

การศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในฟังก์ชันการแจกแจง

แบบปกติที่ค่าไคสแควร์จากการทดสอบภาวะสารูปสัมพัทธ์มีค่าต่ำสุด

Angelo ๖๒๒ ๑๐๙๑๐๘๕:



นางสาว วิไลพร อธรรมเนียมอินทร์

004067

ศูนย์วิทยบรังษยกร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทศึกษาศาสตรมหาบัณฑิต

จุฬาลงกรณ์มหาวิทยาลัย
ภาควิชาสังคม

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. ๒๕๖๔

A comparative study on the methods of estimating
parameters of the Normal distribution giving minimal
chisquare (χ^2) value in the goodness of fit test

Miss Wilaiporn Thumniumintr

A thesis Submitted in Partial Fulfilment of the Requirements

for the Degree of Master of Science

Department of Statistics

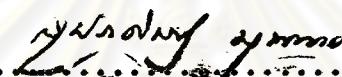
Graduate School

Chulalongkorn University

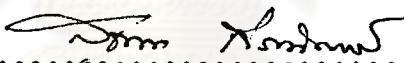
1981

หัวข้อวิทยานิพนธ์	การศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในฟังก์ชันการแจกแจงแบบปกติที่คำไคลสแควร์จากการทดสอบภาวะสาระปัลนิทีมีค่าด้วย
โดย	นางสาววิไลพร ธรรมเมฆอินทร์
ภาควิชา	สถิติ
อาจารย์ที่ปรึกษา	อาจารย์กฤณา เพ็ชรรัตน์
คร. ชว.ลิต	ทิลยากร

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของ
การศึกษาตามหลักสูตรปริญญามหาบัณฑิต

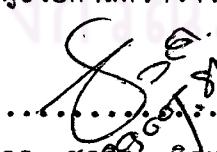
.......... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร. สุประดิษฐ์ บุนนาค)

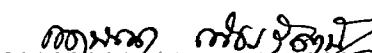
คณะกรรมการสอบวิทยานิพนธ์

.......... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. อุชาดา กีระนันทน์)

.......... กรรมการ
(รองศาสตราจารย์ ดร. สรชัย พิศาลบุตร)

.......... กรรมการ
(ผู้ช่วยศาสตราจารย์ มัลลิกา บุนนาค)

.......... กรรมการ
(ดร. ชาลิต ทิลยากร)

.......... กรรมการ
(อาจารย์กฤณา เพ็ชรรัตน์)

สิชลิกอธิบดีของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

หัวข้อวิทยานิพนธ์	การศึกษาเบรี่ยบเทียบเพื่อบริการประมาณค่าพารามิเตอร์ในฟังก์ชันการแจกแจงแบบปกติที่คำนวณแปรผันจากการทดสอบภาวะสารภูมิที่มีค่าต่ำสุด
ชื่อผู้ลิตร	นางสาววิไลพร ธรรม เมียนอินทร์
อาจารย์ที่ปรึกษา	อาจารย์กฤษณา เพ็ชรรัตน์ ดร. ชาลิต ทิสยากร
ภาควิชา	สถิติ
ปีการศึกษา	๒๕๖๔



บทสรย่อ

ในการศึกษาลักษณะของข้อมูลที่มีการแจกแจงแบบปกติ (Normal Distribution) นั้น เราจะเริ่มต้นด้วยการสร้างกราฟเมทริกแกนของข้อมูล เพื่อถูกัดลักษณะการแจกแจงของข้อมูลโดยคร่าว ๆ เสียก่อน และคำนวณหาค่าพารามิเตอร์ของข้อมูลโดยใช้สูตร $\mu = \frac{\sum x_i}{N}$ และ $\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$ จากนี้จะคำนวณหาค่า χ^2 จากสูตร $\chi^2 = \sum_{i=1}^K \frac{(O_i - E_i)^2}{E_i}$ ซึ่ง O_i เป็นค่าความถี่ในช่วงที่ i ของข้อมูล (Observed Frequency) E_i เป็นค่าคาดหวังของข้อมูล (Expected Frequency) และ $E_i = N \int_{i-1}^{i+1} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

