# บทที่ 1 :
# ทบทวนความรู้พื้นฐานเกี่ยวกับสถิติ

208263 : สถิติเบื้องต้น

โดย... ผศ. ดร.สุคนธ์ ประสิทธิ์วัฒนเสรี
ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

# วัตถุประสงค์

นักศึกษาสามารถทราบถึงความสำคัญ และบทบาทของสถิติศาสตร์ในวิชาการและวิชาชีพอื่น

นักศึกษาสามารถอธิบายลักษณะข้อมูลทางสถิติ การเกิดขึ้นของข้อมูล และวิธีการเก็บรวบรวมข้อมูล

นักศึกษาสามารถจัดการกับข้อมูลโดยการเลือกใช้สถิติเชิงพรรณนาที่เหมาะสม

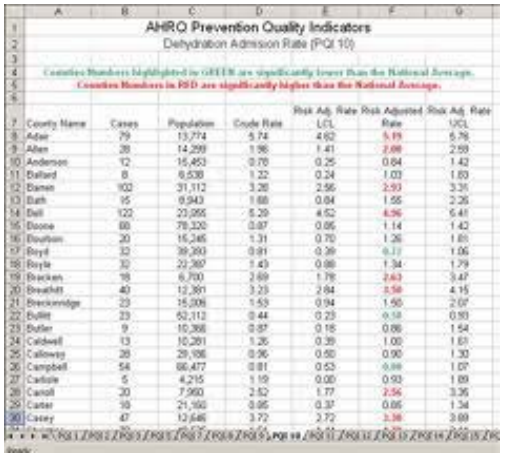
# Outline

- ความสำคัญและบทบาทของสถิติศาสตร์ในวิชาการและวิชาชีพต่าง ๆ
- ลักษณะของข้อมูลทางสถิติ
- วิธีการเก็บรวบรวมข้อมูล
- นิยามศัพท์ทางสถิติ
- การวิเคราะห์ข้อมูลโดยใช้สถิติพรรณนา

# ความสำคัญและบทบาทของสถิติศาสตร์ในวิชาการและวิชาชีพต่าง ๆ

# ความสำคัญของสถิติ

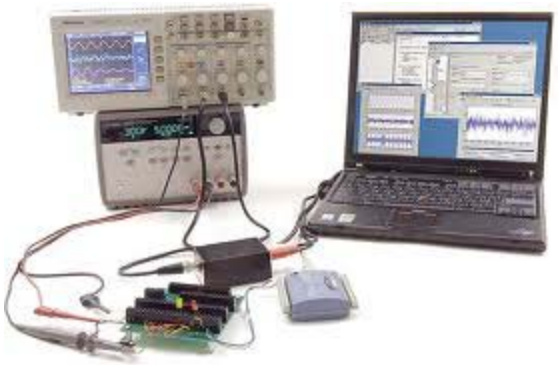
- โลกในยุคปัจจุบัน เป็นโลกแห่งข้อมูลสารสนเทศ
  - ผู้ที่สามารถใช้ประโยชน์จาก**ข้อมูล**ได้อย่างเต็มที่ ย่อมเป็นผู้ที่ได้เปรียบกว่า
  - การจะใช้ประโยชน์จากข้อมูลดังกล่าว จำเป็นต้องอาศัย '**สถิติศาสตร์**'



AHRQ Prevention Quality Indicators

Dehydration Admission Rate (PQI 10)

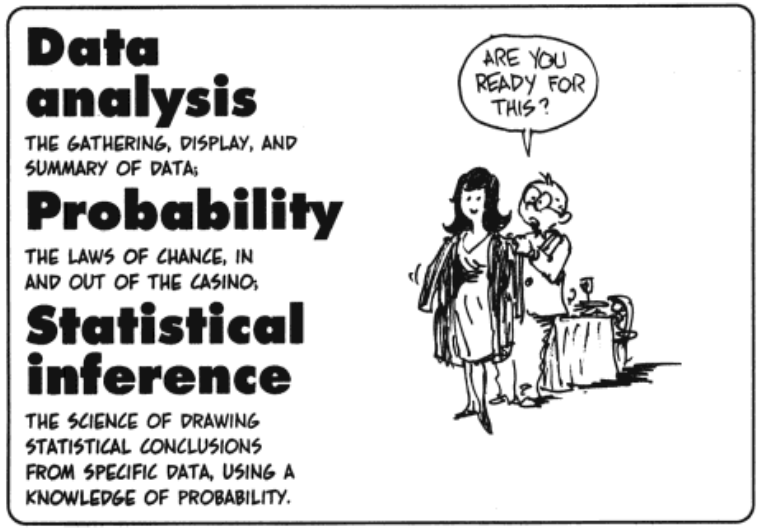
| County Name | Cases | Population | Crude Rate | Risk Adj. Rate LCL | Risk Adjusted Rate | Risk Adj. Rate UCL |
|---|---|---|---|---|---|---|
| Adair | 79 | 13,774 | 5.74 | 4.62 | 5.19 | 5.76 |
| Allen | 38 | 14,299 | 1.96 | 1.41 | 2.68 | 2.59 |
| Anderson | 12 | 16,453 | 0.70 | 0.25 | 0.64 | 1.42 |
| Ballard | 6 | 8,638 | 1.22 | 0.24 | 1.03 | 1.83 |
| Barren | 102 | 31,112 | 3.28 | 2.56 | 2.93 | 3.31 |
| Bath | 15 | 8,043 | 1.68 | 0.84 | 1.55 | 2.26 |
| Bell | 122 | 23,055 | 5.29 | 4.52 | 4.96 | 5.41 |
| Boone | 88 | 78,320 | 0.87 | 0.65 | 1.14 | 1.42 |
| Bourbon | 20 | 15,245 | 1.31 | 0.70 | 1.26 | 1.81 |
| Boyd | 33 | 39,293 | 0.91 | 0.39 | 0.77 | 1.06 |
| Boyle | 32 | 22,397 | 1.43 | 0.88 | 1.34 | 1.79 |
| Bracken | 18 | 6,700 | 2.69 | 1.78 | 2.63 | 3.47 |
| Breathitt | 40 | 12,391 | 3.23 | 2.84 | 3.58 | 4.15 |
| Breckenridge | 23 | 18,306 | 1.53 | 0.94 | 1.56 | 2.07 |
| Bullitt | 23 | 62,112 | 0.44 | 0.23 | 0.38 | 0.93 |
| Butler | 9 | 10,368 | 0.87 | 0.18 | 0.86 | 1.54 |
| Caldwell | 13 | 10,281 | 1.26 | 0.39 | 1.00 | 1.61 |
| Calloway | 28 | 29,106 | 0.96 | 0.50 | 0.90 | 1.30 |
| Campbell | 54 | 86,477 | 0.81 | 0.53 | 0.88 | 1.07 |
| Carlisle | 5 | 4,215 | 1.19 | 0.00 | 0.93 | 1.89 |
| Carroll | 20 | 7,960 | 2.52 | 1.77 | 2.56 | 3.36 |
| Carter | 18 | 21,160 | 0.85 | 0.37 | 0.85 | 1.34 |
| Casey | 47 | 12,646 | 3.72 | 2.72 | 3.38 | 3.89 |



# ความหมายของสถิติ

- **สถิติ**

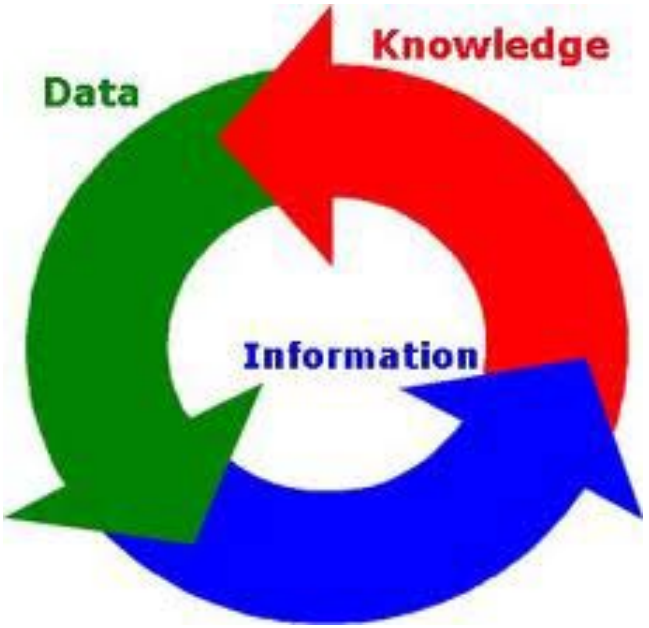
  คือวิทยาศาสตร์ของการเก็บรวบรวมข้อมูล การบริหารจัดการ การวิเคราะห์ และการตีความข้อมูล เพื่อใช้ช่วยในการตัดสินใจ



# ความหมายของสถิติ

- **สถิติ** จำแนกได้เป็น
  - **ข้อมูลสถิติ** ข้อเท็จจริงที่ทำการรวบรวมมาได้ อาจอยู่ในรูปตัวเลข หรือตัวอักษร
  - **สถิติศาสตร์** ขบวนการต่าง ๆ ที่นำมาช่วยในการจัดการกับข้อมูลทางสถิติ

**สถิติศาสตร์** จะช่วยในการจัดการกับ**ข้อมูล (Data)** เพื่อให้ได้มาซึ่ง **สารสนเทศ (Information)** ที่สำคัญ จนได้ **องค์ความรู้ (Knowledge)** เพื่อช่วยในการแก้ปัญหาหรือการตัดสินใจต่อไป



# บทบาทความสำคัญของสถิติ

- ในการตัดสินใจจัดการกับข้อคำถามหรือปัญหาที่เกิดขึ้น ส่วนใหญ่มักเป็นกรณีที่ไม่อาจทราบผลลัพธ์ที่แน่นอนได้ นั่นคือมี ‘**ความไม่แน่นอน (Uncertainty)**’ เกิดขึ้น
  - การจัดการกับการระบาดของโรคไข้เลือดออกในพื้นที่เชียงใหม่
  - การตัดสินใจทำการรักษาผู้ป่วยด้วยวิธีการผ่าตัด หรือใช้รังสีรักษา
  - …
- การวิเคราะห์ข้อมูลเพื่อให้ได้สารสนเทศสำหรับช่วยในการหาคำตอบของเหตุการณ์ภายใต้ความไม่แน่นอน คือ ‘**สถิติศาสตร์**’

# บทบาทความสำคัญของสถิติ

- **บทบาทของสถิติต่อชีวิตประจำวัน**

  การอาศัย**ข้อมูลทางสถิติ** ช่วยในการตัดสินใจสำหรับการใช้ชีวิตประจำวัน

- **บทบาทของสถิติต่อวิชาชีพต่าง ๆ**

  การอาศัย**สถิติศาสตร์** ช่วยในงานวิจัยเพื่อหาข้อสรุปสำหรับใช้เป็นแนวทางในการตัดสินใจขององค์กร หน่วยงาน

# บทบาทของสถิติต่อชีวิตประจำวัน

- การอาศัยข้อมูลทางสถิติช่วยในการตัดสินใจ เช่น
  - การตัดสินใจเลือกซื้อหลอดไฟ

# บทบาทของสถิติต่อชีวิตประจำวัน

- การตัดสินใจเลือกร้านอาหาร

# บทบาทของสถิติต่อวิชาชีพต่าง ๆ

- **ด้านวิศวกรรม**

  การพัฒนาอุปกรณ์สำหรับใช้ทดแทนภายในเครื่องจักรกล วัสดุที่ใช้ ?

การสร้างเครื่องมือต้นแบบสำหรับตรวจวัดคุณภาพอากาศ

ความถูกต้อง?

ความคุ้มค่า?

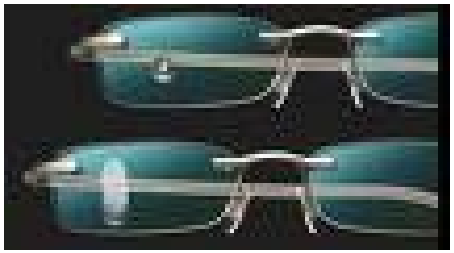
# บทบาทของสถิติต่อวิชาชีพต่าง ๆ

- **ด้านการแพทย์**

  การพัฒนายารักษาโรค ชนิดยา/ความเข้มข้น/ช่วงเวลา ?

# บทบาทของสถิติต่อวิชาชีพต่าง ๆ

- **ด้านวิทยาศาสตร์**
  การพัฒนาสารเคลือบวัสดุนาโน ความทนทาน/คุณภาพ ?

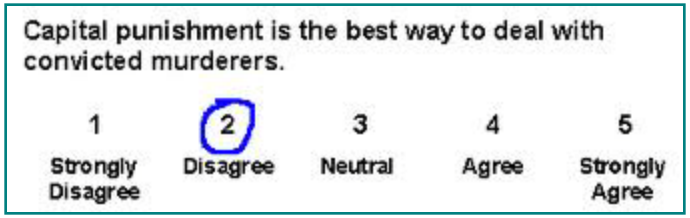


# ลักษณะของข้อมูลทางสถิติ

# ชนิดของข้อมูล (Types of data)

- **ข้อมูลคุณภาพ (Qualitative data)**

  ข้อมูลที่มีค่าประกอบด้วย ข้อความ คำบรรยาย หรือจำนวนที่ไม่ได้สื่อความหมายเชิงตัวเลข




- **ข้อมูลปริมาณ (Quantitative data)**

  ข้อมูลที่มีเป็นตัวเลขที่ได้จากการวัดหรือการนับ

# Ex : ตารางแสดงราคาต่อหน่วยของโทรศัพท์มือถือรุ่นต่าง ๆ

| รุ่น | ราคา |
|---|---|
| 1110 | 1590 |
| 2610 | 2850 |
| 2626 | 3450 |
| 5300 | 8800 |
| 5500 | 9900 |
| 6080 | 4870 |

ข้อมูลที่มีจัดเป็นข้อมูลคุณภาพหรือปริมาณ ?

