

# การจัดกลุ่มจังหวัดในประเทศไทยด้วยปัจจัยประชากรศาสตร์และระบาดวิทยา ของโควิด-19

Clustering Provinces in Thailand Based on Demographic and COVID-19  
Epidemiological Factors

วัฒนา ขยธวัช

Vadhana Jayathavaj

สาขาการแพทย์แผนไทย คณะสหเวชศาสตร์ มหาวิทยาลัยพทุมธานี

Thai Traditional Medicine, Faculty of Allied Health Sciences, Pathumthani University

E-mail : vadhana.j@ptu.ac.th

(Received: 22 January 2023, Revised: 3 March 2023, Accepted: 8 March 2023)

<https://doi.org/10.57260/stc.2023.450>

## บทคัดย่อ

โรคติดเชื้อไวรัสโคโรนา 2019 (โควิด - 19) ได้มีการระบาดอย่างต่อเนื่องตั้งแต่ต้นวาคม 2562 การจัดกลุ่มพื้นที่เป็นมาตรการสำคัญในการบริหารจัดการโรคติดเชื้อนี้ การศึกษานี้จึงมีวัตถุประสงค์เพื่อทำการวิเคราะห์กลุ่มจังหวัดจากข้อมูลความหนาแน่นของประชากร ขนาดครัวเรือน รายได้ครัวเรือนเฉลี่ยต่อเดือน อัตราการติดเชื้อ และอัตราการเสียชีวิต จากการวิเคราะห์กลุ่มโดยใช้ K-mean clustering พบว่า สามารถจัดกลุ่มจังหวัดได้เป็น 4 กลุ่ม จำแนกตามอัตราการติดเชื้อและอัตราการเสียชีวิต กลุ่ม 1 ปานกลาง จำนวน 23 จังหวัด กลุ่ม 2 และ 4 ต่ำ จำนวน 28 และ 21 จังหวัด ตามลำดับ และกลุ่ม 3 สูง จำนวน 5 จังหวัด เมื่อทำการแจกแจงจังหวัดตามการวิเคราะห์กลุ่มและการจัดจังหวัดตามพื้นที่สถานการณ์ย่อยโรคติดเชื้อไวรัสโคโรนา 2019 (24 มกราคม 2565) ของศูนย์บริหารสถานการณ์ โควิด-19 ที่จัดเป็นพื้นที่พื้นที่ควบคุม (สีส้ม) 44 จังหวัด พื้นที่เฝ้าระวังสูง (สีเหลือง) 25 จังหวัด และ พื้นที่นาร่องทองเที่ยว (สีฟ้า) 8 จังหวัด ทำให้สามารถจำแนกระดับของจังหวัดพื้นที่ในแต่ละสีว่ายังมีอัตราการติดเชื้อและอัตราการตายอีกสี่ละ 3 ระดับ (สูง กลุ่ม 3 กลาง กลุ่ม 1 และ ต่ำ กลุ่ม 2 และ 4) ด้วย ซึ่งสามารถใช้เป็นข้อมูลในการป้องกันตนเองจากการติดเชื้อโควิด-19 ได้

**คำสำคัญ:** การวิเคราะห์กลุ่ม โรคติดเชื้อไวรัสโคโรนา 2019 (โควิด - 19) อัตราการติดเชื้อ

## Abstract

The Coronavirus Disease 2019 (COVID-19) outbreak has continued since December 2019. Clustering of areas is an important measure in managing this infectious disease. The purpose of this study was to analyze the provinces from population density data, household size, average monthly household income, infection rate, and death rate. From cluster analysis using K-mean clustering, it was found that provinces could be grouped into 4 groups classified by infection rates and mortality rates: group 1, medium (23 provinces), group 2 and 4, low (28 and 21 provinces, respectively), and group 3, high (5 provinces). When classifying provinces according to cluster analysis and provincial classification according to the sub-situation area of the Coronavirus Disease 2019 (January 24, 2022) of the Center for COVID-19 Situation Administration, classified as controlled areas (orange), 44 provinces, monitoring areas cautiously high (yellow), 25 provinces and pilot area tourism (blue), 8 provinces, allowing to classify the level of provinces in each color that there are infection rates and mortality rates in 3 levels (high group 3, medium group 1, and low group 2 and 4) in each color as well, these three levels can be used as supporting information to protect oneself from being infected with COVID-19.

**Keywords:** Cluster analysis, Coronavirus disease 2019 (COVID-19), Prevalence rate

## บทนำ

โรคติดเชื้อไวรัสโคโรนา 2019 (โควิด - 19) มีการระบาดและกลายพันธุ์อย่างต่อเนื่องตั้งแต่ธันวาคม 2562 จนถึงระลอกที่ 6 ในช่วงปลายเดือนกรกฎาคม 2565 (สำนักข่าว Hfocus เจาะลึกระบบสุขภาพ, 2565) การเปิดประเทศรับนักท่องเที่ยวในต้นปี 2566 ตามมาตรการผ่อนคลายเป็นที่ดึงดูดผู้เดินทางจากทั่วโลก เพื่อพลิกฟื้นเศรษฐกิจของประเทศ ภายใต้คณะกรรมการบริหารจัดการป้องกันควบคุมโรคโควิด 19 รองรับการเดินทางเข้าประเทศแทนศูนย์บริหารสถานการณ์โควิด 19 (ศบค.) (สำนักสารนิเทศ สำนักงานปลัดกระทรวงสาธารณสุข, 2566) ดังนั้น การมีข้อมูลความรุนแรงของการระบาดเชิงพื้นที่เพื่อการป้องกันตนเองยังมีความจำเป็นสำหรับประชาชน

การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคการแบ่งคน สัตว์ สิ่งของ องค์กร และอื่น ๆ (objects) ออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม โดยให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่นำมาจำแนกเหมือนกันหรือคล้ายกัน แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่นำมาจำแนกต่างกัน (Tan et al., 2018; Everitt et al., 2011) การจำแนกเขตของรัฐบาลราชอาณาจักรอินเดียที่ได้รับผลกระทบจากโควิด-19 ออกเป็นกลุ่มตามความคล้ายคลึงกันภายในกลุ่มและความแตกต่างระหว่างกลุ่ม เพื่อการดำเนินนโยบายของรัฐบาลในการตัดสินใจเกี่ยวกับสิ่งอำนวยความสะดวกทางการแพทย์ (เครื่องช่วยหายใจ ชุดทดสอบ หน้ากาก การรักษา ฯลฯ) สามารถลดจำนวนผู้ติดเชื้อและผู้เสียชีวิต (Kumar, 2020) การจัดกลุ่ม 28 รัฐและ 9 สหภาพดินแดนของประเทศอินเดียออกเป็น 3 กลุ่ม จากจำนวนทั้งหมดของผู้ป่วยยืนยัน จำนวนผู้ป่วยรักษาหาย/ออกจากโรงพยาบาล และจำนวนผู้เสียชีวิตทั้งหมด โดยใช้ซอฟต์แวร์ PAST (version 3) (Zargar, et al., 2021) การจัดอันดับและให้คะแนน 180 ประเทศจากจำนวนผู้ติดเชื้อและการเสียชีวิตจาก COVID-19 ในปี พ.ศ. 2563 เพื่อแสดงความพร้อมของระบบและประสิทธิภาพในการต่อสู้กับโรคระบาดแต่ละกลุ่มประเทศ (Sadeghi et al., 2021) การวิเคราะห์กลุ่มเขตท้องที่ต่าง ๆ จากจำนวนผู้ป่วยยืนยัน จำนวนประชากรความหนาแน่นและจำนวนโรงพยาบาลที่ใช้รักษาโควิด-19 ใช้ในการจัดอันดับตามภาระของระบบการรักษาพยาบาล (Sengupta et al., 2021) การวิเคราะห์การจัดกลุ่มอนุกรมเวลาของข้อมูลโควิด-19 ที่ทำให้เป็นมาตรฐานทั่วทุกรัฐและเขตในอินเดียโดยใช้ตัวบ่งชี้ เช่น ผู้ป่วยที่ยังมีชีวิต ผู้ป่วยรายใหม่รายวัน และผู้เสียชีวิตในระดับระดับอำเภอ และระดับประเทศ ทำให้ค้นพบคลัสเตอร์จำแนกตามผู้ป่วยรายใหม่รายวันต่อประชากร ในช่วงระลอกแรกในอินเดีย (ระหว่างเดือนมีนาคม 2020 ถึงเดือนมกราคม 2021) มีรูปแบบที่แตกต่างกันสี่รูปแบบ: การเพิ่มขึ้นอย่างรวดเร็วและการลดลง การเพิ่มขึ้นอย่างต่อเนื่องและการลดลง เพิ่มขึ้นอย่างต่อเนื่องและเพิ่มขึ้นลดลงเป็นยอดเขาหลายยอดและพบว่ารูปแบบรูปร่างเหล่านี้มีความสัมพันธ์กับลักษณะทางภูมิศาสตร์สามารถนำไปใช้ในการจัดการทรัพยากรเพื่อควบคุมการแพร่กระจายของโรคได้อย่างเหมาะสม (Raj et al., 2022)

ประเทศไทยจัดอยู่ในกลุ่มประเทศที่มีความเสี่ยงต่ำโดยได้คะแนนร้อยละ 70 จากตัวชี้วัด 4 ด้าน ได้แก่ จำนวนผู้ติดเชื้อที่ยังดูแลอยู่ (Active case) ต่อประชากร ผู้ที่รักษาหายแล้วต่อผู้ติดเชื้อ จำนวนการตรวจต่อจำนวนผู้ติดเชื้อ และจำนวนการตรวจต่อประชากร ส่วนคะแนนอีกร้อยละ 30 เป็นคะแนนคงที่ที่ได้มาจากดัชนี

ความมั่นคงด้านสุขภาพโลก (The Global Health Security Index: GHS) ที่ประเมินโดยมหาวิทยาลัย Johns Hopkins University ประเทศสหรัฐอเมริกา (กรมควบคุมโรค, 2563) การจัดกลุ่มประเทศและดินแดนตามความเสี่ยงสถานการณ์โรคโควิด-19 เกณฑ์ที่ใช้ในการจัดกลุ่ม ได้แก่ IHR-JEE, GCI และผู้ป่วยรายใหม่เฉลี่ย (จากข้อมูลย้อนหลัง 14 วัน) และปรับปรุงข้อมูล ทุกวันที่ 15 และ 30 ทุกเดือน จัดเป็น 3 กลุ่ม ได้แก่ ประเทศและดินแดนที่มีความเสี่ยงสูง เสี่ยงปานกลาง และเสี่ยงต่ำ (กรมควบคุมโรค, 2564) และ กระทรวงสาธารณสุข ประกาศยกเลิกโควิด-19 จากการเป็นโรคติดต่ออันตราย และกำหนดให้เป็นโรคติดต่อที่ต้องเฝ้าระวังตั้งแต่ 1 ตุลาคม 2565 เป็นต้นไป (รัฐบาลไทย, 2565) สถานการณ์ผู้ป่วยโควิด-19 ภายในประเทศ รายสัปดาห์ ปรับปรุงล่าสุด 1 มกราคม 2566 สัปดาห์ที่ 52 (25-31 ธันวาคม 2565) ผู้ป่วยรายใหม่รักษาตัวในโรงพยาบาล 2,111 คน (เฉลี่ย 302 คนต่อวัน) เสียชีวิตเฉลี่ย 10 คน/วัน เมื่อสถานการณ์โควิด-19 ยังดำเนินต่อไป (โครงการ SAT COVID Dashbord, 2566) จึงจำเป็นต้องมีการจัดกลุ่มจังหวัดตามระดับความรุนแรงของการระบาดและความหนาแน่นของประชากรเพิ่มขึ้นด้วย การจัดกลุ่มจังหวัดตามปัจจัยตัวชี้วัดทางประชากรศาสตร์ และข้อมูลการระบาดของโควิด-19 จะเป็นข้อมูลเพิ่มเติมในการระมัดระวังและป้องกันตนเองเพิ่มขึ้น เมื่อต้องเข้าไปในพื้นที่ที่มีการจัดกลุ่มจังหวัดของศูนย์บริหารสถานการณ์โควิด-19 ที่ได้จำแนกออกเป็น พื้นที่ควบคุม (สีส้ม) 44 จังหวัด พื้นที่เฝ้าระวังสูง (สีเหลือง) 25 จังหวัด และ พื้นที่นำร่องท่องเที่ยว (สีฟ้า) 8 จังหวัด (ศูนย์บริหารสถานการณ์ โควิด-19, 2565)

## วัตถุประสงค์

การศึกษานี้มีวัตถุประสงค์เพื่อ ทำการจัดแบ่งกลุ่มจังหวัดตามปัจจัยตัวชี้วัดทางประชากรศาสตร์และข้อมูลการระบาดของโควิด-19

## ระเบียบวิธีวิจัย

ข้อมูลที่ใช้ในการวิเคราะห์กลุ่ม เป็นข้อมูลรายจังหวัด 77 จังหวัด มีแหล่งที่มาดังนี้

1. ประชากรต่อตารางกิโลเมตร จำนวนบ้านต่อตารางกิโลเมตร จำนวนประชากรจากการทะเบียนชาย หญิง เนื้อที่ ความหนาแน่น และบ้าน จำแนกตามภาค และจังหวัด พ.ศ. 2555-2564 จากสำนักงานสถิติแห่งชาติ (2565ก) และ รายได้ของครัวเรือนเฉลี่ยต่อปี จาก 8 สถิติรายได้และรายจ่ายของครัวเรือน รายได้เฉลี่ยต่อเดือนของครัวเรือน เป็นรายภาค และจังหวัด พ.ศ. 2547 – 2564 (สำนักงานสถิติแห่งชาติ, 2565ข) จำนวนประชากรเดือนมิถุนายน พ.ศ. 2565 ได้จากสถิติประชากรรายจังหวัด ณ เดือน มิถุนายน 2565 จากสำนักบริหารการทะเบียน กรมการปกครอง (2565)
2. จำนวนผู้ติดเชื้อและจำนวนผู้เสียชีวิตรายจังหวัด ณ วันที่ 30 ธันวาคม 2565 (กรมควบคุมโรค, 2566)

## การคำนวณ

อัตราการติดเชื้อโควิด-19 ต่อประชากร 1,000 คน คำนวณจากจำนวนผู้ติดเชื้อ (กรมควบคุมโรค, 2566) หารด้วยจำนวนประชากรเดือนมิถุนายน พ.ศ. 2565 (สำนักบริหารการทะเบียน กรมการปกครอง, 2565)

