

การศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในฟังก์ชันการแจกแจง

แบบปกติที่ค่าโคสแควร์จากการทดสอบภาวะสภาวะปกติมีค่าต่ำสุด

: AS 622 ๑๐๗๐๕๖



นางสาว วิไลพร ธรรมเนียมอินทร์

004067


ศษญ์วิทยทรพยากร  
วิทยานิพนธ์นี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาสถิติ  
จุฬาลงกรณ์มหาวิทยาลัย

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. ๒๕๖๔

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



Miss Wilaiporn Thumnumintr

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 . การศึกษา เปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในฟังก์ชันการแจกแจงแบบปกติที่ค่าโคสแควร์จากการทดสอบภาวะสารูปสนิทที่มีค่าต่ำสุด

ชื่อนิสิต นางสาววิไลพร ธรรมเนียมอินทร์

อาจารย์ที่ปรึกษา อาจารย์กฤษณา เพ็ชรรัตน์

ดร. ขวลิต ทิสยากร

ภาควิชา สถิติ

ปีการศึกษา ๒๕๒๔



### บทคัดย่อ

ในการศึกษาลักษณะของข้อมูลที่มีการแจกแจงแบบปกติ (Normal Distribution)

นั้น เรามักจะเริ่มต้นด้วยการสร้างกราฟฮิสโตแกรมของข้อมูล เพื่อดูลักษณะการแจกแจงของ

ข้อมูลโดยคร่าว ๆ เสียก่อน แล้วคำนวณหาค่าพารามิเตอร์ของข้อมูลโดยใช้สูตร  $\mu = \frac{\sum_{i=1}^N x_i}{N}$

และ  $\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$  จากนั้นจึงจะคำนวณหาค่า  $\chi^2$  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  $E_i = N \int_i^{i+1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

ขั้นตอนการคำนวณหาพารามิเตอร์ของข้อมูลดังกล่าวข้างบนนี้เป็นวิธีมาตรฐานที่ใช้กันโดยแพร่หลาย (Conventional Method) ซึ่งเรียกกันโดยทั่วไปว่า "การทดสอบภาวะสารูปสนิท" (Goodness of fit test) และเป็นที่ยอมรับกันว่าเมื่อค่า  $\chi^2$  ที่คำนวณได้นี้มีค่าอย่างยิ่งดี เพราะหมายความว่า  $O_i$  จะเข้าใกล้  $E_i$  มากที่สุดด้วย และโอกาสที่ค่า  $\chi^2$  ที่คำนวณได้นี้จะต่ำกว่าค่า  $\chi^2$  ที่เปิดจากตารางย่อมจะมีมากขึ้นด้วย นั่นคือ จะช่วยลดความผิดพลาดแบบที่ ๑ (Type-I-Error) ให้น้อยลงได้

วิทยานิพนธ์ฉบับนี้แสดงถึงการใช้ Univariate Search (เป็น Optimization Technique วิธีหนึ่งที่อยู่ร่วมกันเป็นที่แพร่หลาย) คำนวณหาค่าประมาณของพารามิเตอร์  $\bar{X}^*$  และ  $S^*$  ที่จะให้ค่า  $\chi^2$  ต่ำสุดได้ และจากการทดลองใช้ Univariate Search กับข้อมูลที่มีการแจกแจงแบบปกติ (Normal Distribution) ซึ่งสุ่มตัวอย่างมาจำนวน ๘๐ ชุด นั้น เราพบว่าจำนวนความผิดพลาดแบบที่ ๑ (Type-I-Error) ได้ลดลงตามต้องการ

เป็นที่น่าสนใจว่า  $\chi^2$  ( $\bar{X}^*$ ,  $S^*$ ) ต่ำสุดที่คำนวณได้นี้ จะมีผลกระทบต่อคุณภาพของประชากรหรือไม่ ซึ่งวิทยานิพนธ์ฉบับนี้ได้ใช้ t-test ทำการทดสอบหาความสัมพันธ์ระหว่างตัวประมาณค่าของพารามิเตอร์ ( $\bar{X}^*$ ,  $S^*$ ) และตัวพารามิเตอร์ ( $\mu, \sigma$ ) ของประชากรเดิม ปรากฏผลไม่แตกต่างกันอย่างมีนัยสำคัญ



ศูนย์วิจัยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

Thesis Title: A comparative study on the methods of estimating parameters of the Normal distribution giving minimal chisquare ( $\chi^2$ ) 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  $\mu = \frac{\sum_{i=1}^N X_i}{N}$  and  $\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$ . From here on, we also compute  $\chi^2$  from  $\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$

where  $O_i$  represents Observed Frequency of the  $i$ th. interval,  $E_i$  is the Expected Frequency and  $E_i = N \int_{i-1}^{i+1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

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  $\chi^2$  indicates how  $O_i$  appropades  $E_i$ , which also implies reducing Type-I-Error. In other words, lower  $\chi^2$  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  $\chi^2$ . About 80 sets Computer Generated Normal Distribution Data has been tested by Univariate 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  $(\mu, \sigma)$  of the assumed main population.