# มาตรวัดข้อมูล พิจารณาจากค่าที่เป็นไปได้ของข้อมูล

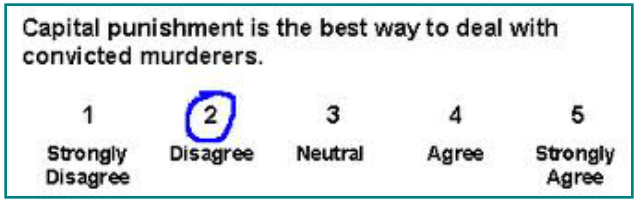
- **มาตรนามบัญญัติ (Norminal scale)**

ค่าข้อมูลอยู่ในรูปข้อความ ซึ่งสามารถบอกได้ถึงความแตกต่าง หรือความเหมือนกัน

เช่น รูปทรงของวัสดุ: – ทรงกลม – วงรี – หกเหลี่ยม

- มาตรเรียงลำดับ (Ordinal scale)

ค่าข้อมูลอยู่ในรูปข้อความ ที่สามารถจัดลำดับความสำคัญ/ความแตกต่างได้




เช่น ความสำคัญของการเป็นคนซื่อสัตย์:

– น้อยมาก – ไม่เห็นถึงความสำคัญ

– สำคัญมาก – สำคัญพอสมควร

# มาตรวัดข้อมูล พิจารณาจากค่าที่เป็นไปได้ของข้อมูล

- มาตรอันตรภาค (Interval scale)

  ค่าข้อมูลเป็นตัวเลขที่ได้จากการวัดหรือนับ ไม่มีศูนย์ที่แท้จริง สามารถนำค่าข้อมูลมาบวกลบกันได้

  เช่น อุณหภูมิรายวันของจังหวัดเชียงใหม่

- มาตรอัตราส่วน (Ratio scale)

  ค่าข้อมูลเป็นตัวเลขมีศูนย์ที่แท้จริง สามารถทำการบวกลบคูณหารกันได้

  เช่น น้ำหนักของผู้ป่วย

# ลองทำ: พิจารณาชนิดของข้อมูลต่อไปนี้

| คนที่ | เพศ | อายุ (ปี) | คะแนนสอบ | ผลการเรียน |
|---|---|---|---|---|
| 1 | ชาย | 20 | 30 | ปานกลาง |
| 2 | หญิง | 21 | 40 | ดี |
| 3 | หญิง | 19 | 50 | ดีมาก |

ก. ข้อมูลชุดนี้ประกอบด้วยตัวแปรกี่ตัว

ข. ให้ระบุชนิดของค่าตัวแปรแต่ละตัว

# ชุดข้อมูล (Data set)

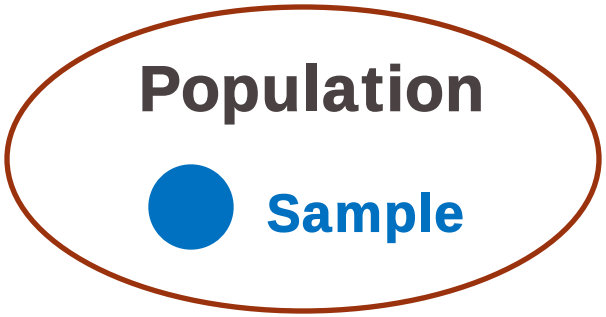
จำแนกเป็น

- ชุดข้อมูล “ประชากร (Population)”

  การเก็บรวบรวมทั้งหมดของผลลัพธ์ คำตอบ การวัด หรือการนับในสิ่งที่สนใจศึกษา

- ชุดข้อมูล “ตัวอย่าง (Sample)”

  ส่วนหนึ่งของประชากร




ประชากร จำแนกเป็น

- ประชากรจำกัดหรืออันตะ (Finite population)

  สามารถนับสมาชิกในประชากรได้ครบถ้วน

- ประชากรไม่จำกัดหรืออนันต์ (Infinite population)

# Problem

- จากการสำรวจคนวัยหนุ่มสาว 3002 คน ว่าเคยอ่านข่าวผ่านทางอินเตอร์เน็ตอย่างน้อย 1 ครั้ง/สัปดาห์

  ผลปรากฏว่า มี 600 คน ตอบว่าเคย

  ให้ระบุถึงประชากร และตัวอย่างในการสำรวจครั้งนี้

# พารามิเตอร์ และสถิติ

เมื่อคำนวณค่าจากชุดข้อมูล **ประชากร**/**ตัวอย่าง** จะมี 2 เทอมเข้ามาเกี่ยวข้อง

- **พารามิเตอร์ (Parameter)**

  ตัวเลขหรือค่าที่สื่อถึงลักษณะของประชากร

- **สถิติ (Statistic)**

  ตัวเลขหรือค่าที่สื่อถึงลักษณะของตัวอย่าง

ตัวเลขหรือค่า ข้างต้น อาจเป็น
- ค่าเฉลี่ย, ความแปรปรวน ของตัวแปรที่สนใจ
- ค่าสัดส่วนหรือร้อยละของเหตุการณ์ที่สนใจ
- ค่าแสดงความสัมพันธ์ระหว่างตัวแปร 2 ตัว

# Problem

- ให้พิจารณาว่าค่าตัวเลขที่กล่าวถึงคือค่าพารามิเตอร์ หรือค่าสถิติ
  - ผลจากการสำรวจรายได้ของตัวอย่างบัณฑิต ม.ช. จำนวน 3000 คน พบว่า ค่าเฉลี่ยของรายได้เริ่มต้นมีค่าน้อยกว่า **50,000** บาท/ปี
  - รายได้เริ่มต้นของบัณฑิตทั้งหมด 60 คนที่สำเร็จการศึกษาสาขาสถิติ ม.ช. เพิ่มสูงขึ้นกว่าปีที่ผ่านมา **8.5**%

# พารามิเตอร์ และสถิติ

**สัญลักษณ์** แทนค่าพารามิเตอร์ และค่าสถิติ

| ค่า | พารามิเตอร์ | สถิติ |
|---|---|---|
| ค่าเฉลี่ย | $\mu$ (mu) | $\bar{x}$ |
| ความแปรปรวน | $\sigma^2$ (sigma square) | $s^2$ |
| ค่าสัดส่วน | $P$ | $\hat{p}$ |
| ค่าความสัมพันธ์ (สัมประสิทธิ์สหสัมพันธ์) | $\rho$ (rho) | $r$ |

# วิธีการเก็บรวบรวมข้อมูล

# การเกิดขึ้นของข้อมูล

- ข้อเท็จจริงของเรื่องใดเรื่องหนึ่งที่สนใจ
  - อาจเกิดขึ้นแล้ว รอเพียงผู้วิจัยทำการรวบรวมหรือสำรวจ เช่น
    - ลักษณะทางสังคมประชากรของผู้ติดเชื้อ HIV

  - ยังไม่เกิดขึ้น ผู้วิจัยต้องเป็นผู้ดำเนินการสร้างและรวบรวมเอง
    - ผลการรักษาผู้ติดเชื้อ HIV ด้วยยาสูตรใหม่

# การเก็บรวบรวมข้อมูล

- **การสำมะโน (Census)**

  การเก็บรวบรวมข้อมูลทั้งหมดในประชากร

- **การศึกษาหรือสำรวจด้วยตัวอย่าง (Sample survey)**

  การเลือกรวบรวมข้อมูลเพียงบางส่วนของประชากร

- **การทดลอง (Experimental)**

  การเสาะหาคำตอบหรือผลลัพธ์ที่เกิดขึ้นภายใต้เงื่อนไขต่าง ๆ โดยมีกระบวนการในการวางแผนงานที่ชัดเจน

- **การจำลองแบบ (Simulation)**

  การทดลองเสมือนจริงโดยอาศัยวิธีการจำลองสถานะการณ์หรือเงื่อนไขต่าง ๆ ขึ้น

# การเก็บรวบรวมข้อมูล (ต่อ)

- การสำมะโน (Census)

  เช่น

  - การสำมะโนครัวประชากร (สภาพสังคมเศรษฐกิจของแต่ละครัวเรือน)
  - การสำมะโนอุตสาหกรรม

  เป็นวิธีที่เหมาะสำหรับหน่วยงานของรัฐ

# การเก็บรวบรวมข้อมูล (ต่อ)

- **การศึกษาหรือสำรวจด้วยตัวอย่าง** (Sample survey)

  ต้องมีการกำหนด

  - **ประชากร** : สมาชิกในกรอบที่กำลังศึกษาทั้งหมด
  - **หน่วยประชากร** : แต่ละสมาชิกของประชากร
  - **ตัวอย่าง** : สมาชิกส่วนหนึ่งของประชากร
  - **หน่วยตัวอย่าง** : แต่ละสมาชิกของตัวอย่าง
  - **วิธีการเก็บรวบรวมตัวอย่าง (Sampling techniques)** :
    การสุ่มเลือกอย่างง่าย/การเลือกแบบกลุ่ม/
    การเลือกแบบชั้นภูมิ/การเลือกแบบมีระบบ/ ...

เช่น การสำรวจความสุขของคนในชุมชน

การศึกษาเปรียบเทียบค่าใช้จ่ายต่อวันของนศ. ชายกับหญิง

# การเก็บรวบรวมข้อมูล (ต่อ)

- **การทดลอง (Experimental)**

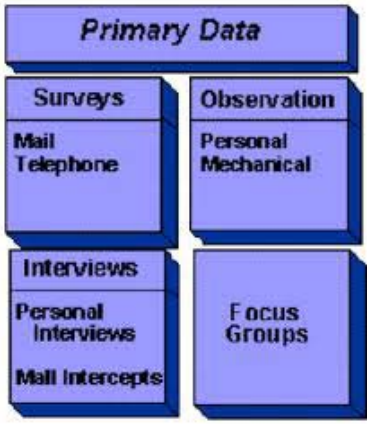
  เช่น การเปรียบเทียบประสิทธิภาพระหว่างเครื่องจักร 2 รุ่น

- **การจำลองแบบ (Simulation)**

  เช่น การจำลองสถานการณ์น้ำท่วมใน อ.เมืองเชียงใหม่

ข้อมูลที่ได้จากการรวบรวมด้วยตนเอง – **ข้อมูลปฐมภูมิ (Primary data)**

ข้อมูลที่ได้มาจากการรวบรวมของผู้อื่นหรือหน่วยงานอื่น – **ข้อมูลทุติยภูมิ (Secondary data)**



# ทดลองทำ

- ในการศึกษาพฤติกรรมการจัดการกับความเครียดของนักศึกษา มช. จึงทำการเก็บรวบรวมข้อมูลจากตัวอย่างนักศึกษาจำนวน 1000 คน

อยากทราบวิธีการเก็บรวบรวมข้อมูลที่ใช้

ก. การสำรวจด้วยตัวอย่าง

ข. การทดลอง

ค. การสำมะโน

ชนิดของข้อมูลที่ได้

ก. ข้อมูลปฐมภูมิ ข. ข้อมูลทุติยภูมิ

# ทดลองทำ

- ในการศึกษาเปรียบเทียบผลการรักษาวัณโรคโดยใช้ยา A กับยา B จึงทำการสุ่มตัวอย่างผู้ป่วยวัณโรคจำนวน 100 คน แบ่งเป็น 2 กลุ่ม กลุ่มที่ 1 ได้รับยา A กลุ่มที่ 2 ได้รับยา B

  อยากทราบวิธีการเก็บรวบรวมข้อมูลที่ใช้

  ก. การสำรวจด้วยตัวอย่าง

  ข. การทดลอง

  ค. การสำมะโน

  ชนิดของข้อมูลที่ได้

  ก. ข้อมูลปฐมภูมิ ข. ข้อมูลทุติยภูมิ

# นิยามศัพท์ทางสถิติ

- สถิติ
- ข้อมูลสถิติ
- สถิติศาสตร์
- ชนิดข้อมูล และมาตรวัดข้อมูล
- ประชากร และตัวอย่าง
- ประชากรจำกัด และประชากรไม่จำกัด
- พารามิเตอร์ และค่าสถิติ
- การเก็บรวบรวมข้อมูล
  (การสำมะโน, การสำรวจด้วยตัวอย่าง, การทดลอง, การจำลองแบบ)
- ข้อมูลปฐมภูมิ และข้อมูลทุติยภูมิ

# การวิเคราะห์ข้อมูลโดยใช้สถิติพรรณนา

# สถิติศาสตร์

- เป็นศาสตร์ที่ว่าด้วยการจัดการกับข้อมูลทางสถิติอย่างมีกระบวนการ โดยอาศัย**ระเบียบวิธีทางสถิติ**
- **ระเบียบวิธีทางสถิติ**
  - **การวางแผน** - การกำหนดขอบข่ายและวัตถุประสงค์ของเรื่องที่จะศึกษา
  - **การรวบรวมข้อมูล** - รวบรวมข้อความหรือตัวเลขของสิ่งที่จะศึกษา
  - **การนำเสนอข้อมูล** - จัดหมวดหมู่ข้อมูลให้ดูง่าย
  - **การวิเคราะห์ข้อมูล**
  - **การตีความหมาย** - สรุปผลการวิเคราะห์

# ประเภทของสถิติศาสตร์

ในการจัดการกับข้อมูล อาจเลือกดำเนินการได้ 2 รูปแบบ คือ

- **สถิติพรรณนา (Descriptive Statistics)**

  เป็นการอธิบายและนำเสนอลักษณะของข้อมูลที่ศึกษา โดยจะนำเสนอค่าแนวโน้มเข้าสู่ส่วนกลาง การกระจายของข้อมูล หรือนำเสนอในรูปของกราฟต่าง ๆ

- **สถิติอนุมาน (Inferential Statistics)**

  เป็นการนำลักษณะที่พบในข้อมูลกลุ่มตัวอย่าง (Sample) ผ่านวิธีการทางสถิติ เพื่อไปอธิบายถึงลักษณะของข้อมูลประชากร (Population)

# Descriptive Statistics

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 15.7 | 14.5 | 14.5 | 15.3 | 14.8 | 14.1 | 14.5 | 16.8 | 15.3 | 16.6 |
| 15.1 | 15.8 | 15.8 | 17.7 | 16.7 | 14.8 | 15.4 | 15.7 | 14.8 | 16.4 |
| 17.0 | 14.2 | 14.8 | 14.9 | 13.8 | 14.8 | 15.2 | 17.2 | 15.7 | 16.1 |
| 17.3 | 14.2 | 16.5 | 15.6 | 14.6 | 14.9 | 15.4 | 15.8 | 14.6 | 15.0 |
| 17.6 | 16.2 | 14.0 | 15.9 | 15.5 | 14.4 | 15.1 | 15.4 | 15.4 | 15.7 |

**All points** – Population

mean = 15.5, S.D. = 0.95 → **Parameters**

**5 points - Samples**

14.5 15.3 15.4 14.8 15.5

Mean = 15.1, S.D. = 0.35 → **Statistics**

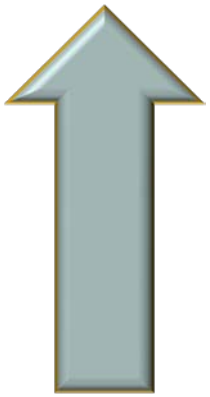
# Inferential Statistics

**Population**

mean = **?**, S.D. = **?**

Proportion = **?**

**Parameters ?**



**การอนุมาน**

**Samples**

mean = 15.6

S.D. = 0.94

Proportion = 0.80

**Statistics**

- การประมาณค่า
- การทดสอบสมมุติฐาน

# Descriptive Statistics

## *Descriptive* Statistics

- การนำเสนอด้วยตาราง (Tabulations)
- ตารางแจกแจงความถี่ (Frequency distribution table)
- กราฟหรือแผนภูมิ (Graphs)

## Descriptive *Statistics*

- การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of central tendency)
- การวัดตำแหน่งและการกระจาย (Measures of dispersion and variability)

# การนำเสนอโดยตาราง

- เมื่อต้องทำการสรุปความในข้อมูลที่มี**ขนาดใหญ่** โดยปกติมักนำเสนอในรูปของตาราง

**Qualitative data :**

ผลการสำรวจหมู่เลือดของน.ศ. 100 คน เป็นดังนี้

O, A, O, O, A, A, A, AB, B, A, O, O, A, O, A, A, A, O, A, O
A, O, A, O, A, A, A, AB, B, O, A, O, O, O, A, A, AB, O, O, A
A, A, O, O, A, A, A, AB, B, O, O, O, O, A, O, A, B, O, O, O
O, A, A, O, A, A, A, AB, B, O, O, A, O, O, A, A, B, O, O, A
O, A, O, O, A, A, A, AB, B, O, O, A, O, A, A, O, O, O, A

# การนำเสนอโดยตาราง

**ตาราง** ผลการสำรวจกลุ่มเลือดของน.ศ. 100 คน

| Group | Number of students |
|---|---|
| A | 43 |
| B | 7 |
| AB | 6 |
| O | 44 |
| Total | 100 |

ตารางแจกแจงความถี่

# การนำเสนอโดยตาราง

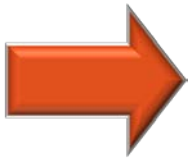
**Quantitative data**

ผลการสำรวจน้ำหนัก (กรัม) ของทารกแรกเกิด 25 คน เป็นดังนี้

| | | | | |
|---|---|---|---|---|
| 3990 | 1250 | 3500 | 1900 | 1500 |
| 2000 | 1900 | 2800 | 1900 | 2400 |
| 2490 | 3200 | 2800 | 2490 | 2400 |
| 2500 | 2400 | 3250 | 2400 | 2800 |
| 2500 | 2800 | 3450 | 2800 | 2500 |

3990 1250 3500 1900 1500
2000 1900 2800 1900 2400
2490 3200 2800 2490 2400
2500 2400 3250 2400 3250
2500 2800 3450 2800 2500

## ตาราง นน. ทารกแรกเกิด



| น้ำหนัก | จำนวน | น้ำหนัก | จำนวน |
|---|---|---|---|
| 1250 | 1 | 2800 | 4 |
| 1500 | 1 | 3200 | 1 |
| 1900 | 3 | 3250 | 2 |
| 2000 | 1 | 3450 | 1 |
| 2400 | 4 | 3500 | 1 |
| 2490 | 2 | 3990 | 1 |
| 2500 | 3 | | |

