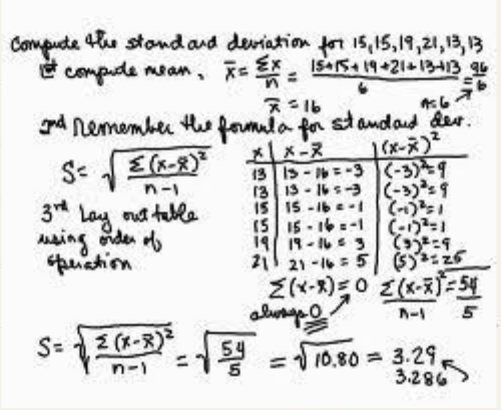
04-01-201 สถิติทางธุรกิจ (Business Statistics)

ภาคเรียนที่ 2 ปีการศึกษา 2558



ผู้สอน อาจารย์สุมาลี สมนึก

คำอธิบายรายวิชา

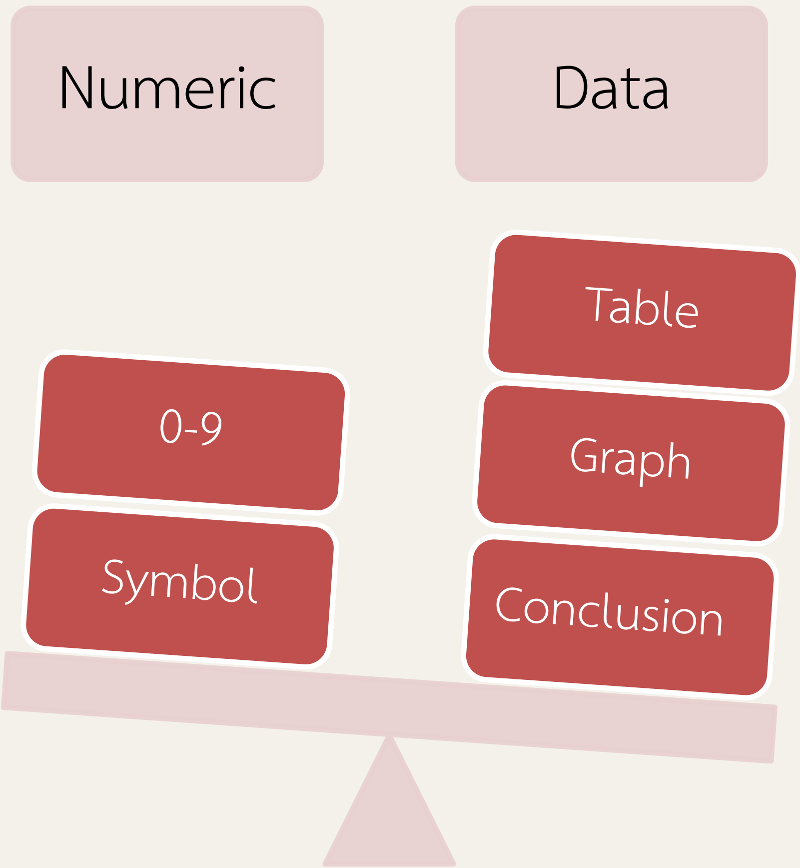
ระเบียบวิธีสถิติและการวิเคราะห์ข้อมูลเบื้องต้น ตัวแปรสุ่มและการแจกแจง ความน่าจะเป็นของตัวแปรสุ่ม การประมาณค่า และการทดสอบสมมติฐาน การทดสอบไคร์สแคว การวิเคราะห์ความแปรปรวน การวิเคราะห์การถดถอยและสหสัมพันธ์ การวิเคราะห์อนุกรมเวลา

What is Statistics?

Why do you know it ?

When do you use it?

Where do you use it ?



สถิติกับการประยุกต์ทางธุรกิจ

1 สถิติกับการประยุกต์ทางธุรกิจ
2 การนำสถิติไปประยุกต์ทางธุรกิจ
3 การเก็บรวบรวมข้อมูล
4 การนำเสนอข้อมูล

ความรู้ทั่วไปเกี่ยวกับสถิติ

ความหมายของสถิติ

ตัวเลข หรือ ข้อมูลที่ใช้แทนข้อความจริงเกี่ยวกับเรื่องต่างๆ ตัวเลข หรือ ข้อมูลเกี่ยวกับเรื่องต่างๆทางธุรกิจที่พบเห็นกันอยู่เสมอๆเป็นประจำ เช่น

- ราคาน้ำมันเชื้อเพลิงต่างๆของแต่ละยี่ห้อในแต่ละครั้งที่มีการเปลี่ยนแปลงราคา
- อัตราการแลกเปลี่ยนเงินตราต่างประเทศสกุลเงินที่เทียบกับเงินบาทรายวัน
- ดัชนีราคาหลักทรัพย์และมูลค่าการซื้อขายรายวัน
- ราคาขายส่งขายปลีกรายวันของสินค้าเกษตร
- ต้นทุนการผลิต ปริมาณการผลิต ยอดขายของสินค้าหรือบริการแต่ละชนิด

ความรู้ทั่วไปเกี่ยวกับสถิติ

ข้อมูลสารสนเทศ

เมื่อนำข้อมูลที่มีอยู่แล้ว หรือ ที่เก็บรวบรวมได้ใหม่มาวิเคราะห์ด้วยวิธีการวิเคราะห์ที่เหมาะสม จะสามารถนำไปใช้ในการตัดสินใจและวางแผนด้านต่างๆได้ตามวัตถุประสงค์ที่ต้องการ ซึ่งส่วนใหญ่จะดีกว่าการใช้ข้อมูลที่เกี่ยวข้องกับเรื่องที่ต้องตัดสินใจและวางแผนเพียงอย่างเดียว เช่น

- การตัดสินใจและวางแผนเกี่ยวกับการเลือกชนิดน้ำมันเชื่อเพลิงมาใช้ในการผลิตสินค้า หรือ ใช้ในการขนส่งสินค้า

ข้อมูล + **วิเคราะห์เชิงสถิติ** → **สารสนเทศ**

ตัวอย่างความสัมพันธ์ระหว่างข้อมูล วิธีวิเคราะห์ข้อมูลเชิงสถิติ และสารสนเทศ

ข้อมูล	วิธีวิเคราะห์ข้อมูลเชิงสถิติ	สารสนเทศ
อายุของลูกค้ากลุ่มเป้าหมายแต่ละคน	วิธีหาค่าเฉลี่ย (ค่าเฉลี่ยเลขคณิต)	อายุเฉลี่ยของลูกค้าเป้าหมาย
	วิธีการแจกแจงความถี่	จำนวนและร้อยละของลูกค้ากลุ่มเป้าหมาย จำแนกตามวัย (วัยเด็ก วัยรุ่น วัยผู้ใหญ่ วัยสูงอายุ)

ตัวอย่างความสัมพันธ์ระหว่างข้อมูล

วิธีวิเคราะห์ข้อมูลเชิงสถิติ และสารสนเทศ

ข้อมูล	วิธีวิเคราะห์ข้อมูลเชิงสถิติ	สารสนเทศ
สื่อที่ลูกค้ากลุ่มเป้าหมายแต่ละคนรับทราบเกี่ยวกับตัวสินค้า	วิธีหาค่าเฉลี่ยแจกแจงความถี่	จำนวนและร้อยละของลูกค้ากลุ่มเป้าหมาย จำแนกตามสื่อที่รับทราบ
	วิธีทดสอบไคสแควร์	ความสัมพันธ์ระหว่างเพศ อายุ ระดับการศึกษา รายได้ ของลูกค้ากับชนิดของสื่อที่รับทราบ

ตัวอย่างความสัมพันธ์ระหว่างข้อมูล วิธีวิเคราะห์ข้อมูลเชิงสถิติ และสารสนเทศ

ข้อมูล	วิธีวิเคราะห์ข้อมูลเชิงสถิติ	สารสนเทศ
ค่าใช้จ่ายรายเดือนของการใช้โทรศัพท์มือถือของนักศึกษากับผู้ที่อยู่ในวัยทำงาน	วิธีทดสอบสมมติฐานเชิงสถิติ	กลุ่มนักศึกษาหรือกลุ่มผู้ที่อยู่ในวัยทำงานมีค่าใช้จ่ายรายเดือนของการใช้โทรศัพท์มือถือมากกว่ากัน

ตัวอย่างความสัมพันธ์ระหว่างข้อมูล วิธีวิเคราะห์ข้อมูลเชิงสถิติ และสารสนเทศ

ข้อมูล	วิธีวิเคราะห์ข้อมูลเชิงสถิติ	สารสนเทศ
ยอดขายสินค้าหรือบริการรายเดือน	วิธีวิเคราะห์อนุกรมเวลา	แนวโน้มยอดขายสินค้าหรือบริการในอนาคต
		ดัชนีฤดูการรายเดือนของยอดขายสินค้าหรือบริการ

เนื้อหาสถิติที่สำคัญ

นำไปใช้ในการตัดสินใจและการวางแผนงานด้านต่างๆ

ส่วนของวิธีการเก็บรวบรวมข้อมูล
ส่วนของวิธีการวิเคราะห์ข้อมูล
ส่วนของวิธีการนำเสนอข้อมูล

Link mind mapping

ประโยชน์ของสถิติ

- การใช้สถิติเพื่อการวางแผนและพัฒนาเศรษฐกิจของประเทศ
- การใช้สถิติเพื่อการวางแผนการทดลอง
- การใช้สถิติเพื่อการประกันภัยและประกันชีวิต
- การใช้สถิติเพื่อหาคำตอบของศาสตร์แขนงอื่นๆ
- ใช้ในการตรวจสอบและควบคุมคุณภาพ
- ใช้ในการวิจัยตลาด
- ใช้ในด้านการเงินการธนาคาร

การนำความรู้ทางสถิติไปประยุกต์ทางธุรกิจ

ข้อมูลทางธุรกิจ

- ข้อมูลปฐมภูมิ (Primary Data)
- ข้อมุลทุติยภูมิ (Secondary Data)

การวิเคราะห์ข้อมูลทางธุรกิจ

- แจกแจงความถี่ (frequency)
- ร้อยละ (percentage)
- สัดส่วน (proportion)
- อัตราส่วน (ratio)
- ค่าเฉลี่ย (mean)
- การกระจาย (Dispersion)
- การประมาณค่า (estimation)
- การทดสอบสมมติฐาน (hypothesis)
- การหาความสัมพันธ์ (association)
- การพยากรณ์ (prediction)


break

การเก็บรวบรวมข้อมูล

ความหมายของข้อมูล

ตัวเลข หรือ ข้อความ ซึ่งเป็นความจริง และสามารถสื่อความหมายเกี่ยวกับขนาดได้

เช่น ....

....

....

....

....

ประเภทของข้อมูล

ข้อมูลแบ่งตามวิธีเก็บรวบรวมข้อมูล

- **ข้อมูลปฐมภูมิ (Primary data)** คือ ข้อมูลที่ผู้วิเคราะห์จะต้องเก็บรวบรวมข้อมูลจากผู้ให้ข้อมูลโดยตรง กล่าวคือ ผู้วิเคราะห์ต้องทำการสัมภาษณ์ วัด นับ หรือ สังเกตจากหน่วยที่ให้ข้อมูลที่ต้องการโดยที่ข้อมูลเหล่านี้ไม่เคยมีผู้ใดเก็บรวบรวมไว้ก่อน การเก็บรวบรวมประเภทนี้ทำได้2 วิธี คือ การทำสำมะโน (census) และการสำรวจจากตัวอย่าง (sample survey)
- **ข้อมูลทุติยภูมิ (Secondary Data)** คือ ข้อมูลที่ผู้อื่นเก็บรวบรวมไว้แล้ว ซึ่ง อาจเป็นบุคคล หรือ หน่อยงานก็ได้ การเก็บรวบรวมข้อมูลดังกล่าวก็เพื่อใช้ประโยชน์ในการบริหารงานของหน่วยงานนั้นๆ ในรูปของทะเบียบ (record) หรือ ทะเบียน (registration) หรือรายงาน (report)

ประเภทของข้อมูล

ข้อมูลแบ่งตามลักษณะของข้อมูล

- **ข้อมูลเชิงปริมาณ (Quantitative data)** เป็นข้อมูลที่ใช้แทนขนาดหรือปริมาณซึ่งวัดออกมาเป็นค่าตัวเลขที่สามารถนำมาใช้เปรียบเทียบขนาดได้โดยตรง
- **ข้อมูลเชิงคุณภาพ (Qualitative data)** เป็นข้อมูลที่ไม่สามารถวัดออกเป็นค่าตัวเลขได้โดยตรง แต่วัดออกมาในรูปจำนวน หรือ ความถี่ของข้อมูล

ประเภทของข้อมูล

ข้อมูลแบ่งตามเวลาที่ใช้ในการเก็บรวบรวมข้อมูล

- **ข้อมูลภาคตัดขวาง (Cross sectional data)** เป็นข้อมูล ณ จุดหนึ่งหรือช่วงของเวลา เช่น ณ สิ้นปี พศ. 2556 ธนาคารแห่งหนึ่งมีสาขารวมทั้งสิ้น 452 แห่ง ในจำนวนนี้เป็นสาขาในประเทศไทย 447 แห่ง และสาขาต่างประเทศ 5 แห่ง
- **ข้อมูลอนุกรมเวลา (Time series data)** เป็นข้อมูลที่มีการเปลี่ยนแปลงเมื่อเวลาเปลี่ยนไป กล่าวคือ เป็นข้อมูลภาคตัดขวางต่อเนื่องกันหลายจุดของเวลา หรือหลายช่วงเวลา เช่น สินทรัพย์รายปีของธนาคารแห่งประเทศไทยระหว่าง พศ. 2550-2556

วิธีการเก็บรวบรวมข้อมูล

- การเก็บรวบรวมข้อมูลจากทะเบียนหรือรายงาน
- การเก็บรวบรวมข้อมูลจากการสำรวจ
- การเก็บรวบรวมข้อมูลจากการสังเกต
- การเก็บรวบรวมข้อมูลจากการทดลอง

ข้อควรระวังในการเลือกวิธีเก็บรวบรวมข้อมูล

- ตรวจสอบข้อมูล
- ลักษณะหรือข้อมูลของประชากร
- การเลือกตัวอย่างแบบชั้นภูมิ
- กรอบตัวอย่าง
- การเก็บรวบรวมข้อมูลจากการสังเกต
- การเก็บรวบรวมข้อมูลจากการทดลอง
- วิธีที่ใช้ในการเก็บรวมรวบข้อมูล
- การให้ความร่วมมือจากผู้ให้ข้อมูล
- แบบสอบถาม
- ความสมบูรณ์ของการตอบแบบสอบถาม

