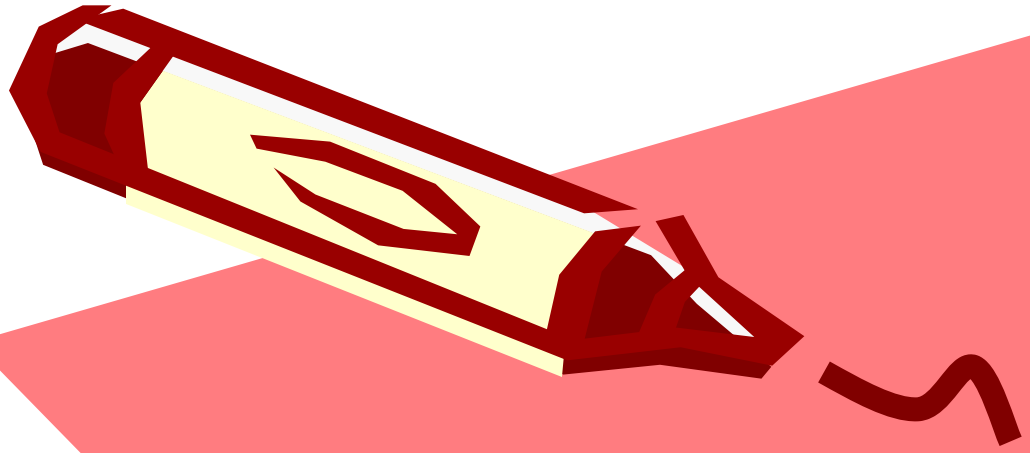


บทที่ 1
ความรู้เบื้องต้นเกี่ยวกับสถิติ

ความหมายสถิติ

คำว่า สถิติ (Statistics) จำแนกได้ 2 ความหมาย คือ

สถิติ หมายถึง ข้อมูลหรือตัวเลขที่แทนข้อเท็จจริงเกี่ยวกับเรื่องต่างๆที่เราสนใจหรือที่อยู่รอบตัวเรา

สถิติ หมายถึง ศาสตร์หรือวิชาที่ว่าด้วยหลักการและวิธีการทางสถิติ ซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การจัดระเบียบข้อมูล หรือการนำเสนอข้อมูล การวิเคราะห์และการแปลความหมายข้อมูล

ตามสถิติในแง่ที่เป็นศาสตร์ สถิติแบ่งออกเป็น 2 ประเภทคือ

1) สถิติเชิงพรรณนา (descriptive statistics)
2) สถิติเชิงอนุมาน (inferential statistics)

ข้อมูลและระดับการวัดข้อมูล

ข้อมูล (data) หมายถึง ข้อเท็จจริงต่างๆ ที่เราสนใจจะศึกษา ซึ่งอาจเป็นตัวเลขหรือมิใช่ตัวเลขก็ได้ โดยแบ่งได้เป็น 2 ประเภท คือ

- ข้อมูลเชิงปริมาณ คือ ค่าที่วัดได้ออกมาเป็นตัวเลข
- ข้อมูลเชิงคุณภาพ คือ ค่าที่วัดได้ไม่สามารถวัดออกมาเป็นตัวเลขได้

ระดับการวัดข้อมูล (levels of measurement)

ระดับการวัดข้อมูลแบ่งออกเป็น 4 ระดับหรือมาตรา ดังนี้

- มาตรานามบัญญัติ
- มาตราเรียงลำดับ
- มาตราอันตรภาค
- มาตราอัตราส่วน

ระดับการวัดข้อมูล

มาตรานามบัญญัติ (nominal scale)

เป็นมาตราการวัดที่ใช้กับข้อมูลที่มีลักษณะหยาบหรือต่ำสุด อาจเป็นการกำหนดสัญลักษณ์หรือตัวเลขเพื่อจำแนกประเภทสิ่งของหรือคุณลักษณะต่างๆ เท่านั้น ไม่สามารถแสดงให้เห็นปริมาณมากน้อยหรือสูงต่ำแต่อย่างใด ดังนั้น จึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หารได้ เช่น เพศ คือ เพศชายและหญิง

มาตราเรียงอันดับ (ordinal scale)

เป็นมาตราการวัดที่มีความละเอียดการวัดเพิ่มขึ้นหรือสูงกว่ามาตรานามบัญญัติ เพราะสามารถบอกลำดับและความแตกต่าง แต่ไม่สามารถบอกได้ว่าคุณลักษณะหรือคุณสมบัติเหล่านี้มีปริมาณมากน้อยกว่ากันเท่าใด กล่าวอีกนัยหนึ่ง ข้อมูลในระดับนี้ไม่สามารถนำมาคำนวณทางคณิตศาสตร์ได้เช่นเดียวกับมาตรานามบัญญัติ เช่น การประกวดนางสาวไทย

อันดับที่ 1 คือ นางสาวไทย

อันดับที่ 2 คือ รองนางสาวไทยคนที่ 1

และอันดับที่ 3 คือ รองนางสาวไทยคนที่ 2

ระดับการวัดข้อมูล

มาตราอันตรภาค (interval scale)

เป็นมาตราการวัดที่สามารถทราบได้ว่าสิ่งที่จะวัดมีช่วงมากน้อยเท่าใด โดยแต่ละช่วงมาตรานี้มีค่าเท่าๆ กัน ทำให้เราทราบถึงความแตกต่างที่ห่างกันเป็นช่วงได้ และค่าที่ได้จากการวัดสามารถนำมาคำนวณทางคณิตศาสตร์ได้ เช่น เราจะบอกความแตกต่างของน้ำร้อนระหว่าง 60° C กับ 80° C เท่ากับความแตกต่างระหว่าง 100° C กับ 120° C โดยดูจากช่วงที่ห่างกันเท่ากับ20° C

มาตราอัตราส่วน (ratio scale)

เป็นมาตราการวัดที่ดีที่สุดและวัดได้อย่างละเอียดที่สุด ตัวเลขที่วัดได้สามารถสื่อความหมายตรงตามค่าของสิ่งที่วัด และเป็นมาตราวัดที่ข้อมูลมีค่าเป็นศูนย์แท้ คือ ถ้าค่าตัวเลขที่วัดได้มีค่าเป็นศูนย์ ก็แปลว่า สิ่งที่วัดนั้นมีค่าศูนย์ด้วย ข้อมูลที่อยู่ในมาตราวัดระดับนี้ ได้แก่ เวลา อายุ น้ำหนัก ส่วนสูง ระยะทาง เป็นต้น ข้อมูลที่วัดได้สามารถนำมาทำการคำนวณทางคณิตศาสตร์ได้

แหล่งของข้อมูล

แหล่งของข้อมูล (Source of data) หมายถึง สิ่งที่ทำให้ได้มาซึ่งข้อมูลทางสถิติ แบ่งเป็น 2 ประเภท คือ

แหล่งปฐมภูมิ (Primary source) เป็นแหล่งของข้อมูลที่เป็นต้นกำเนิดของข้อมูลนั้นๆ เราเรียกข้อมูลนี้ว่า ข้อมูลปฐมภูมิ เช่น ข้อมูลที่ได้จากการสัมภาษณ์ ข้อมูลที่ได้จากการทดลอง ข้อมูลที่ได้จากการกรอกแบบสอบถาม เป็นต้น

แหล่งทุติยภูมิ (Secondary source) เป็นแหล่งของข้อมูลที่ได้รวบรวมเอาข้อมูลเอาไว้เมื่อผู้สนใจต้องการศึกษาข้อมูลที่ได้รวบรวมเอาไว้ เราเรียกข้อมูลนี้ว่า ข้อมูลทุติยภูมิ เช่น ข้อมูลที่คัดเลือกมาจากทะเบียนราษฎร์ ข้อมูลที่คัดมาจากบัญชีรายชื่อผู้ป่วย ข้อมูลจากหนังสือพิมพ์ เป็นต้น

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูล แบ่งเป็น 3 วิธี ดังนี้

1. การเก็บรวบรวมข้อมูลจากงานทะเบียน เช่น มหาวิทยาลัยจะมีการบันทึกหมายเหตุรายวันในเรื่องจำนวนอาจารย์ เจ้าหน้าที่ที่มาปฏิบัติราชการ, จำนวนอาจารย์ เจ้าหน้าที่ที่ลา, จำนวนอาจารย์ที่ไปราชการ เป็นต้น

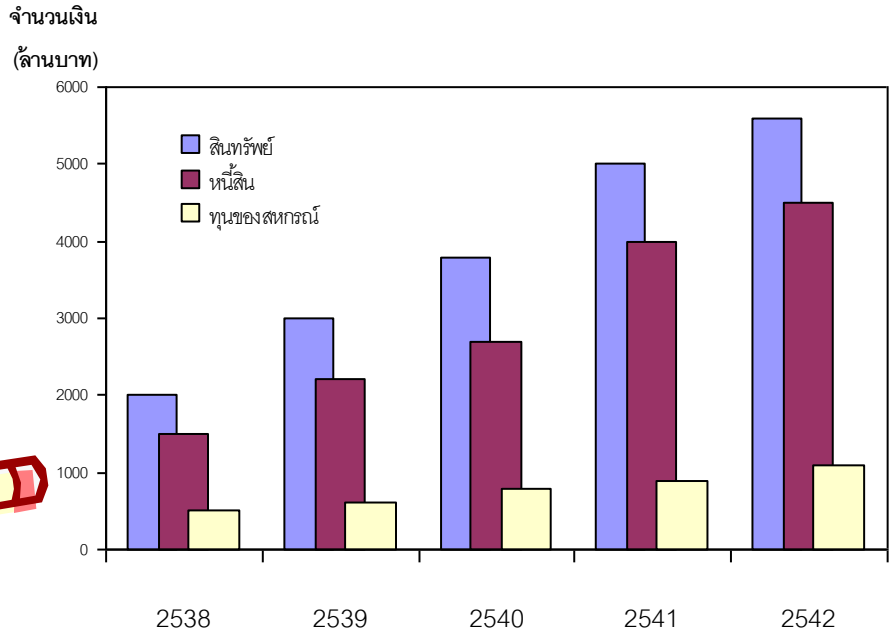
2. การเก็บรวบรวมข้อมูลโดยการสำรวจ เป็นการเก็บรวบรวมข้อมูลจากหน่วยที่ศึกษาโดยตรง เช่น การสำรวจความคิดเห็นของประชาชนในการร่างรัฐธรรมนูญ ซึ่งหน่วยที่ศึกษา คือ ประชาชนคนไทย การเก็บรวบรวมข้อมูลโดยการสำรวจจะแบ่งออกเป็น 2 ประเภท

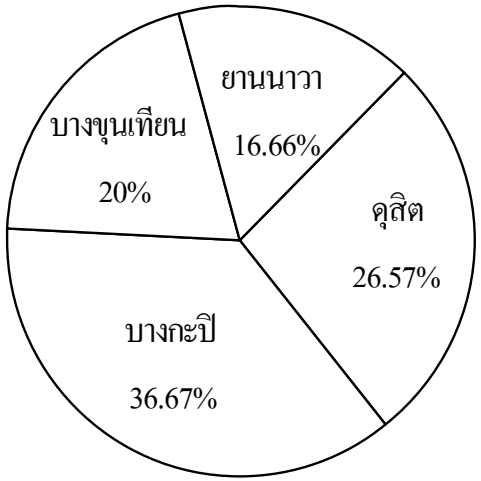
- **การสำมะโน** เก็บรวบรวมข้อมูลจากทุกๆหน่วยศึกษาของประชากร
- **การสำรวจตัวอย่าง** เก็บรวบรวมข้อมูลจากบางหน่วยศึกษาของประชากร

3. การเก็บรวบรวมข้อมูลจากการทดลอง ข้อมูลบางประเภทไม่สามารถหาได้จากการสำรวจ แต่จัดทำได้จากการทดลอง เช่น การศึกษาวิธีการปลูกพืชที่แตกต่างกัน 3 วิธี

การนำเสนอข้อมูล

- การนำเสนอด้วยบทความ
- การนำเสนอด้วยตาราง
- การนำเสนอด้วยกราฟ
- การนำเสนอด้วยแผนภูมิ อาจเป็นแผนภูมิแท่ง, วงกลม หรือรูปภาพ





การแจกแจงความถี่ด้วยตาราง

การแจกแจงความถี่ด้วยวิธีนี้ เป็นการนำเสนอข้อมูลเชิงปริมาณที่เก็บรวบรวมได้ให้อยู่ในรูปของตารางที่เรียกว่า **ตารางแจกแจงความถี่** (Frequency table) การแจกแจงวิธีนี้สามารถทำได้ 3 วิธี คือ

1. การแจกแจงจัดเรียง (Listed distribution) เป็นวิธีการนำเสนอข้อมูลเชิงปริมาณแบบง่ายที่สุด โดยมีข้อมูลแตกต่างกันไม่มากนักและไม่มีข้อมูลค่าใดมีค่าซ้ำกัน โดยนำข้อมูลทุกค่าที่มีความถี่เท่ากับ 1 มาเรียงตามลำดับ โดยทั่วไปเรียงจากน้อยไปมาก ทำให้เราทราบว่าค่าใดเป็นค่าสูงสุด และต่ำสุด

2. การแจกแจงความถี่ชนิดไม่จัดข้อมูลเป็นหมวดหมู่ (Ungrouped frequency distribution) ในกรณีที่ข้อมูลแตกต่างกันไม่มากนักและมีบางค่าซ้ำกัน ในการนำเสนอข้อมูลดังกล่าวจะสร้างสดมภ์ที่แสดงรอยขีด (Tally) เพื่อแสดงความถี่ของข้อมูลแต่ละค่าและการคำนวณหาความถี่ของข้อมูลแต่ละค่าจะทำได้ง่ายและไม่สับสน

การแจกแจงความถี่ด้วยตาราง

ตัวอย่าง จงสร้างตารางแจกแจงความถี่จากข้อมูลจำนวนหนังสือที่เรียบเรียงโดยอาจารย์ 100 คน ในมหาวิทยาลัยแห่งหนึ่ง ต่อไปนี้

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

วิธีทำ ข้อมูลมีค่าสูงสุดเป็น 9 และค่าต่ำสุดเป็น 0 ซึ่งแตกต่างกันไม่มากนักและมีข้อมูลบางค่าซ้ำกัน จึงสร้างตารางแจกแจงความถี่ได้เป็นดังนี้

จำนวนหนังสือ	รอยขีด	ความถี่ (f)
0	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ///	33
1	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30
2	𝍸 𝍸 𝍸 ///	18
3	𝍸 //	7
4	𝍸	5
5	///	3
6	//	2
7	/	1
8		0
9	/	1
	รวม	100

การแจกแจงความถี่ด้วยตาราง

3. การแจกแจงความถี่ชนิดจัดข้อมูลเป็นหมวดหมู่ (Grouped frequency distribution) ในกรณีที่มีข้อมูลแตกต่างกันมาก แล้วสดมภ์แรกของการแจกแจงความถี่จะมีความยาวมาก ดังนั้น เราสามารถสร้างตารางแจกแจงความถี่ของข้อมูลดังกล่าวได้กะทัดรัดยิ่งขึ้น โดยการจัดหมวดหมู่ให้แก่ข้อมูล ดังเช่น

น้ำหนักสัมภาระ(กิโลกรัม)	จำนวนสัมภาระ (f)
7 – 9	2
10 - 12	8
13 - 15	14
16 - 18	19
19 - 21	7
รวม	50

ขีดจำกัดล่าง

ขีดจำกัดบน

ค่ากึ่งกลางชั้น คือ 14

ขอบเขตบน คือ 15.5

ขอบเขตล่าง คือ 12.5

ผลต่างระหว่างขอบเขตล่างและบนของแต่ละชั้น คือ ความกว้างของชั้น

การแจกแจงความถี่ด้วยตาราง

วิธีการสร้างตารางแจกแจงความถี่

ตัวอย่าง คะแนนของนักศึกษา 150 คน เป็นดังนี้

27 79 69 40 51 88 55 48 36 61 53 44 12 51 65 42 58 55 69 63 70 48 61
55 60 25 47 78 61 54 57 76 73 62 36 67 40 51 59 68 27 46 62 43 54 83
59 94 72 57 82 45 54 52 71 53 82 69 60 35 41 65 62 75 60 42 55 34 49
45 49 64 40 61 73 44 59 46 71 86 43 69 54 31 56 51 75 44 66 53 80 71
53 56 91 60 41 29 56 57 35 51 43 39 56 27 62 44 85 61 59 89 60 51 71
53 58 26 77 68 62 57 48 69 76 52 49 45 54 41 33 61 80 57 42 45 59 44
68 73 55 70 39 58 69 51 85 46 55 67

วิธีทำ 1. กำหนดจำนวนชั้น ในที่นี้ต้องการตารางแจกแจงความถี่ที่มี 9 ชั้น

2. คำนวณค่าพิสัย (Range) ของข้อมูล เมื่อพิสัย คือ ผลต่างระหว่างค่าสูงสุดและค่าต่ำสุด

จากข้อมูล มีค่าสูงสุดและค่าต่ำสุดเป็น **94** และ **12** ดังนั้น พิสัย = 94 – 12 = 82

3. คำนวณความกว้างของแต่ละชั้น โดยความกว้างของแต่ละชั้น คือ ผลหารระหว่างพิสัยและ

จำนวนชั้น คือ $82/9 = 9.11$ เพื่อความสะดวกจะใช้ความกว้างของแต่ละชั้นเป็น 10

การแจกแจงความถี่ด้วยตาราง

4. กำหนดขีดจำกัดล่างของชั้นแรก แล้วคำนวณขีดจำกัดของแต่ละชั้น ชั้นแรกเป็น 10 ขอบเขตล่างของชั้นแรกจึงเท่ากับ 9.5 บวกค่าดังกล่าวด้วยความกว้างของชั้นเท่ากับ 10 จะได้ขอบเขตบนของชั้นแรกเป็น 19.5 และขีดจำกัดบนของชั้นแรกเป็น 19 ดังนั้นขีดจำกัดของชั้นแรกคือ 10 – 19

5. เขียนรอยขีด

6. คำนวณความถี่ โดยการนับรอยขีดในชั้นนั้นๆ ซึ่งผลรวมของความถี่ทั้งหมดทุกชั้นต้องเท่ากับจำนวนข้อมูล จากวิธีการข้างต้นได้ตารางแจกแจงความถี่ดังนี้

คะแนน	รอยขีด	ความถี่ (f)
10 - 19	/	1
20 - 29	𝍸 /	6
30 - 39	𝍸 ////	9
40 - 49	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 /	31
50 - 59	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	42
60 – 69	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	32
70 – 79	𝍸 𝍸 𝍸 //	17
80 – 89	𝍸 𝍸	10
90 - 99	//	2
	รวม	150

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ค่าเฉลี่ย (Mean : $\bar{X}$) เป็นค่ากลางที่คำนวณได้จากข้อมูลทุกค่าที่มีในชุดข้อมูล โดยวิธีการหาได้ 2 วิธี ดังนี้

การหาค่าเฉลี่ยในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่

ค่าเฉลี่ย = ผลรวมทั้งหมดหารด้วยจำนวนข้อมูล

หรือ $$\bar{x} = \frac{1}{N}\sum x_i = \frac{\sum x}{N}$$

ตัวอย่าง 1 จงหาค่าเฉลี่ยของข้อมูลที่กำหนดให้ต่อไปนี้ 22 23 24 24 25 25 21 23 15 20

วิธีทำ $$\bar{x} = \frac{1}{N}\sum x_i = \frac{22+23+24+24+25+25+21+23+15+20}{10}$$

$$= \frac{222}{10} = 22.2$$

นั่นคือ ค่าเฉลี่ยของข้อมูลนี้เท่ากับ 22.2 คะแนน

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การหาค่าเฉลี่ยในกรณีที่ข้อมูลแจกแจงความถี่แล้ว

ในการหาค่าเฉลี่ยของข้อมูลที่แจกแจงความถี่แล้ว ซึ่งไม่ทราบค่าที่แท้จริงของข้อมูลในแต่ละชั้น ดังนั้นถ้าให้ค่ากึ่งกลางของแต่ละชั้นเป็นตัวแทนของข้อมูลในชั้นนั้นๆ แล้ว

$$\bar{x} = \frac{1}{\sum f_i} \sum f_i x_i = \frac{\sum fx}{\sum f}$$

เมื่อกำหนดให้ x_i คือ ค่ากึ่งกลางของชั้นคะแนนที่ i ในชุดข้อมูล

f_i คือ ความถี่ของชั้นที่ i

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 2 จงหาค่าเฉลี่ยของข้อมูลที่กำหนดให้ต่อไปนี้

ขีดจำกัดคะแนน	ค่ากึ่งกลาง x_i	ความถี่ (f)	ผลรวมคะแนนในชั้น fx_i
19.5 – 24.5	22	4	88
24.5 – 29.5	27	6	162
29.5 – 34.5	32	12	384
34.5 – 39.5	37	16	592
39.5 – 44.5	42	12	504
44.5 – 49.5	47	7	329
49.5 – 54.5	52	3	156
รวม		60	2,215

วิธีทำ คำนวณค่าเฉลี่ยได้

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{2,215}{60} = 36.92$$

นั่นคือ ค่าเฉลี่ยของข้อมูลชุดนี้ คือ 36.92

การวัดแนวโน้มเข้าสู่ส่วนกลาง

มัธยฐาน (Median : Me) คือ ค่าที่บอกตำแหน่งกึ่งกลางของข้อมูลทั้งหมดเมื่อได้จัดเรียงข้อมูลจากค่าน้อยไปหาค่ามาก หรือจากค่ามากไปหาค่าน้อย โดยวิธีการหาได้ 2 วิธี ดังนี้

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่

- ถ้าข้อมูลเป็นจำนวนคี่ มัธยฐาน คือ ค่าของข้อมูลตัวที่ $\frac{N+1}{2}$

- ถ้าข้อมูลเป็นจำนวนคู่ คำนวณมัธยฐานโดยการเปรียบเทียบระหว่างข้อมูลตัวที่ $\frac{N}{2}$ และ $\frac{N}{2}+1$ การหามัธยฐานให้นำเอาคะแนนที่อยู่ตรงตำแหน่งกึ่งกลางทั้งสองค่านั้นมาบวกกันหารสอง

ตัวอย่าง 3 จงหามัธยฐานของข้อมูลที่กำหนดให้ 1, 3, 2, 2, 5, 3, 4, 4, 3

วิธีทำ จัดเรียงข้อมูลใหม่ได้ดังนี้ 1, 2, 2, 3, 3, 3, 4, 4, 5

เนื่องจากข้อมูลเป็นจำนวนคี่ ดังนั้น ตำแหน่งกึ่งกลาง คือ $\frac{N+1}{2} = \frac{9+1}{2} = 5$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ 3

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 4 จงหามัธยฐานของข้อมูลที่กำหนดให้ต่อไปนี้

6.5, 12, 14.9, 10, 7.9, 21.9, 9.2, 14.5

วิธีทำ จัดเรียงลำดับข้อมูลใหม่ 6.5, 7.9, 9.2, 10, 12, 14.5, 14.9, 21.9

เนื่องจากข้อมูลเป็นจำนวนคู่ ดังนั้น ตำแหน่งกึ่งกลาง คือ $\frac{N}{2} = \frac{8}{2} = 4$ ซึ่งมีค่าเท่ากับตำแหน่งที่ 4 ตรงกับค่าคะแนน 10

และตำแหน่งกึ่งกลาง คือ $\frac{N}{2}+1 = \frac{8}{2}+1 = 5$ ซึ่งมีค่าเท่ากับตำแหน่งที่ 5 ตรงกับค่าคะแนน 12

ดังนั้นมัธยฐานของข้อมูลชุดนี้ คือ $\frac{10+12}{2} = 11$

การวัดแนวโน้มเข้าสู่ส่วนกลาง

กรณีข้อมูลที่แจกแจงความถี่แล้ว

คำนวณหามัธยฐานได้จากสูตร ต่อไปนี้

$$Me = L_1 + \left[\frac{\frac{N}{2} - F}{f_1} \right] I$$

เมื่อ L_1 คือ ขอบเขตล่างของชั้นที่มัธยฐานอยู่

N คือ จำนวนข้อมูลทั้งหมด หรือความถี่ทั้งหมด

F คือ ผลรวมของความถี่ของทุกชั้นที่มีข้อมูลต่ำกว่าชั้นที่มีมัธยฐานอยู่

f_1 คือ ความถี่ของชั้นที่มีมัธยฐานอยู่

I คือ ความกว้างของชั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 5 จงหามัธยฐานของข้อมูลที่กำหนดให้ต่อไปนี้

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

การวัดแนวโน้มเข้าสู่ส่วนกลาง

วิธีทำ เนื่องจาก $\frac{N}{2} = \frac{60}{2} = 30$ ดังนั้นมัธยฐานอยู่ในชั้น 34.5 – 39.5

และได้ว่า $L_1 = 34.5,\ F = 22,\ f_1 = 16,\ I = 5$

จาก $$Me = L_1 + \left[\frac{\frac{N}{2} - F}{f_1}\right] I$$

จะได้ $$Me = 34.5 + \left[\frac{\frac{60}{2} - 22}{16}\right](5)$$

$$= 37$$

นั่นคือ มัธยฐานของข้อมูลชุดนี้ คือ 37

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ฐานนิยม (Mode : Mo) คือ ค่าของข้อมูลที่ปรากฏย่อยครั้งที่สุดหรือมีความถี่สูงสุด ทั้งนี้ข้อมูลชุดหนึ่งอาจไม่มีฐานนิยมเลยก็ได้ แต่ถ้ามีก็อาจมีได้มากกว่าหนึ่งค่า โดยวิธีการหาได้ 2 วิธี ดังนี้

กรณีข้อมูลที่ไม่ได้แจกแจงความถี่

ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด

ตัวอย่าง 6 จงหาฐานนิยมของข้อมูลต่อไปนี้

ก. 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12 และ 18

Mo = 9

ข. 3, 5, 8, 10, 12, 15, 16

ข้อมูลชุดนี้ไม่มี Mo เนื่องจากไม่มีข้อมูลค่าใดมีความถี่สูงสุด

ค. 2, 3, 4, 4, 4, 5, 5, 7, 7, 7 และ 9

Mo = 4 และ 7

การวัดแนวโน้มเข้าสู่ส่วนกลาง

กรณีข้อมูลที่แจกแจงความถี่แล้ว

คำนวณหาฐานนิยมได้จากสูตร ต่อไปนี้

$$Mo = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] I$$

เมื่อ L_1 คือ ขอบเขตล่างของชั้นที่ฐานนิยมอยู่

Δ_1 คือ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลต่ำกว่า

Δ_2 คือ ผลต่างระหว่างความถี่ของชั้นที่มีฐานนิยมอยู่ กับชั้นติดกันที่มีข้อมูลสูงกว่า

I คือ ความกว้างของชั้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง

ตัวอย่าง 7 จงหาฐานนิยมของคะแนนนักศึกษา 150 คน

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

วิธีทำ เนื่องจากความถี่สูงสุด = 16 ดังนั้นฐานนิยมอยู่ในชั้น 34.5 – 39.5

และได้ว่า $L_1 = 34.5,\ \Delta_1 = 16-12 = 4,\ \Delta_2 = 16-12 = 4,\ I = 5$

จะได้ว่า $$Mo = L_1 + \left[\frac{\Delta_1}{\Delta_1+\Delta_2}\right]I = 34.5 + \left[\frac{4}{4+4}\right](5) = 37$$

นั่นคือ ฐานนิยมของข้อมูลชุดนี้ คือ 37

การวัดการกระจายของข้อมูล

พิจารณาข้อมูล 3 ชุด ต่อไปนี้

ชุดที่ 1 : 10, 8, 9, 10, 11, 12

ชุดที่ 2 : 5, 6, 8, 10, 12, 14, 15

ชุดที่ 3 : 1, 2, 5, 10, 15, 18, 19

จะพบว่า ข้อมูลทั้ง 3 ชุด มี $\mu = 10$ เท่ากัน และอาจจะเข้าใจว่าข้อมูลทั้ง 3 ชุด มีลักษณะคล้ายคลึงกัน แต่จะเห็นว่าข้อมูลชุดที่ 1 มีค่าของข้อมูลแต่ละค่าใกล้เคียงกับ $\mu = 10$ มากที่สุด นั่นคือ จะได้ว่า ในข้อมูลชุดที่ 1 ค่าของข้อมูลแต่ละตัว **มีการกระจายห่างจากค่าเฉลี่ยน้อยที่สุด**

การวัดการกระจายของข้อมูล

การวัดการกระจายมีอยู่หลายวิธี ดังต่อไปนี้

พิสัย (Range) คือ ผลต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูล

นั่นคือ ถ้ามีข้อมูล $X_1, X_2, \ldots, X_N$ รวม N ตัว แล้ว **พิสัย = ค่าสูงสุด - ค่าต่ำสุด**

ตัวอย่าง 8 จงหาพิสัยของข้อมูลที่กำหนดให้ดังต่อไปนี้ 2, 8, 6, 7, 10, 5, 6, 3

วิธีทำ จากพิสัย = ค่าสูงสุด – ค่าต่ำสุด

= 10 – 2 = 8

นั่นคือ พิสัยของข้อมูลชุดนี้ คือ 8

การวัดการกระจายของข้อมูล

พิสัยกึ่งควอไทล์ คือ ผลต่างระหว่างควอไทล์ที่ 3 และควอไทล์ที่ 1 นั่นคือ

$$Q.R. = Q_3 - Q_1$$

การหาพิสัยกึ่งควอไทล์ในกรณีที่ข้อมูลไม่ได้แจกแจงความถี่

ตำแหน่งของควอไทล์ที่ r หรือ Q_r คือ $\frac{r(N+1)}{4}$; $r = 1, 2, 3$

ตัวอย่าง 9 จงหาพิสัยกึ่งควอไทล์ของข้อมูลที่กำหนดให้ ต่อไปนี้

10, 16, 2, 19, 30, 15, 23, 13

วิธีทำ ก่อนอื่นเรียงลำดับข้อมูลจากน้อยไปหามาก ได้ดังนี้

2, 10, 13, 15, 16, 19, 23, 30

ตำแหน่งของ Q_1 คือ $\frac{1(8+1)}{4} = 2.25$

ดังนั้น

$$Q_1 = 10 + (0.25)(13 - 10) = 10.75$$

การวัดการกระจายของข้อมูล

และตำแหน่งของ Q_3 คือ $\frac{3(8+1)}{4} = 6.75$

จะได้ว่า

$$Q_3 = 19 + (0.75)(23-19) = 22$$

ดังนั้น $Q.R. = Q_3 - Q_1 = 22 - 10.75 = 11.25$

นั่นคือ พิสัยกึ่งควอไทล์ของข้อมูลเท่ากับ 11.25

การวัดการกระจายของข้อมูล

ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation ; S.D.) คือ ค่ารากที่สองที่เป็นบวกของค่าเฉลี่ยของกำลังสองของผลต่างของคะแนนใดๆ กับค่าเฉลี่ยของข้อมูลนั้น นั่นคือ

$$S.D. = \sqrt{\frac{\sum(x-\bar{x})^2}{N}}$$

ในกรณีที่ข้อมูลแจกแจงความถี่แล้ว

ส่วนเบี่ยงเบนมาตรฐาน หาได้จากสูตรดังนี้

$$S.D. = \sqrt{\left(\frac{\sum fx^2}{\sum f}\right)-\left(\frac{\sum fx}{\sum f}\right)^2}$$

การวัดการกระจายของข้อมูล

ตัวอย่าง 10 จงหา S.D. ของข้อมูลที่กำหนดให้ ต่อไปนี้ 2, 3, 4, 4, 5, 5, 1, 3

วิธีทำ 1) คำนวณหา $\overline{X}$ ของข้อมูลได้ดังนี้

$$\overline{x} = \frac{2+3+4+4+5+5+1+3}{8} = 3.38$$

2) หากำลังสองของผลต่างของ x และ $\overline{X}$ หรือหา $(x-\overline{x})^2$ ได้ดังนี้

$(2-3.38)^2 = (-1.38)^2 = 1.90$ $(3-3.38)^2 = (-0.38)^2 = 0.14$

$(4-3.38)^2 = (0.62)^2 = 0.38$ $(4-3.38)^2 = (0.62)^2 = 0.38$

$(5-3.38)^2 = (1.62)^2 = 2.62$ $(5-3.38)^2 = (1.62)^2 = 2.62$

$(1-3.38)^2 = (-2.38)^2 = 5.66$ $(3-3.38)^2 = (-0.38)^2 = 0.14$

3) คำนวณค่า S.D. ดังนี้

$$S.D. = \sqrt{\frac{1.90+0.14+0.38+0.38+2.62+2.62+5.66+0.14}{8}}$$

$$= \sqrt{\frac{13.84}{8}} = \sqrt{1.73} = \pm 1.315$$

การวัดการกระจายของข้อมูล

ตัวอย่าง 11 จงหา S.D. ของข้อมูลที่กำหนดให้ ต่อไปนี้

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

การวัดการกระจายของข้อมูล

วิธีทำ สร้างตารางวิเคราะห์ข้อมูลก่อน ดังนี้

ขีดจำกัดคะแนน	ค่ากึ่งกลาง x_i	ความถี่ (f)	ผลรวมคะแนนในชั้น fx_i	fx^2
19.5 – 24.5	22	4	88	1936
24.5 – 29.5	27	6	162	4374
29.5 – 34.5	32	12	384	12,288
34.5 – 39.5	37	16	592	21,904
39.5 – 44.5	42	12	504	21,168
44.5 – 49.5	47	7	329	15,463
49.5 – 54.5	52	3	156	8,112
รวม		60	2,215	85,245

แทนค่าในสูตร จะได้

$$S.D. = \sqrt{\left(\frac{\sum fx^2}{\sum f}\right) - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{\left(\frac{85,245}{60}\right) - \left(\frac{2,215}{60}\right)^2} = \sqrt{57.91}$$

ดังนั้น ค่าเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ คือ ± 7.61

การวัดการกระจายของข้อมูล

ความแปรปรวน (Variance : s^2) ความแปรปรวนเป็นกำลังสองของส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดเดียวกัน นั่นคือ

$$s^2 = \frac{\sum(x-\bar{x})^2}{N}$$

ในกรณีที่ข้อมูลแจกแจงความถี่แล้ว

ความแปรปรวน หาได้จากสูตร

$$s^2 = \left(\frac{\sum fx^2}{\sum f}\right) - \left(\frac{\sum fx}{\sum f}\right)^2$$

การวัดการกระจายของข้อมูล

ตัวอย่าง 12 จงหาความแปรปรวนของข้อมูลที่กำหนดให้ ต่อไปนี้

ขีดจำกัดคะแนน	ความถี่	ความถี่สะสม
19.5 – 24.5	4	4
24.5 – 29.5	6	10
29.5 – 34.5	12	22
34.5 – 39.5	16	38
39.5 – 44.5	12	50
44.5 – 49.5	7	57
49.5 – 54.5	3	60

การวัดการกระจายของข้อมูล

วิธีทำ สร้างตารางวิเคราะห์ข้อมูลก่อน ดังนี้

ขีดจำกัดคะแนน	ค่ากึ่งกลาง x_i	ความถี่ (f)	ผลรวมคะแนนในชั้น fx_i	fx^2
19.5 – 24.5	22	4	88	1936
24.5 – 29.5	27	6	162	4374
29.5 – 34.5	32	12	384	12,288
34.5 – 39.5	37	16	592	21,904
39.5 – 44.5	42	12	504	21,168
44.5 – 49.5	47	7	329	15,463
49.5 – 54.5	52	3	156	8,112
รวม		60	2,215	85,245

