บทที่ 2
การศึกษาความสัมพันธ์ของตัวแปรต่าง ๆ

2.1 มาตรการวัด (Measurement Scales)

ข้อมูลต่าง ๆ จะใช้วิธีการทางสถิติแบบใดในการวิเคราะห์นั้น ต้องศึกษาให้ถ่องแท้ก่อนว่าข้อมูลที่เราสนใจศึกษานั้นอยู่ในมาตราวัดระดับไหน มาตราวัดบางระดับจะไม่สามารถนำข้อมูลที่เป็นตัวเลขมาทำการคำนวณทางคณิตศาสตร์ได้ ในขณะที่บางมาตราวัดสามารถจัดระดับได้ แต่ไม่สามารถบอกรายละเอียดของความแตกต่างของข้อมูลได้ เราสามารถจำแนกมาตรา การวัดข้อมูลออกได้เป็น 4 มาตรา คือ

1. มาตรานามบัญญัติ (Nominal Scales) เป็นมาตราการวัดที่ใช้กับข้อมูลที่มีลักษณะหยาบหรือต่ำที่สุด เป็นการกำหนดสัญลักษณ์หรือตัวเลขเพื่อจำแนกประเภทสิ่งของหรือคุณลักษณะต่าง ๆ เท่านั้น ไม่สามารถแสดงให้เห็นปริมาณมากน้อยหรือสูงต่ำแต่อย่างใด ดังนั้น จึงไม่สามารถนำตัวเลขเหล่านั้นมาบวก ลบ คูณ หาร ได้ เช่น เพศ ที่แบ่งออกได้เป็นแค่ 2 ประเภทเท่านั้น คือ ชายและหญิง โดยให้ 1 เป็นสัญลักษณ์แทนเพศชาย และ 2 เป็นสัญลักษณ์แทนเพศหญิงเท่านั้น ไม่สามารถนำสัญลักษณ์ที่เป็นตัวเลข 1 กับ 2 มาบวก ลบ คูณ หรือหารได้แต่อย่างใด หรือไม่สามารถสรุปได้ว่า 2 มากกว่า 1 เป็นต้น

2. มาตราอันดับ (Ordinal Scale) เป็นมาตราการวัดที่มีความละเอียดการวัดเพิ่มขึ้นหรือสูงกว่ามาตรานามบัญญัติ เพราะสามารถบอกลำดับและความแตกต่าง แต่ไม่สามารถบอกได้ว่าคุณลักษณะหรือคุณสมบัติเหล่านี้มีปริมาณมากน้อยกว่ากันเท่าใด กล่าวอีกนัยหนึ่ง ข้อมูลในระดับนี้ไม่สามารถนำมาคำนวณทางคณิตศาสตร์ได้เช่นเดียวกับมาตรานามบัญญัติ เช่น การประกวดนางสาวไทย เราได้อันดับที่ 1 คือ นางสาวไทย อันดับ 2 คือ รองนางสาวไทยคนที่ 1 และอันดับ 3 คือ รองนางสาวไทยคนที่ 2 เช่นนี้เรารู้ถึงอันดับและความแตกต่างของความสวยของนางสาวไทยทั้ง 3 คน เราไม่สามารถจัดจำแนกได้ว่า อันดับ 1 สวยมากกว่าอันดับ 2 เท่าใด เป็นต้น

3. มาตราช่วง (Interval Scale) ในระดับเราสามารถทราบได้ว่าสิ่งที่จะวัดมีช่วงห่างกันมากน้อยเท่าใด โดยแต่ละช่วงมาตรานี้มีค่าเท่า ๆ กัน ทำให้เราทราบถึงความแตกต่างที่ห่างกันเป็นช่วงได้ และค่าที่ได้จากการวัดสามารถนำมาคำนวณทางคณิตศาสตร์ได้ เช่น เราจะบอกความแตกต่างของน้ำร้อนระหว่าง 60° C กับ 80° C เท่ากับความแตกต่างระหว่าง 100° C กับ 120° C โดยดูจากช่วงที่ห่างกันเท่ากับ 20° C แต่อย่างไรก็ตามยังมาสามารถบอกได้ว่า ที่อุณหภูมิ 120° C มีความร้อนเป็น 2 เท่าของอุณหภูมิ 60° C ทั้ง ๆ ที่ 120 มีค่าเป็น 2 เท่าของ 60 ทั้งนี้เพราะว่าที่ 0° C ไม่ได้แปลว่าไม่มีความร้อนเลยหรือเราไม่สามารถบอกได้ว่าคนที่สอบได้คะแนน 100 คะแนนเต็มมีความรู้เป็นสองเท่ามากกว่าคนที่สอบได้คะแนน 50 เพราะคนที่สอบได้ 0 คะแนนไม่ได้แปลว่าไม่มีความรู้เลย ทั้งนี้เนื่องจากในมาตรา การวัดระดับนี้จะไม่มีศูนย์แท้นั่นเอง

4. มาตราอัตราส่วน (Ratio Scale) เป็นมาตราการวัดที่ดีที่สุดและวัดได้อย่างละเอียดที่สุด ตัวเลขที่วัดได้สามารถสื่อความหมายตรงตามค่าของสิ่งที่วัด และเป็นมาตราวัดที่ข้อมูลมีค่าเป็นศูนย์แท้ คือถ้าค่าตัวเลขที่วัดได้มีค่าเป็นศูนย์ ก็แปลว่าสิ่งที่วัดนั้นมีค่าเป็นศูนย์ด้วย ข้อมูลที่อยู่ในมาตราวัดระดับนี้ได้แก่ เวลา อายุ น้ำหนัก ความสูง ระยะทาง เป็นต้น ข้อมูลที่วัดได้สามารถนำมาทำการคำนวณทางคณิตศาสตร์ได้ ตัวอย่างเช่น คนที่มีน้ำหนัก 60 กิโลกรัม ย่อมแปลว่ามีน้ำหนักมากกว่า หรือหนักเป็นสองเท่าของคนที่มีน้ำหนักเพียง 30 กิโลกรัม ทั้งนี้เพราะว่าถ้าคนที่ไม่มีน้ำหนักเลยจะเท่ากับศูนย์ หรือระยะทางของถนนที่ยาว 20 เมตร ย่อมยาวเป็น 4 เท่าของถนนที่มีระยะทางยาวเพียง 5 เมตร เป็นต้น

