ความน่าจะเป็นและสถิติ

Probability and Statistics

ธนัชยศ จำปาหวาย

คณะครุศาสตร์

มหาวิทยาลัยราชภัฏสวนสุนันทา

2565

ชื่อหนังสือ : ความน่าจะเป็นและสถิติ

ชื่อผู้แต่ง : ผู้ช่วยศาสตราจารย์ ดร.ธนัชยศ จำปาหวาย

จัดทำโดย : ผู้ช่วยศาสตราจารย์ ดร.ธนัชยศ จำปาหวาย

พิมพ์ครั้งที่ : 1

เดือนและปี พ.ศ. ที่จัดพิมพ์ : ตุลาคม 2565

ราคา 250 บาท

ISBN (e-book) : 978-616-594-463-2

คำนำ

ตำราความน่าจะเป็นและสถิติเล่มนี้ ครอบคลุมเนื้อหารายวิชาความน่าจะเป็นและสถิติ รหัสวิชา MAC1304 เป็นวิชาเอกเลือกในหลักสูตรครุศาสตรบัณฑิต สาขาวิชาคณิตศาสตร์ (4 ปี) (หลักสูตรปรับปรุง พ.ศ.2562) โดยมีวัตถุประสงค์เพื่อเป็นแหล่งเรียนรู้ของรายวิชาดังกล่าว ประกอบไปด้วยเนื้อหาที่แบ่งออกเป็น 10 บท คือ ความรู้พื้นฐานเกี่ยวกับสถิติ การวิเคราะห์ข้อมูลเบื้องต้น ความน่าจะเป็น ตัวแปรสุ่ม การแจกแจงความน่าจะเป็นไม่ต่อเนื่อง การแจกแจงความน่าจะเป็นต่อเนื่อง การสุ่มตัวอย่าง การประมาณค่า การทดสอบสมมติฐาน และการถดถอยเชิงเส้นและสหสัมพันธ์ โดยคำศัพท์คณิตศาสตร์ที่ใช้ในเล่มนี้ ใช้ตามพจนานุกรมศัพท์คณิตศาสตร์ ฉบับราชบัณฑิตยสภา พ.ศ. 2559

เนื้อหาในตำราเล่มนี้ได้เรียบเรียงมาจากประสบการณ์ในการสอนวิชาความน่าจะเป็นและสถิติ ในบางส่วนต้องอาศัยความรู้แคลคูลัสเบื้องต้น โดยเฉพาะการพิสูจน์ที่มาของแต่ละทฤษฎีบทซึ่งจะทำให้ผู้เรียนเข้าใจที่มาได้ชัดเจนยิ่งขึ้น ผู้เขียนมุ่งหวังว่าตำราเล่มนี้จะเป็นเอกสารประกอบการเรียนของนักศึกษา สาขาวิชาคณิตศาสตร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา เพื่อนำไปเรียนในวิชาดังกล่าวได้ดียิ่งขึ้น

เนื้อหาที่กล่าวในตำราวิชา "ความน่าจะเป็นและสถิติ" เป็นคณิตศาสตร์เชิงประยุกต์ซึ่งจะเป็นพื้นฐานสำคัญในการเรียนรู้คณิตศาสตร์ประยุกต์ในด้านต่าง ๆ เช่น การวิจัยทางการศึกษา คณิตศาสตร์การเงิน คณิตศาสตร์ประกันภัย เป็นต้น

ตำราเล่มนี้จะเกิดขึ้นไม่ได้ถ้าไม่มีครูอาจารย์ที่คอยอบรมสั่งสอน และเป็นต้นแบบของการสอนวิชาความน่าจะเป็นและสถิติแก่ผู้เขียน คณาจารย์ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย สุดท้ายขอบคุณบิดามารดาและพี่สาวที่คอยเป็นกำลังเสมอมา

ธนัชยศ จำปาหวาย
2565

สารบัญ

หน้า

สารบัญตาราง

ตารางที่ หน้า

สารบัญรูป

รูปที่ หน้า

สารบัญรูป (ต่อ)

รูปที่ หน้า

บทที่ 1

ความรู้พื้นฐานเกี่ยวกับสถิติ

สถิติเกี่ยวข้องกับชีวิตประจำวันเราอยู่เสมอไม่ว่าทางตรงหรือทางอ้อม เช่น แม่ต้องทราบปริมาณการทานอาหารของครอบครัวในแต่ละมื้อเพื่อที่จะได้ทำเตรียมอาหารได้เพียงพอกับคนในครอบครัว หรือการจดจำเวลาที่เราเดินทางจากบ้านไปยังมหาวิทยาลัยในช่วงภาวะการต่าง ๆ เพื่อที่จะได้นำข้อมูลที่มีใช้ในการจัดการเดินทางในครั้งถัดไป ให้มีประสิทธิภาพมากขึ้น เป็นต้น จะเห็นได้ว่าสถิติมีความสำคัญในการแก้ปัญหาและพัฒนากระบวนการบางอย่างให้ดียิ่งขึ้น ในหัวข้อนี้จะกล่าวถึงความรู้พื้นฐานเกี่ยวกับสถิติที่ควรทราบ ก่อนจะศึกษาในรายละเอียดอื่น ๆ ในบทถัดไป

1.1 ความหมายของสถิติ

คำว่า **สถิติ** ตรงกับคำในภาษาอังกฤษ Statistics ซึ่งเป็นคำที่แปลมาจากศัพท์ Statistik ในภาษาเยอรมัน เป็นคำที่มีรากศัพท์เดียวกับคำว่า “State” ซึ่งแปลว่า “รัฐ” มีความหมายถึงข้อมูลหรือข่าวสาร ซึ่งเป็นประโยชน์ต่อการบริหารงานของรัฐในด้านต่าง ๆ สถิติเริ่มรู้จักในวงการวิชาการช่วงต้นศตวรรษที่ 18 โดยอับราฮัม เดอ มัวร์ (Abraham de Moive : 1667-1754) ได้นำเสนอสมการการแจกแจงปกติ ซึ่งต่อมา คาร์ล ฟรีดริช เกาส์ (Carl Fredrich Gauss : 1777-1855) ก็ได้สมการนี้จากการศึกษาเรื่องความคลาดเคลื่อนในการวัดปริมาณหลาย ๆ ครั้ง ในปัจจุบันมีนักวิชาการให้ความหมายของคำว่าสถิติไว้หลายความหมาย เช่น

1. สถิติหมายถึงตัวเลขที่แสดงข้อเท็จจริงเกี่ยวกับเรื่องใดเรื่องหนึ่ง เช่น
   - สถิติที่แสดงปริมาณน้ำฝนในกรุงเทพมหานครในเดือนมกราคม พ.ศ. 2563
   - สถิติอุบัติเหตุที่เกิดในประเทศไทย พ.ศ.2562
   - สถิตินักเรียนที่มาสายของโรงเรียนแห่งหนึ่งในปีการศึกษา 2560
2. สถิติหมายถึงค่าตัวเลขที่คำนวณได้จากข้อมูลกลุ่มตัวอย่าง ค่าที่คำนวณได้ออกมานั้นเรียกว่า ค่าสถิติ เช่น ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน ค่าฐานนิยม และค่าส่วนเบี่ยงเบนมาตรฐาน เป็นต้น
3. สถิติหมายถึงศาสตร์ที่ว่าด้วยการเก็บรวมรวมข้อมูล การนำเสนอข้อมูล และการวิเคราะห์ข้อมูล

จากความหมายของสถิติ จะเห็นว่าการตัดสินใจด้านต่าง ๆ ไม่ว่าจะเป็นองค์กรของรัฐ หรือเอกชน หรือหน่วยงานต่าง ๆ จำเป็นต้องใช้ข้อมูลทางสถิติเข้ามาช่วยในการตัดสินใจเพื่อให้การตัดสินใจนั้นถูกต้องมากยิ่งขึ้น เช่นสาขาวิชาคณิตศาสตร์ของมหาวิทยาลัยแห่งหนึ่ง ต้องการจะเพิ่มอาจารย์ในสาขา จะพิจารณาจากอัตราจำนวนอาจารย์ต่อนักศึกษาตามเกณฑ์สำนักงานคณะกรรมการอุดมศึกษา (สกอ.) คือ 1:30 นั่นคืออาจารย์ 1 คน ต่อนักศึกษา 30 คน จากข้อมูลของสาขาพบว่ามีอัตราส่วน 1:46 ดังนั้น สาขาวิชาต้องเพิ่มอาจารย์ในสาขาให้ได้สัดส่วนตามเกณฑ์มาตรฐานเป็นต้น

สถิติแบ่งออกเป็น 2 ประเภทคือ สถิติเชิงพรรณา และ สถิติเชิงอนุมาน ซึ่งจะกล่าวรายละเอียดในหัวข้อการวิเคราะห์ข้อมูล

1.2 ข้อมูลทางสถิติ

ข้อมูล (Data) คือ สิ่งที่แสดงถึงลักษณะของข้อเท็จจริงเกี่ยวกับบุคคล สิ่งของหรือเหตุการณ์ในรูปแบบของตัวเลข ภาพ ตัวอักษร และสัญลักษณ์ต่าง ๆ ซึ่งอาจเป็นตัวเลขหรือไม่ก็ได้ และต้องจำนวนมากพอที่จะแสดงลักษณะของเรื่องนั้นได้ เช่น อายุ น้ำหนัก ความสูง รายได้ เพศ ระดับการศึกษา อาชีพ เป็นต้น ข้อมูลที่เก็บรวบรวมมาได้เรียกว่า **ข้อมูลดิบ** (raw data) จากนั้นจะนำข้อมูลไปวิเคราะห์และประมวลผลต่อไป ซึ่งจะเรียกว่า **สารสนเทศ** (Information) ดังนั้นสารสนเทศหมายถึง ข้อมูลได้ผ่านการเปลี่ยนแปลงหรือมีการประมวลผลหรือวิเคราะห์สรุปผลด้วยวิธีการต่าง ๆ

แสดงตัวอย่างดังนี้

ตารางที่ 1.1: ตัวอย่างข้อมูลดิบและสารสนเทศ

ข้อมูลดิบ	สารสนเทศ
- คะแนนนักเรียนในห้อง	- นักเรียนส่วนใหญ่ในห้องนี้มีคะแนนสูงกว่าเกณฑ์
- รายได้ของครูในพื้นที่ กทม.	- ครูในพื้นที่ กทม. มีรายได้เฉลี่ย 35,000 ต่อเดือน
- จำนวนพายุที่ผัดผ่านประเทศไทย	- โดยเฉลี่ยแล้วพายุจะผัดผ่านประเทศไทยปีละ 15 ครั้ง
- จำนวนคนที่กดไลค์เฟสบุค	- คนที่กดไลค์เพจแมทโคตรส่วนใหญ่มีอายุช่วง 20-30 ปี

ข้อมูลที่นำมาวิเคราะห์ประกอบด้วยข้อมูลหลายประเภท อาจพิจารณาตามลักษณะต่าง ๆ ดังนี้

แบ่งตามลักษณะของข้อมูล

1. **ข้อมูลเชิงปริมาณ** (Quantitative data) คือข้อมูลที่วัดค่าได้ว่ามากหรือน้อย จึงแสดงเป็นตัวเลข เช่น รายได้ น้ำหนัก ส่วนสูง คะแนนสอบ อายุ เป็นต้น แบ่งออกเป็น 2 แบบคือ

   1.1 **ข้อมูลแบบไม่ต่อเนื่อง** (Discrete data) หมายถึงข้อมูลที่มีค่าเป็นจำนวนเต็มหรือจำนวนนับ

   1.2 **ข้อมูลแบบต่อเนื่อง** (Continunous data) หมายถึงข้อมูลที่มีค่าได้ทุกค่าในช่วงที่กำหนดที่มีความหมาย

แสดงตัวอย่างดังนี้

ตารางที่ 1.2: ตัวอย่างข้อมูลแบบไม่ต่อเนื่องและข้อมูลแบบต่อเนื่อง

ข้อมูลแบบไม่ต่อเนื่อง	ข้อมูลแบบต่อเนื่อง
- จำนวนไข่ไก่ในตะกร้า	- น้ำหนักของไข่ไก่ 1 ฟอง
- จำนวนเด็กในห้องเรียน	- ส่วนสูงของนักเรียน
- จำนวนพายุที่ผัดผ่านในฤดูกาลหนึ่ง	- ความเร็วลม
- จำนวนคนที่กดไลค์เฟสบุค	- อุณหภูมิของน้ำ

2. **ข้อมูลเชิงคุณภาพ** (Qualitative data) หรือข้อมูลจำแนกประเภท (Categorical data) หรือบางครั้งเรียกว่าข้อมูลเชิงกลุ่มเป็นข้อมูลที่ไม่สามารถระบุค่าได้ว่ามากหรือน้อย มักจะเป็นข้อความ เช่น ระดับการศึกษา การนับถือศาสนา เพศ เป็นต้น

แบ่งตามแหล่งที่มาของข้อมูล

1. **ข้อมูลปฐมภูมิ** (Primary data) คือข้อมูลที่เก็บรวบรวมด้วยตัวเองหรือจากแหล่งที่ให้ข้อมูลโดยตรง เช่น การทดลอง การสอบถาม การทำสำมะโน การสัมภาษณ์ เป็นต้น ข้อมูลปฐมภูมิจะเป็นข้อมูลที่มีรายละเอียดตรงตามที่ผู้ใช้ต้องการ อาจเสียเวลาและค่าใช้จ่ายมาก และเป็นข้อมูลยังไม่ทำการวิเคราะห์

2. **ข้อมูลทุติยภูมิ** (Secondary data) คือข้อมูลที่รวบรวมได้จากแหล่งอื่นที่มีผู้รวบรวมไว้ให้ เช่น รายงานต่าง ๆ ของหน่วยราชการหรือเอกชน เป็นต้น ผู้ใช้ข้อมูลเพียงนำข้อมูลมาใช้เท่านั้นจึงประหยัดทั้งเวลาและค่าใช้จ่าย ดังนั้นข้อมูลทตุยภูมิคือข้อมูลที่ทำการวิเคราะห์เบื้องต้นมาแล้ว การนำข้อมูลชนิดนี้มาใช้อาจมีรายละเอียดไม่เพียงพอ หรือไม่ตรงตามต้องการ ดังนั้นผู้ใช้อาจไม่ทราบถึงข้อผิดพลาดของข้อมูล จึงควรระมัดระวังในการใช้

แสดงตัวอย่างดังนี้

ตารางที่ 1.3: ตัวอย่างข้อมูลปฐมภูมิและข้อมูลทุติยภูมิ

ข้อมูลปฐมภูมิ	ข้อมูลทุติยภูมิ
- คะแนนที่ได้จากการสอบถามเพื่อน ๆ ในห้อง	- คะแนนที่ได้จากการบันทึกของครูผู้สอน
- อุณหภูมิของสารต่าง ๆ ที่ได้จากทดลอง	- ค่าความชื้นในอากาศที่ได้จากเว็ปไซค์ของกรมอุตุนิยมวิทยา
- จำนวนนักเรียนที่มาสายในแต่ละวันจากการสังเกตหน้าโรงเรียน	- น้ำหนักของผู้สูงวัยใน กทม. ที่ได้จากสำนักสถิติแห่งชาติ

หากต้องการทราบข้อมูลทางสถิติบางอย่าง เช่น รายได้ของคนไทย สิ่งแรกที่ต้องทำคือการสำรวจหรือเก็บข้อมูลของรายได้ของคนไทย จะเรียกคนไทยทุกคนว่า **ประชากร** (Population) ของการสำรวจในครั้งนี้ ถ้าสนใจน้ำหนักของนักเรียนในกรุงเทพมหานคร ประชากรคือ นักเรียนทุกคนในกรุงเทพมหานคร ดังนั้นประชากรในทางสถิติจึงมีความหมายกว้างกว่าความคำว่าประชากรโดยทั่วไป ดังจะนิยามดังนี้

บทนิยาม 1.2.1 ประชากร คือเซตของค่าสังเกตทั้งหมดในการทดลอง หรือในการสำรวจทางสถิติ

จำนวนของค่าสังเกตทั้งหมดของประชากร คือตัวเลขที่บอกขนาดของประชากร เช่นสำรวจคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้น ม. 3 ของโรงเรียนแห่งหนึ่งที่มีจำนวนทั้งหมด 450 คน จะได้ว่าขนาดของประชากรในการสำรวจครั้งนี้คือ 450 ประชากรอาจแบ่งตามขนาดได้ 2 ชนิดคือ

1. **ประชากรที่มีจำนวนแน่นอน** (Finite population) เช่น จำนวนรถโดยสารประจำทางในกรุงเทพมหานคร จำนวนนักศึกษาสาขาวิชาคณิตศาสตร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา เป็นต้น

2. **ประชากรที่มีจำนวนอนันต์** (Infinite population) เช่น จำนวนข้าวที่เก็บเกี่ยวได้ในปีหนึ่ง ๆ และจำนวนครั้งที่ทอดลูกเต๋าจนได้แต้ม 5 เป็นต้น

จะเห็นว่าการเก็บรวบรวมข้อมูลทุกหน่วยในประชากรที่มีขนาดใหญ่ อาจเกิดความยุ่งยาก ใช้เวลามากและค่าใช้จ่ายสูง ผู้วิจัยอาจเก็บข้อมูลบางส่วนของประชากร ซึ่งจะเรียกว่า **ตัวอย่าง** (Sample) เช่น ต้องการหาอายุเฉลี่ยของประชากรไทย ประชากรคือคนไทยทุกคน ตัวอย่างคือคนไทยบางส่วนที่ถูกเลือกเป็นตัวอย่าง หรือนิยามได้ดังนี้

บทนิยาม 1.2.2 ตัวอย่าง คือเซตย่อยของประชากรที่ไม่ใช่เซตว่าง

การเก็บรวบรวมข้อมูลเพื่อนำมาวิเคราะห์ อาจประกอบด้วยข้อมูลปฐมภูมิ และข้อมูลทุติยภูมิ โดยข้อมูลที่ได้อาจเป็นข้อมูลเชิงปริมาณ และเชิงคุณภาพ ในกรณีการใช้ข้อมูลปฐมภูมิหน่วยงานจะเป็นผู้เก็บข้อมูล ซึ่งมีวิธีเก็บรวบรวม 3 วิธีดังนี้

1. การเก็บรวบรวมข้อมูลจากทะเบียนหรือการบันทึก
   คือการจดบันทึกหรืองานทะเบียนของหน่วยงาน หรือองค์กรต่าง ๆ เช่น การบันทึกจำนวนนักเรียนที่มาเรียนในแต่ละวัน ของโรงเรียนแห่งหนึ่ง การบันทึกข้อมูลผู้มารักษาในโรงพยาบาลแห่งหนึ่ง เช่น อายุ เพศ และหมู่เลือด เป็นต้น ห้างสรรพสินค้าจำบันทึกยอดขายของสินค้าแต่ละแผนกทุกวัน

2. การเก็บรวมรวมข้อมูลโดยการสำรวจ
   ทำได้ 2 แบบคือ

   2.1 **การทำสำมะโน** (Census) คือข้อมูลที่เก็บรวบรวมข้อมูลทุก ๆ หน่วยของประชากร

   2.2 **การสำรวจด้วยตัวอย่าง** (Sample survey) คือการเก็บข้อมูลบางหน่วยที่เลือกมาเป็นตัวแทนจากทุก ๆ หน่วยของประชากร

3. การเก็บรวบรวมข้อมูลโดยการทดลอง
   ในบางครั้งเรื่องที่เราสนใจไม่สามารถทำการสำรวจได้ แต่ต้องเก็บข้อมูลโดยทำการทดลอง เช่น เปรียบเทียบการสอนวิชาคณิตศาสตร์เรื่องจำนวนเต็ม 2 รูปแบบ หรือศึกษาผลสัมฤทธิ์ของวิชาคณิตศาสตร์เรื่องพื้นที่ผิวและปริมาตรของนักเรียนชั้น ม.4 โรงเรียนแห่งหนึ่งโดยใช้สื่อประสมร่วมกับโปรแกรม GeoGebra

การเก็บรวบรวมข้อมูลทำได้หลายแบบ เช่น

1. การสัมภาษณ์ (Interview)
   การส่งพนักงานไปสัมภาษณ์หน่วยต่าง ๆ ในประชากรหรือตัวอย่าง พนักงานจะเป็นผู้จดบันทึกคำตอบในแบบสอบถาม

2. การส่งไปรณีย์ (Mail)
   การส่งแบบสอบถามทางไปรษณีย์ไปให้ผู้ตอบแบบสอบถามที่ถูกเลือก เมื่อผู้ตอบแบบสอบถามแล้วให้ส่งคืนมาทางไปรษณีย์

3. การทอดแบบ
   การนำแบบสอบถามไปให้ผู้ตอบ แล้วนัดมารับแบบสอบถามคืน

4. โทรศัพท์
   พนักงานใช้โทรศัพท์ไปสอบถามผู้ตอบ หรือผู้ที่ถูกเลือกเป็นตัวอย่าง

5. การชั่ง ตวง วัด หรือนับ
   การศึกษาข้อมูลบางเรื่องที่ต้องการตัวเลขที่แน่นอน เช่นน้ำหนักของเด็กแรกเกิด ขนาดขแงแปลงเพราะปลูก จึงต้อง ชั่ง ตวง วัด หรือนับ จึงจะได้ข้อมูลที่ถูกต้อง

6. การสังเกต
   การเก็บรวบรวมข้อมูลที่พนักงานสนามต้องไปสังเกตการณ์ในปฏิกิริยาต่าง ๆ ส่วนใหญ่วิธีนี้จะใช้เมื่อไม่สามารถใช้วิธีอื่นได้ เช่นการสังเกตความพอใจในรสชาติของกาแฟยี่ห้อหนึ่ง การสังเกตความพึงพอใจต่อการจัดการเรียนการสอนของวิชาหนึ่ง เป็นต้น

หลังจากการเก็บรวบรวมข้อมูล จะต้องทำการวิเคราะห์ข้อมูลเพื่อนำไปสรุปผล การวิเคราะห์ข้อมูลอาจแบ่งออกเป็น 2 ชนิดคือ การวิเคราะห์ข้อมูลเบื้องต้น และการวิเคราะห์ข้อมูลขั้นสูง

1. การวิเคราะห์ข้อมูลเบื้องต้นหรือเรียกว่า **สถิติเชิงพรรณา** (Descriptive Statistics) เป็นการวิเคราะห์ขั้นต้นที่มุ่งวิเคราะห์เพื่อหาลักษณะกว้าง ๆ ของข้อมูลชุดนั้น ซึ่งข้อสรุปและผลที่ได้จะพรรณนาลักษณะหรือแจกแจงข้อมูลตามที่ได้รวบรวมมาเท่านั้น มักนำเสนอในรูปของตาราง แผนภาพ แผนภูมิ ร้อยละ เปอร์เซ็นไทล์ การแจกแจงความถี่ การหาค่าเฉลี่ย เป็นต้น จะกล่าวในบทที่ 2

2. การวิเคราะห์ข้อมูลขั้นสูงหรือเรียกว่า **สถิติเชิงอนุมาน** (Inferential Statistics) เป็นการวิเคราะห์ข้อมูลที่เก็บรวบรวมได้จากตัวอย่างเพื่ออ้างอิงไปถึงข้อมูลทั้งหมดหรือประชากร การวิเคราะห์ในขั้นนี้ได้แก่ การประมาณค่า การทดสอบสมมติฐาน การวิเคราะห์ความแปรปรวน เป็นต้น จะกล่าวในบทที่ 4-10

การนำเสนอข้อมูลสถิติแบ่งออกเป็น 2 แบบ

1. การนำเสนอข้อมูลสถิติโดยปราศจากแบบแผน (Informal Presentation)
   - การนำเสนอข้อมูลสถิติเป็นบทความ
   - การนำเสนอข้อมูลสถิติเป็นบทความกึ่งตาราง
2. การนำเสนอข้อมูลสถิติโดยมีแบบแผน (Formal Presentation)

- การเสนอข้อมูลสถิติด้วยตาราง (Tabular Presentation)
- การเสนอข้อมูลสถิติด้วยกราฟและรูป (Graphic Presentation)

1.3 การแจกแจงความถี่

ในการเก็บรวบรวมคะแนนหรือข้อมูลต่าง ๆ ที่ได้จากการวัด เพื่อให้มีการจัดระเบียบเตรียมข้อมูลจัดให้เป็นหมวดหมู่ โดยการทำให้แปลความหรือนำไปใช้ได้ง่าย เช่น การเรียงข้อมูลจากมากไปหาน้อยหรือน้อยไปหามาก การทำตารางแจกแจงความถี่ การเขียนแผนภาพต้นใบ เป็นต้น

การสร้างตารางแจกแจงความถี่ของข้อมูลเชิงปริมาณสิ่งแรกคือจำนวนชั้นของ การตัดสินใจว่าควรมีจำนวนชั้นเท่าใดขึ้นอยู่กับจำนวนข้อมูล ไม่ควรมีน้อยเกินไปหรือมากเกินไป อย่างไรก็ตามโดยทั่วไปจะมีอย่างน้อย 5 ชั้น แต่ไม่เกิน 15 ชั้น กรณีที่ข้อมูลมีการกระจายมาก ๆ ควรกำหนดชั้นให้น้อยเพื่อไม่ให้บางชั้นไม่มีข้อมูลตกอยู่ในชั้นนั้น และในกรณีที่ไม่ทราบว่าจะกำหนดจำนวนชั้นเป็นเท่าใด ให้ k แทนจำนวนชั้น อาจใช้ **สูตรของสตูรเกส** (Sturges' formula) คำนวณดังนี้

$$k = 1 + 3.3\ell\text{og}N \qquad \text{เมื่อ } N \text{ เป็นจำนวนข้อมูลทั้งหมด}$$

ตัวอย่างการคำนวณจำนวนชั้นแสดงดังตารางต่อไปนี้

ตารางที่ 1.4: ตัวอย่างจำนวนชั้นเมื่อทราบจำนวนข้อมูล

จำนวนข้อมูล (N)	จำนวนชั้น (k)	จำนวนข้อมูล (N)	จำนวนชั้น (k)
10	5	45	7
15	5	50	7
20	6	60	7
25	6	70	8
30	6	80	8
35	7	90	8
40	7	100	8

ต่อไปคือองค์ประกอบของการสร้างตารางแจกแจงความถี่

1. **พิสัย** (Range) คือผลต่างระหว่างข้อมูลที่มีค่ามากสุดและค่าน้อยสุด เขียนแทนด้วย R นั่นคือ

$$R = \text{ ค่าสูงสุด } - \text{ ค่าต่ำสุด}$$

ถ้าข้อมูลมี N ประกอบไปด้วย $X_1, X_2, X_3, ..., X_N$ กำหนดให้

$$X_{\text{max}} = \text{ ค่าสูงสุดของข้อมูล } = \text{max}\{X_1, X_2, X_3, ..., X_N\}$$
$$X_{\text{min}} = \text{ ค่าต่ำสุดของข้อมูล } = \text{min}\{X_1, X_2, X_3, ..., X_N\}$$

จะได้ว่า

$$R = X_{\text{max}} - X_{\text{min}}$$

2. **อันตรภาคชั้น** (Class interval) คือช่วงคะแนนของแต่ละชั้นที่กำหนด ความกว้างของอันตรภาคชั้น เขียนแทนด้วย I

$$I = \frac{\text{พิสัย}}{\text{จำนวนชั้น}} = \frac{R}{k}$$

โดยที่ I เป็นจำนวนเต็มบวก กรณีที่ไม่เป็นจำนวนเต็มให้ใช้จำนวนเต็มใกล้สุดที่มากกว่าค่านั้น เช่น 5.7 ให้ใช้ 6 และ 9.1 ให้ใช้ 10 เป็นต้น

หมายเหตุ ในกรณีที่ผู้สร้างตารางกำหนดความกว้างชั้นไว้ก่อน จะหาจำนวนชั้นได้จาก

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

โดยที่ k เป็นจำนวนเต็มบวก ถ้าไม่เป็นจำนวนเต็มอาศัยหลักการเดียวกับการหา I

3. **ขีดจำกัดชั้น** (Class limit) การคำนวณขีดจำกัดล่างของชั้นแรก (ชั้นที่มีค่าต่ำสุด) ครอบคลุมข้อมูลที่มีค่าต่ำสุด และขีดจำกัดบนของชั้นสุดท้าย (ชั้นที่มีค่าสูงสุด) ครอบคลุมข้อมูลที่มีค่ามากสุด อาจใช้สูตรได้ดังนี้

$$\text{ขีดจำกัดล่างของชั้นแรก } = X_{\text{min}} - \frac{Ik - R}{2}$$

และปัดเศษให้มีลักษณะเหมือนข้อมูลจริง ดังนั้น

$$\text{ขีดจำกัดบนของชั้นแรก } = \text{ ขีดจำกัดล่างของชั้นแรก } + I - 1$$

4. **ขอบเขตจำกัดชั้น** (Class boundary) การหาขอบเขตชั้นจะกำหนดให้ขอบเขตชั้นมีจำนวนหลักหลังจุดทศนิยมมากกว่าค่าของข้อมูลจริงอยู่หนึ่งหลักเสมอ โดยคำนวนได้จาก

$$\text{ขอบเขตจำกัดชั้น } = \frac{\text{ขีดจำกัดบนของชั้น } + \text{ ขีดจำกัดล่างของชั้นถัดไป}}{2}$$

มักใช้ U_i แทนขอบเขตจำกัดบนของชั้นที่ i และ L_i แทนขอบเขตจำกัดล่างของชั้นที่ i ถ้าทุก ๆ ชั้นความกว้างเท่ากันคือ I จะได้ว่า

$$I = U_i - L_i$$

5. **ค่ากึ่งกลางชั้น** (Midpoint) คือข้อมูลที่อยู่ตรงกลางชั้นนั้น มักใช้ x_i แทนจุดกึ่งกลางของชั้นที่ i หาได้จาก

$$x_i = \frac{L_i + U_i}{2}$$

6. **ความถี่** (Frequency) คือจำนวนที่แสดงว่าค่าที่เป็นไปได้แต่ละค่าที่เกิดกี่ครั้ง มักใช้ f_i แทนความถี่ของชั้นที่ i หมายถึงจำนวนค่าที่เป็นไปได้ของชั้นที่ i

ตัวอย่าง 1.3.1 คะแนนสอบย่อยวิชาแคลคูลัส 1 ของนักศึกษาจำนวน 20 คน แสดงดังนี้

1	1	1	2	5	8	9	10	10	11
11	12	13	14	15	16	17	18	18	18

จงสร้างตารางแจกแจงความถี่ของข้อมูลชุดนี้

1. จำนวน 5 ชั้น

   แนวคำตอบ กำหนดให้ $k = 5$ จะเห็นว่า $X_{\text{max}} = 18$ และ $X_{\text{min}} = 1$ ดังนั้นความกว้างชั้นหาได้จาก

   $$I = \frac{R}{k} = \frac{X_{\text{max}} - X_{\text{min}}}{5} = \frac{18-1}{5} = 3.4 \approx 4$$

   จะได้ว่า

   $$\text{ขีดจำกัดล่างของชั้นแรก } = 1 - \frac{4(5)-17}{2} = -0.5 \approx 0$$
   $$\text{ขีดจำกัดบนของชั้นแรก } = 0 + 4 - 1 = 3$$

   ดังนั้น

คะแนนสอบย่อยวิชาแคลคูลัส 1	รอยขีด	จำนวนคน
$0-3$	$\vert\vert\vert\vert$	4
$4-7$	$\vert$	1
$8-11$	$\cancel{\vert\vert\vert\vert}$ $\vert$	6
$12-15$	$\vert\vert\vert\vert$	4
$16-19$	$\cancel{\vert\vert\vert\vert}$	5

2. ไม่ได้กำหนดจำนวนชั้น

   แนวคำตอบ เนื่องจาก $N = 20$ จากตาราง 1.4 จะได้ว่า $k = 6$ ดังนั้นความกว้างชั้นหาได้จาก

   $$I = \frac{R}{k} = \frac{X_{\text{max}} - X_{\text{min}}}{6} = \frac{18-1}{6} = 2.83 \approx 3$$

   จะได้ว่า

   $$\text{ขีดจำกัดล่างของชั้นแรก } = 1 - \frac{3(6)-17}{2} = 0.5 \approx 1$$
   $$\text{ขีดจำกัดบนของชั้นแรก } = 1 + 3 - 1 = 3$$

   ดังนั้น

คะแนนสอบย่อยวิชาแคลคูลัส 1	รอยขีด	จำนวนคน
$1-3$	$\vert\vert\vert\vert$	4
$4-6$	$\vert$	1
$7-9$	$\vert\vert$	2
$10-12$	$\cancel{\vert\vert\vert\vert}$	5
$13-15$	$\vert\vert\vert$	3
$16-18$	$\cancel{\vert\vert\vert\vert}$	5

3. กำหนดความกว้างแต่ละชั้นเท่ากับ 5

   แนวคำตอบ กำหนดให้ $I = 5$ จะได้จำนวนชั้นคือ

$$k = \frac{R}{I} = \frac{X_{\text{max}} - X_{\text{min}}}{5} = \frac{18-1}{5} = 3.4 \approx 4$$

   จะได้ว่า

$$\text{ขีดจำกัดล่างของชั้นแรก } = 1 - \frac{5(4)-17}{2} = -0.5 \approx 0$$

$$\text{ขีดจำกัดบนของชั้นแรก } = 0 + 5 - 1 = 4$$

   ดังนั้น

คะแนนสอบย่อยวิชาแคลคูลัส 1	รอยขีด	จำนวนคน
$0-4$	||||	4
$5-9$	|||	3
$10-14$	~~||||~~ ||	7
$15-19$	~~||||~~ |	6

ตัวอย่าง 1.3.2 อุณหภูมิ (หน่วยองศาเซลเซียส) ของสสารทั้ง 30 ชนิด แสดงดังนี้

0.5	0.8	1.5	1.6	1.8	1.9	2.0	2.1	2.3	2.5
2.7	3.0	3.3	3.4	3.4	3.5	3.7	3.8	3.9	3.9
4.0	4.0	4.1	4.2	4.2	4.3	4.5	4.8	4.9	5.0

จงสร้างตารางแจกแจงความถี่ของข้อมูลชุดนี้

แนวคำตอบ พิจารณาข้อมูลเป็นจำนวนเต็ม โดยใช้หน่วย $\times 10^{-1}$ องศาเซลเซียส แสดงข้อมูลในรูปจำนวนเต็มดังนี้

5	8	15	16	18	19	20	21	23	25
27	30	33	34	34	35	37	38	39	39
40	40	41	42	42	43	45	48	49	50

เนื่องจาก $N = 30$ จากตาราง 1.4 ฉะนั้น $k = 6$ จะเห็นว่า $X_{\text{max}} = 50$ และ $X_{\text{min}} = 5$ ดังนั้น ความกว้างชั้นหาได้จาก

$$I = \frac{R}{k} = \frac{X_{\text{max}} - X_{\text{min}}}{6} = \frac{50-5}{6} = 7.5 \approx 8$$

จะได้ว่า

$$\text{ขีดจำกัดล่างของชั้นแรก } = 5 - \frac{8(6)-45}{2} = 3.5 \approx 4$$

$$\text{ขีดจำกัดบนของชั้นแรก } = 4 + 8 - 1 = 11$$

ดังนั้น

อุณหภูมิ ($\times 10^{-1}$ องศาเซลเซียส)	รอยขีด	จำนวนสสาร
$4-11$	$\vert\vert$	2
$12-19$	$\vert\vert\vert\vert$	4
$20-27$	$\cancel{\vert\vert\vert\vert}$	5
$28-35$	$\cancel{\vert\vert\vert\vert}$	5
$36-43$	$\cancel{\vert\vert\vert\vert}\ \cancel{\vert\vert\vert\vert}$	10
$44-51$	$\vert\vert\vert\vert$	4

หรืออาจสร้างตารางในหน่วยองศาเซลเซียสได้ดังนี้

อุณหภูมิ (องศาเซลเซียส)	รอยขีด	จำนวนสสาร
$0.4-1.1$	$\vert\vert$	2
$1.2-1.9$	$\vert\vert\vert\vert$	4
$2.0-2.7$	$\cancel{\vert\vert\vert\vert}$	5
$2.8-3.5$	$\cancel{\vert\vert\vert\vert}$	5
$3.6-4.3$	$\cancel{\vert\vert\vert\vert}\ \cancel{\vert\vert\vert\vert}$	10
$4.4-5.1$	$\vert\vert\vert\vert$	4

ข้อสังเกตในการกำหนดจำนวนและความกว้างของอันตรภาคชั้น

1. ความกว้างของแต่ละอันตรภาคชั้นไม่จำเป็นต้องเท่ากัน ขึ้นอยู่กับวัตถุประสงค์ของการใช้ข้อมูล

2. ค่าที่สังเกตได้บางค่าอาจต่างไปจากค่าอื่นมาก เช่น ในการสอบครั้งหนึ่งมีผู้สอบได้ 2 คะแนน ในขณะที่คนอื่น ได้คะแนน มากกว่า 50 คะแนน ควรกำหนดอันตรภาคชั้นแรกเป็นอันตรภาคชั้นเปิด (Open-Ended Class Interval) ดังตัวอย่าง ตารางแสดงจำนวนผู้ป่วยของโรงพยาบาลแห่งหนึ่ง จำแนกตามระยะเวลาป่วย เป็นดังนี้

ระยะเวลาป่วย (วัน)	จำนวนผู้ป่วย (คน)
น้อยกว่า 7 วัน	$444,250$
$7-14$	$51,210$
$15-30$	$24,105$
$31-60$	$17,220$
$61-90$	$34,550$
มากกว่า 90 วัน	$11,150$
รวม	$182,485$

จากตัวอย่างนี้จะเห็นได้ชัดว่าความกว้างของแต่ละอันตรภาคชั้นไม่เท่ากัน และมีอันตรภาคชั้นเปิด

3. การกำหนดจำนวนอันตรภาคชั้นในการสร้างตารางแจกแจงความถี่ไม่มีกฎเกณฑ์แน่นอนตายตัว ขึ้นอยู่กับลักษณะการกระจายของข้อมูลรายละเอียดของข้อมูลที่ต้องการทราบด้วย เช่น ถ้าค่าที่สังเกตได้มีความแตกต่างกันมากมักจะกำหนดให้มีอันตรภาคชั้นน้อย เพื่อไม่ให้มีอันตรภาคชั้นที่มีความถี่เป็นศูนย์หรือหากต้องการทราบรายละเอียดของข้อมูลอย่างละเอียดก็ควรกำหนดให้มีจำนวนอันตรภาคชั้นมาก

1.4 ฮิสโทแกรม

ฮิสโทแกรม (Histogram) คือกราฟที่แสดงถึงการกระจายของข้อมูล มีลักษณะเป็นรูปสี่เหลี่ยมมุมฉากวางเรียงติดต่อกันบนแกนนอนโดยมีแกนนอนแทนค่าของตัวแปร ความกว้างของรูปสี่เหลี่ยมมุมฉากแทนความกว้างของอันตรภาคชั้น ความสูงของรูปสี่เหลี่ยมมุมฉากจะแสดงความถี่

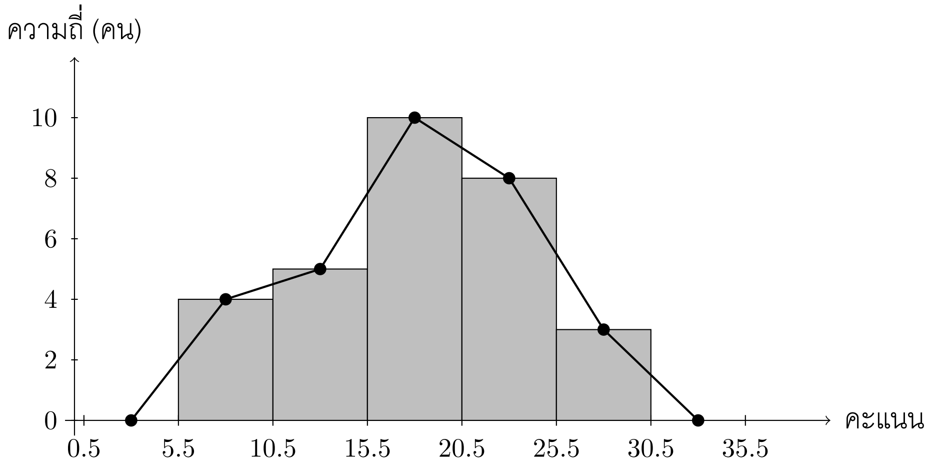
รูปหลายเหลี่ยมความถี่ (Frequency Polygon) คือรูปกราฟหลายเหลี่ยมที่เกิดจากการโยงจุดกึ่งกลางของยอดแท่งของสี่เหลี่ยมของฮิสโทแกรมด้วยเส้นตรง หรือแผนภูมิเส้นที่แสดงความถี่ของคะแนนแต่ละชั้น โดยเพิ่มฮิสโทแกรมอีก 2 ชั้นคือชั้นต่ำสุดและชั้นสูงสุดมีค่าความถี่เท่ากับ 0

ตัวอย่าง 1.4.1 ข้อมูลชุดหนึ่งแสดงตารางแจกแจงความถี่ดังนี้

คะแนน	ความถี่ (คน)	ขอบเขตจำกัดล่าง–ขอบเขตจำกัดบน	ค่ากึ่งกลาง
$6-10$	4	$5.5-10.5$	8
$11-15$	5	$10.5-15.5$	13
$16-20$	10	$15.5-20.5$	18
$21-25$	8	$20.5-25.5$	23
$26-30$	3	$25.5-30.5$	28

จงสร้างฮิสโทแกรมและรูปหลายเหลี่ยมความถี่ของข้อมูลชุดนี้

แนวคำตอบ สร้างฮิสโทแกรมและรูปหลายเหลี่ยมความถี่ ได้ดังนี้

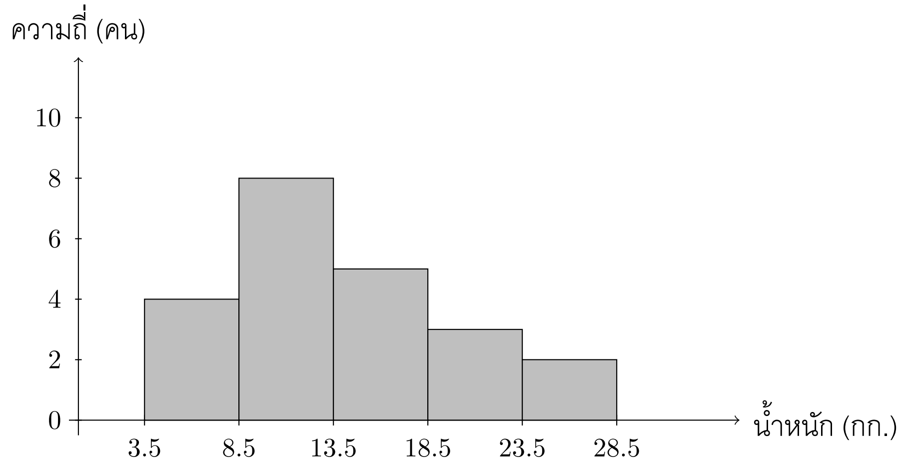


ตัวอย่าง 1.4.2 น้ำหนักของนักเรียนชั้นอนุบาลของโรงเรียนแห่งหนึ่ง เมื่อนำไปสร้างตารางแจงแจงความถี่ที่มีความกว้างเท่ากันทุกชั้นคือ 5 โดยมีขีดจำกัดล่างชั้นแรก 4 กิโลกรัม สร้างได้ทั้งหมด 5 ชั้น เรียงจากน้อยไปมาก โดยมีจำนวนนักเรียนแต่ละชั้น 4, 8, 5, 3, 2 ตามลำดับ จงสร้างฮิสโทแกรมของข้อมูลชุดนี้

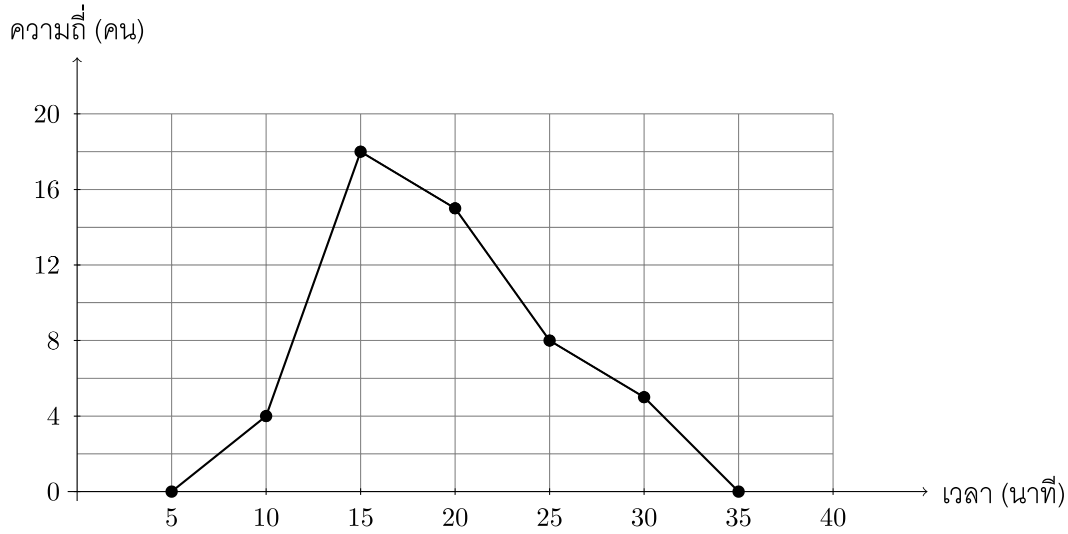
แนวคำตอบ จากข้อมูลเบื้องต้นนำไปสร้างตารางแจกแจงความถี่ได้ดังนี้

น้ำหนัก (กก.)	ความถี่ (คน)	ขอบเขตจำกัดล่าง−ขอบเขตจำกัดบน
$4-8$	4	$3.5-8.5$
$9-13$	8	$8.5-13.5$
$14-18$	5	$13.5-18.5$
$19-23$	3	$18.5-23.5$
$24-28$	2	$23.5-28.5$

สร้างฮิสโทแกรมและรูปหลายเหลี่ยมความถี่ ได้ดังนี้



ตัวอย่าง 1.4.3 ข้อมูลเวลา (นาที) ที่ใช้ในการเดินทางมาโรงเรียนกับจำนวนนักเรียนแสดงด้วยรูปหลายเหลี่ยมความถี่ดังนี้



1. จงหาจำนวนชั้นและความกว้างของอันตรภาคชั้น
2. จงหาจำนวนนักเรียนในโรงเรียนแห่งนี้
3. นักเรียนที่ใช้เวลาเดินทางมาโรงเรียนน้อยกว่า 18 นาทีมีกี่เปอร์เซ็นต์

แนวคำตอบ จากรูปหลายเหลี่ยมความถี่ คู่อันดับ (จุดกึ่งกลางชั้น, จำนวนนักเรียน) โดยไม่พิจารณาความถี่เป็น 0 ประกอบด้วย

$$(10, 4), (15, 18), (20, 15), (25, 8) \text{ และ } (30, 5)$$

ดังนั้นจำนวนชั้นเท่ากับ 5 ความกว้างของอันตรภาคชั้นเท่ากับ $15 - 10 = 5$
และจำนวนนักเรียนในโรงเรียนแห่งนี้ $4 + 18 + 15 + 8 + 5 = 50$ คน
พิจารณา

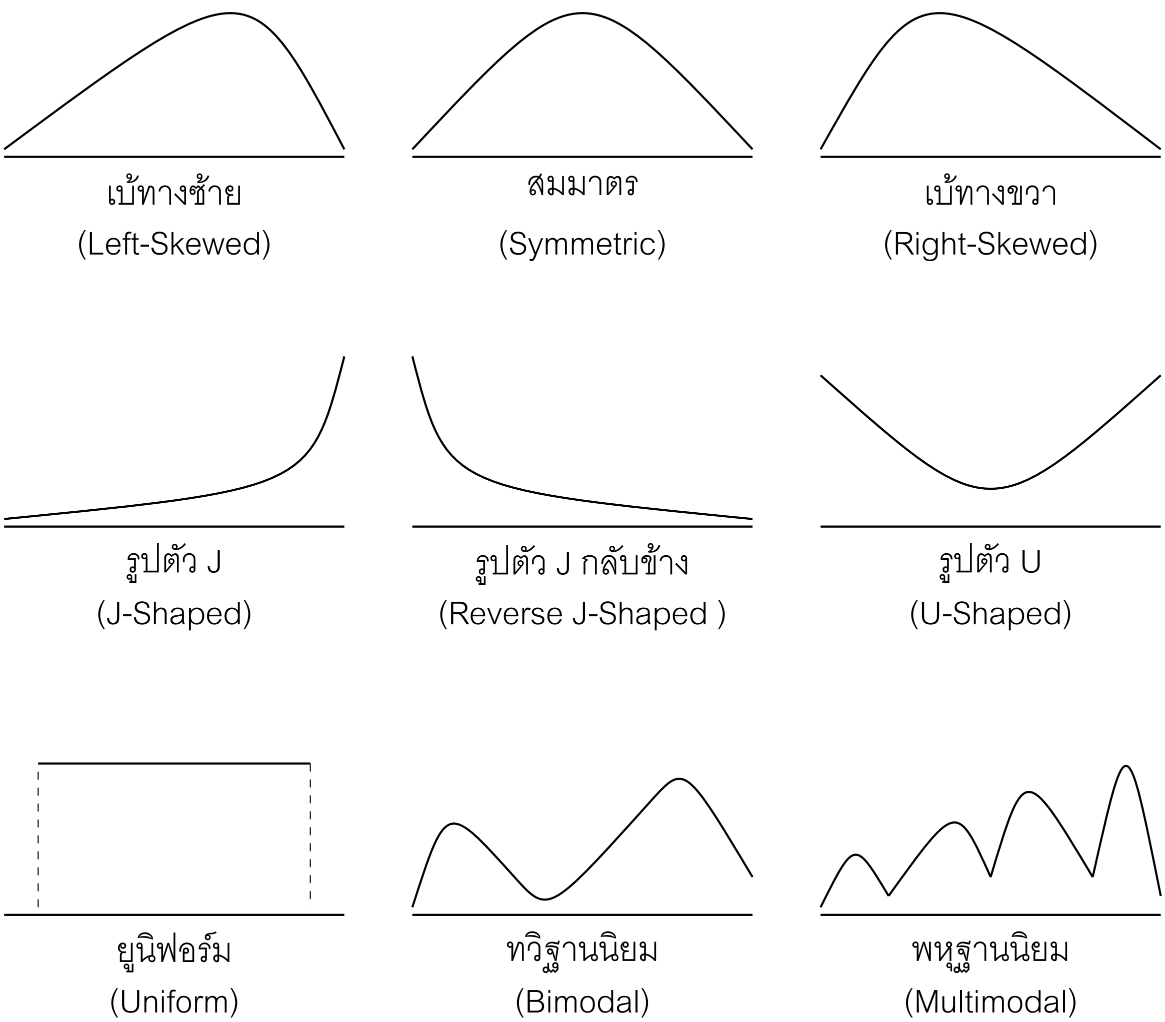
$$\text{ขอบเขตจำกัดบนชั้นที่สาม} = \frac{15 + 20}{2} = 17.5$$

ดังนั้นนักเรียนที่ใช้เวลาเดินทางมาโรงเรียนน้อยกว่า 18 นาทีอยู่ในชั้นที่ 1 และ 2 มีจำนวนเท่ากับ $4 + 18 = 22$ คิดเป็น

$$\frac{22}{50} \times 100 = 44\%$$

บางครั้งเมื่อสร้างรูปหลายเหลี่ยมความถี่แล้วสามารถปรับเส้นโค้งความถี่ให้เรียบได้เรียกว่า **เส้นโค้งความถี่** (Frequency curve) ซึ่งจะแสดงถึงลักษณะข้อมูล เส้นโค้งความถี่อาจอยู่ในรูปแบบต่าง ๆ ดังนี้

รูปที่ 1.1: รูปแบบต่าง ๆ ของเส้นโค้งความถี่



ความถี่สะสม (Cumulative frequency) คือผลรวมความถี่ของอันตรภาคชั้นนั้นกับความถี่ของทุก ๆ อันตรภาคชั้นที่อยู่ต่ำกว่าหรือสูงกว่าอย่างใดอย่างหนึ่ง มักใช้ F_i แทนความถี่สะสมของชั้นที่ i **ความถี่สัมพัทธ์** (Relative frequency) คือความถี่ของอันตรภาคชั้นนั้นหารด้วยผลรวมของความถี่ทั้งหมด **ความถี่สะสมสัมพัทธ์** (Cumulative relative frequency) คือผลรวมความถี่สะสมของอันตรภาคชั้นนั้นกับความถี่สะสมของทุก ๆ อันตรภาคชั้นที่อยู่ต่ำกว่าหรือสูงกว่าอย่างใดอย่างหนึ่ง หรืออาจบอกในรูปร้อยละหรือเปอร์เซ็นต์เมื่อคูณด้วย 100

ตัวอย่าง 1.4.4 ข้อมูลชุดหนึ่งแสดงตารางแจกแจงความถี่ดังนี้

คะแนน	ความถี่ (คน)
$6-10$	3
$11-15$	5
$16-20$	15
$21-25$	10
$26-30$	7

จากตารางแจกแจงความถี่จงหาความถี่สะสม ความถี่สัมพัทธ์ และร้อยละความถี่สะสมสัมพัทธ์

แนวคำตอบ

คะแนน	ความถี่ (คน)	ความถี่สะสม	ความถี่สัมพัทธ์	ร้อยละความถี่สะสมสัมพัทธ์
$6-10$	3	3	$\frac{3}{40}=0.075$	$\frac{3}{40}\times 100=7.5$
$11-15$	5	8	$\frac{5}{40}=0.125$	$\frac{8}{40}\times 100=20$
$16-20$	15	23	$\frac{15}{40}=0.375$	$\frac{23}{40}\times 100=57.5$
$21-25$	10	33	$\frac{10}{40}=0.250$	$\frac{33}{40}\times 100=82.5$
$26-30$	7	40	$\frac{7}{40}=0.175$	$\frac{40}{40}\times 100=100$

ตัวอย่าง 1.4.5 ข้อมูลชุดหนึ่งมีบางส่วนถูกนำเสนอในตารางต่อไปนี้

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์
$2-\ 6$			
$7-11$		11	0.2
$12-16$		14	
$17-21$	6		0.3

จงเติมตารางให้สมบูรณ์

แนวคำตอบ ให้ข้อมูลชุดนี้มี N จำนวน จากความถี่สัมพัทธ์ของชั้นสุดท้ายจะได้ว่า

$$\frac{6}{N}=0.3$$

ฉะนั้น $N=20$ ถ้าให้ x แทนความถี่ชั้นที่ 2 จะได้ว่า $\frac{x}{20}=0.2$ นั่นคือ $x=4$

พิจารณาค่าต่าง ๆ ได้ดังนี้

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์
$2-\ 6$	$11-4=7$	7	$\frac{7}{20}=0.35$
$7-11$	$x=4$	11	0.2
$12-16$	$14-11=3$	14	$\frac{3}{20}=0.15$
$17-21$	6	$N=20$	0.3

ดังนั้นเติมตารางให้สมบูรณ์ได้ดังนี้

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์
$2-\ 6$	7	7	0.35
$7-11$	4	11	0.2
$12-16$	3	14	0.15
$17-21$	6	20	0.3

เส้นโค้งความถี่สะสม (Cumulative frequency curve) **หรือเส้นโค้งโอจีฟ** (Ogive curve) เป็นโค้งที่แสดงความถี่สะสมของข้อมูลตั้งแต่ค่าต่ำสุดถึงค่าสูงสุด หลักการเขียนเส้นโค้งของความถี่สะสมทำได้ดังนี้

1. ให้แกนนอนเป็นคะแนน และแกนตั้งเป็นความถี่สะสม
2. หาตำแหน่งของจุด (ขอบเขตจำกัดบน, ความถี่สะสม) ของอันตรภาคชั้นแต่ละชั้นที่ไม่ใช่ชั้นแรก โดยชั้นแรกให้ใช้จุด (ขอบเขตจำกัดล่างชั้นแรก, 0)
3. ลากเส้นเชื่อมแต่ละจุดที่ติดกัน แล้วปรับโค้งให้เรียบ

ตัวอย่าง 1.4.6 ข้อมูลชุดหนึ่งแสดงตารางแจกแจงความถี่ดังนี้

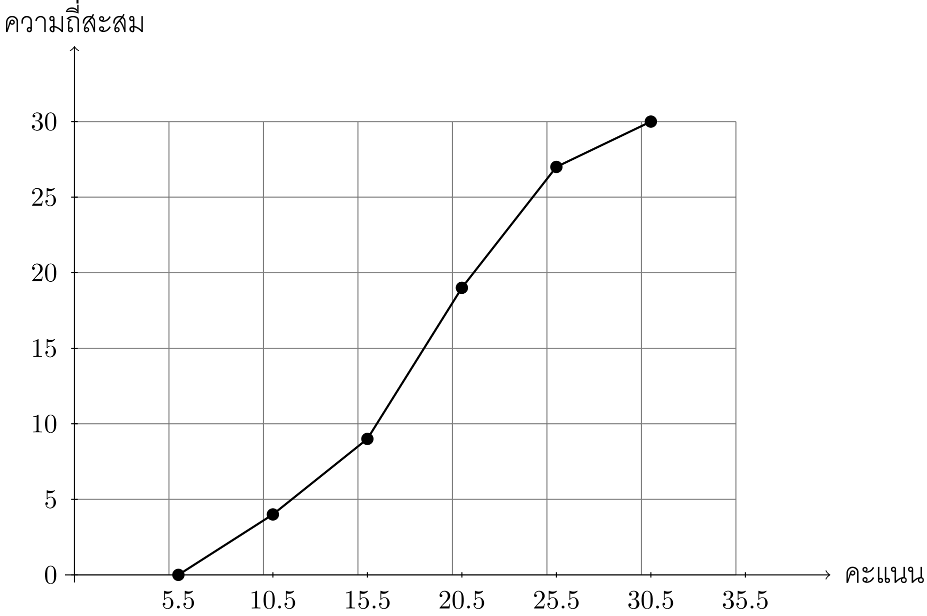
คะแนน	ความถี่ (คน)	ความถี่สะสม
$6-10$	4	4
$11-15$	5	9
$16-20$	10	19
$21-25$	8	27
$26-30$	3	30

จงสร้างเส้นโค้งความถี่สะสมของข้อมูลชุดนี้

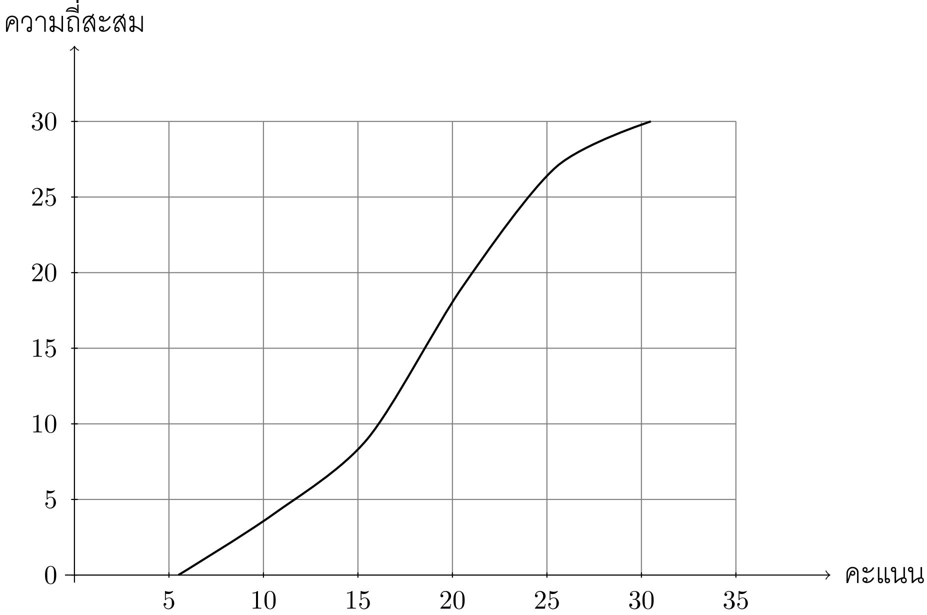
แนวคำตอบ พิจารณาคู่อันดับ (ขอบเขตจำกัดบน, ความถี่สะสม) ดังต่อไปนี้

คะแนน	จำนวนคน (ความถี่สะสม)	(ขอบเขตจำกัดบน, ความถี่สะสม)
น้อยกว่า 5.5	0	$(5.5,0)$
น้อยกว่า 10.5	4	$(10.5,4)$
น้อยกว่า 15.5	9	$(15.5,9)$
น้อยกว่า 20.5	19	$(20.5,19)$
น้อยกว่า 25.5	27	$(25.5,27)$
น้อยกว่า 30.5	30	$(30.5,30)$

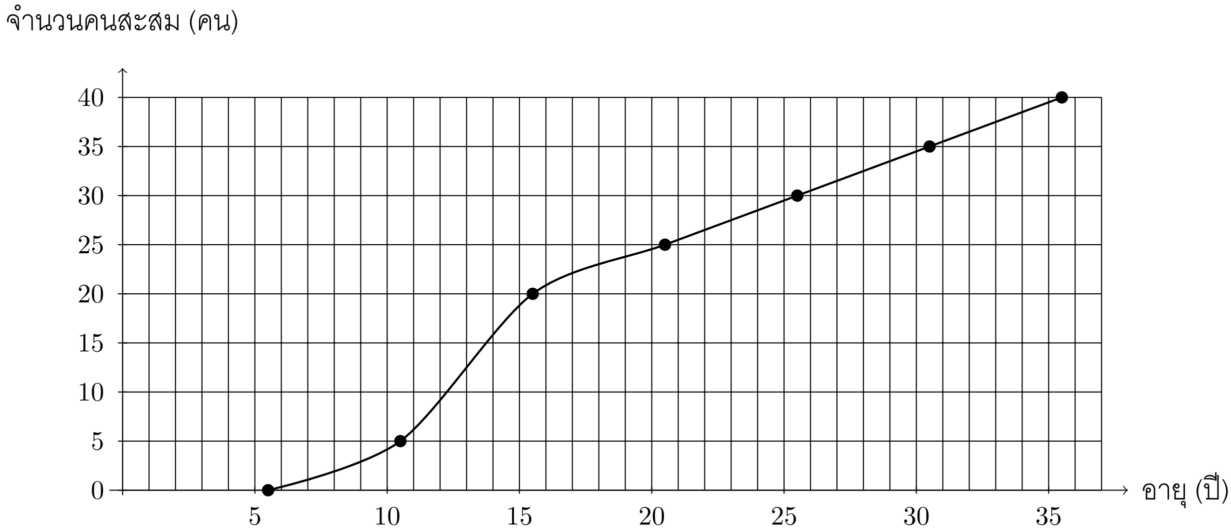
นำคู่อันดับดังกล่าวเขียนกราฟได้ดังนี้



จากนั้นปรับโค้งให้เรียบได้ดังนี้

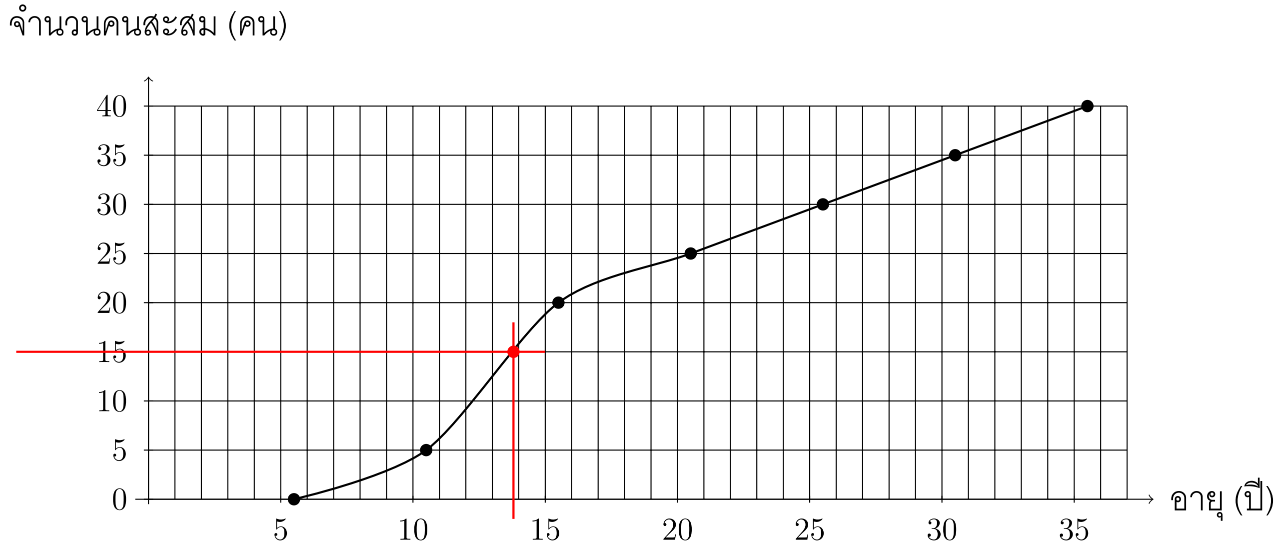


ตัวอย่าง 1.4.7 ข้อมูลอายุของชุมชนแห่งหนึ่ง แสดงด้วย**เส้นโค้งโอจีฟ**ดังนี้



จงหาช่วงอายุที่มีจำนวนคนสูงสุด (ความถี่สูงสุด) และจงประมาณอายุของคนลำดับที่ 15 เมื่อเรียงอายุจากน้อยไปมาก

แนวคำตอบ จากเส้นโค้งโอจีฟ จะเห็นว่า $(10.5, 5)$ และ $(15.5, 20)$ เป็นจุดบนกราฟที่ห่างกันมากที่สุด ดังนั้นช่วงอายุที่มีจำนวนคนสูงสุดคือ 11 - 15 ปี มีจำนวน $20 - 5 = 15$ คน และเมื่อลากเส้นแนวนอนที่ 15 คน ไปตัดเส้นโค้งโอจีฟ ดังกราฟ



จะได้ว่า ลำดับที่ 15 เมื่อเรียงอายุจากน้อยไปมากมีอายุประมาณ 13.8 ปี

1.5 แผนภาพต้น-ใบ

ในการจัดข้อมูลที่มีอยู่ให้เป็นกลุ่ม ๆ เพื่อความสะดวกในการนำไปวิเคราะห์ข้อมูลอาจทำได้โดยใช้ตารางแจกแจงความถี่และใช้กราฟ เช่น การสร้างฮิสโทแกรม จะเห็นว่าวิธีดังกล่าวอาจทำให้ไม่สามารถบอกได้ว่าข้อมูลที่มีอยู่มีค่าใดบ้าง เนื่องจากได้จัดแบ่งข้อมูลที่มีอยู่เป็นช่วง ๆ

การจัดข้อมูลเป็นกลุ่มมีอีกวิธีหนึ่งคือการสร้างแผนภาพเพื่อแจกแจงความถี่และวิเคราะห์ข้อมูลเบื้องต้นไปพร้อมกันคือ **แผนภาพต้น-ใบ** (stem and leaf diagram) ซึ่งทำได้ดังตัวอย่างต่อไปนี้

ตัวอย่าง 1.5.1 น้ำหนักของนักเรียนในโรงเรียนแห่งหนึ่งจำนวน 20 คน ดังนี้

34	35	36	38	39	39	40	40	41	42
42	43	44	45	46	50	51	51	52	54

แนวคำตอบ แบ่งข้อมูลออกเป็น 3 กลุ่มคือ

กลุ่มที่ 1 คือน้ำหนักตั้งแต่ $30 - 39$ ประกอบด้วย $34, 35, 36, 38, 39, 39$

กลุ่มที่ 2 คือน้ำหนักตั้งแต่ $40 - 49$ ประกอบด้วย $40, 40, 41, 42, 42, 43, 44, 45, 46$

กลุ่มที่ 3 คือน้ำหนักตั้งแต่ $50 - 59$ ประกอบด้วย $50, 51, 51, 52, 54$

ต้น	ใบ								
3	4	5	6	8	9	9			
4	0	0	1	2	2	3	4	5	6
5	0	1	1	2	4				

เขียนได้อีกแบบคือ

3	4	5	6	8	9	9			
4	0	0	1	2	2	3	4	5	6
5	0	1	1	2	4				

ตัวอย่าง 1.5.2 คะแนนของวิชาคณิตศาสตร์และฟิสิกส์ของนักเรียนในโรงเรียนแห่งหนึ่งจำนวน 20 คน ดังนี้

วิชาคณิตศาสตร์

12	15	16	22	23	24	26	30	31	33
33	35	35	39	40	41	42	44	44	45

วิชาฟิสิกส์

10	11	15	19	20	21	21	23	25	27
29	32	35	35	36	37	38	40	45	49

แนวคำตอบ แบ่งข้อมูลวิชาคณิตศาสตร์ออกเป็น 4 กลุ่มคือ
กลุ่มที่ 1 คือน้ำหนักตั้งแต่ 10 – 19 ประกอบด้วย 12, 15, 16
กลุ่มที่ 2 คือน้ำหนักตั้งแต่ 20 – 29 ประกอบด้วย 22, 23, 24, 26
กลุ่มที่ 3 คือน้ำหนักตั้งแต่ 30 – 39 ประกอบด้วย 30, 31, 33, 33, 35, 35, 39
กลุ่มที่ 4 คือน้ำหนักตั้งแต่ 40 – 49 ประกอบด้วย 40, 41, 42, 44, 44, 45

แบ่งข้อมูลวิชาฟิสิกส์ออกเป็น 4 กลุ่มคือ
กลุ่มที่ 1 คือน้ำหนักตั้งแต่ 10 – 19 ประกอบด้วย 10, 11, 15, 19
กลุ่มที่ 2 คือน้ำหนักตั้งแต่ 20 – 29 ประกอบด้วย 21, 20, 23, 25, 21, 29, 27
กลุ่มที่ 3 คือน้ำหนักตั้งแต่ 30 – 39 ประกอบด้วย 32, 35, 35, 36, 37, 38
กลุ่มที่ 4 คือน้ำหนักตั้งแต่ 40 – 49 ประกอบด้วย 40, 45, 49

ใบ (วิชาคณิตศาสตร์)	ต้น	ใบ (วิชาฟิสิกส์)
6 5 2	1	0 1 5 9
6 4 3 2	2	0 1 1 3 5 7 9
9 5 5 3 3 1 0	3	2 5 5 6 7 8
5 4 4 2 1 0	4	0 5 9

เขียนได้อีกแบบคือ

วิชาคณิตศาสตร์		วิชาฟิสิกส์
6 5 2	1	0 1 5 9
6 4 3 2	2	0 1 1 3 5 7 9
9 5 5 3 3 1 0	3	2 5 5 6 7 8
5 4 4 2 1 0	4	0 5 9

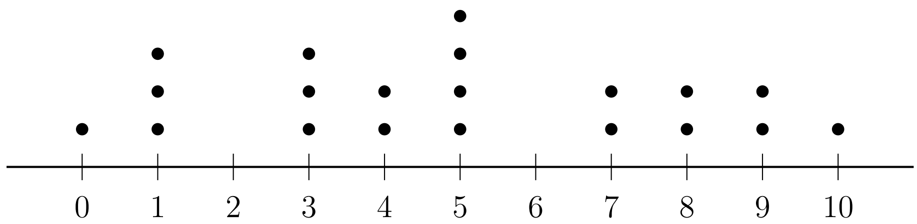
1.6 แผนภาพจุด

แผนภาพจุด (Dot Plot) เป็นรูปแบบหนึ่งของการนำเสนอข้อมูลเชิงปริมาณที่ทำได้ไม่ยาก โดยจะเขียนจุดแทนข้อมูลแต่ละตัวไว้เหนือเส้นในแนวนอนที่มีสเกลให้ตรงกับตำแหน่งที่แสดงค่าของข้อมูลนั้น แผนภาพจุดช่วยให้เห็นภาพรวมของข้อมูลได้รวดเร็วกว่าการพิจารณาจากข้อมูลโดยตรง โดยเฉพาะอย่างยิ่งเมื่อสนใจจะพิจารณาลักษณะของข้อมูลว่ามีการกระจายมากน้อยเพียงใด

ตัวอย่าง 1.6.1 จงสร้างภาพจุดของคะแนนสอบย่อยวิชาแคลคูลัส 1 ที่มีคะแนนดังต่อไปนี้

0	5	1	3	7	5	7	10	9	1
5	3	8	9	4	5	4	3	1	8

แนวคำตอบ นำคะแนนแทนในแผนภาพได้ดังนี้

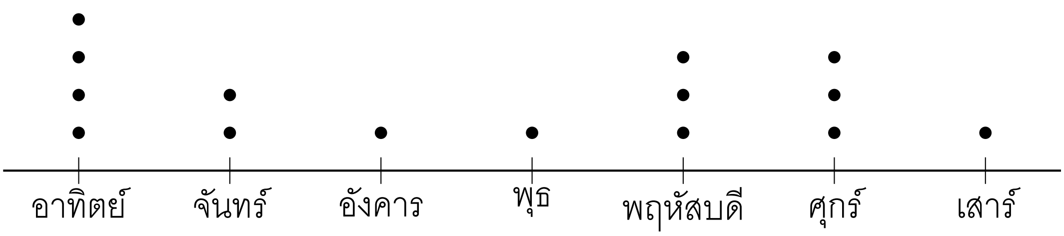


ตัวอย่าง 1.6.2 จากการสำรวจวันเกิดของนักเรียนในห้องเรียนหนึ่ง ปรากฏดังนี้

ลำดับที่	วันเกิด	ลำดับที่	วันเกิด	ลำดับที่	วันเกิด
1	อาทิตย์	6	เสาร์	11	ศุกร์
2	จันทร์	7	ศุกร์	12	จันทร์
3	อาทิตย์	8	อาทิตย์	13	อาทิตย์
4	อังคาร	9	พฤหัสบดี	14	ศุกร์
5	พุธ	10	พฤหัสบดี	15	พฤหัสบดี

จงสร้างภาพจุดแสดงวันเกิดของนักเรียนในห้องนี้

แนวคำตอบ นำวันเกิดแทนในแผนภาพได้ดังนี้



สรุป

ความรู้พื้นฐานเกี่ยวกับสถิติประกอบด้วยความหมายของสถิติซึ่งหมายถึงตัวเลขที่แสดงข้อเท็จจริงเกี่ยวกับเรื่อง ๆ หนึ่ง หรือหมายถึงศาสตร์ที่ว่าด้วยการเก็บรวบรวมข้อมูล เป็นต้น ข้อมูลทางสถิติคือสิ่งที่แสดงถึงลักษณะข้อเท็จจริงเกี่ยวกับบุคคล สิ่งของหรือเหตุการณ์ ซึ่งอาจแบ่งได้ตามลักษณะของข้อมูล ได้ 2 แบบ คือข้อมูลเชิงปริมาณ และข้อมูลเชิงคุณภาพ ถ้าแบ่งจากแหล่งที่มาคือ ข้อมูลปฐมภูมิ และข้อมูลทุติยภูมิ จากนั้นกล่าวถึง ประชากรซึ่งคือเซตของค่าสิ่งเกตุที่สนใจ และตัวอย่างคือเซตย่อยของประชากร ถัดจากนั้นกล่าวถึงการสำรวจซึ่งทำได้ 2 แบบ คือสำรวจทั้งประชากรเรียกว่า การทำสำมโน และการสำรวจบางส่วนคือ การสำรวจด้วยตัวอย่าง ถ้ามีข้อมูลเชิงปริมาณเราอาจนำไปแบ่งเป็นกลุ่ม ๆ เรียกว่าการแจกแจงความถี่ แล้วนำข้อมูลจากตารางแจกแจงความถี่ไปสร้างฮิสโทรแกรมและเส้นโค้งความถี่ เพื่อดูการกระจายของข้อมูลในเบื้องต้น อาจได้กราฟสมมาตร หรือไม่สมมาตร ถ้ากราฟที่สมมาตรเราจะเรียกว่าข้อมูลมีการแจกแจงปกติ สุดท้ายกล่าวถึงการสร้างแผนภาพต้น-ใบ จากข้อมูลเชิงปริมาณที่มีหลักสิบซ้ำกัน และการสร้างแผนภาพจุดเพื่อนำเสนอข้อมูลที่ซ้ำกันแต่มีจำนวนไม่มากนัก เหมาะทั้งข้อมูลเชิงปริมาณและเชิงคุณภาพ

แบบฝึกหัดบทที่ 1

1. ความหมายของคำว่าสถิติที่เหมาะสมที่สุดสำหรับคุณคือความหมายใด เพราะเหตุใด
2. จงยกตัวอย่างประโยชน์ของสถิติในงานด้านต่าง ๆ มาอย่างน้อย 2 ด้าน พร้อมเหตุผลประกอบ
3. จงยกตัวอย่างข้อมูลแบบต่อเนื่อง 2 ตัวอย่าง
4. จงบอกข้อดี และข้อเสีย ของข้อมูลปฐมภูมิและข้อมูลทุติยภูมิ
5. จงยกตัวอย่างการเก็บข้อมูลที่จำเป็นต้องใช้ การทำสำมะโน มาอย่างน้อย 2 ตัวอย่าง
6. จงบอกข้อดี และข้อเสีย ของการสำมะโน และการสำรวจด้วยตัวอย่าง
7. จงยกตัวอย่างการเก็บข้อมูลที่ยังไม่ได้กล่าวถึงในบทนี้
8. จงบอกข้อแตกต่างระหว่างสถิติเชิงพรรณา และสถิติเชิงอนุมาน
9. คะแนนจากการสอบวิชาคณิตศาสตร์ของนักเรียน 60 คน ซึ่งมีคะแนนเต็ม 30 คะแนน เป็นดังนี้

28	22	20	17	16	25	18	22	28	17	19	22	22	21	19
28	26	21	18	24	21	24	22	20	22	24	28	16	23	22
17	28	24	27	23	22	22	29	16	20	21	21	26	27	28
27	27	25	23	24	25	24	22	25	21	24	28	16	23	22

จากข้อมูลข้างต้นจงสร้างตารางแจกแจงความถี่จำนวน 5 ชั้น พร้อมทั้งแสดงข้อมูลด้วยฮิสโทแกรม รูปหลายเหลี่ยมความถี่ และเส้นโค้งโอจีฟ

10. จงเติมตารางให้สมบูรณ์

คะแนน	ความถี่ (คน)	ความถี่สะสม	ร้อยละความถี่สัมพัทธ์	จุดกึ่งกลางชั้น
$11-20$	20			
$21-30$	120			
$31-40$			25	
$41-50$	80			
	50			55.5
$61-70$		390		
$71-80$	10			

11. ข้อมูลคะแนนสอบวิชาสถิติ ของนักเรียน 36 คน ดังนี้

72	83	82	92	70	91	71	33	42	51	55	75
38	40	75	49	53	41	86	89	51	57	66	92
38	96	85	93	60	75	55	48	85	85	54	56

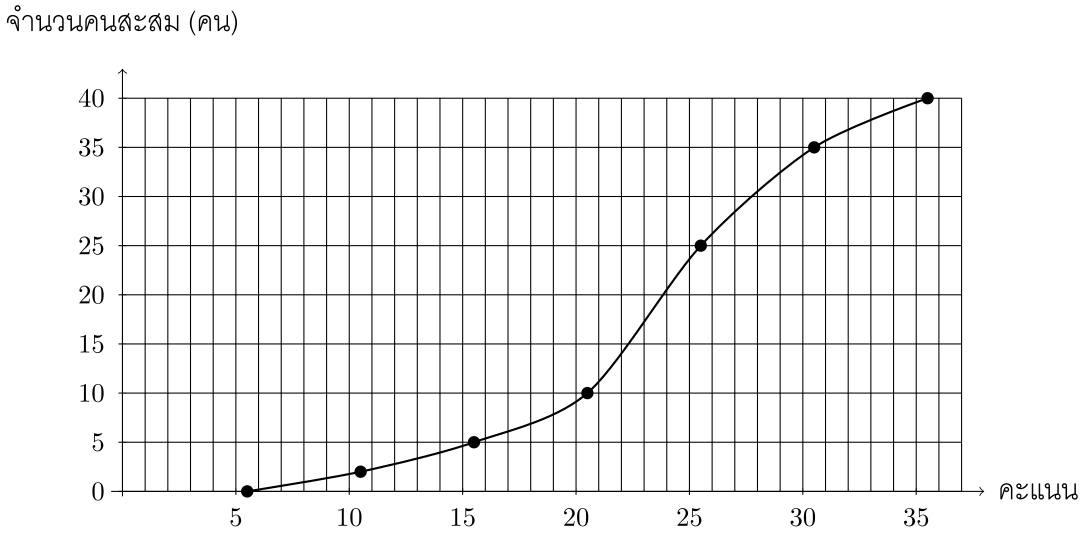
11.1 จงสร้างตารางแจกแจงความถี่ ความถี่สะสม ความถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์ ที่มีอันตรภาคชั้นเป็น $30-39$, $40-49$, $50-59$, $60-69$, $70-79$, $80-89$, $90-99$

11.2 จงหาร้อยละของจำนวนนักเรียนที่ได้คะแนนระหว่าง $70-79$ คะแนน

11.3 จงหาจำนวนนักเรียนที่ได้คะแนนมากกว่า 59 คะแนน

11.4 จงหาร้อยละของนักเรียนที่ได้คะแนนน้อยกว่าหรือเท่ากับ 59 คะแนน

12. ข้อมูลคะแนนเก็บวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง แสดงด้วยเส้นโค้งโอจีฟดังนี้



12.1 จงหาช่วงคะแนนที่มีจำนวนคนสูงสุด (ความถี่สูงสุด)

12.2 จงประมาณคะแนนของคนลำดับที่ 10 เมื่อเรียงคะแนนจากมากไปน้อย

12.3 นักเรียนที่ได้คะแนนมากกว่า 24 คะแนน ประมาณกี่เปอร์เซ็นต์

13. จงเขียนแผนภาพต้น-ใบ ของข้อมูลต่อไปนี้

44	52	46	59	104	101	46	55	43	60
66	48	54	56	100	109	74	84	49	70
44	59	88	84	40	79	71	104	49	101

14. จงสร้างแผนภาพต้น-ใบ ของข้อมูลการสำรวจนักเรียนในการนำเงินติดตัวมาโรงเรียนกลุ่มหนึ่งจำนวน 40 คน

100	150	200	160	150	140	160	180	150	170
200	180	125	140	170	190	150	120	150	140
180	190	150	130	180	140	140	160	190	160
180	190	140	120	130	110	190	160	170	130

15. ราคาปิด (บาท) ของหุ้น BTS ของ บริษัท บีทีเอส กรุ๊ป โฮลดิ้งส์ จำกัด (มหาชน) ย้อนหลัง 30 วัน นับจากวันที่ 9 เมษายน 2564 (ข้อมูลจากตลาดหลักทรัพย์แห่งประเทศไทย) ดังนี้

9.20	9.20	9.25	9.40	9.60	9.70	9.65	9.70	9.75	9.45
9.45	9.45	9.40	9.40	9.40	9.40	9.40	9.40	9.45	9.40
9.45	9.55	9.35	9.35	9.45	9.40	9.50	9.30	9.30	9.40

จงสร้างแผนภาพต้น-ใบ โดยใช้ทศนิยม 1 ตำแหน่งเป็นต้น และทศนิยมหลักที่ 2 เป็นใบ เช่น 9.2 | 0 แทนราคาปิด 9.20 บาท

16. แผนภาพต้น-ใบที่กำหนดให้แสดงค่าที่นักเรียน 30 คนประมาณความสูงของเสาธงของโรงเรียน โดยที่ 5 | 5 แทนความสูง 5.5 เมตร

5	2	7	5	5	5	8	9			
6	0	1	1	2	2	2	3	5	6	8
7	1	2	3	9						

16.1 มีนักเรียนทั้งหมดกี่คนที่ประมาณความสูงเสาธงมากกว่า 61 เมตร

16.2 ถ้าเสาธงสูง 6 เมตร มีนักเรียนร้อยละเท่าใดที่ประมาณผิดพลาดเกิน ± 20 เซนติเมตร

17. จงเขียนแผนภาพต้น-ใบ ของข้อมูลต่อไปนี้

ข้อมูลชุด A

61	63	74	84	95	67	71	77	88	92
62	98	68	65	81	74	77	69	96	60

ข้อมูลชุด B

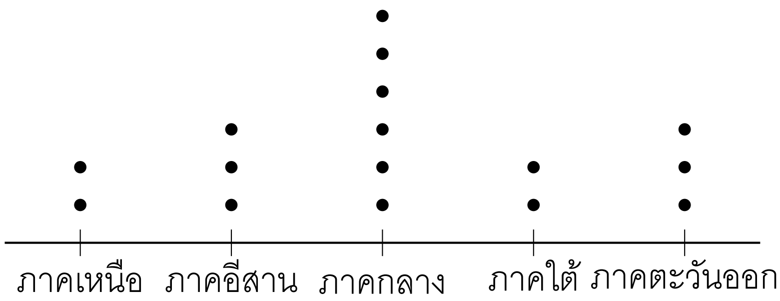
62	63	70	98	74	64	62	65	82	84
90	60	63	72	73	75	69	67	85	97

18. จากการสำรวจเดือนเกิดของนักเรียนในห้องเรียนหนึ่ง ปรากฏดังนี้

เลขที่	เดือนเกิด	เลขที่	เดือนเกิด	เลขที่	เดือนเกิด	เลขที่	เดือนเกิด
1	สิงหาคม	6	พฤษภาคม	11	ธันวาคม	16	พฤศจิกายน
2	มกราคม	7	สิงหาคม	12	ธันวาคม	17	เมษายน
3	พฤษภาคม	8	กันยายน	13	มีนาคม	18	มีนาคม
4	ธันวาคม	9	ตุลาคม	14	สิงหาคม	19	มิถุนายน
5	กรกฎาคม	10	มิถุนายน	15	สิงหาคม	20	เมษายน

จงสร้างภาพจุดแสดงเดือนเกิดของนักเรียนในห้องนี้

19. จากการสำรวจภูมิลำเนาของผู้ป่วยโควิด-19 แสดงเป็นแผนภาพจุดดังนี้
โดย ● แทนจำนวน 5 คน



19.1 จำนวนผู้ป่วยที่สำรวจในครั้งนี้มีจำนวนกี่คน

19.2 จำนวนผู้ป่วยที่มาจากภาคกลางสูงกว่าภาคเหนือคิดเป็นกี่เปอร์เซ็นต์

19.3 ถ้าผู้ป่วยในการสำรวจครั้งนี้มีภูมิลำเนาในกรุงเทพฯ 80% ของผู้ป่วยมีภูมิลำเนาในภาคกลาง แล้วผู้ป่วยมีภูมิลำเนาในกรุงเทพฯ คิดเป็นกี่เปอร์เซ็นต์ของการสำรวจครั้งนี้

บทที่ 2

การวิเคราะห์ข้อมูลเบื้องต้น

การใช้ประโยชน์จากสถิตินั้นในเบื้องต้นคือการนำข้อมูลที่มีมาวิเคราะห์เพื่ออธิบายลักษณะของข้อมูลชุดนั้น ๆ เช่นการนำคะแนนสอบในห้องมาหาค่าเฉลี่ยได้เป็น 8 จากคะแนนเต็ม 10 นั่นอาจแปลความได้ว่านักเรียนส่วนใหญ่ในห้องนี้มีคะแนนสูง ดังนั้นการวิเคราะห์ข้อมูลเบื้องต้นจึงมีความสำคัญในการแปลความหมายของข้อมูล เพื่อนำไปใช้ประโยชน์ต่อไป

2.1 การวัดแนวโน้มสู่ศูนย์กลาง

การวัดแนวโน้มสู่ศูนย์กลาง เป็นการคำนวณค่ากลางของข้อมูลซึ่ง คือค่าที่ใช้เป็นตัวแทนของข้อมูล บอกถึงลักษณะของข้อมูล ทำให้ผู้ใช้สามารถทราบถึงการแจกแจงข้อมูลว่าเป็นอย่างไร เพื่อความสะดวกในการสรุปเกี่ยวกับข้อมูลนั้น ๆ ซึ่งจะช่วยทำให้เกิดการวิเคราะห์ข้อมูลถูกต้องขึ้น

การหาค่ากลางของข้อมูลมีหลายวิธี แต่ละวิธีมีข้อดีและข้อเสีย และมีความเหมาะสมในการนำไปใช้ไม่เหมือนกัน ขึ้นอยู่กับลักษณะข้อมูลและวัตถุประสงค์ของผู้ใช้ข้อมูลนั้น ค่ากลางที่สำคัญมี 3 ชนิดดังนี้ ค่าเฉลี่ย (Mean) มัธยฐาน (Median) และ ฐานนิยม (Mode)

ในหัวข้อนี้จะกล่าวถึงการบวกของข้อมูล ดังนั้นเริ่มต้นด้วยสัญลักษณ์แทนการบวกที่เรียกว่า**ซิกมา** $\sum$ (sigma) นิยามโดย

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

มีสมบัติเบื้องต้นดังนี้ เมื่อ c เป็นค่าคงตัว

(ก) $\displaystyle\sum_{i=1}^{n} c = cn$

(ข) $\displaystyle\sum_{i=1}^{n} x_i = \sum_{i=1}^{m} x_i + \sum_{i=m+1}^{n} x_i$ เมื่อ $m < n$

(ค) $\displaystyle\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$

(ง) $\displaystyle\sum_{i=1}^{n}(x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$

และมีผลบวกที่สำคัญดังนี้

1. $\sum_{i=1}^{n} i = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$ สูตรของเกาส์ (Gauss' formula)

2. $\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$

3. $\sum_{i=1}^{n} i^3 = 1^3 + 2^3 + 3^3 + \cdots + n^3 = \left[\frac{n(n+1)}{2}\right]^2$

1. ค่าเฉลี่ย (Mean)

ค่าเฉลี่ย คือค่าที่ได้จากการเฉลี่ยข้อมูลทั้งหมดซึ่งคำนวณได้หลายแบบ แบ่งออกเป็น 3 ชนิดหลัก ๆ ซึ่งแต่ละค่าเหมาะกับข้อมูลแต่ชนิดแตกต่างกัน

1. ค่าเฉลี่ยเลขคณิต (Arithematic mean)
2. ค่าเฉลี่ยเรขาคณิต (Geometric mean)
3. ค่าเฉลี่ยฮาร์โมนิก (Hamonic mean)

บทนิยาม 2.1.1 ข้อมูลเชิงปริมาณชุดหนึ่งประกอบด้วยข้อมูลทั้งหมด

$$X_1,\ X_2,\ X_3,\ ...,\ X_N$$

ค่าเฉลี่ยเลขคณิตของประชากร เขียนแทนด้วย μ หาได้จาก

$$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N} \quad \text{เขียนได้ย่อ ๆ คือ} \quad \mu = \frac{\sum X_i}{N}$$

เมื่อ N แทนขนาดของประชากร และเมื่อเลือกตัวอย่างขนาด n จากประชากร ได้ข้อมูลดังนี้

$$x_1,\ x_2,\ x_3,\ ...,\ x_n$$

ค่าเฉลี่ยเลขคณิตของตัวอย่าง เขียนแทนด้วย $\overline{X}$ หาได้จาก

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{เขียนได้ย่อ ๆ คือ} \quad \overline{X} = \frac{\sum x_i}{n}$$

ค่าเฉลี่ยเลขคณิตเหมาะที่จะนำมาใช้เป็นค่ากลางของข้อมูลเมื่อข้อมูลนั้นๆ ไม่มีค่าใดค่าหนึ่งซึ่งสูงหรือต่ำกว่าค่าอื่น ๆ ที่เหลืออย่างผิดปกติ ค่าเฉลี่ยเลขคณิตเป็นค่าเฉลี่ยที่นิยมใช้มากที่สุดในกลุ่มค่าเฉลี่ยทั้งหมด บางครั้งอาจเรียกสั้น ๆ ว่า ค่าเฉลี่ย

การหาค่าเฉลี่ยของอายุคนไทย ถ้าขณะนี้มีคนไทย 67 ล้านคน นั่นคือ $N = 67$ ล้าน และ X_i อายุของคนไทยคนที่ i เมื่อ $i = 1, 2, ..., 67$ ล้าน ในทางปฏิบัติทำได้ยาก เราอาจทำได้โดยการเลือกตัวอย่างมา 100 คน นั่นคือ $n = 100$ คนแทนขนาดของตัวอย่าง เพื่อคำนวณหาอายุเฉลี่ย และในบทที่ 8 จะกล่าวถึงการยอมรับว่าค่าเฉลี่ยจากตัวอย่างที่ได้เหมาะสมที่จะเป็นค่าเฉลี่ยของประชากรหรือไม่

ตัวอย่าง 2.1.2 ราคาปิด (บาท) ของหุ้น AOT ของ บริษัท ท่าอากาศยานไทย จำกัด (มหาชน) ย้อนหลัง 10 วัน นับจากวันที่ 31 มีนาคม 2564 (ข้อมูลจากตลาดหลักทรัพย์แห่งประเทศไทย) ดังนี้

69.0 69.00 68.25 67.75 67.00 67.00 67.25 67.25 67.25 68.25

1. จงหาค่าเฉลี่ยเลขคณิตของประชากร

   แนวคำตอบ จะได้ว่า
   $$\begin{aligned}\mu &= \frac{69.0+69.00+68.25+67.75+67.00+67.00+67.25+67.25+67.25+68.25}{10} \\ &= 67.80\end{aligned}$$

   ดังนั้นค่าเฉลี่ยของราคาปิดของ AOT ย้อนหลัง 10 วัน นับจากวันที่ 31 มีนาคม 2564 เท่ากับ 67.80 บาท

2. จงหาค่าเฉลี่ยเลขคณิตของตัวอย่าง เมื่อสุ่มมา 4 วัน ประกอบด้วย 69.00, 68.25, 67.25, 67.00

   แนวคำตอบ จะได้ว่า
   $$\overline{X} = \frac{69.00+68.25+67.25+67.00}{4} = 67.875$$

   ดังนั้นค่าเฉลี่ยของราคาปิดของ AOT จากการสุ่มมา 4 วัน เท่ากับ 67.875 บาท

ตัวอย่าง 2.1.3 โรงงานแห่งหนึ่งผลิตหน้ากากอนามัยที่ใช้ป้องกันโรคโควิด-19 ผลิตสินค้าใส่กล่อง ๆ ละ 100 ชิ้น ถ้าวันนี้ผลิตได้ 1,000 กล่อง และมีการสุ่มตัวอย่างสินค้ามาตรวจคุณภาพ 10 กล่อง พบว่าสินค้าบกพร่องต่อกล่อง 1, 2, 5, 3, 0, 1, 0, 3, 2 และ 6 ชิ้น จงหา

1. ร้อยละของสินค้าที่บกพร่อง

   แนวคำตอบ จำนวนสินค้าบกพร่องโดยเฉลี่ยต่อกล่องเท่ากับ
   $$\overline{X} = \frac{1+2+5+3+0+1+0+3+2+6}{10} = 2.3$$

   นั่นคือโดยเฉลี่ยจะมีสินค้าบกพร่อง 2.3 ชิ้นต่อกล่อง เนื่องจาก 1 กล่องมี 100 ชิ้น ดังนั้นร้อยละของสินค้าที่บกพร่องเท่ากับ 2.3 หรือ 2.3%

2. จำนวนสินค้าบกพร่องที่เกิดจากผลิตในวันนี้

   แนวคำตอบ เนื่องจากวันนี้โรงงานผลิตได้ 100 กล่อง โดยมีสินค้าบกพร่องเฉลี่ย 2.3 ชิ้นต่อกล่อง ดังนั้นจำนวนสินค้าบกพร่องที่คาดว่าจะเกิดขึ้นในวันนี้เท่ากับ
   $$2.3 \times 1000 = 230 \text{ ชิ้น}$$

   จากการผลิตทั้งหมด $100 \times 1000 = 100,000$ ชิ้น

ตัวอย่าง 2.1.4 ข้อมูลของประชากรชุดหนึ่ง ประกอบด้วย $x_1, x_2, x_3, ..., x_{10}$ ซึ่ง $\sum_{i=1}^{10} x_i^2 = 100$ และ $\sum_{i=1}^{10}(x_i-4)^2 = 20$ จงหาค่าเฉลี่ยของข้อมูลชุดนี้

แนวคำตอบ พิจารณา

$$\begin{aligned}
\sum_{i=1}^{10}(x_i^2-8x_i+16) = \sum_{i=1}^{10}(x_i-4)^2 &= 20 \\
\sum_{i=1}^{10}x_i^2 - 8\sum_{i=1}^{10}x_i + \sum_{i=1}^{10}16 &= 20 \\
100 - 8\sum_{i=1}^{10}x_i + 16(10) &= 20 \\
\sum_{i=1}^{10}x_i &= 30
\end{aligned}$$

ดังนั้นค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ $\mu = \dfrac{30}{10} = 3$

ตัวอย่าง 2.1.5 ในการสอบวิชาสถิติของเด็กห้องหนึ่ง ค่าเฉลี่ยเลขคณิตของเด็กห้องนี้เท่ากับ 53 คะแนน แต่จากการตรวจสอบพบว่าข้อมูลนักเรียนสองคนยังไม่ได้ตรวจ เมื่อตรวจเสร็จปรากฏว่าค่าเฉลี่ยเลขคณิตของเด็กห้องนี้เป็น 55 คะแนน และผลรวมของคะแนนสอบเพิ่มขึ้นอีก 180 คะแนน จำนวนนักเรียนห้องนี้เท่ากับเท่าใด

แนวคำตอบ ให้จำนวนนักเรียนห้องนี้เท่ากับ n มีคะแนนสอบวิชาสถิติคือ $x_1, x_2, ..., x_n$ ซึ่งมีค่าเฉลี่ยเลขคณิตเท่ากับ 53 จะได้ว่า

$$\sum_{i=1}^{n}x_i = 53n$$

พบว่าข้อมูลนักเรียนสองคนยังไม่ได้ตรวจให้มีคะแนนสอบคือ x_{n+1}, x_{n+2} โดยที่ $x_{n+1}+x_{n+2} = 180$ และ

$$\begin{aligned}
\sum_{i=1}^{n+2}x_i &= 55(n+2) \\
\sum_{i=1}^{n}x_i + x_{n+1} + x_{n+2} &= 55n+110 \\
53n+180 &= 55n+110 \\
n &= 35
\end{aligned}$$

ดังนั้นจำนวนนักเรียนห้องนี้เท่ากับ $35+2=37$ คน

ตัวอย่าง 2.1.6 จงหาค่าเฉลี่ยเลขคณิตของข้อมูล $x_1, x_2, x_3, ..., x_{100}$ เมื่อ

$$x_n = \begin{cases} n^2 & \text{เมื่อ } n \text{ เป็นจำนวนคี่} \\ 2n & \text{เมื่อ } n \text{ เป็นจำนวนคู่} \end{cases}$$

แนวคำตอบ ค่าเฉลี่ยเลขคณิตคือ

$$\begin{aligned}
\mu = \frac{\sum_{i=1}^{n}}{100} &= \frac{1^2+2(2)+3^2+2(4)+5^2+2(6)+\cdots+99^2+2(100)}{100} \\
&= \frac{(1^2+3^2+5^2+\cdots+99^2)+4(1+2+3+\cdots+50)}{100} \\
&= \frac{\sum_{i=1}^{50}(2i-1)^2+4\sum_{i=1}^{50} i}{100} \\
&= \frac{\sum_{i=1}^{50}(4i^2-4i+1)+\sum_{i=1}^{50} 4i}{100} \\
&= \frac{\sum_{i=1}^{50}(4i^2-4i+1+4i)}{100} \\
&= \frac{\sum_{i=1}^{50}(4i^2+1)}{100} \\
&= \frac{4\sum_{i=1}^{50} i^2+\sum_{i=1}^{50} 1}{100} \\
&= \frac{4\cdot\frac{50(51)(101)}{6}+1(50)}{100} \\
&= \frac{171750}{100} = 1717.5
\end{aligned}$$

ดังนั้นเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ 1,717.5

ทฤษฎีบท 2.1.7 ข้อมูล $x_1, x_2, x_3, ..., x_n$ ให้ A เป็นค่าคงที่ และ

$$d_i = x_i - A \qquad \text{เมื่อ} \quad i = 1, 2, 3, ..., n$$

ถ้า $\overline{D}$ เป็นค่าเฉลี่ยเลขคณิตของ $d_1, d_2, d_3, ..., d_n$
และ $\overline{X}$ เป็นค่าเฉลี่ยเลขคณิตของ $x_1, x_2, x_3, ..., x_n$ แล้ว

$$\overline{X} = \overline{D} + A$$

ข้อสังเกต เห็นได้ชัดว่าถ้า $A = \overline{X}$ แล้ว $\overline{D} = 0$

บทพิสูจน์. จากความสัมพันธ์ $d_i = x_i - A$ จะได้ว่า $x_i = d_i + A$ ดังนั้น

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n}(d_i+A)}{n} = \frac{\sum_{i=1}^{n} d_i + \sum_{i=1}^{n} A}{n} = \frac{\sum_{i=1}^{n} d_i + nA}{n} = \overline{D} + A$$

ดังนั้น $\overline{X} = \overline{D} + A$ □

ตัวอย่าง 2.1.8 จงหาค่าเฉลี่ยเลขคณิตของข้อมูล 10, 12, 11, 12, 13, 14 โดยใช้ $A = 10$

แนวคำตอบ

x_i	10	12	11	12	13	14	รวม
$d_i = x_i - 10$	0	2	1	2	3	4	12

ดังนั้น $\overline{X} = \overline{D} + A = \frac{12}{6} + 10 = 12$

ตัวอย่าง 2.1.9 จงหาค่าเฉลี่ยเลขคณิตของข้อมูล โดยใช้ $A = 74$

73 73 72 77 70 71 80 78 83 74 69 68

แนวคำตอบ

x_i	73	73	72	77	70	71	80	78	83	74	69	68	รวม
$d_i = x_i - 74$	-1	-1	-2	3	-4	-3	6	4	9	0	-5	-6	0

ดังนั้น $\overline{X} = \overline{D} + A = \frac{0}{12} + 74 = 74$

บทนิยาม 2.1.10 ค่าเฉลี่ยเลขคณิตสำหรับข้อมูลที่จัดกลุ่ม
ข้อมูลเชิงปริมาณชุดหนึ่งมีขนาดประชากร N มีการแจงแจงความถี่ k ชั้น แล้วค่าเฉลี่ยเลขคณิตของประชากรเมื่อทราบค่าข้อมูลทุกค่าของประชากร

$$\mu = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{\sum f_i X_i}{\sum f_i}$$

X_i คือค่ากึ่งกลางของชั้นที่ i

f_i คือความถี่ของชั้นที่ i

N คือขนาดของประชากร หรือ $N = \sum f_i$

กรณีที่ทราบข้อมูลบางหน่วย n คือขนาดตัวอย่างซึ่ง $n = \sum f_i$ จะได้ว่า

$$\overline{X} = \frac{\sum f_i x_i}{\sum f_i}$$

ตัวอย่าง 2.1.11 นักเรียนกลุ่มหนึ่งจำนวน 10 คนมีอายุดังนี้ 7, 10, 12, 10, 15, 8, 9, 10, 17, 8 ปี

1. จงหาอายุเฉลี่ยของนักเรียนกลุ่มนี้โดยใช้ข้อมูลดิบ

   แนวคำตอบ อายุเฉลี่ยของนักเรียนกลุ่มนี้เท่ากับ

$$\mu = \frac{7 + 10 + 12 + 10 + 15 + 8 + 9 + 10 + 17 + 8}{10} = \frac{106}{10} = 10.6 \qquad \text{ปี}$$

2. จงหาค่าเฉลี่ยเลขคณิตของอายุของนักเรียนกลุ่มนี้โดยใช้ตารางแจกแจงความถี่ที่กำหนดให้

แนวคำตอบ

อายุ (ปี)	ความถี่ (f_i)	ค่ากึ่งกลางชั้น (x_i)	f_ix_i
$7-9$	4	8	32
$10-12$	4	11	44
$13-15$	1	14	14
$16-18$	1	17	17
รวม	10		107

ดังนั้นค่าเฉลี่ยเลขคณิตของอายุของนักเรียนกลุ่มนี้โดยใช้ตารางแจกแจงความถี่เท่ากับ

$$\mu = \frac{107}{10} = 10.7 \quad \text{ปี}$$

ตัวอย่าง 2.1.12 ในการทดสอบความถนัดของนักเรียนกลุ่มหนึ่ง มีตารางแจกแจงความถี่ของผลการสอบดังนี้

ช่วงคะแนน	ความถี่ (คน)
$0-\ 4$	4
$5-\ 9$	5
$10-14$	x
$15-19$	7

ถ้าค่าเฉลี่ยเลขคณิตของคะแนนสอบเท่ากับ 11 แล้วนักเรียนที่ได้คะแนนในช่วง $5-14$ คะแนน มีจำนวนคิดเป็นร้อยละของนักเรียนกลุ่มนี้เท่ากับเท่าใด

แนวคำตอบ พิจารณา

ช่วงคะแนน	ความถี่ (f_i)	ค่ากึ่งกลางชั้น (x_i)	f_ix_i
$0-\ 4$	4	2	8
$5-\ 9$	5	7	35
$10-14$	x	12	$12x$
$15-19$	7	17	119
รวม	$16+x$		$12x+162$

จะได้ว่า

$$\begin{aligned}
\frac{12x+162}{16+x} &= 11 \\
12x+162 &= 11(16+x) \\
x &= 14
\end{aligned}$$

จำนวนนักเรียนที่ได้คะแนนในช่วง $5-14$ มีจำนวน $5+x=19$ คน จากทั้งหมด $16+x=30$ คน คิดเป็น

$$\frac{19}{30} \times 100 = 63.33\ \%$$

ทฤษฎีบท 2.1.13 ข้อมูลเชิงปริมาณชุดหนึ่งมีขนาดประชากร N มีการแจงแจงความถี่ k ชั้น แล้วค่าเฉลี่ยเลขคณิตของประชากรเมื่อทราบค่าข้อมูลทุกค่าของประชากร และ A เป็นค่าคงตัว ซึ่ง $d_i = X_i - A$ เมื่อ $i = 1, 2, ..., k$ และ

$$\overline{D} = \frac{\sum_{i=1}^{k} f_i d_i}{N}$$

แล้วค่าเฉลี่ยเลขคณิตคือ $\overline{X} = \overline{D} + A$

X_i คือค่ากึ่งกลางของชั้นที่ i

f_i คือความถี่ของชั้นที่ i

กรณีที่ทราบข้อมูลบางหน่วย n คือขนาดตัวอย่าง เมื่อ $d_i = x_i - A$ จะได้ว่า

$$\overline{X} = A + \frac{\sum f_i d_i}{n}$$

บทพิสูจน์. เป็นแบบฝึกหัด □

ตัวอย่าง 2.1.14 คะแนนวิชาคณิตศาสตร์ของนักเรียนในชั้นเรียนที่มีนักเรียน 30 เป็นดังนี้

คะแนน	ความถี่ (f_i)
$1-5$	5
$6-10$	8
$11-15$	15
$16-20$	2

จงหาค่าเฉลี่ยเลขคณิตของคะแนนของนักเรียนในชั้นนี้โดยใช้ $A = 12$

แนวคำตอบ พิจารณา

คะแนน	ความถี่ (f_i)	ค่ากึ่งกลางชั้น (x_i)	$d_i = x_i - 12$	$f_i d_i$
$1-5$	5	3	-9	-45
$6-10$	8	8	-4	-32
$11-15$	15	13	1	15
$16-20$	2	18	6	12
รวม	30			-50

ดังนั้นคะแนนเฉลี่ยของวิชาคณิตศาสตร์ของนักเรียนห้องนี้เท่ากับ

$$\frac{-50}{30} + 12 = 10.33$$

ทฤษฎีบท 2.1.15 สำหรับข้อมูลเชิงปริมาณชุดหนึ่ง เมื่อเพิ่มอีกหนึ่งค่าซึ่งเท่ากับค่าเฉลี่ยเลขคณิตของกลุ่มนี้ จงแสดงว่าค่าเฉลี่ยเลขคณิตใหม่จะไม่เปลี่ยนแปลง

บทพิสูจน์. ให้ประชากรขนาด n ประกอบด้วย $x_1, x_2, ..., x_n$ ซึ่งมีค่าเฉลี่ยเลขคณิตเท่ากับ μ นั่นคือ

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

สมมติค่าที่เพิ่มคือ $y = \mu$ ค่าเฉลี่ยเลขคณิตของ $x_1, x_2, ..., x_n, y$ เท่ากับ

$$\frac{\sum_{i=1}^{n} x_i + y}{n+1} = \frac{n\mu + \mu}{n+1} = \frac{\mu(n+1)}{n+1} = \mu$$

ดังนั้นค่าเฉลี่ยเลขคณิตใหม่จะไม่เปลี่ยนแปลง □

ทฤษฎีบท 2.1.16 ข้อมูล $x_1, x_2, x_3, ..., x_n$ มีค่าเฉลี่ยเลขคณิตคือ μ แล้ว $\sum_{i=1}^{n}(x_i - \mu) = 0$

บทพิสูจน์. $\sum_{i=1}^{n}(x_i - \mu) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \mu = n\mu - n\mu = 0$ □

ตัวอย่าง 2.1.17 ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, x_3, ..., x_{10}$ มีค่าเฉลี่ยเลขคณิตเท่ากับ μ ซึ่ง

$$\sum_{i=1}^{10} x_i = 70 \qquad \text{และ} \qquad \sum_{i=1}^{10}(x_i - 4)^2 = 150$$

จงหา $\sum_{i=1}^{10}(x_i - \mu)^2$

แนวคำตอบ จะเห็นว่า $\mu = \frac{\sum x_i}{10} = \frac{70}{10} = 7$ พิจารณา

$$\sum_{i=1}^{10}[(x_i - 7) + 3]^2 = \sum_{i=1}^{10}(x_i - 4)^2 = 150$$

$$\sum_{i=1}^{10}[(x_i - 7)^2 + 6(x_i - 7) + 9] = 150$$

$$\sum_{i=1}^{10}(x_i - 7)^2 + 6\sum_{i=1}^{10}(x_i - 7) + \sum_{i=1}^{10} 9 = 150$$

โดยทฤษฎีบท 2.1.16 จะได้ว่า $\sum_{i=1}^{10}(x_i - 7) = 0$ นั่นคือ

$$\sum_{i=1}^{10}(x_i - 7)^2 + 6(0) + 9(10) = 150$$

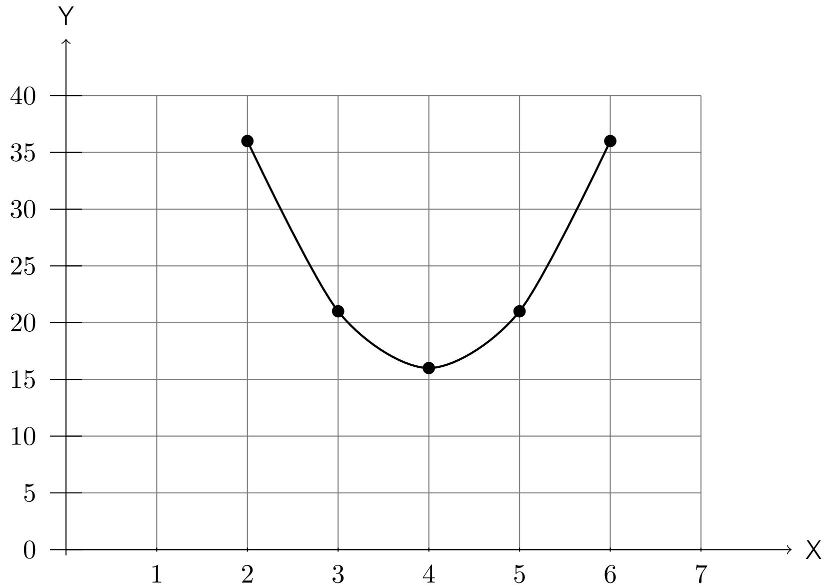
$$\sum_{i=1}^{10}(x_i - 7)^2 = 60$$

ดังนั้น $\sum_{i=1}^{10}(x_i - \mu)^2 = 60$

จากข้อมูล 2, 3, 3, 5, 7 ค่าเฉลี่ยเลขคณิตเท่ากับ $\mu = \frac{2+3+3+5+7}{5} = 4$ จะได้ว่า

คะแนน (x_i)	$(x_i-2)^2$	$(x_i-3)^2$	$(x_i-\mu)^2$	$(x_i-5)^2$	$(x_i-6)^2$
2	0	1	4	9	16
3	1	0	1	4	9
3	1	0	1	4	9
5	9	4	1	0	1
7	25	16	9	4	1
รวม	36	21	16	21	36

กำหนดให้ $y = \sum_{i=1}^{5}(x_i - x)^2$ เมื่อ $x = 2, 3, 4, 5, 6$ นำไปเขียนกราฟแล้วลากเส้นไประหว่างจุดปรับโค้งให้เรียบ ได้ดังนี้



จะเห็นได้ว่ากราฟที่ได้เป็นพาราโบลาหงายที่มีจุดยอดอยู่ที่ $x = \mu$ ซึ่งจะแสดงการพิสูจน์ได้จากทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2.1.18 ข้อมูล $x_1, x_2, x_3, ..., x_n$ มีค่าเฉลี่ยเลขคณิตคือ μ แล้ว

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

บทพิสูจน์. สำหรับจำนวนจริง x ใด ๆ กำหนดให้

$$f(x) = \sum_{i=1}^{n}(x_i - x)^2$$

เราจะพิสูจน์โดยอาศัยความรู้แคลคูลัสในเรื่องค่าสูงสุดต่ำสุด โดยการทดสอบอนุพันธ์ อันดับสอง พิจารณา

$$0 = f'(x) = \sum_{i=1}^{n} 2(x_i - x)(-1) = -2\sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n} x = -2n\mu + 2nx$$

ฉะนั้น $x = \mu$ เป็นจุดวิกฤต เนื่องจาก

$$f''(x) = 2n > 0 \qquad \because n > 0$$

นั่นคือ $f''(\mu) > 0$ ดังนั้น $f(\mu) = \sum_{i=1}^{n}(x_i - \mu)^2$ เป็นค่าต่ำสุด □

ทฤษฎีบท 2.1.19 สมบัติเชิงเส้น (Linear property)
ให้ข้อมูลชุด Y ประกอบด้วย $y_1, y_2, ..., y_n$ และข้อมูลชุด X ประกอบด้วย $x_1, x_2, ..., x_n$ โดยที่

$$y_i = a + bx_i \qquad \text{เมื่อ } i = 1, 2, ..., n \quad \text{และ } a, b \text{ เป็นค่าคงตัว แล้ว}$$

$$\mu_Y = a + b\,\mu_X$$

เมื่อ μ_X และ μ_Y เป็นค่าเฉลี่ยเลขคณิตของข้อมูลชุด X และ Y ตามลำดับ

บทพิสูจน์.

$$n\mu_Y = \sum_{i=1}^{n} y_i = \sum_{i=1}^{n}(a + bx_i) = \sum_{i=1}^{n} a + b\sum_{i=1}^{n} x_i = na + n\mu_X$$

ดังนั้น $\mu_Y = a + b\,\mu_X$ □

หมายเหตุ สมบัติเชิงเส้นเป็นจริงทั้งข้อมูลแบบตัวอย่างและประชากร

ตัวอย่าง 2.1.20 ข้อมูลชุด X คือ 9, 15, 10, 9, 12 สัมพันธ์กับชุด Y คือ $y_i = 1 + 2x_i$ จงหา

1. ค่าเฉลี่ยเลขคณิตของข้อมูลชุด Y โดยเปลี่ยนทุกค่า

   แนวคำตอบ พิจารณา

x_i	9	9	10	12	15	$\sum x_i = 55$
$y_i = 1 + 2x_i$	19	19	21	25	31	$\sum y_i = 115$

   ดังนั้น $\mu_Y = \dfrac{115}{5} = 23$

2. ค่าเฉลี่ยเลขคณิตของข้อมูลชุด Y โดยใช้สมบัติเชิงเส้น

   แนวคำตอบ เนื่องจาก $\mu_X = \dfrac{55}{5} = 11$ โดยสมบัติเชิงเส้นจะได้ว่า $\mu_Y = 1 + 2(11) = 23$

ตัวอย่าง 2.1.21 ในการสำรวจน้ำหนักของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งซึ่งมี 3 ห้อง มีจำนวนนักเรียน 44, 46 และ 42 ตามลำดับ ปรากฏว่ามีค่าเฉลี่ยเลขคณิตเท่ากับ 50 กิโลกรัม แต่พบว่าเครื่องชั่งน้ำหนักได้ตัวเลขสูงเกินจริงคนละ 1 กิโลกรัม ดังนั้นค่าเฉลี่ยที่ถูกต้องของน้ำหนักของนักเรียนชั้นมัธยมศึกษาปีที่ 6 นี้เท่ากับกี่กิโลกรัม

แนวคำตอบ น้ำหนักของนักเรียนทั้ง 3 ห้องประกอบด้วย $x_1, x_2, x_3, ..., x_{132}$ ซึ่งมี $\mu_X = 50$ แต่พบว่าเครื่องชั่งน้ำหนักได้ตัวเลขสูงเกินจริงคนละ 1 กิโลกรัม ดังนั้นน้ำหนักที่แท้จริงคือ $y_i = x_i + 1$ ดังนั้น

$$\mu_Y = \mu_X + 1 = 50 + 1 = 51$$

ดังนั้นค่าเฉลี่ยที่ถูกต้องของน้ำหนักของนักเรียนชั้นมัธยมศึกษาปีที่ 6 นี้เท่ากับก 51 กิโลกรัม

ทฤษฎีบท 2.1.22 ให้ข้อมูลทั้งหมด m ชุดแต่ละชุดมีขนาดเท่ากับ $N_1, N_2, ..., N_m$ และค่าเฉลี่ยเลขคณิต $\overline{X}_1, \overline{X}_2, ..., \overline{X}_m$ ตามลำดับแล้ว **ค่าเฉลี่ยเลขคณิตรวม** (combined arithematic mean) ของข้อมมูลทั้งหมดเขียนแทนด้วย $\overline{X}_{\text{com}}$ คือ

$$\overline{X}_{\text{com}} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + \cdots + N_m\overline{X}_m}{N_1 + N_2 + \cdots + N_m}$$

บทพิสูจน์. เป็นแบบฝึกหัด □

ตัวอย่าง 2.1.23 จากการสำรวจรายได้ของอาชีพต่าง ๆ หารายได้เฉลี่ยต่อเดือนของบุคคลากรในอาชีพต่าง ๆ ดังนี้

อาชีพ	จำนวนตัวอย่าง (คน)	รายได้ต่อเดือน (บาท)
ค้าขายออนไลน์	60	25,000
ครูอัตราจ้าง	30	9,000
พนักงานธุรกิจเอกชน	40	27,000
ผู้ปฏิบัติงานในโรงงาน	100	15,000
นักคณิตศาสตร์ประกันภัย	50	45,000

จงหารายได้เฉลี่ยต่อคนต่อเดือน

แนวคำตอบ พิจารณา

$$\begin{aligned}\overline{X}_{\text{com}} &= \frac{60(25000) + 30(9000) + 40(27000) + 100(15000) + 50(45000)}{60 + 30 + 40 + 100 + 50} \\ &= 23571.43\end{aligned}$$

ดังนั้นรายได้เฉลี่ยต่อคนต่อเดือนเท่ากับ 23,571.43 บาท

ค่าเฉลี่ยเลขคณิตที่ผ่านมาคำนวณจากค่าสังเกตที่มีน้ำหนักเท่ากันทุกค่า แต่ในหัวข้อนี้เราจะกำหนดค่าเฉลี่ยเลขคณิตโดยที่แต่ละค่าสังเกตไม่จำเป็นต้องมีน้ำหนักเท่ากัน เราเรียกค่าเฉลี่ยประเภทนี้ว่า **ค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนัก** (weighted arithematic mean)

บทนิยาม 2.1.24 ให้ $w_1, w_2, ..., w_n$ เป็นน้ำหนักของ $x_1, x_2, ..., x_n$ ตามลำดับ แล้วค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนักคำนวณได้จาก

$$\overline{X}_w = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n}$$

ตัวอย่าง 2.1.25 ถ้าอาจารย์ผู้สอนวิชาสถิติให้น้ำหนักคะแนนสอบกลางภาครวมกับปลายภาคเป็น 3 เท่าของคะแนนสอบย่อยแต่ละครั้ง ถ้านักศึกษาคนหนึ่งมีคะแนนสอบกลางภาค 60 คะแนน คะแนนสอบปลายภาค 80 คะแนน และสอบย่อย 4 ครั้ง คะแนน 8, 9, 4 และ 7 ตามลำดับ จงหาคะแนนเฉลี่ยของนักศึกษาคนนี้

แนวคำตอบ ให้ x_i แทนคะแนนสอบย่อยครั้งที่ 1, 2, 3, 4 และ x_5 แทนผลรวมของคะแนนสอบกลางภาครวมและปลายภาค นั่นคือ $x_1 = 8, x_2 = 9, x_3 = 4, x_4 = 7$ และ $x_5 = 60 + 80 = 140$ ถ้าให้ w_i แทนน้ำหนักของคะแนนสอบ จะได้ว่า $w_1 = w_2 = w_3 = w_4 = 1$ และ $w_5 = 3$ ฉะนั้น

$$\begin{aligned}\overline{X}_w &= \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5}{w_1 + w_2 + w_3 + w_4 + w_5} \\ &= \frac{1(8) + 1(9) + 1(4) + 1(7) + 3(140)}{1 + 1 + 1 + 1 + 3} \\ &= 64\end{aligned}$$

ดังนั้นคะแนนเฉลี่ยของนักศึกษาคนนี้เท่ากับ 64 คะแนน

ตัวอย่าง 2.1.26 ผลการสอบวิชาคณิตศาสตร์ของนายการันย์ในชั้นมัธยมศึกษาปีที่ 4 เป็นดังนี้

รหัสวิชา	ค4101	ค42101	ค41102	ค41202
จำนวนหน่วยกิต	1	1.5	1	1.5
เกรด	2.5	3	3.5	2

จงหาเกรดเฉลี่ยวิชาคณิตศาสตร์ของนายการันย์

แนวคำตอบ พิจารณา

$$\overline{X}_w = \frac{1(2.5) + 1.5(3) + 1(3.5) + 1.5(2)}{1 + 1.5 + 1 + 1.5} = 2.70$$

ดังนั้นเกรดเฉลี่ยวิชาคณิตศาสตร์ของนายการันย์ในชั้นมัธยมศึกษาปีที่ 4 เท่ากับ 2.70

ทฤษฎีบท 2.1.27 ข้อมูลที่ประกอบด้วย $x_1, x_2, ..., x_n$ ซึ่งมีน้ำหนักเท่ากัน แล้วค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนักของข้อมูลชุดนี้เท่ากับ

$$\overline{X}_w = \overline{X}$$

บทพิสูจน์. ให้น้ำหนักของ $x_1, x_2, ..., x_n$ เท่ากันเท่ากับ w จะได้ว่า

$$\overline{X}_w = \frac{wx_1 + wx_2 + \cdots + wx_n}{w + w + \cdots + w} = \frac{w\sum_{i=1}^{n} x_i}{wn} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$$

ดังนั้น $\overline{X}_w = \overline{X}$ □

ตัวอย่าง 2.1.28 ผลคะแนนภาคการศึกษาหนึ่งในระดับมหาวิทยาลัยของนายปลื้ม เป็นดังนี้

ชื่อวิชา	**หน่วยกิต**	**ระดับคะแนน**
แคลคูลัส ๑	3	C
หลักการคณิตศาสตร์	3	B^+
ประวัติและพัฒนาการของคณิตศาสตร์	3	B
ภาษาเพื่อการสื่อสาร	3	A
คุณธรรม จริยธรรม สำหรับครู	3	A

กำหนดให้ $C = 2 \quad B = 3 \quad B^+ = 3.5 \quad A = 4$

จงหาคะแนนเฉลี่ยของนายปลื้มในภาคการศึกษานี้

แนวคำตอบ เนื่องจากน้ำหนักของระดับคะแนนเท่ากัน โดยทฤษฎีบท 2.1.27 จะได้ว่า

$$\overline{X}_w = \frac{2 + 3.5 + 3 + 4 + 4}{5} = 3.30$$

ดังนั้นคะแนนเฉลี่ยของนายปลื้มในภาคการศึกษานี้เท่ากับ 3.30

สำหรับข้อมูลที่มีค่าสังเกตเป็นค่าบวกทุกค่าและมีค่าแตกต่างกันมาก ๆ การใช้ค่ากลางของข้อมูลเป็นค่าเฉลี่ยเลขคณิตจึงไม่เหมาะสม ข้อมูลลักษณะแบบนี้เราจะใช้ค่ากลางของข้อมูลที่เรียกว่า **ค่าเฉลี่ยเรขาคณิต** (Geometric mean)

บทนิยาม 2.1.29 ให้ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, ..., x_n$ เมื่อ $x_i > 0$ ทุก ๆ $i = 1, 2, ..., n$ **ค่าเฉลี่ยเรขาคณิต** เขียนแทนด้วย GM นิยามโดย

$$\text{GM} = \sqrt[n]{x_1 \cdot x_2 \cdot ... \cdot x_n} \qquad \text{หรือ} \qquad \ell\text{og GM} = \frac{1}{n}\sum_{i=1}^{n} \ell\text{og} x_i$$

ตัวอย่าง 2.1.30 จากข้อมูลแต่ละข้อต่อไปนี้ จงหาค่าเฉลี่ยเลขคณิตและค่าเฉลี่ยเรขาคณิต

1. $1, 2, 2, 4$

   แนวคำตอบ จะได้ว่า

$$\overline{X} = \frac{1 + 2 + 2 + 4}{4} = 2.25$$
$$\text{GM} = \sqrt[4]{1 \cdot 2 \cdot 2 \cdot 4} = 1.68$$

2. $1, 3, 243$

   แนวคำตอบ จะได้ว่า

$$\overline{X} = \frac{1 + 3 + 243}{3} = 82.33$$
$$\text{GM} = \sqrt[3]{1 \cdot 3 \cdot 243} = 9$$

ตัวอย่าง 2.1.31 นาย ก ได้รับเงินเดือนเพิ่มขึ้นปีที่แล้ว 5% และปีนี้ เพิ่มขึ้น 15% จงหาเปอร์เซนต์เฉลี่ยของเงินเดือนที่เพิ่มขึ้น

1. จงหาเปอร์เซนต์เฉลี่ยของเงินเดือนที่เพิ่มขึ้น ทั้งค่าเฉลี่ยเลขคณิตและค่าเฉลี่ยเรขาคณิต

   แนวคำตอบ ปีที่แล้วเงินเดือนนาย ก เพิ่มขึ้น 5% ดังนั้นเงินเดือนปีที่แล้วเท่ากับ 1.05 ของเงินเดือนก่อนปีที่แล้ว ปีนี้เงินเดือนเพิ่มขึ้นอีก 15% เงินเดือนของนาย ก ปีนี้เท่ากับ 1.15 ของปีที่แล้ว ดังนั้น

   $$\text{ค่าเฉลี่ยเลขคณิต} = \frac{1.05 + 1.15}{2} = 1.10$$

   นั่นคือเงินของนาย ก เพิ่มขึ้นโดยเฉลี่ย 10%

   $$\text{ค่าเฉลี่ยเรขาคณิต} = \sqrt{(1.05)(1.15)} = 1.09886$$

   นั่นคือเงินของนาย ก เพิ่มขึ้นโดยเฉลี่ย 9.886%

2. ควรใช้ค่าเฉลี่ยเลขคณิต หรือค่าเฉลี่ยเรขาคณิต เป็นตัวแทนของข้อมูลดังกล่าว

   แนวคำตอบ สมมติว่านาย ก ก่อนมีเงินเดือนก่อนขึ้นเงินเดือนเป็น 10,000 บาท
   การขึ้นเงินเดือนครั้งแรก 5% ทำให้นาย ก มีเงินเดือนเพิ่มขึ้นเท่ากับ

   $$10000(0.05) = 500 \quad \text{บาท}$$

   การขึ้นเงินเดือนครั้งที่สอง 15% ทำให้นาย ก มีเงินเดือนเพิ่มขึ้นเท่ากับ

   $$10500(0.15) = 1575 \quad \text{บาท}$$

   รวมเงินที่เพิ่มขึ้น 2 ครั้งเท่ากับ $500 + 1575 = 2075$ บาท
   กรณีใช้ค่าเฉลี่ยเลขคณิต

   $$\begin{aligned}
   \text{ครั้งที่ 1 เงินเดือนเพิ่ม} &= 10000(0.1) = 1000 \\
   \text{ครั้งที่ 2 เงินเดือนเพิ่ม} &= 11000(0.1) = 1100 \\
   \text{เงินรวมที่เพิ่มขึ้นทั้งหมด} &= 1000 + 1100 = 2100
   \end{aligned}$$

