SKYBOOK
ชื่อหนังสือ ปริญญาโทสถิติ พ.ศ.ดร. อนุสรณ์ และคณะ
บาร์โค้ด 9789743893100
ISBN 974-389-310-5

เส้นทางสู่ปริญญาโท

The Road to Master's Degree

คู่มือและแนวข้อสอบเข้าศึกษาต่อ ปริญญาโท

สถิติ

STATISTICS

- จุฬาลงกรณ์ฯ
- ม. ธรรมศาสตร์
- นิด้า (NIDA)
- ม. รามคำแหง
- ม. กรุงเทพ
- ม. หอการค้า
- และมหาวิทยาลัยอื่น ๆ ที่เกี่ยวข้อง

สำหรับ

- นักศึกษาระดับอุดมศึกษาใช้ทบทวน เพื่อเตรียมสอบประจำภาค หรือ เพื่อเตรียมสอบเข้าศึกษาต่อระดับปริญญาโทของทุกมหาวิทยาลัย

พลาดไม่ได้! ถ้าคุณ! ต้องการศึกษาต่อปริญญาโท

โดย... พ.ศ. ดร. อนุสรณ์ สรพรหม และคณะ

คู่มือและแนวข้อสอบ เข้าศึกษาต่อปริญญาโท

สถิติ

โดย... ผ.ศ. ดร. อนุสรณ์ สรพรหม และคณะ

บริษัท สกายบุ๊กส์ จำกัด
SKYBOOK COMPANY LIMITED
515/276-8 ถ.รังสิต-ปทุมธานี ต.ประชาธิปัตย์ อ.ธัญบุรี จ.ปทุมธานี 12130
โทร. 0-2958-1125-7, 0-2567-5119 โทรสาร. 0-2567-5105
E-mail: skybook1992@hotmail.com
www.skybook.co.th

"คู่มือและแนวข้อสอบเข้าศึกษาต่อปริญญาโท สถิติ"

พิมพ์ครั้งที่ 1 มกราคม 2546

สงวนลิขสิทธิ์ตามกฎหมาย
ห้ามคัดลอกถ่ายเอกสารหรือพิมพ์
หรือวิธีหนึ่งวิธีใดของหนังสือเล่มนี้ก่อนได้รับอนุญาต
จากบริษัท สกายบุ๊กส์ จำกัด

ราคา 195 บาท

ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ

อนุสรณ์ สรพรหม

คู่มือและแนวข้อสอบเข้าศึกษาต่อปริญญาโท สถิติ -- กรุงเทพฯ : สกายบุ๊กส์, 2546.

280 หน้า

1. คณิตศาสตร์สถิติ -- ข้อสอบและเฉลย I. ชื่อเรื่อง

519.5076

ISBN: 974-389-310-5

S7901-30-01-03

จัดพิมพ์และจำหน่ายโดย

บริษัท สกายบุ๊กส์ จำกัด
SKYBOOK COMPANY LIMITED
515/276-8 ถ.รังสิต-ปทุมธานี ต.ประชาธิปัตย์ อ.ธัญบุรี จ.ปทุมธานี 12130
โทร. 0-2958-1125-7, 0-2567-5119 โทรสาร. 0-2567-5105
E-mail: skybook1992@hotmail.com
www.skybook.co.th

พิมพ์ที่ บริษัท สยามสปอร์ต ซินดิเคท จำกัด
459 ซอยพิบูลย์อุปถัมภ์ (ลาดพร้าว 48) แขวงสามเสนนอก เขตห้วยขวาง กรุงเทพฯ 10310
โทรศัพท์ : 0-2694-3010

คำนำ

การศึกษาต่อในระดับปริญญาโทโดยเฉพาะสถาบันอุดมศึกษาของรัฐ มีการแข่งขันกันค่อนข้างสูง เนื่องจากจำนวนผู้ประสงค์จะศึกษาต่อมีมากกว่าจำนวนที่สถาบันจะรับได้หลายเท่า ด้วยเหตุนี้กระบวนการคัดเลือกนิสิตนักศึกษาจึงเป็นไปอย่างเข้มงวด โดยขั้นตอนที่สำคัญขั้นตอนหนึ่งคือการสอบคัดเลือกนั่น ผู้ที่ไม่ได้เตรียมตัวสอบเป็นอย่างดีก็มักจะพลาดโอกาสที่จะได้ศึกษาต่อ

สำนักพิมพ์สกายบุ๊กส์ตระหนักถึงความสำคัญดังกล่าว จึงขอให้ข้าพเจ้าและคณะช่วยกันสร้างผลงานในรูปคู่มือและแนวข้อสอบจำนวน 10 เรื่อง เพื่อช่วยให้การเตรียมการสอบคัดเลือกในระดับปริญญาโทมีประสิทธิภาพและประสิทธิผลมากขึ้น โดยเนื้อหาของหนังสือทั้ง 10 เรื่องจะเน้นในการทบทวนความรู้ในวิชาต่าง ๆ ที่จำเป็นต่อการสอบ เนื่องจากความเชื่อที่ว่า "ความรู้เป็นอาวุธที่ดีที่สุดสำหรับเอาชนะข้อสอบ" ไม่ใช่ความจำ และมีแนวข้อสอบให้ฝึกทำบ้างเพื่อทดสอบความรู้ของตน แถมด้วยข้อมูลเกี่ยวกับสถาบันการศึกษาทั้งของรัฐและเอกชนที่เปิดหลักสูตรต่าง ๆ ซึ่งผู้สนใจสามารถใช้ประโยชน์จากข้อมูลเหล่านี้ได้โดยตรง