## ตารางความถี่

| น้ำหนัก | จำนวน | น้ำหนัก | จำนวน |
|---|---|---|---|
| 1250 | 1 | 2800 | 4 |
| 1500 | 1 | 3200 | 1 |
| 1900 | 3 | 3250 | 2 |
| 2000 | 1 | 3450 | 1 |
| 2250 | 4 | 3500 | 1 |
| 2400 | 2 | 3990 | 1 |
| 2500 | 3 | | |

## ตารางแจกแจงความถี่

| น้ำหนัก | จำนวน |
|---|---|
| 1000 – 1499 | 1 |
| 1500 – 1999 | 4 |
| 2000 – 2499 | 7 |
| 2500 – 2999 | 8 |
| 3000 – 3499 | 3 |
| 3500 – 3999 | 2 |
| รวม | 25 |

**ขีดจำกัดชั้น (Class limit)**

**ขีดจำกัดล่าง**

**ขีดจำกัดบน**

# ตารางแจกแจงความถี่

| น้ำหนัก | จำนวน | ขอบเขตชั้น |
|---|---|---|
| 1000 – 1499 | 1 | 999.5 – 1499.5 |
| 1500 – 1999 | 4 | 1499.5 – 1999.5 |
| 2000 – 2499 | 7 | 1999.5 – 2499.5 |
| 2500 – 2999 | 8 | 2499.5 – 2999.5 |
| 3000 – 3499 | 3 | 2999.5 – 3499.5 |
| 3500 – 3999 | 2 | 3499.5 – 3999.5 |
| **รวม** | **25** | |

จุดกลางชั้น (Mid point)
= (1499.5 + 999.5)/2
= (1499 + 1000)/2
= 1249.5

ค่าคงที่ ขึ้นกับค่าสังเกต
- 0.5 เมื่อ xxx
- 0.05 เมื่อ xxx.x

**ขอบเขตชั้น** (Class boundary)
ชั้นที่ 1:
- ขอบเขตบน = (1499+1500)/2 = 1499 + 0.5 =1499.5
- ขอบเขตล่าง = (1000+999)/2 = 1000 - 0.5 = 999.5

**อันตรภาคชั้น** (Class interval, I) = ขอบเขตบน – ขอบเขตล่าง
ชั้นที่ 1: I = 1499.5 – 999.5 = 500

# ความถี่สะสม (Cumulative frequency)

## ตารางแจกแจงความถี่ และความถี่สะสม

| ขอบเขตชั้น (น้ำหนักทารก) | จำนวน | ความถี่สะสม | |
|---|---|---|---|
| | | น้อยกว่า | มากกว่า |
| 999.5 – 1499.5 | 1 | 1 | 25 |
| 1499.5 – 1999.5 | 4 | 5 | 24 |
| 1999.5 – 2499.5 | 7 | 12 | 20 |
| 2499.5 – 2999.5 | 8 | 20 | 13 |
| 2999.5 – 3499.5 | 3 | 23 | 5 |
| 3499.5 – 3999.5 | 2 | 25 | 2 |
| **รวม** | **25** | | |

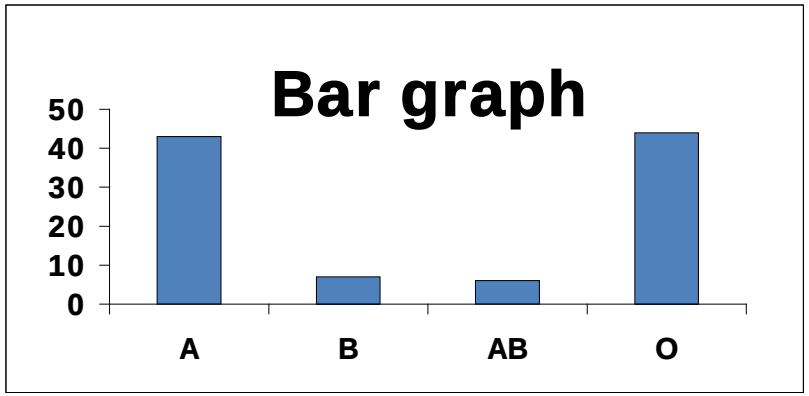
# ความถี่สัมพัทธ์ (Relative frequency)

## ตารางแจกแจงความถี่ และความถี่สัมพัทธ์

| น้ำหนักทารก | จำนวน | ความถี่สัมพัทธ์ (ร้อยละ) |
|---|---|---|
| 1000 – 1499 | 1 | (1/25)*100=4 |
| 1500 – 1999 | 4 | 16 |
| 2000 – 2499 | 7 | 28 |
| 2500 – 2999 | 8 | 32 |
| 3000 – 3499 | 3 | 12 |
| 3500 – 3999 | 2 | 8 |
| **รวม** | **25** | **100** |

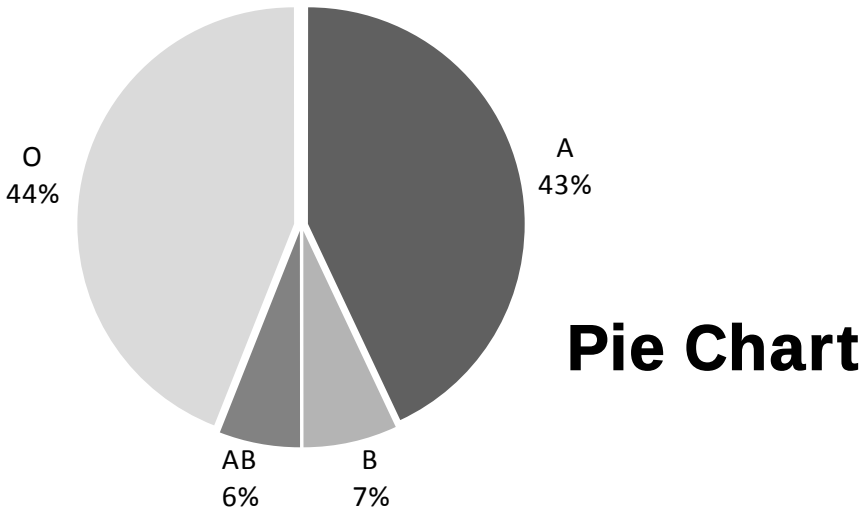
# Graphing Qualitative Data Sets

- Bar graph
- Pie Chart




**ตาราง** กลุ่มเลือดของน.ศ. 100 คน

| Group | N |
|---|---|
| A | 43 |
| B | 7 |
| AB | 6 |
| O | 44 |



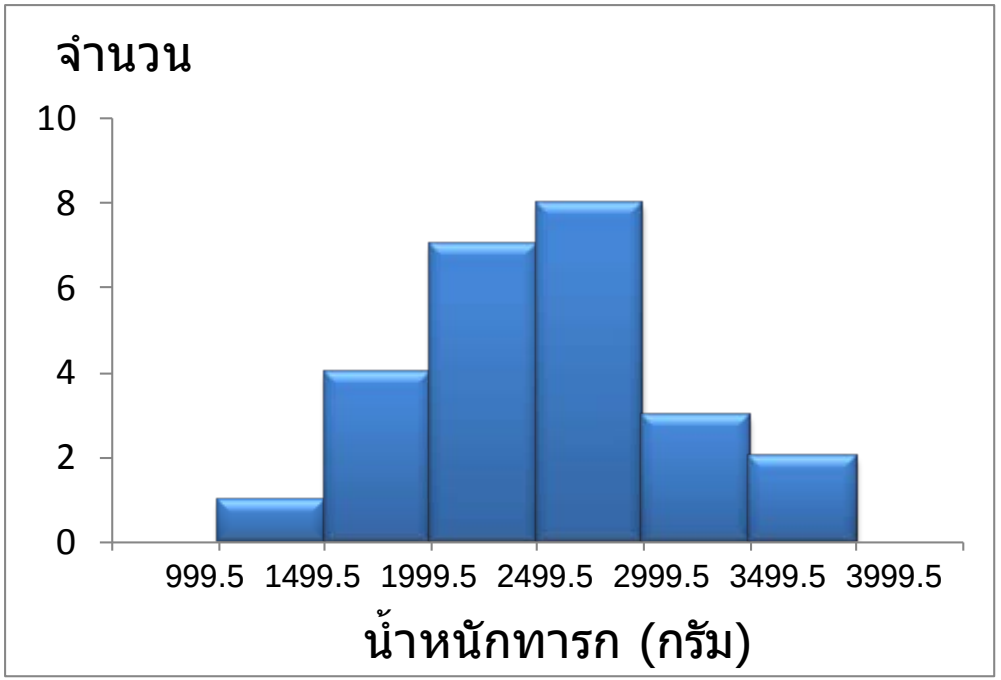
# Graphing Quantitative Data Sets

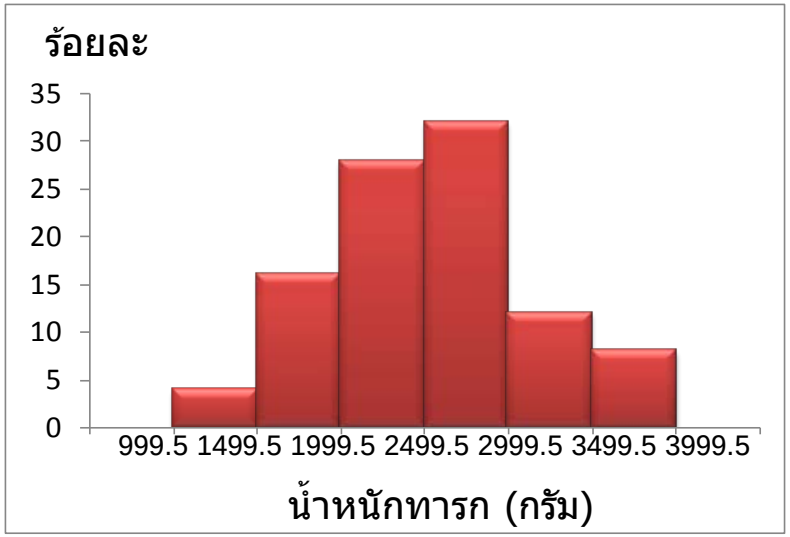
- Histogram
- รูปหลายเหลี่ยมความถี่
- โค้งความถี่
- Stem-and-leaf Plot
- Box Plot

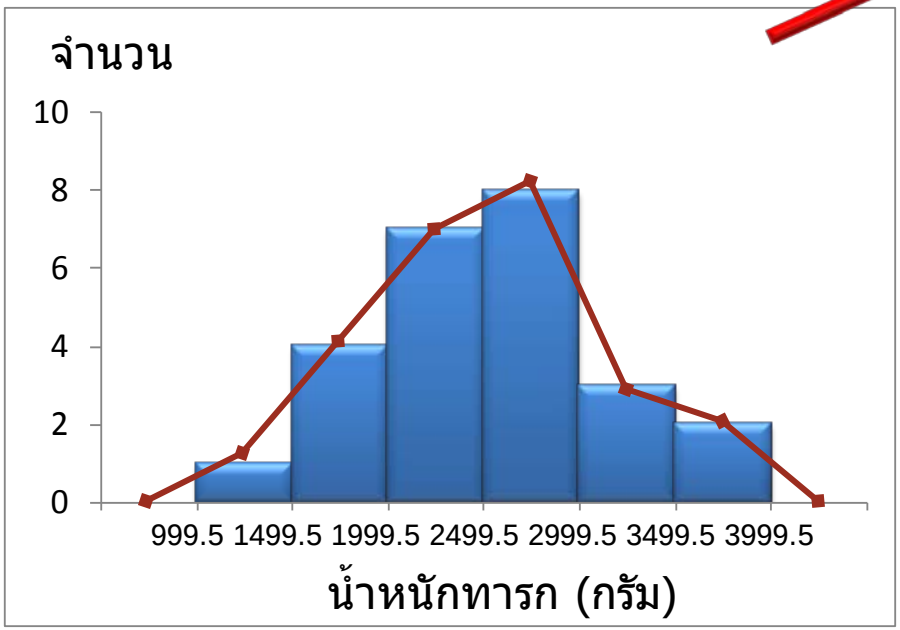
# Histogram

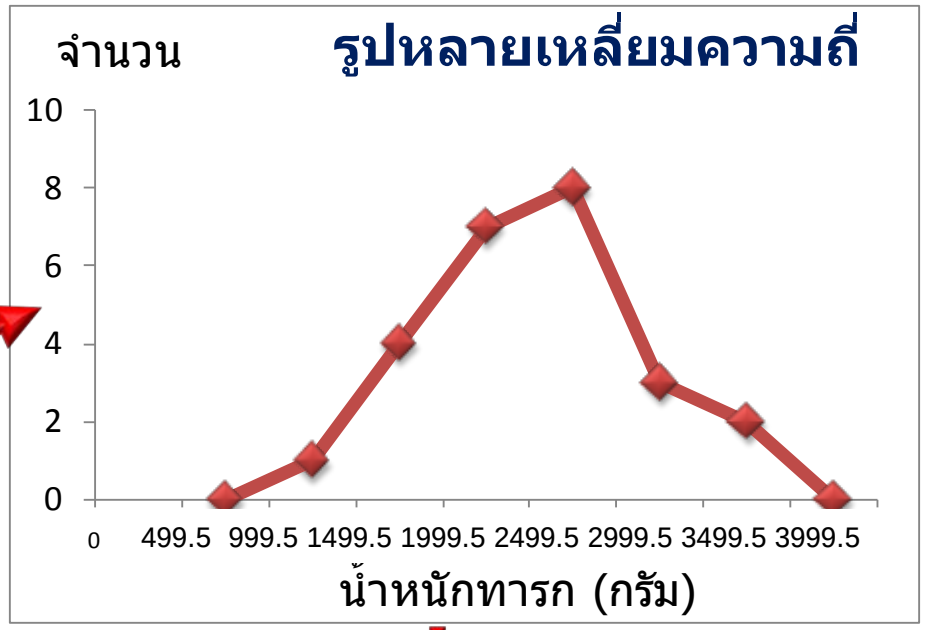
## ตารางแจกแจงความถี่

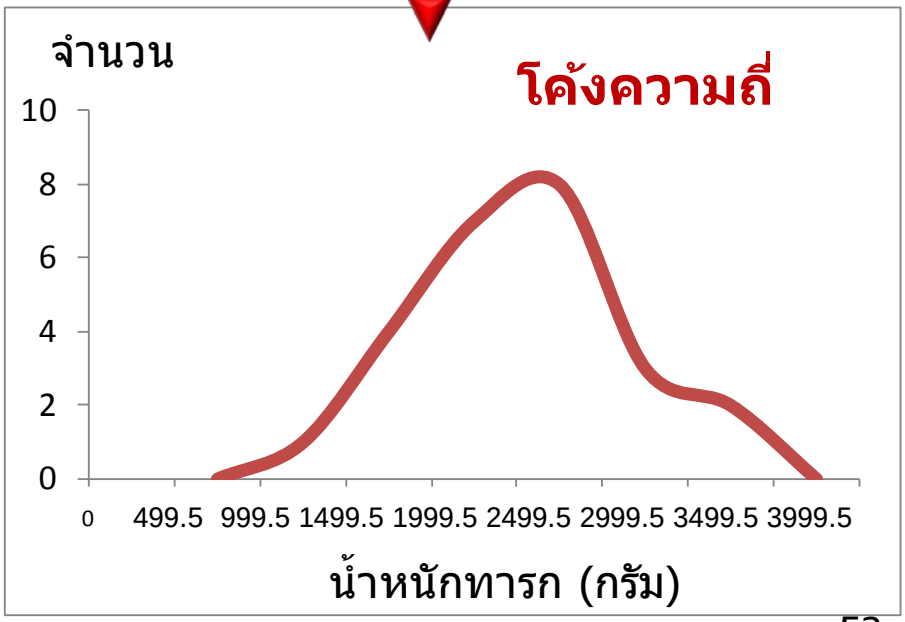
| ขอบเขตชั้น | จำนวน | ร้อยละ |
|---|---|---|
| 999.5 – 1499.5 | 1 | 4 |
| 1499.5 – 1999.5 | 4 | 16 |
| 1999.5 – 2499.5 | 7 | 28 |
| 2499.5 – 2999.5 | 8 | 32 |
| 2999.5 – 3499.5 | 3 | 12 |
| 3499.5 – 3999.5 | 2 | 8 |
| รวม | 25 | 100 |




