บทที่ 4

สถิติพรรณนาเพื่อการบรรยายลักษณะข้อมูล

ในปัจจุบันการวิจัยมีอยู่หลายประเภทดังได้กล่าวมาแล้วในบทที่ 1 แต่ถ้าหากเป็นงานวิจัยที่เกี่ยวข้องกับข้อมูลที่เป็นตัวเลขหรือเป็นข้อมูลที่อยู่ในรูปแบบของข้อมูลเชิงปริมาณในการวิเคราะห์ข้อมูลต้องอาศัยหลักการของสถิติมาใช้ในการวิเคราะห์ ซึ่งในบทนี้จะกล่าวถึงสถิติเบื้องต้นที่ใช้ในการวิจัยประกอบไปด้วย ประชากรและกลุ่มตัวอย่าง การสุ่มตัวอย่าง ตัวแปรและชนิดของตัวแปร ระดับการวัดของตัวแปร ความหมายและประเภทของสมมติฐาน วิธีการทางสถิติที่ใช้ในการวิจัย สถิติเชิงพรรณนา และสถิติเชิงอ้างอิง เพื่อให้นักศึกษาสามารถนำหลักการของสถิติเบื้องต้นไปใช้ในการวิจัยได้อย่างถูกต้องและเหมาะสม

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร (Population)

ธีรยุทธ พึ่งเทียร (2545 : 63) กล่าวถึงความหมายของประชากรไว้ว่า ประชากร หมายถึง เซตของค่าที่สังเกตได้ของหน่วยต่างๆที่อยู่ในขอบข่ายของการทดลอง หรือการสำรวจทางสถิติหรือในวิชาคณิตศาสตร์ ประชากรหมายถึง กลุ่มของสิ่งที่ต้องการศึกษา

ปรีชา อัศวเดชานุกร และคณะ (2550 : 101) กล่าวถึงความหมายของประชากรไว้ว่า ประชากร หมายถึง หน่วยของข้อมูลทั้งหมดที่ครอบคลุมสิ่งที่สนใจศึกษา

พันทิพา สุนทรารชุน (2542 : 7) กล่าวถึงความหมายของประชากรไว้ว่า ประชากร หมายถึง หน่วยทุกหน่วยรวมกันซึ่งเป็นสิ่งที่เราสนใจศึกษาและพยายามหาข้อสรุป

มยุรี ศรีชัย (2541 : 3) กล่าวถึงความหมายของประชากรไว้ว่า ประชากร หมายถึง กลุ่มสมาชิกทั้งหมดของสิ่งต่างๆที่ต้องการสรุปอ้างอิงถึง อาจจะเป็น คน สัตว์ สิ่งของ เหตุการณ์ ปรากฏการณ์ หรือพฤติกรรมใดๆก็ได้ ซึ่งอยู่ภายในขอบเขตที่กำหนด

ซึ่งจากความหมายดังกล่าวข้างต้น สามารถสรุปได้ว่า ประชากร คือ ขอบเขตของข้อมูลทั้งหมดที่เรากำลังศึกษา หรืออาจหมายถึงทุกๆหน่วยที่อยู่ในขอบข่ายกรอบการศึกษา โดยประชากรจะแบ่งออกเป็น 2 ประเภท คือ 1) ประชากรในพื้นที่ที่มีจำนวนแน่นอน เช่น จำนวนนักศึกษาสาขาวิชาสถิติประยุกต์ที่กำลังศึกษาอยู่ในปีการศึกษา 2558 และ 2) ประชากรที่มีไม่แน่นอน เช่น จำนวนคนที่เข้าออกจากสถานีขนส่งจังหวัดบุรีรัมย์ในแต่ละวัน และค่าที่ได้จากประชากรจะถูกเรียกว่า *ค่าพารามิเตอร์ (Parameter)*

1.2 กลุ่มตัวอย่าง (Sample)

ปรีชา อัศวเดชานุกร และคณะ (2550 : 102) กล่าวถึงความหมายของกลุ่มตัวอย่างไว้ว่า กลุ่มตัวอย่าง หมายถึง ส่วนหนึ่งของประชากร

พันทิพา สุนทรารชุน (2542 : 7) กล่าวถึงความหมายของกลุ่มตัวอย่างไว้ว่า กลุ่มตัวอย่าง หมายถึง บางส่วนของประชากร

มยุรี ศรีชัย (2541 : 3) กล่าวถึงความหมายของกลุ่มตัวอย่างไว้ว่า กลุ่มตัวอย่าง หมายถึง สมาชิกของประชากรที่ถูกเลือกมาด้วยวิธีการต่างๆเพื่อทำการศึกษาวิเคราะห์ แล้วนำผลสรุปที่ได้ไปใช้อ้างอิงถึงประชากร ถ้ากลุ่มสมาชิกที่เลือกมานั้นเป็นตัวแทนที่ดีของประชากรและมีจำนวนมากพอแล้ว ค่าที่ใช้สรุปอ้างอิงถึงประชากรจะมีความถูกต้อง หรือใกล้เคียงกับลักษณะหรือคุณสมบัติของประชากรมาก

สำนักงานสถิติแห่งชาติ (2556) กล่าวถึงความหมายของกลุ่มตัวอย่างไว้ว่า กลุ่มตัวอย่าง หมายถึงกลุ่มของหน่วยตัวอย่างที่สุ่มเลือกมาจากกรอบตัวอย่าง เพื่อใช้เป็นตัวแทนในการศึกษาหรือสรุปอ้างอิงถึงลักษณะของประชากร

ซึ่งจากความหมายดังกล่าวข้างต้น สรุปได้ว่า กลุ่มตัวอย่าง คือ สมาชิกของประชากรที่ถูกเลือกมาด้วยวิธีการต่างๆเพื่อใช้เป็นตัวแทนในการศึกษาหรือสรุปอ้างอิงถึงลักษณะของประชากร ดังนั้นกลุ่มตัวอย่างจึงควรเป็นตัวแทนที่ดีของประชากร และผลการเลือกตัวอย่างจะส่งผลต่อการศึกษาที่จะสามารถอ้างอิงไปยังประชากรได้ ฉะนั้นตัวอย่างต้องเป็นตัวแทนที่ดีของประชากรทั้งหมดจึงมีความจำเป็นที่จะต้องอาศัยเทคนิคการสุ่มตัวอย่างเพื่อช่วยคัดเลือกตัวอย่างซึ่งจะกล่าวถึงในหัวข้อถัดไป และสำหรับค่าที่ได้จากตัวอย่างจะถูกเรียกว่า *ค่าสถิติ (Statistics)*

2. ตัวแปรและชนิดของตัวแปร

2.1 ความหมายของตัวแปร

ตัวแปร หมายถึง คุณลักษณะที่เปลี่ยนแปลงได้ ขึ้นอยู่กับความแตกต่างเฉพาะบุคคลหรือกลุ่มตัวอย่าง เช่น อุณหภูมิของร่างกาย การนับถือศาสนา รายได้ อายุ ความสูง เป็นต้น ตัวแปรคุณลักษณะเหล่านี้จะเปลี่ยนแปลงไปในแต่ละบุคคล เช่น ในการสุ่มตัวอย่างนักศึกษาระดับปริญญาตรีของมหาวิทยาลัยแห่งหนึ่งมา 5 คน ได้ข้อมูลดังนี้