ในโอกาสนี้ ข้าพเจ้าขอขอบคุณทุกฝ่ายของสำนักพิมพ์สกายบุ๊กส์ที่ทำให้หนังสือเล่มนี้เสร็จสมบูรณ์ ขอขอบคุณผู้ร่วมทีมทุกท่านดังรายนามบนปกหลังของหนังสือทุกเล่มในชุดนี้ และที่จะขาดไม่ได้คือ คุณพัชราวลัย และเด็กชายสรณ์พัชร สรพรหม ภรรยาและบุตรที่เข้าใจและเอาใจใส่ข้าพเจ้าด้วยดีตลอดมา

อนุสรณ์ สรพรหม
6 มกราคม 2546

โทรศัพท์ 01-259-4830
E-mail: sornpohm@ksc.th.com

สารบัญ

หน้า

บทที่ 1 บทนำ..1
แบบฝึกหัด..6
เฉลยแบบฝึกหัด..8

บทที่ 2 สถิติพรรณนา : การวัดแนวโน้มสู่ส่วนกลาง..9
แบบฝึกหัด..33
เฉลยแบบฝึกหัด..37

บทที่ 3 สถิติพรรณนา : การวัดการกระจาย..45
แบบฝึกหัด..58
เฉลยแบบฝึกหัด..61

บทที่ 4 วิธีเรียงสับเปลี่ยนและวิธีจัดหมู่..66
แบบฝึกหัด..74
เฉลยแบบฝึกหัด..76

บทที่ 5 ความน่าจะเป็น..82
แบบฝึกหัด..89
เฉลยแบบฝึกหัด..92

บทที่ 6 การแจกแจงความน่าจะเป็น..99
แบบฝึกหัด..130
เฉลยแบบฝึกหัด..134

บทที่ 7 การชักตัวอย่าง..142
แบบฝึกหัด..167
เฉลยแบบฝึกหัด..170

บทที่ 8 การประมาณค่าเชิงสถิติ..174
แบบฝึกหัด..189
เฉลยแบบฝึกหัด..192

บทที่ 9 การทดสอบสมมุติฐาน..196
แบบฝึกหัด..219
เฉลยแบบฝึกหัด..223

บทที่ 10 การวิเคราะห์ความแปรปรวน..232
แบบฝึกหัด..244
เฉลยแบบฝึกหัด..248

บทที่ 11 การทดสอบไคกำลังสอง..253
แบบฝึกหัด..268
เฉลยแบบฝึกหัด..271

1 บทนำ

1.1. ความหมายของสถิติ

วิชาสถิติ (Statistics) เป็นวิชาที่ศึกษาเกี่ยวกับการออกแบบ การเก็บรวบรวมข้อมูล แล้วนำมาจัดให้เป็นหมวดหมู่ สรุป ตีความ และนำเสนอข้อมูลที่ได้เพื่อช่วยในการตัดสินใจ

1.2 คำศัพท์ที่สำคัญทางสถิติ

ข้อมูล (Data) คือ เซตของจำนวนซึ่งแสดงลักษณะทั้งที่เป็นตัวเลขและไม่เป็นตัวเลขของสิ่งใดสิ่งหนึ่ง เช่น จำนวนผู้หญิง ผู้ชาย และเด็กในหมู่บ้านแห่งหนึ่งเป็นการแสดงลักษณะที่เป็นตัวเลข ขณะที่เชื้อชาติของคนในหมู่บ้านเป็นการแสดงลักษณะที่ไม่เป็นตัวเลข เป็นต้น

พารามิเตอร์ (Parameter) คือ จำนวนที่ใช้พรรณนาลักษณะของประชากร เช่น ราคาเฉลี่ยของสินค้าทั้งหมดที่วางขายในห้างสรรพสินค้า หรือสัดส่วนของหลอดไฟที่เสียจากการผลิตของโรงงาน A เป็นต้น

ประชากร (Population) คือ สมาชิกทั้งหมดในกลุ่มซึ่งต้องการศึกษา เช่น ราคาสินค้าทุกชนิดที่วางขายในห้างสรรพสินค้า หรือจำนวนหลอดไฟทั้งหมดที่ผลิตโดยโรงงาน A เป็นต้น

ตัวอย่าง (Sample) คือ เซตย่อยของประชากรซึ่งชักมาจากประชากรที่เราต้องการศึกษา เช่น ราคาสินค้าบางชนิดที่ชักมาจากสินค้าทั้งหมดที่วางขายในห้างสรรพสินค้า หรือจำนวนหลอดไฟจำนวนหนึ่งที่ชักมาจากหลอดไฟทั้งหมดที่ผลิตโดยโรงงาน A เป็นต้น

สถิติ (Statistic) คือ จำนวนที่ใช้พรรณนาลักษณะของตัวอย่าง เช่น ราคาเฉลี่ยของสินค้าในตัวอย่าง หรือสัดส่วนของหลอดไฟที่เสียในตัวอย่าง เป็นต้น

การอนุมานเชิงสถิติ (Statistical Inference) คือ กรรมวิธีในการประมาณค่าพารามิเตอร์ของประชากรจาก *สารสนเทศ* (Information) ของตัวอย่าง เช่น เราต้องการหาราคาเฉลี่ยของสินค้าทั้งหมด 6,000 ชนิดในห้างสรรพสินค้าแห่งหนึ่ง แทนที่จะเสียเวลาในการตรวจสอบราคาเฉลี่ยที่แท้จริงว่าเท่ากับเท่าไร เราอาจใช้วิธีการทางสถิติเพื่อชักตัวอย่างสินค้ามาจำนวน 100 ชนิด แล้วหาราคาเฉลี่ยของสินค้าในตัวอย่างดังกล่าวซึ่งสมมุติว่าได้เท่ากับ 1,300 บาท และโดยอาศัยการอนุมานทางสถิติเราสามารถสรุปได้ว่าราคาเฉลี่ยที่แท้จริงของสินค้าทั้งหมดจะมีค่าประมาณ 1,300 บาท เป็นต้น