   กรณีใช้ค่าเฉลี่ยเรขาคณิต

   $$\begin{aligned}
   \text{ครั้งที่ 1 เงินเดือนเพิ่ม} &= 10000(0.09886) = 988.60 \\
   \text{ครั้งที่ 2 เงินเดือนเพิ่ม} &= 10988.6(0.09886) = 1086.33 \\
   \text{เงินรวมที่เพิ่มขึ้นทั้งหมด} &= 988.60 + 1086.33 = 2074.93
   \end{aligned}$$

   จะเห็นว่าเงินรวมกรณีใช้ค่าเฉลี่ยเรขาคณิตมีค่าใกล้เคียงค่าจริงคือ 2,075 บาท มากกว่ากรณีใช้ค่าเฉลี่ยเลขคณิต ดังนั้นควรใช้ค่าเฉลี่ยเรขาคณิต

ค่าเฉลี่ยฮาร์โมนิก (Hamonic mean) เป็นค่ากลางขอข้อมูลที่เหมาะสมสำหรับข้อมูลเฉพาะอย่าง เช่นระยะทางต่อเวลา จำนวนหน่วยต่อเวลา เป็นต้น กล่าวคือข้อมูลลักษณะเป็นอัตราส่วน

บทนิยาม 2.1.32 ให้ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, ..., x_n$ เมื่อ $x_i > 0$ ทุกๆ $i = 1, 2, ..., n$ **ค่าเฉลี่ยฮาร์โมนิก** เขียนแทนด้วย HM นิยามโดย

$$\text{HM} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \qquad \text{หรือ} \qquad \text{HM} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1}$$

ตัวอย่าง 2.1.33 อัตราเร็วของรถเด็กเล่นจำนวน 4 คัน คือ 2, 3, 4 และ 6 เมตรต่อวินาที จงหาค่าเฉลี่ยของอัตราเร็วของรถเด็กเล่น

แนวคำตอบ เนื่องจากอัตราเร็วของรถเด็กเล่นเป็นอัตราส่วนของระยะทางต่อเวลา ดังนั้นหาค่าเฉลี่ยโดยใช้ค่าเฉลี่ยฮาร์โมนิก จะได้ว่า

$$\text{HM} = \left[\frac{1}{4}\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{6}\right)\right]^{-1} = 3.2$$

ดังนั้นอัตราเร็วของรถเด็กเล่นโดยเฉลี่ยเท่ากับ 3.2 เมตรต่อวินาที

ตัวอย่าง 2.1.34 ในการทำงานอย่างเดียวกันใช้เวลาทำงานนี้เสร็จ 1 ชิ้น ของเด็ก 5 คน ใช้เวลาต่างกันดังนี้

เด็กคนที่	เวลาที่ใช้ทำงาน (วัน)
1	4
2	10
3	6
4	5
5	12

จงหาค่าเฉลี่ยของการทำงานของเด็กทั้ง 5 คน

แนวคำตอบ เนื่องการทำงานใช้เวลาต่อการทำงานนี้เสร็จ 1 ชิ้น ดังนั้นหาค่าเฉลี่ยโดยใช้ค่าเฉลี่ยฮาร์โมนิก จะได้ว่า

$$\text{HM} = \left[\frac{1}{5}\left(\frac{1}{4} + \frac{1}{10} + \frac{1}{6} + \frac{1}{5} + \frac{1}{12}\right)\right]^{-1} = 6.25$$

ดังนั้นการทำงานอย่างเดียวกันใช้เวลาทำงานนี้เสร็จ 1 ชิ้นของเด็ก 5 คน โดยเฉลี่ย 6.25 วัน

2. มัธยฐาน (Median)

บทนิยาม 2.1.35 มัธยฐาน (Median) ของข้อมูลชุดหนึ่งคือ ค่าของข้อมูลที่อยู่ตำแหน่งตรงกลาง เมื่อเรียงข้อมูลจากน้อยไปมากหรือมากไปน้อย เขียนแทนด้วย Med

ให้ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, ..., x_n$ ซึ่ง $x_1 \leq x_2 \leq ... \leq x_n$ แล้ว

$$\text{Med} = \begin{cases} x_{\frac{n+1}{2}} & \text{เมื่อ } n \text{ เป็นจำนวนคี่} \\ \frac{1}{2}\left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right] & \text{เมื่อ } n \text{ เป็นจำนวนคู่} \end{cases}$$

นั่นคือข้อมูลขนาด n เมื่อเรียงจากน้อยไปมากหรือมากไปน้อย Med จะมีค่าเท่ากับข้อมูลที่อยู่ตำแหน่งของ $\dfrac{n+1}{2}$ ซึ่งเรียกว่าตำแหน่งของมัธยฐาน

ตัวอย่าง 2.1.36 จงหามัธยฐานของข้อมูลต่อไปนี้

1. 3, 1, 5, 6, 6

แนวคำตอบ ข้อมูลมีขนาด 5 และเมื่อเรียงจากน้อยไปมากจะได้ว่า

1, 3, 5, 6, 6

ตำแหน่งของมัธยฐานคือ $\dfrac{5+1}{2} = 3$ ดังนั้น Med $= 5$

2. 2, 2, 4, 5, 7, 7, 8, 10

แนวคำตอบ เนื่องจากข้อมูลเรียงจากน้อยไปมาก

2, 2, 4, 5, 7, 7, 8, 10

ข้อมูลมีขนาด 8 โดยที่ตำแหน่งของมัธยฐานคือ $\dfrac{8+1}{2} = 4.5$ ดังนั้น Med $= \dfrac{5+7}{2} = 6$

ตัวอย่าง 2.1.37 จากแผนภาพต้น-ใบ ของข้อมูลชุดหนึ่งเป็นดังนี้

2	1	7	9	9		
3	0	1	2	2	5	7
4	2	3	4	5		
6	0	3				

จงหามัธยฐานของข้อมูลชุดนี้

แนวคำตอบ เนื่องจากข้อมูลเรียงจากน้อยไปมาก และข้อมูลมีขนาด 16 โดยที่ตำแหน่งของมัธยฐานคือ $\dfrac{16+1}{2} = 8.5$ ดังนั้น Med $= \dfrac{32+35}{2} = 33.5$

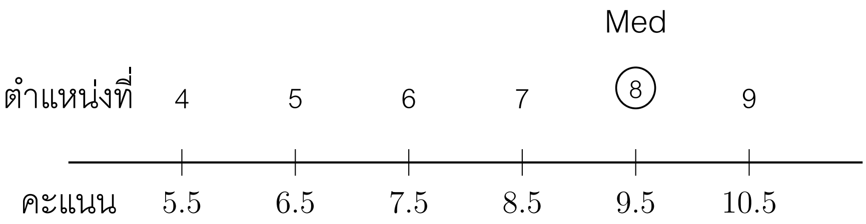
สำหรับข้อมูลที่จัดกลุ่มไว้แล้วที่มีจำนวนข้อมูล N จำนวน จะเป็นข้อมูลที่เรียงลำดับจากน้อยไปมาก ดังนั้นชั้นที่มีมัธยฐานอยู่คือชั้นแรกที่มีความถี่สะสมมากกว่า $\frac{N}{2}$ ซึ่งคำนวณได้โดยการเทียบบัญญัติไตรยางค์ ซึ่งพิจารณาจาก

1. ตำแหน่งของมัธยฐานเท่ากับ $\frac{N}{2}$
2. ตำแหน่งสุดท้ายของชั้นนั้นจะมีค่าเท่ากับขอบเขตจำกัดบนของชั้นนั้นเสมอ
3. ในแต่ละชั้นจะต้องแบ่งช่องของค่าข้อมูลออกเท่า ๆ กันเท่ากับจำนวนความถี่ของชั้นนั้น

ตัวอย่าง 2.1.38 จงหามัธยฐานของข้อมูลต่อไปนี้

คะแนน	ความถี่ (คน)	ความถี่สะสม
$1-5$	4	4
$6-10$	5	9
$11-15$	4	13
$16-20$	3	16

แนวคำตอบ เนื่องจาก $N = 16$ จะได้ว่ามัธยฐานอยู่ตำแหน่ง $\frac{16}{2} = 8$ ซึ่งอยู่ในอันตรภาคชั้นที่ 2 มีจำนวน 5 คน ดังนั้น แบ่งออกเป็น 5 ช่องเท่า ๆ กัน ได้ช่องละ $\frac{I}{5} = \frac{5}{5} = 1$ พิจารณาเส้นจำนวน

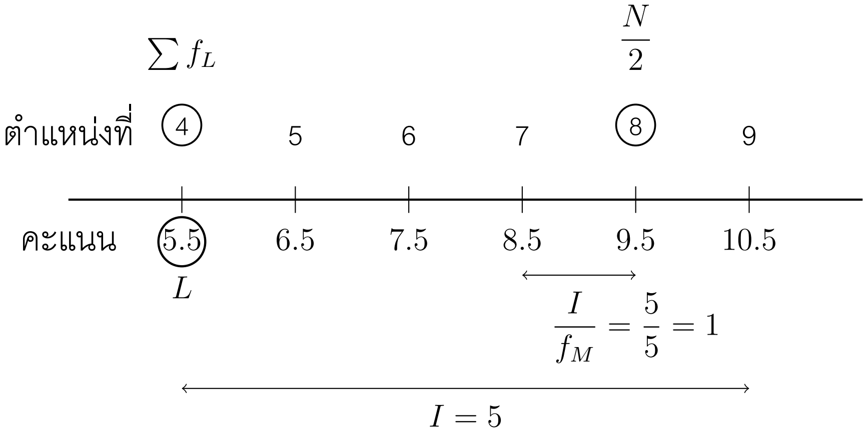


ดังนั้น Med $= 9.5$ คะแนน หรือค่ามัธยฐานเท่ากับ

ขอบเขตจำกัดล่างของชั้นมัธยฐาน + ความกว้าง 1 ช่อง $\times$ จำนวนช่องก่อนถึงมัธยฐาน

ให้ $L =$ ขอบเขตจำกัดล่างของชั้นมัธยฐาน $= L$ ความกว้าง 1 ช่อง $= \frac{I}{f_M}$

และจำนวนช่องจนถึงมัธยฐาน $= \frac{N}{2} - \sum f_L$ แสดงได้ดังเส้นจำนวน



จะได้ข้อสรุปดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2.1.39 สำหรับข้อมูลที่จัดกลุ่มในรูปตารางแจกแจงความถี่ที่มีขนาด N เมื่อเรียงข้อมูลจากน้อยไปมาก (ชั้นที่มาที่หลังมีค่ามากกว่าชั้นก่อนหน้า) จะได้ว่า

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

L คือขอบเขตจำกัดล่างของชั้นมัธยฐาน

I คือความกว้างของอันตรภาคของชั้นมัธยฐาน

f_M คือความถี่ของชั้นมัธยฐาน

$\sum f_L$ คือความถี่สะสมก่อนถึงชั้นมัธยฐาน

ตัวอย่าง 2.1.40 จงหามัธยฐานของข้อมูลต่อไปนี้

คะแนน	ความถี่ (คน)
$11-14$	3
$15-18$	8
$19-22$	12
$23-26$	7

แนวคำตอบ เนื่องจาก $N = 30$ จะได้ว่ามัธยฐานอยู่ตำแหน่ง $\frac{30}{2} = 15$ ซึ่งอยู่ในอันตรภาคชั้นที่ 3

คะแนน	ความถี่ (คน)	ความถี่สะสม
$11-14$	3	3
$15-18$	8	11
$19-22$	12	23
$23-26$	7	30

จะได้ว่า $L = 18.5$, $I = 4$, $f_M = 12$ และ $\sum f_L = 11$ ดังนั้น

$$\begin{aligned}\text{Med} &= 18.5 + \frac{4}{12}\left(\frac{30}{2} - 11\right)\\ &= 19.83\end{aligned}$$

3. ฐานนิยม (Mode)

บทนิยาม 2.1.41 ฐานนิยม (Mode) คือค่าของข้อมูลที่เกิดขึ้นบ่อยที่สุดหรือมีความถี่สูงสุด

ฐานนิยมมักใช้เป็นค่ากลางของข้อมูลเชิงคุณภาพ เขียนแทนฐานนิยมด้วย Mod

ตัวอย่าง 2.1.42 จงหาฐานนิยมของข้อมูลต่อไปนี้

1. $3, 2, 5, 6, 6$

   แนวคำตอบ เนื่องจาก 6 เกิดขึ้นบ่อยสุด ดังนั้นฐานนิยมเท่ากับ 6

2. $7, 9, 5, 7, 2, 3, 9$

   แนวคำตอบ เนื่องจาก 6 และ 9 เกิดขึ้นบ่อยสุด ดังนั้นฐานนิยมเท่ากับ 6 และ 9

3. $1, 3, 10, 8, 6, 9$

   แนวคำตอบ เนื่องจากไม่มีค่าใดซ้ำกัน ดังนั้นไม่มีฐานนิยม

ตัวอย่าง 2.1.43 ข้อมูลชุดที่หนึ่ง 4 จำนวนมีค่ามัธยฐานเท่ากับ 22 ฐานนิยมเท่ากับ 20 และพิสัยเท่ากับ 8 จะมีค่าเฉลี่ยเลขคณิตเท่ากับข้อใด

แนวคำตอบ จากข้อมูล ฐานนิยมเท่ากับ 20 จะได้ว่าข้อมูลเรียงจากน้อยไปมากคือ

20, 20, x, y

เนื่องจากพิสัยเท่ากับ 8 นั่นคือ $y - 20 = 8$ ดังนั้น $y = 28$ เนื่องจากมัธยฐานเท่ากับ 22 จะได้ว่า $\dfrac{20+x}{2} = 22$ ดังนั้น $x = 24$ ฉะนั้นค่าเฉลี่ยเลขคณิตเท่ากับ

$$\frac{20+20+24+28}{4} = 23$$

ตัวอย่าง 2.1.44 อายุของเด็ก 10 คน มีพิสัยเท่ากับ 1 และค่าเฉลี่ยเลขคณิตเท่ากับ 7.6 โดยมีอายุต่ำสุดในกลุ่มนี้เท่ากับ 7 ปี ถ้าทราบว่าอายุของเด็กทั้ง 10 เป็นจำนวนเต็ม จงหาฐานนิยมของเด็กกลุ่มนี้

แนวคำตอบ เนื่องจากพิสัยเท่ากับ 1 และอายุต่ำสุดในกลุ่มนี้เท่ากับ 7 ปี จะได้ว่า $X_{\text{max}} - 7 = 1$ นั่นคือ $X_{\text{max}} = 8$ ให้เด็กที่มีอายุ 7 ปีมีจำนวน n คน ดังนั้นเด็กที่มีอายุ 8 ปี มีจำนวน $10 - x$ คน ฉะนั้น

$$\begin{aligned} 7n + 8(10-n) &= 10(7.6) \\ n &= 4 \end{aligned}$$

อายุของเด็ก 10 คนคือ 7, 7, 7, 7, 8, 8, 8, 8, 8, 8 ดังนั้นฐานนิยมเท่ากับ 8 ปี

ทฤษฎีบท 2.1.45 สำหรับข้อมูลที่จัดกลุ่มในรูปตารางแจกแจงความถี่ที่มีขนาด N ฐานนิยมจะเป็นค่าที่อยู่ในชั้นที่มีความถี่สูงที่สุด จะได้ว่า

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

L คือขอบล่างของชั้นฐานนิยม

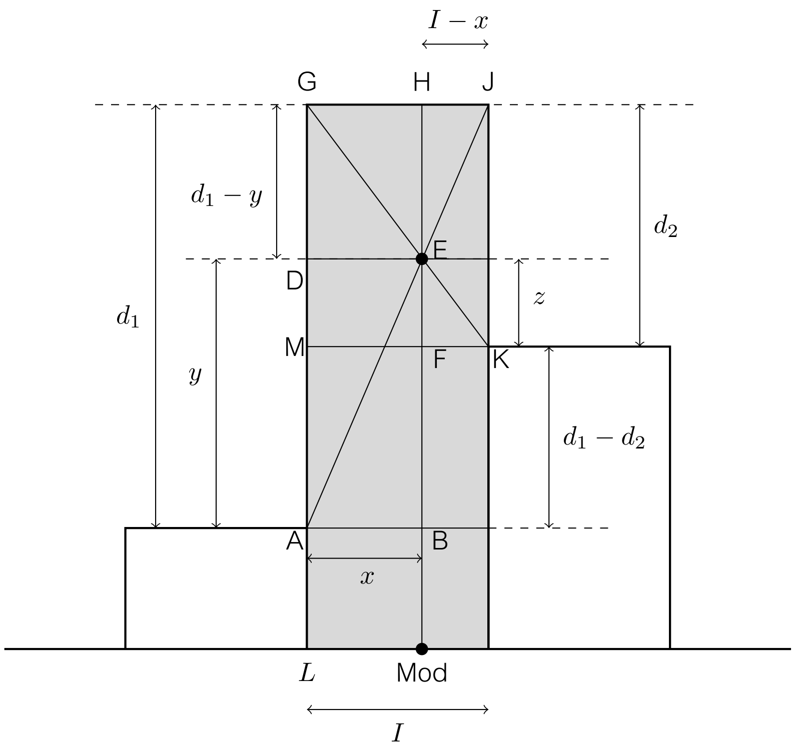
I คือความกว้างของอันตรภาคของชั้นฐานนิยม

d_1 คือผลต่างของความถี่ชั้นฐานนิยมกับชั้นก่อนชั้นฐานนิยม

d_2 คือผลต่างของความถี่ชั้นฐานนิยมกับชั้นหลังชั้นฐานนิยม

บทพิสูจน์. พิจารณาฮิสโทแกรมชั้นที่มีความถี่สูงสุด ดังรูป

รูปที่ 2.1: ฮิสโทแกรมแสดงการหาค่าฐานนิยม



จากรูปที่ 2.1 จะได้ว่า Mod $= L + x$
เนื่องจาก $\Delta ABE \sim \Delta JHE$ ดังนั้น $\dfrac{y}{x} = \dfrac{d_1 - y}{I - x}$ หรือ

$$y(I - x) = x(d_1 - y) = xd_1 - xy$$
$$yI - yx = xd_1 - xy$$
$$y = \frac{xd_1}{I}$$

เนื่องจาก $\Delta KFE \sim \Delta KMG$ ดังนั้น $\dfrac{z}{I-x} = \dfrac{d_2}{I}$ หรือ $z = \dfrac{d_2(I-x)}{I}$

เนื่องจาก $y = z + (d_1 - d_2)$ ฉะนั้น

$$\begin{aligned}\frac{xd_1}{I} &= \frac{d_2(I-x)}{I} + d_1 - d_2\\ d_1x &= d_2I - d_2x + d_1I - d_2I\\ x &= I\left(\frac{d_1}{d_1+d_2}\right)\end{aligned}$$

ดังนั้น Mod $= L + I\left(\dfrac{d_1}{d_1+d_2}\right)$ □

ตัวอย่าง 2.1.46 จงหาฐานนิยมของคะแนนสอบย่อยวิชาสถิติ ซึ่งข้อมูลแสดงดังตารางแจกแจงความถี่ต่อไปนี้

คะแนน	จำนวนคน
$10-14$	10
$11-19$	12
$20-24$	15
$25-29$	13
$30-34$	10

แนวคำตอบ จะเห็นว่าความถี่สูงสุดอยู่ชั้นที่ 3 ดังนั้นฐานนิยมอยู่ในชั้นที่ 3 โดยมี $L = 19.5$, $I = 5$, $d_1 = 15 - 12 = 3$ และ $d_2 = 15 - 13 = 2$ ดังนั้น

$$\text{Mod} = 19.5 + 5\left(\frac{3}{3+2}\right) = 22.5$$

แบบฝึกหัด 2.1

1. จากการเลือกตัวอย่างนักเรียนชั้น ม.3 มาจำนวน 20 คน เพื่อคำนวนคะแนนเฉลี่ยของวิชาคณิตศาสตร์ในการสอบ ONET ประจำปีการศึกษา 2560 เพื่อเป็นตัวแทนนักเรียนชั้น ม.3 ทั้งประเทศได้คะแนนดังต่อไปนี้ 45, 50, 51, 36, 24, 60, 90, 97, 100, 64, 23, 81, 77, 66, 64, 25, 50, 50, 60, 49 จงหาค่าเฉลี่ยของคะแนนของตัวอย่างดังกล่าว

2. คะแนนสอบ ONET วิชาวิทยาศาสตร์ ของนักเรียนห้องหนึ่ง แสดงดังแผนภาพต้น-ใบ ดังนี้

2	1	1	1	3	5	7	9	9
4	0	1	2	3	4			
6	2	3	5	6				
9	1	3	6					

จงหาเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม ของคะแนนสอบครั้งนี้

3. จากข้อมูลที่กำหนด จงหา x และ y

กลุ่มข้อมูล	ชุดที่ 1	ชุดที่ 2	ชุดที่ 2	รวมทั้ง 3 ชุด
ค่าเฉลี่ยเลขคณิต	x	25	15	24
จำนวนข้อมูล	30	y	40	100

4. ในการสอบวิชาภาษาไทยของนักเรียน 20 คน มีค่าเฉลี่ยเลขคณิตเป็น 12 คะแนน ต่อมาปรากฏว่ากรอกผิดไปสองคนโดยกรอก 10 และ 12 แต่ที่ถูกต้องคือ 20 และ 13 ตามลำดับ แล้วค่าเฉลี่ยเลขคณิตที่ถูกต้องเท่ากับเท่าใด

5. ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, x_3, ..., x_N$ ซึ่ง $\sum_{i=1}^{N} x_i^2 = A$ และ $\sum_{i=1}^{N}(x_i+1)^2 = B$ จงหาค่าเฉลี่ยของข้อมูลชุดนี้ในรูป N, A และ B

6. จงหาค่าเฉลี่ยเลขคณิตและมัธยฐานของข้อมูลชุดนี้ $x_1, x_2, x_3, ..., x_{1000}$ เมื่อ

$$x_n = \begin{cases} n & \text{เมื่อ } n \text{ เป็นจำนวนคี่} \\ 2n & \text{เมื่อ } n \text{ เป็นจำนวนคู่} \end{cases}$$

7. ในการแข่งขันกีฬามหาวิทยาลัยโลกครั้งที่ 24 ซึ่งประเทศไทยเป็นเจ้าภาพ มีการส่งรายชื่อนักกีฬาจากประเทศไทย 379 คน มีอายุเฉลี่ย 22 ปี ถ้ามีการถอนตัวนักกีฬาไทยออก 4 คน ซึ่งมีอายุ 24, 25, 25 และ 27 ปี และมีการเพิ่มนักกีฬาไทยอีก 5 คน ซึ่งมีอายุเฉลี่ย 17 ปี แล้วอายุเฉลี่ยของนักกีฬาจากประเทศไทยจะเท่ากับเท่าใด

8. จงหาค่าเฉลี่ยเลขคณิตของข้อมูล 49, 48, 50, 52, 60, 55, 52, 53 โดยใช้ทฤษฎีบท 2.1.7 เมื่อ $A = 50$

9. ในการสำรวจน้ำหนักตัว ของนักเรียนในชั้นเรียนที่มีนักเรียน 30 เป็นดังนี้

น้ำหนัก (กิโลกรัม)	ความถี่สะสม (คน)
$30-49$	10
$50-69$	26
$70-89$	30

ค่าเฉลี่ยเลขคณิตของน้ำหนักตัวของนักเรียนชั้นเรียนนี้เท่ากับกี่กิโลกรัม

10. ตารางต่อไปนี้เป็นข้อมูลเกี่ยวกับอายุของพนักงานจำนวน 50 คน

อายุไม่เกิน (ปี)	จำนวน (คน)
25	9
30	17
35	24
40	37
45	43
50	50

ถ้าอายุต่ำสุดของพนักงานคือ 21 ปี แล้วค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับเท่าใด

11. ข้อมูลชุดหนึ่งมี 99 จำนวนเรียงจากน้อยไปมากได้เป็น $x_1, x_2, x_3, ..., x_{99}$ ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้มีค่าเท่ากับมัธยฐาน จงแสดงว่า $\sum\limits_{i=1}^{49} |x_{50} - x_i| = \sum\limits_{i=51}^{99} |x_{50} - x_i|$

12. อายุของเด็กกลุ่มหนึ่ง มีการแจกแจงดังนี้

อายุ (ปี)	จำนวนเด็ก (คน)
$1-3$	3
$4-6$	a
$7-9$	6
$10-12$	4

ถ้ามัธยฐานของอายุเด็กคือ 7 ปี แล้ว a มีค่าเท่ากับเท่าใด

13. เมื่อสร้างตารางแจกแจงความถี่ของคะแนนนักเรียนจำนวน 36 คนโดยใช้ความกว้างของแต่ละอันตรภาคชั้นเป็น 10 แล้วปรากฏว่ามัธยฐานของคะแนนทั้งหมดอยู่ในช่วงคะแนน $50-59$ ถ้านักเรียนที่สอบได้คะแนนต่ำกว่า 49.5 คะแนน มีอยู่จำนวน 12 คน และนักเรียนได้คะแนนต่ำกว่า 59.5 มีอยู่จำนวน 20 คน แล้วมัธยฐานของคะแนนสอบครั้งนี้เท่ากับเท่าใด

14. ข้อมูล $x_1, x_2, x_3, ..., x_n$ มีมัธยฐานคือ Med จงพิสูจน์ว่า $\sum\limits_{i=1}^{n} |x_i - \text{Med}|$ มีค่าน้อยที่สุด

15. ข้อมูลชุด X คือ 11, 22, 33, 44, 55, 66, 77, 88, 99 ถ้าข้อมูลชุด Y คือ $y_i = \dfrac{3x_i - 1}{4}$ แล้วค่าเฉลี่ยเลขคณิตของข้อมูลชุด Y คือข้อใด

16. นักเรียนกลุ่มหนึ่งเป็นนักเรียนชาย m คน เป็นนักเรียนหญิง n คน หาค่าเฉลี่ยเลขคณิตของนักเรียนทั้งหมดได้ x ปี ถ้านักเรียนชายมีอายุเฉลี่ย y ปี แล้วนักเรียนหญิงจะมีค่าเฉลี่ยเลขคณิตของอายุเท่ากับเท่าใด

17. จงหาค่าเฉลี่ยเรขาคณิตของข้อมูล 1, 2, 4, 8, ..., 4096 เรียงกันเป็นลำดับเรขาคณิต

18. จงหามัธยฐานของข้อมูล 3, 7, 11, 15, ..., 799 เรียงเป็นลำดับเลขคณิต

19. โรงงานแห่งหนึ่งมีพนักงานจำนวน 40 คน และตารางแจกแจงความถี่สะสมของอายุพนักงานเป็นดังนี้

อายุ (ปี)	ความถี่สะสม
$11 - 20$	6
$21 - 30$	14
$31 - 40$	26
$41 - 50$	36
$51 - 60$	40

ถ้าผู้จัดการมีอายุ 48.5 ปี แล้วพนักงานที่มีอายุระหว่างมัธยฐานของพนักงานและอายุของผู้จัดการมีจำนวนประมาณเท่าใด

20. จากสมบัติเชิงเส้น (ทฤษฎีบท 2.1.19) เป็นจริงสำหรับค่าเฉลี่ย คุณคิดว่าจะเป็นจริงสำหรับมัธยฐาน และฐานนิยม หรือไม่เพราะเหตุใด จงยกตัวอย่างและให้เหตุผลประกอบคำตอบ

21. ครอบครัวหนึ่งมีบุตร 4 คน บุตร 2 คนมีน้ำหนักเท่ากันและมีน้ำหนักน้อยกว่าบุตรอีก 2 คน ถ้าน้ำหนักของบุตรทั้ง 4 คนมีค่าฐานนิยม มัธยฐาน และพิสัย เท่ากับ 45, 47.5 และ 7 กิโลกรัมตามลำดับ แล้วค่าเฉลี่ยเลขคณิตของน้ำหนักของบุตรทั้ง 4 คน มีค่าเท่ากับเท่าใด

22. ข้อมูลชุดหนึ่งเรียงลำดับจากน้อยไปมาก 10, 20, 30, 30, a, b, 60, 60, 90, 120 ถ้าฐานนิยมและมัธยฐานคะแนนชุดนี้เป็น 30 และ 40 ตามลำดับ แล้วข้อมูลชุดต่อไปนี้คือ

$$11,\ 22,\ 33,\ 34,\ a+5,\ b+6,\ 67,\ 68,\ 99,\ 130$$

จงหาค่าเฉลี่ยเลขคณิต

23. ข้อมูล $x_1, x_2, x_3, \ldots, x_n$ มีค่าเฉลี่ยเลขคณิตคือ $\overline{X}$ ถ้า $\displaystyle\sum_{i=1}^{n}(x_i - A) = B$ เมื่อ A, B เป็นค่าคงที่ จงแสดงว่า $\overline{X} = \dfrac{B}{n} + A$

24. ค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งเป็น 43 คะแนน ถ้าคิดค่าเฉลี่ยคะแนนสอบของนักเรียนชายและนักเรียนหญิงจะได้เป็น 45 และ 40 คะแนนตามลำดับ จงหาอัตราส่วนระหว่างจำนวนนักเรียนชายและจำนวนนักเรียนหญิง

25. กำหนดตารางแจกแจงความถี่แสดงอายุของคนในหมู่บ้านแห่งหนึ่งดังนี้

อายุ (ปี)	$0-9$	$10-19$	$20-29$	$30-39$	$40-49$	$50-59$
จำนวน (คน)	5	10	A	20	10	10

ถ้าอายุเฉลี่ยของคนในหมู่บ้านนี้เท่ากับ 33.33 ปี แล้วจำนวนคนในหมู่บ้านนี้เท่ากับเท่าใด

26. ข้อมูลชุดหนึ่งประกอบด้วยจำนวน 11, 3, 6, 3, 5, 3, x ให้ S เป็นเซตของ x ที่เป็นไปได้ทั้งหมดซึ่งทำให้ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลชุดนี้แตกต่างกันทั้งหมด และในบรรดาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม เหล่านี้นำมาเรียงใหม่จากน้อยไปหามากแล้วเป็นลำดับเลขคณิต จงหาผลบวกของสมาชิกทั้งหมดในเซต S

27. ในการวิจัยในชั้นเรียนเรื่องศึกษาผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์เรื่องพีทาโกรัสโดยใช้โปรแกรม Geogebra ผู้วิจัยให้นักเรียนทำแบบทดสอบก่อนเรียนและหลังเรียน (คะแนนเต็ม 10 คะแนน) ให้ผลดังนี้

คนที่	1	2	3	4	5	6	7	8	9	10	11	12
คะแนนก่อนเรียน	3	4	5	5	2	6	6	7	9	3	4	6
คะแนนหลังเรียน	5	6	4	9	7	8	9	10	10	9	9	8

27.1 จงหาค่าเฉลี่ยของคะแนนก่อนสอบ และหลังสอบ

27.2 จงหาค่าเฉลี่ยของคะแนนผลต่างระหว่างก่อนสอบและหลังสอบ

27.3 จงบอกความสัมพันธ์ของคะแนนเฉลี่ยก่อนสอบ คะแนนเฉลี่ยชหลังสอบ และคะแนนเฉลี่ยผลต่าง

2.2 การวัดตำแหน่งของข้อมูลเชิงปริมาณ

สำหรับข้อมูลเชิงปริมาณเมื่อเรียงลำดับจากน้อยบไปมาก เราจะทราบได้ว่าค่าใดมีค่ามากเพียงใดซึ่งระบุด้วยตำแหน่ง แต่ในหัวข้อนี้เราสนใจการแบ่งข้อมูลออกส่วน ๆ เท่ากัน ๆ กัน ที่นิยมใช้กันคือการแบ่งข้อมูลออกเป็น 4, 10 และ 100 ส่วน ที่เรียกว่า ควอไทล์ เดไซล์ และ เปอร์เซ็นต์ไทล์ ประโยชน์ของการแบ่งข้อมูลเพื่ออ้างอิงข้อมูลที่สนใจกับข้อมูลทั้งหมด เช่น นาย ก มีคะแนนสอบคณิตศาสตร์เท่ากับเปอร์เซ็นต์ไทล์ที่ 80 หมายถึงนาย ก ได้คะแนนสูงสุดของกลุ่ม 80% คะแนนต่ำสุด หรือมีนักเรียนที่เขาสอบได้คะแนนน้อยกว่านาย ก อยู่ 80%

1. ควอไทล์ (Quartile)

บทนิยาม 2.2.1 ควอไทล์ คือค่าของข้อมูลในตำแหน่งของข้อมูลแต่ละส่วนที่เกิดจากการแบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กันเมื่อเรียงลำดับจากน้อยไปมาก นิยมหาควอไทล์ 3 ตำแหน่ง คือ Q_1 (first quartile หรือ lower quartile), Q_2 (second quartile หรือ median) และ Q_3 (third quartile หรือ upper quartile)

ข้อสังเกต จะเห็นว่า Q_2 เท่ากับค่ามัธยฐานของข้อมูล

ตัวอย่าง 2.2.2 จงหาควอไทล์ทั้ง 3 ตำแหน่งของข้อมูล 2, 2, 3, 3, 5, 6, 9

แนวคำตอบ แบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน ดังต่อไปนี้

2	2	3	3	5	6	9
	Q_1		Q_2		Q_3	

ดังนั้น $Q_1 = 2$, $Q_2 = 3$ และ $Q_3 = 6$

ตัวอย่าง 2.2.3 จงหาควอไทล์ทั้ง 3 ตำแหน่งของ คะแนนสอบ ONET วิชาคณิตศาสตร์ 10 คนที่มีคะแนนต่อไปนี้

5	0	0	2	8		
6	0	2	3	5	5	7

แนวคำตอบ แบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน ดังต่อไปนี้

50	50	52	58	60		62	63	65	65	67
		Q_1			Q_2			Q_3		

ดังนั้น $Q_1 = 52$, $Q_2 = \dfrac{60+62}{2} = 61$ และ $Q_3 = 65$

สำหรับข้อมูลจัดกลุ่มการหาควอไทล์ของประชากรหรือตัวอย่างอาศัยแนวคิดเช่นเดียวกับการหาค่ามัธยฐาน ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2.2.4 สำหรับข้อมูลจัดกลุ่มมีขนาด N จำนวน เมื่อเรียงข้อมูลจากน้อยไปมากแล้วควอไทล์ที่ r จะมีตำแหน่งที่ $\frac{rN}{4}$ อยู่ในชั้นแรกที่มีความถี่สะสมมากกว่า $\frac{rN}{4}$ และ

$$Q_r = L + \frac{I}{f_{Q_r}}\left(\frac{rN}{4} - \sum f_L\right)$$

L คือขอบล่างของชั้น Q_r

I คือความกว้างของอันตรภาคของชั้น Q_r

f_{Q_r} คือความถี่ของชั้น Q_r

$\sum f_L$ คือความถี่สะสมก่อนถึงชั้น Q_r

ตัวอย่าง 2.2.5 จงหาควอไทล์ทั้ง 3 ตำแหน่งของ คะแนนสอบย่อยครั้งหนึ่งของวิชาสถิติ แสดงดังตารางต่อไปนี้

คะแนน	ความถี่ (คน)
$1-5$	5
$6-10$	20
$11-15$	35
$16-20$	10

แนวคำตอบ จะเห็นว่าควอไทล์ที่ 1, 2 และ 3 อยู่ตำแหน่ง $\frac{70}{4} = 17.5$, $\frac{2(70)}{4} = 35$ และ $\frac{3(70)}{4} = 52.5$, ตามลำดับ

คะแนน	ความถี่ (คน)	ความถี่สะสม
$1-5$	5	5
$6-10$	20	25
$11-15$	35	60
$16-20$	10	70

ควอไทล์ที่ 1 อยู่ตำแหน่ง 17.5 นั่นคืออยู่ชั้น 2 ดังนั้น

$$Q_1 = 5.5 + \frac{5}{20}(17.5 - 5) = 8.625$$

ควอไทล์ที่ 2 อยู่ตำแหน่ง 35 นั่นคืออยู่ชั้น 3 ดังนั้น

$$Q_2 = 10.5 + \frac{5}{35}(35 - 25) = 11.929$$

ควอไทล์ที่ 3 อยู่ตำแหน่ง 52.5 นั่นคืออยู่ชั้น 3 ดังนั้น

$$Q_3 = 10.5 + \frac{5}{35}(52.5 - 25) = 14.429$$

2. เดไซล์ (Decile)

บทนิยาม 2.2.6 เดไซล์ คือค่าของข้อมูลในตำแหน่งของข้อมูลแต่ละส่วนที่เกิดจากการแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กันเมื่อเรียงลำดับจากน้อยไปมาก เรานิยมหาเดไซล์ 9 ตำแหน่ง คือ $D_1, D_2, D_3, ..., D_9$

ข้อสังเกต จะเห็นว่า $D_5 = Q_2 =$ Med และ D_r หมายถึงมีจำนวนที่น้อยกว่า D_r อยู่ $r \times 10\%$ หรือมากกว่าอยู่ $(100 - 10r)\%$

ตัวอย่าง 2.2.7 จงหาเดไซล์ทั้ง 4, 6 และ 9 ตำแหน่งของข้อมูล 1, 2, 2, 3, 5, 7, 9, 10, 12, 15

แนวคำตอบ แบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน ดังต่อไปนี้

1	2	2	3	5	7	9	10	12	15

D_4 D_6 D_9

ดังนั้น $D_4 = \dfrac{3+5}{2} = 4$, $D_6 = \dfrac{7+9}{2} = 8$ และ $D_9 = \dfrac{12+15}{2} = 13.5$

ตัวอย่าง 2.2.8 เดไซล์ทั้ง 3, 7 และ 8 ตำแหน่งของคะแนนสอบย่อยแคลคูลัสจำนวน 19 คน ดังนี้

0, 1, 1, 2, 3, 3, 4, 4, 5, 5, 6, 6, 6, 7, 7, 8, 9, 10, 10

แนวคำตอบ แบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน ดังต่อไปนี้

0 1 1 2 3 3 4 4 5 5 6 6 6 7 7 8 9 10 10

D_3 D_7 D_8

ดังนั้น $D_3 = 3$, $D_7 = 7$ และ $D_8 = 8$

ตัวอย่าง 2.2.9 จากการแจกแจงข้อมูลเงินเดือนของพนักงานบริษัทแห่งหนึ่งพบว่า

เดไซล์ที่	1	3	5	7	9
เงินเดือน (บาท)	$10,000$	$15,000$	$20,000$	$25,000$	$40,000$

ถ้านายเอกและนายยศมีเงินเดือนรวมกันเท่ากับ $40,000$ บาท และมีจำนวนพนักงานที่ได้เงินเดือนมากกว่านายยศอยู่ประมาณ 30% ของพนักงานทั้งหมด แล้วเปอร์เซ็นต์ของพนักงานที่ได้เงินเดือนน้อยกว่านายเอกเท่ากับเท่าใด

แนวคำตอบ จำนวนพนักงานที่ได้เงินเดือนมากกว่านายยศอยู่ประมาณ 30% ของพนักงานทั้งหมด นั่นคือเงินเดือนของนายยศเท่ากับเดไซล์ที่ 7 หรือมีเงินเดือนเท่ากับ 25,000 บาท เนื่องจากนายเอกและนายยศมีเงินเดือนรวมกันเท่ากับ $40,000$ บาท ฉะนั้นนายเอกมีเงินเดือนเท่ากับ 15,000 บาท ดังนั้นเงินเดือนของนายเอกเท่ากับเดไซล์ที่ 3 หรือมีพนักงานได้เงินเดือนน้อยกว่านายเอกอยู่ประมาณ 30%

สำหรับข้อมูลจัดกลุ่มการหาเดไซล์ของประชากรหรือตัวอย่างอาศัยแนวคิดเช่นเดียวกับการหาค่ามัธยฐาน ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2.2.10 สำหรับข้อมูลจัดกลุ่มมีขนาด N จำนวน เมื่อเรียงข้อมูลจากน้อยไปมากแล้วเดไซล์ที่ r จะมีตำแหน่งที่ $\frac{rN}{10}$ อยู่ในชั้นแรกที่มีความถี่สะสมมากกว่า $\frac{rN}{10}$ และ

$$D_r = L + \frac{I}{f_{D_r}}\left(\frac{rN}{10} - \sum f_L\right)$$

L คือขอบล่างของชั้น D_r

I คือความกว้างของอันตรภาคของชั้น D_r

f_{D_r} คือความถี่ของชั้น D_r

$\sum f_L$ คือความถี่สะสมก่อนถึงชั้น D_r

ตัวอย่าง 2.2.11 จงหา D_2, D_6 และ D_9 ของคะแนน ONET วิชาวิทยาศาสตร์ แสดงดังตาราง

คะแนน	ความถี่ (คน)
$50 - 59$	10
$60 - 69$	20
$70 - 79$	15
$80 - 89$	5
$90 - 99$	2

แนวคำตอบ เดไซล์ที่ 2, 6 และ 9 อยู่ตำแหน่ง $\frac{2(52)}{10} = 10.4$, $\frac{6(52)}{10} = 31.2$ และ $\frac{9(52)}{10} = 46.8$, ตามลำดับ

คะแนน	ความถี่ (คน)	ความถี่สะสม
$50 - 59$	10	10
$60 - 69$	20	30
$70 - 79$	15	45
$80 - 89$	5	50
$90 - 99$	2	52

เดไซล์ที่ 2 อยู่ตำแหน่ง 10.4 นั่นคืออยู่ชั้น 2 ดังนั้น

$$D_2 = 59.5 + \frac{10}{20}(10.4 - 10) = 59.7$$

เดไซล์ที่ 6 อยู่ตำแหน่ง 31.2 นั่นคืออยู่ชั้น 3 ดังนั้น

$$D_6 = 69.5 + \frac{10}{15}(31.2 - 30) = 70.3$$

เดไซล์ที่ 9 อยู่ตำแหน่ง 46.8 นั่นคืออยู่ชั้น 4 ดังนั้น

$$D_9 = 79.5 + \frac{10}{5}(46.8 - 45) = 83.1$$

3. เปอร์เซ็นไทล์ (Percentile)

บทนิยาม 2.2.12 เปอร์เซ็นไทล์ คือค่าของข้อมูลในตำแหน่งของข้อมูลแต่ละส่วนที่เกิดจากการแบ่งข้อมูลออกเป็น 100 ส่วนเท่า ๆ กันเมื่อเรียงลำดับจากน้อยไปมาก เรานิยมหาเปอร์เซ็นไทล์ 99 ตำแหน่ง คือ $P_1, P_2, P_3, ..., P_{99}$

ข้อสังเกต จะเห็นว่า $P_{50} = D_5 = Q_2 =$ Med และ $P_{25} = Q_1$, $P_{75} = Q_3$ โดยที่ P_r หมายถึงมีจำนวนที่น้อยกว่า P_r อยู่ $r\%$ หรือมากกว่าอยู่ $(100 - r)\%$

ตัวอย่าง 2.2.13 ผลการสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง (เรียงจากน้อยไปมาก) เป็นดังนี้

40, 42, 45, 50, 51, 55, 56, 61, 63, 70, 75, 80, 81, 90

จงหาเปอร์เซ็นไทล์ที่ 20, 40, 60 และ 80

แนวคำตอบ แบ่งข้อมูลออกเป็น 5 ส่วนเท่า ๆ กัน ดังต่อไปนี้

40	42	45	50	51	55	56	61	63	70	75	80	81	90
		P_{20}			P_{40}			P_{60}			P_{80}		

ดังนั้น $P_{20} = 45$, $P_{40} = 55$, $P_{60} = 63$ และ $P_{80} = 80$

ตัวอย่าง 2.2.14 จากการตรวจสอบลำดับที่ของคะแนนสอบของนาย ก และนาย ข ในวิชาคณิตศาสตร์ที่มีผู้เข้าสอบ 400 คน ปรากฏว่านาย ก สอบได้คะแนนอยู่ในตำแหน่งควอไทล์ที่ 3 และนาย ข สอบได้คะแนนอยู่ในตำแหน่งเปอร์เซ็นไทล์ที่ 60 จำนวนนักเรียนที่สอบได้คะแนนระหว่างคะแนนนาย ก และนาย ข มีประมาณกี่คน

แนวคำตอบ นาย ก สอบได้คะแนนอยู่ในตำแหน่งควอไทล์ที่ 3 นั่นคือคะแนนอยู่ในตำแหน่งเปอร์เซ็นไทล์ที่ 75 และนาย ข สอบได้คะแนนอยู่ในตำแหน่งเปอร์เซ็นไทล์ที่ 60 ดังนั้นมีนักเรียนได้คะแนนระหว่างคะแนนนาย ก และนาย ข ประมาณ 15% หรือ $0.15 \times 400 = 60$ คน

สำหรับข้อมูลจัดกลุ่มการหาเปอร์เซ็นไทล์ของประชากรหรือตัวอย่างอาศัยแนวคิดเช่นเดียวกับการหาค่ามัธยฐาน ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2.2.15 สำหรับข้อมูลจัดกลุ่มมีขนาด N จำนวน เมื่อเรียงข้อมูลจากน้อยไปมากแล้วเปอร์เซ็นไทล์ที่ r จะมีตำแหน่งที่ $\frac{rN}{100}$ อยู่ในชั้นแรกที่มีความถี่สะสมมากกว่า $\frac{rN}{100}$ และ

$$P_r = L + \frac{I}{f_{P_r}}\left(\frac{rN}{100} - \sum f_L\right)$$

L คือขอบล่างของชั้น P_r

I คือความกว้างของอันตรภาคของชั้น P_r

f_{P_r} คือความถี่ของชั้น P_r

$\sum f_L$ คือความถี่สะสมก่อนถึงชั้น P_r

ตัวอย่าง 2.2.16 จงหา P_{40}, P_{54} และ P_{85} ของคะแนนเก็บวิชาภาษไทย แสดงดังนี้

คะแนน	ความถี่สะสม
$31-40$	7
$41-50$	15
$51-60$	25
$61-70$	45
$71-80$	50

แนวคำตอบ จะเห็นว่าเปอร์เซ็นไทล์ที่ 40, 54 และ 85 อยู่ตำแหน่ง $\frac{40(50)}{100} = 20$, $\frac{54(50)}{100} = 27$ และ $\frac{85(50)}{100} = 42.5$, ตามลำดับ

คะแนน	ความถี่ (คน)	ความถี่สะสม
$31-40$	7	7
$41-50$	8	15
$51-60$	10	25
$61-70$	20	45
$71-80$	5	50

เปอร์เซ็นไทล์ที่ 40 อยู่ตำแหน่ง 20 นั่นคืออยู่ชั้น 3 ดังนั้น

$$P_{40} = 50.5 + \frac{10}{10}(20-15) = 55.5$$

เปอร์เซ็นไทล์ที่ 54 อยู่ตำแหน่ง 27 นั่นคืออยู่ชั้น 4 ดังนั้น

$$P_{54} = 60.5 + \frac{10}{20}(27-25) = 61.5$$

เปอร์เซ็นไทล์ที่ 85 อยู่ตำแหน่ง 42.5 นั่นคืออยู่ชั้น 4 ดังนั้น

$$P_{85} = 60.5 + \frac{10}{20}(42.5-25) = 69.25$$

ตัวอย่าง 2.2.17 ในการสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งมีจำนวน 30 คน ปรากฏว่ามีนักเรียน 17 คน สอบได้คะแนนในช่วง $10-39$ คะแนน มีนักเรียน 10 คน สอบได้คะแนนในช่วง $40-49$ คะแนน และมีนักเรียน 3 คน สอบได้คะแนนในช่วง $50-59$ คะแนน ถ้าแบ่งคะแนนออกเป็นเกรด 3 ระดับ คือเกรด A เกรด B และ เกรด C โดยที่ 10% ของนักเรียนได้เกรด A และ 20% ของนักเรียนได้เกรด B คะแนนสูงสุดของเกรด C เท่ากับกี่คะแนน

แนวคำตอบ จากข้อมูลเขียนตารางได้ดังนี้

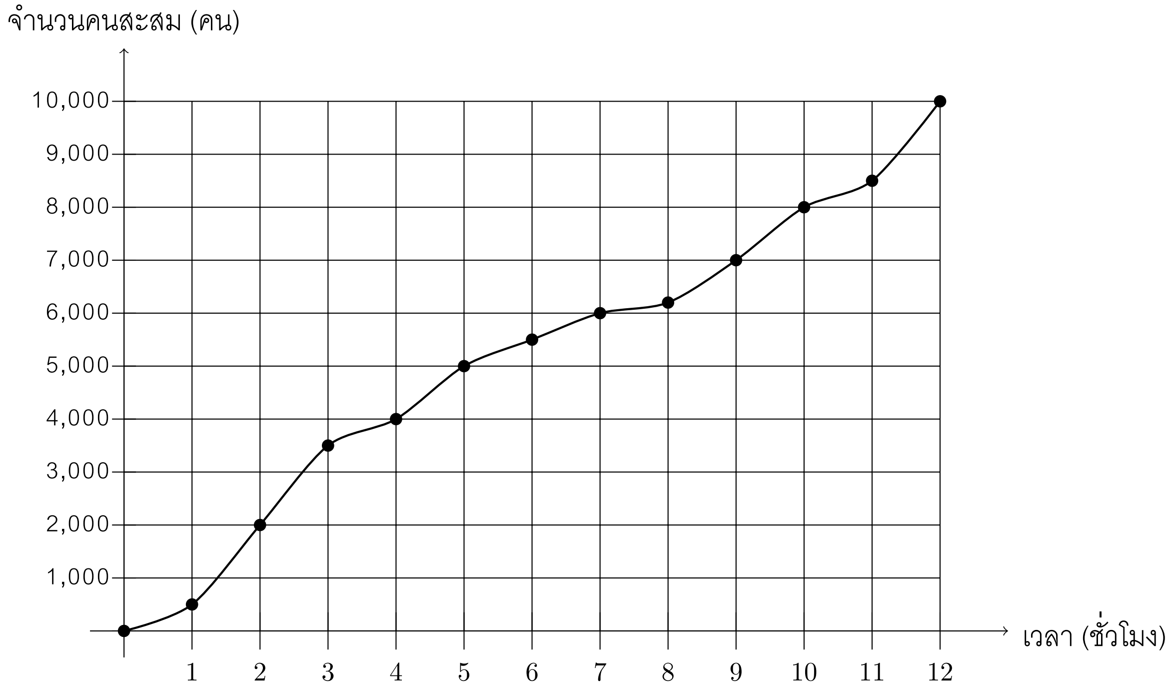
ช่วงคะแนน	จำนวนคน	จำนวนคนสะสม
$10-39$	17	17
$40-49$	10	27
$50-59$	3	30

เนื่องจาก 10% ของนักเรียนได้เกรด A และ 20% ของนักเรียนได้เกรด B ดังนั้นจำนวนนักเรียนที่ได้เกรด C เท่ากับ 70% ซึ่งหมายถึง 70% แรกของจำนวนนักเรียนทั้งหมด นั่นคือ P_{70}
ตำแหน่งของ P_{70} คือ $70 \cdot \dfrac{30}{100} = 21$ อยู่ในชั้นที่ 2 โดยมี $L = 39.5$, $I = 10$, $\sum f_L = 17$ และ $f_{P_{70}} = 10$ ดังนั้น

$$P_{70} = 39.5 + \frac{10}{10}(21 - 17) = 43.5$$

ดังนั้นคะแนนสูงสุดของเกรด C เท่ากับ 43.5 คะแนน

ตัวอย่าง 2.2.18 จากการสำรวจจำนวนชั่วโมงของลูกค้าที่มา shopping ที่ห้างสรรพสินค้าแห่งหนึ่งจำนวน 10,000 คน แสดงดังกราฟต่อไปนี้



1. จงประมาณค่าเปอร์เซ็นไทล์ที่ 75

   แนวคำตอบ เปอร์เซ็นไทล์ที่ 75 ตรงกับคนในลำดับที่ 7,500 จากกราฟจะได้ว่าประมาณ 9.5 ชั่วโมง
   ดังนั้น $P_{75} = 9.5$

2. จงประมาณค่าคนที่มา shopping ใช้เวลา 3 ชั่วโมง ตรงกับเปอร์เซ็นไทล์ที่เท่าใด

   แนวคำตอบ จากกราฟคนที่มา shopping ใช้เวลา 3 ชั่วโมง มีจำนวนประมาณ 3,500 คน ซึ่งตรงกับเปอร์เซ็นไทล์ที่ 35

3. ถ้านาย ก มา shopping ใช้เวลา 7 ชั่วโมง มีคนที่มา shopping ใช้เวลามากกว่านาย ก กี่เปอร์เซ็นต์

   แนวคำตอบ จากกราฟคนที่มา shopping ใช้เวลา 7 ชั่วโมง มีจำนวนประมาณ 6,000 คน ซึ่งตรงกับเปอร์เซ็นไทล์ที่ 60 ดังนั้นมีคนที่มา shopping ใช้เวลามากกว่านาย ก คิดเป็น 40 เปอร์เซ็นต์

แบบฝึกหัด 2.2

1. จงหาควอไทล์ทั้ง 3 ตำแหน่งของข้อมูลต่อไปนี้

1	2	5	5	5	8	9				
2	0	1	1	2	2	2	3	5	6	8
3	1	2	3							

2. จงหาควอไทล์ทั้ง 3 ตำแหน่งของคะแนนเก็บวิชาภาษาไทย

คะแนน	ความถี่สะสม
$61-70$	3
$51-60$	28
$41-50$	40
$31-40$	50
$21-30$	75
$11-20$	80

3. ข้อมูลชุดหนึ่งมี 5 จำนวน ถ้าควอไทล์ที่หนึ่ง ควอไทล์ที่สอง และควอไทล์ที่สามเท่ากับ 18, 25 และ 28 ตามลำดับ แล้วค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับเท่าใด

4. จงหา D_3, D_7, D_8 ของข้อมูลต่อไปนี้

1	1	2	5	5	8	9			
2	0	2	2	2	3	3	3	5	6
4	0	3	4						

5. จงหา D_2, D_6, D_8 ของข้อมูลต่อไปนี้

คะแนน	ความถี่ (คน)
$11-14$	3
$15-18$	8
$19-22$	12
$23-26$	7

6. จงหาตำแหน่งของเปอร์เซ็นไทล์ที่มีคะแนนเป็น 30.5, 45 และ 60 ของข้อมูลต่อไปนี้

คะแนน	ความถี่สะสม
$61-70$	3
$51-60$	28
$41-50$	40
$31-40$	50
$21-30$	75
$11-20$	80

7. ข้อมูลชุดหนึ่งประกอบด้วย 19 จำนวนต่อไปนี้

6 8 9 12 12 15 15 16 18 19
20 20 21 22 23 24 25 30 30

คอไทล์ที่ 3 มีค่าต่างจากเปอร์เซ็นไทล์ที่ 45 เท่ากับเท่าใด

8. คะแนนของผู้เข้าสอบ 15 คนเป็นดังนี้

45 54 59 60 62 64 65 68 70 72 73 75 76 80 81

ถ้าเกณฑ์ในการสอบผ่านคือต้องได้คะแนนไม่ต่ำกว่าเปอร์เซ็นไทล์ที่ 60 จงหาคะแนนต่ำสุดของผู้ที่สอบผ่าน

9. เมื่อพิจารณาผลการสอบวิชาคณิตศาสตร์ของนักเรียน 39 คน พบว่าเปอร์เซ็นไทล์ที่ 25 ของคะแนนสอบเท่ากับ 35 คะแนน และมีนักเรียน 30 คน ได้คะแนนน้อยกว่าหรือเท่ากับ 80 คะแนน ถ้ามีนักเรียนที่สอบได้ 35 คะแนนเพียงคนเดียว แล้วจำนวนนักเรียนที่สอบได้คะแนนในช่วง $35 - 80$ คะแนนเท่ากับเท่าใด

10. นักเรียนกลุ่มหนึ่งจำนวน 50 คน มีส่วนสูงแสดงดังตารางต่อไปนี้

ส่วนสูง (เซนติเมตร)	จำนวนนักเรียน (คน)
$156 - 160$	6
$161 - 165$	15
$166 - 170$	21
$171 - 175$	8

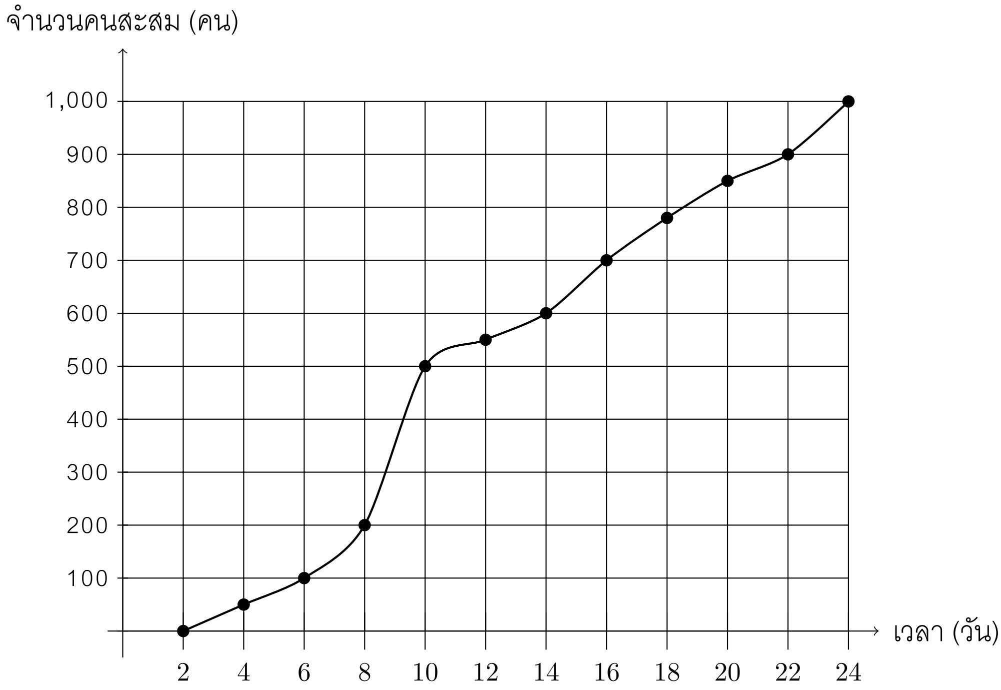
ถ้า a เป็นส่วนสูงโดยที่มีจำนวนนักเรียน 75% ของนักเรียนทั้งหมดที่มีส่วนสูงน้อยกว่า b แล้ว a มีค่าเท่าใด

11. จากตารางแจกแจงความถี่ต่อไปนี้

ค่าจ้าง (บาท)	จำนวนคน
$300 - 309$	1
$310 - 319$	3
$320 - 329$	x
$330 - 339$	5
$340 - 349$	8
$350 - 359$	y
$360 - 369$	4
$370 - 379$	10

ถ้าข้อมูลชุดนี้มี $P_{25} = 339.5$ และ $Q_3 = 359.5$ แล้วจำนวนคนงานที่ได้ค่าจ้างรายวันต่ำกว่า 349.5 บาทเท่ากับเท่าใด

12. จากการสำรวจจำนวนวันที่ต้องพักฟื้นในโรงพยาบาลของผู้ป่วยโควิด-19 จำนวน 1,000 คน แสดงดังกราฟต่อไปนี้



12.1 มีผู้ป่วยกี่เปอร์เซ็นต์ที่พักฟื้นไม่เกิน 10 วัน

12.2 จงหาเปอร์เซ็นไทล์ที่ 60

12.3 ผู้ป่วยที่พักฟื้นมากกว่า 14 วัน มากกว่าผู้ป่วยที่พักฟื้นน้อยกว่า 7 วัน ร้อยละเท่าใด

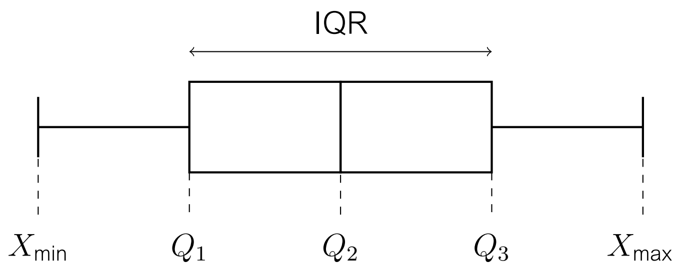
2.3 การวัดการกระจาย

บทนิยาม 2.3.1 พิสัยควอไทล์ (Inter Quartile Range) ของข้อมูลชุดหนึ่ง เขียนแทนด้วย IQR หมายถึง $Q_3 - Q_1$

แผนภาพกล่อง (Box plot) คือแผนภาพที่แสดงค่าสำคัญ 5 ค่าคือ

$$X_{\text{min}},\quad Q_1,\quad Q_2,\quad Q_3 \quad \text{และ} \quad X_{\text{max}}$$

แสดงได้ดังแผนภาพนี้



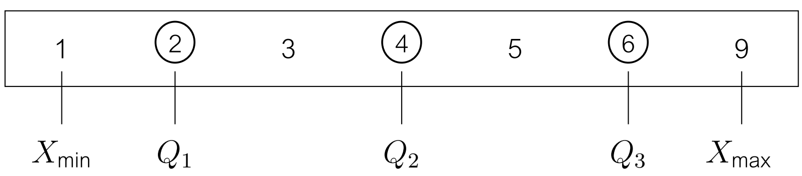
จากการกำหนดค่าดังกล่าวคือการแบ่งข้อมูลทั้งหมดออกเป็น 4 ส่วนเท่า ๆ กันส่วนละ 25% โดยมีเงื่อนไข

1. X_{min} เป็นค่าต่ำสุดยังไม่ต่ำผิดปกติ ถ้ามีค่าไม่น้อยกว่า $Q_1 - 1.5$IQR
2. X_{max} เป็นค่าสูงสุดยังไม่สูงผิดปกติ ถ้ามีค่าไม่เกินกว่า $Q_3 + 1.5$IQR

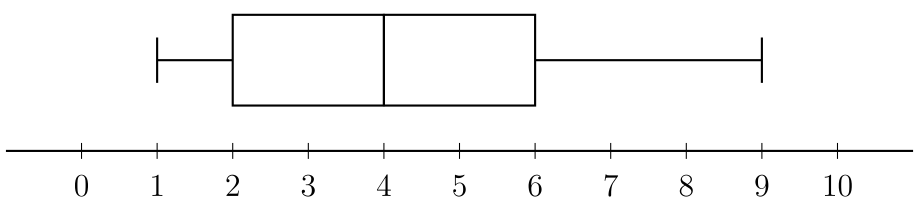
และเรียกว่า IQR ว่าความกว้างของกล่อง

ตัวอย่าง 2.3.2 จงสร้างแผนภาพกล่องของข้อมูล 1, 2, 3, 4, 5, 6, 9

แนวคำตอบ แบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน ดังต่อไปนี้



จะได้ว่า IQR $= Q_3 - Q_1 = 6 - 2 = 4$ โดยที่ $1 > -4 = 2 - 1.5(4)$ และ $9 < 12 = 6 + 1.5(4)$ นำไปเขียนแผนภาพกล่องได้ดังนี้



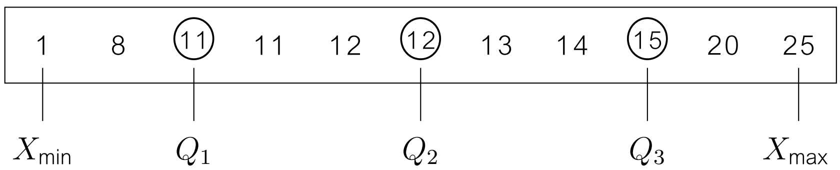
ในกรณีที่มี x เป็นข้อมูลหนึ่งในกลุ่ม ถ้า

$$x < Q_1 - 1.5\text{IQR} \quad \text{หรือ} \quad x > Q_3 + 1.5\text{IQR}$$

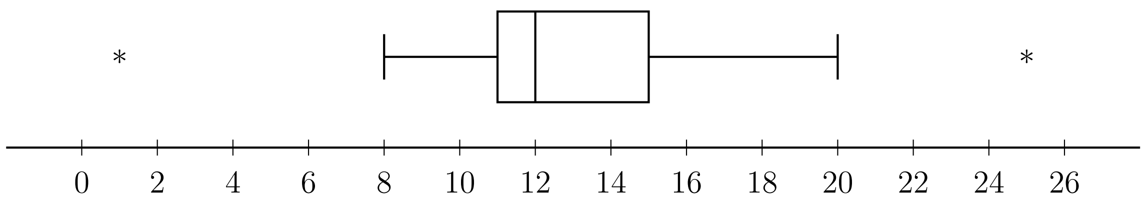
เรียก x ว่า **ค่าผิดปกติ** (Outlier) หรือกล่าวว่า x มีค่า**ต่ำผิดปกติ**หรือ**สูงผิดปกติ** ตามลำดับ ในการสร้างแผนภาพกล่องให้เขียนแทนด้วย $*$ ถ้า x มีค่าต่ำผิดปกติให้ใช้ค่าต่ำสุดตัวแรกที่ไม่ผิดปกติ และถ้า x มีค่าสูงผิดปกติให้ใช้ค่าสูงสุดตัวสุดท้ายที่ไม่ผิดปกติ

ตัวอย่าง 2.3.3 จงสร้างแผนภาพกล่องของข้อมูล 1, 8, 11, 11, 12, 12, 13, 14, 15, 20, 25

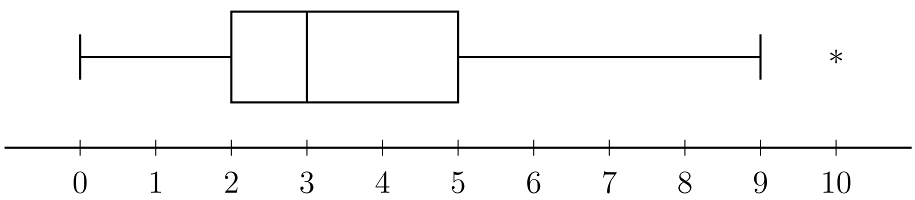
แนวคำตอบ แบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน ดังต่อไปนี้



จะได้ว่า IQR $= Q_3 - Q_1 = 15 - 11 = 4$ เนื่องจาก $1 < 5 = 11 - 1.5(4)$ ดังนั้น 1 เป็นค่าต่ำผิดปกติ และ $25 > 21 = 15 + 1.5(4)$ ดังนั้น 1 เป็นค่าสูงผิดปกติ
นำไปเขียนแผนภาพกล่องได้ดังนี้



ตัวอย่าง 2.3.4 คะแนนสอบย่อยวิชาแคลคูลัส ๒ ของนักศึกษาจำนวน 30 คน แสดงแผนภาพกล่องได้ดังนี้



1. จงหาคะแนนต่ำสุดและสูงสุดของการสอบย่อยครั้งนี้

   แนวคำตอบ จากแผนภาพกล่องจะได้ว่า $X_{\text{min}} = 0$ และ $X_{\text{max}} = 10$

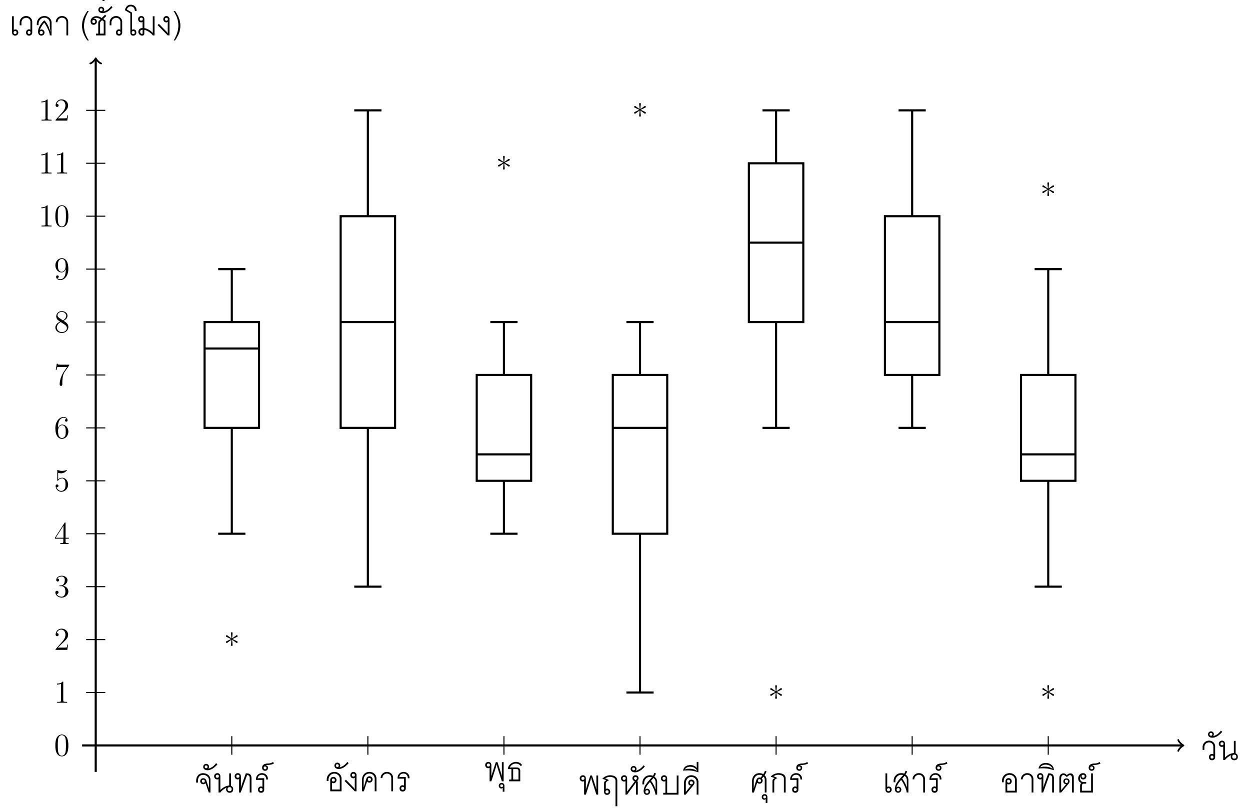
2. จงหาพิสัยควอไทล์

   แนวคำตอบ จากแผนภาพกล่องจะได้ว่า $Q_1 = 2$ และ $Q_3 = 5$ ดังนั้น IQR $= 5 - 2 = 3$

3. ถ้านาย ก อยู่ในกลุ่มนี้ และมีคนได้คะแนนมากกว่านาย ก อยู่ประมาณ 25% แล้วนาย ก มีคะแนนเท่าใด

   แนวคำตอบ จะได้ว่านาย ก มีคะแนนเท่ากับควอไทล์ที่ 3 ดังนั้น นาย ก มีคะแนนเท่ากับ 5 คะแนน

ตัวอย่าง 2.3.5 ในสถานการณ์โรคระบาดโควิด-19 โรงพยาบาลแห่งหนึ่ง จึงสำรวจเวลา (ชั่วโมง) ของผู้มาใช้บริการโรงพยาบาลในแต่ละวันเพื่อปรับปรุงการให้บริการ แสดงได้ดังนี้



1. ในสัปดาห์นี้ผู้มาใช้บริการโรงพยาบาลที่ใช้เวลานานสุดวันใดบ้าง และเป็นเวลากี่ชั่วโมง

   แนวคำตอบ วันอังคาร พฤหัสบดี ศุกร์ และเสาร์ ใช้เวลา 12 ชั่วโมง

2. ถ้านาย ก และนาย ข ไปใช้บริการโรงพยาบาลแห่งนี้วันอังคารและวันอาทิตย์ตามลำดับ ทั้งคู่ใช้เวลาตรงกับควอไทล์ที่ 3 ของแต่ละวัน ถามว่าใครใช้เวลาในโรงพยาบาลมากกว่ากันและมากกว่าเท่าใด

   แนวคำตอบ นาย ก ใช้เวลา 10 ชั่วโมง และนาย ข ใช้เวลา 7 ชั่วโมง ดังนั้นนาย ก ใช้เวลามากกว่านาย ข 3 ชั่วโมง

3. จงหาค่าผิดปกติทั้งหมดของการบริการครั้งนี้

   แนวคำตอบ

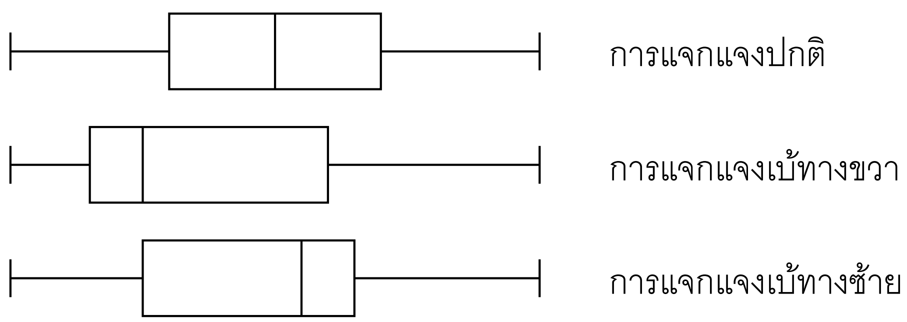
วันจันทร์	มีค่าต่ำผิดปกติคือ	2	ชั่วโมง
วันศุกร์	มีค่าต่ำผิดปกติคือ	1	ชั่วโมง
วันอาทิตย์	มีค่าต่ำผิดปกติคือ	1	ชั่วโมง
วันพุธ	มีค่าสูงผิดปกติคือ	11	ชั่วโมง
วันอาทิตย์	มีค่าสูงผิดปกติคือ	11	ชั่วโมง
วันพฤหัสบดี	มีค่าสูงผิดปกติคือ	12	ชั่วโมง

แผนภาพกล่องนั้นอาจทำให้บอกถึงลักษณะการกระจายของข้อมูลหรือการแจกแจงของข้อมูล ซึ่งมี 3 ชนิดหลัก ๆ คือ

1. การแจกแจงปกติหรือการแจกแจงสมมาตร (Normal or Symmetric distribution)
2. การแจกแจงเบ้ทางขวา (Right-skewed distribution)
3. การแจกแจงเบ้ทางซ้าย (Left-skewed distribution)

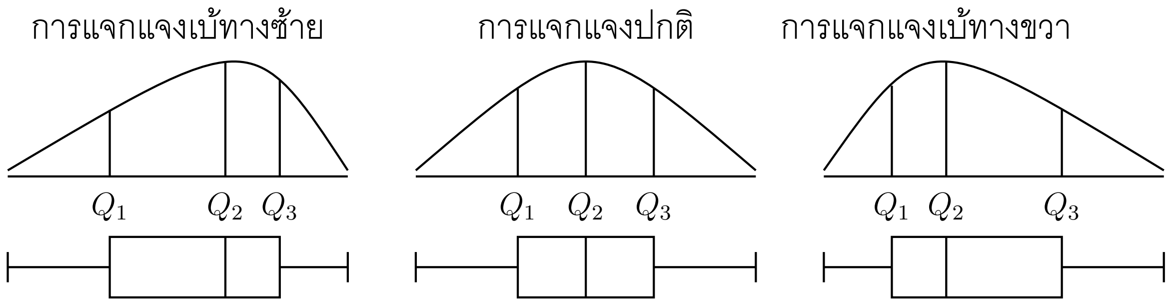
ความสัมพันธ์ของการแจกแจงกับแผนภาพกล่องแสดงได้ดังนี้

รูปที่ 2.2: ความสัมพันธ์ของการแจกแจงกับแผนภาพกล่อง



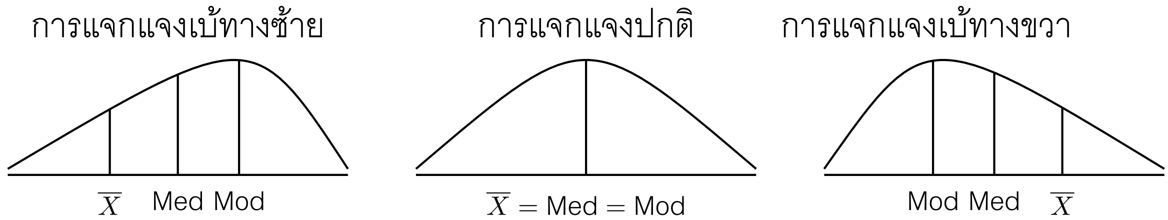
จากรูปที่ 2.2 แผนภาพกล่องที่มีลักษณะสมมาตรจะมีการแจกแจงปกติ แผนภาพกล่องที่มีจำนวนข้อมูลส่วนใหญ่มีค่าน้อยจะมีการแจกแจงเบ้ทางขวา และแผนภาพกล่องที่มีจำนวนข้อมูลส่วนใหญ่มีค่ามากจะมีการแจกแจงเบ้ทางซ้าย และแผนภาพกล่องอาจจะสัมพันธ์กับเส้นโค้งความถี่แสดงได้ดังรูป ต่อไปนี้

รูปที่ 2.3: ความสัมพันธ์ของเส้นโค้งความถี่กับแผนภาพกล่อง



เส้นโค้งความถี่ทั้ง 3 รูปแบบนี้จะสัมพันธ์กับค่ากลาง 3 ค่าคือ ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน และค่าฐานนิยม แสดงดังรูปต่อไปนี้

รูปที่ 2.4: ความสัมพันธ์ของเส้นโค้งความถี่กับค่ากลางของข้อมูล



การนำเสนอข้อมูลด้วยแผนภาพกล่องก็อาจทำให้ทราบเพียงลักษณะการกระจายเท่านั้นแต่ยังไม่มีค่าบ่งชี้ว่ากระจายมากหรือน้อย ต่อไปนี้จะกล่าวค่าในการวัดการกระจายของข้อมูลทั้งแบบไม่เปรียบกับกลุ่มอื่น และเปรียบเทียบกับกลุ่มอื่น ซึ่งมี 2 แบบคือ การวัดการกระจายสัมบูรณ์ (Absolute Variation) และ การวัดการกระจายสัมพัทธ์ (Relative Variation)

1. การวัดกระจายสัมบูรณ์

การวัดการกระจายสัมบูรณ์ คือการวัดการกระจายของข้อมูลเพียงชุดเดียวเพื่อดูว่าข้อมูลชุดนี้แต่ละค่ามีความแตกต่างกันมากหรือน้อยเพียงใด การวัดการกระจายแบบสัมบูรณ์ไม่สามารถนำไปเปรียบเทียบกับข้อมูลชุดอื่นได้ การวัดการกระจายสัมบูรณ์ที่นิยมใช้มี 4 วิธี คือ

1. พิสัย (Range)
2. ส่วนเบี่ยงเบนควอไทล์ (Quartile Deviation)
3. ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation)
4. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

พิสัยคือค่าสูงสุดลบค่าต่ำสุดใช้ในการวัดการกระจายของข้อมูลได้ไม่ละเอียดนัก นิยมใช้เมื่อต้องการความรวดเร็วเท่านั้น ข้อเสียของพิสัยของข้อมูลแต่ละชุด คือมีการใช้เฉพาะคะแนนสูงสุดและคะแนนต่ำสุดเท่านั้น บางครั้งทำให้เกิดการเข้าใจถึงลักษณะของข้อมูลผิดไปได้ แต่ก็ใช่ว่าพิสัยจะไม่มีข้อดีเสียเลย ถ้าพิสัยมีค่าน้อยก็อาจคาดได้ว่าข้อมูลส่วนใหญ่ก็ไม่ห่างกันมากนัก เช่น คะแนนสอบย่อยวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 10 คะแนน ถ้าพิสัยมีค่าเท่ากับ 2 คะแนน แสดงว่าคะแนนของนักเรียนต่างกันไม่เกิน 2 คะแนน ยิ่งไปกว่านั้นถ้าพิสัยมีค่าเท่ากับ 0 ก็จะหมายความว่านักเรียนทุกคนได้คะแนนเท่ากันนั่นเอง

ถ้าข้อมูลใดใช้ค่ากลางเป็นมัธยฐานเป็นตัวแทนข้อมูล มักจะนิยมใช้ส่วนเบี่ยงเบนควอไทล์เป็นการวัดการกระจายของข้อมูล โดยที่ **ส่วนเบี่ยงเบนควอไทล์** คือครึ่งหนึ่งของพิสัยควอไทล์ เขียนแทนด้วย $Q.D.$ นั่นคือ

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

ตัวอย่าง 2.3.6 คะแนนสอบย่อยวิชาแคลคูลัส ๑ ของนักเรียนกลุ่มหนึ่ง ประกอบด้วย 0, 1, 3, 3, 10 จงหาพิสัยและส่วนเบี่ยงเบนควอไทล์ พร้อมเลือกค่าการวัดที่คิดว่าเหมาะสมโดยให้เหตุผลประกอบ

แนวคำตอบ พิสัย $= 10 - 0 = 10$ เนื่องจาก $Q_1 = 0.5$ และ $Q_3 = 6.5$ จะได้ว่า

$$Q.D. = \frac{6.5 - 0.5}{2} = 2.5$$

ถ้าเลือกพิสัยซึ่งเท่ากับ 10 เป็นค่าการกระจายจะเห็นว่าข้อมูลนี้รู้เพียงว่าค่ามากสุดกับค่าต่ำสุดต่างกัน 10 คะแนน ถ้าคะแนนเต็ม 10 จะทราบได้ทันทีว่ามีคนได้ 0 คะแนน และได้ 10 คะแนน เนื่องจากคะแนนในกลุ่มนี้ส่วนมากได้คะแนนต่ำค่ากลางที่เหมาะสมควรเป็นมัธยฐาน จึงควรใช้การวัดการกระจายโดยใช้ส่วนเบี่ยงเบนควอไทล์ซึ่งเท่ากับ 2.5 จะเหมาะสมกว่า

พิสัยและส่วนเบี่ยงควอไทล์นั้นเป็นการวัดที่ใช้ข้อมูลบางส่วนเท่านั้นอาจทำให้เกิดความเข้าใจลักษณะของข้อมูลที่ผิดพลาดไปได้ จึงมีการวัดโดยใช้ทุกข้อมูลเข้ามาเกี่ยวข้องซึ่งอาศัยหลักการที่ว่าถ้าข้อมูลชุดนั้นเหมาะกับค่ากลางที่เป็นค่าเฉลี่ยเลขคณิต แล้วจะวัดผลต่างของทุกข้อมูลแล้วหาค่าเฉลี่ยเหล่านั้นเรียกว่า ส่วนเบี่ยงเบนเฉลี่ย และเป็นที่นิยมอย่างมากคือหารากที่สองของค่าเฉลี่ยของผลต่างกำลังสองจะเรียกว่า ส่วนเบี่ยงเบนมาตรฐาน ดังจะกล่าวต่อไปนี้

บทนิยาม 2.3.7 กำหนดให้ประชากรขนาด N คือ $X_1, X_2, X_3, ..., X_N$ และสุ่มตัวอย่างขนาด n จากประชากรดังกล่าวมาได้เป็น $x_1, x_2, x_3, ..., x_n$ โดยที่ประชากรและตัวอย่างมีค่าเฉลี่ยเลขคณิตเท่ากับ μ และ $\overline{X}$ ตามลำดับ

1. **ส่วนเบี่ยงเบนเฉลี่ย** เขียนแทนด้วย M.D. คือค่านิยามโดย

   สำหรับตัวอย่าง: $M.D. = \dfrac{\sum_{i=1}^{n} |x_i - \overline{X}|}{n}$ สำหรับประชากร: $M.D. = \dfrac{\sum_{i=1}^{N} |X_i - \mu|}{N}$

2. **ส่วนเบี่ยงเบนมาตรฐาน**
   สำหรับตัวอย่าง ให้ S แทนสวนเบี่ยงเบนมาตรฐาน คำนวณได้จาก

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

   สำหรับประชากร ให้ σ แทนสวนเบี่ยงเบนมาตรฐาน คำนวณได้จาก

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

   กำลังสองของส่วนเบี่ยงเบนมาตรฐานเรียกว่า **ความแปรปรวน** (Variance) สำหรับตัวอย่างความแปรปรวนคือ S^2 สำหรับประชากรความแปรปรวนคือ σ^2

ตัวอย่าง 2.3.8 จากการสอบถามเวลา (นาที) ในการเดินทางมายังโรงเรียนของนักเรียน 10 คน

30 35 40 45 50 60 60 70 70 90

จงหาส่วนเบี่ยงเบนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน เมื่อ

1. ข้อมูลเป็นประชากร

   แนวคำตอบ จะได้ว่า

$$\mu = \frac{30+35+40+45+50+60+60+70+70+90}{10} = \frac{550}{10} = 55$$

ดังนั้น

$$M.D. = \frac{|30-55| + |35-55| + \cdots + |90-55|}{10} = \frac{150}{10} = 15$$

และ

$$\sigma = \sqrt{\frac{(30-55)^2 + (35-55)^2 + \cdots + (90-55)^2}{10}} = \sqrt{\frac{3100}{10}} = \sqrt{310} = 17.61$$

2. ข้อมูลเป็นตัวอย่าง

แนวคำตอบ จะเห็นว่า $\overline{X} = 55$ ดังนั้น

$$M.D. = \frac{|30-55| + |35-55| + \cdots + |90-55|}{10} = \frac{150}{10} = 15$$

และ

$$S = \sqrt{\frac{(30-55)^2 + (35-55)^2 + \cdots + (90-55)^2}{10-1}} = \sqrt{\frac{3100}{9}} = 18.56$$

ตัวอย่าง 2.3.9 ข้อมูลประชากรชุดหนึ่งมีการแจกแจงปกติ โดย

มัธยฐานเท่ากับ 12 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 8 และ $\sum_{i=1}^{N}(x_i - 10)^2 = 5440$

แล้ว N มีค่าเท่ากับเท่าใด

แนวคำตอบ เนื่องจากประชากรชุดมีการแจกแจงปกติ จะได้ว่า $\mu = \text{Med} = 12$ และ $\sigma = 8$ ฉะนั้น

$$\frac{\sum_{i=1}^{N}(x_i - 12)^2}{N} = 8^2 \qquad \text{นั่นคือ} \qquad \sum_{i=1}^{N}(x_i - 12)^2 = 64N$$

พิจารณา

$$\begin{aligned}
\sum_{i=1}^{N}[(x_i - 12) + 2]^2 &= \sum_{i=1}^{N}(x_i - 10)^2 \\
\sum_{i=1}^{N}[(x_i - 12)^2 + 4(x_i - 12) + 4] &= 5440 \\
\sum_{i=1}^{N}(x_i - 12)^2 + 4\sum_{i=1}^{N}(x_i - 12) + \sum_{i=1}^{N} 4 &= 5440 \\
64N + 4(0) + 4N &= 5440 \\
68N &= 5440 \\
N &= 80
\end{aligned}$$

ดังนั้นประชากรชุดนี้มีขนาด 80 หรือ $N = 80$

ตัวอย่าง 2.3.10 ข้อมูลประชากรเรียงจากน้อยไปมากดังนี้

$$1 \quad 2 \quad 2 \quad 3 \quad 5 \quad a \quad 9 \quad 12 \quad 13 \quad 15$$

เมื่อ a เป็นจำนวนจริง ถ้าส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 4.4 จงหา a

แนวคำตอบ เนื่องจากข้อมูลเรียงจากน้อยไปมากจะได้ว่า $5 \leq a \leq 9$ และ

$$\mu = \frac{1+2+2+3+5+a+9+12+13+15}{10} = \frac{62+a}{10} = 6.2 + \frac{a}{10}$$

จาก $5 \leq a \leq 9$ ฉะนั้น

$$\begin{aligned} 0.5 &\leq \frac{a}{10} \leq 0.9 \\ 6.2+0.5 &\leq 6.2+\frac{a}{10} \leq 6.2+0.9 \\ 6.7 &\leq \mu \leq 7.1 \end{aligned}$$

จาก $M.D. = 4.4$ และ $6.7 \leq \mu \leq 7.1$ จะได้ว่า

$$\begin{aligned} M.D. &= \frac{\sum |x_i - \mu|}{10} \\ 4.4(10) &= |1-\mu| + |2-\mu| + |2-\mu| + |3-\mu| + |5-\mu| + |a-\mu| \\ &\quad + |9-\mu| + |12-\mu| + |13-\mu| + |15-\mu| \\ 44 &= (\mu-1) + (\mu-2) + (\mu-2) + (\mu-3) + (\mu-5) + |a-\mu| \\ &\quad + (9-\mu) + (12-\mu) + (13-\mu) + (15-\mu) \\ 44 &= |a-\mu| + \mu + 36 \\ 8 &= |a-\mu| + \mu \end{aligned}$$

กรณี $\mu \geq a$ จะได้ว่า

$$\begin{aligned} 8 &= -(a-\mu) + \mu = -a + 2\mu = -a + 2\left(6.2 + \frac{a}{10}\right) = 12.4 - 0.9a \\ 0.9a &= 12.4 - 8 = 4.4 \\ a &= \frac{44}{9} = 4.89 < 5 \qquad \text{เป็นไปไม่ได้} \end{aligned}$$

กรณี $\mu > a$ จะได้ว่า

$$8 = (a-\mu) + \mu = a$$

ดังนั้น $a = 8$

ทฤษฎีบท 2.3.11 กำหนดให้ประชากรขนาด N คือ $X_1, X_2, X_3, ..., X_N$ และสุ่มตัวอย่างขนาด n จากประชากรดังกล่าวมาได้เป็น $x_1, x_2, x_3, ...x_n$ โดยที่ประชากรและตัวอย่างมีค่าเฉลี่ยเลขคณิตเท่ากับ μ และ $\overline{X}$ ตามลำดับ จะได้ว่า

$$\text{สำหรับตัวอย่าง } S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n(\overline{X})^2}{n-1} \qquad \text{และ} \qquad \text{สำหรับประชากร } \sigma^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu^2$$

บทพิสูจน์. สำหรับตัวอย่าง

$$\begin{aligned}
S^2 &= \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i^2 - 2\overline{X}x_i + \overline{X}^2)}{n-1} \\
&= \frac{\sum_{i=1}^{n} x_i^2 - 2\overline{X}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \overline{X}^2}{n-1} \\
&= \frac{\sum_{i=1}^{n} x_i^2 - 2\overline{X}(n\overline{X}) + n\overline{X}^2}{n-1} \\
&= \frac{\sum_{i=1}^{n} x_i^2 - n(\overline{X})^2}{n-1}
\end{aligned}$$

สำหรับประชากร

$$\begin{aligned}
\sigma^2 &= \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N} = \frac{\sum_{i=1}^{N}(X_i^2 - 2\mu X_i + \mu^2)}{N} \\
&= \frac{\sum_{i=1}^{N} X_i^2 - 2\mu\sum_{i=1}^{N} X_i + \sum_{i=1}^{N} \mu^2}{N} \\
&= \frac{\sum_{i=1}^{N} X_i^2 - 2\mu(n\mu) + N\mu^2}{N} \\
&= \frac{\sum_{i=1}^{N} X_i^2 - N\mu^2}{N} \\
&= \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu^2
\end{aligned}$$

$\square$

ตัวอย่าง 2.3.12 ถ้าความยาวรัศมีวงกลม 10 วงมีค่าเฉลี่ยเลขคณิตเท่ากับ 3 และมีความแปรปรวนเท่ากับ 5 แล้วผลรวมของพื้นที่วงกลมทั้ง 10 วงนี้มีค่าเท่ากับเท่าใด ถ้าข้อมูลดังกล่าวเป็นประชากร

แนวคำตอบ ให้ $r_1, r_2, ..., r_{10}$ เป็นรัศมีของวงกลมทั้ง 10 วง จะได้ว่า $\mu_r = 3$ และ $\sigma_r^2 = 5$ จะได้ว่า

$$\frac{\sum_{i=1}^{10} r_i^2}{10} - 3^2 = 5 \qquad \text{นั่นคือ} \qquad \sum_{i=1}^{10} r_i^2 = 10(5+9) = 140$$

ดังนั้นผลรวมของพื้นที่วงกลมทั้ง 10 วงเท่ากับ

$$\sum_{i=1}^{10} \pi r_i^2 = \pi \sum_{i=1}^{10} r_i^2 = 140\pi$$

ตัวอย่าง 2.3.13 ค่าเฉลี่ยเลขคณิตของคะแนนสอบของนักเรียนกลุ่มหนึ่งเท่ากับ 72 คะแนน ความแปรปรวน (ประชากร) เท่ากับ 600 ถ้ามีนักเรียนมาเพิ่มอีก 1 คน ซึ่งสอบได้ 60 คะแนน ทำให้ค่าเฉลี่ยเปลี่ยนไปเป็น 70 คะแนน ความแปรปรวนของข้อมูลชุดใหม่เท่ากับเท่าใด

แนวคำตอบ ให้ประชากรขนาด n ประกอบด้วย $x_1, x_2, ..., x_n$ ซึ่งมีค่าเฉลี่ยเลขคณิตเท่ากับ 72 ความแปรปรวนเท่ากับ 600 นั่นคือ

$$\sum_{i=1}^{n} x_i = 72n$$

และ

$$600 = \frac{\sum_{i=1}^{n} x_i^2}{n} - 72^2 \qquad \text{ฉะนั้น} \qquad \sum_{i=1}^{n} x_i^2 = (600 + 72^2)n = 5784n$$

พิจารณาข้อมูลชุดใหม่คือ $x_1, x_2, ..., x_n, 60$ จะได้ว่า

$$\begin{aligned} 70 &= \frac{\sum_{i=1}^{n} x_i + 60}{n+1} = \frac{72n+60}{n+1} \\ 70(n+1) &= 72n + 60 \\ 70n + 70 &= 72n + 60 \\ n &= 5 \end{aligned}$$

และ $\sum_{i=1}^{5} x_i^2 = 5784(5) = 28920$ ดังนั้นความแปรปรวนชุดใหม่เท่ากับ

$$\sigma^2 = \frac{\sum_{i=1}^{5} x_i^2 + 60^2}{6} - 70^2 = \frac{28920 + 3600}{6} - 4900 = 520$$

สำหรับข้อมูลจัดกลุ่ม ในรูปตารางแจกแจงความถี่ที่มีจำนวน k ชั้น แต่ละชั้นมีความถี่ f_i และ X_i เป็นจุดกึ่งกลางของชั้นที่ i เมื่อ $i = 1, 2, ..., k$ จะได้ว่า

สำหรับตัวอย่าง	สำหรับประชากร
$M.D. = \dfrac{\sum_{i=1}^{k} f_i\lvert X_i - \overline{X}\rvert}{n}$	$M.D. = \dfrac{\sum_{i=1}^{k} f_i\lvert X_i - \mu\rvert}{N}$
$S = \sqrt{\dfrac{\sum_{i=1}^{k} f_i(X_i - \overline{X})^2}{n-1}}$	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{k} f_i(X_i - \mu)^2}{N}}$

ตัวอย่าง 2.3.14 จงหาส่วนเบี่ยงเบนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานและความแปรปรวน ของคะแนนสอบกลางภาคของวิชาภาษาอังกฤษ ซึ่งแสดงดังตารางต่อไปนี้

คะแนน	จำนวนคน
$11-15$	4
$16-20$	8
$21-25$	16
$26-30$	8
$31-35$	4

แนวคำตอบ พิจารณาค่าจากตารางต่อไปนี้

คะแนน	ความถี่ (f_i)	X_i	f_iX_i	$f_i\lvert X_i-\mu\rvert$	$f_i(X_i-\mu)^2$
$11-15$	4	13	52	42	441
$16-20$	8	18	144	44	242
$21-25$	16	23	368	8	4
$26-30$	8	28	224	36	162
$31-35$	4	33	132	38	361
รวม	40		920	168	1210

จะได้ว่า $\mu = \dfrac{920}{40} = 23.5$ ฉะนั้น

$$M.D. = \frac{168}{40} = 4.2 \qquad \text{และ} \qquad \sigma = \sqrt{\frac{1210}{40}} = \sqrt{30.25} = 5.5$$

ดังนั้นส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 4.2 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.5 และความแปรปรวนเท่ากับ 30.25

ทฤษฎีบท 2.3.15 ข้อมูลชุด X คือ $x_1, x_2, , ..., x_n$ และข้อมูลชุด Y คือ $y_1, y_2, , ..., y_n$ ถ้า

$$y_i = ax_i + b \quad \text{เมื่อ} \quad i = 1, 2, ..., n \text{ และ } a, b \text{ เป็นจำนวนจริง}$$

1. สำหรับข้อมูลแบบตัวอย่าง
   ให้ S_X และ S_Y เป็นส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด X และ Y ตามลำดับ จะได้ว่า

$$S_Y = |a|S_X \qquad \text{หรือ} \qquad S_Y^2 = a^2 S_X^2$$

2. สำหรับข้อมูลแบบประชากร
   ให้ σ_X และ σ_Y เป็นส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด X และ Y ตามลำดับ จะได้ว่า

$$\sigma_Y = |a|\sigma_X \qquad \text{หรือ} \qquad \sigma_Y^2 = a^2 \sigma_X^2$$

บทพิสูจน์. สำหรับข้อมูลแบบตัวอย่าง ชุด X มีค่าเฉลี่ยเท่ากับ $\overline{X}$ โดยทฤษฎีบท 2.1.19 จะได้ว่า $\overline{Y} = a\overline{X} + b$ และ

$$\begin{aligned}
S_Y^2 &= \frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n-1} = \frac{\sum_{i=1}^{n}[(ax_i + b) - (a\overline{X} + b)]^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}(ax_i - a\overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n} a^2(x_i - \overline{X})^2}{n-1} \\
&= a^2 \cdot \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1} = a^2 S_X^2
\end{aligned}$$

พิสูจน์ในทำนองเดียวกันสำหรับข้อมูลแบบประชากร (เป็นแบบฝึกหัด) □

ตัวอย่าง 2.3.16 ในการทำธุรกิจเล็ก ๆ อย่างหนึ่งในแต่ละวัน ความสัมพันธ์ของต้นทุน X (บาท) และกำไร Y (บาท) เป็นแบบเชิงเส้นคือ

$$y = 1.25x - 500$$

ถ้าสุ่มตัวอย่างการลงทุนมา 5 วัน เป็นเงินลงทุนดังนี้ 500, 550, 650, 700, 1000 บาท จงหาความแปรปรวนของกำไรในการลงทุนครั้งนี้

แนวคำตอบ ค่าเฉลี่ยของการลงทุน 5 วันคือ

$$\overline{X} = \frac{500 + 550 + 650 + 700 + 1000}{5} = \frac{3400}{5} = 680$$

ความแปรปรวนของการลงทุนของตัวอย่าง 5 วันคือ

$$\begin{aligned}
S_X^2 &= \frac{(500-680)^2 + (550-680)^2 + (650-680)^2 + (700-680)^2 + (1000-680)^2}{5-1} \\
&= \frac{153000}{4} = 38250
\end{aligned}$$

โดยทฤษฎีบท 2.3.15 จะได้ว่า $S_Y^2 = 1.25^2 S_X^2 = 1.25^2(38250) = 59765.625$
ดังนั้นความแปรปรวนของกำไรในการลงทุน 5 วันนี้เท่ากับ 59765.625

ตัวอย่าง 2.3.17 ข้อมูลชุดหนึ่งมี 10 จำนวน คือ $x_1, x_2, x_3, ..., x_{10}$ ซึ่งข้อมูลชุดนี้มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.3 ถ้าข้อมูลชุดที่สองคือ

$$3x_1 + 174,\ 3x_2 + 174,\ 3x_3 + 174,\ ..., 3x_{10} + 174$$

แล้วส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดที่สองนี้จะเท่ากับเท่าใด

แนวคำตอบ เนื่องจากข้อมูลชุดที่สอง Y สัมพันธ์เชิงเส้นกับชุดที่หนึ่งคือ

$$y_i = 3x_i + 173 \quad \text{เมื่อ } i = 1, 2, ..., 10$$

จะเห็นว่าส่วนเบี่ยงเบนมาตรฐานชุดที่หนึ่งเท่ากับ 2.3 โดยทฤษฎีบท 2.3.15 ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดที่สองเท่ากับ $3(2.3) = 6.9$

ทฤษฎีบท 2.3.18 สำหรับข้อมูลแบบประชากร
ข้อมูลชุดที่ 1 มีจำนวน N_1 มีค่าเฉลี่ยเลขคณิตเท่ากับ μ_1 มีความแปรปรวนเท่ากับ σ_1^2
ข้อมูลชุดที่ 2 มีจำนวน N_2 มีค่าเฉลี่ยเลขคณิตเท่ากับ μ_2 มีความแปรปรวนเท่ากับ σ_2^2
ความแปรปรวนรวมของข้อมูลทั้งสองชุด เขียนแทนด้วย σ_{com}^2

1. ถ้า $\mu_1 = \mu_2$ แล้ว $\sigma_{com}^2 = \dfrac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2}$

2. ถ้า $\mu_1 \neq \mu_2$ แล้ว $\sigma_{com}^2 = \dfrac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2} + N_1N_2\left(\dfrac{\mu_2 - \mu_1}{N_1 + N_2}\right)^2$

บทพิสูจน์. ให้ข้อมูลชุดที่ 1 มีจำนวน N_1 ประกอบด้วย $x_1, x_2, ..., x_{N_1}$
ให้ข้อมูลชุดที่ 2 มีจำนวน N_2 ประกอบด้วย $y_1, y_2, ..., y_{N_2}$
ให้ μ แทนค่าเฉลี่ยรวมของทั้งสองประชากร จะได้ว่า

$$\mu = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2}$$

และ

$$\sigma_1^2 = \frac{\sum_{i=1}^{N_1} x_i^2}{N_1} - \mu_1^2 \quad \longrightarrow \quad \sum_{i=1}^{N_1} x_i^2 = (\sigma_1^2 + \mu_1^2)N_1$$

$$\sigma_2^2 = \frac{\sum_{i=1}^{N_2} y_i^2}{N_2} - \mu_2^2 \quad \longrightarrow \quad \sum_{i=1}^{N_2} y_i^2 = (\sigma_2^2 + \mu_2^2)N_2$$

จะได้ว่า

$$
\begin{aligned}
\sigma^2_{com} &= \frac{\sum_{i=1}^{N_1} x_i^2 + \sum_{i=1}^{N_2} y_i^2}{N_1+N_2} - \mu^2 \\
&= \frac{\sum_{i=1}^{N_1} x_i^2 + \sum_{i=1}^{N_2} y_i^2}{N_1+N_2} - \left(\frac{N_1\mu_1 + N_2\mu_2}{N_1+N_2}\right)^2 \\
&= \frac{(\sigma_1^2+\mu_1^2)N_1 + (\sigma_2^2+\mu_2^2)N_2}{N_1+N_2} - \left(\frac{N_1\mu_1 + N_2\mu_2}{N_1+N_2}\right)^2 \\
&= \frac{(N_1\sigma_1^2+N_2\sigma_2^2) + (N_1\mu_1^2+N_2\mu_2^2)}{N_1+N_2} - \left(\frac{N_1\mu_1 + N_2\mu_2}{N_1+N_2}\right)^2 \\
&= \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + \frac{N_1\mu_1^2+N_2\mu_2^2}{N_1+N_2} - \left(\frac{N_1\mu_1 + N_2\mu_2}{N_1+N_2}\right)^2 \\
&= \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + \frac{(N_1\mu_1^2+N_2\mu_2^2)(N_1+N_2)}{(N_1+N_2)^2} - \left(\frac{N_1\mu_1 + N_2\mu_2}{N_1+N_2}\right)^2 \\
&= \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + \frac{N_1^2\mu_1^2+N_1N_2\mu_2^2+N_1N_2\mu_1^2+N_2^2\mu_2^2}{(N_1+N_2)^2} - \frac{N_1^2\mu_1+2N_1N_2\mu_1\mu_2+N_2^2\mu_2}{(N_1+N_2)^2} \\
&= \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + \frac{N_1N_2\mu_2^2+N_1N_2\mu_1^2-2N_1N_2\mu_1\mu_2}{(N_1+N_2)^2} \\
&= \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + \frac{N_1N_2(\mu_2^2-2\mu_1\mu_2+\mu_2^2)}{(N_1+N_2)^2} \\
&= \frac{N_1\sigma_1^2+N_2\sigma_2^2}{N_1+N_2} + N_1N_2\left(\frac{\mu_2-\mu_1}{N_1+N_2}\right)^2
\end{aligned}
$$

ในกรณีที่ $\mu_1 = \mu_2$ จะได้ว่าข้อ 1 ถูกต้องด้วย □

ตัวอย่าง 2.3.19 ข้อมูลสรุปของระดับคะแนน (ประชากร) ของนักเรียน 3 ห้อง แสดงได้ดังนี้

ห้อง	จำนวนนักเรียน	ระดับคะแนนเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
ก	20	2.96	0.75
ข	30	3.25	0.50
ค	25	3.25	0.30

1. จงหาความแปรปรวนรวมของห้อง ข และ ค

   แนวคำตอบ เนื่องจากห้อง ข และ ค มีระดับคะแนนเฉลี่ยเท่ากัน ดังนั้นความแปรปรวนรวมเท่ากับ

   $$\frac{30(0.50)^2 + 25(0.30)^2}{30+25} = 0.1773$$

2. จงหาความแปรปรวนรวมของห้อง ก และ ข

   แนวคำตอบ เนื่องจากห้อง ข และ ค มีระดับคะแนนเฉลี่ยไม่เท่ากัน ดังนั้นความแปรปรวนรวมเท่ากับ

   $$\frac{20(0.75)^2 + 30(0.50)^2}{30+20} + 20(30)\left(\frac{3.25-2.96}{20+30}\right)^2 = 0.3952$$

2. การวัดกระจายสัมพัทธ์

การวัดการกระจายสัมพัทธ์ คือการหาค่าเพื่อเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่าหนึ่งชุด โดยใช้อัตราส่วนหรือเปอร์เซ็นต์ การเปรียบเทียบการกระจายของข้อมูลระหว่างชุดที่นิยมใช้มี 4 ชนิดคือ

1. **สัมประสิทธิ์ของพิสัย** (Coefficient of Range)

$$C.R = \frac{X_{\text{max}} - X_{\text{min}}}{X_{\text{max}} + X_{\text{min}}}$$

2. **สัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์** (Coefficient of Quartile Deviation)

$$C.Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

3. **สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย** (Coefficient of Mean Deviation)

$$C.M.D.\ \text{สำหรับตัวอย่าง} = \frac{M.D.}{\overline{X}} \quad \text{และ}\ C.M.D.\ \text{สำหรับประชากร} = \frac{M.D.}{\mu}$$

4. **สัมประสิทธิ์ของการแปรผัน** (Coefficient of Varian)

$$C.V.\ \text{สำหรับตัวอย่าง} = \frac{S}{\overline{X}} \qquad \text{และ} \qquad C.V.\ \text{สำหรับประชากร} = \frac{\sigma}{\mu}$$

ตัวอย่าง 2.3.20 จากการสำรวจสุ่มตัวอย่าง ระยะรอคอย (waiting period) หน่วยเป็นวัน ของประกันโควิด-19 จำนวน 5 บริษัท เป็นดังนี้ 7, 10, 14, 14, 20 จงหาสัมประสิทธิ์ทั้ง 4 ชนิด

แนวคำตอบ จากข้อมูลจะได้ว่า $\overline{X}_{\text{max}} = 20$, $\overline{X}_{\text{min}} = 7$

$$Q_1 = \frac{7+10}{2} = 8.5,\ \text{และ}\ \ Q_3 = \frac{14+20}{2} = 17\ \text{เนื่องจาก}\ \overline{X} = \frac{7+10+14+14+20}{5} = 13$$

จะได้ว่า

$$M.D. = \frac{|7-13| + |10-13| + |14-13| + |14-13| + |20+13|}{5} = 3.6$$

$$S = \sqrt{\frac{(7-13)^2 + (10-13)^2 + (14-13)^2 + (14-13)^2 + (20+13)^2}{5-1}} = 4.8990$$

ดังนั้น

$$C.R. = \frac{20-7}{20+7} = 0.4815 \qquad \text{หรือ} \quad 48.15\%$$

$$C.Q.D = \frac{17-8.5}{17+8.5} = 0.3333 \qquad \text{หรือ} \quad 33.33\%$$

$$C.M.D = \frac{3.6}{13} = 0.2769 \qquad \text{หรือ} \quad 27.69\%$$

$$C.V. = \frac{4.8990}{13} = 0.3768 \qquad \text{หรือ} \quad 37.68\%$$

ตัวอย่าง 2.3.21 ข้อมูลเกี่ยวกับน้ำหนักของเด็กแรกเกิดของโรงพยาบาลต่างๆดังนี้

โรงพยาบาล	จำนวนนักเรียน	น้ำหนักเฉลี่ย (กรัม)	ส่วนเบี่ยงเบนมาตรฐาน (กรัม)
A	100	2,400	300
B	100	2,500	500
C	200	3,000	300
D	150	2,400	200
E	250	2,800	700

จงตอบคำถามต่อไปนี้

ก. โรงพยาบาลใดมีการกระจายของน้ำหนักเด็กแรกเกิดน้อยที่สุด

ข. โรงพยาบาลใดมีการกระจายของน้ำหนักเด็กแรกเกิดมากที่สุด

แนวคำตอบ คำนวณค่าสัมประสิทธิ์ของการแปรผันของน้ำหนักแต่ละโรงพยาบาลได้ดังตารางต่อไปนี้

โรงพยาบาล	น้ำหนักเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ของการแปรผัน
A	2,400	300	$\frac{300}{2400} \times 100 = 12.50$%
B	2,500	500	$\frac{500}{2500} \times 100 = 20.00$%
C	3,000	300	$\frac{300}{3000} \times 100 = 10.00$%
D	2,400	200	$\frac{200}{2400} \times 100 = 8.33$%
E	2,800	700	$\frac{700}{2800} \times 100 = 25.00$%

จากตารางค่าสัมประสิทธิ์ของการแปรผันของน้ำหนักของโรงพยาบาล D มีค่าน้อยที่สุด และ E มีค่ามากสุด ดังนั้นโรงพยาบาลใดมีการกระจายของน้ำหนักเด็กแรกเกิดน้อยที่สุดคือโรงพยาบาล D และมากสุดคือโรงพยาบาล E

ตัวอย่าง 2.3.22 ในการสอบวิชาคณิตศาสตร์พบว่า คะแนนสอบของนักเรียนมีการแจกแจงปกติ ส่วนเบี่ยงเบนควอไทล์เท่ากับ 6 สัมประสิทธิ์ควอไทล์เท่ากับ 0.6 คะแนนเฉลี่ยของการสอบครั้งนี้เท่ากับเท่าใด

แนวคำตอบ เนื่องจากส่วนเบี่ยงเบนควอไทล์เท่ากับ 6 สัมประสิทธิ์ควอไทล์เท่ากับ 0.6 จะได้ว่า

$$\frac{Q_3 - Q_1}{2} = 6 \qquad \text{และ} \qquad \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0.6$$

นั่นคือ $Q_3 + Q_1 = \frac{Q_3 - Q_1}{0.6} = \frac{2(6)}{0.6} = 20$ เนื่องจากคะแนนสอบของนักเรียนมีการแจกแจงปกติ นั่นคือมีการแจงแจงสมมาตรดังนั้น

$$\overline{X} = \text{Med} = Q_2 = \frac{Q_1 + Q_3}{2} = \frac{20}{2} = 10$$

สรุปได้ว่าคะแนนเฉลี่ยของการสอบครั้งนี้เท่ากับ 10 คะแนน

แบบฝึกหัด 2.3

1. จงสร้างแผนภาพกล่องของคะแนนสอบวิชาสังคม 2 ห้อง (เขียนในแผนภาพเดียวกัน)
ห้อง ก : $11,\ 12,\ 12,\ 14,\ 15,\ 16,\ 20$
ห้อง ข : $11,\ 12,\ 13,\ 15,\ 15,\ 15,\ 16,\ 17,\ 18,\ 20,\ 20$

2. จงหาพิสัยและส่วนเบี่ยงเบนควอไทล์ของข้อมูลต่อไปนี้

คะแนน	จำนวนคน
1	5
2	8
4	10
6	8
10	4

3. จากการสำรวจน้ำหนักของนักเรียนกลุ่มหนึ่งจำนวน 4 คน มี 2 คน น้ำหนักเท่ากันและหนักน้อยกว่าอีก 2 คนที่เหลือ ถ้าฐานนิยม มัธยฐาน และพิสัยของนักเรียน 4 คนนี้คือ 45, 46 และ 6 กิโลกรัม ตามลำดับ แล้วความแปรปรวนของน้ำหนักนักเรียน 4 คนเท่ากับเท่าใด

4. นักเรียนชั้น ม.5 ของโรงเรียนแห่งหนึ่งจำนวน 2 ห้อง มีความแปรปรวนของคะแนนสอบทั้งสองนั้นห้องเท่ากับ 12 นักเรียนห้อง ก. มีจำนวน 20 คน คำนวณส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบได้เท่ากับ 3 คะแนน นักเรียนห้อง ข. จะมีจำนวนเท่ากับเท่าใดถ้าส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบห้องนี้เท่ากับ 4 คะแนน โดยทั้งสองห้องมีค่าเฉลี่ยเลขคณิตเท่ากัน

5. กำหนดให้ข้อมูลชุดหนึ่งซึ่งประกอบด้วย $x_1, x_2, .., x_{10}$ มีค่าเฉลี่ยเลขคณิต $\bar{x}$ และข้อมูลชุดที่สองประกอบไปด้วย $y_1, y_2, .., y_{20}$ มีค่าเฉลี่ยเลขคณิต $\bar{y}$ โดยที่

$$\sum_{i=1}^{10}(x_i-\bar{x})^2 = 160, \quad \sum_{i=1}^{20}(y_i-\bar{y})^2 = 110 \quad \text{และ} \quad \bar{x}=\bar{y}$$

ถ้านำข้อมูลทั้งสองชุดมารวมเป็นชุดเดียวกัน แล้วส่วนเบี่ยงเบนมาตรฐาน (ประชากร) ของข้อมูลชุดใหม่เท่ากับเท่าใด

6. ข้อมูลชุดหนึ่งเรียงจากน้อยไปหามากดังนี้ $a,\ 3,\ 5,\ 7,\ b$ ถ้าข้อมูลชุดนี้มีค่าเฉลี่ยเลขคณิตเท่ากับ 7 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ $2\sqrt{10}$ แล้วค่าของ $2a+b$ มีค่าเท่ากับเท่าใด

7. ข้อมูลชุดหนึ่งมี 11 จำนวนดังนี้

$$15 \quad 10 \quad 12 \quad 15 \quad 16 \quad x \quad 16 \quad 19 \quad 13 \quad 17 \quad 15$$

ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ 15 แล้วกำลังสองของส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับเท่าใด

8. มีข้อมูล 5 จำนวนซึ่งเรียงจากน้อยไปมากคือ $x_1, x_2, , x_3, x_4, x_5$ โดยมี $x_1 = 7$ ค่าเฉลี่ยเลขคณิตเท่ากับ μ และความแปรปรวนเท่ากับ 16 ถ้ากำหนดตารางแสดงค่า $x_i - \mu$ ดังนี้

i	$x_i - \mu$
1	$7-\mu$
2	-3
3	-1
4	3
5	6

แล้วค่าของ μ เท่ากับเท่าใด

9. ข้อมูลชุดหนึ่งมีค่าสังเกต (x) และร้อยละของความถี่สะสมสัมพัทธ์ แสดงดังตารางต่อไปนี้

ค่าสังเกต (x)	ร้อยละของความถี่สะสมสัมพัทธ์
1	20
2	40
a	70
6	90
10	100

เมื่อ a เป็นจำนวนจริง ถ้าข้อมูลชุดนี้มีค่าเฉลี่ยเลขคณิตเท่ากับ 4 แล้วความแปรปรวนของข้อมูลชุดนี้เท่ากับเท่าใด

10. โรงเรียนอนุบาลแห่งหนึ่งมีนักเรียน 80 คน โดยการแจกแจงอายุของนักเรียนเป็นดังตาราง

อายุ (ปี)	3.5	4	4.5	5	5.5	6
จำนวนนักเรียน (คน)	a	15	10	20	b	5

ถ้าค่าเฉลี่ยเลขคณิตของอายุนักเรียนมีค่า 4.5 ปี แล้วส่วนเบี่ยงเบนเลี่ยของอายุนักเรียนมีค่าเท่ากับเท่าใด

11. คะแนนสอบของวิชาคณิตศาสตร์ของนักเรียนจำนวน 30 คน มีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 25 และ 5 คะแนน ตามลำดับ ถ้านำคะแนนของนายสายชลและนางสาวฟ้าซึ่งสอบได้ 20 และ 30 คะแนน ตามลำดับ มารวมด้วย แล้วส่วนเบี่ยงเบนมาตรฐานจะเท่ากับเท่าใด

12. ข้อมูลชุดหนึ่งมี 5 จำนวน และมีมัธยฐานเท่ากับฐานนิยมซึ่งเท่ากับ 15 ค่าเฉลี่ยเลขคณิตเท่ากับ 16 ควอไทล์ที่ 1 เท่ากับ 14 และพิสัยเท่ากับ 7 แล้วความแปรปรวนของข้อมูลชุดนี้เท่ากับเท่าใด

13. มีนักเรียน 5 คน ร่วมกันบริจาคเงิน ได้เงินรวม 360 บาท ความแปรปรวน (ประชากร) เท่ากับ 660 ถ้ามีนักเรียนอีก 1 คนมาร่วมบริจาคเป็นเงิน 60 บาท ความแปรปรวนจะเพิ่มขึ้นหรือลดลงเท่ากับเท่าใด

14. ถ้าข้อมูลชุดหนึ่งมีสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 0.12 ส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 6 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 แล้วสัมประสิทธิ์ความแปรผันมีค่าเท่ากับเท่าใด

15. เมื่อสองปีก่อน นักเรียนห้องหนึ่งมี 30 คน แบ่งออกได้เป็นสองกลุ่ม กลุ่มที่หนึ่งมี 10 คน ทุกคนมีอายุ 10 ปี กลุ่มที่สองมี 20 คน มีอายุเฉลี่ย 8.5 ปี ถ้าความแปรปรวนของอายุนักเรียนกลุ่มที่สอง เท่ากับ 0 แล้ว ปัจจุบันความแปรปรวนของอายุนักเรียนห้องนี้ เท่ากับเท่าใด

16. ครอบครัวหนึ่งมีสมาชิก 6 คน มีอายุเฉลี่ย 34 ปี ส่วนเบี่ยงเบนมาตรฐานของอายุเท่ากับ 8 ปี อีก 6 ปีต่อมามีญาติสองคนมาขออยู่อาศัยด้วย โดยที่ญาติทั้งสองคนมีอายุเท่ากันเท่ากับค่าเฉลี่ยของคนทั้งคน 6 คนในครอบครัวนี้พอดี สัมประสิทธิ์การแปรผันของอายุของคนทั้ง 8 คนนี้เท่ากับเท่าใด

17. ถ้าคะแนนสอบคะแนนวิชาสถิติและวิชาคณิตศาสตร์ของนักเรียนกลุ่มหนึ่งมีดังนี้

นักเรียนคนที่	1	2	3	4	5
คะแนนสอบวิชาสถิติ	8	5	4	2	1
คะแนนสอบวิชาคณิตศาสตร์	9	6	5	3	2

แล้วอัตราส่วนสัมประสิทธิ์ความแปรผันระหว่างของคะแนนสอบวิชาสถิติและคณิตศาสตร์เท่ากับเท่าใด

18. กำหนดข้อมูล 2 ชุดดังนี้ ชุดที่หนึ่งคือ 5, 8, 6, 7, 9 ชุดที่สองคือ x_1, x_2, x_3, x_4, x_5 ถ้าส่วนสัมประสิทธิ์ความแปรผันของข้อมูลชุดที่หนึ่งเป็น 2 เท่าของข้อมูลชุดที่สอง และความแปรผันของข้อมูลชุดที่สองเท่ากับ 9 แล้วค่าเฉลี่ยเลขคณิตของข้อมูลชุดที่สองเท่ากับเท่าใด

19. พิจารณาข้อมูลชุดหนึ่งซึ่งเรียงลำดับจากน้อยไปมากดังต่อไปนี้ $8, a, 12, 17, 22, b, 26$ ถ้าค่าเฉลี่ยเลขคณิตเท่ากับ 17 และควอไทลที่ 1 เท่ากับ 10 แล้วสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยและสัมประสิทธิ์ของส่วนเบี่ยงเบนควอไทล์ตามลำดับ เท่ากับเท่าใด

สรุป

สำหรับการวิเคราะห์ข้อมูลเบื้องต้น อันดับแรกจะกล่าวถึงการวัดแน้วโน้มสู่ศูนย์กลาง ซึ่งเป็นสิ่งที่ใช้เป็นตัวแทนของข้อมูลเพื่อบอกลักษณะข้อมูลของกลุ่ม ๆ นั้น เรียกว่าค่ากลางของข้อมูล ซึ่งนิยมใช้ 3 ชนิดคือ ความเฉลี่ย มัธยฐาน และฐานนิยม จากนั้นกล่าวถึงการวัดตำแหน่งของข้อมูล โดยเรียงข้อมูลจากน้อยไปมาก หรือมากไน้อย เมื่อแบ่งออกเป็น 4, 10 และ 100 ส่วนเท่า ๆ กันเรียกแต่ละส่วนว่า ควอไทล์ เดไซน์ และเปอร์เซ็นต์ไทล์ ตามลำดับ ถัดไปนำควอไทล์ 4 ค่า และค่าต่ำสุดและสูงสุดไปสร้างแผนภาพกล่อง เพื่อดูลักษณะการกระจายของข้อมูล สุดท้ายจะกล่าวถึงการวัดการกระจายสัมบูรณ์เพื่อใช้เป็นค่าแสดงการกระจายของข้อมูลกลุ่ม ๆ หนึ่ง มี 4 แบบ คือพิสัย ส่วนเบี่ยงเบนควอไทล์ ส่วนเบี่ยนเบนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน และสำหรับการวัดการกระจายที่สามารถเทียบกับข้อมูลกลุ่มอื่น ๆ ได้ เรียกว่าการวัดการกระจายสัมพัทธ์ ประกอบด้วย สัมประสิทธ์พิสัย สัมประสิทธิ์ส่วนเบี่ยงเบนควอไทล์ สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ย และสัมประสิทธิ์ความแปรผัน

แบบฝึกหัดบทที่ 2

1. ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, x_3, ..., x_{50}$ ซึ่ง

$$\sum_{i=1}^{50}(x_i - 99) = 250$$

จงหาค่าเฉลี่ยของข้อมูลชุดนี้

2. จงหาค่าเฉลี่ยเลขคณิตและมัธยฐานของข้อมูลชุดนี้ 1, 2, 2, 3, 3, 3, 4, 4, 4, 4,...,10

3. ส่วนสูงของพี่น้อง 2 คน มีพิสัยเท่ากับ 12 เซนติเมตร มีค่าเฉลี่ยเลขคณิตเท่ากับ 171 เซนติเมตร จงหาส่วนสูงของพี่และน้อง

4. ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, x_3, ..., x_{10}$ ซึ่ง

$$\sum_{i=1}^{10} x_i = 30 \qquad \text{และ} \qquad \sum_{i=1}^{10}(x_i - 4)^2 = 18$$

จงหาค่าของ $\sum_{i=1}^{10}(x_i - \mu)^2$

5. ข้อมูลชุดหนึ่งประกอบด้วย $x_1, x_2, x_3, ..., x_n$ มีค่าเฉลี่ยเลขคณิตเท่ากับ μ ซึ่ง $\sum_{i=1}^{n}(x_i - A)^2 = K$ จงพิสูจน์ว่า

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

6. ค่าเฉลี่ยเลขคณิตของน้ำหนักของพนักงานบริษัทแห่งหนึ่งเท่ากับ 48.01 กิโลกรัม บริษัทนี้มีพนักงานชาย 43 คนและพนักงานหญิง 57 คน ถ้าค่าเฉลี่ยเลขคณิตของน้ำหนักพนักงานหญิงเท่ากับ 45 กิโลกรัม แล้วน้ำหนักเฉลี่ยของพนักงานชายทั้งหมดรวมกันเท่ากับเท่าใด

7. เจ้าของคอกหมูเลี้ยงหมูไว้ 10 ตัว เมื่อนำไปขายปรากฏข้อมูลดังนี้

น้ำหนัก (กิโลกรัม)	29	33	37	42
จำนวนหมู (ตัว)	1	4	3	2

น้ำหนักเฉลี่ยของหมูในคอกนี้ตัวละกี่กิโลกรัม

8. ให้ข้อมูลทั้งหมด m ชุดแต่ละชุดมีขนาดเท่ากัน และค่าเฉลี่ยเลขคณิต $\overline{X}_1, \overline{X}_2, ..., \overline{X}_m$ ตามลำดับ จงแสดงว่าค่าเฉลี่ยเลขคณิตรวมของข้อมมูลทั้งหมดเท่ากับ

$$\overline{X}_{\text{com}} = \frac{\sum_{i=1}^{m} \overline{X}_i}{m}$$

9. กำหนดข้อมูลชุดหนึ่ง $10,\ 3,\ x,\ 6,\ 6$ ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้มีค่าเท่ากับมัธยฐาน จงหาค่า x

10. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียน 50 มีการแจกแจงความถี่ดังนี้

ช่วงคะแนน	จำนวนนักเรียน (คน)
$1 - 20$	3
$21 - 40$	5
$41 - 60$	13
$61 - 80$	20
$81 - 100$	19

ค่าเฉลี่ยเลขคณิตของคะแนนสอบนี้เท่ากับเท่าใด

11. จากตารางแจกแจงความถี่ค่าเฉลี่ยเลขคณิตเท่ากับ 94.5 อันตรภาคชั้นที่มีความถี่สะสม P มีความถี่เท่ากับเท่าใด

คะแนน	ความถี่สะสม
$100 - 104$	20
$95 - 99$	35
$90 - 94$	45
$85 - 89$	53
$80 - 84$	P
$75 - 79$	60

12. ความสัมพันธ์ระหว่างกำไร (y) และราคาทุน (x) ของสินค้าในร้านแห่งหนึ่งเป็นไปตามสมการ $y = 2x - 30$ ถ้าราคาทุนสินค้า 5 ชนิดคือ 31, 34, 35, 36 และ 39 บาท แล้วค่าเฉลี่ยเลขคณิตของกำไรในการขายสินค้า 5 ชนิดนี้ เท่ากับเท่าใด

13. ชายคนหนึ่งตักปลาที่เลี้ยงไว้ในกระชังเพื่อส่งขายจำนวน 500 ตัว ซึ่งมีน้ำหนักโดยเฉลี่ยตัวละ 700 กรัม ในจำนวนนี้เป็นปลาจากกระชังที่หนึ่ง 300 ตัว และจากกระชังที่สอง 200 ตัว ถ้าปลาในกระชังที่หนึ่งมีน้ำหนักเฉลี่ยต่อตัวมากกว่าในกระชังที่สอง 50 กรัม แล้วเขาตักปลาจากกระชังที่สองมากี่กิโลกรัม

14. อายุเฉลี่ยของคนกลุ่มหนึ่งเท่ากับ 31 ปี ถ้าอายุเฉลี่ยของผู้หญิงในกลุ่มนี้เท่ากับ 35 ปี และอายุเฉลี่ยของผู้ชายกลุ่มนี้เท่ากับ 25 ปี อัตราส่วนระหว่างจำนวนผู้หญิงต่อจำนวนผู้ชายในกลุ่มนี้เท่ากับเท่าใด

15. ข้อมูลชุดหนึ่ง 5 จำนวน ค่าเฉลี่ยเลขคณิตเป็น 30.8 มัธยฐานเป็น 27 ฐานนิยมเป็น 25 และค่าเฉลี่ยเรขาคณิตเป็น 30 จะได้พิสัยของข้อมูลเป็นเท่าใด

16. เงินเดือนของพนักงาน 50 คนของบริษัทแห่งหนึ่งมีการแจกแจงความถี่ดังนี้

เงินดือน (บาท)	จำนวนพนักงาน (คน)
$10,000 - 19,999$	5
$20,000 - 29,999$	10
$30,000 - 49,999$	25
$50,000 - 59,999$	10

จงหาค่ามัธฐานและฐานนิยมของข้อมูลชุดนี้

17. ข้อมูลชุดหนึ่งมี 6 จำนวนคือ 2, 3, 6, 11, a, b ถ้าค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ 8 และมัธยฐานเท่ากับ 7 แล้ว $|a - b|$ มีค่าเท่าใด