อัตราการตายจากโควิด-19 ต่อประชากร 10,000 คน คำนวณจากจำนวนผู้เสียชีวิต (กรมควบคุมโรค, 2566) หารด้วยจำนวนประชากรเดือนมิถุนายน พ.ศ. 2565 (สำนักบริหารการทะเบียน กรมการปกครอง, 2565)

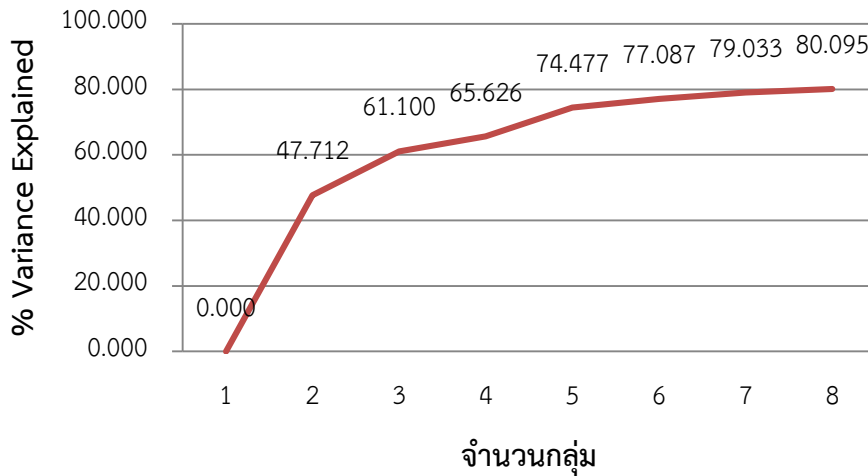
ข้อมูลปัจจัยที่นำมาพิจารณาทำการจัดกลุ่มรายจังหวัด ดังแสดงในตารางที่ 2, 3, 4, และ 4

## การทำให้เป็นค่ามาตรฐาน

เนื่องจากตัวแปรมีขนาดและช่วงแตกต่างกันมาก การมีตัวแปรบางตัวที่มีช่วง/สเกลต่างกันมากมักจะสร้างปัญหา กล่าวคือ "ผลลัพธ์" อาจมีอิทธิพลของค่าขนาดใหญ่ เพื่อหลีกเลี่ยงปัญหาดังกล่าว จึงได้ทำให้ข้อมูลเป็นมาตรฐานมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ก็จะสามารถลดอิทธิพลจากขนาดลงได้ (Evgeniou, n.d.)

## การวิเคราะห์ข้อมูล

ใช้โปรแกรม Past (version 4.09) เป็นซอฟต์แวร์ฟรีสำหรับการวิเคราะห์ข้อมูลทางวิทยาศาสตร์ พร้อมด้วยฟังก์ชันสำหรับการจัดการข้อมูล การลงจุด สถิติหลายตัวแปรและหลายตัวแปร การวิเคราะห์ระบบนิเวศน์ อนุกรมเวลาและการวิเคราะห์เชิงพื้นที่ สันฐานวิทยาและชั้นหิน (Hammer, 2022) โดยใช้ตัวเลือกการวิเคราะห์กลุ่ม (Cluster Analysis) k-means clustering ที่จัดจังหวัดให้มาเกาะกลุ่มกันตามระยะห่างทางเรขาคณิตของข้อมูลที่น่ามาวิเคราะห์ และการจัดกลุ่มแบบลำดับชั้น (The hierarchical clustering) สร้างแผนภาพต้นไม้ (Dendrogram) ที่แสดงวิธีจัดกลุ่มจุดข้อมูล (แถว) ค่าเฉลี่ยของกลุ่มคู่ที่ไม่ได้ถ่วงน้ำหนัก (Unweighted pair-group average - UPGMA) กลุ่มจะเข้าร่วมตามระยะห่างเฉลี่ยระหว่างสมาชิกทั้งหมดในสองกลุ่ม ซึ่งเป็นโปรแกรมเดียวกันกับที่ Zargar, et al. (2021) ใช้การจัดกลุ่ม 28 รัฐและ 9 สหภาพดินแดนของประเทศอินเดียออกเป็น 3 กลุ่ม จากจำนวนทั้งหมดของผู้ป่วยยืนยัน จำนวนผู้ป่วยรักษาหาย/ออกจากโรงพยาบาล และจำนวนผู้เสียชีวิตทั้งหมด แต่เป็นรุ่นที่ใหม่กว่า



ภาพที่ 1 แสดงจำนวนกลุ่มที่บ่งชี้ของความแปรปรวนที่อธิบายได้

## ผลการวิจัย

ปัจจัยที่ใช้ในการวิเคราะห์จัดกลุ่มประกอบไปด้วยข้อมูลรายจังหวัด คือ ขนาดครัวเรือน (จำนวนคนต่อครัวเรือน) ความหนาแน่นของประชากร (คนต่อตารางกิโลเมตร) ความหนาแน่นของครัวเรือน (ครัวเรือนต่อตารางกิโลเมตร) รายได้ครัวเรือนเฉลี่ยต่อเดือน (บาท/เดือน) อัตราการติดเชื้อโควิด-19 ต่อประชากร 1,000 คน และ อัตราการเสียชีวิตต่อประชากร 10,000 คน ดังแสดงในตารางที่ 2, 3, 4, และ 5 นำมาทำให้เป็นค่ามาตรฐาน (Normalized) ใช้วิธีการ K-means clustering ที่จัดจังหวัดให้มาเกาะกลุ่มกันตามระยะห่างทางเรขาคณิตของข้อมูลที่น่ามาวิเคราะห์ เลือกจำนวนกลุ่มด้วย The elbow method จากการพิจารณาค่า % Variance explained ที่เพิ่มขึ้น เมื่อจำนวนกลุ่มเพิ่มขึ้นแล้ว จึงได้เลือก 4 กลุ่ม เนื่องจากมีการเพิ่มขึ้นของค่า % Variance explained เริ่มลดลงเมื่อจำนวนกลุ่มเพิ่มขึ้น (The Data Science Lab, 2013) ดังแสดงในภาพที่ 1 สามารถจัดจังหวัดทั้ง 77 จังหวัดออกเป็น 4 กลุ่ม (โปรแกรม PAST แสดงค่า WGSS 156.74 F 1.9092 Var% 65.626) ทำให้ได้กลุ่มจังหวัดดังแสดงในตารางที่ 2, 3, และ 4 และเมื่อนำมาแจกแจงความถี่สรุปเป็นรายเขต/ภาคแล้ว (เขต คือ ตัวเลขหลักแรกของรหัสจังหวัด) ดังแสดงในตารางที่ 1

การนำข้อมูลการปรับระดับของพื้นที่สถานการณ์ย่อยในพื้นที่ทั่วราชอาณาจักรของโรคติดเชื้อไวรัสโคโรนา 2019 (24 มกราคม 2565) จัดกลุ่มจังหวัดออกเป็น 3 กลุ่ม จำแนกออกเป็น พื้นที่ควบคุม (สีส้ม) 44 จังหวัด พื้นที่เฝ้าระวังสูง (สีเหลือง) 25 จังหวัด และ พื้นที่นำร่องท่องเที่ยว (สีฟ้า) 8 จังหวัด (ศูนย์บริหารสถานการณ์ โควิด-19, 2565) เมื่อนำสีมาระบายลงบนรหัสและชื่อจังหวัดที่ได้จากการวิเคราะห์กลุ่มจำแนกตามเขตและกลุ่ม ดังแสดงในตารางที่ 5 และ 6

