เอกสารประกอบการค้นคว้า

3000 – 1524 วิชาสถิติ (Statistics) 3 (3)

ศักดิ์สิทธิ์ วัชรารัตน์

วิทยาลัยสารพัดช่างพิษณุโลก

สำนักงานคณะกรรมการการอาชีวศึกษา กระทรวงศึกษาธิการ

พ.ศ. 2553

เอกสารประกอบการค้นคว้า

3000 – 1524 วิชาสถิติ (Statistics) 3 (3)

โดย

ศักดิ์สิทธิ์ วัชรารัตน์

ครูชำนาญการพิเศษ

วิทยาลัยสารพัดช่างพิษณุโลก

สำนักงานคณะกรรมการการอาชีวศึกษา กระทรวงศึกษาธิการ

พ.ศ. 2553

คำนำ

การวิจัยการพัฒนาชุดโปรแกรมช่วยสอนวิชาสถิติ เพื่อสร้างโจทย์ฝึกทักษะและแบบทดสอบนี้ จะประกอบด้วยชุดของเอกสารประกอบการค้นคว้าจำนวน 2 เรื่อง คือ วิชาสถิติ และการเลือกใช้ตัวสถิติ และนวัตกรรม 1 ชุด คือ ชุดโปรแกรมช่วยสอนวิชาสถิติ จำนวน 12 โปรแกรมย่อย บรรจุอยู่ภายในซีดี 1 แผ่น ซึ่งได้พัฒนาขึ้นมาจากการเขียนสูตรสถิติในโปรแกรม Microsoft Excel 2007 แล้วบันทึกเป็นแฟ้ม Microsoft Excel 2007 และ 2003

เอกสารประกอบการค้นคว้าวิชาสถิตินี้ ก็เป็นข้อมูลจากการตรวจเอกสารในการวิจัยดังกล่าวที่ได้เรียบเรียงจากบทความ วารสาร เอกสาร ตำรา และเว็บไซต์ต่าง ๆ และหวังเป็นอย่างยิ่งที่เอกสารประกอบการค้นคว้าวิชาสถิตินี้ จะเป็นประโยชน์ต่อครูผู้สอน ผู้เรียน และผู้สนใจทั่วไป ที่จะได้นำมาใช้เป็นข้อมูลในการเรียนการสอน และการวิจัยทางการศึกษาต่อไป

ท้ายนี้ ขอขอบพระคุณครูบาอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้แก่ข้าพเจ้า ทั้งขอขอบคุณและขออนุญาตผู้เขียนบทความ หนังสือ ตำรา และเอกสารทุก ๆ ท่านที่ข้าพเจ้าได้รวบรวมเรียบเรียงและได้นำข้อความรูปภาพบางส่วนมาประกอบในเอกสารค้นคว้าเหล่านี้ สำหรับส่วนดีที่มีคุณค่าทั้งหลายที่เกิดขึ้น ขออุทิศส่วนกุศลให้กับคุณบิดามารดา เจ้ากรรมนายเวร เทพเจ้าทั้งหลาย พระยายมราชญาติพี่น้อง เพื่อสนิทมิตรสหาย ทั้งที่มีชีวิตและที่ล่วงลับไปแล้ว หากมีข้อผิดพลาดประการใด กรุณาแจ้งมาทางอีเมล (saksit2500@gmail.com) เพื่อจะได้พัฒนาปรับปรุงแก้ไขให้ถูกต้องและมีความสมบูรณ์ยิ่ง ๆ ขึ้นต่อไปและจักขอบคุณเป็นอย่างยิ่ง

ศักดิ์สิทธิ์ วัชรารัตน์

กันยายน 2553

สารบัญ

สารบัญตาราง

สารบัญรูป

สารบัญรูป

หลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) พุทธศักราช 2545 (ปรับปรุง พ.ศ. 2546)

สำนักงานคณะกรรมการการอาชีวศึกษา กระทรวงศึกษาธิการ

1. หลักการ

1.1 เป็นหลักสูตรที่มุ่งผลิตและพัฒนาแรงงานระดับผู้ชำนาญการเฉพาะสาขาอาชีพสอดคล้องกับตลาดแรงงาน สภาพเศรษฐกิจ สังคม วัฒนธรรม เทคโนโลยีและสิ่งแวดล้อม สามารถเป็นหัวหน้างานหรือเป็นผู้ประกอบการได้

1.2 เป็นหลักสูตรที่มุ่งเน้นให้ผู้เรียนมีสมรรถนะในการประกอบอาชีพ มีความรู้เต็มภูมิ ปฏิบัติได้จริง และเข้าใจชีวิต

1.3 เป็นหลักสูตรที่เปิดโอกาสให้ผู้ประกอบการวิชาชีพมีส่วนร่วมในการเรียนการสอนวิชาชีพ สามารถถ่ายโอนประสบการณ์การเรียนรู้จากสถานประกอบการ และสามารถสะสมการเรียนรู้และประสบการณ์ได้

2. จุดหมาย

2.1 เพื่อให้มีความรู้และทักษะพื้นฐานในการดำรงชีวิต สามารถศึกษาค้นคว้าเพิ่มเติมหรือศึกษาต่อในระดับที่สูงขึ้น

2.2 เพื่อให้มีทักษะและสมรรถนะในงานอาชีพตามมาตรฐานวิชาชีพ

2.3 เพื่อให้สามารถบูรณาการความรู้ ทักษะจากศาสตร์ต่าง ๆ ประยุกต์ใช้ในงานอาชีพสอดคล้องกับการเปลี่ยนแปลงทางเทคโนโลยี

2.4 เพื่อให้มีเจตคติที่ดีต่ออาชีพ มีความมั่นใจและภาคภูมิใจในงานอาชีพ รักงาน รักองค์กร สามารถทำงานเป็นหมู่คณะได้ดี และมีความภาคภูมิใจในตนเองต่อการเรียนวิชาชีพ

2.5 เพื่อให้มีปัญญา ใฝ่รู้ ใฝ่เรียน มีความคิดสร้างสรรค์ มีความสามารถในการจัดการ การตัดสินใจ และการแก้ปัญหา รู้จักแสวงหาแนวทางใหม่ ๆ มาพัฒนาตนเอง ประยุกต์ใช้ความรู้ในการสร้างงานให้สอดคล้องกับวิชาชีพ และการพัฒนางานอาชีพอย่างต่อเนื่อง

2.6 เพื่อให้มีบุคลิกภาพที่ดี มีคุณธรรม จริยธรรม ซื่อสัตย์ มีวินัย มีสุขภาพสมบูรณ์ แข็งแรงทั้งร่างกายและจิตใจ เหมาะสมกับการปฏิบัติในอาชีพนั้น ๆ

2.7 เพื่อให้เป็นผู้มีพฤติกรรมทางสังคมที่ดีงาม ทั้งในการทำงาน การอยู่ร่วมกัน มีความรับผิดชอบต่อครอบครัว องค์กร ท้องถิ่นและประเทศชาติ อุทิศตนเพื่อสังคม เข้าใจและเห็นคุณค่าของศิลปวัฒนธรรมไทย ภูมิปัญญาท้องถิ่น ตระหนักในปัญหาและความสำคัญของสิ่งแวดล้อม

2.8 เพื่อให้ตระหนักและมีส่วนร่วมในการพัฒนาและแก้ไขปัญหาเศรษฐกิจของประเทศ โดยเป็นกำลังสำคัญในด้านการผลิตและให้บริการ

2.9 เพื่อให้เห็นคุณค่าและดำรงไว้ซึ่งสถาบันชาติ ศาสนา และพระมหากษัตริย์ ปฏิบัติตนในฐานะพลเมืองดี ตามระบอบประชาธิปไตย อันมีพระมหากษัตริย์ทรงเป็นประมุข

3. โครงสร้าง

โครงสร้างหลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง พุทธศักราช 2545 (ปรับปรุง พ.ศ. 2546) แบ่งเป็น 3 หมวดวิชา ฝึกงาน และกิจกรรมเสริมหลักสูตร ดังนี้

3.1 หมวดวิชาสามัญ

3.1.1 วิชาสามัญทั่วไป เป็นวิชาที่เป็นพื้นฐานในการดำรงชีวิต

3.1.2 วิชาสามัญพื้นฐานวิชาชีพ เป็นวิชาที่เป็นพื้นฐานสัมพันธ์กับวิชาชีพ

3.2 หมวดวิชาชีพ แบ่งเป็น

3.2.1 วิชาชีพพื้นฐาน เป็นกลุ่มวิชาชีพสัมพันธ์ที่เป็นพื้นฐานที่จำเป็นในประเภทวิชานั้น ๆ

3.2.2 วิชาชีพสาขาวิชา เป็นกลุ่มวิชาชีพหลักในสาขาวิชานั้น ๆ

3.2.3 วิชาชีพสาขางาน เป็นกลุ่มวิชาชีพที่มุ่งให้ผู้เรียนมีความรู้และทักษะเฉพาะด้านในงานอาชีพตามความถนัดและความสนใจ

3.2.4 โครงการ

3.3 หมวดวิชาเลือกเสรี

3.4 ฝึกงาน

3.5 กิจกรรมเสริมหลักสูตร

จำนวนหน่วยกิต และรายวิชาของแต่ละหมวดวิชาตลอดหลักสูตร ให้เป็นไปตามกำหนดไว้ในโครงสร้างของแต่ละประเภทวิชาและสาขาวิชา ส่วนรายวิชาแต่ละหมวด สถานศึกษาสามารถจัดตามที่กำหนดไว้ในหลักสูตร หรือจัดตามความเหมาะสมของสภาพท้องถิ่น ทั้งนี้สถานศึกษาต้องกำหนดรหัสวิชา จำนวนคาบเรียนและจำนวนหน่วยกิต ตามระเบียบที่กำหนดไว้ในหลักสูตร

การจัดการเรียนการสอนวิชาคณิตศาสตร์และสถิติ
หลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) พุทธศักราช 2545 (ปรับปรุง พ.ศ. 2546)

1. คำอธิบายรายวิชาคณิตศาสตร์และสถิติ

รายวิชาสถิติ เป็นรายวิชาที่อยู่ในหมวดวิชาสามัญ วิชาสามัญพื้นฐานวิชาชีพ สังกัดกลุ่มวิชาคณิตศาสตร์ ดังนี้

รหัสวิชา ชื่อวิชา - หน่วยกิต (ชั่วโมง)

3000-1520 คณิตศาสตร์ 1 - 3 (3)

3000-1521 คณิตศาสตร์ 2 - 3 (3)

3000-1522 คณิตศาสตร์ 3 - 3 (3)

3000-1523 คณิตศาสตร์ 4 - 3 (3)

3000-1524 สถิติ - 3 (3)

3000-1520 คณิตศาสตร์ 1 3 (3) (Mathematics 1) (บริหารธุรกิจ)

จุดประสงค์รายวิชา

1. เพื่อให้มีความรู้ความเข้าใจเรื่อง ตรรกศาสตร์ ความน่าจะเป็น พื้นที่ใต้โค้งปกติ การสุ่มตัวอย่าง

2. เพื่อให้สามารถนำความรู้เรื่อง ตรรกศาสตร์ ความน่าจะเป็น พื้นที่ใต้โค้งปกติ การสุ่มตัวอย่างไปใช้ประกอบในวิชาชีพ

3. เพื่อให้มีเจตคติที่ดี และเกิดความคิดรวบยอดเกี่ยวกับ ตรรกศาสตร์ ความน่าจะเป็น พื้นที่ใต้โค้งปกติ การสุ่มตัวอย่าง

มาตรฐานรายวิชา

1. มีความรู้ ความเข้าใจ เกี่ยวกับตรรกศาสตร์และนำไปใช้สรุปในการอ้างเหตุผลได้

2. มีความคิดรวบยอดเกี่ยวกับความน่าจะเป็น และคำนวณความน่าจะเป็นของเหตุการณ์ได้

3. คำนวณพื้นที่ใต้โค้งปกติและนำไปคำนวณความน่าจะเป็นของตัวแปรที่ต่อเนื่องได้

4. มีความคิดรวบยอดเกี่ยวกับการสุ่มตัวอย่าง และการแจกแจงค่าเฉลี่ยที่ได้จากตัวอย่างประชาได้

คำอธิบายรายวิชา

ศึกษาเกี่ยวกับความรู้เบื้องต้นเกี่ยวกับเซต ประพจน์ การเชื่อมและค่าความจริงของประพจน์ ตารางค่าความจริง ประพจน์ที่สมมูลกัน ประโยคเปิด และตัวแปรปริมาณตัวเดียว การอ้างเหตุผล กฎเกณฑ์เบื้องต้นเกี่ยวกับการนับ แฟกทอเรียล วิธีเรียงสับเปลี่ยน วิธีจัดหมู่ การทดลองสุ่ม และแซมเปิลสเปซ เหตุการณ์และความน่าจะเป็นของเหตุการณ์ ความน่าจะเป็นแบบมีเงื่อนไข และเหตุการณ์ที่เป็นอิสระต่อกันพื้นที่ภายใต้โค้งปกติ การสุ่มตัวอย่าง วิธีและขั้นตอนของการสุ่มตัวอย่าง ค่าพารามิเตอร์และค่าสถิติ การแจกแจงค่าเฉลี่ยที่ได้จากตัวอย่างประชากร

3000-1521 คณิตศาสตร์ 2 3 (3) (Mathematics 2) (อุตสาหกรรม)

จุดประสงค์รายวิชา

1. เพื่อให้มีความรู้ความเข้าใจเรื่อง ฟังก์ชันแบบต่าง ๆ เมตริกซ์และดีเทอร์มินันท์ และเรขาคณิตวิเคราะห์

2. เพื่อให้สามารถนำความรู้เรื่อง ฟังก์ชันแบบต่าง ๆ เมตริกซ์และดีเทอร์มินันท์ และเรขาคณิตวิเคราะห์ ไปใช้ประกอบในวิชาชีพ

3. เพื่อให้มีเจตคติที่ดี และเกิดความคิดรวบยอดเกี่ยวกับ ฟังก์ชันแบบต่าง ๆ เมตริกซ์ และดีเทอร์มินันท์ และเรขาคณิตวิเคราะห์

มาตรฐานรายวิชา

1. มีความรู้ความเข้าใจเกี่ยวกับฟังก์ชันแบบต่าง ๆ และนำไปใช้ในการแก้ปัญหา

2. มีความรู้ความเข้าใจทฤษฎีบททวินามและเศษส่วนย่อย และนำไปใช้ในการแก้ปัญหา

3. มีความรู้ความเข้าใจเกี่ยวกับเมตริกซ์และดีเทอร์มินันท์ และนำไปใช้ในการแก้ปัญหา

4. มีความรู้ความเข้าใจเกี่ยวกับเรขาคณิตวิเคราะห์ และนำไปใช้ในการแก้ปัญหา

5. สามารถนำความรู้เกี่ยวกับฟังก์ชันแบบต่าง ๆ เมตริกซ์และดีเทอร์มินันท์ และเรขาคณิตวิเคราะห์ไปใช้เป็นพื้นฐานประกอบในวิชาชีพ

คำอธิบายรายวิชา

ศึกษาเกี่ยวกับฟังก์ชันเอกซ์โปเนนเชียล และลอกริทึม ฟังชั่นตรีโกนมิติ และอินเวอร์ส ฟังก์ชันตรีโกนมิติ กฎของไซน์และกฎของโคไซน์ ทฤษฎีบททวินาม เศษส่วนย่อย ชนิดของ เมทริกซ์ การบวกและลบเมตริกซ์ การคูณเมตริกซ์ด้วยจำนวนจริง การคูณเมตริกซ์ด้วยเมตริกซ์ ดีเทอร์มินันท์ อินเวอร์สการคูณเมทริกซ์ การแก้สมการเชิงเส้นด้วยเมทริกซ์และดีเทอร์มินันท์ ระยะทางระหว่างจุดสองจุด จุดกึ่งกลางระหว่างจุดสองจุด ความชัน รูปแบบของสมการเส้นตรง ระยะทางระหว่างจุดกับเส้นตรง ระยะทางระหว่างเส้นตรงกับเส้นตรง ภาคตัดกรวยที่มีจุดศูนย์กลางหรือจุดยอดอยู่ที่จุดใด ๆ ในระนาบ

3000-1522 คณิตศาสตร์ 3 3 (3) (Mathematics 3) (คหกรรม-ศิลปกรรม)

จุดประสงค์รายวิชา

1. เพื่อให้มีความรู้ความเข้าใจ เรื่อง สถิติเบื้องต้น เซต ตรรกศาสตร์ ความน่าจะเป็น

2. เพื่อให้สามารถนำความรู้เรื่อง สถิติเบื้องต้น เซต ตรรกศาสตร์ ความน่าจะเป็น ไปใช้ประกอบในวิชาชีพ และชีวิตประจำวัน

3. เพื่อให้มีเจตคติที่ดี และเกิดความคิดรวบยอดเกี่ยวกับ สถิติเบื้องต้น เซต ตรรกศาสตร์ ความน่าจะเป็น

มาตรฐานรายวิชา

1. นำวิธีการทางสถิติไปใช้ในการวิเคราะห์ข้อมูลได้

2. มีความคิดรวบยอดในเรื่องเซต การดำเนินการของเซต และการแก้ปัญหาโดยใช้เซต

3. มีความรู้ความเข้าใจเกี่ยวกับตรรกศาสตร์ และนำไปใช้สรุปในการอ้างเหตุผลได้

4. มีความคิดรวบยอดเกี่ยวกับความน่าจะเป็น และคำนวณความน่าจะเป็นของเหตุการณ์ได้

5. นำความรู้เกี่ยวกับความน่าจะเป็น ช่วยในการตัดสินใจ และแก้ปัญหาได้

คำอธิบายรายวิชา

ศึกษาเกี่ยวกับค่ากลาง ส่วนเบี่ยงเบนมาตรฐาน สัมประสิทธิ์ของการแปรผัน เซต การดำเนินการของเซต การแก้โจทย์ปัญหาโดยใช้เซต ประพจน์ การเชื่อมและการหาค่าความจริงของประพจน์ ตารางแสดงค่าความจริง ประพจน์ที่สมมูลกัน ประโยคเปิด การอ้างเหตุผล กฎเกณฑ์เบื้องต้นเกี่ยวกับการนับ แฟคทอเรียล วิธีเรียงสับเปลี่ยน วิธีจัดหมู่ การทดลองสุ่ม แซมเปิลสเปซ เหตุการณ์ และความน่าจะเป็นของเหตุการณ์

3000-1523 คณิตศาสตร์ 4 3 (3) (Mathematics 4) (เกษตรกรรม)

จุดประสงค์รายวิชา

1. เพื่อให้มีความรู้ความเข้าใจเรื่อง เมทริกซ์ ตรรกศาสตร์ กำหนดการเชิงเส้น และการคำนวณภาษี

2. เพื่อให้สามารถนำความรู้เรื่อง เมทริกซ์ ตรรกศาสตร์ กำหนดการเชิงเส้น และการคำนวณภาษีไปใช้ประกอบในวิชาชีพและชีวิตประจำวัน

3. เพื่อให้มีเจตคติที่ดี และเกิดความคิดรวบยอดเกี่ยวกับ เมทริกซ์ ตรรกศาสตร์ กำหนดการ เชิงเส้น และการคำนวณภาษี

มาตรฐานรายวิชา

1. มีความรู้ความเข้าใจเกี่ยวกับเมตริกซ์ และดีเทอมินันท์ และนำไปใช้ในวิชาชีพได้

2. มีความรู้ความเข้าใจเกี่ยวกับตรรกศาสตร์ และนำไปใช้สรุปในการอ้างเหตุผลได้

3. มีความรู้ความเข้าใจเกี่ยวกับกำหนดการเชิงเส้น และนำไปใช้ในวิชาชีพได้

4. สามารถคำนวณภาษีเงินได้บุคคลธรรมดา และภาษีมูลค่าเพิ่มอย่างถูกต้อง

คำอธิบายรายวิชา

ศึกษาเกี่ยวกับความรู้เบื้องต้นเกี่ยวกับเซต ประพจน์ การเชื่อมประพจน์ และค่าความจริงของประพจน์ ตารางค่าความจริง การอ้างเหตุผล ความหมายของเมตริกซ์ การบวกและลบเมตริกซ์ การคูณเมตริกซ์ การดำเนินการเปลี่ยนแถวเชิงมูลฐาน เมตริกซ์ลดรูปเป็นขั้นแบบแถว การหาคำตอบของสมการเชิงเส้นโดยวิธีเกาส์ เมตริกซ์ผกผัน การหาคำตอบโดยใช้เมตริกซ์ผกผัน ดีเทอร์มินันท์ กฎของคราเมอร์ ความหมายของกำหนดการเชิงเส้น การคาดหมายโดยใช้ กำหนดการเชิงเส้นด้วยวิธีการใช้กราฟ และวิธีการคำนวณแบบง่าย ภาษีเงินได้บุคคลธรรมดาและภาษีมูลค่าเพิ่ม

3000-1524 สถิติ 3 (3) (Statistics)

จุดประสงค์รายวิชา

1. เพื่อให้มีความรู้ความเข้าใจ เรื่อง ความหมายของสถิติ และวิธีการคำนวณทางสถิติไปใช้ในการวิเคราะห์เบื้องต้น

2. เพื่อให้สามารถนำความรู้เรื่อง ความหมายของสถิติ และวิธีการคำนวณทางสถิติไปใช้ในการวิเคราะห์เบื้องต้น ไปใช้ในงานอาชีพ และชีวิตประจำวัน

3. เพื่อให้มีเจตคติที่ดี และเกิดความคิดรวบยอดเกี่ยวกับ ความหมายของสถิติ และวิธีการคำนวณทางสถิติไปใช้ในการวิเคราะห์เบื้องต้น

มาตรฐานรายวิชา

1. มีความรู้ความเข้าใจเกี่ยวกับสถิติ นำไปใช้ในงานอาชีพและชีวิตประจำวันได้

2. สามารถแปลความหมายจากค่าสถิติที่คำนวณได้

3. นำสถิติไปวิเคราะห์ข้อมูล และสรุปผล ในงานอาชีพและชีวิตประจำวันได้

คำอธิบายรายวิชา

ศึกษาเกี่ยวกับความหมายและขอบข่ายของสถิติ การวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดการกระจายของข้อมูล คะแนนมาตรฐาน การประมาณค่าเฉลี่ยประชากร การทดสอบสมมติฐานทางสถิติ การทดสอบ

เกี่ยวกับค่าเฉลี่ยของประชากร และสัดส่วนของประชากร การทดสอบ ไคสแควร์ การวิเคราะห์ถดถอย และสหสัมพันธ์

2. การจัดการเรียนการสอนวิชาสถิติ

รายวิชาสถิติ มีรายละเอียดดังนี้ ชื่อวิชา สถิติ รหัส 3000-1524 (ท-ป-น) 3-0-3 ระดับ ปวส. หน่วยกิต 3 หน่วย จัดการเรียนการสอนภาคเรียนละ 18 สัปดาห์ เป็นทฤษฎี 3 ชั่วโมงต่อสัปดาห์ รวมจำนวน 54 ชั่วโมง

จากคำอธิบายรายวิชา นำมาจัดทำเป็นเนื้อหารายวิชาได้ 10 หน่วย ดังนี้

หน่วยที่ 1 ความหมายและข้อมูลทางสถิติ

หน่วยที่ 2 การวิเคราะห์ข้อมูลเบื้องต้น

หน่วยที่ 3 ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น

หน่วยที่ 4 การแจกแจงของค่าสถิติ

หน่วยที่ 5 การอนุมานเชิงสถิติ

หน่วยที่ 6 การประมาณค่า

หน่วยที่ 7 การทดสอบสมมติฐาน

หน่วยที่ 8 การวิเคราะห์ความแปรปรวน

หน่วยที่ 9 การทดสอบไคสแควร์

หน่วยที่ 10 การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายและสหสัมพันธ์

3. การจัดการเรียนการสอนด้านเนื้อหาวิชาสถิติ

3.1 หน่วยที่ 1 ความหมายและข้อมูลทางสถิติ

3.1.1 ความหมายและขอบข่ายของสถิติ

สถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistik มีรากศัพท์มาจาก Stat หมายถึงข้อมูลหรือสารสนเทศ

สถิติ (Statistics) โดยทั่วไปอาจจะแบ่งเป็น 3 ความหมาย คือ

1) สถิติ หมายถึง ตัวเลขแทนปริมาณจำนวนข้อมูล หรือข้อเท็จจริงของสิ่งต่าง ๆ ที่คนโดยทั่วไปต้องการศึกษาหาความรู้ เช่น ต้องการทราบปริมาณน้ำฝนที่ตกในจังหวัดพิษณุโลก ปี 2552 เป็นต้น

2) สถิติ หมายถึง ค่าตัวเลขที่เกิดจากการคำนวณมาจากกลุ่มตัวอย่าง (Sample) หรือคิดมาจากนิยามทางคณิตศาสตร์ เช่นคำนวณหาค่าเฉลี่ย ค่าความแปรปรวน ค่าที่คำนวณได้เรียกว่า ค่าสถิติ (A Statistic) ส่วนค่าสถิติทั้งหลายเรียกว่า ค่าสถิติหลาย ๆ ค่า (Statistics)

3) สถิติ หมายถึง วิชาการแขนงหนึ่งที่จัดเป็นวิชาวิทยาศาสตร์ และเป็นทั้งวิทยาศาสตร์บริสุทธิ์และวิทยาศาสตร์ประยุกต์ และยังหมายรวมถึงระเบียบวิธีการสถิติอันประกอบไปด้วยขั้นตอน 4

ขั้นตอนที่ใช้ในการศึกษา ได้แก่ การเก็บรวบรวมข้อมูล (Collection of Data) การนำเสนอข้อมูล (Presentation of Data) การวิเคราะห์ข้อมูล (Analysis of Data) และการตีความหมายของข้อมูล (Interpretation of Data)

ขอบข่ายหรือประเภทของสถิติ สามารถแบ่งตามลักษณะของข้อมูลได้เป็น 2 ประเภท คือ

1) สถิติเชิงอนุมาน (Inductive Statistics) หมายถึง สถิติที่ใช้จัดกระทำกับข้อมูลที่ได้มาเพียงบางส่วนของข้อมูลทั้งหมด

2) สถิติเชิงบรรยาย (Descriptive Statistics) หมายถึง สถิติที่ใช้จัดกระทำกับข้อมูลที่ได้มาเฉพาะเรื่องใดเรื่องหนึ่ง

3.1.2 ความหมายและประเภทของข้อมูล

ข้อมูล (Data) หมายถึง รายละเอียดข้อเท็จจริง (facts) ที่เกี่ยวกับเรื่องต่าง ๆ ทั้งที่เป็นรูปธรรม และนามธรรม ซึ่งตรงกับสิ่งที่ผู้ต้องการศึกษา ตัวอย่างข้อเท็จจริงที่เป็นตัวเลข เช่น จำนวนผู้ป่วยที่ติดเชื้อ HIV ในหมู่บ้าน ราคาของพืชผักและผลไม้ต่าง ๆ ในหมู่บ้าน เป็นต้น ตัวอย่างข้อเท็จจริงที่ไม่ใช่ตัวเลข เช่น การศึกษา หรือ อาชีพของคนในหมู่บ้าน เป็นต้น

การจำแนกประเภทของข้อมูล สามารถจำแนกได้หลายลักษณะของการจัดแบ่ง ดังนี้

1) เมื่อจำแนกตามลักษณะของข้อมูล สามารถแบ่งออกได้เป็น 2 ชนิดคือ

1.1) ข้อมูลเชิงคุณภาพ (Qualitative Data) หมายถึง ข้อมูลที่ไม่สามารถบอกได้ว่า มีค่ามากหรือน้อย แต่จะสามารถบอกได้ว่าดีหรือไม่ดี หรือบอกลักษณะความเป็นกลุ่มของ ข้อมูล เช่น เพศ ศาสนา สีผม คุณภาพสินค้า ความพึงพอใจ ฯลฯ

1.2) ข้อมูลเชิงปริมาณ (Quantitative Data) หมายถึง ข้อมูลที่สามารถวัดค่าได้ ว่ามีค่ามากหรือน้อยซึ่งสามารถวัดค่าออกมาเป็นตัวเลขได้ เช่น คะแนนสอบ อุณหภูมิ ส่วนสูง น้ำหนัก ปริมาณต่าง ๆ ฯลฯ ซึ่งข้อมูลเชิงปริมาณยังสามารถแบ่งออกได้เป็นอีก 2 ลักษณะคือ

1.2.1) ข้อมูลเชิงปริมาณแบบต่อเนื่อง (Continues Data) หมายถึง ข้อมูลที่เป็นจำนวนจริงซึ่งสามารถบอกหรือระบุได้ทุกค่าที่กำหนดเช่น จำนวน 0 – 1 ซึ่งมีค่ามากมายนับไม่ถ้วน และเป็นเส้นจำนวนแบบไม่ขาดตอน

1.2.2) ข้อมูลเชิงปริมาณแบบไม่ต่อเนื่อง (Discrete Data) หมายถึงข้อมูลที่เป็นจำนวนเต็ม หรือจำนวนนับ เช่น 0 , 1 , 2 , … ,…, 100 ฯลฯ หรือ 0.1 , 0.2 , 0.3 , … , … ซึ่งในช่องว่างของแต่ละค่าของข้อมูลจะไม่มีค่าอื่นใดมาแทรก

2) เมื่อจำแนกตามแหล่งที่มาของข้อมูล สามารถแบ่งออกได้เป็น 2 ชนิด คือ

2.1) ข้อมูลปฐมภูมิ (Primary Data) หมายถึง ข้อมูลที่ผู้ใช้เป็นผู้เก็บรวบรวมข้อมูลขึ้นเอง เช่น การเก็บแบบสอบถาม การทดลองในห้องทดลอง ฯลฯ

2.2) ข้อมูลทุติยภูมิ (Second Data) หมายถึง ข้อมูลที่ผู้ใช้นำมาจากหน่วยงานอื่น หรือผู้อื่น ที่ได้ทำการเก็บรวบรวมมาแล้วในอดีต เช่น รายงานประจำปีของหน่วยงานต่าง ๆ ข้อมูลท้องถิ่นเป็นผู้รวบรวมไว้ ฯลฯ

3) เมื่อจำแนกตามระดับการวัด สามารถแบ่งออกได้เป็น 4 ชนิด คือ

3.1) ข้อมูลระดับนามบัญญัติ (Nominal Scale) หมายถึง ข้อมูลที่แบ่งเป็นกลุ่มเป็นพวก เช่น เพศ อาชีพ ศาสนา ผิวสี ฯลฯ ไม่สามารถนำมาจัดลำดับ หรือนำมาคำนวณได้

3.2) ข้อมูลระดับอันดับ (Ordinal Scale) หมายถึง ข้อมูลที่สามารถแบ่งเป็นกลุ่มได้ แล้วยังสามารถบอกอันดับที่ของความแตกต่างได้ แต่ไม่สามารถบอกระยะห่างของอันดับที่แน่นนอนได้ หรือไม่สามารถเปรียบเทียบได้ว่าอันดับที่จัดนั้นมีความแตกต่างกันของระยะห่างเท่าใด เช่น อันดับที่ของการสอบของนักศึกษา อันดับที่ของผู้เข้าประกวดนางสาวไทย ฯลฯ

3.3) ข้อมูลระดับช่วงชั้นอันตรภาค (Interval Scale) หมายถึง ข้อมูลที่มีช่วงห่าง หรือระยะห่างเท่าๆ กัน สามารถวัดค่าได้แต่เป็นข้อมูลที่ไม่มีศูนย์แท้ เช่น อุณหภูมิ คะแนนสอบ GPA คะแนน I.Q. ฯลฯ

3.4) ข้อมูลระดับอัตราส่วน (Ratio Scale) หมายถึง ข้อมูลที่มีมาตราวัดหรือระดับการวัดที่สูงที่สุด คือนอกจากสามารถแบ่งกลุ่มได้ จัดอันดับได้ มีช่วงห่างของข้อมูลเท่าๆกันแล้ว ยังเป็นข้อมูลที่มีศูนย์แท้เช่น น้ำหนัก ส่วนสูง ระยะทาง รายได้ จำนวนต่าง ๆ ฯลฯ

3.1.3 การเก็บรวบรวมข้อมูล

1) การเก็บรวบรวมข้อมูลจากข้อมูลปฐมภูมิ สามารถรวบรวมจากผู้ที่ให้ข้อมูลหรือแหล่งที่มาโดยตรง

1.1) การสำมะโน (Census) คือ การเก็บรวบรวมข้อมูลจากทุกหน่วยของประชากรที่ต้องการศึกษา

1.2) การสำรวจจากกลุ่มตัวอย่าง (Sample Survey) คือ การเก็บรวบรวมข้อมูลที่ประกอบด้วยตัวแทนจากทุกลักษณะของประชากรที่ต้องการศึกษา

ในทางปฏิบัติ ไม่ว่าจะทำการสำมะโนหรือการสำรวจ นิยมปฏิบัติอยู่ 5 วิธี คือ

ก. การสัมภาษณ์ (Interview) นิยมใช้กันมาก เพราะจะได้คำตอบทันที นอกจากนี้หากผู้ตอบไม่เข้าใจก็สามารถอธิบายเพิ่มเติมได้ แต่ผู้สัมภาษณ์ต้องซื่อสัตย์ และเข้าใจจุดมุ่งหมายของการเก็บข้อมูลอย่างแท้จริง

ข. การแจกแบบสอบถาม (Questionnaire) วิธีนี้ประหยัดเวลาและค่าใช้จ่ายมาก สะดวกและสบายใจต่อการตอบแบบสอบถาม แต่ก็มีข้อเสียหลายประการ เช่น ต้องใช้ในเฉพาะผู้ที่มีการศึกษา มีไปรษณีย์ไปถึง คำถามต้องชัดเจน อาจจะไม่ได้รับคืนตามเวลาหรือจำนวนที่ต้องการ จึงต้องส่งแบบสอบถามออกไปเป็นจำนวนมาก ๆ หรือไปแจกและเก็บด้วยตนเอง

ค. การสังเกต (Observation) เป็นข้อมูลที่ได้จากการสังเกตแล้วบันทึกสิ่งที่เราสนใจเอาไว้ ต้องใช้การสังเกตเป็นช่วง ๆ ของเวลาอย่างต่อเนื่องกัน ข้อมูลจะน่าเชื่อถือได้มากน้อยขึ้นอยู่กับความเข้าใจและความชำนาญของผู้สังเกต เช่น ข้อมูลเกี่ยวกับการใช้บริการต่าง ๆ เช่น บริการรถโดยสาร การบริการสหกรณ์ ความหนาแน่นของการใช้ถนนสายต่างๆ เป็นต้น วิธีนี้นิยมใช้ประกอบกับการเก็บข้อมูลวิธีอื่น ๆ

ง. การทดลอง (Experimental Design) เป็นการเก็บรวบรวมข้อมูลที่มีการทดลอง ซึ่งมักจะใช้เวลาในการทดลองนาน ๆ ทำซ้ำ ๆ