คนที่	คณะที่ศึกษา	เพศ	ชั้นปี	อายุ	GPA	จำนวนหน่วยกิตที่เรียน
1	วิทยาศาสตร์	หญิง	1	17	2.45	18
2	การจัดการ	หญิง	2	19	3.10	15
3	เทคโนโลยีอุตสาหกรรม	ชาย	3	21	2.75	18
4	เทคโนโลยีการเกษตร	ชาย	2	18	2.00	21
5	มนุษยศาสตรและสังคมศาสตร์	หญิง	4	22	3.14	12

ตัวแปรที่อยู่ในตัวอย่างนี้ประกอบด้วยตัวแปรคณะที่ศึกษา เพศ ชั้นปี อายุ GPA และจำนวนหน่วยกิตที่เรียน จากคุณลักษณะของตัวแปรแต่ละตัว ตัวแปรคณะที่ศึกษาแบ่งได้มากเท่ากับจำนวนคณะที่เปิดสอนในระดับปริญญาตรีของมหาวิทยาลัยแห่งนี้ ตัวแปรที่สองที่สอบวัดกลุ่มตัวอย่าง คือ เพศ สามารถเป็นไปได้เพียง 1 ประเภทใน 2 ประเภทคือเพศชายหรือเพศหญิง ตัวแปรเพศเป็นตัวแปรที่ไม่ให้ค่าเป็นตัวเลข ซึ่งคล้ายกับตัวแปรที่สามคือ ชั้นปี ตัวแปรชั้นปีจะแบ่งได้ 4 ประเภท (ชั้นปีที่ 1, 2, 3 และ 4) สำหรับตัวแปร อายุ GPA และ จำนวนหน่วยกิตที่เรียน จะให้ค่าเป็นตัวเลข เป็นชุดของตัวแปรเชิงปริมาณ โดยถ้าในตัวอย่างหนึ่งๆมีเพียงตัวแปรเดียวที่ถูกวัด เช่นวัดเฉพาะตัวแปร GPA ข้อมูลที่ได้เราจะเรียกว่า ***Univariate data*** แต่ถ้าตัวแปร 2 ตัวถูกวัด เช่น ความสูงและน้ำหนัก ข้อมูลที่ได้จะถูกเรียกว่า ***bivariate data*** และถ้าตัวแปรมากกว่า 2 ตัวขึ้นไปถูกวัด เช่น ในตารางนี้มีตัวแปรที่ถูกวัด 6 ตัว ข้อมูลที่ได้จะถูกเรียกว่า ***multivariate data***

2.2 ชนิดของตัวแปร

จากตัวอย่างที่กล่าวมาแล้วข้างต้นในหัวข้อความหมายของตัวแปร พบว่า ตัวแปรต่าง ๆ ที่อยู่ในตารางจะมีทั้งที่เป็นตัวเลขและไม่ใช่ตัวเลข ดังนั้น จึงสามารถแบ่งชนิดของตัวแปรได้เป็น 2 ประเภทคือ

2.2.1 ตัวแปรเชิงคุณภาพ (Categorical or Qualitative)

เป็นตัวแปรที่ไม่ใช่ตัวเลข แต่เป็นตัวแปรที่ข้อมูลมีลักษณะเป็นการแบ่งประเภทให้เห็นถึงความแตกต่างของกลุ่มตัวอย่างแต่ละกลุ่ม เช่น ศาสนา, อาชีพ, สถานภาพสมรส, ระดับการศึกษา ฯลฯ และจากตารางในตัวอย่างข้างต้น ตัวแปรคณะที่ศึกษา เพศ และ ชั้นปี ต่างก็เป็นตัวแปรเชิงคุณภาพทั้งสิ้น

2.2.2 ตัวแปรเชิงปริมาณ (Numerical or Quantitative)

เป็นตัวแปรที่ถูกวัดมามีค่าเป็นตัวเลขแสดงให้เห็นถึงปริมาณความมากน้อย โดยถ้าตัวแปรที่วัดค่าได้เป็นเลขจำนวนนับ (0, 1, 2, ...) เช่น จำนวนบุคคลในครอบครัว, จำนวนสินค้าที่จำหน่ายไป, จำนวนนักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์ปีการศึกษา 2558 ฯลฯ จะเรียกว่า ***ตัวแปรไม่ต่อเนื่อง (Discrete)*** แต่ถ้าตัวแปรที่วัดค่าได้เป็นเลขจำนวนจริง เช่น ส่วนสูง น้ำหนัก ระยะเวลา อุณหภูมิ ฯลฯ จะเรียกว่า ***ตัวแปรต่อเนื่อง (Continuous)***

3. ระดับการวัดของตัวแปร

ระดับการวัดของตัวแปร หมายถึง การแปลความหมายของตัวเลขซึ่งเป็นค่าของตัวแปรว่ามีความหมายในระดับใด ในทางสถิติสามารถแบ่งระดับการวัดของตัวแปรได้เป็น 4 ระดับ ดังนี้

1. ระดับนามบัญญัติ (Nominal Scale) หมายถึง การกำหนดตัวเลขแทนค่าตัวแปร โดยให้ตัวเลขนั้นเป็นเพียงสัญลักษณ์หรือจำแนกตัวแปรออกเป็นพวก ๆ ไม่สามารถนำตัวเลขนั้นมาคิดคำนวณ ในเชิงปริมาณได้ เช่น เพศ เราแบ่งเป็นเพศชายและเพศหญิง โดยแทนเพศชายด้วย 1 เพศหญิง เป็น 2 แต่เลข 1 และ2 ไม่ได้มีความหมายเป็นจำนวนเป็นเพียงแค่สัญลักษณ์แทนเพศเท่านั้น

2. ระดับจัดลำดับ (Ordinal Scale) หมายถึง การกำหนดตัวเลขที่บอกอันดับของสิ่งที่นำมาเปรียบเทียบกันโดยไม่สามารถบอกปริมาณของความแตกต่างกันได้ และไม่สามารถนำตัวเลขนั้น ๆ มาคิดคำนวณในเชิงปริมาณได้ เช่น การประกวดนางงาม การแข่งขันต่างๆ ได้รางวัลชนะเลิศ รองชนะเลิศอันดับ 1 รองชนะเลิศอันดับ 2 สามารถบอกอันดับได้ แต่บอกปริมาณไม่ได้ว่าชนะเลิศอันดับ 1 และรองชนะเลิศอันดับ 1 นั้นสวยไพเราะหรือดีมากกว่ากันเท่าใด

3. ระดับอันตรภาค (Interval Scale) หมายถึง การกำหนดตัวเลขแทนคุณลักษณะของสิ่งต่าง ๆ โดยมีปริมาณแทนสิ่งนั้น ๆ ด้วย ซึ่งตัวเลขต่างๆเหล่านี้สามารถเปรียบเทียบเพื่อบอกความแตกต่างกัน ทั้งยังสามารถนำไปคิดคำนวณในเชิงปริมาณได้ แต่ตัวเลขที่นำมาใช้แทนนั้นเป็นตัวเลขที่สมมติให้มีความเป็นศูนย์ที่เท่ากัน หรืออาจเรียกอีกอย่างหนึ่งว่า ศูนย์ไม่แท้ เช่น อุณหภูมิ หรือการสอบวัดความรู้ของนักเรียนในวิชาหนึ่งๆ นักเรียนคนหนึ่งได้ศูนย์คะแนน (0) แต่ไม่ได้หมายความว่านักเรียนไม่มีความรู้ในวิชานั้นเลย อาจเป็นเพียงแค่ความรู้ไม่ตรงกับข้อสอบ หรือวัดไม่ตรงความรู้ของเขานั่นเอง อย่างนี้เรียกว่าศูนย์ไม่แท้

