

## บทที่ 7

### สถิติที่ใช้ในการวิเคราะห์ข้อมูล

#### บทนำ

สถิติ (statistics) เป็นเรื่องที่อยู่คู่กับมวลมนุษยเรามาตั้งแต่เกิดทั้งที่เรารู้ตัวและไม่รู้ตัว อนึ่ง ผู้เขียนมีความเชื่อว่าหลายคนอาจจะเบื่อนหน้าหนังสือสถิติ (statistics) นั้นเพราะวิชาสถิติ มักจะเต็มไปด้วยตัวเลข (datum) และมากด้วยข้อมูล (data) เพื่อแสดงให้เห็นถึงความชัดเจนและความแตกต่าง ซึ่งหากสังเกตรอบตัวเราจะเห็นได้ว่าการใช้สถิติตั้งแต่เรลั้มตามาตุโลก เช่น การจดบันทึกน้ำหนัก (weight) ส่วนสูง (height) เมื่อแรกคลอด ตลอดจนการจดบันทึกข้อมูลทุกเดือนเพื่อเฝ้าดูการเจริญเติบโตของร่างกาย เป็นต้น สถิติ (statistics) เป็นเครื่องมือสำคัญที่ถูกนำมาใช้ในทุองค์กร สถาบัน หน่วยงาน ครั้วเรือนหรือแม้แต่ตัวเราเองเพื่อทำการจดเก็บหรือบันทึก (record) ข้อมูลต่าง ๆ ให้อยู่ในรูปแบบตัวเลขไว้สำหรับใช้ในการเปรียบเทียบ (comparison) การวัด (measurement) และการประมาณค่า (estimation) ต่าง ๆ หรือใช้สถิติมาเป็นเกณฑ์มาตรฐาน สถิติ (statistics) ช่วยให้เราสามารถเห็นความเปลี่ยนแปลงของอดีตและปัจจุบันได้ชัดเจนยิ่งขึ้นด้วยเหตุนี้จึงมีการนำสถิติมาใช้ในด้านต่าง ๆ มากมายเพื่อใช้วางแผนการทำงานในอนาคตการวิเคราะห์ข้อมูล (data analysis) เพื่อหาแนวทางแก้ไขปัญหา การนำสถิติมาตีความหมายและประมวลผลเพื่อการตัดสินใจที่แม่นยำหรือกำหนดทิศทางในเรื่องต่าง ๆ ให้มีประสิทธิภาพมากยิ่งขึ้นไม่ว่าจะอยู่ในช่วงอายุไหน ประกอบอาชีพอะไร สถิติ (statistics) คือเรื่องใกล้ตัวที่เราสามารถนำไปประยุกต์ใช้ได้ในชีวิตประจำวัน โดยไม่จำเป็นต้องวิเคราะห์ข้อมูลเชิงลึก อีกทั้งยังส่งผลดีที่จะทำให้เกิดการพัฒนาในด้านนั้น ๆ ยิ่งขึ้นในอนาคต สามารถวางแผนกลยุทธ์ (strategic plan) เพื่อให้บรรลุเป้าหมายได้อย่างแม่นยำ แต่ทั้งนี้ต้องขึ้นอยู่กับข้อมูลด้วยว่าเราสามารถเก็บสถิติได้ละเอียดมาก-น้อยแค่ไหนและมีการวางแผนปรับปรุงโดยการอ้างอิงจากสถิติในอดีตได้อย่างไรถึงจะทำให้เกิดผลลัพธ์ที่ดีขึ้นกว่าเดิมวิชาสถิติ (statistics) นับว่าเป็นศาสตร์สาขาหนึ่งของวิชาคณิตศาสตร์ (mathematics) เป็นวิชาการที่เกี่ยวข้องกับชีวิตคนเราเพราะสถิติคือข้อมูลสำคัญที่ถูกจัดอยู่ในรูปแบบของตัวเลขเพื่อแสดงข้อเท็จจริงเกี่ยวกับเรื่องใดเรื่องหนึ่ง เช่น สถิติอุบัติเหตุช่วงเทศกาลสงกรานต์สถิติจำนวนอาสาสมัครเพื่อช่วยเหลือประชาชนช่วงเทศกาลสำคัญสถิติการเข้ารับบริการในหน่วยงานของรัฐ เป็นต้น ซึ่งมีเกณฑ์หรือหลักการสำคัญในการเก็บสถิติอยู่ 4 ขั้นตอน คือ 1) การเก็บรวบรวมข้อมูล 2) การนำเสนอข้อมูล 3) การวิเคราะห์ข้อมูล และ 4) การตีความหมายข้อมูล กล่าวได้ว่าวิชาสถิติ (statistics) มีประโยชน์ต่อการศึกษาในทุกแขนงวิชาและทุกหน่วยงาน เพราะทุกแขนงวิชาจะมีการศึกษาข้อมูล (data) เพื่อสรุปผลไปใช้ในการวางแผนและการตัดสินใจอยู่เสมอ นอกจากนั้นหน่วยงานต่าง ๆ ไม่ว่าจะเป็ภาครัฐบาลหรือภาคเอกชนยังใช้สถิติเพื่อการวางแผนดำเนินงานและตัดสินใจสำหรับปัญหาต่าง ๆ อยู่เสมอ ซึ่งโดยสรุปสถิติ (statistics) มีประโยชน์สำหรับการทำงานในด้านต่าง ๆ มากมาย เช่น

1) ด้านการพัฒนาประเทศ (national development) สำหรับในการพัฒนาประเทศนั้นไม่ว่าจะเป็นทางด้านเศรษฐกิจหรือด้านสังคม รัฐบาลต้องรู้สถิติ (statistics) ข้อมูล (data) เกี่ยวกับ

ประชากรและปัจจัยที่เกี่ยวข้อง เช่น สถิติประชากร สถิติการศึกษา สถิติแรงงาน สถิติการเกษตร สถิติการเกิด สถิติการตาย สถิติการส่งสินค้าออกหรือการนำเข้าสินค้า เป็นต้น เพื่อให้สามารถวางแผนหรือกำหนดนโยบายในการพัฒนาประเทศให้ดียิ่งขึ้น

2) ด้านการพัฒนาธุรกิจ (business development) ในการประกอบธุรกิจนักธุรกิจจำเป็นต้องมีการวางแผนในการดำเนินงานในด้านต่าง ๆ เช่น การวางแผนการผลิต (manufacturing) การจำหน่าย (distribution) การบริหารจัดการ (management) เป็นต้น นอกจากนั้นในบางครั้งอาจเกิดปัญหาเกี่ยวกับธุรกิจ เช่น ปัญหาด้านการผลิต ปัญหาด้านการตลาด ปัญหาเกี่ยวกับพนักงาน ปัญหาการควบคุมคุณภาพสินค้า เป็นต้น นักธุรกิจจึงต้องทำการตัดสินใจเพื่อแก้ปัญหาหรือดำเนินการตามแผนการธุรกิจ ซึ่งการวางแผนที่ดีและการตัดสินใจที่ถูกต้องย่อมส่งผลให้การดำเนินทางธุรกิจนั้นราบรื่นและมีโอกาสประสบความสำเร็จสูง ดังนั้นในการวางแผน (planning) และการตัดสินใจทางธุรกิจจำเป็นต้องอาศัยข้อมูลที่ได้จากการวิเคราะห์ด้วยวิธีการทางสถิติ (analysis using statistical methods) ที่น่าเชื่อถือ

3) ด้านการพัฒนาการเกษตร (agricultural development) ในการทำการเกษตรเกษตรกร (agriculturist) ต้องการทำให้ผลผลิตมากที่สุดหรือสูงกว่าเดิมในระยะเวลาที่ต้องการหรือต้องการให้มีผลผลิตนอกฤดูกาล เช่น เกษตรกรผู้เลี้ยงโคนมต้องการให้แม่โคมีปริมาณน้ำนมมากขึ้นและมีระยะเวลายาวขึ้น เป็นต้น ดังนั้นในการทำการเกษตรจึงต้องอาศัยข้อมูลต่าง ๆ ที่เกี่ยวข้องเพื่อวางแผนในการดำเนินการให้ได้ผลผลิตตามที่คาดหวัง

4) ด้านการพัฒนาการศึกษา (educational development) ในด้านการศึกษา นั้นสถิติ (statistics) ถูกนำมาใช้เพื่อการพัฒนาคุณภาพการศึกษา เช่น การศึกษาหาวิธีการสอนวิชาสถิติเพื่อให้นักศึกษามีผลสัมฤทธิ์ทางการเรียนสูงขึ้น ต้องใช้สถิติ (statistics) ข้อมูลเกี่ยวกับนักศึกษาและกระบวนการสอน เป็นต้น นอกจากนั้นในงานด้านอื่น ๆ

สำหรับในประเด็นเรื่องสถิติที่ (statistics) ใช้ในการวิเคราะห์ข้อมูล (สำหรับการวิจัย) ผู้เขียนจะกล่าวให้ครอบคลุมในประเด็นดังต่อไปนี้ คือ

- 7.1 สถิติกับการวิจัย
- 7.2 ประเภทของสถิติ
- 7.3 แนวความคิดพื้นฐานทางสถิติอ้างอิง
- 7.4 ข้อมูล ตัวแปร และระดับการวัดตัวแปร
- 7.5 การแจกแจงความถี่
- 7.6 การวัดแนวโน้มเข้าสู่ส่วนกลาง
- 7.7 การวัดการกระจาย
- 7.8 เปอร์เซนไทล์ (percentile) เดไซล์ (decile) และควอไทล์ (quartile)

### 7.1 สถิติกับการวิจัย (statistics and research)

#### 1. ความหมาย

วิชาสถิติ (statistics) นั้น นับว่ามีประโยชน์ต่อการศึกษาในทุกแขนงวิชาและทุกหน่วยงาน เพราะทุกแขนงวิชาจะมีการศึกษาข้อมูล (data) เพื่อสรุปผลไปใช้ในการวางแผน (planning) และการตัดสินใจ (decision making) อยู่เสมอ นอกจากนั้นหน่วยงานต่าง ๆ ไม่ว่าจะเป็นภาครัฐบาลหรือ

ภาคเอกชนยังคงต้องใช้ข้อมูลทางสถิติ (statistical data) เพื่อการวางแผน (planning) ในการดำเนินงานการปฏิบัติงาน (perform) และการตัดสินใจสำหรับปัญหาต่าง ๆ อยู่เสมอ

วรรณดี แสงประทีปทอง (2554, หน้า 23-24) อธิบายว่า คำว่า “สถิติ” มาจากศัพท์ภาษาอังกฤษ คือ “statistics” ซึ่งมีรากศัพท์มาจากคำว่า “state” ดังนั้นตามความหมายดั้งเดิม “สถิติ” จึงหมายถึงข้อมูลหรือข่าวสาร (information) ที่เป็นประโยชน์ต่อรัฐในด้านต่าง ๆ เช่น ข้อมูลในการวางแผนกำลังคน (manpower planning) การเก็บภาษีอากร (taxation) การประกันสังคม (social security) การจัดการศึกษา (educational management) และการสาธารณสุข (public health) เป็นต้น ต่อมาคำว่า “สถิติ” ได้มีความหมายกว้างขวางขึ้น ซึ่งในปัจจุบันสถิติมีความหมายต่าง ๆ ดังนี้

1) สถิติในความหมายของข้อมูลสถิติ หมายถึง ตัวเลขที่ใช้แทนข้อเท็จจริงของสิ่งต่าง ๆ เช่น สถิติการเข้าชั้นเรียน ปริมาณการขายสินค้า สถิติจำนวนบุคลากรของหน่วยงานต่าง ๆ เป็นต้น

2) สถิติในความหมายของระเบียบวิธีการทางสถิติ หมายถึง ระเบียบวิธีการทางสถิติ ซึ่งประกอบด้วยการเก็บรวบรวมข้อมูล (data collection) การนำเสนอข้อมูล (data presentation) การวิเคราะห์ข้อมูล (data analysis) และการแปลความหมายข้อมูล (data Interpretation) ดังนั้นสถิติตามความหมายนี้ จึงเป็นเครื่องมือสำคัญ (master key) ของนักวิจัย นักวิชาการ และนักบริหาร

3) สถิติในความหมายของค่าสถิติ หมายถึง ค่าตัวเลขที่คำนวณได้จากข้อมูล (data) กลุ่มตัวอย่าง (sample) เช่น ค่าเฉลี่ยเลขคณิต (mean) ค่าส่วนเบี่ยงเบนมาตรฐาน (standard deviation) เป็นต้น

4) สถิติในความหมายของวิชาสถิติ หมายถึง วิชาวิทยาศาสตร์แขนงหนึ่งซึ่งมีเนื้อหาและรากฐานมาจากวิชาคณิตศาสตร์ (mathematics) และตรรกวิทยา (logic) โดยสถิติเป็นศาสตร์ของการตัดสินใจภายใต้สถานการณ์ที่ยังไม่แน่นอน

จากความหมายของสถิติในฐานะที่เป็น “ข้อมูล” กับ สถิติในฐานะที่เป็น “วิธีการ” มีข้อแตกต่างกันดังนี้ (อ้างถึงใน ศิริชัย กาญจนวาสี, 2537, หน้า 25)

สถิติในฐานะที่เป็นข้อมูล	สถิติในฐานะที่เป็นวิธีการ
1) เป็นปริมาณ (quantitative)	1) เป็นวิธีการปฏิบัติ (operational)
2) มักอยู่ในรูปข้อมูลดิบ (raw state)	2) สถิติช่วยในกระบวนการจัดกระทำข้อมูลดิบ (processing)
3) เป็นการบรรยาย (descriptive)	3) สถิติเป็นเครื่องมือในการวิเคราะห์ (tool of analysis)
4) เป็นวัตถุดิบที่ถูกป้อนเข้าสู่กระบวนการเมื่อข้อมูลถูกจัดกระทำแล้ว สามารถนำมาช่วยในการตัดสินใจ	4) สถิติเป็นระเบียบวิธีทางวิทยาศาสตร์ของการวิเคราะห์และตีความหมายข้อมูล
5) ตัวของมันเองไม่ค่อยมีความหมายถ้าขาด	5) เครื่องมือในการวิเคราะห์จะไร้ความหมาย

## วิชาการเปรียบเทียบวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

เครื่องมือในการวิเคราะห์	ถ้าปราศจากข้อมูลที่จะนำมาวิเคราะห์
6) ธรรมชาติของข้อมูลจะเป็นตัวกำหนดวิธีการวิเคราะห์	6) เครื่องมือที่ใช้จะเป็นตัวกำหนดวิธีการเก็บรวบรวมข้อมูล

ตารางที่ 1.1 ข้อแตกต่างระหว่างความหมายของสถิติในฐานะทางข้อมูลกับวิธีการ

### 2 หลักสถิติ

สถิติ (statistics) เป็นศาสตร์ (science) ที่ใช้เป็นเครื่องมือช่วยในการตัดสินใจ (choice) อย่างมีเหตุผล การแก้ปัญหาส่วนใหญ่มีความจำเป็นที่ต้องใช้ข้อมูล (data) สารสนเทศ (information) และกระบวนการทางสถิติ (statistical process) มาช่วยในการสรุปผล (conclude) หลักสถิติที่ใช้อธิบายจะแบ่งออกเป็น 2 นัยยะ คือ

1) สถิติที่เป็นตัวเลข (numeric statistics) หมายถึง ตัวเลขที่ได้จากการเก็บรวมข้อมูลมาด้วยวิธีการใด ๆ ก็ตามจากข้อมูลที่มีเป็นจำนวนมาก เช่น สถิติผู้สำเร็จการศึกษาในระดับปริญญาตรี ปริญญาโท และปริญญาเอกของมหาวิทยาลัยมหามกุฏราชวิทยาลัย ประจำปีการศึกษา 2561 สถิติการเกิดอุบัติเหตุของการจราจรทางบกในเทศกาลวันสงกรานต์ของสำนักงานตำรวจแห่งชาติ เป็นต้น

2) สถิติที่เป็นศาสตร์ (statistical science) หมายถึง ศาสตร์ (science) หรือวิชาที่ศึกษาเกี่ยวกับวิธีการเก็บรวบรวมข้อมูล การสุ่มตัวอย่าง วิเคราะห์ข้อมูล ทดสอบสมมติฐาน และการสรุปผลแบ่งเป็น 2 ประเภท คือ

ก. สถิติพรรณนา (descriptive statistics) เป็นสถิติที่กล่าวถึงวิธีการที่จะบรรยายลักษณะของข้อมูลที่เก็บรวบรวมมาได้โดยจะต้องเก็บข้อมูลมาทั้งหมดหรือทุกหน่วยของประชากรเป้าหมาย ดังนั้นสถิติพรรณนาจะไม่ได้ใช้ผลข้อมูลที่ปรากฏไปอ้างอิงกับข้อมูล (data) กลุ่มอื่น ๆ โดยจะเป็นการใช้วิธีการวิเคราะห์ข้อมูลอย่างง่าย ๆ มาบรรยายลักษณะของข้อมูล (data characteristics) เช่น การนำเสนอข้อมูล (presentation) การแจกแจงความถี่ (frequencies) การวัดแนวโน้มเข้าสู่ส่วนกลาง หรือการหาค่าการกระจายของข้อมูล (dispersion)

ข. สถิติอ้างอิง (inference statistics) เป็นสถิติที่กล่าวถึงการนำข้อมูลที่ได้จากตัวอย่าง (sample) หรือบางส่วนของกลุ่มประชากรเป้าหมาย (population) มาศึกษาโดยอาจใช้วิธีการของสถิติพรรณนามาช่วยวิเคราะห์แล้วนำข้อเท็จจริงที่ได้ไปอธิบายหรือสรุปผลลักษณะของกลุ่มประชากร (population) เป้าหมายทั้งหมด ซึ่งจะเรียกว่าเป็นการอ้างอิง (reference) หรือการอนุมาน (inference statistics)

