แผนภูมิควบคุมรอยตำหนิสำหรับกระบวนการผลิตที่มีศูนย์มาก

วิลาสินี ปิยะจิตร[1] และ ธิดาเดียว มยุรีสวรรค์[2]

แผนภูมิควบคุมรอยตำหนิ (Nonconforming Control Charts) มีพื้นฐานการสร้างมาจากการแจกแจงแบบปัวส์ซอง (Poisson distribution) ในการสุ่มตัวอย่างผลิตภัณฑ์จากกระบวนการผลิตที่มีคุณภาพดีขึ้นมาตรวจสอบแต่ละครั้ง จะพบจำนวนรอยตำหนิที่เกิดขึ้นในหน่วยตัวอย่างเป็นศูนย์มาก ซึ่งเรียกการแจกแจงแบบปัวส์ซองที่มีศูนย์มากว่า Zero-Inflated Poisson (ZIP) โดยตัวแบบ ZIP ในงานวิจัยนี้ ได้ทำการศึกษาประสิทธิภาพของแผนภูมิควบคุมรอยตำหนิที่ใช้ควบคุมกระบวนการผลิต เมื่อกระบวนการผลิตมีสัดส่วนของจำนวนรอยตำหนิเป็นศูนย์มาก โดยทำการศึกษาแผนภูมิควบคุมรอยตำหนิวิธี ZIP (C_{ZIP}-Chart) แผนภูมิควบคุมรอยตำหนิวิธี Jeffreys prior interval (C_J-Chart) และนำไปเปรียบเทียบกับแผนภูมิควบคุมรอยตำหนิพื้นฐาน (C-Chart) เมื่อกระบวนการผลิตมีสัดส่วนของจำนวนรอยตำหนิที่เป็นศูนย์เท่ากับ 0.3, 0.4, 0.5 และ 0.6 ทั้งกรณีที่กระบวนการผลิตมีค่าเฉลี่ยของรอยตำหนิอยู่ภายใต้การควบคุม และเปลี่ยนแปลงไป โดยใช้เกณฑ์ในการประเมินประสิทธิภาพของแผนภูมิควบคุมคือ ค่าความยาววิ่งเฉลี่ย (Average Run Length: ARL) และค่าความน่าจะเป็นครอบคลุม(Coverage Probability: CP) ผลการวิจัยพบว่า C_J-Chart มีประสิทธิภาพมากที่สุด ในกรณีที่กระบวนการผลิตอยู่ภายใต้การควบคุม โดยมีสัดส่วนของจำนวนรอยตำหนิที่เป็นศูนย์เท่ากับ 0.3, 0.4 และ 0.5 มีค่าเฉลี่ยของจำนวนรอยตำหนิเท่ากับ 4.0 และ 5.5 แต่เมื่อกระบวนการผลิตมีค่าเฉลี่ยของจำนวนรอยตำหนิเท่ากับ 4.5 และ 5.0 พบว่า C_{ZIP}-Chart จะมีประสิทธิภาพมากที่สุด และเมื่อกระบวนการผลิตมีสัดส่วนของจำนวนรอยตำหนิที่เป็นศูนย์เท่ากับ 0.6 ทุกระดับค่าเฉลี่ยของจำนวนรอยตำหนิที่กำหนดในแผนการทดลอง พบว่า C_J-Chart มีประสิทธิภาพมากที่สุด แต่เมื่อกระบวนการผลิตมีการเปลี่ยนแปลงในค่าเฉลี่ยของจำนวนรอยตำหนิทุกระดับ ที่กำหนดในแผนการทดลอง พบว่า C-Chart มีประสิทธิภาพในการตรวจจับการเปลี่ยนแปลงดีที่สุด แต่ให้ค่าความน่าจะเป็นครอบคลุมแย่ที่สุด

© 2010 by the AMM2010 All rights reserve.

แผนภูมิควบคุมโดยการสุ่มตัวอย่างแบบกลุ่มลำดับ

ณัฐสุวัชร์ ถาวรธิรา[1] และ ธิดาเดียว มยุรีสวรรค์[2]

แผนภูมิควบคุม (Control Chart) เป็นเทคนิคทางสถิติที่นำมาใช้ในการควบคุมกระบวนการผลิตอย่างแพร่หลาย โดยทั่วไป ข้อมูลจากการผลิตที่นำมาสร้างแผนภูมิควบคุมจะได้จากการสุ่มตัวอย่างแบบอย่างง่าย (Simple Random Sampling: SRS) หรือการสุ่มตัวอย่างแบบระบบ (Systematic Sampling: SS) ต่อมาได้มีนักวิจัยนำเสนอการสุ่มตัวอย่างแบบกลุ่มลำดับ (Ranked-Set Sampling: RSS) และทำการเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยของประชากรที่ได้จากการสุ่มตัวอย่างแบบ SRS และแบบ RSS และสร้างช่วงความเชื่อมั่น 100(1-α)% ของค่าเฉลี่ยของประชากร แต่ยังไม่มีการนำช่วงความเชื่อมั่นนี้มาประยุกต์ใช้ในการสร้างแผนภูมิควบคุมค่าเฉลี่ย งานวิจัยนี้ได้นำการสุ่มตัวอย่างแบบ RSS และช่วงความเชื่อมั่น 100(1-α)% ของค่าเฉลี่ยของประชากรที่ได้จากการสุ่มตัวอย่างแบบ RSS มาประยุกต์ใช้ในการสร้างแผนภูมิควบคุมค่าเฉลี่ย ($\bar{x}$-RSS chart) และแผนภูมิควบคุมค่ามัธยฐาน (M-RSOS chart) และได้ออกแบบวิธีการสุ่มตัวอย่างขึ้นมาอีก 1 วิธี นั่นคือการสุ่มตัวอย่างแบบกลุ่มมัธยฐาน (Median-Set Sampling: MSS) จากการสุ่มตัวอย่างแบบ RSS และนำข้อมูลที่ได้จากการสุ่มตัวอย่างที่ออกแบบมาสร้างแผนภูมิควบคุมค่าเฉลี่ย ($\bar{x}$-MSS chart) และแผนภูมิควบคุมค่ามัธยฐาน (M-MSOS chart) ทำการเปรียบเทียบประสิทธิภาพของแผนภูมิควบคุมกับแผนภูมิควบคุมค่าเฉลี่ยที่ได้จากการสุ่มตัวอย่างแบบพื้นฐานคือแบบ SRS ($\bar{x}$-chart) ผลการวิจัยพบว่า ในกรณีที่ค่าเฉลี่ยของกระบวนการผลิตไม่มีการเปลี่ยนแปลง M-RSOS chart มีประสิทธิภาพในการตรวจสอบกระบวนการผลิตดีที่สุด ในกรณีที่ค่าเฉลี่ยของกระบวนการผลิตมีการเปลี่ยนแปลงเล็กน้อย M-MSOS chart มีประสิทธิภาพในการตรวจจับการเปลี่ยนแปลงของกระบวนการผลิตดีที่สุด ในกรณีที่ค่าเฉลี่ยของกระบวนการผลิตมีการเปลี่ยนแปลงมาก $\bar{x}$-RSS chart, M-RSOS chart, $\bar{x}$-MSS chart และ M-MSOS chart มีประสิทธิภาพในการตรวจจับการเปลี่ยนแปลงของกระบวนการผลิตใกล้เคียงกัน

© 2010 by the AMM2010 All rights reserve.

การประมาณฟังก์ชันความหนาแน่นน่าจะเป็นของข้อมูลจากการแจกแจงผสม

เบญจวรรณ อยู่ถาวร[1] วิจิตรา พลเยี่ยม[2] และ เสาวณิต สุขภารังษี[3]

การศึกษารูปแบบการแจกแจงของข้อมูล เป็นสิ่งสำคัญในการวิเคราะห์ข้อมูลเพื่อให้ได้ผลสรุปที่ถูกต้องและแม่นยำ แต่อาจมีบางข้อมูลที่ไม่ทราบถึงรูปแบบการแจกแจงที่แน่ชัด จึงต้องทำการประมาณฟังก์ชันความหนาแน่นน่าจะเป็น (Probability Density Function : PDF) เพื่อให้ทราบถึงรูปแบบการแจกแจงของข้อมูลนั้นๆ โดยมีนักวิจัยได้ทำการศึกษาวิธีการประมาณฟังก์ชันความหนาแน่นน่าจะเป็นที่เหมาะสม เมื่อข้อมูลมีลักษณะต่างๆ เช่น ข้อมูลที่มีลักษณะสมมาตร (Symmetric) ข้อมูลลักษณะการแจกแจงที่มี ตัวแปรหลายตัว (Multivariate Distribution) เป็นต้น โดยวิธีโมเมนต์ (Moment Method) เป็นวิธีที่นิยมใช้ประมาณฟังก์ชันความหนาแน่นน่าจะเป็น แบบพาราเมตริกซ์ ซึ่งใช้ได้ดีกับข้อมูลที่มีลักษณะเป็นการแจกแจงแบบ หางยาว แต่ยังไม่มีการนำข้อมูลที่เกิดจากการผสมกันมาประมาณฟังก์ชันความหนาแน่นน่าจะเป็นด้วยวิธีการนี้ งานวิจัยนี้ได้นำข้อมูลการแจกแจงผสม (Mixture Distribution) จากการรวมตัวกันของการแจกแจงแบบไคสแควร์ (Chi-square Distribution) ที่มีพารามิเตอร์ต่างกันมาเปรียบเทียบวิธีการประมาณฟังก์ชันความหนาแน่นน่าจะเป็นด้วยเทคนิคโมเมนต์ 3 วิธีได้แก่ โมเมนต์รอบจุดกำเนิด ลำดับที่ 1 (First Ordinary Moment) โมเมนต์รอบจุดกำเนิด ลำดับที่ 2 (Secondary Ordinary Moment) และ โมเมนต์รอบจุดศูนย์กลาง ลำดับที่ 2 (Secondary Central Moment) โดยประเมินประสิทธิภาพของฟังก์ชันความหนาแน่นน่าจะเป็นด้วยค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE) และฟังก์ชันความน่าจะเป็นสะสม (Cumulative Distribution Function : CDF) ผลการศึกษาพบว่า ในกรณีที่มีพารามิเตอร์เท่ากันวิธีโมเมนต์รอบจุดกำเนิด ลำดับที่ 1 เป็นวิธีที่ประมาณฟังก์ชันความหนาแน่นน่าจะเป็นได้ใกล้เคียงกับฟังก์ชันความหนาแน่น น่าจะเป็นแท้จริงมากที่สุด ในกรณีที่มีพารามิเตอร์แตกต่างกันเล็กน้อยวิธีโมเมนต์รอบจุดศูนย์กลางลำดับที่ 2 เป็นวิธีประมาณฟังก์ชันความหนาแน่นน่าจะเป็นได้ใกล้เคียงกับฟังก์ชันความหนาแน่นน่าจะเป็นที่แท้จริงมากที่สุด และในกรณีที่มีพารามิเตอร์แตกต่างกันมาก

© 2010 by the AMM2010 All rights reserve.

