 โตกว่า 20 เราจะใช้วิธีข้อ (ข) ทดสอบดูความสุ่ม เนื่องจากในตาราง r ใช้ได้เฉพาะค่า n_1 และ n_2 ที่มีค่า n_1 และ n_2 ตั้งแต่ 2 ถึง 20 เราจึงต้องใช้การประมาณค่าเฉลี่ย และความคลาดเคลื่อนมาตรฐานของ r และประมาณว่าการแจกแจงของ r เป็นการแจกแจงแบบปกติ แล้วพิจารณาจากค่าของ Z ว่ามีนัยสำคัญหรือไม่ที่ α ที่กำหนดให้

ค่าเฉลี่ยของ r $(\mu_r) = \dfrac{2n_1n_2}{n_1+n_2} + 1$

ความคลาดเคลื่อนมาตรฐานของ r $(\sigma_r) = \sqrt{\dfrac{2n_1n_2(2n_1n_2-n_1n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$

และ $Z = \dfrac{r-\mu_r}{\sigma_r}$

ตัวอย่างที่ 2 จงทดสอบว่าคะแนน 2 ชุด ต่อไปนี้ มีความแตกต่างกันหรือไม่

ชุด A : 20, 55, 29, 24, 75, 56, 31, 45

ชุด B : 23, 8, 24, 15, 8, 6, 15, 15, 21, 23, 16, 15, 24, 15, 21, 15, 18, 14, 22, 15, 14

เรียงคะแนนทั้ง 2 ชุดใหม่ ได้ดังนี้

$$\underbrace{\begin{matrix}6, & 8, & 8, & 14, & 14, & 15, & 15, & 15, & 15, & 15, & 15, & 16, & 18, \\ B & B & B & B & B & B & B & B & B & B & B & B & B\end{matrix}}_{1}\ \underbrace{\begin{matrix}20, \\ A\end{matrix}}_{2}$$

$$\underbrace{\begin{matrix}21, & 21, & 22, & 23, & 23, & 24, \\ B & B & B & B & B & B\end{matrix}}_{3}\ \underbrace{\begin{matrix}24, \\ A\end{matrix}}_{4}\ \underbrace{\begin{matrix}24, \\ B\end{matrix}}_{5}\ \underbrace{\begin{matrix}29, & 31, & 45, & 55, & 56, & 75 \\ A & A & A & A & A & A\end{matrix}}_{6}$$

$$\therefore \quad r = 6$$

$$n_1 = 8, \quad n_2 = 21$$

หาค่า $\mu_r = \dfrac{2n_1n_2}{n_1 + n_2} + 1$

แทนค่า n_1 และ n_2 จะได้

$$\mu_r = \frac{2 \times 8 \times 21}{8 + 21} + 1$$

$$= \frac{336}{29} + 1 = \frac{336 + 29}{29}$$

$$= \frac{365}{29} = 12.59$$

$$\sigma_r = \sqrt{\frac{(2)(8)(21)\left[(2)(8)(21)-8-21\right]}{(8+21)^2\ (8+21-1)}}$$

$$= \sqrt{\frac{336 \times (336 - 8 - 21)}{(29)(29)\ (28)}}$$

$$= \sqrt{\frac{336 \times 307}{29 \times 29 \times 28}} = \sqrt{\frac{103152}{23848}}$$

$$\doteq \sqrt{4.38} = 2.09$$

จาก $$Z = \frac{r - \mu_r}{\sigma_r}$$

$$= \frac{6 - 12.59}{2.09} = \frac{-6.59}{2.09} = -3.15$$

ใช้ $\alpha = .05$

$\therefore$ จะปฏิเสธ H_o เมื่อ $Z < 1.96$ และ $Z > 1.96$

ดังนั้นเราไม่ยอมรับว่าคะแนนทั้ง 2 ชุด มีลักษณะเหมือนกันแสดงว่าคะแนนทั้ง 2 ชุด นี้แตกต่างกัน

ตัวอย่างที่ 3 ผู้จัดการห้างสรรพสินค้าแห่งหนึ่งสนใจที่จะทราบว่าการเข้ามาซื้อสินค้าของลูกค้าที่เป็นผู้ชาย (ช) กับลูกค้าที่เป็นเพศหญิง (ญ) เป็นการจัดเรียงแบบสุ่มหรือไม่ จากการสังเกตลูกค้า 50 รายได้ข้อมูลดังนี้