แทนค่าในสูตร จะได้

$$S^2 = \left(\frac{\sum fx^2}{\sum f}\right) - \left(\frac{\sum fx}{\sum f}\right)^2 = \left(\frac{85,245}{60}\right) - \left(\frac{2,215}{60}\right)^2 = 57.91$$

ดังนั้น ความแปรปรวนของข้อมูลชุดนี้ คือ 57.91

แบบฝึกหัดบทที่ 1

1. ผลการสอบวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนแห่งหนึ่งจำนวน 30 คน ได้คะแนนสอบ ดังนี้

50	49	27	46	48	26	30
47	24	17	31	42	45	33
44	29	35	43	20	37	22
19	38	30	18	39	30	40
21	25					

จงสร้างตารางแจกแจงความถี่ (หาขีดจำกัด, ขอบเขต, จำนวนข้อมูล, ความถี่, ความถี่สะสม, ความถี่สัมพัทธ์, จุดกึ่งกลางชั้น)

2. จงหาพิสัย มัธยฐาน ความแปรปรวน และค่าเบี่ยงเบนมาตรฐานประชากรของข้อมูล ต่อไปนี้

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

แบบฝึกหัดบทที่ 1

3. จากการสำรวจอายุการเข้าทำงานของพนักงานในบริษัทแห่งหนึ่ง ดังนี้

อายุของพนักงาน	จำนวน
15 - 17	4
18 - 20	7
21 - 23	9
24 - 26	12
27 - 29	10
30 - 32	5
33 - 35	3

3.1 จงหาความถี่สะสมและความถี่สัมพัทธ์

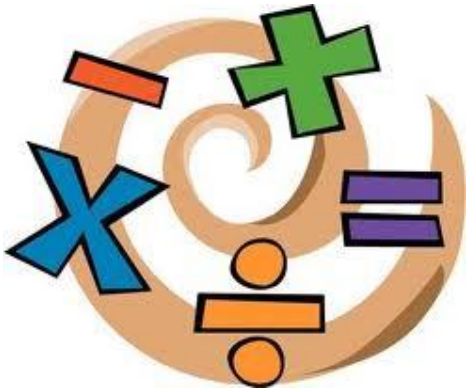
3.2 จงหาค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และมัธยฐาน

4. จงหาค่าเฉลี่ยและมัธยฐานของตัวอย่าง ต่อไปนี้

5.5 6.7 7.3 7.7 6.8 6.9 7.2

6.4 8.1 7.5

บทที่ 2
ความน่าจะเป็นและ
การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม



ความน่าจะเป็น

ความน่าจะเป็นจัดว่าเป็นพื้นฐานสำคัญของสถิติ จำเป็นต้องมีความรู้ การนิยาม คุณสมบัติและทฤษฎีเบื้องต้นของความน่าจะเป็น เพื่อนำไปประยุกต์ใช้ในการคำนวณหาค่าความน่าจะเป็นของการแจกแจงตัวแปรสุ่ม

นิยาม 2.1 ความน่าจะเป็นของเหตุการณ์ใด

ความน่าจะเป็นของเหตุการณ์ใด เท่ากับอัตราส่วนของจำนวนผลที่จะเกิดเหตุการณ์นั้นต่อจำนวนผลทั้งหมดที่อาจจะเกิดขึ้นได้

$$\text{หรือ ความน่าจะเป็นของเหตุการณ์ใด} = \frac{\text{จำนวนผลที่จะเกิดขึ้นในเหตุการณ์นั้น}}{\text{จำนวนผลทั้งหมดที่อาจจะเกิดขึ้นได้}}$$

เมื่อผลทั้งหมดที่อาจจะเกิดขึ้นจากการทดลองสุ่ม แต่ละตัวมีโอกาสเกิดขึ้นได้เท่า ๆ กัน

ความน่าจะเป็น

กำหนดให้ E เป็นเหตุการณ์ที่เราสนใจ

$P(E)$ เป็นความน่าจะเป็นของเหตุการณ์นั้น

$n(S)$ เป็นจำนวนสมาชิกทั้งหมดที่เกิดขึ้นได้จากการทดลองสุ่ม

$n(E)$ เป็นจำนวนสมาชิกของเหตุการณ์ที่เราสนใจ

ดังนั้น

$$P(E) = \frac{n(E)}{n(S)}$$

คุณสมบัติที่สำคัญของความน่าจะเป็น มีดังนี้

1. ความน่าจะเป็นของเหตุการณ์ใด ๆ จะมีค่าได้ตั้งแต่ 0 ถึง 1 เท่ากับ $0 \leq P(E) \leq 1$

2. ผลรวมของความน่าจะเป็นของเหตุการณ์ที่เป็นไปได้ทั้งหมดทุก ๆ เหตุการณ์มีค่าเท่ากับ 1

ความน่าจะเป็น

ตัวอย่าง 2.1 บริษัทสินค้าเบ็ดเตล็ด จำกัด ผลิตสินค้าเบ็ดเตล็ดได้จำนวน 6,000 ชิ้น ในจำนวนที่ผลิตได้นี้มีสินค้าเบ็ดเตล็ดชำรุด 500 ชิ้น จงหาความน่าจะเป็นที่บริษัทสินค้าเบ็ดเตล็ดชำรุด

วิธีทำ จำนวนสมาชิกของเหตุการณ์ที่สนใจ $n(E)$

คือ จำนวนสินค้าเบ็ดเตล็ดชำรุด = 500 ชิ้น

จำนวนสมาชิกของเหตุการณ์ทั้งหมดที่เป็นไปได้ $n(S)$

คือ จำนวนสินค้าเบ็ดเตล็ดที่ผลิตได้ = 6,000 ชิ้น

ดังนั้น ความน่าจะเป็นที่บริษัทจะผลิตสินค้าเบ็ดเตล็ดชำรุด

$$P(E) = \frac{500}{6,000} = 0.083$$

ความน่าจะเป็น

ตัวอย่าง 2.2 จากการสำรวจโรงงานที่ตั้งอยู่ในจังหวัดหนึ่ง ปรากฏว่ามีจำนวนโรงงานที่มีคนงานจำนวนต่าง ๆ ดังนี้

จำนวนคนงาน	จำนวนโรงงาน
ต่ำกว่า 30	48
30-44	22
45-59	10
60-74	37
75-89	9
ตั้งแต่ 90 คนขึ้นไป	15

ความน่าจะเป็น

จงหาความน่าจะเป็นที่โรงงานแห่งหนึ่งในจังหวัดนี้จะมีจำนวนคนงาน

2.1 ต่ำกว่า 60 คน

2.2 ตั้งแต่ 30 คน ถึง 89 คน

2.3 ตั้งแต่ 90 คนขึ้นไป

ตัวแปรสุ่ม

ตัวแปรสุ่ม (Random variable) คือ ค่าหรือตัวเลขที่ใช้แทนเหตุการณ์ต่าง ๆ ที่อาจเกิดขึ้นได้ทั้งหมดจากการทดลอง

โดยทั่วไปนิยมใช้อักษรภาษาอังกฤษตัวใหญ่เป็นสัญลักษณ์แทนตัวแปรสุ่ม เช่น X, Y หรือ Z และใช้อักษรภาษาอังกฤษตัวเล็กเป็นสัญลักษณ์แทนค่าของตัวแปรสุ่ม

เช่น x, y หรือ z

ตัวอย่าง 2.3

1. ถ้าให้ X เป็นตัวแปรสุ่ม แทนเหตุการณ์ที่เป็นจำนวนวันซึ่งร้านน้ำฝนพาณิชย์เปิดขายสินค้าในเดือน มีนาคม 2560 โดยที่ x = 0 (ไม่เปิดขายสินค้าเลย)

x = 1, 2, 3, 4, 5, ... , 50 (เปิดขายสินค้าทุกวัน)

ตัวแปรสุ่ม

2. ถ้าให้ Y เป็นตัวแปรสุ่ม แทนเหตุการณ์ที่เป็นจำนวนลูกค้าวันเข้ามาซื้อสินค้าในวันที่ผ่านมาของร้านโอเอการค้า โดยที่ y = 0 (ไม่มีลูกค้าเข้ามาซื้อสินค้าเลย)

y = 1, 2, 3, 4, 5, ... (มีลูกค้าเข้ามาซื้อสินค้า)

3. ถ้าให้ X เป็นตัวแปรสุ่ม แทนเหตุการณ์ที่เป็นจำนวนสินค้าในโรงงานแห่งหนึ่งที่ผลิตได้

โดยที่ x = 0 (ไม่มีจำนวนสินค้าที่ผลิตได้)

x = 1, 2, 3, ..., 100 (มีจำนวนสินค้าที่ผลิตได้)

ชนิดของตัวแปรสุ่ม

ตัวแปรสุ่มแบ่งได้ 2 ชนิด คือ ตัวแปรสุ่มไม่ต่อเนื่องและตัวแปรสุ่มต่อเนื่อง

ตัวแปรสุ่มไม่ต่อเนื่อง (Discrete random variable) คือ ตัวแปรสุ่มที่ประกอบด้วยค่าหรือตัวเลข อาจมีจำนวนจำกัดหรือจำนวนไม่จำกัดก็ได้ โดยทั่วไปตัวแปรสุ่มไม่ต่อเนื่องคือ จำนวนนับ เช่น 0, 1, 2, 3, ...

ตัวอย่าง 2.4

1. X = 5, 6, 8, 10, 11 *เป็นตัวแปรสุ่มไม่ต่อเนื่อง*
2. X แทนจำนวนสินค้าในโรงงานแห่งหนึ่ง *เป็นตัวแปรสุ่มไม่ต่อเนื่อง*
3. *X = 0.1, 0.5, 1.0, 1.5, 2.0 เป็นตัวแปรสุ่มไม่ต่อเนื่อง*
4. Y แทนจำนวนคนงานในบริษัทฟู้ดเซ็นเตอร์ *เป็นตัวแปรสุ่มไม่ต่อเนื่อง*

ชนิดของตัวแปรสุ่ม

ตัวแปรสุ่มต่อเนื่อง (Continuous random variable) ตัวแปรสุ่มที่ประกอบด้วยค่าหรือตัวเลขนับไม่ถ้วน โดยทั่วไปตัวแปรสุ่มต่อเนื่องอยู่ในรูปของช่วง ซึ่งจะมีจำนวนนับไม่ถ้วนหรือไม่สามารถนับจำนวนได้

ตัวอย่าง 2.5

1. $0 \leq Z \leq 5$ *เป็นตัวแปรสุ่มต่อเนื่อง*

2. $N \geq 1$ *เป็นตัวแปรสุ่มต่อเนื่อง*

3. X แทน ปริมาณน้ำที่ใช้ในการผลิตน้ำอัดลมชนิดหนึ่งเมื่อเดือนที่แล้ว *เป็นตัวแปรสุ่มต่อเนื่อง*

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม คือ การแจกแจงของเหตุการณ์ทุกเหตุการณ์ที่เป็นไปได้ และความน่าจะเป็นที่เหตุการณ์เหล่านั้นจะเกิดขึ้น

ตัวอย่าง 2.6 ในการลงทุนเปิดร้านขายอุปกรณ์เสริมสวยต่าง ๆ ผลจากการลงทุนมีเหตุการณ์ที่เป็นไปได้ 3 เหตุการณ์ คือ กำไร เท่าทุนและขาดทุน โดยในรอบปีที่ผ่านมา ถ้าจากการสอบถามร้านอุปกรณ์เสริมสวยต่าง ๆ จำนวน 500 ร้าน ปรากฏว่าได้กำไร 250 ร้าน เท่าทุน 140 ร้าน และขาดทุน 110 ร้าน

ความน่าจะเป็นที่ลงทุนเปิดร้านอุปกรณ์เสริมสวยต่าง ๆ แล้วได้กำไร $\frac{250}{500} = 0.5$

ความน่าจะเป็นที่ลงทุนเปิดร้านอุปกรณ์เสริมสวยต่าง ๆ แล้วเท่าทุน $\frac{140}{500} = 0.28$

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

ความน่าจะเป็นที่ลงทุนเปิดร้านอุปกรณ์เสริมสวยต่าง ๆ แล้วขาดทุน $\frac{110}{500} = 0.22$

การแจกแจงความน่าจะเป็นของผลจากการลงทุนเปิดร้านขายอุปกรณ์เสริมสวยต่าง ๆ อาจเขียนสรุปอยู่ในรูปตารางได้ดังนี้

ผลการลงทุน	ความน่าจะเป็น
กำไร	0.50
เท่าทุน	0.28
ขาดทุน	0.22
รวม	**1.00**

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

การโยนเหรียญ 1 อัน 3 ครั้ง เมื่อทั้ง 8 ผลลัพธ์มีโอกาสเกิดขึ้นเท่า ๆ กัน จะได้ตารางดังนี้

X เป็นจำนวนครั้งของการเกิดหัว

ผลลัพธ์	ความน่าจะเป็น	x
(H,H,H)	1/8	3
(H,H,T)	1/8	2
(H,T,H)	1/8	2
(T,H,H)	1/8	2
(H,T,T)	1/8	1
(T,H,T)	1/8	1
(T,T,H)	1/8	1
(T,T,T)	1/8	0

เมื่อ P(X = x) คือ สัญลักษณ์แทนความน่าจะเป็นที่ตัวแปรสุ่ม X มีค่า x จะได้ตารางแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง X ดังนี้

x	P(X = x)
0	1/8
1	3/8
2	3/8
3	1/8

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ค่าเฉลี่ยหรือค่าคาดหมายของตัวแปรสุ่ม (Expected Value)

ค่าเฉลี่ยของตัวแปรสุ่มหาได้จากผลบวกของผลคูณระหว่างค่าแต่ละของตัวแปรสุ่มกับความน่าจะเป็นของแต่ละค่านั้น นั่นคือ

$$\mu_x = \mathrm{E}(X) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \ldots + x_N \cdot P(x_N)$$

$$= \sum_{i=1}^{N} x_i P(x_i)$$

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ตัวอย่าง 2.7 อัตราผลตอบแทนรายเดือนต่อหุ้นของบริษัทแห่งหนึ่ง มีค่าดังตารางต่อไปนี้

อัตราผลตอบแทน (บาท)	1,000	5,200	3,700	2,900
ความน่าจะเป็น	0.3	0.1	0.4	0.2

จงหาอัตราผลตอบแทนรายเดือนโดยเฉลี่ยควรจะเป็นเท่าใด

<u>วิธีทำ</u> อัตราผลตอบแทนรายปีโดยเฉลี่ย

$$\mu_x = \mathrm{E(X)} = \sum_{x=1}^{4} x_i P(x_i)$$

$$= x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + x_3 \cdot P(x_3) + x_4 \cdot P(x_4)$$

$$= 1,000(0.3) + 5,200(0.1) + 3,700(0.4) + 2,900(0.2)$$

$$= 2,880 \quad \text{บาท}$$

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ตัวอย่าง 2.8 ถ้าความน่าจะเป็นที่พนักงานขายคนหนึ่งขายเตาไมโครเวฟในราคาเครื่องละ 6,500 บาท และ 5,200 บาท เท่ากับ 0.4 และ 0.6 ตามลำดับ จงหาราคาเตาไมโครเวฟโดยเฉลี่ยที่เขาขายให้ลูกค้า

<u>วิธีทำ</u> ให้ X เป็นตัวแปรสุ่มซึ่งแทนราคาเตาไมโครเวฟ

นั่นคือ $x_1 = 6{,}500,\ x_2 = 5{,}200,\ P(x_1) = 0.4,\ P(x_2) = 0.6$

ดังนั้นราคาเตาไมโครเวฟโดยเฉลี่ย

$$\mu_x = \mathrm{E(X)} = \sum_{x=1}^{2} x_i P(x_i) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2)$$

$$= 6{,}500(0.4) + 5{,}200(0.6)$$

$$= 5{,}720 \quad \text{บาท}$$

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ตัวอย่าง 2.9 ถ้าจำนวนเครื่องปรับอากาศที่บริษัทแห่งหนึ่งขายได้ในช่วงเวลา 100 วัน ที่ผ่านมาเป็นดังนี้

จำนวนเครื่องปรับอากาศ (เครื่อง/วัน)	จำนวนวันที่ขายได้
1	10
2	17
3	25
4	32
5	7
6	9

จงหาว่าบริษัทขายเครื่องปรับอากาศได้โดยเฉลี่ยวันละเท่าไร

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ความแปรปรวนของตัวแปรสุ่ม (Variance of Random variable)

ความแปรปรวนของตัวแปรสุ่มเป็นค่าที่แสดงการกระจายของข้อมูล โดยวัดจากค่าเฉลี่ยของผลต่างกำลังสองระหว่างค่าตัวแปรสุ่มแต่ละค่ากับค่าเฉลี่ยหรือค่าคาดหมายของตัวแปรสุ่มนั้น

โดยทั่วไปการวัดการกระจายของข้อมูลซึ่งประกอบด้วยค่าต่าง ๆ ที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม นิยมวัดโดยใช้ค่าเบี่ยงเบนมาตรฐาน ซึ่งเป็นรากที่สองของความแปรปรวน

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ถ้า X เป็นตัวแปรสุ่มไม่ต่อเนื่อง ซึ่งประกอบด้วย $x_1, x_2, x_3, \ldots, x_N$ และ $P(x_1), P(x_2), P(x_3), \ldots, P(x_N)$ เป็นความน่าจะเป็นของ $x_1, x_2, x_3, \ldots, x_N$ ตามลำดับ ความแปรปรวนของ X คือ

$$\sigma_x^2 = \left[x_1 - E(x)\right]^2 \cdot P(x_1) + \ldots + \left[x_N - E(x)\right]^2 \cdot P(x_N)$$

$$= \sum_{i=1}^{n} \left[x_i - E(x)\right]^2 P(x_i)$$

$$= \sum_{i=1}^{n} \left[x_i - \mu_x\right]^2 P(x_i)$$

นั่นคือ ค่าเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X คือ $\sqrt{\sigma_x^2}$ ซึ่งเท่ากับ σ_x

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ตัวอย่าง 2.10 จงหาความแปรปรวนและค่าเบี่ยงเบนมาตราฐานของ X โดยกำหนดตารางแจกแจงความน่าจะเป็นของ X ดังต่อไปนี้

X	0	1	2	3
$P(x)$	1/30	9/30	15/30	5/30

<u>วิธีทำ</u> $\mu_x = E(X) = \sum_{x=0}^{3} xP(x)$

$$= 0\left(\frac{1}{30}\right) + 1\left(\frac{9}{30}\right) + 2\left(\frac{15}{30}\right) + 3\left(\frac{5}{30}\right)$$

$$= 1.8$$

$$\sigma_x^2 = \sum_{x=0}^{3} (x-1.8)^2 P(x)$$

$$= (0-1.8)^2\left(\frac{1}{30}\right) + (1-1.8)^2\left(\frac{9}{30}\right) + (2-1.8)^2\left(\frac{15}{30}\right) + (3-1.8)^2\left(\frac{5}{30}\right) = 0.56$$

$$\therefore \sigma_x = \sqrt{0.56} = 0.748$$

ค่าเฉลี่ยและความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง

ตัวอย่าง 2.11 ถ้าความน่าจะเป็นที่พนักงานขายคนหนึ่งขายเครื่องครัวได้ในราคาชุดละ 4,500 บาท และ 5,000 บาท เท่ากับ 0.8 และ 0.2 ตามลำดับ (มีความหมายว่าในการขายเครื่องครัวแก่ลูกค้า 100 ราย จะขายในราคา 4,500 บาทได้ 80 ราย และขายในราคา 5,000 บาทได้ 20 ราย) จงหาความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของราคาเครื่องครัว

$x_1 = 4,500, \quad x_2 = 5,000$

$P(x_1) = 0.8, \quad P(x_2) = 0.2$

<u>วิธีทำ</u> จาก $\sigma_x^2 = \sum_{x=1}^{2} (x - \mu_x)^2 P(x)$

$$= (4,500 - 4,600)^2 (0.8) + (5,000 - 4,600)^2 (0.2)$$

$$= 8,000 + 32,000 = 40,000$$

$$\therefore \sigma_x = \sqrt{40,000} = 200$$

$$\mu_x = \sum_{x=1}^{2} x_i P(x_i) = 4,600$$

แบบฝึกหัดบทที่ 2

1. จงพิจารณาว่าตัวแปรสุ่มต่อไปนี้เป็นตัวแปรสุ่มไม่ต่อเนื่องหรือตัวแปรสุ่มต่อเนื่อง

   1.1 จำนวนสินค้าในห้างสรรพสินค้าแห่งหนึ่งผลิตได้ในรอบปีที่ผ่านมา

   1.2 ปริมาณน้ำฝนที่ตกในเขตกรุงเทพและปริมณฑล

   1.3 จำนวนลูกค้าที่ใช้บริการในธนาคารแห่งหนึ่งระหว่างเวลา 08.30 – 12.00 น. ของวันที่ 31 มกราคม พ.ศ. 2561

   1.4 รายได้เจ้าของธุรกิจต่าง ๆ ในจังหวัดนครราชสีมาในเดือนที่ผ่านมา

   1.5 จำนวนรถยนต์ที่ตัวแทนจำหน่ายขายได้ในรอบปีที่ผ่านมา

   1.6 ยอดขายผลิตภัณฑ์เพื่อสุขภาพในเดือนธันวาคมของร้านค้าแห่งหนึ่ง

แบบฝึกหัดบทที่ 2

2. ในรอบปีที่ผ่านมา โรงงานอุตสาหกรรมแห่งหนึ่งซึ่งมีคนงาน 890 คน มีคนลาออก 50 คน จงหาความน่าจะเป็นที่คนงานคนใดคนหนึ่งจะลาออก

3. จากการเก็บรวบรวมข้อมูลเกี่ยวกับจำนวนวันที่ผลิตของเล่นได้ของบริษัทแห่งหนึ่งเป็นดังนี้

จำนวนของเล่นที่ผลิต (ชิ้น/วัน)	จำนวนวันที่ผลิตได้
ต่ำกว่า 2,000 ชิ้น	6
2,000 – 2,499 ชิ้น	2
2,500 – 2,999 ชิ้น	3
3,000 – 3,499 ชิ้น	12
ตั้งแต่ 3,500 ชิ้นขึ้นไป	1

แบบฝึกหัดบทที่ 2

จงหาความน่าจะเป็นที่บริษัทจะผลิตของเล่นได้

3.1 ตั้งแต่ 2,500 – 2,999 ชิ้น/วัน

3.2 ต่ำกว่า 2,000 ชิ้น/วัน

3.3 ตั้งแต่ 2,000 ถึง 3,499 ชิ้น/วัน

4. จากตารางการแจกแจงความน่าจะเป็นของ X ต่อไปนี้

X	1	2	3	4
$P(x)$	0.2	0.4	0.3	0.1

จงหา 4.1 ค่าเฉลี่ย $\left(\mu_x\right)$

4.2 ความแปรปรวน $\left(\sigma_x^2\right)$

แบบฝึกหัดบทที่ 2

5. ถ้าความน่าจะเป็นที่ลูกค้าจะได้รับส่วนลดในอัตราต่าง ๆ จากร้านค้าแห่งหนึ่ง เป็นดังนี้

ส่วนลดการค้า (x)	ความน่าจะเป็น P(x)
5%	0.25
10%	0.12
15%	0.34
25%	0.09
35%	0.20

จงหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของส่วนลดการค้าที่ลูกค้าได้รับจากร้านค้าแห่งนี้

บทที่ 3
การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

3.1 การแจกแจงแบบแบร์นูลลี

นิยาม 3.1 การทดลองแบบแบร์นูลลี คือ การทดลองที่มีลักษณะ ดังนี้

1. ผลลัพธ์แต่ละครั้งของการทดลอง มีเพียง 2 ประเภท คือ ความสำเร็จและความไม่สำเร็จ
2. ในแต่ละครั้งของการทดลอง ความน่าจะเป็นที่จะเกิดความสำเร็จเท่ากับ p และความน่าจะเป็นที่เกิดความไม่สำเร็จเท่ากับ q โดย $p + q = 1$

นิยาม 3.2 ถ้า X เป็นตัวแปรสุ่ม ที่มีค่าเป็น

$$x = \begin{cases} 0 & \text{เมื่อเกิดความไม่สำเร็จ} \\ 1 & \text{เมื่อเกิดความสำเร็จ} \end{cases}$$

แล้วเรียก X ว่า **เป็นตัวแปรสุ่มแบร์นูลลี** และเรียกการแจกแจงความน่าจะเป็นของ X ว่า การแจกแจงแบบแบร์นูลลี โดยมีฟังก์ชันความน่าจะเป็น คือ

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

ทฤษฎี 3.1 ค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบแบร์นูลลี คือ

$$\mu = p \quad , \quad \sigma^2 = pq$$

ตัวอย่าง 3.1 โรงงานทอผ้าแห่งหนึ่งมีพนักงานร้อยละ 80 เป็นเพศหญิง ถ้าให้ X เป็นตัวแปรสุ่ม โดย $x = 1$ เมื่อพนักงานเป็นเพศหญิง และ $x = 0$ เมื่อพนักงานเป็นเพศชาย จงหาค่าเฉลี่ยและความแปรปรวนของ X

วิธีทำ เนื่องจาก $p = 0.8$

ดังนั้น $\mu = p = 0.8$

$$\sigma^2 = pq = 0.8(1-0.8) = 0.16$$

3.2 การแจกแจงแบบทวินาม

นิยาม 3.3 การทดลองแบบทวินาม คือ การทดลองที่ประกอบด้วยการทดลองแบบแบร์นูลลีที่เป็นอิสระต่อกันรวม n ครั้ง

ถ้าทำการทดลองแบบแบร์นูลลี n ครั้ง โดยอิสระต่อกัน จะได้ตัวแปรสุ่มแบบแบร์นูลลีรวม n ตัว คือ $Y_1, Y_2, \ldots, Y_n$

โดย Y_i เป็นตัวแปรสุ่มแบบแบร์นูลลีของการทดลองครั้งที่ i ซึ่งมีค่าเป็น

$$y_i = \begin{cases} 0 & \text{เมื่อเกิดความไม่สำเร็จ} \\ 1 & \text{เมื่อเกิดความสำเร็จ} \end{cases}$$

ให้ $Y_1 + Y_2 + \ldots + Y_n = X$

แล้วเรียก X ว่า **เป็นตัวแปรสุ่มแบบทวินาม** ซึ่งมีค่าเป็น $0, 1, 2, \ldots, n$

ถ้าทำการทดลองแบบทวินามขนาด n แล้วได้ความสำเร็จ x ครั้ง และความไม่สำเร็จ n – x ครั้ง จำนวนวิธีที่เป็นไปได้ทั้งหมด คือ ${}^{n}C_{x} = \frac{n!}{(n-x)!\,x!}$ วิธี ซึ่งแต่ละวิธีเกิดร่วมกันไม่ได้

นิยาม 3.4 ถ้า X เป็นตัวแปรสุ่มแบบทวินาม แล้วจะเรียกการแจกแจงความน่าจะเป็นของ X ว่าการแจกแจงทวินาม ซึ่งแทนด้วย $X \sim B(n\ ;\ p)$ และฟังก์ชันความน่าจะเป็นของ X คือ

$$b(x;\ n,\ p) = {}^{n}C_{x}\, p^{x} q^{n-x} \quad , \quad x = 0,\ 1,\ 2,\ \ldots,\ n$$

ทฤษฎี 3.2 ค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบทวินาม คือ

$$\mu = \text{np} \quad , \quad \sigma^{2} = \text{npq}$$

ตัวอย่าง 3.2 ในการทอดลูกเต๋าเที่ยงตรง 1 ลูก 5 ครั้ง จงหา

ก) ความน่าจะเป็นที่ลูกเต๋าหงายแต้ม 1 ในการทอดลูกเต๋าครั้งแรกและครั้งสุดท้าย

ข) ความน่าจะเป็นที่ลูกเต๋าจะหงายแต้ม 1 รวม 2 ครั้ง

ค) ความน่าจะเป็นที่ลูกเต๋าจะหงายแต้ม 1 อย่างมาก 2 ครั้ง

ง) จำนวนครั้งโดยเฉลี่ยที่ลูกเต๋าจะหงายแต้ม 1

วิธีทำ ให้ Y_i เป็นตัวแปรสุ่มแบบแบร์นูลลีของการทอดลูกเต๋าครั้งที่ i ซึ่งมีค่าเป็น

$$y_i = \begin{cases} 0 & \text{เมื่อลูกเต๋าหงายแต้มอื่นๆ ด้วย } q = 5/6 \\ 1 & \text{เมื่อลูกเต๋าหงายแต้ม 1 ด้วย } p = 1/6 \end{cases}$$

ก) ความน่าจะเป็นที่ลูกเต๋าหงายแต้ม 1 ในการทอดลูกเต๋าครั้งแรกและครั้งสุดท้าย

$$P(Y_1 = 1, Y_2 = 0, Y_3 = 0, Y_4 = 0, Y_5 = 1)$$
$$= P(Y_1 = 1)\,P(Y_2 = 0)\,P(Y_3 = 0)\,P(Y_4 = 0)\,P(Y_5 = 1)$$
$$= (1/6)(5/6)(5/6)(5/6)(1/6)$$
$$= 0.016$$

ข) ให้ X เป็นจำนวนครั้งที่ลูกเต๋าหงายแต้ม 1 จากการทอดลูกเต๋า 1 ลูก 5 ครั้ง

ดังนั้น $X \sim B(5\ ;\ 1/6)$

$$P(X = 2) = b(2\ ;\ 5,\ 1/6)$$
$$= {}^{5}C_{2}(1/6)^{2}(5/6)^{3}$$
$$= 0.1608$$

ค) $$P(X \leq 2) = \sum_{x=0}^{2} b\left(x; 5, 1/6\right)$$

$$= {}^{5}C_{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{5} + {}^{5}C_{1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{4} + {}^{5}C_{2}\left(\frac{1}{6}\right)^{2}\left(\frac{5}{6}\right)^{3}$$

$$= 0.4019 + 0.4019 + 0.1608$$

$$= 0.9646$$

ง) $$E(X) = np = (5)(1/6) = 0.833$$

การใช้ตารางการแจกแจงความน่าจะเป็นแบบทวินาม

ใช้ตารางที่เรียกว่า Cumulative Binomial Probability Distribution ซึ่งแสดงความน่าจะเป็นในลักษณะ

$$P(X \geq r) = P(X \geq r \mid n, p) = \sum_{x=r}^{n} b(x; n, p)$$

โดย n มีค่าตั้งแต่ 5 – 20 และ 25 ส่วน p มีค่าเริ่มจาก 0.01 ถึง 0.5 เมื่อ r และ s เป็นจำนวนเต็มบวกใด ๆ โดย $r < s$ มีวิธีใช้ดังนี้

ก. เมื่อความน่าจะเป็นที่เกิดความสำเร็จ p มีค่าเริ่มจาก 0.01 ถึง 0.5 ได้

1. $P(X = r) = \sum_{x=r}^{n} b(x; n, p) - \sum_{x=r+1}^{n} b(x; n\ p)$

2. $P(X \leq r) = 1 - \sum_{x=r+1}^{n} b(x; n, p)$

3. $P(r \leq X \leq s) = \sum_{x=r}^{n} b(x; n, p) - \sum_{x=s+1}^{n} b(x; n, p)$

ข. เมื่อความน่าจะเป็นที่เกิดความสำเร็จ p มีค่ามากกว่า 0.5 ได้

$$P(X \geq r) = 1 - \sum_{x=0}^{r-1} b(x; n, p)$$

$$= 1 - \sum_{x=n-r+1}^{n} b(x; n, p^*)$$

โดยที่ $p^* = q = 1 - p$ ซึ่งมีค่าน้อยกว่า 0.5

ตัวอย่าง 3.3 ผู้จัดการร้านจำหน่ายเครื่องใช้ไฟฟ้าแห่งหนึ่งทราบว่า จากลูกค้าที่เข้ามาเลือกชมสินค้าเพียงร้อยละ 28 ที่ซื้อสินค้า ถ้ามีลูกค้าเข้ามาเลือกชมสินค้า 12 คน จงหาความน่าจะเป็นที่ลูกค้าซื้อสินค้า

ก) อย่างน้อย 6 คน

ข) อย่างมาก 3 คน

ค) 5 คน

วิธีทำ ให้ X เป็นจำนวนลูกค้าที่ซื้อสินค้า โดย $X \sim B(12\ ;\ 0.28)$

ก) $P(X \geq 6) = \sum_{x=6}^{12} b(x;\ 12,\ 0.28) = 0.0887$

ข) $P(X \leq 3\ ) = 1 - \sum_{x=4}^{12} b(x;\ 12,\ 0.28) = 1 - 0.4452 = 0.5548$

ค) $P(X = 5) = \sum_{x=5}^{12} b(x;\ 12,\ 0.28) - \sum_{x=6}^{12} b(x;\ 12,\ 0.28)$

$= 0.2254 - 0.0887 = 0.1367$

ตัวอย่าง 3.4 ถั่วเหลืองมีอัตราการงอกร้อยละ 80 ถ้าทำการเพาะถั่วเหลืองดังกล่าว 15 เมล็ด จงหาความน่าจะเป็นที่จะมีถั่วเหลืองงอก

ก) อย่างมาก 8 เมล็ด

ข) ตั้งแต่ 9 ถึง 11 เมล็ด

วิธีทำ ให้ X เป็นจำนวนเมล็ดถั่วเหลืองที่งอก โดย $X \sim B(15\ ;\ 0.8)$

ก) $$P(X \leq 8) = \sum_{x=7}^{15} b(x;\ 15,\ 0.2) = 0.0181$$

ข) $$P(9 \leq X \leq 11) = \sum_{x=4}^{15} b(x;\ 15,\ 0.2) - \sum_{x=7}^{15} b(x;\ 15,\ 0.2)$$

$$= 0.3518 - 0.0181$$

$$= 0.3337$$

3.3 การแจกแจงแบบปัวส์ซอง

ให้ X เป็นตัวแปรสุ่มแทนจำนวนครั้งของเหตุการณ์ที่เกิดขึ้นในช่วงเวลาหนึ่งหรืออาณาบริเวณหนึ่งที่สนใจ ซึ่งช่วงเวลานั้นอาจอยู่ในรูปของปี เดือน สัปดาห์ วัน ชั่วโมง หรือนาที และอาณาบริเวณดังกล่าวอาจอยู่ในรูปของพื้นที่ หรือปริมาตร ตัวอย่าง เช่น

- X เป็นจำนวนลูกค้าที่มาใช้บริการเครื่องฝาก-ถอนเงินอัตโนมัติ แห่งหนึ่งใน 1 วัน
- X เป็นจำนวนครั้งที่มีโทรศัพท์ต่อเข้ามาที่หมายเลขหนึ่งระหว่างเวลา 9.00 – 10.00 น.
- X เป็นจำนวนแบคทีเรียในน้ำ 1 หยด
- X เป็นจำนวนคำที่พิมพ์ผิดในหนังสือ 1 หน้า

แล้วจะเรียก X ว่า **ตัวแปรสุ่มแบบปัวส์ซอง** ซึ่งได้จากการทดลองแบบปัวส์ซอง

นิยาม 3.5 ถ้า X เป็นจำนวนครั้งของเหตุการณ์ที่เกิดขึ้นในช่วงเวลา หรืออาณาบริเวณที่สนใจ ที่ได้จากการทดลองแบบปัวส์ซอง แล้วจะเรียก X ว่า ตัวแปรสุ่มแบบปัวส์ซอง และเรียกการแจงแจงความน่าจะเป็นของ X ซึ่งมีฟังก์ชันความน่าจะเป็น $f(x\ ;\ \lambda)$ ว่า **การแจกแจงแบบปัวส์ซอง** (Poisson distribution) โดย