การประมาณฟังก์ชันความหนาแน่นน่าจะเป็นของข้อมูลจากการแจกแจงผสม

Density Estimations of Data from Mixture Distributions.

เบญจวรรณ อยู่ถาวร*, วิจิตรา พลเยี่ยม, เสาวณิต สุขภารังษี

Benjawan Yoothaworn*, Wichitra Phonyiem, Saowanit Sukparungsee

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* E-mail : bbcstat@gmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาวิธีประมาณฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงผสมไคสแควร์ที่พารามิเตอร์ต่างกัน และการแจกแจงเอ็กซ์โพเนนเชียลที่พารามิเตอร์ต่างกัน โดยวิธี Moment Matching 4 วิธี ได้แก่ โมเมนต์ที่ 1 รอบจุดกำเนิด (First Ordinary Moment) โมเมนต์ที่ 2 รอบจุดกำเนิด (Secondary Ordinary Moment) โมเมนต์ที่ 2 รอบจุดศูนย์กลาง (Secondary Central Moment) และโมเมนต์ที่ 3 รอบจุดศูนย์กลาง (Third Central Moment) วัดประสิทธิภาพของฟังก์ชันความหนาแน่นน่าจะเป็นด้วยค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (Mean Integrate Square Error: MISE) จากการศึกษาพบว่า ในกรณีที่การแจกแจงผสมเป็นการแจกแจงไคสแควร์ที่มีความเบ้มากวิธีโมเมนต์ที่ 2 รอบจุดศูนย์กลาง เป็นวิธีที่เหมาะสม และที่มีความเบ้น้อยวิธีโมเมนต์ที่ 2 รอบจุดกำเนิดเป็นวิธีที่เหมาะสม และในกรณีการแจกแจงไฮเปอร์เอ็กซ์โพเนนเชียล เมื่อการแจกแจงมีความเบ้มากวิธีโมเมนต์ที่ 1 รอบจุดกำเนิด เป็นวิธีที่เหมาะสม และเมื่อความเบ้ของการแจกแจงน้อยวิธีโมเมนต์ที่ 2 รอบจุดศูนย์กลางเป็นวิธีที่เหมาะสม

คำสำคัญ: ฟังก์ชันความหนาแน่นน่าจะเป็น, เทคนิคโมเมนต์, การแจกแจงผสม, การแจกแจงแบบไคสแควร์ , การแจกแจงไฮเปอร์เอ็กซ์โพเนนเชียล

การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าความยาววิ่งเฉลี่ยของแผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักด้วยเอกซ์โพเนนเชียล เมื่อข้อมูลมีการแจกแจงแบบทวินาม ด้วยเทคนิคมอนติคาร์โล และวิธีลูกโซ่มาร์คอฟ

An Efficient Comparison of Average Run Length Approximation methods of Exponentially Weighted Moving Average Control Chart when Observations are Binomial Distributed by Monte Carlo Simulation and Markov Chain Approach

นิธิยา ทองยวน*, เสาวณิต สุขภารังษี, ยุพาภรณ์ อารีพงษ์

Nitiya Thongyoun*, Saowanit Sukparungsee, Yupaporn Areepong

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* E-mail : amm_oh@hotmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการประมาณค่าประสิทธิภาพความยาววิ่งเฉลี่ยของแผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักด้วยเอกซ์โพเนนเชียล เพื่อตรวจจับจำนวนของเสียที่เพิ่มขึ้น ด้วยวิธีลูกโซ่มาร์คอฟ และเปรียบเทียบความถูกต้องของผลลัพธ์ที่ได้จากเทคนิคการจำลองมอนติคาร์โล โดยวิธีลูกโซ่มาร์คอฟสามารถหาความยาววิ่งเฉลี่ยเพื่อใช้ในการตรวจจับการเปลี่ยนแปลงที่เกิดขึ้นได้รวดเร็วกว่าเทคนิคการจำลองมอนติคาร์โล เมื่อกำหนดขนาดการเปลี่ยนแปลง และค่าความยาววิ่งเฉลี่ยของกระบวนการอยู่ภายใต้การควบคุม (ARL_0) โดยกำหนดขนาดตัวอย่าง n = 30, 50 และ 100 สัดส่วนของเสียเมื่อกระบวนการอยู่ภายใต้การควบคุม α_0 =0.01 และกำหนดให้สัดส่วนของเสียเพิ่มขึ้น 1%, 2%, … , 10% เมื่อกระบวนการอยู่นอกเหนือการควบคุม โดยผลลัพธ์ที่ได้จากวิธีดังกล่าวมีความถูกต้องและแม่นยำ เทียบเท่ากับเทคนิคการจำลองมอนติคาร์โล แต่ใช้เวลาในการคำนวณน้อยกว่า

คำสำคัญ: สำหรับแผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักเอกซ์โพเนนเชียล, จำนวนของเสีย, ลูกโซ่มาร์คอฟ, เทคนิคการจำลองมอนติคาร์โล, ค่าความยาววิ่งเฉลี่ย

การเปรียบเทียบวิธีการประมาณค่าสูญหาย สำหรับตัวแบบการพยากรณ์

A Comparison of Missing Value Estimation Methods for Forecasting Models

จิรกานต์ นวลละออง* และ เสาวณิต สุขภารังษี

Jirakarn Nunlaong and Saowanit Sukparungsee

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* E-mail : nunlaong@hotmail.com

บทคัดย่อ

ในการวิเคราะห์ข้อมูลอนุกรมเวลาและการพยากรณ์นั้นมักไม่มีความแม่นยำ ถูกต้องและน่าเชื่อถือเมื่อมีข้อมูลสูญหายเกิดขึ้น งานวิจัยนี้มีวัตถุประสงค์ในการศึกษาวิธีเพื่อประมาณค่าสูญหายเพื่อใช้ในการวิเคราะห์ตัวแบบพยากรณ์ โดยทำการเปรียบเทียบวิธี การประมาณค่าสูญหาย 3 วิธีเดี่ยว คือ วิธีค่าเฉลี่ย, วิธีกำลังสองน้อยสุด, วิธีการใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่า และเสนอวิธีการใหม่เรียกว่า "การประมาณค่าสูญหายร่วม" โดยนำวิธีการประมาณค่าสูญหายเดี่ยว 3 วิธีมารวมกันด้วยตัวถ่วงน้ำหนักโดยมีวิธีการค่าถ่วงน้ำหนัก 3 วิธี คือ วิธีการถ่วงน้ำหนักที่เท่ากัน, วิธีถ่วงน้ำหนักโดยค่าสัมบูรณ์ต่ำสุด และวิธีการถ่วงน้ำหนักโดยความแปรปรวนต่ำสุด โดยจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลด้วยโปรแกรม R ซ้ำ 50,000 ครั้งและกำหนดขนาดตัวอย่าง 30, 50 และ 100 ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1, 2, 5 และ 10 เปอร์เซ็นต์การสูญหายเท่ากับ 5%, 10%, 15% และ 20% ขนาดความสัมพันธ์ของตัวแปรอิสระ คือ 0, 0.2 และ 0.5 เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์

วิธีที่เหมาะสมกับข้อมูลภาคตัดขวางคือวิธีค่าเฉลี่ย วิธีที่เหมาะสมกับข้อมูลอนุกรมเวลา คือ วิธีกำลังสองน้อยสุด สำหรับทุกกรณีศึกษามีเพียง 3 กรณี วิธีค่าเฉลี่ยเหมาะสมกับข้อมูล

แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบชี้กำลัง สำหรับข้อมูลที่มีการแจกแจงแบบล็อกนอร์มอล เพื่อตรวจจับการเปลี่ยนแปลงพารามิเตอร์ของกระบวนการ

Lognormal EWMA Control Chart for Monitoring of Changes in a Parameter

วรพล สมานันตกุล*, ยุพาภรณ์ อารีพงษ์ และ เสาวณิต สุขภารังษี

Worapon Samanuntakul, Yupaporn Areepong and Saowanit Sukparungsee

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

E-mail: pyyn@hotmail.com*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษากระบวนการควบคุมคุณภาพสำหรับข้อมูลที่มีการแจกแจงแบบล็อกนอร์มอล (Lognormal Distribution) ด้วยแผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบชี้กำลัง (Exponential Weight Moving Average Control Chart: EWMA) กรณีที่ขนาดการเปลี่ยนแปลงของพารามิเตอร์มีขนาดเล็ก โดยวัดประสิทธิภาพของแผนภูมิด้วยค่าความยาววิ่งเฉลี่ย (Average Run Length: ARL) โดยปกติแล้วเมื่อกระบวนการอยู่ภายใต้การควบคุม (in-control process) กำหนดให้ค่า ARL = ARL_0 แต่ในกรณีที่กระบวนการอยู่นอกเหนือการควบคุม (out-of-control process) กำหนดให้ ARL = ARL_1 ซึ่งปัจจุบันวิธีที่เป็นที่นิยมใช้ทั่วไป คือ วิธีการจำลองมอนติคาร์โล (Monte Carlo Simulations: MC) โดยวิธีนี้ให้ผลลัพธ์ที่ได้เป็นค่าที่ถูกต้อง และสามารถนำมาเปรียบเทียบกับผลลัพธ์ที่หาได้จากวิธีศึกษาอื่น ๆ แต่เป็นที่รู้กันดีว่าใช้เวลาเป็นอย่างมากในการประมาณค่า ARL ผู้วิจัยจึงได้พัฒนาวิธีการหาค่า ARL_0 และ ARL_1 โดยวิธีสมการปริพันธ์ (Integral Equations Method: IE) ด้วยกฎของซิมป์สัน (Simpson's Rule) และ กฎค่ากลาง (Midpoint's Rule) ผลการวิจัยพบว่าผลลัพธ์ที่ได้จากวิธี IE โดยกฎของซิมป์สัน (Simpson's Rule) ค่าที่ได้มีความถูกต้องเทียบเท่ากับวิธี MC แต่ใช้เวลาในการประมวลผลน้อยกว่าวิธี MC นอกจากนี้ยังศึกษาเปรียบเทียบประสิทธิภาพของแผนภูมิ EWMA และแผนภูมิควบคุมผลรวมสะสม (Cumulative Sum Control Chart: CUSUM) พบว่าแผนภูมิ EWMA นั้นมีประสิทธิภาพดีกว่า แผนภูมิ CUSUM เมื่อ ARL_0 มีค่าต่ำ แต่ในกรณีที่ค่า ARL_0 สูงขึ้นนั้น แผนภูมิ EWMA จะมีประสิทธิภาพดีกว่าเมื่อขนาดของการเปลี่ยนแปลงมีขนาดเล็ก

