การสุ่มตัวอย่างแบบชั้นภูมิโดยไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ

นายสุวัฒน์ ปลื้มอารมณ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

STRATIFIED SAMPLING WITH UNKNOWN POPULATION SIZE IN EACH STRATUM

Mr. Suwat Pleumarom

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2015

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การสุ่มตัวอย่างแบบชั้นภูมิโดยไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ
โดย	นายสุวัฒน์ ปลื้มอารมณ์
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร. พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์)

.....กรรมการ

(อาจารย์ ดร. นัท กุลวานิช)

.....กรรมการภายนอกมหาวิทยาลัย

(อาจารย์ ดร. อรุณี กำลัง)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สุวัฒน์ ปลื้มอารมณ์ : การสุ่มตัวอย่างแบบชั้นภูมิโดยไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ (STRATIFIED SAMPLING WITH UNKNOWN POPULATION SIZE IN EACH STRATUM) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. เสกสรร เกียรติสุไพบูลย์, 49 หน้า.

วิทยานิพนธ์ฉบับนี้นำเสนอวิธีการใหม่เพื่อเป็นทางเลือกในการประมาณค่าเฉลี่ยประชากรสำหรับการสุ่มตัวอย่างแบบชั้นภูมิเมื่อไม่ทราบขนาดของประชากรในแต่ละชั้นภูมิด้วยวิธีเบย์ งานวิจัยนี้พิจารณาเฉพาะกรณีที่ประชากรมีสองชั้นภูมิโดยมีชั้นภูมิใดชั้นภูมิหนึ่งมีค่าสังเกตเป็นบวก และอีกชั้นภูมิหนึ่งมีค่าสังเกตเป็นศูนย์ ระเบียบวิธีชั้นภูมิเบย์ที่นำเสนอจะประมาณค่าเฉลี่ยประชากรด้วยค่าคาดหวังภายหลังของค่าเฉลี่ยถ่วงน้ำหนักระหว่างค่าเฉลี่ยตัวอย่างจากสองชั้นภูมิ โดยกำหนดให้การแจกแจงก่อนเป็นการแจกแจงแบบสม่ำเสมอ และฟังก์ชันความควรจะเป็นเป็นฟังก์ชันแจกแจงแบบไฮเปอร์จีออเมตริก และค่าประมาณค่าเฉลี่ยประชากรคำนวณด้วยวิธีเชิงตัวเลข จากการประเมินประสิทธิภาพของวิธีชั้นภูมิเบย์เปรียบเทียบกับการสุ่มตัวอย่างแบบง่ายโดยพิจารณาจากความคลาดเคลื่อนมาตรฐานเป็นเกณฑ์ พบว่าการประมาณค่าแบบชั้นภูมิเบย์จะมีประสิทธิภาพดีกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายเมื่อชั้นภูมิทั้งสองมีความแตกต่างกันอย่างชัดเจน ซึ่งอาจเกิดขึ้นเมื่อค่าเฉลี่ยประชากรของชั้นภูมิที่มีค่าสังเกตเป็นบวกมีค่าสูง หรือค่าความแปรปรวนประชากรของชั้นภูมิที่มีค่าสังเกตเป็นบวกมีค่าต่ำ ผู้วิจัยประยุกต์ระเบียบวิธีดังกล่าวกับข้อมูลค่าใช้จ่ายด้านการวิจัยและพัฒนาของประเทศไทยประจำปี 2556 ซึ่งสำรวจโดยสำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ เทคโนโลยีและนวัตกรรมแห่งชาติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2558

5681606626 : MAJOR STATISTICS
KEYWORDS: SAMPLING / STRATIFIED SAMPLING / UNKNOWN POPULATION SIZE

SUWAT PLEUMAROM: STRATIFIED SAMPLING WITH UNKNOWN POPULATION SIZE IN EACH STRATUM. ADVISOR: ASSOC. PROF. SEKSAN KIATSUPAIBUL, Ph.D., 49 pp.

This thesis introduces a new approach to estimate population mean for stratified sampling with unknown strata sizes, based on Bayesian estimation. We focus on the case of two strata when one stratum contains positive observed values and the other stratum contains zero observed values. The proposed Bayesian stratified sampling approach estimates the population mean by the posterior expected value of the weighted average between the sample means from the two strata. The prior distribution is a uniform distribution and the likelihood function is a hypergeometric distribution. The estimate is evaluated by a numerical procedure. Employing standard error as the performance measure, we compare the performance of the proposed method with that of the simple random sampling approach. From the experimental results, we find that the Bayesian stratified sampling approach offers a better performance when the two strata are clearly distinguished from each other. This condition can be met either by that the population mean of the positive-valued stratum is of high value or by that the population variance of the positive-valued stratum is of low value. We apply the proposed methodology to the data set of Thailand's 2013 national research and development expenditures collected by National Science Technology and Innovation Policy Office.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department: Statistics Student's Signature .....
Field of Study: Statistics Advisor's Signature .....
Academic Year: 2015

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้คงไม่อาจสำเร็จสมบูรณ์เช่นนี้ หากปราศจากความเมตตากรุณาจากท่านอาจารย์ รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์ที่ทำให้ผู้เขียนได้มีหัวข้อในการทำวิทยานิพนธ์ ตลอดจนได้ให้ข้อมูลและคำแนะนำต่างๆ โดยเฉพาะการวางเค้าโครง แนวทางการเขียนเนื้อหาและบทวิเคราะห์ ตลอดจนการกำหนดกรอบเวลาในการเสนอความคืบหน้าของงาน ซึ่งถือเป็นแรงกระตุ้นให้แก่ผู้เขียนได้อย่างดียิ่ง ผู้เขียนรู้สึกซาบซึ้งใจและสำนึกในพระคุณของท่านอาจารย์เป็นอย่างยิ่งจึงขอกราบขอบพระคุณท่านอาจารย์ไว้ ณ ที่นี้

ผู้เขียนขอกราบขอบพระคุณ อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช ประธานกรรมการวิทยานิพนธ์ และอาจารย์ ดร. นัท กุลวานิช กรรมการวิทยานิพนธ์ ที่ท่านได้กรุณาชี้แนะแนวทางและคำแนะนำต่างๆ และขอกราบขอบพระคุณคณาจารย์ต่างๆ ซึ่งผู้เขียนมิได้เอ่ยนามที่ได้อบรมสั่งสอนให้ความรู้ทางด้านวิชาการจนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้

ผู้เขียนขอกราบขอบพระคุณสำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ เทคโนโลยีและนวัตกรรมแห่งชาติ (สวทน.) ที่ได้เอื้อเฟื้อข้อมูลอันเป็นประโยชน์สำหรับใช้ในการเขียนประกอบวิทยานิพนธ์ฉบับนี้

ผู้เขียนขอขอบคุณกัลยาณมิตรของผู้เขียนทุกๆ ท่านที่ได้ให้ความช่วยเหลือและเป็นกำลังใจให้ผู้เขียนตลอดมา ขอขอบคุณน้อง ๆ ภาควิชาสถิติ ทุกท่านที่คอยช่วยเหลือตลอดระยะเวลาที่ศึกษา โดยเฉพาะอย่างยิ่งนางสาวศรัณยา สมทรง ผู้ซึ่งคอยให้ความช่วยเหลือและให้คำแนะนำแก่ผู้เขียนด้วยความเต็มใจยิ่งเสมอมา

สุดท้ายผู้เขียนขอกราบขอบพระคุณ คุณพ่อมานิตย์ และคุณแม่ศิริ ปลื้มอารมณ์ที่ท่านอบรมสั่งสอนให้ตระหนักถึงความสำคัญของการศึกษา ให้การสนับสนุนและเป็นกำลังใจสำคัญในการทำวิทยานิพนธ์ฉบับนี้จนสำเร็จลุล่วงลงได้ โดยหากวิทยานิพนธ์ฉบับนี้มีประโยชน์และคุณค่าทางการศึกษาอยู่บ้าง ผู้เขียนขอยกคุณงามความดีทั้งหมดแด่ท่านอาจารย์ รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์และกรรมการสอบวิทยานิพนธ์ทุกท่าน รวมทั้งกราบเป็นกตเวทิตาแก่บิดามารดา คณาจารย์และผู้มีพระคุณที่ได้อบรมเลี้ยงดู ให้ความรู้ ความเมตตาแก่ผู้เขียน แต่หากวิทยานิพนธ์ฉบับนี้มีความบกพร่องประการใด ผู้เขียนขอน้อมรับความผิดพลาดไว้แต่เพียงผู้เดียว

นายสุวัฒน์ ปลื้มอารมณ์

พ.ศ. 2559

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและความสำคัญของปัญหา 1
1.2 บริบทของปัญหา 5
1.3 วัตถุประสงค์ 5
1.4 สมมติฐานงานวิจัย 6
1.5 ขอบเขตของงานวิจัย 6
1.6 วิธีการดำเนินการวิจัย 6
1.7 ประโยชน์ที่คาดว่าจะได้รับ 7
บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง 8
2.1 การสุ่มตัวอย่างแบบชั้นภูมิ (stratified random sampling) 8
2.2 ทฤษฎีบทของเบย์ (Bayes' Theorem) 10
2.3 การแจกแจงแบบไฮเปอร์จีออเมตริก (Hypergeometric Distribution) 13
2.4 การแจกแจงก่อน (Prior Distribution) 13
บทที่ 3 วิธีการดำเนินการศึกษา 15
3.1 ขอบเขตของการศึกษา 15
3.2 ขั้นตอนการดำเนินการศึกษา 15

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

บทที่ 4 ผลการวิจัย 19
4.1 ผลการศึกษาตัวแบบของพารามิเตอร์ 19
4.2 ผลการศึกษาการเปรียบเทียบการประมาณค่าแบบชั้นภูมิเบย์กับการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย 23
บทที่ 5 อภิปรายผลและข้อเสนอแนะ 42
5.1 อภิปรายผล 42
5.2 ข้อเสนอแนะ 42
รายการอ้างอิง 43
ภาคผนวก 44
ประวัติผู้เขียนวิทยานิพนธ์ 48

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่ หน้า

1.1 ความคลาดเคลื่อนมาตรฐานของค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาโดยวิธีการสุ่มตัวอย่างแบบชั้นภูมิและการสุ่มตัวอย่างแบบง่ายในกรณีสมมติที่มีจำนวนบริษัทที่มีกิจกรรมวิจัยและพัฒนาแตกต่างกัน....2

4.2.1 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลสมมติที่สามารถพบได้ในสถานการณ์จริง.... 24

4.2.2 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 โดยรวมทุกขนาดกิจการ.... 31

4.2.3 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดเล็ก.... 34

4.2.4 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดกลาง.... 37

4.2.5 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดใหญ่.... 40

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

ภาพที่ หน้า

2.2.1 หลักการโดยสรุปของทฤษฎีบทของเบย์ (Bayes' Theorem) 10

2.2.2 แผนภาพแสดงเหตุการณ์ A และ B ที่อยู่ในแซมเปิลสเปซ S 11

4.2.1 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในกรณีที่มีขนาดตัวอย่างในชั้นภูมิที่ 1 แตกต่างกันจำนวน 35 กรณี.....25

4.2.2 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในกรณีที่มีค่าคาดหวังของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัทแตกต่างกันจำนวน 7 กรณี.....27

4.2.3 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในกรณีที่มีส่วนเบี่ยงเบน-ซมาตรฐานของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัทแตกต่างกันจำนวน 7 กรณี.....29

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการสำรวจวิจัยต้องมีการวางแผนการสุ่มตัวอย่างโดยเลือกแผนการสุ่มตัวอย่างที่เหมาะสมกับประชากรที่ต้องการศึกษา กรอบตัวอย่างที่มี งบประมาณ เวลาและกำลังคน สำหรับการสุ่มตัวอย่างแบบชั้นภูมิ (stratified random sampling) เป็นการสุ่มตัวอย่างที่เหมาะกับประชากรที่มีจำนวนมากและมีความแตกต่างกันโดยเราจะแบ่งประชากรเป็นกลุ่มย่อยหรือชั้นภูมิซึ่งในแต่ละชั้นภูมิมีสมาชิกที่มีลักษณะคล้ายคลึงกัน และระหว่างชั้นภูมิมีความแตกต่างกัน จากนั้นทำการสุ่มตัวอย่างในแต่ละชั้นภูมิอีกต่อหนึ่ง ข้อได้เปรียบของการสุ่มตัวอย่างแบบชั้นภูมิ คือ ความคลาดเคลื่อนมาตรฐานของตัวประมาณพารามิเตอร์ (ค่าเฉลี่ยของประชากร) จะลดลง อย่างไรก็ตามการประมาณพารามิเตอร์ต้องทราบขนาด หรือน้ำหนักของชั้นภูมิเป็นสำคัญ

การสุ่มตัวอย่างแบบง่าย (simple random sampling) ถือว่าทุกๆ หน่วย หรือทุกๆ สมาชิกในประชากรมีลักษณะเดียวกัน และมีโอกาสจะถูกเลือกเท่าๆ กันซึ่งจะละเลยคุณสมบัติของประชากรที่มีความแตกต่างกัน ซึ่งเป็นข้อได้เปรียบที่พบในการสุ่มตัวอย่างแบบชั้นภูมิ ดังตารางที่ 1.1 ซึ่งแสดงถึงความแตกต่างของค่าความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยประชากรระหว่างการสุ่มตัวอย่างแบบชั้นภูมิ และการสุ่มตัวอย่างแบบง่ายจากข้อมูลชุดเดียวกัน

ในตารางที่ 1.1 เป็นการพิจารณาสถานการณ์สมมติในการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาของบริษัทเอกชนที่เมื่อทำการสุ่มตัวอย่างในแต่ละประเภทกิจการสามารถแบ่งบริษัทออกเป็น 2 ชั้นภูมิ ได้แก่ บริษัทที่มีกิจกรรมวิจัยและพัฒนา และบริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนา โดยมีจำนวนบริษัทเอกชนทั้งหมด 50 บริษัท และสุ่มตัวอย่างบริษัทเอกชนได้ทั้งหมด 10 บริษัท ซึ่งพบว่าเป็นบริษัทที่มีค่าใช้จ่ายด้านการวิจัยและพัฒนา 2 บริษัท และบริษัทที่ไม่มีค่าใช้จ่ายด้านการวิจัยและพัฒนา 8 บริษัท โดยทั้ง 2 บริษัทดังกล่าวที่มีค่าใช้จ่ายด้านการวิจัยและพัฒนาได้เท่ากับ 1,020,000 บาท และ 980,000 บาท

