เฉลย Assignment 1
MAC1304 ความน่าเป็นและสถิติ

หัวข้อ สถิติพื้นฐานและค่ากลางของข้อมูล **สัปดาห์ที่** 1 **คะแนนเต็ม** 10 คะแนน

ผู้สอน ผศ.ดร.ธนัชยศ จำปาหวาย สาขาวิชาคณิตศาสตร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา

---

1. คะแนนวิชาแคลคูลัส 1 ของนักศึกษา 50 คน ซึ่งมีคะแนนเต็ม 100 คะแนน เป็นดังนี้

30	35	36	39	39	41	42	42	43	44	45	46	46	48	49	50	50
50	50	50	54	55	58	58	59	59	61	62	62	62	63	66	68	69
69	71	72	72	73	74	76	78	80	82	82	84	88	90	92	98	

จากข้อมูลข้างต้นจงสร้างตารางแจกแจงความถี่พร้อมทั้งแสดงข้อมูลด้วยฮิสโทแกรม รูปหลายเหลี่ยมความถี่ และเส้นโค้งโอจีฟ

แนวคำตอบ เนื่องจาก $N = 50$ จากสูตรสเตอเกส $k = 1 + 3.3\log 50 = 6.61 \approx 7$ โดยที่ $X_{\text{max}} = 98$ และ $X_{\text{min}} = 30$ นั่นคือ $R = 98 - 30 = 68$ และความกว้างชั้นหาได้จาก

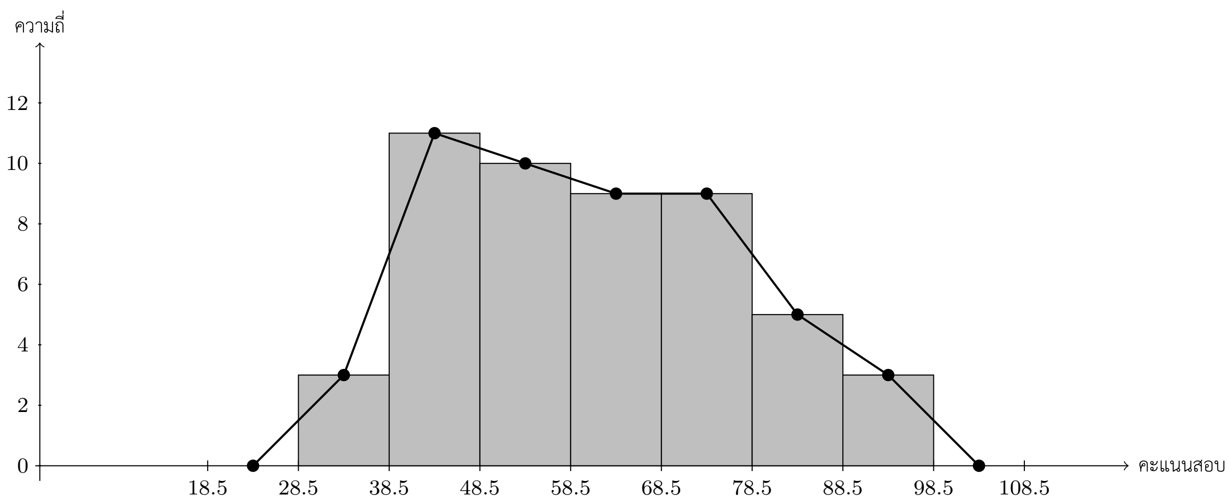
$$I = \frac{R}{k} = \frac{68}{7} = 9.71 \approx 10$$

