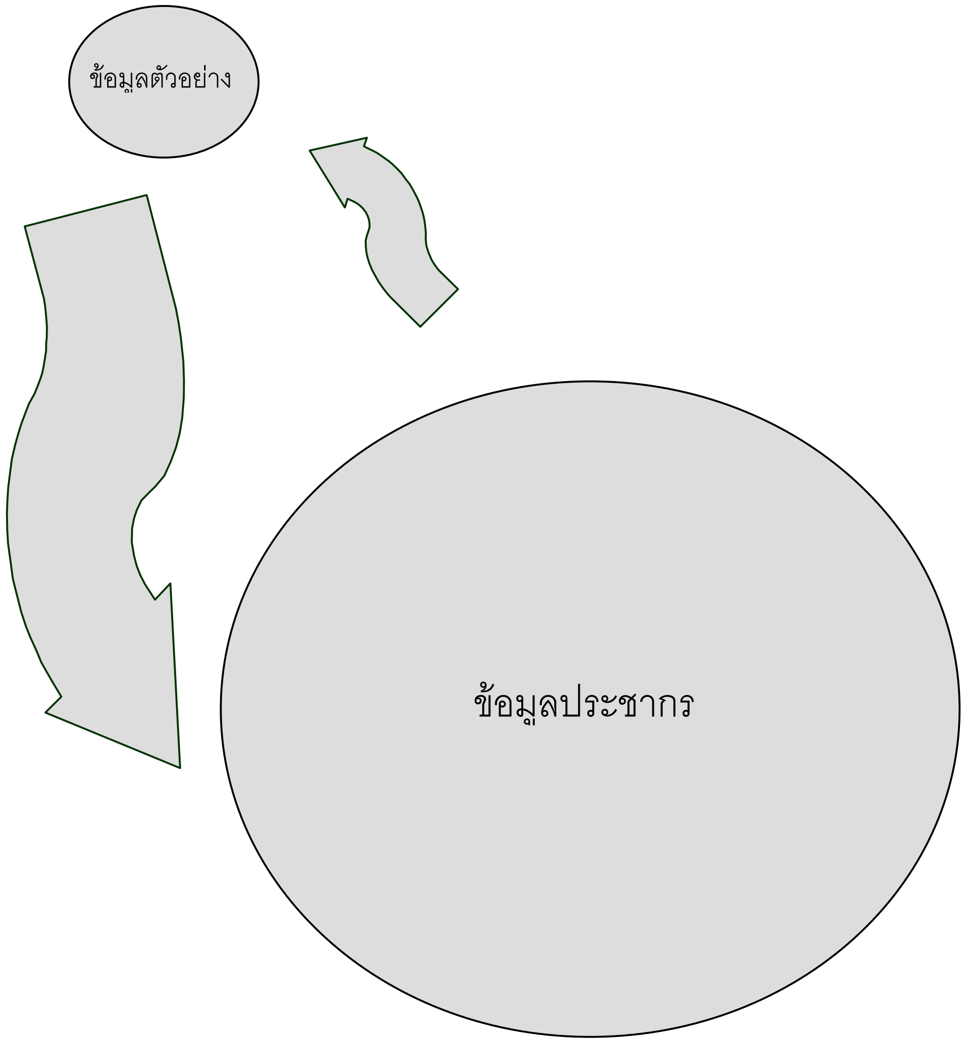
สถิติธุรกิจ



ผู้ช่วยศาสตราจารย์ถิรเดช พิมพ์ทองงาม

มหาวิทยาลัยราชภัฏเทพสตรี

Tiradate Pimtongngam Version 1.0 (2008)

สถิติธุรกิจ

-หน้านี้ใช้เป็นปกใน-

ถิรเดช พิมพ์ทองงาม

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเทพสตรี

ISBN 974-9606-02-7

พ.ศ.2549

Tiradate Pimtongngam Version 1.0 (2008)

สถิติธุรกิจ

พิมพ์ครั้งที่ 1 พ.ศ.2546

พิมพ์ครั้งที่ 2 พ.ศ.2548

พิมพ์ครั้งที่ 3 พ.ศ.2549

พิมพ์ที่ ศูนย์ตำราและเอกสารทางวิชาการ
มหาวิทยาลัยราชภัฏเทพสตรี ลพบุรี

ISBN 974-9606-02-7

-หน้านี้อยู่หลังปกนอก-

คำนำ

การจัดพิมพ์ครั้งที่สามสถิติธุรกิจ เพื่อให้นักศึกษาสาขาบริหารธุรกิจ การจัดการทั่ว บัญชี ตลอดจนคอมพิวเตอร์ธุรกิจ ทั้งในมหาวิทยาลัยฯ วิทยาเขตและศูนย์การศึกษาต่างๆ ได้ใช้ประกอบการเรียนวิชาสถิติธุรกิจ

ถิรเดช พิมพ์ทองงาม

14 มกราคม 2549

คำนำ

(พิมพ์ครั้งที่ 2)

การจัดพิมพ์ครั้งที่สองสถิติธุรกิจเล่มนี้ ได้แก้ไขและมีการพิสูจน์อักษรเพิ่มเติมหวังเป็นอย่างยิ่งสิ่งที่ได้ปรับปรุงแก้ไขจะมีประโยชน์สำหรับผู้ใช้มากขึ้น

ถิรเดช พิมพ์ทองงาม

14 มกราคม 2548

คำนำ

(พิมพ์ครั้งที่ 1)

สถิติธุรกิจเล่มนี้จัดทำขึ้นโดยรวบรวมขึ้นจากประสพการณ์ การสอนรายวิชาสถิติธุรกิจพบกับผู้เรียนมีพื้นฐานทางวิทยาศาสตร์และคณิตศาสตร์ค่อนข้างน้อย ทำให้การทำความเข้าใจและเรียนรู้หลักเกณฑ์ทางสถิติ รวมถึงการประยุกต์ใช้ให้เหมาะสมกับสาขาของผู้เรียน เช่นสาขาบริการธุรกิจ สาขาการตลาด สาขานิเทศศาสตร์และอื่นๆ วิชาสถิติธุรกิจนี้บางสาขากำหนดเป็นวิชาชีพบังคับและหลายสาขาก็กำหนดเป็นวิชาเลือกบังคับ เอกสารฉบับนี้เริ่มต้นจากเรื่องข้อมูลและการจัดระบบข้อมูล การวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูล การวัดการกระจายของข้อมูล การแจกแจงปรกติของข้อมูล ซึ่งเป็นความรู้พื้นฐานการบรรยายลักษณะทั่วไปของข้อมูลโดยเฉพาะข้อมูลทางธุรกิจ ด้วยการใช้หลักเกณฑ์คณิตศาสตร์อย่างง่ายๆและสามารถนำไปใช้ศึกษาโดยใช้คณิตศาสตร์ชั้นสูงในโอกาสต่อไป นำเสนอความรู้เบื้องต้นเกี่ยวกับสถิติอนุมานภายใต้ทฤษฎีความน่าจะเป็นเชื่อมโยงไปหาการประยุกต์ใช้งานแก้ปัญหาทางธุรกิจที่จำเป็น

หวังเป็นอย่างยิ่งว่าเอกสารฉบับนี้จะเป็นประโยชน์ สำหรับผู้เรียนและผู้สนใจนำไปใช้เป็นพื้นฐานสำหรับการศึกษาชั้นสูงต่อไป ความดีของเอกสารฉบับนี้ขอมอบให้ครูอาจารย์ผู้ประสิทธิประศาสตร์วิชาการและบุพการีทุกสายโดยเฉพาะมารดาผู้ล่วงลับ แต่หากมีข้อผิดพลาดด้วยประการใดๆก็ตามข้าพเจ้าขอน้อมรับไว้แต่เพียงผู้เดียวและพร้อมที่จะแก้ไขในโอกาสต่อไป

ถิรเดช พิมพ์ทองงาม

14 มกราคม 2546

-หน้านี้ใช้เป็นแผ่นรองปกหลัง-

พิมพ์ที่ ศูนย์ตำราและเอกสารทางวิชาการ
มหาวิทยาลัยราชภัฏเทพสตรี ลพบุรี
โทรศัพท์: 0-3642-2607-9 โทรสาร: 0-3642-2610

Tiradate Pimtongngam Version 1.0 (2008)

----แผ่นนี้ใช้ประกอบกับปกนอกเป็นส่วนสันหนังสือ----

สถิติธุรกิจ ถิรเดช พิมพ์ทองงาม 2549

สารบัญ

บทที่ 1
ข้อมูลและการจัดระบบข้อมูล

เมื่อกล่าวถึงข้อมูลเป็นคำกล่าวรวม ๆ ที่แสดงถึงข้อเท็จจริงหรือการบอกลักษณะของเรื่องราวเรื่องใดเรื่องหนึ่ง สิ่งใดสิ่งหนึ่งหรือสถานการณ์ใดสถานการณ์หนึ่งอันได้แก่ ข่าวสาร (information) เป็นข้อเท็จจริงในลักษณะการบรรยายความ เช่น รายงานอากาศประจำวัน รายงานอุบัติเหตุบนท้องถนน เป็นต้น ระเบียนหรือทะเบียน (record or registration) เป็นข้อเท็จจริงที่บ่งบอกเรื่องราวใด ๆ ที่มีการปรับแก้และทันสมัยอยู่เสมอ เช่น ทะเบียนบ้าน ระเบียนผลการศึกษา (transcripts) เป็นต้น และข้อมูลเชิงสถิติ (statistical data) ในเบื้องต้นนี้เพื่อความเข้าใจถึงการบรรยายคุณลักษณะประชากรที่ต้องการศึกษาจะถือว่าข้อมูลเชิงสถิติที่อยู่ในมือเป็นประชากร จากที่กล่าวมาจะเห็นได้ว่าไม่ว่าจะเป็นข่าวสาร ระเบียน ต้องอาศัยข้อมูลทั้งสิ้น ดังนั้นในบทนี้จะขอกล่าวถึงข้อมูลเชิงสถิติ แหล่งข้อมูล การแจกแจงความถี่ของข้อมูล (frequency distribution of data) ตลอดจนการนำเสนอการแจกแจงความถี่ของข้อมูลเพื่อการเสนอวิเคราะห์หรือตัดสินใจในเรื่องราวต่าง ๆ

1.1 ข้อมูลเชิงสถิติ

ข้อมูลเชิงสถิติเป็นตัวเลขที่แสดงข้อเท็จจริงอันเกิดจากผลการดำเนินการเบื้องต้นด้วยการชั่ง การตวง การวัด หรือการนับ เช่น เจ้าของร้านค้าย่อยนับจำนวนชิ้นสินค้าชนิดหนึ่งที่จำหน่ายได้แต่ละวันแล้วจดบันทึกแต่ละวัน ตัวเลขที่แสดงจำนวนสินค้าเป็นข้อเท็จจริงอันเกิดจากการประมวลผลเบื้องต้นด้วยการนับ เป็นต้น เราอาจเรียกข้อมูลเชิงสถิตินี้ว่าข้อมูลดิบ (raw data) หรือ คะแนนดิบ (raw score) การกล่าวถึงข้อมูลใด ๆ นิยมใช้อักษรภาษาอังกฤษ X , Y หรือ Z แทน ส่วนค่าของข้อมูลนั้น ๆ ซึ่งได้จากการนับหรือการวัดจะใช้ x, y หรือ z แทน หากมีข้อมูลและตัวอักษรจำนวนมาก ข้อมูลและตัวอักษรนี้ไม่เพียงพอสำหรับเขียนแทน จึงนิยมใช้ดัชนีล่าง (subscript) เช่น ข้อมูลชุดแรกเป็นน้ำหนักนักศึกษาชาย 5 คน เราอาจเขียนแทนนักศึกษาชายด้วย x_1 , x_2 , x_3 , x_4 และ x_5 แล้วแทนน้ำหนักนักศึกษาหญิง 5 คน

ด้วย y_1 , y_2 , y_3 , y_4 และ y_5 สำหรับค่าตัวเลข 1, 2, 3, 4 และ 5 ที่เขียนข้าง ๆ x หรือ y คือ ดัชนีล่าง ถ้าหากต้องการกล่าวถึงข้อมูลค่าหนึ่งค่าใด จะใช้สัญลักษณ์ x_i หรือ y_i แทน นอกจากนี้ดัชนีล่างอาจพบในลักษณะอื่นที่บอกให้ทราบว่าข้อมูลที่แตกต่างกัน เช่น x_{11} , x_{12} , x_{13} , ... เป็นข้อมูลที่แตกต่างจาก x_{12} , x_{22} , x_{23} ,...... เป็นต้น ข้อมูลเชิงสถิติที่ใช้ในงานวิจัยด้านต่าง ๆ จะต้องมีคำหรือข้อความที่เกี่ยวข้อง โดยจะกล่าวถึงดังต่อไปนี้

1.1.1 ข้อมูลเชิงสถิติสำหรับการคำนวณ ในการศึกษาสถิติศาสตร์อาจเรียกข้อมูลเชิงสถิติสำหรับการคำนวณในขอบเขตทั้งหมดหรือบางส่วนที่ต้องการศึกษาแตกต่างกันตามที่มาของ ข้อมูลหรือสภาวะการณ์เกิดข้อมูลดังต่อไปนี้

1) ค่าสังเกต (observation) หมายถึงตัวเลขที่ได้จากการวัดหน่วยหรือสิ่งของทั้งหลายที่เราสนใจ เช่น นำกระต่ายมาเลี้ยงไว้ในห้องทดลองระยะหนึ่งแล้วนำกระต่ายมาชั่งน้ำหนักทีละตัวน้ำหนักกระต่าย 1 ตัว คือ 1 ค่าสังเกตหรือข้อมูล 1 ค่า

2) ตัวแปร (variable) หมายถึงข้อความหรือตัวเลขที่แสดงคุณลักษณะของประชากรหรือสิ่งตัวอย่างที่ยังมิได้สรุปคุณลักษณะรวม ๆ อย่างชัดเจน

3) ตัวแปรสุ่ม (random variable) มีความหมายเช่นเดียวกับตัวแปรแต่ได้มาโดยอาศัยกระบวนการชักตัวอย่างสุ่ม (random sampling) ข้อมูลเชิงสถิติอนุมานจะเป็น ตัวแปรชนิดนี้ เพราะตัวแปรที่ได้สุ่มจากประชากรสำหรับเป็นสิ่งตัวอย่างในการอนุมาน

4) ค่าตอบสนอง (response variable) ตัวเลขที่ได้จากการวัดหน่วยหรือสิ่งของอันเกิดจากอิทธิพลของปัจจัยอย่างหนึ่งผลการเติมโตของกระต่าย 1 ตัวนับจากระยะแรกที่เริ่มให้อาหารในช่วงระยะเวลาที่กำหนดเป็น 1 ค่าตอบสนองหรือข้อมูล 1 ค่า

1.1.2 ประเภทของข้อมูลเชิงสถิติ ข้อมูลที่แสดงค่าด้วยตัวเลขสามารถชั่ง ตวง วัด และนับได้โดยตรง หรืออาจกำหนดขึ้นการแปลงข้อมูล (data coding) ที่เป็นข้อความเพื่อให้ค่าแตกต่างกัน ดังนั้นจึงอาจจำแนกได้เป็น 2 ประเภท คือ

1) ข้อมูลจำแนกตามคุณภาพ (qualitative data) เป็นข้อมูลประเภท ข้อความ เช่น เพศ การนับถือศาสนา อาชีพ เชื้อชาติ จะกำหนดเป็นค่าตัวเลขให้เห็นแตกต่างกันไม่ได้แสดงปริมาณมากน้อย กรณีเพศอาจกำหนดเลข 1 แทนเพศชายและกำหนด 2 แทนเพศหญิง เป็นต้น

2) ข้อมูลจำแนกตามปริมาณ (quantitative data) เป็นข้อมูลที่ได้จากการชั่ง ตวง วัดหรือนับได้ ค่าตัวเลขจะแสดงปริมาณมากน้อย เช่น น้ำหนักของทารกแรกเกิด ความสูงนักศึกษา ความยาวของทางเดิน ยอดจำหน่ายสินค้ารายวัน เป็นต้น

1.1.3 ลักษณะข้อมูลเชิงสถิติ สำหรับการคำนวณค่าเชิงคณิตศาสตร์อาจจะกำหนดขึ้นหรือแปลงข้อมูลที่เป็นข้อความหรือข่าวสารเป็นตัวเลขรวมถึงการได้ตัวเลขมาโดยตรง การคำนวณจะเรียกว่าตัวแปร ค่าตัวแปรที่พบจะมี 2 ลักษณะ คือ

1) ตัวแปรวิยุต (discrete variables) คือค่าที่แสดงออกมาเป็นตัวเลขจำนวนเต็มและในสถานการณ์ไม่สามารถแสดงด้วยเลขทศนิยมหรือเศษส่วนหรือแสดงได้แต่ไม่มีความหมายเดิมจะเรียกตัวแปรลักษณะนี้ว่าตัวแปรวิยุต เช่น จำนวนบุตรของครอบครัวจำนวนผู้จำหน่ายตรงในสาย จำนวนหนังสือในห้องสมุด จำนวนไข่ในกล่อง เป็นต้น ซึ่งกรณีจำนวนบุตรของครอบครัวจะเป็น 0, 1, 2, 3, 4, ... คน ในสถานการณ์จริงจะไม่พบครอบครัวใดที่มีจำนวนบุตร 2.5 คน 3.25 คน อย่างแน่นอน

2) ตัวแปรต่อเนื่อง (continuous variables) คือค่าที่แสดงออกมา เป็นตัวเลขในรูปของช่วง (interval) ภายในขอบเขตที่กำหนดได้อย่างมีความหมายและเป็นไปได้ทุกค่าตัวเลขที่แสดงของช่วง เช่น ปริมาณน้ำมันที่สูญหายแต่ละวันของสถานีจ่ายน้ำมันแห่งหนึ่งเป็น 0-15 ลิตร น้ำหนักนักศึกษาชายเป็น 70-100 กิโลกรัม ในสถานการณ์จริง อาจพบค่าที่แสดงเป็นเศษส่วนหรือทศนิยม กรณีปริมาณน้ำมันที่สูญหายในแต่ละวันเป็น 0, 2.8, 5.22, 10.215,…ลิตรและกรณีน้ำหนักนักศึกษาอาจเป็น 58.2, 59.3, 48.7,…กิโลกรัม เป็นต้น

1.1.4 การจัดระดับตัวแปร ก่อนที่จะดำเนินการคำนวณตามระเบียบวิธีทางสถิติที่เหมาะสมกับตัวแปรวิยุตหรือตัวแปรต่อเนื่องที่ได้มาอาจได้มาโดยการชั่ง ตวง วัด การนับโดยตรงหรือกำหนดค่าขึ้นเองจึงกำหนดระดับการวัด (scale) จากระดับหยาบที่สุดถึงระดับละเอียดที่สุดจัดระดับตัวแปรที่ได้เป็น 4 ระดับ คือ

1) ระดับนามบัญญัติ (nominal scale) เป็นตัวเลขที่แสดงการแบ่งประเภทจำแนกกลุ่มตามลักษณะหรือคุณสมบัติที่กำหนด เช่น เพศ กำหนด 1 คือ เพศชาย 2 คือ เพศหญิง อาชีพ กำหนด 1 คือค้าขาย 2 คือรับราชการ 3 คือพนักงานรัฐวิสาหกิจ 4 คือ อาชีพอื่น ๆ

2) ระดับอันดับ (ordinal scale) เป็นตัวเลขแสดงการจัดลำดับ การเรียงคุณสมบัติตามปริมาณตามความสำคัญหรือตามความพอใจ เช่น ลำดับความชอบสินค้า ลำดับความชอบสิ่งของ ลำดับความรักเพื่อน ลำดับค่านิยามในสิ่งใดสิ่งหนึ่ง เช่น จัดลำดับความชอบออกมาเป็นลำดับ น้อยที่สุดเป็นค่า 1 น้อยเป็นค่า 2 ปานกลางเป็นค่า 3 มากเป็นค่า 4 และมากที่สุดเป็นค่า 5 เป็นต้น

3) ระดับช่วง (interval scale) เป็นตัวเลขแสดงความแตกต่างระหว่างค่าที่ได้

แต่ละช่วงของการวัด ค่าที่ได้จะไม่มีจำนวนศูนย์แท้จริง แต่ทราบความมากน้อยแตกต่างกันจำนวนเท่าใด เช่น คะแนนสอบ อุณหภูมิ เป็นต้น

4) ระดับอัตราส่วน (ratio scale) เป็นตัวเลขแสดงค่าที่เกิดจากการชั่ง ตวงวัดหรือการนับที่บ่งบอกถึงขนาดที่แน่นอนและมีจำนวนศูนย์แท้จริง เช่น ปริมาตร ความเร็ว ความยาว น้ำหนัก พื้นที่ เป็นต้น

ตัวอย่างที่ 1.1 จงบอกลักษณะตัวแปรและระดับตัวแปรจากสถานการณ์ต่อไปนี้

ก. ความสูงของนักศึกษาชั้นปีที่ 1 ของสถาบันราชภัฏเทพสตรี

ข. จำนวนหนังสือที่นักศึกษาชั้นปีที่ 1 สถาบันราชภัฏฯ ยืมจากห้องสมุด

วิธีทำ ก. ความสูงของนักศึกษาชั้นปีที่ 1 ของสถาบันราชภัฏเทพสตรีเป็นตัวแปรต่อเนื่อง ระดับอัตราส่วน ค่าตัวแปรที่เป็นไปได้ เช่น 152.2, 168.0, 157.5, …เซนติเมตร

ข. จำนวนหนังสือที่นักศึกษาชั้นปีที่ 1 สถาบันราชภัฏฯ ยืมจากห้องสมุดเป็นตัวแปรเต็มหน่วย ระดับอัตราส่วน ค่าตัวแปรที่เป็นไปได้ เช่น 3, 0, 12, 5, 0, 4, ... เล่ม

1.2 แหล่งข้อมูล

ในการเก็บข้อมูลในงานธุรกิจและงานวิจัยด้านต่าง ๆ จะต้องมีที่มาของข้อมูลซึ่งข้อมูลตามแหล่งที่มา (source of data) แบ่งออกได้เป็น 2 ประเภท คือ

1.2.1 ข้อมูลปฐมภูมิ (primary data)
เป็นข้อมูลที่ได้จากการเก็บจากต้นกำเนิดโดยตรง ได้แก่ ข้อมูลจากการบันทึก การทดลอง การสังเกต การส่งแบบสอบถาม การสัมภาษณ์หรือใช้เครื่องมืออื่นใด

1.2.2 ข้อมูลทุติยภูมิ (secondary data) เป็นข้อมูลที่เก็บจากข้อมูลที่มีอยู่แล้วตามหน่วยงานต่าง ๆ เพื่อประโยชน์สำหรับหน่วยงานนั้น ๆ การใช้ข้อมูลประเภทนี้ ผู้รวบรวมจะต้องคำนึงถึงความสะดวก ความประหยัดเวลาและค่าใช้จ่าย แต่ตามปกติแล้วข้อมูลประเภทนี้มักถูกนำออกเผยแพร่แก่บุคคลทั่วไปในรูป บทความ งานวิจัย หรือในระบบเครื่อข่ายอินเทอร์เน็ต

อยู่แล้ว

1.3 การนำเสนอข้อมูล

หลังจากการได้ข้อมูลมาไม่ว่าจากแหล่งปฐมภูมิหรือจากแหล่งทุติยภูมิก็ตาม จะต้องนำมาจัดระบบข้อมูล (organization of data) ซึ่งการจัดระบบข้อมูลนี้มีจุดมุ่งหมาย 2 ประการ คือเพื่อนำเสนอเบื้องต้นโดยการจัดระบบให้สะดวกในการพิจารณาและเข้าใจได้ง่ายต่อผู้อื่นอันเป็นการจูงใจให้น่าสนใจด้วยวิธีการที่เหมาะสม เนื้อหาสาระไม่ขาดหายไปและเพื่อแสดงให้เห็นธรรมชาติของข้อมูลด้วยการแจกแจงความถี่ของข้อมูล สำหรับการนำเสนอข้อมูลโดยการจัดระบบให้สะดวกในการพิจารณาหรือการนำไปใช้ นิยมนำเสนอ 6 แบบ คือ

1.3.1 การนำเสนอข้อมูลโดยบทความ (text presentation) เป็นการบรรยายความเกี่ยวกับสภาพของข้อมูล การแสดงตัวเลข เช่น รายงานอุณหภูมิอากาศประเทศไทยเมื่อเวลา 07.00 น. ภาคเหนือที่จังหวัดเชียงใหม่ 20 องศาเซลเซียส ภาคใต้ที่จังหวัดชุมพร 32 องศาเซลเซียส เป็นต้น

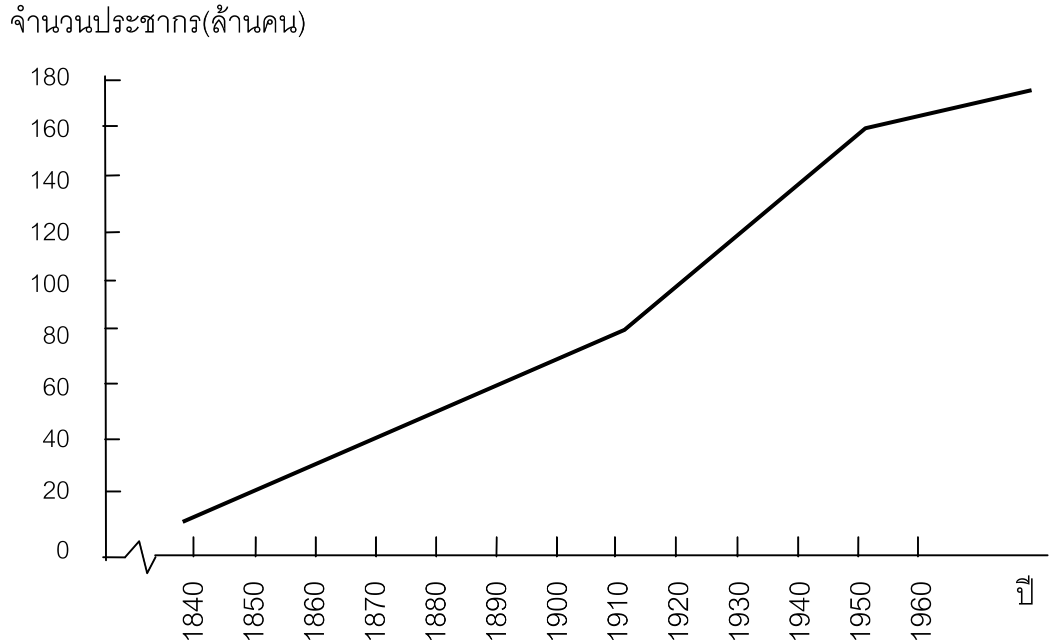
1.3.2 การนำเสนอข้อมูลโดยตาราง (tabular presentation) เป็นการนำเสนอข้อมูลแบบหนึ่งที่นิยมนำเสนอข้อมูลเชิงสถิติ เพราะอ่านง่าย เข้าใจง่าย วิธีการคือจัดข้อมูลให้สัมพันธ์กันในแนวตั้งหรือสดมภ์ (column) และแนวนอนหรือแถว (row) ดังตารางที่ 1.1 เป็นการนำเสนอข้อมูลของจำนวนพนักงานแผนกต่าง ๆ ตามเพศ

ตารางที่ 1.1 การนำเสนอข้อมูลของจำนวนพนักงานแผนกต่าง ๆ ตามเพศ

ประเภทพนักงาน	ชาย	หญิง	รวม
พนักงานแผนกเครื่องเขียน	65	35	100
พนักงานแผนกเครื่องดื่ม	-	2	2
พนักงานแผนกเสื้อผ้า	22	16	38
พนักงานแผนกเครื่องใช้ไฟฟ้า	4	-	4
รวมทั้งสิ้น	91	53	144

1.3.3 การนำเสนอข้อมูลโดยกราฟเส้น (broken-line graph presentation) นิยมนำ

เสนอข้อมูลที่ต้องการให้เห็นต่อเนื่อง (continuity) และเปรียบเทียบปริมาณ 2 อย่าง เช่น จำนวนคนกับเวลา จำนวนสินค้าส่งออกกับจำนวนสินค้าที่ผลิตได้ ดังแสดงในภาพที่ 1.1

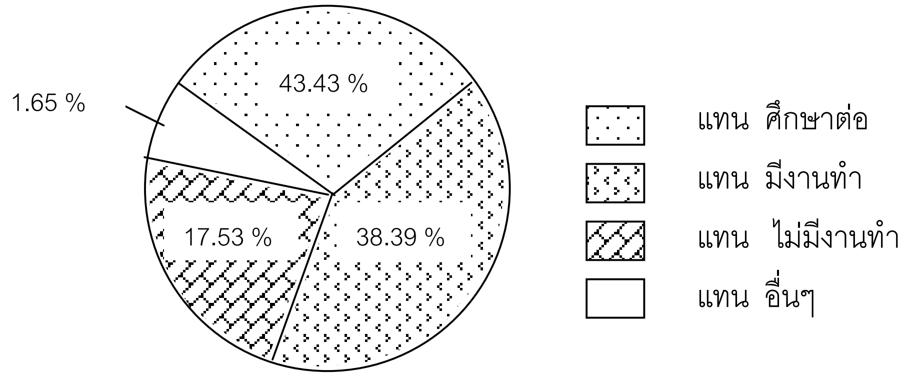


ภาพที่ 1.1 การนำเสนอข้อมูลโดยกราฟเส้น

1.3.4 การนำเสนอข้อมูลโดยกราฟแท่ง (bar-chart presentation) เป็นการแสดงความสัมพันธ์ของปริมาณ 2 อย่างเช่นเดียวกับกราฟเส้น แต่ความสัมพันธ์ที่ปรากฏแก่ผู้พบเห็นจะมีลักษณะเป็นแท่ง ๆ อาจระบายสี มีลวดลายที่สุภาพในแท่งกราฟแต่ละแท่งควรมีฐานกว้างเท่ากันทุกแท่ง

1.3.5 การนำเสนอข้อมูลโดยกราฟรูปภาพ (pictograph presentation) เป็นกราฟที่สามารถดึงดูดความสนใจจากผู้พบเห็นและสามารถเข้าใจความหมายได้ทันที เพราะกราฟรูปภาพจะใช้ภาพจริง ภาพเหมือน หรือการ์ตูนสุภาพ เช่น ข้อมูลเกี่ยวกับคนจะใช้ภาพคน หรือรูปทรงของคน ข้อมูลเกี่ยวกับรถยนต์จะใช้ภาพรถยนต์ เป็นต้น วิธีการนำเสนอก็เหมือนกับการนำเสนอโดยกราฟเส้น

1.3.6 การนำเสนอข้อมูลโดยกราฟวงกลม (pie-chart presentation) บางครั้งเรียกว่า แผนภาพกง เป็นวิธีนำเสนอโดยบรรจุข้อมูลลงในวงกลม ซึ่งมุมรอบจุดศูนย์กลางเป็น 360 องศา เทียบกับจำนวนข้อมูลทั้งหมดที่มีอยู่ร้อยละ 100 โดยนำข้อมูลแต่ละประเภทที่จะนำเสนอเทียบเป็นร้อยละและค่ามุมรอบจุดศูนย์กลางวงกลมครบ 360 องศา ดังภาพที่ 1.2



ภาพที่ 1.2 การนำเสนอข้อมูลโดยกราฟวงกลม

1.4 การแจกแจงความถี่ของข้อมูล

เมื่อมีการเก็บข้อมูลใด ๆ ที่สนใจศึกษาหรือเหตุการณ์ใด ๆ ที่ให้ความสนใจและได้มีการรวบรวมข้อมูลโดยวิธีการใดก็ตาม เช่น การสัมภาษณ์ การสำรวจ การทดลอง การสังเกตหรืออื่น ๆ เรานิยมนำมาจัดระบบด้วยการจัดกลุ่มข้อมูลตามค่ามากน้อย เรียกกลุ่มเหล่านี้ว่า ชั้น (class) ช่วงชั้นหรืออันตรภาคชั้น (class interval) เพื่อดูธรรมชาติข้อมูลที่มีอยู่ว่ามีลักษณะการแจกแจง (distributions) เป็นอย่างไรซึ่งวิธีการนี้เรียกว่า การแจกแจงความถี่ (frequency distribution) ซึ่งโดยทั่วไปจะจัดกระทำในรูปของตารางแจกแจงความถี่อันประกอบด้วยสดมภ์ตัวแปร (column variable) สดมภ์รอยขีด (tally column) สดมภ์ความถี่ (frequency column) หรือ อื่น ๆ ที่เกี่ยวข้อง แล้วนำเสนอการแจกแจงความถี่ในลักษณะของฮิสโทแกรม (histogram) รูปหลายเหลี่ยมของความถี่ (frequency polygon) และกราฟความถี่ (frequency curve) ตามความเหมาะสมสำหรับตัวแปรนั้น ๆ การแจกแจงความถี่ของข้อมูลจัดแบ่งได้ เป็น 2 ประเภทดังนี้

1.4.1 การแจกแจงความถี่ของข้อมูลเชิงคุณภาพ ข้อมูลเชิงคุณภาพ เช่น กลุ่มสายงาน กลุ่มอาชีพ การนับถือศาสนา เพศ ความรู้สึกต่อสิ่งใดสิ่งหนึ่ง เป็นต้น ดังตัวอย่างตารางการแจกแจงความถี่ดังตารางที่ 1.2 ถึงตารางที่ 1.4

ตารางที่ 1.2 การแจกแจงความถี่ของข้อมูลเชิงคุณภาพ

เพศ	รอยขีด	ความถี่ (*f*)
ชาย	𝍸 𝍸	185
หญิง	𝍸 𝍸	115
	รวม	300

ตารางที่ 1.3 การแจกแจงความถี่ของข้อมูลเชิงคุณภาพที่ตัดสดมภ์รอยขีดออก

อาชีพ	ความถี่ (*f*)
รับราชการ	70
ค้าขาย	90
เกษตรกร	62
กรรมกร	53
อื่นๆ	25
รวม	300

ตารางที่ 1.4 การแจกแจงความถี่ของข้อมูลเกี่ยวกับระดับความพอใจที่ตัดสดมภ์รอยขีดออก

ระดับความพอใจ	ความถี่ (*f*)
พอใจมาก	22
ค่อนข้างพอใจมาก	48
เฉยๆ	51
ไม่พอใจ	47
ไม่พอใจอย่างมาก	22
รวม	190

1.4.2 การแจกแจงความถี่ของข้อมูลเชิงปริมาณ ปริมาณที่เกิดจากการวัด การนับ ไม่ว่าจะกระทำด้วยวิธีการใด ๆ จะแสดงขนาดหรือจำนวนตัวเลข การแจกแจงความถี่จะนำข้อมูลทั้งหมดมาจัดแบ่งเป็นกลุ่มย่อยหรือชั้นย่อย โดยมีขอบเขต (bounded) เป็นตัวเลขเริ่มต้นและสิ้นสุดของกลุ่มย่อยหรือชั้นย่อย เราเรียกว่า แบ่งเป็นช่วงชั้น หรืออันตรภาคชั้น ขนาดของช่วงชั้นหรือขนาดของอันตรภาคชั้น อาจเป็น 1 หน่วยหรือมากกว่าก็ได้ขึ้นอยู่กับข้อมูลนั้นจะมีลักษณะเป็นตัวแปรเต็มหน่วยหรือตัวแปรต่อเนื่อง กล่าวคือ กรณีข้อมูลตัวแปรเต็มหน่วยจะกำหนดช่วงชั้นเป็น 1 หน่วย ส่วนกรณีข้อมูลเป็นตัวแปรต่อเนื่อง นิยมกำหนดขนาดของช่วงชั้นมากกว่า 1 หน่วย โดยขอยกตัวอย่างการดำเนินการแจกแจงความถี่และตารางแจกแจงความถี่ของข้อมูลทั้งสองกรณีดังนี้

1) ข้อมูลที่มีค่าเต็มหน่วย เช่น ข้อมูลเป็นจำนวนบุตรจาก 50 ครอบครัว การจดบันทึกจำนวนบุตรแต่ละครอบครัวเป็นตัวแปร X จะพบว่า ตัวแปร X เป็นตัวแปรวิยุตหรือตัวแปรเต็มหน่วยเพราะว่าเราสามารถนับจำนวนบุตร ตั้งแต่ 1, 2, 3,… สมมติว่าเราบันทึกค่าตัวแปร x จำนวน 50 ค่าได้เป็นดังนี้

2, 3, 2, 0, 5, 1, 4, 0, 1, 1,

7, 2, 1, 2, 5, 4, 5, 1, 2, 0 ,

0, 1, 2, 2, 1, 0, 2, 5, 4, 3,

3, 3, 3, 0, 4, 3, 3, 5, 7, 3,

2, 2, 0, 2, 3, 5, 0, 0, 4, 2.

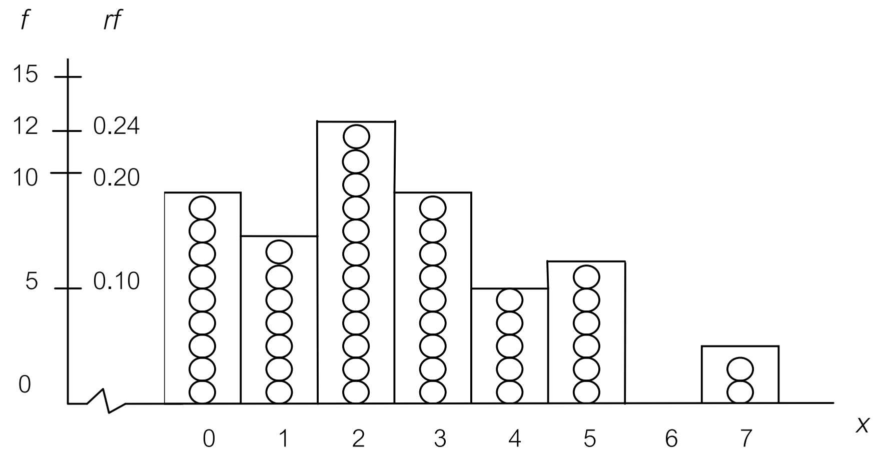
ในการตรวจนับค่าจากข้อมูลดังกล่าวข้างต้น สามารถเขียนอยู่ในรูปของตารางการแจกแจงความถี่ที่ประกอบด้วยสดมภ์ 3 สดมภ์คือ จำนวนบุตร รอยขีด ความถี่ที่แสดงการตรวจนับและสดมภ์ที่ 4 เป็นการแสดงความถี่สัมพัทธ์ (relative frequency) ดังแสดงในตารางที่ 1.5 เช่น จำนวนจำนวนบุตรเป็น 3 มีความถี่สัมพัทธ์เท่ากับ $\frac{9}{50}$ เท่ากับ 0.18 หรือร้อยละ 18 ซึ่งโดยทั่วไปความถี่สัมพัทธ์ของ x ได้จากการนำความถี่ของ x หารด้วยจำนวนข้อมูลทั้งหมดของการแจกแจงนั้น ถ้ากำหนด f เป็นความถี่ของ x โดย n เป็นจำนวนข้อมูลทั้งหมด $rf(x)$ เป็นความถี่สัมพัทธ์ของ x จะได้สมการ (1–1)

$$rf(x) = \frac{f}{n} \qquad \ldots(1\text{-}1)$$

เราสามารถนำเสนอการแจกแจงความถี่และการแจกแจงความถี่สัมพัทธ์ควบคู่กันโดยกำหนดแนวนอนและแนวตั้งหรือสดมภ์ได้ดังตารางที่ 1.5 จากข้อมูลในสดมภ์ความถี่สัมพัทธ์เราอาจแสดงเป็นร้อยละได้ เช่น 0.10 คือ ร้อยละ 10 และ 0.20 คือ ร้อยละ 20 เป็นต้น และจากตารางที่ 1.5 เมื่อนำเสนอข้อมูลในรูปของกราฟจะได้ดังแสดงในภาพที่ 1.3

ตารางที่ 1.5 การแจกแจงความถี่และความถี่สัมพัทธ์ของจำนวนบุตร

จำนวนบุตร (x)	รอยขีด	ความถี่ (f)	ความถี่สัมพัทธ์ (rf)
0	~~////~~ ////	9	0.18
1	~~////~~ //	7	0.14
2	~~////~~ ~~////~~ //	12	0.24
3	~~////~~ ////	9	0.18
4	~~////~~	5	0.10
5	~~////~~ /	6	0.12
6	…	0	0
7	//	2	0.04
	รวม	$n = 50$	1.00



ภาพที่ 1.3 ความสัมพันธ์ของความถี่และความถี่สัมพัทธ์ของจำนวนบุตร

2) ข้อมูลที่มีค่าต่อเนื่อง เช่น การจดบันทึกน้ำหนักนักศึกษาชายจำนวน 50 คน ในหน่วยกิโลกรัม จะพบว่าตัวแปรอาจมีค่าเป็น 64.00, 64.32, 65.00, 65.50, 60.00,... กิโลกรัม เข้าลักษณะตัวแปรแบบต่อเนื่อง การกำหนดช่วงของตัวแปรในแต่ละกลุ่มหรือชั้น นิยมกำหนดช่วงมากกว่า 1 หน่วย หรือขนาดช่วงชั้นมากกว่า 1 หน่วย สำหรับช่วงชั้นที่มีขนาดมากกว่า 1 หน่วย ควรมีลักษณะดังต่อไปนี้

ก. ชั้นหนึ่ง ๆ ควรมีวัตถุประสงค์เพื่อ 1 ค่าสังเกตหรือตัวแปร

ข. แต่ละช่วงชั้นควรมีขนาดเท่ากันหมด เช่นชั้นที่ 1 กำหนด 30 – 39 ช่วงชั้นขนาด 10 ชั้นต่อ ๆ ไปถึงชั้นสุดท้ายก็ควรมีช่วงชั้นขนาด 10 ด้วย

ค. ชั้นต่าง ๆ ควรต่อเนื่องกันหมดตลอดการแจกแจง เช่น ชั้นที่ 1 กำหนดให้เป็น 30 – 39 ชั้นที่ 2 จะเป็น 40 - 49 และชั้นที่ 3 จะเป็น 50 - 59 เป็นต้น

ง. ขอบเขตของชั้น (class boundary) มีความหมายรวมถึงขีดจำกัดของชั้น (class limit) เช่น ชั้นที่ 2 กำหนดขอบเขตของชั้นเป็น 40 - 49 จะมีขีดจำกัดของชั้นเป็น 39.5 – 49.5 เราเรียกค่าต่ำสุด (39.5) ว่าขีดจำกัดล่าง (lower limit) และเรียกค่าสูงสุด (49.5) ว่าขีดจำกัดบน (upper limit)

จ. จุดกลางชั้น (mid point or middle point) คือตัวแทนของแต่ละชั้นที่เป็นค่ากึ่งกลางของแต่ละชั้น เช่น จุดกลางชั้นที่ของ 40 – 49 คือ $\frac{40+49}{2}$ เท่ากับ 44.5 หรือ $\frac{39.5+49.5}{2}$ เท่ากับ 44.5 นั่นหมายความว่า 44.5 สามารถใช้เป็นตัวแทนข้อมูลหรือตัวแปรที่มีค่าระหว่าง 40 – 49 หรือ 39.5 – 49.5 สำหรับวิธีดำเนินการแจกแจงความถี่ สามารถดำเนินการตามลำดับขั้นตอนได้ดังต่อไปนี้

(1) ขั้นที่ 1 หาผลต่างระหว่างตัวแปรค่าสูงสุดกับตัวแปรค่าต่ำสุด ถ้ากำหนด x_n เป็นค่าตัวแปรสูงสุด x_1 เป็นตัวแปรต่ำสุด และ R เป็นผลต่างจะได้สมการ (1-2)

$$R = x_n - x_1 \qquad \ldots(1\text{-}2)$$

(2) ขั้นที่ 2 กำหนดจำนวนชั้น โดยทั่วไปจะกำหนดประมาณ 4 ชั้น แต่ไม่เกิน 12 ชั้น ทั้งนี้ควรกำหนดชั้นให้เหมาะสมสำหรับการนำเสนอการแจกแจงความถี่ในโอกาสต่อไป หรืออาจจะกำหนดจำนวนชั้นโดยประมาณด้วยสมการ (1-3) (Watson & Croft, 1990, p. 24)

$$c = 1 + 3.3\,(\log n) \qquad \ldots(1\text{-}3)$$

เมื่อ n คือ จำนวนข้อมูลหรือจำนวนตัวแปรทั้งหมด

c คือ จำนวนชั้น

log คือ ลอการิทึมสามัญ จากตารางภาคผนวกที่ 2

(3) ขั้นที่ 3 การหาขนาดของอันตรภาคชั้น (I) หาได้จากสูตร

$$I = \frac{R}{c} \qquad \ldots(1\text{-}4)$$

(4) ขั้นที่ 4 หาค่าเริ่มต้นของชั้นที่ 1 (L)

$$L = x_1 - \frac{1}{2}(cI - R) \qquad \ldots(1\text{-}5)$$

(5) ขั้นที่ 5 ตรวจนับจำนวนตัวเลขของแต่ละชั้นโดยใช้รอยขีด

ตัวอย่างที่ 1.2 จงเขียนตารางการแจกแจงความถี่ของความสูงของเด็กอายุ 1 ขวบ จำนวน 50 คน ที่มีหน่วยเป็นเซนติเมตรดังนี้

60	33	85	52	65	77	84	65	57	74
71	81	35	50	35	64	74	47	68	54
80	41	61	91	55	73	59	53	45	77
41	78	55	48	69	85	67	39	76	60
94	66	98	66	73	42	65	94	89	88

วิธีทำ

ขั้นที่ 1 หาผลต่างระหว่างค่าตัวแปรค่าสูงสุดและค่าต่ำสุด : R

$R = x_n - x_1$

$x_n = 98$ และ $x_1 = 33$

$R = 98 - 33 = 65$

ขั้นที่ 2 กำหนดจำนวนชั้น : c

$c = 1 + 3.3\,(log\ n)$

$n = 50$ และ $log\ 50 = 1.699$

$c = 1 + 3.3\,(1.699)$

$= 6.607$

≈ 7

ขั้นที่ 3 หาจำนวนอันตรภาคชั้น : I

$$I = \frac{R}{c}$$

$R = 65$ และ $c = 7$

$$I = \frac{65}{7}$$

$= 9.25 = 10$

ขั้นที่ 4 หาค่าเริ่มต้นของชั้นที่ 1

$$L = x_1 - \frac{1}{2}(cI - R)$$

$$x_1 = 33 \quad , \quad c = 7$$

$$I = 10 \quad , \quad R = 65$$

$$L = 33 - \frac{1}{2}[(7)(10) - 65]$$

$$L = 33 - 2.5$$

$$= 30.5$$

$$\approx 30$$

ขั้นที่ 5 ตรวจนับจำนวนตัวเลขของแต่ละชั้น โดยใช้รอยขีด

ตารางที่ 1.6 การแจกแจงความถี่และความถี่สัมพัทธ์ของความสูงของเด็กอายุ 1 ขวบ

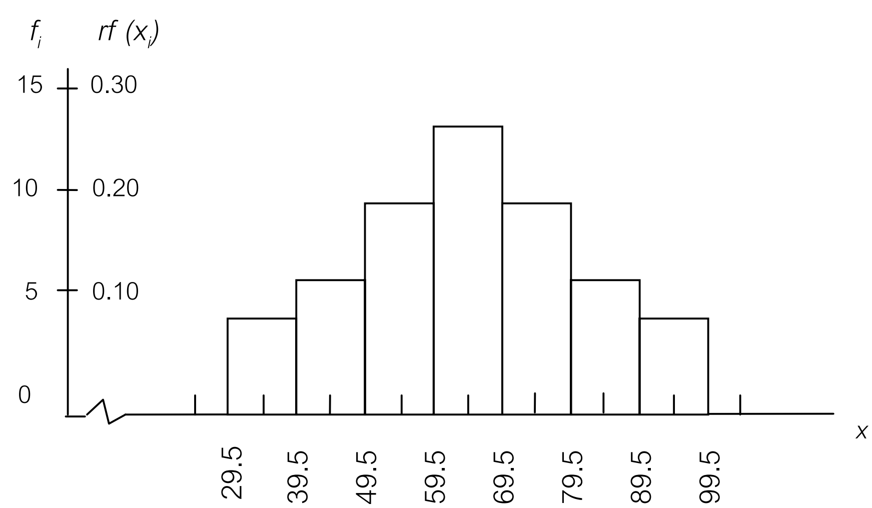
ความสูง(เซนติเมตร)	รอยขีด	ความถี่ (*f*)	ความถี่สัมพัทธ์ (*rf*)
30-39	////	4	0.08
40-49	~~////~~ /	6	0.12
50-59	~~////~~ ///	8	0.16
60-69	~~////~~ ~~////~~ //	12	0.24
70-79	~~////~~ ////	9	0.18
80-89	~~////~~ //	7	0.14
90-99	////	4	0.08
	รวม	$n = 50$	1.00

ได้ตารางการแจกแจงความถี่เป็นดังตารางที่ 1.6

1.5 การนำเสนอการแจกแจงความถี่และความถี่สัมพัทธ์ของข้อมูล

เมื่อดำเนินการแจกแจงความถี่และความถี่สัมพัทธ์แล้ว จะเห็นว่าสดมภ์ความถี่บอกจำนวนตัวแปรในแต่ละชั้น ซึ่งแต่ละชั้นนี้คือข้อมูลและมักจะเป็นปริมาณพร้อมด้วยหน่วย เช่น ความสูง (เซนติเมตร) คะแนนสอบ (คะแนน) น้ำหนัก (กรัม กิโลกรัม หรือ ปอนด์) หรืออาจใช้สัญลักษณ์แทนตัวแปรด้วย x_i การนำเสนอการแจกแจงความถี่และความถี่สัมพัทธ์ จึงเป็นการแสดงความสัมพันธ์ระหว่างค่าข้อมูลหรือค่าตัวแปรกับความถี่และ/หรือความถี่สัมพัทธ์ด้วยกราฟหรือแผนภูมิแบบต่าง ๆ โดยกำหนดแกนนอนเป็นค่าตัวแปร x แกนตั้งเป็นความถี่และ/หรือความถี่สัมพัทธ์ สำหรับกราฟการนำเสนอความถี่และความถี่สัมพัทธ์มีอยู่ 3 แบบคือ

1.5.1 ฮิสโทแกรม เป็นการสร้างกราฟแท่ง ความสูงของแท่งเป็นความถี่และ/หรือความถี่สัมพัทธ์ ความกว้างของแต่ละแท่งเป็นขนาดของแต่ละอันตรภาคชั้น และเพื่อไม่ให้เกิดช่องว่างระหว่างแต่ละแท่งของกราฟ จึงใช้ขีดจำกัดล่างและขีดจำกัดบนของชั้นเป็นตัวกำหนดความกว้างของแท่ง ทั้งนี้เพราะขีดจำกัดบนของชั้นใด ๆ จะเป็นขีดจำกัดบนของชั้นถัดไปด้วยดังภาพที่ 1.4

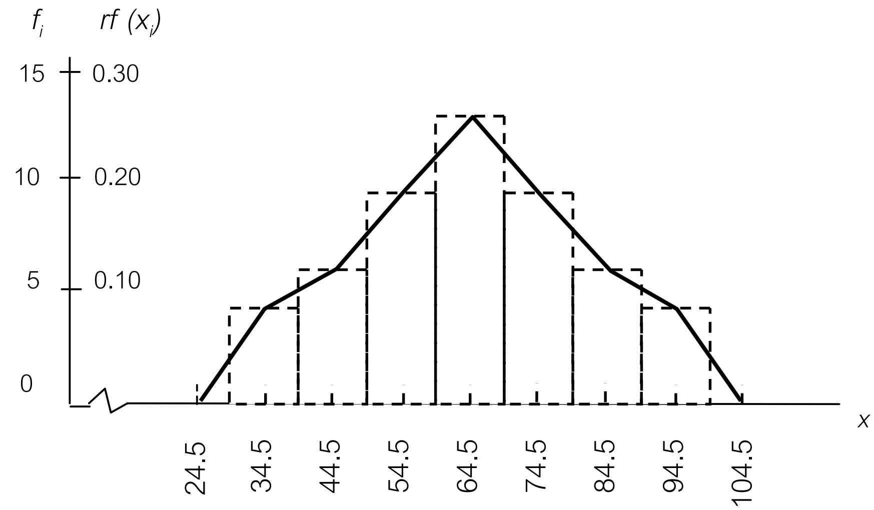


ภาพที่ 1.4 ฮิสโทแกรมการนำเสนอความถี่และความถี่สัมพัทธ์ของข้อมูลจากตัวอย่างที่ 1.2

1.5.2 รูปหลายเหลี่ยมของความถี่ รูปหลายเหลี่ยมของความถี่เกิดจากการลากเส้น เชื่อมโยงระหว่างจุดกลางยอดของแท่งสี่เหลี่ยมผืนผ้าของฮิสโทแกรมทุกแท่ง โดยปกติจะกำหนดจุดกลางชั้นของชั้นที่ 1 และจุดกลางชั้นของชั้นถัดจากชั้นสูงสุดของตารางแจกแจงความถี่เพื่อให้ได้รูปหลายเหลี่ยมของความถี่สมบูรณ์ เพราะความถี่และความถี่สัมพัทธ์ของชั้นดังกล่าวเป็นศูนย์ จึงทำให้รูปหลายเหลี่ยมของความถี่ปกคลุมพื้นที่เหนือแกนและแสดงถึงรูปทรงการแจกแจงอันมีประโยชน์สำหรับการวิเคราะห์ทางสถิติและการปรับปรุงรูปหลายเหลี่ยมของความถี่ อาจเริ่มด้วยการสร้างฮิสโทแกรมก่อนแล้วลากเส้นเชื่อมโยงระหว่างจุดกลางยอดแท่งเริ่มตั้งแต่ จุดกลางชั้นไปจนถึงชั้นถัดจากชั้นสุดท้ายที่หามาเพิ่ม หรืออาจกำหนดจุดกลางชั้นของทุกชั้นบนแนวนอนกับจุดความถี่แล้วลากเส้นเชื่อมโยงจะได้รูปหลายเหลี่ยมความถี่เช่นเดียวกัน

1.5.3 โค้งความถี่ เกิดจากการปรับรูปหลายเหลี่ยมของความถี่ โดยมิได้เสียพื้นที่รูปหลายเหลี่ยมของความถี่ปกคลุมอยู่เสียไปเราเรียกการทำเช่นนี้ว่าการปรับโค้งให้เรียบ (smoothing a curve) ทำได้โดยการปรับความถี่การแจกแจงเพื่อให้เกิดวามโค้งหรือลดการหักเหลี่ยมให้น้อยลงหรือไม่มีเลย วิธีการปรับโค้งให้เรียบ เช่น ข้อมูลจากตัวอย่างที่ 1.2 ชั้นที่ 3 เป็น 50 – 59 ความถี่ก่อนปรับเท่ากับ 8 ความถี่ที่ปรับแล้วของชั้นนี้จะเป็น $\frac{6+8+12}{3}$ เท่ากับ 8.66 จะเห็นได้ว่า 6 และ 12 มาจากความถี่ของชั้นใกล้เคียงคือชั้นที่ 2 และชั้นที่ 4 ตามลำดับ ดังแสดงด้วยกราฟในภาพที่ 1.5 และตารางที่ 1.7 จะแสดงความถี่ที่ปรับแล้ว สังเกตในชั้น 20 – 29 และชั้น 100 – 109 ซึ่งเป็นชั้นก่อนหน้าชั้นที่ 1 และชั้นถัดจากชั้นสุดท้าย ความถี่ที่ปรับแล้วของ 20 – 29 เท่ากับ $\frac{0+0+4}{3}$ เท่ากับ 1.33 ส่วนความถี่ที่ปรับแล้วของ 100 - 109 เท่ากับ $\frac{4+0+0}{3}$ เท่ากับ 1.33 ส่วนชั้นอื่น ๆ การหาความถี่ที่ปรับแล้วทำเช่นเดียวกับชั้นที่ 3 เพราะมีชั้นที่ใกล้เคียงลักษณะเดียวกัน

เมื่อปรับโค้งให้เรียบแล้วจะได้เส้นโค้งลักษณะต่าง ๆ ที่แตกต่างกันออกไปและลักษณะความถี่เหล่านี้จะถูกกำหนดและตั้งชื่อลักษณะการแจกแจงของ X เช่น การแจกแจงปรกติ (norminal distribution or symmetric distribution) การแจกแจงเบ้ทางบวกของ X (positively skewed distribution) การแจกแจงเบ้ทางลบของ X (negatively skewed distribution) การแจกแจงทวิฐานนิยมของ X (bimodal distribution) หรืออื่น ๆ ดังแสดงในภาพที่ 1.6 ภาพที่ 1.7 ภาพที่ 1.8 ภาพที่ 1.9 และภาพที่ 1.10 เป็นต้น



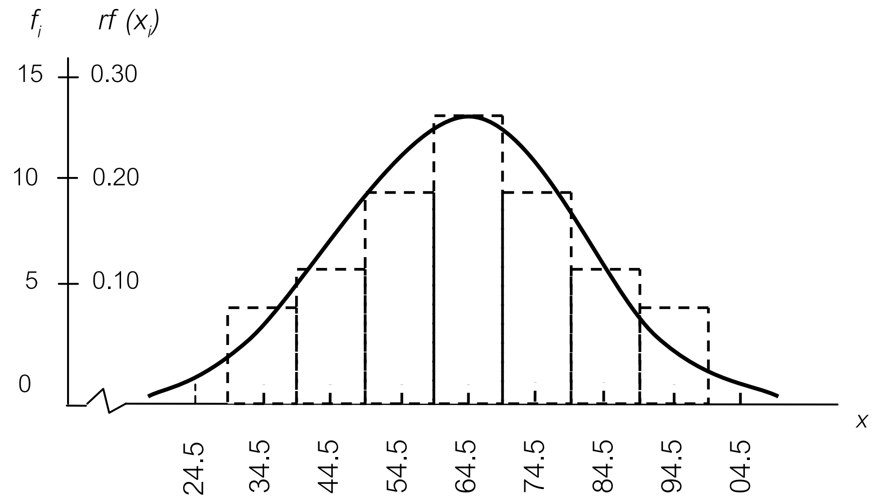
ภาพที่ 1.5 รูปหลายเหลี่ยมของความถี่นำเสนอความถี่และ
ความถี่สัมพัทธ์ของข้อมูลจากตัวอย่างที่ 1.2

ตารางที่ 1.7 การเปรียบเทียบความถี่เดิมและความถี่ที่ปรับแล้วจากข้อมูลตัวอย่างที่ 1.2

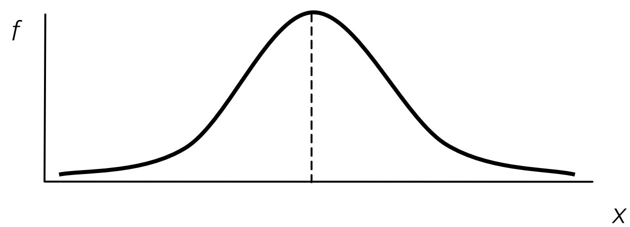
ชั้น	ความถี่เดิม (*f*)	ความถี่ที่ปรับแล้ว
20 – 29	0	1.33
30 –39	4	0.33
40 –49	6	6.00
50 – 59	8	8.67
60 –69	12	9.67
70 – 79	9	9.33
80 –89	7	6.67
90 –99	4	3.67
100 – 109	0	1.33
รวม	50	50.00

Xxx(**หน้านี้ไม่ใช้แต่ต้องพิมพ์**)

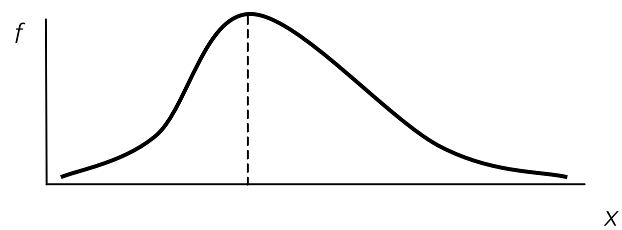
xxxxxxxxxxxxxxxxx



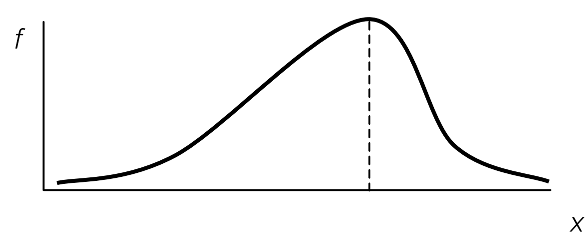
ภาพที่ 1.6 โค้งความถี่ข้อมูลจากความถี่ที่ปรับแล้วจากตารางที่ 1.7



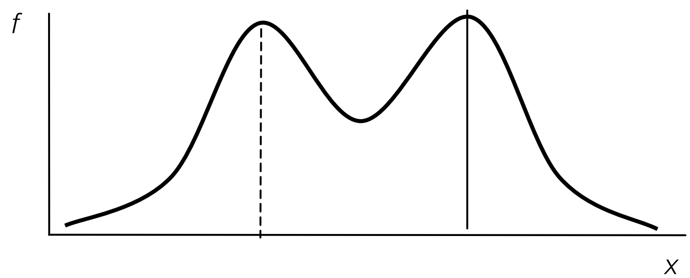
ภาพที่ 1.7 การแจกแจงปรกติของ X



ภาพที่ 1.8 การแจกแจงเบ้ทางบวกของ X



ภาพที่ 1.9 การแจกแจงเบ้ทางลบของ X



ภาพที่ 1.10 การแจกแจงทวิฐานนิยมของ *X*

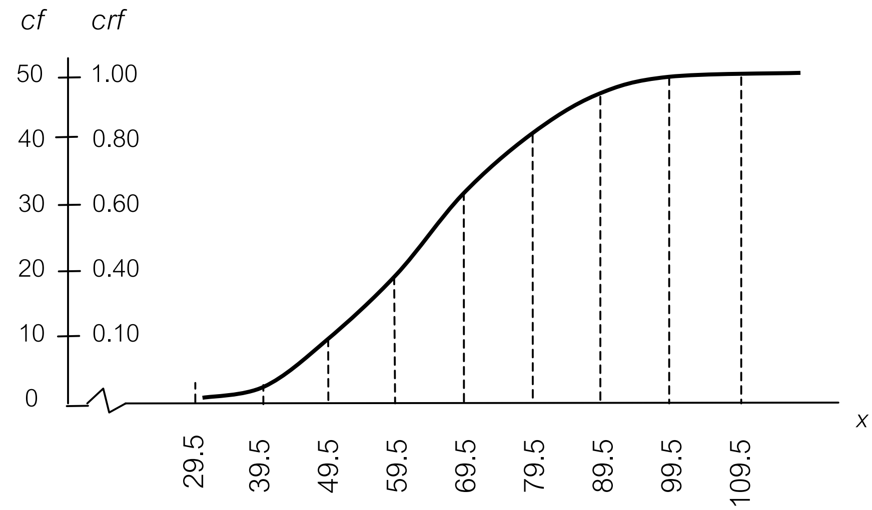
1.6 ความถี่สะสมและความถี่สะสมสัมพัทธ์ของข้อมูล

ความถี่สะสม (cumulative frequency : *cf*) คือการรวบรวมความถี่ของแต่ละชั้นนับตั้งแต่ชั้นที่มีค่าน้อยไปถึงชั้นที่มีค่ามากที่สุดและจะพบว่าความถี่สะสมของชั้นสุดท้ายเท่ากับจำนวนข้อมูลทั้งหมดดังแสดงในตารางที่ 1.8 สำหรับความถี่สะสมสัมพัทธ์ (relative comulative frequency : *crf*) เป็นการรวมความถี่สัมพัทธ์ ซึ่งผลรวมจะได้เท่ากับ 1.00 ดังแสดงในตารางที่ 1.8

ตารางที่ 1.8 การเปรียบเทียบความถี่สะสมและความถี่สะสมสัมพัทธ์ของข้อมูล

ชั้น	*f*	*cf*	*rf*	*crf*
30 –39	4	4	0.08	0.08
40 –49	6	10	0.12	0.20
50 – 59	8	18	0.16	0.36
60 –69	12	30	0.24	0.06
70 – 79	9	39	0.18	0.78
80 –89	7	46	0.14	0.92
90 –99	4	50	0.08	1.00
	50		1.00	

สำหรับการนำเสนอความถี่สะสมและความถี่สะสมสัมพัทธ์ ในรูปของกราฟ จะกำหนดแกนนอนเป็นข้อมูล แกนตั้งเป็นความถี่สะสมและ/หรือความถี่สะสมสัมพัทธ์ ดังภาพที่ 1.11



ภาพที่ 1.11 กราฟการนำเสนอความถี่สะสมและความถี่สะสมสัมพัทธ์ ของข้อมูลจากตารางที่ 1.8

1.7 บทสรุป

เมื่อกล่าวถึงข้อมูลจะทำให้นึกถึงการบอกเรื่องราวหรือเหตุการณ์ใดเหตุการณ์หนึ่งที่มีการปรับเปลี่ยนอยู่ตลอดเวลา และต้องมีการจัดเก็บข้อมูลต่าง ๆ เหล่านั้นไว้ สำหรับทางสถิติที่กล่าวถึงในบทนี้ ซึ่งสรุปข้อมูลและการแจกแจงความถี่ได้ดังนี้

1.7.1 ข้อมูลเชิงสถิติ จะเป็นตัวเลขที่แสดงข้อเท็จจริงที่เกิดจากการประมวลผลเบื้องต้นด้วยการวัดหรือการนับ ซึ่งอาจเรียกข้อมูลสถิติสำหรับการคำนวณหรือนิยามเกี่ยวกับประชากรหรือสิ่งตัวอย่าง เป็นค่าสังเกต ตัวแปร ตัวแปรสุ่ม และค่าตอบสนอง

1.7.2 แหล่งข้อมูล ที่ใช้คือข้อมูลปฐมภูมิและข้อมูลทุติยภูมิ

1.7.3 การนำเสนอข้อมูลเบื้องต้น มีวิธีการนำเสนอได้ดังนี้

1) การนำเสนอข้อมูลเป็นบทความ

2) การนำเสนอข้อมูลโดยตาราง

3) การนำเสนอข้อมูลโดยกราฟเส้น

4) การนำเสนอข้อมูลโดยกราฟแท่ง

5) การนำเสนอข้อมูลโดยกราฟรูปภาพ

6) การนำเสนอข้อมูลโดยกราฟวงกลม

1.7.4 การแจกแจงความถี่ข้อมูล เป็นการจัดระบบข้อมูล โดยการจัดแบ่งเป็นกลุ่ม ๆ แบ่งออกเป็น

1) การแจกแจงความถี่ข้อมูลเชิงคุณภาพ

2) การแจกแจงความถี่ข้อมูลเชิงปริมาณ

สำหรับการแจกแจงความถี่ จะต้องมีการการจัดชั้นข้อมูล การหาความถี่สัมพัทธ์ ความถี่สะสม

1.7.5 การนำเสนอการแจกแจงความถี่และความถี่สัมพัทธ์ เมื่อมีการแจกแจงความถี่แล้วจะต้องมาจัดกระทำโดยการเขียนกราฟ ซึ่งมีดังนี้

1) ฮิสโทแกรม

2) รูปหลายเหลี่ยมความถี่

3) โค้งความถี่

1.8 คำถามทบทวน

1. การบันทึกรายจ่ายประจำวันของท่านจัดเป็นข้อมูลเชิงสถิติหรือไม่ อธิบายมาพอสังเขป
2. จากประสบการณ์ของท่านจงแยกสิ่งที่กล่าวต่อไปนี้ออกเป็นตัวแปรวิยุตหรือตัวแปรต่อเนื่อง

    ก. จำนวนวันลาในรอบปีของพนักงาน 20 คน

    ข. ระยะเวลา(ชั่วโมง)ที่เครื่องจักรหยุดการผลิตฉุกเฉินในรอบ 180 วัน

    ค. จำนวนลูกค้าที่เข้าใช้บริการร้านอาหารแต่ละวันเป็นเวลา 30 วัน

    ง. น้ำหนักแรกเกิดของทารก 500 คน

3. จงนำเสนอข้อมูลจากตารางที่ 1.1 ด้วยกราฟแท่ง
4. จงนำเสนอข้อมูลจากตารางที่ 1.3 ด้วยกราฟวงกลม
5. จากตารางที่ 1.9 แสดงความถี่และความถี่สะสมของคะแนนสอบวิชาสถิติธุรกิจจาก

นักศึกษาจำนวน 20 คน แล้วตอบคำถามต่อไปนี้

ตารางที่ 1.9 ความถี่และความถี่สะสมของคะแนนสอบวิชาสถิติธุรกิจ

ช่วงคะแนน	*f*	*cf*
40 – 44	2	2
45 - 49	0	2
50 – 54	6	8
55 – 59	8	16
60 – 64	4	20
	20	

ก. ช่วงคะแนนที่มีความถี่ 6 คือคะแนนช่วงใด

ข. คะแนนระหว่าง 39.5 – 59.5 มีความถี่เท่าใด

ค. คะแนนช่วงใดที่มีความถี่สะสม 8

ง. ช่วงคะแนนใดที่ไม่มีผู้สอบได้เลย

6. จงแจกแจงความถี่และจงหาค่าความถี่สะสมและความถี่สะสมสัมพัทธ์ของค่าสังเกตต่อไปนี้ โดยมีชั้นค่าน้อยที่สุดเป็น 10 ถึง 19

77	44	49	33	38	33	76	55	68	39
44	59	36	55	47	61	53	32	65	51
29	41	32	45	83	58	73	47	40	26
59	43	66	44	41	25	39	75	37	55
34	47	66	53	55	58	49	45	61	41
55	92	83	77	45	62	45	36	78	48
54	50	51	66	80	73	57	61	56	50
45	82	71	48	46	69	38	72	56	64
38	45	51	44	41	68	45	92	43	12
37	16	44	57	63	71	40	64	57	51

7. จากการเก็บข้อมูลความสูงของพนักงานหญิง 40 คน ในหน่วยเป็นเซนติเมตร ได้ผลดังนี้

138	164	150	132	144	125	149	157
146	158	140	147	136	148	153	144
168	126	138	176	163	119	154	165
146	173	142	147	135	153	140	135
161	145	135	142	150	156	145	128

ก. จงสร้างตารางแจกแจงความถี่ของข้อมูลโดยมีชั้นค่าน้อยที่สุดเป็น 117 ถึง 126

ข. จงสร้างฮิสโทแกรมของข้อมูล

ค. จงสร้างรูปหลายเหลี่ยมความถี่ของข้อมูล

บทที่ 2
การวัดแนวโน้มสู่ส่วนกลางของข้อมูล

เมื่อนำข้อมูลหรือค่าสังเกตของเรื่องราวใดเรื่องราวหนึ่งจำนวนหลาย ๆ ค่ามาดำเนินการแจกแจงความถี่ของค่าสังเกตจะพบลักษณะที่คล้ายคลึงกันอยู่สองประการ กล่าวคือ ประการแรก ค่าสังเกตในกลุ่มที่ศึกษามีความแตกต่างกัน (variation) ประการที่สอง ค่าสังเกตในกลุ่มที่ศึกษามีความโน้มเอียงสู่ส่วนกลาง หากเราเลือกค่าสังเกตมาค่าหนึ่งเพื่อใช้เป็นตัวแทนของค่าสังเกต ทั้งหมดในขอบเขตที่เราศึกษาอยู่นั้นแน่นอนที่สุดเราจะต้องเลือกตัวกลาง (medium) หรือตัวที่พบบ่อยที่สุดของข้อมูลเหล่านั้น การวัดแนวโน้มสู่ส่วนกลางของข้อมูลที่กล่าวในบทนี้จึงเป็นการนิยามถึงเครื่องมือหรือเครื่องวัดค่ากลางของข้อมูล

2.1 ผลรวมของข้อมูล

ก่อนกล่าวถึงการวัดแนวโน้มสู่ส่วนกลาง จำเป็นต้องทำความเข้าใจกับ ผลรวมของค่าสังเกตทั้งหมด หรือ การรวมยอดตัวแปร หรือ การหา ผลรวมของตัวแปร สำหรับผลรวมหรือยอดรวม ของค่าสังเกตของข้อมูลหรือของตัวแปร จำนวนมาก ๆ จะใช้สัญลักษณ์ Σ (capital sigma : ซิกมาซึ่งเป็นอักษรกรีก) อ่านว่า ผลรวม รวมยอด ผลบวก หรือ ซัม (sum)

ถ้า $x_1, x_2, x_3, ..., x_n$ เป็นค่าสังเกตจำนวน n ค่า ผลรวมของค่าสังเกตทั้งหมด คือ

$$\sum_{i=1}^{n} X_i = x_1 + x_2 + ...+ x_n$$

หรือ $$\textstyle\sum_{i=1}^{n} X_i = x_1 + x_2 + ...+ x_n \qquad ...(2-1)$$

$\sum_{i=1}^{n} X_i$ อ่านว่า ผลรวมของ X_i เมื่อ i มีค่าตั้งแต่ 1 ถึง n ซึ่งเท่ากับ x_1 รวมกับ x_2 รวมกับ x_3 .... รวมกับ x_n หรือกรณีหาผลรวมของทั้งหมดอาจใช้สัญลักษณ์ ΣX แทนผลรวม

ของ X ตั้งแต่ตัวแรกถึงตัวสุดท้าย ก็ได้

ตัวอย่างที่ 2.1 กำหนดตัวแปร X ดังต่อไปนี้ $x_1 = 2, x_2 = 3, x_3 = 4, x_4 = (-1)$

ก. จงหา $\sum_{i=1}^{4} X_i$

ข. จงหา $\sum_{i=2}^{4} X_i$

วิธีทำ

ก. $\sum_{i=1}^{4} X_i$ หมายถึง ผลรวมของ X_i , i ตั้งแต่ 1 ถึง 4

$$\sum_{i=1}^{4} X_i = x_1 + x_2 + x_3 + x_4$$

$$= 2 + 3 + 4 + (-1)$$

$$= 8$$

ข. $\sum_{i=2}^{4} X_i$ หมายถึง ผลรวมของ X_i , i ตั้งแต่ 2 ถึง 4

$$\sum_{i=2}^{4} X_i = x_2 + x_3 + x_4$$

$$= 3 + 4 + (-1)$$

$$= 6$$

ตัวอย่างที่ 2.2 กำหนดให้

$x_1 = 3$, $x_2 = (-2)$, $x_3 = 1$, $x_4 = 0$, $x_5 = 2$

$y_1 = 7$, $y_2 = (-4)$, $y_3 = 8$, $y_4 = 5$, $y_5 = (-3)$

$f_1 = 2$, $f_2 = 8$, $f_3 = 20$, $f_4 = 12$, $f_5 = 3$

จงหา ก. $\sum_{i=1}^{5} X_i$ ข. $\sum_{i=1}^{2} X_i Y_i$

ค. $\sum_{i=2}^{3} X_i^2$ ง. $\sum_{i=2}^{4} f_i X_i$

จ. $\sum_{i=4}^{5} (X_i - Y_i)$ ฉ. $\sum_{i=4}^{5} (X_i - Y_i)^2$

วิธีทำ

ก. $\sum_{i=1}^{5} X_i$ หมายถึง ผลรวมของ X_i, i ตั้งแต่ 1 ถึง 5

$$\sum_{i=1}^{5} X_i = x_1 + x_2 + x_3 + x_4 + x_5$$

$$= 3 + (-2) + 1 + 0 + 2$$

$$\sum_{i=1}^{5} X_i = 4$$

ข. $\sum_{i=1}^{2} X_i Y_i$ หมายถึง ผลรวมของ X_i คูณ Y_i , i ตั้งแต่ 1 ถึง 2

$$\sum_{i=1}^{2} X_i Y_i = x_1 y_1 + x_2 y_2$$

$$= (3)(7) + (-2)(-4)$$

$$\sum_{i=1}^{2} X_i Y_i = 21 + 8 = 29$$

ค. $\sum_{i=2}^{3} X_i^2$ หมายถึง ผลรวมของ X_i^2 , i ตั้งแต่ 2 ถึง 3

$$\sum_{i=2}^{3} X_i^2 = x_2^2 + x_3^2$$

$$= \quad (-2)^2 - (1)^2$$

$$\sum_{i=2}^{3} X_i^2 \quad = \quad 4 + 1$$

$$= \quad 5$$

ง. $\sum_{i=2}^{4} f_i X_i$ หมายถึง ผลรวมของ f_i คูณ X_i , i ตั้งแต่ 2 ถึง 4

$$\sum_{i=2}^{4} f_i X_i \quad = \quad f_2x_2 + f_3x_3 + f_4x_4$$

$$= \quad (8)(-2) + (20)(1) + (12)(0)$$

$$= \quad (-16) + 20 + 0$$

$$\sum_{i=2}^{4} f_i X_i \quad = \quad 4$$

จ. $\sum_{i=4}^{5} (X_i - Y_i)$ หมายถึง ผลรวมของ X_i ลบด้วย Y_i (หรือผลบวกของผลต่างระหว่าง X_i กับ Y_i) เมื่อ i มีค่าตั้งแต่ 4 ถึง 5

$$\sum_{i=4}^{5} (X_i - Y_i) \quad = \quad (x_4 - y_4) + (x_5 - y_5)$$

$$= \quad (0-5) + [2-(-3)]$$

$$= \quad (-5) + 5$$

$$\sum_{i=4}^{5} (X_i - Y_i) \quad = \quad 0$$

ฉ. $\sum_{i=4}^{5} (X_i - Y_i)^2$ หมายถึง ผลรวมของ X_i ลบด้วย Y_i ยกกำลังสอง (หรือผลบวกของผลต่างระหว่าง X_i กับ Y_i ยกกำลังสอง) i ตั้งแต่ 4 ถึง 5

$$\sum_{i=4}^{5}(X_i - Y_i)^2 \quad = \quad (x_4\text{-}y_4)^2 + (x_5\text{-}y_5)^2$$

$$= \quad (0\text{-}5)^2 + [2\text{-}(\text{-}3)]^2$$

$$= \quad 25 + 25$$

$$\sum_{i=4}^{5}(X_i - Y_i)^2 \quad = \quad 50$$

จากหลักเกณฑ์ที่กล่าวมา ถ้ากำหนดให้ X_i ,Y_i, , Z_i เป็นตัวแปรจำนวน n ตัว และ a เป็นค่าคงตัว (constant) สามารถนำไปประยุกต์ใช้กับการหาผลรวมของหลายตัวแปร ดังสมการ (2-2) ถึงสมการ (2-5)

$$\sum_{i=1}^{n}(X_i + Y_i + Z_i) \quad = \quad \sum_{i=1}^{n}X_i + \sum_{i=1}^{n}Y_i + \sum_{i=1}^{n}Z_i \qquad \text{…(2-2)}$$

$$\sum_{i=1}^{n}(X_i - Y_i - Z_i) \quad = \quad \sum_{i=1}^{n}X_i - \sum_{i=1}^{n}Y_i - \sum_{i=1}^{n}Z_i \qquad \text{…(2-3)}$$

$$\sum_{i=1}^{n}aX_i \quad = \quad a\sum_{i=1}^{n}X_i \qquad \text{…(2-4)}$$

$$\sum_{i=1}^{n}a \quad = \quad an \qquad \text{…(2-5)}$$

2.2 มัชฌิมเลขคณิตของข้อมูล

มัชฌิมเลขคณิต (arithematic mean) เป็นการวัดแนวโน้มสู่ส่วนกลางที่อาจเรียกชื่อและใช้สัญลักษณ์แตกต่างกันแต่เกิดจากข้อกำหนดอย่างเดียวกัน เช่น มัชฌิมเลขคณิต ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ย (mean) ส่วนเฉลี่ย (average) และ ค่าคาดหมาย (expectation value)

สำหรับสัญลักษณ์ที่ใช้แทนมัชฌิมเลขคณิตกรณีประชากรใช้ μ (อักษรกรีกชนิดตัวอักษรเล็ก ชื่อ mu : มิว.) และกรณีมัชฌิมเลขคณิตสิ่งตัวอย่างใช้ $\overline{X}$ (อ่านว่า เอกซ์บาร์) ส่วนกรณีค่าคาดคะเนซึ่งเป็นค่าเฉลี่ยความน่าจะเป็นใช้ $E(X)$ แทนค่าเฉลี่ยหรือค่าคาดหมายของ ตัวแปรสุ่ม X

สถิติพรรณนาหรือการมองภาพรวมของค่าสังเกตเป็นประชากรตามหลักทฤษฎี (theoretical model) ทางคณิตศาสตร์และเพื่อเป็นการทราบความหมายเบื้องต้นของมัชฌิมเลขคณิต จึงขอใช้ μ แทนมัชฌิมเลขคณิตซึ่งมีข้อกำหนดดังสมการ (2-6)

ถ้า $x_1, x_2, x_3, ... x_n$ เป็นค่าสังเกตจำนวน n ค่า แล้วจะได้ (Keller & Warrack, 2000, p. 90)

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad ...(2\text{-}6)$$

จากข้อกำหนดเบื้องต้นนี้ สามารถเปลี่ยนรูปการหาค่ามัชฌิมเลขคณิต ในกรณีต่าง ๆ ตามความเหมาะสม ซึ่งพิจารณาได้ดังต่อไปนี้

2.2.1 กรณีข้อมูลไม่ได้แจกแจงความถี่ (ungrouped data) หาค่ามัชฌิมเลขคณิตได้ดังสมการ (2-7) และสมการ (2-8)

$$\mu = \frac{x_1 + x_2 + x_3 + x_4 .... + x_n}{n} \qquad ...(2\text{-}7)$$

$$\mu = x_1\left(\frac{1}{n}\right) + x_2\left(\frac{1}{n}\right) + x_3\left(\frac{1}{n}\right) + x_4\left(\frac{1}{n}\right) + ... + x_n\left(\frac{1}{n}\right) \qquad ...(2\text{-}8)$$

ตัวอย่างที่ 2.3 จงหามัชฌิมเลขคณิตของ 8, 12, 4, 7 และ 5

วิธีทำ หาค่ามัชฌิมเลขคณิตได้จากสมการ (2-6) ถึง สมการ (2-8) ดังนี้

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

$$\mu = \frac{1}{5}(8+12+4+7+5)$$

$$\mu = \frac{1}{5}(36)$$

$$\mu = 7.2$$

$$\mu = \frac{x_1+x_2+x_3+x_4.........+x_n}{n}$$

$$= \frac{8+12+4+7+5}{5}$$

$$= \frac{36}{5} = 7.2$$

$$\mu = x_1\left(\frac{1}{n}\right)+x_2\left(\frac{1}{n}\right)+x_3\left(\frac{1}{n}\right)+x_4\left(\frac{1}{n}\right)+.......+x_n\left(\frac{1}{n}\right)$$

$$= 8\left(\frac{1}{5}\right)+12\left(\frac{1}{5}\right)+4\left(\frac{1}{5}\right)+7\left(\frac{1}{5}\right)+5\left(\frac{1}{5}\right)$$

$$= \left(\frac{8}{5}\right)+\left(\frac{12}{5}\right)+\left(\frac{4}{5}\right)+\left(\frac{5}{5}\right)$$

$$\mu = \frac{36}{5} = 7.2$$

2.2.2 กรณีมีการข้อมูลแจกแจงความถี่ (grouped data) ตามตารางที่ 2.1 ให้ x_j เป็นค่าสังเกตหรือตัวแปรของแต่ละชั้น ถ้าหากช่วงชั้นขนาดมากกว่า 1 หน่วย x_j จะเป็นจุดกลางชั้นหาค่ามัชฌิมเลขคณิตได้ดังสมการ (2-9)

$$\mu = \frac{\sum_{j=1}^{k} f_j x_j}{\sum_{j=1}^{k} f_j} \qquad \ldots(2\text{-}9)$$

ตารางที่ 2.1 วิธีการหาความถี่ของแต่ละชั้น

x_j	f_j
x_1	f_1
X_2	f_2
x_3	f_3
⋮	⋮
⋮	⋮
x_k	f_k
$\sum_{j=1}^{k} x_j$	$\sum_{j=1}^{k} f_j = n$

สำหรับ $\left(\frac{f_1}{n}\right)$ เป็นความถี่สัมพัทธ์ของ x_1 หรือ $rf(x_1)$

$\left(\frac{f_2}{n}\right)$ เป็นความถี่สัมพัทธ์ของ x_2 หรือ $rf(x_2)$

$\left(\frac{f_3}{n}\right)$ เป็นความถี่สัมพัทธ์ของ x_3 หรือ $rf(x_3)$

⋮

⋮

⋮

⋮

และ $\left(\frac{f_k}{n}\right)$ เป็นความถี่สัมพัทธ์ของ x_k หรือ $rf(x_k)$

หามัชฌิมเลขคณิตได้ดังดังสมการ (2-10)

$$\mu \quad = \quad x_1\left(\frac{f_1}{n}\right)+x_2\left(\frac{f_2}{n}\right)+x_3\left(\frac{f_3}{n}\right)+\ldots\ldots +x_n\left(\frac{f_k}{n}\right) \qquad \ldots(2\text{-}10)$$

หรือเขียนในรูปความถี่สัมพัทธ์เป็นดังสมการ (2-11)

$$\mu \quad = \quad x_1\, rf\,(x_1) + x_2\ rf\,(x_2) +x_3\, rf\,(x_3) \ldots\ldots\ldots\ldots + x_k\ rf\,(x_k) \qquad \ldots(2\text{-}11)$$

จากสมการ (2-10) อาจเขียนได้เป็นดังสมการ (2-12)

$$\mu \quad = \quad \sum_{j=1}^{k} x_j \;\; rf\,(x_j) \qquad \ldots(2\text{-}12)$$

ในการเลือกใช้สูตรสำหรับหาค่ามัชฌิมเลขคณิตควรคำนึงถึงความเหมาะสมและความรวดเร็วในการคิดคำนวณ

ตัวอย่างที่ 2.4 จากตารางที่ 2.2 จงหาค่าคะแนนเฉลี่ยของผลการสอบวิชาวิทยาศาสตร์ทั่วไปจากคะแนนเต็ม 30 คะแนน และมีนักศึกษาจำนวน 55 คน

ตารางที่ 2.2 ผลการสอบวิชาวิทยาศาสตร์ทั่วไป

คะแนน	28	27	26	25	24	23	22	21	20	19	18	17	16
จำนวนนักศึกษา	2	3	4	5	4	6	11	5	3	4	3	3	2

วิธีทำ จากสมการ (2-9)

$$\mu \quad = \quad \frac{\sum_{j=1}^{k} f_j\, x_j}{\sum_{j=1}^{k} f_j}$$

$$= \frac{1,220}{55}$$

$$= 22.18$$

คะแนนเฉลี่ยของผลการสอบวิชาวิทยาศาสตร์ทั่วไปมีค่าเท่ากับ 22.18 คะแนน

ตารางที่ 2.3 การแจกแจงความถี่ของผลการสอบวิชาวิทยาศาสตร์ทั่วไป

x_j	f_j	$x_j f_j$
28	2	56
27	3	81
26	4	104
25	5	125
24	4	96
23	6	138
22	11	242
21	5	105
20	3	60
19	4	76
18	3	54
17	3	51
16	2	32
รวม	$\sum_{j=1}^{k} f_j = 55$	$\sum_{j=1}^{k} f_j x_j = 1,220$

คะแนนเฉลี่ยผลการสอบวิชาวิทยาศาสตร์ทั่วไป เท่ากับ 22.18 คะแนน

เราอาจจะหาค่า μ อีกวิธีหนึ่งจากสมการ (2-12) ดังนี้

$$\mu = \sum_{j=1}^{k} x_j \ rf(x_j)$$

$$\mu = \frac{1,220}{55}$$

$$= 22.18$$

คะแนนเฉลี่ยของผลการสอบวิชาวิทยาศาสตร์ทั่วไปมีค่าเท่ากับ 22.18 คะแนน

ตารางที่ 2.4 การแจกแจงความถี่และความถี่สัมพัทธ์ของผลการสอบวิชาวิทยาศาสตร์ทั่วไป

x_j	f_j	$rf(x_j)$	$x_j\, rf(x_j)$
28	2	2/55	56/55
27	3	3/55	81/55
26	4	4/55	104/55
25	5	5/55	125/55
24	4	4/55	96/55
23	6	6/55	136/55
22	11	11/55	242/55
21	5	5/55	105/55
20	3	3/55	160/55
19	4	4/55	76/55
18	3	3/55	54/55
17	3	3/55	51/55
16	2	2/55	32/55
รวม	$\sum_{j=1}^{k} f_j = 55$	$\sum_{j=1}^{k} rf(x_j) = 1$	$\sum_{j=1}^{k} x_j \ rf(x_j) = \frac{1,220}{55}$

คะแนนเฉลี่ยผลสอบวิชาวิทยาศาสตร์ทั่วไป เท่ากับ 22.18

ตัวอย่างที่ 2.5 จงหาความสูงเฉลี่ยของนักเรียนชั้นอนุบาลจำนวน 41 คน จากข้อมูล ใน

ตารางที่ 2.5

ตารางที่ 2.5 ความสูงของนักเรียนชั้นอนุบาล

ความสูง (เซนติเมตร)	จำนวน
127-129	9
124-126	10
121-123	12
118-120	7
115-117	3
รวม	41

วิธีทำ (1) จากตารางที่ 2.5 นำมาแจกแจงความถี่จะได้ดังตารางที่ 2.6

ตารางที่ 2.6 การแจกแจงความถี่ความสูงของนักเรียนชั้นอนุบาล

ความสูง (เซนติเมตร)	x_j	f_j	$x_j f_j$
127-129	128	9	1,152
124-126	125	10	1,250
121-123	122	12	1,464

118-120	119	7	833
115-117	116	3	348
		$\sum_{j=1}^{k} f_j = 41$	$\sum_{j=1}^{k} f_j x_j = 5{,}047$

และจากสมการ (2-9) จะได้

$$\mu = \frac{\sum_{j=1}^{k} f_j x_j}{\sum_{j=1}^{k} f_j} = \frac{5,047}{41}$$

$$\mu = 123.10$$

ความสูงเฉลี่ยของนักเรียนเท่ากับ 123.10 เซนติเมตร

วิธีทำ (2) จากข้อมูลในตารางที่ 2.6 นำมาแจกแจงความถี่สะสมสัมพัทธ์ของข้อมูล

ตารางที่ 2.7 การแจกแจงความถี่สะสมสัมพัทธ์ความสูงของนักเรียนชั้นอนุบาล

ความสูง (เซนติเมตร)	x_j	f_j	$rf(x_j)$	$x_j\,rf(x_j)$
127-129	128	9	$\frac{9}{41}$	$\frac{1,152}{41}$
124-126	125	10	$\frac{10}{41}$	$\frac{1,250}{41}$
121-123	122	12	$\frac{12}{41}$	$\frac{1,464}{41}$
118-120	119	7	$\frac{7}{41}$	$\frac{833}{41}$
115-117	116	3	$\frac{3}{41}$	$\frac{348}{41}$
		$\sum_{j=1}^{k} f_j$ = 41	$\sum_{j=1}^{k} f_j x_j$ = 1	$\sum_{j=1}^{k} x_j\ rf(x_j) = \frac{5,047}{41}$

จากสมการ (2-12)

$$\mu = \sum_{j=1}^{k} x_j\ rf(x_j)$$

$$\mu = \frac{5,047}{41}$$

$$\mu \quad = \quad 123.10$$

ความสูงเฉลี่ยของนักเรียนเท่ากับ 123.10 เซนติเมตร

จะเห็นได้ว่าในการหาค่าเฉลี่ยไม่ว่าเราจะใช้สูตรในการคำนวณหรือใช้ตารางแจกแจงความถี่จะได้คำตอบเท่ากัน

ตัวอย่างที่ 2.6 จงหาเกรดเฉลี่ยของนักศึกษาผู้หนึ่งที่มีผลการสอบดังแสดงไว้ในตารางที่ 2.8

ตารางที่ 2.8 ผลการสอบของนักศึกษา

วิชา	หน่วยกิต	เกรด
คณิตศาสตร์	4	2
ชีวะวิทยา	4	0
สถิติ	3	4
สังคมวิทยา	3	2
	14	

วิธีทำ จากตารางที่ 2.8 จัดทำตารางการแจกแจงความถี่ ได้เป็นตารางที่ 2.9

ตารางที่ 2.9 การแจกแจงความถี่ของผลการสอบของนักศึกษา

วิชา	x_j	f_j	$x_j f_j$
คณิตศาสตร์	2	4	8
ชีวะวิทยา	0	4	0
สถิติ	4	3	12
สังคมวิทยา	2	3	6
		14	26

จากสมการ (2-9)

$$\mu = \frac{\sum_{j=1}^{k} f_j x_j}{\sum_{j=1}^{k} f_j}$$

$$\mu = \frac{26}{14}$$

$$\mu = 1.857$$

เกรดเฉลี่ยของนักศึกษาผู้นี้เท่ากับ 1.86

สำหรับตัวอย่างที่ 2.6 ค่าสังเกต X_j มีความถี่ f_j เป็นค่าสังเกตที่มีน้ำหนัก f_j จึงมักแทนน้ำหนักของ X_j ด้วย W_j (weight of X_j) และเรียกค่าเฉลี่ยที่ได้ในกรณีนี้ว่าค่าเฉลี่ยถ่วงน้ำหนัก ถ้าหาก X_1 มีน้ำหนัก W_1 , X_2 มีน้ำหนัก W_2 , X_3 มีน้ำหนัก W_3 …………*และ* X_k มีน้ำหนัก W_k แล้วค่าเฉลี่ยของ X_j จะเป็นสมการ (2-13) หรือสมการ (2–14)

$$\mu = \frac{\sum_{j=1}^{k} W_j X_j}{\sum_{j=1}^{k} W_j} \qquad ...(2\text{-}13)$$

หรือ

$$\mu = \frac{W_1 X_1 + W_2 X_2 + W_3 X_3 \cdots\cdots + W_k X_k}{W_1 + W_2 + W_3 + \cdots\cdots\cdots + W_k} \qquad ...(2\text{-}14)$$

นอกจากข้อกำหนดเบื้องต้นและการประยุกต์ใช้มัชฌิมเลขคณิตหรือค่าเฉลี่ยแล้วยังมีคุณสมบัติของมัชฌิมเลขคณิตหรือค่าเฉลี่ยของตัวแปรใด ๆ ที่น่าสนใจ เช่นถ้า X_1 , X_2 , X_3 ,…… X_n เป็นค่าสังเกตจำนวน n หาค่าเฉลี่ยได้เท่ากับ μ , a และ b เป็นค่าคงตัวใด ๆ จะได้คุณสมบัติที่น่าสนใจดังสมการ (2-15) ถึงสมการ (2-19)

$$\sum_{i=1}^{n}(X_i - \mu) \quad = \quad 0 \qquad \text{...(2-15)}$$

พิสูจน์ จากสมการ (2-3)

$$\sum_{i=1}^{n}(X_i - \mu) \quad = \quad \sum_{i=1}^{n}X_i - \sum_{i=1}^{n}\mu$$

จากสมการ (2-5) และ (2-6) จะได้

$$\sum_{i=1}^{n}(X_i - \mu) \quad = \quad n\mu - n\mu$$

$$= \quad 0$$

ถ้าค่าสังเกต Y_1 , Y_2 , Y_3 ,...... Y_n จำนวน n หาต้องการหาค่าเฉลี่ย ($\overline{Y}$) และพบว่า Y_i แต่ละตัวเขียนอยู่ในรูปของสมการได้ดังนี้

$$Y_i \quad = \quad a + bX_i$$

แล้ว

$$\overline{Y} \quad = \quad a + b\mu \qquad \text{...(2-16)}$$

พิสูจน์ $\sum_{i=1}^{n}Y_i \quad = \quad \sum_{i=1}^{n}(a + bX_i)$

จากสมการ (2-2) $= \quad \sum_{i=1}^{n}a + \sum_{i=1}^{n}bX_i$

$$= \quad an + b\sum_{i=1}^{n}X_i$$

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

จะได้ $\overline{Y} \quad = \quad a + b\mu$

ในทำนองเดียวกัน

ถ้า $Y_i = bX_i$ แล้วจะได้ $\bar{Y} = b\mu$ ...(2-17)

ถ้า $Y_i = a + X_i$ แล้วจะได้ $\bar{Y} = a + b\mu$ ...(2-18)

และ ถ้า $W_i = Y_i + X_i$ แล้วจะได้ $\bar{W} = \bar{Y} + \mu$ ...(2-19)

พิสูจน์

$$\sum_{i=1}^{n} W_i = \sum_{i=1}^{n} (Y_i + X_i)$$

$$= \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} X_i$$

จากสมการ (2-2) จะได้

$$\sum_{i=1}^{n} W_i = \frac{\sum_{i=1}^{n} Y_i}{n} + \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\bar{W} = \bar{Y} + \mu$$

คุณสมบัติที่กล่าวมา อาจนำไปใช้ได้ในหลายกรณี เช่น การแปลงข้อมูล การเปลี่ยนข้อมูลค่ามาก ๆ ให้อยู่ในรูป 10 ยกกำลังเช่น 1,000 เปลี่ยนเป็น $(1)(10^3)$ และ 1,500 เปลี่ยนเป็น $(1.5)(10^3)$เป็นต้น แต่หลักสำคัญข้อมูลที่เปลี่ยนแปลงไปจะต้องมีจำนวนเท่าเดิม สังเกตได้ดังตัวอย่างที่ 2.7

ตัวอย่างที่ 2.7 กำหนดค่าสังเกต X_i เป็น 2, 3, 4, 5, 6, 7, และ 8 หาค่าเฉลี่ยได้เท่ากับ 5 จงหาค่าเฉลี่ยของข้อมูลต่อไปนี้

ก. 25, 35, 45, 55, 65, 75 และ 85.

ข. 200, 300, 400, 500, 600, 700, และ 800.

ค. 11, 12, 13, 14, 15, 16, และ 17.

ง. 225, 335, 445, 555, 665, 775, และ 885

วิธีทำ

ก. ค่าของ 25, 35, 45, 55, 65, 75, และ 85 อาจจะเขียนได้อีกรูปแบบหนึ่งคือ 5+(10)(2), 5+(10)(3), 5+(10)(4), 5+(10)(5), 5+(10)(6), 5+(10)(7) และ 5+(10)(8) ซึ่งเขียนสมการได้เป็น

$$Y_i = a + bX_i$$

แล้ว จากสมการ (2-16) จะได้ว่า

$$\overline{Y} = a + b\mu$$

เมื่อ $a = 5$, $b = 10$ และ $\mu = 5$ จะได้

$$\overline{Y} = 5 + (10)(5)$$

$$= 55$$

ค่าเฉลี่ยของข้อมูลมีค่าเท่ากับ 55

ข. ค่าของ 200, 300, 400, 500, 600, 700, และ 800 อาจจะเขียนได้อีกรูปแบบหนึ่งคือ 100(2), 100(3), 100(4), 100(5), 100(6), 100(7), และ 100(8) และเขียนสมการได้เป็น

$$Y_i = bX_i$$

แล้วจากสมการ (2-17) จะได้ว่า

$\overline{Y} = b\mu$ เมื่อ $b = 100$ และ $\mu = 5$

$$\overline{Y} = 100\ (5)$$

$$\overline{Y} = 500$$

ค่าเฉลี่ยของข้อมูลมีค่าเท่ากับ 500

ค. ค่าของ 11, 12, 13, 14, 15, 16, และ 17 อาจจะเขียนได้อีกรูปแบบหนึ่ง คือ (9+2), (9+3), (9+4), (9+5), (9+6), (9+7), และ (9+8)

เขียนสมการได้เป็น ถ้า $Y_i = a+X_i$

แล้วจากสมการ (2-18) จะได้

$$\bar{Y} = a + \mu \quad \text{เมื่อ} \quad b = 100 \ \text{และ} \ \mu = 5$$

$$\bar{Y} = 9+5$$

$$= 14$$

ค่าเฉลี่ยของข้อมูลมีค่าเท่ากับ 14

ง. ค่าของ 225, 335, 445, 555, 665, 775, และ 885 อาจจะเขียนได้อีกรูปแบบหนึ่งคือ (200+25), (300+35), (400+45), (500+55), (600+65), (700+75), และ (800+85).

เขียนได้เป็นสมการ

$$W_i = Y_i + X_i$$

แล้วจากสมการ (2-19) จะได้

$$\bar{W} = \bar{Y} + \mu$$

แทนค่า $\bar{Y}$ และ μ ที่ได้จากข้อ ข. และข้อ ก. แล้วจะได้

$$\bar{W} = 500+55$$

$$\bar{W} = 550$$

ค่าเฉลี่ยของข้อมูลมีค่าเท่ากับ 550

2.3 มัธยฐานของข้อมูล

มัธยฐาน (median) เป็นมาตรวัดแนวโน้มสู่ส่วนกลางของตัวแปรมีสัญลักษณ์เป็น Me , Med หรือ Mdn. กำหนดว่า มัธยฐาน คือ ค่าสังเกตหรือตัวแปรที่แบ่งจำนวนตัวแปรออกเป็นส่วนที่ค่ามากกว่ามัธยฐานและส่วนที่ค่าน้อยกว่ามัธยฐานได้ในจำนวนเท่า ๆ กัน การหามัธยฐานของตัวแปรจากประชากรหรือสิ่งตัวอย่าง สามารถดำเนินการได้ ดังต่อไปนี้

2.3.1 กรณีข้อมูลไม่ได้แจกแจงความถี่ หาค่ามัธยฐานได้โดยจัดเรียงลำดับค่าสังเกตจากน้อยไปหามาก (หรือจากมากไปหาน้อย) ลำดับที่ได้คือตำแหน่งของค่าสังเกตที่ 1, 2, 3,.. ค่าสังเกตที่อยู่ตำแหน่งกึ่งกลางของตำแหน่งทั้งหมดคือ มัธยฐาน

ตัวอย่างที่ 2.8 จงหามัธยฐานของ 8, 12, 4, 6, และ 10

วิธีทำ

เรียงลำดับ X_i	4	6	8	10	12
ตำแหน่ง X_i	1	2	3	4	5

พบว่า ตำแหน่งกึ่งกลางของตำแหน่งทั้งหมด คือ ตำแหน่งที่ 3 ตรงกับ 8
นั่นคือ มัธยฐานของข้อมูลนี้ คือ 8

ตัวอย่างที่ 2.9 จงหามัธยฐานของ 3, 4, 11, 8, 10, และ 5

วิธีทำ

เรียงลำดับ X_i	3	4	5	8	10	11
ตำแหน่ง X_i	1	2	3	4	5	6

พบว่า ตำแหน่งกึ่งกลางของตำแหน่งทั้งหมด คือ ตำแหน่งที่ 3 กับ ตำแหน่งที่ 4 หรือ ตำแหน่งที่ 3.5 ตรงกับเลข 5 และ 8 หรือ $\frac{5+6}{2} = 5.5$

นั่นคือ มัธยฐานของข้อมูลนี้ คือ 5.5

จึงสรุปได้ว่า เมื่อต้องการหามัธยฐานของประชากรหรือสิ่งตัวอย่างและเรียงลำดับค่าตัวแปร จำนวนทั้งหมด n ตัวแปรแล้ว

$$\text{มัธยฐาน คือ ค่าตัวแปรในตำแหน่งที่ } \frac{n+1}{2} \qquad \text{...(2-20)}$$

2.3.2 กรณีข้อมูลแจกแจงความถี่แล้ว เมื่อแจกแจงความถี่ของข้อมูลแล้วสามารถใช้ความถี่สะสมหรือความถี่สะสมสัมพัทธ์ประกอบในการหาค่ามัธยฐาน เพราะสดมภ์ความถี่สะสมหรือความถี่สะสมสัมพัทธ์ จะบอกจำนวนตัวแปรที่มีค่าต่ำกว่าค่าตัวแปรในชั้นที่กล่าวถึง สำหรับชั้นที่กล่าวถึงในที่นี้คือชั้นที่มีความถี่สะสมหรือความถี่สะสมสัมพัทธ์เป็นร้อยละ 50 หรือ $\frac{n}{2}$ และชั้นนี้ก็คือชั้นที่มีมัธยฐานบรรจุอยู่

ตารางที่ 2.10 ช่วงชั้นข้อมูลน้ำหนักของพนักงานขายที่ค่ามัธยฐานบรรจุอยู่

ช่วงขั้นข้อมูล (กิโลกรัม)	*f*	*cf*	
29.5-39.5	4	4	
39.5-49.5	6	10	
49.5-59.5	8	18	
59.5-69.5	12	30	⇐ ชั้นมัธยฐาน
69.5-79.5	9	39	
79.5-89.5	7	46	
89.5-99.5	4	50	
	50		

จากตารางที่ 2.10 เป็นตารางการแจกแจงความถี่ข้อมูล น้ำหนักของพนักงานขายจำนวน 50 คน สดมภ์ช่วงชั้นที่แสดงเป็นขีดจำกัดล่างและขีดจำกัดบนของแต่ละชั้น (วัดในหน่วยกิโลกรัม) ชั้นที่มัธยฐานบรรจุอยู่หรือชั้นที่มีความถี่สะสมเป็นร้อยละ 50 คือชั้นที่ 4 (59.5-69.5) ซึ่งมีความถี่เท่ากับ 12 กรณีเช่นนี้เราถือว่าน้ำหนักพนักงานขายคนที่ 25 เป็นมัธยฐาน นั่นคือน้ำหนักระหว่าง 59.6 ถึง 69.5 กิโลกรัม มีพนักงานขายจำนวน 12 คน อยู่ในตำแหน่ง

ที่ 19, 20, 21,.....30. แต่ตำแหน่งในชั้นเป็นคนที่ 1, 2,......,12. ดังตารางที่ 2.10 และจากตารางที่ 2.11 จะพบว่า น้ำหนักของพนักงานขายคนที่ 25 ตรงกับในชั้นตำแหน่งที่ 7 เป็นมัธยฐาน และต้องมีค่าระหว่าง 59.5 ถึง 69.5 กิโลกรัม อาศัยวิธีการประมาณค่าในช่วงของตัวเลขดังกล่าวและมีวิธีการคิดดังนี้

ตำแหน่งในชั้นต่างกัน 12 ตำแหน่ง มีค่าต่างกัน $= 69.5 - 59.5$ กิโลกรัม

$= 10$ กิโลกรัม

ตำแหน่งในชั้นต่างกัน 7 ตำแหน่ง มีค่าต่างกัน $= \frac{10}{12}(25 - 18)$ กิโลกรัม

$= 5.83$ กิโลกรัม

ดังนั้น มัธยฐาน เท่ากับ $59.5 + 5.83 = 65.33$ กิโลกรัม

ตารางที่ 2.11 ตำแหน่งและค่ามัธยฐานของน้ำหนักของพนักงานขาย

คนที่	19	20	21	22	23	24	25	26	27	28	29	30
ตำแหน่งในชั้น	1	2	3	4	5	6	7	8	9	10	11	12
น้ำหนัก	59.5	<-	----	----	----	----	*Me*	----	----	----	--->	69.5

จากหลักเกณฑ์ที่กล่าวมา เราสามารถนำไปใช้หามัธยฐานกรณีข้อมูลแจกแจงความถี่แล้วได้

ถ้ากำหนด Me คือ มัธยฐาน

L คือ ขีดจำกัดล่างของชั้นมัธยฐานบรรจุอยู่

n คือ จำนวนข้อมูลทั้งหมด

f_m คือ ความถี่ของชั้นมัธยฐานบรรจุอยู่

I คือ ขนาดของอันตรภาคชั้นที่มัธยฐานบรรจุอยู่

cf_m คือ ความถี่สะสมของชั้นก่อนถึงชั้นมัธยฐานบรรจุอยู่

จะได้สมการ (2-21)

$$Me \quad = \quad L + \frac{I}{f_m}\left(\frac{n}{2} - cf_m\right) \qquad \text{...(2-21)}$$

ตัวอย่างที่ 2.10 จากตารางที่ 2.12 จงหามัธยฐานของแต้มความสามารถของพนักงานขาย 88 คน

ตารางที่ 2.12 แต้มความสามารถของพนักงานขาย

แต้ม	ความถี่
2-6	18
7-11	21
12-16	26
17-21	15
22-26	8
	88

วิธีทำ จากสมการ (2-21) $Me \quad = \quad L + \frac{I}{f_m}\left(\frac{n}{2} - cf_m\right)$

$$L = 11.5,\ I = 5,\ \frac{n}{2} = 44,\quad f = 39\ \ ,f_m = 26$$

$$Me = 11.5 + \frac{5}{26}(44 - 39)$$

$$= 11.5 + 0.961$$

$$= 12.46$$

ตารางที่ 2.13 ความถี่สะสมและชั้นมัธยฐาน

ช่วงชั้นแต้ม	f	cf	
2-6	18	18	
7-11	21	39	
12-16	26	65	$\Leftarrow$ ชั้นมัธยฐาน
17-21	15	80	
22-26	8	88	
	88		

นั่นคือ มัธยฐานของแต้มความสามารถของพนักงานขายครั้งนี้คือ 12.46 แต้ม และอยู่ในช่วงชั้นแต้ม 12 – 16 ดังแสดงในตารางที่ 2.13

2.4 ฐานนิยมของข้อมูล

ในการคิด ค่าเฉลี่ย ค่ามัธยฐาน อาจไม่เหมาะสมสำหรับเป็นแนวโน้มสู่ส่วนกลางสำหรับสถานการณ์บางสถานการณ์ เช่น รายจ่ายของนักศึกษาหลาย ๆ คนต่อวัน จำนวนชิ้นงานที่พนักงาน ผลิตได้ต่อวัน จำนวนบุคคลที่ชอบอาหารรสเผ็ด หรือ อื่น ๆ ทำนองเดียวกัน การวัดแนวโน้มสู่ส่วนกลางควรจะเป็นค่าที่พบบ่อยหรือมีความถี่สูงที่สุด จึงกำหนดฐานนิยม (mode) ขึ้นเป็นเครื่องมือวัดอีกอันหนึ่ง โดยกำหนดว่า ฐานนิยม คือ ค่าสังเกตที่พบบ่อยที่สุดในจำนวนค่าสังเกตทั้งหมดที่มีอยู่ และมีสัญลักษณ์เป็น *Mo* หรือ *Mod* ซึ่งการหาฐานนิยมของประชากรหรือสิ่งตัวอย่างอาจดำเนินการได้ดังต่อไปนี้

2.4.1 กรณีข้อมูลไม่ได้แจกแจงความถี่ กรณีเช่นนี้เราไม่จำเป็นต้องเรียงลำดับค่าสังเกตเสียก่อน เพียงแต่ตรวจดูอย่างละเอียดถี่ถ้วน ว่าค่าสังเกตหรือข้อมูลตัวใดพบบ่อยที่สุด ค่าของข้อมูลตัวนั้นจะเป็นฐานนิยม

ตัวอย่างที่ 2.11 จงหาฐานนิยม เมื่อพนักงาน 10 คนสามารถขายสินค้าต่อวัน บันทึกจำนวนสินค้าที่ขายได้ดังนี้

15, 18, 18, 17, 15, 16, 16, 15, 14, 12.