กำหนดให้

N = จำนวนประชากร (บริษัททั้งหมด)

N_1 = จำนวนประชากรในชั้นภูมิที่ 1 (บริษัทที่มีกิจกรรมวิจัยและพัฒนา)

N_2 = จำนวนประชากรในชั้นภูมิที่ 2 (บริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนา)

m = จำนวนตัวอย่าง (ตัวอย่างบริษัททั้งหมด)

m_1 = จำนวนตัวอย่างในชั้นภูมิที่ 1 (บริษัทที่มีกิจกรรมวิจัยและพัฒนา)

m_2 = จำนวนตัวอย่างในชั้นภูมิที่ 2 (บริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนา)

ในวิทยานิพนธ์ฉบับนี้มีข้อกำหนดการใช้สัญลักษณ์ดังนี้ ตัวอักษรหนาจะใช้แทนตัวแปรสุ่ม และตัวอักษรปกติใช้แทนค่าคงที่หรือค่าสังเกตของตัวแปรสุ่มที่สอดคล้องกัน ยกตัวอย่างเช่น $\mathbf{N_1}$ เป็นตัวแปรสุ่มของขนาดประชากรในชั้นภูมิที่ 1 ซึ่งมี N_1 เป็นค่าสังเกตของตัวแปรสุ่มของขนาดประชากรในชั้นภูมิที่ 1 เป็นต้น

จากตัวอย่างข้างต้นและสัญลักษณ์ที่กำหนดไว้จะได้ว่า $N = 50$, $m = 10$, $m_1 = 2$ และ $m_2 = 8$ เมื่อพิจารณากรณีสมมติที่เกิดขึ้นทุกกรณี โดยในทุกกรณีดังกล่าวกรณีดังกล่าวเราทราบขนาดประชากรในชั้นภูมิที่ 1 เป็น N_1 และขนาดประชากรในชั้นภูมิที่ 2 เป็น $N_2 = N - N_1$ ผู้วิจัยสามารถหาความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาจากการสุ่มตัวอย่างแบบชั้นภูมิและการสุ่มตัวอย่างแบบง่ายด้วยสูตรพื้นฐานยกตัวอย่างเช่น สำหรับการสุ่มตัวอย่างแบบชั้นภูมิใช้สูตร 3.3.2′ หน้า 81 และสำหรับการสุ่มตัวอย่างแบบง่ายใช้สูตร 2.2.2 หน้า 41 จาก *Kish (*1965)

พบว่าการสุ่มตัวอย่างแบบชั้นภูมิซึ่งทราบจำนวนประชากรในแต่ละชั้นภูมิจะได้ความคลาดเคลื่อนมาตรฐานของค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาของการสุ่มตัวอย่างแบบชั้นภูมิจะมีค่าต่ำกว่าการสุ่มตัวอย่างแบบง่ายในทุกกรณี (รายละเอียดดังตารางที่ 1.1)

<u>ตารางที่ 1.1</u> ความคลาดเคลื่อนมาตรฐานของค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาโดยวิธีการสุ่มตัวอย่างแบบชั้นภูมิและการสุ่มตัวอย่างแบบง่ายในกรณีสมมติที่มีจำนวนบริษัทที่มีกิจกรรมวิจัยและพัฒนาแตกต่างกัน

รายละเอียด						**ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนา**	
N	m	m_1	m_2	N_1	N_2	**การสุ่มตัวอย่างแบบชั้นภูมิ**	**การสุ่มตัวอย่างแบบง่าย**
50	10	2	8	2	48	0	1,330,000
50	10	2	8	3	47	97.98	
50	10	2	8	6	44	277.13	
50	10	2	8	7	43	334.66	
50	10	2	8	8	42	391.92	
50	10	2	8	9	41	449.00	

ตารางที่ 1.1 (ต่อ) ความคลาดเคลื่อนมาตรฐานของค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาโดยวิธีการสุ่มตัวอย่างแบบชั้นภูมิและการสุ่มตัวอย่างแบบง่ายในกรณีสมมติที่มีจำนวนบริษัทที่มีกิจกรรมวิจัยและพัฒนาแตกต่างกัน

รายละเอียด						**ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนา**	
N	m	m_1	m_2	N_1	N_2	**การสุ่มตัวอย่างแบบชั้นภูมิ**	**การสุ่มตัวอย่างแบบง่าย**
50	10	2	8	10	40	505.96	1,330,000
50	10	2	8	11	39	562.85	
50	10	2	8	13	37	676.46	
50	10	2	8	14	36	733.21	
50	10	2	8	15	35	789.94	
50	10	2	8	16	34	846.64	
50	10	2	8	17	33	903.33	
50	10	2	8	19	31	1,016.66	
50	10	2	8	20	30	1,073.31	
50	10	2	8	21	29	1,129.96	
50	10	2	8	22	28	1,186.59	
50	10	2	8	23	27	1,243.22	
50	10	2	8	24	26	1,299.85	
50	10	2	8	25	25	1,356.47	
50	10	2	8	26	24	1,413.08	
50	10	2	8	27	23	1,469.69	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 1.1 (ต่อ) ความคลาดเคลื่อนมาตรฐานของค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาโดยวิธีการสุ่มตัวอย่างแบบชั้นภูมิและการสุ่มตัวอย่างแบบง่ายในกรณีสมมติที่มีจำนวนบริษัทที่มีกิจกรรมวิจัยและพัฒนาแตกต่างกัน

รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนา	
N	m	m_1	m_2	N_1	N_2	การสุ่มตัวอย่างแบบชั้นภูมิ	การสุ่มตัวอย่างแบบง่าย
50	10	2	8	28	22	1,526.30	1,330,000
50	10	2	8	29	21	1,582.91	
50	10	2	8	30	20	1,639.51	
50	10	2	8	32	18	1,752.71	
50	10	2	8	33	17	1,809.31	
50	10	2	8	34	16	1,865.90	
50	10	2	8	35	15	1,922.50	
50	10	2	8	36	14	1,979.09	
50	10	2	8	37	13	2,035.68	
50	10	2	8	38	12	2,092.27	
50	10	2	8	39	11	2,148.86	
50	10	2	8	40	10	2,205.45	

ในทางปฏิบัติเราอาจพบการสุ่มตัวอย่างแบบชั้นภูมิในกรณีที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ เช่น การประมาณค่าพารามิเตอร์ของค่าใช้จ่ายด้านการวิจัยและพัฒนาโดยไม่ทราบจำนวนบริษัทที่มีการวิจัยและพัฒนาทำให้ผู้วิเคราะห์ต้องพบกับปัญหาที่ไม่สามารถหาขนาด หรือน้ำหนักของชั้นภูมิได้ ซึ่งอาจแก้ปัญหาดังกล่าวด้วยวิธีการสุ่มตัวอย่างแบบ 2 ชั้น (double sampling for stratification) โดยหลังจากสุ่มตัวอย่างจากประชากรที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิแล้วนำค่าสังเกตของตัวอย่างดังกล่าวมาทำการแบ่งเป็นชั้นภูมิแล้วทำการประมาณค่าถ่วงน้ำหนักในแต่ละ

ชั้นภูมิก่อนจากนั้นจึงสุ่มตัวอย่างย่อย (subsample) ในแต่ละชั้นภูมิจากการสุ่มตัวอย่างในครั้งแรกอีกครั้ง (Okafor, 1994) ซึ่งวิธีนี้จะมีความยุ่งยากในการเลือกค่าสังเกตของตัวอย่างมาแบ่งเป็นชั้นภูมิและการประมาณค่าพารามิเตอร์ของตัวประมาณ (Najmussehar & Bari, 2002) ดังนั้นในงานวิจัยฉบับนี้ได้เสนอวิธีการแก้ปัญหาดังกล่าวอีกวิธีหนึ่งคือ การอนุมานทางสถิติแบบเบย์ (Bayesian inference)

การอนุมานทางสถิติแบบเบย์นั้นจะสามารถใช้ได้อย่างมีประสิทธิภาพเมื่อความแปรปรวนของตัวอย่างมีค่าต่ำโดย วิธีการอนุมานทางสถิติแบบเบย์ให้ผลดีในตัวอย่างที่มีความแปรปรวน และรากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (root mean squared error, RMSE) ต่ำ มีอัตราการครอบคลุมข้อมูล (coverage rate) สูง (Si, Pillai, & Gelman, 2015) และควรเป็นตัวอย่างที่ได้ผ่านการจัดระเบียบข้อมูลให้เสร็จสมบูรณ์แล้วอีกด้วย (Gelman, 2007)

งานวิจัยนี้ (1) นำเสนอวิธีการประมาณค่าเฉลี่ยประชากรและความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยประชากรแบบชั้นภูมิเบย์ เพื่อแก้ไขปัญหาการสุ่มตัวอย่างแบบชั้นภูมิโดยไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ โดยพิจารณาเฉพาะกรณีที่ประชากรมี 2 ชั้นภูมิ โดยมีชั้นภูมิใดชั้นภูมิหนึ่งมีค่าสังเกตเป็นบวก และอีกชั้นภูมิหนึ่งมีค่าสังเกตเป็นศูนย์ (2) ทำการวิเคราะห์ความไวของค่าประมาณพารามิเตอร์ตามตัวแบบเทียบกับตัวแปรต่างๆ (3) จากนั้นนำวิธีการที่ได้ไปประยุกต์ใช้งานกับข้อมูลจริงที่เป็นข้อมูลค่าใช้จ่ายค่าใช้จ่ายด้านการวิจัยและพัฒนาของบริษัทเอกชนในประเทศไทย

1.2 บริบทของปัญหา

ในการสุ่มตัวอย่างเมื่อไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ แต่ทราบขนาดตัวอย่างในชั้นภูมิซึ่งให้ข้อมูลเกี่ยวกับน้ำหนักของชั้นภูมิ จะทำการกำหนดระเบียบวิธีคำนวณค่าถ่วงน้ำหนักของชั้นภูมิให้เหมาะสมโดยอาศัยหลักการอนุมานทางสถิติแบบเบย์ (Bayesian inference)

1.3 วัตถุประสงค์

เพื่อกำหนดระเบียบวิธีในการประมาณค่าเฉลี่ยประชากรสำหรับการสุ่มตัวอย่างแบบชั้นภูมิในกรณีที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ

1.4 สมมติฐานงานวิจัย

ระเบียบวิธีที่กำหนดขึ้นจะให้การประมาณค่าเฉลี่ยประชากรที่มีความคลาดเคลื่อนมาตรฐานที่ต่ำกว่าการแก้ปัญหาโดยการสุ่มตัวอย่างแบบง่าย

1.5 ขอบเขตของงานวิจัย

1.5.1 พิจารณาการสุ่มตัวอย่างแบบชั้นภูมิเมื่อมี 2 ชั้นภูมิในกรณีที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ โดยที่ชั้นภูมิที่ 1 มีค่าสังเกตเป็นบวก และชั้นภูมิที่ 2 มีค่าสังเกตเป็นศูนย์

1.5.2 พารามิเตอร์ที่ต้องการประมาณ คือ ค่าเฉลี่ยของประชากร

1.5.3 งานวิจัยนี้จะทำการวิเคราะห์ทางสถิติเชิงอนุมานโดยอาศัยการอนุมานทางสถิติแบบเบย์ (Bayesian inference)

1.5.4 ประยุกต์ระเบียบวิธีกับกรณีศึกษาซึ่งเป็นข้อมูลที่ได้จากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนา

1.6 วิธีการดำเนินการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.6.1 ศึกษาและพัฒนาระเบียบวิธีในการประมาณค่าพารามิเตอร์ของประชากรสำหรับการสุ่มตัวอย่างแบบชั้นภูมิในกรณีที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ โดยสมมติค่าสังเกตของตัวอย่างและแบ่งชั้นภูมิของตัวอย่างที่เราศึกษาออกเป็น 2 ชั้นภูมิ คือ ชั้นภูมิที่ 1 มีค่าสังเกตเป็นบวก และชั้นภูมิที่ 2 มีค่าสังเกตเป็นศูนย์ ยกตัวอย่างเช่น การสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาสามารถแบ่งออกเป็น 2 ชั้นภูมิ คือ ชั้นภูมิที่ 1 เป็นบริษัทที่มีกิจกรรมวิจัยและพัฒนาซึ่งมีค่าสังเกตของค่าใช้จ่ายเป็นบวก และชั้นภูมิที่ 2 เป็นบริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนาซึ่งมีค่าสังเกตของค่าใช้จ่ายเป็นศูนย์

1.6.2 คำนวณขนาดประชากรที่เป็นไปได้ในแต่ละชั้นภูมิที่มีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ในข้อ 1.6.1

1.6.3 ใช้การอนุมานทางสถิติแบบเบย์ (Bayes & Price, 1763), (Berger, 1980) คำนวณหาการแจกแจงหลัง (posterior distribution) เมื่อการแจกแจงก่อน (prior distribution) เป็นการแจกแจงแบบสม่ำเสมอ (uniform distribution) ของพารามิเตอร์แทนประชากรในชั้นภูมิที่ 1

1.6.4 ทำการคำนวณค่าประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์เมื่อกำหนดว่าเรารู้ขนาดตัวอย่างในแต่ละชั้นภูมิ

1.6.5 ทำการคำนวณความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์เมื่อกำหนดว่าเรารู้ขนาดตัวอย่างในแต่ละชั้นภูมิ

1.6.6 นำระเบียบวิธีที่ได้มาใช้ในข้อมูลจริงเพื่อเป็นกรณีศึกษาของงานวิจัย

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ทำให้ทราบระเบียบวิธีใหม่ในการประมาณค่าเฉลี่ยประชากรของการสุ่มตัวอย่างแบบชั้นภูมิเมื่อไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ เฉพาะกรณี 2 ชั้นภูมิ โดยชั้นภูมิที่ 1 มีค่าสังเกตเป็นบวกและชั้นภูมิที่ 2 มีค่าสังเกตเป็นศูนย์ เพื่อนำไปวิเคราะห์ข้อมูลเชิงสถิติที่มีประโยชน์ต่อการปฏิบัติงานได้อย่างถูกต้องและเหมาะสม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

2.1 การสุ่มตัวอย่างแบบชั้นภูมิ (stratified random sampling)