ดังนั้นเลือก $k = 7$ และ $I = 10$ จะได้ว่า

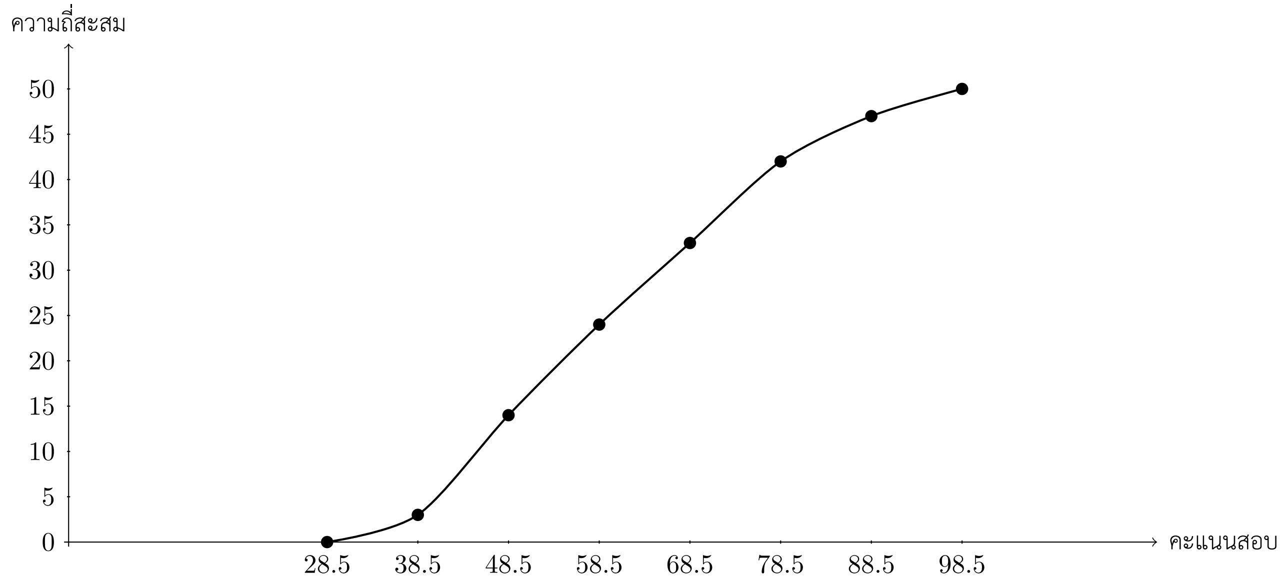
$$\begin{aligned} \text{ขีดจำกัดล่างของชั้นแรก} &= X_{\min} - \frac{Ik - R}{2} = 30 - \frac{10(7) - 68}{2} = 29 \\ \text{ขีดจำกัดบนของชั้นแรก} &= 29 + 10 - 1 = 38 \end{aligned}$$

คะแนนวิชาแคลคูลัส 1	รอยขีด	จำนวนคน	จำนวนคนสะสม
$29 - 38$	|||	3	3
$39 - 48$	||||/ ||||/ |	11	14
$49 - 58$	||||/ ||||/	10	24
$59 - 68$	||||/ ||||	9	33
$69 - 78$	||||/ ||||	9	42
$79 - 88$	||||/	5	47
$89 - 98$	|||	3	50

แสดงฮิสโทรแกรมได้ดังนี้



และแสดงเส้นโค้งโอจีฟได้ดังนี้



2. สำหรับข้อมูลที่จัดกลุ่มในรูปตารางแจกแจงความถี่ที่มีขนาด N ฐานนิยมจะเป็นค่าที่อยู่ในชั้นที่มีความถี่สูงที่สุด จงพิสูจน์ว่า