2.2 ความหมายและชนิดของตัวแปร

ตัวแปร (Variable) หมายถึง คุณลักษณะหรือคุณสมบัติของสิ่งต่าง ๆ ที่สามารถแปรค่าได้ เช่น น้ำหนัก ส่วนสูง อายุ เพศ ผลสัมฤทธิ์ทางการเรียน ระดับสติปัญญา เชื้อชาติ เป็นต้น ในการวิจัยโดยทั่ว ๆ ไป มักจะแบ่งตัวแปรออกเป็น 2 ชนิดคือ

2.2.1 ตัวแปรต้น หรือตัวแปรอิสระ (Independent Variable) เป็นตัวแปรที่เป็นสาเหตุที่ก่อให้เกิดผล หรือก่อให้เกิดการแปรผันของปรากฏการณ์ เป็นตัวแปรที่ผู้วิจัยกำหนดหรือจัดกระทำได้ เพื่อศึกษาผลที่เกิดขึ้นจากตัวแปรนี้

2.2.2 ตัวแปรตาม (Dependent Variable) เป็นตัวแปรที่เป็นผลมากจากการเปลี่ยนแปลงค่าของตัวแปรอิสระ เป็นตัวแปรที่ผู้วิจัยมุ่งวัดเพื่อเป็นข้อมูลสำหรับนำมาวิเคราะห์เพื่อตอบคำถามของการวิจัยว่าเป็นผลมากจากสิ่งใด

นอกจากนี้ยังมีตัวแปรอื่นที่ผู้วิจัยมิได้มุ่งศึกษาโดยตรง แต่เป็นตัวแปรที่อาจมีผลกระทบต่อตัวแปรตามได้ ตัวแปรนี้เรียกว่า ตัวแปรเกิน หรือตัวแปรแทรกซ้อน (Extraneous Variable) หรือตัวแปรควบคุม (Control Variable) ผู้วิจัยจะต้องทราบว่ามีตัวแปรใดบ้างที่อาจส่งผลกระทบต่อตัวแปรตามและหาวิธีการควบคุมอิทธิพลจากตัวแปรแทรกซ้อน เหล่านี้ จากการตรวจสอบเอกสารตัวแปรเกินอาจเกิดขึ้นจากแหล่งต่าง ๆ ดังนี้

1. จากกลุ่มตัวอย่างหรือกลุ่มประชากร เป็นตัวแปรที่กลุ่มตัวอย่างมีมาก่อนจะมีการวิจัย เช่น อายุ เพศ ระดับสติปัญญา ความถนัด เชื้อชาติ บุคลิกภาพ สภาพครอบครัว เป็นต้น

2. จากวิธีดำเนินการทดลอง และการทดสอบในการวิจัยเชิงทดลอง เช่น ความผิดพลาดในวิธีดำเนินการ คุณภาพเครื่องมือที่ใช้ทดสอบ เวลาที่ใช้ทดสอบ เป็นต้น

3. จากแหล่งภายนอกหรือสิ่งแวดล้อม เช่นเสียงรบกวน สถานที่ไม่เหมาะสมและมีตัวแปรอีกประเภทหนึ่ง ที่อาจมีผลกระทบต่อตัวแปรตาม แต่เราไม่อาจรู้ได้ล่วงหน้าว่าจะเกิดขึ้นหรือไม่จึงไม่สามารถควบคุมได้ ตัวแปรเหล่านี้เรียกว่า ตัวแปรสอดแทรก (Intervening Variable) เช่น ภาวะสุขภาพ ความวิตกกังวล ความตื่นเต้น ความโกรธ แรงจูงใจ เป็นต้น

ตัวอย่างที่ 2.1

งานวิจัยเรื่อง เปรียบเทียบผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่ได้รับการสอน โดยวิธีสอน แบบค้นพบแบบนิรนัย

ตัวแปรต้น วิธีการสอน 2 วิธี คือ แบบค้นพบ และแบบนิรนัย

ตัวแปรตาม ผลสัมฤทธิ์ทางการเรียน

ตัวแปรเกิน ระดับสติปัญญา เพศ คุณภาพของแบบทดสอบ ฯลฯ

ตัวแปรสอดแทรก ความวิตกกังวล แรงจูงใจ แรงจูงใจใฝ่สัมฤทธิ์ ฯลฯ

ตัวอย่างที่ 2.2

งานวิจัยเรื่อง การเปรียบเทียบผลการประมาณค่าความสามารถด้วยวิธีการทดสอบแบบเทเลอร์รูปพีระมิด ขนาดขั้นคงที่ และรูปพีระมิดข้างตัด : การทดสอบทักษะกระบวนการทางวิทยาศาสตร์

ตัวแปรต้น วิธีการทดสอบแบบเทเลอร์ 2 รูปแบบ คือ รูปพีระมิดขนาดขั้นคงที่และรูปพีระมิด

ตัวแปรตาม ความสามารถทักษะกระบวนการทางวิทยาศาสตร์

ตัวแปรเกิน ระดับสติปัญญา เพศ จำนวนครั้งในการสอบ คุณภาพของแบบทดสอบ

ตัวแปรสอดแทรก แรงจูงใจใฝ่สัมฤทธิ์ การปรับตัว ความตื่นเต้น ฯลฯ

วิธีการควบคุมตัวแปรเกิน

1. ใช้สมาชิกที่มีลักษณะใกล้เคียงกันมากที่สุด (Homogeneous Group) ในลักษณะของตัวแปรเกิน แต่การควบคุมตัวแปรวิธีนี้จำกัดขอบเขตการอ้างอิงผลการวิจัยไปใช้ให้แคบลง

2. จัดสมาชิกเข้ากลุ่มโดยการสุ่ม (Random Assignment) การจัดสมาชิกเข้ากลุ่มโดยการสุ่มจะทำให้ โอกาสหรือ ความน่าจะเป็นที่ค่าตัวแปรตามของ กลุ่มทดลอง หรือกลุ่มควบคุมไม่แตกต่างกัน มีมากกว่าโอกาสที่จะแตกต่างกันก่อนทำการทดลอง