จ. การเก็บข้อมูลในทางอื่น ๆ เช่น ทางโทรศัพท์ (Telephone Survey) เป็นวิธีที่ง่าย เสียค่าใช้จ่ายน้อย ต้องเป็นการสัมภาษณ์อย่างสั้น ๆ ตอบได้ทันทีโดยไม่ต้องเสียเวลาค้นหาหลักฐาน ใช้ได้เฉพาะส่วนที่มีโทรศัพท์เท่านั้น ทางอินเทอร์เน็ต (Internet Survey) เป็นวิธีที่นิยมใช้กันมากในปัจจุบัน โดยการใช้แบบสอบถามแสดงบนหน้าเว็บ ซึ่งจะได้กลุ่มผู้ตอบแบบสอบถามที่มีความหลากหลายมากกว่าและประหยัดค่าใช้จ่ายเป็นอย่างมาก จึงมีความสะดวกรวดเร็ว

2) การเก็บรวบรวมข้อมูลจากข้อมูลทุติยภูมิ คือ ข้อมูลที่ต้องเก็บรวบรวมจากผู้ที่ให้ข้อมูลหรือแหล่งที่มาโดยตรง แต่ได้จากรายงานหรือข้อมูลที่มีผู้อื่นเก็บรวบรวมไว้แล้ว (Reporting System) โดยมีแหล่งที่มาของข้อมูลทุติยภูมิที่สำคัญมีอยู่ 2 แหล่ง คือ

2.1) รายงานต่าง ๆ ของหน่วยราชการและองค์การของรัฐบาล เช่น ทะเบียนประวัติบุคลากร ประวัติคนไข้ ทะเบียนนักเรียนนักศึกษา เป็นต้น

2.2) รายงานและบทความจากหนังสือ หรือรายงานจากหน่วยงานเอกชน ซึ่งจะมีการพิมพ์เผยแพร่เฉพาะในส่วนของข้อมูลที่เผยแพร่ได้ในรูปของรายงานต่าง ๆ

3.1.4 คุณสมบัติของข้อมูลที่ดี

ข้อมูลที่ดีจะต้องประกอบด้วยคุณสมบัติที่สำคัญ ๆ ดังนี้ คือ

1) ความถูกต้องแม่นยำ (accuracy) ข้อมูลที่ดีควรจะมีความถูกต้องแม่นยำสูง หรือถ้ามีความคลาดเคลื่อน (errors) ปนอยู่บ้าง ก็ควรที่จะสามารถควบคุมขนาดของความคลาดเคลื่อนที่ปนมาให้มีความคลาดเคลื่อน น้อยที่สุด

2) ความทันเวลา (timeliness) เป็นข้อมูลที่ทันสมัย (up to date) และทันต่อความต้องการของ ผู้ใช้ ถ้าผลิตข้อมูลออกมาช้า ก็ไม่มีคุณค่าถึงแม้จะเป็นข้อมูลที่ถูกต้องแม่นยำก็ตาม

3) ความสมบูรณ์ครบถ้วน (completeness) ข้อมูลที่เก็บรวบรวมมาต้องเป็นข้อมูลที่ให้ข้อเท็จจริง (facts) หรือข่าวสาร (information) ที่ครบถ้วนทุกด้านทุกประการ มิใช่ขาดส่วนหนึ่งส่วนใดไปทำให้นำไปใช้การไม่ได้

4) ความกะทัดรัด (conciseness) ข้อมูลที่ได้รับส่วนใหญ่จะกระจัดกระจาย ควรจัดข้อมูลให้อยู่ใน รูปแบบที่กะทัดรัดไม่เยิ่นเย้อ สะดวกต่อการใช้และค้นหา ผู้ใช้มีความเข้าใจได้ทันที

5) ความตรงกับความต้องการของผู้ใช้ (relevance) ข้อมูลที่จัดทำขึ้นมาควรเป็นข้อมูลที่ผู้ใช้ ข้อมูลต้องการใช้ และจำเป็นต้องรู้/ทราบ หรือเป็นประโยชน์ต่อการจัดทำแผน กำหนดนโยบายหรือตัดสินปัญหาในเรื่องนั้นๆ ไม่ใช่เป็นข้อมูลที่จัดทำขึ้นมาอย่างมากมาย แต่ไม่มีใครต้องการใช้หรือไม่ตรงกับความต้องการของผู้ใช้ข้อมูล

6) ความต่อเนื่อง (continuity) การเก็บรวบรวมข้อมูล ควรอย่างยิ่งที่จะต้องดำเนินการอย่างสม่ำเสมอและต่อเนื่องในลักษณะของอนุกรมเวลา (time-series) เพื่อจะได้นำไปใช้ประโยชน์ในด้านการวิเคราะห์วิจัยหรือหาแนวโน้มในอนาคต

3.1.5 การนำเสนอข้อมูลสถิติ (Statistical Presentation)

การนำเสนอข้อมูลสถิติแบ่งออกเป็น 2 แบบใหญ่ ๆ คือ

1) การนำเสนอข้อมูลสถิติโดยปราศจากแบบแผน (Informal Presentation)

1.1) การนำเสนอข้อมูลสถิติเป็นบทความ

1.2) การนำเสนอข้อมูลสถิติเป็นบทความกึ่งตาราง

2) การนำเสนอข้อมูลสถิติโดยมีแบบแผน (Formal Presentation)

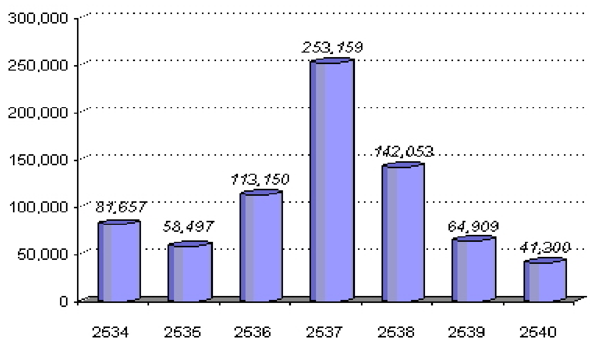
2.1) การเสนอข้อมูลสถิติด้วยตาราง (Tabular Presentation)

2.2) การเสนอข้อมูลสถิติด้วยกราฟและรูป (Graphic Presentation)

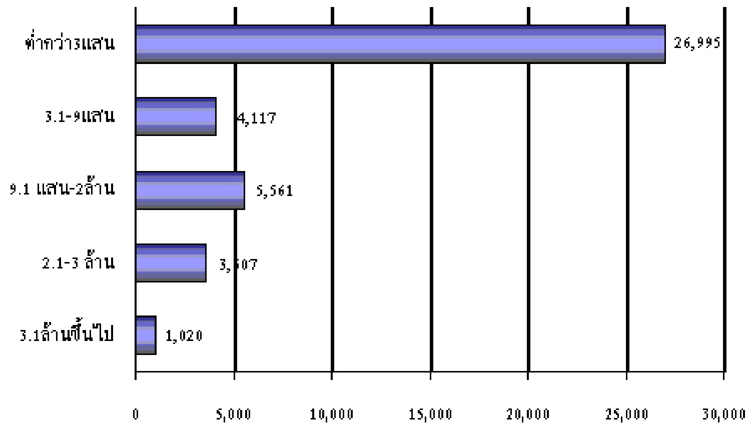
เทคนิคการนำเสนอข้อมูลสถิติด้วยกราฟและรูป

1) เมื่อต้องการเสนอข้อมูลสถิติโดยข้อมูลที่จะนำเสนอนั้นมีเพียงชุดเดียว

1.1) แผนภูมิแท่งเชิงเดียว (Simple Bar Chart)

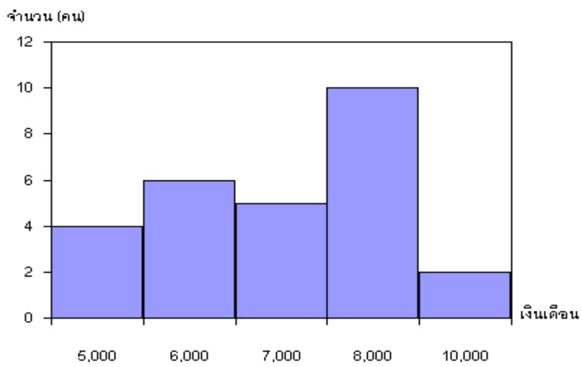


รูปที่ 1 ที่อยู่อาศัยในเขตกทม.และปริมณฑล



<u>รูปที่ 2</u> เปรียบเทียบจำนวนที่อยู่อาศัยที่เปิดขายตามระดับราคาต่าง ๆ ในเขตกรุงเทพฯ และปริมณฑลปี 2540

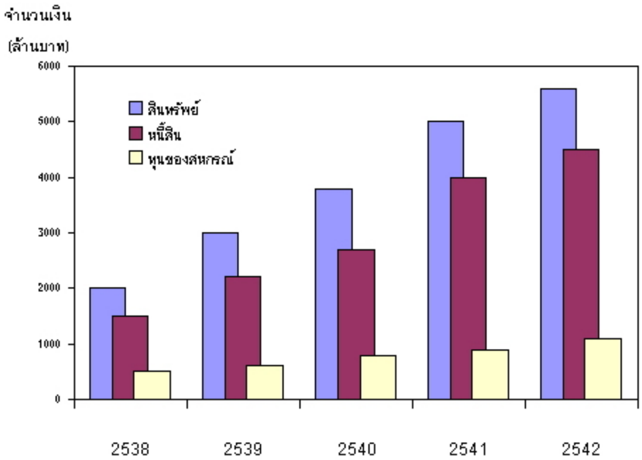
1.2) ฮิสโตแกรม (Histogram) ฮิสโตแกรมจะมีลักษณะเหมือนแผนภูมิแท่งทุกประการ ต่างกันเฉพาะตรงที่ฮิสโตแกรมนั้นแต่ละแท่งจะติดกัน



<u>รูปที่ 3</u> ฮิสโตแกรมแสดงเงินเดือนของพนักงานในบริษัทแห่งหนึ่ง

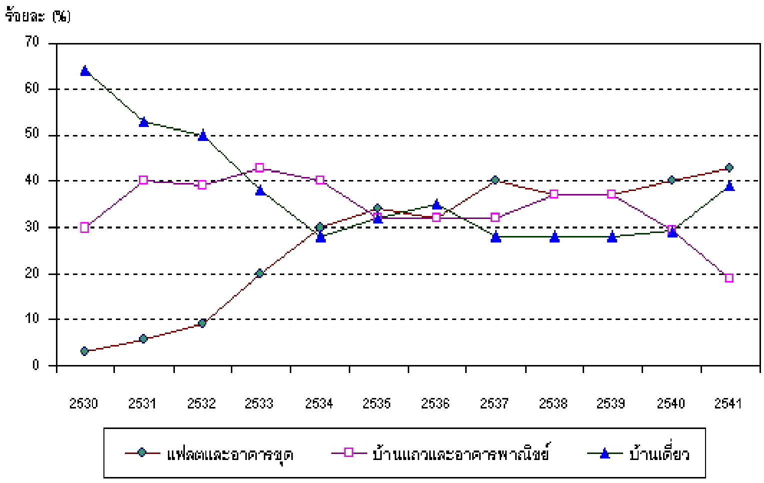
2) เมื่อต้องการนำเสนอข้อมูลสถิติในเชิงเปรียบเทียบ เมื่อต้องการนำเสนอในเชิงเปรียบเทียบข้อมูลตั้งแต่ 2 ชุดขึ้นไป ควรนำเสนอข้อมูลด้วยกราฟ ดังนี้

2.1) แผนภูมิแท่งเชิงซ้อน (Multiple Bar Chart) ข้อมูลสถิติที่จะนำเสนอด้วยแผนภูมิแท่งต้องเป็นข้อมูลประเภทเดียวกันหน่วยของตัวเลขเป็นหน่วยเดียวกันและควรใช้เปรียบเทียบข้อมูลเพียง 2 ชุดเท่านั้น ซึ่งอาจเป็นแผนภูมิในแนวตั้งหรือแนวนอน ก็ได้สิ่งที่สำคัญต้องมีกุญแจ (Key) อธิบายว่าแท่งใดหมายถึงข้อมูลชุดใดไว้ที่กรอบล่างของกราฟ



รูปที่ 4 แผนภูมิแท่งแสดงสินทรัพย์ หนี้สินทุนของสหกรณ์ออมทรัพย์มหาวิทยาลัยเกษตรศาสตร์

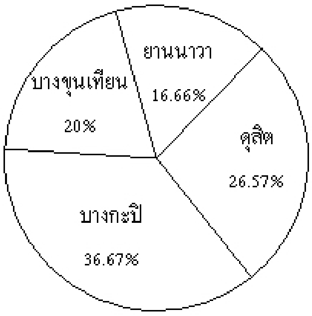
2.2) แผนภูมิเส้นหลายเส้น (Multiple Line Chart) ถ้าต้องการเปรียบเทียบข้อมูลสถิติหลายประเภทพร้อมๆกันควรจะนำเสนอด้วยแผนภูมิเส้นซึ่งสามารถนำเสนอข้อมูลที่มีหน่วยเหมือนกันหรือมีหน่วยต่างกันได้



รูปที่ 5 แผนภูมิเส้นแสดงการเปรียบเทียบสัดส่วนประเภทที่อยู่อาศัยสร้างเสร็จปี 2530 – 2541

3) เมื่อต้องการนำเสนอข้อมูลสถิติในเชิงส่วนประกอบ การนำเสนอข้อมูลในเชิงส่วนประกอบมีวิธีเสนอได้ 2 แบบ คือ

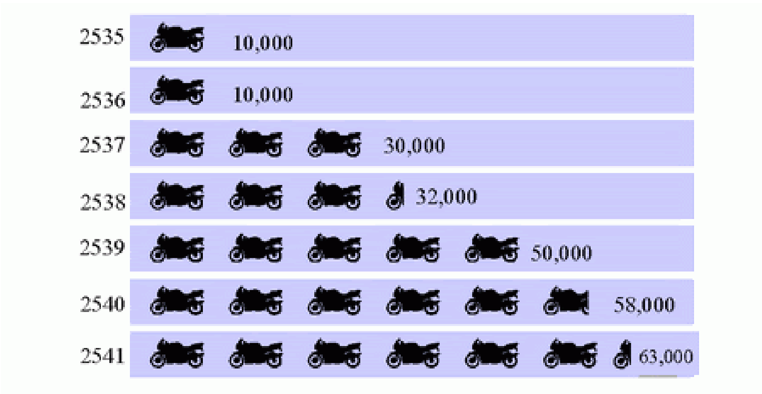
3.1) แผนภูมิวงกลม (Pie Chart)



รูปที่ 6 แผนภูมิวงกลมแสดงเขตที่พักอาศัยของลูกค้าที่มีเงินฝากธนาคารเกินกว่า 50,000,000 บาท

3.2) แผนภูมิแท่งเชิงประกอบ (Component Bar Chart) แผนภูมิแท่งเชิงประกอบเหมาะจะนำไปใช้เสนอข้อมูลเชิงเปรียบเทียบ วิธีทำคือเมื่อคิดองค์ประกอบต่างๆเป็นร้อยละของทั้งหมดแล้ว จะให้ความสูงของแผนภูมิแท่ง แทนองค์ประกอบทั้งหมดความสูงขององค์ประกอบแต่ละส่วนเป็นไปตามสัดส่วนขององค์ประกอบนั้น ๆ จะเรียงลำดับองค์ประกอบที่มีความสำคัญมากให้อยู่ข้างล่าง

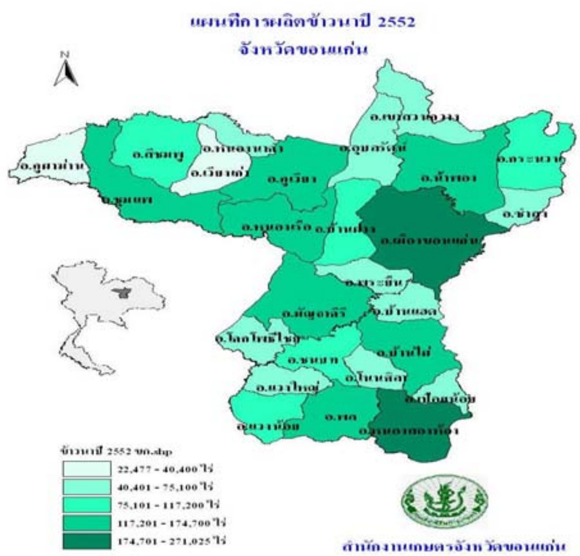
4) การนำเสนอข้อมูลสถิติด้วยแผนภูมิภาพ (Pictograph) การนำเสนอข้อมูลสถิติด้วยวิธีนี้จึงเป็นการเสนอสถิติที่เข้าใจง่ายที่สุด



รูปที่ 7 แผนภูมิภาพจำนวนรถจักรยานยนต์ที่เพิ่มขึ้นในแต่ละปี

(ที่มา: กองทะเบียนขนส่งทางบก)

5) การเสนอข้อมูลสถิติด้วยแผนที่สถิติ เป็นการนำเสนอข้อมูลที่เกี่ยวข้องกับสภาพภูมิศาสตร์หรือสถานที่ เช่น สถิติเกี่ยวกับความหนาแน่นของประชากรตามภูมิภาคต่าง ๆ สถิติจำนวนผู้ป่วยเป็นไข้ทรพิษที่ระบาดในประเทศบังกลาเทศ เป็นต้น



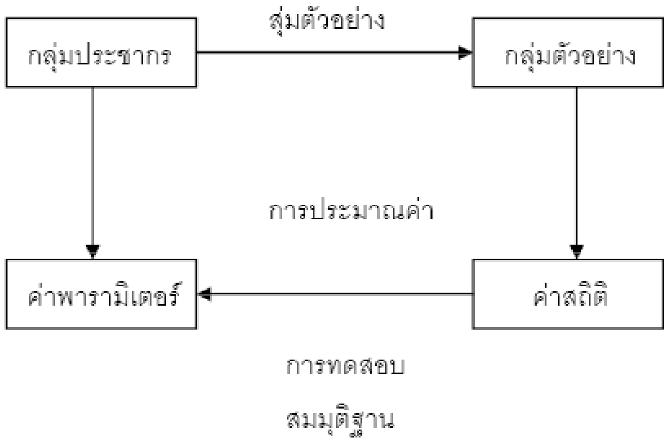
รูปที่ 8 แผนที่สถิติการผลิตข้าวนาปีจังหวัดขอนแก่น ในปี 2552
(ที่มา: สำนักงานเกษตรจังหวัดขอนแก่น)

3.1.6 ประชากรและกลุ่มตัวอย่าง

ประชากร (Population) หมายถึง หน่วยทุกหน่วย (ซึ่งอาจมีชีวิตหรือไม่มีชีวิตก็ได้) ที่เราสนใจ เช่น จำนวนคนไทยที่เป็นเพศชาย ประชากรคือคนไทยทุกคนที่เป็นเพศชาย จำนวนรถยนต์ในจังหวัดพิษณุโลก ประชากรคือ รถยนต์ทุกคันที่อยู่ในจังหวัดพิษณุโลก ฯลฯ

ตัวอย่าง (Sample) หมายถึง หน่วยย่อยของประชากรที่เราสนใจ เช่น จำนวนรถยนต์ที่วิ่งในจังหวัดพิษณุโลก ซึ่งตัวอย่างจะต้องเป็นรถยนต์ที่กำลังวิ่งอยู่ในจังหวัดพิษณุโลก ฯลฯ

ค่าต่าง ๆ ที่คำนวณได้จากประชากรจะเรียกว่าค่า พารามิเตอร์ (Parameter) ส่วนค่าต่างที่คำนวณได้จากกลุ่มตัวอย่างจะเรียกว่าค่าสถิติ (Statistics)



รูปที่ 9 ความสัมพันธ์ของกลุ่มประชากรกับกลุ่มตัวอย่าง

ตารางที่ 1 สัญลักษณ์ของค่าประชากรและตัวอย่าง

ค่าคำนวณ	ประชากร	ตัวอย่าง
1. ค่าเฉลี่ย	μ	$\overline{X}$
2. ส่วนเบี่ยงเบนมาตรฐาน	σ	S. , S.D. , s
3. ความแปรปรวน	σ^2	S^2, SD^2
4. จำนวนข้อมูล	N	n
5. สหสัมพันธ์	ρ	R , r
6. สัดส่วน	P	p

3.1.7 การใช้ประโยชน์ข้อมูลสถิติ

1) ประโยชน์ที่เกิดขึ้นตามประเภทของผู้ใช้ข้อมูลสถิติ ข้อมูลสถิติต่าง ๆ ที่ได้จัดเก็บรวบรวมมานั้น ไม่ใช่เป็นประโยชน์เฉพาะผู้ใช้ในวงการราชการเท่านั้น แต่ยังเป็นประโยชน์ต่อผู้ใช้ในวงการอื่น ๆ อีกมากมาย ซึ่งสามารถจัดกลุ่มผู้ใช้ข้อมูลสถิติออกเป็นกลุ่มใหญ่ได้ ดังนี้

1.1) กลุ่มผู้ใช้ในวงการราชการ ซึ่งได้แก่กระทรวง กรม และองค์การต่างๆ ของรัฐ ซึ่งรับผิดชอบ หรือเกี่ยวข้องกับการวางแผนงาน และการควบคุม ดูแล บริหารงานในด้านต่าง ๆ

1.2) กลุ่มผู้ใช้ในวงการเอกชน ซึ่งได้แก่ ธนาคารพาณิชย์ นักลงทุนทางธุรกิจอุตสาหกรรม ทั้งในประเทศและต่างประเทศ สถาบันวิจัยธุรกิจ การโฆษณา สื่อมวลชน

1.3) สถาบันการศึกษาและวิจัย ซึ่งได้แก่หน่วยงานที่ดำเนินการทางด้านการศึกษา และการวิจัยด้านต่าง ๆ

1.4) องค์การระหว่างประเทศ โดยเฉพาะองค์การระหว่างประเทศ หรือหน่วยชำนาญพิเศษของ สหประชาชาติที่เกี่ยวข้องกับการส่งเสริมสนับสนุนช่วยเหลือทางด้านความก้าวหน้าทางเศรษฐกิจสังคม การยกระดับการกินดีอยู่ดีของประชากรของโลก เช่น FAO ILO WHO IMF เป็นต้น

2) ประโยชน์ของข้อมูลสถิติ ข้อมูลสถิติมีความสำคัญและจำเป็นต่อการบริหารงานและพัฒนาประเทศ เป็นเครื่องมือ สำหรับ ผู้บริหาร ใช้เป็นแนวทางประกอบการตัดสินใจในการจัดทำแผนงานกำหนดนโยบายหรือแก้ไขปัญหาต่าง ๆ ดังนั้นประโยชน์ของข้อมูลสามารถจำแนกตามการใช้ที่สำคัญ ๆ ได้ดังนี้

2.1) ข้อมูลสถิติที่ใช้ในการบริหาร เป็นข้อมูลสถิติที่หน่วยงานต่าง ๆ ได้ผลิตขึ้นมาเพื่อใช้ในการบริหารและควบคุมการดำเนินงานประจำในสายงานต่าง ๆ หรือตรวจสอบผลการบริหารงานเช่น ข้อมูลสถิติจากระบบทะเบียนราษฎร สามารถนำไปใช้ในการกำหนดเขตการเลือกตั้ง การเกณฑ์ทหารหรือการเข้าเกณฑ์การศึกษาภาคบังคับ เป็นต้น

2.2) ข้อมูลสถิติที่ใช้ในการพัฒนา ข้อมูลสถิติมีบทบาทสำคัญยิ่งต่อการพัฒนาเศรษฐกิจและสังคมของประเทศ ประโยชน์ของข้อมูลที่ใช้ในการพัฒนานั้น สามารถแยกพิจารณาได้ 3 กรณี คือ

ก) การใช้ข้อมูลสถิติ สำหรับการจัดทำแผนพัฒนาเศรษฐกิจและสังคม โดยอาศัยข้อมูลสถิติเป็นพื้นฐานในการจัดทำแผน การกำหนดเป้าหมาย และทิศทางของการพัฒนา เช่น การกำหนด หรือ การวางนโยบายเกี่ยวกับการศึกษาภาคบังคับ การวางนโยบายเกี่ยวกับงบประมาณแผ่นดิน การวางนโยบายเกี่ยวกับ การค้าทั้งในประเทศและนอกประเทศ อัตราค่าจ้างแรงงาน การเก็บภาษีอากร เป็นต้น ในช่วงภาวะวิกฤติเศรษฐกิจเช่นในปัจจุบันนี้ ข้อมูลสถิติเป็นสิ่งที่มีความจำเป็นอย่างยิ่งต่อการกำหนดนโยบาย และแก้ไขปัญหาต่างๆ ของรัฐบาล โดยเฉพาะใช้เป็นเครื่องเตือนภัยล่วงหน้า เพื่อรัฐบาลจะได้กำหนดนโยบายหรือแผนงานต่างๆ ให้สอดคล้องกับภาวะเศรษฐกิจ

ข) การใช้ข้อมูลสถิติ สำหรับการติดตามความก้าวหน้าของแผนพัฒนา หรือโครงการต่างๆ ซึ่งรัฐบาลและหน่วยงานราชการได้จัดทำโครงการพัฒนาเป็นจำนวนมาก ซึ่งเป็นแผนระยะสั้นและระยะยาว ฉะนั้นจึงจำเป็นต้องมีข้อมูลเพื่อทำการตรวจสอบและติดตามความก้าวหน้าของโครงการดังกล่าวว่า ได้ผลมากน้อยเพียงใด เพื่อผู้บริหาร สามารถนำไปแก้ไขปรับปรุงแผนการดำเนินงานได้อย่างถูกต้องและทันเวลา หรือเพื่อนำผลที่ได้ไปใช้ประโยชน์ สำหรับการวางแผนโครงการอื่น ๆ ที่มีลักษณะคล้าย ๆ กัน

ค) การใช้ข้อมูลสถิติ สำหรับการประเมินผลแผนพัฒนาหรือโครงการพัฒนา เมื่อการดำเนินงานตามแผนงาน/โครงการพัฒนาได้เสร็จสิ้นลงแล้ว จำเป็นต้องมีการประเมินผลหรือวัดผลการพัฒนาว่าได้ผลตามวัตถุประสงค์หรือเป้าหมายที่ตั้งไว้เพียงไร จึงจำเป็นต้องใช้ข้อมูลสถิติเป็นเครื่องมือที่ชี้บอก ความสำเร็จหรือประสิทธิภาพและประสิทธิผลของการพัฒนา

3.2 หน่วยที่ 2 การวิเคราะห์ข้อมูลเบื้องต้น

3.2.1 ศัพท์ที่ใช้ในตารางการแจกแจงความถี่

1) อันตรภาคชั้น (Class Interval) หรือเรียกสั้น ๆ ว่า ชั้น หมายถึง ช่วงของคะแนนในแต่ละพวกที่แบ่ง

อันตรภาคชั้นต่ำสุด หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่าต่ำสุดอยู่

อันตรภาคชั้นสูงสุด หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่าสูงสุดอยู่

อันตรภาคชั้นต่ำกว่า หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่าน้อยกว่า

อันตรภาคชั้นสูงกว่า หมายถึง อันตรภาคชั้นของข้อมูลที่มีค่ามากกว่า

2) ความถี่ (Frequency: f) หมายถึง จำนวนข้อมูลที่มีอยู่ในแต่ละอันตรภาคชั้น

3) ตารางแจกแจงความถี่ (Frequency Distribution Table) หมายถึง ตารางที่เขียนเรียงลำดับข้อมูล และแสดงให้เห็นว่าแต่ละข้อมูล หรือกลุ่มข้อมูลมีความถี่เท่าใด

4) ขอบล่าง (Lower Boundary) ของอันตรภาคชั้น หมายถึง ค่ากึ่งกลางระหว่างค่าที่น้อยที่สุดของอันตรภาคชั้นนั้นกับค่าที่มากที่สุดของอันตรภาคชั้นที่ต่ำกว่าหนึ่งชั้น หรือขอบล่างเท่ากับค่าที่น้อยที่สุดของอันตรภาคชั้นบวกกับค่าที่มากที่สุดของอันตรภาคชั้นที่ต่ำกว่าหนึ่งชั้นแล้วหารด้วย 2

5) ขอบบน (Upper Boundary) ของอันตรภาคชั้น หมายถึง ค่ากึ่งกลางระหว่างค่าที่มากที่สุดของอันตรภาคชั้นนั้นกับค่าที่น้อยที่สุดของอันตรภาคชั้นที่สูงกว่าหนึ่งชั้น หรือขอบบนเท่ากับค่าที่มากที่สุดของอันตรภาคชั้นนั้นบวกกับค่าที่น้อยที่สุดของอันตรภาคชั้นที่สูงกว่าหนึ่งชั้น แล้วหารด้วย 2 จะเห็นว่าขอบบนของอันตรภาคชั้นหนึ่งย่อมเท่ากับขอบล่างของอันตรภาคชั้นที่สูงกว่าหนึ่งชั้นเสมอ

6) ความกว้างของอันตรภาคชั้น (Interval: i) หมายถึง ผลต่างระหว่างขอบบนและขอบล่างของอันตรภาคชั้นนั้น

7) ขีดจำกัดชั้นที่แท้จริง (Exacts Limit) หมายถึง ค่าในแต่ละชั้นที่มีค่าต่อเนื่องกัน

8) จุดกึ่งกลางชั้น (Middle Point) ของอันตรภาคชั้น หมายถึง ค่ากึ่งกลางระหว่างขอบล่างและขอบบนของอันตรภาคชั้นนั้น นิยมใช้สัญลักษณ์ X

3.2.2 การแจกแจงความถี่ของข้อมูล (Frequency Distribution)

การแจกแจงความถี่ของข้อมูล เป็นวิธีการทางสถิติอย่างหนึ่งที่ใช้ในการจัดข้อมูลที่มีอยู่หรือที่เก็บรวบรวมมาได้ให้อยู่เป็นกลุ่มๆ เพื่อสะดวกในการวิเคราะห์ข้อมูลเหล่านั้น

การแจกแจงความถี่ จัดเป็น 2 ลักษณะ ดังนี้

1) การแจกแจงความถี่แบบไม่จัดเป็นอันตรภาคชั้นหรือแบบไม่จัดกลุ่ม (Ungrouped Data) ใช้กับข้อมูลที่มีค่าสูงสุดและต่ำสุดของข้อมูลไม่แตกต่างกันมากนัก หรือข้อมูลที่มีค่าของจำนวนที่ต่างกันมีไม่มาก

2) การแจกแจงความถี่แบบจัดเป็นอันตรภาคชั้นหรือแบบจัดกลุ่ม (Grouped Data) ใช้กับข้อมูลที่มีค่าสูงสุดและต่ำสุดของข้อมูลแตกต่างกันมาก หรือการแจกแจงไม่สะดวกที่จะใช้ค่าสังเกตทุกๆค่า เพื่อความสะดวกจึงใช้วิธีแจกแจงความถี่ของค่าที่เป็นไปได้แทน โดยแบ่งค่าที่เป็นไปได้ออกเป็นช่วง หรืออันตรภาคชั้น (Interval)

การแจกแจงความถี่สะสม (Cumulative Frequency) ของค่าที่เป็นไปได้ค่าใดหรืออันตรภาคชั้นใด หมายถึง ผลรวมของความถี่ของค่านั้นหรืออันตรภาคชั้นนั้น กับความถี่ของค่าหรือของอันตรภาคชั้นที่มีช่วงคะแนนต่ำกว่าทั้งหมด หรือสูงกว่าทั้งหมดอย่างใดอย่างหนึ่ง (นิยมใช้ความถี่สะสมแบบต่ำกว่า)

3.2.3 การสร้างตารางแจกแจงความถี่

ควรทำเป็นขั้นตอนดังนี้

1) หาพิสัย (Range: R) โดย พิสัย = ค่าสูงสุด - ค่าต่ำสุด

2) กำหนดจำนวนอันตรภาคชั้นหรือจำนวนชั้น (k) โดยทั่วไปจะกำหนด 5 – 20 ชั้น แต่ต้องพิจารณาลักษณะการกระจายของข้อมูลด้วย หรือสามารถกำหนดได้ด้วยสูตร

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

3) ถ้ารู้จำนวนชั้นแล้ว เราต้องคำนวณหาความกว้างของแต่ละอันตรภาคชั้น โดยใช้หลักเกณฑ์ดังนี้

$$\text{ความกว้างของอันตรภาคชั้น (i)} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}}$$

แต่ถ้ารู้ความกว้างของอันตรภาคชั้น ก็สามารถหาจำนวนชั้นได้เช่นกัน

4) เขียนอันตรภาคชั้นเรียงตามลำดับ แล้วดูว่าค่าจากการสังเกตแต่ละค่าของข้อมูลอยู่ในอันตรภาคชั้นใด ก็ให้ขีด “ | ” ลงในอันตรภาคชั้นไปเรื่อย ๆ จนครบทุกค่าจากการสังเกตของข้อมูล

5) นับจำนวนขีดในแต่ละอันตรภาคชั้นและสรุปออกมาเป็นจำนวน ซึ่งจำนวนดังกล่าวคือความถี่ (f)

6) คำนวณขีดจำกัดชั้นที่แท้จริงจากสูตร

$$\text{ขีดจำกัดชั้นที่แท้จริง} = \frac{\text{ขีดบน+ขีดล่างของชั้นถัดไป}}{2}$$

ตารางที่ 2 ตัวอย่างตารางแจกแจงความถี่

ค่าจ้าง	ขีดจำกัดชั้นที่แท้จริง	จุดกึ่งกลาง	ความถี่	ความถี่สะสม
140-144	139.5-144.5	142	4	4
145-149	144.5-149.5	147	10	14
150-154	149.5-154.5	152	13	27
155-159	154.5-159.5	157	12	39
160-164	159.5-164.5	162	8	47
165-169	164.5-169.5	167	10	57
170-174	169.5-174.5	172	3	60

3.2.4 การวัดค่ากลางของข้อมูล

การหาค่ากลางของข้อมูลที่เป็นตัวแทนของข้อมูลทั้งหมดเพื่อความสะดวกในการสรุปเรื่องราวเกี่ยวกับข้อมูลนั้น ๆ จะช่วยทำให้เกิดการวิเคราะห์ข้อมูลถูกต้องดีขึ้น การหาค่ากลางของข้อมูลมีวิธีหาหลายวิธี แต่ละวิธีมีข้อดีและข้อเสีย และมีความเหมาะสมในการนำไปใช้ไม่เหมือนกัน ขึ้นอยู่กับลักษณะข้อมูลและวัตถุประสงค์ของผู้ใช้ข้อมูลนั้น ๆ

ค่ากลางของข้อมูลที่สำคัญ มี 6 ชนิด คือ

1. ค่าเฉลี่ยหรือมัชฌิมเลขคณิต (Arithmetic mean)
2. มัธยฐาน (Median)

3. ฐานนิยม (Mode)

4. ตัวกลางเรขาคณิต (Geometric mean)

5. ตัวกลางฮาโมนิค (Harmonic mean)

6. ตัวกึ่งกลางพิสัย (Mid-range)

แต่ในวิชานี้จะศึกษาค่ากลางของข้อมูลที่สำคัญเพียง 3 ชนิด คือ ค่าเฉลี่ยเลขคณิต (Arithmetic mean) มัธยฐาน (Median) และฐานนิยม (Mode)

1) ค่าเฉลี่ยหรือมัชฌิมเลขคณิต (Arithmetic me $\overline{X}$, Average:) จัดว่าเป็นค่าที่มีความสำคัญมากในวิชาสถิติ เพราะค่าเฉลี่ยเลขคณิตเป็นค่ากลางหรือเป็นตัวแทนของข้อมูลที่ดีที่สุด เพราะ

