L คือ ขีดจำกัดล่างของอันตรภาคชั้นที่ D_x ตกอยู่

i คือ ความกว้างของอันตรภาคชั้น

F_L คือ ความถี่สะสมของชั้นที่อยู่ติดกับ Q_X และเป็นช่วงคะแนนที่มีค่าต่ำกว่า

f_{D_x} คือ ความถี่ในชั้นที่ D_X ตกอยู่

การหาเปอร์เซ็นต์ไตล์

$$Px = L + \frac{i}{f_{P_x}}\left[\frac{X}{100}N - F_L\right]$$

เมื่อ $\frac{X}{100}N$ คือ ตำแหน่งเปอร์เซ็นต์ไตล์ที่ X

P_x คือ ค่าเปอร์เซ็นต์ไตล์ที่ X

L คือ ขีดจำกัดล่างของอันตรภาคชั้นที่ P_x ตกอยู่

i คือ ความกว้างของอันตรภาคชั้น

F_L คือ ความถี่สะสมของชั้นที่อยู่ติดกับ P_x และเป็นช่วงคะแนนที่มีค่าต่ำกว่า

ความสัมพันธ์ระหว่างควอไตล์ (Q_x) เดไซล์ (D_x) และ เปอร์เซ็นต์ไตล์ (P_x)

$Q_1 = P_{25}$

$Q_2 = P_{50} = D_5$

$Q_3 = P_{75}$

$D_1 = P_{10}$

$D_2 = P_{20}$

$\vdots \qquad \vdots$

$D_9 = P_{90}$

ตัวอย่าง จงหาเปอร์เซ็นต์ไตล์ที่ 20 ของคะแนนสอนของนักเรียน 125 คน ซึ่งมีการแจกแจงความถี่ดังนี้

คะแนน	ความถี่ (f)	ความถี่สะสม (cmf)
95 - 99	1	125
90 - 94	5	124
85 - 89	6	119
80 - 84	10	113
75 - 79	20	103
70 - 74	35	83
65 - 69	18	48
60 - 64	12	30
55 - 59	10	18
50 - 54	5	8
45 - 49	2	3
40 - 44	1	1
รวม	125	

วิธีทำ เปอร์เซ็นต์ไตล์ที่ 20 จะอยู่ในตำแหน่งที่ $\frac{20}{100}$ (125) = 25

ตำแหน่งที่ 25 อยู่ระหว่างความถี่สะสม 18 และ 30 ของอันตรภาคชั้น 50 - 54 และ 55 - 59

ความถี่สะสมเพิ่มขึ้น 30 - 18 = 12 คะแนนเพิ่มขึ้น 59.5 - 54.5 = 5 คะแนน

ความถี่สะสมเพิ่มขึ้น 25 - 18 = 7 คะแนนเพิ่มขึ้น $\frac{7}{5} \times 5$ = 2.92 คะแนน

ดังนั้น เปอร์เซ็นต์ไตล์ที่ 20 จะตรงกับคะแนน = 54.5 + 2.92

= 57.42 คะแนน

หรือ หาคะแนนเปอร์เซ็นต์ไตล์ที่ 20 จากสูตร

$$P_{20} = L + I\left[\frac{\frac{XN}{100} - F_L}{f_{P_x}}\right]$$

$L = 54.5,\ I = 5,\ X = 20,\ N = 125,\ F_L = 18 \qquad f_{P_x} = 12$

$$\therefore\ P_{20} = 54.5 + 5\left[\frac{\frac{20 \times 125}{100} - 18}{12}\right]$$

$= 57.42$ คะแนน

$\therefore$ เปอร์เซนต์ไตล์ที่ 20 ของคะแนนสอบของนักเรียนตรงกับคะแนน 57.42 คะแนน

การวัดการกระจายของข้อมูล

การวัดการกระจายของข้อมูลแบ่งออกเป็น 2 แบบ คือ

1. การวัดการกระจายสัมบูรณ์
2. การวัดการกระจายสัมพัทธ์

การวัดการกระจายสัมบูรณ์ เป็นการวัดการกระจายของข้อมูลเพียงชุดเดียวเพื่อดูว่าข้อมูลแต่ละค่าในชุดนั้นมีความแตกต่างกันมากน้อยเพียงใด มีอยู่ 4 ชนิด ได้แก่

1. พิสัย คือ ค่าที่ใช้วัดการกระจายที่ได้จากผลต่างระหว่างข้อมูลที่มีค่าสูงสุดและข้อมูลที่มีค่าต่ำสูง

$$X_{max} - X_{min}$$

2. ส่วนเบี่ยงเบนควอไตล์ (Q.D.) คือ ค่าที่ใช้วัดการกระจายที่หาได้จากครึ่งหนึ่งของความแตกต่างระหว่างควอไตล์ที่ 3 (Q_3) และควอไตล์ที่ 1 (Q_3)

นั่นคือส่วนเบี่ยงเบนควอไตล์ (Q.D.) $= \dfrac{Q_3 - Q_1}{2}$

3. ส่วนเบี่ยงเบนเฉลี่ย (M.D.) คือ ค่าที่ใช้วัดการกระจายของข้อมูลที่ได้จากการเฉลี่ยค่าสัมบูรณ์ของความแตกต่างระหว่างค่าของข้อมูลแต่ละค่าจากค่ากลางของข้อมูลชุดนั้น

– ข้อมูลไม่ได้แจกแจงความถี่

$$\text{ส่วนเบี่ยงเบนเฉลี่ย} = \frac{\sum_{i=1}^{N} \left| X_i - \overline{X} \right|}{N}$$

– ข้อมูลที่แจกแจงความถี่

$$\text{ส่วนเบี่ยงเบนเฉลี่ย} = \sum_{i=1}^{N} f_i \frac{\left| X_i - \overline{X} \right|}{N}$$

4. ส่วนเบี่ยงเบนมาตรฐาน (S.D.)