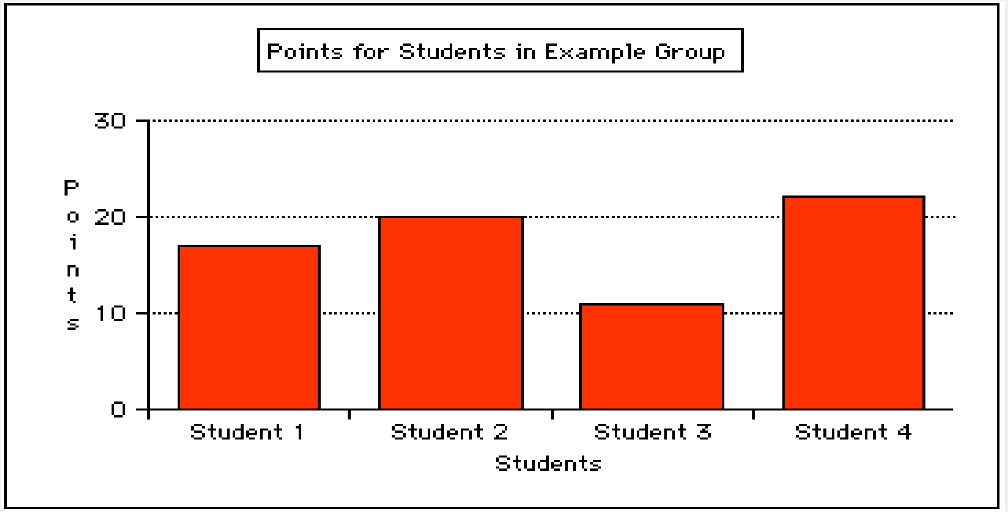
การวิเคราะห์ข้อมูลเบื้องต้น

1. การแจกแจงความถี่
2. การวัดค่ากลาง
3. การวัดการกระจาย
4. ข้อควรระวังในการเลือกวิธีวิเคราะห์สถิติพรรณนามาใช้

การวิเคราะห์สถิติพรรณนา (Descriptive statistics)

เป็นการวิเคราะห์ข้อมูลที่เก็บรวบรวมมาได้ ไม่ว่าจะเป็นข้อมูลปฐมภูมิหรือข้อมูลทุติยภูมิ เพื่อจำแนกข้อมูลตามลักษณะต่างๆที่ต้องการเปรียบเทียบ หรือเพื่อหาขนาดตัวแทน และการกระจายของข้อมูลชุดนั้นๆสำหรับนำไปใช้ในการวิเคราะห์ขั้นต่อไป

- การแจกแจงความถี่
- การหาค่ากลาง
- การวัดการกระจาย



การแจกแจงความถี่ของข้อมูล

อายุ (ปี)	จำนวนประชากร (ความถี่)
1-10	6
11-20	3
21-30	4
31-40	2
41-50	5
51-60	3
61-70	2
71-80	3
81-90	1
91-100	1
รวม	**30**

เป็นการจำแนกข้อมูลตามลักษณะต่างๆ เพื่อสะดวกต่อการเปรียบเทียบและสะดวกต่อการนำไปใช้ในการวิเคราะห์เรื่องอื่นๆต่อไป เช่น การหาค่ากลาง หรือ ค่าเฉลี่ย การหาค่าส่วนเบี่ยงเบนมาตรฐาน ฯลฯ การแจกแจงความถี่นิยมใช้กับข้อมูลที่มีอยู่เป็นจำนวนมากหรือมีค่าซ้ำกันอยู่มากเท่านั้น การแจกแจงความถี่มี 2 ชนิด คือ การแจกแจงความถี่ของค่าแต่ละค่าที่เป็นไปได้ทั้งหมด และการแจกแจงความถี่ของค่าในแต่ละช่วงหรือแต่ละอัตราภาคชั้น

การแจกแจงความถี่ของค่าแต่ละค่าที่เป็นไปได้ทั้งหมด

เป็นการจำแนกข้อมูลตามลักษณะต่างๆซึ่งมีจำนวนลักษณะไม่มากนัก เช่น การแจกแจงความถี่ของคนงานในโรงงานอุตสาหกรรมแห่งหนึ่งจำแนกตามอัตราค่าแรงงานต่อวันซึ่งมี 5 อัตรา คือ 180 บาท 200 บาท 220 บาท 250 บาท และ 300 บาท

อัตราค่าแรง	ความถี่
180 บาท	50
200 บาท	30
220 บาท	20
250 บาท	20
300 บาท	10

การแจกแจงความถี่ของค่าในแต่ละช่วงหรือแต่ละอันตรภาคชั้น

เป็นการแจกแจงความถี่โดยการแบ่งค่าที่เป็นไปได้ทั้งหมดออกเป็นช่วงๆโดยให้แต่ละช่วงประกอบด้วยค่าที่เป็นไปได้หลายๆค่า แล้วหาความถี่ของข้อมูลที่ตกอยู่ในแต่ละช่วง หรือ อัตรภาคชั้น เช่น การแจกแจงความถี่ของคนงานในโรงงานแห่งหนึ่งตามอัตราค่าแรงต่อวันซึ่งมี 7 ช่วง

อัตราค่าแรง	ความถี่
160-169 บาท	12
170-179 บาท	18
180-189 บาท	23
190-199 บาท	34
200-209 บาท	36
210-219บาท	47
220-229 บาท	50

16 12 7 14 8

4 15 12 16

15 18 20 15

12 13

รายละเอียดการแจกแจงความถี่ที่น่าสนใจ

อันตรภาคชั้นของการแจกแจงความถี่

- อันตรภาคชั้นเปิด
- อันตรภาคชั้นปิด
- ขีดจำกัดล่างของอันตรภาคชั้น
- ขีดจำกัดบนของอันตรภาคชั้น

การแจกแจงความถี่สะสม

การแจกแจงความถี่สัมพัทธ์

การแจกแจงความถี่โดยใช้กราฟ

การแจกแจงความถี่สะสมโดยใช้กราฟ

อันตรภาคชั้นของการแจกแจงความถี่ คือ ช่วงค่าที่เป็นไปได้ของข้อมูลที่เกิดขึ้นจากการแบ่งค่าที่เป็นไปได้ทั้งหมดของข้อมูลออกเป็นช่วงๆ เช่น รายได้ต่อเดือนของนักศึกษา แบ่งเป็น ต่ำกว่า 5,000 บาท 5,001-10,000 บาท 10001-15000 บาท 15001 ขึ้นไป

อันตรภาคชั้นเปิด คือ อันตรภาคชั้นที่ไม่มีค่าต่ำสุด หรือ ไม่มีค่าสูงสุด

เช่น..................................

อันตรภาคชั้นปิด คือ อันตรภาคชั้นที่มีค่าต่ำสุดและค่าสูงสุดของอันตรภาคชั้นที่แน่นอน

เช่น..................................

ขีดจำกัดล่างของอันตรภาคชั้น คือ ค่าเฉลี่ยระหว่างค่าต่ำสุดของอันตรภาคชั้นนั้นกับค่าสูงสุดของอันตรภาคชั้นก่อนหน้าหนึ่งชั้น เช่น ....

ขีดจำกัดบนของอันตรภาคชั้น คือ ค่าเฉลี่ยระหว่างค่าสูงสุดของอัตรภาคชั้นนั้นกับค่าต่ำสุดของอันตรภาคชั้นถัดไปหนึ่งชั้น เช่น ....

16 12 7 14 8
4 15 12 16
15 18 20 15
12 13

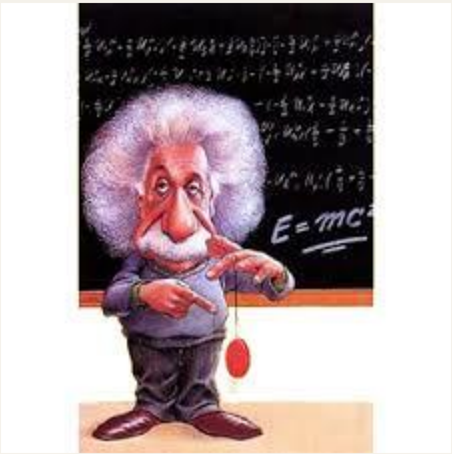
ตารางการแจกแจงความถี่

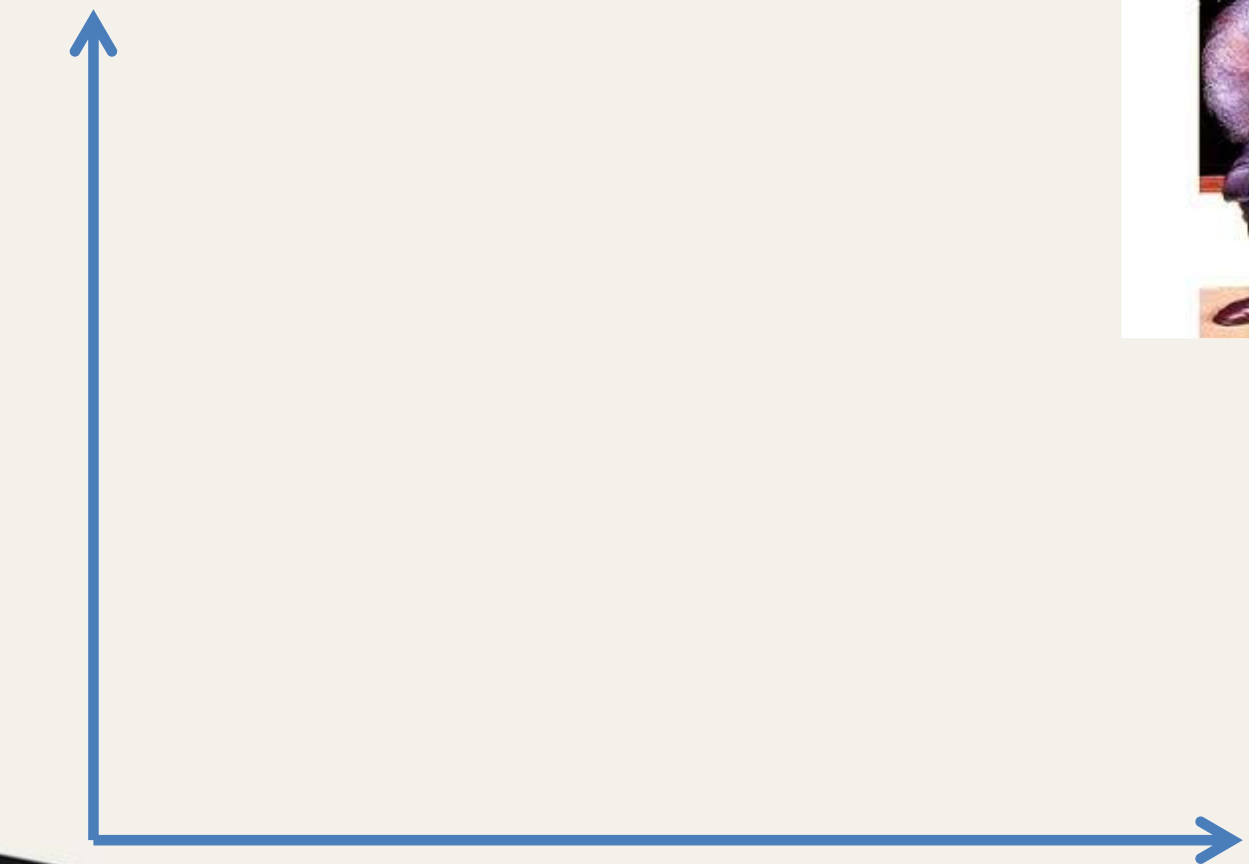
คะแนน	ความถี่	ขอบเขตล่าง	ขอบเขตบน	ความถี่สะสม

การแจกแจงความถี่สะสม คือ การจำแนกข้อมูลตามลักษณะต่างๆในรูปผลรวมของความถี่ของค่าหรืออันตรภาคชั้นนั้นกับค่าหรืออันตรภาคชั้นที่ต่ำกว่าทั้งหมด

การแจกแจงความถี่สัมพัทธ์ คือ การแจกแจงข้อมูลตามลักษณะต่างๆที่อยู่ในรูปอัตราส่วนระหว่างความถี่ หรืออันตรภาคชั้นนั้นกับผลรวมของความถี่ทั้งหมด

การแจกแจงความถี่โดยใช้กราฟ เช่น ฮีสโทแกรม เส้นโค้งความถี่ เป็นต้น

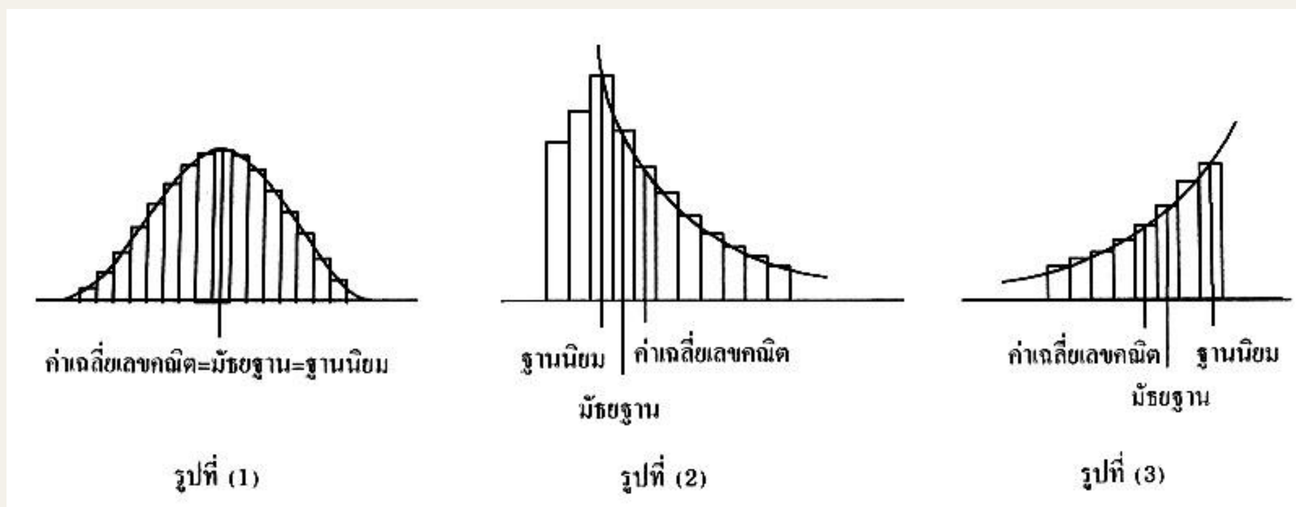




ฮีสโทแกรม

การวัดค่ากลางของข้อมูล

คือ ค่าที่ใช้เป็นตัวแทนของค่าทั้งหมดในข้อมูลชุดนั้น ค่ากลางที่ดีของข้อมูลชุดใดควรเป็นค่าใกล้เคียงกับค่าส่วนใหญ่ของข้อมูลชุดนั้น โดยค่ากลางมีหลายชนิด แต่ที่นิยมใช้กันเสมอๆ มีอยู่ 3 ชนิด คือ ค่าเฉลี่ย มัธยฐาน และฐานนิยม



ค่าเฉลี่ยเลขคณิต

ของข้อมูลที่ไม่ได้แจกแจงความถี่ แบบไม่ถ่วงน้ำหนัก

สูตร $$\bar{x} = \frac{1}{N}(X_1 + X_2 + X_3 + \dots X_N)$$

หรือ $$= \frac{1}{N}\sum_{i=1}^{N} X_i$$

ค่าเฉลี่ยเลขคณิต

ของข้อมูลที่ไม่ได้แจกแจงความถี่ แบบถ่วงน้ำหนัก

สูตร $$\overline{x} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots + w_N x_N}{w_1 + w_1 + w_1 + \ldots + w_N}$$

หรือ $$= \frac{\sum_{i=1}^{N} W_i X_i}{\sum_{i=1}^{N} W}$$

1,1,3,3,3,3,3

,3,3,5

การหาค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

(ข้อมูลแต่ละตัวมีการซ้ำกัน)

1 มี 2 ตัว

3 มี 7 ตัว

5 มี 1 ตัว

1 , 1 , 3 , 3 , 3 , 3 , 3 , 3 , 3 , 5

ค่าเฉลี่ยเลขคณิต

ของข้อมูลที่แจกแจงความถี่

$$\bar{x} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \ldots\ldots f_N X_N}{f_1 + f_2 + f_3 + \ldots\ldots f_N} = = \frac{1}{N} \sum_{i=1}^{N} f_i X_i$$