- เป็นค่าที่ไม่เอนเอียง
- เป็นค่าที่มีความคงเส้นคงวา
- เป็นค่าที่มีความแปรปรวนต่ำที่สุด
- เป็นค่าที่มีประสิทธิภาพสูงสุด

แต่ค่าเฉลี่ยเลขคณิตก็มีข้อจำกัดในการใช้ เช่น ถ้าข้อมูลมีการกระจายมาก หรือข้อมูลบางตัวมีค่ามากหรือน้อยจนผิดปกติ หรือข้อมูลมีการเพิ่มขึ้นเป็นเท่าตัว ค่าเฉลี่ยเลขคณิตจะไม่สามารถเป็นค่ากลางหรือเป็นตัวแทนที่ดีของข้อมูลได้

1.1) การหาค่าเฉลี่ยเลขคณิตของข้อมูลที่ไม่แจกแจงความถี่ให้ $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูล N ค่า

$$\overline{x} = \frac{\sum_{i=1}^{N} X_i}{N} \text{ หรือ } \overline{x} = \frac{\sum X}{N}$$

1.2) ค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่ ถ้า $f_1, f_2, f_3, \ldots, f_k$ เป็นความถี่ของค่าจากการสังเกต $x_1, x_2, x_3, \ldots, x_k$

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots + f_kx_k}{f_1 + f_2 + f_3 + \ldots + f_k}$$

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{\sum fx}{\sum f}$$

<u>ตารางที่ 3</u> ตัวอย่างการหาค่าเฉลี่ยเลขคณิตของข้อมูลที่แจกแจงความถี่

ค่าจ้าง	จุดกึ่งกลาง(X_i)	ความถี่(f_i)	f_iX_i
140-144	142	4	568
145-149	147	10	1,470
150-154	152	13	1,976
155-159	157	12	1,884
160-164	162	8	1,296
165-169	167	10	1,670
170-174	172	3	516
ผลรวม		60	9,380

จากตารางค่าเฉลี่ยเลขคณิตหรือค่าจ้างเฉลี่ย = 9,380/60 = 156.33 บาท

2) มัธยฐาน (Median: Med) เป็นค่ากลางของข้อมูลที่ได้จากการพิจารณาตำแหน่งของข้อมูลที่อยู่ตรงกลาง โดยที่ข้อมูลต้องทำการเรียงลำดับตามปริมาณจากมากไปน้อย หรือจากน้อยไปมากก็ได้ และค่ามัธยฐานยังสามารถใช้เป็นตัวแทนของข้อมูลได้เป็นอย่างดี ในกรณีที่ข้อมูลมีการกระจายที่ผิดปกติ ซึ่งอาจเกิดจากการที่มีข้อมูลบางตัวมีค่ามากหรือน้อยจนผิดปกติ

2.1) การหาค่ามัธยฐานของข้อมูลที่ไม่แจกแจงความถี่ โดยมีขั้นตอน ดังนี้

2.1.1) เรียงข้อมูลที่มีอยู่ทั้งหมดจากน้อยไปมาก หรือมากไปน้อยก็ได้

2.1.2) ตำแหน่งมัธยฐาน คือ ตำแหน่งกึ่งกลางข้อมูล

2.1.3) ค่ามัธยฐาน คือ ข้อมูลที่อยู่ในตำแหน่งมัธยฐาน

- ถ้าข้อมูลเป็นเลขคี่ คือ ค่าข้อมูลที่อยู่ตรงกลางจากสูตร $\frac{N+1}{2}$

- ถ้าข้อมูลเป็นเลขคู่ คือ ค่าเฉลี่ยของข้อมูลสองค่าที่อยู่ตรงกลางจากสูตร

กับ $+\frac{N}{2}$ $\frac{N}{2}$

เมื่อ N คือ จำนวนข้อมูลทั้งหมด

ตัวอย่าง จงหาค่ามัธยฐานของข้อมูล 2 ชุดมีดังนี้

ชุดที่ 1: 8 5 2 6 4

ชุดที่ 2: 5 9 3 4 7 3

วิธีทำ 1. เรียงข้อมูลใหม่

ชุดที่ 1: 2 4 5 6 8

ชุดที่ 2: 3 3 4 5 7 9

2. ค่ามัธยฐานชุดที่ 1 อยู่ในตำแหน่งที่ = (5+1)/2 = 3 นั่นคือ 5

3. ค่ามัธยฐานชุดที่ 2 อยู่ในตำแหน่งที่ 1 = 6/2 = 3 และ ตำแหน่งที่ 2 = (6/2)+1 = 3+1 = 4 นั่นคือค่าเฉลี่ยข้อมูล 4 กับ 5 = (4+5)/2 = 4.5 ซึ่งจะไม่ปรากฏในข้อมูล

2.2) การหาค่ามัธยฐานของข้อมูลที่มีการแจกแจงความถี่ โดยมีขั้นตอน ดังนี้

2.2.1) สร้างตารางความถี่สะสม

2.2.2) หาตำแหน่งของมัธยฐาน คือ N/2 เมื่อ N เป็นจำนวนของข้อมูลทั้งหมด

2.2.3) ถ้า N/2 เท่ากับความถี่สะสมของอันตรภาคชั้นใด อันตรภาคชั้นนั้นเป็นชั้นมัธยฐาน และมีมัธยฐานเท่ากับขอบบนของอันตรภาคชั้นนั้น ถ้า N/2 ไม่เท่าความถี่สะสมของอันตรภาคชั้นใดเลย อันตรภาคชั้นแรกที่มีความถี่สะสมมากกว่า N/2 เป็นชั้นของมัธยฐาน และหามัธยฐานได้จากการเทียบบัญญัติไตรยางค์ หรือใช้สูตร ดังนี้

$$Med = L + i\left(\frac{\frac{N}{2} - \sum f_L}{f_M}\right)$$

เมื่อ L คือ ขอบล่างที่แท้จริงของคะแนนที่มีมัธยฐาน

i คือ ความกว้างของอันตรภาคชั้นที่มีมัธยฐาน

$\sum f_L$ คือ ความถี่สะสมของชั้นคะแนนต่ำสุดถึงชั้นก่อนที่มีมัธยฐาน

f_M คือ ความถี่ของคะแนนในชั้นที่มีมัธยฐาน

N คือ จำนวนความถี่หรือข้อมูลทั้งหมด

3) ฐานนิยม (Mode: Mod) เป็นค่ากลางซึ่งจะนำมาใช้ในกรณีที่ข้อมูลมีการซ้ำกันมาก ๆ จนผิดปกติ ซึ่งค่าฐานนิยมจะเป็นค่ากลางหรือตัวแทนของข้อมูลที่สามารถอธิบายลักษณะที่เกิดขึ้นได้ดีกว่าค่าเฉลี่ยเลขคณิตและค่ามัธยฐาน นอกจากนี้ค่าฐานนิยมยังมีข้อพิเศษมากกว่าค่าเฉลี่ยและมัธยฐาน ตรงที่สามารถใช้ได้กับข้อมูลที่เป็นข้อมูลเชิงคุณภาพ (qualitative) และข้อมูลเชิงปริมาณ (quantitative) และค่าฐานนิยมยังสามารถมีค่าได้มากกว่า 1 ค่าอีกด้วย

3.1) การหาค่าฐานนิยมของข้อมูลที่ไม่แจกแจงความถี่ โดยการนับจำนวนข้อมูล ซึ่งข้อมูลชุดใดมีจำนวนซ้ำกันมากที่สุดก็จะเป็นค่าฐานนิยม ซึ่งค่าฐานนิยมอาจจะไม่มีหรือมีมากกว่า 1 ค่าก็ได้

3.2) การหาค่าฐานนิยมของข้อมูลที่มีการแจกแจงความถี่ โดยใช้สูตร ดังนี้

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

เมื่อ L คือ ขอบล่างที่แท้จริงของคะแนนที่มีมัธยฐาน

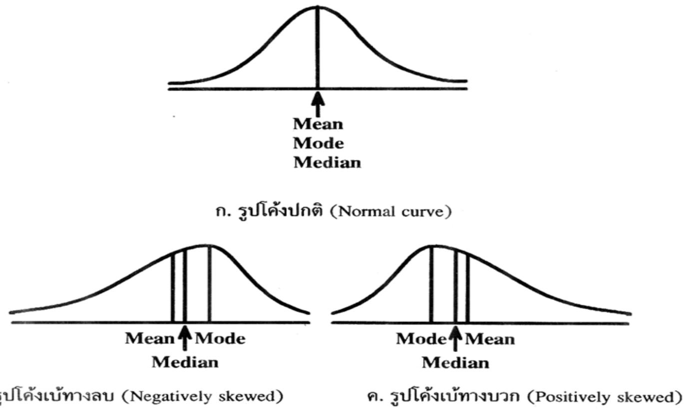
i คือ ความกว้างของอันตรภาคชั้นที่มีมัธยฐาน

d_1 คือ ความถี่ของชั้นที่มีฐานนิยมอยู่ความถี่ของชั้นติดกันที่มีช่วงคะแนนต่ำกว่า

d_2 คือ ความถี่ของชั้นที่มีฐานนิยมอยู่ความถี่ของชั้นติดกันที่มีช่วงคะแนนสูงกว่า

โดยปกติจะใช้ค่าเฉลี่ยเลขคณิต (Arithmetic mean) เป็นค่ากลางที่สื่อความเข้าใจง่าย และมีคุณสมบัติที่จะโยงไปสู่การอนุมานสถิติต่อไปได้ดี

ถ้าการแจกแจงความถี่เป็นแบบสมมาตร จุดค่าเฉลี่ยเลขคณิต มัธยฐานและฐานนิยมจะอยู่ที่จุดเดียวกัน แต่ถ้าการแจกแจงเป็นแบบเบ้ จุดทั้ง 3 จุดจะไม่ใช่จุดเดียวกัน สำหรับการแจกแจงความถี่แบบเบ้บวก เราจะพบว่าค่าค่าเฉลี่ยเลขคณิตจะมากกว่ามัธยฐานและมากกว่าฐานนิยม ถ้าการแจกแจงเป็นเบ้ลบค่าที่ได้จะตรงกันข้าม



<u>รูปที่ 10</u> ตำแหน่งของค่าเฉลี่ยเลขคณิต มัธยฐานและฐานนิยมในการแจกแจงแบบต่าง ๆ
(ชูศรี วงศ์รัตนะ, 2550: 42)

3.2.5 การเลือกใช้ค่ากลางชนิดต่าง ๆ

ปัจจัยที่ควรพิจารณาในการเลือกค่ากลางที่เป็นตัวแทนของข้อมูลมีดังนี้

1) วัตถุประสงค์ของการนำไปใช้

ถ้าต้องการทราบข้อมูลที่มีลักษณะเด่น มีความถี่สูงสุดใช้ฐานนิยม ถ้าต้องการทราบตำแหน่งของข้อมูลที่เป็นค่ากลางใช้มัธยฐาน และถ้าต้องการตัวแทนของข้อมูลเพื่อนำไปใช้ในสถิติขึ้นสูงใช้ค่าเฉลี่ยเลขคณิต

2) ลักษณะของข้อมูลที่นำมาหาค่ากลาง

ถ้าข้อมูลเป็นข้อมูลเชิงคุณภาพ คือ วัดในมาตรานามบัญญัติ ใช้ฐานนิยมในการวัดค่ากลาง แต่ถ้า ข้อมูลเป็นข้อมูลเชิงปริมาณ ซึ่งวัดในมาตราเรียงลำดับ ใช้มัธยฐานเป็นค่ากลาง และถ้าข้อมูลวัดในมาตราช่วงหรือมาตราอัตราส่วน ใช้ค่าเฉลี่ยเลขคณิตเป็นค่ากลาง

3) การกระจายของข้อมูลที่นำมาหาค่ากลาง

ข้อมูลที่ต้องการหาค่ากลางมีการกระจายมากหรือน้อย ถ้าข้อมูลมีการกระจายมาก ควรใช้ค่ามัธยฐานเป็นค่ากลาง แต่ถ้าข้อมูลมีการกระจายน้อย ควรใช้ค่าค่าเฉลี่ยเลขคณิต หรือมัธยฐานเป็นค่ากลาง

4) ความสะดวก รวดเร็วในการหาค่ากลาง

ถ้าต้องการหาค่ากลางอย่างคร่าว ๆ อาจใช้ค่าฐานนิยม แต่ถ้าต้องการใช้ค่ากลางที่ละเอียดขึ้นควรใช้ค่าค่าเฉลี่ยเลขคณิต หรือมัธยฐาน

5) ความถูกต้องเชื่อถือได้ของค่ากลาง

ถ้าต้องการค่ากลางที่ถูกต้องเชื่อถือได้ ต้องพิจารณาวัตถุประสงค์ของการนำไปใช้และการกระจายของข้อมูลประกอบกัน

3.2.6 ตำแหน่งเปอร์เซ็นไทล์ เดไซล์ และควอไทล์

ในการเปรียบเทียบตำแหน่งต่าง ๆ ของข้อมูลคนละชุดนั้น ควรจะแปลงข้อมูลให้อยู่ในลักษณะเดียวกันก่อน

1) ตำแหน่งเปอร์เซ็นไทล์ (Percentile: P) หมายถึง ตำแหน่งที่บอกให้ทราบว่ามีข้อมูลอยู่กี่ส่วนจากร้อยส่วน

2) ตำแหน่งเดไซล์ (Decile: D) หมายถึง ตำแหน่งที่บอกให้ทราบว่ามีข้อมูลอยู่กี่ส่วนจากสิบส่วน

3) ตำแหน่งควอไทล์ (Quartile: Q) หมายถึง ตำแหน่งที่บอกให้ทราบว่ามีข้อมูลอยู่กี่ส่วนจากสี่ส่วน

การคำนวณหาตำแหน่งเปอร์เซ็นไทล์ เดไซล์ และควอไทล์ ของคะแนนที่กำหนด

1) สูตรการคำนวณหาตำแหน่งเปอร์เซ็นไทล์

$$P = \frac{100}{N}\left[F + \frac{f(X-L)}{i}\right]$$

2) สูตรการคำนวณหาตำแหน่งเดไซล์

$$D = \frac{10}{N}\left[F + \frac{f(X-L)}{i}\right]$$

3) สูตรการคำนวณหาตำแหน่งควอไทล์

$$Q = \frac{4}{N}\left[F + \frac{f(X-L)}{i}\right]$$

การคำนวณหาคะแนน ณ ตำแหน่งเปอร์เซ็นไทล์ เดไซล์ และควอไทล์ ที่กำหนด

1) สูตรการคำนวณหาคะแนน ณ ตำแหน่งเปอร์เซ็นไทล์ที่กำหนด

$$X = L + \left[\frac{N(P/100) - F}{f}\right]i$$

2) สูตรการคำนวณหาคะแนน ณ ตำแหน่งเดไซล์ ที่กำหนด

$$X = L + \left[\frac{N(D/10) - F}{f}\right]i$$

3) สูตรการคำนวณหาคะแนน ณ ตำแหน่งควอไทล์ ที่กำหนด

$$X = L + \left[\frac{N(Q/4) - F}{f} \right] i$$

F = ความถี่สะสมของคะแนนต่ำสุดถึง L

f = ความถี่ของคะแนนในชั้นที่ X อยู่

L = ขีดจำกัดล่างที่แท้จริงที่คะแนน X อยู่

i = อันตรภาคชั้น

N = จำนวนข้อมูลทั้งหมด

3.2.7 การวัดการกระจายของข้อมูล (Measures of Variation)

เป็นการวัดว่าข้อมูลแต่ละค่าในข้อมูลชุดใดชุดหนึ่งมีความแตกต่างกันมากน้อยเพียงใด การวัดการกระจายของข้อมูลชุดหนึ่ง ๆ สามารถทำได้หลายวิธี ได้แก่ พิสัย (Range) ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation or Average Deviation) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

1) พิสัย (Range) หมายถึง ความแตกต่างระหว่างข้อมูลที่มีค่ามากที่สุดและข้อมูลที่มีค่าน้อยที่สุด สามารถหาค่าพิสัย ได้ดังนี้

1.1) การหาพิสัยของข้อมูลที่ไม่ได้แจกแจงความถี่

พิสัย = ค่าสูงสุด - ค่าต่ำสุด

ตัวอย่าง ข้อมูลชุดหนึ่งมีดังนี้ 31 34 37 40 46 50 52 55 จงหาพิสัยของข้อมูลชุดนี้

พิสัย = ค่าสูงสุด - ค่าต่ำสุด

= 55 - 31

= 24

ดังนั้น พิสัยของข้อมูลชุดนี้คือ 24

1.2) การหาพิสัยของข้อมูลที่มีการแจกแจงความถี่

พิสัย = ขีดจำกัดบนที่แท้จริงของอันตรภาคชั้นที่มีค่ามากที่สุด - ขีดจำกัดล่างที่แท้จริงของอันตรภาคชั้นที่มีค่าน้อยที่สุด

ตัวอย่าง จงหาพิสัยของข้อมูลต่อไปนี้

ข้อมูล	จำนวน
31 - 35	5
36 - 40	8
41 - 45	12
46 - 50	9
51 - 55	6

พิสัย = 55.5 - 30.5 = 25

ดังนั้น พิสัยของข้อมูลชุดนี้คือ 25

ข้อสังเกต การวัดการกระจายด้วยพิสัยเป็นการวัดการกระจายอย่างคร่าว ๆ ใช้ในกรณีที่ต้องการความรวดเร็ว การหาพิสัยใช้เพียงค่าสูงสุดกับค่าต่ำสุดของข้อมูลเพียง 2 ค่าเท่านั้น บางครั้งจึงอาจทำให้เราเข้าใจลักษณะของข้อมูลผิดไป

2) ส่วนเบี่ยงเบนควอไทล์ (Quartile Deviation: Q.D.) หมายถึง ค่าครึ่งหนึ่งของผลต่างควอไทล์ที่ 3 กับควอไทล์ที่ 1 จะใช้วัดการกระจายของข้อมูล ที่วัดค่ากลางด้วยค่ามัธยฐาน

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2}$$

3) ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation: M.D.) หมายถึง ค่าเฉลี่ยของค่าเบี่ยงเบนของข้อมูลแต่ละตัวที่เบี่ยงเบนไปจากค่าเฉลี่ยของข้อมูลชุดนั้น โดยไม่คำนึงถึงทิศทางหรือเครื่องหมายของการเบี่ยงเบน ซึ่งมักใช้ควบคู่กับค่าเฉลี่ย

$$\text{M.D.} = \frac{\sum \left| x - \bar{x} \right|}{n}$$ กรณีข้อมูลที่ไม่ได้มีการแจกแจงความถี่

$$\text{M.D.} = \frac{\sum f \left| x - \bar{x} \right|}{n}$$ กรณีข้อมูลที่มีการแจกแจงความถี่

4) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) และความแปรปรวน (Variance)

4.1) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) เป็นการวัดว่าโดยเฉลี่ยแล้วข้อมูลแต่ละค่าแตกต่างไปจากค่ามัชฌิมเลขคณิตเพียงใด การวัดการกระจายโดยส่วนเบี่ยงเบนมาตรฐานนิยมใช้กันมากที่สุด

กรณีข้อมูลที่ไม่ได้มีการแจกแจงความถี่

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

กรณีข้อมูลที่มีการแจกแจงความถี่

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

4.2) ความแปรปรวน (Variance) เป็นค่าส่วนเบี่ยงเบนมาตรฐานยกกำลังสอง

กรณีข้อมูลที่ไม่ได้มีการแจกแจงความถี่

$$S^2 = \frac{\sum (x-\bar{x})^2}{n-1} \qquad , S^2 = \frac{\sum x^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$

กรณีข้อมูลที่มีการแจกแจงความถี่

$$S^2 = \frac{\sum f(x-\bar{x})^2}{n-1} \qquad , S^2 = \frac{\sum fx^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$

การหาค่าการกระจายของข้อมูล ได้แก่ พิสัย และส่วนเบี่ยงเบนมาตรฐานของข้อมูล ดังได้กล่าวมาแล้ว เป็นการหาการวัดการกระจายของข้อมูลใดชุดหนึ่ง ซึ่งเรียกได้ว่า การวัดการกระจายสัมบูรณ์ ในกรณีที่มีข้อมูลหลายชุด และต้องการเปรียบเทียบว่าข้อมูลชุดใดกระจายมากกว่ากันก็สามารถเปรียบเทียบได้ โดยการวัดการกระจายสัมพัทธ์ ซึ่งคำนวณได้จากสัมประสิทธิ์ของพิสัย และสัมประสิทธิ์ของความแปรผัน

5) สัมประสิทธิ์ของพิสัย (Coefficient of Range) ใช้ในกรณีที่ต้องการเปรียบเทียบการกระจายของข้อมูลตั้งแต่ 2 ชุดขึ้นไป ซึ่งวัดค่าการกระจายของข้อมูลแต่ละชุดด้วยค่าพิสัย การเปรียบเทียบพิจารณาจากสัมประสิทธิ์ของพิสัย ถ้าสัมประสิทธิ์ของพิสัยของข้อมูลชุดใดมีค่าน้อยกว่า แสดงว่าข้อมูลชุดนั้นมีการกระจายน้อยกว่าข้อมูลอีกชุดหนึ่ง สัมประสิทธิ์ของพิสัยคำนวณได้จากสูตร

$$\text{สัมประสิทธิ์ของพิสัย} = \frac{\text{ค่าสูงสุด - ค่าต่ำสุด}}{\text{ค่าสูงสุด + ค่าต่ำสุด}}$$

6) สัมประสิทธิ์ของความแปรผัน (Coefficient of Variation: C.V.) ในกรณีที่ข้อมูลวัดการกระจายด้วยค่าส่วนเบี่ยงเบนมาตรฐาน และต้องการเปรียบเทียบการกระจายของข้อมูล 2 ชุดขึ้นไป สามารถเปรียบเทียบได้โดยพิจารณาจากสัมประสิทธิ์ของความแปรผัน ถ้าข้อมูลชุดใดมีค่าสัมประสิทธิ์ของความแปรผันน้อยกว่า แสดงว่าข้อมูลชุดนั้นมีการกระจายน้อยกว่าข้อมูลอีกชุดหนึ่ง สัมประสิทธิ์ของความแปรผันคำนวณได้จากสูตร

$$\text{สัมประสิทธิ์ของความแปรผัน (C.V.)} = \frac{\text{ส่วนเบี่ยงเบนมาตรฐาน}}{\text{ค่าเฉลี่ยเลขคณิต}}$$

ส่วนใหญ่จะอธิบายเป็นร้อยละ โดยนำผลลัพธ์ที่ได้คูณด้วย 100

3.3 หน่วยที่ 3 ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น

3.3.1 การทดลองเชิงสุ่ม แซมเปิลสเปซ แซมเปิลพ้อย และเหตุการณ์

1) การทดลองเชิงสุ่ม (Random Experiment) หมายถึง การทดลองใด ๆ ที่ไม่สามารถทำนายผลลัพธ์ของการทดลองได้อย่างถูกต้องแน่นอน เนื่องจากผลลัพธ์อาจเกิดได้หลายอย่าง เช่น การโยนเหรียญ 1 เหรียญ 1 ครั้ง อาจจะออกหัวหรือก้อยก็ได้ หรือการโยนลูกเต๋า 1 ลูก 1 ครั้ง อาจจะออกแต้ม 1,2,3,4,5,6 ก็ได้ ฯลฯ

2) แซมเปิลสเปซ และแซมเปิลพ้อย (Sample Space and Sample Point) แซมเปิลสเปซ หมายถึง เซตของผลลัพธ์ที่อาจเป็นไปได้ทั้งหมดของการทดลองเชิงสุ่มเขียนแทนด้วยสัญลักษณ์ S และเรียกสมาชิกที่เป็นไปได้แต่ละตัวว่าแซมเปิลพ้อย เช่น

การทดลองเชิงสุ่ม: โยนเหรียญ 1 เหรียญ 1 ครั้ง

แซมเปิลสเปซ: S = {H, T}

แซมเปิลพ้อย: H, T

3) เหตุการณ์ (Event) หมายถึง เซตของผลลัพธ์ที่สนใจจากการทดลองเชิงสุ่ม เช่น การทดลองเชิงสุ่ม: โยนเหรียญ 1 เหรียญ 1 ครั้ง S = {H, T} จะมีเหตุการณ์ 2 เหตุการณ์ คือ

E_1: = {H}

E_2: = {T}

3.3.2 การนับจำนวนแซมเปิลพ้อย

1) หลักการเกี่ยวกับการนับ ในการทดลองใด ๆ ที่สามารถแบ่งการกระทำออกเป็น r ขั้นตอน โดยที่ขั้นตอนที่หนึ่งสามารถกระทำได้ n_1 วิธี ในแต่ละวิธีกระทำได้ในขั้นตอนที่สองอีก n_2 วิธี ในแต่ละวิธีกระทำได้ในขั้นตอนที่สามอีก n_3 วิธี เป็นเช่นนี้ไปเรื่อย ๆ จนถึงขั้นตอนที่ r ซึ่งสามารถกระทำได้ n_r วิธี

การคำนวณจำนวนวิธีทั้งหมดที่สามารถกระทำได้เท่ากับผลคูณของ n_1 n_2 n_3 n_r วิธี เช่น โยนเหรียญ 1 เหรียญ 3 ครั้ง จะได้ทั้งหมด = 2 x 2 x 2 = 8 วิธี

2) แฟคทอเรียล (n Factorial) คือ ผลคูณของเลขจำนวนเต็มบวกที่เรียงกันตามลำดับตั้งแต่ n ลงไปจนถึง 1 เขียนแทนด้วยสัญลักษณ์ n! โดยที่

$n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots\ldots\ldots \cdot 3 \cdot 2 \cdot 1$ และ $0! = 1$

เช่น $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$

3) การจัดลำดับ (Permutation) หมายถึง วิธีการเรียงลำดับสิ่งของจำนวนหนึ่ง โดยถือลำดับเป็นสำคัญ เช่น การเรียงอักษร 2 ตัว คือ A กับ B จะเรียงได้ 2 วิธี คือ AB กับ BA

ถ้าการจัดเรียงสิ่งของ n สิ่งที่แตกต่างกันทั้งหมด โดยจัดครั้งละ r สิ่ง ($r \leq n$) จะมีวิธีการจัดลำดับได้เท่ากับ $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots\ldots\ldots \cdot (n - r + 1)$ วิธี เขียนแทนด้วยสัญลักษณ์ ${}^nP_r$ โดยที่ ${}^nP_r = n!/(n-r)!$

เช่นอักษร 3 ตัว คือ A B C นำมาเรียง 2 ตัว จะได้กี่วิธี

$^{3}P_{2}$ = 3!/(3-2)! = 3! = 6 วิธี

4) การจัดหมู่ (Combination) หมายถึง วิธีการจัดเรียงสิ่งของ โดยไม่ถือลำดับเป็นสำคัญ บางครั้งจึงเรียกว่าการเลือก เขียนแทนด้วยสัญลักษณ์ $^{n}C_{r}$ โดยที่ $^{n}C_{r}$ หรือ $\binom{n}{r}$ = n!/(n-r)! r!

เช่น ในการจัดเรียงอักษร 3 ตัว คือ A B C นำมาเรียง 2 ตัว จะได้การจัดลำดับ 6 วิธี แต่จะได้การจัดหมู่เพียง 3 วิธี คือ หนึ่ง AB หรือ BA สอง CB หรือ BC สาม AC หรือ CA นั่นคือ

$^{3}C_{2}$ = 3!/(3-2)! 2! = 3 วิธี

5) การจัดหมู่ของหลายประเภท ในกรณีที่ของ n สิ่งที่แตกต่างกัน สามารถแบ่งเป็น k ประเภท แต่ละประเภทมีจำนวน $n_1, n_2, n_3, \ldots, n_k$ สิ่ง ถ้านำมาจัดหมู่ประเภทละ $r_1, r_2, r_3, \ldots, r_k$ สิ่ง ตามลำดับ ($r_i \leq n_i$) จะมีวิธีการจัดหมู่ได้ $\binom{n_1}{r_1}\binom{n_2}{r_2}\ldots\binom{n_k}{r_k}$ วิธี

3.3.3 ความน่าจะเป็นของเหตุการณ์ (Probability of an Event)

ถ้า S เป็นแซมเปิลสเปซของการทดลองเชิงสุ่ม ประกอบด้วยสมาชิก N ตัว สมาชิกแต่ละตัวมีโอกาสเกิดขึ้นเท่า ๆ กัน และ E เป็นเหตุการณ์ที่เกิดจากการทดลองนี้ ประกอบด้วยสมาชิก n ตัว จะได้ว่าความน่าจะเป็นของเหตุการณ์ E เขียนแทนด้วยสัญลักษณ์ P(E) จะมีค่าเท่ากับ n/N โดยมีคุณสมบัติ

1) ความน่าจะเป็นของเหตุการณ์ จะมีค่าตั้งแต่ 0 ถึง 1 เสมอ

2) P(S) = 1

3.3.4 ตัวแปรสุ่ม (Random Variable)

1) ตัวแปรสุ่ม หมายถึง ฟังก์ชันที่มีค่าเป็นเลขจำนวนจริง ซึ่งกำหนดจากผลลัพธ์ที่ได้จากการทดลองเชิงสุ่ม หรือเป็นตัวแปร ซึ่งกำหนดค่าให้กับผลการทดลองต่าง ๆ ที่เกิดขึ้น โดย

X คือ ชื่อของตัวแปรสุ่ม

x เป็นค่าของตัวแปรสุ่ม

2) ชนิดของตัวแปรสุ่ม ตัวแปรสุ่มแบ่งออกเป็น 2 ชนิด ดังนี้

2.1) ตัวแปรสุ่มแบบไม่ต่อเนื่อง (Discrete Random Variable)

- ค่าของตัวแปรสุ่มจะไม่ต่อเนื่อง มีค่าได้เป็นบางค่า
- จำนวนค่าของตัวแปรสุ่มแบบไม่ต่อเนื่อง อาจมีจำนวนจำกัดหรือไม่จำกัดก็ได้
- ส่วนมากจะมีค่าเป็นจำนวนเต็ม (จำนวนนับ) แต่อาจมีค่าเป็นทศนิยมได้

เช่น ในการตรวจสอบสินค้า 10 ชิ้น ว่ามีสินค้าเสียปนมาด้วยหรือไม่ ถ้าให้ X เป็นจำนวนสินค้าเสีย X จะเป็นตัวแปรสุ่มแบบไม่ต่อเนื่อง ค่าที่เป็นไปได้ของ X คือ 0, 1, 2, …, 10

2.2) ตัวแปรสุ่มแบบต่อเนื่อง (Continuous Random Variable)

- ค่าของตัวแปรจะเป็นเซทของเลขจำนวนจริงที่ต่อเนื่องกัน

- จำนวนค่าของตัวแปรสุ่มแบบต่อเนื่อง มีจำนวนไม่จำกัด

- ส่วนมากเกี่ยวข้องกับมาตราชั่ง ตวง วัด เช่น น้ำหนัก ส่วนสูง ระยะเวลา อุณหภูมิ เป็นต้น

เช่น ระยะเวลาที่เครื่องจักรใช้งานได้ (หน่วยเป็นนาที)

X อาจมีค่าได้ 0, 0.0001, 1, 2.57, … หรือเขียนได้ว่า $X \geq 0$

3.3.5 ฟังก์ชันความน่าจะเป็น (Probability Function)

นิยาม

“ถ้าให้ S เป็นสเปซตัวอย่าง (sample space) ฟังก์ชันความน่าจะเป็น เป็นฟังก์ชันซึ่งแสดงว่าตัวแปรสุ่มจะมีค่าเท่ากับค่าที่กำหนดให้จากการทดลองใน S ด้วยความน่าจะเป็นเท่าใด”

- เป็นฟังก์ชันซึ่งแสดงค่าความน่าจะเป็นที่ตัวแปรสุ่มนั้นจะมีค่าใดค่าหนึ่งหรือมีค่าอยู่ในช่วงใดช่วงหนึ่งที่กำหนด

- จะใช้สัญลักษณ์ ดังนี้ P(X=x), P(x) หรือ f(x)

1) ฟังก์ชันของมวลความน่าจะเป็น (Probability Mass Function: P.M.F.)

- เป็นฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มแบบไม่ต่อเนื่อง เขียนย่อได้ว่า P.M.F.

- มีคุณสมบัติ ดังนี้

(1) $f(x) \geq 0 \quad \forall x$ (สำหรับทุกค่าของ x)

(2) $P(a \leq X \leq b) = \sum_{x=a}^{b} f(x)$

$P(a < X < b) = \sum_{x=a+1}^{b-1} f(x)$

(3) $\sum_{\forall x} f(x) = 1$

2) ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function: P.D.F.)

- เป็นฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่อง เขียนย่อได้ว่า P.D.F.

- มีคุณสมบัติ ดังนี้

(1) 1) $f(x) \geq 0 \quad \forall x$ (สำหรับทุกค่าของ x)

(2) $P(X = a) = 0$

(3) $P(a < X < b) = P(a \leq X \leq b)$

$= P(a \leq X < b)$

$= P(a < X \leq b)$

$= \int_a^b f(x)dx$

ซึ่งก็คือพื้นที่ใต้กราฟของ f(x) ที่อยู่ในช่วง a , b

(4) $\int_{\forall x} f(x)dx = 1$ (คือ พื้นที่ใต้กราฟทั้งหมด)

3.3.6 ฟังก์ชันการแจกแจงความน่าจะเป็นแบบสะสม (Cumulative Probability Distribution Function: C.D.F.)

นิยาม

"ฟังก์ชันการแจกแจงความน่าจะเป็นแบบสะสมของตัวแปรสุ่ม X คือฟังก์ชันซึ่งแสดงถึงค่าความน่าจะเป็นสะสมเมื่อ X มีค่าน้อยกว่าค่าใดค่าหนึ่ง"

- เขียนย่อได้ว่า C.D.F. หรือ c.d.f.
- C.D.F. ของ X คือ $F(x) = P(X \leq x)$

สำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง (DRV)

$$F(x) = \sum_{x=a}^{x} f(x) \quad ; x = a, a+1, \ldots, b$$

หรือ $= P(X= a) + P(X= a+1) + \ldots + P(X= x)$

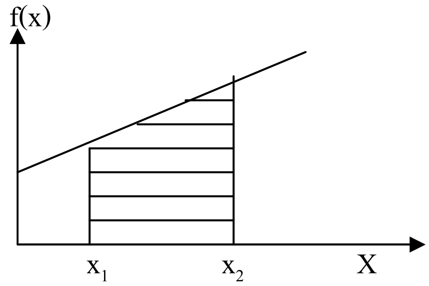
$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1 - 1)$

หรือ $= F(x_2) - F(x_1) + P(X= x_1)$

สำหรับตัวแปรสุ่มแบบต่อเนื่อง (CRV)

$$F(x) = \int_a^x f(x)dx \quad ; \quad a < x < b$$

$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$



<u>รูปที่ 11</u> ฟังก์ชันการแจกแจงความน่าจะเป็นแบบสะสมของตัวแปรสุ่ม X

3.3.7 การแจกแจงความน่าจะเป็นร่วม (Joint Probability Distribution)

- ในกรณีที่มีตัวแปรสุ่ม 2 ตัวที่เราต้องการศึกษา เราสามารถหาความน่าจะเป็นร่วมได้จาก

f(x,y) = ฟังก์ชันความน่าจะเป็นร่วมของตัวแปรสุ่ม X และ Y (Joint Probability Function)

คุณสมบัติ

1) $f(x,y) \geq 0$ $\forall (x,y)$ (สำหรับทุกค่า x , y)

2) $\sum_y \sum_x f(x, y) = 1$ สำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง (D.R.V.)

$$\int_y \int_x f(x,y)dxdy = 1$$ สำหรับตัวแปรสุ่มแบบต่อเนื่อง (C.R.V.)