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

กิตติกรรมประกาศ



ผู้เขียนขอกราบขอบพระคุณ อาจารย์กฤษณา เพ็ชรรัตน์ ที่กรุณาเป็นอาจารย์ที่ปรึกษา  
ขอขอบพระคุณเป็นอย่างยิ่งที่ ดร. ขวสิต ทิสยากร กรุณาเป็นอาจารย์ที่ปรึกษาร่วมให้คำแนะนำ  
และช่วยเหลือในการทำวิทยานิพนธ์เล่มนี้มาโดยตลอด และผู้ช่วยศาสตราจารย์ นพรัตน์ ทิสยากร  
ที่กรุณาได้ให้ความช่วยเหลือในการให้สัมภาษณ์เอกสารและตำราที่มีประโยชน์ในการค้นคว้าแก่ผู้เขียน  
จนเป็นผลสำเร็จ

วิไลพร ธรรมเนียมอินทร์

ศูนย์วิจัยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญ

	หน้า
บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	ฉ
กิตติกรรมประกาศ .....	ช
รายการตารางประกอบ .....	ฉ
รายการรูปประกอบ .....	ฉ
<b>บทที่ ๑</b>	
๑.๑ บทนำ .....	๑
๑.๒ ที่มาของปัญหา .....	๒
๑.๓ แนวความคิดการกำหนดค่าพารามิเตอร์และประวัติ .....	๕
๑.๔ วัตถุประสงค์ของการวิจัย .....	๑๑
๑.๕ ขอบเขตของการวิจัย .....	๑๒
๑.๖ ประโยชน์ที่คาดว่าจะได้รับ .....	๑๓
<b>บทที่ ๒</b>	
๒ ทฤษฎีที่ใช้ในการวิจัย .....	๑๔
๒.๑ การทดสอบแบบไคสแควร์ (Chi-square Test) .....	๑๔
๒.๒ การแจกแจงแบบปกติ (Normal Distribution) .....	๑๗
๒.๓ การทดสอบแบบสตีวเด็นท์ (t-test) .....	๒๑
๒.๔ เทคนิคการค้นหาตัวแปรเดียว (Univariate Search Technique) .....	๒๕
๒.๕ วิธีทิศทางสังยุคของพาวเวลล์ .....	๓๐
๒.๖ ข้อผิดพลาดบางประการในวิธีทิศทางสังยุคของพาวเวลล์ .....	๓๗



## สารบัญ (ต่อ)

	หน้า
บทที่ ๓ เครื่องมือที่ใช้ในการวิจัย .....	๔๐
๓.๑ การหาค่าคาดหมายของจำนวนข้อมูล .....	๔๐
๓.๒ การสร้างเลขสุ่ม .....	๕๐
๓.๓ การสุ่มตัวอย่างแบบมีระบบ .....	๕๓
๓.๔ โปรแกรมที่ใช้งานทั้งหมด .....	๕๖
บทที่ ๔ ขั้นตอนการวิจัยและผลการวิจัย .....	๕๗
๔.๑ การเตรียมข้อมูลสำหรับการวิจัย .....	๕๗
๔.๒ การตรวจสอบข้อมูลที่จัดเตรียมได้และผลการตรวจสอบ .....	๕๗
๔.๓ การคำนวณค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐาน ( $\bar{X}^*$ , $S^*$ ) ที่จะให้ค่าไคสแควร์ต่ำสุด .....	๖๒
๔.๔ ผลการวิจัย .....	๗๕
๔.๕ ข้อสังเกต .....	๘๐
บทที่ ๕ สรุปผลการวิจัย .....	๘๑
๕.๑ สรุปผลการวิจัย .....	๘๑
๕.๒ ข้อเสนอแนะ .....	๘๒
บรรณานุกรม .....	๘๓
ภาคผนวก .....	๘๕
ประวัติผู้เขียน .....	๑๓๓