ช่วงข้อมูล	ความถี่ (f_i)	ค่ากลางชั้น (x_i)
1-9	6	
10-19	11	
20-29	17	
30-39	8	

มัธยฐาน

คือ ค่าที่มีจำนวนมากกว่าและน้อยกว่าค่านี้อยู่ประมาณเท่าๆกันของชุดข้อมูลนั้น การหาค่ามัธยฐานควรจะหาจากข้อมูลที่ยังไม่ได้แจกแจงความถี่ เนื่องจากการหาค่ามัธยฐานจากข้อมูลที่แจกแจงความถี่แล้วมีความยุ่งยาก และมีความเชื่อถือได้ค่อนข้างน้อย สำหรับข้อมูลที่ไม่ได้แจกแจงความถี่

1) ถ้าข้อมูลเป็นจำนวนคี่ มัธยฐาน คือ ค่าที่อยู่ตรงกลางของข้อมูลทั้งหมดเมื่อเรียงข้อมูลจากน้อยไปหาค่ามาก

2) ถ้าข้อมูลเป็นจำนวนคู่ มัธยมฐาน คือ ค่าเฉลี่ยของข้อมูล 2 ค่าซึ่งอยู่ตรงกลางของข้อมูลทั้งหมดเมื่อเรียงข้อมูลจากค่าน้อยไปหาค่ามาก

16 12 7 14 8 4 15 12 16

15 18 20 15 12 13

ฐานนิยม

คือ ค่าที่มีความถี่สูงสุดของข้อมูลชุดนั้น การหาฐานนิยมควรจะหาจากข้อมูลที่ยังไม่ได้แจกแจงความถี่ เนื่องจากการหาฐานนิยมจากข้อมูลที่แจกแจงความถี่แล้วมีความยุ่งยากและเชื่อถือได้น้อยเช่นเดียวกันกับการหามัธยฐาน

1) ถ้าข้อมูลมีความถี่สูงสุดค่าเดียว ฐานนิยม คือ ค่าของข้อมูลที่มีความถี่สูงสุด
2) ถ้าข้อมูลมีความถี่สูงสุดเท่ากัน 2 ค่า ถือว่า...
3) ถ้าข้อมูลมีความถี่สูงสุดเท่ากันมากกว่า 2 ค่า ถือว่า...............................

16 12 7 14 8 4 15 12 16

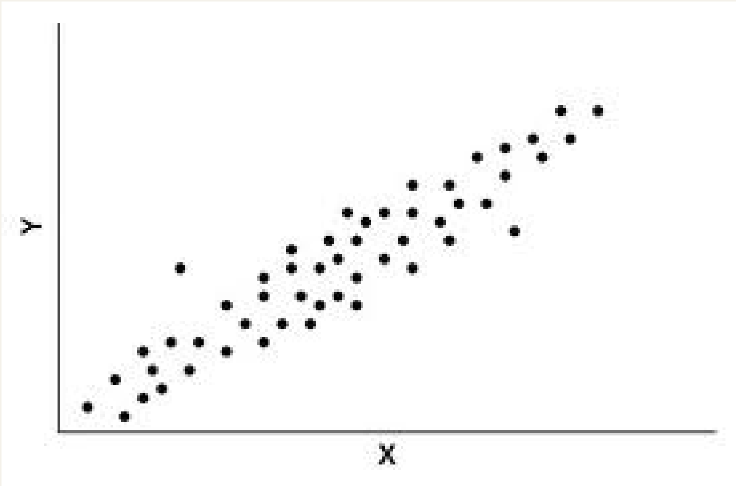
15 18 20 15 12 13

ค่ากึ่งกลางพิสัย = ½ (ค่าสูงสุด+ค่าต่ำสุด)

การวัดค่าการกระจายของข้อมูล

คือ ความแตกต่างระหว่างค่าแต่ละค่าของข้อมูลชุดนั้นกับค่าอื่นๆที่เหลือ โดยการวัดค่าการกระจายของข้อมูลสามารถวัดได้ 2 วิธี

1. การวัดการกระจายสัมบูรณ์
2. การวัดการกระจายสัมพัทธ์



การวัดการกระจายสัมบูรณ์

คือ การวัดค่าการกระจายโดยพิจารณาจากความแตกต่างระหว่างค่าแต่ละค่าของข้อมูลชุดนั้นกับค่าอื่นๆที่เหลือเพียงอย่างเดียว ไม่ได้พิจารณาถึงขนาดของข้อมูลแต่ละค่าด้วยว่ามีค่ามากหรือน้อยเพียงไร

วิธีที่นิยมในการวัดการกระจายสัมบูรณ์มี 3 วิธี

ได้แก่ พิสัย , ส่วนเบี่ยงเบนมาตรฐาน , ความแปรปรวน

พิสัย

คือ ค่าที่ใช้วัดการกระจายของข้อมูลอย่างหยาบๆ นิยมใช้เมื่อต้องการทราบค่าการกระจายของข้อมูลอย่างรวดเร็ว

พิสัยของข้อมูลที่ไม่ได้แจกแจงความถี่ = $x_{max} - x_{min}$

พิสัยของข้อมูลที่แจกแจงความถี่ = ขีดจำกัดบนของอันตรภาคชั้นที่มีค่าสูงสุด – ขีดจำกัดล่างของอันตรภาคชั้นที่มีค่าต่ำสุด

* การแจกแจงความถี่ของอันตรภาคชั้นเปิดไม่สามารถนำมาหาค่าพิสัยได้

ส่วนเบี่ยงเบนมาตรฐาน

คือ ค่าที่ใช้วัดการกระจายของข้อมูล ซึ่งเป็นที่ยอมรับกันโดยทั่วไปว่ามีความเชื่อถือได้มากที่สุด แต่การสัดค่าการกระจายโดยใช้ส่วนเบี่ยงเบนมาตรฐานมีความยุ่งยากกว่าการใช้พิสัย (S,S.D)

สูตรของข้อมูลที่ไม่ได้แจกแจงความถี่ $S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$

สูตรของข้อมูลที่แจกแจงความถี่ $S = \sqrt{\frac{1}{N}\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}$

ส่วนเบี่ยงเบนมาตรฐาน

แทนค่าในสูตรที่ 1 เมื่อ x_i แทนค่าของข้อมูล

N แทนจำนวนข้อมูล

$\bar{x}$ แทนค่าเฉลี่ยเลขคณิต

แทนค่าในสูตรที่ 2 เมื่อ x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i แทนความถี่ของอันตรภาคชั้นที่ i

k แทนจำนวนอันตรภาคชั้น

N แทนจำนวนข้อมูลทั้งหมด ซึ่งเท่ากับผลรวมของความถี่ทุกอันตรภาคชั้น

$\bar{x}$ แทนค่าเฉลี่ยเลขคณิต

ความแปรปรวน

คือ ค่าที่ได้จากการนำเอาส่วนเบี่ยงเบนมาตรฐานมายกกำลังสอง ดังนั้น ความแปรปรวนจึงเป็นค่าที่ใช้วัดการกระจายเช่นเดียวกัน แต่ขนาดของค่าการกระจายที่วัดได้ไม่ได้เป็นขนาดที่แท้จริง เป็นค่าการกระจายที่ยกกำลังสอง (S^2)

การวัดการกระจายสัมพัทธ์

คือ เป็นการวัดการกระจายโดยพิจารณาจากความแตกต่างระหว่างค่าแต่ละค่าของข้อมูลนั้นกับค่าอื่นๆที่เหลือเมื่อเทียบกับขนาดของข้อมูลชุดนั้น

วิธีที่นิยมในการวัดการกระจายสัมบูรณ์มี 2 วิธี

ได้แก่ สัมประสิทธิ์ของพิสัย , สัมประสิทธิ์การแปรผัน

สัมประสิทธิ์ของพิสัย

คือ อัตราส่วนระหว่างผลต่างกับผลบวกของค่าสูงสุด (x_{max}) และค่าต่ำสุด (x_{min}) ของข้อมูลชุดเดียวกัน

$$\text{สัมประสิทธิ์ของพิสัย} = \frac{X_{max} - X_{min}}{X_{max} + X_{min}}$$

สัมประสิทธิ์ของพิสัยใช้วัดค่าการกระจายสัมพัทธ์อย่างหยาบๆ ในกรณีที่ข้อมูลมีการกระจายน้อย หรือ ค่าแต่ละค่าของข้อมูลแตกต่างกันน้อยสัมประสิทธิ์ของพิสัยจะใช้วัดค่าการกระจายได้ดีพอสมควร

สัมประสิทธิ์การแปรผัน

คือ อัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดเดียวกัน

สัมประสิทธิ์การแปรผัน หรือ $\mathrm{C.V.} = \frac{S}{\bar{x}}$

โดยทั่วไปสัมประสิทธิ์การแปรผันใช้วัดค่าการกระจายสัมพันธ์ได้ถูกต้องกว่าค่าสัมประสิทธิ์ของพิสัยและนิยมวัดในรูปร้อยละ

ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น

1 ความน่าจะเป็น
2 ตัวแปรสุ่ม
3 การแจกแจงความน่า
4 ประชากรและการสุ่มตัวอย่าง

ความน่าจะเป็นของเหตุการณ์

คือ อัตราส่วนของจำนวนสมาชิกของเหตุการณ์นั้นกับจำนวนสมาชิกของเหตุการณ์ทั้งหมดที่อาจเกิดขึ้นได้ โดยที่สมาชิกแต่ละตัวมีโอกาสเกิดขึ้นได้เท่าๆกัน

ถ้าเหตุการณ์ที่สนใจศึกษา ความน่าจะเป็นของเหตุการณ์ E นิยมเขียนแทนด้วยสัญลักษณ์ P(E)

โดย $$P(E) = \frac{n}{N}$$

ตัวอย่าง

ในเดือนสิงหาคม 2557 บริษัทตุ๊กตาไทย จำกัด ผลิตตุ๊กตาได้จำนวน 5,000 ตัว ในจำนวนที่ผลิตได้มีตุ๊กตาชำรุด 80 ตัว จงหาความน่าจะเป็นที่บริษัทจะผลิตตุ๊กตาชำรุด

n คือ เหตุการณ์ที่สนใจ จากโจทย์ บริษัทผลิตตุ๊กตาแล้วชำรุด = 80

N คือ จำนวนเหตุการณ์ทั้งหมดที่เป็นไปได้ คือ บริษัทผลิตตุ๊กตาทั้งหมด(ชำรุดและไม่ชำรุด) = 5000

แทนค่า $P(E) = \frac{n}{N} = \frac{80}{5000} = 0.016$ (ไม่มีหน่วย)

ตัวอย่าง

จากการเก็บข้อมูลเกี่ยวกับจำนวนวันที่บริษัทผลิตตุ๊กตาได้ในรอบเดือน ตุลาคม 2557 ดังนี้

จำนวนตุ๊กตาที่ผลิต (ตัว/วัน)	จำนวนวันที่ผลิตได้
ต่ำกว่า 4,000	1
4,000-4,499	6
4,500-4,999	18
5,000-5,499	4
ตั้งแต่ 5,500 ขึ้นไป	2

ตัวอย่าง

จงหาความน่าจะเป็นที่บริษัทจะผลิตเสื้อได้ตั้งแต่ 5,000 ถึง 5,499 ตัว/วัน

แทนค่า P(E) $\frac{n}{N} = \frac{\mathbf{4}}{\mathbf{31}}$ โดย $\frac{\mathbf{4}\text{ มาจากจำนวนวันของเหตุการณ์ที่เราสนใจ}}{\mathbf{31}\text{ มาจากจำนวนวันของเหตุการณ์ทั้งหมด}}$

จงหาความน่าจะเป็นที่บริษัทจะผลิตเสื้อได้ตั้งแต่ 5,000 ตัวขึ้นไป/ต่อวัน

แทนค่า P(E) $\frac{n}{N} = \frac{\mathbf{4+2}}{\mathbf{31}}$ โดย $\frac{\mathbf{4+2}\text{ มาจากจำนวนวันของเหตุการณ์ที่เราสนใจ}}{\mathbf{31}\text{ มาจากจำนวนวันของเหตุการณ์ทั้งหมด}}$

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

ตัวแปรสุ่ม คือ ค่าหรือตัวเลขที่ใช้แทนเหตุการณ์ต่างๆที่อาจเกิดขึ้นได้ทั้งหมดจากการทดลองหรือสภาพที่เป็นจริงเกี่ยวกับเรื่องที่ศึกษา

โดยทั่วไปนิยมใช้พยัญชนะภาษาอังกฤษตัวใหญ่แทนตัวแปรสุ่ม เช่น X Y Z

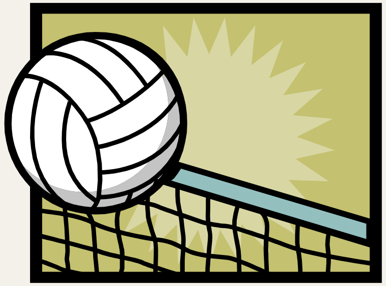
และนิยมใช้พยัญชนะภาษาอังกฤษตัวเล็กแทนค่าที่เป็นไปได้ของตัวแปรสุ่มนั้น เช่น x_1 x_2 x_3 y_1 y_2 y_3 z_1 z_2 z_3

ตัวอย่าง

1. ถ้าให้ X เป็นตัวแปรสุ่มที่ใช้แทนเหตุการณ์ที่เป็นจำนวนวันซึ่งร้านอุทัยพานิชย์เปิดขายสินค้าในเดือน มิถุนายน 2557 อาจเป็น 0 (ไม่เปิดขายสินค้าเลย) ,1,2,3,4,5,6,7,......30 (เปิดขายทุกวัน)
2. ถ้าให้ Y เป็นตัวแปรสุ่มที่ใช้แทนเหตุการณ์ที่เป็นจำนวนลูกค้าซึ่งเข้ามาซื้อสินค้าในวันที่ผ่านมาของร้านอรุณการค้า Y อาจเป็น 0 (ไม่มีลูกค้ามาซื้อสินค้าเลย) ,1,2,3,4,5......
3. ถ้าให้ Z เป็นตัวแปรสุ่มที่ใช้แทนเหตุการณ์ที่เป็นจำนวนเครื่องจักรชำรุดของโรงงานอุตสาหกรรมไทยพลาสติกในเดือนที่ผ่านมา ซึ่งมีจำนวนเครื่องจักรทั้งหมด 25 เครื่อง Zอาจเป็น 0 (ไม่มีเครื่องจักรชำรุด)1,2,3,4,5......25 (เครื่องจักรชำรุดทั้งหมด)

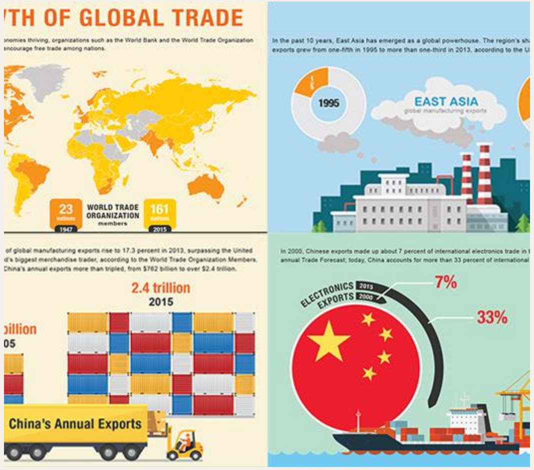
ตัวแปรสุ่มมี 2ชนิด คือ

- ตัวแปรสุ่มชนิดไม่ต่อเนื่อง (Discrete random variable)
- ตัวแปรสุ่มชนิดต่อเนื่อง (Continuous random variable)



ตัวแปรสุ่มชนิดไม่ต่อเนื่อง (Discrete random variable)

คือ ตัวแปรสุ่มที่ประกอบด้วยค่าหรือตัวเลขทั้งหมดที่เป็นไปได้จำนวนจำกัด หรือถ้ามีจำนวนไม่จำกัดก็จะต้องไม่มีตัวเลขอื่นใดนอกเหนือจากตัวเลขดังกล่าวอยู่ระหว่างตัวเลขสองตัวของตัวเลขทั้งหมดที่เป็นไปได้นี้ โดยทั่วไปตัวแปรสุ่มชนิดไม่ต่อเนื่องคือ จำนวนนับ เช่น 0,1,2,3,4.....................................