คำสำคัญ: ความยาววิ่งเฉลี่ย, แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักแบบชี้กำลัง, แผนภูมิควบคุมผลรวมสะสม, สมการปริพันธ์, การแจกแจงแบบล็อกนอร์มอล, กฎของซิมป์สัน

การเปรียบเทียบวิธีการคำนวณค่า ARL ด้วยวิธี MC และ MCA สำหรับแผนภูมิ EWMA เมื่อข้อมูลมีการแจกแจงแบบปัวส์ซอง

Comparison of ARL by Monte Carlo Simulation and Markov Chain Approach for Poisson EWMA Control Chart

อุมา รัตนเทพี วิชัย สุรเชิดเกียรติ และ เสาวณิต สุขภารังษี

Uma Rattanathepee Wichai Suracherdkiati and Saowanit Sukparungsee

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

E-mail: eoagnon@gmail.com; wsc@kmutnb.ac.th; snws@kmutnb.ac.th

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาคู่พารามิเตอร์ที่เหมาะสมกับแผนภูมิ EWMA นั่นคือ ค่าสถิติของ EWMA (λ) และค่าขีดจำกัดควบคุมบนของแผนภูมิ EWMA สำหรับข้อมูลที่มีการแจกแจงแบบปัวส์ซอง เพื่อตรวจจับการเปลี่ยนแปลงค่าเฉลี่ยในกระบวนการ (α) ได้อย่างรวดเร็ว เมื่อกำหนดระดับการเปลี่ยนแปลงค่าเฉลี่ย $\alpha = 1.05,\ 1.10,\ 1.15,\ 1.20,\ 1.25,\ 1.30,\ 1.40,\ 1.50,\ 1.75$ และ 2.00 ตามลำดับ โดยได้ทำการศึกษาวิธีการคำนวณค่าความยาววิ่งเฉลี่ย (Average Run Length: ARL) เมื่อกระบวนการอยู่ภายใต้การควบคุม (in-control state) แทนด้วยสัญลักษณ์ ARL_0 และเมื่อกระบวนการอยู่นอกเหนือการควบคุม (out-of-control state) แทนด้วยสัญลักษณ์ ARL_1 2 วิธี นั่นคือ วิธีการจำลองมอนติคาร์โล (Monte Carlo Simulation: MC) และวิธีลูกโซ่มาร์คอฟ (Markov Chain Approach: MCA) ซึ่งจากการศึกษาพบว่าทั้งสองวิธีนั้นให้ค่า ARL ใกล้เคียงกัน แต่วิธี MCA นั้นใช้เวลาในการประมวลผลน้อยกว่าวิธี MC จึงเลือกใช้วิธี MCA ในการคำนวณค่า ARL_1 เพื่อหาคู่พารามิเตอร์ที่เหมาะสม ในการตรวจจับการเปลี่ยนแปลงค่าเฉลี่ยในกระบวนการ ซึ่งเกณฑ์ในการเลือกคู่พารามิเตอร์ที่เหมาะสมนั้น จะทำการเลือกคู่พารามิเตอร์ (λ, H) ที่ให้ค่า ARL_1 ที่ต่ำที่สุดในแต่ละระดับ α ผลการวิจัยพบว่าเมื่อค่า α เพิ่มขึ้น ค่าพารามิเตอร์ (λ, H) นั้นมีค่าเพิ่มมากขึ้นเช่นเดียวกัน

คำสำคัญ: แผนภูมิควบคุมค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนักด้วยเอกซโพเนนเชียล, ความยาววิ่งเฉลี่ย, การแจกแจงปัวส์ซอง, วิธีการลูกโซ่มาร์คอฟ, วิธีการจำลองมอนติคาร์โล

Confidence Intervals for the Different Means Using Jackknife Method after ANOVA for Data with Outliers

Jeerapa Sappakitkamjorn* and Sa-aat Niwitpong
Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bangkok
Email: jsj@kmutnb.ac.th*, snw@kmutnb.ac.th

Abstract

In this study, the results from a continuing investigation into the performance of confidence intervals for the different means based on the jackknife method for multiple comparisons after the ANOVA F-test are presented. The main focus was to examine the jackknife based confidence interval in detecting the differences between the means in multiple comparisons tests, especially when datasets contain outliers. A Monte Carlo simulation with 5,000 runs under various situations was conducted to investigate the performance of the jackknife based intervals. As equal sample sizes were considered in this study, we compared the performance of the jackknife based intervals with those obtained by the Tukey's HSD (Honesty Significant Difference), a method that allows the comparison of all pairs of means, in terms of coverage probability and expected length. The results from a simulation study show that the jackknife based intervals enable us to better detect differences between the means when outliers are present in the datasets. The intervals constructed by the jackknife method not only have shorter lengths but also offer control over the probability of a Type I error.

Keywords: ANOVA, Confidence Interval, Jackknife Method, Multiple Comparisons, Outliers

1. Introduction

In the presence of outliers, one should not simply remove them from a dataset unless there is a proof that they are false values. Outliers are often true values that are important since they contain valuable information [1]. However, taking outliers into account, particularly in the one-way ANOVA, may cause unequal variances and non-normality. As a result, assumptions to be satisfied are violated when ANOVA is applied. In order to protect against the effects of these violations, one should apply the balanced designs where sample sizes are equal.

As reviewed in [2], when a balanced design is applied, the ANOVA F-test is relatively resistant to violations of assumptions. Similarly, the multiple comparisons tests after the rejection of null hypothesis in the one-way ANOVA will be resistant to those violations. Due to availability of several powerful multiple comparison procedures, such as Tukey's HSD, Scheffé's and Bonferroni methods, the Tukey's HSD is used in this study because it allows the comparison of all pairs of means when used with equal sample sizes.

According to our previous studies [4,5], the confidence intervals constructed by the jackknife method perform well in detecting true differences between the means in multiple comparisons tests. However, outliers were not considered in those studies. Therefore, it is of interest to further investigate the accuracy of the jackknife method for making paired comparisons in the presence of outliers.

The outline of this paper is as follows: in Section 2, a formula for a $(1-\alpha)100\%$ confidence interval for the difference between two means based on the jackknife method is given. In section 3, details of the simulation study are presented. The results and conclusion are given in Section 3 and Section 4 respectively.

2. The Jackknife based confidence interval

Following our preliminary studies that proposed a confidence interval and an adjusted confidence interval based on the jackknife method [4,5], the $(1-\alpha)100\%$ confidence interval for the difference between two means is given by

$$(\bar{x}_{i(JK)} - \bar{x}_{j(JK)}) \pm 4t_{\alpha/2,N-r}\sqrt{MSE_{(JK)}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \quad (1)$$

where

$$MSE_{(JK)} = \frac{1}{r}\sum_{h=1}^{r} Var_{(JK)}(\bar{x}_h) \quad (2)$$

$$Var_{(JK)}(\bar{x}_h) = \frac{(n_h-1)}{n_h}\sum_{k=1}^{n_h}\left(\bar{x}_{h(k)} - \bar{x}_{h(.)}\right)^2 \quad (3)$$

Confidence Intervals for the Mean of an Unknown Mean AR(1) Process Following the Dickey-Fuller Unit Root Test

Sasithorn Labnarin* and Sa-aat Niwitpong
Department of Applied Statistics,
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
Email: annarm_as@hotmail.com* , snw@kmutnb.ac.th

Abstract

This paper presents new confidence intervals for the mean of an unknown mean AR(1) process with an almost unit root. We use the unit root test to choose between a stationary time series and the time series with an almost unit root. If the unit root test does not reject the null hypothesis, the new confidence interval for the mean are constructed based on the differenced series. We also propose the new adjusted confidence interval for the mean of this process so that its minimum coverage probability is $1-\alpha$. The Monte Carlo simulation is used to investigate the behavior of these new confidence intervals compared to the existing confidence interval based on their coverage probabilities and expected lengths.

Keywords: AR(1), confidence interval, the Dickey-Fuller unit root test

On Coverage Probability of a Prediction Interval for an Unknown Mean AR(1) Process Using the Residuals Model

Wararit Panichkitkosolkul* and Sa-aat Niwitpong
Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
Email: wararit@mathstat.sci.tu.ac.th*

Abstract

A new prediction interval for an unknown mean first-order autoregressive process (AR(1)) using the residuals model is investigated in this paper. According to the Monte Carlo simulation result, a new prediction interval has a desired minimum coverage probability $1-\alpha$. Moreover, this new one is better than the existing one based on the random walk for all the autoregressive parameter values and all sample sizes considered in this paper.