จำนวน
10
8
6
4
2
0
999.5 1499.5 1999.5 2499.5 2999.5 3499.5 3999.5
น้ำหนักทารก (กรัม)


จำนวน
รูปหลายเหลี่ยมความถี่
10
8
6
4
2
0
0 499.5 999.5 1499.5 1999.5 2499.5 2999.5 3499.5 3999.5
น้ำหนักทารก (กรัม)

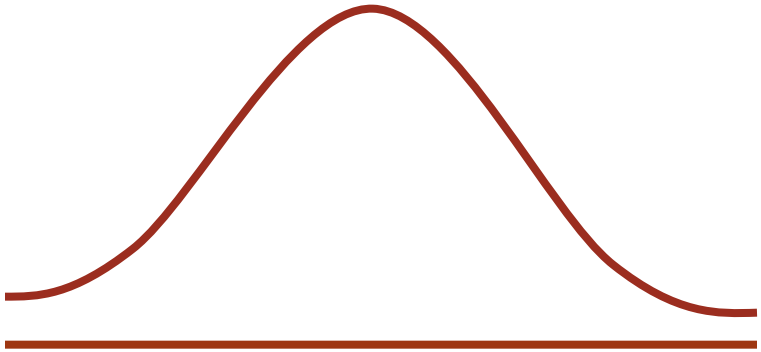

จำนวน
โค้งความถี่
10
8
6
4
2
0
0 499.5 999.5 1499.5 1999.5 2499.5 2999.5 3499.5 3999.5
น้ำหนักทารก (กรัม)

# ลักษณะการแจกแจงจากโค้งความถี่ข้อมูล

- **โค้งรูประฆังคว่ำ**

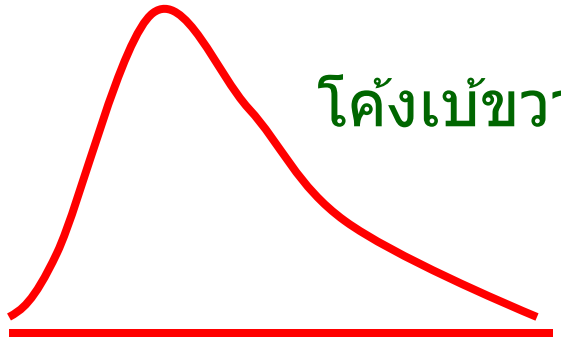
  เป็นลักษณะการแจกแจงของข้อมูลที่พบโดยทั่ว ๆ ไป

โค้งสมมาตร



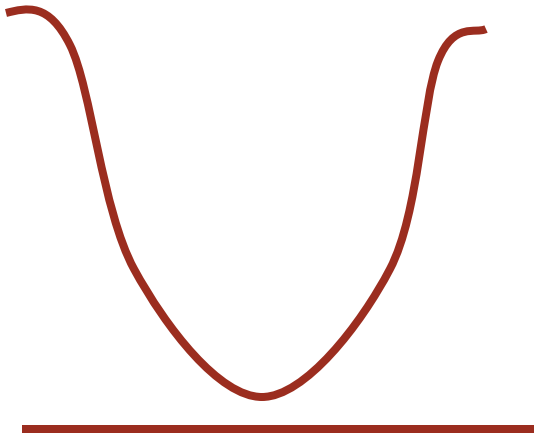
ข้อมูลที่มีค่าปานกลางมีเป็นจำนวนมาก

โค้งเบ้ขวา



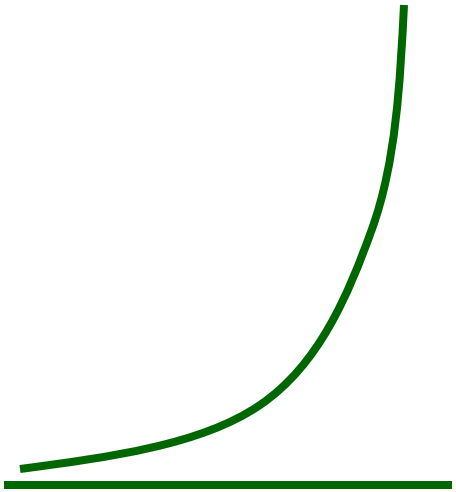
โค้งเบ้ซ้าย

- **โค้งรูปตัวยู**



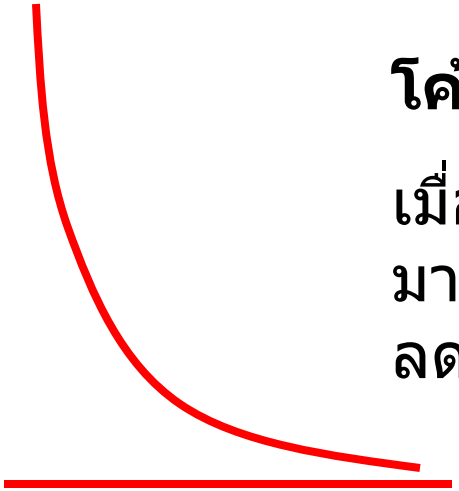
ข้อมูลที่มีค่าปานกลางมีเป็นจำนวนน้อย

- **โค้งรูปตัวเจและเจกลับ**



**โค้งรูปตัวเจ**

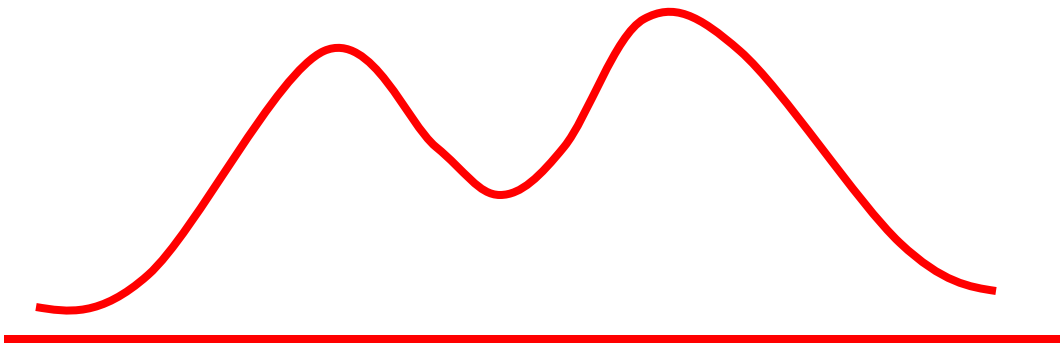
เมื่อข้อมูลมีค่ามากขึ้น ความถี่จะมากขึ้นเช่นกัน



**โค้งรูปตัวเจกลับ**

เมื่อข้อมูลมีค่ามากขึ้น ความถี่จะลดลง

- **โค้งที่มีจุดยอดหลายแห่ง**

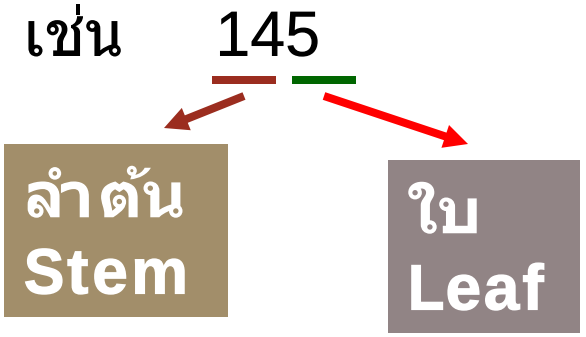


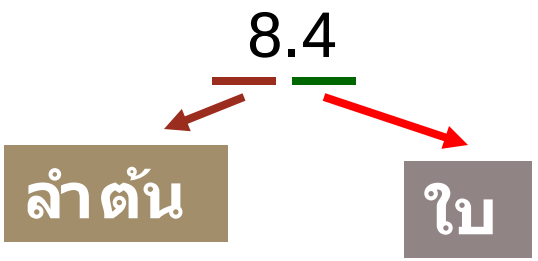
Bi-modal Curve

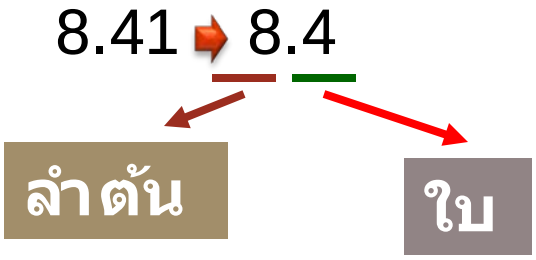
ข้อมูลมีตำแหน่งที่มีความถี่สูง ๆ 2 จุด

# Stem & Leaf Plot

เป็นการนำเสนอสำหรับข้อมูลปริมาณ โดยทำการ plot ค่าข้อมูลแต่ละตัว ด้วยการแบ่งตัวเลขนั้นเป็น 2 ส่วน คือ
- ลำต้น ใช้ตัวเลขส่วนหน้า
- ใบ ใช้ตัวเลขส่วนหลัง




**ขั้นตอนการ plot**

- เรียงค่าข้อมูลจากน้อยไปหามาก
- กำหนดแนวลำต้น เรียงค่าน้อยไปมากในแนวตั้ง (บนลงล่าง)
- นำใบที่มีลำต้นเดียวกัน มาเรียงค่าจากน้อยไปมากในแนวนอน (ซ้ายไปขวา)

# Stem & Leaf Plot

น้ำหนัก (กรัม) ของทารกแรกเกิด 25 คน

| | | | | |
|---|---|---|---|---|
| 3990 | 1250 | 3500 | 1900 | 1500 |
| 2000 | 1900 | 2800 | 1900 | 2400 |
| 2490 | 3200 | 2800 | 2490 | 2400 |
| 2500 | 2400 | 3250 | 2400 | 2800 |
| 2500 | 2800 | 3450 | 2800 | 2500 |

Leaf Unit = 10

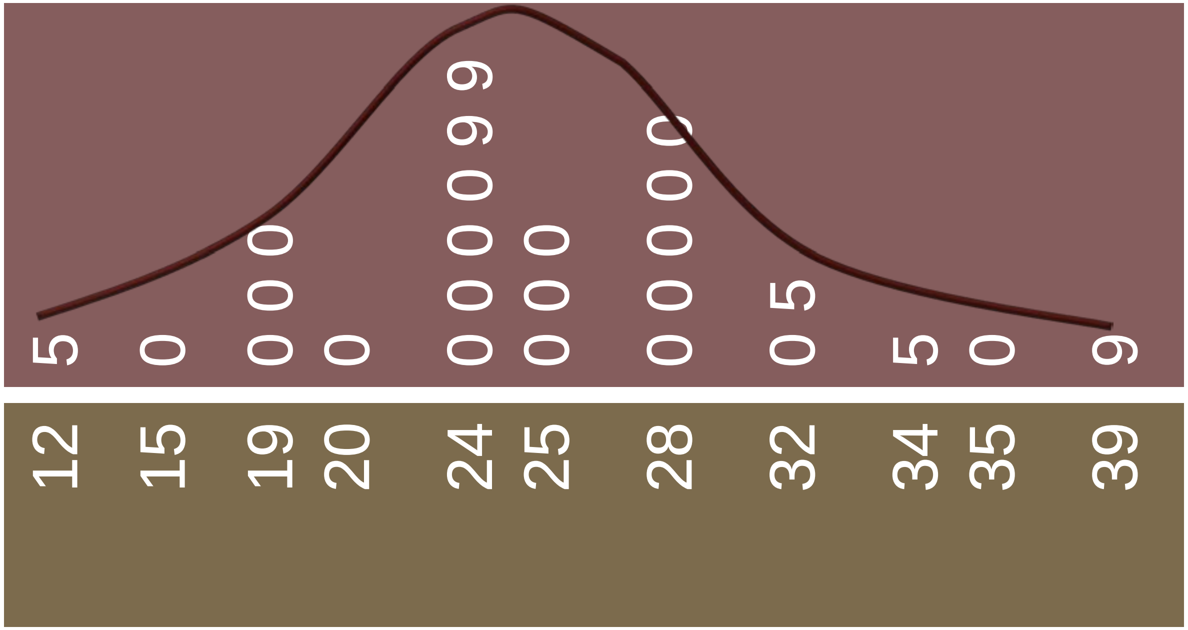
| | | | | |
|---|---|---|---|---|
| 125 | 150 | 190 | 190 | 190 |
| 200 | 240 | 240 | 240 | 240 |
| 249 | 249 | 250 | 250 | 250 |
| 280 | 280 | 280 | 280 | 280 |
| 320 | 325 | 345 | 350 | 399 |

| ลำต้น Stem | ใบ Leaf |
|---|---|
| 12 | 5 |
| 15 | 0 |
| 19 | 0 0 0 |
| 20 | 0 |
| 24 | 0 0 0 0 9 9 |
| 25 | 0 0 0 |
| 28 | 0 0 0 0 0 |
| 32 | 0 5 |
| 34 | 5 |
| 35 | 0 |
| 39 | 9 |

# Stem & Leaf Plot

| ลำต้น | ใบ |
|---|---|
| 12 | 5 |
| 15 | 0 |
| 19 | 0 0 0 |
| 20 | 0 |
| 24 | 0 0 0 0 9 9 |
| 25 | 0 0 0 |
| 28 | 0 0 0 0 0 |
| 32 | 0 5 |
| 34 | 5 |
| 35 | 0 |
| 39 | 9 |

**โค้งความถี่จากแผนภูมิลำต้นและใบ**



# Stem & Leaf Plot

น้ำหนัก (kg) ของทารกแรกเกิด 10 คน

3.9 1.2 3.5 1.9 1.5
2.0 1.9 2.8 1.9 2.4

Leaf Unit = 0.1

12 15 19 19 19
20 24 28 35 39

| ลำต้น Stem | ใบ Leaf |
|---|---|
| 1 | 2 5 9 9 9 |
| 2 | 0 4 8 |
| 3 | 5 9 |

| ลำต้น Stem | ใบ Leaf |
|---|---|
| 1 | 2 |
| 1* | 5 9 9 9 |
| 2 | 0 4 |
| 2* | 8 |
| 3 | 5 9 |

# Stem & Leaf Plot

ผลคะแนนสอบของนักศึกษา

| ลำต้น | ใบ |
|---|---|
| 1 | 2 |
| 1 | 5 |
| 1 | 999 |
| 2 | 0 |
| 2 | 4 |
| 3 | 5 |