### 3. ค่าสถิติ

ค่าสถิติ (statistics) หมายถึง ค่าที่ได้จากข้อมูลตัวอย่าง (sample) หรือข้อมูลบางส่วน ค่าสถิตินี้จะแสดงถึงคุณลักษณะ (attribute) ของข้อมูลตัวอย่าง (sample) โดยในการหาค่าสถิตินั้นอาจจะหาโดยการคำนวณ (calculation) หรือวิธีอื่น ๆ ก็ได้ เช่น การคำนวณค่าเฉลี่ย แต่เป็นการนำ

ข้อมูลบางส่วนมาดำเนินการ ดังนั้น ค่าสถิติ (statistics) จึงมีความหมายที่ครอบคลุมค่าต่าง ๆ ที่คำนวณมาจากกลุ่มตัวอย่าง (sample) และจะเป็นค่าที่เปลี่ยนแปลงได้ตามกลุ่มตัวอย่างที่ได้เลือกสุ่มมา จึงถือว่าเป็นค่าตัวแปรสุ่ม (random variable) ข้อมูลที่รวบรวมมาจากการวิจัยมีตัวเลขจำนวนมาก การนำสถิติมาจัดตัวเลขเหล่านั้นให้เป็นระเบียบจะทำให้ผู้อ่านเข้าใจได้ถูกต้องตรงความเป็นจริงในเวลาอันรวดเร็ว

#### 4. ค่าพารามิเตอร์ (parameter)

ค่าพารามิเตอร์ (parameter) หมายถึง ค่าของสิ่งต่าง ๆ ที่คำนวณมาจากกลุ่มประชากร (population) จะถือเป็นค่าคงตัว กล่าวคือเราคำนวณกี่ครั้ง ๆ ก็จะไม่เปลี่ยนแปลงค่าพารามิเตอร์ (parameter) คือค่าที่ได้จากข้อมูลประชากร (population) หรือข้อมูลทั้งหมด ซึ่งค่าพารามิเตอร์ (parameter) จะแสดงถึงคุณลักษณะของข้อมูลประชากรโดยในการหาค่าสถิตินั้น จะหาโดยวิธีการเดียวกันกับการหาค่าสถิติ แต่เป็นการนำข้อมูลทั้งหมดหรือทุกจำนวนมาดำเนินการ ซึ่งในความเป็นจริงคงจะทำได้ยากและถ้าทำเช่นนี้ได้ก็ไม่จำเป็นต้องมีการหาค่าสถิติ และจะไม่มี การอนุมานทางสถิติ ดังนั้น อาจจะกล่าวได้อีกประเด็นหนึ่งว่า “การอนุมานทางสถิติ” ก็คือการศึกษาค่าสถิติเพื่อไปอธิบายค่าพารามิเตอร์ (parameter) นั้นเอง เช่น การหาค่ารายได้เฉลี่ยของคนไทยที่เลือกมาแบบวิธีการสุ่มเพื่อไปอธิบายรายได้เฉลี่ยของคนไทยทั้งประเทศ ซึ่งสัญลักษณ์ต่าง ๆ ของค่าสถิติและค่าพารามิเตอร์แสดงได้ในตารางที่ 1.2 ดังนี้

ค่าที่ได้จากการวิเคราะห์ข้อมูล	ค่าสถิติ (statistics)	ค่าพารามิเตอร์ (parameter)
ค่าเฉลี่ย (Mean)	$\bar{X}$	$\mu$
ค่าสัดส่วน (Proportion)	P	$\pi$
ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)	S	$\sigma$
ค่าความสัมพันธ์ (Correlation)	R	$\rho$
ค่าสัมประสิทธิ์การถดถอย (Regression)	B	$\beta$
ค่าอื่น ๆ	$\hat{\theta}$	$\theta$

ตารางที่ 1.2 สัญลักษณ์ค่าสถิติ (statistics) และค่าพารามิเตอร์ (parameter)

ในการอ้างอิง (reference) หรือการอนุมาน (inference) ด้วยวิธีการทางสถิติตามระเบียบวิธีวิจัย เราสามารถที่จะจำแนกการอนุมานออกได้เป็น 2 ประเภท คือ การอนุมานแบบพารามิเตอร์ (parametric inference) และการอนุมานแบบนอนพารามิเตอร์ (non-parametric inference)

#### 5. การอนุมานแบบพารามิเตอร์

วิธีการทางสถิติการอนุมานแบบพารามิเตอร์ (parametric inference) เป็นวิธีการอนุมานแบบที่มีข้อกำหนดเกี่ยวกับลักษณะของข้อมูล (data characteristics) ที่จะนำมาอนุมานหลายประการและจะแตกต่างกันออกไปสำหรับการอนุมาน (inference) แต่ละแบบ แต่ก็พอจะสรุป

คุณสมบัติ (qualification) ทัวไปว่าข้อมูลที่สามารถนำเอาวิธีการอนุมานแบบพาราเมตริก (parametric) ไปใช้นั้นควรมีคุณลักษณะ ดังนี้

- 1) ระดับการวัดของข้อมูลควรอยู่ในระดับช่วง (interval) หรืออัตราส่วน (ratio)
- 2) ข้อมูลควรมีการแจกแจงแบบปกติ (normal distribution)
- 3) ข้อมูลที่ถูกเลือกเป็นตัวอย่างควรจะได้มาจากการสุ่มที่ไม่มีความเอนเอียง

#### 6. การอนุมานแบบนอนพาราเมตริก

การอนุมานแบบนอนพาราเมตริก (non-parametric inference) เป็นวิธีการอนุมานที่จะใช้กับข้อมูลที่มีคุณสมบัติที่ไม่เป็นไปตามเงื่อนไข (condition) และข้อกำหนด (requirements) ตามวิธีการอนุมานแบบพาราเมตริก (parametric) เช่น ข้อมูลไม่ได้อยู่ในระดับช่วง (interval) หรืออัตราส่วน (ratio) หรือไม่ทราบลักษณะการแจกแจงของข้อมูล นอกจากนี้อาจจะใช้วิธีการอนุมานแบบนอนพาราเมตริก (non-parametric inference) กับกลุ่มข้อมูลตัวอย่าง (sample) ที่เลือกมานั้นมีขนาดเล็กหรือจำนวนน้อย

#### 7. มาตรการวัด (measure)

ในการเตรียมข้อมูลเพื่อการวิเคราะห์ เราจะต้องเข้าใจข้อมูลที่เราได้ ซึ่งข้อมูลที่เก็บรวบรวมมาได้นั้นอาจมีลักษณะต่างกันเพื่อการเลือกใช้สถิติที่เหมาะสมในการประมวลผล (data processing) ซึ่งสามารถจัดระดับของข้อมูลได้ 4 ระดับ ตามวิธีการวัดค่าดังต่อไปนี้

##### 1) มาตรฐานนามบัญญัติ (nominal scale)

มาตรฐานนามบัญญัติ (nominal scale) เป็นระดับของข้อมูลที่ได้จากวัดแบบง่ายที่สุดคือ เป็นการแบ่งแยกประชากร (population) ที่จะศึกษาออกเป็นกลุ่มหรือเป็นพวก โดยแต่ละกลุ่มและแต่ละพวกมีความเท่าเทียมกัน (equal) เช่น แบ่งประชากรโดยใช้เพศเป็นตัวแบ่ง คือ ชายและหญิง แบ่งประชากรโดยใช้ภาคเป็นตัวแบ่ง คือ ภาคเหนือ ภาคใต้ ภาคกลาง ภาคตะวันออกเฉียงเหนือ และภาคตะวันออกเฉียงเหนือจะเห็นว่าแต่ละกลุ่มแยกออกจากกันและกันแสดงถึงความแตกต่างของประชากรในการนำไปใช้อาจกำหนดตัวเลขหรือสัญลักษณ์แทนกลุ่ม เช่น ถ้าเป็นเพศชาย กำหนดให้เป็น M เพศหญิง กำหนดให้เป็น F หรือการกำหนดเบอร์ให้กับนักฟุตบอล ผู้รักษาประตูเป็น 1 กองหน้าเป็น 2, 3, เป็นต้น ตัวเลขหรือสัญลักษณ์ (symbol) เหล่านี้เป็นเพียงชื่อที่แทนกลุ่มเท่านั้น ไม่สามารถนำมาใช้ในการคำนวณทางคณิตศาสตร์ได้

##### 2. มาตรการเรียงลำดับ (ordinal scale)

มาตรการเรียงลำดับ (ordinal scale) เป็นระดับของข้อมูลที่กำหนดรายละเอียดของการวัดเพิ่มขึ้นจากระดับนามบัญญัติ กล่าวคือนอกจากจะแบ่งแยกข้อมูลออกเป็นกลุ่มแล้วยังสามารถหาระดับความแตกต่างระหว่างกลุ่มได้ด้วย (differences between groups) ซึ่งระบบการวัดแบบนี้ใช้หลักของความมากกว่า ความน้อยกว่า เช่น การแบ่งประชากรโดยใช้ความคิดเห็นซึ่งอาจมีระดับต่าง ๆ เช่น เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่มีความเห็น ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง การแบ่งสินค้าโดยใช้คุณภาพของสินค้า คือ ดีมาก ดี พอใช้ ต้องปรับปรุงจากตัวอย่างดังกล่าว สามารถบอกความแตกต่างของแต่ละกลุ่มข้อมูลได้โดยจัดอันดับของข้อมูล แต่ไม่สามารถกำหนดปริมาณความน้อยกว่าหรือมากกว่าออกมาเป็นตัวเลขที่แน่นอนได้ เช่น สามารถบอกได้ว่าสินค้าที่มีคุณภาพนั้นย่อมจะดีกว่าสินค้าพอใช้ แต่ไม่สามารถบอกได้ว่าดีกว่าเป็นตัวเลขเท่าไร การใช้ตัวเลขหรือสัญลักษณ์ใด ๆ



กำหนดอันดับความมากกว่าหรือน้อยกว่า จะไม่มีผลต่อข้อมูล เช่น การให้ร้อยตรีติดดาว 1 ดวง ร้อยโทติดดาว 2 ดวง และร้อยเอกติดดาว 3 ดวง อาจกำหนดใหม่ให้ร้อยเอกติดดาว 1 ดวง ร้อยโทติดดาว 2 ดวง และร้อยตรีติดดาว 3 ดวง โดยที่อันดับก่อน และหลังย่อมไม่เปลี่ยนแปลง ตัวเลขที่แทนข้อมูลระดับนี้ยังไม่สามารถนำมาคำนวณทางคณิตศาสตร์ได้

### 3. มาตราอันตรภาค (interval scale)

มาตราอันตรภาค (interval scale) เป็นมาตราวัดที่มีคุณสมบัติเพิ่มขึ้นจากมาตราวัดแบบที่ 2 คือ ทราบระยะห่างของความแตกต่างระหว่างกลุ่ม 2 กลุ่ม อย่างชัดเจน เช่น ความแตกต่างระหว่างค่า 20 และ 30 เท่ากับ 10 และเท่ากับ ความแตกต่างระหว่างค่า 30 กับ 40 มาตราวัดนี้มักวัดค่าเป็นเชิงปริมาณ (quantitative) แต่ไม่มีจุดศูนย์ (none zero) หรือจุดเริ่มต้นที่แท้จริง ตัวอย่างที่นิยมใช้กับมาตราวัดนี้ คือ อุณหภูมิ, ค่า I.Q., ความดัน, คะแนนให้แก่ความสามารถต่าง ๆ (ส่วนมากเป็นคะแนนด้านจิตวิทยา) ข้อมูลเกี่ยวกับอุณหภูมิ เช่น องศาเซลเซียส 0 °C คือจุดเยือกแข็ง ในขณะที่ 32 °F เป็นจุดเยือกแข็งของฟาเรนไฮต์ ความแตกต่างระหว่าง 30 °C กับ 10 °C นั้นทราบเพียงว่าห่างกัน = 20 °C แต่ไม่มีความหมายว่าเป็น 3 เท่าระหว่างกัน ดังนั้นมาตราวัดนี้ใช้ได้เพียงเครื่องหมายบวกและลบในเชิงพีชคณิต (algebra) เท่านั้น และยังไม่รวมกับการใช้เครื่องหมายคูณและหาร

### 4 มาตรฐานอัตราส่วน (ratio scale)

มาตรฐานส่วน ratio scale เป็นระดับของข้อมูล (data) ที่ถือว่ามีความสมบูรณ์ที่สุด และเป็นการวัดระดับสูงสุดมีจุดเริ่มต้นเป็นธรรมชาติ คือมีศูนย์แท้ (true zero) ที่หมายความว่า การไม่มีค่า เช่น น้ำหนักส่วนสูง อายุ ฯลฯ ข้อมูลบางชนิดไม่สามารถวัดได้ถึงระดับนี้ เช่น ข้อมูลทางพฤติกรรม (behavior) ทศนคติ (attitude) ข้อมูลในระดับนี้สามารถนำไปคำนวณ (calculate) ในทางคณิตศาสตร์ได้

หลังจากที่ได้เก็บหรือรวบรวมข้อมูลและดำเนินการจัดระเบียบข้อมูลให้อยู่ในสภาพที่เรียบร้อยพร้อมที่จะนำไปวิเคราะห์ได้แล้ว งานในขั้นต่อไปของผู้วิจัยคือการตัดสินใจว่าจะนำเสนออะไรมาใช้ ซึ่งในกระบวนการนี้ผู้วิจัยจะต้องทราบตั้งแต่แรกว่าข้อมูลที่มีอยู่ในลักษณะใดและต้องการเสนอผลการวิเคราะห์อะไรข้อมูลที่เก็บรวบรวมมาได้บางครั้งยังมีรูปแบบที่กระจัดกระจายเป็นรายบุคคลไม่เป็นระบบ จำเป็นต้องมีกระบวนการจัดกระทำข้อมูลเหล่านั้นให้เป็นระบบหรือเป็นหมวดหมู่เกิดเป็นสารสนเทศที่สามารถนำไปใช้ประโยชน์ เพื่อสรุปอ้างอิงไปยังประชากรต่อไป ศาสตร์ที่ถูกนำเข้ามาช่วยในขั้นตอนของการตรวจสอบคุณภาพของเครื่องมือไปจนถึงการอ้างอิงเหล่านี้ เรียกว่า สถิติ (statistics)

## 7.2 ประเภทของสถิติ

นักคณิตศาสตร์ (mathematician) นักสถิติ (Statistician) และนักวิจัย (researcher) ส่วนมากมักนิยมที่จะสถิติแบ่งเป็น 2 ประเภท (ออนไลน์, 2561)

### 1. สถิติพรรณนา (descriptive statistics)

เมื่อกล่าวถึงสถิติพรรณนา (descriptive statistics) จะเป็นสถิติที่ใช้ในการสรุปข้อมูลที่ได้มาจากกลุ่มตัวอย่าง (sample) โดยไม่มีการอ้างอิงไปยังประชากร (population) แต่เป็นการบรรยายลักษณะข้อมูลเท่านั้น เช่น การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจายของข้อมูล ฯลฯ ส่วนการนำเสนอผลการวิเคราะห์ข้อมูลโดยการพรรณนามักจะอยู่ในรูป

ของตาราง (table) และแผนภูมิ (Chart) ชนิดต่าง ๆ เป็นสถิติที่ใช้อธิบายคุณลักษณะ (attribute) ของสิ่งที่ต้องการศึกษากลุ่มใดกลุ่มหนึ่ง ไม่สามารถอ้างอิงไปยังกลุ่มอื่น ๆ ได้ สถิติที่อยู่ในประเภทนี้ เช่น ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) ค่าฐานนิยม (mode) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ค่าพิสัย (range) ฯลฯ

## 2. สถิติอ้างอิง (inferential statistics)

สถิติอ้างอิง (inferential statistics) หรือสถิติอนุมาน เป็นสถิติที่ใช้เพื่อนำผลสรุปที่คำนวณได้จากการสุ่มตัวอย่าง (sample) ไปอธิบายหรือสรุปลักษณะของประชากร (population) ทั้งหมด วิธีที่ใช้ในการสรุปอ้างอิงไปยังกลุ่มประชากรนั้น คือ การประมาณค่า (estimation) และการทดสอบสมมติฐาน (hypothesis testing) การวิเคราะห์ความแปรปรวน (variance analysis) การวิเคราะห์ความถดถอยและสหสัมพันธ์ (regression and correlation analysis) สถิติอ้างอิงจำแนกเป็น 2 ชนิด คือ

1) แบบอ้างอิงพารามิเตอร์ (parametric statistics) เช่น การทดสอบสมมติฐานโดยใช้สถิติ t-test, z-test, ANOVA, regression analysis เป็นต้น ส่วนตัวแปรที่ต้องการวัดจะเป็นมาตราอันตรภาค (interval scale) กลุ่มตัวอย่าง (sample) จะต้องมีการแจกแจงเป็นโค้งปกติ กลุ่มประชากร (population) จะต้องมีความแปรปรวน (variance) เท่ากัน

2) แบบไม่อ้างอิงพารามิเตอร์ (nonparametric statistics) เช่น การใช้สถิติ Chi-square, median test, sign test กลุ่มตัวอย่างเป็น free distribution เป็นกลุ่มตัวอย่างขนาดเล็ก และเราไม่ทราบลักษณะการแจกแจงของประชากร (population) ที่สนใจจะศึกษา สถิติที่ใช้เป็นเครื่องอธิบายคุณลักษณะของสิ่งที่ต้องการศึกษากลุ่มใดกลุ่มหนึ่งหรือหลายกลุ่มแล้วสามารถอ้างอิงไปยังกลุ่มประชากรได้ โดยกลุ่มที่นำมาศึกษาจะต้องเป็นตัวแทนที่ดี (representation) ของประชากร ตัวแทนที่ดีของประชากรที่ได้มาโดยวิธีการสุ่มตัวอย่าง (random) และตัวแทนที่ดีของประชากรเรียกว่า “กลุ่มตัวอย่าง” หรือ “sample group”

สถิติอ้างอิง (reference statistics) แบ่งได้เป็น 2 ประเภท คือ

1 สถิติพารามิเตอร์ (parametric statistics) เป็นวิธีการทางสถิติ (method) ที่จะต้องเป็นไปตามข้อตกลงเบื้องต้น 3 ประการ ดังนี้ (พารามิเตอร์ หมายถึง ค่าที่ใช้อธิบายคุณลักษณะประชากร เช่น ค่าเฉลี่ยของประชากร เป็นต้น)