1.3 ชนิดของประชากร

ในการสำรวจตัวอย่างนั้นเราจะเริ่มจากการกำหนด *กรอบ* (Frame) ซึ่งเป็นรายการของสมาชิกทั้งหมดในประชากรขึ้นก่อน ตัวอย่างเช่น กรณีศึกษาอัตราการว่างงานของประเทศ A กรอบของการศึกษาคือรายการของผู้ที่อยู่ในกำลังแรงงานทั้งหมด เป็นต้น เมื่อได้กรอบที่ต้องการแล้ว เราจึงทำการสำรวจ *สำมะโนประชากร* (Census) เพื่อสำรวจหาสมาชิกทั้งหมดที่อยู่ในกรอบ ทั้งนี้ประชากรที่ต้องการศึกษาอาจจำแนกออกเป็น *สารสนเทศเชิงปริมาณ* (Quantitative Information) เช่น ราคาสินค้าทั้งหมดที่ขายในห้างสรรพสินค้า หรือจำนวนหลอดไฟที่ชำรุดทั้งหมด ฯลฯ ซึ่งเป็นจำนวนที่สามารถมีค่าเป็นบวก ลบ หรือศูนย์ก็ได้ กับ *สารสนเทศเชิงคุณภาพ* (Qualitative Information) เช่น ระดับการศึกษาของคนงาน เพศ หรือเชื้อชาติ ฯลฯ ซึ่งไม่เป็นตัวเลข

นอกจากนั้นเราอาจแบ่งประชากรออกเป็น *ประชากรจำกัด* (Finite population) เมื่อจำนวนของสมาชิกของประชากรมีจำกัดซึ่งรู้ว่าสมาชิกสุดท้ายของประชากรเป็นเท่าไร เช่น จำนวนแรงงานในตำบลใดตำบลหนึ่ง หรือจำนวนปลาที่เลี้ยงไว้ในบ่อหลังบ้าน ฯลฯ และ *ประชากรไม่จำกัด* (Infinite population) เมื่อจำนวนสมาชิกของประชากรมีไม่จำกัดหรือไม่รู้ว่าสมาชิกสุดท้ายของประชากรเป็นเท่าไร เช่น จำนวนแรงงานทั้งหมดของประเทศจีน หรือจำนวนปลาในมหาสมุทร เป็นต้น

1.4 ชนิดของข้อมูล

เราอาจแบ่งชนิดของข้อมูลออกได้เป็น 4 ชนิดตามมาตราส่วนของการวัด ดังนี้

(1) **ข้อมูลตามตัวเลข** (Nominal Data) คือ ค่าสังเกตซึ่งจำแนกออกตามประเภทใดประเภทหนึ่งโดยเฉพาะ เช่น บริษัทแห่งหนึ่งต้องการลงทุนทำศูนย์การค้าโดยเลือกว่าจะสร้างที่ไหนจากสถานที่สองแห่ง (A กับ B) ข้อมูลตามตัวเลขของสถานที่สองแห่งนี้อาจเป็นจำนวนผู้ชายและผู้หญิงที่อาศัยอยู่ในละแวกไม่เกิน 5 กิโลเมตรจากสถานที่แต่ละแห่ง การจัดประเภทของข้อมูลในลักษณะนี้จะต้องให้คำนิยามของข้อมูลอย่างชัดเจน (ในที่นี้คือผู้ชายและผู้หญิง) เพื่อที่ค่าสังเกตแต่ละค่า (ในที่นี้คือคนแต่ละคน) จะถูกจัดอยู่ในประเภทใดประเภทหนึ่งเท่านั้น

(2) **ข้อมูลเชิงอันดับที่** (Ordinal Data) คือ ค่าสังเกตซึ่งมีลำดับที่ สามารถระบุการเป็นที่หนึ่ง ที่สอง ที่สาม และลำดับต่อ ๆ ไปได้ เช่น อันดับของเมืองที่น่าอยู่ที่สุดในโลก คือ A, B, C, . . . ซึ่งจะบอกให้เราทราบว่าเมือง A น่าอยู่กว่าเมือง B และเมือง B น่าอยู่กว่าเมือง C ไปเรื่อย ๆ แต่ไม่สามารถบอกออกมาเป็นตัวเลขระหว่างค่าสังเกต A กับ B และ B กับ C ว่าต่างกันมากน้อยเพียงไร หรือผลการแข่งขันว่ายน้ำชาย 4 x 100 เมตร คือ นาย ก. นาย ข. และนาย ค. ซึ่งได้เหรียญทอง เหรียญเงิน

และเหรียญทองแดงตามลำดับ ข้อมูลนี้บอกเพียงแต่ว่าใครได้อันดับที่เท่าไร แต่ไม่สามารถบอกได้ว่านาย ก. ชนะนาย ข. มากหรือน้อย เช่นเดียวกับไม่สามารถบอกได้ว่านาย ข. ชนะนาย ค. มากหรือน้อยเพียงไร เป็นต้น

(3) **ข้อมูลแบบช่วง** (Interval Data) คือ ค่าสังเกตที่วัดบนมาตราส่วนซึ่งมีระยะห่างที่เท่ากัน โดย 0 เป็นค่าที่กำหนดขึ้น เช่น การวัดอุณหภูมิเป็นองศาฟาเรนไฮต์เป็นข้อมูลแบบช่วงซึ่งค่า 0 ถูกกำหนดให้เป็นอุณหภูมิที่เกิดจากการผสมหิมะและเกลือในสัดส่วนที่เท่ากัน (0 องศาไม่ได้หมายความว่าไม่มีอุณหภูมิเลย) โดยระยะห่างที่เท่ากันของจำนวนในแต่ละช่วงแสดงว่าการเปลี่ยนแปลงของค่าสังเกตเท่ากัน เช่น อุณหภูมิ 60 องศาจะห่างจากอุณหภูมิ 70 องศาเท่ากับอุณหภูมิ 70 องศาห่างจากอุณหภูมิ 80 องศา เป็นต้น