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

เมื่อ e = 2.71828... และ λ คือ จำนวนครั้งของเหตุการณ์ที่เกิดขึ้นโดยเฉลี่ยในช่วงเวลาหรืออาณาบริเวณที่สนใจ โดย $\lambda > 0$

ทฤษฎี 3.3 ค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบปัวส์ซอง คือ

$$\mu = \lambda \quad , \quad \sigma^2 = \lambda$$

ตัวอย่าง 3.5 จากการสำรวจพบว่า ถนนในชนบทสายหนึ่งมีรถยนต์วิ่งผ่าน โดยเฉลี่ยชั่วโมงละ 24 คัน จงหาความน่าจะเป็นที่ถนนสายนี้จะ

ก) มีรถยนต์วิ่งผ่านอย่างมาก 30 คัน ใน 1 ชั่วโมง

ข) มีรถยนต์วิ่งผ่าน 6 คัน ใน 20 นาที

วิธีทำ ก) ให้ X เป็นจำนวนรถยนต์ที่วิ่งผ่านถนนสายนี้ใน 1 ชั่วโมง

$$f(x\,;\,24) = \frac{e^{-24}\,24^{x}}{x!}\;;\; x = 0, 1, 2, \ldots$$

$$P(X \leq 30) = 1 - \sum_{x=31}^{\infty} f(x\,;\,24)$$

$$= 1 - 0.096$$

$$= 0.904$$

ข) ให้ X เป็นจำนวนรถยนต์ที่วิ่งผ่านถนนสายนี้ใน 20 นาที

โดย $\lambda = \frac{(24)(20)}{60} = 8$ คัน ใน 20 นาที

และ $$f(x\,;\,8) = \frac{e^{-8}\,8^{x}}{x!} \quad ; \quad x = 0, 1, 2, \ldots$$

$$P(X = 6) = \sum_{x=6}^{\infty} f(x\,;\,8) - \sum_{x=7}^{\infty} f(x\,;\,8)$$

$$= 0.809 - 0.687$$

$$= 0.122$$

แบบฝึกหัดท้ายบท บทที่ 3

1. ในมหาวิทยาลัยแห่งหนึ่ง พบว่า 20 % ของนักศึกษาที่ถนัดมือซ้าย ถ้าเราสุ่มนักศึกษามา 20 คน จงหา

   ก) ความน่าจะเป็นที่พบนักศึกษาที่ถนัดมือซ้าย 5 คน

   ข) ค่าเฉลี่ยของจำนวนนักศึกษาที่ถนัดมือซ้าย

   ค) ความแปรปรวนของจำนวนนักศึกษาที่ถนัดมือซ้าย

2. กำหนดให้ตัวแปรสุ่ม $X \sim B(10\ ;\ 0.3)$ จงหา

   ก) $P(X = 5)$

   ข) $P(X \leq 7)$

3. จำนวนอุบัติเหตุรถชนกันในมหาวิทยาลัยเชียงใหม่ โดยเฉลี่ยเกิดขึ้น 4 รายต่อวัน จงหาความน่าจะเป็นที่

   ก) ไม่เกิดอุบัติเหตุรถชนกันเลยในวันใด ๆ

   ข) มีอุบัติเหตุเกิดขึ้น 6 ราย

4. ผู้จัดการศูนย์การค้าแห่งหนึ่ง สังเกตพบว่า โดยเฉลี่ยแล้วจะมีลูกค้ามาชำระเงินที่พนักงานจ่ายเงิน จำนวน 16 คน ใน 10 นาที ดังนั้นเพื่อให้การจัดการเกี่ยวกับจำนวนพนักงานจ่ายเงินและจัดตารางเวลาที่ทำงานให้เหมาะสมจึงต้องการทราบความน่าจะเป็นดังต่อไปนี้

ก) ความน่าจะเป็นที่จะมีลูกค้าอย่างมาก 5 คน มาชำระเงินที่พนักงานจ่ายเงินใน 1 นาทีใด ๆ

ข) โอกาสที่จะไม่มีลูกค้าเลยมาชำระเงินที่พนักงานจ่ายเงินใน 1 นาทีใด ๆ

ค) ความน่าจะเป็นที่จะมีลูกค้าอย่างน้อย 2 คน มาชำระเงินในช่วงเวลา 3 นาทีใด ๆ ของวันพรุ่งนี้

บทที่ 4
การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง

การแจกแจงแบบปกติ

นิยาม 4.1 ถ้า X เป็นตัวแปรสุ่มแบบปกติ (Normal random variable) ที่มีฟังก์ชันความน่าจะเป็น

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

โดยพารามิเตอร์ $-\infty < \mu < \infty$ และ $\sigma > 0$ ค่าคงตัว $\pi \approx 3.14159$ และ $e \approx 2.71828$

แล้วจะเรียก การแจกแจงความน่าจะเป็นของ X ว่า **การแจกแจงปกติ (Normal distribution)**

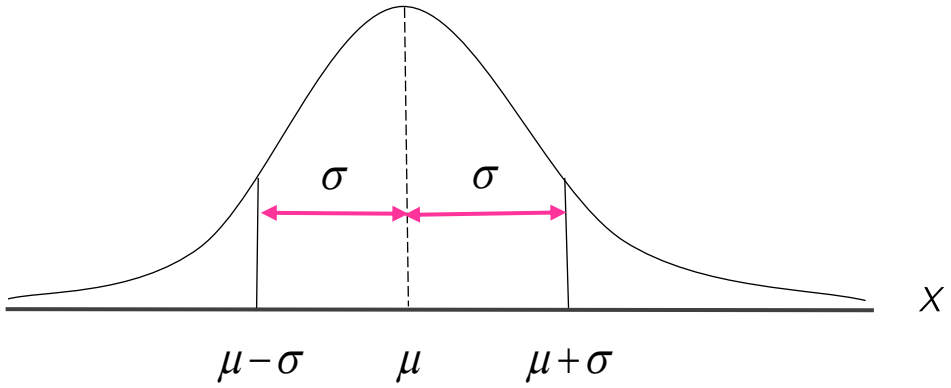
ทฤษฎีบท 4.1 ค่าเฉลี่ยและความแปรปรวนของการแจกแจงปกติ คือ

$$E(X) = \mu \qquad Var(X) = \sigma^2$$

ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 จะเขียนแทนด้วย $X \sim N(\mu, \sigma^2)$

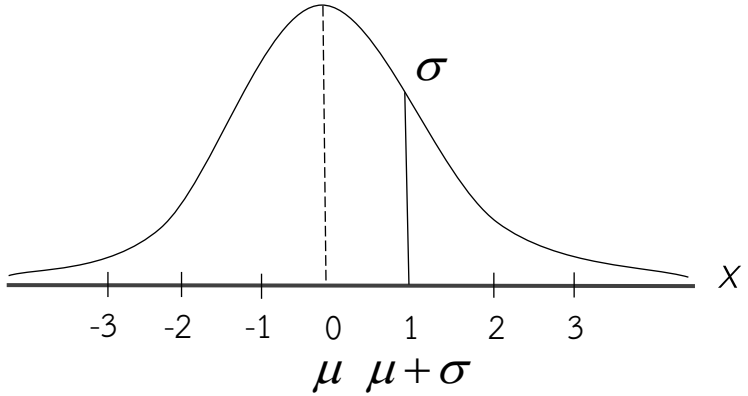
เส้นโค้งปกติ (Normal curve)

จากฟังก์ชันความน่าจะเป็น $f(x)$ ของตัวแปรสุ่มแบบปกติ X ถ้าทราบค่าพารามิเตอร์ คือค่าเฉลี่ย μ และความแปรปรวน σ^2 โดยแทนค่า x หลาย ๆ ค่าก็จะได้ค่าความน่าจะเป็น หรือ $f(x)$ ของแต่ละค่าของตัวแปรสุ่ม X และหากนำค่าดังกล่าวมาลงจุดก็จะได้กราฟเส้นโค้ง 1 เส้น ซึ่งเรียกว่า **โค้งปกติ (Normal curve)** โดยค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานจะแสดงถึงตำแหน่งและรูปร่างของโค้งปกติ



คุณสมบัติของโค้งปกติ

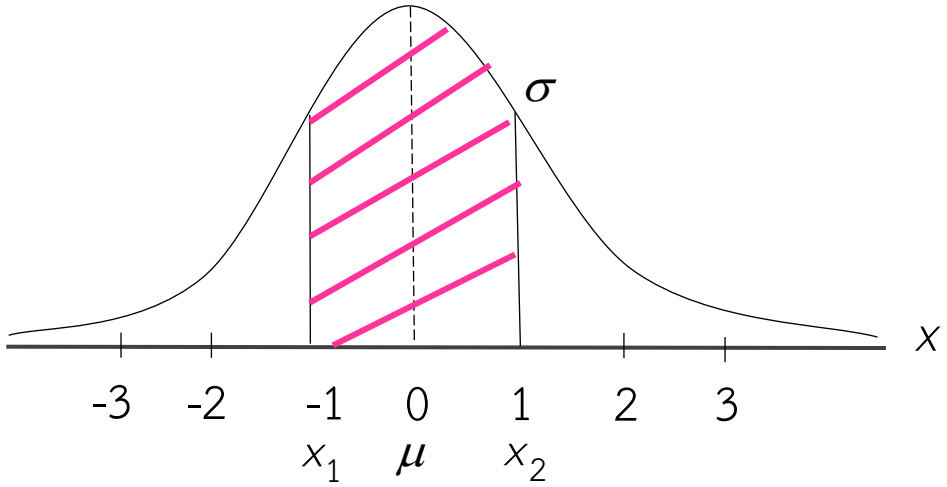
1. เป็นโค้งรูประฆังคว่ำ โดยปลายโค้งทั้งสองข้างใกล้แกน X แต่ไม่พบแกน X
2. โค้งมีค่าสูงสุด เมื่อ $x = \mu$
3. โค้งสมมาตรรอบค่าเฉลี่ย μ
4. โค้งมีค่าเฉลี่ย มัธยฐาน และ ฐานนิยมเท่ากัน
5. จุดเปลี่ยนโค้งอยู่ที่ $x = \mu \pm \sigma$



6. พื้นที่ทั้งหมดใต้โค้งที่เหนือแกน X รวมกัน เท่ากับ 1 และมีความหมายในเชิงความน่าจะเป็น นั่นคือ

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1 \quad \text{และ} \quad \int_{-\infty}^{\mu} f(x)\mathrm{d}x = 0.5 = \int_{\mu}^{\infty} f(x)\mathrm{d}x$$

พื้นที่ใต้โค้งปกติ



พื้นที่ทั้งหมดใต้โค้งที่เหนือแกน X รวมกัน เท่ากับ 1

โดยที่

$$P\left(x_1 < X < x_2\right) = \int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

แปลงค่าตัวแปรสุ่มปกติ X ที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 ไปเป็นตัวแปรสุ่มใหม่เรียกว่า **ตัวแปรสุ่มแบบปกติมาตรฐาน** (Standard normal random variable) ซึ่งเขียนแทนด้วย Z โดย

$$Z = \frac{X - \mu}{\sigma}$$

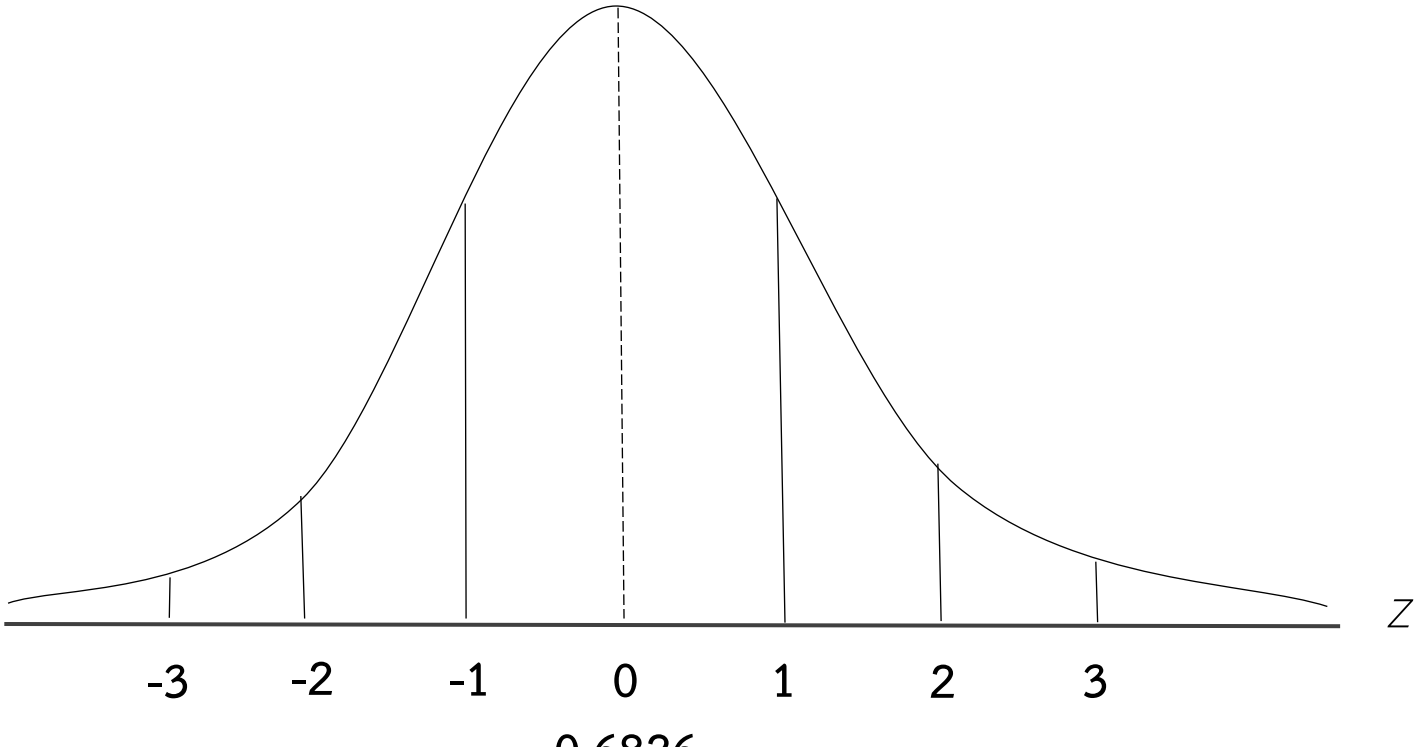
ซึ่ง $\mu_Z = 0$ และ $\sigma_Z^2 = 1$

โดยตัวแปรสุ่มแบบปกติมาตรฐาน Z จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และความแปรปรวน 1 ซึ่งเรียกว่า **การแจกแจงแบบปกติมาตรฐาน**

นิยาม 4.2 ถ้า Z เป็นตัวแปรสุ่มแบบปกติมาตรฐานที่มีค่าเฉลี่ย 0 และความแปรปรวน 1 แล้วจะเรียกการแจกแจงความน่าจะเป็นของ Z ซึ่งมีฟังก์ชันความน่าจะเป็น $f(z)$ว่า การแจกแจงปกติมาตรฐาน (Standard normal distribution) ซึ่งเขียนแทนด้วย $Z \sim N(0,1)$ โดย

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

จากฟังก์ชันความน่าจะเป็น $f(z)$ ของการแจกแจงแบบปกติมาตรฐานจะได้โค้งปกติมาตรฐาน (Standard normal curve) ซึ่งมีพื้นที่ใต้โค้งในช่วงต่าง ๆ ดังรูป

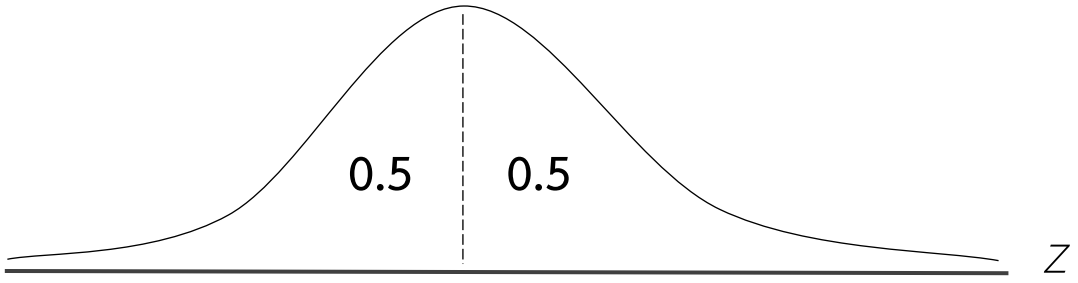

Z
-3
-2
-1
0
1
2
3
0.6826
0.9544
0.9974

จาก $X \sim N(\mu, \sigma^2)$ เมื่อต้องการคำนวณความน่าจะเป็นในช่วงระหว่าง x_1 และ x_2 จะได้

$$P\left(x_1 < X < x_2\right) = P\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}\right)$$

$$= P\left(z_1 < Z < z_2\right)$$

$P(0 < Z < z)$ คือ พื้นที่ใต้โค้งปกติมาตรฐานที่ Z มีค่าอยู่ระหว่าง 0 ถึง z และโดยที่โค้งปกติมาตรฐาน**สมมาตรรอบค่าเฉลี่ย 0** ซึ่งทำให้พื้นที่ซีก**ซ้ายและขวา**มีค่าเท่ากับ 0.5



ตัวอย่าง 4.1 ถ้า $Z \sim N(0,1)$ จงหาความน่าจะเป็นต่อไปนี้

1. $P(Z \leq 1.17)$ *Ans* 0.8790
2. $P(Z > 2.23)$ *Ans* 0.0129
3. $P(0.42 < Z < 1.61)$ *Ans* 0.2835
4. $P(-1.61 < Z < 0.59)$ *Ans* 0.6687

ตัวอย่าง 4.2 ถ้า $X \sim N(12,4)$ จงหาความน่าจะเป็นเมื่อ X อยู่ระหว่าง 11.26 และ 18.44

<u>วิธีทำ</u> $P(11.26 < X < 18.44) = P\left(\frac{11.26-12}{2} < \frac{X-\mu}{\sigma} < \frac{18.44-12}{2}\right)$

$= P(-0.37 < Z < 3.22)$

$= 0.5 + 0.1443 = 0.6443$

ตัวอย่าง 4.3 ถ้าส่วนสูงของนักศึกษาชายในมหาวิทยาลัยแห่งหนึ่ง มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 67.56 นิ้ว และค่าเบี่ยงเบนมาตรฐาน 2.57 นิ้ว จงหาความน่าจะเป็นที่นักศึกษาชายจะมีส่วนสูงมากกว่า 62 นิ้ว

<u>วิธีทำ</u> ให้ X เป็นส่วนสูงของนักศึกษาชายในมหาวิทยาลัยดังกล่าว หน่วย : นิ้ว

$X \sim N(67.56 , 2.57^2)$ ต้องการหาความน่าจะเป็นที่นักศึกษาชายจะสูงมากกว่า 62 นิ้ว

$$P(X > 62) = P\left(\frac{X-\mu}{\sigma} > \frac{62-67.56}{2.57}\right)$$

$$= P(Z > -2.16)$$

$$= 0.5 + 0.4846 = 0.9846$$

นั่นคือ ความน่าจะเป็นที่นักศึกษาชายจะมีส่วนสูงมากกว่า 62 นิ้ว เท่ากับ 0.9846

ตัวอย่าง 4.4 รายได้ต่อสัปดาห์ของพนักงานที่ทำงานในโรงงานผลิตขวดแก้วมีการแจกแจงปกติที่มีค่าเฉลี่ย 2,000 บาท และมีค่าเบี่ยงเบนมาตรฐาน 100 บาท โดยโรงงานแห่งนี้มีระยะเวลาทำงานสัปดาห์ละ 5 วันและพนักงานสามารถทำงานล่วงเวลาได้ไม่เกิน 2 ชั่วโมงต่อวัน ซึ่งจะได้รับค่าล่วงเวลาชั่วโมงละ 60 บาท จงหาความน่าจะเป็นที่พนักงานจะมีรายได้อยู่ระหว่าง 1,950 ถึง 2,100 บาท

<u>วิธีทำ</u> ให้ X เป็นรายได้ต่อสัปดาห์ของพนักงาน หน่วย : บาท $X \sim N(2{,}000\ ,\ 100^2)$

ต้องการหา $P\left(1{,}950 < X < 2{,}100\right) = P\left(\frac{1{,}950-2{,}000}{100} < \frac{X-\mu}{\sigma} < \frac{2{,}100-2{,}000}{100}\right)$

$= P\left(-0.5 < Z < 1\right)$

$= 0.1915+0.3413 = 0.5328$

นั่นคือ ความน่าจะเป็นที่พนักงานจะมีรายได้อยู่ระหว่าง 1,950 ถึง 2,100 บาท เท่ากับ 0.5328

ตัวอย่าง 4.5 ถ้าน้ำหนักของมะเขือเทศในเข่งใบหนึ่งซึ่งมีจำนวน 300 ผล มีการแจกแจงปกติที่มีค่าเฉลี่ย 68 กรัม และค่าเบี่ยงเบนมาตรฐาน 3 กรัม จงหาว่ามีมะเขือเทศอยู่กี่ผลที่คาดว่าจะมีน้ำหนักตั้งแต่ 64 ถึง 72 กรัม

<u>วิธีทำ</u> ให้ X เป็นน้ำหนักของมะเขือเทศในเข่งดังกล่าว หน่วย : กรัม $X \sim N(68, 3^2)$

ต้องการหา $P(64 < X < 72) = P\left(\frac{64-68}{3} < \frac{X-\mu}{\sigma} < \frac{72-68}{3}\right)$

$= P(-1.33 < Z < 1.33)$

$= 0.4082+0.4082$

$= 0.8164$

นั่นคือ จำนวนมะเขือเทศที่คาดว่าจะมีน้ำหนักตั้งแต่ 64 ถึง 72 กรัม มี 300(0.8164) = 244.92 ประมาณ 245 ผล

ตัวอย่าง 4.6 ถ้า $Z \sim N(0,1)$ จงหาค่าของ z_0 ต่อไปนี้

1. $P(Z < z_0) = 0.975 \quad : \quad z_0 = 1.96$
2. $P(Z \leq z_0) = 0.1977 \quad : \quad z_0 = -0.85$
3. $P(Z > z_0) = 0.05 \quad : \quad z_0 = 1.645$

การประมาณค่าการแจกแจงแบบทวินามด้วยการแจกแจงแบบปกติ

ถ้า n มี**ค่ามาก** และ p มีค่า**ไม่เข้าใกล้** 0 หรือ 1 แล้วสามารถประมาณค่าการแจกแจงแบบทวินามด้วยการแจกแจงแบบปกติได้ ดังต่อไปนี้

ทฤษฎีบท 4.2 ให้ X เป็นตัวแปรสุ่มแบบทวินาม ที่มีค่าเฉลี่ย $\mu = np$ และความแปรปรวน $\sigma^2 = npq$ ถ้า n มีค่ามาก และ p มีค่า**ไม่เข้าใกล้** 0 หรือ 1 แล้วตัวแปรสุ่ม

$$Z = \frac{X - np}{\sqrt{npq}}$$

มีการแจกแจงใกล้เคียงแบบปกติมาตรฐาน ซึ่งเขียนแทนด้วย $Z \sim N(0, 1)$

เนื่องจากตัวแปรสุ่มแบบทวินาม X เป็นตัวแปรสุ่มชนิดไม่ต่อเนื่อง จะต้องมีการปรับแก้เพื่อความต่อเนื่องของ X โดยการ**บวกหรือลบด้วย 0.5** เพื่อให้ครอบคลุมพื้นที่หรือความน่าจะเป็นที่ต้องการ

ความน่าจะเป็นของตัวแปรสุ่ม x	วิธีปรับเป็นค่าต่อเนื่อง
$P(X = a)$	$P(a-0.5 \leq X \leq a+0.5)$
$P(X \leq b)$	$P(X \leq b+0.5)$
$P(a \leq X \leq b)$	$P(a-0.5 \leq X \leq b+0.5)$

ตัวอย่าง 4.7 เป็นที่ทราบกันว่าแม่วัวที่ได้รับการผสมเทียมมีอัตราการตั้งท้อง 40% ถ้าแม่วัว 15 ตัวได้รับการผสมเทียม จงหาความน่าจะเป็นที่จะมีแม่วัวตั้งท้องตั้งแต่ 7 ถึง 10 ตัว

<u>วิธีทำ</u> ให้ X เป็นจำนวนแม่วัวที่ตั้งท้อง $X \sim B(15, 0.4)$

ต้องการหา $P(7 \leq X \leq 10) = \sum_{x=7}^{10} b(x; 15, 0.4)$

$$= \sum_{x=7}^{15} b(x; 15, 0.4) - \sum_{x=11}^{15} b(x; 15, 0.4)$$

$$= 0.3902 - 0.0093$$

$$= 0.3809$$

โดยการประมาณค่าด้วยการแจกแจงปกติ ที่มี $\mu = np = (15)(0.4) = 6$ และ

$$\sigma = \sqrt{npq} = \sqrt{(15)(0.4)(0.6)} = 1.9$$

ดังนั้น $P(7 \le X \le 10) = \sum_{x=7}^{10} b(x ; 15 , 0.4)$

$$\approx P\left(\frac{(7-0.5)-6}{1.9} \le \frac{X-np}{\sqrt{npq}} \le \frac{(10+0.5)-6}{1.9}\right)$$

$$= P(0.26 \le Z \le 2.37)$$

$$= 0.4911 - 0.1026$$

$$= 0.3885$$

จะเห็นได้ว่ามีค่าใกล้เคียงกับค่าที่แท้จริงมาก

ตัวอย่าง 4.8 ถ้าร้อยละ 30 ของนักเรียนระดับประถมเป็นโรคฟันผุ จงหาความน่าจะเป็นของการตรวจสุขภาพฟันนักเรียนระดับประถม 1,000 คน แล้วพบนักเรียนเป็นโรคฟันผุ

1. อย่างมาก 279 คน
2. 316 คน

<u>วิธีทำ</u> ให้ X เป็นจำนวนนักเรียนระดับประถมที่เป็นโรคฟันผุ $X \sim B(1{,}000\ ,\ 0.3)$

ดังนั้น $\mu = np = (1000)(0.3) = 300$

และ $\sigma = \sqrt{npq} = \sqrt{(1000)(0.3)(0.7)} = 14.49$

1. $$P(X \leq 279) = \sum_{x=0}^{279} b(x\ ;1{,}000\ ,0.3)$$

$$\approx P\left(\frac{X-np}{\sqrt{npq}} \leq \frac{(279+0.5)-300}{14.49}\right)$$

$$= P(Z \leq -1.41)$$

$$= 0.5-0.4207$$

$$= 0.0793$$

2. $P(X = 316) = b(316 \ ; \ 1{,}000 \ , \ 0.3)$

$$\approx P\left(\frac{(316-0.5)-300}{14.49} \leq \frac{X-np}{\sqrt{npq}} \leq \frac{(316+0.5)-300}{14.49}\right)$$

$$= P(1.07 \leq Z \leq 1.14)$$

$$= 0.3729 - 0.3577$$

$$= 0.0152$$

แบบฝึกหัดบทที่ 4

1. ถ้า $Z \sim N(0,1)$ จงหาความน่าจะเป็นต่อไปนี้

   1.1 $P(Z \leq 1.52)$

   1.2 $P(-0.46 < Z < 2.21)$

   1.3 $P(0.81 < Z < 1.94)$

   1.4 $P(Z > -1.28)$

2. ถ้า $Z \sim N(0,1)$ จงหาค่าของ z_0 ต่อไปนี้

   2.1 $P(Z < z_0) = 0.8621$

   2.2 $P(Z > z_0) = 0.025$

   2.3 $P(Z < z_0) = 0.1314$

3. ถ้าค่าจ้างแรงงานต่อวันของพนักงานในนิคมอุตสาหกรรมแห่งหนึ่ง มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย 190 บาท และความแปรปรวน 64 บาท จงหาความน่าจะเป็นที่พนักงานจะได้รับค่าจ้างแรงงานต่อวัน

3.1 น้อยกว่า 205 บาท

3.2 ตั้งแต่ 180 ถึง 220 บาท

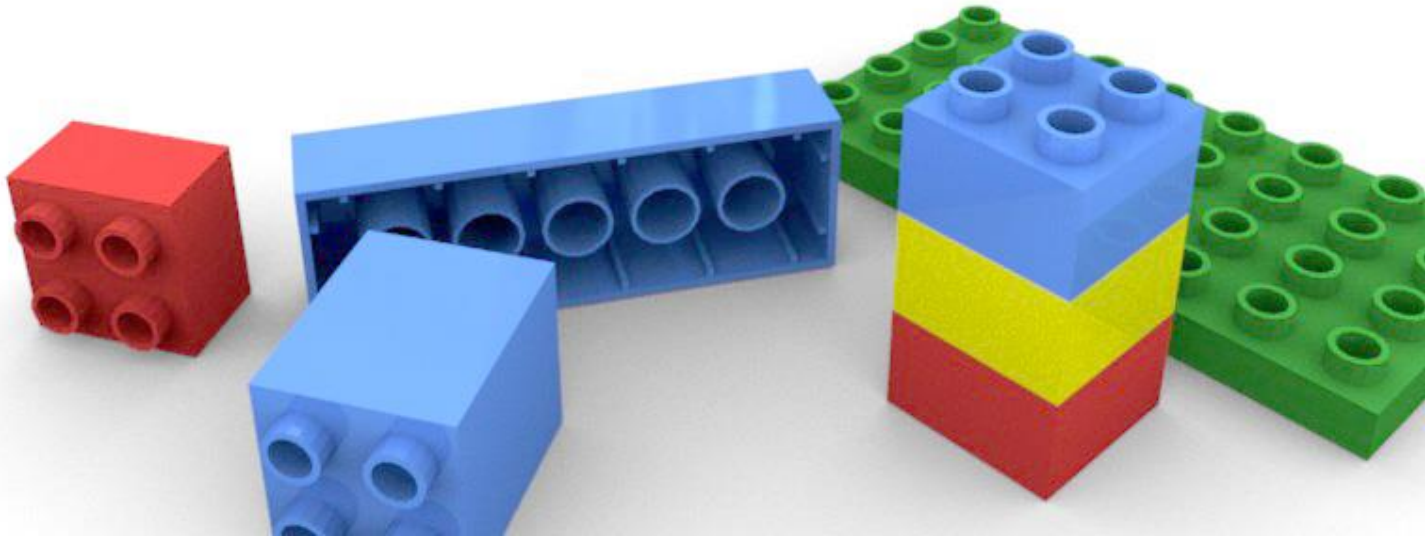
4. นักบาสเก็ตบอลผู้หนึ่งมีความแม่นยำในการโยนลูกบอลลงห่วง 70% ในฤดูกาลแข่งขันครั้งหนึ่ง ถ้านักบาสเก็ตบอลผู้นี้พยายามโยนลูกบอลลงห่วง 120 ครั้ง จงหาความน่าจะเป็นที่จะโยนได้สำเร็จ

4.1 อย่างมาก 90 ครั้ง

4.2 88 ครั้ง

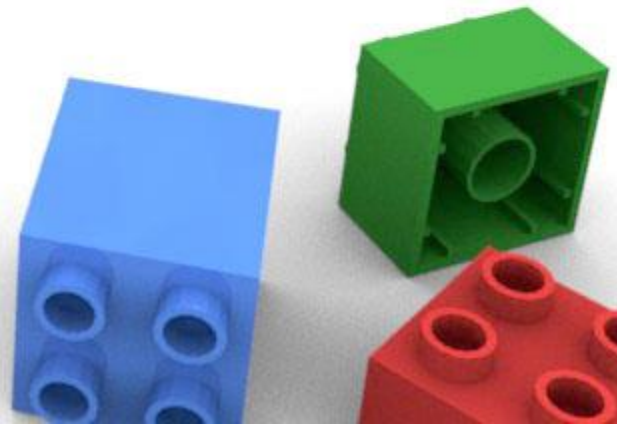
บทที่ 5

การกำหนดกลุ่มตัวอย่าง



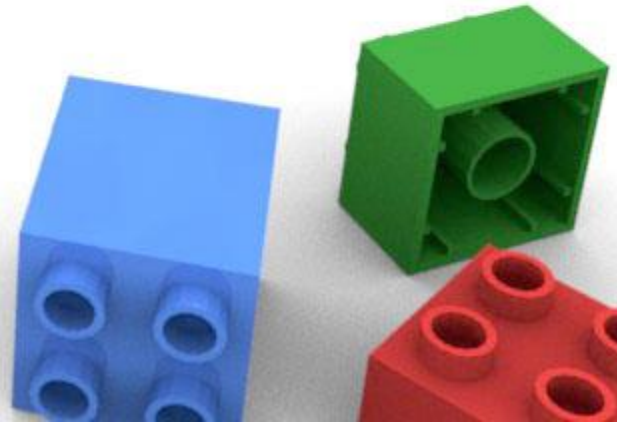
การกำหนดประชากรเป้าหมายเพื่อใช้ในการวิจัยเรื่องหนึ่ง ๆ ถือว่าเป็นตอนที่สำคัญอีกขั้นตอนหนึ่ง และเนื่องจากงานวิจัยบางเรื่องประชากรที่จะศึกษามีจำนวนมาก ผู้วิจัยไม่สามารถเก็บรวบรวมข้อมูลจากประชากรทั้งหมดได้อย่างครบถ้วนภายใต้ข้อจำกัดในเรื่องเวลา แรงงาน และงบประมาณ ดังนั้น ผู้วิจัยจึงนิยมที่จะศึกษาข้อมูลเพียงบางส่วนหรือศึกษาจากตัวแทนของประชากรซึ่งเรียกว่า **กลุ่มตัวอย่าง** ที่ผู้วิจัยเลือกมาจากประชากร และกลุ่มตัวอย่างนั้นจะต้องเป็นตัวแทนของประชากรได้ โดยกลุ่มตัวอย่างจะต้องมีลักษณะเหมือนประชากรและมีขนาดใหญ่พอที่จะเป็นตัวแทนของประชากรได้ตลอดจนสามารถนำผลการวิจัยอ้างอิงไปสู่ประชากรได้อย่างถูกต้องตรงตามเป็นจริง ผู้วิจัยจึงจำเป็นต้องกำหนดขนาดของกลุ่มตัวอย่างและวิธีการสุ่มที่เหมาะสมอีกด้วย

การกำหนดกลุ่มตัวอย่างมีความจำเป็นอย่างยิ่ง ทั้งนี้เนื่องจากการเก็บข้อมูลกับประชากรทุกหน่วยอาจทำให้เสียเวลาและค่าใช้จ่ายที่สูงมากและบางครั้งเป็นเรื่องที่ต้องตัดสินใจภายในเวลาจำกัด การเลือกศึกษาเฉพาะบางส่วนของประชากรจึงเป็นเรื่องที่มีความจำเป็น เพื่อให้มีความเข้าใจในการเลือกตัวอย่าง จะขอนำเสนอความหมายของคำที่เกี่ยวข้อง ดังนี้



