การพัฒนาแบบจำลองทางสถิติ เพื่อการพยากรณ์สุขภาพ ผู้สูงอายุ

โดย กองพยากรณ์สุขภาพ
กรมอนามัย

บทนำ

1) ความเป็นมา

กองพยากรณ์สุขภาพ กรมอนามัย มีภารกิจสำคัญในการพัฒนาระบบและกลไกการพยากรณ์ด้านส่งเสริมสุขภาพและอนามัยสิ่งแวดล้อม ออกแบบเครื่องมือเพื่อการพยากรณ์สุขภาพ เทคโนโลยีและแนวทางการบริหารจัดการตามมาตรฐานและกฎหมายที่เกี่ยวข้อง ตลอดจนพัฒนาสมรรถนะกำลังคนด้านการพยากรณ์สุขภาพ รวมทั้งสนับสนุนการบูรณาการให้หน่วยงานในสังกัดกรมอนามัยและภาคเครือข่ายนำระบบพยากรณ์สุขภาพไปใช้ในการดำเนินงานให้มีประสิทธิภาพ โดยที่ผ่านมาได้มีการรวบรวมข้อมูลความคิดเห็นจากผู้เข้าร่วมการประชุมเชิงปฏิบัติการที่ผ่านมาของกองพยากรณ์สุขภาพ ซึ่งมีข้อเสนอแนะต่อทีมงานกองพยากรณ์สุขภาพ ในการจัดกิจกรรมการพัฒนาศักยภาพในครั้งต่อไป คือ การพยากรณ์สุขภาพ ด้านส่งเสริมสุขภาพและอนามัยสิ่งแวลด้อมในรายประเด็น เช่น ผลกระทบต่อสุขภาพรายกลุ่มวัย

ในปีงบประมาณ 2567 กองพยากรณ์สุขภาพ ได้ดำเนินโครงการพัฒนาแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพของผู้สูงอายุ ที่ให้ความสำคัญกับการประเมินคาดการณ์ด้านส่งเสริมสุขภาพและอนามัยสิ่งแวดล้อม โดยใช้ข้อมูลที่เกี่ยวข้องกับกลุ่มวัยสูงอายุ และใช้เครื่องมือแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพ ทั้งนี้ สิ่งสำคัญในการดำเนินงานนี้คือ การพัฒนาศักยภาพบุคลากรกรมอนามัยให้มีความรู้ความเข้าใจ มีทักษะด้านการวิเคราะห์ข้อมูลและการพยากรณ์สุขภาพ ที่สามารถนำไปประยุกต์ใช้ในงานที่เกี่ยวข้องและนำไปสู่การขับเคลื่อนการดำเนินงานด้านส่งเสริมสุขภาพและอนามัยสิ่งแวดล้อมของกรมอนามัยได้อย่างมีประสิทธิภาพ

กองพยากรณ์สุขภาพ กรมอนามัย จึงได้จัดประชุมเชิงปฏิบัติการพัฒนาแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพของผู้สูงอายุ ในวันที่ 18 – 20 ธันวาคม 2566 ณ ห้องประชุม 1 กองประเมินผลกระทบต่อสุขภาพ กรมอนามัย และผ่านระบบสื่ออิเล็กทรอนิกส์ โดยมีวัตถุประสงค์เพื่อเสริมสร้างความรู้ ความเข้าใจแก่บุคลากรกรมอนามัย เกี่ยวกับการจัดการและบริหารคุณภาพข้อมูล และการใช้เครื่องมือแบบจำลองทางสถิติ ในการวิเคราะห์ข้อมูลเพื่อการพยากรณ์สุขภาพ รวมถึงแลกเปลี่ยนเรียนรู้และฝึกปฏิบัติโดยใช้เครื่องมือทางสถิติต่าง ๆ และตัวอย่างข้อมูลในการพยากรณ์สุขภาพได้ สำหรับรูปแบบการประชุมเชิงปฏิบัติการในครั้งนี้ มีทั้งการบรรยาย และฝึกปฏิบัติ มีผู้เข้าร่วมประชุมประกอบด้วยเจ้าหน้าที่จากหน่วยงานส่วนกลางกรมอนามัย และส่วนภูมิภาค (ศูนย์อนามัยที่ 1 – 12 และ สถาบันพัฒนาสุขภาวะเขตเมือง รวมประมาณ 60 คน โดยได้เรียนรู้จากวิทยากร ผู้เชี่ยวชาญจากหน่วยงานที่เกี่ยวข้องและสถาบันการศึกษา โดยผลที่คาดว่าจะได้รับ คือผู้เข้าร่วมการประชุมมีความรู้ความเข้าใจเกี่ยวกับการพัฒนาแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพ และสามารถประยุกต์ใช้ความรู้เพื่อการพยากรณ์ด้านส่งเสริมสุขภาพและอนามัยสิ่งแวดล้อมได้ ต่อไป

2. วัตถุประสงค์

2.1 เพื่อเสริมสร้างความรู้ ความเข้าใจแก่บุคลากรกรมอนามัย เกี่ยวกับการบริหารคุณภาพข้อมูลเพื่อการพยากรณ์สุขภาพ

2.2 เพื่อเสริมสร้างความรู้ ความเข้าใจแก่บุคลากรกรมอนามัย เกี่ยวกับการใช้เครื่องมือทางสถิติในการวิเคราะห์ข้อมูลเพื่อการพยากรณ์ และการแปลผลทางสถิติ

2.3 เพื่อให้เกิดการแลกเปลี่ยนเรียนรู้ และฝึกปฏิบัติใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพ

3. วิทยากร

3.1 ผศ.ดร.ณัฐนารี เอมยงค์ ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดลล

3.2 ดร.จิราลักษณ์ นนทารักษ์ ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิด

3.3 นายแพทย์ศุภณัฐ วงศานุพัทธ์ หัวหน้ากลุ่มพัฒนาระบบข่าวกรองและเฝ้าระวังโรคไม่ติดต่อ (iNCD) กองระบาดวิทยา กรมควบคุมโรค

3.4 นายฤทธิไกร นามเกษ นักวิชาการสาธารณสุขปฏิบัติการ กองระบาดวิทยา กรมควบคุมโรค

4. รูปแบบการประชุม

4.1 การบรรยาย และฝึกปฏิบัติ มีทั้ง onsite และ online ผ่านระบบออนไลน์ Zoom Meetings https://zoom.us/j/99089607689?pwd=VDlGR2NYd3pCVnpJbDNnTWZGMkRkQT09#succe

5. ตารางหัวข้อการอบรม วัตถุประสงค์ และผลที่คาดว่าจะได้รับ

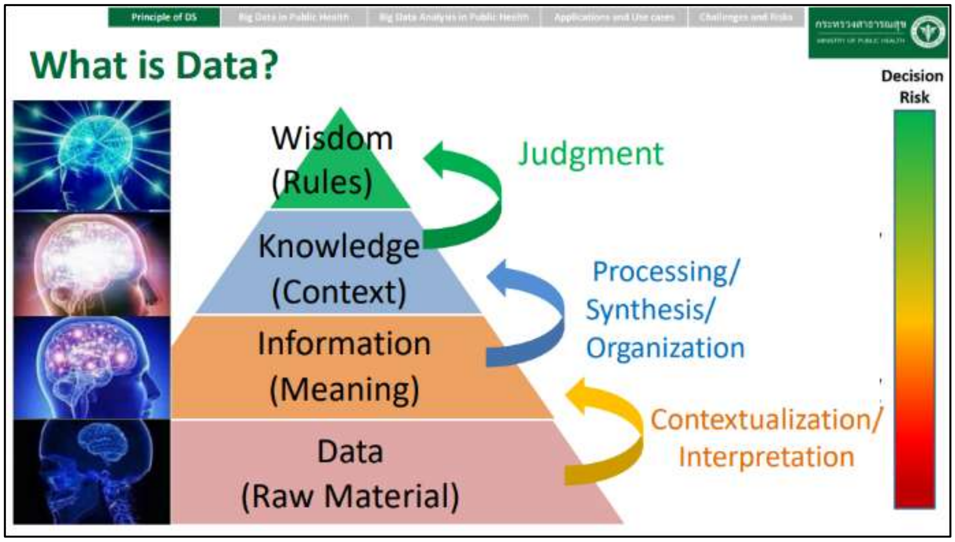
	หัวข้อ	**วัตถุประสงค์**	**ผลที่คาดว่าจะได้รับ**
18 ธ.ค. 66 09.00 – 09.30 น.	- พิธีเปิดการประชุม		
09.30 – 10.30 น. นายแพทย์ศุภณัฐ วงศานุพัทธ์ กองระบาดวิทยา กรมควบคุมโรค	**1) หลักวิธีคิดและแนวทางการวิเคราะห์ข้อมูลที่อาศัยการขับเคลื่อนการดำเนินงานด้วยข้อมูล (Data Driven Mindset)**	เพื่อให้ผู้เรียนมีความเข้าใจเกี่ยวกับหลักวิธีคิดและแนวทางการวิเคราะห์ข้อมูลที่อาศัยการขับเคลื่อนการดำเนินงานด้วยข้อมูล	ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับหลักวิธีคิดและแนวทางการวิเคราะห์ข้อมูลที่อาศัยการขับเคลื่อนการดำเนินงานด้วยข้อมูล เพื่อให้ได้ข้อมูลที่พร้อมใช้สำหรับงานพยากรณ์สุขภาพ
10.30 – 12.00 น. นายแพทย์ศุภณัฐ วงศานุพัทธ์ กองระบาดวิทยา กรมควบคุมโรค	**2) ความรู้เบื้องต้นเกี่ยวกับการวิเคราะห์ข้อมูลเชิงพยากรณ์ (Predictive Analytics) และแบบจำลองทางสถิติ**	เพื่อให้ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูล การหาความสัมพันธ์ และการวิเคราะห์เชิงพยากรณ์ ด้วยวิธีทางสถิติ รวมถึงเทคนิคการสร้างแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพ	ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูล การหาความสัมพันธ์ และการวิเคราะห์เชิงพยากรณ์ ด้วยวิธีทางสถิติ รวมถึงเทคนิคการสร้างแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพ

	หัวข้อ	วัตถุประสงค์	ผลที่คาดว่าจะได้รับ
13.00 – 16.00 น. นายฤทธิไกร นามเกษ กองระบาดวิทยา กรมควบคุมโรค	**3) การใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพ**	เพื่อให้ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับเครื่องมือการพยากรณ์สุขภาพ	ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับเครื่องมือการพยากรณ์สุขภาพ และสามารถนำความรู้ไปประยุกต์ใช้กับข้อมูลด้านส่งเสริมสุขภาพและอนามัยสิงแวดล้อมได้
19 ธ.ค. 66 09.30 – 12.00 น. ดร.จิราลักษณ์ นนทารักษ์ ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ ม.มหิดล	**4) การใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพผู้สูงอายุด้วย Software ที่เหมาะสม**	เพื่อให้ผู้เรียน ได้เรียนรู้หลักการการใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพ จากข้อมูลตัวอย่าง ที่เกี่ยวกับกลุ่มวัยสูงอายุ	ผู้เรียน มีความรู้เบื้องต้นเกี่ยวกับหลักการการ ใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพจากข้อมูลตัวอย่างเกี่ยวกับ กลุ่มวัยสูงอายุ
13.00 – 16.00 น. ดร.จิราลักษณ์ นนทารักษ์ ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ ม.มหิดล	**5) การใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพผู้สูงอายุด้วย Software ที่เหมาะสม และฝึกปฏิบัติ**	เพื่อให้ผู้เรียนได้ฝึกปฏิบัติใช้เครื่องมือสำหรับการพยากรณ์สุขภาพผู้สูงอายุ	ผู้เรียนสามารถใช้เครื่องมือสำหรับการพยากรณ์ สุขภาพได้
20 ธ.ค. 66 09.30 – 11.30 น. ผศ.ดร.ณัฐนารี เอมยงค์ ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ ม.มหิดล	**6) เทคนิคการแปลผลจากการวิเคราะห์และการพยากรณ์ข้อมูลด้วยเครื่องมือทางสถิติ**	เพื่อให้ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับหลักการ และวิธีในการแปลผลข้อมูล จากการวิเคราะห์หรือการพยากรณ์ข้อมูล รวมทั้งการนำเสนอผลจากการวิเคราะห์ข้อมูลทางสถิติ	ผู้เรียนได้มีความรู้เกี่ยวกับการแปลผลและการนำเสนอ ผล จากการวิเคราะห์ข้อมูลทางสถิติ
11.30 - 12.00	สรุปผลการประชุม ฯ และปิดการประชุม	เพื่อสรุปผลภาพรวมและรับฟังความคิดเห็นจาก ผู้เข้าร่วมประชุม ต่อภาพรวมการประชุมฯ	ได้ผลสรุปการประชุมและ ผู้เข้าร่วมสามารถนำความรู้ไปประยุกต์ใช้ รวมทั้งได้ข้อเสนอแนะต่อการพัฒนาศักยภาพการพยากรณ์สุขภาพต่อไป

บรรยาย เรื่อง หลักวิธีคิดและแนวทางการวิเคราะห์ข้อมูลที่อาศัยการขับเคลื่อนการดำเนินงานด้วยข้อมูล (Data Driven Mindset) และความรู้เบื้องต้นเกี่ยวกับการวิเคราะห์ข้อมูลเชิงพยากรณ์ (Predictive Analytics) และแบบจำลองทางสถิติ

โดย นายแพทย์ศุภณัฐ วงศานุพัทธ์ หัวหน้ากลุ่มพัฒนาระบบข่าวกรองและเฝ้าระวังโรคไม่ติดต่อ (iNCD) กองระบาดวิทยา กรมควบคุมโรค

การขับเคลื่อนองค์กรสาธารณสุขด้วยข้อมูล (Data-Driven Public Health) หมายถึง การใช้ข้อมูลและหลักฐานทางวิทยาศาสตร์เป็นเครื่องมือหลักในการตัดสินใจ วางแผน กำหนดนโยบาย ดำเนินการ และประเมินผลงานขององค์กรสาธารณสุข เพื่อให้บรรลุเป้าหมายในการพัฒนาสุขภาพ อย่างมีประสิทธิภาพและยั่งยืน



- **What is Data ?**

Data หรือข้อมูล คือ ข้อเท็จจริงที่ยังไม่ได้รับการจัดการ จัดรูปแบบ หรือผ่านกระบวนการจัดการข้อมูล

1. ข้อมูลดิบ (Data) คือ ข้อเท็จจริงที่ยังไม่ได้รับการจัดการ จัดรูปแบบ หรือผ่านกระบวนการใดๆ เปรียบเสมือนวัตถุดิบดิบที่ยังไม่ได้ผ่านการปรุงแต่ง ข้อมูลเหล่านี้สามารถอยู่ในรูปแบบต่างๆ เช่น ตัวอักษร ตัวเลข รูปภาพ เสียง ข้อความ หรือแม้แต่การสังเกต

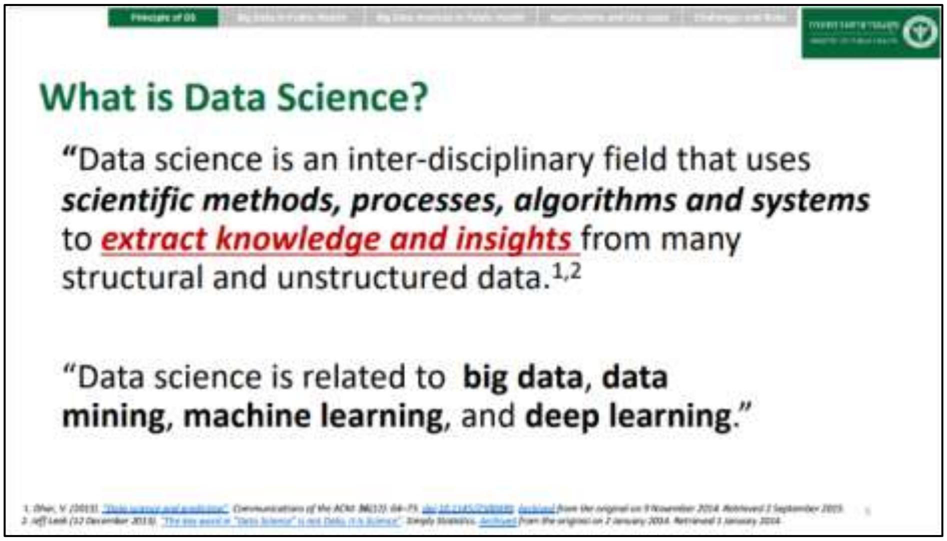
2. ข้อมูล (Information) คือ ข้อมูลที่ผ่านการประมวลผล จัดกลุ่ม และตีความ มีความหมาย และสามารถนำไปใช้ประโยชน์ได้

3. ความรู้ (Knowledge) คือ ข้อมูลที่ผ่านการตีความ วิเคราะห์ และประมวลผล จนเข้าใจความหมายและสามารถนำไปใช้ประโยชน์ได้

4. ภูมิปัญญา (Wisdom) คือ ความรู้ที่ลึกซึ้ง เกิดจากประสบการณ์ การเรียนรู้ และการไตร่ตรอง

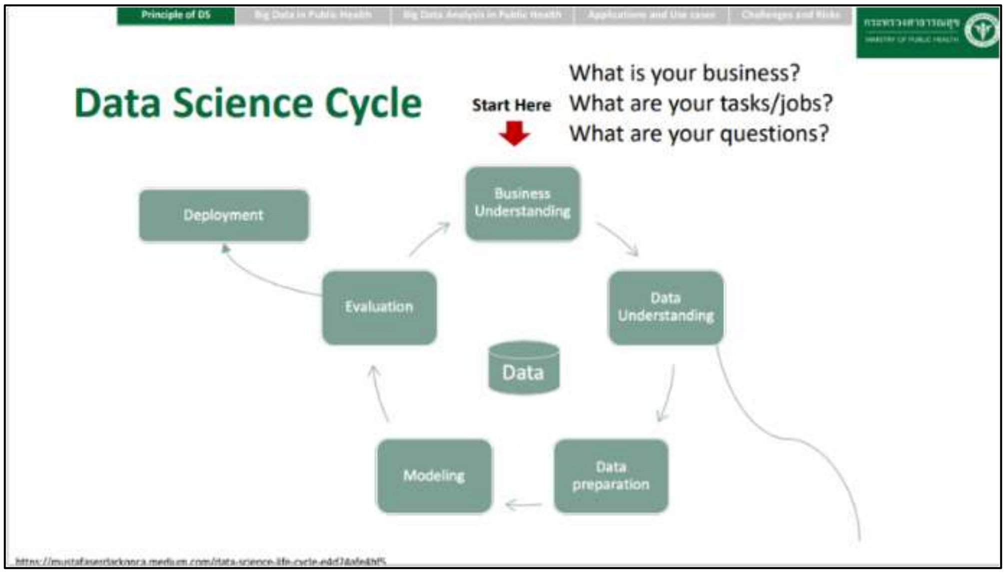
ระดับของข้อมูลมีความสำคัญและมีความสัมพันธ์กัน

- ข้อมูลดิบ (Data) จะถูกนำไปประมวลผล จัดกลุ่มและตีความ เพื่อสร้างเป็นข้อมูล
- ข้อมูล (Information) จะถูกนำไปวิเคราะห์และประมวลผล เพื่อสร้างเป็นความรู้
- ความรู้ (Knowledge) จะถูกนำไปไตร่ตรอง และประมวลผลเพื่อสร้างเป็นภูมิปัญญา
- ภูมิปัญญา (Wisdom) จะถูกนำไปใช้ในการตัดสินใจ



● Data Science

Data Science หรือ “วิทยาการข้อมูล” คือ ศาสตร์การนำข้อมูลมาใช้ประโยชน์ ผ่านวิธีการต่างๆ ที่ครอบคลุมตั้งแต่การจัดเก็บ จัดการ และวิเคราะห์ ไปจนถึงการนำเสนอข้อมูลที่วิเคราะห์ได้ออกมาในรูปแบบของข้อมูลเชิงลึก (Insight) เพื่อนำมาใช้ประโยชน์ทั้งทางด้านธุรกิจและด้านอื่นๆ ในแง่ของการตอบคำถาม แก้ปัญหา ทำนายผล รวมไปถึงคาดการณ์ความเสี่ยงที่อาจเกิดขึ้นในอนาคต

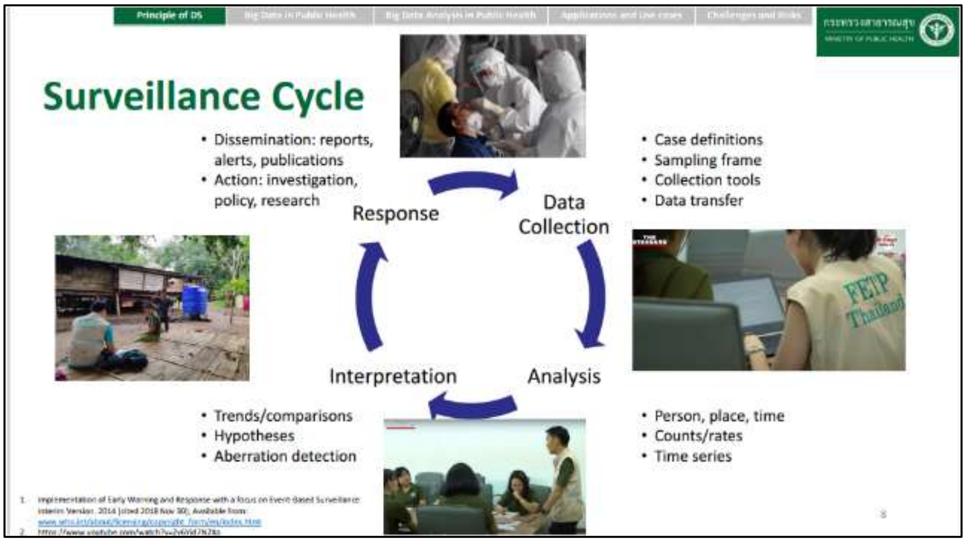


Data Science Cycle คือ ลำดับของกระบวนการจัดการข้อมูล ตั้งแต่ตอนเก็บรวบรวม วิเคราะห์ ประมวลผล และนำไปใช้ จนถึงตอนเริ่มเก็บข้อมูลใหม่อีกครั้ง ประกอบด้วย 6 ขั้นตอน

1. การทำความเข้าใจเกี่ยวกับกระบวนการ (Business Understanding)
2. การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)
3. การเตรียมข้อมูล (Data Preparation)
4. การสร้างแบบจำลอง (Modeling)
5. การประเมินผล (Evaluation)
6. การนำไปใช้งาน (Deployment)

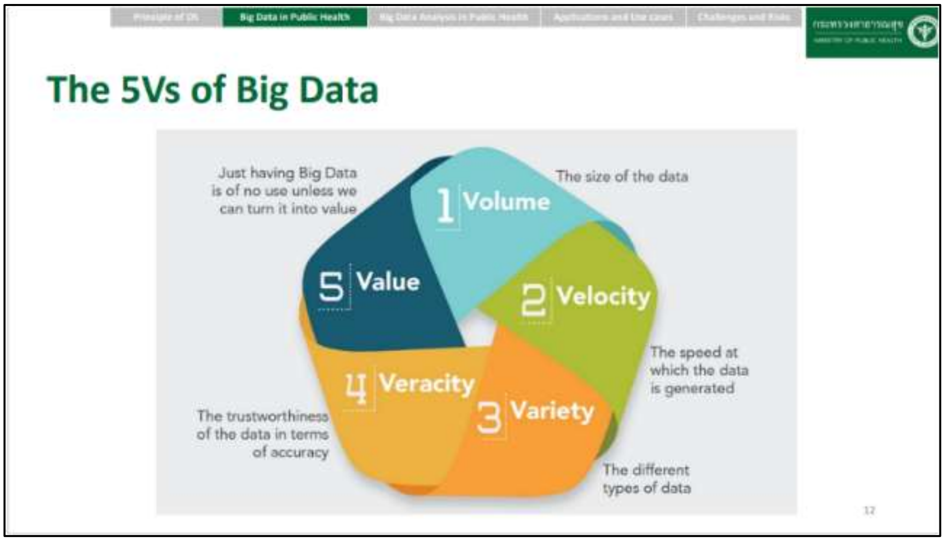
Surveillance Cycle คือ กระบวนการต่อเนื่องของการรวบรวม วิเคราะห์ และใช้ข้อมูล เพื่อติดตามสถานการณ์หรือประชากรกลุ่มเป้าหมายที่เฉพาะเจาะจง เป็นกระบวนการแบบวนซ้ำ ที่มีขั้นตอนสำคัญหลายขั้นตอน ซึ่งช่วยให้มั่นใจได้ถึงการปรับปรุงอย่างต่อเนื่องและการติดตามที่มีประสิทธิภาพ

1. Data Collection คือ กระบวนการที่เราทำการรวบรวมข้อมูลจากแหล่งต่างๆ เพื่อใช้ในการวิเคราะห์ การตัดสินใจ หรือการทำงาน
2. Analysis คือ การนำข้อมูลมาวิเคราะห์เพื่อหาความหมาย รูปแบบ ความสัมพันธ์ หรือแนวโน้ม
3. Interpretation คือ การอธิบายความหมาย สรุปผล หรือหาข้อสรุปจากข้อมูลที่วิเคราะห์แล้ว
4. Response คือ การสื่อสารผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล



- **Big data**

Big data เป็นข้อมูลขนาดใหญ่ มีความซับซ้อนที่จะจัดการหรือวิเคราะห์ด้วยคอมพิวเตอร์ แต่ปัจจุบันมีการพัฒนาและสามารถจัดการได้ดีขึ้น



5 Vs ของ Big Data คือ คุณลักษณะสำคัญ 5 ประการที่อธิบายถึงความท้าทายในการจัดการข้อมูลจำนวนมหาศาล

1) ปริมาณข้อมูล (Volume) ปริมาณข้อมูลที่สร้างขึ้นในปัจจุบันมีจำนวนมหาศาลและมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง ข้อมูลจำนวนมหาศาลนี้เรียกว่า Big Data
2) ความเร็วในการสร้างข้อมูล (Velocity) ข้อมูลจำนวนมหาศาลถูกสร้างขึ้นอย่างรวดเร็ว ข้อมูลเหล่านี้มาจากแหล่งต่างๆ เช่น เครือข่ายสังคมออนไลน์ เว็บไซต์
3) ความหลากหลายของข้อมูล (Variety) ข้อมูลจำนวนมหาศาลมีรูปแบบที่หลากหลาย เช่น ข้อความ รูปภาพ วิดีโอ เสียง และข้อมูลเชิงพื้นที่
4) ความถูกต้อง (Veracity) มักมีข้อผิดพลาดและข้อมูลที่ไม่สมบูรณ์ ซึ่งอาจส่งผลต่อผลลัพธ์การวิเคราะห์ การทำความสะอาดและตรวจสอบคุณภาพข้อมูล จำเป็นต้องมีกระบวนการที่รัดกุมเพื่อให้มั่นใจว่าข้อมูลมีความถูกต้องและเชื่อถือได้
5) คุณค่าของข้อมูล (Value) ข้อมูลจำนวนมหาศาลมีคุณค่ามหาศาลสำหรับองค์กรต่างๆ ข้อมูลเหล่านี้สามารถนำมาใช้เพื่อปรับปรุงประสิทธิภาพการทำงาน ตัดสินใจได้อย่างชาญฉลาด และสร้างนวัตกรรมใหม่ๆ

ประโยชน์ของ Big Data Big Data มีประโยชน์มากมายสำหรับองค์กรต่างๆ ประโยชน์เหล่านี้รวมถึง

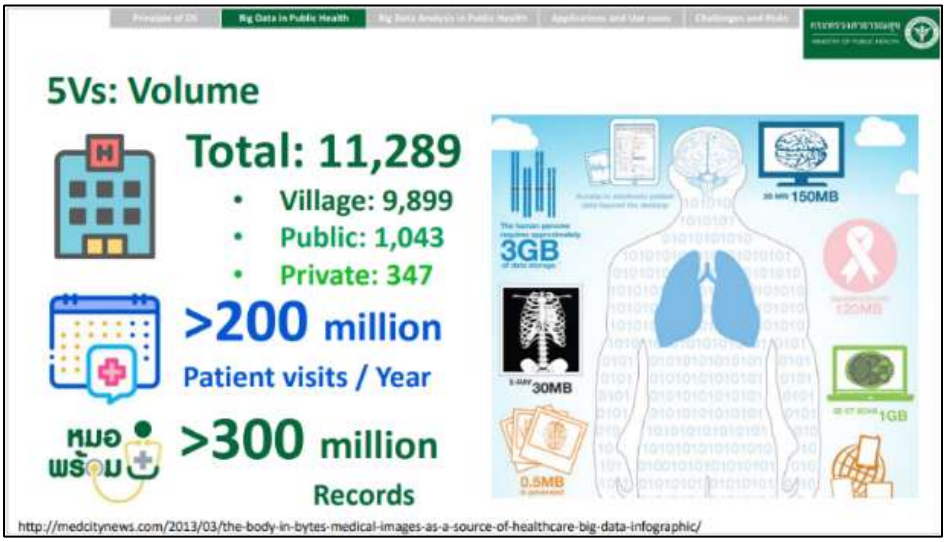
- การปรับปรุงประสิทธิภาพการทำงาน: Big Data สามารถนำมาใช้เพื่อระบุโอกาสในการปรับปรุงประสิทธิภาพการทำงาน ตัวอย่างเช่น องค์กรสามารถใช้ Big Data เพื่อวิเคราะห์รูปแบบการขายและระบุผลิตภัณฑ์หรือบริการที่ขายดีที่สุด
- การตัดสินใจได้อย่างชาญฉลาด: Big Data สามารถนำมาใช้เพื่อให้ข้อมูลเชิงลึกแก่ผู้ตัดสินใจ ตัวอย่างเช่น องค์กรสามารถใช้ Big Data เพื่อวิเคราะห์ความคิดเห็นของลูกค้าและระบุพื้นที่ที่ต้องปรับปรุง

- การสร้างนวัตกรรมใหม่ๆ: Big Data สามารถนำมาใช้เพื่อสร้างนวัตกรรมใหม่ๆ ตัวอย่างเช่น องค์กรสามารถใช้ Big Data เพื่อพัฒนาผลิตภัณฑ์และบริการใหม่ๆ ที่ตรงกับความต้องการของลูกค้า

Do you consider Public Health Data as Big Data?