ช ญ ช ญ ช ช ช ญ ญ ช ญ ช ญ ช ญ ช ช ช ช ญ ช ญ ช ญ ช ช ญ ญ ญ ช ญ

ช ญ ช ญ ช ช ญ ช ช ญ ช ช ช ช ญ ช ญ ช ช

ผลจากการสังเกตจะเป็นอย่างไรใช้ $\alpha = .05$

จากโจทย์ ได้ค่า $r = 35$, $n_1 = 30$, $n_2 = 20$

$$\mu_r = \frac{(2)(30)(20)}{30 + 20} + 1$$

$$= \frac{1200}{50} + 1 = \frac{1250}{50}$$

$$= 25$$

$$\sigma_r = \sqrt{\frac{(20)(30)(20)\ (2)(30)(20) - 30 - 20}{(30+20)^2\ \ (30+20-1)}}$$

$$= \sqrt{\frac{(1200)(1150)}{(50)(50)(49)}}$$

$$= \sqrt{\frac{13800}{1225}} = \sqrt{11.27}$$

$$= 3.36$$

จาก $z = \dfrac{r - \mu_r}{\sigma_r}$

แทนค่า μ_r, σ_r และ r จะได้

$$z = \frac{35 - 25}{3.36} = \frac{10}{3.36} = -.976$$

$\therefore$ ที่ $\alpha = .05$ เราจึงไม่ยอมรับว่าการเข้ามาซื้อสินค้าหาลูกค้าตามเพศเป็นแบบสุ่ม แสดงว่าการเข้ามาซื้อสินค้าของลูกค้าตามเพศไม่เป็นแบบสุ่มนั่นเอง

13.2.2 Runs Above and Below the Median

เป็นการทดสอบความสุ่มของตัวอย่าง โดยอาศัยอันดับที่ของค่าจากตัวอย่างที่เลือกสุ่มมา ค่าแต่ละค่าจะแทนด้วยอักษร a หรือ b ตามแต่ว่ามันจะมีค่ามากกว่า หรือน้อยกว่ามัธยฐานของตัวอย่าง แล้วใช้การทดสอบของ Runs ในแบบ 13.2.1 ประยุกต์เข้ากับอนุกรมของ a และ b ถ้าจำนวนตัวอย่างเป็นจำนวนคู่ และ $n > 25$ และประชากรเป็นแบบต่อเนื่องแล้วการแจกแจงของจำนวน run (R) จะเป็นแบบปกติที่มี

$$\mu_r = \frac{n}{2} + 1$$

$$\sigma_r = \sqrt{\frac{n(n-2)}{4(n-1)}}$$

และ $z = \dfrac{r - \mu_r}{\sigma_r}$

<u>ตัวอย่างที่ 4</u> สมมติได้ข้อมูลจากตัวอย่างขนาด 26 เป็นดังนี้

97 89 25 81 11 83 16 96 44 32 98 19 68

33 25 54 74 82 17 49 33 22 62 20 92 80

หามัธยฐานของตัวอย่างได้โดยเรียงลำดับข้อมูลจากน้อยไปมาก

11 16 17 19 20 22 25 25 32 33 33 44 49 54

62 68 74 80 81 82 83 89 92 96 97 98

ตำแหน่งที่มัธยฐานอยู่ คือ $\frac{n+1}{2} = \frac{26+1}{2} = \frac{27}{2} = 13.5$

$\therefore$ มัธยฐานมีค่าเท่ากับ $\frac{49+54}{2} = 51.5$

จากข้อมูลเราแทนค่าสังเกตด้วย a หรือ b โดยที่ a จะเป็นค่าที่สูงกว่ามัธยฐาน b เป็นค่าที่ต่ำกว่ามัธยฐาน ได้ดังนี้

a a b a b a b a b b a b a b b a a a b b b b a b a a

$n = 26, \quad n_1 = 9, \quad n_2 = 8, \quad r = 17$

$$\mu_r = \frac{n}{2} + 1 = \frac{26}{2} + 1$$

$$= \frac{28}{2} = 14$$

$$\sigma_r = \sqrt{\frac{n(n-2)}{4(n-1)}}$$

$$= \sqrt{\frac{(26)(24)}{(4)(25)}}$$

$$= \sqrt{9.24}$$

$$= 2.498$$

จาก
$$Z = \frac{r - \mu_r}{\sigma_r}$$

$$= \frac{17 - 14}{2.498} = \frac{3}{2.498}$$

$$= 1.2$$

เมื่อ $\alpha = .05$ แสดงว่าตัวอย่างที่ได้มาเป็นตัวอย่างสุ่ม

13.2.3 Runs Up and Down

เป็นการทดสอบความสุ่มของตัวอย่าง โดยอาศัยอันดับที่ของค่าสังเกตุ โดยพิจารณาค่าสังเกตทีละคู่ที่สุ่มได้ติดต่อกันถ้าค่าหลังมากกว่าค่าแรกก็แทนด้วยเครื่องหมายบวก และน้อยกว่าก็แทนด้วยเครื่องหมายลบ