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

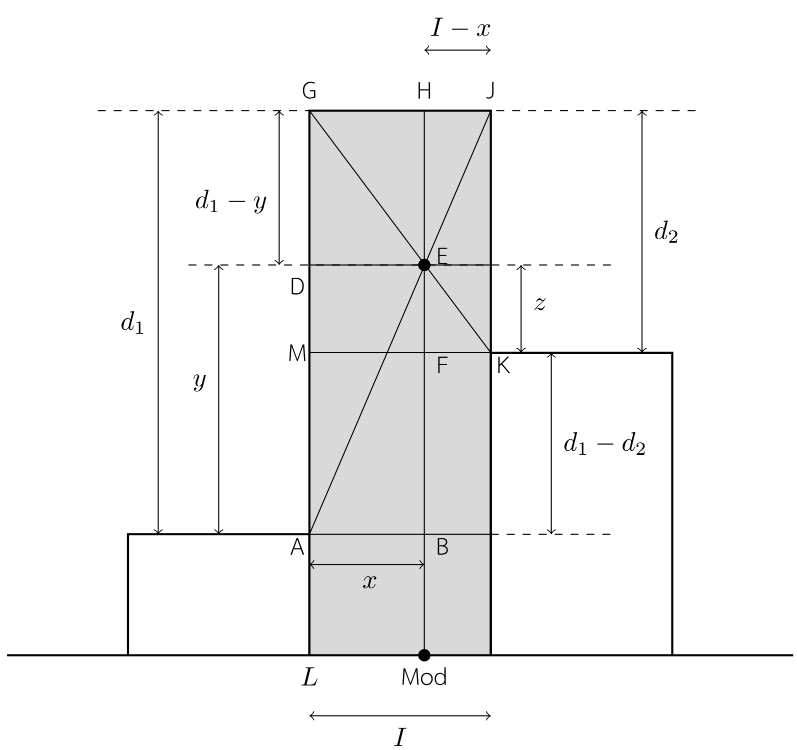
L คือขอบล่างของชั้นฐานนิยม

I คือความกว้างของอันตรภาคของชั้นฐานนิยม

d_1 คือผลต่างของความถี่ชั้นฐานนิยมกับชั้นก่อนชั้นฐานนิยม

d_2 คือผลต่างของความถี่ชั้นฐานนิยมกับชั้นหลังชั้นฐานนิยม

บทพิสูจน์. พิจารณาฮิสโทแกรมชั้นที่มีความถี่สูงสุด ดังรูป



จากรูปที่ **??** จะได้ว่า Mod $= L + x$

เนื่องจาก $\Delta ABE \sim \Delta JHE$ ดังนั้น $\dfrac{y}{x} = \dfrac{d_1 - y}{I - x}$ หรือ

$$\begin{aligned} y(I - x) &= x(d_1 - y) = xd_1 - xy \\ yI - yx &= xd_1 - xy \\ y &= \frac{xd_1}{I} \end{aligned}$$

เนื่องจาก $\Delta KFE \sim \Delta KMG$ ดังนั้น $\dfrac{z}{I - x} = \dfrac{d_2}{I}$ หรือ $z = \dfrac{d_2(I - x)}{I}$

เนื่องจาก $y = z + (d_1 - d_2)$ ฉะนั้น

$$\begin{aligned} \frac{xd_1}{I} &= \frac{d_2(I - x)}{I} + d_1 - d_2 \\ d_1 x &= d_2 I - d_2 x + d_1 I - d_2 I \\ x &= I\left(\frac{d_1}{d_1 + d_2}\right) \end{aligned}$$

ดังนั้น Mod $= L + I\left(\dfrac{d_1}{d_1 + d_2}\right)$ □

3. สำหรับข้อมูลเชิงปริมาณชุดหนึ่ง เมื่อเพิ่มอีกหนึ่งค่าซึ่งเท่ากับค่าเฉลี่ยเลขคณิตของกลุ่มนี้ จงแสดงว่าค่าเฉลี่ยเลขคณิตใหม่จะไม่เปลี่ยนแปลง

บทพิสูจน์. ให้ประชากรขนาด n ประกอบด้วย $x_1, x_2, ..., x_n$ ซึ่งมีค่าเฉลี่ยเลขคณิตเท่ากับ μ นั่นคือ

$$\sum_{i=1}^{n} x_i = n\mu$$

สมมติค่าที่เพิ่มคือ $y = \mu$ ค่าเฉลี่ยเลขคณิตของ $x_1, x_2, ..., x_n, y$ เท่ากับ

$$\frac{\sum_{i=1}^{n} x_i + y}{n + 1} = \frac{n\mu + \mu}{n + 1} = \frac{\mu(n + 1)}{n + 1} = \mu$$

ดังนั้นค่าเฉลี่ยเลขคณิตใหม่จะไม่เปลี่ยนแปลง □

4. จงหาค่าเฉลี่ยเลขคณิตของข้อมูล $x_1, x_2, x_3, ..., x_{100}$ เมื่อ

$$x_n = \begin{cases} n^2 & \text{เมื่อ } n \text{ เป็นจำนวนคี่} \\ 2n & \text{เมื่อ } n \text{ เป็นจำนวนคู่} \end{cases}$$

แนวคำตอบ ค่าเฉลี่ยเลขคณิตคือ

$$\begin{aligned}
\mu = \frac{\sum_{i=1}^{n}}{100} &= \frac{1^2 + 2(2) + 3^2 + 2(4) + 5^2 + 2(6) + \cdots + 99^2 + 2(100)}{100} \\
&= \frac{(1^2 + 3^2 + 5^2 + \cdots + 99^2) + 4(1 + 2 + 3 + \cdots + 50)}{100} \\
&= \frac{\sum_{i=1}^{50}(2i-1)^2 + 4\sum_{i=1}^{50} i}{100} \\
&= \frac{\sum_{i=1}^{50}(4i^2 - 4i + 1) + \sum_{i=1}^{50} 4i}{100} \\
&= \frac{\sum_{i=1}^{50}(4i^2 - 4i + 1 + 4i)}{100} \\
&= \frac{\sum_{i=1}^{50}(4i^2 + 1)}{100} \\
&= \frac{4\sum_{i=1}^{50} i^2 + \sum_{i=1}^{50} 1}{100} \\
&= \frac{4 \cdot \frac{50(51)(101)}{6} + 1(50)}{100} \\
&= \frac{171750}{100} = 1717.5 \quad \#
\end{aligned}$$