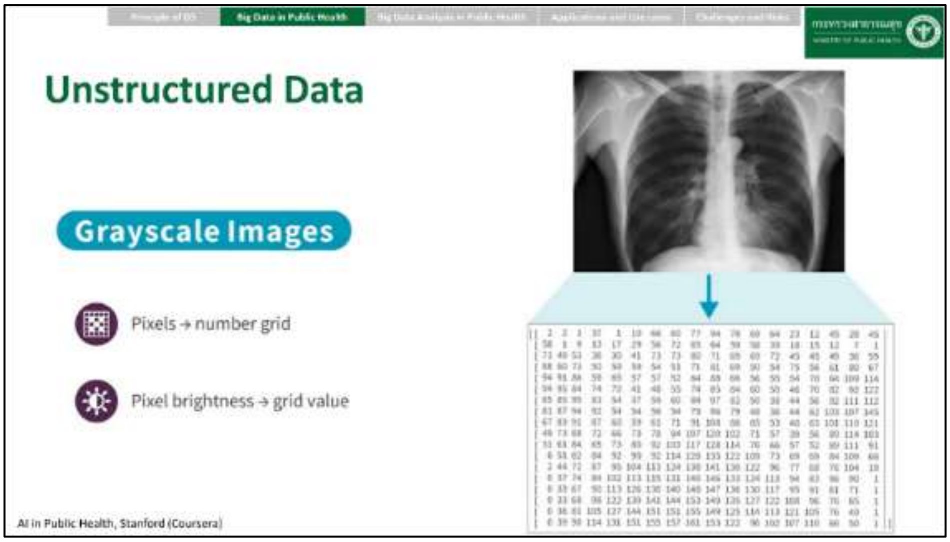
ตัวอย่างข้อมูล Bid data ในภาคสาธารณสุข ได้แก่

- "หมอพร้อม" เป็นข้อมูลที่มีขนาดใหญ่
- ข้อมูล 43 แฟ้ม คือข้อมูลที่เก็บไว้เพื่อการวินิจฉัยรักษา และเพื่อการเบิกจ่าย
- Consumer Data ข้อมูลที่ผู้ป่วยเก็บเอง เช่น smart watch เครื่องวัดน้ำตาล ผู้ป่วยกรอกข้อมูลตัวเองโดยตรง
- ข้อมูลจาก Social Media เช่น twitter Facebook สามารถตรวจจับประชาชนว่าท้องเสียได้
- ข้อมูลน้ำเสีย เช่น ผลตรวจน้ำเสียจากเครื่องบิน โรงแรม เพื่อเฝ้าระวังโรคระบาดในต่างประเทศ



ลักษณะของข้อมูล มี 2 ประเภทคือ

1) **Structured data** ข้อมูลแบบมีโครงสร้าง เช่น เป็นตาราง
2) **Unstructured data** เช่น ข้อมูลประวัติผู้ป่วย ข้อมูลแบบ free text เป็นข้อมูลที่ท้าทายวิเคราะห์ค่อนข้างยาก

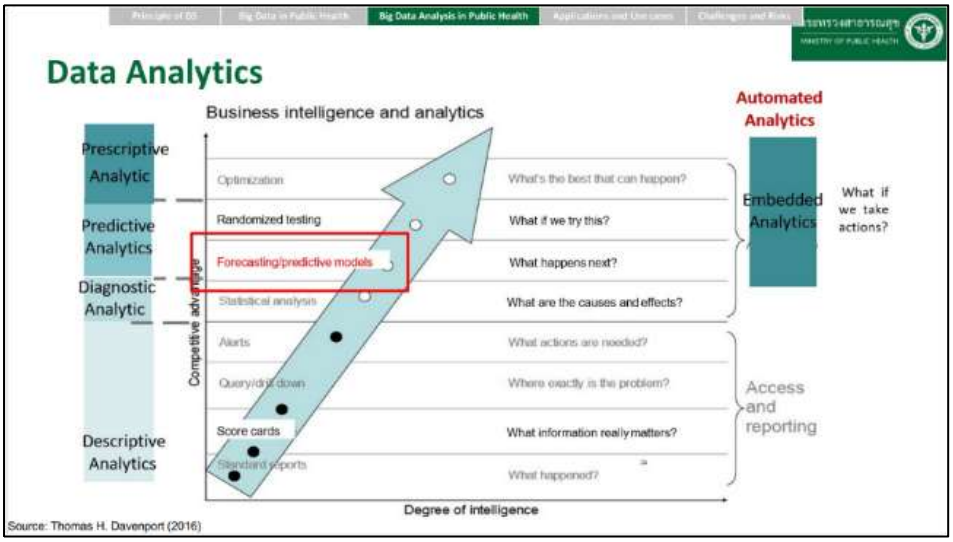


Data Science to drive Public Health Decision Making

ปลายศตวรรษที่ 18 เอ็ดเวิร์ด เจนเนอร์ (Edward Jenner) นายแพทย์ชาวอังกฤษสังเกตว่า หญิงรีดนมวัวที่ติดเชื้อไวรัสฝีดาษวัว (Cowpox) ไม่เคยป่วยเป็นไข้ทรพิษเลย เขาจึงทดลองนำหนองจากแผลของผู้ป่วยฝีดาษวัวใส่เข้าไปใต้ผิวหนังของเด็กชาย แล้วจึงนำเชื้อไข้ทรพิษเข้าสู่ร่างกาย ปรากฏว่า เด็กไม่มีอาการล้มป่วย จึงเป็นข้อพิสูจน์ได้ว่า เชื้อฝีดาษวัวช่วยสร้างภูมิคุ้มกันได้

หมอ John snow คือ อายุรแพทย์ชาวอังกฤษและผู้นำในการพัฒนาด้านวิสัญญีวิทยาและสุขศาสตร์ทางการแพทย์ เขาได้รับการยกย่องให้เป็นหนึ่งในผู้บุกเบิกทางวิทยาการระบาดสมัยใหม่ซึ่งเป็นผลจากงานของเขาส่วนหนึ่งเกี่ยวกับการระบุแหล่งกำเนิดของการแพร่ระบาดของอหิวาตกโรคในเขตโซโฮ ลอนดอนใน ค.ศ. 1854 นักวิจัยจากมหาวิทยาลัยออกซฟอร์ดระบุถึงการค้นพบของสโนว์ว่า เป็นแรงบันดาลใจให้กับการยอมรับด้านวิสัญญีวิทยาตลอดจนการเปลี่ยนแปลงรากฐานของระบบน้ำ กับการบำบัดน้ำเสียในลอนดอนและการเปลี่ยนแปลงที่คล้ายกันในเมืองอื่น ๆ รวมถึงการปรับปรุงอย่างมีนัยสำคัญทางสาธารณสุขพื้นฐานทั่วโลก

- Data Analytics



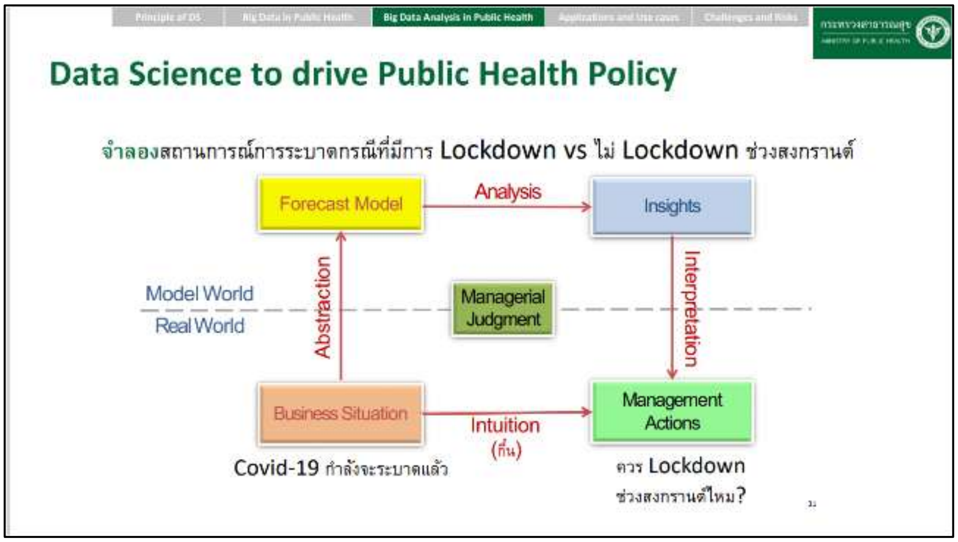
การวิเคราะห์ข้อมูลขึ้นกับคำถามที่กำหนดไว้ โดยการวิเคราะห์ข้อมูลประกอบด้วย

1. **Descriptive Analytic** (การวิเคราะห์เชิงพรรณนา) คือ การวิเคราะห์ขั้นพื้นฐาน ช่วยให้เห็นภาพรวมและความสัมพันธ์ของข้อมูล ชี้ให้เห็นว่าเกิดเหตุการณ์อะไรขึ้นบ้างในอดีต ทำให้หหน่วยงานสามารถวางแผนและตัดสินใจการดำเนินงานที่จะเกิดขึ้นในอนาคตได้
2. **Diagnostic Analytic** (การวิเคราะห์เชิงวินิจฉัย) คือ การวิเคราะห์และหาความสัมพันธ์ของข้อมูลเพื่อหาสาเหตุของสิ่งที่เกิดขึ้น ช่วยให้หน่วยงานทราบสาเหตุของผลลัพธ์ และวางแผนหลีกเลี่ยงผลลัพธ์ในทางลบ เช่น cohort case control
3. **Predictive Analytic** (การวิเคราะห์เชิงทำนาย) คือ ผลผนวกจากการวิเคราะห์เชิงพรรณนาและการวิเคราะห์เชิงวินิจฉัย โดยใช้การวิเคราะห์เชิงพรรณนาเพื่อหาภาพรวมและความสัมพันธ์ของข้อมูลในอดีต แล้วหารูปแบบของข้อมูล เพื่อคาดเดาผลลัพธ์ที่จะเกิดขึ้นในอนาคต ส่วนการวิเคราะห์เชิงวินิจฉัยถูกนำมาใช้สำหรับอธิบายสาเหตุของผลลัพธ์ที่คาดการณ์ว่าจะเกิดขึ้น เช่น อยากรู้อนาคตผู้ป่วยกี่ราย
4. **Prescriptive Analytic** (การวิเคราะห์เชิงให้คำแนะนำ) คือ การวิเคราะห์แนวโน้มของสิ่งที่จะเกิดขึ้น เป็นการวิเคราะห์ข้อมูลระดับสูง เพื่อคาดการณ์ผลลัพธ์ที่อาจเกิดขึ้นของแนวโน้มทั้งหมด พร้อมให้คำแนะนำว่าแนวโน้มไหนจะดีที่สุด หรือแนะนำแนวทางในการรับมือและแก้ไขปัญหา ส่งผลให้การวิเคราะห์ประเภทนี้เป็นการวิเคราะห์ที่ซับซ้อนที่สุด มักนำ Machine Learning and Simulation มาใช้ เพื่อวิเคราะห์แทนมนุษย์

ตัวอย่างวิธีการใช้ Data Science ในการขับเคลื่อนนโยบายสาธารณสุข

การใช้ Data Science ในการขับเคลื่อนนโยบายสาธารณสุข มีประโยชน์หลายประการ เช่น

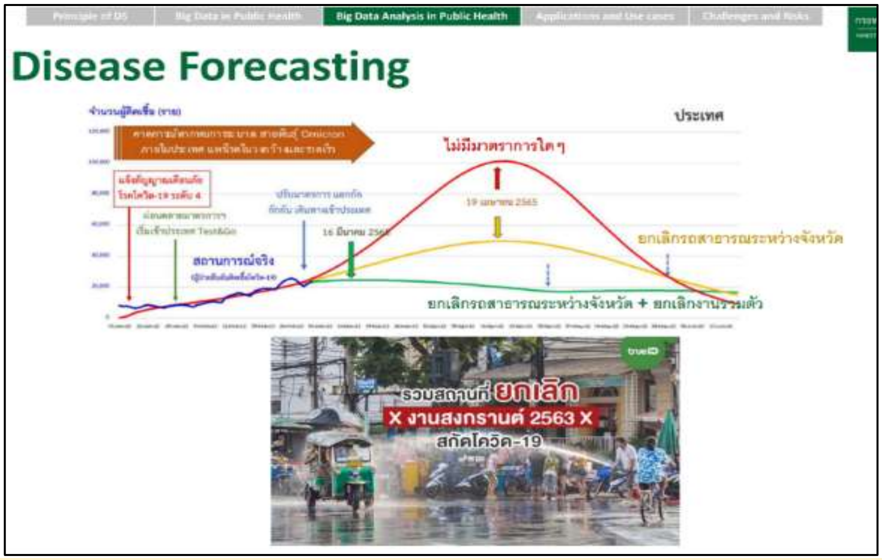
- ช่วยให้ตัดสินใจได้อย่างมีประสิทธิภาพ จากข้อมูลที่มี
- ช่วยในการคาดการณ์ผลลัพธ์ของนโยบายต่าง ๆ ได้
- ช่วยในการจัดสรรทรัพยากรได้อย่างมีประสิทธิภาพ
- ช่วยในการติดตามผลและประเมินผลนโยบาย



ตัวอย่างที่แสดงในภาพ คือ การจำลองสถานการณ์การระบาดของ COVID-19 กรณีมีการ Lockdown ช่วงสงกรานต์ vs ไม่ Lockdown มีวิธีการดังนี้

1) **การวิเคราะห์ข้อมูล**: รวบรวมข้อมูลเกี่ยวกับ COVID-19 เช่น จำนวนผู้ติดเชื้อ อัตราการเสียชีวิต อัตราการฉีดวัคซีน ฯลฯ
2) **การสร้างโมเดล**: พัฒนาโมเดลทางคณิตศาสตร์เพื่อจำลองสถานการณ์การระบาด
3) **การวิเคราะห์ผลลัพธ์**: วิเคราะห์ผลลัพธ์จากโมเดลเพื่อคาดการณ์ว่าอะไรจะเกิดขึ้นหากมีการ Lockdown ช่วงสงกรานต์ vs ไม่ Lockdown
4) **การตัดสินใจ**: ใช้ข้อมูลและ insights จากโมเดลเพื่อตัดสินใจว่าควรมีการ Lockdown ช่วงสงกรานต์หรือไม่

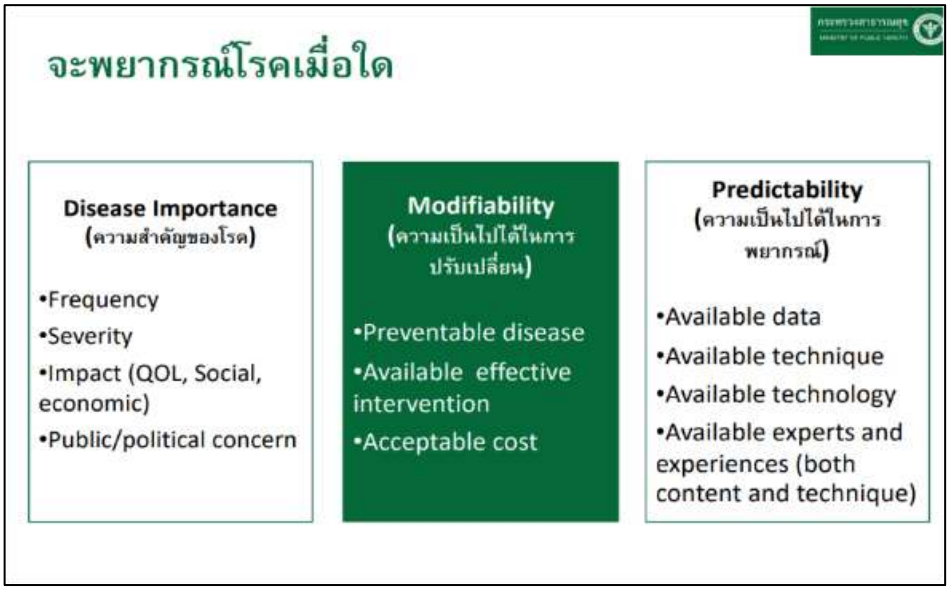
การวิเคราะห์ข้อมูล Big Data เกี่ยวกับ COVID-19 ในประเทศไทย ข้อมูลถูกนำมาใช้เพื่อจำลองสถานการณ์การระบาดของ COVID-19 ภายใต้ scenarios ต่างๆ จากภาพผลลัพธ์ที่ได้จากโมเดลแสดงให้เห็นว่า scenarios การ Lockdown ช่วงสงกรานต์จะช่วยลดจำนวนผู้ติดเชื้อและผู้เสียชีวิตจาก COVID-19 ได้อย่างมีนัยสำคัญ



- **การใช้ข้อมูล Forecasting เพื่อตัดสินใจ**

Forecasting คือ การใช้ข้อมูลในอดีตและปัจจุบันมาทำนายอนาคต โดยกฎของการพยากรณ์ มี 3 ข้อคือ

1) **LAW 1 การพยากรณ์ส่วนมากจะผิด** เพราะฉะนั้นเวลาสื่อสาร เราจะไม่บอกจำนวน แต่เราจะบอกแนวโน้ม บอกค่าความคาดเคลื่อน บอกค่า 95% confident interval
2) **LAW 2 การพยากรณ์ที่แยกย่อยจะแม่นยำน้อยกว่าแบบหยาบ** เช่น การพยากรณ์ล่วงหน้าที่แยกย่อยมากๆ เช่น พยากรณ์เป็นรายวันจะมีโอกาสผิดแต่ละวันมาก แต่การพยากรณ์เป็นรายสัปดาห์ รายเดือน หรือรายปีมีโอกาสผิดน้อยกว่า เช่น ค่าความคลาดเคลื่อนของการพยากรณ์อากาศว่าสัปดาห์นี้ฝนจะตก ตกวันไหนก็ถือว่าฝนตก แต่ถ้าบอกว่าพรุ่งนี้ฝนจะตกแสดงว่าพรุ่งนี้ต้องตกจริง
3) **LAW 3 ยิ่งพยากรณ์ไปไกลยิ่งแม่นยำน้อยลง** เช่น การพยากรณ์ว่าพรุ่งนี้ฝนจะตก กับวันจันทร์หน้าจะมีฝนตก พบว่าการพยากรณ์สำหรับวันพรุ่งนี้จะแม่นยำกว่า



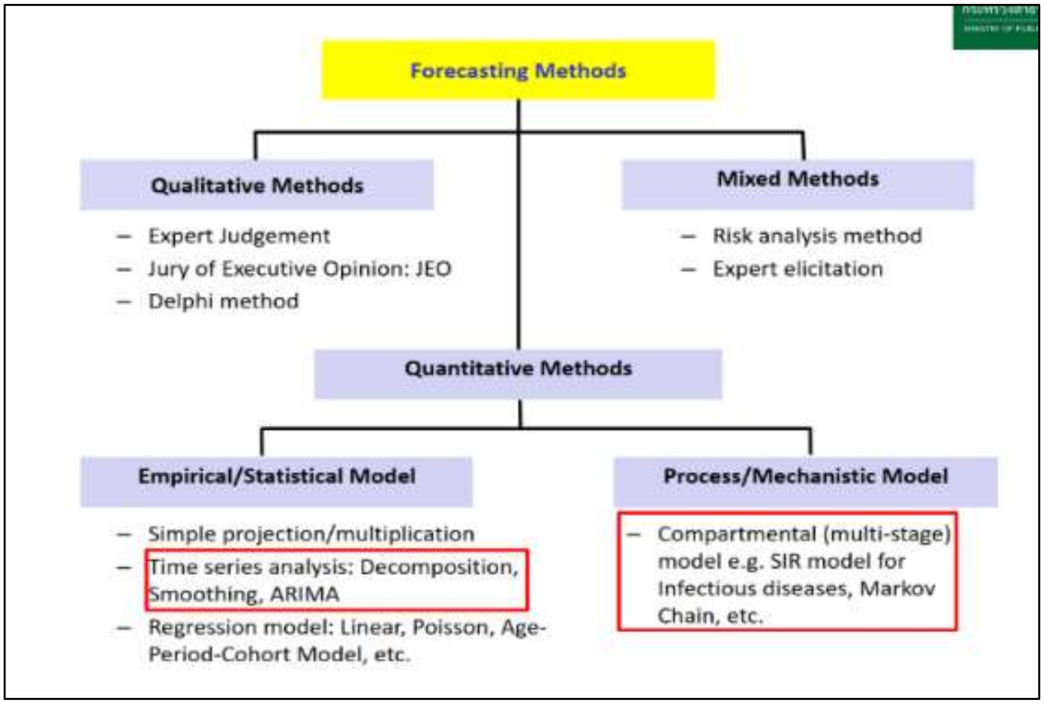
องค์ประกอบที่ดีของการพยากรณ์

1) Compatible (มีความสอดคล้อง)
2) Meaningful (มีความหมาย นำไปใช้ประโยชน์ได้)
3) Useful time horizon (มีช่วงเวลาที่จะพยากรณ์)
4) Reliable (เชื่อถือได้) ถ้ามีข้อมูลดี พัฒนา Model ดี ประเมินแล้วผ่านถือว่าเชื่อถือได้
5) Accurate (มีความแม่นยำ)
6) Easy to understand and use (ง่ายต่อการเข้าใจและใช้งาน) สามารถสื่อสารกับประชาชนได้เข้าใจง่าย หรือผู้บริหารใช้ข้อมูลการพยากรณ์ได้ง่าย โดยเลือกมาทำ visualize ที่สื่อสาร ให้ดูง่าย

ประโยชน์ของการพยากรณ์โรค

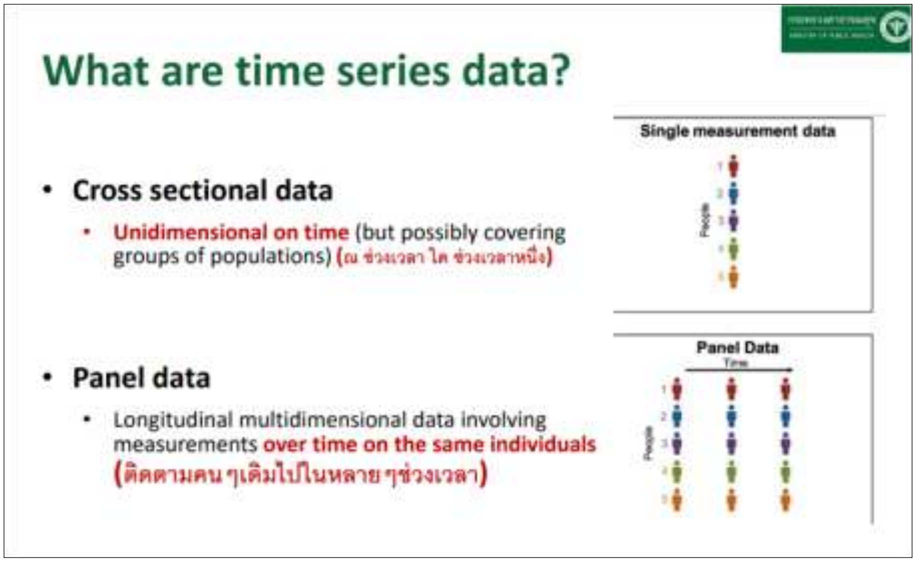
- ช่วยในการเตรียมความพร้อม และวางแผนทางทรัพยากร
- ช่วยในการตัดสินใจเชิงนโยบาย
- ช่วยในการประเมินประสิทธิภาพของนโยบาย
- แจ้งเตือนการระบาด
- ควบคุมการระบาด
- สร้างความตระหนัก

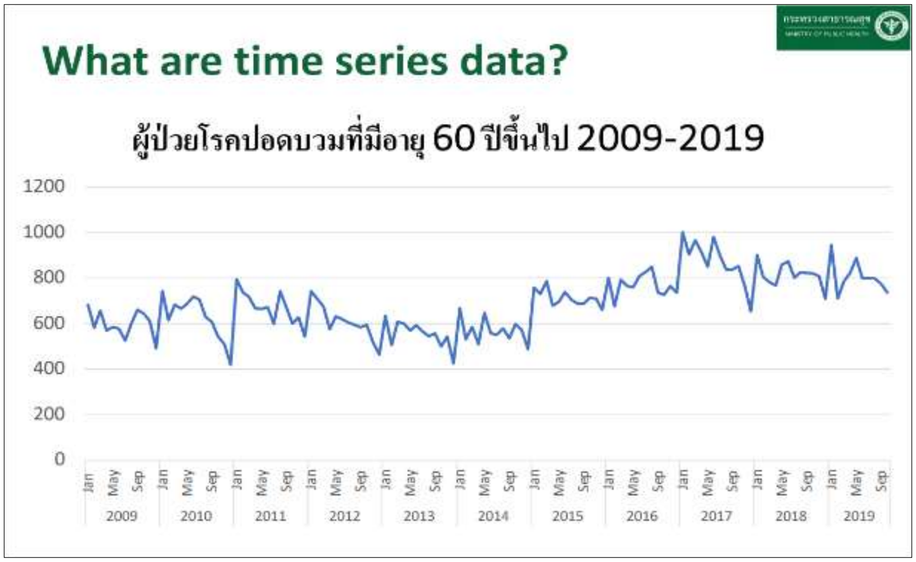
วิธีการพยากรณ์โรค (Forecasting Methods)



การพยากรณ์แบบอนุกรมเวลา (Time series Model) เหมาะสำหรับประเภทข้อมูลต่อไปปนี้

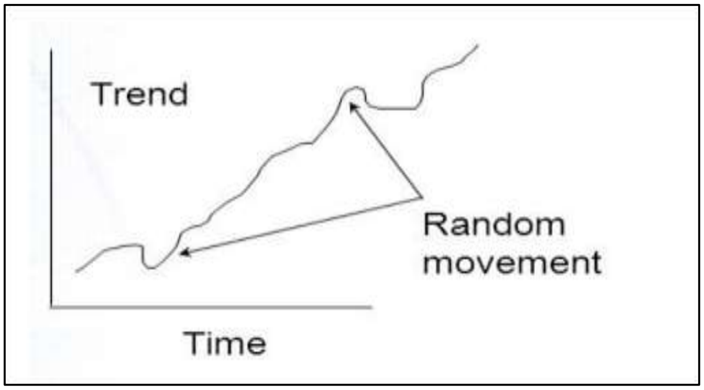
1) Cross sectional data ข้อมูล ณ ช่วงเวลาใด เวลาหนึ่ง
2) Panel data ข้อมูลจากการติดตามคน ๆ เดิมไปหลาย ๆ ช่วงเวลา
3) Time series data ข้อมูลจากการติดตามกลุ่มประชากรกลุ่มหนึ่งอย่างต่อเนื่องในหลายช่วงเวลา ไม่จำเป็นต้องเป็นคนเดิม โดยข้อมูลจะเก็บต่อเนื่องในหน่วยเวลาที่เท่า ๆ กัน เช่น ทุกชั่วโมง ทุกวัน ทุกสัปดาห์