จากตัวอย่างที่ 4 จะได้

– – + – + – + – – + – + – – + + + – + – – – + – + – –

ซึ่งจะได้ $r = 19$ แล้วใช้การทดสอบของ Runs ในแบบที่ 13.2.1 ประยุกต์เข้ากับอนุกรมของ + และ – หรือใช้สมมติฐานที่ว่าเป็นตัวอย่างสุ่ม และประชากรแบบต่อเนื่อง เมื่อตัวอย่างมีขนาดโต ($n > 20$) แล้วจำนวน Run Up และ Down จะมีการแจกแจงแบบปกติที่มี

$$\mu_r = \frac{1}{3}(2n - 1)$$

$$\sigma_r = \sqrt{\frac{1}{90(16n-29)}}$$

และ

$$Z = \frac{r - \mu_r}{\sigma_r}$$

จากตัวอย่างได้ค่า $r = 19$, $n = 26$

$$\therefore \quad \mu_r = \frac{1}{3}\left[2(26) - 1\right]$$

$$= 17$$

$$\sigma_r = \sqrt{\frac{1}{90\left[(16)(26) - 29\right]}}$$

$$= \sqrt{4.3} = 2.07$$

$$\therefore \quad Z = \frac{19 - 17}{2.07}$$

$$= \frac{2}{2.07} = 0.97$$

เมื่อ $\alpha = .05$ แสดงว่าตัวอย่างเป็นตัวอย่างสุ่ม

แบบฝึกหัด

1. จากการทดสอบวิชาคณิตศาสตร์ของนักเรียนชั้นประถมปีที่ 3 เป็นเด็กผู้ชาย 12 คน เด็กผู้หญิง 12 คน ได้คะแนนดังต่อไปนี้

ชาย : 86 69 72 65 113 65 118 45 141 104 41 50

หญิง : 55 40 22 58 16 7 9 16 26 26 20 15

จงตรวจสอบว่ามีความแตกต่างระหว่างคะแนนของเด็กผู้ชายและเด็กผู้หญิงหรือไม่ ใช้ $\alpha = .05$

2. จงทดสอบความสุ่มตัวอย่างของการวัดเส้นผ่าศูนย์กลางของเส้นลวด 15 เส้น โดยเลือกหยิบมาทีละอันทุก ๆ ครึ่งชั่วโมงจากโรงงาน และได้ข้อมูลมาดังนี้

ตัวอย่างชุดที่ 1	ตัวอย่างที่ 2	ตัวอย่างที่ 3
.501	.502	.505
.502	.501	.501
.502	.500	.502
.503	.503	.502
.502	.501	.506
.503	..502	.502
.503	.503	.507
.504	.509	.505
.505	.508	.503
.505	.509	.504
.506	.509	.502

(ต่อข้อ 2)

ตัวอย่างชุดที่ 1	ตัวอย่างชุดที่ 2	ตัวอย่างชุดที่ 3
.505	.510	.505
.507	.507	.506
.507	.510	.502

ใช้ α = .05

3. จากข้อมูลของจำนวน SO_2 ในอากาศใน 44 วัน มีดังนี้

วัน	SO_2	วัน	SO_2
1	.057	23	.051
2	.040	24	.063
3	.059	25	.060
4	.063	26	.049
5	.061	27	.040
6	.040	28	.044
7	.009	29	.058
8	.003	30	.032
9	.031	31	.018
10	.067	32	.017
11	.071	33	.017
12	.083	34	.030
13	.081	35	.053
14	.093	36	.054
15	.065	37	.085
16	.023	38	.081

วัน	SO_2	วัน	SO_2
17	.029	39	.041
18	.018	40	.037
19	.001	41	.063
20	.010	42	.073
21	.055	43	.055
22	.056	44	.048

จงทดสอบความสุ่มของตัวอย่าง โดยใช้ Runs Above and Below the Median ที่ $\alpha = .05$ และถ้า Median = 0.050

4. จงทดสอบความสุ่มของตัวอย่าง โดยใช้ Runs Up and Down จากข้อมูลต่อไปนี้ โดยใช้ $\alpha = .05$

76 88 01 35 34 49 17 89 19 41 14 99 13

23 79 40 15 19 01 66 33 31 15 16 54 03

11 93 78 87 50 23 46 14 27 12 38 12 20

15