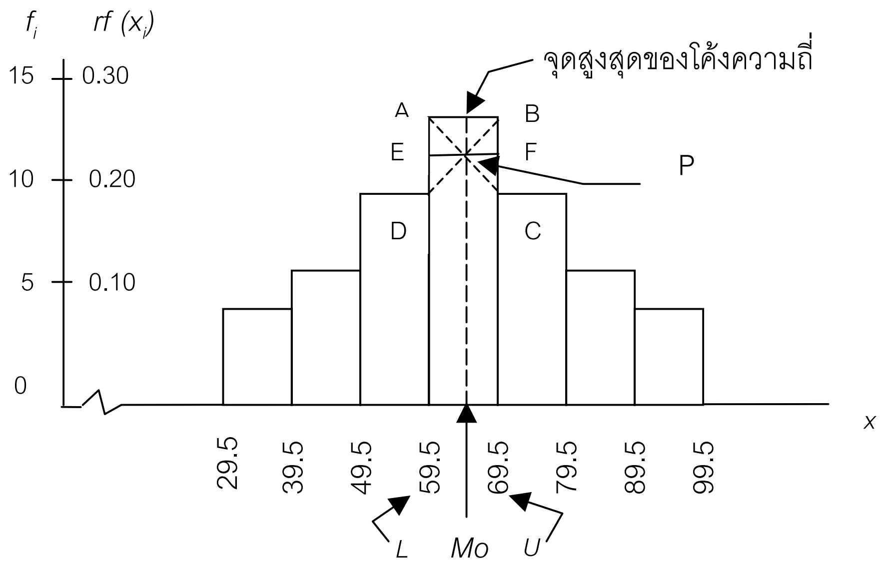
วิธีทำ

ตรวจดูอย่างละเอียดแล้ว พบว่า จำนวนสินค้าค้าที่ขายได้ 15 ชิ้น พบบ่อยที่สุดคือ 3 ครั้งดังนั้น ฐานนิยม คือ 15

ในสถานการณ์บางอย่างอาจไม่พบว่ามีค่าสังเกตตัวใดซ้ำเลย เช่น ข้อมูลประกอบด้วย 3, 6, 4, 8, 7. จะถือว่าข้อมูลนี้ไม่มีฐานนิยม แต่ในบางครั้งอาจพบฐานนิยมมากกว่า 1 ค่า เช่น ข้อมูลประกอบด้วย 3, 6, 7, 10, 8, 3, 9, 6, 9. มีฐานนิยมถึง 3 ค่า คือ 3, 6 และ 9 เพราะพบบ่อยที่สุด 2 ครั้งเหมือน ๆ กัน กรณีเช่นนี้ควรหลีกเหลี่ยงการใช้ฐานนิยมเป็นแนวโน้มสู่ส่วนกลางและอาจถือได้ว่า ข้อมูลดังกล่าวไม่มีฐานนิยมหรือเรียกอย่างอื่น เช่น ทวิฐานนิยม ในกรณีที่มีฐานนิยม 2 ค่า เป็นต้น

2.4.2 กรณีข้อมูลมีการแจกแจงความถี่ การแจกแจงความถี่จะทำให้สามรถหาค่าฐานนิยมได้ง่ายกว่าข้อมูลที่ไม่ได้แจกแจงความถี่ เพราะหากช่วงชั้นขนาด 1 หน่วยใดมีความถี่สูงสุดค่าช่วงชั้นนั้น คือ ฐานนิยมส่วนอันตรภาคชั้นขนาดมากกว่า 1 หน่วย ค่าจุดกลางชั้นนั้น คือ ฐานนิยม โดยทั่วไปจะเรียกชั้นที่มีความถี่สูงสุดนี้ว่า ชั้นฐานนิยม (modal class)

เมื่อนำข้อมูลในตารางที่ 2.13 ไปสร้างฮิสโทแกรมและโค้งความถี่จุดศูนย์กลางของโค้งความถี่ จุดสูงสุดของโค้งความถี่ จะชี้บอกค่าข้อมูลที่เป็นฐานนิยม เพื่อความสะดวกในการสังเกตจุดสูงสุดของโค้งความถี่ ได้จากภาพที่ 2.1



ภาพที่ 2.1 ฐานนิยมจากฮิสโทแกรมและจุดสูงสุดของโค้งความถี่

พิจารณาภาพที่ 2.1 กำหนดให้

Mo คือ ฐานนิยม

L คือ ขีดจำกัดล่างของชั้นฐานนิยม

U คือ ขีดจำกัดบนของชั้นฐานนิยม

I คือ $U - L$ หรือ ขนาดของอันตรภาคชั้น

d_1 คือ ผลต่างความถี่ของชั้นฐานนิยมกับชั้นก่อนหน้า

d_2 คือ ผลต่างความถี่ของชั้นฐานนิยมกับชั้นถัดลงไป

จากรูปทรงเรขาคณิตที่เกิดขึ้นพบว่า ΔADP และ ΔBPC เป็นสามเหลี่ยมคล้าย ซึ่งจากสมบัติของสามเหลี่ยมคล้ายนี้ จะได้

$$\frac{EP}{AD} = \frac{PF}{BC}$$

แต่ $$EP = Mo - L$$

$$PF = U - Mo$$

ตารางที่ 2.14 ชั้นฐานนิยมที่ได้จากกราฟของข้อมูลในภาพที่ 2.1

ช่วงชั้นข้อมูล	f	
29.5-39.5	4	
39.5-49.5	6	
49.5-59.5	8	
59.5-69.5	12	$\Leftarrow$ ชั้นฐานนิยม
69.5-79.5	9	
79.5-89.5	7	
89.5-99.5	4	
	50	

$$AD = d_1$$

$$BC = d_2$$

แทนค่า

$$\frac{M_0 - L}{d_1} = \frac{U - M_0}{d_2}$$

$$d_2\,(M_o - L) = d_1\,(U - M_o)$$

$$d_2 M_o - d_2 L = d_1 U - d_1 M_o$$

$$M_o\,(d_1 + d_2) = d_1 U + d_2 L$$

$$= d_1 L + d_1 I + d_2 L$$

$$M_o = \frac{L(d_1 + d_2)}{d_1 + d_2} + \frac{Id_1}{d_1 + d_2}$$

ดังนั้น

$$M_o = L + \frac{Id_1}{d_1 + d_2} \qquad ...(2\text{-}22)$$

ตัวอย่างที่ 2.12 จงหาฐานนิยมของข้อมูลของแต้มความสามารถของพนักงานขายจากตารางที่ 2.13

วิธีทำ จากตารางที่ 2.13 ชั้นฐานนิยม คือชั้นที่ 4 (59.5-69.5)

และจากสมการ (2-22) $$Mo = L + \frac{ld_1}{d_1 + d_2}$$

$$L = 59.5 , l = 10, d_1 = 12 - 8 = 4 , d_2 = 12 - 9 = 3$$

แทนค่า $$Mo = 59.5 + \frac{(10)(4)}{(4+3)}$$

$$= 59.5 + 5.71$$

$$Mo = 65.21$$

ดังนั้น ฐานนิยม คือ 65.21

2.5 แนวโน้มสู่ส่วนกลางอื่น ๆ ของข้อมูล

สำหรับเครื่องมือวัดแนวโน้มสู่ส่วนกลางอื่น ๆ ในที่นี้จะขอกล่าวเพียงข้อกำหนดเบื้องต้นและตัวอย่างการหาซึ่งเป็นพื้นฐานการศึกษาขั้นสูงต่อไป ดังต่อไปนี้

2.5.1 มัชฌิมเรขาคณิต (geometric mean : GM.) กำหนด $X_1, X_2 , X_3,\ldots\ldots X_n$ เป็นค่าสังเกตจำนวน n ค่า จะได้ตามสมการ (2-23 และ (2-24) (Keller & Warrack, 2000, p. 100)

$$GM. = \sqrt[n]{X_1 \bullet X_2 \bullet X_3 \bullet X_3 \ldots. \bullet X_n} \qquad \ldots(2\text{-}23)$$

$$GM. = (X_1 \bullet X_2 \bullet X_3 \bullet X_3 \ldots \bullet X_n)^{1/n} \qquad \ldots(2\text{-}24)$$

ตัวอย่างที่ 2.2 จงหาค่ามัชฌิมเรขาคณิตของ 3 , 1 , และ 9

วิธีทำ จากสมการ (2-23)

$$\text{GM.} = \sqrt[n]{X_1 \bullet X_2 \bullet X_3 \bullet X_3 .... \bullet X_n}$$

แทนค่า $\text{GM.} = \sqrt[3]{(3)(1)(9)} = 3$

ดังนั้น มัชฌิมเรขาคณิตของข้อมูลนี้เท่ากับ 3

การหาค่ามัชฌิมเรขาคณิตจะเห็นว่ามีความซับซ้อน หากมีค่าสังเกตจำนวนเท่าใด จะต้องนำมาคูณกันทุกตัวแล้วถอดรากที่เท่ากับจำนวนค่าสังเกตทั้งหมด เช่น ค่าสังเกตมีจำนวน 10 ค่า จะต้องนำมาคูณกันทั้งหมดแล้วถอดรากที่ 10 จึงจะเป็นมัชฌิมเรขาคณิต เพื่อให้ง่ายยิ่งขึ้นจึงนิยมใช้ลอการิทึมสามัญในการคำนวณ (ซึ่งแสดงในตารางภาคผนวกที่ 2) ดังนั้นจากสมการ (2-2) เราสามารถเขียนให้อยู่ในรูปลอการิทึมสามัญ ได้เป็น

$$\log \text{GM.} = \log \left(X_1 \bullet X_2 \bullet X_3 \bullet X_3 ... \bullet X_n\right)^{1/n}$$

$$= \frac{1}{n} \log \left(X_1 \bullet X_2 \bullet X_3 \bullet X_3 ... \bullet X_n\right)$$

$$\log \text{GM.} = \frac{1}{n} \left(\log X_1 + \log X_2 + \log X_3 + \ldots + \log X_n\right) \qquad \ldots(2\text{-}25)$$

เช่น จากตัวอย่างที่ 2.2 อาจเขียนได้เป็น

$$\log \text{GM.} = \frac{1}{3} \left(\log 3 + \log 1 + \log 9\right)$$

จากตารางภาคผนวกที่ 2 อ่านได้

$$\log 3 = 0.4771$$

$$\log 1 = 0$$

$$\log 9 = 0.9542$$

$$log\ GM. = (0.4771 + 0 + 0.9542) = 0.4771$$

พบว่าที่ $log\ N = 0.4771$; $N = 3$ ดังนั้น GM. = 3 เป็นต้น

2.5.2 มัชฌิมฮาร์มอนิก (harmonic mean : *HM.*) กำหนด $X_1, X_2, X_3, \ldots\ldots X_n$ เป็นค่าสังเกตจำนวน n ค่า จะได้สมการ (2-26) (Solomon, 1996, p. 16)

$$HM. = \frac{n}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots\ldots\ldots\ldots+\frac{1}{X_n}} \qquad \ldots(2\text{-}26)$$

ตัวอย่างที่ 2.14 จงหามัชฌิมฮาร์มอนิกของ 2, 7, และ 3

วิธีทำ จากสมการ (2-26)

$$HM. = \frac{n}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots\ldots\ldots\ldots+\frac{1}{X_n}}$$

เมื่อ $n = 3, x_1 = 2, x_2 = 7, x_3 = 3$

แทนค่า

$$HM. = \frac{3}{\frac{1}{2}+\frac{1}{7}+\frac{1}{3}}$$

$$= \frac{3}{\frac{21+6+14}{42}}$$

$$= (3)\frac{42}{41} = 3.07$$

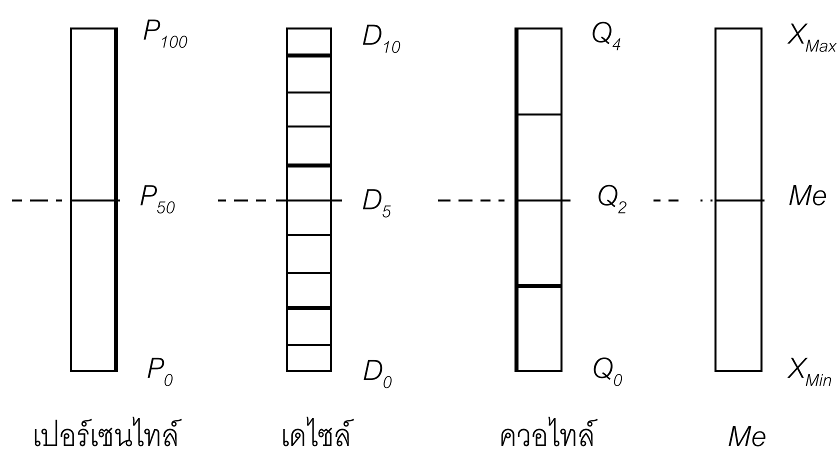
ดังนั้นมัชฌิมฮาร์มอนิกของข้อมูลนี้เท่ากับ 3.07

2.5.3 เปอร์เซนไทล์ เดไซล์และควอร์ไทล์ (percentile: P, decile : D and quartile : Q) สำหรับเปอร์เซนไทล์ เดไซล์และควอร์ไทล์ เป็นเครื่องมือที่กำหนดขึ้นเพื่อเรียกตำแหน่ง (position) ของข้อมูล เมื่อจัดเรียงลำดับข้อมูลทั้งหมด จากค่าน้อยไปหาค่ามากแล้ว โดยกำหนดตำแหน่งเริ่มต้นเป็น เปอร์เซนไทล์ 0 (P_0) เดไซล์ 0 (D_0) และควอร์ไทล์ 0 (Q_0) อธิบายได้ดังนี้ (Solomon, 1996, p. 10-11)

1) เปอร์เซนไทล์ จะบอกตำแหน่งข้อมูล เมื่อข้อกำหนดทั้งหมดถูกแบ่งเป็น 100 ส่วน ๆ ละเท่า ๆ กัน (เซนติมีค่าเท่ากับ $\frac{1}{100}$) ตำแหน่งเปอร์เซนไทล์จะเริ่มต้นตั้งแต่ที่ 0 , 1, 2 , ... , 100. เขียนแทนว่า P_0 , P_1, P_2, P_3, ... และ P_{100}.

2) เดไซล์ จะบอกตำแหน่งข้อมูล เมื่อข้อมูลทั้งหมดถูกแบ่งเป็น 10 ส่วน ๆ ละเท่า ๆ กัน (เดซิมีค่าเท่ากับ $\frac{1}{10}$) ตำแหน่งเดไซล์จึงเริ่มต้นตั้งแต่ตำแหน่งที่ 0, 1, 2, 3, ... , 10 เขียนแทนว่า D_0, D_1, D_2, D_3........ และ D_{10}

3) ควอร์ไทล์ จะบอกตำแหน่งข้อมูล เมื่อข้อมูลทั้งหมดถูกแบ่งออกเป็น 4 ส่วน ๆ ละเท่า ๆ กัน (ควอเตอร์มีค่าเท่ากับ $\frac{1}{4}$) ตำแหน่งควอร์ไทล์เริ่มต้นตั้งแต่ตำแหน่งที่ 0 , 1 , 2 , 3 และ 4 เขียนแทนว่า Q_0, Q_1, Q_2, Q_3, และ Q_4



ภาพที่ 2.2 การเปรียบเทียบ P_{50} , D_5 , Q_2 และ Me

การวัดแนวโน้มสู่ส่วนกลางจะได้แก่ ข้อมูลตัวที่มีตำแหน่งกึ่งกลางของข้อมูลทั้งหมด คือ P_{50}, D_5,, และ Q_2 ซึ่งเข้าลักษณะเดียวกับมัธยฐานนั่นเอง สำหรับข้อมูลชุดเดียวกันจะพบว่าข้อมูลในตำแหน่ง $P_{50} = D_5 = Q_2 = Me$ พิจารณาได้จากภาพที่ 2.2

ตัวอย่างที่ 2.15 จงหา P_{50} , D_5 และ Q_2 จาก 2, 9, 10, 6, 3, 7, 5 และ 4

วิธีทำ

เรียงลำดับ : 2 3 4 5 6 7 9 10

ตำแหน่งที่ : 1 2 3 4 5 6 7 8

หา P_{50}

ตำแหน่งเปอร์เซนไทล์ที่ 100 ตรงกับข้อมูลจริงตำแหน่งที่ 8

ตำแหน่งเปอร์เซนไทล์ที่ 50 ตรงกับข้อมูลจริงตำแหน่งที่ $\dfrac{(8)(50)}{100} = 4$

พบว่าข้อมูลจริงตำแหน่งที่ 4 คือ 5

ดังนั้น P_{50} ของข้อมูลนี้เท่ากับ 5

หา D_5

ตำแหน่งเดไซล์ที่ 10 ตรงกับข้อมูลจริงตำแหน่งที่ 8

ตำแหน่งเดไซล์ที่ 5 ตรงกับข้อมูลจริงตำแหน่งที่ $\dfrac{(8)(5)}{10} = 4$

พบว่าข้อมูลจริงตำแหน่งที่ 4 คือ 5

ดังนั้น D_5 ของข้อมูลนี้เท่ากับ 5

หา Q_2

ตำแหน่งควอร์ไทล์ที่ 4 ตรงกับข้อมูลจริงตำแหน่งที่ 8

ตำแหน่งควอร์ไทล์ที่ 2 ตรงกับข้อมูลจริงตำแหน่งที่ $\frac{(8)(2)}{4} = 4$

พบว่าข้อมูลจริงตำแหน่งที่ 4 คือ 5

ดังนั้น Q_2 ของข้อมูลนี้เท่ากับ 5

สถานการณ์จริงบางครั้งตำแหน่ง P_{50} , D_5 และ Q_2 ที่คำนวณได้ตรงกับข้อมูลจริงในตำแหน่งที่ไม่เป็นจำนวนเต็ม เช่น P_{50} ตรงกับข้อมูลจริงตำแหน่งที่ 2.5 หรือ D_5 ตรงกับข้อมูลจริงตำแหน่งที่ 3.4 เป็นต้น เราจะอาศัยวิธีประมาณค่าในช่วงโดยเทียบอัตราส่วนเช่นเดียวกับมัธยฐาน ดังตัวอย่างที่ 2.16 ต่อไปนี้

ตัวอย่างที่ 2.16 จงหา P_{50} , D_5 และ Q_2 ของ 14, 15, 12, 9, 13, 6, 4, 5, 3.

วิธีทำ P_{50} , D_5 และ Q_2 ของข้อมูลนี้เท่ากันจึงอาจใช้วิธีการหาเปอร์เซนไทล์ 50 เดไซล์ 5 หรือ ควอร์ไทล์ 2 ก็ได้

หา P_{50}

เรียงลำดับ : 3 4 5 6 9 12 13 14 15
ตำแหน่งที่ : 1 2 3 4 5 6 7 8 9

ตำแหน่งเปอร์เซนไทล์ที่ 100 ตรงกับข้อมูลจริงตำแหน่งที่ 9

ตำแหน่งเปอร์เซนไทล์ที่ 50 ตรงกับข้อมูลจริงตำแหน่งที่ $\frac{(9)(50)}{100} = 4.5$

พบว่าข้อมูลจริงตำแหน่งที่ 4.5 หรือ 4 ถึง คือ 6 กับ 9 อาศัยวิธีการประมาณค่าในช่วงได้เป็น

ข้อมูลจริงตำแหน่งต่างกัน 1 ตำแหน่งค่าข้อมูลต่างกัน = 9 - 6 = 3

ข้อมูลจริงตำแหน่งต่างกัน (4.5-4) = 0.5 ตำแหน่งค่าข้อมูลต่างกัน $= \frac{(3)(0.5)}{1} = 1.5$

ดังนั้น P_{50} ของข้อมูลนี้เท่ากับ 6 + 1.5 = 7.5

2.6 บทสรุป

บทนี้เป็นการกล่าวถึงมาตรการโน้มสู่ส่วนกลางดังนี้

2.6.1 การหาผลรวมของข้อมูล เป็นการนำข้อมูลทั้งหมดมารวมกัน

2.6.2 มัชฌิมเลขคณิตของข้อมูล เป็นการหาค่าเฉลี่ย โดยหาได้จากการหาผลรวมของข้อมูล ทั้งหมดมาหารด้วยจำนวนข้อมูลทั้งหมด ทั้งในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่และมีการแจกแจงความถี่ (จัดกระทำข้อมูลเป็นตาราง) ซึ่งสามารถบอกความถี่ ความถี่สะสม ความถี่สัมพัทธ์ ความถี่สะสมสัมพัทธ์ ได้

2.6.3 มัธยฐานของข้อมูล เป็นค่าสังเกตหรือตัวแปรที่แบ่งจำนวนค่าสังเกตออกเป็นส่วนที่มากกว่าหรือส่วนที่น้อยกว่ามัธยฐาน

2.6.4 ฐานนิยมของข้อมูล เป็นค่าสังเกตที่พบบ่อยที่สุดในจำนวนค่าสังเกตทั้งหมดที่มีอยู่

นอกจากการหาค่าดังกล่าวมาแล้วเราอาจหาค่ามาตรการโน้มสู่ส่วนกลางได้อีก เช่น การหามัชฌิมเรขาคณิต มัชฌิมฮาร์มอนิก ตำแหน่งเปอร์เซนไทล์ เดไซล์ และควอร์ไทล์ของข้อมูลได้

2.7 คำถามทบทวน

1. จงกระจายค่าต่อไปนี้โดยไม่ให้มีสัญลักษณ์ผลรวม

ก. $\sum_{i=1}^{n} X_i$

ข. $\sum_{i=1}^{n} X_i^2$

ค. $\sum_{i=1}^{n}(X_i + Y_i)$

ง. $\sum_{i=1}^{n}(X_i + a)$

จ. $\sum_{i=1}^{n}X_iY_i$

ฉ. $\sum_{i=1}^{n}X_if_i$

2. จงแสดงค่าเหล่านี้ให้อยู่ในสัญลักษณ์ของผลรวม

ก. $z_1 + z_2 + z_3 + \ldots + z_{20}$

ข. $x_1 f_1 + x_2 f_2 + \ldots + x_6 f_6$

ค. $x_2^3 y_2 + x_3^3 y_3 + \ldots + x_7^3 y_7$

ง. $(x_1 - y_1) + (x_2 - y_2) + \ldots + (x_m - y_m)$

3. จงพิสูจน์ว่า $\sum_{i=1}^{n}(X_i - k) = \sum_{i=1}^{n}X_i - nk$

4. จงแสดงให้เห็นว่า $(\sum_{i=1}^{n}X_i)^2 = \sum_{i=1}^{n}(X_i)^2$ หรือไม่เพราะเหตุใด

5. จงพิสูจน์ $\sum_{i=1}^{n}(X_i - k)^2 = \sum_{i=1}^{n}(X_i)^2 - 2k\sum_{i=1}^{n}X_i + nk^2$

6. ถ้ามัชฌิมเลขคณิตของข้อมูลจำนวน 36 ตัวเท่ากับ 55 จงหาผลรวมของข้อมูลชุดนี้

7. ประชากรหนึ่งประกอบด้วย 6.2, 5.1, 8.7, 6.2, 4.1, 3.3, 5.4, 6.2, 6.7 และ 9.6

ก. จงหามัชฌิมเลขคณิต มัธยฐานและฐานนิยมของประชากร

ข. คูณทุกตัวด้วย 10 แล้วหามัชฌิมเลขคณิตของประชากร

ค. บวกทุกตัวด้วย 20 แล้วหามัชฌิมเลขคณิตของประชากร

8. จากตัวเลขต่อไปนี้

50, 55, 60, 65, 70, 75, 80, 85

จงหา

ก. มัชฌิมเรขาคณิตของตัวเลข

ข. มัชฌิมฮาร์มอนิกของตัวเลข

9. แต้มความขยันของพนักงานเป็นดังต่อไปนี้ จงหา P_{50} , D_5 และ Q_2 ของ แต้มความขยันพนักงาน

20, 17, 16, 15, 19, 19, 18, 12,

13, 13, 17, 16, 14, 14, 15 , 15

10. จากข้อมูลอายุของพนักงานประจำร้านอาหารแห่งหนึ่งจำนวน 7 คน เป็นดังนี้

19, 19, 65, 20, 21, 18, และ 18

จงหา

ก. ค่าเฉลี่ยของอายุพนักงาน

ข. ค่ามัฐยฐานของอายุพนักงาน

ค. ค่าฐานนิยมของอายุพนักงาน

ง. ค่าแนวโน้มสู่ส่วนกลางของอายุที่เหมาะสม ถ้าต้องการตัดพนักงานที่มีอายุมากออก

11. จากตารางที่ 2.14 จงหาค่าต่อไปนี้

ก. ค่าเฉลี่ยจำนวนรถที่มีของนักธุรกิจ

ข. ค่ามัฐยฐานของจำนวนรถที่มีนักธุรกิจ

ค. ค่าฐานนิยมของจำนวนรถที่มีนักธุรกิจ

ตารางที่ 2.15 จำนวนรถของนักธุรกิจ 20 คน

จำนวนรถที่มี (คัน)	นักธุรกิจ (คน)
0	3
1	10
2	4
3	2
4	1

บทที่ 3
การวัดการกระจายข้อมูล

ข้อมูล ค่าสังเกต หรือตัวแปรทั้งหลาย อันเป็นคุณลักษณะของประชากรหรือตัวอย่างจะมีแนวโน้มสู่ส่วนกลางที่เลือกเป็นคุณลักษณะของประชากรหรือตัวอย่างนั้น ค่าแนวโน้มสู่ส่วนกลางอาจเป็น มัชฌิมเลขคณิต มัธยฐาน ฐานนิยมหรืออื่น ๆ เมื่อต้องการเปรียบเทียบประชากรหรือสิ่งตัวอย่าง 2 กลุ่ม หรือมากกว่า 2 กลุ่ม อาจพบว่ามีค่าแนวโน้มสู่ส่วนกลางเท่ากัน แต่เมื่อพิจารณาอย่างละเอียดแล้วคุณลักษณะในกลุ่ม แตกต่างกัน เช่น ในกรณีที่พนักงาน 2 กลุ่ม ๆ ละ 10 คนทำการบรรจุสินค้าลงหีบห่อแล้วพบว่าความสามารถในการบรรจุสินค้าลงหีบห่อต่อชั่วโมงของพนักงานทั้งสองกลุ่มเป็นดัง ตารางที่ 3.1

ตารางที่ 3.1 การเปรียบเทียบความสามารถในการบรรจุสินค้าลงหีบห่อได้ของพนักงาน 2 กลุ่ม

กลุ่มที่1	
จำนวนหีบห่อ	จำนวนคน
7	1
8	2
9	4
10	2
11	1

กลุ่มที่ 2	
จำนวนหีบห่อ	จำนวนคน
3	1
4	1
5	1
6	1
9	2
12	1
13	1
14	1
15	1

พิจารณาจากตารางที่ 3.1 เมื่อเราหามัชฌิมเลขคณิตกรณีข้อมูลแจกแจงความถี่แล้ว จะพบว่า ความสามารถในการบรรจุสินค้าลงหีบห่อของพนักงาน 2 กลุ่มนี้มีมัชฌิมเลขคณิตเป็น 9 หีบห่อต่อชั่วโมงเท่ากัน แต่เมื่อพิจารณาคุณลักษณะอย่างละเอียดแล้วจะเห็นว่ามีความแตกต่างในลักษณะที่เรียกว่าการแปรผัน หรือการกระจาย (scatter) ของความสามารถของพนักงานในแต่ละกลุ่ม ด้วยเหตุผลดังกล่าว จำเป็นอย่างยิ่งที่จะต้องกำหนดเครื่องมือสำหรับวัดการกระจายของค่าสังเกต เพื่อบอกให้ทราบว่าค่าสังเกตเหล่านั้นมีการแปรผันการกระจายอย่างไร หรือเพื่อให้ทราบว่าค่าสังเกตทั้งหลายมีการเบี่ยงเบนจากค่าแนวโน้มสู่ส่วนกลางที่ถูกเลือกเอาไว้อย่างไรบ้างในบทนี้จึงเป็นการกล่าวถึงการกำหนดมาตรวัดการกระจาย (measures of dispersion)

3.1 **พิสัยของข้อมูล**

พิสัย (range) เป็นการหาความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของค่าสังเกตทั้งหมด สัญลักษณ์แทนพิสัย คือ R หรือ RNG กำหนดให้ x_1 , x_2 , ..., x_n เป็นค่าสังเกตจำนวน n ค่า ถ้ากำหนดให้

x_U	เป็นขีดจำกัดสูงสุด
x_L	เป็นขีดจำกัดต่ำสุด
R	เป็นพิสัย

จะได้สมการ (3-1) (Loether & McTavish,1980, p.153)

$$R = x_U - x_L \qquad \ldots(3\text{-}1)$$

เปรียบเทียบสมการ (3-1) กับสมการ (1-2) ในบทที่ 1 พบว่าผลต่างระหว่างตัวแปรค่าสูงสุดกับตัวแปรต่ำสุดเราจัดเป็นพิสัยอย่างคร่าว ๆ โดยใช้ x_n แทนค่าตัวแปรค่าสูงสุดและ x_1 แทนค่าตัวแปรค่าต่ำสุด จะได้ สมการ (3-2)

$$R = x_n - x_1 \qquad \ldots(3\text{-}2)$$

อย่างไรก็ดี ค่าพิสัยเหมาะสมสำหรับการวัดการกระจายของค่าสังเกตจำนวนน้อย ๆ เพราะง่ายต่อการคำนวณและสามารถบอกการกระจายค่าสังเกตแต่ละกลุ่มว่าแตกต่างกันหรือไม่ ตลอดจนเป็นเหตุผลนำไปสู่มาตรการกระจายอื่น ๆ ต่อไป

ตัวอย่างที่ 3.1 จงหาพิสัยของจำนวนสินค้าที่พนักงาน 32 คนจำหน่ายได้ต่อวันต่อไปนี้

3 , 3 ,4 , 4 , 4 , 4 ,8 ,8 ,8 ,8 ,8 ,8 ,

8 ,8 ,8 ,8 ,8 ,8 , 9 ,9 , 9 , 9 , 9 , 9 ,

10 , 10 , 10 , 11 , 11 , 11 , 13 ,13 .

วิธีทำ

ก. กรณีข้อมูลไม่ได้แจกแจงความถี่

จากสมการ (3-1) $R = x_U - x_L$

$$x_U = 13.5$$

$$x_L = 2.5$$

$$R = 13.5 - 2.5 = 11$$

ข. กรณีข้อมูลแจกแจงความถี่แล้ว ในที่นี้กำหนดการแจกแจงความถี่ที่มีขนาดของอันตรภาคชั้นเท่ากับ 2 ดังนั้นจากสมการ (3-1) จะได้

$$R = x_U - x_L$$

ตารางที่ 3.2 การแจกแจงความถี่ที่บอกจุดกลางชั้น

ชั้น	จุดกลางชั้น (x_j)	ความถี่ (f_j)
3 - 4	3.5	6
5 - 6	5.5	0
7 - 8	7.5	12
9 - 10	9.5	9
11 - 12	11.5	3
13 - 14	13.5	2
		32

x_U = 14.5

x_L = 2.5

R = 14.5 - 2.5 = 12

จากตัวอย่างที่ 3.1 ใน ก. และ ข. จะเห็นว่า มีค่า R ที่แตกต่างกันเพราะการจัดชั้นในการแจกแจงความถี่ข้อมูล

3.2 พิสัยระหว่างควอร์ไทล์ของข้อมูล

พิสัยระหว่างควอร์ไทล์ (interquatile range : Q) เป็นพิสัยของค่าสังเกตช่วงที่บรรจุค่าสังเกตเอาไว้จำนวนร้อยละ 50 ของค่าสังเกตทั้งหมด หรือถ้ากำหนด Q_3 เป็นตำแหน่ง ควอร์ไทล์ที่ 3 (เปอร์เซนไทล์ 75 หรือ เดไซล์ 7.5) Q_1 เป็นตำแหน่งควอร์ไทล์ที่ 1 (เปอร์เซนไทล์ที่ 25 หรือ เดไซล์ 2.5) และ Q เป็นพิสัยระหว่างควอร์ไทล์จะได้สมการ (3-3) (Keller & Warrack , 2000, p.119)

$$Q = Q_3 - Q_1$$

…(3-3)

ตัวอย่างที่ 3.2 จากข้อมูลในตัวอย่างที่ 4.1 จงหาพิสัยระหว่างควอร์ไทล์

วิธีทำ ก. กรณีข้อมูลไม่ได้แจกแจงความถี่ แบ่งค่าสังเกตออกเป็น 4 ส่วนเพื่อกำหนดตำแหน่งควอร์ไทล์ ดังนี้

3	3	4	4	4	4	8	8	8	8	8	8	8	8	8	8
Q_0							Q_1								Q_2
8	8	9	9	9	9	9	9	10	10	10	11	11	11	13	13
							Q_3								Q_4

จากสมการ (3-3) $Q = Q_3 - Q_1 = 9 - 8 = 1$

พิสัยระหว่างควอร์ไทล์มีค่าเท่ากับ 1

ข. กรณีที่ข้อมูลมีการแจกแจงความถี่ วิธีการแจกแจงความถี่ข้อมูลดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 การแจกแจงความถี่ของข้อมูลที่บอกค่าจุดกลางชั้นและความถี่สะสม

ชั้น	จุดกลางชั้น (x_j)	ความถี่ (f_j)	ความถี่สะสม (cf_j)	
3 – 4	3.5	6	6	
5 – 6	5.5	0	6	
7 – 8	7.5	12	18	⇐ มีจำนวนข้อมูล $\frac{1}{4}$ ของข้อมูลทั้งหมด
9 – 10	9.5	9	27	⇐ มีจำนวนข้อมูล $\frac{3}{4}$ ของข้อมูลทั้งหมด
11 - 12	11.5	3	30	
13 - 14	13.5	2	32	
		32		

จากความถี่สะสมในตารางที่ 3.3 จะพบว่า ตำแหน่ง Q_1 ซึ่งมีจำนวนข้อมูล $\frac{1}{4}$ ของข้อมูลทั้งหมดมีค่าอยู่ในชั้นที่ 3 ระหว่างค่า 7 - 8 และตำแหน่ง Q_3 ซึ่งมีจำนวนข้อมูล $\frac{3}{4}$ ของข้อมูล ทั้งหมดมีค่าอยู่ในชั้นที่ 3 ระหว่าง 9 ถึง 10 อาศัยวิธีประมาณค่าในช่วงทำนองเดียวกับฐานนิยมหา Q_1 และ Q_3 ได้ดังนี้

$$Q_1 = 6.5 + \frac{2}{12}\left[\frac{1}{4}(32) - 6\right] = 6.83$$

$$Q_3 = 8.5 + \frac{2}{9}\left[\frac{3}{4}(32) - 18\right] = 9.83$$

ดังนั้น $Q = 9.83 - 6.83 = 3$

พิสัยระหว่างควอร์ไทล์มีค่าเท่ากับ 3

จากตัวอย่างที่ 3.2 ในกรณีที่ข้อมูลมีการแจกแจงความถี่ ถ้ากำหนดให้

L_3 , L_1 เป็นขีดจำกัดล่างของชั้น Q_3 และ Q_1 ตามลำดับ

I เป็นขนาดอันตรภาคชั้น

cf_3 , cf_1 เป็นความถี่สะสมชั้นก่อนถึงชั้น Q_3 และ Q_1 ตามลำดับ

f_3 , f_1 เป็นความถี่ของชั้น Q_3 และ Q_1 ตามลำดับจะได้สมการการหาค่า Q_1 และ Q_3 ดังสมการ (3-4) และสมการ (3-5)

$$Q_1 = L_1 + \frac{I}{f_1}\left(\frac{n}{4} - cf_1\right) \qquad \ldots(3\text{-}4)$$

$$Q_3 = L_3 + \frac{I}{f_3}\left(\frac{3n}{4} - cf_3\right) \qquad \ldots(3\text{-}5)$$

3.3 ส่วนเบี่ยงเบนเฉลี่ยของข้อมูล

ส่วนเบี่ยงเบนเฉลี่ย (average deviation) เป็นมาตรที่กำหนดขึ้นเพื่อบอกให้ทราบว่าค่าสังเกตตัวใด ๆ ที่มีอยู่มีค่าสมบูรณ์ของการแปรผันจากมัชฌิมเลขคณิตหรือศูนย์กลางการแจกแจงโดยเฉลี่ยเท่าใดจึงทำให้บางครั้งเรียกคำนี้ว่าค่าเบี่ยงเบนเฉลี่ยสัมบูรณ์ (average absolute deviation) หรือค่าเบี่ยงเบนเฉลี่ย สัญลักษณ์แทนค่านี้ คือ *AD.*

กำหนดให้ x_1 , x_2 , ... , x_n เป็นค่าสังเกตจำนวน n ค่า หามัชฌิมเลขคณิตได้เท่ากับ μ แล้วจะได้สมการ (3-6)

$$AD. = \frac{\sum_{i=1}^{n} |x_i - \mu|}{n} \qquad \ldots(3\text{-}6)$$

$|x_i - \mu|$ คือ ค่าสัมบูรณ์ของการเบี่ยงเบนจากมัชฌิมเลขคณิตของค่าสังเกตแต่ละตัว (absolute value of deviation)

ตัวอย่างที่ 3.3 จงหาค่าเบี่ยงเบนเฉลี่ยของข้อมูลต่อไปนี้

2 , 4 , 6 , 8 , 2 , 14 , 20

วิธีทำ

ก. กรณีข้อมูลไม่ได้แจกแจงความถี่

จากสมการ (3-6)

$$AD. = \frac{\sum_{i=1}^{n}|x_i - \mu|}{n}$$

$$\mu = \frac{2+4+6+8+2+14+20}{7}$$

$$\mu = \frac{56}{7} = 8$$

$$\sum_{i=1}^{n}|x_i - \mu| = |2-8| + |4-8| + |6-8| + |8-8| + |2-8| + |14-8| + |20-8|$$

$$= 6 + 4 + 2 + 0 + 6 + 6 + 12 = 36$$

$$AD. = \frac{36}{7}$$

$$= 5.14$$

ค่าเบี่ยงเบนเฉลี่ยมีค่าเท่ากับ 5.14

ข. กรณีข้อมูลแจกแจงความถี่แล้วดังตารางที่ 3.4 จะได้เป็น

$$\mu = \frac{56}{7}$$

และ $AD. = \frac{36}{7} = 5.14$

ค่าเบี่ยงเบนเฉลี่ยมีค่าเท่ากับ 5.14

ตารางที่ 3.4 การแจกแจงความถี่ของข้อมูลเพื่อหาค่าเบี่ยงเบนเฉลี่ย

x_j	f_j	$x_j f_j$	$\lvert x_j - \mu \rvert$	$f_j \lvert x_j - \mu \rvert$
2	2	4	6	12
4	1	4	4	4
6	1	6	2	2
8	1	8	0	0
14	1	14	6	12
20	1	20	12	12
	7	56		36

พิจารณาจากตัวอย่างที่ 3.3 กรณีข้อมูลแจกแจงความถี่แล้ว ถ้ากำหนดให้

x_j เป็นจุดกลางชั้นที่ 1 ถึงชั้น k

f_j เป็นความถี่ชั้นที่ 1 ถึงชั้น k

และ

$$n = \sum_{j=1}^{k} f_j$$

จะได้สมการ (3-7)

$$AD. = \frac{\sum_{j=1}^{k} f_j \lvert x_j - \mu \rvert}{n} \qquad \text{...(3-7)}$$

และจากสมการ (3-7) นี้จะพบว่า $\frac{f_j}{n}$ คือความถี่สัมพัทธ์ของชั้นที่ 1 ถึงชั้นที่ k เขียนแทนด้วย $rf(x_j)$ ได้สมการ (3-8)

$$AD. = \sum_{j=1}^{k} f_j \lvert x_j - \mu \rvert \, rf(x_j) \qquad \text{...(3-8)}$$

ตัวอย่างที่ 3.4 จงหาค่าเบี่ยงเบนเฉลี่ยจากข้อมูลตัวอย่างที่ 3.3 กรณีข้อมูลแจกแจง

ความถี่และความถี่สัมพัทธ์แล้ว

วิธีทำ จากข้อมูลในตารางที่ 3.3 นำมาจัดกระทำเพื่อหาค่าเบี่ยงเบนเฉลี่ยจะได้ดังตารางที่ 3.5 ดังนี้

ตารางที่ 3.5 การหาค่าเบี่ยงเบนเฉลี่ยกรณีข้อมูลแจกแจงความถี่และมีค่าความถี่สัมพัทธ์แล้ว

x_j	f_j	$rf(x_j)$	$\lvert x_j - \mu \rvert$	$\lvert x_j - \mu \rvert rf(x_j)$
2	2	$\frac{2}{7}$	6	$\frac{12}{7}$
4	1	$\frac{1}{7}$	4	$\frac{4}{7}$
6	1	$\frac{1}{7}$	2	$\frac{2}{7}$
8	1	$\frac{1}{7}$	0	0
14	1	0	$\frac{1}{7}$	$\frac{6}{7}$
20	1	$\frac{1}{7}$	12	$\frac{12}{7}$
	7			$\frac{36}{7}$

จากสมการ (3-8) $AD. = \sum_{j=1}^{k} \lvert x_j - \mu \rvert rf(x_j)$

$$= \frac{36}{7} \qquad = 5.14$$

ค่าเบี่ยงเบนเฉลี่ยมีค่าเท่ากับ 5.14

3.4 ความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของข้อมูล

ค่าความแปรปรวน (variance) และค่าเบี่ยงเบนมาตรฐาน (standard deviation) เป็นค่าที่บอกการเบี่ยงเบนจากมัชฌิมเลขคณิตต่อหนึ่งหน่วยค่าสังเกต ทำนองเดียวกับค่า

เบี่ยงเบนเฉลี่ย สัญลักษณ์ทั่วไป ที่ใช้แทนค่าความแปรปรวน คือ $Var(X)$ หรือ $V(X)$ กรณีกล่าวถึงค่าความแปรปรวนของประชากรจะใช้ σ^2 (อักษรกรีกชนิดตัวอักษรเล็ก ชื่อ

sigma: ซิกมา) ส่วนกรณีกล่าวถึงค่าความแปรปรวนของสิ่งตัวอย่างจะใช้ S^2 หรือ s^2 เพื่อความเข้าใจเบื้องต้นจะใช้มัชฌิมเลขคณิต (μ) เป็นหลักในการหาการเบี่ยงเบน $|x_i - \mu|$ กำหนด x_1, x_2 ,............,x_n เป็นค่าสังเกตจำนวน n ค่ามีมัชฌิมเลขคณิตเท่ากับ μ แล้วสมการ (3-9)

$$V(X) = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} \qquad \text{…(3-9)}$$

$\sum_{i=1}^{n}(x_i - \mu)^2$ เรียกว่า ผลบวกกำลังสองของการเบี่ยงเบน (sum of the squared deviation : *SS*) ทำให้สามารถเรียกค่าความแปรปรวนว่า ค่าเฉลี่ยของผลบวกกำลังสองของการเบี่ยงเบน (mean of the sum of squared deviation : *MS*) สำหรับค่าความแปรปรวนของสิ่งตัวอย่าง จะแตกต่างออกไปตามสมการ (3-10)

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1} \qquad \text{…(3-10)}$$

เมื่อ $\overline{X}$ เป็นค่าเฉลี่ยสิ่งตัวอย่าง ซึ่งมีเหตุผลการใช้ค่าเช่นนี้ กล่าวโดยละเอียดในเรื่องการประมาณค่าเกี่ยวกับประชากร

มีข้อสังเกตว่า หน่วยของความแปรปรวนเป็นหน่วยยกกำลังสองซึ่งไม่ตรงกับหน่วยความจริงของค่าสังเกต จึงถอดรากที่สองของค่าความแปรปรวนและเรียกค่านี้ว่า ค่าเบี่ยงเบนมาตรฐาน สัญลักษณ์ที่ใช้แทนค่าเบี่ยงเบนมาตรฐานเช่น *SD.* หรือ *sd.* กรณีประชากรใช้ σ และกรณีสิ่งตัวอย่างใช้ S หรือ s เป็นต้น

ถ้ากำหนด σ เป็นค่าเบี่ยงเบนมาตรฐาน จากค่า $V(X)$ จะได้สมการ (3-11)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$ …(3-11)

ตัวอย่างที่ 3.5 จงหาค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของข้อมูลต่อไปนี้

15, 9, 10 , 6 , 18 , 13 ,21 , 8

วิธีทำ (1) จากสมการ (3-9)

$$V(X) = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} \quad ; \quad n = 8$$

$$\mu = (15+9+10+6+18+13+21+8)/8 = 12.5$$

$$\sum_{i=1}^{n}(x_i - \mu)^2 = (15-12.5)^2 + (9-12.5)^2 + (10 -12.5)^2 + (6 -12.5)^2 +$$

$$(8 -12.5)+(13- 12.5)^2+(21-12.5)^2+(8-12.5)^2$$

$$= (-2.5)^2 + (-3.5)^2+(-2.5)^2+(-6.5)^2+(5.5)^2+(0.5)^2 + (8.5)^2$$

$$+(-4.5)^2$$

$$= 190$$

$$V(X) = \frac{190}{8} = 23.75$$

$$\sigma = \sqrt{23.75}$$

$$= 4.87$$

ค่าเบี่ยงเบนมาตรฐานของข้อมูลมีค่าเท่ากับ 4.87

วิธีทำ (2) จากตารางที่ 3.6 เมื่อหาค่า $|x_i - \mu|$ และ $(x_i - \mu)^2$ แล้วเราสามารถหาค่าที่ต้องการได้ดังนี้

$$\mu = \frac{100}{8}$$

$$\mu = 12.5$$

$$\sum_{i=1}^{n}(x_i - \mu)^2 = 190$$

$$V(X) = \frac{190}{8}$$

$$= 23.75$$

ตารางที่ 3.6 การหาค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของ x_i

x_i	$\lvert x_i - \mu \rvert$	$(x_i - \mu)^2$
6	-6.5	42.25
8	-4.5	20.25
9	-3.5	12.25
10	-2.5	9.25
13	0.5	0.25
15	-3.5	6.25
18	5.5	30.25
21	8.5	72.25
100	0.0	190.00

และ $\sigma = \sqrt{23.75} = 4.87$

ค่าเบี่ยงเบนมาตรฐานของข้อมูลมีค่าเท่ากับ 4.87

นอกจากที่กล่าวมาเราอาจอาศัยคุณสมบัติของผลรวม หาได้ดังสมการต่อไปนี้

$$\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} = \frac{\sum(x_i - \mu)^2}{n}$$

$$= \frac{\sum (x^2 - 2x\mu + \mu^2)}{n}$$

$$= \frac{\sum x^2 - 2\sum x\mu + \sum \mu^2}{n}$$

$$\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} = \frac{\sum x^2}{n} - \frac{2\mu\sum x}{n} + \frac{\sum \mu^2}{n}$$

$$\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} = \frac{\sum x^2}{n} - 2\mu^2 + \mu^2$$

ดังนั้น $E(X) = \frac{\sum x^2}{n} - \mu^2$ …(3-12)

หรือ $V(X) = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$ …(3-13)

และ $\sigma = \sqrt{\frac{\sum x^2}{n} - \mu^2}$ …(3-14)

หรือ $\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$ …(3-15)

จากข้อมูลตามตัวอย่างที่ 3.5 เราอาจใช้สมการ (3-12) หาค่าความแปรปรวนและใช้สมการ (3-14) หาค่าเบี่ยงเบนมาตรฐานได้ดังต่อไปนี้

$$V(X) = \frac{\sum_{i=1}^{n} x^2}{n} - \mu^2$$

$$\mu^2 = \left(\frac{100}{8}\right)^2 = \frac{10,000}{64}$$

$$\sum_{i=1}^{n} x^2 = 6^2 + 8^2 + 9^2 + 10^2 + 13^2 + 15^2 + 18^2 + 21^2$$

$$= 36+64+81+100+169+225+324+441$$

$$= \quad 1,440$$

และ $n = 8$

$$V(X) = \frac{1,440}{8} - \frac{10,000}{64}$$

$$V(X) = \frac{11,520}{64} - \frac{10,000}{64}$$

$$V(X) = 23.75$$

และ $\sigma = \sqrt{23.75} = 4.87$

กรณีข้อมูลแจกแจงความถี่แล้ว ถ้าค่าสังเกตถูกจัดเป็น k ชั้นโดย x_1 , x_2 , ... และ x_k เป็นจุดกลางชั้นที่ 1 ถึงชั้นที่ k มีความถี่เป็น f_1 , f_1 , ... และ f_k ทำนองเดียวกันกับการหาค่าเบี่ยงเบนเฉลี่ย กรณีข้อมูลเราแจกแจงความถี่แล้ว เราจะได้ สมการ (3-16)

$$V(X) = \frac{\sum_{j=1}^{k} f_j \left(x_j - \mu\right)^2}{n} \qquad \text{...(3-16)}$$

หรือ $$V(X) = \sum_{j=1}^{k} f_j \left(x_j - \mu\right)^2 rf(x_j) \qquad \text{...(3-17)}$$

เพราะว่า $\frac{f_j}{n}$ คือ $rf(x_j)$ เป็นความถี่สัมพัทธ์ของชั้นที่ 1 ถึงชั้นที่ k และจากกรณีข้อมูลแจกแจงความถี่แล้วเทอม $\frac{\sum_{j=1}^{k} x^2}{n}$ ก็คือ $\frac{\sum_{j=1}^{k} f_j \; x_j^2}{n}$ เท่ากับ $\sum_{j=1}^{k} x^2 \; rf(x_j)$ จึงได้สมการ (3-18)

$$V(X) = \sum_{j=1}^{k} x^2 \; rf(x_j) - \mu^2 \qquad \text{...(3-18)}$$

เรามีข้อสังเกตสำหรับค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของประชากรหลายกลุ่มที่มีมัชฌิมเลขคณิตเท่ากัน จากตารางที่ 3.7 จะพบว่า ประชากรกลุ่ม ง มีการกระจายมากที่สุด

ดังนั้นคุณสมบัติของค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐาน เมื่อค่าสังเกตเปลี่ยนแปลงไปทุกค่า จะอยู่ในลักษณะถูกบวก ลบ คูณ หรือหาร ด้วยค่าคงตัว

กำหนด a เป็นค่าคงตัวและ $x_1, x_2, \ldots\ldots, x_n$ เป็นค่าสังเกตจำนวน n ค่าคำนวณมัชฌิมเลขคณิตได้เท่ากับ μ ค่าความแปรปรวนเท่ากับ $V(X)$ และค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_x จะมีคุณสมบัติของค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐาน เมื่อค่าสังเกตเปลี่ยนแปลงไป โดยมีการพิสูจน์สมการการหาค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐานได้ดังต่อไปนี้

ตารางที่ 3.7 ประชากร 4 กลุ่มที่มีมัชฌิมเลขคณิตเท่ากัน

ประชากร	ก	ข	ค	ง
ค่าสังเกต (x_j)	6 6 6	4 5 9	3 4 11	2 3 13
Σx_j	18	18	18	18
μ	6	6	6	6
$V(X)$	0	$\frac{14}{3}$	$\frac{38}{3}$	$\frac{74}{3}$
σ	0	2.16	3.55	4.96

3.4.1 ถ้า (x_1+a) , (x_2+a) , (x_3+a) , ... , (x_n+a) มีค่าความแปรปรวนเท่ากับ $V(X+a)$ และค่าเบี่ยงเบนมาตรฐานเท่ากับ $\sigma_{(x+a)}$ แล้วจะได้สมการ (3-19) และสมการ (3-20)

$$V(X+a) = V(X) \quad \ldots(3\text{-}19)$$

และ

$$\sigma_{(X+a)} = \sigma_x \quad \ldots(3\text{-}20)$$

พิสูจน์

จากสมการ(3-9) $V(X) = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$

ดังนั้น $V(X+a) = \frac{\sum_{i=1}^{n}((x_i + a) - (\mu + a))^2}{n}$

$V(X+a) = \frac{\sum_{i=1}^{n}(x_i - \mu + a - a)^2}{n}$

Xxxxxxxxxxx(หน้านี้ไม่ใช้แต่ต้องพิมพ์)

$$V(X+a) = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

$$= V(X)$$

3.4.2 ถ้า $(x_1 - a)$, $(x_2 - a)$, $(x_3 - a)$, ... , $(x_n - a)$ มีค่าความแปรปรวนเท่ากับ $V(X-a)$ และค่าเบี่ยงเบนมาตรฐานเท่ากับ $\sigma_{(x-a)}$ แล้วจะได้สมการ (3-21) และสมการ (3-22)

$$V(X-a) = V(X) \qquad \text{...(3-21)}$$

และ

$$\sigma_{(X-a)} = \sigma_x \qquad \text{...(3-22)}$$

พิสูจน์ ในทำนองเดียวกับหัวข้อ 3.4.1

3.4.3 ถ้า ax_1 , ax_2 , ax_2 ,........................., ax_n มีค่าความแปรปรวนเท่ากับ $V(aX)$ และค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_{ax} แล้วจะได้สมการ (3-23) และสมการ (3-24)

$$V(aX) = a^2V(X) \qquad \text{...(3-23)}$$

และ

$$\sigma_{(ax)} = a\,\sigma_x \qquad \text{...(3-24)}$$

พิสูจน์ จากสมการ(3-9) $V(X) = \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$

$$V(aX) = \frac{\sum_{i=1}^{n}(ax_i-a\mu)^2}{n}$$

$$= \frac{\sum_{i=1}^{n}a^2(x_i-\mu)^2}{n}$$

$$V(aX) \quad = \quad \frac{a^2\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

$$= \quad a^2\, V(X)$$

3.4.4 ถ้าค่า $V(a) \quad = \quad 0$ …(3-25)

และ $\sigma_{(a)} \quad = \quad 0$ …(3-26)

พิสูจน์ จากสมการ(3-9) $V(X) \quad = \quad \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$

ในที่นี้ X คือ a เป็นค่าคงตัว และ μ คือ $\frac{an}{n}$ เท่ากับ a

ดังนั้น $V(a) \quad = \quad \frac{\sum_{i=1}^{n}(a-a)^2}{n}$

$$= \quad 0$$

3.5 ค่ามาตรฐานของข้อมูล

ค่ามาตรฐานหรือคะแนนมาตรฐาน (standard score or Z- score) กำหนดขึ้นใช้สำหรับการเปรียบเทียบค่าสังเกตจากประชากรหรือค่าสังเกตจากสิ่งตัวอย่างสองกลุ่มหรือมากกว่าที่มีการแจกแจงลักษณะเดียวกันโดยเฉพาะการแจกแจงปรกติซึ่งจะกล่าวถึงโดยละเอียดในบทที่ 4 โดยกำหนดว่าค่ามาตรฐานคือตัวเลขบอกถึงการเบี่ยงเบนของแต่ละค่าสังเกตจากศูนย์กลางการแจกแจง (มัชฌิมเลขคณิต) ต่อค่าเบี่ยงเบนมาตรฐานหนึ่งหน่วย นั่นคือ ถ้า X_1 , X_2 , X_3 ,..............., X_n เป็นค่าสังเกตจำนวน n ค่า มีมัชฌิมเลขคณิตเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_x และ Z_i เป็นค่ามาตรฐานของ X_i แล้ว ได้สมการ (3-27) (Solomon, 1996, p. 172)

$$z_i \quad = \quad \frac{X_i-\mu}{\sigma_x} \qquad \text{…(3-27)}$$

$(X_i - \mu)$ คือผลต่างหรือการเบี่ยงเบนของแต่ละค่าสังเกตกับมัชฌิมเลขคณิตและเราสามารถพิสูจน์ได้สมการ (3-28) ดังนี้

$$\sum_{i=1}^{n}(X_i - \mu) = 0 \qquad \text{...(3-28)}$$

พิสูจน์ จากสมการ (3-6) จะได้ $\sum_{i=1}^{n} X_i$ คือ $n\mu$ ดังนั้น

$$\sum_{i=1}^{n}(X_i - \mu) = (X_1 - \mu) + (X_2 - \mu) + (X_3 - \mu) + \ldots\ldots\ldots + (X_n - \mu)$$

$$= (X_1 + X_2 + X_3 + \ldots\ldots\ldots + X_n) - n\mu$$

$$= n\mu - n\mu$$

$$= 0$$

จากสมการ (3-28) ทำให้เกิดคุณสมบัติของค่ามาตรฐานที่น่าสนใจและใช้เป็นพื้นฐานสำหรับการแจกแจงลักษณะอื่น ๆ

กำหนดให้ Z_1, Z_2, Z_3 ,............., Z_n เป็นค่ามาตรฐานของค่าสังเกต X_1, X_2, X_3 ,............, X_n ตามลำดับ และ

$\bar{Z}$ เป็นตัวกลางมัชฌิมของค่ามาตรฐาน

σ_z^2 เป็นค่าความแปรปรวนของค่ามาตรฐาน

σ_Z เป็นค่าเบี่ยงเบนมาตรฐานของค่ามาตรฐาน

จะได้คุณสมบัติของ Z_i ดังสมการ (3-29) ถึงสมการ (3-35)

$$\sum_{i=1}^{n} Z_i = 0 \qquad \text{...(3-29)}$$

พิสูจน์ $$\sum_{i=1}^{n} Z_i = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma_x}$$

$$\sum_{i=1}^{n} Z_i = \frac{1}{\sigma_x}\sum_{i=1}^{n}(X_i - \mu)$$

$$= 0$$

$$\overline{Z} = 0 \qquad \ldots(3\text{-}30)$$

พิสูจน์ $$\overline{Z} = \frac{\frac{1}{\sigma_x}\sum_{i=1}^{n}(X_i - \mu)}{n}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \mu)}{n\sigma_x}$$

$$= 0$$

$$\sigma_z^2 = 1 \qquad \ldots(3\text{-}31)$$

และ $$\sigma_z = 1 \qquad \ldots(3\text{-}32)$$

พิสูจน์ จากสมการ (3-9) เป็นการหาค่าความแปรปรวนของ X เมื่อเราเปลี่ยน X เป็น Z จึงสามารถหาค่าความแปรปรวนของ Z (σ_z^2) ได้เป็น

$$\sigma_z^2 = \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})^2}{n}$$

แต่ $$\overline{Z} = 0$$

ดังนั้น $$\sigma_z^2 = \frac{\sum_{i=1}^{n}(Z_i)^2}{n}$$

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

$$\sigma_z^2 = \frac{1}{\sigma_x^2}\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$$

$$= \frac{1}{\sigma_x^2}[\sigma_x^2]$$

$$= 1$$

$$\sigma_z = \sqrt{1}$$

$$= 1$$

$$\sum_{i=1}^{n} Z_i^2 = n \qquad \text{…(3-35)}$$

พิสูจน์ $$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma_x}\right)^2$$

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

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

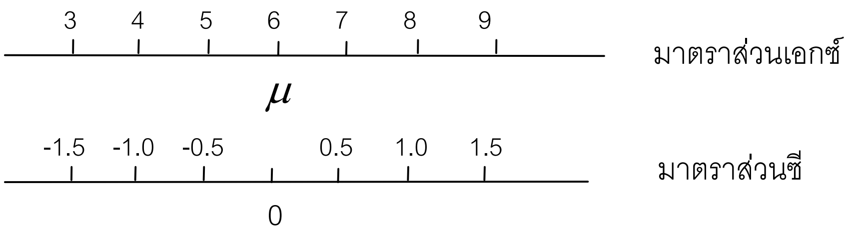
$$\sum_{i=1}^{n} Z_i^2 = n$$

การเปลี่ยนค่าสังเกตใด ๆ ไปเป็นค่ามาตรฐานตามสมการ (3-27) ลักษณะการแจกแจงของค่ามาตรฐานที่ได้จะยังคงรักษารูปทรงการแจกแจงเดิมเอาไว้ ทั้งนี้เพราะค่ามาตรฐานที่เปลี่ยนมาจากค่าสังเกตใด ๆ จะมีค่าขึ้นลงตามกัน สังเกตได้จากตารางที่ 3.8 ซึ่งจะพบว่า มาตราส่วนเอกซ์ (X - scale) มีค่ามัชฌิมเลขคณิตเท่ากับ 6 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 2 ส่วนมาตราส่วนซี (Z – scale) จะมีค่ามัชฌิมเลขคณิตเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1

จากภาพที่ 3.1 จะเห็นชัดเจนว่า ค่ามาตรฐานจะเป็นบวก (+) ถ้าค่าสังเกตตัวนั้นมากกว่า μ ตรงกันข้ามค่ามาตรฐานจะเป็นลบ (-) ถ้าค่าสังเกตตัวนั้นน้อยกว่า μ, ค่ามาตรฐานเหล่านี้จึงใช้ในการเปรียบเทียบค่าสังเกตจากประชากร หรือตัวอย่างสุ่มสองกลุ่มหรือมากกว่าสองกลุ่มได้ดังตัวอย่างที่ 3.6

ตารางที่ 3.8 การเปรียบเทียบมาตราส่วนเอกซ์และมาตราส่วนซี

มาตราส่วนเอกซ์	มาตราส่วนซี
3	-1.5
4	-1.0
5	-0.5
6	0.0
7	0.5
8	1.0
9	1.5
42	0.0



ภาพที่ 3.1 เส้นจำนวนของมาตราส่วนเอกซ์และมาตราส่วนซีของข้อมูลในตารางที่ 3.8

ตัวอย่างที่ 3.6 รายได้ต่อวันของพนักงานโรงงานเอคนหนึ่งเป็น 84 บาท ขณะที่ค่าเฉลี่ยรายได้พนักงานโรงงานเอเป็น 76 บาท ค่าเบี่ยงเบนมาตรฐานเป็น 10 บาทและพนักงานโรงงานบีคนหนึ่งเป็น 90 บาทขณะที่ค่าเฉลี่ยรายได้พนักงานโรงงานบีเป็น 82 บาท ค่าเบี่ยงเบนมาตรฐานเป็น 16 บาท อยากทราบว่าพนักงานคนใดรายได้ดีกว่า

วิธีทำ เปลี่ยนรายได้ของพนักงานโรงงานเอและรายได้ของพนักงานโรงงานบีเป็นค่ามาตรฐาน

จากสมการ(3-27) $Z_i = \dfrac{X_i - \mu}{\sigma}$

สำหรับรายได้ของพนักงานโรงงานเอ $X_1 = 84$,μ_1 =76 และ $\sigma_1 = 10$

$$Z_1 = \frac{84-76}{10} = 0.8$$

สำหรับรายได้ของพนักงานโรงงานบี $X_2 = 90$,$\mu_2 = 82$ และ $\sigma_2 = 16$

$$Z_2 = \frac{X_2 - \mu}{\sigma}$$

$$Z_2 = \frac{90-82}{16}$$

$$Z_2 = 0.5$$

สรุปรายได้ของพนักงานโรงงานเอดีกว่ารายได้ของพนักงานโรงงานบี

3.6 สัมประสิทธิ์การแปรผันของข้อมูล

ค่าสัมประสิทธิ์ของการแปรผัน (coefficient of variation : *CV.* หรือ *V.*) ใช้สำหรับการเปรียบเทียบการกระจายของประชากรหรือสิ่งตัวอย่างสองกลุ่มหรือมากกว่าสองกลุ่ม เช่น เปรียบเทียบค่าสังเกตประเภทเดียวกันแต่มาจากสองแหล่งหรือหลาย ๆ แหล่ง เปรียบเทียบค่าสังเกตจากการทดลองสองครั้งหรือมากกว่าสองครั้ง หรือเปรียบเทียบค่าสังเกตที่วัดด้วยหน่วยที่แตกต่างกัน เป็นต้น อย่างไรก็ดีในงานทดลองค่าสัมประสิทธิ์ของการแปรผันจะใช้ประเมินประสิทธิภาพของงานทดลองนั้น ๆ ด้วย

กำหนด X_1 , X_2 ,.............., X_n เป็นค่าสังเกตจำนวน n ค่า มีมัชฌิมเลขคณิตเท่ากับ μ และค่าเบี่ยงเบนมาตรฐานเท่ากับ σ แล้วจะได้สมการ (3-36) (Watson et al., 1990, p. 92)

$$CV. = \frac{\sigma}{\mu} \qquad \text{…(3-36)}$$

ในสถานการณ์ทั่วไปค่า *CV.* ที่คำนวณได้จะมีค่าไม่เกิน 1 หรือเป็นทศนิยม จึงคำนวณค่า *CV.* เท่ากับ ค่าเบี่ยงเบนมาตรฐานคิดเป็นร้อยละของมัชฌิมเลขคณิตได้เป็นสมการ (3-37)

$$CV. = \frac{\sigma}{\mu}\,(100)\ \% \qquad \text{…(3-37)}$$

ตัวอย่างที่ 3.7 น้ำหนักเฉลี่ยนักศึกษาชายกลุ่มหนึ่งเท่ากับ 60 กิโลกรัมค่าเบี่ยงเบนมาตรฐานเท่ากับ 6 กิโลกรัม และน้ำหนักเฉลี่ยนักศึกษาชายกลุ่มสองเฉลี่ยเท่ากับ 250 กิโลกรัม ค่าเบี่ยงเบนมาตรฐาน 7 กิโลกรัม จงหาว่าน้ำหนักเฉลี่ยของนักศึกษาชายกลุ่มใดมีการกระจายมากกว่ากัน

วิธีทำ $$CV. = \frac{\sigma}{\mu}\,(100)\ \%$$

คำนวณค่า *CV.* จากสองแหล่งแล้วเปรียบเทียบกัน ถ้าค่า *CV.* สูงกว่า
น้ำหนักนักศึกษาชายกลุ่มหนึ่ง

$$\mu_1 = 60$$

$$\sigma_1 = 6$$

$$CV._1 = \frac{6}{60}\,(100)\ \%$$

$$= 10\ \%$$

น้ำหนักเฉลี่ยนักศึกษาชายกลุ่มสอง

$$\mu_2 = 250$$

$$\sigma_2 = 7$$

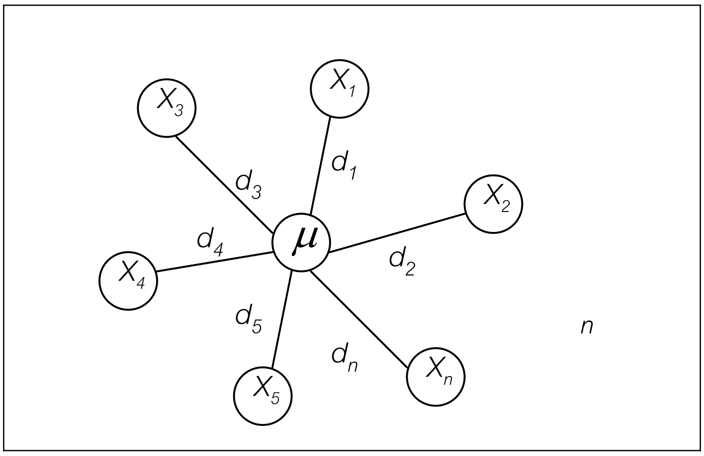
$$CV._2 = \frac{7}{250}(100)\ \%$$

$$= 2.8\ \%$$

สรุปว่าการกระจายของน้ำหนักนักศึกษาชายกลุ่มหนึ่งมากกว่าน้ำหนักนักศึกษาชายกลุ่มสอง

3.7 โมเมนต์รอบศูนย์กลางของข้อมูล

ในการบอกค่าสังเกตแต่ละค่าว่ามีความสมดุลรอบ ๆ มัชฌิมเลขคณิตอย่างไรนั้น เราบอกได้ด้วย ค่าโมเมนต์รอบศูนย์กลาง (central moment) ดังแสดงในภาพที่ 3.2 เมื่อเรากำหนด X_1 , X_2 ,............,X_n เป็นค่าสังเกตจำนวน n ค่ามีมัชฌิมเลขคณิตเท่ากับ μ และ d_i เป็นการเบี่ยงเบนหรือความแตกต่างระหว่างค่าสังเกตใด ๆ กับมัชฌิมเลขคณิต (Casella & Berger, 1990, p. 58)



ภาพที่ 3.2 โมเมนต์ที่เกิดจาก X_i รอบ ๆ μ

$$d_1 = (X_1 - \mu) \qquad \text{...(3-38)}$$

และเราทราบจากสมการ (3-28) ว่าผลรวมการเบี่ยงเบนเป็นศูนย์ จึงได้เป็น

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (X_i - \mu) = 0 \qquad \text{...(3-39)}$$

เมื่อเราใช้โมเมนต์รอบศูนย์กลางวัดการกระจายของค่าสังเกตกลุ่มใดเราจะกำหนดขนาดของโมเมนต์เป็น M_1 , M_2 , M_3 ,.................. นั่นคือ

กำหนด M_1 เป็นโมเมนต์ที่ 1 (the first moment or the first central moment) จะได้สมการ (3-40)

$$M_1 = \frac{\sum_{i=1}^{n} d_i}{n} \qquad ...(3\text{-}40)$$

และพบว่า เมื่อ $\sum_{i=1}^{n} d_i = 0$ จึงทำให้ M_1 เป็นศูนย์ ไปด้วย แต่ถ้ากำลังของ d_i เพิ่มขึ้นเป็น d_i^2

กำหนด M_2 เป็นโมเมนต์ที่ 2 (the second moment or the second central moment) จะได้สมการ (3-41)

$$M_2 = \frac{\sum_{i=1}^{n} d_i^2}{n} \qquad ...(3\text{-}41)$$

และจะพบว่า M_2 นี้คือข้อกำหนดของความแปรปรวนตามสมการ (3-9) นั่นเอง

ในทำนองเดียวกันหากกำลังของ d_i เพิ่มขึ้นเป็น d_i^3 และ d_i^4 พร้อมกับให้ M_3 และ M_4 เป็นโมเมนต์ที่ 3 และโมเมนต์ที่ 4 (the third and the fourth moment) ตามลำดับ จะได้สมการ (3-42)

$$M_3 = \frac{\sum_{i=1}^{n} d_i^3}{n} \qquad ...(3\text{-}42)$$

$$M_4 = \frac{\sum_{i=1}^{n} d_i^4}{n} \qquad ...(3\text{-}43)$$

ตัวอย่างที่ 3.8 จงหาโมเมนต์ที่ 1 ถึงที่ 4 ของค่าสังเกตต่อไปนี้

3, 4, 5, 6, 7, 8 , 9

วิธีทำ จากข้อมูลนำไปแทนค่าในสมการ (3-40) ถึง (3-43) แล้วจะได้ดังแสดงในตารางที่ 3.9

ตารางที่ 3.9 การหาค่าโมเมนต์ที่ 1 ถึงโมเมนต์ที่ 4

X_i	การเบี่ยงเบน $d_i = (X_i - \mu)$	กำลังของ d_i			
		1^{st}	2^{nd}	3^{rd}	4^{th}
3	-3	-3	9	-27	81
4	-2	-2	4	-8	16
5	-1	-1	1	-1	1
6	0	0	0	0	0
7	1	1	1	1	1
8	2	2	4	8	16
9	3	3	9	27	81
42	0	0	28	0	196

$$M_1 = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{0}{7} = 0$$

$$M_2 = \frac{\sum_{i=1}^{n} d_i^2}{n} = \frac{28}{7} = 4$$

$$M_3 = \frac{\sum_{i=1}^{n} d_i^3}{n} = \frac{0}{7} = 0$$

$$M_4 = \frac{\sum_{i=1}^{n} d_i^4}{n} = \frac{196}{7} = 28$$

ค่าโมเมนต์ 1 ถึงค่าโมเมนต์ 4 มีค่าตามลำดับดังนี้ 0, 4, 0 และ 28

จากตัวอย่างที่ 3.3 จะเห็นว่ากำลังของ d_i ที่เพิ่มขึ้นจากกำลัง k จะได้ว่า

$$M_k = \frac{\sum_{i=1}^{n} d_i^k}{n} \qquad \text{...(3-44)}$$

โดยที่ k = 1,2,3,.......

เราเรียก M_k ตามสมการ (3-44) ว่า โมเมนต์ที่ k ของค่าสังเกต X_i รอบศูนย์กลาง กรณีศูนย์กลางเป็นศูนย์ หรือมัชฌิมเลขคณิตของค่าสังเกต X_i เท่ากับศูนย์ เราจะเรียก M_k กรณีนี้ว่า โมเมนต์ที่ k รอบจุดกำเนิด หรือ k^{th} รอว์โมเมนต์ (raw moment) และเขียนเป็นสมการ (3-45)

$$M_k = \frac{\sum_{i=1}^{n} X_i^k}{n} \qquad \text{...(3-45)}$$

ถ้า k เป็นจำนวนคี่ (odd number) จะทำให้ M_1, M_3,.. มีค่าเป็นศูนย์ บวก หรือ ลบ แต่หาก k เป็นจำนวนคู่ (even number) จะทำให้ M_2, M_4,......... มีค่าไม่เป็นศูนย์หรือบวกเท่านั้น อย่างไรก็ดีเราอาจสรุปเกี่ยวกับการใช้โมเมนต์รอบศูนย์กลางวัดการกระจายของค่าสังเกตเพียงเบื้องต้นคือ โมเมนต์ที่ 1 โมเมนต์ที่ 2 โมเมนต์ที่ 3 และโมเมนต์ที่ 4 ได้ดังนี้

โมเมนต์ที่ 1 (M_1) บอกลักษณะเฉพาะของมัชฌิมเลขคณิตของค่าสังเกต

โมเมนต์ที่ 2 (M_2) บอกความแปรปรวนของค่าสังเกต

โมเมนต์ที่ 3 (M_3) บอกความเบ้ของโค้งการแจกแจงของค่าสังเกต (skewness of frequency curve)

โมเมนต์ที่ 4 (M_4) บอกความโด่งของโค้งการแจกแจงของค่าสังเกต (kurtosis of frequency curve)

เพื่อให้เห็นชัดเจนและสะดวกยิ่งขึ้น โค้งการแจกแจงความถี่ที่ใช้พิจารณาการกระจายของค่าสังเกตกลุ่มใด ๆ เราจะใช้โค้งความถี่สัมพัทธ์ (relative frequency curve) เพราะมีพื้นที่ใต้โค้งเท่ากับ 1 หน่วย นำไปสู่การวัดค่าความเบ้ (skewness : *Sk*) และค่าความโด่ง (kurtosis : *Ku* or peakedness) ของการแจกแจงของค่าสังเกตต่อไปดังนี้

3.7.1 การวัดค่าความเบ้ ค่าความเบ้หรือเรียกว่าสัมประสิทธ์ความเบ้ (coefficient of skewness) เป็นค่าอัตราส่วนของโมเมนต์ที่ 3 ต่อกำลังสามของรากที่สองของโมเมนต์ที่ 2 ตามสมการ (3-46)

$$Sk = \frac{M_3}{\left(M_2\right)^{\frac{3}{2}}} \quad \text{...(3-46)}$$

หรือ

$$Sk = \frac{M_3}{M_2\sqrt{M_2}} \quad \text{...(3-47)}$$

และจากสมการ (3-9) ประกอบกับสมการ (1-41) เราพบว่า $\sqrt{M_2}$ หรือ $M_2^{1/2}$ คือค่าเบี่ยงเบนมาตรฐาน σ ดังนั้นค่า *Sk* จะเป็นดังสมการ (3-48)

$$Sk = \frac{M_3}{\sigma^2} \quad \text{...(3-48)}$$

ค่า *Sk* จะอยู่ระหว่าง -1 ถึง +1 กรณีการแจกแจงสมมาตร (the symmetric distribution) มี $M_3 = 0$ จึงทำให้ $Sk = 0$ ไปด้วยกรณีการแจกแจงเบ้ทางขวา (the positively skewned distribution) หรือเรียกว่าการแจกแจงมีการลาดทางขวา ค่า *Sk* จะเป็นบวก ส่วนกรณีการแจกแจงเบ้ทางซ้าย (the negatively skewed distribution) หรือเรียกว่าการแจกแจงมีการลาดทางซ้าย ค่า *Sk* จะเป็นลบ

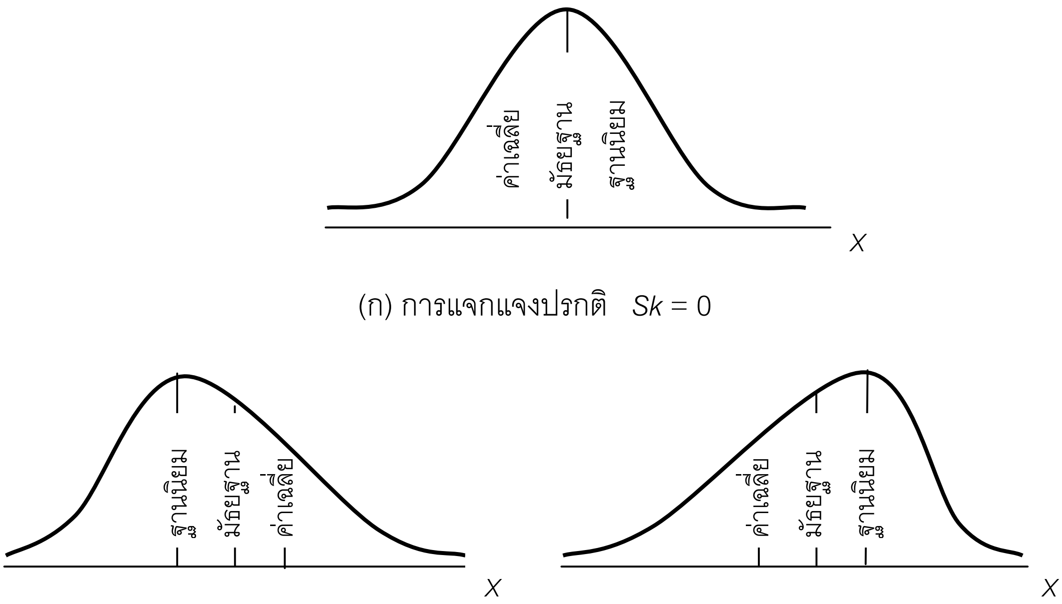
3.7.2 การวัดค่าความโด่ง ค่าความโด่งเป็นอัตราส่วนระหว่างค่าโมเมนต์ 4 ต่อโมเมนต์ 2 ดังสมการ (3-49)

$$Ku = \frac{M_4}{M_2^2} \quad \text{...(3-49)}$$

การพิจารณาความโด่งของโค้งที่จะกำหนดเป็นโค้งปรกติมาตรฐาน (the standard normal curve) มีรูปทรงเป็นระฆังอันสมบูรณ์ไว้เป็นมาตรฐานความโด่งหรือยอด (peak) ของโค้ง และเรียกโค้งนี้ว่า มีโซเคอร์ทิก (mesokurtic) การแจกแจงใดได้เป็นโค้งเช่นนี้จะเรียกการแจกแจงนั้นว่าการแจกแจงปรกติ (the normal or mesokurtic distribution) เราสามารถหาได้ว่าความโด่ง มีค่า *Ku* เท่ากับ 3 ซึ่งจะกล่าวถึงพื้นที่ใต้โค้งอย่างละเอียดในบทที่ 4

ดังนั้น โค้งการแจกแจงสมมาตรใดที่มีค่า *Ku* สูง (มากกว่า 3) ถือว่ามีความโด่งมากกว่าปกติหรือมียอดสูงมากกว่าปกติ และเรียกโค้งลักษณะนี้ว่า เลปโทเคอร์ทิก

(leptokurtic) ตรงกันข้ามหากโค้งการแจกแจงสมมาตรใดมีค่า *Ku* ต่ำ (น้อยกว่า 3) จะถือว่ามีความโด่งน้อยกว่าปกติหรือมียอดต่ำกว่าปกติและเรียกโค้งลักษณะหลังนี้ว่า พลาทิเคอร์ทิก (platykurtic)

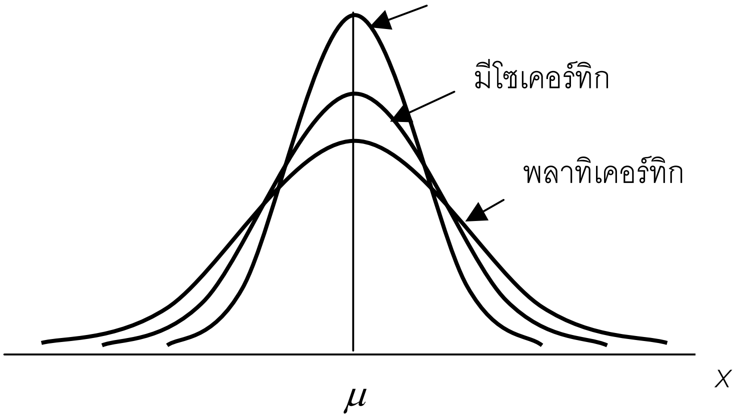


(ก) การแจกแจงปรกติ $Sk = 0$

(ข) การแจกแจงเบ้ทางขวา $Sk > 0$

(ค) การแจกแจงเบ้ทางซ้าย $Sk < 0$

ภาพที่ 3.3 ตำแหน่งค่าเฉลี่ย (μ) มัธยฐาน (*Me*) ฐานนิยม (*Mo*) และค่าความเบ้ (*Sk*)
ที่มา : (Keller & Warrack , 2000, p. 95)



ภาพที่ 3.4 เลปโทเคอร์ทิก มีโซเคอร์ทิก และ พลาทิเคอร์ทิก
ที่มา : (Casella & Berger,1990, p. 119)

เราอาจสรุปได้ว่า ไม่ว่าโค้งที่เกิดขึ้นจะมีลักษณะอย่างไร พื้นที่ใต้โค้งจะยังคงเท่ากับ 1 อยู่เสมอ จึงสามารถบ่งบอกการกระจายของค่าสังเกตกลุ่มนั้น ๆ ได้ กล่าวคือ เลปโทเคอร์ทิก เป็นค่าสังเกตที่มีกระจายน้อยกว่าปกติ แต่พลาทิเคอร์ทิกเป็นค่าสังเกตที่มีกระจายมากกว่าปกติ

3.8 บทสรุป

ในบทนี้เป็นการกล่าวถึงนิยามหลักของเครื่องมือสำหรับวัดการกระจายของข้อมูล เริ่มตั้งแต่พิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนเฉลี่ย ความแปรปรวนและค่าเบี่ยงเบนมาตรฐาน ค่ามาตรฐาน สัมประสิทธิ์ของการแปรผัน โมเมนต์รอบศูนย์กลางของข้อมูล ซึ่งสรุปได้ดังนี้

3.8.1 พิสัย เป็นการหาค่าความแตกต่างระหว่างค่าสูงสุดและค่าต่ำสุดของค่าสังเกตทั้งหมด

3.8.2 พิสัยระหว่างควอไทล์ เป็นพิสัยของค่าสังเกตในช่วงที่มีค่าสังเกตจำนวนร้อยละ 50 ของค่าสังเกตทั้งหมด

3.8.3 ส่วนเบี่ยงเบนเฉลี่ย เป็นมาตรที่ใช้ในการบอกค่าสังเกตที่มีอยู่มีค่าสมบูรณ์ของการแปรผันจากมัชฌิมเลขคณิตหรือศูนย์กลางการแจกแจงโดยเฉลี่ยเท่าใด

3.8.4 ความแปรปรวน เป็นค่าที่ใช้ในการการเบี่ยงเบนจากมัชฌิมเลขคณิตต่อหนึ่งหน่วยค่าสังเกต และค่ารากที่สองของค่าความแปรปรวน เรียกว่าค่าเบี่ยงเบนมาตรฐาน

3.8.5 ค่ามาตรฐานหรือคะแนนมาตรฐาน เป็นค่าที่ใช้ในการเปรียบเทียบค่าสังเกตจากประชากรหรือค่าสังเกตจากตัวอย่างสองกลุ่มหรือมากกว่าที่มีการแจกแจงลักษณะเดียวกัน

3.8.6 สัมประสิทธิ์ของการแปรผัน เป็นเปรียบเทียบการกระจายของประชากรหรือตัวอย่างสองกลุ่ม หรือมากกว่าสองกลุ่ม ซึ่งใช้ในการประเมินประสิทธิภาพของการทดลองนั้น ๆ

3.8.7 โมเมนต์รอบศูนย์กลาง เป็นการหาว่าค่าสังเกตแต่ละค่ามีความสมดุลรอบ ๆ มัชฌิมเลขคณิตอย่างไร

3.9 คำถามทบทวน

1. จากข้อมูลประชากรต่อไปนี้ 40, 45 , 50 , 55 , 60 , 65 , 70 , 75 ,80 , 85 , 90 และ 95 จงหา

   ก. $\sum_{i=1}^{n}\left|x_i-\mu\right|$

   ข. $\sum_{i=1}^{n}(x_i-\mu)^2$

   ค. พิสัย (R)

   ง. พิสัยระหว่างควอร์ไทล์ (Q)

   จ. ความแปรปรวน $V(X)$

   ฉ. ค่าเบี่ยงเบนมาตรฐาน (σ_x)

2. ประชากรประกอบด้วย 1.8 , 2.0 , 1.8 , 1.9 , 2.0 จงหา

   ก. ความแปรปรวนของประชากร (ทศนิยม 3 ตำแหน่ง)

   ข. ความแปรปรวนของประชากรเมื่อคูณทุกตัวด้วย 10^3

   ค. ความแปรปรวนของประชากรเมื่อบวกทุกตัวด้วย 1.5

3. ข้อมูลของกลุ่มตัวอย่างชุดหนึ่งเป็น 20, 19, 15, 17, 22, 19, 21, 18, และ 18 จงหา

   ก. ความแปรปรวนของกลุ่มตัวอย่าง (ทศนิยม 3 ตำแหน่ง)

   ข. ความแปรปรวนของกลุ่มตัวอย่างเมื่อคูณทุกตัวด้วย 10^3

   ค. ความแปรปรวนของกลุ่มตัวอย่างเมื่อบวกทุกตัวด้วย 1.5

4. จากบันทึกประจำวันราคาสินค้าเกษตรเดือนเมษายนหลาย ๆ ปี ราคาพริกสดเฉลี่ยกิโลกรัมละ 80 บาท ค่าเบี่ยงเบนมาตรฐาน 20 บาท ราคาหอมแบ่งเฉลี่ยกิโลกรัมละ 18 บาท ค่าเบี่ยงเบนมาตรฐาน 2 บาท ในเดือนเมษายน ปีนี้ท่านควรเลือกปลูกอะไรในสองอย่างนี้จึงจะได้ราคาที่แน่นอนกว่า

5. จงเปรียบเทียบการกระจายของรายได้พนักงานต่อปีแต่คนจาก 2 โรงงานจากข้อมูลในตารางที่ 3.10

ตารางที่ 3.10 ข้อมูลรายได้ของพนักงานสายการผลิตต่อปีจากโรงงาน 2 โรงงาน

	โรงงาน ก.	โรงงาน ข.
ก. รายได้เฉลี่ยต่อปีแต่ละคน	30,000	50,000
ข. ค่าเบี่ยงเบนมาตรฐาน	5,000	700

6. จากข้อมูลในตารางที่ 3.11 จงหา
   ก. โมเมนต์ 2 , 3 และ 4 ของข้อมูล
   ข. ค่าความเบ้และค่าความโด่งของข้อมูล

ตารางที่ 3.11 ความถี่ของอายุของนักเรียนกลุ่มหนึ่ง

อายุ (ปี)	ความถี่
15	1
14	1
13	2
12	2
11	5
10	6
9	9
8	3
7	3
6	2
5	1
	35

7. ในการสอบปลายภาควิชาสถิติของนักศึกษาจำนวน 50 คน ได้ข้อมูลดังตารางที่ 3.12 จงหา

ก. ความเบี่ยงเบนมาตรฐานของการสอบ

ข. โมเมนต์ 2, 3 และ 4 ของการสอบ

ค. ค่าความเบ้และค่าความโด่งของการสอบ

ตารางที่ 3.12 ผลการสอบปลายภาควิชาสถิติของนักศึกษาจำนวน 50 คน และความถี่ของข้อมูล

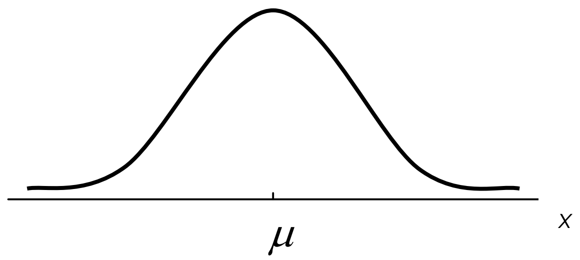
ผลการสอบ	ความถี่
45	2
44	3
43	1
40	5
39	4
35	10
30	5
29	10
28	7
27	2
26	1
	50

บทที่ 4
การแจกแจงปรกติของข้อมูล

การวัดการกระจายของค่าสังเกตในบทที่ 3 ที่ได้กล่าวถึงการแจกแจงปรกติจนถึงค่ามาตรฐาน ซึ่งเป็นที่ยอมรับกันว่าความเป็นจริงในธรรมชาติและความเป็นจริงในชีวิตประจำวันของมนุษย์ ค่าสังเกตที่พบอยู่เสมอ เช่น คะแนนสอบวิชาต่าง ๆ ค่าที่วัดจากห้องปฏิบัติการทางชีววิทยา เคมี ฟิสิกส์ ทางการแพทย์ ทางธุรกิจหรืออื่น ๆ มีการแจกแจงปรกติ หรือถือว่าการแจกแจงปรกติในขอบเขตของสถิติพรรณนาเป็นการกล่าวถึงการแจกแจงปรกติของความถี่ของประชากรที่ศึกษา

4.1 เส้นโค้งปรกติ

การเขียนกราฟบนพื้นระนาบที่ประกอบด้วยแกนนอน X และแกนตั้ง Y โดยที่ค่า y เปลี่ยนแปลงตามค่า x [มักเรียกว่า y เป็นฟังก์ชันของ x หรือแทนว่า $f(x)$] จะได้ทางเดินของจุดความสัมพันธ์ระหว่าง x และ y เป็นรูปทรงต่าง ๆ บนพื้นระนาบนั้น เช่น เส้นตรง พาราโบลา ไฮเพอร์โบลา วงกลม เส้นโค้ง หรือ อื่น ๆ เส้นโค้งปรกติ (normal curve) เป็นทางเดินของจุดความสัมพันธ์ระหว่าง x กับ y มีรูปลักษณะเป็นระฆัง (bell-shape curve) และมี $x = \mu$ ค่าหนึ่งเป็นศูนย์กลางสมมาตร (center of symmetry) ดังภาพที่ 4.1

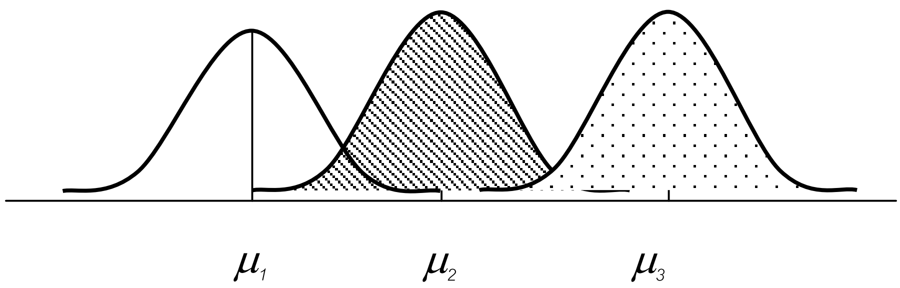


ภาพที่ 4.1 แสดงเส้นโค้งปรกติ

อะบราฮัม เดอ มัวฟ์ (Abraham de Moivre:1667-1754) นักคณิตศาสตร์ชาวฝรั่งเศส ได้พบสมการเส้นโค้งปรกตินี้เป็นคนแรกประมาณปี ค.ศ. 1733 แต่ไม่ได้นำไปใช้ในทางปฏิบัติ จนกระทั่ง คาร์ล ฟรีดริช เกาส์ (Carl Friedrich Gauss :1777-1855) ชาวเยอรมัน ได้นำสมการเส้นโค้งปรกติไปใช้หาค่าผิดพลาดในการวัดระยะทางดาราศาสตร์ โดยกำหนดแกน *X* (abscissa : พิกัดที่หนึ่ง) เป็นค่าผิดพลาดในการวัด และแกน *Y* (ordinate : พิกัดที่สอง) เป็นความถี่ของค่าผิดพลาดที่พบ ซึ่งอาจมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ หรือ *X* มีจำนวนมากจนนับไม่ได้ (uncountably infinite or non-denumerable) ค่าเฉลี่ยเท่ากับ μ และค่าความแปรปรวนเท่ากับ σ^2 (Keller & Warrack , 2000, p. 241)

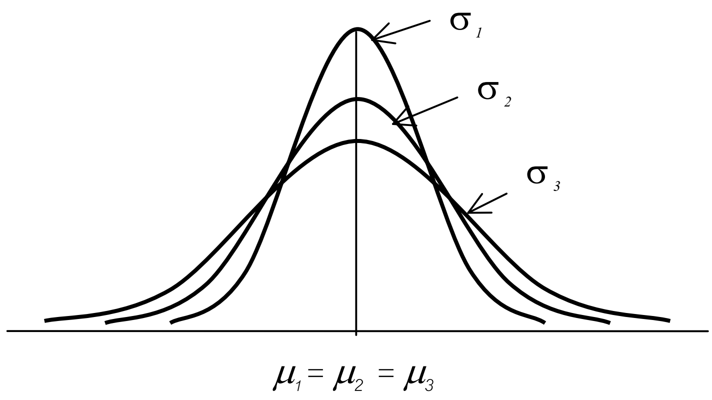
$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad -\infty \le X \le +\infty \qquad \ldots(4\text{-}1)$$

ศูนย์กลางความสมมาตรบนแกน *X* คือ μ ซึ่งค่า $\frac{1}{\sqrt{2\pi}}$ เป็นค่าคงตัวที่ทำให้พื้นที่ใต้โค้งปรกตินี้เท่ากับ 1 หน่วย *e* เป็นฐานของลอการิทึมธรรมชาติ (natural logarithm or napierian logarithm)ซึ่งมีค่าประมาณ 2.71828 และ π มีค่าประมาณ 3.1416 ให้คนทั่วไปรู้จักเส้นโค้งปรกติในนาม การแจกแจงของเกาส์ (Gaussian distribution or the normal function of error) สำหรับค่า σ^2 จะบอกค่าความโด่ง (*Ku*) ตามสมการ (3-49) ดังแสดงในภาพที่ 4.2 และ 4.3



ภาพที่ 4.2 การเปรียบเทียบเส้นโค้งปรกติกรณี $\mu_1 < \mu_2 < \mu_3$ แต่ $\sigma_1 = \sigma_2 = \sigma_3$

ที่มา : (ดัดแปลงมาจาก Casella & Berger, 1990, p. 116)



ภาพที่ 4.3 การเปรียบเทียบเส้นโค้งปรกติกรณี $\mu_1 = \mu_2 = \mu_3$ แต่ $\sigma_1 < \sigma_2 < \sigma_3$

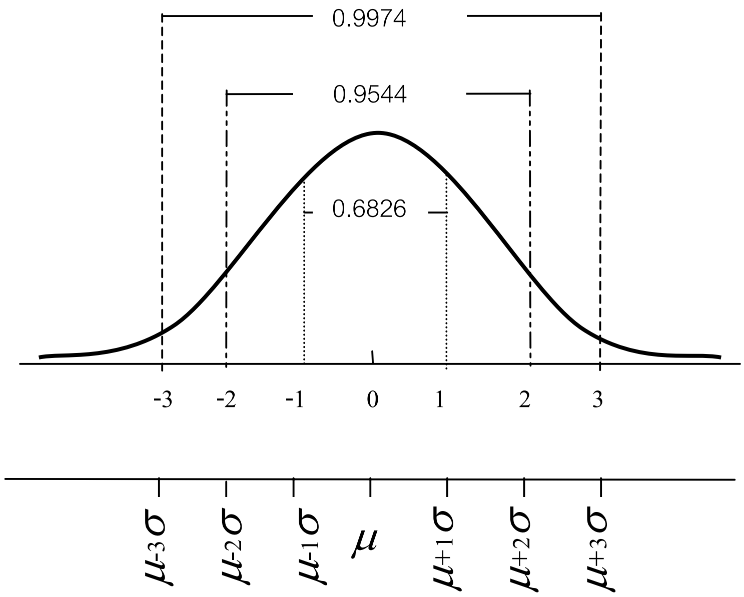
ที่มา : (ดัดแปลงมาจาก Casella & Berger, 1990, p. 110)

สำหรับเส้นโค้งปรกติที่มี $\mu = 0$ และ $\sigma = 1$ เราเรียกว่า เส้นโค้งปรกติมาตรฐาน (the standard normal curve) เพื่อให้แตกต่างจากเส้นปกติทั่วไป จึงกำหนดมาตราบนแกน X ของกราฟ จากมาตราส่วนเอกซ์เป็นมาตราส่วนซี ดังนั้นศูนย์กลางความสมมาตรของเส้นโค้งปรกติมาตรฐาน จึงอยู่ที่ $Z = 0$ หรือ $X = \mu = 0$ และจากสมการ (3-27) จึงทำให้สามารถเขียน สมการเส้นโค้งปรกติมาตรฐานได้เป็นสมการ (4-2)

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z)^2} \quad ; \quad -\infty \leq Z \leq +\infty \qquad \text{...(4-2)}$$

4.2 พื้นที่ใต้เส้นโค้งปรกติ

พื้นที่ใต้เส้นโค้งปรกติและพื้นที่ใต้เส้นโค้งปรกติมาตรฐานจะมีเท่ากับ 1 หน่วยเหมือนกัน จะแตกต่างกันเฉพาะศูนย์กลางความสมมาตรและค่าเบี่ยงเบนมาตรฐานเท่านั้น จึงสามารถเปรียบเทียบมาตราส่วนเอกซ์และมาตราส่วนซี ได้ดังภาพที่ 4.4 และจะพบว่าพื้นที่ใต้เส้นโค้งปรกติ 0.6826 หน่วยหรือร้อยละ 68.26 ของพื้นที่ทั้งหมดอยู่ระหว่างค่า Z = -1 กับ +1 (ค่า x ระหว่าง $\mu-\sigma$ กับ $\mu+\sigma$) พื้นที่ 0.9544 หน่วยหรือร้อยละ 95.44 ของพื้นที่ทั้งหมดอยู่ระหว่างค่า Z = -2 กับ +2 (ค่า X ระหว่าง $\mu - 2\sigma$ กับ $\mu + 2\sigma$) และพื้นที่ 0.9974 หน่วยหรือร้อยละ 99.74 ของพื้นที่ทั้งหมดอยู่ระหว่างค่า Z = -3 กับ +3 (ค่า X ระหว่าง $\mu - 3\sigma$ กับ $\mu + 3\sigma$)



ภาพที่ 4.4 การเปรียบเทียบพื้นที่ใต้เส้นโค้งปรกติมาตรฐานของ
มาตราส่วนเอกซ์และมาตราส่วนซี
ที่มา : (ดัดแปลงมาจาก Wonnacott & Wonnacott ,1985, p.105)

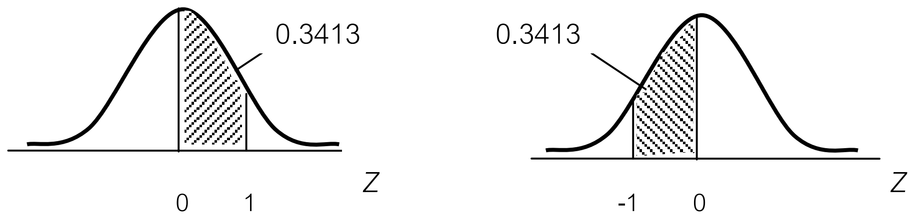
ตารางภาคผนวกที่ 5 เป็นตารางการแสดงการแจกแจงปรกติ ซึ่งจะบอกพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ระหว่าง Z = 0.00 กับ Z = 3.09 และบอก พื้นที่ใต้เส้นโค้งปรกติไม่เกิน 0.5 หน่วยเฉพาะพื้นที่ทางซีกขวาของศูนย์ความสมมาตร (Z = 0) เท่านั้น แต่เนื่องจากเส้นโค้งปรกติมีความลาดทางซีกขวาและซีกซ้ายเหมือนกัน ตารางนี้จึงสามารถบอกพื้นที่ใต้เส้นโค้งปรกติมาตรฐานทางซีกซ้ายระหว่าง Z = 0.00 กับ -3.09 ได้เช่นกัน

ตัวอย่างที่ 4.1 จงหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อค่า Z อยู่ระหว่าง

ก. 0 กับ +1

ข. 0 กับ -1

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้โค้งปรกติมาตรฐาน ได้ดังแสดงในภาพที่ 4.5



(ก) $0<Z<+1$ (ข) $-1<Z<0$

ภาพที่ 4.5 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อค่า Z อยู่ระหว่าง 0 กับ 1 และ 0 กับ -1

เมื่ออ่านพื้นที่ใต้เส้นโค้งปรกติมาตรฐานจากตารางภาคผนวกที่ 5 จะได้ค่าดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 ค่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน จากตารางภาคผนวกที่ 5 เมื่อค่า Z อยู่ระหว่าง 0 กับ 1 และ 0 กับ -1

Z	.00	.01	………….. ………………………………….	.09
0.00	.0000			
0.10	.			
0.20	.			
.	.			
.	$\Downarrow$			
1.00	$\Rightarrow$ .3413			

นั่นคือพื้นที่ใต้เส้นโค้งปรกติมาตรฐานตามตารางที่ 4.1 เมื่อ

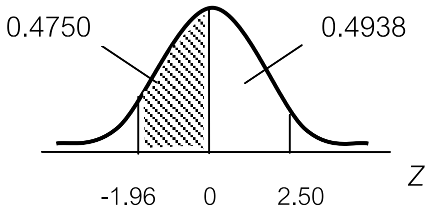
ก. $0 < Z < +1$ มี 0.3413 หน่วย

ข. $-1 < Z < 0$ มี 0.3413 หน่วย

จากตัวอย่างที่ 4.1 จะเห็นได้ว่า เมื่อ Z มีค่าระหว่าง -1 กับ +1 มีพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เป็น 0.3413 + 0.3413 = 0.6826 หน่วยหรือร้อยละ 68.26 ของพื้นที่ทั้งหมดที่แสดงไว้ในภาพที่ 4.4

ตัวอย่างที่ 4.2 จงหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ -1.96< *Z* <+ 2.5 (อ่านว่า *Z* มากกว่า -1.96 และน้อยกว่า + 2.50 หรือ *Z* มีค่าระหว่าง -1.96 กับ +2.50)

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ได้ดังแสดงใน ภาพที่ 4.6



ภาพที่ 4.6 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อค่า *Z* อยู่ระหว่าง –1.96 กับ 2.50

เมื่ออ่านพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน จากตารางภาคผนวกที่ 5 จะได้ค่าดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 ค่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน จากตารางภาคผนวกที่ 5 เมื่อค่า *Z* อยู่ระหว่าง 0 กับ–1.96 และ 0 กับ 2.50

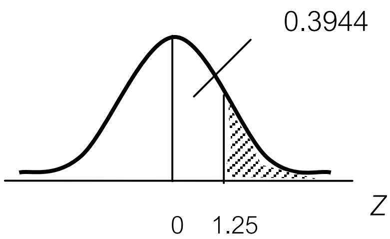
Z	.00	.01	……	.06	………………………………………	.09
0.00	.0000			.0239		
0.10	.			.		
0.20	.			.		
.	.			.		
.	.			⇓		
1.90	……………………⇒			.4750		
.	⇓					
2.50	⇒ .4938					

นั่นคือพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $-1.96 < Z < 0$ มี 0.4750 หน่วย และ $0.00 < Z < +2.50$ มี 0.4938 หน่วย

ดังนั้นพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $-1.96 < Z < +2.50$ จึงมีค่าเท่ากับ $0.4750 + 0.4938 = 0.9688$ หน่วย

ตัวอย่างที่ 4.3 จงหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $Z \geq +1.25$

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ได้ดังแสดงในภาพที่ 4.7



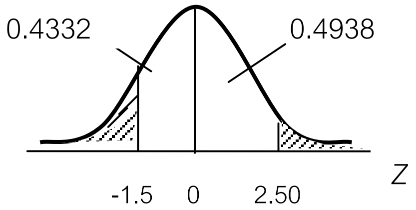
ภาพที่ 4.7 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อค่า Z อยู่ระหว่าง 0 กับ 1.25

เมื่ออ่านพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน จากตารางภาคผนวกที่ 5 นั่นคือพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $0 \leq Z \leq +\infty$ มี 0.5000 หน่วย แต่ $0 \leq Z \leq +1.25$ มีค่าเท่ากับ 0.3944 หน่วย

ดังนั้นพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $Z \geq +1.25$ จึงมีเท่ากับ $0.5000 - 0.3944$ เท่ากับ 0.1056 หน่วย หรือร้อยละ 10.56 ของพื้นที่ทั้งหมด

ตัวอย่างที่ 4.4 จงหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $-1.5 \geq Z$ และ $Z \geq +2.5$

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ได้ดังแสดงในภาพที่ 4.8



ภาพที่ 4.8 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อค่า Z อยู่ระหว่าง –1.5 กับ 2.50

เมื่ออ่านพื้นที่ใต้เส้นโค้งปรกติมาตรฐานจากตารางภาคผนวกที่ 5 จะได้ว่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $0 \leq Z \leq +\infty$ มี 0.5 หน่วย แต่ $0 \leq Z \leq +2.5$ มี 0.4938 หน่วย

ดังนั้นเมื่อ $Z \geq +2.5$ จึงเหลือพื้นที่ 0.5 - 0.4938 เท่ากับ 0.0062 หน่วย และพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $-\infty \leq Z \leq 0$ มี 0.5 หน่วยแต่เมื่อ $-1.5 \leq Z \leq 0$ มี 0.4332 หน่วย ดังนั้นเมื่อ $Z \leq -1.5$ จึงเหลือพื้นที่ 0.5 – 0.4332 เท่ากับ 0.0668 หน่วย

ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 ค่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน จากตารางภาคผนวกที่ 5 เมื่อค่า Z อยู่ระหว่าง 0 กับ -1.5 และ 0 กับ 2.5

Z	.00	.01	…… .06	………………………………… .09
0.00	.0000			
0.10	.			
0.20	.			
.	.			
.	.			
1.50	⇒ .4332			
.	.			
	⇓			
2.50	⇒ .4938			

นั่นคือพื้นที่ใต้เส้นโค้งปรกติ เมื่อ $-1.5 \geq Z$ และ $Z \geq +2.5$ จึงมีเท่ากับ 0.0062+0.0668 = 0.0730 หน่วย หรือ ร้อยละ 7.3 ของพื้นที่ทั้งหมด

4.3 คุณสมบัติสำคัญของการแจกแจงปรกติของข้อมูล

การแจกแจงความถี่ของค่าสังเกตหรือตัวแปร X ของกลุ่มใด ๆ ที่จำนวนค่าสังเกตหรือตัวแปรมีมากมายจนนับไม่ได้ มีค่าเฉลี่ยเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ และจุดสูงสุดของโค้งปรกติตรงกับ μ โค้งค่อย ๆ ลาดลงทางซ้ายและทางขวาเหมือน ๆ กัน ส่วนปลายเข้าใกล้แกน X จนเป็นรูประฆังคว่ำศูนย์กลางความสมมาตรอยู่ที่ μ เราจะเรียกว่า ค่าสังเกตหรือตัวแปร X มีการแจกแจงปรกติ ในกรณีที่ X_1, X_2 ,............,X_n เป็นค่าสังเกตจำนวน n ค่าที่แจกแจงปรกติ ตัวมัชฌิมคณิตเท่ากับ μ และความแปร σ^2 แล้วเขียนสัญลักษณ์แทนเป็นสมการ (4-3)

$$X \sim N(\mu, \sigma^2) \qquad \ldots(4\text{-}3)$$

สมการ (4-3) อ่านว่า X มีการแจกแจงปรกติ ค่าเฉลี่ยเท่ากับ μ ความแปรปรวนเท่ากับ σ^2 สมมติว่า X เป็นคะแนนสอบวิชาสถิติธุรกิจของนักศึกษาหลายพันคนแจกแจงปรกติคะแนนเฉลี่ย 52 ความแปรปรวน 14 เขียนสัญลักษณ์แทนด้วย $X \sim N$ (52,14) เป็นต้น

กรณีที่ X มีการแจกแจงปรกติ $\mu = 0$ และ $\sigma^2 = 1$ เราจะเรียกว่า X มีการแจกแจงปรกติมาตรฐาน (standard normal distribution) แต่ในสถานการณ์ทั่วไปมาตราส่วนเอกซ์จะมี $\mu \neq 0$ และ $\sigma^2 \neq 1$ จะมีเฉพาะมาตราส่วนซีเท่านั้น ทำให้การแจกแจงปรกติมาตรฐานของความถี่จึงมีเฉพาะกรณีแจกแจงปรกติมาตรฐานเขียนแทนว่าด้วยสมการ (4-4)

$$Z \sim N(0,1) \qquad \ldots(4\text{-}4)$$

สมการ (4-4) อ่านว่า Z มีการแจกแจงปรกติมาตรฐาน (ค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 1) ด้วยเหตุนี้จะได้ว่า การแจกแจงปรกติมาตรฐาน และพื้นที่ใต้โค้งการแจกแจงความถี่ของค่ามาตรฐานก็คือ ความหนาแน่นของความถี่สัมพัทธ์ของ Z (relative frequency density of Z) หรือโอกาสพบ Z (probability of Z) ถ้า X_1 , X_2 ,....,X_n มีการแจกแจงปรกติ เราสามารถเปลี่ยนมาตราส่วนเอกซ์มาเป็นมาตราส่วนซีทำให้พื้นที่ใต้โค้งการ

แจกแจงความถี่ของ X คือ ความหนาแน่นของความถี่สัมพัทธ์ของ X (relative frequency density of X) คือ ความหนาแน่นของ X_i เมื่อเทียบกับทั้งหมด เท่ากับ $\frac{f_i}{n}$ หรือ โอกาสพบ X เช่นเดียวกัน

อย่างไรก็ดี เมื่อค่าสังเกต X มีการแจกแจงปรกติ จะได้เส้นโค้งความถี่หรือเส้นโค้งความถี่สัมพัทธ์เป็นดังนี้

4.3.1 ลักษณะของเส้นโค้งเป็นรูประฆังคว่ำทรงสมมาตร กล่าวคือเมื่อลากเส้นตรงจากจุดสูงสุดของโค้งไปตั้งฉากกับฐาน (แกน X) แล้วพับตามรอยเส้นตั้งฉากนี้โค้งทั้งซ้ายและขวาจะทับกันสนิททุกจุด

4.3.2 ความโด่งของโค้งจะขึ้นอยู่กับค่า σ^2 กล่าวคือถ้า σ^2 มีค่าน้อยโค้งจะมีความโด่งมาก ตรงข้ามถ้า σ^2 มีค่ามากโค้งจะมีความโด่งน้อย

4.3.3 จุดสมมาตรของโค้ง คือ ค่าเฉลี่ย มัธยฐานและฐานนิยมของค่าสังเกตกลุ่มนั้น ๆ

4.3.4 ปลายโค้งทั้งซ้ายและขวาจะไม่จดกับฐาน แต่จะมีลักษณะเข้าใกล้ฐาน นั่นคือ X มีค่าไม่จำกัดทั้งซ้ายขวา $(-\infty \leq X \leq +\infty)$

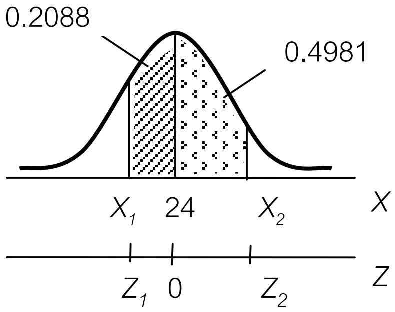
4.3.5 พื้นที่ใต้เส้นโค้งความถี่ คือ ความถี่สัมพัทธ์รวมเท่ากับ 1

4.4 การประยุกต์ใช้พื้นที่ใต้โค้งปรกติมาตรฐาน

จากคุณสมบัติของการแจกแจงปรกติของข้อมูล พื้นที่ใต้เส้นโค้งปรกติ คือ ความถี่สัมพัทธ์รวมเท่ากับ 1 หรือร้อยละ 100 ของจำนวนค่าสังเกตทั้งหมดอยู่ใต้เส้นโค้งปรกติมาตรฐาน จึงสามารถหาจำนวนของค่าสังเกตค่าใดค่าหนึ่ง หรือหลายค่า หรือหาจำนวนของค่าสังเกตค่าช่วงใดช่วงหนึ่ง โดยอาศัยตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐานในตารางภาคผนวกที่ 5 ดังตัวอย่างที่ 4.5

ตัวอย่างที่ 4.5 ค่าสังเกต X กลุ่มหนึ่งมีการแจกแจงปรกติ มีค่าเฉลี่ยเท่ากับ 24 ค่าเบี่ยงเบนมาตรฐานเท่ากับ 12 จงหาจำนวนค่าสังเกตระหว่าง X_1 =17.4 กับ X_2 = 58.8

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ได้ดังแสดงในภาพที่ 4.9



ภาพที่ 4.9 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อเปลี่ยนค่าจากมาตราส่วนเอกซ์ เป็นมาตราส่วนซีเมื่อ X_1 =17.4 กับ X_2= 58.8

การเปลี่ยนค่าสังเกต X_i ไปเป็นค่ามาตรฐาน Z_i

$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

เมื่อ μ = 24 , σ = 12

ค่ามาตรฐานของ X_1 =17.4 คือ

$$Z_1 = \left(\frac{17.4 - 24}{12}\right)$$

$$= -0.55$$

ค่ามาตรฐานของ X_2 = 58.8 คือ

$$Z_2 = \left(\frac{58.8 - 24}{12}\right)$$

$$= +2.90$$

จากตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $-0.55 \le Z \le 0$ มีพื้นที่ใต้เส้นโค้งปรกติ 0.2088 หน่วยและ $0 \le Z \le +2.90$ มีพื้นที่ใต้เส้นโค้งปรกติ 0.4981 หน่วย