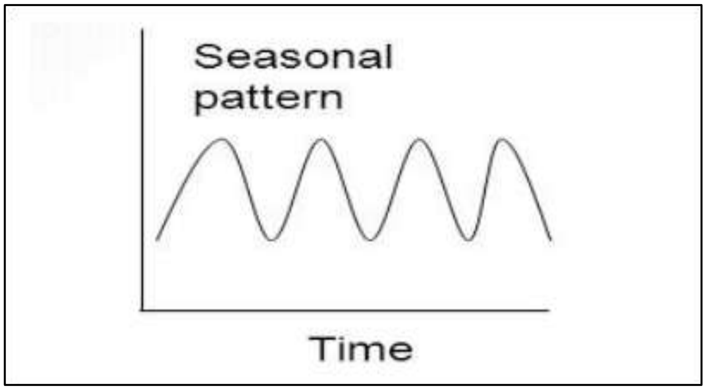


องค์ประกอบของอนุกรมเวลา (Time series)

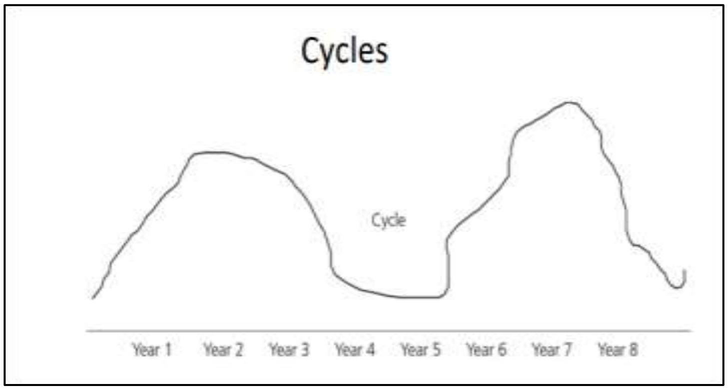
1) Trend มีแนวโน้ม มีการบอกจุดนี้ถึงจุดนั้น เช่น มีแนวโน้มเพิ่มขึ้น



2) Seasonality มีฤดูกาล โดยปกติฤดูกาลยาวได้ไม่เกิน 1 ปี ถ้าเกิน 1 ปี จะถือว่าเป็นรอบ (Cycle) หรืออาจเป็นช่วงสั้น ๆ ช่วงเวลาของวัน ได้ เช่น จำนวนรถที่วิ่งบนถนนในแต่ละช่วงเวลาใน 1 วัน หรือจำนวนผู้ป่วยรายวันก็ได้



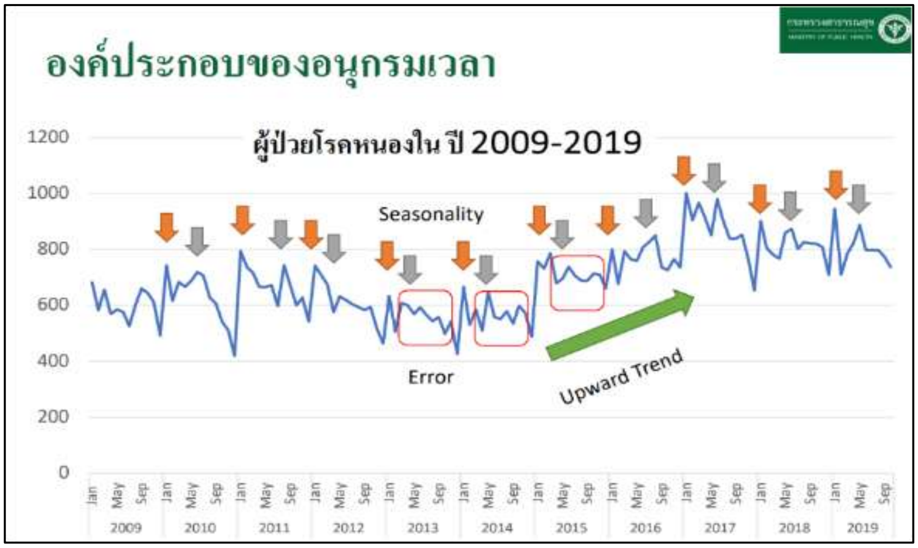
3) Cycles (รอบ) คล้ายกับฤดูกาลแต่จะเป็นช่วงเวลา 1 ปี หรือ 2 ปี เป็นต้น



4) Error / Irregularity / Residual (มีค่าความคลาดเคลื่อน) อาจเกิดขึ้นจากเหตุการณ์ที่ไม่คาดคิด เช่น ภัยพิบัติ ข่าวลือ

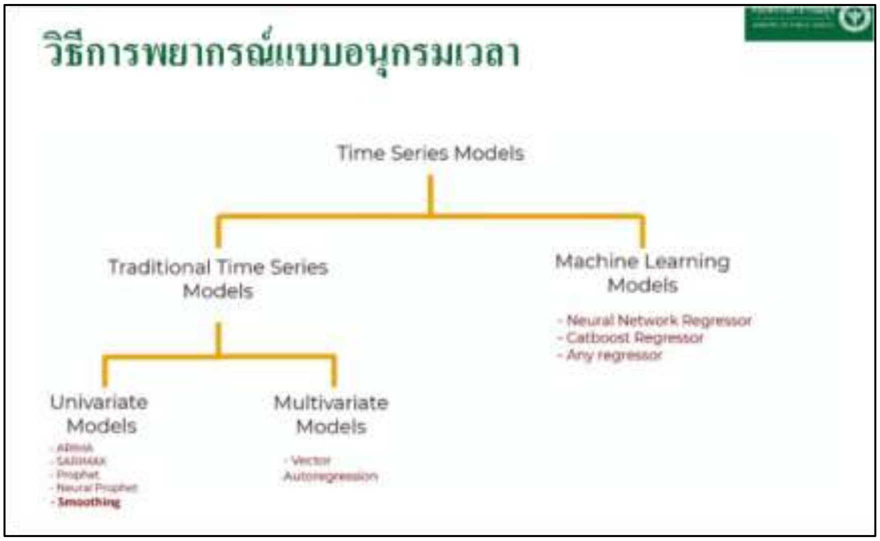
ตัวอย่างองค์ประกอบของอนุกรมเวลา

เช่น ผู้ป่วยโรคหนองในปี 2009 - 2019 ที่แสดงเป็น Trend มีแนวโน้มเพิ่มขึ้น และมี Seasonality



วิธีการพยากรณ์แบบ Time Series Models

โมเดลอนุกรมเวลา เป็นเครื่องมือทางคณิตศาสตร์และสถิติ สำหรับวิเคราะห์และพยากรณ์ข้อมูลที่เรียงลำดับตามเวลา ช่วยให้เข้าใจรูปแบบและแนวโน้มข้อมูล ซึ่งสามารถนำไปใช้เพื่อคาดการณ์ในอนาคตได้ โดยปัจจุบันวิธีการพยากรณ์แบบ Time Series Models มี 2 วิธีได้แก่ Traditional Time series Models และ Machine Learning Models



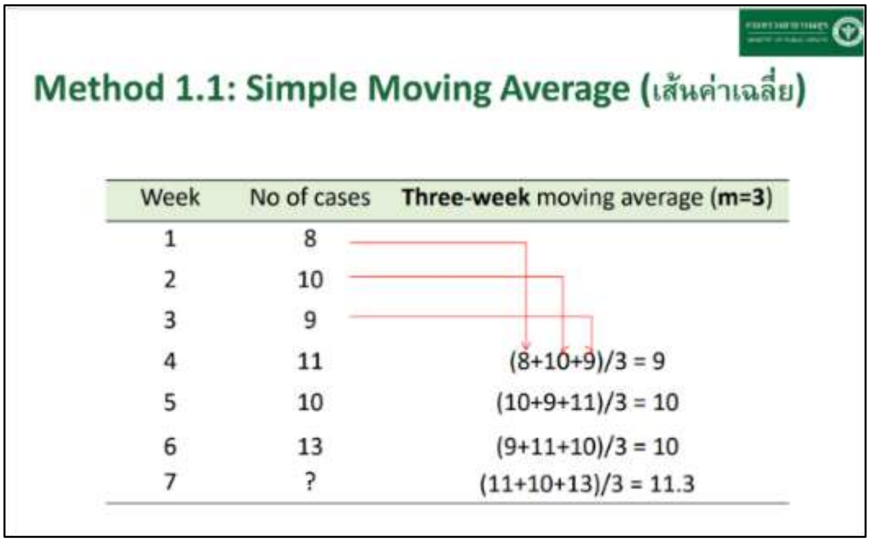
1) Traditional Time series Models มี 2 แบบคือ

- Univariate Models คือ การพยากรณ์ด้วยตัวแปรเดียว เช่น การพยากรณ์ไข้หวัดใหญ่ ปี 2023 จะใช้ข้อมูลโรคจากอดีตและปัจจุบันมาพยากรณ์
- Multivariate Models คือ การพยากรณ์โดยใช้หลายตัวแปร เช่น การพยการณ์จำนวนผู้ป่วยด้วยโรคที่เกี่ยวข้องกับ PM2.5 จะใช้ข้อมูลค่าฝุ่น PM2.5 ในการพยากรณ์ด้วย

2) Machine Learning Models เหมาะสำหรับข้อมูลที่มีรูปแบบและแนวโน้มที่ซับซ้อน

ตัวอย่าง Traditional Time series Models แบบ Univariate Models ได้แก่

1) Simple Moving Average (เส้นค่าเฉลี่ย) การหาค่าเฉลี่ยจากข้อมูลย้อนหลัง เพื่อจัดการความคาดเคลื่อน ตัวอย่างเช่น การพยากรณ์จำนวนผู้ป่วยในสัปดาห์ที่ 7 ใช้วิธีคือ นำข้อมูลเฉลี่ยของ 3 สัปดาห์ย้อนหลัง (สัปดาห์ที่ 4 5 และ 6) มาคำนวณค่าเฉลี่ยของสัปดาห์ที่ 7



Week	No of cases	Three-week moving average (m=3)
1	8	
2	10	
3	9	
4	11	(8+10+9)/3 = 9
5	10	(10+9+11)/3 = 10
6	13	(9+11+10)/3 = 10
7	?	(11+10+13)/3 = 11.3

2) Single Exponential Smoothing & Double & Triple เทคนิคนี้สามารถใช้ excel

3) Autoregressive Integrated Moving Average (ARIMA) ใน SPSS โดย Autoregressive (AR) คือการพยากรณ์ที่สนใจในค่าของตัวมันเองในอดีต และ Moving Average (MA) คือการพยากรณ์โดยใช้ค่าความคลาดเคลื่อนในอดีต

สำหรับการเลือกวิธีการพยากรณ์ที่ดีที่สุดนั้น ขึ้นอยู่กับสถานการณ์เฉพาะของแต่ละบุคคล ควรทดสอบวิธีการพยากรณ์หลายๆ วิธีเพื่อดูว่าวิธีไหนแม่นยำที่สุด การพยากรณ์เป็นเพียงการประมาณการเท่านั้น ไม่สามารถรับประกันความถูกต้องได้ 100%

บรรยาย เรื่องการใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพ

โดย นายฤทธิไกร นามเกษ นักวิชาการสาธารณสุขปฏิบัติการ
กองระบาดวิทยา กรมควบคุมโรค

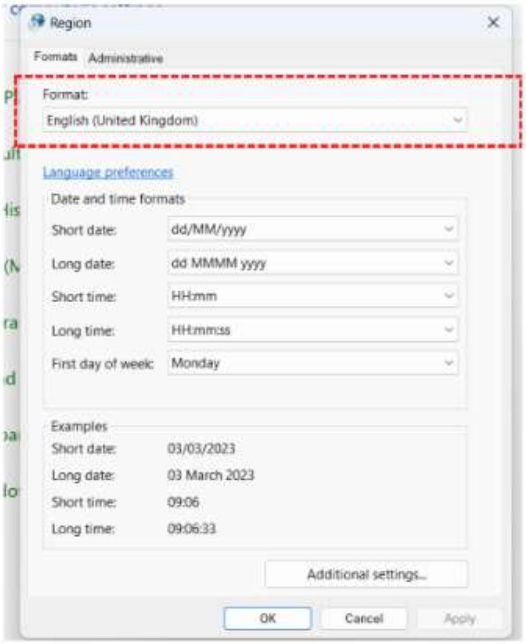
ในการวิเคราะห์ข้อมูลและการพยากรณ์สุขภาพ สิ่งที่ต้องคำนึงถึงอันดับแรกคือจัดการข้อมูลและเตรียมข้อมูลให้พร้อมใช้ ซึ่งการอบรมในหัวข้อนี้วัตถุประสงค์เพื่อให้ผู้เรียนมีความรู้ ความเข้าใจเกี่ยวกับเครื่องมือการพยากรณ์สุขภาพ และการใช้ Excel เพื่อจัดการข้อมูล โดยผู้สอนมีไฟล์ตัวอย่างประกอบการฝึกปฏิบัติ สรุปสาระสำคัญดังนี้

1. Data Cleaning and Preparation

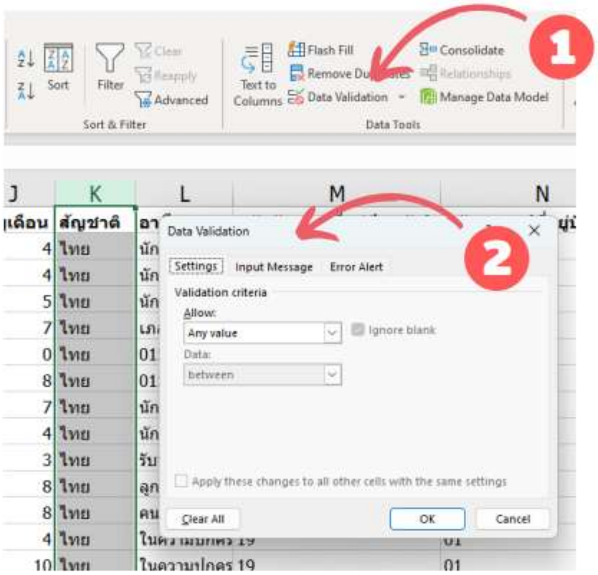
คือ การทำความสะอาดข้อมูล และจัดเตรียมข้อมูล ซึ่งเป็นเป็นขั้นตอนที่สำคัญมากในการวิเคราะห์ข้อมูลเพื่อให้ได้ข้อมูลที่พร้อมใช้และมีคุณภาพ โดยมีตัวอย่างการทำด้วยสูตรหรือฟังก์ชันใน Excel ดังนี้

1.1 Set Format คือฟังก์ชันที่ใช้ปรับรูปแบบของข้อมูลที่เป็นตัวแปร “วันที่” ให้อยู่ในรูปแบบที่พร้อมสำหรับการวิเคราะห์ข้อมูล ดังนี้

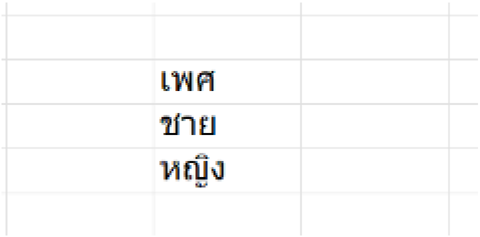
1) พิมพ์ค้นหา "Control Panel"
2) เลือก Region
3) ตั้งค่าในแถบ Format เป็น English (United Kingdom)



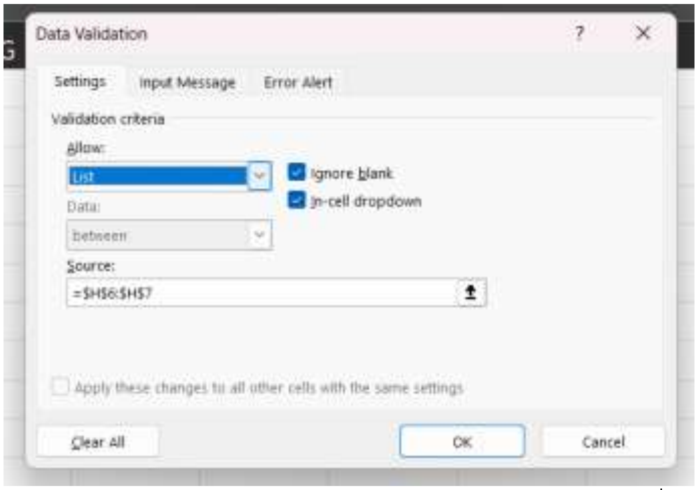
2. **Data validation and verification** คือฟังก์ชันที่ใช้ตรวจสอบข้อมูล โดยใช้ไฟล์ประกอบการฝึกปฏิบัติ ชื่อ *“severe_ex”* และมีวิธีการตรวจสอบ ข้อมูลดังนี้



1) เลือกเซลล์หรือช่องที่ต้องการตรวจสอบข้อมูล
2) ไปที่แท็บ "Data" ในแถบเมนู
3) คลิกที่ "Data Validation" ใน "Data Tools"
4) ในหน้าต่าง "Data Validation" สามารถกำหนดเงื่อนไขต่าง ๆ ได้ โดย
   - แถบ "Settings": สามารถกำหนดประเภทของเงื่อนไขที่ต้องการ เช่น ตรวจสอบข้อมูลจำนวนเต็ม หรือข้อความที่ไม่เป็นค่าว่าง
   - แถบ "Input Message": กำหนดข้อความที่จะแสดง เมื่อผู้ใช้คลิกเพื่อป้อนข้อมูลในเซลล์นี้
   - แถบ "Error Alert": กำหนดข้อความแจ้งเตือนที่จะปรากฏ เมื่อข้อมูลที่ป้อนไม่ตรงกับเงื่อนไขที่กำหนด
5) การทำตัวเลือก dropdown list ดังตัวอย่าง ตัวแปรเพศ ได้แก่ ชาย, หญิง
   - สร้างตัวแปร เพศ



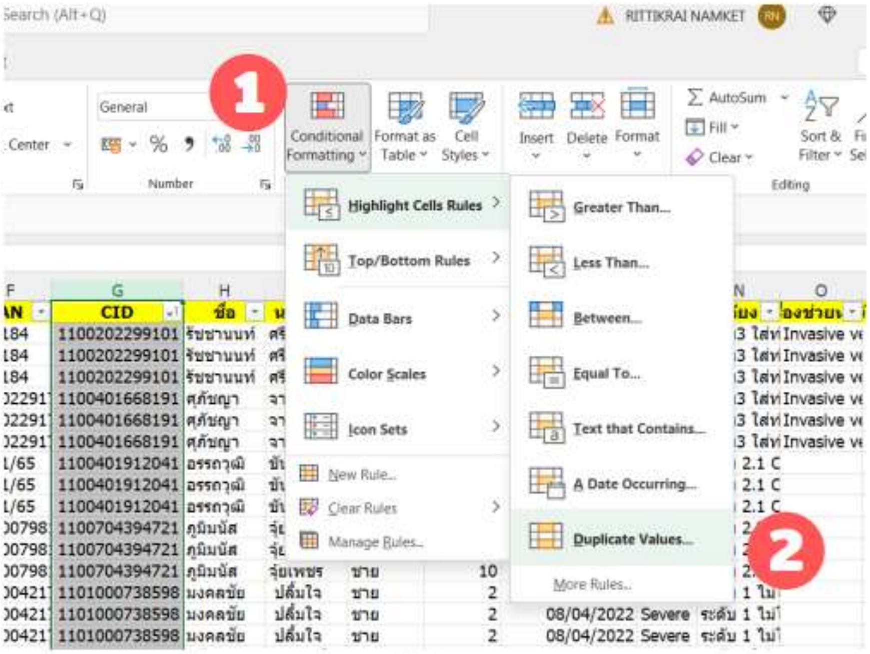
- คลิก Data Validation เลือก List และใส่แหล่งข้อมูลที่สร้างไว้



- ผลลัพธ์ จะเห็นว่ามี List รายการ category ของตัวแปร “เพศ”ที่ต้องการ ได้แก่ ชาย หญิง วิธีนี้จะช่วยลดการผิดพลาดของการกรอกข้อมูลได้

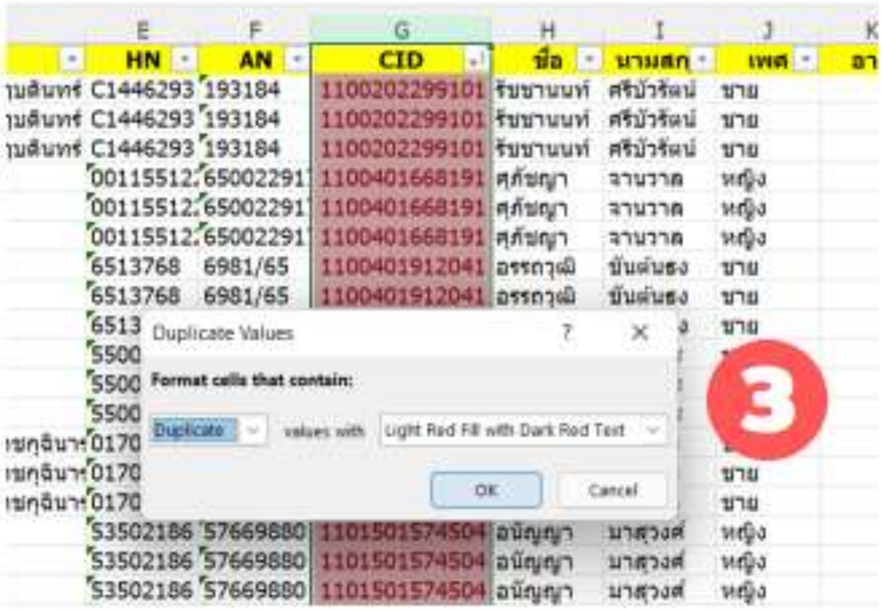
2.1 Remove duplicates คือฟังก์ชันที่ใช้ตรวจสอบข้อมูลที่ซ้ำกัน

1) คลิกที่เมนู Conditional Formatting (ตามหมายเลข 1) และเลือก -> Highlight Cells Rules -> Duplicate Values (ตามหมายเลข 2)

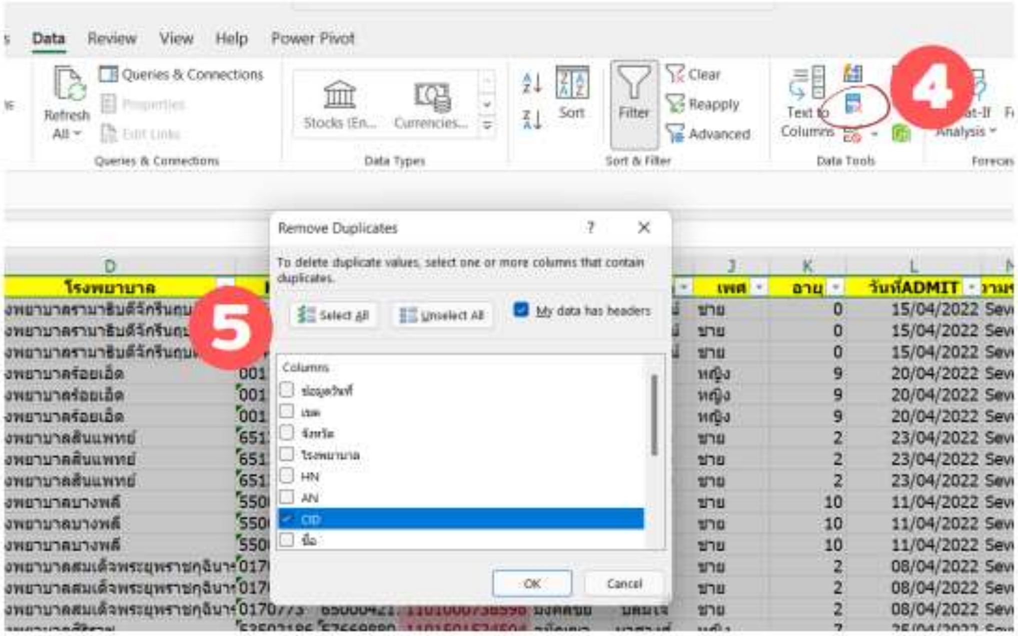


(ใช้ data severe_ex.xlsx)

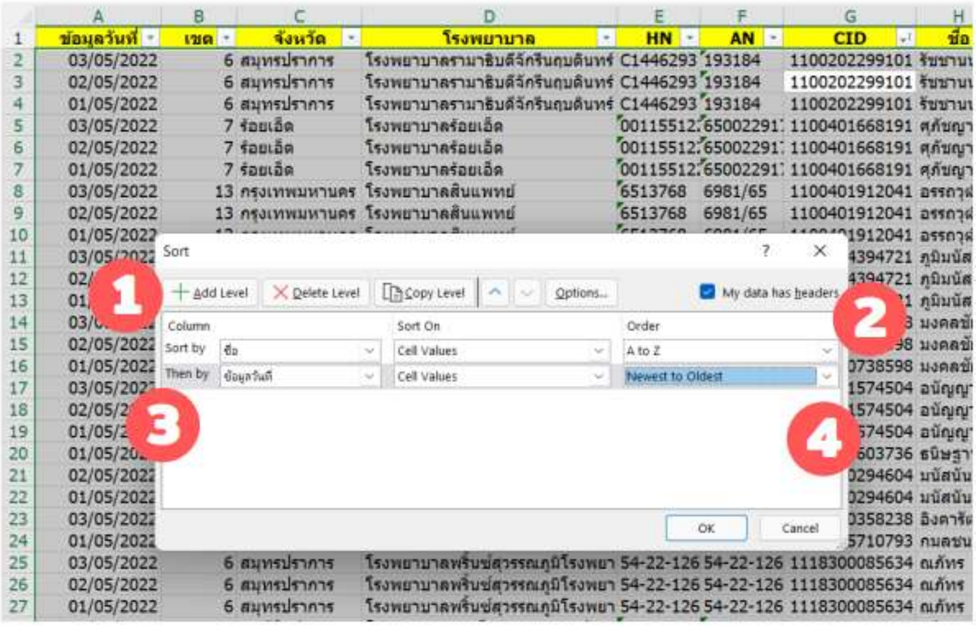
2) เลือกการแสดงผลข้อมูลที่ซ้ำกัน โดยการเทสีที่พื้นหลังและเปลี่ยนสีตัวอักษร โดยเลือก Duplicate และเลือก “Light Red Fill with Dark Red Text” ในช่อง values with (ตามหมายเลข 3)



3) ใช้เครื่องมือ Remove Duplicate (ตามหมายเลข 4) -> เลือกตัวแปรที่ต้องการตัดซ้ำ แล้วกด “OK” (ตามหมายเลข 5)

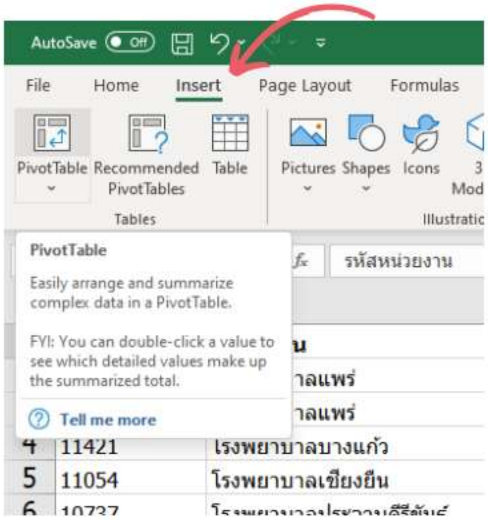


4) การตัดซ้ำข้อมูลแบบมีเงื่อนไขด้วยตัวแปร “วันที่” โดยคลิกที่ Data-> Sort
   - เลือกตัวแปร “ชื่อ” แล้ว เรียงลำดับจาก A to Z หรือ ก-ฮ (ตามหมายเลข 1 และ 2)
   - เลือกข้อมูล “วันที่” จาก Newest to oldest (ตามหมายเลข 3)
   - ทำการตัดซ้ำข้อมูล “ชื่อ” โดยใช้เครื่องมือ Remove Duplicates ในแถบเมนู Data โดยวิธีนี้จะเก็บข้อมูลไว้เฉพาะบรรทัดบนสุดเท่านั้น ดังนั้นต้องเรียงข้อมูลซ้ำกันไว้บนสุดและเรียงตามวันที่จากใหม่ไปเก่า (ตามหมายเลข 4)
   - กด OK จะได้ผลลัพธ์ข้อมูลที่ตัดซ้ำเรียบร้อยแล้ว