ส่วนเบี่ยงเบนมาตรฐานเป็นวิธีการวัดการกระจายของข้อมูล เป็นวิธีที่นักสถิติทั่ว ๆ ไปยอมรับว่าเป็นวิธีที่ใช้วัดการกระจายได้ดีที่สุด เมื่อเทียบกับวิธีการวัดการกระจายทั้งสามวิธีที่กล่าวมาแล้ว

การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่ได้แจกแจงความถี่

ถ้า $X_1, X_2, X_3, \ldots, X_N$ เป็นข้อมูล N จำนวนและมีค่าเฉลี่ยเป็น $\overline{X}$

$$\text{ส่วนเบี่ยงเบนมาตรฐาน (S.D.)} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N}}$$

$$\text{หรือ} \quad \text{S.D.} = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \overline{X}^2}$$

ตัวอย่าง จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลต่อไปนี้ 3, 6, 9, 12, 15

วิธีทำ

$$\text{MI} \quad \overline{X} = \sum_{i=1}^{N} X_i \Big/ N$$

$$\sum_{i=1}^{N} (X_i - \overline{X})^2 = \frac{3+6+9+12+15}{5} = \text{g}$$

$$= (3-9)^2 + (6-9)^2 + (9-9)^2 + (12-9)^2 + (15-9)^2$$

$$= 36 + 9 + 0 + 9 + 36$$

$$= \text{go}$$

$$\sum_{i=1}^{N}(X_i-\overline{X})^2 \Big/ N = \frac{90}{5} = 18$$

$$\therefore \text{S.D.} = \sqrt{18}$$

ส่วนเบี่ยงเบนมาตรฐานของข้อมูลนี้คือ 4.24

การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่แจกแจงความถี่แล้ว

– โดยวิธีตรง

$$\text{ส่วนเบี่ยงเบนมาตรฐาน (S.D.)} = \sqrt{\frac{\sum_{i=1}^{N} f_i X_i^2 - \overline{X}^2}{N}}$$

$$\text{หรือ S.D.} = \sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2}{N} - \overline{X}^2}$$

เมื่อ f_i คือ ความถี่ของข้อมูลของอันตรภาคชั้นที่ i

X_i เป็นจุดกึ่งกลางของอันตรภาคชั้นที่

k จำนวนอันตรภาคชั้น

N จำนวนข้อมูลทั้งหมด หรือผลรวมของความถี่ของทุก ๆ อันตรภาคชั้น

$\overline{X}$ เป็นค่าเฉลี่ยเลขคณิตของข้อมูลทั้งหมด

– การหาส่วนเบี่ยงเบนมาตรฐานโดยวิธีทอนค่าข้อมูลหรือวิธีลัด

$$\text{ส่วนเบี่ยงเบนมาตรฐาน (S.D.)} = I\sqrt{\frac{\sum_{i=1}^{k} f_i d_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i d_i}{N}\right)^2}$$

I คือ ความกว้างของอันตรภาคชั้น

$d_i = \frac{X_i - a}{I}$ เมื่อ a คือ จุดกึ่งกลางชั้นที่มีความถี่สูงที่สุด

X_i คือ จุดกึ่งกลางชั้นของแต่ละชั้น

ตัวอย่าง จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลจากตารางแจกแจงความถี่ต่อไปนี้

คะแนน	ความถี่
43 - 47	1
38 - 42	3
33 - 37	11
28 - 32	15
23 - 27	8
18 - 22	2
รวม	40

วิธีทำ หาส่วนเบี่ยงเบนมาตรฐานโดยวิธีตรงโดยจัดเรียงคะแนนข้อมูลใหม่ดังนี้

คะแนน	ความถี่ (f_i)	จุดกึ่งกลาง X_i	$f_i X_i$	$f_i X_i^2$
18 - 22	2	20	40	800
23 - 27	8	25	200	5,000
28 - 32	15	30	450	13,500
33 - 37	11	35	385	13,475
38 - 42	3	40	120	4,800
43 - 47	1	45	45	2,025
	$\sum_{i=1}^{6} f_i = 40$		$\sum_{i=1}^{6} f_i X_i = 1,240$	$\sum_{i=1}^{6} f_i X_i^2 = 39,600$

$$\overline{X} = \sum_{i=1}^{k} f_i X_i \Big/ N = \frac{1240}{40} = 31$$

$$S.D. = \sqrt{\frac{\sum_{i=1}^{k} f_i X_i^2}{N} - \overline{X}^2}$$

$$= \sqrt{\frac{39,600}{40} - (31)^2}$$

$$= \sqrt{990 - 961}$$

$$= \sqrt{29}$$

$$S.D. = 5.39$$

ส่วนเบี่ยงเบนมาตรฐานโดยวิธีตรงของคะแนน คือ 5.39

การหาส่วนเบี่ยงเบนมาตรฐานโดยวิธีลัด

คะแนน	ความถี่ (f)	จุดกึ่งกลาง X_i	$d_i = \frac{X_i - a}{I}$	$f_i d_i$	$f_i d_i^2$
18 - 22	2	20	-2	-4	8
23 - 27	8	25	-1	-8	8
28 - 32	15	30	0	0	0
33 - 37	11	35	1	11	11
38 - 42	3	40	2	6	12
43 - 47	1	45	3	3	9
	$\sum_{i=1}^{k} f_i = 40$			$\sum_{i=1}^{6} f_i d_i = 8$	$\sum_{i=1}^{6} f_i d_i^2 = 48$

จาก $S.D. = I\sqrt{\frac{\sum_{i=1}^{k} f_i d_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i d_i}{N}\right)^2}$

$$= 5\sqrt{\frac{48}{40} - \frac{64}{1600}}$$

$$= 5\ (1.08)$$

$$S.D. = 5.39$$

ส่วนเบี่ยงเบนมาตรฐานโดยวิธีลัดของคะแนนคือ 5.39

ความแปรปรวน(Variance) คือ ส่วนเบี่ยงเบนมาตรฐานยกกำลังสอง นั่นคือ ความแปรปรวนของข้อมูลที่ไม่ได้แจกแจงความถี่คือ

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N}$$

ความแปรปรวนของข้อมูลที่แจกแจงความถี่ คือ

$$S^2 = \frac{\sum_{i=1}^{k} f_i\ (X_i - \overline{X})^2}{N}$$

หรือวิธีลัด

$$S^2 = I^2 \left\{ \frac{\sum_{i=1}^{k} f_i d_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i d_i}{N}\right)^2 \right\}$$

ถ้าข้อมูลมี k ชุด โดยมี N เป็นจำนวนข้อมูลและ S_t^2 เป็นความแปรปรวนของข้อมูล k ชุดแล้ว

$$\text{ความแปรปรวนรวม}\ (S_t^2) = \frac{\sum_{i=1}^{N} X_i^2\ \text{รวม}}{N} - \overline{X}^2\ \text{รวม}$$

$$\text{หรือ}\quad S_t^2 = \sum_{i=1}^{k} \frac{N_i}{N} \left[S_i^2 + (\overline{X}_i - \overline{X}_t)^2 \right]$$

