

ชื่อวิทยานิพนธ์	การประมาณในช่วงโดยใช้ฟังก์ชันเสมือนพหุนามและภาพฉายสำหรับข้อมูลเชิงประชากรและข้อมูลเชิงพื้นที่
ผู้เขียน	นายอัครชัย เอื้ออนันตสันต์
สาขาวิชา	วิธีวิทยาการวิจัย
ปีการศึกษา	2555

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อสร้างการแสดงผลเชิงภาพที่มีความชัดเจนและถูกต้องและเพื่อแสดงสารสนเทศจากข้อมูลที่มีให้มากที่สุดสำหรับสนับสนุนการรับรู้และเข้าใจข้อมูลของผู้อ่านข้อมูล วิทยานิพนธ์ฉบับนี้ประกอบด้วยงานวิจัย 4 เรื่องโดยมีวัตถุประสงค์ในการแสดงตัวอย่างจริงที่เกี่ยวข้องกับการรับรู้ข้อมูลของมนุษย์ โดยใช้โปรแกรม R ในการวิเคราะห์เชิงสถิติและประยุกต์การสร้างกราฟสำหรับแสดงผล

งานวิจัยเรื่องแรกคือการประมาณค่าในช่วงสำหรับข้อมูลทางประชากร การประมาณค่าในช่วงที่นิยมใช้กับข้อมูลทางประชากรได้แก่ฟังก์ชันเสมือนพหุนามธรรมชาติลำดับสาม (natural cubic spline function) มีข้อกำหนดทั้งเงื่อนไขขอบ และข้อกำหนดที่ฟังก์ชันต้องเป็นฟังก์ชันทางเดียวแบบไม่ลดค่า เนื่องจากข้อมูลทางประชากรมักเป็นข้อมูลที่ไม่เป็นจำนวนลบ จากข้อกำหนดข้างต้น การศึกษาในหัวข้อนี้จึงใช้เทคนิคเพิ่มจุดข้อมูลเสมือนเพื่อใช้ในการประมาณค่าในช่วงจากฟังก์ชันเสมือนพหุนามธรรมชาติลำดับสามกับข้อมูลตัวอย่าง 3 กลุ่ม ได้แก่ ข้อมูลการให้กำเนิดบุตรของหญิงในประเทศอิตาลี ในปี พ.ศ. 2498 ข้อมูลประชากรชายในประเทศไทยปี พ.ศ. 2543 และข้อมูลการเสียชีวิตของหญิงในประเทศออสเตรเลียในปี พ.ศ. 2444 ผลการศึกษาพบว่าการเพิ่มจุดข้อมูลเสมือนสำหรับฟังก์ชันในการประมาณค่าในช่วงเป็นไปตามเงื่อนไขที่กำหนดไว้สำหรับฟังก์ชันเสมือนพหุนามธรรมชาติลำดับสาม รวมถึงเส้นโค้งของฟังก์ชันมีความเรียบต่อเนื่อง นอกจากนี้การเปรียบเทียบกับผลการศึกษาของนักวิจัยอื่นพบว่าเทคนิคเพิ่มจุดข้อมูลเสมือนให้การปรับเรียบของเส้นโค้งที่ดีกว่าและใช้การคำนวณที่มีความซับซ้อนน้อยกว่า

งานวิจัยเรื่องที่สองเกี่ยวข้องกับการสร้างแผนที่แสดงข้อมูลการเสียชีวิตที่ไม่สามารถระบุสาเหตุที่มีการลงทะเบียนไว้ในทะเบียนราษฎร์ของประเทศไทยระหว่างปี พ.ศ. 2542 ถึงปี พ.ศ. 2544 วิเคราะห์ข้อมูลด้วยแบบจำลองการถดถอยลอจิสติก เพื่อหาอัตราส่วนของข้อมูลการเสียชีวิตนอกโรงพยาบาลที่ถูกบันทึกเป็นไม่สามารถระบุสาเหตุได้ เพื่อทดสอบหาคุณภาพของข้อมูล

ทั้งหมด สร้างช่วงความเชื่อมั่นของอัตราส่วนด้วย weighted sum contrasts เพื่อแสดงจำนวนการเสียชีวิตที่ไม่สามารถระบุสาเหตุได้ ช่วงความเชื่อมั่นที่มากกว่า เท่ากับ และน้อยกว่าค่าเฉลี่ยรวมใช้ในการกำหนดสี 3 สีที่แตกต่างกันเพื่อใช้เป็นสัญลักษณ์แทน แต่ละสีประยุกต์ใช้กับแผนที่เพื่อระบุระดับของข้อมูลการเสียชีวิตที่ไม่สามารถระบุสาเหตุได้ในทั้งประเทศและในกรุงเทพมหานคร ผลการศึกษาชี้ว่า การใช้ตัวแบบการถดถอยลอจิสติกเหมาะสมกับชนิดของข้อมูลเป็นอย่างดี และพบว่าการบันทึกสาเหตุการเสียชีวิตที่ไม่สามารถระบุสาเหตุได้นั้นพบในหญิงมากกว่าชาย และการใช้แผนที่เฉพาะเรื่องที่มีการระบุสีแทนระดับของข้อมูลทำให้เกิดความเข้าใจกับผู้รับข้อมูลได้ดีกว่าการใช้กราฟแสดงช่วงความเชื่อมั่น

