

ใบความรู้ที่ 3

เรื่อง การวัดค่ากลางของข้อมูล (ฐานนิยม)

ฐานนิยม (Mode) คือ ค่าของข้อมูลที่มีความถี่สูงสุด ใช้เป็นค่ากลางของข้อมูลอีกชนิดหนึ่ง นอกเหนือจากค่าเฉลี่ยเลขคณิต และมัธยฐานที่ได้กล่าวมาแล้ว ส่วนมากฐานนิยมจะใช้กับข้อมูลเชิงคุณภาพมากกว่าข้อมูลเชิงปริมาณ

ฐานนิยม เหมาะที่จะนำมาใช้เป็นค่ากลางของข้อมูล เมื่อข้อมูลนั้นๆ เป็นค่ามาตรฐาน เช่น ขนาดรองเท้า ขนาดยางรถยนต์ ฯลฯ หรือข้อมูลที่แจกแจงความถี่แล้วตามกลุ่มหรือช่วงต่างๆ โดยเฉพาะเมื่อมีข้อมูลที่มีค่าสูงหรือต่ำผิดปกติรวมอยู่ด้วย

การหาฐานนิยมของข้อมูลที่ไม่ได้แจกแจงความถี่

หาได้จากการศึกษาว่าข้อมูลค่าใดจากข้อมูลที่มีอยู่ทั้งหมด มีความถี่สูงสุดหรือปรากฏบ่อยครั้งที่สุด ข้อมูลนั้นจะเป็นฐานนิยมของข้อมูลชุดนั้น

ตัวอย่างที่ 1 จงหาฐานนิยมของอายุนักเรียนที่เข้ามาเข้าค่ายคณิตศาสตร์ จำนวน 15 คน ดังนี้ 5, 8, 7, 6, 7, 8, 12, 11, 10, 11, 8, 6, 8, 7 และ 8 ปี

วิธีทำ ฐานนิยมของอายุนักเรียนที่เข้ามาเข้าค่ายคณิตศาสตร์ทั้ง 15 คน คือ 8 ปี เพราะนักเรียนที่เข้ามาเข้าค่ายคณิตศาสตร์มีอายุ 8 ปี มากที่สุด คือ 5 คน กล่าวคือ นักเรียนที่เข้ามาเข้าค่ายคณิตศาสตร์ อายุ 8 ปี มีจำนวนมากที่สุด

การหาฐานนิยมโดยวิธีดังกล่าว ข้อมูลบางชุดอาจจะไม่มีฐานนิยมเลยก็ได้ หรืออาจจะมีฐานนิยมเกินกว่าหนึ่งค่าก็ได้ เช่น

- ข้อมูลที่ประกอบด้วย 5, 8, 9, 10, 12, 18, 16, 20 จะไม่มีฐานนิยมเลย เพราะข้อมูลแต่ละค่ามีความถี่เท่ากันหมด

- ข้อมูลที่ประกอบด้วย 13, 16, 20, 25, 20, 26, 25 มีฐานนิยมสองค่า คือ 20 และ 25 เนื่องจากทั้งสองค่านี้มีความถี่สูงสุดเท่ากัน คือ 2 ในกรณีที่ข้อมูลชุดใดมีฐานนิยมมากกว่า 2 ค่า อาจจะถือได้ว่า ข้อมูลชุดนั้นไม่มีฐานนิยมได้

การหาฐานนิยมของข้อมูลที่แจกแจงความถี่แล้ว

ถ้าเขียนเส้นโค้งของความถี่ของข้อมูลที่แจกแจงความถี่แล้ว ฐานนิยมคือ ค่าในแกน x ที่อยู่ตรงกับจุดสูงสุดบนเส้นโค้งของความถี่ แต่ถ้าเส้นโค้งของความถี่มีจุดสูงสุดสองจุด ข้อมูลชุดนั้นจะมีฐานนิยม 2 ค่า

สำหรับการคำนวณหาฐานนิยมของข้อมูลที่แจกแจงความถี่ไว้แล้ว ทำได้หลายวิธี วิธีหนึ่งคือ หาจุดกึ่งกลางของอันตรภาคชั้นที่มีฐานนิยมอยู่ ค่าที่หาได้จะเป็นค่าของฐานนิยมโดยประมาณ ส่วนการหาว่าฐานนิยมอยู่ในอันตรภาคชั้นใดนั้น จะต้องพิจารณาด้วยว่า อันตรภาคแต่ละชั้นมีความกว้างเท่ากันหรือไม่ ในกรณีที่ความกว้างของอันตรภาคชั้นทุกชั้นเท่ากัน อันตรภาคชั้นที่มีฐานนิยม คือ อันตรภาคชั้นที่มีความถี่สูงสุด ส่วนกรณีที่ความกว้างของอันตรภาคชั้นไม่เท่ากันทุกชั้น ให้หารความถี่ด้วยความกว้างของแต่ละอันตรภาคชั้น อันตรภาคชั้นที่ผลหารมีค่ามากที่สุด จะเป็นอันตรภาคชั้นที่มีฐานนิยมอยู่