3) ถ้า A เป็นเหตุการณ์ใด ๆ

$$P(A) = \sum_{(x,y)\in A}\sum f(x,y)$$ สำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง (D.R.V.)

$$= \int_{(x,y)\in A}\int f(x,y)dxdy$$ สำหรับตัวแปรสุ่มแบบต่อเนื่อง (C.R.V.)

f(x=2, y=1) = P(X=2 , Y=1) = ความน่าจะเป็นร่วม เมื่อ X = 2 และ Y = 1 สำหรับตัวแปรสุ่มแบบไม่ต่อเนื่อง (D.R.V.)

3.3.8 ความน่าจะเป็นเดี่ยว (Marginal Probability)

- ความน่าจะเป็นเดี่ยวของ X เขียนแทนด้วย $f_x(x)$ โดยที่

$$f_x(x) = \sum_y f(x,y)$$ สำหรับ D.R.V.

$$= \int_y f(x,y)dy$$ สำหรับ C.R.V.

- ความน่าจะเป็นเดี่ยวของ Y เขียนแทนด้วย $f_y(y)$ โดยที่

$$f_y(y) = \sum_x f(x,y)$$ สำหรับ D.R.V.

$$= \int_x f(x,y)dx$$ สำหรับ C.R.V.

- ให้ h(x/y) และ h(y/x) เป็นความน่าจะเป็นร่วมแบบมีเงื่อนไข จะได้ว่า $h(x/y) = \dfrac{f(x,y)}{f_y(y)}$ และ $h(y/x) = \dfrac{f(x,y)}{f_x(x)}$

h(x=1/y=2) = P(X=1/Y=2)

- ถ้า X และ Y เป็นอิสระจากกัน (independent) จะได้ว่า

$$f(x,y) = f_x(x)\, f_y(y)$$ สำหรับ $\forall$ (x,y)

3.3.9 ค่าคาดหวัง (Expected Value)

- ถ้าให้ X เป็นตัวแปรสุ่ม (R.V.) ค่าคาดหวังของ X หรือ E(x) ก็คือค่าเฉลี่ยของ X นั่นเอง

$$E(x) = \sum_{\forall x} xf(x)$$, เมื่อ X เป็น D.R.V.

$$= \int_{\forall x} xf(x)dx$$, เมื่อ X เป็น C.R.V.

โดยที่ f(x) เป็นฟังก์ชันการแจกแจงความน่าจะเป็น

- สำหรับ D.R.V. อาจเขียนค่าคาดหวังของ X ได้ว่า

$$E(x) = \sum_{\forall x} xP(X = x)$$

ค่าคาดหวังของฟังก์ชัน h(x) และ h(x,y)

$$E[h(x)] = \sum_{\forall x} h(x)f(x) \quad \text{เมื่อ x เป็น D.R.V.}$$

$$= \int_{\forall x} h(x)f(x)dx \quad \text{เมื่อ x เป็น C.R.V.}$$

$$E[h(x,y)] = \sum_{\forall x}\sum_{\forall y} h(x,y)f(x,y) \quad \text{เมื่อ x และ y เป็น D.R.V.}$$

$$= \int_{\forall y}\int_{\forall x} h(x,y)f(x,y)dxdy \quad \text{เมื่อ x และ y เป็น C.R.V.}$$

กฎของค่าคาดหวัง (Law of Expectation)

1) ถ้า a และ b เป็นค่าคงที่ จะได้

$E[ax+b] = aE(x) + b$

2) ถ้า $h_1(x)$ และ $h_2(x)$ เป็นฟังก์ชันตัวแปรสุ่มของ X จะได้

$E[h_1(x)+ h_2(x)] = E[h_1(x)] + E[h_2(x)]$

$E[h_1(x) – h_2(x)] = E[h_1(x)] – E[h_2(x)]$

3) ถ้า X และ Y เป็นตัวแปรสุ่ม ซึ่งมีความน่าจะเป็นร่วม f(x,y) และ $h_1(x,y)$ กับ $h_2(x,y)$ เป็นฟังก์ชันของ X และ Y

$E[h_1(x,y)+ h_2(x,y)] = E[h_1(x,y)] + E[h_2(x,y)]$

$E[h_1(x,y) – h_2(x,y)] = E[h_1(x,y)] – E[h_2(x,y)]$

4) ถ้า X และ Y เป็นตัวแปรสุ่มที่อิสระจากกัน จะได้

$E[xy] = E(x) E(y)$

3.3.10 โมเมนต์ (Moment) และค่าความแปรปรวน (Variance)

- ถ้า $g(x) = x^k$ หรือ $(x– c)^k$ โดยที่ c เป็นค่าคงที่ จะเรียก E[g(x)] ว่า โมเมนต์

เรียก $E[x^k]$ ว่าโมเมนต์ที่ k ของ x รอบจุดกำเนิด

เรียก $E[(x– c)^k]$ ว่าโมเมนต์ที่ k ของ x รอบ c

- ถ้า c = E(x) จะได้ $E[(x- E(x))^k]$ เรียก $E[(x- E(x))^k]$ ว่าโมเมนต์ที่ k ของ x รอบจุด E(x)

ถ้า k = 2 จะได้ $E[(x- E(x))^2]$ = ค่าความแปรปรวน (Variance)

$$= V(x) = \sigma^2 = \sigma_x^2$$

และจะได้ $\sqrt{E[(x - E(x))^2]}$ = ค่าเบี่ยงเบนมาตรฐาน = σ = σ_x

$$E[(x-E(x))^2] = \sum_{\forall x}(x - E(x))^2 f(x) \qquad \text{สำหรับ D.R.V.}$$

$$= \int_{\forall x}(x - E(x))^2 f(x)dx \qquad \text{สำหรับ C.R.V.}$$

ดังนั้น $E[(x-E(x))^2] = E(x^2) - [E(x)]^2$

-ในกรณีที่มีตัวแปรสุ่ม 2 ตัว (X และ Y) เราจะสนใจค่าความแปรปรวนร่วม (Covariance)

$Cov(x,y) = C_{xy} = E[(x-E(x))(y-E(y))] = E[xy] - E(x)\,E(y)$

ถ้า X และ Y เป็นอิสระจากกัน จะได้ $Cov(x,y) = 0$

คุณสมบัติ

1) ถ้า $h(x) = x + c$ จะได้ $V(x + c) = V(x)$

2) ถ้า $h(x) = ax$ จะได้ $V(ax) = a^2 V(x)$

3) ถ้า X และ Y เป็นตัวแปรสุ่มที่มีฟังก์ชันความน่าจะเป็นร่วม จะได้

$V(ax + by) = a^2 V(x) + b^2V(y) + 2ab\,Cov(x,y)$

$V(ax - by) = a^2 V(x) + b^2V(y) - 2ab\,Cov(x,y)$

4) ถ้า X และ Y เป็นตัวแปรสุ่มที่มีฟังก์ชันความน่าจะเป็นร่วม และ X และ Y เป็นอิสระจากกัน จะได้

$Cov(x,y) = 0$

$V(ax + by) = V(ax - by) = a^2 V(x) + b^2V(y)$

3.3.11 การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่องที่สำคัญ (Discrete Probability Distribution)

1) การแจกแจงความน่าจะเป็นเอกรูป (Uniform Probability Distribution)

ถ้า X เป็นตัวแปรสุ่มที่มีค่าต่างกัน คือ $x_1, x_2, \ldots, x_k$ ด้วยความน่าจะเป็นที่เท่า ๆ กัน จะเรียก X ว่ามีการแจกแจงแบบ Uniform และมีพารามิเตอร์ k โดยมีฟังก์ชันการแจกแจงความน่าจะเป็น f(x) ค่าคาดหวัง E(x) และค่าความแปรปรวน V(x) ดังนี้

$$\boxed{f(x,k) = \frac{1}{k}} \quad , \quad x = x_1, x_2, \ldots, x_k$$

$$E(x) = \frac{1}{k}\sum_{i=1}^{k} x_i$$

$$V(x) = \frac{1}{k}\sum_{i=1}^{k}(x_i - E(x))^2$$

ตัวอย่าง โยนลูกเต๋า 1 ลูก ให้ X แทนแต้มที่ได้จากการโยน จงหาฟังก์ชันการแจกแจงความน่าจะเป็นของ X, E(x) และ V(x)

จะได้ X = 1, 2, 3, 4, 5, 6

P(X=1) = P(X=2) = … = P(X=6) = 1/6

ดังนั้น X มีการแจกแจงแบบ Uniform และมีค่าพารามิเตอร์ k = 6 ซึ่งจะได้

$$f(x,6) = \frac{1}{6} \quad , \quad x = 1 , 2 , \dots , 6$$

$$E(x) = \frac{1}{k}\sum_{i=1}^{k} x_i = \frac{1}{6}\sum_{i=1}^{6} x_i$$

$$= 1/6\ (1+2+3+4+5+6) = 7/2$$

$$V(x) = \frac{1}{k}\sum_{i=1}^{k}(x_i - E(x))^2 = \frac{1}{6}\sum_{i=1}^{6}(x_i - \frac{7}{2})^2$$

$$= 1/6\ [(1 - 7/2)^2 + (2 - 7/2)^2 + \dots + (6 - 7/2)^2] = 35/12$$

2) การแจกแจงความน่าจะเป็นแบบเบอร์นูลลี่ (Bernoulli Probability Distribution)

การแจกแจงความน่าจะเป็นแบบเบอร์นูลลี่ เหมือนการทดลองสุ่มแบบ Binomial แต่เป็นการทดลองเพียงครั้งเดียวเท่านั้น

จาก $f(x, n , p) = \binom{n}{x} p^x q^{n-x}$

เมื่อ n = 1 จะได้ x = 0 , 1

$$\boxed{f(x , p) = p^x q^{1-x}} \quad , \quad x = 0 , 1$$

E(x) = p

V(x) = pq

3) การแจกแจงความน่าจะเป็นแบบทวินาม (Binomial Probability Distribution)

การแจกแจงความน่าจะเป็นแบบทวินาม มีการทดลองลักษณะดังนี้

3.1) การทดลองซ้ำกันหลาย ๆ ครั้ง (n ครั้ง) เช่น โยนเหรียญหนึ่งเหรียญ 10 ครั้ง

3.2) การทดลองแต่ละครั้งมีผลลัพธ์ที่เป็นไปได้ 2 อย่าง คือ สิ่งที่สนใจ (success) และสิ่งที่ไม่สนใจ (failure)

3.3) การทดลองแต่ละครั้งเป็นอิสระกัน

3.4) ความน่าจะเป็นที่จะเกิดสิ่งที่สนใจ มีค่าคงที่และเท่ากันในทุกการทดลอง คือ p นั่นก็คือ P(สิ่งที่สนใจ) = p และ P(สิ่งที่ไม่สนใจ) = 1- p = q

การทดลองใด ๆ ที่สอดคล้องกับลักษณะทั้ง 4 ข้อ เรียกว่า การทดลองแบบทวินาม (Binomial Experiment)

ให้ X เป็นตัวแปรสุ่ม X จะหมายถึง จำนวนครั้งของการทดลองที่เกิดสิ่งที่สนใจ เรียก X ว่า ตัวแปรสุ่มทวินาม

สำหรับการทดลอง n ครั้ง จะได้ X = 0 , 1 , 2 , … , n

ให้ S เป็นสิ่งที่สนใจ และ F เป็นสิ่งที่ไม่สนใจ ถ้าในการทดลอง n ครั้ง เกิด S เป็นจำนวน x ครั้ง ฉะนั้นจะเกิด F เป็นจำนวน n-x ครั้ง

$$\underbrace{SS\ldots S}_{\text{เกิด S จำนวน x ครั้ง}} \quad \underbrace{FF\ldots F}_{\text{เกิด F จำนวน n-x ครั้ง}}$$

เนื่องจาก P(S) = p และ P(F) = q จะได้ ความน่าจะเป็นที่จะเกิด S เป็นจำนวน x ครั้ง และเกิด F เป็นจำนวน n-x ครั้ง มีค่าเป็น $\binom{n}{x}p^x q^{n-x}$

$$\boxed{P(X=x) = f(x, n, p) = \binom{n}{x}p^x q^{n-x}} \quad , \quad x = 0, 1, 2, \ldots, n$$

$E(x) = np$

$V(x) = npq$

4) การแจกแจงความน่าจะเป็นแบบพัวซอง (Poisson Probability Distribution)

ถ้าให้ X หมายถึง จำนวนครั้งของสิ่งที่เราสนใจที่เกิดขึ้นในช่วงเวลาใดเวลาหนึ่งหรือในขอบเขตใดขอบเขตหนึ่ง X จะมีการแจกแจงแบบพัวซอง (Poisson)

ตัวอย่างเช่น X หมายถึงจำนวนคำผิดที่พบในหนังสือ 1 หน้า ซึ่งขอบเขตที่สนใจ คือ 1 หน้า

ให้ λ เป็นจำนวนสิ่งที่สนใจที่เกิดขึ้นโดยเฉลี่ยในช่วงเวลาหรือขอบเขต จะได้ฟังก์ชันการแจกแจงความน่าจะเป็นแบบพัวซอง ดังนี้

$$\boxed{f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}} \quad , \quad x = 0, 1, 2, \ldots$$

$E(x) = \lambda$

$V(x) = \lambda$

5) การแจกแจงความน่าจะเป็นแบบทวินามด้วยการแจกแจงแบบพัวซอง (Poisson Approximation to Binomial Probability Distribution)

ถ้า X เป็นตัวแปรสุ่มที่มีการแจกแจงแบบ Binomial ที่มีค่า n มาก ๆ ($n \rightarrow \infty$) และ p มีค่าน้อยมาก ๆ ($p \rightarrow 0$) แล้ว เราสามารถประมาณค่าของฟังก์ชันความน่าจะเป็นของ X ด้วยฟังก์ชันความน่าจะเป็นแบบ Poisson

โดยทั่วไปจะใช้การแจกแจงแบบ Poisson ประมาณค่าการแจกแจงแบบ Binomial เมื่อ $n \geq 20$ และ $p \leq 0.05$ หรือ $np < 7$

$$\boxed{\binom{n}{x} p^x q^{n-x} \leftarrow \frac{e^{-\lambda}\lambda^x}{x!}}$$ โดยที่ $\lambda = np$

$E(x) = np$

$V(x) = npq$

6) การแจกแจงความน่าจะเป็นแบบไฮเพอร์ยีออเมตริก (Hypergeometric Probability Distribution)

การทดลองแบบ Hypergeometric จะยึดหลักการสุ่มแบบไม่มีการแทนที่ ซึ่งแตกต่างจาก Binomial มีลักษณะ คือ มีสิ่งของ N ชิ้น มีสิ่งที่สนใจ k ชิ้น สิ่งที่ไม่สนใจ N-k ชิ้น สุ่มสิ่งของออกมา n ชิ้น ได้สิ่งที่สนใจ x ชิ้น สิ่งที่ไม่สนใจ n-x ชิ้น

$$\boxed{f(x, N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}} \quad , \quad x = 0, 1, \ldots, \min(k, n)$$

$$E(x) = \frac{nk}{N}$$

$$V(x) = \left(\frac{N-n}{N-1}\right)\left(1-\frac{k}{N}\right)\left(\frac{nk}{N}\right)$$

7) การการแจกแจงความน่าจะเป็นแบบไฮเพอร์ยีออเมตริกด้วยการแจกแจงแบบทวินาม (Binomial to Hypergeometric Probability Distribution)

เป็นการแจกแจงความน่าจะเป็นแบบไฮเพอร์ยีออเมตริกที่มีขนาดตัวอย่างเล็กมากเมื่อเทียบค่าขนาดของประชากรแล้ว นั่นหมายความว่าการสุ่มตัวอย่างแบบใส่คืนกับไม่ใส่คืนจะให้ผลใกล้เคียงกัน ดังนั้นการแจกแจงความน่าจะเป็นแบบไฮเพอร์ยีออเมตริก จะมีค่า

$$\boxed{f(x) = P(X=x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}} \quad ; \binom{k}{x} = \frac{k!}{(k-x)!x!}$$

$E(x) = m = nk/N$

$$V(x) = \frac{N-n}{N-1} \cdot \frac{nk}{N}\left(1-\frac{k}{N}\right)$$

และการแจกแจงแบบทวินาม มีค่า

$$f(x) = P(X=x) = {}^{n}C_{x}\, p^{x}(1-p)^{n-x} = \binom{n}{x} p^{x} q^{n-x} \qquad ; x = 0, 1,2,\ldots,n$$

$$E(x) = m = np = n \cdot \frac{k}{N}$$

$$V(x) = npq = n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

3.3.12 การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง (Continuous Probability Distribution)

1) การแจกแจงความน่าจะเป็นเอกรูป (Uniform Probability Distribution)

การแจกแจงความน่าจะเป็นแบบเอกรูปนี้ ค่าทุกค่าของตัวแปรสุ่มมีโอกาสเกิดได้เท่า ๆ กัน

$$f(x,a,b) = \frac{1}{b-a} \quad , \quad a < x < b \text{ และ}$$

$$f(x,a,b) = 0 \quad , \quad \text{else}$$

$$E(x) = \frac{a+b}{2}$$

$$V(x) = \frac{(b-a)^2}{12}$$

2) การแจกแจงความน่าจะเป็นแบบปกติ (Normal Probability Distribution)

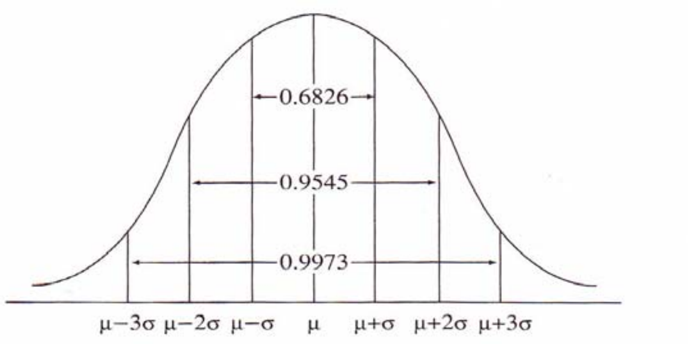
การแจกแจงความน่าจะเป็นแบบปกติ มีการใช้กันมากที่สุด เนื่อง จากข้อมูลส่วนมากมีการแจกแจงแบบปกติหรือใกล้เคียงแบบปกติ

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad , \quad -\infty \leq x \leq \infty$$

โดยที่ $-\infty \leq \mu \leq \infty$ และ $\sigma > 0$

$$E(x) = \mu$$

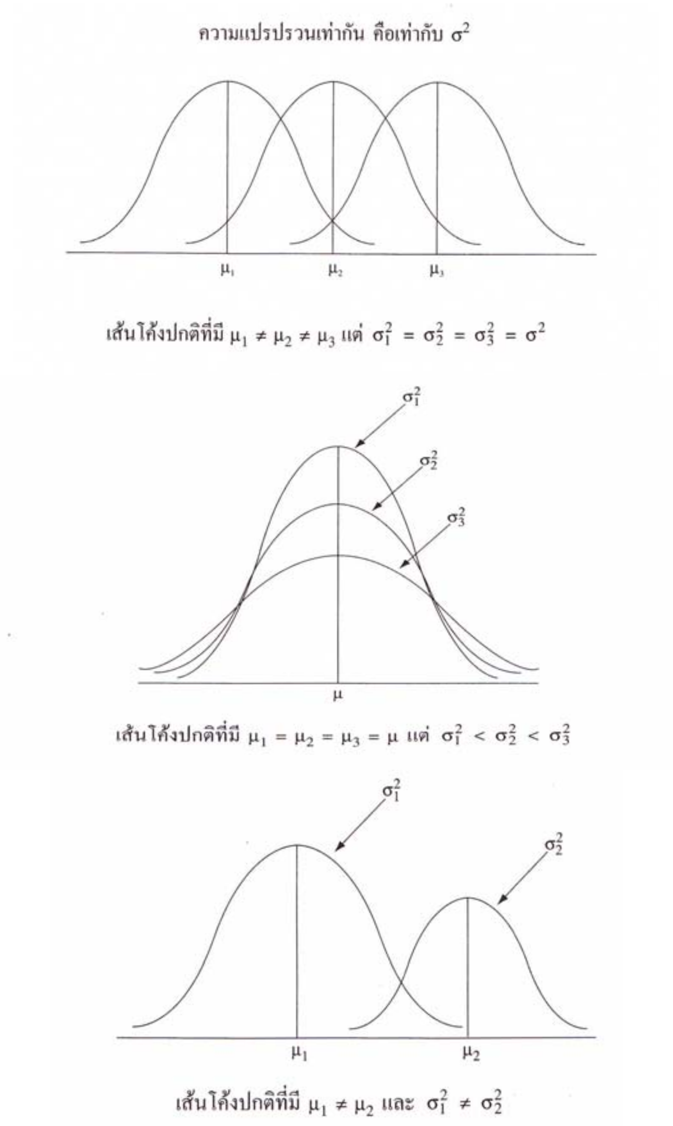
$$V(x) = \sigma^2$$



<u>รูปที่ 13</u> พื้นที่เส้นโค้งปกติที่เป็นรูประฆังคว่ำมีความสมมาตร (พิศมัย หาญมงคลพิพัฒน์, 2550: 66)

คุณสมบัติ

1. กราฟของฟังก์ชันเป็นรูประฆังคว่ำ
2. สมการ x = μ เป็นแกนสมมาตร
3. มีจุดสูงสุดอยู่จุดเดียว (unimodel) ที่ x = μ ค่าสูงสุด คือ $\frac{1}{\sigma\sqrt{2\pi}}$
4. ค่า Mean = Median = Mode
5. $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$
6. พื้นที่ใต้กราฟรวมกันทั้งหมด มีค่าเท่ากับ 1 นั่นคือ $\int_{-\infty}^{\infty} f(x)dx = 1$
7. P(X=x) = 0



<u>รูปที่ 13</u> เส้นโค้งปกติที่มีค่าเฉลี่ย และความแปรปรวนในลักษณะต่าง ๆ

(พิศมัย หาญมงคลพิพัฒน์, 2550: 65)

3) การแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution)

การแจกแจงแบบปกติ ที่ $\mu = 0$ และ $\sigma^2 = 1$

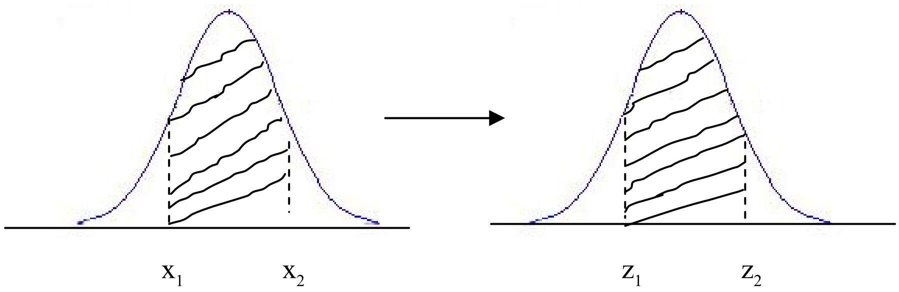
มี z เป็นตัวแปรสุ่ม (standard random variable) ที่แปลงค่ามาจากตัวแปรสุ่มแบบปกติ x โดยที่

$$\boxed{Z = \frac{x - \mu}{\sigma}}$$

เมื่อ $z \sim N(0, 1)$ และ $f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$, $-\infty \leq z \leq \infty$

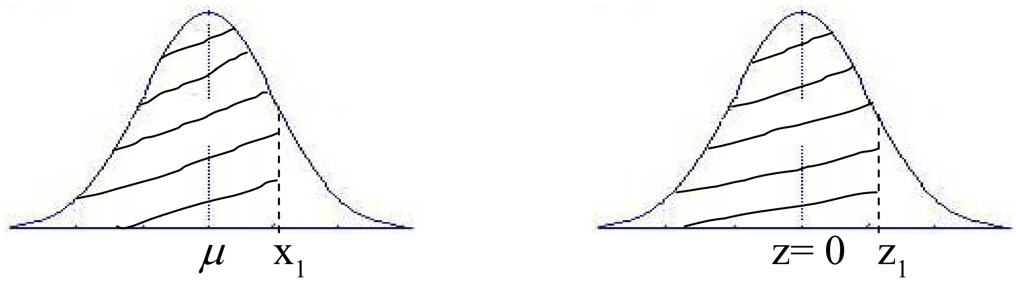
ในการหา $P(x_1 < x < x_2)$ ก็ปรับค่าเป็น $P(z_1 < z < z_2)$

โดยที่ $z_1 = \frac{x_1 - \mu}{\sigma}$ และ $z_2 = \frac{x_2 - \mu}{\sigma}$



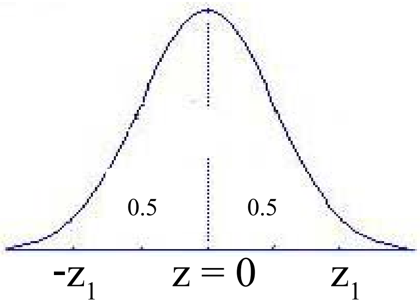
รูปที่ 14 การแจกแจงแบบปกติมาตรฐานที่แปลงค่ามาจากตัวแปรสุ่มแบบปกติ x

การหาค่าความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐาน สามารถหาค่าได้จากตารางการแจกแจงความน่าจะเป็นแบบปกติมาตรฐาน



รูปที่ 15 ค่า $P(x < x_1) = 1 - P(x > x_1)$ ที่มีการแจกแจงแบบปกติมาตรฐาน $P(z < z_1) = P(z < \frac{x_1 - \mu}{\sigma})$

พื้นที่ใต้กราฟทั้งหมดมีค่าเท่ากับ 1 ดังนั้น ครึ่งหนึ่งของพื้นที่ใต้กราฟ จะเท่ากับ 0.5



<u>รูปที่ 16</u> พื้นที่ใต้โค้งปกติทั้งหมดมีค่าเท่ากับ 1

ถ้า $P(z < z_1) = 0.8$ จะได้ $P(z > z_1) = 1 - 0.8 = 0.2$ และ

$P(z < -z_1) = 0.2$

$P(-z_1 < z < z_1) = (0.8 - 0.5)(2) = 0.6$

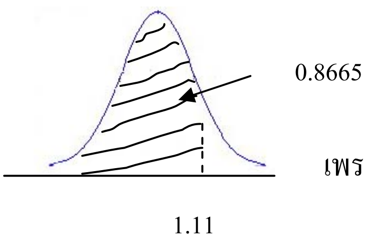
หรือ $= 1 - (0.2)(2) = 0.6$

ตัวอย่าง

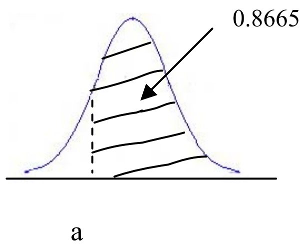
จงหาค่า a ที่ทำให้ ก. $P(z \geq a) = 0.8665$ ข. $P(z \geq a) = 0.0244$ ค. $P(z < a) =$ 0.0025 ง. $P(0 < z < a) = 0.4808$ จ. $P(-a < z < a) = 0.995$ ฉ. $P(z < a) = 0.98$

ก. $P(z \geq a) = 0.8665$

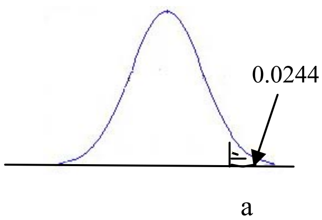
เปิดตาราง จะได้ $P(z < 1.11) = 0.8665$



เพราะฉะนั้น $a = -1.11$

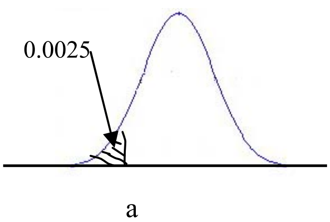


ข. $P(z \geq a) = 0.0244$ ดังนั้น $P(z < a) = 1 - 0.0244 = 0.9756$



จากตาราง จะได้ $a = 1.97$

ค. $P(z < a) = 0.0025$

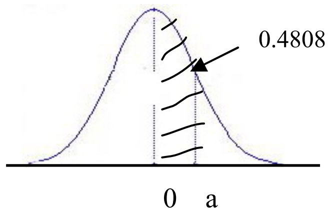


เปิดตารางหาค่า z_1 ที่ทำให้ $P(z < z_1) = 1 - 0.0025 = 0.9975$

จะได้ $z_1 = 2.81$ ดังนั้น $a = -2.81$

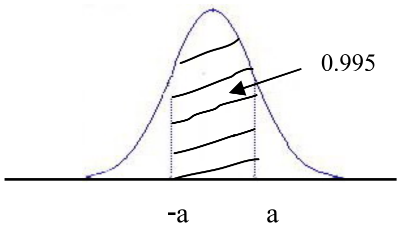
ง. $P(0 < z < a) = 0.4808$

จะได้ $P(z < a) = 0.4808 + 0.5 = 0.9808$



จากตารางจะได้ $a = 2.07$

จ. $P(-a < z < a) = 0.995$

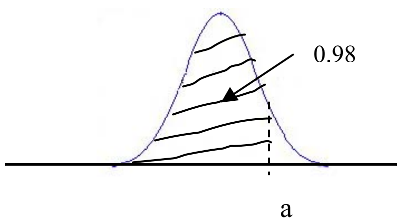


จะได้ว่า $P(z < a) = 0.995 + \frac{(1-0.995)}{2}$

$= 0.9975$

จากตาราง จะได้ $a = 2.81$

ฉ. $P(z < a) = 0.98$



จากตารางจะได้ $P(z < 2.06) = 0.9803$

$P(z < 2.05) = 0.9798$

เพราะฉะนั้น $a = 2.054$

4) การประมาณค่าความน่าจะเป็นแบบทวินาม (Binomial Distribution) ด้วยความน่าจะเป็นแบบปกติ (Normal Distribution)

จะใช้การแจกแจงแบบปกติมาประมาณค่าการแจกแจงแบบทวินามได้ เมื่อ $np \geq 5$ และ $nq \geq 5$ โดยที่

$\mu = np \quad \sigma^2 = npq$

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

$$\boxed{z = \frac{x - np}{\sqrt{npq}}}$$

ในกรณีที่ต้องการหาความน่าจะเป็นเมื่อตัวแปรสุ่มมีการแจกแจงแบบBinomial ถ้าจะนำเอาการแจกแจงแบบปกติมาประมาณค่า จะต้องมีการปรับค่าของตัวแปรสุ่มก่อน เนื่องจากเป็นการประมาณค่าตัวแปรสุ่มแบบไม่ต่อเนื่องด้วยค่าของตัวแปรสุ่มแบบต่อเนื่อง เช่น

ให้ x เป็นตัวแปรสุ่มแบบ Binomial

$P(x < 90)$ ต้องปรับค่าเป็น $P(x \leq 89.5)$

$P(x > 90)$ ต้องปรับค่าเป็น $P(x \geq 90.5)$

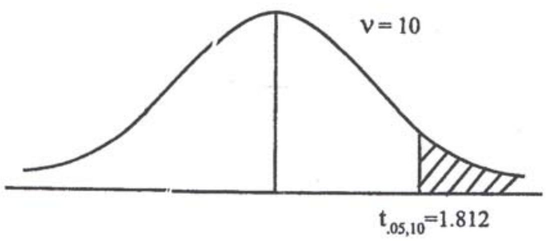
5) การแจกแจงความน่าจะเป็นแบบที (Student-t Probability Distribution)

ถ้า $z \sim N(0,1)$ และ $V \sim \chi^2_{(n)}$ ให้ T เป็นตัวแปรสุ่มที่มีค่า ดังนี้

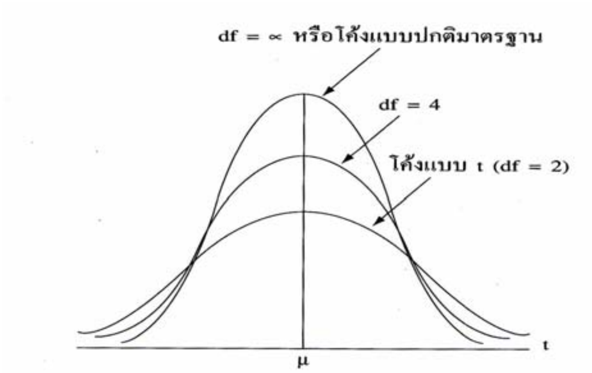
$$T = \frac{z}{\sqrt{V/n}}$$

T จะมีการแจกแจงแบบ Student-t ที่มี degree of freedom = n และฟังก์ชันคือ

$$f(t\,,n) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi\, n}}(1+\frac{t^2}{n})^{-(\frac{n+1}{2})} \quad , \quad -\infty < t < \infty$$



รูปที่ 17 กราฟของฟังก์ชันมีลักษณะเหมือนกับกราฟ Normal (พิศมัย หาญมงคลพิพัฒน์, 2550: 75)



รูปที่ 18 กราฟของฟังก์ชันที่มี df ในค่าต่าง ๆ (พิศมัย หาญมงคลพิพัฒน์, 2550: 75)

คุณสมบัติ

1. กราฟของฟังก์ชันเป็นรูประฆังคว่ำ
2. สมการ t = 0 เป็นแกนสมมาตร
3. มีจุดสูงสุดอยู่จุดเดียว (unimodel) ที่ t = 0
5. $\lim_{t\to\infty} f(t) = \lim_{t\to-\infty} f(t) = 0$
6. พื้นที่ใต้กราฟรวมกันทั้งหมด มีค่าเท่ากับ 1
7. $P(t = t_0) = 0$

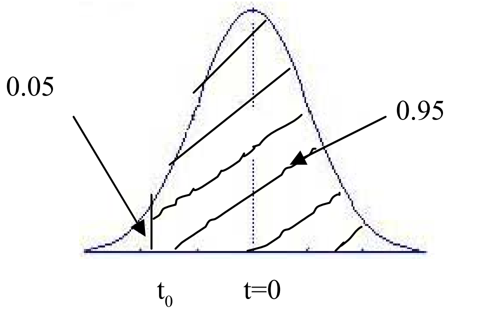
เมื่อทราบค่า α และค่า df ก็สามารถหาค่า t ได้จากตาราง

เช่น α = 0.05 และ df = 23 จากตารางจะได้ t0.05, 23 = 1.714

ตัวอย่าง

จงหาค่า t_0 ที่ทำให้ เมื่อ df = 23 ก. $P(t > t_0) = 0.95$ ข. $P(-t_0 \le t \le t_0) = 0.9$