18. สร้างตารางแจกแจงความถี่ของคะแนนการสอบของนักเรียนกลุ่มหนึ่ง โดยให้ความกว้างแต่ละอันตรภาคชั้นเป็น 10 แล้วปรากฏว่ามัธยฐานของคะแนนสอบเท่ากับ 57 คะแนนซึ่งอยู่ในช่วง $50 - 59$ ถ้านักเรียนที่สอบได้คะแนนต่ำกว่า 49.5 คะแนนอยู่จำนวน 12 คน และมีนักเรียนได้คะแนนต่ำหว่า 59.5 คะแนนอยู่จำนวน 20 คน จงหาว่านักเรียนกลุ่มนี้มีจำนวนกี่คน

19. ข้อมูลชุดหนึ่งถ้าเรียงจากน้อยไปมากแล้วได้เป็นลำดับเลขคณิตต่อไปนี้

$$2,\ 5,\ \ 8, ..., 92$$

ควอไทล์ที่ 3 ของข้อมูลนี้มีค่าเท่ากับเท่าใด

20. ข้อมูลชุดหนึ่งเรียงจากน้อยไปมากดังนี้

$$5,\quad 10,\quad 12,\quad 20,\quad x,\quad 26,\quad 30,\quad 42,\quad 47,\quad y$$

ถ้าข้อมูลชุดนี้มีพิสัยเท่ากับ 45 และค่าเฉลี่ยเลขคณิตเท่ากับ 26.4 แล้วควอไทล์ที่สองของข้อมูลชุดนี้เท่ากับเท่าใด

21. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งแสดงด้วยแผนภาพต้น-ใบ ดังนี้

3	0 4 9
4	0 7 7 8 8 8
5	0 0 1 2 2 3 4 6 6 7 7 8 8 9
6	0 2 3 3 6 8 9
7	0 1

เปอร์เซ็นไทล์ที่ 50 ของคะแนนสอบนี้เท่ากับเท่าใด

22. คะแนนสอบวิชาวิทยาศาสตร์ของนักเรียนห้องหนึ่งจำนวน 119 คนดังนี้

คะแนนที่ได้	52	55	57	60	62	65	70	75	78	80	82
จำนวนนักเรียน (คน)	13	12	17	9	10	6	14	14	7	10	7

คะแนนที่เปอร์เซ็นไทล์ที่ 56 เท่ากับเท่าใด

23. จากตารางแจกแจงความถี่ต่อไปนี้ ถ้า P_{30} เท่ากับ 15 แล้วมีข้อมูลจำนวนกี่ค่าตั้งแต่ $26 - 39$

อันตรภาคชั้น	ความถี่
$5 - 11$	10
$12 - 18$	x
$19 - 25$	6
$26 - 32$	11
$33 - 39$	y
$40 - 46$	7
รวม	50

24. คะแนนสอบของนักเรียนกลุ่มหนึ่งมีการแจกแจงดังนี้

คะแนน	ความถี่
$21 - 30$	2
$31 - 40$	5
$41 - 50$	8
$51 - 60$	24
$61 - 70$	6
$71 - 80$	9
$81 - 90$	6

สุมิตรและนารีรัตน์เป็นนักเรียนในกลุ่มนี้ สุมิตรได้คะแนนในตำแหน่งควอไทล์ที่ 3 และนารีรัตน์ได้คะแนนในตำแหน่งเดไซล์ที่ 9 ถ้าคะแนนเต็ม 100 นารีรัตน์จะได้คะแนนสอบมากกว่าสุมิตรกี่เปอร์เซนต์

25. กำหนดตารางแสดงเงินค่าอาหารกลางวันที่นักเรียนห้องหนึ่งได้รับจากผู้ปกครองดังนี้

ค่าอาหารกลางวัน (บาท)	จำนวนนักเรียน (คน)
$50-54$	1
$55-59$	4
$60-64$	5
$65-69$	5
$70-74$	5

ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน และส่วนเบี่ยงเบนควอไทล์ตามลำดับ มีค่าเท่ากับเท่าใด

26. ข้อมูลชุดหนึ่งประกอบไปด้วย $x_1, x_2, , ..., x_N$ จงพิสูจน์ว่า

$$\sigma = 0 \qquad \text{ก็ต่อเมื่อ} \qquad x_1 = x_2 = ... = x_N$$

เมื่อ σ เป็นส่วนเบี่ยงมาตรฐานของข้อมูลชุดนี้

27. จงหาส่วนเบี่ยงเบนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวน ของข้อมูลต่อไปนี้
ข้อมูลตัวอย่าง : 10, 14, 7, 15, 16, 13, 16
ข้อมูลประชากร : 10, 14, 7, 15, 16, 13, 16

28. จงหาพิสัย ส่วนเบี่ยงเบนควอไทล์ ส่วนเบี่ยงเบนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ของประชากร

คะแนน	ความถี่
$11-20$	4
$21-30$	10
$31-40$	11
$41-50$	5
$51-60$	9
$61-70$	6

29. พิจารณาถูกผิดของข้อความต่อไปนี้ พร้อมให้เหตุผล

29.1 ในการสอบของนักเรียน 3 คน พบว่าค่าเฉลี่ยเลขคณิตของคะแนนสอบเท่ากับ 80 คะแนน ค่ามัธยฐานเท่ากับ 75 และพิสัยเท่ากับ 25 คะแนน คะแนนของนักเรียนที่ได้คะแนนต่ำสุดเท่ากับ 70 คะแนน

29.2 ข้อมูลชุดหนึ่งมี 5 จำนวนคือ x_1, x_2, x_3, x_4, x_5 และข้อมูลชุดที่สองมี 4 จำนวนคือ x_1, x_2, x_3, x_4 โดยที่ค่าเฉลี่ยเลขคณิตของข้อมูลทั้งสองชุดเท่ากัน ถ้า a และ b เป็นส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดที่หนึ่งและสองตามลำดับ แล้ว $\dfrac{b}{a} = \dfrac{\sqrt{5}}{2}$

30. ตารางต่อไปนี้เป็นคะแนนสอบของวิชาหนึ่งของนักเรียน 40 คน

คะแนน	จำนวนนักเรียน (f_i)
$10-14$	4
$15-19$	6
$20-24$	a
$25-29$	8
$30-34$	4
$35-39$	6

โดยมีคะแนนเฉลี่ย (μ) เท่ากับ 24.5 ถ้าส่วนเบี่ยงเบนควอไทล์เท่ากับ b และส่วนเบี่ยงเบนเฉลี่ยเท่ากับ c จงหา $a+b+c$

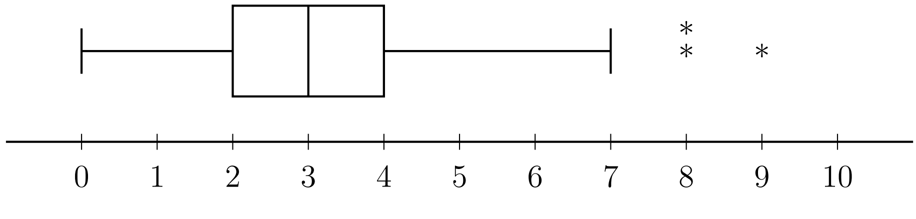
31. ถ้าข้อมูลชุดหนึ่งซึ่งประกอบด้วย $x_1,\ x_2,\ ..,x_n$ มีค่าเฉลี่ยเลขคณิตเท่ากับ 12 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 เมื่อบวกข้อมูลแต่ละจำนวนด้วย 5 ข้อมูลชุดใหม่จะมีค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานเท่ากับเท่าใด

32. กำหนดให้ $x_1, x_2, x_3, ..., x_n$ เป็นข้อมูลชุดหนึ่ง ซึ่งมีค่าเฉลี่ยเลขคณิตเท่ากับ 6 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2 $y_1, y_2, y_3, ..., y_n$ เป็นข้อมูลชุดที่สอง โดยที่

$$y_i = ax_i + b \quad \text{เมื่อ } i = 1, 2, 3, ..., n \text{ และ } a, b \in \mathbb{R} \text{ ซึ่ง } a > 0$$

ถ้านำข้อมูลทั้งสองชุดมารวมกัน $x_1, x_2, ..., x_n, y_1, y_2, ..., y_n$ พบว่าค่าเฉลี่ยเลขคณิตเท่ากับ 7 และความแปรปรวนเท่ากับ 21 แล้ว $a^2 + b^2$ เท่ากับเท่าใด

33. ในการสำรวจจำนวนชั่วโมงของนักเรียนที่เรียนเสริมแบบออนไลน์ จำนวน 50 ตัวอย่าง แสดงแผนภาพกล่องได้ดังนี้



33.1 จงหาวลาที่ใช้เรียนน้อยสุดและมากสุดในการสำรวจครั้งนี้

33.2 จงหาพิสัยควอไทล์

33.3 จงหาค่าผิดปกติของแผนภาพกล่อง

33.4 จำนวนชั่วโมงสูงสุดของ 25% ต่ำสุดของที่เรียนออนไลน์เท่ากับเท่าใด

34. นำข้อมูล 3 จำนวนที่แตกต่างกันมารวมกันเท่ากับ 195 ถ้าข้อมูลชุดนี้มีค่ามัธยฐานและสัมประสิทธิ์ของพิสัยเท่ากับ 60 และ 0.2 ตามลำดับ แล้วความแปรปรวนของข้อมูลชุดนี้เท่ากับเท่าใด

35. ค่าเฉลี่ยเลขคณิตของคะแนนสอบของนักเรียนกลุ่มหนึ่งเท่ากับ 72 คะแนน ความแปรปรวน (ประชากร) เท่ากับ 600 ถ้ามีนักเรียนมาเพิ่มอีก 1 คน ซึ่งสอบได้ 60 คะแนน ทำให้ค่าเฉลี่ยเปลี่ยนไปเป็น 70 คะแนน ความแปรปรวนของข้อมูลชุดใหม่เท่ากับเท่าใด

36. ข้อมูลชุดหนึ่งมี 5 จำนวน มีฐานนิยม มัธยฐาน และค่าเฉลี่ยเลขคณิตเท่ากับ 15, 16 และ 17 ตามลำดับ ถ้าพิสัยของข้อมูลชุดนี้เท่ากับ 7 จงหาความแปรปรวนของข้อมูลชุดนี้

37. ข้อมูล 7 คนมีค่าแตกต่างกันดังนี้

$$9,\ 6,\ 15,\ a,\ 2,\ 4,\ 12 \quad \text{โดยที่} \quad 2 < a < 12$$

ถ้าข้อมูลชุดนี้มีค่าเฉลี่ยเลขคณิตเป็นสองเท่าของส่วนเบี่ยงเบนควอไทล์ จงหาค่าของ a ทั้งหมดที่เป็นไปได้

38. ในการสอบชิงทุนการศึกษาที่มีผู้เข้าสอบ 1,000 คน ซึ่งในจำนวนนี้เป็นนักเรียนชายอยู่ 600 คน ปรากฏว่าได้ค่าเฉลี่ยเลขคณิตของคะแนนสอบนักเรียนชายและหญิงเท่ากันคือ 50 คะแนน แต่ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบของนักเรียนชายและนักเรียนหญิงเท่ากับ 1.5 และ 1 คะแนนตามลำดับ จงหาสัมประสิทธิ์ความแปรผันของคะแนนสอบนักเรียนทั้งหมด

บทที่ 3

ความน่าจะเป็น

ในชีวิตประจำวันเราใช้คำว่าโอกาสในการคาดเดาเหตุการณ์ที่จะเกิดขึ้นในอนาคต ซึ่งอาศัยความรู้สึกหรือประสบการณ์ที่ผ่านมาในการคาดเดาสิ่งเหล่านั้น นั่นแสดงว่าโอกาสที่จะเกิดขึ้นของเหตุการณ์เดียวกันของแต่ละคนอาจไม่เหมือนกัน แต่ในทางคณิตศาสตร์ไม่เป็นเช่นนั้นเพราะโอกาสหรือความน่าจะเป็นจะถูกนิยามอย่างแจ่มชัด ซึ่งจะกล่าวในบทนี้ โดยเริ่มจากปริภูมิตัวอย่าง การนับจุดตัวอย่าง ความน่าจะเป็น ความน่าจะเป็นแบบมีเงื่อนไข และกฎของเบย์

3.1 ปริภูมิตัวอย่าง

ในการโยนเหรียญ 1 อัน ผลลัพธ์ที่เกิดขึ้น (outcome) จาก การทดลอง (experiment) ดังกล่าว คือ หัวหรือก้อย ถ้าทำการทดลองเช่นนี้หลาย ๆ ครั้ง ผลลัพธ์ที่จดบันทึกจะเรียกว่าข้อมูลดิบ จะเห็นว่าการทดลองเช่นนี้ย่อมต้องรู้ขอบเขตของผลลัพธ์ที่เกิดขึ้นเสมอ เราจะเรียกว่าปริภูมิตัวอย่าง ดังนิยามต่อไปนี้

บทนิยาม 3.1.1 เซตที่มีสมาชิกเป็นผลลัพธ์ที่เป็นไปได้ทั้งหมดของการทดลองอย่างหนึ่งเรียกว่า **ปริภูมิตัวอย่าง** (sample space) เขียนแทนด้วย S และสมาชิกที่อยู่ใน S เรียกว่า **จุดตัวอย่าง** (sample point)

ตัวอย่าง 3.1.2 จงเขียนปริภูมิตัวอย่างต่อไปนี้

1. ในการทอดลูกเต๋า 1 ลูก ที่มีแต้ม 1, 2, 3, 4, 5, 6

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

2. ในการโยนเหรียญ 2 อันพร้อมกัน

   แนวคำตอบ ให้ H แทนเหรียญที่ขึ้นหัว และ T แทนเหรียญที่ขึ้นก้อย
   ดังนั้น $S = \{HH, HT, TH, TT\}$

ในการทดลองอย่างหนึ่ง เช่น ทอดลูกเต๋า 1 ลูกที่แต้ม 1, 2, 3, 4, 5, 6 เราอาจสนใจเหตุการณ์ขึ้นแต้มคู่ หรือขึ้นแต้มคี่ เป็นต้น นั่นแสดงว่าเหตุการณ์คือส่วนหนึ่งของปริภูมิตัวอย่าง

บทนิยาม 3.1.3 เหตุการณ์ (event) คือเซตย่อยของปริภูมิตัวอย่าง

ตัวอย่าง 3.1.4 ในการทอดลูกเต๋า 2 ลูก ที่มีแต้ม 1, 2, 3, 4, 5, 6 จงแจกแจงสมาชิกของ

1. ปริภูมิตัวอย่าง

   แนวคำตอบ ให้ (x, y) แทนแต้มของลูกเต๋าทั้งสอง ดังนั้น

$$\begin{aligned} S &= \{(x,y) : x, y \in \{1,2,3,4,5,6\}\} \\ &= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ &\quad (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ &\quad (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \end{aligned}$$

2. เหตุการณ์ที่ผลรวมของแต้มมากกว่า 10

   แนวคำตอบ ให้ E แทนเหตุการณ์ที่ผลรวมของแต้มมากกว่า 10 ดังนั้น $E = \{(5,6),(6,5),(6,6)\}$

ตัวอย่าง 3.1.5 ในการตรวจสอบคุณภาพสินค้าของโรงงานแห่งหนึ่ง ได้หยิบสินค้ามา 3 ชิ้น โดยวิธีสุ่ม (random) แล้วตรวจสอบทีละชิ้นว่าดีหรือชำรุด ให้สินค้าที่ตรวจแล้วมีสภาพดีแทนด้วย N และสินค้าที่ตรวจแล้วมีสภาพชำรุดแทนด้วย D จงแจกแจงสมาชิกของ

1. ปริภูมิตัวอย่าง

   แนวคำตอบ $S = \{DDD, DDN, DND, NDD, DNN, NDN, NND, NNN\}$

2. เหตุการณ์ที่พบจำนวนชำรุดมากกว่า 1 ชิ้น

   แนวคำตอบ $\{DDD, DDN, DND, NDD\}$

3. เหตุการณ์ไม่พบสินค้าชำรุดเลย

   แนวคำตอบ $\{NNN\}$

ตัวอย่าง 3.1.6 ในการตรวจสอบอายุการใช้งานของหลอดไฟ เมื่อให้ t แทนอายุการใช้งาน จงเขียนเซตต่อไปนี้

1. ปริภูมิตัวอย่าง

   แนวคำตอบ $S = \{t \in \mathbb{R} : t \geq 0\}$

2. เหตุการณ์ที่หลอดไฟชำรุดก่อนครบปีที่ 5

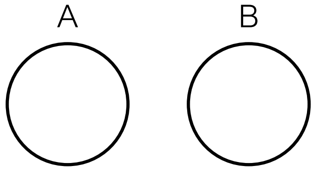
   แนวคำตอบ $E = \{t \in \mathbb{R} : 0 \leq t < 5\}$

3. เหตุการณ์ที่หลอดไฟชำรุดหลังปีที่ 10

   แนวคำตอบ $E = \{t \in \mathbb{R} : t > 10\}$

บทนิยาม 3.1.7 จะกล่าวว่าเหตุการณ์ A และ B **เป็นเหตุการณ์ที่ไม่เกิดร่วมกัน** (mutually exclusive) ถ้า $A \cap B = \varnothing$

รูปที่ 3.1: แผนภาพแสดงเหตุการณ์ที่ไม่เกิดร่วมกัน



3.2 การนับจุดตัวอย่าง

นักศึกษาคนหนึ่งต้องการจะเดินทางหอพักซึ่งอยู่ตามแนวรถไฟฟ้าไปห้างสรรพสินค้า สยามพารากอน (Paragon) หรือ เซ็นทรัลลาดพร้าว (CTL) จะเดินทางโดย รถประจำทาง (Bus) หรือ รถแท็กซี่ (Taxi) หรือ รถไฟฟ้า (Electric trian: ET) สมมติตลอดการเดินทางไปยังจุดหมายใช้วิธีเดียวกัน
ให้ E_1 คือเหตุการณ์ในการเลือกวิธีการเดินทาง นั่นคือ

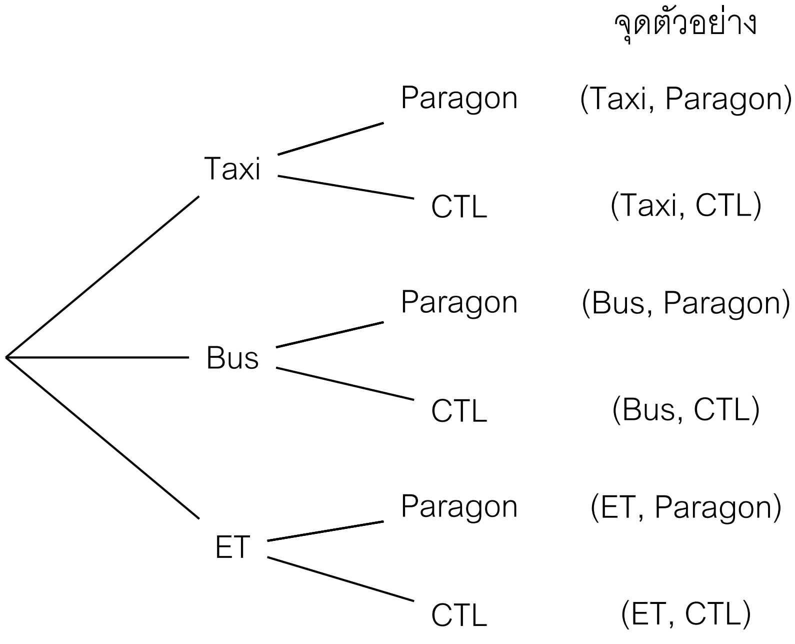
$$E_1 = \{\text{ Taxi , Bus , ET }\}$$

และ E_2 คือเหตุการณ์ในการเลือกจุดหมาย นั่นคือ

$$E_2 = \{\text{ Paragon , CTL }\}$$

เป็นเหตุการณ์ย่อยที่นักศึกษาจะเดินไปยังจุดหมาย จะแสดงวิธีเดินทางไปยังจุดหมายได้ดังนี้

รูปที่ 3.2: แผนภาพแสดงวิธีการเดินทางไปยังจุดหมาย



แผนภาพดังกล่าวเรียกว่า **แผนภาพต้นไม้ (tree diagram)** จะเห็นเหตุการณ์ E_1 เกิดได้ 3 วิธี และเหตุการณ์ E_2 เกิดได้ 2 วิธี จำนวนจุดตัวอย่างดังกล่าวเท่ากับ $3 \times 2 = 6$ วิธี ซึ่งสรุปเป็นหลักการได้ดังนี้

ทฤษฎีบท 3.2.1 หลักการคูณ (The Multiplication Principle)
ถ้าสามารถแยกเหตุการณ์ E เป็น k เหตุการณ์ย่อย $E_1, E_2, ..., E_k$ มีลำดับโดย

เหตุการณ์ E_1 เกิดได้ n_1 วิธี
เหตุการณ์ E_2 เกิดได้ n_2 วิธี
$\vdots$
เหตุการณ์ E_k เกิดได้ n_k วิธี

แล้วจำนวนวิธีที่จะเกิดเหตุการณ์ E คือ

$$n_1 n_2 \cdots n_k$$

ตัวอย่าง 3.2.2 ในการทอดลูกเต๋า 2 ลูกพร้อมกัน จงหาจำนวนจุดตัวอย่างในปริภูมิตัวอย่าง
แนวคำตอบ ให้ E_1 คือเหตุการณ์ในการขึ้นแต้มลูกเต๋าลูกที่หนึ่ง นั่นคือ $E_1 = \{1, 2, 3, 4, 5, 6\}$
E_2 คือเหตุการณ์ในการขึ้นแต้มลูกเต๋าลูกที่สอง นั่นคือ $E_2 = \{1, 2, 3, 4, 5, 6\}$
ดังนั้นจำนวนจุดตัวอย่างในปริภูมิตัวอย่างเท่ากับ $6 \times 6 = 36$ วิธี

ตัวอย่าง 3.2.3 ในการทอดลูกเต๋า 1 ลูก และโยนเหรียญ 1 อัน จงหา

1. จำนวนสมาชิกของปริภูมิตัวอย่าง

   แนวคำตอบ ให้ E_1 คือเหตุการณ์ในการขึ้นแต้มลูกเต๋า นั่นคือ $E_1 = \{1, 2, 3, 4, 5, 6\}$
   E_2 คือเหตุการณ์ในการขึ้นหน้าเหรียญ นั่นคือ $E_2 = \{H, T\}$
   ดังนั้นจำนวนจุดตัวอย่างในปริภูมิตัวอย่างเท่ากับ $6 \times 2 = 12$ วิธี

2. จงหาจำนวนวิธีที่ลูกเต๋าขึ้นแต้มคู่

   แนวคำตอบ ให้ E_1 คือเหตุการณ์ของลูกเต๋าขึ้นแต้มคู่ นั่นคือ $E_1 = \{2, 4, 6\}$
   E_2 คือเหตุการณ์ในการขึ้นหน้าเหรียญ นั่นคือ $E_2 = \{H, T\}$
   ดังนั้นจำนวนวิธีที่ลูกเต๋าขึ้นแต้มคู่เท่ากับ $3 \times 2 = 6$

ตัวอย่าง 3.2.4 จำนวนเต็มสี่หลักโดยแต่ละหลักไม่ซ้ำกัน ที่สร้างจาก 1, 2, 3, 5, 8 มีทั้งหมดกี่จำนวน
แนวคำตอบ ให้ E_1 คือเหตุการณ์ของหลักหน่วย เกิดได้ 5 วิธี
E_2 คือเหตุการณ์ของหลักสิบ เกิดได้ 4 วิธี
E_3 คือเหตุการณ์ของหลักร้อย เกิดได้ 3 วิธี
E_4 คือเหตุการณ์ของหลักพัน เกิดได้ 2 วิธี
ดังนั้นจำนวนเต็มสี่หลักโดยแต่ละหลักไม่ซ้ำกันเท่ากับ $5 \times 4 \times 3 \times 2 = 120$

ตัวอย่าง 3.2.5 นาย ก ตั้งใจจะดูซีรีส์จากแอพพลิเคชั่นหนึ่ง โดยมีซีรีส์เรื่องที่สนใจจะดูจำนวน 10 เรื่อง จงหาจำนวนวิธีที่นาย ก ดูซีรีส์อย่างน้อยหนึ่งเรื่อง
แนวคำตอบ ให้ E_i แทนเหตุการณ์ที่จะดูหรือไม่ดูซีรีส์เรื่องที่ i เมื่อ $i = 1, 2, ..., 10$ จะได้ว่า E_i เกิดได้ 2 วิธี ดังนั้นจำนวนวิธีที่นาย ก ดูซีรีส์อย่างน้อยหนึ่งเรื่อง

$$2^{10} - 1 = 1023 \quad \text{วิธี}$$

ทฤษฎีบท 3.2.6 หลักการบวก (The Addition Principle)
ให้ $k \geq 1$ ถ้ามีเหตุการณ์ $E_1, E_2, ..., E_k$ ซึ่งทุกคู่เป็นเหตุการณ์ไม่เกิดร่วมกัน โดย

เหตุการณ์ E_1 เกิดได้ n_1 วิธี
เหตุการณ์ E_2 เกิดได้ n_2 วิธี
$\vdots$
เหตุการณ์ E_k เกิดได้ n_k วิธี

แล้วจำนวนวิธีที่จะเกิดอย่างน้อยหนึ่งเหตุการณ์จาก $E_1, E_2, ..., E_k$ คือ

$$n_1 + n_2 + \cdots + n_k$$

ตัวอย่าง 3.2.7 จงหาจำนวนคู่อันดับ (x, y) ซึ่ง x และ y เป็นจำนวนเต็ม และ $x^2 + y^2 \leq 3$
แนวคำตอบ ให้

$$\begin{aligned}
E_1 &= \{(x,y) : x^2 + y^2 = 0\} = \{(0,0)\} \\
E_2 &= \{(x,y) : x^2 + y^2 = 1\} = \{(1,0), (-1,0), (0,1), (0,-1)\} \\
E_3 &= \{(x,y) : x^2 + y^2 = 2\} = \{(1,1), (-1,1), (1,-1), (-1,-1)\} \\
E_4 &= \{(x,y) : x^2 + y^2 = 3\} = \varnothing
\end{aligned}$$

ดังนั้นจงหาจำนวนคู่อันดับ (x, y) ทั้งหมดเท่ากับ

$$1 + 4 + 4 + 0 = 9$$

ตัวอย่าง 3.2.8 จำนวนเต็มสี่หลักโดยแต่ละหลักไม่ซ้ำกัน ที่สร้างจาก 0, 2, 3, 5, 8 มีทั้งหมดกี่จำนวน
แนวคำตอบ ให้ E_1 คือเหตุการณ์ที่หลักหน่วยเป็น 0 จะได้ว่า

เหตุการณ์หลักหน่วย	เกิดได้	1	วิธี
เหตุการณ์หลักสิบ	เกิดได้	4	วิธี
เหตุการณ์หลักร้อย	เกิดได้	3	วิธี
เหตุการณ์หลักพัน	เกิดได้	2	วิธี

จะได้ว่า E_1 เกิดได้ $1 \times 4 \times 3 \times 2 = 24$ วิธี
และให้ E_2 คือเหตุการณ์ที่หลักหน่วยไม่ใช่ 0

เหตุการณ์หลักหน่วย	เกิดได้	4	วิธี
เหตุการณ์หลักพัน	เกิดได้	4	วิธี
เหตุการณ์หลักสิบ	เกิดได้	3	วิธี
เหตุการณ์หลักร้อย	เกิดได้	2	วิธี

จะได้ว่า E_2 เกิดได้ $4 \times 4 \times 3 \times 2 = 96$ วิธี
ดังนั้นจำนวนเต็มสี่หลักโดยแต่ละหลักไม่ซ้ำกันเท่ากับ $24 + 96 = 120$

บทนิยาม 3.2.9 วิธีเรียงสับเปลี่ยน (permutation) คือการเรียงลำดับบางสิ่งหรือทุกสิ่งในเซตโดยคำนึงถึงลำดับที่

บทนิยาม 3.2.10 ให้ n เป็นจำนวนนับ $\boldsymbol{n}$ **แฟคทอเรียล** (n-factorial) เขียนแทนด้วย $n!$ หมายถึงผลคูณของจำนวนเต็มตั้งแต่ 1 ถึง n นั่นคือ

$$n! = n(n-1)(n-2)\cdots 2\cdot 1$$

ในกรณี $n = 0$ นิยาม $0! = 1$

ทฤษฎีบท 3.2.11 จำนวนวิธีเรียงสับเปลี่ยนหรือ **วิธีเรียงสับเปลี่ยนเป็นเส้นตรง** (linear permutation) ของ n สิ่งที่แตกต่างกันทั้งหมด โดยนำมาจัดที่ละ n สิ่ง มีค่าเท่ากับ

$$n! \text{ วิธี}$$

บทพิสูจน์. ให้สิ่งของต่างกัน n สิ่ง เมื่อ $n \in \mathbb{N}$ จะได้ว่า

ตำแหน่งที่ 1	เกิดได้	n	วิธี
ตำแหน่งที่ 2	เกิดได้	$n-1$	วิธี
ตำแหน่งที่ 3	เกิดได้	$n-2$	วิธี
		$\vdots$	
ตำแหน่งที่ n	เกิดได้	1	วิธี

โดยหลักการคูณ เหตุการณ์ในการเรียงสิ่งของต่างกัน n เท่ากับ

$$n(n-1)(n-2)\cdots 1 = n!$$

□

ทฤษฎีบท 3.2.12 จำนวนวิธีเรียงสับเปลี่ยนของ n สิ่งที่แตกต่างกันทั้งหมด โดยเรียงสับเปลี่ยนที่ละ r สิ่ง เมื่อ $0 < r < n$ จะได้เท่ากับ

$$\frac{n!}{(n-r)!} \text{ วิธี}$$

เขียนแทนด้วย ${}^{n}P_{r}$

บทพิสูจน์. ให้สิ่งของต่างกัน n สิ่ง เมื่อ $n \in \mathbb{N}$ จะได้ว่า

ตำแหน่งที่ 1	เกิดได้	n	วิธี
ตำแหน่งที่ 2	เกิดได้	$n-1$	วิธี
ตำแหน่งที่ 3	เกิดได้	$n-2$	วิธี
		$\vdots$	
ตำแหน่งที่ r	เกิดได้	$n-(r-1)$	วิธี

โดยหลักการคูณ เหตุการณ์ในการเรียงสิ่งของต่างกัน n โดยเรียงสับเปลี่ยนที่ละ r สิ่ง เท่ากับ

$$n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} = {}^{n}Pr$$

□

โดยทฤษฎีบท 3.2.12 ในกรณี $r = n$ จำนวนวิธีเรียงสับเปลี่ยนของ n สิ่งที่แตกต่างกันทั้งหมด โดยเรียงสับเปลี่ยนที่ละ n เท่ากับ $n!$ วิธี หรืออาจเขียนได้ว่า

$$n! = {}^{n}P_n = \frac{n!}{(n-n)!} = \frac{n!}{0!}$$

เราจึงนิยามได้ว่า $0! = 1$ ดังบทนิยามของแฟคทอเรียล
ในกรณี $r = 0$ จำนวนวิธีเรียงสับเปลี่ยนของ n สิ่งที่แตกต่างกันทั้งหมด โดยเรียงสับเปลี่ยนที่ละ 0 สิ่ง หมายถึงไม่เกิดการเรียงสับเปลี่ยน ซึ่งทำได้เท่ากับ 1 วิธี ซึ่งสอดคล้องที่ว่า

$${}^{n}P_0 = \frac{n!}{(n-0)!} = 1$$

ดังนั้น ${}^{n}P_r$ จึงนิยามได้ เมื่อ $0 \le r \le n$

ตัวอย่าง 3.2.13 ถ้านำอักษรจากคำว่า PRICE มาจัดเรียงใหม่ 3 ตัวโดยไม่สนใจความหมาย จะได้ทั้งหมดกี่คำ

แนวคำตอบ จะเห็นว่ามีตัวอักษร 5 ตัวที่แตกต่างกันนำมาเรียงทีละ 3 ตัว เกิดคำได้ทั้งหมด

$${}^{5}P_3 = \frac{5!}{(5-3)!} = 5 \times 4 = 20$$

ตัวอย่าง 3.2.14 จงหาจำนวนวิธีที่จะจับฉลากรางวัลที่ 1 และรางวัลที่ 2 จากฉลาก 20 ใบ ซึ่งมีใบที่ถูกรางวัลทั้ง 2 ชนิดรวมอยู่ในนั้น

แนวคำตอบ จะเห็นว่ามีฉลาก 20 ใบที่แตกต่างกันนำมาเรียงทีละ 2 ใบ ทำได้ทั้งหมด

$${}^{20}P_2 = \frac{20!}{(20-2)!} = 20 \times 19 = 380$$

ตัวอย่าง 3.2.15 ในการจัดนักศึกษา 5 คนในการเข้าแถวเพื่อเข้าห้องสอบ

1. จงหาจำนวนวิธีในการจัด
   แนวคำตอบ จัดนักศึกษา 5 คนในการเข้าแถวได้ $5! = 120$ วิธี
2. ถ้ามีนักศึกษา 2 คน ไม่ยอมยืนติดกัน จะจัดได้กี่วิธี
   แนวคำตอบ สมมติว่านักศึกษาที่ไม่ยืนติดกันคือ นาย ก และ นาย ข

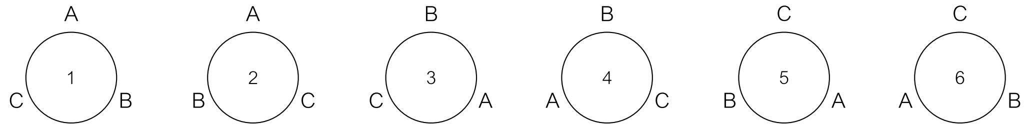
จัด 3 คนที่ไม่ใช่ นาย ก และ นาย ข	เกิดได้	$3!$	วิธี
จัด นาย ก ไปแทรกระหว่าง 3 คนแรกที่จัดไว้	เกิดได้	4	วิธี
จัด นาย ข ไปแทรกระหว่าง 3 คนแรกที่จัดไว้ ไม่แทรกตำแหน่งที่มี นาย ก	เกิดได้	3	วิธี

โดยหลักการคูณ เหตุการณ์มีนักศึกษา 2 คน ไม่ยอมยืนติดกันเท่ากับ

$$3! \times 4 \times 3 = 72$$

พิจารณาการเขียนแผนภาพแสดงจำนวนวิธีเรียงสับเปลี่ยน A, B, C รอบวงกลม

รูปที่ 3.3: แสดงการเรียงสับเปลี่ยน A, B, C รอบวงกลม



จากรูปที่ 3.3 แสดงการเรียงสับเปลี่ยนโดยการแจกแจงเช่นเดียวกับเรียงแบบเส้นตรง แต่จะเห็นว่า

แบบที่ 1, 4 และ 5 เป็นแบบเดียวกัน นั่นคือ ABC, BCA, CAB
แบบที่ 2, 3 และ 6 เป็นแบบเดียวกัน นั่นคือ ACB, BAC, CBA

เนื่องจากการเรียงแบบวงกลมสามารถหมุนได้ โดยการนำหัวและท้ายมาต่อกัน แสดงได้ว่า

ABC, BCA, CAB นับเป็น 1 วิธี

ตัวอย่าง 3.2.16 จงหาจำนวนวิธีเรียงสับเปลี่ยนเป็นวงกลมที่ละ 3 สิ่งจาก $A = \{a, b, c, d\}$

แนวคำตอบ พิจารณาวิธีเรียงสับเปลี่ยนเป็นเส้นตรงได้ ${}^4P_3 = 24$ วิธี จัดเป็นกลุ่มได้ 8 กลุ่ม ๆ ละ 3 สิ่ง ได้ดังนี้

abc	cab	bca	acb	bac	cba
abd	dab	bda	adb	bad	dba
acd	dac	cda	adc	cad	dca
bcd	dbc	cdb	bdc	cbd	dcb

จะเห็นว่า $\{abc, cab, bca\}$ แทนการเรียงสับเปลี่ยนวงกลม 1 แบบ จำนวนวิธีเรียงสับเปลี่ยนวงกลมที่ละ 3 สิ่ง ของเซต A เขียนแทน ${}^4Q_3$ ฉะนั้น

$${}^4Q_3 = \frac{24}{3} = 8$$

จากตัวอย่าง 3.2.16 ขยายแนวคิดไปกรณีทั่วไปได้ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 3.2.17 จำนวน **วิธีเรียงสับเปลี่ยนแบบวงกลม** (circular permutation) ที่ละ r สิ่งที่ต่างกันจากของ n สิ่ง สิ่งที่แตกต่างกันทั้งหมด เขียนแทนด้วย ${}^nQ_r$ มีค่าเท่ากับ

$${}^nQ_r = \frac{{}^nP_r}{n} \quad \text{วิธี}$$

ในกรณีที่ $r = n$ จะได้ว่าจำนวนวิธีเรียงสับเปลี่ยนเป็นวงกลมของของ n สิ่งที่แตกต่างกันเท่ากับ

$${}^nQ_n = (n-1)!$$

ตัวอย่าง 3.2.18 ในการรับประทานอาหารของครอบครัวหนึ่งนั่งรอบโต๊ะกลมประกอบไปด้วย 4 คน คือ พ่อ แม่ ลูกชาย และลูกสาว

1. จงหาจำนวนวิธีในการจัดดังกล่าว
   แนวคำตอบ จัดคน 4 คนนั่งรอบโต๊ะกลมได้ ${}^4Q_4 = 3! = 6$ วิธี

2. ถ้าพ่อและแม่ต้องติดกันเสมอ จะจัดได้กี่วิธี
   แนวคำตอบ ขั้นแรก จัด แม่ ลูกชาย และลูกสาว นั่งรอบโต๊ะกลมได้ $(3-1)!$
   ลำดับถัดไป จัด พ่อ ลงไปนั่งติดกับแม่ ทำได้ 2 วิธี
   โดยหลักการคูณจะได้ว่าจำนวนวิธีนั่งรอบโต๊ะกลมโดยที่พ่อและแม่ต้องติดกันเสมอเท่ากับ

$$(3-1)! \times 2 = 4$$

3. ถ้าลูกทั้งสองห้ามนั่งติดกัน จะจัดได้กี่วิธี
   แนวคำตอบ ขั้นแรก พ่อและแม่ นั่งรอบโต๊ะกลมได้ $(2-1)!$
   ลำดับถัดไป จัดลูกชายนั่งแทรกระหว่างพ่อและแม่ ทำได้ 2 วิธี
   ลำดับถัดสุดท้าย จัดลูกสาวนั่งที่ว่างที่เหลือ ทำได้ 1 วิธี
   โดยหลักการคูณจะได้ว่าจำนวนวิธีนั่งรอบโต๊ะกลมโดยที่ลูกทั้งสองห้ามนั่งติดกัน

$$(2-1)! \times 2 \times 1 = 2$$

หลักคอมพลีเมนท์ (Principle of Complementation)
ถ้า A เป็นเซตย่อยของเอกภพสัมพัทธ์จำกัด $\mathcal{U}$ และ A^c เป็น**คอมพลีเมนท์** (complement) ของ A แล้ว

$$n(A^c) = n(\mathcal{U}) - n(A)$$

ตัวอย่าง 3.2.19 จำนวนเต็มสี่หลักที่มี 7 ปะปนอยู่อย่างน้อยหนึ่งหลัก มีทั้งหมดกี่จำนวน

แนวคำตอบ ให้ A แทนเหตุการณ์ของจำนวนเต็มสี่หลักที่มี 7 ปะปนอยู่อย่างน้อยหนึ่งหลัก จะได้ว่า A^c คือเหตุการณ์ของจำนวนเต็มสี่หลักที่มีไม่มี 7 ปะปนอยู่แม้แต่หลักเดียว ซึ่งสร้างจำนวนเต็มสี่หลักจาก $\{0, 1, 2, 3, 4, 5, 6, 8, 9\}$ จะได้ว่า

เหตุการณ์หลักหน่วย	เกิดได้	9	วิธี
เหตุการณ์หลักสิบ	เกิดได้	9	วิธี
เหตุการณ์หลักร้อย	เกิดได้	9	วิธี
เหตุการณ์หลักพัน	เกิดได้	8	วิธี

จะได้ว่า $n(A^c) = 9 \times 9 \times 9 \times 8$ และจำนวนเต็มสี่หลักทั้งหมดเท่ากับ $n(\mathcal{U}) = 9 \times 10 \times 10 \times 10$ โดยหลักการคอมพลีเมนท์

$$n(A) = 9 \times 10 \times 10 \times 10 - 9 \times 9 \times 9 \times 8 = 3168$$

ดังนั้นจำนวนเต็มสี่หลักที่มี 7 ปะปนอยู่อย่างน้อยหนึ่งหลัก มีทั้งหมด 3168 จำนวน

บทนิยาม 3.2.20 การจัดหมู่ (Combination) ของเซตของสิ่งของ คือการหาเซตย่อยใด ๆ ของเซต โดยไม่คำนึงถึงลำดับที่ในเซตย่อยนั้น

ให้ $A = \{a, b, c, d\}$ จัดหมู่ที่ละ 3 สิ่งได้ดังนี้

$$\{a, b, c\},\ \{a, b, d\},\ \{a, c, d\},\ \{b, c, d\}$$

ทำได้ 4 วิธี เขียนแทนด้วย ${}^4C_3$ หรือ $\binom{4}{3}$ นั่นคือ ${}^4C_3 = 4$

ให้ $n(A) = n$ และ $0 \le r \le n$ การเรียงสับเปลี่ยนที่ละ r สิ่งของเซต A ทำได้ 2 ขั้นตอนคือ

ขั้นที่ 1 สร้างการจัดหมู่ที่ละ r สิ่งของเซต A เกิดได้ ${}^nC_r$ วิธี
ขั้นที่ 2 เรียงสับเปลี่ยนของ r นั้นเป็นเส้นตรง เกิดได้ $r!$ วิธี

จะได้ว่า ${}^nP_r = {}^nC_r \cdot r!$ ดังนั้นจำนวนวิธีในการจัดหมู่ที่ละ r สิ่งจากเซต A ที่มีสมาชิกต่างกัน n เท่ากับ

$$\binom{n}{r} = \frac{{}^nP_n}{r!} = \frac{n!}{(n-r)!r!}$$

และจะเห็นได้ว่า $\binom{n}{r} = \dfrac{n!}{(n-r)!r!} = \dfrac{n!}{(n-r)!(n-(n-r))!} = \binom{n}{n-r}$

ตัวอย่าง 3.2.21 จากคนในกลุ่มหนึ่งมีนักเรียนชาย 6 คน นักเรียนหญิง 4 คน ต้องการเลือกกรรมการที่ประกอบด้วยชาย 4 หญิง 3 ทำได้ทั้งหมดกี่วิธี

แนวคำตอบ ให้ A เป็นเซตของทีมฟุตบอลทั้งหมด นั่นคือ $n(A) = 20$ จะได้ว่าจำนวนวิธีในการจัดหมู่ที่ละ 2 ทีมจากเซต A เท่ากับ

ขั้นที่ 1 การจัดหมู่นักเรียนชายที่ละ 4 คน จาก 6 คน เกิดได้ ${}^6C_4$ วิธี
ขั้นที่ 2 การจัดหมู่นักเรียนหญิงที่ละ 3 คน จาก 4 คน เกิดได้ ${}^4C_3$ วิธี

ดังนั้นจำนวนวิธีในการเลือกกรรมการที่ประกอบด้วยชาย 4 หญิง 3 ทำได้

$${}^6C_4 \cdot {}^4C_3 = \frac{6!}{2!4!} \cdot \frac{4!}{1!3!} = 60$$

ตัวอย่าง 3.2.22 ในการแข่งขันฟุตบอลระดับเขตมีทีมเข้าแข่งขัน 20 ทีม ถ้าจัดแบบพบกันหมด จัดได้ทั้งหมดกี่คู่

แนวคำตอบ ให้ A เป็นเซตของทีมฟุตบอลทั้งหมด นั่นคือ $n(A) = 20$ จะได้ว่าจำนวนวิธีในการจัดหมู่ที่ละ 2 ทีมจากเซต A เท่ากับ

$$\binom{20}{2} = \frac{20!}{18!2!} = \frac{20 \times 19}{2} = 190$$

ดังนั้น การแข่งขันฟุตบอลแบบพบกันหมดจัดได้ 190 คู่

ตัวอย่าง 3.2.23 กล่องใบหนึ่งมีลูกบอลสีแดง 3 ลูกต่างกัน สีเขียว 3 ลูกต่างกัน และสีดำ 2 ลูกต่างกัน เมื่อหยิบลูกบอลออกมาพร้อมกัน 3 ลูก จงหาจำนวนวิธีทั้งหมดที่หยิบได้สีแดง n ลูก

แนวคำตอบ จะได้ว่า $n = 0, 1, 2, 3$ โดยแบ่งผลการหยิบได้ 2 กลุ่มคือ ลูกบอลสีแดงจำนวน 3 ลูก และลูกบอลที่ไม่ใช่สีแดงจำนวน 5 ลูก แสดงวิธีได้ดังตาราง

จำนวลูกบอลสีแดง (n)	จำนวนวิธี
0	${}^3C_0 \cdot {}^5C_3 = 10$
1	${}^3C_1 \cdot {}^5C_2 = 30$
2	${}^3C_2 \cdot {}^5C_1 = 15$
3	${}^3C_3 \cdot {}^5C_0 = 1$

หรือตอบในรูปทั่วไปได้เป็น

$${}^3C_n \cdot {}^5C_{3-n} \quad \text{เมื่อ} \quad n = 0, 1, 2, 3$$

ตัวอย่าง 3.2.24 ให้ n และ r เป็นจำนวนเต็มบวกซึ่ง $r \leq n$ จงแสดงว่า

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

บทพิสูจน์. จะได้ว่า

$$\begin{aligned}
\binom{n-1}{r} + \binom{n-1}{r-1} &= \frac{(n-1)!}{(n-r-1)!r!} + \frac{(n-1)!}{(n-r)!(r-1)!} \\
&= \frac{(n-1)! \cdot (n-r)}{(n-r) \cdot (n-r-1)!r!} + \frac{(n-1)! \cdot r}{(n-r)! \cdot r \cdot (r-1)!} \\
&= \frac{(n-1)! \cdot (n-r)}{(n-r)!r!} + \frac{(n-1)! \cdot r}{(n-r)!r!} \\
&= \frac{(n-1)!}{(n-r)!r!} [(n-r) + r] \\
&= \frac{(n-1)! \cdot n}{(n-r)!r!} = \frac{n!}{(n-r)!r!} = \binom{n}{r}
\end{aligned}$$

□

ตัวอย่าง 3.2.25 จงเรียงสับเปลี่ยนตัวอักษร 5 ตัว a, a, a, b, c ได้ทั้งหมดกี่วิธี

แนวคำตอบ ให้ N เป็นจำนวนวิธีในการเรียงสับเปลี่ยน a, a, a, b, c
พิจารณาการเรียงสับเปลี่ยนวิธีหนึ่งคือ $baaca$ โดยมองว่า a ทั้งสามแตกต่างกัน นั่นคือ a_1, a_2, a_3
นั่นหมายความว่าวิธีเรียงสับเปลี่ยน $ba_1a_2ca_3$ เท่ากับ $3!$
แต่ทั้งหมดเท่ากับการเรียงสับเปลี่ยน $\{a_1, a_2, a_3, b, c\}$ ดังนั้น

$$\begin{aligned}
N \cdot 3! &= 5! \\
N &= \frac{5!}{3!} = 20
\end{aligned}$$

ทฤษฎีบท 3.2.26 ในการเรียงสับเปลี่ยน n สิ่งซึ่งมี

n_1 สิ่งเหมือนกัน n_2 สิ่งเหมือนกัน .... n_k สิ่งเหมือนกัน

จะได้จำนวนวิธีเรียงสับเปลี่ยนเท่ากับ

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

โดยที่ $n_1+n_2+\cdots+n_k=n$

บทพิสูจน์. เป็นแบบฝึกหัด □

ตัวอย่าง 3.2.27 จงหาจำนวนวิธีในการเรียงตัวอักษร APPEARS ใหม่จะจัดได้กี่วิธี โดยที่

1. ไม่มีเงื่อนไข
2. A ต้องติดกัน
3. P ไม่ติดกัน

แนวคำตอบ จะเห็นว่ามี A 2 ตัว P 2 ตัว E R และ S อย่างละ 1 ตัว

1. จำนวนวิธีในการเรียงตัวอักษร APPEARS เท่ากับ

$$\frac{7!}{2!2!1!1!1!}=1260$$

2. เนื่องจาก A ติดกัน นั่นคือมอง A 2 ตัว เป็น 1 กลุ่มที่ติดกันไป ดังนั้นจำนวนวิธีเรียงตัวอักษร APPEARS โดยที่ A ติดกัน เท่ากับ

$$\frac{6!}{1!2!1!1!1!}=360$$

3. พิจารณา P ติดกัน เนื่องจาก P มี 2 ตัว ในทำนองเดียวกับข้อ 2. จะได้เรียงได้ทั้งทั้งหมด 360 วิธี โดยหลักคอมพลีเมนต์ ดังนั้นจำนวนวิธีเรียงตัวอักษร APPEARS โดยที่ P ไม่ติดกัน เท่ากับ

$$1260-360=900$$

ตัวอย่าง 3.2.28 ถ้านำอักษรจากคำว่า APPLE มาจัดเรียงใหม่ 3 ตัวโดยไม่สนใจความหมาย จะได้ทั้งหมดกี่คำ

แนวคำตอบ จะเห็นว่ามี P 2 ตัว A L และ E อย่างละ 1 ตัว พิจารณาการจัดเรียง 3 ตัว โดย

กรณีที่ 1 ไม่มี P นั่นคือการเรียง ALE ทำได้ $3!=6$ วิธี

กรณีที่ 2 มี P 1 ตัว นั่นคืออีก 2 ตัว เลือกจาก A,L,E ทำได้

$${}^3C_2\cdot 3!=18 \text{ วิธี}$$

กรณีที่ 3 มี P 2 ตัว นั่นคืออีก 1 ตัว เลือกจาก A,L,E ทำได้

$${}^3C_1\cdot\frac{3!}{2!1!}=9 \text{ วิธี}$$

ดังนั้นจำนวนวิธีทั้งหมดเท่ากับ $6+18+9=33$

ตัวอย่าง 3.2.29 จงหาจำนวนวิธีแจกดินสอไม้ 10 แท่งที่เหมือนกันให้เด็ก 13 คน คนละแท่งโดยมี 3 คนไม่ได้รับดินสอ

แนวคำตอบ เลือกเด็ก 10 คนจาก 13 คน จัดได้ ${}^{13}C_{10}$ จากนั้นจัดดินสอไม้ให้เด็กได้ 1 วิธี เนื่องจากดินสอไม้เหมือนกัน ดังนั้นจำนวนวิธีจัดเท่ากับ

$${}^{13}C_{10} \cdot 1 = 286$$

ตัวอย่าง 3.2.30 ต้องการจะแบ่งกลุ่มสมาชิกของ $\{a, b, c, d, e\}$ ออกเป็น 2 กลุ่ม กลุ่มละ 2 ตัว และกลุ่มละ 3 ตัว ทำได้กี่วิธี

แนวคำตอบ พิจารณาการจัดกลุ่มจากตารางต่อไปนี้

$\{a, b, c\}$	$\{d, e\}$	$\{b, c, d\}$	$\{a, e\}$
$\{a, b, d\}$	$\{c, e\}$	$\{b, c, e\}$	$\{a, d\}$
$\{a, b, e\}$	$\{d, c\}$	$\{b, d, e\}$	$\{a, c\}$
$\{a, c, d\}$	$\{b, e\}$	$\{c, e, d\}$	$\{a, b\}$
$\{a, c, e\}$	$\{b, d\}$	$\{a, d, e\}$	$\{b, c\}$

ทำได้อีกวิธีคือ จัดกลุ่มละ 2 ตัวจาก 5 ตัว ได้ ${}^5C_2$ วิธี จากนั้นจัดกลุ่มละ 3 ตัวจาก 3 ตัวที่เหลือ ดังนั้นจำนวนวิธีที่จัดได้เท่ากับ

$${}^5C_2 \cdot \ {}^3C_3 = \frac{5!}{3!2!} \cdot \frac{3!}{0!3!} = \frac{5!}{2!3!} = 10$$

เขียนแทนด้วย $\begin{pmatrix} 5 \\ 2, 3 \end{pmatrix} = \dfrac{5!}{2!3!} = 10$

ตัวอย่าง 3.2.31 ต้องการจะแบ่งกลุ่มนักเรียน 9 คน ออกเป็น 3 กลุ่ม กลุ่มละ 2 คน กลุ่มละ 3 คน และ 4 คน ทำได้กี่วิธี

แนวคำตอบ จัดกลุ่มละ 2 คนจาก 9 คน ได้ ${}^9C_2$ วิธี จากนั้นจัดกลุ่มละ 3 คนจาก 7 คนที่เหลือจากการจัดครั้งแรก สุดท้ายจัดกลุ่มละ 4 ตัวจาก 4 คนที่เหลือ ดังนั้นจำนวนวิธีที่จัดได้เท่ากับ

$$\begin{pmatrix} 9 \\ 2, 3, 4 \end{pmatrix} = \ {}^9C_2 \cdot \ {}^7C_3 \cdot \ {}^4C_4 = \frac{9!}{7!2!} \cdot \frac{7!}{4!3!} \cdot \frac{4!}{0!4!} = \frac{9!}{2!3!4!} = 1260$$

ในกรณีทั่วไป สรุปได้ว่า

ทฤษฎีบท 3.2.32 จำนวนวิธี **แบ่งกลุ่ม** (partitioning) n สิ่งต่าง ๆ กัน ออกเป็น k กลุ่ม ๆ ละ $n_1, n_2, ..., n_k$ โดยที่จำนวนสมาชิกของแต่ละกลุ่มไม่ซ้ำกัน เท่ากับ

$$\begin{pmatrix} n \\ n_1, n_2, ..., n_k \end{pmatrix} = \frac{n!}{n_1!n_2! \cdots n_k!} \quad \text{วิธี}$$

ตัวอย่าง 3.2.33 ต้องการจะแบ่งกลุ่มนักเรียน 10 คน ออกเป็น 2 กลุ่ม กลุ่มละ 4 คนและ 6 คน ทำได้กี่วิธี

แนวคำตอบ จัดได้ $\begin{pmatrix} 10 \\ 4, 6 \end{pmatrix} = \dfrac{10!}{4!6!} = 210$ วิธี

ตัวอย่าง 3.2.34 ต้องการจะแบ่งกลุ่มนักเรียน 9 คน ออกเป็น 3 กลุ่ม กลุ่มละ 2, 3 และ 4 คน เพื่อเข้าห้องพักที่จัดไว้ 3 ห้องแต่ละห้องพักได้สูงสุด 4 คน ทำได้ทั้งหมดกี่วิธี

แนวคำตอบ จัดได้ 2 ขั้นตอนโดยเริ่มจากการจัดกลุ่ม 3 กลุ่ม จัดได้ $\begin{pmatrix} 9 \\ 2,3,4 \end{pmatrix}$ วิธี จากนั้นจัดแต่ละกลุ่มเข้าห้องพัก 3 ห้อง ได้ $3 \cdot 2 \cdot 1$ ดังนั้นจำนวนวิธีที่จัดได้ทั้งหมดคือ

$$\begin{pmatrix} 9 \\ 2,3,4 \end{pmatrix} \cdot 3 \cdot 2 \cdot 1 = \frac{9!}{2!3!4!} \cdot 3! = 7560$$

ตัวอย่าง 3.2.35 ต้องการจะแบ่งกลุ่มสมาชิกของ $\{a, b, c, d\}$ ออกเป็น 2 กลุ่ม ๆ ละ 2 ตัว ทำได้กี่วิธี

แนวคำตอบ พิจารณาการจัดกลุ่มจากตารางต่อไปนี้

$\{a, b\}$	$\{c, d\}$
$\{a, c\}$	$\{b, d\}$
$\{a, d\}$	$\{b, c\}$
$\{b, c\}$	$\{a, d\}$
$\{b, d\}$	$\{a, c\}$
$\{c, d\}$	$\{a, b\}$

$\rightarrow$

$\{a, b\}$	$\{c, d\}$
$\{a, d\}$	$\{b, c\}$
$\{b, d\}$	$\{a, c\}$

ให้ N เป็นจำนวนแบ่งกลุ่ม $\{a, b, c, d\}$ ออกเป็น 2 กลุ่ม ๆ ละ 2 ตัว
พิจารณาการแบ่งกลุ่มวิธีหนึ่งคือ $\{a, b\}$ และ $\{c, d\}$ ถ้าสนใจลำดับของกลุ่มที่ได้จากจากวิธีนี้ จัดได้ $2!$ วิธี ดังนั้น

$$\begin{aligned} N \cdot 2! &= \frac{4!}{2!2!} \\ N &= \frac{4!}{2!2!2!} = \frac{4!}{(2!)^2 2!} = 3 \end{aligned}$$

ในกรณีทั่วไป สรุปได้ดังนี้

ทฤษฎีบท 3.2.36 จำนวนวิธีแบ่งกลุ่ม n สิ่งต่างกัน ออกเป็น k กลุ่มเท่า ๆ กัน กลุ่มละ m เท่ากับ

$$\frac{n!}{(m!)^k \cdot k!}$$

เมื่อ $mk = n$

บทพิสูจน์. เป็นแบบฝึกหัด □

ตัวอย่าง 3.2.37 ต้องการจะแบ่งกลุ่มนักเรียน 10 คน ออกเป็น 3 กลุ่ม กลุ่มละ 3, 3 และ 4 คน ทำได้กี่วิธี

แนวคำตอบ จัดได้ $\begin{pmatrix} 10 \\ 3,3,4 \end{pmatrix} = \dfrac{10!}{(3!)^2 2! \cdot 4!} = 2100$ วิธี

แบบฝึกหัด 3.2

1. นาย ก, ข และ ค ต้องการขึ้นลิฟท์ที่มีทั้งหมด 3 ตัว มีกี่วิธีที่นาย ก และนาย ข ขึ้นด้วยกัน แต่นาย ค ขึ้นคนเดียว

2. สนามกีฬามีประตูเข้าออก 6 ประตู ชายคนหนึ่งจะเลือกเข้าออกจากสนามกีฬาได้ทั้งหมดกี่วิธี

   2.1 ไม่มีเงื่อนไข

   2.2 เข้าประตูใดจะออกประตูนั้นไม่ได้

   2.3 เข้าประตูใดต้องออกประตูนั้น

3. จำนวนเต็มคู่ที่อยู่ระหว่าง 3,000 – 7,000 และเลขในแต่ละหลักไม่ซ้ำกัน มีทั้งหมดกี่จำนวน

4. ต้องการสร้างจำนวนเต็มบวก 4 หลัก จาก 0 – 9 โดยให้แต่ละหลักไม่ซ้ำกัน และหารด้วย 2 หรือ 5 ลงตัว จะสร้างได้กี่จำนวน และมีกี่จำนวนซึ่งมีค่าไม่เกิน 6,500

5. ถ้าต้องการสร้างระหัสตัวอักษร A, B โดยกำหนดให้แต่ละระหัสอาจมีอักษร 1 ตัว หรือ 2 ตัว หรือ 3 ตัว หรือ 4 ตัว เท่านั้น จะสร้างได้หมดกี่ระหัส

6. จัดคน 8 คน ซึ่งมีสมชาย สมคิด และสมศรี รวมอยู่ด้วย เข้านั่งเรียงแถวตรงโดยที่สมศรีนั่งตรงกลางติดกับสมชายและสมคิดเสมอ จำนวนวิธีการจัดที่นั่งดังกล่าวมีค่าเท่ากับข้อใดต่อไปนี้

7. คุณลุง คุณป้า ลูกชาย และลูกสาว มาเยี่ยมครอบครัวเราซึ่งมี 4 คนคือ คุณพ่อ คุณแม่ ตัวฉัน และน้องชาย ในการจัดที่นั่งรอบโต๊ะอาหารกลมที่มี 8 ที่นั่ง โดยให้คุณลุงนั่งติดคุณพ่อ คุณป้านั่งติดกับคุณแม่ ลูกชายของคุณลุงนั่งติดกับน้องชายของฉัน และลูกสาวของคุณลุงนั่งติดกับฉัน จะมีจำนวนวิธีจัดได้เท่ากับข้อใดต่อไปนี้

8. จงหาจำนวนวิธีทั้งหมดในการจัดชาย 5 คนและหญิง 5 คนให้นั่งรอบโต๊ะกลมโดยให้ชายหญิงสลับกันและนาย ก และ นางสาว ข ต้องนั่งติดกัน

9. เลือกอักษร 2 ตัวจากคำว่า NEWS มาเรียงเป็นคำใหม่โดยไม่สนใจความหมาย จะเขียนได้ทั้งหมดกี่วิธี

10. มีเก้าอี้ 6 ตัวเรียงต่อกันเป็นแนวเส้นตรง จำนวนวิธีในการจัดคน 3 คนนั่งเก้าอี้ 6 ตัวนี้โดยที่ไม่มีสองคนใดเลยนั่งเก้าอี้ติดกัน

11. ชาย 6 คน หญิง 6 คนยืนเรียงกันเป็นเส้นตรงได้กี่วิธี ถ้า

    11.1 ชายและหญิงต้องยืนสลับกันคนต่อคน

    11.2 ชายและหญิงต้องสลับกันที่ละ 2 คน

    11.3 ชายและหญิงต้องสลับกันที่ละ 3 คน

    11.4 ชายหญิงคู่หนึ่งต้องยืนติดกันเสมอ

    11.5 หญิงต้องไม่ยืนติดกันเลย

11.6 หัวแถวและท้ายแถวต้องเป็นเพศเดียวกัน

12. จงหาว่ามีจำนวนเต็มคู่กี่จำนวนที่มีค่าอยู่ระหว่าง 20000 และ 70000 และไม่มีเลขโดดซ้ำกัน

13. ให้ S เป็นเซตของจำนวนนับที่เลขโดดเป็นสมาชิกในเซต $\{1, 3, 5, 7\}$ และไม่มีเลขโดดในหลักใดซ้ำกัน จงหาจำนวนสมาชิกของเซต S และผลบวกทั้งหมดของสมาชิกในเซต S

14. หนังสือคณิตศาสตร์เหมือนกัน 3 เล่ม เคมีเหมือนกัน 2 เล่ม ชีววิทยาเหมือนกัน 2 เล่ม จงหาวิธีการเรียงหนังสือบนชั้น ถ้า

    14.1 ไม่มีเงื่อนไข

    14.2 วิชาเดียวกันต้องอยู่ติดกัน

    14.3 คณิตศาสตร์อยู่ติดกันหมด

    14.4 คณิตศาสตร์อยู่แยกกันหมด

15. เด็ก 10 คนเข้าแถวเป็นวงกลมครั้งละ 6 คนได้กี่วิธี

16. ในการประชุมครั้งหนึ่งมีผู้แทนจาก 3 ประเทศเข้าร่วมประชุมโดยมีผู้แทนประเทศละ 3 คน จำนวนวิธีทั้งหมดที่จะให้ผู้แทนแต่ละชาติต้องนั่งติดกันในการประชุมรอบโต๊ะกลมเท่ากับเท่าใด

17. ต้องการจัดอักษร AAABBC เป็นวงกลม จะมีกี่วิธี ถ้า

    17.1 ไม่มีเงื่อนไข

    17.2 อักษรที่ซ้ำอยู่ติดกัน

    17.3 อักษร A อยู่ติดกัน

    17.4 อักษร B แยกกันหมด

18. ต้องการจัดอักษร AABBBB เป็นวงกลม จะมีวิธีจัดได้กี่วิธี

19. ในชมรมส่งเสริมคุณธรรมมีสมาชิกเป็นพ่อบ้าน 4 คน แม่บ้าน 2 คน และเยาวชน 5 คน ต้องการเลือกกรรมการชุดหนึ่งที่มี 3 คน โดยต้องมีเยาวชนอย่างน้อย 1 คนจำนวนวิธีเลือกเท่ากับข้อใดต่อไปนี้

20. สามีภรรยาคู่หนึ่งมีลูกชาย 2 คน มีทรัพย์สินอยู่ 10 อย่าง โดยที่จัดกลุ่มตามมูลค่าที่ใกล้เคียงกันจะแบ่งออกเป็นสองกลุ่ม กลุ่มที่หนึ่ง มี 4 อย่าง กลุ่มที่สอง มี 6 อย่าง และได้เขียนพินัยกรรมโดยระบุว่าลูกชายคนโตได้รับทัพย์สินจากกลุ่มที่หนึ่ง 2 อย่าง และจากกลุ่มที่สอง 3 อย่าง ผู้จัดการมรดกจะแบ่งทัพย์สินให้ลูกคนโตได้กี่วิธี

21. สุ่มเลือกจำนวนตั้งแต่ 1 ถึง 15 มา 5 จำนวน จงหาจำนวนวิธีที่จะได้จำนวนซึ่งมีผลบวกของทั้ง 5 จำนวนหารด้วย 3 ลงตัว

22. ต้องการจะแบ่งกลุ่มนักเรียน 12 คน ออกเป็น 3 กลุ่ม กลุ่มละ 3, 4 และ 5 คน ทำได้กี่วิธี

23. ไพ่สำรับหนึ่งมี 52 ใบ สุ่มหยิบมา 2 ใบ จงหาจำนวนวิธีที่

    23.1 หยิบได้ K ทั้งคู่

    23.2 หยิบได้โพแดงทั้งคู่

    23.3 หยิบได้ K และ Q อย่างละใบ

24. กล่องใบหนึ่งมีลูกบอลสีแดง 3 ลูกต่างกัน สีเขียว 3 ลูกต่างกัน และสีน้ำเงิน 2 ลูกต่างกัน หยิบออกมา 4 ลูก จงหาจำนวนวิธีที่

    24.1 สีแดง 2 ลูก สีเขียว 1 ลูก และสีน้ำเงิน 1 ลูก

    24.2 ต้องได้สีแดงอย่างน้อย 1 ลูก

    24.3 ไม่ได้สีน้ำเงินเลย

25. นักเรียนห้องหนึ่ง 18 คน เป็นชาย 10 คน จะเลือกคณะกรรมการห้อง 5 คน ได้กี่วิธี ถ้า

    25.1 เป็นชาย 3 คน

    25.2 เป็นชายอย่างน้อย 1 คน

    25.3 เป็นชายอย่างมาก 3 คน

26. จะเลือกอักษร 4 ตัวจากคำว่า SAPHIRE มาสร้างคำใหม่โดยไม่สนใจความหมายได้กี่คำ ถ้า

    26.1 คำนั้นต้องไม่มี R

    26.2 คำนั้นต้องมี R

    26.3 คำนั้นต้องขึ้นต้นด้วย S

    26.4 คำนั้นต้องมีสระและพยัญชนะอย่างละ 2 ตัว

27. บนเส้นรอบวงของวงกลมหนึ่งมีจุด 8 จุดที่แตกต่างกัน จงหา

    27.1 จำนวนคอร์ดทั้งหมด

    27.2 จำนวนสามเหลี่ยมทั้งหมดที่มีจุดเหล่านั้นเป็นจุดยอด

    27.3 จำนวนรูปเหลี่ยมทั้งหมดที่มีจุดเหล่านั้นเป็นจุดยอด

    27.4 จำนวนจุดตัดที่มากที่สุดของคอร์ด (ไม่มีคอร์ด 3 เส้นตัดที่จุดเดียวกัน)

28. จะมีวิธีการเลือกอักษร 4 อักษรจากคำว่า MISSISSIPPI ได้กี่วิธี

29. แจกเหรียญเหมือนกัน 6 เหรียญ (แจกหมด) ให้เด็ก 3 คน (ก) ได้รับทุกคน (ข) ไม่สนใจว่าได้รับทุกคน

30. จงหาจำนวนนับ n ที่สอดคล้องสมการ

$$6\binom{9}{n}^{-1} = 6\binom{10}{n}^{-1} + 11\binom{11}{n}^{-1}$$

3.3 ความน่าจะเป็น

ในปริภูมิตัวอย่างเราจะกำหนดค่าของ **น้ำหนัก** (weight) ให้แก่จุดตัวอย่าง ซึ่ง**ผลบวกเหล่านั้นเท่ากับ 1** ถ้าเรามีเหตุผลที่จะเชื่อได้ว่าจุดตัวอย่างใดมีโอกาสเกิดขึ้นได้แน่นอน นั่นแสดงว่าจุดตัวอย่างที่ไม่มีโอกาสเกิดขึ้นเลยจะให้น้ำหนักเท่ากับ 0

บทนิยาม 3.3.1 ให้ S เป็นปริภูมิตัวอย่าง และ A เป็นเหตุการณ์ใด ๆ **ความน่าจะเป็น** (probability) ของเหตุการณ์ A คือ**ผลบวกของน้ำหนักทุก ๆ จุดตัวอย่างในเหตุการณ์** A เขียนแทนด้วย $P(A)$

$$P(A) = \text{ผลรวมของน้ำหนักของทุกจุดตัวอย่างใน } A$$

จากบทนิยามทำให้ได้ว่า $P(\varnothing) = 0, \quad P(S) = 1$ และ $0 \le P(A) \le 1$

ตัวอย่าง 3.3.2 โยนเหรียญ 1 อัน 2 ครั้ง (เที่ยงตรง) จงหาความน่าจะเป็นที่เหรียญจะขึ้นหัวอย่างน้อยหนึ่งครั้ง

แนวคำตอบ ปริภูมิตัวอย่างเท่ากับ $S = \{HH, HT, TH, TT\}$
ให้ w แทนน้ำหนักของแต่ละจุดตัวอย่าง (เที่ยงตรง) จะได้ว่า

จุดตัวอย่าง	HH	TH	HT	TT
น้ำหนัก	w	w	w	w

ทำให้ได้ว่า $w + w + w + w = 1$ นั่นคือ $w = \dfrac{1}{4}$
ให้ A คือเหตุการณ์ที่เหรียญจะขึ้นหัวอย่างน้อยหนึ่งครั้ง นั่นคือ $A = \{HT, TH, HH\}$ ดังนั้น

$$P(A) = P(\{HT, TH, HH\}) = w + w + w = 3w = 3 \cdot \frac{1}{4} = \frac{3}{4}$$

ตัวอย่าง 3.3.3 ลูกเต๋าลูกหนึ่งถูกถ่วงน้ำหนักโดยทำให้แต้มคู่มีโอกาสขึ้นเป็น 2 เท่าของแต้มคี่ จงหาความน่าจะเป็นที่ลูกเต๋าจะขึ้นแต้มน้อยกว่า 4

แนวคำตอบ ปริภูมิตัวอย่างเท่ากับ $S = \{1, 2, 3, 4, 5, 6\}$
ให้ w แทนน้ำหนักของลูกเต๋าที่ขึ้นแต้มคี่ จะได้ว่า

จุดตัวอย่าง	1	2	3	4	5	6
น้ำหนัก	w	$2w$	w	$2w$	w	$2w$

ทำให้ได้ว่า $w + 2w + w + 2w + w + 2w = 9$ นั่นคือ $w = \dfrac{1}{9}$
ให้ A คือเหตุการณ์ที่ที่ลูกเต๋าจะขึ้นแต้มน้อยกว่า 4 นั่นคือ $A = \{1, 2, 3\}$ ดังนั้น

$$P(A) = P(\{1, 2, 3\}) = w + 2w + w = 4w = 4 \cdot \frac{1}{9} = \frac{4}{9}$$

ตัวอย่าง 3.3.4 ในการทดลองสุ่มอย่างหนึ่งมีปริภูมิตัวอย่างเท่ากับ $S = \{a, b, c, d, e\}$ และมีน้ำหนักของจุดตัวอย่างดังตาราง

จุดตัวอย่าง	a	b	c	d	e
น้ำหนัก	w	$3w$	$5w$	$3w$	w

จงหาน้ำหนักของจุดตัวอย่างแต่ละตัว

แนวคำตอบ เนื่องจากน้ำหนักรวมของจุดตัวอย่างในปริภูมิตัวอย่างเท่ากับ 1 ทำให้ได้ว่า

$$w + 3w + 5w + 3w + w = 1 \text{ นั่นคือ } w = \frac{1}{13}$$

ดังนั้น

จุดตัวอย่าง	a	b	c	d	e
น้ำหนัก	$\frac{1}{13}$	$\frac{3}{13}$	$\frac{5}{13}$	$\frac{3}{13}$	$\frac{1}{13}$

ในกรณีที่**น้ำหนักทุกจุดตัวอย่างเท่ากัน** และปริภูมิตัวอย่าง S เป็นเซตจำกัด ให้ $n(S) = N$ และน้ำหนักของจุดตัวอย่างเท่ากับ w ดังนั้น $Nw = 1$ นั่นคือ

$$w = \frac{1}{N}$$

ให้ A เป็นหตุการณ์ซึ่ง $n(A) = k$ จะได้ว่า

$$P(A) = kw = k \cdot \frac{1}{N} = \frac{n(A)}{n(S)}$$

ตัวอย่าง 3.3.5 จงหาความน่าจะเป็นที่ครอบครัวตัว อ มีลูก 3 คน เป็นลูกชาย 2 คน

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = 2 \cdot 2 \cdot 2 = 8$ และ A เป็นเหตุการณ์ที่ครอบครัวตัว อ มีลูกชาย 2 คน นั่นคือ

$$A = \{ \text{ชชญ, ชญช, ญชช} \}$$

ดังนั้น

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{8}$$

ตัวอย่าง 3.3.6 ทอดลูกเต๋าเที่ยงตรง 2 ลูก จงหาความน่าจะเป็นผลบวกของแต้มไม่เกิน 10

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = 6 \cdot 6 = 36$ และ A เป็นเหตุการณ์ที่ผลบวกของแต้มไม่เกิน 10 พิจารณาตารางผลบวกของแต้มลูกเต๋า

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

จากตาราง $n(A) = 30$ ดังนั้น $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{30}{36} = \dfrac{5}{6}$

ตัวอย่าง 3.3.7 การโยนเหรียญ 1 อัน 10 ครั้ง จงหาความน่าจะเป็นที่เหรียญจะขึ้นหัว 2 ครั้ง

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = 2^{10}$ และ A เป็นเหตุการณ์ที่เหรียญจะขึ้นหัว 2 ครั้ง พิจารณา $HHTTTTTTTT$ จะเห็นได้ว่าแต่ละแบบคือเรียงสับเปลี่ยนแบบของซ้ำ นั่นคือ

$$n(A) = \frac{10!}{2!8!} = 45$$

ดังนั้น

$$P(A) = \frac{n(A)}{n(S)} = \frac{45}{2^{10}} = \frac{45}{1024}$$

ทฤษฎีบท 3.3.8 ให้ปริภูมิตัวอย่างเป็นเซตจำกัด ถ้า A และ B เป็นเหตุการณ์สองเหตุการณ์ จะได้ว่า

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

บทพิสูจน์. ให้ S เป็นปริภูมิตัวอย่างที่เป็นเซตจำกัด จะได้ว่า

$$\begin{aligned} n(A \cup B) &= n(A) + n(B) - n(A \cap B) \\ \frac{n(A \cup B)}{n(S)} &= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)} \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

□

บทแทรก 3.3.9 ให้ปริภูมิตัวอย่างเป็นเซตจำกัด ถ้า A และ B เป็นเหตุการณ์ที่ไม่เกิดร่วมกัน จะได้ว่า

$$P(A \cup B) = P(A) + P(B)$$

บทพิสูจน์. เนื่องจาก A และ B เป็นเหตุการณ์ที่ไม่เกิดร่วมกัน นั่นคือ $A \cap B = \varnothing$ จะได้ว่า $P(A \cap B) = 0$ ดังนั้นบทแทรกนี้เป็นจริงโดยทฤษฎีบท 3.3.8 □

ตัวอย่าง 3.3.10 ให้ปริภูมิตัวอย่างเป็นเซตจำกัด ถ้า A และ B เป็นเหตุการณ์ จงพิสูจน์ว่า

$$P(B - A) = P(B) - P(A \cap B)$$

แนวคำตอบ เนื่องจาก $B = (B - A) \cup (A \cap B)$ และ $(B - A) \cap (A \cap B) = \varnothing$ โดยบทแทรก 3.3.9 จะได้ว่า

$$P(B) = P(B - A) + P(A \cap B)$$

ดังนั้น $P(B - A) = P(B) - P(A \cap B)$

ตัวอย่าง 3.3.11 ความน่าจะเป็นของนักศึกษาที่จะสอบผ่านวิชาแคลคูลัส 2 เท่ากับ 0.35 ความน่าจะเป็นที่เขาจะสอบผ่านวิชาความน่าจะเป็นและสถิติ เท่ากับ 0.52 และความน่าจะเป็นที่เขาจะสอบผ่านอย่างน้อยหนึ่งวิชาเท่ากับ 0.83

1. จงหาความน่าจะเป็นที่เขาจะสอบผ่านสองวิชา
2. จงหาความน่าจะเป็นที่เขาสอบวิชาเดียว

แนวคำตอบ ให้ A แทนเหตุการณ์ที่นักศึกษาจะสอบผ่านวิชาแคลคูลัส 2
B แทนเหตุการณ์ที่นักศึกษาจะสอบผ่านวิชาความน่าจะเป็นและสถิติ
จะได้ว่า $P(A) = 0.35$, $P(B) = 0.52$ และ $P(A \cup B) = 0.83$

1. จะเห็นว่า $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.35 + 0.52 - 0.83 = 0.04$
   ดังนั้นความน่าจะเป็นที่เขาจะสอบผ่านสองวิชาเท่ากับ 0.04
2. เนื่องจาก $(A \cup B) \cap (A \cap B) = A \cap B$ จะเห็นว่า

$$\begin{aligned} P((A \cup B) - (A \cap B)) &= P(A \cup B) - P((A \cup B) \cap (A \cap B)) \\ &= P(A \cup B) - P(A \cap B) \\ &= 0.83 - 0.04 = 0.79 \end{aligned}$$

   ดังนั้นความน่าจะเป็นที่เขาสอบผ่านวิชาเดียวเท่ากับ 0.79

ตัวอย่าง 3.3.12 ในการทอดลูกเต๋า 2 ลูกพร้อมกัน 1 ครั้ง จงหาความน่าจะเป็นที่จะได้ผลบวกของแต้มเป็น 5 หรือ 11

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = 6 \cdot 6 = 36$
A เป็นเหตุการณ์ที่ผลบวกของแต้มเป็น 5 และ B เป็นเหตุการณ์ที่ผลบวกของแต้มเป็น 11
พิจารณาตารางผลบวกของแต้มลูกเต๋า

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

จากตาราง $n(A) = 4$ และ $n(B) = 2$ เนื่องจาก $A \cap B = \varnothing$ ดังนั้น

$$P(A \cup B) = P(A) + P(B) = \frac{4}{36} + \frac{2}{36} = \frac{1}{6}$$

ทฤษฎีบท 3.3.13 หลักการคอมพลีเมนต์
ให้ปริภูมิตัวอย่างเป็นเซตจำกัด ถ้า A เป็นเหตุการณ์ จะได้ว่า

$$P(A) = 1 - P(A^c)$$

บทพิสูจน์. ให้ S เป็นปริภูมิตัวอย่างที่เป็นเซตจำกัด เนื่องจาก $A \cup A^c = S$ และ $A \cap A^c = \varnothing$ ดังนั้น

$$P(S) = P(A) + P(A^c)$$

ดังนั้น $P(A) = 1 - P(A^c)$ $\square$

ตัวอย่าง 3.3.14 การโยนเหรียญ 1 อัน 6 ครั้ง จงหาความน่าจะเป็นที่เหรียญจะขึ้นหัวอย่างน้อย 1 ครั้ง

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = 2^6$ และ A เป็นเหตุการณ์ที่เหรียญจะขึ้นหัวอย่างน้อย 1 ครั้ง จะได้ว่า A^c เป็นเหตุการณ์ที่เหรียญจะขึ้นไม่ขึ้นหัวแม้แต่ครั้งเดียว นั่นคือ $n(A^c) = 1$ ดังนั้น

$$P(A) = 1 - P(A^c) = 1 - \frac{1}{2^6} = \frac{63}{64}$$

ตัวอย่าง 3.3.15 มีบัตรอยู่ 10 ใบ มีบัตรที่จะได้รางวัลอยู่ 3 ใบ ถ้านาย ก ซื้อมา 4 ใบ จงหาความน่าจะเป็นที่นาย ก จะถูกรางวัล

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = {}^{10}C_4$ และ A เป็นเหตุการณ์ที่นาย ก จะถูกรางวัล จะได้ว่า A^c เป็นเหตุการณ์ที่นาย ก จะไม่ถูกรางวัล นั่นคือ $n(A^c) = {}^{7}C_4$ ดังนั้น

$$P(A) = 1 - P(A^c) = 1 - \frac{{}^{7}C_4}{{}^{10}C_4} = 1 - \frac{1}{6} = \frac{5}{6}$$

ตัวอย่าง 3.3.16 น้องมิกต้องการขอเงินจากพี่แมก เพื่อไปซื้อขนมที่มีราคา 5 บาท แต่พี่แมกเสนอให้น้องมิกเล่นเกมเสี่ยงทาย โดยการโยนเหรียญ 1 เหรียญพร้อมกับลูกเต๋า 2 ลูก ถ้าขึ้นหัวจะได้เงินเท่ากับผลบวกของแต้มจากลูกเต๋าทั้งสอง ถ้าขึ้นก้อยจะได้เงินเท่ากับผลต่างของแต้มจากลูกเต๋าทั้งสอง จงหาโอกาสที่น้องมิกจะได้รับเงินในการเกมนี้

แนวคำตอบ ให้ S เป็นปริภูมิตัวอย่าง ซึ่ง $n(S) = 2 \cdot 6 \cdot 6 = 72$ และ A เป็นเหตุการณ์ที่น้องมิกจะได้รับเงินในการเกมนี้ จะได้ว่า A^c เป็นเหตุการณ์ที่น้องมิกจะไม่ได้รับเงินในการเกมนี้ นั่นคือ

$$A^c = \{(T,1,1),(T,2,2),(T,3,3),(T,4,4),(T,5,5),(T,6,6)\}$$

ดังนั้น

$$P(A) = 1 - P(A^c) = 1 - \frac{6}{72} = \frac{1}{12}$$

แบบฝึกหัด 3.3

1. ลูกเต๋าลูกหนึ่งถูกถ่วงน้ำหนักโดยทำให้แต้มคี่มีโอกาสขึ้นมากกว่าแต้มคู่อยู่ 40% จงหาความน่าจะเป็นที่ลูกเต๋าจะขึ้นแต้มอย่างน้อย 3

2. ในการโยนเหรียญ 1 เหรียญ 2 ครั้งที่ถ่วงน้ำหนักการเกิดน้อยกว่าการเกิดหัว 45% จงหาความน่าจะเป็นที่หัวทั้งสองครั้ง

3. ในการทดลองสุ่มอย่างหนึ่งมีปริภูมิตัวอย่างเท่ากับ $S = \{x, y, z, s, t\}$ และมีนำหนักของจัดตัวอย่างดังตาราง

จุดตัวอย่าง	x	y	z	s	t
น้ำหนัก	w	w	$2w$	k	$2k$

ถ้า $P(\{x, s\}) = \dfrac{4}{15}$

3.1 จงหาน้ำหนักของจุดตัวอย่างแต่ละตัว

3.2 จงหาความน่าจะเป็นที่ไม่เกิด x และ t

3.3 จงหา $P(\{x, s, t\})$

3.4 โอกาสที่จะเกิด z ต่างจาก t เท่าใด

4. มีถุงยังชีพ 5 ถุงต้องการแจกให้ครอบครัวที่ถูกน้ำท่วม 4 ครอบครัว ครอบครัวละไม่เกิน 2 ถุง ความน่าจะเป็นที่คอบครัวของสมชายซึ่งเป็นหนึ่งในสี่ครอบครัวนั้นไม่ได้รับของแจกเลยเท่ากับข้อใดต่อไปนี้

5. ในจำนวนเด็ก 12 คน มีเด็กถนัดซ้าย 4 คน ถ้าเลือกเด็ก 5 คนจากเด็กเหล่านี้ แล้วความน่าจะเป็นที่มีเด็กถนัดซ้ายในกลุ่มที่ถูกเลือกเท่ากับข้อใดใดต่อไปนี้

6. ถ้าชาย 3 คนและหญิง 4 คนเข้าคิวในแนวเดียวกันเพื่อซื้อตั๋วรถไฟขบวนหนึ่ง ความน่าจะเป็นที่หญิงทั้ง 4 คนยืนเรียงติดกันทั้งหมดในแถวเดียวกันเท่ากับเท่าใด

7. มีแคปซูลบรรจุยาชนิดหนึ่งจำนวน 5 แคปซูล ปนกับแคปซูลซึ่งบรรจุแป้งจำนวน 10 แคปซูล ถ้าหยิบมา 2 แคปซูลอย่างสุ่ม ความน่าจะเป็นที่จะได้แคปซูลซึ่งบรรจุยาทั้งสองแคปซูลเท่ากับเท่าใด

8. ตะกร้าใบหนึ่งมีส้ม น้อยหน่า และมะม่วงรวม 10 ลูก โดยที่จำนวนส้มเป็นสองเท่าของจำนวนน้อยหน่า และมีมะม่วง 1 ลูก ถ้าหยิบผลไม้ขึ้นมา 3 ลูกอย่างไม่เจาะจง ความน่าจะเป็นที่จะได้ผลไม้คนละชนิดกันเท่ากับเท่าใด

9. กล่องใบหนึ่งมีลูกแก้วขนาดเดียวกัน 13 ลูก เป็นสีแดง 6 ลูก สีขาว 4 ลูก นอกนั้นเป็นสีเหลือง สุ่มหยิบลูกแก้วมา 2 ลูก ความน่าจะเป็นที่จะได้ลูกแก้วสีต่างกันเท่ากับข้อใดใดต่อไปนี้

10. ในการเลือกกรรมการชุดหนึ่งจากชาย 6 คน และหญิง 4 คน ความน่าจะเป็นที่จะเลือกกรรมการ 3 คนให้มีทั้งหญิงและชาย

11. เลขานุการผู้หนึ่งพิมพ์จดหมาย 4 ฉบับถึงคน 4 คนพร้อมจ่าหน้าซอง ถ้าเขานำจดหมายใส่ซองอย่างสุ่มโดยไม่สนใจว่าจดหมายในซองจะถูกต้องหรือไม่ ความน่าจะเป็นที่มีจดหมาย 2 ฉบับใส่ในซองที่ถูกที่ถูกต้อง

12. ในการเลือกจำนวนเต็มหนึ่งจำนวนจากจำนวนตั้งแต่ 10 ถึง 59 จงหาความน่าจะเป็นที่ลขจำนวนนั้นหารด้วย 7 ลงตัว

13. ในการทอดลูกเต๋า 2 ลูกพร้อมกัน 1 ครั้ง ความน่าจะเป็นที่ผลรวมของแต้มบนลูกเต๋าทั้งสองลูกเป็นเลขที่หารด้วย 4 ไม่ลงตัว มีค่าเท่ากับเท่าใด

14. ข้าวสารบรรจุถุงแล้วกองหนึ่งประกอบไปด้วย ข้าวหอมมะลิ 4 ถุง ข้าวเสาไห 3 ถุง ข้าวตาแห้ง 2 ถุง และข้าวมัสบาตี 1 ถุง สุ่มหยิบข้าวจากกองนี้มา 4 ถุง ความน่าจะเป็นที่จะได้ข้าวครบทุกชนิดเท่ากับเท่าใด

15. ในลิ้นชักมีถุงเท้าสีขาว 4 คู่ สีดำ 3 คู่ และสีน้ำเงิน 2 คู่ แต่ไม่ได้จัดเรียงไว้เป็นคู่ๆ ถ้าสุ่มหยิบถุงเท้ามา 2 ข้างๆ ความน่าจะเป็นที่จะได้ถุงเท้าสีเดียวกันเท่ากับเท่าใด

16. กล่องใบหนึ่งมีบัตร 10 ใบแต่ละใบมีหมายเลข 0, 1, 2, ..., 9 บัตรละหนึ่งหมายเลข ถ้าหยิบบัตรจากกล่องพร้อมกัน 3 ใบ ความน่าจะเป็นที่จะได้บัตรหมายเลขคู่ทุกใบ และมีแต้มรวมกันมากกว่า 10 มีค่า เท่ากับเท่าใด

17. ถุงใบหนึ่งบรรจุลูกแก้วสีแดง 5 ลูก สีเขียว 4 ลูก และสีเหลือง 3 ลูก ถ้าหยิบลูกแก้วมาจากถุงทีละลูก 3 ครั้งโดยไม่ใส่คืน แล้วความน่าจะเป็นที่จะหยิบได้ลูกแก้วลูกที่หนึ่ง สอง และสาม เป็นสีแดง สีเขียว สีเหลือง ตามลำดับ เท่ากับเท่าใด

18. ถุงใบหนึ่งมีลูกกวาดขนาดเดียวกันสีแดง 24 เม็ด ที่เหลือเป็นสีขาวและสีเขียว ถ้าสุ่มหยิบลูกกวาดขึ้นมา 1 เม็ด ความน่าจะเป็นที่จะได้ลูกกวาดสีขาวหรือสีเขียวเท่ากับ 5/6 ความน่าจะเป็นที่จะได้ลูกกวาดสีเขียวหรือสีแดงเท่ากับ 3/4 แล้วจำนวนลูกกวาดสีเขียวเท่ากับเท่าใด

19. โยนเหรียญบาท (เที่ยงตรง) หนึ่งเหรียญจำนวน 10 ครั้ง ความน่าจะเป็นที่จะได้หัวอย่างน้อย 2 ครั้งติดกันเท่ากับข้อใดต่อไปนี้

20. ความน่าจะเป็นที่สมศักดิ์จะสอบผ่านวิชาคณิตศาสตร์และวิชาเคมีเป็น 2/3 และ 4/9 ตามลำดับ ถ้าความน่าจะเป็นที่เขาสอบผ่านทั้งสองวิชานี้คือ 1/4 แล้วความน่าจะเป็นที่เขาจะสอบไม่ผ่านทั้งสองวิชาเท่ากับข้อใดต่อไปนี้

21. ให้ปริภูมิตัวอย่างเป็นเซตจำกัด ถ้า A และ B เป็นเหตุการณ์ ถ้า $B \subseteq A$ จงพิสูจน์ว่า

$$P(B - A) = P(B) - P(A)$$

22. ให้ A เป็นเหตุการณ์โดยที่ $P(A) = 3P(A^c)$ จงหา $P(A)$

3.4 ความน่าจะเป็นแบบมีเงื่อนไข

ในการสุ่มหยิบสินค้ามา 2 ชิ้น โดยหยิบที่ละชิ้น จากทั้งหมด 100 ชิ้น ในจำนวนนี้มี 20 ชิ้นที่บกพร่อง

A แทนเหตุการณ์ที่หยิบได้ชิ้นแรกพบข้อบกพร่อง

และ B แทนเหตุการณ์ที่หยิบได้ชิ้นสองพบข้อบกพร่อง

จงหา $P(A)$ และ $P(B)$ เมื่อ

1. หยิบแล้วใส่คืน (sampling with replacement)
   จะได้ว่า $P(A) = \dfrac{20}{100}$ และ $P(B) = \dfrac{20}{100}$ เห็นได้ว่าเหตุการณ์ A ไม่มีผลกับเหตุการณ์ B
2. หยิบแล้วไม่ใส่คืน (sampling without replacement)
   จะได้ว่า $P(A) = \dfrac{20}{100}$ แต่ $P(B)$ นั้นขึ้นกับเหตุการณ์ A นั่นคือ
   ถ้าครั้งแรกหยิบได้ชิ้นไม่บกพร่อง จะได้ว่า $P(B) = \dfrac{20}{99}$
   ถ้าครั้งแรกหยิบได้ชิ้นบกพร่อง จะได้ว่า $P(B) = \dfrac{19}{99}$ หมายถึงความน่าจะเป็นที่จะหยิบได้ชิ้นสองพบข้อบกพร่องเมื่อหยิบชิ้นแรกพบข้อบกพร่อง เขียนแทนด้วย $P(B \mid A)$

ให้ A และ B เป็นเหตุการณ์สองเหตุการณ์ ความน่าจะเป็นของเหตุการณ์ B ที่เกิดขึ้นเมื่อเหตุการณ์ A เกิดขึ้น เรียกว่า **ความน่าจะเป็นแบบมีเงื่อนไข** (conditional probability) เขียนแทนด้วย

$$P(B \mid A)$$

อ่านว่า ความน่าจะเป็นที่ B จะเกิดขึ้น เมื่อ A เกิดขึ้นแล้ว

ตัวอย่าง 3.4.1 ในการทอดลูกเต๋า 2 ลูก กำหนดให้ A แทนเหตุการณ์ที่ได้แต้มลูกที่หนึ่งมากกว่าลูกที่สอง และ B แทนเหตุการณ์ผลรวมของแต้มเท่ากับ 10 จงหา

1. $P(A \mid B)$
2. $P(B \mid A)$

แนวคำตอบ จะได้ว่า

$$\begin{aligned} A &= \{(2,1),(3,1),(3,2),(4,1),(4,2),(4,3),(5,1),(5,2),(5,3),(5,4) \\ &\quad (6,1),(6,2),(6,3),(6,4),(6,5)\} \\ B &= \{(4,6),(6,4),(5,5)\} \end{aligned}$$

1. $P(A \mid B)$ หมายถึงความน่าจะเป็นที่ได้แต้มลูกที่หนึ่งมากกว่าลูกที่สอง เมื่อผลรวมของแต้มเท่ากับ 10 เกิดขึ้นแล้ว นั่นคือ $(6,4)$ เกิดขึ้นในเหตุการณ์ B ดังนั้น
   $$P(A \mid B) = \frac{1}{3}$$
2. $P(B \mid A)$ หมายถึงความน่าจะเป็นที่ผลรวมของแต้มเท่ากับ 10 เมื่อได้แต้มลูกที่หนึ่งมากกว่าลูกที่สองเกิดขึ้นแล้ว นั่นคือ $(6,4)$ เกิดขึ้นในเหตุการณ์ A ดังนั้น
   $$P(B \mid A) = \frac{1}{15}$$

จากตัวอย่าง 3.4.1 จะเห็นว่า $P(B \mid A)$ หมายถึงความน่าจะเป็นของเหตุการณ์ของ $A \cap B$ ใน A นั่นคือ

$$P(B \mid A) = \frac{n(A \cap B)}{n(A)} = \frac{P(A \cap B)}{P(A)}$$

โดยที่ $A \neq \varnothing$

บทนิยาม 3.4.2 ความน่าจะเป็นของเหตุการณ์ B ที่เกิดขึ้นเมื่อเหตุการณ์ A เกิดขึ้น เขียนแทนด้วย $P(B \mid A)$ คือ

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

เมื่อ $P(A) \neq 0$

ตัวอย่าง 3.4.3 จากการสำรวจจำนวนอุบัติเหตุที่เกิดขึ้นของผู้ขับขี่รถยนต์ในช่วงเวลา 3 ปี จำนวน 1000 คน แสดงได้ดังตาราง

อายุ (ปี)	18-25	26-39	40-45	มากกว่า 55	รวม
0-1 ครั้ง	100	150	250	75	575
2-3 ครั้ง	150	25	125	25	325
มากกว่า 3 ครั้ง	50	25	25	0	100
รวม	300	200	400	100	1000

เมื่อสุ่มผู้ขับขี่จากกลุ่มนี้มา 1 คน

1. พบว่าเกิดอุบัติเหตุมากกว่า 3 ครั้ง จงหาความน่าจะเป็นที่เขาจะมีอายุ 26-39 ปี
   แนวคำตอบ จะได้ว่า

$$\begin{aligned} P(\text{อายุ 26-39 ปี} \mid \text{อุบัติเหตุ} > 3 \text{ ครั้ง}) &= \frac{P(\text{อายุ 26-39 ปี และอุบัติเหตุ} > 3 \text{ ครั้ง})}{P(\text{อุบัติเหตุ} > 3 \text{ ครั้ง})} \\ &= \frac{25/1000}{100/1000} = \frac{1}{4} \end{aligned}$$

2. พบว่ามีอายุ 40-55 ปี จงหาความน่าจะเกิดอุบัติเหตุไม่เกิน 1 ครั้ง
   แนวคำตอบ จะได้ว่า

$$\begin{aligned} P(\text{อุบัติเหตุไม่เกิน 1 ครั้ง} \mid \text{อายุ 40-55 ปี}) &= \frac{P(\text{อุบัติเหตุไม่เกิน 1 ครั้ง และอายุ 40-55 ปี})}{P(\text{อายุ 40-55 ปี})} \\ &= \frac{250/1000}{400/1000} = \frac{5}{8} \end{aligned}$$

3. จงหาความน่าจะเกิดอุบัติเหตุ 2-3 ครั้ง โดยที่เขามีอายุมากกว่า 55 ปี
   แนวคำตอบ จะได้ว่า

$$\begin{aligned} P(\text{อุบัติเหตุ 2-3 ครั้ง} \mid \text{อายุ} > 55 \text{ ปี}) &= \frac{P(\text{อุบัติเหตุ 2-3 ครั้ง และอายุ} > 55 \text{ ปี})}{P(\text{อายุ} > 55 \text{ ปี})} \\ &= \frac{25/1000}{100/1000} = \frac{1}{4} \end{aligned}$$

ทฤษฎีบท 3.4.4 ทฤษฎีบทการคูณความน่าจะเป็น (Multiplication Theorem of Probability) ในการทดลองครั้งหนึ่ง ถ้าเหตุการณ์ A และ B ที่เกิดขึ้นทั้งสองเหตุการณ์ จะได้ว่า

$$P(A \cap B) = P(A)P(B \mid A) \qquad \text{เมื่อ } P(A) \neq 0$$
$$P(A \cap B) = P(B)P(A \mid B) \qquad \text{เมื่อ } P(B) \neq 0$$

บทพิสูจน์. เห็นได้ชัดจากบทนิยามของความน่าจะเป็นแบบมีเงื่อนไข □

ตัวอย่าง 3.4.5 ในการสุ่มหยิบสินค้ามา 2 ชิ้น เมื่อหยิบที่ละชิ้นโดยไม่ใส่คืน จากทั้งหมด 100 ชิ้น ในจำนวนนี้มี 20 ชิ้นที่บกพร่อง จงหาความน่าจะเป็นที่หยิบได้สินค้าข้อบกพร่องทั้งสองชิ้น โดยใช้ทฤษฎีบทการคูณความน่าจะเป็น

แนวคำตอบ A แทนเหตุการณ์ที่หยิบได้ชิ้นแรกพบข้อบกพร่อง
และ B แทนเหตุการณ์ที่หยิบได้ชิ้นสองพบข้อบกพร่อง
จะได้ว่า $P(A) = \frac{20}{100}$ และ $P(B \mid A) = \frac{19}{99}$ ดังนั้น

$$P(A \cap B) = P(A)P(B \mid A) = \frac{20}{100} \cdot \frac{19}{99} = \frac{19}{495}$$

จากตัวอย่างเริ่มต้นที่กล่าวถึงการสุ่มหยิบสินค้ามา 2 ชิ้น โดยหยิบที่ละชิ้นเมื่อหยิบแล้วใส่คืน จากทั้งหมด 100 ชิ้น ในจำนวนนี้มี 20 ชิ้นที่บกพร่อง
A แทนเหตุการณ์ที่หยิบได้ชิ้นแรกพบข้อบกพร่อง
และ B แทนเหตุการณ์ที่หยิบได้ชิ้นสองพบข้อบกพร่อง
จะได้ว่า $P(B \mid A) = \dfrac{20}{100} = P(B)$ เนื่องจากเหตุการณ์ A ไม่มีผลกับเหตุการณ์ B เรียกว่าเหตุการณ์ B เป็นอิสระจากเหตุการณ์ A จึงนิยามดังนี้

บทนิยาม 3.4.6 เหตุการณ์ B เป็น **อิสระ (independent)** จากเหตุการณ์ A ก็ต่อเมื่อ

$$P(B \mid A) = P(B)$$

ทฤษฎีบท 3.4.7 เหตุการณ์ B เป็นอิสระจากเหตุการณ์ A ก็ต่อเมื่อ

$$P(A \cap B) = P(A)P(B)$$

บทพิสูจน์. โดยทฤษฎีบทการคูณความน่าจะเป็นและบทนิยามเหตุการณ์ที่เป็นอิสระต่อกัน □

ตัวอย่าง 3.4.8 ดึงไพ่ 3 ใบจากสำรับซึ่งมี 52 ใบ โดยดึงที่ละใบ จงหาความน่าจะเป็นที่จะได้ A ทั้ง 3 ใบ เมื่อ

1. ดึงแล้วใส่คืนสำรับก่อนดึงไพ่ใบต่อไป
2. ดึงแล้วไม่ใส่คืนสำรับก่อนดึงไพ่ใบต่อไป

แนวคำตอบ E แทนเหตุการณ์ที่หยิบได้ A ทั้ง 3 ใบ
และ E_i แทนเหตุการณ์ที่หยิบได้ A ในครั้งที่ i เมื่อ $i = 1, 2, 3$
จะได้ว่า $E = E_1 \cap E_2 \cap E_3$ และ $P(E_i) = \dfrac{4}{52}$

1. ในกรณีแต่เหตุการณ์ E_i เป็นอิสระจากกัน ดังนั้น

$$P(E) = P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3) = \frac{4}{52} \cdot \frac{4}{52} \cdot \frac{4}{52} = \frac{1}{2197}$$

2. โดยใช้ทฤษฎีบทการคูณความน่าจะเป็น

$$\begin{aligned} P(E) &= P((E_1 \cap E_2) \cap E_3) \\ &= P(E_1 \cap E_2)P(E_3 \mid E_1 \cap E_2) \\ &= P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1 \cap E_2) \\ &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} = \frac{1}{5525} \end{aligned}$$

ตัวอย่าง 3.4.9 ในการสอบถามวันเกิดของนักเรียน 2 คน จงหาความน่าจะเป็นที่จะเกิดวันเดียวกัน

แนวคำตอบ ให้ E เป็นเหตุการณ์ที่นักเรียน 2 คนเกิดวันเดียวกัน
พิจารณากรณีที่เกิดวันจันทร์เหมือนกัน เนื่องจากเหตุการณ์ที่นักเรียนทั้ง 2 จะเกิดเป็นอิสระจากกัน ดังนั้นความน่าจะเป็นที่นักเรียนทั้ง 2 จะเกิดวันจันทร์เท่ากับ

$$\frac{1}{7} \cdot \frac{1}{7} = \frac{1}{49}$$

เมื่อรวมทั้ง 7 วัน จึงสรุปได้ว่า

$$P(E) = 7 \cdot \frac{1}{49} = \frac{1}{7}$$

นั่นคือความน่าจะเป็นที่จะเกิดวันเดียวกันเท่ากับ $\frac{1}{7}$

บทนิยาม 3.4.10 จะเรียก $\{B_1, B_2, ..., B_n\}$ ว่าเป็น **ผลแบ่งกั้น** (partition) ของเซต S ก็ต่อเมื่อ

1. $\displaystyle\bigcup_{i=1}^{n} B_i = B_1 \cup B_2 \cup ... \cup B_n = S$
2. $B_i \cap B_j = \varnothing$ ทุก ๆ $i \neq j$

ทฤษฎีบท 3.4.11 ให้ $\{B_1, B_2, ..., B_n\}$ เป็นเซตของเหตุการณ์ซึ่งเป็นผลแบ่งกั้นของปริภูมิตัวอย่าง S และ $P(B_i) \neq 0$ ทุก ๆ ค่าของ $i = 1, 2, ..., n$ ถ้า A เป็นเหตุการณ์หนึ่ง จะได้ว่า

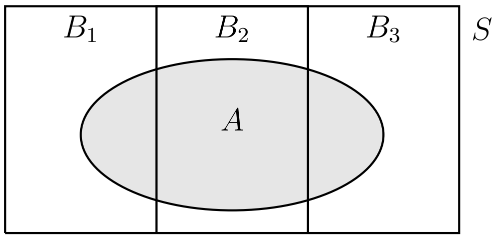
$$P(A) = \sum_{i=1}^{n} P(B_i)P(A \mid B_i)$$

บทพิสูจน์. เนื่องจาก $A = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_n)$
และ $(A \cap B_1), (A \cap B_2), ..., (A \cap B_n)$ เป็นเหตุการณ์ที่ไม่เกิดร่วมกันเป็นคู่ ๆ ดังนั้น

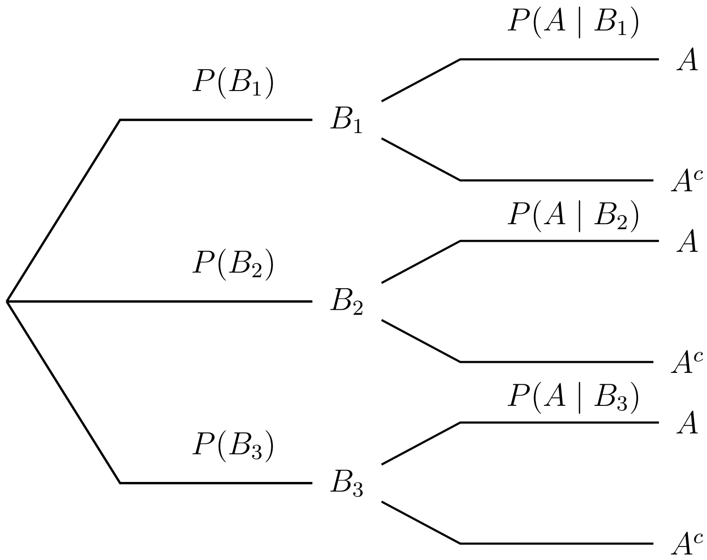
$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n) \\ &= P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + \cdots + P(B_n)P(A \mid B_n) \\ &= \sum_{i=1}^{n} P(B_i)P(A \mid B_i) \end{aligned}$$

□

จากทฤษฎีบท 3.4.11 พิจารณากรณี $n = 3$ แสดงด้วยแผนภาพเวนน์ได้ดังนี้



หรือแสดงด้วยแผนภาพต้นไม้ได้เป็น



จะได้ว่า $P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + P(B_3)P(A \mid B_3)$
ข้อสังเกต $B_1 \cup B_2 \cup B_3 = S$ นั่นคือ $P(B_1) + P(B_2) + P(B_3) = 1$

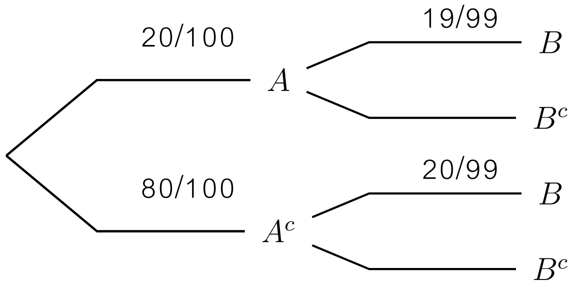
ตัวอย่าง 3.4.12 ในการสุ่มหยิบสินค้ามา 2 ชิ้น เมื่อหยิบที่ละชิ้นโดยไม่ใส่คืน จากทั้งหมด 100 ชิ้น ในจำนวนนี้มี 20 ชิ้นที่บกพร่อง ให้

A แทนเหตุการณ์ที่ของชิ้นแรกมีข้อบกพร่อง

B แทนเหตุการณ์ที่ของชิ้นสองมีข้อบกพร่อง

จงหา $P(B)$ โดยใช้ทฤษฎีบท 3.4.11

แนวคำตอบ เขียนแผนภาพต้นไม้ได้ดังนี้



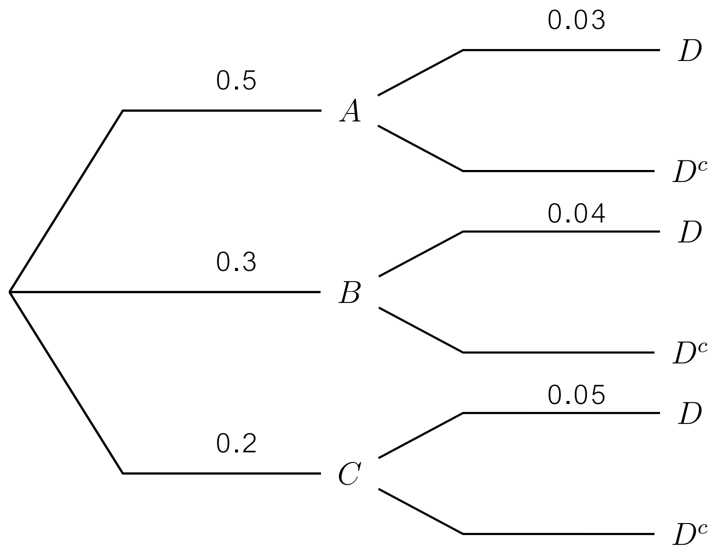
ดังนั้น

$$P(B) = P(A)P(B \mid A) + P(A^c)P(B \mid A^c)$$
$$= \frac{20}{100} \cdot \frac{19}{99} + \frac{80}{100} \cdot \frac{20}{99} = \frac{1}{5}$$

ตัวอย่าง 3.4.13 โรงงานแห่งหนึ่งมีเครื่องจักร 3 เครื่อง A, B และ C ซึ่งสามารถผลิตสินค้าได้ 50%, 30% และ 20% ของปริมาณสินค้าทั้งหมดที่ผลิตจากโรงงานนี้ ตามลำดับ เปอร์เซนต์ของสินค้าที่พบข้อบกพร่องซึ่งผลิตโดยเครื่องจักรทั้งสามคือ 3%, 4 % และ 5% ตามลำดับ ถ้าเลือกสินค้ามาชิ้นหนึ่งโดยวิธีสุ่มแล้วตรวจสภาพ จงหาความน่าจะเป็นที่ของชิ้นนี้จะมีข้อบกพร่อง

แนวคำตอบ A แทนเหตุการณ์ที่สินค้าผลิตจากเครื่องจักร A
B แทนเหตุการณ์ที่สินค้าผลิตจากเครื่องจักร B
C แทนเหตุการณ์ที่สินค้าผลิตจากเครื่องจักร C
D แทนเหตุการณ์ที่ตรวจพบสินค้าบกพร่อง

เขียนแผนภาพต้นไม้ได้ดังนี้



จะได้ว่า

$$
\begin{aligned}
P(D) &= P(A)P(D \mid A) + P(B)P(D \mid B) + P(C)P(D \mid C) \\
&= 0.5(0.03) + 0.3(0.04) + 0.2(0.05) \\
&= 0.037
\end{aligned}
$$

ดังนั้นความน่าจะเป็นที่ของชิ้นนี้จะมีข้อบกพร่องเท่ากับ 0.037

แบบฝึกหัด 3.4

1. ครอบครัวสันติสุข มีบุตร 4 คน จงหาความน่าจะเป็นที่

   1.1 บุตรเป็นชายทั้งหมด

   1.2 มีบุตรสาวอย่างน้อยครึ่งหนึ่ง

2. โยนเหรียญบาท (เที่ยงตรง) หนึ่งเหรียญจำนวน 10 ครั้ง ความน่าจะเป็นที่จะได้หัวอย่างน้อย 2 ครั้งติดกันเท่ากับข้อใดต่อไปนี้

3. จากข้อมูลของบริษัทแห่งหนึ่งที่มีพนักงานจำนวน 500 คน ปรากฏดังนี้

ระดับการศึกษา	ต่ำกว่าปริญญาโท	ปริญญาโทหรือสูงกว่า	รวม
ชาย	130	70	200
หญิง	270	30	300
รวม	400	100	500

   ถ้าสุ่มเลือกพนังงานโรงงานแห่งนี้มา 1 คน

   3.1 พบว่าเป็นผู้หญิง จงหาความน่าจะเป็นที่เขาจะมีระดับการศึกษาต่ำกว่าปริญญาโท

   3.2 พบว่าเป็นผู้ชาย จงหาความน่าจะเป็นที่เขาจะมีระดับการศึกษาต่ำกว่าปริญญาโท

   3.3 พบว่าระดับการศึกษาปริญญาโทหรือสูงกว่า จงหาความน่าจะเป็นที่เขาจะผู้ชาย

4. โรงงานแห่งหนึ่งมีเครื่องจักร 4 เครื่อง A, B, C และ D ซึ่งสามารถผลิตสินค้าได้ 40%, 30%, 20% และ 10% ของปริมาณสินค้าทั้งหมดที่ผลิตจากโรงงานนี้ ตามลำดับ เปอร์เซนต์ของสินค้าที่พบข้อบกพร่องซึ่งผลิตโดยเครื่องจักรทั้งสามคือ 1%, 2 %, 3% และ 4% ตามลำดับ ถ้าเลือกสินค้ามาชิ้นหนึ่งโดยวิธีสุ่มแล้วตรวจสภาพ จงหาความน่าจะเป็นที่ของชิ้นนี้จะมีข้อบกพร่อง

5. จำนวน CD ทั้งหมด 100 แผ่นที่ผลิตมาจากสองเครื่อง มี 20 แผ่นที่ไม่มีคุณภาพ แผ่นที่ผลิตในเครื่องที่หนึ่งมี 5 แผ่นที่ไม่มีคุณภาพ โดยเครื่องที่หนึ่งผลิตได้ 3 เท่าของเครื่องที่สอง ถ้าสุ่มหยิบแผ่น CD มา 1 จงหาความน่าจะเป็นที่จะได้แผ่นที่ไม่มีคุณภาพ

6. เหรียญอันหนึ่งถูกโยนจนกระทั่งขึ้นหัว หรือโดยจนครบสามครั้ง ถ้าในการโยนครั้งแรกไม่ขึ้นหัว จงหาความน่าจะเป็นที่จะโยนครบสามครั้ง

7. มหาวิทยาลัยแห่งหนึ่ง 4% ของนักศึกษาชาย และ 1% ของนักศึกษาหญิง สูงกว่า 165 เซนติเมตร 60% ขอมหาวิทยาลัยแห่งนี้เป็นผู้หญิง ถ้าสุ่มนักศึกษามาหนึ่งคนปรากฏว่าเป็นนักศึกษาที่สูงกว่า 165 เซนติเมตร จงหาความน่าจะเป็นที่นักศึกษาคนนั้นเป็นผู้หญิง

8. One box has 7 red balls and 3 white balls; a second box has 6 red balls and 4 white balls. A pair of dice are tossed. If the sum of the dice are less than five, a ball is selected from the first box, otherwise the ball is selected from the second box. Find the probability of getting a red ball.

3.5 กฎของเบย์

ทฤษฎีบท 3.5.1 กฎของเบย์ (Bay's Rule)
ให้ $\{B_1, B_2, ..., B_n\}$ เป็นเซตของเหตุการณ์ซึ่งเป็นผลแบ่งกั้นของปริภูมิตัวอย่าง S และ $P(B_i) \neq 0$ ทุก ๆ ค่าของ $i = 1, 2, ..., n$ ถ้า A เป็นเหตุการณ์หนึ่ง เมื่อ $P(A) \neq 0$ จะได้ว่า

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)}$$

โดยที่ $k = 1, 2, ..., n$

บทพิสูจน์. โดยใช้ทฤษฎีบทการคูณความน่าจะเป็น และทฤษฎีบท 3.4.11 จะได้ว่า

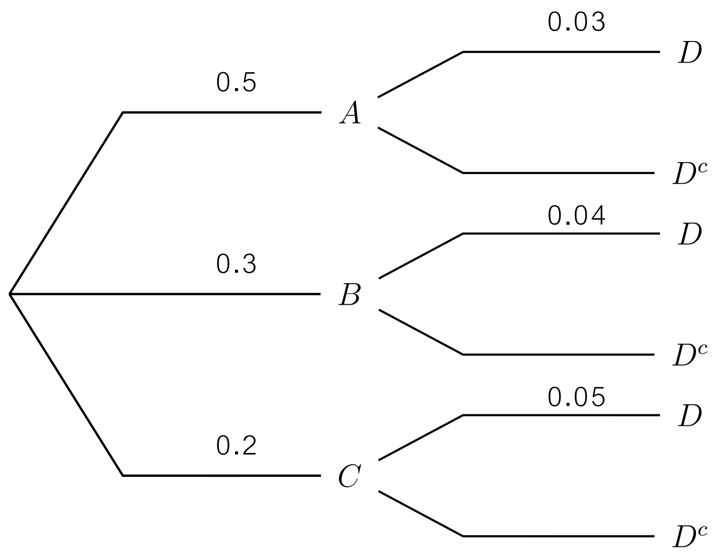
$$\begin{aligned} P(B_k \mid A) &= \frac{P(B_k \cap A)}{P(A)} \\ &= \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)} \end{aligned}$$

□

ตัวอย่าง 3.5.2 โรงงานแห่งหนึ่งมีเครื่องจักร 3 เครื่อง A, B และ C ซึ่งสามารถผลิตสินค้าได้ 50%, 30% และ 20% ของปริมาณสินค้าทั้งหมดที่ผลิตจากโรงงานนี้ ตามลำดับ เปอร์เซนต์ของสินค้าที่พบข้อบกพร่องซึ่งผลิตโดยเครื่องจักรทั้งสามคือ 3%, 4 % และ 5% ตามลำดับ ถ้าเลือกสินค้ามาชิ้นหนึ่งโดยวิธีสุ่มแล้วตรวจสภาพ และพบว่าสินค้าชิ้นนั้นมีข้อบกพร่อง จงหาความน่าจะเป็นที่ของชิ้นนั้นผลิตจากเครื่องจักร A

แนวคำตอบ A แทนเหตุการณ์ที่สินค้าผลิตจากเครื่องจักร A
B แทนเหตุการณ์ที่สินค้าผลิตจากเครื่องจักร B
C แทนเหตุการณ์ที่สินค้าผลิตจากเครื่องจักร C
D แทนเหตุการณ์ที่ตรวจพบสินค้าบกพร่อง

เขียนแผนภาพต้นไม้ได้ดังนี้



โดยกฎของเบย์จะได้ว่า

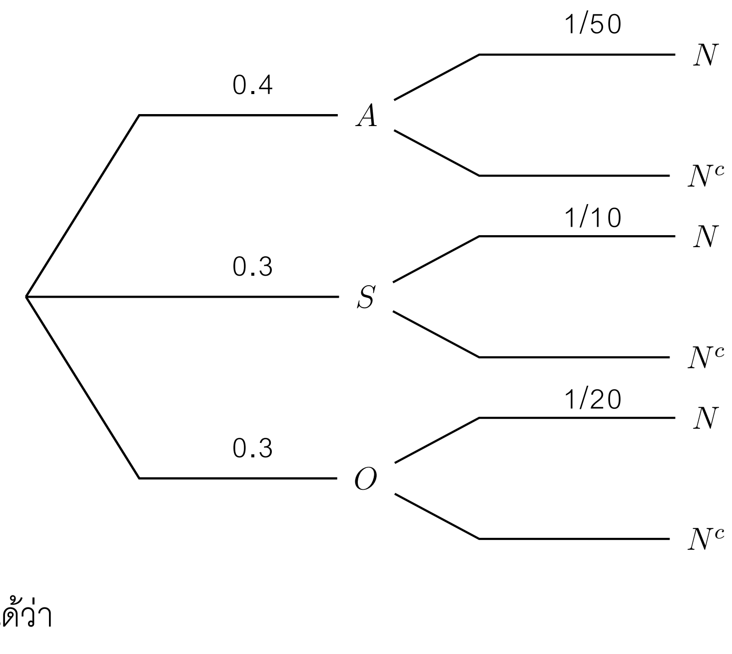
$$\begin{aligned}P(A \mid D) &= \frac{P(A)P(D \mid A)}{P(A)P(D \mid A) + P(B)P(D \mid B) + P(C)P(D \mid C)} \\ &= \frac{0.5(0.03)}{0.5(0.03) + 0.3(0.04) + 0.2(0.05)} \\ &= \frac{15}{37} = 0.4054\end{aligned}$$

ดังนั้นถ้าพบว่าสินค้าชิ้นนั้นมีข้อบกพร่อง ความน่าจะเป็นที่ของชิ้นนั้นผลิตจากเครื่องจักร A เท่ากับ 0.4054

ตัวอย่าง 3.5.3 ในเทศกาลปีใหม่ร้านค้าแห่งหนึ่งจัดผู้ห่อของขวัญไว้ 3 คน คือ สิงหา กันยา และ ตุลา สิงหาห่อของขวัญ 40% ของของขวัญทั้งหมด และลืมเอาป้ายราคาออกก่อนห่อของ 1 ใน 50 ครั้ง กันยาห่อของขวัญ 30% ของของขวัญทั้งหมด และลืมเอาป้ายราคาออกก่อนห่อของ 1 ใน 10 ครั้ง ตุลาห่อของขวัญที่เหลือและลืมเอาป้ายราคาออกก่อนห่อของ 1 ใน 20 ครั้ง สมมติมีลูกค้าคนหนึ่งมาต่อว่าร้านไม่เอาป้ายติดราคาออกก่อนห่อของขวัญ จงหาความน่าจะเป็นที่ของขวัญชิ้นนั้นจะถูกห่อโดยสิงหา

แนวคำตอบ A แทนเหตุการณ์ที่ของขวัญถูกห่อโดย สิงหา
S แทนเหตุการณ์ที่ของขวัญถูกห่อโดย กันยา
O แทนเหตุการณ์ที่ของขวัญถูกห่อโดย ตุลา
N แทนเหตุการณ์ที่ลืมเอาป้ายราคาออก

เขียนแผนภาพต้นไม้ได้ดังนี้



โดยกฎของเบย์จะได้ว่า

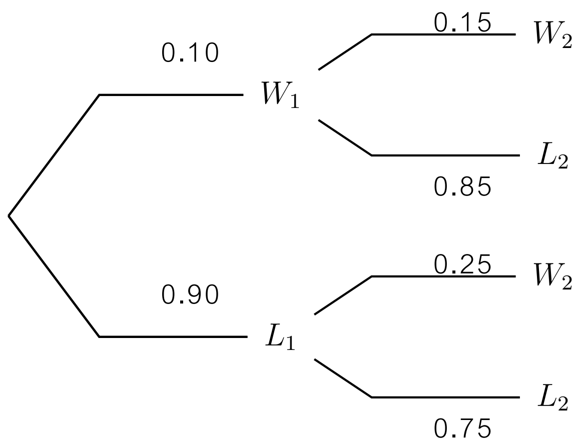
$$\begin{aligned}P(A \mid N) &= \frac{P(A)P(N \mid A)}{P(A)P(N \mid A) + P(S)P(N \mid S) + P(O)P(N \mid O)} \\ &= \frac{0.4(1/50)}{0.4(1/50) + 0.3(1/10) + 0.3(1/20)} \\ &= \frac{8}{53} = 0.1509\end{aligned}$$

ดังนั้นถ้าลูกค้าคนหนึ่งมาต่อว่าร้านไม่เอาป้ายติดราคาออกก่อนห่อของขวัญ ความน่าจะเป็นที่ของขวัญชิ้นนั้นจะถูกห่อโดยสิงหา เท่ากับ 0.1509

ตัวอย่าง 3.5.4 A basketball team is to play two games in a tournament. The probability of winning the first game is 0.10. If the first game is won, the probability of winning the second game is 0.15. If the first game is lost, the probability of winning the second game is 0.25. What is the probability the first game was won if the second game is lost ?