- 1) ข้อมูลที่เก็บรวบรวมได้จะต้องอยู่ในระดับช่วงขึ้นไป (interval scale)
- 2) ข้อมูลที่เก็บรวบรวมได้จากกลุ่มตัวอย่างจะต้องมีการแจกแจงเป็นโค้งปกติ
- 3) กลุ่มประชากรแต่ละกลุ่มที่นำมาศึกษาจะต้องมีความแปรปรวนเท่ากัน

สถิติที่อยู่ในประเภทนี้ เช่น t-test, Z-test, ANOVA, Regression ฯลฯ

2 สถิติไร้พารามิเตอร์ (nonparametric statistics) เป็นวิธีการทางสถิติที่สามารถนำมาใช้ได้โดยปราศจากข้อตกลงเบื้องต้นทั้ง 3 ประการข้างต้น สถิติที่อยู่ในประเภทนี้ เช่น Chi-square, median test, Sign test ฯลฯ

## 3. ลักษณะข้อมูลที่ใช้ในงานวิจัย

1) ข้อมูลเชิงปริมาณ (quantitative data) แบ่งเป็นข้อมูลต่อเนื่องกล่าวคือ ค่าที่มีจุดทศนิยมได้ และข้อมูลไม่ต่อเนื่อง (discrete data) คือค่าที่เป็นจำนวนเต็มหรือจำนวนนับ

2) ข้อมูลเชิงคุณภาพ (qualitative data) เป็นข้อมูลที่แสดงถึงสถานภาพ (status) คุณลักษณะ (attribute) หรือคุณสมบัติ (qualification) เช่น เพศ ตำแหน่ง หรือจำแนกตัวแปร



(variable) ตามระดับการวัด ได้แก่ นามบัญญัติ (nominal scale) เรียงอันดับ (ordinal scale) อันตรภาคหรือระดับช่วง (interval scale) อัตราส่วน (ratio scale)

3) จำแนกตามหน้าที่ ได้แก่ ตัวแปรอิสระ (independent variables) และตัวแปรตาม (dependent variables) ซึ่งเป็นตัวแปรที่ต้องการศึกษา นอกจากนั้นอาจมีตัวแปรที่ไม่ได้ต้องการศึกษาแต่ต้องควบคุม เช่น ตัวแปรภายนอก (ตัวแปรเกินหรือตัวแปรแทรกซ้อน) และตัวแปรเชื่อมโยง (หรือตัวแปรสอดแทรก)

#### 4. สถิติพรรณนาที่ใช้อธิบายข้อมูลเชิงปริมาณ

สำหรับสถิติพรรณนาที่ใช้อธิบายข้อมูลเชิงปริมาณ (descriptive statistics) ที่นักสถิติ (statistician) และนักวิจัย (researcher) ส่วนใหญ่นำมาใช้ในการอธิบายข้อมูลเชิงปริมาณ (quantitative data) มีองค์ประกอบที่สำคัญ ดังนี้ (อ้างถึงใน วรเชษฐ์ โทอิน, 2560, หน้า 125)

1) การแจกแจงข้อมูล ได้แก่ ความถี่ (frequency) ร้อยละ (percentage)

2) วัดแนวโน้มเข้าสู่ส่วนกลาง ได้แก่ ค่าเฉลี่ย (mean) นำข้อมูลทั้งหมดมารวมกันแล้วหารด้วยจำนวนข้อมูลมัธยฐาน (median) เป็นสถิติในการจัดอันดับข้อมูล เป็นค่าที่อยู่ตรงกลางเมื่อนำค่าที่ได้จากการวัดที่นำมาเรียงลำดับจากมากไปน้อยหรือน้อยไปมากฐานนิยม (mode) หรือคะแนนที่มีความถี่สูงสุด

3) บอกตำแหน่งของข้อมูล ได้แก่ เปอร์เซนต์ไทล์ (percentile) เดไซล์ (decide) ควอไทล์ (quartile)

5) วัดการกระจายของข้อมูล ได้แก่ พิสัย (range) หรือค่าสูงสุด-ค่าต่ำสุด (max-min) ส่วนเบี่ยงเบนควอไทล์ (quartile deviation) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ความแปรปรวนของข้อมูล (variance)

#### 5. สถิติพรรณนาที่ใช้อธิบายข้อมูลเชิงคุณภาพ

สถิติพรรณนาที่ใช้อธิบายข้อมูลเชิงคุณภาพ (descriptive statistics) ที่นักสถิติและนักวิจัย (researcher) ส่วนใหญ่นำมาใช้ในการอธิบายข้อมูลเชิงปริมาณ (qualitative data) มีองค์ประกอบที่สำคัญ ดังนี้

1) ค่าร้อยละ (percentage)

2) สัดส่วน (proportion)

3) อัตราส่วน (ratio)

4) ฐานนิยม (mode)

#### 6. สถิติพรรณนาที่ใช้อธิบายความสัมพันธ์ระหว่างข้อมูลเชิงปริมาณ

สถิติพรรณนาที่ใช้อธิบายความสัมพันธ์ระหว่างข้อมูลเชิงปริมาณประกอบด้วย

1) ค่าสัมประสิทธิ์สหสัมพันธ์เชิงเส้นของเพียร์สัน (Pearson's correlation coefficient)

2) ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน (Spearman's correlation coefficient)

#### 7. สถิติพรรณนาที่ใช้อธิบายความสัมพันธ์ระหว่างข้อมูลเชิงคุณภาพ

สถิติพรรณนาที่ใช้อธิบายความสัมพันธ์ระหว่างข้อมูลเชิงคุณภาพประกอบด้วย

1) สัมประสิทธิ์สหสัมพันธ์ของเครเมอร์วี (Cramer's V)

2) สร้างตารางไขว้แสดงความสัมพันธ์ระหว่างตัวแปร (cross tabulation table)

### 7.3 แนวความคิดพื้นฐานทางสถิติอ้างอิง

ในการเก็บรวบรวมข้อมูลการวิจัย (collecting research data) สิ่งแรกที่นักวิจัยจะต้องกำหนด คือ ประชากร (population) ที่ต้องการศึกษา จากนั้นนักวิจัย (researcher) ต้องพิจารณาต่อไปว่าจะสามารถรวบรวมข้อมูลจากประชากรทั้งหมดหรือไม่ ถ้าไม่ได้ก็ต้องทำการศึกษาเพียงบางส่วน of ประชากรเท่านั้น การที่ประชากรที่นักวิจัยสนใจมีขนาดใหญ่คือมีจำนวนมาก งานวิจัยจึงไม่สามารถศึกษาทุกหน่วยของประชากรได้และเป็นสาเหตุให้นักวิจัยจะต้องเลือกกลุ่มตัวแทน (representative sample) ของประชากรมาใช้ในการศึกษา ซึ่งเราเรียกกันโดยทั่วไปว่า กลุ่มตัวอย่าง (sample) โดยที่ค่าต่าง ๆ ที่คำนวณได้จึงมีชื่อเรียกตามกลุ่มตัวอย่าง (sample) และประชากร (population) กลุ่มสถิติที่นักวิจัยใช้ในการศึกษา แบ่งออกเป็น 2 ชนิด คือ ค่าพารามิเตอร์ (parameter) และค่าสถิติ (Statistic) (อ้างอิงใน ธานินทร์ ศิลป์จารุ, 255 หน้า 54-55)

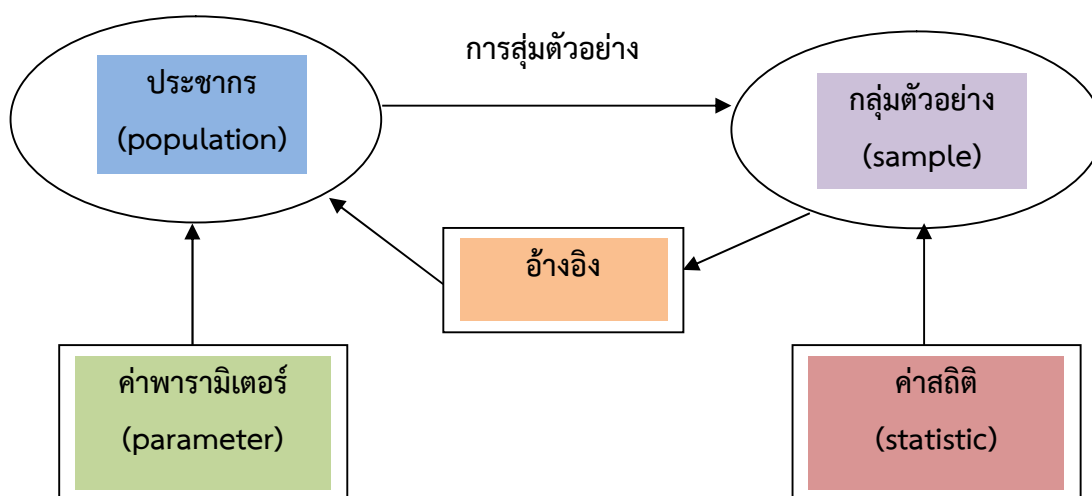
1. พารามิเตอร์ (Parameter) คือ ค่าของสิ่งต่าง ๆ ที่รวบรวมมาจากประชากรหรือคำนวณได้จากประชากร (population) ใช้อักษรกรีกเป็นสัญลักษณ์ ดังนี้ (อ้างอิงใน วรเชษฐ์ โทอิน, 2560, 129-130)

- $\mu$  แทนค่า ค่าเฉลี่ยเลขคณิต
- $\sigma$  แทนค่า ส่วนเบี่ยงเบนมาตรฐาน
- $\sigma^2$  แทนค่า ความแปรปรวน
- $\rho$  แทนค่า สัมประสิทธิ์สหสัมพันธ์

2. ค่าสถิติ (Statistic) คือค่าของสิ่งต่าง ๆ ที่รวบรวมมาจากกลุ่มตัวอย่างหรือที่คำนวณได้จากกลุ่มตัวอย่าง (sample) ใช้ตัวภาษาอังกฤษเป็นสัญลักษณ์ ดังนี้

- $\bar{x}$  แทนค่า ค่าเฉลี่ยเลขคณิต
- $s$  แทนค่า ส่วนเบี่ยงเบนมาตรฐาน
- $s^2$  แทนค่า ความแปรปรวน
- $r$  แทนค่า สัมประสิทธิ์สหสัมพันธ์

### 3. แนวคิดพื้นฐานของการวิเคราะห์ข้อมูลทางสถิติ



#### 4. สิ่งที่ต้องพิจารณาในการเลือกใช้ชนิดทางสถิติ

ในการพิจารณาถึงการเลือกใช้สถิตินั้น นักวิจัยจะต้องมีการคำนึงถึง (consideration) จุดมุ่งหมาย (aim) หรือวัตถุประสงค์ของการวิจัย (research objectives) ซึ่งโดยทั่วไปจะแบ่งจุดมุ่งหมายหรือวัตถุประสงค์ของการวิจัย ดังนี้

1. เพื่อบรรยายลักษณะตัวแปร (variable description) ในกลุ่มตัวอย่าง (sample) หรือประชากร (population) เป็นการใช้สถิติเชิงบรรยาย (descriptive statistics) มาบรรยายภาพรวมของกลุ่มตัวอย่าง ประกอบด้วย

1) การแจกแจงความถี่ (frequency) และค่าร้อยละ (percentage) และนำผลจากการแจกแจงความถี่หรือค่าร้อยละของข้อมูลบรรยายเพื่อแสดงภาพรวมของข้อมูลที่ได้ ส่วนรูปแบบในการนำเสนอข้อมูลนั้นมักจะนิยมใช้ตาราง (table) และแผนภูมิ (chart) มากกว่าคำบรรยาย (description) เพียงอย่างเดียว

2) การวัดแนวโน้มเข้าสู่ส่วนกลาง ได้แก่ ค่าเฉลี่ยเลขคณิต (mean) มัชยฐาน (median) ฐานนิยม(mode)

3) การวัดการกระจาย ได้แก่ ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

2. เพื่อเปรียบเทียบหาความแตกต่างและสรุปอ้างอิงหาความแตกต่างจากกลุ่มตัวอย่าง (sample) กลับไปยังประชากร (population) ที่ศึกษา ได้แก่

1) การเปรียบเทียบค่าเฉลี่ย (average comparison) ของกลุ่มตัวอย่าง 2 กลุ่ม ที่เป็นอิสระกันด้วย independent t-test

2) การเปรียบเทียบค่าเฉลี่ย (average comparison) ของกลุ่มตัวอย่าง 2 กลุ่ม ที่ไม่เป็นอิสระต่อกันด้วย Pair t-test

3) การเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างตั้งแต่ 2 กลุ่มด้วย ANOVA

4) การเปรียบเทียบความถี่และสัดส่วนด้วยไคสแควร์ (Chi-square)

3. เพื่อบรรยายความสัมพันธ์ระหว่างตัวแปร (relationship between variables) ได้แก่ การใช้สหสัมพันธ์อย่างง่ายในการบรรยายความสัมพันธ์ระหว่าง 2 ตัวแปร เช่น การวิเคราะห์ค่าสหสัมพันธ์แบบเพียร์สัน (Pearson's product moment coefficient of correlation) และสหสัมพันธ์แบบสเปียร์แมน (Spearman's Correlation) และการใช้สหสัมพันธ์พหุคูณ (multiple correlations) ในการบรรยายความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป

#### 5. สัญลักษณ์

$\mu$  แทนค่า ค่าเฉลี่ยเลขคณิต

$\sigma$  แทนค่า ส่วนเบี่ยงเบนมาตรฐาน

$\sigma^2$  แทนค่า ความแปรปรวน

$\rho$  แทนค่า สัมประสิทธิ์สหสัมพันธ์

$\bar{x}$  แทนค่า ค่าเฉลี่ยเลขคณิต

$s$  แทนค่า ส่วนเบี่ยงเบนมาตรฐาน

$s^2$  แทนค่า ความแปรปรวน

$r$  แทนค่า สัมประสิทธิ์สหสัมพันธ์

$N$  แทนค่า ประชากร

n แทนค่า กลุ่มตัวอย่าง

X แทนค่า สถิติที่คำนวณได้

## 7.4 ข้อมูล ตัวแปร และระดับการวัดตัวแปร

### 1. ข้อมูล (Data)

คำว่า “ข้อมูล” เป็นคำกลาง ๆ ที่ใช้เรียกข่าวสาร (message) หรือข้อเท็จจริง (fact) ที่อาจจะเป็นตัวเลขหรือไม่เป็นตัวเลขก็ได้

1) ข้อมูลที่เป็นตัวเลข (numeric data) เช่น อายุ รายได้ อุณหภูมิ น้ำหนัก เป็นต้น ข้อมูลชนิดนี้จะมีหน่วยการวัดที่แสดงเป็นตัวเลขได้

2) ข้อมูลที่ไม่เป็นตัวเลข (non-numeric data) เช่น ความคิดเห็น อาชีพ เพศ คุณภาพของสินค้า เป็นต้น ข้อมูลชนิดนี้จะไม่มีการวัดที่เป็นรูปธรรมเหมือนข้อมูลที่เป็นตัวเลข

### 2 ตัวแปร (variable)

ข้อมูล (data) หรือสิ่งที่เราสนใจศึกษาซึ่งในการวิจัยจะมีชื่อเรียกอีกอย่างหนึ่งว่า “ตัวแปร” (variable) ซึ่งจะมีค่าแตกต่างกันได้ในกลุ่มที่เราไปศึกษา เช่น เพศ (gender) ของนายทหารมีทั้งหญิงและชาย ความคิดเห็นอาจมีทั้งเห็นด้วย ไม่เห็นด้วย และไม่เห็นความเห็น เมื่อเราพิจารณาข้อมูลเรื่องใดเรื่องหนึ่ง ข้อมูลแต่ละเรื่องนั้นก็นับได้ว่าเป็นตัวแปร (variable) ตัวหนึ่ง เช่น ข้อมูลเกี่ยวกับนายทหารประกอบด้วย เพศ อายุ ความสูง น้ำหนัก ชันยศ สังกัด เป็นต้น ข้อมูลเกี่ยวกับนายทหารแต่ละเรื่องดังกล่าวเป็นสิ่งที่เปลี่ยนแปลงค่าได้ในหมู่ทหารแต่ละกลุ่ม ข้อมูลดังกล่าวจัดเป็นตัวแปร โดยเพศแปรได้ 2 ค่า คือ ชายและหญิง กองทัพแปรได้ 3 ค่า คือ กองทัพบก กองทัพเรือ และกองทัพอากาศ ชั้นปีที่ศึกษาอาจแปรได้ 4 ค่า คือ ชั้นปีที่ 1, 2, 3 และ 4 หรืออาจกำหนดให้กองทัพแปรค่าได้เพียง 2 ค่า คือ กองทัพอากาศ และอื่น ตามที่ผู้ศึกษาจะกำหนดขึ้น

### 3 ระดับการวัดของข้อมูล

ข้อมูล (data) นั้นเราสามารถเก็บรวบรวมได้ทั้งจากแหล่งปฐมภูมิ (primary source) ซึ่งเป็นแหล่งที่เกิดของข้อมูลและเก็บมาจากแหล่งทุติยภูมิ (secondary source) ซึ่งเป็นแหล่งที่รวบรวมข้อมูลเอาไว้ก่อนแล้ว ข้อมูลที่รวบรวมได้มาทั้งที่มีค่าเป็นตัวเลขและไม่เป็นตัวเลข สามารถจำแนกประเภทหรือระดับการวัดได้ 4 ระดับ คือ ระดับกลุ่ม (nominal scale) ระดับจัดอันดับ (ordinal scale) ระดับช่วง (interval scale) และระดับอัตราส่วน (ratio scale) (อ้างถึงใน วรเชษฐ์ โทอิน, 2560, หน้า 132)