(4) **ข้อมูลแบบอัตราส่วน** (Ratio Data) คือ ค่าสังเกตที่วัดบนมาตราส่วนซึ่งระยะห่างที่เท่ากัน โดย 0 แสดงว่าค่านั้นไม่มี เช่น อายุของคนเป็นข้อมูลแบบอัตราส่วน เนื่องจาก 0 เป็นการเริ่มต้นถือกำเนิด และคนที่มีอายุ 60 ปีจะแก่กว่าคนที่มีอายุ 50 ปีเท่ากับคนที่มีอายุ 50 ปีแก่กว่าคนที่มีอายุ 40 ปี เป็นต้น นอกจากนี้ระยะทางระหว่างจังหวัดสองจังหวัด หรือจำนวนผลไม้ในสวน หรือจำนวนเวลาที่ใช้ในการเดินทางจากบ้านถึงที่ทำงาน ต่างก็เป็นข้อมูลแบบอัตราส่วนเช่นกัน

ความแตกต่างระหว่างข้อมูลแบบช่วงกับข้อมูลแบบอัตราส่วนอยู่ตรงที่ข้อมูลแบบอัตราส่วนสามารถคำนวณหาอัตราส่วนระหว่างจำนวนได้ เช่น คนที่มีอายุ 80 ปีจะมีอายุเป็น 2 เท่าของคนที่มีอายุ 40 ปี หรือคนที่มีอายุ 15 ปีจะมีอายุเป็น ½ เท่าของคนที่มีอายุ 30 ปี ฯลฯ แต่ข้อมูลแบบช่วงนั้นไม่สามารถคำนวณหาอัตราส่วนระหว่างกันได้ เช่น อุณหภูมิ 80 องศาฟาเรนไฮต์ไม่ถือว่าเป็น 2 เท่าของอุณหภูมิ 40 องศาฟาเรนไฮต์เนื่องจาก 0 เป็นค่าที่สมมติขึ้นให้เป็นค่าใด ๆ ก็ได้ โดยข้อมูลแบบช่วงและแบบอัตราส่วนอาจเป็น *ข้อมูลต่อเนื่อง* (Continuous Data) ซึ่งการสังเกตจะรวมทุก ๆ ค่าที่อยู่ในช่วงหรือ *ข้อมูลไม่ต่อเนื่อง* (Discrete Data) ซึ่งการสังเกตจะทำเฉพาะค่าที่จำกัดเฉพาะจุดต่าง ๆ ในช่วง เช่น ความยาวของเท้ามีหน่วยเป็นเซนติเมตรเป็นข้อมูลต่อเนื่องเพราะว่าเราสามารถระบุเป็นจำนวนทศนิยมของความยาวได้ทุกค่า ส่วนขนาดของรองเท้าเช่น เบอร์ 8 กับ 8 ½ นั้นเป็นตัวอย่างของข้อมูลไม่ต่อเนื่อง เพราะว่าไม่มีขนาดของรองเท้าระหว่างค่าทั้งสอง เป็นต้น

1.5 สถิติพรรณนาและสถิติวิเคราะห์

สถิติพรรณนา (Descriptive Statistics) มุ่งเน้นไปที่การสรุปและพรรณนาข้อมูลชุดใดชุดหนึ่ง เช่น ทุกๆ 10 ปี สำนักงานสถิติแห่งชาติจะทำการสำรวจสำมะโนประชากรโดยการเก็บข้อมูลเกี่ยวกับ

จำนวน อายุ อาชีพ การศึกษา ฯ ของประชากร เนื่องจากข้อมูลดิบที่เก็บได้มีมาก จึงจำเป็นที่จะต้องทำการย่อและตีความสารสนเทศเหล่านี้เพื่อสามารถนำไปใช้ประโยชน์ต่อไป

สถิติวิเคราะห์ (Analytical Statistics) มุ่งเน้นไปที่การใช้เทคนิคต่าง ๆ เพื่อช่วยในการตัดสินใจอย่างมีเหตุมีผลภายใต้ความไม่แน่นอน เช่น การวิเคราะห์ข้อมูลเพื่อใช้ในการตัดสินใจทางธุรกิจ ทางเศรษฐศาสตร์ หรือการตัดสินใจดำเนินงานของรัฐบาล ซึ่งจะถูกต้องมากน้อยแค่ไหนนั้นขึ้นอยู่กับวิธีการจัดเก็บข้อมูลดังกล่าวเป็นต้น

1.6 วิธีการแก้ปัญหาทางสถิติ

ขั้นตอนในการแก้ปัญหาทางสถิติประกอบด้วย

(1) **ระบุปัญหา** (Identifying the Problem) ผู้ที่ศึกษาจะต้องเข้าใจอย่างแจ่มแจ้งและนิยามได้อย่างถูกต้องว่าเราต้องการอะไรจากการศึกษา เช่น ต้องให้แน่ใจว่าเป้าหมายของการศึกษาประชากรนั้นเพื่ออะไร เพื่อประมาณค่าเฉลี่ยหรือร้อยละ หรือเพื่อดำเนินการบางอย่างกับกลุ่ม แล้วจึงวัดการตอบสนองว่าเป็นอย่างไร เป็นต้น

(2) **รวบรวมข้อมูล** (Gathering the Data) การเก็บข้อมูลนั้นถ้าเป็นไปได้จะต้องเป็นข้อมูลปฐมภูมิ (เป็นการเก็บข้อมูลครั้งแรกโดยตรง) มากกว่าข้อมูลปฐมภูมิ (ซึ่งมีผู้เก็บมาก่อนแล้ว) ทั้งนี้การเก็บจะต้องมี *ความเที่ยง* (Precision) ทำได้ในเวลาที่เหมาะสม สมบูรณ์ที่สุดเท่าที่เป็นไปได้ และเกี่ยวเนื่องกับปัญหาที่เราต้องการแก้