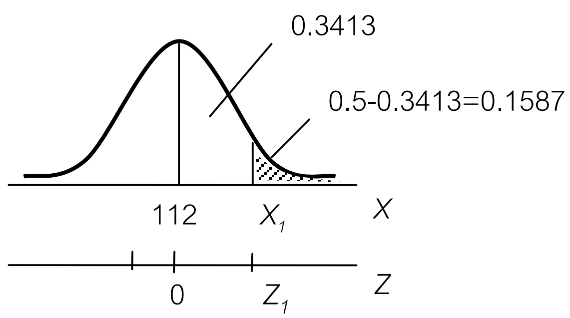
ดังนั้นพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน $-0.55 \le Z \le +2.90$ คือ พื้นที่ใต้เส้นโค้งปรกติมาตรฐานเมื่อ $17.4 \le X \le 58.8$ คือ พื้นที่ $0.2088 + 0.4981 = 0.7069$ หน่วย

นั่นคือ จำนวนค่าสังเกตระหว่าง $X_1 = 17.4$ กับ $X_2 = 58.8$ มีร้อยละ 70.69 ของค่าสังเกตทั้งหมด

จะเห็นว่าในตัวอย่างที่ 4.5 เราสามารถบอกจำนวนค่าสังเกตเป็นร้อยละของค่าสังเกตทั้งหมด ถ้าหากทราบจำนวนค่าสังเกตทั้งหมด (n) ก็จะสามารถบอกจำนวนค่าสังเกตระหว่าง X_1 กับ X_2 เช่น $n = 200$ จะได้ว่าจำนวนค่าสังเกตระหว่าง 17.4 กับ 58.8 มีจำนวนเท่ากับ 200(0.7068) =141.34 ค่า และสมมุติว่าค่าสังเกตจำนวนนี้เป็นน้ำหนักพนักงาน 200 คน จะแสดงให้เห็นว่าพนักงานที่หนักระหว่าง 17.4 ถึง 58.8 กิโลกรัม มีจำนวน 141 คน (141.34 $\approx$ 141 คน)

ตัวอย่างที่ 4.6 คะแนนสอบวิชาสถิติธุรกิจของนักศึกษา 200 คน มีการแจกแจงปรกติ คะแนนเฉลี่ย 112 คะแนน ค่าเบี่ยงเบนมาตรฐาน 2 คะแนน จงหาจำนวนนักศึกษาที่ได้คะแนนตั้งแต่ 114 คะแนนขึ้นไป

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ได้ ดังแสดงในภาพที่ 4.10



ภาพที่ 4.10 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อเปลี่ยนค่าจากมาตราส่วนเอกซ์เป็นมาตราส่วนซีของคะแนนสอบวิชาสถิติธุรกิจของนักศึกษา

การเปลี่ยนค่าสังเกต X_i ไปเป็นค่ามาตรฐาน Z_i

$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

เมื่อ $\mu = 112 , \sigma = 2$

ค่ามาตรฐานของ X_1 =114 คือ

$$Z_1 = \left(\frac{114-112}{2}\right) = 1$$

เปิดตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 อ่านได้เมื่อ $0< Z \leq 1$ มีพื้นที่ 0.3413 หน่วยพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $Z \geq 1$ คือ พื้นที่ใต้เส้นโค้งปรกติเมื่อ $X \geq 114$ คือ พื้นที่ 0.5 - 0.3413 เท่ากับ 0.1587 หน่วย

นั่นคือ จำนวนนักศึกษาที่ได้คะแนนตั้งแต่ 114 คะแนนขึ้นไปมีจำนวนเท่ากับ $200(0.1587) = 31.74 \approx 32$ คน

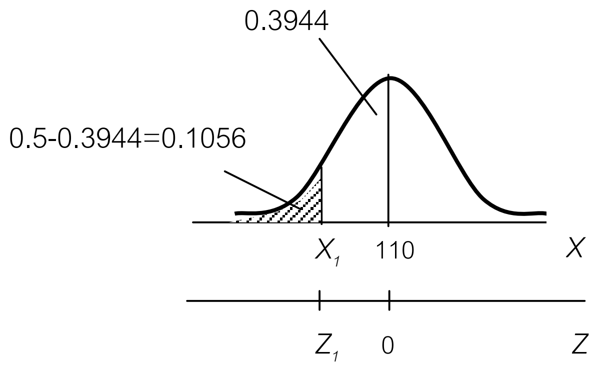
ตัวอย่างที่ 4.7 จากการศึกษาพบว่ารายได้พนักงาน 10,000 คนในบริษัทหนึ่งมีการแจกแจงปรกติ เฉลี่ยวันละ 110 บาทต่อคน ค่าเบี่ยงเบนมาตรฐาน 8 บาทต่อคน จงหา

ก. จำนวนพนักงานที่มีรายได้ไม่เกินวันละ 100 บาท

ข. จำนวนเพนักงานที่มีรายได้มากกว่าวันละ 106 บาท

ค. จำนวนพนักงานที่มีรายได้ระหว่างวันละ 100 ถึง 120 บาท

วิธีทำ ก. เมื่อจำนวนพนักงานที่มีรายได้ไม่เกินวันละ 100 บาท ($X \leq 100$)



ภาพที่ 4.11 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อเปลี่ยนค่าจากมาตราส่วนเอกซ์

เป็นมาตราส่วนซีของรายได้ของพนักงาน

การเปลี่ยนค่าสังเกต X_i ไปเป็นค่ามาตรฐาน Z_i

$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

เมื่อ $\mu = 110$, $\sigma = 8$

ค่ามาตรฐานของ $X_1 = 100$ คือ

$$Z_1 = \left(\frac{100 - 110}{8}\right)$$

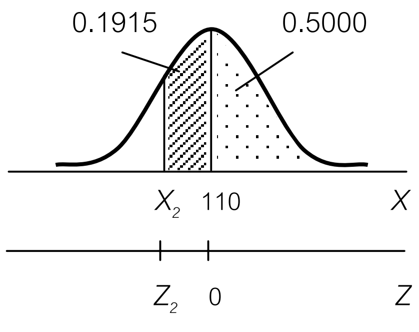
$$= -1.25$$

จากตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 เมื่อ $-\infty \le Z \le 0$ มีพื้นที่ใต้เส้นโค้ง 0.5000 หน่วย แต่ $-1.25 \le Z \le 0$ มีพื้นที่ใต้เส้นโค้ง 0.3944 หน่วย

นั่นคือพื้นที่ใต้เส้นโค้งปรกติมาตรฐานเมื่อ $Z \le -1.25$ คือพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $X \le 100$ คือพื้นที่ 0.5 - 0.3944 เท่ากับ 0.1056 หน่วย

ดังนั้นจำนวนพนักงานที่มีรายได้ไม่เกินวันละ 100 บาท มีจำนวนเท่ากับ 10,000(0.1056) เท่ากับ 1,056 คน

ข. เมื่อจำนวนพนักงานที่มีรายได้มากกว่าวันละ 106 บาท ($X > 106$)



ภาพที่ 4.12 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อเปลี่ยนค่าจากมาตราส่วนเอกซ์เป็น

มาตราส่วนซีของรายได้ของพนักงานที่มีรายได้มากกว่าวันละ 106 บาท

การเปลี่ยนค่าสังเกต X_i ไปเป็นค่ามาตรฐาน Z_i

$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

เมื่อ $\mu = 110$, $\sigma = 8$

ค่ามาตรฐานของ $X_2 = 106$ คือ

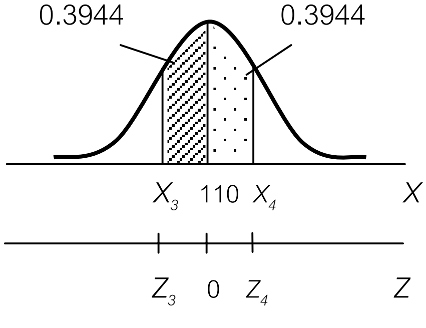
$$Z_2 = \frac{106-110}{8} = -0.5$$

จากตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 เมื่อ $0 < Z < +\infty$ มีพื้นที่ 0.5000 หน่วยและ $-0.5 < Z < 0$ มีพื้นที่ 0.1915 หน่วย

ดังนั้นพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $Z > -0.5$ คือพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $X > 106$ คือ พื้นที่ใต้เส้นโค้งปรกติ 0.5 + 0.1915 เท่ากับ 0.6915 หน่วย

นั่นคือ จำนวนพนักงานที่มีรายได้มากกว่าวันละ 106 บาท มีจำนวนเท่ากับ 10,000 (0.6915) เท่ากับ 6,915 คน

ค. หาจำนวนพนักงานที่มีรายได้ระหว่างวันละ 100 ถึง 120 บาท ($100 \leq X \leq 120$ หรือ ระหว่าง $X_3 = 100$ กับ $X_4 = 120$)



ภาพที่ 4.13 พื้นที่ใต้เส้นโค้งปรกติมาตรฐานเมื่อเปลี่ยนค่าจาก

มาตราส่วนเอกซ์เป็นมาตราส่วนซีของรายได้ ของ
พนักงานที่มีรายได้ระหว่างวันละ 100 ถึง 120 บาท

การเปลี่ยนค่าสังเกต X_i ไปเป็นค่ามาตรฐาน Z_i

$$Z_i = \frac{X_i - \mu}{\sigma}$$

เมื่อ $\mu = 110$, $\sigma = 8$

ค่ามาตรฐานของ $X_3 = 100$ คือ

$$Z_3 = \left(\frac{100-110}{8}\right)$$

$$= -1.25$$

ค่ามาตรฐานของ $X_4 = 120$ คือ

$$Z_4 = \left(\frac{120-110}{8}\right)$$

$$= 1.25$$

จากตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐานในตารางภาคผนวกที่ 5 เมื่อ $-1.25 \leq Z \leq 0$ มีพื้นที่ใต้เส้นโค้ง 0.3944 หน่วยและ $0 \leq Z \leq 1.25$ มีพื้นที่ใต้เส้นโค้ง 0.3944 หน่วย

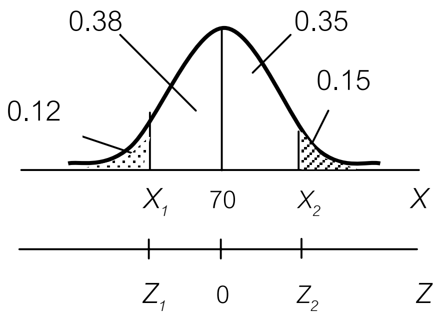
ดังนั้นพื้นที่ใต้เส้นโค้งปรกติมาตรฐานเมื่อ $-1.25 \leq Z \leq 1.25$ คือพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ $100 \leq X \leq 120$ คือพื้นที่ 0.3944 + 0.3944 = 0.7888 หน่วย

นั่นคือ จำนวนพนักงานที่มีรายได้ระหว่างสัปดาห์ละ 100 ถึง 120 บาท มีจำนวนเท่ากับ 10,000(0.7888) เท่ากับ 7,888 คน

ตัวอย่างที่ 4.8 น้ำหนักของนักศึกษาชายกลุ่มหนึ่งการแจกแจงปรกติน้ำหนักเฉลี่ย 70 กิโลกรัม ค่าเบี่ยงเบนมาตรฐาน 6.3 กิโลกรัม หากต้องการแยกนักศึกษาน้ำหนักเกณฑ์ต่ำ 12 % และน้ำหนักเกณฑ์สูง 15 % ของจำนวนนักศึกษาชายทั้งหมด จงหาค่าน้ำหนักสูงสุดของพวกน้ำหนักเกณฑ์ต่ำ และน้ำหนักต่ำสุดของพวกเกณฑ์สูง

Xxxx (หน้านี้ไม่ใช้แต่ต้องพิมพ์)

วิธีทำ ใช้ตารางภาคผนวกที่ 5 ในการหาพื้นที่ใต้เส้นโค้งปรกติมาตรฐานได้ดังแสดงใน ภาพที่ 4.14



ภาพที่ 4.14 พื้นที่ใต้เส้นโค้งปรกติมาตรฐานของน้ำหนักของนักศึกษาชาย

กำหนดให้

$X_1 = Z_1$ คือ ค่าน้ำหนักสูงสุดของนักศึกษาเกณฑ์ต่ำ

$X_2 = Z_2$ คือ น้ำหนักต่ำสุดของของนักศึกษาเกณฑ์สูง

พื้นที่ใต้โค้งปรกติมาตรฐานระหว่าง 0 กับ Z_1 มีพื้นที่เท่ากับ 0.5 - 0.12 = 0.38 หน่วย ระหว่าง 0 กับ Z_2 มีพื้นที่เท่ากับ 0.5 - 0.15 = 0.35 หน่วย อาศัย ตารางพื้นที่ใต้โค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 หาค่า Z_1 และ Z_2

ตารางที่ 4.4 ค่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อ ค่า Z อยู่ระหว่าง 0 กับ 1.04 และระหว่าง 0 กับ 1.17

Z	.00	.01	…0.4………	0.7…………………	. 09
0.00	.0000		⇑	⇑	
0.10			.	.	
0.20			.	.	
.			.	.	
1.00	⇐…… ……………		3508	.	
1.10	⇐	……………………		3790	

จากพื้นที่ใกล้เคียงที่สุด คือ .3790 $\approx$ .38 หน่วย และพื้นที่ .3508 $\approx$ .35 หน่วย จึงได้ Z_1 = -1.17 และ Z_2 = 1.04 เปลี่ยนค่ามาตรฐาน Z_1 และ Z_2 ไปเป็น X_1 และ X_2

$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right); \quad \mu = 70\,,\ \sigma = 6.3$$

$$-1.17 = \left(\frac{X_1 - 70}{6.3}\right)$$

$$X_1 = (-1.17)\,(6.3) + 70$$

$$= (-7.371) + 70$$

$$= 62.63$$

และ

$$1.04 = \left(\frac{X_2 - 70}{6.3}\right)$$

$$X_2 = 1.04\,(6.3) + 70$$

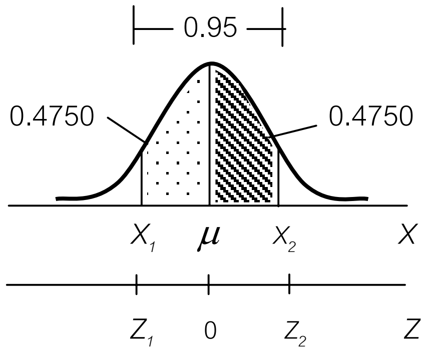
$$= 6.55 + 70$$

$$= 76.55$$

นั่นคือค่าน้ำหนักของนักศึกษาเกณฑ์ต่ำ ประมาณ 62.63 กิโลกรัม และค่าน้ำหนักต่ำสุดของของนักศึกษาเกณฑ์สูงประมาณ 76. 55 กิโลกรัม

ตัวอย่างที่ 4.9 ในการแจกแจงปรกติ ค่าเฉลี่ยเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ จงหาว่าร้อยละ 95 ของประชากรมีค่าอยู่ในช่วงใด

วิธีทำ



ภาพที่ 4.15 พื้นที่ใต้เส้นโค้งปรกติมาตรฐานของร้อยละ 95 ของประชากร

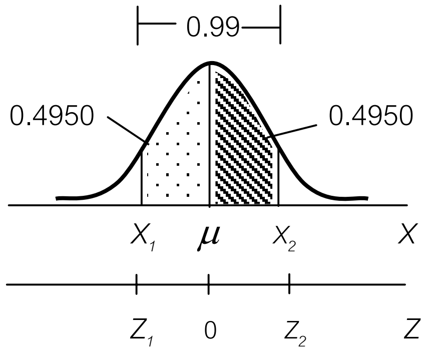
จากภาพที่ 4.15 จำนวนร้อยละ 95 ของประชากรมีค่าอยู่ระหว่าง X_1 กับ X_2 หรือ Z_1 กับ Z_2 เห็นได้ว่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ระหว่าง 0 กับ Z_1 มีพื้นที่เท่ากับ 0.4750 หน่วย ระหว่าง 0 กับ Z_2 มีพื้นที่เท่ากับ 0.4750 หน่วย อาศัย ตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 3 หาค่า Z_1 และ Z_2 พบว่า Z_1 เท่ากับ -1.96 และ Z_2 เท่ากับ 1.96

นั่นคือ ในการแจกแจงปรกติร้อยละ 95 ของประชากร (หรือร้อยละ 95 ของค่าสังเกตทั้งหมด) จะมีค่าระหว่าง Z_1 เท่ากับ -1.96 กับ Z_2 เท่ากับ 1.96 และในการแจกแจงปรกติ ค่าเฉลี่ยเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ จะได้ว่าร้อยละ 95 ของประชากรมีค่าระหว่าง X_1 เท่ากับ μ -1.96 σ กับ X_2 เท่ากับ μ + 1.96 σ

($\mu - 1.96\sigma < X < \mu + 1.96\sigma$)

ตัวอย่างที่ 4.10 ในการแจกแจงปรกติจากตัวอย่างที่ 4.9 จงหาว่าร้อยละ 99 ของประชากรมีค่าอยู่ในช่วงใด

วิธีทำ



ภาพที่ 4.16 พื้นที่ใต้เส้นโค้งปรกติมาตรฐานของร้อยละ 99 ของประชากร

จากภาพที่ 4.16 จำนวนร้อยละ 99 ของประชากรมีค่าอยู่ระหว่าง X_1 กับ X_2 หรือ Z_1 กับ Z_2 เห็นได้ว่า พื้นที่ใต้โค้งปรกติมาตรฐาน ระหว่าง 0 กับ Z_1 มีพื้นที่เท่ากับ 0.4950 หน่วยระหว่าง 0 กับ Z_2 มีพื้นที่เท่ากับ 0.4950 หน่วย จากตารางภาคผนวกที่ 5 เป็นตารางแสดงพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน เมื่อต้องการหาค่า Z_1 และ Z_2 พบว่าพื้นที่ใกล้เคียงกับพื้นที่ 0.4950 คือพื้นที่ 0.4949 หรือ พื้นที่ 0.4951 ได้ค่า Z เท่ากับ 2.57 และ 2.58 ตามลำดับ เราอาจเลือกค่า Z ค่าหนึ่งค่าใดในสองค่านี้หรืออาจใช้ค่าเฉลี่ย Z เท่ากับ $\frac{(2.57+2.58)}{2}$ เท่ากับ 2.575

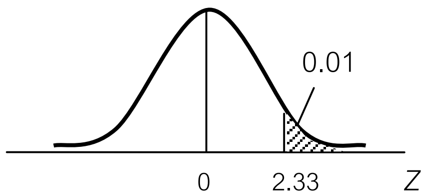
นั่นคือ ในการแจกแจงปรกติร้อยละ 99 ของประชากร (หรือร้อยละ 99 ของค่าสังเกตทั้งหมด) จะมีค่าระหว่าง Z_1 เท่ากับ -2.575 กับ Z_2 เท่ากับ 2.575 และในการแจกแจงปรกติค่าเฉลี่ยเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ จะได้ว่าร้อยละ 99 ของประชากรมีค่าระหว่าง X_1 เท่ากับ (μ - 2.575 σ) กับ X_2 เท่ากับ (μ +2.575 σ)

การแจกแจงความถี่จะเรียงลำดับประชากร (หรือค่าสังเกต) จากค่าน้อยไปหาค่ามากทางซ้ายไปทางขวาเสมอ จากตัวอย่างที่ 4.9 การกล่าวถึงร้อยละ 95 ของจำนวนประชากรโดยไม่ได้ระบุว่าเป็นร้อยละ 95 ทางค่าน้อยหรือค่ามากจะถือว่าประชากรส่วนใหญ่มีค่าอยู่ใกล้ ๆ กับศูนย์กลางการแจกแจงและประชากรจำนวนน้อยๆ อยู่ทางปลายหางซ้ายและขวาของการแจกแจง จึงแบ่งพื้นที่ใต้เส้นโค้งปรกติทางปลายหางทั้งสองข้างออกข้างละเท่า ๆ กัน คือร้อยละ $\frac{5}{2} = 2.5$ หรือพื้นที่ $\frac{0.05}{2} = 0.025$ หน่วย ทำให้พื้นที่ส่วนกลางเป็น .95 หน่วย หรือ ร้อยละ 95 ตัวอย่างที่ 5.10 ก็ทำนองเดียวกัน จำนวนร้อยละ 99 ของประชากรจะแบ่งพื้นที่ใต้เส้นโค้งปรกติทางปลายหางทั้งสองออกข้างละร้อยละ $\frac{1}{2} = 0.5$ หรือพื้นที่ $\frac{0.01}{2} = 0.005$ หน่วย ทำให้พื้นที่ส่วนกลางเป็น 0.99 หน่วยหรือร้อยละ 99

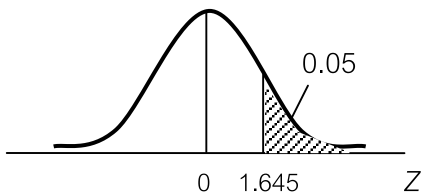
หากระบุให้แบ่งส่วนปลายหางของการแจกแจงทางซ้าย ทางขวาหรือทั้งทางซ้ายและทางขวาเท่า ๆ กันเราจะสามารถใช้ตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 หาค่ามาตรฐาน Z ตรงจุดแบ่งพื้นที่ออกไปได้ ซึ่งการหาจุดแบ่งนี้จะเป็นพื้นฐานสำหรับการหาค่าวิกฤต (critical value) ของการแจกแจงปรกติและการแจกแจงแบบ อื่น ๆ

อย่างไรก็ดี ขอให้ระลึกอยู่เสมอว่า จำนวนประชากรหรือค่าสังเกตที่นำมาแจกแจงความถี่จะต้องมีจำนวนมากพอ ที่จะแสดงให้เห็นรูปร่างการแจกแจงปรกติของประชากรหรือค่าสังเกตเหล่านั้นด้วย สุดท้ายนี้ขอให้พิจารณาภาพที่ 4.17 และ 4.18 ในสถานการณ์ที่ระบุให้แบ่งส่วนปลายหางของการแจกแจงปรกติทางขวาร้อยละ 1 (พื้นที่ปลายหางทางขวามี

0.01 หน่วย) ตรงกับค่ามาตรฐาน *Z* ประมาณ 2.33 และในสถานการณ์ที่ระบุให้แบ่งส่วนหางของการแจงแจกปรกติทางขวาร้อยละ 5 (พื้นที่ปลายหางทางขวามี 0.05 หน่วย) ตรงกับค่ามาตรฐาน *Z* ประมาณ 1.645 ไว้สำหรับเปรียบเทียบกับค่าวิกฤตของการแจกแจงแบบอื่น ๆ ในกระบวนการทดสอบสมมุติฐานหรือกระบวนการอื่นใดของสาขาสถิติอนุมานต่อไป



ภาพที่ 4.17 พื้นที่ปลายหางทางขวาของการแจกแจงปรกติ
ร้อยละ 1 ของทั้งหมด



ภาพที่ 4.18 พื้นที่ปลายหางทางขวาของการแจกแจงปรกติ
ร้อยละ 5 ของทั้งหมด

4.5 บทสรุป

บทนี้เป็นการกล่าวถึงการแจกแจงปรกติของตัวแปร *X* โดยเฉพาะและจะเป็นพื้นฐานสำหรับการศึกษาการแจกแจงสิ่งตัวอย่างในสถิติอนุมานต่อไป สรุปได้ดังนี้

4.5.1 เส้นโค้งปรกติ มีลักษณะเป็นรูประฆังคว่ำ และมีตำแหน่งหนึ่งที่อยู่กึ่งกลางของรูประฆังคว่ำนี้และเรียกว่า ศูนย์กลางความสมมาตร

4.5.2 พื้นที่ใต้เส้นโค้งปรกติ และพื้นที่ใต้เส้นโค้งปรกติมาตรฐานจะเป็นการบอกค่าความน่าจะเป็นของข้อมูลที่ต้องการที่จะศึกษา

ในการที่จะศึกษาพื้นที่ใต้เส้นโค้งปรกตินั้นหลักการที่สำคัญที่สุดคือการอ่านพื้นที่ใต้เส้นโค้งปรกติมาตรฐานดังตารางภาคผนวกที่ 5 โดยจะต้องการหาค่า *Z* เมื่อกำหนดพื้นที่ใต้เส้นโค้งปรกติไว้ก่อน

4.6 คำถามทบทวน

1. ประชากรการแจกแจงปรกติ จงหาว่าร้อยละ 95.44 ของประชากร มีค่าอยู่ช่วงใด
2. ประชากรการแจกแจงปรกติ จงหาว่าร้อยละ 98 ของประชากร มีค่าอยู่ช่วงใด
3. ชั่งน้ำหนักพนักงานจำนวน 1,000 คนพบว่าการแจกแจงของน้ำหนักเป็นโค้งปรกติ ถ้าคัดเฉพาะพนักงานที่น้ำหนักมาก ๆ มาแจกแจงจำนวน 100 คนการแจกแจงของน้ำหนักจะเป็นโค้งปรกติหรือไม่ เพราะเหตุใด
4. จงหาพื้นที่ใต้เส้นโค้งปรกติ ในกรณีต่าง ๆ ดังต่อไปนี้
   - ก. ระหว่างค่าเฉลี่ยกับ Z =1.50
   - ข. ระหว่างค่าเฉลี่ยกับ Z = -2.56
   - ค. เมื่อ Z >1.25
   - ง. เมื่อ Z <2.50
   - จ. เมื่อ $Z = \pm 1.96$
   - ฉ. ระหว่าง Z = 1.25 กับ Z = 3.00
   - ช. เมื่อ Z > -1.64
   - ซ. ระหว่าง Z = -1.65 กับ Z = 1.75
5. รายได้ต่อวันของพนักงานจำนวน 2,000 คน แจกแจงปรกติรายได้เฉลี่ยวันละ 155 บาท ค่าเบี่ยงเบนมาตราฐาน 20 บาท จงหาจำนวนพนักงานที่มีรายได้กรณีต่อไปนี้
   - ก. รายได้ตั้งแต่ไม่เกินวันละ 100 บาท
   - ข. รายได้ระหว่างวันละ 120 บาทถึง 130 บาท
   - ค. รายได้ระหว่างวันละ 150 บาท.ถึง 175 บาท
   - ง. รายได้ตั้งแต่วันละ 200 บาทขึ้นไป
6. ไอคิวของผู้สมัครสอบคัดเลือกจำนวน 600 คนมีการแจกแจงปรกติเฉลี่ยเท่ากับ 115 ค่าเบี่ยงเบนมาตรฐาน 12 ถ้าต้องการรับผู้มีไอคิวอย่างน้อยร้อยละ 95 อยากทราบว่า จำนวนผู้สมัครสอบคัดเลือกที่มีคุณสมบัติไม่ได้ตามต้องการกี่คน
7. คะแนนสอบไล่วิชาสถิติของนักศึกษาทั้งหมดมีการแจกจงปรกติเฉลี่ย 76 คะแนน ค่าเบี่ยงเบนมาตราฐาน 15 คะแนน ถ้ากำหนดให้ร้อยละ 15 ของจำนวนนักศึกษาทั้งหมดได้เกรด A และร้อยละ 10 ของจำนวนนักศึกษาทั้งหมดได้เกรด F จงหา
   - ก. คะแนนต่ำสุดของเกรด A
   - ข. คะแนนสูงสุดของเกรด F (สอบตก)

8. อายุการใช้งานของรถยนต์ที่ผลิตจากบริษัทแห่งหนึ่งมีการแจกแจงปรกติ เฉลี่ย 10 ปี และค่าเบี่ยงเบนมาตรฐาน 2 ปี ถ้าบริษัทนี้ต้องการซ่อมให้ฟรีเพียงร้อยละ 5 ของจำนวนรถที่ผลิตออกจำหน่ายทั้งหมด ควรจะรับประกันการซ่อมฟรีไว้นานกี่ปี

บทที่ 5
ทฤษฎีและการแจกแจงตัวแปร

เราทราบว่าตัวอย่างสุ่มเป็นส่วนหนึ่งของประชากรที่ต้องการทราบคุณลักษณะและได้มาโดยอาศัยกระบวนการเลือกสุ่ม (random selection) จึงจำเป็นต้องกล่าวถึงโอกาสที่พึงเป็นไปได้ของตัวอย่าง โอกาสพบตัวอย่างหรือความน่าจะเป็น (probability) ของตัวอย่าง ทฤษฎีความน่าจะเป็นคือการกล่าวถึงโอกาสที่พึงเป็นไปได้ ด้วยการแสดงค่าออกมาเป็นตัวเลขอันเป็นตัวบ่งบอกให้ทราบว่า เหตุการณ์ที่เรากำลังสนใจอยู่มีโอกาสพบหรือมีโอกาสเกิดขึ้นมากน้อยเพียงใด ถ้าความน่าจะเป็นของเหตุการณ์ใดมีค่าสูงแสดงว่าเหตุการณ์นั้นมีโอกาสพบบ่อยครั้งหรือมีโอกาสเกิดขึ้นได้มาก ตรงกันข้ามถ้าหากความน่าจะเป็นของเหตุการณ์ใดมีค่าต่ำแสดงว่าเหตุการณ์นั้นมีโอกาสพบน้อยครั้งหรือมีโอกาสเกิดขึ้นได้น้อย ทั้งนี้เพื่อประโยชน์ในการตัดสินใจ

5.1 ทฤษฎีความน่าจะเป็น

การแสดงค่าความน่าจะเป็นด้วยตัวเลข แสดงการเปรียบเทียบจำนวนหนทางที่เกิดเหตุการณ์ที่เรากำลังสนใจอยู่กับจำนวนหนทางที่จะเกิดเหตุการณ์ที่พึงเป็นไปได้ทั้งหมด

ถ้ากำหนด N : เป็นจำนวนหนทางที่จะเกิดเหตุการณ์ที่พึงเป็นไปได้ทั้งหมด

a : เป็นจำนวนหนทางที่เกิดเหตุการณ์ที่เรากำลังสนใจอยู่

A : เป็นเหตุการณ์ที่เรากำลังสนใจอยู่

และ $P(A)$: เป็นความน่าจะเป็นเหตุการณ์ A จะได้สมการ (5-1)

$$P(A) = \frac{a}{N} \qquad \text{…(5-1)}$$

จำนวนหนทางที่จะเกิดเหตุการณ์ที่พึงเป็นไปได้ทั้งหมด (N) นิยมแสดงไว้ในรูปของเซต (set) เรียกว่า เซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด (all possible event set or all possible outcomes set or sample space) เขียนแทนเซตนี้ด้วย S เช่น ผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดของสินค้าชิ้นที่วางขายเป็น S คือ {ขายได้, ขายไม่ได้} ผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดของผลการเรียนนักศึกษาผู้หนึ่งเป็น S คือ {A, B, C, D, E} ผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดของการออกรางวัลเลขท้ายสองท้ายสองตัวสำหรับหนึ่งปีน S คือ {00, 01, 02, 03, ... , 99} เป็นต้น

เหตุการณ์ที่เรากำลังสนใจอยู่ (A) จะเป็นเซตย่อย (subset) ของเซต S นี้ซึ่งอาจประกอบด้วยสมาชิก (element) เพียงตัวเดียวหรือหลายตัวของเซต S ก็ได้ เบื้องต้นนี้เราจะกล่าวถึงกรณีเหตุการณ์ที่เรากำลังสนใจอยู่เป็นเซตย่อยของเซต S และประกอบด้วยสมาชิกเพียงหนึ่งของเซต S เราเรียกเหตุการณ์เช่นนี้ว่า เหตุการณ์เชิงเดี่ยว (simple event) เช่น

กำหนดให้ $S = \{1, 2, 3, 4, 5, 6\}$

A : เป็นเหตุการณ์ที่เรากำลังสนใจอยู่ คือ เลข 2

จะได้ $A = \{2\}$ (เป็นเหตุการณ์เชิงเดี่ยว)

$P(A)$: เป็นความน่าจะเป็นเหตุการณ์ A ตามนิยามสมการ (5-1)

$N = 6$ และ $a = 1$

ดังนั้น $$P(A) = \frac{1}{6} = 0.166$$

อ่านว่า ความน่าจะเป็นเหตุการณ์ A เท่ากับ $\frac{1}{6}$ หรือ 0.166

ถ้ากำหนดให้ $A_1, A_2, \ldots\ldots\ldots, A_6$ เป็นเหตุการณ์เชิงเดี่ยวของเซต และ S เป็นเหตุการณ์จนครบในผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด กล่าวคือ A_1 คือ {1}, A_2 คือ {2} , ..…………, A_6 คือ {6} และ $P(A_1)$ เท่ากับ $\frac{1}{6}$, $P(A_2)$ เท่ากับ $\frac{1}{6}$, ... , $P(A_6)$ เท่ากับ $\frac{1}{6}$ จะได้เป็นสมการ

$$P(A_1) + P(A_2) + \ldots\ldots\ldots\ldots\ldots + P(A_6) = 1$$

ในกรณีของเหตุการณ์เชิงเดี่ยว N เหตุการณ์ และเมื่อกำหนดให้ $A_1, A_2, \ldots\ldots\ldots\ldots\ldots, A_N$ เป็นเหตุการณ์เชิงเดี่ยวทุกเหตุการณ์ของผลลัพธ์ที่พึงจะเป็นไปได้ทั้งหมดและแต่ละเหตุการณ์มีความน่าจะเป็น $P(A_1)$, $P(A_2)$, ...,$P(A_N)$ จะได้สมการ (5-2)

$$P(A_1) + P(A_2) + \ldots\ldots\ldots\ldots\ldots + P(A_N) = 1 \qquad \ldots(5\text{-}2)$$

ตัวอย่างที่ 5.1 ซื้อสลากกินแบ่งรัฐบาลมาหนึ่งฉบับมีเลขท้ายเป็น 65 จงหาโอกาสที่จะถูกรางวัลเลขท้ายสองตัว

วิธีทำ $S = \{00, 01, 02, \ldots\ldots, 64, 65, 66, \ldots, 99\}$

A : เป็นเหตุการณ์ที่เรากำลังสนใจอยู่

$A = \{65\}$: เป็นเหตุการณ์เชิงเดี่ยว

$P(A)$: เป็นโอกาสถูกรางวัลหรือความน่าจะเป็นของเหตุการณ์

จากสมการ (6-1) $P(A) = \dfrac{a}{N}$; N คือ 100 , a คือ 1

ดังนั้น $P(A) = \dfrac{1}{100}$

นั่นคือ โอกาสที่จะถูกรางวัลเลขท้ายสองตัว (65) เท่ากับ $\dfrac{1}{100}$ หรือ 1 ใน 100

เราอาจกล่าวได้ว่าความน่าจะเป็นหรือโอกาสคือความถี่สัมพัทธ์ของเหตุการณ์ที่กำลังสนใจอยู่ในจำนวนเหตุการณ์ที่พึงเป็นไปได้ทั้งหมด ทำให้ความน่าจะเป็นมีลักษณะเฉพาะ ดังนี้ คือ

5.1.1 ความน่าจะเป็นของเหตุการณ์ใด ๆ มีค่าอยู่ระหว่าง 0 ถึง 1

5.1.2 ผลรวมความน่าจะเป็นของทุกเหตุการณ์ในจำนวนเหตุการณ์ที่พึงเป็นไปได้ทั้งหมด เท่ากับ 1

5.1.3 ความน่าจะเป็นของเหตุการณ์ใดๆจะมีความหมายเป็น ความน่าจะเป็น หรือ โอกาส ก็ต่อเมื่อ เหตุการณ์นั้นยังไม่เกิด หรือเหตุการณ์นั้นเกิดขึ้นแล้วแต่ไม่ทราบผลแน่ชัด

5.2 การหาผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด

จำนวนผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดหรือจำนวนหนทางที่พึงเป็นไปได้ทั้งหมด อาจหาได้หลายวิธีทั้งนี้เพื่อประโยชน์สำหรับการหาความน่าจะเป็น ที่แม่นยำและถูกต้องเท่าที่พบมานิยมหากันอยู่ 2 วิธี คือ

5.2.1 การทดลองหรือการสำรวจ เป็นวิธีการหาผลลัพธ์ที่เป็นไปได้ทั้งหมด ที่นิยมใช้ในทางปฏิบัติมากโดยเฉพาะงานวิจัยในสังคมและธุรกิจ เช่น ต้องการทราบโอกาสจำหน่ายสินค้า 60 ชิ้น จะได้ S คือ { ขายได้, ขายไม่ได้, ขายได้, ขายไม่ได้ .... } เซตนี้มีจำนวนสมาชิกอยู่ 60 รายการ ตามนิยามความน่าจะเป็นของสมการ (5 -1) เราได้ N คือ 60 สมมุติว่าสินค้าขายได้ 12 ชิ้น ถ้า A เป็นเหตุการณ์ที่เรากำลังสนใจอยู่คือสินค้าขายได้ 12 ชิ้นและได้ a คือ 12 ดังนั้นจากสมการ (5 –1) โอกาสที่จะสินค้าขายได้ 12 ชิ้นคือ $P(A)$ เท่ากับ $\frac{12}{60}$ เท่ากับ 0.2 กรณีต้องการทราบโอกาสสินค้าหายเสียระหว่างการขนส่งโดยทดลองขนส่งสินค้า 5,000 ชิ้น ปรากฏว่าสินค้าเสียหายระหว่างการขนส่ง 1,000 ชิ้น $P(A)$ เท่ากับ $\frac{1,000}{5,000}$ เท่ากับ 0.2 เป็นต้น

5.2.2 การใช้หลักเกณฑ์ทางทฤษฎี วิธีนี้ใช้ตัวแบบทางคณิตศาสตร์ ซึ่งสามารถแสดงผลลัพธ์ที่เป็นไปได้ทั้งหมดอย่างครอบคลุมและแน่นนอนกว่าวิธีแรก ทำให้ความน่าจะเป็นที่คำนวณได้แสดงออกมาเป็นตัวเลขแน่นอนและชัดเจนยิ่งขึ้น เช่น ต้องการทราบว่า เมื่อซื้อสลากกินแบ่งรัฐบาลมาหนึ่งฉบับมีโอกาสถูกรางวัลเลขท้ายสองตัวเท่าใด สามารถแสดงผลลัพธ์ที่เป็นไปได้ทั้งหมดคือ S คือ {00, 01, 02, 03, ... , 99} ซึ่ง N มีค่าเท่ากับ 100 ดังนั้นโอกาสถูกรางวัลเลขท้ายสองตัว เท่ากับ $\frac{1}{100}$ เท่ากับร้อยละ 0.01 เป็นต้น

5.3 ตัวแบบทางคณิตศาสตร์สำหรับการหาผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด

ตัวแบบหลักสำหรับการผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด 3 ลักษณะ คือแผนภูมิต้นไม้ (tree diagram) วิธีเรียงสับเปลี่ยน (permutation) และวิธีจัดหมู่ (combination) ดังจะกล่าวถึงต่อไปนี้

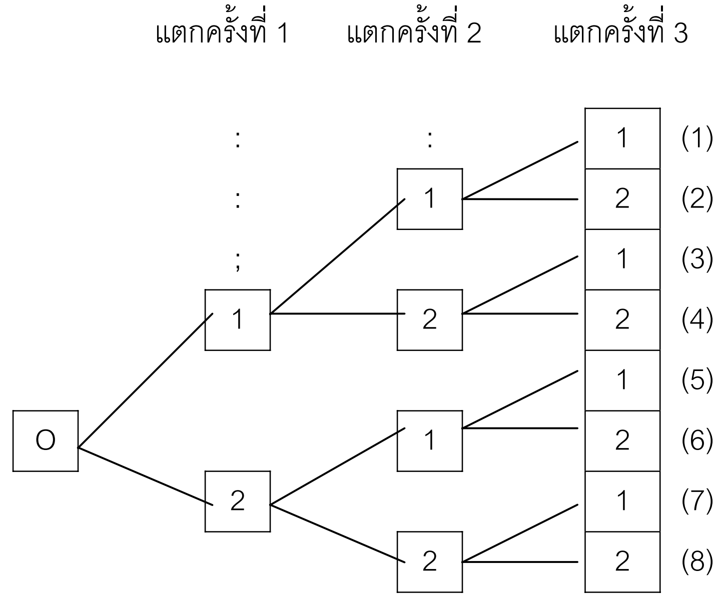
5.3.1 แผนภูมิต้นไม้ เป็นการหาผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดโดยอาศัยหลักจากธรรมชาติการแตกกิ่งของต้นไม้ เช่น มีกิ่งไม้อยู่ 1 กิ่ง แล้วแตกกิ่งออกจะได้เป็น 2 กิ่งทุกครั้งที่มีการแตกกิ่ง (Keller & Warrack,2000, pp.169-170) พิจารณาภาพที่ 5.1 จะเห็นว่าการแตกกิ่งครั้งที่หนึ่ง จะมีจำนวนกิ่งไม้ทั้งหมด เป็น 2 กิ่ง การแตกกิ่งครั้งที่สอง จะมีจำนวนกิ่งไม้ทั้งหมด 4 กิ่ง และการแตกกิ่งครั้งที่สาม จะมีจำนวนกิ่งไม้ทั้งหมด 8 กิ่ง มีข้อที่ควรสนใจ คือ 2 มาจาก 2^1 4 มาจาก 2^2 และ 8 มาจาก 2^3 ถ้ามีการแตกกิ่งครั้งที่สี่ จะมีจำนวนกิ่งทั้งหมด

2^4 เท่ากับ 16 กิ่ง และการแตกกิ่งครั้งที่ห้า จะมีจำนวนกิ่งทั้งหมด 2^5 เท่ากับ 32 กิ่ง เป็นต้น

ถ้ากำหนดให้มีการแตกกิ่งจำนวน n ครั้ง ๆ ละ 2 กิ่งจะได้สมการ (5-3)

$$\text{จำนวนกิ่งของการแตกกิ่งครั้งที่ } n = 2^n \text{ กิ่ง} \qquad \ldots(5\text{-}3)$$

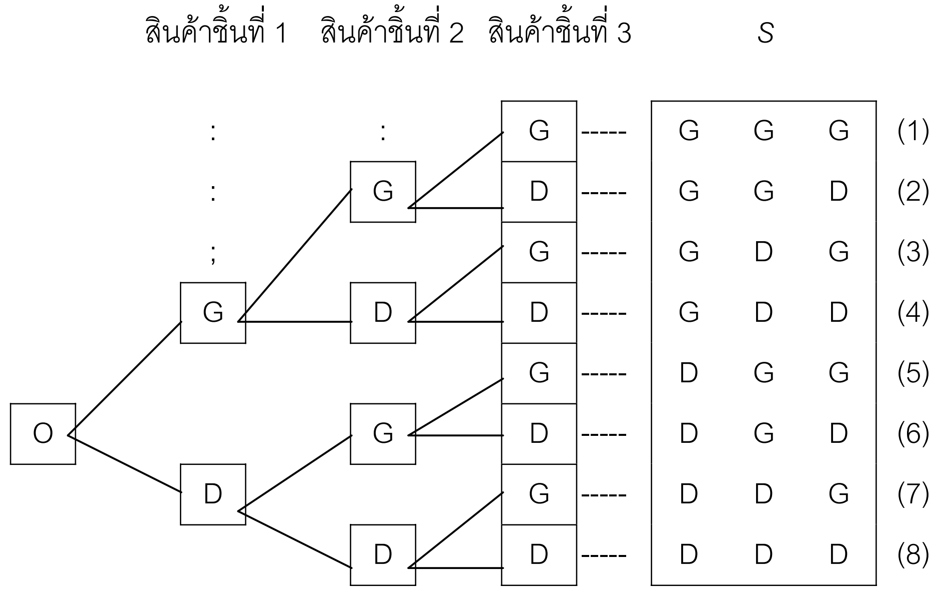
จำนวนกิ่งของการแตกกิ่งครั้งที่ n ตามสมการ (5-3) นี้ก็คือ จำนวนหนทางที่พึงเป็นไปได้ทั้งหมดหรือผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดของเหตุการณ์ที่เรากำลังสนใจอยู่ 2 เหตุการณ์ และปล่อยให้เกิดอย่างอิสระ (มีโอกาสเกิดขึ้นเท่า ๆ กันใน 2 เหตุการณ์นี้) จำนวน n ครั้งนั่นเอง



ภาพที่ 5.1 แผนภูมิต้นไม้จำนวนกิ่งไม้ทั้งหมดเมื่อแตกกิ่ง 3 ครั้ง ๆ ละ 2 กิ่ง

ตัวอย่างที่ 5.2 จงหาผลลัพธ์หรือจำนวนหนทางจะพึงเป็นไปได้ทั้งหมด เมื่อวางสินค้า 3 ชิ้นเพื่อขาย

วิธีทำ ผลลัพธ์ที่พึงเป็นไปได้ของสินค้าแต่ละชิ้นมี 2 ทาง คือขายได้ (G) และขายไม่ได้ (D) สามารถเขียนแผนภูมิต้นไม้แสดงผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดดังภาพที่ 5.2



ภาพที่ 5.2 แผนภูมิต้นไม้จำนวนหนทางที่พึงเป็นไปได้ทั้งหมด เมื่อวางสินค้า 3 ชิ้น

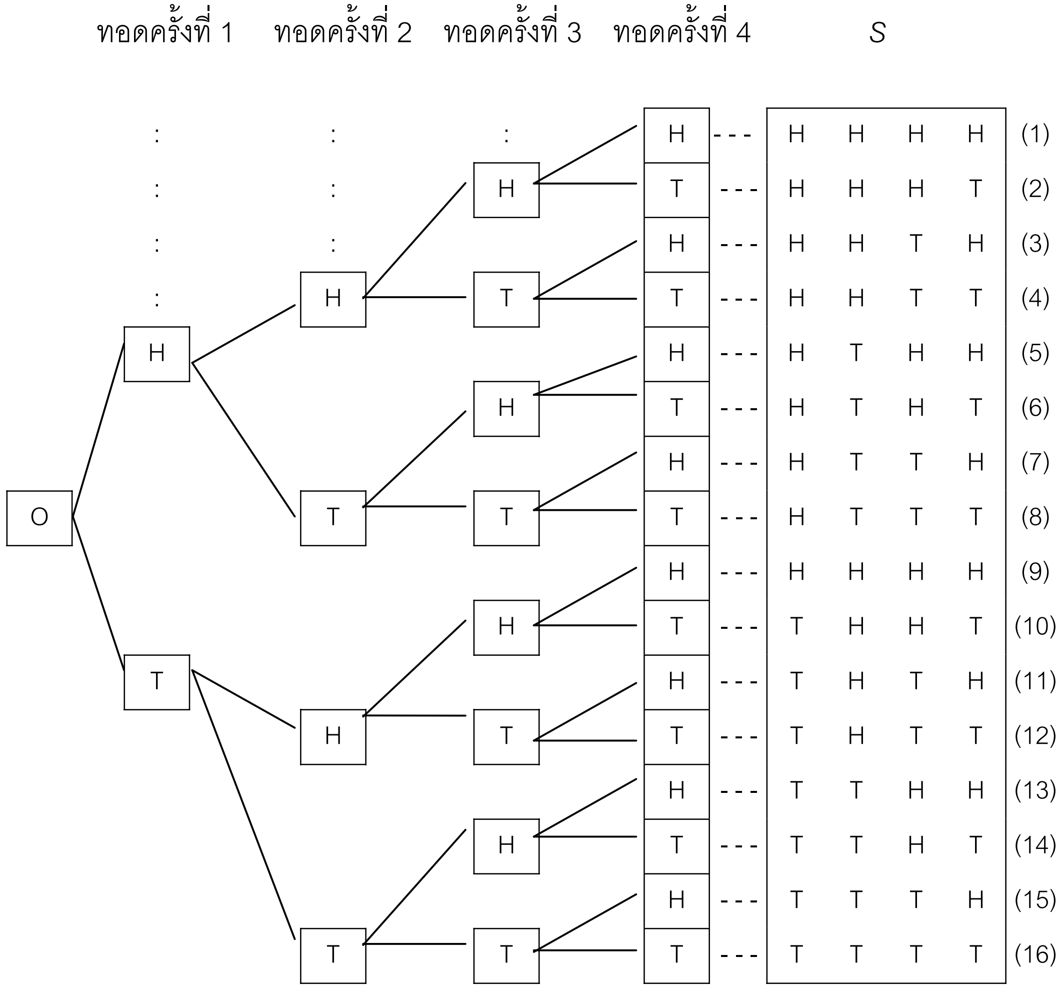
ดังนั้นผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด หรือจำนวนหนทางที่พึงเป็นไปได้ทั้งหมด เมื่อขายสินค้า 3 ชิ้นในท้องตลาดมี 8 หนทาง

จำนวนหนทางที่พึงเป็นไปได้ทั้ง 8 หนทางตามตัวอย่างที่ 5.2 มาจาก 2^n และ n เท่ากับ 3 จึงได้หนทางที่พึงเป็นไปได้ทั้งหมดเท่ากับ 2^3 เท่ากับ 8 หนทางที่ต่างกัน เช่นจากภาพที่ 5.2 หนทางที่ 1 “G G G“ หมายความว่า สินค้าชิ้นที่ 1 ขายได้ สินค้าชิ้นที่ 2 ขายได้ และสินค้าชิ้นที่ 3 ขายได้ และหนทางที่ 4 “G D D” หมายความว่า สินค้าชิ้นที่ 1 ขายได้ สินค้าชิ้นที่ 2 ขายไม่ได้ และสินค้าชิ้นที่ 3 ขายไม่ได้ เป็นต้น แสดงว่า หากขายสินค้าจำนวน 4 ชิ้น จะได้จำนวนหนทางทั้งหมดที่พึงเป็นไปได้เท่ากับ 2^4 เท่ากับ 16 หนทาง หากขายสินค้าจำนวน 5 ชิ้น จะได้จำนวนหนทางทั้งหมดที่พึงเป็นไปได้เท่ากับ 2^5 เท่ากับ 32 หนทางหรือสินค้าจำนวนมากขึ้นก็จะพบจำนวนหนทางที่พึงเป็นไปได้มากขึ้นเป็น $2^{\text{(จำนวนสินค้า)}}$ (สองยกกำลังจำนวนสินค้า) ทั้งนี้เลข 2 เป็นจำนวนหนทางที่พึงเป็นไปได้ 2 หนทางคือ ขายได้ (G) และขายไม่ได้ (D)

นอกจากนี้หลักการแตกกิ่งไม้หรือแผนภูมิกิ่งไม้ครั้งละ 2 กิ่งที่กล่าวมายังสามารถประยุกต์ใช้กับสถานการณ์ การกระทำ การเลือกหรืออื่น ๆ ที่การเกิดสถานการณ์แต่ละครั้งมีทางเกิดได้ 2 อย่าง การกระทำแต่ละครั้งกระทำได้ 2 วิธี การเลือกแต่ละครั้งมีทางเลือกได้ 2 ทาง หรืออื่น ๆ ในทำนองเดียวกัน

ตัวอย่างที่ 5.3 จงแสดงผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด เมื่อทอดเหรียญบาทจำนวน 4 ครั้ง

วิธีทำ ผลลัพธ์ที่พึงเป็นไปได้เมื่อทอดเหรียญบาท แต่ละครั้งจะได้ หัว (H) หรือ ก้อย (T) ซึ่งสามารถเขียนแผนภูมิต้นไม้แสดงผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดดังภาพที่ 5.3



ภาพที่ 5.3 การออกหัวและก้อยของการทอดเหรียญบาทจำนวน 4 ครั้ง

ข้อสังเกต ในการใช้แผนภูมิต้นไม้หาผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดจากตัวอย่างที่ 5.3 คือ ผลลัพธ์ที่พึงเป็นไปได้ทั้ง 16 หนทางนี้ เป็นจำนวนหนทางที่เป็นไปได้ ถ้าหากทอดเหรียญ 4 ครั้ง และหนทางเหล่านี้แตกต่างกันทั้งหมด 16 หนทาง ในสถานการณ์จริงการทอดเหรียญ 4 ครั้งนี้สามารถเกิดเหตุการณ์ได้เพียง 1 ใน 16 หนทางนี้ (เช่น เป็นหัวทั้งหมด 4 ครั้ง ตามหนทางที่ 1 หรือครั้งที่ 1 เป็นหัว ครั้งที่ 2 เป็นหัว ครั้งที่ 3 เป็นก้อย ครั้งที่ 4 เป็นหัว ตาม

หนทางที่ 3 หรืออื่น ๆ เป็นต้น) เมื่อทอดครั้งที่ 4 สามารถทอดครั้งที่ 5 ครั้งที่ 6 .... จนไม่มีที่สิ้นสุด แต่ผลได้แต่ละครั้งยังเป็นหัวหรือก้อยเท่านั้น การถามถึงผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดของการทอดเหรียญเพียง 4 ครั้ง จึงเป็นการยกตัวอย่างว่า (ถ้าผลได้เป็น 2 หนทาง คือ หัวและก้อย) กระทำ 4 ครั้งจะได้ผลอย่างไรบ้าง เราก็สามารถแสดงผลได้ทั้งหมดที่อาจเป็นไปได้อย่างชัดเจนและครบถ้วน หากเราย้อนกลับไปพิจารณาตัวอย่างที่ 5.2 สินค้าวางขาย 3 ชิ้น ก็เป็นการยกตัวอย่างของสินค้าจำนวนนับไม่ถ้วน [ถ้าผลได้เป็น 2 ทาง คือขายได้ (G) และขายไม่ได้ (D)] หากมีสินค้า 3 ชิ้นจะได้ผลอย่างไร เราสามารถใช้แผนภูมิต้นไม้แสดงผลได้ ทั้งหมดที่อาจเป็นไปได้อย่างชัดเจนและครบถ้วนเช่นกันจากข้อสังเกตดังกล่าว สามารถนำไปประยุกต์ใช้กับการชักตัวอย่างสุ่มขนาด n สำหรับการสุ่มตัวอย่างจะกล่าวโดยละเอียดในบทที่ 6 ตัวอย่างขนาด n และประชากรขนาดอนันต์ (infinite population) ต่อไป

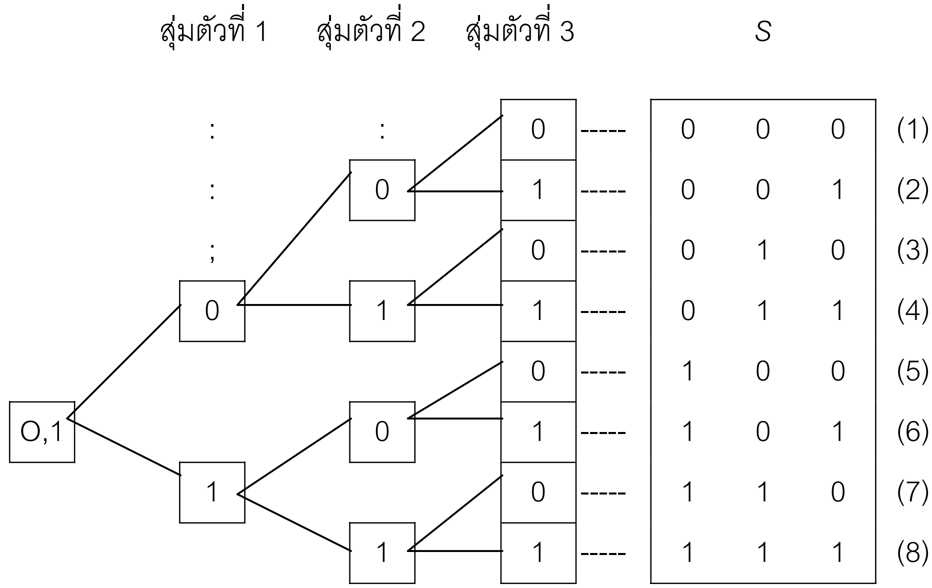
จากประชากรขนาดใหญ่ M หรือประชากรขนาดอนันต์ แต่คุณลักษณะที่เราสนใจมี 2 อย่าง เช่น รอด—ตาย ดี—เลว สอบได้—สอบตก ชอบ—ไม่ชอบ หัว—ก้อย บวก—ลบ หรืออื่น ๆ ในทำนองเดียวกันและนำไปประยุกต์ใช้กับการชักตัวอย่างสุ่มขนาด n จากประชากรขนาดอนันต์ ที่มีความแตกต่างหรือมีข้อแตกต่างกันภายในประชากรเพียง 2 ประชากร เช่น ลูกปัดกองใหญ่มีสีขาวกับสีแดง ฉลากจำนวนนับไม่ถ้วนมีเบอร์ 4 กับเบอร์ 7 เป็นต้น เซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด คือ เซตตัวอย่างขนาด n ที่พึงเป็นไปได้ทั้งหมด และจำนวนตัวอย่างขนาด n ที่พึงเป็นไปได้ทั้งหมด สามารถคำนวณได้เช่นเดียวกับการหาจำนวนกิ่งไม้ที่แตกกิ่งครั้งละ 2 กิ่ง จำนวน n ครั้ง ตามสมการ (5-3) และถ้าประชากรขนาดอนันต์มีคุณลักษณะที่เราสนใจหรือมีข้อแตกต่างกันภายใน 2 ประการ จะได้สมการ (6-4)

$$\text{จำนวนตัวอย่างขนาด } n \text{ ที่พึงเป็นไปได้ทั้งหมด} = 2^n \qquad \ldots(5\text{-}4)$$

ตัวอย่างที่ 5.4 ประชากรขนาดอนันต์ ประกอบด้วย 0 และ 1 จงแสดงจำนวนตัวอย่างสุ่มขนาด $n = 3$ ที่พึงเป็นไปได้ทั้งหมด

วิธีทำ ประชากรขนาดอนันต์ที่มีข้อแตกต่างกัน 2 ประการ คือ 0 และ 1 เมื่อต้องการตัวอย่างขนาด $n = 3$

ดังนั้น จำนวนตัวอย่างที่พึงเป็นไปได้ทั้งหมดเท่ากับ $2^3 = 8$ ตัวอย่าง ถ้ากำหนดให้ S เป็นเซตตัวอย่างขนาด $n = 3$ ที่พึงเป็นไปได้ทั้งหมด



ภาพที่ 5.4 แผนภูมิต้นไม้จำนวนตัวอย่างสุ่มขนาด $n = 3$ จากประชากรขนาดอนันต์

จำนวนตัวอย่างสุ่มขนาด $n = 3$ แสดงด้วยแผนภูมิต้นไม้ได้ดังภาพที่ 5.4

ในสถานการณ์ทั่วไป ประชากรขนาดอนันต์จะมีคุณลักษณะที่เราสนใจหรือข้อแตกต่างกันภายในประชากรไม่ได้มีเพียง 2 ประชากรเสมอไป ได้แก่ 3 ประชากร 4 ประชากร 5 ประชากร .... m ประชากร จึงเป็นการแตกกิ่งไม้ครั้งละ 3 กิ่ง 4 กิ่ง 5 กิ่ง ...... และ m กิ่ง ตามลำดับ เช่น

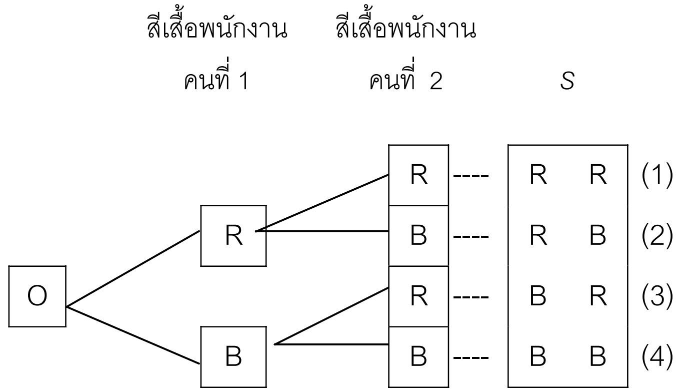
1) กรณีประชากรขนาดอนันต์มีคุณลักษณะที่เราสนใจ หรือข้อแตกต่างกันภายในประชากร 3 ประชากรตัวอย่างขนาด n เท่ากับ 1, 2, 3, ... , n จะมีจำนวนตัวอย่างที่พึงเป็นไปได้ทั้งหมดเท่ากับ $3^1, 3^2, 3^3, \ldots, 3^n$ ตามลำดับ

2) กรณีประชากรอนันต์มีคุณลักษณะที่เราสนใจ หรือข้อแตกต่างกันภายในประชากร 4 ประการ ตัวอย่างขนาด n เท่ากับ $4^1, 4^2, 4^3, \ldots, 4^n$ ตามลำดับ เป็นต้น จึงสรุปได้ว่า ถ้าประชากรขนาดอนันต์มีคุณลักษณะที่เราสนใจ หรือข้อแตกต่างกันภายในประชากร m ประการ จะได้สมการ (5-5)

$$\text{จำนวนตัวอย่างขนาด } n \text{ ที่พึงเป็นไปได้ทั้งหมด} = m^n \qquad \ldots(5\text{-}5)$$

ตัวอย่างที่ 5.5 จงแสดงวิธีจัดพนักงาน 2 คนให้ยืนประจำหน้าร้านเพื่อเชิญชวนลูกค้าให้เข้ามาในร้าน โดยพนักงานมีจำนวนมากแต่ใส่เสื้อสองสีเท่านั้นคือสีแดง (R) และสีน้ำเงิน (B)

วิธีทำ กำหนดให้จัดพนักงาน 2 คนยืนประจำหน้าร้าน จำนวนแบบที่พึงเป็นไปได้ทั้งหมดเท่ากับ $2^2 = 4$ แบบ กำหนด S เป็นเซตที่พึงเป็นไปได้ทั้งหมดของการจัดพนักงาน

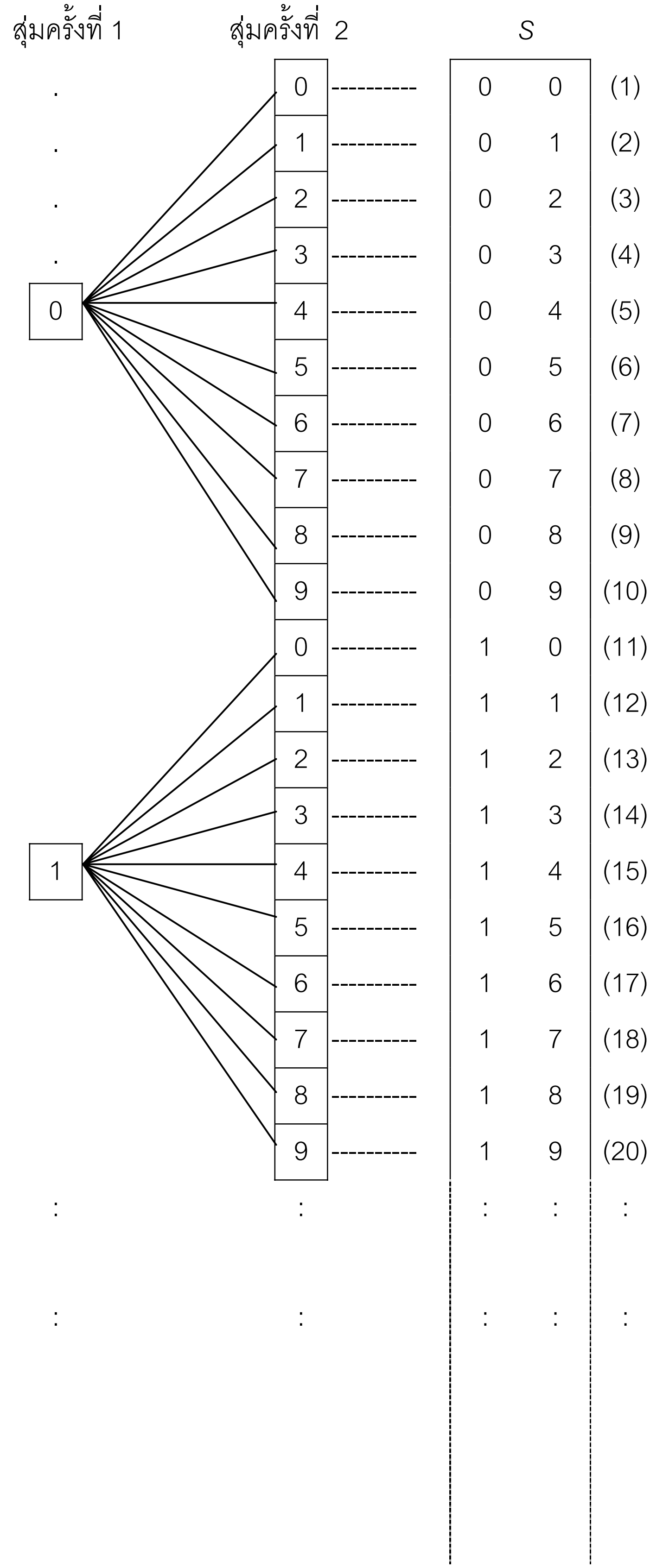


ภาพที่ 5.5 แผนภูมิต้นไม้ของการจัดพนักงานสองคนให้ยืนหน้าร้าน
ที่พึงเป็นไปได้ทั้งหมด

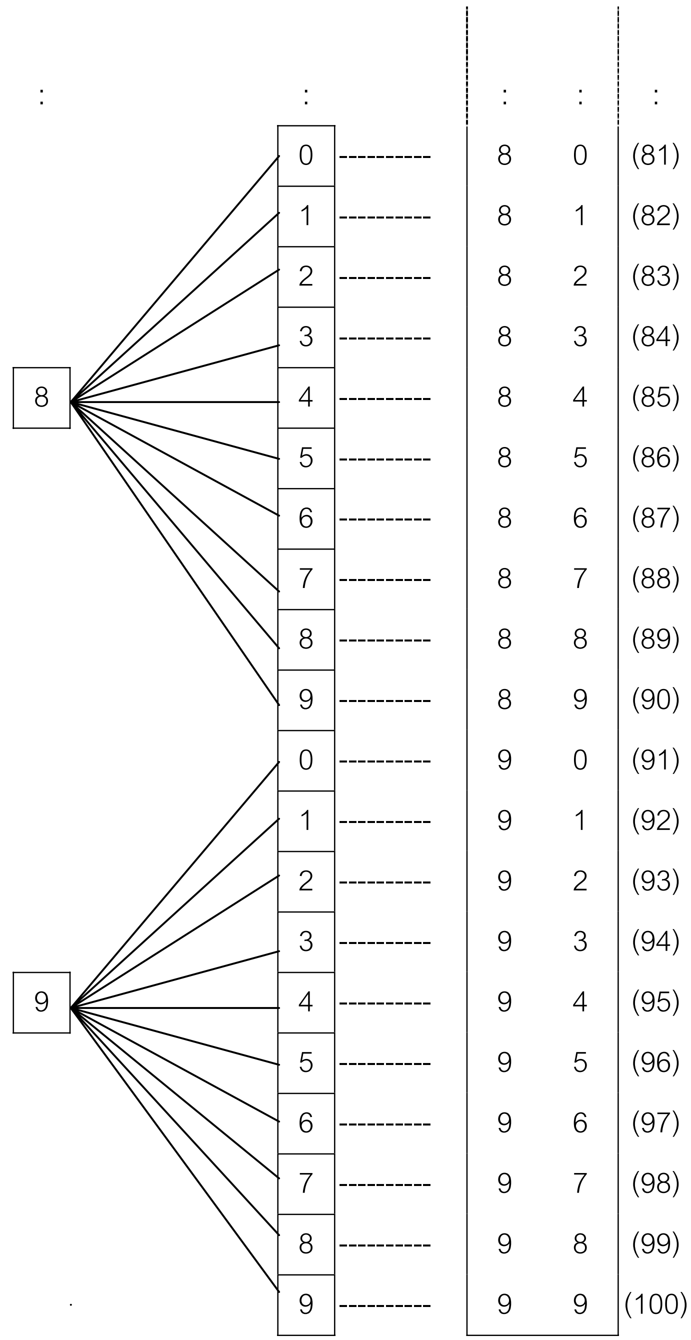
แสดงด้วยแผนภูมิต้นไม้ เมื่อ R และ B เป็นสีเสื้อของพนักงานได้ 4 วิธีดังภาพที่ 5.5

ตัวอย่างที่ 5.6 ประชากรขนาดอนันต์ ประกอบด้วย 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 จงแสดงจำนวนตัวอย่างสุ่มขนาด $n = 2$ ที่พึงเป็นไปได้ทั้งหมด

วิธีทำ ประชากรขนาดอนันต์มีข้อแตกต่างกันภายใน 10 ประการ เมื่อต้องการชักตัวอย่างสุ่มขนาด n เป็น 2



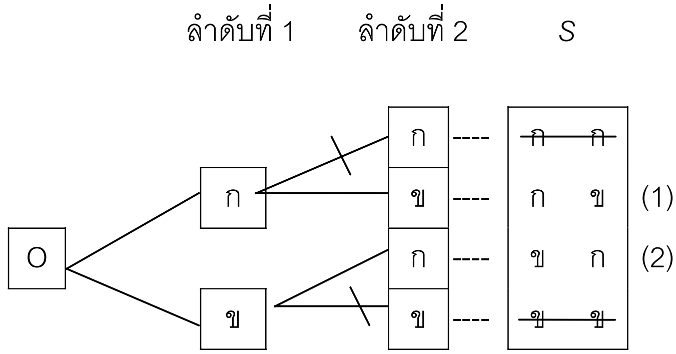
สุ่มครั้งที่ 1
สุ่มครั้งที่ 2
S
0
1
0 0 (1)
0 1 (2)
0 2 (3)
0 3 (4)
0 4 (5)
0 5 (6)
0 6 (7)
0 7 (8)
0 8 (9)
0 9 (10)
1 0 (11)
1 1 (12)
1 2 (13)
1 3 (14)
1 4 (15)
1 5 (16)
1 6 (17)
1 7 (18)
1 8 (19)
1 9 (20)



ภาพที่ 5.6 แผนภูมิต้นไม้จำนวนตัวอย่างสุ่มที่พึงเป็นไปได้ทั้งหมด
จากประชากรขนาดอนันต์

จำนวนตัวอย่างของประชากรที่พึงเป็นไปได้ทั้งหมดเท่ากับ 100 ดังแสดงในภาพที่ 5.6

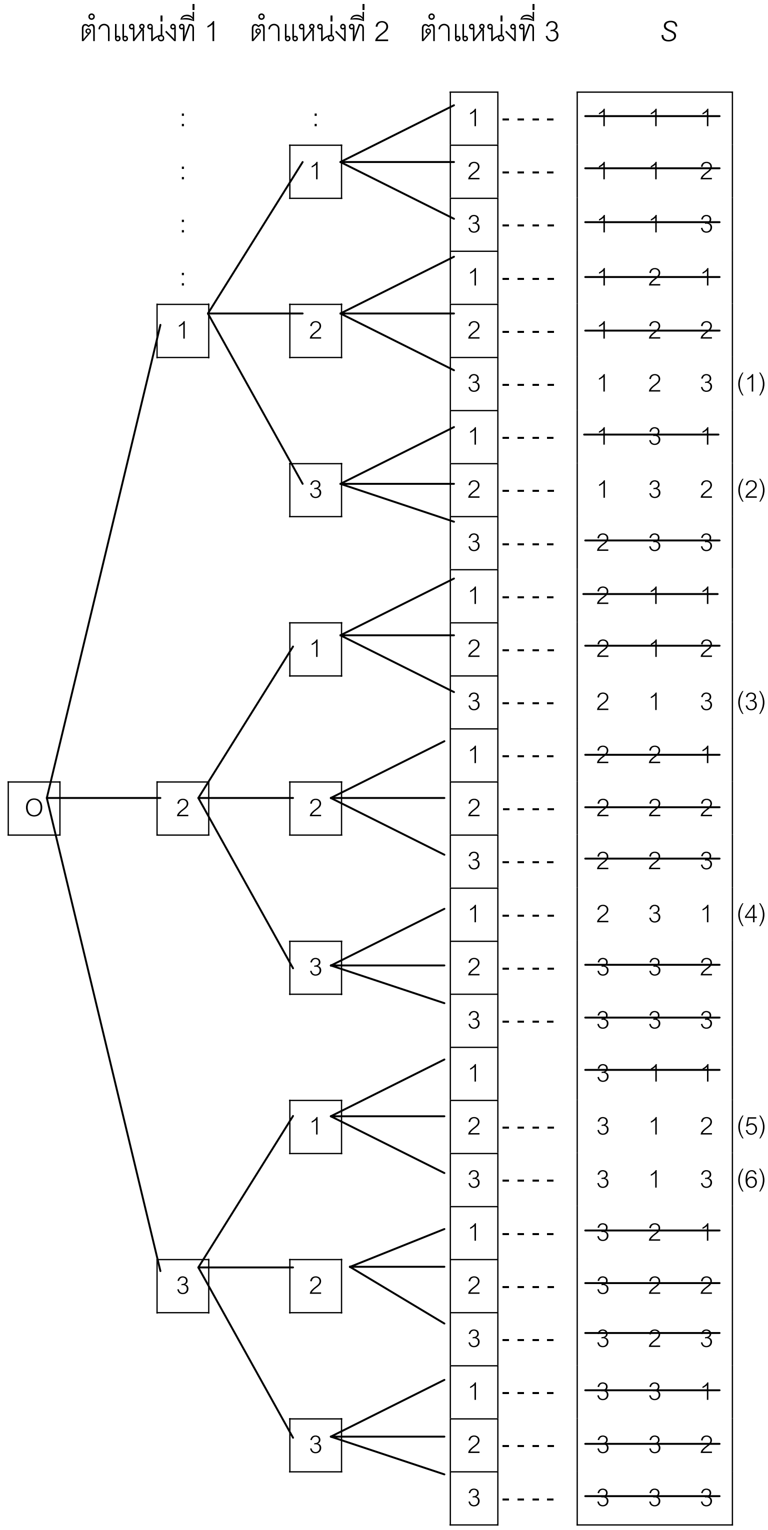
5.3.2 วิธีเรียงสับเปลี่ยน การหาผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดด้วยวิธีเรียงสับเปลี่ยนนี้มีพื้นฐานมาจากการหาจำนวนวิธีที่จะจัดลำดับตำแหน่ง หรือจัดลำดับการเสนอชื่อคน สัตว์ สิ่งของ ซึ่งมีจำนวนตั้งแต่ 2 ขึ้นไปว่า สามารถกระทำได้กี่วิธีที่ไม่ซ้ำกัน เช่น มีคนจำนวน 2 คน คือ นาย ก กับนาย ข อาจจัดลำดับการเสนอชื่อได้เป็น วิธีที่ 1 : (ก , ข) วิธีที่ 2 : (ข , ก) แสดงว่ามีวิธีเรียงสับเปลี่ยน ได้จำนวน 2 วิธี ที่ไม่ซ้ำกัน ซึ่งจำนวนวิธีการเรียงสับเปลี่ยนทั้งหมดก็คือผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดหรือจำนวนหนทางที่พึงเป็นไปได้ทั้งหมด และอาจใช้แผนภูมิต้นไม้แสดงจำนวนวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้ทั้งหมด แต่ตัดกิ่งที่แสดงการจัดเรียงสับเปลี่ยนที่เป็นไปไม่ได้ออกหรือสับเปลี่ยนแล้วมีความหมายเหมือนเดิม สังเกตจากวิธีเรียงสับเปลี่ยนลำดับการเสนอชื่อ นาย ก และนาย ข ด้วยแผนภูมิต้นไม้ ดังภาพที่ 5.7



ภาพที่ 5.7 แผนภูมิต้นไม้วิธีเรียงสับเปลี่ยนลำดับการเสนอชื่อ นาย ก นาย ข และตัดวิธีเรียงสับเปลี่ยนที่เป็นไปไม่ได้ออก

จากภาพที่ 5.7 แผนภูมิต้นไม้มีการแตกกิ่งครั้งละ 2 กิ่ง (นาย ก, นาย ข) จำนวน 2 ครั้ง (ลำดับที่ 1 , ลำดับที่ 2) แต่ตัดวิธีเรียงสับเปลี่ยนที่เป็นไปไม่ได้ คือ (ก , ก) หมายถึง นาย ก อยู่ลำดับที่ 1 และนาย ก อยู่ลำดับที่ 2 และ (ข , ข) หมายถึงนาย ข อยู่ลำดับที่ 2 ในเวลาเดียวกันย่อมเป็นไปไม่ได้ ในการเรียงสับเปลี่ยนลำดับนั้นในที่นี้จะขอยกตัวอย่างวิธีการเรียงสับเปลี่ยนในกรณีที่มีหนังสือจำนวน 3 เล่มที่แตกต่างกัน มาพอเป็นสังเขปดังนี้

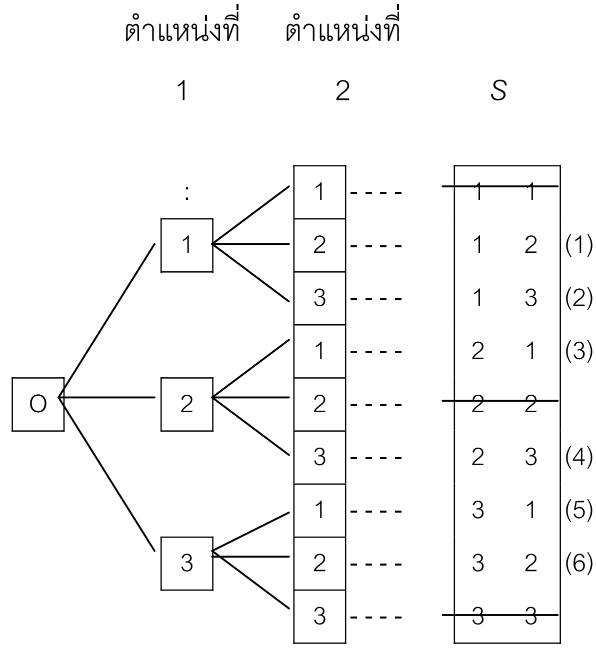
1) กรณีมีหนังสือจำนวน 3 เล่มเรียงสับเปลี่ยนครั้งละ 3 เล่ม คือหนังสือเบอร์ 1 เบอร์ 2 เบอร์ 3 ต้องการหาจำนวนวิธีเรียงสับเปลี่ยนตำแหน่งว่างที่พึงเป็นไปได้ทั้งหมด เช่น แผนภูมิต้นไม้กรณีนี้มีการแตกกิ่งครั้งละ 3 กิ่ง (หนังสือเบอร์ 1 หนังสือเบอร์ 2 หนังสือเบอร์ 3)



ภาพที่ 5.8 แผนภูมิต้นไม้วิธีเรียงสับเปลี่ยนหนังสือเบอร์ 1 เบอร์ 2 และ เบอร์ 3 และตัดวิธีที่เป็นไปไม่ได้ออก

จำนวน 3 ครั้ง (หรือตำแหน่งที่หนึ่ง ตำแหน่งที่สอง ตำแหน่งที่สาม) จะได้จำนวนวิธีเรียงสับเปลี่ยนตำแหน่งว่างที่พึงเป็นไปได้ทั้งหมด 6 วิธี จากกิ่งทั้งหมด 27 กิ่ง ดังภาพที่ 5.8

2) กรณีมีหนังสือจำนวน 3 เล่มเรียงสับเปลี่ยนครั้งละ 2 เล่ม สมมุติว่ามีหนังสือจำนวน 3 เล่ม คือหนังสือเบอร์ 1 หนังสือเบอร์ 2 และหนังสือเบอร์ 3 ต้องการหาจำนวนวิธีเรียงสับเปลี่ยนตำแหน่งว่างที่พึงเป็นไปได้ทั้งหมดทีละ 2 เล่ม แผนภูมิต้นไม้จะเป็นการแตกกิ่งทีละ 3 กิ่ง (หนังสือเบอร์ 1 หนังสือเบอร์ 2 และหนังสือเบอร์ 3 เช่นเดียวกัน) จำนวน 2 ครั้ง (ตำแหน่งที่หนึ่ง ตำแหน่งที่สอง) จะได้จำนวนวิธีเรียงสับเปลี่ยนตำแหน่งว่างที่พึงเป็นไปได้ทั้งหมด 6 วิธี จากกิ่งทั้งหมด 9 กิ่ง ดังภาพที่ 5.9



ภาพที่ 5.9 แผนภูมิต้นไม้วิธีเรียงสับเปลี่ยนหนังสือเบอร์ 1 เบอร์ 2 และ เบอร์ 3 ครั้งละ 2 เล่ม และตัดวิธีที่เป็นไปไม่ได้ออก

เมื่อแผนภูมิต้นไม้สามารถประยุกต์ใช้กับการสุ่มตัวอย่างขนาด n จากประชากรขนาดอนันต์ จึงประยุกต์วิธีเรียงสับเปลี่ยนใช้กับการสุ่มตัวอย่างขนาด n จากประชากรขนาดจำกัด ทั้งนี้เพราะ วิธีเรียงสับเปลี่ยนจะถูกจำกัดด้วยจำนวนคน สัตว์ สิ่งของที่มีอยู่ ซึ่งเปรียบเทียบเป็นคุณลักษณะที่เราสนใจหรือข้อแตกต่างกันภายในประชากร เมื่อคุณลักษณะใดหรือข้อแตกต่างกันภายในประชากรถูกหยิบยกออกมากล่าวแล้ว (หยิบออกแล้วไม่ใส่คืนก่อนหยิบครั้งต่อไป)

ซึ่งจะไม่นำมากล่าวถึงอีก ส่วนการกำหนดให้นำสิ่งที่มีอยู่ ครั้งละกี่สิ่งมาเรียงสับเปลี่ยน เช่น มีหนังสืออยู่ 3 เล่ม แตกต่างกันตามภาพที่ 5.8 นำออกมาเรียงสับเปลี่ยนตำแหน่งการวางครั้งละ 2 เล่ม คือตัวอย่างขนาด *n* เป็น 2 หรือกำหนดให้นำออกมาเรียงสับเปลี่ยนตำแหน่งการวางครั้งละ 3 เล่ม (ทั้งหมด) คือ ตัวอย่างขนาด *n* เป็น 3 เป็นต้น

จากที่กล่าวมา การหาวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้ทั้งหมดด้วยการอาศัยแผนภูมิต้นไม้แต่ตัดส่วนที่เป็นไปไม่ได้ด้วยวิธีเรียงสับเปลี่ยนออกไม่สะดวกเท่าที่ควรจึงขอเสนอกฎเกณฑ์ทางคณิตศาสตร์เกี่ยวกับการหาจำนวนวิธีเรียงสับเปลี่ยน ในกรณีมีสิ่งของ (คน สัตว์หรืออื่น ๆ) ที่แตกต่างกัน *m* ชิ้น กำหนดให้เรียงสับเปลี่ยนลำดับตำแหน่งหรือลำดับการเสนอชื่อ ครั้งละ *n* ชิ้น จะได้สมการ (5-6)

จำนวนวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้ทั้งหมดเท่ากับ P_n^m …(5-6)

อ่านว่า *m* การเรียงสับเปลี่ยน *n* หรือ *m* การจัดลำดับ *n* โดยมีข้อกำหนดดังสมการ (5-7)

$$P_n^m = \frac{m!}{(m-n)!} \qquad ...(5\text{-}7)$$

และ *m*! อ่านว่า *m* แฟกทอเรียล (*m* factorial) คือ ผลคูณของตัวเลขตั้งแต่ค่าที่กำหนดกับค่าที่ลดหลั่นลงไปครั้งละ 1 หน่วยเรื่อย ๆ ไปจนถึง 1 เช่น $2! = (2)(1)$ (อ่านว่า 2 แฟกทอเรียลเท่ากับ 2 คูณ 1)

$$10! = (10)(9)(8) \text{ .... } (3)(2)(1)$$

(อ่านว่า 10 แฟกทอเรียลเท่ากับ 10 คูณ 9 คูณ 8 คูณ .... คูณ 3 คูณ 2 คูณ 1) หรือกรณี *m* เป็นเลขจำนวนเต็มบวกใด ๆ จะได้

$$m! = m(m-1)(m-2)(m-3) \text{ .... } (3)(2)(1) \qquad ..(5\text{-}8)$$

และ $$0! = 1 \qquad ...(5\text{-}9)$$

ตัวอย่าง 5.7 จงหาจำนวนวิธีเรียงสับเปลี่ยนตำแหน่งวางหนังสือ 3 เล่ม ครั้งละ 3 เล่ม

วิธีทำ จำนวนวิธีเรียงสับเปลี่ยนที่พึ่งเป็นไปไดทั้งหมด P_n^m วิธี

$$P_3^3 = \frac{3!}{(3-3)!} \quad ; \quad m = 3 \; , \; n = 3$$

$$= \frac{(3)(2)(1)}{1} \quad ; \quad 0\,! = 1$$

$$P_3^3 = 6$$

จำนวนวิธีเรียงสับเปลี่ยนตำแหน่งวางหนังสือ 3 เล่มนี้มี 6 วิธี

ตัวอย่างที่ 5.8 จงหาจำนวนวิธีเรียงสับเปลี่ยนตำแหน่งวางหนังสือ 3 เล่ม (แตกต่างกัน) ครั้งละ 2 เล่ม

วิธีทำ จำนวนวิธีเรียงสับเปลี่ยนที่เป็นไปได้ทั้งหมด P_n^m วิธี เมื่อ $m = 3$ และ $n = 2$

$$P_2^3 = \frac{3!}{(3-2)!}$$

$$= \frac{(3)(2)(1)}{1} \quad ; \quad 1! = 1$$

$$= 6 \quad \text{วิธี}$$

จำนวนวิธีเรียงสับเปลี่ยนตำแหน่งวางหนังสือ 3 เล่ม ครั้งละ 2 เล่ม มี 6 วิธี

จากตัวอย่างที่ 5.7 มีคำตอบเท่ากับจำนวนวิธีที่พึงเป็นไปได้ทั้งหมดในภาพที่ 5.8 สำหรับตัวอย่างที่ 5.8 มีคำตอบเท่ากับจำนวนวิธีที่พึงเป็นไปได้ทั้งหมดในภาพที่ 5.9

ตัวอย่างที่ 5.9 หัวเทียนรถยนต์ 4 ตัวจะสามารถสับเปลี่ยนตำแหน่งการใช้กับลูกสูบเครื่องยนต์ได้กี่วิธี

วิธีทำ จำนวนวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้ทั้งหมด P_n^m วิธี

เมื่อ $m = 4$ และ $n = 4$

$$P_4^4 = \frac{4!}{(4-4)!}$$

$$= \frac{(4)(3)(2)(1)}{1}$$

$$= 24$$

จำนวนวิธีเรียงสับเปลี่ยนมีทั้งหมด 24 วิธี

ตัวอย่างที่ 5.10 ลูกบอลจำนวน 8 ลูก เบอร์ 1 , เบอร์ 2 , ... , เบอร์ 8 บรรจุไว้ในภาชนะ สุ่มหยิบครั้งละ 1 ลูกจนได้ 3 ลูก จงหาผลลัพธ์ที่เป็นไปได้ทั้งหมด

วิธีทำ การสุ่มหยิบครั้งละ 1 ลูก (ไม่ใส่กลับคืน) จนได้ 3 ลูก แสดงว่าจำนวนลูกบอลถูกจำกัดเช่นเดียวกับวิธีเรียงสับเปลี่ยน จึงได้ว่า จำนวนวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้ทั้งหมดคือผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด

จำนวนผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด P_n^m วิธี

$m = 8$ และ $n = 3$

$$P_3^8 = \frac{8!}{(8-3)!}$$

$$= \frac{(8)(7)(6)(5)(4)(3)(2)(1)}{(5)(4)(3)(2)(1)}$$

$$= 336$$

จำนวนวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้ทั้งหมดมี 336 วิธี

ตัวอย่างที่ 5.11 ประชากรประกอบด้วยคุณลักษณะ 4 ประการคือ 3, 4, 5 และ 9 ต้องการตัวอย่างขนาด $n = 2$ สุ่มหยิบครั้งละ 1 หน่วยจนครบ 2 หน่วยตามต้องการ (หยิบได้แล้วไม่ใส่กลับคืน) จงหาจำนวนตัวอย่างที่เป็นไปได้ทั้งหมด

วิธีทำ การสุ่มหยิบครั้งละ 1 หน่วย (ไม่ใส่กลับคืน) จนได้ตัวอย่างขนาด $n = 2$ แสดงว่าจำนวนประชากรถูกจำกัดโดยวิธีเรียงสับเปลี่ยนจึงได้จำนวนวิธีเรียงสับเปลี่ยนที่พึงเป็นไปได้คือจำนวนตัวอย่างขนาด $n = 2$ ที่พึงเป็นไปได้ทั้งหมด

จำนวนตัวอย่างขนาด $n = 2$ ที่พึงเป็นไปได้ทั้งหมด P_n^m ตัวอย่าง

$m = 4$ และ $n = 2$

$$P_2^4 = \frac{4!}{(4-2)!}$$

$$= \frac{(4)(3)(2)(1)}{(2)(1)} = 12$$

จำนวนตัวอย่างที่เป็นไปได้ทั้งหมดทั้งหมด 12 แบบ

5.3.3 วิธีจัดหมู่ การหาผลลัพธ์ที่เป็นไปได้ทั้งหมดด้วยวิธีจัดหมู่ เป็นการจัดหมู่ คน สัตว์ สิ่งของจำนวนเดียวกันกับวิธีเรียงสับเปลี่ยน แต่ไม่ถือว่าการจัดลำดับตำแหน่งหรือลำดับการเสนอชื่อก่อนหลังเป็นสาระสำคัญในการแยกหมู่หรือแยกกลุ่ม เช่น กลุ่ม (ก, ข) ไม่ถือว่าแตกต่างจากกลุ่ม (ข, ก) การจัดหนังสือจำนวน 3 เล่มเข้าที่ วางครั้งละ 3 เล่มตามวิธีเรียงสับเปลี่ยนได้ถึง 6 วิธีที่ไม่ซ้ำกัน ดังภาพที่ 5.7 ได้แก่ S คือ {1 2 3 , 1 3 2 , 2 1 3 , 2 3 1 , 3 1 2 , 3 2 1} เมื่อเป็นวิธีการจัดหมู่จะถือว่ามีวิธีเดียวคือ S คือ {1 2 3} เท่านั้น หรือการจัดหนังสือจำนวน 3 เล่มเขาที่วางครั้งละ 2 เล่มตามวิธีเรียงสับเปลี่ยนได้ถึง 6 วิธีที่ไม่ซ้ำกัน ดังภาพที่ 5.8 ได้แก่ S คือ { 1 2 , 1 3 , 2 1 , 2 3 , 3 1 , 3 2 } เมื่อเป็นวิธีจัดหมู่จะถือว่ามี 3 วิธีคือ S คือ { 1 2 , 1 3 , 2 3 } เป็นต้น

ถ้ามีสิ่งของ คน สัตว์หรืออื่น ๆ ที่แตกต่างกัน m ชิ้น กำหนดให้เป็นหมู่หรือเป็นกลุ่ม ๆ ละ n ชิ้น จะได้สมการ (5-10) และสมการ (5-11)

จำนวนวิธีจัดหมู่ที่พึงเป็นไปได้ทั้งหมด $= \binom{m}{n}$ วิธี ...(5-10)

อ่านว่า m การจัดหมู่ n (m combination n) โดยกำหนดว่า

$$\binom{m}{n} = \frac{m!}{(m-n)!\,n!} \qquad \text{...(5-11)}$$

ตัวอย่างที่ 5.12 จงหาวิธีจัดหมู่หนังสือ 3 เล่มที่มีสีปกที่แตกต่างกัน ครั้งละ 3 เล่ม

วิธีทำ จำนวนวิธีจัดหมู่ที่พึงเป็นไปได้ทั้งหมด $\binom{m}{n}$ วิธี

$m = 3$ และ $n = 3$

$$\binom{3}{3} = \frac{3!}{(3-3)!3!}$$

$$= \frac{(3)(2)(1)}{(1)(3)(2)(1)}$$

$$= 1$$

ดังนั้นในการจัดหมู่ที่พึงเป็นไปได้ทั้งหมด 1 วิธี คือ $S = \{1\ 2\ 3\}$

ตัวอย่างที่ 5.13 จงหาวิธีจัดหมู่หนังสือ 3 เล่ม ที่มีสีปกที่แตกต่างกัน ครั้งละ 2 เล่ม

วิธีทำ จำนวนวิธีจัดหมู่ที่พึงเป็นไปได้ทั้งหมด $\binom{m}{n}$ วิธี

$m = 3$ และ n =- 2

$$\binom{3}{2} = \frac{3!}{(3-2)!2!}$$

$$= \frac{(3)(2)(1)}{(1)(2)(1)} = 3$$

ดังนั้นในการจัดหมู่ที่พึงเป็นไปได้ทั้งหมด 3 วิธี คือ $S = \{1\ 2, 1\ 3, 2\ 3\}$

วิธีจัดหมู่นี้ถูกนำไปประยุกต์ใช้อย่างแพร่หลายในชีวิตประจำวัน เช่น การจัดหมู่นักศึกษา การจัดทีมฟุตบอล การจัดกรรมการในงานใดงานหนึ่งหรืออื่น ๆ โดยเฉพาะอย่างยิ่งในกระบวนการสุ่มตัวอย่างขนาด n จากประชากรที่มีคุณลักษณะที่เราสนใจหรือประกอบด้วยข้อแตกต่างกันภายในประชากร m ประการ ด้วยเทคนิคการชักตัวอย่างสุ่ม

แบบไม่คืนที่ ซึ่งจะกล่าวในหัวข้อ 6.2.2 บทที่ 6 ต่อไป อย่างไรก็ดีการหาผลลัพธ์ที่เป็นไปได้ทั้งหมดไม่ว่าวิธีใดใน 3 วิธีดังที่กล่าวมาแล้วในหัวข้อ 5.3.1 ถึง 5.3.3 เป็นการหาผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดสำหรับการหาค่าความน่าจะเป็นของเหตุการณ์เชิงเดี่ยวซึ่งเป็นสมาชิกของเซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดนั่นเอง

5.4 ความน่าจะเป็นของเหตุการณ์

การหาค่าความน่าจะเป็นของเหตุการณ์ เป็นการคำนวณค่าน่าจะเป็นหรือโอกาสของการเกิดเหตุการณ์ใดเหตุการณ์หนึ่งหรือหลายเหตุการณ์ในจำนวนผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดตามนิยามความน่าจะเป็นในสมการ (5-1) ซึ่งหาได้โดยเงื่อนไขดังนี้

5.4.1 ความน่าจะเป็นของเหตุการณ์เชิงเดี่ยว (probabilities of simple event) เหตุการณ์เชิงเดี่ยวเป็นเหตุการณ์ที่มีสมาชิกเพียงหนึ่งตัวของเซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด จะสามารถบอกค่าความน่าจะเป็นของเหตุการณ์เชิงเดี่ยวได้ก็ต่อเมื่อจำนวนผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดก่อน เช่น

1) กรณีวางสินค้าจำนวน 1 ชิ้น จำนวนผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด เป็นเหตุการณ์ A_1 คือขายสินค้าได้ (G) และเหตุการณ์ A_2 คือขายสินค้าไม่ได้ (D) ค่าความน่าจะเป็นหาได้ดังตารางที่ 5.1

ตารางที่ 5.1 ความน่าจะเป็นของการขายสินค้าที่มีอยู่ในร้านจำนวน 1 ชิ้น

เหตุการณ์เชิงเดี่ยว	ความน่าจะเป็น
(G) : A_1	$\frac{1}{2}$
(D) : A_2	$\frac{1}{2}$
	1.00

2) กรณีวางสินค้าขายจำนวน 3 ชิ้น จากตัวอย่าง 5.2 ซึ่งพบว่ามีจำนวนเหตุการณ์เชิงเดี่ยวที่พึงเป็นไปได้ทั้งหมดถึง 8 เหตุการณ์ คือ เหตุการณ์ A_1 (GGG) เหตุการณ์ A_2 (GGD) … และเหตุการณ์ A_8 (DDD) ค่าความน่าจะเป็นดังตารางที่ 5.2

ตารางที่ 5.2 ความน่าจะเป็นของขายสินค้าที่มีอยู่ในร้านจำนวน 3 ชิ้น

เหตุการณ์เชิงเดี่ยว	ความน่าจะเป็น
(GGG) : A_1	$\frac{1}{8}$
(GGD) : A_2	$\frac{1}{8}$
(GDG) : A_3	$\frac{1}{8}$
(GDD) : A_4	$\frac{1}{8}$
(DGG) : A_5	$\frac{1}{8}$
(DGD) : A_6	$\frac{1}{8}$
(DDG) : A_7	$\frac{1}{8}$
(DDD) : A_8	$\frac{1}{8}$
	1

จากตารางที่ 5.2 ถ้าเราสนใจเหตุการณ์เชิงเดี่ยวอันใดอันหนึ่ง (Wannacott & Wannacott, 1985, pp. 55-57) เช่นเหตุการณ์ในการจำหน่ายสินค้าจำนวนหนึ่งพบว่า สินค้าชิ้นที่ 1 ขายได้ สินค้าชิ้นที่ 2 ขายไม่ได้ สินค้าชิ้นที่ 3 ขายได้ ตรงกับ เหตุการณ์ A_3 ถามว่าเหตุการณ์นี้มีความน่าจะเป็นเท่าใด ? หรือ เหตุการณ์นี้มีโอกาสเป็นไปได้เท่าไร คำตอบคือความน่าจะเป็นเหตุการณ์ A_3 หรือเขียนแทนว่า P(GDG) หรือ $P(A_3)$ เท่ากับ 1/8

ลักษณะเหตุการณ์ที่เรากำลังสนใจถามถึงดังที่กล่าวมา เป็นเหตุการณ์ที่ไม่สามารถเกิดพร้อมกับเหตุการณ์อื่น ๆ ได้ เราเรียกเหตุการณ์เช่นนี้ว่า เหตุการณ์ที่ไม่เกิดร่วม (mutually exclusive events)

3) กรณีการออกรางวัลเลขท้ายหนึ่งตัวจะพบว่าจำนวนเหตุการณ์เชิงเดี่ยวที่พึงเป็นไปได้ทั้งหมด 10 เหตุการณ์คือเหตุการณ์ A_1:0 เหตุการณ์ A_2:1 เหตุการณ์ A_3 : 2 , ………… และเหตุการณ์ A_{10} : 9 . จึงได้ความน่าจะเป็นดังตารางที่ 5.3

ตารางที่ 5.3 ความน่าจะเป็นของการออกรางวัลเลขท้ายหนึ่งตัว

เหตุการณ์เชิงเดี่ยว	ความน่าจะเป็น
0 : A_1	$\frac{1}{10}$
1 : A_2	$\frac{1}{10}$
2 : A_3	$\frac{1}{10}$
3 : A_4	$\frac{1}{10}$
4 : A_5	$\frac{1}{10}$
5 : A_6	$\frac{1}{10}$
6 : A_7	$\frac{1}{10}$
7 : A_8	$\frac{1}{10}$
8 : A_9	$\frac{1}{10}$
9 : A_{10}	$\frac{1}{10}$
	1

ตัวอย่างที่ 5.13 จงหาความน่าจะเป็นของการได้หัว (H) ทั้ง 4 ครั้ง เมื่อทอดเหรียญบาทจำนวน 4 ครั้ง

วิธีทำ เหตุการณ์ที่เรากำลังสนใจอยู่ คือ (HHHH) มีอยู่ 1 ในจำนวน 16 เหตุการณ์

จึงได้ $P\,(\text{HHHH}) = \frac{1}{16}$

ความน่าจะเป็นที่จะออกหัว (H) ในการทอดเหรียญบาท 4 ครั้ง เท่ากับ $\frac{1}{16}$ หรือ 0.0625

5.4.2 ความน่าจะเป็นของเหตุการณ์เชิงประกอบ (probabilities of compound events) จากตารางที่ 5.2 ถ้า E เป็นเหตุการณ์สินค้าขายไม่ได้อย่างน้อย 2 ชิ้น จะเห็นว่าเหตุการณ์ E ประกอบด้วยเหตุการณ์เชิงเดี่ยว A_4, A_6, A_7 หรือ A_8 จึงจัดว่า E ประกอบด้วยเหตุการณ์เชิงประกอบ และเหตุการณ์เชิงประกอบจะยังคงเป็นเซตย่อยของเซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดอยู่ นั่นคือ

$$E = \{A_4 , A_6 , A_7 , A_8\}$$

เมื่อต้องการทราบค่าความน่าจะเป็นของ E เป็นเท่าใด หรือ โอกาสที่จะเป็น E เท่าใด แนวทางหนึ่งที่จะตอบคำถามนี้ได้โดยใช้สมการ (5-2) แล้วจะได้

$$P(\mathrm{E}) = P(\mathrm{A}_4) + P(\mathrm{A}_6) + P(\mathrm{A}_7) + P(\mathrm{A}_8)$$

$$P(\mathrm{E}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

$$= \frac{4}{8} = 0.5$$

เพื่อทำความเข้าใจอย่างชัดเจนยิ่งขึ้น จึงขอเสนอความรู้เกี่ยวกับพีชคณิตของเหตุการณ์ (algebra of events) หรือพีชคณิตของเซต (algebra of sets) ที่สำคัญ ได้แก่ การผนวกเหตุการณ์ (union of events) เหตุการณ์ร่วม (intersection of events) และเหตุการณ์ส่วนเติมเต็ม (complement of events) พร้อมทั้งใช้แผนภาพเวนน์ (Venn diagram) ซึ่ง จอห์น เวนน์ (John Venn :1834–1923) นักตรรกวิทยาชาวอังกฤษ (Keller & Warrack, 2000, pp.171-172) ใช้อธิบายความสัมพันธ์ระหว่างเซตด้วยการวาดภาพเซตและสมาชิกเซตบนพื้นระนาบดังต่อไปนี้

1) การผนวกเหตุการณ์ การผนวกเหตุการณ์หรือการรวมเหตุการณ์นิยามในรูปเซตได้ว่า

$A \cup B$ = เซตของสมาชิก เซต A หรือ เซต B …(5-12)

เมื่อ A : เป็นเหตุการณ์ที่เรากำลังสนใจอยู่

B : เป็นเหตุการณ์ที่เรากำลังสนใจอยู่อีกอันหนึ่ง

และ A , B เป็น เหตุการณ์ที่ไม่เกิดร่วม (mutually exclusive events) เช่น เหตุการณ์ A_1 ถึง A_8 จากตารางที่ 5.2 เหตุการณ์ A_1 คือ สินค้าชิ้นที่ 1 ขายได้ สินค้าชิ้นที่ 2 ขายได้ สินค้าชิ้นที่ 3 ขายได้ จะเกิดพร้อมกับเหตุการณ์ A_4 คือ สินค้าชิ้นที่ 1 ขายได้ สินค้าชิ้นที่ 2 ขายไม่ได้ สินค้าชิ้นที่ 3 ขายไม่ได้ เป็นต้น

$A \cup B$ อ่านว่า A ยูเนียน B หรือ A ผนวก B หรือ A หรือ B คือ เซตที่ประกอบด้วยสมาชิกของเซต A ทั้งหมด และสมาชิกของเซต B ทั้งหมด เช่น

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$A = \{1, 3, 4, 5\}$$

$$B = \{2, 6\}$$

$$A \cup B = \{1, 3, 4, 5\} \cup \{2, 6\}$$

$$= \{1, 2, 3, 4, 5, 6\}$$

แสดงด้วยแผนภาพเวนน์ ได้ดังภาพที่ 5.10

ความน่าจะเป็น A, B และ $A \cup B$ คือ

$$P(A) = \frac{4}{10}, \quad P(B) = \frac{2}{10}$$

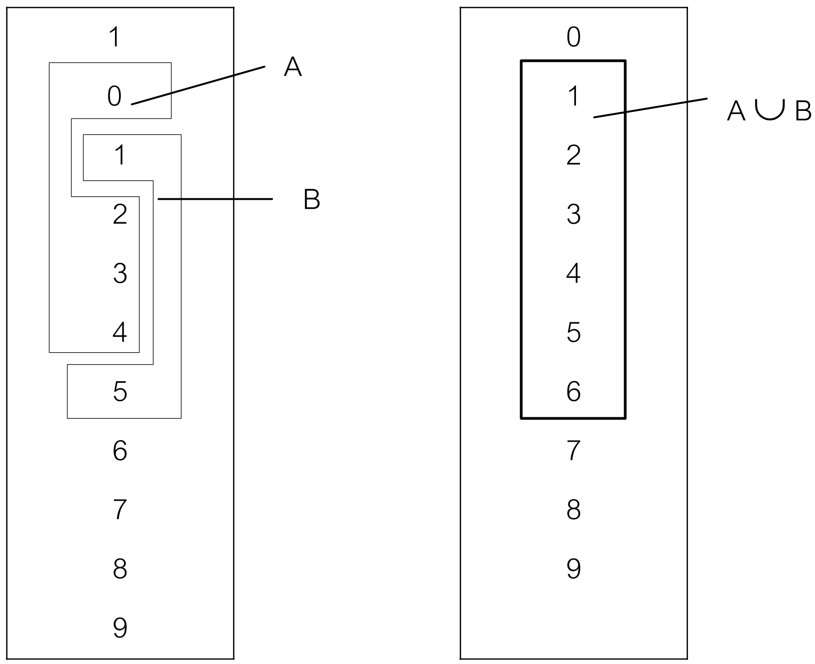
และ $P(A \cup B) = \dfrac{6}{10}$

หรือ $P(A \cup B) = P(A) + P(B)$ …(5-13)

$$= \frac{4}{10} + \frac{2}{10}$$

$$= \frac{6}{10}$$

$$= 0.6$$



(ก) (ข)

ภาพที่ 5.10 แผนภาพเวนน์ (ก) เซต A เซต B (ข) A ∪ B

เราเรียก กฎการผนวกเหตุการณ์ว่า กฎการรวม (addition rule) ถ้า $A_1, A_2, \ldots$ เป็นเหตุการณ์เชิงเดี่ยวของเซต S และ เป็นเหตุการณ์ที่ไม่เกิดร่วม มีความน่าจะเป็นเท่ากับ $P(A_1)$, $P(A_2)$, … ตามลำดับ จะได้สมการ (5-14)

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots \qquad \ldots(5\text{-}14)$$

ตัวอย่างที่ 5.14 จากตารางที่ 5.2 จงหาความน่าจะเป็นของเหตุการณ์ต่อไปนี้

ก. F : สินค้าชิ้นที่ 2 ขายไม่ได้ แต่สินค้าชิ้นที่ 3 ขายได้

ข. K : สินค้าขายได้ไม่น้อยกว่า 2 ชิ้น

ค. E : สินค้าขายไม่ได้ไม่น้อยกว่า 2 ชิ้น

วิธีทำ จากตารางที่ 5.2 A_1(GGG) , A_2 (GGD) , A_3 (GDG), A_4 (GDD), A_5 (DGG) , A_6 (DGD) , A_7 (DDG) และ A_8 (DDD) เป็นเหตุการณ์ที่ไม่เกิดร่วม เมื่อ G แทนการขายสินค้าได้และ D แทนการขายสินค้าไม่ได้

ก. สินค้าชิ้นที่ 2 ขายไม่ได้ แต่สินค้าชิ้นที่ 3 ขายได้ ดังนั้นจะได้ตรงกับ A_3 (GDG) และ A_7 (DDG) นั่นคือจะได้ว่า

$$F = \{A_3 \cup A_7\}$$

$$P(F) = P(A_3) + P(A_7)$$

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

$$= \frac{2}{8} = 0.25$$

ความน่าจะเป็นที่สินค้าชิ้นที่ 2 ขายไม่ได้ แต่สินค้าชิ้นที่ 3 ขายได้ มีค่าเท่ากับ 0.25

ข. สินค้าขายได้ไม่น้อยกว่า 2 ชิ้น จะตรงกับเหตุการณ์ A_1(GGG) , A_2 (GGD) , A_3 (GDG), และ A_5 (DGG) นั่นคือจะได้ว่า

$$F = \{A_1 \cup A_2 \cup A_3 \cup A_5\}$$

$$P(K) = P(A_1) + P(A_2) + P(A_3) + P(A_5)$$

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

$$= \frac{4}{8} = 0.5$$

ความน่าจะเป็นที่สินค้าขายได้ไม่น้อยกว่า 2 ชิ้น มีค่าเท่ากับ 0.5

ค. สินค้าขายไม่ได้ไม่น้อยกว่า 2 ชิ้น จะตรงกับเหตุการณ์ A_4 (GDD), A_5 (DGG) , A_7 (DDG) และ A_8 (DDD) นั่นคือจะได้ว่า

$$H = \{A_4 \cup A_6 \cup A_7 \cup A_8\}$$

$$P(H) = P(A_4) + P(A_6) + P(A_7) + P(A_8)$$

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

$$P(\mathrm{H}) = \frac{4}{8} = 0.5$$

ความน่าจะเป็นที่สินค้าขายไม่ได้ไม่น้อยกว่า 2 ชิ้น มีค่าเท่ากับ 0.5

2) เหตุการณ์ร่วม เมื่อเหตุการณ์ที่เรากำลังสนใจอยู่มีความหลากหลาย พิจารณาตารางที่ 5.2 ถ้าเหตุการณ์ถ้าที่เรากำลังสนใจอยู่ คือ

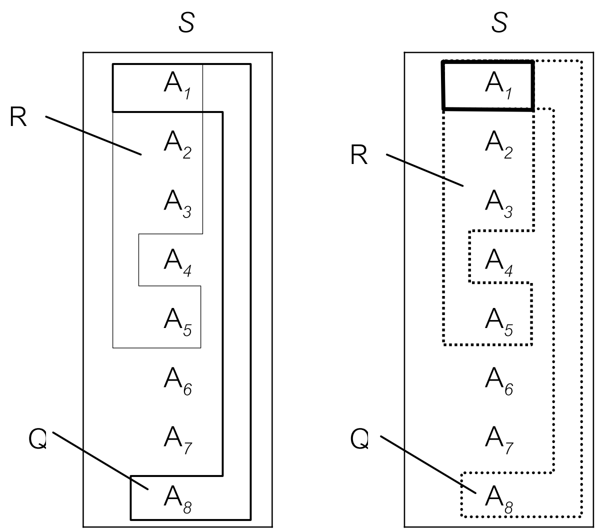
R : สินค้าขายได้ไม่น้อยกว่า 2 ชิ้น

Q : สินค้ามีสภาพเดียวกันทุกชิ้น

จะพบว่าใน R มีเหตุการณ์ปลาอยู่ขายได้ทั้งหมดอยู่ในเหตุการณ์ Q ด้วยเราเรียกว่า R และ Q เป็นเหตุการณ์ร่วม นิยามในรูปของเซตตามสมการ (5-15)

$R \cap Q$ = เซตของสมาชิก เซต R และ เซต Q ในเวลาเดียวกัน …(5-15)

เมื่อ R และ Q อิสระต่อกัน $R \cap Q$ อ่านว่า R เหตุการณ์ร่วม Q (R intersection Q) หรือ R และ Q



(ก) $R \cup Q$ (ข) $R \cap Q$

ภาพที่ 5.11 แผนภาพเวนน์ของเหตุการณ์ (ก) $R \cup Q$ และ (ข) $R \cap Q$

จากภาพที่ 5.11 (ข) จะได้ว่า

$$R \cap Q = \{A_1\}$$

และความน่าจะเป็นของ $R \cap Q$ คือ

$$P(R \cap Q) = \frac{1}{8}$$

เมื่อเป็นเช่นนี้หากถามว่าความน่าจะเป็นของ R ผนวก Q เท่ากับเท่าใด หรือ $P(R \cap Q)$ เท่ากับเท่าใด เราอาจพบคำตอบจากภาพที่ 5.11 (ก)

$$P(R \cup Q) = \frac{5}{8}$$

ได้คำตอบที่ไม่ตรงกับหลักเกณฑ์ตามสมการ (5-13)

$$P(R \cup Q) \neq P(R) + P(Q)$$

$$\frac{5}{8} \neq \frac{4}{8} + \frac{2}{8}$$

สาเหตุที่คำตอบไม่ตรงหลักเกณฑ์ตามสมการ (5-13) เพราะการรวม $P(R)$ และ $P(Q)$ เรานับ A_1 ถึงสองครั้ง คำตอบที่ถูกต้องควรจะเป็นสมการ (5-16)

$$P(R \cup Q) = P(R) + P(Q) - P(R \cap Q) \qquad \ldots(5\text{-}16)$$

$$= \frac{1}{8} + \frac{1}{8} - \frac{1}{8}$$

$$= \frac{1}{8}$$

เห็นได้ว่าการผนวกเหตุการณ์ A กับ B ตามสมการ (6-13) และ การผนวกเหตุการณ์ R และ Q ตามสมการ (5-16) นี้ไม่ขัดแย้งกัน ทั้งนี้เพราะเรากำหนดไว้ก่อนแล้วว่าเหตุการณ์ A และ B เป็นเหตุการณ์ที่ไม่เกิดร่วมกัน หมายความว่า $A \cap B$ คือ { } อ่านว่า เซตว่าง

หรือ ชุดว่าง (empty set or null set or void set) คือ เซตที่ไม่มีสมาชิก หรือเซตเหตุการณ์ที่ไม่มีในเซต S ทำให้ $P(A \cap B)$ เท่ากับศูนย์