1) ข้อมูลระดับกลุ่ม (nominal scale) เป็นข้อมูลที่บอกความแตกต่างของสิ่งที่วัดเป็นกลุ่มเป็นพวก เช่น เพศ สังกัด หมายเลขโทรศัพท์ เลขประจำตัว ตัวเลขแบบนี้ไม่มีค่ามาก ค่าน้อยที่แตกต่างกัน จะนำมาเปรียบเทียบกันในเชิงพีชคณิต (algebra) ไม่ได้

2) ข้อมูลระดับจัดอันดับหรือเรียงลำดับ (ordinal scale) เป็นข้อมูลที่ระบุตำแหน่งที่ตั้งเรียงตามลำดับมากหรือน้อย เช่น ยศของทหาร แบ่งได้เป็น นายสิบ นายร้อย นายพัน และนายพล หรือลำดับขั้นของข้าราชการ แบ่งได้เป็น C1 (ซีหนึ่ง) C2 (ซีสอง) ถึง C11 (ซีสิบเอ็ด) ตัวแปรประเภทนี้แบ่งเป็นกลุ่มย่อย ๆ ซึ่งสามารถเปรียบเทียบว่ากลุ่มใดสูงกว่ากันได้ แต่ช่วงห่างของแต่ละกลุ่มไม่เท่ากันจะนำมาเปรียบเทียบในเชิงพีชคณิต (algebra) ไม่ได้

3) ข้อมูลระดับช่วงหรือระดับอันตรภาค (interval scale) เป็นข้อมูลที่มีค่าเป็นตัวเลขมีช่วงห่างของข้อมูลกลุ่มย่อยเท่า ๆ กันแต่ยังไม่มีค่าศูนย์ที่แท้จริง (none zero) เช่น อุณหภูมิ

ของอากาศแบ่งเป็นกลุ่มย่อยทั้งที่อุณหภูมิเป็นบวกและลบคะแนนจากการสอบซึ่งมีค่าต่าง ๆ ตั้งแต่ 0 ขึ้นไป ข้อมูลระดับนี้สามารถนำมาเปรียบเทียบ (compare) ความแตกต่างโดยการบวกหรือลบกันได้แต่ยังเปรียบเทียบเป็นจำนวนเท่าไม่ได้กล่าวคือผู้ที่สอบได้ 100 คะแนนไม่ได้หมายถึงว่ามีความรู้เป็น 2 เท่าของผู้ที่สอบได้ 50 คะแนนและผู้ที่ได้ 0 คะแนนก็ไม่ได้หมายถึงไม่มีความรู้ในเรื่องที่สอบเลย

4) ข้อมูลระดับอัตราส่วน (ratio scale) เป็นข้อมูลที่มีค่าเป็นตัวเลขมีช่วงห่างของข้อมูลกลุ่มย่อยเท่า ๆ กันและมีศูนย์แท้เช่นรายได้ความยาวความสูงน้ำหนักข้อมูลระดับนี้สามารถนำมาบวกลบคูณหารในเชิงพีชคณิต (algebra) ได้ เช่น ผู้ที่มีรายได้เดือนละ 3 หมื่นบาทเป็นผู้มีรายได้ 3 เท่าของผู้ที่มีรายได้ 1 หมื่นบาท และผู้ที่หนัก 100 กิโลกรัมหนักเป็น 2 เท่าของผู้ที่หนัก 50 กิโลกรัม เป็นต้น

ระดับการวัด	คุณสมบัติ				การจัดกระทำ	ทางพีชคณิต
ของข้อมูล	จัดกลุ่มได้	เรียงลำดับได้	ช่วงเท่ากัน	มีศูนย์แท้	บวก/ลบ	คูณ/หาร
ระดับกลุ่ม	ใช่	ไม่ใช่	ไม่ใช่	ไม่ใช่	ไม่ได้	ไม่ได้
ระดับจัดอันดับ	ใช่	ใช่	ไม่ใช่	ไม่ใช่	ไม่ได้	ไม่ได้
ระดับช่วง	ใช่	ใช่	ใช่	ไม่ใช่	ได้	ไม่ได้
ระดับอัตราส่วน	ใช่	ใช่	ใช่	ใช่	ได้	ได้

ตารางที่ 1-3 แสดงคุณสมบัติของข้อมูลระดับต่าง ๆ

ระดับการวัดของข้อมูลทั้ง 4 ระดับดังกล่าวเรียงลำดับจากการวัดที่เรียบง่ายสุดถึงละเอียดที่สุดได้ ดังนี้ 1) ระดับกลุ่ม (nominal scale) 2) ระดับจัดอันดับ (ordinal scale) 3) ระดับช่วง (interval scale) และ 4) ระดับอัตราส่วน (ratio scale) ข้อมูลจากการวัดเรื่องหนึ่ง ๆ อาจจำแนกระดับการวัดได้หลายระดับ ทั้งนี้ จะขึ้นอยู่กับความต้องการของผู้จำแนกประเภท เช่น ผลการสอบอาจระบุในลักษณะต่าง ๆ ดังนี้

- ถ้าระบุว่า สอบ ผ่าน ไม่ผ่าน รายวิชาสถิติ เป็นการวัดระดับ กลุ่ม
- ถ้าระบุว่า สอบ ได้ที่ 1 2 3 เป็นการวัดระดับ จัดอันดับ
- ถ้าระบุว่า สอบได้ 50 75 89 คะแนน เป็นการวัดระดับ ช่วง

#### 4 การเลือกใช้สถิติเพื่อวิเคราะห์ข้อมูล

โดยทั่วไปแล้วการวิเคราะห์ข้อมูลที่เก็บรวบรวมมาได้อีกเพื่อจะตอบคำถาม (issue) หรือตอบวัตถุประสงค์ (objective) ที่ตั้งไว้ซึ่งอาจจำแนกได้ 2 ลักษณะใหญ่ ๆ ได้แก่ เพื่อบรรยายลักษณะของข้อมูล (characteristics of data) หรืออธิบายความสัมพันธ์ของข้อมูล (data relationship) และเพื่อเปรียบเทียบความแตกต่างของข้อมูล (compare data differences) วิธีการเลือกใช้สถิติให้เหมาะสมนอกจากจะพิจารณาวัตถุประสงค์ของการวิเคราะห์ข้อมูลแล้วต้องพิจารณาว่าข้อมูล (data) หรือตัวแปร (variable) มีกี่ตัวและตัวแปรวัดในระดับใดเนื่องจากสถิติที่ใช้วิเคราะห์จะแตกต่างกันไป

##### ก. สถิติสำหรับวิเคราะห์ข้อมูลกรณีมีตัวแปรเดียว

การบรรยายลักษณะของข้อมูล (data description) มุ่งบรรยายข้อมูลทางสถิติในประเด็นการแจกแจงความถี่ (frequency distribution) การวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจาย (distribution measurement) การวัดความเบ้ (skewness measurement) และการวัดความโด่ง (measure of prominence) ในประเด็น ต่อไปนี้ (อ้างถึงใน ศักดิ์สิทธิ์ วัชรารัตน์, 2552, หน้า 128)

## วิชาการเปรียบเทียบวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

1) สถิติที่ใช้บรรยายลักษณะของตัวแปรระดับกลุ่ม ได้แก่

(1) การแจกแจงความถี่ สถิติที่เหมาะสมคือความถี่ของข้อมูลแต่ละกลุ่มในรูปของจำนวนครั้ง (frequency) และจำนวนร้อยละ (percent)

(2) การวัดแนวโน้มเข้าสู่ส่วนกลาง สถิติที่เหมาะสมคือฐานนิยม (mode)

(3) การวัดการกระจาย (distribution measurement) สถิติที่เหมาะสมคือความถี่สัมพัทธ์ (relative cumulative frequency) ของข้อมูลที่มีความถี่สูงสุด

2) สถิติที่ใช้บรรยายลักษณะตัวแปรระดับจัดอันดับ ได้แก่

(1) การแจกแจงความถี่ สถิติที่เหมาะสมคือความถี่ของข้อมูลแต่ละกลุ่มในรูปของจำนวนครั้ง (frequency) และจำนวนร้อยละ (percent)

(2) การวัดแนวโน้มเข้าสู่ส่วนกลาง สถิติที่เหมาะสมคือมัธยฐาน (median)

(3) การวัดการกระจาย สถิติที่เหมาะสมคือส่วนเบี่ยงเบนควอไทล์ (quartile deviation)

3) สถิติที่ใช้บรรยายลักษณะตัวแปรระดับช่วงหรือสูงกว่า ได้แก่

(1) การแจกแจงความถี่ สถิติที่เหมาะสมคือความถี่ของข้อมูลแต่ละช่วงในรูปของจำนวน ร้อยละ และเปอร์เซ็นต์ไทล์ (percentile) เดไซล์ (decide) และควอไทล์ (quartile)

(2) การวัดแนวโน้มเข้าสู่ส่วนกลาง สถิติที่เหมาะสมคือค่าเฉลี่ยเลขคณิตหรือมัธมิมเลขคณิต (arithmetic mean) หรือมัธยฐาน

(3) การกระจาย สถิติที่เหมาะสมคือส่วนเบี่ยงเบนมาตรฐาน (standard deviation) และพิสัย (range)

(4) ความเบ้ สถิติที่ใช้วัด คือ ค่าความเบ้ (skewedness)

(5) ความโด่ง สถิติที่ใช้วัด คือ ค่าความโด่ง (kurtosis)

การเลือกใช้สถิติเพื่อบรรยายข้อมูลกรณีมีข้อมูลหรือตัวแปรตัวเดียวต้องพิจารณาระดับของข้อมูลและประเด็นที่ต้องการบรรยาย ซึ่งสรุปได้ดังตารางที่ 1-3

ระดับของตัวแปร	สถิติที่เหมาะสม			
	การแจกแจงความถี่	การวัดแนวโน้มเข้าสู่ส่วนกลาง	การวัดการกระจาย	รูปร่างการกระจายตัว
ระดับกลุ่ม	จำนวน ร้อยละ	ฐานนิยม	ความถี่สัมพัทธ์	ไม่มี
ระดับจัดอันดับ	จำนวน ร้อยละ	มัธยฐาน	ส่วนเบี่ยงเบนควอไทล์	ไม่มี
ระดับช่วงหรือสูงกว่า	เปอร์เซ็นต์ไทล์	ค่าเฉลี่ยเลขคณิต มัธยฐาน	ส่วนเบี่ยงเบนมาตรฐาน พิสัย	ความเบ้ ความโด่ง

ตารางที่ 1-4 แสดงสถิติที่เหมาะสมในการบรรยายข้อมูลที่มีตัวแปรเดียว สำหรับกรณีที่มีข้อมูลหรือตัวแปรตัวเดียว จะไม่สามารถอธิบายความสัมพันธ์ของข้อมูล (data relationship) หรือเปรียบเทียบความแตกต่าง (compare differences) ของข้อมูลได้



**ข. สถิติสำหรับวิเคราะห์ข้อมูลกรณีตัวแปรสองตัว**

สถิติสำหรับวิเคราะห์ข้อมูล (data analysis) กรณีตัวแปรสองตัว มีวัตถุประสงค์ของการวิเคราะห์ข้อมูลจำแนกได้ 2 ลักษณะ คือ เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปร (describe the relationship between variables) และเพื่อเปรียบเทียบความแตกต่างระหว่างตัวแปร (compare the differences between variables) สิ่งที่ต้องพิจารณาต่อไปคือระดับการวัดของตัวแปร ซึ่งกรณีที่มีตัวแปรสองตัว จำแนกกรณีต่าง ๆ ที่เป็นไปได้ ได้ 6 กรณี ดังนี้

- 1) ตัวแปรทั้งสองตัวแปรระดับกลุ่ม
- 2) ตัวแปรทั้งสองเป็นตัวแปรระดับจัดอันดับ
- 3) ตัวแปรทั้งสองเป็นตัวแปรระดับช่วงหรือสูงกว่า
- 4) ตัวแปรตัวหนึ่งเป็นตัวแปรจัดกลุ่มและอีกตัวแปรหนึ่งเป็นตัวแปรจัดอันดับ
- 5) ตัวแปรตัวหนึ่งเป็นตัวแปรจัดกลุ่มและอีกตัวแปรหนึ่งเป็นตัวแปรระดับช่วงหรือระดับที่สูงกว่า
- 6) ตัวแปรตัวหนึ่งเป็นตัวแปรจัดอันดับ และอีกตัวแปรหนึ่งเป็นตัวแปรระดับช่วงหรือระดับที่สูงกว่า

**5. หลักการและวิธีใช้สถิติที่เหมาะสมสำหรับการวิเคราะห์ตัวแปร**

นักสถิติ (statistician) และนักวิจัย (researcher) จะมีหลักการและมีวิธีใช้สถิติที่เหมาะสมสำหรับการวิเคราะห์ตัวแปรของแต่ละกรณีศึกษา ดังนี้ (อ้างถึง สายชล สีนสมบูรณ์ทอง, 2560, หน้า 73-14)

- 1) ตัวแปรทั้งสองเป็นตัวแปรระดับกลุ่มสถิติที่ใช้วิเคราะห์เช่นฟี (the Phi coefficient) โคครานคิว (Cochran Q) เครมเมอร์สวี (Creamer's V) สัมประสิทธิ์คอนทินเจนซี (the Contingency coefficient) และไคสแควร์ (Chi-square) เป็นต้น
- 2) ตัวแปรทั้งสองเป็นตัวแปรระดับจัดอันดับสถิติที่ใช้วิเคราะห์เช่นซอมเมอร์สดี (Somers' d) สเปียร์แมนโร (Spearman's rho) เทา (tau) แกมมา (gamma) และคิมส์ดี (Kim's d) เป็นต้น
- 3) ตัวแปรทั้งสองตัวแปรระดับช่วงหรือสูงกว่าสถิติที่ใช้วิเคราะห์เช่นเพียร์สันโปรดักโมเมนต์ (Pearson's Product Moment) ไบซีเรียล (Biserial) พอยท์ไบซีเรียล (Point Piserial) เททราโคริก (Tetrachoric) การทดสอบซี (Z-test) และการทดสอบที (t-test) เป็นต้น
- 4) ตัวแปรตัวหนึ่งเป็นตัวแปรจัดกลุ่มและอีกตัวแปรหนึ่งเป็นตัวแปรจัดอันดับสถิติที่ใช้วิเคราะห์ เช่น การทดสอบของแมน-วิทนี (The Mann-Whitney U Test) ซอมเมอร์ส ดี (Somer's d) การทดสอบมัธยฐาน (Median test) การทดสอบครุสคัลและวอลลิส (Kruskal-Wallis test) เป็นต้น
- 5) ตัวแปรตัวหนึ่งเป็นตัวแปรจัดกลุ่ม และอีกตัวแปรหนึ่งเป็นตัวแปรช่วงหรือสูงกว่าสถิติที่ใช้วิเคราะห์ เช่น อีต้าสแควร์ ( $\text{Eta}^2$ ) โอเมก้าสแควร์ ( $\text{Omega}^2$ ) เป็นต้น
- 6) ตัวแปรตัวหนึ่งเป็นตัวแปรจัดอันดับ และอีกตัวแปรหนึ่งเป็นตัวแปรช่วงหรือสูงกว่า สถิติที่ใช้วิเคราะห์ เช่น การหาค่าสหสัมพันธ์แบบมัลติซีเรียลของแจนสเปน (Janspen's Coefficient of multiserial correlation)

อนึ่ง ในการเลือกใช้สถิติในการวิจัย นอกจากพิจารณาวัตถุประสงค์ของการวิเคราะห์ข้อมูล (data analysis objectives) จำนวนตัวแปรและระดับการวัดตัวแปรแล้ว นักวิจัย (researcher)

ยังต้องพิจารณาเงื่อนไข (condition) อื่น ๆ เป็นองค์ประกอบด้วยเช่นกัน ได้แก่ 1) ลักษณะของตัวแปร (มีความต่อเนื่องหรือไม่ต่อเนื่อง) 2) ประเภทของตัวแปร (ตัวแปรอิสระหรือตัวแปรตาม) และ 3) ความสัมพันธ์ของตัวแปร (เป็นเส้นตรงหรือเส้นโค้ง) เป็นต้น นอกจากนี้กรณีที่มีตัวแปรมากกว่าสองตัว วิธีการที่ใช้ก็เพิ่มความซับซ้อนมากขึ้น

## 7.5 การแจกแจงความถี่

การแจกแจงความถี่ (frequency distribution) เป็นการจัดข้อมูลตามค่าของข้อมูล (data value) โดยข้อมูลที่มีค่าเท่ากันจะต้องจัดให้อยู่ในกลุ่มเดียวกัน แล้วนับจำนวนข้อมูลแต่ละกลุ่มว่ามีจำนวนเท่าใด การแจกแจงความถี่ของข้อมูล ทำได้ 2 วิธี คือ การแจกแจงความถี่ของค่าแต่ละค่าของข้อมูล และการแจกแจงความถี่ของค่าในแต่ละช่วง (อ้างถึงใน สำเริง จันทรสวรรณ, 2537, หน้า 68)

### 1 การแจกแจงความถี่ของค่าแต่ละค่าของข้อมูล

การแจกแจงความถี่ของค่าแต่ละค่าของข้อมูลหรือการแจกแจงความถี่ของข้อมูลไม่จัดกลุ่ม การแจกแจงข้อมูลวิธีนี้เป็น การนำเอาข้อมูลทั้งหมดมาจัดเรียงตามค่าของข้อมูลโดยอาจเรียงจากข้อมูลที่มีค่ามากที่สุดไปหาน้อยหรือเรียงจากข้อมูลที่มีค่าน้อยไปหาค่ามากที่สุดได้ แล้วนับจำนวนข้อมูลแต่ละค่าโดยทำเครื่องหมายขีด (/) แทนจำนวนข้อมูลแต่ละค่าและทุก ๆ ค่าที่ 5 ทำเครื่องหมายขีดขวางบนเครื่องหมายขีดทั้งสี่ก่อนหน้านี้ทั้งนี้เพื่อความสะดวกในการนับรอยขีดของข้อมูลแต่ละค่า ดังตัวอย่าง 2-1 จากการสอบถามพนักงานของหน่วยงานแห่งหนึ่งเกี่ยวกับระยะเวลาที่ทำงานในหน่วยงานนั้น ปรากฏว่าพนักงาน 30 คน ทำงานคิดระยะเวลาเป็นปีได้ ดังนี้

