

Contract Number BRG4980012

ชื่อโครงการ: ระเบียบวิธีเชิงสถิติสำหรับการวิเคราะห์ข้อมูลทางชีววิทยา สิ่งแวดล้อมและการแพทย์:
ทฤษฎีและการประยุกต์

หัวหน้าโครงการ: รองศาสตราจารย์ ดร. มนต์ทิพย์ เทียนสุวรรณ

ที่อยู่: ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ถนนพระรามที่ 6 กรุงเทพมหานคร
10400

โทรศัพท์: 02-2015538, 02-2015340, Fax: 02-2015343, e-mail: scmts@mahidol.ac.th

บทคัดย่อ

การวิเคราะห์ข้อมูลคือภาคจำนวนเต็มของสถิติและวิธีเชิงสถิติในการวิเคราะห์ข้อมูลที่เกิดขึ้นจากแหล่งซึ่งแตกต่างกันมีมากมายในเอกสารข้อมูลหรือวรรณกรรม อย่างไรก็ตามความต้องการวิธีการใหม่ ๆ สำหรับการวิเคราะห์ข้อมูลหรือการประยุกต์วิธีธรรมดาที่นิยมไปสู่บริเวณใหม่จะเกิดขึ้นเสมอๆ เพราะว่าข้อมูลใหม่ปรากฏออกมาและวิธีการเก็บรวบรวมข้อมูลยากและปัญหาเชิงสถิติใหม่ๆ ในบริบทของวิทยาศาสตร์สิ่งแวดล้อมผู้วิจัยให้ความสนใจในการประเมินค่าสภาพแวดล้อมบนฐานของมลพิษทางอากาศและคุณภาพน้ำ โดยการประยุกต์วิธีการตัดสินใจเกณฑ์พหุคูณและการแก้ไขเพิ่มเติมในการคำนวณดัชนีมลพิษทางอากาศทั้งหมดและพารามิเตอร์คุณภาพน้ำ สำหรับจัดการตัดสินใจระหว่างสถานี ดังนั้นการเปรียบเทียบและการจัดลำดับของบริเวณจึงเป็นไปได้ในเทอมของมลพิษทางอากาศและคุณภาพน้ำสำหรับการป้องกันสภาพแวดล้อม การวิเคราะห์อนุกรมเวลาได้ถูกเลือกใช้เป็นเทคนิคสำหรับการพยากรณ์ปริมาณน้ำฝนรายเดือนและจำนวนวันที่ฝนตก เทคนิคบอซ-เจนกินส์ได้ถูกนำมาใช้ในการระบุพารามิเตอร์ของตัวแบบการถดถอยในตัวเองรวมค่าเฉลี่ยเคลื่อนที่ (ARIMA) เกณฑ์ที่ใช้ในการพิจารณาความเหมาะสมของตัวแบบ คือ เกณฑ์ข้อสนเทศของอาากิเคะ (AIC) เกณฑ์ซวาร์ชเบเชียน (SBC) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) การวิเคราะห์แผนภาพเป็นคาบได้ถูกนำมาใช้เพื่อยืนยันการมีอยู่ของคาบฤดูกาลในตัวแบบ ARIMA ตัวแบบ ARIMA แบบฤดูกาลอาจจะใช้ทำนายปริมาณน้ำฝนรายเดือนและจำนวนวันที่ฝนตก หนึ่งปีล่วงหน้าด้วยความแม่นยำที่ยอมรับได้ ผู้วิจัยใช้ การทดสอบลอการิทึม-จัดลำดับและ Cox's proportional hazard models ในการเปรียบเทียบการรอดชีพและหาปัจจัยที่มีผลต่อการรอดชีพของผู้ป่วยเอชไอวี/เอดส์ ในประเทศไทย ผู้วิจัยได้แสดงการอนุพัทธ์ตัวประมาณค่าแบบเบสส์ของพารามิเตอร์ของการแจกแจงปรกติสองตัวแปรภายใต้ข้อจำกัดค่าเฉลี่ยร่วมหรือความแปรปรวนร่วม ผู้วิจัยได้เสนอตัวอย่างการประยุกต์ทางการแพทย์และสิ่งแวดล้อม ผลการศึกษาการจำลองบนพื้นฐานของตัวอย่างขนาดเล็กโดยเปรียบเทียบระหว่างวิธีภาวะน่าจะเป็นสูงสุดกับวิธีของเบสส์ส่วนใหญ่มีลักษณะสมมูล นอกจากนี้ผู้วิจัยได้แสดงการอนุพัทธ์ตัวประมาณค่าแบบเบสส์ของพารามิเตอร์ในการแจกแจงแบบเลขชี้กำลัง และฟังก์ชันความน่าเชื่อถือได้ที่สอดคล้องกับการแจกแจงนี้สำหรับชนิดของข้อมูลที่แตกต่างกัน โดยใช้สถิติแบบอันดับ

