

# Lecture #2

## สถิติเชิงพรรณนา (ต่อ)

การนำเสนอตาราง, การคำนวณค่ากลาง & ค่ากระจายตัวของข้อมูล

# หัวข้อ

1. การนำเสนอข้อมูลด้วยตาราง
2. การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)
3. การวัดการกระจายตัว (Variability)
4. การหาค่ามาตรฐาน (Standard Value หรือ Z-score)

# การนำเสนอข้อมูลด้วยตาราง

1. ตารางแบบทางเดียว (One-way Table)
2. ตารางแบบสองทาง (Two-way Table)
3. ตารางแบบหลายทาง (Multiple-way Table)
4. ตารางแจกแจงความถี่ (Frequency Distribution Table)
5. ตารางแจกแจงความถี่สัมพัทธ์ (Relative Frequency Distribution Table)

# การนำเสนอข้อมูลด้วยตาราง

ตารางที่ 1: ปริมาณนำเข้าทางนมผง

ปี	ปริมาณ (ตัน)
2541	11,620
2542	14,504
2543	21,199
2544	23,937
2545*	17,441
<b>รวม</b>	<b>88,701</b>

\* ตัวเลขเดือนมกราคม - ตุลาคม 2545

ตารางที่ 2: ต้นทุนการผลิตนมพาสเจอร์ไรซ์  
จำแนกตามวัตถุดิบที่ใช้

ค่าใช้จ่าย (ร้อยละ)	นมสดพาสเจอร์ไรซ์ (ขนาด 225 ซีซี)	
	นํ้านมดิบ	หางนมผง
วัตถุดิบ	80	68
ภาชนะ	9	13
อื่นๆ	11	19
<b>รวม</b>	<b>100</b>	<b>100</b>

ที่มา: กองวิจัยเศรษฐกิจอุตสาหกรรม

ตารางที่ 3: ต้นทุน (ร้อยละ) ของวัตถุดิบในการผลิตนมพาสเจอร์ไรซ์และนมยูเอเอสที

ค่าใช้จ่าย (ร้อยละ)	นมสดพาสเจอร์ไรซ์ (ขนาด 225 ซีซี)		นมยูเอเอสที (ขนาด 225 ซีซี)	
	นํ้านมดิบ	หางนมผง	นํ้านมดิบ	หางนมผง
วัตถุดิบ	80	68	53	40
ภาชนะ	9	13	39	48
อื่นๆ	11	19	8	12
<b>รวม</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

ที่มา: กองวิจัยเศรษฐกิจอุตสาหกรรม

# การนำเสนอข้อมูลด้วยตาราง

ตารางที่ 4: จำนวนคนไข้ที่เข้ามา  
รักษาพยาบาลจำแนกตามเพศ

เพศ	จำนวนคนไข้ (คน)
ชาย	259
หญิง	376
<b>รวม</b>	<b>635</b>

ตารางที่ 5: ความถี่สัมพัทธ์ (ร้อยละ) ของคนไข้จำแนกตามเพศ

เพศ	จำนวนคนไข้ (คน)	ความถี่สัมพัทธ์ (ร้อยละ)
ชาย	259	40.8%
หญิง	376	59.2%
<b>รวม</b>	<b>635</b>	<b>100.0%</b>

# การสร้างตารางแจกแจงความถี่ & Histogram (p.61)

ตัวอย่างที่ 2.4.1.5 ตารางที่ 2.4.7 เป็นข้อมูลเกี่ยวกับค่าวัดมลภาวะของอากาศในเมืองใหญ่ 57 เมืองทั่วโลก ให้สร้างตารางแจกแจงความถี่ของข้อมูลเหล่านี้

ตารางที่ 2.4.7 จุลวัตถุ (ไมโครกรัมต่ออากาศ 1 ลูกบาศก์เมตร) ที่มีอยู่ในอากาศจากเมืองใหญ่ 57 เมือง

ลำดับที่	ค่าที่วัดได้	ลำดับที่	ค่าที่วัดได้	ลำดับที่	ค่าที่วัดได้
1	68	21	36	41	16
2	63	22	42	42	24
3	42	23	28	43	69
4	27	24	31	44	47
5	30	25	28	45	23
6	36	26	25	46	22
7	28	27	45	47	43
8	32	28	12	48	27
9	79	29	57	49	49
10	27	30	51	50	28
11	22	31	12	51	23
12	23	32	32	52	19
13	24	33	49	53	46
14	25	34	38	54	30
15	44	35	42	55	43
16	65	36	27	56	49
17	43	37	31	57	12
18	25	38	50		
19	74	39	38		
20	51	40	21		