10	5	8	4	12	25	2	9	20	15
5	12	1	3	5	7	17	10	7	14
5	8	13	9	24	5	6	6	6	4

ระยะเวลา (ปี)	จำนวนข้อมูล หรือรอยขีด	ความถี่
1	/	1
2	/	1
3	/	1
4	//	2
5	///	5
6	////	3
7	//	2
8	//	2
9	//	2
10	//	2
12	//	2
13	/	1

14	/	1
15	/	1
17	/	1
20	/	1
24	/	1
25	/	1
รวม	-	30

ตารางที่ 2-1 ระยะเวลาในการทำงานของพนักงานในหน่วยงานแห่งหนึ่ง

ตัวอย่าง 2-2 จากข้อมูลการอบรมหลักสูตรครูทหาร รุ่นที่ 27 ปรากฏว่านายทหารนักเรียนที่ลงทะเบียนจำแนกตามเพศและสังกัดได้ ดังนี้

สังกัด	เพศ		รวม
	ชาย	หญิง	
โรงเรียนนายร้อยพระจุลจอมเกล้า	5	3	8
โรงเรียนนายเรือ	10	-	10
โรงเรียนนายเรืออากาศ	1	4	5
อื่น ๆ	9	-	9
รวม	25	7	32

ตารางที่ 2-2 จำนวนนายทหารนักเรียนที่เข้าอบรมหลักสูตรครูทหาร รุ่นที่ 27 จำแนกตามเพศ และสังกัด

#### หมายเหตุ\*

ตารางแจกแจงความถี่ตามตัวอย่าง 1-2 เรียกได้ว่า “ตารางทางเดียว” (one-way table) เพราะมีข้อมูลที่สนใจเพียงเรื่องเดียวหรือลักษณะเดียว คือระยะเวลาการทำงาน เท่านั้น ส่วนตัวอย่าง 2-2 ที่มีข้อมูลที่สนใจ 2 เรื่อง คือ เพศและสังกัด ตาราง 2-2 จึงเรียกว่า “ตารางสองทาง” (two-way table) สำหรับตารางที่มีการนำเสนอข้อมูลมากกว่า 2 เรื่อง เรียกได้ว่าเป็น “ตารางหลายทาง” (multi-way table)

#### 2. การแจกแจงความถี่ของค่าในแต่ละช่วง

การแจกแจงความถี่ของค่าในแต่ละช่วงหรือการแจกแจงความถี่ของข้อมูล การจัดกลุ่มข้อมูลบางเรื่องมีค่าที่เป็นไปได้จำนวนมาก ถ้าแจกแจงความถี่ของข้อมูลทุกค่าจะทำให้เสียเวลาในการรวบรวมข้อมูลและไม่สะดวกในการนำเสนอข้อมูล กรณีที่ข้อมูลมีรายละเอียดมากสามารถจัดกลุ่มค่าที่เป็นไปได้เป็นช่วง ๆ จะทำให้ลดค่าที่เป็นไปได้ลงและสะดวกที่จะนำไปวิเคราะห์ต่อไป (อ้างถึงใน วรเชษฐ์ โทอิน, 2560, หน้า 127)

### 3. วิธีการสร้างตารางแจกแจงความถี่ของค่าในแต่ละช่วง มีขั้นตอน ดังนี้

1) พิจารณาการกระจายของข้อมูล (data distribution) หรือความแตกต่างระหว่างค่าของข้อมูลว่ามีค่าแตกต่างกันมากหรือน้อย ซึ่งจะเกี่ยวข้องกับการกำหนดช่วงหรืออันตรภาคชั้นของข้อมูล กล่าวคือถ้าข้อมูลมีการกระจายมาก ควรกำหนดจำนวนอันตรภาคชั้นให้น้อยเพื่อให้มีข้อมูลทุกช่วง

2) กำหนดจำนวนช่วงหรืออันตรภาคชั้น (class interval) ของข้อมูลจำนวน ช่วงจะมีมากหรือน้อยขึ้นอยู่กับผู้กำหนด โดยทั่วไปมักกำหนดจำนวนช่วง ประมาณ 7-12 ช่วง

3) คำนวณความกว้างของช่วงแต่ละช่วง ( $i$ ) โดยคำนวณจากสูตร

$$\text{ความกว้างของช่วง (i)} = \frac{\text{ข้อมูลที่มีค่าสูงสุด} - \text{ข้อมูลที่มีค่าต่ำสุด}}{\text{จำนวนช่วง}}$$

กรณีความกว้างของช่วงที่คำนวณได้เป็นเลขทศนิยม ให้ปัดขึ้นเป็นจำนวนเต็ม (ไม่ว่าทศนิยมจะมีค่ามากกว่าหรือน้อยกว่า .5 ก็ตาม) ถ้า  $i = 2.4$  ให้ใช้  $i = 3$  เป็นต้น

4) เขียนช่วงข้อมูลในแต่ละชั้นโดยอาจเริ่มจากชั้นของข้อมูลที่มีค่าต่ำสุดไปหาชั้นของข้อมูลที่มีค่าสูงสุดหรือเริ่มจากชั้นของข้อมูลที่มีค่ามากไปหาชั้นของข้อมูลที่มีค่าน้อยก็ได้

5) นับจำนวนค่าของข้อมูลที่ตกอยู่ในแต่ละช่วง โดยทำเครื่องหมายขีด (/) แทนค่าของข้อมูลเช่นเดียวกับการสร้างตารางแจกแจงความถี่ของค่าแต่ละค่า

6) หาจำนวนความถี่ของข้อมูลในแต่ละช่วง โดยนับจากรอยขีดจากข้อมูลในตัวอย่าง 2-1 นำมาสร้างตารางแจกแจงความถี่ของค่าในแต่ละช่วงได้ ดังนี้

ขั้นที่ 1 หากการกระจายของข้อมูล

$$\text{ข้อมูลที่มีค่าสูงสุด} - \text{ข้อมูลที่มีค่าต่ำสุด} = 25 - 1 = 24$$

ขั้นที่ 2 กำหนดจำนวนช่วง

กำหนดให้ตารางแจกแจง

ขั้นที่ 3 คำนวณความกว้างของช่วง

$$\text{ความกว้างของช่วง} = \frac{24}{7} = 3.43$$

ดังนั้น ความกว้างของช่วง คือ 4 (จำนวนชั้นอันตรภาค มี 7 ชั้น)

ขั้นที่ 4-6 แสดงในตารางแจกแจงความถี่ ดังนี้

ระยะเวลา (ปี)	รอยขีด	ถวามถี่
1-4	///	5
5-8	/// //	12
9-12	/// /	6
13-16	///	3
17-20	//	2
21-24	/	1
25-28	/	1
รวม		30

ตารางที่ 2-3 ระยะเวลาในการทำงานของพนักงานในหน่วยงานแห่งหนึ่ง  
การนำเสนอตารางแจกแจงความถี่ จะนิยมเสนอข้อมูลของตัวแปร (variable) ที่สนใจและจำนวนหรือความถี่โดยไม่แสดงรอยขีด

ตัวอย่างการบรรยาย

จากตารางแจกแจงข้อมูลระยะเวลาการทำงานของพนักงานในองค์กรแห่งหนึ่ง พบว่ามีพนักงานทำงานในช่วง 1-4 ปี จำนวน 5 คน มีพนักงานที่ทำงานในช่วง 5-8 ปี จำนวน 12 คน พนักงานที่ทำงานในช่วง 9-12 ปี, 13-16 ปี, 17-20 ปี, 21-24 ปี และ 25-28 ปี จำนวน 6, 3, 2, 1 และ 1 คน ตามลำดับ

#### 4. คำที่เกี่ยวข้องกับตารางแจกแจงความถี่

1) อันตรภาคชั้นหรือช่วงของข้อมูล (class interval) หมายถึง ช่วงของข้อมูลในแต่ละชั้น เช่น อันตรภาคชั้น 1-4 อันตรภาคชั้น 5-8 เป็นต้น (อ้างถึงใน วรเชษฐ์ โทอิน, 2560, หน้า 130)

2) จุดกึ่งกลาง (mid-point) คือ ค่าเฉลี่ยระหว่างขอบเขตล่างและขอบเขตบนของแต่ละอันตรภาคชั้น เช่น อันตรภาคชั้น 5-8 มีจุดกึ่งกลางเป็น  $\frac{5+8}{2} = 6.5$

3) ขีดจำกัดชั้น (class limit)

- ขีดจำกัดล่าง (lower limit) หมายถึง ค่าต่ำสุดของอันตรภาคชั้นและ
- ขีดจำกัดบน (upper limit) หมายถึง ค่าสูงสุดของอันตรภาคชั้น
- (เช่น อันตรภาคชั้น 5-8 มี 5 เป็นขีดจำกัดล่าง และ 8 เป็นขีดจำกัดบน)

4) ขอบเขต (class boundary)

ขอบเขตล่าง (lower boundary) หมายถึง ขีดจำกัดล่างที่แท้จริงของข้อมูลซึ่งเป็นจุดเริ่มต้นของชั้น

ขอบเขตบน (upper boundary) หมายถึง ขีดจำกัดบนที่แท้จริงของข้อมูลซึ่งเป็นจุดสิ้นสุดของชั้น เช่น อันตรภาคชั้น 5-8 มี 4.5 เป็นขอบเขตล่าง และ 5.8 เป็นขอบเขตบน โดยทั่วไป ขอบเขตล่าง (lower boundary) หาได้จาก “ค่าเฉลี่ยระหว่างค่าที่น้อยที่สุดของอันตรภาคชั้น” กับ “ค่าที่มากที่สุดของอันตรภาคชั้นที่อยู่ถัดลงไป” และขอบเขตบน (upper boundary) หาได้จาก “ค่าเฉลี่ยระหว่างค่าที่มากที่สุดของอันตรภาคชั้น” กับ “ค่าที่น้อยที่สุดของอันตรภาคชั้นที่อยู่ถัดขึ้นไป” จากข้อมูลการทำงานของพนักงานตามตัวอย่างที่ 2-3 แสดงค่าที่เกี่ยวข้องกับตารางแจกแจงความถี่ได้ดังนี้

อันตรภาคชั้น	จุดกึ่งกลาง	ขีดจำกัดล่าง	ขีดจำกัดบน	ขอบเขตล่าง	ขอบเขตบน
1-4	2.5	1	4	0.5	4.5
5-8	6.5	5	8	4.5	8.5
9-12	10.5	9	12	8.5	12.5
13-16	14.5	13	16	12.5	16.5
17-20	18.5	17	20	16.5	20.5
21-24	22.5	21	24	20.5	24.5
25-28	26.5	25	28	24.5	28.5

ตารางที่ 2-4 จุดกึ่งกลางและขอบเขตของระยะเวลาการทำงานของพนักงานในหน่วยงานแห่งหนึ่ง

## 5. การแจกแจงความถี่สะสม

ข้อมูลจากตารางแจกแจงความถี่จะบอกให้ทราบว่าข้อมูลในแต่ละชั้นมีจำนวนอยู่เท่าใด ถ้าต้องการรายละเอียดเพิ่มขึ้นว่ามีข้อมูลจำนวนเท่าใด มีค่าสูงกว่าหรือต่ำกว่าค่าใดค่าหนึ่งสามารถบอกได้จากค่าความถี่สะสม (cumulative frequency) ความถี่สะสมของข้อมูลใดคือผลรวมของความถี่ของชั้นที่กำหนดให้กับความถี่ของชั้นที่มีค่าต่ำกว่าทั้งหมด เช่น จากข้อมูลระยะเวลาการทำงานของพนักงาน ในตาราง 2-3 สามารถแจกแจงความถี่สะสมได้ดังนี้ (อ้างอิงใน วรเชษฐ์ โทอิน, 2560, หน้า 132)

ระยะเวลา (ปี)	ความถี่	ความถี่สะสม
1-4	5	5
5-8	12	17
9-12	6	23
13-16	3	26
17-20	2	28
21-24	1	29
25-28	1	30
รวม		30

ตาราง 2-5 ความถี่และความถี่สะสมของระยะเวลาการทำงานของพนักงาน

จากตาราง 5-2 เราสามารถบรรยายได้ว่า พนักงานที่ทำงาน 1-8 ปี มีจำนวน 17 คน พนักงานที่ทำงานระหว่าง 5-16 ปี มีจำนวน 21 คน (12+6+3) พนักงานที่ทำงานไม่เกิน 20 ปี มี 28 คน (5+12+6+3+2) เป็นต้น

## 6. การแจกแจงความถี่สัมพัทธ์(relative cumulative frequency)

ความถี่สัมพัทธ์ของข้อมูลใดหมายถึงสัดส่วนของความถี่ของข้อมูลนั้นกับความถี่ทั้งหมด ความถี่สัมพัทธ์อาจแสดงในรูปทศนิยมหรือร้อยละได้ แต่โดยทั่วไปนิยมแสดงในรูปร้อยละจากข้อมูลในตาราง 2-5 แสดงค่าความถี่สัมพัทธ์และความถี่สะสมสัมพัทธ์ได้ ดังนี้

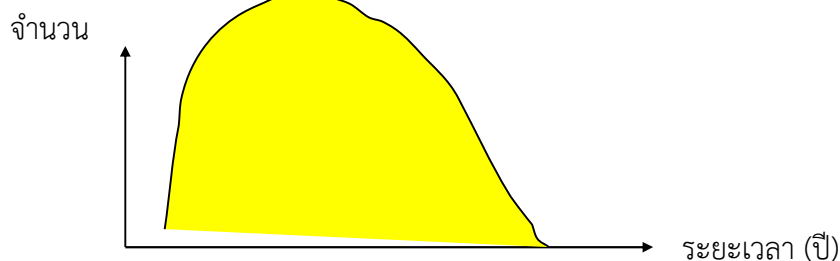
ระยะเวลา (ปี)	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์	ความถี่สะสมสัมพัทธ์
1-4	5	5	16.67	16.67
5-8	12	17	40.00	56.67
9-12	6	23	20.00	76.67
13-16	3	26	10.00	86.67
17-20	2	28	6.67	93.33
21-24	1	29	3.33	96.67
25-28	1	30	3.33	100.00
รวม	30		100.00	

ตารางที่ 2-6 ความถี่สัมพัทธ์และความถี่สะสมสัมพัทธ์ของระยะเวลาในการทำงานของพนักงาน



## 7. โค้งความถี่

โค้งความถี่ (frequency curve) คือเส้นโค้งที่ได้จากปรับปรุงรูปหลายเหลี่ยมของความถี่ให้เรียบ การปรับใช้หลักการทำให้พื้นที่ภายใต้เส้นโค้งที่ปรับแล้วมีขนาดใกล้เคียงกับพื้นที่ของรูปหลายเหลี่ยมของความถี่ให้มากที่สุด รูปเส้นโค้งของความถี่ของระยะเวลาในการทำงานของพนักงานหน่วยงานแห่งหนึ่ง แสดงดังภาพ 2-3

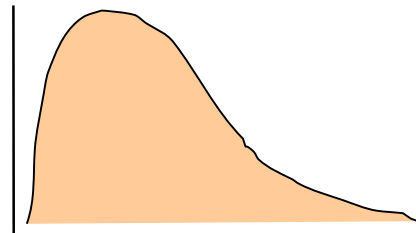


โค้งของความถี่ ซึ่งแสดงลักษณะการแจกแจงของข้อมูล มีหลายรูปแบบ ดังนี้

- 1) โค้งปกติ (normal curve) โค้งมีลักษณะคล้ายรูปประฆังคว่ำ ความถี่ตรงส่วนกลางจะมากและค่อยๆ ลดลงทั้งสองข้าง เส้นโค้งการแจกแจงมีลักษณะสมการ (symmetry)
- 2) โค้งเบ้ทางบวก (positive skewed curve) โค้งมีพื้นที่ใต้โค้งด้านซ้ายน้อยกว่าด้านขวา
- 3) โค้งเบ้ทางลบ (negative skewed curve) โค้งมีพื้นที่ใต้โค้งด้านซ้ายมากกว่าด้านขวา
- 4) โค้งรูปตัวยู (u-shaped) โค้งมีลักษณะคล้ายอักษรตัวยู
- 5) โค้งรูปตัวเจ (j-shaped) โค้งมีลักษณะคล้ายอักษรตัวเจ
- 6) โค้งรูปตัวเอส (s-shaped) โค้งมีลักษณะคล้ายอักษรตัวเอส



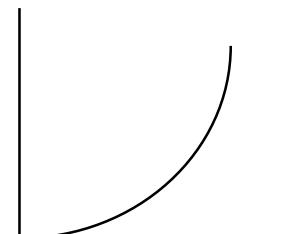
โค้งปกติ



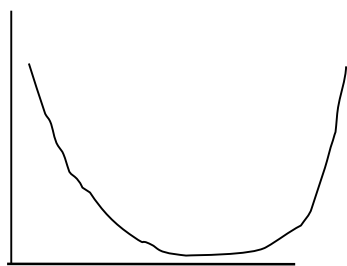
โค้งเบ้ทางบวก



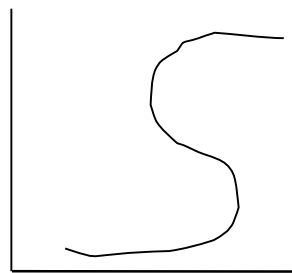
โค้งเบ้ทางลบ



โค้งรูปตัวเจ



โค้งรูปตัวยู



โค้งรูปตัวเอส

## 8 ข้อพิจารณาในการแจกแจงความถี่

1) การแจกแจงความถี่ (frequency distribution) สำหรับค่าแต่ละค่าจะทำให้ผลการวิเคราะห์ข้อมูลมีความถูกต้องมากกว่าการแจกแจงความถี่เป็นช่วง

2) การวิเคราะห์ข้อมูลที่ได้จากการแจกแจงความถี่เป็นช่วง ถ้าช่วงหรือความกว้างของอันตรภาคชั้นมีน้อย จะให้ผลการวิเคราะห์ที่ถูกต้องมากกว่าเมื่อความกว้างของอันตรภาคชั้นมีมาก