ขั้นตอนการคำนวณพารามิเตอร์ของข้อมูลทั้งกล่าวข้างบนนี้เป็นวิธีมาตรฐานที่ใช้กันโดยแพร่หลาย (Conventional Method) ซึ่งเรียกว่าโดยทั่วไปว่า "การทดสอบภาวะสารภูมิที่ดี" (Goodness of fit test) และเป็นที่ทราบกันดีว่า เมื่อค่า χ^2 ที่คำนวณได้มีค่าต่ำยิ่งตัวมีความหมายความถี่ว่า O_i จะเข้าใกล้ E_i มากที่สุดด้วย และโอกาสที่ค่า χ^2 ที่คำนวณได้มีจะต่ำกว่าค่า χ^2 ที่เปิดจากตารางย่อมจะมีมากขึ้นด้วย นั่นคือ จะช่วยลดความผิดพลาดแบบที่ ๑ (Type-I-Error) ให้น้อยลงได้

วิทยานิพนธ์ฉบับนี้แสดงถึงการใช้ Univariate Search (เป็น Optimization Technique หรือที่รู้จักกันเป็นที่แพร่หลาย) คำนวณหาค่าประมาณของพารามิเตอร์ \bar{X}^* และ S^* ที่จะให้ค่า χ^2 ต่ำสุดได้ และจากการทดลองใช้ Univariate Search กับข้อมูลที่มีการแจกแจงแบบปกติ (Normal Distribution) ซึ่งสูมศ้าอย่างมากจำนวน ๘๐ ชุด นั้น เรายพบว่าจำนวนความผิดพลาดแบบที่ ๑ (Type-I-Error) ได้ลดลงตามท้องการ

เป็นที่น่าสนใจว่า χ^2 (\bar{X}^* , S^*) คำสูตรที่คำนวณได้นี้ จะมีผลกระทบต่อคุณภาพของประชากรหรือไม่ ซึ่งวิทยานิพนธ์ฉบับนี้ได้ใช้ t-test ทำการทดสอบหาความสัมพันธ์ระหว่าง ศักยภาพค่าของพารามิเตอร์ (\bar{X}^* , S^*) และศักยภาพพารามิเตอร์ (μ, σ) ของประชากร เนื่อง ปรากฏผลไม่แตกต่างกันอย่างมีนัยสำคัญ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Thesis Title A comparative study on the methods of estimating
 parameters of the Normal distribution giving minimal
 chisquare (χ^2) value in the goodness of fit test
 Name Wilaiporn Thumnumint
 Thesis Advisor Krissana Petcharatana
 Dr. Chavalit Thisayakorn
 Department Statistics
 Academic Year 1981

ABSTRACT

In conducting a study on Normal Distribution Data, we normally begin by drafting a rough histogram of the data. Then, we proceed to compute important parameters of the data utilizing formula $\mu = \frac{1}{N} \sum_{i=1}^N X_i / N$ and $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2}$. From here on, we also compute χ^2 from $\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$ where O_i represents Observed Frequency of the ith. interval, E_i is the Expected Frequency and $E_i = N \int_{\frac{i-1}{2}\sigma}^{\frac{i+1}{2}\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$.

The above conventional Procedures are employed to compute parameters of the Normal Distribution Data under studied and is called "Goodness of Fit Test". It is well known that this conventional χ^2 indicates how O_i approades E_i , which also implies reducing Type-I-Error. In other words, lower χ^2 value is always preferred.

This thesis showed how to use a popular optimization technique named "Univariate Search" in computing \bar{X}^* and S^* which yields minimized χ^2 . About 80 sets Computer Generated Normal Distribution Data has been tested by Univariated Search Technique. Type-I-Error is found to decrease as desired.

By using Student t-test, research results indicate no significant deviation between estimated (\bar{X}^*, S^*) and (μ, σ) of the assumed main population.