$$1. \text{ จำนวนจำนวนชั้น } k = 1 + 3.322 (\log_{10} 57)$$

$$k = 1 + 3.322 (1.7559)$$

$$= 7$$

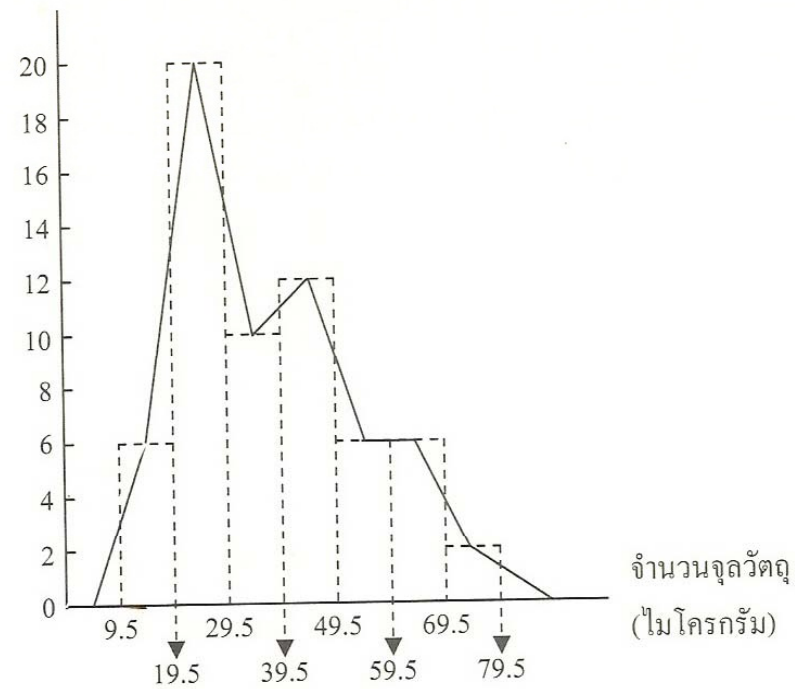
2. จำนวนความกว้างของชั้น โดยสูตร  $w = \text{พิสัย} / \text{จำนวนชั้น}$

$$\text{แทนค่าจะได้ } w = \frac{79 - 12}{7}$$

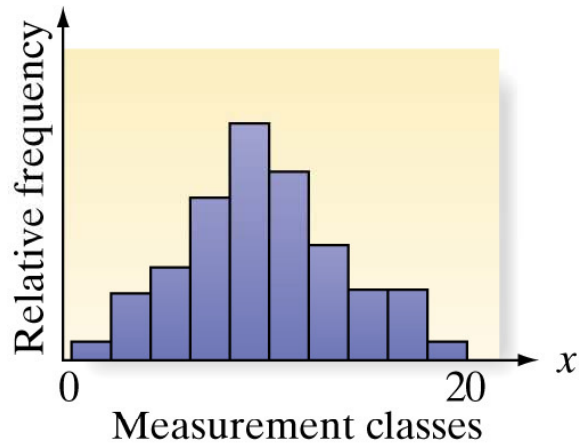
$$= 9.6 \approx 10$$

(1)	(2)	(3)	(4) = (2)/57	(5) = (3)/57
อันดับภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์	ความถี่สัมพัทธ์สะสม
10 – 19	5	5	0.0877	0.0877
20 – 29	19	24	0.3333	0.4210
30 – 39	10	34	0.1754	0.5964
40 – 49	13	47	0.2281	0.8245
50 – 59	4	51	0.0702	0.8847
60 – 69	4	55	0.0702	0.9649
70 – 79	2	57	0.0351	1.000
รวม	57		1.0000	