3. จับคู่สมาชิกบนพื้นฐานของตัวแปรเกิน แล้วจัดสมาชิกของแต่ละคู่เข้ากลุ่มโดยการสุ่ม ตัวแปรที่จะมาใช้เป็นเกณฑ์ในการจับคู่นั้น ควรจะเป็นตัวแปรที่มีความสัมพันธ์กับตัวแปรตามโดยมีค่าสัมประสิทธิ์สหสัมพันธ์ 0.50 หรือมากกว่า เช่น I.Q. มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียน

4. ควบคุมสภาพการณ์ในการทดลองให้มีความคงที่

5. นำตัวแปรเกินมาใช้ในการวิจัยโดยพิจารณาให้เป็นตัวแปรอิสระอีกตัวหนึ่ง

6. ควบคุมด้วยสถิติโดยการใช้การวิเคราะห์ความแปรปรวนร่วม (ANCOVA)

สรุป การควบคุมความแปรผันของตัวแปรหรือความแปรปรวน ซึ่งก็คือ การจัดกระทำให้ความแปรผันของตัวแปรนั้น ๆ มีค่ามากน้อยตามความต้องการอันจะทำให้เราทราบอิทธิพลของตัวแปรต่าง ๆ ที่เกี่ยวข้องกับตัวแปรที่เราต้องการศึกษาได้แน่ชัด หรือขจัดอิทธิพลนี้ออกไป จะช่วยให้การตีความผล การทดลอง หรือการวิจัยเป็นไปอย่างชัดเจนไม่คลุมเครือ และจะนำไปสู่การสรุป

ผลการวิจัยได้อย่างถูกต้อง ใกล้เคียงความเป็นจริง มากที่สุดซึ่ง ความแปรปรวน ในที่นี้หมายถึง ความแปรปรวนที่จะเกิดกับตัวแปรตามอันเป็นผลมาจากตัวแปรต้น

2.3 ความสัมพันธ์ระหว่างตัวแปร

ความสัมพันธ์ระหว่างตัวแปร หมายถึง การพรรณนาเกี่ยวกับการปรากฏขึ้นพร้อมกันของตัวแปร

ตัวอย่างที่ 2.3 การศึกษาสาเหตุในการเกิดอาชญากรรม

ตัวแปรต้น		**ตัวแปรตาม**
การศึกษา รายได้ ภูมิลำเนา	→	การประกอบอาชญากรรม

ตัวอย่างที่ 2.4 การศึกษาทัศนคติของนิสิตจุฬาฯ ที่มีต่อรัฐบาล

ตัวแปรต้น		**ตัวแปรตาม**
คณะ ชั้นปี	→	ทัศนคติ

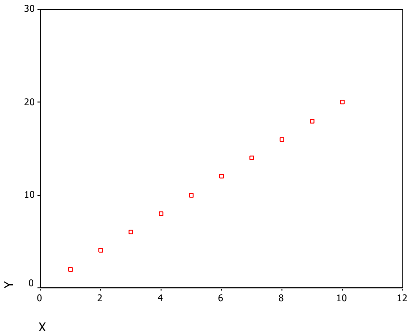
2.4 ลักษณะของความสัมพันธ์ของตัวแปร มี 2 ลักษณะ คือ

1. ความสัมพันธ์ทิศทางเดียวกัน เป็นความสัมพันธ์เมื่อค่าตัวแปรหนึ่งสูงขึ้น ตัวแปรอีกตัวหนึ่งก็จะมีค่าสูงตามกันไปด้วย หรือเรียกว่า ความสัมพันธ์เชิงบวก

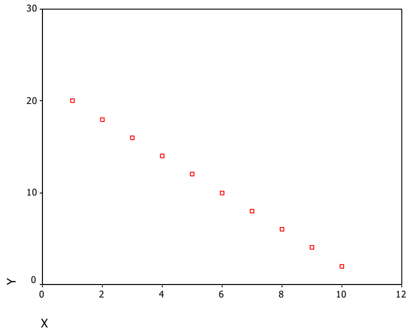
2. ความสัมพันธ์ทิศทางตรงกันข้าม เป็นความสัมพันธ์ในลักษณะเมื่อค่าตัวแปรหนึ่งเพิ่มขึ้น อีกตัวแปรหนึ่งจะลดลง หรือเรียกว่า ความสัมพันธ์เชิงลบ

แผนภาพการกระจาย

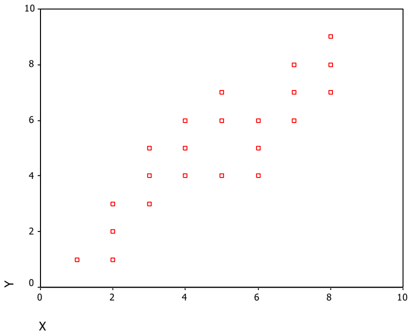
ในการศึกษาความสัมพันธ์ของตัวแปร 2 ตัวแปร คือตัวแปร x และตัวแปร y เมื่อนำค่าตัวแปรทั้งสองมาลงจุดคู่อันดับบนระนาบ โดยให้แกนนอนเป็น x และแกนตั้งเป็น y จะเรียกว่า แผนภาพการกระจาย จากแผนภาพการกระจายนี้จะทำให้เห็นความสัมพันธ์ของตัวแปรทั้งสองอย่างคร่าว ๆ ดังนี้



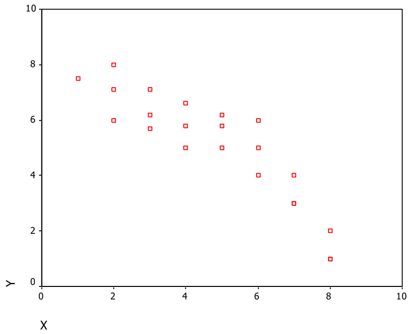
ก. ความสัมพันธ์ทิศทางเดียวกันอย่างสมบูรณ์