แนวคำตอบ W_1 แทนเหตุการณ์ที่ชนะเกมที่ 1
L_1 แทนเหตุการณ์ที่แพ้เกมที่ 1
W_2 แทนเหตุการณ์ที่ชนะเกมที่ 2
L_2 แทนเหตุการณ์ที่แพ้เกมที่ 2

เขียนแผนภาพต้นไม้ได้ดังนี้



โดยกฎของเบย์จะได้ว่า

$$\begin{aligned} P(W_1 \mid L_2) &= \frac{P(W_1)P(L_2 \mid W_1)}{P(W_1)P(L_2 \mid W_1) + P(L_1)P(L_2 \mid L_1)} \\ &= \frac{0.1(0.85)}{0.1(0.85) + 0.90(0.75)} \\ &= \frac{17}{152} = 0.1118 \end{aligned}$$

ดังนั้นความความน่าจะเป็นที่จะชนะเกมที่ 1 ถ้าแพ้เกมที่ 2 เท่ากับ 0.1118

แบบฝึกหัด 3.5

1. นักเรียนห้องที่ 1, 2, 3 และ4 มีจำนวน 50 คน , 50 คน , 40 คน และ 60 คนตามลำดับใน แต่ละห้องมีนักเรียนที่มีเกรดเฉลี่ยเกิน 3.5 อยู่ 30% , 20% , 10% และ 5% ตามลำดับ หาก สุ่มนักเรียนมา 1 คน แล้วพบว่านักเรียนคนนั้นมีเกรดเฉลี่ยเกิน 3.5 จงหา

   1.1 ความน่าจะเป็นที่จะเป็นนักเรียนจากห้องที่ 1

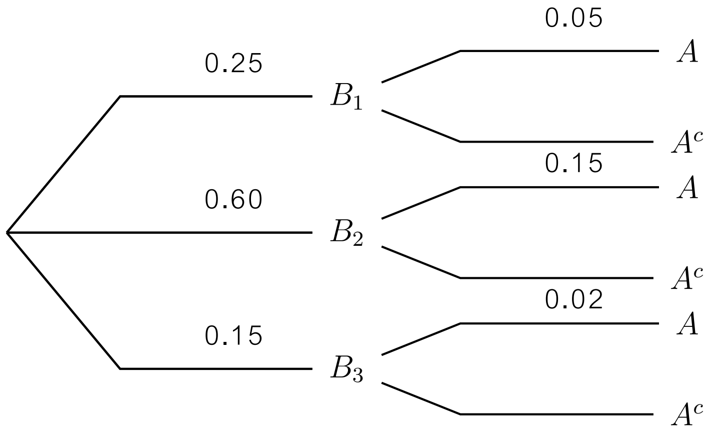
   1.2 ความน่าจะเป็นที่จะเป็นนักเรียนจากห้องที่ 2

   1.3 ความน่าจะเป็นที่จะเป็นนักเรียนจากห้องที่ 3

   1.4 ความน่าจะเป็นที่จะเป็นนักเรียนจากห้องที่ 4

2. บริษัทผลิตขวดน้ำส่งออก มีเครื่องจักร 4 เครื่อง โดยเครื่องที่ 1, 2 , 3 และ 4 ผลิตขวดน้ำได้ 25% , 45% , 20% และ 10% ตามลำดับ จากข้อมูลการผลิตพบว่าแต่ละเครื่องจะมีขวดน้ำที่ ผลิตไม่ได้คุณภาพ ซึ่งเครื่องที่ 1, 2, 3 และ 4 มีขวดน้ำที่ผลิตแล้วไม่ได้คุณภาพอยู่ 1%, 3%, 2% และ 1% ตามลำดับ หากสุ่มหยิบขวดน้ำมา 1 ขวด พบว่าเป็นชิ้นที่ได้คุณภาพ จงหาความ น่าจะเป็นที่ขวดน้ำดังกล่าวจะผลิตด้วยเครื่องที่ 3

3. จากแผนภาพต้นไม้แสดงความน่าจะเป็นดังต่อไปนี้



   จงหา

3.1 $P(A)$	3.4 $P(B_2 \mid A)$	3.7 $P(A \cup B_1)$
3.2 $P(A^c)$	3.5 $P(A^c \mid B_3)$	3.8 $P(A \cap (B_1 \cup B_2))$
3.3 $P(B_1 \mid A)$	3.6 $P(B_3 \mid A^c)$	3.9 $P(B_3 - A)$

4. โรงงานผลิตชาเขียวพร้อมดื่มมีโรงงานผลิต 3 โรงงาน โดยโรงงานที่ 1, 2 และ3 ผลิตสินค้าได้ 10%, 30% และ 60% ของปริมาณการผลิตทั้งหมดของโรงงาน ในการผลิตชาเขียวจะมีบาง ขวดที่มีปริมาตรบรรจุไม่เป็นไปตามที่ต้องการ ซึ่งโรงงานที่ 1 , 2 และ 3 มีจำนวนชาเขียวที่มี ปริมาตรบรรจุไม่เป็นไปตามที่ต้องการอยู่ 2% , 4% และ 1% ตามลำดับ หากสินค้าทั้งหมดที่ ผลิตได้ถูกเก็บรวมไว้ในโกดังสินค้า ผู้จัดการทั่วไปสุ่มหยิบชาเขียวมา 1 ขวดจากโกดงพบว่า มีปริมาตรบรรจุที่ไม่เป็นไปตามที่ต้องการ จงหาความน่าจะเป็นที่ชาเขียวขวดนั้นจะผลิตจาก โรงงานที่ 2

สรุป

ในบทนี้จะกล่าวถึงความน่าจะเป็นของเหตุการณ์ ซึ่งเริ่มต้นด้วยปริภูมิตัวอย่าง ซึ่งหมายถึงเซตของผลลัพธ์ที่เกิดขึ้นจากการทดลองโดยสมาชิกเรียกว่าจุดตัวอย่าง เซตย่อยของปริภูมิตัวอย่างเรียกว่าเหตุการณ์ จากนั้นกล่าวถึงการนับจุดตัวอย่างที่ใช้หลักการคูณ และหลักการบวก วิธีเรียงสับเปลี่ยนแบบเส้นตรง และแบบวงกลม ในแบบที่ซ้ำกันและไม่ซ้ำกัน การจัดหมู่และการแบ่งกลุ่ม จากนั้นกล่าวถึงนิยามความน่าจะเป็นของเหตุการณ์คือผลรวมของน้ำหนักของจุดตัวอย่าง ทำให้ได้ว่า $0 \leq P(E) \leq 1$ สุดท้ายกล่าวถึงความน่าจะเป็นแบบมีเงื่อนไข นำไปสู่การนิยามเหตุการณ์ที่อิสระต่อกัน และกฎของเบย์ซึ่งเป็นผลที่เกิดจากทฤษฎีบทเกี่ยวกับความน่าจะเป็นแบบมีเงื่อนไขโดยใช้ผลแบ่งกั้น

แบบฝึกหัดบทที่ 3

1. มีกิ่งไม้ 3 กิ่ง มีนก 6 ตัว จะมีวิธีเกาะกิ่งไม้ของนกทั้งหมดกี่วิธี
2. มีหนังสือ 10 เล่มที่แตกต่างกัน นำไปแจกให้เด็ก 4 คนได้กี่วิธี
3. จำนวนเต็มเต็มที่อยู่ระหว่าง 6,500 – 9,999 และเลขในแต่ละหลักไม่ซ้ำกัน มีทั้งหมดกี่จำนวน
4. สร้างจำนวนเต็มบวก 5 หลัก จาก 0 – 9 โดยไม่ใช้หลักซ้ำกัน และสลับเลขคู่กับเลขคี่ได้กี่จำนวน
5. จงหาจำนวนวิธีทั้งหมดในการจัดชาย 3 คนและหญิง 3 คน ซึ่งมีนาย ก. และนางสาว ข. รวมอยู่ด้วย ให้ยืนเป็นแถวตรง 2 แถว ๆ ละ 3 คน โดยที่ นาย ก. และนางสาว ข. ไม่ได้ยืนติดกันในแถวเดียวกัน
6. สร้างเลข 4 หลักโดยใช้เลข 6 – 9 และผลรวมของตัวเลขแต่ละหลักมีค่าเท่ากับ 34 จะมีวิธีการสร้างเท่ากับข้อใดต่อไปนี้
7. จงหาจำนวนวิธีสลับอักษรในคำว่า HYPERBOLA โดยให้ขึ้นต้นด้วยสระและลงท้ายด้วยพยัญชนะ
8. มีตำแหน่งว่างอยู่ 5 ตำแหน่งต่างกัน เป็นตำแหน่งเฉพาะชาย 3 ตำแหน่ง และเป็นตำแหน่งเฉพาะหญิง 2 ตำแหน่ง มีผู้สมัครงาน 10 คน เป็นชาย 6 คน และหญิง 4 คน จะมีกี่วิธีในการเลือกคนเข้าทำงานทั้ง 5 ตำแหน่งนี้
9. สามีภรรยา 4 คู่นั่งชมภาพยนต์บนเก้าอี้แถวหนึ่ง ซึ่งมี 9 ตัว โดยที่สามีภรรยาทุกคู่ต้องนั่งติดกันเสมอทำได้กี่วิธี
10. ต้องการเรียงเลข 3 หลักที่มีค่ามากกว่า 300 จากเลข 0, 1, 2, 3, 4, 5 โดยตัวเลขเหล่านั้นนำมาใช้ได้เพียงครั้งเดียว มีกี่จำนวน
11. มีชาย 7 คน หญิง 3 คน จงหาจำนวนวิธีในการจัดแถวคนทั้งหมดตามเงื่อนไขต่อไปนี้

    11.1 หญิงทั้ง 3 คนยืนติดกัน

11.2 หญิงทั้ง 3 คนยืนแยกกันหมดและผู้ที่ยืนในตำแหน่งหัวและท้ายเป็นผู้ชาย

12. ต้องการสลับอักษรคำว่า ENGINEER จะสลับได้กี่วิธี ถ้า

    12.1 ไม่มีเงื่อนไข

    12.2 สระอยู่ติดกันหมด

    12.3 พยัญชนะอยู่แยกกันหมด

    12.4 ขึ้นต้นและลงท้ายด้วยอักษรเดียวกัน

13. ในการประชุมตัวแทนจากประเทศต่างๆในเอเชียครั้งหนึ่งมีตัวแทนจากประเทศไทย 3 คน จากประเทศมาเลเซีย 4 คน จากประเทศฟิลิปปิน 3 คน และจากอินโดนิเซีย 5 คน จะจัดให้นั่งรอบโต๊ะกลมได้กี่วิธีโดยที่ตัวแทนจากประเทศไทยต้องนั่งแยกกัน

14. จะจัดคน 6 คนนั่งรอบโต๊ะกลมได้กี่วิธี ถ้า

    14.1 ไม่มีเงื่อนไข

    14.2 มีคน 2 คนนั่งติดกันเสมอ

    14.3 มีคน 2 คนนั่งแยกกันเสมอ

    14.4 มีคน 2 คนนั่งต้องข้ามกันเสมอ

15. จังหวัดแห่งหนึ่งมีอำเภอ 6 อำเภอแต่ละอำเภอส่งผู้แทนอำเภอละ 2 คนเป็นชาย 1 คนและเป็นหญิง 1 คน ถ้าต้องการคัดเลือกกรรมการ 4 คน เป็นชาย 2 คนและหญิง 2 คน จากตัวแทนทั้ง 12 คน โดยในบรรดากรรมการ 4 คนนี้ต้องเป็นชายและหญิงอย่างน้อย 1 คู่ มาจากอำเภอเดียวกันจะมีวิธีการคัดเลือกกี่วิธี

16. มีข้อสอบปรนัย 20 ข้อ คะแนนเต็ม 50 คะแนน โดยกำหนดข้อ 1 – 10 ข้อละ 4 คะแนน และข้อ 11 – 20 ข้อละ 1 คะแนน ถ้าหากนักเรียนตอบข้อใดถูกต้องจะได้คะแนนเต็มของข้อนั้น แต่ถ้าตอบผิดหรือไม่ตอบจะได้คะแนนเป็น 0 คะแนน จะมีกี่วิธีที่นักเรียนคนหนึ่งจะทำข้อสอบได้คะแนนรวม 45 คะแนน

17. มีคน 10 คน มีนาย ก นาย ข และนาย ค รวมอยู่ด้วย แบ่งเข้าห้องพัก 3 ห้อง แต่ละห้องพักได้ 3, 3, 4 ตามลำดับ จะจัดได้กี่วิธี

    17.1 ไม่มีเงื่อนไข

    17.2 นาย ก ข ค พักห้องเดียวกัน

    17.3 นาย ก ข ค พักคนละห้อง

    17.4 นาย ก และนาย ข พักห้องเดียวกัน แต่ไม่พักห้องเดียวกับ นาย ค

18. แบ่งดินสอสี 10 สีแตกต่างกัน ออกเป็น 4 กลุ่ม กลุ่มละ 3 แท่ง 2 กลุ่ม กลุ่มละ 2 แท่ง 2 กลุ่ม จะแบ่งได้กี่วิธี

19. จะจัดพนักงาน 6 คนแบ่งเป็น 3 กลุ่มไปทำงาน 3 งานซึ่งแตกต่างกัน โดยทุกงานต้องมีคนทำงานอย่างน้อย 1 คน และพนักงาน 6 คน ต้องไม่มีใครว่างงาน

20. ในการสัมภาษณ์ผู้สมัครเข้าทำงานในสำนักงานแห่งหนึ่งมีผู้สมัครชาย 5 คน เป็นหญิง 5 คน ถ้าผู้สัมภาษณ์ตัดสินใจเรียกผู้สมัครมาสัมภาษณ์เพียง 5 คน โดยเลือกชาย 3 คนและหญิง 2 คน จากผู้สมัครโดยการสุ่ม การจัดลำดับเข้าสัมภาษณ์ทีละคนโดยให้เป็นผู้สมัครชายเข้าห้องสอบติดต่อกันจะมีจำนวนวิธีเท่าใด

21. บนเส้นรอบวงของวงกลมหนึ่งมีจุด 8 จุดที่แตกต่างกัน จงหาจำนวนรูปสามเหลี่ยมที่มีจุดยอดเกิดจากจุดดังกล่าว

22. วงกลม 5 วง และเส้นตรง 4 เส้น ตัดกันได้มากที่สุดกี่จุด

23. ให้ $r, n \in \mathbb{N}$ โดยที่ $r \leq n$ จงพิสูจน์เอกลักษณ์ต่อไปนี้

    23.1 ${}^{n}P_r = n \cdot {}^{n-1}P_{r-1}$

    23.2 ${}^{n}P_r = (n-r+1) \cdot {}^{n}P_{r-1}$

    23.3 ${}^{n+1}P_r = {}^{n}P_r + r \cdot {}^{n}P_{r-1}$

    23.4 $\binom{n}{r} = \frac{n}{r}\binom{n-1}{r-1}$

    23.5 $\binom{n}{r} = \frac{n-r+1}{r}\binom{n}{r-1}$

24. กล่องใบหนึ่งมีลูกแก้วขนาดเดียวกัน 13 ลูกเป็นสีแดง 6 ลูก สีขาว 4 ลูกและนอกนั้นสีเหลือง ถ้าสุ่มหยิบมา 2 ลูกพร้อมกัน จงหาความน่าจะเป็นที่จะได้ลูกแก้วสีต่างกัน

25. มีหนังสือที่แตกต่างกัน 5 เล่มคือ หนังสือ ก หนังสือ ข หนังสือ ค หนังสือ ง และหนังสือ จ สุ่มเลือกหนังสือเหล่านี้มาครั้งละ 3 เล่ม ความน่าจะเป็นที่จะได้หนังสือ ก หรือหนังสือ ข เท่ากับเท่าใด

26. ในการแข่งขันตอบปัญหารายการหนึ่งผู้เข้าแข่งขันแต่ละคนต้องเลือกปัญหา 4 ข้อในแต่ละครั้ง ถ้าในกล่องที่ใส่ปัญหา มีปัญหาอย่างยากให้เลือกอยู่ 8 ข้อและมีปัญหาอย่างง่ายมีให้เลือกอยู่ 6 ข้อ ความน่าจะเป็นที่ผู้เข้าแข่งขันจะเลือกได้ปัญหาอย่างยากและอย่างง่ายอย่างละ 2 ข้อเท่ากับเท่าใด

27. ก, ข, ค ตกลงทำธุรกิจกันทั้ง 3 คนช่วยกันหาสถานที่ที่จะเปิดสำนักงาน ก และ ข เสนอสถานที่ที่จะถูกเลือกเป็นสำนักงานด้วยความน่าจะเป็นเท่ากัน ส่วน ค เสนอสถานที่ที่คาดว่าจะถูกเลือกด้วยความน่าจะเป็น 2 เท่าของ ก ความน่าจะเป็นที่สถานที่ ค เสนอไม่ได้ถูกเลือกเป็นเท่าใด

28. กิตติและสมานกับเพื่อนๆรวม 7 คนไปเที่ยวต่างจังหวัดด้วยกันในการค้างแรมที่มีบ้านพัก 3 หลัง หลังแรกพักได้ 3 คน ส่วนหลังที่สองและสามพักได้หลังละ 2 คน ซึ่งแต่ละหลังมีความแตกต่างกัน พวกเขาจึงตกลง ที่จะจับสลากว่าใครจะได้พักบ้านหลังใด ความน่าจะเป็นที่กิตติและสมานจะได้พักบ้านหลังเดียวกันในหลังที่หนึ่งหรือหลังที่สามเท่ากับเท่าใด

29. ถุงใบหนึ่งบรรจุลูกกวาดรสสตรอเบอรี่ 5 ลูก รสชอคโกเลต 4 ลูก และรสกาแฟและรสมินท์อย่างละ 2 ลูก หาสุ่มหยิบลูกกวาดจากถุงนี้มา 3 ลูก ความน่าจะเป็นที่จะหยิบได้ลูกกวาดต่างรสกันทั้งหมดเท่ากับเท่าใด

30. ให้ S เป็นเซตของจุด 10 จุดบนวงกลมวงหนึ่ง ซึ่งมีสมบัติดังนี้ เมื่อลากเส้นตรงเชื่อมระหว่างจุด 2 จุดใดๆใน S จะมีเพียง 3 เส้นเท่านั้นที่ผ่านจุดศูนย์กลางของวงกลมนี้ ถ้าสร้างรูปสามเหลี่ยมโดยเลือกจุด 3 จุดใน S มาเป็นจุดยอดของสามเหลี่ยม ความน่าจะเป็นที่จะได้รูปสามเหลี่ยมมุมฉาก เท่ากับเท่าใด

31. ให้ปริภูมิตัวอย่างเป็นเซตจำกัด ถ้า $A_1, A_2, ..., A_n$ เป็นเหตุการณ์ที่ไม่เกิดร่วมกันเป็นคู่ ๆ หรือ $A_i \cap A_j$ ทุก ๆ $i \neq j$ จงพิสูจน์ว่า

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

32. จากข้อมูลของโรงเรียนแห่งหนึ่งมีครูจำนวน 200 คน ปรากฏดังนี้

สาระวิชา	คณิตศาสตร์	วิทยาศาสตร์	ภาษาไทย	ภาษาต่างประเทศ	สังคม	อื่น ๆ	รวม
ชาย	25	30	5	8	15	20	103
หญิง	15	20	20	17	10	15	97
รวม	40	50	25	25	25	35	200

ถ้าสุ่มเลือกครูจากโรงงานแห่งนี้มา 1 คน

32.1 พบว่าเป็นผู้ชาย จงหาความน่าจะเป็นที่เขาจะครูสาระวิชาคณิตศาสตร์

32.2 พบว่าเป็นผู้หญิง จงหาความน่าจะเป็นที่เขาจะเป็นครูสาระภาษาต่างประเทศ

32.3 พบว่าเขาเป็นครูสาระสังคม จงหาความน่าจะเป็นที่เขาจะผู้ชาย

32.4 พบว่าเขาเป็นครูสาระวิทยาศาสตร์ จงหาความน่าจะเป็นที่เขาจะผู้หญิง

33. ร้านอาหารวางแผนที่จะซื้อเตาอบ 100 เตา จากผู้ผลิตระบุว่าเตาอบเหล่านี้ผ่านการ ตรวจสอบรายละเอียดเนื่องจากค่าใช้จ่ายในการตรวจสูง เตาอบ 5 เตาถูกเลือกโดยการสุ่มการตรวจ เตาอบ 100 เตาจะถูกซื้อถ้าหนึ่งในห้าเตาที่เลือกไม่บกพร่อง สมมติว่ามี 8 เตาบกพร่องในชุด 100 เตานี้ จงหาความน่าจะเป็นที่ผู้ซื้อจะตัดสินใจซื้อเตาอบทั้งชุดดังกล่าว

34. ในการทดลองสุ่มหยิบกล่อง 1 1 กล่องและหยิบลูกบอลมา 1 ลูก โดยที่

กล่องใบที่ 1 มีลูกบอลสีแดง 1 ลูก สีดำ 3 ลูก และสีขาว 6 ลูก
กล่องใบที่ 2 มีลูกบอลสีแดง 2 ลูก สีดำ 3 ลูก และสีขาว 5 ลูก
กล่องใบที่ 3 มีลูกบอลสีแดง 3 ลูก สีดำ 4 ลูก และสีขาว 3 ลูก
กล่องใบที่ 4 มีลูกบอลสีแดง 2 ลูก สีดำ 2 ลูก และสีขาว 6 ลูก

หากพบว่าได้ลูกบอลเป็นสีดำ จงหาความน่าจะเป็นที่ลูกบอลนั้นจะมาจากกล่องใบที่ 2

35. โรงงานผลิตจอโทรทัศน์ มีเครื่องจักรที่ใช้ผลิต 3เครื่อง คือ เครื่องที่ 1 , 2 และ3 ซึ่งผลิตจอได้ร้อยละ40 ร้อยละ 25 และร้อยละ35 ของกำลังผลิตทั้งหมด เครื่องที่ 1, 2 และ 3 มีจอที่ผลิตแล้วไม่ผ่านมาตรฐานอยู่ร้อยละ 3 ร้อยละ5 และร้อยละ 1 ตามลำดับ หากสุ่มจอโทรทัศน์มา 1 ชิ้น แล้วพบว่าไม่ผ่านมาตรฐาน จงหาความน่าจะเป็นที่จอดังกล่าวจะผลิตด้วยเครื่องที่ 1

36. Suppose we send 30% of our products to company A and 70% of our products to company B. Company A reports that 5% of our products are defective and company B reports that 4% of our products are defective.

    36.1 Find the probability that a product is sent to company A and it is defective.

    36.2 Find the probability that a product is sent to company A and it is not defective.

    36.3 Find the probability that a product is sent to company B and it is defective.

    36.4 Find the probability that a product is sent to company B and it is not defective.

37. A small manufacturing company has rated 75% of its employees as satisfactory (S) and 25% as unsatisfactory (S'). Personnel records show that 80% of the satisfactory workers had previous work experience (E) in the job they are now doing, while 15% of the unsatisfactory workers had no work experience (E') in the job they are now doing. If a person who has had previous work experience is hired, what is the approximate empirical probability that this person will be an unsatisfactory employee ?

38. To evaluate a new test for detecting Hansen's disease, a group of people 5% of which are known to have Hansen's disease are tested. The test finds Hansen's disease among 98% of those with the disease and 3% of those who don't. What is the probability that someone testing positive for Hansen's disease under this new test actually has it ?

39. One urn has 4 red balls and 1 white ball; a second urn has 2 red balls and 3 white balls. A single card is randomly selected from a standard deck. If the card is less than 5 (aces count as 1), a ball is drawn out of the first urn; otherwise a ball is drawn out of the second urn. If the drawn ball is red, what is the probability that it came out of the second urn ?

บทที่ 4

ตัวแปรสุ่ม

4.1 นิยามของตัวแปรสุ่ม

การทดลองสุ่ม (random trial) หมายถึงการกระทำที่ไม่สามารถทราบผลลัพธ์ล่วงหน้าจนกว่าจะได้ทำเสร็จสิ้นไปแล้ว เช่นการโยนเหรียญเราทราบเพียงว่าเกิดขึ้น 2 แบบคือ หัว และก้อย แต่ก่อนโดยไม่สามารถทราบแน่ชัดว่าจะเกิดอะไร สมมติว่าโยนเหรียญ 1 อัน 2 ครั้ง จะได้ปริภูมิตัวอย่างคือ

$$S = \{TT, TH, TH, HH\}$$

ถ้าเราสนใจจำนวนหัวที่เกิดขึ้นในการทดลองครั้งนี้ นั่นคือจำนวนหัวเท่ากับ 0, 1 หรือ 2 ซึ่งเป็นผลที่เป็นไปได้ทั้งหมดที่จะเกิดขึ้น ถ้าให้ X เป็นจำนวนหัวที่เกิดขึ้น เมื่อโยนเหรียญ 1 อัน 2 ครั้ง ค่าของที่เป็นไปได้คือ 0, 1, 2 และจะเรียก X ว่าตัวแปรสุ่ม

- X มีค่าเท่ากับ 0 ก็ต่อเมื่อ ผลการทดลองเป็น TT
- X มีค่าเท่ากับ 1 ก็ต่อเมื่อ ผลการทดลองเป็น TH หรือ HT
- X มีค่าเท่ากับ 2 ก็ต่อเมื่อ ผลการทดลองเป็น HH

บทนิยาม 4.1.1 ตัวแปรสุ่ม (random variable) คือฟังก์ชันที่มีค่าเป็นจำนวนจริง ซึ่งกำหนดโดยแต่ละสมาชิกในปริภูมิตัวอย่าง

เรามักนิยมใช้ตัวแปร X, Y, Z แทนตัวแปรสุ่ม และ x, y, z แทนค่าของตัวแปรสุ่ม

ตัวอย่าง 4.1.2 หยิบลูกบอลอย่างสุ่มที่ละลูก 2 ครั้ง โดยหยิบแล้วไม่ใส่คืน จากกล่องที่ลูกบอลสีแดง 3 ลูก (ต่างกัน) สีขาว 2 ลูก (ต่างกัน) ให้ X เป็นจำนวนลูกบอลสีแดงที่หยิบได้ จงแสดงว่า X เป็นตัวแปรสุ่ม

แนวคำตอบ ให้ R_i แทนลูกบอกสีแดงลูกที่ i เมื่อ $i = 1, 2, 3$
และ W_i แทนลูกบอกสีขาวลูกที่ i เมื่อ $i = 1, 2$
แสดงได้ดังตาราง

X	จุดตัวอย่าง
0	W_1W_2, W_2W_1
1	R_1W_1, R_1W_2, R_2W_1, R_2W_2, R_3W_1, R_3W_2 W_1R_1, W_1R_2, W_1R_3, W_2R_1, W_2R_2, W_2R_3
2	R_1R_2, R_1R_3, R_2R_3, R_2R_1, R_3R_1, R_3R_2

เขียนได้ว่า $X(W_1W_2) = 0$ และ $X = 0$ จะหมายถึงเหตุการณ์ W_1W_2 และ W_2W_1

จากตัวอย่าง 4.1.2 จะเรียกตัวแปรสุ่มลักษณะนี้ว่า **ตัวแปรสุ่มไม่ต่อเนื่อง** (discrete random variable) ซึ่งหมายถึงจำนวนค่าของตัวแปรสุ่มมีจำนวนจำกัดหรือนับได้ ปริภูมิตัวอย่างที่มีจุดตัวอย่างเป็นตัวแปรสุ่มไม่ต่อเนื่องเรียกว่า **ปริภูมิตัวอย่างไม่ต่อเนื่อง** (discrete sample space)

ตัวอย่าง 4.1.3 จงแจกแจงค่าตัวแปรสุ่มไม่ต่อเนื่อง X ต่อไปนี้

1. ในการโยนเหรียญ 7 อัน
   ให้ X คือจำนวนหัวที่เกิดขึ้น
   จะได้ว่า $X = 0, 1, 2, ..., 7$
2. ในการทอดลูกเต๋า 1 ลูก จนกระทั่งเกิดแต้ม 5 เป็นครั้งแรก
   ให้ X คือจำนวนครั้งที่เกิดแต้ม 5
   จะได้ว่า $X = 0, 1$
3. ในการหยิบลูกบอล 5 ลูกพร้อมกันจากกล่องที่มีลูกบอลสีแดง 8 ลูก และสีเขียว 5 ลูก
   ให้ X คือจำนวนลูกบอลสีแดง
   จะได้ว่า $X = 0, 1, 2, 3, 4, 5$

ตัวแปรสุ่มต่อเนื่อง (continuous random variable) ซึ่งหมายถึงจำนวนค่าของตัวแปรสุ่มมีเป็นจำนวนนับไม่ได้ ปริภูมิตัวอย่างที่มีจุดตัวอย่างเป็นตัวแปรสุ่มไม่ต่อเนื่องเรียกว่า **ปริภูมิตัวอย่างต่อเนื่อง** (continuous sample space)

ตัวอย่าง 4.1.4 จงเขียนตัวแปรสุ่มต่อเนื่อง X ต่อไปนี้

1. ในการสุ่มหยิบหลอดไฟ เพื่อตรวจสอบอายุการใช้งาน หน่วยเป็นปี $S = \{t : t > 0\}$
   ให้ X คือเวลาที่หลอดไฟมีอายุไม่เกิน 1 ปี
   จะได้ว่า $X = t$ เมื่อ $t < 1$
2. ในการตรวจสอบน้ำหนักของเด็กแรกเกิด หน่วยเป็นกรัม $S = \{w : w > 0\}$
   ให้ X คือน้ำหนักของเด็กแรกเกิดอย่างน้อย 3000 กรัม
   จะได้ว่า $X = w$ เมื่อ $w \geq 3000$

แบบฝึกหัด 4.1

1. จงจำแนกตัวแปรสุ่มต่อไปนี้ว่าเป็นตัวแปรสุ่มต่อเนื่องหรือไม่ต่อเนื่อง

   1.1 จำนวนอุบัติเหตุทางรถยนต์ในกรุงเทพมหานครในรอบปี ๆ หนึ่ง

   1.2 ระยะเวลาที่ใช้เล่นกอล์ฟ 18 หลุมของชายคนหนึ่ง

   1.3 ปริมาณน้ำนมที่แม่วัวตัวหนึ่งให้ได้ในหนึ่งปี

   1.4 จำนวนไข่จากแม่ไก่หนึ่งตัวในแต่ละเดือน

   1.5 จำนวนอาคารที่ได้รับอนุญาตให้สร้างของเมือง ๆ หนึ่งในแต่ละเดือน

   1.6 น้ำหนักเป็นกิโลกรัมของพืชสวนที่ผลิตในเนื้อที่ 1 ไร่

   1.7 ระยะเวลาที่พนักงานของบริษัทหนึ่งมาทำงานสายในแต่ละวัน

   1.8 ระยะทางที่ขับไปส่งพนักงานของบริษัทในแต่ละวัน

2. จงแจกแจงค่าตัวแปรสุ่มไม่ต่อเนื่องต่อไปนี้

   2.1 จำนวนหัวที่เกิดขึ้นในการโยนเหรียญ 10 อัน

   2.2 จำนวนครั้งในการโยนเหรียญ 1 เหรียญจนกว่าจะขึ้นหัว

   2.3 จำนวนครั้งในการโยนเหรียญ 1 เหรียญจนกว่าจะขึ้นหัวเป็นครั้ง 3

   2.4 จำนวนครั้งที่เกิดแต้ม 3 ในการทอดลูกเต๋า 1 ลูก จนกระทั่งเกิดแต้ม 3 เป็นครั้งแรก

   2.5 จำนวนลูกบอลสีแดง ในการหยิบลูกบอล 4 ลูกพร้อมกันจากกล่องที่มีลูกบอลสีแดง 10 ลูก และสีเขียว 5 ลูก

3. จงเขียนตัวแปรสุ่มต่อเนื่องต่อไปนี้

   3.1 น้ำหนักเป็นกิโลกรัมของข้าวที่ผลิตในเนื้อที่ 1 ไร่

   3.2 ระยะเวลาในการทำการบ้านของนักเรียนในแต่ละวัน

4.2 การแจกแจงความน่าจะเป็นไม่ต่อเนื่อง

บทนิยาม 4.2.1 ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องที่มีค่าเป็น $x_1, x_2, ..., x_n, ...$ ด้วยความน่าจะเป็น $f(x_1), f(x_2), ..., f(x_n), ...$ ตามลำดับ ถ้าเหตุการณ์ A เป็นเซตย่อยของ $\{x_1, x_2, ..., x_n, ...\}$ จะได้ว่าความน่าจะเป็นของเหตุการณ์ A คือ

$$P(A) = \sum_x f(x) \qquad \text{ทุกค่า } x \text{ ของเหตุการณ์ } A$$

บทนิยาม 4.2.2 ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องใด ๆ กำหนดให้

$$f(x) = P(X = x)$$

โดยที่ $f(x) \geq 0$ และ $\sum_x f(x) = 1$ ทุกค่า x ของ X

เรียก f ว่า **ฟังก์ชันความน่าจะเป็น** (probability mass function) เขียนย่อ p.m.f หรือ **การแจกแจงความน่าจะเป็น** (probability distribution) ของตัวแปรสุ่มไม่ต่อเนื่อง X

ตัวอย่าง 4.2.3 โยนเหรียญเที่ยงตรง 1 อัน 3 ครั้ง ให้ X แทนจำนวนหัวที่เกิดขึ้น
จงสร้างตารางแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X

แนวคำตอบ จะเห็นว่าปริภูมิตัวอย่าง S มี $n(S) = 2^3 = 8$ และ $X = 0, 1, 2, 3$ แสดงตารางได้ดังนี้

X	0	1	2	3
จุดตัวอย่าง	TTT	HTT, THT, TTH	HHT, HTH, THH	HHH
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

ตัวอย่าง 4.2.4 ในการทอดลูกเต๋า 2 ลูกพร้อมกัน ให้ X แทนจำนวนลูกเต๋าที่ขึ้นหน้า 1 หรือ 2
จงสร้างตารางแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X

แนวคำตอบ จะเห็นว่าปริภูมิตัวอย่าง S มี $n(S) = 6 \cdot 6 = 36$ และ $X = 0, 1, 2$ แสดงตารางได้ดังนี้

X	0	1	2
จุดตัวอย่าง	(x, y) เมื่อ $x, y \in \{3, 4, 5, 6\}$	(x, y) $x \in \{1, 2\}$ และ $y \in \{3, 4, 5, 6\}$ หรือ $x \in \{3, 4, 5, 6\}$ และ $y \in \{1, 2\}$	(x, y) $x, y \in \{1, 2\}$
$P(X = x)$	$\frac{16}{36}$	$\frac{16}{36}$	$\frac{4}{36}$

บทนิยาม 4.2.5 ฟังก์ชันความน่าจะเป็นสะสม (cumulative probability function: C.D.F.) ของตัวแปรสุ่มไม่ต่อเนื่อง ซึ่งมีฟังก์ชันความจะเป็น f คือ

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

เมื่อ $x \in \mathbb{R}$ และ $P(a < X \leq b) = F(b) - F(a)$

ข้อสังเกต $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$ และ $P(X = a) = P(a \leq X \leq a)$

ตัวอย่าง 4.2.6 โยนเหรียญเที่ยงตรง 1 อัน 3 ครั้ง ให้ X แทนจำนวนหัวที่เกิดขึ้น
จงหาการแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่ม X พร้อมวาดกราฟประกอบ

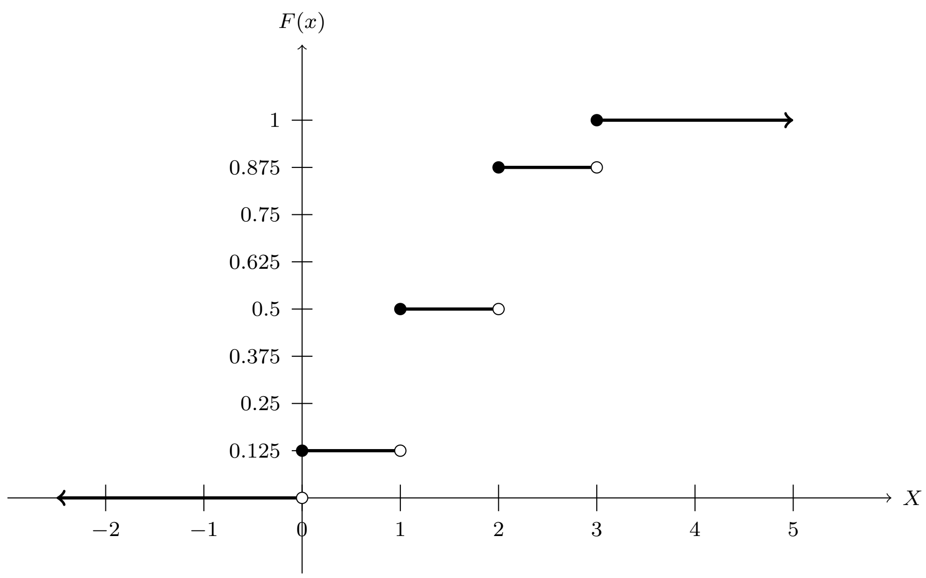
แนวคำตอบ จากตัวอย่าง 4.2.3 จะได้ว่า

X	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$F(x)$	$\frac{1}{8} = 0.125$	$\frac{4}{8} = 0.5$	$\frac{7}{8} = 0.875$	1

ให้ $x \in \mathbb{R}$ จะได้ว่าการแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่ม X คือ

$$F(x) = \begin{cases} 0 & \text{เมื่อ } x < 0 \\ 0.125 & \text{เมื่อ } 0 \leq x < 1 \\ 0.5 & \text{เมื่อ } 1 \leq x < 2 \\ 0.875 & \text{เมื่อ } 2 \leq x < 3 \\ 1 & \text{เมื่อ } x \geq 3 \end{cases}$$

แสดงกราฟได้ดังนี้



ตัวอย่าง 4.2.7 โยนเหรียญเที่ยงตรง 5 อันพร้อมกัน ครั้ง ให้ X แทนจำนวนหัวที่เกิดขึ้น จงหา

1. $P(X = 0)$
2. $P(1 < X \leq 4)$
3. $P(1 \leq X < 3)$
4. $P(X > 3)$
5. $P(-1 < X < 2)$
6. $P(X \leq \pi)$

แนวคำตอบ แสดงตารางแจกแจงความน่าจะเป็นได้ดังนี้

X	0	1	2	3	4	5
$f(x) = P(X = x)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$F(x)$	$\frac{1}{32}$	$\frac{6}{32}$	$\frac{16}{32}$	$\frac{26}{32}$	$\frac{31}{32}$	1

1. $P(X = 0) = \dfrac{1}{32}$
2. $P(1 < X \leq 4) = F(4) - F(1) = \dfrac{31}{32} - \dfrac{6}{32} = \dfrac{25}{32}$
3. $P(1 \leq X < 3) = P(0 < X \leq 2) = F(2) - F(0) = \dfrac{16}{32} - \dfrac{1}{32} = \dfrac{15}{32}$
4. $P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = 1 - \dfrac{26}{32} = \dfrac{6}{32}$
5. $P(-1 < X < 2) = P(-1 < X \leq 1) = F(1) - F(-1) = \dfrac{6}{32} - 0 = \dfrac{6}{32}$
6. $P(X \leq \pi) = P(X \leq 3) = F(3) = \dfrac{26}{32}$

ตัวอย่าง 4.2.8 ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่องที่มีค่าเป็น 1, 2, 3, 4, 5, 6, 7, 8 และ f เป็นการแจกแจงความน่าจะเป็น แสดงได้ดังตาราง

x	1	2	3	4	5	6	7	8
$f(x)$	k	$2k$	$3k$	$2k$	k	k	$3k$	$2k$

1. จงหาค่า k

   แนวคำตอบ จะได้ว่า

$$\begin{aligned}\sum f(x) &= 1 \\ k + 2k + 3k + 2k + k + k + 3k + 2k &= 1 \\ 15k &= 1 \\ k &= \frac{1}{15}\end{aligned}$$

2. จงหา $F(x)$

   แนวคำตอบ

x	1	2	3	4	5	6	7	8
$f(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{3}{15}$	$\frac{2}{15}$
$F(x)$	$\frac{1}{15}$	$\frac{3}{15}$	$\frac{6}{15}$	$\frac{8}{15}$	$\frac{9}{15}$	$\frac{10}{15}$	$\frac{13}{15}$	1

3. $P(2 \leq X \leq 6)$

แนวคำตอบ

$$\begin{aligned} P(2 \leq X \leq 6) &= P(1 < X \leq 6) \\ &= P(X \leq 6) - P(X \leq 1) \\ &= F(6) - F(1) \\ &= \frac{10}{15} - \frac{1}{15} = \frac{3}{5} \end{aligned}$$

4. $P(X < 3 \text{ หรือ } X > 6)$

แนวคำตอบ

$$\begin{aligned} P(X < 3 \text{ หรือ } X > 6) &= P(X < 3) + P(X > 6) \\ &= P(X \leq 2) + 1 - P(X \leq 6) \\ &= F(2) + 1 - F(6) \\ &= \frac{3}{15} + 1 - \frac{10}{15} = \frac{8}{15} \end{aligned}$$

5. $P(X \neq 4)$

แนวคำตอบ

$$\begin{aligned} P(X \neq 4) &= 1 - P(X = 4) \\ &= 1 - f(4) \\ &= 1 - \frac{2}{15} = \frac{13}{15} \end{aligned}$$

แบบฝึกหัด 4.2

1. ในการทอดลูกเต๋า 2 ลูกพร้อมกับโยนเหรียญ 1 เหรียญ ให้ X แทนจำนวนลูกเต๋าที่ขึ้นแต้มคู่ จงสร้างตารางแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X และหาความน่าจะเป็นสะสมพร้อมวาดกราฟประกอบ

2. สมาร์ท TV จำนวน 6 เครื่อง มีเครื่องบกพร่อง 2 เครื่อง โรงแรมแห่งหนึ่งต้องการซื้อ 3 เครื่อง ถ้า X เป็นจำนวนเครื่องที่บกพร่องที่โรงแรมรับมา จงแจกแจงความน่าจะเป็นพร้อมแสดงเป็นตาราง และหา

   2.1 $P(X = 1)$ 2.2 $P(X > 1)$ 2.3 $P(0 < X \leq 2)$

3. กล่องใบหนึ่งบรรจุลูกบอลสีแดง 2 ลูก สีน้ำเงิน 5 ลูก และสีเขียว 3 ลูก หยิบลูกบอลครั้งละ 1 ลูกอย่างสุ่ม 10 ครั้ง โดยหยิบแล้วใส่กล่องกลับคืนก่อนจะหยิบครั้งต่อไป ให้ X คือจำนวนครั้งที่หยิบได้ลูกบอลสีเขียว จงหาสูตรของความน่าจะเป็น

4. นาย ก ซื้อหลอดไฟมา 5 หลอด เขานำไปทดสอบที่ละหลอด ถ้าหลอดไฟที่เขาซื้อมานั้นมีหลอดเสียอยู่ 1 หลอด ให้ X เป็นจำนวนที่หลอดไฟที่ต้องการทดสอบเพื่อหาหลอดเสีย

   4.1 จงสร้างตารางแจกแจงความน่าจะเป็นของ X

   4.2 จงหาความน่าจะเป็นสะสมของ X

5. กำหนดค่าของการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้

x	1	2	3	4	5	6	7
$f(x)$	a	$2a$	$2a$	$3a$	a^2	$2a^2$	$7a^2 + a$

   จงหา

   5.1 ค่าของ a

   5.2 $F(x)$

   5.3 $P(X \geq 5)$

   5.4 $P(X < 3)$

   5.5 $P(-3 \leq X \leq 3)$

   5.6 $P(X \neq 1)$

   5.7 $P(X < 4$ หรือ $X > 5)$

   5.8 ค่าต่ำสุดของ k ถ้า $P(X \leq k) = \dfrac{1}{2}$

6. ให้ f เป็นฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X ที่มีค่าเป็น 2, 3, 4 ที่นิยามโดย

$$f(x) = A\cos^2\left(\frac{\pi}{x}\right) \quad \text{เมื่อ} \quad x = 2, 3, 4$$

   จงหา

6.1 ค่าของ A

6.2 $F(x)$ พร้อมเขียนกราฟ

6.3 $P(1 < X \leq 3)$

6.4 $P(X \geq 1)$

6.5 $P(X \neq 1)$

6.6 $P(X > A)$

7. ให้ f เป็นฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X ที่มีค่าเป็น 0, 1, 2, 3 ที่นิยามโดย

$$f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{3-x} \quad \text{เมื่อ} \quad x = 0, 1, 2, 3$$

จงหา

7.1 $P(X = 2)$

7.2 $P(X < 1)$

7.3 $P(X \geq 2)$

7.4 $P(X \geq 1$ และ $X \neq 2)$

7.5 $P(X \geq 1$ และ $X < 3)$

7.6 $\sum_{x=1}^{3} f(x)$

7.7 $P(X \neq 1$ หรือ $X < 2)$

8. ให้ F ความน่าจะเป็นสะสมของตัวแปรสุ่ม X ที่มีค่าเป็น 1, 2, 3, 4, 5 ที่นิยามโดย

$$F(x) = kx^2 \quad \text{เมื่อ} \quad x = 1, 2, 3, 4, 5$$

จงหา

8.1 ค่าของ k

8.2 $f(x)$ เมื่อ $x = 1, 2, 3, 4, 5$

8.3 $P(X < 3)$

8.4 $P(4 > X > 2)$

8.5 $P(X \neq 4)$

8.6 ค่าต่ำสุดของ m ถ้า $P(X > m) = 0.5$

4.3 การแจกแจงความน่าจะเป็นต่อเนื่อง

บทนิยาม 4.3.1 เรียก f ว่าเป็น **ฟังก์ชันความหนาแน่นของความน่าจะเป็น** (probability density function) เขียนย่อว่า p.d.f หรือเรียกสั้น ๆ ว่า **ฟังก์ชันความน่าจะเป็น** ของตัวแปรสุ่มต่อเนื่อง X ก็ต่อเมื่อ

1. $f(x) \geq 0$ ทุก ๆ $x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $\displaystyle P(a < X < b) = \int_a^b f(x)\,dx$

สำหรับตัวแปรสุ่มต่อเนื่อง X จะกำหนดให้มีเป็นจำนวนจริงทุกค่า หรือกล่าวได้ว่ามีค่าใน $\mathbb{R}$ หรือในช่วง $(-\infty, \infty)$ และจะเห็นว่า

$$P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b)$$
$$P(X = c) = P(c \leq X \leq c) = \int_c^c f(x)\,dx = 0$$

ตัวอย่าง 4.3.2 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad \text{ทุก ๆ } x \in \mathbb{R}$$

1. จงแสดงว่า $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$
2. จงหา $P(-1 < X < 1)$

แนวคำตอบ จะได้ว่า

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{0} \frac{1}{\pi(1+x^2)}\,dx + \int_0^{\infty} \frac{1}{\pi(1+x^2)}\,dx \\
&= \lim_{t\to-\infty} \int_t^0 \frac{1}{\pi(1+x^2)}\,dx + \lim_{s\to\infty} \int_0^s \frac{1}{\pi(1+x^2)}\,dx \\
&= \lim_{t\to-\infty} \left[\frac{1}{\pi}\text{arctan}x\right]_t^0 + \lim_{s\to\infty} \int_0^s \left[\frac{1}{\pi}\text{arctan}x\right]_0^s \\
&= \lim_{t\to-\infty} \left[-\frac{1}{\pi}\text{arctan}t\right] + \lim_{s\to\infty} \int_0^s \left[\frac{1}{\pi}\text{arctan}s\right] \\
&= -\frac{1}{\pi}\left(-\frac{\pi}{2}\right) + \frac{1}{\pi}\left(\frac{\pi}{2}\right) = 1
\end{aligned}$$

และ

$$\begin{aligned}
P(-1 < X < 1) &= \int_{-1}^{1} \frac{1}{\pi(1+x^2)}\,dx \\
&= \left[\frac{1}{\pi}\text{arctan}x\right]_{-1}^{1} \\
&= \frac{1}{\pi}\text{arctan}(1) - \frac{1}{\pi}\text{arctan}(-1) \\
&= \frac{1}{\pi}\left(\frac{\pi}{4}\right) - \frac{1}{\pi}\left(-\frac{\pi}{4}\right) = \frac{1}{2}
\end{aligned}$$

ตัวอย่าง 4.3.3 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \begin{cases} k(3+2x) & \text{เมื่อ } 2 < x < 4 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

1. จงหาค่า k

   แนวคำตอบ

   $$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_2^4 k(3+2k)\,dx = k\left[3x + x^2\right]_2^4 = k[(12+16)-(6+4)] = 18k$$

   ดังนั้น $k = \dfrac{1}{18}$

2. จงหา $P(X > 3)$

   แนวคำตอบ

   $$\begin{aligned} P(X > 3) &= \int_3^{\infty} f(x)\,dx = \int_3^4 k(3+2k)\,dx \\ &= \frac{1}{18}\left[3x + x^2\right]_3^4 = \frac{1}{18}[(12+16)-(9+9)] = \frac{5}{9} \end{aligned}$$

ตัวอย่าง 4.3.4 ถ้าเวลา (ชั่วโมง) ที่คอมพิวเตอร์เครื่องหนึ่งทำงานก่อนจะเสียเป็นตัวแปรสุ่มต่อเนื่องที่มี p.d.f คือ

$$f(x) = \begin{cases} re^{-\frac{x}{100}} & \text{เมื่อ } x \ge 0 \\ 0 & \text{เมื่อ } x < 0 \end{cases}$$

จงหาความน่าจะเป็นที่เครื่องคอมพิวเตอร์เครื่องหนึ่งจะทำงานได้ระหว่าง 50 ถึง 150 ชั่วโมงก่อนจะเสีย

แนวคำตอบ หา r จากเงื่อนไขที่ว่า

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)\,dx = \lim_{t\to\infty} \int_0^t re^{-\frac{x}{100}}\,dx \\ &= \lim_{t\to\infty} \left[-100re^{-\frac{x}{100}}\right]_0^t \\ &= \lim_{t\to\infty} \left[-100re^{-\frac{t}{100}} + 100r\right] \\ &= 100r \end{aligned}$$

ดังนั้น $r = \dfrac{1}{100}$ จะได้ว่า

$$\begin{aligned} P(50 < X < 150) &= \int_{50}^{150} f(x)\,dx = \int_{50}^{150} \frac{1}{100}e^{-\frac{x}{100}}\,dx \\ &= \left[-e^{-\frac{x}{100}}\right]_{50}^{150} = -e^{-1.5} + e^{-0.5} \\ &= \frac{1}{\sqrt{e}} - \frac{1}{e\sqrt{e}} = \frac{e-1}{e\sqrt{e}} = 0.3834 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่เครื่องคอมพิวเตอร์เครื่องหนึ่งจะทำงานได้ระหว่าง 50 ถึง 150 ชั่วโมงก่อนจะเสียเท่ากับ 0.3834

บทนิยาม 4.3.5 การแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่มต่อเนื่อง X ที่มี p.d.f เป็น f คือ

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

จากนิยามของการแจกแจงความน่าจะเป็นสะสม จะได้ว่า

1. $P(a < X < b) = F(b) - F(a)$
2. $P(X > a) = 1 - F(a)$
3. $\dfrac{dF(x)}{dx} = f(x)$
4. $\lim\limits_{x\to-\infty} F(x) = 0$ และ $\lim\limits_{x\to\infty} F(x) = 1$

ตัวอย่าง 4.3.6 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad \text{ทุก ๆ } x \in \mathbb{R}$$

จงหาการแจกแจงความน่าจะเป็นสะสม $F(x)$ และใช้หาค่าต่อไปนี้

1. $P(X < 1)$ 2. $P(X > -1)$ 3. $P(-2 < X \le 10)$

แนวคำตอบ ให้ $x \in \mathbb{R}$ จะได้ว่า

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,dt = \lim_{s\to-\infty} \int_{s}^{x} \frac{1}{\pi(1+t^2)}\,dx \\ &= \lim_{s\to-\infty} \left[\frac{1}{\pi}\text{arctan}t\right]_{s}^{x} = \lim_{s\to-\infty} \left[\frac{1}{\pi}\text{arctan}x - \frac{1}{\pi}\text{arctan}s\right] \\ &= \frac{1}{\pi}\text{arctan}x - \frac{1}{\pi}\left(-\frac{\pi}{2}\right) = \frac{1}{\pi}\text{arctan}x + \frac{1}{2} \end{aligned}$$

ดังนั้น

$$F(x) = \frac{1}{\pi}\text{arctan}x + \frac{1}{2} \quad \text{เมื่อ} \quad x \in \mathbb{R}$$

1. $P(X < 1) = F(1) = \frac{1}{\pi}\text{arctan}1 + \frac{1}{2} = \frac{1}{\pi}\left(\frac{\pi}{4}\right) + \frac{1}{2} = \frac{3}{4}$
2. $P(X > -1) = 1 - P(X < -1) = 1 - \left[\frac{1}{\pi}\text{arctan}(-1) + \frac{1}{2}\right] = 1 - \left[\frac{1}{\pi}\left(-\frac{\pi}{4}\right) + \frac{1}{2}\right] = \frac{3}{4}$
3. $$\begin{aligned} P(-2 < X \le 10) &= F(10) - F(-2) \\ &= \left[\frac{1}{\pi}\text{arctan}(10) + \frac{1}{2}\right] - \left[\frac{1}{\pi}\text{arctan}(-2) + \frac{1}{2}\right] \\ &= \frac{1}{\pi}\text{arctan}(10) - \frac{1}{\pi}\text{arctan}(-2) \\ &= 0.8207 \end{aligned}$$

ตัวอย่าง 4.3.7 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \begin{cases} \dfrac{1}{18}(3+2x) & \text{เมื่อ } 2 < x < 4 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาการแจกแจงความน่าจะเป็นสะสม $F(x)$ พร้อมเขียนกราฟ และใช้หาค่า $P(-3 < X < 3)$

แนวคำตอบ ให้ $x \in \mathbb{R}$

กรณี $x \leq 2$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$$

กรณี $2 < x < 4$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{2} f(t)\,dt + \int_{2}^{x} f(t)\,dt \\ &= \int_{-\infty}^{2} 0\,dt + \int_{2}^{x} \frac{1}{18}(3+2t)\,dt \\ &= 0 + \frac{1}{18}\left[3t+t^2\right]_2^x = \frac{1}{18}[(3x+x^2)-(6+4)] = \frac{1}{18}(x^2+3x-10) \end{aligned}$$

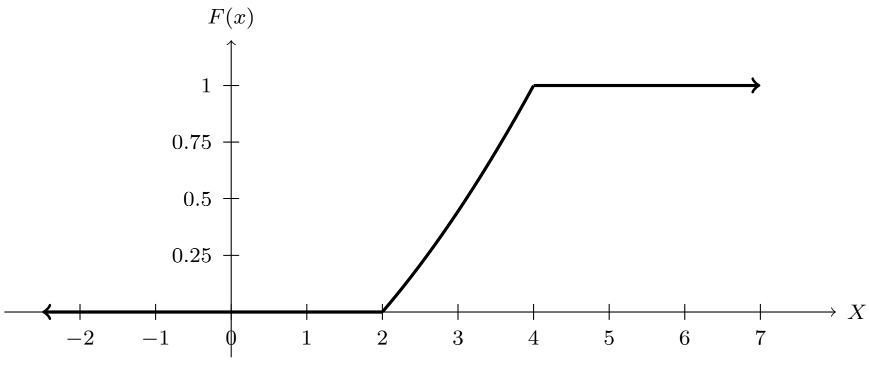
กรณี $x \geq 4$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{2}^{4} f(t)\,dt + \int_{2}^{x} f(t)\,dt = 1 + \int_{2}^{x} 0\,dt = 1 + 0 = 1$$

ดังนั้น

$$F(x) = \begin{cases} 0 & \text{เมื่อ } x \leq 2 \\ \frac{1}{18}(x^2+3x-10) & \text{เมื่อ } 2 < x < 4 \\ 1 & \text{เมื่อ } x \geq 4 \end{cases}$$

เขียนกราฟได้ดังนี้



จะได้ว่า

$$P(-3 < X < 3) = F(3) - F(-3) = \frac{1}{18}(9+9-10) - 0 = \frac{4}{9}$$

แบบฝึกหัด 4.3

1. ตัวแปรสุ่มต่อเนื่อง X ซึ่งสมมติว่ามีค่าอยู่ระหว่าง $x = 1$ และ $x = 3$ มี p.d.f. เป็น $f(x) = \dfrac{1}{2}$

    1.1 จงหา $P(2 < X < 2.5)$

    1.2 จงหา $P(X > 2)$

2. ตัวแปรสุ่มต่อเนื่อง X ซึ่งสมมติว่ามีค่าอยู่ระหว่าง $x = 2$ และ $x = 5$ มี p.d.f. เป็น $f(x) = \dfrac{2}{27}(1 + x)$

    2.1 จงหา $F(x)$

    2.2 จงหา $P(3 < X < 4)$

    2.3 จงหา $P(X < 4)$

    2.4 จงหา $P(X > 3)$

3. ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ $f(x) = \begin{cases} e^{-x} & \text{เมื่อ } x > 0 \\ 0 & \text{เมื่อ } x \le 0 \end{cases}$

    3.1 จงแสดงว่า $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

    3.2 จงหา $P(0 < X < 3)$

    3.3 จงหา $P(X > 1)$

    3.4 จงหา $P(X \le \ell\text{n}2)$

4. ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ $f(x) = \begin{cases} \dfrac{1}{10} & \text{เมื่อ } 0 < x < 10 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$

    4.1 จงหา $F(x)$

    4.2 จงหา $P(X < 2)$

    4.3 จงหา $P(X > 1)$

    4.4 จงหา $P(-2 < X \le 5)$

5. ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ $f(x) = \begin{cases} \dfrac{1}{2} - ax & \text{เมื่อ } 0 < x < 4 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$

    5.1 จงหา a

    5.2 จงหา $F(x)$

    5.3 จงหา $P(1 < X < 2)$

    5.4 จงหา $P(X \ge 2)$

6. ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ $f(x) = \begin{cases} \dfrac{200}{9}x & \text{เมื่อ } 0 \le x \le 0.3 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$

    6.1 จงหา $F(x)$

    6.2 จงหา $P(0.1 < X \le 0.2)$

7. ให้การแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่มต่อเนื่อง X คือ

$$F(x) = \begin{cases} 0 & \text{เมื่อ } x < 0 \\ x & \text{เมื่อ } 0 \le x \le 1 \\ 1 & \text{เมื่อ } x > 1 \end{cases}$$

จงหาฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X

4.4 การแจกแจงความน่าจะเป็นร่วมกัน

บทนิยาม 4.4.1 ให้ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง กำหนดให้

$$f(x,y) = P(X = x, Y = y)$$

โดยที่ $f(x,y) \geq 0$ และ $\sum_x \sum_y f(x,y) = 1$ ทุกค่า (x,y)
เรียก f ว่า **ฟังก์ชันความน่าจะเป็นร่วมกัน** (joint probability mass function) ของตัวแปรสุ่มไม่ต่อเนื่อง X และ Y ความน่าจะเป็นของเหตุการณ์ A เขียนแทนด้วย $P[(x,y) \in A]$ นิยามโดย

$$P[(x,y) \in A] = \sum_{(x,y)} \sum_{\in A} f(x,y)$$

ในกรณี $A = \{(x,y) : a < X < b$ และ $c < Y < d\}$ เขียน $P[(x,y) \in A]$ แทนด้วย

$$P(a < X < b, c < Y < d)$$

ตัวอย่าง 4.4.2 กล่องใบหนึ่งมีลูกบอลสีดำ 3 ลูก สีแดง 2 ลูก และสีเขียว 3 ลูก หยิบลูกบอล 2 ลูกอย่างสุ่ม ให้ X เป็นจำนวนลูกบอลสีดำ และ Y เป็นจำนวนลูกบอลสีแดงที่หยิบได้ จงสร้างตารางของฟังก์ชันความน่าจะเป็นร่วมกันของ X และ Y

แนวคำตอบ หยิบลูกบอล 2 ลูก ได้ลูกบอลสีดำ x เมื่อ $x = 0, 1, 2$
ได้ลูกบอลสีแดง y เมื่อ $y = 0, 1, 2$
ได้ลูกบอลสีเขียว $2 - x - y$ เมื่อ $0 \leq x + y \leq 2$

จะได้ว่า

$$f(x,y) = P(X = x, Y = y) = \frac{\binom{3}{x}\binom{2}{y}}{\binom{8}{2}} \quad \text{เมื่อ} \quad x, y = 0, 1, 2 \text{ และ } 0 \leq x + y \leq 2$$

แสดงค่าดังตารางต่อไปนี้

		$x=0$	$x=1$	$x=2$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{6}{28}$	$\frac{6}{28}$	$-$
	2	$\frac{1}{28}$	$-$	$-$

ตัวอย่าง 4.4.3 โยนลูกเต๋า 2 ลูก 1 ครั้ง จงหาการแจกแจงความน่าจะเป็นร่วมของจำนวนแต้มคี่ที่ได้จากการโยนลูกเต๋า 2 ลูก กับ ผลบวกของแต้มที่ได้จากการโยนลูกเต๋า 2 ลูก

แนวคำตอบ ให้ X คือจำนวนแต้มคี่ที่ได้จากการโยนลูกเต๋า 2 ลูก

Y คือผลบวกของแต้มที่ได้จากการโยนลูกเต๋า 2 ลูก

แสดงจุดตัวอย่างค่าดังตาราง

		จำนวนแต้มคี่		
		0	1	2
ผลบวกแต้ม	2	–	–	$(1,1)$
	3	–	$(1,2),(2,1)$	–
	4	$(2,2)$	–	$(1,3),(3,1)$
	5	–	$(1,4),(3,1),(4,1),(1,4)$	–
	6	$(2,4),(4,2)$	–	$(1,5),(5,1),(3,3)$
	7	–	$(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$	–
	8	$(2,6),(4,4),(6,2)$	–	$(1,7),(7,1)$
	9	–	$(3,6),(4,5),(5,4),(6,3)$	–
	10	$(4,6),(6,4)$	–	$(5,5)$
	11	–	$(5,6),(6,5)$	–
	12	$(6,6)$	–	–

การแจกแจงความน่าจะเป็นร่วมของ X และ Y แสดงได้ดังนี้

		x		
		0	1	2
y	2	–	–	$\frac{1}{36}$
	3	–	$\frac{2}{36}$	–
	4	$\frac{1}{36}$	–	$\frac{2}{36}$
	5	–	$\frac{4}{36}$	–
	6	$\frac{2}{36}$	–	$\frac{3}{36}$
	7	–	$\frac{6}{36}$	–
	8	$\frac{3}{36}$	–	$\frac{2}{36}$
	9	–	$\frac{4}{36}$	–
	10	$\frac{2}{36}$	–	$\frac{1}{36}$
	11	–	$\frac{2}{36}$	–
	12	$\frac{1}{36}$	–	–

บทนิยาม 4.4.4 เรียก f ว่าเป็น **ฟังก์ชันความหนาแน่นร่วมกัน** (joint density function) ของตัวแปรสุ่มต่อเนื่อง X และ Y ก็ต่อเมื่อ

1. $f(x,y) \geq 0$ ทุก ๆ (x,y)
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$
3. $\displaystyle P[(x,y) \in A] = \iint_A f(x,y)\,dxdy$

ตัวอย่าง 4.4.5 ให้ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & \text{เมื่อ } 0 < x < 2 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

1. จงแสดงว่า $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$

   แนวคำตอบ

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy &= \int_0^1\int_0^2 \frac{x(1+3y^2)}{4}\,dx\,dy \\
&= \int_0^1 \frac{1+3y^2}{4}\left[\frac{x^2}{2}\right]_{x=0}^{x=2} dy \\
&= \int_0^1 \frac{1+3y^2}{2}\,dy \\
&= \left[\frac{y}{2} + \frac{y^3}{2}\right]_0^1 = \left[\frac{1}{2} + \frac{1}{2}\right] - 0 = 1
\end{aligned}$$

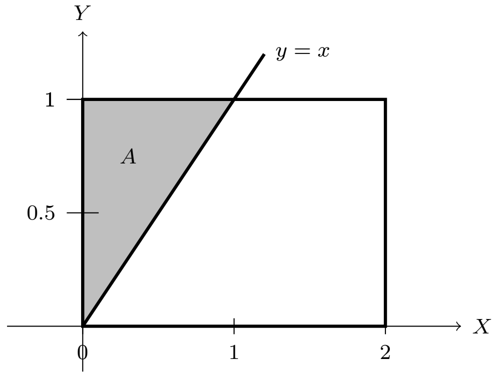
2. จงหา $P[(x,y) \in A]$ เมื่อ $A = \{(x,y) : 0 < x < 1 \text{ และ } \frac{1}{4} < y < \frac{1}{2}\}$

   แนวคำตอบ

$$\begin{aligned}
P[(x,y) \in A] &= \iint_A f(x,y)\,dxdy \\
&= \int_{\frac{1}{4}}^{\frac{1}{2}}\int_0^1 f(x,y)\,dx\,dy \\
&= \int_{\frac{1}{4}}^{\frac{1}{2}}\int_0^1 \frac{x(1+3y^2)}{4}\,dx\,dy \\
&= \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{1+3y^2}{4}\left[\frac{x^2}{2}\right]_{x=0}^{x=1} dy \\
&= \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{1+3y^2}{8}\,dy \\
&= \left[\frac{y}{8} + \frac{y^3}{8}\right]_{\frac{1}{4}}^{\frac{1}{2}} = \left[\frac{1}{16} + \frac{1}{64}\right] - \left[\frac{1}{32} + \frac{1}{512}\right] = \frac{23}{512}
\end{aligned}$$

3. จงหา $P[(x, y) \in A]$ เมื่อ $A = \{(x, y) : x < y\}$

แนวคำตอบ พิจารณา A จากกราฟ



นั่นคือ $A = \{(x, y) : 0 < x < y$ และ $0 < y < 1\}$ ดังนั้น

$$\begin{aligned}
P[(x, y) \in A] &= \iint_A f(x, y)\, dxdy \\
&= \int_0^1 \int_0^y f(x, y)\, dx\, dy \\
&= \int_0^1 \int_0^y \frac{x(1 + 3y^2)}{4}\, dx\, dy \\
&= \int_0^1 \frac{1 + 3y^2}{4} \left[\frac{x^2}{2}\right]_{x=0}^{x=y} dy = \int_0^1 \frac{1 + 3y^2}{4} \left[\frac{y^2}{2}\right] dy \\
&= \int_0^1 \frac{y^2 + 3y^4}{8}\, dy \\
&= \left[\frac{y^3}{24} + \frac{3y^5}{40}\right]_0^1 = \left[\frac{1}{24} + \frac{1}{40}\right] - 0 = \frac{1}{15}
\end{aligned}$$

ตัวอย่าง 4.4.6 ให้ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x, y) = \begin{cases} kxy & \text{เมื่อ } 0 < x < 3 \text{ และ } 0 < y < 4 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหา

1. ค่าของ k

แนวคำตอบ หา k จากเงื่อนไขที่ว่า

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^4 \int_0^3 kxy\, dx\, dy \\
&= \int_0^4 \left[\frac{kx^2 y}{2}\right]_{x=0}^{x=3} dy = \int_0^4 \frac{9ky}{2}\, dy \\
&= \left[\frac{9ky^2}{4}\right]_0^4 = 36k
\end{aligned}$$

ดังนั้น $k = \dfrac{1}{36}$

2. $P(0 < X < 1, 0 < Y < 1)$

แนวคำตอบ

$$\begin{aligned} P(0 < X < 1, 0 < Y < 1) &= \int_0^1 \int_0^1 f(x,y)\,dx\,dy = \int_0^1 \int_0^1 \frac{1}{36}xy\,dx\,dy \\ &= \int_0^1 \left[\frac{x^2y}{72}\right]_{x=0}^{x=1} dy = \int_0^1 \frac{y}{72}\,dy \\ &= \left[\frac{y^2}{144}\right]_0^1 = \frac{1}{144} \end{aligned}$$

3. $P(X > 1, Y < 1)$

แนวคำตอบ

$$\begin{aligned} P(X > 1, Y < 1) &= \int_1^\infty \int_{-\infty}^1 f(x,y)\,dy\,dx = \int_1^3 \int_0^1 \frac{1}{36}xy\,dy\,dx \\ &= \int_1^3 \left[\frac{xy^2}{72}\right]_{y=0}^{y=1} dx = \int_1^3 \frac{x}{72}\,dx \\ &= \left[\frac{x^2}{144}\right]_1^3 = \frac{9}{144} - \frac{1}{144} = \frac{1}{18} \end{aligned}$$

4. $P(X < 2)$

แนวคำตอบ

$$\begin{aligned} P(X < 2) &= \int_{-\infty}^2 \int_{-\infty}^\infty f(x,y)\,dy\,dx = \int_0^2 \int_0^4 \frac{1}{36}xy\,dy\,dx \\ &= \int_0^2 \left[\frac{xy^2}{72}\right]_{y=0}^{y=4} dx = \int_0^2 \frac{2x}{9}\,dx \\ &= \left[\frac{x^2}{9}\right]_0^2 = \frac{4}{9} \end{aligned}$$

5. $P(Y > 3)$

แนวคำตอบ

$$\begin{aligned} P(Y > 3) &= \int_3^\infty \int_{-\infty}^\infty f(x,y)\,dx\,dy = \int_3^4 \int_0^3 \frac{1}{36}xy\,dx\,dy \\ &= \int_3^4 \left[\frac{x^2y}{72}\right]_{x=0}^{x=3} dx = \int_3^4 \frac{y}{8}\,dx \\ &= \left[\frac{y^2}{16}\right]_3^4 = \frac{16}{16} - \frac{9}{16} = \frac{7}{16} \end{aligned}$$

6. $P(X > 0, Y = 1)$

แนวคำตอบ

$$\begin{aligned} P(X > 0, Y = 1) &= P(X > 0, 1 \le Y \le 1) \\ &= \int_0^\infty \int_1^1 f(x,y)\,dy\,dx = \int_0^3 0\,dx = 0 \end{aligned}$$

บทนิยาม 4.4.7 ฟังก์ชัน f_X เป็น**การแจกแจงมาร์จินัล** (marginal distribution) **ของ** X และ f_Y เป็น**การแจกแจงมาร์จินัลของ** Y นิยามโดย

1. สำหรับ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$f_X(x) = \sum_y f(x,y) \quad \text{และ} \quad f_Y(y) = \sum_x f(x,y)$$

2. สำหรับ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{และ} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

ตัวอย่าง 4.4.8 จงหาการแจกแจงมาร์จินัลของ X และ Y ถ้ามีการแจกแจงความน่าจะเป็นร่วมกันดังตาราง

		x		
		0	1	2
y	0	0	0.3	0.1
	1	0.1	0	0.2
	2	0.1	0.2	0

แนวคำตอบ โดยนิยามจะได้ว่า

$$\begin{aligned}
f_X(0) &= f(0,0) + f(0,1) + f(0,2) = 0 + 0.1 + 0.1 = 0.2 \\
f_X(1) &= f(1,0) + f(1,1) + f(1,2) = 0.3 + 0 + 0.2 = 0.5 \\
f_X(2) &= f(2,0) + f(2,1) + f(2,2) = 0.1 + 0.2 + 0 = 0.3 \\
f_Y(0) &= f(0,0) + f(1,0) + f(2,0) = 0 + 0.3 + 0.1 = 0.4 \\
f_Y(1) &= f(0,1) + f(1,1) + f(2,1) = 0.1 + 0 + 0.2 = 0.3 \\
f_Y(2) &= f(0,2) + f(1,2) + f(2,2) = 0.1 + 0.2 + 0 = 0.3
\end{aligned}$$

หรืออาจแสดงค่าฟังก์ชันมาร์จินัลดังตาราง

		x			
		0	1	2	$f_Y(y)$
y	0	0	0.3	0.1	0.4
	1	0.1	0	0.2	0.3
	2	0.1	0.2	0	0.3
	$f_X(x)$	0.2	0.5	0.3	1

ตัวอย่าง 4.4.9 จากตัวอย่าง 4.4.5 ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & \text{เมื่อ } 0 < x < 2 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหา $f_X(x)$ และ $f_Y(y)$

แนวคำตอบ

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 \frac{x(1+3y^2)}{4}\,dy = x\left[\frac{y+y^3}{4}\right]_{y=0}^{y=1} = \frac{x}{2}$$

ดังนั้น $f_X(x) = \dfrac{x}{2}$ เมื่อ $0 < x < 2$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{x(1+3y^2)}{4}\,dx = \frac{1+3y^2}{4}\left[\frac{x^2}{2}\right]_{x=0}^{x=2} = \frac{1}{2}(1+3y^2)$$

ดังนั้น $f_Y(y) = \dfrac{1}{2}(1+3y^2)$ เมื่อ $0 < y < 1$

ถ้าให้ A เป็นเหตุการณ์ที่ตัวแปรสุ่ม $X = x$ และ B เป็นเหตุการณ์ที่ตัวแปรสุ่ม $Y = y$ ความน่าจะเป็นแบบมีเงื่อนไขคือ

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

หรือเขียนแทนด้วย

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} \qquad \text{เมื่อ } f_X(x) > 0$$

ในทำนองเดียวกัน

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} \qquad \text{เมื่อ } f_Y(y) > 0$$

ตัวอย่าง 4.4.10 จากตัวอย่าง 4.4.8 จงหา

1. $f_{X|Y}(x \mid y)$ และ $f_{Y|X}(y \mid x)$
2. $P(X \neq 1 \mid Y = 1)$

แนวคำตอบ จากตาราง

		$x=0$	$x=1$	$x=2$	$f_Y(y)$
y	0	0	0.3	0.1	0.4
	1	0.1	0	0.2	0.3
	2	0.1	0.2	0	0.3
	$f_X(x)$	0.2	0.5	0.3	1

จากนิยาม $f_{Y|X}(y \mid x) = \dfrac{f(x,y)}{f_X(x)}$ และ $f_{X|Y}(x \mid y) = \dfrac{f(x,y)}{f_Y(y)}$ จะได้ว่า

	$f_{Y\|X}(y \mid x)$	$x=0$	1	2
	0	0	$\frac{3}{5}$	$\frac{1}{3}$
y	1	$\frac{1}{2}$	0	$\frac{2}{3}$
	2	$\frac{1}{2}$	$\frac{2}{5}$	0

	$f_{X\|Y}(x \mid y)$	$x=0$	1	2
	0	0	$\frac{3}{4}$	$\frac{1}{4}$
y	1	$\frac{1}{3}$	0	$\frac{2}{3}$
	2	$\frac{1}{3}$	$\frac{2}{3}$	0

ดังนั้น

$$\begin{aligned} P(X \neq 1 \mid Y=1) &= P(X=0 \mid Y=1) + P(X=2 \mid Y=1) \\ &= f_{X|Y}(0 \mid 1) + f_{X|Y}(2 \mid 1) \\ &= \frac{1}{3} + \frac{2}{3} = 1 \end{aligned}$$

ตัวอย่าง 4.4.11 จากตัวอย่าง 4.4.5 ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & \text{เมื่อ } 0 < x < 2 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหา

1. $f_{X|Y}(x \mid y)$ และ $f_{Y|X}(y \mid x)$

   แนวคำตอบ จากตัวอย่าง 4.4.9 จะได้ว่า

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{x(1+3y^2)}{4}}{\frac{(1+3y^2)}{2}} = \frac{x}{2} \qquad \text{เมื่อ } 0 < x < 2$$

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{x(1+3y^2)}{4}}{\frac{x}{2}} = \frac{(1+3y^2)}{2} \qquad \text{เมื่อ } 0 < y < 1$$

2. $P(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3})$

   แนวคำตอบ จะได้ว่า

$$\begin{aligned} P\left(\frac{1}{4} < X < \frac{1}{2} \middle| Y = \frac{1}{3}\right) &= \int_{\frac{1}{4}}^{\frac{1}{2}} f_{X|Y}\left(x \middle| \frac{1}{3}\right) dx \\ &= \int_{\frac{1}{4}}^{\frac{1}{2}} \frac{x}{2} dx \\ &= \left[\frac{x^2}{4}\right]_{\frac{1}{4}}^{\frac{1}{2}} \\ &= \frac{1}{16} - \frac{1}{64} = \frac{3}{64} \end{aligned}$$

บทนิยาม 4.4.12 ให้ X และ Y เป็นตัวแปรสุ่มซึ่งมีฟังก์ชันความน่าจะเป็นร่วมกันคือ f จะกล่าวว่าตัวแปรสุ่ม X และ Y เป็น**อิสระต่อกัน** ก็ต่อเมื่อ

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \text{สำหรับทุกค่า } x \text{ และ } y$$

ตัวอย่าง 4.4.13 จากตัวอย่าง 4.4.8 จงตรวจสอบว่าตัวแปรสุ่มทั้งคู่เป็นอิสระต่อกันหรือไม่
แนวคำตอบ จากตาราง

		$x=0$	$x=1$	$x=2$	$f_Y(y)$
y	0	0	0.3	0.1	0.4
	1	0.1	0	0.2	0.3
	2	0.1	0.2	0	0.3
	$f_X(x)$	0.2	0.5	0.3	1

จะเห็นว่า $f(0,0) = 0 \neq 0.08 = 0.2(0.4) = f_X(0)f_Y(0)$
ดังนั้นตัวแปรสุ่ม X และ Y ไม่เป็นอิสระต่อกัน

ตัวอย่าง 4.4.14 จากตัวอย่าง 4.4.5 ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & \text{เมื่อ } 0 < x < 2 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงตรวจสอบว่าตัวแปรสุ่ม X และ Y เป็นอิสระต่อกันหรือไม่
แนวคำตอบ จากตัวอย่าง 4.4.9 สำหรับ $0 < x < 2$ และ $0 < y < 1$ จะได้ว่า

$$f(x,y) = \frac{x(1+3y^2)}{4} = \frac{x}{2} \cdot \frac{1+3y^2}{2} = f_X(x) \cdot f_Y(y)$$

ดังนั้นตัวแปรสุ่ม X และ Y เป็นอิสระต่อกัน

แบบฝึกหัด 4.4

1. สุ่มเลือกผลไม้ 4 ผล จากถุงใบหนึ่งที่บรรจุผลไม้ 3 ชนิด ส้ม 3 ผล มะม่วง 2 ผล มังคุด 3 ผล จงหาการแจกแจงความน่าจะเป็นร่วมของจำนวนส้มกับจำนวนมะม่วงที่หยิบได้

2. ตัวแปรสุ่ม X และ Y มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y) = \begin{cases} 4xy & \text{เมื่อ } 0 < x < 1 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

   2.1 จงหาความน่าจะเป็นที่ $0 \le X \le \frac{3}{4}$ และ $\frac{1}{8} \le X \le \frac{1}{2}$

   2.2 จงหาความน่าจะเป็นที่ $X > Y$

3. ตัวแปรสุ่ม X และ Y มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y) = \begin{cases} k(x^2+y^2) & \text{เมื่อ } 0 < x < 2 \text{ และ } 1 < y < 4 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

   3.1 จงหา k

   3.2 จงหา $P(1 \le X \le 2)$

   3.3 จงหา $P(1 < X < 2, 2 < Y \le 3)$

   3.4 จงหา $P(X + Y > 4)$

4. ตัวแปรสุ่ม X และ Y มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y) = \begin{cases} x^2 + \frac{1}{3}xy & \text{เมื่อ } 0 < x < 1 \text{ และ } 0 < y < 2 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

   4.1 จงหา $P(X > \frac{1}{2})$

   4.2 จงหา $P(Y < X)$

5. ให้ X และ Y เป็นตัวแปรสุ่ม ที่มีการแจกแจงความน่าจะเป็นร่วมกันดังตาราง

		x = 2	x = 4
y	0	0.10	0.15
	1	0.20	0.30
	2	0.10	0.15

   จงหาการแจกแจงมาร์จินัล และตรวจสอบว่า X และ Y เป็นอิสระต่อกันหรือไม่

6. ตัวแปรสุ่ม X, Y และ Z มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y,z) = \begin{cases} \frac{4}{9}xyz^2 & \text{เมื่อ } 0 < x < 1,\ 0 < y < 1,\ 0 < z < 3 \\ 0 & \text{เมื่อ } x, y, z \text{ มีค่าอื่น ๆ} \end{cases}$$

   6.1 จงหามาร์จินัลของ Y

   6.2 จงหา $P(\frac{1}{4} < X < \frac{1}{2}, Y > \frac{1}{3}, 1 < Z < 2)$

4.5 การคาดคะเนทางคณิตศาสตร์

บทนิยาม 4.5.1 ให้ X เป็นตัวแปรสุ่มที่มีฟังก์ชันความน่าจะเป็น f **ค่าคาดคะเน** (Expected value) ของ X เขียนแทนด้วย $E(X)$ นิยามโดย

1. $E(X) = \sum_x xf(x)$ เมื่อ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง
2. $E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$ เมื่อ X เป็นตัวแปรสุ่มต่อเนื่อง

ในกรณี $E(X) = 0$ หมายถึง **เกมยุติธรรม** (fair game) หรือไม่ได้เปรียบและไม่เสียเปรียบ

ตัวอย่าง 4.5.2 เลือกกรรมการ 3 คนอย่างสุ่มจากผู้สมัครทั้งหมด 7 คน ซึ่งเป็นครูคณิตศาสตร์ 4 คน และครูวิทยาศาสตร์ 3 คน จงหาค่าคาดคะเนของจำนวนครูคณิตศาสตร์ที่ได้รับเลือกเป็นกรรมการ

แนวคำตอบ ให้ X คือจำนวนครูคณิตศาสตร์ที่ได้รับเลือกเป็นกรรมการ มีค่าเป็น $0, 1, 2, 3$ และ

x	0	1	2	3
$f(x)$	$\frac{{}^4C_0 \cdot {}^3C_3}{{}^7C_3} = \frac{1}{35}$	$\frac{{}^4C_1 \cdot {}^3C_2}{{}^7C_3} = \frac{12}{35}$	$\frac{{}^4C_2 \cdot {}^3C_1}{{}^7C_3} = \frac{18}{35}$	$\frac{{}^4C_3 \cdot {}^3C_0}{{}^7C_3} = \frac{4}{35}$
$xf(x)$	0	$\frac{12}{35}$	$\frac{36}{35}$	$\frac{12}{35}$

นั่นคือ

$$E(X) = 0f(0) + 1f(1) + 2f(1) + 3f(3) = 0 + \frac{12}{35} + \frac{36}{35} + \frac{12}{35} = \frac{60}{35} = 1.7$$

ดังนั้นค่าคาดคะเนของจำนวนครูคณิตศาสตร์ที่ได้รับเลือกเป็นกรรมการ 1.7 คน

ตัวอย่าง 4.5.3 เล่นเกมโยนเหรียญ 3 อันครั้งหนึ่ง ผู้เล่นจะได้รับเงิน 5 บาท จากเจ้ามือถ้าปรากฏว่าเหรียญขึ้นหัวทั้งหมดหรือขึ้นก้อยทั้งหมด ในกรณีอื่น ๆ เขาจะต้องจ่ายให้เจ้ามือ 3 บาท จงหาค่าคาดคะเนของเงินที่ได้รับ

แนวคำตอบ ให้ X คือจำนวนเงินที่ได้รับในการเล่นเกม มีค่าเป็น $5, -3$ และ

x	5	-3
$f(x)$	$\frac{1}{4}$	$\frac{3}{4}$
$xf(x)$	$\frac{5}{4}$	$-\frac{9}{4}$

นั่นคือ

$$E(X) = 5f(5) - 3f(-3) = \frac{5}{4} - \frac{9}{4} = -1$$

ดังนั้นค่าคาดคะเนของเงินที่ได้รับเท่ากับ -1 บาท หรือเสีย 1 บาทในการเล่นแต่ละครั้ง

ตัวอย่าง 4.5.4 ในการโยนลูกเต๋า 1 ลูก จะได้รับเงินเท่ากับจำนวนแต้มที่ขึ้น (หน่วยเป็นบาท) จงหาเงินที่คาดว่าจะได้รับ

แนวคำตอบ ให้ X คือจำนวนเงินที่ได้รับในการโยนลูกเต๋า มีค่าเป็น $1, 2, 3, 4, 5, 6$ โดยที่

$$f(x) = P(X = x) = \frac{1}{6} \qquad \text{เมื่อ} \quad x = 1, 2, 3, 4, 5, 6$$

จะได้ว่า

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{3} = 2.33$$

ดังนั้นเงินที่คาดว่าจะได้รับเท่ากับ 2.33 บาท

ตัวอย่าง 4.5.5 ถ้าสลากชนิดหนึ่งแบ่งขายในราคา 5 บาทต่อฉบับเป็นจำนวน 8,000 ฉบับ และมีรางวัลเพียง 1 รางวัลเป็นเงิน 12,000 บาท นายแดงซื้อสลากมา 2 ฉบับ จงหาเงินที่คาดว่านายแดงจะได้รับ

แนวคำตอบ ให้ X คือจำนวนเงินที่นายแดงได้รับจากการซื้อสลาก

$$\text{ถูกรางวัล} \quad X = 12000 - 10 = 11990 \text{ บาท} \quad \text{และ } P(X = 11990) = \frac{2}{8000}$$

$$\text{ไม่ถูกรางวัล} \quad X = -10 \text{ บาท} \quad \text{และ } P(X = -10) = \frac{7998}{8000}$$

จะได้ว่า

$$E(X) = 11990 \cdot \frac{2}{8000} - 10 \cdot \frac{7998}{8000} = -7$$

ดังนั้นเงินที่คาดว่านายแดงจะได้รับ -7 บาท หรือเสีย 7 บาท

ตัวอย่าง 4.5.6 บริษัทประกันชีวิตแห่งหนึ่งต้องการหาเบี้ยประกันชีวิตต่อปีที่จะเก็บลูกค้าที่มีทุนประกัน 100,000 บาท ของผู้เอาประกันชายที่มีอายุ 25 ปี ถ้าคุ้มครอง 1 ปี จากตารางชีพพบว่าชายอายุ 25 ปี จำนวน 10,000 คนจะตายก่อนครบอายุ 26 ปี จำนวน 3 คน จงหาเบี้ยประกันชีวิตต่อปีที่จะเก็บจากชายอายุ 25 ปี (โดยไม่พิจารณาค่าบริหารงานของบริษัท)

แนวคำตอบ ให้ X คือกำไรที่จะได้รับในการทำประกันภัยของชายอายุ 25 ปีคุมครอง 1 ปี
และ m คือเบี้ยประกันภัย

x	m	$m - 100000$
$f(x)$	0.9997	0.0003

นั่นคือ

$$\begin{aligned} E(X) &= m(0.9997) + (m - 100000)(0.0003) \\ 0 &= m - 30 \\ m &= 30 \end{aligned}$$

ดังนั้นต้องเก็บเบี้ยประกันภัยอย่างน้อย 30 บาท

ตัวอย่าง 4.5.7 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \begin{cases} \dfrac{2000}{x^3} & \text{เมื่อ } x > 100 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาค่าคาดคะเนของ X

แนวคำตอบ

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} xf(x)\,dx = \int_{100}^{\infty} x \cdot \frac{2000}{x^3}\,dx \\ &= \lim_{t\to\infty} \int_{100}^{t} 2000x^{-2}\,dx = \lim_{t\to\infty} \left[-\frac{2000}{x}\right]_{100}^{t} \\ &= \lim_{t\to\infty} \left[-\frac{2000}{t} + \frac{2000}{100}\right] = 20 \end{aligned}$$

บทนิยาม 4.5.8 ให้ X เป็นตัวแปรสุ่มที่มีฟังก์ชันความน่าจะเป็น f และ u เป็นฟังก์ชันของ X ค่าคาดคะเนของ $u(x)$ เขียนแทนด้วย $E[u(X)]$ นิยามโดย

1. $E[u(X)] = \sum_x u(x)f(x)$ เมื่อ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง
2. $E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)$ เมื่อ X เป็นตัวแปรสุ่มต่อเนื่อง

ตัวอย่าง 4.5.9 ให้ X เป็นจำนวนหัวที่เกิดจากการโยนเหรียญที่เที่ยงตรง 3 อัน โดยที่

$$u(X) = X^2 + 1 \quad \text{และ} \quad v(X) = (X-1)^2$$

จงหา $E[u(X)]$ และ $E[v(X)]$

แนวคำตอบ จากตารางแจกแจงความน่าจะเป็น

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

จะได้ว่า

$$\begin{aligned} E[u(X)] &= E[X^2+1] = \sum_{x=0}^{3} (x^2+1)f(x) \\ &= (0^2+1)f(0) + (1^2+1)f(1) + (2^2+1)f(2) + (3^2+1)f(3) \\ &= 1\cdot\frac{1}{8} + 2\cdot\frac{3}{8} + 5\cdot\frac{3}{8} + 10\cdot\frac{1}{8} = 4 \\ E[v(X)] &= E[(X-1)^2] = \sum_{x=0}^{3} (x-1)^2 f(x) \\ &= (0-1)^2 f(0) + (1-1)^2 f(1) + (2-1)^2 f(2) + (3-1)^2 f(3) \\ &= 1\cdot\frac{1}{8} + 0\cdot\frac{3}{8} + 1\cdot\frac{3}{8} + 4\cdot\frac{1}{8} = 1 \end{aligned}$$

ตัวอย่าง 4.5.10 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \begin{cases} \dfrac{x^2}{3} & \text{เมื่อ } -1 < x < 2 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาค่าคาดคะเน $E(2X-1)$

แนวคำตอบ จะได้ว่า

$$\begin{aligned} E(2X-1) &= \int_{-\infty}^{\infty} (2x-1) f(x)\, dx \\ &= \int_{-1}^{2} (2x-1) \cdot \frac{x^2}{3}\, dx = \int_{-1}^{2} \frac{2x^3}{3} - \frac{x^2}{3}\, dx \\ &= \left[\frac{x^4}{6} - \frac{x^3}{9} \right]_{-1}^{2} \\ &= \left[\frac{16}{6} - \frac{8}{9} \right] - \left[\frac{1}{6} + \frac{1}{9} \right] = \frac{3}{2} \end{aligned}$$

ทฤษฎีบท 4.5.11 ให้ X เป็นตัวแปรสุ่ม ถ้า a และ b เป็นค่าคงตัว จะได้ว่า

$$E(aX+b) = aE(X) + b$$

บทพิสูจน์. พิสูจน์เฉพาะกรณีตัวแปรสุ่มต่อเนื่อง

$$\begin{aligned} E(aX+b) &= \int_{-\infty}^{\infty} (ax+b) f(x)\, dx \\ &= \int_{-\infty}^{\infty} axf(x) + bf(x)\, dx \\ &= a\int_{-\infty}^{\infty} xf(x)\, dx + b\int_{-\infty}^{\infty} f(x)\, dx \\ &= aE(X) + b(1) = aE(X) + b \end{aligned}$$

□

ทฤษฎีบท 4.5.12 ให้ X เป็นตัวแปรสุ่ม โดยที่ u และ v เป็นฟังก์ชันของ X จะได้ว่า

$$E[u(X) \pm v(X)] = E[u(X)] \pm E[v(X)]$$

บทพิสูจน์. พิสูจน์เฉพาะกรณีตัวแปรสุ่มต่อเนื่อง

$$\begin{aligned} E[u(X) \pm v(X)] &= \int_{-\infty}^{\infty} (u(x) \pm v(x)) f(x)\, dx \\ &= \int_{-\infty}^{\infty} u(x)f(x) \pm v(x)f(x)\, dx \\ &= \int_{-\infty}^{\infty} u(x)f(x)\, dx \pm \int_{-\infty}^{\infty} v(x)f(x)\, dx \\ &= E[u(X)] \pm E[v(X)] \end{aligned}$$

□

ตัวอย่าง 4.5.13 ให้ X เป็นตัวแปรสุ่มซึ่ง $E(X) = 4$ และ $E(X^2) = 19$ จงหา

1. $E(2X+1)$
2. $E[(X-1)^2]$
3. $E[(X-E(X))^2]$

แนวคำตอบ จะได้ว่า

$$\begin{aligned}
E(2X+1) &= 2E(X)+1 = 2(4)+1 = 9 \\
E[(X-1)^2] &= E(X^2-2X+1) = E(X^2)-2E(X)+1 \\
&= 19-2(4)+1 = 12 \\
E[(X-E(X))^2] &= E[(X-4)^2] = E(X^2-8X+16) \\
&= E(X^2)-8E(X)+16 \\
&= 19-8(4)+16 = 3
\end{aligned}$$

บทนิยาม 4.5.14 ให้ X และ Y เป็นตัวแปรสุ่มที่มีฟังก์ชันความน่าจะเป็นร่วมกันคือ f **ค่าคาดคะเน**ของฟังก์ชัน $u(X,Y)$ เขียนแทนด้วย $E[u(X,Y)]$ นิยามโดย

1. $E[u(X,Y)] = \sum_x \sum_y u(x,y)f(x,y)$ เมื่อ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง
2. $E[u(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y)f(x,y)\,dxdy$ เมื่อ X และ Y เป็นตัวแปรสุ่มต่อเนื่อง

ตัวอย่าง 4.5.15 จากตัวอย่าง 4.4.8 การแจกแจงความน่าจะเป็นร่วมกันของ X และ Y ดังตาราง

$$\begin{array}{cc|ccc}
 & & & x & \\
 & & 0 & 1 & 2 \\
\hline
 & 0 & 0 & 0.3 & 0.1 \\
y & 1 & 0.1 & 0 & 0.2 \\
 & 2 & 0.1 & 0.2 & 0
\end{array}$$

จงหา $E(XY)$

แนวคำตอบ

$$\begin{aligned}
E(XY) &= \sum_x \sum_y (xy)f(x,y) \\
&= (0\cdot 0)f(0,0)+(0\cdot 1)f(0,1)+(0\cdot 2)f(0,2)+ \\
&\quad (1\cdot 0)f(1,0)+(1\cdot 1)f(1,1)+(1\cdot 2)f(1,2)+ \\
&\quad (2\cdot 0)f(2,0)+(2\cdot 1)f(2,1)+(2\cdot 2)f(2,2) \\
&= 1\cdot 0+2\cdot 0.2+2\cdot 0.2+4\cdot 0 = 0.8
\end{aligned}$$

ตัวอย่าง 4.5.16 จากตัวอย่าง 4.4.5 ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & \text{เมื่อ } 0 < x < 2 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหา $E\left(\dfrac{Y}{X}\right)$

แนวคำตอบ

$$\begin{aligned} E\left(\frac{Y}{X}\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y}{x} f(x,y)\,dx\,dy = \int_0^1\int_0^2 \frac{y}{x}\cdot\frac{x(1+3y^2)}{4}\,dx\,dy \\ &= \int_0^1\int_0^2 \frac{(y+3y^3)}{4}\,dx\,dy = \int_0^1 \left[\frac{(y+3y^3)x}{4}\right]_{x=0}^{x=2} dy \\ &= \int_0^1 \frac{y+3y^3}{2}\,dy = \left[\frac{y^2}{4}+\frac{3y^4}{8}\right]_0^1 \\ &= \left[\frac{1}{4}+\frac{3}{8}\right] - 0 = \frac{5}{8} \end{aligned}$$

กรณี $u(X,Y) = X$ จะได้ว่า

1. สำหรับ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$E(X) = \sum_x\sum_y x f(x,y) = \sum_x x f_X(x)$$

2. สำหรับ X และ Y เป็นตัวแปรสุ่มต่อเนื่อง

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dxdy = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

กรณี $u(X,Y) = Y$ จะได้ว่า

1. สำหรับ X และ Y เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$E(Y) = \sum_x\sum_y y f(x,y) = \sum_y y f_Y(y)$$

2. สำหรับ X และ Y เป็นตัวแปรสุ่มต่อเนื่อง

$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dxdy = \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

ทฤษฎีบท 4.5.17 ให้ X และ Y เป็นตัวแปรสุ่ม โดยที่ u และ v เป็นฟังก์ชันของ X และ Y จะได้ว่า

$$E[u(X,Y) \pm v(X,Y)] = E[u(X,Y)] \pm E[v(X,Y)]$$

บทพิสูจน์. พิสูจน์เฉพาะกรณีตัวแปรสุ่มต่อเนื่อง

$$\begin{aligned}
E[u(X,Y) \pm v(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(u(x,y) \pm v(x,y))f(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y)f(x,y) \pm v(x,y)f(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y)f(x,y)\,dxdy \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v(x,y)f(x,y)\,dxdy \\
&= E[u(X,Y)] \pm E[v(X,Y)]
\end{aligned}$$

□

ทฤษฎีบท 4.5.18 ถ้า X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน จะได้ว่า

$$E(XY) = E(X)E(Y)$$

บทพิสูจน์. พิสูจน์เฉพาะกรณีตัวแปรสุ่มต่อเนื่อง สมมติ X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน จะได้ว่า $f(x,y) = f_X(x) \cdot f_Y(y)$ ดังนั้น

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x) \cdot f_Y(y)\,dxdy \\
&= \int_{-\infty}^{\infty} yf_Y(y)\left(\int_{-\infty}^{\infty} xf_X(x)\,dx\right)dy \\
&= \int_{-\infty}^{\infty} yf_Y(y)E(X)\,dy \\
&= E(X)\int_{-\infty}^{\infty} yf_Y(y)\,dy \\
&= E(X)E(Y)
\end{aligned}$$

□

ตัวอย่าง 4.5.19 จากตัวอย่าง 4.4.5 ตัวแปรสุ่ม X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน มีฟังก์ชันความหนาแน่นร่วมกันคือ

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & \text{เมื่อ } 0 < x < 2 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงแสดงว่า $E(XY) = E(X)E(Y)$

แนวคำตอบ

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy = \int_0^1\int_0^2 xy\cdot\frac{x(1+3y^2)}{4}\,dx\,dy \\ &= \int_0^1\int_0^2 \frac{x^2(y+3y^3)}{4}\,dx\,dy = \int_0^1 \left[\frac{x^3(y+3y^3)}{12}\right]_{x=0}^{x=2} dy \\ &= \int_0^1 \frac{2(y+3y^3)}{3}\,dy = \left[\frac{y^2}{3}+\frac{y^4}{2}\right]_0^1 \\ &= \left[\frac{1}{3}+\frac{1}{2}\right] - 0 = \frac{5}{6} \end{aligned}$$

และ

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dx\,dy = \int_0^1\int_0^2 x\cdot\frac{x(1+3y^2)}{4}\,dx\,dy \\ &= \int_0^1\int_0^2 \frac{x^2(1+3y^2)}{4}\,dx\,dy = \int_0^1 \left[\frac{x^3(1+3y^2)}{12}\right]_{x=0}^{x=2} dy \\ &= \int_0^1 \frac{2(1+3y^2)}{3}\,dy = \left[\frac{2y}{3}+\frac{2y^3}{3}\right]_0^1 \\ &= \left[\frac{2}{3}+\frac{2}{3}\right] - 0 = \frac{4}{3} \\ E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dx\,dy = \int_0^1\int_0^2 y\cdot\frac{x(1+3y^3)}{4}\,dx\,dy \\ &= \int_0^1\int_0^2 \frac{x(y+3y^3)}{4}\,dx\,dy = \int_0^1 \left[\frac{x^2(y+3y^3)}{8}\right]_{x=0}^{x=2} dy \\ &= \int_0^1 \frac{(y+3y^3)}{2}\,dy = \left[\frac{y^2}{4}+\frac{3y^4}{8}\right]_0^1 \\ &= \left[\frac{1}{4}+\frac{3}{8}\right] - 0 = \frac{5}{8} \end{aligned}$$

จะได้ว่า

$$E(X)E(Y) = \frac{4}{3}\cdot\frac{5}{8} = \frac{5}{6} = E(XY)$$

ตัวอย่าง 4.5.20 ถ้าเก็บข้อมูลครัวเรือนตัวอย่างในกรุงเทพฯมา 200 ครอบครัว และถามถึงจำนวนบุตรที่มี (X) และจำนวนบุตรชายที่มี (Y) ได้ดังข้อมูล

จำนวนบุตรชาย (Y)	จำนวนบุตร (X) 0	1	2	3	รวม
0	50	5	10	5	70
1	0	40	10	0	50
2	0	0	55	5	60
3	0	0	0	20	20
รวม	50	45	75	30	200

1. จงสร้างตารางความน่าจะเป็นร่วมกันของ X และ Y

   แนวคำตอบ

y \ x	0	1	2	3	$f_Y(y)$
0	0.25	0.025	0.05	0.025	0.35
1	0	0.2	0.05	0	0.25
2	0	0	0.275	0.025	0.30
3	0	0	0	0.1	0.10
$f_X(x)$	0.25	0.225	0.375	0.15	1

2. จงหาโอกาสที่ครอบครัวหนึ่งจะมีบุตร 2 คน

   แนวคำตอบ เนื่องจาก

$$P(X=2) = \sum_{y=0}^{3} f(2,y) = f_X(2) = 0.375$$

   ดังนั้นโอกาสที่ครอบครัวหนึ่งจะมีบุตร 2 คน เท่ากับ 0.375

3. จงหาโอกาสที่ครอบครัวหนึ่งจะมีบุตรชาย 1 คน

   แนวคำตอบ เนื่องจาก

$$P(Y=1) = \sum_{x=0}^{3} f(x,1) = f_Y(1) = 0.25$$

   ดังนั้นโอกาสที่ครอบครัวหนึ่งจะมีบุตรชาย 1 คน เท่ากับ 0.25

4. จงหาโอกาสที่ครอบครัวหนึ่งจะมีบุตรชาย 2 คน โดยกำหนดว่าเขามีบุตร 3 คน

   แนวคำตอบ เนื่องจาก

$$P(Y=2 \mid X=3) = \frac{f(3,2)}{f_X(3)} = \frac{0.025}{0.15} = \frac{1}{6} = 0.1667$$

   ดังนั้นโอกาสที่ครอบครัวหนึ่งจะมีบุตรชาย 2 คน ถ้าเขามีบุตร 3 คน เท่ากับ 0.1667

5. จงหาค่าคาดคะเนจำนวนบุตรสาวของแต่ละครอบครัว

   แนวคำตอบ จำนวนบุตรสาวของแต่ละครอบครัวเท่ากับ $X - Y$ และ

$$\begin{aligned}
E(X-Y) &= E(X) - E(Y) \\
&= \sum_{x=0}^{3} x f_X(x) - \sum_{y=0}^{3} y f_Y(y) \\
&= 0(0.25) + 1(0.225) + 2(0.375) + 3(0.15) \\
&\quad - 0(0.35) - 1(0.25) - 2(0.30) - 3(0.10) \\
&= 0.275
\end{aligned}$$

   ดังนั้นค่าคาดคะเนจำนวนบุตรสาวของแต่ละครอบครัว เท่ากับ 0.275

แบบฝึกหัด 4.5

1. ในการลงทุนของชายคนหนึ่งปรากฏว่าใน 1 ปีเขามีกำไร 60,000 บาท ด้วยความน่าจะเป็น 0.3 หรือขาดทุน 20,000 บาท ด้วยความน่าจะเป็น 0.7 จงหาค่าคาดคะเนของการลงทุนนี้

2. การเล่นเกมพนันเกมหนึ่ง ถ้าชายคนหนึ่งดึงไพ่ออกจากสำรับซึ่งมีไพ่ 52 ใบ เป็น J หรือ Q เขาจะได้รับเงิน 2 บาท และถ้าดึงไพ่ออกมาเป็น K หรือ A เขาจะได้รับเงิน 5 บาท แต่ถ้าดึงไพ่ใบอื่น ๆ เขาจะไม่ได้รับเงินจากเจ้ามือ อยากทราบว่าเขาควรจ่ายเงินค่าเกมนี้เท่าใดจึงจะทำให้เกมยุติธรรม

3. ถ้าบริษัทประกันภัยแห่งหนึ่งเก็บเบี้ยประกันต่อปีเป็นเงิน 29,000 บาท สำหรับผู้เอาประกันชีวิตที่มีเบี้ยประกันตลอดชีวิตที่มีทุนประกัน 1 ล้านบาท ถ้าตารางชีพแสดงโอกาสที่ผู้เอาประกันที่ทำประกันข้างต้นจะเสียชีวิตในปีหน้าเป็น 0.001 จงหากำไรที่คาดว่าจะได้เงินจากกรมธรรม์ดังกล่าว

4. ถ้านำรถยนต์ราคา 500,000 บาทไปทำประกันภัยในกรณีรถหายกับบริษัทประกันภัยแห่งหนึ่ง ซึ่งทุนประกัน 500,000 บาท โดยต้องจ่ายเบี้ยประกัน D บาทต่อปี ถ้าโอกาสที่รถจะถูกขโมยในปีที่ทำประกันเป็น 1% ทางบริษัทประกันภัยควรคิดเบี้ยประกันเท่าใด จึงจะได้กำไรเฉลี่ย 10,000 บาทต่อปี

5. ถ้าตัวแปรสุ่ม X มีฟังก์ชันความน่าจะเป็นคือ

$$f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{3-x} \quad \text{เมื่อ} \quad x = 0, 1, 2, 3$$

   จงหา $E(X)$

6. จงหาค่าคาดคะเนของตัวแปรสุ่ม X ซึ่งมี p.d.f. คือ $f(x) = \begin{cases} 2(1-x) & \text{เมื่อ } 0 < x < 1 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$

7. กำหนดให้ X แทนแต้มที่ได้จากการทอดลูกเต๋าที่เที่ยงตรงหนึ่งลูก จงหา $E(2X^2 - 5)$

8. ให้ X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันมีการแจกแจงมาร์จินัลเป็น

$$f_X(x) = \frac{8}{x^3} \quad \text{เมื่อ } x > 2 \qquad \text{และ} \qquad f_Y(y) = 2y \quad \text{เมื่อ } 0 < y < 1$$

   จงหาค่าคาดคะเนของ XY

9. การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังตาราง

x	-3	0	6	9
$f(x)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$

9.1 $E(X)$

9.2 $E(X^2)$

9.3 $E[(2X-1)^2]$

9.4 $E[E(X)-X]$

10. ให้ X แทนแต้มที่ได้จากการทอดลูกเต๋าสีแดง และ Y แทนแต้มที่ได้จากการทอดลูกเต๋าสีเขียว จงหา

10.1 $E(X+Y)$

10.2 $E(X-Y)$

10.3 $E(XY)$

10.4 $E(X^2+Y^2)$

11. ให้ X เป็นตัวแปรสุ่มซึ่ง $E(X+2) = 10$ และ $E[(X-1)(X+1)] = 49$ จงหา

11.1 $E(X+2)$

11.2 $E(X^2)$

11.3 $E[(X+1)^2]$

11.4 $E[(1-X)(2+X)]$

12. ถ้าเก็บข้อมูลตัวอย่างจากนักศึกษาในมหาวิทยาลัยแห่งหนึ่ง 50 คน และถามถึงจำนวนโทรศัพท์ที่ใช้งาน (X) และจำนวนโทรศัพท์ยี่ห้อ iphone ที่มี (Y) ได้ดังข้อมูล

จำนวนโทรศัพท์ยี่ห้อ iphone (Y)	จำนวนโทรศัพท์ (X)			รวม
	0	1	2	
0	1	10	5	16
1	0	30	2	32
2	0	0	2	2
รวม	1	40	9	50

12.1 จงสร้างตารางความน่าจะเป็นร่วมกันของ X และ Y

12.2 จงหาโอกาสที่นักศึกษาจะมีโทรศัพท์ 2 เครื่อง

12.3 จงหาโอกาสที่นักศึกษาจะมีโทรศัพท์ iphone 1 เครื่อง ถ้ามีโทรศัพท์ 2 เครื่อง

12.4 จงหาค่าคาดคะเนของจำนวนโทรศัพท์ที่ไม่ใช่ iphone

4.6 ความแปรปรวนของตัวแปรสุ่ม

บทนิยาม 4.6.1 ให้ X เป็นตัวแปรสุ่ม เรียก $E(X)$ ว่า **ค่าเฉลี่ย (mean)** ของตัวแปรสุ่ม X เขียนแทนด้วย μ นั่นคือ

$$\mu = E(X)$$

ความแปรปรวน (variance) ของตัวแปรสุ่ม X เขียนแทนด้วย σ^2 หรือ σ_X^2 หรือ var(X) นิยามโดย

$$\sigma^2 = E[(X-\mu)^2]$$

และเรียก σ ว่า **ส่วนเบี่ยงเบนมาตรฐาน** (standard deviation) ของตัวแปรสุ่ม X

ในกรณี X เป็นตัวแปรสุ่มไม่ต่อเนื่อง

$$\mu = E(X) = \sum_x xf(x) \qquad \text{และ} \qquad \sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

ในกรณี X เป็นตัวแปรสุ่มต่อเนื่อง

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx \qquad \text{และ} \qquad \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

ทฤษฎีบท 4.6.2 ถ้า X เป็นตัวแปรสุ่ม จะได้ว่า $\sigma^2 = E(X^2) - \mu^2$

บทพิสูจน์.

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2] \\ &= E(X^2) - 2\mu E(X) - \mu^2 \\ &= E(X^2) - 2\mu^2 - \mu^2 = E(X^2) - \mu^2\end{aligned}$$

□

ตัวอย่าง 4.6.3 เลือกกรรมการ 2 คนอย่างสุ่มจากผู้สมัครทั้งหมด 5 คน ซึ่งเป็นครูคณิตศาสตร์ 3 คน และครูวิทยาศาสตร์ 2 คน จงหาค่าความแปรปรวนของจำนวนครูคณิตศาสตร์ที่ได้รับเลือกเป็นกรรมการ

แนวคำตอบ ให้ X คือจำนวนครูคณิตศาสตร์ที่ได้รับเลือกเป็นกรรมการ มีค่าเป็น $0, 1, 2$ และ

x	0	1	2	รวม
$f(x)$	$\dfrac{{}^2C_2 \cdot {}^3C_0}{{}^5C_2} = \dfrac{1}{10}$	$\dfrac{{}^2C_1 \cdot {}^3C_1}{{}^5C_2} = \dfrac{3}{5}$	$\dfrac{{}^4C_2 \cdot {}^3C_1}{{}^7C_3} = \dfrac{3}{10}$	1
$xf(x)$	0	$\dfrac{3}{5}$	$\dfrac{3}{5}$	$\dfrac{6}{5}$
$x^2f(x)$	0	$\dfrac{3}{5}$	$\dfrac{6}{5}$	$\dfrac{9}{5}$

นั่นคือ

$$\sigma^2 = E(X^2) - \mu^2 = \frac{9}{5} - \left(\frac{3}{5}\right)^2 = \frac{9}{25} = 0.36$$

ดังนั้นความแปรปรวนของจำนวนครูคณิตศาสตร์ที่ได้รับเลือกเป็นกรรมการเท่ากับ 0.36

ตัวอย่าง 4.6.4 ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ

$$f(x) = \begin{cases} 2(x-1) & \text{เมื่อ } 1 < x < 2 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่ม X

แนวคำตอบ

$$\begin{aligned} \mu = E(X) &= \int_{-\infty}^{\infty} x f(x)\, dx \\ &= \int_1^2 2x^2 - 2x\, dx = \left[\frac{2x^3}{3} - x^2\right]_1^2 \\ &= \left[\frac{16}{3} - 4\right] - \left[\frac{2}{3} - 1\right] = \frac{5}{3} \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\, dx \\ &= \int_1^2 2x^3 - 2x^2\, dx = \left[\frac{x^4}{2} - \frac{2x^3}{3}\right]_1^2 \\ &= \left[\frac{16}{2} - \frac{16}{3}\right] - \left[\frac{1}{2} - \frac{2}{3}\right] = \frac{17}{6} \\ \sigma^2 = E(X^2) - \mu^2 &= \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18} \end{aligned}$$

ตัวอย่าง 4.6.5 ให้ X เป็นตัวแปรสุ่ม โดยที่

$$E(X) = 5 \quad \text{และ} \quad E[(X+3)^2] = 75$$

จงหาความแปรปรวนของตัวแปรสุ่ม X

แนวคำตอบ เนื่องจาก $\mu = E(X) = 5$ และ

$$\begin{aligned} 75 &= E[(X+3)^2] \\ &= E[((X-5)+8)^2] \\ &= E[(X-5)^2 + 16(X-5) + 64] \\ &= E[(X-5)^2] + 16E(X-5) + 64 \\ &= E[(X-5)^2] + 16(E(X)-5) + 64 \\ &= E[(X-5)^2] + 16(5-5) + 64 \\ &= E[(X-5)^2] + 64 \end{aligned}$$

จะได้ว่า $E[(X-5)^2] = 11$ ดังนั้นความแปรปรวนของตัวแปรสุ่ม X เท่ากับ 11

บทนิยาม 4.6.6 ความแปรปรวนร่วม (covariance) ของตัวแปรสุ่ม X และ Y เขียนแทนด้วย σ_{XY} หรือ cov(X, Y) นิยามโดย

$$\text{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

เมื่อ $\mu_X = E(X)$ และ $\mu_Y = E(Y)$

จากบทนิยามจะได้ว่า

$$\text{cov}(X,Y) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)f(x,y) \qquad \text{เมื่อ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง}$$

$$\text{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)\,dxdy \quad \text{เมื่อ X เป็นตัวแปรสุ่มต่อเนื่อง}$$

ทฤษฎีบท 4.6.7 ถ้า X และ Y เป็นตัวแปรสุ่ม โดยที่ $\mu_X = E(X)$ และ $\mu_Y = E(Y)$ จะได้ว่า

$$\text{cov}(X,Y) = E(XY) - \mu_X\mu_Y$$

บทพิสูจน์.

$$\begin{aligned}
\text{cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\
&= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y \\
&= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y
\end{aligned}$$

□

ข้อสังเกต cov$(X, X) = \sigma_{XX} = 0$

บทแทรก 4.6.8 ให้ X และ Y เป็นตัวแปรสุ่ม จะได้ว่า

X และ Y เป็นอิสระต่อกัน ก็ต่อเมื่อ cov$(X, Y) = 0$

บทพิสูจน์. เห็นได้ชัดจากทฤษฎีบท 4.6.7 เนื่องจาก $E(XY) = E(X)E(Y) = \mu_X\mu_Y$ สำหรับ X และ Y ที่เป็นอิสระต่อกัน □

ตัวอย่าง 4.6.9 การแจกแจงความน่าจะเป็นร่วมกันของ X และ Y แสดงดังตาราง

		$x=0$	$x=1$	$x=2$
y	0	0.1	0.2	0.1
	1	0.1	0	0.2
	2	0.1	0.2	0

จงหา cov(X, Y)

แนวคำตอบ พิจารณาตารางต่อไปนี้

		$x=0$	$x=1$	$x=2$	$f_Y(y)$
y	0	0.1	0.2	0.1	0.4
	1	0.1	0	0.2	0.3
	2	0.1	0.2	0	0.3
	$f_X(x)$	0.3	0.4	0.3	1

จะได้ว่า

$$
\begin{aligned}
E(XY) &= \sum_x \sum_y xyf(x,y) \\
&= (1\cdot 1)f(1,1) + (1\cdot 2)f(1,2) + (2\cdot 1)f(2,1) + (2\cdot 2)f(2,2) \\
&= 0 + 2\cdot 0.2 + 2\cdot 0.2 + 0 = 0.8 \\
E(X) &= \sum_x xf_X(x) = 1f_X(1) + 2f_X(2) = 1\cdot 0.4 + 2\cdot 0.3 = 1 \\
E(Y) &= \sum_y yf_Y(y) = 1f_Y(1) + 2f_Y(2) = 1\cdot 0.3 + 2\cdot 0.3 = 0.9
\end{aligned}
$$

ดังนั้น

$$\text{cov}(X,Y) = 0.8 - 1(0.9) = -0.1$$

ตัวอย่าง 4.6.10 ตัวแปรสุ่ม X และ Y มีฟังก์ชันความหนาแน่นร่วมกันคือ

$$f(x,y) = \begin{cases} 2 & \text{เมื่อ } 0 < x < y \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหา σ_{XY}

แนวคำตอบ พิจารณา

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dxdy = \int_0^1\int_0^y 2xy\,dxdy \\ &= \int_0^1 \left[x^2y\right]_{x=0}^{x=y} dy = \int_0^1 y^3\,dy = \left[\frac{y^4}{4}\right]_0^1 = \frac{1}{4} \\ E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dxdy = \int_0^1\int_0^y 2x\,dxdy \\ &= \int_0^1 \left[x^2\right]_{x=0}^{x=y} dy = \int_0^1 y^2\,dy = \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{3} \\ E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dxdy = \int_0^1\int_0^y 2y\,dxdy \\ &= \int_0^1 \left[2xy\right]_{x=0}^{x=y} dy = \int_0^1 2y^2\,dy = \left[\frac{2y^3}{3}\right]_0^1 = \frac{2}{3} \end{aligned}$$

ดังนั้น

$$\sigma_{XY} = \frac{1}{4} - \frac{1}{3}\left(\frac{2}{3}\right) = \frac{1}{36}$$

ให้ $u(X)$ เป็นฟังก์ชันของตัวแปรสุ่ม X นิยามค่าเฉลี่ยและความแปรปรวนของ $u(X)$ คือ

$$\mu_{u(X)} = E[u(X)] \qquad \text{และ} \qquad \sigma^2_{u(X)} = E[(u(X) - \mu_{u(X)})^2]$$

ในทำนองเดียวกับทฤษฎีบท 4.6.2 จะได้ว่า

$$\sigma^2_{u(X)} = E[u^2(X)] - \mu^2_{u(X)}$$

ทฤษฎีบท 4.6.11 ถ้า X เป็นตัวแปรสุ่ม โดยที่ a, b เป็นค่าคงตัว จะได้ว่า

$$\sigma^2_{aX+b} = a^2\sigma^2_X$$

โดยเฉพาะกรณีที่ $a = 1$ จะได้ว่า $\sigma^2_{X+b} = \sigma^2_X$

บทพิสูจน์.

$$\begin{aligned}
\sigma^2_{aX+b} &= E[(aX+b)^2] - E^2(aX+b) \\
&= E[a^2X^2 + 2abX + b^2] - (aE(X) + b)^2 \\
&= a^2E(X^2) + 2abE(X) + b^2 - (a^2E^2(X) + 2abE(X) + b^2) \\
&= a^2(E(X^2) - E^2(X)) \\
&= a^2\sigma^2_X
\end{aligned}$$

□

ทฤษฎีบท 4.6.12 ถ้า X และ Y เป็นตัวแปรสุ่ม โดยที่ a, b เป็นค่าคงตัว จะได้ว่า

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\text{cov}(X, Y)$$

บทพิสูจน์.

$$\begin{aligned}
\sigma^2_{aX+bY} &= E[(aX+bY)^2] - E^2(aX+bY) \\
&= E[a^2X^2 + 2abXY + b^2Y^2] - (aE(X) + bE(Y))^2 \\
&= a^2E(X^2) + 2abE(XY) + b^2E(Y^2) - (a^2E^2(X) + 2abE(X)E(Y) + b^2E^2(Y)) \\
&= a^2(E(X^2) - E^2(X)) + b^2(E(Y^2) - E^2(Y)) + 2ab(E(XY) - E(X)E(Y)) \\
&= a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\text{cov}(X, Y)
\end{aligned}$$

□

บทแทรก 4.6.13 ถ้า X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน โดยที่ a, b เป็นค่าคงตัว จะได้ว่า

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

บทพิสูจน์. ถ้า X และ Y เป็นตัวแปรสุ่มที่เป็นอิสระต่อกัน จะได้ว่า $\sigma_{XY} = 0$
เห็นได้ชัดจากทฤษฎีบท 4.6.12 □

ตัวอย่าง 4.6.14 ให้ X และ Y เป็นตัวแปรสุ่ม ซึ่งมี $\sigma_X^2 = 2, \sigma_Y^2 = 3$ และ $\sigma_{XY} = 0.5$ จงหาความแปรปรวนของ Z เมื่อ

1. $Z = 2X + 5$
2. $Z = X + Y$
3. $Z = 3X + 2Y + 1$
4. $Z = 2X - Y + 3$

แนวคำตอบ

$$\begin{aligned}
\sigma_{2X+5} &= 2^2\sigma_X^2 = 4(2) = 8 \\
\sigma_{X+Y} &= \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} = 2 + 3 + 2(0.5) = 6 \\
\sigma_{3X+2Y+1} &= \sigma_{3X+2Y} = 3^2\sigma_X^2 + 2^2\sigma_Y^2 + 2(3)(2)\sigma_{XY} = 9(2) + 4(3) + 12(0.5) = 36 \\
\sigma_{2X-Y+3} &= \sigma_{2X-Y} = 2^2\sigma_X^2 + (-1)^2\sigma_Y^2 + 2(2)(-1)\sigma_{XY} = 4(2) + 1(3) - 4(0.5) = 9
\end{aligned}$$

ตัวอย่าง 4.6.15 ให้ X และ Y เป็นตัวแปรสุ่มซึ่งเป็นอิสระต่อกัน ซึ่งมี $\sigma_X^2 = 0.1$ และ $\sigma_Y^2 = 0.3$ จงหา

1. σ_{2X+Y}^2
2. σ_{5X-Y+1}^2

แนวคำตอบ

$$\begin{aligned}
\sigma_{2X+Y} &= 2^2\sigma_X^2 + \sigma_Y^2 = 4(0.1) + 0.3 = 0.7 \\
\sigma_{5X-Y+1} &= \sigma_{5X-Y} = 5^2\sigma_X^2 + (-1)^2\sigma_Y^2 = 25(0.1) + 1(0.3) = 2.8
\end{aligned}$$

ทฤษฎีบท 4.6.16 ถ้า $X_1, X_2, ..., X_n$ เป็นตัวแปรสุ่ม จะได้ว่า

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{var}(X_i) + \sum_{i\neq j} \text{cov}(X_i, X_j)$$

บทพิสูจน์. ทำเป็นแบบฝึกหัด $\square$

แบบฝึกหัด 4.6

1. จงหาค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่ม X เมื่อการแจกแจงความน่าจะเป็นคือ

x	0	2	4	6
$f(x)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$

2. ให้ฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X คือ

$$f(x) = \begin{cases} 2x^2 + \dfrac{1}{3} & \text{เมื่อ } 0 < x < 1 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่ม X

3. การแจกแจงความน่าจะเป็นร่วมกันของ X และ Y ดังตาราง

x	0	0	1	1
y	1	-1	0	1
$f(x,y)$	0.4	0.3	0.2	0.1

จงหา $E(X)$, $E(Y)$, σ_X^2, σ_Y^2 และ $\text{cov}(X,Y)$

4. ตัวแปรสุ่ม X และ Y เป็นตัวแปรสุ่ม มีฟังก์ชันความหนาแน่นร่วมกันคือ

$$f(x,y) = \begin{cases} x+y & \text{เมื่อ } 0 < x < 1 \text{ และ } 0 < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหา σ_{XY}

5. ให้ X และ Y เป็นตัวแปรสุ่มซึ่งมี $\sigma_X = 0.16, \sigma_Y = 0.25$ และ $\sigma_{XY} = 0.01$ จงหา σ_Z^2 เมื่อ

5.1 $Z = 2X + 3Y$

5.2 $Z = X - 5Y + 2$

5.3 $Z = X - 2Y$

5.4 $Z = 2X - Y + 5$

6. ถ้า X และ Y เป็นตัวแปรสุ่มโดยที่ a, b เป็นค่าคงตัว จงแสดงว่า $\text{cov}(aX, bY) = ab\text{cov}(X,Y)$

7. พนักงานขายคนหนึ่งได้เงินเดือน 10000 บาท และได้ค่าคอมมิชชันในการขายสินค้า 500 บาทต่อสินค้า 1 ชิ้น โดยเฉลี่ยแล้วพนักงานขายคนนี้ขายสินค้าได้ 15 ชิ้น ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 ชิ้น จงหารายได้ที่คาดไว้ต่อเดือน และส่วนเบี่ยงเบนมาตรฐานรายได้ต่อเดือนของพนักงานคนนี้

8. เลือกกรรมการ 3 คนอย่างสุ่มจากผู้สมัครทั้งหมด 10 คน ซึ่งเป็นนักเรียนชาย 6 คน และนักเรียนหญิง 4 คน จงหาค่าความแปรปรวนของจำนวนนักเรียนชายที่ได้รับเลือกเป็นกรรมการ

9. สลากจำนวน 20 ใบ มีรางวัล 2 ใบ รางวัลละ 100 บาท ผู้ออกสลากขายใบละ 20 บาท ถ้านายเอฟซื้อสลากมา 2 ใบ จงหาความแปรปรวนของเงินรางวัลที่นายเอฟจะได้รับ

10. ถ้าเก็บข้อมูลตัวอย่างจากนักศึกษาในมหาวิทยาลัยแห่งหนึ่ง 50 คน และถามถึงจำนวนโทรศัพท์ที่ใช้งาน (X) และจำนวนโทรศัพท์ยี่ห้อ iphone ที่มี (Y) ได้ดังข้อมูล

จำนวนโทรศัพท์ยี่ห้อ iphone (Y)	จำนวนโทรศัพท์ (X)			รวม
	0	1	2	
0	1	10	5	16
1	0	30	2	32
2	0	0	2	2
รวม	1	40	9	50

จงหาความแปรปรวนร่วมของตัวแปรสุ่ม X และ Y

11. ในการลงทุนในหุ้นและพันธบัตรมีข้อมูลดังนี้

หุ้น

ผลตอบแทน (บาท)	โอกาส
5,000	0.3
100,000	0.2
-60,000	0.5

พันธบัตร

ผลตอบแทน (บาท)	โอกาส
10,000	0.8
5,000	0.2

11.1 จงหาผลตอบแทนเฉลี่ยและความแปรปรวนของการลงทุนในหุ้น

11.2 จงหาผลตอบแทนเฉลี่ยและความแปรปรวนของการลงทุนในพันธบัตร

สรุป

ในบทนี้จะกล่าวถึงนิยามของตัวแปรสุ่มซึ่งคือฟังก์ชันค่าจริงซึ่งถูกกำหนดค่าโดยสมาชิกในปริภูมิตัวอย่าง โดยมีตัวแปรสุ่ม 2 ชนิดคือ ตัวแปรสุ่มไม่ต่อเนื่องและตัวแปรสุ่มต่อเนื่อง สำหรับตัวแปรสุ่มไม่ต่อเนื่อง X จะมี f เป็นฟังก์ชันความน่าจะเป็น (p.m.f.) ซึ่ง $f(x) \geq 0$ และ $\sum f(x) = 1$ ทุกค่าใน X ถ้า X เป็นตัวแปรสุ่มต่อเนื่อง จะได้การแจกแจงความน่าจะเป็นต่อเนื่อง จะมี f เป็นฟังก์ชันความหนาแน่นความน่าจะเป็น (p.d.f.) ซึ่ง $f(x) \geq 0$ และ $\int_{-\infty}^{\infty} f(x) = 1$ ถัดไปกล่าวถึงความน่าจะเป็นร่วมกัน คือการพิจารณาตัวแปรสุ่ม 2 ตัวแปรซึ่งจะนิยามคล้ายคลึงกับตัวแปรสุ่มตัวเดียว จากนั้นกล่าวถึงตัวแปรสุ่ม X และ Y ที่อิสระต่อกันเมื่อ $f(x, y) = f_X(x) f_Y(y)$ จากนั้นกลาวถึงการคาดคะเนทางคณิตศาสตร์ของตัวแปรสุ่ม X คือ $E(X) = \sum x f(x)$ หรือ $\int_{-\infty}^{\infty} x f(x)\, dx$ สุดท้ายกล่าวถึงความแปรปรวนของตัวแปรสุ่ม $\sigma_X = E[(X - \mu_X)^2]$ และความแปรปรวนร่วม $\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

แบบฝึกหัดบทที่ 4

1. จงหาสูตรความน่าจะเป็นของตัวแปรสุ่ม X เมื่อ X คือจำนวนแต้มที่ได้จากการทอดลูกเต๋าลูกหนึ่ง 1 ครั้ง
2. ซื้อปากกามา 4 ด้ามโดยหยิบจากกล่องซึ่งมีอยู่ 9 ด้าม ในจำนวนนี้มีปากกาที่บกพร่อง 4 ด้าม ถ้า X คือจำนวนจำนวนปากกาที่บกพร่องที่ซื้อมา จงหาสูตรความน่าจะเป็นพร้อมแสดงเป็นตารางค่าความน่าจะเป็น
3. ให้ X เป็นตัวแปรสุ่มต่อเนื่อง ซึ่งมี p.d.f คือ $f(x) = \begin{cases} \dfrac{200}{9}x & \text{เมื่อ } 0 \leq x \leq 0.3 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$

   3.1 จงหา $F(x)$

   3.2 จงหา $P(0.1 < X \leq 0.2)$
4. ตัวแปรสุ่ม X และ Y มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x, y) = \begin{cases} x^2 + \frac{1}{3}xy & \text{เมื่อ } 0 < x < 1 \text{ และ } 0 < y < 2 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

   4.1 จงหา $P(X > \frac{1}{2})$

   4.2 จงหา $P(Y < X)$
5. ให้ X และ Y เป็นตัวแปรสุ่ม ที่มีการแจกแจงความน่าจะเป็นร่วมกันดังตาราง

		x		
		1	2	3
y	1	0	$\frac{1}{6}$	$\frac{1}{12}$
	2	$\frac{1}{5}$	$\frac{1}{9}$	0
	3	$\frac{2}{15}$	$\frac{1}{4}$	$\frac{1}{18}$

   จงหาการแจกแจงมาร์จินัล และตรวจสอบว่า X และ Y เป็นอิสระต่อกันหรือไม่

6. ให้ X_1 และ X_2 เป็นตัวแปรสุ่มต่อเนื่องซึ่งเป็นอิสระต่อกัน โดยที่แต่ละ p.d.f คือ

$$f(x) = \begin{cases} e^{-x} & \text{เมื่อ } x > 0 \\ 0 & \text{เมื่อ } x \leq 0 \end{cases}$$

จงหา $P(X_1 < 2,\, 1 < X_2 < 3)$

7. ให้ฟังก์ชันความหนาแน่นร่วมกันของตัวแปรสุ่มต่อเนื่อง X และ Y คือ

$$f(x,y) = \begin{cases} \dfrac{1}{2} & \text{เมื่อ } 0 \leq x \leq y \text{ และ } 0 \leq y \leq 2 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

7.1 จงหา $P(X \leq \frac{1}{2} \mid Y = \frac{1}{3})$

7.2 จงตรวจสอบว่าตัวแปรสุ่ม X และ Y เป็นอิสระต่อกันหรือไม่

8. ให้ X และ Y เป็นตัวแปรสุ่มต่อเนื่องที่มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{เมื่อ } x > 0 \text{ และ } y > 0 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

8.1 จงหา $f_X(x)$ และ $f_Y(y)$

8.2 จงหา $f_{X|Y}(x \mid y)$ และ $f_{Y|X}(y \mid x)$

8.3 จงหา $P(X = 1 \mid Y = y)$

8.4 X และ Y เป็นอิสระต่อกันหรือไม่

9. บริษัทผู้ผลิตสินค้าชนิดหนึ่งออกจำหน่าย คาดว่าจะได้กำไร 10,000 บาท ด้วยความน่าจะเป็น 0.4 กำไร 30,000 บาท และขาดทุน 5,000 บาท ด้วยความน่าจะเป็น 0.3 เท่ากัน ผู้จัดการบริษัทคิดว่าถ้ากำไรเฉลี่ยมากกว่า 10,000 บาท จึงจะทำการผลิตผลิตสินค้าออกจำหน่าย เขาควรตัดสินใจอย่างไร

10. ในการแข่งขันครั้งหนึ่ง นักขับรถแข่งจะประกันรถของเขาเป็นจำนวนเงิน 200,000 บาท บริษัทประกันประมาณว่า ความน่าจะเป็นที่จะจ่ายเงินทั้งหมดคือ 0.002 ความน่าจะเป็นที่จะจ่ายเงิน 50% คือ 0.010 และความน่าจะเป็นที่จะจ่ายเงินเพียง 25% คือ 0.100 หากไม่คิดค่าใช้จ่ายปลีกย่อยอื่น ๆ บริษัทประกันควรเก็บเบี้ยประกันจากผู้เอาประกันจำนวนเท่าไรจึงจะได้กำไร 2,000 บาท

11. จงหาค่าคาดคะเนของตัวแปรสุ่ม X ซึ่งมี p.d.f. คือ $f(x) = \begin{cases} \dfrac{4}{\pi(x^2+1)} & \text{เมื่อ } 0 < x < 1 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$

12. ให้ X และ Y เป็นตัวแปรสุ่มซึ่ง $E(X) = 10$, $E(Y) = 15$ และ $E(XY) = 49$ จงหา

12.1 $E(X+Y)$

12.2 $E[X - 5Y]$

12.3 $E[2X + 3Y]$

12.4 $E[(X - E(X))(Y - E(Y))]$

13. ถ้าตัวแปรสุ่ม X มีฟังก์ชันความน่าจะเป็นคือ

$$f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{3-x} \quad \text{เมื่อ} \quad x = 0, 1, 2, 3$$

จงหาความแปรปรวนของ X

14. ในการทอดลูกเต๋า 2 ลูกพร้อมกัน จงหาความแปรปรวนของผลรวมของแต้มที่มากกว่า 9

15. ในการโยนเหรียญ 2 เหรียญ ถ้าขึ้นหน้าชนิดเดียวกันจะได้รับเงิน 10 บาท แต่ถ้าขึ้นหน้าชนิดต่างกันจะเสียเงิน 5 บาท จงหาความแปรปรวนของเงินที่จะได้รับในการเล่มเกมนี้

16. ข้อมูลผลตอบแทนของกองทุน A และ B ในภาวะเศรษฐกิจต่าง ๆ เป็นดังนี้

สภาพเศรษฐกิจ	โอกาส	ผลตอบแทนของกองทุน	ผลตอบแทนของกองทุน
ดีมาก	0.25	20%	5%
ปานกลาง	0.50	10%	10%
ถดถอย	0.25	5%	16%