4. ระดับอัตราส่วน (Ratio Scale) หมายถึง ตัวเลขที่ใช้แทนคุณลักษณะของสิ่งต่าง ๆ ตามความเป็นจริงทางธรรมชาติ หรืออาจเรียกว่า มีศูนย์แท้ ส่วนใหญ่เป็นการวัดทางกายภาพ เช่นส่วนสูง น้ำหนัก เวลา เป็นต้น ตัวอย่างเช่น น้ำหนัก 0 กิโลกรัม หมายถึงไม่มีน้ำหนักเลย เรียกว่าเป็นศูนย์แท้

4. ความหมายและประเภทของสมมติฐาน

ในการดำเนินการวิจัย "สมมติฐาน" (hypothesis) เป็นสิ่งที่มีความสำคัญมาก เนื่องจากการวิจัยเป็นกระบวนการแก้ปัญหาหรือค้นหาคำตอบด้วยวิธีการทางวิทยาศาสตร์ ซึ่งจะเริ่มต้นโดยการกำหนดปัญหา จากนั้นจะพยายามคาดคะเนคำตอบของปัญหานั้น การคาดคะเนคำตอบก็คือสมมติฐาน ดังนั้นสมมติฐานการวิจัย คือ คำตอบหรือข้อสรุปของผลการวิจัยที่ผู้วิจัยคาดการณ์ หรือคาดคะเนไว้ล่วงหน้าอย่างมีเหตุ

และผล โดยอาศัยรากฐานของแนวคิดทฤษฎี ผลการศึกษาค้นคว้า ผลการวิจัยรวมถึงประสบการณ์ของผู้วิจัยเอง ซึ่งสมมติฐานนี้สามารถใช้เป็นแนวทางในการค้นคว้า ตลอดจนเป็นแนวทางในการเก็บรวบรวมข้อมูล และวิเคราะห์ข้อมูลว่าสิ่งที่ผู้วิจัยศึกษาอยู่นั้นเป็นไปตามที่คาดการณ์ไว้หรือไม่ ทั้งนี้สมมติฐานที่ตั้งไว้อาจเป็นจริงหรือไม่เป็นจริงตามที่ผู้วิจัยคาดคะเนก็ได้ ขึ้นอยู่กับการทดสอบสมมติฐานโดยอาศัยข้อมูลที่เก็บรวบรวมได้และวิธีการทางสถิติ

4.1 ประเภทของสมมติฐาน

สมมติฐานแบ่งออกได้เป็น 2 ประเภท คือ

4.1.1 สมมติฐานการวิจัย (research hypothesis) เป็นสมมติฐานที่เขียนอยู่ในรูปของข้อความที่แสดงถึงความสัมพันธ์ของตัวแปรที่ศึกษากับคำตอบที่ผู้วิจัยคาดคะเนโดยใช้ภาษาที่เข้าใจง่ายสามารถสื่อความหมายได้โดยตรง

4.2.2 สมมติฐานทางสถิติ (statistical hypothesis) เป็นสมมติฐานที่เปลี่ยนรูปมาจากสมมติฐานการวิจัย โดยใช้สัญลักษณ์ทางคณิตศาสตร์ที่แทนคุณลักษณะเกี่ยวกับค่าพารามิเตอร์ของประชากร (population parameter) มาเขียนอธิบายความสัมพันธ์ของตัวแปร สมมติฐานทางสถิติจะประกอบด้วย 2 ลักษณะ ควบคู่ไปเสมอ คือ

4.2.2.1 สมมติฐานว่าง หรือสมมติฐานหลัก (null hypothesis) แทนสัญลักษณ์ด้วย H_0 เป็นสมมติฐานแสดงข้อความที่เป็นกลาง โดยระบุถึงความสัมพันธ์ของตัวแปรว่าเท่ากัน ไม่แตกต่างกันหรือไม่มีความสัมพันธ์กัน

4.2.2.2 สมมติฐานทางเลือกหรือสมมติฐานรอง (alternative hypothesis) แทนสัญลักษณ์ด้วย H_1 หรือ H_a เป็นสมมติฐานแตกต่างหรือตรงข้ามกับสมมติฐานหลัก โดยระบุถึงความสัมพันธ์ของตัวแปรว่าไม่เท่ากัน แตกต่างกัน มากกว่า น้อยกว่า หรือมีความสัมพันธ์กัน

4.2 วิธีการตั้งสมมติฐาน

สมมติฐานเป็นการคาดคะเนคำตอบของปัญหาที่ทำการศึกษา ดังนั้นการตั้งสมมติฐานจึงต้องเริ่มคิดก่อนว่าจะมีจุดมุ่งหมายอย่างไร แล้วจึงตั้งสมมุติฐานขึ้น โดยวิธีการตั้งสมมติฐานจะมี 2 ลักษณะ คือ

4.2.1 สมมติฐานแบบมีทิศทาง (directional hypothesis) เป็นการเขียนโดยระบุทิศทางของความสัมพันธ์ของตัวแปรว่าสัมพันธ์ในทางใด หรือถ้าเป็นการเปรียบเทียบก็สามารถระบุถึงทิศทางของความแตกต่างได้ เช่น มากกว่า น้อยกว่า มีความสัมพันธ์ทางบวก ทางลบ เช่น

$$H_0 : \mu = 35$$

$$H_1 : \mu > 35$$

4.2.2 สมมติฐานแบบไม่มีทิศทาง (nondirectional hypothesis) เป็นการเขียนที่ไม่ได้ระบุทิศทางของความสัมพันธ์ของตัวแปรหรือทิศทางของความแตกต่างเพียงแต่ระบุว่ามีความสัมพันธ์กันหรือแตกต่างกันเท่านั้น เช่น

$$H_0 : \mu = 35$$

$$H_1 : \mu \neq 35$$

ในการตั้งสมมติฐานอาจจะตั้งแบบมีทิศทางหรือไม่มีทิศทางก็ได้ ขึ้นอยู่กับผู้วิจัยจะมีข้อมูลเกี่ยวกับเรื่องที่ศึกษามากน้อยเพียงใด ถ้ามีข้อมูลมากพอที่จะยืนยัน ก็ตั้งแบบมีทิศทาง ถ้าข้อมูลไม่พอหรือไม่แน่ใจก็ตั้งแบบไม่มีทิศทาง