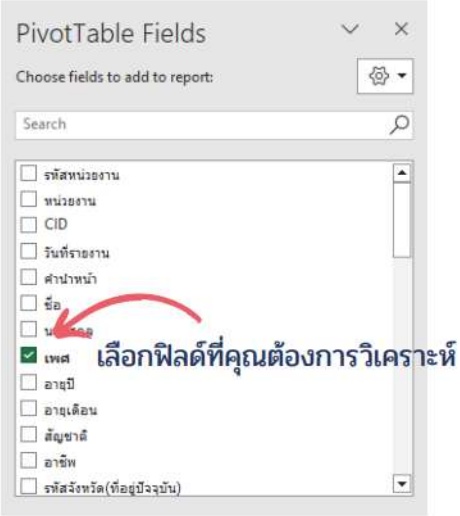


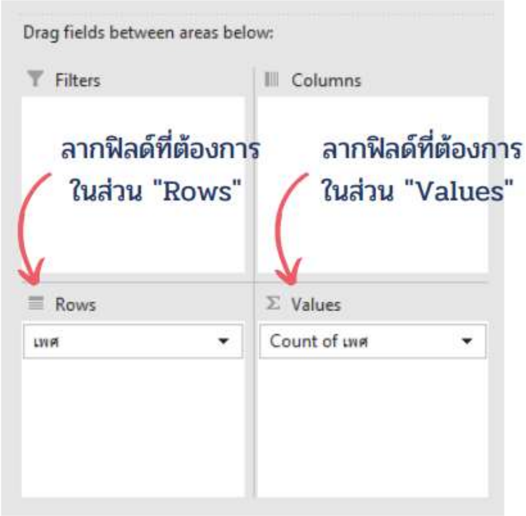
2.2 PivotTable คือฟังก์ชันที่ใช้ในการวิเคราะห์ สรุปผล คำนวณ นำข้อมูลจำนวนมาก ๆ มาสรุปผล ในรูปแบบของตาราง ดังนี้

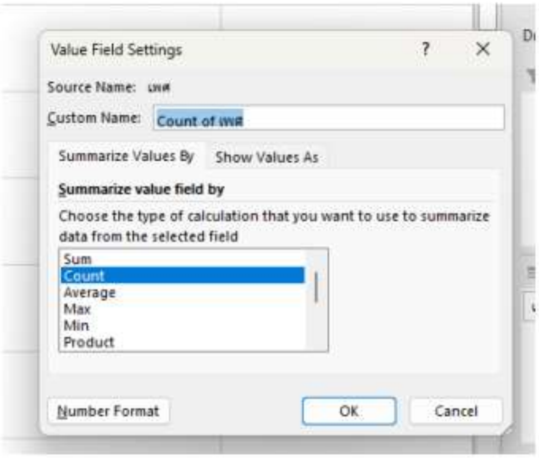
1) เลือกแถบเมนู “Insert”
2) คลิก “Pivotable”
3) ในหน้าต่าง “Create Pivot table” ตรวจสอบว่าช่องข้อมูลที่ถูกเลือกใน “Table/Range” ถูกต้องหรือไม่ สามารถเลือกช่องข้อมูลใหม่โดยคลิกที่ปุ่มเลือกช่องข้อมูล และเลือกช่องข้อมูลที่ต้องการ
4) คลิก “OK”



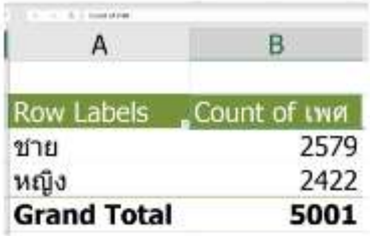
5) เลือกฟิลด์ (ตัวแปร) ที่ต้องการวิเคราะห์ และเลือกวิธีวิเคราะห์ที่ต้องการ เช่น count จำนวนแยกรายเพศ ดังรูป







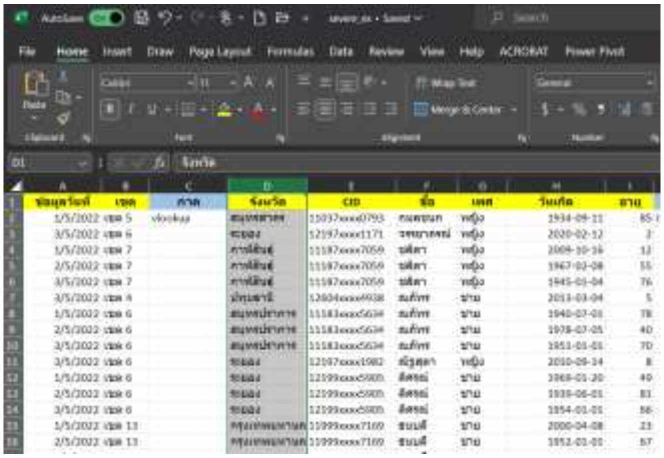
6) ผลลัพธ์ ที่ได้แสดงดังรูป



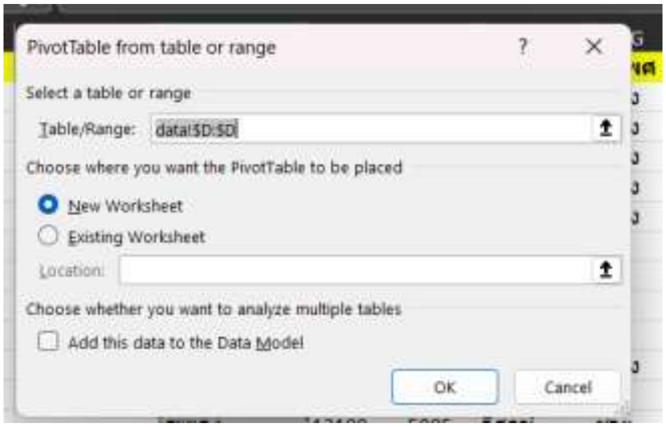
A	B
Row Labels	Count of เพศ
ชาย	2579
หญิง	2422
Grand Total	**5001**

ทั้งนี้ PivotTable จะสรุปข้อมูลตามที่กำหนด โดยสามารถแก้ไขหรือปรับแต่งได้ตามต้องการ และสามารถหาผลรวม (Sum) หรือนับจำนวน (Count) หาค่าเฉลี่ย (Average) ค่ามากที่สุด (Max) น้อยที่สุด (Min) ฯลฯ ได้โดยการคลิกขวาในช่อง Value

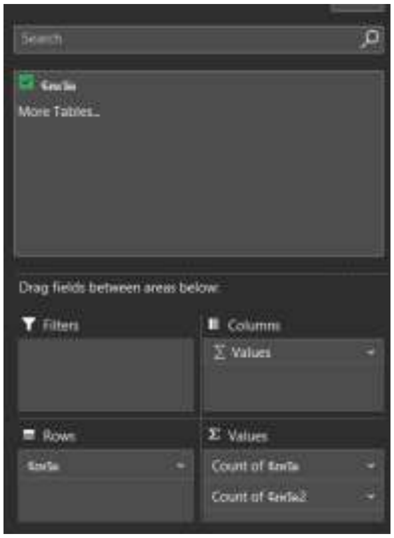
7) ตัวอย่างการใช้เครื่องมือ PivotTable วิเคราะห์ข้อมูลในไฟล์ตัวอย่าง เพื่อหาจำนวนผู้ป่วยรายจังหวัด



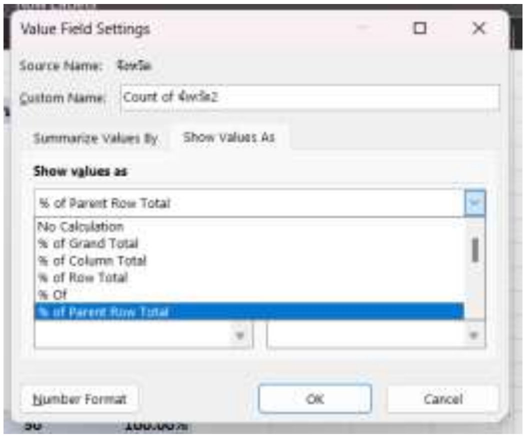
- เลือกคอลัมน์จังหวัด
- เลือกแถบเมนู Insert -> เลือก PivotTable และเลือกสร้าง sheet ใหม่ -> OK

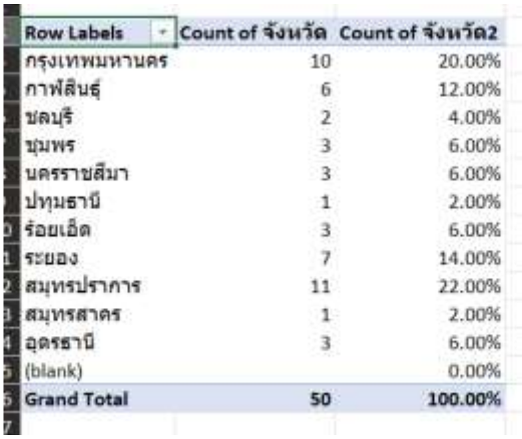


- เลือกฟิลด์ (ตัวแปร) ที่ต้องการ คือ จังหวัด (เนื่องจากข้อมูลแต่ละแถว (row) เป็นข้อมูลผู้ป่วยรายบุคคล ที่ระบุวันที่ป่วย จังหวัด และตัวแปรอื่นๆ) ส่วนในช่อง “∑ Values” เลือกเป็น “count of จังหวัด” เพื่อหาจำนวนผู้ป่วยต่อจังหวัด



- หากจะวิเคราะห์เป็นร้อยละให้เลือก Value Field Setting โดยเลือกแถบเมนู Show Values As และเลือกรายการแสดงผลการวิเคราะห์ที่ต้องการ เช่น ตัวอย่างเลือก “% of Parent Row Total” จะได้ผล “% ของผู้ป่วยรายจังหวัด” ดังรูป





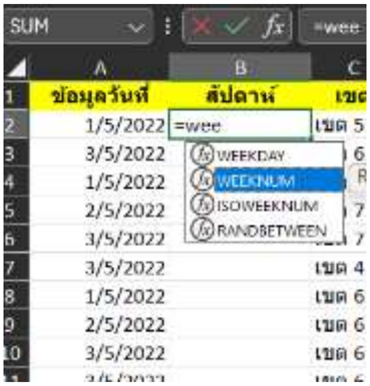
Row Labels	Count of จังหวัด	Count of จังหวัด2
กรุงเทพมหานคร	10	20.00%
กาฬสินธุ์	6	12.00%
ชลบุรี	2	4.00%
ชุมพร	3	6.00%
นครราชสีมา	3	6.00%
ปทุมธานี	1	2.00%
ร้อยเอ็ด	3	6.00%
ระยอง	7	14.00%
สมุทรปราการ	11	22.00%
สมุทรสาคร	1	2.00%
อุดรธานี	3	6.00%
(blank)		0.00%
Grand Total	50	100.00%

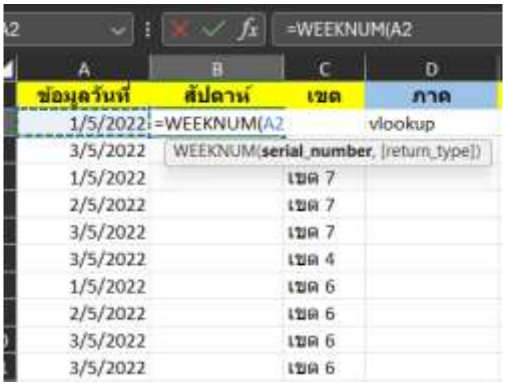
2.3 ฟังก์ชัน Day, Week, Month , Year ใช้แยกข้อมูล เป็น Day, Month, Year, Week โดยเขียนสูตรดังนี้

- **Day: = DAY(serial_number)**
- **Month: = MONTH(serial_number)**
- **Year: = YEAR(serial_number)**
- **Week: เพื่อแยกข้อมูลเป็น Week สามารถใช้สูตรที่ให้ผลลัพธ์เป็นหมายเลขของสัปดาห์ได้โดยใช้ WEEKNUM(serial_number)**

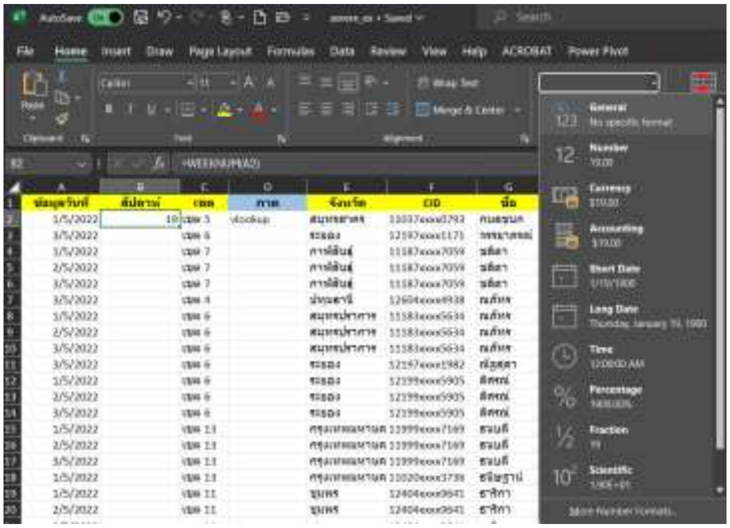
- **ตัวอย่างการแปลงวันที่เป็นสัปดาห์โดยใช้ Week**
    - เพิ่มคอลัมน์และให้ชื่อคอลัมน์ว่า “สัปดาห์”

    - ช่องใต้ชื่อคอลัมน์“สัปดาห์” ใส่สูตร “=WEEKNUM” ตามด้วยข้อมูลในช่อง “ข้อมูลวันที่”





    - จะได้ค่าในช่อง “สัปดาห์” และไปกำหนดรูปแบบให้เป็นข้อความ (General) ได้ผลดังรูป



2.4 ฟังก์ชัน IF logic เป็นฟังก์ชันที่ใช้ในการสร้างเงื่อนไข ตามที่ต้องการกำหนด เช่น

- เปลี่ยนข้อมูล จาก Column C = 1 ให้เป็นคำว่า “Yes” และอื่นๆ ที่เหลือเป็นคำว่า “No”
- เปลี่ยนจากอายุ (ปี) เป็นกลุ่มอายุ โดยมีตัวอย่างการเขียนสูตรเพื่อวิเคราะห์ “อายุ” (column E) เป็น “กลุ่มอายุ” ที่ประกอบไปด้วยกลุ่มอายุ 0-9 ปี, 10-19 ปี, 20-29 ปี, 30-39 ปี, 40-49 ปี, 50–59 ปี และ 60 ปีขึ้นไป ดังนี้

ตัวอย่าง

=IF(C2=1,“Yes”,“No”)

ตัวอย่างการใช้ IF ในการหากลุ่มอายุ

=(IF(E2=0,"",IF(E2<9,"0-9",IF(E2<=19,"10-19",
IF(E2<=29,"20-29",IF(E2<=39,"30-39",
IF(E2<=49,"40-49",IF(E2<=59,"50-59","60+")))))))

2.5 ฟังก์ชัน DATEDIF เป็นฟังก์ชันที่ใช้จัดการวันที่

DATEDIF(start_date,end_date,unit)

"Y" = จำนวนปีแบบเต็มในช่วงเวลา
"M" = จำนวนเดือนแบบเต็มในช่วงเวลา
"D" = จำนวนวันในช่วงเวลา

ตัวอย่าง

=DATEDIF("12/04/1991","29/06/2022","y")
=DATEDIF(U5,NOW(),"y")
(U5 = ตัวแปรวันที่ เช่น วันเกิด)

(ใช้ data severe_ex.xlsx)

- ตัวอย่างการสร้างตัวแปรใหม่เพื่อหาอายุปี เดือน วัน ณ ปัจจุบัน

1. สูตรการคำนวณอายุ ปี =DATEDIF(ช่องเซล,TODAY(),“Y”)
2. สูตรการคำนวณอายุ เดือน =DATEDIF(ช่องเซล,TODAY(),“YM”)
3. สูตรการคำนวณอายุ วัน =DATEDIF(ช่องเซล,TODAY(),“MD”)

	Q	R	S
	อายุปีปัจจุบัน	อายุเดือนปัจจุบัน	อายุวันปัจจุบัน
22	=DATEDIF(I2,TODAY(),"Y")		ฟังก์ชัน DATEDIF
22	3		
22	14		
22	56		
22	79		
22	10		
22	83		
22	45		
22	73		
22	13		

2.6 ฟังก์ชัน VLOOKUP

VLOOKUP ย่อมาจาก Vertical Lookup เป็นฟังก์ชันที่ใช้ค้นหาข้อมูลในตาราง โดยดึงค่าจากคอลัมน์หนึ่งมาแสดงในอีกคอลัมน์หนึ่ง โดยมีรูปแบบการทำงานดังนี้

- พิจารณาจากซ้ายไปขวา
- ตั้งค่า Range Lookup เป็น TRUE หรือ 1 หรือเว้นไว้ไม่กรอก แปลว่าแม้ไม่เจอค่าที่ต้องการหาก็ยังส่งค่าบางอย่างกลับมาได้
- ตั้งค่า Range Lookup เป็น FALSE หรือ 0 แปลว่าต้องเจอค่า/คำ ที่ต้องการเท่านั้น จึงจะส่งค่าที่ตรงกันกลับมา

= VLOOKUP(lookup_value, table_array, col_index_num,[range_lookup])

= VLOOKUP(ค้นหาคำนี้,จากช่วงข้อมูลในตารางนี้,เมื่อเจอแล้วให้เอาค่าในคอลัมน์ที่.... ของตารางกลับมาแสดงผล (นับซ้ายไปขวา)

- ตัวอย่างเปลี่ยน “เขตสุขภาพ”(Column C) เป็น “ภาค” (Column D)

  ใช้สูตรใน “ภาค” = VLOOKUP(C2,เขต!A1:B14,2,FALSE)

=VLOOKUP(C2,เขต!A1:B14,2,FALSE)

B	C	D	E	
สัปดาห์	เขต	ภาค	จังหวัด	
19	เขต 5	= VLOOKUP(C2,เขต!A1:B14,2,FAL		
	เขต 6	ภาคตะวันออก	ระยอง	1
	เขต 7	ภาคตะวันออกเ	กาฬสินธุ์	1
	เขต 7	ภาคตะวันออกเ	กาฬสินธุ์	1
	เขต 7	ภาคตะวันออกเ	กาฬสินธุ์	1
	เขต 4	#N/A	ปทุมธานี	1
	เขต 6	ภาคตะวันออก	สมุทรปราการ	1
	เขต 6	#N/A	สมุทรปราการ	1
	เขต 6	#N/A	สมุทรปราการ	1
	เขต 6	#N/A	ระยอง	1
	เขต 6	#N/A	ระยอง	1
	เขต 6	#N/A	ระยอง	1
	เขต 6	#N/A	ระยอง	1
	เขต 13	ภาคกลาง	กรุงเทพมหานค	1
	เขต 13	#N/A	กรุงเทพมหานค	1
	เขต 13	#N/A	กรุงเทพมหานค	1
	เขต 13	#N/A	กรุงเทพมหานค	1
	เขต 11	#N/A	ชุมพร	1
	เขต 11	#N/A	ชุมพร	1

2.7 Text Function ใช้สำหรับแปลงค่าต่าง ๆ ให้เป็นรูปแบบข้อความ โดยมีตัวอย่างการใช้งานดังนี้

1) เทคนิคการต่อข้อมูลโดยใช้สูตร CONCAT เช่น ตัวอย่างต่อคำว่า “จังหวัด” ในคอลัมน์ D กับชื่อจังหวัดในคอลัมน์ E

```
     D        E       F
1 จังหวัด  กระบี่
2 จังหวัด  ชลบุรี
คอลัมป์ F
   แถว 1 = CONCAT(D1,H1) ---> จังหวัดกระบี่
   แถว 2 = CONCAT(D2,H2) ---> จังหวัดชลบุรี
```

2) รวมข้อความโดยใช้ TEXTJOIN

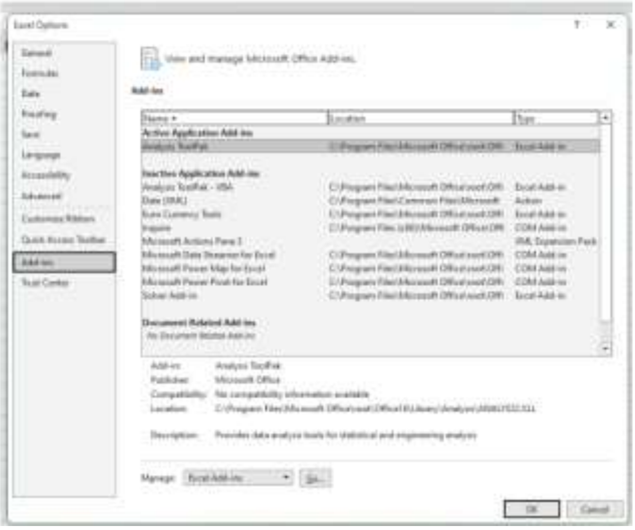
```
=TEXTJOIN(" ", TRUE, F1:F2)
ผลลัพธ์ -> จังหวัดกระบี่ จังหวัดกระบี่

=TEXTJOIN(" , ", TRUE, F1:F2)
ผลลัพธ์ -> จังหวัดกระบี่ , จังหวัดกระบี่
```

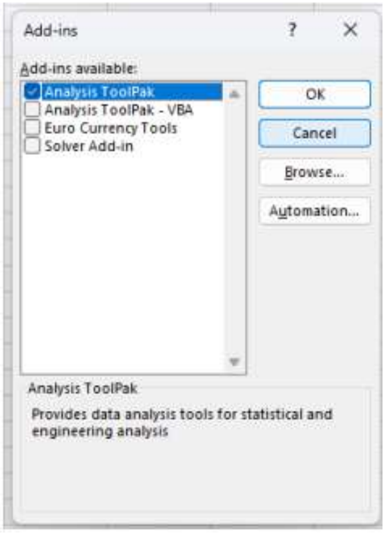
3. การวิเคราะห์ข้อมูลทางสถิติ ด้วยฟังก์ชัน Analysis ToolPak

Analysis Toolpak เป็นเครื่องมือใน Excel ที่ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลทางสถิติอย่างง่าย โดยไม่ต้องเขียนโค้ด โดยมีตัวอย่างการใช้งานดังนี้

1) ขั้นตอนการติดตั้ง Analysis Toolpak

- คลิกแท็บ File คลิก Option แล้วคลิก Add-in
- ใน Manage ให้เลือก Excel Add-ins แล้ว คลิก Go...
- เลือก Analysis Toolpak

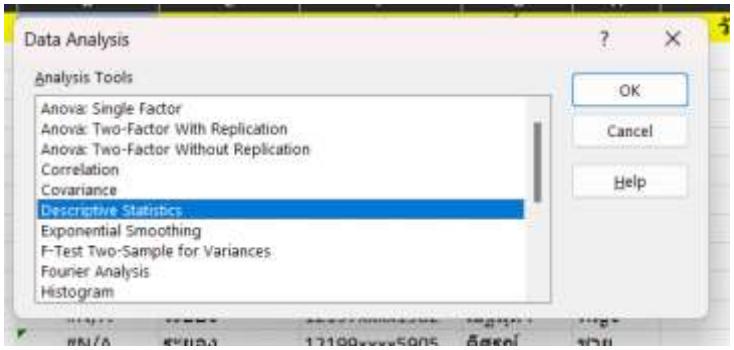




2) **การวิเคราะห์ข้อมูลทางสถิติเชิงพรรณา** (Descriptive Analysis)

2.1) ตัวอย่างการสรุปข้อมูลเบื้องต้นทางสถิติ Summary statistics (Max, Min, mean, median, SD)

- เลือก Data Analysis ที่แท็บ Data -> Descriptive Statistics และเลือกช่วงข้อมูลที่ต้องการวิเคราะห์ เช่น ในที่นี้เลือกคอลัมน์อายุ



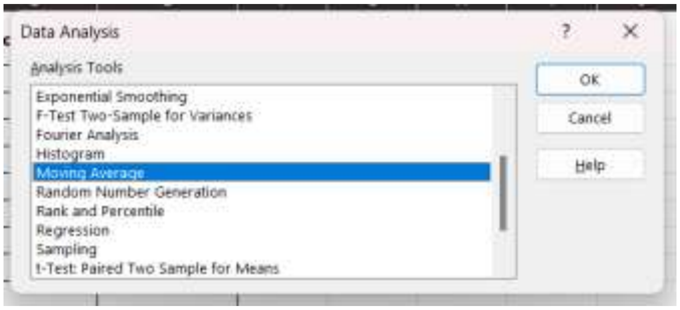
- ได้ผลลัพธ์ดังรูป

	A	B
1	อายุ	
2		
3	Mean	50.66
4	Standard Error	3.83162
5	Median	54.5
6	Mode	55
7	Standard Deviation	27.09365
8	Sample Variance	734.0657
9	Kurtosis	-1.05382
10	Skewness	0.43504
11	Range	88
12	Minimum	1
13	Maximum	89
14	Sum	2533
15	Count	50
16	Confidence Level(95.0%)	7.699929

3. Time Series Model

3.1 การวิเคราะห์ Time Series Model แบบ Moving Average ใช้ไฟล์ชื่อ “Time Series”

- เลือก Data Analysis ที่แท็บเมนู Data
- เลือก Moving Average และเลือกช่วงที่ต้องการ ในที่นี้เราต้องการหาจำนวนผู้ป่วยในเดือนธันวาคม

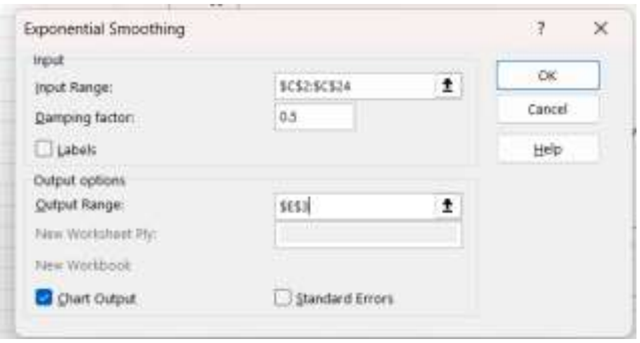


ได้ผลลัพธ์ ดังรูป

ปี	เดือน	จำนวนผู้ป่วย	Moving Avg	Exponential smoothing
2564	Jan	43		
	Feb	24	#N/A	#N/A
	Mar	20	#N/A	43.00
	Apr	24	29	33.50
	May	38	22.67	26.75
	Jun	41	27.33	25.38
	Jul	36	34.33	31.69
	Aug	41	38.33	36.34
	Sep	37	39.33	36.17
	Oct	30	38.00	38.59
	Nov	30	36.00	37.79
	Dec	19	32.33	33.90
2565	Jan	12	26.33	31.95
	Feb	15	20.33	25.47
	Mar	24	15.33	18.74
	Apr	28	17.00	16.87
	May	32	22.33	20.43
	Jun	25	28.00	24.22
	Jul	22	28.33	28.11
	Aug	30	26.33	26.55
	Sep	35	25.67	24.28
	Oct	42	29.00	27.14
	Nov	48	35.67	31.07
	Dec		41.67	36.53

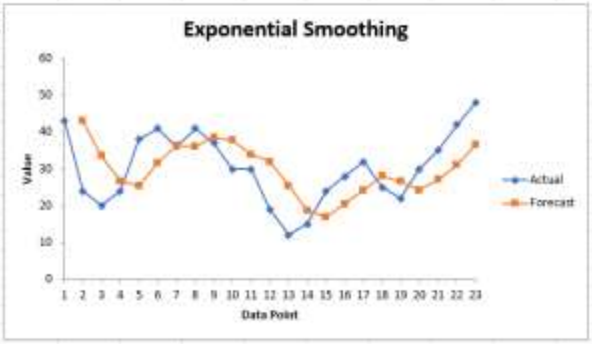
3.2 การวิเคราะห์ Time Series Model แบบ Exponential smoothing

- เลือก Data Analysis ที่แท็บเมนู Data
- เลือก Exponential smoothing และเลือกช่วงที่ต้องการ เช่น หาจำนวนผู้ป่วยในเดือน ธ.ค.