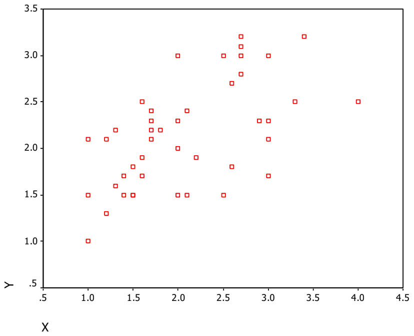
ข. ความสัมพันธ์ทิศทางตรงกันข้ามอย่างสมบูรณ์



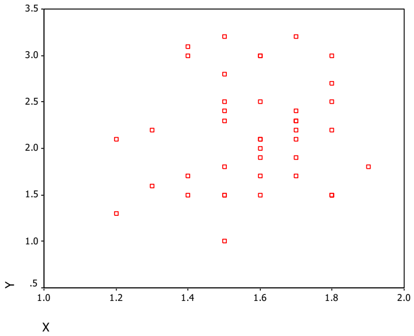
ค. ความสัมพันธ์ทิศทางเดียวกันในระดับสูง



ง. ความสัมพันธ์ทิศทางตรงกันข้ามในระดับสูง



จ. ความสัมพันธ์ทิศทางเดียวกันในระดับต่ำ



ฉ. ไม่มีความสัมพันธ์กัน

2.5 การวัดความสัมพันธ์ระหว่าง 2 ตัวแปร

2.5.1 การวัดความสัมพันธ์ระหว่าง 2 ตัวแปร ระดับมาตรานามบัญญัติ (Nominal Scale) การวัดโดยค่าสถิติไคสแควร์ (Chi-Square) และ Cramer's V

รูปแบบตารางการแจกแจงความถี่แบบ 2 ทาง (Contingency Table)

Row Variable	Column Variable				Total
	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
Total	c_1	c_2	...	c_c	$n = \sum r_i = \sum c_j$

เมื่อ O_{ij} คือ จำนวนความถี่ที่เกิดขึ้นจริงจากข้อมูลตัวอย่างในแถวที่ i หลักที่ j

โดยที่ $i = 1, 2, 3, \ldots, r$ และ $j = 1, 2, 3, \ldots, c$

การทดสอบความเป็นอิสระกันของตัวแปรทั้งสอง จะมีสมมติฐานในการทดสอบดังนี้

H_0 : ตัวแปรเป็นอิสระกันหรือไม่มีความสัมพันธ์กัน

H_1 : ตัวแปรไม่เป็นอิสระกันหรือมีความสัมพันธ์กัน

สถิติทดสอบ

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

เมื่อ E_{ij} คือ ความถี่ที่คาดว่าจะอยู่ในแถวที่ i หลักที่ j โดยที่ $E_{ij} = \frac{r_i c_j}{n}$

และจะปฏิเสธสมมติฐาน H_0 ก็ต่อเมื่อ $\chi^2 > \chi^2_{1-\alpha}$ โดยที่ค่า $\chi^2_{1-\alpha}$ เป็นค่าที่ได้จากการเปิดตารางไคสแควร์ ที่ระดับนัยสำคัญ α และองศาความเป็นอิสระ $df = (r-1)(c-1)$ นั่นคือ เมื่อ $\chi^2 > \chi^2_{1-\alpha}$ จะปฏิเสธสมมติฐาน H_0 ที่ว่าตัวแปรเป็นอิสระกันหรือไม่มีความสัมพันธ์กัน แล้วมายอมรับสมมติรอง H_1 ตัวแปรไม่เป็นอิสระกันหรือมีความสัมพันธ์กันอย่างมีนัยสำคัญ

การวัดระดับความสัมพันธ์ มี 4 ค่า ดังนี้

1. Contingency Coefficient
2. Phi and Cramer's V
3. Lambda
4. Uncertainty Coefficient

แต่ในรายวิชานี้จะขอกล่าวเพียง Contingency Coefficient ซึ่งเป็นค่าที่นิยมใช้กันมากโดยใช้วัดระดับความสัมพันธ์ได้แต่ไม่สามารถระบุทิศทางความสัมพันธ์ได้ โดยสามารถใช้กับตารางการแจกแจงขนาดใดก็ได้ โดยมีสถิติทดสอบเป็น $C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$ โดย $0 < C < 1$

โดยที่ χ^2 = Pearson Chi–Square

- ถ้า C มีค่าเข้าใกล้ 0 แสดงว่าตัวแปร 2 ตัวดังกล่าว มีความสัมพันธ์กันน้อย
- ถ้า C มีค่าเข้าใกล้ 1 แสดงว่าตัวแปร 2 ตัวดังกล่าว มีความสัมพันธ์กันสูง

ตัวอย่างที่ 2.5 จากตารางการแจกแจงความคิดเห็นของคนงานที่มีต่อการหยุดงานในวันเสาร์จำแนกตามเพศ เป็นดังนี้

เพศ	**ความคิดเห็น**			**รวม**
	เห็นด้วย	**ไม่เห็นด้วย**	**ไม่มีความคิดเห็น**	
ชาย	42	100	8	150
หญิง	74	20	6	100
รวม	116	120	14	250

จงทดสอบว่าที่ระดับนัยสำคัญ 0.05 เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานมีความสัมพันธ์กันหรือไม่

วิธีทำ **<u>สมมติฐาน</u>**

H_0 : เพศกับความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานเป็นอิสระหรือไม่มีความสัมพันธ์

H_1 : เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานไม่เป็นอิสระหรือมีความสัมพันธ์กัน

<u>สถิติทดสอบ</u>

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

เนื่องจาก $E_{ij} = \frac{r_i c_j}{n}$

$$E_{11} = \frac{r_1c_1}{n} = \frac{150\times116}{250} = 69.6$$
$$E_{12} = \frac{r_1c_2}{n} = \frac{150\times120}{250} = 72.0$$
$$E_{13} = \frac{r_1c_3}{n} = \frac{150\times14}{250} = 8.4$$
$$E_{21} = \frac{r_2c_1}{n} = \frac{100\times116}{250} = 46.4$$
$$E_{22} = \frac{r_2c_2}{n} = \frac{100\times120}{250} = 48.0$$
$$E_{23} = \frac{r_2c_3}{n} = \frac{100\times14}{250} = 5.6$$