การแสดงสถิติเชิงพรรณนาของทั้งสี่กลุ่มจากการวิเคราะห์กลุ่ม แสดงในตารางที่ 7

ตารางที่ 1 แสดงการแจกแจงความถี่จำแนกจังหวัดตามเขตและกลุ่ม

เขต	ภาค	กลุ่ม				รวม
		1	2	3	4	
เขต 1	กลาง	6	1	3		10
เขต 2	ตะวันออก	7	1			8
เขต 3	อีสานใต้		2		8	10
เขต 4	อีสานเหนือ		2		8	10
เขต 5	เหนือตอนบน		9			9
เขต 6	เหนือตอนล่าง		7		1	8
เขต 7	ตะวันตก	5	2	1		8
เขต 8	ใต้ตอนบน	5	1	1		7
เขต 9	ใต้ตอนล่าง		3		4	7
	รวม	23	28	5	21	77

ตารางที่ 2 แสดงข้อมูลรายจังหวัดเขต 1 ภาคกลาง และ เขต 2 ภาคตะวันออก

รหัส จังหวัด	จังหวัด	กลุ่ม	คนต่อครัว เรือน	คนต่อ ตร.กม.	ครัวเรือนต่อ ตร.กม.	รายได้ ครัวเรือนเฉลี่ย	อัตรา การติดเชื้อ	อัตราการ เสียชีวิต
10	กรุงเทพมหานคร	3	1.8	3,524	2,006	40,201	179	15
11	สมุทรปราการ	3	1.9	1,351	730	32,914	182	14
12	นนทบุรี	3	1.8	2,071	1,171	41,129	113	4
13	ปทุมธานี	1	1.8	780	433	39,507	74	9
14	พระนครศรีอยุธยา	1	2.4	321	134	31,416	86	7
15	อ่างทอง	1	2.7	284	106	24,510	103	8
16	ลพบุรี	1	2.4	119	49	26,676	48	7
17	สิงห์บุรี	1	2.6	249	97	26,598	60	8
18	ชัยนาท	2	2.5	130	52	23,326	23	1
19	สระบุรี	1	2.2	180	81	33,660	85	9
20	ชลบุรี	1	1.5	363	250	32,356	149	9
21	ระยอง	1	1.4	212	148	28,142	120	7
22	จันทบุรี	1	2.2	85	39	30,878	82	4
23	ตราด	1	2.1	81	39	29,917	85	10
24	ฉะเชิงเทรา	1	2.2	135	60	23,196	112	6
25	ปราจีนบุรี	1	2.2	104	47	27,233	106	6
26	นครนายก	1	2.5	123	49	29,547	96	10
27	สระแก้ว	2	2.5	78	31	20,709	72	3

ตารางที่ 3 แสดงข้อมูลรายจังหวัดเขต 7 ภาคตะวันตก เขต 5 ภาคเหนือตอนบน และ เขต 6 ภาคเหนือตอนล่าง

รหัส จังหวัด	จังหวัด	กลุ่ม	คนต่อครัว เรือน	คนต่อ ตร.กม.	ครัวเรือนต่อ ตร.กม.	รายได้ ครัวเรือน	อัตรา ผู้ติดเชื้อ	อัตรา ผู้เสียชีวิต
70	ราชบุรี	1	2.6	167	64	25,995	101	6
71	กาญจนบุรี	2	2.5	46	18	25,940	59	4
72	สุพรรณบุรี	2	2.7	156	58	17,574	62	5
73	นครปฐม	1	2.2	425	193	38,788	91	9
74	สมุทรสาคร	3	1.9	673	347	27,591	294	18
75	สมุทรสงคราม	1	2.6	458	177	24,964	137	12
76	เพชรบุรี	1	2.2	78	36	26,922	93	6
77	ประจวบคีรีขันธ์	1	2.0	87	43	26,463	82	5
50	เชียงใหม่	2	2.1	89	42	22,963	39	3
51	ลำพูน	2	2.2	89	41	25,377	18	2
52	ลำปาง	2	2.4	58	24	19,858	20	2
53	อุตรดิตถ์	2	2.6	57	22	21,664	31	3
54	แพร่	2	2.4	66	28	20,367	27	2
55	น่าน	2	2.7	41	15	20,556	33	1
56	พะเยา	2	2.4	73	31	16,338	20	2
57	เชียงราย	2	2.3	111	49	15,619	9	2
58	แม่ฮ่องสอน	2	2.4	23	9	15,496	21	3
60	นครสวรรค์	2	2.4	108	44	20,527	42	4
61	อุทัยธานี	2	2.5	48	19	18,428	47	5
62	กำแพงเพชร	2	2.5	83	33	20,941	37	3
63	ตาก	4	3.0	41	14	22,632	58	7
64	สุโขทัย	2	2.6	89	34	22,461	39	4
65	พิษณุโลก	2	2.4	78	33	20,913	37	2
66	พิจิตร	2	2.6	117	44	22,192	22	3
67	เพชรบูรณ์	2	2.7	77	29	26,151	27	2



ตารางที่ 4 แสดงข้อมูลรายจังหวัดเขต 3 อีสานเหนือ เขต 4 อีสานใต้เขต 8 ภาคใต้ตอนบน และ เขต 9 ภาคใต้ตอนล่าง

รหัสจังหวัด	จังหวัด	กลุ่ม	คนต่อครัวเรือน	คนต่อตร.กม.	ครัวเรือนต่อตร.กม.	รายได้ครัวเรือน	อัตราผู้ติดเชื้อ	อัตราผู้เสียชีวิต
30	นครราชสีมา	2	2.6	129	50	24,779	33	3
31	บุรีรัมย์	4	3.3	153	47	23,045	44	1
32	สุรินทร์	4	3.4	169	50	24,009	39	2
33	ศรีสะเกษ	4	3.6	165	46	16,569	35	2
34	อุบลราชธานี	4	3.0	119	40	20,969	35	3
35	ยโสธร	4	3.0	128	43	19,667	38	3
36	ชัยภูมิ	2	2.8	88	32	20,280	30	2
37	อำนาจเจริญ	4	3.1	119	39	20,814	29	2
38	บึงกาฬ	4	3.0	98	33	24,345	32	1
39	หนองบัวลำภู	4	3.3	132	40	23,867	32	2
40	ขอนแก่น	2	2.7	165	60	18,028	48	2
41	อุดรธานี	4	2.9	134	46	22,607	36	2
42	เลย	2	2.8	56	20	26,532	35	3
43	หนองคาย	4	2.8	171	60	21,575	55	2
44	มหาสารคาม	4	3.1	179	58	26,542	43	2
45	ร้อยเอ็ด	4	3.2	156	48	17,714	44	3
46	กาฬสินธุ์	4	3.1	140	46	19,398	36	3
47	สกลนคร	4	2.9	119	41	22,508	26	2
48	นครพนม	4	3.0	130	43	18,525	27	2
49	มุกดาหาร	4	2.9	81	28	20,608	27	2
80	นครศรีธรรมราช	1	2.6	156	59	31,053	84	4
81	กระบี่	1	2.4	102	42	29,841	54	4
82	พังงา	2	2.2	64	29	24,594	68	2
83	ภูเก็ต	3	1.5	771	510	33,328	178	8
84	สุราษฎร์ธานี	1	2.1	83	40	34,402	46	3
85	ระนอง	1	2.1	59	29	21,991	107	7
86	ชุมพร	1	2.1	85	41	30,598	63	4
90	สงขลา	2	2.5	194	76	22,691	74	3
91	สตูล	4	3.0	131	43	23,867	61	6
92	ตรัง	2	2.7	130	48	24,457	45	3
93	พัทลุง	2	2.6	153	58	24,084	74	4
94	ปัตตานี	4	3.7	376	103	20,692	82	8
95	ยะลา	4	3.1	120	38	19,182	108	8
96	นราธิวาส	4	3.6	181	50	17,512	63	5