การวิจัยเรื่องที่สามเป็นการปรับปรุงเทคนิคที่ใช้ในงานวิจัยเรื่องที่สอง เพื่อแสดงถึงระดับปริมาณข้อมูลที่ไม่สามารถแสดงได้ในแผนที่สองมิติอันเนื่องมาจากข้อจำกัดของมิติของข้อมูล การแสดงผลในงานวิจัยนี้ประยุกต์ใช้แผนที่สามมิติที่เกิดจาก Google Earth ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลการบาดเจ็บและเสียชีวิตอันเนื่องมาจากความไม่สงบในสามจังหวัดชายแดนใต้และสี่อำเภอในจังหวัดสงขลา วิเคราะห์หาตัวแบบการถดถอยถูกนำมาใช้ในการสร้างอัตราอุบัติการณ์สำหรับเพศ อายุ และพื้นที่ที่เกิดเหตุการณ์ที่แตกต่างกัน สร้างช่วงความเชื่อมั่นด้วยเทคนิค weighted sum contrasts แปลงกราฟช่วงความเชื่อมั่นที่ได้ให้เป็นกราฟแท่งหลายชั้นแบบสามมิติโดยใช้ระดับสี 3 สีที่แสดงถึงปริมาณของอัตราอุบัติการณ์ จากนั้นไปแสดงบนแผนที่สามมิติในกูเกิ้ลเอิร์ธพร้อมกับรูปภาพแสดงข้อมูลอื่น ๆ ในลักษณะของการซ้อนภาพบนหน้าจอ ผลการวิจัยพบว่าการใช้แผนที่สามมิติแสดงข้อมูลได้ครบถ้วนกว่าการใช้กราฟช่วงความเชื่อมั่นและแผนที่สองมิติ

การวิจัยเรื่องที่ดีเป็นการวิเคราะห์ผลการเลือกตั้งสมาชิกสภาผู้แทนราษฎรแบบแบ่งเขตในการเลือกตั้ง ปี พ.ศ. 2548 พ.ศ. 2550 และพ.ศ. 2554 วิเคราะห์ผลการเลือกตั้งเพื่อทำการสร้างแผนที่นำแผนที่มาใช้ในการแสดงผลการเลือกตั้งโดยใช้สีแสดงสัญลักษณ์ของแต่ละพรรคที่ได้คะแนนรวมสูงสุดในแต่ละจังหวัด และมีการเพิ่มจุดสีเพื่อแสดงถึงผู้สมัครที่ได้รับการเลือกตั้งของแต่ละพรรคลงบนแผนที่ ขยายพื้นที่บริเวณกรุงเทพมหานครและปริมณฑลเพื่อแสดงจุดสีที่ซ้อนทับกันให้เห็นชัดเจนมากขึ้น โดยใช้เทคนิค variable-scaled map ผลที่ได้ทำให้สามารถวิเคราะห์ผลการเลือกตั้งได้ดีมากกว่าการใช้ตัวเลขผลคะแนนหรือตัวเลขจำนวนผู้ได้รับเลือกตั้ง ผลการศึกษาพบว่าโดยส่วนใหญ่พรรคที่ได้คะแนนเสียงรวมสูงสุดในแต่ละจังหวัดใดจังหวัดหนึ่งมีจำนวนผู้ได้รับเลือกตั้งมากที่สุดในจังหวัดนั้น แผนที่แสดงผลการเลือกตั้งในปี พ.ศ. 2550 และ 2554 แสดงให้เห็นชัดเจนว่าประชาชนในประเทศไทยมีแนวโน้มเลือกพรรคการเมืองใหญ่เพียงสองพรรค โดยแบ่งตามภูมิศาสตร์อย่างชัดเจนเป็นภาคเหนือและตะวันออก และภาคใต้และตะวันตก นอกจากนี้การ

วิเคราะห์โดยใช้แผนภูมิ bubble เพื่อแสดงถึงคะแนนโหวตที่เปลี่ยนไปในแต่ละการเลือกตั้ง พบว่า พรรคขนาดเล็กได้รับผลคะแนนที่ดีขึ้นในปี พ.ศ. 2550 แต่สูญเสียคะแนนไปมากในการเลือกตั้งในปี พ.ศ. 2554 ผลการวิเคราะห์ทั้งหมดที่เกิดขึ้นแสดงให้เห็นว่าการใช้การแสดงผลเชิงภาพสำหรับข้อมูลการเลือกตั้งสร้างความเข้าใจได้ดีกว่าการใช้ข้อมูลตัวเลขจากตาราง