สำหรับความเป็นอิสระต่อกันของ R และ Q เช่น สินค้าขายไม่ได้น้อยกว่า 2 ชิ้น ไม่ได้บังคับให้ขายสินค้าได้มีสภาพเดียวกันทุกชิ้น (ขายได้ทั้งหมดหรือไม่ได้ทั้งหมด) การทอดเหรียญบาท ครั้งแรกออกหัวไม่ได้บังคับให้การทอดเหรียญครั้งที่สองออกหัว หรือ การออกรางวัลเลขท้าย สองตัวหลักหน่วยจะได้เลขใดไม่อยู่ในการบังคับของหลักสิบ เป็นต้นจึงอาจสรุปเกี่ยวกับเหตุการณ์ร่วมได้ว่า

ถ้า $A_1, A_2, A_3, \ldots$ เป็นเหตุการณ์อิสระต่อกัน มีความน่าจะเป็นเท่ากับ $P(A_1)$, $P(A_2)$, $P(A_3)$, … ตามลำดับ ความน่าจะเป็นของ A_1 และ A_2 และ A_3 … ตามลำดับ คือ

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots) = P(A_1).P(A_2).P(A_3)\ldots.. \qquad \ldots(5\text{-}17)$$

ถึงแม้ว่าเหตุการณ์ร่วมเหล่านั้น เราอาจหาความน่าจะเป็นในรูปของเหตุการณ์เชิงเดี่ยวได้ก็ตาม แต่เมื่อพิจารณาเหตุการณ์เชิงเดี่ยวแล้วพบว่าเกิดจากเหตุการณ์อิสระหลายเหตุการณ์ร่วมกัน เราสามารถนำหลักตามสมการ (5-17) หาความน่าจะเป็นของเหตุการณ์นั้น ๆ ได้ดังตัวอย่างที่ 5.15 ถึงตัวอย่างที่ 5.17

ตัวอย่างที่ 5.15 ถ้าสินค้าแต่ละชิ้นมีความน่าจะเป็นขายได้เท่ากับ $\frac{1}{2}$ จงหาความน่าจะเป็นของสินค้าขายได้ทุกชิ้น เมื่อขายสินค้าจำนวน 3 ตัว

วิธีทำ สินค้าแต่ละตัวอิสระต่อกัน มี $P(G) = \frac{1}{2}$ ทุกตัวดังนั้นความน่าจะเป็นของสินค้าขายได้ทุกชิ้น คือ ความน่าจะเป็นของ G และ G และ G (สินค้าชิ้นที่ 1, 2 และ 3 ตามลำดับ) จึงได้

$$\begin{aligned} P(G \cap G \cap G) &= P(G) \cdot P(G) \cdot P(G) \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \\ &= \frac{1}{8} \end{aligned}$$

ความน่าจะเป็นหรือโอกาสที่สินค้าจะขายได้ทุกชิ้นเท่ากับ $\frac{1}{8}$

จากตัวอย่างที่ 5.15 ถ้า $P(\mathrm{G})$ เป็น 0.2 จะได้ $P(\mathrm{G} \cap \mathrm{G} \cap \mathrm{G})$ เท่ากับ (0.2)(0.2)(0.2) หรือ 0.008

ตัวอย่างที่ 5.16 ถ้าซื้อสลากกินแบ่งรัฐบาลมีเลขท้ายเป็น 65 จงหาความน่าจะเป็นของการถูกรางวัลเลขท้ายสองตัว

วิธีทำ กำหนดให้เหตุการณ์ A เป็นหลักสิบ และเหตุการณ์ B เป็นหลักหน่วย เหตุการณ์ A และเหตุการณ์ B เป็นอิสระต่อกัน ดังนั้น $P(\mathrm{A})$ คือ $\frac{1}{10}$ และ $P(\mathrm{B})$ คือ $\frac{1}{10}$

ความน่าจะเป็นของการออกรางวัล 65 คือ

$$P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{A}) \cdot P(\mathrm{B})$$

$$= \left(\frac{1}{10}\right)\left(\frac{1}{10}\right)$$

$$= \frac{1}{100}$$

ความน่าจะเป็นของการถูกรางวัลเลขท้ายสองตัวกรณีนี้เท่ากับ $\frac{1}{100}$

3) เหตุการณ์ส่วนเติมเต็ม จากภาพที่ 5.11 (ก) R คือ $\{A_1, A_2, A_3, A_5\}$ ส่วนเติมเต็มของ R คือ $\{A_4, A_6, A_7, A_8\}$ ใช้สัญลักษณ์ $\overline{R}$ นั่นคือ

$$\overline{R} = \{A_4, A_6, A_7, A_8\}$$

จึงนิยาม เหตุการณ์ส่วนเติมเต็ม ในรูปของเซตว่า

$$\overline{R} = \text{เซตของสมาชิก } S \text{ ที่ไม่อยู่ใน R} \qquad \ldots(5\text{-}18)$$

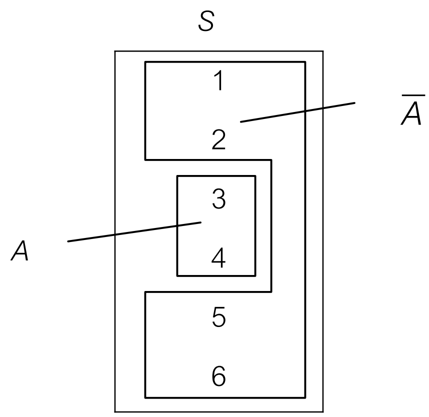
$\overline{R}$ อ่านว่า ส่วนเติมเต็มของ R (complement of R or not R) จะเห็นได้ว่า จำนวนสมาชิกของ $\overline{R}$ รวมกับสมาชิกของ R จะเท่ากับสมาชิกของ S ดังนั้น ความน่าจะเป็นของ $\overline{R}$ รวมกับความน่าจะเป็นของ R

คือ $$P(\overline{R}) + P(\text{R}) = 1 \qquad \ldots(5\text{-}19)$$

หรือ $$P(\overline{R}) = 1 - P(\text{R}) \qquad \ldots(5\text{-}20)$$

ตัวอย่างที่ 5.17 ถ้า S คือ {1, 2, 3, 4, 5 , 6} และ A คือ {3 ,4 } จงหา $\overline{A}$ และ $P(\overline{A})$

วิธีทำ



ภาพที่ 5.12 แผนภาพเวนน์ของเซต A และเซต $\overline{A}$

จากภาพที่ 5.12 จะได้ $\overline{A} = \{1, 2, 5, 6\}$

$$P(\overline{A}) = 1 - P(\text{A})$$

$$= 1 - \frac{2}{6}$$

$$= \frac{4}{6} = \frac{2}{3}$$

$\overline{A}$ และ $P(\overline{A})$ มีค่าเป็น {1, 2, 5, 6 } และ $\frac{2}{3}$

5.5 ค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงความน่าจะเป็น

จากตารางที่ 5.2 ถ้าเหตุการณ์ที่เรากำลังสนใจอยู่คือ จำนวนสินค้าขายได้ (G) เมื่อมีสินค้า 3 ชิ้น โดยกำหนดให้ X เป็นตัวแปรแสดงจำนวนสินค้าที่ขายได้ (G)

X จะเป็นตัวแปรสุ่มและพบว่า X อาจมีค่าเป็น 0 , 1 , 2 , 3 แต่อย่างไรก็ดีเหตุการณ์จำนวนสินค้าที่ขายได้ 0 ชิ้น (ขายไม่ได้) ขายได้ 1 ชิ้น ขายได้ 2 ชิ้น ขายได้ 3 ชิ้น มีความน่าจะเป็นไม่เท่ากัน แต่ยังคงอยู่ในเซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด S เราสามารถสร้างเซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดให้เล็กลง พร้อมทั้งแสดงความน่าจะเป็น X ซึ่งเป็นเหตุการณ์ที่เราสนใจอยู่ ดังตารางที่ 5.4 เมื่อนำความน่าจะเป็น X หรือ $P(X)$ มาสัมพันธ์กับค่า X เราเรียกว่า การแจกแจงความน่าจะเป็น (probability distribution) ดังภาพที่ 5.13

ตารางที่ 5.4 การสร้างเซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดใหม่ตามค่าตัวแปรสุ่ม X ที่กำหนดเป็นจำนวนสินค้าที่ขายได้

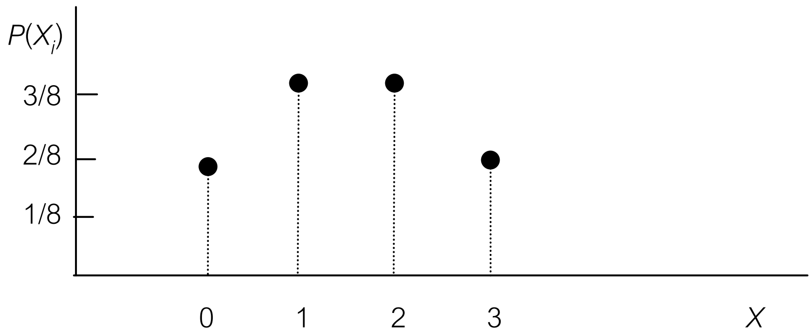
เซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดเดิม

เหตุการณ์ A	$P(A)$
GGG	$\frac{1}{8}$
GGD	$\frac{1}{8}$
GDG	$\frac{1}{8}$
GDD	$\frac{1}{8}$
DGG	$\frac{1}{8}$
DGD	$\frac{1}{8}$
DDG	$\frac{1}{8}$
DDD	$\frac{1}{8}$
	1

$\Rightarrow$

เซตผลลัพธ์ที่พึงเป็นไปได้ทั้งหมดใหม่

X	$P(X)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
	1



ภาพที่ 5.13 ความสัมพันธ์ระหว่าง การแจกแจงความน่าจะเป็นของ X กับค่า X ตามข้อมูลในตารางที่ 5.4

จากความรู้บทที่ 3 และบทที่ 4 เราหาค่าเฉลี่ยและความแปรปรวนของตัวแปรด้วยความถี่สัมพัทธ์ $(\frac{f_i}{n})$หรือ $rf\ (X_i)$ ในทำนองเดียวกันเราสามารถหาค่าเฉลี่ยและค่าความแปรปรวนของตัวแปรสุ่ม X และความน่าจะเป็น $P(X_i)$ได้

ตารางที่ 5.5 การเปรียบเทียบสมการการหาค่าเฉลี่ยและค่าความแปรปรวนของประชากร

	$rf\ (X_i)$	$P(X_i)$
μ	$\sum_{i=1}^{n} X_i . rf\ (X_i)$	$\sum_{i=1}^{n} X_i P(X_i)$
σ_x^2	$\sum_{i=1}^{n} X_i^2\ rf\ (X_i) - \mu^2$ $\sum_{i=1}^{n} (X_i - \mu)^2\ rf\ (X_i)$	$\sum_{i=1}^{n} X_i^2\ P(X_i) - \mu^2$ $\sum_{i=1}^{n} (X_i - \mu)^2\ P(X_i)$

จึงกำหนดว่า $X_1, X_2, X_3, \ldots, X_n$ เป็นตัวแปรสุ่ม มีความน่าจะเป็น $P(X_1), P(X_2), P(X_3), \ldots, P(X_n)$ ตามลำดับ

จะได้ $$\mu = \sum_{i=1}^{n} X_i P(X_i) \qquad ...(5\text{-}21)$$

เปรียบเทียบกับสมการ (2-14)

และ $$\sigma_x^2 = \sum_{i=1}^{n} (X_i - \mu)^2 P(X_i) \qquad ...(5\text{-}22)$$

เปรียบเทียบกับสมการ (3-16)

หรือ $$\sigma_x^2 = \sum_{i=1}^{n} X_i^2 \, rf(X_i) - \mu^2 \qquad ...(5\text{-}23)$$

เปรียบเทียบกับสมการ (3-17) ดังตารางที่ 5.5

สำหรับค่าเฉลี่ยตัวแปรสุ่ม X เช่นแต้มเฉลี่ยลูกเต๋าเมื่อบันทึกผลจากการทอดหลาย ๆ ครั้ง ผลเฉลี่ยเลขคณิตเลขท้ายสองตัวที่ออกมาเวลายาวนานหรือค่าเฉลี่ยผลการสุ่มเป็นเวลานาน ๆ จะเรียกว่า ค่าคาดหมายของ X หรือ $E(X)$ ทำให้สมการ (5-21) เขียนได้เป็นสมการ (5-24)

$$E(X) = \sum_{i=1}^{n} X_i P(X_i) \qquad ...(5\text{-}24)$$

ทำให้คุณสมบัติค่าเฉลี่ยเลขคณิตตามสมการ (2-15) ถึงสมการ (2-19) เป็นคุณสมบัติของค่าคาดหมาย (properties of expectation value) กล่าวคือ ถ้า a , b , c เป็นค่าคงตัวใด ๆ ได้สมการ (5-25) ถึงสมการ (5-27)

$$E(c) = c \qquad ...(5\text{-}25)$$

$$E(cX) = cE(X) \qquad ...(5\text{-}26)$$

$$E(cX + b) = cE(X) + b \qquad ...(5\text{-}27)$$

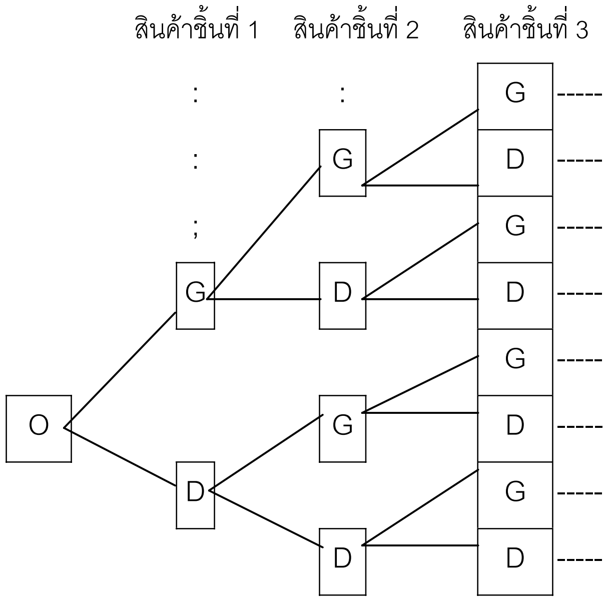
สำหรับความแปรปรวนตัวแปรสุ่ม X จะใช้สัญลักษณ์ $Var(X)$ และแทน σ_x^2 ในสมการ (5.23) ได้เป็น

$$Var(X) = \sum_{i=1}^{n} X_i^2 P(X_i) - [E(X)]^2 \qquad ...(5\text{-}28)$$

5.6 การแจกแจงทวินาม

จากเหตุการณ์การขายสินค้า 1 ชิ้น จะขายได้หรือขายไม่ได้นั้น เมื่อเราสนใจถึงการขายสินค้าได้ ถ้าขายได้ถือว่า สำเร็จ (success) และการขายไม่ได้ถือว่า ไม่สำเร็จ (failure) หรือใช่กับไม่ใช่ สถานการณ์เช่นนี้เรียกว่า สถานการณ์มีการแจกแจงแบบแบร์นูลลี (Bernoulli distribution or print binomial) ภายใต้ข้อตกลงเบื้องต้นว่าต้องทราบความน่าจะเป็นหรือโอกาส ที่จะสำเร็จและความน่าจะเป็นหรือโอกาสที่จะไม่สำเร็จคือ 1 - p ใช้สัญลักษณ์ q ตามตัวอย่างที่หัวข้อ 5.5 ให้โอกาสที่สินค้าแต่ละชิ้นจะขายได้ คือ $p = \frac{1}{2}$ หรือ 0.5 สินค้า 1 ชิ้น ถือว่า $n = 1$

ในกรณีที่วางขายสินค้า 3 ชิ้น และกำหนดให้ X คือ เลขแสดงจำนวนสินค้าที่ขายได้ ค่าของ X จึงเป็น 0, 1, 2 หรือ 3 เรียกว่า X เป็นตัวแปรสุ่มที่แจกแจงทวินาม (binomial distribution) หรือ X เป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นแบบทวินาม หากวางขายสินค้า n ชิ้น ค่า x จะเท่ากับ 0, 1, 2,….,n ซึ่งเป็นตัวแปรวิยุต



S	*P(X=x)*	*ค่า x*
G G G	p^3	3
G G D	p^2q	2
G D G	p^2q	2
G D D	pq^2	1
D G G	pq^2	2
D G D	p^2q	2
D D G	pq^2	1
D D D	q^3	0

ภาพที่ 5.14 ความน่าจะเป็นในการขายสินค้า 3 ชิ้น

ที่มา : (ดัดแปลงมาจาก Keller & Warrack, 2000, p.211)

เมื่อแสดงให้อยู่ในรูปความน่าจะเป็นสินค้าขายได้คือ p และความน่าจะเป็นสินค้าขายไม่ได้ q จะพบเหตุการณ์ของความน่าจะเป็น X ดังแสดงในภาพที่ 5.12 เมื่อ $X = 0$ ความน่าจะเป็น คือ q^3 เขียนแทนด้วย $P(0) = P(X = 0) = q^3$ เมื่อ $X = 1$ ความน่าจะเป็น คือ $3pq^2$ เขียนแทนด้วย $P(1) = P(X = 1) = 3pq^2$ เมื่อ $X = 2$ ความน่าจะเป็น คือ $3p^2q$ เขียนแทนด้วย $P(2) = P(X = 2) = 3p^2q$ และเมื่อ $X = 3$ ความน่าจะเป็น คือ $P(3) = P(X = 3) = p^3$ ดังตารางที่ 5.6

ตารางที่ 5.6 การแจกแจงทวินามของ X เมื่อ $n = 3$

X	$P(X=x)$
0	q^3
1	$3pq^2$
2	$3p^2q$
3	p^3

พิจารณาค่าสัมประสิทธิ์ของ $P(x)$ เป็น 1, 3, 3 และ 1 เขียนให้อยู่ในรูปของจำนวนวิธีการจัดหมู่

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

โดย $n! = n(n-1)(n-2)\ldots(2)(1)$ และ $0!$ เท่ากับ 1 ดังตารางที่ 5.7 เป็นการแสดงการคำนวณสัมประสิทธิ์ความน่าจะเป็นของการแจกแจงทวินามเมื่อ X มีค่าแตกต่างกัน

ดังนั้น จึงเขียนรูปทั่วไปของการแจกแจง X ที่มีการแจกแจงทวินาม หรือ $X \sim \mathrm{B}(n,p)$ คือ

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad \text{โดยที่ } x = 0, 1, 2, \ldots, n \qquad \ldots(5\text{-}29)$$

ตารางที่ 5.7 สัมประสิทธิ์ของความน่าจะเป็นของการแจกแจงทวินามของ X

X	สัมประสิทธิ์ความน่าจะเป็น $\binom{n}{x}$
0	$\binom{3}{0} = \frac{3!}{0!(3-0)!} = \frac{3.2.1}{1.(3.2.1)} = 1$
1	$\binom{3}{1} = \frac{3!}{1!(3-1)!} = \frac{3.2.1}{1.(2.1)} = 3$
2	$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3.2.1}{2.(1)} = 3$
3	$\binom{3}{3} = \frac{3!}{3!(3-3)!} = \frac{3.2.1}{3.2.1(1)} = 1$

จากภาพที่ 5.14 และตารางที่ 5.6 คำนวณหาค่าความน่าจะเป็นของ X เมื่อ $p = 0.5$ ได้เป็น

$$P(X=0) = \frac{3!}{0!3!}(0.5)^0(0.5)^3 = 0.1250$$

$$P(X=1) = \frac{3!}{1!3!}(0.5)^1(0.5)^2 = 0.3750$$

$$P(X=2) = \frac{3!}{2!1!}(0.5)^2(0.5)^1 = 0.3750$$

$$P(X=3) = \frac{3!}{3!0!}(0.5)^3(0.5)^0 = 0.1250$$

และเขียนได้ดังตารางที่ 5.8 และความน่าจะเป็นสะสมของ X ที่แจกแจงทวินาม เมื่อ $n = 3$ และ $p = 0.5$ แสดงได้ดังตารางที่ 5.9

เมื่อต้องการหาความน่าจะเป็นสะสมทุกค่าตั้งแต่ $x = 0, 1, 2, 3$ เขียนแทนด้วยสมการ (5-30)

$$P(X \leq 3) = \sum_{x=0}^{3} P(X) = P(X=0) + P(X=1) + P(X=2) + P(X=3) \quad \ldots (5\text{-}30)$$

ตารางที่ 5.8 การแจกแจงความน่าจะเป็นทวินามของ X เมื่อ $p = 0.5$, $n = 3$

X	$P(X)$
0	0.1250
1	0.3750
2	0.3750
3	0.1250
	1.0000

ตารางที่ 5.9 ความน่าจะเป็นสะสมของ X ที่มีการแจกแจงทวินาม เมื่อ $p = 0.5$, $n = 3$

X	$P(X \leq x)$
0	0.1250
1	0.5000
2	0.8750
3	1.0000

เมื่อต้องการหาค่าเฉลี่ย X ใช้สมการ (5-24) และหาความแปรปรวน X ตามสมการ (5-28) จะได้ (Hogg, 1989, pp. 116-117)

$$E(X) = \sum_{x=0}^{n} \frac{xn!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \frac{xn(n-1)!}{x(x-1)!(n-x-1)!} pp^{x-1} q^{n-x-1}$$

$$= np \sum_{x=0}^{n-1} \frac{(n-1)!}{(x-1)!(n-x-1)!} p^{x-1} q^{n-x-1}$$

$$E(X) = np \quad \ldots (5\text{-}31)$$

และจาก $E[X(X\text{-}1)] \;=\; \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}\,p^{x}q^{n-x}$

สองเทอมแรกเมื่อ $x = 0$ และ 1 ในสมการเป็น $(0)(0\text{-}1)P(X{=}0) = 0$ และ $(1)(1\text{-}1)\,P(X{=}1) = 0$ จะได้

$$E[X(X\text{-}1)] \;=\; \sum_{x=2}^{n}\frac{n!}{(x-2)!(n-x)!}\,p^{x}q^{n-x}$$

แต่ $\frac{x(x-1)!}{x!} = \frac{1}{(x-2)!}$ เมื่อ $x > 1$ และกำหนดให้ $k = x-2$

$$E[X(X\text{-}1)] \;=\; \sum_{k=0}^{n-2}\frac{n!}{k!(n-k-2)!}\,p^{k+2}q^{n-k-2}$$

$$= \; n(n\text{-}1)p^{2}\sum_{k=0}^{n-2}\frac{(n-2)!}{k!(n-2-k)!}\,p^{k}q^{n-2-k}$$

$$= \; n(n\text{-}1)p^{2}$$

เมื่อ $\quad E[X(X\text{-}1)] \;=\; E(X^{2}) - E(X)$

และ $\quad Var(X) \;=\; E(X^{2}) - [E(X)]^{2} \;=\; E(X^{2}) - E(X) + E(X) - [E(X)]^{2}$

$$= \; n(n\text{-}1)p^{2} + np - (np)^{2}$$

$$= \; -np^{2} + np \;=\; np(1\text{-}p)$$

$$Var(X) \;=\; npq \qquad \ldots (5\text{-}32)$$

สำหรับหาความน่าจะเป็นสะสมที่ค่า p = 0.10, 0.20, 0.25,…,0.90 เมื่อ n = 2, 3, 4, 5, 6, 7, 8, 9, 10,15 และ 20 จากตารางภาคผนวกที่ 3 เมื่อต้องการหาความน่าจะเป็นสะสม เมื่อ $X \leq 2$ กรณี p = 0.25 และ n = 5 หรือ $P(X \leq 3) = P(X{=}0) + P(X{=}1) + P(X{=}2) + P(X{=}3)$ ดังนั้นจากตารางที่ 5.10 จะได้ว่า $P(X \leq 3)$ = 0.7759

ตารางที่ 5.10 การหาค่า $P(X \leq 3)$ เมื่อ $p = 0.25$ และ $n = 5$

n	*x*	0.25
5	0	.
	1	.
	2	$\Downarrow$
	3	………………..$\Rightarrow$ 0.7759
	4	
	5	

ตัวอย่างที่ 5.18 ผลวิจัยทราบว่าร้อยละ 40 ของนักศึกษาสถาบันแห่งหนึ่งใช้โทรศัพท์มือถือ หากสุ่มนักศึกษาแห่งนี้จำนวน 5 คน จงหาโอกาสพบ

ก. นักศึกษา 2 คนใช้โทรศัพท์มือถือ

ข. นักศึกษาไม่ถึง 4 คนใช้โทรศัพท์มือถือ

ค. นักศึกษาอย่างน้อย 1 คนใช้โทรศัพท์มือถือ

วิธีทำ

ก. $P(X = 2) = \binom{5}{2}(0.4)^2(0.6)^3 = 0.3456$

โอกาสพบนักศึกษา 2 คนใช้โทรศัพท์มือถือมีค่าเท่ากับ 0.3456

ข. $P(X < 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$$= \binom{5}{0}(0.4)^0(0.6)^5 + \binom{5}{1}(0.4)^1(0.6)^4 + \binom{5}{2}(0.4)^2(0.6)^3$$

$$+ \binom{5}{3}(0.4)^3(0.6)^2$$

$P(X < 4) \quad = \quad 0.9130$

โอกาสพบนักศึกษาไม่ถึง 4 คนใช้โทรศัพท์มือถือมีค่าเท่ากับ 0.9130

ค. $P(X \geq 1) \quad = \quad P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$$= \quad 1 - P(X=0) \qquad = \quad 1 - \binom{5}{0}(0.4)^0(0.6)^5$$

$$= \quad 0.9224$$

โอกาสพบนักศึกษาอย่างน้อย 1 คนใช้โทรศัพท์มือถือมีค่าเท่ากับ 0.9224

ตัวอย่างที่ 5.19 จงใช้ตารางภาคผนวกที่ 3 หาค่า $P(X \leq 2)$ เมื่อ $p = 0.30$ และ $n = 10$

วิธีทำ จากตารางภาคผนวกที่ 3 ในส่วนที่ $n = 2$ แสดงได้ดังตารางที่ 5.11

ตารางที่ 5.11 การหาค่า $P(X \leq 2)$ เมื่อ $p = 0.30$ และ $n = 10$

n	*x*	0.30
10	0	.
	1	$\Downarrow$
	2	……………….. $\Rightarrow$ 0.3828
	.	
	.	
	10	

นั่นคือ $P(X \leq 2) \quad = P(X=0) + P(X=1) + P(X=2) \quad = \quad 0.3828$

ค่า $P(X \leq 2)$ เท่ากับ 0.3828

5.7 การแจกแจงปัวซง

ถ้าค่า X เป็นการนับสิ่งของหรือจำนวนเรื่องราวที่เกิดขึ้นในช่วงเวลาหนึ่ง กรอบหนึ่งหรือผิวพื้นที่หนึ่ง เช่น การเฝ้านับจำนวนอุบัติเหตุที่สี่แยกในรอบ 1 วัน จำนวนครั้งของโทรศัพท์ที่เรียกเข้าในช่วงเวลาครั้งหนึ่งของวัน โดย x = 0,1,2,… การแจกแจงปัวซง (Poisson distribution)สามารถเขียนความน่าจะเป็น X เมื่อ $X = x$ ค่าหนึ่ง ได้เป็น (Keller & Warrack , 2000, pp. 219-220)

$$P(X = x) \;=\; \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \text{เมื่อ } x = 0,1,2,\ldots \text{และ } \lambda > 0 \qquad \ldots(5\text{-}33)$$

และ λ คือ จำนวนครั้งเฉลี่ยของความสำเร็จหรือใช่ และ e = 2.71828 ซึ่งเป็นของฐานลอกาลิทึมธรรมชาติ (logarithms base of natural) ของค่าเฉลี่ยของ X และความแปรปรวนของ X อาศัยหลักเกณฑ์จากสมการ (5-24)และสมการ(5-28)ได้ค่าเท่ากัน คือ ถ้า X แจกแจงปัวซง หรือ $X \sim$ Poi (λ) จะได้ (Hogg, 1989, p. 133)

$$E(X) \;=\; \sum_{x=0}^{\infty} x\frac{\lambda^{x}e^{-\lambda}}{x!} \;=\; e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x}}{(x-1)!}$$

เนื่องจากเทอมแรก $P(X=0) = 0$ และ $\dfrac{x}{x!} = \dfrac{1}{(x-1)!}$

เมื่อ $x > 0$ กำหนดให้ $k = x-1$ จะได้

$$E(X) \;=\; e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{k+1}}{k!} \;=\; \lambda\, e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{k}}{k!}$$

$$= \lambda\, e^{-\lambda}e^{\lambda} \;=\; \lambda$$

สำหรับ $$E[X(X-1)] \;=\; \sum_{x=0}^{\infty} x(x-1)\frac{\lambda^{x}e^{-\lambda}}{x!} \;=\; e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x}}{(x-2)!}$$

สองเทอมแรกเป็น $(0)(0-1)P(X=0) = 0$, $(1)(1-1)P(X=0) = 0$ และ $\dfrac{x(x-1)}{x!} = \dfrac{1}{(x-2)!}$

เมื่อ $x > 1$ กำหนดให้ $k = x-2$ จะได้

$$E[X(X\text{-}1)] = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{k+2}}{k!} \qquad = \lambda^2 e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{k}}{k!}$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} \qquad = \lambda^2$$

เมื่อ $E[X(X\text{-}1)] = E(X^2) - E(X)$

และ $Var(X) = E(X^2) - [E(X)]^2 \qquad = E(X^2) - E(X) + E(X) - [E(X)]^2$

$$= \lambda^2 + \lambda - \lambda^2 \qquad = \lambda$$

ดังนั้น $E(X) = Var(X) \qquad = \lambda$ ...(5-34)

ตัวอย่างที่ 5.20 จากบันทึกของเจ้าหน้าที่รักษาความปลอดภัยโรงงานแห่งหนึ่ง พบว่ามีพนักงานเดินผ่านประตูเล็กด้านหลังโรงงานช่วงพักกลางวัน 360 คนในเวลา 1 ชั่งโมง

ก. จงใช้ความน่าจะเป็นแบบปัวซงหาความน่าจะเป็นพนักงานจะผ่านประตูนี้ 2 คน ในเวลา 1 นาทีช่วงพักกลางวัน

ข. จงใช้ตารางภาคผนวกที่ 4 หาค่าความน่าจะเป็นตามข้อ ก.

ค. จงหาค่าความน่าจะเป็นที่พนักงานจะผ่านประตูเล็กแห่งนี้อย่างน้อย 4 คน ในเวลา 1 นาทีช่วงพักกลางวัน

วิธีทำ หาค่า $\lambda = \frac{360}{6} \qquad = 6$ คน / นาที

ก. $P(X = 2) = \frac{e^{-6}6^2}{2!} \qquad = \frac{(0.00248)(36)}{(2)(1)}$

$$= 0.0446$$

ความน่าจะเป็นที่พนักงานจะผ่านประตู 2 คนในเวลา 1 นาที คือ 0.0446

ข. จากตารางภาคผนวกที่ 4 หาค่า $P(X \leq 2)$ ได้ดังตารางที่ 5.12

ตารางที่ 5.12 การหาค่า $P(X \leq 2)$

x	λ			
	4.2	4.4	. . .	6.0
0				.
1				$\Downarrow$
2	………………………………………$\Rightarrow$			0.062
.				

$P(X = 2) \quad = \quad P(X \leq 2) - P(X \leq 1) \quad = \quad 0.062 - 0.017 \quad = \quad 0.045$

เมื่อใช้ตารางหาค่าความน่าจะเป็นตามข้อ ก. จะได้เท่ากับ 0.045

ค. จากตารางภาคผนวกที่ 4 เมื่อ $P(X \leq 4)$ ได้ดังตารางที่ 5.13

ตารางที่ 5.13 การหาค่า $P(X \leq 4)$

x	4.2	. . . . .	6.0
0			.
1			.
2			.
3			$\Downarrow$
4	………………………………$\Rightarrow$		0.285

$P(X \leq 4) \quad = \quad 0.285$

ความน่าจะเป็นที่พนักงานจะผ่านประตูเล็กแห่งนี้อย่างน้อย 4 คน คือ 0.285

ถ้า n มีขนาดโตที่จะสามารถใช้การแจกแจงทวินามแทนการแจกแจงปัวซงได้ ดังตารางที่ 5.14 (Keller & Warrack, 2000, p. 223) เมื่อ $p \leq .05$ และ $\lambda = np$

ตารางที่ 5.14 การเปรียบเทียบความน่าจะเป็นของการแจกแจงทวินามและการแจกแจงปัวซง ของ X ที่ $p = 0.02$ และ $n = 50$

x	ความน่าจะเป็น x แจกแจงทวินาม ($n = 50$, $p = .02$)	ความน่าจะเป็น x แจกแจงปัวซง ($\lambda = np = 1$)
0	0.364	0.368
1	0.372	0.368
2	0.186	0.184
3	0.061	0.061
4	0.014	0.015
5	0.003	0.003
6	0.001	0.001

5.8 การแจกแจงไฮเพอร์จีออเมตริก

ในกรณีที่มีสินค้าทั้งหมดจำนวน N ชิ้น ทราบว่ามีสินค้าชำรุดหากผู้ซื้อพบจะไม่ขายจำนวน k ชิ้น ดังนั้นสินค้าที่เหลือที่จะนำไปขายได้มีจำนวน $N - k$ ชิ้น เมื่อนำสินค้าออกเสนอขายครั้งละ n ชิ้นซึ่งยังประกอบด้วยสินค้าชำรุดปนอยู่ ให้ $X = x$ เป็นจำนวนสินค้าที่ขายได้ และ $n - x$ เป็นจำนวนสินค้าที่ขายไม่ได้ ดังนั้นค่าของ X จึงมีค่าเป็น $0,1,2,3,\ldots\ n$ และเราเรียก X นี้ว่า การแจกแจงไฮเพอร์จีออเมตริก (Hypergeometric distribution) หรือ $X \sim Hyp(N,k,n)$ เขียนความน่าจะเป็น X เมื่อ $X = x$ ค่าหนึ่ง ได้เป็น

$$P(X{=}x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad ; \quad N,k,n \geq 0 \text{ และ } k-(N-n) \leq x \leq n \quad \ldots(5\text{-}35)$$

ค่าเฉลี่ยของ X และความแปรปรวนของ X อาศัยหลักเกณฑ์จากสมการ (5-24) และสมการ (5-28) จะได้ (Hogg,1989, pp.83-84)

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = n\frac{k}{N}\sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N-1}{n-1}}$$

ให้ $y = x$-1 จะได้

$$E(X) = n\frac{k}{N}\sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{N-1-k+1}{n-1-y}}{\binom{N-1}{n-1}} = n\frac{k}{N} \qquad \ldots(5\text{-}36)$$

จาก $$E[X(X-1)] = \sum_{x=0}^{n} x(x-1) \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$E[X(X-1)] = n(n-1)\frac{k(k-1)}{N(N-1)} \sum_{x=2}^{n} \frac{\binom{k-2}{x-2}\binom{N-k}{n-x}}{\binom{N-2}{n-2}}$$

ให้ $y = x$-2 จะได้

$$E[X(X-1)] = n(n-1)\frac{k(k-1)}{N(N-1)} \sum_{y=0}^{n-2} \frac{\binom{k-2}{y}\binom{N-2-k+2}{n-2-y}}{\binom{N-2}{n-2}}$$

$$= n(n-1)\frac{k(k-1)}{N(N-1)}$$

เมื่อ $E[X(X-1)] \;=\; E(X^2) - E(X)$

และ $Var(X) \;=\; E(X^2) - [E(X)]^2 \;=\; E(X^2) - E(X) + E(X) - [E(X)]^2$

$$= n(n-1)\frac{k(k-1)}{N(N-1)} + \frac{nk}{N} - \left(\frac{nk}{N}\right)^2$$

$$Var(X) = \frac{nk}{N}\left[\frac{(N-k)(N-n)}{N(N-1)}\right] \qquad \text{...(5-37)}$$

ตัวอย่างที่ 5.21 ร้านขายผ้าไหมแห่งหนึ่งมีผ้าไหมอยู่จำนวน 24 ชิ้น โดยผู้ขายทราบว่ามีผ้าไหมที่ชำรุด 18 ชิ้น ถ้าพบว่าผ้าชิ้นไหนชำรุดผู้ขายจะไม่ขายให้กับลูกค้า หากนำผ้าไหมออกเสนอขายครั้งละ 3 ชิ้น จงหาโอกาสที่จะขายผ้าไหมได้ทั้ง 3 ชิ้น

วิธีทำ ให้ X แทนจำนวนผ้าไหมที่ขายได้ ดังนั้น X เป็นการแจกแจงไฮเพอร์จีออเมตริก และมีค่าเป็น 0, 1, 2, 3

$N = 24$, $k = 18$, $N-k = 24-18 = 6$ และ $n = 3$

จากสมการ (5-35) จะได้

$$P(X=x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{18}{3}\binom{24-18}{3-3}}{\binom{24}{3}}$$

$$= \frac{(816)(1)}{(2{,}024)} = 0.403$$

โอกาสที่จะขายผ้าไหมได้ทั้ง 3 ชิ้นเท่ากับ 0.403

5.9 บทสรุป

ทฤษฎีความน่าจะเป็น นับว่ามีความสำคัญสำหรับการศึกษาสถิติอนุมานเป็นอย่างยิ่ง เนื้อหาทฤษฎีความน่าจะเป็นที่นำมากล่าวในบทนี้ มีจุดประสงค์เพื่อนำไปสู่การศึกษาเกี่ยวกับการแจกแจงตัวอย่างและการแจกแจงปรกติของตัวอย่างโดยผู้ศึกษามีพื้นความรู้เรื่องเซตอยู่ก่อนแล้ว ซึ่งสรุปได้ดังนี้

5.9.1 ความน่าจะเป็นเป็นการเปรียบเทียบจำนวนหนทางที่เกิดเหตุการณ์ กับจำนวนเหตุการณ์ที่เกิดขึ้นทั้งหมด ซึ่งในการหาผลจำนวนผลลัพธ์ที่พึงเป็นได้ทั้งหมด มีวิธีการ 2 วิธีคือ การทดลองและการสำรวจ และ การใช้หลักเกณฑ์ทางทฤษฎี

5.9.2 ในการเขียนต้นแบบทางคณิตศาสตร์มีตัวแบบหลักสำหรับการหาผลลัพธ์ที่เป็นไปได้ 3 ลักษณะคือ

1) แผนภูมิต้นไม้
2) วิธีเรียงสับเปลี่ยน
3) วิธีจัดหมู่

5.9.3 ในการคำนวณค่าความน่าจะเป็นหรือโอกาสของการเกิดเหตุการณ์นั้นมีวิธีการคิดดังนี้

1) ความน่าจะเป็นของเหตุการณ์เดี่ยว
2) ความน่าจะเป็นของเหตุการณ์เชิงประกอบ
3) ความน่าจะเป็นของเหตุการณ์

5.9.4 ในการแจกแจงความน่าจะเป็นจะต้องคำนึงถึง

1) ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่ม
2) คุณสมบัติของค่าคาดคะเน

5.95 การแจกแจงตัวแปรวิยุต

1) การแจกแจงทวินาม
2) การแจกแจงแบบปัวซง
3) การแจกแจงไฮเปอร์จีออเมตริก

5.10 คำถามทบทวน

1. จงแสดงผลลัพธ์ที่พึงเป็นไปได้ทั้งหมด เมื่อสำรวจครอบครัวที่มีบุตร 3 คน กำหนด M คือบุตรเพศชาย W คือบุตรเพศหญิง

2. การเดินทางไป – กลับระหว่างบ้านของนายมอสกับสถาบันราชภัฏเทพสตรี มีรถวิ่งผ่าน 3 สายคือสาย ก. สาย ข. และสาย ค. จงหา

   ก. จำนวนวิธีเดินทางไป – กลับทั้งหมด

   ข. ถ้าการเดินทางทั้งไปและกลับด้วยรถที่ไม่ซ้ำสายกัน จะมีวิธีเดินทางไป กลับได้กี่ได้วิธี

   ค. ถ้าการเดินทางเป็นไปตาม ข. กี่วันจึงจะมีวิธีเดินได้ครบ เมื่อเดินทางไปกลับทุกวัน

3. ข้อสอบแบบปรนัยจำนวน 10 ข้อ แต่ละข้อมีตัวเลือก 4 ตัว และมีตัวเลือกถูกเพียง 1 ตัว ถ้านายเอกทำข้อสอบชุดนี้ด้วยวิธีเดาคำตอบ จงหา

   ก. โอกาสที่เขาจะตอบถูก 3 ข้อ

   ข. โอกาสที่เขาจะตอบถูก 8 ข้อขึ้นไป

   ค. โอกาสที่เขาจะตอบถูก 4 ถึง 6 ข้อ

4. นายโทต้องการซื้อรถยนต์ เขามีทางเลือกเครื่องยนต์ 3 ชนิด ตัวถัง 7 แบบและสีรถ 7 สี จงแสดงแผนภูมิต้นไม้เกี่ยวกับลักษณะรถทั้งหมดที่เขามีโอกาสเลือก

5. จากคำถามข้อ 5.1 จงหา

   ก. โอกาสที่จะพบบุตรในครอบครัวมีเพศเดียวกันหมด

   ข. โอกาสที่จะพบครอบครัวที่มีบุตรคนแรกเป็นเพศชาย

   ง.โอกาสที่จะพบครอบครัวที่มีบุตรคนที่ 3 เป็นเพศชาย

6. จงหาค่า

   ก. $\binom{5}{2}$ ข. $\binom{7}{7}$

   ค. $\binom{7}{6}$ ง. $\binom{6}{4}$

   ซ. $\binom{6}{2}$

7. กรณี X แจกแจงทวินาม $p = 0.6$, $n = 5$ จงหาค่าต่อไปนี้

ก. $P(X \leq 2)$

ข. ค่าเฉลี่ย X และความแปรปรวน X

8. ถ้าการแจกแจงทวินามด้วย $p = 0.30$ และ $n = 20$ จงหาค่าต่อไปนี้

ก. $P(X \geq 2)$

ข. $P(X \leq 2)$

ค. $P(X = 2)$

ง. ค่าเฉลี่ย X และความแปรปรวน X

9. ถ้า X มีการแจกแจงปัวซง ด้วย $\lambda = 4$ จงหาค่าต่อไปนี้

ก. $P(2 \leq X \leq 5)$

ข. $P(X \geq 3)$

ง. $P(X \leq 3)$

10. ถ้า X เป็นการแจกแจงไฮเพอร์จีออเมตริกแล้ว $N = 50$, $k = 3$ และ $n = 5$ จงหา

ก. $P(X = 1)$

ข. $P(X \geq 2)$

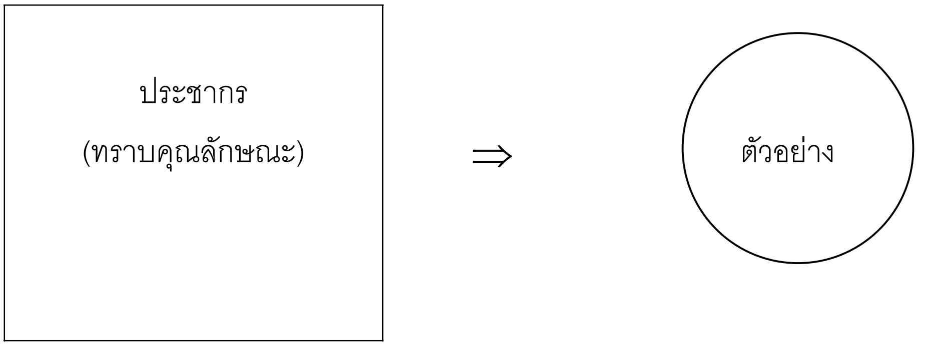
บทที่ 6
การชักตัวอย่างและการแจกแจงตัวอย่าง

จากที่กล่าวมาแล้วในบทที่ 2 ถึงบทที่ 4 เป็นการกล่าวถึงข้อมูลหรือค่าสังเกตซึ่งเป็นประชากร ตามแนวทางคณิตศาสตร์ การสรุปเกี่ยวกับคุณลักษณะของประชากรจึงเป็นการสรุปจากหลักเกณฑ์ การนิรนัย (deduction) หรือ การพรรณนา (description) สำหรับบทนี้จะกล่าวถึง กระบวนการสรุป คุณลักษณะของประชากรด้วยประสบการณ์และอาศัยทฤษฎีความน่าจะเป็นจากบทที่ 5 ในรูปการอุปนัย (induction) หรือการอนุมาน (inference)

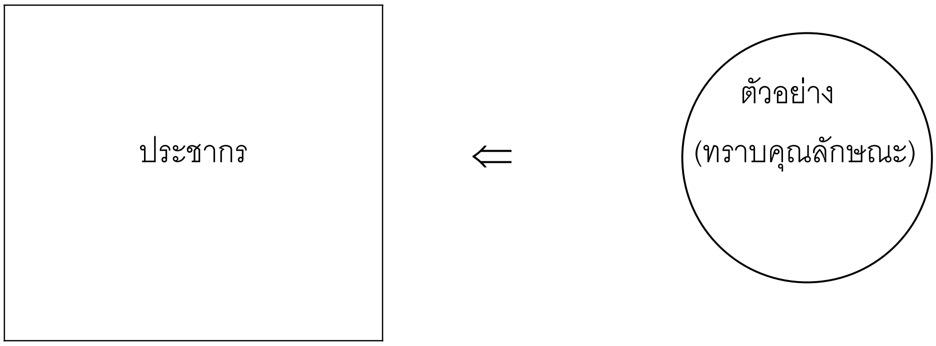
6.1 การอ้างสถิติแบบฉบับ

ขอบเขตของประชากรที่กว้างใหญ่ไพศาล การศึกษาเพื่อให้ทราบถึงคุณลักษณะของประชากรทั้งหมดทำได้ค่อนข้างยากและลำบากในการเก็บข้อมูล มีค่าใช้จ่ายสูงมาก เสียเวลามาก ใช้บุคลากรมาก ตลอดจนความผิดพลาดย่อมเกิดขึ้นมากในกระบวนการวัดได้ ดังนั้นมีการใช้ตัวอย่างซึ่งถูกเลือกมาจากประชากรโดยยึดหลักเกณฑ์ที่ว่าจะต้องให้รายละเอียดของประชากรเสียไปน้อยที่สุดเท่าที่กระทำได้แล้วนำผลที่ได้ไปอนุมานคุณลักษณะของประชากรถือเป็นการอ้างสถิติแบบฉบับ (classical statistics) อันประกอบด้วยกระบวนการประมาณค่า การทดสอบสมมุติฐานและการสรุปคุณลักษณะของประชากร

ธรรมชาติของประชากรและตัวอย่างที่สำคัญคือ คุณลักษณะของตัวอย่างย่อมเหมือนกับของประชากรที่ทราบคุณลักษณะอยู่ก่อน และในทางกลับกันถ้าต้องการทราบคุณลักษณะของตัวอย่างก่อนย่อมทราบคุณลักษณะของประชากรนั่นเอง ดังแสดงในภาพที่ 6.1 เป็นการแสดงการนิรนัยและการอุปนัยของประชากร



(ก) การนิรนัย



(ข) การอุปนัย

ภาพที่ 6.1 การนิรนัยและการอุปนัย

การนิรนัยหรือการพรรณนารสชาติของน้ำในทะเล-มหาสมุทร (ประชากร) ด้วยรสชาติน้ำในแก้ว (ตัวอย่าง) ได้ฉันใด การอุปนัยรสชาติของน้ำในแก้ว (ตัวอย่าง) เป็นรสชาติของน้ำทะเล-มหาสมุทร (ประชากร) ย่อมได้ฉันนั้น อย่างไรก็ดีจะต้องวางพื้นฐานของความไม่อคติ (unbias) ของตัวอย่างที่ได้มา

การชักตัวอย่างสุ่มหรือการสุ่มตัวอย่างจึงเป็นการได้ตัวอย่างมาจากประชากรโดยอาศัยกระบวนการสุ่ม(randomization)ซึ่งจะเป็นการทดลองสุ่ม(randomized experiments or experiments trial) หรือการสังเกตสุ่ม (randomized observation or observation studies) แล้วแต่โอกาสหรือสถานภาพของสิ่งที่เรากำลังสนใจอยู่จะเอื้ออำนวยให้สำหรับคุณลักษณะของประชากรที่ถูกเลือกสุ่มมาเป็นคุณลักษณะของตัวอย่างจะเรียกว่า ตัวแปรสุ่ม (random variable) เพื่อให้เห็นเด่นชัดว่าตัวแปรสุ่มเหล่านั้นเป็นคุณลักษณะของประชากรหรือของตัวอย่าง จึงกำหนดค่าที่บ่งบอกคุณลักษณะเช่น ค่าเฉลี่ย ผลรวม สัดส่วน ค่าเบี่ยงเบนมาตรฐานหรืออื่น ๆ กรณีประชากรเรียกว่า ตัวแปรเสริมหรือพารามิเตอร์ (parameter) กรณีตัวอย่างจะเรียกว่า ตัวสถิติ ค่าสถิติ หรือ สถิติ เมื่อมีตัวอย่างสุ่ม

มากกว่า 1 ตัวอย่างจากประชากรเดียว อาจให้ตัวสถิติที่แตกต่างกันก็ได้ ทั้งนี้ขึ้นอยู่กับขนาดของประชากรว่ามีขนาดจำกัด หรือมีขนาดอนันต์ และขนาดของตัวอย่างสุ่ม เทคนิคการชักตัวอย่างสุ่ม (random sampling technique) ที่จะกล่าวในหัวข้อต่อไปนี้จะสามารถกำหนดขนาดของประชากรให้เป็นขนาดจำกัดหรือขนาดอนันต์ก็ได้

6.2 การชักตัวอย่างสุ่ม

จากความรู้เรื่องความน่าจะเป็นของเหตุการณ์ในหัวข้อ 5.4 บทที่ 5 เหตุการณ์ที่เราสนใจอยู่ จะนำมาเป็นเซตของค่าสังเกตหรือตัวแปรสุ่มจำนวน n หน่วยซึ่งได้จากประชากรที่มีคุณลักษณะที่เราสนใจหรือมีข้อแตกต่างกันภายในประชากร m ประชากร และเราเรียกเซตนี้ว่า ตัวอย่างขนาด n หรือ ตัวอย่างขนาด n (a sample of size n) เมื่อได้ตัวอย่างมาด้วยวิธีการสุ่ม จึงเรียกตัวอย่างขนาด n ว่า ตัวอย่างสุ่มขนาด n หรือ ตัวอย่างสุ่มขนาด n (a random sample of size n) จะขอกล่าวถึงเทคนิคการชักตัวอย่าง 2 แบบดังต่อไปนี้

6.2.1 การชักตัวอย่างสุ่มแบบคืนที่ (sampling with replacment or W/R) มีลักษณะพอสรุปได้ว่า ถ้าตัวอย่างจำนวน 1 หน่วยถูกสุ่มหยิบจากประชากรที่มีคุณลักษณะที่เราสนใจหรือมีข้อแตกต่างกันภายในประชากร m ประชากรหรืออาจเรียกว่ามีหน่วยสุ่ม (sampling units) จำนวน m หน่วย แล้วใส่กลับคืนลงไปอีกก่อนสุ่มหยิบตัวอย่างหน่วยต่อไป เราเรียกเทคนิคการสุ่มตังอย่างนี้ว่า การชักตัวอย่างแบบคืนที่ จะพบว่า จำนวนหน่วยของประชากรหรือหน่วยสุ่มไม่เปลี่ยนแปลงตลอดระยะเวลาการสุ่ม ทำให้ประชากรนี้ถูกเรียกว่าประชากรชนิดไม่จำกัดหรือประชากรขนาดอนันต์ ความน่าจะเป็นของตัวอย่างจำนวน 1 หน่วยมีค่าคงตัวหรือที่เรียกว่าการกระจายความน่าจะเป็นตัวอย่างมีความสม่ำเสมอ (uniform distribution) เช่น การทอดเหรียญบาทที่สมบูรณ์ (ไม่ได้ถ่วงน้ำหนักด้านใดด้านหนึ่ง) ความน่าจะเป็นของการเกิดหัว (H) เท่ากับ 1/2 เมื่อต้องการทราบผลการเกิดหัว (H) ในการทอดเหรียญจำนวน 100 ครั้ง จะได้ความน่าจะเป็นการเกิดหัว ครั้งที่ 1, ครั้งที่ 2,...........และครั้งที่ 100 เท่ากับ (1/2)(1/2)..........(1/2) เท่ากับ $(1/2)^{100}$ จำนวนครั้งของการเกิดหัวอาจจะเป็น 0, 1, 2,............หรือ 100 ครั้งก็ได้ จำนวนครั้งดังกล่าวเราจัดเป็นตัวแปรวิยุต ถ้ากำหนด X เป็นตัวแปรวิยุต จากประชากรที่มีหน่วยสุ่มจำนวน m หน่วย เมื่อชักตัวอย่างสุ่มขนาด n อันประกอบด้วย X_1 , X_2 ,........และ X_n โดยใช้เทคนิคการชักตัวอย่างสุ่มแบบคืนที่ ความน่าจะเป็นของ X จะเท่ากันหมด

$$P(X_1) = P(X_2) = \text{.............} = (X_n) = \frac{1}{m} \qquad \text{…(6-1)}$$

ความน่าจะเป็นของตัวอย่างขนาด n นี้คือความน่าจะเป็นของ (X_1 , X_2 , X_3 และ..... X_n) แล้วได้เป็นสมการ (6-2)

$$P(X_1 \cap X_2 \cap \text{............} \cap X_n) = P(X_1)\ P(X_2)\text{..................}P(X_n) \qquad \text{…(6-2)}$$

ตัวอย่างที่ 6.1 สุ่มหยิบลูกบอลจำนวน 2 ลูกจากภาชนะที่บรรจุลูกบอลเบอร์ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 ทีละลูกแบบคืนที่ก่อนสุ่มหยิบครั้งต่อไป จงหาว่าความน่าจะเป็นของการได้ลูกบอลเบอร์ 5 และเบอร์ 6

วิธีทำ ประชากรมีหน่วยสุ่มตัวอย่าง 10 หน่วย กำหนดให้ X_0,X_1 , X_2 ,X_3 , X_4 , X_5 , X_6 ,X_7 , X_8 ,X_9 เป็นเหตุการณ์ที่ได้ลูกบอลเบอร์ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 ตามลำดับ

$$P(X_0) = P(X_1) = P(X_2) = \text{…..............} = P(X_9) = \frac{1}{10}$$

ดังนั้น ความน่าจะเป็นของการได้ลูกบอลเบอร์ 5 และเบอร์ 6 คือ

$$P(X_5 \cap X_6) = \left(\frac{1}{10}\right)\left(\frac{1}{10}\right) = \frac{1}{100}$$

ความน่าจะเป็นของการได้ลูกบอลเบอร์ 5 และเบอร์ 6 เท่ากับ $\frac{1}{100}$ **ตอบ**

6.2.2 การชักตัวอย่างสุ่มแบบไม่คืนที่ (sampling without repleacement or W/O) มีลักษณะสำคัญพอสรุปได้ว่า ถ้าตัวอย่างจำนวน 1 หน่วยถูกหยิบสุ่มจากประชากรที่มีคุณลักษณะที่เราสนใจหรือมีข้อแตกต่างกันภายในประชากร m ประการ (มีหน่วยสุ่มจำนวน m หน่วย) แล้วไม่ใส่คืนก่อนการหยิบครั้งต่อไป เราเรียกเทคนิคการชักตัวอย่างสุ่มนี้ว่า การชักตัวอย่างสุ่มแบบไม่คืนที่ จะพบว่าตัวอย่างสุ่มของประชากร โดยเฉพาะกรณีประชากรขนาดจำกัดจะลดจำนวนลง 1 หน่วยทุกครั้ง ที่มีการหยิบสุ่ม ในที่สุดก็จะหมดไป ยังผลทำให้ความน่าจะเป็นของตัวอย่างหน่วยต่อ ๆ ไปเพิ่มขึ้น เช่น ประชากรขนาด m เป็น 30,000 หน่วย ต้องการตัวอย่างขนาด n คือ 100 สุ่มหยิบครั้งละ 1 หน่วยแบบไม่คืนที่

สามารถบอกความน่าจะเป็นของตัวอย่างหน่วยที่ 1 และหน่วยที่ 2 และ........ และหน่วยที่ 99 และหน่วยที่ 100 เท่ากับ $\left(\frac{1}{30,000}\right)\left(\frac{1}{29,999}\right)\left(\frac{1}{29,998}\right)\ldots\ldots\left(\frac{1}{29,902}\right)\left(\frac{1}{29,901}\right)$

กรณีทั่วไป ถ้ากำหนด X เป็นตัวแปรวิยุตจากประชากรที่มีหน่วยสุ่มจำนวน m หน่วย เมื่อชักตัวอย่างสุ่มขนาด n อันประกอบด้วย X_1 , X_2 , ...และ X_n โดยใช้เทคนิคการชักตัวอย่างแบบไม่คืนที่ ความน่าจะเป็นของตัวอย่างขนาด n นี้คือความน่าจะเป็นของ (X_1 และ X_2 และ..................และ X_n) แล้วจะได้สมการ (6-3)

$$P(X_1 \cap X_2 \cap \ldots\ldots\ldots\ldots \cap X_n) = \left(\frac{1}{m}\right)\left(\frac{1}{m-1}\right)\left(\frac{1}{m-2}\right)\ldots\ldots\ldots \left(\frac{1}{m-n+1}\right) \quad \ldots(6\text{-}3)$$

ตัวอย่างที่ 6.2 จากตัวอย่างที่ 6.1 ถ้าใช้เทคนิคการสุ่มตัวอย่างแบบไม่คืนที่ จงหาความน่าจะเป็นของการได้ลูกบอลเบอร์ 5 และ เบอร์ 6

วิธีทำ

$$P(X_5 \cap X_6) = \left(\frac{1}{10}\right)\left(\frac{1}{9}\right)$$

ความน่าจะเป็นของการได้ลูกบอล เบอร์ 5 และ เบอร์ 6 เท่ากับ $\frac{1}{90}$ **ตอบ**

6.3 ค่าผิดพลาดของการชักตัวอย่างสุ่ม

ค่าผิดพลาดของการชักตัวอย่างสุ่ม หมายถึงความแตกต่างระหว่างตัวสถิติที่ได้จากประชากรเดียวกัน เมื่อมีการชักตัวอย่างสุ่มขนาด n มากกว่าหนึ่งตัวอย่างหรือหมายถึง การได้ค่าตัวแปรสุ่มที่แตกต่างกัน เมื่อชักตัวอย่างมากกว่า 1 ตัวอย่างจากประชากรเดียวกัน นอกจากนี้ยังมีความหมายรวมถึงค่าผิดพลาดของการสุ่ม (random error) ค่าผิดพลาดของการทดลอง (experiment error) และค่าผิดพลาดมาตรฐาน (standard error) สำหรับการวัดค่าผิดพลาดของการชักตัวอย่างสุ่มที่ธรรมดาและทำได้ทั่วไป เรียกว่า ความเชื่อถือ

ได้หรือความเที่ยงของการชักตัวอย่าง (reliability or precision of sampling) กระทำได้โดยการวัดความแปรปรวนของตัวสถิติ (variance of the sample statistic)

อย่างไรก็ดี มีกฎตัวเลขขนาดใหญ่ (the law of the numbers) กล่าวว่า เมื่อขนาดตัวอย่างโตขึ้นค่าผิดพลาดจะลดลงสำหรับเป็นหลักค่าผิดพลาดของการชักตัวอย่าง (Casella & Berger, 1990, p. 216)

6.4 การวางแผนการชักตัวอย่างสุ่ม

การวางแผนชักตัวอย่างสุ่ม จะนำไปใช้กับการวางแผนการทดลองโดยเฉพาะการทดลองสุ่ม กล่าวถึงรูปแบบของการชักตัวอย่าง 5 แบบดังต่อไปนี้

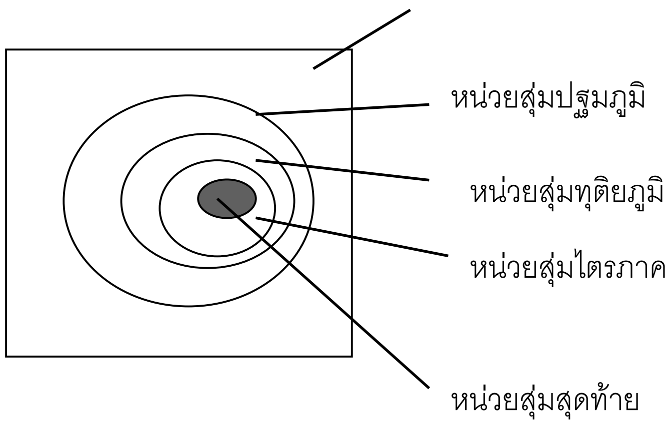
6.4.1 การชักตัวอย่างสุ่มเชิงเดี่ยว (simple random sampling) ตัวอย่างสุ่มเชิงเดี่ยว (simple random sample) เปรียบได้กับเหตุการณ์เชิงเดี่ยว กล่าวคือ ตัวอย่างสุ่มขนาด *n* ที่ได้รับการเลือกจากประชากรที่มีคุณลักษณะที่เราสนใจหรือข้อแตกต่างภายในประชากร *m* ประการ โดยมีข้อตกลงว่า ถ้าทุกสมาชิกของประชากรมีโอกาสจะได้รับเลือกเป็นสมาชิกจากตัวอย่างสุ่มขนาด *n* เท่า ๆ กัน และ ทุกสมาชิกของ *n* ถูกจัดให้มีโอกาสพบในตัวอย่างสุ่มที่เหมือน ๆ กัน สำหรับการชักตัวอย่างสุ่มเดี่ยวในทางปฏิบัติกรณีต้องการทราบความสูงของนักศึกษาทั้งหมด 100 คนด้วยการสุ่มวัดนักศึกษาจำนวน 5 คน (ตัวอย่างขนาด *n* เป็น 5 จากประชากร 100 หน่วย) นิยมใช้กันอยู่ 2 วิธีคือ การจับฉลากและการใช้ตารางสุ่มตัวเลขโดด (random digits) ดังแสดงในตารางภาคผนวกที่ 1 ในการชักตัวอย่างสุ่มเชิงเดี่ยวมักใช้กับกรณีประชากรมีจำนวนน้อย ๆ และมีความสม่ำเสมอ

6.4.2 การชักตัวอย่างสุ่มแบ่งกลุ่ม (cluster random sampling) กรณีประชากรมีขนาดใหญ่มาก ไม่สามารถชักตัวอย่างสุ่มเชิงเดี่ยวได้ทั่วถึง ซึ่งตัวอย่างแบ่งกลุ่ม (a cluster sample) ทำได้โดยแบ่งประชากรออกเป็นกลุ่มและชักตัวอย่างสุ่มเชิงเดี่ยวของกลุ่มนั้นครั้งแรก จากนั้นแบ่งกลุ่มตัวอย่างจากครั้งแรกเป็นกลุ่มและชักตัวอย่างสุ่มเชิงเดี่ยวอีกเป็นครั้งที่สอง กระทำเช่นนี้ต่อไปเรื่อย ๆ จนในที่สุดจะได้ตัวอย่างขนาด *n* ต้องการ

สมาชิกของกลุ่มตัวอย่างครั้งแรกเรียกว่า หน่วยชักตัวอย่างสุ่มปฐมภูมิ (primary sampling units) สมาชิกของกลุ่มตัวอย่างครั้งที่สองเรียกว่า หน่วยการชักตัวอย่างสุ่มทุติยภูมิ (secondary sampling units) และกระทำต่อไปเรื่อย ๆ จนกระทั่งเหลือสมาชิกกลุ่มที่ใช้เป็นตัวอย่างเรียกว่า หน่วยชักตัวอย่างสุ่มสุดท้าย (ultimate sampling units) ดังภาพที่ 6.2

6.4.3 การชักตัวอย่างสุ่มเป็นระบบ (systematic random sampling) วิธีการคือการชักตัวอย่างสุ่มเชิงเดี่ยวจากทุก ๆ หน่วยที่ k (every k^{th} units) ของประชากร โดยตัวเลข k เป็นค่าประมาณของอัตราส่วนตัวอย่างขนาด n กับประชากรขนาด m นั่นคือจะได้สมการ (6-4)

$$k = \frac{m}{n} \qquad \ldots(6\text{-}4)$$



ภาพที่ 6.2 หน่วยการชักตัวอย่างสุ่มแบ่งกลุ่ม

ถ้าต้องการชักตัวอย่างขนาด n = 1,000 หน่วย จากประชากร m = 20,000 หน่วย จะได้

$$k = \frac{20{,}000}{1{,}000} = 20$$

ตัวอย่างจะมาจากการสุ่มทุก ๆ 20 หน่วยของประชากร จะเห็นว่าสามารถทำได้ง่ายมากถ้าประชากรถูกเรียงลำดับหน่วยเอาไว้แล้ว เช่น เลขโทรศัพท์ เลขประจำตัวผู้เสียภาษี เป็นต้น

6.4.4 การชักตัวอย่างสุ่มเป็นชั้นภูมิ (stratified random sampling) เป็นการจัดลำดับความสำคัญของหน่วยประชากรเป็นกลุ่มเป็นชั้นทีเรียกว่าสตราตา (strata) ในแต่ละสตราตัม (stratum) พยายามให้หน่วยของประชากรเป็นเนื้อเดียวกันหรือเป็นเอกพันธ์ (homogeneous) ด้านคุณลักษณะ จากนั้นก็ชักตัวอย่างสุ่มเชิงเดี่ยวทุก ๆ ชั้นที่จัดไว้โดยปกติมักจะเป็นการหาอัตราส่วนตัวอย่างเช่นเดียวกันทุกชั้น เมื่อหาค่าเบี่ยงเบนมาตรฐาน

ของแต่ละชั้น ชั้นที่เป็นเนื้อเดียวกันจะมีค่าเบี่ยงเบนมาตรฐานน้อยกว่า ชั้นที่มีค่าเบี่ยงเบนมาตรฐานต่ำสุดจะนำมาเป็นตัวอย่างสุ่ม

6.4.5 การชักตัวอย่างสุ่มเป็นลำดับ (sequential random sampling) เป็นการชักตัวอย่างสุ่มตามลำดับเหตุการณ์ก่อนหลัง ถูกนำมาใช้อย่างกว้างขวาง โดยเฉพาะการควบคุมคุณภาพ (quality control) และตัวอย่างจะมีขนาดเล็ก ๆ มักใช้เพื่อประกอบการตัดสินใจในการยอมรับหรือปฏิเสธผลงานที่ทำได้ในเบื้องต้นอย่างรวดเร็ว

6.5 แนวคิดของการแจกแจงตัวอย่าง

จากความรู้เกี่ยวความน่าจะเป็นที่ต้องคำนึงถึงหนทางทั้งหมดที่เป็นไปได้ของเหตุการณ์ เป็นการกล่าวถึงความน่าจะเป็นจะเป็นตัวอย่างขนาด n จากประชากรหนึ่งซึ่งมีคุณลักษณะเฉพาะในตัวเองที่เรียกว่า ตัวแปรเสริม เช่น μ และ σ เป็นต้น เมื่อมีการชักตัวอย่างสุ่มขนาด n ตัวแปรสุ่มที่ได้มาจะก่อรูปเป็นตัวสถิติ การนำตัวสถิติเหล่านี้ไปสัมพันธ์กับความน่าจะเป็น เราเรียกว่า การแจกแจงตัวอย่าง (sampling distribution) ซึ่งเป็นรูปแบบการแจกแจงมากกว่า 20 แบบในปัจจุบันและสามารถนำแบบการแจกแจงตัวอย่างไปใช้ในการตัดสินใจเกี่ยวกับการสรุปคุณลักษณะของประชากร อันเป็นแนวทางของสาขาสถิติอนุมาน สำหรับการแจกแจงตัวอย่าง คือการแจกแจงความน่าจะเป็นของตัวสถิติ เช่น ค่าเฉลี่ยตัวอย่าง จะเรียกว่า การแจกแจงของค่าเฉลี่ยตัวอย่าง (sample distribution of sample mean) ความแปรปรวนของตัวอย่าง จะเรียกว่า การแจกแจงของการแปรปรวนตัวอย่าง (sampling distribution of sample variance) เป็นต้น

แบบของการแจกแจงที่สร้างขึ้นมากมายรวมทั้งตารางสำเร็จรูป แสดงความน่าจะเป็นของการแจกแจง จึงมีประโยชน์สำหรับการอนุมานคุณลักษณะของประชากรแตกต่างกัน เช่น การแจกแจงปรกติ ใช้ในการอนุมานคุณลักษณะประชากร เช่น ค่าเฉลี่ย ความแปรปรวน เป็นต้น สำหรับกรณีที่ตัวอย่างมีขนาดใหญ่ การแจกแจงแบบไคกำลังสอง (chi-square distribution) ใช้สำหรับการอนุมานค่าความแปรปรวนของประชากรหรือเทคนิคอื่น ๆ การแจกแจงที ใช้อนุมานค่าเฉลี่ย (μ) ของประชากร กรณีตัวอย่างขนาดเล็ก ($n \leq 20$) หรือกรณีไม่ทราบความแปรปรวนของประชากร เป็นต้น

6.5.1 แนวโน้มสู่ส่วนกลางของการแจกแจงตัวอย่าง (central tendency of sampling distribution) การวัดแนวโน้มสู่ส่วนกลางของประชากร ได้แก่ มัชฌิมเลขคณิต หรืออื่น ๆ แต่สำหรับการวัดแนวโน้มสู่ส่วนกลางของการแจกแจงตัวอย่างเราจะวัดด้วย

ค่าคาดหมายของตัวสถิติ (expectation value of a statistic) เช่น ค่าคาดหมายของค่าเฉลี่ยตัวอย่าง (expectation value of the sample mean : $E(\overline{X})$ ถ้าค่าเฉลี่ยหรือค่าคาดหมายของตัวสถิตินี้เป็นจริง จะนำไปสู่การประมาณค่าพารามิเตอร์ของประชากรต่อไป และจะกล่าวถึงการประมาณค่าอย่างไม่อคติ ในบทที่ 7 โดยจะแสดงให้เห็นว่า ถ้าค่าคาดหมายของ $\overline{X}$ เท่ากับ μ จะถือว่า $\overline{X}$ เป็นตัวประมาณค่า μ ที่ไม่อคติของประชากรแบบจุด เป็นต้น

6.5.2 การแปรผันของการแจกแจงตัวอย่าง (variation of sampling distribution) การแปรผันของประชากร หรือค่าสังเกต X จะวัดด้วยพิสัย ความแปรปรวน ค่าเบี่ยงเบนมาตรฐาน หรืออื่น ๆ เป็นการวัดแปรผันของค่าสังเกตรอบ ๆ พารามิเตอร์ แต่สำหรับการแจกแจงตัวอย่างค่าเบี่ยงเบนมาตรฐานเราจะเรียกว่า ค่าผิดพลาดมาตรฐาน (standard error) เช่น ค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่าง (standard error of sample mean) ค่าผิดพลาดมาตรฐานของความแปรปรวนตัวอย่าง (standard error of sample variance) เป็นต้น

6.6 การแจกแจงค่าเฉลี่ยตัวอย่าง

เราทราบว่าตัวอย่างขนาด n ซึ่งประกอบด้วยตัวแปรสุ่ม X_1, X_2, ...และ X_n (มีค่า x_1 , x_2 , ...และ x_n) จากประชากรขนาดอนันต์หรือขนาดจำกัดใช้เทคนิคการชักตัวอย่างสุ่มแบบคืนที่ สามารถหาตัวอย่างขนาด n ที่พึงเป็นไปได้ทั้งหมดและแตกต่างกันด้วยทฤษฎีความน่าจะเป็น ถ้าทุกตัวแปรสุ่ม X เป็นอิสระ และมีความน่าจะเป็นเท่ากันหมด การแจกแจงของประชากรคือ $P(X_1) = P(X_2) = ... = P(X_n)$ โดยมีค่าเฉลี่ยเท่ากับ μ และค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_x เช่น ประชากรประกอบด้วย {1,3,5,7} ความน่าจะเป็น X ของประชากรคือ

$$P(X_1 = 1) = P(X_2 = 3) = P(X_3 = 5) = P(X_4 = 7) = \frac{1}{4}$$

คำนวณค่าเฉลี่ย (μ) ได้เท่ากับ 4 และค่าเบี่ยงเบนมาตรฐาน (σ_x) เท่ากับ $\sqrt{5}$ เมื่อต้องการชักตัวอย่างขนาด n = 2 จากประชากรนี้โดยใช้เทคนิคการชักตัวอย่างแบบคืนที่จะได้ตัวอย่างที่พึงเป็นไปได้ทั้งหมดและแตกต่างกันจำนวน 16 ตัวอย่าง พร้อมทั้งหาค่าเฉลี่ยตัวอย่าง (sample mean) ดังตารางที่ 6.1 ซึ่งจะพบว่า ใน 16 ตัวอย่างมีเซต

ค่าเฉลี่ยตัวอย่างเป็น {1,2,3,4,5,6,7} โดยมีค่าเฉลี่ยเท่ากับ 1 จำนวน 1 ตัวอย่าง มีค่าเฉลี่ยเท่ากับ 2 จำนวน 2 ตัวอย่าง ค่าเฉลี่ยเท่ากับ 3 จำนวน 3 ตัวอย่าง ค่าเฉลี่ยเท่ากับ 4 จำนวน 4 ตัวอย่าง ค่าเฉลี่ยเท่ากับ 5 จำนวน 3 ตัวอย่าง ค่าเฉลี่ยเท่ากับ 6 จำนวน 2 ตัวอย่าง และค่าเฉลี่ยเท่ากับ 7 จำนวน 1 ตัวอย่าง

ตารางที่ 6.1 ตัวอย่างขนาด $n = 2$ สำหรับตัวแปรสุ่ม X จากประชากร {1,3,5,7} โดยใช้เทคนิคการชักตัวอย่างสุ่มแบบคืนที่

ตัวอย่างที่	X_1	X_2	$\overline{X}$
1	1	1	1
2	1	3	2
3	1	5	3
4	1	7	4
5	3	1	2
6	3	3	3
7	3	5	4
8	3	7	5
9	5	1	3
10	5	3	4
11	5	5	5
12	5	7	6
13	7	1	4
14	7	3	5
15	7	5	6
16	7	7	7

เราสามารถแสดงให้เห็นความน่าจะเป็นของค่าเฉลี่ยตัวอย่างด้วยการแจกแจงของค่าเฉลี่ยตัวอย่างดังตารางที่ 6.2 ซึ่งจะพบว่าแต่ละ $\overline{X}$ ความน่าจะเป็นเท่ากับ $\left(\frac{1}{4}\right)\left(\frac{1}{4}\right)$

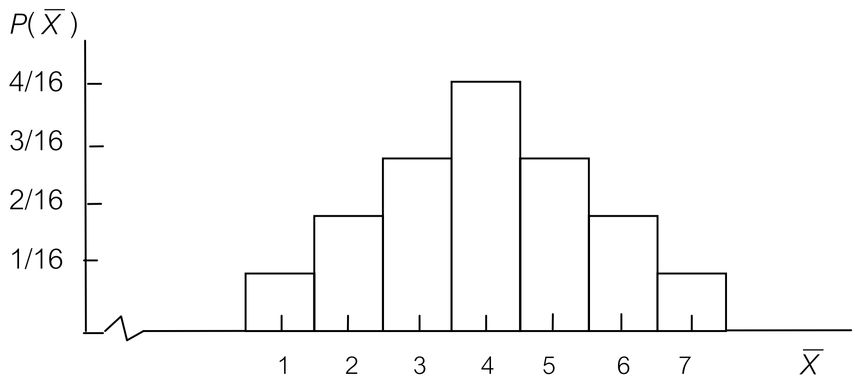
เท่ากับ $\frac{1}{16}$ กรณีมี $\bar{X}$ เท่ากันหลายค่าจะใช้หลักการผนวกเหตุการณ์ในสมการ (5-14) ในบทที่ 5 เช่นกรณี $\bar{X} = 3$ มีจำนวน 3 ตัวอย่าง ดังนั้นความน่าจะเป็นของ $\bar{X} = 3$ หรือ

$$P(\bar{X} = 3) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{3}{16}$$

และสามารถเขียนฮิสโทแกรม การแจกแจงของค่าเฉลี่ยตัวอย่าง ได้ดังภาพที่ 6.3

ตารางที่ 6.2 การแจกแจงของค่าเฉลี่ยตัวอย่างจากตารางที่ 6.1

ค่าเฉลี่ยตัวอย่าง ($\bar{X}$)	จำนวนตัวอย่าง	ความน่าจะเป็น $P(\bar{X})$
1	1	$\frac{1}{16}$
2	2	$\frac{2}{16}$
3	3	$\frac{3}{16}$
4	4	$\frac{4}{16}$
5	3	$\frac{3}{16}$
6	2	$\frac{2}{16}$
7	1	$\frac{1}{16}$
รวม	16	1



ภาพที่ 6.3 ฮิสโทแกรมการแจกแจงของค่าเฉลี่ยตัวอย่างจากตารางที่ 6.2

ในการแจกแจงของค่าเฉลี่ยตัวอย่างมีวิธีการดังนี้

6.6.1 ค่าคาดหมายของค่าเฉลี่ยตัวอย่าง (expectation value of the sample mean) การแจกแจงของค่าเฉลี่ยตัวอย่าง ($\overline{X}$) คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม $\overline{X}$ ค่าเฉลี่ยของตัวแปรสุ่ม $\overline{X}$ ในรูปความน่าจะเป็น จึงเรียกว่า ค่าคาดหมายของ $\overline{X}$ หรือ ค่าคาดหมายของตัวแปรสุ่ม $\overline{X}$ (expectation of $\overline{X}$) ใช้สัญลักษณ์ $E(\overline{X})$

ถ้ากำหนด $\overline{X}_1$, $\overline{X}_2$,............., $\overline{X}_k$ เป็นตัวแปรสุ่ม. มีความน่าจะเป็น $P(\overline{X}_1)$, $P(\overline{X}_2)$,.........,$P(\overline{X}_k)$ ตามลำดับ ทำนองเดียวกับตัวแปรสุ่ม X มีค่าคาดหมายของ X ตามสมการ (5-21) ในบทที่ 5 แล้วจะได้สมการ (6-5)

$$E(\overline{X}) = \sum_{j=1}^{k} \overline{X}_j P(\overline{X}_j) \qquad \text{...(6-5)}$$

ดังนั้น ค่าคาดหมายของ $\overline{X}$ จากการแจกแจงของ $\overline{X}$ ในตารางที่ 6.2 คือ

$$E(\overline{X}) = 1(\frac{1}{16}) + 2(\frac{1}{16})+3(\frac{1}{16})+4(\frac{1}{16})+5(\frac{1}{16})+6(\frac{1}{16})+7(\frac{1}{16})$$

$$= \frac{1}{16} + \frac{4}{16} + \frac{9}{16} + \frac{16}{16} + \frac{15}{16} + \frac{12}{16} + \frac{7}{16}$$

$$E(\overline{X}) = 4$$

เราทราบว่า ค่าเฉลี่ยของประชากรนี้ คือ (1 +3 + 5 + 7)/4 เท่ากับ 4 แสดงว่า ค่าคาดหมายของ $\overline{X}$ กรณีนี้เป็นดังสมการ (6-6)

$$E(\overline{X}) = E(X) = \mu \qquad \text{...(6-6)}$$

X : เป็นตัวแปรสุ่มของประชากร

$\overline{X}$: เป็นค่าเฉลี่ยตัวอย่างสุ่มขนาด n

6.6.2 ค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่าง (standard error of the sample mean) สำหรับค่าเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม $\overline{X}$ เมื่อมีการแจกแจงตัวอย่าง เราเรียกว่า ค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่าง ใช้สัญลักษณ์ $\sigma_{\bar{x}}$ หาได้ทำนองเดียวกับการหาความแปรปรวนของตัวแปรสุ่ม X จากสมการ (5-23) ในบทที่ 5 และค่าคาดหมายของ $\overline{X}$ จากสมการ (6-6) จะได้ความแปรปรวนของค่าเฉลี่ยเป็นดังสมการ (6-7)

$$\sigma^2_{\bar{x}} \quad = \quad \sum_{j=1}^{k} \bar{X}_j^2 P(\bar{X}_j) - [E(\bar{X})]^2 \qquad \ldots(6\text{-}7)$$

จากข้อมูลในตารางที่ 6.2 จะได้

$$\sigma^2_{\bar{x}} \quad = \quad \frac{296}{16} - 4^2 = 2.5$$

และ $\sigma_{\bar{x}} = \sqrt{2.5}$ สำหรับความแปรปรวนของประชากรจากตารางที่ 6.2 หาได้จาก

$$\sigma^2_{\bar{x}} \quad = \quad \sum_{j=1}^{k} \bar{X}_j^2 P(\bar{X}_j) - \mu^2$$

$$= \quad [1^2(\frac{1}{4})+2^2(\frac{1}{4})+5^2(\frac{1}{4})+ 7^2(\frac{1}{4})] - 4^2$$

$$= \quad 5$$

จากตารางที่ 6.2 สามารถคำนวณค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่างได้ดังตารางที่ 6.3

ตารางที่ 6.3 การคำนวณค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่างจากการแจกแจงค่าเฉลี่ยตัวอย่างของตารางที่ 6.2

ค่าเฉลี่ยตัวอย่าง	$P(\bar{X}_j)$	X_j^2	$X_j^2 \; P(\bar{X}_j)$
1	$\frac{1}{16}$	1	$\frac{1}{16}$
2	$\frac{2}{16}$	4	$\frac{8}{16}$
3	$\frac{3}{16}$	9	$\frac{27}{16}$
4	$\frac{4}{16}$	16	$\frac{64}{16}$
5	$\frac{3}{16}$	25	$\frac{75}{16}$
6	$\frac{2}{16}$	36	$\frac{72}{16}$
7	$\frac{1}{16}$	49	$\frac{49}{16}$
	1		$\frac{296}{16}$

จะเห็นว่าความแปรปรวนของค่าเฉลี่ยตัวอย่างจะเท่ากับความแปรปรวนของประชากรหารด้วยขนาดของตัวอย่าง จึงสรุปได้ว่าความแปรปรวนของตัวอย่าง คือสมการ (6-8)

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \qquad \text{…(6-8)}$$

และค่าผิดพลาดมาตรฐานของตัวอย่าง คือ สมการ (6-9)

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \qquad \text{…(6-9)}$$

จากสมการ (6-8) และ (6-9) และคุณสมบัติความแปรปรวนจากสมการ (3-23) ในบทที่ 3 เมื่อกำหนด X เป็นตัวแปรสุ่มอิสระ แต่กรณีนี้ $\overline{X}$ เป็นค่าเฉลี่ยตัวอย่างขนาด n เราอาจพิสูจน์ได้ว่า

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2} Var\ (X_1 + X_2 + \ldots + X_n)$$

$$= \frac{1}{n^2}(\sigma_{x_1}^2 + \sigma_{x_2}^2 + \ldots + \sigma_{x_n}^2)$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2}(n\ \sigma_x^2)$$

$$= \frac{\sigma_x^2}{n}$$

ตัวอย่างที่ 6.3 ถ้าประชากรขนาดอนันต์ คือ ต้นข้าวโพดมีความสูงเฉลี่ย μ = 69 นิ้ว และค่าเบี่ยงเบนมาตรฐาน σ = 3.2 นิ้ว

ก. จงหาค่าคาดหมายของ $\overline{X}$ และค่าผิดพลาดมาตรฐานของ $\overline{X}$ ถ้าสุ่มตัวอย่าง n = 5 มาจำนวนหลายตัวอย่าง

ข. จงอธิบายความสัมพันธ์ระหว่าง $\overline{X}$ กับ μ ถ้าตัวอย่างหนึ่ง มีค่า $\overline{X}$ = 70

วิธีทำ ก. ค่าคาดหมายของ $\overline{X}$ คือ

$$E(\overline{X}) = \mu = 69$$

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

$$= \frac{3.2}{\sqrt{5}}$$

$$= 1.4$$

นั่นคือ ความสูงเฉลี่ยของต้นข้าวโพดทั้งหลายจะกระจายอยู่รอบ ๆ 69 นิ้วด้วยค่าผิดพลาดระหว่าง 69 -1.4 = 67.6 นิ้ว กับ 69 +1.4 = 70.4 นิ้ว **ตอบ**

ข. ตัวอย่างที่มีความสูง 70 นิ้ว ขณะที่ $\mu = 69$ นิ้ว แสดงว่า $\overline{X}$ ตัวอย่างนี้แตกต่างจาก μ เพียง 1 นิ้ว อยู่ในขอบเขตของค่าผิดพลาดมาตรฐานที่คำนวณได้ดังข้อ ก. **ตอบ**

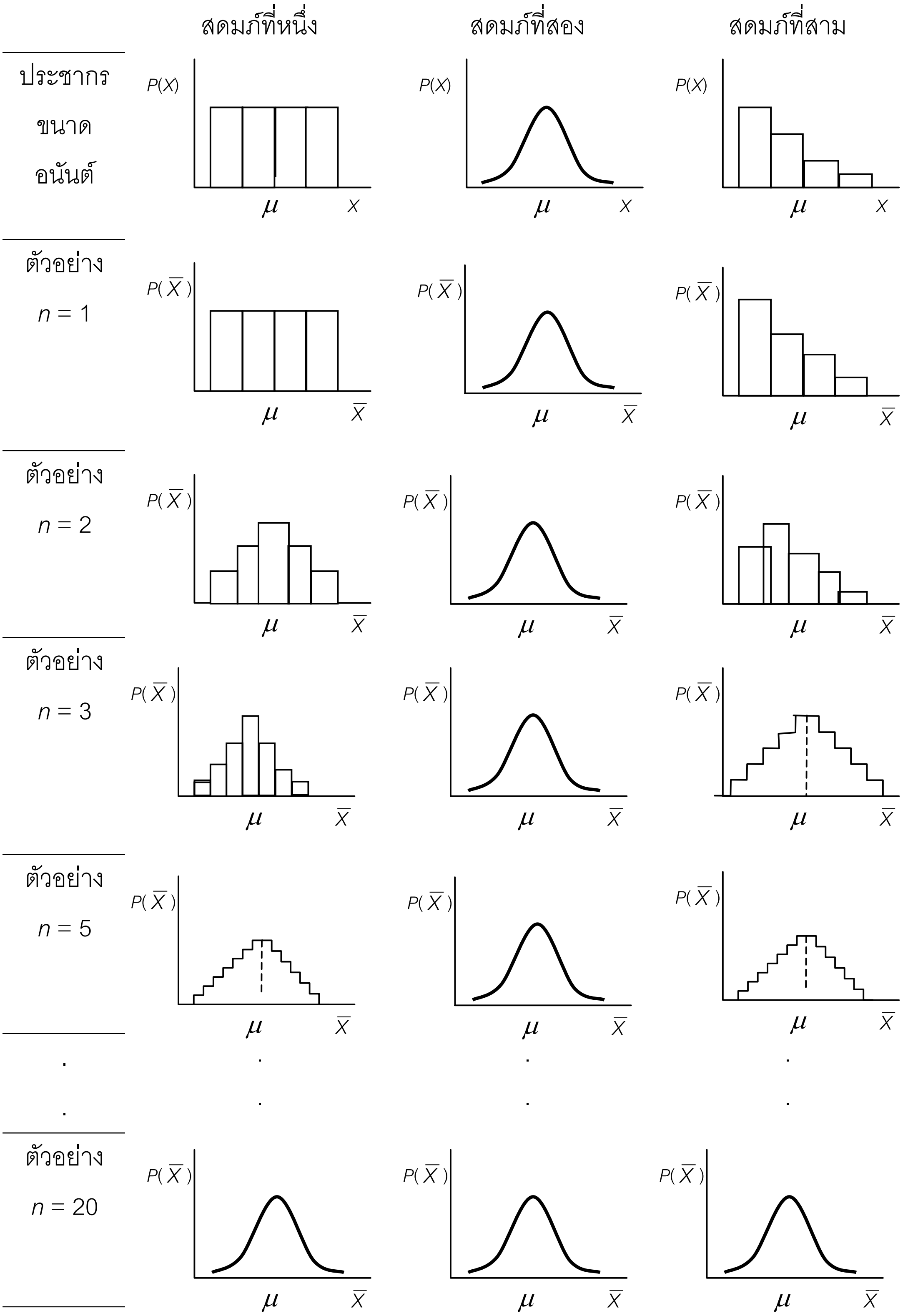
6.7 การแจกแจงปรกติของค่าเฉลี่ยตัวอย่าง

จากฮิสโทแกรมในภาพที่ 6.3 จะเห็นว่ามีการแจกแจงปรกติ และจากสมการ (6-8) ขนาดของตัวอย่างมีผลกระทบต่อความแปรปรวนของ $\overline{X}$ และค่าผิดพลาดมาตรฐานของ $\overline{X}$ ในสมการ (6-9) ดังเช่น

กรณี $n = 2$; $$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{2}}$$

กรณี $n = 3$; $$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{3}}$$

กรณี $n = 100$; $$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{100}}$$



ภาพที่ 6.4 การแจกแจงปรกติของ $\overline{X}$ เมื่อ n มีขนาดต่าง ๆ

ที่มา : (Wonnacott & Wonnacott, 1985, p.163)

ดังนั้นกรณี $n = \infty$ จะทำให้ $\sigma_{\bar{x}}$ เข้าใกล้ 0

หรือ $\sigma_{\bar{x}} = 0$

แสดงว่าค่าคาดหมายของ $\bar{X}$ เข้าใกล้ μ ของประชากรเมื่อขนาดตัวอย่างโตขึ้น และการแจกแจงของ $\bar{X}$ จะเป็นการแจกแจงปรกติ ถึงแม้ว่าประชากรจะมีการแจกแจงใด ๆ ดังแสดงในภาพที่ 6.4

สดมภ์ที่หนึ่ง ประชากรแจกแจงเอกรูป (uniform distribution) หรือที่เรียกว่าการแจกแจงแบบสี่เหลี่ยมผืนผ้า (rectangular distribution) จะพบการแจกแจงของ $\bar{X}$ เข้าสู่การแจกแจงปรกติ เมื่อตัวอย่างมีขนาด $n = 5$ หรือ $n = 10$

สดมภ์ที่สอง ประชากรแจกแจงปรกติ ทำให้การแจกแจงของ $\bar{X}$ แจกแจงปรกติตั้งแต่ตัวอย่างมีขนาด $n = 1$

สดมภ์ที่สาม ประชากรแจกแจงแบบเบ้ จะพบการแจกแจงของ $\bar{X}$ เข้าสู่การแจกแจงปรกติหรือถือว่าแจกแจงปรกติ เมื่อตัวอย่างมีขนาด $n = 10$ หรือ $n = 20$ ปรากฏการณ์เหล่านี้ ได้รับการพิสูจน์และเสนอโดยไลนด์ เกรน (Lind Gren) (Wonnacott & Wonnacott,1985, p.164) และเรียกปรากฏการณ์นี้ว่า ทฤษฎีแนวโน้มสู่ส่วนกลาง (the central limit theorem :CLT) ซึ่งสรุปใจความของกฏได้ว่าการชักตัวอย่างขนาด n จากประชากรที่มีค่าเฉลี่ยเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_x จะได้ค่าเฉลี่ยตัวอย่าง $\bar{X}$ ใด ๆ มีค่าผิดพลาดมาตรฐานเท่ากับ $\frac{\sigma_x}{\sqrt{n}}$ รอบๆ μ พร้อมทั้งการแจกแจงของ $\bar{X}$ จะเป็นการแจกแจงปรกติหรือเข้าใกล้การแจกแจงปรกติ นั่นคือประชากรที่มีการแจกแจงปรกติหรือตัวอย่างขนาดใหญ่ (n =10 หรือ n = 0 ก็เพียงพอ) การแจกแจงของ $\bar{X}$ จะถือว่าการแจกแจงปรกติ เมื่อเราทราบว่าประชากรหนึ่งมีค่าเฉลี่ยเท่ากับ μ ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_x ถ้ากรณีประชากรนี้มีการแจกแจงปรกติหรือตัวอย่างขนาดใหญ่ ($n \geq 20$) การแจกแจงของ $\bar{X}$ มีการแจกแจงปรกติดังกล่าว จึงสามารถนำการแจกแจงของไปสู่การแจกแจงปรกติมาตรฐาน จะได้สมการ (6-10)

$$Z_i = \left(\frac{\bar{X}_i - \mu}{\sigma_{\bar{X}}} \right) \qquad \ldots(6\text{-}10)$$

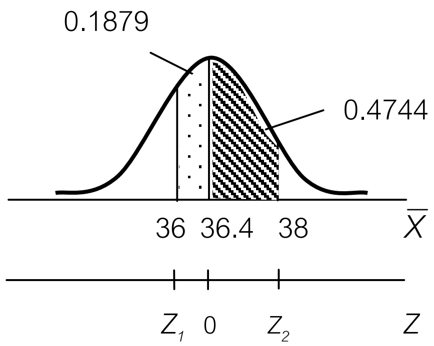
แต่จากสมการ (6-9) $= \frac{\sigma_x}{\sqrt{n}}$

ดังนั้นสมการ (6-10) เขียนใหม่ได้เป็นสมการ (6-11)

$$Z_i = \left(\frac{\overline{X}_i - \mu}{\frac{\sigma_x}{\sqrt{n}}} \right) \qquad \ldots(6\text{-}11)$$

ตัวอย่างที่ 6.4 จากตัวอย่างสุ่มขนาด $n = 40$ ของประชากรที่มีค่าเฉลี่ย μ = 36.4 ค่าเบี่ยงเบนมาตรฐาน σ_x = 5.2 จงหาโอกาสที่จะได้ค่าเฉลี่ยของตัวอย่างนี้ เป็น 36 ถึง 38

วิธีทำ จากการแจกแจงปรกติของ $\overline{X}$ ในภาพที่ 6.5



ภาพที่ 6.5 การแจกแจงปรกติของ $\overline{X}$ เมื่อ $n = 40$

และ $$Z_i = \left(\frac{\overline{X}_i - \mu}{\frac{\sigma_x}{\sqrt{n}}} \right)$$

จะได้ $$Z_1 = \left(\frac{36 - 36.4}{\frac{5.2}{\sqrt{40}}} \right)$$

$$= -0.49$$

$$Z_2 = \left(\frac{38 - 36.4}{\frac{5.2}{\sqrt{40}}} \right)$$

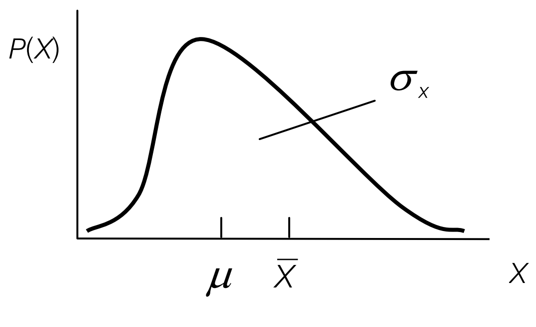
$$= 1.95$$

โอกาสที่จะพบค่าระหว่างเฉลี่ยตัวอย่างระหว่าง 36 กับ 38 เท่ากับ 0.1879+0.4744 = 0.6623 **ตอบ**

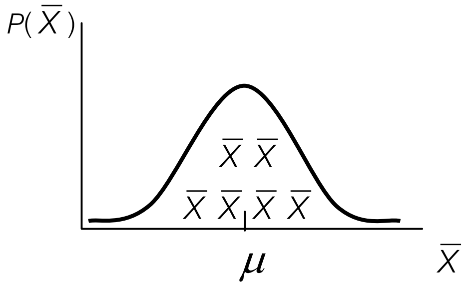
เมื่อตัวแปรสุ่ม $\bar{X}$ ได้จากประชากรแจกแจงปรกติหรือตัวอย่างขนาดใหญ่ ($n \geq 20$) การแจกแจงของ $\bar{X}$ จะเป็นการแจกแจงปรกติหรือแจกแจงปรกติโดยประมาณ (Casella & Berger, 1990, pp. 216-217) ถ้าประชากรนี้มีค่าเฉลี่ยเท่ากับ μ และค่าเบี่ยงเบนมาตรฐานเท่ากับ σ จะเขียนสัญลักษณ์การแจกแจงของ $\bar{X}$ ตามสมการ (6-12)

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \qquad \text{...(6-12)}$$

เปรียบเทียบสมการ (6-12) กับสมการ (4-3) ในบทที่ 4 คุณสมบัติการแจกแจงปรกติของ $\bar{X}$ จะเป็นทำนองเดียวกับการแจกแจงปรกติของตัวแปร X ทุกประการ นอกจากนี้ตัวแปรสุ่ม $\bar{X}$ ยังสามารถประกอบกันเป็นตัวแปรสุ่มใหม่และตัวแปรสุ่มตัวใหม่ก็จะยังคงแจกแจงปรกติหรือแจกแจงปรกติโดยประมาณอยู่ ซึ่งคุณสมบัติเหล่านี้แสดงได้ดังภาพที่ 6.6 และ ภาพที่ 6.7



ภาพที่ 6.6 ประชากรแจกแจงใด ๆ และค่าเฉลี่ยตัวอย่างค่าหนึ่งที่สุ่มได้



ภาพที่ 6.7 การแจกแจงปรกติของตัวอย่างที่พึงเป็นไปได้ทั้งหมด
จากประชากรภาพที่ 6.6

ถ้าตัวแปรสุ่ม $\overline{X}_1$, $\overline{X}_2$ ได้จากประชากรแจกแจงปรกติค่าเฉลี่ย μ_1 , μ_2 และความแปรปรวน σ_1^2 , σ_2^2 หรือตัวอย่างขนาดใหญ่ (n_1 , $n_2 \geq 20$) ตามลำดับ เขียนแทนด้วยสมการ (6-13) และ สมการ (6-14) ดังนี้

$$\overline{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n}) \qquad \ldots (6\text{-}13)$$

และ

$$\overline{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n}) \qquad \ldots (6\text{-}14)$$

จะได้คุณสมบัติของตัวแปรสุ่มใหม่ดังสมการ (6-15) ถึงสมการ (6-17) ดังนี้

1) ถ้าตัวแปรสุ่ม $D = X_1 - X_2$ และจะได้ว่า D แจกแจงปรกติดังสมการ (6-15)

$$D \sim N[\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}] \qquad \ldots (6\text{-}15)$$

2) ถ้าตัวแปรสุ่ม $D = X_1 - X_2$ และ $\sigma_1^2 = \sigma_2^2 = \sigma^2$ แล้วจะได้ว่า D แจกแจงปรกติดังสมการ (6-16)

$$D \sim N[\mu_1 - \mu_2, \sigma^2 (\frac{1}{n_1} + \frac{1}{n_2})] \qquad \ldots (6\text{-}16)$$

3) ถ้าตัวแปรสุ่ม $D = X_1 - X_2$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ และ $n_1 = n_2 = n$ แล้วจะได้ว่า D แจกแจงปรกติดังสมการ (6-17)

$$D \sim N[\mu_1 - \mu_2, \frac{2\sigma^2}{n}] \qquad \text{...(6-17)}$$

การพิสูจน์สมการ (6-15) จะแสดงในหัวข้อ 7.4.1 ในบทที่ 7 สำหรับสมการ (6-16) และ (6-17) ใช้วิธีพีชคณิตธรรมดาก็สามารถแสดงได้ จึงขอยกเว้นไม่แสดงในที่นี้

6.8 บทสรุป

บทนี้เป็นการนำทฤษฎีมูลฐานความน่าจะเป็นมาใช้ในกระบวนการชักตัวอย่างและถือเป็นหลักสำคัญของสถิติอนุมานและการวางแผนการทดลอง ซึ่งสรุปได้ดังนี้

6.8.1 เหตุผลการใช้ตัวอย่าง สิ่งที่จะต้องคำนึงถึง คือ

1) การนิรนัยและการอุปนัย
2) ตัวสถิติ
3) ประชากรขนาดจำกัดและประชากรขนาดอนันต์
4) การชักตัวอย่างสุ่มแบบคืนที่และการชักตัวอย่างสุ่มแบบไม่คืนที่

6.8.2 การชักตัวอย่างสุ่ม มีวิธีการ 2 วิธีคือ การชักตัวอย่างสุ่มแบบคืนที่ และการชักตัวอย่างสุ่มแบบไม่คืนที่

6.8.3 ในการชักตัวอย่าง จะทำให้เกิดความแตกต่างระหว่างตัวสถิติที่ได้จากประชากรเดียวกันซึ่งทำให้เกิดค่าผิดพลาดของการชักตัวอย่างได้ ดังนั้นในการชักตัวอย่างสุ่มจึงต้องมีการวางแผน ซึ่งมีอยู่ 5 รูปแบบคือ

1) การชักตัวอย่างสุ่มเชิงเดี่ยว
2) การชักตัวอย่างสุ่มแบ่งกลุ่ม
3) การชักตัวอย่างสุ่มเป็นระบบ
4) การชักตัวอย่างสุ่มเป็นชั้นภูมิ
5) การชักตัวอย่างสุ่มเป็นลำดับ

6.8.4 ในการแจกแจงปรกติของค่าเฉลี่ยตัวอย่าง มีวิธีการดังนี้

1) การโน้มเข้าสู่ส่วนกลางของการแจกแจงค่าเฉลี่ยตัวอย่าง
2) การแปรผันของการแจกแจงเฉลี่ยตัวอย่าง

6.9 คำถามทบทวน

1. จงอธิบายความแตกต่างระหว่างการนิรนัยกับการอุปนัย

2. จงอธิบายและยกตัวอย่างเกี่ยวกับการชักตัวอย่างสุ่มแบบใส่คืน มีผลทำให้ประชากรขนาด *m* หน่วยเป็นประชากรขนาดอนันต์ได้อย่างไร

3. จงอธิบายความสัมพันธ์ระหว่างกระบวนการชักตัวอย่างสุ่มกับการแจกแจงตัวอย่าง

4. จงอธิบายการชักตัวอย่างเชิงเดี่ยวจากรายชื่อบุคคลในสมุดโทรศัพท์

5. จากการศึกษารายได้ของโรงงานแห่งหนึ่งจำนวนมาก พบว่ามีรายได้แจกแจงปรกติรายได้เฉลี่ยต่อเดือนคนละ 800 บาท ค่าเบี่ยงเบนมาตรฐาน 900 บาท ถ้าตัวอย่างสุ่มเป็นพนักงาน 36 คนและถามถึงรายได้ของเขา จงหา

ก. โอกาสที่จะพบพนักงานมีรายได้เฉลี่ยต่อเดือนไม่เกินคนละ 1000 บาท

ข. โอกาสที่จะพบพนักงานที่มีรายได้เฉลี่ยต่อเดือนต่อคนมีค่า ระหว่าง 790 ถึง 810 บาท

ค. โอกาสที่จะพบพนักงานมีรายได้เฉลี่ยต่อเดือนตั้งแต่คนละ 790 บาท ขึ้นไป

6. ถ้าความสูงของนักศึกษาของวิทยาลัยแห่งหนึ่งจำนวน 3000 คน มีการแจกแจงปรกติความสูงเฉลี่ย 170 ซ.ม. ค่าเบี่ยงเบนมาตรฐาน 7.5 ซม. หากสุ่มวัดความสูงนักศึกษาจำนวน 80 ตัวอย่าง ตัวอย่างสุ่มละ 20 คน จงหาจำนวนตัวอย่างสุ่มที่มีความสูงเฉลี่ยระหว่าง 167 ซม. ถึง 171 ซม.