จากตัวอย่างในการหาค่าส่วนเบี่ยงเบนมาตรฐานจากตารางที่กล่าวมาแล้ว ถ้าเราจะหาค่าความแปรปรวนของข้อมูลดังกล่าว จะเท่ากับกำลังสองของส่วนเบี่ยงเบนมาตรฐานคือ $(5.39)^2 = 29.052$

คุณสมบัติที่สำคัญของส่วนเบี่ยงเบนมาตรฐาน

1. ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็นบวกเสมอ

2. ถ้าคำนวณหาส่วนเบี่ยงเบนมาตรฐานโดยใช้ค่ากลางของข้อมูลชนิดอื่น ๆ ที่ไม่ใช่ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานที่ได้จะมีค่ามากกว่าส่วนเบี่ยงเบนมาตรฐานที่ใช้ค่าเฉลี่ยเลขคณิตเสมอ

3. ถ้าข้อมูล 2 ชุด ประกอบด้วยข้อมูล N_1 และ N_2 จำนวนมีค่าเฉลี่ยเลขคณิตเท่ากันแต่ความแปรปรวนของข้อมูล ชุดที่หนึ่งเป็น s_1^2 และข้อมูลชุดที่สองเป็น s_2^2 ตามลำดับ แล้วความแปรปรวนของข้อมูลทั้งสองชุดนี้ จะเท่ากับ $= \dfrac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2}$

4. ถ้า $y = ax + b$ แล้ว $s_y = |a| \, s_x$

เมื่อ s_x คือ ส่วนเบี่ยงเบนมาตรฐานข้อมูลชุดเดิม

s_y คือ ส่วนเบี่ยงเบนมาตรฐานข้อมูลชุดใหม่

โดยที่ $y = ax + b$

การวัดการกระจายสัมพัทธ์

การวัดการกระจายสัมพัทธ์ใช้ในการเปรียบเทียบข้อมูลตั้งแต่สองชุดขึ้นไปเพื่อตัดสินว่าชุดใดมีการกระจายมากกว่าหรือน้อยกว่า เราวัดการกระจายสัมพัทธ์ด้วยสัมประสิทธิ์ต่อไปนี้

1. สัมประสิทธิ์ของพิสัย $= \dfrac{X_{max} - X_{min}}{X_{max} + X_{min}}$

เมื่อ X_{max} คือ ค่ามากที่สุดของข้อมูล

X_{min} คือ ค่าที่น้อยที่สุดของข้อมูล

2. สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

เมื่อ Q_3 คือ ค่าข้อมูลที่ตรงกับตำแหน่งควอไทล์ที่สาม

และ Q_1 คือ ค่าข้อมูลที่ตรงกับตำแหน่งควอไทล์ที่หนึ่ง

3. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย $= \frac{M.D.}{\overline{X}}$

เมื่อ M.D. คือ ค่าส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่ต้องการหา

และ $\overline{X}$ คือ ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

4. สัมประสิทธิ์ของการแปรผัน $= \frac{S}{\overline{X}}$

เมื่อ S คือ ค่าของส่วนเบี่ยงเบนมาตรฐาน

และ $\overline{X}$ คือ ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

สัมประสิทธิ์ของการแปรผัน เป็นการวัดการกระจายที่นิยมใช้กันมากที่สุดมักจะเขียนในรูปของร้อยละโดยคูณด้วย 100

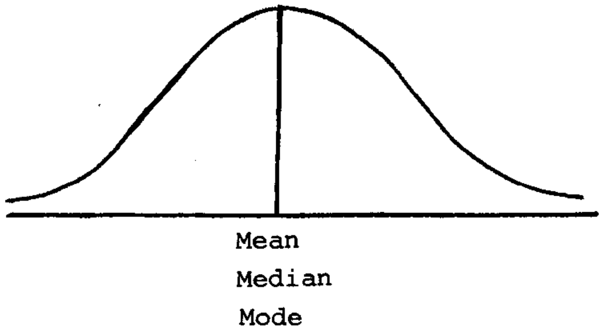
ความสัมพันธ์ระหว่างการแจกแจงความถี่ ค่ากลางและการกระจายของข้อมูล

การแจกแจงความถี่ของตัวแปร โดยทั่วไปเส้นโค้งของการแจกแจงความถี่ แบ่งออกเป็น 3 แบบคือ

1. **เส้นโค้งปกติหรือรูประฆัง** (Normal or bell-shaped curve)

Mean = Median = Mode

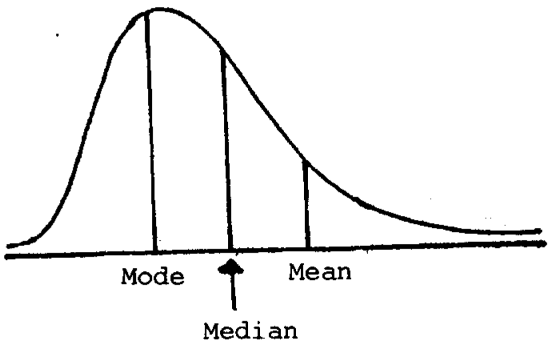
เส้นโค้งปกติ



2. **เส้นโค้งเบ้ทางขวาหรือทางบวก** (Positively skewed curve)

Mode < Median < Mean

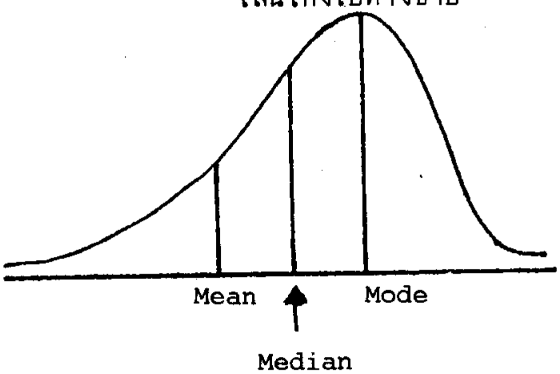
เส้นโค้งเบ้ทางขวา



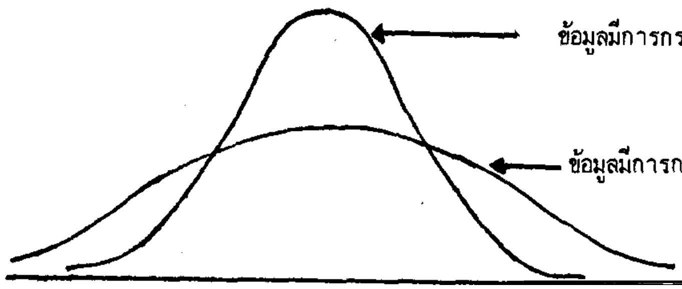
3. **เส้นโค้งเบ้ทางซ้ายหรือทางลบ**

Mean < Median < Mode

เส้นโค้งเบ้ทางซ้าย



เส้นโค้งปกติจะมีความโด่งมากหรือน้อยขึ้นอยู่กับการกระจายของข้อมูล ถ้าข้อมูลมีการกระจายมากเส้นโค้งปกติ จะมีความโด่งน้อย แต่ถ้ามีการกระจายน้อยเส้นโค้งปกติจะมีความโด่งมาก