เป็นการสุ่มตัวอย่างที่แบ่งประชากรออกเป็นชั้นภูมิ (subgroup or strata) บนพื้นฐานของตัวแปรที่สำคัญที่ส่งผลกระทบต่อตัวแปรตาม โดยมีหลักในการจัดแบ่งกลุ่มแต่ละชั้นภูมิมีความเป็นเอกพันธ์ (homogeneous) หรือกล่าวได้ว่าในชั้นภูมิเดียวกันจะมีลักษณะคล้ายคลึงกันตามกลุ่มย่อยของตัวแปรแต่จะมีความแตกต่างระหว่างชั้นภูมิ จำนวนสมาชิกในแต่ละชั้นภูมิจะถูกกำหนดให้เป็นสัดส่วน (proportion) ตามสัดส่วนที่ปรากฏในประชากร ซึ่งเรียกว่า การสุ่มแบบแบ่งชัดโดยใช้สัดส่วน (proportion stratified sampling) การสุ่มตัวอย่างแบบชั้นภูมิจะมีความเหมาะสมกับงานวิจัยที่สนใจความแตกต่างของลักษณะประชากรในระหว่างชั้นภูมิ มีขั้นตอนการดำเนินการ ดังนี้ (Kish, 1965)

1) ศึกษาลักษณะของประชากรที่จะศึกษาอย่างละเอียดว่าคุณลักษณะใดที่จะส่งผลต่อตัวแปรที่จะศึกษาตัวแปรใดบ้าง และคุณลักษณะนั้นๆ สามารถที่จำแนกออกเป็นกลุ่มย่อยได้หรือไม่ยกตัวอย่างเช่น เพศ ระดับการศึกษา และอาชีพ เป็นต้น

2) จำแนกประชากรออกเป็นชั้นภูมิตามคุณลักษณะของกลุ่มย่อยโดยกำหนดให้สมาชิกในแต่ละชั้นภูมิมีความคล้ายคลึงกันให้มากที่สุด และให้มีความแตกต่างระหว่างชั้นภูมิให้มากที่สุดเช่นเดียวกัน แต่ไม่ควรมีจำนวนชั้นมากเกินไปเพราะจะต้องใช้ตัวอย่างจำนวนมาก ซึ่งจะทำให้ลักษณะที่ศึกษามีความแปรปรวนสูง และไม่น่าเชื่อถือ

3) สุ่มตัวอย่างจากแต่ละชั้นภูมิ เพื่อเป็นตัวแทนของสมาชิกในแต่ละชั้นภูมิที่จะศึกษาตามสัดส่วน (proportional allocation) กล่าวคือ ชั้นใดมีประชากรมากควรได้รับการสุ่มตัวอย่างเป็นตัวแทนที่มากกว่า แต่ถ้าตัวอย่างแต่ละชั้นภูมิมีจำนวนที่แตกต่างกันมากควรคำนึงถึงเหตุผลเพื่อให้ได้จำนวนที่เหมาะสมและมีความครอบคลุมลักษณะประชากรที่ไม่จำเป็นต้องใช้สัดส่วนก็ได้ (disproportional allocation)

ข้อดีและข้อจำกัดของการสุ่มตัวอย่างแบบชั้นภูมิ

1) ทำให้ได้ตัวอย่างที่มีคุณลักษณะที่ครอบคลุมทุกลักษณะของประชากรอย่างเป็นระบบ และช่วยลดความแปรปรวนแต่ไม่ต้องลดขนาดของตัวอย่างเหมือนวิธีการสุ่มตัวอย่างแบบง่าย ทำให้การทดสอบทางสถิติมีประสิทธิภาพสูงขึ้น

2) ถ้าจำนวนตัวแปรที่ใช้มีมากเกินไปจะทำให้มีจำนวนชั้นภูมิที่มากและยุ่งยากในการแบ่งชั้น หรือทำให้สมาชิกของแต่ละชั้นภูมิอาจมีจำนวนน้อยไม่เพียงพอ และจะต้องเสียเวลาและใช้ค่าใช้จ่ายสูง

3) ในการประมาณค่าความแปรปรวนจะต้องใช้สูตรการปรับแก้สัดส่วนของตัวอย่างที่ค่อนข้างซับซ้อน

สมมติให้ประชากรถูกแบ่งออกเป็น L ชั้นภูมิ ให้ h แทนชั้นภูมิ

N_h = จำนวนประชากรในชั้นภูมิที่ h

m_h = จำนวนตัวอย่างในชั้นภูมิที่ h

Y_{hi} = ค่าสังเกตประชากรในชั้นภูมิที่ h เมื่อ i = 1, 2, ..., N_h

y_{hi} = ค่าสังเกตตัวอย่างในชั้นภูมิที่ h เมื่อ i = 1, 2, ..., m_h

$\bar{Y}_h$ = ค่าเฉลี่ยประชากรในชั้นภูมิที่ h

$$= \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}$$

$\bar{y}_h$ = ค่าเฉลี่ยตัวอย่างในชั้นภูมิที่ h

$$= \frac{1}{m_h}\sum_{i=1}^{m_h} y_{hi}$$

S_h^2 = ค่าความแปรปรวนของประชากรในชั้นภูมิที่ h

$$= \frac{\sum_{i=1}^{N_h}(Y_{hi} - \bar{Y}_h)^2}{N_h - 1}$$

s_h^2 = ค่าความแปรปรวนของตัวอย่างในชั้นภูมิที่ h

$$= \frac{\sum_{i=1}^{m_h}(y_{hi} - \bar{y}_h)^2}{m_h - 1}$$

f_h = สัดส่วนในการเลือกตัวอย่าง (sampling fraction) ของชั้นภูมิที่ h

$$= \frac{m_h}{N_h}$$

W_h = สัดส่วนของจำนวนประชากรในชั้นภูมิที่ h

$$= \frac{N_h}{N}$$

$\hat{\mu}$ = ตัวประมาณค่าเฉลี่ยประชากร

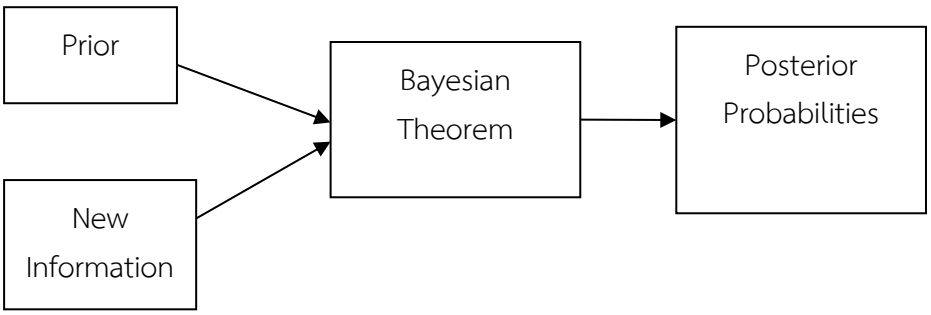
$$= \sum_{h=1}^{L}\left(\frac{N_h}{N}\right)\cdot\bar{y}_h$$

2.2 ทฤษฎีบทของเบย์ (Bayes' Theorem)

Beaumont, Zhang, and Balding (2002) ได้กล่าวถึงทฤษฎีบทของเบย์ไว้ว่าเป็นทฤษฎีที่สามารถนำมาประยุกต์ใช้กับการตัดสินใจแบบมีเงื่อนไขได้โดยมีหลักการว่า ความน่าจะเป็นที่สร้างขึ้นมาครั้งหลังเรียกว่า "posterior probabilities" โดยเป็นความน่าจะเป็นที่ไม่เป็นอิสระต่อกัน สามารถสรุปเป็นภาพที่ 2.2.1 ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาพที่ 2.2.1 หลักการโดยสรุปของทฤษฎีบทของเบย์ (Bayes' Theorem)



จากภาพที่ 2.2.1 ถ้า $A_1, A_2, \ldots, A_n$ เป็นลำดับของเหตุการณ์แต่ละเหตุการณ์ใดๆ ที่ไม่เกิดร่วมกัน และ B เป็นเหตุการณ์ที่เกิดขึ้นก่อนในทุกๆ เหตุการณ์ A (การรวมกันของสับเซตของ A_i) และ $P(B) > 0$ เพราะฉะนั้นจะได้ว่า

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

เมื่อ $P(B) = \sum_{i=1}^{n} P(B_i)\, P(A|B_i)$

นั้นคือ เมื่อกำหนดให้ $A_1, A_2, \ldots, A_n$ เป็นเหตุการณ์ใดๆ ในแซมเปิลสเปซ S ที่ไม่เกิดขึ้นร่วมกัน คือ

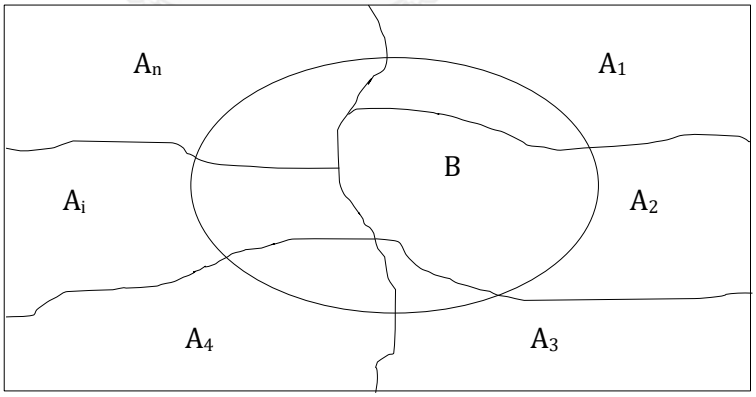
$$\bigcup_{i=1}^{n} A_i = S$$

และ

$$A_i \cap A_j = \emptyset \qquad \text{เมื่อ } i \neq j$$

และกำหนดให้ $B \subset S$ สามารถเขียนเป็นแผนภาพได้ดังนี้

<u>ภาพที่ 2.2.2</u> แผนภาพแสดงเหตุการณ์ A และ B ที่อยู่ในแซมเปิลสเปซ S



จากความน่าจะเป็นแบบมีเงื่อนไข (conditional probability)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

ดังนั้น

$$P(A \cap B) = P(A|B)P(B)$$

ทำนองเดียวกัน

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

และได้ว่า

$$P(A \cap B) = P(A|B)P(A)$$

จากภาพที่ 2.2.2 จะพบว่า

$$B = (A_1 \cup A_2 \cup \ldots \cup A_n) \cap B$$

$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)$$

ดังนั้น

$$P(B) = P(A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)$$

$$= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

$$= \sum_{i=1}^{N} P(B|A_i)P(A_i)$$

จากความน่าจะเป็นแบบมีเงื่อนไข

$$P(A_p|B) = \frac{P(A_p \cap B)}{P(B)}$$

เมื่อ A_p คือ เหตุการณ์ที่สนใจซึ่งเป็นสับเซตของเหตุการณ์ A (การรวมกันของสับเซตของ A_i) จะได้ทฤษฎีบทของเบย์ดังนี้

$$P(A_p|B) = \frac{P(B|A_p)P(A_p)}{\sum_{i=1}^{N} P(B|A_i)P(A_i)}$$

นอกจากนี้มีการนำทฤษฎีบทของเบย์มาประยุกต์ใช้ในการวิเคราะห์สถิติแบบเบย์อีกด้วย กล่าวคือ การพิจารณาการแจกแจงภายหลัง (posterior distribution) จะอาศัยข้อสนเทศจากตัวอย่างที่อยู่ในภาพแบบของฟังก์ชันความควรจะเป็น (likelihood function) และการแจกแจงก่อน (prior distribution) ซึ่งในการแจกแจงก่อน ผู้วิจัยอาศัยจากความรู้ ความเชื่อ และประสบการณ์ของผู้วิจัยและผู้เชี่ยวชาญที่มีต่อค่าพารามิเตอร์ ทั้งนี้ผู้วิจัยสามารถเลือกการแจกแจงก่อนที่เป็นคอนจูเกต (conjugate) กับการแจกแจงของค่าสนเทศ ซึ่งการแจกแจงคอนจูเกตจะมีภาพแบบฟังก์ชันที่คล้ายกับฟังก์ชันภาวะความควรจะเป็น ทำให้การคำนวณเพื่อหาฟังก์ชันภายหลังทำได้ง่าย และฟังก์ชันภายหลังที่ได้ก็จะเป็นการแจกแจงประเภทเดียวกับการแจกแจงก่อนที่กำหนดให้นั่นเองและจากการ

Chulalongkorn University

แจกแจงภาย หลังที่ได้ก็จะถูกนำไปวิเคราะห์เพื่อหาข้อสนเทศที่เกี่ยวข้องกับการแจกแจงภายหลังต่อไป เช่นการหาค่าคาดหวัง ค่าความแปรปรวน เป็นต้น จากการวิเคราะห์แบบเบย์อาศัยข้อมูลจากการสังเกตและจากการแจกแจงเบื้องต้นที่ผู้วิจัย เป็นผู้กำหนดจากการใช้ความเชื่อและประสบการณ์ในปัญหาที่เกี่ยวข้อง ถือว่าเป็นข้อดีอย่างหนึ่งของการดำเนินการแบบเบย์

2.3 การแจกแจงแบบไฮเปอร์จีออเมตริก (Hypergeometric Distribution)

จาก (Berkopec, 2007) ให้ $\mathbf{x}$ เป็นตัวแปรสุ่ม ซึ่งแทนจำนวนความสำเร็จที่ได้จากการสุ่มหยิบของจำนวน m สิ่งจากของ N สิ่งซึ่งประกอบด้วยของ 2 กลุ่มคือ กลุ่มหนึ่งมี N_1 สิ่ง และอีกกลุ่มหนึ่งมี
$\mathrm{N}_2 = \mathrm{N}\text{-}\mathrm{N}_1$ สิ่ง โดยการไม่แทนที่ (without replacement) จะเรียก $\mathbf{x}$ ว่า เป็นตัวแปรสุ่มแบบไฮเปอร์จีออเมตริก (hypergeometric RV.)