7. สุ่มตัวอย่างขนาด 75 จากประชากรซึ่งมีค่าเฉลี่ย 112 ค่าเบี่ยงเบนมาตรฐาน 10 จงหาโอกาสที่จะพบตัวอย่างสุ่มที่มีค่าเฉลี่ยระหว่าง 110.5 กับ 113.5

บทที่ 7
การประมาณค่าเกี่ยวกับประชากร

งานหลักของสถิติอนุมานคือการหาคุณลักษณะเฉพาะของประชากร (characteristics of parameter) โดยอาศัยตัวสถิติจากตัวอย่างการประมาณค่าเกี่ยวกับประชากรภายใต้ หลักการดำเนินการเกี่ยวกับพารามิเตอร์ซึ่งมีวิธีหลักการประมาณค่าเชิงสถิติพารามิเตอร์ และการทดสอบสมมุติฐาน ในบทนี้เป็นการประมาณค่าเชิงสถิติพารามิเตอร์ (parameter statistics or statistical estimation) ซึ่งมีรูปแบบการประมาณค่าด้วยค่าคงตัวหรือการประมาณค่าแบบจุด (point estimation) และการประมาณค่าด้วยช่วง (interval estimation) หรือช่วงความเชื่อมั่น (confidence interval) ก่อนกล่าวถึงการประมาณค่าพารามิเตอร์ ขอให้พิจารณาตารางที่ 7.1 เพื่อให้ระลึกอยู่เสมอว่าตัวอย่างเป็นหน่วยย่อยของประชากรที่มีพารามิเตอร์เป็นค่าคงตัว แต่เราไม่อาจทราบค่าได้อย่างแน่ชัดเปรียบเสมือนเป้ากระสุนระยะไกล เมื่อเรามีคุณลักษณะเฉพาะของตัวอย่างได้แก่ ค่าเฉลี่ยตัวอย่าง ($\overline{X}$) และค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) อยู่ในมือเท่านั้น การประมาณค่าแบบจุดอาจเปรียบเสมือนการยิงกระสุนเข้าสู่เป้า และการประมาณค่าด้วยช่วงอาจเปรียบเสมือนการโยนห่วงหรือเกือกม้าเข้าสู่หลักเป้า ผลที่ได้จึงมักไม่กล่าวถึงความถูกต้อง แต่จะกล่าวถึงความเที่ยง (precision) คือการซ้ำที่เดิมและความแม่น (accuracy) คือ การเข้าใกล้หรือเข้าเป้ามากน้อยเพียงใด

ตารางที่ 7.1 การเปรียบเทียบสถิติและพารามิเตอร์ของตัวอย่างและประชากร

ตัวอย่าง	ประชากร
ใช้ความถี่สัมพัทธ์ $\frac{f}{n}$ หาค่า $\overline{X}$ และ S^2 จัดเป็นตัวสถิติ หรือตัวประมาณค่า	ใช้ความน่าจะเป็น $P(X)$ หาค่า μ และ σ^2 จัดเป็นตัวพารามิเตอร์หรือค่าคงตัว

7.1 การประมาณค่าแบบจุด

ตัวประมาณค่าพารามิเตอร์ที่ดีจะต้องไม่มีอคติ มีความแนบนัย (consistency) คือค่าตัวสถิติ ($\hat{\theta}$) จะต้องเข้าใกล้คุณลักษณะเฉพาะของประชากร (θ) เมื่อ n มีขนาดใหญ่ ๆ มีประสิทธิภาพ (efficiency) คือ ถ้าตัวสถิติตัวใดมีความแปรปรวนน้อยถือว่าตัวสถิติตัวนั้นมีประสิทธิภาพมากกว่า เช่น $V(\overline{X}) = \frac{\sigma^2}{n}$ แต่ $V(Me) = \frac{\pi\sigma^2}{2n}$ จะเรียกว่า $\overline{X}$ มีประสิทธิภาพมากกว่า Me และความเพียงพอ (sufficiency) คือตัวสถิตินั้นจะต้องใช้ได้กับ พารามิเตอร์ทุกตัวของประชากรได้มีหลายประการ ในหัวข้อนี้จะขอกล่าวถึงเฉพาะลักษณะที่เด่นที่สุดคือความไม่อคติ เมื่อใช้ตัวสถิติเป็นตัวประมาณค่าคุณลักษณะประชากรจะไม่อคติสำหรับการประมาณค่า θ ถ้าค่าคาดหมายของ $\hat{\theta}$ คือ θ ซึ่งเขียนได้เป็นสมการ (7-1) (Wonnacott & Wonnacott,1985, p. 161)

$$E(\hat{\theta}) = \theta \qquad \text{...(7-1)}$$

ในการประมาณค่าคุณลักษณะประชากรมีวิธีการดังนี้

7.1.1 การประมาณค่าเฉลี่ยประชากร (population mean estimation) จากตารางที่ 6.1 ในบทที่ 6 การแจกแจงตัวอย่างที่พึงเป็นไปได้ทั้งหมดจะพบว่า ค่าเฉลี่ยตัวอย่าง ($\overline{X}$) เป็นตัวประมาณค่า μ ของประชากร จะได้เป็นสมการ (7-2)

$$\overline{X} \approx \mu \qquad \text{...(7-2)}$$

ถ้า $\overline{X}$ ของ $X_1, X_2, ...X_n$ ที่มีตัวอย่างจำนวน n ค่าจากประชากรที่มีความน่าจะเป็นของ X เท่ากันทุกค่า จะได้

$$E(\overline{X}) = E\left(\frac{\sum X}{n}\right) \qquad \text{...(7-3)}$$

และจากสมการ (5-26) ในบทที่ 5 เมื่อ $c = \frac{1}{n}$ จะได้

$$E(\overline{X}) = \frac{1}{n} E \sum X$$

$$E(\overline{X}) = \frac{1}{n} . E(X_1 + X_2 + \ldots + X_n)$$

$$E(\overline{X}) = \frac{1}{n}.[E(X_1) + E(X_2) + \ldots + E(X_n)]$$

$$= \frac{1}{n}[\, n\, E(X)\,]$$

$$= E(X) = \mu$$

7.1.2 การประมาณค่าความแปรปรวน และค่าเบี่ยงเบนมาตรฐานประชากร (population variance and standard deviation estimation) ความแปรปรวนประชากร จะหาได้จากสมการ (7-4)

$$\sigma^2 = \frac{\sum(X-\mu)^2}{M} \qquad \ldots(7\text{-}4)$$

X คือ ค่าสังเกตประชากร

μ คือ ค่าเฉลี่ยประชากรขนาด M

เปรียบเทียบกับสมการ (3-9) ในบทที่ 3 ถ้าเรามีตัวอย่างสุ่มขนาด n ประกอบด้วย X_1 , X_2 , ..., X_n มีค่าเฉลี่ยเท่ากับ $\overline{X}$ กรณีไม่ทราบค่าเฉลี่ยประชากรเราสามารถประมาณค่า μ ด้วย $\overline{X}$ ได้ในขอบเขตความไม่อคติดังที่กล่าวมาแล้ว ในทำนองเดียวกันถ้าไม่ทราบค่าความแปรปรวนประชากร จะใช้ความแปรปรวนตัวอย่าง (sample variance) สัญลักษณ์ S^2 สำหรับค่าเบี่ยงเบนมาตรฐานตัวอย่างใช้สัญลักษณ์ S ที่มีอยู่ในมือแทนได้หรือไม่ภายในขอบเขตความไม่อคติเราจะต้องดูว่า $E(S^2) = \sigma^2$ หรือไม่ ถ้า $E(S^2) \neq \sigma^2$ แสดงว่าเราใช้ S^2 แทน σ^2 อย่างอคติ พิจารณาตารางที่ 7.2 ซึ่งใช้ตัวอย่างตามตารางที่ 6.1 บทที่ 6 กรณีที่ใช้

$$S^2 = \frac{\sum(X-\overline{X})^2}{n} \qquad \ldots(7\text{-}5)$$

จะได้ $E(S^2) = \dfrac{40}{16} = 2.5$

และในกรณีที่ใช้ $S^2 = \dfrac{\sum(X-\overline{X})^2}{n-1}$ …(7-6)

จะได้ $E(S^2) = \dfrac{80}{16} = 5$

การคำนวณค่าความแปรปรวนตัวอย่างของแต่ละตัวอย่างที่ถูกต้องตามนิยามความแปรปรวนจากสมการ (3-9) ในบทที่ 3 ควรเป็นไปตามสมการ (7-5) แต่ผลคือ $E(S^2) = 5$ ซึ่งไม่เป็นความแปรปรวนประชากร ดังนั้นถ้าใช้ S^2 กรณีนี้จะเป็นการประมาณค่า σ อย่างอคติ เมื่อเราต้องการทั้งความถูกต้องและความไม่อคติ จึงหาผลเฉลี่ยของ S^2 นี้ ด้วยการหารความแปรปรวนตัวอย่างทุกค่า X ด้วย $\frac{(n-1)}{2}$ ค่า (n-1) เรียกว่า ระดับขั้นความเสรีหรือองศาเสรี (degree of freedom) ใช้สัญลักษณ์ df โดยทั่วไป ผลรวมการเบี่ยงเบน $\sum\left(X-\bar{X}\right)$ จะเป็นศูนย์เสมอ ดังนั้นถ้าเราบอกการเบี่ยงเบนของ X จำนวนที่ n –1 เราจะบอกการเบี่ยงเบนของ X จำนวนที่ n ได้ดังสมการ (7-7)

$$S^2 = \left(\frac{n}{n-1}\right)\frac{\sum\left(X-\bar{X}\right)^2}{n} \qquad \text{…(7-7)}$$

หรือ

$$S^2 = \frac{\sum\left(X-\bar{X}\right)^2}{n-1}$$

ซึ่งเหมือนกับสมการ (7-6) นั่นเอง

จากสมการ (7-6) มีผลทำให้ $E(S^2) = 5$ เท่ากับความแปรปรวนประชากร จึงถือว่าความแปรปรวนตัวอย่างที่หาได้ด้วยสมการ (7-6) สามารถนำไปประมาณค่าความแปรปรวนประชากรอย่างไม่อคติ ความจริงข้อนี้ได้รับการพิสูจน์ยืนยันโดยนักสถิติจำนวนหลายท่านและสามารถพิสูจน์ได้ตั้งแต่ตัวอย่างขนาด $n = 2$ ถึงขนาด $n = 200$ จึงเลือกใช้สมการ (7-6) สำหรับหาความแปรปรวนตัวอย่างขนาด n ใด ๆ เพื่อการประมาณค่าความแปรปรวน ประชากรของสาขาสถิติอนุมานและจากสมการ (7-6) สามารถหาค่าเบี่ยงเบนมาตรฐาน ตัวอย่างเมื่อตัวอย่างมีขนาด n ได้เป็นสมการ (7-8)

$$S = \sqrt{\frac{\sum\left(X-\bar{X}\right)^2}{n-1}} \qquad \text{…(7-8)}$$

เมื่อ X เป็นค่าสังเกตตัวอย่าง และ $\bar{X}$ เป็นค่าเฉลี่ยตัวอย่างสุ่มขนาด n

ตารางที่ 7.2 ความแปรปรวนของตัวอย่างสุ่มขนาด n = 2 จาก ตารางที่ 6.1 ในบทที่ 6

ตัวอย่างที่	X_1	X_2	$\overline{X}$	$S^2 = \dfrac{\sum(X-\overline{X})^2}{n}$	$S^2 = \dfrac{\sum(X-\overline{X})^2}{n-1}$
1	1	1	1	0	0
2	1	3	2	1	2
3	1	5	3	1	8
4	1	7	4	9	18
5	3	1	2	1	2
6	3	3	3	0	0
7	3	5	4	1	2
8	3	7	5	4	8
9	5	1	3	4	8
10	5	3	4	1	2
11	5	5	5	0	0
12	5	7	6	1	2
13	7	1	4	9	18
14	7	3	5	4	8
15	7	5	6	1	2
16	7	7	7	0	0
			รวม	40	80

จากสมการ (3-10) ในบทที่ 3 ค่า $\sum(X-\overline{X})^2$ เรียกว่าผลบวกกำลังสองของการเบี่ยงเบน (sum of the squared deviation) เขียนแทนด้วย *SS*. และเรียกความแปรปรวน S^2 ว่าค่าเฉลี่ยของผลบวกกำลังสองของการเบี่ยงเบน (mean of the sum squared deviation : *MS*.) ประกอบกับ *df* คือ n-1 ทำให้เขียนสมการ (7-6) ได้เป็นสมการ (7-9)

$$S^2 = \frac{SS}{df} = MS. \qquad \ldots(7\text{-}9)$$

และเขียนสมการ (7-8) ได้เป็นสมการ (7-10)

$$S = \sqrt{\frac{SS}{df}} = \sqrt{MS} \qquad \ldots(7\text{-}10)$$

เมื่อ $SS = \sum\left(X - \overline{X}\right)^2$

$df = n - 1$, n คือ ขนาดตัวอย่าง

7.1.3 การใช้ค่าเบี่ยงเบนมาตรฐานตัวอย่าง (used of sample standard deviation) การแจกแจงปรกติมาตรฐานของค่าเฉลี่ยตัวอย่างสุ่มขนาด n จากประชากรที่มีการแจกแจงปรกติหรือตัวอย่างสุ่มขนาดใหญ่ ($n \geq 20$) หากเราทราบค่าเฉลี่ยประชากร (μ) และค่าเบี่ยงเบนมาตรฐานประชากร (σ) จะสามารถหาค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่าง ($\sigma_{\overline{X}}$) ได้ดังสมการ (6-9) ในบทที่ 6 แต่กรณีไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร จึงใช้ค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) ที่คำนวณได้ด้วยสมการ (7-8)หรือ (7-10) แทน ทำให้เขียนสมการค่ามาตรฐาน (Z) จากสมการ (6-11) ในบทที่ 6 ได้เป็นสมการ (7-11)

$$Z = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \qquad \ldots(7\text{-}11)$$

และ $\sigma_{\overline{X}} = \dfrac{S}{\sqrt{n}}$ …(7-12)

เพื่อให้เห็นว่าค่า $\sigma_{\overline{X}}$ นี้เราประมาณค่าด้วย $\dfrac{S}{\sqrt{n}}$ จึงใช้สัญลักษณ์ให้แตกต่างออกไปด้วย $S_{\overline{X}}$ จึงได้สมการ (7-13)

$$S_{\overline{X}} = \frac{S}{\sqrt{n}} \qquad \ldots(7\text{-}13)$$

และ $Z = \dfrac{\overline{X} - \mu}{S_{\overline{X}}}$ …(7-14)

สำหรับใช้ในการแจกแจงปรกติมาตรฐานของค่าเฉลี่ยตัวอย่าง จากประชากรที่ทราบค่าเฉลี่ยประชากร

ตัวอย่างที่ 7.1 น้ำหนักรถบรรทุกสินค้าที่วิ่งผ่านถนนสายเอเซียจากนครสวรค์ไปบางประอินจำนวน 100 คัน คำนวณค่าเบี่ยงเบนมาตรฐาน ได้เท่ากับ 1,750 กิโลกรัม จงประมาณค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่างสุ่มทั้งหมด

วิธีทำ ค่าประมาณค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่างสุ่มคือ $S_{\bar{x}}$

จาก $S_{\bar{x}} = \frac{S}{\sqrt{n}}$, S = 1,750 และ n = 100

$$= \frac{1,750}{\sqrt{100}} = 175 \text{ กิโลกรัม}$$

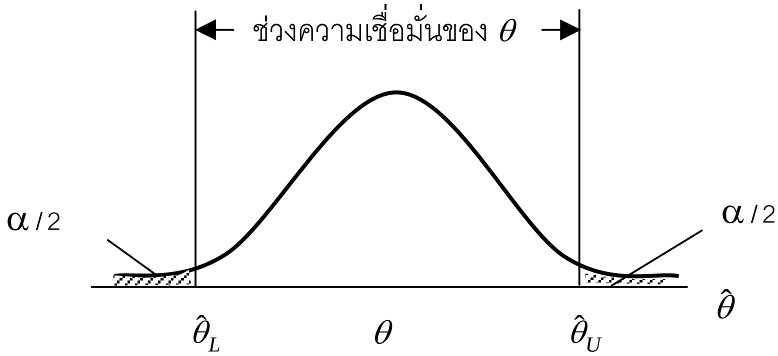
นั่นคือ ค่าเฉลี่ยน้ำหนักรถบรรทุกสุ่มจะมีค่าระหว่าง ค่าเฉลี่ยประชากร -175 กับค่าเฉลี่ยประชากร +175 กิโลกรัม

7.2 การประมาณค่าด้วยช่วงความเชื่อมั่น

การแจกแจงตัวสถิติในลักษณะแบบการแจกแจงใด ๆ ถ้ากำหนดให้ (1-α) เป็นความน่าจะเป็นสูงสุดระหว่างตัวสถิติค่าต่ำ ($\hat{\theta}_L$) กับตัวสถิติค่าสูง ($\hat{\theta}_U$) ซึ่งมีพารามิเตอร์ (θ) ดังภาพที่ 7.1 จะได้ความสัมพันธ์เป็นดังสมการ (7-15)

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1-\alpha \qquad \text{...(7-15)}$$

ช่วงระหว่าง $\hat{\theta}_L$ กับ $\hat{\theta}_U$ เรียกว่า ช่วงความเชื่อมั่นของพารามิเตอร์ (confidence interval of parameter or C. I. of θ) หรือ ขอบเขตความเชื่อมั่นของพารามิเตอร์ตัวนั้น โดยทั่วไปนิยมกำหนดช่วงความเชื่อมั่นเป็นร้อยละ จะได้เป็นสมการ (7-16)



ภาพที่ 7.1 ช่วงความเชื่อมั่นของพารามิเตอร์ (θ)

$$\text{C.I. ของ } \theta = 100\,(1-\alpha)\,\%\text{ ของ } \theta \qquad \ldots(7\text{-}16)$$

ความน่าจะเป็น (1-α) เรียกว่า ระดับความเชื่อมั่น (confidence level) หรือสัมประสิทธิ์ความเชื่อมั่น (coefficient of confidence) สำหรับความน่าจะเป็น α เรียกว่า ระดับนัยสำคัญ (significance level) หรือขนาดของการทดสอบ (size of test)

พิจารณาภาพที่ 7.1 จะพบว่า ช่วงความเชื่อมั่นของพารามิเตอร์ตัวใด เมื่อกำหนด α ก่อนแล้วแบ่งความน่าจะเป็น α ออกเป็น 2 ส่วนเท่า ๆ กัน ให้อยู่ทางส่วนหางทางซ้ายและขวาของการแจกแจงตัวสถิติตัวนั้น ในทางปฏิบัตินักสถิติและนักวิจัยจะกำหนดระดับนัยสำคัญโดยเฉพาะในกระบวนการทดสอบสมมุติฐานเป็น 3 ระดับ คือ

7.2.1 กำหนด α = .01 เรียกว่า มีนัยสำคัญยิ่ง (high significance) และใช้เครื่องหมาย ** (ดอกจัน 2 ดอก) สำหรับการประมาณค่าหรือการทดสอบครั้งนั้น

7.2.2 กำหนด α = .05 เรียกว่า มีนัยสำคัญ (significance) และใช้เครื่องหมาย * (ดอกจัน 1 ดอก) สำหรับการประมาณค่าหรือการทดสอบครั้งนั้น

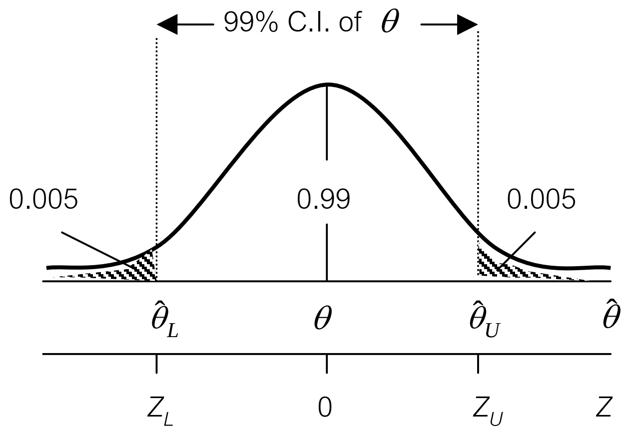
7.2.3 กำหนด α > .05 เรียกว่า ไม่มีนัยสำคัญ (non-significance) และใช้เครื่องหมาย *ns*

สำหรับช่วงความเชื่อมั่นหรือขอบเขตความเชื่อมั่นของพารามิเตอร์ใด ๆ จะพบว่าถ้าช่วงความเชื่อมั่นสูงจะมีความผิดพลาดในการประมาณค่าพารามิเตอร์ตัวนั้นน้อยแต่ค่า พารามิเตอร์ที่ประมาณได้จะมีประโยชน์น้อยลงไปด้วยเช่น สมมุติว่าท่านเห็นชายผู้หนึ่งเดินมา อาจประมาณอายุของเขาเป็น 30-35 ปี (กรณีแรก) หรือประมาณอายุเขาเป็น 30 – 32 ปี (กรณีหลัง) จะเห็นว่าการประมาณกรณีแรกมีโอกาสผิดน้อยกว่ากรณีหลัง เพราะกำหนดช่วงการประมาณไว้กว้างกว่า

กรณีหลัง ถ้าประมาณอายุเขาเป็น 20-60 ปี โอกาสประมาณผิดก็ยิ่งน้อยลงหรือไม่ผิดเลย แต่ค่าประมาณที่ได้ไม่สามารถนำไปใช้ประโยชน์ได้ เป็นต้น

ตัวอย่างที่ 7.2 ตัวสถิติ ($\hat{\theta}$) มีการแจกแจงปรกติเป็นดังภาพที่ 7.2 จงหาช่วงความเชื่อมั่น 99 % ของ พารามิเตอร์ (θ)

วิธีทำ การแจกแจงปรกติของตัวสถิติ ($\hat{\theta}$) เป็นดังภาพที่ 7.2

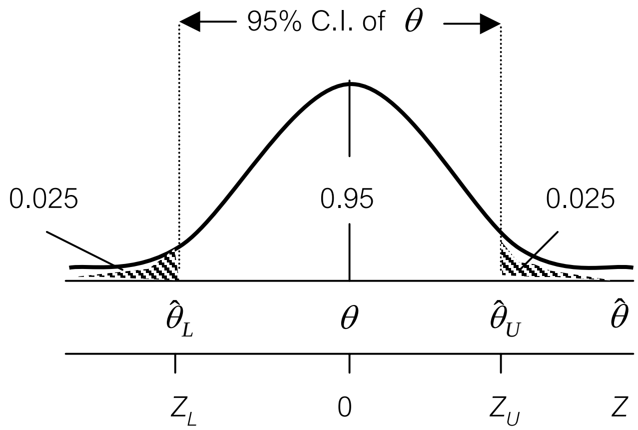


ภาพที่ 7.2 การแจกแจงปรกติ ในช่วงความเชื่อมั่น 99 % ของพารามิเตอร์ (θ)

จากตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 พบว่าพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน 0.99 หน่วย อยู่ระหว่าง Z = -2.575 กับ Z = + 2.575
นั่นคือช่วงความเชื่อมั่น 99 % ของ θ คือ $-2.575 < Z < 2.575$

ตัวอย่างที่ 7.3 ตัวสถิติแจกแจงตามตัวอย่างที่ 7.2 จงหาช่วงความเชื่อมั่น 95% ของพารามิเตอร์ (θ)

วิธีทำ การแจกแจงปรกติของตัวสถิติ ($\hat{\theta}$) เป็นดังภาพที่ 7.3



ภาพที่ 7.3 การแจกแจงปรกติ ในช่วงความเชื่อมั่น 95% ของพารามิเตอร์ θ

95 % C.I. ของ θ = 100 $(1 - \alpha)$ % ของ θ

ดังนั้น $\alpha = 0.05$ และ $\alpha/2 = 0.025$

$$(1-\alpha) = (1-0.05)$$

$$= 0.95$$

จากตารางพื้นที่ใต้เส้นโค้งปรกติมาตรฐาน ตารางภาคผนวกที่ 5 พบว่าพื้นที่ใต้โค้งปรกติมาตรฐาน 0.95 หน่วย อยู่ระหว่าง $Z_L = -1.96$ กับ $Z_U = +1.96$

ช่วงความเชื่อมั่น 95 % ของ θ คือ $-1.96 < Z < 1.96$

7.3 การประมาณค่าเฉลี่ยประชากรด้วยช่วงความเชื่อมั่น

หัวข้อ 7.1.1 เราทราบว่าตัวประมาณค่า μ ของประชากรที่ดีคือ $\overline{X}$ สำหรับการประมาณค่า $\overline{X}$ ด้วยช่วงหรือช่วงความเชื่อมั่นจะนำการแจกแจงของ $\overline{X}$ มาใช้มี 2 กรณีคือ กรณีตัวอย่างขนาดใหญ่และกรณีตัวอย่างขนาดเล็ก

7.3.1 กรณีตัวอย่างมีขนาดใหญ่ สำหรับกรณีตัวอย่างขนาดใหญ่ ($n \geq 20$) หรือประชากรแจกแจงปรกติ ตัวอย่างจะมีการแจกแจงปรกติและนำเข้าสู่การแจกแจงปรกติมาตรฐานได้คือสมการ (7-17)

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \text{...(7-17)}$$

ซึ่งเป็นเช่นเดียวกับสมการ (6-11) ในบทที่ 6 เมื่อ

μ คือ ค่าเฉลี่ยประชากร

$\overline{X}$ คือ ค่าเฉลี่ยตัวอย่างขนาด n

และจากตัวอย่างที่ 7.3 เราได้ช่วงความเชื่อมั่น 95% ของ θ อยู่ระหว่าง Z เท่ากับ -1.96 กับ 1.96 ตามสมการ (7-18)

$$-1.96 < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96 \qquad \text{...(7-18)}$$

คูณตลอดด้วย $\frac{\sigma}{\sqrt{n}}$

$$-1.96 \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1.96 \frac{\sigma}{\sqrt{n}}$$

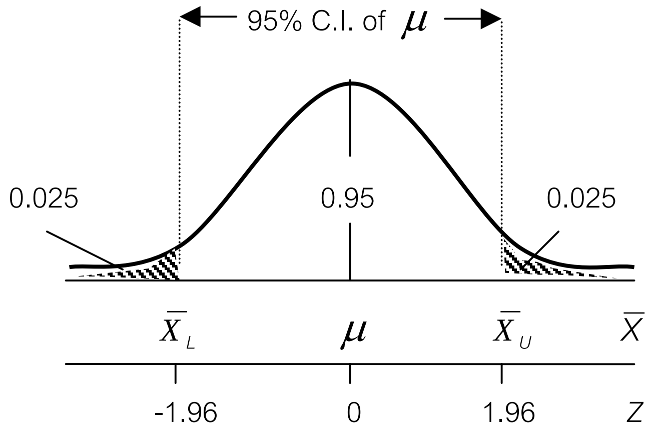
แล้วนำ $\overline{X}$ มาลบออกจากแต่ละเทอมและคูณด้วย (-1) จะได้สมการ (7-19)

$$\overline{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96 \frac{\sigma}{\sqrt{n}} \qquad \text{...(7-19)}$$

สมการ (7-19) นี้บอกให้เราทราบว่า ค่าประมาณของค่าเฉลี่ยประชากรนี้มีค่าระหว่าง $\overline{X} - 1.96 \frac{\sigma}{\sqrt{n}}$ กับ $\overline{X} + 1.96 \frac{\sigma}{\sqrt{n}}$ ที่ช่วงความเชื่อมั่น 95 %หรือเชื่อได้ 95 %

ในทางปฏิบัติเราไม่ทราบค่า σ ของประชากรจึงต้องประมาณค่า σ ของประชากรด้วยค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) ที่คำนวณได้ด้วยสมการ (7-8) หรือ (7-10) จะได้เป็นสมการ (7-20)

$$\overline{X} - 1.96 \frac{S}{\sqrt{n}} < \mu < \overline{X} + 1.96 \frac{S}{\sqrt{n}} \qquad \text{...(7-20)}$$



ภาพที่ 7.4 ช่วงความเชื่อมั่น 95 % ของ μ

ดังนั้นสรุปได้ว่า ช่วงความเชื่อมั่น 95 % ของ μ หรือ 95 % C.I. of μ คือสมการ (7-21) และ สมการ (7-22)

$$\mu = \bar{X} \pm 1.96 \frac{S}{\sqrt{n}} \qquad \text{...(7-21)}$$

หรือ

$$\mu = \bar{X} \pm (Z_{.025}) \frac{S}{\sqrt{n}} \qquad \text{...(7-22)}$$

เมื่อ $Z_{.025}$ หรือ $Z_{.05/2}$ คือค่า Z ที่ตัดพื้นที่ใต้โค้งปรกติมาตรฐานหางทางขวาออก 0.025 หน่วยจะได้เป็น $Z = 1.96$ และค่า Z ที่ตัดพื้นที่ใต้เส้นโค้งปรกติมาตรฐานหางทางซ้ายออก 0.025 หน่วยมีค่าเป็น $Z = -1.96$

ที่ระดับ α อื่น ๆ (รวมถึงช่วงความเชื่อมั่นอื่น ๆ ด้วย) สามารถเขียนสมการ (7-22) ได้เป็น สมการ (7-23) หรือสมการ (7-24)

$$\mu = \bar{X} \pm (Z_{\alpha/2}) \frac{S}{\sqrt{n}} \qquad \text{...(7-23)}$$

หรือ

$$\bar{X} - (Z_{\alpha/2}) \frac{S}{\sqrt{n}} < \mu < \bar{X} + (Z_{\alpha/2}) \frac{S}{\sqrt{n}} \qquad \text{...(7-24)}$$

ตัวอย่างที่ 7.4 สุ่มตัวอย่างเครื่องชั่งน้ำหนักที่ขายในท้องตลาดจำนวน 40 เครื่อง มีค่าเฉลี่ยการบอกน้ำหนักผิดพลาดจากมาตรฐาน 52.2 กรัม ค่าเบี่ยงเบนมาตรฐาน 8.52 กรัม จงประมาณค่าเฉลี่ยการบอกน้ำหนักผิดพลาดของเครื่องชั่งน้ำหนักทั้งหมด ด้วยช่วงความเชื่อมั่น 95%

วิธีทำ กำหนด μ เป็นค่าเฉลี่ยการบอกน้ำหนักผิดพลาดจากมาตรฐานของเครื่องชั่งทั้งหมด

จากสมการ (7-21) ช่วงความเชื่อมั่น 95% ของ μ คือ

$$\mu = \overline{X} \pm 1.96 \frac{S}{\sqrt{n}}$$

$$\overline{X} = 52.2 , S = 8.52 , n = 40$$

ดังนั้น $$\mu = 52.2 \pm (1.96) \frac{8.52}{\sqrt{40}}$$

$$= 52.2 \pm (1.96) \frac{8.52}{6.32}$$

$$= 52.2 \pm 2.64$$

$$= (49.56 , 54.84)$$

นั่นคือ ค่าเฉลี่ยการบอกน้ำหนักผิดพลาดของเครื่องชั่งทั้งหมดที่มีขายในท้องตลาด มีค่าระหว่าง 49.56 กรัม กับ 54.84 ด้วยช่วงความเชื่อมั่น 95% หรือ เชื่อได้ 95 %

ตัวอย่างที่ 7.5 สุ่มชั่งน้ำหนักรถบรรทุกสินค้าที่วิ่งผ่านถนนสายธนบุรี - ปากท่อ จำนวน 120 คัน พบว่ามีน้ำหนักเฉลี่ย 25 ตัน, ค่าเบี่ยงเบนมาตรฐาน 1.75 ตัน จงประมาณน้ำหนักเฉลี่ยรถบรรทุกสินค้าที่วิ่งผ่านถนนสายนี้ทั้งหมดที่เชื่อได้ 95%

วิธีทำ กำหนดให้ μ เป็นน้ำหนักเฉลี่ยรถบรรทุกสินค้าทั้งหมวดที่วิ่งผ่านถนนสายธนบุรี - ปากท่อ

จากสมการ (7-21) ช่วงความเชื่อมั่น 95% ของ μ คือ

$$\mu \quad = \quad \overline{X} \pm 1.96 \frac{S}{\sqrt{n}}$$

$$\overline{X} \quad = \quad 25 \text{ , } S = 1.75 \text{ , } n = 120$$

ดังนั้น

$$\mu \quad = \quad 25 \pm (1.96)\frac{1.75}{\sqrt{120}}$$

$$\mu \quad = \quad 25 \pm (1.96)\frac{1.75}{10.95} \quad = \quad 25 \pm 0.3132$$

$$= \quad (24.6868 \text{ , } 25.3132)$$

นั่นคือ น้ำหนักเฉลี่ยรถบรรทุกสินค้าทั้งหมดที่วิ่งผ่านถนนสายธนบุรี - ปากท่อระหว่าง 24.6868 ตัน กับ 25.3132 ตัน เชื่อได้ 95 %

7.3.2 กรณีที่ตัวอย่างขนาดเล็ก เราทราบมาแล้วว่า ถ้าตัวอย่างขนาดใหญ่หรือประชากรแจกแจงปรกติ ค่าเฉลี่ยตัวอย่างที่พึงเป็นไปได้ทั้งหมดมีการแจกแจงปรกติและ นำไปสู่การแจกแจงปรกติมาตรฐานได้ ค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) สามารถใช้แทนค่าเบี่ยงเบนมาตรฐานประชากร (σ) ได้อย่างเหมาะสม แต่ในกรณีที่ตัวอย่างขนาดเล็ก การแจกแจงของค่าเฉลี่ยตัวอย่างที่ใช้ค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) แทนค่าเบี่ยงเบนมาตรฐานประชากร (σ) มีลักษณะการแจกแจงไม่ปรกติ แต่ยังรักษาความสมมาตรอยู่ ศึกษาและนำเสนอโดยวิลเลียม เอส กอสเสต (William S. Gossett : 1876-1937) ด้วยการศึกษาการแจกแจงค่าเฉลี่ยตัวอย่างขนาดเล็ก ($n < 20$) หลาย ๆ ค่า (Watson et al.,1990, p. 33) ท่านนำเสนอความคิดเห็นนี้ในนามปากกาว่า สติวเดนต์ (student)" จึงเรียกการแจกแจงนี้ว่า การแจกแจงที

ถ้า X_1 , X_2 , X_3 , ..., X_n เป็นค่าสังเกตของตัวอย่างขนาดเล็ก ($n < 20$ จากค่าของ CLT. ที่กล่าวในหัวข้อ 6.7 บทที่ 6 ประชากรที่มีการแจกแจงปรกติค่าเฉลี่ย μ เป็น

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

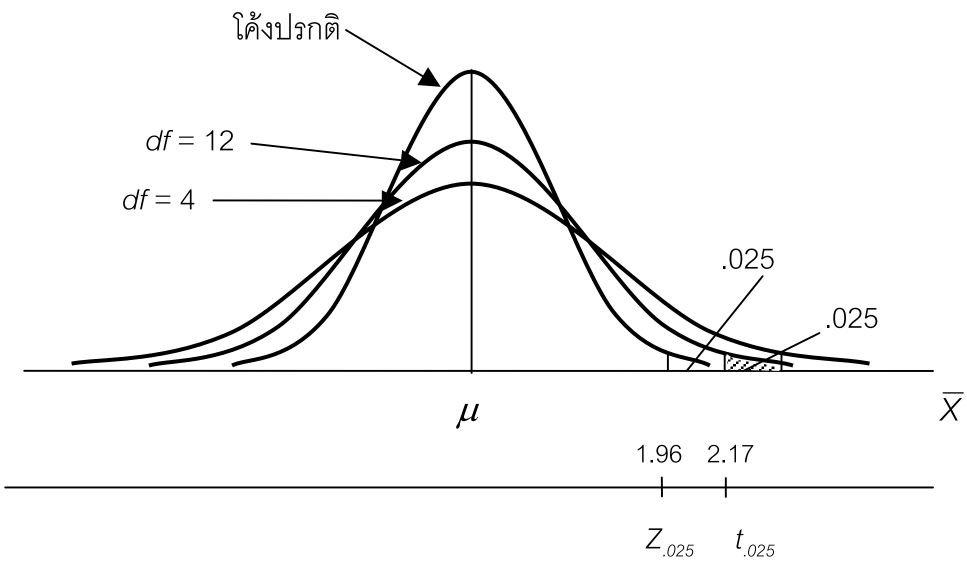
และ

$$S \quad = \quad \sqrt{\frac{\sum (X - \overline{X})^2}{n - 1}}$$

แล้วจะได้ $$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$ …(7-25)

สมการ (7-25) นี้จะบอกให้เราทราบว่า t คือ ความแตกต่าง ระหว่างค่าเฉลี่ยตัวอย่างกับค่าเฉลี่ยประชากร เมื่อเทียบกับค่าผิดพลาดมาตรฐาน ($S_{\bar{x}}$) ซึ่งประมาณค่าด้วยค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) หารด้วย $\sqrt{n}$ และใช้ได้กรณีเดียวคือตัวอย่างมากจากประชากรแจกแจงปรกติเท่านั้น

ช่วงความเชื่อมั่นการแจกแจงของ $\bar{X}$ ที่รวมถึงการแจกแจงของตัวสถิติอื่น ๆ ที่มีรูปแบบเป็นการแจกแจงทีซึ่งขึ้นอยู่กับค่าสังเกตจำนวนที่ n-1 ของตัวอย่างหรือ df ตัวอย่าง และเมื่อ n เข้าสู่อนันต์ การแจกแจงทีของ X นี้จะกลายเป็นการแจกแจงปรกติมาตรฐาน ดังภาพที่ 7.5



ภาพที่ 7.5 การแจกแจงทีของ $\bar{X}$ ที่ df แตกต่างกัน
ที่มา : (Watson et al., 1990, p. 332)

จากภาพที่ 7.5 จะพบว่าช่วงความเชื่อมั่น 95 % ของ μ และ ค่า $Z_{.025}$ มีค่าน้อยกว่าค่า $t_{.025,df}$ เมื่อนำค่า $t_{.025,df}$ ไปใช้แทนค่า $Z_{.025}$ ในสมการ (7-22) จะได้ช่วงความเชื่อมั่น 95 % ของ μ คือ

$$\mu = \bar{X} \pm (t_{.025,df}) \frac{S}{\sqrt{n}} \qquad \text{...(7-26)}$$

โดยค่า $t_{.025}$ หรือ $t_{.05/2}$ อื่น ๆ หาได้จากตารางภาคผนวกที่ 6 และ df ก็คือ ค่าเดียวกันกับที่ใช้คำนวณค่าเบี่ยงเบนมาตรฐานตัวอย่างนั่นเอง พึงสังเกตว่าค่า $t_{.025}$ เหมือนกันแต่ df จะต่างกันและเมื่ออ่านค่า $t_{.025}$ จะได้ไม่เท่ากัน เช่น ตัวอย่างขนาด $n = 5$ จะมี $df = 4$ อ่านค่า $t_{.025,4}$ ได้เท่ากับ 2.776 สำหรับตัวอย่างขนาด $n = 13$ จะมี $df = 12$ อ่านค่า $t_{.025,12}$ ได้เท่ากับ 2.179 เป็นต้น

ตัวอย่างที่ 7.6 จงประมาณค่าเฉลี่ยความยาวรากถั่วเขียวในธรรมชาติเพาะเมล็ดเป็นเวลา 2 วัน ที่ช่วงความเชื่อมั่นได้ 95 % โดยสุ่มวัดความยาวรากถั่วเขียวจำนวน 3 ราก ได้ความยาวเป็น 1.7 นิ้ว 2.4 นิ้วและ 2.5 นิ้ว

วิธีทำ กำหนด μ เป็นค่าเฉลี่ยความยาวรากถั่วเขียวในธรรมชาติเพาะเมล็ดเป็นเวลา 2 วัน จากสมการ (7-26) ช่วงความเชื่อมั่น 95% ของ μ คือ

$$\mu = \bar{X} \pm (t_{.025,2}) \frac{S}{\sqrt{n}}$$

เมื่อ $n = 3, \quad df = 3-1 = 2$

ดังนั้น $\bar{X} = (1.7 + 2.4 + 2.5)/3 = 2.2$

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

$$= \frac{(1.7 - 2.2)^2 + (2.4 - 2.2)^2 + (2.5 - 2.2)^2}{2}$$

$$= 0.19$$

และ $S = \sqrt{0.19}$

$= 0.44$

จากตารางภาคผนวกที่ 6 ที่ $df = 2$, $t_{.025} = 4.303$

จะได้ $$\mu = 2.2 \pm 4.30 \frac{0.44}{\sqrt{3}}$$

$$= 2.2 \pm 1.09$$

$$= (1.11 , 3.29)$$

สรุปว่าค่าเฉลี่ยความยาวรากถั่วเขียวในธรรมชาติเพาะเมล็ดเป็นเวลา 2 วันจะมีค่าระหว่าง 1.11 นิ้ว ถึง 3.29 นิ้ว เชื่อได้ 95 %

ตัวอย่างที่ 7.7 วิทยาลัยเอกชนแห่งหนึ่งต้องการทราบรายได้ต่อสัปดาห์ของผู้สำเร็จการศึกษาระดับ ชั้น ปวช. (ทุกสาขาวิชา) และมีงานทำ โดยสุ่มถามผู้สำเร็จการศึกษาระดับชั้น ดังกล่าว จำนวน 16 คน พบว่ามีรายได้เฉลี่ยสัปดาห์ละ 800 บาท ค่าเบี่ยงเบนมาตรฐาน 120 บาทจงประมาณรายได้ของผู้สำเร็จการศึกษาชั้น ปวช. ทั้งหมดที่ช่วงความเชื่อมั่น 95 %

วิธีทำ กำหนด μ เป็นรายได้เฉลี่ยต่อสัปดาห์ของผู้สำเร็จการศึกษาระดับชั้น ปวช. ทั้งหมด จากสมการ (7-26) ช่วงความเชื่อมั่น 95% ของ μ คือ

$$\mu = \overline{X} \pm (t_{.025,df}) \frac{S}{\sqrt{n}}$$

เมื่อ $n = 16,\ df = 16\text{-}1 = 15$

$\overline{X} = 800$

และ $S = 120$

จากตารางภาคผนวกที่ 6 ที่ $df = 15$, $t_{.025} = 2.131$

ดังนั้น $$\mu = 800 \pm (2.13)\ \frac{120}{14}$$

$$= 800 \pm 63.9$$

$$\mu = (736.1\ ,863.9)$$

นั่นคือรายได้ต่อสัปดาห์ของผู้สำเร็จการศึกษาระดับชั้น ปวช. (ทุกสาขาวิชา) อยู่ระหว่าง 736.1 กับ 863.9 บาท เชื่อได้ 95 %

7.4 การประมาณค่าเฉลี่ยแตกต่างของสองประชากรด้วยช่วงความเชื่อมั่น

หากเรากำลังสนใจประชากร 2 ประชากร โดยประชากรแรกมีค่าเฉลี่ยเท่ากับ μ_1 ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_1 ส่วนประชากรหลังมีค่าเฉลี่ยเท่ากับ μ_2 ค่าเบี่ยงเบนมาตรฐานเท่ากับ σ_2 ค่าเฉลี่ยแตกต่างกันของ 2 ประชากรนี้ คือ (μ_1-μ_2) ตัวอย่างสุ่มขนาด n_1, n_2 จากประชากรแรกและประชากรหลัง ตัวอย่างสุ่มจะมีค่าเฉลี่ยแตกต่างกันเป็น ($\bar{X}_1$-$\bar{X}_2$) ซึ่งในที่นี้จะกล่าวถึงในกรณีต่าง ๆ ดังนี้

7.4.1 กรณีสองตัวอย่างอิสระกัน (independent sample) ตัวอย่างอิสระกัน หมายถึง ค่าสังเกตของตัวอย่างขนาด n_1 ไม่เกี่ยวข้องหรือมีผลกระทบต่อค่าสังเกตของตัวอย่างขนาด n_2 ค่าผิดพลาดมาตรฐานของ $\bar{X}_1$ และ $\bar{X}_2$ คือ

$$\sigma_{\bar{X}_1} = \frac{\sigma_1}{\sqrt{n_1}} \quad \text{และ} \quad V(\bar{X}_1) = \frac{\sigma_1^2}{n_1}$$

$$\sigma_{\bar{X}_2} = \frac{\sigma_2}{\sqrt{n_2}} \quad \text{และ} \quad V(\bar{X}_2) = \frac{\sigma_2^2}{n_2}$$

จะได้ $$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \ldots(7\text{-}27)$$

จากสมการ (3-23) $V(aX) = a^2\ V(X)$

จะได้ $V(aX+bY) = a^2 V(X) + b^2 V(Y)$

ถ้า $a = 1, b = (-1)$ จะได้

$$V(X\text{-}Y) = (1)^2 V(X) + (-1)^2 V(Y)$$

เมื่อ $V(X) = V(\bar{X}_1),\ V(Y) = V(\bar{X}_2)$

และ $$V(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad \ldots(7\text{-}28)$$

โดย $\sigma_{\bar{X}_1 - \bar{X}_2}$ เป็นค่าผิดพลาดมาตรฐานของ $(\bar{X}_1 - \bar{X}_2)$

ถ้า n_1 และ n_2 มีขนาดใหญ่และทราบค่าเบี่ยงเบนมาตรฐานประชากร σ_1 และ σ_2 ช่วงความเชื่อมั่น 95 % ของ $(\mu_1 - \mu_2)$ จากสมการ (7-22) คือสมการ (7-29)

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm Z_{.025} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \ldots(7\text{-}29)$$

จากสมการ (7-29) ถ้าไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร σ_1 และ σ_2 จะประมาณค่าด้วยเบี่ยงเบนมาตรฐานตัวอย่าง S_1 และ S_2 หรือความแปรปรวนตัวอย่าง S_1^2 และ S_2^2 จะได้สมการ (7-30)

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm Z_{.025} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \qquad \ldots(7\text{-}30)$$

สำหรับในกรณีที่ทราบว่าประชากรทุกประชากรแจกแจงปรกติหรือประชากรแจกแจงปรกติโดยประมาณ ค่าเบี่ยงเบนมาตรฐาน $\sigma_1 = \sigma_2 = \sigma$ สมการ (7-29) เขียนได้เป็นสมการ (7-31)

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm Z_{.025}\ \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad \ldots(7\text{-}31)$$

ถ้า n_1 และ n_2 มีขนาดเล็ก (n_1, $n_2 \le 20$) , ประชากรทุกประชากรแจกแจงปรกติหรือแจกแจงปรกติโดยประมาณและไม่ทราบค่า σ ของประชากร จะต้องประมาณค่าด้วยสัญลักษณ์ S_P และเปลี่ยน $Z_{.025}$ เป็น $t_{.(025,df)}$ ทำให้สมการ (7-31) เขียนได้เป็น

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{.(025,df)} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

เมื่อประชากรทั้งสอง มีความแปรปรวน σ^2 เหมือนกัน ค่า S_p^2 จะต้องประมาณโดยอาศัยตัวอย่างทั้งคู่ เราเรียกค่าประมาณนี้ว่าความแปรปรวนร่วม (pooled variance) หาได้โดยใช้ $\sum\left(X_1 - \bar{X}_1\right)^2 + \sum\left(X_2 - \bar{X}_2\right)^2$ แล้วหารด้วย df ของตัวอย่างทั้งคู่คือ $(n_1\text{-}1) + (n_2\text{-}1)$ หรือ $(n_1 + n_2\text{-}2)$ ในทำนองเดียวกันกับการหาความแปรปรวนตัวอย่างเดียว (S^2) จากสมการ (7-6) จะได้สมการ (7-32)

$$S_P^2 = \frac{\sum\left(X_1 - \bar{X}_1\right)^2 + \sum\left(X_2 - \bar{X}_2\right)^2}{\left(n_1 - 1\right) + \left(n_2 - 1\right)} \qquad \ldots(7\text{-}32)$$

แต่ความแปรปรวนตัวอย่างที่ 1 หาค่า

$$S_1^2 = \frac{\sum\left(X_1 - \bar{X}_1\right)^2}{\left(n_1 - 1\right)}$$

จึงได้ $$S_1^2\,(n_1 - 1) = \sum\left(X_1 - \bar{X}_1\right)^2 \qquad \ldots(7\text{-}33)$$

และความแปรปรวนตัวอย่างที่ 2 หาค่า

$$S_2^2 = \frac{\sum\left(X_2 - \bar{X}_2\right)^2}{\left(n_2 - 1\right)}$$

จึงได้ $$S_2^2\,(n_2 - 1) = \sum\left(X_2 - \bar{X}_2\right)^2 \qquad \ldots(7\text{-}34)$$

นำเอาสมการ(7-33) และ (7-34) แทนใน (7-32) จะได้สมการ (7-35)

$$S_P^2 = \frac{\left(n_1 - 1\right)S_1^2 + \left(n_2 - 1\right)S_2^2}{\left(n_1 + n_2 - 2\right)} \qquad \ldots(7\text{-}35)$$

และ $$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}} \qquad \ldots(7\text{-}36)$$

โดย $(n_1 + n_2 - 2)$ เป็น df สำหรับการแจกแจงที

ตัวอย่างที่ 7.8 ต้องการเปรียบเทียบอายุการใช้งานหลอดไฟฟ้าซึ่งผลิตโดยบริษัท ก.กับ บริษัท ข. ที่ช่วงความเชื่อมั่นได้ 95 % สุ่มตัวอย่างหลอดไฟฟ้าจากบริษัท ก. จำนวน 4 หลอดพบว่ามีอายุการใช้งานเป็น 64 , 66 , 89 และ 77 ชั่วโมง หลอดไฟฟ้าจากบริษัท ข. จำนวน 3 หลอดพบว่ามีอายุการใช้งานเป็น 56 , 71 และ 53 ชั่วโมง

วิธีทำ

ตารางที่ 7.3 อายุการใช้งานหลอดไฟฟ้าซึ่งผลิตโดยบริษัท ก.กับ บริษัท ข.

อายุการใช้งานหลอดไฟฟ้า ผลิตโดยบริษัท ก.		
X_1	$(X_1 - \bar{X}_1)$	$(X_1 - \bar{X}_1)^2$
64	-10	100
66	-8	64
89	15	225
77	8	9
296	0	398

อายุการใช้งานหลอดไฟฟ้า ผลิตโดยบริษัท ข.		
X_2	$(X_2 - \bar{X}_2)$	$(X_2 - \bar{X}_2)^2$
56	-4	16
71	11	121
53	-7	49
180	0	186

$$\bar{X}_1 = \frac{296}{4} = 74 \qquad S_1^2 = \frac{398}{3} = 132.66 \qquad \bar{X}_2 = \frac{180}{3} = 60 \qquad S_2^2 = \frac{186}{2} = 93.00$$

คำนวณหาความแปรปรวนร่วม จากสมการ (7-35)

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$$

$$S_P^2 = \frac{(4-1)132.66+(3-1)93.00}{(4+3-2)}$$

$$= \frac{(3)132.66+(2)93.00}{5}$$

$$= \frac{397.98+186}{5} = 116.79$$

$$S_P = \sqrt{116.79}$$

จากตารางภาคผนวกที่ 6

$$df = 5, \quad t_{.025} = 2.571$$

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{.025,df}\, S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(\mu_1 - \mu_2) = (74\text{-}60) \pm 2.57(\sqrt{116.79})\sqrt{\frac{1}{4} + \frac{1}{3}}$$

$$= 14 \pm 21 = (-7 \;, 35)$$

นั่นคือ อายุการใช้งานหลอด/ไฟฟ้าที่ผลิตโดยบริษัท ก. กับบริษัท ข. ($\mu_{ก}$-$\mu_{ข}$) แตกต่างกัน 35 ชั่วโมงหรืออายุการใช้งานหลอดไฟฟ้าที่ผลิตโดยบริษัท ข. กับ บริษัท ก. ($\mu_{ข}$-$\mu_{ก}$) แตกต่างกัน 7 ชั่วโมง เชื่อได้ 95 %

7.4.2 กรณีสองตัวอย่างไม่อิสระกัน (dependent sample) ตัวอย่างไม่อิสระกัน หรืออาจเรียกว่า “ตัวอย่างชนิดคู่” (paired sample) หมายถึง ค่าสังเกตของตัวอย่างขนาด n มีผลกระทบหรือเกี่ยวข้องกับตัวอย่างขนาด n เท่ากับอีกตัวอย่างหนึ่ง เช่น คะแนนสอบ 2 วิชาของนาย ก. นาย ข. นาย ค. และ นาย ง. มักพบว่าคนที่ได้คะแนนวิชาวิทยาศาสตร์สูง มักได้คะแนนวิชาคณิตศาสตร์สูงด้วยการสอบก่อนเรียนกับการสอบหลังจากการเรียนจบแล้ว ของแต่ละคนมักพบว่าคนที่ได้คะแนนก่อนการเรียนสูงมักจะได้คะแนนสอบหลังจากการเรียนจบแล้วสูงด้วย หรือน้ำหนักสุกรขุนตัวที่ 1 ตัวที่ 2 และ ตัวที่ 3 ก่อนให้ฮอร์โมนกับหลังให้ฮอร์โมน เป็นต้น

ในที่นี้จะขอกล่าวเฉพาะตัวอย่างทั้งคู่มีขนาดเล็ก (n_1 = n_2 = $n \le$ 20) จึงต้องอาศัยการแจกแจงที เป็นหลักและอยู่ภายใต้ข้อสมมุติว่า ประชากรทั้งคู่แจกแจงปรกติหรือแจกแจงปรกติโดยประมาณ เริ่มต้นด้วยการหาความแตกต่างค่าสังเกตของทุกหน่วยตัวอย่าง $D = (X_1 - X_2)$ เช่น X_1 คะแนนสอบก่อนเรียน ของนาย ก. X_2 จะเป็นคะแนนสอบหลังจากการเรียนจบแล้วของนาย ก. เช่นกัน เป็นต้น เราจะได้ D เป็นเสมือนตัวอย่างเดียว จากนั้นหาค่าเฉลี่ยความแตกต่างซึ่งเสมือนเป็นค่าเฉลี่ยตัวอย่างเดียว ($\overline{D} = \overline{X}_1 - \overline{X}_2$) และหาความแปรปรวนทำนองเดียวกันได้

$$S_D^2 \quad = \quad \frac{\sum (D - \overline{D})^2}{(n-1)} \qquad \text{…(7-37)}$$

ช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยแตกต่างกันของประชากร(Δ) ซึ่ง

$$\Delta \quad = \quad \mu_1 - \mu_2$$

และจากสมการ (7-26) จะได้เป็นสมการ (7-38)

$$\Delta \quad = \quad \overline{D} \pm (t_{.025,df}) \frac{S_n}{\sqrt{n}} \qquad \text{…(7-38)}$$

หรือ

$$\mu_1 - \mu_2 \quad = \quad (\overline{X}_1 - \overline{X}_2) \pm (t_{.025,df}) \frac{S_n}{\sqrt{n}} \qquad \text{…(7-39)}$$

ตัวอย่างที่ 7.9 จงประมาณค่าเฉลี่ยแตกต่างคะแนนสอบก่อนฝึกอบรมกับหลังฝึกอบรมของพนักงานขายทั้งหมดที่เข้ารับการอบรมวิชาชีพเทคนิคการตลาดเชิงรุก โดยสุ่มสอบผู้เข้ารับการฝึกอบรม ก่อนและหลังฝึกอบรมจำนวน 4 คน คะแนนเป็นดังตารางที่ 7.4

วิธีทำ ที่ $df = 3$, $t_{.025} = 3.182$ ช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยแตกต่างกันของประชากร คือ

ตารางที่ 7.4 คะแนนสอบก่อนฝึกอบรมกับหลังฝึกอบรมของพนักงานขาย

	นาย อ.	นาย พ.	นาย ช.	นาย ท.
ก่อนอบรม	54	54	70	62
หลังอบรม	64	66	89	77

$$\Delta = \overline{D} \pm (t_{.025,df}) \frac{S_n}{\sqrt{n}}$$

$$= 14 \pm 3.18 \frac{\sqrt{15.3}}{\sqrt{4}} \qquad = 14 \pm 6$$

$$\Delta = (8 \ , \ 20)$$

นั่นคือคะแนนก่อนการอบรมและหลังการฝึกอบรมของพนักงานต่างกันระหว่าง 8 ถึง 20 คะแนนด้วยช่วงความเชื่อมั่น 95 %

ตารางที่ 7.5 ความแตกต่างของคะแนนสอบก่อนฝึกอบรมและหลังฝึกอบรมของพนักงานขาย

ความแตกต่างคะแนนก่อนอบรม-หลังอบรม		
$D = (X_1 - X_2)$	$(D - \overline{D})$	$(D - \overline{D})^2$
10	-4	16
12	-2	4
19	5	25
15	1	1
56	0	46

$$\overline{D} = \frac{56}{4} \qquad\qquad S_D^2 = \frac{46}{3}$$

$$= 14 \qquad\qquad = 15.3$$

นั่นคือค่าเฉลี่ยแตกต่างคะแนนก่อนสอบก่อนฝึกอบรมกับหลังฝึกอบรมของพนักงานขายทั้งหมดที่เข้ารับการอบรมฯ อยู่ระหว่าง 8 ถึง 20 คะแนน เชื่อได้ 95 %

7.5 การประมาณค่าสัดส่วนประชากร

สำหรับประชากรขนาดใหญ่ประกอบด้วย X ที่มีค่าเท่ากับ 0 และ 1 ค่าเฉลี่ยของ X เรียกว่า สัดส่วนประชากร (the population proportion : π) จะได้ว่าค่าความแปรปรวนของ X เป็น $n\pi(1-\pi)$ (Wonnacott & Wonnacott, 1985, p. 177) ถ้าชักตัวอย่างสุ่มขนาด n ตัวอย่าง ค่าเฉลี่ยของ X จะเรียกว่า สัดส่วนตัวอย่าง (the sample proportion : p) การแจกแจงสัดส่วนตัวอย่างทำได้เช่นเดียวกับการแจกแจงค่าเฉลี่ยตัวอย่างดังที่กล่าวมาแล้วในหัวข้อ 6.6 นั่นคือ

$$E(p) = \pi$$

$$Var(p) = \pi(1-\pi)/n$$

และใช้ค่าเฉลี่ยตัวอย่างประมาณค่าเฉลี่ยประชากรจากสมการ (7-2) จากนั้นใช้สัดส่วนตัวอย่างประมาณค่าสัดส่วนประชากรแล้วจะได้ดังสมการ (7-40)

$$p \approx \pi \qquad \text{... (7-40)}$$

ในทำนองเดียวกับการประมาณค่าเฉลี่ยประชากรด้วยช่วงความเชื่อมั่น (1-α) จากสมการ (7-23) และสมการ (7-24) จะได้ช่วงความเชื่อมั่น (1- α) ของ π และช่วงความเชื่อมั่น (1- α)100% ของ π เป็นดังสมการ (7-41) และสมการ (7-42)

$$\pi = p \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}} \qquad \text{... (7-41)}$$

หรือ

$$p - Z_{\alpha/2}\sqrt{\frac{pq}{n}} < \pi < p + Z_{\alpha/2}\sqrt{\frac{pq}{n}} \qquad \text{... (7-42)}$$

ตัวอย่างที่ 7.10 สุ่มแจกเอกสารโฆษณา จำหน่ายรถยนต์ 344 ใบ ปรากฏว่ามีผู้ที่ซื้อตามเอกสารโฆษณา 83 ราย จงประมาณค่าสัดส่วนผู้ที่ซื้อรถยนต์ หลังได้รับเอกสารโฆษณาที่เชื่อได้ 90%

วิธีทำ ให้ π = สัดส่วนผู้ที่ซื้อรถยนต์ หลังได้รับเอกสารโฆษณา

เรามี $n = 344$

$$p = \frac{83}{344}$$

$$= 0.241$$

$$Z_{\alpha/2} = Z_{.10/2}$$

$$= Z_{.05}$$

$$= 1.645$$

ดังนั้น ช่วงความเชื่อมั่นที่ 90 % ของ π คือ

$$\pi = p \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$= 0.241 \pm 1.645\sqrt{\frac{(0.241)(0.759)}{344}}$$

$$= (0.203\ ,\ 0.279)$$

นั่นคือช่วงความเชื่อมั่น 90 % ที่ผู้ได้รับเอกสารโฆษณา จะซื้อรถยนต์มีค่าร้อยละ 20.30 ถึง ร้อยละ 27.90

7.6 การประมาณค่าความแปรปรวนประชากร

จากหัวข้อ 7.1 เมื่อสุ่มตัวอย่างขนาด n จากประชากร ตัวประมาณค่าที่ดีของ σ_X^2 คือ S_X^2 หรือ $E(S_X^2) = \sigma_X^2$ โดยใช้สมการ (7-8) จะได้

$$S_X^2 = \frac{1}{n-1}\sum(X-\overline{X})^2$$

หรือ $(n-1)\,S_X^2 = \sum(X-\overline{X})^2$ … (7-43)

$$\frac{(n-1)S_x^2}{\sigma_x^2} = \frac{\sum (X - \overline{X})^2}{\sigma_x^2} \quad \ldots (7\text{-}44)$$

ถ้าสมการ (7-43) หรือสมการ (7-44) มีการแจกแจงไคกำลังสอง (chi-square distribution) ด้วยความเป็นอิสระ n-1 เขียนแทนด้วย $\chi^2_{(n-1)df}$ (Keller & Warrack, 2000, p. 364) สำหรับความเป็นมาของการแจกแจงไคกำลังสอง ในปี 1876 เอฟ อาร์ เฮลเมนต์ (F.R.Helment) ได้นำเสนอการแจกแจงไคกำลังสอง และในปี 1900 คาร์ล เพียร์สัน เป็นผู้พัฒนาการแจกแจงไคกำลังสองนี้ใช้ในการทดสอบสมมุติฐาน (Chao, 1969, p. 269)

กรณีที่ X มีการแจกแจงปรกติด้วยค่าเฉลี่ย μ ความแปรปรวน σ^2 และแปลง X เป็นค่ามาตรฐาน Z

$$Z = \left(\frac{X-\mu}{\sigma_X}\right) \sim N(0,1) \quad \ldots(7\text{-}45)$$

ค่า $Z^2 = \frac{(X - \mu)^2}{\sigma_x^2}$

จะมีการแจกแจงไคกำลังสองด้วย df = 1 ใช้สัญลักษณ์

$$Z^2 = \chi^2_{(1)df} \quad \ldots(7\text{-}46)$$

เมื่อ X_1 และ X_2 ต่างแจกแจงปรกติ ด้วยค่าเฉลี่ย μ_1 ความแปรปรวน σ_1^2 และ μ_2 ความแปรปรวน σ_2^2 ตามลำดับ แปลง X_1 และ X_2 เป็นค่ามาตรฐาน ได้

$$Z_1^2 = \frac{(X_1 - \mu)^2}{\sigma_1^2}$$

และ $Z_2^2 = \frac{(X_2 - \mu)^2}{\sigma_2^2}$

ผลรวม $Z_1^2 + Z_2^2 = \chi^2_{(2)df}$ …(7-47)

มีการแจกแจงไคกำลังสอง $df = 2$

เมื่อมี X_1 , X_2 , … , X_n ด้วยค่าเฉลี่ย μ_1 ความแปรปรวน σ_1^2 ค่าเฉลี่ย μ_2 ความแปรปรวน σ_2^2 . . . และค่าเฉลี่ย μ_n ความแปรปรวน σ_n^2 จะได้

$$\sum_{i=1}^{n} Z_i^2 = \chi^2_{(n-1)df}$$

$$E(\chi_n^2) = (n\text{-}1)df$$

$$Var(\chi_n^2) = 2(n\text{-}1)df \qquad \ldots(7\text{-}48)$$

เมื่อมีการประมาณค่าเฉลี่ยประชากรด้วย $\overline{X}$ ให้พิจารณา

$$\Sigma\left(X-\overline{X}\right)^2 = \sum\left[\left(X-\overline{X}\right)-\left(\overline{X}-\overline{\mu}\right)\right]^2$$

$$= \sum\left[\left(X-\overline{X}\right)^2 - 2\left(X-\overline{X}\right)\left(\overline{X}-\overline{\mu}\right) + \left(\overline{X}-\overline{\mu}\right)^2\right]$$

$$= \sum\left(X-\overline{X}\right)^2 + n\left(\overline{X}-\overline{\mu}\right)^2$$

หารด้วย σ_x^2

$$\frac{\sum(X-\mu)^2}{\sigma_x^2} = \frac{\sum(X-\overline{X})^2}{\sigma_x^2} + \frac{n(X-\mu)^2}{\sigma_x^2}$$

จากเทอมสุดท้ายจะได้

$$\frac{(X-\mu)^2}{\frac{\sigma_x^2}{n}} \sim \chi^2_{(1)df} \qquad \ldots(7\text{-}49)$$

ดังนั้น

$$\frac{\sum(X-\overline{X})^2}{\sigma_x^2} \sim \chi^2_{(n-1)df} \qquad \ldots(7\text{-}50)$$

ค่าของ $\chi^2_{(1)df}$ จากสมการ (7-50) ที่มีช่วงค่าวิกฤตที่ระดับใด ๆ จากตารางภาคผนวกที่ 11 จะมีช่วงเป็น $\chi^2_{df,\alpha}$ และ $\chi^2_{df,1-\alpha}$ เช่น $\chi^2_{6,.05}$ และ $\chi^2_{6,.95}$ จะมีค่าเป็น

$$\chi^2_{6,.05} = 12.592$$

$$\chi^2_{6,.95} = 1.635$$

ดังนั้น ช่วงความเชื่อมั่น 90 % ของ χ^2_6 เป็นดังสมการ (7-51)

$$\chi^2_{6,.95} < \chi^2_6 < \chi^2_{6,.05} \qquad \text{...(7-51)}$$

หรือ ช่วงความเชื่อมั่น (1-α) 100% หา χ^2_{df} เป็นดังสมการ (7-52)

$$\chi^2_{df,1-\alpha/2} < \chi^2_{df} < \chi^2_{df,\alpha/2} \qquad \text{...(7-52)}$$

เมื่อแทนค่า $\chi^2_{df} = \dfrac{(n-1)S^2_x}{\sigma^2_x}$ จากสมการ (7-50)

จะได้ $\chi^2_{df,1-\alpha/2} < \dfrac{(n-1)S^2_x}{\sigma^2_x} < \chi^2_{df,\alpha/2}$

ดังนั้นค่าช่วงความเชื่อมั่น $(1-\alpha)$ 100 % ของ σ^2_X คือ

$$\frac{(n-1)S^2_x}{\chi^2_{df,\alpha/2}} < \sigma^2_X < \frac{(n-1)S^2_x}{\chi^2_{df,(1-\alpha/2)}} \qquad \text{...(7-53)}$$

ตัวอย่างที่ 7.11 สุ่มตรวจสอบความยาวสลักเกลียวที่ผลิตจากเครื่องจักรเครื่องหนึ่ง 15 ตัว พบค่าเบี่ยงเบนมาตรฐานความยาวเป็น 0.8 นิ้ว จงประมาณค่าการเบี่ยงเบนมาตรฐานความยาวสลักเกลียว สลักเกลียวทั้งหมดผลิตจากเครื่องจักรเครื่องนี้เชื่อได้ 90 %

วิธีทำ

ให้ $n = 15$

$df = 15 - 1$

$= 14$

$S_x^2 = (0.8)^2$

$= 0.64$

กำหนดให้ $\alpha = .10$

$\frac{\alpha}{2} = .05$

$1-\frac{\alpha}{2} = .95$

จาก

$\chi^2_{14,.05} = 23.68$

$\chi^2_{14,.95} = 6.57$

ช่วงความเชื่อมั่น 90 % ของความแปรปรวน คือ

$$\frac{(n-1)S_x^2}{\chi^2_{df,\alpha/2}} < \sigma_x^2 < \frac{(n-1)S_x^2}{\chi^2_{df,(1-\alpha/2)}}$$

$$\frac{(14)(.64)}{23.68} < \sigma_x^2 < \frac{(14)(.64)}{6.57}$$

$$0.378 < \sigma_x^2 < 1.364$$

$$0.61 < \sigma_x < 1.17$$

ช่วงความเชื่อมั่น 90 % ส่วนเบี่ยงเบนมาตรฐานความยาวสลักเกลียวมีค่าระหว่าง 0.61 นิ้ว ถึง 1.17 นิ้ว

7.7 บทสรุป

บทนี้ได้กล่าวถึงการประมาณค่าเกี่ยวประชากรด้วยตัวอย่างที่มีอยู่ในมือ กล่าวไว้ 2 ประการคือการประมาณค่าแบบจุดและการประมาณค่าด้วยช่วงทั้งกรณีประชากรเดียวและสองประชากร สำหรับเป็นพื้นฐานกรณีประมาณค่าเกี่ยวกับหลาย ๆ ประชากรในโอกาสต่อไป สรุปได้ดังนี้

7.7.1 การประมาณค่าแบบจุด มีดังนี้

1) การประมาณค่าเฉลี่ยประชากร

2) การประมาณค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐานประชากร

3) การใช้ค่าเบี่ยงเบนมาตรฐานตัวอย่าง

7.7.2 การประมาณค่าด้วยช่วง มีการกำหนดระดับนัยสำคัญ (α) 3 ระดับ ซึ่งใช้ในการทดสอบสมมุติฐานด้วย ดังนี้

1) กำหนด α = .01 เรียกว่า มีนัยสำคัญยิ่ง

2) กำหนด α = .05 เรียกว่า มีนัยสำคัญ

3) กำหนด α > .05 เรียกว่า ไม่มีนัยสำคัญ

7.7.3 ในการประมาณค่าเฉลี่ยประชากรด้วยช่วงความเชื่อมั่น มีการพิจารณา 2 กรณีคือ กรณีที่กลุ่มตัวอย่างมีขนาดใหญ่และกลุ่มตัวอย่างที่มีขนาดเล็ก

7.7.4 เมื่อเราต้องการหาค่าเฉลี่ยแตกต่างกันของสองประชากรมีวิธีการคิดเป็นโดยใช้ค่าดังต่อไปนี้

1) ความแปรปรวนร่วม

2) ตัวอย่างชนิดคู่

7.8 คำถามทบทวน

1. จงอธิบายความหมายของสิ่งต่อไปนี้
   ก. ประชากร
   ข. ตัวอย่าง
   ค. ตัวประมาณค่าที่ดี
   ง. ช่วงความเชื่อมั่น
   จ. การประมาณค่าอย่างไม่อคติ

2. จงอธิบายความแตกต่างระหว่างการประมาณค่าแบบจุดและการประมาณค่าด้วยช่วงหรือช่วงความเชื่อมั่น

3. สำนักงานมาตรฐานอุตสาหกรรม สุ่มชั่งน้ำหนักผงซักฟอกชนิดบรรจุ 1 กิโลกรัม (เขียนข้างภาชนะบรรจุ) จำนวน 36 กล่อง พบว่ามีน้ำหนักเฉลี่ยเป็น 0.95 กก. ค่าเบี่ยงเบนมาตรฐาน 0.335 กิโลกรัม จงประมาณค่าเฉลี่ยน้ำหนักผงซักฟอกชนิดดังกล่าวที่มีวางขายในท้องตลาด ด้วยช่วงความเชื่อมั่น 95%

4. จงประมาณค่าเฉลี่ยรายได้เกษตรกรไทยต่อเดือนแต่ละครอบครัวที่เชื่อได้ 95% เมื่อตัวอย่างสุ่มเป็นรายได้ของเกษตรกร 35 ครอบครัว พบรายได้เฉลี่ยต่อเดือนแต่ละครอบครัวเป็น 3,500 บาท ค่าเบี่ยงเบนมาตรฐาน 325 บาท

5. จากข้อมูลในตารางที่ 7.6 จงประมาณรายได้แตกต่างกันของพนักงานจากโรงงานทั้งสองที่เชื่อได้ 95%

ตารางที่ 7.6 ขนาดตัวอย่าง รายได้ต่อวัน และความแปรปรวนรายได้ของพนักงานโรงงานทอผ้าและโรงงานยาสูบ

พนักงาน	ขนาดตัวอย่าง	รายได้ต่อวัน	ความแปรปรวน
โรงงานทอผ้า	25	45	100
โรงงานยาสูบ	35	70	105

6. จงหาช่วงความเชื่อมั่น 95 % ของสัดส่วนผู้ใช้ผงซักฟอกยี่ห้อ A สุ่มตัวอย่างโดยสอบถามผู้ใช้ผงซักฟอก 142 ราย พบว่า ใช้ผงซักฟอกยี่ห้อ A 83 ราย

7. สุ่ม วัดความหนาของแผ่นพลาสติกที่ผลิตเพื่อทำแผ่นรองพื้น มีความหนาเป็นมิลลิเมตร ได้ดังนี้

19.8 , 21.2 , 18.6 , 20.4 , 21.6 , 19.8 , 19.9 , 20.3 และ 20.8

ถ้าการแจกแจงความหนาแผ่นพลาสติกที่ผสมได้มีการแจกแจงปรกติ จงประมาณค่าเบี่ยงเบนมาตรฐานความหนาของที่แผ่นพลาสติกด้วยช่วงความเชื่อมั่น 95 %

บทที่ 8
การทดสอบสมมุติฐานเกี่ยวกับประชากร

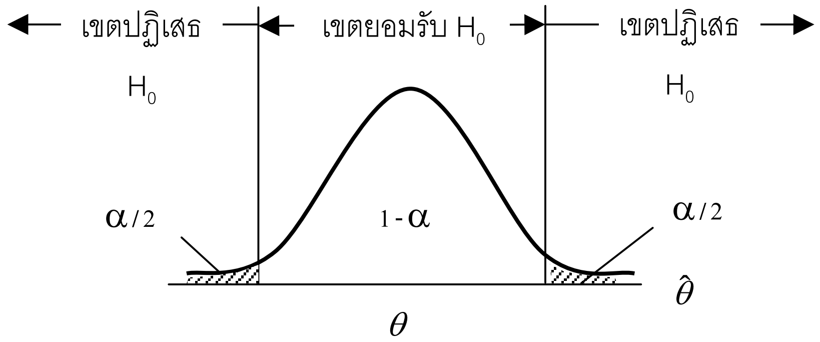
เมื่อมีปัญหาหรือคำถามเกิดขึ้น ข้อความหรือคำกล่าวที่ตั้งขึ้นเพื่อตอบปัญหาหรือ คำถามโดยอาศัยประสบการณ์ ปรากฏการณ์ต่าง ๆ ที่เคยเกิดขึ้นมาแล้วหรือเคยถูกค้นพบมาแล้วพิจารณาประกอบ เราจะเรียกข้อความหรือคำกล่าวนั้นว่า สมมุติฐาน (hypothesis) อาจกล่าวได้ว่า สมมุติฐาน คือการเดาคำตอบของปัญหาหรือคำถามเหตุผลและกลยุทธอันแยบคาย การตอบปัญหาหรือคำถามด้วยวิธีตั้งสมมุติฐานนั้นสมมุติฐานจะจริงหรือไม่จริงจะต้องตรวจหรือมีการทดสอบตามกระบวนการที่เหมาะสม และเรียกกระบวนการนี้ว่า การทดสอบสมมุติฐาน (hypothesis testing or test of hypothesis) อย่างไรก็ตามผู้ตั้งสมมุติฐานจะตั้งใจไว้ว่า สมมุติฐานตั้งขึ้นเป็น สมมุติฐานว่างหรือสมมุติฐานที่น่าจะจริง (null hypothesis) นิยมใช้สัญลักษณ์ H_o จนกว่าจะมีสมมุติฐานอื่นที่มีเหตุผลเพียงพอสำหรับการตัดสินใจ (decision making) ยอมรับหรือปฏิเสธ H_o (accept or reject H_o) นิยมเรียกสมมุติฐานหลังนี้ว่า สมมุติฐานเลือกหรือสมมุติฐานแย้ง (alternative hypothesis) ใช้สัญลักษณ์ H_1 หรือ H_a

8.1 สมมุติฐานเชิงสถิติ

สมมุติฐานเชิงสถิติ (statistical hypothesis) เป็นเทคนิคการทดสอบสมมุติฐานที่นำตัวสถิติมาใช้ในการประมาณค่าพารามิเตอร์แบบจุดด้วยตัวอย่างสุ่ม และใช้ในการตัดสินใจเกี่ยวกับสมมุติฐานที่ตั้งไว้ว่าควรปฏิเสธหรือยอมรับ และจะนำช่วงความเชื่อมั่น ($1-\alpha$) มากำหนดเป็นข้อยุติ (conclusion) ของการตัดสินใจ กล่าวคือ เราจะยอมรับ H_o ถ้าตัวสถิติเอื้ออำนวยต่อการยอมรับ H_o หรือปฏิเสธ H_o ถ้าตัวสถิติไม่เอื้ออำนวยต่อการยอมรับ H_o ในระดับความเชื่อมั่น ($1-\alpha$) ที่กำหนด

สำหรับตัวสถิติที่ใช้ในกระบวนการทดสอบสมมุติฐานจะเรียกว่า ตัวสถิติทดสอบ (test statistic) ส่วนค่าตัวสถิติทดสอบที่พึงเป็นไปได้ทั้งหมด บางค่าที่ถูกกำหนดด้วยระดับความเชื่อมั่น ($1-\alpha$) และแบ่งขอบเขตในการตัดสินใจยอมรับ H_o หรือปฏิเสธ H_o เรียกว่า

ค่าวิกฤต (critical value) หรือ จุดวิกฤต (critical point) โดยการแบ่งความน่าจะเป็นของการแจกแจงตัวสถิติที่ปฏิเสธ H_o ทางส่วนปลายหางของการแจกแจง (ส่วนปลายหางทางซ้าย ส่วนปลายหางทางขวา หรือส่วนปลายหางทางซ้ายและขวา) เพียงเล็กน้อยด้วยระดับ α จึงเรียกระดับ α ในการทดสอบสมมุติฐานว่า ขนาดของการทดสอบ (size of test) และเรียกความน่าจะเป็นส่วนน้อย ๆ ระดับ α นี้ว่า เขตปฏิเสธ หรือ เขตวิกฤต (rejection region or critical region) และเรียกขอบเขตการยอมรับ H_o ว่า เขตยอมรับ H_o (acceptance region) ดังแสดงในภาพที่ 8.1 ซึ่งแบ่งระดับ α ออกเป็นส่วนปลายหางทางซ้ายและขวาเท่า ๆ กัน



ภาพที่ 8.1 เขตยอมรับ H_o และค่าวิกฤต

ที่มา : (ดัดแปลงมาจาก Watson et al., 1990, p. 391)

เมื่อยุติการทดสอบ ผู้ทดสอบมักจะสรุปด้วยระดับ α เช่น ผลการทดสอบมีระดับนัยสำคัญระดับ α = .05 จะหมายความว่า หากใช้ระดับ α =.05 (ระดับความเชื่อมั่น .95 หรือช่วงความเชื่อมั่น 95 %) เป็นเกณฑ์ตัดสิน ตัวสถิติจะต้องตกในเขตปฏิเสธ H_o ซึ่งมีโอกาสหรือความน่าจะเป็นเพียง .05 จึงจะเชื่อว่า H_o ไม่เป็นจริง ผิด หรือถูกปฏิเสธ เป็นต้น

8.2 ลำดับขั้นของการทดสอบสมมุติฐาน

เมื่อจะดำเนินการทดสอบสมมุติฐานเกี่ยวกับพารามิเตอร์ จะมีลำดับขั้นตอนในการดำเนินการทดสอบสมมุติฐานดังต่อไปนี้

8.2.1 ขั้นที่ 1 ตั้งสมมุติฐานว่างและตั้งสมมุติฐานแย้งดังนี้

1) ตั้ง H_o โดยตั้งสมมุติฐานว่าพารามิเตอร์ของประชากร (θ) เท่ากับตัวสถิติที่ประมาณไว้ (θ_0) จะได้

$$H_o : \theta = \theta_0$$

2) ตั้ง H_1 โดยจะเลือกได้จาก 3 ลักษณะ ดังนี้

$$H_1 : \theta < \theta_0$$

หรือ $$H_1 : \theta > \theta_0$$

หรือ $$H_1 : \theta \neq \theta_0$$

8.2.2 ขั้นที่ 2 กำหนดระดับ α ขนาดการทดสอบ หรือ ระดับนัยสำคัญ โดยทั่วไปจะกำหนดเพียง 3 ระดับดังที่กล่าวในหัวข้อ 7.2 ในบทที่ 7

8.2.3 ขั้นที่ 3 กำหนดตัวสถิติทดสอบและตั้งขอบเขตหรือตั้งเกณฑ์การตัดสินใจ ตัวสถิติที่จะเลือกนั้นจะต้องเป็นตัวสถิติที่มีการแจกแจงเหมาะสมและสอดคล้องกับพารามิเตอร์นั้น ๆ เช่น การแจกแจงปรกติหรือการแจกแจงที เหมาะสมกับการทดสอบเกี่ยวกับ μ เป็นต้น แล้วแบ่งเขตการยอมรับ H_0 และ เขตปฏิเสธ H_0 โดยบอกขนาดของเขตปฏิเสธ H_0 ด้วยระดับนัยสำคัญ α สำหรับความผิดพลาดชนิดที่ 1 (type I error) หรือบอกขนาดของเขตยอมรับ H_0 ด้วยระดับนัยสำคัญ β สำหรับความผิดพลาดชนิดที่ 2 (type II error) การวางเขตปฏิเสธ H_0 จะวางไว้ทางส่วนหางของการแจกแจง ซึ่งมีการวาง 3 ลักษณะขึ้นอยู่กับ สมมุติฐานทางเลือก H_1 ว่าจะกล่าวแย้งในทิศทางใดอันถือเป็นการกำหนดทิศทางของการทดสอบด้วย กล่าวคือ

กรณี $$H_1 : \theta < \theta_0$$

ดังแสดงในภาพที่ 8.2 (ก) เป็นการทดสอบหางเดียวทางซ้าย (one-tailed test : left)

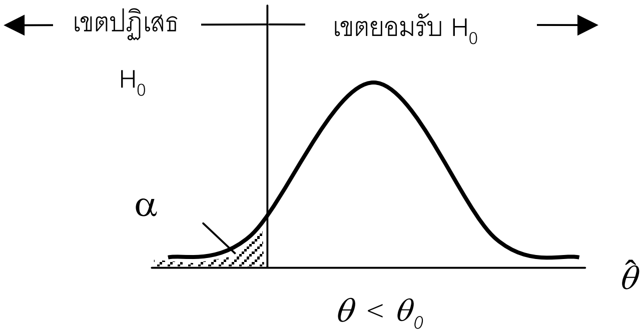
สำหรับกรณีที่

$$H_1 : \theta > \theta_0$$

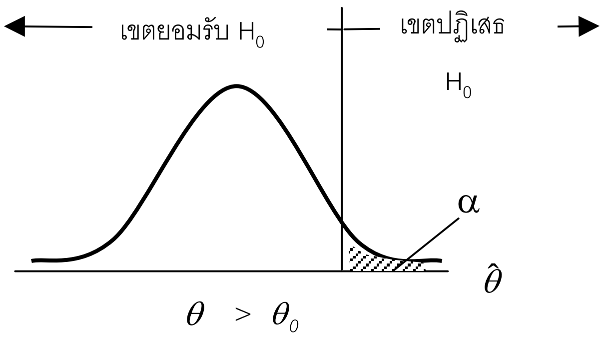
ดังแสดงในภาพที่ 8.2 (ข) เป็นการทดสอบหางเดียวทางขวา (one-tailed test : right)

และในกรณี $$H_1 : \theta \neq \theta_0$$

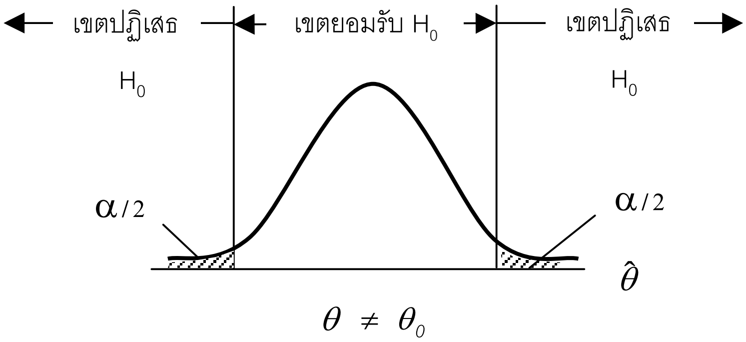
แสดงในภาพที่ 8.2 (ค) เป็นการทดสอบสองหาง (two-tailed test)



(ก)



(ข)



(ค)

ภาพที่ 8.2 ทิศทางการทดสอบสมมุติฐาน ที่ระดับ α

8.2.4 ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ เช่น คำนวณค่า Z หรือ Z_c คำนวณค่า t หรือ t_c เป็นต้น

8.2.5 ขั้นที่ 5 ถ้าตัวสถิติที่คำนวณได้จากขั้นที่ 5 ตกอยู่ในเขตปฏิเสธ H_0 หรือตกอยู่ในเขตวิกฤตเราจะตัดสินใจปฏิเสธ H_0 ที่ระดับ α แต่ถ้าตัวสถิติที่คำนวณได้จากขั้นที่ 5 ไม่ตกอยู่ในเขตวิกฤต แสดงว่าตัวสถิติไม่เอื้ออำนวยหรือไม่มีหลักฐานเพียงพอในการปฏิเสธ H_0 ระดับ α หรือต้องยอมรับ H_0 ที่ระดับ α เช่น ต้องการทดสอบสมมุติฐาน ที่ระดับ α = .05

$$H_0: \quad \theta = \theta_0$$

$$H_1: \quad \theta \neq \theta_0$$

ถ้าตัวสถิติที่คำนวณได้ตกอยู่ในเขตวิกฤต เราจะต้องปฏิเสธ H_0 ที่ระดับ α = .05 หมายความว่า θ สมมุติฐานของประชากรไม่เท่ากับ θ_0 สมมุติฐานที่ตั้งไว้ที่ช่วงความเชื่อมั่น 95% หรือเชื่อได้ 95%

8.3 การทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ยประชากร

ค่าเฉลี่ยประชากรในที่นี้จะหมายถึงประชากรเดียวและมีค่าเฉลี่ยเพียงตัวเดียว การแจกแจงค่าเฉลี่ยตัวอย่าง $(\bar{X})$ จะเป็นตัวสถิติสำคัญในการทดสอบ โดยจะกำหนดให้ค่าเฉลี่ยประชากรมีค่าเท่ากับค่าคงตัวค่าหนึ่ง เมื่อ

$$H_0: \mu = \mu_0$$

ที่ μ_0 เป็นค่าเฉลี่ยประชากรที่ตั้งไว้ว่าน่าจะเป็นจริง ซึ่งนิยมแสดงเป็นตัวเลข ค่าเฉลี่ยตัวอย่างจะกระจายอยู่รอบ ๆ μ_0

ในกรณีที่ประชากรแจกแจงปรกติหรือตัวอย่างขนาดใหญ่ และทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ) ตัวสถิติทดสอบ จะได้สมการ (8-1)

$$Z_c = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad \ldots(8\text{-}1)$$

และถ้าไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ) เราจะประมาณค่าเบี่ยงเบนมาตรฐานประชากรนี้ด้วยค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) ตัวสถิติทดสอบ เป็นดังสมการ (8-2)

$$Z_c = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \qquad \text{…(8-2)}$$

กรณีประชากรแจกแจงปรกติและไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ) จะประมาณค่านี้ด้วยค่าเบี่ยงเบนมาตรฐานตัวอย่าง (s) ตัวสถิติทดสอบ คือ

$$t_c = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \qquad \text{…(8-3)}$$

โดยที่ Z หรือ t จะกระจายอยู่รอบ ๆ ศูนย์ ซึ่งตรงกับจุดที่ $\bar{X} = \mu = \mu_0$

ตัวอย่างที่ 8.1 จงทดสอบที่ระดับ α = .05 ว่า ความยาวของสลักเกลียวที่ผลิตจากเครื่องจักรเครื่องหนึ่งยังคงเป็น 2.5 นิ้ว และค่าเบี่ยงเบนมาตรฐานไม่เกิน 0.08 นิ้ว ตัวอย่างสุ่มประกอบด้วยสลักเกลียว 800 ตัวจาก การผลิตของเครื่องจักรเครื่องนี้ คำนวณค่าเฉลี่ยความยาวสลักเกลียวได้ 2.604 นิ้ว

วิธีทำ ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 H_0 : $\mu_0 = 2.5$

H_1 : $\mu_0 \neq 2.5$

ขั้นที่ 2 ระดับนัยสำคัญ α = .05 , $\alpha/2$ = .025

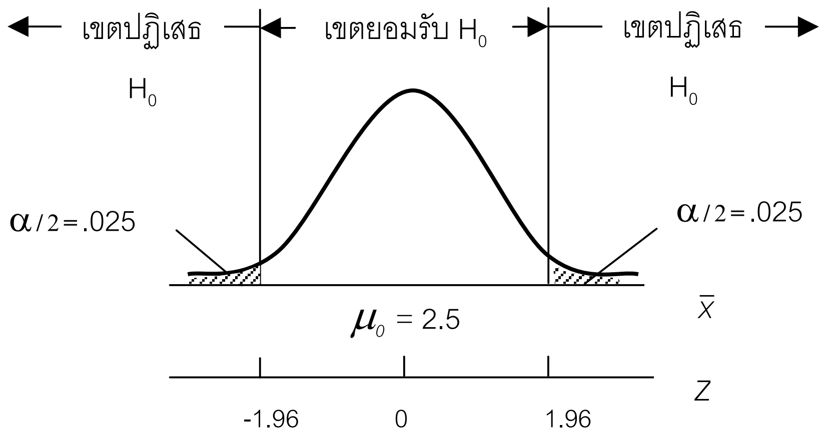
ขั้นที่ 3 สมมุติฐานเกี่ยวกับ μ ของประชากร กรณีนี้ ตัวอย่างขนาดใหญ่และทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ) ตัวสถิติทดสอบคือ Z และกรณีนี้เป็นการทดสอบสองหาง ที่ระดับ α =.05 จากตารางภาคผนวกที่ 5 Z = 1.96 ดังนั้น เขตปฏิเสธ H_0 คือ $Z > 1.96$ หรือ $Z < -1.96$

ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ : Z_c

$$Z_c = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$\bar{X} = 2.604 , \quad \sigma = 0.08 , \quad n = 800$$

$$Z_c = \frac{2.604 - 2.5}{0.08/\sqrt{800}} = \frac{0.104}{0.08/28.28} = 37$$



ภาพที่ 8.3 ทิศทางการทดสอบสมมุติฐานที่ระดับ $\alpha = .05$
ของความยาวของสลักเกลียวที่ผลิตจากเครื่องจักร

ขั้นที่ 5 ตัดสินใจปฏิเสธ H_0 เพราะ $Z_c > 1.96$ ตกอยู่ในเขตปฏิเสธ H_0 ที่ระดับ $\alpha = .05$
นั่นคือเครื่องจักรเครื่องนี้ผลิตสลักเกลียวยาวกว่า 2.5 นิ้ว ที่ระดับ $\alpha = .05$

ตัวอย่างที่ 8.2 จงทดสอบว่าน้ำหนักลูกไก่อายุหนึ่งวันเป็น 12 กรัม หรือไม่ที่ระดับ $\alpha = .05$ ตัวอย่างสุ่มเป็นน้ำหนักลูกไก่อายุหนึ่งวันจำนวน 15 ตัว ดังนี้ 11.8, 11.8, 11.9, 12.1, 11.9, 11.9, 12.1, 12.0, 12.1, 12.1, 12.2, 12.3, 12.3, 12.4 และ 12.4 กรัม

วิธีทำ หาค่าเฉลี่ยตัวอย่าง : $\bar{X}$

$$\bar{X} = (11.8 + 11.8 + 11.9 + \ldots\ldots\ldots\ldots + 12.4 + 12.4)/15$$

$$\bar{X} = 12.1 \text{ กรัม}$$

หาค่าเบี่ยงเบนมาตรฐานตัวอย่าง : S

$$S^2 = (11.8-12.1)^2 + (11.8-12.1)^2 + \ldots\ldots\ldots\ldots$$

$$\ldots\ldots + (12.4-12.1)^2 + (12.4-12.1)^2/14$$

$$S^2 = 0.04$$

$$S = \sqrt{0.04} = 0.2$$

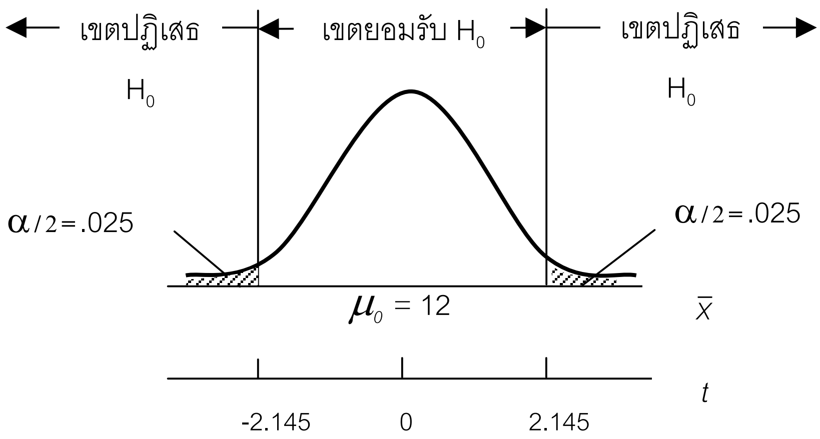
ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 $H_0 : \mu = 12.0$

$H_1 : \mu \neq 12.0$

ขั้นที่ 2 ระดับนัยสำคัญ, α = .05

ขั้นที่ 3 สมมุติฐานเกี่ยวกับ μ ของประชากร กรณีนี้ ตัวอย่างขนาดเล็ก ประมาณค่าเบี่ยงเบนมาตรฐานประชากรด้วยค่าด้วยค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S) และประชากรแจกแจงปรกติโดยประมาณ ตัวสถิติทดสอบคือ t_c และเป็นการทดสอบสองหาง ที่ระดับ α = .05, $\alpha/2$ = .025, จากตารางภาคผนวกที่ 6 df = 14 , $t_{.025,14}$ = 2.145 ดังนั้น เขตปฏิเสธ H_0 คือ $t_c < -2.145$ หรือ $t_c > 2.145$



ภาพที่ 8.4 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05 ของน้ำหนักลูกไก่

ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ : t_c

$$t_c = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$t_c = \frac{12.1 - 12.0}{\frac{0.2}{\sqrt{15}}}$$

$$t_c = 1.94$$

ขั้นที่ 5 ตัดสินใจยอมรับ H_0 เพราะ $t_c < 2.14$ ตกอยู่ในเขตยอมรับ H_0 ที่ระดับ $\alpha = .05$

นั่นคือ ยอมรับน้ำหนักลูกไก่อายุหนึ่งวันทั้งหมดเป็น 12 กรัม ที่ระดับ $\alpha = .05$ เชื่อได้ 95%

ตัวอย่างที่ 8.3 ตัวอย่างสุ่มประกอบด้วยแยมสับปะรด 30 ขวด น้ำหนักเฉลี่ย 119 กรัม ป้ายข้างขวดเขียนว่าน้ำหนักสุทธิ 120 กรัมและถ้าฝ่ายมาตรฐานสินค้ากำหนดค่าเบี่ยงเบนมาตรฐานไม่เกิน 3 กรัม จงทดสอบน้ำหนักเฉลี่ยของแยมนี้ว่าน้อยกว่า 120 กรัม หรือไม่

วิธีทำ ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 $H_0 : \mu = 120$

$H_1 : \mu < 120$

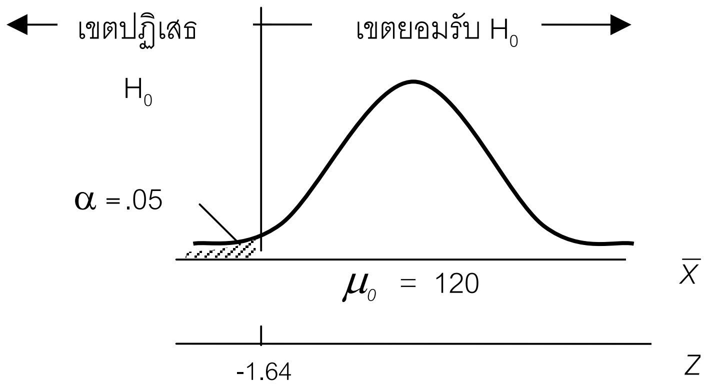
ขั้นที่ 2 ระดับนัยสำคัญ $\alpha = .05$

ขั้นที่ 3 สมมุติฐานเกี่ยวกับ μ ของประชากร กรณีนี้ทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ) ตัวสถิติทดสอบคือ Z และกรณีนี้เป็นการทดสอบสองหางเดียวทางซ้าย ที่ระดับ $\alpha = .05$ จากตาราง ภาคผนวกที่ 5 $Z_{.05} = -1.64$ ($Z_{.01} = -2.31$) ดังนั้น เขตปฏิเสธ H_0 คือ $Z_C < -1.64$

ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ : Z_C

$$Z_C = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$\overline{X} = 119 , \quad \mu_0 = 120 , \sigma = 3 , \; n = 30$$



ภาพที่ 8.5 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05
ของน้ำหนักแยมสับปะรด

$$Z_c = \frac{119 - 120}{3/\sqrt{30}}$$

$$= -1.83$$

ขั้นที่ 5 ตัดสินใจปฏิเสธ H_0 เพราะ $Z_c < -1.64$ ตกอยู่ในเขตปฏิเสธ H_0 ที่ระดับ α = .05 (แต่อาจยอมรับ H_0 ที่ระดับ α = .01)
นั่นคือ เชื่อได้ 95% ว่า น้ำหนักสุทธิของแผนสับปะรดน้อยกว่า 120 กรัม (แต่อาจเชื่อได้ 99% ว่า น้ำหนักสุทธิของแยมสับปะรดยังคงเป็น 120 กรัม)

8.4 การทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ยประชากรสองประชากร

หากประชากรแรกมีค่าเฉลี่ย μ_1 ค่าเบี่ยงเบนมาตรฐาน σ_1 ส่วนประชากรหลังมีค่าเฉลี่ย μ_2 ค่าเบี่ยงเบนมาตรฐาน σ_2 สมมุติฐานเกี่ยวกับค่าเฉลี่ยประชากรทั้งสองนี้ คือ

$H_0 : \mu_1 = \mu_2$ (หรือ $\mu_1 - \mu_2 = 0$)

$H_1 : \mu_1 > \mu_2$ (หรือ $\mu_1 - \mu_2 > 0$)

หรือ $H_1 : \mu_1 < \mu_2$ (หรือ $\mu_1 - \mu_2 < 0$)

หรือ $H_1 : \mu_1 \neq \mu_2$ (หรือ $\mu_1 - \mu_2 \neq 0$)

8.4.1 ในกรณีที่มี 2 ตัวอย่างอิสระกันหากตัวอย่างจากประชากรทั้งสองมีขนาดใหญ่ ($n_1, n_2 \geq 20$) และทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ_1, σ_2). ตัวสถิติทดสอบจะเป็นดังสมการ (8-4)

$$Z_c = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \ldots(8\text{-}4)$$

ที่ $\mu_1 = \mu_2$; $\mu_1 - \mu_2 = 0$

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

และถ้าไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ_1, σ_2) เราจะประมาณด้วยค่าเบี่ยงเบนมาตรฐานตัวอย่าง (S_1 , S_2) ตัวสถิติทดสอบ เป็นดังสมการ (8-5)

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \ldots(8\text{-}5)$$

ตัวอย่างที่ 8.4 สุ่มชั่งน้ำหนักไก่กระทง 2 พันธุ์ที่เลี้ยงด้วยอาหารชนิดเดียวกันเป็นเวลาสามสิบวัน โดยสุ่มชั่งน้ำหนักไก่พันธุ์ ก. จำนวน 100 ตัว หาค่าน้ำหนักเฉลี่ยได้ 952 กรัม ค่าเบี่ยงเบนมาตรฐาน 85 กรัม สุ่มชั่งไก่พันธุ์ ข. จำนวน 50 ตัว หาค่าน้ำหนักเฉลี่ยได้ 987 กรัม, ค่าเบี่ยงเบนมาตรฐาน 92 กรัม จงทดสอบความแตกต่างน้ำหนักของไก่ 2 พันธุ์นี้ที่ระดับ $\alpha = .05$

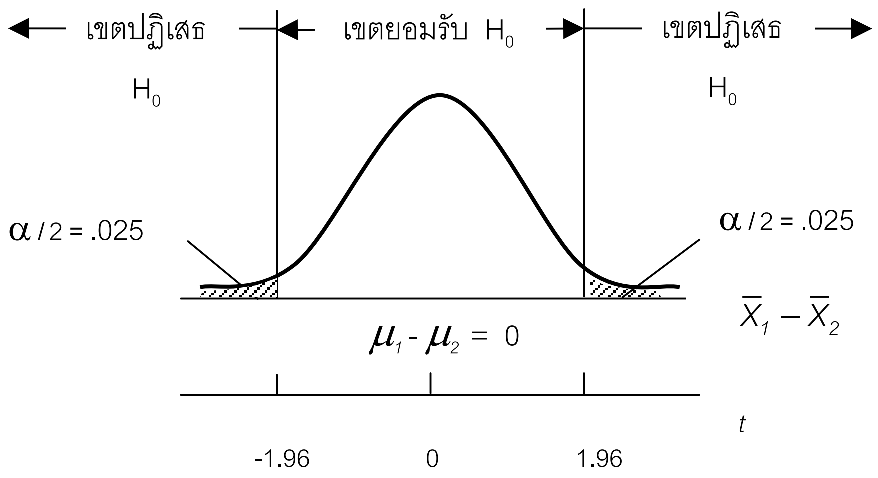
วิธีทำ ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 \neq \mu_2$

ขั้นที่ 2 ระดับนัยสำคัญ $\alpha = .05$, $\alpha/2 = .025$

ขั้นที่ 3 สมมุติฐานเกี่ยวกับความแตกต่างของค่าเฉลี่ย 2 ประชากร (μ_1-μ_2) เมื่อตัวอย่างจากประชากรทั้ง 2 มีขนาดใหญ่ (n_1, $n_2 \geq 20$) แต่การประมาณค่าความแปรปรวนกับประชากรทั้งสอง ตัวสถิติทดสอบคือ t_c และเป็นการทดสอบสองหาง ที่ระดับ α = .05 จากตารางภาคผนวกที่ 6 และในตารางภาคผนวกที่ 5 เมื่อ *df* มีค่าอนันต์ ค่า $Z_{.025}$ = 1.96 ดังนั้น เขตปฏิเสธ H_0 คือ $t_c < -1.96$ หรือ $t_c > 1.96$



ภาพที่ 8.6 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05 ของน้ำหนักไก่กระทง 2 พันธุ์

ขั้นที่ 4 คำนวณตัวสถิติทดสอบ : t_c

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

ที่ $\bar{X}_1 = 952,\ n_1 = 100,\ S_1^2 = 85^2$

$\bar{X}_2 = 987,\ n_2 = 50,\ S_2^2 = 92^2$

$$t_c = \frac{(952 - 987)}{\sqrt{\frac{85^2}{100} + \frac{92^2}{50}}}$$

$$t_c \quad = \quad \frac{35}{15.5} \quad = \quad -2.26$$

ขั้นที่ 5 ตัดสินใจปฏิเสธ H_0 เพราะ $t_c < -1.96$ ตกอยู่ในเขตปฏิเสธ H_0 ที่ระดับ $\alpha = .05$

นั่นคือ น้ำหนักไก่กระทงทั้ง 2 พันธุ์นี้เลี้ยงด้วยอาหารชนิดเดียวกันเป็นเวลาสามสิบวัน น้ำหนักแตกต่างกันที่ระดับ $\alpha = .05$ (เชื่อได้ 95%)

สำหรับกรณีสองตัวอย่างอิสระกันนี้ หากประชากรแจกแจงปรกติหรือประชากรแจกแจงปรกติโดยประมาณและตัวอย่างจากประชากรทั้งสองมีขนาดเล็ก (n_1, $n_2 < 20$) พร้อมกับไม่ทราบค่าเบี่ยงเบนมาตรฐานประชากร (σ_1, σ_2) จะต้องประมาณค่าผิดพลาดมาตรฐานประชากรทั้งสองด้วย S_p ตัวสถิติทดสอบ คือ

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \ldots(8\text{-}6)$$

ที่ S_p^2 คือ ความแปรปรวนร่วม (the pooled variance) หาได้สมการ (7-32) หรือ (7-35) ในบทที่ 7

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{SS_1 + SS_2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \ldots(8\text{-}7)$$

เมื่อ SS_1 คือ $\sum (X_1 - \bar{X}_1)^2$

และ SS_2 คือ $\sum (X_2 - \bar{X}_2)^2$

ตัวอย่างที่ 8.5 จงทดสอบความแตกต่างอายุการใช้งานหลอดไฟฟ้าที่ผลิตโดยบริษัท ก. กับ บริษัท ข. จากตัวอย่างที่ 7.8 ในบทที่ 7 ว่าแตกต่างกันที่ระดับ $\alpha = .05$ หรือไม่

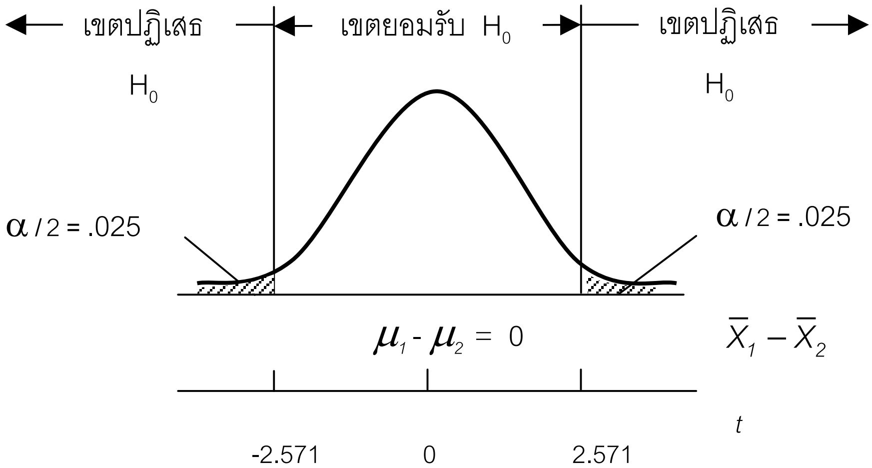
วิธีทำ ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 H_0 : $\mu_1 = \mu_2$

H$_1$: $\mu_1 \neq \mu_2$

ขั้นที่ 2 ระดับนัยสำคัญ α = .05 และ $\alpha/2$ = .025

ขั้นที่ 3 สมมุติฐานเกี่ยวกับความแตกต่างของค่าเฉลี่ย 2 ประชากร ($\mu_1 - \mu_2$) ตัวอย่างมีขนาดเล็กและประชากรแจกแจงปรกติโดยประมาณ ตัวสถิติทดสอบคือ t_C และเป็นการทดสอบสองหาง ที่ระดับ α = .05 และ $\alpha/2$ = .025 จากตารางภาคผนวกที่ 6 df = 4 + 3 - 2 = 5 , $t_{.025}$ = 2.571 หรือ $t_{1-.025} = t_{.975}$ = -2.571



ภาพที่ 8.7 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05
ของอายุการใช้งานหลอดไฟฟ้า

ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ : t_c

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$(\bar{X}_1 - \bar{X}_2)$ = 74-60 = 14, $S_p = \sqrt{116.79}$

n_1 = 4, n_2 = 3

$$t_c = \frac{14}{\sqrt{116.79}\sqrt{\frac{1}{4}+\frac{1}{3}}}$$

$$= \frac{14}{8.26} = 1.69$$

ขั้นที่ 5 ตัดสินใจยอมรับ H_0 เพราะ t_c ไม่ตกอยู่ในเขตปฏิเสธ H_0 นั่นคือ อายุการใช้งานหลอดไฟฟ้าที่ผลิตโดยบริษัท ก. กับที่ผลิตโดยบริษัท ข. ไม่แตกต่างกันที่ระดับ α = .05

8.4.2 กรณีสองตัวอย่างไม่อิสระกัน ความแตกต่างค่าเฉลี่ยประชากรกรณีนี้อาศัยความรู้เกี่ยวกับการประมาณ ค่าความแตกต่างค่าเฉลี่ยประชากรด้วยช่วงในหัวข้อ 7.4.2 ในที่นี้ขอกล่าวเฉพาะประชากรทั้งสองแจกแจงปรกติหรือแจกแจงปรกติโดยประมาณและตัวอย่างจากประชากรทั้งสองมีขนาดเล็ก (n_1, n_2 < 20) ตัวสถิติทดสอบ เป็นดังสมการ (8-8)

$$t_c = \frac{\overline{D} - \Delta_0}{\frac{S_D}{\sqrt{n}}} \qquad \text{…(8-8)}$$

ที่ Δ_0 เป็นความแตกต่างค่าเฉลี่ยประชากรทั้งคู่ และ ตั้งเป็นสมมุติฐานที่น่าจะจริง (H_0)

$$\overline{D} = (\overline{X}_1 - \overline{X}_2)$$

และ

$$S_D = \sqrt{\frac{\sum (D - \overline{D})^2}{n - 1}}$$

สำหรับสมมุติฐานเกี่ยวกับ Δ_0 ผู้ตั้งสมมุติฐานมักจะกำหนดค่าความแตกต่าง Δ_0 นี้ไว้ก่อนเป็นตัวเลข เช่น ความแตกต่างคะแนนหลังฝึกอบรมกับคะแนนก่อนฝึกอบรม เป็น 30 คะแนน จะได้สมมุติฐานที่ทำการทดสอบเป็น

H_0 : $\Delta_0 = 30$

H_1 : $\Delta_0 > 30$

หรือ H_1 : $\Delta_0 < 30$

หรือ H$_1$: $\Delta_0 \neq 30$ เป็นต้น

ตัวอย่างที่ 8.6 จงทดสอบว่าความแตกต่างคะแนนสอบหลังฝึกอบรมก่อนฝึกอบรมของพนักงานจากข้อมูลในตัวอย่างที่ 7.9 ในบทที่ 7 ว่าคะแนนแตกต่างกันอย่างน้อย 16 คะแนน หรือไม่ ที่ระดับ $\alpha = .05$

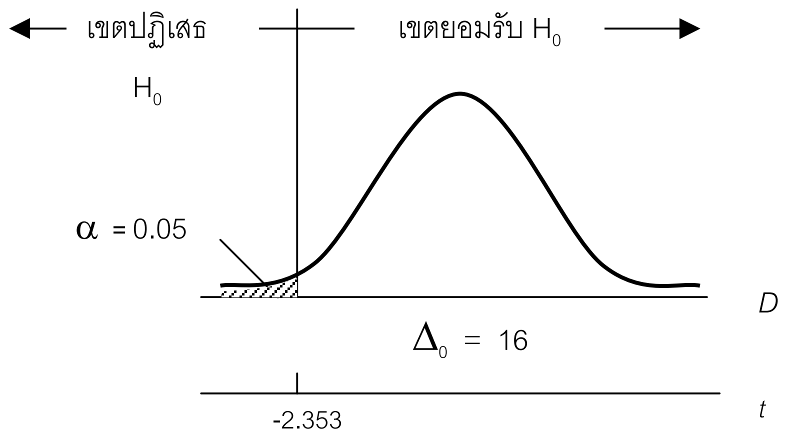
วิธีทำ ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 H$_0$: $\Delta_0 = 16$

H$_1$: $\Delta_0 < 16$

ขั้นที่ 2 ระดับนัยสำคัญ $\alpha = .05$

ขั้นที่ 3 กรณีนี้เป็นการทดสอบหางเดียว : ทางซ้าย จากตารางภาคผนวกที่ 6 $df = 4-1 = 3$, $t_{.05} = -2.353$



ภาพที่ 8.8 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05 ของความแตกต่างของคะแนนสอบหลังฝึกอบรมและก่อนฝึกอบรม

ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ : t_c

$$t_c = \frac{\overline{D} - \Delta_0}{\frac{S_D}{\sqrt{n}}}$$

$$\overline{D} = 14\ ,\ \Delta_0 = 16\ ,\ S_D = \sqrt{15.3} = 3.91\ ,\ n = 4$$

$$= \frac{14 - 16}{3.91/\sqrt{4}} = -1.02$$

ขั้นที่ 5 ตัดสินใจยอมรับ H_0 เพราะ t_c ไม่ตกอยู่ในเขตปฏิเสธ H_0

นั่นคือ ความแตกต่างคะแนนสอบหลังฝึกอบรมกันก่อนฝึกอบรมไม่น้อยกว่า 16 คะแนน ที่ระดับ α = .05

8.5 การทดสอบสมมุติฐานเกี่ยวกับสัดส่วนประชากร

จากหัวข้อ 7.5 บทที่ 7 การประมาณค่าสัดส่วนประชากร กำหนด π เป็นสัดส่วนประชากรและ p เป็นสัดส่วนตัวอย่าง สมมุติฐานที่ดำเนินทดสอบคือ

H_0 : $\pi = \pi_0$

และ H_1 : $\pi \neq \pi_0$

หรือ H_1 : $\pi > \pi_0$

หรือ H_1 : $\pi < \pi_0$

และจากการแจกแจงสัดส่วนตัวอย่างขนาด n ได้ตัวสถิติทดสอบดังสมการ (8-9)

$$Z_c = \frac{p - \pi_0}{\sqrt{\frac{p(1-p)}{n}}} \qquad \ldots(8\text{-}9)$$

ตัวอย่างที่ 8.7 ในปีแรกของการผลิตน้ำดื่มตราลูกเป็ดผู้ผลิตทราบว่ามีส่วนแบ่งการตลาดร้อยละ 12 สำหรับปีต่อมาผู้ผลิตปรับปรุงแผนการตลาดใหม่เพื่อหวังว่าจะมีส่วนแบ่งการตลาดเพิ่มขึ้น สุ่มถามผู้ซื้อ 400 รายพบว่าซื้อน้ำดื่มตราลูกเป็ด 58 ราย จงทดสอบว่าน้ำดื่มตราลูกเป็ดมีส่วนแบ่งการตลาดเพิ่มจากเดิม ที่ระดับ α = .10 หรือไม่

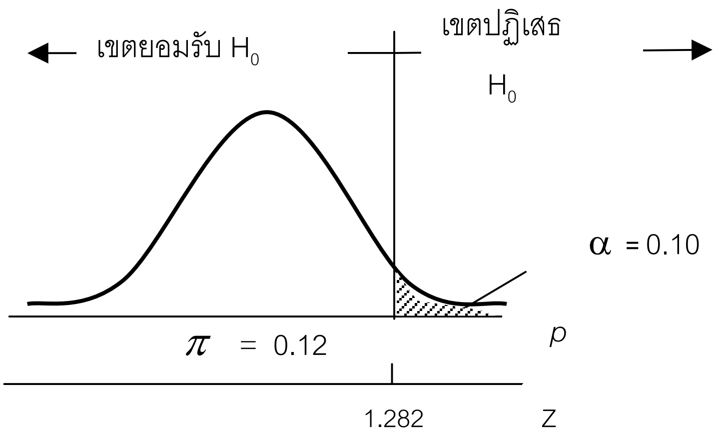
วิธีทำ ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 H_0 : $\pi = 0.12$

H_1 : $\pi > 0.12$

ขั้นที่ 2 ระดับนัยสำคัญ α = .10

ขั้นที่ 3 สมมุติฐานเกี่ยวกับสัดส่วนประชากร ตัวสถิติทดสอบคือ Z_c และเป็นการทดสอบหางเดียวทางขวา ที่ระดับ α = .10 จากตารางภาคผนวกที่ 5 หรือ จากตารางภาคผนวกที่ 6 $df = \infty$, $t_{.10} = Z_{.10} = 1.282$



ภาพที่ 8.9 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .10

ขั้นที่ 4 คำนวณค่าตัวสถิติทดสอบ : Z_c

$$Z_c = \frac{p - \pi_0}{\sqrt{\frac{p(1-p)}{n}}}$$

ให้ p = สัดส่วนผู้ที่ซื้อน้ำดื่มตราลูกเป็ด (สัดส่วนตัวอย่าง)

เรามี n = 400

$$p = \frac{58}{400} = 0.145 , \quad (1-p) = 0.855$$

$$Z_\alpha = Z_{.10} = 1.282$$

$$Z_c = \frac{0.145 - 0.12}{\sqrt{\frac{(0.145)(0.855)}{400}}}$$

$$= 1.54$$

ขั้นที่ 5 ตัดสินใจปฏิเสธ H_0 เพราะ Z_c > 1.282 ตกอยู่ในเขตปฏิเสธ H_0 นั่นคือ เชื่อว่าน้ำดื่มตราลูกเป็ดมีส่วนแบ่งการตลาดมากกว่าเดิม ที่ระดับ α = .10

8.6 ความผิดพลาดในการทดสอบสมมุติฐาน

ในหัวข้อ 8.2 ได้กล่าวพิงถึง ความผิดพลาดชนิดที่ 1 และ ความผิดพลาดชนิดที่ 2 เกี่ยวกับการบอกขนาดของเขตปฏิเสธ H_0 ด้วยระดับ α และเพื่อความเข้าใจสำหรับความผิดพลาดในการทดสอบสมมุติฐาน ขอให้กลับไปพิจารณาตัวอย่างที่ 8.3 อีกครั้ง จะพบว่าเขตปฏิเสธ H_0 ยิ่งอยู่ปลายสุดของการแจกแจงค่าเฉลี่ย ($\overline{X}$) เป็นโอกาสที่จะเกิดความผิดพลาดในการ ตัดสินใจเกี่ยวกับ μ ยิ่งน้อยเมื่อใดที่ H_0 ที่ตั้งไว้เป็นจริง (ความเป็นจริงของประชากรหรือความจริงในธรรมชาติ) และจุดที่โอกาสเกิดขึ้นน้อยเหมาะสมสำหรับการปฏิเสธ H_0 มากกว่าการยอมรับ H_0 เป็นสิ่งที่ควรคำนึงถึงในการเลือกระดับ เช่น สมมติว่าเราตั้งระดับ α = .25 และสมมุติฐาน H_0 เป็นจริง ในตัวอย่างนี้ตัวสถิติที่คำนวณได้จะตกอยู่ในเขตปฏิเสธ H_0 เป็นการบอกว่าการทดลองเพียงร้อยละ 25 หรือ $\frac{1}{4}$ ครั้ง จะนำไปสู่ข้อสรุปที่ผิดพลาดว่าสมมุติฐาน H_0 ผิด (ปฏิเสธ H_0) จึงเป็นการเสี่ยงภัย (risk) สูงเกินไปที่จะสรุปเช่นนั้น

แต่ถ้าเราเลือกระดับ α = .001 และสมมุติฐาน H_0 ไม่จริง (เป็นเท็จของประชากรหรือเป็นเท็จในธรรมชาติ) ถ้าตัวสถิติที่คำนวณได้จะตกอยู่ในเขตยอมรับ H_0 เป็นการบอกให้ทราบว่าการทดลอง $\frac{1}{1{,}000}$ ครั้ง ก็ยังไม่มีหลักฐาน เพียงพอที่จะปฏิเสธ H_0 และต้องยอมรับ H_0 ซึ่งเป็นการเสี่ยงภัยสูงเกินไปอีกเช่นกันที่จะยอมรับ H_0 ทั้ง ๆ ที่ H_0 ผิด

ถึงแม้ว่าจะเป็นการเสี่ยงภัยมากหรือน้อย ผู้ทดสอบจะต้องตัดสินใจ ซึ่งอาจ ผิดพลาด ในกระบวนการทดสอบสมมุติฐานเชิงสถิติเราเรียกการตัดสินใจที่ผิดพลาดนี้ว่า ความผิดพลาดในการทดสอบสมมุติฐาน (error in hypothesis testing) มีอยู่ 2 ชนิด คือ

8.5.1 ความผิดพลาดชนิดที่ 1 หรืออาจเรียกว่า การเสี่ยงภัยแบบที่ 1 (alpha risk) คือ การตัดสินใจปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 ที่ตั้งไว้เป็นจริง ในการทดสอบสมมุติฐานจะแทนระดับนัยสำคัญหรือขนาดของการทดสอบด้วย α โดยที่ α คือโอกาสที่จะพบตัวสถิติทดสอบตกอยู่ในเขตปฏิเสธ H_0 เมื่อ H_0 เป็นจริง

8.5.2 ความผิดพลาดชนิดที่ 2 หรืออาจเรียกว่า การเสี่ยงภัยแบบที่ 2 (beta risk) คือ การตัดสินใจยอมรับ H_0 ทั้ง ๆ ที่ H_0 ที่ตั้งไว้ไม่จริงหรือเป็นเท็จ ในการทดสอบสมมุติฐาน

จะแทนระดับนัยสำคัญหรือขนาดของการทดสอบด้วย β (beta : บีตา) โดยที่ β คือโอกาสที่จะพบตัวสถิติทดสอบตกอยู่ในเขตยอมรับ H_0 เมื่อ H_0 ไม่จริง (H_1 จริง)

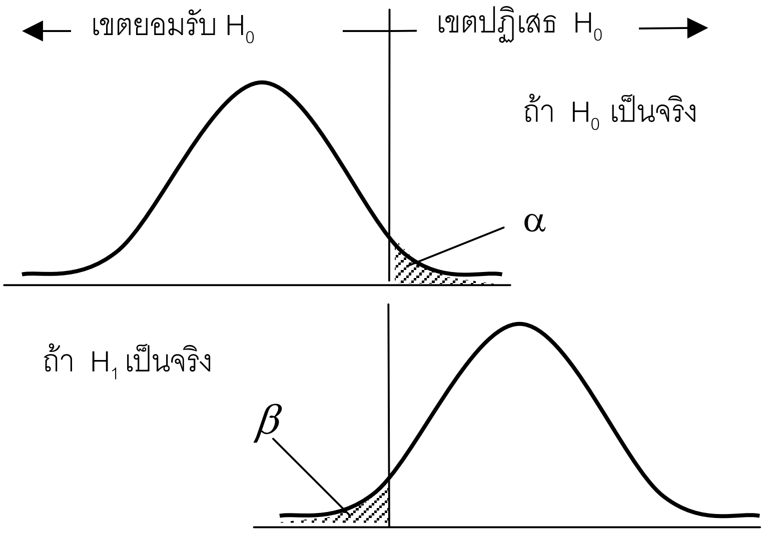
ตารางที่ 8.1 โอกาสหรือความน่าจะเป็นในการตัดสินใจเกี่ยวกับสมมุติฐาน

การตัดสินใจถูก	ตั้ง H_0 ไว้เป็นจริง	ตั้ง H_0 ไว้ไม่จริง
ปฏิเสธ H_0	α	$1 - \beta$
ยอมรับ H_0	$1 - \alpha$	β

ความผิดพลาดชนิดที่ 1 หรือการเสี่ยงภัยแบบที่ 1 เช่น ความจริงนาย ก. ฆ่าคนตาย แต่ตำรวจหรือพนักงานอัยการหาหลักฐานไม่เพียงพอที่จะแสดงให้ศาลเห็นว่านาย ก. เป็นผู้กระทำผิดจริง ศาลตัดสินยกฟ้อง (ให้โอกาสตัดสินใจยกฟ้องไว้ระดับ α) สำหรับความผิดพลาดชนิดที่ 2 หรือการเสี่ยงภัยแบบที่ 2 เช่น ความจริงนาย ก. ไม่ได้ฆ่าคนตาย แต่สภาพแวดล้อม พยานหลักฐานที่ตำรวจหรือพนักงานอัยการนำมาแสดงต่อศาล พอที่จะให้ศาลเชื่อว่านาย ก. เป็นผู้กระทำผิด ศาลตัดสินลงโทษนาย ก. (ให้โอกาสศาลตัดสินใจลงโทษไว้ระดับ β)

จะเห็นว่า ความผิดพลาดชนิดที่ 1 หรือการเสี่ยงภัยแบบที่ 1 ทำให้เกิดความเสียหายน้อยกว่าความผิดพลาดชนิดที่ 2 หรือการเสี่ยงภัยแบบที่ 2 (แม้ว่าจะเสียหายทั้ง 2 ชนิด) ในทางปฏิบัติจึงพิจารณาความผิดพลาดชนิดที่ 1 เป็นหลักในกระบวนการทดสอบสมมุติฐาน กล่าวคือ ถ้ากำหนดระดับ $\alpha = .05$ จะมีความหมายว่าเราจะยอมให้มีโอกาสตัดสินใจผิดพลาดเพียง 0.05 (5%) หรือเราจะยอมให้มีโอกาสตัดสินใจถูกต้อง 1 - 0.05 = 0.95 (95%) หรือเราเชื่อมั่นว่าตัดสินใจถูกร้อยละ 95

อย่างไรก็ดี ความน่าจะเป็นหรือโอกาสที่จะยอมรับ H_0 ที่เป็นจริง ($1 - \alpha$) โดยที่ คือโอกาสที่จะปฏิเสธ H_0 ซึ่งเป็นจริงในความผิดพลาดชนิดที่ 1 เราเรียกว่า สภาพไวของการทดสอบ (sensitivity of test) ส่วนความน่าจะเป็นหรือโอกาสที่จะปฏิเสธ H_0 ที่ไม่จริงหรือเป็นเท็จ ($1-\beta$) โดยที่ β คือโอกาสที่จะยอมรับ H_0 ที่เป็นจริง (ปฏิเสธ H_1 ที่เป็นจริง) ในความผิดพลาดชนิดที่ 2 เราเรียกว่า อำนาจของการทดสอบ (power of test)



ภาพที่ 8.10 ความผิดพลาดชนิดที่ 1 และความผิดพลาดชนิดที่ 2
ที่มา : (Wonnacott & Wonnacott, 1985, p. 727)

8.7 บทสรุป

เนื้อหาบทนี้ต่อเนื่องจากการประมาณค่าเกี่ยวกับประชากรด้วยช่วงและการประมาณค่าแบบจุด โดยนำวิธีการประมาณค่าทั้งสองมาประกอบกันด้วยการตั้งสมมุติฐานที่น่าจะจริง (H_0) เอาไว้และดำเนินการทดสอบสมมุติฐานดังกล่าว การสรุปเกี่ยวกับประชากรด้วยการยอมรับหรือปฏิเสธสมมุติฐานที่น่าจะจริงเช่นว่านั้นด้วยเหตุผลหรือหลักฐานคือตัวอย่างในมือ พร้อมทั้งบอกโอกาสของการสรุปผิดด้วยระดับนัยสำคัญ α (เน้นความผิดพลาดชนิดที่ 1 เป็นหลัก) สิ่งที่ควรรู้และทำความเข้าใจเป็นพิเศษคือ

8.7.1 วิธีการตั้งสมมุติฐานที่น่าจะจริง

1) ตัวสถิติทดสอบ
2) ค่าวิกฤติหรือจุดวิกฤต

8.7.2 สมมุติฐานเกี่ยวกับค่าเฉลี่ยประชากร

1) ประชากรเดียว
2) สองประชากร
3) ตัวอย่างขนาดเล็ก
4) ตัวอย่างอิสระกัน
5) ตัวอย่างไม่อิสระกันหรือตัวอย่างชนิดคู่

8.8 คำถามทบทวน

1. การทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ยกับช่วงความเชื่อมั่นเกี่ยวกับ ค่าเฉลี่ยประชากรแตกต่างกันหรือไม่ อธิบาย

2. จงตั้งสมมุติฐาน H_0 และ H_1 ในสถานการณ์ดังต่อไปนี้

ก. รายได้ของเกษตรกรไทยมากกว่า 5,000 บาทต่อเดือน

ข. รายได้ต่อวันของครอบครัวชาวนาไทยภาคอีสาน น้อยกว่า 200 บาท

ค. ปุ๋ยยี่ห้อ ก. จำหน่ายได้มากกว่าปุ๋ยยี่ห้อ ข.