การเก็บอาจทำได้โดยการสัมภาษณ์ การส่งแบบสอบถาม ฯลฯ รวมทั้งมี *การชักตัวอย่าง* (Sampling) จากประชากรโดยใช้ระเบียบวิธีทางสถิติ

(3) **จัดประเภทและสรุปข้อมูล** (Classifying and Summarizing the Data) หลังจากรวบรวมข้อมูลที่ต้องการได้แล้ว ขั้นต่อไปจะเป็นการจัดหมวดหมู่หรือจัดกลุ่มข้อมูลที่ได้เพื่อใช้ในการศึกษา เช่น ข้อมูลการผลิตที่เก็บมาได้ จะถูกจำแนกออกตามผลผลิต ตามแหล่งผลิต และตามกระบวนการผลิต เป็นต้น แล้วจึงสรุปในรูปของตาราง แผนภูมิ การวัดแนวโน้มเข้าสู่ส่วนกลาง (เช่น ค่ามัชฌิม) และการวัดการกระจาย (เช่น ค่าเบี่ยงเบนมาตรฐาน)

(4) **นำเสนอและวิเคราะห์ข้อมูล** (Presenting and Analyzing the Data) การสรุปข้อมูลในขั้นที่ผ่านมาทำให้เข้าใจปัญหาได้ง่ายขึ้น สารสนเทศดังกล่าวจะช่วยระบุความสัมพันธ์และช่วยให้ผู้ทำการศึกษาสามารถนำเสนอจุดสำคัญ ๆ ให้ผู้อื่นได้ทราบด้วย สำหรับการวิเคราะห์ข้อมูลทำได้โดยการตี

ความผลลัพธ์จากขั้นที่ 3 และใช้การคำนวณเพื่อเป็นพื้นฐานในการอนุมานเชิงสถิติ รวมทั้งการใช้เครื่องมือทางสถิติเพื่อระบุแนวทางในการแก้ปัญหา

(5) **ทำการตัดสินใจ** (Making the Decision) ขั้นสุดท้ายคือการเปรียบเทียบแนวทางในการแก้ปัญหาแล้วตัดสินใจเลือกแนวทางที่ดีที่สุดเพื่อให้สำเร็จตามเป้าหมายที่ตั้งไว้ โดยการเลือกนี้จะถูกต้องหรือไม่ย่อมขึ้นอยู่กับความช่ำชองของผู้ทำการศึกษาและคุณภาพของสารสนเทศเป็นสำคัญ

แบบฝึกหัด

1. ถ้าบริษัทน้ำมันแห่งหนึ่งต้องการหาว่าสถานีบริการน้ำมันของบริษัทจำนวนทั้งหมด 520 แห่ง มีร้อยละเท่าไรที่สามารถทำรายได้มากกว่า 500,000 บาทต่อเดือน ท่านคิดว่าข้อความต่อไปนี้ข้อใดที่ผิด
   (ก) ประชากรที่กำลังศึกษาเป็นประชากรจำกัด
   (ข) ประชากรที่กำลังศึกษาเป็นสารสนเทศเชิงปริมาณ
   (ค) สมาชิกของประชากรที่กำลังศึกษาคือสถานีบริการแต่ละสถานี
   (ง) กรอบของประชากรที่กำลังศึกษาคือรายชื่อของสถานีบริการทั้งหมด

2. เกษตรกรที่ได้รับผลกระทบจากการสร้างเขื่อนจำนวน 257 คน ได้ทำการประท้วงรัฐบาลหน้าที่ทำการอำเภอแห่งหนึ่ง ถ้าผู้สื่อข่าวเข้าไปสัมภาษณ์เกษตรกรแต่ละคนว่าพวกเขาคิดว่าการประท้วงนี้ทางรัฐบาลจะให้ความช่วยเหลือเป็นอย่างดีหรือไม่ ข้อความต่อไปนี้ข้อใดถูก
   (ก) ผลการสัมภาษณ์ถือว่าเป็นตัวแทนของเกษตรกรทั้งประเทศ
   (ข) ความคิดเห็นที่ได้เป็นความคิดเห็นที่ *ไม่เอียงเอน* (Unbiased)
   (ค) ประชากรที่กำลังศึกษาเป็นประชากรจำกัด
   (ง) ประชากรที่กำลังศึกษาเป็นสารสนเทศเชิงปริมาณ

3. ข้อมูลต่อไปนี้ข้อใดเป็นข้อมูลเชิงอันดับที่
   (ก) จำนวนผู้เข้าชมภาพยนตร์จำแนกเป็นชาย หญิง เด็กผู้ชาย และเด็กผู้หญิง
   (ข) รายชื่อผู้ได้ตำแหน่งนางงามสุริยจักรวาล และรองทั้งสี่
   (ค) คะแนนเฉลี่ยของนักศึกษาในวิทยาลัยแห่งหนึ่ง
   (ง) ระดับน้ำในแม่น้ำเจ้าพระยาในแต่ละวันของปี 2545

4. ข้อมูลต่อไปนี้ข้อใดที่เป็นข้อมูลแบบช่วงและไม่ต่อเนื่อง
   (ก) จำนวนนักศึกษาจำแนกตามชั้นปีที่หนึ่ง ชั้นปีที่สอง ชั้นปีที่สาม และชั้นปีที่สี่
   (ข) วันเกิดของนักศึกษาคณะสังคมศาสตร์ มหาวิทยาลัยมหิดล