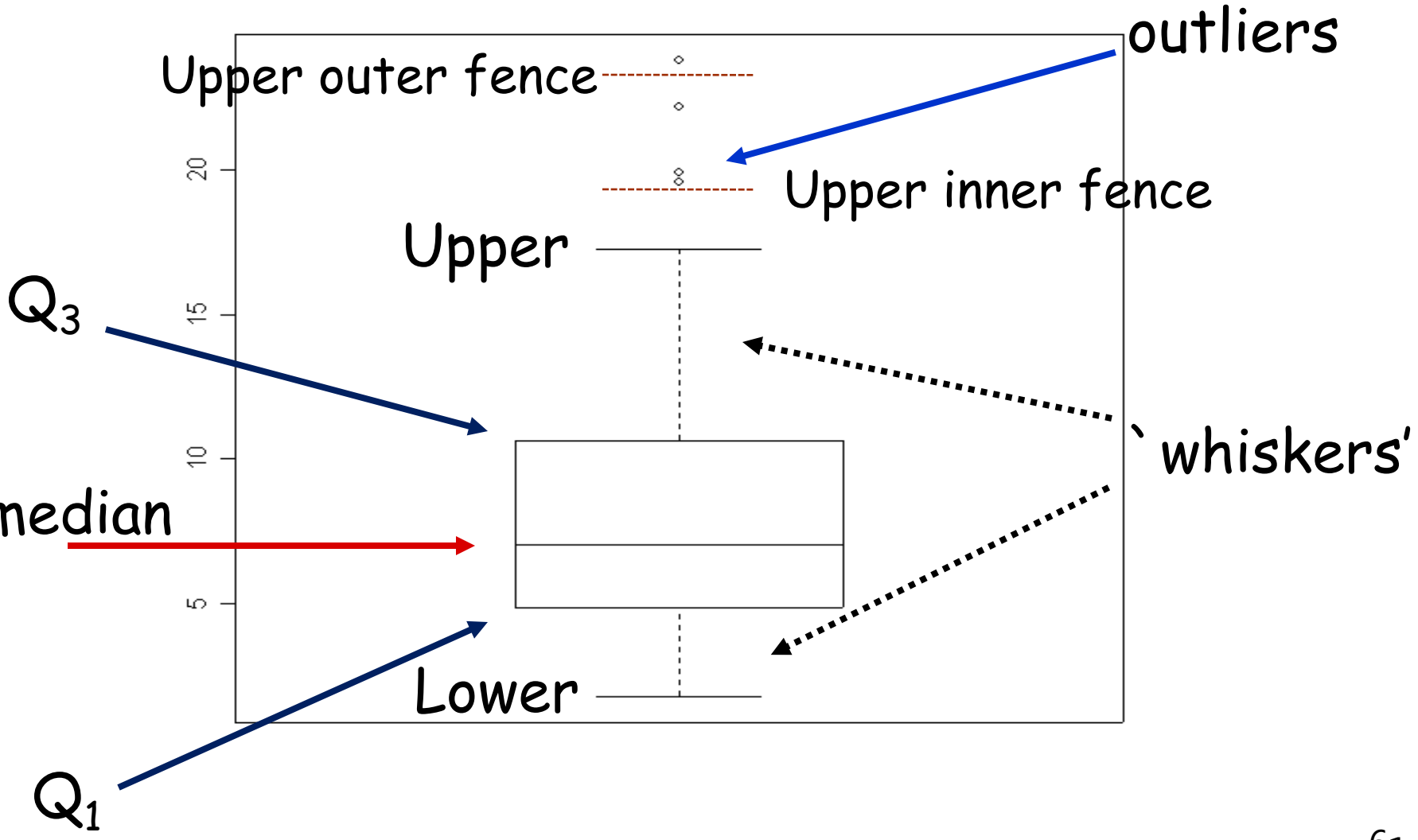
Leaf Unit = 10

จงตอบคำถามต่อไปนี้

ก. จำนวนนศ.ทั้งหมด

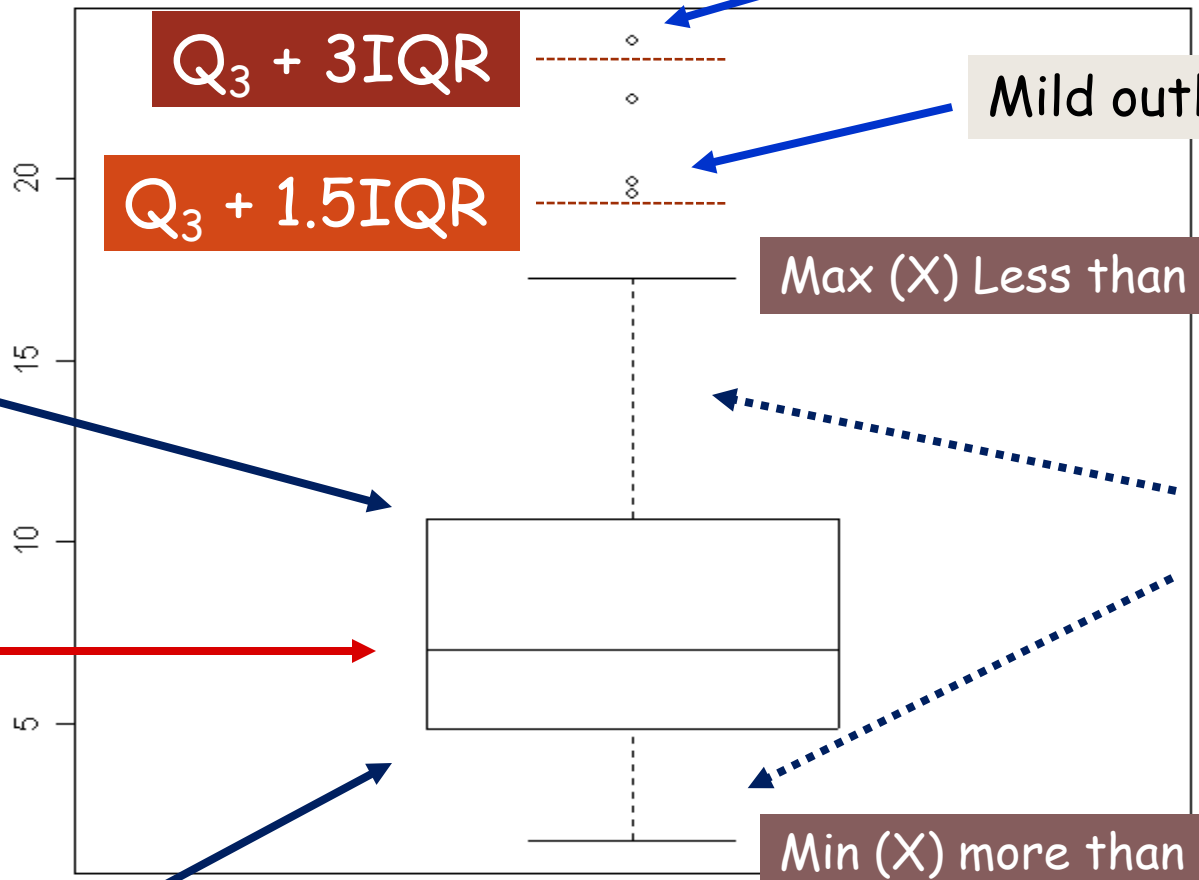
ข. ให้ระบุคะแนนสอบของนักศึกษาทั้งหมด

# Boxplot


outliers
Upper outer fence
Upper inner fence
Upper
$Q_3$
`whiskers'
median
Lower
$Q_1$
20
15
10
5

# ขั้นตอนการสร้าง Boxplot

$IQR = Q_3 - Q_1$



# ตัวอย่าง

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 6 | 24 | 13 | 11 | 34 | 32 | 28 | 19 |
| 30 | 22 | 30 | 58 | 15 | 5 | 17 | 25 |