ตารางที่ 5 แสดงการจำแนกจังหวัดตามเขตและกลุ่ม เขต 1, 2, 3, และ 4

กลุ่ม 1		กลุ่ม 2		กลุ่ม 3		กลุ่ม 4	
รหัส	จังหวัด	รหัส	จังหวัด	รหัส	จังหวัด	รหัส	จังหวัด
13	ปทุมธานี	18	ชัยนาท	10	กรุงเทพมหานคร		
14	พระนครศรีอยุธยา			11	สมุทรปราการ		
15	อ่างทอง			12	นนทบุรี		
16	ลพบุรี						
17	สิงห์บุรี						
19	สระบุรี						
20	ชลบุรี	27	สระแก้ว				
21	ระยอง						
22	จันทบุรี						
23	ตราด						
24	ฉะเชิงเทรา						
25	ปราจีนบุรี						
26	นครนายก						
		30	นครราชสีมา			31	บุรีรัมย์
		36	ชัยภูมิ			32	สุรินทร์
						33	ศรีสะเกษ
						34	อุบลราชธานี
						35	ยโสธร
						37	อำนาจเจริญ
						38	บึงกาฬ
						39	หนองบัวลำภู
		40	ขอนแก่น			41	อุดรธานี
		42	เลย			43	หนองคาย
						44	มหาสารคาม
						45	ร้อยเอ็ด
						46	กาฬสินธุ์
						47	สกลนคร
						48	นครพนม
						49	มุกดาหาร

หมายเหตุ: พื้นที่ควบคุม (สีส้ม) พื้นที่เฝ้าระวังสูง (สีเหลือง) และ พื้นที่นำร่องท่องเที่ยว (สีฟ้า) (ศูนย์บริหารสถานการณ์ โควิด-19, 2565)

ตารางที่ 6 แสดงการจำแนกจังหวัดตามเขตและกลุ่ม เขต 5,6,7,8, และ 9

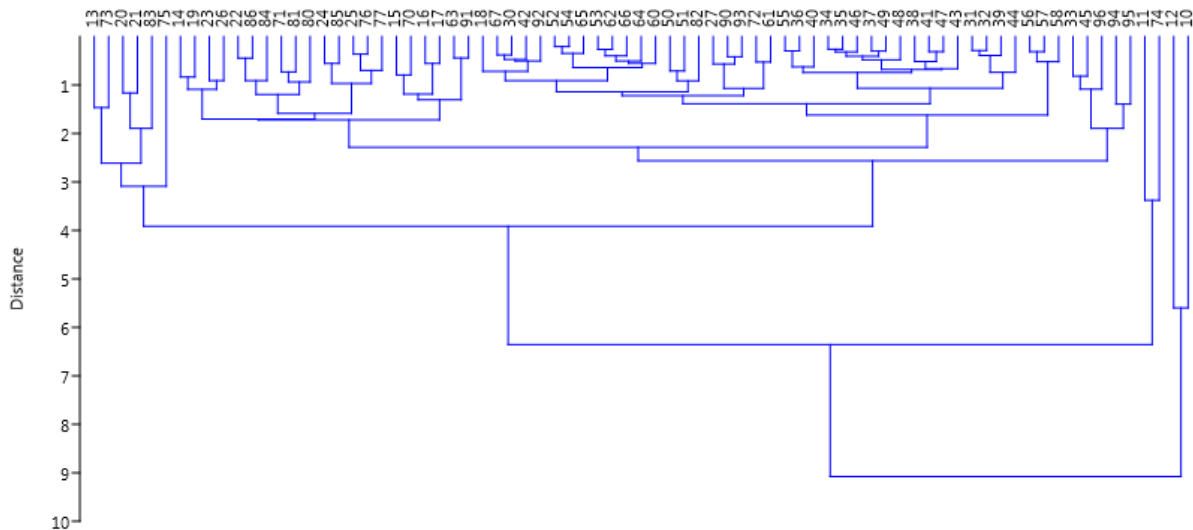
กลุ่ม 1		กลุ่ม 2		กลุ่ม 3		กลุ่ม 4	
รหัส	จังหวัด	รหัส	จังหวัด	รหัส	จังหวัด	รหัส	จังหวัด
		50	เชียงใหม่				
		51	ลำพูน				
		52	ลำปาง				
		53	อุตรดิตถ์				
		54	แพร่				
		55	น่าน				
		56	พะเยา				
		57	เชียงราย				
		58	แม่ฮ่องสอน				
		60	นครสวรรค์			63	ตาก
		61	อุทัยธานี				
		62	กำแพงเพชร				
		64	สุโขทัย				
		65	พิษณุโลก				
		66	พิจิตร				
		67	เพชรบูรณ์				
70	ราชบุรี	71	กาญจนบุรี	74	สมุทรสาคร		
73	นครปฐม	72	สุพรรณบุรี				
75	สมุทรสงคราม						
76	เพชรบุรี						
77	ประจวบคีรีขันธ์						
80	นครศรีธรรมราช	82	พังงา	83	ภูเก็ต		
81	กระบี่						
84	สุราษฎร์ธานี						
85	ระนอง						
86	ชุมพร						
		90	สงขลา			91	สตูล
		92	ตรัง			94	ปัตตานี
		93	พัทลุง			95	ยะลา
						96	นราธิวาส

หมายเหตุ: พื้นที่ควบคุม (สีส้ม) พื้นที่เฝ้าระวังสูง (สีเหลือง) และ พื้นที่นำร่องท่องเที่ยว (สีฟ้า) (ศูนย์บริหารสถานการณ์ โควิด-19, 2565)