(ค) ราคาน้ำมันของสถานีบริการ A แต่ละวันในรอบปีที่ผ่านมา

(ง) เวลาที่ใช้ในการเดินทางจากบ้านไปวิทยาลัยของนักศึกษา 65 คน

5. การแก้ปัญหาทางสถิติประกอบด้วย (1) นำเสนอและวิเคราะห์ข้อมูล (2) จัดประเภทและสรุปข้อมูล (3) รวบรวมข้อมูล (4) ระบุปัญหา (5) ทำการตัดสินใจ ถ้าจะเรียงลำดับขั้นตอนให้ถูกต้องคำตอบควรเป็นข้อใด

   (ก) 1, 2, 3, 4, 5

   (ข) 2, 4, 5, 1, 3

   (ค) 3, 2, 1, 4, 5

   (ง) 4, 3, 2, 1, 5

เฉลยแบบฝึกหัด

1. ตอบข้อ (ข)

   เนื่องจากประชากรที่กำลังศึกษาเป็นสารสนเทศเชิงคุณภาพ เพราะสถานีบริการแต่ละสถานีจะให้ข้อมูลเพียงแต่ยอดจำหน่ายว่ามากกว่าหรือน้อยกว่า 500,000 บาทต่อเดือนเท่านั้น

2. ตอบข้อ (ค)

   เนื่องจากประชากรมีจำนวนที่แน่นอนคือ 257 คน

   ส่วนข้อ (ก) ผิดเพราะว่าผลการสัมภาษณ์ถือเป็นตัวแทนของกลุ่ม 257 คนนี้เท่านั้น ข้อ (ข) ผิดเพราะว่าความคิดเห็นที่ได้น่าจะเอนเอียง เนื่องจากเป็นการสัมภาษณ์ผู้ที่ได้รับความเดือดร้อนและกำลังประท้วงอยู่ ข้อ (ง) ผิดเพราะว่าเป็นสารสนเทศเชิงคุณภาพ

3. ตอบข้อ (ข)

   ส่วนข้อ (ก) เป็นข้อมูลตามตัวเลข ข้อ (ค) และ (ง) เป็นข้อมูลแบบอัตราส่วน

4. ตอบข้อ (ข)

   เนื่องจากวันที่ศูนย์ในกรณีนี้ไม่ได้หมายถึงปีที่ 0 แต่เป็นการเริ่มต้นปฏิทินซึ่งแต่ละศาสนาก็นับไม่เหมือนกัน และที่ไม่ต่อเนื่องเพราะว่าระบุเป็นช่วงของวันเท่านั้น

   ส่วนข้อ (ก) เป็นข้อมูลตามตัวเลข ข้อ (ค) และ (ง) เป็นข้อมูลแบบอัตราส่วน

5. ตอบข้อ (ข)

2 สถิติพรรณนา : การวัดแนวโน้มสู่ส่วนกลาง

2.1 การจัดองค์ข้อมูลและการแจกแจงความถี่

(1) การจัดองค์ข้อมูล (Data Organization)

ข้อมูลจากแหล่งปฐมภูมิที่เก็บมาได้และยังไม่ได้จัดข้อมูลนั้นเป็น *ข้อมูลดิบ* (Raw Data) ที่ใช้ประโยชน์ในการตัดสินใจไม่ได้มากนัก ตัวอย่างเช่น ข้อมูลดิบเดือนตุลาคมของยอดขายยาปราบศัตรูพืชที่พนักงานขายในบริษัท A จำนวน 30 คนขายได้ในตารางที่ 2.1 ไม่มีความหมายมากนักในเชิงสถิติและยากต่อการเข้าใจ ดังนั้นก่อนที่จะศึกษาต่อไปสิ่งแรกที่ควรทำคือจัดองค์ข้อมูลซึ่งวิธีที่ง่ายที่สุดได้แก่ การจัดข้อมูลดิบให้เป็น *แถวลำดับ* (Array) จากค่าต่ำสุดไปหาค่าสูงสุดหรือจากค่าสูงสุดไปหาค่าต่ำสุดก็ได้ ดังตัวอย่างในตารางที่ 2.2 ซึ่งเป็นการจัดข้อมูลดิบในตารางที่ 2.1 โดยเรียงลำดับจากยอดขายต่ำสุดไปหายอดขายสูงสุด

ตารางที่ 2.1 ยอดขายยาปราบศัตรูพืชของพนักงานขายในบริษัท A (หน่วยเป็นตัน)

พนักงานขาย	ยอดขาย	พนักงานขาย	ยอดขาย	พนักงานขาย	ยอดขาย
1	81	11	50	21	41
2	83	12	81	22	58
3	78	13	86	23	43
4	76	14	74	24	76
5	74	15	79	25	78
6	62	16	71	26	71
7	93	17	61	27	72
8	99	17	75	28	53
9	51	19	73	29	60
10	75	20	71	30	65

ตารางที่ 2.2 แถวลำดับยอดขายยาปราบศัตรูพืชของพนักงานขายในบริษัท A (หน่วยเป็นตัน)

41	71	76
43	71	78
50	71	78
51	72	79
53	73	81
58	74	81
60	74	83
61	75	86
62	75	93
65	76	99

ซึ่งการจัดข้อมูลดิบให้อยู่ในรูปของแถวลำดับในตารางที่ 2.2 จะมีประโยชน์หลายประการ เช่น

(ก) ยอดขายของพนักงานจะมีค่าระหว่าง 41 ถึง 99 ตัน โดยมี *พิสัย* (Range) ซึ่งเป็นผลต่างระหว่างค่าสูงสุดและต่ำสุดของข้อมูลเท่ากับ 99 – 41 = 58 ตัน

(ข) ยอดขายที่ต่ำกว่า 50 % (ดูจากคนที่ขายได้น้อยที่สุดถึงคนที่ 15) จะอยู่ระหว่าง 41 และ 73 ตัน ส่วนยอดขายที่สูงกว่า 50% (ดูจากคนที่ 16 ถึงคนที่ขายได้สูงที่สุด) จะอยู่ระหว่าง 74 และ 99 ตัน