สถิติพรรณนา

เป็นสถิติที่บรรยายให้เห็นคุณลักษณะของสิ่งที่ต้องการศึกษาจากกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ ซึ่งอาจจะเป็นกลุ่มใหญ่หรือกลุ่มเล็กก็ได้ ผลของการศึกษาไม่สามารถนำไปอ้างอิงกลุ่มอื่นได้ มักนำเสนอในรูปของ ตาราง แผนภาพ แผนภูมิ ร้อยละ สัดส่วน เปอร์เซนไทล์ การแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจาย และการวัดความสัมพันธ์ระหว่างตัวแปร เป็นต้น โดยในบทนี้จะขออธิบายเพียงการแจกแจงความถี่ การวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจาย และการวัดความสัมพันธ์ระหว่างตัวแปรซึ่งเป็นสถิติพรรณนาที่นิยมใช้เท่านั้น ดังนี้

การแจกแจงความถี่ (frequency distribution)

ในการเก็บรวบรวมคะแนนหรือข้อมูลต่าง ๆที่ได้จากการวัด การทดสอบหรือ โดยวิธีการใด ๆ ก็ตามที่ยังไม่ได้มีการจัดหมวดหมู่ ข้อมูลมีลักษณะกระจัดกระจายประปนกันอยู่อย่างไม่เป็นระเบียบ จะต้องมีการจัดระเบียบเตรียมข้อมูลให้เป็นหมวดหมู่เสียก่อน เพื่อทำให้การแปลความหรือนำไปใช้ได้ง่าย ซึ่งการเตรียมข้อมูลง่าย ๆ ก็คือ การเรียงข้อมูลจากมากไปหาน้อยหรือน้อยไปหามาก โดยข้อมูลที่ยังไม่มีการจัดหมวดหมู่เรียกว่า Ungrouped Data ซึ่งเป็นค่าจริงข้อมูลนั้น

เช่น น้ำหนักของนักศึกษาระดับปริญญาตรี จำนวน 12 คน มีน้ำหนักดังนี้ 44, 50, 48, 55, 60, 45, 48, 47, 63, 65, 70, 71, เราสามารถจัดเรียงข้อมูลจากน้อยไปหามากได้ ดังนี้ 44, 45, 47, 48, 48, 50, 55, 60, 63, 65, 70, 71 แต่ถ้าหากเมื่อไรก็ตามที่ข้อมูลมีมากจำนวนมาก (มีข้อมูลมากกว่า 30 จำนวน) จะต้องทำการจัดหมวดหมู่ หรือแจกแจงความถี่ ซึ่งเรียกข้อมูลที่จัดหมวดหมู่ว่า Grouped Data

ดังนั้นการแจกแจงความถี่ (Frequency Distributions) ก็คือการทำข้อมูลดิบที่รวบรวมได้มาจัดหมวดหมู่ให้เป็นระเบียบ โดยเรียงข้อมูลจากมากไปหาน้อยหรือจากน้อยไปมาก เพื่อหาคะแนนที่เกิดซ้ำกันในแต่ละกลุ่มหรือช่วงคะแนนว่ามีจำนวนเท่าไรนั่นเอง โดยการสร้างตารางแจกแจงความถี่ มีขั้นตอน ดังนี้

โดยการสร้างตารางแจกแจงความถี่ มีขั้นตอน ดังนี้

ขั้นตอนที่ 1 หาพิสัยของข้อมูล (R)

พิสัย (Range) = ค่าสูงสุด – ค่าต่ำสุด

ขั้นตอนที่ 2 กำหนดจำนวนชั้น (K)

$$K = 1 + 3.3\log N$$

ขั้นตอนที่ 3 คำนวณหาความกว้างของชั้น (Class Interval : I)

$$I = \text{ความกว้างของชั้น} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}}$$

ขั้นตอนที่ 4 คำนวณหาขีดจำกัดล่างของชั้นแรก (Class limit)

ขีดจำกัดล่างของชั้นแรก = ค่าต่ำสุด – (I × K – R)/2

ขั้นตอนที่ 5 สร้างตารางแจกแจงความถี่

ตัวอย่างที่ 2.1 จากข้อมูลคะแนนสอบวิชาสถิติและการวิจัยของนักศึกษา 35 คน ดังนี้

72	83	82	92	70	91	71	33	42	51
55	75	38	95	85	93	60	75	38	40
75	49	53	41	86	89	51	57	66	92
55	48	85	85	54					

จงสร้างตารางแจกแจงความถี่

วิธีทำ *ขั้นตอนที่ 1* หาพิสัยของข้อมูล (R)

พิสัย = 95 – 33

= 62

ขั้นตอนที่ 2 กำหนดจำนวนชั้น (K)

$K = 1 + 3.3\log 35$

$= 1 + 3.3(1.544)$

$= 1 + 5.0952$

$= 6.0952 \approx 7$

ขั้นตอนที่ 3 คำนวณหาความกว้างของชั้น (Class Interval : I)

$$I = \text{ความกว้างของชั้น} = \frac{62}{7} = 8.86 \approx 9$$

ขั้นตอนที่ 4 คำนวณหาขีดจำกัดล่างของชั้นแรก (Class limit)

$$\text{ขีดจำกัดล่างของชั้นแรก} = 33 - (9 \times 7 - 62)/2$$

$$= 33 - 0.5 = 32.5 \approx 33$$

ขั้นตอนที่ 5 สร้างตารางแจกแจงความถี่

ขีดจำกัดชั้น	ขอบเขตชั้น	รอยขีด	ความถี่ (จำนวน)	ความถี่สะสม	ความถี่สัมพัทธ์	จุดกึ่งกลางชั้น
33 – 41	32.5 – 41.5	////	5	5	5/35 = 0.143	(33+41)/2 = 37
42 – 50	41.5 – 50.5	///	3	8	3/35 = 0.086	(42+50)/2 = 46
51 – 59	50.5 – 59.5	//////	7	15	7/35 = 0.2	(51+59)/2 = 55
60 – 68	59.5 – 68.5	//	2	17	2/35 = 0.057	(60+68)/2 = 64
69 – 77	68.5 – 77.5	/////	6	23	6/35 = 0.171	(69+77)/2 = 73
78 – 86	77.5 – 86.5	/////	6	29	6/35 = 0.171	(78+86)/2 = 82
87 – 95	86.5 – 95.5	/////	6	35	6/35 = 0.171	(87+95)/2 = 91
รวม			35		0.999 ≈ 1.00	

การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measure of central tendency)

ในการนำเสนอข้อมูลเบื้องต้น ถ้าข้อมูลมีจำนวนมากบางครั้งอาจทำให้ผู้อ่านเข้าใจยาก ไม่สะดวกต่อการนำไปใช้ประโยชน์ ดังนั้นจึงมีความจำเป็นที่จะต้องใช้ตัวแทนของข้อมูล ซึ่งจะเรียกว่าการวัดแนวโน้มเข้าสู่ส่วนกลาง โดยค่ากลางที่นิยมใช้กันมีอยู่ 3 ชนิด คือ

1. ค่าเฉลี่ย (Mean) ซึ่งมีอยู่หลายชนิดแต่ที่นิยมใช้มากที่สุดคือ ค่าเฉลี่ยเลขคณิตเนื่องจากสามารถสื่อความหมายและทำความเข้าใจได้ง่าย และยังมีสมบัติทางสถิติที่ดีอีกด้วย