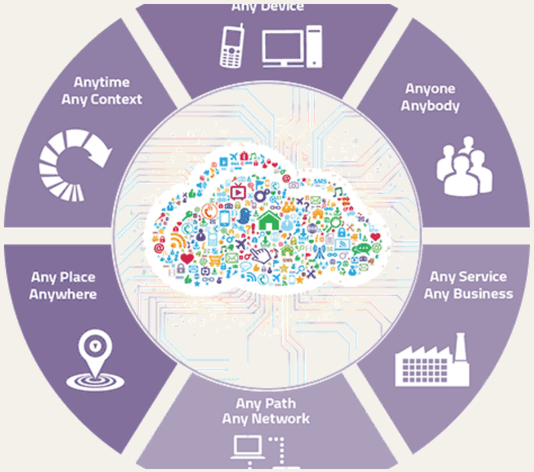


ตัวอย่าง

1. X = 10,11,12,13,14,15
2. Y=จำนวนคนงานในโรงงานแห่งหนึ่ง
3. Z= 0.1,0.3,0.5,0.7,0.9
4. P=จำนวนร้านค้าที่เป็นตัวแทนจำหน่ายสินค้าของบริษัทชัยอนันต์
5. R = จำนวนสินค้าชำรุดที่เครื่องจักรของโรงงานแห่งหนึ่งผลิตได้ในเดือนพฤษภาคม 2557

ตัวแปรสุ่มชนิดต่อเนื่อง (Continuous random variable)

คือ ตัวแปรสุ่มที่ประกอบด้วยค่าหรือตัวเลขนับไม่ถ้วน (infinite) โดยทั่วไปตัวแปรสุ่มชนิดต่อเนื่องอยู่ในรูปช่วงของค่าหรือตัวเลข 2 ตัว ซึ่งจำนวนค่าหรือตัวเลขที่อยู่ระหว่างตัวเลข 2 ตัวนี้จะมีจำนวนที่นับไม่ถ้วนหรือไม่สามารถนับได้



ตัวอย่าง

1. $5 \leq M \leq 10$
2. $N \geq 0$ เป็นตัวแปรสุ่มชนิดต่อเนื่อง
3. P = ราคาสินค้าปลีกชนิดหนึ่ง
4. $Q \geq 1.732$
5. R = ปริมาณน้ำที่ใช้ในการผลิตน้ำอัดลมชนิดหนึ่งในรอบเดือนที่ผ่านมาเป็นตัวแปรสุมชนิดต่อเนื่อง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม คือ การแจกแจงของเหตุการณ์ทุกเหตุการณ์ที่เป็นไปได้ และความน่าจะเป็นที่เหตุการณ์เหล่านั้นจะเกิดขึ้น

ตัวอย่าง เช่น ในการลงทุนเปิดร้านเสริมสวยสตรี ผลจากการลงทุนมีเหตุการณ์ที่เป็นไปได้ 3 เหตุการณ์ก็คือ กำไร เท่าทุน และขาดทุน การแจกแจงความน่าจะเป็นของเหตุการณ์นี้ก็คือ จะต้องหาความน่าจะเป็นที่ลงทุนเปิดร้านเสริมสวยสตรีแล้วได้กำไร ความน่าจะเป็นที่ลงทุนเปิดร้านเสริมสวยสตรีแล้วเท่าทุน และความน่าจะเป็นที่ลงทุนเปิดร้านเสริมสวยสตรีแล้วขาดทุน

การหาความน่าจะเป็นของเหตุการณ์ทั้งสามดังกล่าวนี้อาจทำได้โดยการเก็บรวบรวมข้อมูลเกี่ยวกับผลการลงทุนของร้านเสริมสวยสตรีที่เปิดดำเนินการหรือเคยเปิดดำเนินการมาแล้วในรอบปีที่ผ่านมา ถ้าจากการสอบถามร้านเสริมสวยสตรีจำนวน 1,000 ร้าน ปรากฏว่าได้กำไร 850 บาท เท่าทุน 125 ร้าน และขาดทุน 25 ร้าน

ผลการลงทุน	P(E)	ความน่าจะเป็น
กำไร	$\frac{850}{1000}$	0.850
เท่าทุน	$\frac{125}{1000}$	0.125
ขาดทุน	$\frac{25}{1000}$	0.025
รวม	$\frac{850+125+25}{1000}$	1.000

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

ตัวอย่างเช่น ถ้าจำนวนวันที่บริษัทขายรถยนต์แห่งหนึ่งขายรถได้เป็นจำนวน 0(ขายไม่ได้เลย),1,2,3,4,5,6,7,8,9,และ10คัน ในรอบเดือนที่ผ่านมาเป็นดังนี้

จำนวนรถยนต์ที่ขายได้ (คัน)	0	1	2	3	4	5	6	7	8	9	10
จำนวนวันที่ขายได้ (วัน)	1	1	2	4	6	7	5	2	1	1	1

การแจกแจงความน่าจะเป็นของจำนวนรถยนต์ที่ขายได้

จำนวนรถที่ขายได้ต่อวัน(x)	จำนวนวันที่ขายได้	ความน่าจะเป็นที่จะขายได้P(X)
0	1	
1	1	
2	2	
3	4	
4	6	
5	7	
6	5	
7	2	
8	1	
9	1	
10	1	

ค่าเฉลี่ยและความแปรปรวนของการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

ค่าเฉลี่ยของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

$$E(X) = \sum_{i=1}^{n} x_i P(x_i)$$

บริษัทควรขายรถยนต์ได้เฉลี่ยวันละ หรือ ความแปรปรวน/ความแตกต่างระหว่างจำนวนรถยนต์ที่บริษัทขายได้แต่ละวันมีมากน้อยเพียงไร...

....

....

.

ค่าเฉลี่ยการแจกแจงความน่าจะเป็นของจำนวนรถยนต์ที่ขายได้ต่อวัน

จำนวนรถที่ขายได้ต่อวัน	จำนวนวันที่ขายได้	ความน่าจะเป็นที่จะขายได้	X P (X)
0	1		
1	1		
2	2		
3	4		
4	6		
5	7		
6	5		
7	2		
8	1		
9	1		
10	1		

ตัวอย่าง

ร้านดาวเบเกอรี่รี่รับขนมเค้กมาจำหน่ายครั้งละ 5 กล่อง ถ้าภายในระยะ 3 วันหลังจากรับมาจำหน่ายความน่าจะเป็นที่ขนมเค้กจะไม่เสียเลย เท่ากับ 0.8 เสีย1 กล่องเท่ากับ 0.1 เสีย 2 กล่องเท่ากับ 0.05 เสีย 3 กล่อง เท่ากับ 0.03 เสีย 4 กล่องเท่ากับ 0.01 และเสีย 5 กล่องเท่ากับ 0.01 ตามลำดับ จงหาจำนวนกล่องโดยเฉลี่ยที่ขนมเค้กเสียในการสั่งแต่ละครั้ง ....

....

....

....

ตัวอย่าง

ในการประกอบธุรกิจของนายประสงค์ในช่วง 5 ปีข้างหน้า เขาคาดว่าปีแรกจะขาดทุน 150,000 บาท ปีที่สองจะขาดทุน 50,000 บาท ปีที่สามจะขาดทุน 10,000 บาท ปีที่สี่จะได้กำไร 20,000 บาท และปีที่ห้าจะได้กำไร 40,000 บาท นายประสงค์จะขาดทุนจากการประกอบกิจการโดยเฉลี่ยปีละเท่าไหร่ โดยที่ค่าความน่าจะเป็นปีแรกที่จะขาดทุนเท่ากับ 0.1 ปีที่สองเท่ากับ 0.1ปีที่สามเท่ากับ 0.2 ปีที่สี่เท่ากับ 0.3 และปีที่ห้าเท่ากับ 0.3

ความแปรปรวนของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

ความแปรปรวนของตัวแปรสุ่มเป็นค่าที่แสดงการกระจายของข้อมูลโดยวัดจากค่าเฉลี่ยของผลต่างกำลังสองระหว่างค่าแต่ละค่าของตัวแปรสุ่มกับค่าเฉลี่ย หรือค่าคาดคะเนของตัวแปรสุ่มนั้น

โดยทั่วไปการวัดค่าการกระจายของข้อมูลซึ่งประกอบด้วยค่าต่างๆที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มนิยมวัดโดยใช้ค่าส่วนเบี่ยงเบนมาตรฐาน ซึ่งเป็นรากที่สองของความแปรปรวน

ความแปรปรวนของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

ถ้าให้ X เป็นตัวแปรสุ่มชนิดไม่ต่อเนื่องซึ่งประกอบด้วยค่า $x_1 , x_2 , x_3 , \ldots\ldots, x_n$ ตามลำดับ และ $P(x_1) , P(x_2) , P(x_3) , \ldots\ldots, P(x_n)$ เป็นความน่าจะเป็นของ $x_1 , x_2 , x_3 , \ldots\ldots, x_n$ ตามลำดับ

ดังนั้น ค่าความแปรปรวนของ X คือ

สูตร $\sum_{i=1}^{N} (x_i - \mu_x)^2 \cdot P(x_i)$

ตัวอย่าง

จากข้อมูลเกี่ยวกับการขายรถยนต์ ถ้าจำนวนวันที่ขายรถยนต์แห่งหนึ่งขายรถได้เป็นจำนวน 0,1,2,3,4,5,6,7,8,9,และ10คัน ในรอบเดือนที่ผ่านมาเป็นดังนี้

จำนวนรถยนต์ที่ขายได้ (คัน)	0	1	2	3	4	5	6	7	8	9	10
จำนวนวันที่ขายได้ (วัน)	1	1	2	4	6	7	5	2	1	1	1

ตัวอย่าง

เนื่องจากสูตร $Var(X) = \sigma^2 = \sum (x_i - \bar{x})^2 \cdot P(x_i)$

แทนค่า $X_1 = 0$, $X_2 = 1$, $X_3 = 2$, $X_4 = 3$, $X_5 = 4$, $X_6 = 5$, $X_7 = 6$,

$X_8 = 7$, $X_9 = 8$, $X_{10} = 9$, $X_{11} = 11$,

แทนค่า $P_{(x1)}$ = ... , $P_{(x2)}$ = ... , $P_{(x3)}$ = ... , $P_{(x4)}$ = ... , $P_{(x5)}$ = ...

$P_{(x6)}$ = ... ,$P_{(x7)}$ = ... ,$P_{(x8)}$ = ... $P_{(x9)}$ = ... ,$P_{(x10)}$ = ...

$P_{(x11)}$ = ... ,

ตัวอย่าง

ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของจำนวนรถยนต์ที่บริษัทขายได้

สูตร

....

....

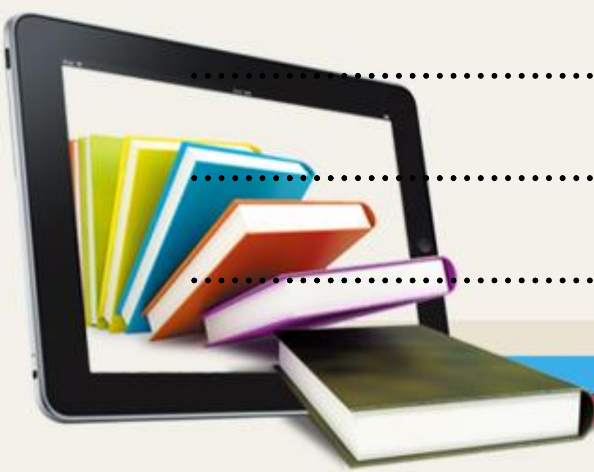
....

....

....

....

....



แบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

การแจกแจงความน่าจะเป็นแบบทวินาม

การทดลองแบบทวินามมีลักษณะดังต่อไปนี้

1. มีจำนวนครั้งของการทดลองแน่นอน
2. ผลการทดลองแต่ละครั้งเกิดได้เพียง 2 อย่าง คือ สำเร็จหรือล้มเหลว
3. ความน่าจะเป็นที่จะให้ผลการทดลองสำเร็จเท่ากันทุกครั้ง
4. ผลการทดลองแต่ละครั้งไม่ขึ้นต่อกัน

แบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

การแจกแจงความน่าจะเป็นแบบทวินาม (Binomial distribution)

สูตรในการหาค่าความน่าจะเป็นแบบทวินาม

$$P(x) = \binom{n}{x} p^{x} q^{n-x}$$

เมื่อ x จำนวนครั้งที่ทำการทดลองสำเร็จ

n จำนวนครั้งที่ทำการทดลอง

p ความน่าจะเป็นที่สำเร็จในการทดลอง

q ความน่าจะเป็นที่ล้มเหลวในการทดลอง $q = 1 - p$

แบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

ค่าเฉลี่ยของการแจกแจงแบบทวินาม $E(X) = \mu = np$

ความแปรปรวนของการแจกแจงแบบทวินาม $Var(X) = \sigma^2 = npq$

ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบทวินาม $\sigma = \sqrt{npq}$

ตัวอย่าง

พนักงานของบริษัททรงชัย จำกัด โทรศัพท์ติดต่อลูกค้าเพื่อขายสินค้าวันละ 10 ราย ถ้าความน่าจะเป็นที่เขาจะขายสินค้าได้สำหรับลูกค้าแต่ละราย เท่ากับ 0.2 จงหาความน่าจะเป็นที่พนักงานขายผู้นี้จะ

1. ขายสินค้าไม่ได้เลย
2. ขายสินค้าให้ลูกค้าได้ 3 ราย
3. ขายสินค้าให้ลูกค้าได้อย่างน้อย 1 ราย
4. จงหาค่าเฉลี่ยและความแปรปรวน

ตัวอย่าง

ให้ x แทนจำนวนลูกค้าที่พนักงานขายสินค้าได้ในหนึ่งวัน

นั่นคือ $x = 0 , 1 , 2 , \ldots , 10$

$n = 10 , p = 0.2 , q = 1- p = 1 - 0.2 = 0.8$

เนื่องจากเป็นการแจกแจงแบบทวินาม

$$P(x) = \binom{n}{x} p^{x} q^{n-x}$$

ตัวอย่าง

ความน่าจะเป็นที่พนักงานขายสินค้าไม่ได้เลย

นั่นคือ ตัวแปรสุ่ม คือ $x = 0$ จะได้

$$P(x=0)=\binom{10}{0}0.2^{0}0.8^{10}$$

$$= 0.1074$$

ดังนั้น ความน่าจะเป็นที่พนักงานขายสินค้าไม่ได้เลย เท่ากับ 0.1074

ตัวอย่าง

ความน่าจะเป็นที่พนักงานขายสินค้าให้ลูกค้าอย่างน้อย 1 ราย

นั่นคือ ตัวแปรสุ่ม คือ $x = 1$ จะได้

....