โดยฟังก์ชันความน่าจะเป็นของ m กำหนดให้ได้ ดังนี้

$$\mathrm{p}(\mathbf{x}) = \frac{\binom{\mathrm{N}_1}{\mathbf{x}}\binom{\mathrm{N}-\mathrm{N}_1}{\mathrm{m}-\mathbf{x}}}{\binom{\mathrm{N}}{\mathrm{m}}}$$

เมื่อ $\mathbf{x} < \mathrm{N}_1$, $\mathbf{x} < \mathrm{m}$ และ $\mathrm{m}\text{-}\mathbf{x} < \mathrm{N}\text{-}\mathrm{N}_1$ และการแจกแจงของความน่าจะเป็นของตัวแปรสุ่ม $\mathbf{x}$ นี้เราเรียกว่า การแจกแจงแบบไฮเปอร์จีออเมตริก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.4 การแจกแจงก่อน (Prior Distribution)

ในการลดค่าความคลาดเคลื่อนของพารามิเตอร์ให้ต่ำลงควรใช้ข้อมูลของพารามิเตอร์ที่เราเคยทราบมาก่อนมาช่วยในการประมาณค่า กล่าวคือ เราควรใช้ข้อมูลของพารามิเตอร์มาพิจารณาด้วย ซึ่งเป็นการอนุมานทางสถิติแบบเบย์ (Gelman, 2002) การอนุมานทางสถิติแบบเบย์มีส่วนประกอบสำคัญในการประมาณค่าพารามิเตอร์ 3 ส่วน คือ

1) ฟังก์ชันความควรจะเป็น (likelihood function) หรือข้อมูลปัจจุบัน เป็นข้อมูลที่ได้จากการสังเกต หรือทดลอง

2) การแจกแจงก่อน (prior distribution) หรือข้อมูลในอดีต เป็นข้อมูลที่เราทราบเพิ่มเติมเกี่ยวกับพารามิเตอร์ที่ต้องการประมาณค่าซึ่งเป็นตัวแปรสุ่มของการแจกแจง โดยนักสถิติบางคนอาจถือว่าการกำหนดความรู้เกี่ยวกับพารามิเตอร์ในอดีตเป็นความคิดแบบอัตวิสัย (subjective opinion) ซึ่งหมายถึงเป็นความคิดเฉพาะบุคคลเท่านั้น

3) การแจกแจงหลัง (posterior distribution) ข้อมูลในอนาคต เป็นข้อมูลที่เป็นผลจากการทราบข้อมูลในปัจจุบันและข้อมูลในอดีต โดยการแจกแจงหลังจะแปรผันกับผลคูณของฟังก์ชันความควรจะเป็นและการแจกแจงก่อน

การแจกแจงก่อนจะมีลักษณะการกระจายที่แตกต่างจากการแจกแจงหลัง โดยการแจกแจงก่อนจะแบ่งออกได้เป็น 2 แบบ คือ

1) การแจกแจงก่อนที่ให้ข้อมูล (informative prior distribution) เป็นการแจกแจงก่อนที่ให้ข้อมูลเกี่ยวกับตัวแปรสุ่มมาอย่างชัดเจน

2) การแจกแจงก่อนที่ไม่ให้ข้อมูล (noninformative prior distribution) เป็นการแจกแจงก่อนที่ให้ข้อมูลเกี่ยวกับตัวแปรสุ่มไม่ชัดเจน (vague) หรือให้ข้อมูลน้อยเมื่อเปรียบกับข้อมูลปัจจุบัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3

วิธีการดำเนินการศึกษา

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอวิธีการประมาณค่าเฉลี่ยประชากรและความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยประชากรแบบชั้นภูมิเบย์เพื่อแก้ไขปัญหาการสุ่มตัวอย่างแบบชั้นภูมิโดยไม่ทราบขนาดประชากรพร้อมทั้งเปรียบเทียบกับการสุ่มตัวอย่างแบบง่ายภายใต้ขอบเขตและวิธีการดำเนินงานดังนี้

3.1 ขอบเขตของการศึกษา

งานวิจัยนี้จะศึกษาภายใต้ขอบเขตดังนี้

3.1.1 พิจารณาเฉพาะกรณีที่ประชากรมี 2 ชั้นภูมิ โดยมีชั้นภูมิใดชั้นภูมิหนึ่งมีค่าสังเกตเป็นบวก และอีกชั้นภูมิหนึ่งมีค่าสังเกตเป็นศูนย์

3.1.2 พารามิเตอร์ที่ต้องการประมาณ คือ ค่าเฉลี่ยของประชากร

3.1.3 งานวิจัยนี้จะทำการวิเคราะห์ทางสถิติเชิงอนุมานโดยอาศัยการอนุมานทางสถิติแบบเบย์ (Bayesian inference)

3.1.4 ประยุกต์ระเบียบวิธีกับกรณีศึกษาซึ่งเป็นข้อมูลที่ได้จากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนา

3.2 ขั้นตอนการดำเนินการศึกษา

3.2.1 ศึกษาและพัฒนาระเบียบวิธีในการประมาณค่าพารามิเตอร์ของการสุ่มตัวอย่างแบบชั้นภูมิในกรณีที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ โดยสมมติข้อมูลค่าใช้จ่ายด้านการวิจัยและพัฒนา และแบ่งชั้นภูมิของประชากรที่เราศึกษาออกเป็น 2 ชั้นภูมิ คือ ชั้นภูมิที่ 1 มีค่าสังเกตเป็นบวก และชั้นภูมิที่ 2 มีค่าสังเกตเป็นศูนย์ ยกตัวอย่างเช่น การสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาสามารถแบ่งออกเป็น 2 ชั้นภูมิ คือ ชั้นภูมิที่ 1 เป็นบริษัทที่มีกิจกรรมวิจัยและพัฒนาซึ่งมีค่าใช้จ่ายเป็นบวก และชั้นภูมิที่ 2 เป็นบริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนาซึ่งมีค่าใช้จ่ายเป็นศูนย์

3.2.2 คำนวณขนาดประชากรที่เป็นไปได้จากการสุ่มจากทั้ง 2 ชั้นภูมิในข้อ 1 โดยขนาดของประชากรดังกล่าวนั้นมีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ส่วนขนาดตัวอย่างมีการแจกแจงการแจกแจงแบบไฮเปอร์จีออเมตริก (hypergeometric distribution) ซึ่งเป็นการแจกแจงที่ใช้อย่างแพร่หลายในกรณีที่ต้องการทดสอบตัวอย่างเพื่อยอมรับหรือไม่ยอมรับ ประชากรจึงมีความใกล้เคียงกับการแจกแจงแบบทวินาม (binomial distribution) ข้อแตกต่างกันก็คือการแจกแจงแบบไฮเปอร์จีออเมตริกจะมีโอกาสของสิ่งที่สนใจซึ่งก็คือ บริษัทที่มีกิจกรรมวิจัยและพัฒนาที่ไม่อิสระเนื่องมาจากวิธีสุ่มตัวอย่างจะเป็นแบบดึงออกมาแล้วไม่ใส่คืน (without replacement selected) (Berkopec, 2007) เราสามารถเขียนสมการคณิตศาสตร์ของโอกาสได้ดังนี้

$$h(m_1; N, m, \mathbf{N_1}) = \frac{\binom{\mathbf{N_1}}{m_1}\binom{\mathbf{N_2}}{m_2}}{\binom{N}{m}}$$

3.2.3 ใช้การอนุมานทางสถิติแบบเบย์ (Berger, 1980; Bayes, 1763) คำนวณหาการแจกแจงหลัง (posterior distribution) เมื่อการแจกแจงก่อน (prior distribution) เป็นการแจกแจงแบบสม่ำเสมอ (uniform distribution) ของพารามิเตอร์แทนประชากรในชั้นภูมิที่ 1 โดยคำนวณได้จาก

$$f(x) = P\left[\mathbf{N_1} = k \,\middle|\, \mathbf{M_1} = m_1\right] = \frac{P(\mathbf{M_1} = m_1 | \mathbf{N_1} = k)P(\mathbf{N_1} = k)}{\sum_{k=0}^{N} P(\mathbf{M_1} = m_1 | \mathbf{N_1} = k)P(\mathbf{N_1} = k)}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

เมื่อ N = ขนาดประชากรทั้งหมด โดย $N = \mathbf{N_1} + \mathbf{N_2}$

$\mathbf{N_1}$ = ตัวแปรสุ่มของขนาดประชากรในชั้นภูมิที่ 1

$\mathbf{N_2}$ = ตัวแปรสุ่มของขนาดประชากรในชั้นภูมิที่ 2

m = ขนาดตัวอย่างทั้งหมด โดย $m = m_1 + m_2$

m_1 = ขนาดตัวอย่างในชั้นภูมิที่ 1

$\mathbf{M_1}$ = ตัวแปรสุ่มของขนาดตัวอย่างในชั้นภูมิที่ 1

$\mathbf{M_2}$ = ตัวแปรสุ่มของขนาดตัวอย่างในชั้นภูมิที่ 2

$\mathbf{W_1}$ = ตัวแปรสุ่มของค่าถ่วงน้ำหนักของตัวอย่างในชั้นภูมิที่ 1

$\mathbf{W_2}$ = ตัวแปรสุ่มของค่าถ่วงน้ำหนักของตัวอย่างในชั้นภูมิที่ 2

$\mathbf{Y_{hi}}$ = ตัวแปรสุ่มของค่าใช้จ่ายด้านการวิจัยและพัฒนาในชั้นภูมิที่ $\mathbf{h}$ เมื่อ $\mathbf{h = 1,2}$ และ บริษัทที่ $\mathbf{i = 1, 2, ..., M_h}$

$\bar{\mathbf{Y}}$ = ตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนา

k = ขนาดของประชากรที่เป็นไปได้ในแต่ละชั้นภูมิ

$E(\bar{\mathbf{Y}}|\mathbf{M_1} = m_1)$ = ค่าประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์

$SD(\bar{\mathbf{Y}}|\mathbf{M_1} = m_1)$ = ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์

ทั้งนี้ $\mathbf{Y_{11}}$, $\mathbf{Y_{12}}$, ... เป็นตัวแปรสุ่มของค่าใช้จ่ายด้านการวิจัยและพัฒนาในชั้นภูมิที่ 1 ซึ่งแต่ละตัวเป็นอิสระกันและมีการแจกแจงเหมือนกัน (independent and identical distribution) ที่มีค่าคาดหวังเท่ากับ μ_1และความแปรปรวนเท่ากับ σ_1^2 และ $\mathbf{N_1}$ มีการแจกแจงแบบสม่ำเสมอ (uniform distribution) บนค่า 1,2, ..., N

3.2.4 นิยาม

$$\mathbf{W_1} = \frac{\mathbf{N_1}}{N}$$

$$\mathbf{W_2} = \frac{\mathbf{N_2}}{N} = \frac{N-\mathbf{N_1}}{N}$$

$$\bar{\mathbf{Y}} = \mathbf{W_1}\left(\frac{\mathbf{Y_{11}}+\mathbf{Y_{12}}+\ldots+\mathbf{Y_{1M_1}}}{\mathbf{M_1}}\right) + \mathbf{W_2}\left(\frac{\mathbf{Y_{21}}+\mathbf{Y_{22}}+\ldots+\mathbf{Y_{2M_2}}}{\mathbf{M_2}}\right)$$

3.2.5 ทำการคำนวณค่าประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์เมื่อกำหนดว่าเรารู้ขนาดตัวอย่างในแต่ละชั้นภูมิดังนี้

$$E(\bar{\mathbf{Y}}|\mathbf{M_1} = m_1) = E\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}}+\mathbf{Y_{12}}+\ldots+\mathbf{Y_{1M_1}}}{\mathbf{M_1}}\right) + \mathbf{W_2}\left(\frac{\mathbf{Y_{21}}+\mathbf{Y_{22}}+\ldots+\mathbf{Y_{2M_2}}}{\mathbf{M_2}}\right)|\mathbf{M_1} = m_1\right)$$

3.2.6 ทำการคำนวณความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์เมื่อกำหนดว่าเรารู้ขนาดตัวอย่างในแต่ละชั้นภูมิดังนี้

$$SD(\bar{\mathbf{Y}}|\mathbf{M_1} = m_1) = \sqrt{\mathrm{Var}\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}}+\mathbf{Y_{12}}+\ldots+\mathbf{Y_{1M_1}}}{\mathbf{M_1}}\right) + \mathbf{W_2}\left(\frac{\mathbf{Y_{21}}+\mathbf{Y_{22}}+\ldots+\mathbf{Y_{2M_2}}}{\mathbf{M_2}}\right)|\mathbf{M_1} = m_1\right)}$$

3.2.7 นำระเบียบวิธีที่ได้มาใช้ในข้อมูลจริงเพื่อเป็นกรณีศึกษาของงานวิจัย

3.2.7.1 ข้อมูลที่ได้จากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาซึ่งมีการสำรวจและพัฒนาฐานข้อมูลด้านการวิจัยและพัฒนาและกิจกรรมนวัตกรรมในภาคอุตสาหกรรมของประเทศไทยมาเป็นกรณีศึกษาของงานวิจัย

3.2.7.2 แบ่งข้อมูลในข้อ 3.2.6.1 ออกเป็น 2 ชั้นภูมิ คือ บริษัทที่มีค่าใช้จ่ายด้านการวิจัยและพัฒนาซึ่งมีค่าใช้จ่ายเป็นบวก และบริษัทที่ไม่มีค่าใช้จ่ายด้านการวิจัยและพัฒนาที่มีค่าใช้จ่ายเป็นศูนย์

3.2.7.3 ทำการประมาณค่าเฉลี่ย และค่าความคลาดเคลื่อนของค่าใช้จ่ายด้านการวิจัยและพัฒนา

3.2.8 สรุปผล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4

ผลการวิจัย

ในการศึกษาครั้งนี้ผู้วิจัยต้องการนำเสนอวิธีการประมาณค่าพารามิเตอร์จากการสุ่มตัวอย่างแบบชั้นภูมิ โดยอาศัยการถ่วงน้ำหนักของชั้นภูมิด้วยการอนุมานทางสถิติแบบเบย์ (Bayesian inference) ซึ่งเป็นวิธีการที่นิยมใช้กับข้อมูลที่มีมิติไม่สูงมากโดยใช้การแจกแจงก่อน (prior distribution) มาหาการแจกแจงหลัง (posterior distribution) ของพารามิเตอร์ในตัวแบบทำให้ประสิทธิภาพในการวิเคราะห์เพิ่มมากขึ้น อีกทั้งยังสามารถหาค่าประมาณแบบจุดและแบบช่วงของพารามิเตอร์ได้อีกด้วย (Kish, 1965)