ดังนั้น
$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$
$$= \frac{(42-69.6)^2}{69.6} + \frac{(100-72.0)^2}{72.0} + \frac{(8-8.4)^2}{8.4} +$$
$$\frac{(74-46.4)}{46.4} + \frac{(20-48.0)^2}{48.0} + \frac{(6-5.6)^2}{5.6}$$
$$= 54.63$$

จะปฏิเสธสมมติฐาน H_0 ก็ต่อเมื่อ $\chi^2 > \chi^2_{1-\alpha}$ โดยที่ค่า $\chi^2_{1-\alpha}$ ที่ได้จากการเปิดตารางที่องศาความเป็นอิสระ $df = (2-1)(3-1) = 2$ และระดับนัยสำคัญ 0.05 เท่ากับ 5.99

เนื่องจากว่า $\chi^2 = 54.63 > \chi^2_{1-\alpha} = 5.99$ จึงปฏิเสธสมมติฐาน H_0 ดังนั้น เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานไม่เป็นอิสระหรือมีความสัมพันธ์กัน

และสามารถวัดระดับความสัมพันธ์ได้ดังนี้

จาก $C = \sqrt{\dfrac{\chi^2}{\chi^2 + n}}$

จะได้ $C = \sqrt{\dfrac{54.63}{54.63 + 250}} = 0.4235$

นั่นคือ เพศและความคิดเห็นเกี่ยวกับการหยุดงานในวันเสาร์ของคนงานมีความสัมพันธ์กันในระดับ 0.4235 โดยทิศทางความสัมพันธ์สามารถวัดได้จากข้อมูลความถี่จากตาราง นั่นคือ

เพศชาย เห็นด้วย 42 ไม่เห็นด้วย 100 (น้อย, มาก)

เพศหญิง เห็นด้วย 74 ไม่เห็นด้วย 20 (มาก, น้อย)

แสดงว่าระดับความคิดเห็นของเพศชายและเพศหญิงไม่ไปในทิศทางเดียวกัน

2.5.2 การพรรณนาข้อมูลโดยการวัดความสัมพันธ์ระหว่าง 2 ตัวแปร ระดับมาตราอันดับ (Ordinal Scale)

ข้อมูลที่แต่ละกรณีศึกษาเป็นลำดับที่ : Spearman's Correlation Coefficient

The Spearman's Correlation Coefficient ใช้สัญลักษณ์ r_S หรือ ρ เป็นวิธีการที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร โดยที่ตัวแปรนั้นจะต้องอยู่ในรูปของข้อมูลมาตราเรียงอันดับ

Spearman นักจิตวิทยาชาวอังกฤษได้คิดวิธีการนี้ขึ้นในปี ค.ศ. 1906 จึงมีผู้เรียกวิธีการนี้อีกอย่างหนึ่งว่า Spearman's ρ (Spearman's Rho) หรือ Rho correlation ซึ่งการหาความสัมพันธ์ดังกล่าวนี้อาจเป็นความสัมพันธ์ระหว่างตัวแปรประเภทเดียวกันของประชากรกลุ่มเดียวกัน แต่มีการจัดอันดับสองครั้ง เช่น การจัดอันดับความสามารถของนักเรียนโดยครู 2 คน หรือความสัมพันธ์ระหว่างอันดับของบุคคลเดียวกันบนสองตัวแปร เช่น ผลการสอบกลางภาคกับผลการสอบปลายภาค เป็นต้น วิธีนี้มักใช้เปรียบเทียบข้อมูล 2 ชุด ว่ามีความสัมพันธ์กันมากน้อยเพียงใดเมื่อมีขนาดข้อมูลเล็ก ๆ หรือไม่มากนัก ประมาณ 20 – 30

ขั้นตอนการทดสอบ ทำตามลำดับดังนี้

1. จัดลำดับคะแนนแต่ละชุด โดยเรียงจากคะแนนต่ำสุดไปหาคะแนนสูงสุดหรือคะแนนสูงสุดไปหาคะแนนต่ำสุดก็ได้ แต่ต้องเป็นแบบเดียวกันทั้ง 2 ชุด ในกรณีที่คะแนนซ้ำกันให้ถืออันดับเฉลี่ยเป็นอันดับของคะแนนแต่ละคะแนนนั้น

ให้ R_x แทน อันดับของคะแนนชุด X

R_y แทน อันดับของคะแนนชุด Y

2. หาผลต่างระหว่างอันดับของคะแนน (d) ซึ่ง $d = R_x - R_y$
3. ยกกำลังสองของผลต่างระหว่างอันดับของคะแนน $\left(d^2\right)$
4. หาผลรวมทั้งหมดของกำลังสองของผลต่างระหว่างอันดับของคะแนน $\left(\sum d^2\right)$
5. คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) โดยใช้สูตร

$$\rho = 1 - \frac{6\sum d^2}{N\left(N^2 - 1\right)}$$

เมื่อ ρ แทน สัมประสิทธิ์สหสัมพันธ์จากผลต่างของอันดับของคะแนน

$\sum d^2$ แทน ผลรวมทั้งหมดของกำลังสองของผลต่างระหว่างอันดับของคะแนนแต่ละคู่

N แทน จำนวนคู่ของอันดับ

ตัวอย่างที่ 2.6 จากการแข่งขันฟุตบอลแห่งสหภาพยุโรปครั้งล่าสุด ซึ่งปรากฏว่าทีมฝรั่งเศสเป็นผู้ชนะเลิศผู้ฝึกสอนและผู้จัดการทีมฟุตบอลฝรั่งเศสได้จัดอันดับให้แก่ผู้เล่นแต่ละคนในทีม จากผู้เล่นดีเด่นอันดับที่ 1 ถึงอันดับที่ 11 ปรากฏผลดังนี้