ง. นักศึกษาระดับมหาวิทยาลัย มีความเห็นพ้องกับนโยบายการบริหารประเทศมากกว่าร้อยละ 50

3. การศึกษารายได้ต่อวันของพนักงานขายของห้างสรรพสินค้าแห่งหนึ่ง ตัวอย่างสุ่มเป็นรายได้ต่อวันของพนักงานจำนวน 1,600 ราย พบรายได้เฉลี่ยต่อวันเท่ากับ 47 บาท ค่าเบี่ยงเบนมาตรฐาน 20 บาท จะสรุปว่าพนักงานขายนี้มีรายได้มากกว่า 50 บาทต่อวัน ที่ระดับ α =.05 ได้ หรือไม่

4. ตั้งสมมุติฐานที่น่าจะจริงเอาไว้ว่า รายได้เฉลี่ยของพนักงานของบริษัทแห่งหนึ่งต่อปีแต่ละคนเป็น 30,000 บาท ตัวอย่างสุ่มเป็นรายได้พนักงานจำนวน 100 ราย ค่าเบี่ยงเบนมาตรฐาน 1,000 บาท จงดำเนินการทดสอบสมมุติฐานเช่นว่านี้อย่างไร จงอธิบายและแสดงวิธีทำมาอย่างละเอียด

5. จากคำถามข้อ 3 ในบทที่ 7 จะสรุปผลการทดสอบสมมุติฐานว่า ผงซักฟอกชนิดบรรจุ 1 กิโลกรัมที่ขายในท้องตลาดมีน้ำหนักตามที่ระบุที่ระดับ α =.01 ได้หรือไม่

6. จากคำถามข้อ 5 ในบทที่ 7 จะสรุปผลการทดสอบสมมุติฐานว่าพนักงานโรงงานทอผ้ามีรายได้สูงกว่าพนักงานโรงงานยาสูบที่ระดับ α = .05 ได้หรือไม่

7. บริษัทผู้จำหน่ายเสื้อยืดเจบีทราบว่ามีผู้นิยมซื้อร้อยละ 15 ถ้าการสำรวจยอดจำหน่ายครั้งล่าสุด โดยสุ่มถามจากร้านค้า 75 แห่งพบว่ามีผู้นิยมซื้อถึงร้อยละ 17 จงทดสอบที่ระดับ α = .05 ว่าเสื้อยืดเจบีมีผู้นิยมสูงขึ้นหรือไม่

8. จากคำถามข้อ 6 บทที่ 7 จะสรุปผลการทดสอบสมมุติฐานว่าสัดส่วนผู้ใช้ผงซักฟอกยี่ห้อ A เกินกว่าร้อยละ 50 หรือไม่ ที่ระดับ α = .05

บทที่ 9
การวิเคราะห์ความแปรปรวน

ในบทที่ 7 และบทที่ 8 เราได้ศึกษาเกี่ยวกับการอนุมานค่าเฉลี่ยประชากรด้วยตัวอย่างที่มีอยู่ในมือสำหรับการวิเคราะห์ความแปรปรวนนี้จะกล่าวถึงการแปรผันของค่าเฉลี่ยตัวอย่าง ($\overline{X}$) จากค่าเฉลี่ยประชากรซึ่งถูกประมาณค่าไว้แล้วตรวจสอบความแปรปรวนระหว่างค่าเฉลี่ยตัวอย่างทั้งหลายที่มีอยู่ โดยนำเทคนิคการวิเคราะห์ความแปรปรวนปัจจัยเดียวหรือการจำแนกทางเดียว (one-factor analysis of variance or one way classification) ของท่านโรนัลด์ ไอเมอร์ ฟิชเชอร์ (Ronald Aylmer Fisher : 1890-1962) มากล่าวเพื่อเป็นแนวทางสำหรับการวางแผนดำเนินการตัดสินใจทางธุรกิจที่มีทางเลือกหลายทางแต่ต้องเลือกแนวทางที่ดีที่สุดต่อไป

9.1 การวิเคราะห์ความแปรปรวนปัจจัยเดียว

คำว่า ปัจจัยเดียว (one-factor) ในที่นี้จะหมายถึงประชากรเดียวที่ถูกชักตัวอย่างขนาด n ออกมาหลาย ๆ ตัวอย่างและหลาย ๆ ตัวอย่างเหล่านี้ก็จะมีค่าเฉลี่ยตัวอย่าง ($\overline{X}_i$) หลาย ๆ ค่าซึ่งมีการแปรผันหรือกระจายอยู่รอบ ๆ ค่าเฉลี่ยประชากร (μ) นั้น

ถ้ากำหนด ตัวอย่างที่ i (the i^{th} sample) มีค่าเฉลี่ยเขียนแทนด้วย $\overline{X}_i$ ค่าสังเกตตัวที่ j (the j^{th} observation) ของตัวอย่างที่ i เขียนแทนด้วย X_{ij} เมื่อมีตัวอย่างทั้งหมด c ตัวอย่าง และแต่ละตัวอย่างมีขนาด n เราสามารถเขียนค่าสังเกตหรือข้อมูลด้วยสัญลักษณ์ X_{ij} เมื่อ $i = 1, 2, \ldots\ldots, c$ และ $j = 1, 2, \ldots\ldots, n$

9.2 การแปรผันระหว่างตัวอย่าง

ถ้าประชากรที่ 1, 2, ..., c. มีค่าเฉลี่ยเป็น μ_1 , μ_2 , ..., μ_c ถูกชักตัวอย่างขนาด n เป็นตัวอย่างที่ 1, 2, ..., c. คำนวณค่าเฉลี่ยตัวอย่างได้ $\overline{X}_1, \overline{X}_2, \ldots\ldots \overline{X}_c$ ตามลำดับ เมื่อพิจารณาปัจจัยเดียว ตัวอย่างเหล่านี้มาจากประชากรเดียวกันหรือ μ_1

เท่ากับ μ_2 เท่ากับ ..... เท่ากับ μ_c ส่วนความแตกต่างกันของค่า $\overline{X}_1, \overline{X}_2, \ldots\ldots \overline{X}_c$ เป็นความแปรผันระหว่างตัวอย่างในที่นี้แสดงด้วยความแปรปรวนของ $\overline{X}_i$ (variance of $\overline{X}_i$: $S^2_{\overline{x}}$) เช่น สุ่มชั่งน้ำหนักสุกรขุนอายุ 4 เดือนจากฟาร์มเลี้ยงสุกร 3 ฟาร์ม ๆ ละ 5 ตัวได้ข้อมูล ดังตารางที่ 9.1

ตารางที่ 9.1 น้ำหนักสุกรขุนจากฟาร์มเลี้ยงสุกร 3 ฟาร์ม ๆ ละ 5 ตัว

	น้ำหนักสุกร (กิโลกรัม)			
	ฟาร์มที่ 1	ฟาร์มที่ 2	ฟาร์มที่ 3	
	47	52	54	
	53	55	50	
	50	54	51	
	49	58	51	
	46	61	49	
รวม	245	280	255	$GT = 780$
$\overline{X}_i$	49	56	51	$GM = \overline{\overline{X}} = 52$
$(\overline{X}_i - \overline{\overline{X}})$	-3	4	-1	$\sum(\overline{X}_i - \overline{\overline{X}}) = 0$
$(\overline{X}_i - \overline{\overline{X}})^2$	9	16	1	$\sum(\overline{X}_i - \overline{\overline{X}})^2 = 26$

การตั้งสมมติฐาน H_0 เกี่ยวกับความไม่แตกต่างของค่าเฉลี่ยประชากร เป็นดังสมการ

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots\ldots\ldots\ldots = \mu_C \qquad \ldots(9\text{-}1)$$

ถ้าหากสมมุติฐานว่างนี้เป็นจริง ตัวอย่างเหล่านี้ถูกชักจากประชากรเดียวกันอันมีค่าเฉลี่ยประชากรเท่ากับ μ และประมาณค่าเฉลี่ยประชากรด้วย $\overline{\overline{X}}$ (grand mean of $\overline{X}_i$: *GM*) และความแปรปรวนของ $\overline{X}_i$ ตามตารางที่ 9.1 คำนวณได้ ดังนี้

$$S^2_{\overline{x}} = \frac{26}{2} = 13 \qquad \ldots(9\text{-}2)$$

สมการ (9-2) นี้คำนวณได้โดยใช้หลักทำนองเดียวกับการหาความแปรปรวนตัวอย่าง (S^2) ตามสมการ (7-6) ของบทที่ 7 เมื่อมีตัวอย่างทั้งหมด *c* ตัวอย่าง คือ *n* และ *c* - 1 คือ *df*

ระหว่างตัวอย่างทั้งหมด ทำให้สามารถเขียนความแปรปรวนของ $\overline{X}_i$ รูปทั่วไปเป็นดังสมการ (9-3)

$$s_{\bar{x}}^2 = \frac{1}{c-1}\sum_{i=1}^{c}(\overline{X}_i - \overline{\overline{X}})^2 \qquad \text{…(9-3)}$$

ถ้าสมมุติฐาน H_O ตามสมการ (9-1) เป็นจริง เราประมาณค่าความแปรปรวนของประชากร (σ^2) ได้โดยนำหลักการสมการ (6-8) จากบทที่ 6 ถ้าตัวอย่างที่ 1, 2, ... , c มีขนาด n เท่ากันแล้ว

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

และ $$\sigma_{\bar{x}}^2 \approx s_{\bar{x}}^2$$

ในที่นี้ จะได้ $$\sigma^2 = n s_{\bar{x}}^2 \qquad \text{…(9-4)}$$

9.3 การแปรผันภายในตัวอย่าง

จากตารางที่ 9.1 จะพบว่าค่าสังเกตในแต่ละตัวอย่างมีการแปรผัน หรือแกว่งอยู่รอบ ๆ $\overline{X}_i$ เราสามารถหาผลบวกกำลังสองของการเบี่ยงเบนภายในแต่ละตัวอย่าง (sum of the squared deviation within sample) เช่น ตัวอย่างที่ 1 (สุกรขุนฟาร์มที่ 1) หาได้เป็นสมการ (9-5)

$$\sum_{j=1}^{n}(X_{1j} - \overline{X}_1)^2 = (47\text{-}49)^2 + (53\text{-}49)^2 + (50\text{-}49)^2 + (49\text{-}49)^2 + (46\text{-}49)^2$$

$$= 30 \qquad \text{…(9-5)}$$

และหาผลบวกของกำลังสองเบี่ยงเบนภายในตัวอย่างที่ 2 ตัวอย่างที่ 3 ในทำนองเดียวกันได้ 50 และ 14 ตามลำดับ นำผลบวกของกำลังสองของการเบี่ยงเบนภายในตัวอย่างทั้ง 3 รวมกันแล้ว หารด้วยผลรวม *df* แต่ละตัวอย่าง (n-1) เท่ากับ (5-1) เท่ากับ 4 ผลได้จะเห็นความแปรปรวนร่วม (pooled variance : s_p^2) เช่นเดียวกับการหาความแปรปรวนร่วมของสองตัวอย่างตามสมการ (7-32) ในบทที่ 7 จะได้เป็น

$$S_p^2 = \frac{30+50+14}{4+4+4} = \frac{96}{12} = 7.83 \qquad \ldots(9\text{-}6)$$

เพื่อให้ง่ายสำหรับการหาค่า S_p^2 ทั่วไป เมื่อเรามีตัวอย่างทั้งหมดเท่ากับ c และแต่ละตัวอย่างประกอบด้วย n ค่าสังเกต จะได้สมการ (9-7)

$$S_p^2 = \frac{\sum_{j=1}^{n}(X_{1j}-\overline{X}_1)^2+\sum_{j=1}^{n}(X_{2j}-\overline{X}_2)^2+\ldots..+\sum_{j=1}^{n}(X_{cj}-\overline{X}_c)^2}{c(n-1)} \qquad \ldots(9\text{-}7)$$

คล้ายกับสมการ (7-32)

ถ้าสมมุติฐาน H_O จากสมการ (9-1) เป็นจริง จึงประมาณค่าความแปรปรวนของประชากร (σ^2) ด้วยค่า S_p^2 นี้ทำนองเดียวกับสมการ (7-32) บทที่ 7 นั้นจะได้สมการ (9-8)

$$\sigma^2 = S_p^2 \qquad \ldots(9\text{-}8)$$

9.4 การเปรียบเทียบการแปรผันระหว่างตัวอย่างกับการแปรผันภายในตัวอย่าง

ถ้าสมมุติฐาน H_O เป็นจริง อัตราส่วนค่าความแปรปรวนประชากร nS_p^2 ตามสมการ (9-4) ต่อค่าความแปรปรวนประชากร S_p^2 ตามสมการ (9-8) ควรจะมีค่าเท่ากับ 1 หรือเข้าใกล้ ๆ 1 และเราเรียกอัตราส่วน nS_p^2 ต่อ S_p^2 ว่า " F " เพื่อเป็นเกียรติแก่ท่านฟิชเชอร์ นั่นคือ

$$F = \frac{nS_{\bar{x}}^2}{S_P^2} \qquad \ldots(9\text{-}9)$$

ฟิชเชอร์ได้นำเสนอการแจกแจงของเอฟ (*F*-distribution) รูปร่างของโค้งการแจกแจงของเอฟขึ้นอยู่กับ *df* ของตัวเศษหรือตัวตั้ง (nummerator degree of freedom: df_1) และ *df* ของตัวส่วนหรือตัวหาร (denominator degree of freedom : df_2) เมื่อ *df* ของตัวเศษเท่ากับ c-1 และ *df* ของตัวส่วนเท่ากับ $c(n$-1) จะเป็น

df ของ F = c-1 และ $c(n$-1)

หรือ df ของ F = c-1 , $c(n$-1)

และใช้ตารางภาคผนวกที่ 7 และตารางภาคผนวกที่ 8 ภาคผนวกสำหรับหาค่าวิกฤตของเขตปฏิเสธ H_0 ที่ระดับ α = .01 และระดับ α = .05 ตามลำดับ

ตัวอย่างที่ 9.1 ข้อมูลตามตารางที่ 9.1

ก. จงหาอัตราส่วน F

ข. จงหา df ของ F

ค. จะปฏิเสธ H_0 ที่ระดับ α = .05 ได้หรือไม่

วิธีทำ พบว่า

ความแปรปรวนระหว่างตัวอย่าง 3 ตัวอย่าง S_p^2 = 13

และความแปรปรวนภายในของ 3 ตัวอย่าง ตัวอย่างละ 5 ค่าสังเกต

$$S_p^2 = 7.83$$

ก. $$F = \frac{nS_{\bar{x}}^2}{S_P^2}$$

$$= \frac{5(13)}{7.83} = 8.3$$

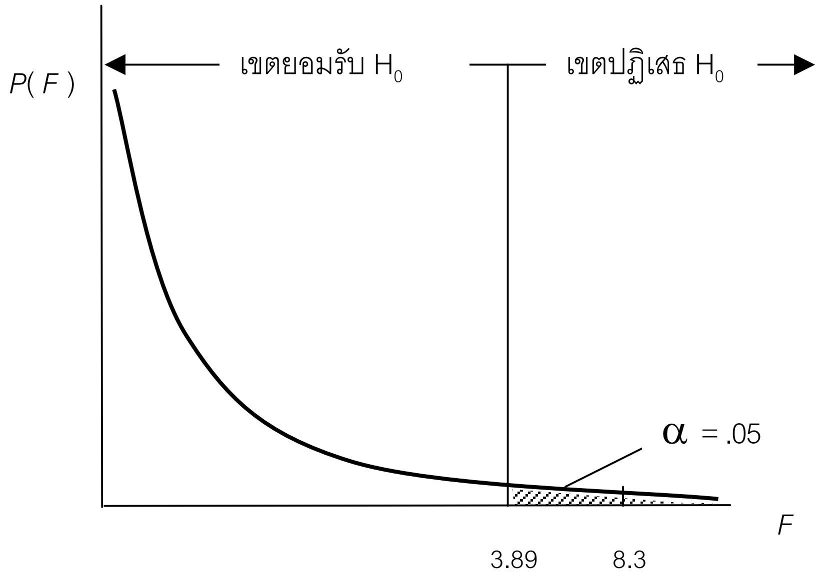
ข. df ของ F = c-1 , $c(n$-1)

$$= (3\text{-}1),3(5\text{-}1) = 2,\ 12$$

ค. สมมุติฐาน $H_O : \mu_1 = \mu_2 = \mu_3$

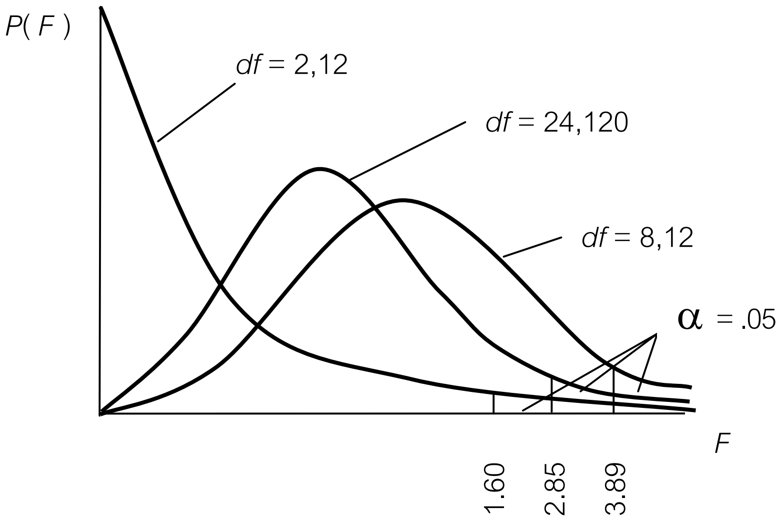
$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

จากภาพที่ 9.1 เมื่อตัวสถิติทดสอบ F_c ที่คำนวณได้ ตกอยู่ในเขตปฏิเสธ H_0 จึงตัดสินใจปฏิเสธ H_0 ที่ระดับ α = .05 ได้ นั่นคือ ตัวอย่างทั้ง 3 ไม่ได้มาจากประชากรเดียวกัน หรือประชากรทั้ง 3 แตกต่างกันนั่นเอง



ภาพที่ 9.1 เขตยอมรับและเขตปฏิเสธ H_0

ที่มา : (ดัดแปลงมาจาก Watson et. al.,1990, p. 415)



ภาพที่ 9.2 การแจกแจงของ *F* ที่มี *df* แตกต่างกันและ

เขตปฏิเสธ H_0 ที่ระดับ $\alpha = .05$

ที่มา : (Wonnacott & Wonnacott, 1985, p. 296)

จากสมการ (9-4) และ (9-8) จะเห็นว่าแหล่งหรือสิ่งที่ทำให้เกิดความแปรปรวนมีอยู่ 2 ลักษณะคือ

9.4.1 ความแปรปรวนที่เกิดขึ้น ระหว่างตัวอย่างกับภายในตัวอย่าง และความแปรปรวนที่เกิดขึ้นภายในตัวอย่าง สำหรับความแปรปรวนระหว่างตัวอย่างเราพอจะอธิบายสาเหตุได้ เช่น ความแปรปรวนเกิดเพราะสุ่มชั่งน้ำหนักสุกรจากฟาร์มต่างกันหรือในงานทดลองอาจอธิบายว่า เกิดจากการปฏิบัติต่อวัตถุทดลองแตกต่างกัน (between treatment) หรือเกิดจากปัจจัยแตกต่างกัน (between factor) เป็นต้น จึงเรียกความแปรปรวนประเภทนี้ว่า ความแปรปรวนที่อธิบายได้ (explained variance)

9.4.2 ความแปรปรวนที่เกิดขึ้น ภายในตัวอย่าง เช่น ภายในฟาร์มเดียวกัน ภายใต้การปฏิบัติเช่นเดียวกันหรือภายใต้ปัจจัยเดียวกันเป็น ความแปรปรวนที่อธิบายไม่ได้ (unexplained variance) หรือที่เรียกว่า ค่าผิดพลาด หรือ ความคลาดเคลื่อน ทำให้อัตราส่วนเอฟ (*F*-ratio หรือ *F*- calculate มีสัญลักษณ์เป็น F_c) เขียนได้เป็น

$$F = \frac{\text{explained variance}}{\text{unexplained variance}} \qquad \ldots(9\text{-}10)$$

9.5 ตารางวิเคราะห์ความแปรปรวน

ตารางวิเคราะห์ความแปรปรวนเป็นตารางที่นิยมใช้ ประกอบด้วย 5 สดมภ์ คือ สดมภ์ที่ 1 เป็นแหล่งความแปรปรวน (source of variance เขียนย่อ *SV.*) สดมภ์ที่ 2 เป็นการแปรผัน หรือผลบวกกำลังสองของการเบี่ยงเบน (*SS.*) สดมภ์ที่ 3 เป็น *df* สดมภ์ที่ 4 เป็นความแปรปรวน สดมภ์ที่ 5 เป็น อัตราส่วนเอฟที่คำนวณได้ นอกจากนี้อาจจะมีสดมภ์ที่ 6 แสดงค่าวิกฤตที่จะปฏิเสธ H_0 ที่ระดับ α ซึ่งอ่านได้จากตารางภาคผนวกที่ 7 หรือ ตารางภาคผนวกที่ 8

สำหรับการวิเคราะห์ความแปรปรวนปัจจัยเดียวข้อมูลจากตารางที่ 9.1 ประกอบด้วยตัวอย่างทั้งหมด 3 ตัวอย่าง แต่ละตัวอย่างมี 5 ค่าสังเกตเท่ากันหมด สามารถแยกส่วนประกอบของตารางวิเคราะห์ความแปรปรวนได้ดังนี้

9.5.1 แหล่งความแปรปรวน แยกเป็น 2 ส่วน คือ

1) ระหว่างฟาร์ม ในที่นี้จะใช้คำว่า ระหว่างสดมภ์ (between columns เขียนย่อเป็น col.)

2) ภายในฟาร์ม ในที่นี้จะใช้คำว่า เศษตกค้าง (residual เขียนย่อ res.) หรือ ค่าผิดพลาด

9.5.2 ผลบวกกำลังสองของการเบี่ยงเบน (*SS.*) มี 3 ส่วน คือ

1) SS_{col} ใช้สมการ (9-3) และสมการ (9-4) จะได้

$$SS_{col} = n\sum_{i=1}^{c}(\overline{X}_i - \overline{\overline{X}})^2 \qquad \ldots(9\text{-}11)$$

$$SS_{col} = (5)(26) = 130$$

2) SS_{res} จากสมการ (9-7)

$$SS_{res} = \sum_{j=1}^{n}(X_{1j} - \overline{X}_1)^2 + \sum_{j=1}^{n}(X_{2j} - \overline{X}_2)^2 + \ldots.. + \sum_{j=1}^{n}(X_{Cj} - \overline{X}_C)^2 \ \ldots(9\text{-}12)$$

$$SS_{res} = 30 + 50 + 14 = 94$$

3) SS_{Tot} หมายถึง ผลบวกกำลังสองของการเบี่ยงเบน (sum squares of total) ของค่าสังเกตทั้งหมดจากค่าเฉลี่ยของทั้งหมด (grand mean of all the X_{ij}: GM หรือ overal mean: $\overline{\overline{X}}$) ดังสมการ (9-13)

$$SS_{Tot} = \sum_{i=1}^{c}\sum_{J=1}^{n}(X_{iJ} - \overline{\overline{X}})^2 \qquad \ldots(9\text{-}13)$$

$$SS_{Tot} = (47\text{-}52)^2 + (53\text{-}52)^2 + (50\text{-}52)^2 + \ldots.. + (51\text{-}52)^2 + (49\text{-}52)^2$$

$$= 25 + 1 + 4 + \ldots\ldots\ldots\ldots + 1 + 9 = 224$$

จากสมการ (9-11), (9-12) และ (9-13) เป็นข้อพิสูจน์ขั้นต้นแสดงให้เห็นว่า

$$224 = 130 + 94$$

นั่นคือ $SS_{Tot} = SS_{col} + SS_{res}$ …(9-14)

9.5.3 *df* มี 3 ส่วน คือ

1) สดมภ์ *df* จากสมการ(9-3)

สดมภ์ $df = c - 1$ ...(9-15)

สดมภ์ $df = 3\text{-}1 = 2$

2) เศษตกค้าง *df* ใช้สมการ (9-7)

$$\text{เศษตกค้าง } df = c(n-1) \quad \ldots(9\text{-}16)$$

$$\text{เศษตกค้าง df} = 3(5-1) = 12$$

3) ผลรวม *df* หาได้โดยใช้สมการ (9-17)

$$\text{ผลรวม } df = N-1 \quad \text{โดย } N \text{ คือ จำนวนค่าสังเกตทั้งหมด} \quad \ldots(9\text{-}17)$$

$$\text{ผลรวม } df = 15-1 = 14$$

9.5.4 ค่าเฉลี่ยของผลบวกกำลังสองของการเบี่ยงเบน (*MS*.)มี 2 ส่วน คือ

1) MS_{col} หรือ nS_p^2 หาได้โดยใช้สมการ (9-18)

$$MS_{col} = \frac{SS_{col}}{c-1} \quad \ldots(9\text{-}18)$$

$$MS_{col} = \frac{130}{2} = 65$$

2) $MS_{res} = S_p^2$ และหาได้โดยใช้สมการ (9-19)

$$MS_{res} = \frac{SS_{col}}{c(n-1)} \quad \ldots(9\text{-}19)$$

$$= \frac{94}{12} = 7.83$$

9.5.5 อัตราส่วน *F* ที่คำนวณได้ : F_c และใช้ได้โดยใช้สมการ (9-20)

$$F = \frac{MS_{col}}{MS_{res}} \quad \ldots(9\text{-}20)$$

$$F_c = \frac{65}{7.83} = 8.3$$

เมื่อ F_c = 8.3 $>$ 6.9266 ; ค่า F_c ตกในเขตปฏิเสธปฏิเสธ H_0 ที่ระดับ α = .01 นิยมใส่เครื่องหมาย ** ไว้เหนือค่า F_c สำหรับแสดงว่า น้ำหนักสุกรขุน 3 ฟาร์มแตกต่างกันอย่างมีนัยสำคัญยิ่ง อย่างน้อยหนึ่งคู่ (2 ดอก) แต่ยังบอกไม่ได้ว่าคู่ใดแตกต่างกันบ้าง ต้องอาศัยการตรวจพินิจค่าเฉลี่ยประชากรซึ่งจะกล่าวในหัวข้อ 9.6 ต่อไป

ตารางที่ 9.2 การวิเคราะห์ความแปรปรวนของข้อมูลจากตารางที่ 9.1

SV.	*df*	*SS.*	*MS.*	F_C
ระหว่างสดมภ์	2	130	65	8.3 **
เศษตกค้าง	12	94	7.83	
รวม	14	224		

ที่ df =2,12; $F_{.01}$ = 6.9266, $F_{.05}$= 3.8853

สำหรับงานทดลองที่มีความมุ่งหมายเพื่อการเปรียบเทียบกรรมวิธี เราสามารถสรุปผลการทดลองได้ 3 ลักษณะจากตารางวิเคราะห์ความแปรปรวนคือ

(1) ถ้า F_c มีค่าน้อยกว่า $F_{.05}$; เราจะสรุปผลการทดลองว่า กรรมวิธี ไม่มีความแตกต่างกันอย่างมีนัยสำคัญ (*ns*)

(2) ถ้า F_c มีค่ามากกว่า $F_{.05}$; แต่ยังน้อยกว่า $F_{.01}$ จะสรุปผลการทดลองว่า กรรมวิธี มีความแตกต่างกันอย่างมีนัยสำคัญ

(3) ถ้า F_c มีค่ามากกว่า $F_{.01}$ (มากกว่า $F_{.05}$ ด้วย) จะสรุปผลการทดลองว่า กรรมวิธี มีความแตกต่างกันอย่างมีนัยสำคัญยิ่ง การสร้างตารางวิเคราะห์ความแปรปรวนสำหรับการวิเคราะห์ความแปรปรวนปัจจัยเดียวของข้อมูลประกอบด้วย c ตัวอย่าง แต่ละตัวอย่างมี n ค่าสังเกตเท่ากันหมด ควรดำเนินการคำนวณค่าต่าง ๆ เพื่อนำบรรจุลงในตารางวิเคราะห์ความแปรปรวนอย่างระมัดระวัง โดยอาศัยรูปทั่วไปของการคำนวณค่าต่าง ๆ ตามสมการ (9-11) ถึง (9-20)

กรณีตัวอย่างที่ 1, 2, 3, ... , c มีขนาด n_1 , n_2,......., n_c ตามลำดับ จะมีโครงสร้างของข้อมูลและกำหนด สดมภ์ 1 สดมภ์ 2 .... สดมภ์ c คือ ตัวอย่างที่ 1,2.... , c ตามลำดับ *GT* คือ ผลรวมข้อมูลทั้งหมดและ *N* คือ จำนวนข้อมูลทั้งหมด ดังตารางที่ 9.3

ตารางที่ 9.3 โครงสร้างของข้อมูล : X_{ij} ตัวอย่างที่ 1, 2, ... , c มีขนาด $n_1 , n_2 \ldots , n_c$ ตามลำดับ

	สดมภ์ 1	สดมภ์ 2	....	สดมภ์ c	
	X_{11} X_{12} : : : : X_{1n1}	X_{21} X_{22} : : : : X_{2n1}	....	X_{c1} X_{c2} : : : : X_{cnc}	
รวม (T_i)	T_1	T_2	....	T_c	$\sum_{i=1}^{c} T_i = GT$
ขนาด (n_i)	n_1	n_2	....	n_c	$\sum_{i=1}^{c} n_i = N$
ค่าเฉลี่ย ($\overline{X}_i$)	$\overline{X}_1$	$\overline{X}_2$	....	$\overline{X}_c$	GM $= \overline{\overline{X}}$
$(\overline{X}_i - \overline{\overline{X}})$	$(\overline{X}_1 - \overline{\overline{X}})$	$(\overline{X}_2 - \overline{\overline{X}})$	....	$(\overline{X}_c - \overline{\overline{X}})$	$\sum_{i=1}^{c} (\overline{X}_i - \overline{\overline{X}}) = 0$
$(\overline{X}_i - \overline{\overline{X}})^2$	$(\overline{X}_1 - \overline{\overline{X}})^2$	$(\overline{X}_2 - \overline{\overline{X}})^2$	....	$(\overline{X}_c - \overline{\overline{X}})^2$	$\sum_{i=1}^{c} (\overline{X}_i - \overline{\overline{X}})^2 \neq 0$

ตารางที่ 9.4 การวิเคราะห์ความแปรปรวนของข้อมูลจากตารางที่ 9.3

SV.	*df*	*SS.*	*MS.*	F_c
ระหว่างสดมภ์	c-1	$SS_{col} = \sum_{i=1}^{c} n_i \; (\overline{X}_i - \overline{\overline{X}})^2$	$\dfrac{SS_{col}}{c-1}$	$\dfrac{MS_{col}}{MS_{res}}$
เศษตกค้าง	N-c	$SS_{res} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2$	$\dfrac{SS_{res}}{(N-c)}$	
รวม	N-1	$SS_{Tot} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (X_{ij} - \overline{\overline{X}})^2$		

จาก $$SS_{Tot} = \sum_{i=1}^{c}\sum_{j=1}^{n_i}(X_{ij} - \overline{\overline{X}})^2$$

$$= \sum_{i=1}^{c}\sum_{j=1}^{n_i}(X_{ij}^2 - 2X_{ij}\overline{\overline{X}} + \overline{\overline{X}}^2)$$

$$= \sum_{i=1}^{c}\sum_{j=1}^{n_i}(X)_{ij}^2 - 2\overline{\overline{X}}\sum_{i=1}^{c}\sum_{j=1}^{n_j}(X_{ij}) + \sum_{i=1}^{c}\sum_{j=1}^{n_j}\overline{\overline{X}}^2)$$

$$= \sum_{i=1}^{c}\sum_{j=1}^{n_i}(X)_{ij}^2 - 2\frac{GT}{N}(GT) + N(\frac{GT}{N})^2$$

$$= \sum_{i=1}^{c}\sum_{j=1}^{n_i}(X)_{ij}^2 - 2\frac{(GT)^2}{N} + \frac{(GT)^2}{N}$$

จึงได้ $$SS_{Tot} = \sum_{i=1}^{c}\sum_{j=1}^{n_i}(X)_{ij}^2 - \frac{(GT)^2}{N} \quad \text{…(9-21)}$$

และจาก $$SS_{col} = \sum_{i=1}^{c}n_i(\overline{X}_i - \overline{\overline{X}})^2$$

$$= \sum_{i=1}^{c}n_i\ (\overline{X}_i^2 - 2\overline{X}_i\overline{\overline{X}} + \overline{\overline{X}}^2)$$

$$= \sum_{i=1}^{c}n_i\ \overline{X}_i^2 - 2\overline{\overline{X}}\sum_{i=1}^{c}n_i\overline{X}_i + \overline{\overline{X}}^2\sum_{i=1}^{c}n_i$$

$$= \sum_{i=1}^{c}n_i(\frac{T_i}{n_1})_i^2 - 2(\frac{GT}{N})GT + (\frac{GT}{N})^2 N$$

จึงได้ $$SS_{col} = \sum_{i=1}^{c}\frac{T_i^2}{n_1} - \frac{(GT)^2}{N} \quad \text{…(9-22)}$$

ประกอบกับสมการ (9-14) ทำให้ได้ดังนี้

$$SS_{res} = SS_{Tot} - SS_{col} \quad \text{…(9-23)}$$

สำหรับเทอม $\frac{(GT)^2}{N}$ ในสมการ (9-21) และ (9-22) เรียกว่า ตัวแก้ หรือ ตัวประกอบแก้ (corrector หรือ correction factor เขียนย่อว่า *CF*) ทางปฏิบัติจึงนำค่า

$$CF \quad = \quad \frac{(GT)^2}{N} \qquad \ldots(9\text{-}24)$$

แทนในสมการ (9-21) และ (9-22) ส่วนเทอม $\sum_{i=1}^{c}\sum_{j=1}^{n_i}(X)_{ij}^2$ ในสมการ (9-21) คือผลบวกของแต่ละค่าสังเกตที่มีอยู่ทั้งหมดยกกำลังสอง ทางปฏิบัติจึงเขียนเป็น

$$SS_{Tot} \;=\; \sum_{All}\left(Each \quad value\right)^2 - CF \qquad \ldots(9\text{-}25)$$

ตารางที่ 9.5 การวิเคราะห์ความแปรปรวนของข้อมูลจากตารางที่ 9.3 และ ตารางที่ 9.4

SV.	*df*	*SS.*	*MS.*	F_C
ระหว่างสดมภ์	c-1	$SS_{col} = \sum_{i=1}^{c}\frac{T_i^2}{n_1} - CF$	$\frac{SS_{col}}{c-1}$	$\frac{MS_{col}}{MS_{res}}$
เศษตกค้าง	N-c	$SS_{res} = SS_{Tot}\text{-}SS_{col}$	$\frac{SS_{res}}{(N-c)}$	
รวม	N-1	$SS_{Tot} = \sum_{All}\left(Each \quad value\right)^2 - CF$		

9.6 การตรวจพินิจความแตกต่างของค่าเฉลี่ยประชากร

การตรวจพินิจความแตกต่างกันของค่าเฉลี่ยประชากร เป็นกระบวนการต่อจากการวิเคราะห์ความแปรปรวน เมื่อผลการวิเคราะห์ความแปรปรวนออกมาแสดงถึงความแตกต่างระหว่างแนวตั้ง อย่างมีนัยสำคัญหรือมีนัยสำคัญยิ่ง กรณีตัวอย่างขนาดเล็กจะต้องอยู่ภายใต้ข้อสมมุติว่า ประชากรแจกแจงปรกติหรือประชากรแจกแจงปกติโดยประมาณและตัวอย่างอิสระกัน ซึ่งการตรวจสอบข้อสมมุติินี้ได้หลายวิธี เช่น การทดสอบของบาร์ตเลตต์ (Bartlett' s test) การทดสอบของฮาร์ตลีย์ (Hartley's test) หรือ การทดสอบของคอกครัน (Cochran's test) ซึ่งจะได้ศึกษาในหัวข้อ 9.7 สำหรับการตรวจพินิจความแตกต่างกระทำโดยการจัดค่าเฉลี่ยตัวอย่างที่มีอยู่เป็นคู่ ๆ ซึ่งเรียกว่า มัลติเพิลคอมแพริซันมีน (multiple comparisons mean) ซึ่งสมมุติฐานที่ทดลองคือ

$$H_O : \mu_1 = \mu_2 \text{ หรือ } \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \quad \text{หรือ} \quad \mu_1 - \mu_2 \neq 0$$

9.6.1 วิธีแอลเอสดี (least significant difference : LSD) เป็นวิธีที่ง่ายที่สุดและนิยมใช้ในทางปฏิบัติมากที่สุด โดยทั่วไปจะไม่ได้วางแผน เกี่ยวกับความแตกต่างกันของค่าเฉลี่ยประชากรให้ล่วงหน้า จึงมักเรียกคู่ความแตกต่าง (unexplained comparisons) กล่าวคือ ถ้า $\bar{X}_1$ และ $\bar{X}_2$ เป็นค่าเฉลี่ยตัวอย่างที่จะนำมาหาความแตกต่างกับของค่าเฉลี่ยประชากร

กรณี $\bar{X}_1 - \bar{X}_2 \geq LSD_{\alpha}$ …(9-26)

แสดงว่าค่าเฉลี่ยประชากรคู่นี้ แตกต่างกันอย่างมีนัยสำคัญหรือแตกต่างกันอย่างมีนัยสำคัญยิ่ง ขึ้นอยู่กับระดับ $\alpha = .05$ หรือระดับ $\alpha = .01$ ตามลำดับ

แต่กรณี $\bar{X}_1 - \bar{X}_2 \leq LSD_{\alpha}$ …(9-27)

แสดงว่าค่าเฉลี่ยประชากรคู่นี้ ไม่แตกต่างกันอย่างมีนัยสำคัญ (*ns*)

ในที่นี้จะขอกล่าวเฉพาะกรณีตัวอย่างขนาดเล็ก ($n < 20$) และข้อสมมุติเป็นจริง ตัวสถิติที่ทดสอบที่เหมาะสมเป็นดังสมการ (9-28)

$$t_c = \frac{(\bar{X}_1 - \bar{X}_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad \text{…(9-28)}$$

สมมุติฐาน คือ $H_O : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

เป็นการทดลองสองทางจะปฏิเสธ H_O ได้

เมื่อ $\bar{X}_1 - \bar{X}_2 \geq t_{\alpha/2}\ S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

หรือ $\bar{X}_1 - \bar{X}_2 \leq -t_{\alpha/2}\ S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

ดังนั้น $$LSD_{\alpha/2} = t_{\alpha/2}\ S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \qquad ...(9\text{-}29)$$

โดย S_p^2 คือ ความแปรปรวนร่วมมีค่าเท่ากับ MS_{res} หรือ MSe (mean squares of residual or error) n_1 , n_2 คือขนาดตัวอย่างที่ให้ค่าเฉลี่ย $\overline{X}_1$ และ $\overline{X}_2$ ตามลำดับ (ในการวางแผนการทดลองคือจำนวนซ้ำของกรรมวิธี) และ $t_{\alpha/2}$ มี df = เศษตกค้าง df หรือค่าผิดพลาด df สำหรับกรณีที่กำหนดระดับ $\alpha = .05$ หรือระดับ $\alpha = .01$

จะได้ $$LSD_{.05/2} = t_{.05/2}\sqrt{MSe\frac{1}{n_1}+\frac{1}{n_2}} \qquad ...(9\text{-}30)$$

และ $$LSD_{.01/2} = t_{.01/2}\sqrt{MSe\frac{1}{n_1}+\frac{1}{n_2}} \qquad ...(9\text{-}31)$$

สำหรับกรณี $n_1 = n_2 = n$

จะได้ $$LSD_{.05/2} = t_{.05/2}\sqrt{\frac{2MSe}{n}} \qquad ...(9\text{-}32)$$

และ $$LSD_{.01/2} = t_{.01/2}\sqrt{\frac{2MSe}{n}} \qquad ...(9\text{-}33)$$

การดำเนินการตรวจพินิจความแตกต่างกันของค่าเฉลี่ยประชากร จะกล่าวโดยสรุปเป็นลำดับขั้นและใช้ข้อมูลจากตารางที่ 9.1 เนื่องผลการวิเคราะห์ความแปรปรวนข้อมูลนี้จากตารางที่ 9.2 แสดงให้เห็นว่าค่าเฉลี่ยประชากรแตกต่างกันอย่างมีนัยสำคัญยิ่ง โดยมีขั้นตอนดังนี้

1) ขั้นที่ 1 จัดเรียงลำดับค่าเฉลี่ยตัวอย่างจากค่าเฉลี่ยน้อยไปหามาก หรือจากมากไปหาน้อย

2) ขั้นที่ 2 คำนวณค่า $LSD_{.05/2}$ และ $LSD_{.01/2}$ กรณีนี้แต่ละตัวอย่างมีขนาดเท่ากัน จึงใช้สมการ (9-32) และ สมการ (9 -33)

เมื่อ $n = 5$, $MSe = 7.83$ และที่ $df = 12$, $t_{.05/2} = 2.179$, $t_{.01/2} = 3.055$

$$LSD_{.05/2} = 2.179\sqrt{\frac{2(7.83)}{5}}$$

$$= 2.179\ (1.77) \qquad = 3.86$$

ตารางที่ 9.6 ลำดับค่าเฉลี่ยตัวอย่าง

ลำดับที่	1	2	3
ชื่อตัวอย่าง	ฟาร์มที่ 2(B)	ฟาร์มที่ 3(C)	ฟาร์มที่ 1(A)
ค่าเฉลี่ย	56	51	49

$$LSD_{.01/2} = 3.055\sqrt{\frac{2(7.83)}{5}}$$

$$= 3.055(1.77) \qquad = 5.41$$

3) ขั้นที่ 3 หาผลต่างของค่าเฉลี่ยทุกคู่ (ไม่ซ้ำคู่) โดยนำค่าเฉลี่ยมากตั้งลบด้วยค่าเฉลี่ยน้อย

$$B–A = 56\text{-}49 = 7^{**}$$

$$B–C = 56\text{-}51 = 5^{*}$$

$$C–A = 51\text{-}49 = 2^{ns}$$

4) ขั้นที่ 4 สรุปผลความแตกต่างกันของค่าเฉลี่ยประชากร

ความหมายของ B–A = 56–49 = 7 มากกว่า $LSD_{.01/2}$ แสดงว่าค่าเฉลี่ยประชากรคู่นี้แตกต่างกันอย่างมีนัยสำคัญยิ่ง จึงใช้เครื่องหมาย ** ; B–C = 56–51 = 5 มากกว่า $LSD_{.05/2}$ แต่ยังน้อยกว่า $LSD_{.01/2}$ แสดงว่าค่าเฉลี่ยประชากรคู่นี้แตกต่างกันอย่างมีนัยสำคัญยิ่ง จึงได้เครื่องหมาย * สำหรับ C–A = 51–49 = 2 น้อยกว่า $LSD_{.05/2}$ (น้อยกว่า$LSD_{.05/2}$ ด้วย) แสดงว่าค่าเฉลี่ยประชากรคู่นี้ไม่แตกต่างกันอย่างมีนัยสำคัญ จึงใช้เครื่องหมาย ^{ns}

หรืออาจสรุปว่าเป็นสัญลักษณ์ $\underline{A \quad C}$ $\quad B$

ซึ่งมีความหมายว่า A และ C ไม่แตกต่างกันอย่างมีนัยสำคัญนอกนั้นแตกต่างกันอย่างมีนัยสำคัญ

9.6.2 วิธีดีเอ็มอาร์ที (duncan's new multiple range test : DMRT) เป็นการทดสอบความแตกต่างกันของค่าเฉลี่ยด้วยวิธีแอลเอสดี จะพบว่าเมื่อค่าเฉลี่ยที่นำมาทดสอบมีจำนวนมาก ๆ จะเกิดความสับสนของคู่แตกต่าง จึงนิยมใช้วิธีแอลเอสดีทดสอบความแตกต่างค่าเฉลี่ยกรณีมีตัวอย่างจำนวนมากในคราวเดียวกัน (ทั่วไปไม่เกิน 20 ตัวอย่าง หรือ

20 กรรมวิธีในงานทดลอง) การดำเนินการทดสอบความแตกต่างกันของค่าเฉลี่ยด้วยวิธีดีเอ็มอาร์ที มีลำดับขั้นดังนี้

1) ขั้นที่ 1 เรียงลำดับค่าเฉลี่ยตัวอย่างเช่นเดียวกับวิธีแอลเอสดี

2) ขั้นที่ 2 คำนวณค่า LSR_{α} (least significant range) โดยอาศัยค่า SSR_{α} (significant studentized range) จากตารางภาคผนวกที่ 9 และตารางภาคผนวกที่ 10

กำหนดให้

$$LSR_{\alpha} = SSR_{\alpha}\sqrt{\frac{2MSe}{n}} \qquad \text{…(9-34)}$$

เมื่อ $\sqrt{\frac{2MSe}{n}}$ คือค่าผิดพลาดมาตราฐานของค่าเฉลี่ยตัวอย่าง และตัวอย่างขนาด n เท่ากันทุกตัวอย่าง เช่น จากข้อมูลในตารางที่ 9.1 เราได้ค่า

$$\sqrt{\frac{2MSe}{n}} = \sqrt{\frac{7.83}{5}} = 1.566 = 1.25$$

การอ่านค่า $SSR_{.05}$ และ $SSR_{.01}$ จากตารางภาคผนวกที่ 9 และ ตารางภาคผนวกที่ 10 ถ้าตัวอย่างหรือจำนวณค่าเฉลี่ยเท่ากับ c แต่ละตัวอย่างมีขนาด n เท่ากันหมด จะได้ เศษตกค้าง df เท่ากับ $c(n\text{-}1)$ หรือ $N\text{-}c$ และเมื่อกำหนด p คือจำนวนค่าเฉลี่ยที่นำมาหาความแตกต่างตามลำดับที่จัดไว้ตามขั้นที่ 1 ซึ่งอย่างน้อยจะต้องมีจำนวน 2 ค่าเฉลี่ยและอาจเขียนได้เป็น

$$p = (\text{ผลต่างลำดับ}) + 1 \qquad \text{…(9-35)}$$

ตามข้อมูลตารางที่ 9.1 มีความผิดพลาด $df = 12$; ค่าเฉลี่ยมีทั้งหมด 3 ค่า จึงคำนวณค่า $LSR_{.05}$ และ $LSR_{.01}$ ตามสมการ (9-34) และมีจำนวนค่าเฉลี่ยในการหาความแตกต่าง 2 ค่าและ 3 ค่า ดังตารางที่ 9.7

3) ขั้นที่ 3 เราสามารถจับคู่หาความแตกต่างกันของค่าเฉลี่ยได้ตามความต้องการ เช่น ต้องหาความแตกต่างกันของ B กับ A เมื่อ B เป็นลำดับที่ 1 และ A เป็นลำดับที่ 3

$$p = (3 - 1) + 1 = 3$$

ค่าเฉลี่ยแตกต่าง : $56 - 49 = 7 > 5.63$ ($LSR_{.01}$)

แสดงว่า B กับ A ค่าเฉลี่ยประชากรแตกต่างกันอย่างมีนัยสำคัญยิ่ง

ตารางที่ 9.7 ค่า $SSR_{.\alpha}$ และ $LSR_{.\alpha}$

p	2	3
$SSR_{.05}$	3.08	3.23
$SSR_{.01}$	4.32	4.50
$LSR_{.05}$	3.85	4.03
$LSR_{.01}$	5.40	5.63

กรณีความแตกต่างกันของ C กับ A เมื่อ C อยู่ลำดับที่ 2 และ A อยู่ลำดับที่ 3

$$p = (3-2)+1 = 2$$

ค่าเฉลี่ยแตกต่าง : $51-49 = 2 < 3.85 \; (LSR_{.05})$

แสดงว่า C กับ A ค่าเฉลี่ยประชากรไม่แตกต่างกันอย่างมีนัยสำคัญ

ในกรณีที่หาความแตกต่างกันของ C กับ B เมื่อ C อยู่ลำดับที่ 2 และ B อยู่ลำดับที่ 1

$$p = (2-1)+1 = 2$$

ค่าเฉลี่ยแตกต่าง : $56-51 = 4 > 3.85 \; (LSR_{.05})$

แสดงว่า C กับ B ค่าเฉลี่ยประชากรแตกต่างกันอย่างมีนัยสำคัญ

4) ขั้นที่ 4 สรุปผลความแตกต่างกันของค่าเฉลี่ยประชากร

<u>C A</u> B

9.6.3 วิธีเปรียบเทียบออร์โธโกนัล (orthogonal comparison) การทดสอบความแตกต่างค่าเฉลี่ยประชากรวิธีนี้ เป็นวิธีที่นิยมใช้ในการหาความแตกต่างกันของกรรมวิธี ในการทดลองและการวางแผนเกี่ยวกับความแตกต่างกันไว้ล่วงหน้า (planed comparisons) ถ้าสมมุติว่า เป็นจริง กล่าวคือตัวอย่าง 1, 2, ..., c ขนาด n_1, n_2, ..., n_C และได้จากประชากรที่มีค่าเฉลี่ย μ_1, μ_2, ..., μ_C ตามลำดับ ประชากรแจกแจงปรกติหรือแจกแจงปรกติโดยประมาณและตัวอย่างอิสระต่อกัน ฟังก์ชันเชิงเส้นของค่าเฉลี่ย

ประชากรทั้งหมด (linear function of population means) ซึ่งอย่างน้อยต้องมี 2 ค่าเฉลี่ย มีรูปเป็น

$$L = \sum_{i=1}^{c} a_i \mu_i$$

โดย L เป็นค่าเปรียบเทียบ หรืออาจใช้คำว่าคู่เปรียบเทียบ หรือ คู่เปรียบเทียบความแตกต่างของค่าเฉลี่ยประชากรและ a_i คือสัมประสิทธิ์ของค่าเปรียบเทียบ (coefficient of comparisons) ซึ่งมีค่าเป็น

$$\sum_{i=1}^{c} a_i = 0$$

ขณะที่เรามีค่าเฉลี่ยตัวอย่าง $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$,...., $\overline{X}_c$ อยู่ในมือ การประมาณค่าที่ไม่อคติของ L (อาจใช้ $\hat{L}$ เพราะเป็นตัวสถิติ) คือ

$$L = \sum_{i=1}^{c} a_i \overline{X}_i$$

ค่าความแปรปรวนของ L คือ $V(L)$ โดยการประมาณค่าหาได้จาก

$$V(L) = V(\sum_{i=1}^{c} a_i \overline{X}_i \;)$$

$$= V(a_1\overline{X}_1 + a_2\overline{X}_2 + a_3\overline{X}_3 ..... + a_c\overline{X}_c)$$

$$= V(a_1\overline{X}_1) + V(a_2\overline{X}_2) + V(a_3\overline{X}_3) ..... + V(a_c\overline{X}_c)$$

กรณีตัวอย่างที่ 1, 2,............. , c, มีขนาด n_1 , n_2,..........., n_c ตามลำดับและข้อสมมุติเป็นจริงประกอบกับสมการ (3-23) ในบทที่ 3

$$V(a_i\overline{X}_i) = a_i^2 V(\overline{X}_i)$$

$$= a_i^2 \frac{\sigma^2}{n_i}$$

$$= \sigma^2 \frac{a_i^2}{n_i}$$

ทำให้ได้ $$V(L) = \sum_{i=1}^{c} \sigma^2 \frac{a_i^2}{n_i}$$

$$= \sigma^2 \sum_{i=1}^{c} \frac{a_i^2}{n_i}$$

เมื่อ σ^2 เป็นค่าคงตัวและค่าประมาณที่ไม่อคติของ σ^2 คือ *Mse* (MS_{res})

ดังนั้น $$V(L) = MSe \sum_{i=1}^{c} \frac{a_i^2}{n_i}$$

สมมุติฐานเกี่ยวกับความไม่แตกต่างกันของค่าเฉลี่ยประชากร (no difference of population means) ในการเปรียบเทียบค่าใดค่าหนึ่ง ดังนี้

$$H_0 : L = 0 \quad \text{หรือ} \quad \sum_{i=1}^{c} a_i \overline{X}_i = 0$$

$$H_1 : L \neq 0 \quad \text{หรือ} \quad \sum_{i=1}^{c} a_i \overline{X}_i \neq 0$$

ตัวสถิติทดสอบคือ $$F_c = \frac{(\sum_{i=1}^{c} a_i^2 \ \overline{X}_i)^2}{MSe \sum_{i=1}^{c} \frac{a_i^2}{n_i}}$$

เทอม $\dfrac{(\sum_{i=1}^{c} a_i^2 \ \overline{X}_i)^2}{\sum_{i=1}^{c} \frac{a_i^2}{n_i}}$ เรียกว่าผลบวกกำลังสองของค่าเปรียบเทียบ (sum square of comparison : *SSL*) และมี *df* เท่ากับ 1 นั่นคือ

$$SSL = \frac{(\sum_{i=1}^{c} a_i^2 \ \overline{X}_i)^2}{\sum_{i=1}^{c} \frac{a_i^2}{n_i}}$$

แต่ $\overline{X}_i = \dfrac{T_i}{n_i}$ (โดย T_i คือผลรวมของแต่ละตัวอย่าง) จะได้

$$SSL = \frac{(\sum_{i=1}^{c} a_i \frac{T_i}{n_i})^2}{\sum_{i=1}^{c} \frac{a_i^2}{n_i}} \qquad \text{…(9-36)}$$

กรณีตัวอย่างที่ 1, 2, ... , c มีขนาด n เท่ากันหมด (คือ n_1 เท่ากับ n_2 เท่ากับ ...... เท่ากับ n_c เท่ากับ n) จะได้

$$SSL = \frac{(\sum_{i=1}^{c} a_i^2 T_i)^2}{n\sum_{i=1}^{c} a_i^2} \qquad \text{…(9-37)}$$

และเขียนตัวสถิติทดสอบได้เป็น

$$F_c = \frac{SSL}{MSe} \qquad \text{…(9-38)}$$

จะปฏิเสธ H_0 ที่ระดับ α และ df = (1, N–c) หรือ (1,c(n-1))

สำหรับความรู้เบื้องต้นนี้ขอนำกรณีตัวอย่างขนาด n เท่ากันหมดกล่าวคือ ถ้าเรากำหนด ค่าเปรียบเทียบ 2 ค่า เช่น L_1เท่ากับ $\sum_{i=1}^{c} a_i \mu_i$ และL_2 เท่ากับ $\sum_{i=1}^{c} b_i \mu_i$ และถ้าหาก $\sum_{i=1}^{c} a_i b_i$ เท่ากับ 0 แล้ว เราจะเรียกว่า L_1 และ L_2 ต่างก็เป็น วิธีเปรียบเทียบออร์โธโกนัล ซึ่งกันและกัน เราสามารถคำนวณค่า L_1 และ L_2 ได้ดังสมการ (9-37) ซึ่งแสดงว่า SS_{col} ที่นี่ df เท่ากับ c–1 ถูกแบ่งแยกออกเป็น c–1 ส่วน, แต่ละส่วนเป็น SSL_1 และ SSL_2 โดยแต่ละส่วนมี df เท่ากับ 1 หรือแสดงว่า SS_{col} เท่ากับ $SSL_1 + SSL_2$

ในทางปฏิบัติถ้าหากเรากำหนด ค่าเปรียบเทียบ m คู่ต่างก็เป็นการเปรียบเทียบออร์โธโกนัล ซึ่งกันและกัน เขียนได้เป็น

$$L_m = \sum_{i=1}^{c} a_{mi} T_i$$

โดย L_m คือ คู่การเปรียบเทียบที่ m (คู่การหาความแตกต่างที่ m) สำหรับ a_{mi} คือสัมประสิทธิ์ที่ให้กับ T_i และ $\sum_{i=1}^{c} a_{mi}$ เท่ากับ 0 กล่าวโดยสรุปคือเมื่อ

คู่การเปรียบเทียบที่ 1 : L_1 ; $a_{11} + a_{12} + \ldots\ldots + a_{1c} = 0$

คู่การเปรียบเทียบที่ 2 : L_2 ; $a_{21} + a_{22} + \ldots\ldots + a_{2c} = 0$

คู่การเปรียบเทียบที่ 3 : L_3 ; $a_{31} + a_{32} + ... + a_{3c} = 0$

: :

: :

: :

และ คู่การเปรียบเทียบที่ m : L_m ; $a_{m1} + a_{m2} + ... + a_{mc} = 0$

แสดงให้เห็นว่า SS_{col} ที่มี df เท่ากับ $c-1$ ถูกแบ่งแยกออกเป็น $c-1$ ส่วน, แต่ละส่วนเป็น SSL_1, SSL_2,........ และ SSL_m (m จะต้องไม่เกิน $c-1$ ด้วย) โดยแต่ละส่วนมี df เท่ากับ 1 หรือแสดงว่า SSL เท่ากับ $SSL_1 + SSL_2 + ... + SSL_m$ นั่นเอง

การดำเนินการหาความแตกต่างค่าเฉลี่ยประชากรโดยวิธีการเปรียบเทียบออร์โธโกนัล อาจจะดำเนินการไปพร้อมกับการวิเคราะห์ความแปรปรวนหรืออาจดำเนินการตามลำดับขั้นดังนี้

1) ขั้นที่ 1 กำหนดการเปรียบเทียบ (ปกติในการทดลองจะกำหนดล่วงหน้าไว้ก่อนแล้ว) ผลรวมน้ำหนักสุกรขุนฟาร์มที่ 1, 2 และ 3 ในตาราง 9.1 คือ T_1 เท่ากับ 245, T_2 เท่ากับ 255 และ T_3 เท่ากับ 255 สมมุติว่าเรากำหนดคู่การเปรียบเทียบเป็น

L_1 : T_1 กับ T_2 , T_3

L_2 : T_2 กับ T_3

2) ขั้นที่ 2 ให้สัมประสิทธิ์ T_i โดยถือหลัก

$$\sum_{i=1}^{c} a_{1i} = \sum_{i=1}^{c} a_{2i} = \ldots = \sum_{i=1}^{c} a_{mi} = 0$$

ขอแนะนำหลักการให้สัมประสิทธิ์กับ T_i ที่ง่าย ๆ และใช้ในทางปฏิบัติ คือ นับจำนวนสมาชิกของแต่ละคู่การเปรียบแล้วหา ค.ร.น. จะได้อัตราส่วนต่ำสุดสำหรับนำมาให้ค่าสัมประสิทธิ์กับ T_i เช่น สมมุติว่า

L_1 : T_1 , T_2 , T_3 กับ T_5 , T_6 , T_7 , T_8

มีจำนวนสมาชิก 3 ตัวกับ 4 ตัว ได้ค่าคูณร่วมน้อยเท่ากับ 12 จึงให้สัมประสิทธิ์กับ T_1 , T_2 , T_3 เป็น 4 (+ หรือ –) และให้สัมประสิทธิ์กับ T_5 , T_6 , T_7 , T_8 เป็น 3 (–หรือ +) เป็นต้น สำหรับข้อมูลจากตารางที่ 9.1 ให้สัมประสิทธิ์ T_i ตามตารางที่ 9.8 สังเกต (-2) + (1) + (1) เท่ากับ 0, (0) + (-1) + (1) เท่ากับ 0 และ (-2)(0) + (1)(-1) + (1)(1) เท่ากับ 0

ตารางที่ 9.8 ค่าสัมประสิทธิ์ T_i

T_i	L_i	
	L_1	L_2
T_1	-2	0
T_2	+1	-1
T_3	+1	+1
ความหมายของการเปรียบเทียบ	T_1 กับ T_2 , T_3	T_2 กับ T_3

3) ขั้นที่ 3 คำนวณค่า SSL ตามสมการ (9-37) ข้อมูลจากตารางที่ 9.1 คำนวณได้ค่า SSL_1 และ SSL_2 ดังนี้

$$SSL_1 = \frac{[(-2)(245)+(1)(280)+(1)(255)]^2}{5[(-2)^2+(1)^2+(1)^2]}$$

$$= \frac{[45]^2}{5(6)} = \frac{2,025}{30} = 67.5$$

$$SSL_2 = \frac{[(0)(245)+(-1)(280)+(1)(255)]^2}{5[(0)^2+(-1)^2+(1)^2]}$$

$$= \frac{[25]^2}{5(2)} = \frac{625}{10} = 62.5$$

4) ขั้นที่ 4 นำค่าที่คำนวณได้จากขั้นที่ 3 บรรจุลงในตารางการวิเคราะห์ความแปรปรวน ดังตารางที่ 9.9

5) ขั้นที่ 5 สรุปผลความแตกต่างกันของค่าเฉลี่ยประชากร

พบว่าค่าเฉลี่ยน้ำหนักสุกรขุนฟาร์มที่ 1 ฟาร์มที่ 2 ฟาร์มที่ 3 และค่าเฉลี่ยน้ำหนักสุกรขุนฟาร์มที่ 2 กับฟาร์มที่ 3 แตกต่างกันอย่างมีนัยสำคัญ

ตารางที่ 9.9 การวิเคราะห์ความแปรปรวนของน้ำหนักสุกรจากฟาร์ม 3 ฟาร์ม

SV.	*df*	*SS.*	*MS.*	F_C
สดมภ์	2	130	65	8.3**
L_1	1	67.7	67.5	8.62*
L_1	1	62.5	62.6	7.98
ค่าคลาดเคลื่อน	12	94	7.83	
ผลรวม	14	224		

ที่ df =2,12; $F_{.01}$ = 6.93, $F_{.05}$=3.89

ที่ df =1,12; $F_{.01}$ = 9.33, $F_{.05}$=4.75

9.7 การทดสอบข้อสมมุติ

การตรวจรูปแบบการแจกแจงของประชากรว่า มีการแจกแจงปรกติหรือแจกแจงปรกติโดยประมาณ และอิสระต่อกันหรือไม่ตามข้อสมมุติ (assumption) การวิเคราะห์ความแปรปรวน อาจกระทำได้หลายทาง ทางหนึ่งที่กระทำได้คือการทดสอบการเท่ากันของความแปรปรวนประชากร ภายใต้สมมุติฐานเกี่ยวกับการเท่ากันของความแปรปรวนประชากรทั้งหลาย คือ

H_0 : $\sigma_1^2 = \sigma_2^2 = \ldots\ldots\ldots\ldots = \sigma_c^2$

H_1 : ความแปรปรวนไม่เท่ากันหมด

9.7.1 การทดสอบของบาร์ตเลตต์ ตัวอย่างที่ 1, 2, ……, c. มีขนาด $n_1, n_2, \ldots, n_c$ ตามลำดับ (ตัวอย่างขนาดไม่เท่ากันหมด) และ N เท่ากับ $n_1 + n_2 + \ldots\ldots\ldots\ldots + n_c$

ตัวสถิติทดสอบ คือ (Kanji, 1999, p. 62)

$$B = (2.3026)\frac{p}{q} \qquad \ldots(9\text{-}39)$$

โดยที่

$$p = (N\text{–}c)\ log\ S_P^2 - \sum_{i=1}^{c}(n_i - 1)logS_i^2$$

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

$$S_P^2 = \frac{\sum_{i=1}^{c}(n_i - 1)S_i^2}{N-c}$$; คล้ายกับสมการ (7-35) ในบทที่ 7

$S_P^2 = MS_{res} = Mse$; ตามสมการ (9-19)

เขตปฏิเสธ H_0 คือ $B > X^2_\alpha$; โดยอ่านค่า X^2_α ได้จากตารางภาคผนวกที่ 11 ค่าวิกฤตของการแจกแจงไคกำลังสอง ที่ $df = c-1$

ตัวอย่างที่ 9.2 ถ้ามีตัวอย่างจำนวน 3 ตัวอย่างขนาด $n_1 = 5$, $n_2 = 6$ และ $n_3 = 6$ จาก 3 ประชากร คำนวณค่า $S_1^2 = 0.0171$, $S_2^2 = 0.1898$ และ $S_3^2 = 0.1510$ จงทดสอบการเท่ากันของความแปรปรวนประชากรทั้ง 3 นี้ ที่ระดับ α=.05

วิธีทำ $c = 3$; $N = 5 + 6 + 6 = 17$

$$S_P^2 = \frac{(5-1)(0.0171) + (6-1)(0.1898) + (6-1)(0.1510)}{17-3}$$

$$= \frac{0.0684) + 0.9490 + 0.7550)}{17-3} = 0.1266$$

$$p = (17-3)log(0.1266) - (4log(0.0171) + 5log(0.1898) + 5log(0.1510)$$

$$= 14(-0.8966) - (4-1.767) + 5(-0.7235) + 5(-0.8239))$$

$$p = -12.5524 + 14.805 = 2.2526$$

(อ่านค่าลอการิทึมสามัญในตารางภาคผนวกที่ 2 ภาคผนวกอย่างประมาณ ถ้าคำนวณด้วยเครื่องคอมพิวเตอร์หรืออื่น ๆ อาจได้ค่าแตกต่างจากนี้เล็กน้อย)

$$q = 1 + \frac{1}{3(3-1)}[(\frac{1}{(5-1)} + \frac{1}{(6-1)} + \frac{1}{(6-1)} - \frac{1}{(17-3)}]$$

$$= 1 + \frac{1}{6}[(\frac{13}{20} - \frac{1}{14}] = 1 + 0.096 = 1.096$$

ได้ค่า $B = (2.3026)\dfrac{2.2526}{1.0960}$

$= 4.7325$

ที่ $df = c\text{-}1 = 2$, ค่า $X^2_{.05} = 5.991$ จึงไม่สามารถปฏิเสธ H_0 ที่ระดับ $\alpha = .05$ จึงยอมรับว่าความแปรปรวนประชากรเท่ากันหมด

9.7.2 การทดสอบของคอกครัน ตัวอย่างที่ 1, 2, ..., c มีขนาด n เท่ากันหมด

ตัวสถิติทดสอบ คือ (Kanji, 1999, p. 66)

$$G = \frac{(Largest)S_i^2}{\sum_{i=1}^{c} S_i^2} \qquad \text{...(9-40)}$$

เขตปฏิเสธ H_0 คือ $G > g_{(n,c)}$ และอ่านค่า $g_{(n,c)}$ จากตารางภาคผนวกที่ 12 หรือ ตารางภาคผนวกที่ 13 ในที่นี้ c คือ k ของตาราง ค่าวิกฤตของการทดสอบของคอกครัน (critical values for cochran's test)

ตัวอย่างที่ 9.3 จงทดสอบการเท่ากันของความแปรปรวนประชากร ตามข้อมูลตารางที่ 9 .1 ที่ระดับ $\alpha = .05$ ด้วยวิธีการทดสอบของคอกครัน

วิธีทำ

พบว่าขนาดตัวอย่าง $n_1 = n_2 = n_3 = 5$

ความแปรปรวนตัวอย่าง $S_1^2 = 7.5$, $S_2^2 = 12.5$, $S_3^2 = 3.5$

$(largest)S^2_1 = 12.5$

$\sum_{i=1}^{c} S_i^2 = 7.5 + 12.5 + 3.5$

$= 23.5$

$$G = \frac{12.5}{23.5} = 0.5319$$

ค่า $g_{(5,3)}$ จากตารางภาคผนวกที่ 13 เท่ากับ 0.7457 ไม่สามารถปฏิเสธ H_0 ที่ระดับ $\alpha = .05$ จึงยอมรับว่าความแปรปรวนประชากรเท่ากันหมด

9.7.3 การทดสอบของฮาร์ตลีย์ ตัวอย่างที่ 1,2, ..., *c*. ขนาด *n* เท่ากันหมด ตัวสถิติทดสอบคือ (Kanji, 1999, p. 64)

$$H = \frac{(Largest)S_i^2}{(Smallest)S_i^2} \quad \text{...(9-41)}$$

เขตปฏิเสธ H_o คือ $H > h_{(c,n)}$ โดยอ่านค่า $h_{(c,n)}$ จากตารางภาคผนวกที่ 14 หรือ ตารางภาคผนวกที่ 15 เมื่อ *c* คือ *r* ของตารางค่าวิกฤตของการทดสอบของฮาร์ตลีย์ (critical values for Hartley's test) สำหรับกรณีกำหนดระดับ α =.05 แสดงว่า 1-α = .95 ใช้ตารางภาคผนวกที่ 14 ถ้าระดับ α = .01 แสดงว่า 1-α = .99 ใช้ตารางภาคผนวกที่ 15

ตัวอย่างที่ 9.4 จงทดสอบการเท่ากันของความแปรปรวนประชากร ตามข้อมูลตารางที่ 9.1 ที่ระดับ α = .05 ด้วยค่าวิกฤตของการทดสอบของฮาร์ตลีย์

วิธีทำ พบว่าขนาดตัวอย่าง $n_1 = n_2 = n_3 = 5$

$$Largest\ S_i^2 = 12.5$$

$$Smallest\ S_i^2 = 3.5$$

$$H = \frac{12.5}{3.5} = 3.5714$$

กรณีต้องการระดับ α = .05 ใช้ตารางภาคผนวกที่ 14 อ่านได้ว่า $h_{(3,5)} = 15.5$ ไม่สามารถปฏิเสธ H_o ที่ระดับ α = .05 จึงยอมรับว่าความแปรปรวนประชากรเท่ากันหมด

จากตัวอย่างทั้งตัวอย่างที่ 9.1 ตัวอย่างที่ 9.3 และตัวอย่างที่ 9.4 เป็นกรณี *n* เท่ากันหมดการทดสอบของคอกครันกับการทดสอบของฮาร์ตลีย์ มักจะได้ผลเหมือน ๆ กัน แต่การทดสอบของคอกครันจะใช้ได้ดีกว่าเพราะใช้ข้อเท็จจริงจากข้อมูลมากกว่า

9.8 การวิเคราะห์ความแปรปรวนสองปัจจัย

จากตารางที่ 9.1 ปัจจัยที่มีอิทธิพลของการเติบโตของสุกรขุนคือฟาร์มทั้งสามฟาร์ม ถ้าแต่ละฟาร์มมีสุกรพันธุ์ A, B, C พันธุ์สุกรก็จะเป็นปัจจัยที่สองที่มีอิทธิพลต่อการเจริญเติบโตของสุกรขุน ซึ่งในการวิเคราะห์ความแปรปรวนจะเรียกปัจจัยที่สองนี้ว่า บล็อก (block, *blk*) มีที่มาจากการออกแบบการทดลอง (experimental design) เพื่อวัดอิทธิพลของปัจจัยที่กำหนดขึ้นต่อคน สัตว์ สิ่งของหรือพืช สำหรับปัญหาในทางธุรกิจที่พิจารณาของปัจจัย เช่น รายได้ของพนักงานต่างบริษัทและต่างลักษณะงาน รายรับของบริษัทพื้นที่เดียวกันและต่างพื้นที่ทำงาน เป็นต้น ดังนั้นตารางข้อมูล โครงสร้างข้อมูล X_{ij} ใช้ได้ทำนองเดียวกันกับตารางที่ 9.3 และตารางวิเคราะห์ความแปรปรวน โดยกำหนดได้ดังนี้

9.8.1 โครงสร้างข้อมูล X_{ij} เมื่อ i คือสดมภ์ที่ 1,2,3…,c และ j คือบล๊อกที่ 1,2,3…,b ดังตารางที่ 9.10

ตารางที่ 9.10 โครงสร้างข้อมูล สดมภ์ที่ 1,2,3…,c และ คือบล๊อกที่ 1,2,3…,b

สดมภ์ที่ / บล็อกที่	1	2	…	c	B_j	$\bar{X}_{.j}$
1	X_{11}	X_{12}	…	X_{c1}	B_1	$\bar{X}_{.1}$
2	X_{12}	X_{22}	…	X_{c2}	B_2	$\bar{X}_{.2}$
3	X_{13}	X_{23}	…	X_{c3}	B_3	$\bar{X}.3$
.	.	.	…	.	.	.
.	.	.	...	.	.	.
.	.	.	…	.	.	.
b	X_{1b}	X_{2b}	…	X_{cb}	B_c	$\bar{X}.b$
ผลรวม : T_i	T_1	T_2	…	T_c	GT	xxxxx
ค่าเฉลี่ย : $\bar{X}_i$	$\bar{X}_1$	$\bar{X}_2$	…	$\bar{X}_c$	xxxxx	$\bar{\bar{X}}$ =GM

เมื่อ $T_i = \sum_{j=1}^{b} X_{ij}$; ผลรวมของสดมภ์ที่ i

$B_j = \sum_{i=1}^{c} X_{ij}$; ผลรวมของบล๊อกที่ j

$GT = \sum_{i=1}^{c} \sum_{j=1}^{b} X_{ij}$; ผลรวมรวมทั้งหมด

$\bar{\bar{X}} = \frac{GT}{bc}$; ค่าเฉลี่ยทั้งหมด

$N = bc$; จำนวนค่าสังเกตทั้งหมด

9.8.2 แหล่งความแปรปรวน แยกออกเป็น 3 ส่วนคือ

1) แหล่งความแปรปรวนระหว่างสดมภ์ มีสัญลักษณ์เป็น *col*

2) แหล่งความแปรปรวนระหว่างบล๊อก มีสัญลักษณ์เป็น *blk*

3) เศษตกค้าง มีสัญลักษณ์เป็น *res*

9.8.3 ผลบวกกำลังสองของการเบี่ยงเบน (*SS*) มี 4 ส่วน คือ

1) SS_{Tot} หมายถึงผลบวกกำลังสองของการเบี่ยงเบนของค่าสังเกตทั้งหมดหาได้ทำนองเดียวกับสมการ (9-21) และ (9–25) ดังนั้น

$$SS_{Tot} = \sum_{i=1}^{c} \sum_{j=1}^{b} (X_{ij} - \bar{\bar{X}})^2 - CF$$

$$= \sum (Each - Value)^2 - CF$$

และมีค่าตัวประกอบแก้ดังสมการ

$$CF = \frac{GT^2}{bc} \qquad \text{…(9-42)}$$

2) SS_{col} หมายถึงผลบวกกำลังสองของการเบี่ยงเบนของสดมภ์ที่มีจำนวนบล๊อกเท่ากับ b โดยที่

$$SS_{sol} = b\sum_{i=1}^{c} (X_i - \bar{\bar{X}})^2$$

ทำนองเดียวกับสมการ (9-22) เมื่อให้ $n_i = b$ จะได้สมการ (9–43)

$$SS_{sol} = \frac{1}{b}\sum_{i=1}^{c} T_i^2 - CF \qquad \text{...(9-43)}$$

3) SS_{blk} หมายถึงผลบวกกำลังสองของการเบี่ยงเบนของบล๊อกตามจำนวน c สดมภ์ หาได้ทำนองเดียวกับสมการ (9 - 21) และ (9 – 25) ดังนั้น

$$SS_{blk} = c\sum_{j=1}^{b}(X_j - \overline{\overline{X}})^2$$

ทำนองเดียวกับสมการ (9-22) เมื่อให้ $n_i = c$ และ $\overline{X}_{.j} = \frac{B_j}{b}$ จะได้สมการ (9–44)

$$SS_{sol} = \frac{1}{c}\sum_{j=1}^{b} B_i^2 - CF \qquad \text{...(9-44)}$$

4) SS_{res} หมายถึงผลบวกกำลังสองของการเบี่ยงเบนของเศษตกค้าง

$$SS_{res} = SS_{Tot} - SS_{col} - SS_{blk} \qquad \text{...(9-45)}$$

9.8.4 ตารางวิเคาระห์ความแปรปรวนที่ใช้ในทางปฏิบัติ จะใช้ตามตารางที่ 9.11

ตารางที่ 9.11 ตารางวิเคราะห์ความแปรปรวนของข้อมูลที่มี c สดมภ์และแต่ละสดมภ์มี b บล๊อก

SV	df	SS	MS	F
ระหว่างสดมภ์	c-1	SS_{col}	$\frac{SS_{Tot}}{c-1} = MS_{col}$	$\frac{MS_{Tot}}{MS_{res}}$
ระหว่างบล๊อก	b-1	SS_{blk}	$\frac{SS_{blk}}{b-1} = MS_{blk}$	$\frac{MS_{blk}}{MS_{res}}$
เศษตกค้าง	(c-1)(b-1)	SS_{res}	$\frac{SS_{res}}{(c-1)(b-1)} = MS_{res}$	
รวม	bc - 1	SS_{Tot}		

เมื่อ $Tot\ df = bc - 1 = N - 1$: รวม df

$col\ df = c - 1$: สดมภ์ df

$blk\ df = b - 1$: บล๊อก df

$res\ df \quad = \quad (c-1)(b-1) \qquad\qquad :$ เศษตกค้าง df

ตัวอย่างที่ 9.5 ฟาร์มแห่งหนึ่งให้อาหารสุกร 4 ชนิดคือ ชนิด ก ข ค และ ง อยากทราบว่าอาหารที่ให้สุกรขุนชนิดใดที่มีอิทธิพลต่อการเจริญเติบโตของสุกรที่มีอายุ 4 เดือน ซึ่งเลือกมาจาก 3 ท้องที่และแต่ละท้องที่นำสุกรมาท้องที่ละ 4 ตัว และบันทึกข้อมูลได้ดังตารางที่ 10. 12

ตารางที่ 9.12 ข้อมูลการเลี้ยงสุกรขุนอายุ 4 เดือนด้วยอาหารชนิด ก ข ค และ ง

อาหาร / ท้องที่	ก	ข	ค	ง	รวม	ค่าเฉลี่ย
1	44.2	40.8	50.0	34.3	169.3	42.33
2	51.5	39.5	52.4	41.7	185.1	46.28
3	53.4	47.8	55.0	52.6	208.8	52.20
รวม	149.1	128.1	157.4	128.6	563.2	xxxxx
ค่าเฉลี่ย	49.7	42.7	52.47	42.87	xxxxx	46.94

คำนวนหาค่าสำหรับตารางวิเคราะห์ความแปรปรวนที่ $N = 12$, $c = 4$, $b = 3$ จะได้

1) df $\qquad Tot\ df = 12-1 = 11$

$col\ df = 4-1 = 3$

$blk\ df = 3-1 = 2$

$res\ df = (4-1)(3-1) = 6$

2) CF $\qquad CF = \dfrac{(563.3)^2}{12} = 26{,}432.85$

3) $\qquad SS_{Tot} = \left[(44.2)^2+(51.5)^2+\ldots+(41.7)^2+(52.6)^2\right] - CF$

$= 26{,}920.08 - 26{,}432.85$

$$SS_{Tot} = 487.23$$

$$SS_{col} = \frac{1}{3}\left[(149.1)^2+(128.1)^2+(157.4)^2+(128.6)^2\right] - CF$$

$$= 26{,}651.05 - 26{,}432.85$$

$$= 218.19$$

$$SS_{blk} = \frac{1}{4}\left[(169.3)^2+(185.1)^2+(208.8)^2\right] - CF$$

$$= 26{,}630.49 - 26{,}432.85$$

$$= 197.64$$

$$SS_{res} = 487.23 - 218.19 - 197.64$$

$$= 71.40$$

4) *MS* $MS_{col} = \frac{218.19}{3} = 72.73$

$$MS_{blk} = \frac{197.64}{3} = 98.82$$

$$MS_{\mathrm{resl}} = \frac{71.40}{3} = 11.90$$

5) *F- ratio* $F_{col} = \frac{72.71}{11.90} = 6.11$

$$\mathrm{F}_{\mathrm{blk}} = \frac{98.82}{11.90} = 8.30$$

6) อ่านค่า *F* จากตารางภาคผนวกที่ 7 และ ตารางภาคผนวกที่ 8 จะได้

$df = 3,6$ $F_{.01} = 9.78$, $F_{.05} = 9.78$

$df = 2,6$ $F_{.01} = 10.92$, $F_{.05} = 5.14$

7) *CV* $CV = \sqrt{\frac{71.4}{46.94}} \times 100$

$= 18.00\ \%$

จากข้อมูลดังกล่าวเขียนตารางวิเคราะห์ความแปรปรวนได้เป็นตารางที่ 9.13

ตารางที่ 9.13 ตารางวิเคราะห์ความแปรปรวนของข้อมูลผลการเจริญเติบโตของสุกรอายุ 4 เดือนด้วยอาหารชนิด ก ข ค และ ง

SV	*df*	*SS*	*MS*	*F-ratio*
ระหว่างอาหาร	3	218.19	72.73	F_{col} = 6.11*
ระหว่างท้องที่	2	197.64	98.82	F_{blk} =8.3*
เศษตกค้าง	6	71.40	11.90	
รวม	11	487.23		

สรุป ผลการวิเคราะห์ความแปรปรวนของการให้อาหารสุกรทั้ง 4 ชนิดมีอิทธิพลทำให้สุกรมีน้ำหนักต่างกันตามสภาพท้องที่ทั้ง 3 ที่มีผลทำให้น้ำหนักสุกรต่างกันด้วย

9.9 บทสรุป

ก่อนศึกษาบทนี้ควรทบทวนความรู้เกี่ยวกับการแจกแจงตัวอย่าง การประมาณค่าเกี่ยวกับประชากรและทดสอบสมมุติฐานเสียก่อน จุดเด่นของเนื้อหาอยู่ที่การสร้างตารางการวิเคราะห์ความแปรปรวน สิ่งที่ควรรู้และทำความเข้าใจเป็นพิเศษคือ

9.9.1 การแจกแจงเอฟ

9.9.2 ตารางวิเคราะห์ความแปรปรวน

1) ความแปรปรวนที่อธิบายได้

2) ค่าผิดพลาดมาตรฐานของค่าเฉลี่ยตัวอย่างในรูป $\sqrt{\frac{MSe}{n}}$

9.9.3 การตรวจพินิจความแตกต่างกันของค่าเฉลี่ยประชากร

1) ข้อสมมุติ

2) วิธีแอลเอสดี

3) วิธีดีเอ็มอาร์ที

4) วิธีเปรียบเทียบออร์โธโกนัล

9.9.4 การทดสอบข้อสมมุติ

1) การทดสอบของบาร์ตเลตต์

2) การทดสอบของฮาร์ต์ลีย์

3) การทดสอบของคอกครัน

9.9.5 ตารางวิเคราะห์ความแปรปรวนสองทาง

9.10 คำถามทบทวน

1. จงอธิบายความแตกต่างระหว่าง ความแปรผันระหว่างตัวอย่างกับความแปรผันภายในตัวอย่าง มาพอสังเขป
2. จงอ่านค่า F ที่กำหนดให้ต่อไปนี้โดยใช้ตารางภาคผนวกที่ 6 หรือตารางภาคผนวกที่ 7

ก. $F_{9,16,.05}$

ข. $F_{9,16,.01}$

ค. $F_{8,25,.05}$

ง. $F_{8,25,.01}$

จ. $F_{12,30,.05}$

ฉ. $F_{12,30,.01}$

3. จงอธิบายว่าเมื่อใดจะใส่เครื่องหมาย * และ ** ไว้ที่อัตราส่วนเอฟของตารางวิเคราะห์ความแปรปรวน
4. จงยกตัวอย่างการทดลองที่ต้องเปรียบเทียบทางธุรกิจ ในสถานการณ์ที่มีแหล่งความแปรปรวนซึ่งอธิบายได้มากกว่า 1 แหล่ง
5. จงทดสอบว่าตัวอย่างเหล่านี้มาจากประชากรเดียวกันหรือไม่ จากข้อมูลตัวอย่างจำนวน 4 ตัวอย่างขนาด n = 4 เท่ากันหมดเป็นดังตารางที่ 9.14

ตารางที่ 9.14 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของตัวอย่างสุ่ม 4 ตัวอย่าง

ตัวอย่างที่	1	2	3	4
ค่าเฉลี่ย	0.261	0.296	0.312	0.135
ค่าเบี่ยงเบนมาตรฐาน	0.21	0.17	0.19	0.08

6. จากข้อมูลในตารางที่ 9.15 ที่เป็นรายจ่ายต่อเดือน (ล้านบาท) ที่สุ่มเก็บจากสาขาย่อย 4 แห่ง

ก. จงทดสอบว่าความแปรปรวนของรายจ่ายต่อเดือนของสาขาย่อยเท่ากันหรือไม่

ข. รายจ่ายต่อเดือนของสาขาย่อยทั้ง 4 แตกต่างกันหรือไม่

ตารางที่ 9.15 รายจ่ายต่อเดือนของสาขาย่อย 4 แห่ง

รายจ่ายสาขาย่อย(ล้านบาท)			
ก	ข	ค	ง
0.6	1.1	2.1	0.5
0.8	1.3	2.0	0.9
0.3	1.5	1.7	1.0
0.5	1.8	1.6	0.7
0.6	0.9	1.0	0.7

7. อยากทราบว่าฮอร์โมน 4 ชนิดคือ a b c และ d มีผลทำให้น้ำหนักของไก่จะแตกต่างกันหรือไม่ เมื่อทดลองเลี้ยงไก่ 4 โรงเรือน โรงเรือนละ 4 ตัว ได้ผลการทดลองดังตารางที่ 16

ตารางที่ 9.16 ผลการทดลองใช้ฮอร์โมน 4 ชนิดเลี้ยงไก่ 4 โรงเรือน โรงเรือนละ 4 ตัว

ฮอร์โมน / โรงเรือน	a	b	c	d
1	0.71	0.95	0.40	0.69
2	0.44	0.68	0.64	0.71
3	0.86	0.85	0.51	0.42
4	0.44	0.41	0.74	0.62

บทที่ 10
เทคนิคเชิงปริมาณเบื้องต้น

การใช้ข้อมูลเชิงตัวเลขที่บ่งบอกถึงประมาณมากน้อยของสิ่งของ จำนวนคน ยอดขายหรืออื่น ๆ ค่าตัวเลขที่แตกต่างกันและไม่คงตัว ข้อมูลเหล่านี้คือตัวแปรที่จำเป็นต้องการทำความเข้าใจเบื้องต้นกับความสัมพันธ์ระหว่างข้อมูลสองชุดหรือมากกว่าสองชุด ซึ่งถือว่าเป็นงานวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอันได้แก่ การศึกษาความสัมพันธ์ระหว่างสองตัวเป็นและการศึกษาความสัมพันธ์ระหว่างตัวแปรหนึ่งกับอีกหลายตัวแปร

10.1 ความหมายของความสัมพันธ์ระหว่างตัวแปร

การกล่าวว่านักลงทุนชาวต่างประเทศที่เข้ามาในประเทศไทย เพศไม่มีความสัมพันธ์กับความนิยมในบริษัทที่จะลงทุน หมายความถึงทั้งเพศชายและเพศหญิงต่างก็มีความนิยมในบริษัทที่จะลงทุนแตกต่างกันหรือคล้าย ๆ กัน หากทราบเพศนักลงทุนไม่สามารถบอกให้ทราบถึงความนิยมในบริษัทที่จะลงทุนของเขาเหล่านั้น และหากทราบความนิยมในบริษัทที่จะลงทุนไม่สามารถบอกได้เพศของนักนักลงทุนเป็นหญิงหรือชาย ถ้าข้อความดังกล่าวจะเป็นจริง อาจกล่าวได้ว่าเพศ นักลงทุนกับความนิยมในบริษัทที่จะลงทุนเป็นอิสระต่อกันในเชิงสถิติหรือกล่าวอีกนัยหนึ่งเพศกับความนิยมในบริษัทที่จะลงทุนไม่มีความสัมพันธ์กัน

สถานการณ์ที่แสดงถึงความอิสระของตัวแปรสองกลุ่มเป็นอิสระต่อกัน คือการที่ทราบค่าของตัวแปรตัวกลุ่มหนึ่งไม่สามารถบอกให้ทราบถึงค่าของตัวแปรอีกกลุ่มหนึ่ง เช่น กรณีตัวท่านสนใจท่านทอดเหรียญบาทอันหนึ่งเพื่อดูผลหัวและก้อยในขณะที่เพื่อนของท่านทอดลูกเต๋าเพื่อดูผลแต้ม 1 2 3 4 5 หรือ 6 การทอดเหรียญออกหัวหรือก้อยไม่มีผลต่อแต้มที่ได้จากการทอดลูกเต๋า ดังนั้น เหตุการณ์ทั้งสองเหตุการณ์จึงไม่มีความสัมพันธ์กันอย่างสิ้นเชิงหรือเป็นอิสระต่อกันนั่นเอง

สถานการณ์ของตัวแปรสองตัวที่มีความสัมพันธ์กัน เช่น จำนวนชั่วโมงที่นักเรียนใช้ในการดูหนังสือมีความสัมพันธ์กับคะแนนที่สอบได้ ลักษณะนิสัยของการกินมีความสัมพันธ์กับน้ำหนัก อายุของพนักงานขายกับยอดขายที่แต่ละคนทำได้ เป็นต้น คำว่ามีความสัมพันธ์กันนี้หมายความว่า ตัวแปรทั้งสองมีบางสิ่งบางอย่างร่วมกันที่สามารถกล่าวเชื่อมโยงถึงกัน

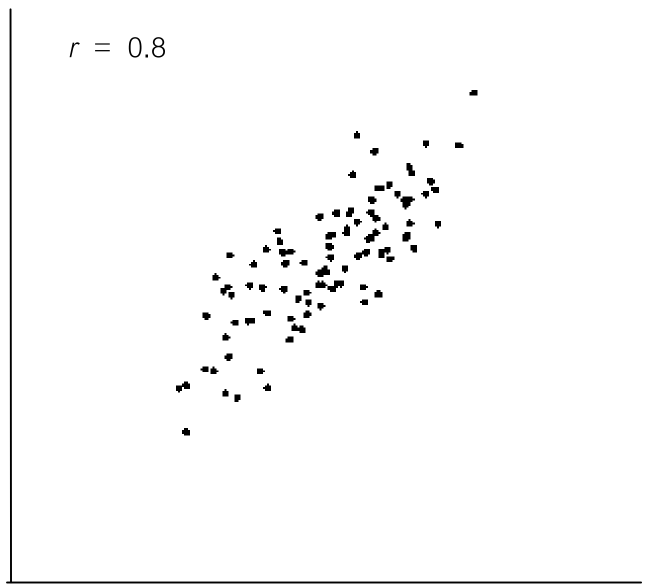
ได้ เช่น ในกรณีแรก อาจกล่าวได้ว่าจำนวนชั่วโมงที่นักเรียนใช้ในการดูหนังสือยิ่งมากเท่าใดก็ยิ่งทำให้คะแนนที่นักเรียนคนนั้นจะได้จากการสอบมากขึ้นเท่านั้น ส่วนในกรณีที่สองก็อาจกล่าวได้ว่าถ้าคนใดชอบกินอาหารทอดน้ำมันและชอบทานอาหารว่างบ่อยเพียงใด ก็จะทำให้คนนั้นมีน้ำหนักมากขึ้นเพียงนั้น กรณีสุดท้ายที่กล่าวไว้พนักงานขายอายุมากกับยอดขายที่เขาทำได้เป็นอย่างไร การทราบความสัมพันธ์ระหว่างตัวแปรนี้ ทำให้สามารถบอกค่าของตัวแปรอีกตัวหนึ่งได้ เช่น ถ้าท่านได้เกรด A ในการสอบวิชาสถิติ ก็อาจบอกได้ว่าท่านดูหนังสือมาก การพยากรณ์เช่นนี้อาจผิดก็ได้ถ้าท่านเป็นผู้ที่สามารถเข้าใจสถิติได้ดีอยู่แล้วโดยไม่ต้องใช้ความพยายามมากนัก แต่โดยเฉลี่ยแล้ว สามารถคาดหวังได้ว่านักเรียนที่ดูหนังสือมาก น่าจะทำคะแนนได้ดีและนักเรียนที่ทำคะแนนได้ดีคือนักเรียนที่ดูหนังสือมาก

อย่างไรก็ดีเมื่อกล่าวว่าตัวแปรสองตัวมีความสัมพันธ์กัน ไม่ได้หมายความว่าตัวแปรตัวหนึ่งเป็นสาเหตุของอีกตัวแปรหนึ่งชัดเจนไม่เปลี่ยนแปลงเหมือนหนึ่งรวมกับสองต้องเท่ากับสามเสมอแต่เป็นเพียงการพยากรณ์ล่วงหน้า หากทราบค่าของตัวแปรตัวหนึ่งมีแนวโน้วที่จะทราบค่าของตัวแปรอีกตัวหนึ่ง เช่น พบว่ารายได้ต่อเดือนของชายโสดผู้หนึ่งมีความสัมพันธ์กับค่าใช้จ่ายต่อเดือนของเขา ไม่ได้หมายความว่าเมื่อเขามีรายได้เพิ่มขึ้นมีผลทำให้ค่าใช่จ่ายของเขาสูงขึ้น ทุก ๆ เดือนและเมื่อเขามีรายได้ลดลงเป็นผลให้ค่าใช้จ่ายของเขาลดลงทุกเดือนไปด้วย แต่มีความหมายว่าการที่รายได้และค่าใช้จ่ายสูงขึ้นอาจเป็นผลมาจากการที่ค่าครองชีพสูงขึ้น ทำให้มีผลกระทบต่อทั้งรายได้และค่าใช้จ่าย

การศึกษาหาความสัมพันธ์ระหว่างตัวแปรสองตัวว่ามีความสัมพันธ์กันหรือไม่มักถามกันอยู่เป็นประจำ เช่น การเลือกตั้งสมาชิกสภาผู้แทนราษฎร คำถามที่น่าสนใจก็คือระดับการศึกษาสัมพันธ์กับพฤติกรรมการเลือกตั้งของผู้ลงคะแนนเสียงหรือไม่ ในทางจิตวิทยาอาจมีการศึกษาว่าสถานภาพสมรสสัมพันธ์กับความสุขหรือไม่ ระยะเวลานอนของคนสัมพันธ์กับน้ำหนักตัวหรือไม่ หรือในทางการตลาดอาจสนใจศึกษาว่าความสัมพันธ์ยอดขายสัมพันธ์กับประสบการณ์การขายของพนักงานขายหรือไม่ ยอดขายสัมพันธ์กับความบ่อยของการโฆษณาทางวิทยุหรือไม่ ความสนใจต่อตัวแปรทั้งสองไม่ได้มุ่งลงเพียงการศึกษาว่ามีความสัมพันธ์กันหรือไม่ แต่ความสนใจนั้นได้เลยไปถึงการศึกษาว่าตัวแปรนั้นมีความสัมพันธ์กันมากน้อยเพียงใด และมีลักษณะของความสัมพันธ์เป็นอย่างไรจึงมีการการกำหนดค่าความสัมพันธ์ระหว่างตัวแปรเรียกว่า ค่าสัมประสิทธิ์สหสัมพันธ์ (coefficient of correlation) สัญลักษณ์ r สำหรับเบื้องต้นนี้จะศึกษาความสัมพันธ์ระหว่างตัวแปรเป็นความสัมพันธ์เชิงเส้น (linear correlation) ค่า r จะมีค่าระหว่าง -1 ถึง 1 และหากค่า r

เป็น 0 แสดงว่าตัวแปรคู่นั้นไม่มีความสัมพันธ์เชิงเส้นกันแสดงให้เห็นด้วยแผนภูมิแสดงความสัมพันธ์ตัวแปรที่อยู่ในแนวตั้งและแนวนอน ดังภาพที่ 10.1 ถึงภาพที่ 10.5

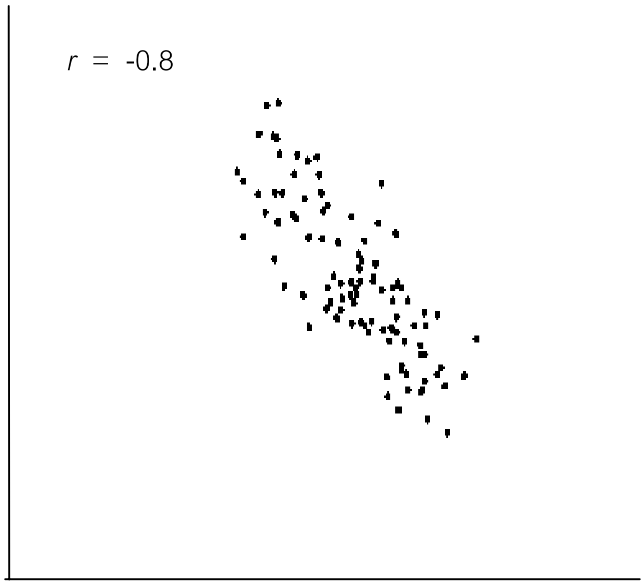
10.1.1 ความสัมพันธ์ในทางบวก เป็นความสัมพันธ์ระหว่างสองตัวแปรที่มีค่าขึ้นลงไปทางเดียวกัน กล่าวคือถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้น ตัวแปรอีกตัวหนึ่งจะมีค่าเพิ่มขึ้นด้วยดังภาพที่ 10.1 ค่า *r* เท่ากับ 0.8



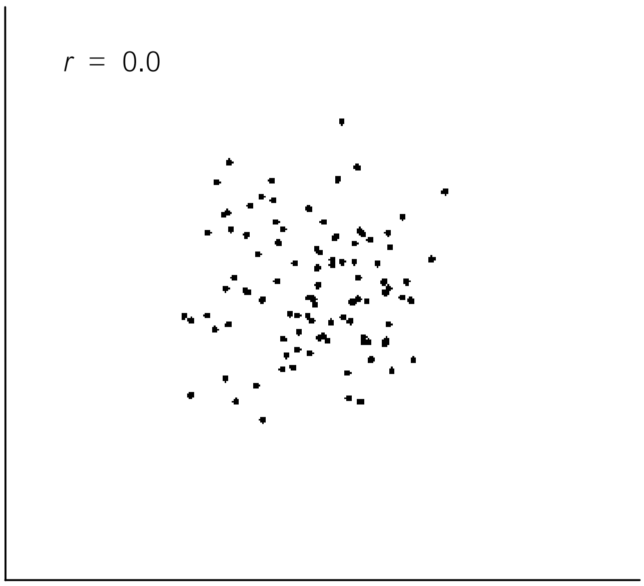
ภาพที่ 10.1 ความสัมพันธ์ระหว่างสองตัวแปรที่ค่า *r* = 0.8
ที่มา : (ดัดแปลงมาจาก Puranen, 2002)

10.1.2 ความสัมพันธ์ในทางลบ เป็นความสัมพันธ์ระหว่างสองตัวแปรที่มีค่าขึ้นลงไปสวนทางกัน กล่าวคือถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้น แต่ตัวแปรอีกตัวหนึ่งจะมีค่าลงดังภาพที่ 10.2 ค่า *r* เท่ากับ -0.8

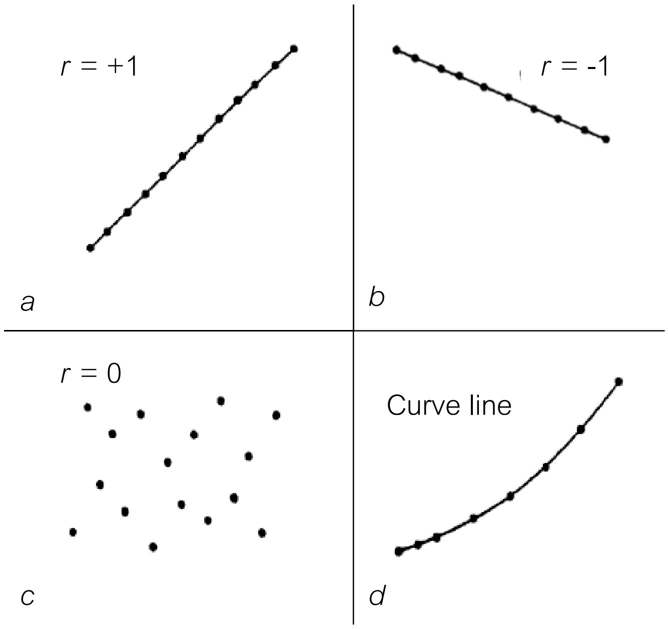
10.1.3 ความสัมพันธ์ไม่เป็นเชิงเส้น เป็นความสัมพันธ์ระหว่างสองตัวแปรที่มีค่าขึ้นลงเมื่อแสดงด้วยแผนภูมิแล้วไม่สามารถบอกทิศทางความสัมพันธ์ให้ชัดเจนได้ ภาพที่ 10.3 ภาพที่ 10.4 และ ภาพที่ 10.5 ค่า *r* เท่ากับ 0



ภาพที่ 10.2 ความสัมพันธ์ระหว่างสองตัวแปรที่ค่า r = -0.8
ที่มา : (ดัดแปลงมาจาก Puranen, 2002)



ภาพที่ 10.3 ความสัมพันธ์ระหว่างสองตัวแปรที่ค่า r = 0.0
ที่มา : (ดัดแปลงมาจาก Puranen, 2002)



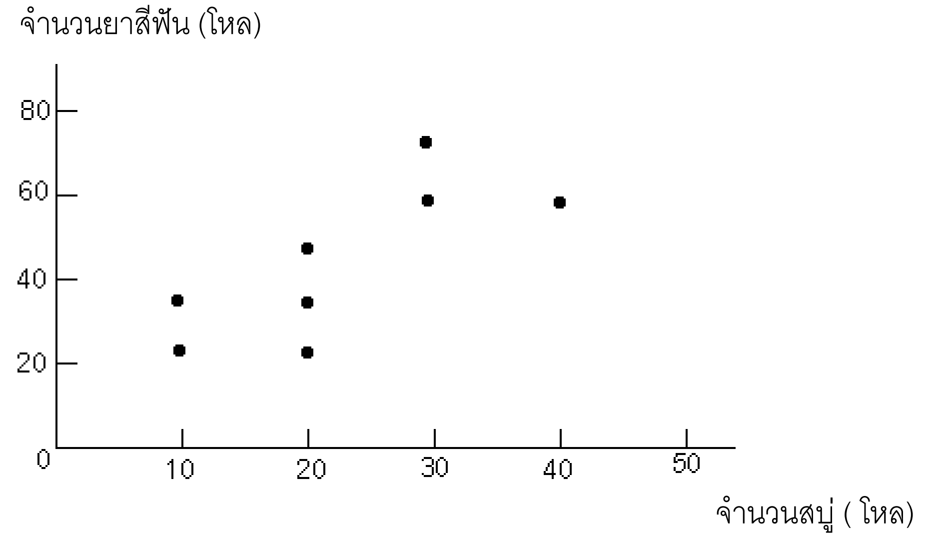
ภาพที่ 10.4 ความสัมพันธ์ระหว่างสองตัวแปรที่ค่า *r* แตกต่างกัน
ที่มา : (ดัดแปลงมาจาก Campbell, 2002)

ตัวอย่างที่ 10.1 จงสร้างแผนภูมิแสดงความสัมพันธ์ระหว่างจำนวนยาสีฟันและสบู่ที่พนักงานขาย 10 คนจำหน่ายได้ในรอบสัปดาห์นี้

ตารางที่ 10.1 จำนวนสบู่และยาสีฟันที่พนักงานขายจำหน่ายได้

ชื่อพนักงาน	จำนวนสบู่(โหล)	จำนวนยาสีฟัน(โหล)
นาย ก	20	30
นางสาว ถ	40	60
นาย ท	20	40
นาย อ	30	60
นาย ธ	10	30
นาย จ	10	40
นาย ฟ	10	40
นาย ว	20	50
นาย ม	20	30
นาง บ	30	70

วิธีทำ สร้างแผนภูมิโดยกำหนดแกนนอนแทนจำนวนสบู่และแกนตั้งแทนจำนวนยาสีฟัน



ภาพที่ 10.5 แผนภูมิความสัมพันธ์ระหว่างจำนวนสบู่และจำนวนยาสีฟัน

แสดงจำนวนยาสีฟันและสบู่ที่พนักงานขาย 10 คนจำหน่ายได้ไม่สัมพันธ์เชิงเส้นกัน

10.2 การวัดค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสองตัวแปร

การพิจารณาความสัมพันธ์ระหว่างสองตัวแปรเป็นเชิงเส้น ที่สะดวกและง่ายต่อการทำความเข้าใจ คือ การพิจารณาตัวแปร X , Y ที่เป็นตัวแปรชนิดคู่ (paired variable) แล้วคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็น +1 หมายถึง มีความสัมพันธ์เชิงเส้นตรงไปในทิศทางบวกอย่างสมบูรณ์ ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็น -1 หมายถึง มีความสัมพันธ์เชิงเส้นตรงไปในทิศทางลบอย่างสมบูรณ์ ค่าสัมประสิทธิ์สหสัมพันธ์นี้มีค่าที่เป็นไปได้อยู่ระหว่าง -1 ถึง +1 สำหรับกรณีที่ค่าสัมประสิทธิ์สหสัมพันธ์เป็น 0 ไม่ได้หมายคามว่า ไม่มีความสัมพันธ์ระหว่างตัวแปรทั้งสอง ค่าสัมประสิทธิ์สหสัมพันธ์นี้เพียงแต่บอกถึงความสัมพันธ์เชงเส้นตรงเท่านั้น ตัวแปรทั้งสองที่มีความสัมพันธ์เชิงเส้นตรงอาจมีความสัมพันธ์ในรูปแบบอื่นที่ไม่ใช่เส้นตรงก็ได้ หลักเกณฑ์การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์กรณีประชากรจะใช้สัญลักษณ์ ρ ส่วนกรณีตัวอย่างจะใช้สัญลักษณ์ r ในที่นี้ขอกล่าวเพียง 2 วิธีตามหัวข้อ 10.2.1-10.2.2 และเน้นการหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง

เพื่อนำไปใช้ประกอบการทดสอบสมมมุติฐานสหสัมพันธ์ระหว่างสองตัวแปรของประชากรต่อไป

10.2.1 วิธีหาค่าสัมประสิทธ์สหสัมพันธ์เพียร์สัน (Pearson coefficient of correlation) หลักการ คือ ถ้ากำหนด X และ Y เป็นตัวแปรของประชากรที่มีค่าเฉลี่ย μ_x และ μ_y ความแปรปรวน σ_x^2 และ σ_y^2 ตามลำดับ ความแปรปรวนร่วมระหว่างตัวแปรทั้งสองคือสมการ 10-1 (Keller & Warrack, 2000, pp. 657-658)

$$Cov\,(X,Y) = E\,[(X-\mu_x)(Y-\mu_y)] \qquad ...(10\text{-}1)$$

จากสมการ (10-1) หากค่า X เพิ่มขึ้น และ Y เพิ่มขึ้นด้วยและถ้าหาก X เพิ่มขึ้น ค่า Y กลับลดลง การหาค่า ρ จะอาศัยความสัมพันธ์ของ X กับ Y คำนวณค่าสัมประสิทธ์ สหสัมพันธ์ของสมการ (10-1) แล้วจะได้เป็นสมการ (10-2)

$$\rho = Cov\,(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

$$= \frac{E[(X-\sigma_x)(X-\sigma_y)]}{E[(X-\sigma_x)^2]E[(X-\sigma_y)^2]} \qquad ..(10\text{-}2)$$