ตารางที่ 7 แสดงสถิติเชิงพรรณนาของกลุ่ม

รายการ	กลุ่ม				รวม
	1	2	3	4	
<b>จำนวนจังหวัด</b>	23	28	5	21	77
<b>จำนวนคนต่อครัวเรือน</b>					
ต่ำสุด	1.43	2.10	1.51	2.84	1.43
สูงสุด	2.68	2.77	1.94	3.66	3.66
เฉลี่ย	2.22	2.51	1.76	3.14	2.55
ส่วนเบี่ยงเบนมาตรฐาน	0.34	0.18	0.16	0.25	0.48
<b>ความหนาแน่นประชากร (คนต่อตารางกิโลเมตร)</b>					
ต่ำสุด	59	23	673	41	23
สูงสุด	780	194	3,524	376	3,524
เฉลี่ย	206	93	1678	145	244
ส่วนเบี่ยงเบนมาตรฐาน	171	41	1173	63	479
<b>ความหนาแน่นครัวเรือน (ครัวเรือนต่อตารางกิโลเมตร)</b>					
ต่ำสุด	29	9	347	14	9
สูงสุด	433	76	2,006	103	2,006
เฉลี่ย	98	37	953	46	117
ส่วนเบี่ยงเบนมาตรฐาน	94	16	665	16	275
<b>รายได้ต่อครัวเรือนเฉลี่ย (พันบาท/เดือน)</b>					
ต่ำสุด	21.991	15.496	27.591	16.569	15.496
สูงสุด	39.507	26.532	41.129	26.542	41.129
เฉลี่ย	29.333	21.530	35.033	21.269	24.666
ส่วนเบี่ยงเบนมาตรฐาน	4.461	3.154	5.627	2.594	5.762
<b>อัตราผู้ติดเชื่อต่อประชากรกลางปี 1,000 คน</b>					
ต่ำสุด	46	9	113	26	9
สูงสุด	149	74	294	108	294
เฉลี่ย	90	39	189	45	66
ส่วนเบี่ยงเบนมาตรฐาน	26	18	65	20	47
<b>อัตราผู้เสียชีวิตต่อประชากรกลางปี 10,000 คน</b>					
ต่ำสุด	3	1	4	1	1
สูงสุด	12	5	18	8	18
เฉลี่ย	7	3	12	3	5
ส่วนเบี่ยงเบนมาตรฐาน	2	1	6	2	3

การจัดทำแผนผังต้นไม้โดยการจัดกลุ่มจังหวัดด้วยวิธี Hierarchical Clustering ใช้ระยะห่าง Euclidean Distance ดังแสดงในภาพที่ 2 รหัสจังหวัดในกลุ่ม 3 คือ 10 กรุงเทพฯ 11 สมุทรปราการ 12 นนทบุรี 74 สมุทรสาคร อยู่เรียงกันด้านบนขวามือ ส่วน 83 ภูเก็ต แยกออกมาอยู่ทางซ้าย ซึ่งการจัดกลุ่มจังหวัดในภาพแผนผังต้นไม้ก็สอดคล้องเป็นส่วนใหญ่กับการจัดกลุ่มด้วยวิธีการ k-means clustering ในตารางที่ 5 และ 6



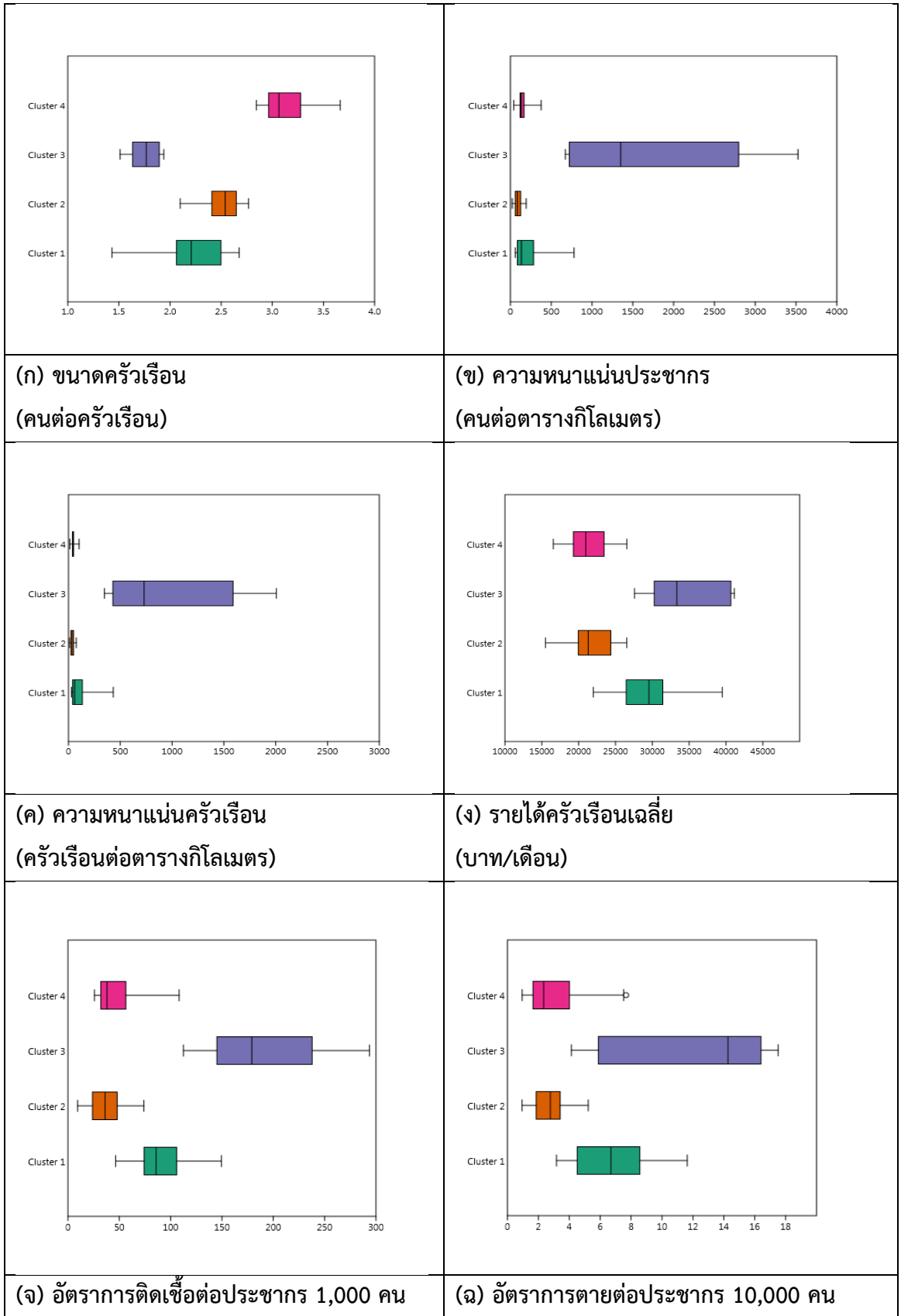
ภาพที่ 2 แสดงแผนผังต้นไม้การจัดกลุ่มจังหวัดด้วยวิธี Hierarchical Clustering

สัมประสิทธิ์สหสัมพันธ์เพียร์สันของปัจจัยที่นำมาวิเคราะห์กลุ่ม ดังแสดงในตารางที่ 8 พบความสัมพันธ์ประชากรต่อตารางเมตรสัมพันธ์กับครัวเรือนต่อตารางเมตร และ อัตราการติดเชื้อมีอัตราการเสียชีวิตมีความสัมพันธ์กันสูง เป็น 1 และ 0.87 ตามลำดับ เมื่อตัดครัวเรือนต่อตารางเมตร และอัตราการเสียชีวิตออกไปผลการวิเคราะห์กลุ่มมีเพียง 2 จังหวัดที่ย้ายกลุ่ม คือ จังหวัดระยองและชลบุรี ซึ่งย้ายจากกลุ่ม 1 เป็นกลุ่ม 2 เท่านั้น จึงคงไว้ทุกปัจจัย