ความหมายของประชากรและกลุ่มตัวอย่าง

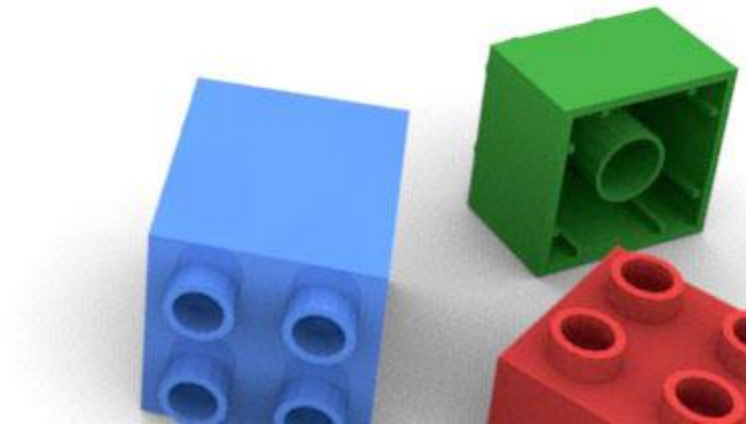
ประชากร (Population) หมายถึง สมาชิกทุกหน่วยของสิ่งที่สนใจศึกษา ซึ่งไม่ได้หมายถึงคนเพียงอย่างเดียว ประชากรอาจจะเป็นสิ่งของ เวลา สถานที่ ฯลฯ เช่น ถ้าสนใจว่าความคิดเห็นของคนไทยที่มีต่อการเลือกตั้ง ประชากร คือคนไทยทุกคน หรือถ้าสนใจอายุการใช้งานของเครื่องคอมพิวเตอร์ยี่ห้อหนึ่ง ประชากรคือเครื่องคอมพิวเตอร์ยี่ห้อนั้นทุกเครื่อง แต่การเก็บข้อมูลกับประชากรทุกหน่วยอาจทำให้เสียเวลาและค่าใช้จ่ายที่สูงมากและบางครั้งเป็นเรื่องที่ต้องตัดสินใจภายในเวลาจำกัด การเลือกศึกษาเฉพาะบางส่วนของประชากรจึงเป็นเรื่องที่มีความจำเป็น เรียกว่า กลุ่มตัวอย่าง



ประเภทของประชากร แบ่งเป็น 2 ประเภท คือ

1. ประชากรที่มีจำนวนจำกัด (Finite population) หมายถึง ประชากรมีปริมาณสามารถนับจำนวนได้ครบถ้วน เช่น จำนวนนักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์ ปีการศึกษา 2549 หรือจำนวนหนังสือที่มีในสำนักวิทยบริการ มหาวิทยาลัยราชภัฏบุรีรัมย์ เป็นต้น

2. ประชากรที่มีจำนวนไม่จำกัด (Infinite population) หมายถึง ประชากรที่มีปริมาณที่ไม่สามารถนับจำนวนได้ครบถ้วน เช่น ดาวบนท้องฟ้า น้ำในมหาสมุทร จำนวนข้าวสารในหนึ่งกระสอบ เป็นต้น



กลุ่มตัวอย่าง (Sample) หมายถึง ประชากรส่วนหนึ่งของประชากรทั้งหมด มีผู้วิจัยเลือกมาเป็นตัวแทนโดยมีคุณสมบัติต่าง ๆ ครบถ้วนเท่าเทียมกัน เพื่อให้เป็นตัวแทนที่ดี ซึ่งหมายถึงกลุ่มตัวอย่างที่ค่าสถิติ (statistic) ที่คำนวณได้ใกล้เคียงหรือเท่าค่าพารามิเตอร์ (parameter) ของประชากร โดยคำนึงถึงกลุ่มตัวอย่างที่เป็นตัวแทนได้จริงและมีจำนวนเหมาะสม คือมีจำนวนมากพอที่จะทดสอบความเชื่อถือโดยวิธีการทางสถิติได้

ค่าสถิติ คือ คำที่ใช้อธิบายตัวแปรในตัวอย่าง โดยคำนวณจากตัวอย่างที่เลือกสุ่มตัวอย่างขึ้นมา

ค่าพารามิเตอร์ คือ ค่าที่ใช้อธิบายตัวแปรในประชากร โดยคำนวณจากค่าของประชากร

โคแครน (Cochran, 1996 : p 2) ได้กล่าวถึงประโยชน์จากการใช้กลุ่มตัวอย่างดังนี้

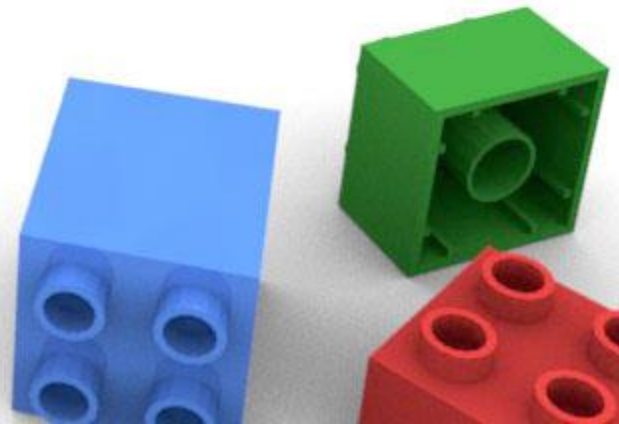
1.) ประหยัดงบประมาณ เวลา และแรงงานในการวิจัย เนื่องจากกลุ่มตัวอย่างจะมีจำนวนน้อยกว่าจำนวนประชากร ทำให้สามารถเก็บรวบรวมข้อมูลได้รวดเร็วและสามารถดำเนินการวิจัยได้เสร็จตามเวลาที่กำหนด

2.) มีความถูกต้อง แม่นยำ และมีความเชื่อมากกว่า เนื่องจากได้จัดกระทำกับกลุ่มที่มีจำนวนน้อยจะง่ายและได้ผลดีกว่าศึกษากับประชากรทั้งหมด

3.) ข้อมูลบางอย่างผู้วิจัยไม่สามารถศึกษากับประชากรทั้งหมดได้ จำเป็นต้องศึกษาจากกลุ่มตัวอย่างแทน เช่น ความคิดเห็นของประชาชนคนไทยที่มีต่อการเลือกซื้อสินค้าในห้างสรรพสินค้า ซึ่งไม่สามารถเก็บข้อมูลจากประชาชนทั้งประเทศ จึงต้องเลือกศึกษาเฉพาะจังหวัดใดจังหวัดหนึ่งเท่านั้น

การกำหนดขนาดของกลุ่มตัวอย่าง

ขนาดของกลุ่มตัวอย่าง (Sample size) หมายถึง จำนวนสมาชิกกลุ่มตัวอย่าง ซึ่งผู้วิจัยจะต้องกำหนดกลุ่มตัวอย่างว่าจะใช้จำนวนเท่าใด หากกำหนดขนาดของกลุ่มตัวอย่างจำนวนมากจะทำให้ผลการวิเคราะห์ข้อมูลมีความเชื่อมั่นสูง เพราะโอกาสที่จะเกิดความคลาดเคลื่อนมีน้อย ซึ่งจะมีค่าใกล้เคียงกับการคำนวณจากประชากรมากกว่ากลุ่มตัวอย่างจำนวนน้อยและในทำนองเดียวกัน ถ้ากลุ่มตัวอย่างมีขนาดเล็ก จะทำให้ผลการวิเคราะห์มีความคลาดเคลื่อนสูง อย่างไรก็ตาม การกำหนดขนาดกลุ่มตัวอย่างจะขึ้นอยู่กับขนาดและลักษณะของประชากร ถ้าประชากรมีลักษณะคล้ายกัน สามารถสุ่มตัวอย่างมาศึกษาเพียงเล็กน้อย แต่ถ้าประชากรมีลักษณะที่แตกต่างกัน จะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ วิธีการกำหนดขนาดกลุ่มตัวอย่าง มีดังนี้



1. การกำหนดขนาดของกลุ่มตัวอย่างโดยใช้เกณฑ์

การกำหนดขนาดของกลุ่มตัวอย่างโดยใช้เกณฑ์ ผู้วิจัยจะต้องทราบจำนวนประชากรที่ค่อนข้างแน่นอน แล้วคำนวณหาจำนวนกลุ่มตัวอย่างจากเกณฑ์ต่อไปนี้

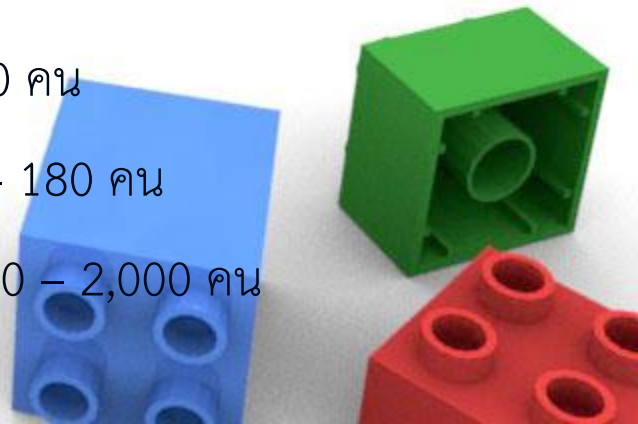
ถ้า	100	<= N<1,000	กำหนดให้ใช้กลุ่มตัวอย่าง 15-30% ของ N
ถ้า	1,000	<= N<10,000	กำหนดให้ใช้กลุ่มตัวอย่าง 10-15% ของ N
ถ้า	10,000	<= N<100,000	กำหนดให้ใช้กลุ่มตัวอย่าง 5-10% ของ N
ถ้า	100,000	<= N<1,000,000	กำหนดให้ใช้กลุ่มตัวอย่าง 1-5% ของ N

เมื่อ N คือ ขนาดหรือจำนวนประชากรทั้งหมด เช่น

ตัวอย่าง จำนวนประชากรมี 300 คน ใช้กลุ่มตัวอย่าง 45 – 90 คน

จำนวนประชากรมี 1,200 คน ใช้กลุ่มตัวอย่าง 120 – 180 คน

จำนวนประชากรมี 20,000 คน ใช้กลุ่มตัวอย่าง 1,000 – 2,000 คน



2. การกำหนดขนาดของกลุ่มตัวอย่างโดยใช้สูตรคำนวณ

การกำหนดขนาดของกลุ่มตัวอย่างโดยใช้สูตรคำนวณ สามารถคำนวณได้ดังนี้

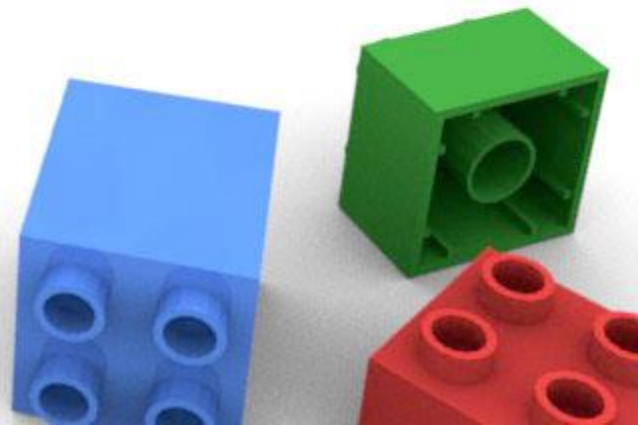
2.1 กรณีทราบจำนวนประชากรและกำหนดค่าความคลาดเคลื่อนโดยใช้สูตรของทาโร ยามาเน่ (Taro Yamane, 1973, P.125)

สูตร $$n = \frac{N}{1 + Ne^2}$$

กำหนดให้ n = ขนาดของกลุ่มตัวอย่าง

N = ขนาดของประชากร

e = ความคลาดเคลื่อนที่ยอมรับได้



ตัวอย่าง ให้การวิจัยเรื่องหนึ่งมีประชากร 5,000 คน และยอมให้เกิดความคลาดเคลื่อนในการสุ่มร้อยละ 5 ขนาดของกลุ่มตัวอย่างจะมีจำนวนเท่าใด

จากสูตร $n = \frac{N}{1+Ne^2}$

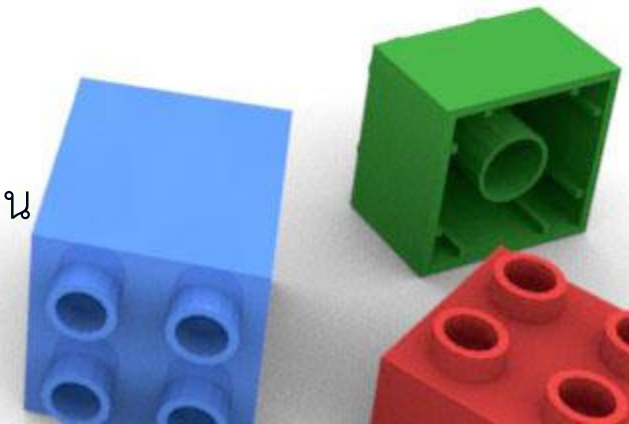
และ $N = 5,000$, $e = 0.05$

ดังนั้น $n = \frac{5,000}{1+5,000(0.05)^2}$

$= \frac{5,000}{1+12.5}$

$= 370.37$

นั่นคือ กลุ่มตัวอย่างที่ใช้ในการศึกษา เท่ากับ 370 คน



2. การกำหนดขนาดของกลุ่มตัวอย่างโดยใช้สูตรคำนวณ

2.2 กรณีไม่ทราบจำนวนประชากร ทราบแต่เพียงว่ามีจำนวนมากใช้สูตรดังนี้ (บุญชม ศรีสะอาด, 2538, หน้า 38)

จากสูตร $n = \frac{P(1-P)Z^2}{e^2}$

กำหนดให้ n = ขนาดของกลุ่มตัวอย่าง

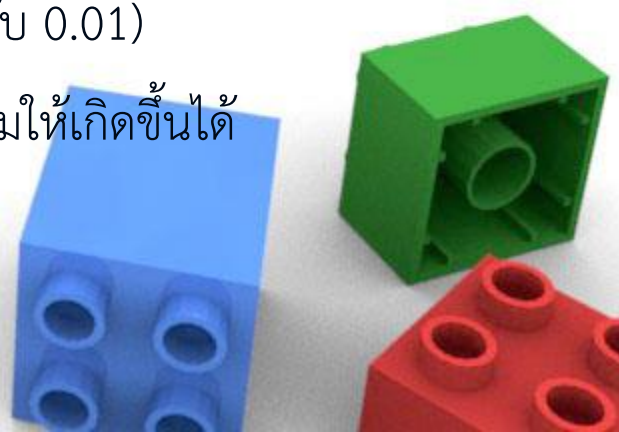
P = สัดส่วนของประชากรที่ผู้วิจัยจะสุ่ม

Z = ระดับความมั่นใจที่ผู้วิจัยกำหนดไว้

Z มีค่าเท่ากับ 1.96 ที่ระดับความเชื่อมั่น 95% (ระดับ 0.05)

Z มีค่าเท่ากับ 2.58 ที่ระดับความเชื่อมั่น 99% (ระดับ 0.01)

e = สัดส่วนของความคลาดเคลื่อนที่ยอมให้เกิดขึ้นได้



ตัวอย่าง ในการวิจัยเรื่องหนึ่ง ผู้วิจัยกำหนดให้สัดส่วนของประชากรเท่ากับ 0.45 ต้องการระดับความเชื่อมั่น 95% และยอมให้มีความคลาดเคลื่อนได้ 5% จะมีขนาดของกลุ่มตัวอย่างเป็นจำนวนเท่าใด

$$\text{สูตร} \quad n = \frac{P(1-P)Z^2}{e^2}$$

จากโจทย์ กำหนดค่า P = 0.45 , ระดับความเชื่อมั่น 95% เป็นค่า Z = 1.96 ความคลาดเคลื่อน 5% เป็นค่า e = 0.05

$$n = \frac{0.45(1-0.45)(1.96)^2}{(0.05)^2}$$

$$= \frac{0.45(0.55)(1.96)^2}{(0.05)^2} = \frac{0.9504}{0.0025} = 380.16$$

ดังนั้น จะใช้กลุ่มตัวอย่างจำนวน 380 คน

2.3 กรณีทราบจำนวนประชากร และมีจำนวนไม่มากใช้สูตรดังนี้

สูตร

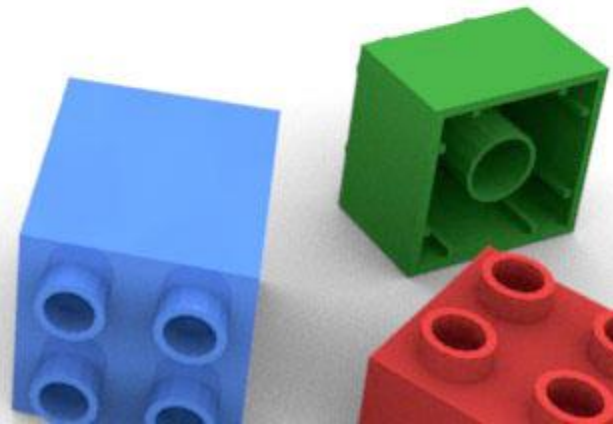
$$n = \frac{P(1-P)}{\frac{e^2}{Z^2} + \frac{P(1-P)}{N}}$$

ตัวอย่าง ในการวิจัยเรื่องหนึ่งมีประชากร 600 คน ถ้ากำหนดสัดส่วนของประชากรเท่ากับ 0.20 ต้องการความเชื่อมั่น 99% โดยยอมให้มีความคลาดเคลื่อนได้ 1% จะต้องใช้กลุ่มตัวอย่างจำนวนเท่าใด

จากโจทย์ กำหนดค่า N = 600 , P = 0.20

ระดับความเชื่อมั่น 99 % เป็นค่า Z = 2.58

ความคลาดเคลื่อน 1% เป็นค่า e = 0.01



สูตร

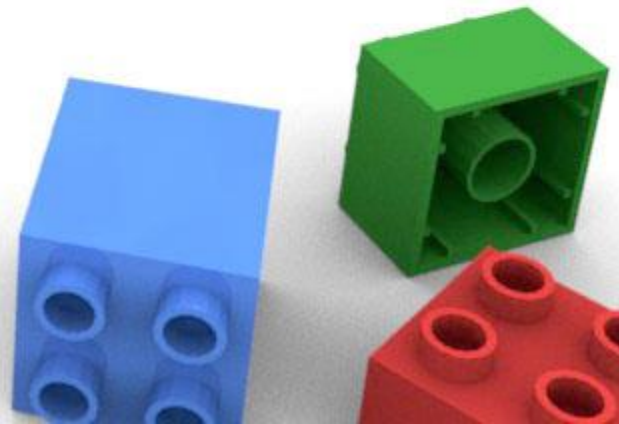
$$n = \frac{(0.20)(1-0.20)}{\frac{(0.01)^2}{(2.58)^2} + \frac{(0.20)(1-0.20)}{600}}$$

$$= \frac{(0.20)(0.80)}{0.000015 + 0.00026}$$

$$= \frac{0.16}{0.000275}$$

$$= 581.81$$

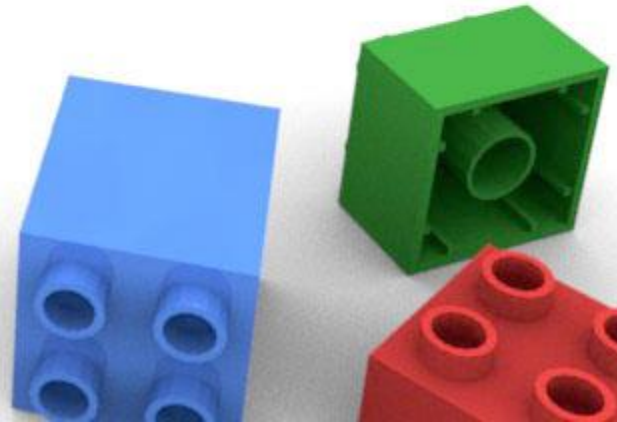
ดังนั้น จะใช้กลุ่มตัวอย่างจำนวน 582 คน



3. การคำนวณขนาดของกลุ่มตัวอย่างโดยใช้ตาราง

การคำนวณขนาดของกลุ่มตัวอย่างโดยใช้ตารางของ เคอร์จี่ และมอร์แกน Krejcie และ Morgan

R.V. Krejcie และ D.W. Morgan ได้จัดทำตารางระบุจำนวนของกลุ่มตัวอย่างที่จะสุ่ม เมื่อทราบจำนวนประชากร ตั้งแต่ประชากร 10 คนไปจนถึง 1 แสนคน เพื่อความสะดวกแก่ผู้วิจัยในการคำนวณขนาดตัวอย่าง (สามารถ Search หาได้จากอินเทอร์เน็ต)



การสุ่มตัวอย่าง (Sampling) หมายถึง กระบวนการได้มาซึ่งกลุ่มตัวอย่างที่มีความเป็นตัวแทนที่ดีของประชากร

เทคนิคการสุ่มกลุ่มตัวอย่าง

เทคนิคการสุ่มตัวอย่างแบ่งเป็น 2 ประเภทใหญ่ ๆ คือ

1. การสุ่มตัวอย่างโดยไม่อาศัยความน่าจะเป็น (Nonprobability sampling)

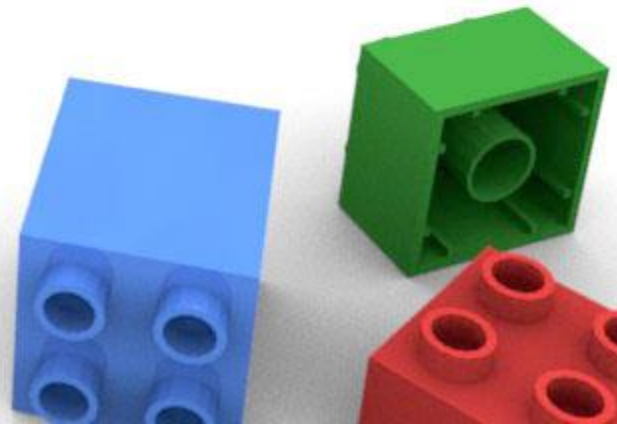
เป็นการเลือกกลุ่มตัวอย่างโดยไม่ใช้วิธีสุ่ม ดังนั้นประชากรจึงมีโอกาสที่จะได้รับเลือกเป็นกลุ่มตัวอย่างไม่เท่ากัน จึงไม่สามารถทราบค่าความน่าจะเป็นของสมาชิกที่ถูกเลือกเป็นตัวอย่าง การสุ่มไม่มีกฎเกณฑ์แน่นอน และไม่สามารถคำนวณความคลาดเคลื่อนที่เกิดขึ้น การสุ่มแบบนี้สามารถสุ่มได้ 4 วิธี คือ

1.1 การสุ่มแบบบังเอิญ (Accidental sampling) เป็นการเลือกที่ไม่มีกฎเกณฑ์จะเลือกใครก็ได้ที่สามารถให้ข้อมูลที่ผู้วิจัยต้องการจนครบจำนวน กลุ่มตัวอย่างจึงจะไม่เป็นตัวแทนที่ดีของประชากร ทำให้ไม่สามารถอ้างอิงถึงประชากรได้อย่างเที่ยงตรง เช่น เดินแจกแบบสอบถามตามห้างสรรพสินค้า โรงอาหาร สวนสาธารณะ สถานีรถไฟ ป้ายรถเมล์ เป็นต้น

1.2 การสุ่มโดยกำหนดสัดส่วน (Quota sampling) ใช้ในกรณีกลุ่มตัวอย่างมีหลายประเภท และมีลักษณะที่แตกต่างกัน จึงต้องกำหนดจำนวนหรือโควตาของสมาชิกแต่ละประเภทว่าจะให้สัดส่วนเท่าใดจึงจะได้จำนวนที่ต้องการ เช่น จะใช้นักศึกษาชั้นปีที่ 1 กี่คน นักศึกษาชั้นปีที่ 2 กี่คน เป็นต้น

1.3 การสุ่มแบบเจาะจง (Purposive sampling) ผู้วิจัยจะต้องใช้ดุลยพินิจพิจารณาว่าสมาชิกในกลุ่มใดน่าจะเป็นตัวแทนที่ดีที่เลือกเอาสมาชิกกลุ่มนั้น สิ่งที่เป็นจุดอ่อนคือผู้วิจัยแต่ละคนอาจจะมีดุลยพินิจที่แตกต่างกัน ซึ่งอาจจะทำให้กลุ่มตัวอย่างอาจไม่ใช่ตัวแทนที่ดีก็ได้ เช่น ผู้วิจัยเลือกเฉพาะผู้ใช้โทรศัพท์มือถือยี่ห้อโนเกียเท่านั้น

1.4 การสุ่มตามสะดวก (Convenience sampling) เป็นการเลือกลุ่มตัวอย่างโดยถือเอาความสะดวกของผู้วิจัยเป็นหลัก และง่ายต่อการรวบรวมข้อมูล เช่น ผู้วิจัยเลือกกลุ่มตัวอย่างเฉพาะห้องที่รับผิดชอบสอนเพราะง่ายและสะดวกได้ข้อมูลครบถ้วน



2. การสุ่มตัวอย่างโดยอาศัยความน่าจะเป็น (Probability sampling)

เป็นการเลือกกลุ่มตัวอย่างโดยที่สมาชิกมีโอกาสถูกเลือกเท่า ๆ กัน โดยอาศัยเทคนิคการสุ่มกลุ่มตัวอย่าง และปราศจากความลำเอียง (bias) วิธีการสุ่มแบบนี้สามารถใช้วิธีการทางสถิติอ้างอิง และคำนวณค่าความแตกต่างระหว่างกลุ่มตัวอย่างสามารถอ้างอิงถึงประชากรได้ การสุ่มแบบนี้สามารถสุ่มได้ 4 วิธี คือ

2.1 การสุ่มอย่างง่าย (Simple random sampling) เป็นการสุ่มตัวอย่างโดยถือว่าทุก ๆ หน่วยหรือทุก ๆ สมาชิกในประชากรมีโอกาสจะถูกเลือกเท่า ๆ กัน การสุ่มวิธีนี้จะต้องมีรายชื่อประชากรทั้งหมดและมีการให้เลขกำกับ วิธีการอาจใช้วิธีการจับสลากโดยทำรายชื่อประชากรทั้งหมด หรือใช้ตารางเลขสุ่มโดยมีเลขกำกับหน่วยรายชื่อทั้งหมดของประชากร

2.2 การสุ่มแบบเป็นระบบ (Systematic sampling) การสุ่มแบบนี้สำหรับประชากรที่มีการเรียงลำดับของสมาชิกไว้แล้วโดยปราศจากความลำเอียง เช่น การจัดเรียงลำดับรายชื่อจาก ก – ฮ หรือมีการจัดเรียงตามบ้านเลขที่ แล้วนำมาหาช่วงของสมาชิก โดยนำเอาจำนวนประชากร (N) หารด้วยจำนวนกลุ่มตัวอย่างที่ต้องการ เช่น มีประชากร 500 คนต้องการกลุ่มตัวอย่าง 50 ช่วงคือ 500/50 = 10 ต่อมาทำฉลากขึ้น 10 แผ่น เขียนเลข 1 – 10 บนกระดาษแล้วพับหรือม้วน จับฉลากขึ้นมาได้ 1 แผ่น สมมติได้เลข 2 กลุ่มตัวอย่างก็จะเริ่มเป็น 2, 12, 22, 32, 42, ... เป็นการนำเลขที่จับฉลากได้บวกกับช่วงคือ 2 + 10 = 12 , 12 + 10 = 22 เป็นต้น ถ้านับแล้วได้จำนวนยังไม่ครบก็ต้องจับฉลากใหม่แล้วบวกกับเลขอีกเช่นเดิม ทำไปจนกว่าจะได้ตัวอย่างครบตามจำนวน

2.3 การสุ่มแบบชั้นภูมิ (Stratified sampling) เป็นการสุ่มตัวอย่างโดยแยกประชากรออกเป็นกลุ่มประชากรย่อย ๆ หรือแบ่งเป็นชั้นภูมิก่อน โดยหน่วยประชากรในแต่ละชั้นภูมิจะมีลักษณะเหมือนกัน (homogenious) แล้วสุ่มอย่างง่ายเพื่อให้ได้จำนวนกลุ่มตัวอย่างตามสัดส่วนของขนาดกลุ่มตัวอย่างและกลุ่มประชากร

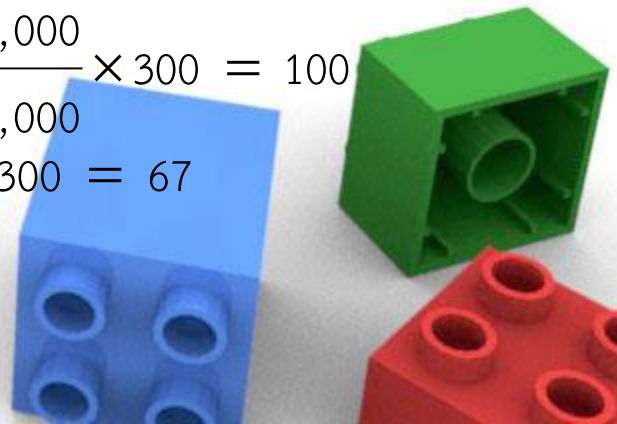
ตัวอย่าง การสุ่มจากประชากรซึ่งเป็นโรงเรียน แบ่งเป็น 3 ขนาด คือ โรงเรียนขนาดใหญ่มีนักเรียน 4,000 คน ขนาดกลางมี 3,000 คน ขนาดเล็กมี 2,000 คน ต้องการสุ่มตัวอย่าง 300 คน จะเลือกนักเรียนแต่ละระดับเท่าใด

โรงเรียนขนาดใหญ่นักเรียน 4,000 คน เลือกตัวอย่าง $\frac{4,000}{9,000} \times 300 = 133$

โรงเรียนขนาดกลางนักเรียน 3,000 คน เลือกตัวอย่าง $\frac{3,000}{9,000} \times 300 = 100$

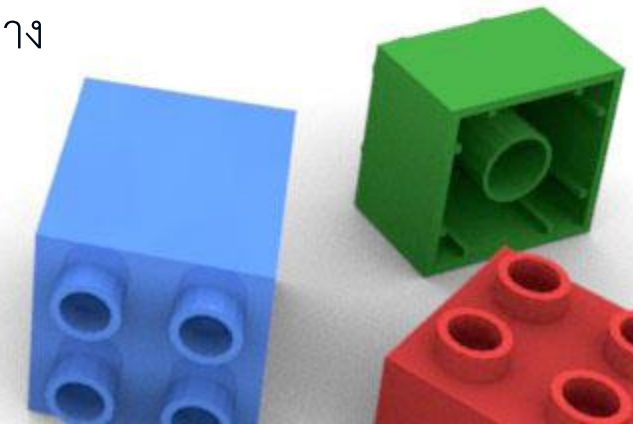
โรงเรียนขนาดเล็กนักเรียน 2,000 คน เลือกตัวอย่าง $\frac{2,000}{9,000} \times 300 = 67$

รวม 9,000 คน เลือกตัวอย่าง 300 คน



2.4 การสุ่มแบบแบ่งกลุ่ม (Cluster sampling) วิธีนี้เหมาะกับประชากรที่มีขนาดใหญ่ อาณาเขตกว้าง หากสุ่มด้วยวิธีสุ่มอย่างง่าย จะได้กลุ่มตัวอย่างกระจัดกระจายกัน ไม่สะดวกในการเก็บข้อมูล เสียแรงงาน เวลา และค่าใช้จ่ายมาก ตัวประชากรมีลักษณะเป็นกลุ่มที่แต่ละกลุ่มมีความคล้ายคลึงกัน การสุ่มมาเพียงบางกลุ่ม แล้วสุ่มสมาชิกภายในกลุ่มอีกครั้ง จะช่วยลดปัญหาในการเก็บข้อมูลได้ และกลุ่มตัวอย่างที่สุ่มมาก็เป็นตัวอย่างที่ดีได้เช่นกัน เช่น

ต้องการศึกษาทัศนคติของคนในชนบทอาจแบ่งคนกลุ่มชนบทออกเป็นกลุ่ม ๆ โดยอาศัยแผนที่ภายในแต่ละกลุ่มมีความคล้ายคลึงกันจำนวนมาก แล้วสุ่มมาจำนวนหนึ่ง แล้วสุ่มประชากรในหมู่บ้านที่ถูกสุ่มไว้แล้วนั้นมาเป็นกลุ่มตัวอย่าง



คำถามท้ายบท

1. จงยกตัวอย่างประชากรที่สามารถนับจำนวนได้ และประชากรที่ไม่สามารถนับจำนวนได้อย่างละ 5 ข้อ

2. การสุ่มตัวอย่างให้ประโยชน์ต่อการวิจัยอย่างไรบ้าง จงอธิบาย

3. ต้องการสุ่มตัวอย่างจากนักธุรกิจค้าปลีก จำนวน 6,000 คน โดยยอมให้เกิดความคลาดเคลื่อนไม่เกิน 5% จะได้กลุ่มตัวอย่างมีขนาดเท่าใด

4. ถ้าต้องการสุ่มตัวอย่างจำนวน 0.40 จากประชากรทั้งหมด ที่ระดับความเชื่อมั่น 99% ยอมให้เกิดความคลาดเคลื่อนไม่เกิน 1% กลุ่มตัวอย่างจะมีขนาดเท่าใด

5. ประชากรจำนวน 800, 1300, 1800, 10000, และ 25,000 คน ถ้าใช้ตารางของ Krejcie และ Morgan จะได้กลุ่มขนาดเท่าใด

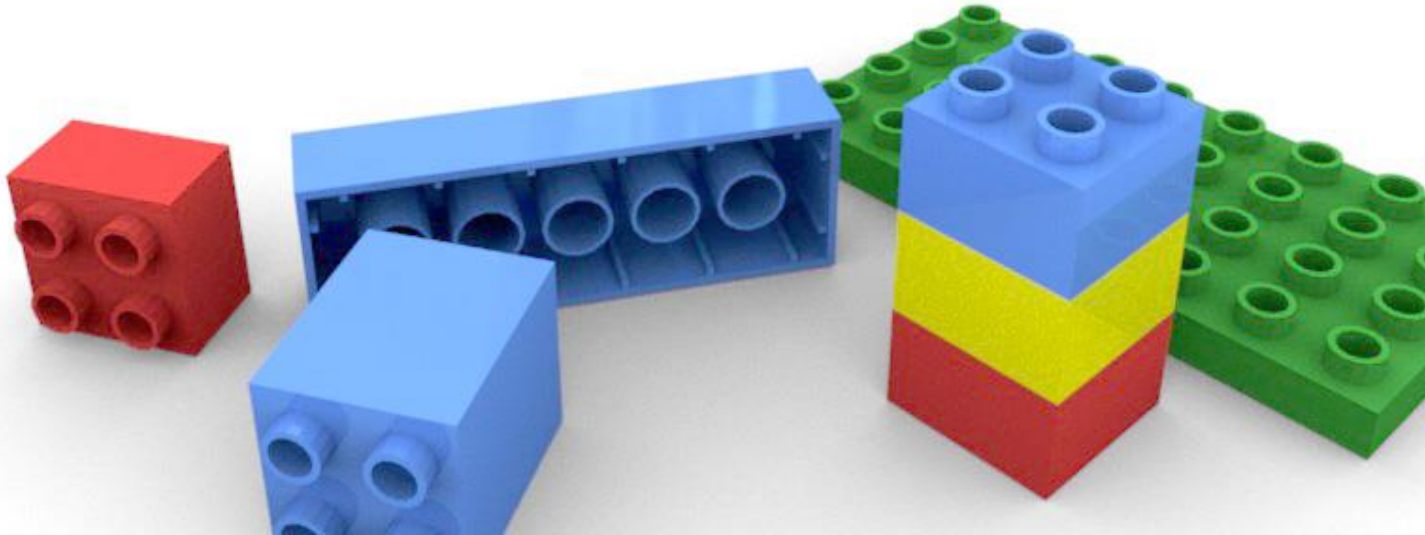
6. การสุ่มตัวอย่างโดยอาศัยความน่าจะเป็นมีข้อดีอย่างไรบ้าง จงอธิบาย