....

....

....

....

ตัวอย่าง

ความน่าจะเป็นที่พนักงานขายสินค้าให้ลูกค้า 3 ราย

นั่นคือ ตัวแปรสุ่ม คือ $x = 3$ จะได้

....

....

....

....

....

ตัวอย่าง

จากโจทย์ $n = 10$, $p = 0.2$ และ $q = 1 - 0.2 = 0.8$

ค่าเฉลี่ยของการแจกแจงแบบทวินาม $\mu_x = E(X) = np = (10)(0.2) = 2$ ราย

ความแปรปรวนของการแจกแจงแบบทวินาม $Var(X) = \sigma^2 = npq$

$= (10)(0.2)(0.8) = 1.6$

ส่วนเบี่ยงเบนมาตรฐาน $= \sqrt{1.6} = 1.26$

แบบการแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดไม่ต่อเนื่อง

การแจกแจงความน่าจะเป็นแบบปัวส์ซอง

สูตรในการหาค่าความน่าจะเป็นแบบปัวส์ซอง

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

เมื่อ $x = 0 , 1 , 2 \ldots$

μ แทนค่าเฉลี่ยของจำนวนครั้งที่สำเร็จ

e แทนค่าคงที่ซึ่งมีค่าเท่ากับ 2.71828

ค่าเฉลี่ยของการแจกแจงแบบปัวส์ซอง $P(X) = \mu$

ความแปรปรวนของการแจกแจงแบบแบบปัวส์ซอง $Var(X) = \mu$

ส่วนเบี่ยงเบนมาตรฐานของการแจกแจงแบบปัวส์ซอง $\sigma^2 = \mu$

ตัวอย่าง

ร้านอาหารแห่งหนึ่งมีบริการส่งอาหารถึงบ้าน เฉลี่ยแล้วมีลูกค้าโทรเข้ามาสั่งอาหาร 3 ราย ต่อ นาที จงหาความน่าจะเป็นที่

1. ลูกค้าโทรเข้ามาสั่งอาหาร 2 ครั้ง ต่อ นาที
2. ลูกค้าไม่โทรเข้ามาสั่งอาหารเลย
3. ลูกค้าโทรเข้ามาสั่งอาหารอย่างน้อย 1 ครั้ง ต่อ นาที

วิธีทำ **เนื่องจากเป็นการแจกแจงแบบปัวส์ซอง**

$x = 0,1,2,\ldots$ $\mu = 3$

การแจกแจงความน่าจะเป็นแบบปัวส์ซอง

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

1.ความน่าจะเป็นที่ลูกค้าโทรเข้ามาสั่งอาหาร 2 ครั้งต่อนาที

$$P(x=2) = \frac{e^{-3}3^{2}}{2!}$$

= 0.2240

ดังนั้น ความน่าจะเป็นที่ลูกค้าโทรเข้ามาสั่งอาหาร 2 ครั้งต่อนาที เท่ากับ 0.2240

2. ความน่าจะเป็นที่ลูกค้าไม่โทรเข้ามาสั่งอาหารเลย

$$P(x=0)=\frac{e^{-3}3^{0}}{0!}$$

$$= 0.0498$$

ดังนั้น ความน่าจะเป็นที่ลูกค้าไม่โทรเข้ามาสั่งอาหารเลย เท่ากับ 0.0498

3. ความน่าจะเป็นที่ลูกค้าโทรเข้ามาสั่งอาหารอย่างน้อย 1 ราย

$$P(x \geq 1) = 1 - P(x = 0)$$

$$= 1 - 0.0498$$

$$= 0.9502$$

ดังนั้น ความน่าจะเป็นที่ลูกค้าโทรเข้ามาสั่งอาหาร

อย่างน้อย 1 ราย เท่ากับ 0.9502

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่อง

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มชนิดต่อเนื่องเป็นการแจกแจงเหตุการณ์ทุกเหตุการณ์ที่เป็นไปได้ และความน่าจะเป็นที่เหตุการณ์เหล่านั้นจะเกิดขึ้นเมื่อตัวแปรสุ่มที่สนใจศึกษาเป็นตัวแปรสุ่มชนิดต่อเนื่อง

จำนวนสินค้าที่พนักงานขาย 50 คน ของบริษัทแห่งหนึ่งขายได้ในรอบเดือนที่ผ่านมา

จำนวนสินค้าที่ขายได้ (หน่วย)	จำนวนพนักงานนายที่ขายได้
22-26	2
27-31	8
32-36	15
37-41	15
42-46	7
47-51	3

ความน่าจะเป็นของตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงปกติ

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

เมื่อ X คือ ตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงแบบปกติ

μ คือ ค่าเฉลี่ยของประชากรที่เกิดจากตัวแปรสุ่ม X ที่มีการแจกแจงแบบปกติ

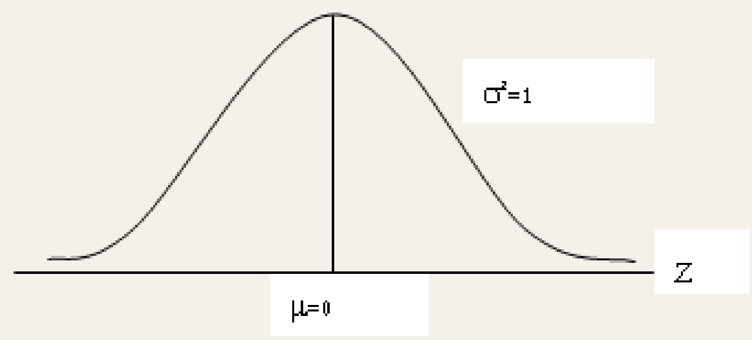
σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรที่เกิดจากตัวแปรสุ่ม X ที่มีการแจกแจงแบบปกติ

π มีค่าเท่ากับ 3.14159...

e มีค่าเท่ากับ 2.71828...

การแจกแจงปกติที่มีค่าเฉลี่ย $(\mu = 0)$ และส่วนเบี่ยงเบนมาตรฐาน $(\sigma = 1)$ เรียกว่า การแจกแจงแบบปกติมาตรฐาน กรณีที่การแจกแจงแบบปกติมีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานไม่เป็นไปตามเงื่อนไขดังกล่าว เราจะต้องแปลงเป็นคะแนนมาตรฐาน จากสูตร

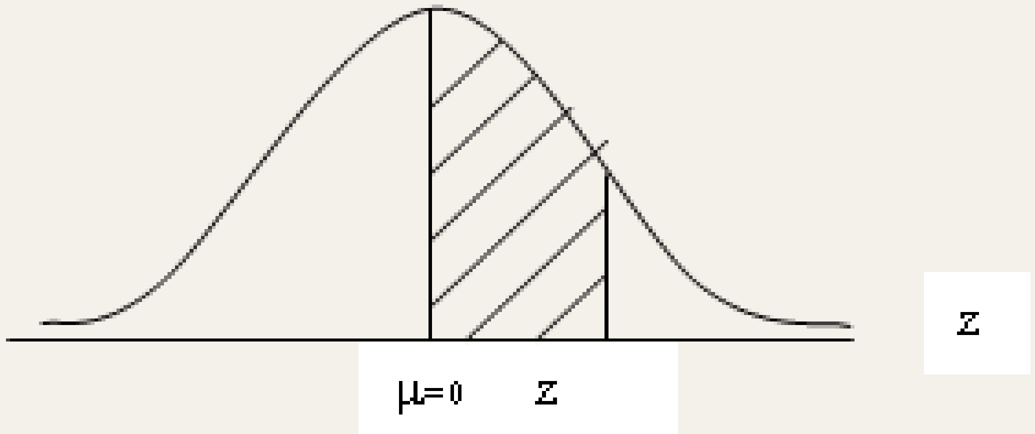
ลักษณะของโค้งปกติมาตรฐาน



การหาค่าความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐาน
โดยอาศัยตาราง

การหาค่าความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐานนอกจากหาค่าโดยการคำนวณโดยตรงแล้ว ยังสามารถหาได้โดยอาศัยตารางสำเร็จรูปของการแจกแจงแบบปกติมาตรฐานที่มีหลายรูปแบบ แต่ในที่นี้จะอาศัยตารางหาค่าความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐานในท้ายเล่มของหนังสือที่บอกค่าพื้นที่ภายใต้โค้งที่สะสมค่า (Cumulative Value) จากจุดที่ค่า $\mu = 0$ ถึงค่า Z ใด ๆ

ดังรูป

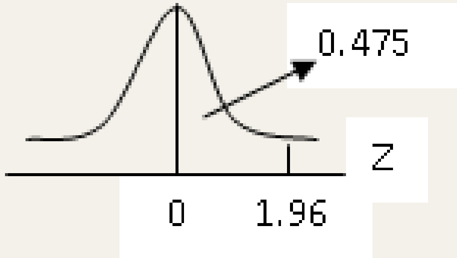


จากตารางการหาค่าความน่าจะเป็นของการแจกแจงแบบปกติมาตรฐาน

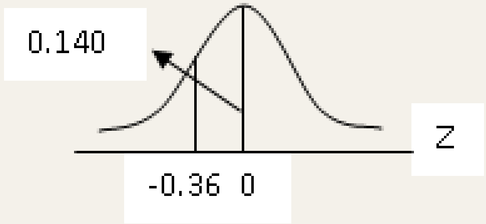
จงหาค่าความน่าจะเป็นของ $P(0 < Z < 1.96)$,

$P(-0.36 < Z < 0)$,$P(Z > 1.52)$,$P(Z < -1.24)$ และ

$P(1.05 < Z < 2.55)$

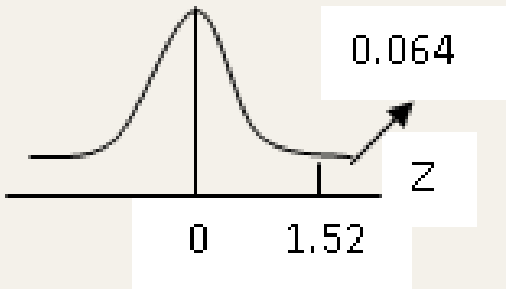


1. $P(0 < Z < 1.96) = 0.4750$



2. $P(-0.36 < Z < 0) = P(0 < Z < 0.36)$

$$= 0.1406$$



3. $P(Z > 1.52) = 0.5 - P(0 < Z < 1.52)$

$$= 0.5 - 0.4357$$

$$= 0.0643$$

ประชากรและการสุ่มตัวอย่าง

1. ประชากรและตัวอย่าง
2. การเลือกตัวอย่าง
3. การกำหนดจำนวนตัวอย่าง
4. ข้อควรระวังในการเลือกตัวอย่าง

ประชากรและตัวอย่าง

การเก็บข้อมูลเกี่ยวกับลักษณะต่างๆของกลุ่มที่ต้องการทราบเพื่อนำมาใช้ในการวิเคราะห์วิจัย สำหรับใช้ในการตัดสินใจแก้ปัญหา หรือใช้วางแผนเกี่ยวกับเรื่องต่างๆที่เกี่ยวกับเรื่องต่างๆที่เกี่ยวข้องกับกลุ่มนั้นๆในอนาคต มีความจำเป็นต้องทราบเกี่ยวกับประชากร (population) ตัวอย่าง (sampling) และหน่วยตัวอย่าง (sampling unit) ก่อน เนื่องจากมีความเกี่ยวข้องกับกระบวนการในการเก็บรวบรวมข้อมูลในขั้นตอนต่างๆโดยตรง

ประชากร

หมายถึง ที่รวมของทุกๆหน่วยซึ่งมีลักษณะที่ผู้วิเคราะห์วิจัยต้องการเก็บรวบรวมข้อมูล ซึ่งอาจเป็นคน สัตว์ สิ่งของ หรือสถานที่ก็ได้ เช่น

ชาวกรุงเทพ ฯ ทุกคน คือ ประชากรที่ผู้วิเคราะห์วิจัยต้องการเก็บรวบรวมข้อมูลเกี่ยวกับพฤติกรรมการบริโภคอาหารจานด่วน

วัยรุ่นสตรีทุกคนในประเทศไทย คือ ประชากรที่ผู้วิเคราะห์วิจัยต้องการเก็บรวบรวมข้อมูลเกี่ยวกับความพึงพอใจที่มีต่อการใช้เครื่องสำอางที่ผลิตในประเทศไทย

ตัวอย่าง

หมายถึง บางหน่วยของประชากรที่เลือกมาเป็นตัวแทนจากทุกๆหน่วยของประชากรนั้นๆ เช่น

ตัวอย่าง ชาวกรุงเทพที่เลือกมาเป็นตัวแทนในการเก็บรวบรวมข้อมูลเกี่ยวกับพฤติกรรมการบริโภคอาหารจานด่วน

ตัวอย่าง วัยรุ่นสตรีในประเทศไทยบางคนที่เลือกมาเป็นตัวแทนในการเก็บรวบรวมข้อมูลเกี่ยวกับความพึงพอใจต่อการใช้เครื่องสำอางค์ที่ผลิตในประเทศไทย

หน่วยตัวอย่าง

หมายถึงหน่วยที่สามารถให้ข้อมูลที่ผู้วิเคราะห์วิจัยต้องการซึ่งอาจจะได้จากการสัมภาษณ์โดยตรง หรือได้จากการสังเกตโดยตรง หรือได้จากการทดลองโดยตรงก็ได้ หน่วยตัวอย่างอาจเป็นตัวอย่างหรือส่วนหนึ่งของตัวอย่างก็ได้ เช่น

ตัวอย่าง คือ ครอบครัว หน่วยตัวอย่างอาจเป็นหัวหน้าครอบครัว คู่สมรสของหัวหน้าครอบครัว หรือ ผู้ใดผู้หนึ่งในครอบครัวที่สามารถให้ข้อมูลที่ผู้วิเคราะห์วิจัยต้องการได้

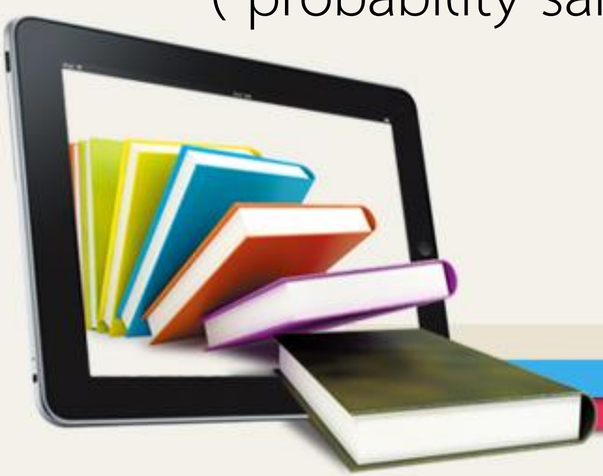
การเลือกตัวอย่าง(วิธีการสุ่มตัวอย่าง)

การเลือกกลุ่มตัวอย่างเพื่อจะมาเป็นตัวแทนในการศึกษาวิจัยมีเทคนิคการเลือกอยู่ 2 ประเภทได้แก่

1.การเลือกโดยไม่ใช้หลักทฤษฎีความน่าจะเป็น (Non probability sampling)