## รายการตารางประกอบ

ตารางที่		หน้า
๔.๑	แสดงค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของข้อมูล ๒๐๐๐ ตัวที่ต้องการเปรียบเทียบกับค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานที่ได้จากโปรแกรม GAUSS .....	๔๘
๔.๒	แสดงค่าความถี่ของข้อมูลและค่าคาดหวังของข้อมูลในแต่ละช่วง .....	๖๓
๔.๓	แสดงค่าประมาณของพารามิเตอร์ในรอบต่าง ๆ ของการค้นหาค่าที่เหมาะสมด้วยวิธีการค้นหาตัวแปรเดียว ( $R = 0.01$ ) .....	๖๗
๔.๔	แสดงค่าคาดหวังของความถี่ใหม่หลังจากใช้วิธีการค้นหาตัวแปรเดียว	๗๑
๔.๕	สรุปผลภาพหลังจากการใช้การค้นหาตัวแปรเดียวกันและสุ่ม ๒๐๐ ตัว โดยใช้ $R = 0.01$ .....	๗๒
๔.๖	ผลการทดสอบตัวประมาณค่าเฉลี่ยที่ได้จากการ Optimize โคสแควร์ โดยใช้สถิติทดสอบ t-test .....	๗๔
๔.๗	เปรียบเทียบค่าสถิติทดสอบแบบโคสแควร์ที่ได้จากวิธี Conventional Method กับวิธี Optimization Method .....	๗๖
๔.๘	เปรียบเทียบผลการยอมรับและปฏิเสธสมมติฐานของค่าโคสแควร์ที่ได้จาก ๒ วิธี ในระดับความมีนัยสำคัญระดับต่าง ๆ .....	๗๗
๔.๙	เปรียบเทียบค่าเฉลี่ยที่ได้จาก ๒ วิธี .....	๗๘

รายการรูปประกอบ

รูปที่	หน้า
๑.๑	ผังแสดงขอบเขตคุณสมบัติของวิธีการค้นหาค่า optimum (Optimization Technique) ประเภท Multidimensional-sequential search method ..... ๑๐
๒.๑	แสดงลักษณะฮิสโตแกรมที่ได้จากความถี่ของข้อมูล ..... ๑๕
๒.๒	แสดงโค้งปกติ $n(x ; \mu , \sigma^2)$ ..... ๑๘
๒.๓	แสดงโค้งปกติที่มี $\mu_1 , \mu_2$ และ $\sigma_1 < \sigma_2$ ..... ๑๘
๒.๔	แสดงรูปการค้นหาตัวแปรทีละตัวในทิศทางแต่ละทิศทาง ..... ๒๖
๒.๕	แสดงให้เห็นช่องการหาค่าของตัวแปรที่ทำให้ฟังก์ชันต่ำสุด ..... ๒๘
๒.๖	แสดงให้เห็นช่องการหาค่าของตัวแปรที่ทำให้ฟังก์ชันต่ำสุด ..... ๒๘
๓.๑	แสดงการหาพื้นที่โดยการใช้พื้นที่สี่เหลี่ยมผืนผ้า ๑ ส่วน ..... ๔๔
๓.๒	แสดงการหาพื้นที่โดยการใช้พื้นที่สี่เหลี่ยมผืนผ้า $n$ ส่วน ..... ๔๔
๓.๓	แสดงการหาพื้นที่โดยใช้กราฟเส้นโค้ง ๒ ส่วน ..... ๔๕
๓.๔	แสดงการหาพื้นที่โดยใช้กราฟเส้นโค้ง ๓ ส่วน ..... ๔๗
๓.๕	ผังภาพแสดงขั้นตอนต่าง ๆ ในการคำนวณของโปรแกรมในการวิจัยทั้งหมด ..... ๕๕
๔.๑	แสดงฮิสโตแกรมของเลขสุ่ม ๒๐๐๐ ตัว ..... ๖๐
๔.๒	แสดงฮิสโตแกรมของข้อมูล ๒๐๐ ตัวที่ใช้โปรแกรม HIST ..... ๖๕
๔.๓	แสดงฮิสโตแกรมที่ได้จากค่าคาดหวังของความถี่ก่อนใช้ UNISER ..... ๖๖
๔.๔	แสดงการใช้ช่องกว้าง (R) ในระดับต่าง ๆ พร้อมทั้งแสดงจำนวน iterate ในแต่ละระดับของ R ในการค้นหาค่าโคสแควร์ต่ำสุด ..... ๖๘
๔.๕	แสดงฮิสโตแกรมที่ได้จากค่าคาดหวังของความถี่ใหม่ที่ได้จากการใช้ UNISER ..... ๗๓
๔.๖	แสดงค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานที่ได้จาก ๒ วิธี โดยใช้ตัวอย่างขนาด ๒๐๐ ตัว จำนวน ๑๐ ชุด ..... ๗๔