จงหาผลตอบแทนเฉลี่ยรวมของกองทุน A และ B

บทที่ 5

การแจกแจงความน่าจะเป็นไม่ต่อเนื่อง

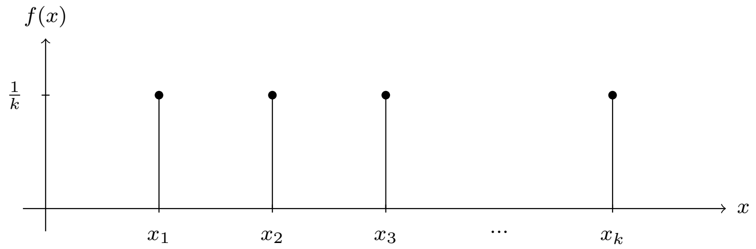
5.1 การแจกแจงยูนิฟอร์ม

บทนิยาม 5.1.1 การแจกแจงของตัวแปรสุ่มซึ่งค่าแต่ละค่าของตัวแปรสุ่มที่จะเกิดเท่า ๆ กันเรียกว่า **การแจกแจงยูนิฟอร์ม** (uniform distribution)

ให้ X เป็นตัวแปรสุ่มที่มีค่าเป็น $x_1, x_2, ..., x_k$ ซึ่งต่างมีความน่าจะเป็นเท่า ๆ กัน ฟังก์ชันความน่าจะเป็นคือ

$$f(x;k) = \frac{1}{k} \qquad \text{เมื่อ} \qquad x = x_1, x_2, ..., x_k$$

หรืออาจแสดงได้ดังกราฟ



ตัวอย่าง 5.1.2 ให้ X เป็นจำนวนหัวที่ได้จากการโยนเหรียญ 1 อัน 1 ครั้ง จงแสดงว่าฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงยูนิฟอร์ม
แนวคำตอบ เนื่องจากโอกาสที่เกิดหัวหรือก้อยเท่า ๆ กันคือ $\frac{1}{2}$ และ

$$f(x;2) = \frac{1}{2} \quad \text{เมื่อ} \quad x = 0, 1$$

ดังนั้น X เป็นการแจกแจงยูนิฟอร์ม

ตัวอย่าง 5.1.3 ในการทอดลูกเต๋า 1 ลูก 1 ครั้ง ให้ X เป็นแต้มที่ได้ จงแสดงว่าฟังก์ชันความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงยูนิฟอร์ม
แนวคำตอบ เนื่องจากโอกาสที่เกิดแต่ละแต้มเท่า ๆ กันคือ $\frac{1}{6}$ และ

$$f(x;6) = \frac{1}{6} \quad \text{เมื่อ} \quad x = 1, 2, 3, 4, 5, 6$$

ดังนั้น X เป็นการแจกแจงยูนิฟอร์ม

ทฤษฎีบท 5.1.4 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของการแจกแจงยูนิฟอร์ม $f(x;k)$ คือ

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i \qquad \text{และ} \qquad \sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2$$

บทพิสูจน์. จะได้ว่า

$$\begin{aligned}\mu &= E(X) = \sum_{i=1}^{k} x_i f(x_i;k) = \sum_{i=1}^{k} x_i \cdot \frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i \\ \sigma^2 &= E[(X-\mu)^2] = \sum_{i=1}^{k}(x_i-\mu) f(x_i;k) = \sum_{i=1}^{k}(x_i-\mu)\cdot\frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)\end{aligned}$$

□

ตัวอย่าง 5.1.5 ในการทอดลูกเต๋า 1 ลูก 1 ครั้ง ให้ X เป็นแต้มที่ได้ จงหาค่าเฉลี่ยและความแปรปรวนของ X

แนวคำตอบ จากตัวอย่าง 5.1.3 จะได้ว่า

$$\begin{aligned}\mu &= \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3.5 \\ \sigma^2 &= \frac{1}{6}\left((1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2\right) \\ &= \frac{35}{12} = 2.92\end{aligned}$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนของแต้มที่ได้จากการทอดลูกเต๋า 1 ลูก 1 ครั้ง เท่ากับ และ 3.5 ตามลำดับ 2.92

ตัวอย่าง 5.1.6 ในการเลือกกรรมการสมาคม 2 คน จากผู้สมัคร 5 คน โดยผู้สมัครแต่ละคนมีโอกาสถูกเลือกเท่า ๆ กัน จงเขียนฟังก์ชันความน่าจะเป็นของกรรมการแต่ละชุด พร้อมทั้งหาค่าเฉลี่ยและความแปรปรวนของการแจกแจงดังกล่าว

แนวคำตอบ จำนวนชุดของกรรมการที่เป็นไปได้ทั้งหมดเท่ากับ ${}^5C_2 = 10$ ชุด
ให้ X เป็นเลขชุดของกรรมการ ซึ่งมีโอกาสถูกเลือกเท่า ๆ กัน ดังนั้น

$$f(x;10) = \frac{1}{10} \quad \text{เมื่อ} \quad x = 1, 2, ..., 10$$

จะได้ว่า

$$\begin{aligned}\mu &= \frac{1}{10}(1+2+3+\cdots+10) = \frac{11}{2} = 5.5 \\ \sigma^2 &= \frac{1}{10}\left((1-5.5)^2+(2-5.5)^2+\cdots+(10-5.5)^2\right) = \frac{33}{4} = 8.25\end{aligned}$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนของเลขชุดกรรมการ เท่ากับ และ 5.5 ตามลำดับ 8.25

5.2 การแจกแจงแบร์นูลลี

การทดลอง 1 ครั้งที่มีผลการทดลองเป็นได้เพียงอย่างใดอย่างหนึ่งใน 2 อย่างนี้คือ

ความสำเร็จ (succees) และ **ความไม่สำเร็จ** (failure)

โดยมีความน่าจะเป็นที่จะเกิดความสำเร็จเท่ากับ p และความน่าจะเป็นที่จะเกิดความไม่สำเร็จเท่ากับ q ซึ่ง $p+q=1$

บทนิยาม 5.2.1 ตัวแปรสุ่มที่แสดงจำนวนครั้งของความสำเร็จที่เกิดขึ้นจากการทดลอง 1 ครั้งเรียกว่า **ตัวแปรสุ่มแบร์นูลลี** (Bernoulli random variable) การแจกแจงความน่าจะเป็นของตัวแปรสุ่มดังกล่าวเรียกว่า **การแจกแจงแบร์นูลลี** (Bernoulli Distribution)

ให้ X เป็นตัวแปรสุ่มแทนจำนวนครั้งที่เกิดความสำเร็จในการทดลอง 1 ครั้ง

ถ้าการทดลองเกิดความสำเร็จแล้ว $X=1$ และถ้าการทดลองเกิดความไม่สำเร็จแล้ว $X=0$

นั่นคือ X มีค่าเป็น 0, 1 แล้วฟังก์ชันความน่าจะเป็นของตัวแปรสุ่มแบร์นูลลี X คือ

$$f(x)=P(X=x)=p^xq^{1-x} \qquad \text{เมื่อ} \qquad x=0,1$$

ตัวอย่าง 5.2.2 ในการโยนเหรียญเที่ยงตรง 1 อัน 1 ครั้ง จงเขียนฟังก์ชันความน่าจะเป็นของการโยนเหรียญได้เหรียญขึ้นหัว

แนวคำตอบ ให้ X เป็นจำนวนครั้งที่ขึ้นหัว นั่นคือโอกาสที่จะขึ้นหัว (สำเร็จ) $p=\frac{1}{2}$ และโอกาสที่ขึ้นหัว (ไม่สำเร็จ) $q=\frac{1}{2}$ ดังนั้น X เป็นตัวแปรสุ่มแบร์นูลลี โดยที่

$$f(x)=\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{1-x} \quad \text{เมื่อ} \quad x=0,1$$

จากตัวอย่าง 5.1.2 จะเห็นว่า X มีการแจกแจงยูนิฟอร์ม นั่นหมายความว่าตัวแปรสุ่มชนิดหนึ่งอาจเป็นได้มากกว่าหนึ่งการแจกแจง

ตัวอย่าง 5.2.3 ในการทอดลูกเต๋า 1 ลูก 1 ครั้ง จงหาแสดงว่าจำนวนครั้งที่ลูกเต๋าขึ้นแต้ม 5 เป็นการแจกแจงแบร์นูลลี

แนวคำตอบ ให้ X เป็นจำนวนครั้งที่ลูกเต๋าขึ้นแต้ม 5 นั่นคือโอกาสที่จะขึ้นแต้ม 5 (สำเร็จ) $p=\frac{1}{6}$ และโอกาสที่ไม่ขึ้นแต้ม 5 (ไม่สำเร็จ) $q=\frac{5}{6}$ ดังนั้น X เป็นตัวแปรสุ่มแบร์นูลลี โดยที่

$$f(x)=\left(\frac{1}{5}\right)^x\left(\frac{5}{6}\right)^{1-x} \quad \text{เมื่อ} \quad x=0,1$$

ทฤษฎีบท 5.2.4 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของการแจกแจงแบร์นูลลี คือ

$$\mu = p \qquad \text{และ} \qquad \sigma^2 = pq$$

บทพิสูจน์. ให้ X เป็นตัวแปรสุ่มแบร์นูลลี โดยมีความน่าจะเป็นที่จะเกิดความสำเร็จเท่ากับ p และความน่าจะเป็นที่จะเกิดความไม่สำเร็จเท่ากับ q ซึ่ง $p+q=1$ จะได้ว่า

$$\mu = E(X) = \sum_{x=0}^{1} xf(x) = \sum_{x=0}^{1} xp^x q^{1-x} = 0\cdot p^0 q^{1-0} + 1\cdot p^1 q^{1-1} = p$$

$$\begin{aligned}\sigma^2 = E(X^2) - \mu^2 &= \sum_{x=0}^{1} xf(x) - p^2 = \sum_{x=0}^{1} xp^x q^{1-x} - p^2 \\ &= 0^2\cdot p^0 q^{1-0} + 1^2\cdot p^1 q^{1-1} - p^2 = p - p^2 = p(1-p) = pq\end{aligned}$$

$\square$

ตัวอย่าง 5.2.5 ในการโยนเหรียญเที่ยงตรง 1 อัน 1 ครั้ง จงหาค่าเฉลี่ยและความแปรปรวนที่จะเหรียญจะขึ้นหัว

แนวคำตอบ จากตัวอย่าง 5.2.2 จำนวนครั้งที่ขึ้นหัวเป็นการแจกแจงแบร์นูลลี โดยที่ $p=\frac{1}{2}$ และ $q=\frac{1}{2}$ ดังนั้น

$$\begin{aligned}\mu &= p = \frac{1}{2} = 0.5 \\ \sigma^2 &= pq = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} = 0.25\end{aligned}$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนที่จะเหรียญจะขึ้นหัวเท่ากับ 0.5 และ 0.25 ตามลำดับ

ตัวอย่าง 5.2.6 การแจกแจงแบร์นูลลีที่มีความแปรปรวนเท่ากับ $\dfrac{2}{9}$ จงหาความน่าจะเป็นที่เกิดความสำเร็จ

แนวคำตอบ จะเห็นว่า $\sigma^2 = \dfrac{2}{9}$ ดังนั้น

$$\begin{aligned}\sigma^2 &= pq = p(1-p) \\ \frac{2}{9} &= p - p^2 \\ 2 &= 9p - 9p^2 \\ 9p^2 - 9p + 2 &= 0 \\ (3p-2)(3p-1) &= 0 \\ p &= \frac{1}{3}, \frac{2}{3}\end{aligned}$$

ดังนั้นความน่าจะเป็นที่เกิดความสำเร็จเท่ากับ $\dfrac{1}{3}$ หรือ $\dfrac{2}{3}$

แบบฝึกหัด 5.1 - 5.2

1. สุ่มหมายเลขโทรศัพท์จากบัญชีหมายเลขโทรศัพท์ ให้ X แทนตัวเลขท้ายของหมายเลขโทรศัพท์ จงหาความน่าจะเป็นที่จะได้ตัวเลขท้าย

   1.1 เป็น 6
   1.2 มากกว่า 3
   1.3 น้อยกว่า 5
   1.4 ระหว่าง 2 กับ 6

2. จงหาค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มข้อ 1

3. ในการทอดลูกเต๋า 1 ลูก 1 ครั้ง จงหาความน่าจะเป็นที่

   3.1 ขึ้นแต้มเป็นจำนวนคู่
   3.2 ขึ้นแต้มมากกว่า 3

4. สุ่มหยิบสินค้ามา 1 ชิ้นจากกล่องที่มีสินค้า 10 ชิ้นซึ่งมีสินค้าชำรุดปนอยู่ 4 ชิ้น จงหาความน่าจะเป็นที่ได้สินค้าชำรุด และหาค่าเฉลี่ยและความแปรปรวนของการทดลองดังกล่าว

5. สุ่มเลือกนักเรียนมา 1 คน จากห้องเรียนหนึ่งที่มี 30 คน มีนักเรียนชาย 10 และนักเรียนหญิง 20 คน จงหาความน่าจะเป็นที่ได้นักเรียนชาย และหาค่าเฉลี่ยและความแปรปรวนของการทดลองดังกล่าว

6. บริษัทแห่งหนึ่งมีพนักงาน 50 คน มีคนที่มีเงินเดือนน้อยกว่า 1,500 บาทต่อเดือนอยู่ 30 คน เมื่อสุ่มเลือกพนักงานมา 1 คน ความน่าจะเป็นที่เขาจะมีเงินเดือนมากกว่าหรือเท่ากับ 1,500 บาทต่อเดือน พร้อมเขียนฟังก์ชันความน่าจะเป็นดังกล่าว

7. จงหาค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มข้อ 6

8. โอกาสที่ผู้ลงทะเบียนจะไม่ผ่านวิชาแคลคูลัส 2 เท่ากับ 5% สุ่มนักศึกษาที่ลงทะเบียนวิชามา 1 คน จงหาค่าเฉลี่ยและความแปรปรวนของจำนวนไม่ผ่านวิชานี้

9. การแจกแจงแบร์นูลลีที่มีความแปรปรวนเท่ากับ 0.1875 จงหาความน่าจะเป็นที่เกิดความไม่สำเร็จ

10. การแจกแจงแบร์นูลลีที่มีค่าเฉลี่ย 0.35 จงหาความน่าจะเป็นที่เกิดความไม่สำเร็จ

11. ให้ X เป็นตัวแปรสุ่มยูนิฟอร์มที่มีค่าเป็น $x_1, x_2, ..., x_{10}$ ถ้าค่าเฉลี่ยเท่ากับ 2 และความแปรปรวนเท่ากับ 16 จงหา $E(X^2)$

12. ให้ X เป็นตัวแปรสุ่มยูนิฟอร์มที่มีค่าเป็น $x_1, x_2, ..., x_k$ ถ้า

$$\text{ค่าเฉลี่ยเท่ากับ 10 และ } \sum_{i=1}^{k}(x_i - 9)^2 = 120k + 1$$

    จงหาความแปรปรวนของ X

5.3 การแจกแจงทวินาม

การทดลองทวินาม (binomial experiment) คือการทดลองแบบแบร์นูลลีซ้ำ ๆ หลายครั้ง มีลักษณะการทดลองดังนั้น

1. ทำการทดลอง n ครั้ง
2. ในการทดลองแต่ละครั้งมีผล 2 อย่างเกิดขึ้นคือ ความสำเร็จ และความไม่สำเร็จ
3. ความน่าจะเป็นที่จะเกิดความสำเร็จเท่ากับ p และความน่าจะเป็นที่จะเกิดความไม่สำเร็จเท่ากับ q ซึ่ง $p+q=1$
4. การทดลองแต่ละครั้งเป็นอิสระต่อกัน

บทนิยาม 5.3.1 ตัวแปรสุ่ม X ที่แสดงจำนวนครั้งของความสำเร็จที่เกิดขึ้นจากการทดลอง n ครั้ง ในการทดลองทวินามเรียกว่า **ตัวแปรสุ่มทวินาม** (binomial random variable) เขียนแทนด้วย Bin(n,p) นั่นคือ X เป็นตัวแปรสุ่มทวินาม หรือ

$$X \sim \text{Bin}(n,p)$$

สำหรับการแจกแจงความน่าจะเป็นของ X เรียกว่า **การแจกแจงทวินาม** (binomial distribution) เขียนแทนด้วย $b(x;n,p)$ เมื่อ p คือความน่าจะเป็นของความสำเร็จที่เกิดจากการทดลองแต่ละครั้ง

ในการทดลองทวินาม ถ้า X คือจำนวนครั้งของความสำเร็จที่เกิดขึ้นจากการทดลอง n ครั้ง จะได้ว่า X มีค่าเป็น $0,1,2,...,n$
สนใจความสำเร็จ x ครั้ง เกิดครั้งที่เท่าไรก็ได้แต่รวมกันแล้วเกิดความสำเร็จ x ครั้ง
ดังนั้นจำนวนครั้งที่ไม่สำเร็จเท่ากับ $n-x$ ครั้ง
ให้ S แทนความสำเร็จ และ F แทนไม่สำเร็จ พิจารณาตัวอย่างการเกิดกรณี

$$\underbrace{SSSSS...S}_{x\text{ ครั้ง}}\underbrace{FFFFF...F}_{n-x\text{ ครั้ง}}$$

ความน่าจะเป็นที่เกิดกรณีนี้เท่ากับ p^xq^{n-x} โดยการเรียงสับเปลี่ยของซ้ำจะได้ครบทุกแบบคือ

$$\frac{n!}{x!(n-x)!}\cdot p^xq^{n-x} = \binom{n}{x}p^xq^{n-x}$$

สรุปได้ว่า

$$b(x;n,p) = \binom{n}{x}p^xq^{n-x} \qquad \text{เมื่อ} \qquad x=0,1,2,...,n$$

ให้ $0 \le r \le n$ จะได้ว่า

$$P(X \le r) = \sum_{x=0}^{r} b(x;n,p) = \sum_{x=0}^{r}\binom{n}{x}p^xq^{n-x}$$

ข้อสังเกต การกระจายทวินามของ $1 = 1^n = (p+q)^n = \displaystyle\sum_{x=0}^{n}\binom{n}{x}p^xq^{n-x}$

ตัวอย่าง 5.3.2 ในการทอดลูกเต๋าคู่หนึ่ง 5 ครั้ง ถ้าให้ความสำเร็จของเหตุการณ์ที่ผลบวกของแต้มบนลูกเต๋าทั้งสองเท่ากับ 5 หรือ 6 จงหา

1. จงแสดงว่าจำนวนครั้งที่สำเร็จเป็นการแจกแจงทวินาม และเขียนกราฟของการแจกแจงทวินาม
2. ความน่าจะเป็นที่จะสำเร็จ 3 ครั้ง
3. ความน่าจะเป็นที่จะสำเร็จอย่างมาก 2 ครั้ง

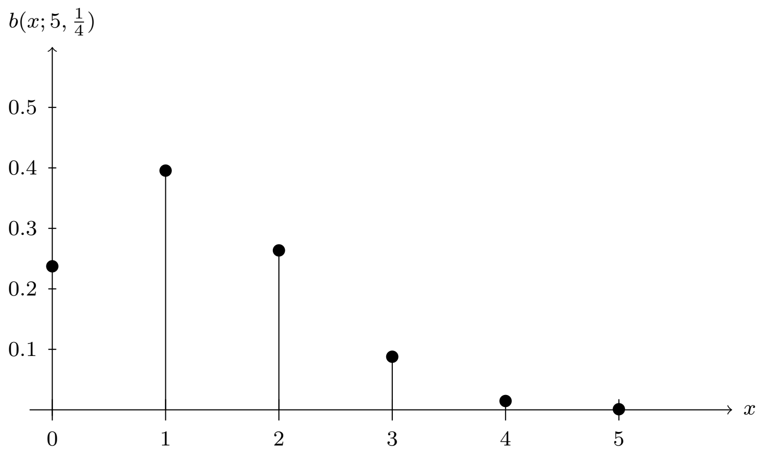
แนวคำตอบ ในการทดลอง 5 ครั้ง คือ และมีผลที่เกิดขึ้นแต่ละครั้ง 2 อย่างคือสำเร็จและไม่สำเร็จ การทอดลูกเต๋าแต่ละครั้งเป็นอิสระต่อกัน ดังนั้นถ้า X เป็นจำนวนครั้งที่สำเร็จ จะได้ว่า $X \sim \text{Bin}(n, p)$ โดยที่ $n = 5$ และความสำเร็จของผลบวกแต้มเป็น 5 หรือ 6 ประกอบด้วยจุดตัวอย่าง

$$(1,4),(2,3),(3,2),(4,1),(1,5),(2,4),(3,3),(4,2),(5,1) \text{ จาก 36 จุดตัวอย่าง}$$

ดังนั้น $p = \frac{9}{36} = \frac{1}{4}$ และ $q = \frac{3}{4}$ แสดงตารางการแจกแจง $\text{Bin}(n = 5, p = \frac{1}{4})$ ได้ดังนี้

X	0	1	2	3	4	5
$b(x; 5, \frac{1}{4})$	${}^5C_0(\frac{3}{4})^5$	${}^5C_1(\frac{1}{4})^1(\frac{3}{4})^5$	${}^5C_2(\frac{1}{4})^2(\frac{3}{4})^3$	${}^5C_3(\frac{1}{4})^3(\frac{3}{4})^2$	${}^5C_4(\frac{1}{4})^4(\frac{3}{4})^1$	${}^5C_5(\frac{1}{4})^5$
	0.2373	0.3955	0.2637	0.0879	0.0146	0.0010

เขียนกราฟได้ดังนี้



จะได้ว่า

$$P(X = 3) = b\left(3; 5, \frac{1}{4}\right) = 0.0879$$

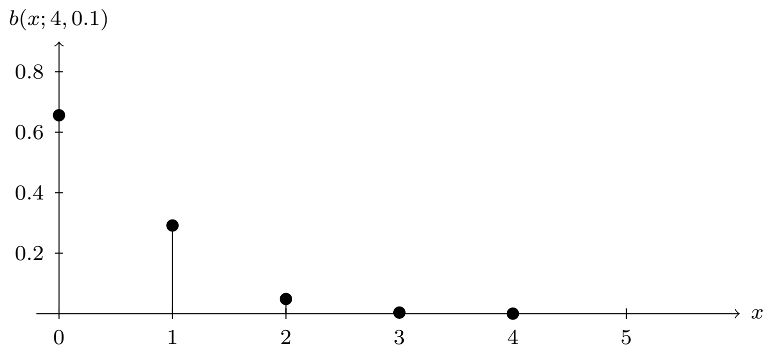
$$P(X \leq 2) = \sum_{x=0}^{2} b(\left(x; 5, \frac{1}{4}\right)) = 0.2373 + 0.3955 + 0.2637 = 0.8965$$

ดังนั้นความน่าจะเป็นที่จะสำเร็จ 3 ครั้งเท่ากับ 0.0878 และความน่าจะเป็นที่จะสำเร็จอย่างมาก 2 ครั้งเท่ากับ 0.8964

ตัวอย่าง 5.3.3 ในการตรวจสอบคุณภาพสินค้าชนิดหนึ่ง โดยการสุ่มสินค้ามา 4 ชิ้นแบบใส่คืนถ้าทราบว่ามีสินค้าเสีย 10% จงเขียนฟังก์ชันความน่าจะเป็นของจำนวนสินค้าตัวอย่างที่เสีย ตารางการแจกแจงความน่าจะเป็น พร้อมแสดงกราฟการแจกแจง
แนวคำตอบ ให้ X เป็นจำนวนสินค้าตัวอย่างที่เสีย จะได้ว่า $X \sim \text{Bin}(n = 4, p = 0.1)$ แสดงตารางการแจกแจงความน่าจะเป็นได้ดังนี้

X	0	1	2	3	4
$b(x; 4, 0.1)$	${}^4C_0(0.9)^5$	${}^4C_1(0.1)^1(0.9)^5$	${}^4C_2(0.1)^2(0.9)^3$	${}^4C_3(0.1)^3(0.9)^2$	${}^4C_4(0.1)^4$
	0.6561	0.2916	0.0486	0.0036	0.0001

เขียนกราฟได้ดังนี้



ตัวอย่าง 5.3.4 โอกาสที่นักศึกษาสาขาวิชาคณิตศาสตร์จะไม่ผ่านวิชาแคลคูลัส ๒ 5% จงหาความน่าจะเป็นที่นักศึกษา 84 คนที่ลงทะเบียนเรียนวิชาแคลคูลัส ๒ จะไม่ผ่านวิชานี้ไม่เกิน 1 คน
แนวคำตอบ ให้ X เป็นจำนวนคนที่ไม่ผ่านวิชาแคลคูลัส ๒ จะได้ว่า $X \sim \text{Bin}(n = 15, p = 0.05)$ จะได้ว่า

$$\begin{aligned} P(X \le 1) &= P(X = 0) + P(X = 1) = b(0; 15, 0.05) + b(1; 15, 0.15) \\ &= \binom{84}{0} 0.05^0(0.95^{84}) + \binom{84}{1} 0.05^1(0.95^{83}) = 0.0135 + 0.0595 = 0.0730 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่นักศึกษา 84 คนที่ลงทะเบียนเรียนวิชาแคลคูลัส ๒ จะไม่ผ่านวิชานี้ไม่เกิน 1 คน เท่ากับ 0.0730

ตัวอย่าง 5.3.5 ในการรักษาโรคชนิดหนึ่งของโรงพยาบาลแห่งหนึ่ง โอกาสที่คนไข้จะหายมี 80% จงหาความน่าจะเป็นที่คนไข้ 10 คน ที่ได้รับการรักษาจากโรงพยาบาลแห่งนี้ จะหายจากโรคนี้อย่างน้อย 9 คน
แนวคำตอบ ให้ X เป็นจำนวนคนไข้ที่หายป่วย จะได้ว่า $X \sim \text{Bin}(n = 10, p = 0.8)$ จะได้ว่า

$$\begin{aligned} P(X \ge 9) &= P(X = 9) + P(X = 10) = b(9; 10, 0.8) + b(10; 10, 0.8) \\ &= \binom{10}{9} 0.8^9(0.2^1) + \binom{10}{10} 0.8^{10}(0.2^0) = 0.2684 + 0.1074 = 0.3758 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่คนไข้ 10 คน ที่ได้รับการรักษาจากโรงพยาบาลแห่งนี้ จะหายจากโรคนี้อย่างน้อย 9 คน เท่ากับ 0.3758

การแจกแจงทวินามมักถูกใช้อย่างกว้างขวาง เราอาจหาค่าได้จากหลากหลายเครื่องมือ เช่น ตารางทวินาม เครื่องคำนวณ และแอพปลิเคชัน

ตารางทวินาม (Binomial Table)

ตารางแสดงค่าความน่าจะเป็นสะสม $\sum_{x=0}^{r} b(x;n,p)$ ถ้า $X \sim \text{Bin}(n,p)$

ตารางที่ 5.1: ตัวอย่างตารางทวินามเมื่อ $n = 5$

		p				
n	r	0.10	0.25	0.50	0.75	0.80
5	0	0.5905	0.2373	0.0312	0.0010	0.0003
	1	0.9185	0.6328	0.1875	0.0156	0.0067
	2	0.9914	0.8965	0.5000	0.1035	0.0579
	3	0.9995	0.9844	0.8125	0.3672	0.2627
	4	1.0000	0.9990	0.9688	0.7627	0.6723
	5	1.0000	1.0000	1.0000	1.0000	1.0000

จากตัวอย่าง 5.3.2 โดยใช้ตารางทวินามได้ดังนี้

$$P(X = 3) = b(3;5,0.25) = \sum_{x=0}^{3} b(x;5,0.25) - \sum_{x=0}^{2} b(x;5,0.25) = 0.9844 - 0.8965 = 0.0879$$

$$P(X \leq 2) = \sum_{x=0}^{2} b(x;5,0.25) = 0.8965$$

ตัวอย่าง 5.3.6 ให้ $X \sim \text{Bin}(n = 5, p = 0.75)$ จงใช้ตารางทวินามเพื่อหาค่าต่อไปนี้

1. $P(X < 4)$
2. $P(X = 4)$
3. $P(X > 2)$
4. $P(1 \leq X < 3)$
5. $P(X \geq 3)$
6. $P(X \neq 2)$

แนวคำตอบ จากตารางทวินามจะได้ว่า

$$P(X < 4) = P(X \leq 3) = \sum_{x=0}^{3} b(x;5,0.75) = 0.3672$$

$$P(X = 4) = \sum_{x=0}^{4} b(x;5,0.75) - \sum_{x=0}^{3} b(x;5,0.75) = 0.7627 - 0.3672 = 0.3955$$

$$P(X > 2) = 1 - P(X \leq 2) = 1 - \sum_{x=0}^{2} b(x;5,0.75) = 1 - 0.1035 = 0.8965$$

$$\begin{aligned} P(1 \leq X < 3) = P(1 \leq X \leq 2) &= \sum_{x=0}^{2} b(x;5,0.75) - \sum_{x=0}^{0} b(x;5,0.75) \\ &= 0.1035 - 0.0010 = 0.1025 \end{aligned}$$

$$P(X \geq 3) = P(X > 2) = 0.8965$$

$$\begin{aligned} P(X \neq 2) = 1 - P(X = 2) &= 1 - \left(\sum_{x=0}^{2} b(x;5,0.75) - \sum_{x=0}^{1} b(x;5,0.75) \right) \\ &= 1 - (0.1035 - 0.0156) = 0.9121 \end{aligned}$$

เครื่องคำนวณ (Calculator)
เครื่องคำนวณมีมากมายหลายแบบที่มีฟังก์ชันการแจกแจงความน่าจะเป็น ในที่นี้ขอยกตัวอย่างการเครื่องคำนวณ CASIO รุ่น fx-991EX โดยทำได้เบื้องต้นดังนี้

MENU → Distribution → Binomial PD / Binomial CD → List / Variable

ซึ่งหาค่าความน่าจะเป็นได้ 4 รูปแบบ คือ

1. หาทีละค่า $b(x;N,p)$ Binomial PD → Variable
2. หา $b(x;N,p)$ หลายค่าพร้อมกัน Binomial PD → List
3. หาทีละค่า $\sum_{x=0}^{r} b(x;N,p)$ Binomial CD → Variable
4. หา $\sum_{x=0}^{r} b(x;N,p)$ หลายค่าพร้อมกัน Binomial CD → List

จากตัวอย่าง 5.3.2 สร้างตารางการแจกแจงทวินามได้โดย

Binomial PD / Binomial CD → List → $x = 0,1,2,3,4,5$ → $N = 5$ และ $p = 0.25$

รูปที่ 5.1: ผลลัพธ์จากเครื่องคำนวณของ Bin$(n = 5, p = 0.25)$

	x	P	Binomal PD
1	0	0.2373	
2	1	0.3955	
3	2	0.2637	
4	3	0.0878	
5	4	0.0146	
6	5	0.0010	

	x	P	Binomal CD
1	0	0.2373	
2	1	0.6328	
3	2	0.8965	
4	3	0.9844	
5	4	0.9990	
6	5	1.0000	

ตัวอย่าง 5.3.7 จงใช้เครื่องคำนวนในการหาค่าต่อไปนี้

1. ถ้า $X \sim$ Bin$(n = 10, p = 0.65)$ จงหาค่าของ $P(X > 5)$

   แนวคำตอบ จะได้ว่า

$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) = 1 - \text{ Binomial CD } (x = 5, N = 10, p = 0.65) \\ &= 1 - 0.2485 = 0.7515 \end{aligned}$$

2. ถ้า $X \sim$ Bin$(n = 15, p = 0.45)$ จงหาค่าของ $P(X < 10)$

   แนวคำตอบ จะได้ว่า

$$P(X < 10) = P(X \leq 9) = \text{ Binomial CD } (x = 9, N = 15, p = 0.45) = 0.9231$$

3. ถ้า $X \sim$ Bin$(n = 20, p = 0.80)$ จงหาค่าของ $P(X \neq 13)$

   แนวคำตอบ จะได้ว่า

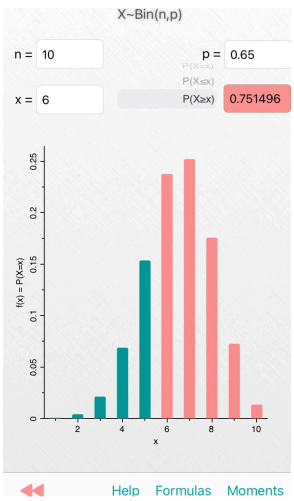
$$\begin{aligned} P(X \neq 13) &= 1 - P(X = 13) = 1 - \text{ Binomial PD } (x = 13, N = 20, p = 0.80) \\ &= 1 - 0.0545 = 0.9455 \end{aligned}$$

แอปพลิเคชัน (Applications)
ปัจจุบันมีแอปพลิเคชันมีบทมากในการศึกษาด้านสถิติมากขึ้น มีการใช้งานอย่างกว้างขวางเนื่องจากการใช้งานที่ง่าย สวยงาม แสดงผลได้รวดเร็ว มีทั้งแบบฟรีและมีค่าใช้จ่าย ในเล่มนี้ผู้เขียนจะขอยกตัวอย่างการแอปพลิเคชันแบบไม่มีค่าใช้จ่ายที่ชื่อ Probability Distribution

รูปที่ 5.2: แอปพลิเคชัน Probability Distribution

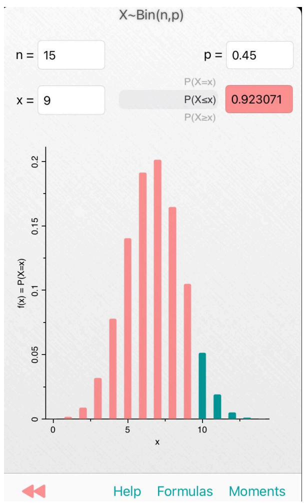
ตัวอย่าง 5.3.8 จงใช้แอปพลิเคชัน Probability Distribution หาค่าต่อไปนี้

1. ถ้า $X \sim \text{Bin}(n = 10, p = 0.65)$
   จงหาค่าของ $P(X > 5)$
   แนวคำตอบ



$$P(X > 5) = P(X \geq 6) = 0.7515$$

2. ถ้า $X \sim \text{Bin}(n = 15, p = 0.45)$
   จงหาค่าของ $P(X < 10)$
   แนวคำตอบ



$$P(X < 10) = P(X \leq 9) = 0.9231$$

ตัวอย่าง 5.3.9 โรงเรียนแห่งหนึ่งกล่าวอ้างว่ามีนักเรียนสอบผ่านเกณฑ์วิชาคณิตศาสตร์พื้นฐาน จำนวน 90% ถ้าสุ่มเลือกนักเรียนมา 20 คน เพื่อสอบถามผลการสอบว่าผ่านเกณฑ์หรือไม่ จงหา

1. ความน่าจะเป็นที่สอบผ่านเกณฑ์จำนวน 15 คน
2. ความน่าจะเป็นที่สอบผ่านเกณฑ์ตั้งแต่ 15 ถึง 18 คน
3. ความน่าจะเป็นที่สอบผ่านเกณฑ์อย่างน้อย 90% ของจำนวนที่เลือกมา

แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่สอบผ่านเกณฑ์ ดังนั้น $X \sim \text{Bin}(n = 20, p = 0.9)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$P(X = 15) = b(15; 20, 0.9) = 0.0319$$

ดังนั้นความน่าจะเป็นที่สอบผ่านเกณฑ์จำนวน 15 คน เท่ากับ 0.0319

$$P(15 \leq X \leq 18) = P(X \leq 18) - P(X \leq 14) = 0.6083 - 0.0113 = 0.5970$$

ดังนั้นความน่าจะเป็นที่สอบผ่านเกณฑ์ตั้งแต่ 15 ถึง 18 คน เท่ากับ 0.5970

$$P(x \geq 0.9(20)) = P(X \geq 18) = 0.6769$$

ดังนั้นความน่าจะเป็นที่สอบผ่านเกณฑ์อย่างน้อย 90% ของจำนวนที่เลือกมา เท่ากับ 0.6769

ตัวอย่าง 5.3.10 ถ้าโอกาสที่นาย ก จะยิงปืนแล้วถูกเป้าที่ตั้งไว้เป็น 70% ถ้านาย ก ยิงปืน 5 ครั้ง จงหา

1. ความน่าจะเป็นที่ถูกเป้า 1 ครั้ง
2. ความน่าจะเป็นที่ถูกเป้าอย่างน้อย 1 ครั้ง

แนวคำตอบ ให้ X เป็นจำนวนครั้งที่นาย ก ยิงเข้าเป้าที่ตั้งไว้ ดังนั้น $X \sim \text{Bin}(n = 5, p = 0.7)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$P(X = 1) = b(1; 5, 0.7) = 0.0283$$

ดังนั้น ความน่าจะเป็นที่ถูกเป้า 1 ครั้ง เท่ากับ 0.0283

$$P(X \geq 1) = 0.9976$$

ดังนั้นความน่าจะเป็นที่ถูกเป้าอย่างน้อย 1 ครั้ง เท่ากับ 0.9976

ทฤษฎีบท 5.3.11 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของตัวแปรสุ่มทวินาม $\mathrm{Bin}(n,p)$ คือ

$$\mu = np \qquad \text{และ} \qquad \sigma^2 = npq$$

บทพิสูจน์. ให้ $X \sim \mathrm{Bin}(n,p)$ แล้ว

$$\begin{aligned}
\mu = E(X) &= \sum_{x=0}^{n} x b(x;n,p) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)!x!} \cdot p^x q^{n-x} = \sum_{x=1}^{n} x \cdot \frac{n(n-1)!}{(n-x)!(x-1)!} \cdot p \cdot p^{x-1} q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{((n-1)-(x-1))!(x-1)!} \cdot p^{x-1} q^{(n-1)-(x-1)} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} \\
&= np \left(\sum_{x=1}^{n-1} \binom{n-1}{x-1} p^{x-1} q^{n-x} + p^{n-1} \right) \\
&= np(p+q)^{n-1} = np(1) = np
\end{aligned}$$

และ

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 b(x;n,p) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^{n} x^2 \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=1}^{n} [(x(x-1)+x)] \cdot \frac{n!}{(n-x)!x!} \cdot p^x q^{n-x} \\
&= \sum_{x=1}^{n} x(x-1) \cdot \frac{n!}{(n-x)!x!} \cdot p^x q^{n-x} + \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)!x!} \cdot p^x q^{n-x} \\
&= \sum_{x=1}^{n} (x-1) \cdot \frac{n!}{(n-x)!(x-1)!} \cdot p^x q^{n-x} + \mu \\
&= \sum_{x=2}^{n} (x-1) \cdot \frac{n!}{(n-x)!(x-1)!} \cdot p^x q^{n-x} + \mu \\
&= \sum_{x=2}^{n} \frac{n!}{(n-x)!(x-2)!} \cdot p^x q^{n-x} + \mu \\
&= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(n-x)!(x-2)!} p^2 \cdot p^{x-2} q^{n-x} + \mu \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!(x-2)!} \cdot p^{x-2} q^{n-x} + \mu \\
&= n(n-1)p^2 (p+q)^{n-2} + \mu = n(n-1)p^2(1) + \mu = n(n-1)p^2 + \mu
\end{aligned}$$

ดังนั้น

$$\sigma^2 = E(X^2) - \mu^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$$

□

ตัวอย่าง 5.3.12 โอกาสที่นักศึกษาของมหาวิทยาลัยแห่งหนึ่งที่จบสาขาวิชาคณิตศาสตร์ คณะครุศาสตร์ จะสอบบรรจุผ่านคิดเป็นร้อยละ 95% จงหาค่าเฉลี่ยและความแปรปรวนของคนที่สอบบรรจุผ่าน เมื่อสุ่มเลือกมาจำนวน 15 คน

แนวคำตอบ ให้ X เป็นจำนวนนักศึกษาที่สอบบรรจุผ่าน ดังนั้น $X \sim \text{Bin}(n = 15, p = 0.95)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$\mu = np = 15(0.95) = 14.25$$
$$\sigma^2 = npq = 15(0.95)(0.05) = 0.7125$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนของคนที่สอบบรรจุผ่าน เมื่อสุ่มเลือกมาจำนวน 15 คน เท่ากับ 14.25 และ 0.7125 ตามลำดับ

ตัวอย่าง 5.3.13 โอกาสที่ผู้ป่วยโควิด-19 ที่ได้รับ ยาฟาวิพิราเวียร์ (Favipiravir) ตั้งแต่เริ่มตรวจพบจะหายร้อยละ 98% ถ้าสุ่มผู้ป่วยโควิดที่เพิ่งตรวจพบมาจำนวน 20 คน จงหาความน่าจะเป็นที่จำนวนผู้ป่วยจะสูงกว่าค่าเฉลี่ยของจำนวนผู้ป่วยที่จะหายจากโรค

แนวคำตอบ ให้ X เป็นจำนวนผู้ป่วยที่จะหายจากโรค ดังนั้น $X \sim \text{Bin}(n = 20, p = 0.98)$ จะได้ว่า

$$\begin{aligned} P(X > \mu) &= P(X > np) = P(X > 20(0.98)) \\ &= P(X > 19.6) = P(X = 20) = 0.6676 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่จำนวนผู้ป่วยจะสูงกว่าค่าเฉลี่ยของจำนวนผู้ป่วยที่จะหายจากโรค เท่ากับ 0.6676

ตัวอย่าง 5.3.14 เนื่องจากมีผู้ติดเชื้อโควิด-19 ลดลงในแต่ละวันในช่วงปลายปี 2564 ทำให้รัฐบาลมีนโยบายเปิดประเทศ ทำให้สถานศึกษาต่าง ๆ เริ่มเปิดเรียนตามปกติ โดยให้นักเรียนมาเรียนในห้องเรียน แต่ต้องทำการตรวจคัดกรองโรคด้วย ATK โรงเรียนแห่งหนึ่งได้ซื้อชุดตรวจ ATK ของบริษัทแห่งหนึ่งซึ่งระบุว่ามีความเม่นยำ 97% ถ้าวันนี้สุ่มตรวจนักเรียนจำนวน 50 คน

1. จงหาค่าเฉลี่ยและความแปรปรวนของจำนวนนักเรียนที่ผลถูกต้อง
2. จงหาความน่าจะเป็นที่ผลไม่ถูกต้องจำนวน 1 คน

แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่ผลตรวจถูกต้อง ดังนั้น $X \sim \text{Bin}(n = 50, p = 0.97)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$\mu = np = 50(0.97) = 48.5$$
$$\sigma^2 = npq = 50(0.97)(0.03) = 1.455$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนของจำนวนนักเรียนที่ผลถูกต้อง เท่ากับ 48.5 และ 1.455 ตามลำดับ ความน่าจะเป็นที่ผลไม่ถูกต้องจำนวน 1 คน หมายถึงความน่าจะเป็นที่ผลถูกต้องจำนวน 49 คน นั่นคือ

$$P(X = 49) = 0.3372$$

ดังนั้นความน่าจะเป็นที่ผลไม่ถูกต้องจำนวน 1 คน เท่ากับ 0.3372

5.4 การแจกแจงเรขาคณิต

การทดลองที่คล้ายกับการทดลองทวินาม แต่สนใจการทดลองซ้ำ ๆ จนพบความสำเร็จครั้งแรก เรียกว่า **การทดลองเรขาคณิต** (geometric experiment) เช่นการโยนเหรียญ 1 อันจนกว่าจะเกิดหัวจึงจะหยุด

บทนิยาม 5.4.1 ตัวแปรสุ่ม X ที่แสดงจำนวนครั้งของการทดลองที่ลงท้ายด้วยความสำเร็จเป็นครั้งแรกในการทดลองเรขาคณิตเรียกว่า **ตัวแปรสุ่มเรขาคณิต** (geometric random variable) เขียนแทนด้วย Geo(p) นั่นคือ X เป็นตัวแปรสุ่มเรขาคณิต หรือ

$$X \sim \text{Geo}(p)$$

สำหรับการแจกแจงความน่าจะเป็นของ X เรียกว่า **การแจกแจงเรขาคณิต** (geometric distribution) เขียนแทนด้วย $g(x;p)$ เมื่อ p คือความน่าจะเป็นของความสำเร็จที่เกิดจากการทดลองแต่ละครั้ง

ถ้าให้ X เป็นจำนวนของการทดลองที่ลงท้ายด้วยความสำเร็จเป็นครั้งแรก ให้ S แทนความสำเร็จซึ่งเกิดขึ้นได้ด้วยความน่าจะเป็น และ F แทนความไม่สำเร็จซึ่งเกิดขึ้นได้ด้วยความน่าจะเป็น ลำดับเหตุการณ์เป็นได้เพียงแบบเดียวคือ

$$\underbrace{FFF...F}_{x-1}S$$

เนื่องจากการทดลองแต่ละครั้งเป็นอิสระต่อกันดังนั้น

$$g(x;p) = pq^{x-1} \qquad \text{เมื่อ} \quad x = 1, 2, 3, ...$$

ตัวอย่าง 5.4.2 นักวิทยาศาสตร์คนหนึ่งต้องการจะเพาะเชื้อในตัวหนูซึ่งมีจำนวนหลายตัว เขาจะฉีดเชื้อเข้าไปตัวละครั้ง แต่ละครั้งที่ฉีดก็จะตรวจดูว่าหนูที่ถูกฉีดเชื้อเข้าไปนั้นติดเชื้อหรือไม่ ถ้าหากไม่พบว่ามันติดเชื้อ เขาก็จะเลิกฉีดตัวต่อไป ถ้าความน่าจะเป็นที่หนูแต่ละตัวจะติดเชื้อชนิดนี้เท่ากับ 0.2 จงหาความน่าจะเป็นที่นักวิทยาศาสตร์คนนี้ใช้หนูเพียง 6 ตัวสำหรับการทดลอง
แนวคำตอบ X เป็นจำนวนหนูที่ถูกฉีดเชื้อเข้าไปในตัวแล้วติดเชื้อเป็นตัวแรก ดังนั้น $X \sim \text{Geo}(p = 0.8)$ จะได้ว่า

$$P(X = 6) = g(6; 0.8) = 0.8(0.2)^5 = 0.0003$$

ดังนั้นความน่าจะเป็นที่นักวิทยาศาสตร์คนนี้ใช้หนูเพียง 6 ตัวสำหรับการทดลองเท่ากับ 0.0655

ตัวอย่าง 5.4.3 ศึกษานิเทศต้องการจะตรวจสอบการอ่านออกเสียง ร ของนักเรียนระดับประถมศึกษาของโรงเรียนแห่งหนึ่ง โดยให้นักเรียนอ่านออกเสียงคำที่กำหนดให้ที่ละคน จะหยุดทดสอบถ้านักเรียนอ่านออกเสียง ร ไม่ถูกต้อง ถ้าความน่าจะเป็นที่นักเรียนแต่ละคนจะออกเสียง ร ถูกต้องเท่ากับ 0.7 จงหาความน่าจะเป็นที่ศึกษานิเทศจะทดสอบนักเรียนเพียง 10 คน
แนวคำตอบ X เป็นจำนวนนักเรียนที่อ่านออกเสียง ร ไม่ถูกต้องเป็นคนแรก ดังนั้น $X \sim \text{Geo}(p = 0.3)$ จะได้ว่า

$$P(X = 10) = g(10; 0.3) = 0.3(0.7)^9 = 0.0121$$

ดังนั้นความน่าจะเป็นที่ศึกษานิเทศจะทดสอบนักเรียนเพียง 10 คนเท่ากับ 0.0121

ทฤษฎีบท 5.4.4 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของตัวแปรสุ่มเรขาคณิต Geo(p) คือ

$$\mu = \frac{1}{p} \qquad \text{และ} \qquad \sigma^2 = \frac{q}{p^2}$$

บทพิสูจน์. ให้ $X \sim$ Geo(p) จะได้ว่า

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} x g(x;p) = \sum_{x=1}^{\infty} xpq^{x-1} \\ E(X^2) &= \sum_{x=1}^{\infty} x^2 g(x;p) = \sum_{x=1}^{\infty} x^2pq^{x-1} \end{aligned}$$

พิจารณาอนุกรมเรขาคณิตอนันต์

$$\begin{aligned} \frac{p}{1-q} &= \sum_{x=1}^{\infty} pq^{x-1} = \sum_{x=1}^{\infty}[x-(x-1)]pq^{x-1} = \sum_{x=1}^{\infty} xpq^{x-1} - \sum_{x=1}^{\infty}(x-1)pq^{x-1} \\ 1 &= E(X) - \sum_{x=2}^{\infty}(x-1)pq^{x-1} = E(X) - \sum_{x=1}^{\infty} xpq^{x} \\ 1 &= E(X) - q\sum_{x=1}^{\infty} xpq^{x-1} = E(X) - qE(X) \\ \mu = E(X) &= \frac{1}{1-q} = \frac{1}{p} \end{aligned}$$

และ

$$\begin{aligned} \frac{1}{p} &= \sum_{x=1}^{\infty} xpq^{x-1} = \sum_{x=1}^{\infty}[(x-1)^2 - x^2 + 3x - 1]pq^{x-1} \\ &= \sum_{x=1}^{\infty}(x-1)^2pq^{x-1} - \sum_{x=1}^{\infty} x^2pq^{x-1} + 3\sum_{x=1}^{\infty} xpq^{x-1} - \sum_{x=1}^{\infty} pq^{x-1} \\ &= \sum_{x=2}^{\infty}(x-1)^2pq^{x-1} - E(X^2) + 3E(X) - 1 \\ &= \sum_{x=1}^{\infty} x^2pq^{x} - E(X^2) + \frac{3}{p} - 1 \\ &= q\sum_{x=1}^{\infty} x^2pq^{x-1} - E(X^2) + \frac{3}{p} - 1 \\ &= (q-1)E(X^2) + \frac{3}{p} - 1 \\ E(X^2) &= \frac{2}{p^2} - \frac{1}{p} \end{aligned}$$

ดังนั้น

$$\sigma^2 = E(X^2) - \mu^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$$

□

แบบฝึกหัด 5.3 - 5.4

1. ในการโยนเหรียญ 1 อัน 10 ครั้ง ให้ X เป็นจำนวนหัวที่เกิดขึ้น

    1.1 X เป็นตัวแปรสุ่มทวินามหรือไม่ จงให้เหตุผลประกอบ

    1.2 จงหาความน่าจะเป็นที่จะขึ้นหัวอย่างน้อยครึ่งหนึ่ง

    1.3 จงหาความน่าจะเป็นที่จะขึ้นหัวไม่เกิน 3 ครั้ง

2. สุ่มหยิบสินค้ามา 3 ชิ้นจากกล่องที่มีสินค้า 10 ชิ้นซึ่งมีสินค้าชำรุดปนอยู่ 4 ชิ้น โดยหยิบที่ละชิ้นมาตรวจสอบแล้วใส่สินค้าล่งในกล่องก่อนหยิบชิ้นต่อไป จงหาความน่าจะเป็นที่ได้สินค้าชำรุด

    2.1 2 ชิ้น

    2.2 ไม่ได้สินค้าชำรุด

    2.3 มากกว่า 3 ชิ้น

    2.4 ชำรุดทุกชิ้น

3. บริษัทผลิตยาสีฟัน A ได้โฆษณาว่าขณะนี้คนไทยใช้ยาสีฟัน A ถึง 30% บริษัทคู่แข่งจึงต้องการตรวจสอบโดยการสุ่มคนไทยมา 20 คน เพื่อสอบถามถึงยี่ห้อยาสีฟันที่ใช้ จงหา

    3.1 ความน่าจะเป็นที่จะมีคนไทยตัวอย่างใช้ยาสีฟัน A จำนวน 10 คน

    3.2 ความน่าจะเป็นที่จะมีคนไทยตัวอย่างใช้ยาสีฟัน A ไม่เกิน 10 คน

    3.3 ความน่าจะเป็นที่จะมีคนไทยตัวอย่างใช้ยาสีฟัน A ตั้งแต่ 8 ถึง 12 คน

4. ถ้าทราบว่าสินค้าที่บรรจุในกล่องใหญ่ใบหนึ่งบรรจุสินค้าเป็นจำนวนมาก ซึ่งมีสินค้าชำรุดอยู่ 5% ถ้าสุ่มหยิบตัวอย่างสินค้าจากหล่องนี้มา 10 ชิ้น แล้วพบว่ามีสินค้าชำรุดตั้งแต่ 2 ชิ้นขึ้นไปจะไม่ยอมรับสินค้านั้นทั้งกล่อง จงหา

    4.1 ความน่าจะเป็นที่ยอมรับสินค้านั้นทั้งกล่อง

    4.2 ความน่าจะเป็นทที่ไม่ยอมรับสินค้านั้นทั้งกล่อง

    4.3 จำนวนสินค้าชุดรุดที่คาดไว้

5. จากการสำรวจครัวเรือนในชุมชนแห่งหนึ่งพบว่า 95% ใช้สมาร์ททีวี เมื่อสุ่มมา 8 ครัวเรือน จงหาความน่าจะเป็นที่มากกว่า 6 ครัวเรือนใช้สมาร์ทีวี

6. ถ้าข้อมูลจากกรมประกันภัยระบุว่า 90% ของรถยนต์ในกรุงเทพฯทำประกันภัยประเภท 3 ไว้ ถ้าสุ่มรถยนต์ในกรุงเทพฯมา 10 คัน จงหาความน่าจะเป็นที่รถยนต์อย่างน้อย 8 คันทำประกันภัยประเภท 3

7. ถ้าทราบว่านักเรียนห้องหนึ่งสอบคณิตศาสตร์ไม่ผ่านเกณฑ์จำนวน 5% ถ้าสุ่มเลือกมา 4 คน จงหาความน่าจะเป็นที่จะสอบผ่านเกณฑ์อย่างน้อย 1 คน

8. จงหาความน่าจะเป็นในการทอดลูกเต๋าเที่ยงตรง 1 ลูกไปเรื่อย ๆ จนกว่าจะขึ้นแต้ม 5 ในครั้งที่ 5 จึงหยุด

9. ถ้าทราบว่าหลอดไฟยี่ห้อ A มีหลอดชำรุดปนอยู่ 10% ถ้าสุ่มหลอดไฟยี่ห้อ A มาตรวจสอบ
   9.1 จงหาความน่าจะเป็นที่จะตรวจพบหลอดชำรุดหลอดแรกในการตรวจสอบหลอดไฟหลอดที่ 4 หรือหลังจากนั้น
   9.2 คาดว่าผู้ตรวจจะต้องตรวจหลอดไฟกี่หลอดจึงจะพบหลอดชำรุดหลอดแรก
10. ถ้าทราบจากลูกค้า 5 คนที่เข้าในร้านขายเฟอร์นิเจอร์แห่งหนึ่งจะมีค้าที่ซื้อ 1 คน จงหาความน่าจะเป็นที่ลูกค้าเข้ามาคนที่ 4 ในวันนี้จะเป็นผู้ซื้อเป็นคนแรก
11. จากสถิติคนที่เกินผ่านร้าน LUX Bergery จำนวน 15 คนจะเข้ามาในร้าน 2 คน ถ้าช่วงเวลาหนึ่งมีคนเดินผ่าน 50 คน จงหาความน่าจะเป็นที่จะมีคนเข้ามาในร้าน 1 คน
12. จงสร้างตารางแจกแจงความน่าจะเป็น และเขียนกราฟของตัวแปรสุ่มทวินาม X ต่อไปนี้
   12.1 $X \sim$ Bin$(n = 5, p = 0.15)$
   12.2 $X \sim$ Bin$(n = 6, p = 0.6)$
13. จงหาค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มต่อไปนี้
   13.1 Bin$(n = 13, p = 0.3)$
   13.2 Bin$(n = 60, p = 0.26)$
   13.3 Bin$(n = 100, p = 0.05)$
   13.4 Geo$(p = 0.15)$
   13.5 Geo$(p = 0.33)$
   13.6 Geo$(p = 0.65)$
14. จงเติมตารางให้สมบูรณ์ เมื่อ $X \sim$ Bin(n, p)

n	p	ค่าเฉลี่ย	ความแปรปรวน	ส่วนเบี่ยงเบนมาตรฐาน
50	0.5			
20		5		
		0.4	100	
	0.2			8

15. ตัวแปรสุ่ม $X \sim$ Geo(p) มีค่าเฉลี่ยเท่ากับ 20 จงหา p และความแปรปรวนของ X
16. ในการโยนเหรียญ 1 อัน n ครั้ง ถ้าความน่าจะเป็นที่จะขึ้นหัวไม่เกิน 30% ของจำนวนครั้งในการโยนเท่ากับ 0.0033 จงหา n

5.5 การแจกแจงทวินามลบ

การทดลองที่คล้ายกับการทดลองทวินาม แต่สนใจการทดลองซ้ำ ๆ จนกระทั่งจำนวนครั้งของความสำเร็จครบตามที่ตั้งใจไว้ ดังนั้นแทนที่ความสำเร็จครั้งแรก ความน่าจะเป็นของการได้ความสำเร็จเป็นครั้งที่ r (จึงหยุด) ในการทดลองครั้งที่ x เรียกว่า **การทดลองทวินามลบ** (negative binomial experiment)

บทนิยาม 5.5.1 ตัวแปรสุ่ม X ที่แสดงจำนวนครั้งของการทดลอง เพื่อให้ได้ความสำเร็จเป็นครั้งที่ r ในการทดลองทิวนามลบ เรียกว่า **ตัวแปรสุ่มทวินามลบ** (negative binomial random variable) เขียนแทนด้วย NB(r, p) นั่นคือ X เป็นตัวแปรสุ่มทวินามลบ หรือ

$$X \sim \text{NB}(r, p)$$

สำหรับการแจกแจงความน่าจะเป็นของ X เรียกว่า **การแจกแจงทิวนามลบ** (negative binomial distribution) หรือ **การแจกแจงปาสคาล** (Pascal distribution) เขียนฟังก์ชันความน่าจะเป็นด้วย $b^*(x; r, p)$ เมื่อ p คือความน่าจะเป็นของความสำเร็จที่เกิดจากการทดลองแต่ละครั้ง

ในการใช้ยาชนิดหนึ่งซึ่งทราบได้ว่าผล 60% จากการใช้ยานี้ถือว่าเกิดความสำเร็จถ้ายาทำให้คนไข้บรรเทาจากอาการป่วย เราต้องการหาความน่าจะเป็นในการที่คนไข้บรรเทาจากอาการป่วยคนที่ 5 เป็นคนไข้คนที่ 7 ที่ได้รับยานี้
ให้ X เป็นจำนวนผู้ป่วยที่ได้รับยา แล้วบรรเทาอาการป่วยเป็นคนที่ 5 โดย $p = 0.6$
พิจารณากรณีหนึ่งที่เกิดขึ้น ของคนไข้บรรเทาจากอาการป่วยคนที่ 5 เป็นคนไข้คนที่ 7 ที่ได้รับยา

$$\begin{array}{ccccccc} 1 & 2 & 3 & & 4 & & 5 \\ S & S & S & F & S & F & S \\ & & & & & & \uparrow \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array}$$

ความน่าจะเป็นที่เกิดกรณีนี้เท่ากับ p^5q^2 โดยการเรียงสับเปลี่ยของซ้ำจะได้ครบทุกแบบคือ

$$\frac{6!}{4!2!} \cdot p^5q^2 = \binom{6}{4} p^5q^2 = \binom{7-1}{5-1} 0.6^5(0.4)^{7-5} = 0.1866$$

ขยายแนวคิดไปยังกรณีทั่วไปได้ดังนี้ ถ้า X เป็นตัวแปรสุมที่แสดงจำนวนครั้งของการทดลอง เพื่อให้ได้ความสำเร็จเป็นครั้งที่ r มีฟังก์ชันการแจกแจงความน่าจะเป็นคือ

$$b^*(x; r, p) = \binom{x-1}{r-1} p^r q^{x-r} \qquad \text{เมื่อ} \quad x = r, r+1, r+2, ...$$

ตัวอย่าง 5.5.2 โยนเหรียญ 2 อัน จงเขียนฟังก์ชันความน่าจะเป็นที่เหรียญขึ้นหัวทั้งคู่เป็นครั้งที่ 3 ในการโยน 3,4,5,... ครั้ง และหาความน่าจะเป็นที่เหรียญจะขึ้นหัวทั้งคู่เป็นครั้งที่ 3 ในการโยน 12 ครั้ง

แนวคำตอบ ให้ X เป็นจำนวนครั้งในการโยนเหรียญ 2 อัน แล้วขึ้นหัวทั้งคู่เป็นครั้งที่ 3 ซึ่งมี $p = \frac{1}{4}$ ดังนั้น $X \sim$ NB$(r = 3, p = 0.25)$ และ

$$b^*(x; 3, 0.25) = \binom{x-1}{2} 0.25^3(0.75)^{x-3} \qquad \text{เมื่อ} \quad x = 3, 4, 5, ...$$

จะเห็นว่า

$$P(X = 12) = b^*(12; 3, 0.25) = \binom{11}{2} 0.25^3(0.75)^9 = 0.0645$$

ดังนั้นความน่าจะเป็นที่เหรียญจะขึ้นหัวทั้งคู่เป็นครั้งที่ 3 ในการโยน 12 ครั้งเท่ากับ 0.0704

ตัวอย่าง 5.5.3 ถ้าทราบว่าสินค้าชนิดหนึ่งมีสินค้าชำรุดปนอยู่ 5% จึงสุ่มสินค้ามาตรวจทีละชิ้น

1. จงหาความน่าจะเป็นที่จะพบสินค้าชำรุดเป็นชิ้นที่ 4 เมื่อตรวจสินค้าชิ้นที่ 10

   แนวคำตอบ ให้ X เป็นจำนวนสินค้าที่ตรวจ แล้วพบสินค้าชำรุดชิ้นที่ 4 แล้ว $X \sim$ NB$(r = 4, p = 0.05)$ จะได้ว่า

$$P(X = 10) = b^*(10; 4, 0.05) = \binom{9}{3} 0.05^4(0.95)^6 = 0.0004$$

   ดังนั้นความน่าจะเป็นที่จะพบสินค้าชำรุดเป็นชิ้นที่ 4 เมื่อตรวจสินค้าชิ้นที่ 10 เท่ากับ 0.0004

2. จงหาความน่าจะเป็นที่จะพบสินค้าชำรุดเป็นชิ้นที่ 5 เมื่อตรวจสินค้าไม่เกิน 15 ชิ้น

   แนวคำตอบ ให้ X เป็นจำนวนสินค้าที่ตรวจ แล้วพบสินค้าชำรุดชิ้นที่ 5 แล้ว $X \sim$ NB$(r = 5, p = 0.05)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$P(X \leq 15) = \sum_{x=5}^{15} b^*(x; 5, 0.05) = 0.0006$$

   ดังนั้นความน่าจะเป็นที่จะพบสินค้าชำรุดเป็นชิ้นที่ 5 เมื่อตรวจสินค้าไม่เกิน 15 เท่ากับ 0.0006

3. จงหาความน่าจะเป็นที่จะพบสินค้าชำรุดเป็นชิ้นที่ 6 เมื่อตรวจสินค้ามากกว่า 20 ชิ้น

   แนวคำตอบ ให้ X เป็นจำนวนสินค้าที่ตรวจ แล้วพบสินค้าชำรุดชิ้นที่ 6 แล้ว $X \sim$ NB$(r = 6, p = 0.05)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$P(X > 20) = P(X \geq 21) = \sum_{x=21}^{\infty} b^*(x; 6, 0.05) = 0.9997$$

   ดังนั้นความน่าจะเป็นที่จะพบสินค้าชำรุดเป็นชิ้นที่ 6 เมื่อตรวจสินค้าชิ้นมากกว่า 20 เท่ากับ 0.9997

ทฤษฎีบท 5.5.4 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของการตัวแปรสุ่มทวินามลบ NB(r,p) คือ

$$\mu = \frac{r}{p} \qquad \text{และ} \qquad \sigma^2 = \frac{rq}{p^2}$$

บทพิสูจน์. เป็นแบบฝึกหัด (ใช้แนวคิดเช่นเดียวกับการแจกแจงทวินาม) □

ตัวอย่าง 5.5.5 ถ้าทราบว่ามีนักเรียนชั้น ม.3 โรงเรียนแห่งหนึ่งสอบวิชาคณิตศาสตร์ไม่ผ่านเกณฑ์ 4% จึงสุ่มถามนักเรียนทีละคน

1. จงหาความน่าจะเป็นที่นักเรียนสอบไม่ผ่านเกณฑ์เป็นคนที่ 3 เมื่อสอบถามคนที่ 5

   แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่สอบถาม แล้วพบนักเรียนที่สอบไม่ผ่านเกณฑ์คนที่ 3
   แล้ว $X \sim$ NB$(r = 3, p = 0.04)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

   $$P(X = 15) = b^*(5; 3, 0.04) = 0.0003$$

   ดังนั้นความน่าจะเป็นที่นักเรียนสอบไม่ผ่านเกณฑ์เป็นคนที่ 3 เมื่อสอบถามคนที่ 5 เท่ากับ 0.0003

2. จงหาจำนวนนักเรียนที่ต้องสอบถามโดยเฉลี่ยจนพบนักเรียนที่ไม่ผ่านเกณฑ์เป็นคนที่ 10

   แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่สอบถาม แล้วพบนักเรียนที่สอบไม่ผ่านเกณฑ์คนที่ 10
   แล้ว $X \sim$ NB$(r = 10, p = 0.04)$ จะได้ว่า

   $$\mu = \frac{r}{p} = \frac{10}{0.04} = 250$$

   ดังนั้นต้องตรวจสอบนักเรียนโดยเฉลี่ย 250 คน จึงจะพบนักเรียนที่ไม่ผ่านเกณฑ์เป็นคนที่ 10

3. จงหาความแปรปรวนของจำนวนนักเรียนที่ต้องสอบถามจนพบนักเรียนที่ไม่ผ่านเกณฑ์เป็นคนที่ 20

   แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่สอบถาม แล้วพบนักเรียนที่สอบไม่ผ่านเกณฑ์คนที่ 20
   แล้ว $X \sim$ NB$(r = 20, p = 0.04)$ จะได้ว่า

   $$\sigma^2 = \frac{rq}{p^2} = \frac{20(0.96)}{0.04^2} = 12000$$

   ดังนั้นความแปรปรวนของจำนวนนักเรียนที่ต้องตรวจสอบจนพบนักเรียนที่ไม่ผ่านเกณฑ์เป็นคนที่ 20 เท่ากับ 12000

5.6 การแจกแจงพหุนาม

การทดลองพหุนามจะกลายเป็น **การทดลองพหุนาม** (multinomial experiment) ถ้าการทดลองแต่ละครั้งมีผลจำแนกได้มากกว่า 2 อย่าง ให้ผลการทดลองแต่ละครั้งมีได้ k อย่างคือ $E_1, E_2, ..., E_k$ ด้วยความน่าจะเป็น $p_1, p_2, ..., p_k$ ตามลำดับ **การแจกแจงพหุนาม** (multinomial distribution) จะให้ความน่าจะเป็นของผลการทดลอง n ครั้ง ที่มี

$$E_1 \text{ เกิดขึ้น } x_1 \text{ ครั้ง } E_2 \text{ เกิดขึ้น } x_2 \text{ ครั้ง .... } E_k \text{ เกิดขึ้น } x_k \text{ ครั้ง}$$

โดยที่ $\sum_{i=1}^{k} x_i = n$ และ $\sum_{i=1}^{k} p_i = 1$ การแจกแจงพหุนามเขียนแทนด้วย $f(x_1, x_2, ..., x_k; p_1, p_2, ..., p_k)$

ในการทดลองพหุนาม ความน่าจะเป็นของตัวแปรสุ่ม $X_1, X_2, ..., X_k$ ซึ่งแทนจำนวนครั้งที่ได้ $E_1, E_2, ..., E_k$ ในการทดลอง n ครั้งที่เป็นอิสระต่อกัน ให้ S_i แทนความสำเร็จของเหตุการณ์ E_i พิจารณาตัวอย่างการเกิดกรณี

$$\underbrace{S_1S_1S_1...S_1}_{x_1 \text{ ครั้ง}}\underbrace{S_2S_2S_2...S_2}_{x_2 \text{ ครั้ง}}\cdots\underbrace{S_kS_kS_k...S_k}_{x_k \text{ ครั้ง}}$$

ความน่าจะเป็นที่เกิดกรณีนี้เท่ากับ $p_1^{x_1}p_1^{x_1}\cdots p_k^{x_k}$ โดยการเรียงสับเปลี่ยนของซ้ำจะได้ครบทุกแบบคือ

$$\frac{n!}{x_1!x_2!\cdots x_k!}\cdot p_1^{x_1}p_1^{x_1}\cdots p_k^{x_k}$$

สรุปได้ว่า

$$f(x_1, x_2, ..., x_k; p_1, p_2, ..., p_k) = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$$

ตัวอย่าง 5.6.1 ในการทอดลูกเต๋า 2 ลูก 7 ครั้ง จงหาความน่าจะเป็นของเหตุการณ์ที่ได้ผลบวกของแต้มทั้งคู่เป็น 7 จำนวน 3 ครั้ง ผลบวกของแต้มทั้งคู่เป็น 11 จำนวน 1 ครั้ง ทั้งคู่ขึ้นหน้าเดียวกันจำนวน 2 ครั้ง และมีอยู่ 1 ครั้งที่ออกมานอกเหนือจากที่กล่าวมา

แนวคำตอบ ให้

E_1	เป็นเหตุการณ์ที่ได้ผลบวกแต้มเป็น 7	เกิด 3 ครั้ง	จะได้	$p_1 = \frac{6}{36} = \frac{1}{6}$
E_2	เป็นเหตุการณ์ที่ได้ผลบวกแต้มเป็น 11	เกิด 1 ครั้ง	จะได้	$p_2 = \frac{2}{36} = \frac{1}{18}$
E_3	เป็นเหตุการณ์ที่ขึ้นแต้มเดียวกัน	เกิด 2 ครั้ง	จะได้	$p_3 = \frac{6}{36} = \frac{1}{6}$
E_4	เป็นเหตุการณ์นอกเหนือจากที่กล่าวมา	เกิด 1 ครั้ง	จะได้	$p_4 = \frac{22}{36} = \frac{11}{18}$

ทำให้ได้ว่า

$$\begin{aligned} f\left(3,1,2,1;\frac{1}{6},\frac{1}{18},\frac{1}{6},\frac{11}{18}\right) &= \frac{7!}{3!1!2!1!}\left(\frac{1}{6}\right)^3\left(\frac{1}{18}\right)\left(\frac{1}{6}\right)^2\left(\frac{11}{18}\right) \\ &= \frac{385}{209952} = 0.0018 \end{aligned}$$

ดังนั้นความน่าจะเป็นดังกล่าวเท่ากับ 0.0018

แบบฝึกหัด 5.5 - 5.6

1. โยนเหรียญ 2 อัน จงเขียนฟังก์ชันความน่าจะเป็นที่เหรียญขึ้นหน้าต่างกันเป็นครั้งที่ 5 ในการโยน 5,6,7,... ครั้ง และหาความน่าจะเป็นที่เหรียญจะขึ้นหน้าต่างกันเป็นครั้งที่ 5 ในการโยน 15 ครั้ง
2. จากการตรวจสอบของภาคชีววิทยาพบว่า 1/4 ของแมลงจะมีตาสีขาว อีก 3/4 จะมีตาสีแดง จงหา

   2.1 ความน่าจะเป็นที่จะพบแมลงตาสีขาวเป็นตัวที่ 3 เมื่อตรวจสอบแมลงตัวที่ 12

   2.2 จำนวนแมลงที่ต้องตรวจสอบโดยเฉลี่ยจนพบแมลงที่มีตาสีขาวเป็นตัวที่ 13

   2.3 ส่วนเบี่ยงเบนมาตรฐานที่ต้องตรวจสอบจนพบแมลงที่มีตาสีขาวเป็นตัวที่ 15
3. ในการทอดลูกเต๋า 1 ลูก 7 ครั้ง จงหาคามน่าจะเป็นที่จะขึ้นแต้ม 6 ครั้งที่ 5 ในการทอดลูกเต๋าครั้งที่ 7
4. ฝ่ายการตลาดของบริษัทผู้ผลิต smart TV แห่งหนึ่งคาดว่ามีลูกค้า 70% ที่ซื้อ smart TV จากห้างสรรพสินค้า จึงสุ่มลูกค้ามา 15 คน สอบถามว่าซื้อ smart TV มาจากที่ใด จงหาความน่าจะเป็นที่จะซื้อ smart TV จากร้านค้าทั่ว ๆ ไปเป็นคนที่ 8 จากลูกค้า 15 คน
5. ถ้าโยนลูกเต๋าเที่ยงตรง 1 ลูก 12 ครั้ง จงหาความน่าจะเป็นที่จะได้แต่ละแต้ม 2 ครั้ง
6. กล่องใบหนึ่งหยิบมีลูกบอลสีแดง 5 ลูก สีขาว 5 ลูก และสีเขียว 4 ลูก หยิบลูกบอลจากกล่องใบนี้มา 7 ลูก โดยการหยิบที่ละลูกแบบไม่ใส่คืน จงหาความน่าจะเป็นที่หยิบได้ลูกบอลสีแดง 5 ลูก สีขาว 1 ลูก และสีเขียว 1 ลูก
7. จากการเก็บข้อมูลพบว่าลูกค้าที่เข้ามาในร้านสมาร์ทโฟนจะซื้อสินค้า 2 ใน 10 คน จงหาความน่าจะเป็นที่ลูกค้าจะซื้อสินค้าเป็นคนที่ 5 เมื่อเข้ามาในร้านเป็นลำดับที่ 10
8. ถ้านักศึกษาลงทะเบียนเรียนวิชาแคลคูลัส 1 จำนวน 70 คน ประกอบด้วยชั้นปีที่ 1 จำนวน 63 คน ชั้นปีที่ 2 จำนวน 5 คน ชั้นปีที่ 3 จำนวน 2 คน ถ้าเลือกนักศึกษามา 5 คน จงหาความน่าจะเป็นที่จะเป็นนักศึกษาชั้นปีที่ 1 จำนวน 3 และชั้นปีที่ 2 และ 3 อย่างละคน
9. ในการดึงไพ่จากสำหรับหนึ่ง 5 ครั้งโดยไม่ใส่คืน จงหาความน่าจะเป็นที่จะได้โพดำ 2 ใบ โพแดง 1 ข้าวเหลามตัด 1 และดอกจิก 1
10. ถ้าทราบว่ามีนักศึกษาสอบวัดระดับภาษาอังกฤษไม่ผ่านเกณฑ์มีมากถึง 55% จึงสุ่มถามนักศึกษาที่ละคน

    10.1 จงหาความน่าจะเป็นที่นักศึกษาจะสอบไม่ผ่านเกณฑ์เป็นคนที่ 7 เมื่อสอบถามคนที่ 10

    10.2 จงหาจำนวนนักศึกษาที่ต้องสอบถามโดยเฉลี่ยจนพบนักเรียนที่ไม่ผ่านเกณฑ์เป็นคนที่ 15

    10.3 จงหาความแปรปรวนของจำนวนนักศึกษาที่ต้องสอบถามจนพบคนที่ไม่ผ่านเกณฑ์เป็นคนที่ 8

5.7 การแจกแจงไฮเพอร์จีออเมตริก

การทดลองไฮเพอร์จีออเมตริก (hypergeometric experiment) มีลักษณะดังต่อไปนี้

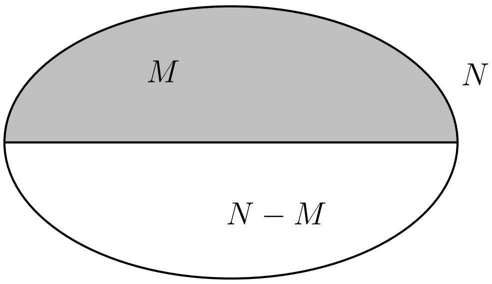
1. สุ่มตัวอย่างขนาด n จากประชากรที่มีขนาด N
2. ใน N รายการ มีอยู่ M รายการที่จำแนกไว้ในพวกที่เรียกว่าความสำเร็จ และ $N-M$ รายการที่เหลือจำแนกอยู่ในพวกความไม่สำเร็จ

บทนิยาม 5.7.1 ตัวแปรสุ่ม X ที่แสดงจำนวนรายการที่เป็นความสำเร็จที่สุ่มได้ในการทดลองไฮเพอร์จีออเมตริก เรียกว่า **ตัวแปรสุ่มไฮเพอร์จีออเมตริก** (hypergeometric random variable) เขียนแทนด้วย HG(n, N, M) นั่นคือ X เป็นตัวแปรสุ่มไฮเพอร์จีออเมตริก หรือ

$$X \sim \text{HG}(n, N, M)$$

สำหรับการแจกแจงความน่าจะเป็น X เรียกว่า **การแจกแจงไฮเพอร์จีออเมตริก** (hypergeometric distribution) เขียนฟังก์ชันความน่าจะเป็นด้วย $h(x; N, n, k)$ คือความน่าจะเป็นขึ้นกับรายการความสำเร็จ x ที่ได้จากการสุ่มตัวอย่างขนาด n จากประชากรขนาด N ซึ่งมี รายการที่เป็นความสำเร็จ M

ให้ X เป็นตัวแปรสุ่มไฮเพอร์จีออเมตริกของจำนวนรายการที่เป็นความสำเร็จที่ได้จากการสุ่มตัวอย่างขนาด n จากประชากรขนาด N ซึ่งมีรายการที่เป็นความสำเร็จ M และ ความไม่สำเร็จ $N-M$ รายการ



การแจกแจงตัวแปรสุ่ม X คือ

$$h(x; n, N, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad \text{เมื่อ} \qquad x = 0, 1, 2, ..., \min\{n, M\}$$

ตัวอย่าง 5.7.2 คณะกรรมการชุดหหนึ่งประกอบด้วยสมาชิก 5 คน ซึ่งเลือกมาอย่างสุ่มจากชาย 3 คน หญิง 6 คน จงหาการแจกแจงความน่าจะเป็นของจำนวนชายที่ได้รับคัดเลือกเข้าเป็นกรรมการ พร้อมเขียนกราฟ

แนวคำตอบ ต้องการคณะกรรมการ 5 คน จากชาย 3 คน และหญิง 6 คน ให้ $n = 5$, $N = 9$ และ $M = 3$

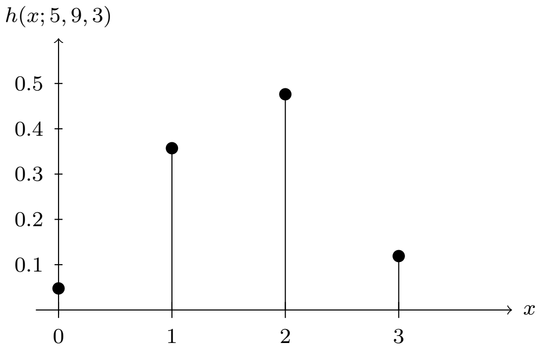
ให้ X เป็นจำนวนนักเรียนชายที่ได้รับคัดเลือก จะได้ว่า $X \sim$ HG$(n = 5, N = 9, M = 3)$ ดังนั้น

$$h(x; 5, 9, 3) = \frac{\binom{3}{x}\binom{6}{5-x}}{\binom{9}{5}} \qquad \text{เมื่อ} \qquad x = 0, 1, 2, 3$$

แสดงตารางแจกแจงความน่าจะเป็นได้ดังนี้ (คำนวณค่าโดยใช้แอปพลิเคชัน)

X	0	1	2	3
	0.0476	0.3571	0.4762	0.1190

เขียนกราฟได้ดังนี้



ตัวอย่าง 5.7.3 จากการสุ่มนักเรียน 3 คนจากนักเรียน 20 คนมาสอบถามคะแนน ONET วิชาคณิตศาสตร์ ถ้าทราบมีนักเรียนไม่ผ่านเกณฑ์ 50% จำนวน 5 คน จงหาความน่าจะเป็นที่พบนักเรียนไม่ผ่านเกณฑ์อย่างน้อย 1 คน

แนวคำตอบ จะได้ว่า $n = 3$, $N = 20$ และ $M = 5$

ให้ X เป็นจำนวนนักเรียนชายที่ไม่ผ่านเกณฑ์ จะได้ว่า $X \sim$ HG$(n = 3, N = 20, M = 5)$ และ

$$P(X \geq 1) = 1 - P(X = 0) = 1 - h(0; 3, 20, 5) = 1 - \frac{\binom{15}{3}}{\binom{20}{3}} = \frac{127}{228} = 0.6009$$

ดังนั้นความน่าจะเป็นที่พบนักเรียนไม่ผ่านเกณฑ์อย่างน้อย 1 คน เท่ากับ 0.6009

ทฤษฎีบท 5.7.4 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของตัวแปรสุ่มไฮเพอร์จีออเมตริก HG(n, N, M) คือ

$$\mu = \frac{nM}{N} \qquad \text{และ} \qquad \sigma^2 = \frac{N-n}{N-1} \cdot \frac{nM}{N}\left(1 - \frac{M}{N}\right)$$

บทพิสูจน์. เป็นแบบฝึกหัด □

ตัวอย่าง 5.7.5 ถ้าเครื่องเล่นเกม Play Station ที่ผลิตในเดือนนี้ 50 เครื่อง มี 40 เครื่องที่มีคุณภาพดี ถ้าซื้อมา 5 เครื่อง

1. จงหาโอกาสที่จะได้เครื่องคุณภาพดีอย่างน้อย 4 เครื่อง
2. จงหาค่าเฉลี่ยและความแปรปรวนของจำนวนสินค้าคุณภาพดีที่ซื้อมา

แนวคำตอบ จะได้ว่า $n = 5, N = 50$ และ $M = 40$
ให้ X เป็นจำนวนนักเรียนชายที่ไม่ผ่านเกณฑ์ จะได้ว่า $X \sim$ HG$(n = 5, N = 50, M = 40)$ คำนวณค่าโดยใช้แอปพลิเคชัน จะได้ว่า

$$P(X \geq 4) = \sum_{x=4}^{5} h(x; 5, 50, 40) = 0.7419$$

ดังนั้นโอกาสที่จะได้เครื่องคุณภาพดีอย่างน้อย 4 เครื่อง เท่ากับ 0.7419

$$\mu = \frac{5(40)}{50} = 4$$
$$\sigma^2 = \frac{50-5}{50-1} \cdot \frac{5(40)}{50}\left(1 - \frac{40}{50}\right) = \frac{36}{49} = 0.7347$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนของจำนวนสินค้าคุณภาพดีที่ซื้อมา เท่ากับ 4 และ 0.7374 ตามลำดับ

ในกรณีที่ $n < M$ และ n มี**ค่าน้อยมาก**เมื่อเทียบกับ N นั่นคือ

$$\frac{N-n}{N-1} \approx 1$$

จะเห็นว่าความน่าจะเป็นที่จะสำเร็จคือ $p = \dfrac{M}{N}$ จะได้ว่า

$$\mu = \frac{nM}{N} = np \qquad \text{และ} \qquad \sigma^2 = \frac{nk}{N}\left(1 - \frac{M}{N}\right) = npq$$

ค่าเฉลี่ยและความแปรปรวนดังกล่าวเท่ากับการปจกแจงทวินาม นั่นคือเราจะแทนการแจกแจงไฮเพอร์จีออเมตริกด้วยการแจกแจงทวินาม หรือ

$$\text{HG}(n, N, M) \sim \text{Bin}\left(n, \frac{M}{N}\right)$$

ตัวอย่าง 5.7.6 เลือกนักศึกษาอย่างสุ่ม 20 คน จากมหาวิทยาลัยแห่งหนึ่ง ซึ่งมีนักศึกษาชาย 4900 คน และนักศึกษาหญิง 2100 คน เพื่อสำรวจความคิดเห็นอย่างหนึ่ง จงหาความน่าจะเป็นที่จะสุ่มได้นักศึกษาชาย 12 คน โดย

1. ใช้การแจกแจงแจกแจงไฮเพอร์จีออเมตริก
2. การประมาณค่าโดยใช้การแจกแจงทวินาม

แนวคำตอบ จะได้ว่า $n = 20, N = 7000$ และ $M = 4900$
ให้ X เป็นจำนวนนักเรียนชายที่สุ่มได้ จะได้ว่า $X \sim$ HG$(n = 20, N = 7000, M = 4900)$ คำนวณค่าโดยใช้แอปพลิเคชัน จะได้ว่า

$$P(X = 12) = h(12; 20, 7000, 4900) = 0.114436$$

ถ้า HG$(n = 20, N = 7000, M = 4900) \sim$ Bin$(20, 0.7)$ จะได้ว่า

$$P(X = 12) = h(12; 20, 7000, 4900) \approx b(12; 20, 0.7) = 0.114397$$

ดังนั้นความน่าจะเป็นที่จะสุ่มได้นักศึกษาชาย 12 คนจากการประมาณถูกต้องทศนิยม 3 ตำแหน่ง เนื่องจากค่า n มีค่าน้อยมากเมื่อเทียบกับ N หรือ

$$\frac{N-n}{N-1} = \frac{7000-20}{7000-1} = 0.9972853 \approx 1$$

ตัวอย่าง 5.7.7 จากตัวอย่าง 5.7.5 จงใช้การประมาณค่าของโอกาสที่จะได้เครื่องคุณภาพดีอย่างน้อย 4 เครื่อง พร้อมเปรียบเทียบทั้ง 2 วิธี
แนวคำตอบ ถ้า HG$(n = 5, N = 50, M = 40) \sim$ Bin$(5, 0.8)$ จะได้ว่า

$$P(X \geq 4) = \sum_{x=4}^{5} h(x; 5, 50, 40) \approx \sum_{x=4}^{5} b(x; 5, 0.8) = 0.7373$$

แต่ค่าจริง 0.7419 ทศนิยมถูกต้องเพียง 1 ตำแหน่งเท่านั้น วิธีการประมาณนี้อาจไม่เหมาะสมหากต้องการความถูกต้องมากกว่านี้ ก่อนตัดสินใจประมาณค่าเราอาจตรวจสอบอัตราส่วน

$$\frac{N-n}{N-1} = \frac{50-5}{50-1} = 0.9183673$$

ว่ามีค่าใกล้ 1 มากเพียงใด อาจทำให้การตัดสินใจประมาณค่าผิดพลาดน้อยลงได้

5.8 การแจกแจงปัวส์ซง

การทดลองที่มีค่าของตัวแปรสุ่มซึ่งแสดงจำนวนครั้งของความสำเร็จในช่วงเวลาที่กำหนดให้ หรือภายในอาณาบริเวณหนึ่งโดยเฉพาะ การดลองนี้เรียกว่า **การทดลองปัวส์ซง** (Poisson experiment) ซึ่งมีลักษณะดังต่อไปนี้

1. จำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลาใดเวลาหนึ่ง หรืออาณาบริเวณใดบริเวณหนึ่ง เป็นอิสระกับจำนวนครั้งของความสำเร็จที่เกิดในช่วงเวลาอื่น หรืออาณาบริเวณอื่น
2. ความน่าจะเป็นของการได้ความสำเร็จหนึ่งครั้งในช่วงเวลาสั้นมากช่วงหนึ่ง หรืออาณาบริเวณที่เล็กมากบริเวณหนึ่ง แปรผันตรงกับช่วงเวลาหรืออาณาบริเวณนั้น และไม่ขึ้นกับจำนวนครั้งความสำเร็จที่เกิดขึ้นนอกเวลาหรือนอกอาณาบริเวณดังกล่าว
3. ความน่าจะเป็นของการได้ความสำเร็จที่เกิดขึ้นมากกว่าหนึ่งครั้งในช่วงเวลาที่สั้นมาก หรือภายในอาณาบริเวณที่เล็กมาก มีค่าน้อยมากจนสามารถตัดทิ้งได้

การทดลองปัวส์ซงอาจกล่าวได้ว่าคือการทดลองทวินามของตัวแปรสุ่ม Bin(n, p) โดยที่ตัวอย่างที่เลือกมามีค่าไม่จำกัด หรือ $n \to \infty$ จะทำให้ $p \to 0$ โดยที่ค่าเฉลี่ยเท่ากับ np เป็นค่าคงตัว

บทนิยาม 5.8.1 ตัวแปรสุ่ม X ที่แสดงจำนวนครั้งของความสำเร็จที่ได้จากการทดลองปัวส์ซง เรียกว่า **ตัวแปรสุ่มปัวส์ซง** (Poisson random variable) เขียนแทนด้วย Pois(λ) หรือ

$$X \sim \text{Pois}(\lambda)$$

สำหรับการแจกแจงความน่าจะเป็น X เรียกว่า **การแจกแจงปัวส์ซง** (Poisson distribution) โดยที่ λ คือค่าเฉลี่ยของจำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลาหนึ่งหรืออาณาบริเวณหนึ่งที่กล่าวถึง

ทฤษฎีบท 5.8.2 ฟังก์ชันการแจกแจงของตัวแปรสุ่ม $X \sim \text{Pois}(\lambda)$ คือฟังก์ชัน p ที่นิยามโดย

$$p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{เมื่อ} \qquad x = 0, 1, 2, ...$$

เมื่อ λ คือค่าเฉลี่ยของจำนวนครั้งของความสำเร็จที่เกิดขึ้นในช่วงเวลาหนึ่งหรืออาณาบริเวณหนึ่งที่กล่าวถึง

บทพิสูจน์. ให้ X เป็นตัวแปรสุ่มปัวส์ซง จะได้ว่า $X \sim \text{Bin}(n, p)$ เมื่อ $n \to \infty$ ให้ $\lambda = np$ นั่นคือ

$$\begin{aligned}
p(x; \lambda) &= \lim_{n\to\infty} b(x; n, p) = \lim_{n\to\infty} \binom{n}{x} p^x q^{n-x} = \lim_{n\to\infty} \frac{n!}{(n-x)!x!} \cdot p^x (1-p)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \cdot \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} \cdot \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty} 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \cdot \left(1 - \frac{x-1}{n}\right) \cdot \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&= \lim_{n\to\infty} 1 \cdot \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \cdot 1
\end{aligned}$$

จาก $\lim_{n\to\infty}\left(1+\dfrac{x}{n}\right)^n = e^x$ ทำให้ได้ว่า

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

ดังนั้น

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{เมื่อ} \qquad x = 0, 1, 2, \ldots$$

$\square$

ตัวอย่าง 5.8.3 จากการสำรวจอุบัติเหตุที่เกิดขึ้นบนถนนสายหนึ่งพบว่า จะมีอุบัติเหตุเกิดขึ้นโดยเฉลี่ย 4 ครั้งในหนึ่งสัปดาห์ จงหาความน่าจะเป็นที่จะเกิดอุบัติเหตุบนถนนสายนี้ 6 ครั้ง ในหนึ่งสัปดาห์
แนวคำตอบ ให้ X จำนวนอุบัติเหตุที่เกิดขึ้นบนถนนสายนี้แต่ละสัปดาห์ โดยมีอุบัติเหตุเกิดขึ้นโดยเฉลี่ย 4 ครั้งต่อสัปดาห์ ดังนั้น $X \sim \text{Pois}(\lambda = 4)$ จะได้ว่า

$$P(X=6) = p(6;4) = \frac{e^{-4}4^6}{6!} = 0.1042$$

ดังนั้นความน่าจะเป็นที่จะเกิดอุบัติเหตุบนถนนสายนี้ 6 ครั้ง ในหนึ่งสัปดาห์ เท่ากับ 0.1042

ตัวอย่าง 5.8.4 ในการเข้าเรียนวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง พบว่าโดยเฉลี่ยมีนักเรียนเข้าเรียนสายจำนวน 3 คนต่อสัปดาห์ จงหาความน่าจะเป็นที่สัปดาห์ถัดไปที่นักเรียนห้องนี้จะเข้าเรียนวิชาคณิตศาสตร์สายจำนวน 5 คน
แนวคำตอบ ให้ X จำนวนนักเรียนที่เข้าเรียนวิชาคณิตศาสตร์สายแต่ละสัปดาห์ โดยเฉลี่ยมีนักเรียนเข้าเรียนสายจำนวน 3 คนต่อสัปดาห์ ดังนั้น $X \sim \text{Pois}(\lambda = 3)$ จะได้ว่า

$$P(X=5) = p(5;3) = \frac{e^{-3}3^5}{5!} = 0.1008$$

ดังนั้นความน่าจะเป็นที่สัปดาห์ถัดไปที่นักเรียนห้องนี้จะเข้าเรียนวิชาคณิตศาสตร์สายจำนวน 5 คน เท่ากับ 0.1008

ตารางปัวส์ซง (Poisson Table)
ตารางแสดงค่าความน่าจะเป็นสะสม $\sum_{x=0}^{r} p(x;\lambda)$ ถ้า $X \sim \text{Pois}(\lambda)$

ตารางที่ 5.2: ตัวอย่างตารางปัวส์ซง

	λ				
r	0.5	1	3	5	10
0	0.6065	0.3679	0.0498	0.0067	0.0000
1	0.9098	0.7358	0.1991	0.0404	0.0005
2	0.9982	0.9197	0.4232	0.1247	0.0028
3	0.9998	0.9810	0.6472	0.2560	0.0103
4	1.0000	0.9963	0.8153	0.4405	0.0293
5	1.0000	0.9994	0.9161	0.6160	0.0671

จากตัวอย่าง 5.8.4 โดยใช้ตารางปัวส์ซงได้ดังนี้

$$P(X=5)=p(5;3)=\sum_{x=0}^{5}p(x;3)-\sum_{x=0}^{4}p(x;3)=0.9161-0.8153=0.1008$$

ตัวอย่าง 5.8.5 ให้ $X \sim \text{Pois}(\lambda=5)$ จงใช้ตารางปัวส์ซงเพื่อหาค่าต่อไปนี้

1. $P(X<4)$
2. $P(X=2)$
3. $P(X>2)$

แนวคำตอบ จากตารางทวินามจะได้ว่า

$$P(X<4)=P(X\leq 3)=\sum_{x=0}^{3}p(x;5)=0.2560$$

$$P(X=2)=\sum_{x=0}^{2}p(x;5)-\sum_{x=0}^{1}p(x;5)=0.1247-0.0404=0.0843$$

$$P(X>2)=1-P(X\leq 2)=1-\sum_{x=0}^{2}p(x;5)=1-0.1247=0.8753$$

ตัวอย่าง 5.8.6 จากการสำรวจพบว่าอัตราเฉลี่ยที่เครื่องบินจะลงสู่สนามบินแห่งหนึ่งเท่ากับ 5 ลำต่อชั่วโมง จงหาความน่าจะเป็นที่ภายในหนึ่งโมงจะมีเครื่องบินลงสู่สนามบินนี้มากกว่า 10 ลำ
แนวคำตอบ ให้ X จำนวนเครื่องบินที่ลงสู่สนามบินแห่งนี้ในหนึ่งชั่วโมง อัตราเฉลี่ยที่เครื่องบินจะลงสู่สนามบินแห่งนี้เท่ากับ 5 ลำต่อชั่วโมง ดังนั้น $X \sim \text{Pois}(\lambda=5)$ จะได้ว่า

$$P(X>10)=1-P(X\leq 10)=1-\sum_{x=0}^{10}p(x;5)=1-0.9863=0.0137$$

ดังนั้นความน่าจะเป็นที่ภายในหนึ่งโมงจะมีเครื่องบินลงสู่สนามบินนี้มากกว่า 10 ลำ เท่ากับ 0.0137

ตัวอย่าง 5.8.7 จากสถิติของโรงพยาบาลประจำจังหวัดแห่งหนึ่งพบว่าโดยเฉลี่ยจะมีคนไข้ 5 คนเข้ามารักษาที่ห้องฉุกเฉินในเวลา 18.00-21.00 น.

1. จงหาความน่าจะเป็นที่จะมีคนไข้มารับการรักษาที่ห้องฉุกเฉินไม่เกิน 2 คนในช่วงเวลา 18.00-21.00 น.
2. จงหาความน่าจะเป็นที่จะมีคนไข้มารับการรักษาที่ห้องฉุกเฉินมากกว่า 6 คนในช่วงเวลา 18.00-21.00 น. ในวันพรุ่งนี้

แนวคำตอบ ให้ X จำนวนคนไข้ที่เข้ามาที่ห้องฉุกเฉินในเวลา 18.00-21.00 น. โดยเฉลี่ยจะมีคนไข้ 5 คนเข้ามารักษาที่ห้องฉุกเฉินในเวลา 18.00-21.00 น. ดังนั้น $X \sim \text{Pois}(\lambda=5)$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$P(X\leq 2)=0.1247$$

ดังนั้นความน่าจะเป็นที่จะมีคนไข้มารับการรักษาที่ห้องฉุกเฉินไม่เกิน 2 คนในช่วงเวลา 18.00-21.00 น. เท่ากับ 0.1247 และ

$$P(X > 6) = P(X \geq 7) = 0.2378$$

ดังนั้นความน่าจะเป็นที่จะมีคนไข้มารับการรักษาที่ห้องฉุกเฉินมากกว่า 6 คนในช่วงเวลา 18.00-21.00 น. ในวันพรุ่งนี้ เท่ากับ 0.2378

ทฤษฎีบท 5.8.8 ค่าเฉลี่ย μ และความแปรปรวน σ^2 ของตัวแปรสุ่มปัวส์ซอง Pois(λ) คือ

$$\mu = \lambda \qquad \text{และ} \qquad \sigma^2 = \lambda$$

บทพิสูจน์. จากทฤษฎีบท 5.8.2 จะได้ว่า $\lambda = np$ เป็นค่าคงตัว ดังนั้น

$$\begin{aligned} \mu &= \lim_{n\to\infty} np = \lambda \\ \sigma^2 &= \lim_{n\to\infty} npq = \lim_{n\to\infty} \lambda(1-p) = \lim_{n\to\infty} \lambda\left(1 - \frac{\lambda}{n}\right) = \lambda \end{aligned}$$

□

ตัวอย่าง 5.8.9 ให้ตัวแปรสุ่ม $X \sim$ Pois(λ) โดยมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 จงหา

1. $P(X = \lambda)$
2. $P(|X - \lambda| < 5)$

แนวคำตอบ จะได้ว่า $\lambda = \sigma^2 = 5^2 = 25$ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution

$$\begin{aligned} P(X = \lambda) &= P(X = 25) = 0.5266 \\ P(|X - \lambda| < 5) &= P(|X - 25| < 5) = P(-5 < X - 25 < 5) \\ &= P(20 < X < 30) = P(X \leq 29) - P(X \leq 20) \\ &= 0.8179 - 0.1855 = 0.6324 \end{aligned}$$

จากทฤษฎีบท 5.8.2 เราอาจ**ประมาณค่าการแจกแจงทวินามด้วยปัวส์ซง**ได้เมื่อ

$$n \text{ มีค่ามาก ๆ และ } \lambda = np$$

ตัวอย่าง 5.8.10 โรงงานผลิตหลอดไฟแห่งหนึ่ง ได้ทำการทดสอบหลอดไฟที่ได้ผลิตขึ้นในวันหนึ่ง พบว่ามีหลอดเสีย 3% สุ่มตรวจหลอดไฟ 100 หลอดที่ผลิตได้ จงหาความน่าจะเป็นที่จะได้หลอดเสีย 5 หลอด จงหาค่าโดยใช้การแจกแจงทวินาม และประมาณค่าโดยใช้การแจกแจงปัวส์ซงพร้อมทั้งเปรียบเทียบค่าที่ได้

แนวคำตอบ ให้ X เป็นจำนวนหลอดไฟที่เสียจากการสุ่มตรวจ จาก 100 หลอด

นั่นคือ $X \sim$ Bin($n = 100, p = 0.03$) คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution จะได้ว่า

$$\begin{aligned} P(X = 5) &= b(5; 100, 0.03) = 0.1013 \\ P(X = 5) &\approx p(5; 100(0.03)) = p(5, 3) = 0.1008 \end{aligned}$$

จะเห็นว่าค่าประมาณที่ได้ถูกต้องทศนิยมตำแหน่งที่สอง

แบบฝึกหัด 5.7 - 5.8

1. จากการสุ่มสินค้า 6 ชิ้นจากกล่องที่มีสินค้า 20 ชิ้นมาตรวจสอบ ถ้าพบว่ามีสินค้าชำรุดตั้งแต่ 2 ชิ้นขึ้นไปจะไม่ยอมรับสินค้าทั้งกล่อง ถ้ากล่องใบนี้มีสินค้าชำรุด 5 ชิ้น จงหาความน่าจะเป็นที่จะยอมรับสินค้าทั้งกล่อง

2. ในการคัดเลือกพนักงานตำแหน่งต่าง ๆ 3 ตำแหน่งจากผู้สมัคร 10 คน ที่มีคุณสมบัติเหมือนกันโดยผู้สมัคร 10 คนนี้เป็นเพศชาย 6 คน และเพศหญิง 4 คน

   2.1 ความน่าจะเป็นที่ผู้สมัครหญิงจะถูกเลือก 1 คน

   2.2 ความน่าจะเป็นที่ผู้สมัครหญิงไม่ถูกเลือก

   2.3 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของจำนวนสมัครผู้ชายที่ถูกเลือก

3. แม่บ้านผู้หนึ่งต้องการจะปลูกต้นไม้หน้าสนามหน้าบ้าน จึงไปเลือกซื้อต้นไม้จากร้านขายต้นไม้แห่งหนึ่งซึ่งมีกุหลาบ 3 ต้น มะม่วง 3 ต้น ฝรั่ง 3 ต้น ถ้าแม่บ้านผู้นี้ต้องการซื้อต้นไม้ 5 ต้น จงหาความน่าจะเป็นที่แม่บ้านจะซื้อกุหลาบ 1 ต้น และมะม่วงและฝรั่งอย่างละ 2 ต้น

4. เลือกนักศึกษาอย่างสุ่ม 30 คน จากมหาวิทยาลัยแห่งหนึ่ง ซึ่งมีนักศึกษาชาย 4000 คน และนักศึกษาหญิง 6000 คน เพื่อสำรวจความคิดเห็นอย่างหนึ่ง จงหาความน่าจะเป็นที่จะสุ่มได้นักศึกษาหญิง 15 คน

5. สุ่มสินค้ามาจากกล่อง 2 ชิ้นที่มีสินค้า 1000 ชิ้น ถ้าทราบว่าในกล่องนี้มีสินค้าชำรุดปนอยู่ 4 ชิ้น จงหาความน่าจะเป็นที่จะได้สินค้าชำรุด 2 ชิ้น

   5.1 สุ่มหยิบแบบไม่ใส่คืน

   5.2 สุ่มหยิบแบบใส่คืน

6. ในการนับแบคทีเรียโดยใช้กล้องจุลทัศน์พบว่า จำนวนเฉลี่ยของแบคทีเรียที่ปรากฏบนหน้ากล้องเท่ากับ 15 ตัว จงหาความน่าจะเป็นที่จะพบแบคทีเรีย 20 ตัวบนหน้ากล้อง

7. โดยเฉลี่ยแล้วเรือบรรทุกสินค้าจะเข้าจอดที่ท่าเรือทุก ๆ 2 วัน จงหาความน่าจะเป็นที่จะมีเรือบรรทุกสินค้าตั้งแต่ 2 ลำขึ้นไปมาจอดที่ท่าเรือนี้ในวันพรุ่งนี้

8. ถ้าทราบว่าโดยเฉลี่ยแล้วจะมีลูกค้าเข้าร้านขายของเก่าแห่งหนึ่งวันละ 8 คน จงหาความน่าจะเป็นที่จะมีลูกค้าเข้ามาร้านนี้ตั้งแต่ 3 คนขึ้นไปในวันพรุ่งนี้

9. จากการสำรวจลูกค้ารอแถวจ่ายเงินใน supper market แห่งหนึ่งหลัง 17.00 น. พบว่าโดยเฉลี่ยจะมีลูกค้ารอคิว 3 คน

   9.1 จงหาโอกาสที่ไม่มีลูกค้ารอคิวในวันพรุ่งนี้หลัง 17.00 น.

   9.2 จงหาโอกาสที่ลูกค้ารอคิวมากกว่า 3 คนในวันพรุ่งนี้หลัง 17.00 น.

สรุป

ในบทนี้กล่าวถึงการแจกแจงความน่าจะเป็นไม่ต่อเนื่องแบบต่าง ๆ ประกอบด้วย 1. การแจกแจงยูนิฟอร์มคือการแจกแจงตัวแปรสุ่มที่จะเกิดขึ้นเท่า ๆ กัน 2. การแจกแจงแบร์นูลลีคือการแจกแจงของจำนวนครั้งที่สำเร็จจากการทดลอง 1 ครั้ง 3. การแจกแจงทวินามคือการทดลองแบร์นูลลีซ้ำ ๆ หลายครั้ง นั่นคือการทดลอง n ครั้งโดยแต่ละครั้งอิสระต่อกัน และสนใจจำนวนความสำเร็จ 4. การแจกแจงเรขาคณิตคือการแจกแจงของจำนวนครั้งของการทดลองที่ลงท้ายด้วยความสำเร็จเป็นครั้งแรก 5. การแจกแจงทวินามลบคือการแจกแจงจำนวนครั้งของการทดลองเพื่อให้ได้ความสำเร็จครั้งที่ r 6.การแจกแจงพหุนามคือการแจกแจงจำนวนครั้งที่ผลการจำแนกได้มากกว่า 2 อย่างขึ้นไป 7. การแจกแจงไฮเพอร์จีออเมตริกคือการแจกแจงที่แสดงถึงรายการความสำเร็จที่สุ่มตัวอย่างขนาด n จากประชากร N โดยมีรายการความสำเร็จ M รายการ และ 8. การแจกแจงปัวส์ซงคือการแจกแจงได้จากการทดลองทวินามโดยเลือกตัวอย่างที่มีค่าไม่จำกัด

แบบฝึกหัดบทที่ 5

1. โรงงานผลิต smart phone ยี่ห้อหนึ่ง พบว่า smart phone ใช้งานได้จริง 99.9% ถ้าทางโรงงานส่ง smart phone ไป 800 เครื่องให้กับบริษัทแม่ จงหาความน่าจะเป็นที่จะไม่พบ smart phone เสียเลยใน 800 เครื่องข้างต้น

2. หยิบไพ่ 1 ใบ จากไพ่สำรับหนึ่งซึ่งมี52 ใบ จงเขียนฟังก์ชันความน่าจะเป็นของหน้าไพ่ที่หยิบได้

3. ถ้าฝ่ายบัญชีของสหกรณ์ออมทรัพย์แห่งหนึ่งเชื่อว่ามีความผิดพลาดในการเขียนใบเสร็จรับเงินของ สหกรณ์เป็น 10% จึงสุ่มตัวอย่างใบเสร็จมา 25 ใบ จงหาความน่าจะเป็นที่จะพบว่ามีใบเสร็จที่ เขียนผิดอย่างน้อย 1 ใบ

4. ถ้าโรงพยาบาลแห่งหนึ่งใช้รถฉุกเฉินโดยเฉลี่ย 2 ครั้งต่อวัน จงหาความน่าจะเป็น

   4.1 โรงพยาบาลใช้รถฉุกเฉิน 2 ครั้งในวันพรุ่งนี้

   4.2 โรงพยาบาลไม่ใช้รถฉุกเฉินเลยในวันพรุ่งนี้

   4.3 โรงพยาบาลใช้รถฉุกเฉินมากกว่า 2 ครั้งในวันพรุ่งนี้

5. จากการสำรวจล่าสุดของจังหวัดหนึ่งพบว่า 90% ของครัวเรือนติดเคเบิลทีวี ถ้าสุ่มตัวอย่างครัวเรือนมา 9 ครัวเรือน จงหาโอกาสที่

   5.1 ทั้ง 9 ครัวเรือนจะติดเคเบิลทีวี

   5.2 อย่างน้อย 7 ครัวเรือนจะติดเคเบิลทีวี

6. ถ้าการไฟฟ้าคาดว่าจะมีผู้จ่ายค่าไฟฟ้าที่ธนาคาร 5% ถ้าสุ่มผู้ใช้ไฟฟ้ามา 1200 ราย จงหาโอกาสที่อย่างน้อย 5 รายจ่ายค่าไฟฟ้าที่ธนาคาร

7. ในการเรียนวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง พบว่าโดยเฉลี่ยมีนักเรียนไม่ส่งการบ้าน 5 คนต่อครั้ง จงหาความน่าจะเป็นที่นักเรียนห้องนี้จะไม่ส่งการบ้านไม่เกิน 3 คนต่อครั้ง

8. การฉีดวัคซีนป้องกันโรค covid-19 ให้คนจำนวนมาก พบว่า $\frac{48}{50}$ ของคนที่ได้รับการฉีดวัคซีนแล้วมีภูมิคุมกันโรค จงหา

   8.1 ความน่าจะเป็นที่คนที่ 4 เป็นคนที่มีภูมิคุมกันโรค covid-19 เป็นคนที่สอง

   8.2 ความน่าจะเป็นที่คนที่ 4 เป็นคนที่มีภูมิคุมกันโรค covid-19 เป็นคนที่แรก

9. สินค้าที่ผลิตจากโรงงานหนึ่งบกพร่องประมาณ 1% ผู้ตรวจคุณภาพจะตรวจสินค้าที่ละชิ้น ถ้าการตรวจสอบเป็นไปอย่างอิสระ จงหาความน่าจะเป็นที่ต้องตรวจสินค้าอย่างน้อยที่สุด 100 ชิ้น จึงจะพบสินค้าบกพร่อง 1 ชิ้น

10. ความน่าจะเป็นที่คนจะตาบอดสีเท่ากับ 0.001 ถ้าจักษุแพทย์ตรวจคน 700 คน จงหาความน่าจะเป็นที่จะมีคนตาบอดสี 2 หรือ 3 คน

11. นักศึกษาที่เรียนวิชาแคลคูลัส ๒ เป็นนักศึกษาปีที่สาม 5 คน ปีที่สอง 6 คน และปีที่หนึ่ง 8 คน ถ้าอาจารย์ผู้สอนเลือกนักศึกษามาเป็น 7 คน จงหาความน่าจะเป็นที่จะได้นักศึกษาปีที่สอง 3 คน และนักศึกษาปีที่หนึ่ง 4 คน นักศึกษาที่เรียนวิชาสถิติวิเคราะห์เป็นนักศึกษาปีที่สี่ 3 คน ปีที่ 3 5 คน ปีที่ 2 6 คน และปีที่ 1 8 คน ถ้าอาจารย์ผู้สอนเลือกนักศึกษามาเป็น 7 คน จงหาความน่าจะเป็นที่จะได้นักศึกษาปีที่สอง 3 คน และนักศึกษาปีที่หนึ่ง 4 คน

12. สมมติว่าโดยเฉลี่ย 1 ใน 1000 คน ที่ยื่นใบประเมินภาษีรายได้บุคคลธรรมดา คำนวณผิดพลาดและได้รับใบประเมินภาษีกลับคืน ถ้าสุ่มใบประเมินภาษีมา 10000 ราย และตรวจสอบ จงหาความน่าจะเป็นที่จะพบว่ามี 6, 7 แหรือ 8 ราย ที่คำนวณผิดพลาด

13. กรรมการนักเรียนชุดหนึ่งมี 3 คน โดยเลือกจากมัธยมศึกษาปีที่ 5 จำนวน 6 คน และมัธยมศึกษาปีที่ 6 จำนวน 5 คน จงเขียนสูตรการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ซึ่งแทนจำนวนมัธยมศึกษาปีที่ 5

14. โยนเหรียญเที่ยงตรง 5 เหรียญพร้อมกัน 20 ครั้ง จงหาความน่าจะเป็นที่เหรียญจะขึ้นหัวมากกว่า 2 เหรียญ เป็นจำนวนอย่างน้อย 8 ครั้ง

15. ถ้าจำนวนเด็กในภาคใต้ที่ถูกจับเรียกค่าไถ่ในรอบ 10 ปีที่ผ่านมาโดยเฉลี่ยเป็น 4 คน จงหาเปอร์เซ็นต์ที่ด็กในภาคใต้จะถูกเรียกค่าไถ่ 6 คน

16. บริษัทแห่งหนึ่งจะมีผู้โทรศัพท์เข้ามาเฉลี่ย 3 ครั้งต่อนาที จงสร้างตารางแจกแจงความ น่าจะเป็น ที่จะมีผู้ใชโทรศัพท์เข้ามาในช่วงระยะเวลา 9.00-9.01 น. โดยแสดงอย่างน้อย 5 ค่า พร้อมเขียนกราฟ

17. บริษัทสั่งผลิตแบตเตอรี่จากโรงงานที่หนึ่ง 100 ชิ้น และสั่งผลิตแบตเตอรี่จากโรงงานที่ 2 200 ชิ้น หากสุ่มแบตเตอรี่มา 4 ชิ้นโดยไม่ใส่กลับคืน จงหาความน่าจะเป็นที่ทั้ง 4 ชิ้นจะมาจากโรงงานที่ 1 และความน่าจะเป็นที่มีอยางน้อย 1 ชิ้นที่มาจากโรงงานที่ 1

บทที่ 6

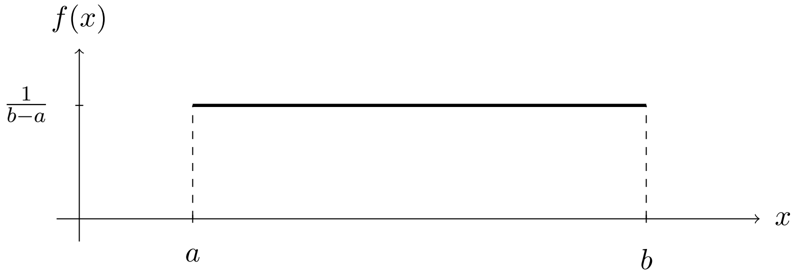
การแจกแจงความน่าจะเป็นต่อเนื่อง

6.1 การแจกแจงยูนิฟอร์ม

ตัวแปรสุ่มต่อเนื่อง X ซึ่งมีค่าได้ทุกค่าจริงในช่วง (a, b) ด้วยความน่าจะเป็นเท่า ๆ กัน เรียกว่า **ตัวแปรสุ่มยูนิฟอร์ม** (uniform random variable) เขียนแทนด้วย Unif(a, b) จะมีฟังก์ชันความน่าจะเป็นดังนี้

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{เมื่อ } a < x < b \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

เขียนกราฟได้ดังนี้



โดยมีค่าเฉลี่ยและความแปรปรวนคือ

$$\mu = \frac{a+b}{2} \qquad \text{และ} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

บทพิสูจน์. จะได้ว่า

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b \frac{x}{b-a} x\,dx = \left[\frac{x^2}{2(b-a)}\right]_a^b$$
$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

และ

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_a^b \frac{x^2}{b-a} x\,dx = \left[\frac{x^3}{3(b-a)}\right]_a^b = \frac{b^3-a^3}{3(b-a)}$$
$$= \frac{(b-a)(b^2+ab+b^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

ดังนั้น

$$\begin{aligned}\sigma_X^2 = E(X^2) - \mu_X^2 &= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{4(a^2+ab+b^2)-3(a+b)^2}{12} \\ &= \frac{a^2+2ab+b^2}{12} = \frac{(b-a)^2}{12}\end{aligned}$$

□

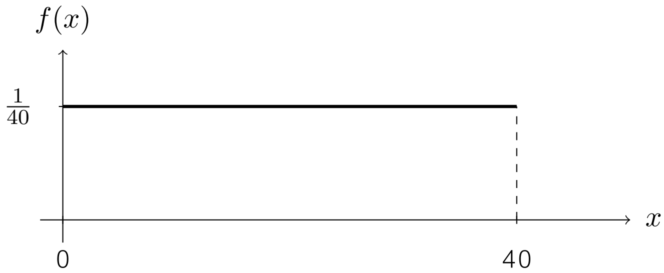
ตัวอย่าง 6.1.1 เมื่อคุณเข้าไปยังอาคารแห่งหนึ่ง และรอลิพอยู่ชั้นล่าง โดยช่วงเวลาในการรอลิพคือ 0 ถึง 40 วินาที สมมติว่าเวลาที่ลิพมาถึงเป็นการแจกแจงยูนิฟอร์มระหว่าง 0 ถึง 40 วินาที หลังจากกดลิพที่ชั้นล่าง

1. จงเขียนฟังก์ชันความน่าจะเป็นพร้อมเขียนกราฟ
2. จงหาความน่าจะเป็นที่รอลิพไม่เกิน 20 วินาที
3. จงหาค่าเฉลี่ยและความแปรปรวนของเวลาในการรอลิพ

แนวคำตอบ ให้ X คือเวลาในการรอลิพในช่วง 0 ถึง 40 วินาที จะได้ว่า $X \sim \text{Unif}(a=0, b=40)$ ดังนั้น

$$f(x) = \begin{cases} \dfrac{1}{40} & \text{เมื่อ } 0 < x < 40 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

เขียนกราฟได้ดังนี้



จะเห็นว่า

$$P(X<20) = \int_{-\infty}^{20} f(x)\,dx = \int_0^{20} \frac{1}{40}\,dx = \left[\frac{1}{40}x\right]_0^{20} = \frac{1}{40}[20-0] = \frac{1}{2} = 0.5$$

ดังนั้นจงหาความน่าจะเป็นที่รอลิพไม่เกิน 20 วินาที เท่ากับ 0.5

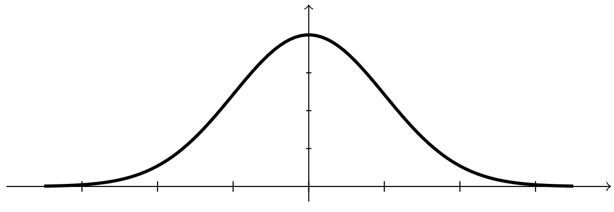
$$\begin{aligned}\mu &= \frac{0+40}{2} = 20 \\ \sigma^2 &= \frac{(40-0)^2}{12} = \frac{400}{3} = 133.33\end{aligned}$$

ดังนั้นค่าเฉลี่ยและความแปรปรวนของเวลาในการรอลิพเท่ากับ 20 และ 133.33 ตามลำดับ

6.2 การแจกแจงปกติ

การแจกแจงปกติ (normal distribution) หรือเรียกอีกอย่างว่า **การแจกแจงแบบเกาส์** (Gaussian distribution) คือการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่ส่วนมากจะมีค่าใกล้เคียงค่าเฉลี่ยของตัวแปรเหล่านั้น จะมีค่าของตัวแปรที่มากกว่าหรือน้อยกว่าค่าเฉลี่ยอย่างมากเป็นส่วนน้อย ตัวแปรสุ่มที่ได้จากการแจกแจงนี้เรียกว่า **ตัวแปรสุ่มปกติ** (normal random variable) เขียนแทนด้วย $\mathsf{N}(\mu, \sigma^2)$ ถ้าให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มปกติคือ μ และ σ ตามลำดับ ฟังก์ชันการแจกแจงของตัวแปรสุ่ม $X \sim \mathsf{N}(\mu, \sigma^2)$ เขียนแทนด้วย $n(x; \mu, \sigma^2)$ และ

$$n(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{เมื่อ} \quad x \in \mathbb{R}$$



สำหรับการแจกแจกปกติจะได้ว่า

$$\text{ค่าเฉลี่ย } E(X) = \mu \qquad \text{และ} \qquad \text{ความแปรปรวน } \mathrm{var}(X) = \sigma^2$$

บทพิสูจน์. ใช้ค่าปริพันธ์ไม่ตรงแบบ

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

ให้ $z = \dfrac{x-\mu}{\sigma}$ จะได้ว่า $x = \sigma z + \mu$ และ $dx = \sigma dz$ ฉะนั้น

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x n(x; \mu, \sigma^2)\, dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \int_{-\infty}^{\infty} (\sigma z + \mu) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, \sigma dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-\frac{1}{2}z^2}\, dz \\
&= \frac{1}{\sqrt{2\pi}} \left(\sigma \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2}\, dz + \mu \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\sigma \cdot 0 + \sqrt{2}\mu \int_{-\infty}^{\infty} e^{-\left(\frac{z}{\sqrt{2}}\right)^2} d\left(\frac{1}{\sqrt{2}} z\right) \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\sqrt{2}\mu \cdot \sqrt{\pi} \right) = \mu
\end{aligned}$$

แสดงได้โดยง่าย $\displaystyle\int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2}\, dz = 0$ (ทำเป็นแบบฝึกหัด)
และความแปรปรวน $\mathrm{var}(X) = \sigma^2$ ทำเป็นแบบฝึกหัด □

ให้ $X \sim \mathsf{N}(\mu, \sigma^2)$ และให้ $Z = \dfrac{X-\mu}{\sigma}$ จะได้ว่า

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}(E(X)-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$\mathsf{var}(Z) = \mathsf{var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathsf{var}(X) = \frac{1}{\sigma^2}\cdot\sigma^2 = 1$$

ดังนั้น

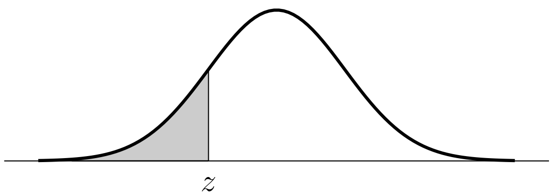
$$E(Z) = 0 \qquad \text{และ} \qquad \mathsf{var}(Z) = 1$$

และพิสูจน์ได้ว่า Z เป็นตัวแปรสุ่มปกติ นั่นคือ $Z \sim \mathsf{N}(0,1)$
เรียกการแจกแจง $n(z;0,1)$ ว่า **การแจกแจงปกติมาตรฐาน** (standard normal distribution)
ดังนั้นตัวแปรสุ่ม X สัมพันธ์กับ Z ตัวอย่างเช่น

$$\begin{aligned} P(x_1 < X < x_2) &= P\left(\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}\right) \\ &= P(z_1 < Z < z_2) \\ &= P(Z < z_2) - P(Z < z_1) \end{aligned}$$

ดังนั้นเราอาจคำนวนค่าความน่าจะเป็นของการแจกแจงปกติ อาศัยการคำนวณจากการแจกแจงปกติมาตรฐาน Z ซึ่งทำได้โดยการเปิดค่าจากตาราง Z ซึ่งบอกค่าพื้นที่ที่หมายถึง $P(Z < z)$ เช่น

ตารางที่ 6.1: ตัวอย่างตาราง z



	0.00	0.01	0.02	0.03	0.04	0.05
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8745
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944

ตัวอย่าง 6.2.1 ให้ X เป็นตัวแปรสุ่มปกติ โดยที่ $\mu = 10$ และ $\sigma = 2$ จงหาความน่าจะเป็นต่อไปนี้

1. $P(X < 12)$
2. $P(12 < X < 12.3)$

แนวคำตอบ

$$P(X < 12) = P\left(\frac{X-\mu}{\sigma} < \frac{12-10}{2}\right) = P(Z < 1) = 0.8413$$

$$\begin{aligned} P(12 < X < 12.3) &= P\left(\frac{12-10}{2} < \frac{X-\mu}{\sigma} < \frac{12.3-12}{2}\right) = P(1 < Z < 1.15) \\ &= P(Z < 1.15) - P(Z < 1) = 0.8745 - 0.8413 = 0.0332 \end{aligned}$$

ตัวอย่าง 6.2.2 ถ้า $X \sim \mathrm{N}(\mu = 50, \sigma^2 = 100)$ จงหาความน่าจะเป็นที่ X มีค่าระหว่าง 45 และ 62
แนวคำตอบ จะได้ว่า $\mu = 50$ และ $\sigma = 10$ หาค่าโดยใช้ตาราง z

$$\begin{aligned} P(45 < X < 62) &= P\left(\frac{45-50}{10} < \frac{X-\mu}{\sigma} < \frac{62-50}{10}\right) = P(-0.5 < Z < 1.2) \\ &= P(Z < 1.5) - P(Z < -0.5) = 0.8849 - 0.3085 = 0.5764 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่ X มีค่าระหว่าง 45 และ 62 เท่ากับ 0.5764

ตัวอย่าง 6.2.3 แบตเตอรี่ชนิดหนึ่งมีอายุการใช้งานเฉลี่ย 3 ปี และส่วนเบี่ยงเบนมาตรฐาน 0.5 ปี สมมติว่าอายุการใช้งานของแบตเตอรี่ชนิดนี้มีการแจกแจงปกติ จงหาความน่าจะเป็นที่แบตเตอรี่จะมีอายุน้อยกว่า 2 ปี
แนวคำตอบ ให้ X เป็นอายุของแบตเตอรี่ จะได้ว่า $X \sim \mathrm{N}(\mu = 3, \sigma^2 = 0.5^2)$ หาค่าโดยใช้ตาราง z

$$P(X < 2) = P\left(\frac{X-\mu}{\sigma} < \frac{2-3}{0.5}\right) = P(Z < -2) = 0.0228$$

ดังนั้นความน่าจะเป็นที่แบตเตอรี่จะมีอายุน้อยกว่า 2 ปี เท่ากับ 0.0228

ตัวอย่าง 6.2.4 อายุของหลอดไฟชนิดหนึ่งมีการแจกแจงปกติที่มีค่าเฉลี่ย 800 ชั่วโมง และส่วนเบี่ยงเบนมาตรฐาน 40 ชั่วโมง จงหาความน่าจะเป็นที่หลอดไฟชนิดนี้จะมีอายุการใช้งานระหว่าง 778 และ 834 ชั่วโมง
แนวคำตอบ ให้ X เป็นอายุการใช้งานของหลอดไฟ จะได้ว่า $X \sim \mathrm{N}(\mu = 800, \sigma^2 = 40^2)$ หาค่าโดยใช้ตาราง z

$$\begin{aligned} P(778 < X < 834) &= P\left(\frac{778-800}{40} < \frac{X-\mu}{\sigma} < \frac{834-800}{40}\right) = P(-0.55 < Z < 0.85) \\ &= P(Z < 0.85) - P(Z < -0.55) = 0.8023 - 0.2912 = 0.5111 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่หลอดไฟชนิดนี้จะมีอายุการใช้งานระหว่าง 778 และ 834 ชั่วโมง เท่ากับ 0.5111

สำหรับตัวอย่างนี้เราสามารถใช้เครื่องคำนวณ CASIO รุ่น fx-991EX ทำได้เบื้องต้นดังนี้

MENU $\rightarrow$ Distribution $\rightarrow$ Normal CD

Normal CD		
Lower	:	778
Upper	:	834
σ	:	40
μ	:	800
P	=	0.5111777713

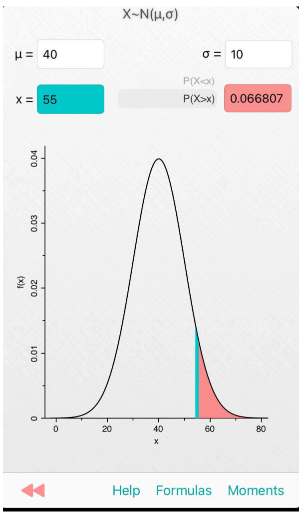
ดังนั้น $P(778 < X < 834) = 0.5111$

ตัวอย่าง 6.2.5 คะแนนเฉลี่ยของวิชาคณิตศาสตร์ ONET ระดับประถมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งเท่ากับ 40 คะแนน และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 คะแนน สมมติว่าคะแนนมีการแจกแจงปกติ ถ้านักเรียนที่เข้าสอบมีจำนวน 500 คน มีนักเรียนประมาณกี่คนที่สอบได้คะแนนมากกว่า 55 คะแนน
แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่เข้าสอบ จะได้ว่า $X \sim \mathrm{N}(\mu = 40, \sigma^2 = 10^2)$
หาค่าโดยใช้ตาราง z

$$P(X > 55) = P\left(\frac{X-\mu}{\sigma} > \frac{55-40}{10}\right) = P(Z > 1.5)$$
$$= 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$$

จะเห็นว่าจำนวนนักเรียนเท่ากับ $500 \times 0.0688 = 33.4$
ดังนั้นมีนักเรียนประมาณกี่คนที่สอบได้คะแนนมากกว่า 55 คะแนน จำนวน 33 คน

สำหรับตัวอย่างนี้เราสามารถใช้แอปพลิเคชัน Probability Distributions



ดังนั้น $P(X > 55) = 0.0668$

ตัวอย่าง 6.2.6 คะแนนสอบของวิชาแคลคูลัสที่มีการแจกแจงปกติ มีคะแนนเฉลี่ยเท่ากับ 60 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 15 คะแนน ถ้าผู้สอนจะให้เกรด A แก่นักศึกษาที่สอบได้คะแนนสูงสุด 5% ของห้อง อยากทราบว่านักศึกษาจะต้องได้คะแนนอย่างน้อยเท่าใดจึงจะได้เกรด A
แนวคำตอบ ให้ X เป็นจำนวนนักเรียนที่เข้าสอบวิชาแคลคูลัส จะได้ว่า $X \sim \mathrm{N}(\mu = 60, \sigma^2 = 15^2)$
ให้ M คือคะแนนต่ำสุดของ 5% สูงสุด นั่นคือ

$$P(X > M) = 0.05$$

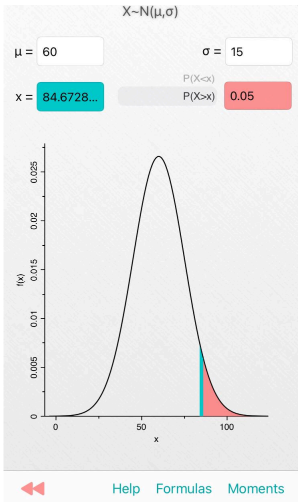
ในตัวอย่างนี้เราจะนำเสนอวิธีหา M ทำได้ 2 แบบคือใช้เครื่องคำนวณ CASIO รุ่น fx-991EX และแอปพลิเคชัน Probability Distributions
ใช้เครื่องคำนวณ CASIO รุ่น fx-991EX ทำได้เบื้องต้นดังนี้