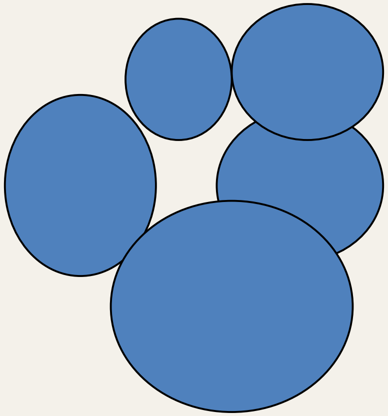
2. การเลือกกลุ่มตัวอย่างโดยใช้หลักทฤษฎีความน่าจะเป็น

(probability sampling)



วิธีการสุ่มตัวอย่าง (ต่อ)

1. **การสุ่มกลุ่มตัวอย่างโดยไม่อาศัยทฤษฎีความน่าจะเป็น (Non probability sampling)** เป็นการสุ่มกลุ่มตัวอย่างที่ไม่สามารถประมาณค่าความน่าจะเป็นของแต่ละหน่วยที่ถูกเลือกเป็นตัวอย่างได้ และไม่มีหลักประกันว่าทุกหน่วยของประชากรจะมีโอกาสได้รับเลือกเป็นกลุ่มตัวอย่างเท่า ๆ กัน



1. การสุ่มกลุ่มตัวอย่างโดยไม่อาศัยทฤษฎีความน่าจะเป็น (ต่อ)

วิธีการสุ่มกลุ่มตัวอย่างโดยไม่อาศัยทฤษฎีความน่าจะเป็น ทำได้หลายวิธีดังนี้

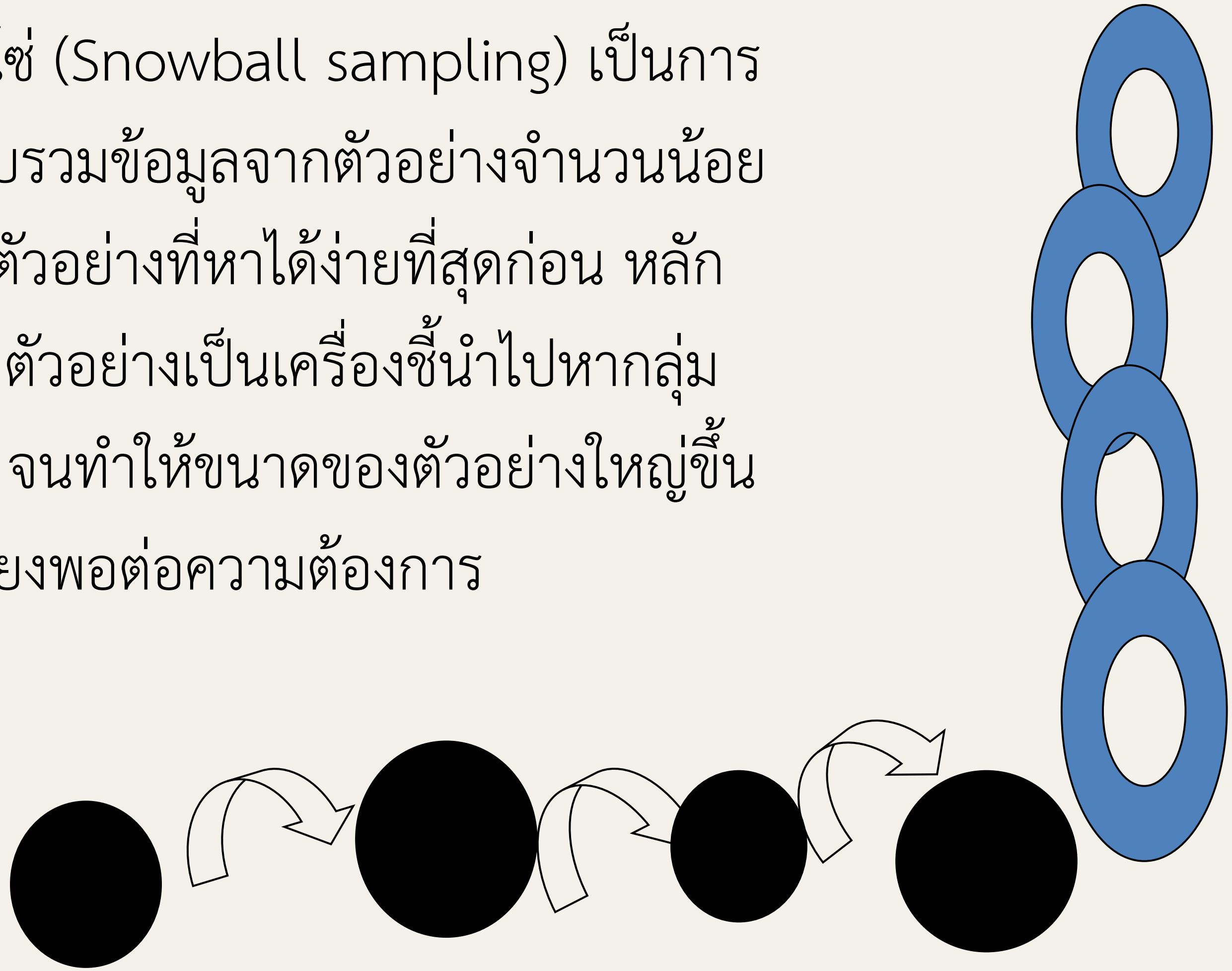
1. การสุ่มแบบบังเอิญ (Accidental sampling) เป็นการเลือกหน่วยประชากรที่พบหรือสามารถหาได้สะดวกจนครบตามจำนวนที่ต้องการ

2. การสุ่มแบบเจาะจงหรือสุ่มแบบมีจุดมุ่งหมาย (Purposive sampling) เป็นการสุ่มตัวอย่างตามจุดมุ่งหมายและชนิดของงานวิจัย

3. การสุ่มแบบโควต้า (Quota sampling) เป็นการสุ่มตัวอย่างโดยจำแนกประชากรออกเป็นส่วน ๆ ก่อนโดยมีหลักจำแนกว่าตัวแปรที่ใช้ในการจำแนกนั้น ควรจะมีความสัมพันธ์กับตัวแปรที่จะรวบรวม

1. การสุ่มกลุ่มตัวอย่างโดยไม่อาศัยทฤษฎีความน่าจะเป็น (ต่อ)

4. การสุ่มแบบลูกโซ่ (Snowball sampling) เป็นการสุ่มโดยการรวบรวมข้อมูลจากตัวอย่างจำนวนน้อย ๆ ซึ่งเป็นกลุ่มตัวอย่างที่หาได้ง่ายที่สุดก่อน หลักจากนั้นใช้กลุ่มตัวอย่างเป็นเครื่องชี้นำไปหากลุ่มตัวอย่างอื่น ๆ จนทำให้ขนาดของตัวอย่างใหญ่ขึ้นเรื่อย ๆ จนเพียงพอต่อความต้องการ



2. การสุ่มกลุ่มตัวอย่างโดยอาศัยทฤษฎีความน่าจะเป็น

เป็นวิธีการสุ่มตัวอย่างที่คำนึงถึงความน่าจะเป็นซึ่งสมาชิกแต่ละหน่วยของประชากรจะทราบโอกาสที่จะได้รับเลือกให้เป็นตัวอย่างของการวิจัย หมายถึง สมาชิกแต่ละหน่วยมีโอกาสได้รับการเลือกมาเป็นกลุ่มตัวอย่างเท่า ๆ กัน

วิธีการสุ่มกลุ่มตัวอย่างโดยอาศัยทฤษฎีความน่าจะเป็น ทำได้หลายวิธีดังนี้

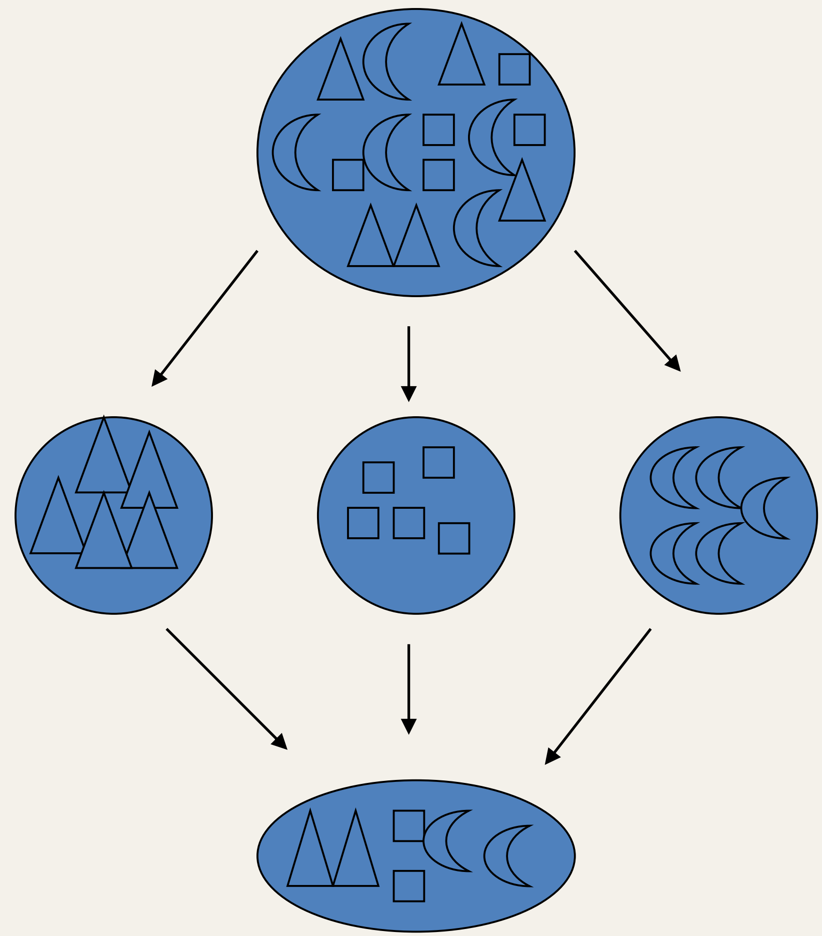
1. การสุ่มอย่างง่าย (Simple random sampling) คือ ให้แต่ละหน่วยตัวอย่างของประชากรมีโอกาสได้รับเลือกเท่า ๆ กัน และการเลือกแต่ละหน่วยตัวอย่างเป็นอิสระจากกัน สามารถดำเนินการได้ 2 วิธีดังนี้

   1.1 การจับฉลาก

   1.2 การใช้ตารางเลขสุ่ม

2. การสุ่มกลุ่มตัวอย่างโดยอาศัยทฤษฎีความน่าจะเป็น(ต่อ)

2. การสุ่มแบบแบ่งชั้น หรือตามระดับชั้น (Stratified random sampling) การสุ่มโดยวิธีนี้มักจะใช้ประชากรที่มีลักษณะแตกต่างกันมากจนสามารถแยกเป็นกลุ่มย่อย ๆ ตามลักษณะที่แตกต่างกัน แล้วทำการสุ่มตัวอย่าง



2. การสุ่มกลุ่มตัวอย่างโดยอาศัยทฤษฎีความน่าจะเป็น(ต่อ)

3. การสุ่มแบบมีระบบ (Systematic random sampling)

ใช้กรณีที่ประชากรจัดเรียงไว้อย่างเป็นระบบอยู่แล้ว เช่น เลขที่บ้าน เลขโทรศัพท์ เลขทะเบียนรถยนต์ เป็นต้น อาจสุ่มโดยกำหนดหมายเลขหรือช่วงห่างการสุ่มได้ เช่น

สมมติประชากรมี 15 รายต้องการตัวอย่าง 5 ราย

ช่วงห่างการสุ่ม = 15 / 5 = 3

วิธีการสุ่ม คือ จับฉลากจุดเริ่มต้นแล้วนับไปที ละ 5 หน่วยเป็นตัวอย่างที่สุ่มได้ตามต้องการ จากภาพ

จับฉลากเริ่มต้นได้ที่หมายเลข 4 จะได้ตัวอย่าง 5 ตัวอย่าง คือ ประชากร

หมายเลข 4, 6, 9, 13, และ 1

1
2
3
4
5
6
7
8
9
10
12
13
14
15

การกำหนดจำนวนตัวอย่างสำหรับวิธีการเลือกตัวอย่างแบบสุ่มอย่างง่าย

$$n = N / 1+Ne^2$$

เมื่อ n คือ ขนาดกลุ่มตัวอย่าง

N คือ ขนาดประชากร

e คือ ความคลาดเคลื่อนของกลุ่มตัวอย่าง

มาตรและการวัดข้อมูล

– **มาตรนามบัญญัติ (Nominal Scale)** เป็นมาตรวัดที่จำแนกข้อมูลที่สามารถจำแนก แบ่งชนิดหรือกลุ่มของตัวแปรเท่านั้น ไม่สามารถนำมาเปรียบเทียบในเชิงปริมาณได้ เช่น เพศ ศาสนา สถานการศึกษา สถานภาพสมรส เป็นต้น

– **มาตรเรียงลำดับ (Ordinal Scale)** เป็นมาตรวัดที่สามารถเปรียบเทียบได้ว่า สิ่งใดมากกว่าหรือน้อยกว่ากัน แต่ไม่สามารถนำมาเปรียบเทียบเชิงปริมาณ ในลักษณะ การบวกลบคูณหารได้ นั่นคือสามารถเรียงลำดับได้เพียงอย่างเดียว

– **มาตรอันตรภาค (Interval Scale)** เป็นมาตรการวัดที่คล้ายกับมาตรเรียงลำดับ แต่แตกต่างตรงที่ข้อมูลสามารถนำมาเปรียบเทียบเชิงปริมาณได้ คำนวณได้ นำมา บวก ลบ คูณ หาร ได้ แต่ว่าเลขศูนย์ในที่นี้ ถือว่าเป็นศูนย์ที่ไม่แท้จริง เช่น อุณหภูมิ เป็นต้น

– **มาตรอัตราส่วน (Ratio Scale)** เป็นมาตรการวัดที่คล้ายกับมาตรอันตรภาค แต่ต่างตรงที่เลขศูนย์ในที่นี้ ถือว่าเป็นศูนย์ที่แท้จริง เช่น ส่วนสูง น้ำหนัก รายได้ อายุ เป็นต้น

การประมาณค่า

1 วิธีการประมาณค่า
2 การประมาณค่าเฉลี่ยของประชากร
3 การประมาณค่าผลต่างของประชากร
4 การประมาณค่าสัดส่วนของประชากร

วิธีการประมาณค่า (ค่าเฉลี่ย , สัดส่วน)

1. **การประมาณค่าแบบจุด**

   สนใจศึกษาลักษณะประชากรเพียงค่าเดียว

2. **การประมาณแบบช่วง**

   สนใจศึกษาลักษณประชากรที่มีค่า 2 ค่า ซึ่งเกิดจากการส่วนเบี่ยงเบนมาตรฐานจากตัวอย่างมาหักออกและรวมเข้ากับค่าประมาณแบบจุดของพารามิเตอร์นั้นๆ

การประมาณค่าเฉลี่ยของประชากร μ

การประมาณค่าเฉลี่ยแบบจุด

สูตร $$\bar{\boldsymbol{x}} = \frac{\mathbf{1}}{\boldsymbol{N}}(X_1 + X_2 + X_3 + \ldots X_N)$$

หรือ $$= \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} X_{\boldsymbol{i}}$$

การประมาณค่าแบบช่วง หรือ ช่วงความเชื่อมั่นของ μ

เป็นการประมาณค่า(พารามิเตอร์) ของประชากรอยู่ระหว่างตัวเลข 2 จำนวน ค่าที่อยู่ระหว่างตัวเลข 2 จำนวน เรียกว่าช่วงความเชื่อมั่น (confidence interval)