ค่ามาตรฐาน

ค่ามาตรฐาน (z) ใช้ในการเปรียบเทียบค่าของข้อมูลของตัวแปรตั้งแต่สองตัวขึ้นไป ว่ามีความแตกต่างกันหรือไม่เพียงใด โดยเปลี่ยนค่าของข้อมูล ของตัวแปรแต่ละตัวให้เป็นค่ามาตรฐานเสียก่อนแล้วจึงเปรียบเทียบกัน

ถ้า x_i เป็นตัวแปรของข้อมูลที่มีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานเป็น $\bar{x}$ และ s ตามลำดับแล้วค่ามาตรฐานของ x_i คือ

$$z_i = \frac{x_i - \bar{x}}{s}$$

หรือ อาจจะกล่าวได้ว่า ค่ามาตรฐานเป็นค่าที่บอกให้ทราบว่าความแตกต่างระหว่างค่าของข้อมูลนั้น ๆ กับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น เป็นกี่เท่าของส่วนเบี่ยงเบนมาตรฐาน

ตัวอย่าง นักเรียนคนหนึ่งสอบวิชาสถิติและวิชาภาษาอังกฤษ ซึ่งมีคะแนนเต็ม 100 คะแนนเท่ากัน สอบวิชาสถิติได้ 72 คะแนน และวิชาภาษาอังกฤษได้ 75 คะแนน ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของวิชาสถิติของนักเรียนในห้องนี้เป็น 70 และ 10 คะแนนตามลำดับ และคะแนนเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานของวิชาภาษาอังกฤษเป็น 73 และ 16 คะแนนตามลำดับ จงเปรียบเทียบดูว่านักเรียนคนนี้เรียนวิชาไหนได้ดีกว่ากัน

วิธีทำ

หาคะแนนมาตรฐานของทั้งสองวิชา

คะแนนมาตรฐานของคะแนนสอบวิชาสถิติ $= \dfrac{x_i - \bar{x}}{s}$

$= \dfrac{72 - 70}{10}$

$= 0.2$

คะแนนมาตรฐานของคะแนนสอบวิชาภาษาอังกฤษ $= \dfrac{75 - 73}{16}$

$= 0.125$

ค่ามาตรฐานของคะแนนสอบวิชาสถิติของนักเรียนคนนี้สูงกว่าค่ามาตรฐานของคะแนนสอบวิชาภาษาอังกฤษ แสดงว่านักเรียนคนนี้เรียนวิชาสถิติได้ดีกว่าวิชาภาษาอังกฤษ

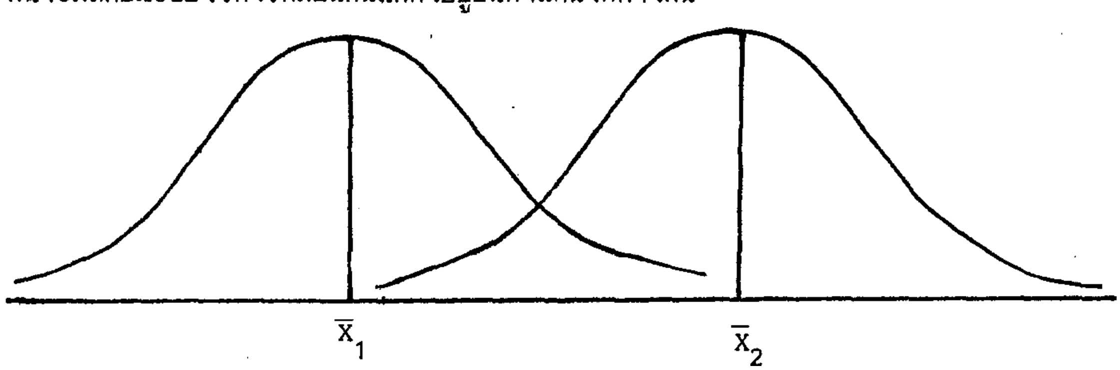
คุณสมบัติของค่ามาตรฐาน

1. $\bar{z} = 0$ (ค่ามาตรฐานมีค่าเฉลี่ยเลขคณิตเป็น 0)
2. $s_z = 1$ (ส่วนเบี่ยงเบนมาตรฐานของค่ามาตรฐานเป็น 1 เสมอ)

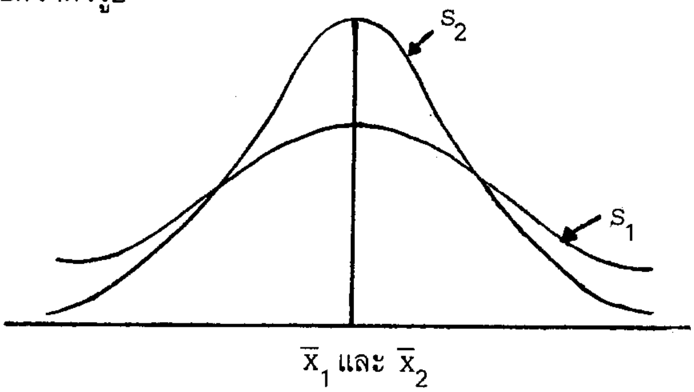
การแจกแจงปกติ

เส้นโค้งของความถี่ที่พบเสมอ ๆ มักจะมีรูปเป็นรูประฆัง ซึ่งมักเรียกว่าเส้นโค้งปกติ และการแจกแจงความถี่ของข้อมูลซึ่งให้เส้นโค้งมีลักษณะเป็นรูประฆังเรียกว่าการแจกแจงปกติ โดยปกติสมการของโค้ง จะขึ้นอยู่กับค่าของส่วนเบี่ยงเบนมาตรฐานและค่าเฉลี่ยเลขคณิต