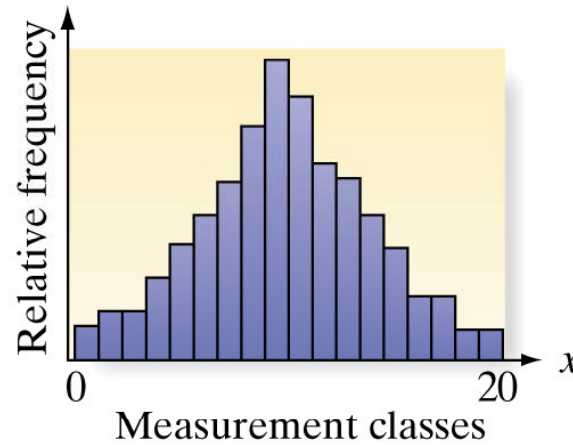
ความถี่สัมพัทธ์



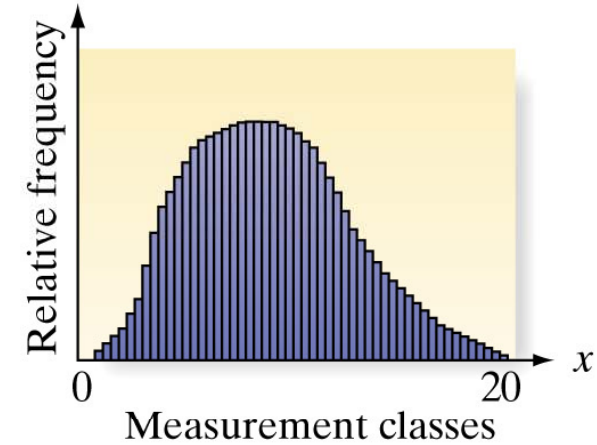
# More on Histogram



a. Small data set



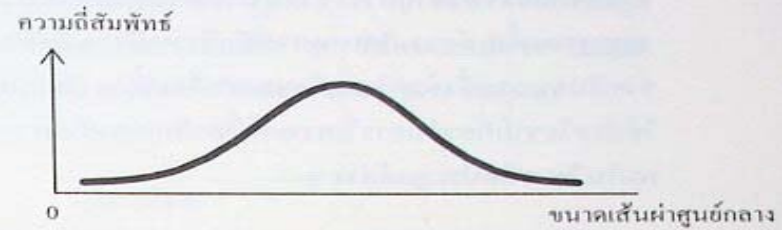
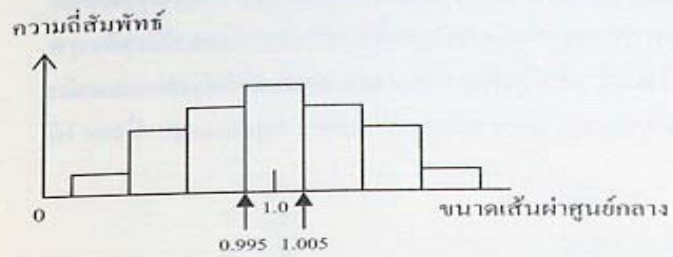
b. Larger data set



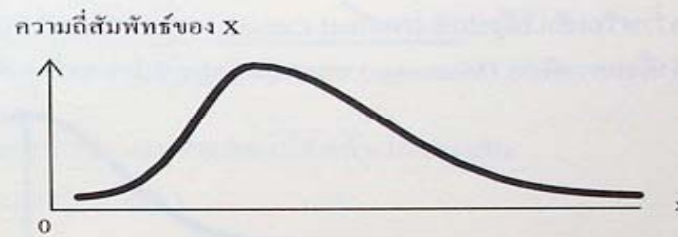
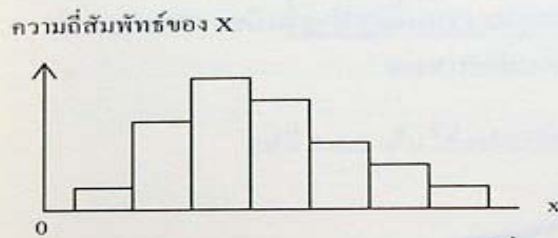
c. Very large data set

Copyright © 2005 Pearson Prentice Hall, Inc.

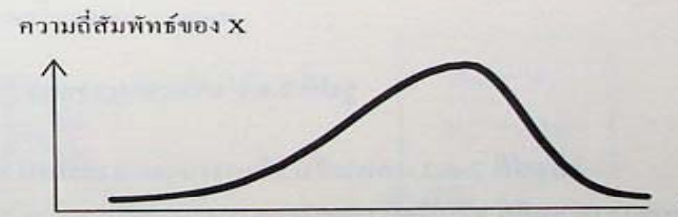
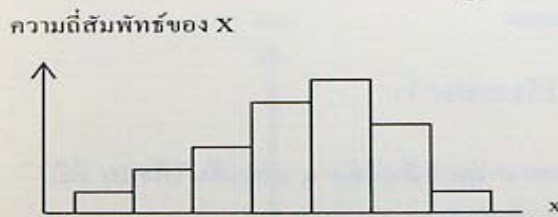
Number of Observations in Data Set	Number of Classes
Less than 25	5-6
25-50	7-14
More than 50	15-20