จะหาได้จากสูตร
$$\text{Mod} = L + \left(\frac{d_1}{d_1 + d_2} \right) I$$

เมื่อ Mod แทน ฐานนิยม

L แทน ขอบล่างของอันตรภาคชั้น

d_1 แทน ผลต่างระหว่างความถี่ของอันตรภาคชั้นที่มีความถี่มากที่สุดกับความถี่ของอันตรภาคชั้นที่มีค่าต่ำกว่าและอยู่ติดกัน

d_2 แทน ผลต่างระหว่างความถี่ของอันตรภาคชั้นที่มีความถี่มากที่สุดกับความถี่ของอันตรภาคชั้นที่มีค่าสูงกว่าและอยู่ติดกัน

I แทน ความถี่ของชั้นที่มีความถี่มากที่สุด

ตัวอย่างที่ 2 จงหาฐานนิยม จากตารางแจกแจงความถี่ต่อไปนี้

น้ำหนัก(กิโลกรัม)	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64
จำนวน	8	12	16	2	8

วิธีทำ

น้ำหนัก (กิโลกรัม)	จำนวน (f)
40 – 44	8
45 – 49	12
50 – 54	16
55 – 59	2
60 – 64	8
65 – 69	4

ฐานนิยมอยู่ในอันดับที่ 50 – 54 เพราะเป็นอันดับที่มีความถี่มากที่สุด

$$\therefore \text{จากสูตร } \text{Mod} = L + \left(\frac{d_1}{d_1 + d_2} \right) I$$

$$\text{จะได้ } L = 49.5, \quad d_1 = 16 - 12 = 4,$$

$$d_2 = 16 - 2 = 14 \quad \text{และ} \quad I = 5$$

$$\text{แทนค่า } \text{Mod} = 49.5 + \left(\frac{4}{4 + 14} \right) \times 5 = 49.5 + 1.11 = 50.61$$

สมบัติของฐานนิยม

1. ฐานนิยมสามารถหาได้จากเส้นโค้งของความถี่ และฮิสโทแกรม
2. ในข้อมูลแต่ละชุดอาจจะมีฐานนิยม หรือไม่มีก็ได้ ถ้ามีอาจจะมีเพียงค่าเดียวหรือหลายค่าก็ได้
3. ถ้าให้ $x_1, x_2, x_3, \dots, x_n$ เป็นข้อมูลชุดหนึ่งที่มีฐานนิยมเท่ากับ a ถ้า k เป็นค่าคงตัว จะได้ว่า $x_1+k, x_2+k, x_3+k, \dots, x_n+k$ เป็นฐานข้อมูลที่มีฐานนิยมเท่ากับ $a+k$
4. ถ้าให้ $x_1, x_2, x_3, \dots, x_n$ เป็นข้อมูลชุดหนึ่งที่มีฐานนิยมเท่ากับ a ถ้า k เป็นค่าคงตัว ซึ่ง $k \neq 0$ จะได้ว่า $kx_1, kx_2, kx_3, \dots, kx_n$ เป็นฐานข้อมูลที่มีฐานนิยมเท่ากับ ka