ก. $P(t > t_0) = 0.95$

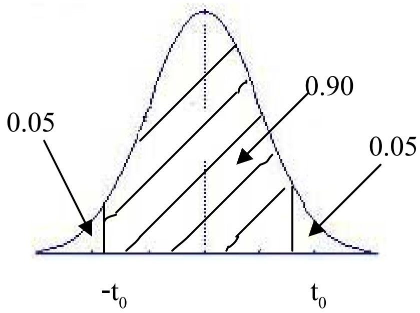


จะได้ว่า $t_0 = -t_{0.05, 23}$

จากตารางที่ α = 0.05 และ df = 23

จะได้ $t_{0.05, 23} = 1.714$ ดังนั้น $t_0 = -1.714$

ข. $P(-t_0 \le t \le t_0) = 0.9$



จะได้ว่า $P(t > t_0) = 0.05$

จากตารางที่ α = 0.05 และ df = 23

จะได้ $t_{0.05, 23} = 1.714$ ดังนั้น $t_0 = 1.714$

6) การแจกแจงความน่าจะเป็นแบบไคสแควร์ (Chi-Square Distribution: χ^2)

การแจกแจงความน่าจะเป็นแบบไคสแควร์ จะเป็นโค้งที่ไม่สมมาตร เรียกว่า Gamma Distribution ที่ $\alpha = \frac{\upsilon}{2}$ และ $\beta = 2$ โดยที่ $\upsilon \in I^+$

- υ เป็น องศาแห่งความเป็นอิสระ (degree of freedom = df)

$$\boxed{f(x, \upsilon) = \frac{1}{2^{\frac{\upsilon}{2}}\Gamma(\frac{\upsilon}{2})} e^{\frac{-x}{2}} x^{\frac{\upsilon}{2}-1}} \quad , \quad x > 0$$

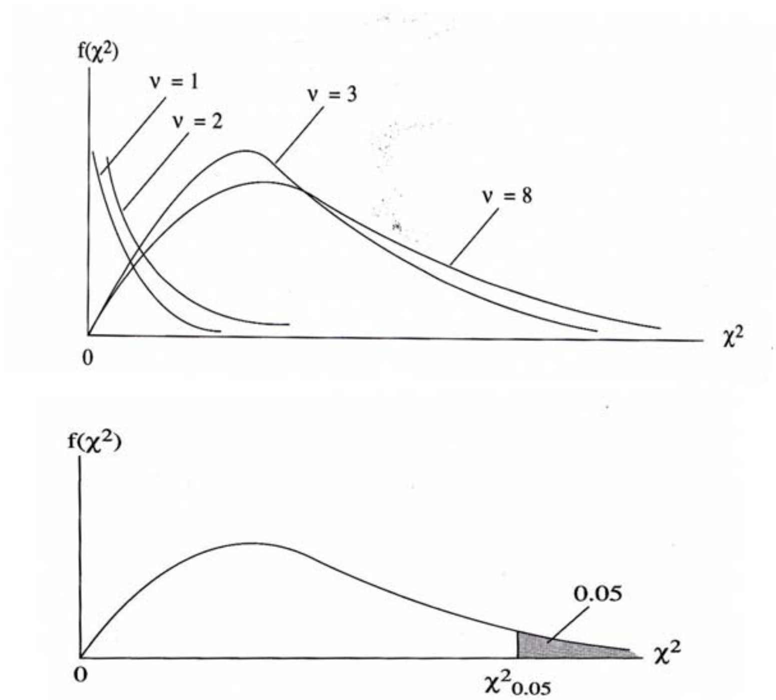
$$= \quad 0 \qquad , \quad \text{else}$$

$E(x) = \upsilon$

$V(x) = 2\upsilon$

- ถ้าตัวแปรสุ่ม $z_1, z_2, \ldots, z_n$ เป็นอิสระต่อกันและมีการแจกแจงแบบปกติมาตรฐาน (standard normal distribution) N(0,1) จะได้

$$z_1^2 + z_2^2 + \ldots + z_n^2 \sim \chi^2_{(n)} \quad (df = n)$$



รูปที่ 19 กราฟการแจกแจงความน่าจะเป็นแบบไคสแควร์ (พิศมัย หาญมงคลพิพัฒน์, 2550: 77-78)

คุณสมบัติ

1. $\chi^2 \geq 0$

2. พื้นที่ใต้กราฟรวมกันทั้งหมดมีค่าเท่ากับ 1

3. กราฟมีลักษณะเบ้ขวา แตกแต่งกันตามค่าของ df

4. P(χ^2= x) = 0

-เราสามารถหาค่าของตัวแปรสุ่มไคสแควร์ χ^2 ได้จากตารางการแจกแจงแบบไคสแควร์ เมื่อเรารู้ค่าความน่าจะเป็นαและค่า df เช่น α = 0.1 และ df = 5 จากตารางจะได้ $\chi^2_{0.1,5} = 9.236$

ตัวอย่าง

P($\chi^2 \leq \chi^2_{0.05,20}$) = 0.95 $\chi^2_{0.05,20}$ มีค่าเท่าไร

นั่นคือ P($\chi^2 \geq \chi^2_{0.05,20}$) = 0.05

จากตาราง ที่ df = 20 และα = 0.05 จะได้ $\chi^2_{0.05,20}$ = 31.41

7) การแจกแจงความน่าจะเป็นแบบเอฟ (F Probability Distribution)

การแจกแจงความน่าจะเป็นแบบเอฟ เป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่องอีกแบบหนึ่ง คือ เป็นการแจกแจงของอัตราส่วนของตัวแปรที่มีการแจกแจงแบบไคสแควร์ 2 ตัว ที่เป็นอิสระต่อกันหารด้วยองศาความเป็นอิสระของตัวมันเอง คือ

$$F(\nu_1,\nu_2) = \frac{\chi^2_{(\nu2)}/\nu_1}{\chi^2_{(\nu1)}/\nu_2}$$

หรือ ถ้า U และ V เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยที่ U ~ $\chi^2_{(n)}$ และ V ~ $\chi^2_{(m)}$ จะได้ว่า

$F = \frac{U/n}{V/m}$ เป็นตัวแปรสุ่มที่มีการแจกแจงแบบ F มี df = n , m

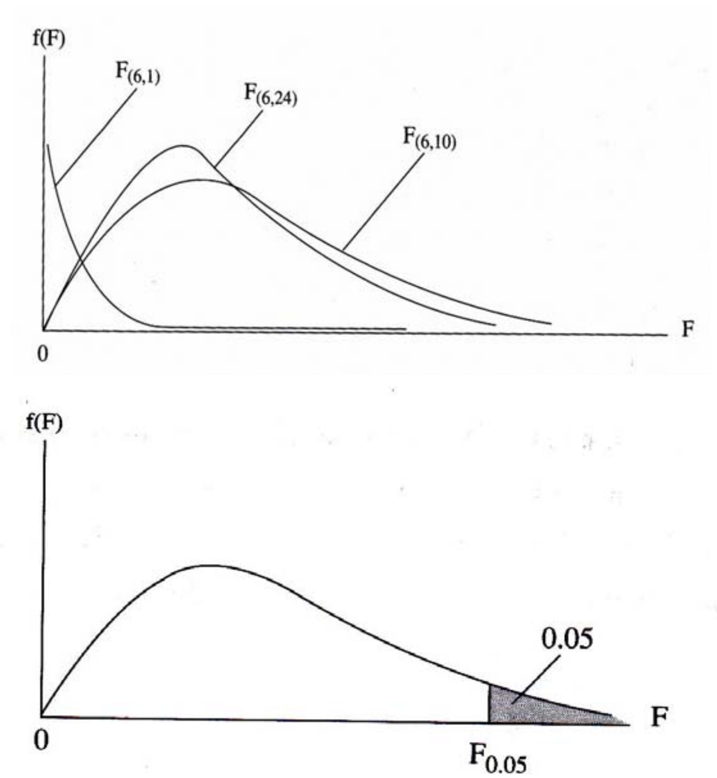
ฟังก์ชันการแจกแจงความน่าจะเป็นของ F คือ

$$G(f, n, m) = \boxed{\frac{\Gamma(\frac{n+m}{2})\left(\frac{n}{m}\right)^{\frac{n}{2}} f^{(\frac{n}{2}-1)}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})\left(1+\frac{nf}{m}\right)^{\frac{n+m}{2}}}} \quad , \quad f > 0$$

$$= 0 \quad , \quad \text{else}$$

คุณสมบัติ

1. กราฟของฟังก์ชันมีลักษณะเบ้ขวา
2. พื้นที่ใต้กราฟรวมกันทั้งหมด มีค่าเท่ากับ 1
3. F $\geq$ 0
4. P(F = f_0) = 0



รูปที่ 20 กราฟการแจกแจงความน่าจะเป็นแบบเอฟ (พิศมัย หาญมงคลพิพัฒน์, 2550: 79-80)

เมื่อทราบค่า α และค่า df (n, m) ก็สามารถหาค่า $F_{\alpha\ n,m}$ ได้จากตาราง เช่น α = 0.05 และ df = 2, 3 จากตารางจะได้ค่า $F_{0.05\ 2,3}$ = 9.55

สำหรับค่า α บางค่า เราไม่สามารถหาค่า $F_{\alpha\ n,m}$ ได้โดยตรงจากตาราง แต่เราอาจหาค่า $F_{\alpha\ n,m}$ ได้ โดยใช้คุณสมบัติ ดังต่อไปนี้

$$\boxed{f_{1-\alpha\ (n,m)} = \frac{1}{f_{\alpha\ (m,n)}}}$$

-ถ้า σ_1^2 และ σ_2^2 เป็นค่าความแปรปรวนของประชากรที่ 1 และ 2 ตามลำดับ โดยที่ s_1^2 เป็นค่าความแปรปรวนของตัวอย่างที่ได้จากการสุ่มตัวอย่างจำนวน m จากประชากรที่ 1 และ s_2^2 เป็นค่าความแปรปรวนของตัวอย่างที่ได้จากการสุ่มตัวอย่างจำนวน n จากประชากรที่ 2 จะได้ว่า

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$ มีการแจกแจงแบบ F ที่ df = m-1 และ n-1

ตัวอย่าง

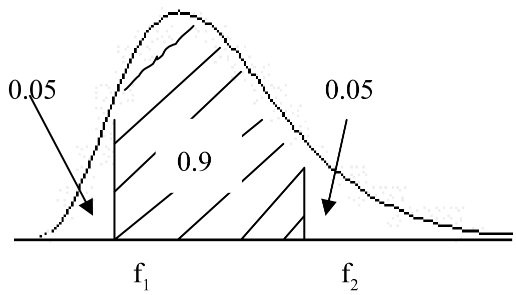
จงหาค่า f_0 ที่ทำให้ $P(F < f_0) = 0.99$ โดยที่ df = 4,5

$P(F < f_0) = 0.99$ นั่นคือ $P(F > f_0) = 0.01$

จะได้ว่า $f_0 = f_{0.01\ 4,5}$ จากตารางจะได้ $f_{0.01\ 4,5} = 11.39$

ตัวอย่าง

จงหาค่า f_1 และ f_2 ที่ทำให้ $P(f_1 < F < f_2) = 0.90$ และ $P(F \le f_1) = 0.05$ ที่ df เท่ากับ 9 และ 15



รูปที่ 21 การหาค่า f_1 และ f_2

จะได้ $P(F>f_2) = 0.05$ และ$P(F>f_1) = 0.95$

ดังนั้น $f_2 = f_{0.05\ 9,15} = 2.59$

และ f1 = $f_{0.95\ 9,15}$

เนื่องจากไม่สามารถหาค่า $f_{0.95\ 9,15}$ จากตารางได้ ดังนั้นจะใช้คุณสมบัติ

$$f_{1-\alpha\ (n,m)} = \frac{1}{f_{\alpha\ (m,n)}}$$

$$f_{0.95\ 9,15} = \frac{1}{f_{0.05\ 15,9}}$$

$$= \frac{1}{3.01} = 0.332$$

3.4 หน่วยที่ 4 การแจกแจงของค่าสถิติ

การแจกแจงของฟังก์ชันที่ได้จากตัวอย่างสุ่ม (Sampling Distribution)

- ประชากร: ทุกหน่วยของสิ่งที่สนใจศึกษา

- ตัวอย่าง: บางส่วนของสิ่งที่สนใจศึกษา (subset ของประชากร)

- ค่าที่บ่งบอกลักษณะของประชากร เรียกว่า ค่าพารามิเตอร์ เช่น ค่าเฉลี่ย (μ) ค่าความแปรปรวน (σ^2) ขนาดประชากร (N)

- ค่าที่คำนวณหาได้จากตัวอย่าง เรียกว่า ค่าสถิติ เช่น ค่าเฉลี่ย ($\bar{x}$) ค่าความแปรปรวน (s^2) ขนาดตัวอย่าง (n)

- การศึกษาข้อมูลต่าง ๆ ของประชากร ถ้าต้องเก็บข้อมูลจากจำนวนประชากรทั้งหมด จะทำให้ต้องเสียค่าใช้จ่ายสูง เสียเวลา บางกรณีเป็นไปได้ยาก จึงต้องเก็บข้อมูลจากตัวอย่างแทน

- ค่าสถิติที่หาได้จากตัวอย่าง สามารถนำมาวิเคราะห์เพื่อหาข้อสรุปเกี่ยวกับค่าพารามิเตอร์ของประชากรได้

- ตัวอย่างที่เลือกในการเก็บข้อมูล ต้องเป็นตัวแทนที่ดีของประชากร

- การเลือกตัวอย่างแบบสุ่มหรือการสุ่มตัวอย่าง (Random Sampling) จะทำให้ได้ตัวอย่างที่ดี ตัวอย่างที่ได้นี้เรียกว่า ตัวอย่างสุ่ม (random sample)

3.4.1 การแจกแจงของค่าเฉลี่ยตัวอย่าง

กำหนดให้ตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ (normal distribution) ที่มีค่าเฉลี่ย (μ) และค่าความแปรปรวน (σ^2)

1) กรณีที่ทราบค่าความแปรปรวนของประชากร σ^2

ถ้าสุ่มตัวอย่าง n จากประชากร แล้วหาค่าเฉลี่ยตัวอย่าง $\bar{x}$

จะได้ $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ และหาค่า z ได้จาก

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad \text{โดยที่ } z \sim N(0\ ,\ 1)$$

2) กรณีไม่ทราบค่าความแปรปรวนของประชากร σ^2 และ n<30

เมื่อไม่ทราบค่า σ^2 และ $n < 30$ จะใช้การแจกแจงแบบ student-t ในการวิเคราะห์ค่าเฉลี่ยตัวอย่าง $\bar{x}$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \qquad , \qquad df = n-1$$

โดยที่ s เป็นส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

3) กรณีไม่ทราบค่าความแปรปรวนของประชากร σ^2 และ n ≥ 30

ในกรณีไม่รู้ค่า σ^2 แต่ n ≥ 30 ค่าเฉลี่ยตัวอย่าง $\bar{x}$ จะมีการแจกแจงแบบปกติ $\bar{x} \sim N(\mu, \frac{s^2}{n})$ และหา z ได้จาก

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

ทฤษฎีลิมิตสู่ส่วนกลาง (Central Limit Theorem)

ถ้าสุ่มตัวอย่างขนาด n ($x_1, x_2, \ldots, x_n$) จากประชากรใดๆที่มีค่าเฉลี่ย μ และค่าความแปรปรวน σ^2 ถ้า n มีขนาดใหญ่ ค่าเฉลี่ยตัวอย่าง $\bar{x}$ จะมีการแจกแจงเข้าสู่การแจกแจงแบบปกติ ที่มีค่าเฉลี่ย μ และค่าความแปรปรวน $\frac{\sigma^2}{n}$ นั่นคือ

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \text{ และ } z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \text{ โดยที่ } z \sim N(0, 1)$$

สรุป

1. สำหรับประชากรที่มีการแจกแจงใดๆ ถ้า n ≥ 30 แล้ว ค่าเฉลี่ยตัวอย่าง $\bar{x}$ จะมีการแจกแจงประมาณแบบปกติ

2. ค่าเฉลี่ยของ $\bar{x}$ จะมีค่าเท่ากับค่าเฉลี่ยของประชากร μ

$$\mu_{\bar{x}} = \mu_x = \mu$$

3. ค่าความแปรปรวนของ $\bar{x}$ จะเท่ากับ $\frac{\sigma^2}{n}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

3.4.2 การแจกแจงของผลต่างของค่าเฉลี่ยตัวอย่าง ($\bar{X}_1 - \bar{X}_2$)

- ใช้ในกรณีที่ศึกษาประชากร 2 ประชากร
- วิเคราะห์ผลต่างของค่าเฉลี่ยของ 2 ประชากร โดยศึกษาจากตัวอย่างของประชากรทั้งสอง

1) ประชากรทั้งสองมีการแจกแจงแบบปกติ และทราบค่า σ_1^2, σ_2^2

- ถ้ามีประชากร 2 ประชากร ประชากรที่ 1 มีค่าเฉลี่ย μ_1 และความแปรปรวน σ_1^2 ประชากรที่ 2 มีค่าเฉลี่ย μ_2 และความแปรปรวน σ_2^2

- สุ่มตัวอย่าง n_1 จากประชากรที่ 1 แล้วหาค่าเฉลี่ยตัวอย่าง $\bar{X}_1$

สุ่มตัวอย่าง n_2 จากประชากรที่ 2 แล้วหาค่าเฉลี่ยตัวอย่าง $\bar{X}_2$

- ผลต่างของค่าเฉลี่ยตัวอย่างจากประชากรทั้งสอง ($\bar{x}_1 - \bar{x}_2$) จะมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น ($\mu_1 - \mu_2$) และค่าความแปรปรวนเป็น ($\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$) นั่นคือ ($\bar{x}_1 - \bar{x}_2$) ~ N($\mu_1 - \mu_2$, $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$) ดังนั้น

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ โดยที่ $z \sim N(0\ ,\ 1)$

2) ประชากรทั้งสองไม่มีการแจกแจงแบบปกติและ $n_1 \geq 30$, $n_2 \geq 30$

- เมื่อรู้ค่าความแปรปรวนของประชากรทั้งสอง σ_1^2, σ_2^2 จะได้ว่า ($\bar{x}_1 - \bar{x}_2$) ~ N($\mu_1 - \mu_2$, $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$) ดังนั้น

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ โดยที่ $z \sim N(0\ ,\ 1)$

- เมื่อไม่รู้ค่า σ_1^2, σ_2^2 จะประมาณค่า σ_1^2, และ σ_2^2 ด้วย s_1^2 และ s_2^2 ตามลำดับ จะได้ว่า ($\bar{x}_1 - \bar{x}_2$) ~ N($\mu_1 - \mu_2$, $\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$) ดังนั้น

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ โดยที่ $z \sim N(0\ ,\ 1)$

3) ประชากรทั้งสองมีการแจกแจงแบบปกติ แต่ไม่ทราบ σ_1^2, σ_2^2 และ $n_1 < 30$, $n_2 < 30$

3.1) เมื่อรู้ว่า $\sigma_1^2 = \sigma_2^2$

- ในกรณีนี้ต้องหาความแปรปรวนร่วม s_p^2 จาก

$$S_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

- ใช้การแจกแจงแบบ student-t ในการวิเคราะห์ผลต่างของค่าเฉลี่ยของสองประชากร โดยที่

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad , \quad df = n_1 + n_2 - 2$$

3.2) เมื่อรู้ว่า $\sigma_1^2 \neq \sigma_2^2$

- จะประมาณค่า σ_1^2 และ σ_2^2 ด้วย s_1^2 และ s_2^2 ตามลำดับ

- ใช้การแจกแจงแบบ student-t ในการวิเคราะห์ผลต่างของค่าเฉลี่ยของสองประชากร

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad , \quad df = \upsilon$$

โดยที่ $$\upsilon = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

3.4.3 การแจกแจงของค่าสัดส่วนตัวอย่าง (Sample Proportion: $\hat{p}$)

- ในบางกรณีเราต้องศึกษาว่าประชากรกี่เปอร์เซนต์ที่มีคุณสมบัติตรงกับสิ่งที่เราสนใจ เช่น เปอร์เซนต์หรือสัดส่วนของชาวจังหวัดอุบลราชธานีที่มีสิทธิลงคะแนนเสียงเลือกตั้ง สัดส่วนของผู้หญิงที่สูบบุหรี่ในประเทศไทย

- ในการสุ่มตัวอย่าง ผลที่ได้จะมีอยู่ 2 อย่าง คือ สัดส่วนตัวอย่างของสิ่งที่เราสนใจ $\hat{p}$ และสัดส่วนตัวอย่างของสิ่งที่เราไม่สนใจ $\hat{q}$

- สมมติว่าประชากรมีสัดส่วนของสิ่งที่เราสนใจ p และมีสัดส่วนของสิ่งที่เราไม่สนใจ q ถ้าจำนวนตัวอย่างที่สุ่ม n มากพอ และทั้ง $np > 5$

และ $nq > 5$ จะได้ว่า สัดส่วนตัวอย่าง $\hat{p}$ จะมีการแจกแจงแบบปกติ มีค่าเฉลี่ย p และค่าความแปรปรวน $\frac{pq}{n}$ นั่นคือ $\hat{p} \sim N(p, \frac{pq}{n})$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \quad , \quad z \sim N(0, 1)$$

3.4.4 การแจกแจงของผลต่างของสัดส่วนตัวอย่าง ($\hat{p}_1 - \hat{p}_2$)

ถ้า $\hat{p}_1$ และ $\hat{p}_2$ มีการแจกแจงแบบปกติ ($n_1p_1 > 5$, $n_1q_1 > 5$ และ $n_2p_2 > 5$, $n_2q_2 > 5$) จะได้ว่า $\hat{p}_1 - \hat{p}_2$ มีการแจกแจงแบบปกติ ด้วยค่าเฉลี่ย $p_1 - p_2$ และค่าความแปรปรวน $\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$ นั่นคือ

$(\hat{p}_1 - \hat{p}_2) \sim N(p_1 - p_2, \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2})$ ดังนั้น

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} \quad , \quad z \sim N(0, 1)$$

3.4.5 การแจกแจงการสุ่มตัวอย่างของ $\frac{(n-1)s^2}{\sigma^2}$

- ในบางกรณีเราต้องการศึกษาความแปรปรวนของประชากร σ^2 โดยการวิเคราะห์ความแปรปรวนของตัวอย่าง s^2

- การแจกแจงของค่าสถิติ s^2 มีใช้น้อยมากในทางปฏิบัติ ดังนั้น จึงใช้การแจกแจงของตัวสถิติ $\frac{(n-1)s^2}{\sigma^2}$ แทน

- ถ้า s^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n ที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติซึ่งมีค่าความแปรปรวน σ^2 จะได้ว่า $\frac{(n-1)s^2}{\sigma^2}$ จะมีการแจกแจงไคสแควร์ มีองศาแห่งความเป็นอิสระ (degree of freedom = df) เท่ากับ n – 1 นั่นคือ

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad , \quad df = n-1$$

3.5 หน่วยที่ 5 การอนุมานเชิงสถิติ

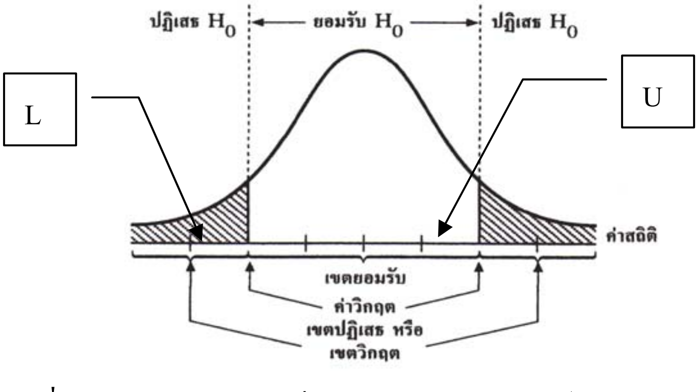
สถิติเชิงอนุมาน (Statistical Inference) จะเกี่ยวข้องกับการสรุปผลเกี่ยวกับพารามิเตอร์ในประชากรโดยอาศัยการวิเคราะห์จากข้อมูลตัวอย่าง ซึ่งประกอบด้วย 2 ส่วน คือ การประมาณค่าพารามิเตอร์ (Estimation of Parameters) และการทดสอบสมมติฐานทางสถิติ (Testing of Statistical Hypothesis)

3.5.1 การประมาณค่าพารามิเตอร์ (Estimation of Parameters)

การประมาณค่าพารามิเตอร์ในประชากรสามารถกระทำได้ใน 2 ลักษณะ คือ การหาค่าประมาณแบบค่าเดียว (point estimation) และการหาค่าประมาณแบบเป็นช่วง (interval estimation)

การหาค่าประมาณแบบค่าเดียว เป็นการคำนวณค่าของตัวสถิติหนึ่งจากข้อมูลตัวอย่างและใช้ค่าที่คำนวณได้เป็นค่าประมาณของพารามิเตอร์ เช่น การใช้ค่าเฉลี่ยของตัวอย่าง ($\overline{X}$) เป็นตัวประมาณค่าเฉลี่ยของพารามิเตอร์ประชากร (μ) เป็นต้น

การหาค่าประมาณแบบเป็นช่วง เป็นการประมาณว่าค่าพารามิเตอร์หนึ่งคาดว่าจะมีค่าอยู่ระหว่างค่าสองค่า คือ L = ขอบเขตต่ำ (lower limit) และ U = ขอบเขตสูง (upper limit) สมมติว่าพารามิเตอร์ที่ต้องการประมาณในประชากรหนึ่ง คือ θ ค่าประมาณแบบเป็นช่วงของ θ นี้ จะอยู่ในรูป $L \le \theta \le U$



รูปที่ 22 ค่าประมาณแบบเป็นช่วงของ θ (ชูศรี วงศ์รัตนะ, 2550: 127)

ค่า L และ U จะขึ้นกับพารามิเตอร์ที่ต้องการประมาณ ตัวประมาณที่ใช้ในการประมาณการแจกแจงของตัวประมาณและระดับความเชื่อมั่นที่ต้องการ สมมติว่าพารามิเตอร์ที่ต้องการประมาณคือ

ค่าเฉลี่ยประชากร (μ) โดยมีข้อสมมติว่าประชากรนี้มีการแจกแจงแบบปกติที่ทราบค่าเบี่ยงเบนมาตรฐานเป็น σ ด้วยความน่าจะเป็น .95 ซึ่งค่า $Z_{.025} = 1.96$ จะเขียนได้ว่า

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95$$

ซึ่งหมายความว่าช่วง $\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}$ ถึง $\overline{X} + 1.96\frac{\sigma}{\sqrt{n}}$ จะครอบคลุมค่าพารามิเตอร์ μ ด้วยความน่าจะเป็น .95 เรียกช่วงดังกล่าวนี้ว่า ช่วงความเชื่อมั่น 95% ของ μ เมื่อประชากรมีการแจกแจงปกติและทราบค่าเบี่ยงเบนมาตรฐานของประชากร (σ)

การเลือกตัวประมาณค่า ในการประมาณค่าพารามิเตอร์ เราต้องพิจารณาว่าจะใช้ค่าสถิติตัวใดเป็นหลักในการประมาณนั้น เช่นการประมาณค่าเฉลี่ยของประชากร ค่าสถิติที่แสดงถึงค่ากลางของข้อมูลมีหลายค่า เป็นต้นว่าค่าเฉลี่ยเลขคณิต ค่าฐานนิยม และค่ามัธยฐาน ดังนั้น เราจะต้องพิจารณาว่าควรจะเลือกตัวใด เป็นตัวประมาณได้ หรือในกรณีที่ต้องการประมาณค่าความแปรปรวนของประชากร เราควรจะเลือกสถิติตัวใดเป็นตัวประมาณที่ดีได้ ซึ่งจำเป็นต้องพิจารณาถึงคุณสมบัติต่าง ๆ ที่สำคัญได้ดังนี้

1) ไม่ลำเอียง (Unbiasesness)

ถ้าให้ $\hat{\theta}$ เป็นตัวประมาณค่าพารามิเตอร์ θ (ค่าของประชากร) $\hat{\theta}$ จะเป็นตัวประมาณที่ไม่เอนเอียง ถ้าค่าเฉลี่ยของ $\hat{\theta}$ ที่คำนวณจากตัวแทนสุ่มที่เป็นไปได้ทุกตัวแทนมีค่าพารามิเตอร์ ซึ่งเขียนแทนด้วย $E(\hat{\theta}) = \theta$ สำหรับทุก ๆ ค่า ของ θ $E(\hat{\theta})$อ่านว่า “Expected Value” หรือค่าคาดหมายของ $\hat{\theta}$ เช่น

$E(\overline{X}) = \mu$ กล่าวได้ว่า $\overline{X}$ เป็นตัวประมาณที่ไม่เอนเอียงของ μ ทั้งนี้เพราะ

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}(E(X_i))$$

$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{i=1}^{n} X_i \cdot \frac{1}{n}\right)$ เป็นโอกาสที่ค่าสังเกตแต่ละตัวถูกสุ่มจากประชากร 1/n

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}(\mu)$$

$$= \frac{1}{n} \cdot n\mu$$

$$= \mu$$

$E(S^2) = \sigma^2$ หรือ S^2 เป็นตัวประมาณที่ไม่เอนเอียงของ σ^2 เพราะ $E(S^2) = \sigma^2$ แต่ $E(S) \neq \sigma$ หมายความว่า ค่าเบี่ยงเบนมาตรฐานของตัวอย่างเป็นค่าประมาณที่เอนเอียงของค่าเบี่ยงเบนมาตรฐานของประชากร

2) ต้องมีประสิทธิภาพมากที่สุด (Efficiency)

ถ้าให้ $\hat{\theta}_1$ และ $\hat{\theta}_2$ ต่างเป็นตัวประมาณที่ไม่เอนเอียงของ θ แต่ถ้าค่าความแปรปรวนของ $\hat{\theta}_1$ $\hat{\theta}$ มีค่าน้อยกว่าค่าความแปรปรวนของ $\hat{\theta}_2$ นั่นคือ $\sigma^2_{e1} < \sigma^2_{e2}$ หมายความว่า $\hat{\theta}_1$ เป็นตัวประมาณที่มี

ประสิทธิภาพสูงกว่า $\hat{\theta}_2$ ตัวอย่างเช่น ในการประมาณค่าเฉลี่ยของประชากร สามารถประมาณได้ด้วยค่าตัวกลางต่าง ๆ เช่น ค่าเฉลี่ยเลขคณิต ($\overline{X}$) ค่ามัธยฐาน ของกลุ่มตัวแทน ทั้งนี้เพราะทั้งค่าเฉลี่ยเลขคณิตและค่ามัธยฐานต่างเป็นตัวประมาณที่ไม่เอนเอียง แต่ความแปรปรวนของค่าเฉลี่ยเลขคณิตน้อยกว่าความแปรปรวนของค่ามัธยฐาน จึงกล่าวได้ว่าค่าเฉลี่ยเลขคณิตเป็นตัวประมาณที่มีประสิทธิภาพดีกว่าการใช้ค่ามัธยฐานเป็นตัวประมาณ

3.5.2 การทดสอบสมมติฐาน (Hypothesis Testing)

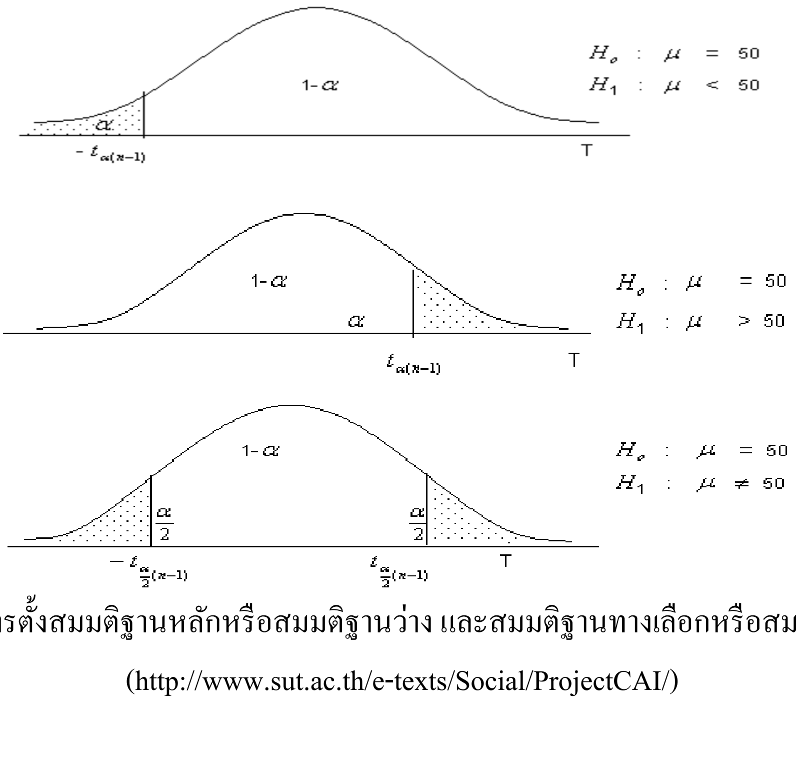
การทดสอบสมมติฐานเป็นการอ้างอิงทางสถิติอย่างหนึ่ง โดยเป็นการตั้งสมมติฐานเกี่ยวกับค่าพารามิเตอร์ของประชากรแล้วสุ่มตัวอย่างเพื่อนำค่าจากตัวอย่างมาเปรียบเทียบกับค่าตามสมมติฐานเพื่อตัดสินใจยอมรับหรือปฏิเสธสมมติฐาน

โดยมีสมมติฐานทางสถิติอยู่ 2 สมมติฐานควบคู่กันไปเสมอ คือ

1. สมมติฐานหลักหรือสมมติฐานว่าง (Null Hypothesis) ใช้สัญลักษณ์ H_0 โดยสมมติฐานหลักนี้จะมีเครื่องหมายเท่ากับอยู่ด้วยเสมอ

2. สมมติฐานทางเลือกหรือสมมติฐานแย้ง (Alternative Hypothesis) ใช้สัญลักษณ์ H_1 หรือ H_A โดยสมมติฐานทางเลือกนี้จะเป็นสมมติฐานที่มีเครื่องหมายตรงกันข้ามกับสมมติฐานหลัก

ตามปกติทั่วไปการตั้งสมมติฐานว่าง จะตั้งเป็นแบบแน่นอนตายตัว ซึ่งเรียกว่าสมมติฐานเชิงเดียว (simple hypothesis) โดยตั้งอยู่ในรูปของพารามิเตอร์ เท่ากับค่าคงตัว ส่วนการตั้งสมมติฐานทางเลือกอื่น จะต้องเป็นแบบประกอบ (composite hypothesis) โดยตั้งอยู่ในรูปเป็นช่วงของพารามิเตอร์ได้แก่ พารามิเตอร์น้อยกว่า หรือมากกว่า หรือไม่เท่ากับ ค่าคงตัวในสมมติฐานว่าง นั่นคือนอกจากจะตั้งสมมติฐานข้างต้นแล้ว อาจจะตั้งเป็น



<u>รูปที่ 23</u> การตั้งสมมติฐานหลักหรือสมมติฐานว่าง และสมมติฐานทางเลือกหรือสมมติฐานแย้ง

(http://www.sut.ac.th/e-texts/Social/ProjectCAI/)