7. การสุ่มแบบแบ่งกลุ่มนิยมใช้ในโอกาสใด จงอธิบายพอสังเขป

บทที่ 6

การประมาณค่าพารามิเตอร์

(Estimation)



เนื้อหาในบทนี้ประกอบด้วยหัวข้อดังนี้

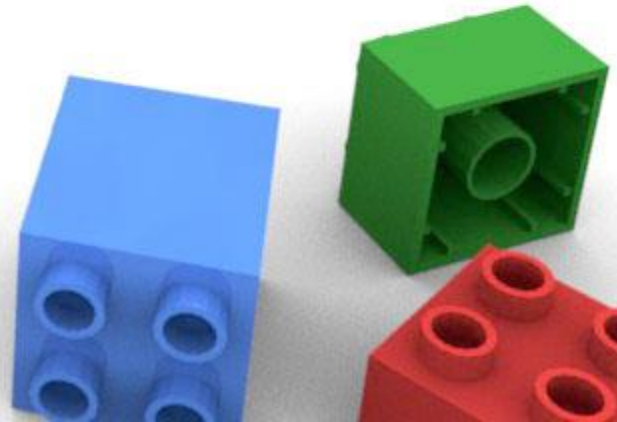
6.1 การประมาณค่าแบบเดี่ยวหรือแบบจุด (Point estimation)

6.2 การประมาณค่าแบบช่วง (Interval estimation)

6.3 ช่วงความเชื่อมั่นของค่าเฉลี่ยของประชากร

6.4 ช่วงความเชื่อมั่นของค่าสัดส่วนของประชากร

6.5 ช่วงความเชื่อมั่นของความแปรปรวนของประชากร



การประมาณค่าพารามิเตอร์ (Estimation)

เป็นการประมาณค่าพารามิเตอร์หรือลักษณะของประชากรโดยใช้ข้อมูลจากตัวอย่าง หรือการประมาณค่าพารามิเตอร์ของประชากรโดยใช้สถิติที่เหมาะสม ซึ่งมีวิธีการอยู่ 2 แบบ คือ

6.1 การประมาณค่าแบบเดี่ยวหรือแบบจุด (Point Estimation)

เป็นการประมาณค่าพารามิเตอร์ของประชากรโดยใช้ข้อมูลจากตัวอย่างซึ่งได้ค่าประมาณเป็นค่าเดียวเท่านั้น เช่น

ค่าเฉลี่ยตัวอย่าง $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ในการประมาณค่าเฉลี่ยประชากร μ

ค่าสัดส่วนตัวอย่าง $\hat{p} = X/n$ ในการประมาณสัดส่วนประชากร p

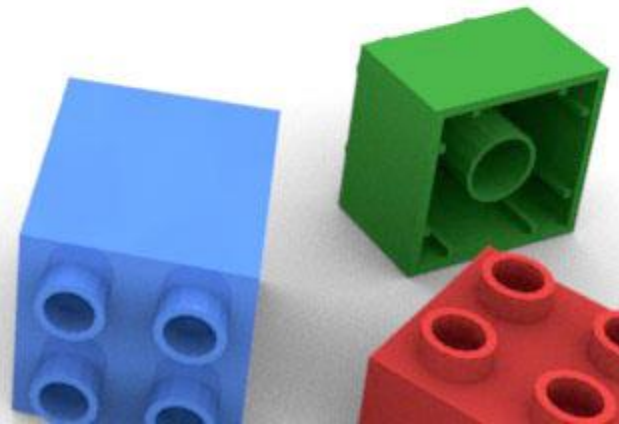
ค่าความแปรปรวนตัวอย่าง $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ ในการประมาณความแปรปรวนประชากร σ^2

6.2 การประมาณค่าแบบช่วง (Interval Estimation)

เป็นการหาตัวแปรสุ่ม L และ U ที่ครอบคลุมค่าพารามิเตอร์ θ ด้วยความน่าจะเป็นระดับหนึ่ง นั่นคือ จะหาตัวแปรสุ่ม L (ลิมิตล่าง) และ U (ลิมิตบน) ที่ทำให้

$$P\left(L<\theta<U\right)=1-\alpha$$

โดยเรียกช่วงระหว่าง L และ U ว่า ช่วงความเชื่อมั่น $(1-\alpha)100\%$ และเรียก $1-\alpha$ ว่าระดับความเชื่อมั่น หรือความน่าจะเป็นที่ช่วงจะครอบคลุมค่าของ θ ซึ่งนิยมใช้ 0.95 หรือ 0.99

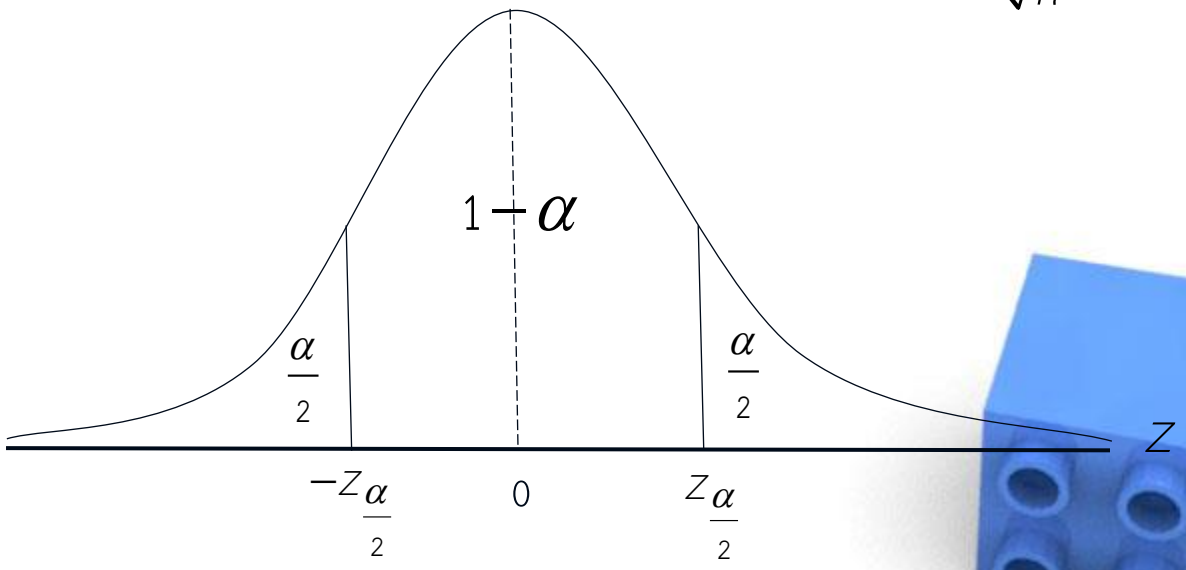


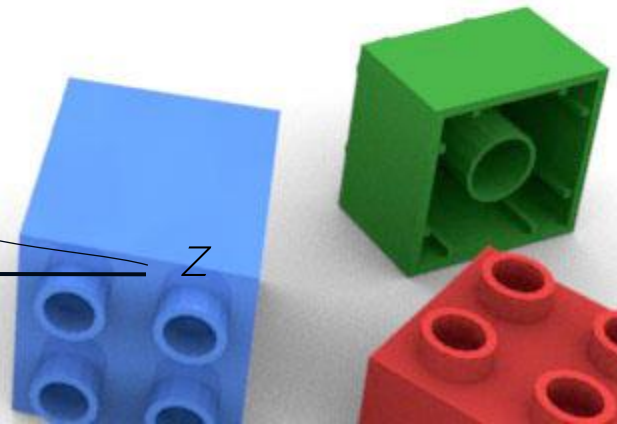
6.3 ช่วงความเชื่อมั่นของค่าเฉลี่ยของประชากร

โดยที่ $\bar{X}$ เป็นตัวประมาณแบบเดี่ยวของ μ โดยวิธีการสร้างช่วงความเชื่อมั่นของ μ มี 2 กรณี คือ

กรณีที่ 1 เมื่อสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงปกติที่มีค่าเฉลี่ย μ ซึ่งไม่ทราบค่า แต่**ทราบความแปรปรวน** σ^2

จาก $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ จะได้ว่า $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$





จากรูปจะได้ว่า $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1-\alpha$

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

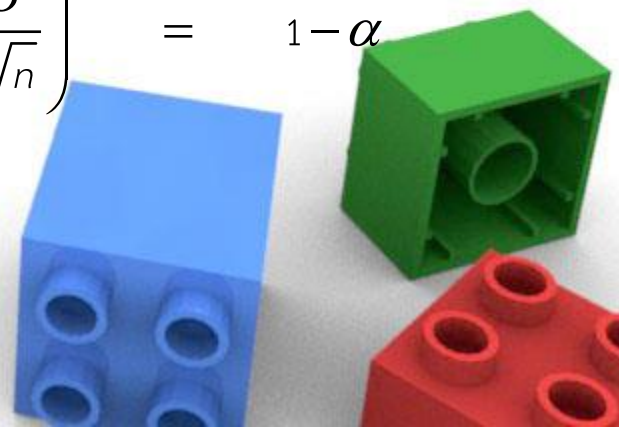
$$P\left(-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

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

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

นั่นคือ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ μ คือ

$$\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$



ตัวอย่าง 6.1 น้ำหนักสุทธิของน้ำตาลที่บรรจุถุงโดยเครื่องจักรเครื่องหนึ่งมีการแจกแจงแบบปกติ ที่มีค่าเบี่ยงเบนมาตรฐาน 1.2 ออนซ์ สุ่มน้ำตาลมา 25 ถุง พบว่ามีน้ำหนักสุทธิเฉลี่ย 19.8 ออนซ์ จงหาความเชื่อมั่น 95% ของน้ำหนักสุทธิเฉลี่ยของน้ำตาลที่บรรจุถุงโดยเครื่องจักรเครื่องนี้ทั้งหมด

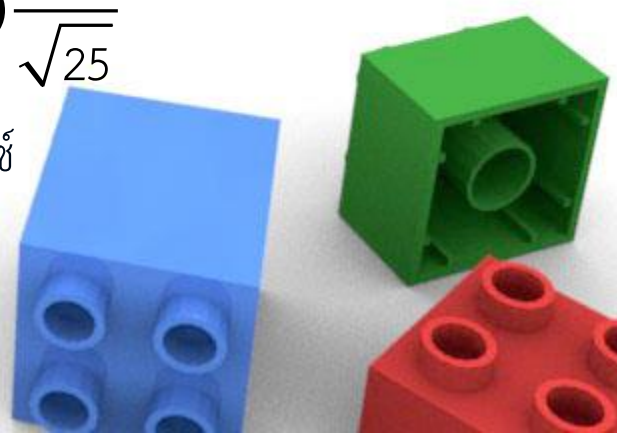
วิธีทำ ให้ μ น้ำหนักสุทธิเฉลี่ยของน้ำตาลที่บรรจุถุงโดยเครื่องจักรเครื่องนี้ทั้งหมด หน่วย :ออนซ์

ช่วงความเชื่อมั่น 95% ของ μ คือ

$$\bar{X} - z_{0.025}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{0.025}\frac{\sigma}{\sqrt{n}}$$

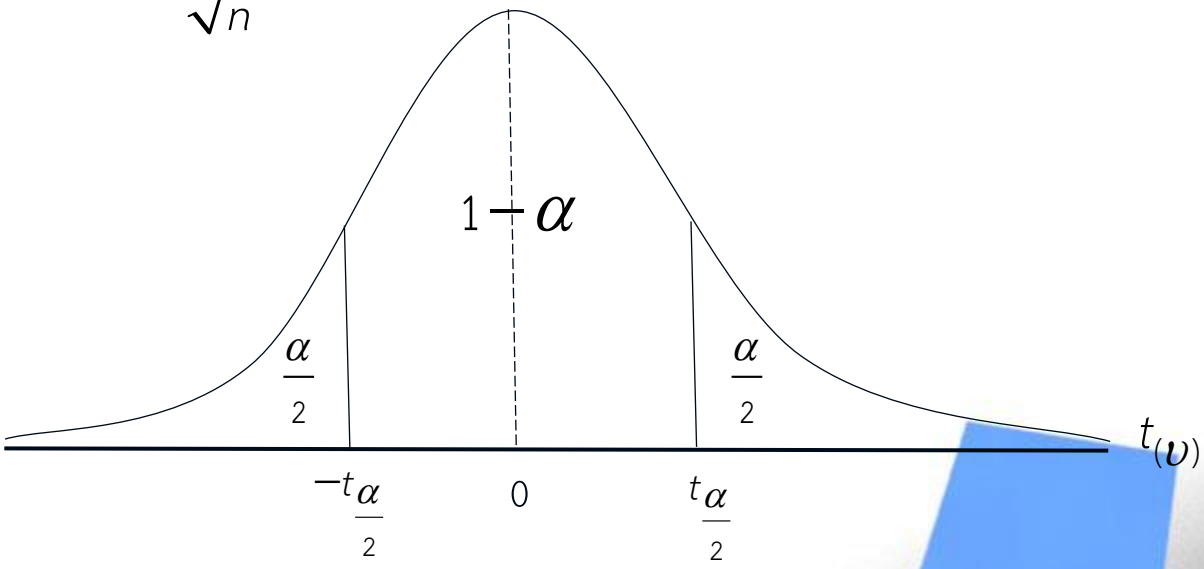
$$19.8 - (1.96)\frac{1.2}{\sqrt{25}} < \mu < 19.8 + (1.96)\frac{1.2}{\sqrt{25}}$$

$$19.33 < \mu < 20.27 \quad \text{ออนซ์}$$



กรณีที่ 2 เมื่อสุ่มตัวอย่างขนาด n จากประชากรที่มีการแจกแจงปกติที่มีค่าเฉลี่ย μ และความแปรปรวน σ^2 ซึ่งไม่ทราบค่า

แล้ว $T = \dfrac{\overline{X} - \mu}{\dfrac{S}{\sqrt{n}}}$ มีการแจกแจงแบบที ที่มีองศาแห่งความเป็นอิสระ $\upsilon = n - 1$



จากรูปจะได้ว่า $P\left(-t_{\frac{\alpha}{2}} < T_{(\upsilon)} < t_{\frac{\alpha}{2}}\right) = 1-\alpha$

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

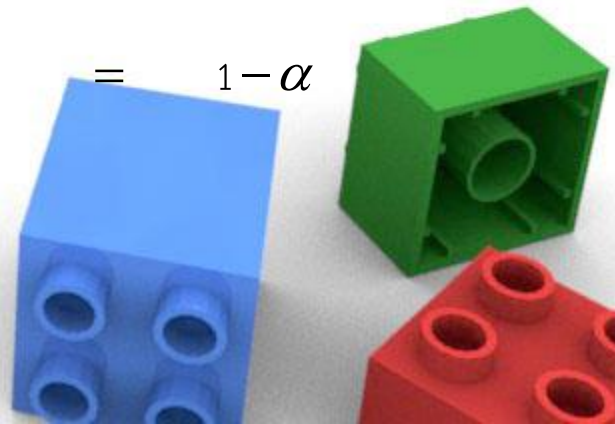
$$P\left(-t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

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

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

นั่นคือ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ μ คือ

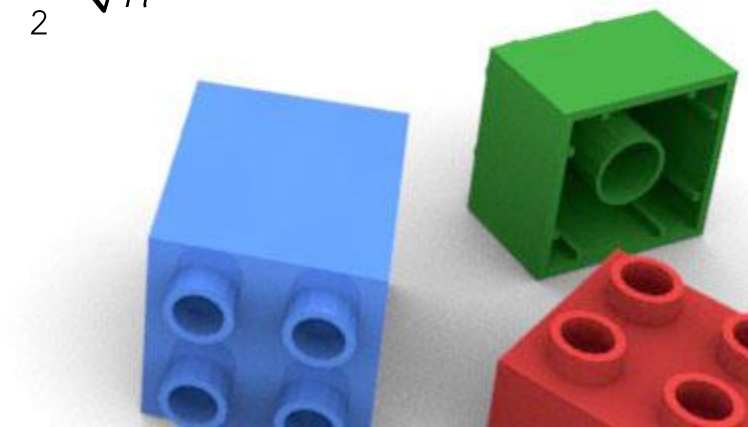
$$\bar{X}-t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{X}+t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}$$



เนื่องจาก **ตัวอย่างที่สุ่มมามีขนาดใหญ่พอ** คือ ถ้ามีค่ามากกว่าหรือเท่ากับ 30 ตัวแปรสุ่ม T จะมีการแจกแจงใกล้เคียงแบบปกติมาตรฐาน จะได้ว่า

นั่นคือ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ μ คือ

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$



สรุปการหาช่วงความเชื่อมั่นของค่าเฉลี่ยของประชากร

ประชากรมีการแจกแจงแบบปกติ

ทราบ σ^2

ใช้ Z, σ

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

ไม่ทราบ σ^2

ตัวอย่างสุ่ม ≥ 30

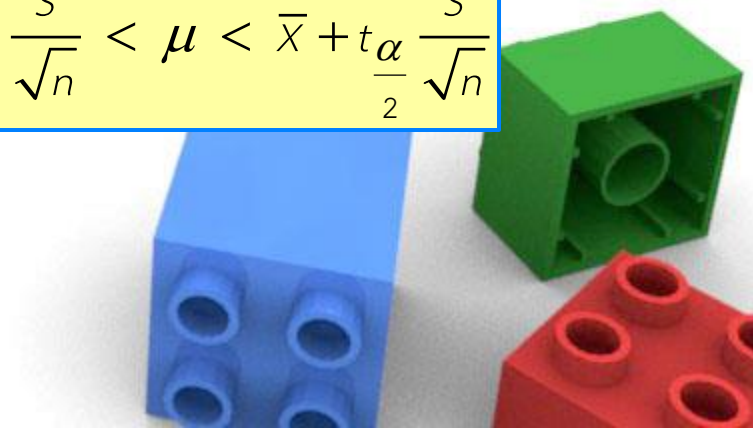
ใช้ Z, S

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

ตัวอย่างสุ่ม < 30

ใช้ t, S

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$



ตัวอย่าง 6.2 ในการศึกษาปริมาณในการใช้น้ำมันในรถบรรทุกรุ่นใหม่ยี่ห้อหนึ่งรวม 10 คัน โดยบันทึกระยะทางที่รถแต่ละคันแล่นได้ในหน่วย : ไมล์/แกลลอน ได้ดังนี้

18.6 18.4 19.2 20.8 19.4 20.5 18.7 19.3 19.6 20.2

จงหาความเชื่อมั่น 99% ของระยะทางโดยเฉลี่ยที่รถบรรทุกรุ่นใหม่ยี่ห้อนี้แล่นได้

วิธีทำ ให้ X ระยะทางที่รถแต่ละคันแล่นได้ในหน่วย : ไมล์/แกลลอน

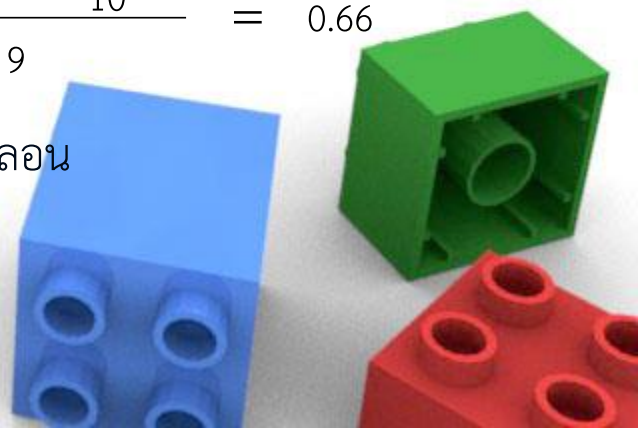
$$\bar{x} = \frac{\sum x}{n} = \frac{194.7}{10} = 19.47$$

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{3,796.79 - \frac{(194.7)^2}{10}}{9} = 0.66$$

จะได้ $s = \sqrt{0.66} = 0.81$ ไมล์/แกลลอน

โดย $\upsilon = n-1 = 10-1 = 9$

ดังนั้น $t_{\frac{\alpha}{2}} = t_{\frac{0.01}{2}} = t_{0.005} = 3.25$



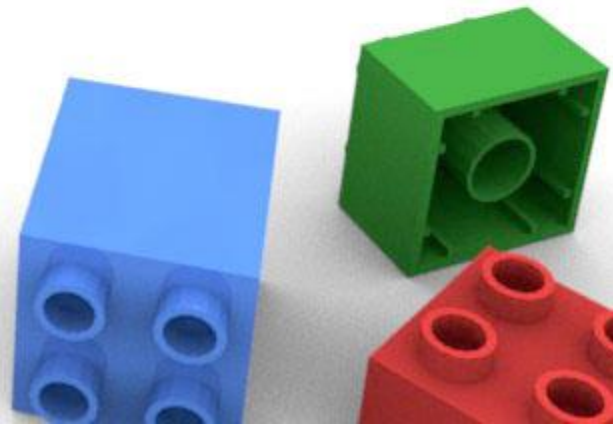
ให้ μ เป็นระยะทางโดยเฉลี่ยที่รถบรรทุกรุ่นใหม่ยี่ห้อนี้แล่นได้ หน่วย : ไมล์/แกลลอน

ช่วงความเชื่อมั่น 99% ของ μ คือ

$$\bar{X} - t_{0.005}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{0.005}\frac{S}{\sqrt{n}}$$

$$19.47 - (3.25)\frac{0.81}{\sqrt{10}} < \mu < 19.47 + (3.25)\frac{0.81}{\sqrt{10}}$$

$$18.64 < \mu < 20.30 \quad \text{ไมล์/แกลลอน}$$



6.4 ช่วงความเชื่อมั่นของสัดส่วนของประชากร

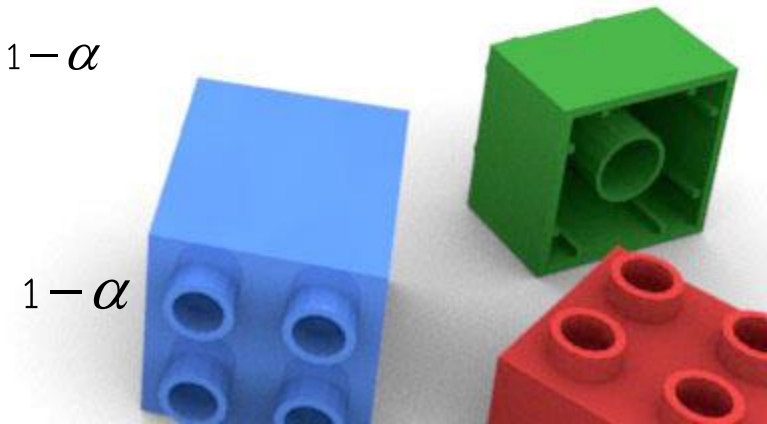
ถ้าประชากรที่สนใจมีความน่าจะเป็นหรือสัดส่วนของลักษณะที่สนใจ p ซึ่งไม่ทราบค่า และ $\hat{p}$ เป็นสัดส่วนของตัวอย่างขนาด n โดยที่ $\hat{p}$ มีการแจกแจงใกล้เคียงแบบปกติที่มีค่าเฉลี่ย $\mu_{\hat{p}} = p$ และความแปรปรวน $\sigma^2_{\hat{p}} = \frac{pq}{n}$ เมื่อตัวอย่าง n มีขนาดใหญ่พอ หรือมีค่ามากกว่าหรือเท่ากับ 30 และได้ว่า

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

จะได้ว่า $$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

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

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

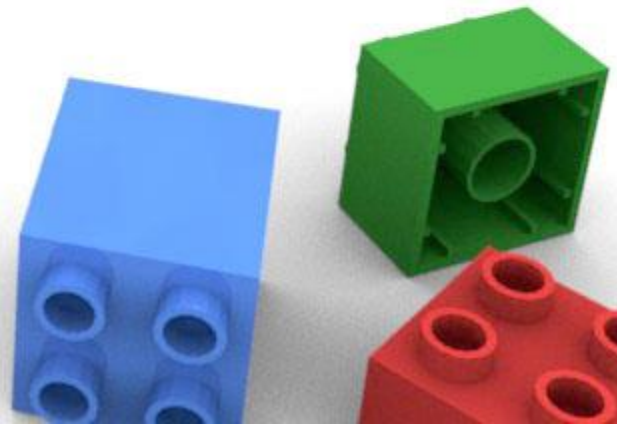


แต่เนื่องจากไม่ทราบค่า p ดังนั้น จะใช้ $S_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$ เป็นตัวประมาณค่าแบบเดี่ยวของ

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ p คือ

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$



ตัวอย่าง 6.3 จากการสุ่มแม่บ้านในจังหวัดนครราชสีมาจำนวน 400 คน แล้วสอบถามผงซักฟอกยี่ห้อที่ใช้เป็นประจำ ปรากฏว่ามีแม่บ้าน 224 คน ที่ใช้ผงซักฟอกยี่ห้อซื่อสัตย์ จงหาช่วงความเชื่อมั่น 90% ของสัดส่วนของแม่บ้านทั้งหมดในจังหวัดนครราชสีมาที่ใช้ผงซักฟอกตราซื่อสัตย์เป็นประจำ

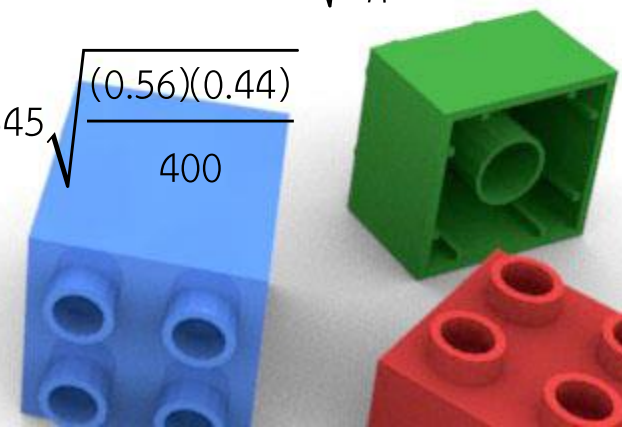
วิธีทำ ให้ p แทนสัดส่วนของแม่บ้านทั้งหมดในจังหวัดนครราชสีมาที่ใช้ผงซักฟอกตราซื่อสัตย์เป็นประจำ

$$\hat{p} = \frac{224}{400} = 0.56 \text{ ดังนั้น } \hat{q} = 1-\hat{p} = 1-0.56 = 0.44$$

ช่วงความเชื่อมั่น 90% ของ p คือ $\hat{p}-z_{0.05}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p}+z_{0.05}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$0.56-1.645\sqrt{\frac{(0.56)(0.44)}{400}} < p < 0.56+1.645\sqrt{\frac{(0.56)(0.44)}{400}}$$

$$0.52 < p < 0.6$$

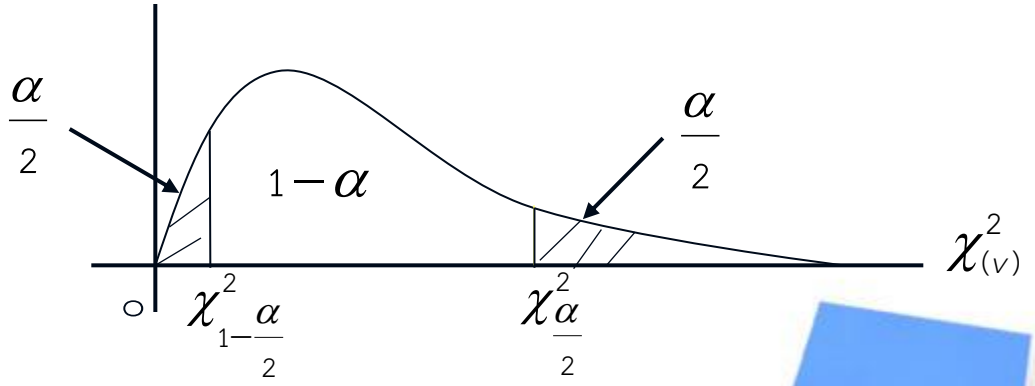


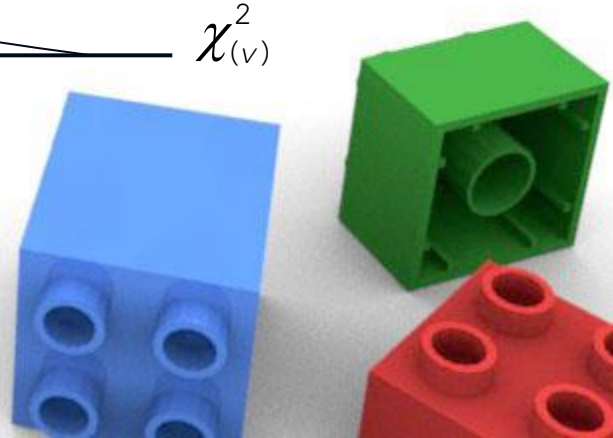
6.5 ช่วงความเชื่อมั่นของความแปรปรวนของประชากร

ถ้า S^2 เป็นความแปรปรวนของตัวอย่างขนาด n ที่สุ่มจากประชากรที่มีการแจกแจงแบบปกติ ที่มีความแปรปรวน σ^2 ซึ่งไม่ทราบค่า แล้วตัวแปรสุ่ม

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

มีการแจกแจงแบบไคสแควร์กำลังสอง ที่มีองศาแห่งความอิสระ $\upsilon = n-1$





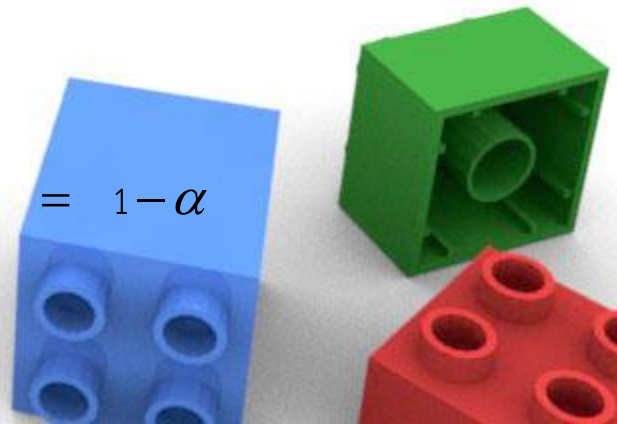
จากรูปจะได้ว่า $$P\left(\chi^2_{1-\frac{\alpha}{2}} < \chi^2_{(\upsilon)} < \chi^2_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\chi^2_{1-\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\frac{\chi^2_{1-\frac{\alpha}{2}}}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{\frac{\alpha}{2}}}{(n-1)S^2}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}} > \sigma^2 > \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}\right) = 1-\alpha$$

หรือ $$P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

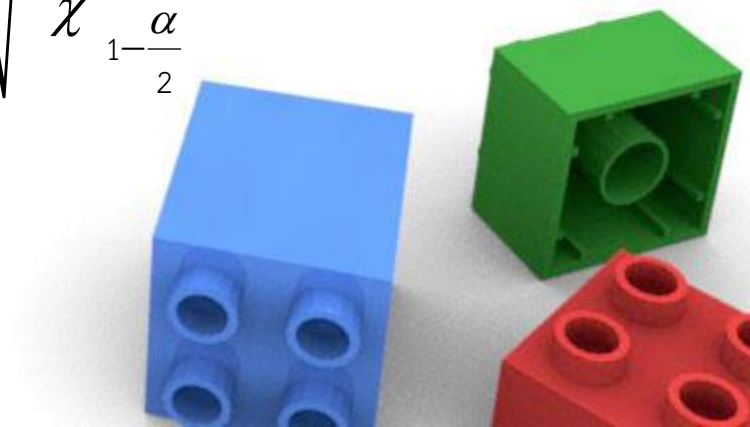


นั่นคือ ช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ σ^2 คือ

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}$$

และได้ว่าช่วงความเชื่อมั่น $(1-\alpha)100\%$ ของ σ คือ

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}}$$



ตัวอย่าง 6.4 โรงงานเคลือบโลหะบนพลาสติกแห่งหนึ่งได้สุ่มพลาสติกที่ผ่านการเคลือบแล้วจากการทำงานในสัปดาห์หนึ่งมาจำนวน 9 แผ่น วัดความหนาของโลหะที่เคลือบได้ ดังนี้ หน่วย : มิลลิเมตร

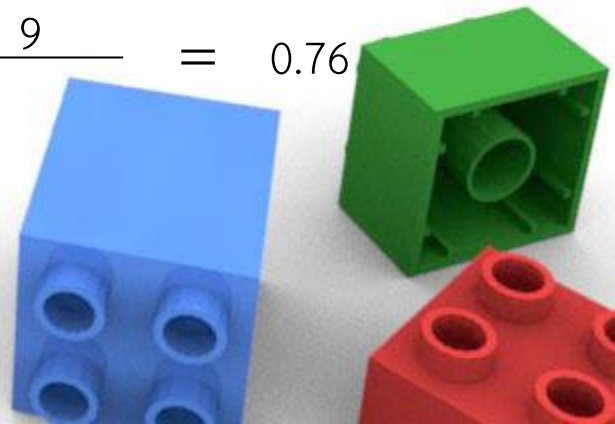
20.5 21.2 18.6 20.4 21.6 19.8 19.9 20.3 20.8

จงหาช่วงความเชื่อมั่น 90% ของค่าเบี่ยงเบนมาตรฐานของความหนาของโลหะที่เคลือบได้ทั้งหมดจากการทำงานในสัปดาห์ดังกล่าว

วิธีทำ ให้ σ เป็นค่าเบี่ยงเบนมาตรฐานของความหนาของโลหะที่เคลือบได้ทั้งหมดจากการทำงานในสัปดาห์ดังกล่าว หน่วย : มิลลิเมตร

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{3,731.15 - \frac{(183.1)^2}{9}}{9-1} = 0.76$$

ที่ $\upsilon = 8$, $\chi^2_{0.95} = 2.73$, $\chi^2_{0.05} = 15.5$

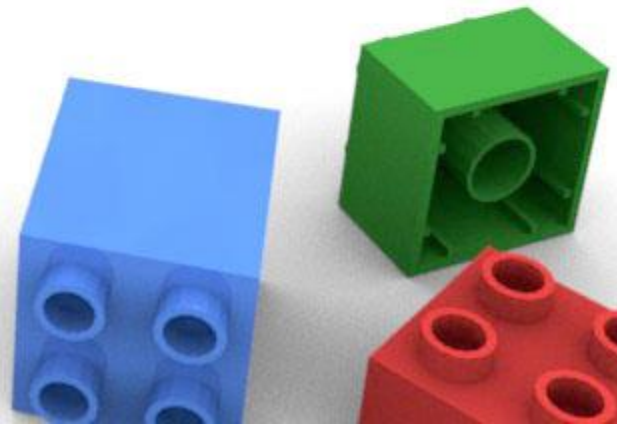


นั่นคือ ช่วงความเชื่อมั่น 90% ของ σ คือ

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{0.05}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{0.95}}}$$

$$\sqrt{\frac{(9-1)(0.76)}{15.5}} < \sigma < \sqrt{\frac{(9-1)(0.76)}{2.73}}$$

$0.63 < \sigma < 1.49$ มิลลิเมตร



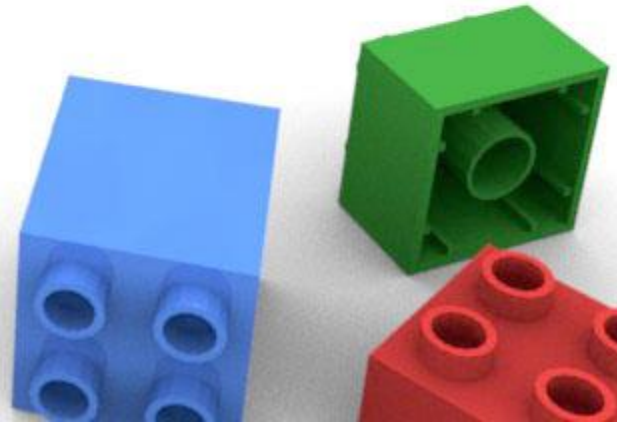
แบบฝึกหัดบทที่ 6

1. จากการสุ่มเก็บน้ำฝนที่ตกในเขตกรุงเทพมหานคร 13 จุด วัดสภาพความเป็นกรด (pH) ได้ดังนี้