ตารางที่ 8 แสดงสหสัมพันธ์เชิงเส้นตรง

หัวข้อ	คนต่อครัวเรือน	คนต่อ ตร.กม.	ครัวเรือนต่อ ตร.กม.	รายได้ ครัวเรือน	อัตราการติดเชื้อ	อัตราการเสียชีวิต
คนต่อครัวเรือน		0.001	0.000	0.000	0.000	0.000
คนต่อตร.กม.	-0.36		0.000	0.000	0.000	0.000
ครัวเรือนต่อ ตร.กม.	-0.41	1.00		0.000	0.000	0.000
รายได้ครัวเรือน	-0.62	0.57	0.59		0.000	0.000
อัตราผู้ติดเชื้อ	-0.51	0.54	0.54	0.54		0.000
อัตราผู้เสียชีวิต	-0.45	0.54	0.53	0.54	0.87	

*p-value* แสดงด้วยอักษรตัวเอน 0.001 สัมประสิทธิ์สหสัมพันธ์แสดงด้วยอักษรตัวตรง เช่น 1.00



ภาพที่ 3 แสดงแผนผังกล่องปัจจัยที่นำมากำหนดการวิเคราะห์กลุ่ม

## การอภิปรายผล

จากการจัดทำแผนภาพกล่องจำแนกปัจจัยที่นำมาวิเคราะห์กลุ่ม ดังแสดงในภาพที่ 3 กลุ่ม 3 จะมีความแตกต่างไปจากกลุ่มอื่นในทุกปัจจัย กลุ่ม 2 และ 4 มีค่า Interquartile range ที่ซ้อนทับในทุก ๆ ปัจจัย กลุ่มที่มีจำนวนจังหวัดน้อยที่สุด คือ กลุ่ม 3 มีจำนวน 5 จังหวัด คือ ภาคกลาง ได้แก่ กรุงเทพมหานคร นนทบุรี สมุทรปราการ ภาคตะวันตก คือ สมุทรสาคร ภาคใต้ฝั่งอ่าวไทย คือ ภูเก็ต จากการจัดกลุ่มจังหวัดเป็น 4 กลุ่ม แล้วแสดงเป็นแผนผังกล่องในภาพที่ 2 การที่กลุ่ม 2 และ 4 มีค่า Interquartile range ที่ซ้อนทับในทุก ๆ ปัจจัยกล่าวได้ว่าสามารถรวมกันได้ ก็จะเหลือเพียง 3 กลุ่ม คือ กลุ่ม 1 อัตราการติดเชื้อและอัตราการเสียชีวิตปานกลาง กลุ่ม 2 และ 4 อัตราการติดเชื้อและอัตราการเสียชีวิตต่ำ และกลุ่ม 3 อัตราการติดเชื้อและอัตราการเสียชีวิตสูง

การจัดจำนวนกลุ่มเป็น 4 กลุ่ม ก็สอดคล้องกับจำนวนกลุ่มที่มีการจัดในประเทศอินโดนีเซียของ Herawati et al. (2022) โดยที่มากกว่าการจัดกลุ่มของประเทศอินเดียโดย Zargar et al. (2021) ที่จัดเพียง 3 กลุ่ม ซึ่งก็สามารถดำเนินมาตรการในการควบคุมและป้องกันได้อย่างเหมาะสมกับจำนวนกลุ่มที่แตกต่างกันในจำนวนกลุ่ม/รูปแบบมาตรการที่ไม่มากจนเกินไป

สหสัมพันธ์เชิงเส้นตรงจากตารางที่ 8 จำนวนคนต่อตารางกิโลเมตร และจำนวนครัวเรือนต่อตารางกิโลเมตร มีค่าสหสัมพันธ์กับอัตราผู้ติดเชื้อ 0.54 และ 0.54 ตามลำดับ และกับอัตราผู้เสียชีวิต 0.54 และ 0.53 ตามลำดับด้วย ซึ่งสอดคล้องกับการจัดกลุ่มพื้นที่ในประเทศอินเดียของ Sengupta et al. (2021) ได้ดำเนินการวิเคราะห์กลุ่มเขตท้องที่ต่าง ๆ จากจำนวนผู้ป่วยยืนยัน จำนวนประชากรความหนาแน่นและจำนวนโรงพยาบาลที่ใช้รักษาโควิด-19 ซึ่งพื้นที่ที่มีความหนาแน่นของประชากรก็จะผู้ติดเชื้อมาก

## บทสรุปและข้อเสนอแนะ

สรุปการประชุมคณะกรรมการบริหารสถานการณ์การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (โควิด - 19) ครั้งที่ 2/2565 วันพฤหัสบดีที่ 20 มกราคม 2565 ได้ปรับระดับของพื้นที่สถานการณ์ย่อยในพื้นที่ทั่วราชอาณาจักรของโรคติดเชื้อไวรัสโคโรนา 2019 (24 มกราคม 2565) จัดกลุ่มจังหวัดออกเป็น 3 กลุ่ม จำแนกออกเป็น พื้นที่ควบคุม (สีส้ม) 44 จังหวัด พื้นที่เฝ้าระวังสูง (สีเหลือง) 25 จังหวัด และ พื้นที่นำร่องท่องเที่ยว (สีฟ้า) 8 จังหวัด (ศูนย์บริหารสถานการณ์ โควิด-19, 2565) ดังแสดงแถบสีบนรหัสและชื่อจังหวัดในตารางที่ 6 และ 7 มีความแตกต่างไปจากการใช้ปัจจัยความหนาแน่นของประชากร ขนาดครัวเรือน รายได้เฉลี่ยต่อเดือน อัตราการติดเชื้อ และอัตราการเสียชีวิต ซึ่งอาจเกิดจากการคำนึงถึงผลกระทบทางเศรษฐกิจจากแหล่งท่องเที่ยวด้วย อย่างไรก็ตาม เมื่อทำการซ้อนการจัดกลุ่มจังหวัดของศูนย์บริหารสถานการณ์ โควิด-19 เป็นพื้นที่ควบคุม พื้นที่เฝ้าระวังสูง กับ พื้นที่นำร่องท่องเที่ยว ก็กับการวิเคราะห์กลุ่มจัดเป็น 4 กลุ่ม ที่สามารถแสดงระดับอัตราการติดเชื้อและอัตราการเสียชีวิตในระดับต่ำกลางและสูงที่แตกต่างกันให้พิจารณาเพิ่มขึ้นอีกมิติหนึ่งด้วย สามารถใช้เป็นข้อมูลการปฏิบัติดูแลตนเองเมื่อเข้าไปในพื้นที่ดังกล่าว

การจัดกลุ่มจังหวัดตามบริบทของหน่วยงานที่รับผิดชอบก็สามารถนำมาจัดใหม่ตามวิธีการทางสถิติที่ใช้ระยะห่างทางเรขาคณิตของข้อมูลหลายมิติโดยการรวบรวมข้อมูลจากปัจจัยที่เกี่ยวข้องกับโรคติดเชื้อทางเดินหายใจนำมาจัดกลุ่มเพื่อให้สอดคล้องกับวัตถุประสงค์ตามบริบทของประชาชนในพื้นที่ซึ่งมีความสำคัญต่อการเฝ้าระวังและป้องกันตนเอง ทั้งนี้ทำการวิเคราะห์ด้วยโปรแกรมสำเร็จรูปที่เป็นที่ยอมรับเพื่อนำไปใช้งานได้เอง นอกจากนี้ การเพิ่มปัจจัยด้านจำนวนนักท่องเที่ยวจากประเทศที่มีความเสี่ยงในระดับต่าง ๆ กันเข้าเป็นปัจจัยพิจารณาในการจัดกลุ่มในแต่ละช่วงเวลา ก็จะทำให้ยังสามารถทำข้อเสนอเพื่อการดำเนินนโยบายของรัฐบาลในการตัดสินใจเตรียมความพร้อมเกี่ยวกับสิ่งอำนวยความสะดวกทางการแพทย์ที่เหมาะสมเช่นเดียวกับงานวิจัยจัดกลุ่มพื้นที่ปกครองในประเทศอินเดีย