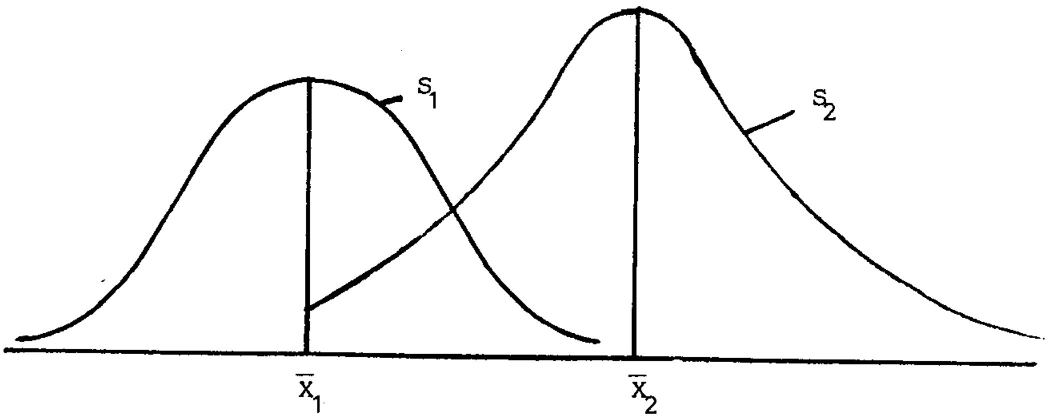
เส้นโค้งปกติ 2 รูป มีส่วนเบี่ยงเบนมาตรฐานเท่ากัน แต่ค่าเฉลี่ยเลขคณิตไม่เท่ากันจะมีลักษณะของโค้งเหมือนกันแต่ตั้งอยู่บนตำแหน่งที่ต่างกัน



เส้นโค้งปกติ 2 รูป ที่มีค่าเฉลี่ยเลขคณิตเท่ากัน แต่ส่วนเบี่ยงเบนมาตรฐานต่างกันจะมีจุดที่แสดงค่าเฉลี่ยเลขคณิตอยู่ที่ตำแหน่งเดียวกันบนแกนนอน แต่ส่วนโค้งที่มีส่วนเบี่ยงเบนมาตรฐานมากกว่าจะเตี้ยกว่าดังรูป



เส้นโค้งปกติ 2 รูป ที่มีค่าเฉลี่ยเลขคณิตต่างกันและส่วนเบี่ยงเบนมาตรฐานต่างกันจะมีลักษณะดังรูป



คุณสมบัติของเส้นโค้งปกติ

1. ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม จะมีค่าเท่ากัน และจะอยู่ ณ จุดที่เส้นตรงลากผ่านจุดโด่งสุดของเส้นโค้งนั้น ตั้งฉากกับแกนนอน

2. เส้นโค้ง จะมีเส้นตั้งฉากที่ลากผ่านค่าเฉลี่ยเลขคณิตเป็นแกนสมมาตร

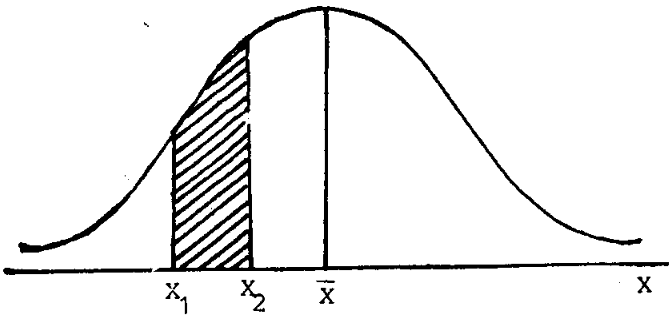
3. เส้นโค้ง จะเข้าใกล้แกนระนาบเมื่อเราต่อปลายเส้นโค้งทั้งสองข้างให้ห่างจากค่าเฉลี่ยเลขคณิตออกไป แต่จะไม่ตัดแกนนอน

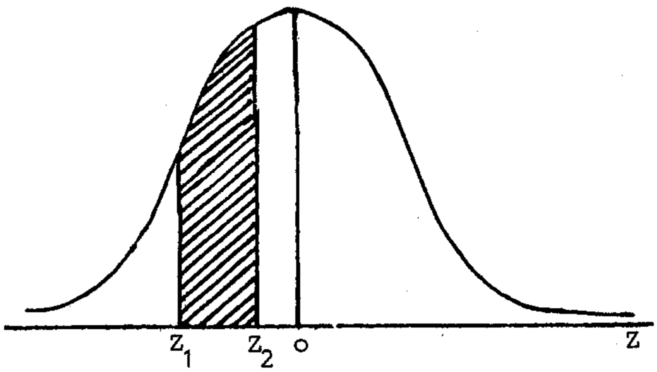
4. พื้นที่ใต้โค้งปกติมีค่าเท่ากับ 1 เสมอ

การหาพื้นที่ใต้โค้งปกติ

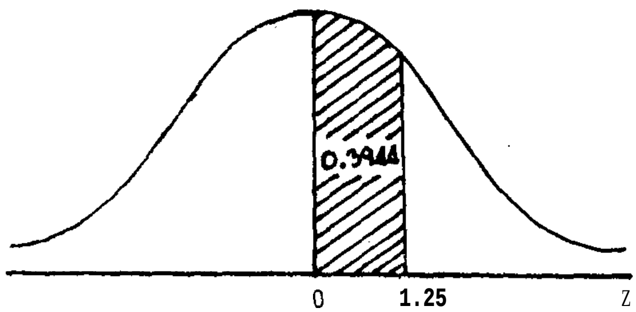
ถ้า x มีการแจกแจงปกติโดยมีค่าเฉลี่ยเลขคณิตเป็น $\bar{x}$ และส่วนเบี่ยงเบนมาตรฐานเป็น s เมื่อเปลี่ยน x ให้เป็น z สามารถจะพิสูจน์ได้ว่า z มีการแจกแจง ปกติโดยมีค่าเฉลี่ยเลขคณิตเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

พื้นที่ใต้เส้นโค้งปกติระหว่าง x และ x_2 จะเท่ากับพื้นที่ใต้เส้นโค้งปกติระหว่าง z_1 และ z_2 เมื่อ $z = \frac{x_1 - x}{s}$ และ $z = \frac{x_2 - \bar{x}}{s}$