MENU → Distribution → Inverse Normal

โดยค่า Area $= P(X <$ xInv$)$ จะเห็นว่า $P(X < M) = 0.95$

Inverse Normal		
Area	:	0.95
σ	:	15
μ	:	60
xInv	=	84.6728

ดังนั้น $M = 84.67$
ใช้แอปพลิเคชัน Probability Distributions



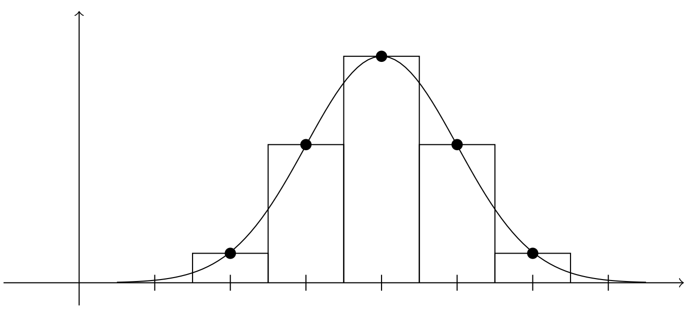
ดังนั้น $M = 84.67$
สรุปได้ว่านักศึกษาจะต้องได้คะแนนอย่างน้อย 84.67 จึงจะได้เกรด A

ทฤษฎีบท 6.2.7 การแจกแจงทวินามประมาณโดยการแจกแจงปกติ
ถ้า X เป็นตัวแปรสุ่มทวินามที่มีค่าเฉลี่ย $\mu = np$ และความแปรปรวน $\sigma^2 = npq$ เมื่อ n มีค่ามากพอ การแจกแจงของ

$$Z = \frac{X - np}{\sqrt{npq}} \qquad \text{จะเป็น} \qquad \mathrm{N}(0,1)$$

บทพิสูจน์. ขอละไว้ในวิชานี้ □

ถ้า $X \sim$ Bin(n,p) ซึ่งเป็นตัวแปรสุ่มไม่ต่อเนื่อง แต่เมื่อ n มีค่ามากพอ จะมีการแจกแจงของตัวแปรสุ่ม N$(0,1)$ ซึ่งเป็นตัวแปรสุ่มต่อเนื่อง แสดงกราฟตัวอย่างเช่น



การนำทฤษฎีบท 6.2.7 ไปใช้ในการประมาณค่าการแจกแจงทวินาม เมื่อ n มีค่ามากพอ

Bin(n, p)	N$(\mu = np, \sigma^2 = npq)$
$P(a < X < b)$	$P(a - 0.5 < X < b + 0.5)$
$P(X > a)$	$P(X > a + 0.5)$
$P(X < a)$	$P(X < a - 0.5)$

ตัวอย่าง 6.2.8 ให้ $X \sim$ Bin(n $= 150, p = 0.2)$ จงหาค่า $P(X < 30)$ และ

1. จงประมาณค่าของ $P(X < 30)$ โดยใช้การแจกแจงปกติ
2. จงประมาณค่าของ $P(X < 30)$ โดยใช้การแจกแจงปัวส์ซง

แนวคำตอบ คำนวณค่าโดยใช้แอปพลิเคชัน Probability Distribution

$$P(X < 30) = P(X \leq 29) = \sum_{x=0}^{29} b(x; 150, 0.2) = 0.467457$$

ประมาณโดยใช้ N$(\mu = np = 30, \sigma^2 = npq = 24)$ จะได้ว่า

$$P(X < 30) \approx P(X < 29.5) = P\left(\frac{X - \mu}{\sigma} < \frac{29.5 - 30}{\sqrt{24}}\right) = P(Z < -0.1021) = 0.459338$$

ประมาณโดยใช้ Pois$(\lambda = np = 30)$ จะได้ว่า

$$P(X < 30) = P(X \leq 29) = \sum_{x=0}^{29} p(x; 30) = 0.475717$$

ตัวอย่าง 6.2.9 ข้อสอบแบบ 4 ตัวเลือกซึ่งมีคำตอบที่ถูกต้องรวมอยู่ 1 ข้อ จำนวน 200 ข้อ ถ้ามีข้อสอบ 80 ข้อ ที่คาดว่านักเรียนไม่มีความรู้ที่จะตอบได้ จงหาโอกาสที่นักเรียนจะตอบถูกโดยการเดาได้ตั้งแต่ 25 ถึง 30 ข้อ
แนวคำตอบ ให้ X เป็นจำนวนข้อสอบที่ตอบถูก นั่นคือ $p = \frac{1}{4}$ จะได้ว่า $X \sim$ Bin(n $= 80, p = 0.25)$

$$P(25 \leq X \leq 30) = P(X \leq 30) - P(X \leq 24) = 0.9954 - 0.8761 = 0.1193$$

เนื่องจาก n มีค่ามาก เราจะประมาณค่าด้วยการแจกแจงปกติ N$(\mu = np = 20, \sigma^2 = npq = 15)$

$$\begin{aligned} P(25 \leq X \leq 30) &\approx P(24.5 \leq X \leq 30.5) \\ &= P\left(\frac{24.5 - 20}{\sqrt{15}} < \frac{X - \mu}{\sigma} < \frac{30.5 - 20}{\sqrt{15}}\right) \\ &= P(1.162 < Z < 2.711) \\ &= 0.119263701 \end{aligned}$$

ดังนั้นโอกาสที่นักเรียนจะตอบถูกโดยการเดาได้ตั้งแต่ 25 ถึง 30 ข้อ เท่ากับ 0.1193

แบบฝึกหัด 6.1 - 6.2

1. ให้ $X \sim \text{Unif}(a = 1, b = 9)$ จงหาความน่าจะเป็นที่ X จะมีค่ามากกว่า 5

2. ให้ X เป็นตัวแปรสุ่มยูนิฟอร์มบนช่วง $(a, 5)$ ถ้าค่าเฉลี่ยของ $\mu = 3$ เท่ากับ

   2.1 จงหาความน่าจะเป็นที่ X จะมีค่าน้อยกว่า 2

   2.2 จงหา $P(1 < X < 3)$

   2.3 จงหาความแปรปรวนของ X

3. ห้างสรรพสินค้าแห่งหนึ่งซึ่งมีความสูง 3 ชั้น เวลาที่ลูกค้าคนหนึ่ง ๆ ต้องรอลิฟอยู่ที่ชั้น 2 มีการแจกแจงแบบยูนิฟอร์ม จาก 0 ถึง 4 นาที

   3.1 จงหาค่าเวลาเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของเวลาที่ลูกค้าต้องรอลิฟอยู่ที่ชั้น 2

   3.2 ถ้าลิฟใช้เวลา 15 นาทีในการเคลื่อนที่จากชั้นหนึ่งไปอีกชั้นหนึ่ง จงหาความน่าจะเป็นที่ลูกค้าคนหนึ่งซึ่งอยู่ที่ชั้น 2 จะถึงชั้น 1 ในเวลา 1.5 นาที

4. จาก $\text{N}(\mu = 40, \sigma^2 = 36)$ จงหา

   4.1 $P(X < 32)$

   4.2 $P(X > 27)$

   4.3 a ซึ่ง $P(X > a) = 0.1539$

   4.4 a ซึ่ง $P(X < a) = 0.2578$

5. คะแนนสอบของนักเรียนจำนวน 1000 คน มีการแจกแจงปกติ มีค่าเฉลี่ยเลขคณิต 150 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 20 คะแนน

   5.1 จำนวนคนที่ได้คะแนนมากกว่า 160

   5.2 จำนวนคนที่ได้คะแนนอยู่ระหว่าง 120 และ 140

   5.3 คนที่ได้คะแนน 180 คะแนน อยู่เปอร์เซนต์ที่เท่าใด

   5.4 มีคนที่ได้คะแนนมากกว่า 170 คะแนน อยู่กี่คน

   5.5 คนที่คะแนนต่ำสุดของ 15% สูงสุด

6. ในการผลิตสินค้าชนิดหนึ่งทราบว่ามีที่ชำรุด 10% ถ้าสุ่มเลือกผลิตภัณฑ์เหล่านี้มา 100 ชิ้น จงหาความน่าจะเป็นที่จะได้สินค้าชำรุดมากกว่า 13 ชิ้น

7. ถ้าคะแนนสอบวิชาสถิติมีการแจกแจงปกติ ค่าเฉลี่ยเลขคณิต 74 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7.9

   7.1 ถ้านักศึกษาสอบได้ F คิดเป็น 10% จากทั้งหมด คะแนนต่ำสุดของนักศึกษาที่ไม่ได้ F เท่ากับเท่าใด

   7.2 คะแนนสูงสุดสำหรับเกรด B ถ้านักศึกษาที่สอบได้คะแนนรวม 5% ได้เกรด A

8. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่งมีการแจกแจงปกติ นาย ก และ นาย ข เป็นนักเรียนในห้องนี้ ถ้ามีนักเรียนในห้องนี้ร้อยละ 9.48 สอบได้คะแนนมากกว่าคะแนนสอบของนาย ก มีนักเรียนร้อยละ 10.64 สอบได้คะแนนน้อยกว่าคะแนนสอบของนาย ข และนาย ข สอบได้คะแนนน้อยกว่าคะแนนสอบของนาย ก อยู่ 51 คะแนน แล้วส่วนเบี่ยงเบนมาตรฐานของคะแนนครั้งนี้เท่ากับเท่าใด

9. น้ำหนักและส่วนสูงของนักเรียนห้องหนึ่งต่างมีการแจกแจงปกติ โดยมีน้ำหนักเฉลี่ย 40 กิโลกรัม ส่วนเบี่ยงเบนมาตรฐาน 2 กิโลกรัม ส่วนสูงเฉลี่ย 150 เซนติเมตร ส่วนเบี่ยงเบนมาตรฐาน 4 เซนติเมตร ถ้ามีนักเรียน a เปอร์เซ็นต์ที่สูงไม่ต่ำกว่า 125 เซนติเมตร และไม่เกิน 155 เซนติเมตร คือกลุ่มนักเรียนที่มีน้ำหนักไม่ต่ำกว่า 36 กิโลกรัมและไม่เกิน b กิโลกรัม จงหาค่าของ a และ b

10. ในการสอบวิชาแคลคูลัสครั้งหนึ่งให้คะแนนไว้เป็นจำนวนเต็ม คะแนนเหล่านี้มีการแจกแจงปกติโดยประมาณ ค่าเฉลี่ยเลขคณิต 82 ส่วนเบี่ยงเบนมาตรฐาน 5 นักศึกษา 8 คน สอบได้คะแนนตั้งแต่ 88 ถึง 94 ได้เกรด B จงหาจำนวนนักศึกษาที่เข้าสอบวิชานี้

11. ผลผลิตของข้าวโพดที่ดินทั้งหมด 1000 ไร่ มีการแจกแจงปกติ ค่าเฉลี่ยเลขคณิต 20.50 ถังต่อไร่ จากการสำรวจพบว่ามีอยู่ 200 ไร่ ที่ให้ผลผลิตไม่มากกว่า 17.75 ถังต่อไร่ จงหาส่วนเบี่ยงเบนมาตรฐานของข้าวโพดเหล่านี้

12. ไอคิวของผู้สมัครเข้าเรียนของมหาวิทยาลัยแห่งหนึ่งมีการแจกแจงปกติ มีค่าเฉลี่ยเลขคณิต 115 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 12 ถ้ามหาวิทยาลัยที่ต้องการผู้มีไอคิวอย่างต่ำที่สุด 95 จะมีผู้สมัครจำนวนเท่าใดที่ไม่ได้เข้าไปเนื่องจากเหตุนี้และไม่พิจารณาคุณสมบัติอื่น

13. ในการโยนเหรียญ ๆ หนึ่ง 400 ครั้ง จงใช้โค้งปกติประมาณความน่าจะเป็นที่จะได้หัว

    13.1 ตั้งแต่ 185 ถึง 210 ครั้ง

    13.2 205 ครั้ง

    13.3 น้อยกว่า 176 หรือมากกว่า 227 ครั้ง

14. ในการตรวจสอบคุณภาพสินค้าโดยการสุ่มสินค้ามาตรวจสอบ 700 ชิ้น ถ้าทราบว่ามีสินค้าเสียปนอยู่ 6% จงหาความน่าจะเป็นที่

    14.1 ได้สินค้าเสียมากกว่า 50 ชิ้น

    14.2 ได้สินค้าน้อยกว่าหรือเท่ากับ 45 ชิ้น

15. บริษัทผู้ผลิตสินค้าชนิดหนึ่ง ได้จ้างบริษัทโฆษณาสินค้าของเขา โดยบริษัทอ้างว่าจะมีคนคุ้นเคยกับสินค้าหลังโฆษณา 30% เพื่อทดสอบคำอ้างของบริษัทจึงสุ่มคนมา 2000 คน จงหาความน่าจะเป็นที่คนอย่างน้อย 527 คนที่รู้จักสินค้า

6.3 การแจกแจงไคสแควร์

ถ้า X เป็นตัวแปรสุ่มปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 จะได้ว่า $Z = \frac{X-\mu}{\sigma}$ เป็นการแจกแจงปกติมาตรฐาน จะเรียก

$$Z^2 = \frac{(X-\mu)^2}{\sigma^2}$$

ว่า**ตัวแปรสุ่มไคสแควร์** (Chi-square random varibale) ที่มี **องศาเสรี** (degree of freedom) เท่ากับ 1 เขียนแทนด้วย df หรือ ν เขียนแทนด้วย $\chi^2(\nu)$

ถ้าสุ่มตัวอย่างขนาด n โดย $X_1, X_2, ..., X_n$ เป็นตัวแปรสุ่มปกติ และ $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ เป็นการแจกแจงไคสแควร์ที่มีองศาเสรี n เขียนแทนด้วย $\chi^2(\nu)$ นั่นคือ

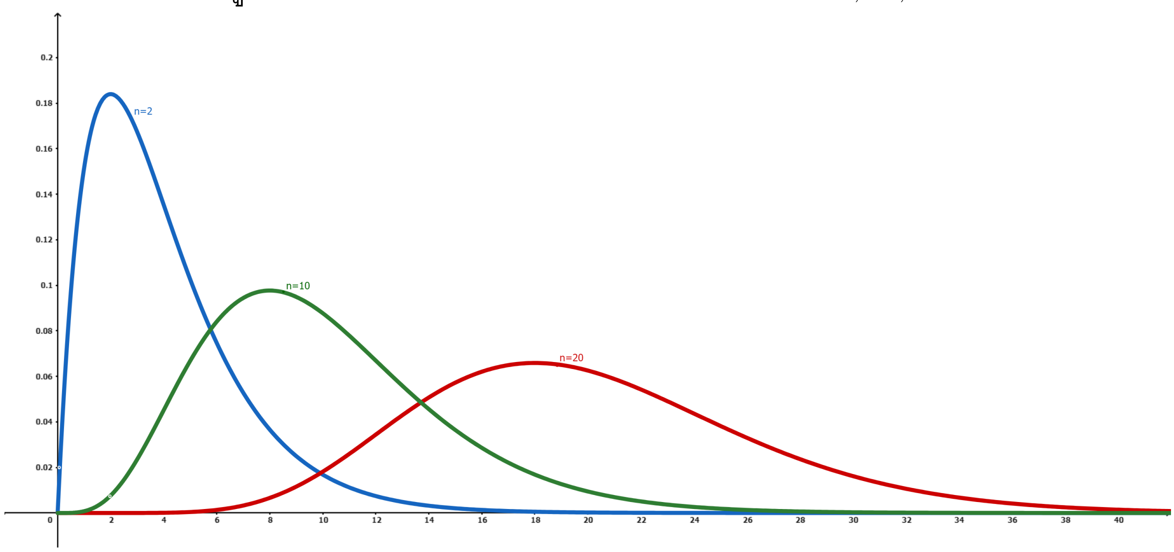
$$\chi^2(\nu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

มีฟังก์ชันความน่าจะเป็นคือ

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{x}{2}}x^{\frac{n}{2}-1} & \text{เมื่อ } x > 0 \\ 0 & \text{เมื่อ } x \text{ มีค่าอื่น ๆ} \end{cases}$$

เมื่อ Γ คือ**ฟังก์ชันแกมมา** (gamma function) นิยามโดย $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ เมื่อ $x > 0$ จะได้ว่าค่าเฉลี่ย $\mu = n$ และค่าความแปรปรวน $\sigma^2 = 2n$

รูปที่ 6.1: กราฟการแจกแจงไคสแควร์ เมื่อ $\nu = 2, 10, 20$



จากฟังก์ชันความน่าจะเป็นของไคสแควว์มีความยุ่งยากในการหาความน่าเป็น เราอาจหาค่าโดยวิธีอื่น ๆ เช่น ใช้ตารางไคสแควร์ และ ใช้แอปพลิเคชัน เป็นต้น

การใช้ตารางไคสแควร์
ตางรางไคสแควร์จะแสดงพื้นที่ด้านขวามือ นั่นคือ $P(\chi^2 > a)$ ซึ่งมักเขียนแทนด้วย

$$\chi^2_{\alpha,\nu} \quad \text{หรือ} \quad \chi^2_\alpha \text{ เป็นค่าที่ทำให้ } P(\chi^2 > \chi^2_\alpha) = \alpha$$

ตารางที่ 6.2: ตัวอย่างตารางไคสแควร์

α

ν	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.0000	0.0002	0.0010	0.0039	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.201	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	11.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750

ตัวอย่าง 6.3.1 จากตาราง 6.2 จงหาค่าต่อไปนี้

1. $\chi^2_{0.05,3}$
   เนื่องจาก $P(\chi^2 > \chi^2_{0.05,3}) = P(\chi^2 > 7.815) = 0.05$ ดังนั้น $\chi^2_{0.05,3} = 7.815$
2. $P(\chi^2 > \chi^2_{0.01,2})$
   จะได้ว่า $P(\chi^2 > \chi^2_{0.01,2}) = 0.01$
3. $P(\chi^2 < \chi^2_{0.975,5})$ จะได้ว่า $P(\chi^2 < \chi^2_{0.975,5}) = 1 - P(\chi^2 > \chi^2_{0.975,5}) = 1 - 0.975 = 0.025$
4. $P(\chi^2 > 4.605)$ เมื่อ $\nu = 2$
   จะได้ว่า $P(\chi^2 > 4.605) = P(\chi^2 > \chi^2_{0.1,2}) = 0.1$
5. $P(\chi^2 < 0.115)$ เมื่อ $\nu = 3$
   จะได้ว่า $P(\chi^2 < 0.115) = 1 - P(\chi^2 > 0.115) = 1 - P(\chi^2 > \chi^2_{0.99,3}) = 1 - 0.99 = 0.01$

ตัวอย่าง 6.3.2 ให้ X เป็นตัวแปรสุ่มไคสแควร์ที่มีองศาเสรีเท่ากับ 5 จงหาค่าของ a และ b ที่ทำให้ $P(X < a) = 0.05$ และ $P(a < X < b) = 0.9$
แนวคำตอบ เนื่องจาก

$$P(X > a) = 1 - P(X < a) = 1 - 0.05 = 0.95 = P(\chi^2 > 1.145)$$

ดังนั้น $a = 1.145$ และ

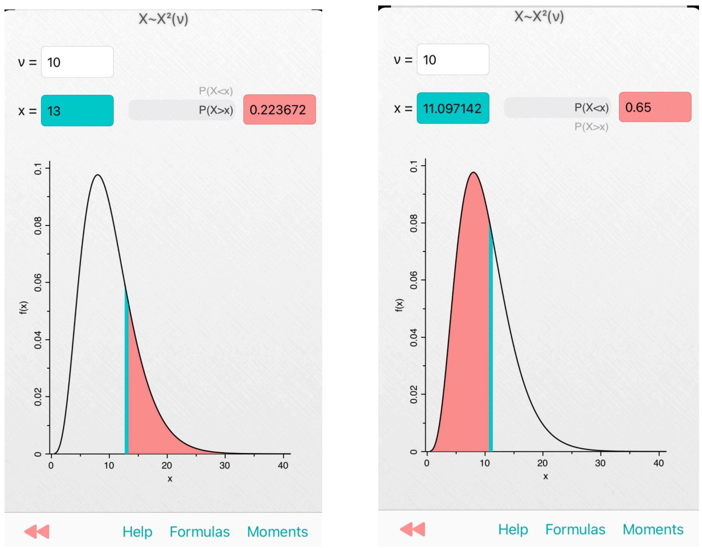
$$\begin{aligned}
P(a < X < b) &= P(X > a) - P(X > b) = P(X > 1.145) - P(X > b) \\
0.9 &= 0.95 - P(X > b) \\
P(X > b) &= 0.05 = P(\chi^2 > 11.070)
\end{aligned}$$

ดังนั้น $b = 11.070$

การใช้แอปพลิเคชัน

สำหรับ $X \sim \chi^2(\nu)$ ขอยกตัวอย่างการใช้แอปพลิเคชัน **Probability Distribution** ทำได้ 2 ลักษณะ คือการหาพื้นที่ใต้กราฟเมื่อกำหนดค่า x และหาค่า x เมื่อกำหนดพื้นที่ใต้กราฟมาให้
เช่นถ้า $X \sim \chi^2(10)$ จงหาค่า $P(X > 13)$ และ a เมื่อ $P(X < a) = 0.65$

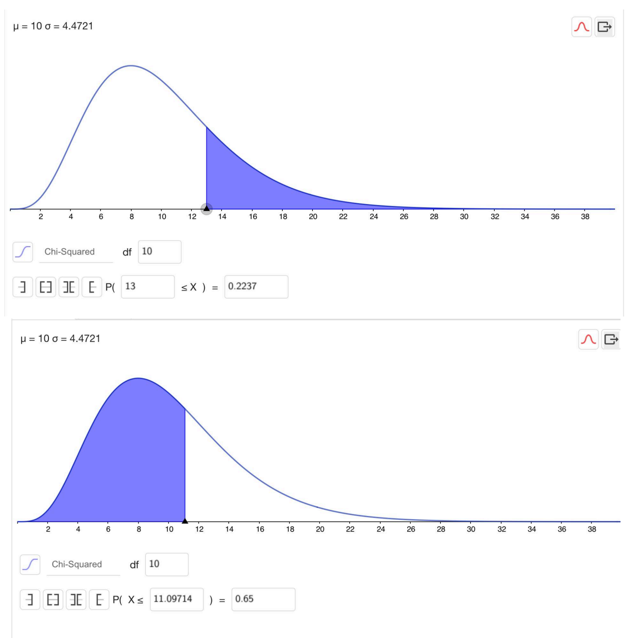
รูปที่ 6.2: ตัวอย่างการใช้แอปพลิเคชัน Probability Distribution ในการแจกแจงไคสแควร์



จะได้ว่า $P(X > 13) = 0.223672$ และ $a = 11.097142$
อาจเลือกใช้แอปพลิเคชัน **Geogebra Classic** ทำได้โดยกด $\equiv$ จากนั้น

View $\rightarrow$ Probability Calculator $\rightarrow$ Distribution $\rightarrow$ Chi-Squared

รูปที่ 6.3: ตัวอย่างการใช้แอปพลิเคชัน Geogebra Classic ในการแจกแจงไคสแควร์



6.4 การแจกแจงที

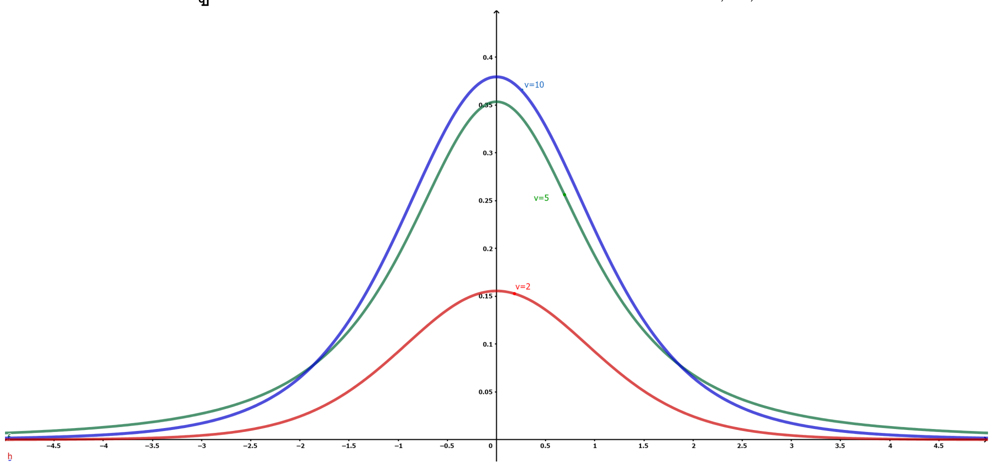
ให้ Z เป็นตัวแปรสุ่มมีการแจกแจงปกติมาตรฐาน และ V เป็นตัวแปรสุ่มไคสแควร์ที่มีองศาเสรีเป็น ν ถ้า Z และ V เป็นอิสระต่อกัน การแจกแจงของตัวแปรสุ่ม $T = \dfrac{Z}{\sqrt{\frac{V}{\nu}}}$ เขียนแทนด้วย $t(\nu)$

โดยมีฟังก์ชันความน่าจะเป็นคือ

$$h(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{เมื่อ } t \in \mathbb{R}$$

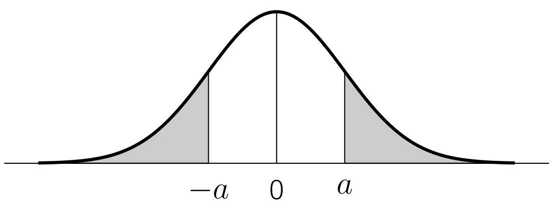
เรียกว่า **การแจกแจงสติวเดนท์ที** (Student's t-distribution) เรียกย่อ ๆ ว่า **การแจกแจงที** ตามชื่อผู้คิดค้นคือ วิลเลียม กอสเสต (William Sealey Gosset, 1876-1937) โดยใช้ชื่อในการตีพิมพ์ว่า Student เนื่องจากขณะนั้นเขาทำงานในโรงงานต้มกลั่นของพวกไอริช ซึ่งห้ามเจ้าหน้าที่โรงงานพิมพ์เผยแพร่ผลงานวิจัย

รูปที่ 6.4: กราฟการแจกแจงที เมื่อ $\nu = 2, 5, 10$



จะได้ว่าค่าเฉลี่ย $\mu = 0$ และค่าความแปรปรวน $\sigma^2 = \dfrac{\nu}{\nu-2}$ เมื่อ $\nu \geq 3$

จะเห็นได้ว่าการแจกแจงทีคล้ายกับการแจกแจงปกติ โดยเมื่อ n มีค่ามากพอการแจกแจงทีจะเป็นการแจกแจงปกติ พิจารณาจากกราฟจะเห็นว่าเป็นแบบสมมาตร ถ้า $a > 0$ จะได้ว่า



$$\begin{aligned}
P(T > a) &= P(T < -a) \\
P(|T| < a) &= 1 - 2P(T > a) = 1 - 2P(T < -a) \\
P(|T| > a) &= 2P(T > a)
\end{aligned}$$

จากฟังก์ชันความน่าจะเป็นของทีมีความยุ่งยากในการหาความน่าเป็น เราอาจหาค่าโดยวิธีอื่น เช่น ใช้ตารางที และ ใช้แอปพลิเคชัน เป็นต้น

การใช้ตารางที
ตางรางทีจะแสดงพื้นที่ด้านขวามือ นั่นคือ $P(T > a)$ ซึ่งมักเขียนแทนด้วย

$t_{\alpha,\nu}$ หรือ t_α เป็นค่าที่ทำให้ $P(T > t_\alpha) = \alpha$

ตารางที่ 6.3: ตัวอย่างตารางที

α

ν	0.10	0.05	0.025	0.01	0.005
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	2.353	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

ตัวอย่าง 6.4.1 จากตาราง 6.3 จงหาค่าต่อไปนี้

1. $t_{0.05,4}$
   เนื่องจาก $P(T > t_{0.05,4}) = P(T > 2.132) = 0.05$ ดังนั้น $t_{0.05,4} = 2.132$

2. $P(T > t_{0.01,3})$
   จะได้ว่า $P(T > t_{0.01,3}) = 0.01$

3. $P(T < t_{0.95,5})$ จะได้ว่า $P(T > t_{0.95,5}) = 1 - P(T > t_{0.95,5}) = 1 - 0.95 = 0.05$

4. $P(T > 1.533)$ เมื่อ $\nu = 4$
   จะได้ว่า $P(T > 1.533) = P(T > t_{0.1,4}) = 0.1$

5. $P(|T| < 4.541)$ เมื่อ $\nu = 3$
   จะได้ว่า $P(|T| < 4.541) = 1 - 2P(T > 4.541) = 1 - 2(0.01) = 0.98$

ตัวอย่าง 6.4.2 จากตาราง 6.3 ในการแจกแจงทีซึ่งมีองศาเสรีเท่ากับ 5 จงหาค่า a ที่ทำให้

$$P(|T| < a) = 0.90$$

แนวคำตอบ เนื่องจาก

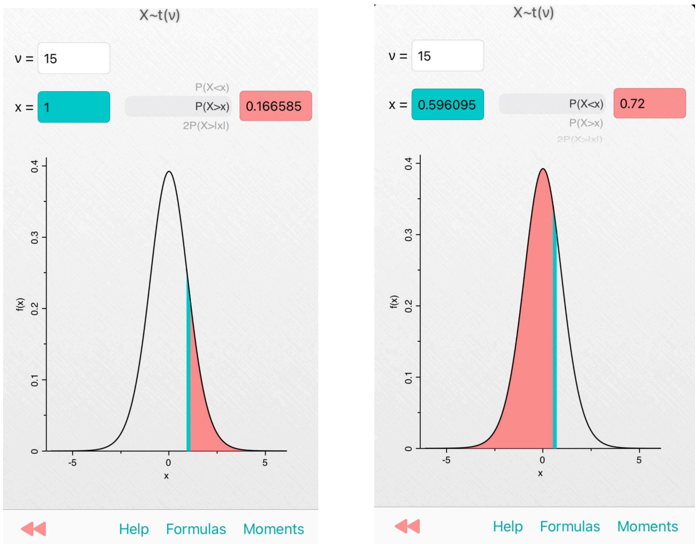
$$0.90 = P(|T| < a) = 1 - 2P(T > a)$$
$$P(T > a) = 0.05 = P(T > 2.015)$$

ดังนั้น $a = 2.015$

การใช้แอปพลิเคชัน

สำหรับ $X \sim t(\nu)$ ขอยกตัวอย่างการใช้แอปพลิเคชัน **Probability Distribution** ทำได้ 2 ลักษณะคือ การหาพื้นที่ใต้กราฟเมื่อกำหนดค่า t_α และหาค่า t_α เมื่อกำหนดพื้นที่ใต้กราฟมาให้
เช่นถ้า $X \sim t(15)$ จงหาค่า $P(X > 1)$ และ a เมื่อ $P(X < a) = 0.72$

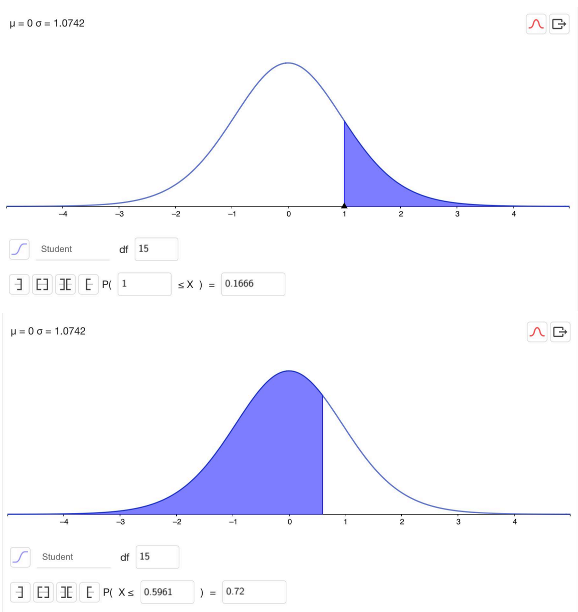
รูปที่ 6.5: ตัวอย่างการใช้แอปพลิเคชัน Probability Distribution ในการแจกแจงที



จะได้ว่า $P(X > 1) = 0.166585$ และ $a = 0.596059$
อาจเลือกใช้แอปพลิเคชัน **Geogebra Classic** ทำได้โดยกด $\equiv$ จากนั้น

View $\rightarrow$ Probability Calculator $\rightarrow$ Distribution $\rightarrow$ Student

รูปที่ 6.6: ตัวอย่างการใช้แอปพลิเคชัน Geogebra Classic ในการแจกแจงที



ในการสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงปกติ $n(x;\mu,\sigma^2)$ ด้วยค่าเฉลี่ย μ

เมื่อ**ไม่ทราบความแปรปรวน** σ^2 เราจะประมาณค่าความแปรปรวนตัวอย่าง S^2 จะได้

$$T = \frac{X-\mu}{S} \quad \text{มีการแจกแจงแบบที ที่องศาเสรี } n-1$$

ตัวอย่าง 6.4.3 ถ้าทราบน้ำหนักของลำไยกระป๋องยี่ห้อหนึ่งมีการแจกแจงแบบปกติที่มีน้ำหนักเฉลี่ย 135 กรัม ถ้าสุ่มลำไยกระป๋องยี่ห้อนี้มา 12 กระป๋อง แล้วชั่งน้ำหนักแต่ละกระป๋อง คำนวณหาค่าส่วนเบี่ยงเบนมาตรฐานของลำไยกระป๋องตัวอย่าง 10 กรัม ถ้าเลือกลำไยกระป๋องมา 1 กระป๋อง จากกลุ่มตัวอย่าง 12 กระป๋องข้างต้น จงหาความน่าจะเป็นที่ลำไยกระป๋องนั้นจะมีน้ำหนักมากกว่า 149 กรัม

แนวคำตอบ ให้ X เป็นน้ำหนักของลำไยกระป๋องยี่ห้อหนึ่ง ซึ่งมีการแจกแจงปกติ $X \sim$ N($\mu = 135, \sigma^2$) แต่เนื่องจากไม่ทราบ จะประมาณตัวแปรสุ่มด้วยที $t(12-1)$ โดยใช้ความแปรปรวน $S = 10$ และองศาเสรี $\nu = 12-1 = 11$ จะได้ว่า

$$P(X > 149) \approx P\left(\frac{X-\mu}{S} > \frac{149-135}{10}\right) = P(T > 1.4) = 0.0957$$

ดังนั้นความน่าจะเป็นที่ลำไยกระป๋องนั้นจะมีน้ำหนักมากกว่า 149 กรัม เท่ากับ 0.09573

6.5 การแจกแจงเอฟ

ตัวแปรสุ่มเอฟ (F random variable) คืออัตราส่วนระหว่างตัวแปรสุ่มไคสแควร์ 2 ค่าที่เป็นอิสระต่อกันและต่างหารด้วยองศาเสรี เขียนแทนด้วย $F(\nu_1, \nu_2)$ นั่นคือ

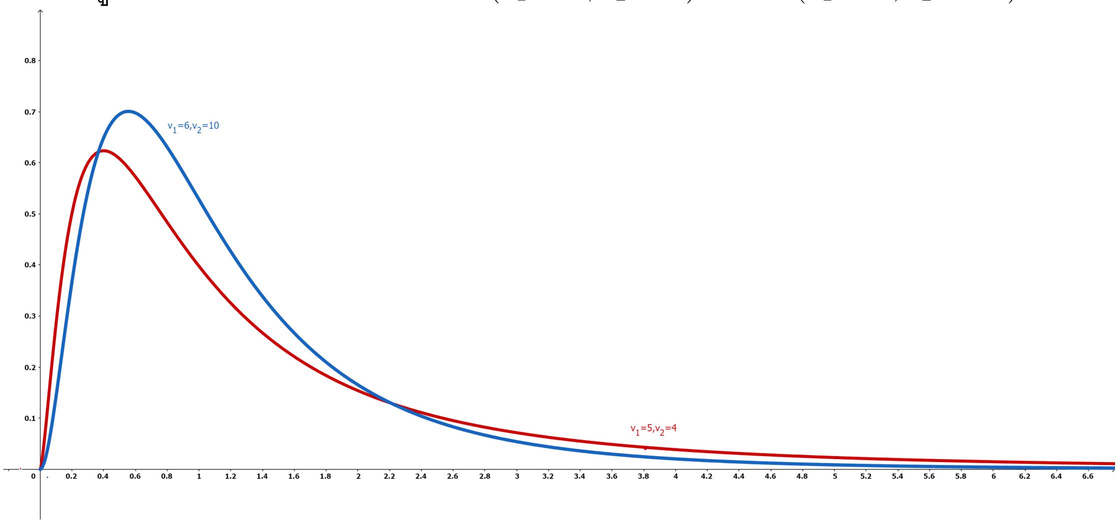
$$F = \frac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}}$$

ฟังก์ชันความน่าจะเป็นคือ

$$f(x) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\cdot\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\left(1+\frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}} \qquad \text{เมื่อ } x > 0$$

เรียกว่า **การแจกแจงเอฟ** (F-distribution) ที่มีองศาเสรี ν_1 และ ν_2

รูปที่ 6.7: กราฟการแจกแจง $F(\nu_1 = 5, \nu_2 = 4)$ และ $F(\nu_1 = 6, \nu_2 = 10)$



จะได้ว่าค่าเฉลี่ย $\mu = \dfrac{\nu_2}{\nu_2 - 2}$ เมื่อ $\nu_2 \geq 2$ และค่าความแปรปรวน $\sigma^2 = \dfrac{2\nu_2^2(\nu_1+\nu_2-2)}{(\nu_2-4)\nu_1(\nu_2-2)^2}$ เมื่อ $\nu_2 > 4$

จากฟังก์ชันความน่าจะเป็นของเอฟมีความยุ่งยากในการหาความน่าเป็น เราอาจหาค่าโดยวิธีอื่น เช่น ใช้ตารางที และ ใช้แอปพลิเคชัน เป็นต้น

การใช้ตารางเอฟ

ตางรางทีจะแสดงพื้นที่ด้านขวามือ นั่นคือ $P(F > a)$ ซึ่งมักเขียนแทนด้วย

$$f_{\alpha,(\nu_1,\nu_2)} \text{ หรือ } f_\alpha \text{ เป็นค่าที่ทำให้ } P(F > f_\alpha) = \alpha$$

ตารางที่ 6.4: ตัวอย่างตารางเอฟ $\alpha = 0.05$

ν_1

ν_2	1	2	3	4	5	6	7	8	9
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77

ตัวอย่าง 6.5.1 จากตาราง 6.4 จงหาค่าต่อไปนี้

1. $f_{0.05,(5,4)}$
   เนื่องจาก $P(F > f_{0.05,(5,4)}) = P(t > 6.26) = 0.05$ ดังนั้น $f_{0.05,(5,4)} = 6.26$

2. $f_{0.05,(4,5)}$
   เนื่องจาก $P(F > f_{0.05,(4,5)}) = P(t > 5.19) = 0.05$ ดังนั้น $f_{0.05,(4,5)} = 5.19$

3. $P(F < f_{0.05,(3,7)})$ จะได้ว่า $P(F < f_{0.05,(3,7)}) = 1 - P(F > f_{0.05,(3,7)}) = 1 - 0.05 = 0.95$

ทฤษฎีบท 6.5.2 ถ้า $f_{\alpha,(\nu_1,\nu_2)}$ คือค่าของ f_α ซึ่งมีองศาเสรีคือ ν_1 และ ν_2 จะได้ว่า

$$f_{1-\alpha,(\nu_1,\nu_2)} = \frac{1}{f_{\alpha,(\nu_2,\nu_1)}}$$

บทพิสูจน์. ขอละไว้ในวิชานี้ □

ตัวอย่าง 6.5.3 จากตาราง 6.4 จงหาค่าต่อไปนี้

1. $f_{0.95,(3,4)}$
   จะได้ว่า $f_{0.95,(3,4)} = \frac{1}{f_{0.05,(4,3)}} = \frac{1}{9.12} = 0.1096$

2. $f_{0.95,(4,5)}$
   จะได้ว่า $f_{0.95,(4,5)} = \frac{1}{f_{0.05,(5,4)}} = \frac{1}{6.26} = 0.1597$

ตัวอย่าง 6.5.4 กำหนดให้ F เป็นตัวแปรสุ่มแอฟที่มีองศาเสรีเท่ากับ 9 และ 5 ตามลำดับ จงหาค่าต่อไปนี้โดยใช้ตาราง

1. $P(0 < F < 6.68)$
   แนวคำตอบ จะได้ว่า

   $$P(0 < F < 6.68) = 1 - P(F > 6.68) = 1 - P(F > f_{0.025,(9,5)}) = 1 - 0.025 = 0.975$$

2. ค่า k ที่ทำให้ $P(F > k) = 0.01$
   แนวคำตอบ เนื่องจาก

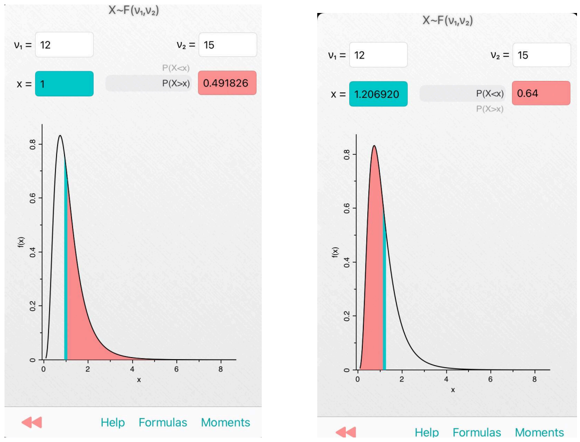
   $$P(F > k) = 0.01 = P(F > f_{0.01,(9,5)}) = P(F > 10.16)$$

   ดังนั้น $k = 10.16$

การใช้แอปพลิเคชัน

สำหรับ $X \sim F(\nu_1, \nu_2)$ ขอยกตัวอย่างการใช้แอปพลิเคชัน **Probability Distribution** ทำได้ 2 ลักษณะ คือการหาพื้นที่ใต้กราฟเมื่อกำหนดค่า f_α และหาค่า f_α เมื่อกำหนดพื้นที่ใต้กราฟมาให้ เช่นถ้า $X \sim F(\nu_1 = 12, \nu_2 = 15)$ จงหาค่า $P(X > 1)$ และ a เมื่อ $P(X < a) = 0.64$

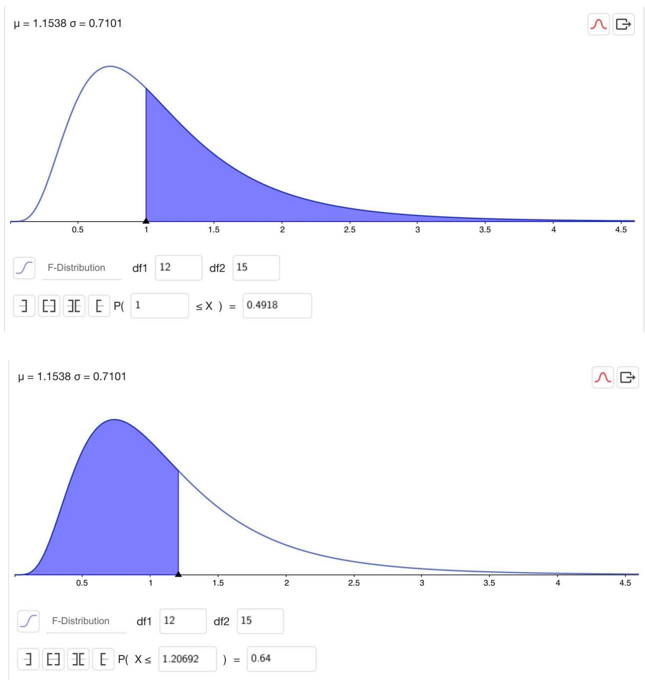
รูปที่ 6.8: ตัวอย่างการใช้แอปพลิเคชัน Probability Distribution ในการแจกแจงเอฟ



จะได้ว่า $P(X > 1) = 0.491826$ และ $a = 1.206920$
อาจเลือกใช้แอปพลิเคชัน **Geogebra Classic** ทำได้โดยกด $\equiv$ จากนั้น

View $\rightarrow$ Probability Calculator $\rightarrow$ Distribution $\rightarrow$ F-distribution

รูปที่ 6.9: ตัวอย่างการใช้แอปพลิเคชัน Geogebra Classic ในการแจกแจงเอฟ



แบบฝึกหัด 6.3 - 6.5

1. ให้ $X \sim \chi^2(20)$ จงหาค่า

   1.1 $\chi^2_{0.975}$

   1.2 a และ b ที่ทำให้ $P(X < a) = 0.075$ และ $P(a < X < b) = 0.90$

2. จงหาค่าต่อไปนี้ โดยใช้ตารางแจกแจงที

   2.1 $t_{0.95,9}$

   2.2 $t_{0.05,12}$

   2.3 จงหา a ที่ทำให้ $P(|T| < a) = 0.95$ ถ้าองศาเสรีเท่ากับ 18

3. จงหาค่าต่อไปนี้ โดยใช้ตารางแจกแจงเอฟ

   3.1 $f_{0.99,(9,15)}$

   3.2 $t_{0.01,(14,20)}$

   3.3 จงหา a และ b ที่ทำให้ $P(a < F < b) = 0.90$ และ $P(F < a) = 0.05$ ถ้าองศาเสรีเท่ากับ 9 และ 15 ตามลำดับ

4. จงใช้ตารางหาค่าต่อไปนี้

   4.1 $\chi^2_{0.99,23}$

   4.2 $\chi^2_{0.90,38}$

   4.3 $t_{0.975,17}$

   4.4 $t_{0.01,10}$

   4.5 $f_{0.95,(10,9)}$

   4.6 $f_{0.01,(20,7)}$

5. ในการแจกแจงไคสแควร์ จงหาค่า a ที่ทำให้ $P(\chi^2 < a) = 0.05$ ที่องศาเสรี 16

6. ในการแจกแจงที จงหาค่า k ที่ทำให้ $P(|T| < k) = 0.9$ ที่องศาเสรี 23

7. ถ้าทราบว่าน้ำหนักของคนไทยที่อายุ 20 ปี มีการแจกแจงปกติที่มีค่าเฉลี่ย 47 กิโลกรัม ถ้าสุ่มตัวอย่างคนไทยที่มีอายุ 20 ปีมา 10 คน แล้วชั่งน้ำหนักและคำนวณค่าส่วนเบี่ยงเบนมาตรฐานได้ 5.4 กิโลกรัม ถ้าสุ่มคนไทยมา 1 คนจากกลุ่มนี้ จงหาความน่าจะเป็นที่คนที่สุ่มได้จะมีน้ำหนักมากกว่า 50 กิโลกรัม

สรุป

การแจกแจงความน่าจะเป็นต่อเนื่องที่ใช้กันอย่างแพร่หลายประกอบด้วย การแจกแจงยูนิฟอร์ม คือการแจกแจงของตัวแปรสุ่มที่ทุกค่าในช่วง (a,b) มีความน่าจะเป็นเท่า ๆ กัน จากนั้นกล่าวถึง การแจกแบบเกาส์หรือที่รู้จักกันดีในชื่อ การแจกแจงปกติ $X \sim N(\mu,\sigma^2)$ ซึ่งมีความน่าจะเป็น ของตัวแปรสุ่มที่ส่วนมากมีค่าใกล้เคียงค่าเฉลี่ย ซึ่งสามารถแปลงเป็น การแจกแจงปกติมาตรฐาน $Z \sim N(0,1)$ ด้วย $Z = \frac{X-\mu}{\sigma}$ แล้วนำไปใช้ในการอธิบายความเป็นไปได้ของเหตุการณ์ต่าง ๆ ที่สนใจ โดยอ่านค่าจากตาราง หรือแอปพลิเคชั่น สำหรับการแจกแจงทวินาม Bin(n,p) ที่ n มีค่ามาก ๆ อาจประมาณด้วยการแจกแจงปกติได้โดย $\mu = np$ และ $\sigma^2 = np(1-p)$ จากนั้นกล่าวถึงการแจกแจงไคสแควร์ ที่เกิดจากผลรวมของกำลังสองของการแจกแจงปกติมาตรฐาน n ชุด โดยเรียกค่านี้ว่าองศาเสรี ถัดมากล่าวถึงการแจกแจงทีซึ่งเกิดจากของตัวแปรสุ่มปกติและไคสแควร์ซึ่งมีกราฟลักษณะเดียวกับการแจกแจงปกติ แต่จะเปลี่ยนไปตามองศาเสรี และสุดท้ายกล่าวถึงการแจกแจงเอฟคือ อัตราส่วนระหว่างตัวแปรสุ่มไคสแควร์ 2 ค่าที่เป็นอิสระต่อกันและต่างหารด้วยองศาเสรี

แบบฝึกหัดบทที่ 6

1. จงหาฟังก์ชันการแจกแจงความน่าจะเป็นสะสมของ $X \sim$ Unif(a,b)

2. ให้ X เป็นตัวแปรสุ่มยูนิฟอร์มบนช่วง $(2,7)$ จงหาความน่าจะเป็นที่ X จะมีค่าน้อยกว่า 3

3. ถ้าอายุการใช้งานของเครื่องซักผ้ายี่ห้อ A มีการแจกแจงแบบปกติด้วยค่าเฉลี่ย 5.10 ปี และส่วนเบี่ยงเบนมาตรฐานเป็น 1.54 ปี ถ้าเครื่องซักผ้ายี่ห้อ A รับประกันไว้ 2 ปี จงหาโอกาสที่เครื่องซักผ้ายี่ห้อ A จะเสียในช่วงเวลาที่รับประกัน

4. ถ้าเวลาที่ใช้รอลิพของอาคารแห่งหนึ่งมีการแจกแจงยูนิฟอร์มระหว่าง 0 ถึง 30 วินาที จงหาความแปรปรวนของเวลาที่ใช้รอลิพ

5. ให้ X เป็นตัวสุ่มต่อเนื่องแบบยูนิฟอร์มบนช่วง $(1,b)$ ถ้า $E(X)=3$ จงหา b

   เมื่อ $x \in \mathbb{R}$ แล้ว σ^2_{2X+1} มีค่าเท่าใด

6. ถ้า f ฟังก์ชันการแจกแจงของตัวแปรสุ่ม X และ $f(x) = \dfrac{1}{5\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-20}{5}\right)^2}$ เมื่อ $x \in \mathbb{R}$ แล้ว $P(X>25)$ มีค่าเท่าใด

7. ถ้าเชื่อกันว่าอัตราผลตอบแทนของการลงทุนมีการแจกแจงปกติซึ่งมีค่าเฉลี่ย 20% ค่าส่วนเบี่ยงเบนมาตรฐาน 3% จงหาความน่าจะเป็นที่อัตราผลตอบแทนจะ

   7.1 มากกว่า 26% หรือน้อยกว่า 14%

   7.2 อย่างน้อย 6%

   7.3 มากกว่า 28.5%

8. บริษัทผลิตหลอดไฟต้องการรับประกันคุณภาพผลิตภัณฑ์ของบริษัท โดยจะเปลี่ยนหลอดไฟใหม่ถ้าหลอดเดิมชำรุดบริษัทจะรับประกันไม่เกิน 4.1% ของจำนวนที่ผลิต หลอดไฟมีอายุการใช้งานเฉลี่ย 2500 ชั่วโมง มีสัมประสิทธิ์ของความแปรผันเท่ากับ 0.2 ถ้าคาดว่าตามปกติคนจะใช้หลอดไฟวันละ 5 ชั่วโมง บริษัทนี้ควรกำหนดเวลาประกันมากที่สุดกี่วัน

9. กำหนดให้ข้อมูลชุดหนึ่งมีการแจกแจงปกติ ถ้าหยิบข้อมูล x และ y จากข้อมูลชุดนี้มาพิจารณาพบว่า 13.14% ของข้อมูลมีค่ามากกว่า x และ x มากกว่า y อยู่ 2% ของส่วนเบี่ยงเบนมาตรฐาน แล้วจำนวนข้อมูล (คิดเป็นเปอร์เซ็นต์) ที่มีค่าน้อยกว่า y เท่ากับเท่าใด

10. ในการตรวจสอบคุณภาพสินค้าโดยการสุ่มสินค้ามาตรวจสอบ 700 ชิ้น ถ้าทราบว่ามีสินค้าเสียปนอยู่ 6% จงหาความน่าจะเป็นที่

    10.1 ได้สินค้าเสียมากกว่า 50 ชิ้น

    10.2 ได้สินค้าน้อยกว่าหรือเท่ากับ 45 ชิ้น

11. เส้นผ่านศูนย์กลางส่วนในวงแหวนมีการแจกแจงปกติ ค่าเฉลี่ยเลขคณิต 4.00 นิ้ว ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.01 นิ้ว

    11.1 จงหาสัดส่วนของวงแหวนที่มีเส้นผ่านศูนย์กลางส่วนในมากกว่า 4.025 นิ้ว

    11.2 จงหาความน่าจะเป็นวงแหวนเหล่านี้มีเส้นผ่านศูนย์กลางส่วนในอยู่ระหว่าง 3.99 และ 4.01 นิ้ว

    11.3 เส้นผ่านศูนย์กลางส่วนในที่มีวงแหวนอยู่ 15% ต่ำกว่าค่าใด

12. ในการเข้าศึกษาต่อหลักสูตรนานาชาติทั้งในประเทศและต่างประเทศ มักใช้ผลสอบ SAT Math ประกอบการพิจารณาเพื่อเข้าศึกษา ซึ่งในปี 2017-2019 (ในประเทศไทย) รายงานคะแนนดังต่อไปนี้

ปี	2017	2018	2019
คะแนนสูงสุด	800	800	800
คะแนนต่ำสุด	660	660	670
คะแนนเฉลี่ย	752	762	770

สมมติว่าคะแนนในแต่ละปีมีการแจกแจงปกติ ในปี 2019 พบว่ามีนักเรียนที่สอบได้คะแนนมากกว่า 780 คะแนนอยู่ 15%

    12.1 จงหาความแปรปรวนของคะแนนสอบในปี 2019

    12.2 อยากทราบว่ามีนักเรียนอยู่กี่เปอร์เซ็นต์ที่ได้คะแนนมากกว่าคะแนนต่ำสุด 90 คะแนนขึ้นไป ในปี 2019

    12.3 ในปี 2019 ถ้ามีนักเรียนเข้าสอบทั้งหมด 1000 คน มีนักเรียนประมาณกี่คนที่ได้คะแนนมากกว่าคะแนนต่ำสุด 50 คะแนน แต่น้อยกว่าคะแนนสูงสุด 50 คะแนน

13. ระดับคะแนน (GPAX) ของนักศึกษาคณะครุศาสตร์ ประจำปีการศึกษาหนึ่งมีการแจกแจงปกติ โดยที่ระดับคะแนนเฉลี่ยเท่ากับ 2.65 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.30 จงหา

    13.1 จำนวนคนที่มีระดับคะแนนมากกว่า 2.50 คิดเป็นกี่เปอร์เซ็นต์

    13.2 ถ้าคนที่มีระดับคะแนนอยู่ระหว่าง 3.00 ถึง 3.50 มีจำนวน 300 คน จงหาจำนวนนักศึกษาทั้งหมด

14. คะแนนสอบของนักเรียนจำนวน 1000 คน มีการแจกแจงปกติ มีค่าเฉลี่ยเลขคณิต 45 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 คะแนน จงหา

    14.1 จำนวนคนที่ได้คะแนนน้อยกว่า 60

    14.2 จำนวนคนที่ได้คะแนนอยู่ระหว่าง 30 และ 50

    14.3 คนที่คะแนนต่ำสุดของ 15% สูงสุด

15. บริษัทผลิตหลอดไฟต้องการรับประกันคุณภาพผลิตภัณฑ์ของบริษัท โดยจะเปลี่ยนหลอดไฟใหม่ถ้าหลอดเดิมชำรุดบริษัทจะรับประกันไม่เกิน 4.1% ของจำนวนที่ผลิต หลอดไฟมีอายุการใช้งานเฉลี่ย 2500 ชั่วโมง มีสัมประสิทธิ์ของความแปรผันเท่ากับ 0.2 ถ้าคาดว่าตามปกติคนจะใช้หลอดไฟวันละ 5 ชั่วโมง บริษัทนี้ควรกำหนดเวลาประกันมากที่สุดกี่วัน

16. นายเดชาทราบว่าเวลาที่เขาเดินทางจากบ้านมายังมหาวิทยาลัยมีการแจกแจงปกติที่มีค่าเฉลี่ย 90 นาที ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 นาที นายเดชาควรออกจากบ้านกี่โมงในตอนเช้า ที่จะทำให้เขามั่นใจ 95% ว่าจะมาถึงมหาวิทยาลัยในเวลา 8 โมงเช้า

17. ข้อสอบแบบ 4 ตัวเลือกซึ่งมีคำตอบที่ถูกต้องรวมอยู่ 1 ข้อ จำนวน 200 ข้อ ถ้ามีข้อสอบ 80 ข้อที่คาดว่านักเรียนไม่มีความรู้ที่จะตอบได้จริง ๆ จงหาโอกาสที่นักเรียนจะตอบถูกโดยการเดาได้ตั้งแต่ 25 ถึง 30 ข้อ (ทำโดยใช้ การแจกแจงทวินามและประมาณด้วยการแจกแจงปกติ)

18. คะแนนสอบของวิชาคณิตศาสตร์ ONET ระดับประถมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งมีการแจกแจงปกติ โดยมีคะแนนเฉลี่ยเท่ากับ 65 คะแนน ถ้าสุ่มนักเรียนที่สอบมา 15 คน คำนวณส่วนเบี่ยงเบนมาตรฐานได้ 10 คะแนน จงหาความน่าจะเป็นที่นักเรียนที่เข้าสอบมีคะแนนไม่น้อยกว่า 50 คะแนน แต่ไม่เกิน 80 คะแนน

19. ให้ t เป็นการแจกแจงที และ χ^2 เป็นการแจกแจงไคสแควร์ มีองศาเสรีเท่ากันคือ $\nu = 10$ ถ้า a และ b เป็นค่าคงที่ซึ่งทำให้ $P(|t| < a - b) = 0.99$ และ $P(\chi^2 < a + b) = 0.10$ จงหา a และ b

20. ในการแจกแจง $F(\nu_1, \nu_2)$ กำหนดให้ $f_{0.94,(\nu_1,\nu_2)} = 0.37$ จงหาค่าของ $f_{0.06,(\nu_2,\nu_1)}$

21. ในการแจกแจงไคสแควร์จงหา a และ b ที่ทำให้ $P(\chi^2 < a) = 0.025$ และ $P(a < \chi^2 < b) = 0.97$ ถ้าองศาเสรีเท่ากับ 10

22. ในการแจกแจงเอฟ จงหา a และ b ที่ทำให้ $P(a < F < b) = 0.90$ และ $P(F < a) = 0.05$ ถ้าองศาเสรีเท่ากับ 8 และ 16 ตามลำดับ

บทที่ 7

การสุ่มตัวอย่าง

เมื่อเราต้องการทราบข้อมูลทางสถิติบางอย่าง เช่น รายได้ของคนไทย สิ่งแรกที่ต้องทำคือการสำรวจหรือเก็บข้อมูลของรายได้ของคนไทย เราจะเรียกคนไทยทุกคนว่า **ประชากร** (Population) ของการสำรวจในครั้งนี้ ถ้าเราสนใจน้ำหนักของนักเรียนในกรุงเทพมหานคร ประชากรคือ นักเรียนทุกคนในกรุงเทพมหานคร ดังนั้นประชากรในทางสถิติจึงมีความหมายกว้างกว่าความคำว่าประชากรโดยทั่วไป

จำนวนของค่าสังเกตทั้งหมดของประชากร คือตัวเลขที่บอกขนาดของประชากร เช่นสำรวจคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้น ม. 3 ของโรงเรียนแห่งหนึ่งที่มีจำนวนทั้งหมด 450 คน จะได้ว่าขนาดของประชากรในการสำรวจครั้งนี้คือ 450 ประชากรอาจแบ่งตามขนาดได้ 2 ชนิดคือ

1. **ประชากรที่มีจำนวนแน่นอน** (Finite population) เช่น จำนวนรถโดยสารประจำทางในกรุงเทพมหานคร จำนวนนักศึกษาสาขาวิชาคณิตศาสตร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา เป็นต้น

2. **ประชากรที่มีจำนวนอนันต์** (Infinite population) เช่น จำนวนข้าวที่เก็บเกี่ยวได้ในปีหนึ่ง ๆ และจำนวนครั้งที่ทอดลูกเต๋าจนได้แต้ม 5 เป็นต้น

จะเห็นว่าการเก็บรวบรวมข้อมูลทุกหน่วยในประชากรที่มีขนาดใหญ่ อาจเกิดความยุ่งยาก ใช้เวลามากและค่าใช้จ่ายสูง ผู้วิจัยอาจเก็บข้อมูลบางส่วนของประชากร ซึ่งจะเรียกว่า **ตัวอย่าง** (Sample) เช่น ต้องการหาอายุเฉลี่ยของประชากรไทย ประชากรคือคนไทยทุกคน ตัวอย่างคือคนไทยบางส่วนที่ถูกเลือกเป็นตัวอย่าง ตัวอย่างจะเป็นตัวแทนที่ดีของประชากรหรือไม่นั้น ก็อยู่ที่ว่าลักษณะของตัวอย่างเหมือนกับประชากรมากน้อยเพียงใด และผลสรุปจะถูกต้องเพียงใด ผลสรุปจะถูกต้องเพียงใดขึ้นอยู่กับ วิธีการเลือกตัวอย่าง และวิธีกการประมาณค่าต่าง ๆ ทางสถิติ

การสุ่มตัวอย่างที่มักนิยมใช้มี 4 แบบ คือ

1. **การสุ่มตัวอย่างแบบธรรมดา** (sample random sampling) หมายถึงการเลือกตัวอย่างโดยไม่เจาะจงจากประชากรทั้งหมด ซึ่งอาจทำได้โดยใช้ตารางเลขสุ่ม หรือการจับฉลาก

2. **การสุ่มตัวอย่างแบบแบ่งพวก** (stratified random sampling) ทำได้โดยการแบ่งประชากรออกเป็นพวก (stratum) ตามลักษณะของของประชากร เช่นแบ่งนักศึกษาคณะครุศาสตร์ออกเป็นสาขาวิชา เป็นต้น แล้วสุ่มตัวอย่างแบบธรรมดา

3. **การสุ่มตัวอย่างมีระบบ** (systematic random sampling) ทำได้โดยการกำหนดหมายเลขให้แก่ประชากรเสียก่อน จากนั้นเลือกสมาชิกตัวแรกด้วยวิธีสุ่ม (random start) จากหมายเลข 1 ถึง k ต่อจากนั้นเลือกทุก ๆ หน่วยที่ k เมื่อ $k = \frac{N}{n}$ เมื่อ N คือขนาดประชากร และ n ขนาดของตัวอย่าง เช่น $k = 7$ สุ่มตัวแรกได้เลข 2 ตัวเลขที่ต้องการประกอบด้วยค่าสังเกตของสมาชิกหมายเลข 2, 9, 16, 23, ...

4. **การสุ่มตัวอย่างแบบแบ่งกลุ่ม** (cluster random sampling) เริ่มด้วยการแบ่งประชากรออกเป็นกลุ่ม ๆ เสียก่อนแล้วเลือกเอาเพียงบางกลุ่มเป็นตัวอย่าง โดยอาจจะใช้ประชากรทั้งหมดในกลุ่มนั้น หรือเลือกแต่เพียงบางส่วนของประชากรในกลุ่มที่ถูกเลือกมาเป็นตัวอย่าง

7.1 ทฤษฎีการสุ่มตัวอย่าง

จุดประสงค์ของการสุ่มตัวอย่างคือการประมาณค่า **พารามิเตอร์** (parameter) ซึ่งเป็นค่าคำนวณของประชากร และค่าที่คำนวณได้จากตัวอย่างเรียกว่า **สถิติ** (statistics) เนื่องจากตัวอย่างที่สุ่มจากประชากรมีได้หลายชุดต่าง ๆ กัน ค่าสถิติจึงเปลี่ยนไปตามตัวอย่าง
สัญลักษณ์ที่ใช้แทนค่าสถิติและพารามิเตอร์ที่มักพบบ่อยครั้ง

	สถิติ ($\hat{\theta}$)	พารามิเตอร์ (θ)
ค่าเฉลี่ยเลขคณิต (Arithematic Mean)	$\bar{X}$	μ
ค่าความแปรปรวน (Varian)	S^2	σ^2
ค่าสัดส่วน (Proprotion)	$\hat{P}$	p

บทนิยาม 7.1.1 ถ้า $X_1, X_2, ..., X_n$ แทนตัวอย่างสุ่มขนาด n แล้ว
ค่าเฉลี่ยเลขคณิตของตัวอย่าง (samaple mean) คือสถิติ

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

ความแปรปรวนของตัวอย่าง (sample varian) คือสถิติ

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

และเรียก S ว่าส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
และ**สัดส่วนของตัวอย่าง** (sample proprotion) คือสถิติ

$$\hat{P} = \frac{X}{n}$$

เมื่อ X จำนวนที่สนใจในการสำรวจที่ได้จากตัวอย่าง

ตัวอย่าง 7.1.2 คะแนน Quiz1 วิชาแคลคูลลัส ๒ ของนักศึกษา 5 คน ประกอบด้วย

5, 6, 8, 9, 10

จงหา ค่าเฉลี่ย ความแปรปรวน และสัดส่วนของคนที่ได้คะแนนมากกว่า 8 ของตัวอย่างทุกชุด ที่เกิดขึ้นจากการสุ่มตัวอย่างขนาด 3 จากประชากรดังกล่าว พร้อมหาค่าทั้ง 3 ของประชากร
แนวคำตอบ ตัวอย่างขนาด $n = 3$ แสดงดังตาราง

ชุดที่	ตัวอย่าง	ค่าเฉลี่ย : $\bar{X}$	ความแปรปรวน : S^2	สัดส่วน : $\hat{P}$
1	$5, 6, 8$	6.33	2.33	0
2	$5, 6, 9$	6.67	4.33	1/3
3	$5, 6, 10$	7.00	7.00	1/3
4	$5, 8, 9$	7.33	4.33	1/3
5	$5, 9, 10$	8.00	7.00	2/3
6	$5, 8, 10$	7.67	6.33	1/3
7	$6, 8, 9$	7.67	2.33	1/3
8	$6, 8, 10$	8.00	4.00	1/3
9	$6, 9, 10$	8.33	4.33	2/3
10	$8, 9, 10$	9.00	1.00	2/3

สำหรับค่าทั้ง 3 จากประชากร 5, 6, 8, 9, 10 คือ

ค่าเฉลี่ย $\mu = 7.6$ ความแปรปรวน $\sigma^2 = 3.44$ และสัดส่วน $p = \frac{2}{3}$

ทฤษฎีบท 7.1.3 ถ้า S^2 เป็นความแปรปรวนของตัวอย่าง $X_1, X_2, ..., X_n$ จะได้ว่า

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

บทพิสูจน์. จะได้ว่า

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2 - 2\bar{X}X_i + \bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}(n\bar{X}) + n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right) = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}
\end{aligned}$$

□

7.2 การแจกแจงของค่าเฉลี่ยเลขคณิตของตัวอย่าง

การสุ่มตัวอย่างขนาด n จากประชากรขนาด N โดยการทำการทดลองทั้งหมด m ครั้ง นั่นคือ

$$^{N}C_n = m$$

เมื่อคำนวณค่าเฉลี่ยเลขคณิตของตัวอย่างที่สุ่มได้แต่ละครั้ง

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, ..., \bar{x}_m$$

ค่าที่ได้อาจแตกต่างกันไป นั่นหมายถึงค่าเฉลี่ยเลขคณิตมีการกระจายที่แตกต่างกัน สิ่งที่สนใจในหัวข้อนี้คือ การแจกแจงของค่าเฉลี่ยเลขคณิตของตัวอย่าง

บทนิยาม 7.2.1 ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นของสถิติ $\hat{\theta}$ เรียกว่า

ความคลาดเคลื่อนมาตรฐาน (standard error) หรือ **ส่วนเบี่ยงเบนมาตรฐาน**

ของสถิติ $\hat{\theta}$ นั้น เขียนแทนด้วย $\sigma_{\hat{\theta}}$

ตัวแปร $\hat{\theta}$ ในบทนิยามนี้หมายถึง $\bar{X}$, S^2 และ $\hat{P}$ เป็นต้น

กำหนดประชากรทั้งหมดคือ $x_1, x_2, ..., x_N$ มีค่าเฉลี่ยเลขคณิตของประชากรเท่ากับ μ และความแปรปรวนของประชากรเท่ากับ σ^2 ดังนั้น

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{และ} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2$$

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างขนาด n ซึ่งสุ่มแบบธรรมดาจากประชากร นั้นคือ สำหรับแต่ละค่าของ $i = 1, 2, ..., n$

$$P(X_i = x_j) = \frac{1}{N} \quad \text{สำหรับแต่ละค่าของ } j = 1, 2, ..., N$$

จะได้ว่า

$$E(X_i) = \sum_{j=1}^{N} x_j P(X_i = x_j) = \sum_{j=1}^{N} x_j \cdot \frac{1}{N} = \frac{1}{N}\sum_{j=1}^{N} x_j = \mu$$

$$\begin{aligned} \text{var}(X_i) &= E(X_i^2) - E^2(X_i) = \sum_{j=1}^{N} x_j^2 P(X_i = x_j) - \mu^2 \\ &= \sum_{j=1}^{N} x_j^2 \cdot \frac{1}{N} - \mu^2 = \frac{1}{N}\sum_{j=1}^{N} x_j^2 - \mu^2 = \sigma^2 \end{aligned}$$

ให้ $i, j = 1, 2, ..., n$ พิจารณาความแปรปรวนร่วมของแต่ละคู่ จะได้ว่า

$$P(X_i = x_r, X_j = x_s) = \frac{1}{N(N-1)} \quad \text{สำหรับแต่ละค่าของ } r, s = 1, 2, ..., N \text{ ซึ่ง } r \neq s$$

คือ

$$\begin{aligned}\text{cov}(X_i, X_j) &= E(X_iX_j) - E(X_i)E(X_j)\\ &= \sum_{r\neq s} x_rx_sP(X_i = x_r, X_j = x_s) - \mu^2\\ &= \sum_{r\neq s} x_rx_s \cdot \frac{1}{N(N-1)} - \mu^2\\ &= \frac{1}{N(N-1)} \cdot \sum_{r\neq s} x_rx_s - \mu^2\end{aligned}$$

พิจารณา

$$\begin{aligned}\sum_{r\neq s} x_rx_s &= x_1(x_1 + x_2 + \cdots + x_N) - x_1^2\\ &\quad + x_2(x_1 + x_2 + \cdots + x_N) - x_2^2\\ &\quad + x_3(x_1 + x_2 + \cdots + x_N) - x_3^2\\ &\quad\ \vdots\\ &\quad + x_N(x_1 + x_2 + \cdots + x_N) - x_N^2\\ &= (x_1 + x_2 + \cdots + x_N)(x_1 + x_2 + \cdots + x_N) - (x_1^2 + x_2^2 + \cdots + x_N^2)\\ &= (x_1 + x_2 + \cdots + x_N)^2 - (x_1^2 + x_2^2 + \cdots + x_N^2)\\ &= \left(\sum_{j=1}^{N} x_j\right)^2 - \sum_{j=1}^{N} x_j^2\\ &= (N\mu)^2 - (\sigma^2 + \mu^2)N\\ &= N(N\mu^2 - \mu^2 - \sigma^2)\end{aligned}$$

ทำให้ได้ว่า

$$\begin{aligned}\text{cov}(X_i, X_j) &= \frac{1}{N(N-1)} \cdot N(N\mu^2 - \mu^2 - \sigma^2) - \mu^2\\ &= \frac{1}{N-1} \cdot (N\mu^2 - \mu^2 - \sigma^2 - (N-1)\mu^2)\\ &= -\frac{1}{N-1}\sigma^2\end{aligned}$$

ดังนั้นสำหรับ $i, j = 1, 2, ..., n$ เมื่อ $i \neq j$ จะได้ว่า

1. $E(X_i) = \mu$
2. $\text{var}(X_i) = \sigma^2$
3. $\text{cov}(X_i, X_j) = -\dfrac{1}{N-1}\sigma^2$

ทฤษฎีบท 7.2.2 สุ่มตัวอย่างขนาด n จากประชากรขนาด N ค่าเฉลี่ยเลขคณิตและความแปรปรวนของ $\bar{X}$ คือ $\mu_{\bar{X}}$ และ $\sigma_{\bar{X}}^2$ ตามลำดับคำนวณได้จาก

$$\mu_{\bar{X}} = \mu \qquad \text{และ} \qquad \sigma_{\bar{X}}^2 = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

บทพิสูจน์. ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างขนาด n ซึ่งสุ่มแบบธรรมดาจากประชากรขนาด N

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

และ

$$\begin{aligned}
\sigma_{\bar{X}}^2 &= \text{var}(\bar{X}) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n}\text{var}(X_i) - \sum_{i\neq j}\text{cov}(X_i, X_j)\right] \\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n}\sigma^2 + \sum_{i\neq j}\frac{1}{N-1}\sigma^2\right] = \frac{1}{n^2}\left[n\sigma^2 + \frac{1}{N-1}\sigma^2 \cdot n(n-1)\right] \\
&= \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}
\end{aligned}$$

□

เมื่อขนาดประชากร N มีค่ามาก หรือการสุ่มตัวอย่างโดยการกระทำโดยหยิบแล้วใส่คืนกลับที่เดิม จะได้ค่าเฉลี่ยของค่าเฉลี่ย และความแปรปรวนของค่าเฉลี่ยเท่ากับ

$$\mu_{\bar{X}} = \mu \qquad \text{และ} \qquad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

ในทางปฏิบัติเราจะไม่คำนึงถึงตัวคูณ $\frac{N-n}{N-1}$ เมื่อค่าของ $\frac{n}{N}$ มีค่าไม่เกิน 5% ถ้าสูงเกินไปอาจมีผลต่อการคำนวนค่าประมาณ $\sigma_{\bar{X}}^2$ ผิดไป แต่ถ้าประชากรนั้นมีการแจกแจงปกติ ไม่ว่าจะเลือกตัวอย่างอย่างไรก็ตามถ้าขนาดตัวอย่างมีค่ามาก การแจกแจงของค่าเฉลี่ยของตัวอย่างประมาณได้ด้วยการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ $\frac{\sigma^2}{n}$

ทฤษฎีบท 7.2.3 **ทฤษฎีบทลิมิตสู่ส่วนกลาง** (the Central Limit Theorem)
ถ้า $\bar{X}$ เป็นค่าเฉลี่ยของตัวอย่างสุ่มขนาด n จากประชากรซึ่งมีค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2 แล้วการแจกแจงตัวแปรสุ่ม

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

เมื่อ $n \to \infty$จะประมาณด้วยการแจกแจงปกติมาตรฐาน $n(z; 0, 1)$

บทพิสูจน์. ขอละไว้ในวิชานี้ □

ถ้า $n < 30$ จะประมาณได้ผลดีก็ต่อเมื่อประชากรมีการแจกแจงใกล้เคียงการแจกแจงปกติ แต่ถ้าประชากรแจกแจงปกติอยู่แล้ว การแจกแจงค่าเฉลี่ยเลขคณิตของตัวอย่างจะมีการแจกแจงปกติโดยไม่สนใจขนาดของตัวอย่าง

ตัวอย่าง 7.2.4 โรงงานผลิตหลอดไฟแห่งหนึ่ง ผลิตหลอดไฟซึ่งอายุการใช้งานมีการแจกแจงปกติที่มีค่าเฉลี่ยเลขคณิต 800 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 40 ชั่วโมง สุ่มตัวอย่างของหลอดไฟ 16 หลอด จงหา

1. ความน่าจะเป็นที่หลอดไฟจะมีอายุเฉลี่ยน้อยกว่า 775 ชั่วโมง
2. ความน่าจะเป็นที่หลอดไฟจะมีอายุเฉลี่ยอย่างน้อย 810 ชั่วโมง

แนวคำตอบ อายุหลอดไฟมีการแจกแจงปกติ โดยที่ $\mu = 800$ และ $\sigma = 40$
สุ่มตัวอย่างขนาด $n = 16$ เนื่องจากอายุการใช้งานของหลอดไฟมีการแจกแจงปกติ จะได้ว่า

$$P(\bar{X} < 775) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{775-800}{\frac{40}{\sqrt{16}}}\right) = P(Z < -2.5) = 0.0062$$

ดังนั้นความน่าจะเป็นที่หลอดไฟจะมีอายุเฉลี่ยน้อยกว่า 775 ชั่วโมง เท่ากับ 0.0062

$$P(\bar{X} \geq 810) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{810-800}{\frac{40}{\sqrt{16}}}\right) = P(Z \geq 1) = 0.1587$$

ดังนั้นความน่าจะเป็นที่หลอดไฟจะมีอายุเฉลี่ยอย่างน้อย 510 ชั่วโมง เท่ากับ 0.1587

ตัวอย่าง 7.2.5 จากการสำรวจส่วนสูงของนักศึกษาชายชั้นปีที่ 1 ของมหาวิทยาลัยแห่งหนึ่ง มีส่วนสูงเฉลี่ย 168 เซนติเมตร และส่วนเบี่ยงเบนมาตรฐาน 10 เซนติเมตร ถ้าส่วนสูงของนักศึกษาชายกลุ่มนี้มีการแจกแจงปกติ สุ่มตัวอย่างนักศึกษาชายชั้นปีที่ 1 มา 50 คน จงหาความน่าจะเป็นที่จะได้ส่วนสูงเฉลี่ยของนักศึกษาชาย 50 คนนี้ มีค่ามากกว่า 170 เซนติเมตร
แนวคำตอบ ส่วนสูงของนักศึกษาชายปี 1 มีการแจกแจงปกติ โดยที่ $\mu = 168$ และ $\sigma = 10$
สุ่มตัวอย่างขนาด $n = 50$ เนื่องจากส่วนสูงของนักศึกษาชายกลุ่มนี้มีการแจกแจงปกติจะได้ว่า

$$P(\bar{X} > 170) = P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{170-168}{\frac{10}{\sqrt{50}}}\right) = P(Z > 1.41) = 0.0793$$

ดังนั้นความน่าจะเป็นที่จะได้ส่วนสูงเฉลี่ยของนักศึกษาชาย 50 คนนี้ มีค่ามากกว่า 170 เซนติเมตร เท่ากับ 0.0793

ทฤษฎีบท 7.2.6 การแจกแจงของผลต่างของค่าเฉลี่ยของตัวอย่าง
ถ้าสุ่มตัวอย่างขนาด n_1 และ n_2 โดยอิสระต่อกันจากประชากรสองชุดชนิดต่อเนื่องหรือไม่ต่อเนื่องก็ตาม ซึ่งมีค่าเฉลี่ย μ_1 และ μ_2 ความแปรปรวน σ_1^2 และ σ_2^2 ตามลำดับ แล้วการแจกแจงของผลต่างของค่าเฉลี่ยของตัวอย่าง $\bar{X}_1$ และ $\bar{X}_2$ มีการแจกแจงโดยประมาณเป็นการแจกแจงปกติโดยมี

ค่าเฉลี่ยเลขคณิต $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$ และความแปรปรวน $\sigma^2_{\bar{X}_1-\bar{X}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

และตัวแปรสุ่ม Z คือ

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

เมื่อ n_1 และ n_2 มีค่ามากพอ (มากกว่าหรือเท่ากับ 30) จะมีการแจกแจงประมาณเป็นการแจกแจงปกติมาตรฐาน $n(z;0,1)$

บทพิสูจน์. จะได้ว่า

$$\mu_{\bar{X}_1-\bar{X}_2} = E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$
$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \text{var}(\bar{X}_1 - \bar{X}_2) = \text{var}(\bar{X}_1) + \text{var}(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

โดยทฤษฎีบทลิมิตสู่ส่วนกลางจะได้ผลตามทฤษฎีบทนี้ □

ตัวอย่าง 7.2.7 โรงงานที่ 1 ผลิตหลอดไฟ LED ที่มีอายุเฉลี่ย 6.5 ปี ส่วนเบี่ยงเบนมาตรฐาน 0.9 ปี โรงงานที่ 2 ผลิตหลอดไฟ LED ที่มีอายุเฉลี่ย 6.0 ปี ส่วนเบี่ยงเบนมาตรฐาน 0.8 ปี จงหาความน่าจะเป็นที่ตัวอย่างสุ่มขนาด 36 ซึ่งผลิตโดยโรงงานที่ 1 จะมีอายุเฉลี่ยมากกว่า อายุเฉลี่ยของหลอด LED 49 หลอดซึ่งผลิตโดยโรงงานที่ 2 อย่างน้อย 1 ปี
แนวคำตอบ พิจารณาข้อมูล

	โรงงานที่ 1	โรงงานที่ 2
ค่าเฉลี่ย	$\mu_1 = 6.5$	$\mu_2 = 6.0$
ส่วนเบี่ยงเบนมาตรฐาน	$\sigma_1 = 0.9$	$\sigma_2 = 0.8$
ขนาดตัวอย่าง	$n_1 = 36$	$n_2 = 49$

เนื่องจากขนาดของตัวอย่างมากกว่า 30 จะได้ว่าผลต่างของค่าเฉลี่ยอายุทั้ง 2 โรงงานมีการแจกแจงประมาณเป็นการแจกแจงปกติมาตรฐาน

$$\begin{aligned} P(\bar{X}_1 - \bar{X}_2 > 1) &= P\left(\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} > \frac{1 - (6.5 - 6.0)}{\sqrt{\dfrac{0.9^2}{36} + \dfrac{0.8^2}{49}}}\right) \\ &= P(Z > 2.65) = 1 - P(Z < 2.65) = 1 - 0.9960 = 0.0040 \end{aligned}$$

ดังนั้นความน่าจะเป็นที่ตัวอย่างสุ่มขนาด 36 ซึ่งผลิตโดยโรงงานที่ 1 จะมีอายุเฉลี่ยมากกว่า อายุเฉลี่ยของหลอด LED 49 หลอดซึ่งผลิตโดยโรงงานที่ 2 อย่างน้อย 1 ปี เท่ากับ 0.0040

7.3 การแจกแจงสัดส่วนของตัวอย่าง

จากการทดลองทวินามซึ่งมีความน่าจะเป็นของความสำเร็จในการทดลองแต่ละครั้งเท่ากับ p เลือกตัวอย่างสุ่มขนาด n ให้ X เป็นจำนวนผลสำเร็จที่ได้จากการทดลอง x จะได้การแจกแจงของตัวแปรสุ่ม X ซึ่งเป็นจำนวนครั้งของความสำเร็จของตัวอย่างขนาด n นั้น ตัวแปรสุ่ม X ประมาณได้กับการแจกแจงปกติที่มีค่าเฉลี่ย $\mu = np$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma = \sqrt{npq}$ จากทฤษฎีบท 6.2.7 เมื่อ n มีค่ามากพอ จะได้ว่าผลดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 7.3.1 ถ้าเลือกตัวอย่างขนาด n จากประชากรทวินามซึ่งมีค่าเฉลี่ย $\mu = np$ และความแปรปรวน $\sigma^2 = npq$ **การแจกแจงของสัดส่วน** (proportion distribution) ของความสำเร็จ $\hat{P}$ จะประมาณด้วยการแจกแจงปกติที่มี

$$\text{ค่าเฉลี่ยเลขคณิต} \quad \mu_{\hat{P}} = p \quad \text{และส่วนเบี่ยงเบนมาตรฐาน} \quad \sigma_{\hat{P}} = \sqrt{\frac{pq}{n}}$$

และ $\hat{P} = \dfrac{X}{n}$ โดยที่ตัวแปรสุ่ม Z คือ

$$Z = \frac{\hat{P} - p}{\sqrt{\dfrac{pq}{n}}}$$

เมื่อ n มีค่ามากพอ (ในทางปฏิบัติมากกว่าหรือเท่ากับ 30) จะมีการแจกแจงประมาณเป็นการแจกแจงปกติมาตรฐาน $n(z; 0, 1)$

บทพิสูจน์. ให้ $X \sim \text{Bin}(n, p)$ โดยทฤษฎีบท 6.2.7 ให้ X มีค่ามากพอจะประมาณตัวแปรสุ่ม เป็น $\text{N}(\mu = np, \sigma^2 = npq)$ ให้ $\hat{P} = \dfrac{X}{n}$ จะได้ว่า

$$\mu_{\hat{P}} = E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n} \cdot np = p$$

$$\sigma^2_{\hat{P}} = \text{var}(\hat{P}) = \text{var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\text{var}(X) = \frac{1}{n^2} \cdot npq = \frac{pq}{n}$$

□

เนื่องจากการแจกแจงสัดส่วนมาจากการแจกแจงทวินามซึ่งเป็นการแจกแจงของตัวแปรสุ่มไม่ต่อเนื่อง ถ้า $X \sim \text{Bin}(n, p)$ ซึ่งเป็นตัวแปรสุ่มไม่ต่อเนื่อง แต่เมื่อ n มีค่ามากพอ จะมีการแจกแจงของตัวแปรสุ่ม $\text{N}(0, 1)$ ซึ่งเป็นตัวแปรสุ่มต่อเนื่อง โดยนำทฤษฎีบท 6.2.7 ไปใช้ในการประมาณค่าการแจกแจงทวินาม เมื่อ n มีค่ามากพอ กำหนดให้ $\hat{p}_1 = \frac{x_1}{n}$ และ $\hat{p}_2 = \frac{x_2}{n}$ จะได้ว่า

$\text{Bin}(n, p)$	$\text{N}(\mu = np, \sigma^2 = npq)$	
$P(x_1 < X < x_2)$	$P(x_1 - 0.5 < X < x_2 + 0.5)$	$P(\hat{p}_1 - \frac{0.5}{n} < \hat{P} < \hat{p}_2 + \frac{0.5}{n})$
$P(X > x_1)$	$P(X > x_1 + 0.5)$	$P(\hat{P} > \hat{p}_1 + \frac{0.5}{n})$
$P(X < x_1)$	$P(X < x_1 - 0.5)$	$P(\hat{P} < \hat{p}_1 - \frac{0.5}{n})$

ถ้า n มีค่ามากพอ $p \to \frac{1}{2}$ ทำให้เข้าใกล้การแจกแจงปกติ ถึงแต่แม้ p จะไม่เข้าใกล้ $\frac{1}{2}$ การประมาณค่าจะได้ผลดีเช่นกันเมื่อ $n > 30$ และ $np > 5$ และ $nq > 5$

ตัวอย่าง 7.3.2 ถ้าสุ่มเด็กเกิดใหม่ 100 คน จงหาความน่าจะเป็นที่เกิดใหม่จะเป็นชายตั้งแต่ 53% ถึง 62% ถ้าความน่าจะเป็นที่เด็กเป็นเพศชายหรือเพศหญิงเท่า ๆ กัน

แนวคำตอบ ให้ X เป็นจำนวนเด็กชายที่เกิดใหม่ ซึ่งมี $p = \frac{1}{2} = 0.5$ นั่นคือ $X \sim \text{Bin}(n = 100, p = 0.5)$ จะเห็นว่า $n > 30$, $np = nq = 15 > 5$ ดังนั้นการประมาณค่าการแจกแจงสัดส่วนด้วยการแจกแจงปกติ จะได้ว่า

$$
\begin{aligned}
P(0.53 < \hat{P} < 0.62) &\approx P\left(0.53 - \frac{0.5}{100} < \hat{P} < 0.62 + \frac{0.5}{100}\right) \\
&= P(0.525 < \hat{P} < 0.625) \\
&= P\left(\frac{0.525 - 0.5}{\sqrt{\frac{0.5(0.5)}{100}}} < \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} < \frac{0.625 + 0.5}{\sqrt{\frac{0.5(0.5)}{100}}}\right) \\
&= P(0.5 < Z < 2.5) = 0.3023
\end{aligned}
$$

ดังนั้นความน่าจะเป็นที่เกิดใหม่จะเป็นชายตั้งแต่ 53% ถึง 62% เท่ากับ 0.3023

ตัวอย่าง 7.3.3 บริษัทแห่งหนึ่งผลิตแบตเตอรี่โดยใช้เครื่องจักร โดยแบตเตอรี่ที่ผลิตโดยเครื่องจักรนี้ชำรุด 10% ถ้าสุ่มตัวอย่างแบตเตอรี่ของบริษัทแห่งนี้มา 500 อัน จงประมาณค่าความน่าจะเป็นที่สัดส่วนของแบตเตอรี่ตัวอย่างจะชำรุดมากกว่า 11%

แนวคำตอบ ความน่าจะเป็นที่แบตเตอรี่ที่จะชำรุดคือ $p = 0.1$ จะเห็นว่า $n > 30$, $np = 50 > 5$ และ $nq = 450 > 5$ ดังนั้นการประมาณค่าการแจกแจงสัดส่วนด้วยการแจกแจงปกติ จะได้ว่า

$$
\begin{aligned}
P(\hat{P} > 0.11) &\approx P\left(\hat{P} > 0.11 + \frac{0.5}{500}\right) = P(\hat{P} > 0.111) \\
&= P\left(\frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} > \frac{0.111 - 0.1}{\sqrt{\frac{0.1(0.9)}{500}}}\right) \\
&= P(Z > 0.82) = 0.2061
\end{aligned}
$$

ดังนั้นความน่าจะเป็นที่สัดส่วนของแบตเตอรี่ตัวอย่างจะชำรุดมากกว่า 11% เท่ากับ 0.2061

ทฤษฎีบท 7.3.4 ถ้าเลือกตัวอย่าง 2 ชุดขนาด n_1 และ n_2 โดยอิสระต่อกันจากประชากรทวินามซึ่งมีค่าเฉลี่ย $\mu_1 = n_1p_1$ และ $\mu_2 = n_2p_2$ และความแปรปรวน $\sigma_1^2 = n_1p_1q_1$ และ $\sigma_2^2 = n_2p_2q_2$ **การแจกแจงของผลต่างสัดส่วน** (different proportion distribution) ของความสำเร็จ $\hat{P}_1 - \hat{P}_2$ จะประมาณด้วยการแจกแจงปกติที่มี

ค่าเฉลี่ยเลขคณิต $\mu_{\hat{P}_1-\hat{P}_2} = p_1 - p_2$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\hat{P}_1-\hat{P}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$

และตัวแปรสุ่ม Z คือ

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$$

เมื่อ n_1 และ n_2 มีค่ามากพอ (ในทางปฏิบัติมากกว่าหรือเท่ากับ 30) จะมีการแจกแจงประมาณเป็นการแจกแจงปกติมาตรฐาน $n(z; 0, 1)$

บทพิสูจน์. เป็นแบบฝึกหัด □

ตัวอย่าง 7.3.5 สุ่มตัวอย่างมา 2 ชุด จากผู้มีสิทธิ์เลือกตั้งในเขตหนึ่ง ชุดละ 200 คน จงหาความน่าจะเป็นที่ผลต่างของสัดส่วนที่ออกเสียงคนหนึ่งมากกว่า 10% ถ้าผลจากการลงคะแนนในเขตนั้นปรากฏว่าผู้สมัครคนนี้ได้คะแนนเสียง 65%
แนวคำตอบ พิจารณาข้อมูล

	ตัวอย่างที่ 1	ตัวอย่างที่ 2
สัดส่วน	$p_1 = 0.65$	$p_2 = 0.65$
ขนาดตัวอย่าง	$n_1 = 200$	$n_2 = 200$

เนื่องจากขนาดของตัวอย่างมากกว่า 30 จะได้ว่าผลต่างของสัดส่วนของ 2 กลุ่มตัวอย่างเป็นการแจกแจงปกติมาตรฐาน จะได้ว่า

$$\begin{aligned}
P(|\hat{P}_1 - \hat{P}_2| > 0.1) &= 1 - P(|\hat{P}_1 - \hat{P}_2| < 0.1) \\
&= 1 - P(-0.1 < \hat{P}_1 - \hat{P}_2 < 0.1) \\
&= 1 - P\left(\frac{-0.1 - (0.65 - 0.65)}{\sqrt{\frac{0.65(0.35)}{200} + \frac{0.65(0.35)}{200}}} < Z < \frac{0.1 - (0.65 - 0.65)}{\sqrt{\frac{0.65(0.35)}{200} + \frac{0.65(0.35)}{200}}}\right) \\
&= 1 - P(-2.10 < Z < 2.10) \\
&= 1 - 0.9643 \\
&= 0.0357
\end{aligned}$$

ดังนั้นความน่าจะเป็นที่ผลต่างของสัดส่วนที่ออกเสียงคนหนึ่งมากกว่า 10% เท่ากับ 0.0357

แบบฝึกหัด 7.1 - 7.3

1. ส่วนสูงของนักศึกษา 5 คน ประกอบด้วย 155, 167, 168, 169, 175 จงหา ค่าเฉลี่ย ความแปรปรวน และสัดส่วนของคนที่มีส่วนสูงมากกว่า 165 ของตัวอย่างทุกชุด ที่เกิดขึ้นจากการสุ่มตัวอย่างขนาด 4 จากประชากรดังกล่าว พร้อมหาค่าทั้ง 3 ของประชากร

2. จงหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวอย่างต่อไปนี้ 10, 12, 14, 15, 17, 12, 11

3. ประชากรที่มีจำนวนแน่นอนคือ 2, 3, 5, 7, 11 เลือกตัวอย่างขนาด 3 มาได้ทั้งหมดกี่แบบ ประกอบชุดใดบ้าง พร้อมทั้งหาค่าเฉลี่ยเลขคณิตและความแปรปรวนของทุกชุด

4. จงแสดงว่าค่าความแปรปรวนของตัวอย่างไม่เปลี่ยนแปลง ถ้านำค่าคงที่ไปบวกทุก ๆ ค่าสังเกตในตัวอย่าง

5. ลูกปืน 500 ลูก มีน้ำหนักเฉลี่ย 5.02 ออนซ์ และส่วนเบี่ยงเบนมาตรฐาน 0.30 ออนซ์ จงหาความน่าจะเป็นที่จะสุ่มตัวอย่างลูกปืน 100 ลูก ได้น้ำหนักรวมกัน

   5.1 ระหว่าง 496 ถึง 500 ออนซ์

   5.2 น้อยกว่า 510 ออนซ์

6. สุ่มตัวอย่างขนาด 25 จากประชากรปกติ จงหาความน่าจะเป็นที่ $|\bar{X} - \mu| > \dfrac{\sigma}{2}$

7. สุ่มตัวอย่างขนาด 30 จากประชากรชุดที่หนึ่งซึ่งมีส่วนเบี่ยงเบนมารฐาน 0.5 สุ่มตัวอย่างขนาด 30 จากประชากรชุดที่สองซึ่งมีส่วนเบี่ยงเบนมารฐาน 0.4 จากการคำนวณพบว่า 5% ของค่าเฉลี่ยเลขคณิตของตัวอย่างชุดที่หนึ่ง มีค่ามากกว่าค่าเฉลี่ยเลขคณิตของตัวอย่างชุดที่สองอย่างน้อย 0.6 จงหาความแตกต่างของค่าเฉลี่ยเลขคณิตของประชากรชุดที่หนึ่งและสอง

8. ถ้าขนาดตัวอย่าง 36 จะได้ค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยตัวอย่างเป็น 2 อยากทราบว่าจะต้องใช้ตัวอย่างขนาดเท่าใดเพื่อที่จะทำให้ค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยตัวอย่าง ลดลงเป็น 1.2

9. ตัวอย่างขนาด 25 สุ่มจากประชากรปกติอีกพวกหนึ่งซึ่งมีค่าเฉลี่ยเลขคณิต 80 และส่วนเบี่ยงเบนมาตรฐาน 5 ตัวอย่างอีกชุดหนึ่งขนาด 36 สุ่มจากประชากรปกติอีกพวกหนึ่งซึ่งมีค่าเฉลี่ยเลขคณิต 75 และส่วนเบี่ยงเบนมาตรฐาน 3 จงหาความน่าจะเป็นที่ค่าเฉลี่ยเลขคณิตของตัวอย่างขนาด 25 จะมากกว่าค่าเฉลี่ยของตัวอย่างขนาด 36 อยู่อย่างน้อย 3.4 แต่ไม่เกิน 5.9

10. ในการสอบ ONET วิชาคณิตศาสตร์ของนักเรียนชั้น ม. 6 ปีการศึกษาหนึ่ง มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 คะแนน ถ้าสุ่มตัวอย่างนักเรียนมา 36 คน จงหาโอกาสที่คะแนนเฉลี่ยของตัวอย่างจะต่างจากคะเฉลี่ยประชากรมากกว่า 6 คะแนน

11. จงหาความน่าจะเป็นในการโยนเหรียญเที่ตรง 120 ครั้ง แล้วได้หัว

    11.1 ตั้งแต่ 40% ถึง 60%

    11.2 มากกว่าหรือเท่ากับ 5 ใน 8 ของจำนวนครั้งที่โยน

12. จากการตรวจสอบผลิตภัณฑ์ชนิดหนึ่งปรากฏว่ามีที่บกพร่อง 2% จงหาความน่าจะเป็นที่ผลิตภัณฑ์นี้ 400 ชิ้น จะมีส่วนบกพร่อง

    12.1 3% หรือมากกว่า

    12.2 2% หรือน้อยกว่า

13. จากผลการเลือกตั้งคณะกรรมการนักเรียน นาย ก เป็นผู้รับสมัครได้คะแนน 46% สุ่มตัวอย่างจากนักเรียนที่ไปเลือกตั้ง

    13.1 จากตัวอย่างขนาด 50 คน จงหาความน่าจะเป็นที่นาย ก ได้รับคะแนนเกินครึ่ง

    13.2 จากตัวอย่างขนาด 100 คน จงหาความน่าจะเป็นที่นาย ก ได้รับคะแนนเกินครึ่ง

14. ร้านสะดวกซื้อประมาณการว่าลูกค้าที่เข้ามาในร้านจะซื้อสินค้าประมาณ ร้อยละ 40 สุ่มลูกค้ามา 100 คน จงหาความน่าจะเป็นที่ลูกค้าซื้อสินค้าไม่น้อยกว่าร้อยละ 30

15. จากผลการเลือกตั้งสมาชิกสภาผู้แทนราษฎรเขตหนึ่ง สมมติว่าเปอร์เซนต์ของชายและหญิงที่ชอบพรรค A เป็น 30% และ 40% ตามลำดับ ถ้าสัมภาษณ์ชาย 500 คน และหญิง 500 คน ที่สุ่มตัวอย่างมา จงหาความน่าจะเป็นที่ชายจะชอบพรรค A มากกว่าหญิง 10%

16. ระบบสายส่งไฟฟ้าประกอบด้วยจุดบำรุงรักษา 1500 จุด แต่ละจุดมีค่าใช้จ่ายด้านการ บำรุงรักษาเฉลี่ยทั้งระบบเท่ากับ 1000 บาท และมีส่วนเบี่ยงเบนมาตรฐานของค่าใช้จ่ายด้านการบำรุงรักษาเท่ากับ 120 บาท หากสุ่มเลือกจุดบำรุงรักษามา 80 จุด จงหาความน่าจะเป็นที่ค่าบำรุงรักษาเฉลี่ยของทั้ง 80 จุด จะมีค่าระหว่าง 975 บาท ถึง 1025 บาท

17. มหาวิทยาลัยแห่งหนึ่งทำการสำรวจพบว่าสัดส่วนของบัณฑิตรุ่นที่หนึ่งที่ได้ทำงานตรงสายวิชาชีพเท่ากับ 0.68 สัดส่วนของบัณฑิตรุ่นที่สองได้ทำงานตรงสายวิชาชีพเท่ากับ 0.74 หากสุ่มตัวอย่างชื่อบัณฑิตมา รุ่นละ 100 คน จงหาความน่าจะเป็นที่สัดส่วนของตัวอย่างของบัณฑิตรุ่นที่หนึ่งได้ทำงานตรงสายวิชาชีพ น้อยกว่าตัวอย่างของบัณฑิตรุ่นที่สองมากกว่าหรือเท่ากับ 10%

18. ฝรั่ง 100 กรัม เทียบเท่าวิตามินซี 228 กรัม ถ้าผู้ที่รับประทานฝรั่ง 100 กรัม เป็นประจำ มีความน่าจะเป็นที่จะเป็นหวัด 0.10 ส่วนกลุ่มที่ไม่รับประทานฝรั่งวันละ 100 กรัม เป็น ประจำ มีโอกาสเป็นหวัด 0.20 ถ้าเลือกตัวอย่างสุ่มมากลุ่มละ 40 คน จงหาความน่าจะเป็นของความแตกต่างของสัดส่วนตัวอย่างของผู้ที่เป็นหวัดทั้งสองกลุ่มว่าจะแตกต่างกันไม่เกิน 0.05

7.4 การแจกแจงค่าสถิติอื่น ๆ

ในกรณีที่**ไม่ทราบความแปรปรวนของประชากร** เราจะประมาณความแปรปรวนของประชากรด้วยความแปรปรวนของตัวอย่าง

1. ถ้า $n \geq 30$ จะประมาณด้วยการแจกแจงปกติซึ่ง $Z = \dfrac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}}$

2. ถ้า $n < 30$ จะประมาณด้วยการแจกทีซึ่ง $T = \dfrac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}}$ และมีองศาเสรี $\nu = n - 1$

ตัวอย่าง 7.4.1 โรงงานผลิตหลอดไฟแห่งหนึ่งผลิตหลอดไฟซึ่งอายุการใช้งานมีการแจกแจงปกติ โดยกล่าวอ้างว่าหลอดไฟที่ผลิตใช้ได้นานโดยเฉลี่ย 10000 ชั่วโมง เพื่อรักษาค่าเฉลี่ยไว้ตามที่กล่าวอ้างจึงสุ่มตัวอย่างหลอดไฟมาทดสอบจากตัวอย่างสุ่ม 16 หลอด ได้ส่วนเบี่ยงเบนมาตรฐาน 400 ชั่วโมง จงหา k ที่ทำให้ $P(|\bar{X} - \mu| < k) = 0.95$ เมื่อ μ เป็นค่าเฉลี่ยของหลอดไฟทั้งหมด และ $\bar{X}$ เป็นค่าเฉลี่ยของหลอดไฟตัวอย่าง

แนวคำตอบ ประชากรมีการแจกแจงปกติไม่ทราบค่า σ^2 และตัวอย่างมีขนาดเล็ก $n = 16$ ประมาณตัวอย่างด้วยการแจกแจงที ซึ่งมี $\nu = 16 - 1 = 15$ โดยมี $\mu = 10000$ และ $S = 400$ จะได้ว่า

$$
\begin{aligned}
0.95 &= P(|\bar{X} - \mu| < k) = P\left(\left|\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}\right| < \frac{k}{\frac{400}{\sqrt{16}}}\right) \\
0.95 &= P\left(|T| < \frac{k}{100}\right) = 1 - P\left(|T| > \frac{k}{100}\right) \\
0.95 &= 1 - 2P\left(T > \frac{k}{100}\right) \\
P\left(T > \frac{k}{100}\right) &= 0.025
\end{aligned}
$$

จากตารางทีจะได้ว่า $P(T > 2.131) = 0.025$ ฉะนั้น

$$
\begin{aligned}
\frac{k}{100} &= 2.131 \\
k &= 213.1
\end{aligned}
$$

ดังนั้น $k = 213.1$ ชั่วโมง

สุ่มตัวอย่าง 2 ชุดขนาด n_1 และ n_2 ซึ่งอิสระต่อกัน เมื่อ**ไม่ทราบค่า**ของ σ_1^2 และ σ_2^2

1. กรณี n_1 และ n_2 มากกว่าหรือเท่ากับ 30
   จะประมาณด้วยการแจกแจงปกติ และประมาณค่า σ_1^2 และ σ_2^2 ด้วย S_1^2 และ S_2^2 ตามลำดับ โดยที่
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

2. กรณี n_1 หรือ n_2 น้อยกว่า 30

   • สำหรับ $\sigma_1^2 = \sigma_2^2$ จะประมาณด้วยการแจกแจงที โดยที่
$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$
   เมื่อ
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$
   และองศาเสรีคือ $\nu = n_1 + n_2 - 2$

   • สำหรับ $\sigma_1^2 \neq \sigma_2^2$ จะประมาณด้วยการแจกแจงที โดยที่
$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$
   เมื่อองศาเสรีคือ
$$\nu = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\left(\dfrac{S_1^2}{n_1}\right)^2 \dfrac{1}{n_1 - 1} + \left(\dfrac{S_2^2}{n_2}\right)^2 \dfrac{1}{n_2 - 1}}$$

ตัวอย่าง 7.4.2 โรงงานผลิตหลอดไฟ 2 ชนิด หลอดไฟชนิดที่ 1 มีอายุเฉลี่ย 520 ชั่วโมง หลอดไฟที่ 2 มีอายุเฉลี่ย 420 ชั่วโมง ทุก ๆ เดือนโรงงานจะสุ่มตัวอย่างหลอดไฟชนิดที่ 1 จำนวน 16 หลอด ซึ่งมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 40 ชั่วโมง และสุ่มตัวอย่างหลอดไฟชนิดที่ 2 จำนวน 9 หลอด ซึ่งมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 35 ชั่วโมง จงหาความน่าจะเป็นที่ค่าเฉลี่ยของตัวอย่างจากหลอดไฟชนิดที่ 1 มากกว่าค่าเฉลี่ยของตัวอย่างจากหลอดไฟชนิดที่ 2 อย่างน้อย 133 ชั่วโมง

1. กรณีความแปรปรวนของหลอดไฟทั้ง 2 ชนิดมีค่าเท่ากัน
2. กรณีความแปรปรวนของหลอดไฟทั้ง 2 ชนิดมีค่าไม่เท่ากัน

แนวคำตอบ พิจารณาข้อมูล

	หลอดไฟชนิดที่ 1	หลอดไฟชนิดที่ 2
ค่าเฉลี่ย	$\mu_1 = 520$	$\mu_2 = 420$
ส่วนเบี่ยงเบนมาตรฐาน(ตัวอย่าง)	$S_1 = 40$	$S_2 = 35$
ขนาดตัวอย่าง	$n_1 = 16$	$n_2 = 9$

1. กรณีความแปรปรวนของหลอดไฟทั้ง 2 ชนิดมีค่าเท่ากัน

   เนื่องจาก $\sigma_1^2 = \sigma_2^2$ จะได้ว่า

$$S_p^2 = \frac{(16-1)40^2 + (9-1)35^2}{16+9-2} = 1469.56$$

   และองศาเสรีคือ $\nu = 16 + 9 - 2 = 23$ ฉะนั้น

$$\begin{aligned} P(\bar{X}_1 - \bar{X}_2 \geq 133) &= P\left(\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \geq \frac{133 - (520 - 420)}{\sqrt{1469.56\left(\frac{1}{16} + \frac{1}{9}\right)}}\right) \\ &= P(T \geq 2.07) = 0.0249 \quad \text{(อ่านค่าจากแอฟพลิเคชั่น)} \end{aligned}$$

   ดังนั้นความน่าจะเป็นที่ค่าเฉลี่ยของตัวอย่างจากหลอดไฟชนิดที่ 1 มากกว่าค่าเฉลี่ยของตัวอย่างจากหลอดไฟชนิดที่ 2 อย่างน้อย 133 ชั่วโมง เท่ากับ 0.0249

2. กรณีความแปรปรวนของหลอดไฟทั้ง 2 ชนิดมีค่าไม่เท่ากัน

   เนื่องจาก $\sigma_1^2 \neq \sigma_2^2$ มีองศาเสรีคือ

$$\nu = \frac{\left(\frac{40^2}{16} + \frac{35^2}{9}\right)^2}{\left(\frac{40^2}{16}\right)^2 \cdot \frac{1}{16-1} + \left(\frac{35^2}{9}\right)^2 \cdot \frac{1}{9-1}} = 18.69 \approx 19$$

   จะได้ว่า

$$\begin{aligned} P(\bar{X}_1 - \bar{X}_2 \geq 133) &= P\left(\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \geq \frac{133 - (520 - 420)}{\sqrt{\frac{40^2}{16} + \frac{35^2}{9}}}\right) \\ &= P(T \geq 2.15) = 0.0224 \quad \text{(อ่านค่าจากแอฟพลิเคชั่น)} \end{aligned}$$

   ดังนั้นความน่าจะเป็นที่ค่าเฉลี่ยของตัวอย่างจากหลอดไฟชนิดที่ 1 มากกว่าค่าเฉลี่ยของตัวอย่างจากหลอดไฟชนิดที่ 2 อย่างน้อย 133 ชั่วโมง เท่ากับ 0.0224

การทำงานทางสถิติเกี่ยวกับข้อมูลประชากรปกติ 2 ชุดที่เปรียบเทียบกันได้ (ไม่อิสระต่อกัน) สามารถเก็บข้อมูลเป็นคู่ ๆ

ลำดับที่	ประชากรชุดที่ 1	ประชากรชุดที่ 2	ผลต่างของประชากรสองชุด
1	x_1	y_1	d_1
2	x_1	y_1	d_2
3	x_1	y_1	d_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	d_n

ให้ μ_D คือค่าเฉลี่ยเลขคณิตของผลต่าง และ σ_D^2 คือค่าส่วนเบี่ยงเบนมาตรฐานของผลต่าง

เมื่อขนาดตัวอย่าง n มีขนาดไม่มาก และไม่ทราบความแปรปรวนประชากร σ_D^2 อาจประมาณด้วยความแปรปรวนตัวอย่าง S_d^2 และจะประมาณด้วยการแจกแจงทีซึ่ง

$$T = \frac{\bar{D} - \mu_D}{\frac{S_d}{\sqrt{n}}}$$

โดยมีองศาอิสระคือ $\nu = n - 1$

ตัวอย่าง 7.4.3 ในการทดสอบความรวดเร็วในการให้บริการแก่ผู้มาติดต่อฝ่ายประชาสัมพันธ์ของ 2 มหาวิทยาลัย โดยให้อาสาสมัคร 7 คน ไปติดต่อสอบถามเรื่องเดียวกันที่ฝ่ายประชาสัมพันธ์ทั้ง 2 แห่งแล้วบันทึกเวลาที่ใช้บริการพบว่ามีความแปรปรวน 25 นาที^{2} สมมติว่าข้อมูลประชากรมีการแจกแจงปกติ จงหาความน่าจะเป็นที่ค่าเฉลี่ยของผลต่างของเวลาจากตัวอย่างคนที่ใช้บริการจาก 2 แหล่งเกิน 3 นาที

แนวคำตอบ ตัวอย่างเป็นข้อมูลเป็นคู่ ๆ จากประชากรปกติ โดยที่ $n = 7$ และ $S_d^2 = 25$ ประมาณด้วยการแจกแจงทีซึ่งมีองศาเสรี $\nu = 7 - 1 = 6$ โดยที่ $\mu_D = 0$ จะได้ว่า

$$\begin{aligned}
P(|\bar{D}| > 3) &= P\left(\left|\frac{\bar{D} - \mu_D}{\frac{S_d}{\sqrt{n}}}\right| > \frac{3-0}{\frac{5}{\sqrt{7}}}\right) \\
&= P(|T| > 1.59) \\
&= 2P(T > 1.59) \\
&= 2(0.0815) \qquad \text{(อ่านค่าจากแอฟพลิเคชั่น)} \\
&= 0.1630
\end{aligned}$$

ดังนั้นความน่าจะเป็นที่ผลต่างของเวลาจากตัวอย่างคนที่ใช้บริการจาก 2 แหล่งเกิน 3 นาที เท่ากับ 0.1630

ทฤษฎีบท 7.4.4 ถ้า S^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n จากประชากรปกติที่มีความแปรปรวน σ^2 จะได้ตัวแปรสุ่ม

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

มีการแจกแจงไคสแควร์มีองศาเสรี $\nu = n-1$

บทพิสูจน์. ให้ประชากรมีการแจกแจงปกติโดยมีค่าเฉลี่ยเลขคณิต μ และความแปรปรวน σ^2 สุ่มตัวอย่างขนาด n ประกอบด้วย $X_1, X_2, ..., X_n$ ได้ความแปรปรวน S^2 พิจารณา

$$\begin{aligned}
\sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_{i=1}^{n}[(X_i-\bar{X})+(\bar{X}-\mu)]^2 \\
&= \sum_{i=1}^{n}[(X_i-\bar{X})^2+2(X_i-\bar{X})(\bar{X}-\mu)+(\bar{X}-\mu)^2] \\
&= \sum_{i=1}^{n}(X_i-\bar{X})^2+2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X})+\sum_{i=1}^{n}(\bar{X}-\mu)^2 \\
&= (n-1)S^2+2(\bar{X}-\mu)\cdot 0+n(\bar{X}-\mu)^2 \\
&= (n-1)S^2+n(\bar{X}-\mu)^2
\end{aligned}$$

ดังนั้น

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

เป็นการแจกแจงไคสแควร์ที่มีองศาเสรี $\nu = n-1$ □

ตัวอย่าง 7.4.5 ตัวอย่างสุ่มขนาด 25 จากประชากรปกติ ซึ่งมีความแปรปรวนของประชากรเท่ากับ 6 จงหาความน่าจะเป็นที่ความแปรปรวนของตัวอย่างจะมีค่ามากกว่า 9.104

แนวคำตอบ สุ่มตัวอย่างขนาด $n = 25$ จากประชากรปกติ และ $\sigma^2 = 6$ เมื่อ $\nu = 25-1 = 24$ จะได้ว่า

$$\begin{aligned}
P(S^2 > 9.104) &= P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(25-1)9.104}{6}\right) \\
&= P(\chi^2 > 36.416) = 0.05 \qquad \text{(อ่านค่าจากตารางไคสแควร์)}
\end{aligned}$$

ดังนั้นความน่าจะเป็นที่ความแปรปรวนของตัวอย่างจะมีค่ามากกว่า 9.104 เท่ากับ 0.05

ตัวอย่าง 7.4.6 สุ่มนักศึกษาชั้นปี 1 จำนวน 15 เพื่อสอบถามคะแนน GPA ในเทอมนี้ ถ้าประชากรมีการแจกแจงปกติ ซึ่งมีความแปรปรวนของประชากรเท่ากับ 0.16 จงหาความน่าจะเป็นที่ความแปรปรวนของตัวอย่างจะอยู่ระหว่าง 0.09 ถึง 0.25

แนวคำตอบ สุ่มตัวอย่างขนาด $n = 15$ จากประชากรปกติ และ $\sigma^2 = 16$ เมื่อ $\nu = 15-1 = 14$ จะได้ว่า

$$\begin{aligned}
P(0.09 < S^2 < 0.25) &= P\left(\frac{(15-1)0.09}{0.16} < \frac{(n-1)S^2}{\sigma^2} < \frac{(15-1)0.25}{0.16}\right) \\
&= P(7.875 < \chi^2 < 21.875) \\
&= P(\chi^2 > 7.875) - P(\chi^2 > 21.875) \\
&= 0.8957 - 0.0812 = 0.8145 \qquad \text{(อ่านค่าจากแอฟพลิเคชัน)}
\end{aligned}$$

ดังนั้นความน่าจะเป็นที่ความแปรปรวนของตัวอย่างจะอยู่ระหว่าง 0.09 ถึง 0.25 เท่ากับ 0.8145

ทฤษฎีบท 7.4.7 ถ้า S_1^2 และ S_2^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n_1 และ n_2 จากประชากรปกติที่มีความแปรปรวน σ_1^2 และ σ_2^2 ตามลำดับ จะได้ตัวแปรสุ่ม

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

มีการแจกแจงเอฟ ที่มีองศาเสรี $\nu_1 = n_1 - 1$ และ $\nu_2 = n_2 - 1$

บทพิสูจน์. สมมติสุ่มตัวอย่าง 2 ชุดโดยอิสระจากกัน ขนาด n_1 และ n_2 จากประชากรปกติที่มี σ_1^2 และ σ_2^2 ตามลำดับ จะได้ตัวแปรสุ่มไคสแควร์ 2 ชุดที่อิสระต่อกันคือ

$$\chi_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \quad \text{และ} \quad \chi_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2}$$

โดยมีองศาเสรีคือ $\nu_1 = n_1 - 1$ และ $\nu_2 = n_2 - 1$ ดังนั้น

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

โดยมีองศาเสรี $\nu_1 = n_1 - 1$ และ $\nu_2 = n_2 - 1$ □

ตัวอย่าง 7.4.8 ให้ S_1^2 และ S_2^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด 25 และ 31 โดยสุ่มอย่างอิสระต่อกันจากประชากรปกติ ที่มีความแปรปรวนเป็น $\sigma_1^2 = 10$ และ $\sigma_2^2 = 15$ จงหาค่าของ

$$P\left(\frac{S_1^2}{S_2^2} > 1.26\right)$$

แนวคำตอบ สุ่มตัวอย่างขนาด $n_1 = 25, n_2 = 31$ จากประชากรปกติ และ $\sigma_1^2 = 10, \sigma_2^2 = 15$ เมื่อ $\nu_1 = 25 - 1 = 24$ และ $\nu_2 = 31 - 1 = 30$ จะได้ว่า

$$\begin{aligned} P\left(\frac{S_1^2}{S_2^2} > 1.26\right) &= P\left(\frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} > 1.26 \cdot \frac{15}{10}\right) \\ &= P(F > 1.89) = 0.05 \qquad \text{(อ่านค่าจากตารางเอฟ)} \end{aligned}$$

ตัวอย่าง 7.4.9 จงหาความน่าจะเป็นที่ความแปรปรวนของตัวอย่างน้ำหนักแรกเกิดเพศชายน้อยกว่าเพศหญิง เมื่อสุ่มตัวอย่างเด็กชาย 10 คน และเด็กหญิง 12 สมมติประชากรทั้งสองมีการแจกแจงปกติและมีความแปรปรวน 2500 และ 1600 กรัม² ตามลำดับ

แนวคำตอบ พิจารณาข้อมูล

	เด็กแรกเกิดเพศชาย	เด็กแรกเกิดเพศหญิง
ความแปรปรวน	$\sigma_1^2 = 2500$	$\sigma_2^2 = 1600$
ขนาดตัวอย่าง	$n_1 = 10$	$n_2 = 12$
องศาเสรี	$\nu_1 = 9$	$\nu_2 = 11$

$$P(S_1^2 < S_2^2) = P\left(\frac{S_1^2}{S_2^2} < 1\right) = P\left(\frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} > \cdot \frac{1600}{2500}\right) = P(F < 0.64) = 0.2560$$

ดังนั้นความน่าจะเป็นที่ความแปรปรวนของตัวอย่างน้ำหนักแรกเกิดเพศชายน้อยกว่าเพศหญิงเท่ากับ 0.2560

แบบฝึกหัด 7.4

1. น้ำหนักของกาแฟยี่ห้อ ก มีการแจกแจงปกติด้วยค่าเฉลี่ย 102 กรัมต่อขวด ถ้าสุ่มตัวอย่างมาก 25 ขวด มีส่วนเบี่ยงเบนมาตรฐาน 7 กรัมต่อขวด จงหาความน่าจะเป็นที่นำหนักเฉลี่ยของตัวอย่างจากกมากกว่า 110 กรัมต่อขวด

2. ผู้จัดการโรงงานซึ่งผลิตสินค้า 2 แบบ ต้องการเปรียบเทียบประสิทธิภาพของคนงานที่ผลิตสินค้าทั้ง 2 แบบ จึงสุ่มตัวอย่างคนงานที่ผลิตสินค้าชนิดที่ 1 มา 6 คน และ คนงานที่ผลิตสินค้าชนิดที่ 2 มา 8 คน แล้วบันทึกเวลาทำงานในหน่วยนาทีแล้วคำนวณค่าความแปรปรวนได้เป็น 5.55 นาที และ 1.81 นาทีตามลำดับ สมมติว่าเวลาที่ผลิตสินค้าทั้ง 2 แบบ มีการแจกแจงปกติโดยค่าเฉลี่ยเท่ากัน จงหาโอกาสที่คนงานตัวอย่างซึ่งผลิตสินค้าชนิดที่ 2 จากใช้เวลามากกว่าชนิดที่ 1 อย่างน้อย 2 นาที

   2.1 เมื่อความแปรปรวนของประชากรทั้งสองเท่ากัน

   2.2 เมื่อความแปรปรวนของประชากรทั้งสองไม่เท่ากัน

3. บริษัทผลิตทุเรียนทอดบรรจุกระป๋องแห่งหนึ่ง พบว่าน้ำหนักสุทธิของทุเรียนมีการ แจกแจงปกติ มีความแปรปรวน 7 กรัม^2 ถ้าสุ่มทุเรียนทอดมา 15 กระป๋อง จงหาความน่าจะเป็นที่ความแปรปรวนของทุเรียนมากกว่า 9 กรัม^2

4. ตัวอย่างสุ่มขนาด 36 จากประชากรปกติ ซึ่งมีความแปรปรวนของประชากรเท่ากับ 5 จงหาความน่าจะเป็นที่ความแปรปรวนของตัวอย่างจะมีค่า

   4.1 น้อยกว่า 4.52

   4.2 ระหว่าง 4 และ 6

5. ให้ S_1^2 และ S_2^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด 8 และ 12 โดยสุ่มอย่างอิสระต่อกันจากประชากรปกติ ที่มีความแปรปรวนเท่ากัน จงหาค่าของ

   5.1 $P\left(\frac{S_1^2}{S_2^2} < 5\right)$

   5.2 $P\left(\frac{S_2^2}{S_1^2} < 5\right)$

   5.3 $P\left(S_1^2 > 4S_2^2\right)$

   5.4 $P\left(S_2^2 > 5S_1^2\right)$

   5.5 k ซึ่ง $P\left(\frac{S_1^2}{S_2^2} < k\right) = 0.95$

6. จงหาความน่าจะเป็นที่ความแปรปรวนของตัวอย่างของส่วนสูงคณะ ก มากกว่า 2เท่าของคณะ ข เมื่อสุ่มตัวอย่างคณะ ก มา 14 คน และคณะ ข มา 17 สมมติประชากรทั้งสองมีการแจกแจงปกติและมีความแปรปรวน 25 และ 100 เซนติเมตร^2 ตามลำดับ

สรุป

ในบทนี้จะกล่าวถึงการสุ่มตัวอย่างแต่จะไม่ลงรายเอียดเรื่องวิธีการสุ่ม แต่จะเน้นทฤษฎีการสุ่มตัวอย่าง เริ่มด้วยการแจกแจงของค่าเฉลี่ยเลขคณิตของตัวอย่างขนาด n ซึ่งได้ค่า $\mu_{\bar{X}} = \mu$ และ $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$ เมื่อ n มีค่ามาก การแจกแจงดังกล่าวจะประมาณด้วยการแจกแจงปกติ จากนั้นกล่าวถึงการแจกแจงสัดส่วนของตัวอย่างคือการพิจารณาจากการแจกแจงทวินาม Bin(n, p) ที่ n มีค่ามาก ๆ อาจประมาณด้วยการแจกแจงปกติได้โดย $\mu = np$ และ $\sigma^2 = np(1-p)$ ถัดมาการแจกแจงของค่าเฉลี่ยเลขคณิตของตัวอย่างขนาด n ซึ่งมาน้อยกว่า 30 อาจใช้การแจกทีช่วยในการประมาณค่าเมื่อไม่ทราบความแปรปรวนของประชากร จากนั้นกล่าวถึงการใช้การแจกแจงไคสแควร์และเอฟที่เกิดจากความแปรปรวนของตัวอย่าง

แบบฝึกหัดบทที่ 7

1. ถ้า $X_1, X_2, ..., X_{10}$ แทนตัวอย่างสุ่มขนาด 10 ถ้า $\sum_{i=1}^{10} X_i = 60$ และ $\sum_{i=1}^{10} X_i^2 = 500$ จงหาค่าเฉลี่ยและความแปรปรวนของตัวอย่างดังกล่าว

2. ถ้า $X_1, X_2, ..., X_6$ และ $Y_1, Y_2, ..., Y_5$ เป็นตัวอย่างสุ่มจากประชากรปกติ 2 พวกซึ่งเป็นอิสระต่อกันและมีความแปรปรวนเท่ากัน จากคำนวณเบื้องต้น $\bar{X} = 8.5$, $\sum_{i=1}^{6}(X_i - \bar{X})^2 = 0.34$ และ $\bar{Y} = 8.2$, $\sum_{i=1}^{5}(Y_i - \bar{Y})^2 = 0.1$ จงหาความน่าจะเป็นที่ค่าเฉลี่ยของประชากรกลุ่ม X มากกว่ากลุ่ม Y น้อยกว่า 0.165

3. สมมติว่าน้ำหนักนักเรียนหญิงของโรงเรียนแห่งหนึ่งจำนวน 1500 คน มีการแจกแจงปกติ มีค่าเฉลี่ย 55 กิโลกรัม และส่วนเบี่ยงเบนมาตรฐาน 5 กิโลกรัม เลือกตัวอย่างมา 80 ชุด ประกอบด้วยชุดละ 25 คน จะมีตัวอย่างกี่ชุดที่มีน้ำหนักเฉลี่ย

   3.1 ระหว่าง 40 ถึง 60 กิโลกรัม
   3.2 มากกว่า 50

4. หลอดไฟที่ผลิตโดยโรงงานที่ 1 ที่มีอายุเฉลี่ย 1400 ชั่วโมง และส่วนเบี่ยงเบนมาตรฐาน 200 ชั่วโมง หลอดไฟที่ผลิตโดยโรงงานที่ 2 ที่มีอายุเฉลี่ย 1200 ชั่วโมง และส่วนเบี่ยงเบนมาตรฐาน 100 ชั่วโมง สุ่มตัวอย่างหลอดไฟแต่ละโรงงานมา 125 หลอด และทดลองใช้ จงหาความน่าจะเป็นที่อายุการใช้งานเฉลี่ยของหลอดไฟที่ผลิตจากโรงงานที่ 1 จะมากกว่าโรงงานที่ 2 อย่างน้อย 160 ชั่วโมง

5. บริษัทผลิตน้ำดื่มบรรจุขวดแห่งหนึ่งบรรจุน้ำในขวดที่มีฉลากปิดบอกปริมาตรสุทธิ 600 มิลลิลิตร ถ้าปริมาตรสุทธิของน้ำดื่มมีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเป็น 600 มิลลิลิตรและ ความแปรปรวน 4.2 มิลลิลิตร^2 ในการบรรจุน้ำดื่มลงในแต่ละขวดจะต้องให้ปริมาตรสุทธิใกล้เคียงกับที่บอกไว้บนฉลากมากที่สุด แต่ปริมาตรจริงที่ยอมรับได้เป็น 599 ถึง 601 มิลลิลิตร ถ้า

บริษัทชัก ตัวอย่างน้ำดื่มขนาด 16 ขวด จงหาความน่าจะเป็นที่ค่าเฉลี่ยของปริมาตรสุทธิอยู่ในเกณฑ์ที่ยอมรับได้

6. ถ้า 30% ของผู้ดูทีวีชอบดูช่องหมายเลข 999 มากกว่าช่องอื่น ๆ ถ้าสุ่มตัวอย่างมา 100 คน จงหาความน่าจะเป็นที่สัดส่วนตัวอย่างของผู้ที่ชอบดูช่องหมายเลข 999 จะต่ำกว่า 25%

7. สัดส่วนของผู้ชายที่เห็นด้วยกับการทำแท้งเท่ากับ 0.40 และสัดส่วนของผู้หญิงที่เห็นด้วยกับการทำแท้งเท่ากับ 0.32 ถ้าสุ่มตัวอย่างผู้ชายและผู้หญิงมากลุ่มละ 100 คน จงหาความน่าจะเป็นที่สัดส่วนตัวอย่างที่เห็นด้วยกับการทำแท้งของผู้ชายสูงกว่าผู้หญิงมากกว่า 0.05