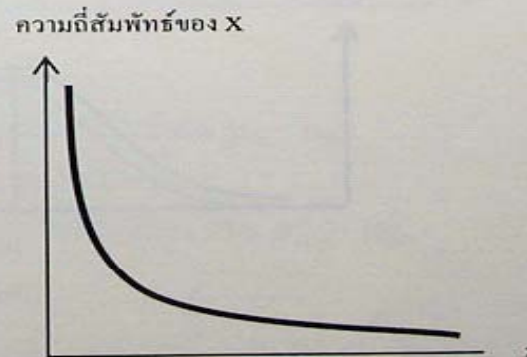
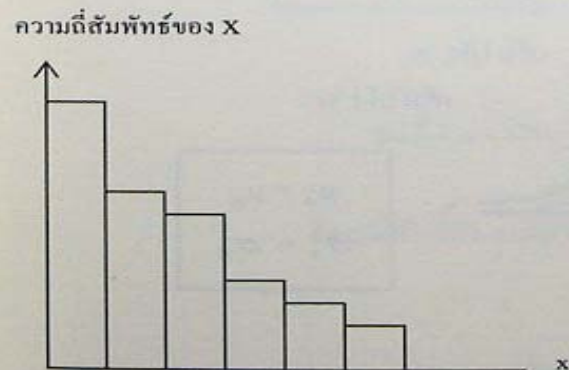
รูปที่ 2.5.1 ฮิสโตแกรมของข้อมูลเกี่ยวกับขนาดเส้นผ่าศูนย์กลาง  
ของแท่งเหล็กจำนวนหนึ่งที่มีการแจกแจงแบบ โกลังระฆังกว่า



รูปที่ 2.5.2 ฮิสโตแกรมแสดงลักษณะเบ้ขวา



รูปที่ 2.5.3 ฮิสโตแกรมแสดงลักษณะเบ้ซ้าย



รูปที่ 2.5.4 ฮิสโตแกรมแสดงถึงการแจกแจงแบบ Exponential

# เครื่องหมายผลรวม ( $\Sigma$ )

- เพื่อง่ายในการสั่งให้หาผลรวมของข้อมูล
- ทุกค่าข้อมูลในชุดข้อมูล จะระบุด้วย ตัวเลขห้อยล่าง (subscript)

$$x_1, x_2, x_3, x_4, x_5, \dots, x_n$$

- เครื่องหมายผลรวม แสดงผลรวมของทุกค่าข้อมูลข้างบนเข้าด้วยกัน คือ

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + x_4 + \dots + x_n$$

# การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)

- ค่าแนวโน้มเข้าสู่ส่วนกลาง เป็นตัวเลขที่เป็นค่ากลางหรือค่าเฉลี่ยของชุดข้อมูล
  - **ค่าสถิติ** (Statistic): คำนวณจากข้อมูลตัวอย่าง
  - **ค่าพารามิเตอร์** (Parameter): คำนวณจากประชากร
- ค่าแนวโน้มเข้าสู่ส่วนกลาง ที่นิยม
  - **ค่าเฉลี่ยเลขคณิต** (Mean)
  - **มัธยฐาน** (Median)
  - **ฐานนิยม** (Mode)

# การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)

- ค่าเฉลี่ย (Mean)

- รวมทุกค่าของข้อมูลเข้าด้วยกัน แล้วหารด้วยจำนวนข้อมูลทั้งหมด
- ค่าเฉลี่ยจากข้อมูลตัวอย่าง แทนด้วยเครื่องหมาย  $\bar{x}$
- ค่าเฉลี่ยจากข้อมูลประชากร แทนด้วยเครื่องหมาย  $\mu$
- สูตรในการคำนวณ

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

และ

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

# การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)

- มัธยฐาน (Median)

- ค่าข้อมูลที่อยู่ตรงกลางของกลุ่มข้อมูลที่เรียงลำดับ
- มัธยฐาน แทนด้วยเครื่องหมาย  $m$
- ถ้าจำนวนข้อมูล ( $n$ ) เป็นเลขคี่ คำนวณตำแหน่งมัธยฐานจาก  $\frac{n}{2} + 0.5$
- ถ้าจำนวนข้อมูล ( $n$ ) เป็นเลขคู่ คำนวณค่ามัธยฐานจากค่าเฉลี่ยของข้อมูลตำแหน่ง  $\frac{n}{2}$  และ  $\frac{n}{2} + 1$

# การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)

- **ฐานนิยม (Mode)**

- ค่าที่เกิดขึ้นบ่อยครั้งหรือมีค่าซ้ำในจำนวนมากที่สุด
- ชุดข้อมูลอาจมีค่า ฐานนิยม หลายค่า หรือ ไม่มีเลยก็ได้
- ข้อมูลที่แสดงด้วย Histogram จะมีช่วงชั้นที่เป็นฐานนิยม (คือช่วงชั้นที่มีความถี่สูงสุด)

## การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency)

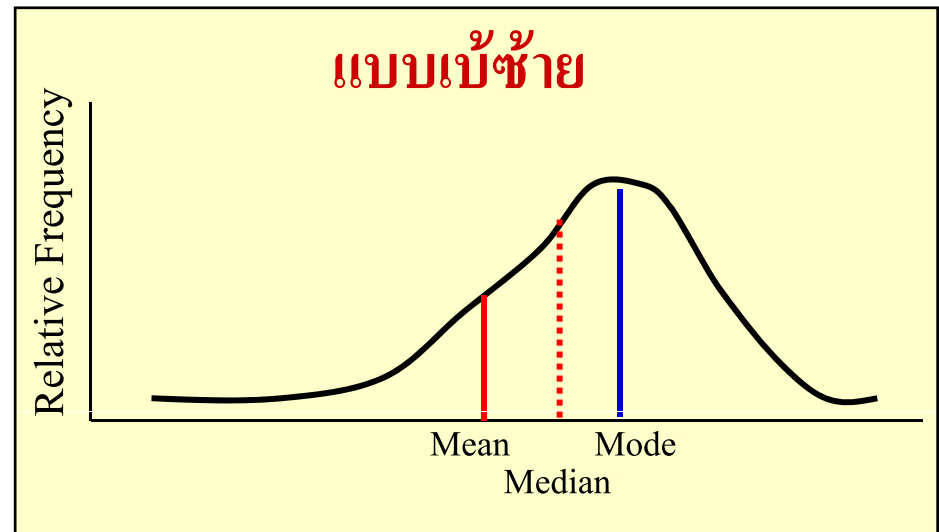
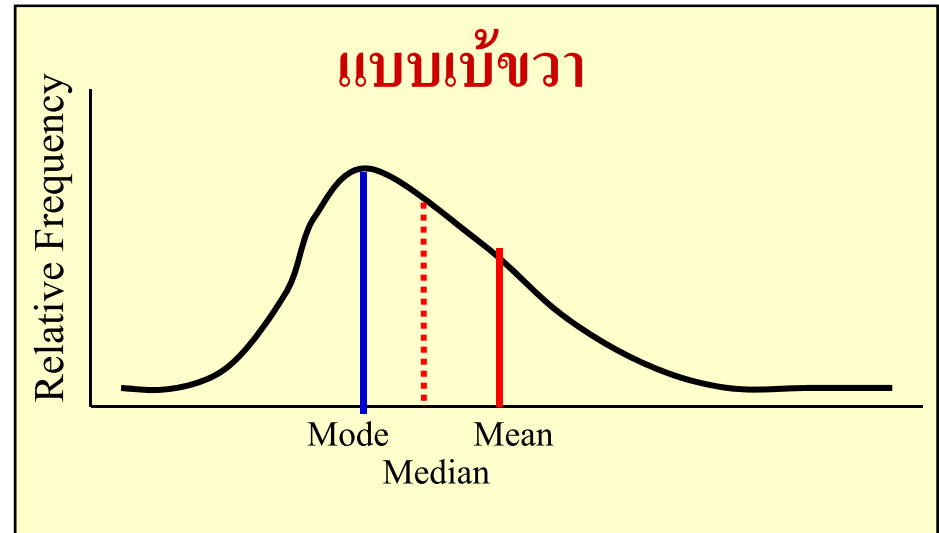
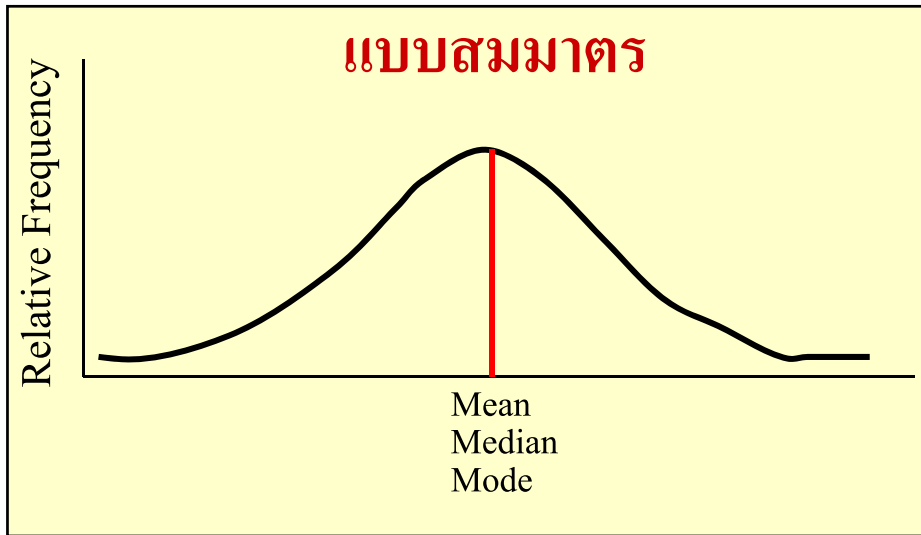
- ชุดข้อมูล 1 3 5 6 8 8 9 11 12