1) ประเภทของความผิดพลาด

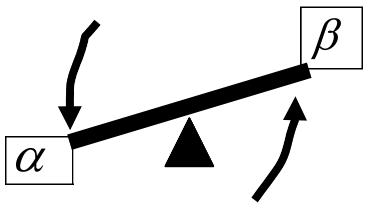
การทดสอบสมมติฐานในงานวิจัยจะเกิดความผิดพลาดขึ้นได้ 2 ลักษณะคือ ความผิดพลาดชนิดที่ 1 (type I error) และความผิดพลาดชนิดที่ 2 (type II error)

1.1) ความผิดพลาดชนิดที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานหลักหรือสมมติฐานว่าง (null hypothesis) เมื่อสมมติฐานหลักเป็นจริง ความน่าจะเป็นที่จะเกิดความผิดพลาดชนิดที่ 1 นี้คือ α เรียกว่า ระดับนัยสำคัญ (level of significance) ส่วนค่า $1-\alpha$ เรียกว่า ระดับความเชื่อมั่น

1.2) ความผิดพลาดชนิดที่ 2 (Type II error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐานหลักหรือสมมติฐานว่าง (null hypothesis) เมื่อสมมติฐานหลักเป็นเท็จ ความน่าจะเป็นที่จะเกิดความผิดพลาดชนิดที่ 2 นี้คือ β

ตารางที่ 4 การทดสอบสมมติฐาน และความผิดพลาด

	H_0 เป็นจริง	**H_0 เป็นเท็จ**
ยอมรับ H_0	ถูกต้อง	ความผิดพลาดชนิดที่ 2
ปฏิเสธ H_0	ความผิดพลาดชนิดที่ 1	ถูกต้อง



รูปที่ 24 ความผิดพลาด 2 ลักษณะที่เกิดขึ้น

ความผิดพลาดทั้งสองชนิดนี้จะแปรผกผันกันเสมอ ซึ่งส่วนใหญ่ผู้วิจัยจะยอมให้เกิดความผิดพลาดชนิดที่ 1 มากกว่าเพราะความเสียหายที่เกิดขึ้นจะน้อยกว่าที่ให้เกิดความเสียหายจากความผิดพลาดชนิดที่ 2 โดยที่ความผิดพลาดชนิดที่ 1(α) ที่นิยมใช้คือ 0.01 และ 0.05

2) การกำหนดขอบเขตการยอมรับสมมติฐาน

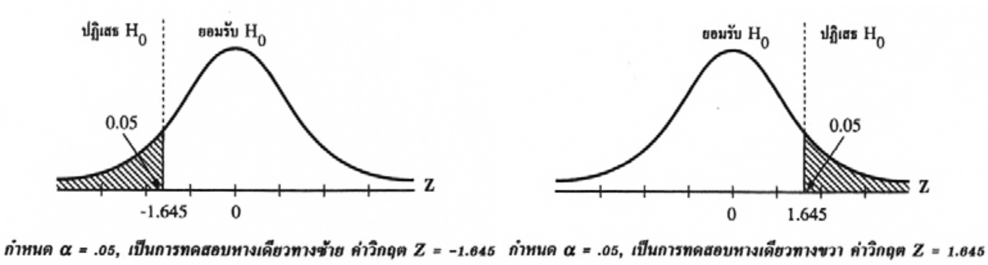
การปฏิเสธหรือยอมรับสมมติฐานหลัก จะขึ้นอยู่กับค่าสถิติที่คำนวณได้จากตัวอย่าง และทำการปรับเป็นค่ามาตรฐานตามชนิดของสถิติที่ทดสอบ เช่น Z test, t test, f test หรือ χ^2 test แล้วนำค่าที่ได้ไปเปรียบเทียบกับค่าวิกฤต (critical value) ที่ได้จากตารางพื้นที่ใต้โค้งตามตัวสถิติที่ใช้ในการทดสอบ ซึ่งค่าวิกฤตจะแบ่งพื้นที่ใต้โค้งเป็นเขตสองเขต คือ เขตปฏิเสธสมมติฐานหลัก และเขตการยอมรับสมมติฐานหลัก

3) ประเภทของการทดสอบสมมติฐาน

3.1) การทดสอบสมมติฐานแบบทางเดียว (One – tail test) แบ่งเป็นการทดสอบทางเดียวด้านขวาหรือข้างมาก และการทดสอบทางเดียวด้านซ้ายหรือข้างน้อย ถ้าจะปฏิเสธสมมติฐานว่าง ค่าพารามิเตอร์จะต้องน้อยกว่าหรือมากกว่าค่าคงตัวในสมมติฐานว่าง

$H_0 : \mu \geq \mu_0$ $\qquad$ $H_0 : \mu \leq \mu_0$

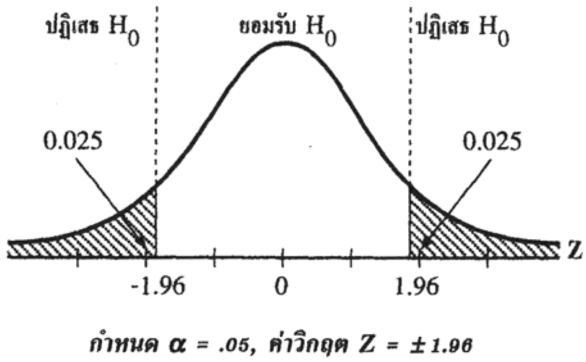
$H_1 : \mu < \mu_0$ $\qquad$ $H_1 : \mu > \mu_0$



<u>รูปที่ 25</u> การทดสอบสมมติฐานแบบทางเดียว (ชูศรี วงศ์รัตนะ, 2550: 130)

3.2) การทดสอบสมมติฐานแบบสองทาง (Two – tail test) ถ้าจะปฏิเสธสมมติฐานว่าง ค่าพารามิเตอร์จะต้องไม่เท่ากับค่าคงตัวในสมมติฐานว่าง

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$



<u>รูปที่ 26</u> การทดสอบสมมติฐานแบบสองทาง (ชูศรี วงศ์รัตนะ, 2550: 129)

4) สรุปขั้นตอนในการทดสอบสมมติฐาน

4.1) ตั้งสมมติฐาน (Hypothesis) ตั้งสมมติฐานหลัก หรือสมมติฐานว่าง (null hypothesis) ใช้สัญลักษณ์ H_0 และตั้งสมมติฐานทางเลือกหรือสมมติฐานแย้ง (alternative hypothesis) ใช้สัญลักษณ์ H_1 หรือ H_A

4.2) กำหนดระดับนัยสำคัญ (Statistical Significance Level:α) ในการทดสอบสมมติฐานทางสถิติ จะยอมรับสมมติฐานที่ตั้งขึ้นก็ต่อเมื่อคำนวณค่าความน่าจะเป็นได้มากกว่าค่าระดับ

นัยสำคัญทางสถิติ ซึ่งใช้แทนด้วยสัญลักษณ์ α และถ้าคำนวณค่าความน่าจะเป็นออกมาได้น้อยกว่า ก็แสดงว่าจะปฏิเสธสมมติฐานที่ตั้ง โดยทั่วไปจะกำหนดระดับนัยสำคัญ α ไว้ที่ 0.05 และ 0.01

4.3) รวบรวมข้อมูล และคำนวณค่าสถิติที่ใช้ในการทดสอบสมมติฐาน

4.4) ตัดสินใจ และสรุปผลว่า ยอมรับสมมติฐาน หรือปฏิเสธสมมติฐานที่ตั้งขึ้น อย่างมีนัยสำคัญทางสถิติ

5) การใช้ค่า P-values ในการทดสอบสมมติฐาน

ถ้าค่า $P \geq \alpha$ จะยอมรับ H_0

ถ้าค่า $P < \alpha$ จะปฏิเสธ H_0

3.6 หน่วยที่ 6 การประมาณค่า

3.6.1 การประมาณค่าเฉลี่ยของประชากร

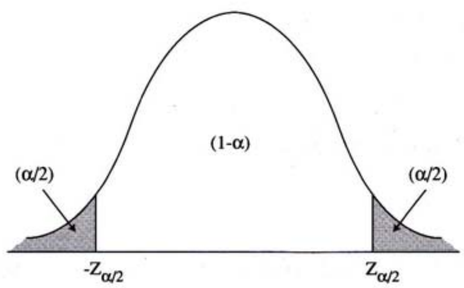
1) การประมาณค่าเฉลี่ยของประชากร 1 กลุ่ม

1.1) การประมาณค่าเฉลี่ยของประชากร เมื่อทราบค่าความแปรปรวน

- ใช้ตัวสถิติ z ในการประมาณค่า

- ถ้าสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงแบบปกติและมีความแปรปรวน σ^2 แล้วหาค่าเฉลี่ยตัวอย่างได้เป็น $\bar{x}$ ช่วงความเชื่อมั่น $100(1-\alpha)$ % ของค่าเฉลี่ยประชากร μ แบบสองด้านหาได้จาก

$$\boxed{\bar{x} + z_{\frac{\alpha}{2}} \sigma/\sqrt{n} \le \mu \le \bar{x} + z_{1-\frac{\alpha}{2}} \sigma/\sqrt{n}}$$



รูปที่ 27 แสดงค่า $P(z_{\frac{\alpha}{2}} \le z \le z_{1-\frac{\alpha}{2}}) = 1-\alpha$ (พิศมัย หาญมงคลพิพัฒน์, 2550: 143)

จะได้ $P(z_{\frac{\alpha}{2}} \le z \le z_{1-\frac{\alpha}{2}}) = 1-\alpha$

จาก $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ ดังนั้น

$$P\left(z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(z_{\frac{\alpha}{2}} \sigma/\sqrt{n} \le \bar{x} - \mu \le z_{1-\frac{\alpha}{2}} \sigma/\sqrt{n}\right) = 1 - \alpha$$

$$P\left(\bar{x} + z_{\frac{\alpha}{2}} \sigma/\sqrt{n} \le \mu \le \bar{x} + z_{1-\frac{\alpha}{2}} \sigma/\sqrt{n}\right) = 1 - \alpha$$

- ช่วงความเชื่อมั่น 100(1–α)% ของค่าเฉลี่ยประชากร μ แบบด้านเดียว หาจาก $\mu \le \bar{x} + z_{1-\alpha} \sigma/\sqrt{n}$ สำหรับด้านบนด้านเดียว และ $\bar{x} + z_{\alpha} \sigma/\sqrt{n} \le \mu$ สำหรับด้านล่างด้านเดียว

1.2) การประมาณค่าเฉลี่ยของประชากร เมื่อไม่ทราบค่าความแปรปรวน

เนื่องจากเราไม่ทราบค่าความแปรปรวนประชากร σ^2 ดังนั้นในการสร้างช่วงความเชื่อมั่น จึงใช้ค่าความแปรปรวนตัวอย่าง s^2 แทน

1.2.1) เมื่อมีขนาดตัวอย่างเล็ก

- ถ้าขนาดตัวอย่าง n มีค่าน้อยกว่า 30 ($n < 30$) จะใช้ตัวสถิติ t ในการหาช่วงความเชื่อมั่น

- ถ้า $n < 30$ ช่วงความเชื่อมั่น 100(1 –α) % ของค่าเฉลี่ยประชากร μ แบบสองด้านหาได้จาก

$$\boxed{\bar{x} - t_{\frac{\alpha}{2}, n-1} s/\sqrt{n} \le \mu \le \bar{x} + t_{\frac{\alpha}{2}, n-1} s/\sqrt{n}}$$

สำหรับช่วงความเชื่อมั่น 100(1–α)% ของค่าเฉลี่ยประชากร μ แบบด้านเดียว หาได้จาก $\mu \le \bar{x} + t_{\alpha, n-1} s/\sqrt{n}$ สำหรับด้านบนด้านเดียว และ $\bar{x} - t_{\alpha, n-1} s/\sqrt{n} \le \mu$ สำหรับด้านล่างด้านเดียว

1.2.2) เมื่อมีขนาดตัวอย่างใหญ่

- ถ้าขนาดตัวอย่าง n มีค่าตั้งแต่ 30 ขึ้นไป ($n \ge 30$) จะใช้ตัวสถิติ Z ในการหาช่วงความเชื่อมั่น

- ในกรณี $n \ge 30$ ช่วงความเชื่อมั่น 100(1 –α) % ของค่าเฉลี่ยประชากร μ แบบสองด้านหาได้จาก

$$\boxed{\bar{x} + z_{\frac{\alpha}{2}} s/\sqrt{n} \le \mu \le \bar{x} + z_{1-\frac{\alpha}{2}} s/\sqrt{n}}$$

2) การประมาณค่าผลต่างของค่าเฉลี่ยของประชากร 2 กลุ่ม ที่เป็นอิสระต่อกัน

2.1) การประมาณค่าผลต่างของค่าเฉลี่ยของประชากร เมื่อทราบค่าความแปรปรวน

- ใช้ตัวสถิติ Z ในการสร้างช่วงความเชื่อมั่น

- ให้ $\overline{x}_1$ เป็นค่าเฉลี่ยของตัวอย่างขนาด n_1 ที่สุ่มมาจากประชากรที่ 1 ที่มีการแจกแจงแบบปกติและมีค่าความแปรปรวน σ_1^2 และให้ $\overline{x}_2$ เป็นค่าเฉลี่ยของตัวอย่างขนาด n_2 ที่สุ่มมาจากประชากรที่ 2 ที่มีการแจกแจงแบบปกติและมีค่าความแปรปรวน σ_2^2

ดังนั้น ช่วงความเชื่อมั่น 100(1–α)% ของผลต่างของค่าเฉลี่ยของสองประชากร $\mu_1 - \mu_2$ แบบสองด้าน หาได้จาก

$$\boxed{(\overline{x}_1 - \overline{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\overline{x}_1 - \overline{x}_2) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- สำหรับช่วงความเชื่อมั่น 100(1–α) % ของผลต่างของค่าเฉลี่ยของสองประชากร $\mu_1 - \mu_2$ แบบด้านเดียว หาได้จาก

$$\mu_1 - \mu_2 \le (\overline{x}_1 - \overline{x}_2) + z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ สำหรับด้านบนด้านเดียว

$$(\overline{x}_1 - \overline{x}_2) + z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2$$ สำหรับด้านล่างด้านเดียว

2.2) การประมาณค่าผลต่างของค่าเฉลี่ยของประชากร เมื่อไม่ทราบค่าความแปรปรวน

- ในกรณีที่ขนาดตัวอย่าง n_1 และ n_2 มีค่าตั้งแต่ 30 ขึ้นไป ($n_1 \ge 30$ และ $n_2 \ge 30$) จะใช้ตัวสถิติ Z ในการสร้างช่วงความเชื่อมั่นและใช้สูตรในข้อ 2.1) โดยแทนความแปรปรวนประชากร σ_1^2 และ σ_2^2 ด้วยความแปรปรวนตัวอย่าง s_1^2 และ s_2^2

- ถ้าขนาดตัวอย่าง n_1 และ n_2 มีค่าน้อยกว่า 30 ($n_1 < 30$ และ $n_2 < 30$) จะใช้ตัวสถิติ t ในการสร้างช่วงความเชื่อมั่น โดยแบ่งออกเป็นสองกรณี ดังนี้

2.2.1) เมื่อทราบว่า $\sigma_1^2 = \sigma_2^2$

- ใช้ความแปรปรวนร่วม s_p^2 แทนความแปรปรวนตัวอย่าง s_1^2 และ s_2^2 ซึ่งสามารถหาค่า s_p^2 ได้จากสูตร

$$\boxed{s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- ช่วงความเชื่อมั่น 100(1–α)% ของผลต่างของค่าเฉลี่ยของสองประชากร $\mu_1 - \mu_2$ แบบสองด้าน เมื่อทราบว่า $\sigma_1^2 = \sigma_2^2$ หาได้จาก

$$(\overline{x}_1 - \overline{x}_2) - t_{\frac{\alpha}{2}, n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le (\overline{x}_1 - \overline{x}_2) + t_{\frac{\alpha}{2}, n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

- สำหรับช่วงความเชื่อมั่น 100(1–α) % ของผลต่างของค่าเฉลี่ยของสองประชากร $\mu_1 - \mu_2$ แบบด้านเดียว เมื่อทราบว่า $\sigma_1^2 = \sigma_2^2$ หาได้จาก

$$\mu_1 - \mu_2 \le (\overline{x}_1 - \overline{x}_2) + t_{\alpha, n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

สำหรับด้านบนด้านเดียว

$$(\overline{x}_1 - \overline{x}_2) - t_{\alpha, n_1+n_2-2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2$$

สำหรับด้านล่างด้านเดียว

2.2.2) เมื่อไม่ทราบว่า $\sigma_1^2 \neq \sigma_2^2$

- กรณีนี้จะใช้ค่าความแปรปรวนตัวอย่าง s_1^2 และ s_2^2 ในการคำนวณ

- ช่วงความเชื่อมั่น 100(1–α)% ของผลต่างของค่าเฉลี่ยของสองประชากร $\mu_1 - \mu_2$ แบบสองด้าน เมื่อรู้ว่า $\sigma_1^2 \neq \sigma_2^2$ หาได้จาก

$$(\overline{x}_1 - \overline{x}_2) - t_{\frac{\alpha}{2}, \upsilon}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\overline{x}_1 - \overline{x}_2) + t_{\frac{\alpha}{2}, \upsilon}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

โดยที่

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1+1}+\frac{(s_2^2/n_2)^2}{n_2+1}} - 2$$

3) การประมาณค่าผลต่างของค่าเฉลี่ยของประชากร ที่ไม่เป็นอิสระต่อกัน

- การเก็บข้อมูลของตัวอย่างจากประชากรทั้งสองประชากรจะกระทำภายใต้สภาวะและเงื่อนไขเดียวกัน โดยมีขนาดตัวอย่างเท่ากัน

- มีการเก็บข้อมูลเป็นคู่ $(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})$ โดยที่

x_{ij} หมายถึงค่าของข้อมูลจากประชากรที่ i ของตัวอย่างที่ j

$d_1 = x_{11} - x_{21}$ = ผลต่างของข้อมูลคู่ที่ 1

$d_2 = x_{12} - x_{22}$ = ผลต่างของข้อมูลคู่ที่ 2 และ

$d_j = x_{1j} - x_{2j}$ = ผลต่างของข้อมูลคู่ที่ j ซึ่ง j = 1, 2, … , n

- สมมติให้ผลต่างของข้อมูลมีการแจกแจงแบบปกติด้วยค่าเฉลี่ย μ_d และค่าความแปรปรวน σ_d^2 ซึ่ง $\mu_d = \mu_1 - \mu_2$ และผลต่างแต่ละคู่เป็นอิสระจากกัน

- ถ้าขนาดตัวอย่างหรือจำนวนคู่ของข้อมูลน้อยกว่า 30 ($n < 30$) ใช้ตัวสถิติ t ในการสร้างช่วงความเชื่อมั่น โดยที่ df = n – 1

- ถ้า $n \geq 30$ ใช้ตัวสถิติ Z ในการสร้างช่วงความเชื่อมั่น

- กำหนดให้ $\bar{d}$ เป็นค่าเฉลี่ยตัวอย่างของผลต่างของข้อมูล n คู่ และ s_d เป็นค่าเบี่ยงเบนมาตรฐานของผลต่างของข้อมูล n คู่ ดังนั้น ช่วงความเชื่อมั่น $100(1-\alpha)\%$ ของค่าเฉลี่ยผลต่าง μ_d แบบสองด้าน หาได้จาก

$$\boxed{\bar{d} - t_{\frac{\alpha}{2}, n-1} \, s_d/\sqrt{n} \;\leq\; \mu_d \;\leq\; \bar{d} + t_{\frac{\alpha}{2}, n-1} \, s_d/\sqrt{n}}$$

3.6.2 การประมาณค่าสัดส่วนของประชากร

1) การประมาณค่าสัดส่วนของประชากร 1 กลุ่ม

- ตัวแปรสุ่มที่ศึกษา มีการแจกแจงแบบทวินาม (binomial distribution)

- ใช้การแจกแจงแบบปกติมาประมาณการแจกแจงแบบทวินามได้เมื่อขนาดตัวอย่าง n มีขนาดใหญ่ หรือ

$np > 5$ และ $nq > 5$ โดยที่ $q = 1 - p$

p เป็นสัดส่วนประชากรที่เกิดสิ่งที่สนใจศึกษา

q เป็นสัดส่วนประชากรที่เกิดสิ่งที่ไม่สนใจศึกษา

- กำหนดให้ $\hat{p}$ เป็นสัดส่วนตัวอย่างขนาด n ของเหตุการณ์ที่สนใจของประชากรที่มีการแจกแจงแบบทวินาม และ $\hat{q} = 1 - \hat{p}$ ดังนั้น ช่วงความเชื่อมั่น $100(1-\alpha)\%$ ของสัดส่วนประชากร p แบบสองด้าน หาได้จาก

$$\boxed{\hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \;\leq\; p \;\leq\; \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

จาก $P(z_{\frac{\alpha}{2}} \leq z \leq z_{1-\frac{\alpha}{2}}) = 1-\alpha$ และ $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$

จะได้ $P\left(z_{\frac{\alpha}{2}} \leq \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} \leq z_{1-\frac{\alpha}{2}} \right) = 1-\alpha$

$$P\left(\hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} \;\leq\; p \;\leq\; \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1-\alpha$$

เนื่องจากไม่ทราบค่า p และ q จึงใช้ $\hat{p}$ และ $\hat{q}$ แทน ซึ่งจะได้

$$P(\hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}) = 1-\alpha$$

- สำหรับช่วงความเชื่อมั่น 100(1−α) % ของสัดส่วนประชากร p แบบด้านเดียว หาได้จาก

$$p \le \hat{p} + z_{1-\alpha}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$ สำหรับด้านบนด้านเดียว

$$\hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p$$ สำหรับด้านล่างด้านเดียว

2) การประมาณค่าผลต่างของค่าสัดส่วนประชากร 2 กลุ่ม

- ในกรณีที่มีประชากรสองประชากรที่มีการแจกแจงแบบทวินาม (binomial distribution) เราสามารถใช้การแจกแจงแบบปกติมาประมาณการแจกแจงแบบทวินามได้เมื่อขนาดตัวอย่าง n_1 และ n_2 มีขนาดใหญ่ หรือ

$n_1p_1 > 5$, $n_1q_1 > 5$ และ $n_2p_2 > 5$, $n_2q_2 > 5$

- กำหนดให้ $\hat{p}_1$ และ $\hat{p}_2$ เป็นสัดส่วนตัวอย่างขนาด n_1 และ n_2 ของเหตุการณ์ที่สนใจที่มาจากประชากรที่ 1 และ2 ตามลำดับและประชากรทั้งสองมีการแจกแจงแบบทวินาม และ $\hat{q}_1 = 1 - \hat{p}_1$ และ $\hat{q}_2 = 1 - \hat{p}_2$ ดังนั้นช่วงความเชื่อมั่น 100(1−α)% ของผลต่างของสัดส่วนสองประชากร $p_1 - p_2$ แบบสองด้าน หาได้จาก

$$\boxed{(\hat{p}_1 - \hat{p}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

- สำหรับช่วงความเชื่อมั่น 100(1−α)% ของผลต่างของสัดส่วนสองประชากร แบบด้านเดียว หาได้จาก

$$p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z_{1-\alpha}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$ สำหรับด้านบนด้านเดียว

$$(\hat{p}_1 - \hat{p}_2) + z_{\alpha}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \le p_1 - p_2$$ สำหรับด้านล่างด้านเดียว

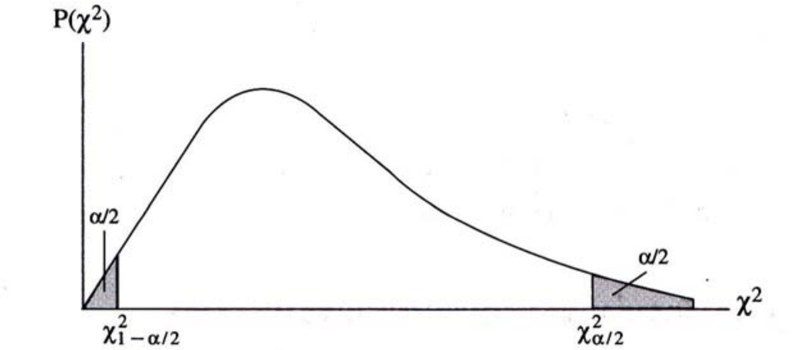
3.6.3 การประมาณค่าความแปรปรวนของประชากร

1) การประมาณค่าความแปรปรวนของประชากร 1 กลุ่ม

- ตัวสถิติที่ใช้สร้างความเชื่อมั่นคือ ไคสแควร์ χ^2

- กำหนดให้ s^2 เป็นค่าความแปรปรวนของตัวอย่างที่มีขนาด n ซึ่งมาจากประชากรที่มีการแจกแจงแบบปกติ ดังนั้นช่วงความเชื่อมั่น 100(1–α)% ของความแปรปรวนของประชากรเดียว σ^2 แบบสองด้าน หาได้จาก

$$\boxed{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},\, n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},\, n-1}}} \qquad df = n - 1$$



รูปที่ 28 แสดงค่า P($\chi^2_{1-\frac{\alpha}{2},\, n-1} \le \chi^2 \le \chi^2_{\frac{\alpha}{2},\, n-1}$) = $1-\alpha$

(พิศมัย หาญมงคลพิพัฒน์, 2550: 161)

$$\mathbf{P}(\chi^2_{1-\frac{\alpha}{2},\, n-1} \le \chi^2 \le \chi^2_{\frac{\alpha}{2},\, n-1}) = 1-\alpha$$

จาก $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$ จะได้

$$P(\chi^2_{1-\frac{\alpha}{2},\, n-1} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},\, n-1}) = 1-\alpha$$

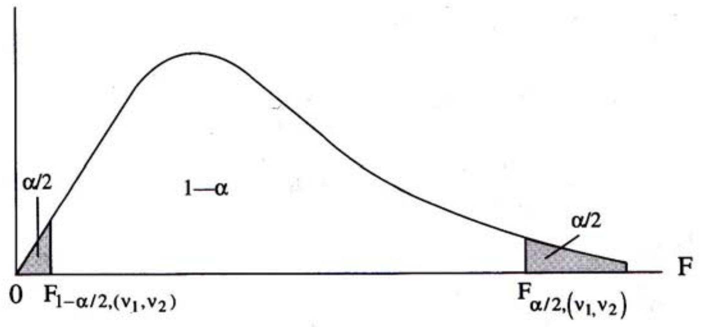
ดังนั้น $P(\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},\, n-1}} \le \sigma^2 \le \dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},\, n-1}}) = 1-\alpha$

- สำหรับช่วงความเชื่อมั่น 100(1–α) % ของความแปรปรวนของประชากรเดียว σ^2 แบบด้านเดียว หาได้จาก

$$\sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha,\, n-1}}$$ สำหรับด้านบนด้านเดียว

$$\frac{(n-1)s^2}{\chi^2_{\alpha,\, n-1}} \le \sigma^2$$ สำหรับด้านล่างด้านเดียว

2) การประมาณค่าอัตราส่วนของความแปรปรวนของประชากร 2 กลุ่ม



รูปที่ 29 แสดงค่า $P(F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} \leq F \leq F_{\frac{\alpha}{2}, n_2-1, n_1-1}) = 1-\alpha$

(พิศมัย หาญมงคลพิพัฒน์, 2550: 163)

- ใช้ตัวสถิติ F ในการสร้างช่วงความเชื่อมั่น

- กำหนดให้ s_1^2 เป็นค่าความแปรปรวนของตัวอย่างที่มีขนาด n_1 ซึ่งมาจากประชากรที่ 1 และ s_2^2 เป็นค่าความแปรปรวนของตัวอย่างที่มีขนาด n_2 ซึ่งมาจากประชากรที่ 2 โดยทั้งสองประชากรมีการแจกแจงแบบปกติ ดังนั้นช่วงความเชื่อมั่น 100(1−α)% ของสัดส่วนของความแปรปรวนสองประชากร $\frac{\sigma_1^2}{\sigma_2^2}$ แบบสองด้าน หาได้จาก

$$\boxed{\frac{s_1^2}{s_2^2} \cdot F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot F_{\frac{\alpha}{2}, n_2-1, n_1-1}}$$

โดยที่ $df = n_2 - 1, n_1 - 1$ และ $F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} = \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}}$

$$P(F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} \leq F \leq F_{\frac{\alpha}{2}, n_2-1, n_1-1}) = 1-\alpha$$

จาก $F = \dfrac{\frac{S_2^2}{\sigma_2^2}}{\frac{S_1^2}{\sigma_1^2}}$, $df = n_2 - 1, n_1 - 1$

ดังนั้น $P(F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} \leq \dfrac{\frac{S_2^2}{\sigma_2^2}}{\frac{S_1^2}{\sigma_1^2}} \leq F_{\frac{\alpha}{2}, n_2-1, n_1-1}) = 1-\alpha$

$$P(\frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1}) = 1-\alpha$$

- สำหรับช่วงความเชื่อมั่น 100(1–α) % ของสัดส่วนของความแปรปรวนสองประชากร $\frac{\sigma_1^2}{\sigma_2^2}$ แบบด้านเดียว หาได้จาก

$$\frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} F_{\alpha,\, n_2-1,\, n_1-1}$$ สำหรับด้านบนด้านเดียว

$$\frac{s_1^2}{s_2^2} F_{1-\alpha,\, n_2-1,\, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2}$$ สำหรับด้านล่างด้านเดียว

3.7 หน่วยที่ 7 การทดสอบสมมติฐาน

3.7.1 การทดสอบสมมติฐานทางสถิติที่เกี่ยวกับค่าเฉลี่ยของประชากร

1) การทดสอบเกี่ยวกับค่าเฉลี่ยประชากร 1 กลุ่ม (One Sample t-test)

การทดสอบเกี่ยวกับค่าเฉลี่ยประชากร 1 กลุ่ม เป็นการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรกลุ่มเดียว ว่ามีความแตกต่างจากค่าเฉลี่ยที่ได้กำหนดไว้ในสมมติฐานหลัก H_0 หรือไม่ โดยมีระดับนัยสำคัญ (α) ตามที่ได้กำหนด[1] ถ้าผลลัพธ์ที่ได้จากการวิเคราะห์มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ จะปฏิเสธสมมติฐานหลัก H_0 ที่ตั้งไว้

กรณีประชากรมีการแจกแจงแบบปกติ และทราบค่าความแปรปรวนของประชากร (σ^2) ประชากรมีการแจกแจงแบบใด ๆ ขนาดตัวอย่างใหญ่ ($n \ge 30$) และทราบค่า σ^2

สมมติฐานว่าง $H_0 : \mu = \mu_0$ โดยที่ μ_0 เป็นค่าคงที่

สถิติทดสอบ $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ ถ้าไม่ทราบค่า σ ใช้ค่า S แทน

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
H_1: $\mu > \mu_0$	$Z > Z_{\alpha}$
H_1: $\mu < \mu_0$	$Z < -Z_{\alpha}$
H_1: $\mu \ne \mu_0$	$\lvert Z \rvert > Z_{\alpha/2}$

กรณีประชากรมีการแจกแจงแบบปกติ ขนาดตัวอย่างเล็ก ($n < 30$) และไม่ทราบค่า σ 2

สมมติฐานว่าง $H_0 : \mu = \mu_0$ โดยที่ μ_0 เป็นค่าคงที่

สถิติทดสอบ $t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$ โดยที่การแจกแจง t มี df = n-1

<u>สมมติฐานแย้ง</u> <u>เขตปฏิเสธ H_0</u>

---

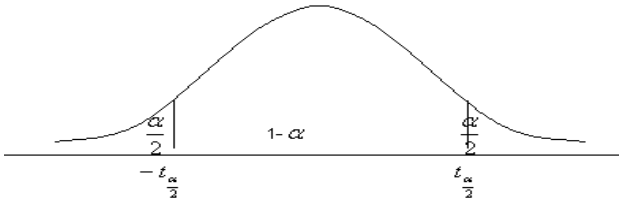
[1] การกำหนดเขตปฏิเสธด้วยสถิติทดสอบสมมติฐานด้วยระดับนัยสำคัญ α หรือ 1-α นั้นมีความหมายหรือค่าเดียวกัน โดยค่า α จะเปิดตารางที่ α = .01, .05 แต่ค่า 1-α จะเปิดตารางที่ .90, .95 ในรายงานการวิจัยนี้ได้ใช้ระดับนัยสำคัญ α เปิดตารางที่ α = .01, .05, .005, .025

H$_1$: $\mu > \mu_0$ $\quad$ $t > t_\alpha$

H$_1$: $\mu < \mu_0$ $\quad$ $t < -t_\alpha$

H$_1$: $\mu \neq \mu_0$ $\quad$ $|t| > t_{\alpha/2}$

การหาช่วงความเชื่อมั่นของ μ อาศัยการแจกแจงแบบ t ที่มี df = n-1 และเมื่อกำหนด (1 - α) 100% จะหาค่า $t_{\alpha/2}$ จากตาราง t ได้



และ $P\left(\bar{x} - t_{\frac{\alpha}{2}} S/\sqrt{n} < \mu < \bar{x} + t_{\frac{\alpha}{2}} S/\sqrt{n}\right) = 1 - \alpha$

รูปที่ 30 ช่วงความเชื่อมั่นของ μ อาศัยการแจกแจงแบบ t

(http://www.sut.ac.th/e-texts/Social/ProjectCAI/)

2) การทดสอบความแตกต่างของค่าเฉลี่ยประชากร 2 กลุ่มที่เป็นอิสระต่อกัน (Independent Samples t-test)

การทดสอบความแตกต่างของค่าเฉลี่ยประชากร 2 กลุ่มที่เป็นอิสระต่อกัน หมายถึงตัวแปรที่ต้องการทราบถึงความแตกต่างของค่าเฉลี่ยโดยแบ่งตัวแปรนั้นออกเป็น 2 กลุ่มตามที่ได้กำหนดและทำการทดสอบถึงความแตกต่างของค่าเฉลี่ยว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่ โดยมีระดับนัยสำคัญ (α) ตามที่ได้กำหนด

กรณีประชากรมีการแจกแจงแบบปกติ และทราบค่าความแปรปรวนของประชากรทั้งสอง (σ^2) ประชากรมีการแจกแจงแบบใด ๆ ขนาดตัวอย่างใหญ่ $(n_1 \geq 30, n_2 \geq 30)$

สมมติฐานว่าง $\quad H_0 : \mu_1 - \mu_2 = d_0 \quad$ โดยที่ d_0 เป็นค่าคงที่

สถิติทดสอบ $\quad Z = \dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \quad$ ถ้าไม่ทราบค่า σ ใช้ค่า S แทน

สมมติฐานแย้ง	เขตปฏิเสธ H$_0$
H$_1$: $\mu_1 - \mu_2 > d_0$	$Z > Z_\alpha$
H$_1$: $\mu_1 - \mu_2 < d_0$	$Z < -Z_\alpha$
H$_1$: $\mu_1 - \mu_2 \neq d_0$	$\lvert Z \rvert > Z_{\alpha/2}$

กรณีประชากรมีการแจกแจงแบบปกติ ขนาดตัวอย่างเล็ก $(n_1 < 30, n_2 < 30)$ และไม่ทราบค่า σ^2

สมมติฐานว่าง $H_0 : \mu_1 - \mu_2 = d_0$ โดยที่ d_0 เป็นค่าคงที่

ถ้า $\sigma_1^2 = \sigma_2^2$

สถิติทดสอบ $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - d_0}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ โดยที่การแจกแจง t มี $df = n_1 + n_2 - 2$