เมื่อชักตัวอย่างสุ่มจะได้ตัวแปร (X , Y) n คู่ ได้เป็น (X_1 , Y_1) , (X_2 , Y_2) , . . . , (X_n , Y_n) คำนวณค่าเฉลี่ย X ได้ $\overline{X}$ และคำนวณค่าเฉลี่ย Y ได้ $\overline{Y}$ ค่าสหสัมพันธ์ระหว่าง X กับ Y ซึ่ง r นำไปสู่ การประมาณค่าสหสัมพันธ์ระหว่าง X กับ Y ของประชากร ρ โดยค่าสหสัมพันธ์ของตัวอย่างโดยอาศัยความแปรปรวนร่วมหาได้จากสมการ (10-3)

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X-\overline{X}\right)^2}$$

และ

$$S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y-\overline{Y}\right)^2}$$

ได้ค่าประมาณของ $Cov\,(X,Y)$ คือ

$$Cov\ (X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) \qquad ...(10\text{-}3)$$

นำไปคำนวณค่า r ได้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันตามสมการ (10-4)

$$r = \frac{\frac{1}{n-1}\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{S_x S_y}$$

$$= \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \qquad ...(10\text{-}4)$$

ใช้วิธีการทางพีชคณิตเข้าช่วยได้ผลสำหรับนำไปใช้ในทางปฏิบัติได้อีกแบบหนึ่ง (computing formular) จะได้ดังสมการ (10-5) หรือสมการ (10-6)

$$r = \frac{\sum_{i=1}^{n}X_iY_i - n\overline{X}\overline{Y}}{\sqrt{\left(\sum X^2 - n\overline{X}^2\right)\left(\sum Y2 - n\overline{Y}^2\ \right)}} \qquad ...(10\text{-}5)$$

$$r = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right]\ \left[N\sum Y^2 - \left(\sum Y\right)^2\right]}} \qquad ...(10\text{-}6)$$

ตัวอย่างที่ 10.2 จากตารางที่ 10.1 แสดงคะแนนสอบสถิติธุรกิจและคะแนนสอบวิชาการจัดการทั่วไปของนักศึกษาจำนวน 10 คน จงหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันคะแนนสอบสถิติธุรกิจและคะแนนสอบวิชาการจัดการทั่วไปนักศึกษาจำนวน 10 คน

ตารางที่ 10.2 คะแนนสอบวิชาสถิติธุรกิจและวิชาการจัดการทั่วไปของนักศึกษาจำนวน 10 คน

นักศึกษาคนที่	คะแนนสอบวิชาสถิติธุรกิจ	คะแนนสอบวิชาการจัดการทั่วไป
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

วิธีทำ จากตารางที่ 10.2 เขียนตารางใหม่เพื่อใช้ในการคำนวณหาค่า *r* ได้เป็นตารางที่ 10.3 แล้วใช้สูตรจากสมการ (10-6) ได้ดังนี้

$$r = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right]\left[N\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

$$= \frac{10(36,853) - (460)(760)}{\sqrt{\left[10(23,634) - (460)^2\right]\left[10(59,816) - (760)^2\right]}}$$

$$= \frac{18,940}{\sqrt{(24,740)(20,560)}}$$

$$= 0.84$$

ตารางที่ 10.3 การหาค่า $\sum X, \sum X^2, \sum Y, \sum Y^2$ และ $\sum XY$

นักศึกษาคนที่	คะแนนสอบวิชาสถิติธุรกิจ		คะแนนสอบวิชาการจัดการทั่วไป		XY
	(X)	X^2	(Y)	Y^2	
1	39	1,521	65	4,225	2,536
2	43	1,849	78	6,084	3,354
3	21	441	52	2,704	1,092
4	64	4,096	82	6,724	5,248
5	57	3,249	92	8,464	5,244
6	47	2,209	89	7,921	4,183
7	28	784	73	5,329	2,044
8	75	5,625	98	9,604	7,350
9	34	1,156	56	3,136	1,904
10	52	2,704	75	5,625	3,900
	$\sum X = 460$	$\sum X^2 = 23,634$	$\sum Y = 760$	$\sum Y^2 = 59,816$	$\sum XY = 36,854$

ดังนั้น ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันของคะแนนสอบวิชสถิติธรกิจกับวิชาการจัดการทั่วไปของนักศึกษาเท่ากับ 0.84

10.2.2 วิธีหาค่าสัมประสิทธ์สหสัมพันธ์ลำดับสเปียร์แมน (Spearman's rank correlation coefficient) หลักการ คือ จัดลำดับค่าของ X จากน้อยไปมากและจัดลำดับ Y จากน้อยไปมาก จากนั้นนำลำดับของ X_i , Y_i แต่ละคู่หาผลต่างกัน นำไปหาค่าสหสัมพันธ์สเปียร์แมน (Newbold, 1999, p. 437) ได้เป็นสมการ (10-7)

$$r = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)} \qquad \text{...(10-7)}$$

r คือ ค่าสหสัมพันธ์สเปียร์แมน

D_i คือ ความแตกต่างลำดับครั้งแรกกับครั้งหลัง

n คือ จำนวน X หรือ Y ซึ่งเท่ากัน

ตัวอย่างที่ 10.3 ในการทดสอบความรู้พนักงานก่อนอบรมและหลังอบรมจำนวน 10 คน ได้คะแนนดังตารางที่ 10.4 จงหาค่าสหสัมพันธ์สเปียร์แมนของคะแนนพนักงานก่อนอบรมและหลังอบรม

ตารางที่ 10.4 คะแนนทดสอบความรู้พนักงานก่อนอบรมและหลังอบรม

คะแนน	ชื่อพนักงาน									
	ก	ข	ค	ง	จ	ฉ	ช	ฌ	ญ	ด
ก่อนอบรม	15	18	17	16	12	14	13	16	15	19
หลังอบรม	14	17	19	14	14	14	15	18	17	18

วิธีทำ จัดอันดับของการสอบก่อนหลังเป็นและสร้างตารางที่ 10.5 และแทนค่าตัวแปรต่าง ๆ ลงในสมการ (10-7) จะได้

ตารางที่ 10.5 การหาค่า $\sum D^2$ เพื่อใช้ในการคำนวณหาค่า r จากสมการ (10-7)

พนักงานชื่อ	ลำดับคะแนนก่อนอบรม	ลำดับคะแนนหลังอบรม	ความแตกต่างก่อนหลัง (D)	D^2
ด	1	2.5	-1.5	2.25
ข	2	5	-3	9
ค	3	1	2	4
ง	4.5	9	-4.5	20.25
ฌ	4.5	2.5	2	4
ก	6.5	9	-2.5	6.25
ญ	6.5	5	1	1
ฉ	8	9	-1	1
ช	9	7	2	4
ด	10	5	5	25
				$\sum D^2$=76.75

ค่าสหสัมพันธ์สเปียร์แมน

$$r = 1 - \frac{(6)(76.75)}{10(10^2 - 1)}$$

$$= 0.534$$

ดังนั้น ค่าสหสัมพันธ์สเปียร์แมนของความแตกต่างอันดับเท่ากับ 0.534

10.3 การวิเคราะห์ความสัมพันธ์ระหว่างสองตัวแปร

จากหัวข้อ 10.2 ได้เน้นการแสดงความสัมพันธ์ระหว่างสองตัวแปรชนิดคู่ด้วยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่างขนาด n ที่ชักจากประชากรขนาดใหญ่ หัวข้อนี้จึงมุ่งเน้นการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ด้วยการทดสอบสมมุติฐานของข้อมูลที่นำมาวิเคราะห์ได้จากการชักตัวอย่างสุ่มจากประชากรที่มีการแจกแจงแบบปกติ การทดสอบความสัมพันธ์ระหว่าง X กับ Y ของประชากร จะใช้สมมุติฐานหลักว่า X กับ Y ไม่สัมพันธ์เชิงเส้นกันหรืออิสระกัน หากปฏิเสธสมมุติฐานหลักจะสรุปว่ามีความสัมพันธ์เชิงเส้นกันหรือไม่อิสระกัน (Keller & Warrack, 2002 ,pp. 657-658) ดังสมการ (10-8)

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0 \qquad \ldots(10\text{-}8)$$

ขั้นตอนการทดสอบความสัมพันธ์ระหว่างสองตัวแปรอาจดำเนินการเช่นเดียวกับตัวอย่างที่ 10.5 หรือดำเนินการตามขั้นตอนดังนี้

ขั้นที่ 1 ตั้งสมมุติฐานหลัก H_0 และสมมุติฐานแย้ง H_1 ในการทดสอบ

ขั้นที่ 2 กำหนดระดับนัยสำคัญ α

ขั้นที่ 3 หาค่าตัวสถิติทดสอบ

ขั้นที่ 4 เปิดตารางค่าวิกฤติตัวสถิติที่จำนวนองศาเสรีและที่ระดับนัยสำคัญ α ที่กำหนด

ขั้นที่ 5 นำค่าสถิติทดสอบที่คำนวณได้ไปเปรียบเทียบกับค่าค่าวิกฤติตัวสถิติตารางจากตารางค่ตัวสถิติที่จำนวนองศาเสรีที่เท่ากันและระดับนัยสำคัญ α ที่กำหนดถ้าค่าสถิติทดสอบที่คำนวณ ได้ตกนอกเขตการยอมรับสมมุติฐานหลักที่ระดับนัยสำคัญ α นั้นแสดงว่าไม่มีเหตุผลเพียงพอที่จะยอมรับว่าตัวแปรทั้งสองเป็นอิสระต่อกัน

หรือมีความสัมพันธ์กัน แต่ถ้าค่าสถิติทดสอบที่คำนวณ ได้ตกในเขตการยอมรับสมมุติฐานหลักและสรุปว่าที่ระดับนัยสำคัญ α แล้วมีไม่เหตุผลเพียงพอที่จะยอมรับว่า ตัวแปรทั้งสองเป็นอิสระต่อกันหรือไม่มีความสัมพันธ์กัน นั่นคือสรุปได้ว่าตัวแปรทั้งสองไม่สัมพันธ์กัน ซึ่งมีวิธีการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปร ดังนี้

10.3.1 การทดสอบความสัมพันธ์ของเพียร์สัน ข้อตกลงเบื้องต้นว่า ประชากร X และ Y แจกแจงแบบปรกติ (Newbold, 1999, p. 433) กำหนดเขตปฏิเสธหลักเมื่อ t_e น้อยกว่าหรือมากค่าวิกฤตทีจากตารางภาคผนวกที่ 6 ($t_e < t_{(n-2),1-\alpha/2}$ หรือ $t_e > \pm t_{(n-2),\alpha/2}$) โดยขนาดตัวอย่างและค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณตามสมการ (10-2) ถึง สมการ (10-6) ใช้ตัวสถิติทดสอบทีดังสมการ (10-9)

$$t_e = r\sqrt{\frac{(n-2)}{(1-r^2)}} \qquad \text{...(10-9)}$$

โดย n คือขนาดตัวอย่าง

r คือ ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ของตัวอย่าง

ตัวอย่างที่ 10.4 จากตารางแสดงคะแนนสอบสถิติธุรกิจและคะแนนสอบวิชาการจัดการทั่วไปของนักศึกษาจำนวน 10 คนเป็นตัวอย่างสุ่มที่ชักจากนักศึกษาที่ลงทะเบียนเรียนจำนวนมาก จงทดสอบว่าคะแนนสอบสถิติธุรกิจและคะแนนสอบวิชาการจัดการทั่วไปของนักศึกษาทั้งสัมพันธ์กันหรือไม่

วิธีทำ

ทดสอบสมมุติฐาน

ขั้นที่ 1 $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

ขั้นที่ 2 $\alpha = .05$

ขั้นที่ 3 คำนวณค่าสถิติทดสอบ

$$t_e = r\sqrt{\frac{(n-2)}{(1-r^2)}}$$

$$t_e = 0.84\sqrt{\frac{8}{1-(.84)^2}}$$

$$= 0.84\sqrt{\frac{8}{.30}}$$

$$= 0.84(5.16)$$

$$= 4.33$$

ขั้นที่ 4 เปิดตารางภาคผนวกที่ 6 ที่ α =.05/2 , $df = n-2 = 8$ ค่าวิกฤต $t_{.975,\,8} = -2.306$ และ $t_{.025,\,8} = 2.306$

ขั้นที่ 5 พบค่า $t_e > t_{.025,\,8}$ จึงปฏิเสธสมมุติฐานหลัก ที่ $\alpha = .05$ สรุปได้ว่า คะแนนสอบสถิติธุรกิจและคะแนนสอบวิชาการจัดการทั่วไปของนักศึกษาที่ลงทะเบียนเรียนทั้งหมดมีความสัมพันธ์กัน

ตัวอย่างที่ 10.5 จากการสุ่มสำรวจรายจ่ายและรายรับของพนักงานธนาคารแห่งหนึ่ง จำนวน 49 ราย พบค่าสหสัมพันธ์ระหว่างรายจ่ายกับรายรับของพนักงาน เป็น 0.43 จงทดสอบ รายจ่ายกับรายรับของพนักงานทั้งหมดมีความสัมพันธ์กันหรือไม่ที่ความเชื่อมั่น 99 %

วิธีทำ $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$n = 49$, $r = 0.43$

คำนวณค่าสถิติทดสอบ

$$t_e = r\sqrt{\frac{(n-2)}{(1-r^2)}}$$

$$= 0.43\sqrt{\frac{49-2}{1-(.43)^2}}$$

$$t_e = 3.265$$

จากตารางภาคผนวกที่ 16 $t_{(n-2),.005} = t_{47,.005} = 2.689$
ดังนั้นเชื่อมั่นได้ร้อยละ 99 ว่ารายจ่ายและรายรับของพนักงานธนาคารแห่งนี้ทั้งหมดสัมพันธ์กัน

10.3.2 การทดสอบความสัมพันธ์ลำดับของสเปียร์แมน สเปียร์แมนได้ให้หลักการทดสอบสมมุติฐานความพันธ์ระหว่างลำดับตัวแปรสองตัวไว้สองแนวคือกรณีใช้ตัวอย่างขนาดใหญ่และกรณีใช้ตัวอย่างขนาดเล็ก เพื่อการใช้งานทางธุรกิจทั่วไปที่มักใช้ตัวอย่างขนาดเล็ก (Newbold, 1999, p. 437) สถิติทดสอบจากสมการ (10-10) โดยใช้สมมุติฐานจากสมการ (10-11) เมื่อกำหนดให้ ρ_s คือค่าสัมประสิทธิ์สหสัมพันธ์ประชากรและ n คือขนาดตัวอย่าง

$$r_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)} \qquad \text{...(10-10)}$$

เมื่อ r คือ ค่าสหสัมพันธ์สเปียร์แมน

d_i คือ ความแตกต่างลำดับครั้งแรกกับครั้งหลัง

n คือ ขนาดตัวอย่างจำนวน x หรือ y ซึ่งเท่ากัน

และตั้งสมมมุติฐานดังสมการ(10-11)

$$H_0 : \rho_s = 0$$

$$H_1 : \rho_s \neq 0 \qquad \text{...(10-11)}$$

จะปฏิเสธ H_0 เมื่อ $r_s < -r_{n,\alpha/2}$ หรือ $r_s > r_{n,\alpha/2}$ ที่ n ใด ๆ จากตารางภาคผนวกที่ 16

ตัวอย่างที่ 10.6 จากการสุ่มสำรวจความสูงและน้ำหนักของพนักงานหญิง จำนวน 15 คน ในตารางที่ 10.6 จงทดสอบว่าลำดับความสูงและน้ำหนักของพนักงานหญิงทั้งหมดมีความสัมพันธ์กันหรือไม่ที่ความเชื่อมั่น 95 %

ตารางที่ 10.6 ความสูงและน้ำหนักของพนักงานหญิง จำนวน 15 คน

พนักงานคนที่	ความสูง (ซม.)	น้ำหนัก(กก.)
1	110	44
2	116	31
3	124	43
4	129	45
5	131	56
6	138	79
7	142	57
8	150	56
9	153	58
10	155	92
11	156	78
12	159	64
13	164	88
14	168	112
15	174	101
รวม	2169	1004
เฉลี่ย	144.6	66.933

วิธีทำ จัดอันดับของการสอบก่อนหลังเป็นและสร้างตารางช่วยคำนวณค่าตามสมการ (10-10) ดังตารางที่ 10.7 แล้วคำนวณได้จากสูตร

$$r_s = 1 - \frac{6(60.5)}{15\left(225 - 1\right)}$$

$$= 0.892$$

ตารางที่ 10.7 การหาค่า $\sum d^2$ เพื่อใช้ในการคำนวณหาค่า r_s จากสมการ (10-10)

พนักงานคนที่	ลำดับ ความสูง	ลำดับน้ำหนัก	ความแตกต่าง ลำดับ(d)	d^2
1	1	3	2	4
2	2	1	-1	1
3	3	2	-1	1
4	4	4	0	0
5	5	5.5	0.5	0.25
6	6	11	5	25
7	7	7	0	0
8	8	5.5	-2.5	6.25
9	9	8	-1	1
10	10	13	3	9
11	11	10	-1	1
12	12	9	-3	9
13	13	12	-1	1
14	14	15	1	1
15	15	14	-1	1
				$\sum d^2$=60.5

จากตารางภาคผนวกที่ 16 เมื่อ $n = 15$ $r_{15,.025} = 0.525$ ดังนั้นคา $r_s = 0.8920$ ตกในเขตปฏิเสธ H_0 สรุป ลำดับความสูงและน้ำหนักของพนักงานหญิงทั้งหมดมีความสัมพันธ์กัน

10.4 การวิเคราะห์การถดถอยเชิงเดียว

ในหัวข้อ 10.3 เป็นการศึกษาเบื้องต้นที่ให้เห็นความสัมพันธ์ระหว่างตัวแปรตัวหนึ่ง ซึ่งเรียกว่าตัวแปรตาม (dependent variable) และตัวแปรอีกตัวหนึ่งเรียกตัวแปรนี้ว่า

ตัวแปรอิสระ (independent variable) หัวข้อนี้มุ่งเน้นการใช้ตัวแปรอิสระกับค่าคงตัวค่าหนึ่งเป็นตัวที่ช่วยในการสร้างตัวแบบ (model) หรือสมการความสัมพันธ์ เพื่อหาค่าของตัวแปรตามเมื่อทราบว่าตัวแปรทั้งสองความสัมพันธ์กัน หากตัวแปรอิสระมีจำนวนหนึ่งตัวเรียกว่าเป็นการวิเคราะห์การถดถอยเชิงเดียว (simple regression analysis) และถ้าตัวแปรอิสระมีตั้งแต่ 2 ตัวขึ้นไปเรียกว่าการวิเคราะห์การถดถอยพหุคูณ (multiple regression analysis) เป้าหมายของการวิเคราะห์คือ การใช้ตัวแบบประมาณค่าตัวแปรตามและหาตัวแบบที่เหมาะสม ดังจะกล่าวถึงต่อไปนี้

10.4.1 ตัวแบบประชากรและข้อสมมุติ ข้อมูลประชากรชนิดคู่ (X, Y) เมื่อได้ระบุค่า X แล้วเราก็จะตามด้วยค่า Y ที่สอดคล้องกับค่า X ข้อตกลงเบื้องต้นเพื่อการวิเคราะห์ ได้แก่ ค่าใดค่าหนึ่งของ X จะมีค่า Y_i ได้หลายค่าและ Y_i นั้น มีการแจกแจงปรกติ ด้วยค่าเฉลี่ยเท่ากับ $\mu_y = \beta_0 + \beta_1 \mu_X$ และค่าแปรปรวน σ^2 แต่ค่าความแปรปรวน σ^2 ของความคลาดเคลื่อน ณ ค่าต่าง ๆ ของ X จะเท่ากัน นั่นคือ $\sigma^2_{y_i/x_1} = \sigma^2_{y_i/x_i} = \ldots = \sigma^2_{y_i/x_n}$ และค่าของ Y_i ซึ่งต้องเป็นตัวแปรเชิงปริมาณระดับอัตราหรือเป็นระดับอัตราโดยอนุโลมได้ สมการการถดถอยประชากรตามสมการ (10-12) แต่เมื่อใช้ ข้อมูลตัวอย่างงานวิเคราะห์ความถดถอยเชิงเดี่ยวนี้จะจบลงที่ว่า สามารถหาสมการการถดถอยได้เป็นสมการ (10 -12) โดยประมาณค่า β_0 และ β ด้วยค่า b_0 และ b พร้อมทั้งแทน Y_i ด้วย $\hat{Y}$ และจำได้เป็นสมการ (10 -13) ต่อไปนี้

$$Y = \beta_0 + \beta X \quad \ldots(10\text{-}12)$$

$$\hat{Y} = b_0 + bX \quad \ldots(10\text{-}13)$$

10.4.2 ขั้นตอนของสร้างตัวแบบการถดถอยเชิงเดี่ยว มีดังนี้

1) ขั้นที่ 1 เขียนกราฟข้อมูลตามคู่ลำดับ (X, Y) เพื่อเป็นการตรวจสอบความสัมพันธ์เชิงเส้นตรงในเบื้องต้น

2) ขั้นที่ 2 หาสมการ $Y = \beta_0 + \beta X$ ด้วยการประมาณ β_0 ด้วย b_0 และประมาณ β ด้วย b ซึ่งคำนวณได้ด้วยสมการ (10-14) และสมการ (10-15) ต่อไปนี้

$$b_0 = \overline{Y} - b\overline{X} \quad \ldots(10\text{-}14)$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2} \quad \ldots(10\text{-}15)$$

ตัวอย่างที่ 10.7 จากข้อมูลระดับเกรดเฉลี่ยกับจำนวนเวลาดูหนังสือต่อสัปดาห์ของนักศึกษาของมหาวิทยาลัยแห่งหนึ่ง ถ้ามหาวิทยาลัยแห่งนั้นต้องการทำนายจำนวนเวลาดูหนังสือต่อสัปดาห์ของนักศึกษาก็จะสร้างสมการในรูป $\hat{Y} = b_0 + bX$ เมื่อ $\hat{Y}$ เป็นจำนวนเวลาดูหนังสือต่อสัปดาห์ของนักศึกษาและ X เป็นระดับเกรดเฉลี่ยของนักศึกษา

วิธีทำ คำนวณค่า b_0 และ b

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

ตารางที่ 10.8 เกรดเฉลี่ยและจำนวนเวลาการดูหนังสือของนักศึกษาจำนวน 6 คน

นักศึกษา	เกรดเฉลี่ย (X)	จำนวนเวลา (Y)	X^2	XY
1	3.90	7	15.21	27.30
2	2.50	2	6.25	5.00
3	3.00	4	9.00	12.00
4	4.00	8	16.00	32.00
5	2.00	0	4.00	0
6	3.80	6	14.40	22.80
รวม	19.20	27	64.90	99.10

$$b = \frac{(6)(99.10) - (19.20)(27)}{(6)(64.90) - (19.20)^2}$$

$$= 3.67$$

$$b_0 = \overline{Y} - b\overline{X}$$

$$= \frac{27}{6} - 3.67\left(\frac{19.20}{6}\right)$$

$$= -7.24$$

สมการความถดถอยคือ $\hat{Y} = -7.24 + 3.67X$

ค่าประมาณจำนวนเวลาดูหนังสือต่อสัปดาห์ของนักศึกษา $\hat{Y}$

ถ้านักศึกษาต้องการเกรดเฉลี่ย 2.25 ควรจะมีจำนวนเวลาดูหนังสือต่อสัปดาห์

$$\hat{Y} = -7.24 + 3.67(2.25)$$

$$= 1.02 \text{ ชั่วโมง}$$

ถ้านักศึกษาต้องการเกรดเฉลี่ย 3.25 ควรจะมีจำนวนเวลาดูหนังสือต่อสัปดาห์

$$\hat{Y} = -7.24 + 3.67(3.25)$$

$$= 4.69 \text{ ชั่วโมง}$$

10.4.3 ขั้นตอนของงานวิเคราะห์การถดถอยเชิงเดี่ยว หลังจากได้ตัวแบบหรือสมการการถดถอยแล้ว จะต้องดำเนินการตรวจสอบถึงตัวแบบประชากรว่าควรยอมรับหรือไม่ ขอกล่าวเป็นหลักการไว้ 4 ขั้นตอน ดังนี้

1) ขั้นที่ 1 ทดสอบความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระซึ่งเป็นการวิเคราะห์ความสัมพันธ์ระหว่างสองตัวแปรที่กล่าวแล้วในหัวข้อ 10.2 แล้ว

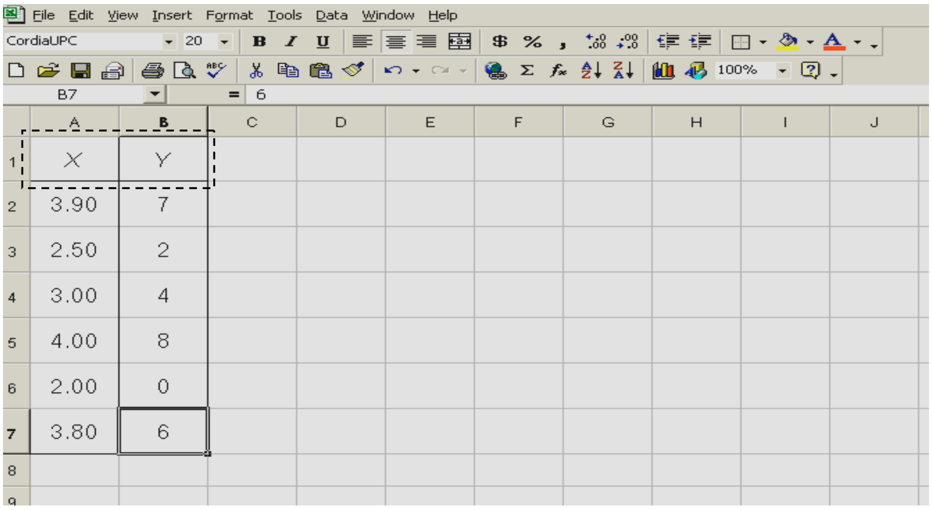
2) ขั้นที่ 2 สร้างตารางวิเคราะห์ความแปรปรวน เพื่อความสะดวกในการหาค่าต่าง ๆ สำหรับการทดสอบสมมุติฐานเกี่ยวกับ β_0 และ β ด้วยสถิติเอฟ

3) ขั้นที่ 3 ประมาณค่าของความเชื่อมั่น β_0 และ β

4) ขั้นที่ 4 การใช้เส้นถดถอยเพื่อการประมาณเป็นช่วงให้กับค่าเฉลี่ยของ Y เมื่อกำหนดค่า X ให้ และค่าประมาณของ Y เมื่อกำหนดค่า X ให้

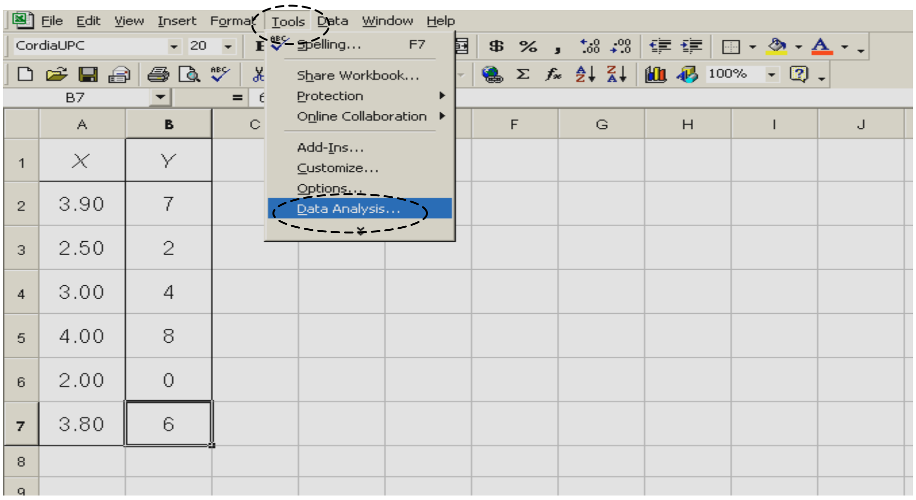
10.4.4 การวิเคราะห์ความถดถอยเชิงเดี่ยวด้วยโปรแกรมสำเร็จรูป สำหรับโปรแกรมสำเร็จรูปที่ใช้สำหรับประมวลผลข้อมูลสถิติ เช่นโปรแกรมเอสพีเอสเอส (SPSS) โปรแกรมเอส.เอ.เอส (S.A.S) โปรแกรมมินิแทป (MINITAP) หรือโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล (Microsoft Excel) เป็นต้น ในที่นี้จะแนะการใช้งานไมโครซอฟเอกเซลซึ่งมีใช้ในสำนักงานทั่วไปและมีเมนูช่วยเหลือ เมนูคำสั่งที่ใช้ในโปรแกรมเป็นภาษาอังกฤษ ดังนั้นขั้นตอนในการใช้ต่อไปนี้จะใช้ ข้อความในการอธิบายการใช้โปรแกรมวิเคราะห์ข้อมูล (โดยใช้ข้อมูลตามตารางที่ 10.8)

1) ขั้นที่ 1 เตรียมข้อมูลในโปรแกรมสำเร็จรูป ด้วยการกำหนดข้อมูล *X* และ *Y* เป็น 2 สดมภ์ ดังภาพที่ 10.6

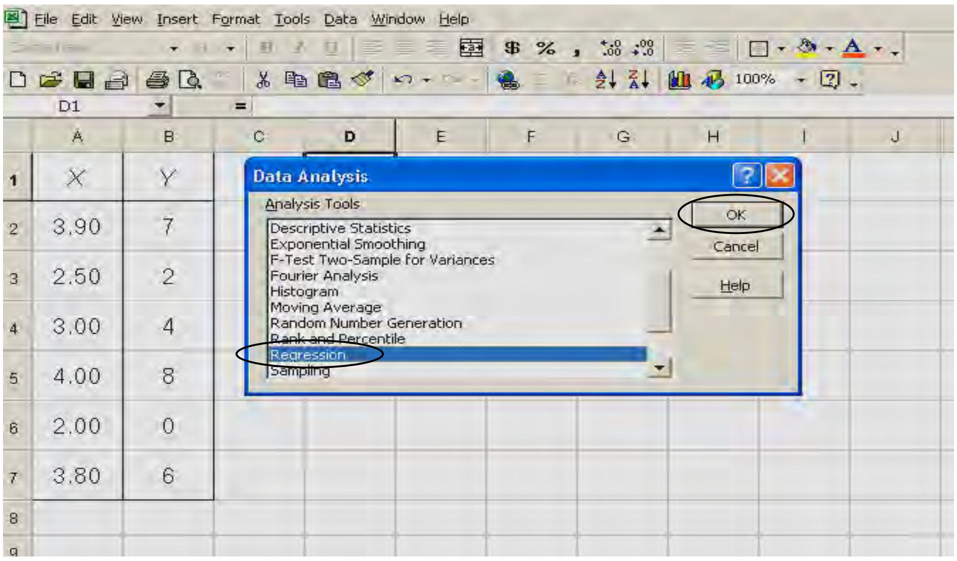


ภาพที่ 10.6 ข้อมูลจากตารางที่ 10.8 ในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

2) ขั้นที่ 2 เลือกเมนูเครื่องมือ (Tools) ตามด้วยเลือกวิเคราะห์ข้อมูล (Data Analysys)และการถดถอย (Regression) แล้วคลิกโอเค (O.K.) ดังภาพที่ 10.7 และภาพที่ 10.8

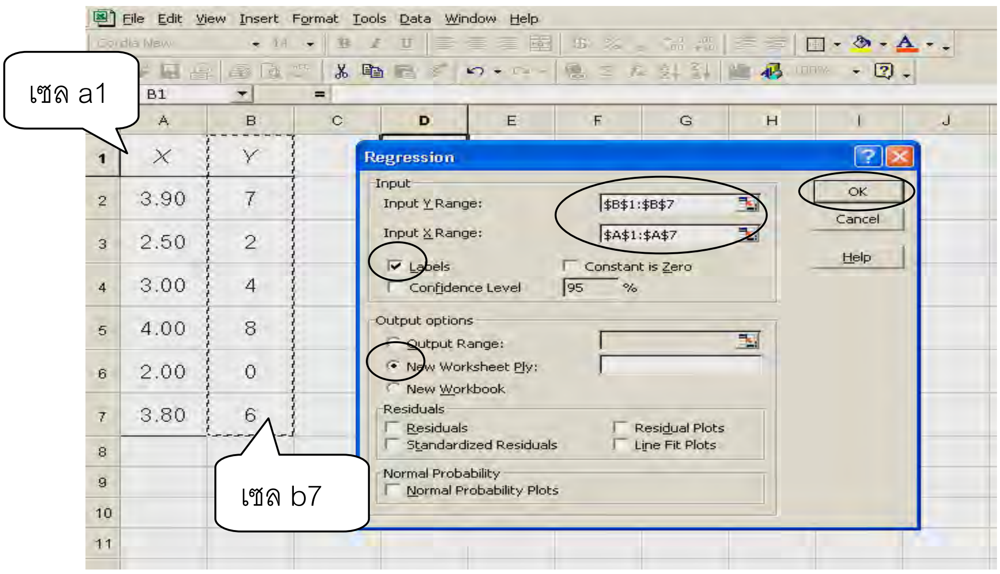


ภาพที่ 10.7 เลือกเมนูคำสั่งเครื่องมือ ตามเลือกวิเคราะห์ข้อมูล
ในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล



ภาพที่ 10.8 เลือกเมนูคำสั่งการถดถอย ในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

3) ขั้นที่ 3 กำหนดขอบเขตตัวแปร *X* เป็นเซลที่ a1 ถึง a7 กำหนดขอบเขตตัวแปร *Y* เป็นเซลที่ b1 ถึง b7 เลือกรหัสตัวแปร (Label) และเลือกส่วนแสดงผล (Output option) เป็นแผ่นงานใหม่ (New Worksheet Ply) แล้วคลิกโอเค (O.K.) ดังภาพที่ 10.9



ภาพที่ 10.9 กำหนดขอบเขตตัวแปร *X* เป็นเซลที่ a1ถึง a7 กำหนดขอบเขตตัวแปร *Y* เป็นเซลที่ b1ถึง b7 เลือกรหัสตัวแปร และเลือกส่วนแสดงผลเป็นแผ่นงานใหม่ ในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

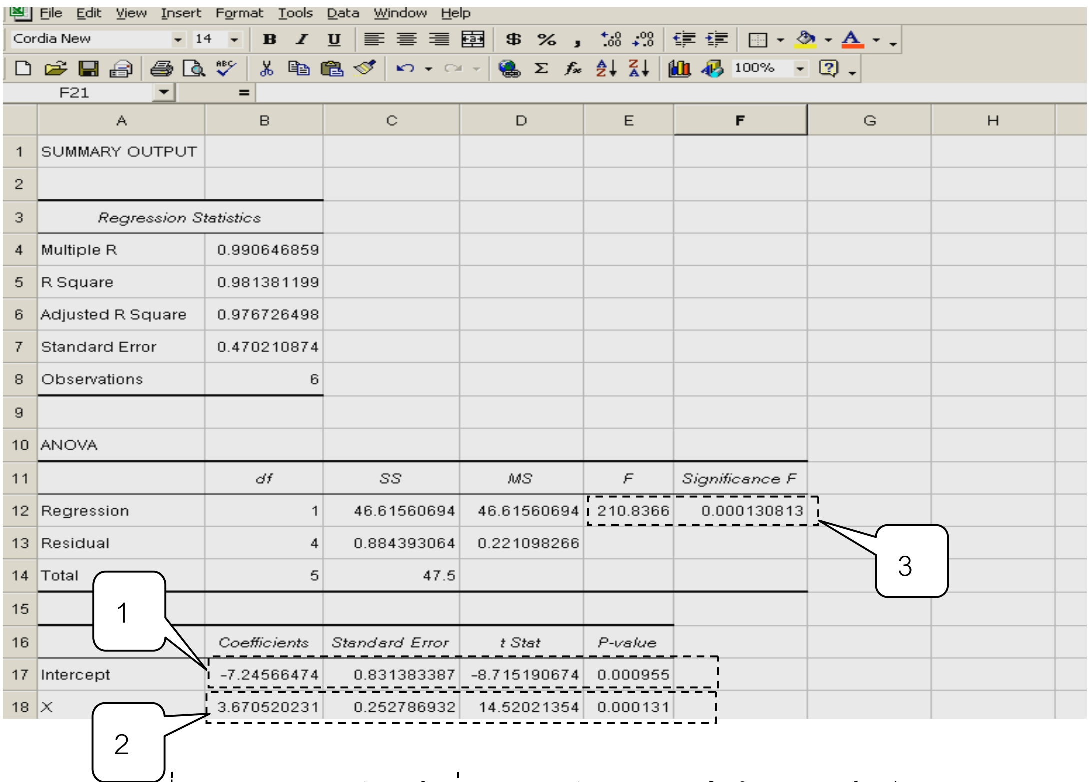
4) ขั้นที่ 4 อ่านผลลัพธ์ส่วนสำคัญเกี่ยวกับสมการการถดถอย 3 ส่วน ดังภาพที่ 10.10 คือ

ส่วนที่ 1 ค่า b_0 สำหรับประมาณ β_0 เท่ากับ -7.24566474 ค่าผิดพลาดมาตรฐานของ b_0 ค่าสถิติทดสอบทีที่คำนวณได้ สำหรับ $H_0 : \beta_0 = 0$ และค่าความน่าเป็น (p-value) ของสถิติทดสอบทีที่คำนวณได้ ถ้าน้อยกว่า .05 หรือ .01 จะปฏิเสธ H_0 หรือยอมรับค่าประมาณ β_0 นี้ที่ระดับ α .05 หรือ .01 ตามลำดับ กรณีนี้ยอมรับค่าประมาณ β_0

ส่วนที่ 2 ค่า b สำหรับประมาณ β เท่ากับ 3.670520231 ค่าผิดพลาดมาตรฐานของ b ค่าสถิติทดสอบทีที่คำนวณได้ สำหรับ $H_0 : \beta = 0$ และความน่าเป็นของสถิติทดสอบทีที่คำนวณได้ กรณีนี้ยอมรับค่าประมาณ β เช่นเดียวกับ β_0

ส่วนที่ 3 ค่าสถิติทดสอบเอฟ สำหรับ $H_0 : \beta = 0$ เมื่อความน่าจะเป็นของสถิติเอฟที่คำนวณได้น้อยกว่า .05 หรือ .01 ทำนองเดียวกับความน่าจะเป็นสถิติทดสอบที กรณีนี้ยอมรับสมการการถดถอย

จึงได้สมการการถดถอยคือ $\hat{Y} = -7.24566474 + 3.670520231X$



	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.990646859				
5	R Square	0.981381199				
6	Adjusted R Square	0.976726498				
7	Standard Error	0.470210874				
8	Observations	6				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	46.61560694	46.61560694	210.8366	0.000130813
13	Residual	4	0.884393064	0.221098266		
14	Total	5	47.5			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	-7.24566474	0.831383387	-8.715190674	0.000955	
18	X	3.670520231	0.252786932	14.52021354	0.000131	

ภาพที่ 10.10 ผลลัพธ์จากคำสั่งวิเคราะห์การถดถอยในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

10.5 การวิเคราะห์การถดถอยพหุคูณ

ในหัวข้อ 10.4 เป็นการวิเคราะห์การถดถอยนั้นมีหลักการว่าค่าของตัวแปรที่จะพยากรณ์หรือตัวแปรตามถูกกำหนดขึ้นโดยค่าของ Y เมื่อมีตัวแปรอิสระหลาย ๆ ตัวและความสัมพันธ์ระหว่างตัวแปรอิสระเหล่านั้นสัมพันธ์กับตัวแปรตาม เช่น ยอดขายสินค้าชนิดหนึ่ง อาจขึ้นอยู่กับค่าใช้จ่ายในการโฆษณา จำนวนตัวแทนขาย จำนวนผู้บริโภคเป้าหมาย เป็นต้น ในที่นี้ปริมาณขายเป็นตัวแปรตาม สำหรับตัวแปรอิสระประกอบด้วยค่าใช้จ่ายในการโฆษณา จำนวนตัวแทนขาย จำนวนผู้บริโภคเป้าหมาย เป็นต้น ซึ่งในที่นี้จะกล่าวถึงโดยสังเขปดังนี้

10.5.1 ตัวแบบความถดถอยพหุคูณ ลักษณะความสัมพันธ์เมื่อกำหนด Y คือยอดขาย X_1 คือค่าใช้จ่ายในการโฆษณา X_2 คือจำนวนตัวแทนขาย X_3 คือจำนวนผู้บริโภคเป้าหมาย อาจแสดงตัวแบบตามสมการ (10-16) เป็นสมการการถดถอยพหุคูณของประชากร

$$Y = \beta_0+\beta_1 X_1 +\beta_2 X_2 +\beta_3 X_3 +\ldots \qquad \ldots(10\text{-}16)$$

การกำหนดตัวแบบของฟังก์ชันซึ่งอาจจะเป็นเชิงเส้นหรือไม่เชิงเส้นก็ได้ ซึ่งมีความสำคัญมากต่อความถูกต้องในการพยากรณ์ค่า Y และมักจะเป็นปัญหาของผู้สร้างตัวแบบพยากรณ์ อย่างไรก็ตามการกำหนดตัวแปรตามและตัวแบบของสมการการถดถอยควรตั้งอยู่บนพื้นฐานของทฤษฎีที่เกี่ยวข้องกับตัวแบบตามกำลังศึกษาอยู่ด้วย เมื่อมีการชักตัวอย่างสุ่ม ตัวแบบการถดถอยจะประมาณค่า β_0 ด้วยค่า b_0 และแทนสัมประสิทธ์ β_1 ด้วยค่า b_1 แทน β_2 ด้วยค่า b_2 แทน β_3 ด้วยค่า b_3 จนครบตามจำนวนตัวแปรอิสระ พร้อมทั้งแทน Y_i ด้วย $\hat{Y}$ ดังสมการ (10.17)

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots \qquad \ldots(10\text{-}17)$$

หลังจากกำหนดตัวแบบขึ้นแล้ว การประมาณค่าของสัมประสิทธิ์ต่าง ๆ อาจกระทำได้หลายลักษณะและหลายวิธีการเช่นอาจกระทำในลักษณะการถดถอยแบบขั้นบันได (stepwise regression) ด้วยวิธีกำลังสองน้อยสุด (least square) เป็นต้น เมื่อได้ประมาณค่าของสัมประสิทธิ์แล้ว การทดสอบความเหมาะสมของสมการถดถอยและสัมประสิทธิ์ก็อาจกระทำได้ ประมาณช่วงความเชื่อมั่นของค่าตัวแปรตาม ก็จะสามารถคำนวณได้เช่นเดียวกัน

10.5.2 การวิเคราะห์การถดถอยพหุคูณด้วยโปรแกรมสำเร็จรูป การวิเคราะห์การถดถอยนี้ได้รับความนิยมอย่างมาก เพราะเป็นการพยากรณ์ตัวแปรตามเพื่อการตัดสินใจเกี่ยวด้านธุรกิจด้านการขยายกำลังการผลิตหรืออื่น ๆ ซึ่งเป็นรูปธรรมที่สุดและใช้เครื่องวัดการวัดความถูกต้องและระดับความเชื่อมั่นของสมการที่จะใช้พยากรณ์และของค่าพยากรณ์และการพยากรณ์ในลักษณะนี้สนองความเชื่อที่ว่าสิ่งที่จะพยากรณ์ถูกกำหนดให้เป็นไปโดยปัจจัยต่าง ๆ ที่อยู่ภายนอกระบบและในระบบ การคำนวณด้วยมือจึงยากลำบากมาก ในที่นี้จะขอแนะนำการใช้งานโปรแกรมไมโครซอฟเอกเซลสำหรับประมาณค่าสัมประสิทธิ์ต่าง ๆ ตามสมการ (10–17) ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 10.8 จากข้อมูลในตารางที่ 10.9 จงสร้างสมการถดถอยของราคาขายห้องชุด

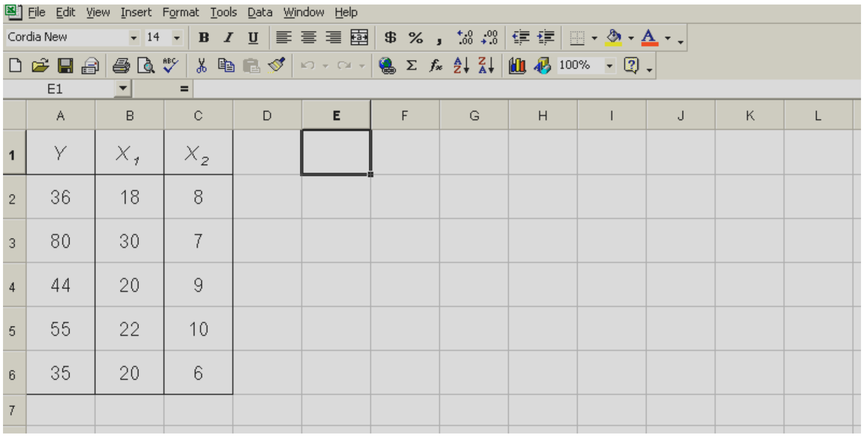
ตารางที่ 10.9 ข้อมูลราคาขายห้องชุด

ราคาห้องชุด (×100,000 บาท)	ขนาดพื้นที่ (ตารางเมตร)	ราคาที่ดินต่อไร่ (×100,000 บาท)
36	18	8
80	30	7
44	20	9
55	22	10
35	20	6

วิธีทำ ให้ Y เป็นราคาห้องชุด
X_1 เป็นขนาดพื้นที่
X_2 เป็นราคาที่ดินต่อไร่

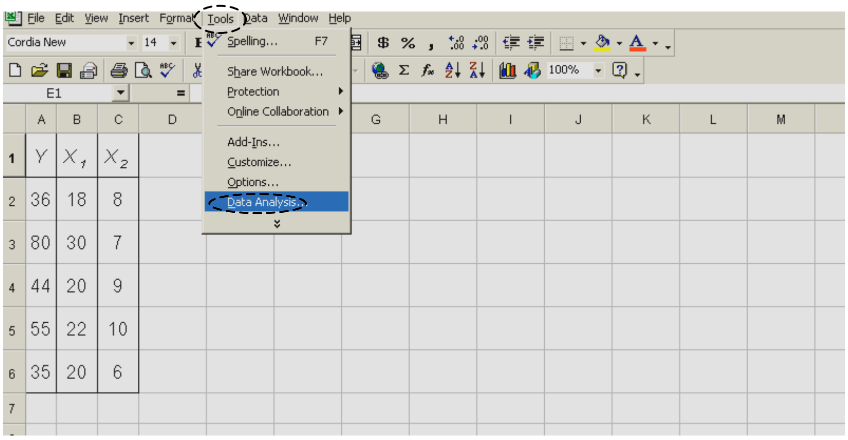
ดำเนินการใช้โปรแกรมสำเร็จรูปไมโครซอฟเอกเซลตามลำดับขั้นตอนดังนี้

ขั้นที่ 1 นำข้อมูลจากตารางที่ 10.9 ไปใส่ในโปรแกรมสำเร็จรูปดังแสดงในภาพที่ 10.11

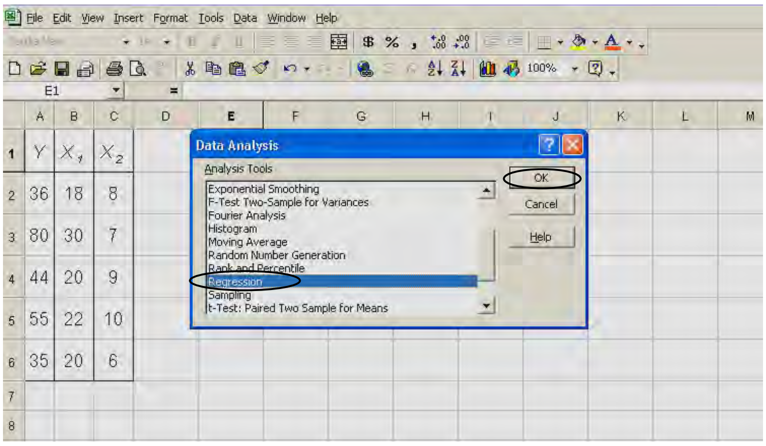


ภาพที่ 10.11 ข้อมูล จากตารางที่ 10.9 ในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

ขั้นที่ 2 เลือกเมนูเครื่องมือ ตามด้วยเลือกวิเคราะห์ข้อมูลและการถดถอย แล้วคลิกโอเค ดังแสดงในภาพที่ 10.12 และ ภาพที่ 10.13



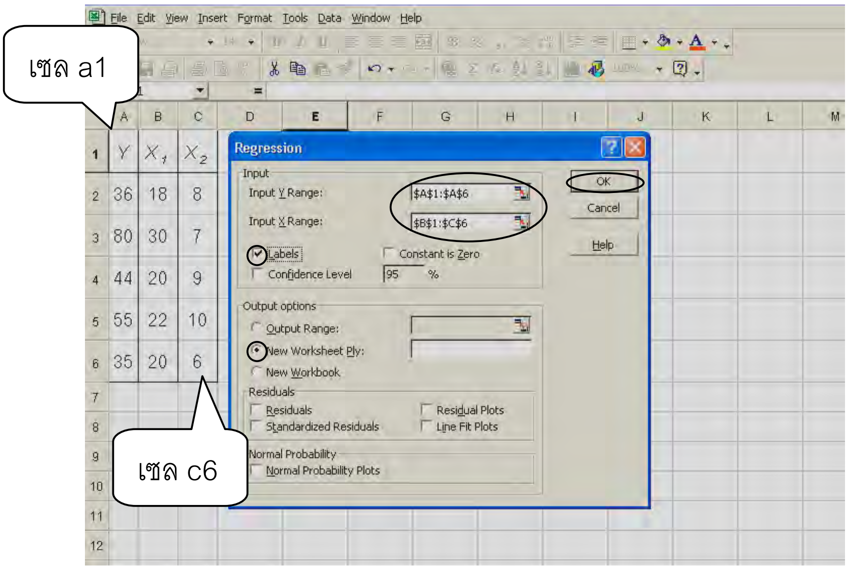
ภาพที่ 10.12 เลือกเมนูคำสั่งตามด้วยเลือกวิเคราะห์ข้อมูล



ภาพที่ 10.13 เลือกเมนูคำสั่งการถดถอยเพื่อวิเคราะห์ข้อมูลของตัวอย่างที่ 10.8

ขั้นที่ 3 กำหนดขอบเขตตัวแปร Y เป็นเซล a1 ถึง a6

กำหนดขอบเขตตัวแปร X_1 และ X_2 เป็นเซล b1 ถึง c6 ดังแสดงใน ภาพที่ 10.14



ภาพที่ 10.14 กำหนดขอบเขตตัวแปร Y เป็นเซลที่ a1 ถึง a6

และ X_1 และ X_2 เป็นเซลที่ b1 ถึง c6 ของตัวอย่างที่ 10.8

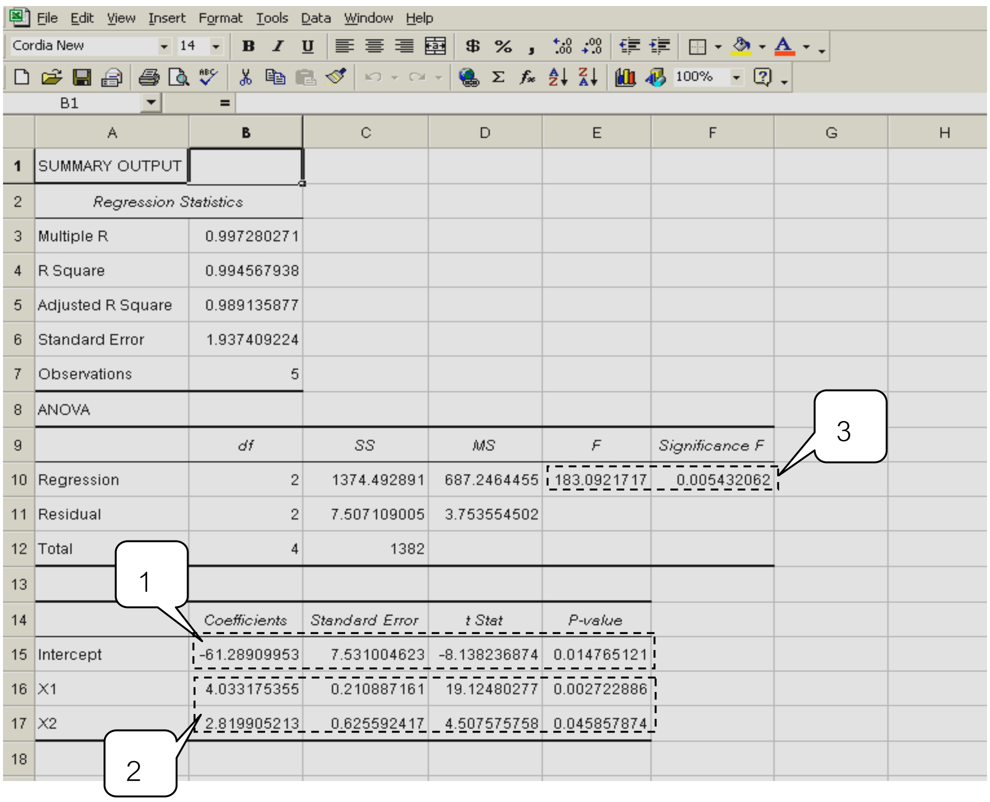
ขั้นที่ 4 อ่านผลลัพธ์ส่วนสำคัญเกี่ยวกับสมการการถดถอย 3 ส่วนดังภาพที่ 10.15 คือ

ส่วนที่ 1 ค่า b_0 สำหรับประมาณ β_0 เท่ากับ -61.289 ค่าผิดพลาดมาตรฐานของ b_0 ค่าสถิติทดสอบทีที่คำนวณได้สำหรับ $H_0 : \beta_0 = 0$ และความน่าเป็นของสถิติทดสอบทีที่คำนวณได้ กรณีนี้ยอมรับค่าประมาณ β_0

ส่วนที่ 2 ค่า b_1 สำหรับประมาณ β_1 เท่ากับ 4.033 และ β_2 เท่ากับ 2.819 ค่าผิดพลาดมาตรฐานของ b_1 และ b_2 สถิติทดสอบทีที่คำนวณได้สำหรับ $H_0 : \beta_1 = 0$ และ $\beta_2 = 0$ และความน่าเป็นของสถิติทดสอบทีที่คำนวณได้ กรณีนี้ยอมรับค่าประมาณ β_1 และ β_2

ส่วนที่ 3 ค่าสถิติทดสอบเอฟ สำหรับ $H_0 : \beta_1 = 0$ และ $\beta_2 = 0$ จะยอมรับสมการการถดถอยหรือปฏิเสธ H_0 เมื่อความน่าจะเป็นของสถิติเอฟที่คำนวณได้น้อยกว่า .05 หรือ .01 กรณีนี้ยอมรับสมการการถดถอย

จึงได้สมการความถดถอยคือ $\hat{Y} = -61.289 + 4.033\,X_1 + 2.819\,X_2$



	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2	*Regression Statistics*					
3	Multiple R	0.997280271				
4	R Square	0.994567938				
5	Adjusted R Square	0.989135877				
6	Standard Error	1.937409224				
7	Observations	5				
8	ANOVA					
9		*df*	*SS*	*MS*	*F*	*Significance F*
10	Regression	2	1374.492891	687.2464455	183.0921717	0.005432062
11	Residual	2	7.507109005	3.753554502		
12	Total	4	1382			
13						
14		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
15	Intercept	-61.28909953	7.531004623	-8.138236874	0.014765121	
16	X1	4.033175355	0.210887161	19.12480277	0.002722886	
17	X2	2.819905213	0.625592417	4.507575758	0.045857874	
18						

ภาพที่ 10.15 ผลลัพธ์จากคำสั่งวิเคราะห์การถดถอยในโปรแกรมสำเร็จรูปไมโครซอฟเอกเซล ของตัวอย่างที่ 10.8

Xxxxxx (หน้านี้ไม่ใช้แต่ต้องพิมพ์)

xxxxxxxxxxxxxx

10.6 การทดสอบความอิสระระหว่างสองตัวแปรด้วยสถิติไคกำลังสอง

สำหรับการชักตัวอย่างตัวแปรระดับนามบัญญัติสองตัวแปรพร้อมๆกัน เช่น ตัวแปรเพศกับตัวแปรระดับการศึกษา ตัวแปรอาชีพกับตัวแปรระดับความพอใจในสินค้าชนิดหนึ่ง เมื่อนำตัวแปรทั้งสองมาแจกแจงพร้อมๆกันในตาราง เรียกว่า ตารางการจรสองทางของข้อมูล (two-way contingency table of data) โดยทั่วไปจะกำหนดตัวแปรหนึ่งในแนวตั้งหรือแนวสดมภ์ อีกตัวแปรหนึ่งอยู่ในแนวนอน ความสัมพันธ์ระหว่างสองตัวมีการแจกแจงไคกำลังสอง (Watson & Croft.,1996, pp.558-559) การทดสอบนี้เรียกว่า การทดสอบความอิสระระหว่างสองตัวแปร (test independent between two variable)

10.6.1 การสร้างตารางการจรสองทางของข้อมูล กรณีตัวแปรในแนวตั้งที่ i เท่ากับ 1, 2, 3,……….,c และตัวแปรในแนวนอนที่ j เท่ากับ 1, 2, 3,……….,r จะได้ตารางการจรสองทาง ตามตารางที่ 10.10

ตารางที่ 10.10 ตารางการจรสองทางของข้อมูล c แนวตั้ง และ r แนวนอน

แนวตั้งที่ / แนวนอนที่	1	2	....	c	รวมแนวนอน R_i
1	O_{11}	O_{21}	…	O_{c1}	R_1
2	O_{12}	O_{22}	…	O_{c2}	R_2
.	.	.	…	.	.
.	.	.	…	.	.
r	O_{1r}	O_{2r}	…	O_{cr}	R_r
รวมแนวนอน C_j	C_1	C_2	…	C_c	n

โดย

c แทนจำนวนค่าตัวแปรในแนวตั้ง

r แทนจำนวนค่าตัวแปรในแนวนอน

O_{ij} แทนจำนวนแจกนับหรือความถี่ที่รวบรวมได้ตรงกับแนวตั้งที่ i และแนวนอนที่ j ของตัวแปรในแนวตั้งและตัวแปรในแนวนอน ตามลำดับ

R_i แทนจำนวนแจกนับหรือความถี่ที่รวบรวมได้ตรงกับแนวตั้งที่ i ของตัวแปรแนวนอนทั้งหมด

C_j แทนจำนวนแจกนับหรือความถี่ที่รวบรวมได้ตรงกับแนวนอนที่ i ของตัวแปรแนวตั้งทั้งหมด

n แทนขนาดตัวอย่าง เท่ากับ $\sum_{i=1}^{c} R_i$ เท่ากับ $\sum_{j=1}^{r} C_j$

10.6.2 การทดสอบความเป็นอิสระระหว่างสองตัวแปร สำหรับการดำเนินการทดสอบสมมุติฐานเกี่ยวกับความเป็นอิสระหรือการทดสอบความสัมพันธ์ระหว่างตัวแปรแนวตั้งกับตัวแปรแนวนอน เช่น ทดสอบความเป็นอิสระกันระหว่างตัวแปรเพศ (ค่า 1 คือชาย ค่า 2 คือ หญิง) กับระดับการศึกษา (ค่า 1 คือต่ำกว่าปริญญาตรี ค่า 2 คือ ระดับปริญญาตรี ค่า 3 คือระดับปริญญาโท และ ค่า 4 คือระดับปริญญาเอก) ทดสอบความเป็นอิสระกันระหว่างตัวแปรอาชีพ (ค่า 1 คือข้าราชการ ค่า 2 คือนักธุรกิจ ค่า 3 คือเกษตรกร) กับระดับความพอใจในการใช้สินค้าชนิดหนึ่ง (ค่า 1 คือพอใจน้อยที่สุด ค่า 2 คือพอใจ ค่า 3 คือพอใจมาก และ ค่า 4 คือพอใจมากที่สุด) เป็นต้น จะตั้งสมมุติฐาน H_o และ H_1 ทำนองเดียวกันกับการทดสอบความสัมพันธ์ระหว่างตัวแปร X กับ Y ตั้งสมมุติฐานตามสมการ (10-8) หรืออาจตั้งสมมุติฐานเป็นข้อความต่อไปนี้

H_0 : ตัวแปรทั้งสองอิสระกัน

H_1 : ตัวแปรทั้งสองไม่อิสระกัน

เช่น

H_0 : ระดับการศึกษาไม่ขึ้นกับเพศ

H_1 : ระดับการศึกษาขึ้นกับเพศ

หรือ

H_0 : ความพอใจในสินค้าไม่ขึ้นกับอาชีพ

H_1 : ความพอใจในสินค้าขึ้นกับอาชีพ เป็นต้น

สถิติทดสอบคือ สถิติไคกำลังสอง ตามสมการ (10-18)

$$\chi^2_{cal} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{r} (O_{ij} - E_{ij})^2}{E_{ij}} \qquad \ldots(10\text{-}18)$$

ให้ E_{ij} แทนจำนวนแจกนับหรือความถี่ที่คาดหวัง ตรงกับแนวตั้งที่ i และแนวนอนที่ j ของตัวแปรในแนวตั้งและตัวแปรในแนวนอน ตามลำดับและเมื่อจัดข้อมูลตามตารางที่ 10.10 คำนวณได้จากสมการ (10–19)

$$E_{ij} = \frac{R_i C_j}{n} \qquad \ldots(10\text{-}19)$$

จะปฏิเสธ H_0 เมื่อค่า χ^2_{cal} ที่คำนวณได้มากกว่า χ^2 จากตารางภาคผนวกที่ 11 อ่านที่ $df = (r\text{-}1)(c\text{-}1)$ และระดับ α = .05 หรือ .01 ที่กำหนดหรือกรณีคำนวณด้วยโปรแกรมสำเร็จรูป ค่าความน่าจะเป็นของ χ^2_{cal} น้อยกว่า .05 หรือ .01 จะปฏิเสธ H_0 เช่นเดียวกัน

ตัวอย่างที่ 10.9 จงทดสอบว่าพฤติกรรมการดื่มสุราของชายสัมพันธ์กับสถานภาพสมรสหรือไม่ โดยอาศัยข้อมูลที่รวบรวมได้ตามตารางที่ 10.11

ตารางที่ 10.11 ตารางการจรสองทางจำนวนแจกนับที่รวบรวมได้จากพฤติกรรมการดื่มสุราของชาย และ สถานภาพสมรส

พฤติกรรมการดื่มสุรา	สถานภาพสมรส				รวม R_i
	โสด	อยู่กับครอบครัว	ไม่อยู่กับครอบครัว	หย่าร้าง	
บ่อย	34	36	16	14	100
นานๆครั้ง	82	64	34	20	200
ไม่ดื่ม	84	50	50	16	200
รวม C_j	200	150	100	50	500

วิธีทำ **กำหนด**

H_0 : พฤติกรรมการดื่มสุราของชายไม่สัมพันธ์สถานภาพสมรส

H_1 : พฤติกรรมการดื่มสุราของชายสัมพันธ์สถานภาพสมรส

สร้างตารางเพื่อหา $E_{ij} = \frac{R_i C_j}{n}$ ดังตารางที่ 10.12

ตารางที่ 10.12 ตารางการจรสองทางของจำนวนแจกนับที่คาดหวังจากพฤติกรรมการดื่มสุราของชาย และสถานภาพสมรส

พฤติกรรมการดื่มสุรา	สถานภาพสมรส				รวม R_i
	โสด	อยู่กับครอบครัว	ไม่อยู่กับครอบครัว	หย่าร้าง	
บ่อย	$\frac{(100)(200)}{500}=40$	$\frac{(100)(150)}{500}=30$	$\frac{(100)(100)}{500}=20$	$\frac{(100)(50)}{500}=10$	100
นานๆครั้ง	$\frac{(200)(200)}{500}=80$	$\frac{(200)(150)}{500}=60$	$\frac{(200)(100)}{500}=40$	$\frac{(200)(50)}{500}=20$	200
ไม่ดื่ม	$\frac{(200)(200)}{500}=40$	$\frac{(200)(150)}{500}=60$	$\frac{(200)(100)}{500}=40$	$\frac{(200)(50)}{500}=20$	200
รวม C_j	200	150	100	50	500

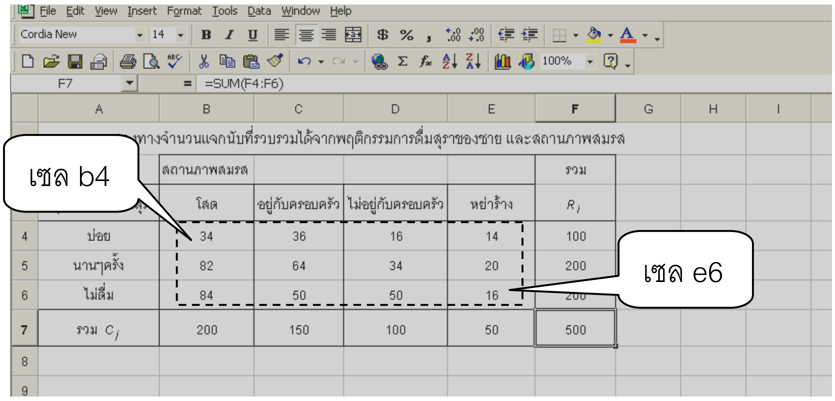
$$\chi^2 = \frac{(34-40)^2}{40} + \frac{(36-30)^2}{30} + \frac{(16-20)^2}{20} + \frac{(14-10)^2}{10}$$
$$+ \frac{(82-80)^2}{80} + \frac{(64-60)^2}{60} + \frac{(34-40)^2}{40} + \frac{(20-20)^2}{20}$$
$$+ \frac{(84-40)^2}{40} + \frac{(50-60)^2}{60} + \frac{(50-40)^2}{40} + \frac{(16-20)^2}{20}$$
$$= 10.88$$

ค่า χ^2_{cal} ที่คำนวนได้น้อยกว่านี้เปรียบเทียบกับค่า χ^2 ที่ df = (3-1)(4-1) และ α = .05 หรือน้อยกว่า $\chi^2_{6,.05}$ = 12.59 จึงไม่สามารถปฏิเสธ H_0 ได้

สรุปได้ว่าพฤติกรรมการดื่มสุราของชาย และ สถานภาพสมรสไม่สัมพันธ์กัน

10.6.3 การทดสอบความอิสระระหว่างสองตัวแปร ด้วยโปรแกรมสำเร็จรูปไมโครซอฟเอกเซลมีฟังชันสถิติและอื่นๆจำนวนมากให้ความสะดวกสำหรับงานวิเคราะห์ข้อมูล ดำเนินการใช้โปรแกรมสำเร็จรูปไมโครซอฟเอกเซลตามลำดับขั้นตอนดังนี้

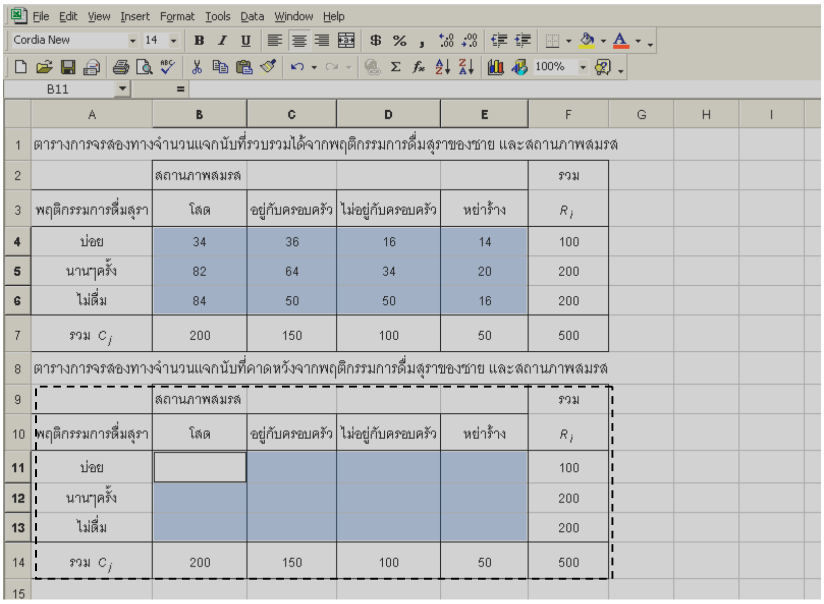
ขั้นที่ 1 นำข้อมูลจากตารางที่ 10.11 ไปใส่ในโปรแกรมสำเร็จรูปดังแสดงในภาพที่ 10.16 สำหรับกำหนดเป็นจำนวนแจกนับที่รวบรวมได้โดยไม่รวมชายขอบหรือผลรวม อยู่ช่วงเซล b4 ถึง e6



	A	B	C	D	E	F
1	ทางจำนวนแจกนับที่รวบรวมได้จากพฤติกรรมการดื่มสุราของชาย และสถานภาพสมรส					
2		สถานภาพสมรส				รวม
3		โสด	อยู่กับครอบครัว	ไม่อยู่กับครอบครัว	หย่าร้าง	R_i
4	บ่อย	34	36	16	14	100
5	นานๆครั้ง	82	64	34	20	200
6	ไม่ดื่ม	84	50	50	16	200
7	รวม C_j	200	150	100	50	500

ภาพที่ 10.16 ข้อมูลจากตารางที่ 10.10 โปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

ขั้นที่ 2 คัดลอกตารางจำนวนแจกนับที่รวบรวมได้จากขั้นที่ 1 ให้เหลือเฉพาะผลรวมขอบตาราง ดังภาพที่ 10.17



	A	B	C	D	E	F
1	ตารางการจรสองทางจำนวนแจกนับที่รวบรวมได้จากพฤติกรรมการดื่มสุราของชาย และสถานภาพสมรส					
2		สถานภาพสมรส				รวม
3	พฤติกรรมการดื่มสุรา	โสด	อยู่กับครอบครัว	ไม่อยู่กับครอบครัว	หย่าร้าง	R_i
4	บ่อย	34	36	16	14	100
5	นานๆครั้ง	82	64	34	20	200
6	ไม่ดื่ม	84	50	50	16	200
7	รวม C_j	200	150	100	50	500
8	ตารางการจรสองทางจำนวนแจกนับที่คาดหวังจากพฤติกรรมการดื่มสุราของชาย และสถานภาพสมรส					
9		สถานภาพสมรส				รวม
10	พฤติกรรมการดื่มสุรา	โสด	อยู่กับครอบครัว	ไม่อยู่กับครอบครัว	หย่าร้าง	R_i
11	บ่อย					100
12	นานๆครั้ง					200
13	ไม่ดื่ม					200
14	รวม C_j	200	150	100	50	500

ภาพที่ 10.17 ผลรวมของค่าที่ได้จำนวนแจกนับที่รวบรวมได้จากขั้นที่ 1

ขั้นที่ 3 สร้างจำนวนแจกนับที่คาดหวังในเซล b11 ถึง e13

โดยไปที่ เซล b11 พิมพ์ = b14*f11/ f14 แล้วกด ↵

เซล b12 พิมพ์ = b14*f12/ f14 แล้วกด ↵

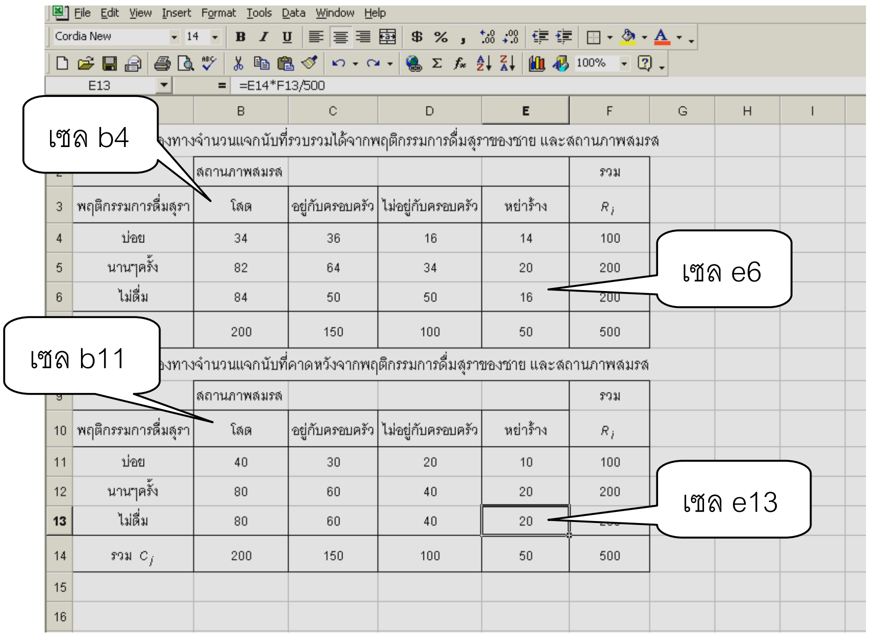
เซล b13 พิมพ์ = b14*f13/ f14 แล้วกด ↵

. . . . .

. . . . .

เซล e13 พิมพ์ = e14*f13/ f14 แล้วกด ↵

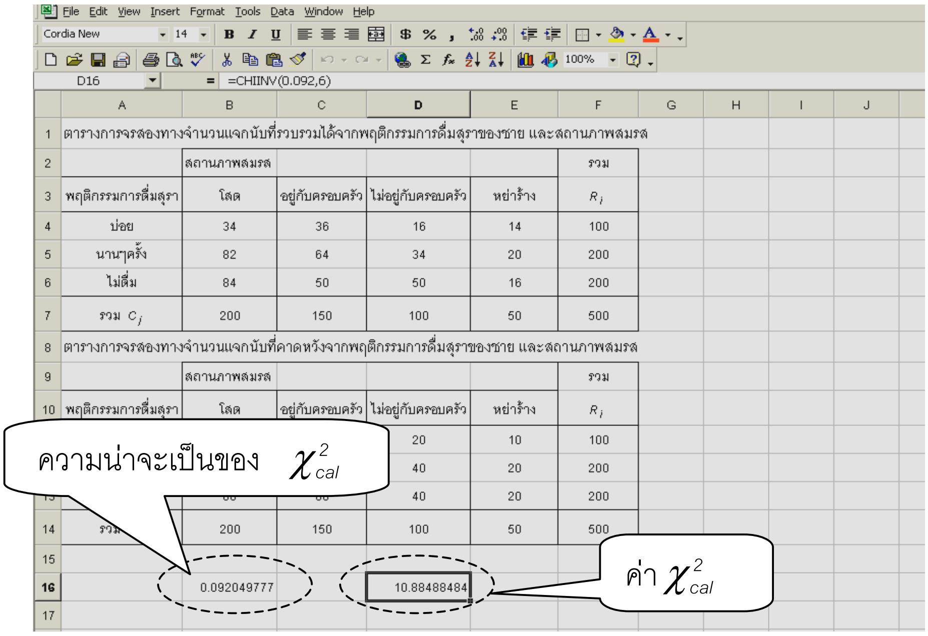
จะได้ผลลัพธ์ตามภาพที่ 10.18



ภาพที่ 10.18 จำนวนแจกนับที่คาดหวังในเซล b11 ถึง e13

ขั้นที่ 4 หาค่าความน่าจะเป็นของ χ^2_{cal} โดยพิมพ์ข้อความ =CHITEST (b4:e6,b11:e13) ลงในเซล b16 ได้เท่ากับ 0.092 และถ้าต้องการทราบค่า χ^2_{cal} ให้นำค่าความน่าจะเป็นที่นี้พิมพ์ในเซล d16 ด้วยข้อความ =CHIINV(0.092,6) จะได้ χ^2_{cal} เท่ากับค่าที่ได้จากการคำนวณตามตัวอย่างที่ 10.9 ดังแสดงในภาพที่ 10.19

ขั้นที่ 5 ขั้นสรุปผลการทดสอบ ถ้าความน่าจะเป็นในเซล b16 น้อยกว่า .05 หรือ .01 จะปฏิเสธ H_0 ที่ระดับ α = .05 หรือ α = .01 ตามลำดับ สำหรับกรณีนี้ค่าในเซล b16 เท่ากับ 0.092 มากกว่า .05 จึงยอมรับ H_0 เช่นเดียวกับตัวอย่างที่ 10.9



ภาพที่ 10.19 ผลลัพธ์จากคำสั่งทดสอบความเป็นอิสระระหว่างสองตัวแปรใน
โปรแกรมสำเร็จรูปไมโครซอฟเอกเซล

10.7 บทสรุป

บทนี้จะเน้นการนำข้อมูลเชิงปริมาณอย่างน้อยสองชุดนำมาวิเคราะห์หาความสัมพันธ์เพื่อพิจารณารูปแบบความสัมพันธ์

10.7.1 ลักษณะของความสัมพันธ์ด้วยใช้แผนภูมิ

10.7.2 ความสัมพันธ์ตัวแปรทั้งสองไม่ได้มุ่งลงเพียงการศึกษาว่ามีความสัมพันธ์กันหรือไม่เท่านั้ แต่ความสนใจนั้นได้เลยไปถึงการศึกษาว่าตัวแปรนั้นมีความสัมพันธ์กันมากน้อยเพียงใด และมีลักษณะของความสัมพันธ์เป็นอย่างไร ด้วยวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันและค่าสัมประสิทธิ์สหสัมพันธ์ลำดับสเปียร์แมน

10.7.3 เสนอแนวทางการสร้างสมการถดถอยเชิงเดียวจากตัวอย่างสุ่ม

10.7.4 แนวทางทดสอบสมมุติฐานเกี่ยวกับสมการการถดถอยเชิงเดี่ยวและการทดสอบสมมุติเกี่ยวพารามิเตอร์ β_0 β_1

10.7.5 การวิเคราะห์ความถดถอยเชิงเดี่ยวด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์

10.7.6 แนวคิดสมการการถดถอยพหุคูณและการทดสอบสมมุติเกี่ยวพารามิเตอร์ทั้งหมด

10.7.7 การวิเคราะห์ความถดถอยพหุคูณด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์

10.7.8 การทดสอบความอิสระระหว่างสองตัวแปรด้วยสถิติไคกำลังสอง

10.8 คำถามทบทวน

1. จงสร้างแผนภูมิแสดงความสัมพันธ์ จากข้อมูลความสูงและความยาวรอบศีรษะของพนักงานจำนวน 15 คน ดังตารางที่ 10.13

2. จงหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันของความสูง และความยาวรอบศีรษะของพนักงานจำนวน 15 คน จากข้อมูลข้อ 10.1

3. จงหาค่าสัมประสิทธ์สหสัมพันธ์ลำดับสเปียร์แมนของความสูงและความยาวรอบศีรษะของพนักงานจำนวน 15 คน จากข้อมูลข้อ 10.1

ตารางที่ 10.13 ความสูงและความยาวรอบศีรษะของพนักงานจำนวน 15 คน

คนที่	ความสูง (cm)	ความยาวรอบศีรษะ(cm)
1	111	44
2	117	31
3	125	43
4	130	45
5	132	56
6	139	79
7	143	57
8	151	56
9	154	58
10	156	92
11	157	78
12	160	64
13	165	88
14	169	112
15	175	101

4. จากการศึกษาจำนวนวันลาในรอบปีของ 2 พนักงาน 16 คนกับระยะทางระหว่างที่พักกับโรงงานดังข้อมูลในตารางที่ 10.14 จงหาค่าสหสัมพันธ์สเปียร์แมนของจำนวนวันลากับระยะทางระหว่างบ้านกับที่พักพนักงาน

ตารางที่ 10.14 จำนวนวันลาและระยะทางระหว่างที่พักกับโรงงาน

พนักงานคนที่	จำนวนวันลา	ระยะทาง(km)
1	21	6.8
2	12	10.3
3	30	1.7
4	8	14.2
5	10	8.8
6	26	5.8
7	42	2.1
8	31	3.3
9	21	4.3
10	15	9.0
11	19	3.2
12	6	12.7
13	18	8.2
14	12	7.0
15	23	5.1
16	34	4.1

5. ประธานบริษัทและรองประธานได้ให้คะแนนลำดับความสำคัญของเจ้าหน้าที่ของบริษัทจำนวน 12 คน เพื่อใช้ตัดสินในการลดจำนวนพนักงานที่มีความสำคัญน้อยที่สุด ปรากฏว่าคะแนนดังตารางที่ 10.15 จงหาว่าความเห็นของประธานบริษัทและรองประธานตรงกันหรือไม่

6. ฝ่ายโฆษณาของบริษัทต้องการตรวจสอบจำนวนครั้งที่โฆษณาต่อเดือนทาง หนังสือพิมพ์สัมพันธ์กับยอดขายรวมหรือไม่ โดยการทดลองโฆษณาเป็นเวลา 9 เดือน ได้ ข้อมูลจำนวนครั้งโฆษณากับยอดขายรวมดังตารางที่ 10.16

ตารางที่ 10.15 การให้คะแนนของประธานบริษัทและรองประธานบริษัทแก่พนักงาน จำนวน 12 คน

พนักงานคนที่	คะแนนจากประธานบริษัท	คะแนนจากรองประธานบริษัท
1	8	2
2	2	9
3	7	5
4	4	10
5	3	11
6	1	12
7	9	4
8	12	3
9	6	7
10	11	4
11	15	8
12	10	1

ตารางที่ 10.16 จำนวนครั้งที่โฆษณาต่อเดือนทางหนังสือพิมพ์สัมพันธ์กับยอดขาย

เดือนที่	จำนวนครั้งโฆษณา	ยอดขายรวม (ล้านบาท)
1	1	2.0
2	2	1.0
3	3	3.0
4	4	3.0
5	5	4.0
6	6	5.0
7	7	6.0
8	8	5.0
9	9	7.0

7. ห้างสรรพสินค้ามีข้อมูลค่าใช้จ่ายในการโฆษณากับรายได้จากการขาย ในช่วง 15 สัปดาห์ที่ผ่านมา เป็นดังตารางที่ 10.17 จงสร้างสมการถดถอยรายได้จากการขายกับค่าใช้จ่ายในการโฆษณา
8. จากข้อมูลในคำถามข้อ 10.5 จงสร้างสมการถดถอยเชิงเดี่ยว
9. จงอธิบายพร้อมยกตัวอย่างแนวการสร้างสมการถดถอยพหุคูณมา 3 ตัวแบบ
10. จงบอกข้อเสียของการพยากรณ์ตัวแปรตามด้วยสมการการถดถอยมาอย่างน้อย 2 ข้อ

ตารางที่ 10.17 ค่าใช้จ่ายในการโฆษณากับรายได้จากการขาย

สัปดาห์ที่	ค่าใช้จ่ายโฆษณา (ล้านบาท)	รายได้รวม (ล้านบาท)
1	3.2	114.6
2	2.5	86.5
3	2.3	95.5
4	2.4	61.8
5	3.5	154.8
6	3.2	120.9
7	2.6	104.2
8	3.0	136.8
9	2.8	143.5
10	2.5	98.7
11	3.0	128.2
12	2.7	99.2
13	2.6	107.3
14	3.1	130.1
15	2.8	150.2

11. จากตารางที่ 10.18 จงสร้างสมการถดถอยเพื่อประมาณยอดขายจากค่าใช้จ่ายในการโฆษณาและจำนวนแหล่งที่วางขายโดยใช้โปรแกรมสำเร็จรูป

ตารางที่ 10.18 ค่าใช้จ่ายในการโฆษณากับจำนวนแหล่งที่วางขาย

สัปดาห์ที่	ยอดขาย (ล้านบาท)	ค่าใช้จ่ายในการโฆษณา (ล้านบาท)	จำนวนแหล่งที่วางขาย
1	8.74	2	2
2	10.53	2	3
3	10.99	2	4
4	11.97	2	5
5	12.74	3	2
6	12.83	3	3
7	14.69	3	4
8	15.30	3	5
9	16.10	4	2
10	16.31	4	3
11	16.46	4	4
12	17.69	4	5
13	19.65	5	2
14	18.86	5	3
15	19.93	5	4
16	20.51	5	5

12. จงทดสอบว่าการลากิจและการลาป่วยของพนักงานในแผนกอาหารสดขึ้นกับเหตุจูงใจที่เข้าทำงานหรือไม่ ข้อมูลแจกนับจากพนักงานจำนวน 552 รายตามตารางที่ 10.19

ตารางที่ 10.19 ข้อมูลแจกนับการลากิจและการลาป่วยของพนักงานในแผนกอาหารสด ขึ้นกับเหตุจูงใจที่เข้าทำงาน

เหตุจูงใจ	การลา		รวม R_i
	ลาป่วย	ลากิจ	
ชอบอาชีพนี้	181	103	284
ได้รับทุนแผนกนี้	113	70	183
ไม่ทราบจะทำอะไร	35	21	56
ทำชั่วคราว	17	12	29
รวม C_j	346	206	552

บรรณานุกรม

สมจิต วัฒนาชยากูล. (2545). **สถิติวิเคราะห์เบื้องต้น.** กรุงเทพมหานคร: ประกายพรึก.

Alonso, M. & Finn, E. (1967). **Fundamental University Physics.** (3 rd ed). New York: Addison-Wesley.

Black, K. (1992). **Business Statistics and introductory course.** New York: West Publishing.

Brown, R. A., & Beck, S. J. (1994). **Medical Statistics on personal computers.** (2 nd ed). London: BMJ Publishing Group.

Campbell, M. J. (1995). **Statistics at square one.** [Online]. Available FTP: http://bmj.com/collections/statsbk/11.shtml.[2002,Nov 9].

Casella, G., & Berger, R. L. (1990). **Statistics Inference.** California: Wadsworth.

Chao, L. L. (1969). **Statistics Methods and analyses.** New York: McGraw-Hill.

Harnett, D. L., & Murphy, J. L. (1975). **Statitical Analysis.** (3 rd ed). New York: Addison-Wesley.

Hogg, R. V., & Tanis, E. A. (1989). **Probability and Statistical Inference.** New York: Macmillan.

Kanji, G. K. (1999). **100 Statistical tests.** London: SAGE Publications.

Keller, G., & Warrack, B. (2000). **Statistics for Management and Economics.** (5 th ed). New York: Duxbury.

Loether, H. J., & McTavish, D. G. (1980). **Descriptive and Inferential Statistics an Introduction.** Boston: Allyn and Bacon.

Mason, R. D., Lind, D. A., & Marchal, W. G. (1999). **Statistical Techniques in Business and Economics.** (10 th ed). New York: McGraw-Hill.

Montgomery, D. D. (1991). **Design and analysis of experiments.** (3 rd ed). NewYork: John Wiley & son.

Newbold, P. (1995). **Statistis for Business & Economics.** (4 th ed). New Jerscy: Prentice-Hall .

Puranen, J. (2002). **WWW & teaching statistics a teacher's point of view.** [Online]. Available FTP: http://noppa5.pc.helsinki.fi/koe/corr/cor5.html. [2002,Nov 15].

Solomon, R. C. (1996). **Statistics.** London: John Murry.

Walpole, R. E. (1974). **Introduction to Statistics**. (2 nd ed). New York: Macmillan.

Watson, C. J., Billingsley, P., Croft, D. J., & Huntsberger, D.V. (1990). **Statistics for Management and Economics.** (4 th ed). Boston: Allyn and Bacon.

Wonnacott, R. J., & Wonnacott, T. H. (1985). **Introductory Statistics**. New York: John Wily & sons.

ภาคผนวก

ตารางภาคผนวกที่ 1 ตารางสุ่มเลขโดด

39 65 76 45 45	19 90 69 64 61	20 26 36 31 62	58 24 97 14 97	95 06 70 99 00
73 71 23 70 90	65 97 60 12 11	31 56 34 19 19	47 83 75 51 33	30 62 38 20 46
72 20 47 33 84	51 67 47 97 19	98 40 07 17 66	23 05 09 51 80	59 78 11 52 49
75 17 25 69 17	17 95 21 78 58	24 33 45 77 48	69 81 84 09 29	93 22 70 45 80
37 48 79 88 74	63 52 06 34 30	01 31 60 10 27	35 07 79 71 53	28 99 52 01 41
02 89 08 16 94	85 53 83 29 95	56 27 09 24 43	21 78 55 09 82	72 61 88 73 61
87 18 15 70 07	37 79 49 12 38	48 13 93 55 96	41 92 45 71 51	09 18 25 58 94
98 83 71 70 15	89 09 39 59 24	00 06 41 41 20	14 36 59 25 47	54 45 17 24 89
10 08 58 07 04	76 62 16 48 68	58 76 17 14 86	59 53 11 52 21	66 04 18 72 87
47 90 56 37 31	71 82 13 50 41	27 55 10 24 92	28 04 67 53 44	95 23 00 84 47
93 05 31 03 07	34 18 04 52 35	74 13 39 35 22	68 95 23 92 35	36 63 70 35 33
21 89 11 47 99	11 20 99 45 18	76 51 94 84 86	13 79 93 37 55	98 16 04 41 67
95 18 94 06 97	27 37 83 28 71	79 57 95 13 91	09 61 87 25 21	56 20 11 32 44
97 08 31 55 73	10 65 81 92 59	77 31 61 95 46	20 44 90 32 64	26 99 76 75 63
69 29 88 86 13	59 71 74 17 32	48 38 75 93 29	73 37 32 04 05	60 82 29 20 25
41 47 10 25 03	87 63 93 95 17	81 83 83 04 49	77 45 85 50 51	79 88 01 97 30
91 94 14 63 62	08 61 74 51 69	92 79 43 89 79	29 18 94 51 23	14 85 11 47 23
80 06 54 18 47	08 52 85 08 40	48 40 35 94 22	72 65 71 08 86	50 03 42 99 36
67 72 77 63 99	89 85 84 46 06	64 71 06 21 66	89 37 20 70 01	61 65 70 22 12
59 40 24 13 75	42 29 72 23 19	06 94 76 10 08	81 30 15 39 14	81 83 17 16 33
63 62 06 34 41	79 53 36 02 95	94 61 09 43 62	20 21 14 68 86	94 95 48 46 45
78 47 23 53 90	79 93 96 38 63	34 85 52 05 09	85 43 01 72 73	14 93 87 81 40
87 68 65 15 43	97 48 72 66 48	53 16 71 13 81	59 97 50 99 52	24 62 20 42 31
47 60 92 10 77	26 97 05 73 51	88 46 38 03 58	72 68 49 29 31	75 70 16 08 24
56 88 87 59 41	06 87 37 78 48	65 88 69 58 39	88 02 84 27 83	85 81 56 39 38
22 17 68 65 84	87 02 22 57 51	68 69 80 95 44	11 29 01 95 80	49 34 35 86 47
19 36 27 59 46	39 77 32 77 09	79 57 92 36 59	89 74 39 82 15	08 58 94 34 74
16 77 23 02 77	28 06 24 25 93	22 45 44 84 11	87 80 61 65 31	09 71 91 74 25
78 43 76 71 61	97 67 63 99 61	80 45 67 93 82	59 73 19 85 23	53 33 65 97 21
03 28 28 26 08	69 30 16 09 05	53 58 47 70 93	66 56 45 65 79	45 56 20 19 47
04 31 17 21 56	33 73 99 19 87	26 72 39 27 67	53 77 57 68 93	60 61 97 22 61
61 06 98 03 91	87 14 77 43 96	43 00 65 98 50	45 60 33 01 07	98 99 46 50 47
23 68 25 26 00	99 53 93 61 28	52 70 05 48 34	56 65 05 61 86	90 92 10 70 80
15 39 25 70 99	93 86 52 77 65	15 33 59 05 28	22 87 26 07 47	86 96 98 29 06
58 71 96 30 24	18 46 23 34 27	85 13 99 24 44	49 18 09 79 49	74 16 32 23 02
93 22 53 64 39	07 10 63 76 35	87 03 04 79 88	08 13 13 85 51	55 34 57 72 69
78 76 58 54 74	92 38 70 96 92	52 06 79 79 45	82 63 18 27 44	69 66 92 19 09
61 81 31 96 82	00 57 25 60 59	46 72 60 18 77	55 66 12 62 11	08 99 55 64 57
42 88 07 10 05	24 98 65 63 21	47 21 61 88 32	27 80 30 21 60	10 92 35 36 12
77 94 30 05 39	28 10 99 00 27	12 73 73 99 12	49 99 57 94 82	96 88 57 17 91

ที่มา : (Wonnacott & Wonnacott,1985, p. 605)

ตารางภาคผนวกที่ 2 ลอการิทึมสามัญ

N	0	1	2	3	4	5	6	7	8	9
0	....	0000	3010	4771	6021	6990	7782	8451	9031	9542
1	0000	0414	0792	1139	1461	1761	2041	2304	2553	2788
2	3010	3222	3424	3617	3802	3979	4150	4314	4472	4624
3	4771	4914	5051	5185	5315	5441	5563	5682	5798	5911
4	6021	6128	6232	6335	6435	6532	6628	6721	6812	6902
5	6990	7076	7160	7243	7324	7404	7482	7559	7634	7709
6	7782	7853	7924	7993	8062	8129	8195	8261	8325	8388
7	8451	8513	8573	8633	8692	8751	8808	8865	8921	8976
8	9031	9085	9138	9191	9243	9294	9345	9395	9445	9494
9	9542	9590	9638	9685	9731	9777	9823	9868	9912	9956
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3463	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5142	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5376	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6091	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
N	0	1	2	3	4	5	6	7	8	9

ตารางภาคผนวกที่ 2 ลอการิทึมสามัญ (ต่อ)

N	0	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	9062	8069	8075	8082	8089	8096	8102	8109	8118	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8798	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	8689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996
100	0000	0004	0009	0013	0017	0022	0026	0030	0035	0039
N	0	1	2	3	4	5	6	7	8	9

ที่มา : (Alonso & Finn,1967, pp. A-10 - A-11)

ตารางภาคผนวกที่ 3 การแจกแจงทวินาม $P(X \le x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$

		p									
n	*x*	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.9280	0.9090	0.7760	0.4870
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1290	0.0625	0.0256	0.0081	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.9069	0.8464	0.5869
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156	0.0040	0.0010	0.0000	0.0000
	1	0.8857	0.6553	0.5339	0.4202	0.2333	0.1094	0.0410	0.0110	0.0020	0.0000
	2	0.9842	0.9011	0.8306	0.7443	0.5443	0.3438	0.1790	0.0700	0.0170	0.0010
	3	0.9987	0.9830	0.9624	0.9295	0.8208	0.6562	0.4560	0.2560	0.0990	0.0160
	4	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906	0.7670	0.5800	0.3450	0.1140
	5	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844	0.9530	0.8820	0.7380	0.4690
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 3 การแจกแจงทวินาม (ต่อ)

		p									
n	*x*	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
7	0	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078	0.0020	0.0000	0.0000	0.0000
	1	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625	0.0190	0.0040	0.0000	0.0000
	2	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266	0.0960	0.0290	0.0050	0.0000
	3	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000	0.2900	0.1260	0.0330	0.0030
	4	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734	0.5800	0.3530	0.1480	0.0260
	5	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375	0.8410	0.6710	0.4230	0.1500
	6	1.0000	1.0000	0.9999	0.9998	0.9984	0.9922	0.9720	0.9180	0.7900	0.5220
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039	0.0010	0.0000	0.0000	0.0000
	1	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352	0.0090	0.0010	0.0000	0.0000
	2	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445	0.0500	0.0110	0.0010	0.0000
	3	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633	0.1740	0.0580	0.0100	0.0000
	4	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367	0.4060	0.1940	0.0560	0.0050
	5	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555	0.6850	0.4480	0.2030	0.0380
	6	1.0000	0.9999	0.9996	0.9987	0.9915	0.9648	0.8940	0.7450	0.4970	0.1870
	7	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961	0.9830	0.9420	0.8320	0.5700
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020	0.0000	0.0000	0.0000	0.0000
	1	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195	0.0040	0.0000	0.0000	0.0000
	2	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898	0.0250	0.0040	0.0000	0.0000
	3	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539	0.0990	0.0250	0.0030	0.0000
	4	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000	0.2670	0.0990	0.0200	0.0010
	5	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461	0.5170	0.2700	0.0860	0.0080
	6	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102	0.7680	0.5370	0.2620	0.0530
	7	1.0000	1.0000	0.9999	0.9996	0.9962	0.9805	0.9290	0.8040	0.5640	0.2250
	8	1.0000	1.0000	1.0000	1.0000	0.9997	0.9980	0.990	0.9600	0.8660	0.6130
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 3 การแจกแจงทวินาม (ต่อ)

		p									
n	*x*	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	0.0000
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ตารางภาคผนวกที่ 3 การแจกแจงทวินาม (ต่อ)

		p									
n	*x*	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ที่มา : (Keller & Warrack,2000, pp. B-1 – B-7)

ตารางภาคผนวกที่ 4 การแจกแจงปัวซง $P(X \leq x) = \sum_{x=0}^{n} \frac{\lambda^{x} e^{-\lambda}}{x!}$

x	λ									
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0	0.905	0.819	0.741	0.670	0.607	0.549	0.497	0.449	0.407	0.368
1	0.995	0.982	0.963	0.938	0.910	0.878	0.844	0.809	0.772	0.736
2	1.000	0.999	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.920
3	1.000	1.000	1.000	0.999	0.998	0.997	0.994	0.991	0.987	0.981
4	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.996
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

x	λ									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.00
0	0.333	0.301	0.273	0.247	0.223	0.202	0.183	0.165	0.150	0.135
1	0.699	0.663	0.627	0.592	0.558	0.525	0.493	0.463	0.434	0.406
2	0.900	0.879	0.857	0.833	0.809	0.783	0.757	0.731	0.704	0.677
3	0.974	0.966	0.957	0.946	0.934	0.921	0.907	0.891	0.875	0.857
4	0.995	0.992	0.989	0.986	0.981	0.976	0.970	0.964	0.956	0.947
5	0.999	0.998	0.998	0.997	0.996	0.994	0.992	0.990	0.987	0.983
6	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.997	0.995
7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ตารางภาคผนวกที่ 4 การแจกแจงปัวซง (ต่อ)

	λ									
x	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
0	0.111	0.091	0.074	0.061	0.050	0.041	0.033	0.027	0.022	0.018
1	0.355	0.308	0.267	0.231	0.199	0.171	0.147	0.126	0.107	0.092
2	0.623	0.570	0.518	0.469	0.423	0.380	0.340	0.303	0.269	0.238
3	0.819	0.779	0.736	0.692	0.647	0.603	0.558	0.515	0.473	0.433
4	0.928	0.904	0.877	0.848	0.815	0.781	0.744	0.706	0.668	0.629
5	0.975	0.964	0.951	0.935	0.916	0.895	0.871	0.844	0.816	0.785
6	0.993	0.988	0.983	0.976	0.966	0.955	0.942	0.927	0.909	0.889
7	0.998	0.997	0.995	0.992	0.988	0.983	0.977	0.969	0.960	0.949
8	1.000	0.999	0.999	0.998	0.996	0.994	0.992	0.988	0.984	0.979
9	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.996	0.994	0.992
10	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ตารางภาคผนวกที่ 4 การแจกแจงปัวซง (ต่อ)

	λ									
x	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0
0	0.015	0.012	0.010	0.008	0.007	0.006	0.005	0.004	0.003	0.002
1	0.078	0.066	0.056	0.048	0.040	0.034	0.029	0.024	0.021	0.017
2	0.210	0.185	0.163	0.143	0.125	0.109	0.095	0.082	0.072	0.062
3	0.395	0.359	0.326	0.294	0.265	0.238	0.213	0.191	0.170	0.151
4	0.590	0.551	0.513	0.476	0.440	0.406	0.373	0.342	0.313	0.285
5	0.753	0.720	0.686	0.651	0.616	0.581	0.546	0.512	0.478	0.446
6	0.867	0.844	0.818	0.791	0.762	0.732	0.702	0.670	0.638	0.606
7	0.936	0.921	0.905	0.887	0.867	0.845	0.822	0.797	0.771	0.744
8	0.972	0.964	0.955	0.944	0.932	0.918	0.903	0.886	0.867	0.847
9	0.989	0.985	0.980	0.975	0.968	0.960	0.951	0.941	0.929	0.916
10	0.996	0.994	0.992	0.990	0.986	0.982	0.977	0.972	0.965	0.957
11	0.999	0.998	0.997	0.996	0.995	0.993	0.990	0.988	0.984	0.980
12	1.000	0.999	0.999	0.999	0.998	0.997	0.996	0.995	0.993	0.991
13	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.996
14	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ตารางภาคผนวกที่ 4 การแจกแจงปัวซง (ต่อ)

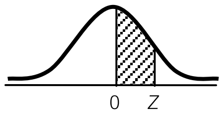
	λ									
x	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0
0	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.011	0.007	0.005	0.003	0.002	0.001	0.001	0.000	0.000	0.000
2	0.043	0.030	0.020	0.014	0.009	0.006	0.004	0.003	0.002	0.001
3	0.112	0.082	0.059	0.042	0.030	0.021	0.015	0.010	0.007	0.005
4	0.224	0.173	0.132	0.100	0.074	0.055	0.040	0.029	0.021	0.015
5	0.369	0.301	0.241	0.191	0.150	0.116	0.089	0.067	0.050	0.038
6	0.527	0.450	0.378	0.313	0.256	0.207	0.165	0.130	0.102	0.079
7	0.673	0.599	0.525	0.453	0.386	0.324	0.269	0.220	0.179	0.143
8	0.792	0.729	0.662	0.593	0.523	0.456	0.392	0.333	0.279	0.232
9	0.877	0.830	0.776	0.717	0.653	0.587	0.522	0.458	0.397	0.341
10	0.933	0.901	0.862	0.816	0.763	0.706	0.645	0.583	0.521	0.460
11	0.966	0.947	0.921	0.888	0.849	0.803	0.752	0.697	0.639	0.579
12	0.984	0.973	0.957	0.936	0.909	0.876	0.836	0.792	0.742	0.689
13	0.993	0.987	0.978	0.966	0.949	0.926	0.898	0.864	0.825	0.781
14	0.997	0.994	0.990	0.983	0.973	0.959	0.940	0.917	0.888	0.854
15	0.999	0.998	0.995	0.992	0.986	0.978	0.967	0.951	0.932	0.907
16	1.000	0.999	0.998	0.996	0.993	0.989	0.982	0.973	0.960	0.944
17	1.000	1.000	0.999	0.998	0.997	0.995	0.991	0.986	0.978	0.968
18	1.000	1.000	1.000	0.999	0.999	0.998	0.996	0.993	0.988	0.982
19	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.994	0.991
20	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.997	0.995
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ตารางภาคผนวกที่ 4 การแจกแจงปัวซง (ต่อ)

	λ									
x	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.003	0.002	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000
4	0.011	0.008	0.005	0.004	0.003	0.002	0.001	0.001	0.001	0.000
5	0.028	0.020	0.015	0.011	0.008	0.006	0.004	0.003	0.002	0.001
6	0.060	0.046	0.035	0.026	0.019	0.014	0.010	0.008	0.006	0.004
7	0.114	0.090	0.070	0.054	0.041	0.032	0.024	0.018	0.013	0.010
8	0.191	0.155	0.125	0.100	0.079	0.062	0.048	0.037	0.029	0.022
9	0.289	0.242	0.201	0.166	0.135	0.109	0.088	0.070	0.055	0.043
10	0.402	0.347	0.297	0.252	0.211	0.176	0.145	0.118	0.096	0.077
11	0.520	0.462	0.406	0.353	0.304	0.260	0.220	0.185	0.154	0.127
12	0.633	0.576	0.519	0.463	0.409	0.358	0.311	0.268	0.228	0.193
13	0.733	0.682	0.628	0.573	0.518	0.464	0.413	0.363	0.317	0.275
14	0.815	0.772	0.725	0.675	0.623	0.570	0.518	0.466	0.415	0.368
15	0.878	0.844	0.806	0.764	0.718	0.669	0.619	0.568	0.517	0.467
16	0.924	0.899	0.869	0.835	0.798	0.756	0.711	0.664	0.615	0.566
17	0.954	0.937	0.916	0.890	0.861	0.827	0.790	0.749	0.705	0.659
18	0.974	0.963	0.948	0.930	0.908	0.883	0.853	0.819	0.782	0.742
19	0.986	0.979	0.969	0.957	0.942	0.923	0.901	0.875	0.846	0.812
20	0.992	0.988	0.983	0.975	0.965	0.952	0.936	0.917	0.894	0.868
21	0.996	0.994	0.991	0.986	0.980	0.971	0.960	0.947	0.930	0.911
22	0.998	0.997	0.995	0.992	0.989	0.983	0.976	0.967	0.956	0.942
23	0.999	0.999	0.998	0.996	0.994	0.991	0.986	0.981	0.973	0.963
24	1.000	0.999	0.999	0.998	0.997	0.995	0.992	0.989	0.984	0.978
25	1.000	1.000	0.999	0.999	0.998	0.997	0.996	0.994	0.991	0.987
26	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.995	0.993
27	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.998	0.997	0.996
28	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.998
29	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999
30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
31	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
32	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
33	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
34	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
35	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

ที่มา : (Watson et al.,1990, pp. 942-947)

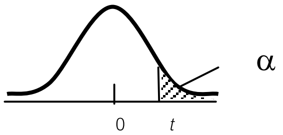
ตารางภาคผนวกที่ 5 พื้นที่ใต้เส้นโค้งปรกติมาตรฐาน



Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.10	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.20	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.30	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.40	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.50	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.60	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.70	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2793	.2823	.2852
0.80	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.90	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.00	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.10	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.20	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	3997	.4015
1.30	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.40	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.50	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.60	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.70	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.80	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.90	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.00	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.10	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.20	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.30	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.40	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.50	.4938	.4940	.4941	.4843	.4945	.4946	.4948	.4949	.4951	.4952
2.60	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.70	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.80	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.90	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.00	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.10	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.20	.4993	.4993	.4994	0494	.4994	.4994	.4994	.4995	.4995	.4995
3.30	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.40	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.50	.4998	.4997	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.60	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.70	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.80	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999

ที่มา : (Watson et al.,1990, p. Appendix C)

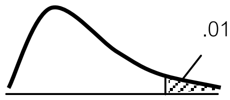
ตารางภาคผนวกที่ 6 ค่าวิกฤตการแจกแจงที



df	α				
	.10	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.323	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

ที่มา : (Watson et al.,1990, p. Appendix C)

ตารางภาคผนวกที่ 7 ค่าวิกฤตการแจกแจงเอฟ $\alpha = 0.01$



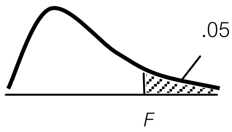
df_1 / df_2	1	2	3	4	5	6	7	8	9
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859	5928.4	5981.1	6022.5
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.364	99.388
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659
5	16.258	13.274	12.060	11.3920	12.967	10.672	10.456	10.289	10.158
6	13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.2460	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.2590	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.5610	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.0440	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7435	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8060	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3031	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.3990	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2987
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.8748	3.5558	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1186	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4073

ตารางภาคผนวกที่ 7 ค่าวิกฤตการแจกแจงเอฟ α = 0.01 (ต่อ)

df_1 / df_2	10	12	15	20	24	30	40	60	120	∞
1	6055.8	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313	6339.4	6365.9
2	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	10.0510	9.8883	9.7222	9.5526	9.4665	9.3793	9.2912	9.2020	9.1180	9.0204
6	7.8741	7.7183	7.5590	7.3958	7.3127	7.2285	7.1432	7.0567	6.9690	6.8800
7	6.6201	6.4691	6.3143	6.1554	6.0743	5.9920	5.9084	5.8236	5.7373	2.6950
8	5.8143	5.6667	5.5151	5.3591	5.2793	5.1981	5.1156	5.0316	4.9461	4.8588
9	5.2565	5.1114	4.9621	4.8080	4.7290	4.6486	4.5666	4.4831	4.3978	4.3105
10	4.8491	4.7059	4.5581	4.4054	4.3269	4.2469	4.1653	4.0819	3.9965	3.9090
11	4.5393	4.3974	4.2509	4.0990	4.0209	3.9411	3.8596	3.7761	3.6904	3.6024
12	4.2961	4.1553	4.0096	3.8584	3.7805	3.7008	3.6192	3.5355	3.4494	3.3608
13	4.1003	4.9603	3.8154	3.6646	3.5868	3.5070	3.4253	3.3413	3.2548	3.1654
14	3.9394	3.8001	3.6557	3.5052	3.4274	3.3476	3.2656	3.1813	3.0942	3.0040
15	3.8049	3.6662	3.5222	3.3719	3.2940	3.2141	3.1319	3.0471	2.9595	2.8684
16	3.6909	3.5527	3.4089	3.2587	3.1808	3.1007	3.0182	2.9330	2.8447	2.7528
17	3.5931	3.4552	3.3117	3.1615	3.0835	3.0032	2.9205	2.8480	2.7459	2.6530
18	3.5282	3.3706	3.2273	3.0771	2.9990	2.9185	2.8354	2.7493	2.6597	2.5660
19	3.4338	3.2965	3.1533	3.0031	2.9249	2.8442	2.7608	2.6742	2.5839	2.4893
20	3.6820	3.2311	3.0880	2.9377	2.8594	2.7785	2.6947	2.6077	2.5168	2.4212
21	3.3098	3.1730	3.0300	2.8796	2.8010	2.7200	2.6359	2.5484	2.4568	2.3603
22	3.3576	3.1209	2.9779	2.8274	2.7488	2.6675	2.5831	2.4951	2.4029	2.3055
23	3.3106	3.0740	2.9311	2.7805	2.7017	2.6202	2.5355	2.4471	2.3542	2.2558
24	3.1681	3.0316	2.8887	2.7380	2.6591	2.5773	2.4923	2.4035	2.3100	2.2107
25	3.1294	2.9931	2.8502	2.6993	2.6203	2.5383	2.4530	2.3637	2.2696	2.1694
26	3.0941	2.9578	2.8150	2.6640	2.5848	2.5026	2.4170	2.3273	2.2325	2.1315
27	3.0618	2.9256	2.7827	2.6316	2.5522	2.4699	2.3840	2.2938	2.1985	2.0965
28	3.0320	2.8959	2.7530	2.6017	2.5523	2.4397	2.3535	2.2629	2.1670	2.0642
29	3.0045	2.8685	2.7256	2.5742	2.4946	2.4118	2.3253	2.2344	2.1379	2.0342
30	2.9791	2.8431	2.7002	2.5487	2.4689	2.3860	2.9920	2.2079	2.1108	2.0062
40	2.0050	2.6648	2.5316	2.3689	2.2880	2.2034	2.1142	2.0194	1.9172	1.8047
60	2.6318	2.4961	2.3523	2.1978	2.1154	2.0285	1.9360	1.8363	1.7263	1.6006
120	2.4721	2.3363	2.1915	2.0346	1.9500	1.8600	1.7628	1.6557	1.5330	1.3805
∞	2.3209	2.1847	2.0385	1.8783	1.7908	1.6964	1.5923	1.4730	1.3246	1.0000

ที่มา : (Wonnacott & Wonnacott,1985, pp. 614-615)

ตารางภาคผนวกที่ 8 ค่าวิกฤตการแจกแจงเอฟ $\alpha = 0.05$



df_1 / df_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8868	8.8452	8.8123
4	7.7086	6.9443	6.5914	6.3883	6.2560	6.1631	6.0942	6.0410	5.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2066	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8378	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8626	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6601	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5763	2.5102	2.4563
19	4.3808	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3513	3.4928	3.0984	2.8661	2.7109	2.5994	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5757	2.4876	2.4205	2.3661
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9751	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1380	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2782	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0848	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2540	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2900	2.1750	2.0867	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

ตารางภาคผนวกที่ 8 ค่าวิกฤตการแจกแจงเอฟ α = 0.05 (ต่อ)

df_1 / df_2	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5844	8.572	8.5494	8.5264
4	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	2.5370	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	2.4935	2.4347	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	2.3779	2.3080	2.3410	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7840	1.7110
26	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

ที่มา : (Wonnacott & Wonnacott,1985, pp. 616-617)