- Lower confidence limit
- Upper confidence limit

ค่าประมาณแบบช่วงของ μ เมื่อข้อมูลของประชากรที่ต้องการประมาณมีการแจกแจงปรกติ หรือ อนุโลมได้ว่ามีการแจกแจงปกติ นั่นคือ $n \geq 30$

ช่วงความเชื่อมั่นของค่าเฉลี่ย $(\mu) = \overline{x} \pm t\left(\frac{\propto}{2}, n-1\right) \frac{s}{\sqrt{n}}$

ตัวอย่าง

จากการสอบถามจำนวนวันในรอบเดือนที่ผ่านมาแม่บ้าน 16 คน ที่เลือกมาจากแม่บ้านทั้งหมดที่อาศัยอยู่ในหมู่บ้านจัดสรรแห่งหนึ่งดูภาพยนต์โฆษณาก่อนรายการข่าวทางโทรทัศน์ปรากฎว่าได้จำนวนวันเฉลี่ย 25 วัน และส่วนเบี่ยงเบนมาตรฐาน 2 วัน จงสร้างช่วงความเชื่อมั่น ณ ระดับนัยสำคัญ 0.05 ของจำนวนวันเฉลี่ยที่แม่บ้านดูภาพยนตร์โฆษณา ถ้าประชากรของข้อมูลมีการแจกแจงปกติ

การประมาณค่าผลต่างระหว่างค่าเฉลี่ยของประชากร 2 ชุด (μ_1-μ_2)

ตัวประมาณแบบจุดของ (μ_1-μ_2)

$$X_1 - X_2 = \frac{1}{n} \sum_{i=1}^{n} X_{1i} - \frac{1}{n} \sum_{i=1}^{n} X_{2i}$$

เมื่อ μ_1และ μ_2 คือ ค่าเฉลี่ยของประชากรชุดที่หนึ่งและชุดที่สอง

X_1 และ X_2 คือ ค่าเฉลี่ยของตัวอย่างที่เลือกมาจากประชากรชุดที่หนึ่งและชุดที่สอง

n_1 และ n_2 คือจำนวนตัวอย่างที่เลือกมาจากประชากรชุดที่หนึ่งและชุดที่สอง

x_{1i} และ x_{2i} คือ ค่าจากตัวอย่างที่ i ซึ่งเลือกมาจากประชากรชุดที่หนึ่งและชุดที่สอง

ตัวอย่าง

หากบริษัทผู้ผลิตน้ำมันพืชปรุงอาหารต้องการทราบค่าประมาณแบบจุดของผลต่างระหว่างราคาขายปลีกของร้านค้าในต่างจังหวัดกับร้านค้าในกรุงเทพมหานคร ถ้าเลือกตัวอย่างร้านค้าในต่างจังหวัด 20 ร้าน ได้ราคาขายปลีกเป็นดังนี้ 35, 37, 34, 36, 35, 38, 34, 32, 34, 35, 38, 40, 34, 35, 38, 36, 34, 32, 35 และ37 บาท ถ้าเลือกตัวอย่างร้านค้าในกรุงเทพมหานคร 12 ร้าน ได้ราคาขายปลีกเป็นดังนี้ 34, 35 , 32 , 36, 34, 34, 33, 35, 36, 32 , 32 และ 34 บาท จงประมาณผลต่างระหว่างราคาขายปลีกน้ำมันพืชปรุงอาหารเฉลี่ยของร้านค้าในต่างจังหวัดกับร้านค้าในกรุงเทพมหานคร

การประมาณค่าผลต่างระหว่างค่าเฉลี่ยของประชากร 2 ชุด (μ_1-μ_2)

ตัวประมาณแบบช่วงของหรือช่วงความเชื่อมั่นของ μ_1-μ_2

$$(X_1 - X_2) \pm t_{\left(\frac{\propto}{2},\ n_1 + n_2 - 2\right)} \sqrt{S_p^{\,2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

โดย $S^2p = \dfrac{(n_1-1)s^2_1 + (n_2-1)s^2_2}{n_1 + n_2 - 2}$

ตัวอย่าง

ในการประมาณค่าผลต่างระหว่างค่าล่วงเวลาเฉลี่ยต่อสัปดาห์ของคนงานชายและหญิงของโรงงานแห่งหนึ่ง ซึ่งเลือกตัวอย่างมา 20 และ 25 คน ตามลำดับ จากการสอบถามเกี่ยวกับค่าล่วงเวลาต่อสัปดาห์ในสัปดาห์ที่ผ่านมาพบว่าค่าล่วงเวลาเฉลี่ยของคนงานชายและหญิงเท่ากับ 500 บาท และ 400 บาท ความแปรปรวนของค่าล่วงเวลาคนงานชายและหญิงเท่ากับ 100 บาทและ 50 บาทตามลำดับ ถ้าค่าล่วงเวลาของคนงานมีการแจกแจงปกติ จงหาค่าประมาณแบบช่วงของผลต่างระหว่างค่าล่วงเวลาเฉลี่ยต่อสัปดาห์คนงานชายและหญิง

ณ ระดับนัยสำคัญ 0.05

การประมาณค่าสัดส่วนของประชากร

การประมาณแบบจุดของ P

ใช้ตัวประมาณ $P = \dfrac{\text{จำนวนตัวอย่างที่มีลักษณะที่สนใจ}}{\text{จำนวนตัวอย่างทั้งหมด}}$

การประมาณแบบช่วงของ P

ใช้ตัวประมาณ $P \pm Z_{(\frac{\alpha}{2})} \sqrt{\frac{P(1-P)}{n}}$

ตัวอย่าง

จากการเลือกตัวอย่างผู้ใช้รถยนต์ส่วนบุคคลจำนวน 100 คน เพื่อสอบถามถึงยี่ห้อรถคันต่อไปถ้าเขาจะเปลี่ยนรถคันที่ใช้อยู่ปัจจุบัน ปรากฏว่ามีอยู่ 30 คน ที่ต้องการเปลี่ยนไปใช้รถยี่ห้อ A จงประมาณสัดส่วนของผู้ใช้รถส่วนบุคคลที่จะเปลี่ยนไปใช้รถยี่ห้อ A ในการเปลี่ยนรถครั้งต่อไปโดย การประมาณค่าแบบจุด และแบบช่วง ณ ระดับความเชื่อมั่น 90%

การประมาณผลต่างระหว่างสัดส่วนของประชากร

การประมาณแบบจุดของ P_1-P_2

ใช้ตัวประมาณ $P_1-P_2 = \dfrac{a_1}{n_1} - \dfrac{a_2}{n_2}$

การประมาณแบบช่วงของ P_1-P_2

ใช้ตัวประมาณ $$P1\text{-}P2 \pm Z_{(\frac{\propto}{2})} \sqrt{\frac{P_1(1-P_1)}{n_1}} + \sqrt{\frac{P_2(1-P_2)}{n_2}}$$

การทดสอบสมมติฐานเชิงสถิติ

1. การตั้งสมมติฐาน
2. ขั้นตอนการทดสอบ
3. การทดสอบเกี่ยวกับค่าเฉลี่ย
4. การทดสอบเกี่ยวกับสัดส่วน

การทดสอบสมมติฐาน

สมมติฐาน

คือ ความเชื่อของบุคคลใดบุคคลหนึ่งหรือกลุ่มบุคคล

สมมติฐานเป็นสิ่งที่คาดว่าจะเกิดขึ้น (ความเชื่ออาจจะจริงหรือไม่จริงก็ได้)

ตัวอย่าง

1.ผู้จัดการแผนกครัวแห่งหนึ่งเชื่อว่าน้ำหนักเฉลี่ยเนื้อที่สั่งมาจะหนักอย่างน้อย 1 กิโลกรัม

2.ฝ่ายต้อนรับลูกค้าของโรงแรมแห่งหนึ่งคาดว่าจะมีลูกค้าเข้ามาพักเฉลี่ยไม่เกิน 1,500 คนต่อวัน

3.บริษัทผลิตแก้วน้ำ เชื่อว่าแก้วที่ผลิตมาได้จะเสียไม่เกิน 3%

สรุป

การทดสอบความเชื่อของสิ่งที่สนใจ เรียกว่า **การทดสอบสมมติฐานทางสถิติ**

การตั้งสมมติฐานทางสถิติ

การกำหนดสมมติฐานเพื่อการทดสอบ ประกอบด้วยสมมติฐาน 2 ชนิด

1.สมมติฐานว่าง/สมมติฐานหลัก(Null Hypothesis) ใช้สัญลักษณ์ H_0

2.สมมติฐานแย้ง/สมมติฐานรอง(Alternative Hypothesis) ใช้สัญลักษณ์ H_1

ตั้งตรงข้าม

ตัวอย่างการตั้งสมมติฐาน

1.ผู้จัดการแผนกครัวแห่งหนึ่งเชื่อว่าน้ำหนักเนื้อที่สั่งมาจะหนัก**อย่างน้อย** 1 กิโลกรัม

: น้ำหนักเนื้อที่สั่งมาจะหนักเฉลี่ย**มากกว่าหรือเท่ากับ** 1 กิโลกรัม

: น้ำหนักเนื้อที่สั่งมาจะหนักเฉลี่ย**น้อยกว่า** 1 กิโลกรัม

2.ฝ่ายต้อนรับลูกค้าของโรงแรมแห่งหนึ่งคาดว่าจะมีลูกค้าเข้ามาพัก**ไม่เกิน** 1,500 คนต่อวัน

: ฝ่ายต้อนรับลูกค้าของโรงแรมแห่งหนึ่งคาดว่าจะมีลูกค้าเฉลี่ยเข้า**มาพักน้อยกว่าหรือเท่ากับ** 1,500 คนต่อวัน

: ฝ่ายต้อนรับลูกค้าของโรงแรมแห่งหนึ่งคาดว่าจะมีลูกค้าเฉลี่ยเข้ามาพัก**มากกว่า**1,500 คนต่อวัน

3.บริษัทผลิตแก้วน้ำ เชื่อว่าแก้วที่ผลิตมาได้จะเสีย**ไม่เกิน** 3 %

: บริษัทผลิตแก้วน้ำ เชื่อว่าแก้วที่ผลิตมาได้จะเสีย**น้อยกว่าหรือเท่ากับ**3 %

: บริษัทผลิตแก้วน้ำ เชื่อว่าแก้วที่ผลิตมาได้จะเสีย**มากกว่า** 3 %

แบบฝึกหัดการตั้งสมมติฐานทางสถิติ

1.บริษัททัวร์แห่งหนึ่งโฆษณาว่าการเดินทางจากกรุงเทพฯไปโรงแรมดุสิตธานีดีทูพัทยาใช้เวลา**เฉลี่ยไม่เกิน** 1 ชั่วโมง

2.เจ้าของร้านกาแฟต้องการทดสอบว่าคนวัยทำงานชอบทานกาแฟ**เฉลี่ยมากกว่า** 65% จริงหรือไม่

3.การท่องเที่ยวแห่งประเทศไทยเชื่อว่า สัดส่วนของคนไทยที่มีอายุตั้งแต่ 55 ปีขึ้นไป **ชอบไปเที่ยว**ต่างประเทศร้อยละ 0.44

4.นักวิจัยท่านหนึ่งเชื่อว่าสัดส่วนครอบครัวไทยควรมีบุตรไม่เกิน 2 คน**มีค่าอย่างน้อย** 0.68

5.ผู้จัดการครัวร้อนเชื่อว่า Lobster ที่สั่งมามีน้ำหนัก**เฉลี่ยตัวละมากกว่า** 1.5 กิโลกรัม

ความผิดพลาดในการทดสอบสมมติฐานเชิงสถิติ

- ความผิดพลาดประกอบด้วย 2 ประเภท

ความผิดพลาดประเภทที่ 1(Type I Error)

เป็นความผิดพลาดเนื่องจากการปฏิเสธ H_0 หรือไม่ยอมรับ H_0 เมื่อ H_0 เป็นจริง และ เรียกความผิดพลาดประเภทนี้ว่า **ระดับนัยสำคัญ (Level of significance)** และใช้สัญลักษณ์ α (Alpha)

ความผิดพลาดประเภทที่ 2(Type II Error)

เป็นความผิดพลาดเนื่องจากการยอมรับ H_0 โดยที่ H_0 ไม่เป็นจริง และใช้สัญลักษณ์ β (Beta)

		ความเป็นจริงของ H_0	
		H_0 เป็นจริง	H_0 เป็นเท็จ
การตัดสินใจ	ยอมรับ H_0	✓ $(1-\alpha)$	Type II error β
	ปฏิเสธ H_0	Type I error α	✓ $(1-\beta)$

ตัวอย่างความผิดพลาด : ประเภทที่ I และ II

- H_0 : รายได้เฉลี่ยของลูกค้าไม่เกิน 15,000 บาท

  หรือ $H_0 : \mu \leq 15,000$

- H_1 : รายได้เฉลี่ยของลูกค้ามากกว่า 15,000 บาท

  หรือ $H_1 : \mu > 15,000$

ถ้าเกิดความผิดพลาดประเภทที่ I

α = P(ปฏิเสธ H_0 | H_0 เป็นจริง)

β = P(ยอมรับ H_0 | H_0 ไม่เป็นจริง)

ถ้าเกิดความผิดพลาดประเภทที่ II

การลดความผิดพลาดทั้งสองประเภทสามารถทำได้โดยการเพิ่มขนาดตัวอย่าง

α แทน ระดับนัยสำคัญ(Level of Significance)

α= 0.05

ผลการทดสอบมีความถูกต้อง 95% และมีโอกาสผิดพลาด 5%

α สามารถเขียนให้อยู่ในรูปความเชื่อมั่นได้ว่า

ถ้า α= 0.05 ดังนั้นระดับความเชื่อมั่นจะเท่ากับ 1 – α

หรือเท่ากับ 1- 0.05 = 95

แบบฝึกหัด

จงเขียนให้อยู่ในรูประดับความเชื่อมั่นพร้อมอธิบายความหมาย

1. α= 0.01 2. α= 0.25 3. α= 0.10 4. α= 0.15 5. α= 0.25

ประเภทของการทดสอบสมมติฐาน

การทดสอบสมมติฐานแบ่งออกเป็น 2 ประเภท

การทดสอบแบบด้านเดียว(One-Sided Test)

แบบที่ 1 การทดสอบด้านขวา(A right-tailed Test)

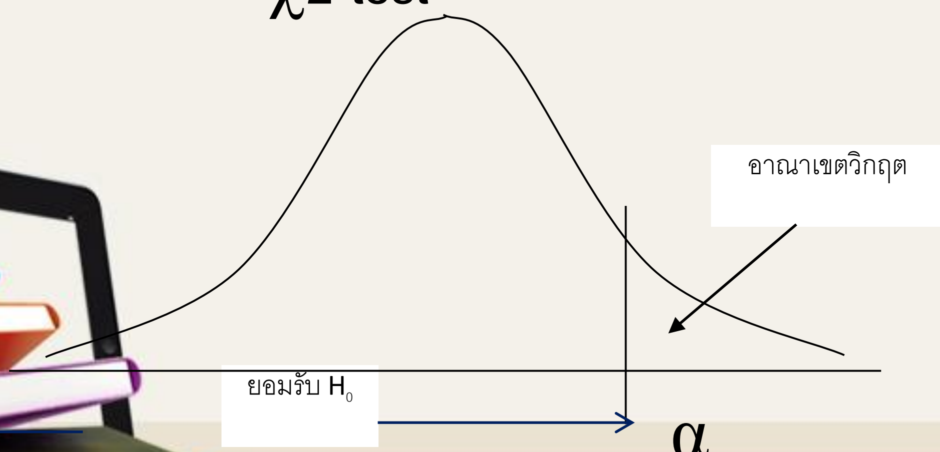
$H_0 : \mu \leq \mu_0$ และ $H_1 : \mu > \mu_0$ โดย แทน μ_0 ค่าคงที่