(ค) แถวลำดับจะแสดงให้เห็นการกระจุกตัวของข้อมูลที่ค่าใดค่าหนึ่งหรือหลายค่า เช่น จากตัวเลขในตารางที่ 2.2 จะเห็นได้ว่ายอดขายที่กระจุกตัวมากที่สุดคือ 71 ตัน (ขายได้ 3 คน) รองลงมาคือ 74, 75, 76, 78 และ 81 ตัน (ขายได้อย่างละ 2 คน)

(2) การแจกแจงความถี่ (Frequency Distribution)

แม้ว่าแถวลำดับจะมีประโยชน์แล้วก็ตาม แต่การจัดองค์ข้อมูลโดยวิธีนี้จะมีปัญหาและดูได้ยากเมื่อข้อมูลมีขนาดใหญ่มาก ด้วยเหตุนี้เราจึงต้องจัดให้อยู่ในรูปที่กระชับขึ้นเพื่อช่วยในการวิเคราะห์และนำเสนอ โดยวิธีที่นิยมใช้กันได้แก่ การแจกแจงความถี่

การแจกแจงความถี่เป็นการจัดกลุ่มข้อมูลให้อยู่ในรูปของ *ชั้น* (Class) แล้วลงจำนวนของข้อมูลในแต่ละชั้น ดังตารางที่ 2.3 เราจัดกลุ่มยอดขายยาปราบศัตรูพืชออกเป็น 5 ชั้น คือ ชั้นที่ 1 ยอดขาย 40 ตันและน้อยกว่า 52 ตัน ชั้นที่ 2 ยอดขาย 52 ตันและน้อยกว่า 64 ตัน ชั้นที่ 3 ยอดขาย 64 ตันและน้อยกว่า 76 ตัน ชั้นที่ 4 ยอดขาย 76 ตันและน้อยกว่า 88 ตัน ชั้นที่ 5 ยอดขาย 88 ตันและน้อยกว่า 100 ตัน แล้วจึงใส่จำนวนพนักงานที่ขายได้ในแต่ละชั้นลงไปในตาราง

ตารางที่ 2.3 การแจกแจงความถี่ยอดขายยาปราบศัตรูพืชของพนักงานขายในบริษัท A

จำนวนที่ขายได้ (ตัน)	จำนวนพนักงานขาย (ความถี่)
40 และน้อยกว่า 52	4
52 และน้อยกว่า 64	5
64 และน้อยกว่า 76	10
76 และน้อยกว่า 88	9
88 และน้อยกว่า 100	2
	30

ข้อดีของตารางที่ 2.3 คือ เราสามารถเห็นภาพรวมของยอดขายในเดือนตุลาคมได้ดียิ่งขึ้น เช่น พนักงานหนึ่งในสามของทั้งหมดขายได้ระหว่าง 64 ถึง 76 ตัน (ความถี่เท่ากับ 10 จากจำนวนเต็ม 30) หรือมีพนักงาน ครึ่งหนึ่งที่ขายได้ระหว่าง 52 ถึง 76 ตัน (ความถี่เท่ากับ 5 + 10 จากจำนวนเต็ม 30 คน) แต่ขณะเดียวกันตารางแจกแจงความถี่จะมีข้อเสียคือเราไม่สามารถบอกได้ว่าพนักงานขายแต่ละคนขายได้คนละเท่าไร และไม่สามารถบอกได้ว่าพิสัยมีค่าเท่ากับ 56 ตันเนื่องจากไม่รู้ค่าสูงสุดและต่ำสุด เพียงแต่บอกได้ว่ามีผู้ขายได้ระหว่าง 88 ถึง 100 ตันอยู่ 2 คนในชั้นที่ 5 และพิสัยจะมีค่าประมาณ 60 (100 - 40) ตัน เป็นต้น

การสร้างตารางแจกแจงความถี่ของข้อมูลใดก็ตามจะต้องกำหนดค่าต่างๆ ต่อไปนี้คือ (1) จำนวนชั้นที่กำหนด ซึ่งโดยทั่วไปจะอยู่ระหว่าง 5 ถึง 18 ชั้น (หรือใช้จำนวน k ชั้นซึ่งคำนวณจาก $2^k \geq$ จำนวนค่าสังเกตทั้งหมด) แล้วแต่ความเหมาะสม ทั้งนี้จะต้องครอบคลุมค่าต่ำสุดและสูงสุด และค่าสังเกตทุกตัวจะต้องจัดลงในชั้นใดชั้นหนึ่งได้เพียงชั้นเดียวเท่านั้น (2) ความกว้างของแต่ละชั้นซึ่งเราเรียกว่า " *อันตรภาคชั้น* (Class Interval)" นั้น สามารถหาได้โดยการแทนค่าในสูตร *ความกว้างของชั้น*

= *ค่าพิสัย / จำนวนชั้น* และจากผลหารที่ได้เราจะปัดให้เป็นจำนวนเต็ม (จากตารางที่ 2.2 ค่าพิสัยเท่ากับ 58 หารด้วยจำนวนชั้นเท่ากับ 5 ผลลัพธ์คือ 11.6 เมื่อปัดให้เป็นจำนวนเต็มจะได้ความกว้างของชั้นเท่ากับ 12)

การแบ่งชั้นตัวอย่างในตารางที่ 2.3 เราสามารถสรุป (1) *ค่ากึ่งกลางของชั้น* (Class Midpoint) ซึ่งเท่ากับผลรวมของขีดจำกัดบนและขีดจำกัดล่างหารด้วยสอง (2) *ขีดจำกัดล่างของชั้น* (The Lower Limit of the Class) และ (3) *ขีดจำกัดบนของชั้น* (The Upper Limit of the Class) ได้ดังต่อไปนี้