ภายใต้ความน่าจะเป็นก่อนแบบเกมมาและแบบเจฟฟรีย์ ผู้วิจัยพบว่าตัวประมาณค่าแบบเบส์ได้ถูกแสดงอย่างชัดเจนและสามารถแสดงในรูปแบบปิดได้ภายใต้ความน่าจะเป็นก่อนแบบเจฟฟรีย์ นิพจน์ของการประมาณสำหรับความเสี่ยงแบบเบส์ได้ถูกแสดงเชิงทฤษฎี และสามารถคำนวณได้เชิงตัวเลขเช่นกัน สุดท้ายผู้วิจัยให้ความสนใจการประมาณค่าพารามิเตอร์และการเปรียบเทียบตัวประมาณค่าสำหรับตัวแบบภายในชั้นด้วยวิธีภาวะน่าจะเป็นสูงสุดกับวิธีของเบส์โดยศึกษาจากตัวอย่างขนาด 1 วงศ์ และ n วงศ์ ตามลำดับ วิธีที่ศึกษาใช้ค่าตลาดเคลื่อนกำลังสองเฉลี่ยจากการจำลองข้อมูล วิธีทั้ง 2 ได้นำมาประยุกต์กับข้อมูลระดับก่อนการเก็บรักษา ATP ในเซลล์เม็ดเลือดแดงโดยวัดจากพี่น้องใน 18 วงศ์

คำสำคัญ: เอดส์, ตัวชี้วัดมลพิษทางอากาศ, สหสัมพันธ์ในตัวเอง, ตัวประมาณค่าแบบเบส์, วิธีของเบส์, ความเสี่ยงแบบเบส์, การแจกแจงปรกติสองตัวแปร, ตัวแบบบอซซ์-เจนกินส์, ค่าเฉลี่ยร่วม, ความแปรปรวนร่วม วิธีอิลเคเตอร์, ความน่าจะเป็นก่อนแบบเกมมา, เอชไอวี, ตัวแบบภายในชั้น, ความน่าจะเป็นก่อนแบบเจฟฟรีย์, วิธีแคพแลน-เมียร์เออร์, วิธีความควรจะเป็นสูงสุด, การตัดสินใจเกณฑ์พหุคูณ, การแจกแจงปรกติหลายตัวแปร, สหสัมพันธ์ในตัวเองบางส่วน, การวิเคราะห์แผนภาพเป็นคาบ, ปริมาณน้ำฝน, ฟังก์ชันความเชื่อถือได้, การวิเคราะห์การรอดชีพ, การวิเคราะห์อนุกรมเวลา, ตัวชี้วัดคุณภาพน้ำ

Contract Number BRG4980012

Project Title: Statistical Methods of data Analysis in Biology Environment and
Medicine: Theory and Applications

Principal Investigator: Assoc. Prof. Dr. Montip Tiensuwan

Address: Department of Mathematics Faculty of Science, Mahidol University,
Rama 6 Road, Bangkok, Thailand 10400

Telephone:02-2015538, 02-2015340, **Fax:** 02-2015343, **e-mail:** scmts@mahidol.ac.th

Abstract

Data analysis is an integral part of statistics, and statistical methods to analyze data arising from various sources abound in the literature. Nevertheless, the need for new methods of data analysis or application of conventional methods to new area often arises because of emerging new and difficult data collection methods and innovative statistical problems. In the context of environmental science, we focused on environmental assessment based on the air pollution and water quality. We applied Multiple Criteria Decision Making (MCDM) method and its modification to compute an overall air pollution index and water quality parameters for making an overall judgment among stations. Therefore, comparing and ranking of regions are feasible in terms of air pollution and water quality for environmental protection. A time series analysis is adopted as a technique for forecasting of monthly rainfall and number of rainy days. The Box-Jenkins technique is used for identifying the parameters of an Autoregressive Integrated Moving Average (ARIMA) model. The Akaike information criterion, the Schwarz's Bayesian criterion and the mean square error are used throughout to test for simplification of any particular model. The periodogram analysis is used to confirm the existence of a seasonal period in the ARIMA model. The ARIMA with seasonal model possibly predicts the monthly rainfall and the number of rainy days one year ahead with acceptable accuracy. Log-rank test and Cox's proportional hazard models are applied to compare survival curves and find the covariates which effect survival time for HIV/AIDS patients in Thailand.

Further, we derive Bayes estimates of the parameters of a bivariate normal population under the constraint of either a common mean or a common variance. Some environmental and medical applications are indicated. Results of a simulation study based on a small sample size to compare the maximum likelihood estimates and the Bayes estimates indicate that they are mostly equivalent. We also derived the Bayes estimators of the parameter of an exponential distribution and the reliability function associated with it based on different type of truncated data sets under the gamma and Jeffreys' priors. The Bayes estimators are derived explicitly and can be obtained in the closed form. Approximate expressions for Bayes risk are also derived theoretically and numerically. Lastly, we focused on estimating parameters and comparing estimators in intraclass models based on the method of maximum likelihood and a Bayesian method for one family and n families, separately. A simulation study is presented, in which the performance of the method of maximum likelihood is compared with the Bayesian method by using mean square error. Further the two methods are also applied to the pre-storage adenosine triphosphate levels data in red blood cells as measures in the siblings of 18 families.

Keywords: AIDS, air pollution indicators, autocorrelation, Bayes estimator, Bayesian method, Bayes risk, bivariate normal population, Box-Jenkins model, common mean, common variance, maximum, electre method, gamma prior, HIV, intraclass models, Jeffreys' prior, Kaplan-Meier method, maximum likelihood method, multiple criteria decision making, multivariate normal distribution, partial autocorrelation, periodogram analysis, rainfall, reliability function, survival analysis, time series analysis, water quality indicators.