เมื่อ $S_p = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
H_1: $\mu_1 - \mu_2 > d_0$	$t > t_\alpha$
H_1: $\mu_1 - \mu_2 < d_0$	$t < -t_\alpha$
H_1: $\mu_1 - \mu_2 \neq d_0$	$\lvert t \rvert > t_{\alpha/2}$

การหาช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$ อาศัยการแจกแจงแบบ t ที่มี $df = n_1 + n_2 - 2$ และกำหนด $(1 - \alpha)$ 100% หาค่า $t_{\alpha/2}$ จากตาราง t ได้ช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$ ดังนี้

$$P\left\{(\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < +t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right\} = 1 - \alpha$$

หรือได้ว่า กำหนด $(1 - \alpha)$ 100% ช่วงความเชื่อมั่นของ $\mu 1 - \mu 2$ คือ

$$(\overline{X}_1 - \overline{X}_2) - t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < +t_{\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1 - \alpha$$

ถ้า $\sigma_1^2 \neq \sigma_2^2$

สถิติทดสอบ $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - d_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ โดยที่การแจกแจง t มี $df = \nu$

เมื่อ $\nu = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$ ค่า ν ที่ได้ปัดเป็นเลขจำนวนเต็ม

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
H_1: $\mu_1 - \mu_2 > d_0$	$t > t_\alpha$
H_1: $\mu_1 - \mu_2 < d_0$	$t < -t_\alpha$
H_1: $\mu_1 - \mu_2 \neq d_0$	$\lvert t \rvert > t_{\alpha/2}$

การหาช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$ อาศัยการแจกแจงแบบ t ที่มี $df = \nu$ และกำหนด $(1 - \alpha)$ 100% หาค่า $t_{\alpha/2}$ จากตาราง t ได้ช่วงความเชื่อมั่นของ $\mu_1 - \mu_2$ ดังนี้

$$(\overline{X}_1-\overline{X}_2)-t_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<(\overline{X}_1-\overline{X}_2)+t_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

3) การทดสอบความแตกต่างของค่าเฉลี่ยประชากร 2 กลุ่มที่ไม่เป็นอิสระต่อกัน (Paired Samples t-test)

การทดสอบความแตกต่างของค่าเฉลี่ยประชากร 2 กลุ่มที่ไม่เป็นอิสระต่อกัน หมายถึง ตัวแปรที่ต้องการทราบถึงความแตกต่างของค่าเฉลี่ยที่เป็นตัวแปรตัวเดียวกัน หน่วยตัวอย่างเดียวกันแต่มีการเก็บรวบรวมข้อมูล 2 ครั้งและทดสอบหาความแตกต่างระหว่างครั้งแรกกับครั้งหลัง หรืออีกลักษณะหนึ่ง ก็คือตัวแปรที่เป็นลักษณะจับคู่ โดยมีข้อจำกัดว่าหน่วยตัวอย่างต้องเป็นหน่วยตัวอย่างเดียวกัน เช่น คะแนนสอบก่อน-หลังเรียน ของนักศึกษาชั้น ปวช. 1 สาขางานยานยนต์ โดยเก็บรวบรวมข้อมูลคะแนนสอบ มาทดสอบสมมติฐานว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่ โดยมีระดับนัยสำคัญ (α) ตามที่ได้กำหนด

สมมติฐานว่าง $H_0:\mu_1-\mu_2=d_0$ หรือ $H_0:\mu_d=d_0$ โดยที่ $\mu_1-\mu_2=\mu_0$

สถิติทดสอบ $t=\dfrac{\overline{d}-d_0}{S_d/\sqrt{n}}$ โดยที่การแจกแจง t มี $df=n-1$

เมื่อ $\overline{d}=\dfrac{\sum_{i=1}^{n}d_i}{n}$ และ $S_d=\sqrt{\dfrac{\sum_{i=1}^{n}(d_i-\overline{d})^2}{n-1}}$

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
H_1: $\mu_d > d_0$	$t > t_\alpha$
H_1: $\mu_d < d_0$	$t < -t_\alpha$
H_1: $\mu_d \neq d_0$	$\lvert t\rvert > t_{\alpha/2}$

การหาช่วงความเชื่อมั่นของ μ_d อาศัยการแจกแจงแบบ t ที่มี df = n - 1 และกำหนด (1 - α) 100% หาค่า $t_{\alpha/2}$ จากตาราง t ได้ช่วงความเชื่อมั่นของ μ_d ดังนี้

$$\overline{d}-t_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}}<\mu_d<\overline{d}+t_{\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}}$$

3.7.2 การทดสอบสมมติฐานทางสถิติที่เกี่ยวกับค่าสัดส่วนของประชากร

1) การทดสอบเกี่ยวกับค่าสัดส่วนประชากร 1 กลุ่ม

ในกรณีตัวแปรสุ่ม X ที่ศึกษา มีการแจกแจงแบบทวินาม (binomial distribution) เราสามารถใช้การแจกแจงแบบปกติมาประมาณการแจกแจงแบบทวินามได้เมื่อขนาดตัวอย่าง n มีขนาดใหญ่ หรือ

$np > 5$ และ $nq > 5$ โดยที่ $q = 1-p$

p เป็นสัดส่วนประชากรที่เกิดสิ่งที่สนใจศึกษา

q เป็นสัดส่วนประชากรที่เกิดสิ่งที่ไม่สนใจศึกษา

ดังนั้นในการทดสอบสมมติฐานเกี่ยวกับสัดส่วนประชากรที่มีการแจกแจงแบบทวินาม ถ้าจำนวนตัวอย่างที่สุ่ม n มากเพียงพอ และทั้ง $np > 5$ และ $nq > 5$ เราสามารถใช้ตัวทดสอบสถิติ Z ได้ โดยที่

$$Z = \frac{x - np}{\sqrt{npq}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \qquad \text{โดยที่} \quad \hat{p} = \frac{x}{n}$$

เมื่อ x คือจำนวนตัวอย่างที่เกิดสิ่งที่สนใจศึกษาจากตัวอย่างที่สุ่มมาทั้งหมด n

สมมติฐานว่าง $H_0 : p = p_0$ โดยที่ p_0 เป็นค่าคงที่

สถิติทดสอบ $Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
$H_1 : p > p_0$	$Z > Z_\alpha$
$H_1 : p < p_0$	$Z < -Z_\alpha$
$H_1 : p \neq p_0$	$\lvert Z \rvert > Z_{\alpha/2}$

การหาช่วงความเชื่อมั่นของ p อาศัยการแจกแจงแบบ Z ที่ระดับ $(1 - \alpha)$ 100% หาค่า $Z_{\alpha/2}$ จากตาราง Z ได้ช่วงความเชื่อมั่นของ p ดังนี้

$$P\left(\hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 1 - \alpha$$

2) การทดสอบเกี่ยวกับผลต่างของค่าสัดส่วนประชากร 2 กลุ่ม

ถ้าตัวแปรสุ่มที่ศึกษามาจากประชากรสองประชากรที่มีการแจกแจงแบบทวินาม (binomial distribution) เราสามารถใช้การแจกแจงแบบปกติมาประมาณการแจกแจงแบบทวินามได้เช่นกัน เมื่อขนาดตัวอย่าง n_1 และ n_2 มีขนาดใหญ่ หรือ

$n_1p_1 > 5$, $n_1q_1 > 5$ และ $n_2p_2 > 5$, $n_2q_2 > 5$

ในกรณีนี้การทดสอบสมมติฐานเกี่ยวกับผลต่างของสัดส่วนประชากรของสองประชากร ยังคงใช้ตัวทดสอบสถิติ Z และ

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

ซึ่ง $\hat{p}_1 = \dfrac{x_1}{n_1}$ $\quad \hat{p}_2 = \dfrac{x_2}{n_2}$ $\quad q_1 = 1 - p_1$ $\quad q_2 = 1 - p_2$

โดยที่ $\hat{p}_1$ และ $\hat{p}_2$ เป็นสัดส่วนตัวอย่างที่มาจากประชากรที่ 1 และ 2 ตามลำดับ

x_1 คือจำนวนตัวอย่างที่เกิดสิ่งที่สนใจศึกษาจากตัวอย่างที่สุ่มมาทั้งหมด n_1 ซึ่งมาจากประชากรที่ 1

x_2 คือจำนวนตัวอย่างที่เกิดสิ่งที่สนใจศึกษาจากตัวอย่างที่สุ่มมาทั้งหมด n_2 ซึ่งมาจากประชากรที่ 2

เนื่องจากเราไม่ทราบค่าที่แท้จริงของสัดส่วนประชากร p_1 และ p_2 ดังนั้นจึงใช้สัดส่วนตัวอย่างมาประมาณค่าแทน นั่นคือ ใช้ $\hat{p}_1$ และ $\hat{q}_1$ มาประมาณค่าแทน p_1 และ q_1 ตามลำดับ และใช้ $\hat{p}_2$ และ $\hat{q}_2$ มาประมาณค่าแทน p_2 และ q_2 ตามลำดับ ดังนั้น สามารถหาค่า Z จาก

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

2.1) การทดสอบเกี่ยวกับผลต่างของค่าสัดส่วนประชากร 2 กลุ่ม กรณี $d_0 = 0$

สมมติฐานว่าง $H_0 : p_1 - p_2 = d_0 = 0$ โดยที่ $d_0 = 0$

สถิติทดสอบ $$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
$H_1 : p_1 > p_2$	$Z > Z_\alpha$
$H_1 : p_1 < p_2$	$Z < -Z_\alpha$
$H_1 : p_1 \neq p_2$	$\lvert Z \rvert > Z_{\alpha/2}$

2.2) การทดสอบเกี่ยวกับผลต่างของค่าสัดส่วนประชากร 2 กลุ่ม กรณี $d_0 \neq 0$

สมมติฐานว่าง $H_0 : p_1 - p_2 = d_0$ โดยที่ d_0 เป็นค่าคงที่

สถิติทดสอบ $$Z = \frac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
$H_1 : p_1 - p_2 > d_0$	$Z > Z_\alpha$
$H_1 : p_1 - p_2 < d_0$	$Z < -Z_\alpha$
$H_1 : p_1 - p_2 \neq d_0$	$\lvert Z \rvert > Z_{\alpha/2}$

การหาช่วงความเชื่อมั่น 100(1–α)% ของผลต่างของสัดส่วนสองประชากร $p_1 - p_2$ ได้ดังนี้

$$(\hat{p}_1 - \hat{p}_2) + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

3.7.3 การทดสอบสมมติฐานทางสถิติที่เกี่ยวกับความแปรปรวนของประชากร

1) การทดสอบเกี่ยวกับความแปรปรวนของประชากร 1 กลุ่ม

สมมติประชากรมีการแจกแจงแบบปกติ มีค่าความแปรปรวน σ^2 และ s^2 เป็นค่าความแปรปรวนของตัวอย่างที่มีขนาด n

สมมติฐานว่าง $H_0 : \sigma_1^2 = \sigma_0^2$ โดยที่ σ_0^2 เป็นค่าคงที่

สถิติทดสอบ $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$, df = n – 1

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
$H_1 : \sigma_1^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha,(n-1)}$
$H_1 : \sigma_1^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha,(n-1)}$
$H_1 : \sigma_1^2 \neq \sigma_0^2$	$\chi^2 > \chi^2_{\alpha/2,(n-1)}$

2) การทดสอบเกี่ยวกับความแปรปรวนของประชากร 2 กลุ่ม

สมมติประชากรทั้งสองประชากรมีการแจกแจงแบบปกติ ประชากรที่ 1 มีความแปรปรวน σ_1^2 และประชากรที่ 2 มีความแปรปรวน σ_2^2

ให้ s_1^2 เป็นความแปรปรวนตัวอย่างที่มีขนาดตัวอย่าง n_1 ซึ่งสุ่มมาจากประชากรที่ 1 และ s_2^2 เป็นความแปรปรวนตัวอย่างที่มีขนาดตัวอย่าง n_2 ซึ่งสุ่มมาจากประชากรที่ 2

สมมติฐานว่าง

สมมติฐานว่าง $H_0 : \sigma_1^2 = \sigma_2^2$

สถิติทดสอบ $F = \dfrac{s_1^2}{s_2^2}$, $df = n_1 - 1, n_2 - 1$

และ $$F_{1-\alpha,(n_1-1,n_2-1)} = \frac{1}{F_{\alpha,(n_2-1,n_1-1)}}$$

<u>สมมติฐานแย้ง</u>	<u>เขตปฏิเสธ H_0</u>
$H_1 : \sigma_1^2 > \sigma_2^2$	$F > F_{\alpha,(n_1-1,n_2-1)}$
$H_1 : \sigma_1^2 < \sigma_2^2$	$F < F_{1-\alpha,(n_1-1,n_2-1)}$
$H_1 : \sigma_1^2 \neq \sigma_2^2$	$F > F_{\alpha/2,(n_1-1,n_2-1)}$

การหาช่วงความเชื่อมั่น 100(1–α)% ของอัตราส่วนของความแปรปรวนสองประชากรที่มีการแจกแจงแบบปกติ ได้ดังนี้

$$\frac{s_1^2}{s_2^2} \cdot F_{1-\frac{\alpha}{2},\, n_1-1,\, n_2-1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot F_{\frac{\alpha}{2},\, n_1-1,\, n_2-1}$$

3.8 หน่วยที่ 8 การวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) หรือ การทดสอบความแตกต่างของค่าเฉลี่ยประชากรมากกว่า 2 กลุ่ม โดยปกติถ้าตัวแปรที่ต้องการทดสอบความแตกต่างของค่าเฉลี่ยมีมากกว่า 2 กลุ่ม ผู้วิจัยจะใช้การทดสอบสมมติฐานพร้อมกันเพียงครั้งเดียว โดยไม่ต้องทำการทดสอบความแตกต่างของค่าเฉลี่ยทีละคู่ ทำให้ประหยัดเวลาในการทดสอบสมมติฐาน ซึ่งเรียกการวิเคราะห์นี้ว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) โดยมีข้อจำกัดที่ว่า ประชากรมีการแจกแจงแบบปกติ ตัวอย่างที่สุ่มมาต้องมีความเป็นอิสระต่อกัน และมีความแปรปรวนเท่ากัน เช่น ต้องการทราบความแตกต่างของวิธีสอนแบบต่าง ๆ ว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่ โดยมีระดับนัยสำคัญ (α) ตามที่ผู้วิจัยกำหนด ซึ่งมีข้อตกลงเบื้องต้น ดังนี้

1. กลุ่มตัวอย่างแต่ละกลุ่มต้องสุ่มมาจากประชากรที่มีการแจกแจงเป็นปกติ
2. กลุ่มตัวอย่างแต่ละกลุ่มต้องสุ่มมาจากประชากรที่ความแปรปรวนไม่แตกต่างกัน
3. หน่วยสมาชิกในกลุ่มตัวอย่างแต่ละหน่วยต้องสุ่มมาอย่างอิสระ
4. กลุ่มตัวอย่างแต่ละกลุ่มต้องเป็นอิสระต่อกัน

3.8.1 การวิเคราะห์ความแปรปรวนแบบมีปัจจัยเดียว (Single-Factor Analysis of Variance: One-Way ANOVA)

การวิเคราะห์ความแปรปรวนแบบมีปัจจัยเดียว เป็นสถิติใช้สำหรับวิเคราะห์ความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ 3 ประชากรหรือ 3 กลุ่มขึ้นไป ที่รับปัจจัยที่ต่างระดับกันตั้งแต่ 3 ระดับขึ้นไป โดยจะต้องมีตัวแปรตามมีระดับการวัดอยู่ในระดับ Interval Scale และตัวแปรอิสระมีเพียงตัวแปรเดียวหรือปัจจัยเดียวอยู่ในระดับ Nominal Scale แบ่งออกเป็น k ระดับ

จากข้อตกลงเบื้องต้นกลุ่มตัวอย่างจะต้องสุ่มมาจากประชากรที่มีความแปรปรวนไม่แตกต่างกัน จึงต้องทดสอบด้วยสถิติต่อไปนี้

1) การทดสอบสมมติฐานทางสถิติที่เกี่ยวกับค่าความแปรปรวนของสองประชากร

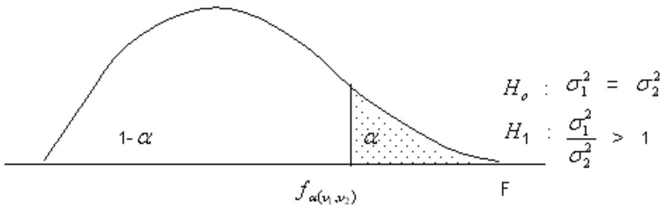
การทดสอบความแตกต่างของค่าความแปรปรวนของสองประชากร มีข้อจำกัดว่าประชากรทั้งสองกลุ่มต้องมีการแจกแจงแบบปกติ ทั้งสองกลุ่มต้องเป็นอิสระจากกัน และไม่ทราบค่า σ^2

สมมติฐานว่าง $H_0 : \sigma_1^2 = \sigma_2^2$

สถิติทดสอบ $F = \dfrac{S_1^2}{S_2^2}$ โดยที่การแจกแจง F มี $df = n_1 - 1, n_2 - 1$

เมื่อ $S_1^2 > S_2^2$ หรือ

สมมติฐานแย้ง $H_1 : \sigma_1^2 > \sigma_2^2$ เขตปฏิเสธ H_0 เมื่อ $F > F_{\alpha; n1-1, n2-1}$



รูปที่ 31 การทดสอบสมมติฐานค่าความแปรปรวนแบบทางเดียวข้างขวา

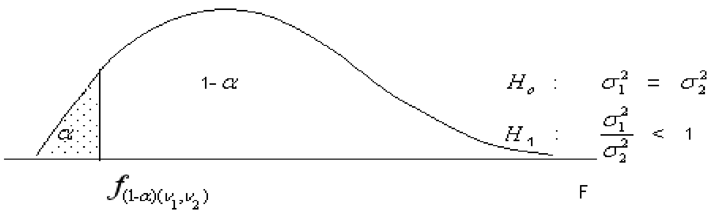
(http://www.sut.ac.th/e-texts/Social/ProjectCAI/)

สมมติฐานว่าง $H_0 : \sigma_1^2 = \sigma_2^2$

สถิติทดสอบ $F = \dfrac{S_2^2}{S_1^2}$ โดยที่การแจกแจง F มี $df = n_2 - 1, n_1 - 1$

เมื่อ $S_2^2 > S_1^2$

สมมติฐานแย้ง $H_1 : \sigma_1^2 < \sigma_2^2$ เขตปฏิเสธ H_0 เมื่อ $F < F_{\alpha;n2-1,\, n1-1}$

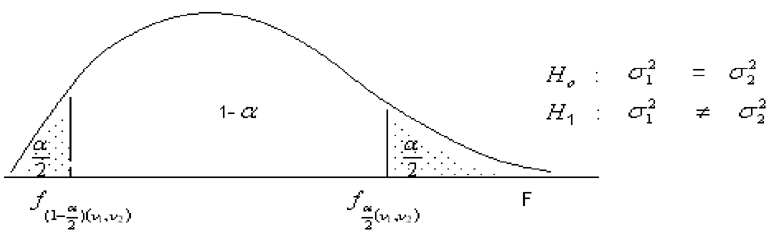


รูปที่ 32 การทดสอบสมมติฐานค่าความแปรปรวนแบบทางเดียวข้างซ้าย

(http://www.sut.ac.th/e-texts/Social/ProjectCAI/)

สมมติฐานว่าง $H_0 : \sigma_1^2 = \sigma_2^2$

สมมติฐานแย้ง $H_1 : \sigma_1^2 \neq \sigma_2^2$ เขตปฏิเสธ H_0 เมื่อ $F > F_{\alpha/2;\, n1-1,\, n2-1}$ หรือ $F > F_{\alpha/2;\, n2-1,\, n1-1}$



รูปที่ 33 การทดสอบสมมติฐานค่าความแปรปรวนแบบสองทาง

(http://www.sut.ac.th/e-texts/Social/ProjectCAI/)

การหาค่าวิกฤตของการแจกแจง F จะมีอยู่ด้วยกัน 2 ค่า คือ F_L และ F_U ซึ่งตารางการแจกแจง F จะมีให้เพียงค่า F_U เท่านั้น โดยที่ค่า F_L ต้องคำนวณโดยอาศัยความสัมพันธ์ $F_L = 1/F_U$ *(degrees of freedom switched)

2) การทดสอบสมมติฐานทางสถิติที่เกี่ยวกับค่าความแปรปรวน k ประชากร

การทดสอบความแตกต่างของค่าความแปรปรวนของประชากรตั้งแต่ 2 ประชากรขึ้นไปเท่ากันหรือไม่ จะใช้สถิติทดสอบของ Bartlett's Test และ Levene's Test โดยแต่ละวิธีมีรายละเอียด ดังนี้

2.1) การทดสอบความเท่ากันของค่าความแปรปรวน k ประชากร โดยใช้สถิติทดสอบของ Bartlett's Test ซึ่งเป็นการทดสอบที่มีประสิทธิภาพ ถ้าแต่ละประชากรมีการแจกแจงแบบปกติ (กัลยา วานิชย์บัญชา, 2550: 145-157)

สมมติฐานว่าง $H_0 : \sigma_{_1}^2 = \sigma_{_2}^2 = ... = \sigma_k^2$

สมมติฐานแย้ง $H_1 : \sigma_{\underset{1}{i}}^2 \neq \sigma_j^2$ อย่างน้อย 1 คู่; $i \neq j$

เขตปฏิเสธ H_0 เมื่อ $B_c > \chi^2_{\alpha}$

สถิติทดสอบ $$B = \left[\ln S_p^2\right]\left[\sum_{i=1}^{k}(n_i - 1)\right] - \sum_{i=1}^{k}(n_i - 1)\ln S_i^2$$

โดยที่

$$i = 1,2,......., k$$

$$S_i^2 = \frac{\sum_{i=1}^{k}\left(X_{ij} - \overline{X}_i\right)}{n_i - 1}$$

$$S_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^2}{\sum_{i=1}^{k}(n_i - 1)}$$

ln = natural logarithm (base e) อาจจะเปลี่ยนเป็น logarithm ฐาน 10 (log) สูตร B จะเปลี่ยนเป็นสูตร B* ดังนี้

$$B^* = 2.30259\left[\left[\log S_p^2\right]\left[\sum_{i=1}^{k}(n_i - 1)\right] - \sum_{i=1}^{k}(n_i - 1)\log S_i^2\right]$$

และ B_c = B*/C ที่มีการแจกแจงแบบไคสแควร์มี df = n-1 โดย C เป็นค่าปรับ (Correction factor) ที่ปรับให้ B* มีการแจกแจงเข้าสู่การแจกแจงแบบไคสแควร์มากขึ้น มีสูตรดังนี้

$$C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^{k} n_i - 1}\right]$$

2.2) การทดสอบความเท่ากันของค่าความแปรปรวน k ประชากร โดยใช้สถิติทดสอบของ Levene's Test ซึ่งเป็นการทดสอบที่ใช้ค่าสัมบูรณ์ (absolute value) เมื่อประชากรมีการแจกแจงแบบหางยาว (long-tailed distribution) หรือมีความเบ้เป็นบวก (positive skewness) (กัลยา วานิชย์บัญชา, 2550: 145-157 และ http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm)

สมมติฐานว่าง $H_0 : \sigma_{_1}^2 = \sigma_{_2}^2 = ... = \sigma_k^2$

สมมติฐานแย้ง $H_1 : \sigma_{\underset{1}{i}}^2 \neq \sigma_j^2$ อย่างน้อย 1 คู่; $i \neq j$

เขตปฏิเสธ H_0 เมื่อ W > F $_{(\alpha,k\text{-}1,N\text{-}k)}$ โดยที่การแจกแจง F มี df = k-1, N-k ณ ระดับนัยสำคัญ = α

สถิติทดสอบ $$W = \frac{(N-k)\sum_{i=1}^{k} N_i\left(\overline{Z}_{i.} - \overline{Z}_{..}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{j-1}^{N_i}\left(Z_{ij} - \overline{Z}_{i.}\right)^2}$$

โดยที่ Z_{ij} มีความหมายหนึ่งในสามความหมาย ดังนี้

1. $Z_{ij} = \left|Y_{ij} - \overline{Y}_{i.}\right|$ เมื่อ $\overline{Y}_{i.}$ คือ ค่าเฉลี่ย (mean)
2. $Z_{ij} = \left|Y_{ij} - \widetilde{Y}_{i.}\right|$ เมื่อ $\widetilde{Y}_{i.}$ คือ ค่ามัธยฐาน (median)
3. $Z_{ij} = \left|Y_{ij} - \overline{Y}'_{i.}\right|$ เมื่อ $\overline{Y}'_{i.}$ คือ 10% of trimmed mean

ซึ่ง $\overline{Z}_{i.}$ คือ ค่าเฉลี่ยของแต่ละกลุ่ม Z_{ij}

และ $\overline{Z}_{..}$ คือ ค่าเฉลี่ยของทุกกลุ่ม Z_{ij}

จากนั้นจึงเปรียบเทียบค่าเฉลี่ยหลายคู่ในครั้งเดียวด้วยสถิติ One-Way ANOVA ในที่นี้ได้ใช้การเปรียบเทียบจากการทดลองแบบสุ่มโดยสมบูรณ์ (completely randomized design: CRD) ซึ่งเป็นการทดลองเพื่อเปรียบเทียบ k ทรีทเม้นต์ โดยการสุ่มตัวอย่าง k กลุ่มอย่างเป็นอิสระกันแล้วกำหนดทรีทเม้นต์ให้ตัวอย่างแต่ละกลุ่มอย่างสุ่ม

สมมติฐานว่าง $H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots\ldots = \mu_k$

สมมติฐานแย้ง $H_1 : \mu_i \neq \mu_j$ อย่างน้อย 1 คู่; $i \neq j$

สถิติทดสอบ $F = \dfrac{MST_{rt}}{MSE}$ โดยที่การแจกแจง F มี df = k-1, n-k

เขตปฏิเสธ H_0 เมื่อ F > $F_{\alpha;\, k\text{-}1,\, n\text{-}k}$

ตารางที่ 5 การคำนวณของ CRD ANOVA

	Sum of Squares	df	Mean Square	F
Between Groups	SST_{rt}	k-1	MST_{rt}	MST_{rt}/ MSE
Within Groups	SSE	n-k	MSE	
Total	SST	n-1		

โดยมีสูตรการคำนวณ ดังนี้

$$CM = \frac{\left(\sum \sum X_{ij}\right)^2}{n} = \frac{\left(\sum T_i\right)^2}{n}$$

$$SST = \sum \sum X_{ij}^2 - CM$$

$$SST_{rt} = \frac{\sum T_i^2}{n_i} - CM$$

$$SSE = SST - SST_{rt}$$

$$MST_{rt} = \frac{SST_{rt}}{k-1}$$

$$MSE = \frac{SSE}{n-k}$$

3.8.2 ความสัมพันธ์ระหว่างสถิติทดสอบ F กับ t

เมื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของ 2 ประชากร (k = 2) จะใช้สถิติทดสอบ t หรือ Z แต่เมื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยตั้งแต่ 3 ประชากร ($k \geq 3$) จะใช้สถิติทดสอบ F แต่ถ้า k = 2 แล้วใช้สถิติทดสอบ F ที่มีองศาอิสระเป็น 1 และ n-2 ก็จะได้ผลลัพธ์เหมือนกับสถิติทดสอบ t แบบสองข้างโดยที่ MSE = S^2 ดังนั้น สถิติทดสอบ t และ F จึงมีความสัมพันธ์กันเป็น

$$F = t^2$$

แต่ถ้าสมมติฐานเป็นการทดสอบแบบข้างเดียว ต้องใช้สถิติทดสอบ t เท่านั้น ไม่สามารถใช้สถิติทดสอบ F ได้

3.9 หน่วยที่ 9 การทดสอบไคสแควร์

3.9.1 การทดสอบสมมติฐานทางสถิติสำหรับข้อมูลที่จำแนกทางเดียว

การทดสอบสมมติฐานทางสถิติสำหรับข้อมูลที่จำแนกทางเดียว หรือบางครั้งเรียกว่า การทดสอบภาวะสารูปสนิทดี (Goodness of Fit Test) เป็นการทดสอบเกี่ยวกับลักษณะหนึ่งของประชากร โดยพิจารณาจากความถี่ในแต่ละระดับ ทดสอบลักษณะต่าง ๆ ของประชากรว่าเป็นไปตามที่คาดไว้หรือไม่ ซึ่งมีตัวแปรที่สนใจตั้งแต่ 3 ค่าหรือ 3 ระดับขึ้นไป

การทดสอบว่าสัดส่วนประชากร k ประชากรเท่ากับสัดส่วนที่คาดไว้หรือไม่

สมมติฐานว่าง $H_0 : p_1 = p_{10}, p_2 = p_{20}, \ldots\ldots, p_k = p_{k0}$

สมมติฐานแย้ง $H_1 : p_i \neq p_{i0}$

เขตปฏิเสธ H_0 เมื่อ $\chi^2 > \chi^2_{\alpha, k-1}$

สถิติทดสอบ $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ มีการแจกแจงแบบไคสแควร์ด้วยองศาอิสระ หรือ df = k -1 โดยที่

O_i (Observed Frequency) = ความถี่หรือจำนวนครั้งที่เกิดในระดับที่ i ที่เกิดขึ้นจริงของตัวอย่างขนาด n

E_i (Expected Frequency) = ความถี่หรือจำนวนครั้งที่ของระดับที่ i ที่คาดว่าจะเกิดภายใต้สมมติฐานว่าง

K = จำนวนกลุ่มหรือจำนวนระดับของตัวแปรหรือลักษณะที่สนใจศึกษา

n = ขนาดตัวอย่างหรือจำนวนครั้งที่ทดลอง และ

$$\sum_{i}^{k} O_i = \sum_{i}^{k} E_i = n$$

เมื่อสมมติฐานว่างเป็นจริงจะได้ $E_i = np_{i0}$

ข้อจำกัดในการทดสอบ มีดังนี้

1. ความถี่ที่คาดไว้ของแต่ละระดับไม่ควรต่ำกว่า 5

2. ในกรณีที่ระดับของข้อมูลมีเพียง 2 ระดับ (k=2) องศาอิสระจะเหลือเพียง 1 (k=1) จะมีผลให้ค่าไคสแควร์สูงกว่าที่ควรจะเป็น จึงต้องใช้สูตรปรับค่าไคสแควร์ ยกเว้นมีขนาดตัวอย่างมากกว่า 50 ตัวอย่าง

สูตรปรับค่าไคสแควร์ $\chi^2 = \sum_{i=1}^{2} \frac{\left(|O_i - E_i| - 0.5\right)^2}{E_i}$

3.9.2 การทดสอบสมมติฐานทางสถิติสำหรับข้อมูลจำแนกแบบสองทาง

การทดสอบสมมติฐานทางสถิติสำหรับข้อมูลที่จำแนกสองทาง หรือสองตัวแปร หรือสองลักษณะ เช่น ผลการเรียนจำแนกตามเพศกับพื้นฐานความรู้ ฯลฯ

การทดสอบความเป็นอิสระกันระหว่างลักษณะสองลักษณะ (Testing for Independence of two Categorical Variable)

สมมติฐานว่าง H_0 : ลักษณะทั้งสองลักษณะเป็นอิสระกัน

สมมติฐานแย้ง H_1 : ลักษณะทั้งสองลักษณะไม่เป็นอิสระกัน

เขตปฏิเสธ H_0 เมื่อ $\chi^2 > \chi^2_{\alpha,(r-1)(c-1)}$

สถิติทดสอบ $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$ มีการแจกแจงแบบไคสแควร์ด้วยองศาอิสระ หรือ df = (r-1)(c -1) โดยที่

O_{ij} = ความถี่ของแถวนอนที่ i และแถวตั้ง (ลักษณะที่ 2) ที่ j

E_{ij} = ความถี่ของข้อมูลที่มีลักษณะที่ 1 ในระดับที่ i และมีลักษณะที่ 2 ในระดับที่ j เมื่อ i = 1,2,……,r; j = 1,2,……,c

$$E_{ij} = \frac{(r_i)(c_j)}{n}$$

r_i = จำนวนข้อมูลที่มีลักษณะที่หนึ่งในระดับที่ i

$$r_i = \sum_{j=1}^{c} O_{ij} \quad \text{เมื่อ } i = 1,2,\ldots\ldots,r$$

c_j = จำนวนข้อมูลที่มีลักษณะที่หนึ่งในระดับที่ j

$$c_j = \sum_{i=1}^{r} O_{ij}$$ เมื่อ j = 1,2,……,c

n = จำนวนข้อมูลทั้งหมด

$$n = \sum_{i=1}^{r} r_i = \sum_{j=1}^{c} c_j = \sum_{j=1}^{c}\sum_{i=1}^{r} O_{ij}$$

ข้อจำกัดในการทดสอบ มีดังนี้

1. ความถี่ที่คาดไว้ของแต่ละระดับไม่ควรต่ำกว่า 5

2. ในกรณีที่ระดับของข้อมูลในแต่ละลักษณะมีเพียง 2 ระดับ (r=2, c=2) องศาอิสระจะเหลือเพียง 1 = (r-1)(c-1) จะมีผลให้ค่าไคสแควร์สูงกว่าที่ควรจะเป็น จึงต้องใช้สูตรปรับค่าไคสแควร์ ยกเว้นมีขนาดตัวอย่างมากกว่า 50 ตัวอย่าง

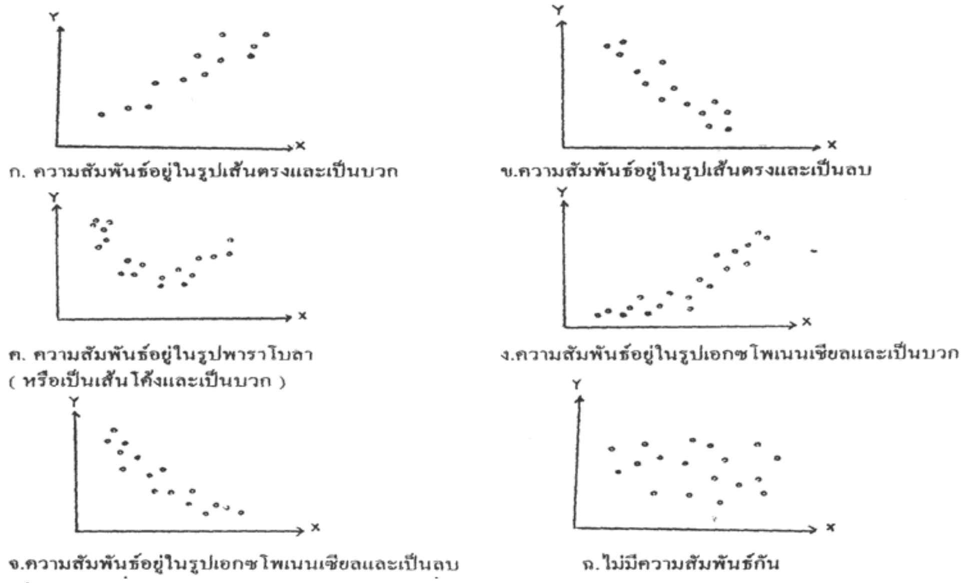
สูตรปรับค่าไคสแควร์ $$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{\left(\left|O_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}}$$