3) ถ้าข้อมูลที่น่ามาวิเคราะห์มีจำนวนน้อย ไม่จำเป็นต้องแจกแจงความถี่ของข้อมูล โดยเฉพาะไม่จำเป็นต้องแจกแจงความถี่ของข้อมูลเป็นช่วง เนื่องจากจะทำให้ผลการวิเคราะห์มีความถูกต้องน้อยลง

4) ในการแจกแจงความถี่ ความกว้างของแต่ละอันตรภาคชั้นไม่จำเป็นต้องเท่ากันทุกชั้น แต่ถ้าความกว้างของอันตรภาคชั้นเท่ากันจะทำให้สะดวกในการวิเคราะห์

5) ในการแจกแจงความถี่ของข้อมูล ไม่ควรให้อันตรภาคชั้นแรกหรือชั้นสุดท้ายเป็นชั้นเปิด เนื่องจากถ้าเป็นชั้นเปิดจะไม่สามารถนำไปวิเคราะห์ข้อมูลชั้นสูงได้เนื่องจากหาจุดกึ่งกลางไม่ได้

## 7.6 การวัดแนวโน้มเข้าสู่ส่วนกลาง

### 1. การวัดแนวโน้มเข้าสู่ส่วนกลาง

การวัดแนวโน้มเข้าสู่ส่วนกลางเป็นระเบียบวิธีทางสถิติ (statistical methods) ในการหาค่าเพียงค่าเดียวที่จะใช้เป็นตัวแทนของข้อมูลทั้งหมด ค่าที่หาได้นี้จะทำให้สามารถทราบถึงลักษณะของข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ ค่าที่หาได้นี้จะเป็นค่ากลาง ๆ เรียกว่า “ค่ากลาง” ประเภทของการวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดแนวโน้มเข้าสู่ส่วนกลางมีอยู่หลายวิธีด้วยกัน ที่นิยมกันมากได้แก่ มัชฌิมเลขคณิต (arithmetic mean) มัธยฐาน (median) และฐานนิยม (mode)

### 2. ค่าเฉลี่ยเลขคณิต (arithmetic mean)

ค่าเฉลี่ยเลขคณิต (arithmetic mean) หมายถึง การหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลทั้งหมดการหาค่าเฉลี่ยเลขคณิตสามารถหาได้ 2 วิธี

1. ค่าเฉลี่ยเลขคณิตของข้อมูลที่ไม่ได้แจกแจงความถี่สามารถคำนวณได้จากสูตร

$$\bar{X} = \frac{\sum x}{n}$$

เมื่อ  $\bar{X}$  (เอ็กซ์บาร์) คือ ค่าเฉลี่ยเลขคณิต

$\sum x$  คือ ผลบวกของข้อมูลทุกค่า

$n$  คือ จำนวนข้อมูลทั้งหมด

## วิชาการเบี่ยงวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

Ex. จากการสอบถามอายุของนักเรียนกลุ่มหนึ่งเป็น ดังนี้ 14, 16, 14, 17, 16, 14, 18, 17  
จงหาค่าเฉลี่ยเลขคณิตของอายุนักเรียนกลุ่มนี้

$$\text{วิธีทำ } \bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1}{8} = \frac{8}{8}$$

$$\bar{x} = 15.75$$

ดังนั้น ค่าเฉลี่ย อายุนักเรียนกลุ่มนี้ = 15.75 ปี

2. ค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่สามารถคำนวณได้จากสูตร

$$\bar{x} = \frac{\sum fx}{n}$$

เมื่อ  $\bar{x}$  (เอ็กซ์บาร์) คือ ค่าเฉลี่ยเลขคณิต

$f$  คือ ความถี่ของข้อมูล

$x$  คือ ค่าของข้อมูล (ในกรณีการแจกแจงความถี่ไม่เป็นอันตรภาคชั้น) หรือจุดกึ่งกลางของอันตรภาคชั้น (ในกรณีการแจกแจงความถี่เป็นอันตรภาคชั้น) หาได้จาก

$$\frac{\text{ค่าสูงสุดของอันตรภาคชั้น} + \text{ค่าต่ำสุดของอันตรภาคชั้น}}{2}$$

$n$  คือ ผลรวมความถี่ทั้งหมด หรือจำนวนข้อมูลทั้งหมด

2.1 การหาค่าเฉลี่ยเลขคณิตข้อมูลที่แจกแจงความถี่ในกรณีที่ข้อมูลไม่เป็นอันตรภาคชั้น

Ex. จากการสอบถามอายุของนักเรียนกลุ่มหนึ่งเป็น ดังนี้ 14, 16, 14, 17, 16, 14, 18, 17  
จงหาค่าเฉลี่ยเลขคณิตของอายุนักเรียนกลุ่มนี้

วิธีทำ สร้างตารางแจกแจงความถี่ข้อมูล

ค่าข้อมูล ( $x$ )	ความถี่ ( $f$ )	( $f$ )
14	3	42
16	2	32
17	3	34
18	1	18
	$n = 8$	$\sum fx = 126$

$$\text{แทนค่าสูตร } \bar{x} = \frac{\sum fx}{n}$$

$$\bar{x} = \frac{126}{8}$$

$$\bar{x} = 15.75$$

ดังนั้น ค่าเฉลี่ย อายุนักเรียนกลุ่มนี้ = 15.75 ปี

2.2 การหาค่าเฉลี่ยเลขคณิตข้อมูลที่แจกแจงความถี่ในกรณีที่ข้อมูลเป็นอันตรภาคชั้น (class interval) หรือเรียกสั้นๆ ว่า ชั้น หมายถึง ช่วงของคะแนนในแต่ละพวกที่แบ่ง

Ex. จากข้อมูลในตารางแจกแจงความถี่ จงหาค่าเฉลี่ยเลขคณิต

คะแนน	ความถี่
5-9	3
10-14	4
15-19	3
20-24	7
25-29	6
30-34	4
35-39	2
40-44	3
	N = 32

วิธีทำ

คะแนน	ความถี่	จุดกึ่งกลาง อันตรภาคชั้น	(f)
5-9	3	7	21
10-14	4	12	48
15-19	3	17	51
20-24	7	22	154
25-29	6	27	162
30-34	4	32	128
35-39	2	37	74
40-44	3	42	126
.	N=32	.	$\Sigma fx$ = 764

$$\text{แทนค่าสูตร } \bar{X} = \frac{\Sigma f}{n}$$

$$\bar{X} = \frac{7}{2}$$

$$\bar{X} = 23.86 \text{ ปี}$$

ดังนั้นค่าเฉลี่ย อายุนักเรียนกลุ่มนี้ = 23.86 ปี

### 3. มัธยฐาน (median)

มัธยฐาน หมายถึง ค่ากึ่งกลางของข้อมูลชุดนั้นหรือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลชุดนั้น เมื่อได้จัดเรียงค่าของข้อมูลจากน้อยที่สุด ไปหามากที่สุด หรือจากมากที่สุดไปหาน้อยที่สุด ค่ากึ่งกลางจะเป็นตัวแทนที่แสดงว่ามีข้อมูลที่มากกว่าและน้อยกว่านี้อยู่ 50 %

การหาค่ามัธยฐาน สามารถหาได้ 2 วิธี

1. การหามัธยฐานของข้อมูลที่ไม่แจกแจงความถี่ ซึ่งมีวิธีหาได้ดังนี้เรียงข้อมูลจากน้อยไปมาก หรือจากมากไปน้อย จากสูตรการคิด ดังนี้

$$\frac{n+1}{2} \text{ โดยที่เมื่อ } n = \text{จำนวนข้อมูลทั้งหมด}$$

Ex. จงหามัธยฐานของข้อมูลต่อไปนี้ 9, 10, 5, 11, 14, 6, 16, 17, 13

วิธีทำ เรียงข้อมูลที่มีค่าน้อยที่สุดไปหาข้อมูลที่มีค่ามากที่สุด คือ 5, 6, 9, 10, 11, 13, 14, 16, 17

$$\text{หาตำแหน่งมัธยฐาน } \frac{n+1}{2} = \frac{9+1}{2} = 5$$

ดังนั้น ค่ามัธยฐานของข้อมูล = 11 (9+1 = 10 ÷ 2 = 5)

Ex. จงหามัธยฐานของข้อมูลต่อไปนี้ 40, 35, 24, 28, 26, 29, 36, 31, 42, 20, 23, 32

วิธีทำ เรียงข้อมูลจากข้อมูลที่มีค่าน้อยที่สุดไปหาข้อมูลที่มีค่ามากที่สุดคือ 20, 23, 24, 26, 28, 29, 31, 32, 35, 36, 40, 42, ซึ่ง  $n = 12$

$$\begin{aligned} \text{ตำแหน่งมัธยฐาน} &= \frac{n+1}{2} \\ &= \frac{12+1}{2} \\ &= 6.5 \end{aligned}$$

มัธยฐานอยู่ในตำแหน่ง ที่ 6.5 หมายถึง อยู่ระหว่าง 29 กับ 31

$$\text{มัธยฐานเท่ากับ } \frac{29+31}{2}$$

มัธยฐานคือ 30

2. การหามัธยฐานของข้อมูลที่แจกแจงความถี่

$$Mdn = L + i \left[ \frac{\frac{n}{2} - \sum fl}{fm} \right]$$

คำนวณได้จากสูตร

เมื่อ Mdn = มัธยฐาน (median)

L = ชีตจำกัดล่างที่แท้จริงของชั้นที่มีมัธยฐานอยู่

i = ความกว้างของอันตรภาคชั้น

$\sum fl$  = ความถี่สะสมชั้นที่อยู่ก่อนชั้นที่มีมัธยฐานไปหาคะแนนน้อย

f = ความถี่ของคะแนนในชั้นที่มีมัธยฐาน

$\frac{n}{2}$  = คือตำแหน่งมัธยฐาน

Ex. จากข้อมูลในตารางแจกแจงความถี่ จงหาค่ามัธยฐาน

คะแนน	ความถี่
5-9	3
10-14	4
15-19	3
20-24	7

25-29	6
30-34	4
35-39	2
40-44	3
	n = 32

#### 4. การแจกแจงความถี่สะสม

ความถี่สะสม (cumulative frequency) ของค่าที่เป็นไปได้ค่าใดหรืออันตรภาคชั้นใด คือ “ผลรวมของความถี่ของค่านั้นหรือของอันตรภาคชั้นนั้น” กับ “ความถี่ของค่าหรือของอันตรภาคชั้นที่มีช่วงคะแนนต่ำกว่าทั้งหมดหรือสูงกว่าทั้งหมดอย่างใดอย่างหนึ่ง” การแจกแจงความถี่สะสมคือการหาผลรวมของความถี่แต่ละชั้นต่อเนื่องกันมาตั้งแต่ต้นตามลำดับ จากชั้นที่มีค่าน้อยที่สุดไปหาชั้นที่มีค่ามากที่สุด หรือจากชั้นที่มีค่ามากที่สุดไปหาชั้นที่มีค่าน้อยที่สุด

ตัวอย่างที่ 1 จงสร้างตารางแจกแจงความถี่สะสม จากตารางแจกแจงความถี่ต่อไปนี้

คะแนน	10-19	20-29	30-39	40-49	50-59	60-69	70-79
ความถี่	7	13	18	22	26	10	4

#### วิธีทำ

คะแนน	ความถี่	ความถี่สะสม
10-19	7	7
20-29	13	$7 + 13 = 20$
30-39	18	$20 + 18 = 38$
40-49	22	$38 + 22 = 60$
50-59	26	$60 + 26 = 86$
60-69	10	$86 + 10 = 96$
70-79	4	$96 + 4 = 100$

จากตารางแสดงว่า

- 1) คะแนนที่มีค่าต่ำกว่า 9.5 มี 0 จำนวน
- 2) คะแนนที่มีค่าต่ำกว่า 19.5 มี 7 จำนวน
- 3) คะแนนที่มีค่าต่ำกว่า 29.5 มี 20 จำนวน

ทั้งนี้ วัตถุประสงค์สำคัญในการสร้างตารางแจกแจงความถี่ก็เพื่อเปรียบเทียบความถี่ของค่าจากการสังเกตทั้งหมดที่ตกอยู่ในแต่ละค่าที่เป็นไปได้หรือในแต่อันตรภาคชั้น



### 5. การแจกแจงความถี่สัมพัทธ์

ความถี่สัมพัทธ์ (relative frequency) ของค่าที่เป็นไปได้ค่าใดหรืออันตรภาคชั้นใด คือ “อัตราส่วนระหว่างความถี่ของค่านั้น” หรือของ “อันตรภาคชั้นนั้น” กับ “ผลรวมของความถี่ทั้งหมด” ซึ่งอาจจะแสดงในรูปเศษส่วน (fraction) ทศนิยม (decimal) หรือร้อยละ (percentage) ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 2 จงสร้างตารางแจกแจงความถี่สัมพัทธ์จากตารางแจกแจงความถี่ในกลุ่มของตัวอย่างที่ 2

#### วิธีทำ

คะแนน	ความถี่	ความถี่สัมพัทธ์	ร้อยละของ ความถี่สัมพัทธ์
10-19	7	0.07 ( $7 \div 100$ )	7.0
20-29	13	0.13 ( $13 \div 100$ )	13.0
30-39	18	0.18 ( $18 \div 100$ )	18.0
40-49	22	0.22 ( $22 \div 100$ )	22.0
50-59	26	0.26 ( $26 \div 100$ )	26.0
60-69	10	0.10 ( $10 \div 100$ )	10.0
70-79	4	0.04 ( $4 \div 100$ )	4.0
รวม	100	1.00	100

ทั้งนี้ วัตถุประสงค์สำคัญในการสร้างตารางแจกแจงความถี่สัมพัทธ์ เพื่อหาค่าความถี่ของแต่ละค่าที่เป็นไปได้หรือแต่ละอันตรภาคชั้นมีความถี่สะสมเป็นจำนวนมากน้อยเพียงใดเมื่อเปรียบเทียบกับความถี่ทั้งหมด

### 6. การแจกแจงความถี่สะสมสัมพัทธ์

การแจกแจงความถี่สะสมสัมพัทธ์ (relative cumulative frequency) ของค่าที่เป็นไปได้ค่าใดหรืออันตรภาคชั้นใด คือ อัตราส่วนระหว่างความถี่สะสมของค่านั้น หรือของอันตรภาคชั้นนั้นกับผลรวมของความถี่ทั้งหมดซึ่งอาจแสดงในรูปเศษส่วน (fraction) ทศนิยม (decimal) หรือร้อยละ (percentage) ดังตัวอย่างต่อไปนี้

คะแนน	ความถี่	ความถี่สะสม	ความถี่สะสมสัมพัทธ์	ร้อยละของ ความถี่สัมพัทธ์
10-20	7	7	0.07	7.0
20-29	13	20	0.20	20.0
30-39	18	38	0.38	38.0
40-49	22	60	0.60	60.0
50-59	26	86	0.86	86.0
60-69	10	96	0.96	96.0
70-79	4	100	1.00	100.0

## วิชาการเบี่ยงวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

ทั้งนี้ ตารางแจกแจงความถี่สะสมสัมพัทธ์ (relative cumulative frequency) ใช้เพื่อหาว่าแต่ละค่าที่เป็นไปได้หรือแต่ละอันตรภาคชั้น (interval scale) มีความถี่สะสมเป็นจำนวนมากน้อยเพียงใดเมื่อเทียบกับความถี่ทั้งหมด เช่น จำนวนนักเรียนที่ทำคะแนนสอบได้ต่ำกว่า 40 คะแนน คิดเป็นร้อยละ 38 ของนักเรียนที่เข้าสอบทั้งหมด

### วิธีทำ

วิธีหาความถี่สะสม

คะแนน	ความถี่	ความถี่สะสม
5-9	3	3
10-14	4	7
15-19	3	10
20-24	7	17
25-29	6	23
30-34	4	27
35-39	2	29
40-44	3	32
	n = 32	

หาตำแหน่งมัธยฐานจาก  $\frac{n}{2} = \frac{32}{2} = 16$  ค่ามัธยฐานที่อยู่ในชั้น 20 - 24

$$Mdn = L + i \left[ \frac{\frac{n}{2} - \sum f_l}{f_m} \right]$$

จากสูตร

$$L = 20 - 0.5 = 19.5$$

$$i = 5$$

$$\sum f_l = 10$$

$$f_m = 7$$

$$Mdn = 19.5 + 5 \left[ \frac{16 - 10}{7} \right] = 19.5 + 4.2 = 23.7$$

แทนค่าในสูตร

มัธยฐาน คือ 23.7

### 7. ฐานนิยม (Mode)

ฐานนิยมหมายถึง ค่าของคะแนนที่ซ้ำกันมากที่สุดหรือ ค่าคะแนนที่มีความถี่สูงที่สุดในข้อมูลชุดนั้นการหาค่าฐานนิยม สามารถหาได้ 2 วิธี

1. ฐานนิยมของข้อมูลที่ไม่แจกแจงความถี่พิจารณาค่าของข้อมูลที่ซ้ำกันมากที่สุด คือฐานนิยม

Ex. จงหาฐานนิยมของข้อมูลต่อไปนี้ 3, 2, 4, 5, 6, 4, 8, 4, 7, 10

1) ข้อมูลที่ซ้ำกันมากที่สุดคือ 4 ฐานนิยมคือ 4

2) ข้อมูลบางชุดอาจมีฐานนิยม 2 ค่า เช่น 10, 14, 12, 10, 11, 13, 12, 14, 12, 10

3) ข้อมูลที่ซ้ำกันมากที่สุดคือ 10 กับ 12 ฐานนิยม คือ 10 กับ 12

4) ข้อมูลบางชุดอาจจะไม่มีฐานนิยมซึ่ง ได้แก่ข้อมูลที่ไม่มีรายการซ้ำกันเลย เช่น 8, 9, 10, 11, 13, 15

## 2. ฐานนิยมของข้อมูลที่แจกแจงความถี่

$$Mo = L + i \left[ \frac{\Delta 1}{\Delta 1 + \Delta 2} \right]$$

คำนวณได้จากสูตร

เมื่อ  $Mo$  = ฐานนิยม (mode)