ผู้เล่นตำแหน่ง	**อันดับของผู้เล่นที่ให้โดย**	
	ผู้ฝึกสอน	**ผู้จัดการ**
ผู้รักษาประตู	4	3
แบล็คขวา	9	6
แบล็คซ้าย	2	5
ฮาล์ฟขวา	5	8
เซนเตอร์ฮาล์ฟ	3	2
ฮาล์ฟซ้าย	8	4
ปีกขวา	10	1
ในขวา	11	11
ศูนย์หน้า	1	7
ในซ้าย	7	10
ปีกซ้าย	6	9

(ชัชวาลย์ เรืองประพันธ์, 2544 : 505)

จงหาค่าสัมประสิทธิ์สหสัมพันธ์ของความคิดเห็นของผู้ฝึกสอนและผู้จัดการทีมฝรั่งเศส

วิธีทำ

ผู้เล่นตำแหน่ง	**อันดับของผู้เล่นที่ให้โดย**		$d = R_x - R_y$	d^2
	ผู้ฝึกสอน : R_x	**ผู้จัดการ :** R_y		
ผู้รักษาประตู	4	3	1	1
แบล็คขวา	9	6	3	9
แบล็คซ้าย	2	5	-3	9
ฮาล์ฟขวา	5	8	-3	9
เซนเตอร์ฮาล์ฟ	3	2	1	1
ฮาล์ฟซ้าย	8	4	4	16
ปีกขวา	10	1	9	81
ในขวา	11	11	0	0
ศูนย์หน้า	1	7	-6	36
ในซ้าย	7	10	-3	9
ปีกซ้าย	6	9	-3	9
			รวม	**180**

แทนค่าในสูตร $\rho = 1 - \frac{6\sum d^2}{N\left(N^2 - 1\right)}$

$$= 1 - \frac{6(180)}{11\left(11^2 - 1\right)}$$

$$= 1 - \frac{1,080}{1,320}$$

$$= 1 - 0.8181$$

$$= 0.1819$$

นั่นคือ ความคิดเห็นของผู้ฝึกสอนและผู้จัดการทีมฝรั่งเศสมีความสัมพันธ์แบบตามกัน เท่ากับ 0.1819

2.5.3 การพรรณนาข้อมูลโดยการวัดความสัมพันธ์ระหว่างตัวแปรระดับมาตราช่วง (Interval Scale) และมาตราอัตราส่วน (Ratio Scale)

2.5.3.1 สมการถดถอยเชิงเส้นตรงอย่างง่าย (Simple linear regression equation) พยากรณ์ตัวแปรตามหนึ่งตัวจากการทราบค่าตัวแปรต้นอีกตัวหนึ่ง

กรณีที่ข้อมูลมีความสัมพันธ์กันแบบเส้นตรงนั้น ซึ่งตามรูปแบบสมการ $Y = a + bX$ ตัวแปร Y จะมีค่ามากหรือน้อยขึ้นอยู่กับตัวแปร X

โดยเรียก ตัวแปร Y ว่า ตัวแปรตาม

ตัวแปร X ว่า ตัวแปรอิสระ

a เป็นระยะตัดแกน Y ของเส้นแนวโน้ม (y - intercept)

b เป็นความชัน (slope) ของเส้นแนวโน้ม ซึ่งหมายถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงไป 1 หน่วย

สำหรับการหาสมการเส้นตรงหรือหาค่า a และ b วิธีการที่นิยมใช้อย่างแพร่หลาย ก็คือ **วิธีกำลังสองน้อยที่สุด** (Least square method) ดังนี้

ให้ Y แทน ค่าจริงของตัวแปร Y

$\hat{Y}$ แทน ค่าประมาณของตัวแปร Y

$e = Y - \hat{Y}$ แทน ค่าความคลาดเคลื่อนของการประมาณ Y

ในการพิจารณาความคลาดเคลื่อนของการประมาณทั้งหลาย จะไม่พิจารณาผลรวมของความคลาดเคลื่อน $\left(\sum e\right)$ ทั้งนี้เนื่องจาก $\sum e$ อาจมีค่าเท่ากับ 0 ทั้ง ๆ ที่การประมาณมีความคลาดเคลื่อน แต่จะพิจารณาผลรวมกำลังสองของความคลาดเคลื่อน $\left(\sum e^2\right)$ ซึ่งถ้า $\sum e^2$ มีค่าน้อย สามารถอธิบายได้ว่ามีความคลาดเคลื่อนน้อย และถ้า $\sum e^2$ มีค่ามากแล้ว ความคลาดเคลื่อนของการประมาณจะมากด้วย โดยหลักการของวิธีกำลังสองน้อยที่สุด จะเป็นการหาค่า a และ b ที่ทำให้ $\sum e^2$ มีค่าน้อยที่สุด ดังนี้

กำหนดให้
$$\begin{aligned} Q &= \sum e^2 \\ &= \sum \left(Y - \hat{Y}\right)^2 \\ &= \sum \left(Y - a - bX\right)^2 \\ \frac{\partial Q}{\partial a} &= -2\sum\left(Y - a - bX\right) = 0 \quad \ldots\ldots\ldots (1) \\ \frac{\partial Q}{\partial b} &= -2\sum\left(Y - a - bX\right)X = 0 \quad \ldots\ldots\ldots (2) \end{aligned}$$

จากสมการที่ (1) และสมการที่ (2) แก้สมการหาค่า a และ b ได้ดังนี้

$$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2} = \frac{\sum\left(X - \bar{X}\right)\left(Y - \bar{Y}\right)}{\sum\left(X - \bar{X}\right)^2}$$

$$a = \bar{Y} - b\bar{X}$$

ตัวอย่างที่ 2.7 ศึกษาความสัมพันธ์ระหว่างน้ำหนักที่ลดได้ด้วยกระบวนการลดน้ำหนักวิธีหนึ่งกับระยะเวลาที่ทำการลดน้ำหนักโดยทำการรวบรวมข้อมูลของผู้เข้ากระบวนการลดน้ำหนักจำนวน 10 คน ดังนี้