5. ข้อมูลชุดหนึ่งมี 99 จำนวนเรียงจากน้อยไปมากได้เป็น $x_1,\ x_2,\ x_3,\ ...,\ x_{99}$ ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้มีค่าเท่ากับมัธฐาน จงแสดงว่า $\displaystyle\sum_{i=1}^{49}|x_{50}-x_i| = \sum_{i=51}^{99}|x_{50}-x_i|$

แนวคำตอบ เนื่องจากข้อมูล $x_1,\ x_2,\ x_3,\ ...,\ x_{99}$ เรียงน้อยไปมาก นั่นคือ

$$x_1 \le x_2 \le x_3 \le \cdots \le x_{99}$$

โดยมี $\mu = \text{Med} = x_{50}$ จะเห็นว่า

$$\begin{array}{llll} x_{50}-x_i \ge 0 & \text{เมื่อ } i=1,2,...,49 & \text{นั่นคือ} & |x_{50}-x_i| = x_{50}-x_i \\ x_{50}-x_i \le 0 & \text{เมื่อ } i=51,52,...,99 & \text{นั่นคือ} & |x_{50}-x_i| = -(x_{50}-x_i) \end{array}$$

จะได้ว่า

$$\begin{aligned} \sum_{i=1}^{49}|x_{50}-x_i| - \sum_{i=51}^{99}|x_{50}-x_i| &= \sum_{i=1}^{49}(x_{50}-x_i) + \sum_{i=51}^{99}(x_{50}-x_i) \\ &= \sum_{i=1}^{49}(x_{50}-x_i) + 0 + \sum_{i=51}^{99}(x_{50}-x_i) \\ &= \sum_{i=1}^{49}(x_{50}-x_i) + (x_{50}-x_{50}) + \sum_{i=51}^{99}(x_{50}-x_i) \\ &= \sum_{i=1}^{99}(x_{50}-x_i) \\ &= \sum_{i=1}^{99}(\mu-x_i) = 0 \qquad \text{จากสมบัติของค่าเฉลี่ยเลขคณิต} \end{aligned}$$

สรุปได้ว่า $\displaystyle\sum_{i=1}^{49}|x_{50}-x_i| = \sum_{i=51}^{99}|x_{50}-x_i|$

6. จงหาค่าเฉลี่ยเลขคณิตและฐานนิยมจากตารางนี้ เมื่อค่ามัธยฐานเท่ากับ 11.75

คะแนน	ความถี่ (คน)
$1-\ 5$	5
$6-10$	x
$11-15$	8
$16-20$	7

แนวคำตอบ จะเห็นว่า $\text{Med} = 11.75$ อยู่ชั้นที่ 3 และ

คะแนน	ความถี่ (คน)	จำนวนคนสะสม
$1-\ 5$	5	5
$6-10$	x	$5+x$
$11-15$	8	$13+x$
$16-20$	7	$20+x$

จะได้ $N = 20 + x$, $L = 10.5$, $f_m = 8$, $\sum f_L = x + 5$ และ $I = 5$ ดังนั้น

$$\begin{aligned}
11.75 &= 10.5 + \frac{5}{8}\left(\frac{20+x}{2} - (5+x)\right) \\
1.25 &= \frac{5}{8}\left(\frac{20+x}{2} - 5 - x\right) \\
10 &= \frac{100+5x}{2} - 25 - 5x \\
20 &= 100 + 5x - 50 - 10x \\
5x &= 30 \\
x &= 6
\end{aligned}$$

จะได้ตารางดังนี้

คะแนน	x_i	f_i	$f_i x_i$
$1 - 5$	3	5	15
$6 - 10$	8	6	48
$11 - 15$	13	8	104
$16 - 20$	18	7	126
รวม		26	293