Keywords: AR(1), Coverage Probability, Prediction Interval, Residuals Model

ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยในตัวแบบการถดถอยเชิงเส้นอย่างง่ายภายหลังการทดสอบสมมติฐาน

Confidence Intervals for the Response Mean of Simple Regression Following Pre –Tests

นันทยา ภารชาตรี และ สอาด นิวิศพงศ์
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: prety_oil@hotmail.com , snw@kmutnb.ac.th

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อนำเสนอช่วงความเชื่อมั่นสำหรับค่าเฉลี่ย ในตัวแบบการถดถอยเชิงเส้นอย่างง่ายภายหลังการทดสอบสมมติฐาน ซึ่งก็คือการทดสอบว่า $H_0:\beta_i=0, i=0,1$ เมื่อ β_i คือ ค่าสัมประสิทธิ์การถดถอย จากการศึกษาพบว่าช่วงความเชื่อมั่นที่เสนอขึ้นใหม่นั้นให้ค่าเฉลี่ยความกว้างของช่วงสั้นกว่าช่วงความเชื่อมั่นเดิม แต่เมื่อ β_i มีค่าเข้าใกล้ศูนย์ ช่วงความเชื่อมั่นใหม่ที่นำเสนอให้ค่าความน่าจะเป็นครอบคลุมต่ำกว่า $1-\alpha$ ดังนั้นผู้วิจัยจึงสนใจที่จะปรับค่าความน่าจะเป็นครอบคลุมของช่วงความเชื่อมั่นที่นำเสนอให้เพิ่มขึ้น โดยการนำแนวความคิดของ Olive [2007,Prediction intervals for regression models, *Computational Statistics & Data Analysis*, 51, 3115-3122] มาปรับช่วงความเชื่อมั่นดังกล่าว พบว่าช่วงความเชื่อมั่นที่ปรับใหม่นั้นให้ค่าความน่าจะเป็นครอบคลุมอย่างน้อย $1-\alpha$ และให้ค่าเฉลี่ยความกว้างของช่วงสั้นที่สุดเมื่อ β_i มีค่าเข้าใกล้ศูนย์

คำสำคัญ: ตัวแบบการถดถอยเชิงเส้นอย่างง่าย, ช่วงความเชื่อมั่น, การทดสอบสมมติฐาน

Abstract

This article presents new confidence intervals for the response mean of simple regression following pre-tests, for the null hypothesis $H_0:\beta_i=0, i=0,1$ when β_i are the regression coefficients. From the experiment we found that, new confidence intervals gave expected length to be shorter than the former expected lengths. However, when the values of β_i approaches zero, new confidence intervals are the coverage probabilities of less than $1-\alpha$ .Therefore, it is of interest to increase the coverage probability of confidence intervals. By using Olive[2007, Prediction intervals for regression models, *Computational Statistics & Data Analysis*, 51, 3115-3122] ,we adjust confidence intervals, so that coverage probabilities of these intervals are at least $1-\alpha$ and their expected lengths are shorter than existing confidence intervals when the value of β_i approaches zero.

Keywords: Simple linear Regression Model, Confidence Interval, Hypothesis Testing

1. บทนำ

การวิเคราะห์การถดถอย เป็นวิธีการทางสถิติอย่างหนึ่งที่นำมาใช้ในการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ(X) และตัวแปรตาม (Y) นอกจากนั้นสมการถดถอยที่ประมาณได้ ยังนำไปประยุกต์ใช้ประมาณค่าเฉลี่ยของ Y เมื่อกำหนดค่า X_0 ได้อีกด้วย พิจารณาตัวแบบถดถอยเชิงเส้นอย่างง่าย (Simple Regression Model)

$$Y_i=\beta_0+\beta_1X_i+\varepsilon_i \qquad (1)$$

A Single-Level Continuous Sampling Plan for High Quality Production Line

Tidadeaw Mayureesawan
Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok
Email: tidadeaw@yahoo.com

Abstract

This paper presents a plan SKIP-CSP-1 for inspection of a high quality continuous production line. The plan is defined by 3 parameters i (the number of consecutive non-defective units that must be produced during a 100% inspection of the line), a fraction f (the specified sampling frequency during a fractional inspection of the line) and k (the number of units for skipping over in an inspection). SKIP-CSP-1 computes 3 performance measures, average fraction inspected (AFI), average outgoing quality (AOQ) and average outgoing quality limit ($AOQL$), for given values of the parameters and incoming fraction of defective units on the line (p). The validity of the performance measure formulas have been tested by extensive simulations. The formulas of performance measures, AFI and AOQ are valid for all the sets of p, i, k, r ($r=1/f$) values. The SKIP-CSP-1 plan has been compared with CSP-1 and CSP-2 plans. On comparing AFI and AOQ, we have found that SKIP-CSP-1 does not give an appreciable difference in the number of units inspected and output quality for low level of p (0.001, 0.003) and for all the sets of i, k, r. Further, for higher levels of p, but low levels of k, it also does not give different results. However, compared with CSP-1, we have found that SKIP-CSP-1 gives lower number of units inspected and output quality than CSP-1, whereas compared with CSP-2, SKIP-CSP-1 gives higher number of units inspected and output quality than CSP-2.

Keywords: Continuous Sampling Plan, High Quality Production Line.

1. Introduction

A continuous sampling plan (CSP) is a plan of sampling inspection for a product consisting of individual units (parts, subassemblies, finished articles etc.) that is manufactured in quantity by an essentially continuous process. A CSP is applicable only to units subject to nondestructive inspection on a GO-NOGO basis. It is intended primarily for use in process inspection of parts, or final inspection of finished articles, where it is desired to have assurance that the percentage of defective units in the accepted product will be less than some prescribed low figure. The original continuous sampling plan (CSP-1) was described by H.F. Dodge and variations of the plan (e.g., CSP-1, CSP-2, CSP-2 CSP-4, CSP-5, CSP-F), have been proposed by many workers.

An inspection procedure in CSP-1 always starts with 100% inspection (screening). This screening is performed until i successive non-defective units are observed. Then the procedure samples only one of the following r units. If the sampled unit is found to be good, then the procedure continues to sample one unit from the next r, etc. As soon as a defective sample unit is observed, the procedure switches back to screening every unit and continues until a further i consecutive good units have been observed. Thus, the inspection procedure consists of alternating periods of 100% inspection and periods with f.100% inspection, where $f = 1/r$. [1]

A CSP-2 differs from a CSP-1 in the sense that during the sampling inspection period the first observed defective unit does not immediately require the procedure to change to 100% inspection. Switching only takes place if another defective is found in the following m sampled units. Frequently, one chooses $m = i$. This choice of m implies that, after discovery of the first nonconforming unit during the f.100% inspection period, the inspection needs to draw only good units in the next i sampled units. [2]

Reviews of other CSPs are now available in textbooks (see, e.g., Duncan [3], Grant [4] and Montgomery [5]).

The main objective of this paper is to develop a CSP that can be used for a high quality production line. The paper describes the following:

1) The design of a continuous sampling plan for a high quality production line which we call SKIP-CSP-1.
2) The development of the theory and formulas for important performance measures in SKIP-CSP-1, such as the average fraction inspected (AFI), the average outgoing quality (AOQ) and the average outgoing quality limit ($AOQL$).

ช่วงความเชื่อมั่นของผลต่างค่าเฉลี่ย เมื่อมีค่าสูญหายและทำการทดสอบเบื้องต้น

Confidence intervals for the difference between two means with missing data following a pretest

ภวัต ภักดิ์สรานุวัต และ สอาด นิวิศพงศ์
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: txexaxpx@hotmail.com , snw@kmutnb.ac.th

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นของผลต่างค่าเฉลี่ยจาก 2 ประชากร เมื่อข้อมูลตัวอย่างมีค่าสูญหาย ช่วงความเชื่อมั่นที่พิจารณาคือ ช่วงความเชื่อมั่นผลต่างค่าเฉลี่ยของ Welch-Satterthwaite (Iws) และช่วงความเชื่อมั่นที่ปรับปรุงโดยการใช้การทดสอบลักษณะสมมาตรเป็นการทดสอบเบื้องต้น (Iadp_s) ช่วงความเชื่อมั่นที่ปรับปรุงขึ้นจะใช้ค่าช่วงความเชื่อมั่นของ Welch-Satterthwaite เมื่อการทดสอบสอบเบื้องต้นไม่ปฏิเสธว่าข้อมูลตัวอย่างมีลักษณะสมมาตร แต่ในกรณีอื่นจะประยุกต์ใช้ช่วงความเชื่อมั่นของ Welch-Satterthwaite กับข้อมูลที่แปลงค่าโดยใส่ล็อคฐานธรรมชาติ แล้วทำการแปลงช่วงความเชื่อมั่นกลับสู่เกณฑ์ปกติ ผลการจำลองได้แสดงว่าเมื่อแทนที่ค่าสูญหายของตัวอย่างสุ่ม $\{x_i; i=1,...,n_x\}$ และ $\{y_j; j=1,...,n_y\}$ โดยวิธี random (hot deck) imputation ช่วงความเชื่อมั่นผลต่างค่าเฉลี่ยที่ปรับปรุงโดยการใช้การทดสอบลักษณะสมมาตรเป็นการทดสอบเบื้องต้นนั้น ใช้งานได้ดีกว่าช่วงความเชื่อมั่นผลต่างค่าเฉลี่ยของ Welch-Satterthwaite

คำสำคัญ: ช่วงความเชื่อมั่น, ความน่าจะเป็นครอบคลุม, การแทนที่ค่าสูญหาย, การทดสอบเบื้องต้น

Abstract

The objective of this study is to compare efficiency of confidence intervals for the difference between two means when item nonresponse occurs in sample. The confidence intervals considered are Welch-Satterthwaite confidence interval (Iws) and the adaptive interval that incorporates a preliminary test (pre-test) of symmetry for the underlying distributions (Iadp_s). The adaptive confidence intervals use Welch-Satterthwaite confidence interval if the pre-test fails to reject symmetry for distributions; otherwise, apply Welch-Satterthwaite confidence interval to the log-transformed data, then transform the interval back. Simulation studies show that the adaptive interval that incorporates the test of symmetry performs better than Welch-Satterthwaite confidence interval when we imputed values for the missing data in a random sample $\{x_i; i=1,...,n_x\}$ and $\{y_j; j=1,...,n_y\}$ based on random hot deck imputation method.