## เอกสารอ้างอิง

- กรมควบคุมโรค. (2566). *สถานการณ์ผู้ติดเชื้อ COVID-19 อัปเดตรายสัปดาห์ ในรูปแบบ API (JSON/CSV Data Format)* ณ. วันที่ 30 ธันวาคม 2565. สืบค้นจาก <https://covid19.ddc.moph.go.th/api/Cases/today-cases-by-provinces>
- กรมควบคุมโรค. (2564). *การจัดกลุ่มประเทศและดินแดนตามความเสี่ยงสถานการณ์โรคโควิด-19*. สืบค้นจาก <https://ddc.moph.go.th/uploads/ckeditor2//files/จัดกลุ่มประเทศตามความเสี่ยง กรมควบคุมโรค 1 สค 64.pdf>
- กรมควบคุมโรค. (2563). *กลุ่มประเทศที่มีความเสี่ยงต่ำ และคำแนะนำการเดินทางที่ปลอดภัย สำหรับคนไทย ที่เดินทางไปต่างประเทศ ในช่วงโควิด-19*. สืบค้นจาก <https://ddc.moph.go.th/dcd/pagecontent.php?page=204&dept=dcd>
- โครงการ SAT COVID Dashboard. (2566). *สถานการณ์ผู้ป่วยโควิด-19 ภายในประเทศ รายสัปดาห์ ปรับปรุงล่าสุด 1 มกราคม 2566*. สืบค้นจาก <https://ddc.moph.go.th/covid19-dashboard/>
- รัฐบาลไทย. (2565). *ข่าวคณะโฆษก: กระทรวงสาธารณสุขออกประกาศ 1 ต.ค. 65 เป็นต้นไป ยกเลิกโควิด-19 จากการเป็นโรคติดต่ออันตราย และกำหนดให้เป็นโรคติดต่อที่ต้องเฝ้าระวัง*. สืบค้นจาก <https://www.thaigov.go.th/news/contents/details/59491>
- สำนักข่าว Hfocus เจาะลึกระบบสุขภาพ. (2565). "หมอยง" ชี้โควิดระบอบรอบนี้ นับเป็นระลอก 6 ยังไม่พีคสุด แต่ผู้ป่วยส่วนใหญ่ไม่มีอาการ. สืบค้นจาก <https://www.hfocus.org/content/2022/07/25500>
- สำนักงานปลัดกระทรวงสาธารณสุข. (2566). *คกก.โรคติดต่อฯ เห็นชอบมาตรการป้องกัน “โควิด” รองรับ การเข้าประเทศ พร้อมตั้งคณะอนุฯ บริหารแบบไร้รอยต่อ*. สืบค้นจาก <https://pr.moph.go.th/?url=pr/detail/2/04/184448/>



- สำนักงานสถิติแห่งชาติ. (2565ก). 1 สถิติประชากรและการเคหะ. จำนวนประชากรจากการทะเบียน ชาย หญิง เนื้อที่ ความหนาแน่น และบ้าน จำแนกตามภาค และจังหวัด พ.ศ. 2555-2564. สืบค้นจาก <http://statbbi.nso.go.th/staticreport/page/sector/th/01.aspx>
- สำนักงานสถิติแห่งชาติ. (2565ข). 8 สถิติรายได้และรายจ่ายของครัวเรือน รายได้เฉลี่ยต่อเดือนของครัวเรือน เป็นรายภาค และจังหวัด พ.ศ. 2547 – 2564. สืบค้นจาก <http://statbbi.nso.go.th/staticreport/page/sector/th/08.aspx>
- สำนักบริหารการทะเบียน กรมการปกครอง. (2565). สถิติประชากรทางการทะเบียนราษฎร (รายเดือน). สืบค้นจาก <https://stat.bora.dopa.go.th/stat/statnew/statMONTH/statmonth/#/displayData>
- ศูนย์บริหารสถานการณ์ โควิด-19 (ศบค.). (2565). สรุปการประชุมคณะกรรมการบริหารสถานการณ์ การแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (โควิด – 19) ครั้งที่ 2/2565 วันพฤหัสบดีที่ 20 มกราคม 2565. สืบค้นจาก [https://media.thaigov.go.th/uploads/public\\_img/source/200165.pdf](https://media.thaigov.go.th/uploads/public_img/source/200165.pdf)
- Everitt, B. S., Landau, S., Leese, M., & Stahl D. (2011). *Cluster Analysis*. 5<sup>th</sup> ed. West Sussex: John Wiley & Sons.
- Evgeniou, T. (n.d.). Cluster Analysis and Segmentation What is this for? Retrieved from <https://inseaddataanalytics.github.io/INSEADAnalytics/CourseSessions/Sessions45/ClusterAnalysisReading.html>
- Hammer, Ø. (2022). *Past 4 - the Past of the Future. Current version (December 2022): 4.12 Version 4.09, 32-bit Windows (7 MB)*. Retrieved from <https://www.nhm.uio.no/english/research/resources/past/>
- Herawati, N., Nisa, K., & Saidi, S. (2022). *Implementation of the trimmed k-means clustering method in mapping the distribution of Covid-19 in Indonesia*. AIP Conference Proceedings, 050012, 1-12. <https://doi.org/10.1063/5.0103175>
- Kumar, S. (2020). Use of cluster analysis to monitor novel coronavirus-19 infections in Maharashtra, India. *Indian J Med Sci*, 72(2), 44–8. DOI: 10.25259/IJMS\_68\_2020. PMID: PMC7485640
- Tan, P-N., Steinbach, M., & Kumar, V. (2018). *Introduction to Data Mining*. 2<sup>nd</sup> ed. London: Pearson+.

- Raj, A., Bhattacharyya, P., & Gupta, G. R. (2022). Clusters of COVID-19 Indicators in India: Characterization, Correspondence and Change Analysis. *SN Comput Sci.*, 3(3), 210. DOI: 10.1007/s42979-022-01083-3. Epub 2022 Apr 5. PMID: 35400015; PMCID: PMC8981186.
- Sadeghi B., Cheung R.C.Y., and Hanbury M. (2021). Using hierarchical clustering analysis to evaluate COVID-19 pandemic preparedness and performance in 180 countries in 2020. *BMJ Open*, 11(11), e049844. doi: 10.1136/bmjopen-2021-049844
- Sengupta, P., Ganguli, B., SenRoy, S., & Chatterjee, A. (2021). An analysis of COVID-19 clusters in India. *BMC Public Health*, 21, 631. <https://doi.org/10.1186/s12889-021-10491-8>
- The Data Science Lab. (2013). *Finding the K in K-Means Clustering*. Retrieved from <https://datasciencelab.wordpress.com/2013/12/27/finding-the-k-in-k-means-clustering/>
- Zargar, S, A., Islam, T., Rehman, I. U., & Pandey, D. (2021). Use of Cluster Analysis to Monitor Novel Corona Virus (COVID-19) Infections in India . *Asian Journal of Advances in Medical Science*, 3(1), 32-38. <https://mbimph.com/index.php/AJOAIMS/article/view/1908>