$L$  = ขีดจำกัดล่างของคะแนนในชั้นที่มีความถี่สูงสุด

$i$  = ความกว้างอันตรภาคชั้น

$\Delta 1$  = ผลต่างของความถี่มากที่สุดกับความถี่ของชั้นก่อนหน้า

$\Delta 1 \Delta 2$  = ผลต่างของความถี่มากที่สุดกับความถี่ของชั้นที่ถัดไปทางคะแนนมาก

Ex. จากข้อมูลในตารางแจกแจงความถี่ จงหาฐานนิยม

คะแนน	ความถี่
5-9	3
10-14	4
15-19	3
20-24	7
25-29	6
30-34	4
35-39	2
40-44	3
	$N = 32$

## วิธีทำ

ค่าฐานนิยมอยู่ในอันตรภาคชั้น 20 - 24 (ค่าที่มีความถี่มากที่สุด)

$$Mo = L + i \left[ \frac{\Delta 1}{\Delta 1 + \Delta 2} \right]$$

จากสูตร

$$1. L = 20 - 0.5 = 19.5$$

$$2. i = 5$$

$$3. \Delta 1 = 7 - 3 = 4$$

$$4. \Delta 2 = 7 - 6 = 1$$

$$Mo = 19.5 + 5 \left[ \frac{4}{4+1} \right]$$

แทนค่าในสูตร

$$= 19.5 + 4 = 23.5$$

ดังนั้น ฐานนิยมของข้อมูลในตารางนี้คือ 23.5

ในกรณีที่หาค่ามัธยฐานและมัธยฐานได้แล้ว สามารถที่จะนำมาคำนวณหาฐานนิยมได้

โดยใช้สูตร  $Mode = 3 \text{ Median} - 2 \text{ Mean}$

## 7.7 การวัดการกระจาย (measure of dispersion)

การใช้สถิติเกี่ยวกับการวัดแนวโน้มเข้าสู่ส่วนกลาง (measure of dispersion) ซึ่งเป็นค่าที่ทำหน้าที่เป็นตัวแทนกลุ่มข้อมูลเพียงอย่างเดียว เมื่อแปลความหมายข้อมูล (interpret data) จึงยังไม่สมบูรณ์ (incomplete) ไม่ชัดเจน (unclear) และมีโอกาสคลาดเคลื่อนได้ (error) สิ่งที่เราควรนำมาพิจารณาควบคู่ไปกับการวัดแนวโน้มเข้าสู่ส่วนกลางก็คือ ลักษณะการกระจายของกลุ่มข้อมูล ซึ่งสถิติที่ใช้คือการวัดการกระจาย (measure of dispersion) การที่ข้อมูลแต่ละชุดมีค่าต่างกัันนั้น เราเรียกว่า “ข้อมูลมีการกระจาย” ถ้าข้อมูลชุดนั้นประกอบด้วยค่าแตกต่างกันมาก เรียกว่า “ข้อมูลมีการกระจายมาก” ถ้าข้อมูลชุดนั้นประกอบด้วยค่าต่าง ๆ แตกต่าง กัันน้อยหรือมีค่าใกล้เคียงกัน เรียกว่า “ข้อมูลมีการกระจายน้อย” ถ้าข้อมูลนั้นประกอบด้วยค่าต่าง ๆ เท่ากันหมด เรียกว่า “ข้อมูลไม่มีการกระจาย” (อ้างถึงใน วรเชษฐ์ โพธิ์, 2560, หน้า 133 )

ข้อมูลชุดที่ 1 : 9, 12, 37, 73, 105

ข้อมูลชุดที่ 2 : 52, 60, 63, 61, 65

ข้อมูลชุดที่ 3 : 35, 35, 35, 35, 35

จากข้อมูลทั้ง 3 ชุด เมื่อเปรียบเทียบแล้วพบว่า ข้อมูลชุดที่ 1 มีการกระจายมากที่สุด ข้อมูลชุดที่ 2 มีการกระจายรองลงมา ส่วนข้อมูลชุดที่ 3 ไม่มีการกระจาย ในการเปรียบเทียบข้อมูลหลายๆ ชุดว่าแตกต่างกันหรือไม่ ควรจะต้องพิจารณาถึงค่าเฉลี่ยและการกระจายของข้อมูลควบคู่กันไปด้วย เพื่อจะช่วยให้สรุปหรือแปลความหมายได้อย่างถูกต้อง เช่น เด็กนักเรียนกลุ่มหนึ่ง วัดคะแนนสอบวิชาภาษาไทยได้ 75, 87, 115, 118, 130 เด็กนักเรียนกลุ่มสอง วัดคะแนนสอบวิชาภาษาไทยได้ 100, 100, 105, 110, 110 ค่าเฉลี่ยของคะแนน 2 ชุดนี้เท่ากัน คือ 105 ถ้าพิจารณาเฉพาะค่าเฉลี่ยจะสรุปได้ว่านักเรียน 2 กลุ่มนี้ มีคะแนนสอบวิชาภาษาไทยอยู่ในระดับเดียวกัน แต่เมื่อพิจารณาจากคะแนนแต่ละชุดจะพบว่า คะแนนสอบวิชาภาษาไทยของนักเรียนกลุ่มหนึ่ง แตกต่างกัันมากกว่า คะแนนสอบวิชาภาษาไทยของนักเรียนในกลุ่มที่สอง นั่นคือตามข้อสรุปแล้วคะแนนสอบวิชาภาษาไทยของนักเรียน 2 กลุ่มนี้แตกต่างกัน ดังนั้น จึงสรุปได้ว่าถ้าต้องการบรรยายลักษณะของข้อมูลให้ถูกต้องสมบูรณ์จะต้องวัดแนวโน้มเข้าสู่ส่วนกลางควบคู่ไปกับการวัดการกระจายด้วยเสมอการวัดการกระจาย แบ่งได้ 2 ประเภท คือ

1. การวัดการกระจายสัมบูรณ์ (absolute variation) เป็นการวัดการกระจายข้อมูลเพียงชุดเดียวมี ดังนี้คือ 1) พิสัย (range : R) และ 2) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation : S.D., S, s)

### 1) พิสัย (range : R)

พิสัย หมายถึง การหาการกระจายของข้อมูลโดยนำข้อมูลที่มีค่าสูงที่สุด ลบกับข้อมูลที่มีค่าต่ำที่สุด เพื่อให้ได้ค่าที่เป็นช่วงของการกระจาย ซึ่งสามารถบอกถึงความกว้างของข้อมูลชุดนั้น ๆ สำหรับสูตรที่ใช้ในการหาพิสัยคือพิสัย (R) =  $X_{\max} - X_{\min}$

Ex 1. จงหาพิสัยจากข้อมูลชุดนี้ 25, 19, 32, 29, 19, 21, 22, 31, 19, 20, 15, 22, 23, 20

### วิธีทำ

$$\begin{aligned}\text{สูตรการหาค่าพิสัย (R)} &= X_{\max} - X_{\min} = 32 - 15 \\ &= 17\end{aligned}$$

ข้อมูลชุดนี้มีพิสัย (R) เท่ากับ 17

ดังนั้น ความแตกต่างของข้อมูลสูงสุดกับข้อมูลต่ำสุดมีค่าเท่ากับ 17

## 2) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation : S.D.,s)

ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) เป็นค่าวัดการกระจายที่สำคัญทางสถิติ เพราะเป็นค่าที่ใช้บอกถึงการกระจายของข้อมูลได้ดีกว่าค่าพิสัย (range) และ ค่าส่วนเบี่ยงเบนเฉลี่ย (standard deviation) การหาส่วนเบี่ยงเบนมาตรฐานสามารถหาได้ 2 วิธี คือ การหาส่วนเบี่ยงเบนมาตรฐาน (S.D.) ในกรณีข้อมูลไม่ได้มีการแจกแจงความถี่ และการหาส่วนเบี่ยงเบนมาตรฐาน (S.D.) ในกรณีข้อมูลมีการแจกแจงความถี่ (อ้างถึงใน วรเชษฐ์ โทอิน, 2560)

1. การหาส่วนเบี่ยงเบนมาตรฐาน (S.D.) ในกรณีข้อมูลไม่ได้มีการแจกแจงความถี่ สามารถหาได้จากสูตร

$$\begin{aligned} \text{สูตรที่ 1} \quad S.D. &= \sqrt{\frac{(x - \bar{x})^2}{n-1}} \quad \text{หรือ} \\ \text{สูตรที่ 2} \quad S.D. &= \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \end{aligned}$$

เมื่อ S.D. คือ ส่วนเบี่ยงเบนมาตรฐาน

$x$  คือ ข้อมูล (ตัวที่ 1,2,3...,n)

$\bar{x}$  คือ ค่าเฉลี่ยเลขคณิต

$n$  คือ จำนวนข้อมูลทั้งหมด

หมายเหตุ ในกรณีที่  $\bar{x}$  เป็นทศนิยมทำให้เกิดความยุ่งยากในการคำนวณ จึงควรเลือกใช้สูตรที่ 2

Ex2. จากข้อมูลต่อไปนี้จงหาส่วนเบี่ยงเบนมาตรฐาน 1, 2, 4, 6, 8, 9

วิธีทำ

ใช้สูตรที่ 2 
$$S.D. = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$$

หาค่า 
$$\sum x^2 = 1^2 + 2^2 + 4^2 + 6^2 + 8^2 + 9^2$$

$$\sum x^2 = 1 + 4 + 16 + 36 + 64 + 81$$

$$\sum x^2 = 212$$

หาค่า  $\sum x = 1 + 2 + 4 + 6 + 8 + 9$

$$\sum x = 30$$

$$(\sum x)^2 = 30^2$$

$$(\sum x)^2 = 900$$

$$n = 6$$

แทนค่าในสูตร 
$$S.D. = \sqrt{\frac{6(212) - 900}{6(6-1)}}$$

$$S.D. = \sqrt{\frac{372}{30}}$$

$$S.D. = \sqrt{12.4}$$

$$S.D. = 3.52$$

3). การหาส่วนเบี่ยงเบนมาตรฐาน (S.D.) ในกรณีข้อมูลมีการแจกแจงความถี่ สามารถหาได้จากสูตร

$$1. \quad S.D. = \sqrt{\frac{\sum f(x-\bar{x})^2}{n-1}}$$

หรือ 2

$$S.D. = \sqrt{\frac{n \sum fx^2 - (\sum x)^2}{n(n-1)}}$$

S.D. คือ ส่วนเบี่ยงเบนมาตรฐาน

$f$  คือ ความถี่

$x$  คือ จุดกึ่งกลางชั้น

$\bar{x}$  คือ ค่าเฉลี่ยเลขคณิต

$n$  คือ จำนวนข้อมูล

Ex3. จากตารางข้อมูลต่อไปนี้จงหาส่วนเบี่ยงเบนมาตรฐาน

คะแนน	ความถี่
5-9	3
10-14	6
15-19	7
20-24	8
25-29	10
30-34	12
35-39	14

วิธีทำใช้สูตรที่ 2 สร้างตารางแจกแจงความถี่

คะแนน	(f)	(x)	$x^2$	(f x)	$f x^2$
5-9	3	7	49	21	147
10-14	6	12	144	72	864
15-19	7	17	289	119	2023
20-24	8	22	484	176	3872
25-29	10	27	729	270	7290
30-34	12	32	1024	384	12288
35-39	14	37	1369	518	19166



	$n = 60$			$\Sigma f = 1190$	$\Sigma f^2 = 45650$
--	----------	--	--	-------------------	----------------------

$$1. \text{ หาค่าเฉลี่ยเลขคณิต จากสูตร } \bar{x} = \frac{\Sigma f}{n} = \frac{1}{6} = 19.83$$

$$2. \text{ หาส่วนเบี่ยงเบนมาตรฐาน จากสูตร } S.D. = \sqrt{\frac{n \Sigma f x^2 - (\Sigma f x)^2}{n(n-1)}} = \sqrt{\frac{60(45650) - (19.83)^2}{60(60-1)}} = \sqrt{\frac{273900 - 393.23}{3540}} = 8.79$$

= ส่วนเบี่ยงเบนมาตรฐาน คือ 8.79

#### หมายเหตุ\*

ค่าความแปรปรวน (variance) สามารถหาได้จากสูตร (S.D)<sup>2</sup> หมายความว่าหากต้องการทราบความแปรปรวนให้นำผลของส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ไปทำเป็นเลขยกกำลังสอง

#### 4. การวัดการกระจายสัมพัทธ์

การวัดการกระจายสัมพัทธ์ (relative Variation) คือ การหาค่าเพื่อเปรียบเทียบของการกระจายระหว่างข้อมูลมากกว่าหนึ่งชุดโดยใช้อัตราส่วน การเปรียบเทียบการกระจายของข้อมูลระหว่างชุดที่นิยมใช้กันมีอยู่ 2 ชนิดคือ 1) ค่าสัมประสิทธิ์ของพิสัย (coefficient of range) และ 2) ค่าสัมประสิทธิ์ของการแปรผัน (coefficient of variation) (อ้างอิงใน วรเชษฐ์ โทอิน, 2560)

##### 1) ค่าสัมประสิทธิ์ของพิสัย

ค่าสัมประสิทธิ์ของพิสัย (coefficient of range) คือ อัตราส่วนระหว่างผลต่างของค่าสูงสุดและค่าต่ำสุด กับ ผลบวกของค่าสูงสุดและต่ำสุดของข้อมูลชุดนั้น

$$\frac{x_{max} - x_{min}}{x_{max} + x_{min}}$$

หาได้จากสูตร

Ex4. จงหาสัมประสิทธิ์พิสัยจากข้อมูลชุดนี้ 25, 19, 32, 29, 19, 21, 22, 31, 19, 20, 15, 22, 23, 20

วิธีทำ

$$\begin{aligned}\text{สูตร} &= \frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}} \\ &= \frac{32 - 15}{32 + 15} \\ &= \frac{1}{4} \\ &= .3617 \text{ ทำเป็นเปอร์เซ็นต์ด้วยการ } \times 100 \\ &= 36.17\%\end{aligned}$$

ดังนั้น สัมประสิทธิ์พิสัยของข้อมูลชุดนี้ = 36.17%

## 2) ค่าสัมประสิทธิ์ของการแปรผัน

ค่าสัมประสิทธิ์ของการแปรผัน(coefficient of variation) ตัวย่อ (C.V.) อัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

$$\text{สูตร } C.V. = \frac{S}{\bar{x}}$$

เมื่อ C.V. คือ สัมประสิทธิ์ของการแปรผัน

S คือ ส่วนเบี่ยงเบนมาตรฐาน

$\bar{x}$  คือ ค่าเฉลี่ยเลขคณิต

จาก Ex. 3. จงหาสัมประสิทธิ์การแปรผัน

จะได้  $\bar{x}$  = 19.83

จะได้ S (S.D.) = 8.79

$$\begin{aligned}C.V. &= \frac{S}{\bar{x}} \\ &= \frac{8.79}{19.83} \\ &= 0.3584 \text{ ทำเป็นเปอร์เซ็นต์ด้วยการ } \times 100\end{aligned}$$

ดังนั้น สัมประสิทธิ์พิสัยของข้อมูลชุดนี้ = 35.84 %

## 7.8. เปอร์เซนไทล์ เดไซล์ และควอไทล์

### 1 เปอร์เซนไทล์ (percentile)

เปอร์เซนไทล์ (percentile) หมายถึง ตำแหน่งที่แสดงให้ทราบว่า มีจำนวนร้อยละเท่าไรของจำนวนคะแนนที่มีค่าต่ำกว่าของคะแนนที่ตำแหน่งนั้น เช่น นักศึกษาคนหนึ่งสอบวิชาภาษาไทยได้ 54 คะแนน และคะแนน 54 นี้ อยู่ตำแหน่งเปอร์เซนไทล์ที่ 60 หมายความว่า ร้อยละ 60 ของนักศึกษาในกลุ่มนั้นได้คะแนนวิชาภาษาไทยต่ำกว่า 54 คะแนนการคำนวณหาค่าเฉลี่ยที่ตำแหน่งเปอร์เซนไทล์ (percentile) ที่กำหนดให้สามารถทำได้ 2 วิธี คือ 1) สำหรับคะแนนที่ไม่ได้จัดหมวดหมู่ และ 2) สำหรับคะแนนที่จัดหมวดหมู่ (อ้างอิงใน วรเชษฐ์ โทอิน, 2560, หน้า 135)

## วิชาการเปรียบเทียบวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

1) การคำนวณหาคะแนนที่ตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่กำหนดให้สำหรับข้อมูลที่ไม่ได้จัดหมวดหมู่

ตัวอย่างที่ 4.1 ผลการสอบวิชาสถิติของนักศึกษาระดับประกาศนียบัตรวิชาชีพชั้นสูงจำนวน 20 คน ได้คะแนนสอบ ดังต่อไปนี้

67   45   73   78   38   54   61   41   53   36  
55   37   74   65   47   34   66   68   77   44

คะแนนเท่าไรที่แสดงว่านักศึกษาร้อยละ 75 ของนักศึกษา 20 คนนี้ ได้คะแนนต่ำกว่าคะแนนนั้น

วิธีทำ

ก. เรียงลำดับคะแนนสอบจากน้อยไปหามาก ดังนี้

34   36   37   38   41   44   45   47   53   54  
55   61   65   66   67   68   73   74   77   78

ข. หาตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่กำหนดให้จากตำแหน่งของคะแนนที่เรียงลำดับแล้วในข้อ 1 คะแนนในตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่ 75 คือ  $(75 \times 20 \div 100 = 15)$  ดังนั้น จากคะแนนที่เรียงลำดับแล้ว ตำแหน่งที่ 15 คือ คะแนนที่ 67

2. การคำนวณหาค่าคะแนนที่ตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่กำหนดให้สำหรับข้อมูลที่จัดหมวดหมู่ การคำนวณหาคะแนนในตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่กำหนดสามารถทำได้ คือ การใช้สูตรเทียบบัญญัติไตรยางค์ และการใช้โค้งความถี่สะสม