4.1 ผลการศึกษาตัวแบบของพารามิเตอร์

เราสามารถหาตัวแบบค่าประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ทั้ง 2 ชั้นภูมิที่มีการแจกแจงก่อน (prior distribution) เป็นการแจกแจงแบบสม่ำเสมอ (uniform distribution) ด้วยวิธีแบบเบย์ดังนี้

ตัวแบบค่าประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเท่ากับ

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \mathrm{E}\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}+Y_{12}+\cdots+Y_{1M_1}}}{\mathbf{M_1}}\right) + \mathbf{W_2}\left(\frac{\mathbf{Y_{21}+Y_{22}+\ldots+Y_{2M_2}}}{\mathbf{M_2}}\right) \mid \mathbf{M_1} = \mathrm{m}_1\right)$$

เนื่องจาก $\mathbf{W_1} = \frac{\mathbf{N_1}}{\mathrm{N}}$ และ $\mathbf{W_2} = \frac{\mathbf{N_2}}{\mathrm{N}} = \frac{\mathrm{N}-\mathbf{N_1}}{\mathrm{N}}$ จะได้ว่า

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \mathrm{E}\left(\frac{\mathbf{N_1}}{\mathrm{N}}\left(\frac{\mathbf{Y_{11}+Y_{12}+\cdots+Y_{1M_1}}}{\mathbf{M_1}}\right) + \frac{\mathbf{N_2}}{\mathrm{N}}\left(\frac{\mathbf{Y_{21}+Y_{22}+\ldots+Y_{2M_2}}}{\mathbf{M_2}}\right) \mid \mathbf{M_1} = \mathrm{m}_1\right)$$

เนื่องจากชั้นภูมิที่ 1 มีค่าสังเกตเป็นบวก เพราะเป็นบริษัทที่มีกิจกรรมวิจัยและพัฒนาซึ่งมีค่าใช้จ่ายเป็นบวก และชั้นภูมิที่ 2 มีค่าสังเกตเป็นศูนย์ เพราะเป็นบริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนาซึ่งมีค่าใช้จ่ายเป็นศูนย์ จะได้ว่า

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \mathrm{E}\left(\frac{\mathbf{N_1}}{\mathrm{N}}\left(\frac{\mathbf{Y_{11}+Y_{12}+\cdots+Y_{1M_1}}}{\mathbf{M_1}}\right) \mid \mathbf{M_1} = \mathrm{m}_1\right)$$

เนื่องจาก $\frac{\mathbf{N_1}}{\mathrm{N}}$ และ $\left(\frac{\mathbf{Y_{11}+Y_{12}+\ldots+Y_{1M_1}}}{\mathbf{M_1}}\right)$ เป็นอิสระต่อกันเมื่อทราบ $\mathbf{M_1} = \mathrm{m}_1$ และต่างก็มีค่าเฉลี่ยจำกัด และจากความน่าจะเป็นแบบมีเงื่อนไข จะได้ว่า

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \mathrm{E}\left(\frac{\mathbf{N_1}}{\mathrm{N}} \mid \mathbf{M_1} = \mathrm{m}_1\right) \mathrm{E}\left(\frac{\mathbf{Y_{11}+Y_{12}+\ldots+Y_{1M_1}}}{\mathbf{M_1}} \mid \mathbf{M_1} = \mathrm{m}_1\right)$$

$$= \frac{1}{\mathrm{N}} \frac{\mathrm{m}_1\mu_1}{\mathrm{m}_1} \mathrm{E}(\mathbf{N_1} \mid \mathbf{M_1} = \mathrm{m}_1)$$

$$= \frac{\mu_1}{\mathrm{N}} \mathrm{E}(\mathbf{N_1} \mid \mathbf{M_1} = \mathrm{m}_1)$$

เนื่องจากขนาดประชากรที่เป็นไปได้ในแต่ละชั้นภูมิ (k) สามารถเกิดขึ้นได้หลายกรณี จะได้ว่า

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=0}^{\mathrm{N}} \mathrm{k}\, \mathrm{P}(\mathbf{N_1} = \mathrm{k} \mid \mathbf{M_1} = \mathrm{m}_1)$$

จากการอนุมานทางสถิติแบบเบย์ และตัวแปรสุ่มแต่ละค่ามีความน่าจะเป็นที่จะเกิดเท่าๆ กัน คือ $\mathrm{P}(\mathbf{N_1} = \mathrm{k}) = \frac{1}{\mathrm{N}}$ และมีการแจกแจงก่อนซึ่งเป็นการแจกแจงแบบสม่ำเสมอ จะได้ว่า

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=0}^{\mathrm{N}} \frac{\mathrm{k}\, \mathrm{P}(\mathbf{M_1} = \mathrm{m}_1 \mid \mathbf{N_1} = \mathrm{k}) \mathrm{P}(\mathbf{N_1} = \mathrm{k})}{\sum_{\mathrm{k}=0}^{\mathrm{N}} \mathrm{P}(\mathbf{M_1} = \mathrm{m}_1 \mid \mathbf{N_1} = \mathrm{k}) \mathrm{P}(\mathbf{N_1} = \mathrm{k})}$$

สังเกตว่าเมื่อทราบ $\mathbf{N_1} = \mathrm{k}$ จะได้ $\mathbf{M_1}$ มีการแจกแจงแบบไฮเปอร์จีออเมตริกโดยสิ่งที่สนใจคือบริษัทที่มีกิจกรรมวิจัยและพัฒนา เนื่องจากวิธีสุ่มตัวอย่างเป็นการสุ่มแบบไม่ใส่คืน (sampling without replacement) (Berkopec, 2007) โดยมีขนาดประชากรที่เป็นไปได้ในแต่ละชั้นภูมิ (k) นั้นเป็นไปได้ตั้งแต่ $\mathrm{m}_1, \mathrm{m}_1 + 1$ ไปเรื่อยๆ จนถึง $\mathrm{N} - (\mathrm{m} - \mathrm{m}_1)$ ดังนั้น

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}} \qquad (1)$$

ตัวแบบความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเท่ากับ

กำหนดให้ $\mathrm{Var}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1)$ คือ ความแปรปรวนของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์

$$\mathrm{SD}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1) = \sqrt{\mathrm{Var}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1)}$$

$$= \sqrt{\mathrm{Var}\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}+Y_{12}+\ldots+Y_{1M_1}}}{\mathbf{M_1}}\right)+\mathbf{W_2}\left(\frac{\mathbf{Y_{21}+Y_{22}+\ldots+Y_{2M_2}}}{\mathbf{M_2}}\right)|\mathbf{M_1}=\mathrm{m}_1\right)}$$

เนื่องจาก $\mathrm{Var}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1) = \mathrm{E}((\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1)^2) - \left(\mathrm{E}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1)\right)^2$ และจากสมการ (1) จะได้ว่า

$$\mathrm{Var}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1) = \mathrm{E}\left(\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}+Y_{12}+\cdots+Y_{1M_1}}}{\mathbf{M_1}}\right)+\mathbf{W_2}\left(\frac{\mathbf{Y_{21}+Y_{22}+\cdots+Y_{2M_2}}}{\mathbf{M_2}}\right)\Big|\mathbf{M_1}=\mathrm{m}_1\right)^2\right)$$

$$-\left(\mathrm{E}\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}+Y_{12}+\cdots+Y_{1M_1}}}{\mathbf{M_1}}\right)+\mathbf{W_2}\left(\frac{\mathbf{Y_{21}+Y_{22}+\cdots+Y_{2M_2}}}{\mathbf{M_2}}\right)\Big|\mathbf{M_1}=\mathrm{m}_1\right)\right)^2 \quad (2)$$

ดังนั้นจากสมการ (1) และ (2) จะได้ว่า

$$\mathrm{Var}(\bar{\mathbf{Y}}|\mathbf{M_1}=\mathrm{m}_1) = \mathrm{E}\left(\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}+Y_{12}+\ldots+Y_{1M_1}}}{\mathbf{M_1}}\right)+\mathbf{W_2}\left(\frac{\mathbf{Y_{21}+Y_{22}+\ldots+Y_{2M_2}}}{\mathbf{M_2}}\right)\right)^2\Bigg|\mathbf{M_1}=\mathrm{m}_1\right)$$

$$-\left(\frac{\mu_1}{\mathrm{N}}\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \mathrm{E}\left(\left(\frac{\mathbf{N_1}}{\mathrm{N}}\left(\frac{\mathbf{Y_{11}+Y_{12}+\ldots+Y_{1M_1}}}{\mathbf{M_1}}\right)\right)^2\Bigg|\mathbf{M_1}=\mathrm{m}_1\right)$$

$$-\left(\frac{\mu_1}{\mathrm{N}}\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$= \mathrm{E}\left(\left(\frac{\mathbf{N_1}}{\mathrm{N}}\right)^2 \mid \mathbf{M_1} = \mathrm{m}_1\right) \mathrm{E}\left(\left(\frac{\mathbf{Y_{11}+Y_{12}+\ldots+Y_{1M_1}}}{\mathbf{M_1}}\right)^2 \mid \mathbf{M_1} = \mathrm{m}_1\right)$$

$$-\left(\frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \mathrm{E}\left(\left(\frac{\mathbf{N_1}}{\mathrm{N}}\right)^2 \mid \mathbf{M_1} = \mathrm{m}_1\right) \mathrm{E}\left(\left(\frac{\mathrm{Y}_{11}+\mathrm{Y}_{12}+\ldots+\mathrm{Y}_{1\mathrm{m}_1}}{\mathrm{m}_1}\right)^2\right)$$

$$-\left(\frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \mathrm{E}\left(\left(\frac{\mathbf{N_1}}{\mathrm{N}}\right)^2 \mid \mathbf{M_1} = \mathrm{m}_1\right) \mathrm{E}(\bar{\mathrm{Y}}^2) - \left(\frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \mathrm{E}\left(\left(\frac{\mathbf{N_1}}{\mathrm{N}}\right)^2 \mid \mathbf{M_1} = \mathrm{m}_1\right) (\mathrm{Var}(\bar{\mathrm{Y}}) + (\mathrm{E}(\bar{\mathrm{Y}}))^2)$$

$$- \left(\frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \mathrm{E}\left(\left(\frac{\mathbf{N_1}}{\mathrm{N}}\right)^2 \mid \mathbf{M_1} = \mathrm{m}_1\right)\left(\frac{{\sigma_1}^2}{\mathrm{m}_1} + {\mu_1}^2\right)$$

$$- \left(\frac{\bar{\mathrm{Y}}_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

Chulalongkorn University

$$= \frac{1}{\mathrm{N}^2}\left(\frac{{\sigma_1}^2}{\mathrm{m}_1} + {\mu_1}^2\right) \mathrm{E}(\mathbf{N_1}^2 \mid \mathbf{M_1} = \mathrm{m}_1)$$

$$- \left(\frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \frac{1}{\mathrm{N}^2}\left(\frac{{\sigma_1}^2}{\mathrm{m}_1} + {\mu_1}^2\right) \left(\sum_{\mathrm{k}=0}^{\mathrm{N}} \mathrm{k}\, \mathrm{P}(\mathbf{N_1} = \mathrm{k} | \mathbf{M_1} = \mathrm{m}_1)\right)^2$$

$$- \left(\frac{\mu_1}{\mathrm{N}} \sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)} \frac{\mathrm{k}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}{\sum_{\mathrm{k}=\mathrm{m}_1}^{\mathrm{N}-(\mathrm{m}-\mathrm{m}_1)}\binom{\mathrm{k}}{\mathrm{m}_1}\binom{\mathrm{N}-\mathrm{k}}{\mathrm{m}-\mathrm{m}_1}}\right)^2$$

$$= \frac{1}{N^2}\left(\frac{{\sigma_1}^2}{m_1} + \mu_1\right)\sum_{k=m_1}^{N-(m-m_1)} k^2\, P(\mathbf{N_1} = k|\mathbf{M_1} = m_1)$$

$$- \left(\frac{\mu_1}{N}\sum_{k=m_1}^{N-(m-m_1)} \frac{k\binom{k}{m_1}\binom{N-k}{m-m_1}}{\sum_{k=m_1}^{N-(m-m_1)}\binom{k}{m_1}\binom{N-k}{m-m_1}}\right)^2$$

$$= \frac{1}{N^2}\left(\frac{{\sigma_1}^2}{m_1} + {\mu_1}^2\right)\sum_{k=m_1}^{N-(m-m_1)} \frac{k^2\binom{k}{m_1}\binom{N-k}{m-m_1}}{\sum_{k=m_1}^{N-(m-m_1)}\binom{k}{m_1}\binom{N-k}{m-m_1}}$$

$$- \left(\frac{\mu_1}{N}\sum_{k=m_1}^{N-(m-m_1)} \frac{k\binom{k}{m_1}\binom{N-k}{m-m_1}}{\sum_{k=m_1}^{N-(m-m_1)}\binom{k}{m_1}\binom{N-k}{m-m_1}}\right)^2$$

ดังนั้น $SD(\bar{\mathbf{Y}}|\mathbf{M_1} = m_1)$ =

$$\sqrt{\frac{1}{N^2}\left(\frac{{\sigma_1}^2}{m_1} + {\mu_1}^2\right)\sum_{k=m_1}^{N-(m-m_1)} \frac{k^2\binom{k}{m_1}\binom{N-k}{m-m_1}}{\sum_{k=m_1}^{N-(m-m_1)}\binom{k}{m_1}\binom{N-k}{m-m_1}} - \left(\frac{\mu_1}{N}\sum_{k=m_1}^{N-(m-m_1)} \frac{k\binom{k}{m_1}\binom{N-k}{m-m_1}}{\sum_{k=m_1}^{N-(m-m_1)}\binom{k}{m_1}\binom{N-k}{m-m_1}}\right)^2} \quad (3)$$

ในการประยุกต์ใช้สูตรในสมการ (1) และ (3) ตามสมการ (4) และ (5) ดังนี้

μ_1 ประมาณค่าด้วย $\hat{\mu}_1 = \frac{Y_{11}+Y_{12}+\ldots+Y_{1m_1}}{m_1}$ (4)

${\sigma_1}^2$ ประมาณค่าด้วย $\hat{\sigma}_1^{\,2} = \frac{(Y_{11}-\hat{\mu}_1)^2+(Y_{12}-\hat{\mu}_1)^2+\cdots(Y_{1m_1}-\hat{\mu}_1)^2}{m_1-1}$ (5)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2 ผลการศึกษาการเปรียบเทียบการประมาณค่าแบบชั้นภูมิเบย์กับการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย

สังเกตว่า ค่าประมาณค่าเฉลี่ย และค่าความคลาดเคลื่อนมาตรฐานของประชากรแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายเป็นฟังก์ชันที่มีตัวแปร 5 ตัว ได้แก่ N, m, m_1, μ_1และ σ_1

เมื่อนำข้อมูลสมมติที่สามารถพบได้ในสถานการณ์จริงมาทำการประมาณค่าพารามิเตอร์ด้วยตัวแบบทั้ง 2 ตัวแบบพบว่าค่าความคลาดเคลื่อนของค่าเฉลี่ยของประชากรโดยวิธีการประมาณค่าแบบชั้นภูมิเบย์มีค่าต่ำกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายเล็กน้อย (รายละเอียดดังตารางที่ 4.2.1)

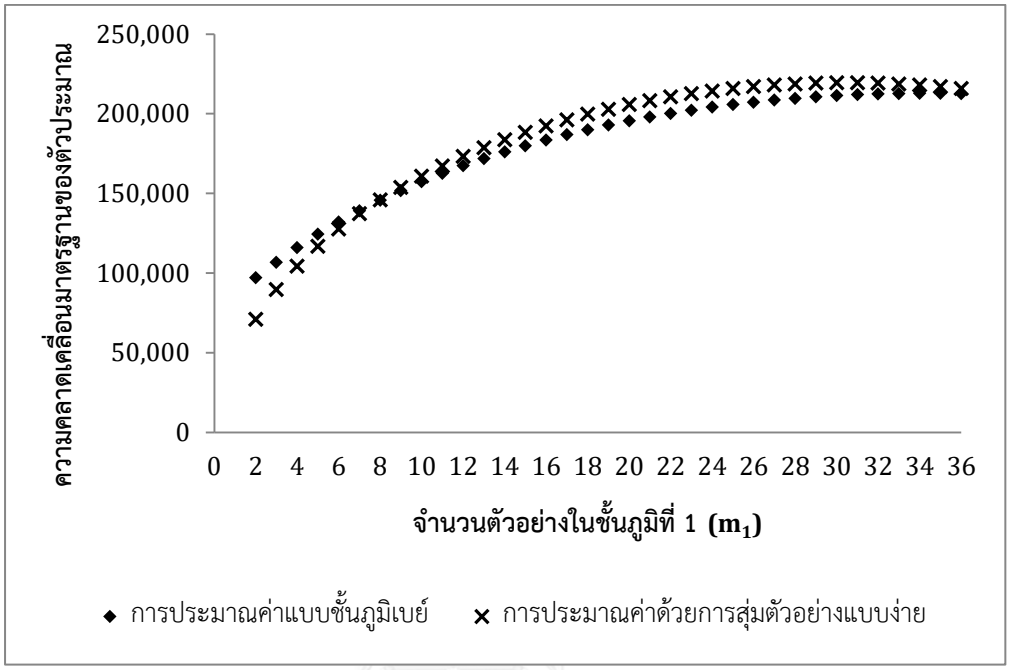
<u>ตารางที่ 4.2.1</u> ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลสมมติที่สามารถพบได้ในสถานการณ์จริง

รายละเอียด						**ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา**	
N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	**การประมาณค่าแบบชั้นภูมิเบย์**	**การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย**
200	38	8	30	1.69×10^6	1.31×10^6	145,700.21	145,956.76

นอกจากนี้เมื่อทำการวิเคราะห์ความไว (sensitivity analysis) ที่จะแสดงให้เห็นความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายจะเปลี่ยนแปลงไปเท่าใดถ้าตัวแปรใดตัวแปรหนึ่งในกลุ่ม N, m, m_1, μ_1 และ σ_1 มีการเปลี่ยนแปลงโดยที่สมมติให้ตัวแปรอื่นๆ คงที่โดยแบ่งออกเป็น 3 กรณีดังนี้

กรณีที่ 1 ทำการวิเคราะห์ความไว เมื่อขนาดตัวอย่างในชั้นภูมิที่ 1 (m_1) เปลี่ยนแปลง ในขณะที่ N, m, μ_1 และ σ_1 คงที่ ดังตารางที่ 4.2.1 โดยในภาพที่ 4.2.1 เราจะสามารถสมมติ m_1 แตกต่างกันได้ทั้งสิ้น 35 กรณี คือ ตั้งแต่ 2, 3, ..., m-2 = 36 ตัวอย่าง ได้ผลดังภาพที่ 4.2.1

<u>ภาพที่ 4.2.1</u> ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในกรณีที่มีขนาดตัวอย่างในชั้นภูมิที่ 1 แตกต่างกันจำนวน 35 กรณี



จากภาพที่ 4.2.1 จะเห็นได้ว่า เมื่อ m_1 เพิ่มขึ้นจะทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะมีค่าต่ำกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย เนื่องจากเมื่อ m_1 เพิ่มขึ้นจะช่วยลดผลกระทบเนื่องจากค่าสังเกตของตัวอย่างไม่ครอบคลุมประชากรได้เป็นผลให้ข้อมูลเกาะกลุ่มกันมากขึ้น ชั้นภูมิทั้ง 2 ชั้นภูมิแยกออกจากกันอย่างชัดเจนทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ลดลง

เมื่อพิจารณาเฉพาะกราฟแสดงความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเห็นได้ว่าเมื่อ m_1 มากขึ้น ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ก็จะมีค่าเพิ่มมากขึ้นเรื่อยๆ โดยในช่วงแรกกราฟจะมีความชันมากเป็นช่วงที่ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์มีการเพิ่มขึ้นอย่างรวดเร็วจนกระทั่งถึงความชันค่อยๆ ลดลง และเมื่อถึงช่วงหนึ่งก็จะหยุด หลังจากผ่านช่วงนั้นแล้วเส้นกราฟก็จะไม่สูงขึ้นอีกต่อไป แสดงว่าความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์แทบจะไม่มีการเปลี่ยนแปลง

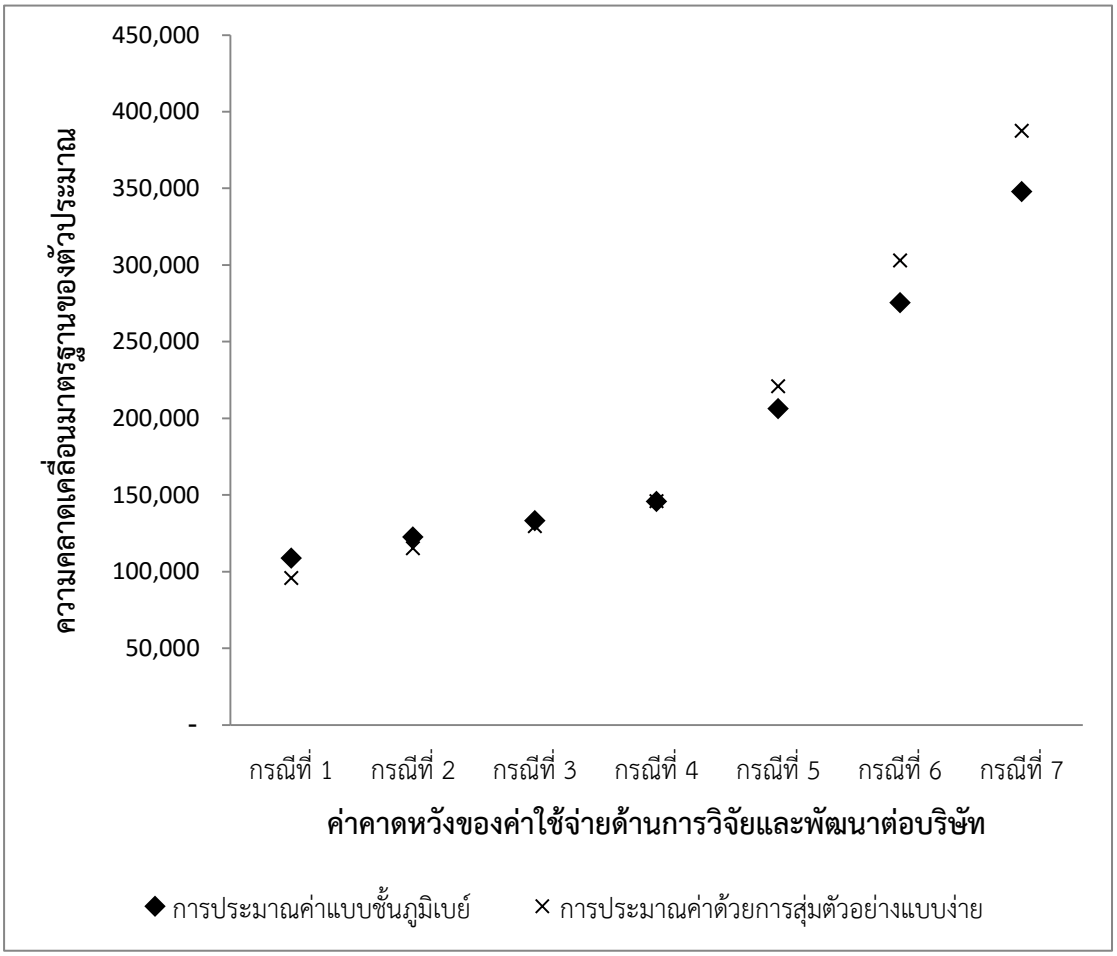
กรณีที่ 2 ทำการวิเคราะห์ความไว เมื่อค่าคาดหวังของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัท (μ_1) เปลี่ยนแปลง ในขณะที่ N, m, m_1และ σ_1 คงที่ ดังตารางที่ 4.2.1 โดยในภาพที่ 4.2.2 เราจะสมมติ μ_1 แตกต่างกันตามส่วนเบี่ยงเบนมาตรฐานของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัท (σ) ในระดับต่างๆ ทั้งสิ้น 7 กรณี คือ

กรณีที่ 1	μ_1- σ_1	มีค่าเท่ากับ	374,569.32	บาท
กรณีที่ 2	$\mu_1$0.5 σ_1	มีค่าเท่ากับ	1,029,784.66	บาท
กรณีที่ 3	μ_1- 0.25 σ_1	มีค่าเท่ากับ	1,357,392.335	บาท
กรณีที่ 4	μ_1	มีค่าเท่ากับ	1,685,000.00	บาท
กรณีที่ 5	μ_1+ σ_1	มีค่าเท่ากับ	2,995,430.68	บาท
กรณีที่ 6	μ_1+ 2 σ_1	มีค่าเท่ากับ	4,305,861.36	บาท
กรณีที่ 7	μ_1+ 3 σ_1	มีค่าเท่ากับ	616,292.04	บาท

ได้ผลดังภาพที่ 4.2.2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

<u>ภาพที่ 4.2.2</u> ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในกรณีที่มีค่าคาดหวังของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัทแตกต่างกันจำนวน 7 กรณี



จากภาพที่ 4.2.2 จะเห็นได้ว่า เมื่อ μ_1 เพิ่มขึ้นจะทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะมีค่าต่ำกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย แต่ในกรณีที่ μ_1ลดลงจะทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะมีค่าสูงกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย เนื่องจากเมื่อ μ_1 เพิ่มขึ้นเป็นผลให้ชั้นภูมิทั้ง 2 ชั้นภูมิแยกออกจากกันอย่างชัดเจนขึ้นทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ลดลง

เมื่อพิจารณาเฉพาะกราฟแสดงความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเห็นได้ว่าเมื่อ μ_1 มากขึ้น ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ก็จะมีค่าเพิ่มมากขึ้นเรื่อยๆ โดยในช่วงแรกที่ μ_1 ลดลงเป็นช่วงที่มีการเพิ่มขึ้นของความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จนกระทั่งช่วงที่ μ_1 เพิ่มขึ้นเป็นช่วงที่ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์เพิ่มขึ้น ในช่วงที่ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ ทำให้ดูเหมือนว่าความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเพิ่มขึ้น

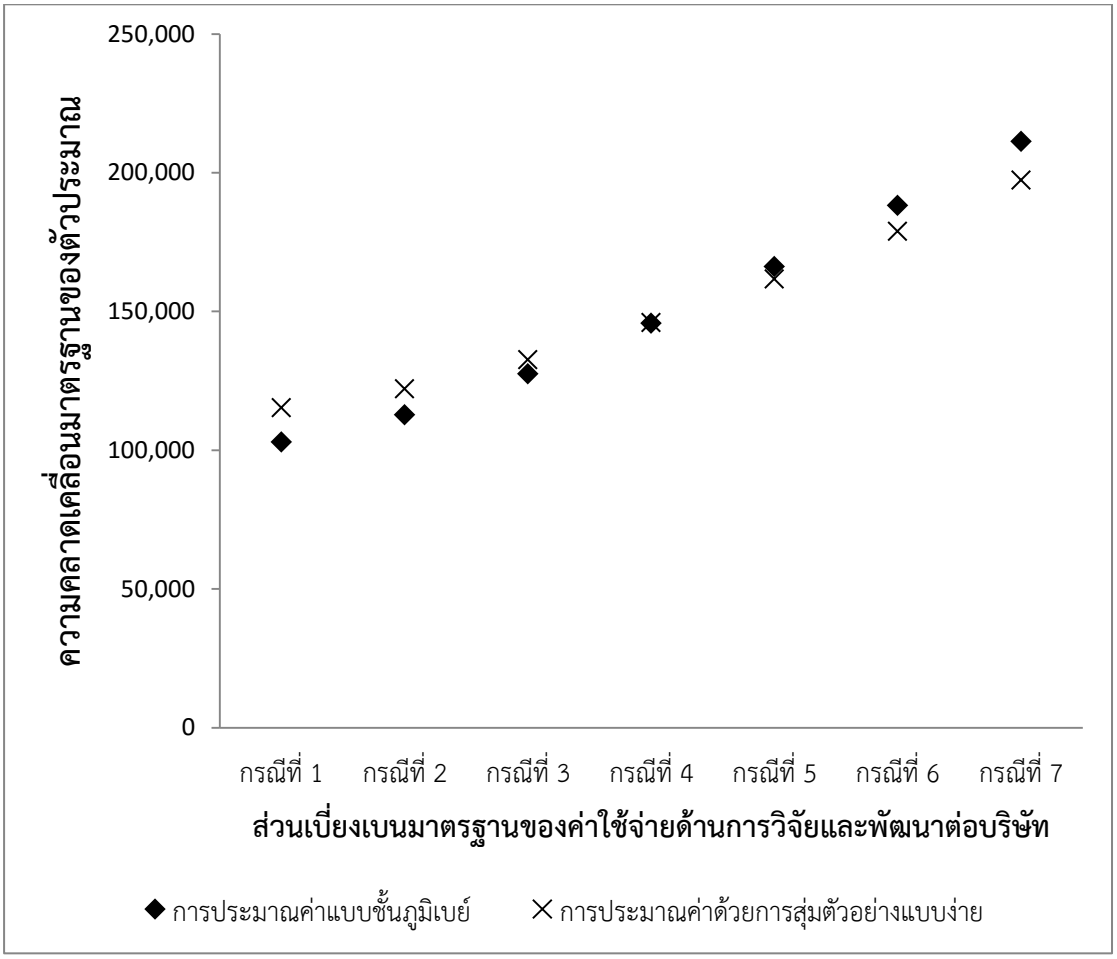
กรณีที่ 3 ทำการวิเคราะห์ความไวเมื่อส่วนเบี่ยงเบนมาตรฐานของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัท (σ_1) เปลี่ยนแปลง ในขณะที่ N, m, m_1และ μ_1 คงที่ ดังตารางที่ 4.2.1 โดยในภาพที่ 4.2.3 เราจะสมมติ S_1 ในระดับต่างๆ ทั้งสิ้น 7 กรณี คือ