ในกรณีที่ข้อมูลไม่ได้ถูกจัดกลุ่ม

ค่าเฉลี่ยประชากร แทนด้วย

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

ค่าเฉลี่ยตัวอย่าง แทนด้วย $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{(x_1 + x_2 + \ldots x_n)}{n}$

ในกรณีที่ข้อมูลถูกจัดกลุ่ม

ค่าเฉลี่ยประชากร แทนด้วย

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

ค่าเฉลี่ยตัวอย่าง แทนด้วย

$$\overline{X} = \frac{\sum_{i=1}^{R} f_i x_i}{n}$$

ตัวอย่างที่ 2.2 จากการบันทึกข้อมูลจำนวนนักศึกษาที่มาใช้บริการห้องคอมพิวเตอร์ของสาขาวิชาวิทยาการคอมพิวเตอร์เป็นเวลา 20 วัน ได้ข้อมูลดังนี้

30	35	27	20	24	32	38	29	19	22
33	37	28	21	38	30	25	27	26	17

จงหาค่าเฉลี่ยเลขคณิต

วิธีทำ จากสูตร

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

$$= \frac{(30 + 35 + 27 + \ldots + 17)}{20}$$

$$= \frac{558}{20} = 27.9$$

จะได้ค่าเฉลี่ยเลขคณิตเท่ากับ 27.9 หมายความว่า จำนวนนักศึกษาเฉลี่ยที่มาใช้บริการห้องคอมพิวเตอร์ของสาขาวิชาวิทยาการคอมพิวเตอร์เป็นเวลา 20 วัน เท่ากับ 28 คน

ตัวอย่างที่ 2.3 จากการบันทึกเวลาที่นักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์ใช้ในการรับประทานอาหารกลางวัน จำนวน 20 คน เป็นดังนี้

หน่วย : นาที

20	25	37	27	28	45	50	55	45	57
28	32	38	41	55	40	45	37	46	39

จงหาค่าเฉลี่ยเลขคณิต

วิธีทำ จากสูตร

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{(x_1 + x_2 + \ldots x_n)}{n}$$

$$= \frac{(20 + 25 + 37 + \ldots + 39)}{20} = \frac{790}{20} = 39.5$$

นักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์ใช้เวลาในการรับประทานอาหารกลางวันเฉลี่ย 39.5 นาที

ตัวอย่างที่ 2.4 จากการตรวจสอบรอยตำหนิบนชิ้นส่วนไมโครชิพจำนวน 50 ชิ้น เป็นดังนี้

รอยตำหนิ	0 - 2	3 - 5	6 - 8	9 - 11	12 - 14
จำนวนชิ้น	25	14	5	4	2

จงคำนวณหาค่าเฉลี่ย

วิธีทำ

รอยตำหนิ	จำนวนชิ้น (f_i)	จุดกึ่งกลางชั้น (X_i)	$f_i X_i$
0 – 2	25	1	25
3 - 5	14	4	56
6 - 8	5	7	35
9 – 11	4	10	40
12 – 14	2	13	26
รวม	50		$\sum f_i X_i = 182$

จากสูตร $$\mu = \frac{\sum_{i=1}^{R} f_i x_i}{N}$$

$$= \frac{182}{50} = 3.64$$

จำนวนรอยตำหนิบนชิ้นส่วนไมโครชิพเฉลี่ยเท่ากับ 3.64 หรือประมาณ 4 รอย

2. มัธยฐาน (Median , Me)

ในกรณีที่ข้อมูลไม่ได้ถูกจัดกลุ่ม

ค่ามัธยฐาน คือ ค่าของข้อมูลที่มีตำแหน่งอยู่ตรงกลางของชุดข้อมูลเมื่อนำชุดข้อมูลเรียงจากน้อยไปหามากหรือมากไปหาน้อย

- ถ้าจำนวนข้อมูลเป็นจำนวนคี่ (n เป็นเลขคี่) ค่ามัธยฐาน คือ ค่าของข้อมูลที่อยู่ที่ตำแหน่ง $\frac{(n+1)}{2}$

- ถ้าจำนวนข้อมูลเป็นจำนวนคู่ (n เป็นเลขคู่) ค่ามัธยฐาน คือ ค่าเฉลี่ยของข้อมูลที่อยู่ที่ตำแหน่ง $\frac{n}{2}$ และ $\frac{(n+2)}{2}$

ในกรณีที่ข้อมูลถูกจัดกลุ่ม *ต้องหาค่ามัธยฐานจากสูตร*

$$M_e = L + I\left(\frac{\frac{n}{2} - \sum f_L}{f}\right)$$

โดยที่ L แทน ขอบเขตล่างของชั้นมัธยฐาน

I แทน ความกว้างของชั้นมัธยฐาน

N แทน จำนวนค่าสังเกตทั้งหมด

$\sum f_L$ แทน ผลรวมความถี่ของอันตรภาคชั้นทุกชั้นที่มีค่าสังเกตต่ำกว่าชั้นมัธยฐานเมื่อเรียงข้อมูลจากน้อยไปหามาก

f แทน ความถี่ของชั้นมัธยฐาน

ตัวอย่างที่ 2.5 จงหาค่ามัธยฐานของข้อมูล

24 28 32 39 27 35 33 29 30

วิธีทำ เรียงข้อมูลจากน้อยไปหามาก ดังนี้

24 27 28 29 **30** 32 33 35 39

ค่ามัธยฐานคือค่าของข้อมูลที่อยู่ตำแหน่งที่ $\frac{(n+1)}{2} = \frac{(9+1)}{2} = 5$

ดังนั้นจะได้ค่ามัธยฐานเท่ากับ 30

ตัวอย่างที่ 2.6 จงหาค่ามัธยฐานของข้อมูล

17 19 12 14 29 22 13 21 20 18

วิธีทำ เรียงข้อมูลจากน้อยไปหามาก ดังนี้

12 13 14 17 **18** **19** 20 21 22 29

ค่ามัธยฐานคือค่าเฉลี่ยของข้อมูลที่อยู่ตำแหน่งที่ $\frac{n}{2} = \frac{10}{2} = 5$ และ $\frac{(n+2)}{2} = \frac{(10+2)}{2} = 6$

ดังนั้นจะได้ค่ามัธยฐานเท่ากับ $\frac{18+19}{2} = 18.5$

ตัวอย่างที่ 2.7 จากการสำรวจน้ำหนักของนักเรียนชั้นประถมศึกษาปีที่ 4 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบุรีรัมย์ได้ข้อมูลดังนี้

น้ำหนัก (กก.)	จำนวน
20 – 24	7
25 – 29	10
30 – 34	12
35 – 39	9
40 – 44	2
รวม	40

จงหามัธยฐานของน้ำหนักของนักเรียนกลุ่มนี้

วิธีทำ หาค่าความถี่สะสม ดังนี้

น้ำหนัก (กก.)	จำนวน	ความถี่สะสม
20 – 24	7	7
25 – 29	10	17
30 – 34	12	29
35 – 39	9	38
40 – 44	2	40
รวม	40	

ชั้นมัธยฐาน

หา $\frac{N}{2} = \frac{40}{2} = 20$ พิจารณาชั้นที่มีความถี่สะสมมากกว่า $\frac{N}{2} = \frac{40}{2} = 20$ เป็นชั้นแรกเมื่อเรียงข้อมูลจากน้อยไปหามาก ดังนั้นจะได้ว่า 30 – 34 คือชั้นมัธยฐาน