เส้นโค้งปกติซึ่งมีค่าเฉลี่ยเลขคณิตเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 เรียกว่าเส้นโค้งปกติมาตรฐาน ในการหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่างค่า z = 0 ถึง z ใด ๆ เราใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐาน ซึ่งแสดงพื้นที่ใต้เส้นโค้งปกติระหว่างค่า z = 0 และ z = 0.00, 0.01, 0.02,..., 3.88, 3.89 เช่นพื้นที่ใต้เส้นโค้งปกติระหว่าง z = 0 และ z = 1.25 ที่อ่านได้จากตารางเท่ากับ 0.3944



ในตารางไม่มีค่า z ที่เป็นจำนวนลบ แต่สามารถหาพื้นที่ใต้เส้นโค้งปกติมาตรฐานระหว่าง z ที่เป็นจำนวนลบ และ z = 0 ได้ เนื่องจากเส้นโค้งมีเส้นตั้งฉากกับแกนนอนเป็นแกนสมมาตร เช่น พื้นที่ใต้เส้นโค้งปกติระหว่าง z = -1.25 และ z = 0 หาได้จากการหาพื้นที่ใต้เส้นโค้งปกติระหว่าง z = 0 ถึง z = 1.25 ซึ่งได้เท่ากับ 0.3944 เนื่องจากพื้นที่ใต้เส้นโค้งปกติเท่ากับ 1 ดังนั้น พื้นที่ทางขวามือของ z = 0 กับพื้นที่ทางซ้ายมือของ z = 0 มีค่าเท่ากันคือ 0.5 อาศัยความรู้ดังกล่าวจะหาพื้นที่ใต้เส้นโค้งปกติทางขวามือหรือซ้ายมือของ z ใด ๆ และพื้นที่ใต้เส้นโค้งปกติระหว่างค่า z สองค่าใด ๆ ได้

การวิเคราะห์ความสัมพันธ์เชิงฟังก์ชันระหว่างข้อมูล

โดยทั่วไปความสัมพันธ์เชิงฟังก์ชันของข้อมูลแบ่งออกเป็น 2 ชนิด

ถ้าข้อมูลประกอบด้วยตัวแปร 2 ตัว คือ x กับ y ซึ่งมีความสัมพันธ์กันเชิงฟังก์ชัน x, y เป็นตัวแปรอิสระและตัวแปรตาม ตามลำดับ มี a, b, c เป็นค่าคงที่

1. **ความสัมพันธ์เชิงฟังก์ชั่นที่เป็นเส้นตรง**

สมการทั่วไป $Y = mx + C$

เมื่อ Y เป็นตัวแปรตาม

X เป็นตัวแปรอิสระ

m คือ ความชัน (Slope) ของเส้นตรง

C ค่าของระยะทางจากจุดเริ่มต้นมายังจุดที่เส้นตรงตัดแกนซึ่งเป็นค่าคงที่

2. **ความสัมพันธ์เชิงฟังก์ชั่นที่ไม่เป็นเส้นตรง** ได้แก่ ความสัมพันธ์เชิงฟังก์ชั่นเป็นแบบพาราโบล่า (Parabola) มีสมการทั่วไปคือ

$$Y = ax^2 + bx + C$$

3. **ความสัมพันธ์เชิงฟังก์ชั่นเป็นแบบเอกซ์โปเนนเชีล** มีสมการทั่วไปคือ

$$Y = ab^x$$

หรือ $\log y = \log a + x\log b$

4. **ความสัมพันธ์เชิงฟังก์ชั่นของข้อมูลที่อยู่ในอนุกรมเวลา** ข้อมูลที่อยู่ในอนุกรมเวลา คือ ข้อมูลที่แสดงความเปลี่ยนแปลงตามลำดับก่อนหลังของเวลาที่ข้อมูลนั้น ๆ เกิดขึ้น

$$Y = f(T)$$

Y คือ ค่าของฟังก์ชั่นเป็นตัวแปรตาม และ T เป็นตัวแปรอิสระ

นอกจากเนื้อหาวิชาสถิติในระดับมัธยมศึกษาตอนต้นและมัธยมศึกษาตอนปลาย โดยสังเขปดังได้กล่าวไปแล้วนั้นยังมีเนื้อหาอีกส่วนหนึ่งที่เป็นทั้งคณิตศาสตร์และสถิติ ได้แก่ วิธีเรียงสับเปลี่ยน (Permutation) วิธีจัดหมู่ (Combination) และทฤษฎีความน่าจะเป็นเบื้องต้น

วิธีเรียงสับเปลี่ยน (Permutation)

วิธีเรียงสับเปลี่ยน (Permutation) คือ การจัดเรียงลำดับของสิ่งของหลาย ๆ สิ่งในลักษณะต่าง ๆ

1. การดำเนินงานอย่างหนึ่งซึ่งมีสองขั้นตอนโดยขั้นที่ 1 มีวิธีเลือกทำได้ n_1 วิธี และแต่ละวิธีที่เลือกทำในขั้นที่ 1 แล้วมีวิธีเลือกทำงานในขั้นที่ 2 ได้ n_2 วิธี

$\therefore$ จำนวนวิธีที่เลือกทำงานทั้งสองขั้นคือ $n_1 \times n_2$ วิธี

2. ในการดำเนินงานอย่างหนึ่งซึ่งมี 2 ขั้นตอน โดยที่ขั้นที่ 1 มีวิธีเลือกทำได้ n วิธี แต่ละวิธีที่เลือกทำในขั้นที่ 1 มีวิธีเลือกทำในขั้นที่ 2 ได้ n_2 วิธี และในแต่ละวิธีเลือกทำในขั้นที่ 1 และขั้นที่ 2 มีวิธีเลือกทำในขั้นที่ 3 ได้ n_3 วิธี.......

แต่ละวิธีที่เลือกทำในขั้นที่ 1, 2, 3 ..., k − 1 มีวิธีเลือกทำในขั้นที่ k ได้ n_k วิธี

$\therefore$ จำนวนวิธีทั้งหมดที่เลือกทำทั้ง k ขั้นตอนคือ $n_1 \times n_2 \times n_3 \times \ldots . n_k$ วิธี

ตัวอย่าง จงหาจำนวนวิธีที่จะจัดให้นักเรียน 3 คน นั่งบนเก้าอี้ 4 ตัว วางเรียงกันเป็นแนวตรง

วิธีทำ นักเรียนคนที่ 1 มีวิธีเลือกนั่งได้ 4 วิธี

นักเรียนคนที่ 2 มีวิธีเลือกนั่งได้ 3 วิธี

นักเรียนคนที่ 3 มีวิธีเลือกนั่งเก้าอี้ได้ 2 วิธี

$\therefore$ จำนวนวิธีที่จัดนักเรียนนั่งเก้าอี้ได้ $4 \times 3 \times 2 = 24$ วิธี