สถิติทดสอบ

Z-test

T-test

χ2-test



หมายเหตุ

1.การปฏิเสธ H_0 **คือ**

การยอมรับ H_1

2.เขตปฏิเสธขึ้นอยู่กับ

ค่า $\boldsymbol{\alpha}$

แบบที่ 2 การทดสอบด้านซ้าย(A Left-tailed Test)

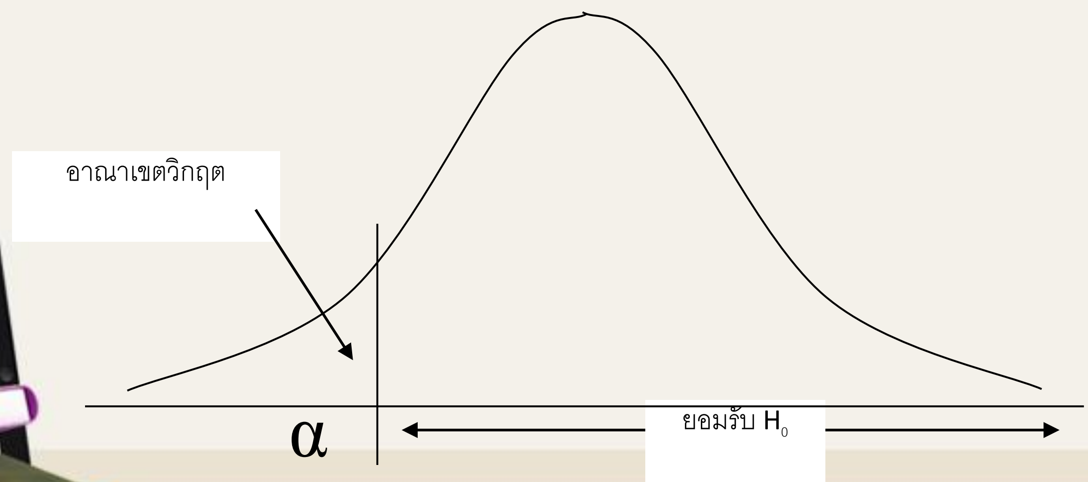
- $H_0 : \mu \geq \mu_0$ และ $H_1 : \mu < \mu_0$

สถิติทดสอบ

Z-test

T-test

χ2-test



การทดสอบแบบสองทาง (two-tailed test)

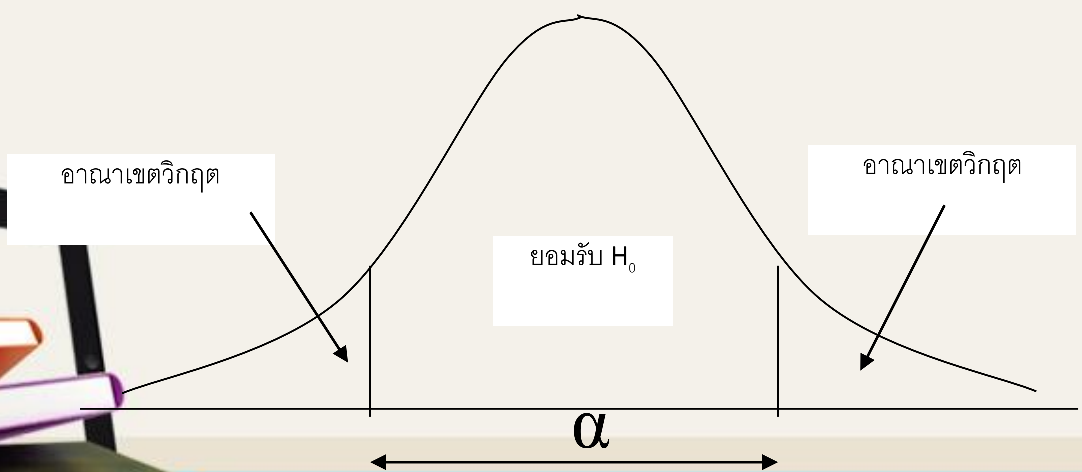
$$H_0 : \mu_1 = \mu_0$$
$$H_1 : \mu_1 \neq \mu_0$$

สถิติทดสอบ

Z-test

T-test

χ2-test



ขั้นตอนของการทดสอบสมมติฐาน

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ

เป็นขั้นตอนในการกำหนด H_0 และ H_1

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

โดยทั่วไปมักใช้ α = 0.10 , 0.05 ,0.01

ขั้นตอนที่ 3 กำหนดสถิติทดสอบ

กำหนดสถิติทดสอบจาก 3 กรณี

ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ

คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไข

ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง

หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

ขั้นตอนที่ 6 สรุปผลการทดสอบ

สรุปผลในการยอมรับ H_0 หรือ H_1

ตัวสถิติเพื่อการทดสอบ

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากร (ประชากรเดียว)

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$ ประชากรขนาดเล็ก หรือ $$z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$ ประชากรมากกว่า 30 ขึ้นไป

การทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างค่าเฉลี่ยของประชากร

(ประชากร 2 กลุ่ม)

กรณีที่ไม่ทราบค่าความแปรปรวน ความแปรปรวนทั้งสองกลุ่มไม่เท่ากัน และตัวอย่างขนาดใหญ่

ค่าเฉลี่ยตัวอย่างกลุ่มที่ 2

ค่าเฉลี่ยตัวอย่างกลุ่มที่ 1

ค่าเฉลี่ยประชากรกลุ่มที่ 2

ค่าเฉลี่ยประชากรกลุ่มที่ 1

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ความแปรปรวนกลุ่มที่ 1

ความแปรปรวนกลุ่มที่ 2

ตัวอย่างกลุ่มที่ 1

ตัวอย่างกลุ่มที่ 2

กรณีที่ทราบค่าความแปรปรวน $\sigma_1^2 = \sigma_2^2$

ตัวสถิติสำหรับการทดสอบ คือ

ค่าเฉลี่ยตัวอย่างกลุ่มที่ 2

ค่าเฉลี่ยตัวอย่างกลุ่มที่ 1

ค่าเฉลี่ยประชากรกลุ่มที่ 2

ค่าเฉลี่ยประชากรกลุ่มที่ 1

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S2_p\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s2_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

ตัวอย่างกลุ่มที่ 1

ตัวอย่างกลุ่มที่ 2

ตัวอย่าง

โรงงานอุตสาหกรรมผลิตหลอดไฟฟ้ายี่ห้อหนึ่งโฆษณาว่าหลอดไฟฟ้าที่ผลิตโดยบริษัทมีอายุการใช้งานเฉลี่ยมากกว่า 400 ชั่วโมง ตัวแทนจำหน่ายจึงต้องการพิสูจน์คำโฆษณาว่าจริงหรือไม่ จึงเลือกตัวอย่างหลอดไฟฟ้ามา 100 หลอด เพื่อวัดอายุการใช้งานปรากฏว่ามีอายุการใช้งานเฉลี่ยเป็น 390 ชั่วโมงและมีส่วนเบี่ยงเบนมาตรฐานของอายุการใช้งานเป็น 20 ชั่วโมง คำโฆษณาของบริษัทผู้ผลิตเป็นจริงหรือไม่ที่ระดับนัยสำคัญ 0.05

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ
เป็นขั้นตอนในการกำหนด H_0 และ H_1

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ
โดยทั่วไปมักใช้ α = 0.10 , 0.05 ,0.01

ขั้นตอนที่ 3 กำหนดสถิติทดสอบ
กำหนดสถิติทดสอบจาก 3 กรณี

ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ
คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไข

ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง
หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

ขั้นตอนที่ 6 สรุปผลการทดสอบ
สรุปผลในการยอมรับ H_0 หรือ H_1

ตัวอย่าง

ผู้ผลิตไอศครีมราย**หนึ่งเชื่อว่าไอศครีมของเขา มีปริมาณแคลอรี่เฉลี่ยในไอศครีมเป็น 500 แคลอรี่ต่อกรัม** เขาจึง**สุ่มไอศครีมก้อนละ 1 กรัมมา 25 ก้อน** ด้วยปริมาณแคลอรี่เฉลี่ยได้ 510 แคลอรี่ต่อ 1 กรัม และมีค่าเบี่ยงเบนมาตรฐาน 23 แคลอรี่ อยากทราบว่า สิ่งที่ผู้ผลิตเชื่อจริงหรือไม่ ที่ระดับนัยสำคัญ 0.10 ถ้าปริมาณแคลอรี่เฉลี่ยต่อไอศครีมหนัก 1 กรัมมีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติ

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ

เป็นขั้นตอนในการกำหนด H_0 และ H_1

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

โดยทั่วไปมักใช้ α = 0.10 , 0.05 ,0.01

ขั้นตอนที่ 3 กำหนดสถิติทดสอบ

กำหนดสถิติทดสอบจาก 3 กรณี

ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ

คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไข

ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง

หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

ขั้นตอนที่ 6 สรุปผลการทดสอบ

สรุปผลในการยอมรับ H_0 หรือ H_1

ตัวอย่าง

โรงงานแห่งหนึ่งมีเครื่องจักร 2 เครื่อง คือ a และ b ที่ใช้ในการผลิตตะปูซึ่งมีความยาว 0.5 นิ้ว แต่เครื่องจักรทั้งสองเครื่อง ส่วนใหญ่จะผลิตตะปูที่มีความยาวเกิน 0.5 นิ้ว เล็กน้อย อย่างไรก็ตามเจ้าของโรงงานนี้มีความเชื่อว่าความยาวเฉลี่ยของตะปูที่ผลิตได้จากเครื่องจักรทั้งสองไม่แตกต่างกัน เพื่อพิสูจน์เขาจึงสุ่มตะปูมาเครื่องละ 100 ตัว แล้ววัดความยาวของตะปูเหล่านั้น พบว่าตะปูที่ผลิตได้จากเครื่องจักร a มีค่าเฉลี่ยเท่ากับ 0.503 นิ้ว และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.004 นิ้ว และตะปูที่ผลิตได้จากเครื่องจักร b มีค่าเฉลี่ยเท่ากับ 0.501 นิ้ว และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.003 นิ้ว ความเชื่อของเจ้าของโรงงานถูกต้องหรือไม่ ที่ระดับนัยสำคัญ 0.05 ถ้าความยาวของตะปูที่ผลิตมีความแปรปรวนเท่ากัน

ขั้นตอนที่ 1 ตั้งสมมติฐานเพื่อการทดสอบ
เป็นขั้นตอนในการกำหนด H_0 และ H_1

ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ
โดยทั่วไปมักใช้ α = 0.10 , 0.05 ,0.01

ขั้นตอนที่ 3 กำหนดสถิติทดสอบ
กำหนดสถิติทดสอบจาก 3 กรณี

ขั้นตอนที่ 4 คำนวณค่าสถิติทดสอบ
คำนวณค่าสถิติทดสอบจากสูตรที่เลือกให้เหมาะสมกับเงื่อนไข

ขั้นตอนที่ 5 การสร้างเขตปฏิเสธสมมติฐาน/เปรียบเทียบค่าคำนวณกับค่าจากตาราง
หาค่าวิกฤติ และกำหนดขอบเขตของการปฏิเสธ

ขั้นตอนที่ 6 สรุปผลการทดสอบ
สรุปผลในการยอมรับ H_0 หรือ H_1

การทดสอบสมมติฐานเกี่ยวกับสัดส่วน

การทดสอบสมมติฐานเกี่ยวกับสัดส่วนของประชากรเดียว

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

ตัวอย่าง โรงงานผลิตกล่องพลาสติกแห่งหนึ่ง **ได้สุ่มตัวอย่างกล่องผลิตออกมา 400 กล่อง** เมื่อตรวจสอบดูแล้ว พบว่ามี 12 กล่องที่ไม่ได้มาตรฐาน จากผลการตรวจสอบนี้จะสรุปได้หรือไม่ว่า**ระบบการผลิตให้สินค้าที่ไม่ได้มาตรฐานมากกว่า 2 เปอร์เซ็นต์** กำหนด α = 0.05

ขั้นตอนที่ 1 ตั้งสมมติฐาน

$$H_0 : p \le 0.02$$

$$H_1 : p > 0.02$$

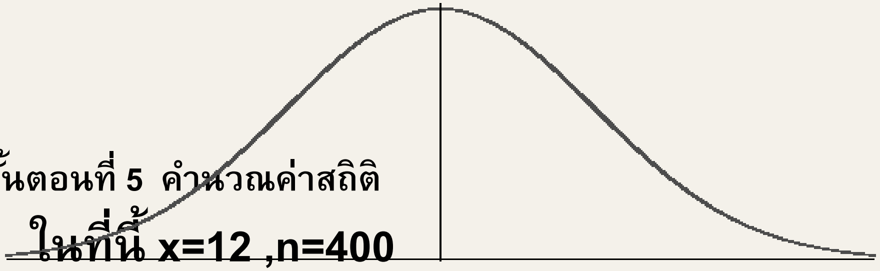
ขั้นตอนที่ 2 กำหนดระดับนัยสำคัญ

$$\alpha = 0.05, n = 400$$

ขั้นตอนที่ 3 เลือกตัวสถิติสำหรับการทดสอบ

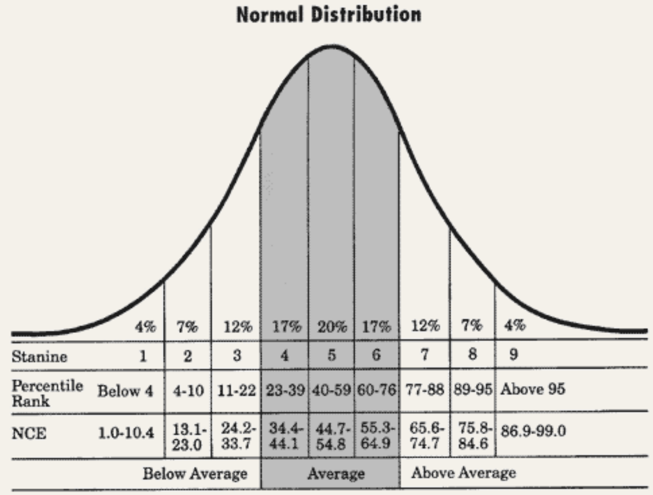
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

ขั้นตอนที่ 4 สร้างขอบเขตที่จะปฏิเสธสมมติฐาน



ขั้นตอนที่ 5 คำนวณค่าสถิติ

ในที่นี้ x=12 ,n=400



A Normal Distribution of Stanines, Percentile Ranks, Normal Curve Equivalents, and Performance Classifications

$$\therefore \hat{p} = \frac{x}{n} = \frac{12}{400} = 0.03$$

ขั้นตอนที่ 6 สรุปผลการทดสอบ

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.03-0.02}{\sqrt{\frac{0.02(1-0.02)}{400}}} = 1.429$$

Z=1.429 น้อยกว่า 1.645 จึงยอมรับ H0 แสดงว่าระบบการผลิตของโรงงานยังใช้ได้ดีเพราะยังผลิตสินค้าที่ไม่ได้มาตรฐานต่ำกว่า 2%

การทดสอบสมมติฐานเกี่ยวกับสัดส่วน

การทดสอบสมมติฐานเกี่ยวกับสัดส่วนของประชากรสองชุด