3.5	5.1	5.0	3.6	4.8	3.6	4.7
4.3	4.2	4.5	4.9	4.7	4.8	

จงหาช่วงความเชื่อมั่น 95% ของสภาพความเป็นกรดโดยเฉลี่ยของน้ำฝนทั้งหมดที่ตกในเขตกรุงเทพฯ

2. จากการศึกษาระยะพักรักษาตัวของคนไข้ที่แผนกหนึ่งในโรงพยาบาลแห่งหนึ่ง ที่สุ่มมา 64 ราย พบว่ามีระยะเวลาที่พักรักษาตัวในแผนกนี้เฉลี่ย 8.25 วัน และมีค่าเบี่ยงเบนมาตรฐาน 3 วัน จงสร้างช่วงความเชื่อมั่นสำหรับประมาณระยะเวลาโดยเฉลี่ยของการพักรักษาตัวของคนไข้ในแผนกนี้ ที่ระดับความเชื่อมั่น 99%



แบบฝึกหัดบทที่ 6

3. ในการสำรวจความนิยมของประชากร 250 คน ที่มีต่อหนังสือพิมพ์รายวันภาษาไทย 2 ฉบับ ปรากฏ ว่ามีประชากร 150 คน ชอบอ่านหนังสือพิมพ์ข่าวไท และ 100 คน ชอบอ่านหนังสือพิมพ์ประชาชาติ จงหาความเชื่อมั่น 90% ของสัดส่วนความนิยมของประชาชนที่มีต่อ

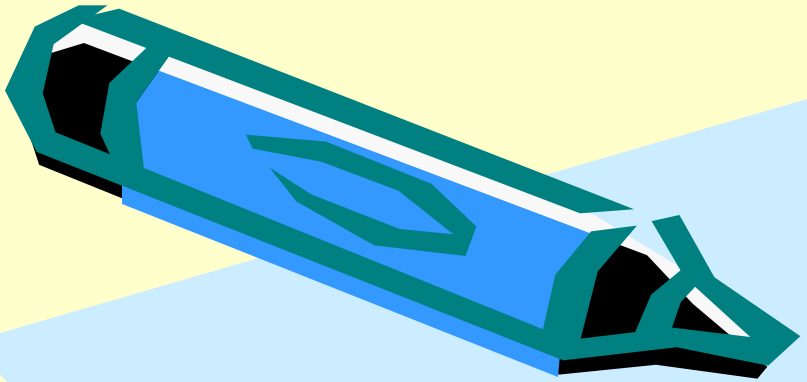
3.1 หนังสือพิมพ์ข่าวไท

3.2 หนังสือพิมพ์ประชาชาติ

4. เพื่อศึกษาเกี่ยวกับมลภาวะของอากาศในกรุงเทพฯ ผู้ศึกษาได้เก็บข้อมูลปริมาณสารเบนซินที่กระจายในอากาศ (หน่วย : ไมโครกรัม/ลูกบาศก์เมตร) จากบริเวณต่างๆของกรุงเทพฯ จำนวน 10 ตัวอย่าง ได้ดังนี้

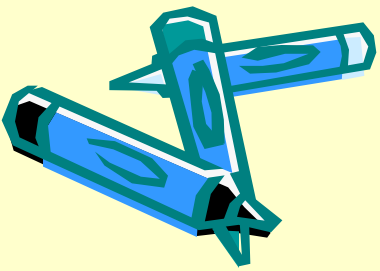
2.5	2.1	2.2	1.3	2.1
3.2	1.9	2.3	2.4	2.0

จงหาช่วงความเชื่อมั่น 95% ของความแปรปรวนที่แท้จริงของปริมาณสารเบนซินที่กระจายในอากาศของกรุงเทพฯ



บทที่ 7

การทดสอบสมมติฐาน



การทดสอบสมมติฐาน

การทดสอบสมมติฐาน ประกอบด้วย การทดสอบสมมติฐานค่าพารามิเตอร์ต่าง ๆ ที่น่าสนใจ และน่าศึกษาเป็นอย่างยิ่ง ดังนี้

7.1 นิยาม

7.2 การทดสอบสมมติฐานแบบทางเดียว และแบบสองทาง

7.3 ขั้นตอนของการทดสอบสมมติฐานพารามิเตอร์

7.4 การทดสอบสมมติฐานค่าเฉลี่ยของหนึ่งประชากร

7.5 การทดสอบสมมติฐานสัดส่วนของหนึ่งประชากร

7.6 การทดสอบสมมติฐานความแปรปรวนของหนึ่งประชากร

ในบทที่ 6 เราทราบถึงวิธีการประมาณค่าพารามิเตอร์หรือค่าประชากรทั้งแบบเดี่ยวและแบบช่วงความเชื่อมั่นแล้ว แต่ถ้าวัตถุประสงค์ของการศึกษาเพื่อต้องการตัดสินใจเกี่ยวกับคุณลักษณะบางอย่างของประชากร เราจะใช้วิธี **“การทดสอบสมมติฐานเชิงสถิติหรือการทดสอบสมมติฐาน”**

7.1 นิยาม

สมมติฐานเชิงสถิติ (Statistical hypothesis) หมายถึง ข้อความเกี่ยวกับประชากรที่ต้องการศึกษา ซึ่งข้อความเกี่ยวกับประชากรนี้ อาจจะเป็นจริงหรือเท็จก็ได้

นิยาม : **สมมติฐาน** เขียนแทนด้วย H มี 2 อย่าง ดังนี้

1. สมมติฐานที่จะทดสอบ เขียนแทนด้วย H_0 เรียกว่า สมมติฐานเพื่อการทดสอบหรือสมมติฐานหลัก (Null hypothesis)
2. สมมติฐานที่แย้งกับสมมติฐานหลัก เขียนแทนด้วย H_1 เรียกว่า สมมติฐานแย้งหรือสมมติฐานรอง (Alternative hypothesis)

การทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ θ

ถ้าให้ θ_0 เป็นค่าของพารามิเตอร์ θ ที่จะพิจารณาใน H_0 และ H_1 ซึ่งขัดแย้งกันเสมอ นั่นคือ ถ้า H_0 เป็นจริงแล้ว H_1 จะไม่จริง และในทางกลับกัน ถ้า H_0 ไม่จริงแล้ว H_1 จะเป็นจริงเสมอ ดังนั้นการขัดแย้งกันของสมมติฐานมี 3 แบบ ดังนี้

แบบที่ 1 $H_0 : \theta = \theta_0$ VS $H_1 : \theta > \theta_0$

แบบที่ 2 $H_0 : \theta = \theta_0$ VS $H_1 : \theta < \theta_0$

แบบที่ 3 $H_0 : \theta = \theta_0$ VS $H_1 : \theta \neq \theta_0$

สรุปว่า สมมติฐานหลัก มีแบบเดียว คือ $H_0 : \theta = \theta_0$

สมมติฐานรอง มี 3 แบบ คือ $H_1 : \theta > \theta_0$

$H_1 : \theta < \theta_0$

$H_1 : \theta \neq \theta_0$

สมมติฐานหลักที่ตั้งขึ้น เพื่อยืนยันความจริงที่เคยทราบมาแล้ว เช่น

- ต้องการทราบว่า วิธีการสอนแบบใหม่มีประสิทธิภาพมากกว่าวิธีการสอนแบบเดิม จริงหรือไม่

H_0 : ประสิทธิภาพของการสอนทั้ง 2 วิธี ไม่แตกต่างกัน

H_1 : วิธีการสอนแบบใหม่มีประสิทธิภาพมากกว่าวิธีการสอนแบบเดิม

แล้วเปลี่ยนข้อความใน H_0 และ H_1 ให้อยู่ในรูปของพารามิเตอร์ ดังนี้

$H_0 : \mu_1 - \mu_2 = 0$ หรือ $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 - \mu_2 > 0$ $H_1 : \mu_1 > \mu_2$

เมื่อ μ_1 , μ_2 แทน คะแนนเฉลี่ยที่ได้จากการสอนวิธีใหม่และวิธีเดิม ตามลำดับ

- สัดส่วนของคนไทยที่มีโทรศัพท์มือถือเท่ากับ 0.53 หรือไม่

$H_0 : p = 0.53$

$H_1 : p > 0.53$ หรือ $H_1 : p < 0.53$ หรือ $H_1 : p \neq 0.53$

เมื่อ p แทน สัดส่วนของคนไทยที่มีโทรศัพท์มือถือ

การทดสอบสมมติฐานทางสถิติ (Statistical hypothesis testing)

หมายถึง กฎเกณฑ์อย่างหนึ่งซึ่งใช้เป็นเกณฑ์ในการตัดสินใจว่า “ จะยอมรับ หรือปฏิเสธ” โดยอาศัยข้อมูลจากตัวอย่างสุ่ม

- ตัวสถิติที่คำนวณได้จากตัวอย่างสุ่ม ที่จะใช้เป็นเครื่องมือในการตัดสินใจว่า ควรจะยอมรับหรือปฏิเสธ H_0 เรียกว่า “**ตัวสถิติทดสอบ** (Test statistic)” ซึ่งอาจเป็นตัวสถิติ Z, T, χ^2 ขึ้นอยู่กับสิ่งที่สนใจศึกษา

- การแจกแจงของตัวสถิติ Z, T, χ^2 จะแบ่งออกเป็น 2 บริเวณ ด้วย “**ค่าวิกฤต** (Critical value)” ได้แก่ **บริเวณยอมรับ** (Acceptance region) และ**บริเวณปฏิเสธ** (Rejection region) หรือ**บริเวณวิกฤต** (Critical region)

- โดยที่ บริเวณยอมรับ (Acceptance region) เป็นบริเวณที่จะทำให้เกิดการยอมรับ H_0 ส่วนบริเวณปฏิเสธ (Rejection region) หรือบริเวณวิกฤต (Critical region) เป็นบริเวณที่จะเกิดการปฏิเสธ H_0

ประเภทของความผิดพลาด

เนื่องจากข้อความใน H_0 อาจเป็นจริงหรือเป็นเท็จก็ได้ และในการตัดสินใจที่จะยอมรับหรือปฏิเสธ H_0 จะขึ้นอยู่กับข้อมูลที่ได้จากตัวอย่างสุ่ม ซึ่งมีความไม่แน่นอนจึงมีโอกาสเกิดความผิดพลาดในการตัดสินใจได้ นั่นคือ เกิดความคลาดเคลื่อนอันเนื่องมาจากตัวอย่างสุ่ม ซึ่งความผิดพลาดที่เกิดขึ้นมี 2 ประเภท ดังนี้

<u>ความผิดพลาดแบบที่ 1 (Type I error)</u>

เป็นความผิดพลาดที่เกิดจากการปฏิเสธ H_0 ทั้งที่ H_0 เป็นจริง และความน่าจะเป็นที่ความผิดพลาดชนิดนี้จะเกิดขึ้น เรียกว่า **ระดับนัยสำคัญ** (Level of significance) หรือ ขนาดของการทดสอบ หรือ การเสี่ยงแบบ 1 (Alpha risk) แทนด้วย α

α = ความน่าจะเป็นที่เกิดความผิดพลาดแบบที่ 1

= P (Type I error)

= P (ปฏิเสธ H_0 | H_0 เป็นจริง)

ความผิดพลาดแบบที่ 2 (Type II error)

เป็นความผิดพลาดที่เกิดจากการยอมรับ H_0 ทั้งที่ H_0 เป็นเท็จ และความน่าจะเป็นที่ความผิดพลาดชนิดนี้จะเกิดขึ้น เรียกว่า การเสี่ยงแบบ 2 (Beta risk) แทนด้วย β

β = ความน่าจะเป็นที่เกิดความผิดพลาดแบบที่ 2

= P (Type II error)

= P (ยอมรับ H_0 | H_0 เป็นเท็จ)

ผลของการตัดสินใจ อาจสรุปได้ดังในตาราง

การตัดสินใจ	สถานการณ์ที่แท้จริง	
	H_0 เป็นจริง	H_0 เป็นเท็จ
ปฏิเสธ H_0	Type I error	No error
ยอมรับ H_0	No error	Type II error

อำนาจการทดสอบ (Power of the test) คือความน่าจะเป็นที่จะปฏิเสธ H_0 เมื่อ H_0 เป็นเท็จ มีค่าเท่ากับ $1-\beta$

7.2 การทดสอบแบบทางเดียว และการทดสอบแบบสองทาง

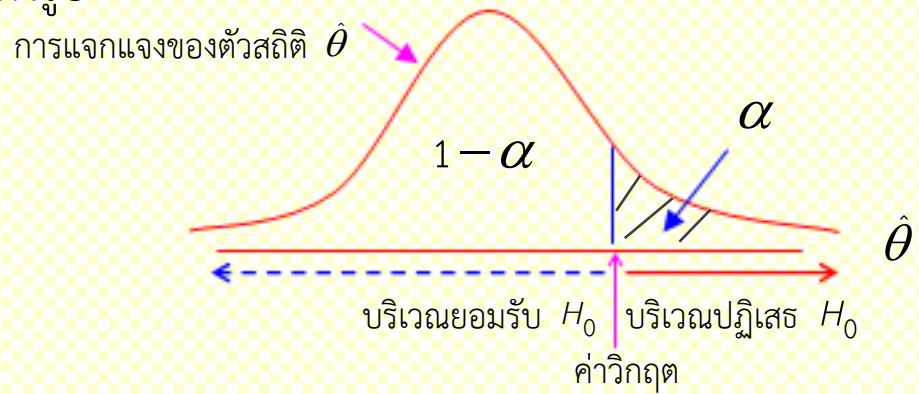
การทดสอบสมมติฐาน แบ่งออกเป็น 2 แบบ คือ การทดสอบแบบทางเดียว และการทดสอบแบบสองทาง ดังนี้

7.2.1 การทดสอบแบบทางเดียว (One-tailed Test) สมมติฐานที่จะทดสอบจะอยู่ใน 2 ลักษณะ ดังนี้

1. $H_0 : \theta = \theta_0$

   $H_1 : \theta > \theta_0$

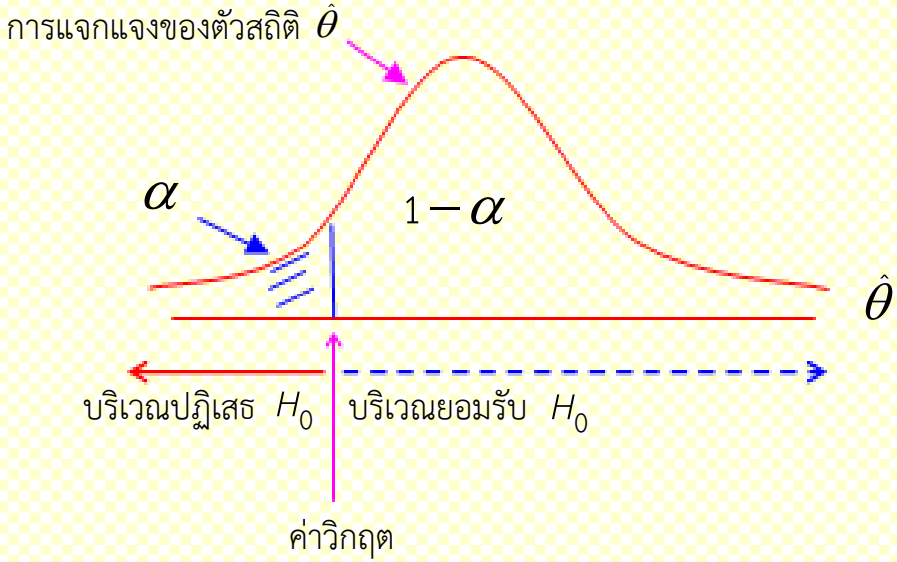
ดังนั้น ถ้า H_0 เป็นจริง แล้ว บริเวณปฏิเสธ H_0 จะอยู่ปลายหางทางขวาของการแจกแจงของตัวสถิติ $\hat{\theta}$ ดังรูป



2. $H_0 : \theta = \theta_0$

$H_1 : \theta < \theta_0$

ดังนั้น ถ้า H_0 เป็นจริง แล้ว บริเวณปฏิเสธ H_0 จะอยู่ปลายหางทางซ้ายของการแจกแจงของตัวสถิติ $\hat{\theta}$ ดังรูป



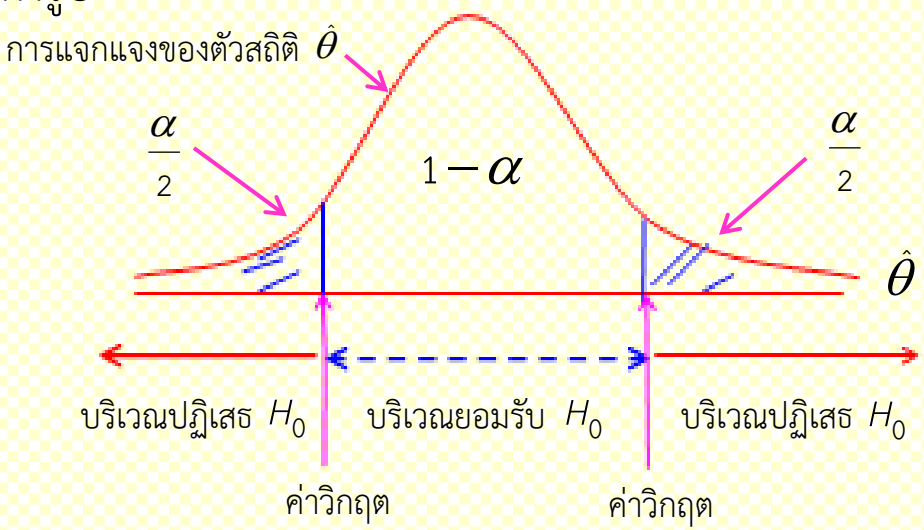
7.2.2 การทดสอบแบบสองทาง (Two-tailed Test)

สมมติฐานที่จะทดสอบอยู่ในลักษณะ ดังนี้

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

ดังนั้น ถ้า H_0 เป็นจริง แล้ว บริเวณปฏิเสธ H_0 จะอยู่ปลายหางทั้งสองของการแจกแจงของตัวสถิติ $\hat{\theta}$ ดังรูป



7.3 ขั้นตอนของการทดสอบสมมติฐาน มีดังนี้

1. ตั้งสมมติฐานหลัก $H_0 : \theta = \theta_0$
2. ตั้งสมมติฐานรอง หรือสมมติฐานแย้ง

   แบบทางเดียว (One-tailed) $H_1 : \theta > \theta_0$ หรือ $H_1 : \theta < \theta_0$

   แบบสองทาง (Two-tailed) $H_1 : \theta \neq \theta_0$
3. กำหนดระดับนัยสำคัญ α
4. เลือกตัวสถิติทดสอบที่เหมาะสม แล้วกำหนดบริเวณปฏิเสธ H_0 ให้สอดคล้องกับ H_1 และ α
5. คำนวณค่าตัวสถิติทดสอบจากตัวอย่างสุ่มขนาด n ที่สุ่มมา
6. สรุปผล ที่ระดับนัยสำคัญ α นั่นคือ ปฏิเสธ H_0 ถ้าค่าสถิติทดสอบที่คำนวณจาก 5. ตกอยู่ในบริเวณปฏิเสธ H_0 หรือยอมรับ H_0 ถ้าค่าสถิติทดสอบที่คำนวณจาก 5. ตกอยู่ในบริเวณยอมรับ H_0

7.4 การทดสอบสมมติฐานค่าเฉลี่ยของหนึ่งประชากร

μ คือ ค่าเฉลี่ยของประชากร และ μ_0 คือ ค่าของค่าเฉลี่ย μ ดังนั้น ต้องการทดสอบว่า **<u>ประชากรที่มีการแจกแจงแบบปกติ</u>** จะมีค่าเฉลี่ย μ เท่ากับ μ_0 หรือไม่ นั่นคือ ทดสอบว่า

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$$

หรือ $$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

หรือ $$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$$

<u>กรณีที่ 1</u> ทราบค่า σ^2 และ n เป็นเท่าใดก็ได้

สมมติฐานหลัก $H_0 : \mu = \mu_0$

ตัวสถิติทดสอบ คือ $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

การทดสอบสมมติฐานที่ใช้ Z เป็นตัวสถิติทดสอบ เรียกว่า “ Z *test* ”

สรุปเกณฑ์การตัดสินใจสำหรับการทดสอบ $H_0 : \mu = \mu_0$ กับสมมติฐานทางเลือกต่าง ๆ กัน โดยกำหนดระดับนัยสำคัญ α ดังตารางต่อไปนี้

H_1	ปฏิเสธ H_0 ถ้า
$H_1 : \mu > \mu_0$	$Z \geq Z_\alpha$
$H_1 : \mu < \mu_0$	$Z \leq -Z_\alpha$
$H_1 : \mu \neq \mu_0$	$Z \leq -Z_{\alpha/2}$ หรือ $Z \geq Z_{\alpha/2}$

หมายเหตุ Z_α และ $Z_{\alpha/2}$ เป็นค่าวิกฤตที่ได้จากการเปิดตารางการแจกแจงปกติมาตรฐาน

<u>กรณีที่ 2</u> ไม่ทราบค่า σ^2 และ $n < 30$

กำหนดสมมติฐานหลัก $H_0 : \mu = \mu_0$

ตัวสถิติทดสอบ คือ $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T_{(\upsilon)}$; องศาความเป็นอิสระ $\upsilon = n-1$

การทดสอบสมมติฐานที่ใช้ T เป็นตัวสถิติในการทดสอบ เรียกว่า " *T test* "

สรุปเกณฑ์การตัดสินใจสำหรับการทดสอบ $H_0 : \mu = \mu_0$ กับสมมติฐานทางเลือกต่างๆกัน โดยกำหนดระดับนัยสำคัญ α ดังตารางต่อไปนี้

H_1	ปฏิเสธ H_0 ถ้า
$H_1 : \mu > \mu_0$	$T \geq t_{\alpha,(n-1)}$
$H_1 : \mu < \mu_0$	$T \leq -t_{\alpha,(n-1)}$
$H_1 : \mu \neq \mu_0$	$T \leq -t_{\alpha/2,(n-1)}$ หรือ $T \geq t_{\alpha/2,(n-1)}$

หมายเหตุ $t_{\alpha,(n-1)}$ และ $t_{\alpha/2,(n-1)}$ เป็นค่าวิกฤตที่ได้จากการเปิดตารางการแจกแจง T ที่มีองศาความเป็นอิสระ $\upsilon = n-1$

<u>กรณีที่ 3</u> ไม่ทราบ σ^2 และ $n \geq 30$

โดยทฤษฎีลิมิตสู่ส่วนกลาง จะได้ $\bar{X} \sim N(\mu, \sigma^2/n)$; นั่นคือ ใช้ S^2 ประมาณ σ^2

ต้องการทดสอบสมมติฐาน $H_0 : \mu = \mu_0$

ตัวสถิติทดสอบ คือ $Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$

การทดสอบสมมติฐานที่ใช้ Z เป็นตัวสถิติทดสอบ เรียกว่า “ *Z test* ”

สรุปเกณฑ์การตัดสินใจสำหรับการทดสอบ $H_0 : \mu = \mu_0$ กับสมมติฐานทางเลือกต่างๆกัน โดยกำหนดระดับนัยสำคัญ α ดังตารางต่อไปนี้

H_1	ปฏิเสธ H_0 ถ้า
$H_1 : \mu > \mu_0$	$Z \geq Z_\alpha$
$H_1 : \mu < \mu_0$	$Z \leq -Z_\alpha$
$H_1 : \mu \neq \mu_0$	$Z \leq -Z_{\alpha/2}$ หรือ $Z \geq Z_{\alpha/2}$

หมายเหตุ Z_α และ $Z_{\alpha/2}$ เป็นค่าวิกฤตที่ได้จากการเปิดตารางการแจกแจงปกติมาตรฐาน

ตัวอย่าง 7.1 เป็นที่ทราบกันว่ารายได้ต่อปีของครัวเรือนเกษตรในจังหวัดนครราชสีมา มีค่าเฉลี่ย 30,000 บาท และค่าเบี่ยงเบนมาตรฐาน 4,000 บาท แต่จากการสุ่มครัวเรือนเกษตรในจังหวัดดังกล่าวมา จำนวน 100 ครัวเรือน พบว่ามีรายได้เฉลี่ยต่อปี 30,800 บาท ที่ระดับนัยสำคัญ 0.05 จะกล่าวได้หรือไม่ว่าครัวเรือนเกษตรในจังหวัดนครราชสีมามีรายได้เฉลี่ยต่อปีสูงขึ้นกว่าเดิม

วิธีทำ ให้ μ เป็นรายได้เฉลี่ยต่อปีของครัวเรือนเกษตรทั้งหมดในจังหวัดนครราชสีมา หน่วย : บาท/ปี

1. $H_0 : \mu = 30,000$
2. $H_1 : \mu > 30,000$
3. $\alpha = 0.05$
4. บริเวณปฏิเสธ H_0 คือ $Z \geq 1.645$
5. ตัวสถิติทดสอบ คือ $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ คำนวณค่า $z = \dfrac{30,800 - 30,000}{4,000/\sqrt{100}} = 2$
6. เพราะว่า $z = 2$ ตกอยู่ในบริเวณปฏิเสธ H_0 ยอมรับ H_1 นั่นคือ ครัวเรือนเกษตรทั้งหมดในจังหวัดนครราชสีมามีรายได้เฉลี่ยต่อปีมากกว่า 30,000 บาท ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 7.2 บริษัทผลิตเลนส์สัมผัสตราหนึ่งโฆษณาว่า เลนส์สัมผัสตราดังกล่าวมีอายุการใช้งานเฉลี่ย 20 เดือน สำนักงานคณะกรรมการคุ้มครองผู้บริโภคมีข้อสงสัยว่าข้อความโฆษณาดังกล่าวอาจเกินความเป็นจริง จึงทำการสุ่มผู้ใช้มา 32 คน พบว่าอายุการใช้งานมีค่าเฉลี่ย 18.9 เดือน และมีค่าเบี่ยงเบนมาตรฐาน 2.7 เดือน ที่ระดับนัยสำคัญ 0.01 จะสรุปได้หรือไม่ว่า ข้อสงสัยของสำนักงานคณะกรรมการคุ้มครองผู้บริโภคเป็นจริง

วิธีทำ ให้ μ เป็นอายุการใช้งานเฉลี่ยของเลนส์สัมผัสตราดังกล่าว หน่วย : เดือน

1. $H_0 : \mu = 20$
2. $H_1 : \mu < 20$
3. $\alpha = 0.01$
4. บริเวณปฏิเสธ H_0 คือ $Z \leq -2.325$
5. ตัวสถิติทดสอบ คือ $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ คำนวณค่า $z = \dfrac{18.9 - 20}{2.7/\sqrt{32}} = -2.3$
6. เพราะว่า $z = -2.3$ ตกอยู่ในบริเวณยอมรับ H_0 นั่นคือ เลนส์สัมผัสตราดังกล่าวมีอายุการใช้งานเฉลี่ยเท่ากับ 20 เดือน ที่ระดับนัยสำคัญ 0.01

ตัวอย่าง 7.3 โรงงานผลิตบุหรี่ตราหนึ่งทราบว่า บุหรี่ที่ผลิตได้มีปริมาณนิโคตินโดยเฉลี่ย 20 มิลลิกรัม จากการสุ่มบุหรี่มาจำนวน 36 มวน พบว่าปริมาณนิโคตินมีค่าเฉลี่ย 22 มิลลิกรัม และค่าเบี่ยงเบนมาตรฐาน 4 มิลลิกรัม ที่ระดับนัยสำคัญ 0.05 จะกล่าวได้หรือไม่ว่า ปริมาณนิโคตินในบุหรี่ตรานี้ได้เปลี่ยนแปลงไปแล้ว

วิธีทำ ให้ μ เป็นปริมาณนิโคตินโดยเฉลี่ยของบุหรี่ตราดังกล่าว หน่วย : มิลลิกรัม

1. $H_0 : \mu = 20$
2. $H_1 : \mu \neq 20$
3. $\alpha = 0.05$
4. บริเวณปฏิเสธ H_0 คือ $Z \leq -1.96$ หรือ $Z \geq 1.96$
5. ตัวสถิติทดสอบ คือ $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ คำนวณค่า $z = \dfrac{22-20}{4/\sqrt{36}} = 3$
6. เพราะว่า $z = 3$ ตกอยู่ในบริเวณปฏิเสธ H_0 ยอมรับ H_1

นั่นคือ บุหรี่ตรานี้มีปริมาณนิโคตินเฉลี่ยแตกต่างจาก 20 มิลลิกรัม ที่ระดับนัยสำคัญ 0.05

ตัวอย่าง 7.4 จากการศึกษาคุณภาพอากาศใน กทม. ก่อนที่จะมีการออกมาตรการพบว่า มีปริมาณก๊าซคาร์บอนมอนนอกไซด์โดยเฉลี่ย 9.4 ส่วนต่อล้านส่วน (ppm) เพื่อตรวจสอบว่ามาตรการดังกล่าวช่วยลดปริมาณก๊าซคาร์บอนมอนนอกไซด์ได้จริง โดยสุ่มอากาศจุดต่างๆทั่ว กทม. รวม 18 จุด วัดปริมาณก๊าซคาร์บอนมอนนอกไซด์ได้ดังนี้ หน่วย : ppm

8.6	6.4	7.2	10.5	8.7	10.7	5.4
5.7	3.9	4.5	3.6	7.6	6.8	10.9
10.2	7.9	9.4	7.9			

ที่ระดับนัยสำคัญ 0.05 มาตรการดังกล่าวได้ผลหรือไม่

วิธีทำ ให้ μ เป็นปริมาณก๊าซคาร์บอนมอนนอกไซด์โดยเฉลี่ยของอากาศใน กทม. หน่วย : ppm

1. $H_0 : \mu = 9.4$
2. $H_1 : \mu < 9.4$
3. $\alpha = 0.05$
4. บริเวณปฏิเสธ H_0 คือ $T \leq -1.74$
5. จาก $\bar{X} = \frac{\sum x}{n}$ จะได้ $\bar{X} = \frac{135.9}{18} = 7.55$

จาก $s^2 = \dfrac{\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}}{n-1}$ จะได้ $S^2 = \dfrac{1{,}117.29 - \dfrac{(135.9)^2}{18}}{18-1} = 5.37$

และ $S = 2.32$

ตัวสถิติทดสอบ คือ $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

คำนวณค่า

$$t = \frac{7.55 - 9.4}{2.32/\sqrt{18}} = -3.38$$

6. เพราะว่า $t = -3.38$ ตกอยู่ในบริเวณปฏิเสธ H_0 ยอมรับ H_1 นั่นคือ ปริมาณก๊าซคาร์บอนมอนนอกไซด์โดยเฉลี่ยมีค่าน้อยกว่า 9.4 ppm ที่ระดับนัยสำคัญ 0.05

7.5 การทดสอบสมมติฐานสัดส่วนของหนึ่งประชากร

ถ้าประชากรมีการแจกแจงแบบทวินาม มีพารามิเตอร์ n และ p เมื่อต้องการทดสอบสมมติฐานว่า สัดส่วนประชากร p มีค่าเท่ากับ p_0 หรือไม่ การทดสอบสมมติฐานอยู่ในลักษณะดังนี้

$$H_0 : p = p_0 \quad \text{VS} \quad H_1 : p > p_0$$

หรือ $$H_0 : p = p_0 \quad \text{VS} \quad H_1 : p < p_0$$

หรือ $$H_0 : p = p_0 \quad \text{VS} \quad H_1 : p \neq p_0$$

สมมติฐานหลัก $H_0 : p = p_0$

ตัวสถิติทดสอบ คือ $Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} \sim N(0,1)$; โดยที่ $q_0 = 1 - p_0$

เมื่อ $\hat{p} = \dfrac{x}{n}$ เป็นสัดส่วนของตัวอย่างสุ่ม n ที่มีขนาดใหญ่

สรุปเกณฑ์การตัดสินใจสำหรับการทดสอบ $H_0 : p = p_0$ กับสมมติฐานทางเลือกต่างๆ กัน โดยกำหนดระดับนัยสำคัญ α ดังตารางต่อไปนี้

H_1	ปฏิเสธ H_0 ถ้า
$H_1 : p > p_0$	$Z \geq Z_\alpha$
$H_1 : p < p_0$	$Z \leq -Z_\alpha$
$H_1 : p \neq p_0$	$Z \leq -Z_{\alpha/2}$ หรือ $Z \geq Z_{\alpha/2}$

หมายเหตุ Z_α และ $Z_{\alpha/2}$ เป็นค่าวิกฤตที่ได้จากการเปิดตารางการแจกแจงปกติมาตรฐาน

ตัวอย่าง 7.5 ผู้จัดการโรงงานแห่งหนึ่งทราบว่า ในแต่ละครั้งของการผลิตจะถือว่าเครื่องจักรทำงานเป็นปกติ หากมีสินค้าชำรุดอย่างมากร้อยละ 3 เพื่อตรวจสอบสภาพการทำงานของเครื่องจักรเครื่องหนึ่ง โดยสุ่มสินค้าที่ผลิตจากเครื่องดังกล่าวมาจำนวน 500 ชิ้น พบว่ามีสินค้าชำรุด 22 ชิ้น ที่ระดับนัยสำคัญ 0.05 ผู้จัดการโรงงานแห่งนี้จะสรุปได้หรือไม่ว่า เครื่องจักรทำงานไม่เป็นปกติ

วิธีทำ ให้ p เป็นสัดส่วนสินค้าชำรุดที่ได้จากการผลิตในครั้งดังกล่าว

1. $H_0 : p = 0.03$
2. $H_1 : p > 0.03$
3. $\alpha = 0.05$
4. บริเวณปฏิเสธ H_0 คือ $Z \geq 1.645$
5. ตัวสถิติทดสอบ คือ $Z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$ จะได้ $z = \dfrac{\dfrac{22}{500} - 0.03}{\sqrt{\dfrac{(0.03)(0.97)}{500}}} = 1.84$
6. เพราะว่า $z = 1.84$ ตกอยู่ในบริเวณปฏิเสธ H_0 ยอมรับ H_1

นั่นคือ สินค้าชำรุดที่ผลิตจากครั้งดังกล่าวมีสัดส่วนมากกว่าร้อยละ 3 ที่ระดับนัยสำคัญ 0.0๕

7.6 การทดสอบสมมติฐานความแปรปรวนของหนึ่งประชากร

เมื่อ σ_0^2 คือ ค่าของความแปรปรวนของประชากรที่มีการแจกแจงปกติ สมมติฐานที่จะทดสอบอยู่ในลักษณะดังนี้

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 > \sigma_0^2$$

$$\text{หรือ} \quad H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 < \sigma_0^2$$

$$\text{หรือ} \quad H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 \neq \sigma_0^2$$

ตัวสถิติทดสอบ คือ $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ ซึ่งเป็นการแจกแจงแบบไคสแควร์ ที่มีองศาแห่งความอิสระ $\upsilon = n-1$

สรุปเกณฑ์การตัดสินใจสำหรับการทดสอบ $H_0 : \sigma^2 = \sigma_0^2$ กับสมมติฐานทางเลือกต่างๆ กัน โดยกำหนดระดับนัยสำคัญ α ดังตารางดังต่อไปนี้