Keywords: Confidence interval, Coverage Probability, Imputation, Preliminary test

1. คำนำ

โดยทั่วไปแล้วช่วงความเชื่อมั่นของผลต่างค่าเฉลี่ย 2 ประชากรของ Welch-Satterthwaite (Welch, 1938; Satterthwaite, 1946) นั้นมีข้อจำกัดว่าข้อมูลทั้ง 2 ประชากรต้องมีการแจกแจงปกติ และความแปรปรวนมีค่าเท่ากันและแตกต่างกัน แต่ Miao และ Chiou (2008) ก็ได้เสนอช่วงความเชื่อมั่นของผลต่างค่าเฉลี่ย 2 ประชากรในกรณีที่ข้อมูล

Adjustment of estimators of the population mean in survey sampling using auxiliary information when some observations are missing

Phanuphong Thaweekaew Kanokwan Channarong and Nuanpan Nangsue
Department of Applied Statistics, Faculty of Applied Science
King Mongkut's University of Technology North Bangkok
Email: x-hollister@hotmail.com

Abstract

In this paper, we compared some estimators for estimating population mean using auxiliary information by the imputation method when some observations are missing. In this study we adjust the Rueda estimator (2005) for estimating a population mean by using the imputation method. We propose two composite imputation methods, called ratio regression imputation and regression ratio imputation, and use them to compare with the three estimators: Rueda's estimator, the estimator that uses the ratio imputation method and the estimator that uses the regression imputation method. Five methods were compared in a simulation study using population (X,Y) values of size (N) 1,000 and 5,000 for different sample sizes and correlation coefficient between X and Y is 0.5 and 0.8. In a sample 10, 20 and 30 percent of the cases will be randomly designated as missing. We fixed the missing of x variable (q) is 10, 20 and 40 percent and the missing of y variable (p) is 10 and 40 percent from each sample unit. Therefore we also report the results of a simulation study that showed that the ratio regression imputation method and the regression ratio imputation method give smaller mean absolute percentage errors than other estimators in every case.

Keywords: auxiliary information, regression imputation method, ratio imputation method, missing data and Rueda estimator.

1. Introduction

Missing data is a common problem in virtually all surveys. Frequently, survey sampling is conducted to gather complete data on all sampling units. However missing data can be contributed to bias and make the analysis harder to be conducted as well as make the results harder to be presented due to a variety of reason. For example, none of data is available or information on one or more variables are missing.

The most frequently used method to compensate for item non response is imputation (Little and Rubin, 1987). Item non response occurs if the questions that should be answered are not answered or the answer are deleted during editing. So we simply discard all the sampling units with missing values and use inference procedures, which can result in actual sample sizes being less than for other estimators, biases in estimators and increases in sampling variance if missing data follows any pattern.

In this paper we propose new estimators for estimation of a population mean using auxiliary information when some observations are missing. The estimators are adjusted from Rueda's estimator(2005) by using imputation methods and using them to compare with other estimators, that use ratio imputation and regression imputation. Specifically, we fix the case where data is missing for some subjects, and the value of an auxiliary variable is missing for other distinct subjects.

2. Methods of Estimation

Let Ω be a population of N units from which a random sample of fixed size is drawn. For this sample we observe the values of two variables, (y_i,x_i), i= 1,...,n, for the estimation of some parameter of y. We assume that only t = (n-p-q) of the observations are complete for the selected units in the sample. We have a set of q observations on the y characteristic in the sample but with the x characteristic missing. Similarly, we have a sample of p observations on the x characteristic but with the y characteristic missing.

Further, p and q are assumed to be integer numbers satisfying $0 < p, q < n/2$. We separate the units of the sample into three sets.

$s_1 = \{i \in s / x_i, y_i \text{ are available}\}$

$s_2 = \{i \in s / x_i \text{ are available, but } y_i \text{ is not available}\}$

$s_3 = \{i \in s / y_i \text{ are available, but } x_i \text{ is not available}\}$

We also let s_4 be the members of the population which are not included in the sample, i.e., no information is available for them.

We compare five methods. The first method we consider uses the estimator that Rueda et al (2005) proposed in an attempt at a post survey prediction of the mean of y. The predictor of $\overline{Y}$ was taken as:

$$T^* = f_{s_1}\bar{y}_{s_1} + f_{s_2}U_2 + f_{s_3}\bar{y}_{s_3} + (1-f_{s_4})U_4 \qquad (1)$$

Analysis of Bioactive Constituents for Lowering Blood Sugar Levels in Herb Juice Extract of *Gynura procumbens* and Development of Herb Juice Spray-Dried Powder

S. Kongtun Janphuk[*1], and W. Suracherdkaiti[2]

[1]Department of Biotechnology,
[2]Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
Email: skongtun@yahoo.com*

Abstract

Gynura procumbens (Asteraceae), has been used traditionally as anti-inflammatory remedy for skin rash and itching. Recently, research of leaves extract of this plant which effect on serum cholesterol and triglyceride level in rat has been reported. In this study, the possible hypoglycaemic constituents in herb juice of *G. Procumbens* leaves were analyzed by chromatographic techniques such as Thin-Layer Chromatography (TLC) and High-Performance Liquid Chromatography (HPLC). These techniques can be used for quality control of plant materials and phytopharmaceutical products. The herb juice from 5 provinces were spray dried to be powdered by using 1%, 1.5% and 3% of maltodextrin as bulking agent. The yields of herb juice powder were compared and analyzed by the least significant difference test (LSD), which differences with $p \leq 0.05$ were considered to be significantly different.

Keywords: hypoglycaemia, chromatography, *Gynura procumbens*

1. Introduction

Gynura procumbens is found in various parts of Southeast Asia. It has been used in Thailand as remedies for inflammation, pain, and allergic conditions. It has been used for the treatment of eruptive fevers, rash and kidney disease (1). Recently, the leaves of this plant have been used as folk medicine to control diabetes mellitus and hyperlipidemia. In some previous research, ethanolic extract of the leaves of *G. procumbens* has shown significantly reducing effect on serum cholesterol and triglyceride level in rats (2).

From phytochemistry investigation of this plant, there are some bioactive compounds which probably has hypoglycemic activities, such as phytosterol and flavonoids. The phytostefol contained in *G. Procumbens* such as stigmaterol, β-sitosterol and their glycoside. Several flavonoids have been found in this plant, such as quercetin, kaempferol and their glycoside, which possess significant pharmacological activities including anti-inflammatory, anti-allergic, and anti-oxidant properties.

The present study was to analyse and evaluate the possible hypoglycaemic constituents in herb juice of *G. Procumbens* leaves by chromatographic technique. The compound G1a (mixture of stigmasterol glucoside and β-sitosterol glucoside), quercetin, and rutin were used as markers for Thin-Layer Chromatography (TLC). The rutin and isoquercitrin content of herb juice from different sources were analyzed by High-Performance Liquid Chromatograpy (HPLC).

2. Material and methods

Preparation of the extract

The aerial part of *G. procumbens* was collected from five provice in Thailand, Supanburi (SP), Nakornsawan (NS), Nakornprathom (NP), Uthai-thani (UT), and Ayudthaya (AY). The fresh leaves of *G. procumbens* were blended and extracted with 95% alcohol (1.5L) until exhaustion. After filtration with cotton wool, the mixture was centrifuged at 10,000 g for 20 min. The supernatant was evaporated at 38°C by a rotavapor (Buchi Labortechnik AG, Switzerland). This solution was then freeze-dried, yielding 16.38 g of light green powder. The extract was suspended in distilled water before use.

The fresh leaves of *G. procumbens* (about 40 kg) from each province were washed and cut, then put them in blender (juice extractor) until herb juice was obtained. The weigh of herb juice (kg) of each source was recorded. Fresh herb juice was storage in freezer at -20 °C. The herb juice was standing in room temperature until it solute to liquid before use.

The herb juice samples (150 g) from each province were dried to be powder by spray-dryer technique with using maltodextrin (1%, 1.5%, and 3% w/v) as bulking agent.

3. Chemical constituents analysis on *G. procumbens*

Gynura procumbens herb juice and ethanolic extract were analyzed for possible hypoglycemic constituents by using chromatographic technique. G1a (mixture of stigmasterol glycoside and β-sitosterol glycoside) rutin (quercetin rutinoside) and quercetin were used as makers for Thin-Layer Chromatography

Confidence Intervals for a Coefficient of Variation Using Re-sampling Methods

Pianpool Kirdwichai * and Suvimol Phanyaem
Department of Applied Statistics
King Mongkut's University of Technology North Bangkok, Thailand
Email: kpp@kmutnb.ac.th*

Abstract

The coefficient of variation (CV), which is the ratio of the standard deviation to the mean, is a dimensionless measure of dispersion that is found to be very useful in many situations. The CV of a population can be estimated from a sample coefficient of variation which is defined as the ratio of sample standard deviation to the sample mean. McKay (1932) used a chi squared distribution as an approximation to the distribution of the sample CV and derived a confidence interval for the population CV. Vangel (1996) proposed a new approximate method of estimating a confidence interval, which he called the Modified McKay method as it was closely related to McKay's method. He found that the Modified McKay method usually gave a more accurate approximation to the exact confidence interval based on a non-central t-distribution (Lehmann, 1986, p. 352) than McKay's method. This paper presents new methods for estimating the confidence intervals for the coefficient of variation by using standard bootstrap, percentile bootstrap and jackknife methods to estimate the sample CV's. We have tested the new methods using Monte Carlo methods to generate samples from normal, chi-square and gamma distributions.

Keywords: Coefficient of Variation, McKay's approximation, Modified McKay's approximation, Standard Bootstrap Method, Percentile Bootstrap Method and Jackknife Method

1. Introduction

The coefficient of variation is usually used as a measure of precision for the dispersion of data sets and is also often used to compare numerical distributions measured on different scales. It has many practical applications, for example, in quality control, for estimating the variability of products produced by a manufacturing process.

The population coefficient of variation for a random variable with mean μ and variance σ^2 is defined by:

$$\kappa \equiv \frac{\sigma}{\mu} \tag{1}$$

Let X_i for $i = 1,2,...,n$ be an independent random sample selected from the population. The sample mean and standard deviation are then given by: $\bar{x} = \sum_{i=1}^{n} x_i / n$ and $s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}$.