ตารางภาคผนวกที่ 9 ค่าเอสเอสอาร์ $\alpha = 0.01$

	p								
df	2	3	4	5	6	7	8	9	10
1	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03	90.03
2	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04	14.04
3	8.261	8.321	8.321	8.321	8.321	8.321	8.321	8.321	8.321
4	6.512	6.677	6.740	6.756	6.756	6.756	6.756	6.756	6.756
5	5.702	5.898	5.989	6.040	6.065	5.074	6.074	6.074	6.074
6	5.243	5.439	5.549	5.614	5.655	5.680	5.694	5.701	5.703
7	4.949	5.145	5.260	5.334	5.383	5.416	5.439	5.454	5.454
8	4.746	4.939	5.057	5.135	5.189	5.227	5.256	5.276	5.291
9	4.596	4.787	4.906	4.986	5.043	5.086	5.118	5.142	5.160
10	4.482	4.671	4.790	4.871	4.931	4.975	5.010	5.037	5.058
11	4.392	4.579	4.697	4.780	4.841	4.887	4.924	4.952	4.975
12	4.320	4.504	4.622	4.706	4.767	4.815	4.852	4.883	4.907
13	4.260	4.442	4.560	4.644	4.706	4.755	4.793	4.824	4.850
14	4.210	4.391	4.508	4.591	4.654	4.704	4.743	4.775	4.802
15	4.168	4.347	4.463	4.547	4.610	4.660	4.700	4.733	4.760
16	4.131	4.309	4.425	4.509	4.572	4.662	4.663	4.696	4.724
17	4.099	4.275	4.391	4.475	4.539	4.589	4.630	4.664	4.693
18	4.071	4.246	4.362	4.445	4.509	4.560	4.601	4.635	4.664
19	4.046	4.230	4.335	4.419	4.483	4.534	4.575	4.610	4.639
20	4.024	4.197	4.312	4.395	4.459	4.510	4.552	4.587	4.617
24	3.956	4.126	4.239	4.322	4.386	4.437	4.480	4.516	4.546
30	3.889	4.056	4.168	4.250	4.314	4.366	4.409	4.445	4.477
40	3.825	3.988	4.098	4.180	4.244	4.296	4.339	4.376	4.408
60	3.762	3.922	4.031	4.111	4.174	4.226	4.270	4.307	4.340
120	3.702	3.858	3.965	4.044	4.107	4.158	4.202	4.239	4.272
∞	3.643	3.796	3.900	3.978	4.040	4.091	4.135	4.172	4.205

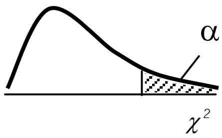
ที่มา : (Walpole,1974, p. 322)

ตารางภาคผนวกที่ 10 ค่าเอสเอสอาร์ $\alpha = 0.05$

	p								
df	2	3	4	5	6	7	8	9	10
1	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.501	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516
4	3.927	4.013	4.033	4.033	4.033	4.033	4.033	4.033	4.033
5	3.635	3.749	3.797	3.814	3.814	3.814	3.814	3.814	3.814
6	3.461	3.587	3.649	3.680	3.694	3.697	3.697	3.697	3.697
7	3.344	3.477	3.548	3.588	3.611	3.622	3.626	3.626	3.626
8	3.261	3.399	3.475	3.521	3.549	3.566	3.575	3.579	3.579
9	3.199	3.339	3.420	3.470	3.502	3.523	3.536	3.544	3.547
10	3.151	3.293	3.376	3.430	3.465	3.489	3.505	3.516	3.522
11	3.113	3.256	3.342	3.397	3.435	3.462	3.480	3.493	3.501
12	3.082	3.225	3.313	3.370	3.410	3.439	3.459	3.474	3.484
13	3.055	3.200	3.289	3.348	3.389	3.419	3.442	3.458	3.470
14	3.033	3.178	3.268	3.329	3.372	3.403	3.426	3.444	3.457
15	3.014	3.160	3.250	3.312	3.356	3.389	3.413	3.432	3.446
16	2.998	3.144	3.235	3.298	3.343	3.376	3.402	3.422	3.437
17	2.984	3.130	3.220	3.285	3.330	3.366	3.392	3.412	3.429
18	2.971	3.118	3.210	3.274	3.321	3.356	3.383	3.405	3.421
19	2.960	3.107	3.199	3.264	3.311	3.347	3.375	3.397	3.415
20	2.950	3.097	3.190	3.255	3.303	3.339	3.368	3.391	3.409
24	2.919	3.066	3.160	3.226	3.276	3.315	3.345	3.370	3.390
30	2.888	3.035	3.131	3.199	3.250	3.290	3.322	3.349	3.371
40	2.838	3.006	3.102	3.171	3.224	3.266	3.300	3.328	3.352
60	2.829	2.976	3.073	3.143	3.198	3.241	3.277	3.307	3.333
120	2.800	2.947	3.045	3.116	3.172	3.217	3.254	3.287	3.314
∞	2.772	2.918	3.017	3.089	3.146	3.193	3.232	3.265	3.294

ที่มา : (Walpole,1974, p. 323)

ตารางภาคผนวกที่ 11 ค่าวิกฤตการแจกแจงไคกำลังสอง



	α							
df	0.995	0.990	0.975	0.950	0.050	0.025	0.010	.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^3393	3.841	5.024	6.635	7.879
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.448	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672

ที่มา : (Watson et al.,1990, p. Appendix C)

ตารางภาคผนวกที่ 12 ค่าวิกฤติการทดสอบของครอกครัน $\alpha = 0.01$

k \ n	2	3	4	5	6	7	8	9	10	11	17	37	145	∞
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5336	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8826	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.8040	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1259
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1811	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1567	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1508	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.2412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0862	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.1033	0.0957	0.0898	0.0853	0.0816	0.0668	0.0503	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.2151	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	0.0178	0.0215	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ที่มา : (Kenji.,1999, p. 184)

ตารางภาคผนวกที่ 13 ค่าวิกฤติการทดสอบของครอกครัน $\alpha = 0.05$

k \ n	2	3	4	5	6	7	8	9	10	11	17	37	145	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.0333
4	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.9800	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.3612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0063
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ที่มา : (Kenji.,1999, p. 185)

ตารางภาคผนวกที่ 14 ค่าวิกฤตการทดสอบของฮาร์ตลีย์ $1-\alpha = 0.95$

	r										
n	2	3	4	5	6	7	8	9	10	11	12
3	39.0	87.5	142	202	266	333	403	475	550	626	704
4	15.40	27.80	39.20	50.70	62.00	72.90	83.50	93.90	104	114	124
5	9.60	15.50	20.60	25.20	29.50	33.60	37.50	41.10	44.60	48.00	51.40
6	7.15	10.80	13.70	16.30	18.70	20.80	22.90	24.70	26.50	28.20	29.90
7	5.82	8.38	10.40	12.10	13.70	15.00	16.30	17.50	18.60	19.70	20.70
8	4.99	6.94	8.44	9.70	108.00	11.80	12.70	13.50	14.30	15.10	15.80
9	4.43	6.00	7.18	8.12	9.03	9.78	10.50	11.10	11.70	12.20	12.70
10	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.30	10.70
11	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
13	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
16	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
21	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
31	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
61	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

ที่มา : (Kenji.,1999, p.182)

ตารางภาคผนวกที่ 15 ค่าวิกฤตการทดสอบของฮาร์ตลีย์ $1-\alpha = 0.99$

	r										
n	2	3	4	5	6	7	8	9	10	11	12
3	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
4	47.5	85	120	151	184	216	249	281	310	337	361
5	23.2	37	49	59	69	79	89	97	106	113	120
6	14.9	22	28	33	38	42	46	50	54	57	60
7	11.1	15.5	19.1	22	25	27	30	32	34	36	37
8	8.89	12.1	14.5	16.5	18.4	20	22	23	24	26	27
9	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
10	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
11	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
13	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
16	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
21	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
31	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
61	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
∞	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

ที่มา : (Kenji.,1999, p. 181)

ตารางภาคผนวกที่ 16 ค่าวิกฤตสัมประสิทธ์สหสัมพันธ์ลำดับสเปียร์แมน

	r			
n	α =0.050	α =0.025	α =0.010	α =0.005
5	.900	---	---	---
6	.829	.886	.943	---
7	.714	.876	.893	---
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.745	.894
11	.523	.623	.736	.818
12	.497	.591	.703	.780
13	.475	.566	.673	.745
14	.457	.545	.646	.716
15	.411	.525	.623	.689
16	.425	.507	.601	.666
17	.412	.490	.582	.645
18	.399	.476	.564	.625
19	.388	.462	.549	.608
20	.377	.450	.534	.591
21	.368	.438	.521	.576
22	.359	.428	.508	.562
23	.351	.418	.496	.549
24	.343	.409	.485	.537
25	.336	.400	.475	.526
26	.329	.392	.465	.515
27	.323	.385	.456	.505
28	.317	.377	.448	.496
29	.311	.370	.440	.487
30	.305	.364	.432	.478

ที่มา : (Keller & Warrack,2000, p. B-22 Appendix B)

เฉลยคำถามทบทวนบทที่ 1

1. เป็นข้อมูลเชิงสถิติ เพราะมีการบันทึกรายจ่ายเป็นตัวเลข เป็นรายวัน เช่น วันที่1,2,... เป็น(ตัวแปร) 50,85,120.5,...บาท

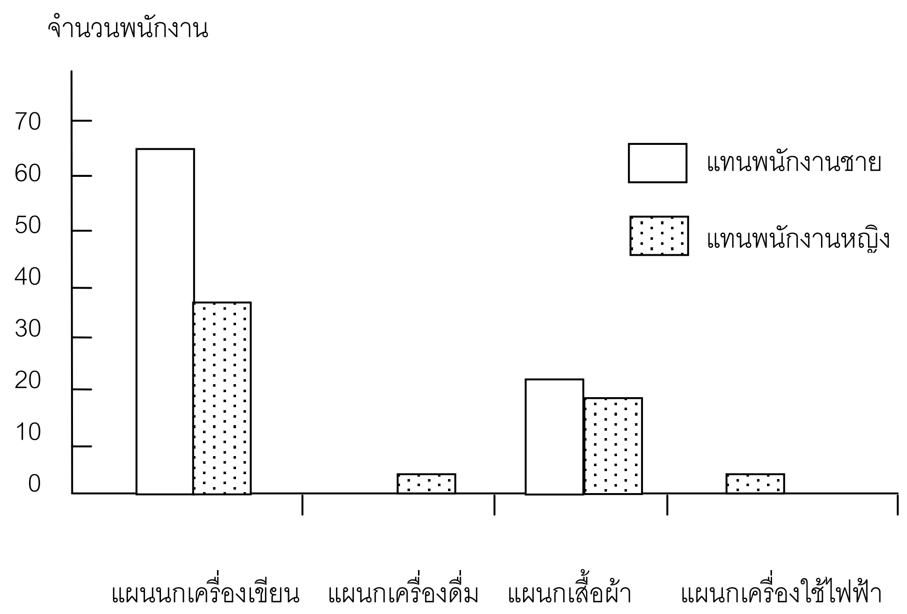
2. ก. ตัวแปรวิยุติ เช่นพนักงานคนที่1,2,...และ 20 มีจำนวนวันลา(ตัวแปร) 5,0,... และ 2 วัน

ข. ตัวแปรต่อเนื่อง เช่น เครื่องจักรหยุด วันที่ 1,2,...,และ 30 เป็นเวลา(ตัวแปร) 0.5,0.2,... และ12.5 ชั่วโมง

ค. ตัวแปรวิยุติ จำนวนลูกค้าที่เข้าใช้บริการ วันที่ 1,2, …และ 30 เป็น(ตัวแปร) 25,30, 32,... และ18 คน

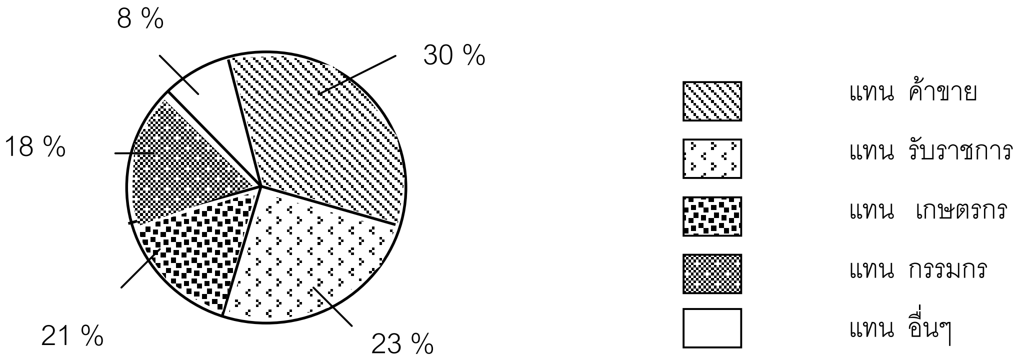
ง. ตัวแปรต่อเนื่อง เช่น น้ำหนักแรกเกิดของทารก คนที่ 1,2,...และ 500 เป็น (ตัวแปร) 2.85,3.25,...และ 3.50 กิโลกรัม

3. การนำเสนอข้อมูลจากตารางที่ 1.1 ด้วยกราฟแท่ง



ภาพภาคผนวกที่ 1 กราฟแท่งของข้อมูลจากตารางที่ 1.1

4. การนำเสนอข้อมูลจากตารางที่ 1.3 ด้วยกราฟวงกลม



ภาพภาคผนวกที่ ผ. 2 กราฟวงกลมของข้อมูลจากตารางที่ 1.3

5. ก. ช่วงคะแนนที่มีความถี่ 6 คือคะแนนช่วง 50-54

ข. คะแนนระหว่าง 39.5 – 59.5 มีความถี่เท่ากับ 16

ค. คะแนนช่วง 50-54 มีความถี่สะสม 8

ง. ช่วงคะแนนใดที่ไม่มีผู้สอบได้เลยคือ 45-49

6. แสดงการแจกแจงความถี่ ความถี่สะสมและความถี่สะสมสัมพัทธ์ของคะแนนสอบวิชาสถิติธุรกิจ

ตารางภาคผนวกที่ ผ. 17 การแจกแจงความถี่ ความถี่สะสมและความถี่สะสมสัมพัทธ์ของคะแนนสอบวิชาสถิติธุรกิจ

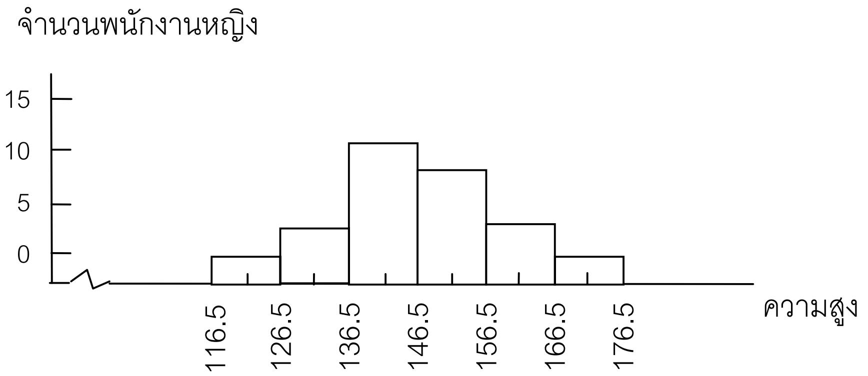
ช่วงคะแนน	ความถี่(f)	ความถี่สะสม(cf)	ความถี่สัมพันธ์(rf)
10-19	2	2	0.02
20-29	3	5	0.03
30-39	14	19	0.14
40-49	28	47	0.28
50-59	23	70	0.23
60-69	14	84	0.14
70-79	10	94	0.10
80-89	4	98	0.04
90-99	2	100	0.02
รวม	n =100		1

7. ก. สร้างการแจกแจงความถี่ ความถี่สะสมและความถี่สัมพัทธ์ของความสูงของพนักงาน

ตารางภาคผนวกที่ 18 การแจกแจงความถี่ ความถี่สะสมและความถี่สัมพัทธ์ของความสูงของพนักงานหญิง

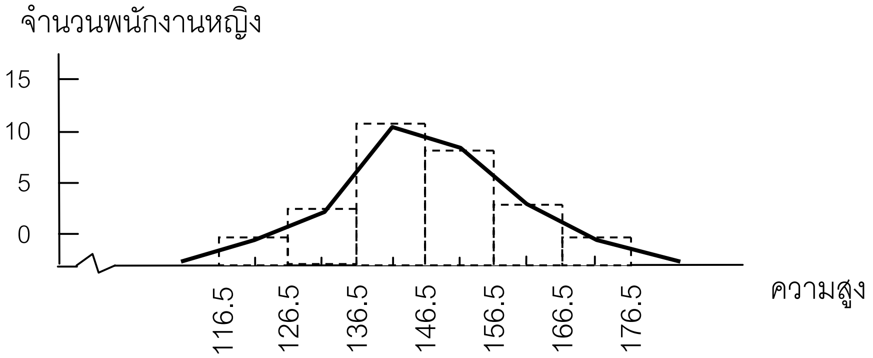
ความสูง(*cm*)	ความถี่(*f*)	ความถี่สัมพันธ์(*rf*)
117-126	3	0.075
127-136	6	0.15
137-146	12	0.30
147-156	10	0.25
157-166	6	0.15
167-176	3	0.075
รวม	n = 40	1.00

ข. สร้างฮิทโทแกรม



ภาพภาคผนวกที่ 3 ฮิทโทแกรมของข้อมูลความสูงของพนักงานหญิง

ค. สร้างรูปหลายเหลี่ยมความถี่ข้อมูล



ภาพภาคผนวกที่ 4 รูปหลายเหลี่ยมความถี่ของข้อมูลความสูงของพนักงานหญิง

เฉลยคำถามทบทวนบทที่ 2

1. ก. $\sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \ldots + X_n$

ข. $\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2$

ค. $\sum_{i=1}^{n} (X_i + Y_i) = (X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3) + \ldots + (X_n + Y_n)$

ง. $\sum_{i=1}^{n} (X_i - a) = (X_1 - a) + (X_2 - a) + (X_3 - a) + \ldots + (X_n - a)$

จ. $\sum_{i=1}^{n} (X_i Y_i) = (X_1 Y_1) + (X_2 Y_2) + (X_3 Y_3) + \ldots + (X_n Y_n)$

ฉ. $\sum_{i=1}^{n} (X_i f_i) = (X_1 f_1) + (X_2 f_2) + (X_3 f_3) + \ldots + (X_n f_n)$

2. ก. $\sum_{i=1}^{20} Z_i = Z_1 + Z_2 + Z_3 + \ldots + Z_{20}$

ข. $\sum_{i=1}^{6} (X_i f_i) = (X_1 f_1) + (X_2 f_2) + (X_3 f_3) + \ldots + (X_6 f_6)$

ค. $\sum_{i=2}^{7} X_i^3 Y_i = X_2^3 Y_2 + X_3^3 Y_3 + \ldots.. + X_7^3 Y_7$

ง. $\sum_{i=1}^{m} (X_i - Y_i) = (X_1 - Y_1) + (X_2 - Y_2) + (X_3 - Y_3) + \ldots + (X_m - Y_m)$

3. $\sum_{i=1}^{n} (X_i - k) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} k$

$= \sum_{i=1}^{n} X_i - nk$

4. ไม่เท่ากัน เพราะ

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2$$

$$(\sum_{i=1}^{n} X_i)^2 = (X_1 + X_2 + X_3 + \ldots + X_n)^2$$

ดังนั้นจะได้ว่า

$$(\sum_{i=1}^{n} X_i)^2 \neq \sum_{i=1}^{n} X_i^2$$

5. จาก $(X_i - k)^2 = X_i^2 - 2X_i k + k^2$

$$\sum_{i=1}^{n} (X_i - k)^2 = \sum_{i=1}^{n} (X_i^2 - 2X_i k + k^2)$$

$$= \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} 2X_i k + \sum_{i=1}^{n} k^2$$

$$= \sum_{i=1}^{n} X_i^2 - 2k\sum_{i=1}^{n} X_i + nk^2$$

6. จาก $\mu = \dfrac{X_1 + X_2 + X_3 + X_4 \ldots\ldots\ldots + X_n}{n}$

$$55 = \frac{X_1 + X_2 + X_3 + X_4 \ldots\ldots\ldots + X_n}{36}$$

ได้ค่า $X_1 + X_2 + X_3 + X_4 \ldots\ldots\ldots + X_n$

$= (55)(36)$

$= 1{,}980$

นั่นคือผลรวมของข้อมูลชุดนี้คือ 1,980

7. ก. หาค่ามัชฌิมเลขคณิตของข้อมูลประชากร

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{10}(61.5) = \mathbf{6.15}$$

หาค่ามัธยฐานของข้อมูลประชากรคือข้อมูลลำดับที่ 5 และ 6

เรียงลำดับข้อมูล	1	2	3	4	5	6	7	8	9	10
ข้อมูล	3.3	4.1	5.1	5.4	6.2	6.2	6.2	6.7	8.7	9.6

$$Med = 6.2$$

หาค่าฐานนิยมของข้อมูลประชากรคือข้อมูลที่พบบ่อยครั้งที่สุด

$$Mo = 6.2$$

ข. ให้ $\overline{Y}$ ค่ามัชฌิมเลขคณิตของข้อมูลประชากรทุกตัวคูณด้วย 10

จาก $\overline{Y} = b\mu \;;\; b = 10 \;,\; \mu = \mathbf{6.15}$

$$\overline{Y} = 10(6.15) = 61.5$$

ค. ให้ $\overline{Y}$ ค่ามัชฌิมเลขคณิตของข้อมูลประชากรทุกตัวบวกด้วย 20

จาก $\overline{Y} = a + \mu \;;\; a = 20 \;,\; \mu = \mathbf{6.15}$

$$\overline{Y} = 20 + 6.15 = 26.15$$

8. ก. จาก log GM. $= \frac{1}{n}\, log\, (X_1 \bullet X_2 \bullet X_3 \bullet X_3 \ldots \bullet X_n)$

$$= \frac{1}{8}\,(\,log\;50 + log\;55 + log\,60 + log\;65 +$$
$$log\,70 + log\;75 + log\;80 + log\;85\,)$$

$$= \frac{1}{8}\,(\,0.6990 + 0.7404 + 0.7782 + 0.8129 + 0.8751 +$$
$$0.9031 + 0.9294\,) = 0.7135$$

พบว่าที่ $log\, N = 0.7135 \;;\; N = 51.7$

$GM \quad = \quad 51.7$

ข. จาก $HM. = \dfrac{n}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\dots\dots\dots+\frac{1}{X_n}}$

$$HM. = \frac{8}{\frac{1}{50}+\frac{1}{55}+\frac{1}{60}+\frac{1}{65}+\frac{1}{70}+\frac{1}{75}+\frac{1}{80}+\frac{1}{85}}$$

$$= \frac{8}{0.1221} \qquad = \quad 65.51$$

9. จัดตำแหน่งข้อมูล

ตำแหน่งข้อมูล	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ข้อมูล	12	13	13	13	14	14	15	15	15	16	17	17	18	19	19	20

หา P_{50}

ตำแหน่ง 100 ตรงกับข้อมูลจริงที่ตำแหน่ง 16

ตำแหน่ง 50 ตรงกับข้อมูลจริงที่ตำแหน่ง $\dfrac{(16)(50)}{100} = 8$

พบว่าข้อมูลจริงตำแหน่งที่ 8 คือ 15 ดังนั้น P_{50} ของข้อมูลนี้เท่ากับ 15

หา D_5

ตำแหน่ง 10 ตรงกับข้อมูลจริงที่ตำแหน่ง 16

ตำแหน่ง 5 ตรงกับข้อมูลจริงที่ตำแหน่ง $\dfrac{(16)(5)}{10} = 8$

พบว่าข้อมูลจริงตำแหน่งที่ 8 คือ 15 ดังนั้น D_5 ของข้อมูลนี้เท่ากับ 15

หา Q_2

ตำแหน่ง 4 ตรงกับข้อมูลจริงที่ตำแหน่ง 16

ตำแหน่ง 2 ตรงกับข้อมูลจริงที่ตำแหน่ง $\dfrac{(16)(2)}{4} = 8$

พบว่าข้อมูลจริงตำแหน่งที่ 8 คือ 15 ดังนั้น Q_2 ของข้อมูลนี้เท่ากับ 15

10. ก. หาค่าเฉลี่ยของอายุพนักงาน

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

$$= \frac{180}{7} = 25.714$$

ดังนั้น ค่าเฉลี่ยของอายุพนักงาน เท่ากับ 25.71 ปี

ข. หาค่ามัธยฐานของอายุพนักงาน

ลำดับอายุ	1	2	3	4	5	6	7
อายุ	18	18	19	19	20	21	65

ดังนั้น มัธยฐานของอายุพนักงาน เท่ากับ 19 ปี

ค. ค่าฐานนิยมของอายุพนักงาน = 18 , 19 ปี

ง. หาแนวโน้มสู่ส่วนกลางที่เหมาะสมโดยตัด 65 ออก

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

$$= 19.16$$

ดังนั้น ค่าเฉลี่ยของอายุพนักงาน 19 ปี

11. ก. หาค่าเฉลี่ยจำนวนรถที่มีของนักธุรกิจ

$$\mu = \frac{\sum_{j=1}^{k} f_j X_j}{\sum_{j=1}^{k} f_j}$$

$$= \frac{0(3) + 1(10) + 2(4) + 2(3) + 1(4)}{20}$$

$$\mu = \frac{28}{20} = 1.4 \text{ คัน}$$

ข. หาค่ามัธยฐานจำนวนรถที่มีของนักธุรกิจ

$$Med = 1 \text{ คัน}$$

ค. หาค่าฐานนิยมของที่มีนักธุรกิจ

$$Mo = 1 \text{ คัน}$$

เฉลยคำถามทบทวนบทที่ 3

1. หาค่าเฉลี่ยของประชากร

$$\mu = \frac{40+45+50+55+60+65+70+75+80+85+90+95}{12}$$

$$= \frac{810}{12}$$

$$= 67.5$$

ก. $$\sum_{i=1}^{n}|X_i - \mu| = |40 - 67.5| + |45 - 67.5| + |50 - 67.5| + |55 - 67.5| +$$

$$|60 - 67.5| + |65 - 67.5| + |70 - 67.5| + |75 - 67.5| +$$

$$|80 - 67.5| + |85 - 67.5| + |90 - 67.5| + |95 - 67.5|$$

$$= 180$$

ข. $\sum_{i=1}^{n}(X_i - \mu)^2$ = $(40 - 67.5)^2 + (45 - 67.5)^2 + (50 - 67.5)^2 + (55 - 67.5)^2 +$

$(60 - 67.5)^2 + (65 - 67.5)^2 + (70 - 67.5)^2 + (75 - 67.5)^2 +$

$(80 - 67.5)^2 + (85 - 67.5)^2 + (90 - 67.5)^2 + (95 - 67.5)^2$

= 3,575

ค. R = 95.5 – 39.5 = 56

ง. Q = $Q_3 - Q_1$

Q_3 = 80

Q_1 = 50

Q = 80 - 50 = 30

จ. $V(X) = \dfrac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}$

$= \dfrac{3575}{12} = 297.92$

ฉ. $\sigma_x = \sqrt{V(X)}$

$= \sqrt{297.92} = 17.26$

2. หาค่าเฉลี่ยประชากร

$\mu = 1.9$

ก. หาความแปรปรวนประชากร

$$V(X) = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n} = \frac{0.04}{5} = 0.008$$

ข. หาความแปรปรวนประชากรเมื่อคูณทุกตัวด้วย 10^3

จาก $V(aX) = a^2V(X) = (10^3)^2(0.008) = (0.008)(10^6)$

ค. หาความแปรปรวนประชากรเมื่อบวกทุกตัวด้วย 1.5

$$V(a+X) = V(X) = 0.008$$

3. หาค่าเฉลี่ยตัวอย่าง

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = 18.78$$

ก. หาความแปรปรวนกลุ่มตัวอย่าง

$$V(X) = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{35.56}{8} = 4.445$$

ข. หาความแปรปรวนกลุ่มตัวอย่างเมื่อคูณทุกตัวด้วย 10^3 ดังนั้น $a = 10^3$

$$V(aX) = a^2V(X) = (10^3)^2(4.445) = (4.445)(10^6)$$

ค. หาความแปรปรวนกลุ่มตัวอย่างเมื่อบวกทุกตัวด้วย 1.5 ดังนั้น $a = 1.5$

$$V(a+X) = V(X) = 4.445$$

4. ให้ CV_1 เป็น สัมประสิทธิ์การแปรผันราคาพริก

CV_2 เป็น สัมประสิทธิ์การแปรผันราคาหอม

$$CV_1 = \frac{\sigma_1}{\mu_1} = \frac{20}{80} = 25\ \%$$

$$CV_2 = \frac{\sigma_2}{\mu_2} = \frac{2}{18} = 11.11\ \%$$

จึงควรเลือกปลูกหอมแบ่งเพราะค่าสัมประสิทธิ์การแปรผันของราคาต่ำกว่า

5. หาค่าสัมประสิทธิ์การแปรผันของพนักงานทั้งสอง

โรงงาน ก. $CV_1 = \dfrac{\sigma_1}{\mu_1} = \dfrac{5{,}000}{30{,}000} = 16.67\ \%$

โรงงาน ข. $CV_2 = \dfrac{\sigma_2}{\mu_2} = \dfrac{700}{50{,}000} = 1.4\ \%$

สรุปได้ว่ารายได้ของพนักงานโรงงาน ก. กระจายมากกว่ารายได้ของพนักงานโรงงาน ข.

6. หาค่าเฉลี่ยของข้อมูล $\mu = 1.9$

ก. หาโมเมนต์ 2, 3 และ 4 ของข้อมูล

ตารางภาคผนวกที่ 19 การคำนวณค่า d_i, d_i^2, d_i^3, d_i^4 ของข้อมูลของอายุของนักเรียนกลุ่มหนึ่งจำนวน 35 คน

X	f	fX	d_i	d_i^2	d_i^3	d_i^4
15	1	15	5.37	28.84	154.85	831.57
14	1	14	4.37	19.10	83.45	364.69
13	2	26	6.74	90.86	612.36	4,127.33
12	2	24	4.74	44.94	212.99	1,009.59
11	5	55	6.85	234.61	1,607.10	11,008.61
10	6	60	2.22	29.57	65.65	145.73
9	9	81	-5.67	289.34	-1,640.56	9,301.97
8	3	24	-4.89	71.74	-350.79	1,715.37
7	3	21	-7.89	186.76	-1,473.51	11,625.97
6	2	12	-7.26	105.42	-75.31	5,556.18
5	1	5	-4.63	21.44	-99.25	459.54
รวม	35	337	0	1,122.59	-1,593.02	46,146.54

$$M_2 = \frac{\sum_{i=1}^{n} d_i^2}{n} = \frac{1,122.59}{35} = 32.07$$

$$M_3 = \frac{\sum_{i=1}^{n} d_i^3}{n} = \frac{-1,593.02}{35} = -45.51$$

$$M_4 = \frac{\sum_{i=1}^{n} d_i^4}{n} = \frac{46,146.54}{3} = 1,318,047$$

ข. หาค่าความเบ้และความโด่งของข้อมูล

$$Sk = \frac{M_3}{M_2\sqrt{M_2}} = \frac{-45.53}{32.07\sqrt{32.07}} = -0.250$$

$$Ku = \frac{M_4}{M_2^2} = \frac{1,318.448}{(32.07)^2} = 1.282$$

7. หาค่าเฉลี่ยของผลการสอบ

$$\mu = 33.74$$

ก. หาค่าเบี่ยงเบนมาตรฐานของผลการสอบ จากข้อมูลในตารางภาคผนวกที่ 20 จะได้

$$M_2 = \frac{\sum_{i=1}^{n} d_i^2}{n}$$

$$= \frac{48,139.46}{50} = 962.79$$

ดังนั้นค่าเบี่ยงเบนมาตรฐานของผลการสอบเท่ากับ $\sqrt{962.79} = 31.02$

ตารางภาคผนวกที่ 20 การคำนวณค่า d_i, d_i^2 ,d_i^3 ,d_i^4 ของข้อมูลของผลการสอบปลายภาควิชาสถิติของนักศึกษาจำนวน 50 คน

X	*f*	*fX*	d_i	d_i^2	d_i^3	d_i^4
45	2	90	22.52	1,014.30	22,842.05	514,403.06
44	3	132	30.78	2,842.23	87,483.69	2,692,748.03
43	1	43	9.26	85.75	794.02	7,352.65
42	5	200	31.30	4,898.45	153,321.49	4,798,962.48
39	4	156	21.04	1,770.73	37,256.08	783,868.00
35	10	350	12.60	1,587.60	20,003.76	252,047.38
30	5	150	-18.70	1,748.42	-32,696.02	611,415.48
29	10	290	-47.40	22,467.60	-1,064,964.24	50,479,304.68
28	7	196	-40.18	11,301.03	-454,075.26	18,244,743.82
27	2	54	-13.48	363.42	-4,898.91	66,037.34
26	1	26	-7.74	59.91	-463.68	3,588.92
รวม	50	1,687	0.00	48,139.46	-1,253,397.01	78,454,472.12

ข. หาโมเมนต์ที่ 2, 3 และ 4 ของผลการสอบ

$$M_2 = 962.79$$

$$M_3 = \frac{\sum_{i=1}^{n} d_i^3}{n}$$

$$= \frac{-1,253,397.01}{50} = -24,707.94$$

$$M_4 = \frac{\sum_{i=1}^{n} d_i^4}{n}$$

$$= \frac{78,454,472.12}{50} = 1,569,089.44$$

ค. หาค่าความเบ้และค่าความโด่งของผลการสอบ

$$Sk = \frac{M_3}{M_2\sqrt{M_2}}$$

$$= \frac{-24{,}707.94}{962.79\sqrt{962.79}}$$

$$= -0.83$$

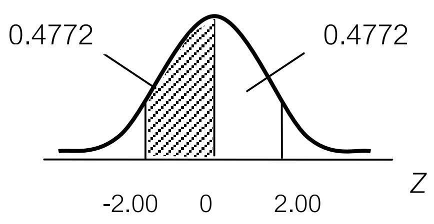
$$Ku = \frac{M_4}{M_2^2}$$

$$= \frac{1{,}569{,}089.44}{(962.79)^2}$$

$$= 1.69$$

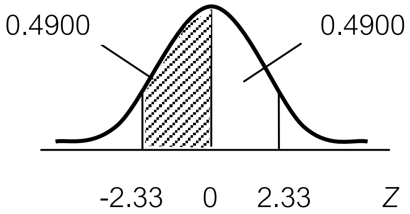
เฉลยคำถามทบทวนบทที่ 4

1. ประชากรแจกแจงปรกติร้อยละ 95.44 ของประชากร มีค่าอยู่ในช่วง $-2.00 \leq Z \leq +2.00$ ดังแสดงในภาพภาคผนวกที่ 5



ภาพภาคผนวกที่ 5 ร้อยละ 95.44 ของประชากรมีค่าอยู่ช่วง $-2.00 \leq Z \leq +2.00$

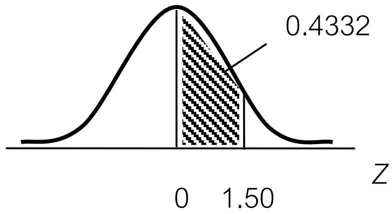
2. ประชากรแจกแจงปรกติร้อยละ 98 ของประชากร มีค่าอยู่ในช่วง $-2.33 \leq Z \leq 2.33$ ดังแสดงในภาพภาคผนวกที่ 6



ภาพภาคผนวกที่ 6 ร้อยละ 98 ของประชากรมีค่าอยู่ในช่วง
$-2.33 \leq Z \leq 2.33$

3. เป็นการแจกแจงแบบปกติเพราะเมื่อนำเฉพาะพนักงานที่มีน้ำหนักมาก ๆ มาแจกแจง จำนวน 100 คน จะแจกแจงปรกติเช่นเดียวกัน แต่ค่าเฉลี่ยจะสูงกว่าการแจกแจงเดิม

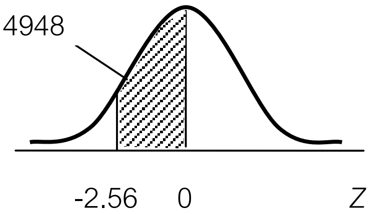
4. ก. พื้นที่ใต้เส้นโค้งปรกติระหว่างค่าเฉลี่ยกับ $Z = 1.5$ ดังแสดงในภาพภาคผนวกที่ 7



ภาพภาคผนวกที่ 7 พื้นที่ใต้เส้นโค้งปรกติระหว่าง
ค่าเฉลี่ยกับ $Z = 1.5$

$$P(0 \leq Z \leq 1.5) = 0.4332$$

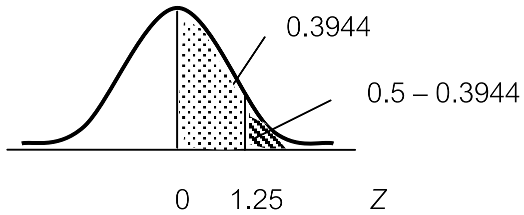
ข. พื้นที่ใต้เส้นโค้งปรกติระหว่างค่าเฉลี่ยกับ $Z = -2.56$ ดังแสดงในภาพภาคผนวกที่ 8



ภาพภาคผนวกที่ 8 พื้นที่ใต้เส้นโค้งปรกติระหว่าง ค่าเฉลี่ยกับ Z = -2.56

$$P(-2.56 \leq Z \leq 0) = 0.4948$$

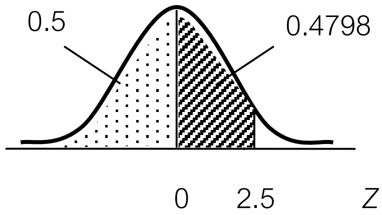
ค. พื้นที่ใต้เส้นโค้งปรกติเมื่อ $Z > 1.25$ ดังแสดงในภาพภาคผนวกที่ 9



ภาพภาคผนวกที่ 9 พื้นที่ใต้เส้นโค้งปรกติเมื่อ $Z > 1.25$

$$P(Z > 1.25) = 0.5 - 0.3944 = 0.1056$$

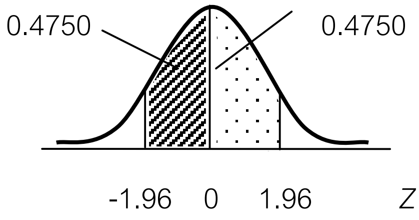
ง. พื้นที่ใต้เส้นโค้งปรกติเมื่อ $Z < 2.50$ ดังแสดงในภาพภาคผนวกที่ 10



ภาพภาคผนวกที่ 10 พื้นที่ใต้เส้นโค้งปรกติเมื่อ $Z < 2.50$

$$P(Z < 2.50) = 0.5 + 0.4798 = 0.9798$$

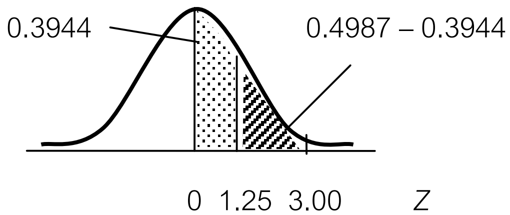
จ. พื้นที่ใต้เส้นโค้งปรกติเมื่อ Z = ± 1.96 ดังแสดงในภาพภาคผนวกที่ 11



ภาพภาคผนวกที่ 11 พื้นที่ใต้เส้นโค้งปรกติเมื่อ $Z = \pm 1.96$

$$P\left(-1.96 \leq P \leq 1.96\right) = 0.4750 + 0.4750 = 0.95$$

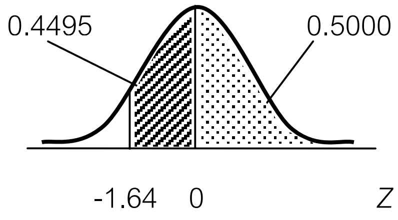
ฉ. พื้นที่ใต้เส้นโค้งปรกติระหว่าง Z = 1.25 กับ Z = 3.00 ดังแสดงในภาพภาคผนวกที่ 12



ภาพภาคผนวกที่ 12 พื้นที่ใต้เส้นโค้งปรกติระหว่าง
$Z = 1.25$ กับ $Z = 3.00$

$$P\left(1.25 \leq Z \leq 3.00\right) = 0.4987 - 0.3944 = 0.1043$$

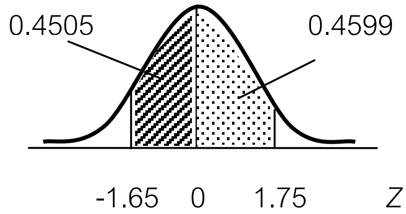
ช. พื้นที่เส้นใต้โค้งปรกติเมื่อ Z > -1.64 ดังแสดงในภาพภาคผนวกที่ 13



ภาพภาคผนวกที่ 13 พื้นที่ใต้เส้นโค้งปรกติเมื่อ Z > -1.64

$P(Z > -1.64) \;=\; 0.5 + 0.4495 = 0.9495$

ซ. พื้นที่ใต้เส้นโค้งปรกติระหว่าง $Z = -1.65$ กับ $Z = 1.75$

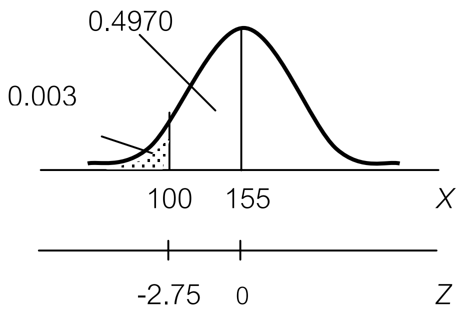


ภาพภาคผนวกที่ 14 พื้นที่ใต้เส้นโค้งปรกติระหว่าง $Z = -1.65$ กับ $Z = 1.75$

$$P(-1.65 \leq Z \leq 1.75) \;=\; 0.4505 + 0.4599 \;=\; 0.9104$$

5. ให้ X เป็นรายได้ต่อวันของพนักงาน

ก. พนักงานมีรายได้ตั้งแต่ 100 บาทลงมา



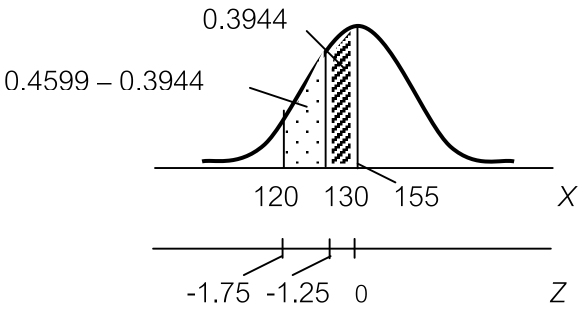
ภาพภาคผนวกที่ 15 พื้นที่ใต้เส้นโค้งปรกติระหว่าง $Z = -2.75$ กับ $Z = 0$

$$P\left(\frac{X_i - \mu}{\sigma} \leq Z \leq \frac{100 - 155}{20}\right)$$

$$P(Z \leq -2.75) \;=\; 0.5 - 0.4970 \;=\; 0.003$$

ดังนั้นจำนวนพนักงานที่มีรายได้ตั้งแต่ 100 บาท ลงมามีจำนวน = 2,000 (0.003) = 6 คน

ข. พนักงานมีรายได้ระหว่าง 120 บาทถึง 130 บาท



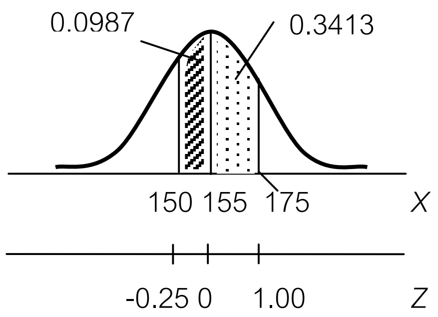
ภาพภาคผนวกที่ 16 พื้นที่ใต้เส้นโค้งปรกติระหว่าง Z = -1.75 กับ Z = -1.25

$$P\left(\frac{120 - 155}{20} \leq Z \leq \frac{100 - 155}{20}\right)$$

$$P(-1.75 \leq Z \leq -1.25) \quad = \quad 0.4599 - 0.3944 \quad = \quad 0.0655$$

จำนวนพนักงานที่มีรายได้ระหว่าง 120 ถึง 130 บาทมีจำนวน = 2,000(0.0655) = 131 คน

ค. พนักงานมีรายได้ระหว่าง 150 ถึง 175 บาท



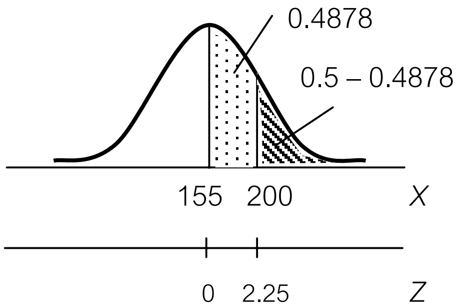
ภาพภาคผนวกที่ 17 พื้นที่ใต้เส้นโค้งปรกติระหว่าง Z = -0.25 กับ Z = 1.00

$$P\left(\frac{150 - 155}{20} \leq Z \leq \frac{175 - 155}{20}\right)$$

$$P(-0.25 \leq Z \leq 1) \quad = \quad 0.0987 + 0.3413 \quad = \quad 0.44$$

จำนวนพนักงานที่มีรายได้ระหว่าง 150 ถึง 175 บาท มีจำนวน = 2,000 (0.44) = 880 คน

ง. พนักงานมีรายได้ตั้งแต่ 200 บาทขึ้นไป



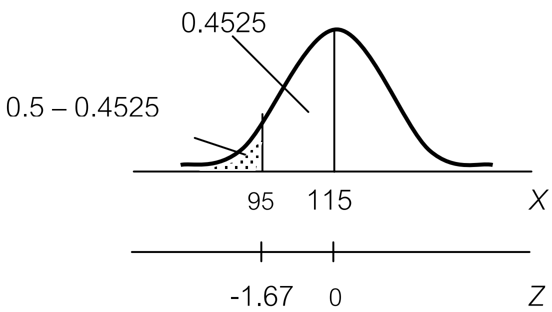
ภาพภาคผนวกที่ 18 พื้นที่ใต้เส้นโค้งปรกติระหว่าง $Z = 0$ กับ $Z = 2.25$

$$P\left(Z \geq \frac{200 - 155}{20}\right) \quad = \quad P(Z \geq 2.25)$$

$$= 0.5 - 0.4878 \quad = \quad 0.0122$$

พนักงานที่มีรายได้ตั้งแต่ 200 บาทขึ้นไป มีจำนวน =2,000(0.122) = 24 คน

6. ให้ X เป็นไอคิวของผู้สมัครสอบ



ภาพภาคผนวกที่ 19 พื้นที่ใต้เส้นโค้งปรกติระหว่าง $Z = -1.67$ กับ $Z = 0$

μ = 155 , σ = 12 , $(X \leq 95)$

$$P(X \leq 95) \quad = \quad P\left(\frac{X\ - \mu}{\sigma} \leq \frac{95 - 115}{12}\right)$$

$$= \quad P(\leq Z\ \text{-}\ 1.67) \qquad = \quad 0.5 - 0.4525$$

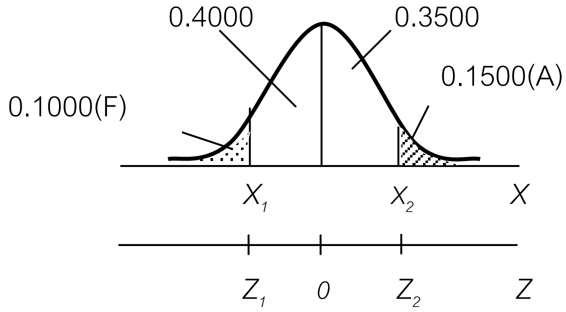
$$= \quad 0.0475$$

จำนวนผู้สมัครสอบคัดเลือกมีคุณสมบัติไม่ได้ตามต้องการจำนวน 600(0.4525) = 272 คน

7. ให้ X เป็นคะแนนสอบไล่วิชาสถิติ

X_1 เป็นคะแนนสอบสูงสุดของผู้ได้เกรด F

X_2 เป็นคะแนนสอบต่ำสุดของผู้ได้เกรด A



ภาพภาคผนวกที่ 20 พื้นที่ใต้เส้นโค้งปรกติเมื่อ Z ระหว่าง 0 กับ Z_1
และ Z ระหว่าง 0 กับ Z_2

ก. มี μ = 76 , σ = 15 และ Z_2 = 1.04

จาก $$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

$$1.04 = \left(\frac{X_2 - \mu}{\sigma}\right)$$

$$X_2 = (1.04)(15) + 76 \qquad = \quad 91.6$$

นั่นคือคะแนนต่ำสุดของเกรด A คือ 91.6 คะแนน

ข. จากตารางภาคผนวกที่ 5 ให้ $Z_1 = -1.28$

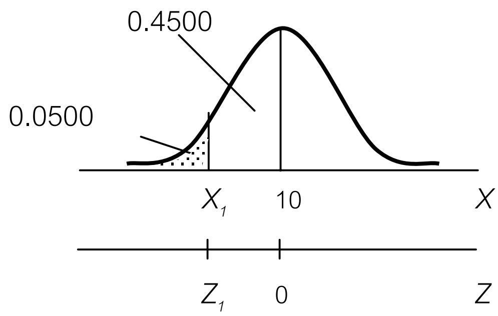
$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

$$-1.28 = \left(\frac{X_1 - \mu}{\sigma}\right)$$

$$X_1 = (-1.28)(15) + 76 = 56.8$$

นั่นคือ คะแนนสูงสุดของเกรด F คือ 56.8 คะแนน

8. ให้ X เป็นอายุการใช้งานของรถยนต์ที่ผลิตจากบริษัท



ภาพภาคผนวกที่ 21 พื้นที่ใต้เส้นโค้งปรกติเมื่อ Z ระหว่าง 0 กับ Z_1

$\mu = 10$, $\sigma = 2$ และ $Z_1 = -1.645$

จาก
$$Z_i = \left(\frac{X_i - \mu}{\sigma}\right)$$

$$-1.645 = \left(\frac{X_1 - \mu}{\sigma}\right)$$

$$X_1 = (-1.645)(12) + 10 = 6.71$$

นั่นคือประกันซ่อมฟรีไม่เกิน 7 ปี

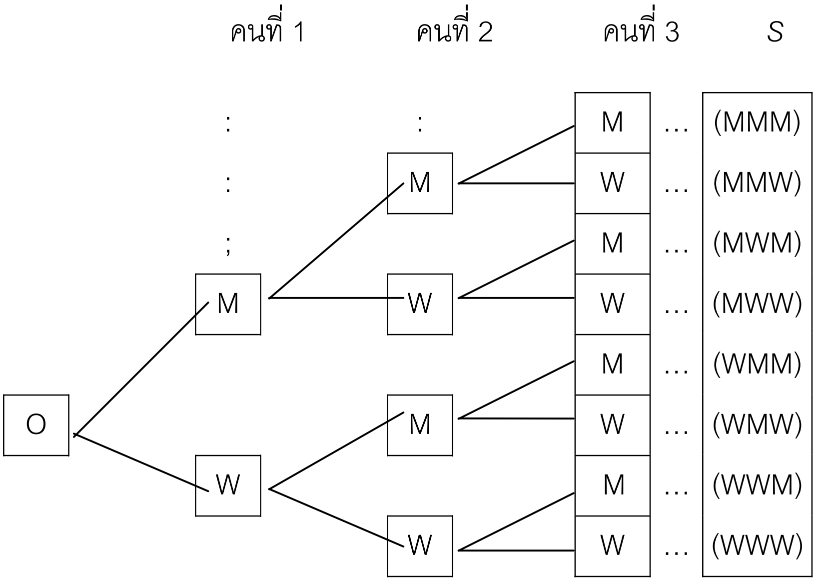
เฉลยคำถามทบทวนบทที่ 5

1. ให้ M เป็น บุตรเพศชาย

   W เป็น บุตรเพศหญิง

   S เป็น เซตจำนวนบุตรทั้งหมด

เขียนแผนภูมิต้นไม้ได้ดังภาพภาคผนวกที่ 22



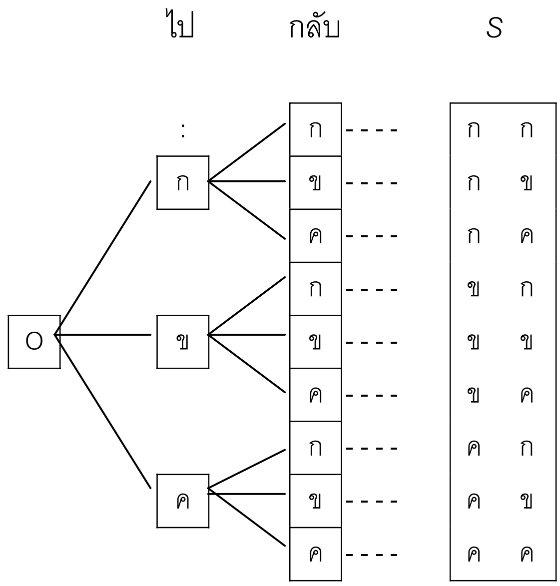
ภาพภาคผนวกที่ 22 แผนภูมิต้นไม้จำนวนบุตร

ผลลัพธ์ที่พึงไปได้ทั้งของการมีบุตร 3 คนของครอบครัวหนึ่งเป็นดังแสดงในภาพภาคผนวกที่ 22

2. ก. ให้ ก, ข, และ ค เป็นเส้นทางไปหรือกลับ

   S เป็น เซตเส้นทางไปและกลับทั้งหมด

เขียนแผนภูมิต้นไม้ได้ดังภาพภาคผนวกที่ 23



ภาพภาคผนวกที่ 23 แผนภูมิต้นไม้แสดงเส้นทางไป - กลับ

จากภาพภาคผนวกที่ 23 จำนวนวิธีการเดินทางไปหรือกลับทั้งหมดมี 9 วิธี

ข. จากภาพภาคผนวกที่ 23 จำนวนวิธีการเดินทางไปและกลับโดยไม่ซ้ำเส้นทางทั้งไปและกลับมี 6 วิธี

ค. การเดินทางไปกลับตาม ข. เมื่อเดินทางไปกลับทุกวันจะใช้เวลา 6 วัน

3. ให้ X เป็น จำนวนข้อที่เดาถูก

p เป็น โอกาสเดาถูกแต่ละข้อ

ก. $P(X = x) = \binom{n}{x} p^{x} q^{n-x}$ โดยที่ $x = 0, 1, 2, \ldots, n$

$$P(X = 3) = \binom{10}{3} (0.25)^{3} (0.75)^{7}$$

$$= 0.2493$$

โอกาสที่เขาจะตอบถูก 3 ข้อ มีค่าเท่ากับ 0.2493

ข. $P(X \geq 8) = \binom{10}{8}(0.25)^8(0.75)^2 + \binom{10}{9}(0.25)^9(0.75) + \binom{10}{10}(0.25)^{10}(0.75)^0$

$= 0.0004$

โอกาสที่เขาจะตอบถูก 8 ข้อขึ้นไปมีค่าเท่ากับ 0.0004

ค. $P(4 \leq X \leq 6) = \binom{10}{4}(0.25)^4(0.75)^6 + \binom{10}{5}(0.25)^5(0.75)^5 + \binom{10}{6}(0.25)^6(0.75)^4$

$= 0.2206$

โอกาสที่เขาจะตอบถูก 4 ถึง 6 ข้อมีค่าเท่ากับ 0.2206

4. ให้ เครื่องยนต์ 3 ชนิดแทนด้วย D,B,G

ตัวถัง 7 ชนิดแทนด้วย ก, ข, ค, ง, จ, ฉ, และ ช

สี 7 สี แทนด้วย a, b, c, d, e, f, และ g

S เป็นลักษณะของรถทั้งหมด

ดังนั้นจะได้

S = {Dกa, Dกb, Dกc, … ,Gชg}

ลักษณะของรถทั้งหมด = (3)(7)(7) = 147 แบบ

5. ก. ให้ A เป็น เซตบุตรเพศเดียวกัน

จากภาพภาคผนวกที่ 22 จะได้

A = {MMM, WWW}

ดังนั้นโอกาสที่จะพบบุตรในครอบครัวเพศเดียวกันหมด มีค่าเท่ากับ $\frac{2}{8}$

ข. ให้ B เป็น เซตบุตรคนแรกเป็นเพศชาย

จากภาพภาคผนวกที่ 22 จะได้

$$B = \{MMM, MMW, MWM, MWW\}$$

ดังนั้นโอกาสที่จะพบบุตรคนแรกในครอบครัวเพศชาย มีค่าเท่ากับ $\frac{4}{8} = \frac{1}{2}$

ค. ให้ C เป็น เซตบุตรคนที่ 3 เป็นเพศชาย

จากภาพภาคผนวกที่ 22 จะได้

$$C = \{MMM, MWM, WMM, WWM\}$$

ดังนั้นโอกาสที่จะพบบุตรคนแรกในครอบครัวเพศชาย มีค่าเท่ากับ $\frac{4}{8} = \frac{1}{2}$

6. จากสูตร $$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

ก. $$\binom{5}{2} = \frac{5!}{2!\,(5-2)!}$$

$$= \frac{(5)(4)(3)}{(2!)\,(3!)} = 10$$

ข. $$\binom{7}{7} = \frac{7!}{7!\,(7-7)!}$$

$$= 1$$

ค. $$\binom{7}{6} = \frac{7!}{6!\,(7-6)!}$$

$$= \frac{(7)(6!)}{(6!)(1!)} = 7$$

ง. $\binom{6}{4} = \frac{6!}{4!(6-4)!}$

$= \frac{(6)(5)(4!)}{(4!)(2!)} = 15$

จ. $\binom{6}{2} = \frac{6!}{2!(6-2)!}$

$= \frac{(6)(5)(4!)}{(2!)(4!)} = 15$

7. จากตารางภาคผนวกที่ 3 ที่ $p = 0.60$, $n = 5$

ก. $P(X \leq 2) = 0.3174$

ข. $E(X) = np$

$= 5(0.6) = 0.30$

$Var(X) = np(1-p)$

$= 5(0.6)(0.4) = 0.12$

8. จากตารางภาคผนวกที่ 3 ที่ $p = 0.30$, $n = 20$

ก. $P(X \geq 2) = 1 - P(X \leq 1)$

$= 1 - 0.0076 = 0.9924$

ข. $P(X \leq 2) = 0.0355$

ค. $P(X = 2) = P(X \leq 2) - P(X \leq 1)$

$= 0.0355 - 0.0076 = 0.0279$

ง. $E(X) = np$

$= 20(0.3) = 6.0$

$Var(X) = np(1-p)$

$= 20(0.3)(0.7) = 4.2$

9. จากตารางภาคผนวกที่ 4 ที่ $\lambda = 4.0$

ก. $P(2 \le X \le 5) = P(X \le 5) - P(X \le 1)$

$= 0.785 - 0.092 = 0.698$

ข. $P(X \ge 3) = 1 - P(X \le 2)$

$= 1 - 0.238 = 0.762$

ค. $P(X \le 5) = 0.238$

10. $N = 50$, $k = 3$ และ $n = 5$

$$P(X=1) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{3}{1}\binom{50-3}{5-1}}{\binom{50}{5}}$$

$$= \frac{\binom{3}{1}\binom{47}{4}}{\binom{50}{5}} = \frac{(3)(178{,}365)}{(2{,}118{,}760)}$$

$$= 0.252$$

$$P(X \geq 2) = 1 - P(X \leq 1)$$

$$= 1 - \left[\frac{\binom{3}{0}\binom{50-3}{5-0}}{\binom{50}{5}} + \frac{\binom{3}{1}\binom{50-3}{5-1}}{\binom{50}{5}} \right]$$

$$= 1 - \left[\frac{(1)(1,533,939)}{(2,118,760)} + \frac{(3)(178,365)}{(2,118,760)} \right]$$

$$= 1 - 0.976$$

$$= 0.023$$

เฉลยคำถามทบทวนบทที่ 6

1. นิรนัยจะเป็นการสรุปเกี่ยวกับคุณลักษณะของประชากรที่ได้ศึกษาทั้งหมดแล้ว แต่การอุปนัยจะกล่าวถึงกระบวนการสรุปคุณลักษณะของประชากรด้วยประสบการณ์และอาศัยการศึกษาตัวอย่างอันเป็นส่วนหนึ่งของประชากร

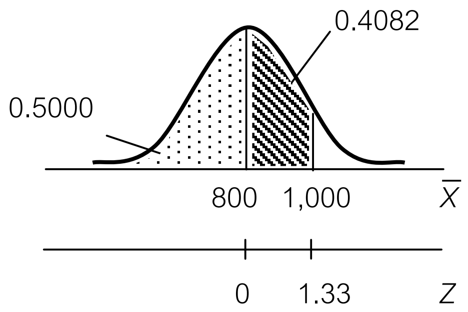
2. การชักตัวอย่างแบบใส่คืน คือเทคนิคการสุ่มหยิบประชากรที่มีคุณลักษณะที่เราสนใจหรือมีข้อแตกต่างกันภายในประชากร และได้ใส่กลับคืนไปอีกก่อนหยิบประชากรหน่วยต่อไป ทำให้หน่วยประชากรขนาดจำกัดกลายเป็นประชากรที่นับไม่รู้จบหรือมีขนาดอนันต์

3. เมื่อมีการชักตัวอย่างขนาด n ตัวแปรสุ่มนี้ได้เป็นตัวสถิติเหล่านี้ไปสัมพันธ์กับความน่าจะเป็นหรือเรียกว่า " การแจกแจงตัวอย่าง" เช่น การแจกแจง $\overline{X}$ เป็นแบบปรกติหรือแบบที การแจกแจง S^2 เป็นแบบ X^2 หรือแบบ F เป็นต้น

4. อาจจะสุ่มคนในสมุดโทรศัพท์ จำนวน 10 คน โดยการจับสลากหาจุดเริ่มต้นได้หน่วยตัวอย่างที่ 1 และหน่วยตัวอย่างต่อ ๆ ไปจะห่างกัน k โดย k เท่ากับขนาดประชากร (m) หารด้วยขนาดตัวอย่าง (n) จนได้จำนวนตัวอย่างตามขนาดที่ต้องการ

5. จากโจทย์ $\mu = 800$, $\sigma = 900$

ก. ให้ $P\left(\bar{X} \leq 1{,}000\right)$ เป็น โอกาสจะพบพนักงานที่มีรายได้เฉลี่ยต่อเดือนไม่เกิน 1,000 บาท



ภาพภาคผนวกที่ 24 การแจกแจง $\bar{X}$ เมื่อ $\bar{X} \leq 1{,}000$

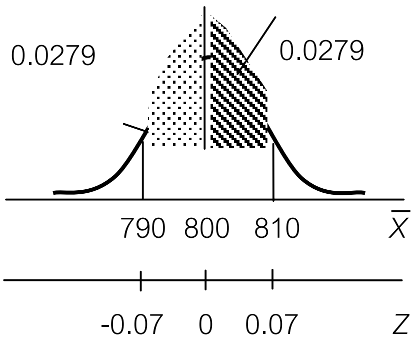
$$P\left(\bar{X} \leq 1{,}000\right) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{1{,}000 - 800}{900/\sqrt{36}}\right)$$

$$= P\left(Z \leq 1.33\right) \qquad = 0.5 + 0.4082$$

$$= 0.9082$$

โอกาสพบพนักงานมีรายได้เฉลี่ยต่อเดือนไม่เกิน 1,000 บาท = 0.9082

ข. ให้ $P\left(790 \leq \bar{X} \leq 810\right)$ เป็น โอกาสจะพบพนักงานที่มีรายได้เฉลี่ยต่อเดือนระหว่าง 790 บาท ถึง 810 บาท



ภาพภาคผนวกที่ 25 การแจกแจง $\overline{X}$ เมื่อ $790 \leq \overline{X} \leq 810$

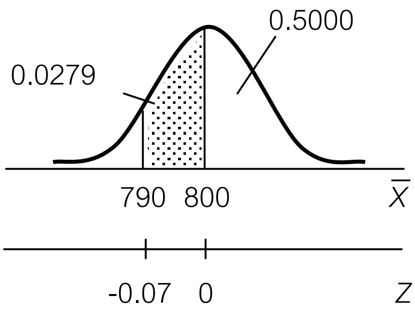
$$P(790 \leq \overline{X} \leq 810) = P\left(\frac{790-800}{900/\sqrt{36}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{810-800}{900/\sqrt{36}}\right)$$

$$= P(-0.07 \leq Z \leq 0.07) = 0.0279 + 0.0279$$

$$= 0.0558$$

โอกาสจะพบพนักงานมีรายได้เฉลี่ยต่อเดือนระหว่าง 790 ถึง 810 บาท = 0.0558

ค. ให้ $P(\overline{X} \geq 790)$ เป็น โอกาสจะพบพนักงานที่มีรายได้เฉลี่ยต่อเดือน 790 บาทขึ้นไป



ภาพภาคผนวกที่ 26 การแจกแจง $\overline{X}$ เมื่อ $\overline{X} \geq 790$

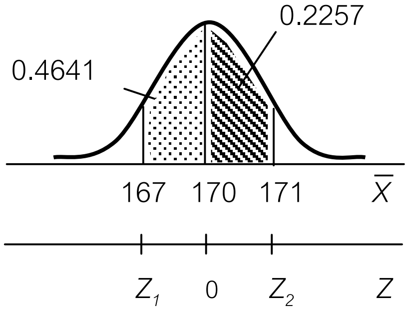
$$P(\overline{X} \geq 790) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{790-800}{900/\sqrt{36}}\right)$$

$$= P(Z \leq -0.07) = 0.5000 + 0.0279$$

$$= 0.5279$$

โอกาสที่จะพบพนักงานมีรายเฉลี่ยต่อเดือนตั้งแต่ 790 บาทขึ้นไป = 0.5279

6. จากโจทย์ $\mu = 170$, $\sigma = 7.5$, $n = 20$



ภาพภาคผนวกที่ 27 การแจกแจง $\overline{X}$ เมื่อ $167 \le \overline{X} \le 171$

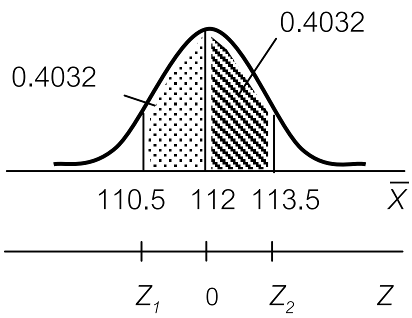
$$P(167 \le \overline{X} \le 171) = P\left(\frac{167-170}{0.75/\sqrt{20}} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \frac{171-170}{0.75/\sqrt{20}}\right)$$

$$= P(-1.80 \le Z \le 0.60) \qquad = 0.4641 + 0.2257$$

$$= 0.6898$$

โอกาสที่จะพบตัวอย่างสุ่มที่มีความสูงเฉลี่ยระหว่าง 167 ถึง 171 เซนติเมตร = (0.6898)(80) = 55 ตัวอย่าง

7. จากโจทย์ $\mu = 112$, $\sigma = 10$, $n = 75$



ภาพภาคผนวกที่ 28 การแจกแจง $\overline{X}$ เมื่อ $110.5 \le \overline{X} \le 113.5$

$$P(110.5 \leq \overline{X} \leq 113.5) \quad = \quad P\left(\frac{110.5-112}{10/\sqrt{75}} \leq \frac{\overline{X}\ -\mu}{\sigma/\sqrt{n}} \leq \frac{113.5-112}{10/\sqrt{75}}\right)$$

$$= \quad P(-1.30 \leq Z \leq 1.30) \qquad = \quad 0.4032 + 0.4032$$

$$= \quad 0.8064$$

โอกาสจะพบตัวอย่างสุ่มที่มีความสูงเฉลี่ยระหว่าง 110.5 กับ 113.5 = 0.8064

เฉลยคำถามทบทวนบทที่ 7

1. ก. ขอบเขตทั้งหมดของข้อมูลที่ต้องการศึกษา

ข. ส่วนหนึ่งของประชากรที่ถูกสุ่มเป็นตัวอย่าง

ค. เป็นตัวสถิติที่เข้าเกณฑ์ไม่อคติ ความแนบนัย มีประสิทธิภาพ ความเพียงพอสำหรับแต่ละพารามิเตอร์ที่เหมาะสม

ง. ช่วงความเชื่อมั่น คือช่วงระหว่างค่าต่ำกับค่าสูงของการแจกแจงตัวสถิติที่เชื่อว่าค่าพารามิเตอร์นั้นตกอยู่

จ. ถ้าการประมาณค่าพารามิเตอร์ใดด้วยตัวประมาณค่าที่ไม่อคติแล้วถ้าค่าคาดหมายของตัวประมาณค่านั้นจะเท่ากับค่าพารามิเตอร์นั้น

2. การประมาณค่าแบบจุด ได้ค่าพารามิเตอร์ที่เราต้องการ ถ้าประมาณค่าผิดก็จะไม่สามารถใช้ประโยชน์ได้ แต่การประมาณค่าพารามิเตอร์แบบช่วงเป็นการประมาณค่าโดยมีช่วงค่าประมาณอยู่ระหว่างค่าต่ำกับค่าสูง และสามารถกำหนดช่วงนี้ด้วยค่าความเสี่ยงที่กำหนดได้

3. ให้ μ เป็น น้ำหนักผงซักฟอกชนิดบรรจุ 1 กิโลกรัมโดยเฉลี่ย

$\overline{X}$ = 0.95, S = 0.335, n = 36, $t_{.025,35}$ = 2.042

ช่วงความเชื่อมั่น 95 % ของ μ คือ

$$\mu \quad = \quad \overline{X} \pm t_{.025,35}\frac{S}{\sqrt{n}} \qquad = \quad 0.95 \pm (2.042)\ \frac{0.335}{6}$$

$$= \quad 0.95 \pm 0.114 \qquad = \quad [0.836,1.064]$$

ดังนั้น น้ำหนักผงซักฟอกชนิดบรรจุ 1 ก.ก อยู่ระหว่าง 0.836 ถึง 1.064 กิโลกรัม ด้วยช่วงความเชื่อมั่น 95%

4. ให้ μ เป็น รายได้ของเกษตรกรไทยต่อเดือนโดยเฉลี่ย

$$\bar{X} = 3{,}500, \quad S = 325, \quad n = 35, \; t_{.025,34} \approx 1.96$$

ช่วงความเชื่อมั่น 95 % ของ μ คือ

$$\mu = \bar{X} \pm t_{.025,34}\frac{S}{\sqrt{n}} = 3{,}500 \pm (1.96)\,\frac{325}{5.916}$$

$$\mu = 3{,}500 \pm 54.935 = [3{,}445.065,\ 3{,}554.935]$$

รายได้เฉลี่ยของเกษตรกรไทยต่อเดือนระหว่าง 3,445.065 บาท กับ 3,554.935 บาท ด้วยช่วงความเชื่อมั่น 95%

5. ให้ μ_1 เป็น รายได้ต่อวันของพนักงานโรงงานทอผ้าโดยเฉลี่ย
μ_2 เป็น รายได้ต่อวันของพนักงานโรงงานยาสูบโดยเฉลี่ย

$$\bar{X}_1 = 45 \qquad \bar{X}_2 = 70$$

$$n_1 = 25 \qquad n_2 = 35$$

$$S_1^2 = 100 \qquad S_2^2 = 105$$

$$df = n_1 + n_2 - 2 \;,\; t_{25+35-2,.025} = 1.96$$

ช่วงความเชื่อมั่น 95 % ของ $\mu_2 - \mu_1$ คือ

$$\mu_2 - \mu_1 = (\bar{X}_2 - \bar{X}_1) \pm t_{df,0.25}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$= (70 - 45) \pm 1.96\sqrt{\frac{100}{25}+\frac{105}{35}}$$

$$= 25 \pm 1.96\sqrt{7}$$

$$\mu_2 - \mu_1 = 25 \pm 5.21 = [9.79, 30.21]$$

สรุปว่า ด้วยช่วงความเชื่อมั่น 95% รายได้ของพนักงานโรงงานยาสูบสูงกว่ารายได้ของพนักงานโรงงานทอผ้าระหว่าง 9.79 ถึง 30.21 บาทต่อวัน

6. ให้ π เป็น สัดส่วนผู้ใช้ผงซักฟอกยี่ห้อ A

$$n = 142$$

$$p = \frac{83}{142} = 0.584$$

$$Z_{\alpha/2} = Z_{.05/2} = Z_{0.25} = 1.96$$

ดังนั้นช่วงความเชื่อมั่นที่ 95 % ของ π คือ

$$\pi = p \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$= 0.584 \pm 1.96\sqrt{\frac{(0.584)(0.416)}{142}}$$

$$= 0.584 \pm 0.08 = [0.504, 0.664]$$

สรุปว่า มีผู้ใช้ผงซักฟอกยี่ห้อ A ร้อยละ 50.4 ถึง ร้อยละ 66.4 ด้วยช่วงความเชื่อมั่น 95%

7. ให้ σ เป็น ค่าเบี่ยงเบนมาตรฐานของความหนาของแผ่นพลาสติกทั้งหมดที่ผลิตได้

$$n = 9$$

$$\overline{X} = \frac{19.8 + 21.2 + 18.6 + 20.4 + 21.6 + 19.8 + 19.9 + 20.3 + 20.8}{9}$$

$$= \frac{182.4}{9} = 20.27$$

$$S^2 = \frac{\sum\left(x - \overline{x}\right)^2}{(n-1)} = 0.787$$

$$S = 0.887$$

$$\alpha = 0.05 , \quad \frac{\alpha}{2} = 0.25 , \quad 1 - \frac{\alpha}{2} = 0.975$$

$$X^{2}_{8,0.25} = 17.535 , \qquad X^{2}_{8,0.975} = 2.180$$

ช่วงความเชื่อมั่นที่ 95 % ของ σ^2 คือ

$$\frac{(n-1)S_x^2}{X^2_{d,\alpha/2}} < \sigma_x^2 < \frac{(n-1)S_x^2}{X^2_{df,(1-\alpha/2)}}$$

$$= \frac{(8)(0.787)}{17.535} < \sigma_x^2 < \frac{(8)(0.787)}{2.180}$$

$$= 0.3591 < \sigma_x^2 < 0.722$$

$$= 0.599 < \sigma_x < 0.849$$

เชื่อได้ 95 % ว่า ค่าเบี่ยงเบนมาตรฐานของความหนาของแผ่นพลาสติกทั้งหมดที่ผลิตได้อยู่ระหว่าง 0.599 ถึง 0.849 มิลลิเมตร

เฉลยคำถามทบทวนบทที่ 8

1. การทดสอบสมมุติฐาน เป็นการหาคำตอบของสมมุติฐานที่เราตั้งขึ้นมา โดยใช้วิธีที่เหมาะสม โดยจะมีการตั้งสมมุติฐานถึงความน่าจะจริง ภายใต้ความเสี่ยงที่กำหนด ส่วนช่วงความเชื่อมั่นคือการกำหนดช่วงของพารามิเตอร์ระหว่างค่าต่ำกับค่าสูงภายใต้ความเสี่ยงเช่นเดียวกัน

2. ก. ให้ μ เป็น รายได้ของเกษตรกรไทยโดยเฉลี่ย

H_0 : $\mu = 5,000$

H_1 : $\mu > 5,000$,

ข. ให้ μ เป็น รายได้ต่อวันของครอบครัวชาวนาไทยอีสานโดยเฉลี่ย

$H_0 \quad : \quad \mu = 200$

$H_1 \quad : \quad \mu < 200$

ค. ให้ μ_1 เป็น ยอดขายปุ๋ยยี่ห้อ ก. โดยเฉลี่ย

μ_2 เป็น ยอดขายปุ๋ยยี่ห้อ ข. โดยเฉลี่ย

$H_0 \quad : \quad \mu_1 = \mu_2$

$H_1 \quad : \quad \mu_1 > \mu_2$

ง. . ให้ p เป็น สัดส่วนของนักศึกษาที่เห็นพ้องกับนโยบาย

$H_0 \quad : \quad p = 0.5$

$H_1 \quad : \quad p > 0.5$

3. ให้ μ เป็น รายได้ต่อวันของพนักงาน

ดำเนินการทดสอบสมมุติฐาน

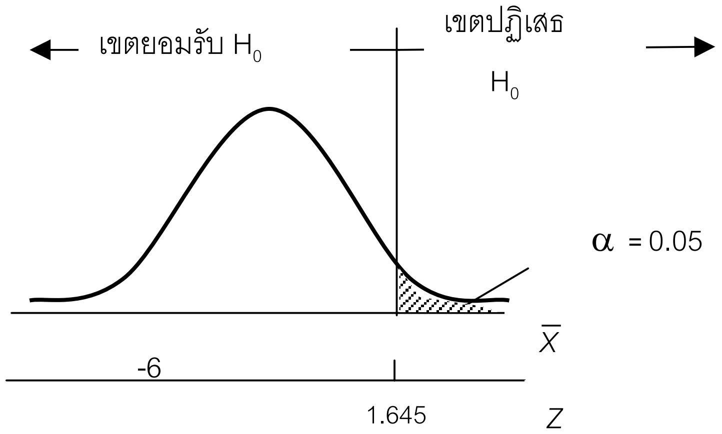
ขั้นที่ 1 $H_0 \quad : \quad \mu_0 = 50$

ขั้นที่ 2 $H_0 \quad : \quad \mu_0 > 50$

ขั้นที่ 3 ระดับนัยสำคัญ $\alpha = .05$

ขั้นที่ 4 กรณีเป็นการทดลองทางเดียว : ทางขวา ที่ระดับนัยสำคัญ $\alpha = .05$
$Z_{.05} = 1.64$ ดังนั้นเขตปฏิเสธ H_0 เมื่อ $Z_C > 1.645$

$\overline{X} = 47$, $S = 20$, $n = 1{,}600$



ภาพภาคผนวกที่ 29 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .50

ขั้นที่ 5 $$Z_C = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

$$Z_C = \frac{47 - 50}{20/\sqrt{1,600}} = -6$$

ขั้นที่ 6 ตัดสินใจปฏิเสธ H_0 เพราะ $Z_C < 1.64$ อยู่ในเขตยอมรับ H_0 ที่ระดับ α = .05 นั่นคือ เชื่อได้ 95% ว่าพนักงานมีรายได้ไม่มากกว่า 50 บาทต่อวัน

4. ให้ μ เป็น รายได้ต่อปีของพนักงานโดยเฉลี่ย

ดำเนินการทดสอบสมมุติฐาน

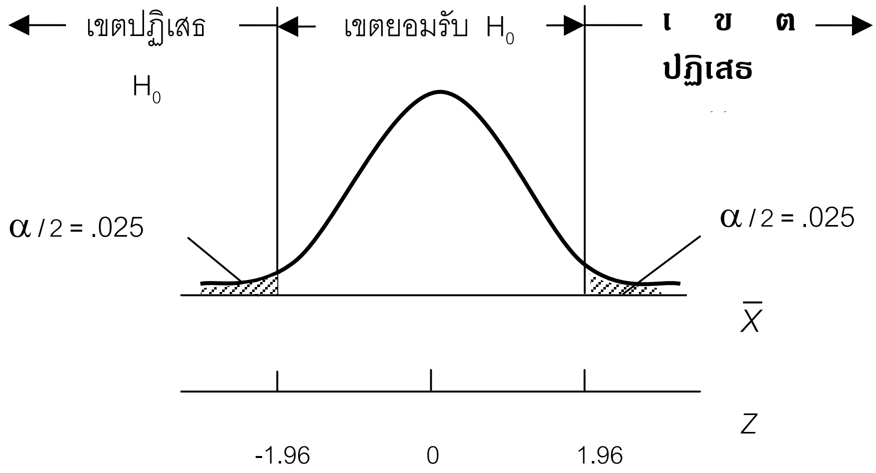
ขั้นที่ 1 H_0 : μ = 30,000

ขั้นที่ 2 H_0 : $\mu \neq$ 30,000

ขั้นที่ 3 ระดับนัยสำคัญ α = .05 หรือ .01

ขั้นที่ 4 กรณีเป็นการทดสอบสองหางดังแสดงใน ภาพภาคผนวกที่ 28 ที่ระดับ $\alpha/2$ = .025 , Z = 1.96 ดังนั้น เขตปฏิเสธ H_0 คือ $Z \leq -1.96$ หรือ $Z \geq 1.96$

แต่ต้องดำเนินการสุ่มตัวอย่างเพื่อให้ได้ค่าเฉลี่ยตัวอย่างจากพนักงาน 100 รายเสียก่อนจึงจะดำเนินการต่อไปได้



ภาพภาคผนวกที่ 30 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05
ของรายได้ต่อปีของพนักงาน

5. ให้ μ เป็น น้ำหนักของผงซักฟอกชนิดบรรจุ 1 กิโลกรัมโดยเฉลี่ย

$\overline{X}$ = 0.95 , S = 0.3355 , n = 36

ดำเนินการตามสมมุติฐานดังนี้

ขั้นที่ 1 $H_0 \; : \; \mu = 1$

ขั้นที่ 2 $H_0 \; : \; \mu \neq 1$

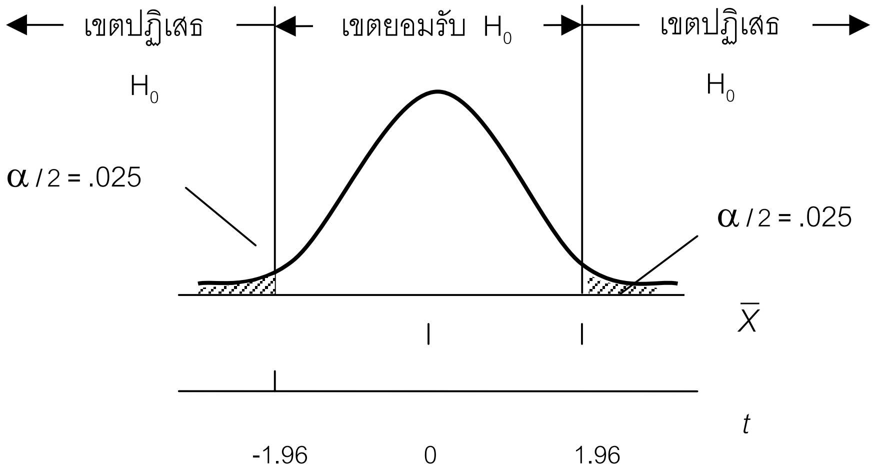
ขั้นที่ 3 ระดับนัยสำคัญ α = .05 /2 = 0.25

ขั้นที่ 4 เขตปฏิเสธ H_0 คือ $t_c < -1.96$ หรือ $t_c > 1.96$

ขั้นที่ 5 ใช้สถิติทดสอบ t_c

$$t_c = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{0.95 - 1.00}{\frac{0.335}{\sqrt{36}}}$$

$$= -0.895$$



ภาพภาคผนวกที่ 31 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .025
น้ำหนักของผงซักฟอกชนิดบรรจุ 1 กิโลกรัม

ขั้นที่ 6 จากภาพภาคผนวกที่ 29 t_c อยู่ในเขตยอมรับ H_0 ดังนั้นเชื่อได้ 95 % ว่าน้ำหนักผงซักฟอกชนิดบรรจุ 1 กิโลกรัมทั้งหมดยังคงมีน้ำหนัก 1 กิโลกรัมตามที่ระบุไว้

6. ให้ μ_1 เป็น รายได้ต่อวันของพนักงานโรงงานทอผ้าโดยเฉลี่ย
μ_2 เป็น รายได้ต่อวันของพนักงานโรงงานยาสูบโดยเฉลี่ย

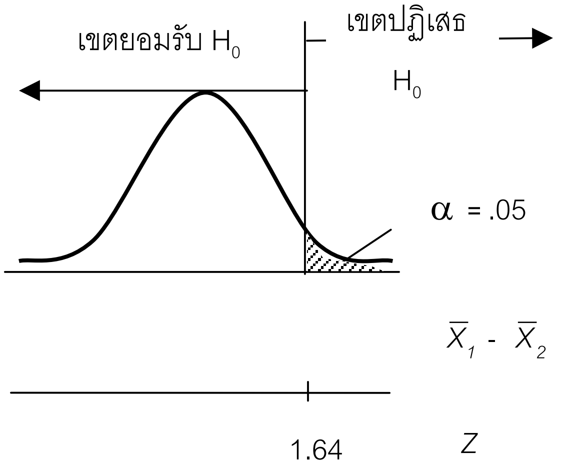
ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 H_0 : $\mu_1 = \mu_2$

ขั้นที่ 2 H_1 : $\mu_1 > \mu_2$

ขั้นที่ 3 ระดับนัยสำคัญ α = .05

ขั้นที่ 4 กรณีทดสอบหางเดียวทางขวา ที่ระดับ α = .05 , $Z_.$ = 1.64 ดังนั้นเขตปฏิเสธ H_0 คือ $Z_C > 1.64$



ภาพภาคผนวกที่ 32 ทิศทางการทดสอบสมมุติฐานที่ระดับ α = .05
ของรายได้ของพนักงานโรงงานทอผ้าและโรงงานยาสูบ

ขั้นที่ 5 คำนวณค่าตัวสถิติทดสอบ : Z_C

$$Z_C = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$\overline{X}_1$ = 45 $\qquad$ $\overline{X}_2$ = 70

n_1 = 25 $\qquad$ n_2 = 35

S_1^2 = 100 $\qquad$ S_2^2 = 105

$$Z_C = \frac{45-70}{\sqrt{\frac{(100)^2}{25} + \frac{(105)^2}{35}}} = -\frac{25}{2.65}$$

$$= -9.43$$

ขั้นที่ 6 ตัดสินใจยอมรับ H_0 เพราะ Z_c < 1.645 ตกอยู่ในยอมรับ H_0 นั่นคือ เชื่อว่ารายได้ของพนักงานโรงงานทอผ้าต่อวันไม่สูงกว่าพนักงานโรงงานยาสูบที่ระดับ α = .05

7. ให้ π เป็น สัดส่วนผู้ซื้อเสื้อยืดเจบีทั้งหมด

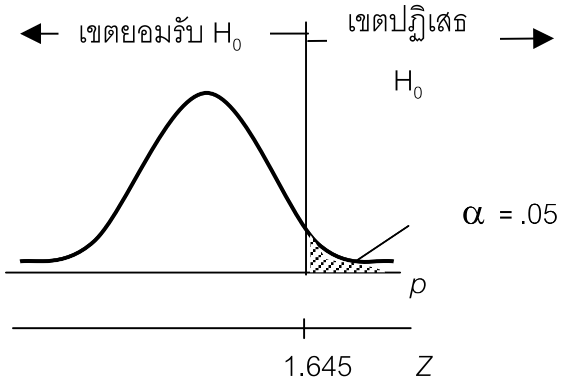
ดำเนินการทดสอบสมมุติฐาน

ขั้นที่ 1 $H_0 : \pi = 0.15$

ขั้นที่ 2 $H_1 : \pi > 0.15$

ขั้นที่ 3 ระดับนัยสำคัญ $\alpha = .10$

ขั้นที่ 4 สมมุติฐานเกี่ยวกับสัดส่วนประชากร ตัวสถิติทดสอบคือ Z_C และเป็นการทดสอบหางเดียวทางขวา ที่ระดับ $\alpha = .05$ จากตารางภาคผนวกที่ 5 หรือ จากตารางภาคผนวกที่ 6 $df = \infty$, $t_{.05} = Z_{.05} = 1.645$



ภาพภาคผนวกที่ 33 ทิศทางการทดสอบสมมุติฐานที่ระดับ $\alpha = .05$ ของผู้ซื้อเสื้อยืดเจบี

ขั้นที่ 5 คำนวณค่าตัวสถิติทดสอบ : Z_C

$$Z_C = \frac{p - \pi_0}{\sqrt{\frac{p(1-p)}{n}}}$$

ให้ p เป็น สัดส่วนการครองตลาดเสื้อยืดเจบี

$$n = 75$$

$$p = 0.17\ , \ (1-p) = 0.83$$

$$Z_\alpha = Z_{.05} = 1.645$$

$$Z_c = \frac{0.17 - 0.15}{\sqrt{\frac{(0.17)(0.83)}{75}}} = 0.465$$

ขั้นที่ 6 ตัดสินใจยอมรับ H_0 เพราะ $Z_c < 1.645$ ตกอยู่ในยอมรับ H_0 นั่นคือ เชื่อว่าเสื้อยืดเจบีไม่มีส่วนแบ่งการตลาดมากกว่าเดิม ที่ระดับ $\alpha = .05$

เฉลยคำถามทบทวนบทที่ 9

1. ความแปรผันระหว่างตัวอย่าง คือความแปรผันระหว่างค่าเฉลี่ยตัวอย่าง ($\overline{X}_i$)

ความแปรผันภายในตัวอย่าง คือ ความแปรผันของค่าสังเกตในแต่ละตัวอย่างมีการแปรผันหรือแกว่งอยู่รอบ ๆ ค่าเฉลี่ย ของแต่ละตัวอย่างนั้น

2. จากตารางภาคผนวกที่ 7 และ 8

ก. $F_{9,16,.05}$ = 2.5377

ข. $F_{9,16,.01}$ = 3.7804

ค. $F_{8,25,.05}$ = 2.3371

ง. $F_{8,25,.01}$ = 3.3239

จ. $F_{12,30,.05}$ = 2.0921

ฉ. $F_{12,30.01}$ = 2.8431

3. ใส่ * เมื่อ F_c มีค่ามากกว่า $F_{.05}$ แต่ยังน้อยกว่า $F_{.01}$ จะสรุปว่ามีความแตกต่างกันอย่างมีนัยสำคัญ

ใส่ ** เมื่อ F_c มีค่ามากกว่า $F_{.01}$ และมากกว่า $F_{.05}$ จะสรุปว่ามีความแตกต่างกันอย่างมีนัยสำคัญยิ่ง

4. ในกรณีที่มี 2 แหล่งความแปรปรวนที่อธิบายได้ เช่น ยอดจำหน่ายรวมเมื่อส่งเริมการขายโดยวิธีลดราคา (แหล่งที่ 1) และเมื่อส่งเสริมการขายโดยวิธีโฆษณาทางวิทยุ (แหล่งที่ 2) เป็นต้น

5. ให้ μ_1 เป็น ค่าเฉลี่ยประชากรที่ 1

μ_2 เป็น ค่าเฉลี่ยประชากรที่ 2

μ_3 เป็น ค่าเฉลี่ยประชากรที่ 3

μ_4 เป็น ค่าเฉลี่ยประชากรที่ 4

H_0 : $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1 : μ ไม่เท่ากันทั้งหมด

จากข้อกำหนด

$df_{total} = 16 - 1 = 15$

$df_{COL} = 4 - 1 = 3$

$df_{RES} = 15 - 3 = 12$

$$\overline{\overline{X}} = \frac{0.261 + 0.296 + 0.312 + 0.135}{4}$$

$$= \frac{1.004}{4} = 0.251$$

$$SS_{COL} = \sum_{i=1}^{c} (\overline{X}_i - \overline{\overline{X}})^2$$

$$= (0.261 - 0.251)^2 + (0.296 - 0.251)^2 + (0.312 - 0.251)^2 + (0.135 - 0.251)^2$$

$$= 0.0193$$

$$SS_{RES} = SS_1 + SS_2 + SS_3 + SS_4$$

$$= (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2 + (n_4 - 1)S_4^2$$

$$= 3(0.21)^2 + 3(0.17)^2 + 3(0.19)^2 + 3(0.08)^2$$

$SS_{RES} = 0.1323 + 0.0867 + 0.1083 + 0.0192$

$= 0.3465$

$SS_{total} = SS_{COL} + SS_{RES} = 0.193 + 0.3465$

$= 0.3658$

$$MS_{COL} = \frac{SS_{col}}{df_{col}} = \frac{1.0193}{3} = 0.0064$$

$$MS_{RES} = \frac{SS_{RES}}{df_{RES}} = \frac{1.3465}{12} = 0.0289$$

$F_c = 0.22$

ตารางภาคผนวกที่ 21 ตารางแอนโนวาของประชากร 4 ประชากร

SV	*df*	*SS*	*MS*	F_c
ระหว่างตัวอย่าง (*COL*)	3	0.0193	0.0064	0.22
เศษ (*RES*)	12	0.3465	0.0289	
ผลรวม	15	0.358		

$F_{.05,3,12} = 3.4903$

$F_{.01,3,12} = 5.7394$

ตัดสินใจยอมรับ H_0 เพราะ F_c อยู่ในเขตยอมรับ H_0
สรุป ประชากรทั้ง 4 กลุ่มไม่แตกต่างกันหรือตัวอย่างเหล่านี้มาจากประชากรเดียวกัน ที่ $\alpha = .01$

6. ให้ μ_1 เป็น รายจ่ายต่อเดือนของสาขาย่อยที่ 1 โดยเฉลี่ย และมีความแปรปรวน σ_1^2

μ_2 เป็น รายจ่ายต่อเดือนของสาขาย่อยที่ 2 โดยเฉลี่ย และมีความแปรปรวน σ_2^2

μ_3 เป็น รายจ่ายต่อเดือนของสาขาย่อยที่ 3 โดยเฉลี่ย และมีความแปรปรวน σ_3^2

μ_4 เป็น รายจ่ายต่อเดือนของสาขาย่อยที่ 4 โดยเฉลี่ย และมีความแปรปรวน σ_4^2 นำข้อมูลมาจัดกระทำจะได้ดังตารางภาคผนวกที่ 22

ตารางภาคผนวกที่ 22 ตารางข้อมูลองรายได้ต่อเดือนของสาขาย่อย 4 สาขา

	รายจ่ายของสาขาย่อย (ล้านบาท)				
	1	2	3	4	
	0.6	1.1	2.1	0.5	
	0.8	1.3	2.0	0.9	
	0.3	1.5	1.7	1.0	
	0.5	1.8	1.6	0.7	
	0.6	0.9	1.0	0.7	
รวม (T_i)	2.8	6.6	8.4	3.8	GT = 21.6
ค่าเฉลี่ย ($\overline{X}_i$)	0.6	1.3	1.7	0.8	GM = 1.08
S_i^2	0.033	0.122	0.187	0.038	$\sum_{i=1}^{c} S_i = 0.38$

ก. ทดสอบสมมุติฐานเกี่ยวกับความเท่ากันของความแปรปรวน

$$H_0 \quad : \quad \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

H_1 : σ^2 ไม่เท่ากันทั้งหมด

$$G = \frac{Largest\ S_i^2}{\sum_{i=1}^{c} S_i^2} = \frac{0.187}{0.380}$$

$$= 0.49$$

จากตารางภาคผนวกที่ 12 หรือ 13 เขตปฏิเสธ H_0 คือ $g_{(n,c)} = g_{5,4,.05} = 0.6287$
หรือ $g_{5,4,.01} = 0.7212$
ตัดสินใจยอมรับ H_0 นั่นคือความแปรปรวนของรายจ่ายต่อเดือนของทุกสาขาเท่ากัน

ข. ทดสอบสมมุติฐาน

$H_0 \quad : \quad \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 \quad : \quad \mu$ ไม่เท่ากันทั้งหมด

$$CF = \frac{(GT)^2}{N} = \frac{(2.16)^2}{20}$$

$$= \frac{466.56}{20} = 23.33$$

$$df_{total} = 20 - 1 = 19$$

$$dfB = 4 - 1 = 3$$

$$dfE = 19 - 3 = 16$$

$$SSB = \sum_{i=1}^{c} \frac{T_i^2}{n_1} - CF$$

$$= \left[\frac{(2.8)^2}{5} + \frac{(6.6)^2}{5} + \frac{(8.4)^2}{5} + \frac{(3.8)^2}{5}\right] - 23.33$$

$$= 27.28 - 23.33$$

$$= 3.95$$

$$SS_{total} = [(0.6)^2 + (0.8)^2 + (0.3)^2 + (0.6)^2 + (1.1)^2 + (1.3)^2 + (1.5)^2 + (1.8)^2 + (0.9)^2 + (2.1)^2 + (2.0)^2 + (1.7)^2 + (1.6)^2 + (1.0)^2 + (0.5)^2 + (0.9)^2 + (0.1)^2 + (0.7)^2 + (0.7)^2] - 23.33$$

$$= 28.80 - 23.33 = 5.47$$

$$SSE = SS_{total} - SSB = 5.47 - 3.95$$

$$= 1.52$$

$$MSB = \frac{SSB}{dfB} = \frac{3.95}{3}$$

$$MSB = 1.32$$

$$MSE = \frac{SSE}{dfE} = \frac{5.47}{16}$$

$$= 0.034$$

$$F_c = \frac{MSB}{MSE} = \frac{1.32}{0.34}$$

$$= 3.88$$

ตารางภาคผนวกที่ 23 ตารางแอนโนวาของรายจ่ายต่อเดือนของสาขาย่อย

SV	*df*	*SS*	*MS*	F_c
รายจ่าต่อเดือน (*B*)	3	3.95	1.32	3.88
เศษ (*E*)	16	1.52	0.34	
ผลรวม	19	23.33		

$$F_{.05,3,16} = 3.2389$$

$$F_{.01,3,16} = 5.2922$$

ตัดสินใจยอมรับ H_0 เพราะ F_c ที่ α = .05 นั่นคือ รายจ่ายต่อเดือนของสาขาทั้ง 4 ไม่แตกต่างกัน

7. จัดโครงสร้างจะได้ตารางภาคผนวกที่ 24 และจากผลของข้อมูลในตารางที่ 24 จะได้ตารางแอนโนวาดังตารางที่ 25

ตารางภาคผนวกที่ 24 ผลการจัดโครงสร้างที่ได้จากทดลองใช้ฮอร์โมน 4 ชนิดเลี้ยงไก่ 4 โรงเรือน โรงเรือนละ 4 ตัว

ฮอร์โมน / โรงเรือน	a	b	c	d	รวม
1	.71	.95	.40	.69	2.75
2	.44	.68	.64	.71	2.47
3	.86	.85	.51	.42	2.64
4	.44	.41	.74	.62	2.21
รวม	2.45	2.89	2.29	2.44	10.07

ตารางภาคผนวกที่ 25 ตารางแอนโนวาที่ได้จากทดลองใช้ฮอร์โมน 4 ชนิดเลี้ยงไก่ 4 โรง เรือน โรงเรือนละ 4 ตัว

SV	*df*	*SS*	MS	*F-ratio*
ฮอร์โมน	3	0.3500	0.1167	$F_{col} = 15.77$
โรงเรือน	3	0.0500	0.167	$F_{blk} = 2.25$
	9	0.0414+0.0253	0.0074	
รวม	15	0.4469		

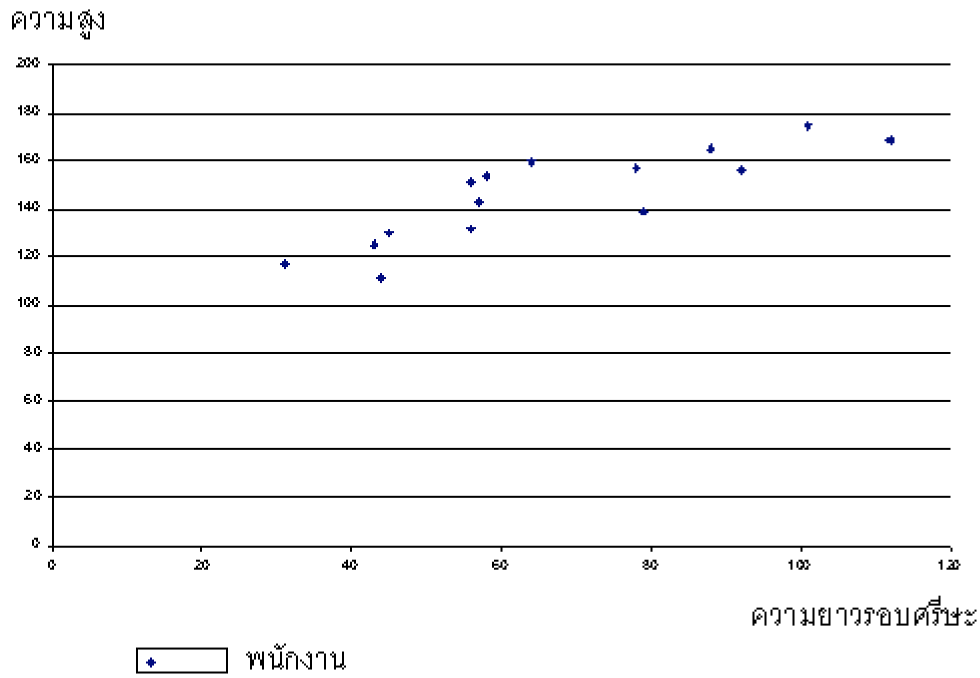
ที่ *df* 3,9 $F_{.01} = 6.99$

$F_{.05} = 3.68$

สรุปผลไดว่า ฮอร์โมน 4 ชนิดมีผลทำให้ไก่เจริญเติบโตต่างกัน แต่โรงเรือนไม่มีผลทำให้ไก่เจริญเติบโตต่างกัน

เฉลยคำถามทบทวนบทที่ 10

1. สร้างแผนภูมิจำนวนความสูงและความยาวรอบศีรษะของพนักงาน 15 คน ดังภาพภาคผนวกที่ 34 พบว่ามีความสัมพันธ์กันในทางบวก



ภาพภาคผนวกที่ 34 ความสัมพันธ์ของความสูงกับความยาวรอบศีรษะ

2. หาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

จากสูตร $$r = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right]\left[N\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

จากตารางภาคผนวกที่ 24 นำมาแทนค่าในสูตรจะได้

$$r = \frac{635{,}550}{\sqrt{(1{,}247{,}484)(360{,}620)}}$$

$$= 0.95$$

ตารางภาคผนวกที่ 26 การหาค่าต่าง ๆ เพื่อหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

พนักงาน	X	X^2	y	Y^2	XY
1	111	12,321	44	2,401	4,884
2	117	13,689	31	961	3,627
3	125	15,625	43	1,849	5,375
4	130	16,900	45	2,025	5,850
5	132	17,424	56	3,136	7.392
6	139	19,321	79	6,241	10,981
7	143	20,499	57	3,249	8,151
8	151	22,801	56	3,136	8,456
9	154	23,716	58	3,364	8,932
10	156	24,336	92	8,464	14,352
11	157	24,649	78	6,084	12,246
12	160	25,600	64	4,096	10,240
13	165	27,225	88	7,744	14,520
14	169	28,561	112	12,544	18,928
15	175	28,900	101	10,201	17,675
รวม	1,569	247,283	785	65,123	124,481

3. หาค่าสัมประสิทธิ์ลำดับสเปียร์แมนโดยนำข้อมูลมาเขียนตารางภาคผนวกที่ 25

ตารางภาคผนวกที่ 27 การหาค่าต่าง ๆ เพื่อหาค่าสัมประสิทธิ์ลำดับสเปียร์แมนของข้อมูลจากตารางที่ 10.19

พนักงาน	X	Y	ลำดับที่ X	ลำดับที่ Y	D_i	D_i^2
1	111	44	1	3	2	4
2	117	31	2	1	-1	1
3	125	43	3	2	-1	1
4	130	45	4	4	0	0
5	132	56	5	5.5	0.5	0.25
6	139	79	6	11	5	25
7	143	57	7	7	0	0
8	151	56	8	5.5	-2.5	6.25
9	154	58	9	8	-1	1
10	156	92	10	13	3	9
11	157	78	11	10	-1	1
12	160	64	12	9	-3	9
13	165	88	13	12	-1	1
14	169	112	14	15	1	1
15	175	101	15	14	-1	1
						$\sum_{i=1}^{n} D_i^2 = 60.5$

$$r = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)} = 1 - \frac{6\left(60.5\right)}{15\left(\left(15\right)^2 - 1\right)}$$

$$= 1 - \frac{363}{3,360} = 0.89$$

4. หาค่าสัมประสิทธิ์ลำดับสเปียร์แมนของจำนวนวันลากับระยะทางฯ โดยนำข้อมูลมาเขียนตารางได้เป็นตารางภาคผนวกที่ 26

ตารางภาคผนวกที่ 28 การหาค่าต่าง ๆ เพื่อหาค่าสัมประสิทธิ์ลำดับสเปียร์แมนของข้อมูลจากตารางที่ 10.10

พนักงาน	**X**	*Y*	ลำดับที่ *X*	ลำดับที่ *Y*	D_i	D_i^2
1	21	6.8	9.5	9	-0.5	0.25
2	12	10.3	4	14	9.5	90.25
3	30	1.7	13	1	-12	144
4	8	14.2	2	16	14	196
5	10	8.8	3	12	9	81
6	26	5.8	12	8	-4	16
7	42	2.1	16	2	-14	196
8	31	3.3	14	4	-10	100
9	21	4.3	9.5	6	-3.5	12.25
10	15	9.0	6	13	7	49
11	19	3.2	8	3	-5	25
12	6	12.7	1	15	14	196
13	18	8.2	7	11	4	16
14	12	7.0	4.5	10	5.5	30.25
15	23	5.1	11	7	-4	16
16	34	4.1	15	5	-10	100
						$\sum_{i=1}^{n} D_i^2 = 1,268$

$$r = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)} = 1 - \frac{6\left(1,268\right)}{16\left(\left(16\right)^2 - 1\right)}$$

$$= 1 - \frac{7,608}{4,080} = -0.86$$

5. ทดสอบความสัมพันธ์ระหว่างความเห็นของประธานกับรองประธานบริษัท

ให้ X เป็น คะแนนจากประธานบริษัท

Y เป็น คะแนนจากรองประธานบริษัท

ρ เป็น ค่าสหสัมพันธ์ระหว่าง X กับ Y

ตารางภาคผนวกที่ 29 การหาค่าต่าง ๆ เพื่อหาค่าสหสัมพันธ์ของคะแนนจากประธานบริษัทและรองประธานบริษัท

พนักงาน	X	X^2	y	Y^2	XY
1	8	64	2	4	16
2	2	4	9	81	18
3	7	49	5	25	35
4	4	16	10	100	40
5	3	9	11	121	33
6	1	1	12	144	12
7	9	81	4	16	36
8	12	144	3	9	36
9	6	36	7	49	42
10	11	121	4	16	44
11	15	225	8	64	120
12	10	100	1	1	10
รวม	88	850	76	630	442

จากสูตร $$r = \frac{N\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[N\sum X^2 - \left(\sum X\right)^2\right]\left[N\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

จากตารางภาคผนวกที่ 27 นำมาแทนค่าในสูตรจะได้

$$r = \frac{1,384}{\sqrt{(2,456)(1,784)}} = -0.66$$

$H_0 \ : \ \rho = 0$

$H_1 \ : \ \rho \neq 0$

$n = 12$, $df = n-2 = 10$

คำนวณค่าสถิติทดสอบ

$$t_e = r\sqrt{\frac{(n-2)}{(1-r^2)}} = -0.66\sqrt{\frac{12-2}{1-(0.66)^2}}$$

$$t_e = -1.567$$

จากตารางภาคผนวกที่ 6 $t_{10,.005} = -3.169$ เมื่อค่า t_e อยู่ในเขตยอมรับ H_0 ดังนั้นเชื่อมั่นได้ร้อยละ 99 ว่าคะแนนจากประธานบริษัทและรองประธานบริษัทไม่สัมพันธ์กัน

6. ทดสอบความสัมพันธ์ระหว่างจำนวนครั้งที่โฆษณากับยอดขายรวม

ให้ X เป็น จำนวนครั้งที่โฆษณา

Y เป็น ยอดขายรวม (ล้านบาท)

ρ เป็น ค่าสหสัมพันธ์ระหว่าง X กับ Y

ตารางภาคผนวกที่ 30 การหาค่าต่าง ๆ เพื่อหาค่าสหสัมพันธ์ของจำนวนครั้งที่โฆษณากับยอดขาย

เดือนที่	X	X^2	y	Y^2	XY
1	1	1	2	4	2
2	2	4	1	1	2
3	3	9	3	9	9
4	4	16	3	9	12
5	5	25	4	16	20
6	6	36	5	25	30
7	7	49	6	36	42
8	8	64	5	25	40
9	9	81	7	49	63
รวม	45	285	36	174	220

จากสูตร

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

จากตารางภาคผนวกที่ 26 นำมาแทนค่าในสูตรจะได้

$$r = \frac{360}{\sqrt{(540)(270)}} = 0.94$$

$H_0 \quad : \quad \rho = 0$

$H_1 \quad : \quad \rho \neq 0$

$n \quad = \quad 9 \,, \quad df = n-2 = 7$

คำนวณค่าสถิติทดสอบ

$$t_e = r\sqrt{\frac{(n-2)}{(1-r^2)}} = 0.94\sqrt{\frac{9-2}{1-(0.94)^2}} = 4.61$$

จากตารางภาคผนวกที่ 6 $t_{7,.005} = 3.499$ เมื่อค่า t_e อยู่ในเขตปฏิเสธ H_0
ดังนั้นเชื่อมั่นได้ร้อยละ 99 ว่าจำนวนครั้งที่โฆษณากับยอดขายสัมพันธ์กัน

7. สร้างสมการถดถอย

ให้ X เป็น ค่าใช้จ่ายในการโฆษณา (ล้านบาท)

Y เป็น รายได้รวม (ล้านบาท)

ตารางภาคผนวกที่ 31 การหาค่าต่าง ๆ เพื่อหาค่าสหสัมพันธ์ของรายได้รวมและค่าใช้จ่ายในการโฆษณา

สัปดาห์ที่	X	X^2	y	XY
1	3.2	10.24	114.6	366.7
2	2.5	6.25	86.5	216.3
3	2.3	5.29	95.5	219.7
4	2.4	5.76	61.8	148.3
5	3.5	12.25	154.8	541.8
6	3.2	10.24	120.9	386.9
7	2.6	6.76	104.2	270.9
8	3.0	9.00	136.8	410.4
9	2.8	7.84	143.5	401.8
10	2.5	6.25	98.7	246.8
11	3.2	10.24	128.2	410.2
12	2.7	7.29	99.2	267.8
13	2.6	6.76	107.3	279.0
14	3.1	9.61	130.1	403.3
15	2.8	7.84	150.2	402.6
รวม	28.5	81.8	1219.1	3,497.7
ค่าเฉลี่ย	2.8		115.5	

สมการถดถอยของรายได้จากการขายกับค่าใช้จ่ายในการโฆษณาคือ

$$\hat{Y} = b_0 + bX$$

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

$$= \frac{15(3,497.7) - (28.5)(1,219.1)}{15(81.8) - (28.5)^2} = 42.7$$

$$b_0 = \bar{Y} - b\bar{X} = 115.5 - (42.7)(2.8) = -5.2$$

$$\hat{Y} = -5.2 + 42.7X$$

8. จากตารางภาคผนวกที่ 27 ในเฉลยคำถามทบทวนข้อ 5 สมการถดถอยของคะแนนรองประธานบริษัท (Y) และประธานบริษัท (X) คือ

$$\hat{Y} = b_0 + bX$$

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

$$= \frac{12(442) - (88)(76)}{12(850) - (88)^2} = -0.564$$

$$b_0 = \bar{Y} - b\bar{X} = 6.3 - (-0.564)(7.33) = 2.165$$

$$\hat{Y} = 2.165 + 0.564X$$

9. กรณียอดขายรวมขึ้นอยู่กับจำนวนพนักงานรวมของบริษัท จำนวนของสินค้า และค่าใช้จ่ายในการโฆษณา

ถ้ากำหนดให้ Y เป็น ยอดขายรวมของบริษัท

X_1 เป็น จำนวนพนักงานรวมของบริษัท

X_2 เป็น จำนวนสินค้า

X_3 เป็น ค่าใช้จ่ายในการโฆษณา

สมการการถดถอยพหุคูณคือ

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

10. ข้อเสียการพยากรณ์ตัวแปรตามด้วยสมการการถดถอยคือ
   1. เกิดข้อผิดพลาดได้สูง เพราะ ตัวแปรตามจะเกิดขึ้นในอนาคต แต่ตัวแปรต้นเป็นตัวแปรที่เกิดขึ้นในอดีต
   2. ภาวะปัจจุบันมีการเปลี่ยนแปลงที่รวดเร็วและบางครั้งไม่เอื้ออำนวยต่อการใช้ข้อมูลในอดีต เป็นต้น

11. สร้างตารางเพื่อหา E_{ij} จะได้ตารางภาคผนวกที่ 32

ตารางภาคผนวกที่ 32 ข้อมูลแจงนับที่คาดหวังการลากิจและลาป่วยของพนักงานในแผนกอาหารสดขึ้นกับเหตุจูงใจที่เข้าทำงาน

เหตุจูงใจ	การลา		รวม R_i
	ลาป่วย	ลากิจ	
ชอบอาชีพนี้	178.01	105.99	284
ได้รับทุนแผนกนี้	114.70	68.29	183
ไม่ทราบว่าจะทำอะไร	35.10	20.90	56
ทำชั่วคราว	18.18	10.82	29
รวม C_j	346	206	552

จากตารางภาคผนวกที่ 32 คำนวณค่า $\chi^2_{cal} = 0.4108$ แต่ $\chi^2_{.05,3} = 7.815$
สรุปได้ว่า การลาป่วยและลากิจของพนักงานไม่ขึ้นเหตุจูงใจที่เข้าทำงาน

ดัชนี