วิธีคำนวณคล้ายกับการคำนวณหาค่ามัธยฐาน (median) ต่างกันที่การหาตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ในคะแนนชุดนั้น จะไม่ใช่คะแนนในตำแหน่งกลาง ( $N/2$ ) เหมือนกับมัธยฐาน ซึ่งใช้สูตรในการคำนวณตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) คือ

$$Pr = L + \{(Fn - F1) / (F2 - F1)\}$$

เมื่อ Pr = คะแนนที่ตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่ต้องการ

L = ขีดจำกัดล่างชั้นที่มี Fn อยู่

I = ความกว้างของอันตรภาคชั้น (ขีดจำกัดบน-ขีดจำกัดล่าง)

Fn = ความถี่สะสมของตำแหน่งที่ต้องการ =  $(P \times N) \div 100$

F1 = ความถี่สะสมของชั้นที่อยู่ถัดจากชั้น Fn ไปทางคะแนนน้อย

F2 = ความถี่สะสมชั้น Fn

ตัวอย่างที่ 4.2 จงคำนวณหาคะแนนในตำแหน่ง P25 ซึ่งคะแนนของนักศึกษา 50 คน ดังนี้

คะแนน	ความถี่
10-19	2
20-29	8
30-39	9
40-49	14
50-59	8
60-69	6
70-79	3
	N = 50

วิธีทำ

ก. สร้างตารางแจกแจงความถี่สะสมจากคะแนนน้อยไปหาคะแนนมาก

คะแนน	ความถี่	ความถี่สะสม
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N = 50	

คำนวณหาค่าความถี่สะสมของตำแหน่งที่ต้องการ ( $F_n$ ) ของคะแนนที่ตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile) P25 ของคะแนนชุดนี้ตรงกับความถี่สะสมของตำแหน่งที่  $= (P \times N) \div 100$   
 $= (25 \times 50 \div 100 = 1.250$  หรือ  $12.5)$   
 $= 12.5$

ดังนั้น  $F_n = 12.5$

หาชั้นของคะแนนในตารางความถี่สะสม ที่มีตำแหน่งที่ต้องการ ( $F_n$ ) อยู่ซึ่งตำแหน่งที่ 2.5 อยู่ในชั้นที่มีคะแนนระหว่าง 30-39 ดังนั้น

$$L = 29.5$$

$$I = 19.5 - 9.5 = 10$$

$$F1 = 10$$

$$F2 = 19$$

คำนวณหาคะแนนในตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่กำหนด จากสูตร

$$Pr = L + I \{ (F_n - F1) / (F2 - F1) \}$$

$$P25 = 29.5 + 10 \{ (12.5 - 10) / (19 - 10) \}$$

$$= 29.5 + 10 \{ 2.5 / 9 \}$$

$$= 29.5 + 2.78$$

$$P25 = 32.28$$

ดังนั้น คะแนนในตำแหน่ง P25 ของคะแนนชุดนี้ คือ 32.28 ซึ่งหมายถึง นักศึกษาร้อยละ 25 ของนักศึกษา 50 คน ได้คะแนนต่ำกว่า 32.28 คะแนน

### 3. การคำนวณหาตำแหน่งของเปอร์เซ็นต์ไทล์ของคะแนนที่กำหนดให้

การคำนวณหาตำแหน่งของเปอร์เซ็นต์ไทล์ของคะแนนที่กำหนดให้สามารถทำได้ 2 วิธี คือ  
1) สำหรับคะแนนที่ไม่ได้จัดหมวดหมู่ และ 2) สำหรับคะแนนที่จัดหมวดหมู่

1. การคำนวณหาตำแหน่งของเปอร์เซ็นต์ไทล์ (percentile) ของคะแนนที่กำหนดให้สำหรับข้อมูลที่ได้จัดหมวดหมู่ให้เรียงลำดับของคะแนนจากน้อยไปมาก แล้วตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) จากสูตร

$$P.R. = (100 R)/N$$

เมื่อ P.R. = ตำแหน่งเปอร์เซ็นต์ไทล์ (percentile)

R = อันดับของคะแนน

N = จำนวนประชากร

ตัวอย่างที่ 4.3 ผลการสอบรายวิชาภาษาอังกฤษของนักศึกษาชั้นปีที่ 3 จำนวน 10 คน มีคะแนนเรียงลำดับจากน้อยไปหามาก ดังนี้ 32, 43, 44, 51, 55, 63, 66, 70, 75, 84 จงหาตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ของคะแนนของนักศึกษา 10 คน ดังนี้

วิธีทำ

คะแนนที่เป็นอันดับที่ 1 คือ คะแนนต่ำสุดของคะแนนชุดนี้ คือ 32 สามารถคำนวณหาเปอร์เซ็นต์ไทล์ (percentile) ได้จากสูตร

$$P.R. = (100 R)/N$$

$$P.R. (1) = (100(1))/10$$

$$= 100/10$$

$$= 10$$

$$P.R. (2) = (100(2))/10$$

$$= 200/10$$

$$= 20$$

ดังนั้น ตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ของคะแนนของนักศึกษา 10 คนนี้ คือ 10, 20, 30, ..., 100

2. การคำนวณหาตำแหน่งของเปอร์เซ็นต์ไทล์ (Percentile) ของคะแนนที่กำหนดให้สำหรับข้อมูลที่จัดหมวดหมู่คำนวณหาตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile) จากสูตร

$$Pr = L + \{(Fn - F1) / (F2 - F1)\}$$

เมื่อ Pr = คะแนนที่ตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile) ที่ต้องการ

L = ชีตจำกัดล่างชั้นที่มี Fn อยู่

I = ความกว้างของอันตรภาคชั้น = ชีตจำกัดบน - ชีตจำกัดล่าง

Fn = ความถี่สะสมของตำแหน่งที่ต้องการ =  $(P \times N)/100$

F1 = ความถี่สะสมของชั้นที่อยู่ถัดจากชั้น Fn ไปทางคะแนนน้อย

F2 = ความถี่สะสมชั้น Fn

## วิชาการเบี่ยงวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

ตัวอย่างที่ 4.4 จงหาว่าคะแนน 35 อยู่ในตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่เท่าไร ซึ่งคะแนนนักศึกษา 50 ดังนี้

คะแนน	ความถี่
10-19	2
20-29	8
30-39	9
40-49	14
50-59	8
60-69	6
70-79	3
	N = 50

วิธีทำ

1) สร้างตารางแจกแจงความถี่สะสมจากคะแนนน้อยไปหาคะแนนมาก

คะแนน	ความถี่	ความถี่สะสม
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N = 50	

จงสร้างตารางความถี่สะสม แสดงว่าคะแนน 35 อยู่ในชั้นคะแนน 30-39 ดังนี้

$$L = 29.5$$

$$I = 19.5 - 9.5 = 10$$

$$F1 = 10$$

$$F2 = 19$$

ข. หาดำแหน่งที่ต้องการ  $F_n$  จากสูตร

$$Pr = L + I \{ (F_n - F1) / (F2 - F1) \}$$

$$35 = 29.5 + 10 \{ (F_n - 10) / (19 - 10) \}$$

$$35 - 29.5 = 10 \{ (F_n - 10) / 9 \}$$

$$5.5 = (10/9) (F_n - 10)$$

$$(5.5)(9) / 10 = F_n - 10$$

$$4.95 = F_n - 10$$

$$F_n = 14.95$$

เนื่องจาก  $F_n = (P \times N)/100$  จะได้

$$14.95 = (P \times 50)/100$$

$$P = 14.95 \times 100 \div 50$$

$$P = 29.9$$

ดังนั้น คะแนน 35 อยู่ในตำแหน่งเปอร์เซ็นต์ไทล์ (percentile) ที่ 29.9

## 2 เดไซล์ (Decile)

เดไซล์ (decile) หมายถึง ตำแหน่งที่แสดงให้ทราบว่า “มีจำนวนข้อมูลเท่าไรในจำนวน 10 ส่วนของจำนวนข้อมูลทั้งหมดที่มีค่าต่ำกว่าข้อมูลที่ตำแหน่งนั้น” เช่น นักศึกษาคนหนึ่งสอบวิชาสังคมศึกษา ได้คะแนน 55 คะแนน และคะแนน 55 นี้ อยู่ในตำแหน่งเดไซล์ที่ 7 (D7) หมายความว่า 7 ใน 10 ของนักศึกษากลุ่มนั้นได้คะแนนวิชาสังคมศึกษาสูงกว่า 55 คะแนน

เปรียบเทียบเดไซล์ (decile) และเปอร์เซ็นต์ไทล์ (percentile) ดังนี้

$$1) D_1 = P_{10}$$

$$2) D_2 = P_{20}$$

$$3) D_{10} = P_{100}$$

ดังนั้น การคำนวณหาเดไซล์ (decile) จึงใช้วิธีเดียวกันกับการคำนวณหาเปอร์เซ็นต์ไทล์ (percentile) ต่างกันที่การแบ่งหมู่เท่านั้น คือ การคำนวณเดไซล์ (decile) ต้องแบ่งหมู่ออกเป็น 10 ส่วน แต่เปอร์เซ็นต์ไทล์ (percentile) แบ่งหมวดหมู่เป็น 100 ส่วน

## 3 ควอไทล์ (Quartile)

ควอไทล์ (quartile) หมายถึง ตำแหน่งที่แสดงให้ทราบว่า “มีจำนวนข้อมูลเท่าไรใน 4 ส่วนของจำนวนข้อมูลทั้งหมดที่มีค่าต่ำกว่าข้อมูลที่ตำแหน่งนั้น” เช่น นักศึกษาคนหนึ่งสอบวิชาสถิติ ได้คะแนน 72 คะแนนและคะแนน 72 อยู่ในตำแหน่งควอไทล์ (quartile) ที่ 3 (Q3) หมายความว่า 3 ใน 4 ของนักศึกษากลุ่มนั้นได้คะแนนวิชาสถิติต่ำกว่า 72 คะแนนและ 1 ใน 4 ของนักศึกษากลุ่มนั้นได้คะแนนวิชาสถิติสูงกว่า 72 คะแนน

เปรียบเทียบควอไทล์ (Quartile) และเปอร์เซ็นต์ไทล์ (Percentile) ดังนี้

$$1) Q_1 = P_{25}$$

$$2) Q_2 = P_{50}$$

$$3) Q_3 = P_{75}$$

$$4) D_{10} = P_{100}$$

ดังนั้น การคำนวณหาควอไทล์ (quartile) จึงใช้วิธีเดียวกันกับการคำนวณหาเปอร์เซ็นต์ไทล์ (percentile) ต่างกันที่การแบ่งหมู่เท่านั้น คือ การคำนวณควอไทล์ (quartile) แบ่งหมวดหมู่ออกเป็น 4 ส่วน แต่เปอร์เซ็นต์ไทล์ (percentile) แบ่งหมวดหมู่ออกเป็น 100 ส่วน



## วิชาการเปรียบเทียบวิธีวิจัย (GS:5007) พระครูสุธีวรสาร, ดร.

ตัวอย่าง

จงหาคำนวนหาคะแนนในตำแหน่ง Q3 ซึ่งคะแนนของนักศึกษา 50 คน ดังนี้

คะแนน	ความถี่
10-19	2
20-29	8
30-39	9
40-49	14
50-59	8
60-69	6
70-79	3
	N = 50

วิธีทำ

สร้างตารางแจกแจงความถี่สะสมจากคะแนนน้อยไปหาคะแนนมาก

คะแนน	ความถี่	ความถี่สะสม
10-19	2	2
20-29	8	10
30-39	9	19
40-49	14	33
50-59	8	41
60-69	6	47
70-79	3	50
	N = 50	

Q3 ของข้อมูลชุดนี้อยู่ในตำแหน่งที่  $= (Q \times N)/4$

$$= (3 \times 50)/4$$

$$= 37.5$$

ตำแหน่งที่ 37.5 อยู่ในชั้นที่มีคะแนนระหว่าง 50-59 ดังนี้

$$F_n = 37.5$$

$$L = 49.5$$

$$I = 10$$

$$F1 = 33$$

$$F2 = 41$$

จากสูตร  $Q_r = L + \{(Fn - F1) / (F2 - F1)\}$

$$Q_3 = 49.5 + 10\{(37.5 - 33) / (41 - 33)\}$$

$$= 49.5 + 10\{(4.5) / (8)\}$$

$$= 49.5 + 5.625$$

$$Q_3 = 55.125$$

ดังนั้น คะแนนที่ตำแหน่ง  $Q_3$  ของข้อมูลชุดนี้ คือ 55.125

การหาตำแหน่งข้อมูลเป็นการบอกถึงตำแหน่งที่ข้อมูลอยู่และยังสามารถบอกได้ด้วยว่ามีข้อมูลจำนวนเท่าไรที่มีค่าต่ำกว่าและข้อมูลจำนวนเท่าไรที่มีค่ามากกว่า การหาตำแหน่งข้อมูล มี 3 ค่า ได้แก่ ควอไทล์ (quartile) เดไซล์ (decile) และเปอร์เซ็นต์ไทล์ (percentile) จะแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน เดไซล์ (decile) จะแบ่งข้อมูลทั้งหมดออกเป็น 10 ส่วน เท่า ๆ กัน และเปอร์เซ็นต์ไทล์ (percentile) จะแบ่งข้อมูลทั้งหมดออกเป็น 100 ส่วนเท่า ๆ กัน

เปอร์เซ็นต์ไทล์ (percentile) เดไซล์ (decile) และควอไทล์ (quartile) เป็นค่าที่แสดงตำแหน่งของข้อมูลเมื่อเทียบกับจำนวนของข้อมูลทั้งหมด ทำให้ทราบว่าข้อมูลที่มีค่าต่ำกว่าหรือเท่ากับตัวมันอยู่เป็นจำนวนเท่าไร และมีข้อมูลที่มีค่าสูงกว่าตัวมันอยู่เป็นจำนวนเท่าไร โดยเปอร์เซ็นต์ไทล์ (percentile) จะแบ่งออกเป็น 100 ส่วนเท่า ๆ กัน เดไซล์ (decile) จะแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน และควอไทล์ (quartile) จะแบ่งออกเป็น 4 ส่วนเท่า ๆ กัน

จากข้อมูลที่เกิดขึ้นมาได้นอกจากจะทำการวัดแนวโน้มเข้าสู่ส่วนกลางแล้วยังสามารถหาค่าเปอร์เซ็นต์ไทล์ (percentile) เดไซล์ (decile) และ ควอไทล์ (quartile) ซึ่งเป็นค่าที่แสดงตำแหน่งของข้อมูลเพื่อเทียบกับจำนวนข้อมูลทั้งหมดจึงเป็นค่าที่แสดงให้ทราบว่าข้อมูลที่มีค่าต่ำกว่าหรือเท่ากับตัวมันอยู่จำนวนเท่าไร และมีข้อมูลที่มีค่าสูงกว่าตัวมันอยู่จำนวนเท่าไร เช่น การจัดอันดับผลของการสอบ เป็นต้น ในหน่วยนี้จะกล่าวถึงเรื่องของเปอร์เซ็นต์ไทล์ (percentile) เดไซล์ (decile) และควอไทล์ (quartile)

### เอกสารอ้างอิงท้ายบท

- ธานินทร์ ศิลป์จารุ. (2557). *การวิจัยและวิเคราะห์ข้อมูลทางสถิติด้วย SPSS และ AMOS*. (พิมพ์ครั้งที่ 15) กรุงเทพมหานคร: บริษัท เอส. อาร์. พรินติ้ง แมสโปรดักส์.
- วรรณดี แสงประทีปทอง. (2544). *การเก็บรวบรวมข้อมูลและการสร้างเครื่องมือในการประเมินทางการศึกษา* ในประมวลสาระชุดวิชาการประเมินและการจัดการโครงการประเมิน หน่วยที่ 5 สาขาวิชาศึกษาศาสตร์, มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- วรเชษฐ์ โทอิน. (2560). *เอกสารประกอบการสอนสถิติและการวิจัยขั้นสูง*. คณะสังคมศาสตร์ สาขาวิชาการปกครอง มหาวิทยาลัยมหามกุฏราชวิทยาลัย วิทยาลัยศาสนศาสตร์ยโสธร
- วรเชษฐ์ โทอิน. (2560). *เอกสารประกอบการสอนวิชาสถิติเพื่อการวิจัยทางสังคมศาสตร์*. คณะสังคมศาสตร์ สาขาวิชาการปกครอง มหาวิทยาลัยมหามกุฏราชวิทยาลัย วิทยาลัยศาสนศาสตร์ยโสธร
- วรเชษฐ์ โทอิน. (2560). *เทคนิคการใช้สถิติเพื่อการวิจัยทางสังคมศาสตร์*. คณะสังคมศาสตร์ สาขาวิชาการปกครอง มหาวิทยาลัยมหามกุฏราชวิทยาลัย วิทยาลัยศาสนศาสตร์ยโสธร
- ศิริชัย กาญจนวาสี. (2537). *การเลือกใช้สถิติที่เหมาะสมสำหรับการวิจัย*. (พิมพ์ครั้งที่ 3) ศักดิ์สิทธิ์ วัชรารัตน์. (2552). *เอกสารประกอบการค้นคว้าวิชาสถิติเพื่อการวิจัย*. วิทยาลัยสารพัดช่างพิษณุโลก สำนักงานคณะกรรมการอาชีวศึกษา กระทรวงศึกษาธิการ
- สายชล สันสมบูรณ์ทอง. (2560). *สถิติเบื้องต้น*. (พิมพ์ครั้งที่ 11), กรุงเทพฯ, จามจุรีโปรดักส์.
- สำเร็จ จันทสุวรรณ และสุวรรณ บัวทวน. (2537). *เอกสารคำสอน สถิติสำหรับการวิจัยทางสังคมศาสตร์*. ภาควิชาสังคมวิทยาและมานุษยวิทยา คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น.
- เทคนิคทางสถิติ (2561). *เทคนิคทางสถิติ*. สืบค้นเมื่อวันที่ 17 กรกฎาคม 2561 จาก <https://ruchareka.wordpress.com>