An estimator of the population coefficient of variation κ is the sample coefficient of variation.

$$\hat{\kappa} \equiv \frac{s}{\bar{x}} \tag{2}$$

An approximate $100(1-\alpha)\%$ confidence interval for the population coefficient of variation is (see, e.g., Vangel, 1996):

$$A = \left\{ \frac{\hat{\kappa}}{\sqrt{t_1(\theta_1 \hat{\kappa}^2 + 1) - \hat{\kappa}^2}}, \frac{\hat{\kappa}}{\sqrt{t_2(\theta_2 \hat{\kappa}^2 + 1) - \hat{\kappa}^2}} \right\} \tag{3}$$

where $\nu = n-1$, $t_1 \equiv \chi^2_{\nu,1-\alpha/2} / \nu$, $t_2 \equiv \chi^2_{\nu,\alpha/2} / \nu$ and $\theta = \theta(\nu, \alpha)$ is a known function selected so that a random variable $W_\nu = \frac{Y_\nu}{\nu}$, where Y_ν has a χ^2_ν distribution, has approximately the same distribution as a pivotal quantity $Q \equiv \frac{K^2(1+\kappa^2)}{(1+\theta K^2)\kappa^2}$. This pivotal quantity can be used to construct hypothesis tests and confidence intervals for κ.

การประมาณช่วงความเชื่อมั่นของสัดส่วนประชากร

Confidence Intervals for Proportion

ไพรวุฒิ อชินีทองคำ* ศรพรหม สิริวัฒนเศรษฐ์ ปรัชญา กาญจนพิบูลย์ ทองคำ ไม้กลัด วิจิตรา พลเยี่ยม
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: nonbug56@hotmail.com *

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการประมาณช่วงความเชื่อมั่นของสัดส่วนประชากรด้วยวิธีการประมาณช่วงค่า 5 วิธี วิธีปกติ วิธีสคอร์ วิธีเอฟ วิธีพัวซอง และวิธีปรับค่าของวาลด์ ใช้วิธีจำลองข้อมูลทวินาม ที่มีพารามิเตอร์ขนาดตัวอย่าง (n) และความน่าจะเป็นที่จะประสบความสำเร็จ (π) ต่างๆ แต่ละสถานการณ์ ทำซ้ำ 5,000 รอบ ผลการวิจัยพบว่า ที่ระดับความเชื่อมั่น 95% เมื่อขนาดตัวอย่างใหญ่ และ π อยู่ระหว่าง 0.1-0.9 วิธีปกติและวิธีสคอร์ให้ค่าสัมประสิทธิ์ความเชื่อมั่น 0.95 ตามที่กำหนด และวิธีสคอร์มีแนวโน้มที่จะให้ความยาวเฉลี่ยของช่วงค่าที่สั้นกว่าวิธีปกติ

คำสำคัญ: ช่วงเชื่อมั่น, สัดส่วนประชากร

Abstract

The objective of this research is to compare confidence interval estimation methods for the population proportion. The five estimation methods under consideration are Normal method, Score method, F method, Poisson method , and Adjusted wald method. The experimental data are generated by the simulation technique with 5,000 repetitions for each sample size (n) and each population proportion (π). The conclusion of this research is that for large sample size and population proportion between 0.1 and 0.9 the 95 % confidence intervals of the Normal and Scores methods give coverages of 95%. The trend of Score method gave average length shorter than the Normal method.

Keywords: confidence interval, population proportion

1. บทนำ

ให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n จากการแจกแจงเบอร์นูลี (Bernoulli distribution) ที่ไม่ทราบค่าพารามิเตอร์ π เมื่อให้ $Y = \sum_{i=1}^{n} X_i$ จะได้ว่าค่าประมาณภาวะความควรจะเป็นสูงสุดซึ่งเป็นค่าประมาณที่ไม่เอนเอียง (Unbiased estimate) และมีความแนบใน (Consistent estimate)ของ π ก็คือ $p = y/n$ เมื่อ y แทนจำนวนความสำเร็จที่พบจากการทดลอง n ครั้ง เรียก p ว่าเป็นสัดส่วนตัวอย่าง การประมาณค่าสัดส่วนของประชากรจึงเป็นการประมาณพารามิเตอร์ π ของการแจกแจงทวินาม ในที่นี้ผู้วิจัยสนใจศึกษาถึงการประมาณค่าแบบช่วงของสัดส่วนประชากร (π) โดยเปรียบเทียบวิธีประมาณช่วงเชื่อมั่นของ π ด้วยวิธีที่แตกต่างกัน

การหาช่วงความเชื่อมั่นของ π มีผู้ศึกษามาอย่างต่อเนื่อง เริ่มแรกในปี ค.ศ. 1934 Clopper และ Pearson ใช้แนวทางการย้อนกลับจากการทดสอบทวินาม (Binomial test) ของสมมติฐาน Ho : $\pi = \pi_0$ แย้งกับ Ha : $\pi \neq \pi_0$ เรียก

The efficiency of statistic methodology for distribution testing in RFID reader fields

Kongtharp, Prateep. [1], Glannamtip, Pathom. [2]
Department of Applied Statistics Faculty of Applied Statistics, King Mongkut University of technology North Bangkok
Email: go_ma09@hotmail.com, Pathom@kmutnb.ac.th

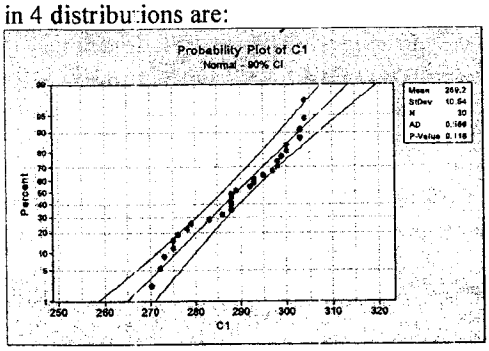
Abstract

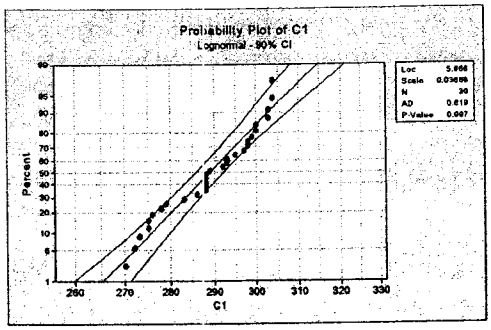
This research has aimed to find out the highest efficiency of goodness of fit statistical methodology in data distribution test. The Probability value and confidence bounds had used for comparison in each data distribution. The most hypothesis regarding the distributional form for accepted or rejected in that hypothesis. Particularly Kolmogorov-Smirnov and Anderson-Darling test have to specific the data distribution before testing. The critical value for hypothesis testing has accepted in many data distributions from in the same group of data. Researcher had put the Confidence Bounds to apply with p-value for Ultra High Frequency (UHF) Radio Frequency Identification (RFID) data of logistic environment. The Kolmogorov -Smirnov test and the Anderson-Darling test have used for distribution hypothesis testing and Confidence Bounds for the Normal Distribution, Lognormal Distribution, Weibull Distribution and Generalized Gamma Distribution had applied for find out the optimization methodology in this situation. Minitab package and R programming were the analysis tools. The result has shown that confidence bounds gave higher efficiency than p-value but not at all. The discussion has explained in the full paper.

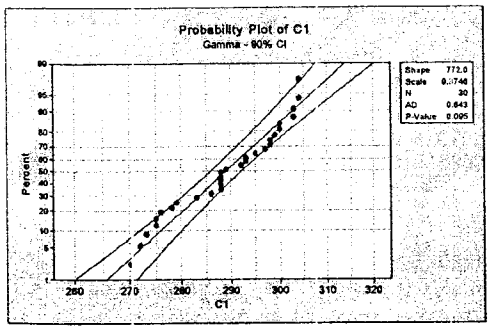
Keywords: Efficiency methodology, RFID reader fields, Distribution test

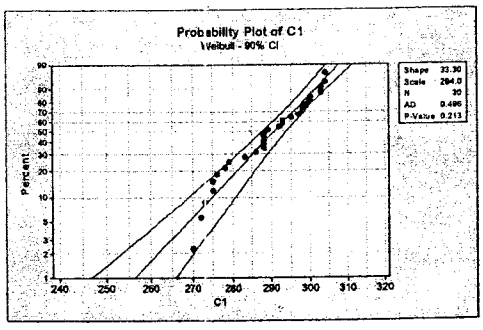
1. Results

By the Minitab package had found that p-value are accepted hypothesis with Anderson-Darling Test in 4 distributions are:









Then we had used R programming for Kolmogorov-Smirnov at the same situation with Anderson-Darling and found that; Almost distributions had the same result with P-value excepted Wiebull distribution CI and P-value had served in the same way and had significant value than other distribution like these;

การเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าขีดจำกัดควบคุมของแผนภูมิควบคุมค่าเฉลี่ย ($\bar{X}$ - Chart) ระหว่างวิธีบูตสแตรป และวิธีมาตรฐาน

A Comparison of Efficiency for Control Limit of Shewhart Control Chart by Estimated Method between Bootstrap Method and Standard Method

สรารัตน์ จารุภูมิ*, วรารัตน์ จันทร์ทอง, อัจฉา แดงจันทร์ และ สุวิมล พันธ์แย้ม
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: mim_2914@hotmail.com*