H_1	ปฏิเสธ H_0 ถ้า
$H_1 : \sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi^2_{\alpha}$
$H_1 : \sigma^2 < \sigma_0^2$	$\chi^2 \leq -\chi^2_{1-\alpha}$
$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi^2 \leq -\chi^2_{1-\alpha/2}$ หรือ $\chi^2 \geq \chi^2_{\alpha/2}$

หมายเหตุ ค่าวิกฤตในกรณีต่างๆ ได้จากการเปิดตารางการแจกแจงไคสแควร์ ที่มีองศาความเป็นอิสระ $\upsilon = n-1$

ตัวอย่าง 7.6 โรงงานผลิตหลอดภาพโทรทัศน์แห่งหนึ่งทราบว่า อายุการใช้งานของหลอดภาพมีการแจกแจงปกติ ที่มีความแปรปรวน 10,000 ซ.ม.^{2} ในการตรวจสอบคุณภาพครั้งหนึ่ง โดยการสุ่มหลอดภาพมา 20 หลอด พบว่าความแปรปรวนของอายุการใช้งานของหลอดภาพเท่ากับ 12,000 ซ.ม.^{2} ที่ระดับนัยสำคัญ 0.05 จะกล่าวได้หรือไม่ว่า ความแปรปรวนของอายุการใช้งานของหลอดภาพไม่เท่ากับ 10,000 ซ.ม.^{2}

วิธีทำ ให้ σ^2 เป็นความแปรปรวนของอายุการใช้งานของหลอดภาพที่ผลิตโดยโรงงานแห่งนี้ หน่วย : ซ.ม.^{2}

1. $H_0 : \sigma^2 = 10{,}000$
2. $H_1 : \sigma^2 \neq 10{,}000$
3. $\alpha = 0.05$
4. บริเวณปฏิเสธ H_0 คือ $\chi^2 \leq 8.91$ หรือ $\chi^2 \geq 32.9$
5. ตัวสถิติทดสอบ คือ $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ จะได้ $\chi^2 = \dfrac{(20-1)(12{,}000)}{10{,}000} = 22.8$
6. เพราะว่า $\chi^2 = 22.8$ ตกอยู่ในบริเวณยอมรับ H_0

นั่นคือ ความแปรปรวนของอายุการใช้งานของหลอดภาพเท่ากับ 10,000 ซ.ม.^{2} ที่ระดับนัยสำคัญ 0.05

แบบฝึกหัด บทที่ 7

1. จากการศึกษาค่าแรงของลูกจ้างในโรงงาน ต้องมีค่าแรงเฉลี่ยขั้นต่ำเท่ากับ 300 บาท เก็บข้อมูลตัวอย่างลูกจ้างจำนวน 25 คน พบว่ามีค่าแรงเฉลี่ย 330 บาท ส่วนเบี่ยงเบนมาตรฐาน 50 บาท ที่ระดับนัยสำคัญ 0.05 จงตรวจสอบว่าค่าแรงเฉลี่ยของลูกจ้างในโรงงานแตกต่างจากค่าแรงขั้นต่ำหรือไม่

2. จากการสำรวจครัวเรือนในเขตเทศบาลนครเชียงใหม่จำนวน 36 ครัวเรือน พบว่า โดยเฉลี่ยครัวเรือนชมรายการโทรทัศน์ 27 ชั่วโมงต่อสัปดาห์ ค่าเบี่ยงเบนมาตรฐาน 4 ชั่วโมง ถ้าจำนวนชั่วโมงการชมโทรทัศน์โดยเฉลี่ยของครัวเรือนทั่วประเทศเท่ากับ 25 ชั่วโมงต่อสัปดาห์ จะกล่าวสรุปได้หรือไม่ว่า จำนวนชั่วโมงการชมรายการโทรทัศน์โดยเฉลี่ยของครัวเรือนในเขตเทศบาลนครเชียงใหม่มากกว่าจำนวนชั่วโมงการชมรายการโทรทัศน์โดยเฉลี่ยของครัวเรือนทั่วประเทศ ที่ระดับนัยสำคัญ 0.01

3. กระบวนการผลิตในโรงงานหนึ่ง จะผลิตสินค้าเสีย 20% ฝ่ายผลิตจึงได้ดำเนินการปรับปรุงกระบวนการผลิตใหม่ แล้วสุ่มตัวอย่างสินค้ามา 100 ชิ้น พบว่าเป็นสินค้าเสีย 16 ชิ้น จะเชื่อได้หรือไม่ว่า การปรับปรุงกระบวนการผลิตจะทำให้สินค้าเสียลดลง ที่ระดับนัยสำคัญ 0.05

แบบฝึกหัด บทที่ 7

4. ตามหลักพันธุกรรมกล่าวว่า พืชที่ได้มาจากการผสมพันธุ์ระหว่างเมล็ดพันธุ์ ก และ ข จะมีลักษณะเตี้ย 80% ได้มีผู้ทดลองผสมพันธุ์พืช 2 ชนิดนี้ คือ ก และ ข พบว่า พืชที่ได้จากการผสมพันธุ์นี้ 200 ต้น ปรากฏว่ามี 64 ต้น ที่ไม่ใช่ต้นเตี้ย จงทดสอบว่า การทดลองนี้จะสอดคล้องตามทฤษฎีดังกล่าวข้างต้นนี้หรือไม่ โดยใช้ระดับนัยสำคัญ 0.10

5. จากการเลือกร้านสรรพสินค้าโดยการสุ่มจำนวน 15 ร้าน จากร้านสรรพสินค้าทั้งหมดในเขตกรุงเทพมหาครมาเก็บรวบรวมข้อมูลเกี่ยวกับราคากระทะไฟฟ้าชนิดหนึ่ง ปรากฏว่าได้ราคาเฉลี่ย 515 บาท และความแปรปรวน 350 จงทดสอบความเชื่อที่ว่า ค่าความแปรปรวนของราคากระทะไฟฟ้าชนิดนี้ต่ำกว่า 320 ที่ระดับนัยสำคัญ 0.05

บทที่ 8
การทดสอบไคสแควร์
(Chi-square Test)

การทดสอบไคสแควร์

สถิติที่ใช้ทดสอบความแตกต่างค่าเฉลี่ยของกลุ่มตัวอย่างที่มีเพียงกลุ่มหรือสองกลุ่ม จะใช้ทดสอบด้วยค่า Z-test หรือ T-test ข้อมูลที่นำมาทดสอบนั้นจะต้องเป็นข้อมูลที่อยู่ในระดับการวัด (Measurement Scale) ระดับอันตรภาคชั้น (Intervel Scale) หรือระดับอัตราส่วน (Ratio Scale) เท่านั้น ในงานวิจัยบางเรื่องข้อมูลอาจอยู่ในรูปของความถี่ที่เป็นอิสระต่อกัน (Discrete Data) เป็นข้อมูลที่อยู่ในระดับนามบัญญัติ (Norminal Scale) หรือข้อมูลเรียงลำดับ (Ordinal Scale) การทดสอบข้อมูลในลักษณะนี้จะเป็นการทดสอบว่า ข้อมูลที่ได้เป็นไปตามคาดหวัง (Expected) ไว้หรือไม่ หรืออาจจะทดสอบว่าตัวแปร (Variable) มีความสัมพันธ์กันหรือไม่ ข้อมูลดังกล่าวไม่สามารถ ทดสอบได้ด้วย Z-test หรือ T-test ซึ่งเป็นสถิติแบบพาราเมตริก (Parametric Statistics) แต่จะ สามารถทดสอบได้ด้วยไคสแควร์ (χ^2) ซึ่งเป็นสถิติแบบนอนพาราเมตริก (Nonpara -metric Statistics) ซึ่งเป็นสถิติที่ไม่คำนึงถึงลักษณะการแจกแจงของประชากร

การทดสอบไคสแควร์

การทดสอบโดยใช้ χ^2 จะมีอยู่ 2 กรณี คือ

1. การทดสอบกรณีตัวแปรเดียว (The χ^2 one - variable case หรือบางครั้งอาจเรียกว่า **การทดสอบภาวสารูปสนิทดี** (Goodness of fit test)

2. การทดสอบกรณีสองตัวแปร (The χ^2 two – variable case) เป็นการทดสอบเพื่อดูว่าตัวแปรสองตัวมีความสัมพันธ์หรือเกี่ยวข้องกันหรือไม่ ดังนั้นบางทีจึงเรียกว่า **การทดสอบความเป็นอิสระ** (The χ^2 test for independence)

การทดสอบภาวสารูปสนิทดี

- ใช้เพื่อทดสอบว่าการแจกแจงความถี่สอดคล้องกับการแจกแจงคาดหวังหรือไม่
- การทดสอบนี้สามารถใช้กับข้อมูลเชิงคุณภาพ

การทดสอบภาวสารูปสนิทดี เป็นการทดสอบสำหรับตัวอย่าง 1 กลุ่ม ทดสอบเกี่ยวกับการแจกแจงของตัวแปร 1 ตัว ว่าเป็นไปตามสัดส่วนที่กำหนดไว้หรือไม่ ตัวแปรมีสเกลการวัดแบบแบ่งประเภท (Nominal scale) ข้อมูลมาจากประชากรกลุ่มเดียว

การทดสอบภาวสารูปสนิทดี

การทดสอบ χ^2 กรณีกลุ่มตัวอย่างกลุ่มเดียว เป็นการทดสอบตัวแปรเพียงด้านเดียวเพื่อต้องการ ทราบว่าความถี่ที่ได้จากการสังเกต (Observed Frequency) จากกลุ่มตัวอย่าง เป็นไปตามความถี่ที่คาดหวัง (Expected Frequency) ไว้หรือไม่ตามนัยสำคัญที่กำหนด การทดสอบโดยการใช้สูตร คำนวณ χ^2 test คือ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = ค่าสถิติไคสแควร์

O_i = ความถี่ที่ได้จากการสังเกต (Observed Frequency)

E_i = ความถี่ที่คาดหวัง (Expected Frequency) ซึ่งมีค่าเท่ากับจำนวนข้อมูลคูณด้วยสัดส่วนที่คาดหวัง (ซึ่ง $E_i = np_i$)

ตัวอย่าง 1 ห้างสรรพสินค้า B ได้ผลิตขนมเค้กยี่ห้อ B ออกจำหน่ายตามสาขาต่าง ๆ ของตน โดยจะขายในราคาที่ถูกกว่ายี่ห้ออื่น ๆ ที่มีชื่อเสียงอีก 4 ยี่ห้อ คือ C, D, E และ F ที่ทางห้างรับมาขาย ทางห้างต้องการทราบว่าสัดส่วนของลูกค้าที่ชอบขนมเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบขนมเค้กยี่ห้อ C, D, E และ F หรือไม่ จึงสุ่มลูกค้ามา 100 คน ให้ชิมขนมเค้กทั้ง 5 ยี่ห้อ แล้วให้ลูกค้าบอกว่าชอบยี่ห้อใดมากที่สุด ได้ผลดังนี้

ยี่ห้อ	B	C	D	E	F
จำนวนลูกค้า	17	27	22	15	19

จงทดสอบสิ่งที่ทางห้างต้องการ ที่ระดับนัยสำคัญ 0.05

วิธีทำ สมมติฐานในการทดสอบ

H_0 : สัดส่วนของลูกค้าที่ชอบเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้ออื่น ๆ อีก 4 ยี่ห้อ

H_1 : สัดส่วนของลูกค้าที่ชอบเค้กยี่ห้อ B ไม่เท่ากับสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้ออื่น ๆ อีก 4 ยี่ห้อ

ยี่ห้อ	จำนวนลูกค้า (O_i)	$E_i = np_i$	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
B	17	20	-3	9	0.45
C	27	20	7	49	2.45
D	22	20	2	4	0.2
E	15	20	-5	25	1.25
F	19	20	-1	1	0.05
รวม	100	100			4.4

ค่าตัวสถิติทดสอบ

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 4.4$$

ค่าวิกฤติ $\chi^2_{\alpha, k-1} = \chi^2_{0.05, 4} = 9.488$

เนื่องจากค่าสถิติทดสอบ = 4.4 น้อยกว่า $\chi^2_{0.05, 4} = 9.488$ อยู่ในบริเวณยอมรับ H_0

สรุปผลการทดสอบ

หมายความว่า สัดส่วนของลูกค้าที่ชอบเค้กยี่ห้อ B เท่ากับสัดส่วนของลูกค้าที่ชอบเค้กยี่ห้ออื่น ๆ อีก 4 ยี่ห้อ ที่ระดับนัยสำคัญ 0.05

การทดสอบความเป็นอิสระ

การทดสอบในกรณีตัวแปรสองตัวนี้เป็นการทดสอบเพื่อดูว่า ตัวแปรสองตัวนี้มีความเกี่ยวข้องหรือสัมพันธ์กันหรือไม่ ถ้าไม่สัมพันธ์กันหมายความว่าเป็นอิสระจากกัน ดังนั้นบางครั้งเรา จึงเรียกว่า **การทดสอบความเป็นอิสระ** (χ^2- test for independence) ข้อมูลที่ได้จะอยู่ในระดับนามบัญญัติ (Norminal scale) ซึ่งอาจเป็นจำนวนความถี่ สัดส่วน ร้อยละก็ได้ โดยแต่ละตัวแปรจะแบ่งเป็น 2 กลุ่ม หรือ 2 ประเภทขึ้นไป เช่น เพศ (ชาย - หญิง) กับวุฒิการศึกษา (ป.ตรี ป .โท ป.เอก) จะได้รูปแบบเป็น 2 x 3 ดังนั้นรูปแบบการวิเคราะห์อาจเป็นได้หลายรูปแบบขึ้นอยู่กับจำนวนกลุ่มของแต่ละตัวแปร (2 x 2, 2 x 4, 3 x 2 เป็นต้น)

ลักษณะข้อมูลจำแนกสองทาง

ในกรณีที่ข้อมูลเรื่องใดเรื่องหนึ่งถูกจำแนกโดย ตัวแปร 2 ตัวแปร ซึ่งโดยทั่วไปข้อมูลที่นำมาทดสอบจะเป็นข้อมูลจำแนก 2 ทาง โดยตัวแปรที่ 1 จะแบ่งเป็น r กลุ่ม (row) และตัวแปรที่ 2 จะแบ่งเป็น c กลุ่ม (column) เราจะเรียกตารางนี้ว่า ตารางการณ์จร ขนาด rxc (rxc contingency table) โดยข้อมูลที่วิเคราะห์อยู่ในรูปแบบตาราง ดังนี้

ตัวแปรที่ 1	**ตัวแปรที่ 2**				**รวม**
	1	2	...	C	
1	O_{11}	O_{12}	...	O_{1c}	r_1
2	O_{21}	O_{22}	...	O_{2c}	r_2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
r	O_{r1}	O_{r2}	...	O_{rc}	r_r
รวม	c_1	c_2	...	c_c	$n = \sum r_i = \sum c_j$

เมื่อ O_{ij} แทน ความถี่ของตัวแปรที่ 1 กลุ่ม i และตัวแปรที่ 2 กลุ่ม j

; i = 1, 2, ..., r , j = 1, 2, ..., c

r_i แทน ความถี่รวมของตัวแปรที่ 1 กลุ่ม i ; i = 1, 2, ..., r

c_i แทน ความถี่รวมของตัวแปรที่ 2 กลุ่ม j ; j = 1, 2, ... , c

และ ความถี่รวมทั้งหมด $n = \sum r_i = \sum c_j = \sum\sum o_{ij}$

หมายเหตุ

ถ้าตัวแปรที่ 1 มี 2 กลุ่ม และตัวแปรที่ 2 มี 2 กลุ่ม เรียก **ตารางการณ์จรขนาด 2×2**

สมมติฐานในการทดสอบ

H_0 : ตัวแปรที่ 1 เป็นอิสระกับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 ไม่เป็นอิสระกับตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่มีความสัมพันธ์กับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีความสัมพันธ์กับตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่มีผลต่อตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 มีผลต่อตัวแปรที่ 2

หรือ

H_0 : ตัวแปรที่ 1 ไม่ขึ้นกับตัวแปรที่ 2

H_1 : ตัวแปรที่ 1 ขึ้นกับตัวแปรที่ 2

ตัวสถิติทดสอบ

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

โดยที่ χ^2 แทน ไคสแควร์ ที่องศาแห่งความเป็นอิสระ (df) = (r-1)(c-1)

O_{ij} แทน ความถี่ที่ได้จากการสังเกตใน cell (i, j)

E_{ij} แทน ความถี่คาดหวังใน cell (i, j) ซึ่ง $E_{ij} = \frac{r_i c_j}{n}$

ตัวอย่าง 2 บริษัทผลิตมือถือแห่งหนึ่งต้องการทราบว่า ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้หรือไม่ จึงทำการเก็บรวบรวมข้อมูลจากผู้ใช้โทรศัพท์มือถือจำนวน 1,000 คน เป็นเพศชาย 500 คน และ หญิง 500 คน ได้ผลดังนี้

เพศ	สีของโทรศัพท์มือถือ		รวม
	สีดำ	สีอื่น ๆ	
ชาย	328	172	500
หญิง	138	362	500
รวม	466	534	1,000

อยากทราบว่า ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้หรือไม่ ที่ระดับนัยสำคัญ 0.05

วิธีทำ สมมติฐานการทดสอบ

H_O : ความชอบสีของโทรศัพท์มือถือไม่มีความสัมพันธ์กับเพศของผู้ใช้

H_1 : ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้

เนื่องจาก

$$O_{11} = 328 \quad , \quad E_{11} = \frac{500 \times 466}{1,000} = 233$$

$$O_{12} = 172 \quad , \quad E_{12} = \frac{500 \times 534}{1,000} = 267$$

$$O_{21} = 138 \quad , \quad E_{21} = \frac{500 \times 466}{1,000} = 233$$

$$O_{22} = 362 \quad , \quad E_{22} = \frac{500 \times 534}{1,000} = 267$$

ค่าตัวสถิติทดสอบ

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

$$= \frac{(328-233)^2}{233}+\frac{(172-267)^2}{267}+\frac{(138-233)^2}{233}+\frac{(362-267)^2}{267}$$

$$= 145.07$$

ค่าวิกฤต $\chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{0.05,(1)(1)} = \chi^2_{0.05,1} = 3.841$

เนื่องจากค่าสถิติทดสอบ $\chi^2 = 145.07$ มากกว่า $\chi^2_{0.05,1} = 3.841$ อยู่ในบริเวณปฏิเสธ H_0

สรุปผลการทดสอบ

หมายความว่า ความชอบสีของโทรศัพท์มือถือมีความสัมพันธ์กับเพศของผู้ใช้ ที่ระดับนัยสำคัญ 0.05

แบบฝึกหัดท้ายบท

1. ฝ่ายบริหารพัสดุของบริษัทแห่งหนึ่ง ทำการศึกษาข้อมูลการเบิกจ่ายสินค้าในวันจันทร์ - ศุกร์ ในช่วง 6 เดือนที่ผ่านมา พบว่ามีความถี่ดังนี้

วัน	จันทร์	อังคาร	พุธ	พฤหัสบดี	ศุกร์	รวม
ความถี่	-35	35	45	40	65	220

จงพิสูจน์ว่าการเบิกจ่ายสินค้าในช่วง 5 วันนี้มีโอกาสที่จะเกิดเท่ากันหรือไม่ ที่ระดับนัยสำคัญ 0.05

แบบฝึกหัดท้ายบท

2. บริษัทแห่งหนึ่งต้องการเปิดทำงานในช่วงวันเสาร์และอาทิตย์ โดยคิดว่าพนักงานต้องการมีรายได้เสริม เพื่อการตัดสินใจที่ถูกต้อง จึงทำการเก็บรวบรวมข้อมูลจากคนงานในบริษัทจำนวน 250 คน โดยแบ่งรายได้ออกเป็น 2 ช่วง คือ รายได้ต่ำกว่า 10,000 บาท และตั้งแต่ 10,000 บาทขึ้นไป และความคิดเห็นเป็น 3 ระดับ ได้ผลดังนี้

รายได้	**ความคิดเห็น**			**รวม**
	เห็นด้วย	**ไม่มีความเห็น**	**ไม่เห็นด้วย**	
ต่ำกว่า 10,000 บาท	68	12	110	190
ตั้งแต่ 10,000 บาทขึ้นไป	48	2	10	60
รวม	116	14	120	250

อยากทราบว่าความคิดเห็นในการทำงานวันเสาร์และอาทิตย์ขึ้นอยู่กับรายได้หรือไม่ ที่ระดับนัยสำคัญ 0.05

บทที่ 9

การวิเคราะห์ข้อมูลด้วยโปรแกรมสำเร็จรูปทางสถิติ

ประเภทของสถิติ

สถิติแบ่งออกเป็น 2 ประเภท คือ

1. สถิติพรรณนา (Descriptive Statistics) เป็นสถิติที่ใช้อธิบายคุณลักษณะของสิ่งที่ต้องการศึกษากลุ่มใดกลุ่มหนึ่ง ไม่สามารถอ้างอิงไปยังกลุ่มอื่น ๆ ได้ สถิติที่อยู่ในประเภทนี้ เช่น ค่าเฉลี่ย ค่ามัธยฐาน ค่าฐานนิยม ส่วนเบี่ยงเบนมาตรฐาน ค่าพิสัย ฯลฯ

2. สถิติอ้างอิง (Inferential Statistics) เป็นสถิติที่ใช้อธิบายคุณลักษณะของสิ่งที่ต้องการศึกษากลุ่มใดกลุ่มหนึ่ง หรือหลายกลุ่ม แล้วสามารถอ้างอิงไปยังกลุ่มประชากรได้ โดยกลุ่มที่นำมาศึกษาจะต้องเป็นตัวแทนที่ดีของประชากร ตัวแทนที่ดีของประชากรได้มาโดยวิธีการสุ่มตัวอย่าง และตัวแทนที่ดีของประชากรเรียกว่า **กลุ่มตัวอย่าง**

ประเภทของสถิติ

สถิติอ้างอิงสามารถแบ่งออกได้เป็น 2 ประเภท คือ

2.1 สถิติแบบใช้พารามิเตอร์ (Parametric Statistics) เป็นวิธีการทางสถิติที่จะต้องเป็นไปตามข้อตกลงเบื้องต้น 3 ประการ ดังนี้

1. ข้อมูลที่เก็บรวบรวมได้ต้องเป็นข้อมูลที่อยู่ในระดับช่วง (Interval Scale)
2. ข้อมูลที่เก็บรวบรวมได้จากกลุ่มตัวอย่าง จะต้องมีการแจกแจงแบบปกติ
3. กลุ่มประชากรแต่ละกลุ่มที่นำมาศึกษาจะต้องมีความแปรปรวนเท่ากัน

สถิติแบบใช้พารามิเตอร์ เช่น t-test, ANOVA, Regression Analysis เป็นต้น

2.2 สถิติแบบไม่ใช้พารามิเตอร์ (Nonparametric Statistics) คือ สถิติที่ไม่อยู่ในข้อตกลงเบื้องต้นทั้ง 3 ประการ สถิติประเภทนี้ ได้แก่ Chi-Square และการวิเคราะห์ความสัมพันธ์

ระดับของการวัด

การวัดเป็นการกำหนดตัวเลขให้กับสิ่งที่ต้องการศึกษา แบ่งออกเป็น 4 ระดับ คือ

1. ระดับนามบัญญัติ (Nominal Scale) เป็นระดับที่ใช้จำแนกความแตกต่างของสิ่งที่ต้องการวัดออกเป็นกลุ่ม เช่น เพศ การศึกษา อาชีพ ภูมิลำเนา โดยในแต่ละกลุ่มจะแทนด้วยตัวเลข เช่น เพศชาย แทนด้วยเลข 1 เพศหญิง แทนด้วยเลข 2

2. ระดับเรียงลำดับ (Ordinal Scales) เป็นระดับที่ใช้สำหรับจัดอันดับที่หรือตำแหน่งของสิ่งที่ต้องการวัด เช่น ระดับความพึงพอใจ

พึงพอใจมากที่สุด = 5 พึงพอใจมาก = 4

พึงพอใจปานกลาง = 3 พึงพอใจน้อย = 2

พึงพอใจน้อยที่สุด = 1

ระดับของการวัด

3. ระดับช่วง (Interval Scale) เป็นระดับที่สามารถกำหนดค่าตัวเลขโดยมีช่วงห่างระหว่างตัวเลขเท่า ๆ กัน แต่ไม่มี 0 (ศูนย์) แท้ มีแต่ 0 (ศูนย์) สมมติ เช่น สอบได้ 0 คะแนน มิได้หมายความว่าเขาไม่มีความรู้ เพียงแต่เขาไม่สามารถทำข้อสอบซึ่งเป็นตัวแทนของความรู้ทั้งหมดได้ หรืออุณหภูมิ 0 องศา มิได้หมายความว่าจะไม่มีความร้อน เพียงแต่มีความร้อนเป็น 0 องศาเท่านั้น จุดที่ไม่มีความร้อนอยู่เลย ก็คือ ที่ -273 องศา ดังนั้น อุณหภูมิ 40 องศาจึงไม่สามารถบอกได้ว่ามีความร้อนเป็น 2 เท่าของอุณหภูมิ 20 องศาเป็นต้น ตัวเลขในระดับนี้สามารถนำมาบวก ลบ คูณ หรือหารกันได้

4. ระดับอัตราส่วน (Ratio Scale) เป็นระดับที่สามารถกำหนดค่าตัวเลขให้กับสิ่งที่ต้องการวัดมี 0 (ศูนย์) แท้ เช่น น้ำหนัก ความสูง อายุ เป็นต้น ระดับนี้สามารถนำตัวเลขมาบวก ลบ คูณ หาร หรือหาอัตราส่วนกันได้ คือสามารถบอกได้ว่าถนนสายหนึ่งยาว 50 กิโลเมตร ยาวเป็น 2 เท่าของถนนอีกสายหนึ่งที่ยาวเพียง 25 กิโลเมตร

ระดับของการวัด

สรุปได้ว่า

ระดับการวัด	ข้อมูล
1. Nominal Scale	ข้อมูลเชิงคุณภาพ
2. Ordinal Scales	
3. Interval Scale	ข้อมูลเชิงปริมาณ
4. Ratio Scale	

ประชากรและกลุ่มตัวอย่าง

ในงานวิจัยโดยมาก ค่าของการวัดจะได้มาจากกลุ่มตัวอย่าง ซึ่งมาจากประชากรที่มีขนาดใหญ่

ประชากร คือ กลุ่มของการวัดทั้งหมดที่สนใจศึกษา

ตัวอย่าง คือ สับเซตของการวัดที่มาจากประชากรที่สนใจศึกษา

หลักการสุ่มกลุ่มตัวอย่าง

- สมาชิกเป็นตัวแทนที่ดีของประชากร
- มีคุณลักษณะสำคัญเหมือนประชากร
- สมาชิกทุกหน่วยมีโอกาสได้เป็นกลุ่มตัวอย่างเท่ากัน
- กลุ่มตัวอย่างต้องมีขนาดใหญ่เพียงพอ

ลักษณะแบบสอบถาม

แบบสอบถามที่ใช้ในการวิจัยเชิงปริมาณนั้น มักจะมีรูปแบบดังนี้

ตอนที่ 1 สอบถามเกี่ยวกับสถานภาพทั่วไปของผู้ตอบแบบสอบถาม มักเป็นคำถามแบบตรวจสอบรายการ (Check-List)

ตอนที่ 2 สอบถามเกี่ยวกับลักษณะโดยทั่วไปของเรื่องที่กำลังทำวิจัย มักเป็นคำถามแบบตรวจสอบรายการ (Check-List)

ตอนที่ 3 สอบถามเกี่ยวกับความพึงพอใจ หรือทัศนคติในประเด็นต่าง ๆ ของเรื่องที่ทำวิจัย ลักษณะคำถามเป็นแบบมาตราส่วนประมาณค่า (Rating Scale)

ตอนที่ 4 สอบถามข้อเสนอแนะเพื่อการปรับปรุงและพัฒนาในเรื่องที่กำลังทำวิจัย ลักษณะคำถามเป็นคำถามปลายเปิด (Open End)

ตัวอย่างแบบสอบถาม

ส่วนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

โปรดเขียนเครื่องหมาย ✓ ลงในช่อง ❒ หน้าข้อความที่ตรงกับท่านมากที่สุด

1. เพศ ❒ ชาย ❒ หญิง

2. อายุ ❒ 20 – 30 ปี ❒ 31 – 40 ปี ❒ 41 – 50 ปี
❒ ตั้งแต่ 51 ปีขึ้นไป

3. ระดับการศึกษา ❒ ประถมศึกษา ❒ มัธยมศึกษาตอนต้น หรือเทียบเท่า
❒ มัธยมศึกษาตอนปลาย หรือเทียบเท่า ❒ อนุปริญญา หรือเทียบเท่า
❒ ปริญญาตรี ❒ สูงกว่าปริญญาตรี

4. สถานภาพสมรส ❒ โสด ❒ สมรส ❒ หย่าร้าง ❒ ม่าย

ตัวอย่างแบบสอบถาม

โปรดกาเครื่องหมาย ✓ ในช่องซึ่งแสดงถึงระดับความสำคัญของแต่ละปัจจัย

ปัจจัย	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
1. ด้านผลิตภัณฑ์					
1.1 ประเภทสินเชื่อมีความหลากหลาย					
1.2 วงเงินที่ได้รับอนุมัติตรงกับความต้องการ					
1.3 ธนาคารมีความน่าเชื่อถือ					
2. ด้านราคา					
2.1 อัตราดอกเบี้ย					
2.2 อัตราค่าธรรมเนียม					
2.3 ระยะเวลาของการชำระคืน					

การสร้างรหัสสำหรับข้อมูล

ตัวแปร

ตัวแปร คือ ชื่อที่ใช้เรียกแทนข้อคำถามในเครื่องมือที่เก็บข้อมูล มักจะตั้งชื่อตัวแปรเป็นภาษาอังกฤษ และมีความยาวไม่เกิน 8 ตัวอักขระ เพื่อให้โปรแกรม SPSS สามารถเข้าใจได้

ชนิดของตัวแปร แบ่งออกเป็น 2 ชนิด คือ

1. ตัวแปรเชิงปริมาณ คือ ตัวแปรที่มีค่าเป็นตัวเลข ที่ระบุได้ว่ามีค่ามากหรือน้อย เช่น รายได้ อายุ น้ำหนัก ส่วนสูง ขนาดของตัวแปร รายได้มี 6 หลัก อายุมี 2 หลัก น้ำหนักมี 3 หลัก ส่วนสูงมี 3 หลัก เป็นต้น

2. ตัวแปรเชิงคุณภาพหรือตัวแปรเชิงกลุ่ม คือ ตัวแปรที่เป็นข้อความ หรือตัวแปรที่ต้องใช้ตัวเลข แทนค่ารหัสของข้อมูล ซึ่งขนาดของตัวแปร ควรจะเท่ากับจำนวนทางเลือกของคำตอบ เช่น เพศ จำแนกเป็น 2 กลุ่ม คือ เพศชาย และเพศหญิง ระดับความคิดเห็น มี 5 ระดับ ขนาดของตัวแปรกำหนด เป็น 1 หลัก แต่ถ้าระดับความคิดเห็นมี 10 ระดับ ขนาดของตัวแปรควรกำหนดเป็น 2 หลัก เป็นต้น

การกำหนดรหัสโดยแบ่งตามชนิดของคำถาม

การกำหนดรหัสของข้อมูลจะต้องคำนึงถึงชนิดของคำถาม โดยชนิดของคำถามแบ่งออกเป็น

1. คำถามปลายปิด แบ่งออกเป็น

- คำถามที่มีคำตอบให้เลือกเพียง 2 คำตอบ

เช่น เพศ มี 1) เพศชาย และ 2) เพศหญิง

- คำถามที่มีคำตอบให้เลือกหลายคำตอบ

เช่น ประเภทบุคลากร มี 1) ข้าราชการ 2) พนักงานมหาวิทยาลัย
3) พนักงานราชการ 4) อื่น ๆ

- คำถามที่สามารถเลือกคำตอบได้หลายคำตอบ หรือตอบได้มากกว่า 1 ข้อ

เช่น ช่วงระยะเวลาที่เลือกซื้อผลิตภัณฑ์ โดยถ้าใช้งานข้อไหน ให้ใส่ตัวเลขเป็น 1 ถ้าไม่ตอบให้ใส่ 0

การกำหนดรหัสโดยแบ่งตามชนิดของคำถาม

- คำถามที่คำตอบให้ใส่ลำดับที่ เช่น ผลิตภัณฑ์ที่ใช้งานบ่อยที่สุด 3 อันดับแรก โดยจะใส่ค่า 1 หน้าผลิตภัณฑ์ที่ใช้งานบ่อยที่สุด และจะใส่ค่า 2 และ 3 หน้าผลิตภัณฑ์ที่ใช้งานบ่อยในลำดับถัดมา ถ้าผลิตภัณฑ์ไหนที่ไม่ได้เลือกให้ใส่ข้อมูลเป็น 0
- คำถามที่ให้แสดงระดับมากน้อย เช่น ระดับความพึงพอใจ แบ่งออกเป็น 5 ระดับ คือ 1 แทนด้วย พึงพอใจน้อยที่สุด 2 พึงพอใจน้อย 3 พึงพอใจปานกลาง 4 พึงพอใจมาก 5 พึงพอใจมากที่สุด

2. คำถามปลายเปิด เช่น ข้อเสนอแนะ โปรแกรม SPSS จะสามารถพิมพ์ข้อความได้ยาวไม่เกิน 255 อักขระ และถ้าพิมพ์เป็นข้อความจะไม่สามารถนำมาประมวลผลได้ จึงจำเป็นต้องสรุปข้อคำตอบนั้น เป็นรหัสตัวเลขแทนข้อความ