	ค่ากึ่งกลางของชั้น	**ขีดจำกัดล่างของชั้น**	**ขีดจำกัดบนของชั้น**
ชั้นที่ 1	46	40	51.999…
ชั้นที่ 2	58	52	63.999…
ชั้นที่ 3	70	64	75.999…
ชั้นที่ 4	82	76	87.999…
ชั้นที่ 5	94	88	99.999…

ทั้งนี้การแบ่งชั้นของข้อมูลอาจทำได้อีกลักษณะหนึ่ง โดยแบ่งอันตรภาคชั้นออกเป็น 40 – 51, 52 – 63, 64 – 75, 76 – 87 และ 88 – 99 ตามลำดับ ซึ่งในกรณีนี้ขนาดของอันตรภาคชั้นยังเท่ากับ 12 เหมือนเดิม แต่ขีดจำกัดล่างและขีดจำกัดบนที่แท้จริงของแต่ละชั้นรวมทั้งค่ากึ่งกลางของชั้นจะเปลี่ยนไปจากเดิมดังนี้

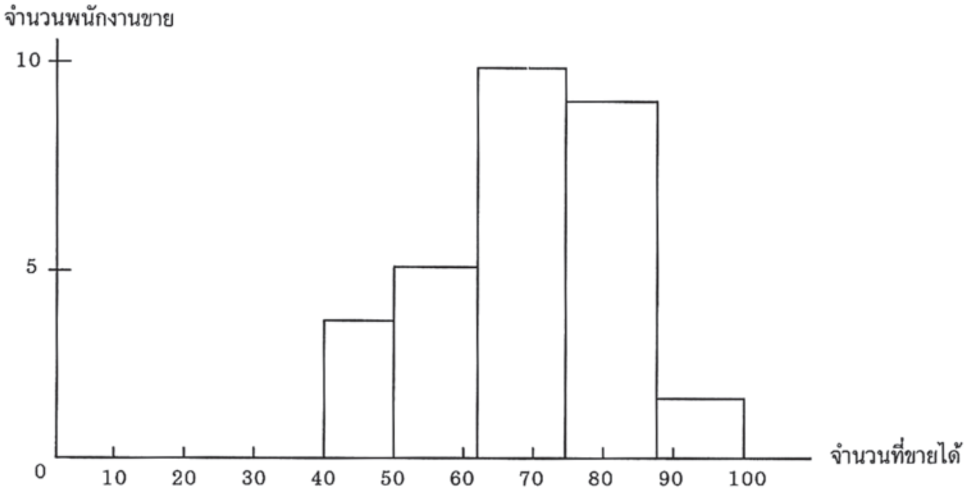
	ค่ากึ่งกลางของชั้น	**ขีดจำกัดล่างของชั้น**	**ขีดจำกัดบนของชั้น**
ชั้นที่ 1	45.5	39.5	51.5
ชั้นที่ 2	57.5	51.5	63.5
ชั้นที่ 3	69.5	63.5	75.5
ชั้นที่ 4	81.5	75.5	87.5
ชั้นที่ 5	93.5	87.5	99.5

2.2 การนำเสนอการแจกแจงความถี่ด้วยกราฟ

การแสดงการแจกแจงความถี่นอกจากจะอยู่ในรูปของตารางแจกแจงความถี่ดังตัวอย่างที่ผ่านมาแล้ว เรายังสามารถแสดงด้วยแผนภูมิหรือกราฟในลักษณะต่าง ๆ กัน เช่น *ฮิสโทแกรม* (Histogram)

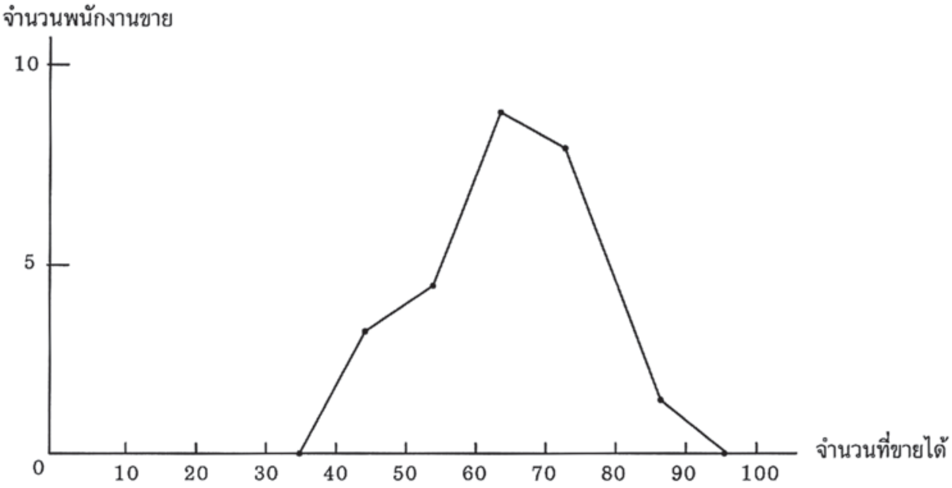
หรือ *รูปหลายเหลี่ยมของความถี่* (Frequency Polygon) และ *กราฟความถี่สะสม* (Cumulative Frequency Graphs) เป็นต้น

(1) ฮิสโทแกรม เป็นแผนภูมิแท่งที่แสดงการแจกแจงความถี่ของข้อมูล ดังรูปที่ 2.1 ซึ่งสร้างมาจากข้อมูลในตารางที่ 2.3



รูปที่ 2.1 ฮิสโทแกรม

(2) รูปหลายเหลี่ยมของความถี่ เป็นแผนภูมิเส้นแสดงการแจกแจงความถี่ของข้อมูล สร้างจากการลากเส้นผ่านจุดกึ่งกลางของแต่ละอันตรภาคชั้น ดังรูปที่ 2.2



รูปที่ 2.2 รูปหลายเหลี่ยมของความถี่