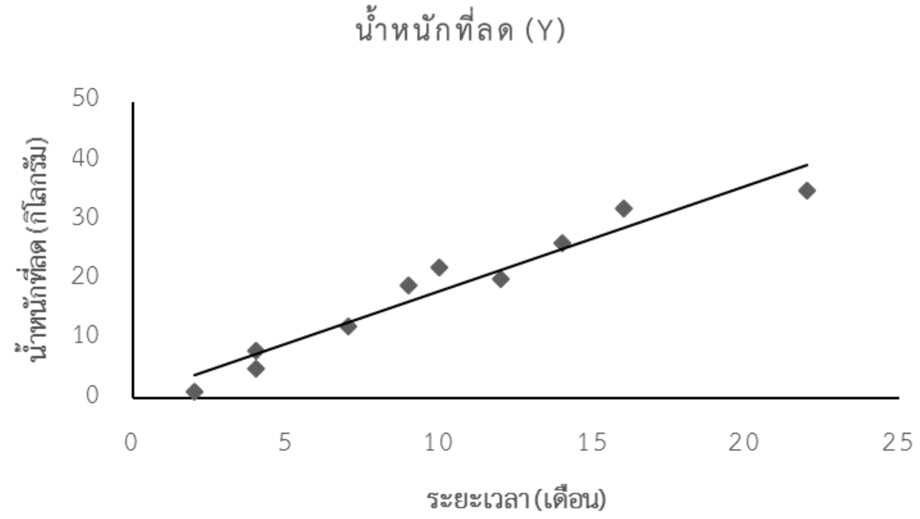
ระยะเวลา (เดือน)	4	16	14	2	10	22	9	12	4	7
น้ำหนักที่ลด (กิโลกรัม)	8	32	26	1	22	35	19	20	5	12

ก. จงสร้างสมการแสดงความสัมพันธ์เพื่อประมาณน้ำหนักที่ลดจากระยะเวลาที่ทำการลดน้ำหนัก
ข. คุณมยุรีจะใช้เวลาลดน้ำหนัก 6 เดือน จงประมาณว่าเธอจะลดน้ำหนักได้ประมาณกี่กิโลกรัม

วิธีทำ ก. สร้างสมการประมาณน้ำหนักที่ลด

ให้ X แทน ระยะเวลาที่ลดน้ำหนัก (เดือน)

Y แทน น้ำหนักที่ลด (กิโลกรัม)



จากแผนภาพการกระจายจะเห็นว่า X และ Y มีแนวโน้มความสัมพันธ์เป็นเส้นตรง

X	Y	X^2	XY
4	8	16	32
16	32	256	512
14	26	196	364
2	1	4	2
10	22	100	220
22	35	484	770
9	19	81	171
12	20	144	240
4	5	16	20
7	12	49	84
100	180	1,346	2,415

$\bar{X} = \frac{100}{10} = 10$ และ $\bar{Y} = \frac{180}{10} = 18$

$$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$= \frac{2,415 - 10(10)(18)}{1,346 - 10(10)^2}$$

$$= 1.7775$$

$$a = \bar{Y} - b\bar{X}$$

$$= 18 - 1.7775(10)$$

$$= 0.225$$

ดังนั้น สมการแสดงความสัมพันธ์ คือ $Y = 0.225 + 1.7775X$

ข. ถ้าลดน้ำหนัก 6 เดือน, $X = 6$

จะได้ว่า $Y = 0.225 + 1.7775(6) = 10.98$

นั่นคือ คุณมยุรีจะลดน้ำหนักที่ได้ประมาณ 10.89 กิโลกรัม

2.5.3.2 สัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson correlation coefficient)

สัมประสิทธ์สหสัมพันธ์ (Coefficient of correlation) เป็นค่าที่ใช้วัดระดับความสัมพันธ์ระหว่างตัวแปร 2 ตัว ในกรณีที่ตัวแปรทั้งสองมีระดับการวัดมาตราช่วง หรือมาตราอัตราส่วน การคำนวณหาสัมประสิทธิ์สหสัมพันธ์เพียร์สัน มีสูตรที่ใช้คำนวณหลายสูตรแต่วิธีการที่นิยมมากที่สุด คือ การคำนวณโดยการใช้คะแนนดิบหรือข้อมูลดิบ เพราะข้อมูลส่วนใหญ่ที่รวบรวมได้จะอยู่ในรูปของคะแนนดิบหรือข้อมูลดิบ การหาสหสัมพันธ์ใช้สัญลักษณ์ r หรือ r_{xy} มีสูตรดังนี้

$$r_{xy} = \frac{\sum XY - n\bar{X}\,\bar{Y}}{\sqrt{\left(\sum X^2 - n\bar{X}^2\right)\left(\sum Y^2 - n\bar{Y}^2\right)}}$$

เมื่อ r_{xy} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y

n แทน จำนวนคู่ของข้อมูล

$\bar{X}$ แทน ค่าเฉลี่ยของตัวแปร X

$\bar{Y}$ แทน ค่าเฉลี่ยของตัวแปร Y

$\sum XY$ แทน ผลรวมทั้งหมดของผลคูณของตัวแปร X และตัวแปร Y

$\sum X^2$ แทน ผลรวมของตัวแปร X แต่ละค่ายกกำลังสอง

$\sum Y^2$ แทน ผลรวมของตัวแปร Y แต่ละค่ายกกำลังสอง

ค่าสัมประสิทธิ์สหสัมพันธ์ r_{xy} มีค่าระหว่าง -1 ถึง +1 ซึ่งโดยทั่วไปมีความหมาย ดังนี้

1. r มีค่าเป็นบวก แสดงว่าตัวแปรทั้ง 2 มีความสัมพันธ์ในทิศทางเดียวกัน

   $r = 1$ แสดงว่าตัวแปรมีความสัมพันธ์กันอย่างสมบูรณ์ในทิศทางเดียวกัน
2. r มีค่าเป็น 0 แสดงว่าตัวแปรทั้ง 2 ไม่มีความสัมพันธ์กันเลย
3. r มีค่าเป็นลบ แสดงว่าตัวแปรทั้ง 2 มีความสัมพันธ์ในทิศทางตรงข้าม