จากสูตร $M_e = L + I\left(\frac{\frac{N}{2} - \sum f_L}{f}\right)$

ในที่นี้จะได้ว่า

$L = 29.5$, $I = 5$, $\frac{N}{2} = \frac{40}{2} = 20$, $\sum f_L = 17$, $f = 12$

แทนค่าต่างๆลงในสูตร ดังนี้

$$M_e = 29.5 + 5\left(\frac{20 - 17}{12}\right)$$

$$= 29.5 + 5(0.25)$$

$$= 29.5 + 1.25 = 30.75$$

ดังนั้นมัธยฐานของน้ำหนักของนักเรียนกลุ่มนี้ คือ 30.75 กิโลกรัม

3. ฐานนิยม (Mode , M_o)

ในกรณีที่ข้อมูลไม่ได้ถูกจัดกลุ่ม

ค่าฐานนิยม คือ ค่าของข้อมูลที่เกิดขึ้นบ่อยที่สุดหรือมีความถี่สูงสุด

ในกรณีที่ข้อมูลถูกจัดกลุ่ม *ต้องหาค่าฐานนิยมจากสูตร*

$$M_0 = L + I\left(\frac{d_1}{d_1 + d_2}\right)$$

โดยที่ L แทน ขอบเขตล่างของชั้นฐานนิยม

d_1 แทน ความแตกต่างระหว่างความถี่ของชั้นฐานนิยมกับชั้นก่อนฐานนิยมเมื่อเรียงข้อมูลจากน้อยไปหามาก

d_2 แทน ความแตกต่างระหว่างความถี่ของชั้นฐานนิยมกับชั้นหลังฐานนิยมเมื่อเรียงข้อมูลจากน้อยไปหามาก

I แทน ความกว้างของชั้นฐานนิยม

ตัวอย่างที่ 2.8 จากข้อมูลต่อไปนี้ จงหาค่าฐานนิยม

12 17 19 10 12 19 20

ฐานนิยม คือ 12

ตัวอย่างที่ 2.9 จากข้อมูลต่อไปนี้ จงหาค่าฐานนิยม

11 10 12 11 13 13 21 25

ฐานนิยม คือ 11 และ 13

ตัวอย่างที่ 2.10 จากข้อมูลต่อไปนี้ จงหาค่าฐานนิยม

4 7 8 12 11 13 5 9

ฐานนิยม คือ ไม่มีฐานนิยม

ตัวอย่างที่ 2.11 จากการสำรวจความสูงของนักศึกษาชายสาขาวิทยาการคอมพิวเตอร์ จำนวน 100 คน ได้ข้อมูลดังตาราง

ความสูงของนักศึกษา (เซนติเมตร)	จำนวน (คน)
134 – 144	8
145 – 154	27
155 – 164	42
165 – 174	18
175 – 184	5
รวม	100

จงหาฐานนิยมของความสูงของนักศึกษากลุ่มนี้

วิธีทำ จากข้อมูลในตารางจะได้ว่าชั้นที่มีความถี่มากที่สุด คือ 155 – 164 ดังนั้นชั้นฐานนิยมคือชั้นนี้ โดยที่ L = 154.5 , I = 10 , d_1 = 42 – 27 = 15 , d_2 = 42 – 18 = 24

$$M_0 = L + I\left(\frac{d_1}{d_1 + d_2}\right)$$

จากสูตร แทนค่าต่างๆลงในสูตร ดังนี้

$$M_0 = 154.5 + 10 \frac{15}{15+24}$$

$$= 154.5 + 10 \frac{15}{39}$$

$$= 154.5 + 10(0.385)$$

$$= 154.5 + 3.85 = 158.35$$

ดังนั้นฐานนิยมของความสูงของนักศึกษากลุ่มนี้ คือ 158.35 เซนติเมตร

การวัดการกระจาย (Measure of Dispersion)

การใช้สถิติเกี่ยวกับการวัดแนวโน้มเข้าสู่ส่วนกลาง ซึ่งเป็นค่าที่ทำหน้าที่เป็นตัวแทนกลุ่มข้อมูลเพียงอย่างเดียว เมื่อแปลความหมายข้อมูลจึงยังไม่สมบูรณ์ ไม่ชัดเจน และ มีโอกาสคลาดเคลื่อนได้ ดังนั้นสิ่งที่ควรนำมาพิจารณาควบคู่ไปกับการวัดแนวโน้มเข้าสู่ส่วนกลางก็คือ ลักษณะการกระจายของกลุ่มข้อมูล ซึ่งสถิติที่ใช้คือ การวัดการกระจาย โดยการที่ข้อมูลแต่ละชุดมีค่าต่าง ๆ กันนั้นเราเรียกว่า ข้อมูลมีการกระจาย ถ้าข้อมูลชุดนั้นประกอบ ด้วยค่าแตกต่างกันมาก เรียกว่า ข้อมูลมีการกระจายมาก ถ้าข้อมูลชุดนั้นประกอบด้วยค่าต่าง ๆ แตกต่างกันน้อย หรือมีค่าใกล้เคียงกันเรียกว่า ข้อมูลมีการกระจายน้อย ถ้าข้อมูลนั้นประกอบด้วยค่าต่าง ๆ เท่ากันหมด เรียกว่า ข้อมูลไม่มีการกระจาย สำหรับการวัดการกระจายที่นิยมใช้มีดังนี้

1. ค่าพิสัย (Range)

ค่าพิสัย หมายถึง การหาการกระจายของข้อมูลโดยนำข้อมูลที่มีค่าสูงที่สุด ลบกับข้อมูลที่มีค่าต่ำที่สุด เพื่อให้ได้ค่าที่เป็นช่วงของการกระจาย ซึ่งสามารถบอกถึงความกว้างของข้อมูลชุดนั้นๆ สำหรับสูตรที่ใช้ในการหาพิสัยคือ

พิสัย (R) = $X_{max} - X_{min}$

ตัวอย่างที่ 2.12 จงหาพิสัยจากข้อมูลชุดนี้

28	30	45	18	27	55	67	58	19	35

วิธีทำ จากสูตร พิสัย (R) = $X_{max} - X_{min}$

$= 67 - 18$

$= 49$ ดังนั้นจะได้ค่าพิสัย เท่ากับ 49

2. ค่าส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation หรือAverage Deviation : M.D.)