3. การทำงานอย่างหนึ่งมีทางเลือก k อย่าง แต่ละอย่างมีวิธีเลือกทำได้ n_1, n_2, n_3, ... n_k วิธีตามลำดับโดยแต่ละขั้นจะเลือกทำพร้อม ๆ กันไม่ได้ จำนวนวิธีที่จะเลือกทำงานทั้งหมดได้ $n_1 + n_2 + n_3 + \ldots n_k$ วิธี

ตัวอย่าง จงหาจำนวนวิธีที่จะสร้างจำนวนเลขคู่ 3 หลักจากเซตของตัวเลข 0, 1, 2, 3, 4, 5, โดยแต่ละหลักใช้เลขซ้ำกันไม่ได้

วิธีทำ การสร้างเลข 3 หลัก แยกออกเป็น 2 ทาง แต่ละทางทำพร้อมกันไม่ได้

ทางแรก เมื่อหลักร้อยเป็นเลขคู่ มีตัวเลข 2 ตัว คือ 2 และ 4 จะเลือกได้ 2 วิธี

หลักหน่วยต้องเป็นเลขคู่เสมอคือ 0, 2, 4, แต่ 2 กับ 4 นำไปใช้หลักร้อยแล้ว 1

∴ หลักหน่วยก็จะเหลือ 2 วิธี

หลักสิบก็จะเลือกได้ 4 วิธี

หลักร้อยจะเลือกได้ 2 วิธี

∴ วิธีที่จะสร้างเลขคู่ 3 หลัก ได้ $2 \times 4 \times 2 = 16$ วิธี

ทางที่สอง เมื่อหลักร้อยเป็นเลขคี่

∴ หลักร้อย จะมีวิธีเลือกได้ 3 วิธี (มีเลข 3 ตัว คือ 1, 3, 5)

หลักหน่วย จะเลือกได้ 3 วิธี

หลักสิบ จะเลือกได้ 4 วิธี

∴ จะสร้างเลขคู่สามหลักได้ $3 \times 3 \times 4 = 36$ วิธี

∴ จะสร้างเลขจำนวนคู่ 3 หลัก ได้ทั้งหมด $16 + 36 = 52$ วิธี

4. แฟคตอเรียล (Factorial)

แฟคตอเรียล n หมายถึง ผลคูณของจำนวนเต็มตั้งแต่ 1 ถึง n เขียนแทนด้วย n! (n ไม่เป็นจำนวนเต็มลบ)

$$n! = n(n-1)(n-2) \ldots 2.1$$

5. การเรียงสับเปลี่ยนที่ เป็นการจัดสิ่งของโดยถือลำดับเป็นสำคัญ

5.1 ถ้ามีสิ่งของ n สิ่งที่แตกต่างกันนำมาจัดลำดับทั้ง n สิ่งจัดได้ n! วิธี

5.2 ถ้ามีสิ่งของ n สิ่งที่แตกต่างกันนำมาจัดลำดับครั้งละ r สิ่งจะจัดได้ $\frac{n!}{(n-r)!}$ วิธี

5.3 การจัดลำดับของ n สิ่งแตกต่างกันนำมาจัดลำดับครั้งละ r สิ่งเขียนแทนด้วย $P_{n,r}$ หรือ $_nP_r$

$$\therefore P_{n,r} = \frac{n!}{(n-r)!}$$

$$P_{n,n} = n!$$

r และ n เป็นจำนวนเต็มและ $0 \leq r \leq n$

6. การจัดลำดับของ n สิ่งที่ไม่แตกต่างกันทั้งหมด ถ้ามีสิ่งของ n สิ่งแบ่งออกเป็น k กลุ่ม

กลุ่มที่ 1 มี n_1 สิ่งซึ่งเหมือนกัน

กลุ่มที่ 2 มี n_2 สิ่งที่เหมือนกัน

กลุ่มที่ 3 มี n_3 สิ่งที่เหมือนกัน

.

.

.

กลุ่มที่ k มี n_k สิ่งที่เหมือนกัน

โดยที่ $n_1 + n_2 + n_3 + \ldots + n_k = n$ แล้ว

$\therefore$ จำนวนวิธีที่จัดของทั้ง n สิ่งจะจัดได้ $\dfrac{n!}{n_1!.n_2!.n_3!....n_k!}$ วิธี

7. วิธีเรียงสับเปลี่ยนเป็นวงกลม

จำนวนวิธีที่จัดสิ่งของซึ่งแตกต่างกันเป็นรูปวงกลม n สิ่งจะมีวิธีจัดเรียงสับเปลี่ยนเป็นวงกลมได้ $(n-1)!$ วิธี

วิธีจัดหมู่ (Combination)

วิธีจัดหมู่เป็นการจัดสิ่งของโดยไม่ถือลำดับเป็นสำคัญ

1. การจัดหมู่สิ่งของ n สิ่งที่แตกต่างกันโดยเลือกครั้งละ r สิ่งจัดได้ $\dfrac{n!}{r!(n-r)!}$ วิธี
2. การจัดหมู่ของ n สิ่งครั้งละ r สิ่งเขียนแทนด้วย $C_{n.r}$ หรือ $\genfrac{}{}{0pt}{}{n}{r}$ หรือ n_{C_r}

$$\therefore \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$ เมื่อ r และ n เป็นจำนวนเต็มและ $0 \leq r \leq n$

3. ถ้ามีสิ่งของ n สิ่งที่แตกต่างกัน k ชนิด แต่ละชนิดมีจำนวน $n_1, n_2, n_3, \ldots n_k$, โดยที่ $n_1+n_2+n_3+\ldots+n_k = n$ เลือกมาจัดหมู่ชนิดละ $r_1, r_2, r_3, \ldots r_k$, สิ่งตามลำดับ $(0 \leqslant r_i \leqslant n_i)$

จำนวนวิธีจัดหมู่ทั้งหมดจะมีวิธีจัดได้ $\binom{n_1}{r_1} \cdot \binom{n_2}{r_2}\binom{n_3}{r_3} \ldots \binom{n_k}{r_k}$ วิธี

4. การจัดลำดับสิ่งของเป็นวงกลม โดยถ้ามีของ n สิ่งที่แตกต่างกันนำมาจัดลำดับเป็นวงกลมทั้ง n สิ่ง