- ค่าเฉลี่ย  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1+3+5+6+8+8+9+11+12}{9} = \frac{63}{9} = 7$

- มัชฌิมฐาน จากสูตร  $\frac{n}{2} + 0.5$  คือตำแหน่งที่ 5 คือ เท่ากับ 8

- ฐานนิยม คือ 8

# ค่า Mean, Median & Mode กับโค้งแจกแจงความถี่



## การวัดการกระจาย (Variability)

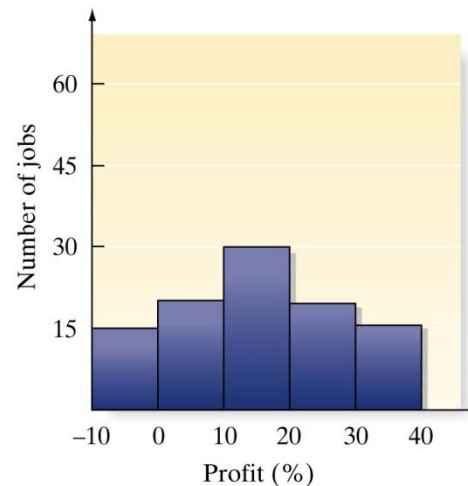
- การกระจาย – ความแตกต่างของค่าข้อมูลซึ่งอยู่ภายในกลุ่มของชุดข้อมูลนั้น
- ค่ากระจายตัวที่นิยม
  - **ค่าพิสัย** (Range)
  - **ค่าแปรปรวน** (Variance)
  - **ค่าเบี่ยงเบนมาตรฐาน** (Standard Deviation)

# การวัดการกระจาย (Variability)

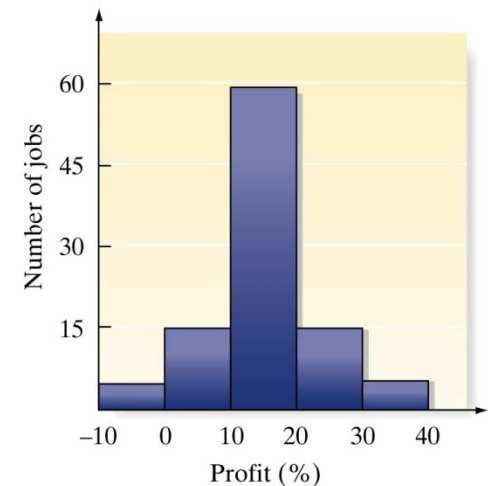
- **พิสัย (Range)**

- ผลต่างระหว่างข้อมูลที่มีค่ามากที่สุดและค่าต่ำสุดของกลุ่มข้อมูลทั้งหมด
- คำนวณจากเพียงสองค่าจากชุดข้อมูล

- การกระจายของทั้งสอง  
มีพิสัยเท่ากัน ?



a. Cost estimator A



b. Cost estimator B

# การวัดการกระจาย (Variability)

- ความแปรปรวนของตัวอย่าง (Sample Variance -  $s^2$ )

- ผลรวม ของทุกค่าข้อมูลลบด้วยค่าเฉลี่ยแล้วยกกำลังสอง (Sum of Square) จากนั้นหารด้วย  $(n-1)$

- สูตรการคำนวณ

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

- ทำไมต้องยกกำลังสอง ค่าความแตกต่าง ?

- ผลรวมทั้งหมดของค่าความแตกต่าง (ทุกค่าข้อมูลลบด้วยค่าเฉลี่ย) จะมีค่าเท่ากับศูนย์

# การวัดการกระจาย (Variability)

- ความเบี่ยงเบนมาตรฐาน ของตัวอย่าง (Sample Standard Deviation - s)
  - Square Root ค่า ความแปรปรวนของตัวอย่าง

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{s^2}$$

- ทำไมต้อง Square Root ?
  - แสดงให้เห็น หน่วยเริ่มแรกของการวัดค่า

# สรุปสูตร Mean, Variance & Standard Deviation

	ตัวอย่าง (Sample)	ประชากร (Population)
Mean	$\bar{X} = \frac{\sum_{i=1}^n x_i}{n-1}$	$\mu = \frac{\sum_{i=1}^N x_i}{N}$
Variance	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$
Standard Deviation	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{s^2}$	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} = \sqrt{\sigma^2}$

# การวัดการกระจายตัวระหว่างกลุ่มข้อมูล

- **ค่าสัมประสิทธิ์แห่งความแปรผัน** (Coefficient of Variation)
- เป็นค่าวัดการกระจายสัมพัทธ์ (Relative Variation) เพื่อเปรียบเทียบการกระจายระหว่างกลุ่มข้อมูล