เป็นค่าเฉลี่ยของความแตกต่างระหว่างค่าสังเกตและค่าเฉลี่ยสำหรับสูตรที่ใช้ในการคำนวณหาค่าส่วนเบี่ยงเบนเฉลี่ยคือ

กรณีข้อมูลที่เก็บรวบรวมจากประชากร

$$MD = \frac{\sum_{i=1}^{N} \left| X_i - \mu \right|}{N}$$

กรณีข้อมูลที่เก็บรวบรวมจากตัวอย่าง

$$MD = \frac{\sum_{i=1}^{n} \left| X_i - \overline{X} \right|}{n}$$

ตัวอย่างที่ 2.13 จากการตรวจสอบเครื่องคอมพิวเตอร์ที่ต้องทำการซ่อมแซมในแต่ละเดือน จำนวน 12 เดือน พบจำนวนเครื่องคอมพิวเตอร์ที่ต้องทำการซ่อมแซมในแต่ละเดือน ดังนี้

5 12 8 4 13 7 10 9 3 2 1 2

จงหาค่าส่วนเบี่ยงเบนเฉลี่ย

วิธีทำ

$$\mu = \frac{5+12+8+\ldots+2}{12} = \frac{76}{12} = 6.33$$

จากสูตร

$$MD = \frac{\sum_{i=1}^{N} |X_i - \mu|}{N}$$

$$= \frac{|5-6.33| + |12-6.33| + |8-6.33| + \ldots + |2-6.33|}{12}$$

$$= \frac{1.33 + 5.67 + 1.67 + 2.33 + 6.67 + 0.67 + 3.67 + 2.67 + 3.33 + 4.33 + 5.33 + 4.33}{12}$$

$$= \frac{42}{12} = 3.5$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 3.5

3. ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.)

ส่วนเบี่ยงเบนมาตรฐานเป็นค่าวัดการกระจายที่สำคัญทางสถิติ เพราะเป็นค่าที่ใช้บอกถึงการกระจายของข้อมูลได้ดีกว่าค่าพิสัย และค่าส่วนเบี่ยงเบนเฉลี่ยและในการนำเสนอข้อมูลเบื้องต้นส่วนใหญ่จะใช้ค่าเฉลี่ยคู่กันกับส่วนเบี่ยงเบนมาตรฐาน สำหรับสูตรที่ใช้ในการคำนวณหาค่าส่วนเบี่ยงเบนมาตรฐานแบ่งออกเป็น 2 กรณีคือ

กรณีข้อมูลไม่ได้จัดกลุ่ม

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากประชากร

สูตรที่ใช้คือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}}{N}}$$

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง

สูตรที่ใช้คือ

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n - 1}}$$

กรณีข้อมูลจัดกลุ่ม

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากประชากร

สูตรที่ใช้คือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{R} f_i X_i^2 - \frac{\left(\sum_{i=1}^{R} f_i X_i\right)^2}{N}}{N}}$$

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง

สูตรที่ใช้คือ

$$S = \sqrt{\frac{\sum_{i=1}^{R} f_i x_i^2 - \frac{\left(\sum_{i=1}^{R} f_i x_i\right)^2}{n}}{n-1}}$$

**เมื่อนำค่าส่วนเบี่ยงเบนมาตรฐานมายกกำลังสองจะกลายเป็นค่าความแปรปรวน

ตัวอย่างที่ 2.14 จากการบันทึกข้อมูลจำนวนนักศึกษาที่เข้ามาใช้บริการห้องคอมพิวเตอร์สาขาวิชาสถิติประยุกต์ มหาวิทยาลัยราชภัฏบุรีรัมย์เป็นเวลาหนึ่งสัปดาห์ได้ข้อมูลดังนี้

12 15 10 7 9 14 7

จงหาค่าความแปรปรวนและค่าส่วนเบี่ยงเบนมาตรฐาน

วิธีทำ จากสูตรการหาค่าส่วนเบี่ยงเบนมาตรฐาน

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}}{N}}$$

หาค่าต่างๆ ดังนี้

$$\sum_{i=1}^{N} X_i^2 = 12^2 + 15^2 + 10^2 + 7^2 + 9^2 + 14^2 + 7^2 = 844$$

$$\sum_{i=1}^{N} X_i = 12 + 15 + 10 + 7 + 9 + 14 + 7 = 74$$

แทนค่าลงในสูตร ดังนี้

$$\sigma = \sqrt{\frac{844 - \frac{(74)^2}{7}}{7}}$$

$$= \sqrt{\frac{844 - 782.286}{7}} \quad = \sqrt{\frac{61.714}{7}} = \sqrt{8.816} = 2.969$$

$$\sigma^2 = 2.969^2 = 8.815$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนเท่ากับ 2.969 และ 8.815 ตามลำดับ

ตัวอย่างที่ 2.15 จงหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลสัมฤทธิ์ทางการเรียนวิชาสถิติสำหรับนักวิทยาศาสตร์ของนักเรียนกลุ่มตัวอย่างจำนวน 10 คน ซึ่งมีค่าดังต่อไปนี้

75	65	87	77	85	92	83	73	61	58

วิธีทำ จากสูตรการหาค่าส่วนเบี่ยงเบนมาตรฐาน

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

หาค่าต่างๆ ดังนี้

$$\sum_{i=1}^{n} x_i^2 = 75^2 + 65^2 + 87^2 + 77^2 + 85^2 + 92^2 + 83^2 + 73^2 + 61^2 + 58^2 = 58,340$$

$$\sum_{i=1}^{n} x_i = 75 + 65 + 87 + 77 + 85 + 92 + 83 + 73 + 61 + 58 = 756$$

แทนค่าลงในสูตร ดังนี้

$$S = \sqrt{\frac{58,340 - \frac{(756)^2}{10}}{10 - 1}}$$

$$= \sqrt{\frac{58,340 - 57,153.6}{9}}$$

$$= \sqrt{\frac{1186.4}{9}} = \sqrt{131.822} = 11.481$$

$$S^2 = 11.481^2 = 131.822$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนเท่ากับ 11.481 และ 131.822 ตามลำดับ

ตัวอย่างที่ 2.16 อาจารย์สอนในรายวิชาปฏิบัติการคอมพิวเตอร์ท่านหนึ่ง ต้องการทราบความสามารถในการใช้โปรแกรมคอมพิวเตอร์ชนิดหนึ่งของนักศึกษาห้องหนึ่ง จึงสุ่มตัวอย่างนักศึกษามา 25 คน และวัดความสามารถในการใช้โปรแกรมของนักศึกษาแต่ละคน คะแนนที่ได้นำมาแจกแจงความถี่ได้ดังต่อไปนี้

ชั้นคะแนน	ความถี่
79 - 81	2
76 - 78	3
73 - 75	4
70 - 72	8
67 - 69	5
64 - 66	2
61 - 63	1
รวม	25

จงคำนวณหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของคะแนนความสามารถในการใช้โปรแกรมของนักศึกษาห้องนี้

วิธีทำ จากสูตร

$$S = \sqrt{\frac{\sum_{i=1}^{R} f_i x_i^2 - \frac{\left(\sum_{i=1}^{R} f_i x_i\right)^2}{n}}{n - 1}}$$

คำนวณค่าต่างๆ ดังนี้

ชั้นคะแนน	ความถี่ (f_i)	จุดกึ่งกลางชั้น (X_i)	X_i^2	$f_i X_i^2$	$f_i X_i$
79 – 81	2	80	6,400	12,800	160
76 - 78	3	77	5,929	17,787	231
73 - 75	4	74	5,476	21,904	296
70 - 72	8	71	5,041	40,328	568
67 - 69	5	68	4,624	23,120	340
64 - 66	2	65	4,225	8,450	130
61 - 63	1	62	3,844	3,844	62
รวม	25			128,233	1,787