ศูนย์วิทยบรังษยการ
จุฬาลงกรณ์มหาวิทยาลัย

กิติกรรมประการ



ผู้เชียนขอทราบขอบพระคุณ อาจารย์กฤณา เพ็ชรรัตน์ ที่กรุณาเป็นอาจารย์ที่ปรึกษา
ขอขอบพระคุณเป็นอย่างยิ่งที่ ดร. ชาลิต ทิสยากร กฤณาเป็นอาจารย์ที่ปรึกษาร่วมให้คำแนะนำ
และช่วยเหลือในการทำวิทยานิพนธ์เล่มนี้มาโดยตลอด และผู้ช่วยศาสตราจารย์ พรรตัน ทิสยากร
ที่กรุณาได้ให้ความช่วยเหลือในการให้ยืมเอกสารและตำราที่มีประโยชน์ในการค้นคว้าแก่ผู้เชียน
จนเป็นผลสำเร็จ

วไลพร ธรรมเนียมอินทร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทศดย์อักษรไทย	๗
บทศดย์อักษรอังกฤษ	๘
กิจกรรมประภาก	๙
รายการตารางประภกอน	๑๐
รายการรูปประภกอน	๑๑
บทที่ ๑	
๑.๑ บทนำ	๑
๑.๒ ทีมงานปัญหา	๒
๑.๓ แนวความคิดการกำหนดค่าพารามิเตอร์และประสิทธิ	๕
๑.๔ วัสดุประสงค์ของการวิจัย	๑๙
๑.๕ ขอบเขตของการวิจัย	๒๒
๑.๖ ประโยชน์ที่คาดว่าจะได้รับ	๒๓
บทที่ ๒ ทฤษฎีที่ใช้ในการวิจัย	๒๔
๒.๑ การทดสอบแบบไชสแควร์ (Chi-square Test)	๒๔
๒.๒ การแจกแจงแบบปกติ (Normal Distribution)	๒๗
๒.๓ การทดสอบแบบสตีว์เด้นท์ (t-test)	๒๙
๒.๔ เทคนิคการค้นหาด้วยประเดียว (Univariate Search Technique)	๓๕
๒.๕ วิธีทิศทางสังยุคของพาวเวลล์	๓๐
๒.๖ ข้อผิดพลาดบางประการในวิธีทิศทางสังยุคของพาวเวลล์	๓๗



สารบัญ (ต่อ)

	หน้า
บทที่ ๓ เครื่องมือที่ใช้ในการวิจัย	๔๐
๓.๑ การหาค่าคาดหมายของจำนวนข้อมูล	๔๐
๓.๒ การสร้างเลขสุ่ม	๔๐
๓.๓ การสุ่มตัวอย่างแบบมีระบบ	๔๗
๓.๔ โปรแกรมที่ใช้งานทั้งหมด	๕๖
บทที่ ๔ ขั้นตอนการวิจัยและผลการวิจัย	๕๗
๔.๑ การเตรียมข้อมูลสำหรับการวิจัย	๕๗
๔.๒ การตรวจสอบข้อมูลที่จัดเตรียมได้และผลการตรวจสอบ	๕๗
๔.๓ การคำนวณค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐาน (\bar{X}^* , S^*) ที่จะให้ค่าไคลแคร์ต้าสูด	๖๑
๔.๔ ผลการวิจัย	๗๕
๔.๕ ข้อสังเกต	๘๐
บทที่ ๕ สรุปผลการวิจัย	๘๙
๕.๑ สรุปผลการวิจัย	๘๙
๕.๒ ข้อเสนอแนะ	๙๒
บรรณานุกรม	๙๓
ภาคผนวก	๙๕
ประวัติผู้เขียน	๑๑๓