$Q_1 = 13.5$

$Q_2 = 23$

$Q_3 = 30$

$IQR = 30 - 13.5 = 16.5$

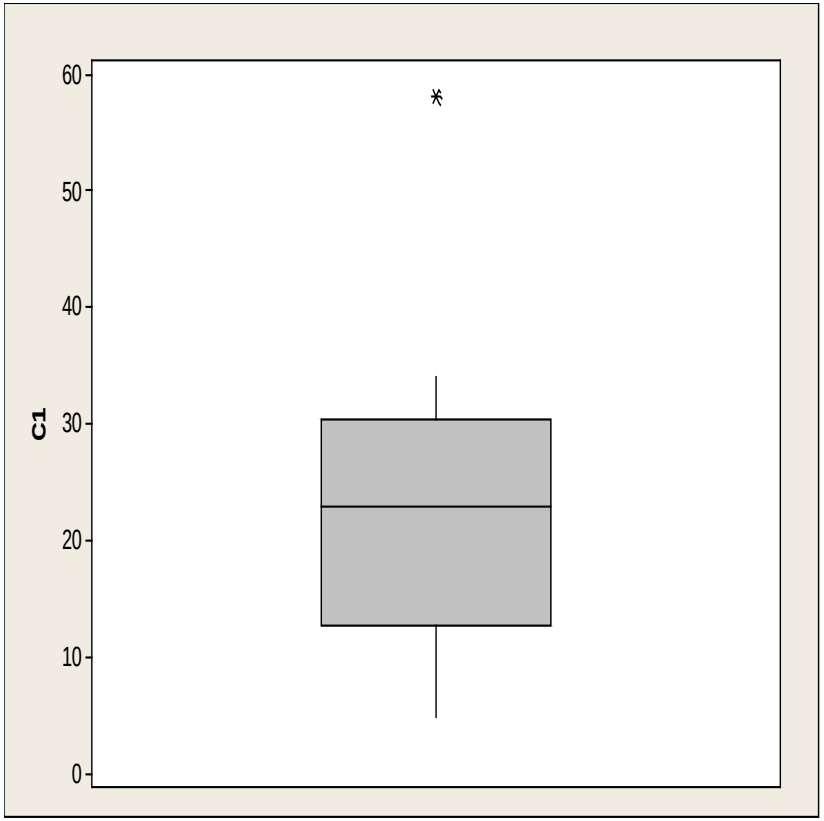
$Q_3 + 1.5IQR = 54.75$

หนวดบน = Max(X) ที่ต่ำกว่า 54.75

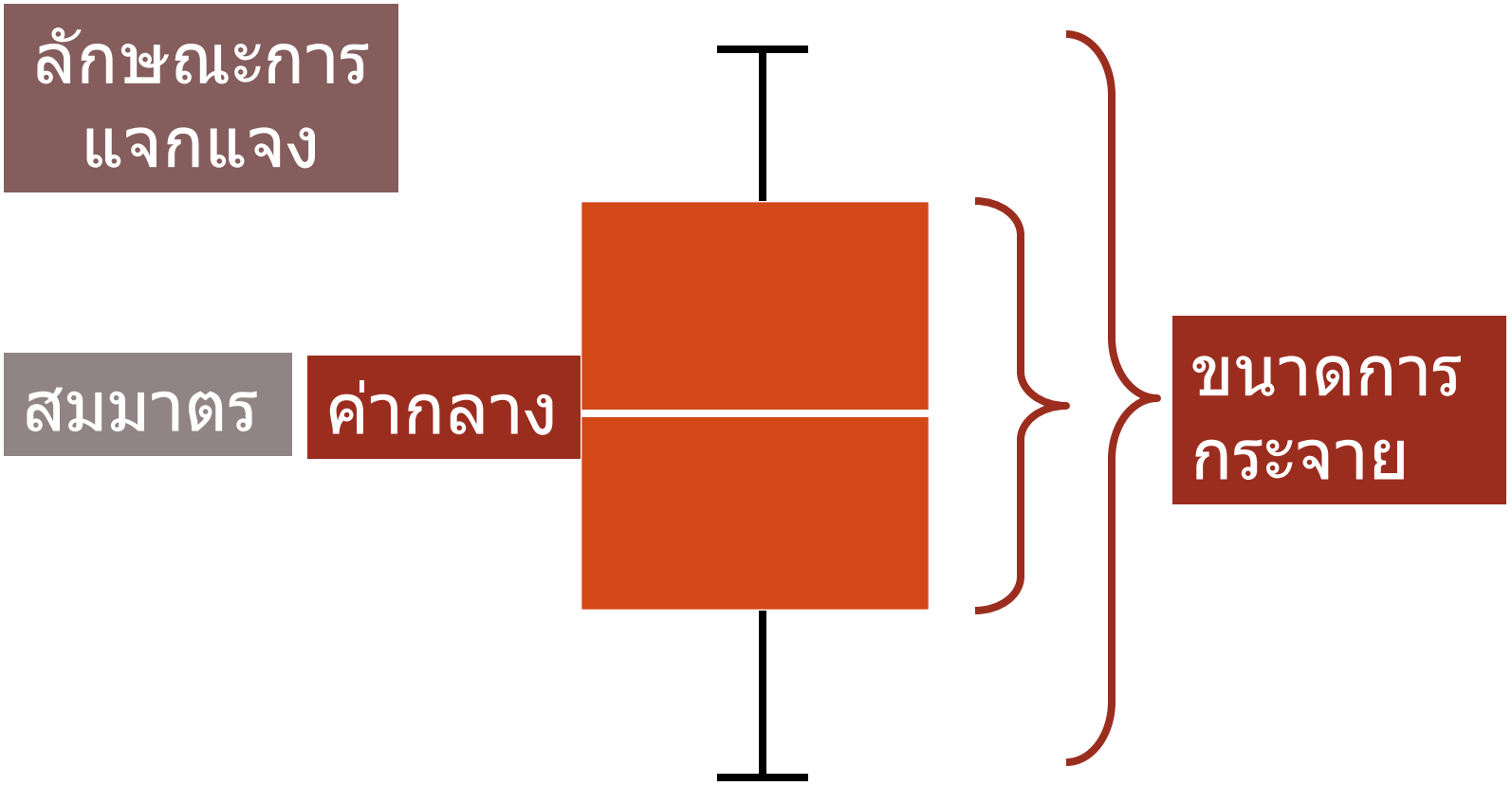
= 34

$Q_1 - 1.5IQR = -11.25$

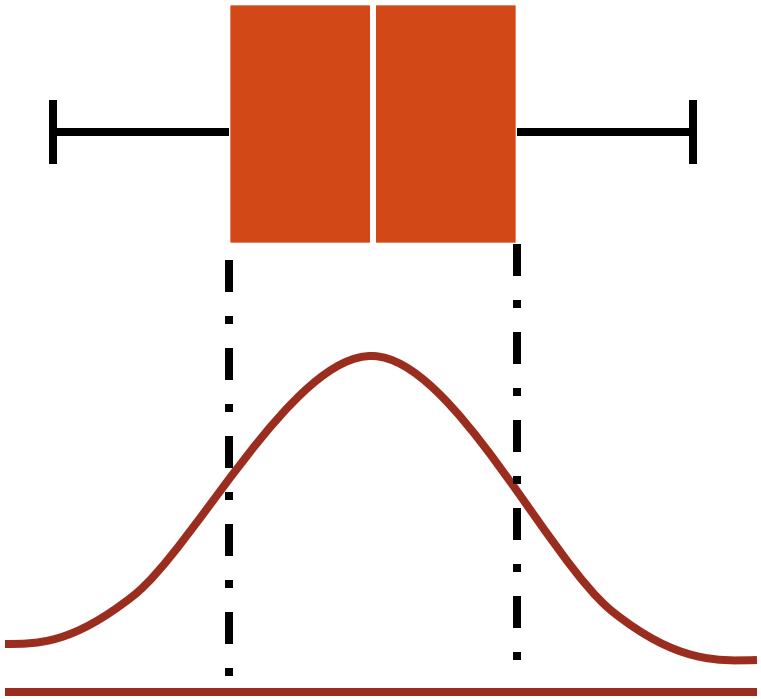
หนวดล่าง = Min(X) ที่สูงกว่า -11.25 = 5

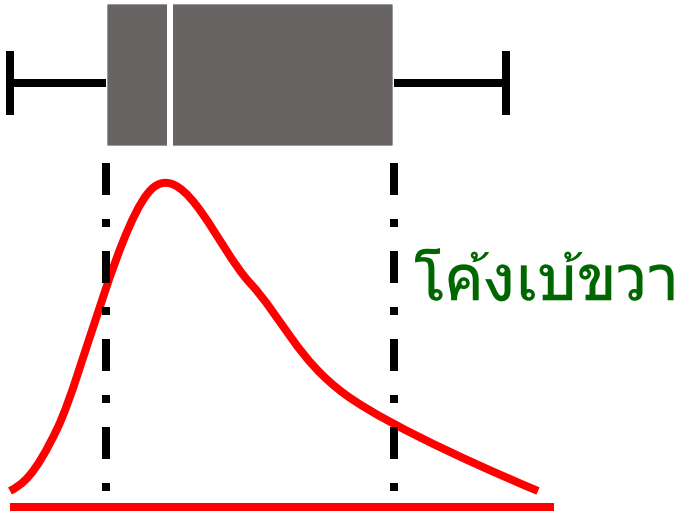


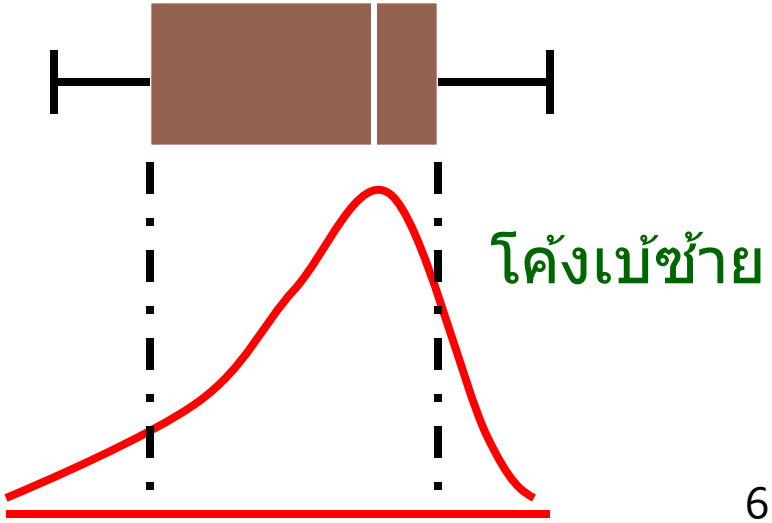
# การอธิบายลักษณะ Box Plot

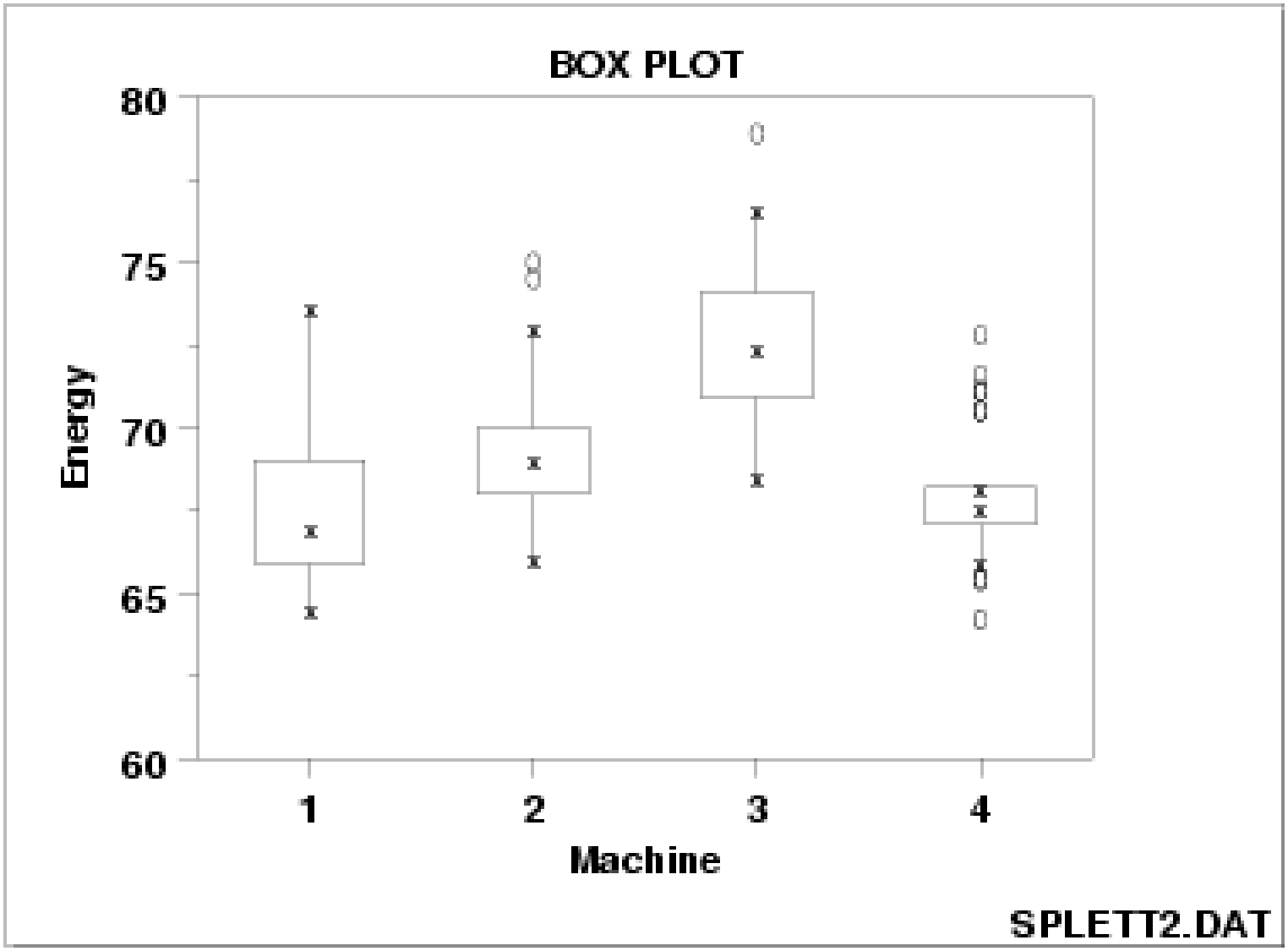


# เปรียบเทียบ Box plot กับโค้งความถี่





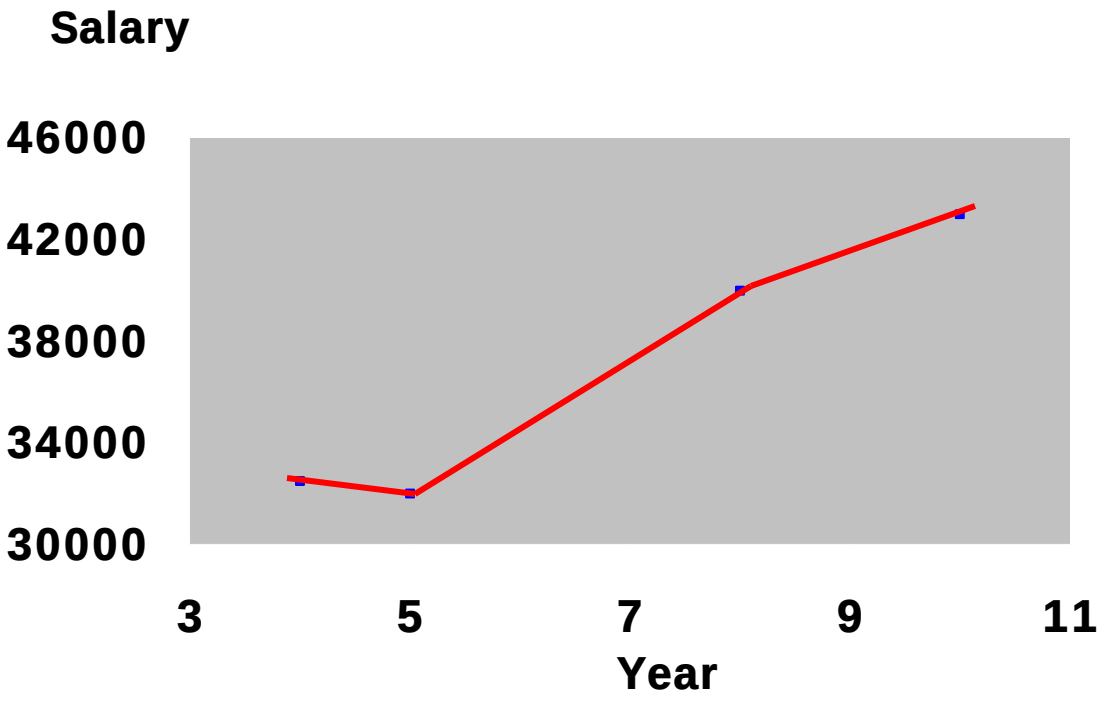




BOX PLOT
80
75
70
65
60
Energy
1
2
3
4
Machine
SPLETT2.DAT

# Graphing Paired Data Sets

- Scatter Plot
- Line Plot

| จำนวนปีที่ทำงาน | รายได้ |
|---|---|
| 4 | 32500 |
| 5 | 32000 |
| 8 | 40000 |
| 10 | 43000 |




Scatter Plot

# ทบทวน

| คนที่ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| คะแนน ($x_i$) | 2 | 4 | 3 | 1 |

จงคำนวณ

1. $\sum_{i=1}^{2} x_i$ 2. $\sum x_i^2$ 3. $\sum (x_i - 1)^2$ 4. $\sum x_i^2 - 10$

$$\sum_{i=1}^{2} x_i = x_1 + x_2 = 2 + 4 = 6$$

$$\sum x_i^2 = 2^2 + 4^2 + 3^2 + 1^2 = 30$$

$$\sum (x_i - 1)^2 = (2-1)^2 + (4-1)^2 + (3-1)^2 + (1-1)^2 = 14$$

$$\sum x_i^2 - 10 = (2^2 + 4^2 + 3^2 + 1^2) - 10 = 20$$

# การวัดแนวโน้มเข้าสู่ส่วนกลาง

- **ค่าเฉลี่ยเลขคณิต** (Arithmetic Mean)

**Population**

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

**Sample**

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Ex จากการชั่งน้ำหนักตัวอย่างนักเรียน 10 คน ได้ผลเป็นดังนี้

30 32 33 34 36 34 33 34 33 36

จงคำนวณค่าเฉลี่ย

$$\sum X_i = 335$$

n = 10

$$\overline{X} = 33.5$$

| 30 | 32 | 33 | 34 | 36 | 34 | 33 | 34 | 33 | 36 |
|---|---|---|---|---|---|---|---|---|---|

- จากข้อมูล สามารถเขียนตารางความถี่ได้เป็น

| $X_i$ | $f_i$ | $f_iX_i$ |
|---|---|---|
| 30 | 1 | 30 |
| 32 | 1 | 32 |
| 33 | 3 | 99 |
| 34 | 3 | 102 |
| 36 | 2 | 72 |
| Tol | 10 | 335 |

สามารถคำนวณค่าเฉลี่ยได้จาก

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n} = \frac{f_1X_1 + f_2X_2 + \ldots + f_kX_k}{n}$$

โดย $$\sum f_i = n = 10$$

$$\sum f_i X_i = 335$$

ดังนั้น $$\overline{X} = \frac{335}{10} = 33.5$$

## ▪ ค่าฐานนิยม (Mode)

คือ ค่าของข้อมูลที่มีความถี่มากที่สุด

เช่น ข้อมูลชุดหนึ่ง 2, 5, 10, 7, 2, 8, 10, 12, 15
สร้างตารางความถี่

| X | ความถี่ |
|---|---|
| 2 | 2 |
| 5 | 1 |
| 7 | 1 |
| 8 | 1 |
| 10 | 2 |
| 12 | 1 |
| 15 | 1 |

ค่าของข้อมูลที่มีความถี่มากสุด คือ

2 และ 10 → ค่าฐานนิยม

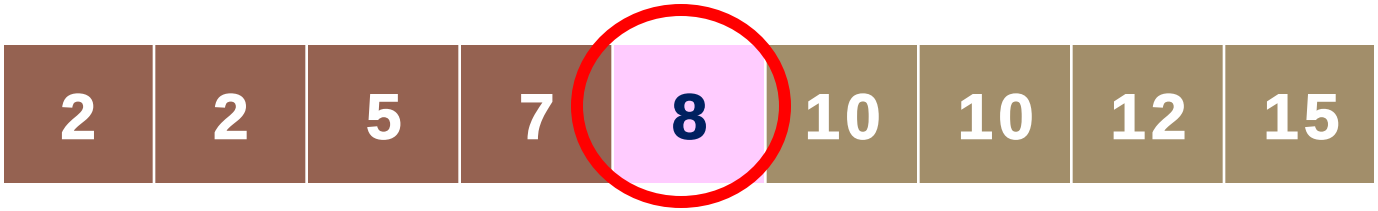
ค่าเฉลี่ย = 7.89

## – ค่ามัธยฐาน (Median)

คือ ค่าข้อมูลที่อยู่ตำแหน่งตรงกลางของข้อมูล ซึ่งจะแบ่งข้อมูลออกเป็น 2 ส่วนเท่า ๆ กัน

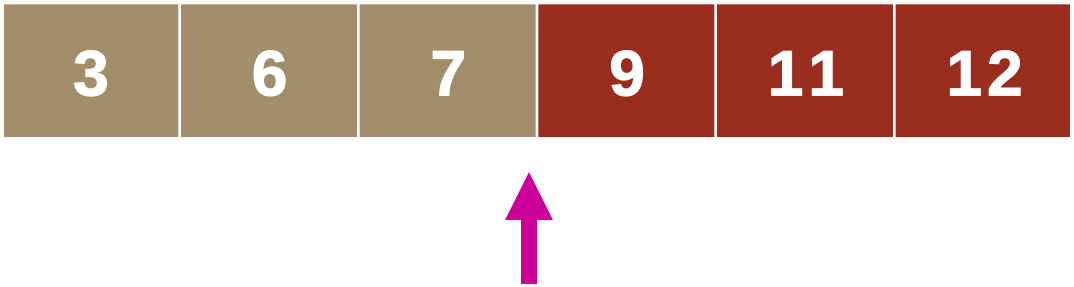
เช่น ข้อมูล 2, 5, 10, 7, 2, 8, 10, 12, 15

หลังจากเรียงลำดับข้อมูลจากน้อยไปหามากแล้ว จะพบว่า




ค่ามัธยฐาน คือ 8 (ค่าของข้อมูลในตำแหน่งที่ 5)

- จากข้อมูล 3, 9, 12, 6, 7, 11 ให้หาค่ามัธยฐาน หลังจากเรียงลำดับข้อมูลจากน้อยไปมาก




ค่ามัธยฐาน คือค่าที่แบ่งข้อมูลเป็น 2 กลุ่ม

ในที่นี้สามารถประมาณค่าได้จาก (7 + 9)/2 = 8

- จากข้อมูลตารางความถี่ จงหาค่ามัธยฐาน

| $x_i$ | $f_i$ |
|---|---|
| 30 | 1 |
| 32 | 1 |
| 33 | 4 |
| 34 | 2 |
| 36 | 2 |
| Tol | 10 |

Med = 33

30 32 33 33 33 33 34 34 36 36

- จากข้อมูลตารางความถี่ จงหาค่ามัธยฐาน

| $x_i$ | $f_i$ |
|---|---|
| 30 | 1 |
| 32 | 1 |
| 33 | 3 |
| 34 | 3 |
| 36 | 2 |
| Tol | 10 |

Med = (33+34)/2

= 33.5

# กรณีข้อมูลอยู่ในรูปตารางแจกแจงความถี่

- การคำนวณค่าแนวโน้มเข้าสู่ส่วนกลาง
  - ค่าเฉลี่ย

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

โดย $X_i$ = จุดกลางชั้นของชั้นที่ $i$

$f_i$ = ความถี่ชั้นที่ $i$

และ $k$ = จำนวนชั้น

  - ค่ามัธยฐาน

$$\text{Med} = L + I\frac{\left[\frac{n}{2} - \sum f_L\right]}{f}$$

โดย $L$ = ขอบเขตล่างของชั้น med

$I$ = ความกว้างของชั้น med

$f$ = ความถี่ชั้น med

และ $\sum f$ = ความถี่สะสมชั้นต่ำกว่า med

  - ค่าฐานนิยม

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

โดย $L$ = ขอบเขตล่างของชั้น mod

$I$ = ความกว้างของชั้น mod

$d_1$ = ผลต่างความถี่ชั้น mod กับชั้นก่อนหน้า

และ $d_2$ = ผลต่างความถี่ชั้น mod กับชั้นถัดไป

# เปรียบเทียบการคำนวณค่า

# กรณีใช้ข้อมูลดิบ กับข้อมูลในรูปตารางแจกแจงความถี่

ผลการสำรวจน้ำหนัก (กรัม) ของทารกแรกเกิด 25 คน เป็นดังนี้

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3990 | 1250 | 3500 | 1900 | 1500 | 2000 | 1900 | 2800 | 1900 |
| 2400 | 2490 | 3200 | 2800 | 2490 | 2400 | 2500 | 2400 | 3250 |
| 2400 | 2800 | 2500 | 2800 | 3450 | 2800 | 2500 | | |

$$\text{Mean} = \frac{3990 + 1250 + \ldots + 2500}{25} = 2574.8$$

Median = 2500

| | | |
|---|---|---|
| 3990 | 1250 | 3500 |
| 1900 | 1500 | 2000 |
| 1900 | 2800 | 1900 |
| 2400 | 2490 | 3200 |
| 2800 | 2490 | 2400 |
| 2500 | 2400 | 3250 |
| 2400 | 3250 | 2500 |
| 2800 | 3450 | 2800 |
| 2500 | | |