8. ในการสุ่มตัวอย่างจากนักเรียนที่เข้าสอบ ONET วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ระดับประถมศึกษาปีที่ 6 จำนวน 50 คน และ 60 คน ตามลำดับ แล้วคำนวนส่วนเบี่ยงเบนมาตรฐานได้เป็น 15 และ 10 คะแนน ตามลำดับ จงหาความน่าจะเป็นที่คะแนนเฉลี่ยของตัวอย่างวิชาคณิตศาสตร์จะสูงกว่าวิชาวิทยาศาสตร์มากกว่า 10 คะแนน

9. ในการสุ่มตัวอย่างลูกค้า 5 คน ที่เข้าคอร์สลดน้ำหนักกับสถาบันเสริมความงามแห่งหนึ่ง เป็นระยะเวลา 3 เดือน ปรากฏว่ามีความแปรปรวนของผลต่างของน้ำหนัก 16 กิโลกรัม^2 ถ้าข้อมูลมีการแจกแจกปกติ โดยสถาบันเสริมความงามอ้างว่าหลังจบคอร์สน้ำหนักจะลดลงโดยเฉลี่ย 8 กิโลกรัม จงหาความน่าจะเป็นที่น้ำหนักหลังจบคอร์สของตัวอย่างจะลดลงโดยเฉลี่ยน้อยกว่า 5 กิโลกรัมเมื่อเทียบกับน้ำหนักก่อนเข้าคอร์ส

10. เวลาที่ครูสอนคณิตศาสตร์ 2 ท่านจะปล่อยช้า (นาที) มีการแจกแจงปกติ ถ้าส่มคาบเรียนแรกที่สอนของแต่ละวันที่ครูทั้ง 2 มีสอนจำนวน 2 สัปดาห์ โดยไม่ตรงวันหยุดราชการและครูทั้ง 2 มีสอนทุกวัน พบว่าความแปรปรวนของผลเวลาที่ปล่อยช้าเท่ากับ 25 นาที^{2} จงหาความน่าจะเป็นของผลต่างของเวลาที่ครูทั้ง 2 ปล่อยช้าไม่เกิน 5 นาที จากการสุ่มดังกล่าว

11. สุ่มตัวอย่างขนาด 10 จากประชากรที่มีการแจกแจงแบบปกติที่มีความแปรปรวนของประชากรเท่ากับ 49 จงประมาณความน่าจะเป็นที่ส่วนเบี่ยงเบนมาตรฐานของตัวอย่างมากกว่า 3 แต่ไม่เกิน 8

12. ตัวอย่างสุ่มขนาด 21 จากประชากรปกติ ซึ่งมีความแปรปรวนของประชากรเท่ากับ 10 จงหาความน่าจะเป็นที่

    12.1 ความแปรปรวนของตัวอย่างจะมีค่ามากกว่า 15

    12.2 ความแปรปรวนของตัวอย่างจะมีค่าน้อยกว่า 9

    12.3 ความแปรปรวนของตัวอย่างจะมีอยู่ระหว่าง 5 ถึง 15

บทที่ 8

การประมาณค่า

หลักการสำคัญอย่างหนึ่งในทางสถิติคือการประมาณค่าพารามิเตอร์ทำได้ 2 แบบคือ การประมาณค่าแบบจุด และการประมาณค่าแบบช่วง

8.1 ชนิดของการประมาณค่า

การประมาณค่าพารามิเตอร์ของประชากรโดยใช้ค่าสถิติตัวอย่างเช่น

1. ประมาณค่าเฉลี่ยของประชากร μ โดยใช้ค่าเฉลี่ยของตัวอย่าง $\bar{X}$

2. ประมาณค่าความแปรปรวนของประชากร σ^2 โดยใช้ค่าความแปรปรวนของตัวอย่าง S^2

เรียกว่า **การประมาณค่าแบบจุด** (point estimation) การประมาณค่าดังกล่าวเราไม่ได้คาดหวังว่าค่าของตัวอย่างจะเท่ากับประชากร แต่ต้องการความผิดพลาดไม่มากนัก เราจึงตัดสินใจล่วงหน้าว่าจะใช้ตัวประมาณตัวใดในการประมาณค่าพารามิเตอร์ โดยพิจารณาจากความไม่ลำเอียง

บทนิยาม 8.1.1 สถิติ $\hat{\theta}$ เป็น **ตัวประมาณค่าไม่ลำเอียง** (unbiased estimator) ของพารามิเตอร์ θ ถ้า

$$E(\hat{\theta}) = \theta$$

โดยทฤษฎีบท 7.2.2 $E(\bar{X}) = \mu$ ดังนั้น $\bar{X}$ เป็นตัวประมาณค่าไม่ลำเอียงของ μ

ตัวอย่าง 8.1.2 จงแสดงว่า S^2 เป็นตัวประมาณค่าไม่ลำเอียงของ σ^2

บทพิสูจน์. ให้ประชากรมีค่าเฉลี่ยเลขคณิต μ และความแปรปรวน σ^2 สุ่มตัวอย่างขนาด n ประกอบด้วย $X_1, X_2, ..., X_n$ ได้ความแปรปรวน S^2 พิจารณา

$$\begin{aligned}
E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) &= \frac{1}{n-1}E\left(\sum_{i=1}^{n}[(X_i-\mu)-(\bar{X}-\mu)]^2\right) \\
&= \frac{1}{n-1}E\left(\sum_{i=1}^{n}[(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2]\right) \\
&= \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + \sum_{i=1}^{n}(\bar{X}-\mu)^2\right)
\end{aligned}$$

เนื่องจาก $\sum_{i=1}^{n}(X_i - \mu) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu = n\bar{X} - n\mu = n(\bar{X} - \mu)$ จะได้ว่า

$$
\begin{aligned}
E(S^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) + n(\bar{X} - \mu)^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} \sigma_{X_i}^2 - n\sigma_{\bar{X}}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} \sigma^2 - n \cdot \frac{\sigma^2}{n}\right) \\
&= \frac{1}{n-1}\left(n\sigma^2 - \sigma^2\right) = \frac{1}{n-1} \cdot (n-1)\sigma^2 = \sigma^2
\end{aligned}
$$

ดังนั้น $E(S^2) = \sigma^2$ หรือกล่าวได้ว่า S^2 เป็นตัวประมาณค่าไม่ลำเอียงของ σ^2 □

บทนิยาม 8.1.3 ในบรรดาตัวประมาณค่าไม่ลำเอียงทั้งหลายของพารามิเตอร์ θ ตัวประมาณค่าซึ่งมี**ความแปรปรวนน้อยที่สุด**คือตัวประมาณค่าที่มี**ประสิทธิภาพมากที่สุด**

การประมาณค่าแบบช่วง (interval estimation) คือการประมาณค่าพารามิเตอร์มักอยู่ในรูป

$$a < \theta < b$$

ซึ่ง a และ b ขึ้นอยู่กับตัวประมาณค่าสถิติ $\hat{\theta}$ ของพารามิเตอร์ θ และการแจกแจงความน่าจะเป็นของ $\hat{\theta}$ เมื่อสุ่มตัวอย่างมาชุดหนึ่งแล้วคำนวนค่าได้เป็น

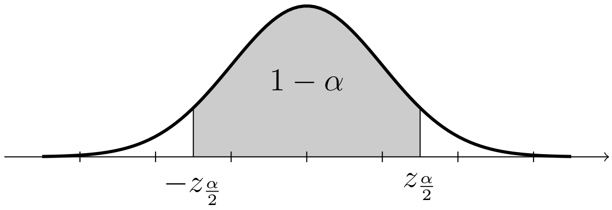
$$P(a < \theta < b) = 1 - \alpha$$

- เรียก (a, b) ว่า**ช่วงความเชื่อมั่น** $(1-\alpha)100\%$ (confidence interval)
- เรียก a และ b ว่า **ลิมิตความเชื่อมั่น** (confidence limits) หรือ **ลิมิตความไว้วางใจ** (fiducial limits)
- เรียก $1-\alpha$ ว่า**สัมประสิทธิ์ความเชื่อมั่น** (confidence coeffient)

8.2 การประมาณค่าเฉลี่ยประชากร

การประมาณค่าแบบจุดของค่าเฉลี่ยประชากร μ คือ $\bar{X}$ การแจกแจงความน่าจะเป็นของ $\bar{X}$ ใกล้เคียงกับการแจกแจงปกติซึ่งมีค่าเฉลี่ย $\mu_{\bar{X}} = \mu$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

รูปที่ 8.1: กราฟการแจกแจงปกติแสดงพื้นที่ $1-\alpha$



กำหนดสัญลักษณ์ z_A ซึ่งเป็นจำนวนบวก หมายถึงค่าที่ทำให้ $P(Z > z_A) = P(Z < -z_A) = A$ เราสนใจ $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1-\alpha$ พิจารณา $-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}$ จะได้ว่า

$$\begin{aligned}
-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}} \\
-z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \\
\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \quad \mu \quad < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

ถ้าสุ่มตัวอย่างขนาด n ที่มีค่าเฉลี่ย $\bar{X}$ จากประชากรซึ่งมีความแปรปรวน σ^2 ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ μ คือ $\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}$ หรือ

$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \ \mu \ < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

ค่า z ที่มักพบในการคำนวณ

1. ช่วงความเชื่อมั่น 99% หรือ $\alpha = 0.01$ จะมี $z_{\frac{\alpha}{2}} = z_{0.005} = 2.58$
2. ช่วงความเชื่อมั่น 95% หรือ $\alpha = 0.05$ จะมี $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$
3. ช่วงความเชื่อมั่น 94% หรือ $\alpha = 0.06$ จะมี $z_{\frac{\alpha}{2}} = z_{0.03} = 1.88$
4. ช่วงความเชื่อมั่น 90% หรือ $\alpha = 0.10$ จะมี $z_{\frac{\alpha}{2}} = z_{0.05} = 1.65$

ตัวอย่าง 8.2.1 ในการสอบ ONET วิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ของปีการศึกษา 2560 มีคนเข้าสอบจากทั่วประเทศจำนวน 707,633 คน มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 16.44 (ข้อมูลจาก สทศ.) โดยคะแนนสอบมีการแจกแจงปกติ สุ่มตัวอย่าง 20 คนปรากฏคะแนนดังนี้

15	16	20	40	50	60	80	90	55	37
72	61	52	49	33	25	32	10	8	30

จงหาช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยคะแนนวิชาคณิตศาสตร์ของการสอบในครั้งนี้
แนวคำตอบ ประชาการมีการแจกแจงปกติซึ่ง $\sigma = 16.44$ จากข้อมูลตัวอย่างมีขนาด $n = 20$
คำนวณค่า $\bar{X} = 41.75$ (โดยเครื่องคิดเลข)
ช่วงความเชื่อมั่น 95% มี $\alpha = 0.5$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $z_{0.025} = 1.96$ จะได้ว่า

$$\bar{X} \pm z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} = 41.75 \pm 1.96 \cdot \frac{16.44}{\sqrt{20}} = 41.75 \pm 7.21$$

ฉะนั้น

$$\begin{aligned} 41.75 - 7.21 &< \mu < 41.75 + 4.21 \\ 34.54 &< \mu < 48.96 \end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยคะแนนวิชาคณิตศาสตร์ของการสอบในครั้งนี้คือ $(34.54, 48.96)$

ตัวอย่าง 8.2.2 ในการสอบ ONET วิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ของปีการศึกษา 2562 มีคนเข้าสอบจากทั่วประเทศจำนวน 692,673 คน (ข้อมูลจาก สทศ.) โดยคะแนนสอบมีการแจกแจงปกติ สุ่มตัวอย่าง 36 คนปรากฏคะแนนดังนี้

18	23	30	41	50	42	33	80	45	38	13	50
25	60	29	30	35	40	43	37	10	29	41	30
15	32	10	32	33	45	70	80	95	65	28	25

จงหาช่วงความเชื่อมั่น 99% ของค่าเฉลี่ยคะแนนวิชาคณิตศาสตร์ของการสอบในครั้งนี้
แนวคำตอบ ประชาการมีการแจกแจงปกติไม่ทราบค่า σ^2
สุ่มตัวอย่างขนาด $n = 36 > 30$ คำนวณค่า $\bar{X} = 38.94$ และ $S = 19.78$ (โดยเครื่องคิดเลข)
ประมาณ σ ด้วย S ช่วงความเชื่อมั่น 99% นั่นคือ $\alpha = 0.01$ ฉะนั้น $\frac{\alpha}{2} = 0.005$ และ $z_{0.005} = 2.58$ จะได้ว่า

$$\bar{X} \pm z_{0.05} \cdot \frac{S}{\sqrt{n}} = 38.94 \pm 2.58 \cdot \frac{19.78}{\sqrt{36}} = 38.94 \pm 8.50$$

ฉะนั้น

$$\begin{aligned} 38.94 - 8.50 &< \mu < 38.94 + 8.50 \\ 30.44 &< \mu < 47.44 \end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น 99% ของค่าเฉลี่ยคะแนนวิชาคณิตศาสตร์ของการสอบในครั้งนี้ $(30.44, 47.44)$

การประมาค่า μ ด้วย $\bar{X}$ เราหวังว่าจะมีค่าไม่ห่างกันมากนัก ค่าความแตกต่างของทั้งสอง เรียกว่า **ค่าความคลาดเคลื่อน** (error) เขียนแทนด้วย ε นั่นคือ

$$|\bar{X} - \mu| = \varepsilon$$

ทฤษฎีบท 8.2.3 การประมาค่า μ ด้วย $\bar{X}$ เราสามารถเชื่อมั่น $(1-\alpha)100\%$ ว่า

$$\text{ความคลาดเคลื่อนจะน้อยกว่า} \qquad z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

บทพิสูจน์. ให้ ε เป็นความคลาดเคลื่อนของ μ กับ $\bar{X}$ จะได้ว่า

$$\begin{aligned}
1-\alpha = P(|\bar{X}-\mu| < \varepsilon) &= P\left(\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| < \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right) \\
&= P\left(|Z| < \varepsilon \cdot \frac{\sqrt{n}}{\sigma}\right) \\
&= P\left(-\varepsilon \cdot \frac{\sqrt{n}}{\sigma} < Z < \varepsilon \cdot \frac{\sqrt{n}}{\sigma}\right)
\end{aligned}$$

นั่นคือ $z_{\frac{\alpha}{2}} = \varepsilon \cdot \dfrac{\sqrt{n}}{\sigma}$ ดังนั้น

$$\varepsilon = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

□

บทแทรก 8.2.4 การประมาค่า μ ด้วย $\bar{X}$ เราสามารถเชื่อมั่น $(1-\alpha)100\%$ ว่าความคลาดเคลื่อนจะน้อยกว่า ε

$$\text{เมื่อขนาดตัวอย่างเท่ากับ} \qquad \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2$$

บทพิสูจน์. เป็นผลโดยตรงจากทฤษฎีบท 8.2.3 □

ตัวอย่าง 8.2.5 ในการสอบ ONET วิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ของปีการศึกษา 2559 มีคนเข้าสอบจากทั่วประเทศจำนวน 724,285 คน มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 18.78 (ข้อมูลจาก สทศ.) โดยคะแนนสอบมีการแจกแจงปกติ ต้องสุ่มนักเรียนที่เข้าสอบมาจำนวนกี่คน จึงจะทำให้คะแนนเฉลี่ยของตัวอย่าง ($\bar{X}$) ของการสอบครั้งนี้คลาดเคลื่อนจาก μ ไม่เกิน 5 คะแนน โดยมีความเชื่อมั่น 95%

แนวคำตอบ ความคลาดเคลื่อนของ $\bar{X}$ จาก μ ไม่เกิน 5 คะแนน นั่นคือ $\varepsilon = 5$
ประชากรมี $\sigma^2 = 18.78$ ช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ ฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $z_{0.025} = 1.96$ จะได้ว่า

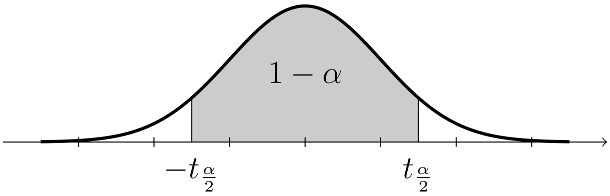
$$n = \left(\frac{1.96(18.78)}{5}\right)^2 = 61.46$$

ดังนั้นต้องสุ่มนักเรียนที่เข้าสอบมาจำนวน 62 คน

ในการประมาณค่าพารามิเตอร์ต่าง ๆ บ่อยครั้งที่เราไม่ทราบค่าความแปรปรวน และขนาดตัวอย่างน้อยกว่า 30 แต่ทราบว่าการแจกแจงของประชากรใกล้เคียงการแจกแจงปกติ เราอาจจะใช้การแจกแจงที และประมาณค่าความแปรปรวนประชากรด้วยค่าความแปรปรวนตัวอย่างซึ่ง

$$T = \frac{\bar{X} - \mu}{\dfrac{S}{\sqrt{n}}} \quad \text{และมีองศาเสรี } \nu = n - 1$$

รูปที่ 8.2: กราฟการแจกแจงทีแสดงพื้นที่ $1 - \alpha$



กำหนดสัญลักษณ์ t_A ซึ่งเป็นจำนวนบวก หมายถึงค่าที่ทำให้ $P(T > t_A) = P(T < -t_A) = A$ เราสนใจ $P\left(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$ พิจารณา $-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}$ จะได้ว่า

$$-t_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{\frac{\alpha}{2}}$$

$$\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} < \quad \mu \quad < \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

ถ้าสุ่มตัวอย่างขนาด n ที่มีค่าเฉลี่ย $\bar{X}$ และความแปรปรวนเท่ากับ S^2
ดังนั้นช่วงความเชื่อมั่น $(1 - \alpha)100$ % ของ μ คือ $\bar{X} \pm t_{\frac{\alpha}{2}} \cdot \dfrac{S}{\sqrt{n}}$ หรือ

$$\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} < \ \mu \ < \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

ตัวอย่าง 8.2.6 น้ำหนักของมะม่วงสุกพันธุ์อกร่องต่อลูก (หน่วยเป็นกรัม) จากสวนแห่งหนึ่งจำนวน 10 ลูก ดังนี้

30 35 40 28 33 25 30 31 32 32

จงหาช่วงความเชื่อมั่น 95% ของค่าน้ำหนักเฉลี่ยของมะม่วงสุกในสวนแห่งนี้ สมมติประชากรปกติ
แนวคำตอบ ประชาการมีการแจกแจงปกติไม่ทราบค่า σ^2 จากข้อมูลตัวอย่างมีขนาด $n = 10 < 30$ จึงประมาณด้วยการแจกแจงทีซึ่งมีองศาเสรี $\nu = 9 - 1 = 8$ คำนวณค่า $\bar{X} = 31.6$ และ $S = 4.03$ ประมาณ σ ด้วย S ช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ ฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $t_{0.025} = 2.262$ จะได้ว่า

$$\bar{X} \pm t_{0.025} \cdot \frac{S}{\sqrt{n}} = 31.6 \pm 2.262 \cdot \frac{4.03}{\sqrt{10}} = 31.36 \pm 2.88$$

ฉะนั้น

$$\begin{array}{rcccl} 31.6 - 2.88 & < & \mu & < & 31.6 + 2.88 \\ 28.72 & < & \mu & < & 34.48 \end{array}$$

ดังนั้นช่วงความเชื่อมั่น 95% ของค่าน้ำหนักเฉลี่ยของมะม่วงสุกในสวนแห่งนี้ $(28.72, 34.48)$

ถ้าสุ่มตัวอย่างขนาด n_1 และ n_2 โดยอิสระต่อกันจากประชากรสองชุด ซึ่งมีค่าเฉลี่ย μ_1 และ μ_2 ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\mu_1 - \mu_2$ พิจารณาได้ดังต่อไปนี้

ทราบค่าของ σ_1^2 และ σ_2^2

การแจกแจงของผลต่างของค่าเฉลี่ยของตัวอย่าง $\bar{X}_1$ และ $\bar{X}_2$ มีการแจกแจงโดยประมาณเป็นการแจกแจงปกติโดยมีตัวแปรสุ่ม Z คือ

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\mu_1 - \mu_2$ คือ

$$(\bar{X}_1 - \bar{X}_2) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ไม่ทราบค่าของ σ_1^2 และ σ_2^2

1. กรณี n_1 และ n_2 มากกว่าหรือเท่ากับ 30
   จะประมาณด้วยการแจกแจงปกติ และประมาณค่า σ_1^2 และ σ_2^2 ด้วย S_1^2 และ S_2^2 ตามลำดับ โดยที่

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

   ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\mu_1 - \mu_2$ คือ

$$(\bar{X}_1 - \bar{X}_2) \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

2. กรณี n_1 หรือ n_2 น้อยกว่า 30

   - สำหรับ $\sigma_1^2 = \sigma_2^2$ จะประมาณด้วยการแจกแจงที โดยที่

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

     เมื่อ

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

     และองศาเสรีคือ $\nu = n_1 + n_2 - 2$ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\mu_1 - \mu_2$ คือ

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2}} \cdot \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- สำหรับ $\sigma_1^2 \neq \sigma_2^2$ จะประมาณด้วยการแจกแจงที โดยที่

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

  เมื่อองศาเสรีคือ

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 \frac{1}{n_1 - 1} + \left(\frac{S_2^2}{n_2}\right)^2 \frac{1}{n_2 - 1}}$$

  ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\mu_1 - \mu_2$ คือ

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

ตัวอย่าง 8.2.7 สุ่มตัวอย่างยาสีฟันสูตร A จำนวน 25 หลอดเพื่อชั่งน้ำหนัก จากประชากรปกติซึ่งมีส่วนเบี่ยงเบนมาตรฐาน 5 คำนวณค่าเฉลี่ยน้ำหนักของตัวอย่างได้ 150 กรัม และสุ่มตัวอย่างยาสีฟันสูตร B จำนวน 36 หลอดเพื่อชั่งน้ำหนัก จากประชากรปกติซึ่งมีส่วนเบี่ยงเบนมาตรฐาน 3 คำนวณค่าเฉลี่ยน้ำหนักของตัวอย่างได้ 145 กรัม จงหาค่าช่วงความเชื่อมั่น 95% สำหรับผลต่างค่าเฉลี่ยน้ำหนักของประชากรยาสีฟันสูตร A และยาสีฟันสูตร B
แนวคำตอบ พิจารณาข้อมูล

	ยาสีฟันสูตร A	ยาสีฟันสูตร B
ส่วนเบี่ยงเบนมาตรฐาน (ประชากร)	$\sigma_A = 5$	$\sigma_B = 3$
ขนาดตัวอย่าง	$n_A = 25$	$n_B = 36$
ค่าเฉลี่ย (ตัวอย่าง)	$\bar{X}_A = 150$	$\bar{X}_B = 145$

ประชากรปกติซึ่งเป็นอิสระต่อกันซึ่งทราบความแปรปรวน สำหรับช่วงความเชื่อมั่น 95% มี $\alpha = 0.5$ ฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $z_{0.025} = 1.96$ จะได้ว่า

$$\begin{aligned}(\bar{X}_A - \bar{X}_B) \pm z_{0.025} \cdot \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} &= (150 - 145) \pm 1.96 \cdot \sqrt{\frac{5^2}{25} + \frac{3^2}{36}} \\ &= 5 \pm 2.19\end{aligned}$$

ฉะนั้น

$$\begin{aligned} 5 - 2.19 \;&<\; \mu_A - \mu_B \;<\; 5 + 2.19 \\ 2.81 \;&<\; \mu_A - \mu_B \;<\; 7.19 \end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น 95% สำหรับผลต่างค่าเฉลี่ยน้ำหนักของประชากรยาสีฟันสูตร A และยาสีฟันสูตร B คือ $(2.81, 7.19)$

ตัวอย่าง 8.2.8 เมื่อสอบถามคะแนนวิชาแคลคูลัส 1 ปรากฏว่ามีนักศึกชาย 75 คน คำนวนค่าเฉลี่ยได้ 82 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 8 คะแนน มีนักศึกษาหญิงจำนวน 50 คน คำนวนค่าเฉลี่ยได้ 76 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 6 คะแนน จงหาค่าช่วงความเชื่อมั่น 94% สำหรับผลต่างค่าเฉลี่ยคะแนนของประชากรนักศึกษาชายและนักศึกษาหญิง
แนวคำตอบ พิจารณาข้อมูล

	นักศึกษาชาย	นักศึกษาหญิง
ขนาดตัวอย่าง	$n_1 = 75$	$n_2 = 50$
ค่าเฉลี่ย (ตัวอย่าง)	$\bar{X}_1 = 82$	$\bar{X}_2 = 76$
ส่วนเบี่ยงเบนมาตรฐาน (ตัวอย่าง)	$S_1 = 8$	$S_2 = 6$

ประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุดไม่ทราบความแปรปรวน และตัวอย่างมีขนาดใหญ่ เราจะประมาณด้วยการแจกแจงปกติ โดยประมาณค่า σ_1, σ_2 ด้วย S_1, S_2 ตามลำดับ สำหรับช่วงความเชื่อมั่น 94% นั่นคือ $\alpha = 0.06$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.03$ และ $z_{0.03} = 1.88$ จะได้ว่า

$$\begin{aligned}(\bar{X}_1 - \bar{X}_2) \pm z_{0.03} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} &= (82 - 76) \pm 1.88 \cdot \sqrt{\frac{8^2}{75} + \frac{6^2}{50}} \\ &= 6 \pm 2.36\end{aligned}$$

ฉะนั้น

$$\begin{aligned}6 - 2.36 \; < \; \mu_1 - \mu_2 \; < \; 6 + 2.36 \\ 3.64 \; < \; \mu_1 - \mu_2 \; < \; 8.36\end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น 94% สำหรับผลต่างค่าเฉลี่ยคะแนนของประชากรนักศึกษาชายและนักศึกษาหญิง คือ $(3.64, 8.36)$

ตัวอย่าง 8.2.9 จากประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุด มีความแปรปรวนเท่ากัน กำหนดให้ตัวอย่างชุดที่ 1 มีขนาด 9 สุ่มมาจากประชากรชุดที่ 1 คือ

20 30 32 25 27 40 37 21 29

ตัวอย่างชุดที่ 2 มีขนาด 16 สุ่มมาจากประชากรชุดที่ 2 คือ

25 33 35 25 41 39 37 21 29 40 41 53 50 30 31 22

จงหาค่าช่วงความเชื่อมั่น 95% สำหรับ $\mu_1 - \mu_2$ เมื่อ μ_1 และ μ_2 คือค่าเฉลี่ยของประชากรชุดที่ 1 และ 2 ตามลำดับ
แนวคำตอบ จากข้อมูลที่ให้

	ประชากรชุดที่ 1	ประชากรชุดที่ 2
ขนาดตัวอย่าง	$n_1 = 9$	$n_2 = 16$
ค่าเฉลี่ย (ตัวอย่าง)	$\bar{X}_1 = 29$	$\bar{X}_2 = 34.5$
ความแปรปรวน (ตัวอย่าง)	$S_1^2 = 45$	$S_2^2 = 87.2$

ประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุดไม่ทราบความแปรปรวนแต่ทราบว่าเท่ากัน ตัวอย่างมีขนาดเล็ก เราจะประมาณด้วยการแจกแจงทีซึ่งมีองศาเสรี $\nu = 9 + 10 - 2 = 23$ และ

$$S_p^2 = \frac{(9-1)45 + (16-1)87.2}{9+16-2} = 72.52$$

สำหรับช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $t_{0.025} = 2.096$ จะได้ว่า

$$\begin{aligned}(\bar{X}_1 - \bar{X}_2) \pm t_{0.025} \cdot \sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} &= (29 - 34.5) \pm 2.096 \cdot \sqrt{72.52\left(\frac{1}{9} + \frac{1}{16}\right)} \\ &= -5.5 \pm 7.44\end{aligned}$$

ฉะนั้น

$$\begin{aligned} -5.5 - 7.44 &< \mu_1 - \mu_2 < -5.5 + 7.44 \\ -12.94 &< \mu_1 - \mu_2 < 1.94 \end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น 95% ของ $\mu_1 - \mu_2$ คือ $(-12.94, 1.94)$

ตัวอย่าง 8.2.10 จากการบันทึก 15 ปี ที่ผ่านมาได้แสดงว่าปริมาณน้ำฝนในเดือนพฤษภาคม ณ ตำบลที่หนึ่งได้ข้อมูลดังนี้ (หน่วยเป็นนิ้ว)

2.40	2.42	1.87	2.50	2.29
1.68	2.57	1.60	1.65	1.41
1.66	1.32	2.43	1.83	1.41

ข้อมูลตำบลที่สองบันทึก 10 ปี ได้ข้อมูลดังนี้ (หน่วยเป็นนิ้ว)

0.79	1.25	0.72	0.84	1.32
1.35	1.29	0.72	0.96	1.13

จงหาค่าช่วงความเชื่อมั่น 95% สำหรับความแตกต่างของค่าเฉลี่ยที่แท้จริงของปริมาณน้ำฝน สมมติว่าข้อมูลเหล่านี้มาจากประชากรที่ค่าความแปรปรวนแตกต่างกัน

แนวคำตอบ จากข้อมูลที่ให้

	ตำบลที่หนึ่ง	ตำบลที่สอง
ขนาดตัวอย่าง	$n_1 = 15$	$n_2 = 10$
ค่าเฉลี่ย (ตัวอย่าง)	$\bar{X}_1 = 1.94$	$\bar{X}_2 = 1.04$
ความแปรปรวน (ตัวอย่าง)	$S_1^2 = 0.20$	$S_2^2 = 0.07$

ประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุดไม่ทราบความแปรปรวนและทราบว่าไม่เท่ากัน ตัวอย่างมีขนาดเล็ก เราจะประมาณด้วยการแจกแจงทีซึ่งมีองศาเสรี

$$\nu = \frac{\left(\frac{0.20}{15} + \frac{0.07}{10}\right)^2}{\left(\frac{0.20}{15}\right)^2 \frac{1}{15-1} + \left(\frac{0.07}{10}\right)^2 \frac{1}{10-1}} = 22.78 \approx 23$$

สำหรับช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $t_{0.025} = 2.096$ จะได้ว่า

$$(\bar{X}_1 - \bar{X}_2) \pm t_{0.025} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = (1.94 - 1.04) \pm 2.069 \cdot \sqrt{\frac{0.20}{15} + \frac{0.07}{10}}$$
$$= 0.9 \pm 0.29$$

ฉะนั้น

$$\begin{aligned} 0.9 - 0.29 \ &< \ \mu_1 - \mu_2 \ < \ 0.9 + 0.29 \\ 0.61 \ &< \ \mu_1 - \mu_2 \ < \ 1.19 \end{aligned}$$

ดังนั้นค่าช่วงความเชื่อมั่น 95% สำหรับความแตกต่างของค่าเฉลี่ยที่แท้จริงของปริมาณน้ำฝน สมมติว่าข้อมูลเหล่านี้มาจากประชากรที่ค่าความแปรปรวนแตกต่างกันคือ $(0.61, 1.19)$

การทำงานทางสถิติเกี่ยวกับข้อมูลประชากรปกติ 2 ชุดที่เปรียบเทียบกันได้ (ไม่อิสระต่อกัน) สามารถเก็บข้อมูลเป็นคู่ ๆ

ลำดับที่	ประชากรชุดที่ 1	ประชากรชุดที่ 2	ผลต่างของประชากรสองชุด
1	x_1	y_1	d_1
2	x_1	y_1	d_2
3	x_1	y_1	d_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	d_n

ให้ μ_D คือค่าเฉลี่ยเลขคณิตของผลต่าง และ σ_D^2 คือค่าส่วนเบี่ยงเบนมาตรฐานของผลต่าง

เมื่อขนาดตัวอย่าง n มีขนาดไม่มาก และไม่ทราบความแปรปรวนประชากร σ_D^2 อาจประมาณด้วยความแปรปรวนตัวอย่าง S_d^2 และจะประมาณด้วยการแจกแจงทีซึ่ง

$$T = \frac{\bar{D} - \mu_D}{\frac{S_d}{\sqrt{n}}}$$

โดยมีองศาอิสระคือ $\nu = n - 1$
เราสนใจ $P\left(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$ พิจารณา $-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}$ จะได้ว่า

$$-t_{\frac{\alpha}{2}} < \frac{\bar{D} - \mu_D}{\frac{S_d}{\sqrt{n}}} < t_{\frac{\alpha}{2}}$$

$$\bar{D} - t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}} < \quad \mu_D \quad < \bar{D} + t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}}$$

ถ้าสุ่มตัวอย่างเป็นคู่ ๆ ขนาด n ที่มีผลต่างค่าเฉลี่ย $\bar{D}$ และความแปรปรวนเท่ากับ S_d^2
ดังนั้นช่วงความเชื่อมั่น $(1 - \alpha)100$ % ของ μ_D คือ $\bar{D} \pm t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}}$ หรือ

$$\bar{D} - t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}} < \ \mu \ < \bar{D} + t_{\frac{\alpha}{2}} \cdot \frac{S_d}{\sqrt{n}}$$

ตัวอย่าง 8.2.11 ในการสุ่มตัวอย่างนักเรียน 5 คน ระดับชั้นประถมศึกษาปีที่ 6 เพื่อสอบถามคะแนนสอบก่อนเรียนและหลังเรียนวิชาภาษาไทยของโรงเรียนแห่งหนึ่ง ปรากฏดังนี้

นักเรียนคนที่	คะแนนสอบก่อนเรียน	คะแนนสอบหลังเรียน
1	9	15
2	10	18
3	9	10
4	15	14
5	7	13

จงหาค่าช่วงความเชื่อมั่น 98% สำหรับค่าเฉลี่ยของผลต่างจริงในการทดสอบครั้งนี้ สมมติประชากรมีการแจกแจงปกติ

แนวคำตอบ พิจารณาผลต่างโดยใช้คะแนนหลังเรียน - คะแนนก่อนเรียน นั่นคือ

นักเรียนคนที่	คะแนนสอบก่อนเรียน (x_i)	คะแนนสอบหลังเรียน (y_i)	$d_i = y_i - x_i$
1	9	15	6
2	10	18	8
3	9	10	1
4	15	14	-1
5	7	13	6

สุ่มตัวอย่างเป็นคู่ ๆ ขนาด $n = 5$ จากประชากรปกติ เราประมาณด้วยการแจกแจงทีซึ่งมีองศาเสรี $\nu = 5 - 1 = 4$ คำนวณค่า $\bar{D} = 4$ และ $S_d = 3.81$ สำหรับช่วงความเชื่อมั่น 98% นั่นคือ $\alpha = 0.02$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.01$ และ $t_{0.01} = 3.75$ จะได้ว่า

$$\bar{D} \pm t_{0.01} \cdot \frac{S_d}{\sqrt{n}} = 4 \pm 3.75 \cdot \frac{3.81}{\sqrt{5}} = 4 \pm 6.39$$

ฉะนั้น

$$\begin{aligned} 4 - 6.39 &< \mu < 4 + 6.39 \\ -2.39 &< \mu < 10.39 \end{aligned}$$

ดังนั้นช่วงความเชื่อมั่น 98% สำหรับค่าเฉลี่ยของผลต่างจริงในการทดสอบครั้งนี้คือ $(-2.39, 10.39)$

แบบฝึกหัด 8.1 - 8.2

1. จงแสดงว่า $\hat{P}$ เป็นตัวประมาณค่าไม่ลำเอียงของ p

2. จงแสดงว่า $\bar{X}$ เป็นตัวประมาณค่าไม่ลำเอียงของ μ ที่มีประสิทธิภาพมากที่สุด

3. ในการสอบ ONET วิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของปีการศึกษา 2560 มีคนเข้าสอบจากทั่วประเทศจำนวน 372,853 คน มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 17.59 (ข้อมูลจาก สทศ.) โดยคะแนนสอบมีการแจกแจงปกติ สุ่มตัวอย่าง 20 คนปรากฏคะแนนดังนี้

8	18	20	30	30	60	40	31	25	20
30	45	60	23	17	19	33	40	38	15

   จงหาช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยคะแนนวิชาคณิตศาสตร์ของการสอบในครั้งนี้

4. อายุหลอดไฟฟ้าของบริษัทอุปกรณ์ไฟฟ้าแห่งหนึ่งมีการแจกแจงปกติซึ่งมีส่วนเบี่ยงเบนมาตรฐาน 40 ชั่วโมง ถ้าต้องการความเชื่อมั่น 95% เพื่อประมาณค่าเฉลี่ย μ จะผิดพลาดได้ไม่เกิน 10 ชั่วโมง ต้องใช้ขนาดตัวอย่างเท่าใด

5. ในการสุ่มนักศึกษาสาขาวิชาคณิตศาสตร์เพื่อสอบถามเกรดเฉลี่ย จำนวน 50 คน คำนวณส่วนเบี่ยงเบนมาตรฐานได้ 0.3 ถ้าต้องการความเชื่อมั่น 95% เพื่อประมาณค่าเฉลี่ย μ จะผิดพลาดได้ไม่เกินท่าใด

6. ถ้าสายการบินแห่งหนึ่งมักจะมีจำนวนที่นั่งว่างแต่ละเที่ยว และทราบว่าค่าส่วนเบี่ยงเบนมาตรฐานของจำนวนที่นั่งว่างเป็น 4.1 ที่นั่ง จึงสุ่มตัวอย่างเที่ยวบินที่เคยบินในปีที่แล้วจำนวน 225 เที่ยวบิน คำนวณหาจำนวนเฉลี่ยที่นั่งที่ว่างได้ 11.6 ที่นั่งต่อเที่ยวบิน จงประมาณจำนวนที่นั่งต่อเครื่องบินที่ว่างโดยเฉลี่ยต่อเที่ยวบิน 1 เที่ยว ที่ระดับความเชื่อมั่น 90 %

7. น้ำหนักของทุเรียนต่อลูก (หน่วยเป็นกิโลกรัม) จากสวนแห่งหนึ่งก่อนส่งไปยังตลาดผลไม้ จำนวน 9 ลูก ดังนี้

3.0	2.5	4.0	3.4	4.2	5.1	3.2	3.5	4.3

   จงหาช่วงความเชื่อมั่น 95% ของค่าน้ำหนักเฉลี่ยของทุเรียนจากสวนแห่งดังกล่าว สมมติว่ามีการแจกแจงปกติ

8. สุ่มเลือกตัวอย่างโลหะรูปทรงกระบอก 9 อัน วัดเส้นผ่านศูนย์กลางดังนี้

1.01	0.97	1.03	0.99	0.98	1.01	1.03

   จงหาช่วงความเชื่อมั่น 99% ของเส้นผ่านศูนย์กลางเฉลี่ยของโลหะรูปทรงกระบอกเหล่านี้

9. ระดับความพึงพอใจในการเรียนวิชาคณิตศาสตร์ของนักเรียนในโรงเรียนแห่งหนึ่ง โดยใช้มาตรวัดของลิเคิร์ท (Likert Scale) ซึ่งใช้เกณฑ์ 5 ระดับคือ

1	2	3	4	5
น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด

เมื่อสุ่มนักเรียนมา 15 คน เพื่อสอบถามความพึงพอใจปรากฏคะแนนดังนี้

4 3 3 2 1 5 4 4 3 3 2 2 3 4 5

ถ้าระดับความพึงพอใจของนักเรียนทั้งหมดมีการแจกแจงปกติ จงหาช่วงความเชื่อมั่น 95% ของระดับความพึงพอใจของนักเรียนในโรงเรียนนี้ พร้อมบอกแปลความหมายที่ช่วงที่ได้

10. เมื่อสอบถามคะแนนวิชาสถิติ ปรากฏว่ามีนักเรียนชาย 80 คน คำนวนค่าเฉลี่ยได้ 75 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 9 คะแนน มีนักเรียนหญิงจำนวน 60 คน คำนวนค่าเฉลี่ยได้ 65 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 คะแนน จงหาช่วงความเชื่อมั่น 99% สำหรับผลต่างค่าเฉลี่ยคะแนนของประชากรนักเรียนชายและนักเรียนหญิง

11. จงประมาณค่าแตกต่างระหว่างระยะทางเฉลี่ยต่อน้ำมัน 1 ลิตร ของรถยนต์ยี่ห้อ A และ B โดยรถยนต์ขนาดเดียวกัน ที่ระดับความเชื่อมั่น 95% สุ่มตัวอย่างรถยนต์ยี่ห้อ A มา 7 คัน เก็บข้อมูลระยะทาง (กิโลเมตร) ที่วิ่งได้ต่อน้ำมัน 1 ลิตร ได้ดังนี้

15.3 16.5 15.5 14.2 13.9 12.7 14.3

สุ่มตัวอย่างรถยนต์ยี่ห้อ B มา 8 คัน เก็บข้อมูลระยะทาง (กิโลเมตร) ที่วิ่งได้ต่อน้ำมัน 1 ลิตร ได้ดังนี้

16.1 15.4 16.3 17.2 16.0 15.5 14.9 15.8

สมมติว่าความแปรปรวนของระยางเฉลี่ยของทั้งสองยี่ห้อแตกต่างกัน

12. ในการสุ่มตัวอย่างลูกค้า 5 คน ที่เข้าคอร์สลดน้ำหนักกับสถาบันเสริมความงามแห่งหนึ่ง เป็นระยะเวลา 3 เดือน ปรากฏดังนี้

นักเรียนคนที่	น้ำหนักก่อนเข้าคอร์ส (กก.)	น้ำหนักหลังจบคอร์ส (กก.)
1	70	68
2	80	77
3	75	74
4	68	67
5	95	85

จงหาค่าช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยของผลต่างจริงของน้ำหนักในการเข้าคอร์สครั้งนี้

8.3 การประมาณค่าสัดส่วนประชากร

ตัวประมาณแบบจุดของสัดส่วน p ในการทดลองทวินามคือ $\hat{P} = \dfrac{X}{n}$ สำหรับการประมาณแบบช่วง พิจารณา $\hat{P}$ ถ้าใกล้เคียงด้วยการแจกแจงปกติ โดยที่

$$\text{ค่าเฉลี่ย} \quad \mu_{\hat{P}} = p \quad \text{และส่วนเบี่ยงเบนมาตรฐาน} \ \sigma_{\hat{P}} = \sqrt{\frac{pq}{n}}$$

โดยที่ตัวแปรสุ่ม Z คือ

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}}$$

ในทางปฏิบัติเราต้องการหา p จึงไม่ทราบค่าดังกล่าวเนื่องจากตัวแปรสุ่มนี้ประมาณด้วยการแจกแจงปกติ เมื่อไม่ทราบ $\sigma_{\hat{P}}$ อาจประมาณด้วย $\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}}$

เราสนใจ $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ พิจารณา $-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}$ จะได้ว่า

$$-z_{\frac{\alpha}{2}} < \frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} < z_{\frac{\alpha}{2}}$$

$$\hat{P} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} < \quad p \quad < \hat{P} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

ถ้าสุ่มตัวอย่างขนาด n ที่มีค่าสัดส่วน $\hat{P}$ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ p คือ

$$\hat{P} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

ตัวอย่าง 8.3.1 จากตัวอย่างสุ่มของครอบครัวในเมืองหลวงแห่งหนึ่งจำนวน 500 ครอบครัว พบว่ามี 160 ครอบครัวอาศัยอยู่ในห้องชุด (Condominium) จงหาค่าช่วงความเชื่อมั่น 95% สำหรับสัดส่วนที่แท้จริงของครอบครัวในเมืองหลวงที่อาศัยในห้องชุด

แนวคำตอบ จากการสุ่มตัวอย่างขนาด $n = 500$ สัดส่วนครอบครัวอาศัยอยู่ในห้องชุดคือ $\hat{P} = \frac{160}{500} = 0.32$ สำหรับช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ ฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $z_{0.025} = 1.96$ จะได้ว่า

$$\hat{P} \pm z_{0.025} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.32 \pm 1.96 \cdot \sqrt{\frac{0.32(0.68)}{500}} = 0.32 \pm 0.0409$$

ฉะนั้น

$$\begin{aligned} 0.32 - 0.0409 \quad &< \quad p \quad < \quad 0.3 + 0.0409 \\ 0.2791 \quad &< \quad p \quad < \quad 0.3609 \end{aligned}$$

สัดส่วนที่แท้จริงของครอบครัวในเมืองหลวงที่อาศัยในห้องชุดอยู่ในช่วง 27.91% ถึง 36.09%

การประมาค่า p ด้วย $\hat{P}$ เราหวังว่าจะมีค่าไม่ห่างกันมากนัก ค่าความแตกต่างของทั้งสอง เรียกว่า **ค่าความคลาดเคลื่อน** (error) เขียนแทนด้วย ε นั่นคือ $|\hat{P}-p|=\varepsilon$

ทฤษฎีบท 8.3.2 การประมาค่า p ด้วย $\hat{P}$ เราสามารถเชื่อมั่น $(1-\alpha)100\%$ ว่า

$$\text{ความคลาดเคลื่อนจะน้อยกว่า} \qquad z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

บทพิสูจน์. ให้ ε เป็นความคลาดเคลื่อนของ μ กับ $\bar{X}$ จะได้ว่า

$$1-\alpha = P(|\hat{P}-\mu|<\varepsilon) = P\left(\left|\frac{\hat{P}-\mu}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}\right| < \frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}\right) = P\left(|Z| < \frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}\right)$$
$$= P\left(-\frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} < Z < \frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}\right)$$

นั่นคือ $z_{\frac{\alpha}{2}} = \frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$ ดังนั้น

$$\varepsilon = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

$\square$

บทแทรก 8.3.3 การประมาค่า p ด้วย $\hat{P}$ เราสามารถเชื่อมั่น $(1-\alpha)100\%$ ว่าความคลาดเคลื่อนจะน้อยกว่า ε

$$\text{เมื่อขนาดตัวอย่างเท่ากับ} \qquad \left(\frac{z_{\frac{\alpha}{2}} \cdot \sqrt{\hat{P}(1-\hat{P})}}{\varepsilon}\right)^2$$

บทพิสูจน์. เป็นผลโดยตรงจากทฤษฎีบท 8.3.2 $\square$

ทฤษฎีบท 8.3.4 การประมาค่า p ด้วย $\hat{P}$ เราสามารถเชื่อมั่น $(1-\alpha)100\%$ ว่าความคลาดเคลื่อนจะน้อยกว่า ε

$$\text{เมื่อขนาดตัวอย่างเท่ากับ} \qquad \left(\frac{z_{\frac{\alpha}{2}}}{2\varepsilon}\right)^2$$

บทพิสูจน์. พิจารณา

$$\hat{P}(1-\hat{P}) = \hat{P}-\hat{P}^2 = \frac{1}{4} - \left(\frac{1}{2}-\hat{P}\right)^2 \leq \frac{1}{4}$$

จะได้ว่า $\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}$ จากบทพิสูจน์ทฤษฎีบท 8.3.2

$$P\left(-\frac{\varepsilon}{\frac{1}{2\sqrt{n}}} < Z < \frac{\varepsilon}{\frac{1}{2\sqrt{n}}}\right) \leq P\left(-\frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} < Z < \frac{\varepsilon}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}\right)$$

นั่นคือ $z_{\frac{\alpha}{2}} = 2\varepsilon\sqrt{n}$ ดังนั้น $n = \left(\frac{z_{\frac{\alpha}{2}}}{2\varepsilon}\right)^2$ $\square$

ตัวอย่าง 8.3.5 จากการสำรวจความชอบวิชาคณิตศาสตร์ของโรงเรียนแห่งหนึ่งในปีที่แล้วพบว่าชอบ 35% ในปีนี้ผู้สำรวจคาดว่าผลที่ได้คงไม่ต่างกันมาก ถ้าต้องการช่วงความเชื่อมั่น 95% ของนักเรียนที่ชอบวิชาคณิตศาสตร์จริง ๆ ต้องสุ่มตัวอย่างนักเรียนจำนวนเท่าใดจึงจะมีความผิดพลาดไม่เกิน 10%
แนวคำตอบ สัดส่วนตัวอย่าง $\hat{P} = 0.35$ สำหรับช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ ฉะนั้น $\frac{\alpha}{2} = 0.025$ และ $z_{0.025} = 1.96$ โดยมีความผิดพลาดของ $\hat{P}$ กับ p ไม่เกิน $\varepsilon = 0.10$ จะได้ว่า

$$n = \left(\frac{1.96 \cdot \sqrt{0.35(0.65)}}{0.10}\right)^2 = 87.40$$

ดังนั้นต้องสุ่มตัวอย่างนักเรียนจำนวน 88 คน จึงจะมีความผิดพลาดไม่เกิน 10%

ตัวอย่าง 8.3.6 จากตัวอย่าง 8.3.5 ถ้าไม่ทราบข้อมูลปีที่แล้ว ถ้าต้องการช่วงความเชื่อมั่น 95% ของนักเรียนที่ชอบวิชาคณิตศาสตร์จริง ๆ ต้องสุ่มตัวอย่างนักเรียนจำนวนเท่าใดจึงจะมีความผิดพลาดไม่เกิน 10%
แนวคำตอบ จะได้ว่า

$$n = \left(\frac{1.96}{2(0.10)}\right)^2 = 96.04$$

ดังนั้นต้องสุ่มตัวอย่างนักเรียนจำนวน 97 คน จึงจะมีความผิดพลาดไม่เกิน 10%

ถ้าเลือกตัวอย่าง 2 ชุดขนาด n_1 และ n_2 โดยอิสระต่อกันจากประชากรทวินามซึ่งมี

ค่าเฉลี่ย $\mu_1 = n_1p_1$ และ $\mu_2 = n_2p_2$ และความแปรปรวน $\sigma_1^2 = n_1p_1q_1$ และ $\sigma_2^2 = n_2p_2q_2$

การแจกแจงของสัดส่วนของความสำเร็จ $\hat{P}_1 - \hat{P}_2$ จะประมาณด้วยการแจกแจงปกติที่มี

ค่าเฉลี่ย $\mu_{\hat{P}_1-\hat{P}_2} = p_1 - p_2$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\hat{P}_1-\hat{P}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$

และตัวแปรสุ่ม Z คือ

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}$$

ถ้า n_1 และ n_2 มีขนาดใหญ่เราอาจประมาณค่า p_1 และ p_2 ด้วย $\hat{P}_1$ และ $\hat{P}_2$ ตามลำดับ

ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $p_1 - p_2$ คือ

$$(\hat{P}_1 - \hat{P}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$$

ตัวอย่าง 8.3.7 เครื่องจักร 2 ชนิดที่ใช้ผลิตหนังสือ โดยสุ่มตัวอย่างจากการผลิตเครื่องจักรชนิดที่ 1 จำนวน 1500 เล่ม มีเล่มที่บกพร่องจากการพิมพ์ 90 เล่ม และสุ่มตัวอย่างจากการผลิตเครื่องจักรชนิดที่ 2 จำนวน 2000 เล่ม มีเล่มที่บกพร่องจากการพิมพ์ 80 เล่ม จงหาค่าช่วงความเชื่อมั่น 90% สำหรับผลต่างที่แท้จริงในสัดส่วนของความบกพร่องระหว่างเครื่องจักรชนิดที่ 1 และ 2
แนวคำตอบ จากข้อมูลที่ให้

	เครื่องจักรชนิดที่ 1	เครื่องจักรชนิดที่ 2
ขนาดตัวอย่าง	$n_1 = 1500$	$n_2 = 2000$
หนังสือบกพร่อง	$X_1 = 90$	$X_2 = 80$
สัดส่วนหนังสือบกพร่อง	$\hat{P}_1 = 0.06$	$\hat{P}_2 = 0.04$

ประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุดและตัวอย่างมีขนาดใหญ่ เราจะประมาณด้วยการแจกแจงปกติ สำหรับช่วงความเชื่อมั่น 90% นั่นคือ $\alpha = 0.10$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.05$ และ $z_{0.05} = 1.65$ จะได้ว่า

$$
\begin{aligned}
(\hat{P}_1 - \hat{P}_2) &\pm z_{0.05}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}} \\
(0.06 - 0.04) &\pm 1.65\sqrt{\frac{0.06(0.94)}{1500} + \frac{0.04(0.96)}{2000}} \\
0.02 &\pm 0.0124
\end{aligned}
$$

ฉะนั้น

$$
\begin{aligned}
0.02 - 0.0124 \;&<\; p_1 - p_2 \;<\; 0.02 + 0.0124 \\
0.0076 \;&<\; p_1 - p_2 \;<\; 0.0324
\end{aligned}
$$

ดังนั้นช่วงความเชื่อมั่น 90% สำหรับผลต่างที่แท้จริงในสัดส่วนของความบกพร่องระหว่างเครื่องจักรชนิดที่ 1 และ 2 คือ 0.76% ถึง 3.24%

แบบฝึกหัด 8.3

1. โยนเหรียญ 1 อัน 40 ครั้ง ขึ้นหัว 24 ครั้ง จงหาช่วงความเชื่อมั่น 95% ของอัตราส่วนที่จะขึ้นหัว เมื่อโยนเหรียญไปเรื่อย ๆ ไม่สิ้นสุด
2. บริษัทผู้ผลิตยางรถยนต์แห่งหนึ่งต้องการประมาณสัดส่วนยางรถยนต์ที่ชำรุดระดับความเชื่อมั่น 95% จึงสุ่มตัวอย่างยางรถยนต์ที่ผลิตได้มา 490 อันตรวจพบว่าชำรุด 27 เส้น
3. จากการสอบถามแม่บ้าน 50 คน เกี่ยวกับการใช้น้ำยาล้างจานยี่ห้อ A และ B พบว่าแม่บ้านใช้น้ำยาล้างจานยี่ห้อ A จำนวน 38 คน จงประมาณสัดส่วนแบบช่วงของแม่บ้านที่ ใช้น้ำยาล้างจานยี่ห้อ A ที่ระดับความเชื่อมั่น 90%
4. สุ่มตัวอย่างผู้สูบบุหรี่ 500 คน 86 คนชอบสูบบุหรี่ไฟฟ้า

   4.1 จงหาค่าช่วงความเชื่อมั่น 90% ของอัตราส่วนของผู้ชอบสูบบุหรี่ไฟฟ้า

   4.2 หาค่าความผิดพลาดถ้าประมาณของอัตราส่วนนี้เป็น 0.172 ด้วยความเชื่อมั่น 90%

   4.3 จงหาขนาดตัวอย่าง ถ้าค่าความผิดพลาดไม่เกิน 0.03 ที่ระดับความเชื่อมั่น 90%
5. จากตัวอย่างสุ่มนักศึกษาจำนวน 500 คน พบว่ามี 300 คน จงหาค่าช่วงความเชื่อมั่น 95% สำหรับสัดส่วนที่แท้จริงของครอบครัวในเมืองหลวงที่อาศัยในห้องชุด
6. ต้องใช้ขนาดตัวอย่างขนาดเท่าใด เพื่อประมาณอัตราส่วนนักเรียนที่ชอบวิชาคณิตศาสตร์ให้มีความเชื่อมั่นอย่างน้อย 99% ที่ค่าประมาณนั้นผิดจากอัตราส่วนจริงไม่เกิน 1%
7. จงหาช่วงความเชื่อมั่น 90% ของอัตราส่วนจำนวนนักเรียนที่สอบไม่ผ่านเกณฑ์วิชาคณิตศาสตร์ ถ้าสุ่มตัวอย่างมา 100 คน พบว่าไม่ผ่านเกณฑ์ 10 คน
8. เพื่อตรวจสอบการบวกเลขของนักเรียนระดับประถมศึกษาปีที่ 3 ของเขตการศึกษาแห่งหนึ่งจะต้องใช้ตัวอย่างขนาดเท่าใดที่จะมีความเชื่อมั่นอย่างน้อย 95% ที่มีค่าความผิดพลาดจากอัตราส่วนที่จริงไม่เกิน 1%
9. จากการสอบถามนักศึกษาชายและหญิงที่จบในระดับ ปวส. ว่าต้องการศึกษาต่อใน ระดับปริญญาตรีหรือไม่ โดยสอบถามนักศึกษาชายจ านวน 125 คน และนักศึกษาหญิง 103 คน พบว่า นักศึกษาชายเรียนต่อจำนวน 83 คน นักศึกษาหญิงเรียนต่อ 80 คน จงประมาณค่าผลต่าง สัดส่วนของนักศึกษาชายและหญิงที่เรียนเข้าศึกษาต่อในระดับปริญญาตรี ที่ระดับความเชื่อมั่น 90%
10. โรงเรียนแห่งหนึ่งต้องการจัดครูไปศึกษาดูงานในภาคเรียนฤดูร้อน จึงสำรวจความต้องการของครู ทั้งหมด 205 คน พบว่ามีครูสนใจไปศึกษาดูงานจำนวน 156 คน จงประมาณค่าสัดส่วนแบบจุด และแบบช่วงของครูที่สนใจไปศึกษาดูงาน ที่ระดับความเชื่อมั่น 99%

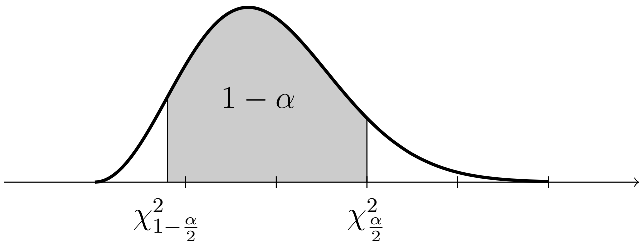
8.4 การประมาณค่าความแปรปรวนประชากร

ตัวประมาณค่าของ σ^2 คือ S^2 ถ้าสุ่มตัวอย่างขนาด n จากประชากรปกติที่มีความแปรปรวน σ^2 จะได้มีการแจกแจงไคสแควร์

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

และมีองศาเสรี $\nu = n-1$ การประมาณค่าความแปรปรวณแบบช่วงทำได้ดังนี้

รูปที่ 8.3: กราฟการแจกแจงไคสแควร์แสดงพื้นที่ $1-\alpha$



กำหนดให้ χ^2_A หมายถึงค่าที่ทำให้ $P(\chi^2 > \chi^2_A) = A$
เราสนใจ $P\left(\chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}\right) = 1-\alpha$ พิจารณา $\chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}$ จะได้ว่า

$$\chi^2_{1-\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}$$

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ σ^2 คือ

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}$$

ตัวอย่าง 8.4.1 สุ่มตัวอย่างคะแนนสอบย่อยครั้งหนึ่งของวิชาสถิติของนักศึกษามาจำนวน 10 คน แสดงได้ดังนี้

9 4 5 6 7 3 1 2 5 5

จงหาช่วงความเชื่อมั่น 95% สำหรับความแปรปรวนที่แท้จริงของคะแนนสอบย่อยวิชาสถิติในครั้งนี้
แนวคำตอบ ประชากรมีการแจกแจงปกติ โดยที่ $n = 10$ และ $S^2 = 5.57$ (โดยเครื่องคิดเลข) และองศาเสรีคือ $\nu = 10 - 1 = 9$
ช่วงความเชื่อมั่น 95% นั่นคือ $\alpha = 0.05$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.025$ จะได้ว่า

$$\chi^2_{0.025} = 19.023 \text{ และ } \chi^2_{0.975} = 2.7 \text{ (เปิดตาราง)}$$

ช่วงความเชื่อมั่นของ σ^2 คือ

$$\begin{aligned}
\frac{(n-1)S^2}{\chi^2_{0.025}} \quad &< \quad \sigma^2 \quad < \quad \frac{(n-1)S^2}{\chi^2_{0.975}} \\
\frac{(10-1)5.57}{19.023} \quad &< \quad \sigma^2 \quad < \quad \frac{(10-1)5.57}{2.7} \\
2.63 \quad &< \quad \sigma^2 \quad < \quad 18.56
\end{aligned}$$

ตัวประมาณค่าอัตราส่วนของความแปรปรวนสองประชากร $\frac{\sigma_1^2}{\sigma_2^2}$ คืออัตราส่วนความแปรปรวนจากตัวอย่าง $\frac{S_1^2}{S_2^2}$ เมื่อ S_1^2 และ S_2^2 เป็นความแปรปรวนของตัวอย่างสุ่มขนาด n_1 และ n_2 จากประชากรปกติที่มีความแปรปรวน σ_1^2 และ σ_2^2 ตามลำดับ จะได้ $F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$ ที่มีองศาเสรี $\nu_1 = n_1 - 1$ และ $\nu_2 = n_2 - 1$ โดยให้ความแปรปรวนที่มากกว่าเป็น S_1^2

กำหนดให้ f_A หมายถึงค่าที่ทำให้ $P(F > f_A) = A$
เราสนใจ $P\left(f_{1-\frac{\alpha}{2}} < F < f_{\frac{\alpha}{2}}\right) = 1 - \alpha$ พิจารณา $f_{1-\frac{\alpha}{2}} < F < f_{\frac{\alpha}{2}}$ จะได้ว่า

$$f_{1-\frac{\alpha}{2}} < \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} < f_{\frac{\alpha}{2}}$$

$$\frac{1}{f_{\frac{\alpha}{2}}} \cdot \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{1-\frac{\alpha}{2}}} \cdot \frac{S_1^2}{S_2^2}$$

ดังนั้นช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ $\frac{\sigma_1^2}{\sigma_2^2}$ คือ

$$\frac{1}{f_{\frac{\alpha}{2}}} \cdot \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{f_{1-\frac{\alpha}{2}}} \cdot \frac{S_1^2}{S_2^2}$$

ตัวอย่าง 8.4.2 สุ่มตัวอย่างคะแนนวิชาแคลคูลัส 1 ของนักศึกษาชายจำนวน 24 คน คือ

90	80	75	60	72	62	72	60
64	52	53	49	50	66	79	61
50	57	60	71	75	80	84	41

และนักศึกษาหญิงจำนวน 10 คน คือ

89	55	70	63	70	69	71	65	62	53

จงหาช่วงความเชื่อมั่น 98% ของ สำหรับอัตราส่วนของความแปรปรวนที่แท้จริงของประชากรทั้งสองชุด

แนวคำตอบ จากข้อมูลที่ให้

	นักศึกษาชาย	นักศึกษาหญิง
ขนาดตัวอย่าง	$n_1 = 24$	$n_2 = 10$
ความแปรปรวน (ตัวอย่าง)	$S_1^2 = 162.03$	$S_2^2 = 100.67$

จะเห็นว่าประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุด และ $\nu_1 = 24 - 1 = 23, \nu_2 = 10 - 1 = 9$ ช่วงความเชื่อมั่น 98% นั่นคือ $\alpha = 0.02$ เพราะฉะนั้น $\frac{\alpha}{2} = 0.01$ ช่วงความเชื่อมั่นของ $\frac{\sigma_1^2}{\sigma_2^2}$ คือ

$$\frac{1}{f_{0.01}} \cdot \frac{S_1^2}{S_2^2} \quad < \quad \frac{\sigma_1^2}{\sigma_2^2} \quad < \quad \frac{1}{f_{0.99}} \cdot \frac{S_1^2}{S_2^2}$$

$$\frac{1}{4.75} \cdot \frac{162.03}{100.67} \quad < \quad \frac{\sigma_1^2}{\sigma_2^2} \quad < \quad \frac{1}{0.303} \cdot \frac{162.03}{100.67}$$

$$0.34 \quad < \quad \frac{\sigma_1^2}{\sigma_2^2} \quad < \quad 5.31$$

แบบฝึกหัด 8.4

1. สุ่มตัวอย่างคะแนนสอบย่อยครั้งหนึ่งของวิชาแคลคูลัส 1 ของนักศึกษามาจำนวน 10 คน แสดงได้ดังนี้ 7, 5, 5, 3, 8, 4, 2, 2, 6, 7 จงหาช่วงความเชื่อมั่น 95% สำหรับความแปรปรวนที่แท้จริงของคะแนนสอบย่อยวิชาแคลคูลัส 1 ในครั้งนี้

2. สุ่มตัวอย่างนักศึกษาจำนวน 10 คน ที่เข้าเรียน 8.00 เพื่อสอบถามเวลาเข้าเรียนแสดงไดดังนี้ 1, 10, 0, -5, -30, 5, -5, -6, -20 ค่าที่เป็นบวกหมายถึงเวลาที่เข้าเรียนสาย ศูนย์หมายถึงทันเวลา 8.00 พอดี และถ้าค่าเป็นลบหมายถึงเข้าเรียนก่อนเวลา จงหาช่วงความเชื่อมั่น 98% สำหรับความแปรปรวนที่แท้จริงของเวลาในการเข้าเรียนในครั้งนี้

3. สุ่มตัวอย่างน้ำหนักของประชาชนที่มาอายุมากกว่า 50 ปี ในจังหวัดหนึ่งจำนวน 100 คน คำนวนความแปรปรวนได้เท่ากับ 10 กิโลกรัม^2 จงหาช่วงความเชื่อมั่น 94% สำหรับความแปรปรวนที่แท้จริงของน้ำหนักของประชาชนที่มาอายุมากกว่า 50 ปี ในจังหวัดนี้

4. จากผลผลิต (หน่วยถังต่อไร่) ของข้าว 2 พันธุ์ ทดลองปลูกพันธุ์ที่ 1 ใน 9 แปลง และปลูกพันธุ์ที่ 2 ใน 11 แปลง ได้ผลดังนี้

พันธุ์ที่ 1	36	32	34	40	36	33	37	32	34		
พันธุ์ที่ 2	34	38	39	38	37	35	42	43	39	38	35

จงหาช่วงความเชื่อมั่น 98% สำหรับอัตราส่วนของความแปรปรวนที่แท้จริงของประชากรทั้งสองชุด

5. ในการสุ่มตัวอย่างคะแนนสอบ ONET ของระดับชั้นมัธยมศึกษาปีที่ 3 วิชาคณิตศาสตร์จำนวน 10 คน และวิทยาศาสตร์จำนวน 15 คน คำนวนความแปรปรวนได้เป็น 13 และ 15 ตามลำดับ จงหาช่วงความเชื่อมั่น 95% สำหรับอัตราส่วนของความแปรปรวนที่แท้จริงของประชากรทั้งสองชุด

6. อายุการใช้งาน (หน่วยเป็นชั่วโมง) ของหลอดไฟ 2 ชนิด เปรียบเทียบโดยการเปิดตลอดเวลาหมดอายุการใช้งาน หลอดไฟชนิดที่ 1 นำมาทดลอง 8 หลอด ได้อายุการใช้งานดังนี้

11122 10641 11208 10097 12087 11759 13104 10638

หลอดไฟชนิดที่ 2 นำมาทดลอง 11 หลอด ได้อายุการใช้งานดังนี้

11371 14971 11714 10760 12759 11607 10796 10963
11981 12032 12265

จงหาช่วงความเชื่อมั่น 90% สำหรับอัตราส่วนของความแปรปรวนที่แท้จริงของประชากรทั้งสองชุด

สรุป

ในบทนี้กล่าวถึงการประมาณพารามิเตอร์ซึ่งทำได้ 2 แบบ 1. การประมาณค่าแบบจุด คือการประมาณค่าพารามอเตอร์ด้วยค่าสถิติ 2. การประมาณค่าแบบช่วงคือการประมาณค่าพารามิเตอร์ในรูปช่วง (a,b) ซึ่งเรียกว่าช่วงความเชื่อมั่น $(1-\alpha)100\%$ ประกอบด้วยการประมาณค่าเฉลี่ยประชากรคือช่วง $\bar{X}\pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ เมื่อทราบความแปรปรวนประชากร ถ้า n น้อยกว่า 30 และไม่ทราบความแปรปรวนประชากร จะได้การแจกแจงทีแทน z จากนั้นประมาณค่าสัดส่วนประชากร โดยการประมาณการแจกแจงทวินามด้วยแจกแจงปกติและหาช่วงเช่นเดิม และสุดท้ายการประมาณค่าความแปรปรวนประชากรด้วยการแจกแจงไคสแควร์และเอฟ สำหรับประชากรกกลุ่มเดียวและสองกลุ่มตามลำดับ

แบบฝึกหัดบทที่ 8

1. อายุหลอดไฟฟ้าของบริษัทอุปกรณ์ไฟฟ้าแห่งหนึ่งมีการแจกแจงปกติซึ่งมีส่วนเบี่ยงเบนมาตรฐาน 40 ชั่วโมง ถ้าเลือกสุ่มมา 30 หลอด คำนวณค่าเฉลี่ยได้เป็น 780 ชั่วโมง จงหาช่วงความเชื่อมั่น 95% ของค่าเฉลี่ย (μ) ของอายุหลอดไฟจากบริษัทนี้

2. ในการสุ่มนักศึกษาของมหาวิทยาลัยแห่งหนึ่งเพื่อสอบถามส่วนสูง จำนวน 60 คน คำนวนส่วนเบี่ยงเบนมาตรฐานได้ 10 เซนติเมตร ถ้าต้องการความเชื่อมั่น 95% ว่าค่าเฉลี่ย μ จะผิดพลาดได้ไม่เกิน 5 เซนติเมตร ต้องใช้ขนาดตัวอย่างอ่างน้อยเท่าใด

3. จากการสำรวจราคาห้องชุด (คอนโดมิเนียม) ที่มีขนาดห้อง 25-30 ตารางเมตร ตามแนวรถไฟฟ้าปรากฎดังนี้ (หน่วยเป็นล้านบาท)

   1.95 1.53 3.25 2.00 1.55 1.99 4.50 5.40 1.57 1.89 2.35 2.15

   จงหาช่วงความเชื่อมั่น 97% ของราคาเฉลี่ยของห้องชุด สมมติว่าประชากรมีการแจกแจงปกติ

4. จากประชากรปกติซึ่งเป็นอิสระต่อกัน 2 ชุด มีความแปรปรวนเท่ากัน กำหนดให้
   ตัวอย่างชุดที่ 1 มีขนาด 10 สุ่มมาจากประชากรชุดที่ 1 คือ

   80 70 82 65 57 60 73 71 69 81

   ตัวอย่างชุดที่ 2 มีขนาด 15 สุ่มมาจากประชากรชุดที่ 2 คือ

   30 31 37 35 41 37 30 29 29 33 40 49 50 38 35

   จงหาค่าช่วงความเชื่อมั่น 95% สำหรับ $\mu_1-\mu_2$ เมื่อ μ_1 และ μ_2 คือค่าเฉลี่ยของประชากรชุดที่ 1 และ 2 ตามลำดับ

5. ในการสุ่มตัวอย่างนักศึกษา 7 คน ชั้นปีที่ 1 เพื่อสอบถามคะแนนสอบ Quiz1 วิชาแคลคูลัส ๑ และแคลคูลัส ๒ ปรากฏดังนี้

นักเรียนคนที่	คะแนนสอบวิชาแคลคูลัส ๑	คะแนนสอบวิชาแคลคูลัส ๒
1	8	10
2	1	8
3	9	5
4	4	6
5	5	5
6	3	2
7	8	5

จงหาค่าช่วงความเชื่อมั่น 95% สำหรับค่าเฉลี่ยของผลต่างจริงในการสอบ Quiz1 ครั้งนี้ สมมติประชากรปกติ

6. จากตัวอย่างของผู้มีสิทธิ์ออกเสียง 100 คน สุ่มเลือกมาจากผู้มีสิทธิ์ออกเสียงทั้งหมดในตำบลแห่งหนึ่งปรากฏว่า 55% ลงคะแนนให้ผู้สมัครรับเลือกตั้งคนหนึ่ง จงหาค่าช่วงความเชื่อมั่น 99% ของอัตราส่วนของคนทั้งหมดในตำบลนี้ ที่จะลงคะแนนเสียงให้กับผู้สมัครรับเลือกตั้งคนนี้

7. สุ่มชาวนาในจังหวัดแห่งหนึ่งมา 100 คน พบว่าเป็นหนี้ธนาคาร ธกส. จำนวน 60 คน จงประมาณช่วงความเชื่อมั่น 95% ของสัดส่วนชาวนาในจังหวัดนี้ที่เป็นหนี้ธนาคาร

8. สุ่มตัวอย่างลูกค้าของร้านเบอเกอรี่จำนวน 100 คน 16 คน ซื้อแซนวิช ภายหลังจากการทำโปรโมชั่นซื้อแซนวิช 1 ชิ้นแถมอีก 1 ชิ้น ได้สุ่มตัวอย่าง 200 คน 80 คน ซื้อแซนวิช อยากทราบว่าการทำโปรโมชั้นมีผลหรือไม่ โดยพิจารณาช่วงตวามเชื่อมั่น 95% ของความแตกต่างระหว่างอัตราส่วนของผู้ซื้อแซนวิช ก่อนและหลังโฆษณา

9. โรงพยาบาลแห่งหนึ่งตรวจสุขภาพผู้สูงวัยที่มีอายุ 60 ปี ขึ้นไป เพื่อหาปริมาณน้ำตาลในเลือด พบว่า เพศหญิงจำนวน 70 คน มีปริมาณน้ำตาลในเลือดเกินเกณฑ์ที่กำหนดจำนวน 37 คน และ เพศชายจำนวน 50 คน มีปริมาณน้ำตาลในเลือดเกินเกณฑ์ที่กำหนดจำนวน 22 คน จงประมาณ ผลต่างสัดส่วนของคนที่มีน้ำตาลในเลือดเกินเกณฑ์ที่กำหนด ที่ระดับความเชื่อมั่น 95%

บทที่ 9

การทดสอบสมมติฐาน

9.1 สมมติฐานสถิติ ความผิดพลาด และการทดสอบ

สมมติฐานสถิติ (Statistical hypothesis) คือข้อสมมติฐานหรือข้อความที่เกี่ยวข้องกับประชากรชุดเดียวหรือมากกว่า ซึ่งอาจเป็นจริงหรือไม่ก็ได้

เพื่อตรวจสอบว่าเหรียญเที่ยงตรงหรือไม่ โดยการทดลองโยนเหรียญนั้น 100 ครั้ง แล้วจดบันทึกจำนวนหัวที่เกิดขึ้น สมมติว่าผู้ทดลองยอมรับว่าเที่ยงตรงถ้าจำนวนหัวอยู่ระหว่าง 40 ถึง 60 ครั้ง ถ้าทดลองแล้วหัวเกิดจำนวน 45 ครั้ง จะสรุปว่าเหรียญเที่ยงตรง แต่ถ้าทดลองแล้วเกิดหัวจำนวน 35 ก็ไม่ยอมรับว่าเหรียญนี้เที่ยงตรง ประโยคที่ว่าเหรียญเที่ยงตรงหรือไม่ คือการ**ตั้งสมมติฐาน** แล้วเราจะทราบได้อย่างไรว่ากระบวนการดังกล่าวสรุปได้ถูกต้องมากน้อยเพียงใด ในบทนี้เราจึงทำการศึกษาเพื่อหาเหตุผลที่ถูกต้องเหมาะสมเพื่อนำไปสรุปผลเกี่ยวกับการยอมรับหรือปฏิเสธสมมติฐาน

การปฏิเสธสมมติฐานหมายถึงสมมติฐานนั้นผิด แต่ถ้ายอมรับสมมติฐานก็แสดงว่าไม่มีหลักฐานเพียงพอที่จะเชื่อเป็นอย่างอื่น ผู้ทำการทดลองมักจะหวังที่จะปฏิเสธสมมติฐาน เช่น

1. ถ้าผลิตวัคซีนชนิดใหม่ สมมติฐานคือประสิทธิภาพวัคซีนเก่าและใหม่ไม่แตกต่างกัน ซึ่งผู้ทดลองมักคาดหวังที่จะปฏิเสธสมมติฐาน

2. การเปรียบเทียบวิธีการสอน 2 แบบ สมมติฐานคือการสอนทั้ง 2 แบบให้ผลสัมฤทธิ์ทางการไม่แตกต่างกัน ซึ่งผู้ทำวิจัยมักคาดหวังที่จะปฏิเสธสมมติฐาน

สมมติฐานที่สร้างขึ้นด้วยความหวังที่จะปฏิเสธเรียกว่า

สมมติฐานหลัก หรือ **สมมติฐานว่าง** (null hypothesis) เขียนแทนด้วย H_0

การปฏิเสธ H_0 ทำให้ยอมรับ **สมมติฐานทางเลือก** (alternative hypothesis) เขียนแทนด้วย H_1 หรือ H_a ตัวอย่างเช่น ถ้า $H_0 : p = 0.5 \quad H_1$ เป็นไปได้หลายแบบเช่น

$$H_1 : p = 0.6 \qquad H_1 : p \neq 0.5 \qquad H_1 : p > 0.5 \qquad H_1 : p < 0.5$$

ในการตัดสินใจจากผลการทดสอบสมมติฐานทางสถิติอาจจะเกิดข้อผดพลาดได้ คือเมื่อปฏิเสธ H_0 แต่ H_0 เป็นจริง หรือเมื่อยอมรับ H_0 แต่ H_0 ไม่จริง

บทนิยาม 9.1.1 ถ้าเราปฏิเสธสมมติฐานหลัก H_0 เมื่อสมมติฐานนั้นเป็นจริง แสดงได้ว่ากระทำ **ความผิดพลาดประเภทที่** 1 (Type I Error)

บทนิยาม 9.1.2 ถ้าเรายอมรับสมมติฐานหลัก H_0 เมื่อสมมติฐานนั้นไม่เป็นจริง แสดงได้ว่ากระทำ **ความผิดพลาดประเภทที่** 2 (Type II Error)

	H_0 เป็นจริง	H_0 เป็นเท็จ
ปฏิเสธ H_0	ความผิดพลาดประเภทที่ 1	-
ยอมรับ H_0	-	ความผิดพลาดประเภทที่ 2

ตัวอย่าง 9.1.3 สัดส่วนของพนักงานที่สำเร็จการศึกษาระดับปริญญาโทของบริษัทแห่งหนึ่งประมาณ $p = 0.3$ เพื่อทดสอบสมมติฐาน $H_0 : p = 0.3$ ได้ตกลงกันว่า ถ้าจำนวนของผู้สำเร็จการศึกษาระดับปริญญาโทเท่ากับ 2 ถึง 7 จากการสุ่มตัวอย่างมา 15 คน เราจะยอมรับ H_0 จงตรวจสอบว่าแต่ละข้อเกิดความผิดพลาดชนิดใด

1. ถ้าความจริง $p \neq 0.3$
   แนวคำตอบ ถ้ายอมรับ $H_0 : p = 0.3$ แต่จริง ๆ แล้ว $p \neq 0.3$ หรือ H_0 เป็นเท็จ จึงเกิดความผิดพลาดชนิดที่ 2 (ถ้าปฏิเสธ H_0 ไม่เกิดความผิดพลาด)

2. ถ้าสุ่มตัวอย่างมา 15 คน ปรากฏว่ามีผู้สำเร็จการศึกษาระดับปริญญาโทจำนวน 9 คน ถ้าความจริง $H_0 : p = 0.3$

   แนวคำตอบ ผลที่ได้จากการสุ่มตัวอย่าง คือปฏิเสธ H_0 แต่จริง ๆ แล้ว $p = 0.3$ จึงเกิดความผิดพลาดชนิดที่ 1

บทนิยาม 9.1.4 บริเวณวิกฤต (critical region) คือบริเวณที่จะทำให้เกิดการปฏิเสธ H_0 หรือยอมรับ H_1 **บริเวณการยอมรับ** (acceptance region) คือบริเวณที่จะทำให้เกิดการยอมรับ H_0 หรือปฏิเสธ H_1 และค่าที่แบ่งบริเวณทั้งสองนี้ เรียกว่า **ค่าวิกฤต** (critical value)

ตัวอย่าง 9.1.5 จากตัวอย่าง 9.1.3 จงหา บริเวณวิกฤต บริเวณการยอมรับ และค่าวิกฤต
แนวคำตอบ ให้ X เป็นจำนวนของผู้สำเร็จการศึกษาระดับปริญญาโทที่ได้จากการสุ่มตัวอย่าง 15 คน
นั่นคือ $X = 0, 1, 2, ..., 15$ จะได้ว่า

บริเวณยอมรับ	$2 \leq X \leq 7$
บริเวณวิกฤต	$X \leq 1$ หรือ $X \geq 8$
ค่าวิกฤต	7.5 และ 1.5 เป็นต้น

บทนิยาม 9.1.6 ความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 1 เรียกว่า

ระดับนัยสำคัญ (level of significance) ของการทดสอบ เขียนแทนด้วย α

นั่นคือ $\alpha = P($ ปฏิเสธ $H_0 \mid H_0$ เป็นจริง $)$

ตัวอย่าง 9.1.7 จากตัวอย่าง 9.1.3 จงหาระดับนัยสำคัญ
แนวคำตอบ ให้ X เป็นจำนวนของผู้สำเร็จการศึกษาระดับปริญญาโทที่ได้จากการสุ่มตัวอย่าง 15 คน
สมมติ H_0 เป็นจริงนั่นคือ $p = 0.3$ จะได้ว่า $X \sim \text{Bin}(n = 15, p = 0.3)$ ดังนั้น

$$\begin{aligned}
\alpha &= P(\text{ ปฏิเสธ } H_0 \mid H_0 \text{ เป็นจริง }) \\
&= P(X \leq 1 \text{ หรือ } X \geq 8 \mid p = 0.3) \\
&= 1 - P(2 \leq X \leq 7 \mid p = 0.3) \\
&= 1 - \sum_{x=2}^{7} b(x; 15, 0.3) \\
&= 1 - \left(\sum_{x=0}^{7} b(x; 15, 0.3) - \sum_{x=0}^{1} b(x; 15, 0.3) \right) \\
&= 1 - (0.9500 - 0.0353) = 0.0853
\end{aligned}$$

ดังนั้นระดับนัยสำคัญเท่ากับ 0.0853

บทนิยาม 9.1.8 β ใช้แทนความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 2 จะหาได้ถ้าทราบค่าที่แน่นอนของพารามิเตอร์ใน H_1 เรียก

$1 - \beta$ ว่า **อำนาจของการทดสอบ** (power of the test)

นั่นคือ $\beta = P($ ยอมรับ $H_0 \mid H_0$ เป็นเท็จ $)$

ตัวอย่าง 9.1.9 จากตัวอย่าง 9.1.3 จงหาอำนาจของการทดสอบ ถ้า $H_1 : p = 0.4$
แนวคำตอบ ให้ X เป็นจำนวนของผู้สำเร็จการศึกษาระดับปริญญาโทที่ได้จากการสุ่มตัวอย่าง 15 คน
สมมติ H_0 เป็นเท็จ นั่นคือยอมรับ H_1 หรือ $p = 0.4$ จะได้ว่า $X \sim \text{Bin}(n = 15, p = 0.4)$ ดังนั้น

$$\begin{aligned}
\beta &= P(\text{ ยอมรับ } H_0 \mid H_0 \text{ ไม่เป็นจริง }) \\
&= P(\text{ ยอมรับ } H_0 \mid H_1 \text{ เป็นจริง }) \\
&= P(2 \leq X \leq 7 \mid p = 0.4) \\
&= \sum_{x=2}^{7} b(x; 15, 0.4) \\
&= \sum_{x=0}^{7} b(x; 15, 0.4) - \sum_{x=0}^{1} b(x; 15, 0.4) \\
&= 0.7869 - 0.0052 = 0.7817
\end{aligned}$$

ดังนั้นจงหาอำนาจของการทดสอบ $1 - 0.7817 = 0.2183$

สรุปความผิดพลาดทั้ง 2 ประเภท

	H_0 เป็นจริง	H_0 เป็นเท็จ
ปฏิเสธ H_0	$P($ ความผิดพลาดประเภทที่ 1 $) = \alpha$	-
ยอมรับ H_0	-	$P($ ความผิดพลาดประเภทที่ 2 $) = \beta$

สรุปสมบัติความผิดพลาดได้ดังนี้

1. ความผิดพลาดทั้งสองประเภทมีความสัมพันธ์กัน ถ้าประเภทหนึ่งลดลงอีกประเภทจะเพิ่มขึ้น
2. บริเวณวิกฤต หรือ α จะลดลงได้โดยการปรับค่าวิกฤต
3. ถ้าขนาดตัวอย่างเพิ่มขึ้นจะทำให้ α และ β ลดลง
4. ถ้า H_0 ไม่จริง β จะมีค่ามากสุดเมื่อค่าที่แท้จริงใกล้กับค่าในสมมติฐานหลัก ถ้าค่าที่แท้จริงอยู่ห่างจากค่าใน H_0 มาก β จะมีค่าน้อย

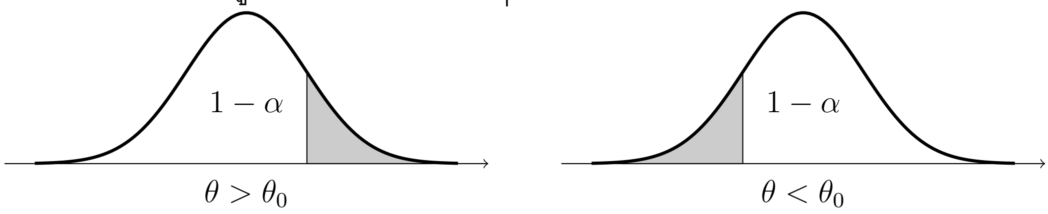
บทนิยาม 9.1.10 จะกล่าวว่า **การทดสอบมีนัยสำคัญ** (significance) เท่ากับ α ถ้าสมมติฐานหลักถูกปฏิเสธที่ระดับนัยสำคัญ α

การทดสอบข้างเดียว (One-tailed Test) เกิดขึ้นเมื่อต้องการทดสอบสมมติฐาน

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta > \theta_0 \end{aligned} \qquad \text{และ} \qquad \begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta < \theta_0 \end{aligned}$$

แสดงตัวอย่างบริเวณวิกฤตดังพื้นที่แรเงาดังกราฟ

รูปที่ 9.1: บริเวณวิกฤตในการทดสอบข้างเดียว

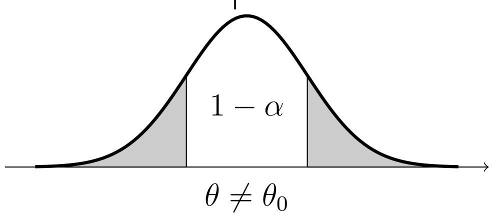


การทดสอบสองข้าง (Two-tailed Test) เกิดขึ้นเมื่อต้องการทดสอบสมมติฐาน

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned} \qquad \text{หรือ} \qquad \begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta < \theta_0 \quad \text{หรือ} \quad \theta > \theta_0 \end{aligned}$$

แสดงตัวอย่างบริเวณวิกฤตดังพื้นที่แรเงาดังกราฟ

รูปที่ 9.2: บริเวณวิกฤตในการทดสอบสองข้าง



ตัวอย่าง 9.1.11 โรงงานผลิตหลอดไฟแห่งหนึ่งอ้างว่าหลอดไฟที่ผลิตในแต่ละรุ่นจะมีชำรุดไม่เกิน 10% เพื่อทดสอบคำกล่าวอ้างนี้ แย้งกับหลอดไฟที่ผลิตจากโรงงานแห่งนี้จะมีชำรุดอยู่มากกว่า 10% จึงตั้งเกณฑ์ไว้ว่า ถ้าหลอดไฟที่สุ่มมา 20 หลอด จากหลอดไฟรุ่นหนึ่งซึ่งมี 60 หลอด มีชำรุดไม่เกิน 2 หลอด จึงยอมรับคำกล่าวอ้าง

1. จงหาความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 1

   แนวคำตอบ ตั้งสมมติฐาน $H_0 : p = 0.1$ ยอมรับเมื่อมีหลอดชำรุดไม่เกิน 2 หลอด
   ให้ X เป็นจำนวนหลอดไฟชำรุดที่ได้จากการสุ่มตัวอย่างมา $n = 20$ จาก $N = 60$ โดยหลอดไฟชำรุด 10% นั่นคือ $M = (0.1)60 = 6$ จะได้ว่า $X \sim \text{HG}(n = 20, N = 60, M = 6)$ ดังนั้น

$$\begin{aligned} P(\text{ ความผิดพลาดชนิดที่ } 1) &= P(\text{ ปฏิเสธ } H_0 \mid H_0 \text{ เป็นจริง }) \\ &= P(X > 2 \mid p = 0.1) \\ &= P(X > 2 \mid n = 20, N = 60, M = 6) \\ &= \sum_{x=3}^{6} h(x; 20, 60, 6) = 0.3136 \end{aligned}$$

   ดังนั้นความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 1 เท่ากับ 0.3136

2. จงหาความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 2
   เมื่อหลอดไฟที่ชำรุดที่แท้จริงเป็น 15%

   แนวคำตอบ ตั้งสมมติฐาน $H_1 : p = 0.15$
   ให้ X เป็นจำนวนหลอดไฟชำรุดที่ได้จากการสุ่มตัวอย่างมา $n = 20$ จาก $N = 60$ โดยหลอดไฟชำรุดแท้จริง 15% นั่นคือ $M = (0.15)60 = 9$ จะได้ว่า $X \sim \text{HG}(n = 20, N = 60, M = 9)$ ดังนั้น

$$\begin{aligned} P(\text{ ความผิดพลาดชนิดที่ } 2) &= P(\text{ ยอมรับ } H_0 \mid H_0 \text{ เป็นเท็จ }) \\ &= P(X \leq 2 \mid p = 0.15) \\ &= P(X \leq 2 \mid n = 20, N = 60, M = 9) \\ &= \sum_{x=0}^{2} h(x; 20, 60, 9) = 0.3621 \end{aligned}$$

   ดังนั้นความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 2 เท่ากับ 0.3621

ตัวอย่าง 9.1.12 ในการทดสอบประสิทธิภาพของวัคซีนป้องกันโควิด-19 ชนิดหนึ่ง สุ่มตัวอย่าง 1000 คน เพื่อให้วัคซีนดังกล่าว ถ้ามีจำนวน 660 ถึง 720 คน เกิดภูมิต้านทานป้องกันโรคโควิด-19 สรุปว่าวัคซีนใช้ได้ผล 70%

1. จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1

2. อำนาจของการทดสอบ เมื่อประสิทธิภาพของการรักษาโรคเป็น 75%

แนวคำตอบ กำหนดให้ X คือจำนวนผู้ป่วย และตั้งสมติฐานดังนี้

H_0 : $p = 0.70$ ยอมรับ $660 \leq X \leq 720$

H_1 : $p \neq 0.70$ นั่นคือ $p = 0.75$

1. จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1

   สุ่มตัวอย่าง 1000 คน วัคซีนใช้ได้ผล 70% นั่นคือ $p = 0.7$ เป็นการแจกแจงทวินาม เนื่องจากขนาดตัวอย่างมีขนาดใหญ่จะประมาณด้วยการแจกแจงปกติซึ่ง

$$\mu = 1000(0.7) = 700 \text{ และ } \sigma = \sqrt{1000(0.7)(0.3)} = \sqrt{210}$$

   จะได้ว่า

$$\begin{aligned}
\alpha &= P(\text{ ปฏิเสธ } H_0 \mid H_0 \text{ เป็นจริง })\\
&= P(X < 660 \text{ หรือ } X > 720 \mid p = 0.7)\\
&= 1 - P(660 \leq X \leq 720 \mid p = 0.7)\\
&\approx 1 - P(659.5 \leq X \leq 720.5)\\
&= 1 - P\left(\frac{659.5 - 700}{\sqrt{210}} \leq \frac{X - \mu}{\sqrt{npq}} \leq \frac{720.5 - 700}{\sqrt{210}}\right)\\
&= 1 - P(-2.79 \leq Z \leq 1.41)\\
&= 1 - P(Z \leq 1.41) + P(Z \leq -2.79)\\
&= 1 - 0.9207 + 0.0026 = 0.0819
\end{aligned}$$

   ดังนั้นความน่าจะเป็นของความผิดพลาดประเภทที่ 1 เท่ากับ 0.0819

2. อำนาจของการทดสอบ เมื่อประสิทธิภาพของการรักษาโรคเป็น 75%

   สุ่มตัวอย่าง 1000 คน วัคซีนใช้ได้ผลจริง ๆ 75% นั่นคือ $p = 0.75$ เป็นการแจกแจงทวินาม เนื่องจากขนาดตัวอย่างมีขนาดใหญ่จะประมาณด้วยการแจกแจงปกติซึ่ง

$$\mu = 1000(0.5) = 750 \text{ และ } \sigma = \sqrt{1000(0.75)(0.25)} = \sqrt{187.5}$$

   จะได้ว่า

$$\begin{aligned}
\beta &= P(\text{ ยอมรับ } H_0 \mid H_0 \text{ ไม่เป็นจริง })\\
&= P(660 \leq X \leq 720 \mid p = 0.75)\\
&\approx P(659.5 \leq X \leq 720.5)\\
&= P\left(\frac{659.5 - 750}{\sqrt{187.5}} \leq \frac{X - \mu}{\sqrt{npq}} \leq \frac{720.5 - 750}{\sqrt{187.5}}\right)\\
&= P(-6.61 \leq Z \leq -2.15)\\
&= P(Z \leq -2.15) - P(Z \leq -6.61)\\
&= 0.0158 - 0 = 0.0158
\end{aligned}$$

   ดังนั้นอำนาจของการทดสอบ เมื่อประสิทธิภาพของการรักษาโรคเป็น 75% เท่ากับ $1 - \beta = 1 - 0.0158 = 0.9842$

แบบฝึกหัด 9.1

1. ครูให้นักเรียนตอบคำถามชนิดถูกผิด 10 ข้อและตั้งสมมติฐานว่านักเรียนตอบคำถามโดยการเดาเพื่อทดสอบสมมติฐานนี้ จึงตั้งกฎไว้ดังนี้

   (ก) ถ้านักเรียนตอบถูกต้องตั้งแต่ 7 ข้อขึ้นไป แสดงว่านักเรียนไม่ได้ตอบคำถามโดยการเดา

   (ข) ถ้าตอบถูกต้องน้อยกว่า 7 ข้อ แสดงว่านักเรียนตอบคำถามโดยการเดา

   จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และหาความน่าจะเป็นของความผิดพลาดประเภทที่ 2 ถ้าความจริงความน่าจะเป็นที่จะตอบคำถามแต่ละข้อได้ถูกต้องเป็น 0.7

2. โรงงานผลิต Smart TV แห่งหนึ่งอ้างว่า Smart TV ที่ผลิตในแต่ละรุ่นจะมีชำรุดไม่เกิน 5% เพื่อทดสอบคำกล่าวอ้างนี้ แย้งกับ Smart TV ที่ผลิตจากโรงงานแห่งนี้จะมีชำรุดอยู่มากกว่า 5% จึงตั้งเกณฑ์ไว้ว่า ถ้า Smart TV ที่สุ่มมา 10 เครื่องจาก Smart TV รุ่นหนึ่งซึ่งมี 100 เครื่อง มีชำรุดไม่เกิน 2 เครื่อง จึงยอมรับคำกล่าวอ้าง

   2.1 จงหาความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 1

   2.2 จงหาความน่าจะเป็นของการกระทำความผิดพลาดประเภทที่ 2 เมื่อหลอดไฟที่ชำรุดที่แท้จริงเป็น 10%

   2.3 อำนาจของการทดสอบดังกล่าวกรณีใดจึงจะมากกว่ากันระหว่างเปอร์เซนต์ของหลอดที่ชำรุด 10% กับ 15%

3. ในโครงการสร้างดาวเทียมเพื่อการสื่อสารจำเป็นต้องบรรจุแบตเตอรี่ที่ดีให้พอเพียง โดยทั่วไปจากการทดสอบคุณภาพของแบตเตอรี่ในรุ่นก่อน ๆ พบว่าจะมีแบตเตอรี่ที่ชำรุดเล็กน้อยเพียง 0.1% และเป็นเกณฑ์พอใจ เพื่อทดสอบค่าสัดส่วน (p) ของแบตเตอรี่ที่ชำรุดของรุ่นใหม่ซึ่งมีจำนวน 10000 ลูก ว่ามากกว่า $p = 0.1\%$ แย้งกับ $p > 0.1\%$ จึงสุ่มตัวอย่างมา 100 ลูก และตรวจสอบโดยตั้งเกณฑ์ว่าพบแบตเตอรี่ที่ชำรุดมากกว่า 1 ลูก จะไม่ยอมรับแบตเตอรี่นี้ทั้งรุ่น จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1

4. สุ่มตัวอย่างประชาชน 400 คน ในเมือง ๆ หนึ่ง เพื่อทดสอบว่าพอใจการเก็บภาษีการค้าอัตราใหม่หรือไม่ ถ้าปรากฏว่ามากกว่า 220 แต่น้อยกว่า 260 คนพอใจ จะสรุปว่า 60% ของประชาชนมีความพอใจในการเก็บภาษีการค้าอัตราใหม่ จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ถ้าอัตราส่วนที่แท้จริงของประชาชนที่พอใจเป็น 60% และจงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 2 ถ้าอัตราส่วนที่แท้จริงของประชาชนที่พอใจเป็น 48%

9.2 การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์

ในการทดสอบสมมติฐานจะต้องทราบความต้องการว่าพารามิเตอร์ตัวใดแล้วจึงตั้งสมมติฐานซึ่งจาก H_1 จะทำให้ทราบว่าเป็นการทดสอบข้างเดียวหรือสองข้าง กำหนดระดับนัยสำคัญ α เลือกสถิติที่เหมาะสมเพื่อทดสอบสมมติฐาน กำหนดบริเวณวิกฤตให้สอดคล้องกับการกำหนด H_1 ค่านัยสำคัญ α และสถิติที่ใช้ จึงคำนวณค่าสถิติที่ได้จากตัวอย่าง สรุปเป็นขั้นตอนการทดสอบพารามิเตอร์ θ ได้ดังนี้

ขั้นตอนการทดสอบพารามิเตอร์

1. กำหนดสมมติฐานหลัก $H_0 : \theta = \theta_0$

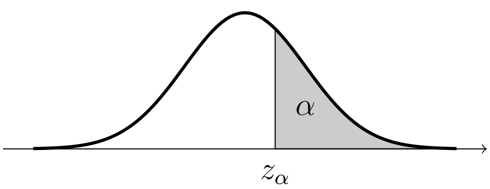
   กำหนดสมมติฐานทางเลือกว่าต้องการ

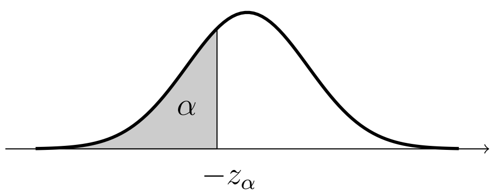
   การทดสอบข้างเดียว $H_1 : \theta > \theta_0$

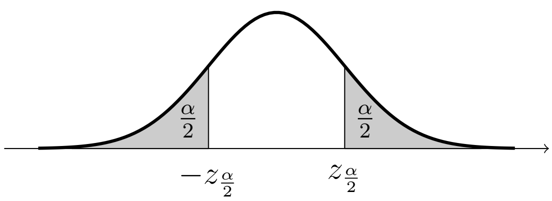
   หรือ $H_1 : \theta < \theta_0$

   การทดสอบสองข้าง $H_1 : \theta \neq \theta_0$
2. เลือกระดับนัยสำคัญ α เช่น 0.05 หรือ 0.01 เป็นต้น
3. ทำการสุ่มตัวอย่าง
4. เลือกค่าสถิติที่เหมาะ เช่นเลือกใช้ Z, T, χ^2, F หรือค่าสถิติอื่น ๆ
5. คำนวณค่าสถิติจากตัวอย่างที่เลือกใช้ในการทดสอบสมมติฐาน เรียกสถิติที่ได้ว่าค่าคำนวณ เช่น $Z_{\text{คำนวณ}}$, $T_{\text{คำนวณ}}$, $\chi^2_{\text{คำนวณ}}$, $F_{\text{คำนวณ}}$
6. เปิดตารางหาค่าวิกฤต และกำหนดบริเวณวิกฤตของการทดสอบ
7. สรุปผล

   ปฏิเสธ H_0 ถ้าค่าสถิติที่คำนวณได้จากข้อ 5 อยู่ในบริเวณที่กำหนดในข้อ 6 และยอมรับ H_1
   ยอมรับ H_0 ถ้าค่าสถิติที่คำนวณได้จากข้อ 5 ไม่อยู่ในบริเวณที่กำหนดในข้อ 6 และปฏิเสธ H_1







9.3 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากร

การทดสอบค่าเฉลี่ยของกลุ่มตัวอย่าง 1 กลุ่ม (One sample mean test) คือการทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยประชากรคือการทดสอบว่าค่าเฉลี่ยของประชากรจะเป็นไปตามที่คาดไว้หรือไม่ โดย μ_0 คือค่าเฉลี่ยของประชากรที่คาดไว้ ซึ่งมีการทดสอบได้ใน 2 ลักษณะดังนี้

H_0	สถิติที่ใช้	H_1	บริเวณวิกฤต
	One Sample Z-test		
$\mu = \mu_0$	ทราบค่า σ^2	$\mu < \mu_0$	$z < -z_\alpha$
	$Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$\mu > \mu_0$	$z > z_\alpha$
		$\mu \neq \mu_0$	$z < -z_{\frac{\alpha}{2}}$ หรือ $z > z_{\frac{\alpha}{2}}$
	One Sample T-test		
$\mu = \mu_0$	ไม่ทราบค่า σ^2 และ $n < 30$	$\mu < \mu_0$	$t < -t_\alpha$
	$T = \dfrac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$	$\mu > \mu_0$	$t > t_\alpha$
	$\nu = n - 1$	$\mu \neq \mu_0$	$t < -t_{\frac{\alpha}{2}}$ หรือ $t > t_{\frac{\alpha}{2}}$

ตัวอย่าง 9.3.1 ผู้ทำการสำรวจในปี 2561 กล่าวอ้างว่าวัยเยาวชนมีการอ่านเฉลี่ย 109 นาทีต่อวัน โดยมีส่วนเบี่ยงเบนมาตรฐาน 20 นาทีต่อวัน เพื่อทดสอบคำกล่าวอ้างดังกล่าวจึงสุ่มตัวอย่างเยาวชน 100 คน เพื่อสอบถามการอ่าน มีค่าเฉลี่ยเท่ากับ 115 นาทีต่อวัน จงทดสอบสมมติฐานว่าค่าเฉลี่ยการอ่านของเยาวชนมากกว่า 109 นาทีต่อวัน จริงหรือไม่ กำหนดระดับนัยสำคัญ 0.05

แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; \mu = 109$

$H_1 \;:\; \mu > 109$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

สุ่มตัวอย่างขนาด $n = 20$ จากประชากรปกติที่มี $\sigma = 20$ โดยมี $\bar{X} = 115$

เนื่องจากประชากรปกติและทราบค่าความแปรปรวน เลือกสถิติ Z นั่นคือ

$$Z_{\text{คำนวณ}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{115 - 109}{\frac{20}{\sqrt{100}}} = 3$$

เป็นการทดสอบข้างเดียว

เนื่องจาก $z_\alpha = z_{0.05} = 1.65$ บริเวณวิกฤตคือ $Z > 1.65$

จะเห็นว่า $Z_{\text{คำนวณ}} = 3 > 1.65$ อยู่ในบริเวณวิกฤต ดังนั้นปฏิเสธ H_0

สรุปได้ว่าค่าเฉลี่ยของเวลาในการอ่านของเยาวชนมากกว่า 109 นาทีต่อวัน ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 9.3.2 อายุการใช้งานหลอดไฟมีการแจกแจงปกติซึ่งมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 400 ชั่วโมง ผู้ผลิตอ้างว่าอายุหลอดไฟมีอายุการใช้งานเฉลี่ย 10000 ชั่วโมง เพื่อทดสอบคำกล่าวอ้างของผู้ผลิต จึงสุ่มเลือกตัวอย่างหลอดไฟมา 25 หลอด ได้ข้อมูลดังนี้

9910	9790	9650	9290	10070
10410	11230	10530	10620	10560
10590	10467	9610	10170	9750
10477	10040	9810	9890	10150
9769	9867	10140	10010	10180

จงทดสอบสมมติฐานว่าค่าเฉลี่ยของอายุการใช้งานของหลอดไฟเป็น 10000 ชั่วโมง จริงหรือไม่ กำหนดระดับนัยสำคัญ 0.10

แนวคำตอบ กำหนดสมมติฐาน

$H_0 \ : \ \mu = 10000$

$H_1 \ : \ \mu \neq 10000$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.10$

สุ่มตัวอย่างขนาด $n = 25$ จากประชากรปกติที่มี $\sigma^2 = 400$ โดยมี $\bar{X} = 10119.2$

เนื่องจากประชากรปกติและทราบค่าความแปรปรวน เลือกสถิติ Z นั่นคือ

$$Z_{\text{คำนวณ}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{10119.2 - 10000}{\frac{20}{\sqrt{25}}} = 1.49$$

เป็นการทดสอบสองข้าง

เนื่องจาก $z_{\frac{\alpha}{2}} = z_{0.05} = 1.65$ บริเวณวิกฤตคือ $Z < -1.65$ หรือ $Z > 1.65$

จะเห็นว่า $-1.65 < Z_{\text{คำนวณ}} < 1.65$ ไม่อยู่ในบริเวณวิกฤต ดังนั้นยอมรับ H_0

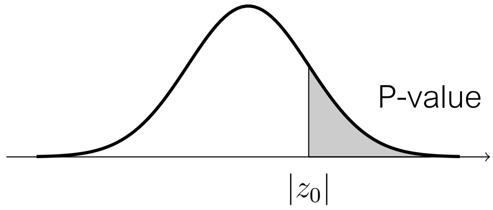
สรุปได้ว่าค่าเฉลี่ยของอายุการใช้งานของหลอดไฟเป็น 10000 ชั่วโมง ที่ระดับนัยสำคัญ 0.10

การทดสอบโดยใช้ค่า P-value ของค่าสถิติ Z

บทนิยาม 9.3.3 ค่า P-value ของค่าสถิติ z_0 หรือ P-value ของ z_0 คือพื้นที่ใต้โค้งทางด้านขวาของเส้นโค้งปกติมาตรฐานบนช่วง $(|z_0|, \infty)$ นั่นคือ

$$\text{P-value ของ } z_0 \text{ เท่ากับ } P(Z > |z_0|)$$

รูปที่ 9.3: แสดงพื้นที่ใต้เส้นโค้งปกติของ P-value

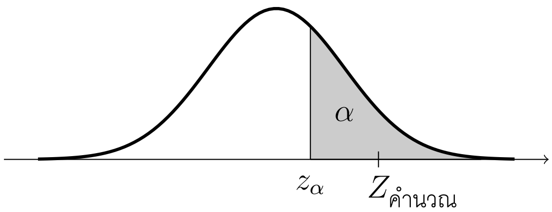


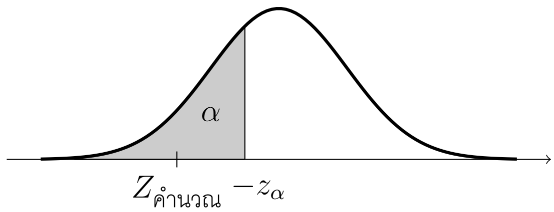
- P-value ของ $z_0 = 1.96$ มีค่าเท่ากับ $P(Z > |z_0|) = P(Z > 1.96) = 0.025$
- P-value ของ $z_0 = -2$ มีค่าเท่ากับ $P(Z > |z_0|) = P(Z > 2) = 0.02275$

ให้ $Z_{\text{คำนวณ}}$ แทนค่าสถิติที่คำนวณได้จากขั้นตอนการทดสอบพารามิเตอร์ ขั้นที่ 4

กรณีการทดสอบทางเดียว

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned} \quad \text{และ} \quad \begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu < \mu_0 \end{aligned}$$

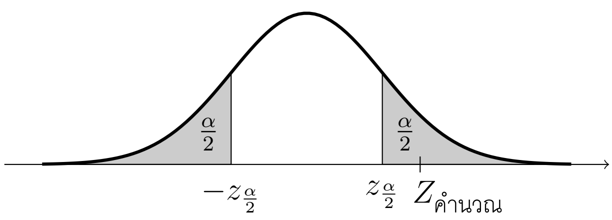




ปฏิเสธ H_0 ก็ต่อเมื่อ $Z_{\text{คำนวณ}}$ ตกในบริเวณวิกฤต หรือ P-value ของ $Z_{\text{คำนวณ}}$ น้อยกว่า α หรือ $P(Z > |Z_{\text{คำนวณ}}|) < \alpha$

กรณีการทดสอบสองข้าง

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned} \quad \text{หรือ} \quad \begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu < \mu_0 \text{ หรือ } \mu > \mu_0 \end{aligned}$$



ปฏิเสธ H_0 ก็ต่อเมื่อ $Z_{\text{คำนวณ}}$ ตกในบริเวณวิกฤต หรือ P-value ของ $Z_{\text{คำนวณ}}$ น้อยกว่า $\frac{\alpha}{2}$ หรือ $P(Z > |Z_{\text{คำนวณ}}|) < \frac{\alpha}{2}$

ตัวอย่าง 9.3.4 คะแนนสอบปรับพื้นฐานวิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งมีการแจกแจงปกติ ซึ่งมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 16 คะแนน ผู้ทำการทดสอบอ้างว่าในการสอบครั้งนี้มีคะแนนเฉลี่ย 30 คะแนน เพื่อทดสอบคำกล่าวอ้างของผู้ทำการทดสอบ จึงสุ่มเลือกตัวอย่าง 20 คนปรากฏคะแนนดังนี้

15	16	20	40	50	60	80	90	55	37
72	61	52	49	33	25	32	10	8	30

จงทดสอบสมมติฐานว่าค่าเฉลี่ยของคะแนนสอบครั้งนี้เป็น 30 คะแนน จริงหรือไม่ กำหนดระดับนัยสำคัญ 0.05

แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; \mu = 30$

$H_1 \;:\; \mu \neq 30$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

สุ่มตัวอย่างขนาด $n = 20$ จากประชากรปกติที่มี $\sigma = 16$ โดยมี $\bar{X} = 41.75$ (เครื่องคิดเลข)

เนื่องจากประชากรปกติและทราบค่าความแปรปรวน เลือกสถิติ Z นั่นคือ

$$Z_{\text{คำนวณ}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{41.75-30}{\frac{16}{\sqrt{20}}} = 3.28$$

เป็นการทดสอบสองข้าง

พิจารณา

$$\begin{aligned}\text{P-value ของ } Z_{\text{คำนวณ}} &= P(Z > |Z_{\text{คำนวณ}}|)\\ &= P(Z > 3.28) = 0.0005 < 0.025 = \frac{\alpha}{2}\end{aligned}$$

ดังนั้นปฏิเสธ H_0

สรุปได้ว่าค่าเฉลี่ยของคะแนนสอบครั้งนี้ไม่เท่ากับ 30 คะแนน ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 9.3.5 คะแนนเฉลี่ย (GPAX) ของนักเรียนของโรงเรียนแห่งหนึ่งมีการแจกแจงปกติ มีผู้กล่าวอ้างว่าคะแนนเฉลี่ย GPAX ของนักเรียนในโรงเรียนเท่ากับ 2.56 เพื่อทดสอบคำกล่าวอ้างดังกล่าว สุ่มตัวอย่างมา 100 คน คำนวณคะแนนเฉลี่ย GPAX ได้เป็น 2.48 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.36 จงทดสอบสมมติฐานว่าคะแนนเฉลี่ย GPAX เป็น 2.56 จริงหรือไม่ กำหนดระดับนัยสำคัญ 0.05

แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; \mu = 2.50$

$H_1 \;:\; \mu > 2.50$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.03$

สุ่มตัวอย่างขนาด $n = 100$ จากประชากรปกติที่มี $\sigma = 16$ โดยมี $\bar{X} = 2.48$ และ $S = 0.36$

เนื่องจากประชากรปกติ ไม่ทราบค่าความแปรปรวน และตัวอย่างมีขนาดใหญ่ เลือกสถิติ Z นั่นคือ

$$Z_{\text{คำนวณ}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{2.48-2.50}{\frac{0.36}{\sqrt{100}}} = -0.56$$

เป็นการทดสอบข้างเดียว
พิจารณา

$$\begin{aligned}\text{P-value ของ } Z_{\text{คำนวณ}} &= P(Z > |Z_{\text{คำนวณ}}|) \\ &= P(Z > 2.56) = 0.2877 > 0.03 = \alpha\end{aligned}$$

ดังนั้นยอมรับ H_0
สรุปได้ว่าคะแนนเฉลี่ย GPAX ของนักเรียนในโรงเรียนไม่สูงกว่า 2.50 ที่ระดับนัยสำคัญ 0.03

ตัวอย่าง 9.3.6 จากข้อมูลที่ผ่านมาการลงทะเบียนเรียนในเว็ปไซค์ของมหาวิทยาลัยแห่งหนึ่งโดยเฉลี่ย 50 นาที ส่วนเบี่ยงเบนมาตรฐาน 10 นาที ขณะนี้ทางมหาวิทยาลัยได้ปรับปรุงเว็ปไซค์ที่ใช้ลงทะเบียนเรียน จากการสุ่มตัวอย่าง 12 คน หาค่าเฉลี่ยที่ใช้ในการลงทะเบียนเท่ากับ 42 นาที ส่วนเบี่ยงเบนมาตรฐาน 11.9 นาที จงทดสอบสมมติฐานที่ว่าเวลาเฉลี่ยที่ใช้ในการลงทะเบียนเรียนหลังปรับปรุงเว็ปไซค์น้อยกว่า 50 นาที ใช้ระดับนัยสำคัญ 0.05 สมมติว่าเวลาที่ใช้ในการลงทะเบียนมีการแจกแจงปกติ
แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; \mu = 50$
$H_1 \;:\; \mu < 50$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$
สุ่มตัวอย่างขนาด $n = 12$ จากประชากรปกติ โดยมี $\bar{X} = 42$ และ $S = 11.9$
เนื่องจากประชากรปกติ ไม่ทราบค่าความแปรปรวน และตัวอย่างมีขนาดเล็ก เลือกสถิติ T โดยมี $\nu = 12 - 1 = 11$ นั่นคือ

$$T_{\text{คำนวณ}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{42 - 50}{\frac{11.9}{\sqrt{12}}} = -2.33$$

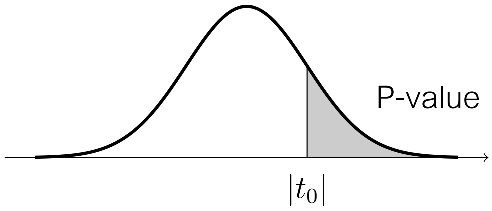
เป็นการทดสอบข้างเดียว
เนื่องจาก $t_{\alpha,\nu} = t_{0.05,11} = 1.796$ บริเวณวิกฤตคือ $T < -1.796$ หรือ $T > 1.796$
จะเห็นว่า $T_{\text{คำนวณ}} = -2.33 < -1.796$ อยู่ในบริเวณวิกฤต ดังนั้นปฏิเสธ H_0
สรุปได้ว่าเวลาเฉลี่ยที่ใช้ในการลงทะเบียนเรียนหลังปรับปรุงเว็ปไซค์น้อยกว่า 50 นาที ที่ระดับนัยสำคัญ 0.05

การทดสอบโดยใช้ค่า P-value ของค่าสถิติ T

บทนิยาม 9.3.7 กำหนดองศาเสรี ν ค่า P-value ของค่าสถิติ t_0 หรือ P-value ของ t_0 คือพื้นที่ใต้โค้งทางด้านขวาของเส้นโค้งทีบนช่วง $(|t_0|, \infty)$ นั่นคือ

$$\text{P-value ของ } t_0 \text{ เท่ากับ } P(T > |t_0|)$$

รูปที่ 9.4: แสดงพื้นที่ใต้เส้นโค้งทีของ P-value



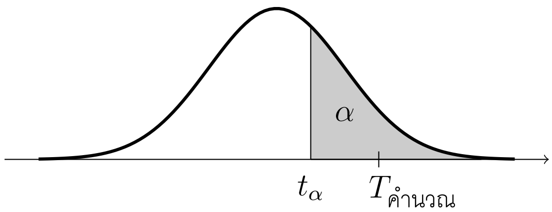
ตัวอย่างเช่น

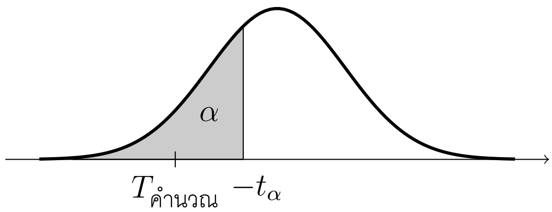
- P-value ของ $t_0 = 1.5$ มีค่าเท่ากับ $P(T > |t_0|) = P(T > 1.5) = 0.0822$ เมื่อ $\nu = 10$
- P-value ของ $t_0 = -2$ มีค่าเท่ากับ $P(T > |t_0|) = P(T > 2) = 0.0296$ เมื่อ $\nu = 20$

ให้ $T_{\text{คำนวณ}}$ แทนค่าสถิติที่คำนวณได้จากขั้นตอนการทดสอบพารามิเตอร์ ขั้นที่ 4

กรณีการทดสอบทางเดียว

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned} \qquad \text{และ} \qquad \begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu < \mu_0 \end{aligned}$$



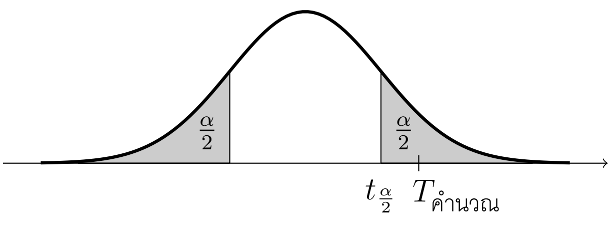


ปฏิเสธ H_0 ก็ต่อเมื่อ $T_{\text{คำนวณ}}$ ตกในบริเวณวิกฤต หรือ P-value ของ $T_{\text{คำนวณ}}$ น้อยกว่า α หรือ

$$P(T > |T_{\text{คำนวณ}}|) < \alpha$$

กรณีการทดสอบสองข้าง

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned} \qquad \text{หรือ} \qquad \begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu < \mu_0 \ \text{ หรือ } \ \mu > \mu_0 \end{aligned}$$



ปฏิเสธ H_0 ก็ต่อเมื่อ $T_{\text{คำนวณ}}$ ตกในบริเวณวิกฤต หรือ P-value ของ $T_{\text{คำนวณ}}$ น้อยกว่า $\frac{\alpha}{2}$ หรือ

$$P(T > |T_{\text{คำนวณ}}|) < \frac{\alpha}{2}$$

ตัวอย่าง 9.3.8 คะแนนสอบย่อยวิชาแคลคูลัส 1 ครั้งหนึ่งของนักศึกษาสาขาวิชาคณิตศาสตร์มีการแจกแจงปกติ ผู้สอนอ้างว่าในการสอบย่อยครั้งนี้มีคะแนนเฉลี่ย 5 คะแนน เพื่อทดสอบคำกล่าวอ้างของผู้ทำการทดสอบ จึงสุ่มเลือกตัวอย่าง 10 คนปรากฏคะแนนดังนี้

5 1 3 7 5 6 2 4 8 7

จงทดสอบสมมติฐานว่าค่าเฉลี่ยของคะแนนสอบครั้งนี้เป็น 5 คะแนน จริงหรือไม่ กำหนดระดับนัยสำคัญ 0.05

แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; \mu = 5$

$H_1 \;:\; \mu \neq 5$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

สุ่มตัวอย่างขนาด $n = 12$ จากประชากรปกติ โดยมี $\bar{X} = 4.8$ และ $S = 2.3$

เนื่องจากประชากรปกติ ไม่ทราบค่าความแปรปรวน และตัวอย่างมีขนาดเล็ก เลือกสถิติ T โดยมี $\nu = 10 - 1 = 9$ นั่นคือ

$$T_{\text{คำนวณ}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{4.8 - 5}{\frac{2.3}{\sqrt{10}}} = -0.27$$

เป็นการทดสอบสองข้าง

พิจารณา

$$\begin{aligned} \text{P-value ของ } T_{\text{คำนวณ}} &= P(T > |T_{\text{คำนวณ}}|) \\ &= P(T > 0.27) = 0.3966 > 0.025 = \frac{\alpha}{2} \end{aligned}$$

ดังนั้นยอมรับ H_0

สรุปได้ว่าค่าเฉลี่ยของคะแนนสอบครั้งนี้เป็น 5 คะแนน ที่ระดับนัยสำคัญ 0.05

การทดสอบค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มที่มีความเป็นอิสระต่อกัน (Two independent samples mean test) คือการทดสอบความแตกต่างของสองประชากรที่เป็นอิสระต่อกันทำได้ 2 ลักษณะดังนี้

H_0	สถิติที่ใช้	H_1	บริเวณวิกฤต
$\mu_1-\mu_2=d_0$	Two independent Samples Z-test ทราบค่า σ_1^2, σ_2^2 $Z=\dfrac{(\bar{X}_1-\bar{X}_2)-d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	$\mu_1-\mu_2<d_0$ $\mu_1-\mu_2>d_0$ $\mu_1-\mu_2\neq d_0$	$z<-z_\alpha$ $z>z_\alpha$ $z<-z_{\frac{\alpha}{2}}$ หรือ $z>z_{\frac{\alpha}{2}}$
$\mu_1-\mu_2=d_0$	Two independent Samples T-test กำหนด $\sigma_1^2=\sigma_2^2$ แต่ไม่ทราบค่า $T=\dfrac{(\bar{X}_1-\bar{X}_2)-d_0}{\sqrt{S_p^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ $\nu=n_1+n_2-2$ $S_p^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$\mu_1-\mu_2<d_0$ $\mu_1-\mu_2>d_0$ $\mu_1-\mu_2\neq d_0$	$t<-t_\alpha$ $t>t_\alpha$ $t<-t_{\frac{\alpha}{2}}$ หรือ $t>t_{\frac{\alpha}{2}}$
	กำหนด $\sigma_1^2\neq\sigma_2^2$ แต่ไม่ทราบค่า $n_1<30, n_2<30$ $T=\dfrac{(\bar{X}_1-\bar{X}_2)-d_0}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$ $\nu=\dfrac{\left(\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}\right)^2}{\left(\dfrac{S_1^2}{n_1}\right)^2\dfrac{1}{n_1-1}+\left(\dfrac{S_2^2}{n_2}\right)^2\dfrac{1}{n_2-1}}$	$\mu_1-\mu_2<d_0$ $\mu_1-\mu_2>d_0$ $\mu_1-\mu_2\neq d_0$	$t<-t_\alpha$ $t>t_\alpha$ $t<-t_{\frac{\alpha}{2}}$ หรือ $t>t_{\frac{\alpha}{2}}$

ตัวอย่าง 9.3.9 คะแนนสอบ ONET วิชาคณิตศาสตร์ของระดับชั้นประถมศึกษาปีที่ 6 ในปีการศึกษา 2557 และ 2558 มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 18 และ 20 คะแนน ตามลำดับ สุ่มตัวอย่างนักเรียนที่สอบในปีการศึกษา 2557 และ 2558 จำนวน 100 คน คำนวนคะแนนเฉลี่ยได้เป็น 39 และ 40 คะแนน ตามลำดับ จงทดสอบว่าคะแนนแฉลี่ยของวิชาคณิตศาสตร์ทั้งสองปีการศึกษานั้นแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.01

แนวคำตอบ จากข้อมูลที่ให้

	คะแนนปี 2557	คะแนนปี 2558
ส่วนเบี่ยงเบนมาตรฐาน (ประชากร)	$\sigma_1 = 18$	$\sigma_2 = 20$
ขนาดตัวอย่าง	$n_1 = 100$	$n_2 = 100$
ค่าเฉลี่ย (ตัวอย่าง)	$\bar{X}_1 = 39$	$\bar{X}_2 = 40$

กำหนดสมมติฐาน

$$H_0 \quad : \quad \mu_1 - \mu_2 = 0$$
$$H_1 \quad : \quad \mu_1 - \mu_2 \neq 0$$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.01$

สุ่มตัวอย่างขนาดจากประชากรปกติซึ่งความแปรปรวน เลือกสถิติ Z นั่นคือ

$$Z_{\text{คำนวณ}} = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(39 - 40) - 0}{\sqrt{\frac{18^2}{100} + \frac{20^2}{100}}} = -0.37$$

เป็นการทดสอบสองข้าง

พิจารณา

$$\begin{aligned}\text{P-value ของ } Z_{\text{คำนวณ}} &= P(Z > |Z_{\text{คำนวณ}}|) \\ &= P(Z > 0.37) = 0.3556 > 0.005 = \frac{\alpha}{2}\end{aligned}$$

ดังนั้นยอมรับ H_0

สรุปได้ว่าคะแนนแฉลี่ยของวิชาคณิตศาสตร์ทั้งสองปีการศึกษานั้นไม่แตกต่างกัน ที่ระดับนัยสำคัญ 0.01

ตัวอย่าง 9.3.10 อายุเฉลี่ย (ปี) ของประชากรในหมู่บ้าน ก และ ข มีการแจกแจงปกติโดยมีความแปรปรวนเท่ากัน ถ้าสุ่มตัวอย่างหมู่บ้านละ 10 คน ได้ข้อมูลดังนี้

หมู่บ้าน ก	5	10	30	45	32	60	15	9	18	20
หมู่บ้าน ข	7	12	32	40	29	55	12	8	14	16

จงทดสอบว่าอายุเฉลี่ยของประชากรหมู่บ้าน ก มากกว่า หมู่บ้าน ข หรือไม่ ที่ระดับนัยสำคัญ 0.05
แนวคำตอบ จากข้อมูลที่ให้

	หมู่บ้าน ก	หมู่บ้าน ข
ขนาดตัวอย่าง	$n_1 = 10$	$n_2 = 10$
ค่าเฉลี่ย (ตัวอย่าง)	$\bar{X}_1 = 24.4$	$\bar{X}_2 = 22.5$
ความแปรปรวน (ตัวอย่าง)	$S_1^2 = 305.6$	$S_2^2 = 253.39$

กำหนดสมมติฐาน

$$H_0 \;:\; \mu_1 - \mu_2 = 0$$
$$H_1 \;:\; \mu_1 - \mu_2 > 0$$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.01$
สุ่มตัวอย่างขนาดจากประชากรปกติซึ่งไม่ทราบความแปรปรวนแต่ทราบว่าเท่ากัน และตัวอย่างทั้ง 2 มีขนาดเล็ก และ

$$S_p^2 = \frac{(10-1)305.6 + (10-1)253.39}{10+10-2} = 279.495$$

เลือกสถิติ T โดยที่มีองศาเสรี $\nu = 10 + 10 - 2 = 18$ นั่นคือ

$$T_{\text{คำนวณ}} = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(24.4 - 22.5) - 0}{\sqrt{279.495\left(\frac{1}{10} + \frac{1}{10}\right)}} = 0.25$$

เป็นการทดสอบข้างเดียว
พิจารณา

$$\begin{aligned} \text{P-value ของ } T_{\text{คำนวณ}} &= P(T > |T_{\text{คำนวณ}}|) \\ &= P(T > 0.25) = 0.4027 > 0.05 = \alpha \end{aligned}$$

ดังนั้นยอมรับ H_0
สรุปได้ว่าอายุเฉลี่ยของประชากรหมู่บ้าน ก ไม่มากกว่า หมู่บ้าน ข ที่ระดับนัยสำคัญ 0.01

การทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่มไม่อิสระต่อกัน (Two dependent samples mean test **หรือ** Paired sample mean test) การทดสอบความแตกต่างของสองประชากรแบบจับคู่ทำได้ดังนี้

H_0	สถิติที่ใช้	H_1	บริเวณวิกฤต
	Two Dedependent Samples T-test		
$\mu_D = d_0$	ข้อมูลเป็นคู่ $n < 30$	$\mu_D < d_0$	$t < -t_\alpha$
	$T = \dfrac{\bar{D} - d_0}{\dfrac{S_d}{\sqrt{n}}}$	$\mu_D > d_0$	$t > t_\alpha$
	$\nu = n - 1$	$\mu_D \neq d_0$	$t < -t_{\frac{\alpha}{2}}$ หรือ $t > t_{\frac{\alpha}{2}}$

ตัวอย่าง 9.3.11 เพื่อทดสอบว่าน้ำหนักเฉลี่ยของนักศึกษาหลังเข้าเรียนในสาขาวิชาคณิตศาสตร์หลังเป็นเวลา 1 ปี มีน้ำหนักเฉลี่ยเพิ่มขึ้นอย่างมีนัยสำคัญที่ 0.05 จึงได้จดบันทึกน้ำหนักของนักศึกษาจำนวน 5 คน ปรากฏดังนี้

นักศึกษาคนที่	น้ำหนักก่อนเข้าศึกษา (กก.)	น้ำหนักหลังศึกษาเป็นเวลา 1 ปี (กก.)
1	45	46
2	55	54
3	75	78
4	62	65
5	52	54

จงสรุปผลดังกล่าวว่าเป็นไปตามข้อคาดการณ์หรือไม่ สมมติประชากรมีการแจกแจงปกติ
แนวคำตอบ กำหนดสมมติฐาน

$H_0 \ : \ \mu_D = 0$

$H_1 \ : \ \mu_D > 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$
สุ่มตัวอย่างขนาด $n = 5$ (ข้อมูลแบบคู่) จากประชากรปกติ
พิจารณาผลต่างโดยใช้น้ำหนักหลัง - น้ำหนักก่อน นั่นคือ