3.10 หน่วยที่ 10 การวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายและสหสัมพันธ์

การวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression Analysis) เป็นการศึกษาถึงข้อมูลที่ประกอบด้วยตัวแปร 2 ตัว หรือเรียกว่า ตัวแปรคู่ (bivariate data) ที่เป็นตัวแปรเชิงปริมาณ หรือข้อมูลต่อเนื่อง (Quantitative variables) โดยการวิเคราะห์จะนำตัวแปรทั้งสองมาพิจารณาพร้อม ๆ กัน เพื่อศึกษาถึงความสัมพันธ์ระหว่างตัวแปร 2 ตัว หรือ 2 ลักษณะ ในลักษณะเชิงเส้นตรง (linear) โดยที่ต้องทราบค่าของตัวแปรตัวหนึ่ง หรือต้องกำหนดค่าของตัวแปรตัวหนึ่งไว้ล่วงหน้า ตัวแปรที่ต้องการทราบค่าหรือต้องการพยากรณ์ เรียกว่า ตัวแปรตามมีสัญลักษณ์แทนด้วย y และตัวแปรตามขึ้นอยู่กับตัวแปรอีกตัวหนึ่งเป็นตัวแปรที่ทราบค่า เรียกว่า ตัวแปรอิสระ มีสัญลักษณ์แทนด้วย x เช่นคะแนนสอบ (y) ขึ้นอยู่กับระยะเวลาในการเตรียมตัวสอบ (x) ฯลฯ การวิเคราะห์ความถดถอยจะเกี่ยวข้องกับการประมาณค่าหรือการพยากรณ์ค่าของตัวแปรตาม เมื่อกำหนดค่าของตัวแปรอิสระโดยพยายามให้ค่าที่ประมาณหรือค่าที่พยากรณ์ได้มีความคลาดเคลื่อนน้อย หรือมีค่าใกล้เคียงกับความเป็นจริงมากที่สุด (กัลยา วานิชย์บัญชา, 2550: 252 - 354)

3.10.1 แผนภาพการกระจาย (Scatter Diagram)

เป็นการแสดงลักษณะความสัมพันธ์ระหว่างตัวแปรอิสระ (X) และตัวแปรตาม (Y) ว่ามีลักษณะแบบใด ซึ่งในการวิเคราะห์ Regression และ Correlation จำเป็นต้องดูลักษณะของความสัมพันธ์ทั้ง X และ Y ว่ามีความสัมพันธ์ในเชิงเส้นตรงหรือไม่ ก่อนจะไปทำการวิเคราะห์ต่อไปโดยนำค่า X และ Y มาทำ Scatter Plot มีลักษณะ ดังนี้



<u>รูปที่ 34</u> แผนภาพการกระจายแสดงลักษณะความสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y

(กัลยา วานิชย์บัญชา, 2550: 253)

จากรูปที่ 34 ก และ ข แสดงความสัมพันธ์ของ X และ Y ในลักษณะเชิงเส้นตรง ที่มีความสัมพันธ์ในทางเดียวกันหรือเป็นบวกในรูปที่ ก และมีความสัมพันธ์ในทางตรงกันข้ามกันหรือเป็นลบในรูปที่ ข จึงเรียกความสัมพันธ์นี้ว่า การวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (Simple linear Regression Analysis)

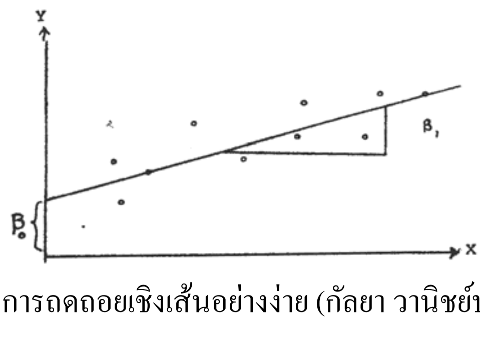
3.10.2 สมการการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (Simple linear Regression Analysis)

จากแผนภาพการกระจายระหว่าง 2 ตัวแปร ที่มีความสัมพันธ์กันในเชิงเส้นตรง จะสามารถเขียนรูปแบบสมการถดถอยเชิงเส้นอย่างง่ายที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระ 1 ตัว และตัวแปรตาม 1 ตัว โดยอาศัยรูปแบบสมการเส้นตรงทางคณิตศาสตร์ ดังนี้

$$\hat{Y} = \mathrm{a} + \mathrm{bX_i}$$

การประมาณค่า β_0 และ β_1 ด้วย a และ b ตามลำดับนั้น มีเป้าหมายเพื่อให้ความคลาดเคลื่อนในการประมาณค่าต่ำสุด โดยใช้วิธีกำลังสองน้อยที่สุดเนื่องจาก

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$



<u>รูปที่ 35</u> สมการถดถอยเชิงเส้นอย่างง่าย (กัลยา วานิชย์บัญชา, 2550: 254)

โดยที่ a เป็น ค่าระยะตัดแกน Y (Y-intercept) คือ ค่าที่เส้นตรงตัดแกน Y หรือ ค่า Y เมื่อ X = 0 นั่นเอง

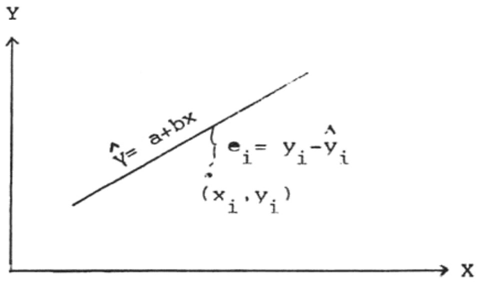
b เป็น ค่าความชัน (slope) ของเส้นตรง คือ ค่าแสดงให้ทราบว่าเมื่อ X มีค่าเปลี่ยนไป 1 หน่วย Y จะเปลี่ยนไปโดยเฉลี่ยเท่าใดหรือเรียกค่านี้ว่า สัมประสิทธิ์ความถดถอย (regression coefficient)

ถ้า b มีค่าเป็นบวก หรือมากกว่า 0 หมายความว่า X และ Y มีความสัมพันธ์ไปในทิศทางเดียวกัน

ถ้า b มีค่าเป็นลบ หรือน้อยกว่า 0 หมายความว่า X และ Y มีความสัมพันธ์ไปในทิศทางตรงกันข้าม

ถ้า b มีค่าเท่ากับ 0 หมายความว่า การเปลี่ยนแปลงของ Y ไม่ขึ้นกับการเปลี่ยนแปลงของ X

e_i เป็นความคลาดเคลื่อนสุ่มของการประมาณ Y โดยที่ $e_i = Y_i - \hat{Y}_i$ และ $\hat{Y} = a + bX_i$



<u>รูปที่ 36</u> ความคลาดเคลื่อนสุ่มของการประมาณ Y (กัลยา วานิชย์บัญชา, 2550: 259)

ข้อสมมติเบื้องต้นของความคลาดเคลื่อน มี ดังนี้

1. ค่า X จะต้องเป็นค่าที่กำหนดไว้ล่วงหน้าหรือทราบค่า
2. ความคลาดเคลื่อน เป็นตัวแปรสุ่มที่มีการแจกแจงปกติ
   โดยที่ $\mu_\varepsilon = 0$ และ $\sigma_\varepsilon^2 = \sigma^2$
3. e_i แต่ละตัวต้องเป็นอิสระต่อกัน นั่นคือ $E(e_i, e_j) = 0$ เมื่อ $i \neq j$

กำหนดให้

$$S = \sum e_i^2$$
$$= \sum \left(Y_i - \hat{Y}_i\right)^2$$
$$= \sum \left(Y_i - a - bX_i\right)^2$$

การหาค่า a และ b ที่ทำให้ S มีค่าต่ำสุด โดยการใช้อนุพันธ์เชิงส่วน (Partial derivation) จะได้สมการปกติ (Normal Equations) ดังนี้

$$\sum_{i=1}^{n} Y_i \quad = \quad \text{na} \; + \; \text{b}\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i \; = \; \text{a}\sum_{\text{i}=1}^{\text{n}} \text{X}_\text{i} + b\sum_{i=1}^{n} X_i^2$$

จากการแก้สมการข้างต้น จะได้

$$b \;=\; \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}$$

$$b \;=\; \frac{n\sum_{i=1}^{n} X_i Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

$$b \;=\; \frac{SS_{xy}}{SS_{xx}}$$

$$a \;=\; \overline{Y} - b\,\overline{X}$$

$$SS_{xx} = \sum_{i=1}^{n}\left(X_i - \overline{X}\right) = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}$$

$$SS_{xy} = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}$$

$$SS_{yy} = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) = \sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$$

โดยที่

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

$$\overline{Y} \;=\; \frac{\sum_{i=1}^{n} Y_i}{n}$$

การประมาณค่า β_0 และ β_1 ด้วยค่า a และ b โดยใช้วิธีกำลังสองน้อยที่สุดนี้ จะทำให้ $\sum\left(Y_i - \hat{Y}_i\right)^2$ มีค่าต่ำสุด และทำให้จุด $\left(\overline{\text{X}}, \overline{\text{Y}}\right)$ เป็นจุดที่อยู่บนเส้นถดถอยด้วย

3.10.3 ช่วงความเชื่อมั่น(Confidence Interval)

ข้อมูลที่ใช้สร้างสมการเส้นตรงที่แสดงความสมพันธ์ระหว่างตัวแปร X และ Y นั้น มาจากตัวอย่าง ค่าต่าง ๆ ที่คำนวณได้จึงเป็นเพียงค่าประมาณแบบจุด การนำค่าประมาณแบบจุดของพารามิเตอร์ไปใช้ อาจทำให้ผู้ใช้เกิดความไม่มั่นใจได้ว่า ค่าประมาณแบบจุดที่ได้ ใกล้เคียงกับค่าจริงเพียงใด

ค่าประมาณแบบช่วงจึงเป็นอีกทางเลือกหนึ่ง ค่าประมาณแบบช่วงของ β_0 β_1 $\mu_{y/x}$ และ Y สามารถหาได้จากสูตร ดังต่อไปนี้

ค่าประมาณแบบช่วงของค่าจุดตัดแกน Y หรือ β_0 ที่ช่วงความเชื่อมั่น (1-α) 100% ของ β_0 หาได้จาก $a \pm t_{\frac{\alpha}{2},n-2}.S_a$ คือ

$$a - t_{\frac{\alpha}{2},n-2}.S_a < \beta_0 < a + t_{\frac{\alpha}{2},n-2}.S_a$$

ค่าประมาณแบบช่วงของ slope ของเส้นตรง หรือ β_1 ที่ช่วงความเชื่อมั่น (1-α) 100% ของ β_1 หาได้จาก $b \pm t_{1-\frac{\alpha}{2},n-2}.S_b$ คือ

$$b - t_{\frac{\alpha}{2},n-2}.S_b < \beta_1 < b + t_{\frac{\alpha}{2},n-2}.S_b$$

ค่าประมาณแบบช่วงของค่าเฉลี่ยของ Y สำหรับค่า X ค่าหนึ่ง หรือ $\mu_{y/x}$ ที่ช่วงความเชื่อมั่น (1-α) 100% ของ $\mu_{y/x}$ หาได้จาก $V(\overline{Y}_x) = \sigma^2_{y/x}\left(\frac{1}{n} + \frac{(X_i - \overline{X})^2}{SS_x}\right)$ เมื่อ $V(\overline{Y}_x)$ เป็นความแปรปรวนของ $\overline{Y}_x$ และ $\overline{Y}_x$ เป็นค่าเฉลี่ยของตัวอย่างที่ได้หาจากการสมการถดถอย เมื่อ X มีค่าใดค่าหนึ่ง คือ

$$\overline{Y}_x - t_{\frac{\alpha}{2},n-2}.\sqrt{V(\overline{Y}_x)} \le \mu_{y/x} \le \overline{Y}_x + t_{\frac{\alpha}{2},n-2}.\sqrt{V(\overline{Y}_x)}$$

ค่าประมาณแบบช่วงของค่า Y ที่ค่า X ค่าหนึ่ง หรือ y_x ที่ช่วงความเชื่อมั่น (1-α) 100% ของ y_x หาได้จาก $V(\overline{Y}_x) = \sigma^2_{y/x}\left(1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{SS_x}\right)$ คือ

$$\hat{Y}_x - t_{\frac{\alpha}{2},n-2}.\sqrt{V(Y_x)} \le Y_x \le \hat{Y}_x + t_{\frac{\alpha}{2},n-2}.\sqrt{V(Y_x)}$$

3.10.4 การทดสอบสมมติฐาน

1) การทดสอบสมมติฐานค่าคงที่ β_0

สมมติฐานว่าง $H_0 : \beta_0 = 0$

สมมติฐานแย้ง $H_1 : \beta_0 \ne 0$

เขตปฏิเสธ H_0 เมื่อ $|t| > t_{\frac{\alpha}{2},n-2}$

สถิติทดสอบ $t = \frac{a - \beta_0}{S_a}$ โดยที่การแจกแจง t มี df = n-2

เมื่อ $S_a = S_{y.x}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum(X - \overline{X})^2}}$

2) การทดสอบสมมติฐาน β_1

สมมติฐานว่าง $H_0 : \beta_1 = 0$

สมมติฐานแย้ง $H_1 : \beta_1 \ne 0$

เขตปฏิเสธ H_0 เมื่อ $|t| > t_{\frac{\alpha}{2},n-2}$

สถิติทดสอบ $t = \dfrac{b-\beta_1}{S_b}$ โดยที่การแจกแจง t มี df = n-2

เมื่อ $S_b = \dfrac{S_{y.x}}{\sqrt{\sum\left(X-\overline{X}\right)^2}}$

3) การทดสอบสมมติฐาน β_1 โดยใช้การทดสอบความแปรปรวน

สมมติฐานว่าง $H_0 : \beta_1 = 0$

สมมติฐานแย้ง $H_1 : \beta_1 \neq 0$

เขตปฏิเสธ H_0 เมื่อ $F > F_{\alpha,(1,n-2)}$

สถิติทดสอบ $F = \dfrac{MSR}{MSE}$ โดยที่การแจกแจง F มี df = 1,n-2

เมื่อ

$$SST = \sum\left(Y_i - \overline{Y}\right)^2$$

$$SSR = \sum\left(\hat{Y}_i - \overline{Y}\right)^2 = bSS_{xy} = \frac{\left(SS_{xy}\right)^2}{SS_{xx}}$$

$$SSE = \sum\left(Y_i - \hat{Y}_i\right)^2 = SS_{yy} - 2bSS_{xy} + bSS_{xy} = SS_{yy} - \frac{\left(SS_{xy}\right)^2}{SS_{xx}}$$

ค่าประมาณความแปรปรวน

$$\sigma^2_{Y.X} = S^2_{y.x} = S^2 = \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{n-2} = \frac{SSE}{n-2}$$

$$\hat{V}(b) = S_b^2 = \hat{V}(SS_{xy}/SS_{xx}) = \frac{S^2}{SS_{xx}}$$

$$S_b = \frac{S}{\sqrt{SS_{xx}}}$$

ค่าความคลาดเคลื่อน $e_i = Y_i - \hat{Y}_i$

<u>ตารางที่ 6</u> การทดสอบความแปรปรวนของ ANOVA

Model	Sum of Squares	df	Mean Square	F
Regression	SSR	1	MSR=SSR/1	MSR/MSE
Residual	SSE	n-2	MSE=SSE/n-2	
Total	SST	n-1		

3.10.5 การทดสอบสัมประสิทธิ์การตัดสินใจหรือสัมประสิทธิ์การอธิบาย (Coefficient of Determination: r^2 or R^2)

สัมประสิทธิ์การตัดสินใจหรือสัมประสิทธิ์การอธิบาย หมายถึงสัดส่วนที่ตัวแปร X สามารถอธิบายการเปลี่ยนแปลงของตัวแปร Y ได้ ดังนั้นถ้า r^2 or R^2 มีค่ามากแสดงว่า Y และ X มีความสัมพันธ์กันมาก หรือ X สามารถอธิบายเปลี่ยนแปลงของตัวแปร Y ได้มาก โดยที่

$R^2 = r^2$ = ความแปรปรวนของ Y ที่เกิดจาก X หารด้วยความแปรปรวนของ Y ทั้งหมด

$$\therefore R^2 = r^2 = \frac{SSR}{SST} = \frac{bSS_{xy}}{SS_{yy}}$$

แต่เนื่องจาก SST = SSR + SSE

$$\therefore R^2 = r^2 = 1 - \frac{SSE}{SST}$$

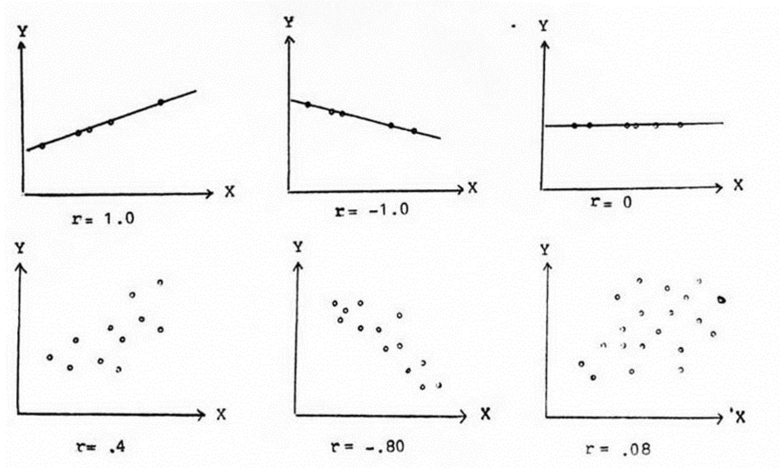
ดังนั้น $0 \le r^2 \le 1$ เนื่องจาก SST > SSR โดย r^2 จะไม่มีหน่วย และมีค่าเข้าใกล้ 1 แสดงว่าเปอร์เซ็นต์ที่ X สามารถอธิบายการเปลี่ยนแปลงของ Y มีค่ามาก หรือ X และ Y มีความสัมพันธ์กันมาก แต่ถ้า r^2 มีค่าเข้าใกล้ 0 แสดงว่าเปอร์เซ็นต์ที่ X สามารถอธิบายการเปลี่ยนแปลงของ Y มีค่าน้อย และมักอ่านค่าเป็นร้อยละ

3.10.6 การทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์ (Correlation, Pearson's Product Moment Correlation, r)

สัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ ก็คือเกณฑ์การตัดสินใจเกี่ยวกับขนาดของสัมประสิทธิ์สหสัมพันธ์ เกณฑ์นี้ไม่สามารถจะระบุชี้ชัดลงไป ทั้งนี้ขึ้นอยู่กับความสำคัญของงานวิจัยหรือตัวแปรที่ศึกษาด้วย สหสัมพันธ์ที่พบระหว่างตัวแปรมีค่า 0.20 อาจจะไม่มีความสำคัญสำหรับนักวิจัยคนหนึ่ง แต่อาจมีสำคัญอย่างใหญ่หลวงกับนักวิจัยอีกคนหนึ่ง โดยเฉพาะงานวิจัยที่เกี่ยวกับทางการแพทย์ เช่น การศึกษาความสัมพันธ์ของยาชนิดใหม่กับความปลอดภัยในชีวิตมนุษย์ เป็นต้น แต่อย่างไรก็ตามได้มีผู้กำหนดเกณฑ์การแปลความหมายของสัมประสิทธิ์สหสัมพันธ์ไว้ดังนี้

Cohen (Runyon and Other, 1996: 238 อ้างถึงใน Cohen, 1988) ได้แนะนำว่า สหสัมพันธ์ที่มีขนาดเล็ก หรือมีความสัมพันธ์กันน้อย ค่าสหสัมพันธ์จะอยู่ระหว่าง -0.29 ถึง -0.10 หรือ 0.10 ถึง 0.29 ส่วนสหสัมพันธ์ที่มีขนาดปานกลาง หรือมีความสัมพันธ์กันปานกลาง ค่าสหสัมพันธ์จะอยู่ระหว่าง -0.49 ถึง -0.30 หรือ 0.30 ถึง 0.49 และสหสัมพันธ์ที่มีขนาดใหญ่ หรือมีความสัมพันธ์กันสูง ค่าสหสัมพันธ์จะอยู่ระหว่าง -1.00 ถึง -0.50 หรือ 0.50 ถึง 1.00

แต่ Devore and Peck (1993: 129) ได้แนะนำเกี่ยวกับขนาดของสหสัมพันธ์ไว้ว่า ถ้าสัมพันธ์กันสูง ค่าสหสัมพันธ์จะมีค่าน้อยกว่า -0.80 หรือมีค่ามากกว่า 0.80 ถ้าสัมพันธ์กันปานกลาง ค่าสหสัมพันธ์จะมีค่าอยู่ระหว่าง -0.50 ถึง -0.80 หรือ 0.80 ถึง 0.50 และสัมพันธ์กันต่ำ ค่าสหสัมพันธ์ควรมีค่าอยู่ระหว่าง -0.50 ถึง 0.50



<u>รูปที่ 37</u> ขนาดต่าง ๆ ของสัมประสิทธิ์สหสัมพันธ์ (กัลยา วานิชย์บัญชา, 2550: 280)

ขนาดของสัมประสิทธิ์สหสัมพันธ์ โดยมีค่า -1 < r < 1 นักสถิติได้อธิบายความหมายของค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้ดังนี้

ถ้า r = 0 แสดงว่าไม่มีความสัมพันธ์กัน (independent)

ถ้า r < .30 แสดงว่ามีความสัมพันธ์กันน้อย

ถ้า .50 < r < .80 แสดงว่ามีความสัมพันธ์ในระดับปานกลาง

ถ้า r > .80 แสดงว่ามีความสัมพันธ์กันสูง

ถ้า r มีค่าห่างจากค่า 0 มาก แสดงว่า มีความสัมพันธ์กัน (dependent)

ถ้า r < 0 หมายถึงตัวแปร X และ Y มีความสัมพันธ์ในทิศทางตรงกันข้าม

ถ้า r > 0 หมายถึงตัวแปร X และ Y มีความสัมพันธ์ในทิศทางเดียวกัน

ถ้า r = 1 เป็น Perfect positive relationship

ถ้า r = -1 เป็น Perfect negative relationship

ความสัมพันธ์ระหว่างสัมประสิทธิ์การตัดสินใจกับสัมประสิทธิ์สหสัมพันธ์ ก็คือการนำค่าสัมประสิทธิ์สหสัมพันธ์มายกกำลังสอง (r^2) จะเรียกว่า สัมประสิทธิ์การตัดสินใจแปลความหมายได้ว่าเปอร์เซ็นต์ของความแปรปรวนในตัวแปรหนึ่งสามารถอธิบายหรือทำนายได้ด้วยตัวแปรอื่น ๆ ตัวอย่างเช่น ความสัมพันธ์ระหว่างความถนัดทางการเรียนกับผลสัมฤทธิ์ทางการเรียน มีค่าสหสัมพันธ์ r = -0.617 ดังนั้นสัมประสิทธิ์การอธิบายจะเท่ากับ $r^2 = 0.381$ เราจะแปลความหมายสัมประสิทธิ์การอธิบายได้ว่า 38.1% ของความแปรปรวนในผลสัมฤทธิ์ทางการเรียนสามารถอธิบายหรือทำนายได้ด้วยความถนัดทางการเรียน

เราสามารถคำนวณสัมประสิทธิ์ของการไม่อธิบาย (coefficient of nondetermination) ซึ่งจะบอกถึงเปอร์เซ็นต์ของความแปรปรวนในตัวแปรหนึ่งไม่สามารถอธิบายหรือทำนายได้ด้วยตัวแปรอื่น ๆ ดังนั้นสัมประสิทธิ์ของการไม่อธิบายจะคำนวณได้ด้วยสูตร ($1 - r^2$) ในตัวอย่างข้างต้น (1 - 0.381) = 0.619 ดังนั้น 61.9% ของความแปรปรวนในผลสัมฤทธิ์ทางการเรียนไม่สามารถอธิบายหรือทำนายได้ด้วยความถนัดทางการเรียน ดังนั้นสัมประสิทธิ์การอธิบายและสัมประสิทธิ์ของการไม่อธิบายจะมีผลรวมเท่ากับ 1.00

การคำนวณสัมประสิทธิ์สหสัมพันธ์ (r)

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

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

การทดสอบสัมประสิทธิ์สหสัมพันธ์ (Test of Correlation) ทดสอบว่าตัวแปรคู่นั้นมีความสัมพันธ์กันหรือไม่

สมมติฐานว่าง $H_1 : \rho = 0$ หรือ ตัวแปรคู่นั้นไม่มีความสัมพันธ์กัน

สมมติฐานแย้ง $H_1 : \rho \neq 0$ หรือ ตัวแปรคู่นั้นมีความสัมพันธ์กัน

เขตปฏิเสธ H_0 เมื่อ $|t| > t_{\alpha/2,n-2}$ (n<30) และ $|Z| > Z_{\alpha/2,n-2}$ (n≥30)

สถิติทดสอบ $t = Z = \frac{r - 0}{\sqrt{V(r)}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ โดยที่การแจกแจง t และ Z มี df = n-2

บรรณานุกรม

กานดา พูนลาภทวี.2530. **สถิติเพื่อการวิจัย.** กรุงเทพฯ: ฟิสิกส์เซ็นเตอร์.

กัลยา วานิชย์บัญชา. 2542. **การวิเคราะห์ข้อมูลด้วย SPSS for Windows.**พิมพ์ครั้งที่ 3.กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

______. 2549. **การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล.** พิมพ์ครั้งที่ 9. กรุงเทพฯ : ธรรมสาร.

______. 2550. **การวิเคราะห์สถิติ: สถิติสำหรับการบริหารและวิจัย.** พิมพ์ครั้งที่ 10. กรุงเทพฯ: ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

กัลยาณี คูณมี. 2538. **สถิติสำหรับเศรษฐศาสตร์และธุรกิจ.**พิมพ์ครั้งที่ 3.กรุงเทพฯ: คณะเศรษฐศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ฉัตรศิริ ปิยะพิมลสิทธิ์. 2544. **บทความสถิติ.** เมษายน-กุมภาพันธ์ 2544. Available. http://www.watpon.com.

ไชยยศ เรืองสุวรรณ. 2526. **เทคโนโลยีทางการศึกษา: หลักการ และแนวปฏิบัติ.** กรุงเทพฯ: วัฒนาพานิช.

______. 2546. **เทคโนโลยีการศึกษา: ทฤษฎีและการวิจัย.** กรุงเทพฯ: โอเดียนสโตร์.

ชูศรี วงศ์รัตนะ. 2550**. เทคนิคการใช้สถิติเพื่อการวิจัย.** พิมพ์ครั้งที่ 10. นนทบุรี: ไทเนรมิตกิจ อินเตอร์ โปรเกรสซิฟ.

ธานินทร์ ศิลป์จารุ. 2551. **การวิจัยและวิเคราะห์ข้อมูลทางสถิติด้วย SPSS**. พิมพ์ครั้งที่ 9. กรุงเทพฯ: บิสซิเนสอาร์แอนด์ดี.

นิศารัตน์ ศิลปเดช. 2542. **เอกสารประกอบการสอนระเบียบวิธีวิจัยทางสังคมศาสตร์เบื้องต้น.** กรุงเทพฯ: สถาบันราชภัฏธนบุรี.

บุญชม ศรีสะอาด. 2545. **วิธีการสร้างสถิติสำหรับการวิจัย**. พิมพ์ครั้งที่ 6. กรุงเทพฯ: สุวีริยาสาส์น.

_____. 2546**. การวิจัยสำหรับครู.** กรุงเทพฯ: ชมรมเด็ก.

_____. 2547. **วิธีการทางสถิติสำหรับการวิจัย.** พิมพ์ครั้งที่ 4. กรุงเทพฯ: สุวีริยาสน์.

บุญชม ศรีสะอาด และบุญส่ง นิลแก้ว. 2535. **การวิจัยเบื้องต้น.** พิมพ์ครั้งที่ 6. มหาสารคาม: มหาวิทยาลัยศรีนครินทรวิโรฒ.

บุญชม ศรีสะอาด และมนตรี อนันตรักษ์. 2549**. เอกสารประกอบวิชา 504702 การสร้างเครื่องมือในการวิจัย.** มหาสารคาม: ภาควิชาการวิจัยและพัฒนาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม.

บุญธรรม จิตต์อนันต์. 2536. **การวิจัยทางสังคมศาสตร์**. กรุงเทพมหานคร: สำนักส่งเสริมและฝึกอบรม, มหาวิทยาลัยเกษตรศาสตร์.

บุญเรียง ขจรศิลป์. 2543. **วิธีวิจัยทางการศึกษา.** พิมพ์ครั้งที่ 5. กรุงเทพฯ: พี.เอ็น.การพิมพ์.

______. 2549. **สถิติวิจัยI.** พิมพ์ครั้งที่ 9. กรุงเทพฯ: ภาควิชาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

ประคอง กรรณสูตร. 2542. **สถิติเพื่อการวิจัยทางพฤติกรรมศาสตร์ (ฉบับปรับปรุงแก้ไข).** กรุงเทพฯ: พิมพ์ครั้งที่ 3. ด่านสุทธาการพิมพ์.

ผกาวดี ศิริรังษี. 2548. **การวิเคราะห์ข้อมูล =Data Analysisм การประยุกต์สถิติในงานวิจัย.** กรุงเทพฯ: พระจอมเกล้าพระนครเหนือ.

เพ็ญแข แสงแก้ว.2541. **การวิจัยทางสังคมศาสตร์.** กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์.

ไพโรจน์ ตีรณธนากุล. 2530. **สถิติเพื่อการวิจัยทางการศึกษา.** กรุงเทพฯ: ศูนย์สื่อเสริมกรุงเทพฯ.

พิศิษฐ ตัณฑวณิช. 2547. **สถิติเพื่องานวิจัยทางการศึกษา.** พิมพ์ครั้งที่ 3. กรุงเทพฯ: บุ๊ค พอยท์.

พิศมัย หาญมงคลพิพัฒน์. 2550. **หลักสถิติ1**. พิมพ์ครั้งที่ 3. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.

ภาควิชาสถิติ. 2549. **หลักสถิติ1.** กรุงเทพฯ: คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

ยุทธพงษ์ กัยวรรณ์. 2543. **พื้นฐานการวิจัย**. กรุงเทพฯ: สุวีริยาสาส์น.

รัตนา ศิริพานิช. 2537. **สถิติและการวิจัยการศึกษา.** กรุงเทพฯ: คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์.

ระพินทร์ โพธิ์ศรี. 2549. **สถิติเพื่อการวิจัย.** กรุงเทพฯ: ด่านสุทธาการพิมพ์.

รวีวรรณ ชินะตระกูล. 2533. **คู่มือการทำ วิจัยทางการศึกษา**. กรุงเทพฯ: ภาพพิมพ์.

ล้วน สายยศ . **เทคนิคการวิจัยทางการศึกษา.** กรุงเทพฯ : สุวีริยาสาส์น, 2543.

ล้วน สายยศ และอังคณา สายยศ. 2522. **สถิติวิทยาทางการศึกษา.** พิมพ์ครั้งที่ 2. กรุงเทพฯ: วัฒนาพานิช.

______. 2539. **สถิติและการวิจัยการศึกษา.** กรุงเทพมหานคร : โรงพิมพ์โอเดียนสโตร์.

______. 2542. **การวัดด้านจิตพิสัย.** กรุงเทพฯ: สุวีริยาสาส์น.

วิเชียร เกตุสิงห์. 2530. **หลักการสร้างและวิเคราะห์เครื่องมือที่ใช้ในการวิจัย.** พิมพ์ครั้งที่ 1. กรุงเทพฯ: ไทยวัฒนาพานิช.

วัฒนา สุนทรธัย. 2551. **เรียนสถิติด้วย SPSS ภาคความรู้เบื้องต้น.** กรุงเทพฯ: วิทยพัฒน์.

วินัย โพธิ์สุรรณ์. 2538. **การพัฒนาโปรแกรมสำเร็จรูปทางด้านสถิติเบื้องต้น.** บทคัดย่อ.

วิรัชช พานิชวงศ์. 2549. **การวิเคราะห์การถดถอย.** พิมพ์ครั้งที่ 4. กรุงเทพฯ: ศูนย์ผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

วาโร เพ็งสวัสดิ์. 2551. **วิธีวิทยาการวิจัย.** กรุงเทพฯ: สุวีริยาสาส์น.

ศักดิ์สิทธิ์ วัชรารัตน์. 2552. **การพัฒนาชุดโปรแกรมช่วยวิเคราะห์ข้อมูลงานวิจัยทางการศึกษา**. วิทยาลัยสารพัดช่างพิษณุโลก. (อัดสำเนา)

______. 2553. **การพัฒนาชุดโปรแกรมช่วยสอนวิชาสถิติ เพื่อสร้างโจทย์ฝึกทักษะและแบบทดสอบ.** วิทยาลัยสารพัดช่างพิษณุโลก. (อัดสำเนา)

สมนึก ภัททิยธนี และคณะ. 2548. **พื้นฐานการวิจัยการศึกษา.** มหาสารคาม : ภาควิชาการวิจัยและพัฒนาการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม.

สมบัติ สินธุเชาวน์. มปป. **วิชา 1302-103: สถิติวิศวกรรม (Engineering Statistic).** ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยอุบลราชธานี. (Online) Available: http://app.eng.ubu.ac.th/~edocs/ (สืบค้นข้อมูล 25 มกราคม 2552).

อังคณา สายยศ. 2536. "การวิเคราะห์ข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ในเชิงปฏิบัติ: อิงกลุ่ม, อิงเกณฑ์, อัตนัย และ ICC", **วารสารการวัดผลการศึกษา**. 14(43) (พฤษภาคม – สิงหาคม 2536), 15-38.

อุทุมพร จามรมาร. 2531. **การสร้างและการพัฒนาเครื่องมือวัดลักษณะผู้เรียน.** พิมพ์ครั้งที่ 2. กรุงเทพฯ: พันนี่พับบลิชชิ่ง.

เว็บไซต์

http://en.wikipedia.org/wiki/Statistics

http://tulip.bu.ac.th/~wathna.s/grading/grade1.htm

http://tulip.bu.ac.th/~wathna.s/item.htm

http://web.uccs.edu/lbecker/SPSS/content.htm

http://www1.graphpad.com/welcome.htm

http://www.biology.ed.ac.uk/research/groups/jdeacon/statistics/tress1.html

http://www.geocities.com/nincoo/new_page_3.htm

http://www.gseis.ucla.edu/courses/ed230bc1/default.html

http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm

http://www.medcalc.be/manual/index.php

http://www.micquality.com/six_sigma_glossary/index.htm

http://www.ou.edu/class/jmc3333/data.htm

http://www.ozgrid.com/

http://www.statisticssolutions.com/

http://www.statsoftinc.com/textbook/streliab.html.

http://www.statsoft.com

http://www.sut.ac.th/e-texts/Social/ProjectCAI/

http://www.une.edu.au/home.html

http://www.union.edu/welcome.html

http://www.uvm.edu/

http://www.vec.go.th/doc/DirectorStr/college_th.php

http://www.wadsworth.com/psychology_d/special_features/ext/workshops/reliability2.html.