# ตารางแจกแจงความถี่

| น้ำหนัก | จำนวน |
|---|---|
| 1000 – 1499 | 1 |
| 1500 – 1999 | 4 |
| 2000 – 2499 | 7 |
| 2500 – 2999 | 8 |
| 3000 – 3499 | 3 |
| 3500 – 3999 | 2 |
| **รวม** | **25** |

| น้ำหนัก | จำนวน(f) | X | f.X |
|---|---|---|---|
| 1000 – 1499 | 1 | 1249.5 | 1249.5 |
| 1500 – 1999 | 4 | 1749.5 | 6998 |
| 2000 – 2499 | 7 | 2249.5 | 15746.5 |
| 2500 – 2999 | 8 | 2749.5 | 21996 |
| 3000 – 3499 | 3 | 3249.5 | 9748.5 |
| 3500 – 3999 | 2 | 3749.5 | 7499 |
| **รวม** | **25** | | **63237.5** |

$$\text{Mean} = 63237.5/25 = 2529.5$$

| น้ำหนัก | จำนวน(f) | F |
|---|---|---|
| 1000 – 1499 | 1 | 1 |
| 1500 – 1999 | 4 | 5 |
| 2000 – 2499 | 7 | 12 |
| 2500 – 2999 | 8 | 20 |
| 3000 – 3499 | 3 | 23 |
| 3500 – 3999 | 2 | 25 |
| **รวม** | **25** | |

## คำนวณค่ามัธยฐาน

1. ตำแหน่งมัธยฐาน = 25/2
2. มัธยฐานตกในชั้นที่ 4
3. ค่าในชั้นที่ 4

   L = 2499.5, I = 500

   f = 8, $\Sigma f_L = 12$

$$\text{Med} = 2499.5 + 500\left(\frac{25/2 - 12}{8}\right) = 2530.75$$

# สรุปเปรียบเทียบระหว่างการคำนวณค่าจากข้อมูลดิบ และตารางแจกแจงความถี่

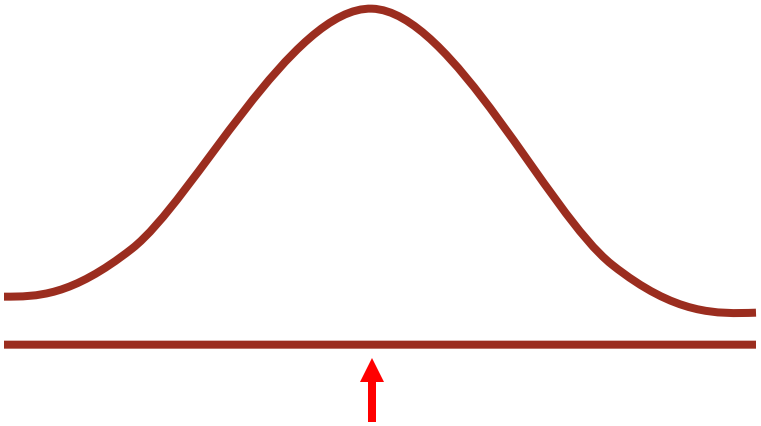
| ค่า | ข้อมูลดิบ | ตารางแจกแจงความถี่ |
|---|---|---|
| Mean | 2574.8 | 2529.5 |
| Median | 2500 | 2530.75 |
| SD | 628.98 | 617.74 |
| Q1 | 2400 | 2088.79 |
| Q3 | 2800 | 2921.38 |

# การเลือกใช้ค่าแนวโน้มเข้าสู่ส่วนกลาง

| เงื่อนไข | Mean | Median | Mode |
|---|---|---|---|
| ข้อมูลปริมาณ<br>- การกระจายสมมาตรหรือใกล้เคียง | ●●● | ●● | ● |
| - การกระจายไม่สมมาตร | ● | ●●● | ●● |
| ข้อมูลคุณภาพ | — | — | ●●● |

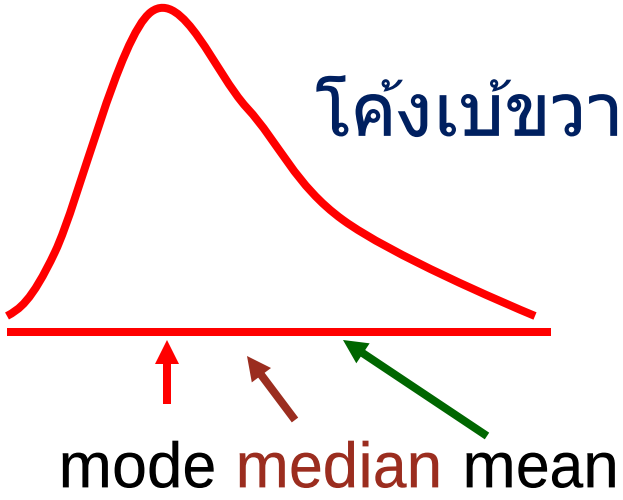
# ความสัมพันธ์ระหว่าง mean, mode และ median

โค้งสมมาตร



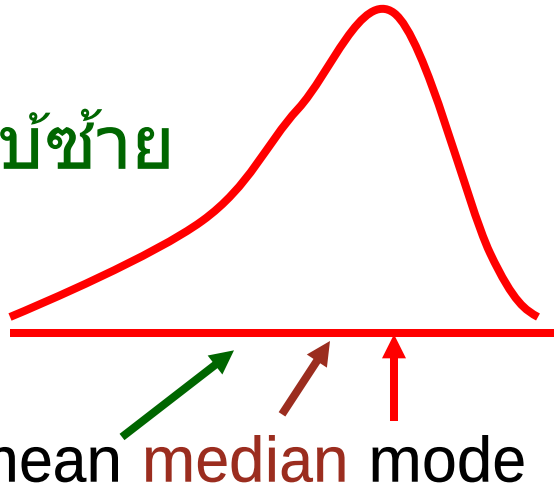
mean
mode
median

โค้งเบ้ขวา



mode median mean

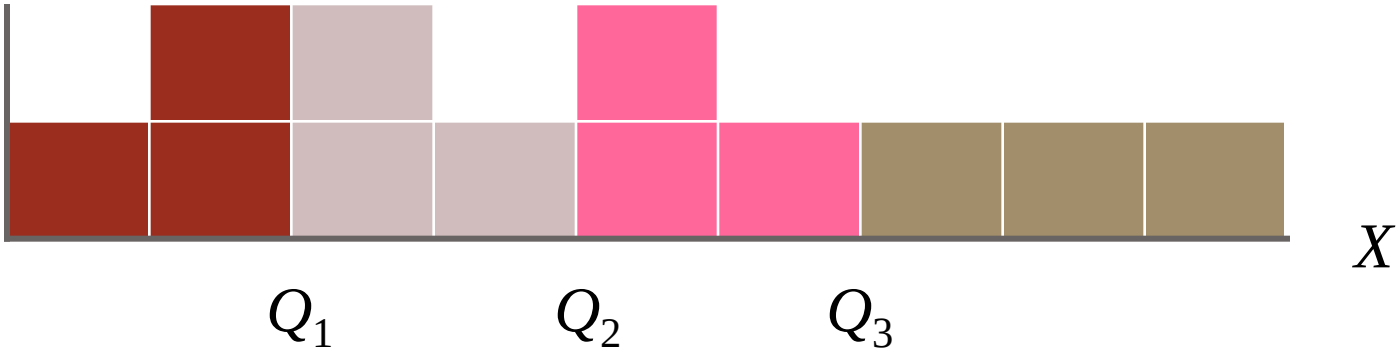
โค้งเบ้ซ้าย



mean median mode

# การวัดตำแหน่งของข้อมูล

- **ควอร์ไทล์ (Quartile)**
  เป็นการแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน




- **เดไซล์ (Decile)** แบ่งข้อมูลเป็น 10 ส่วนเท่า ๆ กัน - $D_1, \ldots, D_9$
- **เปอร์เซ็นไทล์ (Percentile)** แบ่งข้อมูลเป็น 100 ส่วนเท่า ๆ กัน - $P_1, P_2, \ldots, P_{99}$

# การคำนวณค่าตำแหน่งของข้อมูล

- ทำการเรียงลำดับข้อมูลจากน้อยไปหามาก

ค่า $Q_r$ = ค่าของข้อมูล ณ ตำแหน่ง $\frac{r(n+1)}{4}$

ค่า $D_r$ = ค่าของข้อมูล ณ ตำแหน่ง $\frac{r(n+1)}{10}$

ค่า $P_r$ = ค่าของข้อมูล ณ ตำแหน่ง $\frac{r(n+1)}{100}$

เมื่อ $n$ คือจำนวนข้อมูลทั้งหมด

Ex คะแนนวิชาหนึ่งของนักเรียน 12 คน เรียงลำดับได้ดังนี้

| ตำแหน่ง | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 30 | 32 | 43 | 44 | 51 | 55 | 63 | 66 | 70 | 75 | 84 | 86 |

จงหาค่า $Q_3$

จากข้อมูล n = 12

ค่า $Q_3$ คือ ค่าของข้อมูล ณ ตำแหน่ง $\frac{3(12+1)}{4} = 9.75$

ซึ่งอยู่ระหว่างข้อมูลในตำแหน่งที่ 9 และ 10

<u>เทียบบัญญัติไตรยางค์</u>เพื่อหาค่าของข้อมูล ณ ตำแหน่ง 9.75

ตำแหน่งห่างกัน (10-9) = 1 คะแนนห่างกัน (75-70) = 5

ตำแหน่งห่างกัน (9.75-9) = .75 คะแนนห่างกัน = 5 * .75 = 3.75

ดังนั้น $Q_3$ = 70 + 3.75 = 73.75 คะแนน

# กรณีข้อมูลอยู่ในรูปตารางแจกแจงความถี่

- **การคำนวณค่าตำแหน่งของข้อมูล**

ค่าควอร์ไทล์

$$Q_r = L + I\frac{\left[\frac{rn}{4} - \sum f_L\right]}{f}$$

ค่าเดไซล์

$$D_r = L + I\frac{\left[\frac{rn}{10} - \sum f_L\right]}{f}$$

ค่าเปอร์เซ็นไทล์

$$P_r = L + I\frac{\left[\frac{rn}{100} - \sum f_L\right]}{f}$$

โดย $L$ = ขอบเขตล่างของชั้น $Q_r, D_r, P_r$

$I$ = ความกว้างของชั้น $Q_r, D_r, P_r$

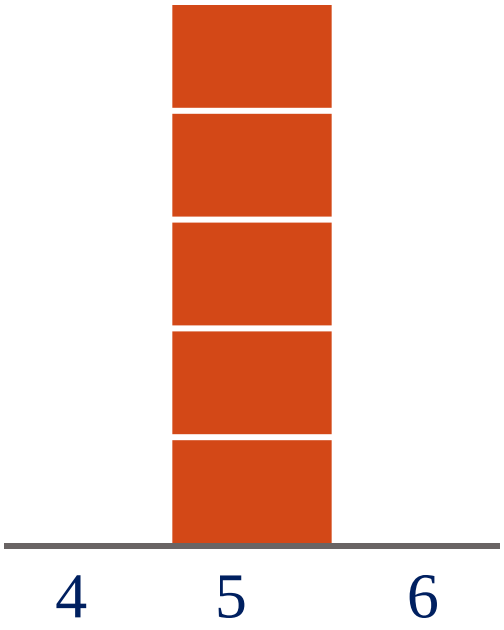
$f$ = ความถี่ชั้น $Q_r, D_r, P_r$

$\sum f_L$ = ความถี่สะสมชั้นต่ำกว่า $Q_r, D_r, P_r$

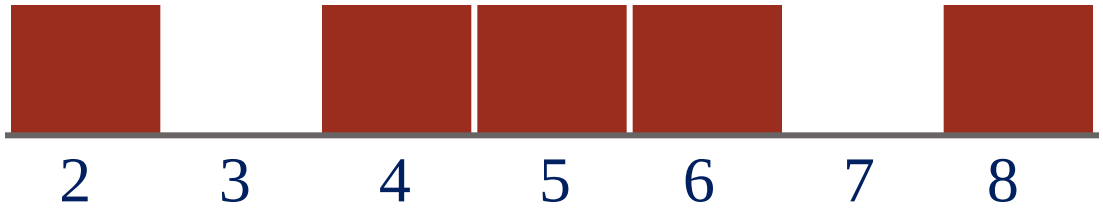
# การวัดการกระจาย

- พิจารณาข้อมูล 2 ชุด ต่อไปนี้

**ข้อมูลชุดที่ 1** Mean = 5




**ข้อมูลชุดที่ 2** Mean = 5



# วิธีการวัดการกระจาย

## 1. พิสัย (Range, R)

R = ค่าสูงสุด - ค่าต่ำสุด

## 2. Quartile Deviation, QD

เป็นค่าวัดการกระจายเพื่อใช้ควบคู่กับค่ามัธยฐาน

$$QD = (Q_3 - Q_1)/2 = (P_{75} - P_{25})/2$$

$Q_3 - Q_1$ = IQR (Interquartile Range)

## 3. ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation, M.D.)

$$M.D. = \frac{\sum_{i=1}^{n} \left| X_i - \bar{X} \right|}{n}$$

## 4. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation, S.D.)

$$S.D. = \sqrt{\text{Variance}} = \sqrt{Var(X)}$$

Variance = ความแปรปรวน

$$Var(X) = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{N} = \frac{\sum_{i=1}^{n} X_i^2 - N\mu^2}{N}$$

Population

$\sigma^2$ (sigma-square)

$$Var(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \left(\sum x\right)^2 / n}{n-1}$$

Sample

$s^2$

**Ex** ให้คำนวณค่าการกระจายของข้อมูลตัวอย่าง 2 ชุดต่อไปนี้

| Sample 1 $X_i$ ($g$) |
|---|
| 1.2 |
| 1.4 |
| 1.6 |
| 1.8 |
| 2.0 |
| 2.2 |
| 2.4 |
| $\sum X_i =$ 12.6 $g$ |

| Sample 2 $X_i$ ($g$) |
|---|
| 1.2 |
| 1.6 |
| 1.7 |
| 1.8 |
| 1.9 |
| 2.0 |
| 2.4 |
| $\sum X_i =$ 12.6 $g$ |

| $X_i$ (g) | $X_i - \bar{X}$ | $\lvert X_i - \bar{X} \rvert$ | $(X_i - \bar{X})^2$ |
|---|---|---|---|
| 1.2 | -0.6 | 0.6 | 0.36 |
| 1.4 | -0.4 | 0.4 | 0.16 |
| 1.6 | -0.2 | 0.2 | 0.04 |
| 1.8 | 0.0 | 0.0 | 0 |
| 2.0 | 0.2 | 0.2 | 0.04 |
| 2.2 | 0.4 | 0.4 | 0.16 |
| 2.4 | 0.6 | 0.6 | 0.36 |
| 12.6 *g* | 0.0 *g* | 2.4 *g* | 1.12 $g^2$ |

$$\bar{X} = \frac{12.6}{7} = 1.8 \ g$$

$$Q_1 = X_{1(7+1)/4} = 1.4 \ g$$

$$Q_3 = X_{3(7+1)/4} = 2.2 \ \text{g}$$

Range = 2.4 – 1.2 = 1.2 *g*

QD. = (2.2 – 1.4)/2 = .4 *g*

**Sample 1**

$$MD = \frac{\sum |X_i - \bar{X}|}{n} = \frac{2.4}{7} = 0.34 \ g$$

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{1.12}{6} = .19 \ g^2$$

$$s = \sqrt{.1867 \ g^2} = .43 \ g$$

| $X_i$ (g) | $X_i - \bar{X}$ | $\lvert X_i - \bar{X} \rvert$ | $(X_i - \bar{X})^2$ |
|---|---|---|---|
| 1.2 | -0.6 | 0.6 | 0.36 |
| 1.6 | -0.2 | 0.2 | 0.04 |
| 1.7 | -0.1 | 0.1 | 0.01 |
| 1.8 | 0.0 | 0.0 | 0 |
| 1.9 | 0.1 | 0.1 | 0.01 |
| 2.0 | 0.2 | 0.2 | 0.04 |
| 2.4 | 0.6 | 0.6 | 0.36 |
| 12.6 *g* | 0.0 *g* | 1.8 *g* | 0.82 $g^2$ |