- ได้ผลลัพธ์ และกราฟเปรียบเทียบค่าจริงและค่าที่ได้จากการพยากรณ์ ดังรูป

ปี	เดือน	จำนวนผู้ป่วย	Moving Avg	Exponential smoothing
2564	Jan	43		
	Feb	24	#N/A	#N/A
	Mar	20	#N/A	43.00
	Apr	24	29	33.50
	May	38	22.67	26.75
	Jun	41	27.33	25.38
	Jul	36	34.33	31.69
	Aug	41	38.33	36.34
	Sep	37	39.33	36.17
	Oct	30	38.00	38.59
	Nov	30	36.00	37.79
	Dec	19	32.33	33.90
2565	Jan	12	26.33	31.95
	Feb	15	20.33	25.47
	Mar	24	15.33	18.74
	Apr	28	17.00	16.87
	May	32	22.33	20.43
	Jun	25	28.00	24.22
	Jul	22	28.33	28.11
	Aug	30	26.33	26.55
	Sep	35	25.67	24.28
	Oct	42	29.00	27.14
	Nov	48	35.67	31.07
	Dec		41.67	36.53



บรรยายเรื่อง การใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพผู้สูงอายุ ด้วย Software ที่เหมาะสม

โดย ดร. จิราลักษณ์ นนทารักษ์
ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล

- Data and Statistics

Data มาจากการเก็บรวบรวมข้อมูลด้วยรูปแบบคำถามงานวิจัยที่เราสนใจ เช่น ข้อมูลพฤติกรรมสุขภาพของประชาชน โดยเป็นการรวบรวมข้อมูลเพื่อได้ข้อมูลสำหรับตอบงานวิจัยที่เฉพาะเจาะจง ข้อมูลได้มาจากกลุ่มตัวอย่างที่เป็นตัวแทนประชากรที่น่าสนใจ

Statistic คือการเอา Data มาวิเคราะห์ตามวัตถุประสงค์ที่กำหนดไว้ ด้วยวิธีที่เหมาะสมกับตัวข้อมูลในสถิติ การวิเคราะห์ข้อมูลแบ่งออกเป็น 2 ระดับ ได้แก่

1) **สถิติเชิงพรรณา (Descriptive Statistics)**

เพื่ออธิบายและสรุปข้อมูลทางสถิติของชุดข้อมูลให้เข้าใจได้ง่าย เพื่อให้คนที่ใช้ข้อมูลเข้าใจลักษณะของข้อมูลหรือคุณสมบัติของเรื่องที่ศึกษานั้น ๆ โดยไม่ต้องดูรายละเอียดของข้อมูลแต่ละตัวทั้งหมดที่มีอยู่ในชุดข้อมูล การวิเคราะห์ทางสถิติเชิงพรรณาเป็นการวิเคราะห์ข้อมูลขั้นต้น มักจะเน้นไปที่การคำนวณ ความถี่ ร้อยละ ค่าเฉลี่ย (ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน ค่าฐานนิยม) ค่าสูงสุด ค่าต่ำสุด และค่าการกระจายของข้อมูล (เช่น ส่วนเบี่ยงเบนมาตรฐาน, ช่วงขอบเขตข้อมูล) เพื่อให้ผู้ที่ใช้ข้อมูลเข้าใจลักษณะสำคัญของข้อมูลนั้น ๆ โดยรวม

2) **สถิติที่ใช้อ้างอิงไปยังกลุ่มตัวอย่างต่าง ๆ (Inferential statistics : Parameter estimation)**

เป็นสถิติการวิเคราะห์ข้อมูลขั้นสูง ที่ได้จากกลุ่มตัวอย่างเพื่อนำไปใช้อ้างอิงถึงลักษณะที่แท้จริงของข้อมูลประชากรที่ตัวอย่างนั้นเป็นตัวแทน สามารถเอาไปประมาณการได้ ประกอบด้วยการประมาณค่า และการทดสอบสมมติฐาน (T-test, Z-test, ทดสอบ Chi-square, ANOVA) การวิเคราะห์สมการถดถอย (Linear, logistic, Poisson Regression Analysis) การทดสอบความสัมพันธ์ หรือทำนายความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม

เหตุผลที่ต้องใช้สถิติ คือ

- เพื่อให้เข้าใจถึงวิธีการเก็บรวบรวมข้อมูล (Encompasses the methods of collecting)
- เพื่อใช้ในการสรุปผล (Summarizing)
- เพื่อวิเคราะห์ข้อมูลได้อย่างถูกต้อง (Analyzing)
- เพื่อสรุปผลจากข้อมูลได้อย่างถูกต้อง (Drawing a conclusions from data)

ประเภทของข้อมูลทางสถิติ แบ่งออกเป็น 3 รูปแบบ

1) **แบ่งตามลักษณะของข้อมูล** ได้แก่

1.1) ข้อมูลเชิงปริมาณ เป็นข้อมูลเชิงตัวเลข (Numerical)

- Discrete เป็นจำนวนเต็ม เช่น จำนวนวันนอนโรงพยาบาล จำนวนผู้ป่วย
- Continuous จำนวนนับที่มีช่วงข้อมูล มีทศนิยมได้ เช่น น้ำหนัก ส่วนสูง

1.2) ข้อมูลเชิงคุณภาพ

- Nominal scale เป็นข้อมูลที่ไม่สามารถเกิดขึ้นพร้อมกันได้ (Mutually exclusive) ไม่สามารถจัดอันดับได้ เช่น เพศ กรุ๊ปเลือด
- Ordinal Scale เป็นข้อมูลที่ไม่สามารถเกิดขึ้นพร้อมกันได้ แต่สามารถจัดลำดับได้ เช่น ระดับความรุนแรงของโรค ระดับการศึกษา

2) **แบ่งตามแหล่งที่เกิดของข้อมูล**

2.1) ข้อมูลปฐมภูมิ คือ ข้อมูลที่เก็บรวบรวมมาจากแหล่งข้อมูลดั้งเดิมโดยตรง ไม่ผ่านการตีความหรือวิเคราะห์จากผู้อื่น เปรียบเสมือนข้อมูลดิบที่นักวิจัยหรือผู้ใช้เป็นผู้เก็บด้วยตนเอง เช่น ข้อมูลจากการสำรวจความคิดเห็นของนักเรียนเกี่ยวกับการเรียนออนไลน์ ข้อมูลจากการทดลองหาประสิทธิภาพของยาชนิดใหม่ ข้อมูลจากการวัดมลพิษทางอากาศในเขตเมือง เป็นต้น

2.2) ข้อมูลทุติยภูมิ คือ ข้อมูลที่เก็บรวบรวมมาจากแหล่งข้อมูลอื่นๆ ไม่ได้เก็บรวบรวมมาจากแหล่งข้อมูลดั้งเดิมโดยตรง เปรียบเสมือนข้อมูลที่ผ่านการตีความ วิเคราะห์ หรือ สรุปผลมาแล้ว โดยผู้ใช้ไม่ได้มีส่วนร่วมในกระบวนการเก็บข้อมูล เช่น รายงานประจำปีของธนาคารแห่งประเทศไทย บทความวิจัยเกี่ยวกับผลกระทบของการใช้โทรศัพท์มือถือต่อสุขภาพ ข้อมูลประชากรจากเว็บไซต์สำนักงานคณะกรรมการพัฒนาเศรษฐกิจและสังคมแห่งชาติ เป็นต้น

3) **การแบ่งตามมาตรของข้อมูล**

- มาตรานามบัญญัติ (Nominal scale) เช่น เพศ อาชีพ วิชา
- มาตราเชิงอันดับ (Ordinal scale) เช่น ส่วนสูง ระดับการศึกษา กลุ่มอายุ
- มาตราอันตรภาค (Interval scale) วัดค่าของข้อมูลเป็นตัวเลขได้และสามารถบอกปริมาณของข้อมูลได้ว่าได้ว่ามีค่ามากน้อยเพียงใด ข้อมูลระดับมาตราอันตรภาค เป็นข้อมูลที่ไม่มีศูนย์แท้ (absolute zero) เช่น อุณหภูมิ จำนวนโคโลนีของเชื้อรา
- มาตราอัตราส่วน (Ratio scale) เป็นมาตรการวัดของข้อมูลเชิงปริมาณเหมือนมาตราอันตรภาค แต่ข้อมูลระดับมาตรอัตราส่วนเป็นข้อมูลที่มีศูนย์แท้ (absolute zero) เช่น ไม่มีเงิน (0 บาท)

ความแตกต่างระหว่างประเภทของข้อมูล

วิธีการทางสถิติขึ้นอยู่กับประเภทของข้อมูล เช่น ข้อมูลแบบกลุ่ม (Categorical data) และข้อมูลตัวเลขต่อเนื่อง (Numerical data) เช่น อายุ รายปี ระดับความเจ็บปวด 10 ระดับ ใช้สถิติแตกต่างกัน มีรายละเอียดดังนี้

1) แหล่งข้อมูลประเภทต่าง ๆ

- ข้อมูลประชากร เช่น ข้อมูลประชากรกลางปี ประชากรเสี่ยง แหล่งข้อมูลสามารถหาได้จากสำนักงานการทะเบียน กรมการปกครอง กระทรวงมหาดไทย หรือสำนักงานสถิติแห่งชาติ
- ข้อมูลการป่วย การตาย เช่น รายงานจากหน่วยงาน/กระทรวง แหล่งข้อมูลสามารถหาได้จากฐานข้อมูล 43 แฟ้ม, ข้อมูลการเบิกจ่าย สปสช. (e-claim)
- ข้อมูลจากโรงพยาบาล เช่น การรักษา ห้องปฏิบัติการ แหล่งข้อมูลคือ ฐานข้อมูล 43 แฟ้ม, e-claim
- ข้อมูลปัจจัยเสี่ยง งานวิจัย การสำรวจสุขภาพ แหล่งข้อมูลคือ การสำรวจสุขภาพประชากรไทยทุก 5 ปี ของคณะแพทยศาสตร์ รามาฯ ม.มหิดล การสำรวจพฤติกรรมสุขภาพ เช่น เหล้า บุหรี่ การออกกำลังกาย อาหาร ของสำนักงานสถิติแห่งชาติ และสำนักโพลต่างๆ เช่น อนามัยโพล

2) คุณภาพข้อมูล (Data quality) หมายถึง ความสามารถของข้อมูลในการตอบสนองต่อความต้องการของผู้ใช้งาน โดยข้อมูลที่ดีควรมีคุณสมบัติดังต่อไปนี้

- ความถูกต้อง (Accuracy): ข้อมูลต้องตรงกับความเป็นจริง
- ความครบถ้วน (Completeness): ข้อมูลต้องมีครบถ้วนสมบูรณ์ ไม่ขาดหาย
- ความสอดคล้องกัน (Consistency): ข้อมูลที่เหมือนกันควรมีค่าเหมือนกัน ไม่ขัดแย้งกัน
- ความถูกต้องตามเวลา (Timeliness): ข้อมูลต้องเป็นปัจจุบัน ทันต่อเหตุการณ์
- ความเป็นเอกลักษณ์ (Uniqueness): ข้อมูลแต่ละชุดต้องไม่ซ้ำกัน
- ความสามารถในการเข้าถึง (Accessibility): ข้อมูลต้องสามารถเข้าถึงได้ง่าย โดยผู้ที่ได้รับอนุญาต
- ความปลอดภัย (Security): ข้อมูลต้องได้รับการป้องกันจากการเข้าถึง แก้ไข หรือลบโดยผู้ที่ไม่ได้รับอนุญาต

3) ความถูกต้องของข้อมูล หมายถึง ข้อมูลที่สะท้อนความเป็นจริงได้อย่างถูกต้อง ตรงไปตรงมา โดยปราศจากข้อผิดพลาดหรือบิดเบือนใดๆ เปรียบเสมือนข้อมูลที่เชื่อถือได้และสามารถนำไปใช้งานได้อย่างมั่นใจได้แก่

- ตรงกับความเป็นจริง: ข้อมูลควรตรงกับสิ่งที่เกิดขึ้นจริง ไม่ควรมีการแต่งเติมหรือดัดแปลง
- ปราศจากข้อผิดพลาด: ข้อมูลไม่มีข้อผิดพลาดทางข้อเท็จจริง ตัวเลข หรือการคำนวณ
- สอดคล้องกัน: ข้อมูลที่เหมือนกันควรมีค่าเหมือนกัน ไม่ขัดแย้งกัน
- ทันต่อเหตุการณ์: ข้อมูลควรเป็นปัจจุบัน ไม่ควรเก่าหรือล้าสมัย
- สามารถตรวจสอบย้อนกลับได้: ข้อมูลควรมีแหล่งที่มาที่ชัดเจน และสามารถตรวจสอบย้อนกลับไปยังแหล่งข้อมูลต้นฉบับได้

4) การวิเคราะห์ข้อมูล

- การวิเคราะห์ข้อมูลต้องมีความชัดเจน สามารถบรรลุวัตถุประสงค์ได้ เช่น เพื่อศึกษาความชุกของโรคซิฟิลิสในกลุ่ม MSM ⟶ จำนวนผู้ป่วยเก่า + จำนวนผู้ป่วยใหม่ : ตัวหารคืออะไร
- การวิเคราะห์เพื่อบรรยายสถานการณ์ เช่น Descriptive analysis (คน เวลา สถานที่): ดูการกระจายของโรค
- การวิเคราะห์เพื่อหาปัจจัยเสี่ยง Analytical analysis (Relative risk, Incident rate ratio, Odds ratio, Hazard ratio)

5) การนำเสนอข้อมูลที่ดี มีดังนี้

- ถูกต้องและชัดเจน (Be accurate and clear)
- ให้ข้อมูลอธิบาย ตัวเอง (Let the data speak) โดยแสดงข้อมูลให้มากที่สุดเท่าที่จะเป็นไปได้ แต่ต้องระวังไม่ให้บดบังข้อความหลัก
- เป็นวิทยาศาสตร์ (Science no sales) โดยหลีกเลี่ยงสิ่งที่ไม่จำเป็น หรือตัดสิ่งที่ไม่จำเป็นออก

6) รูปแบบการนำเสนอ

การนำเสนอต้องมีความง่ายในการทำความเข้าใจและตีความ และอาจนำเสนอด้วยตาราง กราฟ แผนภูมิ แผนที่ ฯลฯ

6.1) การนำเสนอตาราง ตารางเป็นวิธีการนำเสนอข้อมูลที่เป็นระเบียบ ช่วยให้อ่านเข้าใจง่าย เปรียบเทียบข้อมูล และค้นหาข้อมูลที่ต้องการได้รวดเร็ว องค์ประกอบของตารางประกอบด้วยดังนี้

- หัวข้อตาราง : บอกชื่อหรือหัวข้อของข้อมูลที่นำเสนอ
- หัวคอลัมน์ : บอกชื่อหรือประเภทของข้อมูลในแต่ละคอลัมน์
- แถว: บอกรายการข้อมูล
- เซลล์: ช่องที่เกิดจากการตัดกันของหัวคอลัมน์และแถวขอบตาราง: เส้นที่ล้อมรอบตาราง
- ชื่อตารางต้องชัดเจน มีการจัดเรียงลำดับตัวแปร เรียงตามอักษร/เรียงตามขนาด และตารางไม่ซับซ้อนเกินไป ตัวแปรไม่ทับซ้อนกัน จำนวนกลุ่มตัวแปรไม่มากหรือน้อยไป ดังตัวอย่างภาพด้านล่าง

Prepared by Global Health 50/50
Last updated: 01/04/2020. Data reported from last available date of sex-disaggregated reporting based on the 35 countries with the highest number of confirmed cases

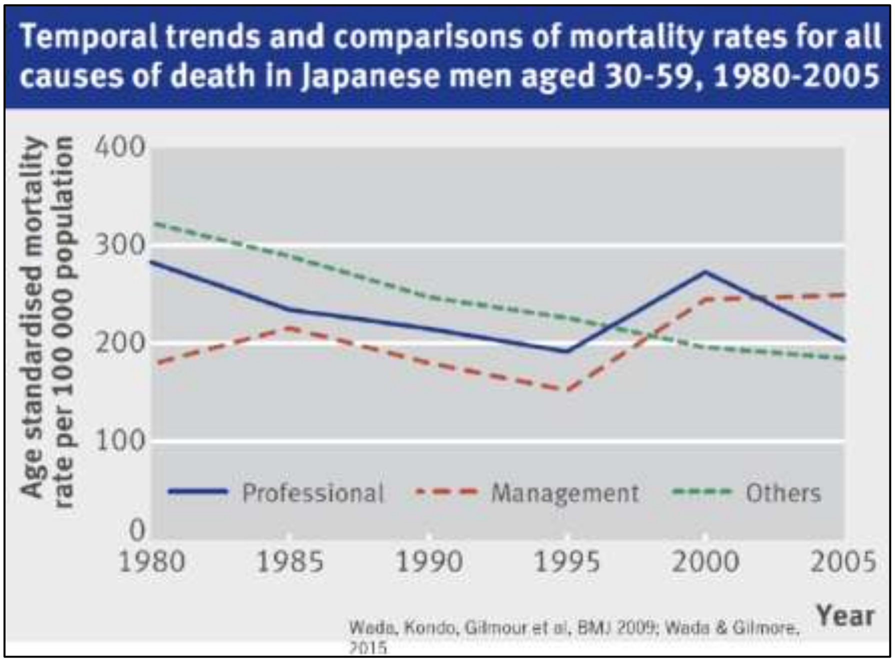
GLOBAL HEALTH 5050

COVID-19: Sex-disaggregated data tracker

Country	Date	Confirmed cases			Deaths			Proportion of deaths among confirmed cases			
		Total	% male	% female	Total	% male	% female	Total	Male	Female	Male:female ratio
USA	31/03	163,539	na	na	2860	na	na	1.7%	na	na	na
Italy	30/03	94,312	56%	44%	10026	69%	31%	10.6%	13.3%	7.4%	1.79
Germany	31/03	61,913	52%	48%	583	66%	34%	0.9%	1.2%	0.7%	1.77
China	28/02	55,924	51%	49%	2114	64%	36%	3.8%	4.7%	2.8%	1.68
Spain	30/03	46,885	50%	50%	2505	65%	35%	5.3%	6.9%	3.8%	1.82
UK	31/03	25,150	na	na	1789	na	na	7.1%	na	na	na
Switzerland	31/03	16,176	48%	52%	373	61%	39%	2.3%	2.9%	1.7%	1.70
Pakistan	31/03	15,709	na	na	25	na	na	0.2%	na	na	na
Iran	17/03	14,991	57%	43%	853	59%	41%	5.7%	5.9%	5.4%	1.09
Belgium	31/03	13,964	47%	52%	828	58%	41%	5.9%	7.2%	4.7%	1.53
The Netherlands	01/04	13,614	50%	50%	1173	61%	39%	8.6%	10.6%	6.6%	1.60
Turkey	31/03	13,531	na	na	214	na	na	1.6%	na	na	na
Austria	01/04	10,372	51%	49%	146	na	na	1.4%	na	na	na
South Korea	01/04	9,887	40%	60%	165	51%	49%	1.7%	2.1%	1.4%	1.56
Portugal	31/03	8,251	45%	54%	187	59%	41%	2.3%	3.0%	1.7%	1.79
Canada	31/03	7,708	51%	49%	89			1.2%	na	na	na
France	15/03	6,378	47%	53%	161	58%	42%	2.5%	3.1%	2.0%	1.55
Brazil	31/03	5,717	na	na	201	na	na	3.5%	na	na	na
Israel	01/04	5,591	na	na	21	na	na	0.4%	na	na	na
Norway	01/04	4,655	51.19%	49%	28	na	na	0.6%	na	na	na
Sweden	31/03	4,435	52.50%	48%	180	58%	42%	4.1%	4.5%	3.6%	1.27
Australia	01/04	4,325	51.40%	49%	20	na	na	0.5%	na	na	na

ig. 1 Sex-disaggregated data of confirmed COVID-19 cases and deaths provided by Global Health 50%50 data tracker as of April 2, 2020 [16]

6.2) การนำเสนอด้วยแผนภาพ การใช้แผนภาพหรือกราฟิกต่างๆ ประกอบการนำเสนอเพื่อช่วยให้ผู้ฟังเข้าใจเนื้อหาได้ดียิ่งขึ้น แผนภาพสามารถช่วยสื่อสารข้อมูลที่ซับซ้อนให้อยู่ในรูปแบบที่เรียบง่าย เข้าใจง่าย และดึงดูดความสนใจ อธิบายตัวเองได้ มีชื่อแผนภาพที่ชัดเจน อธิบายทุกตัวแปร ใช้แกนของกราฟที่เหมาะสม เช่น



- **วิธีการทางสถิติ**

การเลือกใช้วิธีทางสถิติ ขึ้นกับชุดข้อมูลที่ใช้วิเคราะห์ (Data analysis set) ซึ่งเป็นกลุ่มข้อมูลที่เตรียมไว้สำหรับการวิเคราะห์ ประกอบด้วยตัวแปรต่าง ๆ ที่เกี่ยวข้องกับการศึกษานั้น

ขั้นตอนการวิเคราะห์ทางสถิติ (Statistical analysis procedure) เป็นกระบวนการที่ใช้เครื่องมือทางสถิติในการวิเคราะห์ข้อมูล ตั้งแต่การเตรียมข้อมูล การเลือกใช้เทคนิคการวิเคราะห์ที่เหมาะสม ไปจนถึงการตีความผลลัพธ์ ประเภทของการวิเคราะห์ข้อมูลมี ดังนี้

1) การวิเคราะห์เบื้องต้น (Primary analysis): การวิเคราะห์ข้อมูลตามวัตถุประสงค์ที่ตั้งไว้
2) การวิเคราะห์เชิงลึก (Secondary analysis): การวิเคราะห์ข้อมูลเพิ่มเติมจากการวิเคราะห์เบื้องต้นเพื่อหาคำตอบที่ละเอียดขึ้น
3) การวิเคราะห์เชิงสำรวจ (Exploratory analysis): การวิเคราะห์ข้อมูลเบื้องต้นเพื่อทำความเข้าใจข้อมูล และเตรียมข้อมูลสำหรับการวิเคราะห์ขั้นต่อไป ตัวอย่างการวิเคราะห์ข้อมูลเพื่อทำความเข้าใจข้อมูลและเตรียมข้อมูล ได้แก่การวิเคราะห์หา
    - ข้อมูลที่หายไป (Missing Data): ข้อมูลที่ขาดหายไปในบางจุดของชุดข้อมูล
    - ค่าสุดโต่ง (Outliers): ข้อมูลที่มีค่าเบี่ยงเบนไปจากข้อมูลส่วนใหญ่
    - การวิเคราะห์กลุ่มย่อย (Subgroup analysis): การวิเคราะห์ข้อมูลแยกตามกลุ่มต่างๆ เพื่อดูความแตกต่างของผลลัพธ์ระหว่างกลุ่ม
    - การวิเคราะห์ตัวแปรควบคุม (Covariates): ตัวแปรที่อาจส่งผลต่อความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม แต่ไม่ได้เป็นตัวแปรที่สนใจในการศึกษานั้น

ตัวอย่างชุดข้อมูลสำหรับการวิเคราะห์เชิงสำรวจ เช่น

1) ข้อมูลนัดหมาย หมายถึง ข้อมูลที่ระบุวัน เวลา สถานที่ และรายละเอียดอื่นๆ เกี่ยวกับการนัดหมายไว้ล่วงหน้า
2) ข้อมูลครัวเรือน หมายถึง ข้อมูลเกี่ยวกับบุคคลที่อาศัยอยู่ในบ้านหรือที่อยู่อาศัยเดียวกัน โดยไม่คำนึงถึงว่าบุคคลเหล่านั้นมีความสัมพันธ์ฉันญาติพี่น้องหรือไม่ก็ตาม ข้อมูลครัวเรือนมักประกอบด้วย ข้อมูลทั่วไปของครัวเรือน และข้อมูลส่วนบุคคลของสมาชิกในครัวเรือน
3) ข้อมูลการตรวจร่างกาย ข้อมูลที่รวบรวมได้จากการตรวจร่างกายผู้ป่วยโดยแพทย์หรือบุคลากรทางการแพทย์ ข้อมูลเหล่านี้ใช้เพื่อวินิจฉัยโรค ประเมินสุขภาพโดยรวม และติดตามผลการรักษา
4) ข้อมูลแบบสอบถาม หมายถึง ข้อมูลที่รวบรวมได้จากการให้ผู้ตอบแบบสอบถามตอบคำถามในชุดคำถามที่จัดเตรียมไว้ ข้อมูลเหล่านี้ใช้เพื่อวัตถุประสงค์ที่หลากหลาย เช่น

- การวิจัย: ใช้เพื่อศึกษาความคิดเห็น พฤติกรรม ความเชื่อ หรือข้อมูลอื่นๆ ของกลุ่มเป้าหมาย
- การประเมินผล: ใช้เพื่อประเมินผลโครงการ กิจกรรม หรือ นโยบาย
- การพัฒนา: ใช้เพื่อพัฒนาสินค้า บริการ หรือ นโยบาย

5) ข้อมูลทางห้องปฏิบัติการ คือ ข้อมูลที่รวบรวมจากการทดลองหรือการวัดในห้องปฏิบัติการ ข้อมูลนี้สามารถใช้เพื่อวัตถุประสงค์ที่หลากหลาย เช่น การวิจัย การพัฒนา และการควบคุมคุณภาพ ประเภทของข้อมูลทางห้องปฏิบัติการ ได้แก่ :
   - ข้อมูลเชิงปริมาณ: ข้อมูลนี้แสดงในรูปแบบตัวเลข เช่น ความสูง น้ำหนัก หรืออุณหภูมิ
   - ข้อมูลเชิงคุณภาพ: ข้อมูลนี้แสดงในรูปแบบของคำหรือวลี เช่น สี กลิ่น หรือพื้นผิว
   - ข้อมูลกราฟิก: ข้อมูลนี้แสดงในรูปแบบของกราฟ แผนภูมิ หรือรูปภาพ

● การจัดข้อมูลและวิเคราะห์ข้อมูล

สำหรับการจัดข้อมูลและวิเคราะห์ข้อมูล โดยใช้ตัวอย่างจริงในการฝึกการจัดการข้อมูลและวิเคราะห์ข้อมูล จากการศึกษาเรื่อง "ภาวะพึ่งพิงในผู้สูงอายุไทย" อ้างอิงจาก Nontarak, J., Bundhamcharoen, K., Prasitsiriphon, O., & Aekplakorn, W. (2023). The Association of Sociodemographic Variables and Unhealthy Behaviors with Limitations in activities of daily living among Thai older adults: cross-sectional study and projected trends over the next 20 years. Asian/Pacific Island Nursing Journal, 7, e42205.