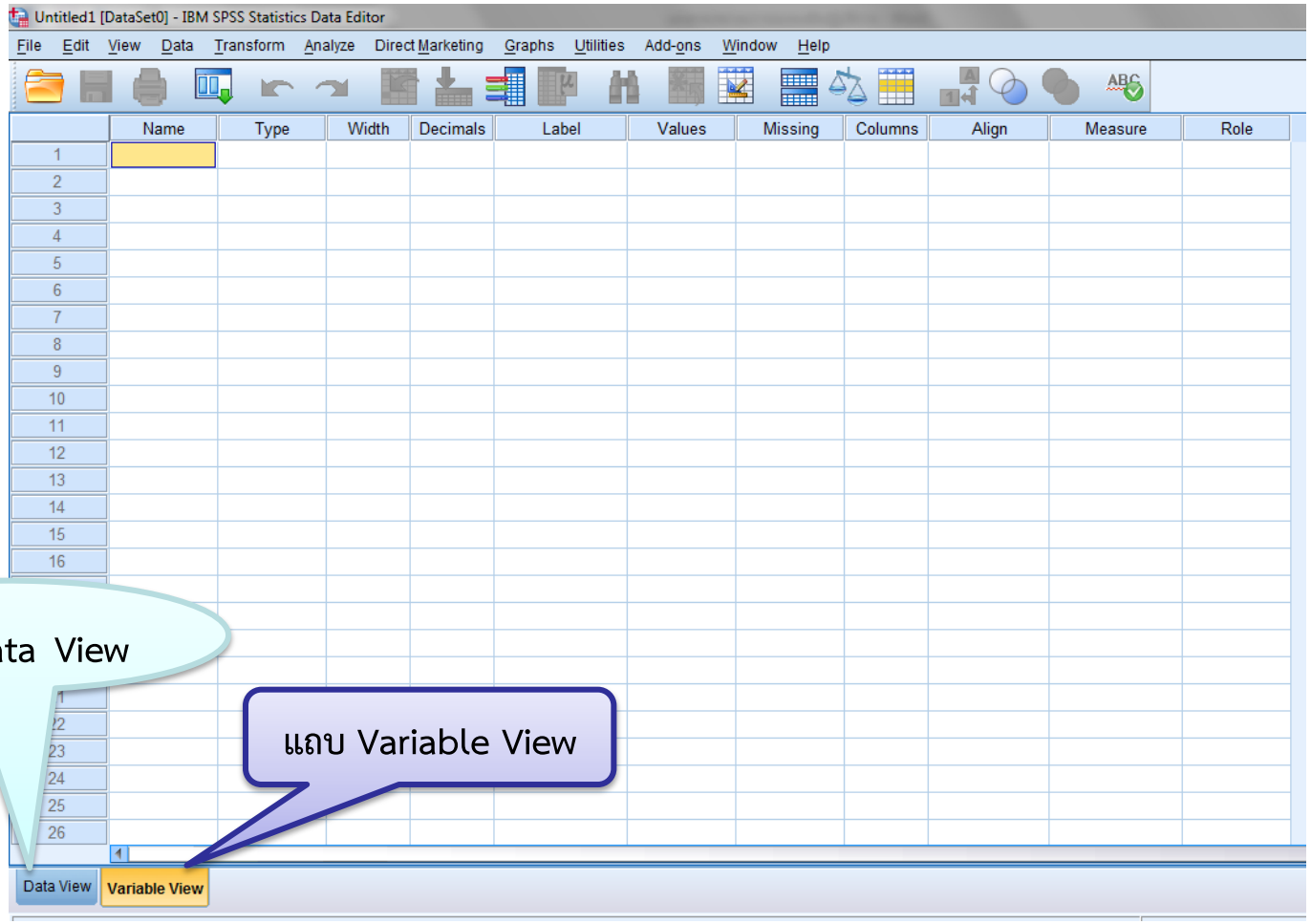
3. คำถามที่ไม่ได้รับคำตอบ (Missing data) จะแทนค่าด้วย 9, 99, 999 ... ขึ้นอยู่กับขนาดของตัวแปรว่าใช้กี่หลัก

การสร้างแฟ้มข้อมูลด้วย SPSS

ส่วนประกอบของ Data Editor

Data Editor ประกอบด้วย 2 หน้าจอ โดยหน้าจอแรก คือ Data View เป็นส่วนใช้ป้อนและแสดงข้อมูลของตัวแปรต่าง ๆ และหน้าจอที่ 2 คือ Variable View เป็นส่วนใช้กำหนดตัวแปรและรายละเอียดต่าง ๆ ทั้งนี้สามารถเลือกเปิดหน้าจอใดหน้าจอหนึ่งของ Data Editor เพื่อกำหนดให้เป็นหน้าจอทำงานตามต้องการ โดยการคลิกที่แถบ Data View หรือ Variable View ซึ่งอยู่มุมล่างด้านซ้ายของ Data Editor

การสร้างแฟ้มข้อมูลด้วย SPSS



การสร้างแฟ้มข้อมูลด้วย SPSS

- **หน้าจอ Data View ของ Data Editor** มีลักษณะสำคัญ ดังนี้

1. แต่ละแถว คือ ชุดข้อมูล 1 ชุด ค่าที่ปรากฏในแถว 1 แถว ถูกแปลงมาจากแบบสอบถาม 1 ชุด ทั้งนี้แต่ละหัวแถวมีหมายเลขแถวแสดงลำดับของชุดข้อมูลใน Data Editor

2. แต่ละสดมภ์ คือ ตัวแปรหนึ่งตัว

การสร้างแฟ้มข้อมูลด้วย SPSS

- **หน้าจอ Variable View ของ Data Editor** ประกอบด้วย 11 สดมภ์ ดังนี้

1. Name ใช้กำหนดชื่อของตัวแปรหรือสัญลักษณแทนตัวแปรนั้น ๆ ความยาวไม่เกิน 8 ตัวอักษร

2. Type ใช้กำหนดชนิดของตัวแปร ได้แก่ Numeric เปนขอมูลที่เป็นตัวเลขและ String เปนขอมูลที่เปนตัวอักษร

3. Width ใช้กำหนดความกวางของตัวแปรหรือจำนวนอักขระหรืออักษรที่ตองการใหใสไดใน Values

4. Decimals ใช้กำหนดจำนวนจุดทศนิยมของแต่ละตัวแปร

5. Labels ใช้คำอธิบายตัวแปรหรือชื่อเต็มของตัวแปรนั้น ๆ จะใชในกรณีที่ผูวิจัยกำหนดชื่อตัวแปรใน column Name เปนอักษรยอแลวตองการอธิบายหรือขยายความไว เชน Name ระบุเปน ID ใน Labels จะระบุเปนลำดับที่ หรือ Name ระบุเป็น Salary ใน Labels จะระบุเป็นรายไดตอเดือน เปนตน

การสร้างแฟ้มข้อมูลด้วย SPSS

6. Values ใช้การกำหนดคาใหกับตัวแปร เชน ตัวแปรเพศ กำหนดให เพศชาย มีคาเทากับ 1 และเพศหญิง มีคาเทากับ 2 เปนตน

7. Missing ใช้กำหนดค่าสูญหายคาของขอมูลหรือแบบสอบถามที่ไม่สมบูรณ์หรือไม่มีคำตอบ

8. Columns ใช้กำหนดความกว้างของ Columns เฉพาะในหน้าจอ Data View

9. Align ใช้กำหนดลักษณะการวางข้อมูลว่าจัดให้ชิดซ้าย ขวา หรืออยู่ตรงกลาง

10. Measures ใช้กำหนดมาตรการวัดของข้อมูล ได้แก่ Nominal, Ordinal และ Scale (หมายถึง Interval และ Ratio)

11. Role เป็นการแสดงบทบาทของตัวแปร

การวิเคราะห์ข้อมูล

หลังจากที่ตรวจสอบความถูกต้องของแบบสอบถามเรียบร้อยแล้ว จะนำข้อมูลในแบบสอบถามมาเปลี่ยนแปลงเป็นรหัสตัวเลข (Code) แล้วบันทึกรหัสลงในเครื่องคอมพิวเตอร์ ด้วยโปรแกรม SPSS

ลักษณะของการวิเคราะห์ข้อมูลจะมีข้อสังเกตดังนี้

– แบบสอบถามเกี่ยวกับสถานภาพทั่วไป ใช้วิธีการหาค่าความถี่ (Frequency) โดยสรุปออกมาเป็นค่าร้อยละ (%)

– แบบสอบถามที่เป็นการสำรวจทัศนคติเกี่ยวกับความพึงพอใจ ใช้วิธีการหาค่าเฉลี่ย ($\overline{X}$) และส่วนเบี่ยงเบนมาตรฐาน (S.D.)

– แบบสอบถามที่เป็นข้อเสนอแนะ ใช้วิธีการหาความถี่ (Frequency)

ความหมายของค่าสถิติ

- Mean คือ ค่าเฉลี่ย
- Median คือ ค่ามัธยฐาน
- Mode คือ ฐานนิยม
- Sum คือ ผลรวม
- Std.deviation คือ ส่วนเบี่ยงเบนมาตรฐาน
- Variance คือ ค่าความแปรปรวน
- Range คือ ค่าพิสัย
- Minimum คือ ค่าต่ำสุด
- Maximum คือ ค่าสูงสุด
- S.E.Mean คือ ค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย
- Skewness คือ ค่าความเบ้
- Kurtosis คือ ค่าความโด่ง

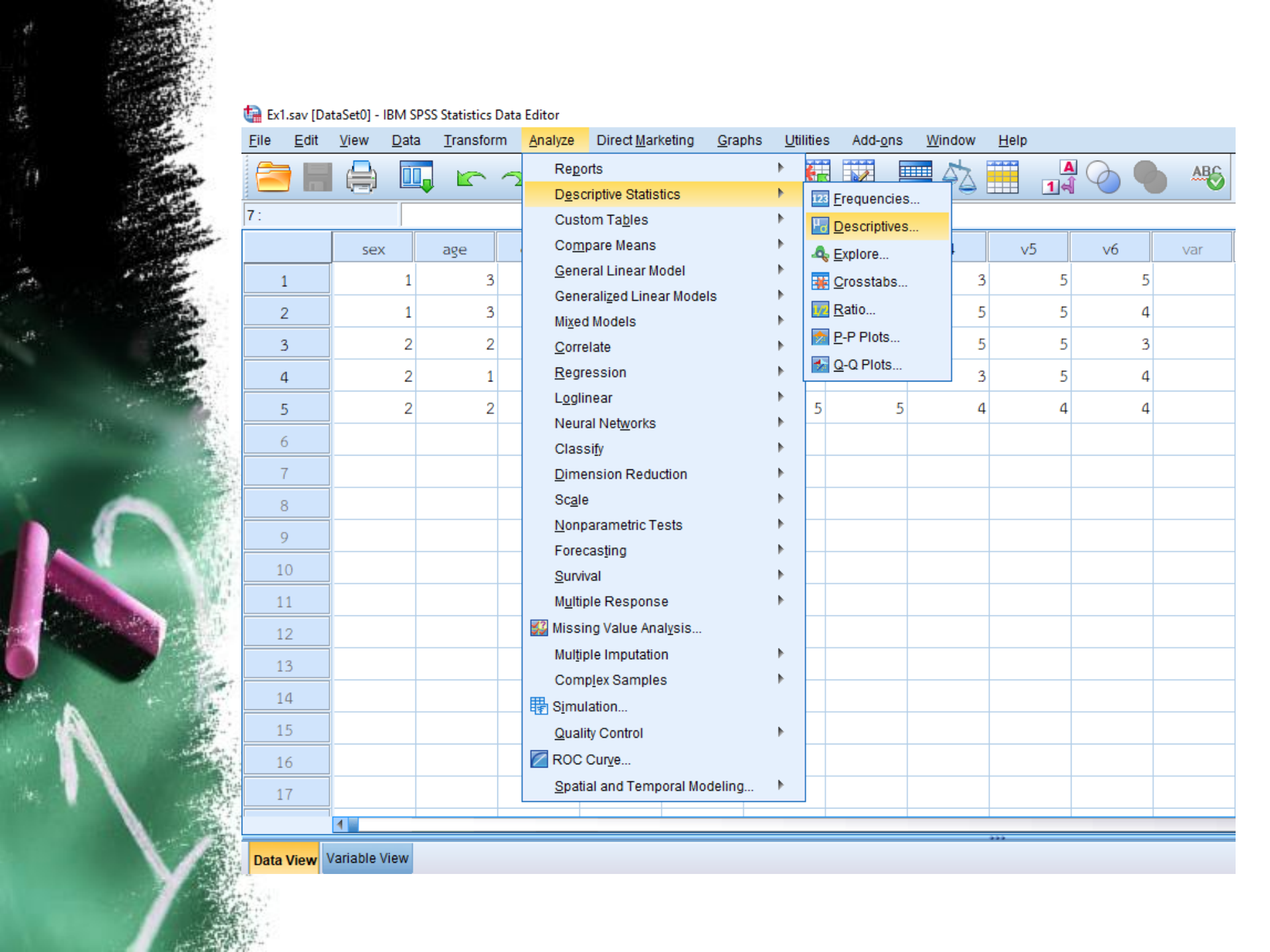
การใช้คำสั่ง Descriptive Statistics

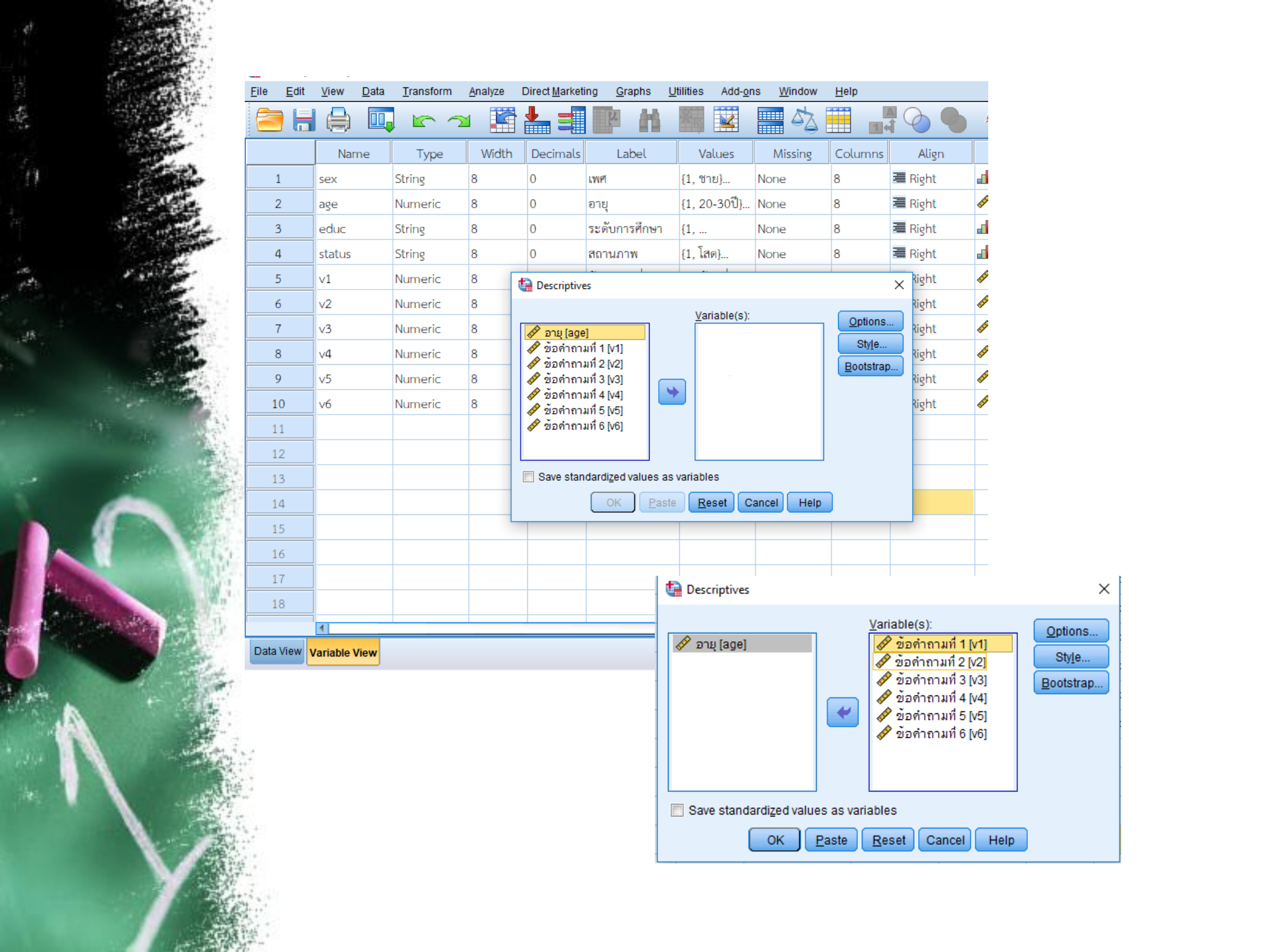
คำสั่ง Descriptive

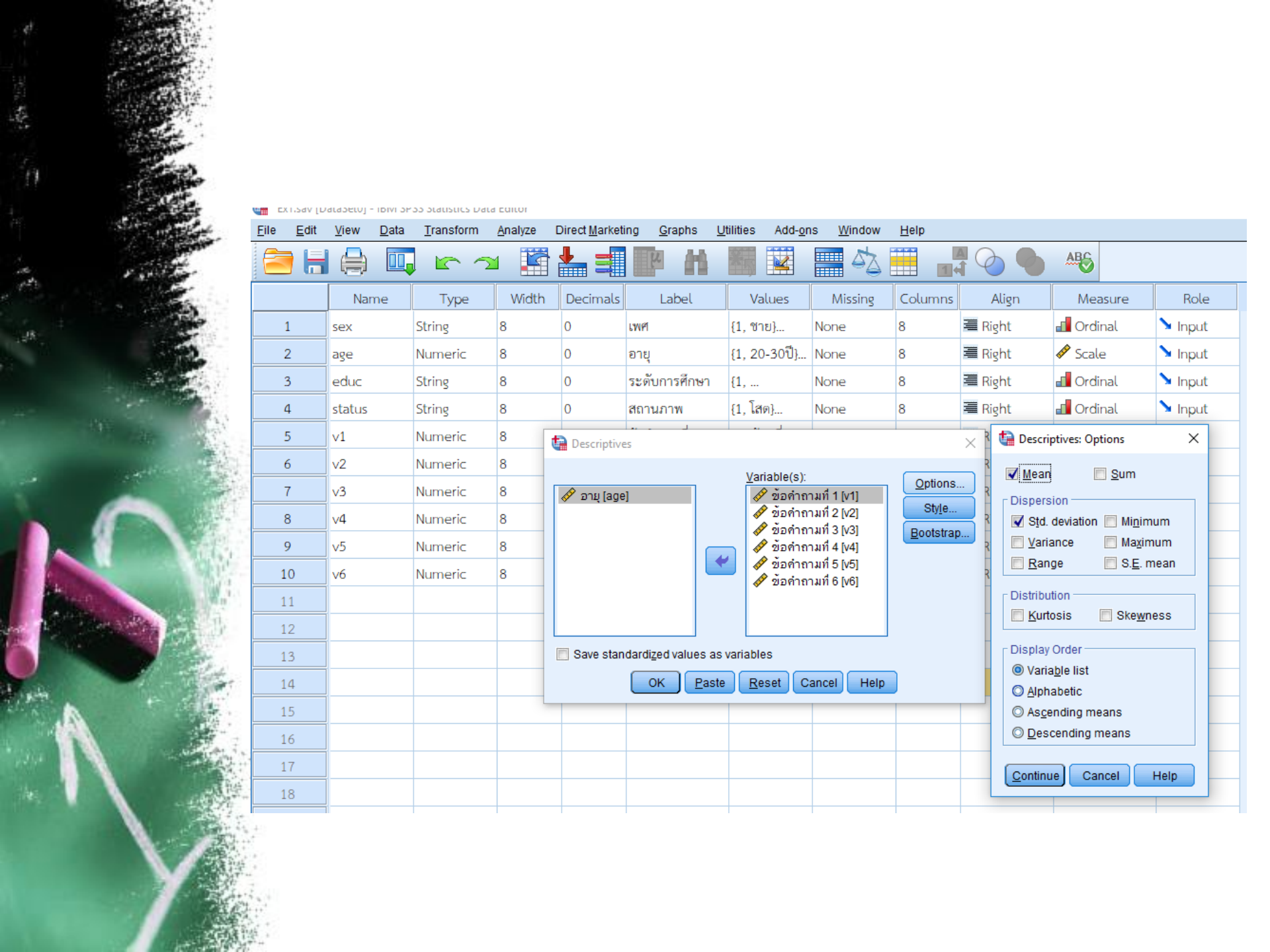
เป็นคำสั่ง ในการอธิบายตัวแปรเชิงปริมาณ ตามสถิติที่ต้องการวิเคราะห์ เช่น Mean, Std. Deviation เช่น หาค่าเฉลี่ยของตัวแปรอายุ เป็นต้น

เลือกเมนู

Analyze -> Descriptive Statistics -> Descriptives


Ex1.sav [DataSet0] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
7 :
sex age v5 v6 var
1 1 3 3 5 5
2 1 3 5 5 4
3 2 2 5 5 3
4 2 1 3 5 4
5 2 2 5 5 4 4 4
6
7
8
9
10
11
12
13
14
15
16
17
Data View
Variable View


File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align
1 sex String 8 0 เพศ {1, ชาย}... None 8 Right
2 age Numeric 8 0 อายุ {1, 20-30ปี}... None 8 Right
3 educ String 8 0 ระดับการศึกษา {1, ... None 8 Right
4 status String 8 0 สถานภาพ {1, โสด}... None 8 Right
5 v1 Numeric 8
6 v2 Numeric 8
7 v3 Numeric 8
8 v4 Numeric 8
9 v5 Numeric 8
10 v6 Numeric 8
Descriptives
อายุ [age]
ข้อคำถามที่ 1 [v1]
ข้อคำถามที่ 2 [v2]
ข้อคำถามที่ 3 [v3]
ข้อคำถามที่ 4 [v4]
ข้อคำถามที่ 5 [v5]
ข้อคำถามที่ 6 [v6]
Variable(s):
Options...
Style...
Bootstrap...
Save standardized values as variables
OK Paste Reset Cancel Help
Data View
Variable View
Descriptives
อายุ [age]
Variable(s):
ข้อคำถามที่ 1 [v1]
ข้อคำถามที่ 2 [v2]
ข้อคำถามที่ 3 [v3]
ข้อคำถามที่ 4 [v4]
ข้อคำถามที่ 5 [v5]
ข้อคำถามที่ 6 [v6]
Options...
Style...
Bootstrap...
Save standardized values as variables
OK Paste Reset Cancel Help


File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 sex String 8 0 เพศ {1, ชาย}... None 8 Right Ordinal Input
2 age Numeric 8 0 อายุ {1, 20-30ปี}... None 8 Right Scale Input
3 educ String 8 0 ระดับการศึกษา {1, ... None 8 Right Ordinal Input
4 status String 8 0 สถานภาพ {1, โสด}... None 8 Right Ordinal Input
5 v1 Numeric 8
6 v2 Numeric 8
7 v3 Numeric 8
8 v4 Numeric 8
9 v5 Numeric 8
10 v6 Numeric 8
Descriptives
อายุ [age]
Variable(s):
ข้อคำถามที่ 1 [v1]
ข้อคำถามที่ 2 [v2]
ข้อคำถามที่ 3 [v3]
ข้อคำถามที่ 4 [v4]
ข้อคำถามที่ 5 [v5]
ข้อคำถามที่ 6 [v6]
Options...
Style...
Bootstrap...
Save standardized values as variables
OK Paste Reset Cancel Help
Descriptives: Options
Mean Sum
Dispersion
Std. deviation Minimum
Variance Maximum
Range S.E. mean
Distribution
Kurtosis Skewness
Display Order
Variable list
Alphabetic
Ascending means
Descending means
Continue Cancel Help

จะปรากฏผลลัพธ์ดังภาพ

Descriptive Statistics

	N	Mean	Std. Deviation
ข้อคำถามที่ 1	5	4.20	.837
ข้อคำถามที่ 2	5	4.20	.837
ข้อคำถามที่ 3	5	4.60	.548
ข้อคำถามที่ 4	5	4.00	1.000
ข้อคำถามที่ 5	5	4.80	.447
ข้อคำถามที่ 6	5	4.00	.707
Valid N (listwise)	5		

เกณฑ์ที่ใช้ในการแปลผลค่าเฉลี่ย ใช้เกณฑ์การแปลผลตามแนวคิดของเบสท์ (Best W. John. 1997) ดังนี้

ค่าเฉลี่ย 4.50 - 5.00 หมายถึง ระดับความ**พึงพอใจมากที่สุด**

ค่าเฉลี่ย 3.50 - 4.49 หมายถึง ระดับความ**พึงพอใจมาก**

ค่าเฉลี่ย 2.50 - 3.49 หมายถึง ระดับความ**พึงพอใจปานกลาง**

ค่าเฉลี่ย 1.50 - 2.49 หมายถึง ระดับความ**พึงพอใจน้อย**

ค่าเฉลี่ย 1.00 - 1.49 หมายถึง ระดับความ**พึงพอใจน้อยที่สุด**

ตัวอย่างการนำเสนอค่าต่าง ๆ และการแปลผล

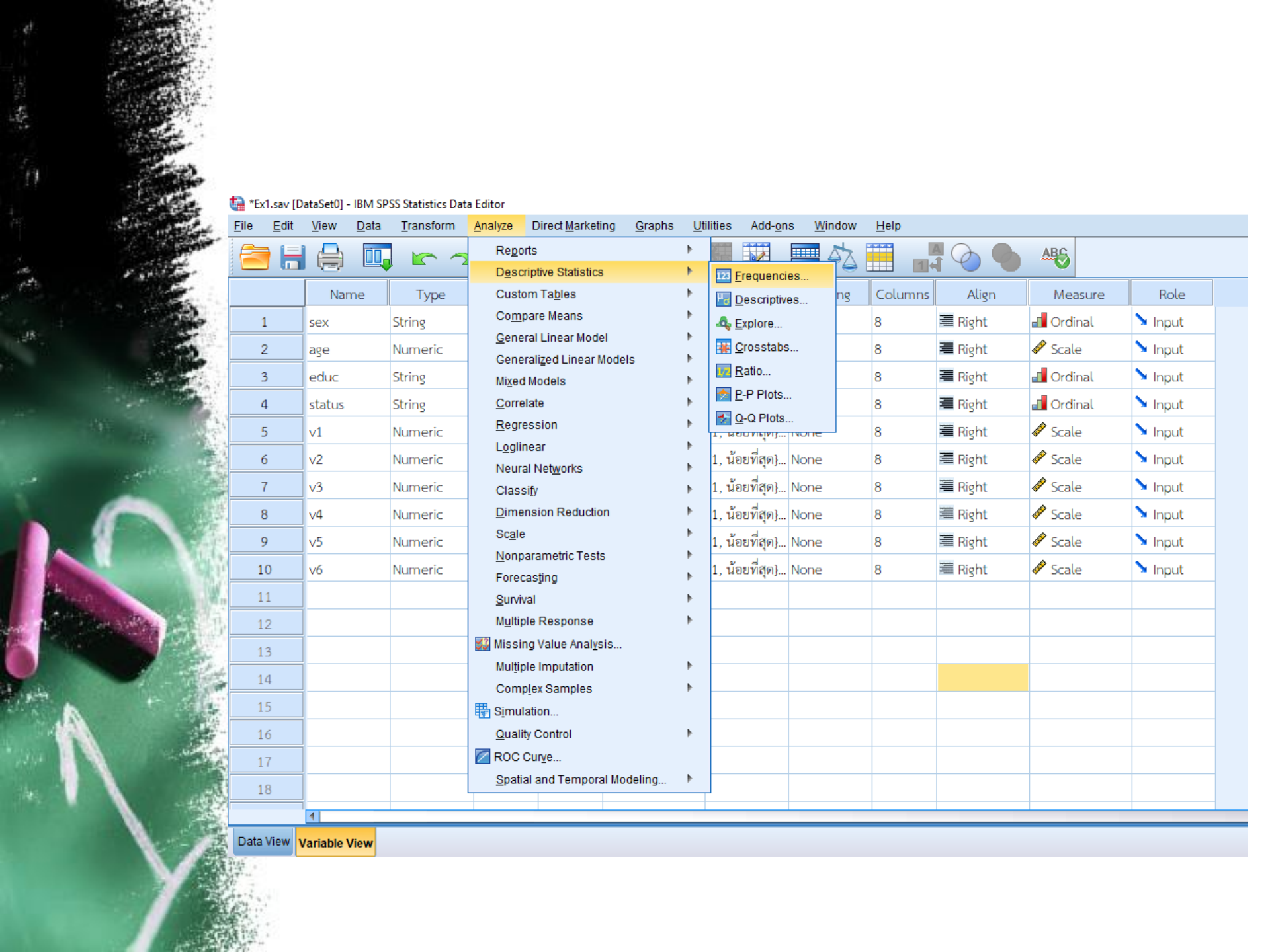
ข้อคำถามที่	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ความหมาย
ข้อคำถามที่ 1	4.20	0.837	มาก
ข้อคำถามที่ 2	4.20	0.837	มาก
ข้อคำถามที่ 3	4.60	0.548	มากที่สุด
ข้อคำถามที่ 4	4.00	1.000	มาก
ข้อคำถามที่ 5	4.80	0.447	มากที่สุด
ข้อคำถามที่ 6	4.00	0.707	มาก

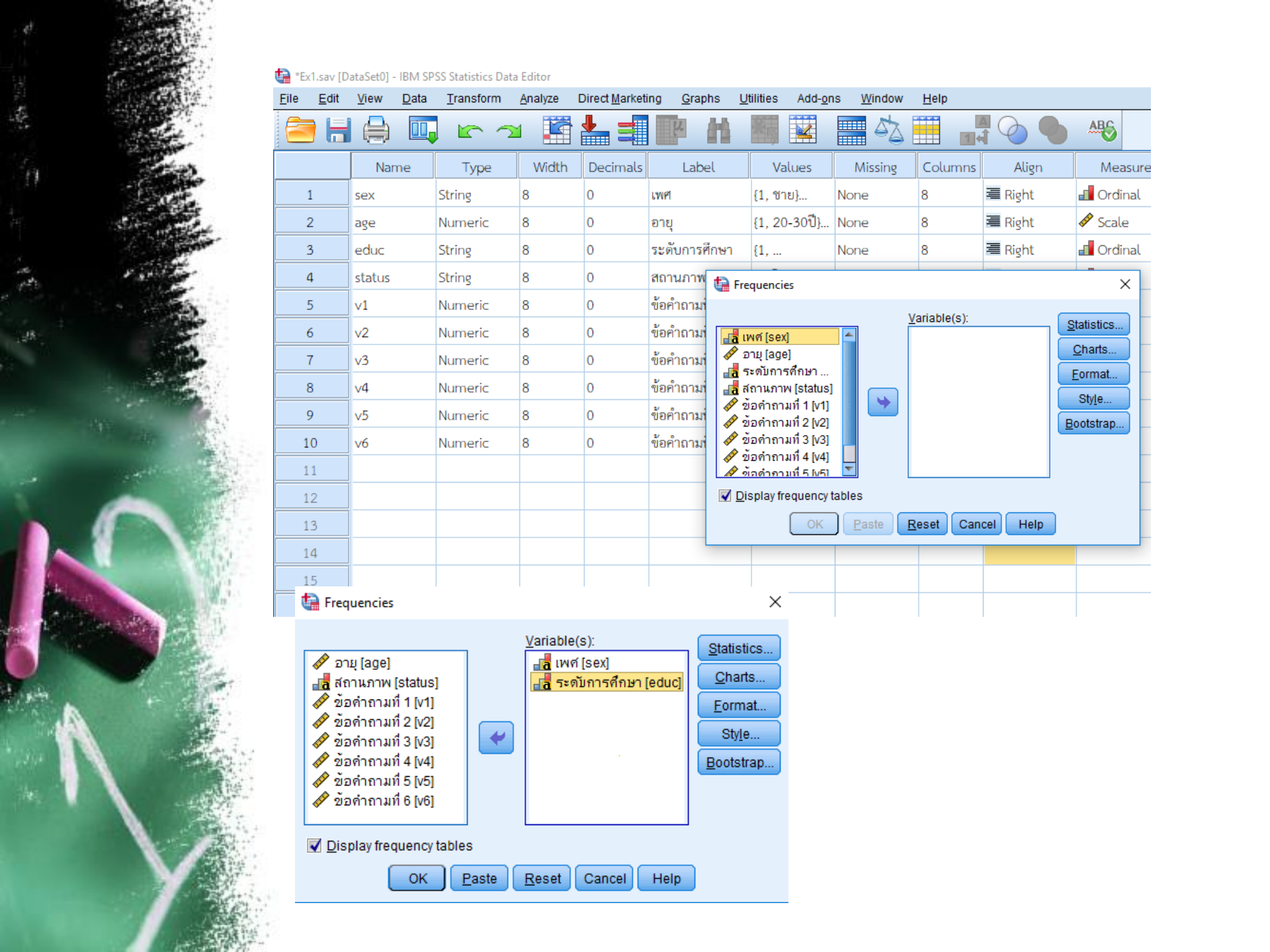
คำสั่ง Frequencies

เป็นคำสั่ง ในการอธิบายทั้งตัวแปรเชิงปริมาณ และตัวแปรเชิงกลุ่ม โดยการแจกแจงความถี่และค่าร้อยละแบบทางเดียว

เลือกเมนู

Analyze -> Descriptive Statistics -> Frequencies


*Ex1.sav [DataSet0] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
Name Type Columns Align Measure Role
1 sex String 8 Right Ordinal Input
2 age Numeric 8 Right Scale Input
3 educ String 8 Right Ordinal Input
4 status String 8 Right Ordinal Input
5 v1 Numeric 8 Right Scale Input
6 v2 Numeric 1, น้อยที่สุด}... None 8 Right Scale Input
7 v3 Numeric 1, น้อยที่สุด}... None 8 Right Scale Input
8 v4 Numeric 1, น้อยที่สุด}... None 8 Right Scale Input
9 v5 Numeric 1, น้อยที่สุด}... None 8 Right Scale Input
10 v6 Numeric 1, น้อยที่สุด}... None 8 Right Scale Input
11
12
13
14
15
16
17
18
Data View
Variable View


*Ex1.sav [DataSet0] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure
1 sex String 8 0 เพศ {1, ชาย}... None 8 Right Ordinal
2 age Numeric 8 0 อายุ {1, 20-30ปี}... None 8 Right Scale
3 educ String 8 0 ระดับการศึกษา {1, ... None 8 Right Ordinal
4 status String 8 0 สถานภาพ
5 v1 Numeric 8 0 ข้อคำถามที่
6 v2 Numeric 8 0 ข้อคำถามที่
7 v3 Numeric 8 0 ข้อคำถามที่
8 v4 Numeric 8 0 ข้อคำถามที่
9 v5 Numeric 8 0 ข้อคำถามที่
10 v6 Numeric 8 0 ข้อคำถามที่
Frequencies
เพศ [sex]
อายุ [age]
ระดับการศึกษา ...
สถานภาพ [status]
ข้อคำถามที่ 1 [v1]
ข้อคำถามที่ 2 [v2]
ข้อคำถามที่ 3 [v3]
ข้อคำถามที่ 4 [v4]
Variable(s):
Statistics...
Charts...
Format...
Style...
Bootstrap...
Display frequency tables
OK Paste Reset Cancel Help
Frequencies
อายุ [age]
สถานภาพ [status]
ข้อคำถามที่ 1 [v1]
ข้อคำถามที่ 2 [v2]
ข้อคำถามที่ 3 [v3]
ข้อคำถามที่ 4 [v4]
ข้อคำถามที่ 5 [v5]
ข้อคำถามที่ 6 [v6]
Variable(s):
เพศ [sex]
ระดับการศึกษา [educ]
Statistics...
Charts...
Format...
Style...
Bootstrap...
Display frequency tables
OK Paste Reset Cancel Help

จะปรากฏผลลัพธ์ดังภาพ

เพศ

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ชาย	2	40.0	40.0	40.0
	หญิง	3	60.0	60.0	100.0
	Total	5	100.0	100.0	

จากตารางเพศ จะเห็นว่ามีเพศชาย 2 คน คิดเป็นร้อยละ 40 และมีเพศหญิง - 3 คน คิดเป็นร้อยละ 60

ระดับการศึกษา

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	มัธยมศึกษาตอนปลาย หรือเทียบเท่า	1	20.0	20.0	20.0
	อนุปริญญา หรือเทียบเท่า	1	20.0	20.0	40.0
	ปริญญาตรี	2	40.0	40.0	80.0
	สูงกว่าปริญญาตรี	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

จากตารางระดับการศึกษา จะเห็นว่าระดับการศึกษาตอนปลายหรือเทียบเท่า 1 คน คิดเป็นร้อยละ 20 ระดับอนุปริญญาหรือเทียบเท่า 1 คน คิดเป็นร้อยละ 20 ระดับปริญญาตรี 2 คน คิดเป็นร้อยละ 40 และสูงกว่าระดับปริญญาตรี 1 คน คิดเป็นร้อยละ 20