ค่าเฉลี่ยเลขคณิตเท่ากับ

$$\mu = \frac{\sum f_i x_i}{n} = \frac{293}{26} = 11.27 \quad \#$$

ฐานนิยมอยู่ในชั้นที่ 3 โดยมี $L = 10.5$, $d_1 = 8 - 6 = 2$, $d_2 = 8 - 7 = 1$ และ $I = 5$ ดังนั้น

$$\text{Mod} = 10.5 + 5\left(\frac{2}{2+1}\right) = 13.83 \quad \#$$

7. ข้อมูล $x_1, x_2, x_3, ..., x_n$ มีค่าเฉลี่ยเลขคณิตคือ μ จงพิสูจน์ว่า $\sum_{i=1}^{n}(x_i - \mu)^2$ มีค่าน้อยที่สุด

บทพิสูจน์. สำหรับจำนวนจริง x ใด ๆ กำหนดให้

$$f(x) = \sum_{i=1}^{n}(x_i - x)^2$$

เราจะพิสูจน์โดยอาศัยความรู้แคลคูลัสในเรื่องค่าสูงสุดต่ำสุด โดยการทดสอบอนุพันธ์ อันดับสอง พิจารณา

$$0 = f'(x) = \sum_{i=1}^{n} 2(x_i - x)(-1) = -2\sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n} x = -2n\mu + 2nx$$

ฉะนั้น $x = \mu$ เป็นจุดวิกฤต เนื่องจาก

$$f''(x) = 2n > 0 \qquad \because n > 0$$

นั่นคือ $f''(\mu) > 0$ ดังนั้น $f(\mu) = \sum_{i=1}^{n}(x_i - \mu)^2$ เป็นค่าต่ำสุด $\square$

8. จงหาค่าเฉลี่ยเลขคณิต ค่ามัธยฐานและฐานนิยม ของประชากรที่ประกอบด้วย

$$1, 2, 2, 3, 3, 3, 4, 4, 4, 4, ..., \underbrace{100, 100, ..., 100}_{100\text{ตัว}}$$

แนวคำตอบ พิจารณาการนับจำนวนข้อมูลโดย

					จำนวน (ตัว)	ผลบวก		
1					1	1	$=$	1^2
2	2				2	$2+2$	$=$	2^2
3	3	3			3	$3+3+3$	$=$	3^2
4	4	4	4		4	$4+4+4+4$	$=$	4^2
$\vdots$					$\vdots$		$\vdots$	
100	100	100	...	100	100	$100+100+\cdots+100$	$=$	100^2

ดังนั้นค่าเฉลี่ยเลขคณิตคือ

$$\mu = \frac{1^2+2^2+3^2+\cdots+100^2}{1+2+3+\cdots 100} = \frac{\frac{1}{6}(100)(101)(201)}{\frac{1}{2}(100)(101)} = 67 \quad \#$$

จะเห็นว่าจำนวนข้อมูลเท่ากับ $N = 1+2+3+\cdots+100 = \dfrac{1}{2}(100)(101) = 5050$

ค่ามัธยฐานเท่ากับคะแนนในตำแหน่ง

$$\frac{5050+1}{2} = 2525.5$$

หา n ที่ทำให้ $1+2+3+\cdots n$ มีค่าใกล้เคียง 2525.5 นั่นคือ

$$1+2+3+\cdots+70 = \frac{1}{2}(70)(71) = 2485$$
$$1+2+3+\cdots+71 = \frac{1}{2}(71)(72) = 2556$$

พิจารณา

						จำนวน (ตัว)	ตำแหน่งที่สุดท้ายแต่ละแถว		
แถวที่ 1	1					1	1	$=$	1
แถวที่ 2	2	2				2	$1+2$	$=$	3
แถวที่ 3	3	3	3			3	$1+2+3$	$=$	6
แถวที่ 4	4	4	4	4		4	$1+2+3+4$	$=$	10
$\vdots$						$\vdots$		$\vdots$	
แถวที่ 70	70	70	70	...	70	70	$1+2+3+\cdots+70$	$=$	2485
แถวที่ 71	71	71	71	...	71	71	$1+2+3+\cdots+71$	$=$	2556

จะได้ว่าตำแหน่งที่ 2525.5 อยู่แถวที่ 71 ดังนั้นค่ามัธยฐานเท่ากับ 71 $\#$

ฐานนิยมมีค่าเท่ากับ 100 เพราะ 100 ซ้ำมากที่สุดคือ 100 ครั้ง