บทคัดย่อ

งานวิจัยนี้ ทำการศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าขีดจำกัดควบคุมของแผนภูมิควบคุมค่าเฉลี่ย ($\bar{X}$ - Chart) โดยอาศัยเทคนิคการสุ่มตัวอย่างซ้ำ วิธีบูตสแตรป (Bootstrap Method) และวิธีมาตรฐาน (Standard Method) ศึกษาในกรณีที่กระบวนการมีลักษณะการแจกแจงแบบปกติและการแจกแจงแบบที ซึ่งการเปรียบเทียบประสิทธิภาพการประมาณค่าขีดจำกัดควบคุมของแผนภูมิควบคุมค่าเฉลี่ย ($\bar{X}$ - Chart) ของทั้ง 2 วิธีนั้น จะพิจารณาจากค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean Squared Error: MSE) ของแต่ละวิธีเป็นเกณฑ์ในการประเมินประสิทธิภาพ ในการศึกษาจะใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) ในการจำลองข้อมูลเพื่อศึกษาในสถานการณ์ต่าง ๆ ผลการวิจัยพบว่าการประมาณค่าขีดจำกัดควบคุมของแผนภูมิควบคุมค่าเฉลี่ย ($\bar{X}$ - Chart) โดยวิธีบูตสแตรป (Bootstrap Method) มีประสิทธิภาพดีกว่าวิธีมาตรฐาน (Standard Method) เนื่องจากให้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean Squared Error : MSE) ต่ำกว่าวิธีมาตรฐาน (Standard Method)

คำสำคัญ: $\bar{X}$ - Chart , Bootstrap Method, Standard Method, Mean Squared Error

1. บทนำ

ในกระบวนการผลิตสินค้าทางด้านอุตสาหกรรม ผู้ผลิตจะต้องผลิตสินค้าที่มีคุณภาพและตรงตามความต้องการของผู้บริโภค ซึ่งเป็นปัจจัยสำคัญของการกำหนดมาตรฐานของสินค้า ดังนั้นเทคนิคหรือวิธีการที่เกี่ยวกับการควบคุมคุณภาพของสินค้าจึงมีความจำเป็นอย่างยิ่งที่จะต้องนำมาใช้เพื่อให้กระบวนการในการผลิตมีความสามารถมากขึ้น โดยผู้ผลิตจะต้องเลือกวิธีการให้เหมาะสมกับคุณสมบัติของสินค้า จึงจะทำให้การควบคุมคุณภาพนั้นประสบผลสำเร็จในกระบวนการผลิต เนื่องจากในกระบวนการผลิตย่อมมีความแปรปรวนเกิดขึ้นได้ตลอดเวลา จึงส่งผลกระทบถึงคุณภาพของสินค้าที่ทำการผลิต ดังนั้นจึงนำวิธีการควบคุมกระบวนการทางสถิติ (Statistical Process Control : SPC) มาใช้เพื่อช่วยในการเตือนให้ทราบถึงการเปลี่ยนแปลงที่เกิดขึ้น หากตรวจสอบความผิดปกติของกระบวนการผลิตได้อย่างรวดเร็วก็จะหาสาเหตุของความผิดปกติได้ทันท่วงที เพื่อที่จะแก้ไขและกำจัดปัญหาที่เกิดขึ้นให้หมดไป

แผนภูมิควบคุมค่าเฉลี่ยเป็นเครื่องมือทางสถิติอย่างหนึ่งที่ถูกนำมาใช้เพื่อควบคุมค่าเฉลี่ยของกระบวนการผลิต ทั้งนี้การสร้างแผนภูมิควบคุมค่าเฉลี่ยโดยทั่วไปจะสร้างจากข้อมูลที่ถูกเก็บรวบรวมมาจากกระบวนการผลิต และทำการสร้างแผนภูมิควบคุมค่าเฉลี่ย โดยวิธีมาตรฐาน (Standard Method) ที่ใช้กันในปัจจุบัน

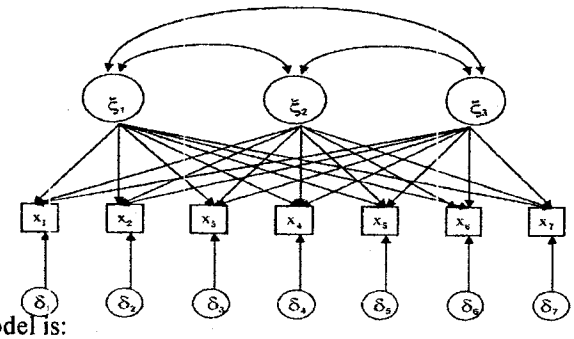
Bradley Efron (1979) ได้เสนอวิธีการประมาณค่าพารามิเตอร์ โดยวิธีบูตสแตรป (Bootstrap Method) ซึ่งมีหลักการสำคัญคือการสุ่มตัวอย่างซ้ำตามจำนวนครั้งที่กำหนด และในการสุ่มตัวอย่างแต่ละครั้งจะทำการสุ่มตัวอย่างที่มีขนาดเท่ากับขนาดตัวอย่างที่ต้องการ โดยทำการสุ่มตัวอย่างแบบใส่คืน (With Replacement)

The Efficiency Test of Factors Analysis model from statistical package between Correlation and Structural Equation Models

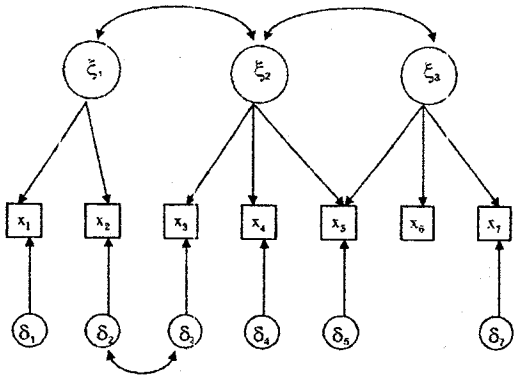
Thaweewan Phoyen*, Natcha Inklub**, Pathom Glannamtip***
Department of Applied Statistics, Faculty of Applied Sciences, King Mongkut University of Technology North Bangkok
Email: chompoo_neung@hotmail.com*, gibpy_kmitnb@hotmail.com**,
Pathom@kmutnb.ac.th***

Abstract

This paper has objected to comparison the efficiency of factor analysis methodology. The correlation model and structural equation model were considered to test the fitted methodology in each experimental design. The correlation factor analysis model is;



And structure equation model is:



The principle components analysis (PCA), unweighted least square (ULS), Generated least square (GLS), Maximum likelihood (ML) and Principal axis factoring (PAF) were the extraction method. The Quartimax, Oblimin Direct, Equamax, Promax, and Varimax for rotated factor methodology had used for factors loading. We also apply Normal, Chi square and Exponential data distributions had generated by R programming for applied in all situations. The one by one experimental design had applied for compared between correlation factor analysis and structural equation model. The MSE is the comparison tester. SPSS and Lisrel statistical package have applied for computational tools in this research. The results had shown that; the factor analysis methodology between correlation and structure equation model have grouping factors at the same number but parameter coefficient are not the same. The structure equation model has more efficiency than correlation model. The autocorrelation between factors are the main problem of factor analysis model. Conclusions, remarks and hints have discussed for optimal testing in the whole factors analysis are given.

Keywords: Optimal testing, Factors analysis, Correlation model, Structural equation model

The performance of time series model testing in free form financial data

Pussara, Kanyapat and Glannamtip, Pathom
Department of applied statistics, King Mongkut's University of Technology North Bangkok, Thailand
E-mail: mangporbin@hotmail.com, ptg@kmitnb.ac.th.

Abstract

This paper has objected to generate the high performance of Time series model in free form data distribution of financial. Distributions of error found by the goodness of fit test had rejected hypothesis at .05 in any methodology. This research was used SARIMA and GARCH methods to find out the performance of seasonal time series and smoothing model in financial field. SARIMA model is; $\phi_p(B)\Phi_P(B^s)(I-B)^d(I-B^s)^D X_t = \theta_q(B)\Theta_Q(B^S)Z_t$ and GARCH model is;

Average equations

$$y = a + by_{t-1} + u_t$$

$$u_t = z\sqrt{h_t}$$

$$z_t \sim i.i.d.N(0,1)$$

Variance equations

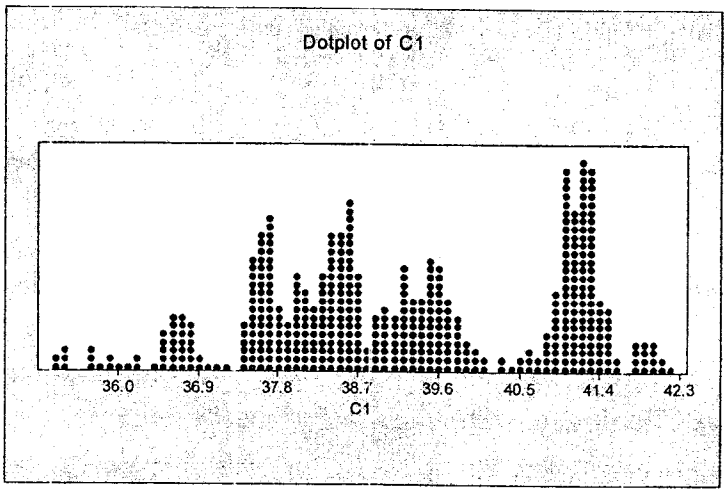
$$h_t = c + e_1u_{t-1}^2 + e_2u_{t-2}^2 + \ldots + e_qu_{t-q}^2 + f_1h_{t-1} + f_2h_{t-2} + \ldots + f_ph_{t-p}$$

The result was shown that SARIMA and GARCH had given the difference performance time series model in some conditions. The solution was described by comparison the performance model with graphs and mathematical equations.