หลักการและเหตุผล

- สังคมผู้สูงอายุ หมายถึง มีสัดส่วนประชากรอายุ 60 ปีขึ้นไป ที่ร้อยละ 20
- ประเทศไทยกำลังเข้าสู่สังคมผู้สูงอายุ และจะเข้าเต็มที่ในปี ค.ศ. 2030 จำนวนผู้สูงอายุที่เพิ่มขึ้นมาพร้อมกับโรคไม่ติดต่อเรื้อรังที่เพิ่มขึ้น เช่น เบาหวาน ความดันโลหิตสูง โรคหัวใจและหลอดเลือด และโรคมะเร็งต่างๆ เป็นต้น
- คำถามงานวิจัยที่สำคัญ คือ การมีชีวิตที่ยืนยาวขึ้นมาพร้อมกับสุขภาพที่ดีหรือไม่ หรือมีข้อจำกัดในการทำกิจวัตรประจำวันหรือไม่ เช่น การอาบน้ำ แต่งตัว เดิน ขึ้นบันได การขับถ่าย เป็นต้น
- การสำรวจสุขภาพประชาชนไทย พบว่า มีผู้สูงอายุ ร้อยละ 11 ที่มีข้อจำกัดในการทำกิจวัตรประจำวัน และมีสัดส่วนสูงขึ้นเมื่ออายุเพิ่มขึ้น (จากร้อยละ 11 เป็น 17.6)
- พฤติกรรมเสี่ยงทางสุขภาพ เช่น การมีกิจกรรมทางกายต่ำ ภาวะอ้วน ภาวะไขมันในเลือดสูง ภาวะน้ำตาลในเลือดสูง มีความสัมพันธ์กับภาวะพิการในผู้สูงอายุ ซึ่งปัจจัยเหล่านี้สามารถป้องกันและควบคุมได้ (Modifiable risk factors)

วัตถุประสงค์ของการศึกษา

- เพื่อหาความชุกของระดับการมีภาวะพึ่งพิงในผู้สูงอายุในประเทศไทย
- เพื่อศึกษาปัจจัยที่มีความสัมพันธ์กับระดับการมีภาวะพึ่งพึงในผู้สูงอายุในประเทศไทย

การออกแบบการศึกษา

จะเห็นว่าการศึกษาครั้งนี้ไม่ใช้การศึกษาความเสี่ยง แต่จะใช้การศึกษาปัจจัยที่มีความสัมพันธ์แทนเนื่องจากรูปแบบการศึกษาเป็นแบบ Cross-sectional study เป็นการศึกษาที่เก็บข้อมูลตัวแปรต้นและตัวแปรตามพร้อมๆกัน ณ ช่วงเวลาใดเวลาหนึ่ง เหมือนเป็น snapshot วัด exposure วัดปัจจัย ตัวแปร x และ y พร้อมกัน เช่น ระดับ ADL กับการออกกำลังกาย เพื่อทดสอบสมมติฐานว่า ผู้สูงอายุมีกิจกรรมทางกายน้อยมีความสัมพันธ์กับการเกิดภาวะพึ่งพิงหรือไม่ ซึ่งเราบอกไม่ได้ว่าอะไรเกิดก่อนหลัง ดังนั้นวัตถุประสงค์ของงานนี้ คือหาปัจจัยที่มีความสัมพันธ์กับระดับการมีภาวะพึ่งพึงในผู้สูงอายุ แต่ไม่สามารถบอกได้ว่าสิ่งใดเกิดก่อน แต่ถ้าอยากรู้ต้องไปศึกษาแบบ Cohort study

วิธีการศึกษา

การศึกษานี้ ใช้รูปแบบการศึกษาแบบภาคตัดขวาง (Cross sectional study) โดยใช้ข้อมูลทุติยภูมิจากการสำรวจสุขภาพประชากรไทย โดยการตรวจร่างกายครั้งที่ 5 พ.ศ. 2557 ของสถาบันวิจัยระบบสาธารณสุข (วิชัย เอกพลากร และคณะ, 2559)

ตัวแปรที่ใช้ศึกษา

ข้อมูลตัวแปรที่ใช้

ตัวแปร	รายละเอียด
ภาวะพึ่งพิง (Barthel index)	การมีข้อจำกัดในกิจวัตรประจำวัน ในการกิน การแต่งตัว การอาบน้ำ การขับถ่าย และการเดิน แบ่งเป็น 3 ระดับ คือ Mild limitation, moderate to severe, no limitation
ข้อมูลลักษณะประชากร	เพศ อายุ ระดับการศึกษา สถานภาพสมรส
สถานะสุขภาพ	ได้แก่ เบาหวาน ภาวะอ้วน ความดันโลหิตสูง
พฤติกรรมสุขภาพ	ได้แก่ การสูบบุหรี่ การดื่มเครื่องดื่มที่มีแอลกอฮอล์ การมีกิจกรรมทางกาย และการบริโภคผักและผลไม้

สถิติที่ใช้

- สถิติเชิงพรรณนา (Descriptive Statistics) หาค่าร้อยละ (Percentage) ค่าความถี่ (Frequency) ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)
- สถิติอนุมาน (Inferential Statistics) เพื่อศึกษาปัจจัยที่สัมพันธ์กับระดับการมีภาวะพึ่งพิงในผู้สูงอายุไทย ใช้การวิเคราะห์ข้อมูลด้วยสถิติถดถอยพหุคุณ Multinomial logistic Regression

สรุปผลการศึกษา

ตารางที่ 1 ลักษณะทางประชากรตามระดับการมีภาวะพึ่งพิง

Characteristics	No ADL limitation (n=5224), n (%)	Mild ADL limitations (n=672), n (%)	Moderate to severe ADL limitations (n=1410), n (%)
Age group (years)			
60-64	1985 (38.0)	180 (26.8)	328 (23.2)
65-69	1358 (26.0)	156 (23.2)	307 (21.8)
70-74	904 (17.3)	140 (20.8)	276 (19.6)
75-79	575 (11.0)	98 (14.6)	264 (18.7)
≥80	402 (7.7)	98 (14.6)	235 (16.7)
Sex			
Female	2748 (52.6)	419 (62.3)	918 (65.1)
Male	2476 (47.4)	253 (37.7)	492 (34.9)
Residential area			
Urban	2737 (52.4)	368 (54.8)	682 (48.4)
Rural	2487 (47.6)	304 (45.2)	728 (51.6)
Education			
No education	4378 (83.8)	587 (87.3)	1282 (90.9)
Having an education	846 (16.2)	85 (12.7)	128 (9.1)

ตารางที่ 2 ปัจจัยที่มีความสัมพันธ์กับระดับภาวะพึ่งพิงผู้สูงอายุไทย

	Mild ADL limitations				Moderate to severe ADL limitations			
	Women		Men		Women		Men	
	AOR[a]	*P* values	AOR	*P* values	AOR	*P* values	AOR	*P* values
Age group (years; reference: 60-64 years)								
65-69	1.250	.10	1.018	.91	1.462	<.001	1.007	.95
70-74	1.642	.003	1.704	.001	2.110	<.001	1.319	.03
75-79	1.894	<.001	1.566	.009	2.914	<.001	1.788	<.001
≥80	2.021	<.001	3.497	<.001	3.164	<.001	1.593	.001
Residential area (reference: urban)								
Rural	0.933	.33	0.905	.36	1.378	<.001	0.964	.62
Education level (reference: no education)								
Having education	0.634	.001	1.009	.94	0.496	<.001	0.700	<.001
Marital status (reference: single)								
Married	0.648	<.001	1.278	.07	0.949	.40	1.650	<.001
Having diabetes mellitus	1.363	.02	1.254	.08	1.185	.02	0.934	.52
Having hypertension	1.015	.86	1.38	.006	0.998	.98	1.777	<.001
BMI category (reference: normal)								
Underweight (BMI<18.5)	1.35	.06	0.916	.70	1.401	.01	1.683	.005
Overweight	0.98	.83	1.49	<.001	1.295	<.001	0.947	.56

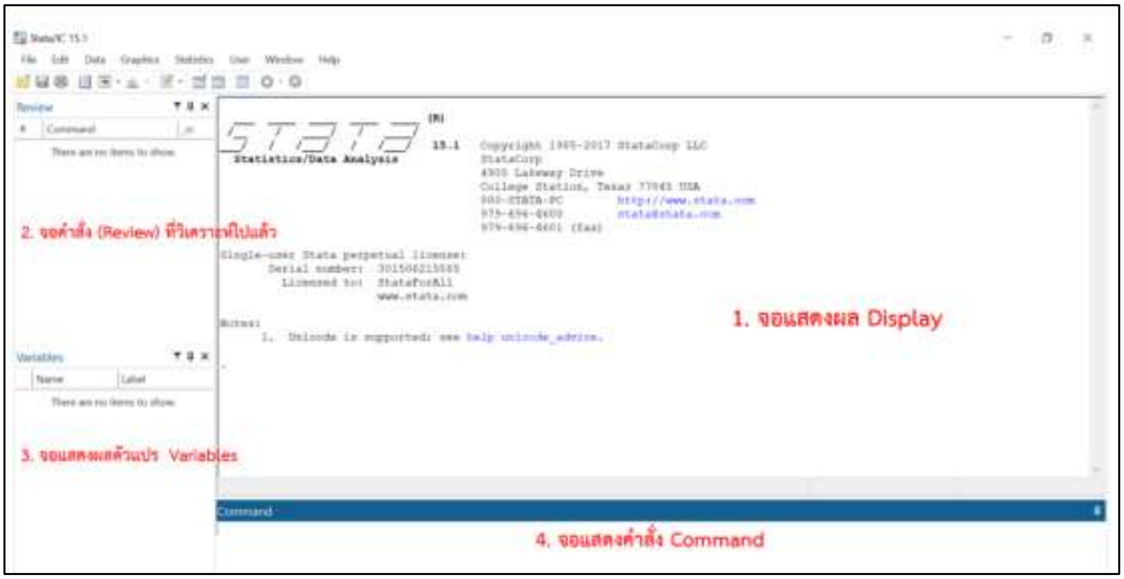
จากตารางที่ 1 และ 2 สรุปผลการศึกษา ดังนี้

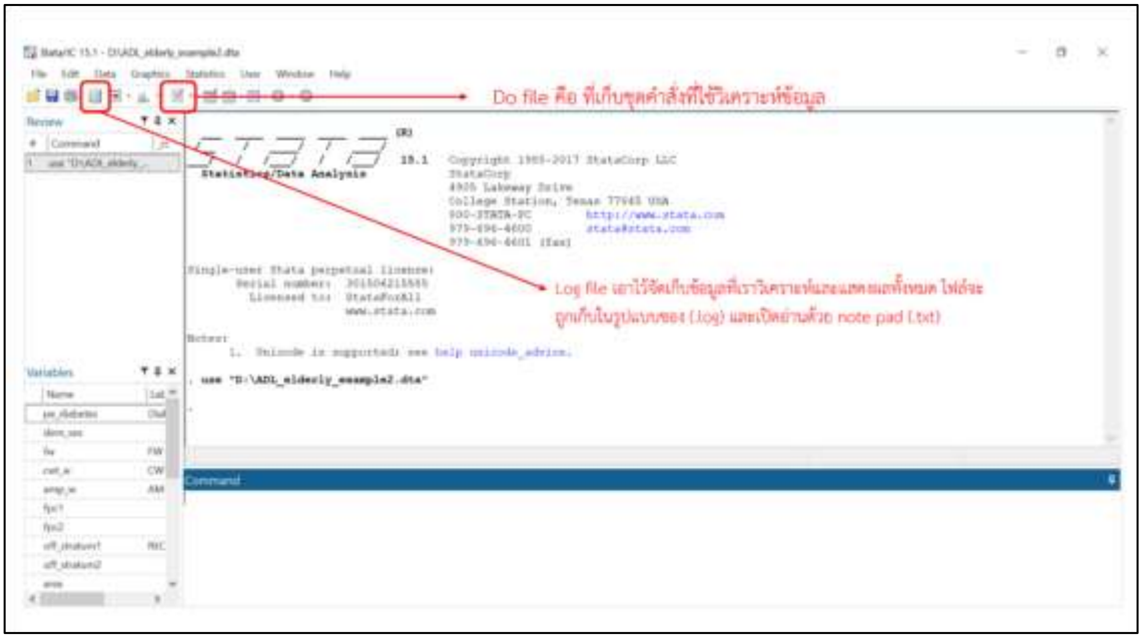
- อายุและการมีกิจกรรมทางกาย มีความสัมพันธ์กับการมีภาวะพึ่งพิง
- ระดับการมีภาวะพึ่งพิง (หรือการมีข้อจำกัดในการทำกิจวัตรประจำวัน) เพิ่มขึ้น เมื่ออายุเพิ่มขึ้น ทั้งในผู้ชายและผู้หญิง โดยผู้หญิงมีระบดับภาวะพึ่งพิงมากกว่า
- การมีโรคเรื้อรัง มีความสัมพันธ์กับการมีภาวะพึ่งพิง

- **การวิเคราะห์ด้วยโปรแกรม STATA เบื้องต้น** ประกอบด้วย

หน้าต่างโปรแกรม STATA มีรายละเอียด ดังนี้

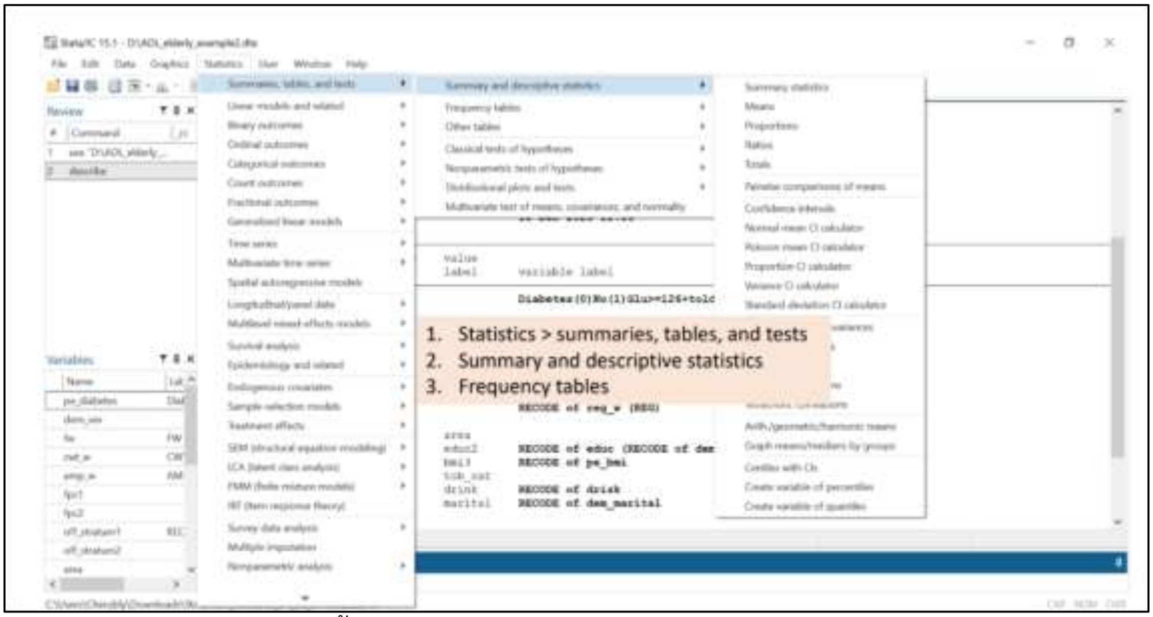
- หน้าจอที่ 1 จอแสดงผล (Display) เป็นเครื่องมือที่มีประโยชน์สำหรับการแสดง จัดรูปแบบบันทึก ส่งออก และโต้ตอบกับผลลัพธ์จากการทำงานของคำสั่ง STATA ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูล ตีความผลลัพธ์ และสื่อสารข้อมูลได้อย่างมีประสิทธิภาพ
- หน้าจอที่ 2 จอพิมคำสั่ง (Review) ใช้สำหรับดูรายละเอียดเกี่ยวกับชุดคำสั่ง STATA ที่ผ่านมา แสดงรายการคำสั่งที่ผู้ใช้พิมพ์ล่าสุด พร้อมทั้งสถานะการทำงาน ตัวแปรที่ใช้ และผลลัพธ์ที่ได้ ช่วยให้ผู้ใช้ติดตามประวัติการทำงาน ตรวจสอบข้อผิดพลาด และเรียนรู้จากคำสั่งที่ผ่านมา
- หน้าจอที่ 3 จอแสดงผลตัวแปร (Variables) ผู้ใช้สามารถดูรายละเอียดเกี่ยวกับตัวแปรในชุดข้อมูล เช่น ชื่อ ประเภท ค่า ค่าที่หายไป สถิติพื้นฐาน ฯลฯ ช่วยให้ผู้ใช้เข้าใจโครงสร้างของข้อมูล และระบุปัญหาที่อาจเกิดขึ้น เช่น ค่าที่หายไป outliers ฯลฯ สามารถแก้ไขค่าตัวแปรได้โดยตรงบนจอแสดงผลตัวแปร และสามารถวิเคราะห์ข้อมูลเบื้องต้นบนจอแสดงผลตัวแปร เช่น หาค่าเฉลี่ย ค่าสูงสุด ค่าต่ำสุด ฯลฯ
- หน้าจอที่ 4 คำสั่งจอแสดง (Command) สามารถพิมพ์คำสั่งลงใน Command Prompt เพื่อสั่งให้โปรแกรมทำงาน คำสั่งเหล่านี้สามารถควบคุมการวิเคราะห์ข้อมูล สร้างกราฟ ตาราง จัดการข้อมูล และอื่นๆ ผลลัพธ์ที่ได้ คือ ค่าทางสถิติ ตาราง กราฟ ข้อความแจ้งเตือน หรือข้อผิดพลาด เป็นต้น
- แถบ Data Editor (Edit) ใช้ซ้สำหรับดู แก้ไข และจัดการข้อมูลชุดข้อมูล
- แถบ Do file คือ ที่เก็บชุดคำสั่งใช้วิเคราะห์ข้มูล
- แถบ Log file เอาไว้จัดเก็บข้อมูลที่เราวิเคราะห์และแสดงผลทั้งหมด ไฟล์จะถูกเก็บในรูปแบบของ ".log" และเปิดอ่านด้วย note pad (.txt)





คำสั่ง STATA เบื้องต้น สำหรับการวิเคราะห์ข้อมูล แบ่งออกได้เป็น 4 กลุ่มหลัก ได้แก่

1) สถิติพรรณนา (Descriptive Statistics) ใช้สำหรับคำนวณสถิติพื้นฐาน เช่น ค่าเฉลี่ย ค่ามัธยฐาน ค่าสูงสุด ค่าต่ำสุด ช่วงระหว่างควอไทล์ ความแปรปรวน ความเบ้ ฯลฯ ช่วยให้ผู้ใช้เข้าใจภาพรวมของข้อมูล ระบุแนวโน้ม และค้นหา outliers
2) การวิเคราะห์ความสัมพันธ์ (Relationship Analysis) ใช้สำหรับวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร ตัวอย่างเช่น การวิเคราะห์ความสัมพันธ์แบบเชิงเส้น การวิเคราะห์ความสัมพันธ์แบบ Spearman's rank correlation การวิเคราะห์ ANOVA การวิเคราะห์ ANCOVA ฯลฯ ช่วยให้ผู้ใช้เข้าใจว่าตัวแปรต่างๆ สัมพันธ์กันอย่างไร และมีอิทธิพลต่อกันหรือไม่
3) การวิเคราะห์เชิงถดถอย (Regression Analysis) ใช้สำหรับสร้างโมเดลเชิงถดถอย ตัวอย่างเช่น โมเดลเชิงเส้นตรงง่ายๆ โมเดลโลจิสติก โมเดล Poisson ฯลฯ ช่วยให้ผู้ใช้คาดการณ์ค่าของตัวแปรหนึ่ง จากค่าของตัวแปรอื่นๆ และอธิบายความสัมพันธ์ระหว่างตัวแปร
4) การวิเคราะห์เพิ่มเติม (Additional Analysis) ประกอบด้วยเครื่องมือวิเคราะห์ทางสถิติอื่นๆ ตัวอย่างเช่น การวิเคราะห์ทางระบาดวิทยา การวิเคราะห์แบบ nonparametric การวิเคราะห์แบบ time series การวิเคราะห์แบบ multivariate ฯลฯ ช่วยให้ผู้ใช้วิเคราะห์ข้อมูลเชิงลึก ตอบคำถามเฉพาะ และทำการวิจัยที่ซับซ้อน



ภาพแสดงขั้นตอนการวิเคราะห์เชิงพรรณนา (Descriptive analysis)

ตัวอย่างคำสั่ง STATA ที่ใช้บ่อย สำหรับการวิเคราะห์ข้อมูลพรรณนา

Command (คำสั่ง)	คำอธิบาย
sort	จัดเรียงข้อมูลตัวแปร
codebook	เรียกดูลักษณะข้อมูล เช่น ประเภทข้อมูล (string, numeric) ช่วงข้อมูล (range) จำนวนตัวแปร ค่า missing
summarize	ดูลักษณะข้อมูลที่เป็น Numeric (เชิงปริมาณ) ในภาพรวม คือ จำนวนตัวอย่าง ค่าเฉลี่ย SD min max
describe	การอธิบายลักษณะของตัวแปร
encode	แปลงค่า string into numeric and vice versa
label	ติดป้ายชื่อตัวแปร หรือคำอธิบายตัวแปร
generate	การสร้างตัวแปรใหม่
replace	การแทนค่าเดิมของตัวแปร Replace contents of existing variable
recode	การจัดกลุ่มใหม่ของตัวแปรที่เป็นกลุ่ม Recode categorical variables
tabulate	การแสดงผลด้วยรูปแบบตาราง (จำนวน, ร้อยละ และ Cumulative percentage)
mean	การแสดงค่าเฉลี่ย
logistic	การวิเคราะห์สหสัมพันธ์ logistic regression
glob	Macro definition and manipulation เพื่อเรียกตัวแปรทุกตัวที่ต้องการวิเคราะห์
foreach	Loop over items เพื่อวิเคราะห์แบบวนลูป

การวิเคราะห์และแปรผลด้วยโปรแกรม STATA

สถิติการถดถอยโลจิสติกแบบพหุกลุ่ม (Multinomial Logistic Regression - MLR) เป็นเทคนิคการวิเคราะห์ทางสถิติที่ใช้สำหรับการประเมินหรือทำนายความน่าจะเป็น ที่ตัวอย่างข้อมูลจะอยู่ในกลุ่มใดกลุ่มหนึ่งจากกลุ่มที่เป็นไปได้มากกว่าสองกลุ่ม MLR มักใช้ในการวิเคราะห์สถานการณ์ที่ผู้ตัดสินใจต้องเผชิญกับทางเลือกมากกว่าสองทาง ใช้ประเมินผลของตัวแปรอิสระที่มีผลต่อตัวแปรตาม ใช้ระบุตัวแปรที่มีความสัมพันธ์อย่างมีนัยสำคัญกับตัวแปรตาม

ตัวอย่างการแปลผลจากสถิติ logistic regression

1) Univariate logistic regression analysis

- ผู้สูงอายุ 65 -69 ปี มีภาวะพึ่งพิงในระดับเล็กน้อย เป็น 2.18 เท่า เมื่อเทียบกับกลุ่มอายุ อายุ 60 - 64 ปี อย่างมีนัยสำคัญทาสถิติ (p-value <0.05)
- ผู้สูงอายุ 65 -69 ปี มีภาวะพึ่งพิงในระดับปานกลางถึงมาก เป็น 1.33 เท่า เมื่อเทียบกับกลุ่มอายุ 60 - 64 ปี แต่ไม่แตกต่างอย่ามีนัยสำคัญทาสถิติ (p-value > 0.05)

```
. mlogit adl3gr i.age4, rrr

Iteration 0:   log likelihood = -1680.2558
Iteration 1:   log likelihood =  -1625.333
Iteration 2:   log likelihood = -1624.1271
Iteration 3:   log likelihood = -1624.1264
Iteration 4:   log likelihood = -1624.1264

Multinomial logistic regression                 Number of obs     =      2,192
                                                LR chi2(8)        =     112.26
                                                Prob > chi2       =     0.0000
Log likelihood = -1624.1264                     Pseudo R2         =     0.0334

------------------------------------------------------------------------------
      adl3gr |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Full         |  (base outcome)
-------------+----------------------------------------------------------------
mild         |   กลุ่มอ้างอิง คือ อายุ60-64 ปี
        age4 |
      65-69  |   2.181029   .4532102     3.75   0.000     1.451392    3.277465
      70-74  |   1.977718   .4713345     2.86   0.004     1.239664    3.155185
      75-79  |   2.930663   .7041648     4.47   0.000     1.829968    4.693407
        80+  |   4.525169   1.098086     6.22   0.000     2.812423    7.280966
             |
       _cons |   .0694006   .0108195   -17.11   0.000     .0511284    .0942031
-------------+----------------------------------------------------------------
mod_severe   |
        age4 |
      65-69  |   1.328481   .2235943     1.69   0.091     .9551927    1.847649
      70-74  |   2.016645   .3510851     4.03   0.000     1.433642    2.836731
      75-79  |   2.945135    .525619     6.05   0.000     2.075826     4.17849
        80+  |   4.366391   .8085042     7.96   0.000     3.037454    6.276762
             |
       _cons |   .1419558   .0159903   -17.33   0.000     .1138339    .1770251
------------------------------------------------------------------------------
Note: _cons estimates baseline relative risk for each outcome.
```

ภาพแสดงการวิเคราะห์แปลผลจากสถิติ Univariate logistic regression analysis

2) Multivariate logistic regression analysis

- ผู้ชายมีระดับภาวะพึ่งพิงในระดับเล็กน้อย (Mild) เป็น 1.79 เท่าเมื่อเทียบกับผู้หญิง อย่างมีนัยสำคัญทางสถิติ (p-value <0.05)
- ผู้สูงอายุ 65-69 ปี มีภาวะพึ่งพิงในระดับเล็กน้อยเป็น 2.32 เท่า เมื่อเทียบกับกลุ่มอายุ 60-64 ปี อย่างมีนัยสำคัญทางสถิติ (p-value <0.05)
- คนที่มีกิจกรรมทางกายไม่พอเพียง มีภาวะพึ่งพิงในระดับเล็กน้อยเป็น 1.53 เท่าเมื่อเทียบกับคนมีกิจกรรมทางกายพอเพียงอย่างมีนัยสำคัญทางสถิติ (p-value <0.05)

```
. mlogit adl3gr sex i.age4 i.area i.educ2 i.marital i.pe_diabetes i.hbp i.tob_cat i.drink i.frveg i.physical i.bmi3, rrr

Iteration 0:   log likelihood = -1284.2199
Iteration 1:   log likelihood =  -1210.195
Iteration 2:   log likelihood = -1207.3112
Iteration 3:   log likelihood = -1207.3002
Iteration 4:   log likelihood = -1207.3002

Multinomial logistic regression                 Number of obs     =      1,733
                                                LR chi2(34)       =     153.84
                                                Prob > chi2       =     0.0000
Log likelihood = -1207.3002                     Pseudo R2         =     0.0599

------------------------------------------------------------------------------
      adl3gr |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Full         |  (base outcome)
-------------+----------------------------------------------------------------
mild         |
         sex |   1.785719   .4195372     2.47   0.014     1.126761    2.830053
             |
        age4 |
      65-69  |   2.322992   .5584965     3.51   0.000      1.45011      3.7213
      70-74  |   1.846769   .5258817     2.15   0.031     1.056882    3.226999
      75-79  |   3.014542    .870583     3.82   0.000     1.711587    5.309378
        80+  |   5.359073   1.600656     5.62   0.000     2.984362    9.623384
             |
        area |
      rural  |   .9311908    .165688    -0.40   0.689     .6570271    1.319757
             |
       educ2 |
        YES  |   1.276027   .3112188     1.00   0.318     .7911423    2.058094
             |
     marital |
    Married  |   1.025228   .1900744     0.13   0.893     .7128699    1.474453
             |
 pe_diabetes |
        yes  |   .9616923   .2006388    -0.19   0.851      .638924    1.447515
             |
         hbp |
        yes  |   1.087405   .1952772     0.47   0.641     .7647717    1.546148
             |
     tob_cat |
     Former  |   1.471344   .3653491     1.56   0.120     .9043836    2.393734
Current regular  |   .9281861   .3136286    -0.22   0.825     .4786511    1.799911
```

ภาพแสดงการแปลผลสถิติ Multivariate logistic regression analysis

บรรยาย เรื่อง เทคนิคการแปลผลจากการวิเคราะห์และการพยากรณ์ข้อมูลด้วยเครื่องมือทางสถิติ

โดย ผศ. ดร. ณัฐนารี เอมยงค์
ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล

- Study Design

1) การศึกษาเชิงสังเกต (Observational study)

หมายถึง การศึกษาวิจัยที่ผู้วิจัยไม่ได้กำหนดหรือควบคุมสิ่งที่จะศึกษา แต่เป็นการสังเกตปรากฏการณ์ตามธรรมชาติโดยไม่มีการกระทำใดๆ เข้าไปแทรกแซง มีสิ่งที่ทำให้เบี่ยงเบนไป สามารถควบคุมได้ด้วยวิธีการศึกษาและสถิติซึ่งมีความซับซ้อนของความสัมพันธ์ของข้อมูล มักใช้ในการศึกษาทางสังคมศาสตร์ พฤติกรรมศาสตร์ รวมถึงระบาดวิทยาเบื้องต้น เช่น การสำรวจความชุก เป็นต้น

ลักษณะสำคัญของการศึกษาเชิงสังเกต คือ

- ไม่มีการสุ่มกลุ่มตัวอย่างหรือควบคุมตัวแปรใดๆ ผู้วิจัยเพียงสังเกตสิ่งที่เกิดขึ้นตามธรรมชาติ
- สามารถศึกษาสาเหตุและผลกระทบได้ แต่ไม่สามารถระบุความสัมพันธ์เชิงเหตุผลได้อย่างแน่ชัด
- เป็นการศึกษาเพื่อทำความเข้าใจปรากฏการณ์ หรือสร้างสมมติฐานสำหรับการศึกษาเชิงทดลองต่อไป
- มีค่าใช้จ่ายต่ำและใช้เวลาน้อยกว่าการทดลอง

2) การศึกษาเชิงทดลอง (Experimental Study)

หมายถึง การวิจัยที่ผู้วิจัยได้กำหนดและควบคุมสิ่งทดลอง โดยมีการจงใจทำให้เกิดการเปลี่ยนแปลง (Treatment) ขึ้นในสภาวะที่ควบคุมได้ และสังเกตผลที่เกิดขึ้น เพื่อศึกษาวามสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม ถือเป็นรูปแบบการวิจัยที่มีน้ำหนักและคุณค่าสูงสุด เนื่องจากสามารถควบคุมตัวแปรต่างๆ ได้อย่างเหมาะสม สามารถอ้างอิงผลการวิจัยไปอธิบายหรือทำนายปรากฏการณ์อื่นๆ ได้ดีกว่าการวิจัยเชิงสำรวจหรือเชิงสังเกต ลักษณะสำคัญของการศึกษาเชิงทดลอง ได้แก่