นักเรียนคนที่	น้ำหนักก่อน (x_i)	น้ำหนักหลัง (y_i)	$d_i = y_i - x_i$
1	45	46	1
2	55	54	-1
3	75	78	3
4	62	65	3
5	52	54	2

จะได้ว่า $\bar{D} = 1.6$ และ $S_d = 1.67$ ใช้ค่าสถิติ T โดยที่องศาเสรี $\nu = 12 - 1 = 11$ และ

$$T_{\text{คำนวณ}} = \frac{\bar{D} - d_0}{\frac{S_d}{\sqrt{n}}} = \frac{1.6 - 0}{\frac{1.67}{\sqrt{5}}} = 2.14$$

เป็นการทดสอบข้างเดียว
พิจารณา

$$\begin{aligned} \text{P-value ของ } T_{\text{คำนวณ}} &= P(T > |T_{\text{คำนวณ}}|) \\ &= P(T > 2.14) = 0.0495 < 0.05 = \alpha \end{aligned}$$

ดังนั้นปฏิเสธ H_0

สรุปได้ว่าน้ำหนักเฉลี่ยของนักศึกษาหลังเข้าเรียนในสาขาวิชาคณิตศาสตร์หลังเป็นเวลา 1 ปี มีน้ำหนักเฉลี่ยเพิ่มขึ้น ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 9.3.12 เพื่อทดสอบว่านักศึกษาสนใจเรียนวิชาแคลคูลัสมากกว่าวิชาสถิติ จากเวลาในการอ่านหนังสือหรือทำแบบฝึกหัดวิชานั้น ๆ ต่อสัปดาห์ ผู้ทดลองจึงสุ่มนักศึกษา 8 คน ปรากฏดังนี้

นักศึกษาคนที่	1	2	3	4	5	6	7	8
เวลาที่ใช้ (ชั่วโมง) วิชาแคลคูลัส	7	5	3	5	10	8	7	2
เวลาที่ใช้ (ชั่วโมง) วิชาสถิติ	7	2	1	3	6	10	5	5

จงสรุปการทดสอบข้างต้น ที่ระดับนัยสำคัญ 0.05 สมมติประชากรมีการแจกแจงปกติ
แนวคำตอบ กำหนดสมมติฐาน

$H_0 \ : \ \mu_D = 0$
$H_1 \ : \ \mu_D > 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$
สุ่มตัวอย่างขนาด $n = 8$ (ข้อมูลแบบคู่) จากประชากรปกติ
พิจารณาผลต่างโดย เวลาที่ใช้ในวิชาแคลคูลัส - เวลาที่ใช้ในวิชาสถิติ นั่นคือ

เวลาที่เข้าเรียนวิชาแคลคูลัส (x_i)	7	5	3	5	10	8	7	2
เวลาที่เข้าเรียนวิชาสถิติ (y_i)	7	2	1	3	6	10	5	5
$d_i = x_i - y_i$	0	3	2	2	4	-2	2	-3

จะได้ว่า $\bar{D} = 1$ และ $S_d = 2.45$ ใช้ค่าสถิติ T โดยที่องศาเสรี $\nu = 8 - 1 = 7$ และ

$$T_{\text{คำนวณ}} = \frac{\bar{D} - d_0}{\frac{S_d}{\sqrt{n}}} = \frac{1 - 0}{\frac{2.45}{\sqrt{8}}} = 1.15$$

เป็นการทดสอบข้างเดียว
พิจารณา

$$\begin{aligned} \text{P-value ของ } T_{\text{คำนวณ}} &= P(T > |T_{\text{คำนวณ}}|) \\ &= P(T > 1.15) = 0.1439 > 0.05 = \alpha \end{aligned}$$

ดังนั้นยอมรับ H_0

สรุปได้ว่านักศึกษาสนใจเรียนวิชาแคลคูลัสไม่มากกว่าวิชาสถิติ ที่ระดับนัยสำคัญ 0.05

แบบฝึกหัด 9.2 - 9.3

1. สุ่มตัวอย่างบุหรี่ชนิดหนึ่ง 8 มวน ปรากฏมีปริมาณเฉลี่ยของนิโคติน 18.6 มิลลิกรัม และส่วนเบี่ยงเบนมตรฐาน 2.4 มิลลิกรัม จะนับว่าเป็นการสนับสนุนคำกล่าวอ้างของบริษัทผู้ผลิตที่ว่า มีปริมาณเฉลี่ยของนิโคตินไม่เกิน 17.5 มิลลิกรัมหรือไม่ โดยใช้ระดับนัยสำคัญ 0.01 และสมมติว่าการแจกแจงของปริมาณนิโคตินเป็นการแจกแจงปกติ

2. สุ่มตัวอย่างขนาด 25 จากประชากรที่มีส่วนเบี่ยงเบนมาตรฐาน 5.2 คำนวนค่าเฉลี่ยตัวอย่างได้เป็น 81 และสุ่มตัวอย่างขนาด 36 จากประชากรที่มีส่วนเบี่ยงเบนมาตรฐาน 3.4 คำนวนค่าเฉลี่ยตัวอย่างได้เป็น 76 จงทดสอบว่าค่าเฉลี่ยของประชากรทั้งสองแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.04

3. สุ่มตัวอย่างขนาด 11 และ 14 จากประชากรปกติสองชุดซึ่งเป็นอิสระต่อกัน คำนวนค่าเฉลี่ยได้เป็น 75 และ 60 ตามลำดับ คำนวณส่วนเบี่ยงเบนมาตรฐาน 6.1 และ 53 ตามลำดับ จงทดสอบว่าค่าเฉลี่ยของประชากรทั้งสองแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.04 สมมติว่าความแปรปรวนของประชากรทั้งสองมีค่าเท่ากัน

4. เพื่อทดสอบการจัดการเรียนการสอนวิชาสถิติ Sec1 เรียนแบบ On Side แตกแต่างจาก เรียนแบบ แตกต่างจาก Sec 2 เรียนแบบ Online หรือไม่ที่ระดับนัยสำคัญ 0.05 ผู้วิจัยจึงสุ่มตัวอย่างนักศึกษามา 10 คน เพื่อสอบถามคะแนนปรากฏดังนี้

คะแนนนักศึกษา Sec1 เรียนแบบ On Side	4	8	3	5	6	6	7	9	6	5
คะแนนนักศึกษา Sec 2 เรียนแบบ Online	7	1	2	6	9	5	2	8	7	5

   จงสรุปการทดสอบข้างต้น

5. เพื่อทดสอบการสอนด้วยวิธีหนึ่ง จึงสุ่มตัวอย่างมา 5 คน โดยคะแนนการทดสอบก่อนเรียนและหลังเรียนเมื่อมีการสอนด้วยวิธีดังกล่าว ปรากฎคะแนนดังนี้

นักศึกษาคนที่	คะแนนก่อนเรียน	คะแนนหลังเรียน
1	5	6
2	4	8
3	3	8
4	2	6
5	7	9

   การสอนวิธีดังกล่าวทำให้คะแนนหลังเรียนสูงขึ้นมากกว่า 3 คะแนนหรือไม่ที่ระดับนัยสำคัญ 0.05 สมมติประชากรมีการแจกแจงปกติ

6. ผู้วิจัยต้องการทราบว่าวิธีการสอนแบบ ก และวิธีการสอนแบบ ข ให้ผลแตกต่างกันหรือไม่ จึงสุ่มนักเรียนมา 2 กลุ่ม ๆ ละ 100 คน เพื่อทำการทดลองสอน หลังจากทดลองสอนด้วยเนื้อหาเดียวกันแต่ใช้วิธีการสอนที่แตกต่างกัน แล้วทำการทดสอบ ปรากฏว่า นักเรียนกลุ่มที่สอบ

แบบ ก ได้คะแนนเฉลี่ย 82 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 20 คะแนน นักเรียนกลุ่มที่สอนแบบ ข ได้ คะแนนเฉลี่ย 75 คะแนน ส่วนเบี่ยงเบนมาตรฐาน 10 คะแนน จะสรุปได้หรือไม่ว่าวิธีการสอนทั้ง สองวิธีนี้ให้ผลเหมือนกัน ($\alpha = 0.05$)

7. นักเรียนสองกลุ่มมีวิธีการเรียนที่แตกต่างกัน กลุ่มที่หนึ่งมีนักเรียน 16 คน ใช้การเรียนตามแบบเรียน กลุ่มที่สองมีนักเรียน 18 คน ใช้การเรียนโดยใช้โปรกแกรมช่วยสอน เมื่อสิ้นภาคเรียนวัดผล สัมฤทธิ์ทางการเรียน ผลปรากฏ ดังนี้

	เรียนตามแบบเรียน	เรียนโดยใช้โปรแกรมช่วยสอน
คะแนนเฉลี่ย ($\bar{X}$)	20	27
ส่วนเบี่ยงเบนมาตรฐาน (S)	7	10

จงทดสอบว่า ผลสัมฤทธิ์ทางการเรียนของนักเรียนสองกลุ่มนี้แตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05 และมีข้อตกลงว่าความแปรปรวนทั้งสองกลุ่มนี้เท่ากัน

8. ในภาคต้นปีการศึกษานี้ นักศึกษาระดับปริญญาตรีในมหาวิทยาลัยแห่งหนึ่ง ใช้เวลาในการลงทะเบียนเฉลี่ย 40 นาที ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 นาที ในการลงทะเบียนเรียนภาคปลาย มหาวิทยาลัยจะใช้วิธีการลงทะเบียนแบบใหม่ จึงสุ่มนักศึกษามา 12 คน เพื่อศึกษาเวลาในการ ลงทะเบียน ผลปรากฏว่านักศึกษาใช้เวลาเฉลี่ย 35 นาที ส่วนเบี่ยงเบนมาตรฐาน 12 นาที จงทดสอบ การการลงทะเบียนแบบใหม่เฉลี่ยแล้วใช้เวลาน้อยกว่า 40 นาทีหรือไม่ ที่ระดับนัยสำคัญ 0.01

9. จากการศึกษาค่าแรงของลูกจ้างในโรงงาน ต้องมีค่าแรงเฉลี่ยขั้นต่ำเท่ากับ 300 บาท เก็บข้อมูลตัวอย่างลูกจ้างจำนวน 25 คน พบว่ามีค่าแรงเฉลี่ย 330 บาท ส่วนเบี่ยงเบนมาตรฐาน 50 บาท ที่ระดับนัยสำคัญ 0.05 จงตรวจสอบว่าค่าแรงเฉลี่ยของลูกจ้างในโรงงานแตกต่างจากค่าแรงขั้นต่ำหรือไม่

10. จากการสำรวจครัวเรือนในเขตเทศบาลนครเชียงใหม่จำนวน 36 ครัวเรือน พบว่าโดยเฉลี่ยครัวเรือนชมรายการโทรทัศน์ 27 ชั่วโมงต่อสัปดาห์ ส่วนเบี่ยงเบนมาตรฐาน 4 ชั่วโมง ถ้าจำนวนชั่วโมง การชมโทรทัศน์โดยเฉลี่ยของครัวเรือนทั่วประเทศเท่ากับ 25 ชั่วโมงต่อสัปดาห์ จะกล่าวสรุปได้หรือไม่ ว่าจำนวนชั่วโมงการชมรายการโทรทัศน์โดยเฉลี่ยของครัวเรือนในเขตเทศบาลนครเชียงใหม่มากกว่า จำนวนชั่วโมงการชมรายการโทรทัศน์โดยเฉลี่ยของครัวเรือนทั่วประเทศ ที่ระดับนัยสำคัญ 0.01

9.4 การทดสอบสมมติฐานเกี่ยวกับความแปรปรวนประชากร

การทดสอบความแปรปรวนของประชากรว่าเป็นไปตามที่คาดการณ์ไว้หรือไม่ และการทดสอบความแตกต่างระหว่างประชากรสองชุดที่เป็นอิสระต่อกัน ทำได้ดังนี้

H_0	สถิติที่ใช้	H_1	บริเวณวิกฤต
$\sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 < \sigma_0^2$	$x < \chi^2_{1-\alpha}$
	$\nu = n-1$	$\sigma^2 > \sigma_0^2$	$x > \chi^2_{\alpha}$
		$\sigma^2 \neq \sigma_0^2$	$x < \chi^2_{1-\alpha}$ หรือ $x > \chi^2_{\alpha}$
$\sigma_1^2 = \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$\sigma^2 < \sigma_0^2$	$f < f_{1-\alpha}$
	$\nu_1 = n_1 - 1, \nu_2 = n_2 - 1$	$\sigma^2 > \sigma_0^2$	$f > f_{\alpha}$
		$\sigma^2 \neq \sigma_0^2$	$f < f_{1-\alpha}$ หรือ $f > f_{\alpha}$

ตัวอย่าง 9.4.1 ผู้ผลิตอ้างว่าอายุการใช้งานของแบตเตอรี่มีการแจกแจงปกติ และมีส่วนเบี่ยงเบนมาตรฐาน 0.9 ปี เพื่อทดสอบคำกล่าวอ้างของผู้ผลิตจึงสุ่มตัวอย่างแบตเตอรี่ 10 ลูกมีอายุการใช้งาน (ปี) ดังนี้

5.25 3.76 5.36 3.67 6.05 3.89 3.39 6.12 6.49 6.03

จงทดสอบว่าความแปรปรวนประชากรจะมากกว่า 0.81 ที่ระดับนัยสำคัญ 0.10

แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; \sigma^2 = 0.81$

$H_1 \;:\; \sigma^2 > 0.81$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.10$

สุ่มตัวอย่างขนาด $n = 10$ จะได้ว่า $S^2 = 1.44$ และ $\nu = 10 - 1 = 9$ เลือกสถิติไคสแควร์ นั่นคือ

$$\chi^2_{\text{คำนวณ}} = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(10-1)1.44}{0.81} = 16$$

เป็นการทดสอบข้างเดียว

เนื่องจาก $\chi^2_{\alpha,\nu} = \chi^2_{0.10,9} = 14.68$ บริเวณวิกฤตคือ $\chi^2 > 14.68$

จะเห็นว่า $\chi^2_{\text{คำนวณ}} = 16 > 14.68$ อยู่ในบริเวณวิกฤต ดังนั้นปฏิเสธ H_0

สรุปได้ว่าความแปรปรวนประชากรจะมากกว่า 0.81 ที่ระดับนัยสำคัญ 0.10

ตัวอย่าง 9.4.2 จากการเก็บข้อมูลระดับคะแนน (เกรด) ของนักเรียน 2 วิชาคือ

วิชาคณิตศาสตร์ 1 :	2	3	4	2	3	
วิชาคณิตศาสตร์ 2 :	1	3	4	4	2	3

จงทดสอบว่าความแปรปรวนระดับคะแนนวิชาคณิตศาสตร์ 1 น้อยกว่าวิชาคณิตศาสตร์ 2 ที่ระดับนัยสำคัญ 0.04

แนวคำตอบ จากข้อมูลที่ให้

	วิชาคณิตศาสตร์ 1	วิชาคณิตศาสตร์ 2
ขนาดตัวอย่าง	$n_1 = 5$	$n_2 = 6$
ความแปรปรวน (ตัวอย่าง)	$S_1^2 = 0.7$	$S_2^2 = 1.37$

กำหนดสมมติฐาน

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

$H_1 \ : \ \sigma_1^2 < \sigma_2^2$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.04$

สุ่มตัวอย่างขนาดเล็กจากประชากรปกติ เลือกใช้สถิติ F โดยที่ $\nu_1 = 5 - 1 = 4$ และ $\nu_2 = 6 - 1 = 5$ จะได้ว่า

$$F_{\text{คำนวณ}} = \frac{S_1^2}{S_2^2} = \frac{0.7}{1.37} = 0.51$$

เป็นการทดสอบข้างเดียว

เนื่องจาก $f_{1-\alpha,(\nu_1,\nu_2)} = f_{0.96,(4,5)} = 0.14$ บริเวณวิกฤตคือ $f < 0.14$

จะเห็นว่า $f_{\text{คำนวณ}} = 0.51 > 0.14$ ไม่อยู่ในบริเวณวิกฤต ดังนั้นยอมรับ H_0

สรุปได้ว่าความแปรปรวนระดับคะแนนวิชาคณิตศาสตร์ 1 ไม่น้อยกว่าวิชาคณิตศาสตร์ 2 ที่ระดับนัยสำคัญ 0.04

9.5 การทดสอบสมมติฐานเกี่ยวกับสัดส่วนประชากร

การทดสอบสัดส่วนของประชากรว่าเป็นไปตามที่คาดการณ์หรือไม่ และการทดสอบผลต่างของสัดส่วนของประชากรทั้งสองชุดทำได้ดังนี้

H_0	สถิติที่ใช้	H_1	บริเวณวิกฤต
$p = p_0$	เมื่อ $n \geq 30$	$p < p_0$	$z < -z_\alpha$
	$Z = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$	$p > p_0$	$z > z_\alpha$
		$p \neq p_0$	$z < -z_{\frac{\alpha}{2}}$ หรือ $z > z_{\frac{\alpha}{2}}$
$p_1 - p_2 = 0$	เมื่อ $n_1 \geq 30$, $n_2 \geq 30$	$p_1 - p_2 < 0$	$z < -z_\alpha$
	$\hat{P} = \dfrac{X_1 + X_2}{n_1 + n_2}$	$p_1 - p_2 > 0$	$z > z_\alpha$
		$p_1 - p_2 \neq 0$	$z < -z_{\frac{\alpha}{2}}$ หรือ $z > z_{\frac{\alpha}{2}}$
	$Z = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$		
$p_1 - p_2 = d_0$	เมื่อ $n_1 \geq 30$, $n_2 \geq 30$	$p_1 - p_2 < d_0$	$z < -z_\alpha$
และ $d_0 \neq 0$		$p_1 - p_2 > d_0$	$z > z_\alpha$
	$Z = \dfrac{(\hat{P}_1 - \hat{P}_2) - d_0}{\sqrt{\dfrac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \dfrac{\hat{P}_2(1-\hat{P}_2)}{n_2}}}$	$p_1 - p_2 \neq d_0$	$z < -z_{\frac{\alpha}{2}}$ หรือ $z > z_{\frac{\alpha}{2}}$

ตัวอย่าง 9.5.1 ผู้ผลิตรายหนึ่งอ้างว่า ผลิตภัณฑ์ของเขามีข้อบกพร่อง 10% หลังจากนั้นเขาปรับปรุงกระบวนการผลิตใหม่ และคาดว่าจะสามารถลดสัดส่วนของการผลิตที่บกพร่องในเหลือน้อยกว่า 10% ในการทดลองผลิตด้วยวิธีใหม่ สุ่มตัวอย่างมา 100 ชิ้น พบว่ามีผลิตภัณฑ์บกพร่อง 5 ชิ้น จากหลักฐานนี้เพียงพอจะสรุปว่ากระบวนการผลิตใหม่ดีกว่าแบบเดิมหรือไม่ ที่ระดับนัยสำคัญ 0.05
แนวคำตอบ กำหนดสมมติฐาน

$H_0 \;:\; p = 0.1$

$H_1 \;:\; p < 0.1$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$
สุ่มตัวอย่างขนาด $n = 100$ ได้ผลิตภัณฑ์บกพร่อง $X = 5$ ชิ้น จะได้ว่า $\hat{P} = \dfrac{5}{100} = 0.05$ เลือกใช้สถิติ Z จะได้ว่า

$$Z_{\text{คำนวณ}} = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.05 - 0.1}{\sqrt{\frac{0.1(0.9)}{100}}} = -1.67$$

เป็นการทดสอบข้างเดียว
พิจารณา

$$\begin{aligned} \text{P-value ของ } Z_{\text{คำนวณ}} &= P(Z > |Z_{\text{คำนวณ}}|) \\ &= P(T > 1.67) = 0.0475 < 0.05 = \alpha \end{aligned}$$

ดังนั้นปฏิเสธ H_0
สรุปได้ว่ากระบวนการผลิตใหม่สัดส่วนของการผลิตที่บกพร่องในเหลือน้อยกว่า 10% ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 9.5.2 ในการหยั่งเสียงนักเรียนนักศึกษาเพื่อต้องการทราบความรู้สึกเกี่ยวกับการเปลี่ยนหลักสูตรครุศาสตร์จาก 5 ปี เป็น 4 ปี ว่าพอใจการเปลี่ยนแปลงหลักสูตรดังกล่าวหรือไม่ เมื่อหยั่งเสียงนักเรียนระดับชั้นมัธยมศึกษาตอนปลายจำนวน 200 คน มีคนพอใจจำนวน 120 คน และหยั่งเสียงนักศึกษาครุศาสตร์จำนวน 500 คน มี 240 คน พอใจ คุณเห็นด้วยหรือไม่ว่านักเรียนระดับชั้นมัธยมศึกษาตอนปลายพอใจมากกว่านักศึกษาครุศาสตร์ ด้วยระดับนัยสำคัญ 0.02

แนวคำตอบ จากข้อมูลที่ให้

	นักเรียน ม.ปลาย	นักศึกษาครุศาสตร์
ขนาดตัวอย่าง	$n_1 = 200$	$n_2 = 500$
จำนวนคนที่พอใจ (ตัวอย่าง)	$X_1 = 120$	$X_2 = 240$
สัดส่วนคนที่พอใจ (ตัวอย่าง)	$\hat{P}_1 = 0.60$	$\hat{P}_2 = 0.48$

กำหนดสมมติฐาน

$H_0 \ : \ p_1 - p_2 = 0$

$H_1 \ : \ p_1 - p_2 > 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.02$

สุ่มตัวอย่างจากประชากรปกติ โดยมีตัวอย่างขนาดใหญ่ เลือกใช้สถิติ Z โดยที่

$$\hat{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0.51$$

จะได้ว่า

$$Z_{\text{คำนวณ}} = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.61 - 0.48}{\sqrt{0.51(0.49)\left(\frac{1}{200} + \frac{1}{500}\right)}} = 2.87$$

เป็นการทดสอบข้างเดียว

พิจารณา

$$\begin{aligned} \text{P-value ของ } Z_{\text{คำนวณ}} &= P(Z > |Z_{\text{คำนวณ}}|) \\ &= P(T > 2.87) = 0.0021 < 0.02 = \alpha \end{aligned}$$

ดังนั้นปฏิเสธ H_0

สรุปได้ว่านักเรียนระดับชั้นมัธยมศึกษาตอนปลายพอใจมากกว่านักศึกษาครุศาสตร์ ที่ระดับนัยสำคัญ 0.02

แบบฝึกหัด 9.4 - 9.5

1. จากการสุ่มคะแนนสอบวิชาสังคมศึกษาระดับประถมศึกษาของโรงเรียนแห่งหนึ่งจำนวน 10 คน ปราฏดังนี้ 80, 85, 70, 55, 70 , 90, 75, 65, 88 และ 80 สมมติคะแนนสอบมีการแจกแจงปกติ จงทดสอบว่าความแปรปรวนของการสอบครั้งนี้มากกว่า 10 หรือไม่ที่ระดับนัยสำคัญ 0.04

2. คะแนนสอบก่อนเรียนวิชาสถิติของนักนักศึกษา 8 คน คือ 5, 4, 4, 5, 3, 6, 2, 1
คะแนนสอบหลังเรียนวิชาสถิติของนักนักศึกษา 6 คน คือ 8, 7, 6, 5, 7, 7
จงทดสอบว่าความแปรปรวนคะแนนสอบก่อนเรียนเท่ากับหลังเรียนหรือไม่ ที่ระดับนัยสำคัญ 0.05

3. จากการเลือกร้านค้าโดยการสุ่มจำนวน 15 ร้าน จากร้านค้าทั้งหมดในเขต กรุงเทพมหาครมาเก็บรวบรวมข้อมูลเกี่ยวกับราคากระทะไฟฟ้าชนิดหนึ่ง ปรากฏว่าได้ราคา เฉลี่ย 515 บาท และความแปรปรวน 350 จงทดสอบความเชื่อที่ว่า ค่าความแปรปรวนของราคา กระทะไฟฟ้าชนิดนี้ต่ำกว่า 320 ที่ระดับนัยสำคัญ 0.05

4. โรงงานผลิตหลอดภาพโทรทัศน์แห่งหนึ่งทราบว่า อายุการใช้งานของหลอดภาพมีการแจกแจงปกติ ที่มีความแปรปรวน 10,000 ชั่วโมง^2 ในการตรวจสอบคุณภาพครั้งหนึ่ง โดยการสุ่มหลอดภาพมา 20 หลอด พบว่าความแปรปรวนของอายุการใช้งานของหลอดภาพเท่ากับ 12,000 ชั่วโมง^2 ที่ระดับนัยสำคัญ 0.05 จะกล่าวได้หรือไม่ว่า ความแปรปรวนของอายุการใช้งานของหลอดภาพไม่เท่ากับ 10,000 ชั่วโมง^2

5. กระบวนการผลิตในโรงงานหนึ่ง จะผลิตสินค้าเสีย 20% ฝ่ายผลิตจึงได้ดำเนินการ ปรับปรุงกระบวนการผลิตใหม่ แล้วสุ่มตัวอย่างสินค้ามา 100 ชิ้น พบว่าเป็นสินค้าเสีย 16 ชิ้น จะเชื่อได้หรือไม่ว่า การปรับปรุงกระบวนการผลิตจะทำให้สินค้าเสียลดลง ที่ระดับนัยสำคัญ 0.05

6. นักยิงเป้าผู้หนึ่งอ้างว่าเขาสามารถยิงถูก 80% คุณจะเห็นด้วยหรือไม่ ถ้าปรากฏวันหนึ่งเขายิงเข้าเป้า 38 ครั้งจากทั้งหมด 50 ครั้ง ที่ระดับนัยสำคัญ 0.05

7. ในมหาวิทยาลัยแห่งหนึ่งประมาณว่า จำนวนนักศึกษาที่เดินทางมาเรียนโดยรถไฟฟ้าไม่เกิน 25% ค่าประมาณนี้ถูกต้องหรือไม่ ที่ระดับนัยสำคัญ 0.05 ถ้าสุ่มตัวอย่างนักศึกษา 90 คน ปรากฏว่า 28 ค นเดินทางมาเรียนโดยรถไฟฟ้า

8. ผู้อำนวยการโรงเรียนแห่งหนึ่ง ทราบข้อมูลจากการประเมินของ สมศ. ว่านักเรียนในโรงเรียนของตนเองมีความสามารถคิดวิเคราะห์ ไม่เกินร้อยละ 3 ของนักเรียนทั้งหมด ผู้อำนวยการจึงสุ่มเลือกตัวอยางนักเรียนมา 500 คน และทดสอบความสามารถในการคิดวิเคราะห์ และพบว่ามีนักเรียนที่มีความสามารถในการคิดวิเคราะห์ 22 คน ที่ระดับนัยสำคัญ 0.05 ผู้อำนวยการโรงเรียนท่านนี้จะเชื่อผลการประเมินของ สมศ.หรือไม่

9. เพื่อต้องการทราบความรู้สึกของประชาชนว่าพอใจการบริหารของผู้ว่าราชการกรุงเทพมหานครหรือไม่ จึงหยั่งเสียงจากผู้มีรายได้น้อย 100 คน มี 40 คนพอใจ จึงหยั่งเสียงจากผู้มีรายได้มาก

100 คน มี 60 คนพอใจ คุณเห็นด้วยหรือไม่ว่าอัตราส่วนของผู้มีรายได้มาก พอใจมากกว่าผู้มีรายได้น้อย กำหนดระดับนัยสำคัญ 0.05

10. สถิติการเข้าศึกษาต่อในระดับปริญญาโทของนักศึกษาส่วนใหญ่เป็นเพศหญิงประมาณ 60% จึงได้ตรวจสอบนักศึกษาระดับปริญญาโทในมหาวิทยาลัยแห่งหนึ่งซึ่งมีจำนวนทั้งหมด 320 คน พบว่าเป็นเพศหญิง 147 คน จงทดสอบว่าสัดส่วนการศึกษาต่อระดับปริญญาโทของนักศึกษาเพศ หญิงว่าแตกต่างจากเดิมหรือไม่ กำหนดระดับนัยสำคัญ 0.01

11. สอบถามผู้สูบบุหรี่ 60 คน พบว่าเป็นมะเร็งที่ปอด 45 คน และสอบถามผู้ที่ไม่สูบบุหรี่ 80คน พบว่าเป็นมะเร็งปอด 48คน จงทดสอบว่าสัดส่วนของผู้ที่สูบบุหรี่แล้วเป็นมะเร็งปอดสูงกว่าสัดส่วน ของผู้ที่ไม่สูบบุหรี่แล้วเป็นมะเร็งปอดหรือไม่ กำหนดระดับนัยสำคัญ 0.05

12. บริษัทแห่งหนึ่งมีเป้าหมายในการผลิตสินค้าให้มีคุณภาพสูงสุด และกำหนดสินค้าที่มีตำหนิไว้ไม่เกิน 10% ในแต่ละเดือน จึงสุ่มสินค้ามา 305 ชิ้น พบว่าสินค้ามีต าหนิ 25 ชิ้น จงทดสอบสัดส่วน สินค้าที่มีตำหนิว่าคุณภาพการผลิตของบริษัทมีปัญหาหรือไม่ ที่ระดับนัยสำคัญ 0.05

13. จากการสอบถามความสนใจที่จะไปทัศนะศึกษาของนักเรียนชายและนักเรียนหญิง โดยถามนักเรียนชาย 124 คน สนใจไปทัศนะศึกษา 83 คน และถามนักเรียนหญิง 103 คน สนใจไปทัศนะ ศึกษา 72 คน ต้องการทดสอบความสนใจในการไปทัศนะศึกษาของนักเรียนชายและนักเรียนหญิงมี สัดส่วนที่แตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05

14. ต้องการทราบว่า นักศึกษาชายและนักศึกษาหญิงชอบเรียนวิชาคณิตศาสตร์ในสัดส่วนที่ เท่ากันหรือไม่ จึงสุ่มตัวอย่างนักศึกษาชายและนักศึกษาหญิงที่เรียนคณิตศาสตร์มา 50 และ 45 คน ตามลำดับ พบว่านักเรียนชายกลุ่มนี้ชอบเรียนคณิตศาสตร์ 45 คน นักเรียนหญิงชอบเรียนคณิตศาสตร์ 30 คน ที่เหลือไม่ชอบเรียนคณิตศาสตร์ จงทดสอบข้อสงสัยดังกล่าว ที่ระดับนัยสำคัญ 0.01

9.6 การทดสอบความเป็นอิสระต่อกัน

การทดสอบว่าเหตุการณ์สองอย่างที่สนใจสัมพันธ์กันหรือไม่ จะใช้การทดสอบไคสแควร์เข้ามาช่วยการในการทดสอบ เช่นสนใจว่าการฉีดวัคซีน A เพื่อป้องกันโรคโควิด-19 ได้ผลจริงหรือไม่ คือดูจากจำนวนผู้ที่ได้รับวัคซีนและไม่ได้รับวัคซีน หายป่วยด้วยจำนวนเท่าใด ถ้าทดสอบแล้วว่าการรับวัคซีนกับไม่ได้รับวัคซีนไม่สัมพันธ์กัน หรือ **อิสระต่อกัน** (independent)

การทดสอบความเป็นอิสระต่อกันของเหตุการณ์ A และ B ทำได้โดยกำหนดสมมติฐาน

H_0 : เหตุการณ์ A และ B เป็นอิสระต่อกัน

H_1 : เหตุการณ์ A และ B ไม่เป็นอิสระต่อกัน

พิจารณา **ตารางการณ์จร** (contigency table) เมื่อตารางนี้ประกอบด้วยจำนวน r แถว และ c หลัก

ตารางที่ 9.1: ตัวอย่างตารางการณ์จร

	A_1	A_2	A_3	$\cdots$	A_c	
B_1	o_{11}	o_{12}	o_{13}	$\cdots$	o_{1c}	R_1
B_2	o_{21}	o_{22}	o_{23}	$\cdots$	o_{2c}	R_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_r	o_{r1}	o_{r2}	o_{r3}	$\cdots$	o_{rc}	R_r
	C_1	C_2	C_3	$\cdots$	C_c	R_r

กำหนดให้ $i \in \{1, 2, 3, ..., r\}$ และ $j \in \{1, 2, 3, ..., c\}$

โดยที่ o_{ij} คือค่าสังเกตในตำแหน่งแถวที่ i หลักที่ j

R_i คือผลรวมของค่าสังเกตแถวที่ i

C_j คือผลรวมของค่าสังเกตหลักที่ j

การหาค่าไคสแควร์คำนวณที่ใช้ในการทดสอบหาได้จาก

$$\chi^2_{\text{คำนวณ}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

เมื่อ $\nu = (r-1)(c-1)$ และ $e_{ij} = \dfrac{R_iC_j}{N}$ โดย N คือผลรวมของความถี่ทั้งหมด

ถ้า $\chi^2_{\text{คำนวณ}} > \chi^2_\alpha$ จะปฏิเสธ H_0 ที่ระดับนัยสำคัญ α

ตัวอย่าง 9.6.1 จากภาวะโรคระบาดโควิด-19 ที่รุนแรงในปี 2020 ผู้ผลิตวัคซีนรายหนึ่งได้ผลิตวัคซีน A เพื่อใช้ในการป้องกันโรคดังกล่าว ได้นำไปทดลองโดยมีอาสามัครจำนวน 750 คน ปรากฏข้อมูลตารางการณ์จร ดังต่อไปนี้

	จำนวนผู้ที่ไม่เป็นโควิด-19	จำนวนผู้ที่เป็นโควิด-19	รวม
จำนวนผู้รับการฉีดวัคซีน A	431	14	445
จำนวนผู้ไม่ได้รับการฉีดวัคซีน A	291	14	305
รวม	722	28	750

จงทดสอบสมมติฐานว่าการฉีดวัคซีน A ป้องกันโรคโควิด-19 ได้ผลหรือไม่ ที่ระดับนัยสำคัญ 0.01

แนวคำตอบ กำหนดสมมติฐาน

H_0 : การฉีดวัคซีน A ไม่มีผลต่อการป้องกัน Covid-19 (อิสระต่อกัน)

H_1 : การฉีดวัคซีน A มีผลต่อการป้องกัน Covid-19 (ไม่อิสระต่อกัน)

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

ทดสอบความเป็นอิสระต่อกันเลือกสถิติไคสแควร์ คำนวณค่าสถิติจากตัวอย่างได้ดังนี้

	ไม่เป็น Covid-19	เป็น Covid-19	รวม
รับวัคซีน A	$o_{11} = 431$	$o_{12} = 14$	$R_1 = 445$
ไม่ได้รับวัคซีน A	$o_{21} = 291$	$o_{22} = 14$	$R_2 = 305$
รวม	$C_1 = 722$	$C_2 = 28$	$N = 750$

จะได้ว่า

e_{ij}	ไม่เป็น Covid-19	เป็น Covid-19
รับวัคซีน A	$e_{11} = \frac{R_1C_1}{N} = 428.39$	$e_{12} = \frac{R_1C_2}{N} = 16.61$
ไม่ได้รับวัคซีน A	$e_{21} = \frac{R_2C_1}{N} = 293.61$	$e_{12} = \frac{R_2C_2}{N} = 11.39$

โดยมี $r = 2$ และ $c = 2$ จะเห็นว่าองศาเสรี $\nu = (2-1)(2-1) = 1$ และ

$$\begin{aligned}
\chi^2_{\text{คำนวณ}} &= \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} \\
&= \frac{(431-428.39)^2}{428.39} + \frac{(14-16.61)^2}{16.61} + \frac{(291-293.61)^2}{293.61} + \frac{(14-11.39)^2}{11.39} \\
&= 1.04
\end{aligned}$$

เนื่องจาก $\chi^2_{\alpha,\nu} = \chi^2_{0.01,1} = 6.63$ บริเวณวิกฤตคือ $\chi^2 > 6.63$

จะเห็นว่า $\chi^2_{\text{คำนวณ}} = 1.04 < 6.63$ ไม่อยู่ในบริเวณวิกฤต ดังนั้นยอมรับ H_0

สรุปได้ว่าการฉีดวัคซีน A ไม่มีผลต่อการป้องกันโรคโควิด-19 ที่ระดับนัยสำคัญ 0.01

ตัวอย่าง 9.6.2 สมมติมีการสอนต่างกัน 3 วิธี ตัวอย่างสุ่มของนักเรียนกลุ่มละ 100 คน ถูกกำหนดให้เรียนแต่ละวิธี เกรดคะแนนสอบครั้งสุดท้ายถูกนำมาเพื่อเปรียบเทียบผลการสอน ได้นำมาลงตารางการณ์จรดังต่อไปนี้

วิธีการสอน	เกรด 4	3	2	1	0	รวม
1	15	34	40	3	8	100
2	13	28	36	19	4	100
3	12	35	38	6	9	100
รวม	40	97	114	28	21	300

จงทดสอบสมมติฐานว่าสัดส่วนของเกรดต่าง ๆ ไม่ขึ้นกับวิธีการสอน ที่ระดับนัยสำคัญ 0.10

แนวคำตอบ กำหนดสมมติฐาน

H_0 : วิธีการสอนไม่มีผลต่อเกรด (อิสระต่อกัน)

H_1 : วิธีการสอนมีผลต่อเกรด (ไม่อิสระต่อกัน)

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

คำนวณค่าสถิติจากตัวอย่าง

วิธีการสอน	เกรด 4	3	2	1	0	รวม
1	$o_{11} = 15$	$o_{12} = 34$	$o_{13} = 40$	$o_{14} = 3$	$o_{15} = 8$	$R_1 = 100$
2	$o_{21} = 13$	$o_{22} = 28$	$o_{23} = 36$	$o_{24} = 19$	$o_{25} = 4$	$R_2 = 100$
3	$o_{31} = 12$	$o_{32} = 35$	$o_{33} = 38$	$o_{34} = 6$	$o_{35} = 9$	$R_3 = 100$
รวม	$C_1 = 40$	$C_2 = 97$	$C_3 = 114$	$C_4 = 28$	$C_5 = 21$	$N = 300$

จะได้ว่า

e_{ij}	4	3	2	1	0
1	$e_{11} = \frac{40}{3}$	$e_{12} = \frac{97}{3}$	$e_{13} = 38$	$e_{14} = \frac{28}{3}$	$e_{15} = 7$
2	$e_{21} = \frac{40}{3}$	$e_{22} = \frac{97}{3}$	$e_{23} = 38$	$e_{24} = \frac{28}{3}$	$e_{25} = 3$
3	$e_{31} = \frac{40}{3}$	$e_{32} = \frac{97}{3}$	$e_{33} = 38$	$e_{34} = \frac{28}{3}$	$e_{35} = 3$

โดยมี $r = 3$ และ $c = 5$ จะเห็นว่า $\nu = (3-1)(5-1) = 8$ และ

$$\chi^2_{\text{คำนวณ}} = \sum_{i=1}^{2}\sum_{j=1}^{5}\frac{(o_{ij}-e_{ij})^2}{e_{ij}} = 18.95$$

เนื่องจาก $\chi^2_{\alpha,\nu} = \chi^2_{0.10,8} = 13.36$ บริเวณวิกฤตคือ $\chi^2 > 13.36$

จะเห็นว่า $\chi^2_{\text{คำนวณ}} = 18.95 > 13.36$ อยู่ในบริเวณวิกฤต ดังนั้นปฏิเสธ H_0

สรุปได้ว่าวิธีการสอนมีผลต่อเกรด ที่ระดับนัยสำคัญ 0.10

แบบฝึกหัด 9.6

1. ในการศึกษาความสัมพันธ์ระหว่างการนับถือศาสนาและถิ่นที่อยู่อาศัย เลือกคนมา 2 กลุ่ม โดยวิธีสุ่ม กลุ่มหนึ่งอาศัยอยู่ในเขตเมืองหลวง อีกกลุ่มอาศัยอยู่ในเขตชานเมือง ปรากฏจำนวนผู้นับถือศาสนาต่าง ๆ ดังนี้

	พุทธ	คริส	อิสลาม	อื่น ๆ	รวม
เขตเมืองหลวง	68	20	10	2	100
เขตชานเมือง	92	5	2	1	100
รวม	160	25	12	3	200

จงทดสอบสมมติฐานว่าการนับถือศาสนาและถิ่นที่อยู่อาศัยมีความสัมพันธ์กันหรือไม่ โดยใช้ระดับนัยสำคัญ 0.10

2. สมมติมีการจัดการเรียนรู้ต่างกัน 2 วิธี คือแบบ Oline และ Onside ตัวอย่างสุ่มของนักนักศึกษากลุ่มละ 40 คน ถูกกำหนดให้เรียนคนละแบบ เกรดคะแนนสอบครั้งสุดท้ายถูกนำมาเพื่อเปรียบเทียบผลการจัดการเรียนรู้ ได้นำมาลงตารางการณ์จรดังต่อไปนี้

การจัดการเรียนรู้	เกรด					รวม
	A	B	C	D	F	
Online	5	10	18	5	2	40
Onside	7	20	8	4	1	40
รวม	12	30	26	9	3	80

จงทดสอบสมมติฐานว่าสัดส่วนของเกรดต่าง ๆ ไม่ขึ้นกับการจัดการเรียนรู้ ที่ระดับนัยสำคัญ 0.05

3. จากการสำรวจผู้ดื่มกาแฟ 311 คน ปรากฏผลตารางการณ์จรดังต่อไปนี้

อายุ (ปี)	จำนวนกาแฟ (ถ้วยต่อวัน)				รวม
	0	1 - 2	3 - 4	มากกว่า 4	
ต่ำกว่า 20	26	30	6	5	67
20 -29	21	30	11	7	69
30-39	25	25	45	9	104
40 ปีขึ้นไป	11	32	17	11	71
รวม	40	83	117	32	311

จงทดสอบว่าจำนวนการดื่มกาแฟกับอายุของผู้ดื่มเป็นอิสระต่อกันหรือไม่ ที่ระดับนัยสำคัญ 0.05

สรุป

การทดสอบสมมติฐานของพารามิเตอร์ประกอบด้วย 7 ขั้นตอน ดังนี้

1. กำหนดสมมติฐานหลัก $H_0 : \theta = \theta_0$

   กำหนดสมมติฐานทางเลือกว่าต้องการ

   การทดสอบข้างเดียว $H_1 : \theta > \theta_0$

   หรือ $H_1 : \theta < \theta_0$

   การทดสอบสองข้าง $H_1 : \theta \neq \theta_0$

2. เลือกระดับนัยสำคัญ α เช่น 0.05 หรือ 0.01 เป็นต้น
3. ทำการสุ่มตัวอย่าง
4. เลือกค่าสถิติที่เหมาะ เช่นเลือกใช้ Z, T, χ^2, F หรือค่าสถิติอื่น ๆ
5. คำนวณค่าสถิติจากตัวอย่างที่เลือกใช้ในการทดสอบสมมติฐาน เรียกสถิติที่ได้ว่าค่าคำนวณ เช่น $Z_{\text{คำนวณ}}$, $T_{\text{คำนวณ}}$, $\chi^2_{\text{คำนวณ}}$, $F_{\text{คำนวณ}}$
6. เปิดตารางหาค่าวิกฤต และกำหนดบริเวณวิกฤตของการทดสอบ
7. สรุปผล
   ปฏิเสธ H_0 ถ้าค่าสถิติที่คำนวณได้จากข้อ 5 อยู่ในบริเวณที่กำหนดในข้อ 6 และยอมรับ H_1
   ยอมรับ H_0 ถ้าค่าสถิติที่คำนวณได้จากข้อ 5 ไม่อยู่ในบริเวณที่กำหนดในข้อ 6 และปฏิเสธ H_1

แบบฝึกหัดบทที่ 9

1. ในการทดลองประสิทธิภาพของยาชนิดใหม่ สุ่มตัวอย่างคนไข้ 400 คน ให้การรักษาโดยใช้ยานี้ ถ้ามีคนไข้มากกว่า 300 คน แต่น้อยกว่า 340 คน หายป่วย จะสรุปว่ายาชนิดใหม่นี้ใช้ได้ผล 80% จงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และจงหาความน่าจะเป็นของความผิดพลาดประเภทที่ 2 ถ้าประสิทธิภาพของการรักษาโรคของยาชนิดนี้เป็น 70%
2. มีคนกล่าวอ้างว่าใน 1 ปี โดยเฉลี่ยรถแต่ละคันจะเดินทางมากกว่า 30000 กิโลเมตร เพื่อทดสอบว่าคำพูดดังกล่าว จึงสุ่มตัวอย่างเจ้าของรถมา 100 คัน และขอให้บันทึกระยะทางการเดินทางตลอดปี ถ้าจากการสุ่มตัวอย่างนี้ได้ระยะทางเฉลี่ย 35000 กิโลเมตร และส่วนเบี่ยงเบนมาตรฐาน 2500 กิโลเมตร จะสรุปผลอย่างไร เมื่อกำหนดระดับนัยสำคัญ 0.01
3. คะแนนการสอบเขียนคำราชาศัพท์ก่อนเรียนและหลังเรียนของนักเรียนเป็นดังนี้

นักเรียนคนที่	1	2	3	4	5	6	7	8	9
คำราชาศัพท์ที่เขียนถูกก่อนเรียน	7	9	4	15	6	3	9	5	6
คำราชาศัพท์ที่เขียนถูกหลังเรียน	5	15	7	11	4	7	8	10	6

จงทดสอบว่าคำราชาศัพท์ที่เขียนได้ถูกต้องก่อนเรียนน้อยกว่าหลังเรียนหรือไม่ ที่ระดับ $\alpha = 0.05$

4. ครูคนหนึ่งเชื่อว่าวิธีการสอนแบบใหม่จะให้ผลสัมฤทธิ์ทางการเรียนสูงกว่าการสอนแบบเก่า ซึ่ง คะแนนเฉลี่ยจากการสอนแบบเก่าเท่ากับ 75 คะแนน จึงสุ่มนักเรียนมา 5 คนเพื่อมาทำการสอน แบบใหม่ และทำการทดสอบ พบว่านักเรียนได้คะแนน ดังนี้ 75 85 80 95 และ 90 ท่านจะ ยอมรับหรือไม่ว่าความเชื่อของครูคนนี้เป็นจริง ($\alpha = 0.01$)

5. หากระดับ urine chloride ในทารกปกติ (full-term infants) มีค่าเฉลี่ยเป็น 210 mEq/24 ชม. และส่วนเบี่ยงเบนมาตรฐานเป็น 20 mEq/24 ชม. นักวิจัยผู้หนึ่งทำการสุ่มตัวอย่างทารกคลอดก่อนกำหนดจำนวน 25 ราย พบว่ามีค่าเฉลี่ย urine chloride เป็น 170 mEq/24 ชม. จะสามารถสรุปได้หรือไม่ว่าระดับ urine chloride ในทารกคลอดก่อนกำหนดมีค่าต่ำกว่า urine chloride ในทารกปกติ ที่ระดับนัยสำคัญ 0.01

6. ตามหลักพันธุกรรมกล่าวว่า พืชที่ได้มาจากการผสมพันธุ์ระหว่างเมล็ดพันธุ์ ก และ ข จะมีลักษณะเตี้ย 80% ได้มีผู้ทดลองผสมพันธุ์พืช 2 ชนิดนี้ คือ ก และ ข พบว่า พืชที่ได้จากการผสมพันธุ์นี้ 200 ต้น ปรากฏว่ามี 64 ต้น ที่ไม่ใช่ต้นเตี้ย จงทดสอบว่า การทดลองนี้จะสอดคล้องตาม ทฤษฎีดังกล่าวข้างต้นนี้หรือไม่ โดยใช้ระดับนัยสำคัญ 0.10

7. คะแนนสอบวิชาทฤษฎีจำนวนมีการแจกแจงปกติ อาจารย์ผู้สอนกล่าวอ้างว่าคะแนนเฉลี่ยในการสอบครั้งนี้เท่ากับ 65 คะแนน เพื่อทดสอบคำกล่าวอ้างดังกล่าว จึงสุ่มเลือกตัวอย่าง 36 คนปรากฏคะแนนดังนี้

30	45	50	45	60	70	75	80	59	55
49	60	65	57	55	90	35	90	78	46
58	62	63	75	73	75	79	56	50	82
60	63	39	45	15	50				

จงทดสอบสมมติฐานว่าค่าเฉลี่ยของคะแนนสอบครั้งนี้เป็น 65 คะแนน จริงหรือไม่ กำหนดระดับนัยสำคัญ 0.05

8. เพื่อต้องการทราบความรู้สึกของนักศึกษาว่าพอใจต่อให้บริการของ One Stop Service หรือไม่ จึงหยั่งเสียงจากนักศึกษาระดับปริญญาตรี 120 คน มี 96 คนพอใจ และหยั่งเสียงจากนักศึกษาที่สูงกว่าระดับปริญญาตรี 100 คน มี 73 คนพอใจ จึงสรุปได้ว่านักศึกษาระดับปริญญาตรีพอใจต่อให้บริการมากกว่านักศึกษาที่สูงกว่าระดับปริญญาตรีเกินกว่า 5% คุณเห็นด้วยหรือไม่ กำหนดระดับนัยสำคัญ 0.10

9. เพื่อทดสอบโปรแกรมลดน้ำหนักชื่อ **โคตรสลิม** ของคลีนิคเสริมความงามแห่งหนึ่งที่กล่าวอ้างว่าหลังจากลูกค้าผู้ที่เข้าโปรแกรมโคตรสลิมจะลดน้ำหนักได้อย่างน้อย 10 กิโลกรัม จึงเลือกตัวอย่างโดยการสุ่มอย่างแบบง่าย (Simple random sampling) จำนวน 10 คน จากลูกค้าที่เข้ารับบริการในโปรแกรมโคตรสลิม ได้ข้อมูลดังนี้ (หน่วยกิโลกรัม)

ลูกค้าคนที่	น้ำหนักก่อนเข้าโปรแกรม	น้ำหนักหลังเข้าโปรแกรม
1	70	60
2	85	80
3	90	80
4	100	85
5	110	80
6	95	93
7	80	75
8	95	82
9	88	80
10	100	95

การกล่าวอ้างเกี่ยวกับโปรแกรมลดน้ำหนักโคตรสลิมของคลีนิคเสริมความงามแห่งนี้ เชื่อถือได้หรือไม่ ที่ระดับนัยสำคัญ 0.05 สมมติประชากรมีการแจกแจงปกติ

10. จากการสำรวจความชื่นชอบนโยบายของรัฐบาล ของกกลุ่มที่อาศัยในกรุงเทพมหานคร และกลุ่มที่ไม่ได้อาศัยในกรุงเทพมหานคร กลุ่มละ 100 คน สอบถามเพียง 3 ระดับเท่านั้น คือ ชอบ ไม่ชอบ และเฉย ๆ ได้นำมาลงตารางการณ์จรดังต่อไปนี้

	ระดับความชื่นชอบ			
เขตที่พักอาศัย	ชอบ	เฉย ๆ	ไม่ชอบ	รวม
อยู่กรุงเทพมหานคร	30	20	50	100
ไม่อยู่กรุงเทพมหานคร	50	10	40	100
รวม	80	30	90	200

จงทดสอบสมมติฐานว่าเขตที่พักอาศัยขึ้นกับระดับความชื่อชอบนโยบายของรัฐบาลหรือไม่ที่ระดับนัยสำคัญ 0.05

บทที่ 10

การถดถอยเชิงเส้นและสหสัมพันธ์

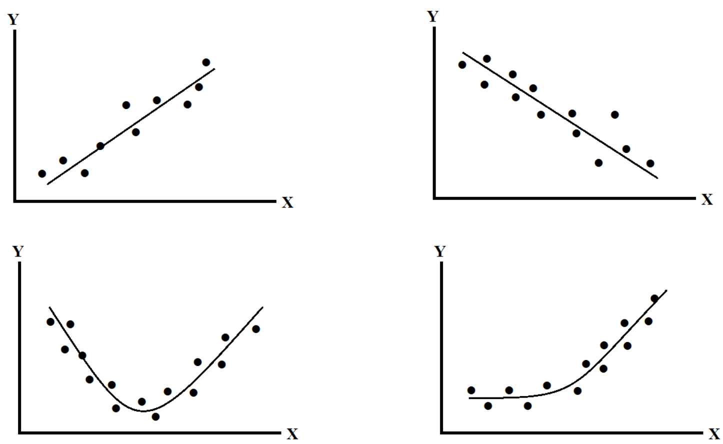
ในทางสถิติตัวแปรที่เกิดจากการทดลองหรือสำรวจ อาจได้จากการวัดมากกว่าหนึ่งชนิด เช่น ส่วนสูง อายุ น้ำหนัก เป็นต้น เมื่อพบว่าตัวแปรที่ได้จากทดลองหรือสำรวจมี 2 ตัวขึ้นไปจึงเกิดความสนใจว่า ตัวแปรเหล่านั้นมีความสัมพันธ์กันหรือไม่ ถ้าสัมพันธ์กันจะอยู่ในรูปใด การศึกษานี้เราเรียกว่า **การวิเคราะห์การถดถอย** (Regression Analysis) ในบทนี้เราจะกล่าวเฉพาะตัวแปร 2 ตัวแปรที่เกี่ยวกับ **การถดถอยเชิงเส้นเชิงเดียว** (Simple linear regression)

ในการสร้างความสัมพันธ์ของตัวแปรที่ทราบค่าเรียกว่าตัวแปรอิสระ (Indepedent variation) หรือเรียกว่าตัวพยากรณ์ (Predictor) นิยมใช้สัญลักษณ์ X ซึ่งสามารถนำมาพยากรณ์ค่าของตัวแปรอีกตัวหนึ่งได้ เรียกว่า ตัวแปรตาม (Depedent variation) ใช้สัญลักษณ์ Y ในการสร้างความสัมพันธ์ของตัวแปรเหล่านั้น ขั้นแรกคือการรวบรวมข้อมูลที่ต้องการพิจารณา เช่น

ข้อมูลชุด X ประกอบด้วย $x_1, x_2, ..., x_n$ และ ข้อมูลชุด Y ประกอบด้วย $y_1, y_2, ..., y_n$

ขั้นต่อไปนำจุด $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ ไปเขียนกราฟจะได้การกระจายของของจุดบนกราฟเรียกว่า แผนภาพกระจาย (Scatter diagram) จากแผนภาพการกระจายจะมองเห็น แนวโน้ม (Trend) ซึ่งอาจจะเส้นตรงหรือเส้นโค้งแบบต่าง ๆ ถ้าเป็นเส้นตรงเรียกว่า ความสัมพันธ์เชิงเส้น (Linear relationship) ถ้าไม่เป็นเส้นตรงเรียกว่า ความสัมพันธ์ไม่เชิงเส้น (Non-linear relationship)

รูปที่ 10.1: กราฟความสัมพันธ์รูปแบบต่าง ๆ



สำหรับความสัมพันธ์ที่ไม่เชิงเส้นเราจะหาสมการเพื่อประมาณเส้นโค้งที่เหมาะสมกับข้อมูลที่กำหนดให้ โดยทั่วไปเรียกว่า การปรับเส้นโค้ง (Curve fitting)

10.1 การถดถอยเชิงเส้น

การพยากรณ์ตัวเลขในอนาคตที่อ่านค่าจากแนวโน้มที่ได้จากข้อมูลในอดีตจะเป็นประโยชน์ในด้านต่าง ๆ แม้ว่าค่าจะผิดพลาดไปจากค่าจริงที่ควรจะเป็น ก็ยังดีกว่าเราไม่ทราบอะไรเลย

ให้ $Y \mid x$ เป็นตัวแปรสุ่ม Y เมื่อกำหนดค่า x ซึ่งคงที่และมีฟังก์ชันการแจกแจงความน่าจะเป็น $f_{Y|X}(y \mid x)$ จะเห็นว่าถ้า $x = x_i$ แล้ว $Y \mid x_i$ เขียนแทนด้วย Y_i นั่นคือ

$$\{Y_i : i = 1, 2, ..., n\} \quad \text{เป็นเซตของตัวแปรสุ่มที่อิสระต่อกันและมีการแจกแจงปกติ}$$

จะเรียกเส้นที่โยงค่า $E(Y \mid x_i) = \mu_{Y|x_i}$ ว่า เส้นโค้งถดถอย (regression curve) ถ้าค่าเฉลี่ยที่ได้เป็นเส้นตรงจะเรียก การถดถอยเชิงเส้น (linear regression) โดยทั่วไปการถดถอยของ 2 ตัวแปรสุ่มเรียกว่า การถดถอยเชิงเดี่ยว (simple regression) กรณีที่มากกวา 2 ตัวแปรเรียก การถดถอยพหุคูณ (multiple regression)

บทนิยาม 10.1.1 ให้ X และ Y เป็นตัวแปรสุ่ม

1. $E(Y \mid x)$ เรียกว่า **เส้นถดถอยของ Y เทียบกับ** x (regression line of Y on x)
2. $E(X \mid y)$ เรียกว่า **เส้นถดถอยของ X เทียบกับ** y (regression line of X on y)

ตัวอย่าง 10.1.2 ให้ตัวแปรสุ่ม X และ Y มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y) = \begin{cases} 2 & \text{เมื่อ } 0 < x < 1 \text{ และ } 0 < x < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาสมการเส้นถดถอยของ Y เทียบกับ x
แนวคำตอบ เนื่องจาก

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^1 2\,dy = 2(1-x)$$

ฉะนั้น

$$\begin{aligned} E(Y \mid x) &= \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\,dy = \int_x^1 y \cdot \frac{f(x,y)}{f_X(x)}\,dy \\ &= \int_x^1 y \cdot \frac{2}{2(1-x)}\,dy = \left[\frac{y^2}{2(1-x)}\right]_{y=x}^{y=1} \\ &= \frac{1-x^2}{2(1-x)} = \frac{1}{2}(x+1) \end{aligned}$$

ดังนั้นสมการเส้นถดถอยของ Y เทียบกับ x คือ $y = \frac{1}{2}(x+1)$

สมการของ**การถดถอยเชิงเส้น**ของประชากรเป็น

$$\mu_{Y|x} = \alpha + \beta x$$

เมื่อ α และ β เป็นพารามิเตอร์เรียกว่า **สัมประสิทธิ์การถดถอย** (regression coefficients) ให้สมการการถดถอยของตัวอย่างเป็น

$$\hat{y} = a + bx$$

เมื่อ a และ b เป็นค่าประมาณของ α และ β ตามลำดับ และ $\hat{y}$ เป็นค่าประมาณของ $\mu_{Y|x}$

10.2 การถดถอยเชิงเส้นเชิงเดียว

ให้ข้อมูลอยู่ในรูป $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ สำหรับประชากรจะได้ว่า

$$Y_i = \mu_{Y|x_i} + E_i = \alpha + \beta x_i + E_i$$

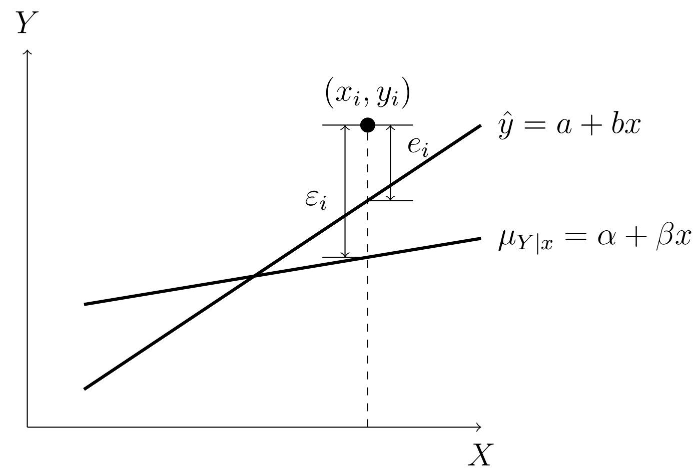
เมื่อตัวแปรสุ่ม E_i คือความผิดพลาดที่มีค่าเฉลี่ยเป็นศูนย์ แต่ละคู่ (x_i, y_i) จะสอดคล้องสมการ

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

ในทำนองเดียวกันสมการถดถอยเชิงเส้นที่ประมาณจากตัวอย่างเป็น $\hat{y} = a + bx$ จะได้ว่าแต่ละคู่สอดคล้องสมการ

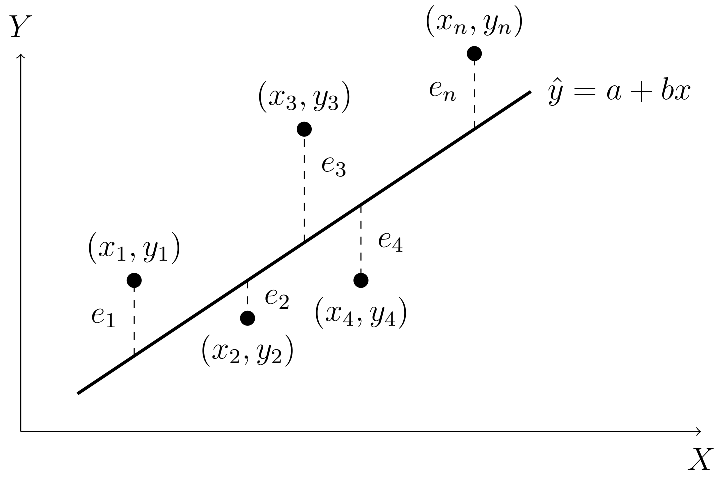
$$y_i = a + bx_i + e_i$$

เรียก e_i ว่า **ส่วนเหลือ** (residual)



จากนี้เราจะประมาณค่าสัมประสิทธิ์ของการถดถอยด้วย **วิธีกำลังสองน้อยสุด** (Least Squares Method) จะได้เส้นถดถอยของข้อมูลที่ดีเพราะวิธีนี้ เป็นการเฉลี่ยค่าความคลาดเคลื่อนของข้อมูลให้มีค่าเหลือน้อยที่สุด จะได้ว่า

$$e_i = y_i - \hat{y}_i \qquad \text{เมื่อ } i = 1, 2, ..., n$$



ผลบวกกำลังสองของความคลาดเคลื่อน (Sum of squares of error) เขียนแทนด้วย SSE คือ

$$\text{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

วิธีกำลังสองน้อยสุดคือสนใจ SEE มีค่าน้อยที่สุด หรือต้องการผลบวกกำลังสองของความคลาดเคลื่อนที่มีค่าน้อยที่สุด

ค่าสังเกตของข้อมูลในรูปคู่อันดับ $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ มีความสัมพันธ์เชิงเส้น

$$\hat{y} = a + bx$$

ให้ $\hat{y}_i = a + bx_i$ ผลบวกกำลังสองของความคลาดเคลื่อนคือ

$$\text{SSE} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

ต่อไปเราจะหาค่าน้อยสุดโดยใช้ทฤษฎีในแคลคูลัส โดยพิจารณา ฟังก์ชัน SSE(a, b) แล้ว

$$\frac{\partial(\text{SSE})}{\partial a} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-1) = 0$$
$$\frac{\partial(\text{SSE})}{\partial b} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) = 0$$

จัดรูปสมการจะได้

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i \tag{10.1}$$
$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \tag{10.2}$$

ดังนั้น (a, b) ที่สอดคล้องสมการ (10.1) และ (10.2) จะให้ค่า SSE ที่น้อยที่สุด

เรียกสมการ (10.1) และ (10.2) ว่า **สมการปกติ (Normal equation)** ของสมการ $\hat{y} = a + bx$ และ b เป็นความชันของสมการนี้ เรียกว่า สัมประสิทธิ์การถดถอย

จากสมการ (10.1) และ (10.2) เราจะได้

$$b = \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) - n\sum_{i=1}^{n} x_i y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \text{และ} \quad a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}$$

จะเห็นว่า $a = \bar{y} - b\bar{x}$ นั่นแสดงว่าจุด $(\bar{x}, \bar{y})$ อยู่บนเส้นถดถอย $\hat{y} = a + bx$

ในทำนองเดียวกันการหาสัมประสิทธิ์ของเส้นถดถอย $\hat{x} = c + dy$ จะได้ว่า

$$d = \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) - n\sum_{i=1}^{n} x_i y_i}{\left(\sum_{i=1}^{n} y_i\right)^2 - n\sum_{i=1}^{n} y_i^2} \quad \text{และ} \quad c = \bar{x} - d\bar{y}$$

หมายเหตุ เพื่อให้การเขียนกระชับมากขึ้นเราจะเขียน $\sum_{i=1}^{n} X_i$ แทนด้วย $\sum X$

ตัวอย่าง 10.2.1 ในการหาความสัมพันธ์เชิงฟังก์ชันระหว่างคะแนนสอบวิชาคณิตศาสตร์ (X) และวิชาฟิสิกส์ (Y) ของนักเรียน 100 คนของโรงเรียนแห่งหนึ่ง ได้พจน์ต่าง ๆ ที่ใช้ในการคำนวณค่าคงตัวสมการถดถอย $\hat{y} = a + bx$ ดังนี้

$$\sum_{i=1}^{100} x_i = \sum_{i=1}^{100} y_i = 1000, \quad \sum_{i=1}^{100} x_i y_i = 2000 \quad \text{และ} \quad \sum_{i=1}^{100} x_i^2 = 5000$$

ถ้าคะแนนสอบวิชาคณิตศาสตร์ของนายสมชายเท่ากับ 15 คะแนน จงหาคาดคะเนคะแนนสอบวิชาฟิสิกส์ของนายสมชาย

แนวคำตอบ พิจารณา

$$b = \frac{\left(\sum x\right)\left(\sum y\right) - 100\sum xy}{\left(\sum x\right)^2 - 100\sum x^2} = \frac{1000(1000) - 100(2000)}{1000^2 - 100(5000)} = 1.6$$

$$a = \frac{\sum y - b\sum x}{100} = \frac{1000 - 1.6(1000)}{100} = -6$$

สมการของเส้นถดถอย y เทียบกับ x คือ $\hat{y} = 1.6x - 6$ สำหรับ $x = 15$ จะได้ว่า

$$\hat{y} = 1.6(15) - 6 = 18$$

ดังนั้นสมชายจะมีคะแนนสอบวิชาฟิสิกส์ 18 คะแนน ถ้าคะแนนสอบวิชาคณิตศาสตร์เท่ากับ 15 คะแนน

ตัวอย่าง 10.2.2 กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	2	3	4	5
y	1	3	2	4	5

จงหาสมการของเส้นถดถอย y เทียบกับ x พร้อมเขียนกราฟประกอบ
แนวคำตอบ พิจารณาตาราง

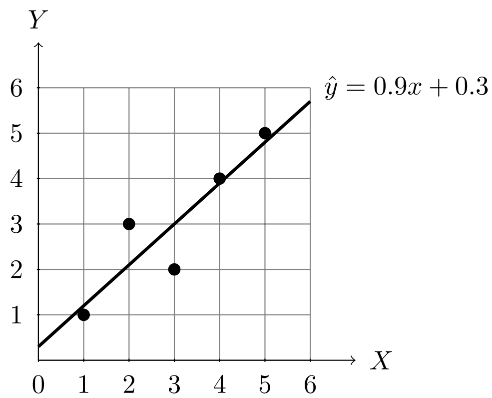
x	1	2	3	4	5	$\sum x = 15$
y	1	3	2	4	5	$\sum y = 15$
x^2	1	4	9	16	25	$\sum x^2 = 55$
xy	1	6	6	16	25	$\sum xy = 54$

จะได้ว่า

$$b = \frac{(\sum x)(\sum y) - 5\sum xy}{(\sum x)^2 - 5\sum x^2} = \frac{15(15) - 5(54)}{15^2 - 5(55)} = 0.9$$

$$a = \frac{\sum y - b\sum x}{5} = \frac{15 - 0.9(15)}{5} = 0.3$$

ดังนั้นสมการของเส้นถดถอย y เทียบกับ x คือ $\hat{y} = 0.9x + 0.3$



ตัวอย่าง 10.2.3 ถ้าระยะเวลากับรายได้ของบริษัทหนึ่งในรอบ 5 ปี มีความสัมพันธ์แบบเส้นตรง ดังข้อมูล

ปี (พ.ศ.)	2561	2562	2563	2564	2565
รายได้ (ร้อยล้านบาท)	1	1.5	2	3	4.5

จงคาดคะเนว่าในปี 2569 บริษัทจะมีรายได้เท่าใด
แนวคำตอบ ให้ $x = -2, -1, 0, 1, 2$ แทนปี 2561, 2562, 2563, 2564, 2565 ตามลำดับ และ y แทนรายได้ของบริษัท หน่วยเป็นร้อยล้านบาท แสดงตารางได้ดังนี้

	x	y	xy	x^2
	-2	1	-2	4
	-1	1.5	-1.5	1
	0	2	0	0
	1	3	3	1
	2	4.5	9	4
รวม	0	12	8.5	10

$$b = \frac{(\sum x)(\sum y) - 5\sum xy}{(\sum x)^2 - 5\sum x^2} = \frac{0(12) - 5(8.5)}{0^2 - 5(10)} = 0.85$$
$$a = \frac{\sum y - b\sum x}{5} = \frac{8.5 - 5(0)}{5} = 2.4$$

สมการของเส้นถดถอย y เทียบกับ x คือ $\hat{y} = 0.85x + 2.4$ สำหรับปี 2569 ตรงกับ $x = 6$ จะได้ว่า

$$\hat{y} = 0.85(6) + 2.4 = 7.5$$

ดังนั้นคาดการณ์ได้ว่าในปี 2569 บริษัทจะมีรายได้ 750 ล้านบาท

การหาค่า a และ b จากเครื่องคำนวณ CASIO รุ่น fx-991EX โดยทำได้เบื้องต้นดังนี้

MENU $\rightarrow$ Statistics $\rightarrow$ $y = a + bx$

จากตัวอย่าง 10.2.2

x	1	2	3	4	5
y	1	3	2	4	5

เราจะนำเข้าตัวเลขได้ดังนี้

รูปที่ 10.2: แสดงหน้าจอเมื่อนำเข้าตัวเลขใส่ในเมนู 2 : $y = a + bx$

	x	y
1	1	1
2	2	3
3	3	2
4	4	4
5	5	5

เมื่อนำเข้าตัวเลขครบถ้วนแล้วให้กดที่เมนู

OPTN $\rightarrow$ Regression Calc

หน้าจอจะแสดงผลดังนี้

$$y = a + bx$$
$$a = 0.3$$
$$b = 0.9$$
$$r = 0.9$$

จะเห็นว่าการใช้เครื่องคำนวณเป็นอีกวิธีที่ทำให้ได้มาซึ่งสมการถดถอยเชิงเส้น แต่มีเครื่องมือที่ใช้ในการคำนวณดังกล่าวอีกอย่าง เช่น Application ทางสถิติ และ โปรแกรม Excel เป็นต้น ซึ่งผู้อ่านอาจศึกษาเพิ่มด้วยตนเองได้

ตัวอย่าง 10.2.4 ความสัมพันธ์ระหว่างความสูงของต้นถั่วชนิดหนึ่งกับเวลาเป็นดังนี้

เวลา (วัน)	1	2	3	4	5
ความสูงของต้นถั่ว (เซนติเมตร)	1	1.3	1.5	1.7	2.5

ถ้าความสัมพันธ์ระหว่างเวลากับความสูงของต้นถั่วแบบเชิงเส้น จงทำนายว่าวันที่ 7 ต้นถั่วชนิดนี้จะมีความสูงกี่เซนติเมตร
แนวคำตอบ ให้ x แทนจำนวนวันเมื่อหลังจากปลูกต้นถั่ว และ y แทนความสูงของต้นถั่ว หน่วยเซนติเมตร แสดงข้อมูลได้ดังนี้

x	1	2	3	4	5
y	1	1.3	1.5	1.7	2.5

นำไปคำนวณในเครื่องคำนวณ CASIO รุ่น fx-991EX ได้ค่า $a = 0.58$ และ $b = 0.34$ ดังนั้น

$$\hat{y} = 0.58 + 0.34x$$

วันที่ 7 ตรงกับ $x = 7$ จะได้ว่า $\hat{y} = 0.58 + 0.34(7) = 2.96$ สรุปได้ว่าวันที่ 7 ต้นถั่วชนิดนี้จะมีความสูง 2.96 เซนติเมตร

ตัวอย่าง 10.2.5 หุ้นน้องใหม่ชื่อย่อ MATH ทำการเข้าขายในตลาดหลักทัพย์ปี 2565 มีความสัมพันธ์ระหว่างวันกับราคาปิดวันสุดท้ายของเดือนที่เปิดซื้อ-ขาย ในตลาด ดังนี้

วันที่ซื้อขายวันสุดท้ายของเดือน	มีนาคม	เมษายน	พฤษภาคม	มิถุนายน
ราคาปิด (บาท)	1.00	0.90	0.82	0.86

ถ้าความสัมพันธ์ระหว่างวันที่ซื้อขายกับราคาปิดเป็นแบบเชิงเส้น จงทำนายว่าปลายปี 2565 (เดือนธันวาคม) หุ้น MATH จะมีราคาปิดประมาณกี่บาท ณ วันสุดท้ายของเดือนที่เปิดซื้อ-ขาย ในตลาด
แนวคำตอบ ให้ $x = 1, 2, 3, 4$ แทนเดือน มีนาคม, เมษายน, พฤษภาคม และ มิถุนายน ตามลำดับ และ y แทนราคาปิด หน่วยบาท แสดงข้อมูลได้ดังนี้

x	1	2	3	4
y	1.00	0.90	0.82	0.86

นำไปคำนวณในเครื่องคำนวณ CASIO รุ่น fx-991EX ได้ค่า $a = 1.02$ และ $b = -0.05$ ดังนั้น

$$\hat{y} = -0.05x + 1.02$$

เดือนธันวาคม ตรงกับ $x = 10$ จะได้ว่า $\hat{y} = -0.05(10) + 1.02 = 0.52$ สรุปได้ปลายปี 2565 หุ้น MATH จะมีราคาปิดประมาณ 0.52 บาท ณ วันสุดท้ายของเดือนที่เปิดซื้อ-ขาย ในตลาด

แบบฝึกหัด 10.1 - 10.2

1. ให้ตัวแปรสุ่ม X และ Y มีฟังก์ชันความน่าจะเป็นร่วมกันคือ

$$f(x,y) = \begin{cases} 2 & \text{เมื่อ } 0 < x < 1 \text{ และ } 0 < x < y < 1 \\ 0 & \text{เมื่อ } x, y \text{ มีค่าอื่น ๆ} \end{cases}$$

จงหาสมการเส้นถดถอยของ X เทียบกับ y

2. จากข้อมูลการปลูกถั่วเขียวของนายถั่วแระทั้งหมด 5 กระถาง ในแต่ละกระถางใส่เมล็ดถั่วเขียวลงไปจำนวนต่างๆกัน หลังจากนั้น 7 วันก็นับจำนวนถั่วเขียวที่งอก ปรากฏดังนี้

กระถางที่	จำนวนถั่วเขียวที่ใส่ลงไป (เมล็ด)	จำนวนถั่วเขียวที่งอกเมื่อนับในวันที่ 7 (เมล็ด)
1	2	1
2	4	2
3	6	5
4	8	6
5	10	8

2.1 จงวาดแผนภาพการกระจาย และระบุความสัมพันธ์ระหว่างจำนวนถั่วเขียวที่ใส่ลงไปและที่จำนวนที่งอก

2.2 จงหาสมการถดถอยของความสัมพันธ์ข้างต้น

2.3 จงทำนายจำนวนถั่วเขียวที่จะงอก เมื่อปลูกจำนวน 9 เมล็ด และจำนวน 20 เมล็ด

3. การทดลองวัดความสัมพันธ์ระหว่างเวลา t (วินาที) กับระยะทาง s (เมตร) ของวัตถุที่เคลื่อนที่เป็นดังนี้

t	1	2	3	4
s	2	8	18	32

ถ้าความสัมพันธ์ระหว่างเวลากับระยะทางเป็นแบบเชิงเส้น จงทำนายว่าระยะทางที่วัตถุเคลื่อนที่ได้ขณะ $t = 1.5$ วินาที เท่ากับกี่เมตร

4. ความสัมพันธ์ระหว่างรายรับและรายจ่ายของครอบครัวในหมู่บ้านแห่งเป็นดังนี้

รายรับ (หน่วยหมื่นบาท)	1.2	1.8	2.5	4.5	6
รายจ่าย (หน่วยหมื่นบาท)	1	1	2.2	3.4	4.4

ถ้าความสัมพันธ์ระหว่างรายรับกับรายจ่ายเป็นแบบเชิงเส้น แล้วครอบครัวที่มีรายรับ $50,000$ บาท จะมีรายจ่ายเท่าใดกี่บาท

5. ถ้าความสัมพันธ์เชิงฟังก์ชันระหว่างตัวแปร x และ y มีกราฟเป็นเส้นตรง โดยที่

$$\sum_{i=1}^{8} x_i = 32,\ \sum_{i=1}^{8} y_i = 16,\quad \sum_{i=1}^{8} x_i y_i = 65 \quad \sum_{i=1}^{8} x_i^2 = 140 \quad \text{และ} \quad \sum_{i=1}^{8} y_i^2 = 34$$

ถ้า $x = 8$ คะแนน แล้วจะประมาณค่า y ได้เท่าใด

6. จากสถิติการขายบ้านของบริษัทแห่งหนึ่ง จำนวนบ้านที่ขายได้มีความสัมพันธ์กับเวลาในลักษณะเส้นตรง

พ.ศ.	2559	2560	2561	2562	2563	2564
จำนวนบ้านที่ขายได้ (ร้อยหลัง)	5	8	12	15	20	25

จงหาว่าจำนวนบ้านที่ขายได้ในปี พ.ศ. 2568

7. พิจารณาข้อมูล x และ y ดังนี้

x	-3	-1	0	1	3
y	0	a	$a+3$	$a+4$	$a+6$

เมื่อ a เป็นค่าคงที่ โดย x และ y มีความสัมพันธ์เป็นกราฟเส้นตรง โดยมีความชันเท่ากับ 1.55 ถ้า $x = 4$ จะประมาณค่า y

8. มูลค่าอุตสาหกรรมสิ่งท้อที่ประเทศไทยส่งออกไปขายยังต่างประเทศระหว่าง พ.ศ. 2520 – 2524 เป็นดังนี้

พ.ศ.	2520	2521	2522	2523	2524
มูลค่า (ล้านบาท)	1	3	4	5	9

ความสัมพันธ์ของการส่งออกกับเวลาเป็นแบบเส้นตรง จงคาดการณ์ว่ามูลค่าการส่งออกกี่ล้านบาทในปี พ.ศ. 2527

10.3 ช่วงความเชื่อมั่นและการทดสอบสมมติฐาน

ข้อมูลในรูป $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ และ

$$Y_i = \alpha + \beta x_i + E_i$$

เมื่อ E_i เป็นตัวแปรสุ่มที่อิสระต่อกัน มีการแจกแจงปกติและมีค่าเฉลี่ยเป็นศูนย์ ความแปรปรวนเท่ากับ σ^2 การหาช่วงความเชื่อมั่นของ α และ β และจะหาค่าประมาณค่าของ Y

ให้ a และ b เป็นค่าประมาณของ α และ β ซึ่งหาได้จากตัวอย่างขนาด n เราอาจคำนวนค่าของ α และ β จากตัวอย่างหลาย ๆ ชุด เป็นค่าของ A และ B ตามลำดับ จากที่ E_i มีการแจกแจงปกติ ทำให้ Y_i ที่เป็นอิสระต่อกันและมีการแจกแจงปกติ $n(Y_i; \alpha + \beta x_i, \sigma^2)$ โดยที่

$$B = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

จากนั้นเราพิสูจน์ได้ว่า $E(B) = \beta$ และ $E(A) = \alpha$ โดยที่

$$\sigma_B^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{และ} \qquad \sigma_A^2 = \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sigma^2$$

จาก $\bar{y} = a + b\bar{x}$ พิจารณา

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n}(y_i - a - bx_i)^2 = \sum_{i=1}^{n}[(y_i - \bar{y}) - b(x_i - \bar{x})]^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2b\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + b^2\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= S_{yy} - 2bS_{xy} + b^2 S_{xx}
\end{aligned}$$

เมื่อ

$$\begin{aligned}
S_{yy} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \\
S_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \\
S_{xx} &= \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2
\end{aligned}$$

เนื่องจาก $b = \frac{S_{xy}}{S_{xx}}$ ดังนั้น

$$\text{SSE} = S_{yy} - bS_{xy}$$

เราพิสูจน์ได้ว่า $E(S^2) = \sigma^2$ เมื่อ

$$S^2 = \frac{\text{SSE}}{n-2}$$

นั่นคือเราใช้ S^2 ประมาณค่า σ^2

เนื่องจาก B เป็นตัวสุ่มปกติ และ $\dfrac{(n-2)S^2}{\sigma^2}$ ตัวแปรสุ่มไคสแควร์ ที่มีองศาอิสระ $n-2$ ดังนั้น

$$T = \frac{B-\beta}{\frac{S}{\sqrt{S_{xx}}}} \qquad \text{มีการแจกแจงที ที่มีองศาอิสระ } n-2$$

ทำให้เราสร้างช่วงความเชื่อมั่น $(1-\alpha)100\%$ สำหรับสัมประสิทธิ์ถดถอย β ได้เป็น

$$b \pm t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{S_{xx}}} \qquad \text{หรือ} \qquad b - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{S_{xx}}} \ < \ \beta \ < \ b + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{S_{xx}}}$$

ตัวอย่าง 10.3.1 กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	2	3	4	5
y	2	3	5	3	6

จงหาช่วงความเชื่อมั่น 95% สำหรับ β

แนวคำตอบ พิจารณาตาราง

x	1	2	3	4	5	$\sum x = 15$
y	2	3	5	3	6	$\sum y = 19$
x^2	1	4	9	16	25	$\sum x^2 = 55$
y^2	4	9	25	9	36	$\sum x^2 = 83$
xy	2	6	15	12	30	$\sum xy = 65$

จะได้ว่า

$$S_{yy} = \sum y^2 - \frac{1}{5}\left(\sum y\right)^2 = 83 - \frac{1}{5}(19)^2 = 10.8$$
$$S_{xy} = \sum xy - \frac{1}{5}\left(\sum x\right)\left(\sum y\right) = 65 - \frac{1}{5}(15)(19) = 8$$
$$S_{xx} = \sum x^2 - \frac{1}{5}\left(\sum x\right)^2 = 55 - \frac{1}{5}(15)^2 = 10$$

จะได้ว่า $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{8}{10} = 0.8$ และ $\text{SSE} = S_{yy} - bS_{xy} = 10.8 - 0.8(8) = 4.4$

ทำให้ได้ว่า

$$S = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{4.4}{3}} = 1.21$$

เนื่องจาก $\alpha = 0.05$ และ $\nu = 5-2 = 3$ ฉะนั้น $t_{\frac{\alpha}{2},\nu} = t_{0.025,3} = 3.18$ (อ่านค่าจาก Application)

ช่วงความเชื่อมั่น 95% สำหรับสัมประสิทธิ์ถดถอย β คือ

$$b - t_{0.025} \cdot \frac{S}{\sqrt{S_{xx}}} \ < \ \beta \ < \ b + t_{0.025} \cdot \frac{S}{\sqrt{S_{xx}}}$$

$$0.8 - 3.18 \cdot \frac{1.21}{\sqrt{10}} \ < \ \beta \ < \ 0.8 + 3.18 \cdot \frac{1.21}{\sqrt{10}}$$

$$-0.42 \ < \ \beta \ < \ 2.02$$

การทดสอบสมมติฐาน

$$H_0 : \beta = \beta_0$$

แย้งกับสมมติฐานที่เหมาะสม ใช้การแจกแจงที ที่มีองศาเสรี $n-2$ โดยใช้ค่าสถิติคำนวณคือ

$$T = \frac{B - \beta_0}{\frac{S}{\sqrt{S_{xx}}}}$$

ตัวอย่าง 10.3.2 ข้อมูลจากสำนักงานสถิติแห่งชาติ (สสช.) เกี่ยวกับการเงินการคลังของประเทศไทย เรื่องการจัดเก็บรายได้ของรัฐบาลและหนี้สาธารณะของปีงบประมาณ 2555-2564 ปรากฏดังนี้

ปีงบประมาณ	2555	2556	2557	2558	2559	2560	2561	2562	2563	2564
รายได้รวม (ล้านล้านบาท)	2.34	2.51	2.71	2.76	2.83	2.82	2.93	3.03	3.23	3.15
หนี้สาธารณะ (ล้านล้านบาท)	4.98	5.51	5.80	5.91	6.08	6.44	6.84	6.97	7.87	9.40

ที่มา : สำนักงบประมาณ สำนักนายกรัฐมนตรี และสำนักงานบริหารหนี้สาธารณะ กระทรวงการคลัง

จงทดสอบว่าสัมประสิทธิ์ถดถอย $\beta < 0.5$ หรือไม่ ของสมการถดถอยปีงบประมาณกับหนี้สาธารณะ ที่ระดับนัยสำคัญ 0.05

แนวคำตอบ ให้ $x = 1, 2, 3, ..., 10$ แทนปีงบประมาณ 2555, 2556, 2557, ...,2564 ตามลำดับ และ y แทนหนี้สาธารณะ หน่วยล้านล้านบาท แสดงตารางได้ดังนี้

x	1	2	3	4	5	6	7	8	9	10
y	4.98	5.51	5.80	5.91	6.08	6.44	6.84	6.97	7.87	9.40

จากเครื่องคำนวณ CASIO รุ่น fx-991EX โดยใช้เมนู

MENU $\rightarrow$ Statistics $\rightarrow$ $y = a + bx$

แล้วนำเข้าข้อมูล (x, y) จากตารางข้างต้น จากนั้นเลือกเมนู

OPTN $\rightarrow$ Regression Calc

ได้ค่า $b = 0.39$ จากนั้นเลือกเมนู

OPTN $\rightarrow$ 2-Variable Calc

เลือกค่าที่ต้องการได้ดังนี้

$$\sum x = 55 \qquad \sum x^2 = 385 \qquad \sum y = 65.8 \qquad \sum y^2 = 447.832 \qquad \sum xy = 394.55$$

กำหนดสมมติฐาน

$H_0 \;:\; \beta = 0.5$

$H_1 \;:\; \beta < 0.5$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

จากข้อมูลข้างต้นเมื่อ $n = 10$ จะได้ว่า

$$\begin{aligned}
S_{xx} &= \sum x^2 - \frac{1}{10}\left(\sum x\right)^2 = 385 - \frac{1}{10}(55)^2 = 82.5 \\
S_{yy} &= \sum y^2 - \frac{1}{10}\left(\sum y\right)^2 = 447.832 - \frac{1}{10}(65.8)^2 = 14.868 \\
S_{xy} &= \sum xy - \frac{1}{10}\left(\sum x\right)\left(\sum y\right) = 394.55 - \frac{1}{10}(55)(65.8) = 32.65 \\
\mathsf{SSE} &= S_{yy} - bS_{xy} = 14.868 - 0.39(32.65) = 2.1345 \\
S &= \sqrt{\frac{\mathsf{SSE}}{n-2}} = \sqrt{\frac{2.1345}{8}} = 0.52
\end{aligned}$$

เนื่องจาก $n = 10$ ดังนั้น $\nu = 10 - 2 = 8$ และ

$$T_{\text{คำนวณ}} = \frac{b - \beta_0}{\frac{S}{\sqrt{S_{xx}}}} = \frac{0.39 - 0.5}{\frac{0.52}{\sqrt{82.5}}} = -1.92$$

เป็นการทดสอบข้างเดียว(ข้างซ้าย) พิจารณา $t_{\alpha,\nu} = t_{0.05,8} = 1.86$ (อ่านค่าจาก Application) ฉะนั้นบริเวณวิกฤตคือ $t < -1.86$

จะเห็นว่า $T_{\text{คำนวณ}} = -1.92 < -1.86$ อยู่ในบริเวณวิกฤต ดังนั้นปฏิเสธ H_0

สรุปได้ว่าสัมประสิทธิ์ถดถอย $\beta < 0.5$ ของสมการถดถอยปีงบประมาณกับหนี้สาธารณะ ที่ระดับนัยสำคัญ 0.05

สำหรับกรณี $\beta_0 = 0$ อาจจะใช้การทดสอบด้วยตัวสถิติที หรือสถิติเอฟ เราจะพิจารณาตารางที่เรียกว่า ANOVA TABLE นั่นคือ

ANOVA TABLE

SOV	SS	DF	MS	F
REGRESSION	SSR	1	MSR	$\frac{\text{MSR}}{\text{MSE}}$
ERROR	SSE	$n-2$	MSE	
TOTAL	SST	$n-1$		

จากตารางจะได้ว่า SST = SSR + SSE โดยที่ SST $= S_{yy}$ และ SSR $= bS_{xy}$
คำนวนค่าสถิติเอฟจาก

$$F_{\text{คำนวณ}} = \frac{\text{MSR}}{\text{MSE}} \qquad \text{โดยมีองศาเสรีคือ } \nu_1 = 1, \nu_2 = n-2$$

เมื่อ $\text{MSR} = \frac{\text{SSR}}{1} = \text{SSR}$ และ $\text{MSE} = \frac{\text{SSE}}{n-2}$

ตัวอย่าง 10.3.3 ปริมาณการส่งออกข้าวหอมมะลิของประเทศไทย แสดงดังตาราง

ปี พ.ศ.	2558	2559	2560	2561	2562
ปริมาณส่งออก (ล้านตัน)	1.39	1 .54	1.61	1.25	1.11

* กรมส่งเสริมการค้าระหว่างประเทศ กระทรวงพานิชย์

1. จงคาดการณ์ว่าปี 2565 ปริมาณส่งออกประมาณเท่าใด
2. จงสร้างตาราง ANOVA
3. จงทดสอบว่าสัมประสิทธิ์ถดถอย $\beta < 0$ หรือไม่ ที่ระดับนัยสำคัญ 0.05 โดยใช้สถิติเอฟ

แนวคำตอบ ให้ x = 1, 2, 3, ... แทนปี พ.ศ. 2558, 2559, 2560, ... ตามลำดับ
และ y แทนปริมาณการส่งออกข้าวหอมมะลิของประเทศไทย หน่วยเป็นล้านตัน แสดงตารางได้ดังนี้

x	1	2	3	4	5
y	1.39	1.54	1.61	1.25	1.11

โดยเครื่องคำนวณจะได้ว่า $a = 1.635$ และ $b = -0.085$ ดังนั้น

$$\hat{y} = 1.635 - 0.085x$$

เนื่องจากปี 2565 คือ $x = 8$ นั่นคือ $\hat{y} = 1.635 - 0.085(8) = 0.955$
ดังนั้นปริมาณการส่งออกข้าวหอมมะลิของประเทศไทยในปี พ.ศ. 2565 ประมาณ 0.955 ล้านตัน
จากตารางโดยใช้เครื่องคำนวณจะได้ว่า $n = 5$ และ $b = -0.085$ โดยที่

$$\sum x = 15 \qquad \sum x^2 = 55 \qquad \sum y = 6.9 \qquad \sum y^2 = 9.6904 \qquad \sum xy = 19.85$$

จะได้ว่า

$$\begin{aligned}
S_{xx} &= \sum x^2 - \frac{1}{5}\left(\sum x\right)^2 = 55 - \frac{1}{5}(15)^2 = 10 \\
S_{yy} &= \sum y^2 - \frac{1}{5}\left(\sum y\right)^2 = 9.6904 - \frac{1}{5}(6.9)^2 = 0.1684 \\
S_{xy} &= \sum xy - \frac{1}{5}\left(\sum x\right)\left(\sum y\right) = 19.85 - \frac{1}{5}(15)(6.9) = -0.85 \\
\text{SSR} &= bS_{xy} = (-0.085)(-0.85) = 0.0723 \\
\text{SSE} &= S_{yy} - bS_{xy} = 0.1684 - (-0.085)(-0.85) = 0.0961 \\
\text{MSE} &= \frac{\text{SSE}}{n-2} = \frac{0.0961}{3} = 0.0320 \\
\text{MSR} &= \frac{\text{SSR}}{1} = 0.0723 \\
\text{F} &= \frac{\text{MSR}}{\text{MSE}} = \frac{0.0723}{0.0320} = 2.2594
\end{aligned}$$

และ $\nu_1 = 1, \nu_2 = 5 - 2 = 3$ แสดงตาราง ANOVA ได้ดังนี้

ANOVA TABLE

SOV	SS	DF	MS	F
REGRESSION	0.0723	1	0.0723	2.2594
ERROR	0.0961	3	0.0320	
TOTAL	0.1684	4		

กำหนดสมมติฐาน

$H_0 \ : \ \beta = 0$

$H_1 \ : \ \beta < 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

เป็นการทดสอบข้างเดียว(ข้างซ้าย) พิจารณา $f_{1-\alpha,(\nu_1,\nu_2)} = f_{1-0.05,(1,3)} = f_{0.95,(1,3)} = 0.0046$ (อ่านค่าจาก Application) หรืออาจใช้ตารางเอฟ

$$f_{0.95,(1,3)} = \frac{1}{f_{1-0.95,(3,1)}} = \frac{1}{f_{0.05,(3,1)}} = \frac{1}{215.707} = 0.0046$$

ฉะนั้น บริเวณวิกฤตคือ $f < 0.0046$

จากตาราง ANOVA จะเห็นว่า $F_{\text{คำนวณ}} = 2.2594$ ซึ่ง $F_{\text{คำนวณ}} > 0.0046$ อยู่ในบริเวณยอมรับ ดังนั้น ยอมรับ H_0

แบบฝึกหัด 10.3

1. กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	2	3	4	5
y	5	6	8	7	9

จงหาช่วงความเชื่อมั่น 95% สำหรับ β

2. จากการสุ่มนักศึกษาเพื่อสอบถามคะแนนวิชาแคลคูลัส 1 (x) และ. แคลคูลัส 2 (y) จำนวน 5 คน แสดงดังตาราง

คะแนนวิชาแคลคูลัส 1	75	60	80	46	55
คะแนนวิชาแคลคูลัส 2	79	59	83	50	65

จงหาช่วงความเชื่อมั่น 90% สำหรับ β

3. ในการสอบถามความขยันในการเรียนคณิตศาสตร์ พบว่าจำนวนนักศึกษาที่สุ่มมา 10 คน โดยสอบถามจำนวนชั่วโมง (x) ที่เข้าศึกษาคณิตศาสตร์ และคะแนนที่เขาได้ (y) ปรากฏดังนี้ $\sum x_i = 20, \;\; \sum y_i = 40, \;\; \sum x_i^2 = 121, \;\; \sum x_i y_i = -82, \;\; \sum y_i^2 = 516, \;\; n = 10$

   3.1 จงหา $\hat{y} = a + bx$

   3.2 จงหาช่วงความเชื่อมั่น 95% สำหรับ β

   3.3 สร้างตาราง ANOVA

   3.4 จงทดสอบสมมติฐาน $H_0 : \beta = -2.5$ แย้งกับ $\beta > -2.5$ ที่ระดับนัยสำคัญ 0.05

4. กำหนดให้ข้อมูล x และ y เป็นตัวอย่างเป็นคู่โดยที่ $n = 25$ และ $\sum x_i = 121.2, \;\; \sum y_i = 67.83, \;\; \sum(x_i - \bar{x})(y_i - \bar{y}) = 36.0678, \;\; \sum(x_i - \bar{x})^2 = 31.4724, \;\; \sum(y_i - \bar{y})^2 = 42.3735$

   4.1 จงหา $\hat{y} = a + bx$

   4.2 สร้างตาราง ANOVA

   4.3 จงทดสอบสมมติฐาน $H_0 : \beta = 0$ แย้งกับ $\beta \neq 0$ ที่ระดับนัยสำคัญ 0.05 โดยใช้สถิติเอฟ

5. กำหนดคู่ของ (x, y) ดังตาราง

x	2	4	6	8	10
y	3	1	7	5	9

   5.1 จงหาสมการเส้นถดถอย $\hat{y} = a + bx$

   5.2 จงหาช่วงความเชื่อมั่น 95% สำหรับ β

   5.3 สร้างตาราง ANOVA แสดงการวิเคราะห์การถดถอย

   5.4 จงทดสอบสมมติฐาน $H_0 : \beta = 0$ แย้งกับ $\beta \neq 0$ ที่ระดับนัยสำคัญ 0.05

10.4 สหสัมพันธ์

สหสัมพันธ์ (Correlation) คือตัววัดความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไปว่ามีความสัมพันธ์กันหรือไม่และมากน้อยเพียงใด

1. ถ้าความสัมพันธ์ของตัวแปรสุ่มเพียง 2 ตัวเรียกว่า **สหสัมพันธ์เชิงเดียว** (simple correlation)

2. ถ้ามีตัวแปรสุ่มมากกว่า 2 ตัวขึ้นไปเรียกว่า **สหสัมพันธ์พหุคูณ** (multiple correlation)

ในหัวข้อนี้จะกล่าวถึงสหสัมพันธ์เชิงเดี่ยวเท่านั้น

สมมติการแจกแจงแบบมีเงื่อนไข $f(y \mid x)$ ของ Y เมื่อกำหนด x เป็นการแจกแจงปกติที่มีค่าเฉลี่ย $\mu_{Y|x} = \alpha + \beta x$ ความแปรปรวน σ^2 และมีการแจกแจงปกติที่มีค่าเฉลี่ย μ_X และความแปรปรวน σ_X^2 ฟังก์ชันความน่าจะเป็นร่วมกันของตัวแปรสุ่ม X และ Y คือ

$$f(x, y) = n(y \mid x; \alpha + \beta x, \sigma^2) \cdot n(x; \mu_X, \sigma_X^2)$$

เมื่อจัดรูปและพิจารณาค่าต่าง ๆ ร่วมด้วยกับ $Y = \alpha + \beta x + E$
จะได้ ρ ที่เรียกว่า **สัมประสิทธิ์สหสัมพันธ์** (correlation coefficient) หาได้จาก

$$\rho^2 = 1 - \frac{\sigma^2}{\sigma_Y^2} = \beta^2 \frac{\sigma_X^2}{\sigma_Y^2}$$

เมื่อ σ_Y^2 คือความแปรปรวน Y ซึ่ง $\sigma_Y^2 \geq \sigma^2$ ดังนั้น

$$-1 \leq \rho \leq 1$$

ในกรณีที่ $\rho = \pm 1$ แสดงว่า $\sigma^2 = 0$ นั่นคือ X และ Y มีความสัมพันธ์กันแบบสมบูรณ์ไม่มีความผิดพลาด มีหมายความว่าถ้า $\rho = 1$ เส้นตรงจะมีความชันเป็นบวก ถ้า $\rho = -1$ เส้นตรงจะมีความชันเป็นลบ อีกทั้งสรุปได้ว่า

1. ถ้า ρ มีค่าเข้าใกล้ 1 มีหมายความว่าความสัมพันธ์ยิ่งใกล้เคียงเส้นตรงที่มีความชันเป็นบวก

2. ถ้า ρ มีค่าเข้าใกล้ -1 มีหมายความว่าความสัมพันธ์ยิ่งใกล้เคียงเส้นตรงที่มีความชันเป็นลบ

3. ถ้า ρ มีค่าเข้าใกล้ 0 มีหมายความว่าแทบไม่มีความสัมพันธ์แบบเส้นตรงหรือไม่มีเลย

ค่าประมาณของสัมประสิทธิ์สหสัมพันธ์ ρ คือ r เรียกว่า **สัมประสิทธิ์สหสัมพันธ์ตัวอย่าง** (sample correlation coefficient) หาได้จาก

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = b^2\frac{S_{xx}}{S_{yy}}$$

เนื่องจาก SSE $= S_{yy} - bS_{xy}$ และ SST $= S_{yy}$ ดังนั้น

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

เนื่องจาก SST $\geq$ SSE ดังนั้น

$$-1 \leq r \leq 1$$

ในกรณีที่ $r = \pm 1$ แสดงว่า SSE $= 0$ นั่นคือ X และ Y มีความสัมพันธ์กันแบบสมบูรณ์ไม่มีความผิดพลาด มีหมายความว่าถ้า $r = 1$ เส้นตรงจะมีความชันเป็นบวก ถ้า $r = -1$ เส้นตรงจะมีความชันเป็นลบ อีกทั้งสรุปได้ว่า

1. ถ้า r มีค่าเข้าใกล้ 1 มีหมายความว่าความสัมพันธ์ยิ่งใกล้เคียงเส้นตรงที่มีความชันเป็นบวก
2. ถ้า r มีค่าเข้าใกล้ -1 มีหมายความว่าความสัมพันธ์ยิ่งใกล้เคียงเส้นตรงที่มีความชันเป็นลบ
3. ถ้า r มีค่าเข้าใกล้ 0 มีหมายความว่าแทบไม่มีความสัมพันธ์แบบเส้นตรงหรือไม่มีเลย

หมายเหตุ ค่าสัมประสิทธิ์สหสัมพันธ์ r จะมีเครื่องหมายเหมือนกับค่า b และ

$r^2 \times 100\%$ คือเปอร์เซ็นต์หรือร้อยละของ x กับ y ที่มีความสัมพันธ์เชิงเส้น

ตัวอย่าง 10.4.1 กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	2	3	4
y	2	3	5	3

จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่างพร้อมบอกความหมาย

แนวคำตอบ พิจารณาตาราง

x	1	2	3	4	$\sum x = 10$
y	2	3	5	3	$\sum y = 13$
x^2	1	4	9	16	$\sum x^2 = 30$
y^2	4	9	25	9	$\sum x^2 = 47$
xy	2	6	15	12	$\sum xy = 35$

จะได้ว่า

$$\begin{aligned}
S_{yy} &= \sum y^2 - \frac{1}{4}\left(\sum y\right)^2 = 47 - \frac{1}{4}(13)^2 = 4.75 \\
S_{xy} &= \sum xy - \frac{1}{4}\left(\sum x\right)\left(\sum y\right) = 35 - \frac{1}{5}(10)(13) = 2.5 \\
S_{xx} &= \sum x^2 - \frac{1}{4}\left(\sum x\right)^2 = 30 - \frac{1}{4}(10)^2 = 5
\end{aligned}$$

จะได้ว่า

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{2.5^2}{5(4.75)} = 0.2631$$

เนื่องจาก $b > 0$ จะได้สัมประสิทธิ์สหสัมพันธ์ตัวอย่าง $r = \sqrt{0.2631} = 0.5130$
โดยมี $r^2 \times 100\% = 26.31\%$ นั่นหมายความว่า x และ y มีความสัมพันธ์เชิงเส้นที่มีความชันเป็นบวก 26.31%

เราอาจหาค่า r ได้จากเครื่องคำนวณ โดยนำเข้าข้อมูล (x, y) จากตารางในตัวอย่าง 10.4.1 จากนั้นเลือกเมนู

OPTN $\rightarrow$ Regression Calc

หน้าจอจะแสดงผลดังนี้

$$y = a + bx$$

$$\begin{aligned} a &= 2 \\ b &= 0.5 \\ r &= 0.512989176 \end{aligned}$$

ตัวอย่าง 10.4.2 กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	3	5	6	7
y	3	1	0	-1	-5

จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่างพร้อมบอกความหมาย
แนวคำตอบ โดยใช้เครื่องคำนวณจะได้ว่าสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง $r = -0.9168$ ฉะนั้น

$$r^2 \times 100\% = (-0.9198)^2 \times 100 = 84.05\%$$

นั่นหมายความว่า x และ y มีความสัมพันธ์เชิงเส้นที่มีความชันเป็นลบ 84.05%

ตัวอย่าง 10.4.3 กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	2	3	4	5
y	1	4	9	4	1

จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่างพร้อมบอกความหมาย
แนวคำตอบ โดยใช้เครื่องคำนวณจะได้ว่าสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง $r = 0$ ฉะนั้น

$$r^2 \times 100\% = 0\%$$

นั่นหมายความว่า x และ y ไม่มีความสัมพันธ์เชิงเส้นกัน

การทดสอบสมมติฐานของค่าสหสัมพันธ์ เราจะอาศัยตัวสถิติ Z_r ที่ทำกำหนดโดย

$$Z_r = \frac{1}{2}\ell\text{n}\left(\frac{1+r}{1-r}\right)$$

ซึ่งมีการแจกแจงปกติและ $E(Z_r) = Z_\rho$ เมื่อ $Z_\rho = \frac{1}{2}\ell\text{n}\left(\frac{1+\rho}{1-\rho}\right)$ เมื่อ $\rho \neq 0$
นำเสนอโดยนักสถิติศาสตร์ชาวอังกฤษชื่อ ฟิเชอร์, โรนัลด์ ไอล์เมอร์ (Fisher, Ronald Aylmer: 1890-1962)
สถิติที่ใช้ทดสอบคำนวณจาก $Z = \dfrac{Z_r - Z_\rho}{\sqrt{\text{var}(Z_r)}}$ จัดรูปจะได้ว่า

$$Z = \frac{\sqrt{n-3}}{2} \cdot \ell\text{n}\left[\left(\frac{1+r}{1-r}\right)\left(\frac{1-\rho}{1+\rho}\right)\right]$$

โดยกำหนดสมมติฐาน $H_0 : \rho = \rho_0$ แย้งกับสมมติฐานที่เหมาะสม

ตัวอย่าง 10.4.4 กำหนดข้อมูลข้างต้นแสดงปริมาณของยา x (เปอร์เซนต์) และเวลาที่ยาออกฤทธิ์ y (วินาที)

x	1	2	3	4	5
y	1	1	2	2	4

จงหาทดสอบสมมติฐาน $H_0 : \rho = 0.95$ แย้งกับ $H_1 : \rho < 0.95$ กำหนดระดับนัยสำคัญ 0.05
แนวคำตอบ ให้ x แทนปริมาณของยา (เปอร์เซนต์) และ y แทนเวลาที่ยาออกฤทธิ์ (วินาที)
กำหนดสมมติฐาน

$H_0 \;:\; \rho = 0.95$
$H_1 \;:\; \rho < 0.95$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$
จากตารางใช้เครื่องคำนวณจะได้ว่า $n = 5$ และ $r = 0.9037$ ใช้สูตรของฟิเชอร์

$$Z_{\text{คำนวณ}} = \frac{\sqrt{5-3}}{2} \cdot \ell\text{n}\left[\left(\frac{1+0.9037}{1-0.9037}\right)\left(\frac{1-0.95}{1+0.95}\right)\right] = -0.48$$

เป็นการทดสอบข้างเดียว (ข้างซ้าย) จะเห็นว่า

$$\text{P-value } = P(Z > |-0.48|) = P(Z > 0.48) = 0.3156 > 0.05 = \alpha$$

ดังนั้นยอมรับ H_0 สรุปได้ว่าปริมาณของยาและเวลาที่ยาออกฤทธิ์มีค่าสหสัมพันธ์ไม่น้อยกว่า 0.95 ที่ระดับนัยสำคัญ 0.05

สำหรับการทดสอบสมมติฐาน $H_0 : \rho = 0$ ใช้ได้ทั้งการทดสอบโดยค่าสถิติ Z และใช้ค่าสถิติที ซึ่งหาได้จาก

$$T = r\sqrt{\frac{n-2}{1-r^2}} \qquad \text{องศาเสรี} \quad \nu = n-2$$

ตัวอย่าง 10.4.5 จากการสุ่มนักศึกษาเพื่อสอบถามคะแนนวิชาแคลคูลัส 1 (x) และ แคลคูลัส 2 (y) จำนวน 5 คน แสดงดังตาราง

คะแนนวิชาแคลคูลัส 1	75	60	80	46	55
คะแนนวิชาแคลคูลัส 2	79	59	83	50	65

จงทดสอบว่าคะแนนของวิชาแคลคูลัส 1 และวิชาแคลคูลัส 2 มีความสัมพันธ์กันเชิงเส้นตรงหรือไม่ ที่ระดับนัยสำคัญ 0.05

แนวคำตอบ ให้ x แทนคะแนนวิชาแคลคูลัส 1 และ y แทนคะแนนคะแนนวิชาแคลคูลัส 2 แสดงได้ดังตาราง

x	75	60	80	46	55
y	79	59	83	50	65

กำหนดสมมติฐาน

$H_0 \;:\; \rho = 0$

$H_1 \;:\; \rho \neq 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

จากตารางโดยใช้เครื่องคำนวณจะได้ว่า $n = 5$ และ $r = 0.9603$ เป็นการทดสอบสองทาง ใช้สถิติที

$$T_{\text{คำนวณ}} = 0.9603\sqrt{\frac{5-2}{1-0.9603^2}} = 5.96$$

จะเห็นว่า $\nu = 5-2 = 3$ และ

$$\text{P-value} = P(T > 5.96) = 0.0047 < 0.025 = \frac{\alpha}{2}$$

ดังนั้นปฏิเสธ H_0 สรุปได้ว่าคะแนนของวิชาแคลคูลัส 1 และวิชาแคลคูลัส 2 มีความสัมพันธ์กันเชิงเส้นตรง ที่ระดับนัยสำคัญ 0.05

เนื่องจากค่าสถิติทีและค่าสถิติเอฟมีความสัมพันธ์กันจาก $r^2 = 1 - \frac{\text{SSE}}{\text{SST}}$ จะได้ว่า

$$\frac{1-r^2}{n-2} = \frac{\text{SSE}}{(n-2)\text{SST}}$$

พิสูจน์ได้ว่า $T^2 = \frac{\text{MSR}}{\text{MSE}} = F$ ดังนั้นสามารถทดสอบได้โดยตัวสถิติเอฟเช่นเดียวกับสัมประสิทธิ์ถดถอย และใช้องศาเสรีคือ $\nu_1 = 1, \nu_2 = n-2$ โดยสร้างตาราง ANOVA ได้เช่นกัน

ตัวอย่าง 10.4.6 ข้อมูลตัวอย่างเป็นคู่ (x, y) แสดงค่าตาราง ANOVA บางส่วนดังนี้

ANOVA TABLE

SOV	SS	DF	MS	F
REGRESSION	112.484	1		
ERROR	2.032	7		
TOTAL	114.516	8		

1. จงเติมค่าในตาราง ANOVA ให้สมบูรณ์
2. จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง
3. จงทดสอบว่า x และ y มีความสัมพันธ์กันเชิงเส้นหรือไม่ ที่ระดับนัยสำคัญ 0.05

แนวคำตอบ จากตารางจะได้ว่า SSR $= 112.484$, SSE $= 2.032$ และ $n - 2 = 7$ ฉะนั้น

$$\text{MSE} = \frac{\text{SSE}}{n-2} = \frac{2.032}{7} = 0.290$$
$$\text{MSR} = \frac{\text{SSR}}{1} = 112.484$$
$$\text{F} = \frac{\text{MSR}}{\text{MSE}} = \frac{112.484}{0.290} = 387.876$$

แสดงตาราง ANOVA ได้ดังนี้

ANOVA TABLE

SOV	SS	DF	MS	F
REGRESSION	112.484	1	112.484	387.876
ERROR	2.032	7	0.290	
TOTAL	114.516	8		

เนื่องจาก

$$r^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{2.032}{114.516} = 0.9823$$

จะได้ว่าสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง $r = \pm\sqrt{0.9823} = \pm 0.9911$ กำหนดสมมติฐาน

$H_0 \;:\; \rho = 0$

$H_1 \;:\; \rho \neq 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

เนื่องจากเป็นการทดสอบสองทาง จากตาราง ANOVA จะได้ค่า $\nu_1 = 1$ และ $\nu_2 = 7$ พิจารณา $f_{\frac{\alpha}{2},(\nu_1,\nu_2)} = f_{0.025,(1,7)} = 8.07$ และ

$$f_{1-\frac{\alpha}{2},(\nu_1,\nu_2)} = f_{0.975,(1,7)} = \frac{1}{f_{0.025,(7,1)}} = \frac{1}{948.22} = 0.00105$$

บริเวณวิกฤตคือ $f < 0.00105$ หรือ $f > 8.07$

จากตาราง ANOVA จะเห็นว่า $F_{\text{คำนวณ}} = 387.876$ และ $F_{\text{คำนวณ}} > 8.07$ อยู่ในบริเวณวิกฤต ดังนั้น ปฏิเสธ H_0 สรุปได้ว่าx และ y มีความสัมพันธ์กันเชิงเส้น ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 10.4.7 จากการสุ่มนักศึกษาเพื่อสอบถามคะแนนสอบย่อยครั้งที่ 1 (x) และ ครั้งที่ 2 (y) จำนวน 5 คน แสดงดังตาราง

คะแนนสอบย่อยครั้งที่ 1	7	6	8	4	9
คะแนนสอบย่อยครั้งที่ 2	5	4	9	5	7

จงสร้างตาราง ANOVA และจงทดสอบว่า คะแนนสอบย่อยครั้งที่ 1 และครั้งที่ 2 มีความสัมพันธ์กันเชิงเส้นตรงหรือไม่ ที่ระดับนัยสำคัญ 0.05

แนวคำตอบ ให้ x แทนคะแนนสอบย่อยครั้งที่ 1 และ y แทนคะแนนสอบย่อยครั้งที่ 2 แสดงได้ดังตาราง

x	7	6	8	4	9
y	5	4	9	5	7

กำหนดสมมติฐาน

$H_0 \;:\; \rho = 0$

$H_1 \;:\; \rho \neq 0$

ทดสอบที่ระดับนัยสำคัญ $\alpha = 0.05$

จากตารางโดยใช้เครื่องคำนวณจะได้ว่า $n = 5$ และ $b = 0.6757$ โดยที่

$$\sum x = 34 \qquad \sum x^2 = 246 \qquad \sum y = 30 \qquad \sum y^2 = 196 \qquad \sum xy = 214$$

จะได้ว่า

$$
\begin{aligned}
S_{xx} &= \sum x^2 - \frac{1}{5}\left(\sum x\right)^2 = 246 - \frac{1}{5}(34)^2 = 14.8 \\
S_{yy} &= \sum y^2 - \frac{1}{5}\left(\sum y\right)^2 = 196 - \frac{1}{5}(30)^2 = 16 \\
S_{xy} &= \sum xy - \frac{1}{5}\left(\sum x\right)\left(\sum y\right) = 214 - \frac{1}{5}(34)(30) = 10 \\
\text{SSR} &= bS_{xy} = 0.6757(10) = 6.757 \\
\text{SSE} &= S_{yy} - bS_{xy} = 16 - 0.6757(10) = 9.243 \\
\text{MSE} &= \frac{\text{SSE}}{n-2} = \frac{9.243}{3} = 3.081 \\
\text{MSR} &= \frac{\text{SSR}}{1} = 6.757 \\
\text{F} &= \frac{\text{MSE}}{\text{MSR}} = \frac{6.757}{3.081} = 2.193
\end{aligned}
$$

และ $\nu_1 = 1, \nu_2 = 5 - 2 = 3$ แสดงตาราง ANOVA ได้ดังนี้

ANOVA TABLE

SOV	SS	DF	MS	F
REGRESSION	6.757	1	6.757	2.193
ERROR	9.243	3	3.081	
TOTAL	16	4		

เนื่องจากเป็นการทดสอบสองทาง พิจารณา $f_{\frac{\alpha}{2},(\nu_1,\nu_2)} = f_{0.025,(1,3)} = 17.44$ และ

$$f_{1-\frac{\alpha}{2},(\nu_1,\nu_2)} = f_{0.975,(1,3)} = \frac{1}{f_{0.025,(3,1)}} = \frac{1}{864.16} = 0.0012$$

บริเวณวิกฤตคือ $f < 0.0012$ หรือ $f > 17.44$
จากตาราง ANOVA จะเห็นว่า $F_{\text{คำนวณ}} = 2.193$ และ $0.0012 < F_{\text{คำนวณ}} < 17.44$ อยู่ในบริเวณยอมรับ ดังนั้นยอมรับ H_0

สรุปได้ว่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0 หรือ คะแนนสอบย่อยครั้งที่ 1 และครั้งที่ 2 ไม่มีความสัมพันธ์กันเชิงเส้นตรง ที่ระดับนัยสำคัญ 0.05

แบบฝึกหัด 10.4

1. กำหนดข้อมูลของตัวอย่างดังตาราง

x	1	2	3	4	5
y	3	5	4	6	7

จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่างและให้ความหมาย

2. กำหนดข้อมูลของตัวอย่างดังตาราง

x	-3	-2	1	2	3
y	5	4	3	3	2

จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่างและให้ความหมาย

3. กำหนดข้อมูลข้างต้นแสดงปริมาณวิตามิน x (เปอร์เซนต์) และเวลาที่วิตามินเข้าสู่กระแสเลือด y (นาที)

x	1	2	3	4	5
y	10	8	6	5	4

จงหาทดสอบสมมติฐาน $H_0 : \rho = -0.8$ แย้งกับ $H_1 : \rho > -0.8$ กำหนดระดับนัยสำคัญ 0.05

4. จากการสุ่มนักศึกษาเพื่อสอบถามคะแนนวิชาทฤษฎีจำนวน (x) และวิชาสถิติ (y) จำนวน 5 คน แสดงดังตาราง

คะแนนวิชาทฤษฎีจำนวน	80	70	60	55	75
คะแนนวิชาสถิติ	75	90	66	60	84

4.1 จงสร้างตาราง ANOVA ให้สมบูรณ์

4.2 จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง

4.3 จงทดสอบว่าคะแนนวิชาทฤษฎีจำนวนและวิชาสถิติ มีความสัมพันธ์กันเชิงเส้นตรงหรือไม่ ที่ระดับนัยสำคัญ 0.05

5. จากข้อมูลแต่ละข้อต่อไปนี้ จงหาสัมประสิทธิ์สหสัมพันธ์ตัวอย่าง และทดสอบ $H_0 : \rho = 0$ แย้งกับ $H_1 : \rho \neq 0$ กำหนดระดับนัยสำคัญ 0.01

5.1 $n = 29, \;\; \sum(y_i - \bar{y})^2 = 64, \;\; \sum(x_i - \bar{x})^2 = 100, \;\; \sum(x_i - \bar{x})(y_i - \bar{y}) = 40$

5.2 $n = 18, \;\; \sum(y_i - \bar{y})^2 = 25, \;\; \sum(x_i - \bar{x})^2 = 144, \;\; \sum(x_i - \bar{x})(y_i - \bar{y}) = 54$

สรุป

ในบทนี้จะกล่าวถึงการถดถอยเชิงเส้นเชิงเดียว โดยใช้วิธีกำลังสองน้อยสุด เมื่อทราบตัวแปรต้นเราจะประมาณค่าหรือคาดคะเนตัวแปรตามได้ จากนั้นกล่าวถึงช่วงความเชื่อมั่นและการทดสอบสมมติฐานของ β โดยใช้การแจกแจงที และใช้การแจงเอฟเมื่อกำหนดตาราง ANOVA สุดท้ายกล่าวถึงสหสัมพันธ์ซึ่งเป็นค่าที่บอกถึงความสัมพันธ์ของตัวแปรต้นและตัวแปรตามว่าสัมพันธ์กันเชิงเส้นเส้นมากแค่ไหนโดยที่ $-1 \leq \rho \leq 1$

1. ถ้า ρ มีค่าเข้าใกล้ 1 มีหมายความว่าความสัมพันธ์ยิ่งใกล้เคียงเส้นตรงที่มีความชันเป็นบวก
2. ถ้า ρ มีค่าเข้าใกล้ -1 มีหมายความว่าความสัมพันธ์ยิ่งใกล้เคียงเส้นตรงที่มีความชันเป็นลบ
3. ถ้า ρ มีค่าเข้าใกล้ 0 มีหมายความว่าแทบไม่มีความสัมพันธ์แบบเส้นตรงหรือไม่มีเลย

แบบฝึกหัดบทที่ 10

1. กำหนดให้ข้อมูลดังตารางต่อไปนี้

x	1	2	3	3
y	1	3	4	6

จงหาสมการถดถอยเชิงเส้นของ Y เทียบ x

2. ข้อมูลส่วนสูง (เซนติเมตร) และน้ำหนัก (กิโลกรัม) ของนักเรียนหญิง 4 คนดังนี้

นักเรียนหญิง	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4
ส่วนสูง (เซนติเมตร)	150	152	154	156
น้ำหนัก (กิโลกรัม)	45	45	48	50

ถ้าส่วนสูงและน้ำหนักของนักเรียนมีความสัมพันธ์คือ $\hat{y} = a + 0.9x$ เมื่อ x เป็นความสูงและ y เป็นน้ำหนัก แล้วนักเรียนที่มีส่วนสูง 155 เซนติเมตร จะมีน้ำหนักกี่กิโลกรัม

3. ในการหาความสัมพันธ์เชิงฟังก์ชันระหว่างปริมาณสารปนเปื้อนชนิดที่ 1 (X) และปริมาณสารปนเปื้อนชนิดที่ 2 (Y) จากตัวอย่างสารอาหารจำนวน 100 ตัวอย่าง พบว่าความแปรปรวนของปริมาณสารชนิดที่ 1 เท่ากับ 1.75 ค่าเฉลี่ยเลขคณิตของปริมาณสารชนิดที่ 2 เท่ากับ 0.5

$$\sum_{i=1}^{100} x_i y_i = 100, \quad \text{และ} \quad \sum_{i=1}^{100} x_i^2 = 200$$

ถ้าสมการถดถอยคือ $\hat{y} = a + bx$ แล้ว เมื่อพบสารปนเปื้อนชนิดที่ 1 อยู่ 4 หน่วย จะพบสารปนเปื้อนชนิดที่ 2 (โดยประมาณ) เท่ากับเท่าใด

4. ถ้าสมการที่ใช้แทนความสัมพันธ์เชิงฟังก์ชันที่ใช้สำหรับการประมาณห้องพักที่มีแขกมาพักจริง (y) จากจำนวนห้องพักที่มีการจองล่วงหน้า (x) คือ

$$\hat{y} = a + 0.75x$$

โดยที่ $\bar{x} = 40$ และ $\bar{y} = 60$ ถ้า $x = 60$ แล้วจำนวนห้องพักที่มีแขกมาพักจริงโดยประมาณเท่ากับเท่าใด

5. ข้อมูลระหว่างเวลากับราคาปิดของหุ้น LUX มีความสัมพันธ์แบบเส้นตรง ดังข้อมูล

พ.ศ.	2550	2551	2552	2553	2554
ราคาปิด (บาท)	2.25	2.50	2.75	3.00	3.50

จงหาสมการถดถอยของความสัมพันธ์ระหว่างเวลากับราคาปิดของหุ้น LUX

6. ปริมาณการส่งออกทุเรียนสดของประเทศไทย แสดงดังตาราง

ปี พ.ศ.	2554	2555	2556	2557	2558	2559	2560
ปริมาณส่งออก (แสนตัน)	2.71	3.51	3.67	3.69	3.58	4.03	4.90

* กรมส่งเสริมการค้าระหว่างประเทศ กระทรวงพาณิชย์

จงทดสอบว่าสัมประสิทธิ์ถดถอย $\beta < 0.5$ หรือไม่ ที่ระดับนัยสำคัญ 0.05

7. ผลิตภัณฑ์มวลรวมในประเทศ (GDP) ของประเทศไทย แสดงดังตาราง

ปี พ.ศ.	2556	2557	2558	2559	2560	2561
GDP (พันล้าน USD)	420.3	407.3	401.3	412.4	455.3	505

* ข้อมูลจากธนาคารแห่งประเทศไทย

จงคาดการณ์ว่าปี 2563 ผลิตภัณฑ์มวลรวมในประเทศประมาณเท่าใด โดยใช้การถดถอยเชิงเส้น และจงทดสอบว่าสัมประสิทธิ์ถดถอย β เท่ากับ 16 หรือไม่ ที่ระดับนัยสำคัญ 0.05

8. ข้อมูลระหว่างเวลากับราคาปิดของหุ้น OR ของบริษัท ปตท. น้ำมันและการค้าปลีก จำกัด (มหาชน) โดยเสนอขายหุ้นใหม่ (IPO) ให้ประชาชนทั่วไปในราคาหุ้นละ 18 บาท เริ่มซื้อขายวันแรก 11 กุมภาพันธ์ 2564 ในตลาดหลักทรัพย์แห่งประเทศไทย โดยมีราคาปิดต่อหุ้นกับวันที่ซื้อขายดังข้อมูล

วันที่ซื้อขายในตลาด	1	2	3	4	5	6	7	8	9	10
ราคาปิด (บาท)	29.25	34.00	32.75	29.50	30.25	31.50	31.75	31.00	30.75	29.50

ตอบคำถามต่อไปนี้

8.1 จงสร้างตาราง ANOVA

8.2 หาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง จากตาราง ANOVA

8.3 จงทดสอบว่าวันที่ซื้อขายในตลาดและราคาปิดมีความสัมพันธ์กันเชิงเส้นตรงหรือไม่ โดยใช้สถิติเอฟ ที่ระดับนัยสำคัญ 0.05

บรรณานุกรม

กฤษณะ เนียมมณี. (2544). **ทฤษฎีความน่าจะเป็น** (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โรงพิมพ์พิทักษ์การพิมพ์.

________________. (2548). **ทฤษฎีความน่าจะเป็นขั้นสูง**. กรุงเทพฯ: โรงพิมพ์พิทักษ์การพิมพ์.

กัลยา วานิชบัญชา. (2552). **หลักสถิติ**. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

คณาจารย์ภาควิชาคณิตศาสตร์และวิทยาคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (2555). **ความน่าจะเป็นและสถิติ** (พิมพ์ครั้งที่ 13). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

โครงการตำราวิทยาศาสตร์และคณิตศาสตร์มูลนิธิ สอวน. (2552). **คอมบินาทอริก** (พิมพ์ครั้งที่ 2). กรุงเทพฯ: บริษัทด่านสุทธาการพิมพ์ จำกัด.

มนตรี สังข์ทอง. (2557). **หลักสถิติ**. กรุงเทพฯ: บริษัท ซีเอ็ดยูเคชัน จำกัด (มหาชน)

ยุทธ ไกยวรรณ์. (2558). **หลักสถิติวิจัยและการใช้โปรแกรม SPSS** (พิมพ์ครั้งที่ 4). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

รวีวรรณ ชินะตระกูล. (2545). **วิธีวิจัยการศึกษา** (พิมพ์ครั้งที่ 3). กรุงเทพฯ: บริษัทเฟื่องฟ้าพริ้นติ้ง จำกัด.

สมชาย ประสิทธิ์จูตระกูล. (2546). **ภินทนคณิตศาสตร์** (พิมพ์ครั้งที่ 2). กรุงเทพฯ: บริษัทด่านสุทธาการพิมพ์ จำกัด.

อิทธิพัทธ์ สุวทันพรกูล. (2562). **การวิจัยทางการศึกษา แนวคิดและการประยุกต์ใช้** (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye. (2017). **Probability and Statistics for Engineers and Scientists.** England : Peason Education, Inc.

Joel Hass, Christopher Heil and Maurice D. Weir. (2019). **Thomas' Calculus: Early Transcendentals.** Fourteenth Edition. USA. Pearson Education, Inc.

Michael Haese, Sandra Haese, Mark umphries, Edward Kemp and Pamela Vollma. (2014). **Mathematics for the international student 10E MYP5 (Extended).** Marleston, Australia : Haese& Harris Publications.

ดัชนี

ประวัติผู้เขียน

นายธนัชยศ จำปาหวาย

- ปริญญาเอก วิทยาศาสตร์ดุษฎีบัณฑิต (คณิตศาสาตร์), จุฬาลงกรณ์มหาวิทยาลัย, 2557
  Ph.D. (Mathematics), Chulalongkorn University, 2014
- ปริญญาโท วิทยาศาสตร์มหาบัณฑิต (คณิตศาสาตร์), จุฬาลงกรณ์มหาวิทยาลัย, 2552
  M.Sc. (Mathematics), Chulalongkorn University, 2009
- ปริญญาตรี วิทยาศาสตร์บัณฑิต (คณิตศาสาตร์, เกียรตินิยมอันดับสอง),
  จุฬาลงกรณ์มหาวิทยาลัย, 2549
  B.Sc. (Mathematics, 2nd class honours),Chulalongkorn University, 2006
- ปัจจุบันดำรงตำแหน่งผู้ช่วยศาสตราจารย์ประจำสาขาวิชาคณิตศาสตร์ คณะครุศาสตร์
  มหาวิทยาลัยราชภัฏสวนสุนันทา

Email: thanatyod_ja@ssru.ac.th
Office: 1144
Facebook: www.facebook.com/Jampawai
Block: www.eledu.ssru.ac.th/thanatyod_ja

ผลงานทางวิชาการ

1. E-book: หลักการคณิตศาสตร์สำหรับครู. (2565). www.mebmarket.com
2. E-book: ทฤษฎีจำนวน. (2565). www.mebmarket.com
3. E-book: พีชคณิตนามธรรม. (2565). www.mebmarket.com
4. หนังสือ: ความจริงที่ต้องพิสูจน์. (2560). ศูนย์หนังสือมหาวิทยาลัยราชภัฏสวนสุนันทา