Keywords: Performance model, Seasonal time series model, Free form distribution,

1. Results

Data testing by Dot plot



Basic Statistics on Mobile

Mitchatharn Chotrasee , Chantira Eamnoi and Sa-aat Niwitpong
Department of Applied Statistics, King Mongkut's University of Technology North Bangkok
Email: ultranan_2527@hotmail.com

บทคัดย่อ

ค่าสถิติพื้นฐานบนโทรศัพท์มือถือ เป็นการพัฒนาขึ้นเพื่อมุ่งหวังให้มีเครื่องมือที่ช่วยอำนวยความสะดวกในการคำนวณค่าสถิติพื้นฐาน โดยมีหลักการทำงานแบ่งออกเป็น 5 ส่วน ดังนี้ ส่วนที่ 1) Data เป็นส่วนของการจัดการกับข้อมูล ส่วนที่ 2) Basic Statistics เป็นส่วนของการคำนวณค่าสถิติพื้นฐาน ซึ่งส่วนที่ 3) Correlation สัมประสิทธิ์สหสัมพันธ์ ส่วนที่ 4) Simple Linear Regression การวิเคราะห์การถดถอยอย่างง่าย และส่วนที่ 5) Statistic formula สูตรคำนวณค่าสถิติพื้นฐาน

1. บทนำ

ปัจจุบันโปรแกรมประยุกต์ทางสถิติ เป็นโปรแกรมสำเร็จรูปที่ใช้กันอย่างกว้างขวาง การวิเคราะห์ข้อมูลทางสถิติมีการคำนวณที่ยุ่งยากซับซ้อน ทั้งนี้โปรแกรมสำเร็จรูปทางด้านสถิติ มักจะใช้อยู่บนเครื่องคอมพิวเตอร์ เมื่อโทรศัพท์มือถือเข้ามามีบทบาทกับชีวิตประจำวันของคนทั่วไปมาก ซึ่งเป็นอุปกรณ์สื่อสารที่มีประสิทธิภาพ ทั้งยังมีการพัฒนาขึ้นอย่างรวดเร็วและต่อเนื่อง ประกอบกับราคาของโทรศัพท์มือถือลดต่ำลง ในขณะที่มีประสิทธิภาพดีขึ้น ทำให้เกิดความนิยมในการใช้โทรศัพท์มือถือ เพื่อการติดต่อสื่อสารมากขึ้น จึงมีความประสงค์ที่จะพัฒนาเทคโนโลยีโทรศัพท์มือถือผนวกกับความต้องการของผู้ใช้งานทางด้านสถิติ เพื่อจัดทำการคำนวณค่าทางสถิติที่มีฟังก์ชันการทำงานเบื้องต้นบนโทรศัพท์มือถือ เพื่ออำนวยความสะดวกแก่ผู้ใช้งาน

2. วิธีการดำเนินงานวิจัย

ค่าสถิติพื้นฐานที่สามารถคำนวณ มีดังนี้

1. ขนาดของตัวอย่าง, ค่าผลรวม, ค่าผลรวมกำลังสอง
2. ค่าเฉลี่ยของตัวอย่าง, มัธยฐาน, ฐานนิยม
3. ค่าผลรวมของผลต่าง, ความแปรปรวนของตัวอย่าง
4. ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
5. ค่าสูงสุดของข้อมูล, ค่าต่ำสุดของข้อมูล
6. พิสัย, ค่าพิสัยระหว่างควอร์ไทล์
7. สัมประสิทธิ์การแปรผันของตัวอย่าง, ค่าคลาดเคลื่อนมาตรฐาน
8. ค่าสัมประสิทธิ์สหสัมพันธ์
9. การวิเคราะห์การถดถอยอย่างง่าย

การจัดอันดับบริษัทประกันภัยยานยนต์ภาคสมัครใจในประเทศไทยปี 2548 วัดผลประสิทธิภาพโดยวิธีการ DEA

The ranking of motor voluntary insurance companies in Thailand 2005 using DEA efficiency score

ศศิวิมล คงพรปรารถนา, ศิริกัญญา อิทธิพิบูล, วิชัย สุรเชิดเกียรติ
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: untiga_t@hotmail.com,however_new@hotmail.com,wsc@kmutnb.ac.th
ภาศกร พันธุ์รอด, วรรณศิริ มิถอย
สำนักงานคณบดี คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อจัดอันดับบริษัทประกันภัยยานยนต์ภาคสมัครใจในประเทศไทยทั้ง 43 แห่ง ตามผลคะแนนประสิทธิภาพโดยวิธีการ DEA ผลในการวิจัยครั้งนี้กล่าวถึงการปรับปรุงค่าปัจจัยเพื่อให้บรรลุถึงความมีประสิทธิภาพในแต่ละบริษัทประกันภัยยานยนต์ที่มีประสิทธิภาพน้อยกว่า 100%

การวิจัยโดยใช้คะแนนประสิทธิภาพของตัวแบบ BCC ภายใต้สมมติฐาน VRS มี 3 ขั้นคือ 1) การเลือกตัวแปรตามแนวคิดของ Wagner and Shimshak (2006) จนได้ตัวแปรปัจจัยผลผลิต 2 ปัจจัยคือ เบี้ยประกันภัยถือเป็นรายได้ และเบี้ยประกันภัยรับตรง ตัวแปรปัจจัยนำเข้า 2 ปัจจัย คือ ค่าใช้จ่ายรวมและจำนวนเงินเอาประกันภัย ที่มีผลต่อค่าคะแนนประสิทธิภาพมากที่สุด 2) ผลคะแนนประสิทธิภาพจากการใช้ตัวแปรปัจจัยที่ถูกเลือกมีบริษัทที่มีคะแนนประสิทธิภาพเต็ม จำนวน 18 บริษัทส่วนบริษัทที่มีคะแนนประสิทธิภาพไม่เต็มต้องทำการปรับปรุงตัวแปรปัจจัยโดยมี 2 ทางคือ บริษัทที่ควรเพิ่มปัจจัยผลผลิตเพียงอย่างเดียวมี 22 บริษัท และบริษัทที่ควรทั้งลดปัจจัยนำเข้าและทั้งเพิ่มปัจจัยผลผลิตมี 3 บริษัท 3) การจัดอันดับตามแนวคิดของ Jahanshahloo et al. (2007) นำมาจัดอันดับได้ 40 อันดับ จากทั้งหมด 43 บริษัท

คำสำคัญ: การประกันภัย, Data envelopment analysis, Motor voluntary insurance

Abstract

The objective of this research is to evaluate and rank forty-three motor voluntary insurance companies in Thailand. The result of efficiency evaluated from the method of DEA reveals the improvements on influence factors for each company, with less than 100 % efficiency, to accomplish the maximum achievable efficiency.

The study is carried out using the efficiency score from the BCC model under the assumption of VRS, which has three steps as follow: 1) Two outputs factors comprises of earning and direct insurance premium and two inputs which are the total expenses and the amount of sum insured, that are the four main influence factors on the efficiency score are obtained by following the concept of Wagner and Shimshak (2006). 2) The efficiency score achieved by the application of these factors, demonstrates that eighteen companies proves to have their full efficiency and the rest should improve on their factors by two means: adjusting inputs and outputs factors. The results shows that twenty-two companies should increase their outputs and three should both increase their outputs and lowers their inputs. 3) The forty-three companies are ranked according to Jahanshahloo et al. (2007) and forty companies rankings were obtained.

Keywords: Insurance, Data envelopment analysis (DEA), Motor voluntary insurance

การปรับตัวแบบการตอบสนองการสุ่มโดยวิธีสุ่มซ้ำ

Improvement of Randomized Response Model by Re-sampling Method

ภาศกร พันธุ์รอด, วิชัย สุรเชิดเกียรติ
ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: phoenix.pk@hotmail.com, wsc@kmutnb.ac.th

บทคัดย่อ

แบบสอบถามที่เป็นคำถามปกปิด ผู้ตอบมักจะไม่ตอบตรงตามความจริง เราจึงแก้ปัญหานี้ด้วยการสร้างตัวแบบการสุ่มตัวอย่างอย่างง่ายสำหรับเทคนิคการตอบสนองการสุ่มขึ้นมา 2 ตัวแบบ คือ ตัวแบบทั่วไปและตัวแบบผสม ซึ่งในการสร้างตัวแบบนั้นเราอาศัยหลักการสุ่มซ้ำ โดยให้ผู้ตอบทุกคนต้องใช้เครื่องมือสุ่มเพื่อทำการสุ่มคำถามแบบใส่คืนเพื่อตอบคนละ 2 ครั้ง ในตัวแบบทั่วไปเครื่องมือสุ่มของเราประกอบด้วย 2 คำถาม คือ คำถามปกปิดที่ให้ตอบว่า "ใช่" หรือ "ไม่ใช่" และ คำถามเปล่าที่เมื่อสุ่มได้ให้ตอบว่า "ใช่" ส่วนในตัวแบบผสมจะทำการคัดแยกกลุ่มผู้ตอบด้วยคำถามทั่วไปก่อน เป็นกลุ่มผู้ที่ตอบว่า "ใช่" และกลุ่มผู้ที่ตอบว่า "ไม่ใช่" โดยแต่ละกลุ่มจะใช้เครื่องมือสุ่มที่ประกอบด้วย 2 คำถามเหมือนกัน คือ คำถามปกปิดที่ให้ตอบว่า "ใช่" หรือ "ไม่ใช่" และ คำถามเปล่าที่เมื่อสุ่มได้ให้ตอบว่า "ใช่" ซึ่งวิธีการของเราให้ประสิทธิภาพดีกว่าในหลายๆ ตัวแบบ

คำสำคัญ: เครื่องมือสุ่ม, การตอบสนองการสุ่ม, การสุ่มซ้ำ, ตัวประมาณ, ความแปรปรวน

Abstract

In sensitive questions usually obtain the responses that not direct to the real situation . So , we solve this problem by making simple random sampling two models for randomized response technique these are general model and mixed model. To design these models, researcher employs re-sampling principle and respondents have to use random device for choosing a random with replacement questions and each respondent has to answer two times. Random devices used in general model consists of two questions are sensitive question and blank question. If sensitive question is chosen, respondent must answer "Yes" or "No" and if blank question is chosen, respondent must answer "Yes". Mixed model uses direct questions for categorized group of respondent. There are two groups of respondent, first group answers "Yes" and second group answers "No". Each group uses random devices consists of two questions are sensitive question that need "Yes" or "No" response and blank question that need "Yes" response. The result shows that both models have more efficiencies than other models.

Keywords: Random Device, Randomized Response, Re-sampling, Estimator, Variance

1. บทนำ

ในการทำวิจัยโดยการเก็บรวบรวมข้อมูลนั้น แบบสอบถามจึงถือเป็นเครื่องมือหนึ่งที่มีความสำคัญในการทำวิจัย การได้มาซึ่งคำตอบที่ไม่ตรงกับความเป็นจริงมีผลกระทบต่อผลการวิจัย ปัญหานี้มักเกิดกับการวิจัยที่เกี่ยวข้องกับคำถามที่มีผลต่อความรู้สึกของบุคคล ซึ่งเรียกคำถามลักษณะนี้ว่า คำถามปกปิด (Sensitive Question) เช่น คำถามที่มีผล