- สูตรในการคำนวณ

$$C.V. = \frac{S}{\bar{X}} (100) \%$$

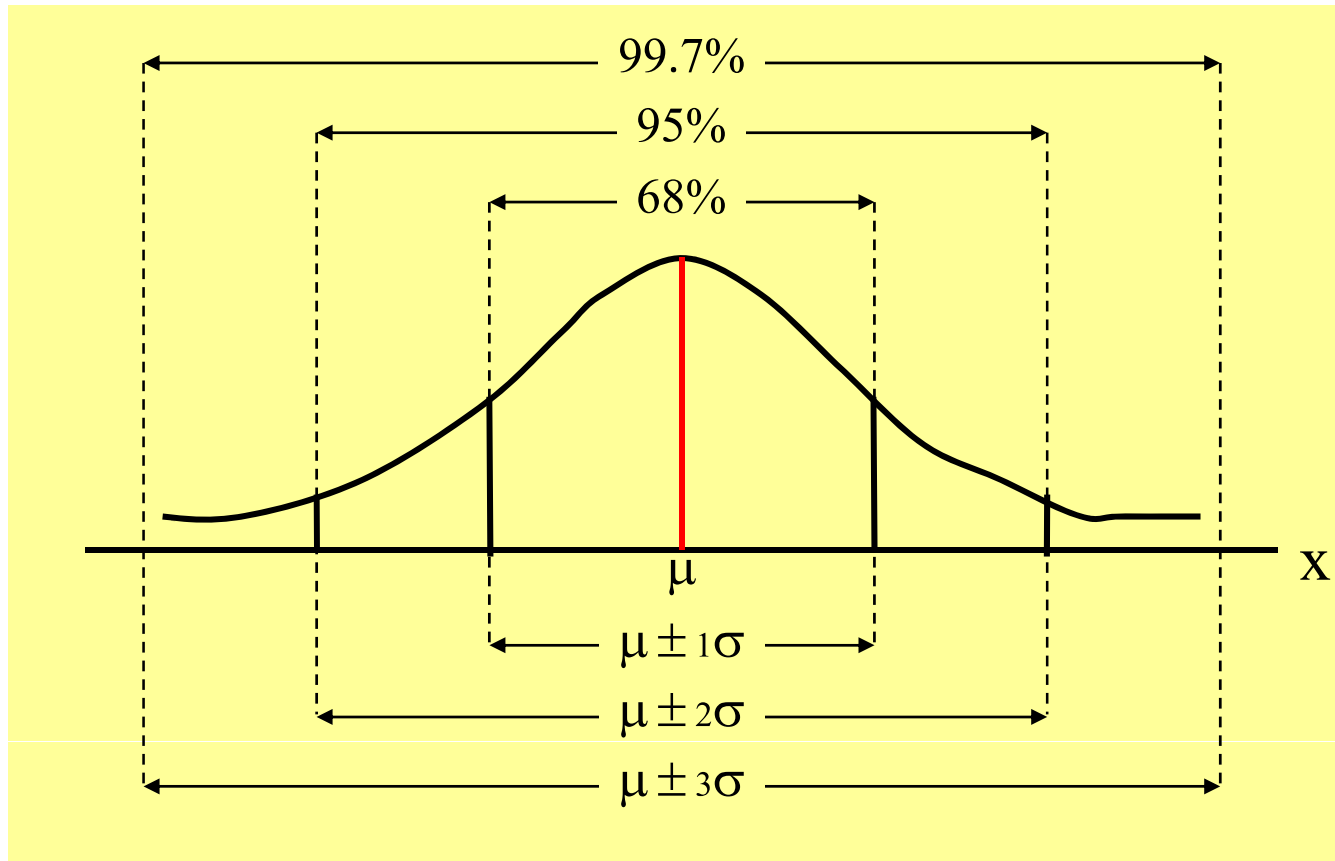
- ตัวอย่าง

ค่าสถิติ	ตัวอย่างชุดที่ 1 (ผู้ใหญ่)	ตัวอย่างชุดที่ 2 (เด็ก)
รายจ่ายเฉลี่ยต่อวัน	145 บาท	80 บาท
ค่าเบี่ยงเบนมาตรฐาน	10 บาท	10 บาท
Coefficient of Variation	$\frac{10}{145} (100) = 6.9\%$	$\frac{10}{80} (100) = 12.5\%$

# กฎ Empirical

- เป็นเครื่องมือที่ใช้บอกสัดส่วนของจำนวนข้อมูลที่อยู่ในช่วงต่างๆ เมื่อเทียบกับจำนวนข้อมูลทั้งหมด
- **เงื่อนไข:** มีการแจกแจงแบบโค้งปกติ (Normal Curve)