แทนค่าต่างๆลงในสูตร ดังนี้

$$S = \sqrt{\frac{128,233 - \frac{(1,787)^2}{25}}{25 - 1}}$$

$$= \sqrt{\frac{128,233 - 127,734.76}{24}}$$

$$= \sqrt{\frac{498.24}{24}} = \sqrt{20.76} = 4.556$$

$$S^2 = 4.556^2 = 20.757$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนเท่ากับ 4.556 และ 20.757 ตามลำดับ

4. ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variance)
หรือ สัมประสิทธิ์การกระจายซึ่งเป็นค่าที่ใช้วัดการกระจายของข้อมูลที่ไม่มีหน่วย
ถ้าข้อมูลชุดใดมีค่า C.V. มากจะมีการกระจายมากกว่าข้อมูลที่มีค่า C.V. น้อย

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากประชากร C.V. $= \frac{\sigma}{\mu} \times 100$
- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง C.V. $= \frac{S}{\overline{X}} \times 100$

การวัดความสัมพันธ์ระหว่างตัวแปร (measure of association)

ในการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัว ถ้าไม่ได้กำหนดว่าตัวแปรใดเป็นตัวแปรตาม ตัวแปรใดเป็นตัวแปรต้นหรือตัวแปรอิสระ จะเป็นเรื่องของการวิเคราะห์สหสัมพันธ์ (correlation analysis) ซึ่งเป็นการศึกษาว่าตัวแปรคู่นั้นมีขนาดของความสัมพันธ์ระดับใด โดยวัดด้วยค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าอยู่ระหว่าง -1 ถึง 1 เมื่อค่าเข้าใกล้ศูนย์

หมายความว่า ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน แต่ถ้าค่าเข้าใกล้ 1 หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์กันอย่างมาก ส่วนเครื่องหมาย + และเครื่องหมาย - จะแสดงถึงทิศทางความสัมพันธ์ ถ้าเป็นเครื่องหมาย + แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางเดียวกัน แต่ถ้าเป็นเครื่องหมาย - แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางตรงกันข้าม

1. สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนแรงค์ (Spearman Rank Correlation)

ถ้าข้อมูลที่ต้องการวัดความสัมพันธ์อยู่ในมาตราเรียงลำดับค่าสัมประสิทธิ์สหสัมพันธ์จะหาได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนแรงค์ โดยสูตรที่ใช้จะเป็นดังนี้

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

เมื่อ ρ แทน ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนแรงค์

d แทน ผลต่างของอันดับที่

n แทน จำนวนคู่ของข้อมูล

ตัวอย่างที่ 2.17 กรรมการ 2 คน ทำการประเมินความสะอาดของร้านอาหาร 12 แห่ง ในย่านชุมชนแห่งหนึ่ง ปรากฏผลดังนี้

ร้านอาหารที่	ผลการตัดสิน (เป็นอันดับที่)	
	กรรมการคนที่ 1	กรรมการคนที่ 2
1	2	3
2	3	4
3	5	5
4	8	6
5	6	8
6	4	2
7	1	1
8	7	7
9	10	9
10	9	11
11	12	10
12	11	12

จงหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการจัดอันดับที่ของกรรมการสองคนที่ได้จัดไว้

วิธีทำ หาค่า d คือ หาผลต่างของอันดับที่ที่กรรมการทั้งสองจัดอันดับให้ของร้านอาหารแต่ละแห่งพร้อมทั้งหาค่ากำลังสองของ d แต่ละค่า แล้วหาผลรวมทั้งหมดของ d^2 จะได้ดังนี้

ร้านอาหารที่	ผลการตัดสิน (เป็นอันดับที่)		d	d^2
	กรรมการคนที่ 1	กรรมการคนที่ 2		
1	2	3	-1	1
2	3	4	-1	1
3	5	5	0	0
4	8	6	2	4
5	6	8	-2	4
6	4	2	2	4
7	1	1	0	0
8	7	7	0	0
9	10	9	1	1
10	9	11	-2	4
11	12	10	2	4
12	11	12	-1	1
รวม				24

แทนค่าต่างลงในสูตร ดังนี้

$$\rho = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

$$= 1 - \frac{6(24)}{12\left(12^2 - 1\right)}$$

$$= 1 - \frac{144}{12(143)}$$

$$= 1 - \frac{144}{1716}$$

$$= 1 - 0.084 = 0.916$$

นั่นคือ การตัดสินของกรรมการทั้งสองมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.916

2. สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation)

ถ้าข้อมูลที่ต้องการวัดความสัมพันธ์อยู่ในมาตราอันตรภาค หรือมาตราอัตราส่วนค่าสัมประสิทธิ์สหสัมพันธ์จะหาได้ด้วยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเขียนแทนด้วยสัญลักษณ์ ρ แต่ในทางปฏิบัติจะหาค่า ρ ไม่ได้ เนื่องจากเป็นการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์จากตัวอย่างจึงต้องทำการประมาณค่าของ ρ ด้วย r

โดยมีสูตรการคำนวณดังนี้

$$r_{XY} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sqrt{\left(\sum X^2 - n\overline{X}^2\right)\left(\sum Y^2 - n\overline{Y}^2\right)}}$$

ตัวอย่างที่ 2.18 จากข้อมูลการผลิตสินค้าของบริษัทแห่งหนึ่ง โดยทำการบันทึกข้อมูลปริมาณการผลิต (X) กับเวลาที่ใช้ในการผลิต (Y) ในช่วงเวลา 15 เดือน ได้ข้อมูลดังนี้

เดือนที่	ปริมาณการผลิต (X)	เวลาที่ใช้ในการผลิต (Y)	X^2	Y^2	XY
1	30	75	900	5625	2250
2	70	152	4900	231 4	10640
3	20	55	400	3025	1100
4	50	110	2500	12100	5500
5	80	175	6400	30625	14000
6	30	73	900	5329	2190
7	20	50	400	2500	1000
8	60	128	3600	16384	7680
9	80	170	6400	28900	13600
10	40	87	1600	7569	3480
11	50	108	2500	11664	5400
12	60	135	3600	18225	8100
13	30	69	900	4761	2070
14	70	148	4 00	21904	10360
15	60	132	36 0	17424	7920
รวม	750	1,667	43,500	209,139	95,290

จงหาค่าสัมประสิทธิ์สหสัมพันธ์

วิธีทำ จากสูตร

$$r_{XY} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sqrt{\left(\sum X^2 - n\overline{X}^2\right)\left(\sum Y^2 - n\overline{Y}^2\right)}}$$

แทนค่าต่างๆลงในสูตร ดังนี้

$$r_{XY} = \frac{95,290 - 15\left(\frac{750}{15}\right)\left(\frac{1,667}{15}\right)}{\sqrt{[(43,500) - 15\left(\frac{750}{15}\right)^2][(209,139) - 15\left(\frac{1,667}{15}\right)^2]}}$$

$$= \frac{95,290 - 83349.75}{\sqrt{[43,500 - 37,500][209,139 - 185,258.155]}}$$

$$= \frac{11,940.25}{\sqrt{(6,000)(23,880.845)}} = \frac{11,940.25}{\sqrt{143,285,070}}$$

$$= \frac{11,940.25}{11,970.174} = 0.9975$$

ดังนั้น แสดงว่า ปริมาณการผลิตและเวลาที่ใช้ในการผลิตมีความสัมพันธ์กัน 0.9975