- มีการกำหนดและควบคุมตัวแปรอิสระหรือตัวแปรต้นให้เกิดการเปลี่ยนแปลง
- มีการสุ่มกลุ่มตัวอย่างเพื่อควบคุมตัวแปรแทรกซ้อนหรือตัวแปรกวน
- มีกลุ่มควบคุมและกลุ่มทดลอง เพื่อเปรียบเทียบผลต่างของการเปลี่ยนแปลง
- สามารถระบุความสัมพันธ์เชิงเหตุผลได้ชัดเจน
- การวิจัยมีขั้นตอนและวิธีการที่เป็นระบบแบบแผน
- การทำในมนุษย์ยังคงจำเป็นที่จะต้องคำนึงถึงอคติต่าง ๆ เช่น ความสนใจ อายุ เพศ
- กรณีอาสาเป็นคนละคน สามารถใช้เปรียบเทียบกันได้ แต่จะไม่สามารถบอกความแตกต่างได้
- ความหมายของการมีนัยสำคัญทางสถิติ ในการเปรียบเทียบก่อน - หลัง ต่างหรือไม่ต่างกัน โปรแกรมจะเซ็ตค่าให้เป็น 0 เสมอ ซึ่งสามารถเปลี่ยนเองได้หากมั่นใจว่าการทดลองเพิ่มขึ้นได้อย่างแน่นอน

หนึ่งในปัญหาสำคัญที่นักวิจัยต้องเผชิญในปัจจุบันคือ การมีข้อมูลจำนวนมหาศาลแต่กลับไม่สามารถระบุได้ว่าควรใช้รูปแบบการศึกษาและการวิเคราะห์แบบใดจึงจะเหมาะสมที่สุดกับข้อมูลชุดนั้นๆ ซึ่งเป็นปัญหาที่ท้าทายที่ส่งผลกระทบอย่างมากต่อคุณภาพของงานวิจัย ในยุคข้อมูลขนาดใหญ่ (Big Data) ที่มีการสร้างและเก็บรวบรวมข้อมูลจำนวนมหาศาลอย่างต่อเนื่อง นักวิจัยมักประสบปัญหาความสับสนว่าควรเลือกใช้รูปแบบการศึกษาเชิงสำรวจ การศึกษาเชิงสังเกต หรือการศึกษาเชิงทดลองในการวิเคราะห์ข้อมูลขนาดใหญ่ เนื่องจากแต่ละรูปแบบการศึกษามีข้อจำกัด และข้อได้เปรียบที่แตกต่างกันในด้านการควบคุมตัวแปร การระบุความสัมพันธ์เชิงเหตุผลและความสามารถในการอ้างอิงผลไปสู่ประชากรกลุ่มใหญ่ การเลือกใช้รูปแบบการศึกษาที่ไม่เหมาะสมอาจทำให้ข้อค้นพบขาดความน่าเชื่อถือและการแปลผลที่คลาดเคลื่อน

ดังนั้น จึงเป็นสิ่งสำคัญที่นักวิจัยจะต้องมีความรู้ความเข้าใจอย่างลึกซึ้งในคุณลักษณะ ข้อจำกัด และความเหมาะสมของรูปแบบการศึกษาแต่ละประเภท เพื่อสามารถเลือกใช้รูปแบบการศึกษาที่สอดคล้องกับลักษณะของข้อมูลและวัตถุประสงค์ของงานวิจัยได้อย่างถูกต้อง นอกจากนี้ ยังจำเป็นต้องมีทักษะในการวิเคราะห์ข้อมูลขนาดใหญ่และแปลผลการวิเคราะห์ได้อย่างมีประสิทธิภาพด้วย เพื่อให้สามารถนำข้อมูลมหาศาลมาใช้ประโยชน์ในเชิงวิชาการได้อย่างเต็มศักยภาพ

- Basic Knowledge

การวัดระดับของตัวแปรหรือ Measurement Scale เป็นสิ่งสำคัญในการวิจัยเชิงปริมาณ เนื่องจากเป็นตัวกำหนดวิธีการวิเคราะห์ข้อมูลและสถิติที่เหมาะสม การเลือกใช้สถิติวิเคราะห์ต้องสอดคล้องกับระดับการวัดของตัวแปร โดยตัวแปรประเภทเดียวกันจะใช้สถิติชนิดเดียวกัน ถ้าใช้สถิติไม่ถูกต้องจะทำให้ผลการวิเคราะห์คลาดเคลื่อน ดังนั้น ควรทำความเข้าใจระดับการวัดให้ชัดเจน ก่อนเลือกสถิติวิเคราะห์ โดยแบ่งออกเป็น 4 ระดับดังนี้

1) ระดับนามบัญญัติ (Nominal Scale) เป็นการกำหนดหมายเลขหรือสัญลักษณ์เพื่อระบุหรือจำแนกประเภท ซึ่งไม่มีคำนวณทางคณิตศาสตร์ เช่น เพศ สัญชาติ สถิติที่ใช้ ได้แก่ ความถี่ ร้อยละ ข้อมูลไม่สามารถเรียงลำดับได้
2) ระดับอันดับ (Ordinal Scale) เป็นการเรียงลำดับข้อมูล โดยหมายเลขที่มากกว่า หมายถึงลักษณะนั้นๆ มีมากกว่า แต่ไม่สามารถบอกระยะห่างได้ เช่น ระดับความพึงพอใจ สถิติที่ใช้ ได้แก่ มัธยฐาน ฐานนิยม
3) ระดับช่วง (Interval Scale) สามารถบอกลำดับและระยะห่างระหว่างคะแนนได้ แต่ไม่มีศูนย์แท้ เช่น อุณหภูมิองศาเซลเซียส สถิติที่ใช้ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน สหสัมพันธ์

หากผล Lab ได้ N/A คือ มีค่าไม่เท่ากับ 0 แต่สามารถแทนที่ด้วยจำนวนใกล้เคียงมากที่สุดที่เครื่องสามารถวัดได้

---

4) ระดับอัตราส่วน (Ratio Scale) สามารถบอกลำดับ ระยะห่าง และมีศูนย์แท้ เช่น น้ำหนัก อายุ รายได้ ใช้สถิติเช่นเดียวกับระดับช่วง แต่สามารถใช้ได้มากกว่า

เรากำหนด measurement เพื่อเตือนใจ การแปลผลการศึกษา มากกว่าการเริ่มต้นใช้สถิติ

สถิติเชิงพรรณนา (Descriptive Statistics)

เป็นวิธีการทางสถิติที่ใช้สำหรับสรุป และนำเสนอข้อมูลในรูปแบบที่เข้าใจง่าย โดยมีวัตถุประสงค์หลักเพื่ออธิบายลักษณะของข้อมูลที่มีอยู่ สถิติเชิงพรรณนาที่สำคัญมีดังนี้

- ค่ากลางของข้อมูล (Measures of Central Tendency) เช่น ค่าเฉลี่ย (Mean) ค่ามัธยฐาน (Median) ค่าที่ถี่ที่สุดหรือฐานนิยม (Mode)

ค่ากลาง คือ ข้อมูลตัวแทนของประชากร จำเป็นต้องมีค่าความคลาดเคลื่อนเพื่ออธิบายว่าของมูลมีลักษณะอย่างไร

- ค่ากระจายของข้อมูล (Measures of Dispersion) เช่น พิสัย (Range) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ความแปรปรวน (Variance)

การแจกแจงประชากรตัวอย่าง (Sampling Distribution) คือ การอธิบายความเป็นตัวแทนของตัวอย่าง ซึ่งได้ค่าต่าง ๆ ออกมาแล้วสามารถเป็นตัวแทนของประชากร ซึ่งจะนำไปสู่การเลือกใช้สถิติที่เหมาะสมในการออกแบบการศึกษา ซึ่งขึ้นอยู่กับผลลัพธ์ของการศึกษา เพื่อให้เป็นไปตามวัตถุประสงค์ของการศึกษา

- การนำเสนอข้อมูลด้วยกราฟและตาราง เช่น ตารางความถี่ (Frequency Distribution Table) กราฟแท่ง (Bar Chart) กราฟวงกลม (Pie Chart) ฮิสโตแกรม (Histogram) เป็นต้น

การทำวิจัยเป็นเรื่องที่ต้องลงทุน ลงแรง การเชื่อ p-value เพียงอย่างเดียวหรือการใช้สถิติขั้นสูง อาจไม่ตอบโจทย์ แต่การทำ Data Visualization (นำเสนอข้อมูลด้วยภาพ) อาจทำให้เห็นภาพมากกว่าสำหรับข้อมูลนั้น ๆ

สถิติเชิงพรรณนา ช่วยสรุปข้อมูลจำนวนมากให้อยู่ในรูปแบบที่กระชับและเข้าใจง่าย สามารถนำไปใช้วิเคราะห์เบื้องต้นและนำเสนอผลการศึกษาได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม สำหรับการอ้างอิงไปสู่ประชากร การทดสอบสมมติฐาน หรือการหาความสัมพันธ์ จำเป็นต้องใช้สถิติเชิงอนุมานประกอบด้วย

สถิติเชิงอนุมาน (Inferential Statistics)

เป็นการนำข้อมูลจากกลุ่มตัวอย่างมาใช้อนุมานหรือคาดคะเนคุณลักษณะของประชากรทั้งหมด โดยมีหลักการสำคัญ ดังนี้

1) การทดสอบสมมติฐาน (Hypothesis Testing) เป็นกระบวนการทางสถิติที่ใช้ในการตัดสินจากข้อมูลตัวอย่างว่าจะยอมรับหรือปฏิเสธสมมติฐานที่ตั้งไว้ เกี่ยวกับค่าพารามิเตอร์ของประชากร โดยกำหนดสมมติฐานหลัก (Null Hypothesis) และสมมติฐานรอง เลือกใช้สถิติทดสอบที่เหมาะสม เช่น Z-test, T-test, ANOVA, Chi-square เป็นต้น เปรียบเทียบค่าสถิติที่คำนวณได้กับค่าวิกฤตภายใต้ระดับนัยสำคัญที่กำหนด และปฏิเสธหรือไม่ปฏิเสธสมมติฐานหลักตามหลักเกณฑ์
2) การประมาณค่าพารามิเตอร์ (Parameter Estimation) ใช้ค่าสถิติจากกลุ่มตัวอย่าง เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และกำหนดช่วงความเชื่อมั่น (Confidence Interval) เพื่อประมาณค่าพารามิเตอร์ของประชากร
3) การวิเคราะห์ความแปรปรวน (Analysis of Variance - ANOVA) ใช้ทดสอบความแตกต่างของค่าเฉลี่ยในประชากรมากกว่า 2 กลุ่ม
4) การวิเคราะห์ถดถอยและสหสัมพันธ์ (Regression and Correlation Analysis) ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม

การประมาณค่าสัมประสิทธิ์ เพื่อนำไปใช้ทำนายหรือพยากรณ์ได้

การประมาณค่า (Estimation)

เป็นกระบวนการทางสถิติที่ใช้ในการคาดคะเนค่าพารามิเตอร์ของประชากร โดยอาศัยข้อมูลจากกลุ่มตัวอย่าง มีวิธีการที่สำคัญ 2 แบบ คือ

1) การประมาณค่าจุด (Point Estimation)

- เป็นการประมาณค่าพารามิเตอร์เป็นค่าเดียว เช่น ค่าเฉลี่ยของประชากร (μ)
- ใช้ค่าสถิติจากตัวอย่าง เช่น ค่าเฉลี่ยของตัวอย่าง ($\bar{x}$) เป็นประมาณค่าเฉลี่ยของประชากร
- ค่าประมาณต้องมีคุณสมบัติที่ดี เช่น ความน่าเชื่อถือ ความเอนเอียง ประสิทธิภาพ เป็นต้น

2) การประมาณค่าช่วงความเชื่อมั่น (Confidence Interval Estimation)

- เป็นการประมาณค่าพารามิเตอร์เป็นช่วงค่า มิใช่ค่าจุดเดียว
- กำหนดระดับความเชื่อมั่น (Confidence Level) เช่น 95% หรือ 99% เป็นต้น
- ช่วงความเชื่อมั่นจะครอบคลุมค่าพารามิเตอร์ที่แท้จริงในสัดส่วนที่เท่ากับระดับความเชื่อมั่นที่กำหนด
- ขอบเขตของช่วงจะกว้างหรือแคบขึ้นอยู่กับขนาดตัวอย่างและระดับความเชื่อมั่น

การเลือกใช้สถิติวิเคราะห์ที่เหมาะสม

1) ดูที่ระดับการวัดของตัวแปร

1.1) Continuous / Primary Outcomes หรือตัวแปรต่อเนื่อง หมายถึง มีระดับการวัดเป็นอันตรภาค (Interval Scale) หรืออัตราส่วน (Ratio Scale) ซึ่งเป็นตัวแปรเชิงปริมาณที่สามารถวัดค่าได้อย่างต่อเนื่องและมีสเกลที่ละเอียด สถิติที่นิยมใช้จึงแบ่งตามวัตถุประสงค์และตัวแปร ดังนี้

- หนึ่งกลุ่ม เช่น ทดสอบค่าเฉลี่ยกับค่ามาตรฐาน ใช้ One Sample t-test
- สองกลุ่ม เช่น เปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม ใช้ Independent t-test
- สามกลุ่มขึ้นไป เช่น เปรียบเทียบความแตกต่างค่าเฉลี่ยหลายกลุ่ม ใช้ ANOVA

1.2) ในทางกลับกัน หากผลลัพธ์หลักเป็นตัวแปรประเภทกลุ่ม (Categorical/Nominal) มีระดับการวัด นามบัญญัติหรืออันดับ ต้องใช้สถิติไม่อิงพารามิเตอร์ เช่น Chi-Square เพื่อทดสอบความสัมพันธ์หรือความแตกต่าง มีการพิจารณาความเป็นไปได้และการจัดกลุ่มข้อมูลอย่างเหมาะสม เช่น การอธิบายความรู้แต่ละระดับซึ่งต้องมีคำนิยามที่ชัดเจนและตอบวัตถุประสงค์ของการศึกษา การไม่ทำความเข้าใจลักษณะข้อมูลแล้วไปใช้งานโมเดลเลย อาจจะทำให้ขาดความเข้าใจในการแปลความผลการศึกษา

2) การเลือกใช้ค่ากลาง (Measurement of Location) การเลือกใช้ค่ากลางอย่างเหมาะสมจะทำให้ได้ข้อสรุปจากข้อมูลที่ถูกต้องและตรงวัตถุประสงค์มากที่สุด โดยคำนึงถึงลักษณะเฉพาะของข้อมูลแต่ละชุด ดังนี้

- ข้อมูลที่มีความเบ้ไปทิศทางใดทิศทางหนึ่ง (Skewed Data) สำหรับข้อมูลประเภทนี้ ค่ามัธยฐาน (Median) จะเหมาะสมกว่าค่าเฉลี่ย เนื่องจากมัธยฐานไม่ได้รับอิทธิพลจากค่าสุดโต่ง
- ข้อมูลที่มีความถี่สูงมากๆ (Data with High Frequency) สำหรับข้อมูลประเภทนี้ ค่าที่ถี่ที่สุด (Mode) จะเป็นตัวแทนที่ดีที่สุด เนื่องจากเป็นค่าที่พบได้บ่อยที่สุดในชุดข้อมูลนั้น
- ข้อมูลที่มีค่าต่างกันมากๆ แต่ไม่ใช่ค่าสุดโต่ง (Data with High Variability without Outliers) สำหรับข้อมูลประเภทนี้ ค่าเฉลี่ย (Mean) จะเหมาะสมที่สุด เนื่องจากสามารถสะท้อนภาพรวมของข้อมูลได้ดี

3) Rate หรืออัตรา คือ การเปรียบเทียบสองปริมาณ โดยที่ตัวตั้งจะเป็นส่วนหนึ่งของตัวหารเสมอ และมักจะมีองค์ประกอบด้านเวลาเข้ามาเกี่ยวข้องด้วยเสมอ ตัวอย่างเช่น

อัตราป่วยต่อประชากรในปี x = จำนวนผู้ป่วย/จำนวนประชากรทั้งหมด

การระบุช่วงเวลาที่ชัดเจน เช่น ต่อปี, ต่อเดือน, ต่อวัน หรือต่อชั่วโมง จะทำให้การแปลความหมายและเปรียบเทียบอัตราต่างๆ มีความถูกต้องและชัดเจนมากขึ้น ดังนั้น ในการนำเสนอหรือวิเคราะห์อัตราใดๆ ก็ตามจำเป็นต้องระบุเวลาหรือช่วงเวลาที่ใช้ในการคำนวณอัตราไว้ด้วยเสมอ เพื่อให้สามารถเข้าใจและแปลความหมายของอัตราได้อย่างถูกต้อง

4) Ratio หรือ อัตราส่วน หมายถึงการเปรียบเทียบสองปริมาณที่ไม่ได้มีความสัมพันธ์กันโดยตรง โดยที่ตัวตั้งและตัวหารไม่จำเป็นต้องเป็นส่วนหนึ่งของกันและกัน เช่น "แพทย์กับจำนวนเตียง" นั้นเป็นอัตราส่วนที่ถูกต้อง เนื่องจากแพทย์และเตียงไม่ได้เกี่ยวข้องโดยตรงกัน แต่การเปรียบเทียบอัตราส่วนระหว่างกัน จะบ่งบอกถึงจำนวนแพทย์ต่อเตียงว่ามีความเหมาะสมหรือไม่ ดังนั้น Ratio และ Rate จึงมีความแตกต่างกัน โดย Ratio เป็นการเปรียบเทียบสองปริมาณโดยตรง ในขณะที่ Rate จะมีความเกี่ยวข้องกับปัจจัยด้านเวลาด้วย

- **การนำเสนอผล**

1) กราฟ

1.1) กราฟใช้สำหรับเปรียบเทียบเชิงปริมาณ การนำเสนอข้อมูลเชิงปริมาณผ่านกราฟเปรียบเทียบนั้นเป็นวิธีการที่มีประสิทธิภาพ เนื่องจากช่วยให้สามารถมองเห็นความแตกต่างและเปรียบเทียบข้อมูลได้อย่างชัดเจนและง่ายต่อการทำความเข้าใจ โดยกราฟที่นิยมใช้ในการเปรียบเทียบเชิงปริมาณ ได้แก่

- กราฟแท่ง (Bar Chart) เหมาะสำหรับข้อมูลประเภทกลุ่ม สามารถเปรียบเทียบค่าของแต่ละกลุ่มได้อย่างชัดเจน
- กราฟวงกลม (Pie Chart) ใช้เปรียบเทียบสัดส่วนของแต่ละส่วนต่อทั้งหมด เหมาะสำหรับข้อมูลที่แบ่งเป็นหมวดหมู่ต่างๆ
- กราฟเส้น (Line Chart) แสดงแนวโน้มของข้อมูลตามช่วงเวลา สามารถเปรียบเทียบแนวโน้มของข้อมูลหลายๆ ชุดได้พร้อมกัน
- กราฟหลายแกน (Dual-Axis Chart) ใช้แสดงข้อมูลสองชุดที่มีหน่วยวัดแตกต่างกันบนแกนเดียวกัน เหมาะสำหรับการเปรียบเทียบความสัมพันธ์ระหว่างข้อมูลสองชุด
- กราฟก้อนสะสม (Stacked Chart) เหมาะสำหรับการเปรียบเทียบสัดส่วนของแต่ละส่วนในแต่ละกลุ่ม และเปรียบเทียบขนาดของแต่ละกลุ่ม

1.2) กราฟใช้เปรียบเทียบเชิงความหมาย มักใช้สำหรับข้อมูลที่เป็นนามธรรมหรือมีความหมายเชิงคุณภาพ ซึ่งกราฟที่เหมาะสมสำหรับการเปรียบเทียบในลักษณะนี้ ได้แก่

- **กราฟแท่งคู่ขนาน (Parallel Bar Chart)** เหมาะสำหรับเปรียบเทียบคุณลักษณะหรือคุณสมบัติของสิ่งต่างๆ โดยใช้แท่งคู่ขนานแสดงระดับของคุณลักษณะนั้นๆ สามารถเปรียบเทียบได้หลายคุณลักษณะพร้อมกัน
- **กราฟเรดาร์ (Radar Chart)** เป็นกราฟวงกลมที่แสดงคุณลักษณะของสิ่งต่างๆ ในหลายๆด้าน เหมาะสำหรับเปรียบเทียบข้อมูลเชิงคุณภาพที่มีหลายตัวแปรพร้อมกัน
- **กราฟกริด (Grid Chart)** ใช้แสดงข้อมูลในลักษณะเมทริกซ์ หรืออาจแสดงด้วยสีหรือรูปร่างต่างๆ เหมาะสำหรับเปรียบเทียบกลุ่มข้อมูลในหลายๆ มิติ

- **กราฟบับเบิล (Bubble Chart)** ใช้วงกลมที่มีขนาดแตกต่างกันแทนขนาดของข้อมูล สามารถเปรียบเทียบข้อมูล 3 ตัวแปรได้พร้อมกัน
- **กราฟลูกบาศก์ (Cubes/Squares Chart)** ใช้สี่เหลี่ยมหรือลูกบาศก์ในการแสดงข้อมูล สามารถเปรียบเทียบความแตกต่างในหลายมิติด้วยการแบ่งสีหรือลวดลาย
- **Boxplot** หรือกราฟแบบกล่องภาพ เป็นกราฟที่สามารถนำเสนอข้อมูลของสถิติเชิงพรรณนาได้อย่างครบถ้วน ทั้งค่ากลาง การกระจายตัว ค่าผิดปกติ รวมถึงขอบเขตของข้อมูลได้อย่างกระชับ ช่วยให้เข้าใจลักษณะของข้อมูลได้ง่ายและรวดเร็ว เช่น
    - ค่ากลางของข้อมูล (Central Tendency) หรือ Median หรือค่ามัธยฐาน จะถูกแสดงด้วยเส้นตรงกลางกล่อง
    - ค่าการกระจายตัวของข้อมูล (Spread) Q1 หรือควอร์ไทล์ที่ 1 และ Q3 หรือควอร์ไทล์ที่ 3 จะถูกแสดงด้วยขอบล่างและขอบบนของกล่อง และ IQR หรือค่าพิสัยระหว่างควอร์ไทล์ จะเป็นความกว้างของกล่อง
    - ค่าผิดปกติ (Outliers) จุดที่อยู่นอกช่วงของแถบข้อมูลจะถูกแสดงเป็นจุดหรือเครื่องหมายพิเศษ แสดงถึงค่าผิดปกติ
    - ค่าสุดโต่ง (Extreme Values) จะถูกแสดงด้วยเครื่องหมายพิเศษ แยกต่างหากจากค่าผิดปกติ
    - คะแนนสูงสุดและต่ำสุด จะถูกแสดงด้วยแถบข้อมูลสั้นๆ ที่ปลายของแกนแสดงค่าสูงสุดและต่ำสุด

การเลือกใช้กราฟในการนำเสนอข้อมูลทางสถิตินั้น จำเป็นต้องให้ความสำคัญกับการสื่อสารความหมายให้ตรงตามวัตถุประสงค์ของการศึกษานั้นๆ เป็นสำคัญ โดยมีประเด็นที่ควรพิจารณาดังนี้

- ประเภทของข้อมูล ควรเลือกกราฟให้เหมาะสมกับลักษณะและประเภทของข้อมูลที่ต้องการนำเสนอ เช่น ข้อมูลเชิงปริมาณ เชิงคุณภาพ ข้อมูลเวลา เป็นต้น
- องค์ประกอบที่ต้องการเน้น เช่น ต้องการเปรียบเทียบความแตกต่างของข้อมูล ต้องการแสดงสัดส่วน หรือต้องการแสดงแนวโน้ม ควรเลือกกราฟที่สามารถสื่อสารจุดนั้นได้อย่างชัดเจนและตรงประเด็น
- ปริมาณข้อมูล หากมีข้อมูลจำนวนมาก อาจต้องใช้กราฟแบบง่ายและตรงไปตรงมา แต่หากมีข้อมูลน้อย สามารถใช้กราฟที่ซับซ้อนและมีรายละเอียดมากขึ้น
- กลุ่มเป้าหมาย ควรคำนึงถึงผู้รับสารด้วย หากเป็นกลุ่มทั่วไป กราฟควรมีความง่ายและชัดเจน แต่หากเป็นผู้เชี่ยวชาญ สามารถใช้กราฟที่ซับซ้อนมากขึ้นได้

หลักการสำคัญในการออกแบบกราฟเพื่อนำเสนอข้อมูลทางสถิติ คือ กราฟต้องสามารถสื่อสารและอธิบายได้ง่ายเมื่อมองในปราดเดียว โดยมีรายละเอียดดังนี้

- **ความง่ายและกระชับ** กราฟที่ดีควรมีรูปแบบที่ง่าย ไม่ซับซ้อนจนเกินไป มีเพียงองค์ประกอบที่จำเป็นในการสื่อสารข้อมูล ทำให้ผู้ชมสามารถจับประเด็นสำคัญได้ง่ายด้วยการมองเพียงปรากฏการณ์แรก
- **ใช้สีและรูปร่างที่เหมาะสม** การเลือกใช้สี รูปร่าง และสัญลักษณ์ต่างๆ ในกราฟอย่างเหมาะสมจะช่วยสื่อสารข้อมูลได้ชัดเจนและง่ายต่อการจดจำ
- **บรรยายสั้นๆ กระชับ** การใช้ชื่อกราฟ คำอธิบายแกนและองค์ประกอบต่างๆ ที่กระชับและตรงประเด็น จะทำให้เข้าใจกราฟได้ง่ายขึ้น
- **ไม่มีรายละเอียดเกินความจำเป็น** การแสดงเพียงข้อมูลสำคัญที่ต้องการสื่อสาร โดยไม่นำเสนอรายละเอียดปลีกย่อยที่ไม่จำเป็น จะทำให้กราฟดูเรียบง่ายและง่ายต่อการรับรู้
- **องค์ประกอบที่สอดคล้องและกลมกลืน** การจัดวางองค์ประกอบต่างๆ เช่น ชื่อกราฟ แกนกราฟ คำอธิบาย ให้มีความสอดคล้องและกลมกลืนกันจะช่วยให้กราฟดูเป็นระเบียบและสบายตา

การเลือกใช้กราฟแต่ละประเภทนั้น ควรพิจารณาจากวัตถุประสงค์ในการสื่อสารข้อมูลเป็นสำคัญ แทนที่จะยึดติดกับกราฟประเภทใดประเภทหนึ่ง ควรรลองทำกราฟหลายๆ แบบก่อน แล้วเลือกกราฟที่อ่านง่ายและสื่อสารได้ตรงประเด็นที่สุด เป็นวิธีการที่ดีที่สุด

การนำเสนอที่ดี

- **ความชัดเจน** ในการนำเสนอต้องมีความชัดเจน ตรงประเด็น ใช้ภาษาที่กระชับ เข้าใจง่าย หลีกเลี่ยงการใช้คำศัพท์ที่ซับซ้อนหรือคลุมเครือ
- **เหมาะสมกับข้อมูล** ควรเลือกวิธีการนำเสนอที่เหมาะสมกับประเภทของข้อมูล เช่น ข้อมูลเชิงปริมาณอาจใช้กราฟหรือแผนภูมิ ข้อมูลเชิงคุณภาพอาจใช้การบรรยายหรือแผนผังความคิด
- **ให้ข้อมูลพูดเอง** (Let the data speak) การนำเสนอที่ดีควรออกแบบให้ข้อมูลสามารถสื่อสารและอธิบายตัวเองได้ โดยไม่ต้องมีการอธิบายเพิ่มเติมจากผู้นำเสนอ ผ่านการใช้กราฟ ตาราง คำอธิบาย ที่ชัดเจนและครบถ้วน
- **มีความเป็นวิทยาศาสตร์** การนำเสนอควรมีความเป็นวิทยาศาสตร์ นำเสนอข้อมูลจากหลักฐานเชิงประจักษ์ มีการอ้างอิงแหล่งที่มา มีการควบคุมตัวแปรต่างๆ อย่างเหมาะสม ไม่นำเสนอด้วยความรู้สึกหรือความคิดเห็นส่วนตัว
- **มีตรรกะและความสมเหตุสมผล** การวิเคราะห์และสรุปผลต้องมีตรรกะและมีเหตุผลรองรับ การเชื่อมโยงประเด็นต่างๆ ต้องสมเหตุสมผลและสอดคล้องกับข้อมูลที่มีอยู่