หรือ  $x \sim \text{Normal}(\mu, \sigma) \rightarrow$  สัดส่วน คือ  $\mu \pm k \cdot \sigma$  (โดย  $k = 1, 2, 3$ )



## กฎ Empirical: ตัวอย่าง

หนึ่งปี (365 วัน) ที่ผ่านมา ร้านแมคโดนัลด์ พบว่าขายแฮมเบอร์เกอร์ได้โดยเฉลี่ย 15.1 ชิ้น/วัน มีค่าเบี่ยงเบนมาตรฐาน 3.1 ชิ้น โดยมีการกระจายตัวลักษณะใกล้เคียงกับเส้นโค้งปกติ (Normal Curve)

ช่วง	สูตร	การคำนวณ	พิสัย	% จน.วันในช่วง
$\pm 1s$	$\bar{x} \pm 1s$	$15.1 \pm 1(3.1)$	12.0 ถึง 18.2 ชิ้น	68%
$\pm 2s$	$\bar{x} \pm 2s$	$15.1 \pm 2(3.1)$	8.9 ถึง 21.3 ชิ้น	95%
$\pm 3s$	$\bar{x} \pm 3s$	$15.1 \pm 3(3.1)$	5.8 ถึง 24.4 ชิ้น	99.7%

# ค่ามาตรฐาน (Standard Value or Z-Score)

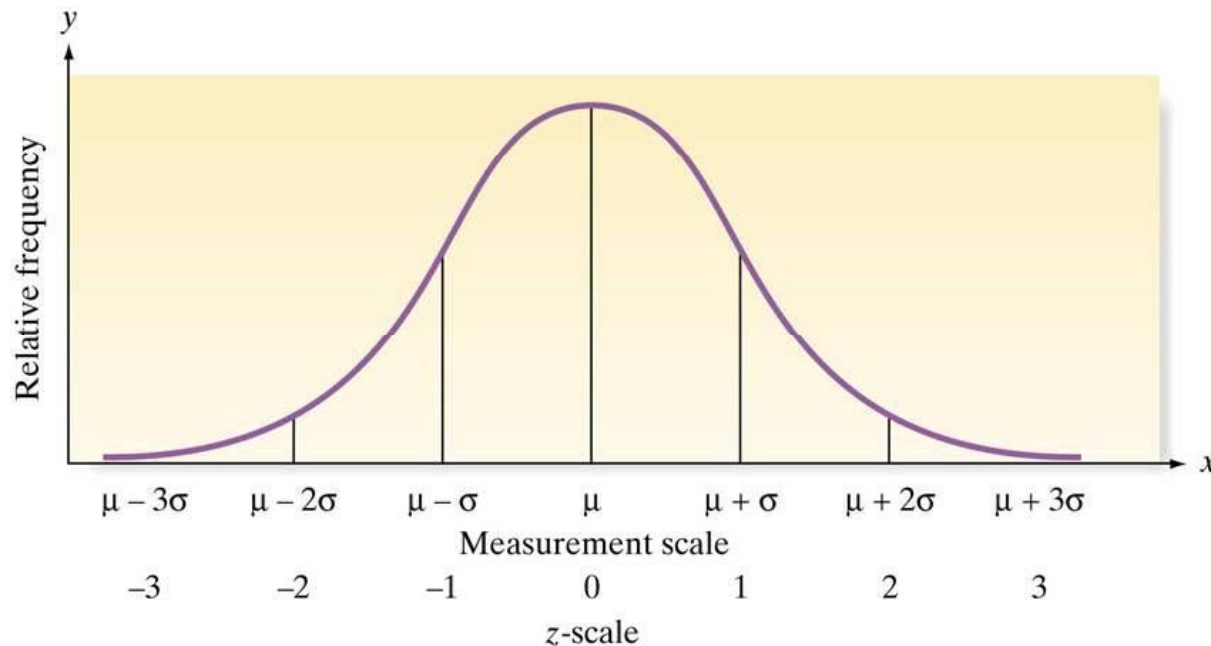
- ค่ามาตรฐาน (z-score) – วัดระยะระหว่างค่าของข้อมูล X และค่าเฉลี่ย

- ใช้เปรียบเทียบค่าที่มาจากประชากรที่มีการแจกแจงที่ต่างกัน หรือชุดข้อมูลที่ต่างกัน

- สูตรในการคำนวณ

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{x - \bar{x}}{s}$$



# ค่ามาตรฐาน (Z-Score): ตัวอย่าง (p.90)

ผู้สมัคร	ข้อสอบ	คะแนนสอบ	คะแนนเฉลี่ย	S.D.	Z-score
วรากร	TOEFL	560	550	28.3	0.3534
วีรจักร	SAT	1268	1255	72	0.1806