**Sample 2**

$$\bar{X} = \frac{12.6}{7} = 1.8 \ g$$

$Q_1 = X_{1(7+1)/4} = 1.6\ g$

$Q_3 = X_{3(7+1)/4} = 2.0$ g

Range = 2.4 – 1.2 = 1.2 *g*

QD. = (2.0 – 1.6)/2 = .2 *g*

M.D. = 1.8/7 = 0.26 *g*

$s^2 = 0.82/6 = .1367\ g^2$

$s = \sqrt{.1367\ g^2} = .37\ g$

| | Sample I | Sample II |
|---|---|---|
| Mean | 1.8 | 1.8 |
| Median | 1.8 | 1.8 |
| Q1 | 1.4 | 1.6 |
| Q3 | 2.2 | 2.0 |
| Range | 1.2 | 1.2 |
| QD | 0.4 | 0.2 |
| MD | 0.34 | 0.26 |
| SD | 0.43 | 0.37 |

# กรณีข้อมูลอยู่ในรูปตารางแจกแจงความถี่

- **การคำนวณค่าการกระจาย**

ค่าพิสัย = ขอบเขตบนชั้นสูงสุด - ขอบเขตล่างชั้นต่ำสุด

ค่าส่วนเบี่ยงเบนเฉลี่ย

$$MD = \frac{\sum f_i \left| X_i - \overline{X} \right|}{n}$$

ค่าส่วนเบี่ยงเบนมาตรฐาน

$$s = \sqrt{\frac{\sum f_i (X_i - \overline{X})^2}{n-1}}$$

โดย $X_i$ = จุดกลางของชั้น $i$

$f$ = ความถี่ชั้น $i$

| คะแนน | ความถี่ |
|---|---|
| 10 – 20 | $f_1$ |
| ... | ... |

## ค่าส่วนเบี่ยงเบนเฉลี่ย

$$MD = \frac{\sum f_i \left| X_i - \overline{X} \right|}{n}$$

## ค่าส่วนเบี่ยงเบนมาตรฐาน

$$s = \sqrt{\frac{\sum f_i (X_i - \overline{X})^2}{n-1}}$$

| $X_i$ | $f_i$ | $f_i (X_i - 33.5)^2$ |
|---|---|---|
| 30 | 1 | 12.25 |
| 32 | 1 | 2.25 |
| 33 | 3 | 0.75 |
| 34 | 3 | 0.75 |
| 36 | 2 | 12.5 |
| | 10 | 28.5 |

สามารถคำนวณ SD ได้จาก

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

$$= \frac{28.5}{9} = 3.167$$

s = 1.780

# Statistics Defined

| ค่าแนวโน้มเข้าสู่ส่วนกลาง | ค่าการกระจาย |
|---|---|
| Median | Quartile deviation |
| Percentile | Range |
| Arithmetic mean | Standard deviation |
| Mode | - |

# กรณีข้อมูล k ชุด

ข้อมูลชุดที่ 1 มี mean = $\overline{X}_1$ variance = $S_1^2$

ข้อมูลชุดที่ 2 มี mean = $\overline{X}_2$ variance = $S_2^2$

...

ข้อมูลชุดที่ k มี mean = $\overline{X}_k$ variance = $S_k^2$

- ค่าเฉลี่ยรวมของข้อมูล k ชุด คือ $\overline{X}_c = \sum n_i \overline{X}_i \Big/ \sum n_i$

- ค่าความแปรปรวนรวมของข้อมูล k ชุด คือ

$$S_c^2 = \frac{\sum n_i (S_i^2 + \overline{X}_i^2)}{\sum n_i} - \overline{X}_c^2$$

# ตัวอย่าง 12

- ตำบลแห่งหนึ่งประกอบด้วย 3 หมู่บ้าน คือ

| | หมู่บ้าน ก | หมู่บ้าน ข | หมู่บ้าน ค |
|---|---|---|---|
| จำนวนครัวเรือน | 150 | 80 | 100 |
| รายได้เฉลี่ยต่อปีต่อครัวเรือน | 10000 | 8000 | 13000 |
| ส่วนเบี่ยงเบนมาตรฐานรายได้ | 1200 | 900 | 1000 |

จงหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของรายได้ในตำบลดังกล่าว

$$\overline{X}_c = \frac{(150\times10000)+(80\times8000)+(100\times13000)}{(150+80+100)} = 10424.24$$

$$S_c = \sqrt{\frac{150(1200^2+10000^2)+\ldots+100(1000^2+13000^2)}{150+80+100} - 10424.24^2} = 2161.24$$

# ข้อสังเกต : ค่าเฉลี่ย และความแปรปรวน

| ข้อมูล | Mean | Variance |
|---|---|---|
| $X_1, X_2, \ldots, X_N$ | $\overline{X}$ | $S^2$ |
| $X_1 \pm a, X_2 \pm a, \ldots, X_N \pm a$ | $\overline{X} \pm a$ | $S^2$ |
| $b.X_1, b.X_2, \ldots, b.X_N$ | $b.\overline{X}$ | $b^2.S^2$ |

# การกระจายสัมพัทธ์ (Relative Dispersion)

- เป็นการคำนวณค่าสัมประสิทธิ์การกระจาย (Coefficient of Dispersion) ที่ทำการปรับหน่วยของข้อมูลออกไป เพื่อใช้ในการเปรียบเทียบการกระจายของข้อมูล 2 ชุดขึ้นไป

| สัมประสิทธิ์การกระจาย | สูตรคำนวณ |
|---|---|
| สัมประสิทธิ์พิสัย (Coefficient of range, Coef. R) | $\frac{X_{\max} - X_{\min}}{X_{\max} + X_{\min}}$ |
| สัมประสิทธิ์ส่วนเบี่ยงเบนควอไทล์ (Coef. QD.) | $\frac{Q_3 - Q_1}{Q_3 + Q_1}$ |
| สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ย (Coef. MD.) | $\frac{M.D.}{\overline{X}}$ |
| สัมประสิทธิ์ส่วนเบี่ยงเบนมาตรฐาน (Coef. SD. หรือ CV.) | $\frac{S.D.}{\overline{X}}$ |

| Sample 1 | $X_i$ (g) |
|---|---|
| | 1.2 |
| | 1.4 |
| | 1.6 |
| | 1.8 |
| | 2.0 |
| | 2.2 |
| | 2.4 |
| Mean | 1.8 |
| SD | 0.43 |

CV = 0.43/1.8 = 0.2389

| Sample 2 | $X_i$ (g) |
|---|---|
| | 1.2 |
| | 1.6 |
| | 1.7 |
| | 1.8 |
| | 1.9 |
| | 2.0 |
| | 2.4 |
| Mean | 1.8 |
| SD | 0.37 |

CV = 0.37/1.8 = 0.2056

# การวัดความเบ้ (Measure of Skewness)

- คำนวณสัมประสิทธิ์ความเบ้ เพื่อใช้ในการพิจารณาลักษณะการแจกแจงของข้อมูล

| สปส. ความเบ้ | สูตรคำนวณ | | เกณฑ์พิจารณา |
|---|---|---|---|
| โดยวิธี Karl Pearson | $\frac{Mean - Mode}{S.D.}$ | $\frac{3(Mean - Median)}{S.D.}$ | สปส. ความเบ้<br>= 0 สมมาตร<br>> 0 เบ้ขวา<br>< 0 เบ้ซ้าย |
| โดยวิธี Bowley | $\frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$ | | |
| โดยวิธีเปอร์เซนต์ไทล์ | $\frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{10}}$ | | |

# การวัดความโด่ง (Measure of Kurtosis)

- คำนวณค่าสปส. ความโด่ง เพื่อพิจารณาลักษณะการแจกแจงของข้อมูลที่มีโค้งความถี่แบบสมมาตร ว่าลักษณะความโด่งของเส้นโค้งเป็นปกติหรือไม่

| สปส.ความโด่ง | เกณฑ์พิจารณา |
|---|---|
| $\frac{Q.D.}{P_{90} - P_{10}}$ | สปส.ความโด่ง<br>= .263 ปกติ<br>< .263 ต่ำกว่าปกติ<br>> .263 สูงกว่าปกติ |
| $\frac{(Q_3 - Q_1)}{2(P_{90} - P_{10})}$ | |

# Ex ให้พิจารณาลักษณะความเบ้และความโด่งของข้อมูลตัวอย่าง 2 ชุดต่อไปนี้

| Sample 1 | $X_i$ ($g$) |
|---|---|
| | 1.2 |
| | 1.4 |
| | 1.6 |
| | 1.8 |
| | 2.0 |
| | 2.2 |
| | 2.4 |
| | 12.6 $g$ |

| Sample 2 | $X_i$ ($g$) |
|---|---|
| | 1.2 |
| | 1.6 |
| | 1.7 |
| | 1.8 |
| | 1.9 |
| | 2.0 |
| | 2.4 |
| | 12.6 $g$ |

## Sample 1

| $X_i$ (*g*) |
|---|
| 1.2 |
| 1.4 |
| 1.6 |
| 1.8 |
| 2.0 |
| 2.2 |
| 2.4 |
| 12.6 *g* |

| | |
|---|---|
| Mean | 1.8 |
| Median | 1.8 |
| Q1 | 1.4 |
| Q3 | 2.2 |

**สปส. ความเบ้**

$$= \frac{[(2.2-1.8)-(1.8-1.4)]}{(2.2-1.4)}$$

$= 0$

**สปส. ความโด่ง**

$$= \frac{(2.2 - 1.4)}{2(2.4 - 1.2)}$$

$= 0.33$

เนื่องจาก $P_{90}$, $P_{10}$ หาค่าไม่ได้

$P_{90} = X_{(.9*8)} = X_{(7.2)}$ จึงประมาณ ~ $X_{(7)} = 2.4$

## Sample 2

| $X_i$ ($g$) |
|---|
| 1.2 |
| 1.6 |
| 1.7 |
| 1.8 |
| 1.9 |
| 2.0 |
| 2.4 |
| 12.6 $g$ |

| | |
|---|---|
| Mean | 1.8 |
| Median | 1.8 |
| Q1 | 1.6 |
| Q3 | 2.0 |

สปส. ความเบ้

$$= \frac{[(2.0-1.8)-(1.8-1.6)]}{(2.0-1.8)}$$

$= 0$

สปส. ความโด่ง

= ...

# กรณีข้อมูลหลายตัวแปร

| Case | Age | Height | Sex | Therapy | Blood Pressure | |
|---|---|---|---|---|---|---|
| | | | | | Before T. | After T. |
| 1 | 17 | 158 | 1(male) | 1(drug) | 90 | 110 |
| 2 | 28 | 165 | 2(female) | 1 | 100 | 110 |
| 3 | 20 | 170 | 1 | 1 | 85 | 120 |
| 4 | 25 | 172 | 2 | 1 | 110 | 140 |
| 5 | 32 | 179 | 1 | 1 | 100 | 105 |
| 6 | 40 | 168 | 1 | 2(operation) | 102 | 124 |
| 7 | 45 | 175 | 1 | 2 | 99 | 108 |
| 8 | 35 | 176 | 1 | 2 | 105 | 135 |
| 9 | 29 | 180 | 2 | 2 | 103 | 145 |
| 10 | 36 | 174 | 2 | 1 | 85 | 121 |
| 11 | 24 | 160 | 2 | 1 | 78 | 109 |
| 12 | 27 | 181 | 2 | 1 | 85 | 140 |
| 13 | 26 | 170 | 1 | 2 | 111 | 138 |
| 14 | 33 | 169 | 2 | 2 | 102 | 132 |
| 15 | 42 | 178 | 2 | 2 | 98 | 107 |

# การนำเสนอด้วยตาราง

ตาราง ลักษณะเบื้องต้นของผู้ป่วย 15 คน

| ลักษณะผู้ป่วย | |
|---|---|
| - อายุ (ปี)[a] | 30.6±8.01 |
| - ส่วนสูง (cm) [a] | 171.7±6.1 |
| -เพศ [b]<br>ชาย<br>หญิง | <br>7 (47)<br>8 (53) |
| -วิธีการรักษา [b]<br>ยา<br>ผ่าตัด | <br>8 (53)<br>7 (47) |

[a] mean±SD; [b] n (%)

| ลักษณะ | รักษาด้วยยา | รักษาด้วยการผ่าตัด |
|---|---|---|
| - อายุ (ปี)[a] | 26.1±6.1 | 35.7±7.0 |
| - ส่วนสูง (cm) [a] | 169.9±8.4 | 173.7±4.7 |
| - เพศ [b] | | |
| ชาย | 38 | 57 |
| หญิง | 62 | 43 |
| - ความดันโลหิต [a] | | |
| ก่อนการรักษา | 91.6±10.7 | 102.9±4.3 |
| หลังการรักษา | 119.4±13.9 | 127.0±14.7 |

# ทดลองทำ 1

- ข้อใดคือข้อมูลสถิติ
  1. หมายเลขโทรศัพท์ในสมุดโทรศัพท์หน้าเหลือง
  2. อุณหภูมิสูงสุดในแต่ละวันประจำเดือนตุลาคม
  3. ร้อยละของคะแนนสอบในชั้นเรียนแห่งหนึ่ง
  4. หมายเลขของผู้เล่นแต่ละคนในทีมฟุตบอล

- สำหรับข้อที่คิดว่าเป็นข้อมูลสถิติ ให้จำแนกข้อมูลตามชนิด (ข้อมูลคุณภาพ/ข้อมูลปริมาณ)

# ทดลองทำ 2

- **ให้ระบุประชากร กลุ่มตัวอย่าง หน่วยตัวอย่าง (หน่วยทดลอง) และค่าสังเกตที่สนใจ ในแต่ละข้อ**
  1. จากการสำรวจเจ้าของนาฬิกายี่ห้อหนึ่งจำนวน 898 คน พบว่า มี 16% เป็นเจ้าของรุ่นเรือนทอง
  2. พยาบาล 38 คนทำงานในเขต จ.เชียงใหม่ ถูกสอบถามถึงความคิดเห็นในการดูแลรักษาผู้ป่วย
  3. จากการสอบถามผู้ที่มีบัตร ATM พบว่า ค่าธรรมเนียมการใช้บริการที่เหมาะสมคือ 100 บาท/ปี
  4. ทำการสำรวจน.ศ. คณะวิศวะฯ มอชอ 1420 คน ถึงหนังสืออ่านเล่นภาษาอังกฤษเล่มโปรด

# ทดลองทำ 3

- **จงเปรียบเทียบรายได้ต่อเดือน (หน่วย 1000 บาท) ของพนักงานบริษัทใน จ.กรุงเทพ และ จ.ชลบุรี**

  กรุงเทพ 34.9 25.7 17.3 16.8 26.8 24.7 29.4 32.7 25.5

  ชลบุรี 25.6 23.2 26.7 27.7 25.4 26.4 18.3 26.1 31.3

# Stem & Leaf Plot

กรุงเทพ N = 9
Leaf Unit = 0.1

LO 168, 173

| | |
|---|---|
| 24 | 7 |
| 25 | 57 |
| 26 | 8 |
| 29 | 4 |
| 32 | 7 |
| 34 | 9 |

สปส. ความเบ้ = - 0.26

ชลบุรี N = 9
Leaf Unit = 0.1

LO 183, 232

| | |
|---|---|
| 25 | 4 |
| 25 | 6 |
| 26 | 14 |
| 26 | 7 |
| 27 | |
| 27 | 7 |

สปส. ความเบ้ = - 0.81

# เปรียบเทียบรายได้

- กรุงเทพ
  - มัธยฐาน = 25.70 พันบาท/เดือน
  - ส่วนเบี่ยงเบนควอไทล์ = 5.02
  - สปส.ส่วนเบี่ยงเบนควอไทล์ = 0.19

- ชลบุรี
  - มัธยฐาน = 26.10 พันบาท/เดือน
  - ส่วนเบี่ยงเบนควอไทล์ = 1.45
  - สปส.ส่วนเบี่ยงเบนควอไทล์ = 0.06