   $r = -1$ แสดงว่าตัวแปรมีความสัมพันธ์กันอย่างสมบูรณ์ในทิศทางตรงข้าม

ถ้าต้องการพิจารณาค่าที่ละเอียดลงไปอาจใช้เกณฑ์อื่น ๆ เช่น

ขนาดของความสัมพันธ์	**ความหมาย**
+0.90 ถึง +1.00	ความสัมพันธ์ทางบวกสูงมาก
+0.70 ถึง +0.90	ความสัมพันธ์ทางบวกสูง
+0.50 ถึง +0.70	ความสัมพันธ์ทางบวกปานกลาง
+0.30 ถึง +0.50	ความสัมพันธ์ทางบวกน้อย
0.00 ถึง +0.30	ความสัมพันธ์ทางบวกน้อยมาก
0.00 ถึง -0.30	ความสัมพันธ์ทางลบน้อยมาก
-0.30 ถึง -0.50	ความสัมพันธ์ทางลบน้อย
-0.50 ถึง -0.70	ความสัมพันธ์ทางลบปานกลาง
-0.70 ถึง -0.90	ความสัมพันธ์ทางลบสูง
-0.90 ถึง -1.00	ความสัมพันธ์ทางลบสูงมาก

(Hinkle et al., 1998, p.120)

ตัวอย่างที่ 2.8 ข้อมูลของบริษัทรถยนต์ที่ใช้แล้ว โดยบันทึกข้อมูลอายุการใช้งานของรถยนต์ (X : หน่วยเป็นปี) และราคาขาย (Y : หน่วยเป็นพันบาท) ของรถจำนวน 6 คัน ดังนี้

X : อายุการใช้งาน (ปี)	4	10	2	1	3	5
Y : ราคาขาย (พันบาท)	895	125	1,395	1,795	1,245	695

จงคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอายุการใช้งานของรถยนต์และราคาขาย

วิธีทำ

X	Y	X^2	Y^2	XY
4	895	16	801,025	3,580
10	125	100	15,625	1,250
2	1,395	4	1,946,025	2,790
1	1,795	1	3,222,025	1,795
3	1,245	9	1,550,025	3,735
5	695	25	483,025	3,475
25	6,150	155	8,017,750	16,625

$$r_{xy} = \frac{\sum XY - n\bar{X}\,\bar{Y}}{\sqrt{\left(\sum X^2 - n\bar{X}^2\right)\left(\sum Y^2 - n\bar{Y}^2\right)}}$$

$$= \frac{16,625 - 6\left(\frac{25}{6}\right)\left(\frac{6,150}{6}\right)}{\sqrt{\left(155 - 6\left(\frac{25}{6}\right)^2\right)\left(8,017,750 - 6\left(\frac{6,150}{6}\right)^2\right)}}$$

$$= \frac{16,625 - 6(4.17)(1,025)}{\sqrt{\left(155 - 6(4.17)^2\right)\left(8,017,750 - 6(1,025)^2\right)}}$$

$$= \frac{16,625 - 25,645.5}{\sqrt{(155 - 104.3334)(8,017,750 - 6,303,750)}}$$

$$= \frac{-9,020.5}{\sqrt{(50.6666)(1,714,000)}}$$

$$= \frac{-9,020.5}{\sqrt{86,842,552}}$$

$$= \frac{-9,020.5}{9,318.935}$$

$$= -0.968$$

ดังนั้น ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างอายุการใช้งานของรถยนต์และราคาขายเท่ากับ -0.968 ซึ่งมีความสัมพันธ์กันค่อนข้างสูงทางลบ โดยถ้าอายุการใช้งานเพิ่ม ราคาขายของรถยนต์จะลดลง

แบบฝึกหัดท้ายบทที่ 2

1. ในการศึกษาความสัมพันธ์ระหว่างพฤติกรรมของเด็กกับวิธีการอบรมเลี้ยงดูในครอบครัว ผู้วิจัยสุ่มตัวอย่างเด็กมาจำนวน 60 คน ปรากฏข้อมูลดังตารางข้างล่างนี้

พฤติกรรมของเด็ก	วิธีการอบรมเลี้ยงดูแบบ			รวม
	เผด็จการ	ปล่อยปละ	ประชาธิปไตย	
เก็บตัว	13	4	4	21
แบบกลาง ๆ	5	9	10	24
แสดงตัว	2	2	11	15
รวม	20	15	25	60

1.1 จงทดสอบว่า พฤติกรรมของเด็กกับวิธีการอบรมเลี้ยงดูมีความสัมพันธ์กันหรือไม่ ทดสอบที่ระดับนัยสำคัญ 0.01

1.2 ถ้าพฤติกรรมของเด็กกับวิธีการอบรมเลี้ยงดูมีความสัมพันธ์กัน จงหาระดับความสัมพันธ์

2. กรรมการ 2 คน ได้จัดอันดับนักฟุตบอลตัวสำรอง 5 คน ดังนี้

นักฟุตบอลคนที่	อันดับที่ที่ให้โดยกรรมการคนที่ 1	อันดับที่ที่ให้โดยกรรมการคนที่ 2
1	3	4
2	2	1
3	5	3
4	4	2
5	1	5

จงหาความสัมพันธ์ของการจัดอันดับนักฟุตบอลตัวสำรองโดยกรรมการ 2 คนนี้

3. จากการศึกษาค่าบำรุงรักษารายปี (Y) ของเครื่องจักรชนิดหนึ่งที่มีอายุการใช้งาน (X) ดังต่อไปนี้

อายุการใช้งาน (ปี)	**ค่าบำรุงรักษารายปี (บาท)**
4	1,480
2	1,280
3	1,330
5	1,540
2	1,180
3	1,450
4	1,430
5	1,590
4	1,420
3	1,270

3.1 จงหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าบำรุงรักษารายปีและอายุการใช้งานของเครื่องจักร

3.2 จงหาสมการแสดงความสัมพันธ์เพื่อประมาณค่าบำรุงรักษารายปีจากอายุการใช้งานของเครื่องจักร

3.3 ถ้าเครื่องจักรเครื่องนี้มีอายุการใช้งาน 8 ปี จงพยากรณ์ค่าบำรุงรักษารายปี