จำนวนวิธีที่จัดลำดับเป็นวงกลมคือ $(n-1)!$ วิธี

จำนวนวิธีที่จัดลำดับเป็นวงกลมใน 3 มิติ คือ $\frac{(n-1)}{2}!$ วิธี

ถ้ามีสิ่งของ k ประเภทจำนวน n สิ่ง แต่ละประเภทซึ่งเหมือนกันมีอยู่ $n_1, n_2, n, \ldots, n_k$ สิ่ง และ $n_1 + n_2 + n_3 + \ldots + n_k = n$ แล้ว

จำนวนวิธีที่นำมาจัดลำดับเป็นวงกลมทั้ง n สิ่ง คือ $\frac{(n-1)!}{n_1! \cdot n_2! \cdot n_3! \ldots n_k!}$

5. การหาจำนวนของการเลือกสิ่งของ n สิ่งที่แตกต่างกันโดยเลือกครั้งละ 1, 2, 3, ..., n สิ่ง ตามลำดับ

จำนวนวิธีที่เลือก $\binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{n}$ วิธี

หรือจำนวนวิธีที่เลือกได้ $2^n - 1$ วิธี

6. การหาจำนวนการเลือกสิ่งของ k ชนิด แต่ละชนิดเหมือนกัน $n_1, n_2, n_3, \ldots, n_k$ สิ่งโดยจะเลือกครั้งละจำนวนกี่สิ่งก็ได้

จำนวนวิธีที่เลือกได้ $(n_1 + 1)(n_2 + 1)(n_3 + 1)\ldots(n_k + 1) - 1$ วิธี

7. การจัดลำดับของ n สิ่ง ที่เหมือนกันออกเป็น r กลุ่ม ถ้าแต่ละกลุ่มมีได้ตั้งแต่ 0, 1, 2, ..., n สิ่งจำนวนวิธีที่จัดได้

จะจัดได้ $C_{n+r-1,\, r-1} = \frac{(n+r-1)!}{(r-1)!\, n!}$ วิธี

แต่ถ้าแต่ละกลุ่มมีอย่างน้อย 1 สิ่ง จำนวนวิธีที่จัด

จะจัดได้ $C_{n-1,\, r-1} = \frac{(n-1)!}{(r-1)!\,(n-r)!}$ วิธี

ทฤษฎีความน่าจะเป็นเบื้องต้น

เนื้อหาโดยสรุป

1. การทดลองสุ่ม คือ การทดลองใด ๆ ซึ่งไม่สามารถทำนายผลลัพธ์ที่เกิดขึ้นได้ถูก

2. ผลลัพธ์ คือ ผลที่เกิดขึ้นจากการทดลองสุ่ม

3. แซมเปิลสเปซ (Sample Space) คือ เซตของผลลัพธ์ซึ่งอาจเป็นไปได้ทั้งหมดจากการทดลองสุ่ม

4. เหตุการณ์ (Event) คือ เซตของผลลัพธ์จากการทดลองสุ่มและเป็นสับเซตของแซมเปิลสเปซ

5. ถ้าในแซมเปิลสเปซมีสมาชิก N ตัว สมาชิกแต่ละตัวมีโอกาสเกิดเท่ากัน และ E เป็นเหตุการณ์ซึ่งประกอบด้วยสมาชิก n ตัว ความน่าจะเป็นของเหตุการณ์ E คือ จำนวนสัมพัทธ์ $\frac{n}{N}$

ความน่าจะเป็นของเหตุการณ์ คือ อัตราส่วนระหว่างจำนวนสมาชิกของเหตุการณ์ที่สนใจกับจำนวนสมาชิกของแซมเปิลสเปซที่มีโอกาสเกิดขึ้นเท่า ๆ กัน

ถ้า E เป็นเหตุการณ์

S เป็นแซมเปิลสเปซ

$\therefore P(E) = \frac{n(E)}{n(S)}$ คือ $\frac{n}{N}$

ทฤษฎีความน่าจะเป็น

1. $0 \leqslant P(E) \leqslant 1$

2. $P(\phi) = 0$

3. $P(S) = 1$

4. $E_1 \subset E_2 \subset S \qquad P(E_1) \leqslant P(E_2)$

5. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

6. $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$

7. $P(E) + P(E') = 1$

8. $P(E_1 - E_2) = P(E_1) - P(E_1 \cap E_2)$

ความน่าจะเป็นแบบเงื่อนไข

ให้ A และ B เป็นเหตุการณ์โดยที่ $P(B) > 0$ และ $P(A/B)$ แทนความน่าจะเป็นของ A เมื่อกำหนดว่าเหตุการณ์ B เกิดขึ้นแล้วจะได้

$$p(A/B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0$$

และถ้า $P(B) = 0$ ให้ $P(A/B) = 0$

ถ้า A และ B เป็นเหตุการณ์ที่อิสระจากกันก็ต่อเมื่อ $P(A/B) = P(A)$ นั่นคือ ถ้า A และ B เป็นเหตุการณ์อิสระจากกัน $P(A \cap B) = P(A) . P(B)$

ที่กล่าวมาแล้วเป็นเนื้อหาโดยสรุปซึ่งในหลักสูตรระดับมัธยมศึกษาตอนต้นและมัธยมศึกษาตอนปลายกำหนดให้เรียน ตัวอย่าง และรายละเอียด นักศึกษาสามารถอ่านเพิ่มเติมได้จากหนังสือแบบเรียนวิชาคณิตศาสตร์ของกระทรวงศึกษาธิการร่วมกับสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี

แบบฝึกหัดท้ายบท

1. จงสรุปขอบข่ายเนื้อหาวิชาสถิติที่กำหนดให้เรียนในหลักสูตรระดับมัธยมศึกษามาเป็นข้อ ๆ พอสังเขป

2. จงบอก ความหมาย ชนิดของข้อมูล ตลอดจนวิธีการเก็บรวบรวมข้อมูล

3. ให้สุ่มตัวอย่างอายุของนักศึกษาคณะศึกษาศาสตร์มา 40 คน แล้วคำนวณหาค่ากลางของข้อมูล และวัดการกระจายของข้อมูลนั้น โดยพิจารณาใช้การวัดค่ากลางของข้อมูล และการวัดการกระจายของข้อมูลที่เหมาะสมกับข้อมูลที่รวบรวมได้

4. จงสรุปเนื้อหาโดยย่อ ในเรื่องต่อไปนี้

- การเรียงสับเปลี่ยน (Permutation)
- การจัดหมู่ (Combination)
- ทฤษฎีความน่าจะเป็นเบื้องต้นในระดับมัธยมศึกษา