ข้อสังเกตและหลักเกณฑ์ที่สำคัญในการใช้ค่ากลางชนิดต่างๆ

1. ค่าเฉลี่ยเลขคณิตเป็นค่ากลางที่ได้จากงานทุกๆ ค่าของข้อมูลมาเฉลี่ยมัธยฐานเป็นค่ากลางที่ได้จากตำแหน่งของข้อมูล และฐานนิยมเป็นค่ากลางที่ได้จากข้อมูลที่มีความถี่มากที่สุด
2. ถ้าในจำนวนทั้งหมดมีข้อมูลบางค่าที่มีค่าสูงหรือต่ำกว่าข้อมูลอื่นๆ มากจะมีผลกระทบต่อค่าเฉลี่ยเลขคณิตแต่จะไม่มีผลกระทบต่อมัธยฐานหรือฐานนิยม ดังนั้น กรณีนี้ ควรใช้มัธยฐาน
3. มัธยฐานและฐานนิยมใช้เมื่อต้องการทราบค่ากลางของข้อมูลทั้งหมด โดยประมาณและรวดเร็ว เนื่องจากการหามัธยฐานและฐานนิยมบางวิธีไม่จำเป็นต้องมีการคำนวณ ซึ่งอาจใช้เวลามาก
4. ถ้าการแจกแจงความถี่ของข้อมูล ประกอบด้วยอันดับชั้นช่วงเปิด ซึ่งอาจเป็นขั้นต่ำสุดหรือขั้นสูงสุด ชั้นใดชั้นหนึ่ง การหาค่ากลางโดยใช้ค่าเฉลี่ยเลขคณิตไม่สามารถหาได้ แต่สามารถหามัธยฐานและฐานนิยมได้
5. การแจกแจงความถี่ที่มีความกว้างของแต่ละอันดับชั้นไม่เท่ากัน จะทำให้ค่าเฉลี่ยเลขคณิตหรือฐานนิยมคลาดเคลื่อน แต่จะไม่กระทบต่อมัธยฐาน
6. ข้อมูลประเภทคุณภาพจะหาฐานนิยมได้ แต่ไม่สามารถหาค่าเฉลี่ยเลขคณิตหรือมัธยฐานได้
7. ในกรณีที่น่าข้อมูลมาเรียงลำดับได้ ควรหาค่ากลาง คือ มัธยฐานก่อน และถ้าเป็นข้อมูลเชิงปริมาณที่มีค่าต่อเนื่องด้วย ควรใช้ค่าเฉลี่ยเลขคณิตแทนมัธยฐานจะเหมาะสมกว่า
8. กรณีที่ข้อมูลมีจำนวนน้อย ฐานนิยมอาจจะมีค่าแตกต่างกันมาก ระหว่างข้อมูลชุดหนึ่งกับข้อมูลอีกชุดหนึ่งที่มีจำนวนเท่ากัน จึงไม่ควรใช้ฐานนิยมในกรณีเช่นนี้

ข้อดีและข้อเสียของค่าเฉลี่ยเลขคณิต มัชฐานและฐานนิยม มีดังนี้

ชนิดของค่ากลางของข้อมูล	ข้อดี	ข้อเสีย
ค่าเฉลี่ยเลขคณิต (\bar{x})	<ol style="list-style-type: none"> 1. หาง่ายอาจจะใช้เครื่องคิดเลขช่วยในการคำนวณได้ 2. ใช้ข้อมูลทุกตัวในการคำนวณ 3. นิยมใช้เป็นค่ากลางของข้อมูล 	<ol style="list-style-type: none"> 1. ใช้กับข้อมูลเชิงปริมาณเท่านั้น 2. ค่าที่คำนวณได้ไม่จำเป็นต้องเป็นค่าของข้อมูลตัวใดตัวหนึ่งเสมอไป 3. ถ้าข้อมูลมีค่าแตกต่างกันมากจะมีผลกระทบต่อค่าเฉลี่ยเลขคณิต
มัชฐาน	<ol style="list-style-type: none"> 1. หาง่ายเพียงแต่เรียงข้อมูลจากน้อยไปหามากหรือจากมากไปหาน้อย 2. ถ้าข้อมูลเป็นจำนวนคี่ มัชฐานจะเป็นค่าในข้อมูล 	<ol style="list-style-type: none"> 1. ใช้ได้กับข้อมูลเชิงปริมาณเท่านั้น 2. ถ้าข้อมูลมีจำนวนมากการจัดเรียงข้อมูลจะเสียเวลามาก 3. ถ้าข้อมูลเป็นจำนวนคู่มัชฐานจะไม่ใช่ค่าที่แท้จริงของข้อมูล
ฐานนิยม	<ol style="list-style-type: none"> 1. ใช้ได้กับข้อมูลเชิงปริมาณและเชิงคุณภาพ 2. หาได้ง่ายโดยนับจำนวนข้อมูลที่ปรากฏมากที่สุด 3. สามารถหาได้จากตารางแจกแจงความถี่ แผนภูมิแบบต่างๆ 	<ol style="list-style-type: none"> 1. ถ้าข้อมูลมีจำนวนน้อยค่าที่ได้จะไม่มีคามหมาย 2. ฐานนิยมอาจมีมากกว่าหนึ่งค่า 3. ข้อมูลบางชุดอาจจะไม่มีฐานนิยม