เนื่องจากความประทับใจแรกของคนเป็นตัวละครต้นการตอบสนองทางความคิด และบุคคลมักเชื่อในสิ่งที่มองเห็นจากการนำเสนอข้อมูลเท่านั้น การนำเสนอข้อมูลเชิงภาพที่ดีไม่เพียงแสดงสารสนเทศที่มีทั้งหมดจากข้อมูล แต่ยังช่วยให้บุคคลสามารถพิจารณา และแสดงการตัดสินใจได้ถูกต้องจากข้อมูลที่ถูกนำเสนอ วิทยานิพนธ์ฉบับนี้ได้พิสูจน์ว่า การใช้เทคนิคที่เหมาะสมช่วยปรับปรุงการนำเสนอเชิงภาพให้แสดงข้อมูลได้ดีขึ้น

Prince of Songkla University
Pattani Campus

Thesis Title Spline Interpolation and Projection for Demographic and Geographic Data

Author Mr. Attachai Ueranantasun

Major Program Research Methodology

Academic Year 2012

ABSTRACT

Using the concept of What You See Is All There Is and System I of thinking modes from the book, Thinking, Fast and Slow by Daniel Kahneman, this thesis aims to create graphical displays to support a person's mind in comprehending as much information as possible from a clear and accurate data visualization. Four different applications are chosen to provide examples of real situations involving human information perception. Statistical analysis and graphical displays in all four applications are implemented using the R package.

The first application is to interpolate demographic data. A preferable interpolation for demographic data is a natural cubic spline function, which requires boundary conditions to be satisfied. Another requirement is that the interpolating spline is monotonic non-decreasing, since age-specific demographic data, such as populations, fertility, and mortality aggregate, are necessarily non-negative. In this study, a natural cubic spline function with additional knots is implemented to fulfill the requirements. Three sets of data are used, comprising Italian fertility in 1955, Thai male population in 2000, and Australian female mortality in 1901. The results show that a method of adding extra knots to a natural cubic spline interpolation fulfills the requirements and the

interpolating functions are smooth. Compared to previous work by other researchers, the results are superior in terms of smoothness and simplicity.

The second application involves registered ill-defined mortality data in Thailand between 1999 and 2001 to be displayed on a thematic map. A logistic regression model is used to approximate the proportion of ill-defined deaths outside hospital in Thailand and thus evaluate the quality of the data. Confidence intervals for proportions are constructed based on weighted sum contrasts to show the magnitude of ill-defined mortality for each super-district. Confidence intervals above, across, and below the overall mean are coded as three different colors, and these colors are applied to corresponding regions to create thematic maps for Bangkok and Thailand. The results show that ill-defined mortality across super-districts in Thailand is reasonably well fitted to a logistic regression model, while show a higher percentage of ill-defined causes of death among females. These thematic maps, better than confidence intervals, enable the viewer to straightforwardly compare results in different regions.

The third application is designed to improve the technique from the second application in order to highlight the magnitudes of the incidences. Due to the limitation of a two-dimensional map, a three-dimensional map is selected and created by freely available software (Google Earth). Incidence rates per 100,000 populations of terrorism events in three southernmost provinces of Thailand and four districts of Songkla are chosen. Regression models are used for creating incidence rates for gender, age group, and region, and weighted sum contrasts are used to obtain confidence intervals for comparing the incidence rates. Confidence intervals are

converted into stacked bar charts with three colored components indicating the magnitudes. With Google Earth, three-dimensional stacked bar charts displaying magnitudes of incidences are placed in a corresponding region for both males and females and other graphics can also be added as a screen overlay. The results show that using three-dimensional maps provides better comprehension than using only confidence intervals and thematic maps.

The fourth application examines Thailand's national election results in 2005, 2007, and 2011. The voting results are analyzed and thematic maps for the whole of Thailand are produced to display a majority vote in a province with color points representing winning candidates for each province. A variable-scaled map is used to clarify a focused area and distinguish otherwise dense plotted points in areas around Bangkok. The graphical results clearly illustrate greater information than using figures and tables. The analysis from the graphics shows that the number of winning candidates coincides with the percentage of majority votes in almost all provinces. The thematic maps of 2007 and 2011 also illustrate that there is a clear split between the North-East and the South-West, indicating a noticeable separation of supports for two major parties. Furthermore, the vote swings are estimated from 2005 to 2007 and 2007 to 2011. This method is implemented by using a simple statistical analysis to distribute gain-loss percentages in proportion, and to present them in the form of bubble plots. The results show that minor parties in 2007 gained benefit from the political turmoil, but lost their significant votes in 2011. This graphical method provides superior information to only a table.

Since a person's first impression stimulates a rapid response and a person tends to believe only what can be observed from the data presented, good data visualization provides not only *all there is* available from information presented, but also accommodate a person's ability in judging and making a correct decision based on the visualization. This thesis has proved that appropriate techniques can offer better data visualization and thus complement a person's responsive and subconscious stage of thinking.

Prince of Songkla University
Pattani Campus