รายการตารางประกอบ

ตารางที่		หน้า
๔.๑	แสดงค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของข้อมูล ๒๐๐๐ ตัวที่ ต้องการเปรียบเทียบกับค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานที่ได้ จากโปรแกรม GAUSS	๔๖
๔.๒	แสดงค่าความถี่ของข้อมูลและค่าคาดหวังของข้อมูลในแต่ละช่วง	๖๗
๔.๓	แสดงค่าประมาณของพารามิเตอร์ในรอบต่าง ๆ ของการค้นหาค่าอุตมะ ด้วยวิธีการค้นหาตัวแปรเดียว ($R = 0.01$)	๖๘
๔.๔	แสดงค่าคาดหวังของความถี่ใหม่หลังจากใช้วิธีการค้นหาตัวแปรเดียว โดยใช้ $R = 0.01$	๗๙
๔.๕	ผลการทดสอบตัวประมาณค่าเฉลี่ยที่ได้จากการ Optimize ไลสแคร์ โดยใช้สถิติทดสอบ $t-test$	๘๔
๔.๖	เปรียบเทียบค่าสถิติทดสอบแบบไลสแคร์ที่ได้จากการ Conventional Method กับวิธี Optimization Method	๘๖
๔.๗	เปรียบเทียบผลการยอมรับและปฏิเสธสมมติฐานของค่าไลสแคร์ที่ได้ จาก ๒ วิธี ในระดับความมั่นยืนลำดับระดับต่าง ๆ	๘๗
๔.๘	เปรียบเทียบค่าเฉลี่ยที่ได้จากการ ๒ วิธี	๘๘

รายการรูปประกอบ

รูปที่		หน้า
๑.๑	ผังแสดงขอบเขตคุณสมบัติของวิธีการค้นหาค่าอุตม化(Optimization Technique) ประเภท Multidimensional-sequential search method	๖๐
๒.๑	แสดงลักษณะข้อดีของข้อมูลที่ได้จากการวิเคราะห์ข้อมูล	๗๕
๒.๒	แสดงโถงปักติ $n(x ; m, \sigma^2)$	๗๖
๒.๓	แสดงโถงปักติที่มี m_1, m_2 และ $\sigma_1 < \sigma_2$	๗๘
๒.๔	แสดงรูปการค้นหาตัวแปรที่ลับซ้อนในทิศทางแต่ละทิศทาง	๗๙
๒.๕	แสดงให้เห็นช่องการหาค่าของตัวแปรที่ทำให้พังก์ชันต่ำสุด	๘๘
๒.๖	แสดงให้เห็นช่องการหาค่าของตัวแปรที่ทำให้พังก์ชันต่ำสุด	๙๙
๓.๑	แสดงการหาพื้นที่โดยการใช้พื้นที่สี่เหลี่ยมผืนผ้า ๑ ส่วน	๔๔
๓.๒	แสดงการหาพื้นที่โดยการใช้พื้นที่สี่เหลี่ยมผืนผ้า m ส่วน	๔๔
๓.๓	แสดงการหาพื้นที่โดยใช้กราฟเส้นโถง ๒ ส่วน	๔๔
๓.๔	แสดงการหาพื้นที่โดยใช้กราฟเส้นโถง ๓ ส่วน	๔๗
๓.๕	ผังภาพแสดงขั้นตอนดำเนินการคำนวณของโปรแกรมในการวิจัยทั้งหมด	๕๕
๔.๑	แสดงข้อมูลของเลขสุ่ม ๒๐๐๐ ตัว	๖๐
๔.๒	แสดงข้อมูลของข้อมูล ๒๐๐ ตัวที่ใช้โปรแกรม HIST	๖๕
๔.๓	แสดงข้อมูลของข้อมูลที่ได้จากการวิจัย UNISER	๖๖
๔.๔	แสดงการใช้ขั้นตอนดำเนินการคำนวณของ R ในระดับเดียว ๑ พร้อมทั้งแสดงจำนวน iterate ในแต่ละระดับของ R ใน การค้นหาค่าไคสแควร์ต่ำสุด	๖๙
๔.๕	แสดงข้อมูลของข้อมูลที่ได้จากการวิจัย UNISER	๗๓
๔.๖	แสดงค่าเฉลี่ยและค่าความเปี่ยงเบนมาตรฐานที่ได้จาก ๒ วิธี โดยใช้ตัวอย่างขนาด ๒๐๐ ตัว จำนวน ๑๐ ครั้ง	๗๙