2) การนำเสนอด้วยตาราง

- การเรียงลำดับตัวแปร ควรเรียงตามวัตถุประสงค์ของการนำเสนอ เช่น จำนวนปัญหา ลำดับตัวอักษร หรือลักษณะทางภูมิศาสตร์
- การจัดกลุ่มข้อมูล ควรจัดกลุ่มให้สอดคล้องกับวัตถุประสงค์การศึกษา เพื่อให้ง่ายต่อการทำความเข้าใจ
- ทศนิยม ควรใช้จำนวนเดียวกันทั้งตาราง เพื่อให้ผู้อ่านเข้าใจความหมายที่ต้องการสื่อได้ง่าย
- การรวมร้อยละ ต้องระวังการรวมร้อยละตามแถวหรือคอลัมน์ เนื่องจากคำอธิบายอาจไม่สอดคล้องกับวัตถุประสงค์การศึกษา เช่น กรณีเปรียบเทียบ n ไม่เท่ากัน หรือ n เท่ากันแต่ต้องการเปรียบเทียบระดับ
- การเขียนอธิบายตาราง การลองเขียนอธิบายความหมายของตารางจะช่วยให้สามารถเลือกการรวมร้อยละข้อมูลได้อย่างเหมาะสม
- การจัดการข้อมูลและการนำเสนอด้วยตารางต้องคำนึงถึงวัตถุประสงค์ของงานเป็นหลัก เพื่อให้ผู้อ่านสามารถเข้าใจข้อมูลได้ง่ายและตรงประเด็น
- การรวมร้อยละของข้อมูลตามแถวหรือคอลัมน์ในตารางนั้นเป็นประเด็นที่ต้องระมัดระวัง เนื่องจากคำอธิบายที่ได้อาจไม่สอดคล้องกับวัตถุประสงค์ของการศึกษา
- ยกตัวอย่างเช่น ในกรณีที่ทำการเปรียบเทียบข้อมูล แต่จำนวนตัวอย่าง (n) ไม่เท่ากัน หรือในกรณีที่จำนวนตัวอย่างเท่ากันแต่ต้องการเปรียบเทียบระดับของข้อมูล การรวมร้อยละของข้อมูลอาจส่งผลให้เกิดความเข้าใจที่คลาดเคลื่อนได้ ดังนั้น การลองเขียนคำอธิบายเพื่ออธิบายความหมายของตารางจะช่วยให้สามารถเลือกวิธีการรวมร้อยละของข้อมูลได้อย่างเหมาะสมและตรงประเด็นมากยิ่งขึ้น

ควรเก็บข้อมูลทุกอย่างในรูปแบบ Continuous เริ่มแรก เนื่องจากสามารถปรับเปลี่ยนข้อมูลได้ตามความต้องการเมื่อนำมาวิเคราะห์ และสามารถนำข้อมูลไปใช้ได้หลากหลายในการศึกษา

- **ตัวอย่างข้อมูลเกี่ยวกับการแปลความหมายของสถิติต่างๆ** ที่ใช้ในการวิเคราะห์ข้อมูลทางสิ่งแวดล้อม มีดังนี้
  - Mean & SD เป็นสถิติเชิงพรรณนา บอกค่ากลางและการกระจายของข้อมูล
  - 95% CI หรือช่วงความเชื่อมั่น อธิบายว่าหากทำการวัดข้อมูลนั้น 100 ครั้ง ค่าเฉลี่ยจะไม่แตกต่างกัน 95 ครั้ง มีโอกาสเพียง 5% ที่ข้อมูลจะแตกต่างออกไป จึงมีความเป็นตัวแทนสูง

- OR หรือ Odd Ratios ถ้าค่าใกล้ 0 หมายถึงโอกาสสัมผัสน้อยมาก ในกรณีนี้หมายถึงไม่มีโอกาสสัมผัสตะกั่วเลย ซึ่งเป็นเป้าหมายสำหรับมาตรการจัดการเรื่องตะกั่ว
- Median และ IQR เนื่องจากข้อมูลการสัมผัสสิ่งแวดล้อมมักเบ้ไปทางใดทางหนึ่ง Median จึงแสดงจุดกึ่งกลาง ส่วน IQR แสดงค่าคนที่ได้รับสัมผัส 25% แรกและ 25% หลัง

● Sampling

- การทดสอบสมมติฐาน
- ค่าที่ตั้งสมมติฐานไม่จำเป็นต้องเท่ากับ 0 เสมอไป สามารถกำหนดเป็นมากกว่า น้อยกว่า หรือเท่ากับค่าอื่นๆ ได้ตามความเหมาะสม
- ขนาดตัวอย่าง (Sample Size) มีผลต่อการกำหนดช่วงความเชื่อมั่น (95% CI) ขนาดตัวอย่างที่ใหญ่พอจะทำให้ช่วงความเชื่อมั่นแคบลง
- โปรแกรมสถิติมีข้อจำกัดในการแสดงผล เช่น SPSS แสดงเพียง P-Value ผู้วิจัยจึงต้องเข้าใจการกำหนดสมมติฐานและวิธีการอ่านค่าที่เหมาะสมกับข้อมูล
- การอ่านผลการทดสอบสมมติฐานจะต้องเลือกระหว่าง "ปฏิเสธสมมติฐาน" หรือ "ไม่สามารถปฏิเสธสมมติฐาน" เท่านั้น ไม่สามารถ "ยอมรับสมมติฐาน" ได้ เนื่องจากมีปัจจัยความคลาดเคลื่อนอื่นๆ ที่อาจเกี่ยวข้อง

● การเลือกใช้สถิติ

การเลือกใช้สถิติที่เหมาะสมนั้นมีความสำคัญอย่างยิ่งต่อการวิเคราะห์ข้อมูลและการสรุปผลการวิจัย เนื่องจากจะส่งผลโดยตรงต่อความถูกต้อง น่าเชื่อถือ และความหมายของผลที่ได้รับ โดยมีประเด็นสำคัญที่ควรพิจารณาดังนี้

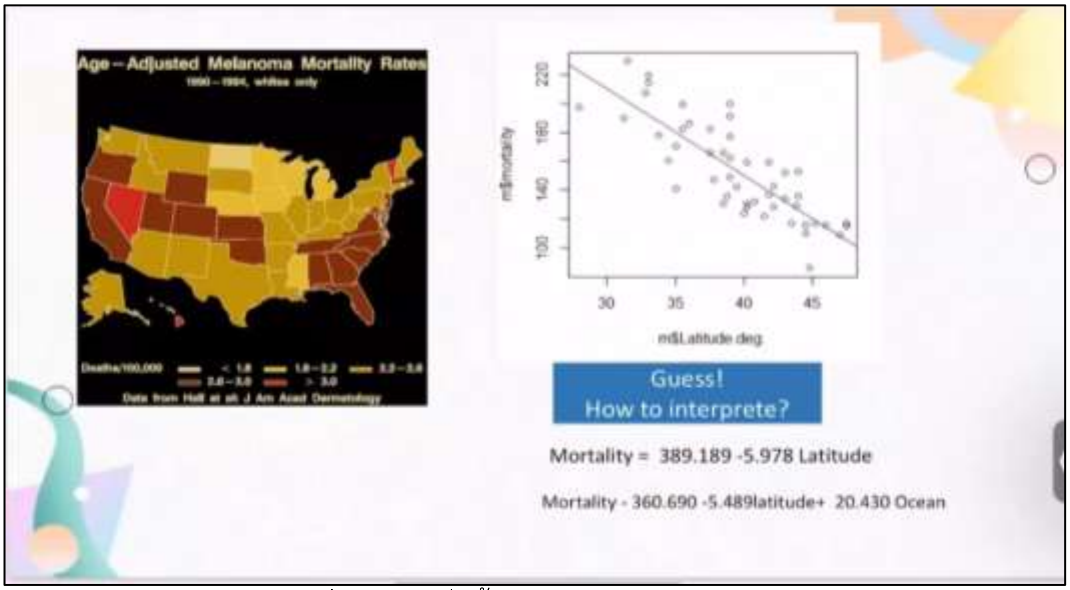
- **ผลลัพธ์ที่ต้องการ** จากการวิเคราะห์ข้อมูลต้องสอดคล้องกับลักษณะของข้อมูลที่มีอยู่ เช่น หากข้อมูลเป็นแบบต่อเนื่อง การใช้สถิติสำหรับข้อมูลกลุ่มอาจไม่เหมาะสม หรือในกรณีที่ข้อมูลเป็นแบบกลุ่ม การใช้สถิติสำหรับข้อมูลต่อเนื่องก็จะนำไปสู่ผลลัพธ์ที่คลาดเคลื่อน
- **วัตถุประสงค์ของการวิจัย** มีบทบาทสำคัญในการเลือกใช้สถิติที่เหมาะสม หากวัตถุประสงค์เป็นการศึกษาหาความสัมพันธ์หรือเปรียบเทียบระหว่างกลุ่ม การใช้สถิติเชิงอนุมานจะเหมาะสมกว่าสถิติเชิงพรรณนาที่ใช้สำหรับการบรรยายข้อมูลเท่านั้น นอกจากนี้ การใช้สถิติขั้นสูงกับข้อมูลที่มีคุณภาพต่ำอาจไม่สร้างประโยชน์เท่ากับการใช้สถิติพื้นฐานที่เหมาะสมกับข้อมูลมากกว่า
- **การศึกษาทบทวนวรรณกรรมและวิธีการวิจัยที่เกี่ยวข้อง**มีความสำคัญมาก เนื่องจากพื้นที่ศึกษาแต่ละแห่งอาจมีลักษณะเฉพาะที่แตกต่างกัน ส่งผลให้ความเป็นตัวแทนของข้อมูลในแต่ละ

พื้นที่ไม่เหมือนกัน การทบทวนวรรณกรรมช่วยให้เข้าใจบริบทและเลือกใช้สถิติให้เหมาะสมกับข้อมูลในพื้นที่ศึกษาได้ดียิ่งขึ้น

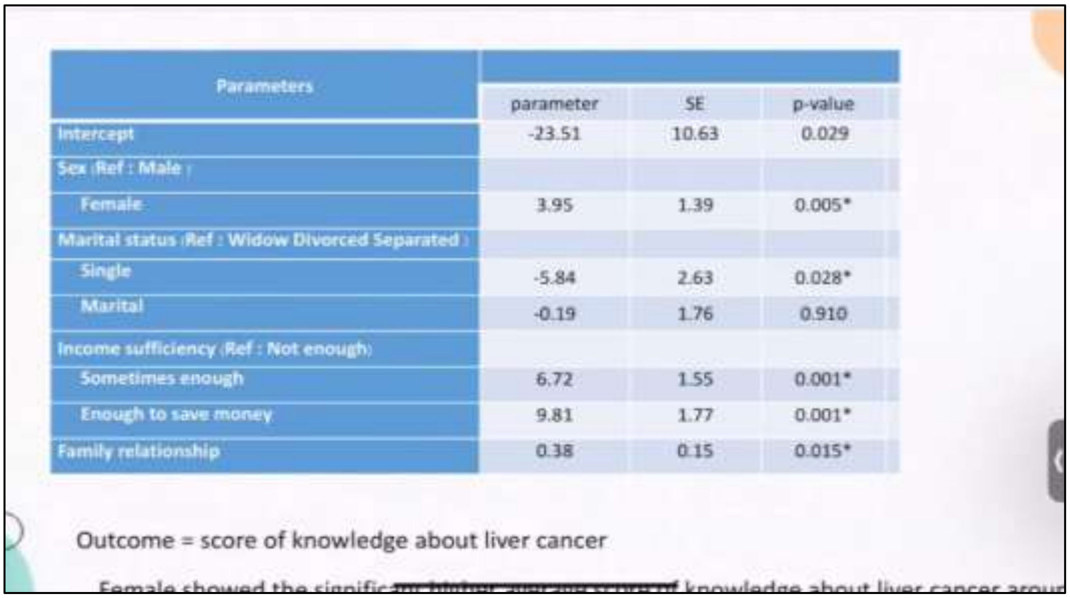
- **Assumption Checking** การตรวจสอบสมมติฐานเป็นสิ่งสำคัญเพราะจะทำให้ผลการวิเคราะห์ข้อมูลมีความน่าเชื่อถือและถูกต้องมากขึ้น ดังนี้
  - การกระจายตัวของข้อมูล (Distribution) ควรตรวจสอบว่าข้อมูลมีการกระจายตัวปกติหรือไม่ เพราะหากไม่เป็นการแจกแจงปกติ อาจต้องใช้สถิตินอนพาราเมตริก
  - ความคงที่ของความแปรปรวน (Homogeneity of Variance) ควรตรวจสอบว่าความแปรปรวนของแต่ละกลุ่มข้อมูลมีความคงที่หรือไม่ เพราะหากไม่คงที่ต้องปรับวิธีคำนวณค่าสถิติบางตัว
  - ความเป็นอิสระของข้อมูล (Independence) ข้อมูลแต่ละค่าควรเป็นอิสระต่อกัน ไม่มีความสัมพันธ์กัน มิเช่นนั้นต้องปรับวิธีการวิเคราะห์
  - ระดับการวัดของข้อมูล (Level of Measurement) สถิติบางตัวต้องการระดับการวัดของตัวแปรในลักษณะเฉพาะ เช่น อันดับขั้น หรืออัตราส่วน เป็นต้น
  - เนื่องจากสถิติแต่ละแบบมีเงื่อนไขการใช้งานที่แตกต่างกัน ดังนั้น ควรตรวจสอบสถิติพื้นฐานให้เรียบร้อยก่อนเลือกใช้สถิติที่เหมาะสม เช่น สถิติเชิงพรรณา ต่าง ๆ
- **P-value**
  - p-value หรือ ค่าความน่าจะเป็น แสดงถึงระดับนัยสำคัญในการทดสอบสมมติฐาน โดยเป็นค่าที่บอกโอกาสที่จะได้ค่าสถิติที่มากกว่าหรือเท่ากับค่าสถิติที่คำนวณได้จากข้อมูลตัวอย่าง ภายใต้สมมติฐานหลัก (H_0)
  - ในกรณีที่ข้อมูลมีการแจกแจงแบบปกติ p-value แปลความหมายได้ดังนี้
    - ถ้า p-value มีค่ามาก (เช่น 0.5) หมายความว่ามีโอกาสสูงมากที่จะสุ่มได้ค่าสถิติในข้อมูลชุดนี้ แม้ว่าสมมติฐานหลัก (H_0) จะเป็นจริง ดังนั้นจึงไม่มีเหตุผลเพียงพอที่จะปฏิเสธ H_0
    - ถ้า p-value มีค่าต่ำมาก (เช่น 0.01 หรือ 0.001) หมายความว่ามีโอกาสต่ำมากที่จะสุ่มได้ค่าสถิตินี้จากข้อมูล หากสมมติฐานหลักเป็นจริง ดังนั้นจึงมีหลักฐานเพียงพอที่จะปฏิเสธ H_0 และยอมรับสมมติฐานรอง (H_1)
  - โดยทั่วไป จะกำหนดค่า p-value ที่ระดับนัยสำคัญที่ 0.05 หรือ 0.01 เป็นเกณฑ์ในการตัดสินว่าจะปฏิเสธหรือไม่ปฏิเสธสมมติฐานหลัก แต่อาจมีการกำหนดเกณฑ์อื่นขึ้นอยู่กับลักษณะของการวิจัยหรืองานที่ทำ
- สรุปคือ p-value เป็นตัวบ่งชี้หลักฐานในการสนับสนุนหรือคัดค้านสมมติฐาน โดยมีค่าต่ำแสดงถึงโอกาสในการปฏิเสธสมมติฐานหลักสูง

- **ตัวอย่าง**

เช่น กรณีข้อมูลมีการแจกแจงปกติ โอกาสที่จะปฏิเสธ H_0 มีมากน้อยแค่ไหน หรือโอกาสที่จะสุ่มไปโดนข้อมูลฝั่งใดฝั่งหนึ่งมีมากน้อยแค่ไหน



- จากกราฟเส้น แสดงให้เห็นว่ายิ่งละจิจูดเพิ่มขึ้น 1 หน่วย อัตราการตายจะลดลง 5.978 คน/100,000 ประชากร



Parameters	parameter	SE	p-value
Intercept	-23.51	10.63	0.029
Sex (Ref : Male)			
Female	3.95	1.39	0.005*
Marital status (Ref : Widow Divorced Separated)			
Single	-5.84	2.63	0.028*
Marital	-0.19	1.76	0.910
Income sufficiency (Ref : Not enough)			
Sometimes enough	6.72	1.55	0.001*
Enough to save money	9.81	1.77	0.001*
Family relationship	0.38	0.15	0.015*

- จากตาราง แสดงค่าความสัมพันธ์ระหว่างตัวแปร คะแนนความรู้เกี่ยวกับมะเร็งตับกับตัวแปรเพศ พบว่ามีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติ โดยมีค่าพารามิเตอร์ที่ 3.95

ตารางที่ 4.23 เปรียบเทียบความสัมพันธ์ระหว่างปฏิสัมพันธ์ระหว่างบุคคล กับพฤติกรรมการกินอาหารเสี่ยง

ปฏิสัมพันธ์ระหว่างบุคคล	n	พฤติกรรมเสี่ยงสูง		Sig.
		จำนวน	ร้อยละ	
อิทธิพลจากครอบครัว				
สูง	156	124	64.2	
ปานกลาง	69	60	31.1	0.003
ต่ำ	18	9	4.7	
อิทธิพลจากเพื่อน				
สูง	57	50	26.0	
ปานกลาง	94	72	37.5	0.202
ต่ำ	91	70	36.5	

- จากตารางแสดงลักษณะข้อมูลที่แบ่งกลุ่ม เช่น ปฏิสัมพันธ์ระหว่างบุคคล กับพฤติกรรมการกินอาหารเสี่ยง (ซึ่งมีการจัดกลุ่มว่าเสี่ยง กับไม่เสี่ยง) ไม่จำเป็นต้องแสดงค่าของทั้ง 2 กลุ่ม แต่สามารถแสดงค่าใดค่าหนึ่งได้ ซึ่งในตารางแสดงจำนวน และร้อยละของพฤติกรรมการกินอาการที่เสี่ยง
- นอกจากนี้ การแปลผลควรพิจารณาจาก % ด้วย เพราะสามารถอธิบายความหมายของความสัมพันธ์ที่ตอบวัตถุประสงค์ได้ด้วย

Table 4.4.1 Statistics of the Logistic Regression model for the level of low compliance

	OR	LB	UB	p-value
Gender (Ref.=Female)	1.585	0.200	1.986	0.431
Age	3.716	0.764	18.081	0.104
Marital status(Ref.= Married)	1.147	0.435	1.748	0.699
Education level (Ref.= high school and lower)	1.166	0.406	1.816	0.690

Male showed the higher risk of low compliance 1.6 times higher than female bu
risks did not present the statistically significant (p-value=0.431).

Increasing one year of age, the risk of low compliance increased 3.716 times.

- จากตัวอย่างตารางแสดงผลจาก Logistic Regression Model เพื่อแสดงความสัมพันธ์ระหว่างปัจจัยส่วนบุคคล กับการมี Low compliance โดยพบว่า ผู้ชายมีโอกาสที่จะมี Low compliance มากกว่า เพศหญิง 1.6 เท่า อย่างไม่มีนัยสำคัญทางสถิติ (P-value 0.431) อายุเพิ่มขึ้น 1 ปี มีโอกาสที่จะมี Low compliance มากกว่า 3.716 เท่า อย่างไม่มีนัยสำคัญทางสถิติ (P-value 0.104)

- **ภาคผนวก: สรุปผลการประเมินความพึงพอใจ ต่อการประชุมเชิงปฏิบัติการพัฒนาแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพ ในวันที่ 18 – 20 ธันวาคม 2566**

จากการประเมินความพึงพอใจต่อการจัดประชุมเชิงปฏิบัติการพัฒนาแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพของผู้สูงอายุ ในวันที่ 18 – 20 ธันวาคม 2566 จากการตอบแบบประเมินออนไลน์ โดยผู้เข้าร่วมการอบรมทั้งแบบ onsite ณ ห้องประชุม 1 กองประเมินผลกระทบต่อสุขภาพ กรมอนามัย และ online ผ่านระบบสื่ออิเล็กทรอนิกส์ สรุปผลการตอบแบบประเมินความพึงพอใจได้ ดังนี้

ส่วนที่ 1 ความพึงพอใจต่อการจัดประชุม

1.1 ระดับความพึงพอใจต่อเนื้อหาการประชุมเชิงปฏิบัติการ

พบว่า ผู้เข้าร่วมประชุมมีความพึงพอใจต่อเนื้อหาการประชุมเชิงปฏิบัติการ ในระดับมากที่สุดในทุกหัวข้อของการอบรม โดยมีหัวข้อการใช้เครื่องมือวิเคราะห์ข้อมูลอย่างง่ายด้วย Excel ที่มีระดับพึงพอใจมากที่สุด ร้อยละ 84.62 รายละเอียดดังตารางที่ 1.1

ตารางที่ 1.1 แสดงความพึงพอใจต่อเนื้อหาการประชุมเชิงปฏิบัติการ

หัวข้อ	จำนวน (ร้อยละ)					ระดับความพึงพอใจ
	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด	
1) การขับเคลื่อนงานสาธารณสุขด้วยข้อมูล (Data-Driven Public Health) ความรู้เบื้องต้นเกี่ยวกับการวิเคราะห์ข้อมูลเชิงพยากรณ์ (Predictive Analytics) (นายแพทย์ศุภณัฐ วงศานุพัทธ์)	-	-	7.69	30.77	61.54	มากที่สุด
2) การใช้เครื่องมือวิเคราะห์ข้อมูลอย่างง่ายด้วย Excel (นายฤทธิไกร นามเกษ)	-	-	7.69	7.69	**84.62**	มากที่สุด
3) การใช้เครื่องมือทางสถิติเพื่อการพยากรณ์สุขภาพ (STATA) (อาจารย์จิราลักษณ์ นนทารักษ์)	-	-	7.69	38.46	53.85	มากที่สุด
4) เทคนิคการแปลผลและเขียนผลทางสถิติ (อาจารย์ณัฐนารี เอมยงค์)	-	7.69	7.69	30.77	53.85	มากที่สุด

1.2 ความรู้/ความเข้าใจในการประชุมเชิงปฏิบัติการ

เมื่อพิจารณาความพึงพอใจต่อความรู้/ความเข้าใจในการประชุมเชิงปฏิบัติการแบบรายข้อพบว่า ผู้ร่วมประชุมมีความพึงพอใจในระดับมากที่สุด ในด้านการเปิดโอกาสให้ผู้เข้าร่วมประชุมแสดงความคิดเห็น ร้อยละ 76.92 และรองลงมาคือ ได้รับความรู้ความเข้าใจเพิ่มขึ้นหลังการประชุมในระดับมากที่สุด และความรู้ที่ได้จากการประชุมนี้เป็นประโยชน์ต่อการทำงาน ในระดับมากที่สุด เท่ากันที่ร้อยละ 53.85 รายละเอียดดังตารางที่ 1.2

ตารางที่ 1.2 แสดงความพึงพอใจต่อความรู้/ความเข้าใจในการประชุมเชิงปฏิบัติการ

หัวข้อ	จำนวน (ร้อยละ)					ระดับความพึงพอใจ
	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด	
1) ความรู้ ความเข้าใจ ก่อนการประชุม	7.69	-	30.77	30.77	30.77	มากที่สุด
2) ท่านได้รับความรู้ความเข้าใจเพิ่มขึ้นเพียงใดหลังการประชุม	-	-	-	46.15	53.85	มากที่สุด
3) เปิดโอกาสให้ผู้เข้าร่วมประชุมแสดงความคิดเห็น	-	-	-	23.08	**76.92**	มากที่สุด
4) ความรู้ที่ได้จากการประชุมนี้เป็นประโยชน์ต่อการทำงานของท่านในระดับใด	-	-	-	46.15	53.85	มากที่สุด

1.3 ด้านสถานที่/ระยะเวลา/รูปแบบการจัดประชุมเชิงปฏิบัติการ

เมื่อพิจารณาด้านสถานที่/ระยะเวลา/รูปแบบการจัดประชุมเชิงปฏิบัติการแบบรายข้อ พบว่า ผู้เข้าร่วมประชุมส่วนใหญ่มีความพึงพอใจในระดับมากที่สุด โดยเฉพาะด้านการติดต่อ/ประสานงาน/อำนวยความสะดวกและด้านเอกสารประกอบการประชุม มากที่สุดที่ร้อยละ 61.54 ยกเว้นในด้านรูปแบบการประชุมที่ส่วนใหญ่ (ร้อยละ 61.54) พึงพอใจอยู่ในระดับมาก รายละเอียดดังตารางที่ 1.3

ตารางที่ 1.3 แสดงความพึงพอใจด้านสถานที่/ระยะเวลา/รูปแบบการจัดประชุมเชิงปฏิบัติการ

หัวข้อ	จำนวน (ร้อยละ)					ระดับความพึงพอใจ
	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด	
1) รูปแบบการประชุม	-	-	7.69	61.54	30.77	มาก
2) การติดต่อ / ประสานงาน / อำนวยความสะดวก	-	-	-	38.46	61.54	มากที่สุด
3) ระยะเวลาการจัดการประชุม	-	-	7.69	38.46	53.85	มากที่สุด
4) การประชาสัมพันธ์การประชุม	-	-	7.69	46.15	46.15	มากที่สุด
5) เอกสารประกอบการประชุม	-	-	-	38.46	61.54	มากที่สุด

<u>ส่วนที่ 2 ข้อเสนอแนะ</u>

ผู้เข้าร่วมการประชุม ฯ ได้ให้ข้อเสนอแนะ ตามคำถามปลายเปิด สรุปได้ดังนี้

2.1 ท่านมีข้อเสนอแนะให้กองพยากรณ์สุขภาพ จัดกิจกรรมพัฒนาศักยภาพในประเด็นใดในครั้งต่อไป

- งานวิจัยของผู้สูงอายุจะเน้นเชิงคุณภาพมากๆ ตัวแปรเชิงปริมาณจะน้อยกว่า โดยมีโรค NCD มากกว่าโรคติดต่อ วิทยากรถ่ายทอดดีแต่อาจไม่ตรงเทรนด์โลก น่าจะให้อาจารย์ที่ทำวิจัยเชิงคุณภาพมากกว่ามาบรรยาย
- อยากให้มีการจัดกิจกรรมนอกสถานที่ และอยากให้มีการ workshop แบบแบ่งกลุ่ม เพื่อให้เกิดการทำงานเป็นทีม มีการช่วยเหลือกันภายในกลุ่ม มีประเด็นการแลกเปลี่ยนกันในกลุ่มที่หลากหลาย
- อยากให้จัดสอนประเด็นการใช้สถิติ logistic regression และ linear regression พร้อม workshop โดยเฉพาะ
- การทำ dashboard
- อยากให้ผู้จัด จัดอีกค่ะ ดีมาก แต่อยากให้ใช้ชื่อหัวข้อว่า สถิติเพื่อการทำวิจัยเชิงปริมาณ
- การอบรมนี้ดีเอาไปใช้ในอนาคตได้ ควรเน้นกิจกรรมกลุ่มมากกว่า การคำนวนทางสถิติ

2.2 ท่านสนใจที่จะนำแนวคิดหลักการ และเครื่องมือเทคโนโลยีการพยากรณ์ไปใช้ในการดำเนินงานของท่านในเรื่องใด (เพื่อเป็นข้อมูลประกอบในการวางแผนกำหนดประเด็นเพื่อการพยากรณ์สุขภาพ)

- งานสิ่งแวดล้อม ที่มีความสัมพันธ์กับผู้สูงอายุในหลายแง่มุม เช่น ปัจจัยที่มีผลกับการล้ม ฝุ่นควัน และภาวะสมองเสื่อม
- ความเสี่ยงต่อการหกล้มในผู้สูงอายุที่นอนโรงพยาบาล
- การทำวิจัย และการนำเสนอข้อมูล

ส่วนที่ 3 ภาพกิจกรรมการประชุมเชิงปฏิบัติการพัฒนาแบบจำลองทางสถิติเพื่อการพยากรณ์สุขภาพ

นายแพทย์ศุภณัฐ วงศานุพัทธ์ ฤทธิไกร นามเกษ

กองระบาดวิทยา กรมควบคุมโรค

ดร.จิราลักษณ์ นนทารักษ์ ผศ. ดร. ณัฐนารี เอมยงค์

ภาควิชาระบาดวิทยา คณะสาธารณสุขศาสตร์ มหาวิทยาลัยมหิดล

ดร.เบญจวรรณ ธวัชสุภา ผู้อำนวยการกองพยากรณ์สุขภาพ กรมอนามัย