กรณีที่ 1	$0.25\sigma_1$	มีค่าเท่ากับ	1.07×10^{11}	บาท
กรณีที่ 2	$0.50\sigma_1$	มีค่าเท่ากับ	4.29×10^{11}	บาท
กรณีที่ 3	$0.75\sigma_1$	มีค่าเท่ากับ	9.66×10^{11}	บาท
กรณีที่ 4	σ_1	มีค่าเท่ากับ	1.72×10^{12}	บาท
กรณีที่ 5	$1.25\sigma_1$	มีค่าเท่ากับ	2.68×10^{12}	บาท
กรณีที่ 6	$1.50\sigma_1$	มีค่าเท่ากับ	3.86×10^{12}	บาท
กรณีที่ 7	$1.75\sigma_1$	มีค่าเท่ากับ	5.26×10^{12}	บาท

ได้ผลดังภาพที่ 4.2.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

<u>ภาพที่ 4.2.3</u> ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในกรณีที่มีส่วนเบี่ยงเบนมาตรฐานของค่าใช้จ่ายด้านการวิจัยและพัฒนาต่อบริษัทแตกต่างกันจำนวน 7 กรณี



จากภาพที่ 4.2.3 จะเห็นได้ว่า เมื่อ σ_1ลดลงจะทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะมีค่าต่ำกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย แต่ในกรณีที่ σ_1 เพิ่มขึ้นจะทำให้ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะมีค่าสูงกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย ทั้งนี้เนื่องจากส่วนเบี่ยงเบนมาตรฐานของประชากรลดลง เป็นผลให้ข้อมูลเกาะกลุ่มกันมากขึ้น ชั้นภูมิทั้ง 2 ชั้นภูมิจึงแยกออกจากกันอย่างชัดเจนและทำให้ความคลาดเคลื่อน-มาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ลดลงในท้ายที่สุด

เมื่อพิจารณาเฉพาะกราฟแสดงความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเห็นได้ว่า เมื่อ $\bar{Y}_1$ มากขึ้น ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์ก็จะมีค่าเพิ่มมากขึ้นเรื่อยๆ โดยในช่วงที่ σ_1 ลดลง และ ช่วงที่ σ_1เพิ่มขึ้นเป็นช่วงที่ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์มีแนวโน้มของการเพิ่มขึ้นค่อนข้างคงที่ และดูเหมือนว่าความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์จะเพิ่มขึ้น

ข้อมูลด้านค่าใช้จ่ายด้านการวิจัยและพัฒนามีความสำคัญในแง่ที่ใช้เป็นดัชนีที่สำคัญอย่างหนึ่งที่ใช้วัดขีดความสามารถด้านการวิจัยและพัฒนาของประเทศ ซึ่งสะท้อนให้เห็นถึงระดับความเจริญก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยีของภาคอุตสาหกรรมไทย เนื่องจากการวิจัยและพัฒนาเป็นรากฐานของการสร้างความรู้และวิทยาการใหม่ๆ ที่สามารถพัฒนาประเทศให้เจริญก้าวหน้า เช่น ข้อมูลค่าใช้จ่ายด้านการวิจัยและพัฒนาในปี 2551 ภาคอุตสาหกรรมไทยมีบริษัทที่ดำเนินกิจกรรมวิจัยและพัฒนาจำนวน 655 กิจการ มีค่าใช้จ่ายด้านการวิจัยและพัฒนาเป็นเงิน 11.11 ล้านบาทต่อบริษัท เป็นต้น (สำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ เทคโนโลยีและนวัตกรรมแห่งชาติ, 2554) ซึ่งในการสำรวจข้อมูลค่าใช้จ่ายด้านการวิจัยและพัฒนาดังกล่าวต้องทำการสุ่มตัวอย่างบริษัทจากรายชื่อบริษัทที่ปรากฏอยู่ในฐานข้อมูลของทางราชการเพื่อทำการส่งแบบสอบถามและติดตามผลทางโทรศัพท์ต่อไป

เมื่อนำข้อมูลจริงที่ได้การสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสาหกรรมของประเทศไทยประจำปี 2556 จำนวนทั้งสิ้น 27 ประเภทกิจการมาเป็นกรณีศึกษาของงานวิจัยแล้วทำการประมาณค่าพารามิเตอร์ของค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาด้วยตัวแบบทั้ง 2 ตัวแบบจะพบว่าความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์มีค่าต่ำกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายในบางประเภทกิจการโดยในตารางที่ 4.2.2 เป็นการพิจารณาข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสาหกรรมของประเทศไทยประจำปี 2556 โดยรวมทุกขนาดกิจการได้แก่ กิจการขนาดเล็กขนาดกลางและขนาดใหญ่ พบว่าการสุ่มตัวอย่างแบบชั้นภูมิซึ่งไม่ทราบขนาดประชากรในแต่ละชั้นภูมิจะมีความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาที่มีค่าต่ำกว่าการสุ่มตัวอย่างแบบง่ายจำนวน 12 ประเภทจากทั้งหมด 27 ประเภท ได้แก่ กิจการประเภทที่ 1, 2, 3, 4, 6, 7, 10, 11, 14, 19, 24 และ 27 โดยความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาด้วยการประมาณค่าแบบชั้นภูมิเบย์ลดลงจากการสุ่มตัวอย่างแบบง่ายมีค่าร้อยละการลดลงอยู่ในช่วงร้อยละ 0.11 ถึงร้อยละ 4.33 หรือคิดเป็นค่าเฉลี่ยร้อยละการลดลงเท่ากับร้อยละ 1.92 (รายละเอียดดังตารางที่ 4.2.2)

ตารางที่ 4.2. 2 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 โดยรวมทุกขนาดกิจการ

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
1	108	45	21	24	6,916,444.52	11,621,449.32	1,254,121.00	1,278,599.00
2	232	67	35	32	3,570,618.12	5,189,736.81	494,159.00	505,254.70
3	210	61	24	37	1,327,925.29	1,608,442.62	147,997.30	152,549.70
4	72	34	14	20	8,306,147.14	16,226,082.46	1,882,744.00	1,886,744.00
5	710	62	17	45	6,424,229.68	10,429,486.41	799,749.40	771,246.10
6	891	117	77	40	5,582,870.36	22,134,463.76	1,672,737.00	1,674,506.00
7	192	54	17	37	4,228,420.09	4,340,268.97	407,346.10	421,997.30
8	615	69	10	59	9,707,600.00	21,371,435.20	1,143,285.00	1,023,638.00
9	403	53	14	39	6,879,703.57	9,777,117.87	819,206.40	792,354.50
10	345	60	22	38	3,987,726.27	4,759,766.89	439,606.80	443,831.10
11	268	65	30	35	4,801,953.01	4,579,972.81	464,530.60	485,560.10
12	266	53	28	25	16,000,161.10	55,954,194.88	5,694,904.00	5,647,954.00
13	184	55	26	29	10,129,096.65	26,923,807.21	2,577,915.00	2,564,246.00
14	232	73	41	32	7,431,877.53	10,712,774.50	1,004,730.00	1,030,660.00

<u>ตารางที่ 4.2.2 (ต่อ)</u> ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 โดยรวมทุกขนาดกิจการ

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
15	505	75	22	53	6,486,488.12	8,317,586.22	618,783.30	616,141.00
16	1,822	79	26	53	6,548,259.75	8,015,715.04	626,083.80	618,108.90
17	870	77	23	54	43,605,219.55	174,040,302.27	11,355,994.00	10,913,890.00
18	2,144	73	21	52	16,710,852.00	38,160,543.86	2,622,618.00	2,517,137.00
19	**174**	**48**	**10**	**38**	**3,856,005.00**	**3,930,724.55**	**336,245.60**	**337,364.70**
20	695	83	33	50	12,401,130.61	33,842,505.78	2,451,453.00	2,415,406.00
21	1751	86	27	59	14,696,515.94	32,733,392.58	2,146,076.00	2,087,649.00
22	677	79	23	56	18,394,230.22	36,103,787.11	2,417,658.00	2,355,631.00
23	618	52	13	39	3,759,623.08	3,259,731.53	320,281.70	316,301.40
24	**833**	**63**	**14**	**49**	**2,041,482.85**	**1,129,831.61**	**124,267.10**	**125,963.60**
25	474	73	22	51	12,987,851.55	25,945,014.34	1,826,440.00	1,784,034.00
26	1,335	52	8	44	107,391,744.05	263,713,352.94	17,005,581.00	14,594,600.00
27	**127**	**48**	**26**	**22**	**39,755,178.12**	**59,315,401.20**	**6,695,647.00**	**6,880,180.00**

ตารางที่ 4.2.3 เป็นการพิจารณาข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสาหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดเล็ก พบว่าการสุ่มตัวอย่างแบบชั้นภูมิซึ่งไม่ทราบขนาดประชากรในแต่ละชั้นภูมิจะมีความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาที่มีค่าต่ำกว่าการสุ่มตัวอย่างแบบง่ายจำนวน 13 ประเภทจากทั้งหมด 27 ประเภท ได้แก่ กิจการประเภทที่ 1, 2, 3, 4, 6, 7, 10, 11, 14, 19, 20, 22 และ 27 โดยความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาด้วยการประมาณค่าแบบชั้นภูมิเบย์ลดลงจากการสุ่มตัวอย่างแบบง่ายมีค่าร้อยละการลดลงอยู่ในช่วงร้อยละ 0.38 ถึงร้อยละ 10.43 หรือคิดเป็นค่าเฉลี่ยร้อยละการลดลงเท่ากับร้อยละ 4.13 (รายละเอียดดังตารางที่ 4.2.3)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.2. 3 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดเล็ก

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
1	69	16	3	13	927,060.00	273,739.75	86,629.24	96,711.61
2	149	22	6	16	1,490,400.00	886,317.14	166,481.30	171,703.30
3	123	21	7	14	556,332.86	234,872.58	58,890.33	65,015.63
4	48	16	3	13	527,153.33	277,502.52	55,287.15	58,856.24
5	501	19	2	17	2,370,000.00	1,598,061.33	250,207.50	191,981.80
6	641	34	14	20	1,417,975.21	1,210,176.62	177,025.40	178,119.50
7	117	20	3	17	430,045.00	175,333.42	37,192.89	37,454.44
8	434	22	3	19	4,812,000.00	4,823,145.86	612,543.60	480,166.80
9	337	24	4	20	2,442,062.50	3,056,170.44	359,620.70	294,574.90
10	227	19	2	17	741,275.00	570,316.97	82,693.37	61,856.19
11	203	28	12	16	1,969,970.25	1,216,899.68	225,198.80	238,215.10
12	224	25	9	16	1,402,943.33	882,213.78	164,702.80	171,092.20
13	125	24	10	14	2,230,643.00	3,011,472.76	452,816.70	447,711.70
14	147	27	12	15	6,085,999.42	6,621,049.24	1,004,686.00	1,019,155.00

ตารางที่ 4.2.3 (ต่อ) ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดเล็ก

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
15	313	24	5	19	1,365,424.00	716,049.01	131,895.40	130,708.40
16	1,270	22	3	19	4,862,583.33	7,317,315.09	848,250.20	603,646.00
17	657	25	2	23	744,575.00	59,503.04	43,752.25	41,304.19
18	1,734	23	2	21	1,900,000.00	1,979,898.99	224,425.20	144,135.50
19	**96**	**21**	**4**	**17**	**1,311,262.50**	**499,396.60**	**113,412.40**	**122,628.00**
20	**490**	**28**	**10**	**18**	**1,332,437.00**	**1,020,297.09**	**165,176.50**	**165,800.50**
21	1,386	28	8	20	870,193.80	912,132.48	122,745.30	115,875.80
22	**308**	**25**	**4**	**21**	**1,627,720.76**	**474,980.05**	**124,004.30**	**126,353.00**
23	540	21	2	19	569,200.00	119,076.78	40,341.84	37,810.44
24	680	26	6	20	1,858,108.31	1,152,182.76	191,466.30	186,351.90
25	181	23	6	17	7,627,325.67	8,722,383.95	1,209,057.00	1,123,235.00
26	1,221	21	2	19	1,036,200.00	998,717.62	125,385.50	83,670.86
27	**84**	**21**	**6**	**15**	**25,171,290.16**	**19,035,350.87**	**3,181,954.00**	**3,283,120.00**

ตารางที่ 4.2.4 เป็นการพิจารณาข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสาหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดกลาง พบว่าการสุ่มตัวอย่างแบบชั้นภูมิซึ่งไม่ทราบขนาดประชากรในแต่ละชั้นภูมิจะมีความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาที่มีค่าต่ำกว่าการสุ่มตัวอย่างแบบง่ายจำนวน 16 ประเภทจากทั้งหมด 27 ประเภท ได้แก่ กิจการประเภทที่ 1, 2, 3, 4, 6, 7, 10, 11,12, 13, 14, 17, 19, 23, 24 และ 27 โดยความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาด้วยการประมาณค่าแบบชั้นภูมิเบย์ลดลงจากการสุ่มตัวอย่างแบบง่ายมีค่าร้อยละการลดลงอยู่ในช่วงร้อยละ 1.84 ถึงร้อยละ 43.18 หรือคิดเป็นค่าเฉลี่ยร้อยละการลดลงเท่ากับร้อยละ 10.99 (รายละเอียดดังตารางที่ 4.2.4)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.2. 4 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดกลาง

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
1	15	11	6	5	3,489,700.00	4,520,051.54	1,043,930.00	1,109,329.00
2	41	17	6	11	2,182,416.67	1,008,594.88	239,684.30	294,419.40
3	53	21	7	14	1,458,375.00	1,314,233.94	207,198.00	219,786.70
4	8	6	3	3	1,644,200.00	558,398.71	224,395.30	394,913.80
5	146	19	2	17	27,044,958.82	23,359,921.38	3,181,541.00	2,328,667.00
6	142	33	25	8	1,801,378.64	1,490,664.34	251,538.90	262,917.00
7	21	11	2	9	2,668,100.00	237,729.30	220,118.00	326,209.50
8	121	25	2	23	1,020,000.00	593,969.70	73,448.29	61,469.99
9	44	14	2	12	1,257,950.00	709,864.50	138,867.50	132,943.40
10	52	20	10	10	1,892,156.80	1,674,738.62	311,345.90	336,952.40
11	33	19	9	10	6,005,502.78	5,524,649.89	982,671.80	1,101,585.00
12	19	14	9	5	3,868,213.89	5,019,286.79	1,100,548.00	1,171,177.00
13	29	13	5	8	7,907,650.40	7,872,999.64	1,605,784.00	1,680,086.00
14	49	24	15	9	3,022,692.39	3,218,326.65	557,904.80	596,487.60

ตารางที่ 4.2.4 (ต่อ) ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดกลาง

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
15	106	24	7	17	5,819,457.14	8,794,612.56	1,140,554.00	1,070,005.00
16	369	29	12	17	5,867,675.83	7,071,625.45	1,001,704.00	987,783.40
17	**125**	**26**	**6**	**20**	**1,345,376.00**	**519,537.96**	**111,235.80**	**122,182.80**
18	259	25	8	17	3,803,537.14	7,161,093.86	926,069.40	854,077.50
19	**30**	**12**	**4**	**8**	**5,405,000.00**	**5,124,186.18**	**1,069,443.00**	**1,089,467.00**
20	120	25	6	19	2,300,091.67	2,420,930.98	313,429.10	298,409.30
21	229	26	7	19	7,509,110.00	11,444,336.81	1,407,128.00	1,285,587.00
22	123	18	2	16	1,272,268.00	714,556.86	123,235.90	105,226.50
23	**58**	**20**	**7**	**13**	**4,404,471.43**	**3,144,129.50**	**569,084.30**	**623,193.00**
24	**110**	**23**	**6**	**17**	**2,148,693.33**	**1,058,834.02**	**210,306.50**	**227,029.80**
25	91	21	5	16	2,377,100.00	2,128,745.29	317,221.60	307,262.70
26	67	17	4	13	21,930,950.00	36,021,111.79	5,222,455.00	4,440,684.00

ตารางที่ 4.2.5 เป็นการพิจารณาข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสาหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดใหญ่ พบว่าการสุ่มตัวอย่างแบบชั้นภูมิซึ่งไม่ทราบขนาดประชากรในแต่ละชั้นภูมิจะมีความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาที่มีค่าต่ำกว่าการสุ่มตัวอย่างแบบง่ายจำนวน 17 ประเภทจากทั้งหมด 27 ประเภท ได้แก่ กิจการประเภทที่ 1, 2, 3, 4, 5, 6, 7, 9, 10, 11,12, 13, 14, 15, 20, 23 และ 27 โดยความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าใช้จ่ายเฉลี่ยด้านการวิจัยและพัฒนาด้วยการประมาณค่าแบบชั้นภูมิเบย์ลดลงจากการสุ่มตัวอย่างแบบง่ายมีค่าร้อยละการลดลงอยู่ในช่วงร้อยละ 0.62 ถึงร้อยละ 17.77 หรือคิดเป็นค่าเฉลี่ยร้อยละการลดลงเท่ากับร้อยละ 5.51 (รายละเอียดดังตารางที่ 4.2.5)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 4.2. 5 ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดใหญ่

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
1	24	18	12	6	10,127,162.92	14,450,792.02	2,826,633.00	2,974,463.00
2	42	28	23	5	4,475,423.23	6,217,148.14	1,073,985.00	1,110,691.00
3	34	19	10	9	1,776,725.20	2,173,270.51	386,658.80	409,895.60
4	16	12	8	4	13,721,500.00	20,257,788.34	4,839,919.00	5,056,293.00
5	63	24	13	11	3,875,537.46	4,324,339.53	718,134.00	754,085.20
6	108	50	38	12	9,605,128.92	31,169,773.04	3,845,474.00	3,875,034.00
7	54	23	12	11	5,438,067.21	4,627,760.55	814,351.30	894,978.40
8	60	22	5	17	16,120,000.00	30,146,815.99	3,575,370.00	3,168,878.00
9	22	15	8	7	10,503,962.50	11,754,094.23	2,345,719.00	2,562,573.00
10	66	21	10	11	6,732,586.00	5,867,608.21	1,068,428.00	1,141,522.00
11	32	18	9	9	7,374,380.27	4,730,357.95	967,606.20	1,176,745.00
12	23	14	10	4	40,056,409.58	91,467,281.67	20,676,958.00	20,950,050.00
13	30	18	11	7	18,319,257.35	40,504,730.32	7,566,826.00	7,635,893.00
14	36	22	14	8	13,309,614.29	15,606,443.68	2,774,224.00	2,967,401.00

ตารางที่ 4.2.5 (ต่อ) ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาแบบชั้นภูมิเบย์และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายของข้อมูลจริงจากการสำรวจค่าใช้จ่ายด้านการวิจัยและพัฒนาในภาคอุตสหกรรมของประเทศไทยประจำปี 2556 เฉพาะกิจการขนาดใหญ่

กิจการประเภทที่	รายละเอียด						ความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา	
	N	m	m_1	m_2	$\hat{\mu}_1$	$\hat{\sigma}_1$	การประมาณค่าแบบชั้นภูมิเบย์	การประมาณค่าด้วยการสุ่มตัวอย่างแบบง่าย
15	**86**	**27**	**10**	**17**	**9,513,941.87**	**9,194,256.19**	**1,320,435.00**	
16	183	28	11	17	7,750,444.87	9,587,171.90	1,335,375.00	1,321,533.00
17	88	26	15	11	63,406,586.67	210,034,780.49	31,751,243.00	31,454,940.00
18	151	25	11	14	27,617,480.00	49,440,159.25	7,147,865.00	6,969,181.00
19	48	15	2	13	5,847,500.00	4,012,830.98	676,869.10	599,091.00
20	**85**	**30**	**17**	**13**	**22,477,199.41**	**45,447,844.60**	**6,460,662.00**	**6,501,103.00**
21	136	32	12	20	28,106,717.50	45,744,088.14	5,505,605.00	5,401,485.00
22	246	36	17	19	24,353,639.76	40,590,111.37	5,051,924.00	5,014,454.00
23	**20**	**11**	**4**	**7**	**4,226,350.00**	**3,831,578.19**	**831,289.90**	**902,068.30**
24	43	14	2	12	2,269,975.00	1,994,076.48	323,489.30	265,297.80
25	202	29	11	18	20,734,843.64	35,131,135.19	4,493,553.00	4,337,574.00
26	47	14	2	12	384,668,876.19	526,609,114.91	78,025,687.00	54,013,640.00
27	**22**	**13**	**8**	**5**	**81,663,871.00**	**91,150,716.67**	**20,901,955.00**	**22,457,810.00**

บทที่ 5

อภิปรายผลและข้อเสนอแนะ

5.1 อภิปรายผล

การสุ่มตัวอย่างแบบชั้นภูมิเป็นเทคนิคการสุ่มตัวอย่างที่ใช้กันอย่างแพร่หลายในการประมาณค่าพารามิเตอร์ของประชากรซึ่งต้องทราบขนาดประชากรในแต่ละชั้นภูมิแต่ในทางปฏิบัติอาจพบกรณีไม่ทราบขนาดประชากรในแต่ละชั้นภูมิ ด้วยข้อจำกัดเช่นนี้เองบทความนี้จึงได้เสนออีกทางเลือกหนึ่งในการประมาณค่าพารามิเตอร์ของประชากรสำหรับการสุ่มตัวอย่างแบบชั้นภูมิเมื่อไม่ทราบขนาดประชากรในแต่ละชั้นภูมิโดยใช้การอนุมานทางสถิติแบบเบย์ และทำการศึกษาในกรณีที่ประชากรมี 2 ชั้นภูมิ โดยมีชั้นภูมิใดชั้นภูมิหนึ่งมีค่าสังเกตเป็นบวก และอีกชั้นภูมิหนึ่งมีค่าสังเกตเป็นศูนย์ เมื่อเปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ของประชากรโดยวิธีการประมาณค่าแบบชั้นภูมิเบย์ และการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายซึ่งพิจารณาจากความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยประชากรพบว่าการประมาณค่าแบบชั้นภูมิเบย์จะมีประสิทธิภาพดีกว่าการประมาณค่าด้วยการสุ่มตัวอย่างแบบง่ายเมื่อชั้นภูมิทั้ง 2 ชั้นภูมินั้นแยกออกจากกันอย่างชัดเจนซึ่งอาจเกิดจากการที่ค่าเฉลี่ยประชากรของชั้นภูมิที่มีค่าสังเกตเป็นบวกมีค่าสูงหรือความแปรปรวนของประชากรในชั้นภูมิที่มีค่าสังเกตเป็นบวกมีค่าต่ำ ดังนั้นวิธีที่นำเสนอจะเหมาะสมกับข้อมูลที่ไม่ทราบขนาดประชากรในแต่ละชั้นภูมิโดยที่แต่ละชั้นภูมิมีความแตกต่างกันอย่างชัดเจน

5.2 ข้อเสนอแนะ

จากวิทยานิพนธ์ฉบับนี้ผู้สนใจอาจนำไปศึกษาและพัฒนาระเบียบวิธีต่อได้อีกในเรื่องของ

1. การประมาณค่าพารามิเตอร์ของประชากรสำหรับการสุ่มตัวอย่างแบบชั้นภูมิเมื่อไม่ทราบขนาดประชากรในแต่ละชั้นภูมิแบบชั้นภูมิเบย์ในกรณีที่ประชากรมีมากกว่า 2 ชั้นภูมิ

2. กรณีที่การแจกแจงก่อนไม่ได้อยู่ในรูปการแจกแจงแบบสม่ำเสอ หรือการแจกแจงหลังไม่ได้อยู่ในรูปการแจกแจงแบบไฮเปอร์จีออเมเตริก

รายการอ้างอิง

ภาษาไทย

สำนักงานคณะกรรมการนโยบายวิทยาศาสตร์ เทคโนโลยีและนวัตกรรมแห่งชาติ (2554). *รายงานผลการสำรวจการวิจัยและพัฒนา และกิจกรรมนวัตกรรมภาคอุตสาหกรรมของประเทศไทย ประจำปี 2552*. กรุงเทพฯ.

ภาษาอังกฤษ

Bayes, M., & Price, M. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London,* 53(0), 370-418. doi:10.1098/rstl.1763.0053

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian Computation in Population Genetics. *GENETICS,* 162 (4), 2025-2035.

Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis*. New York Springers New York.

Berkopec, A. (2007). HyperQuick algorithm for discrete hypergeometric distribution. *Journal of Discrete Algorithms,* 5(2), 341-347. doi:10.1016/j.jda.2006.01.001

Gelman, A. (2002). Prior Distribution. *Encyclopedia of Environmetrics,* 3, 1634-1637.

Gelman, A. (2007). Struggles with Survey Weighting and Regression Modeling. *Statistical Science,* 22(2), 153-164. doi:10.1214/088342306000000691

Kish, L. (1965). *Survey Sampling*. New York, London John Wiley & Sons.

Najmussehar, M. J. A., & Bari, A. (2002). Double sampling for stratification with subsampling the non-respondents: a dynamic programming approach. *Aligarch Journal of Statistics,* 22, 27-41.

Okafor, F. C. (1994). On double sampling for stratification with sub-sampling the non-respondents. *Aligarch Journal of Statistics,* 14, 13-23.

Si, Y., Pillai, N. S., & Gelman, A. (2015). Bayesian Nonparametric Weighted Sampling Inference. *Bayesian Analysis,* 10(3), 605-625. doi:10.1214/14-ba924

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โปรแกรม R ที่ใช้ในวิเคราะห์ข้อมูล

1. การหาค่าประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนา และความคลาดเคลื่อนมาตรฐานของตัวประมาณแบบชั้นภูมิเบย์

$$\mathrm{E}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \mathrm{E}\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}} + \mathbf{Y_{12}} + \cdots + \mathbf{Y_{1M_1}}}{\mathbf{M_1}}\right) + \mathbf{W_2}\left(\frac{\mathbf{Y_{21}} + \mathbf{Y_{22}} + \cdots + \mathbf{Y_{2M_2}}}{\mathbf{M_2}}\right) \mid \mathbf{M_1} = \mathrm{m}_1\right)$$

$$\mathbf{SD}(\bar{\mathbf{Y}} \mid \mathbf{M_1} = \mathrm{m}_1) = \sqrt{\mathrm{Var}\left(\mathbf{W_1}\left(\frac{\mathbf{Y_{11}} + \mathbf{Y_{12}} + \ldots + \mathbf{Y_{1M_1}}}{\mathbf{M_1}}\right) + \mathbf{W_2}\left(\frac{\mathbf{Y_{21}} + \mathbf{Y_{22}} + \ldots + \mathbf{Y_{2M_2}}}{\mathbf{M_2}}\right) \mid \mathbf{M_1} = \mathrm{m}_1\right)}$$

เป็นวิเคราะห์ข้อมูลด้วยโปรแกรม R ด้วยความเร็ว 5-10 วินาที ต่อ 1 ชุดข้อมูลด้วยเครื่องคอมพิวเตอร์ Intel® Core™ i3-2350M Processor 2.30 GHz ที่ติดตั้ง Windows® 10

```
bayes <- function(n,m,m1,mu,sigma2){
b<-c()
d<-c()
m2<-m-m1
result1<-c()
result2<-c()
no<- m1:(n-m2)

for(i in 1:length(no)){
        b[i]<-choose(no[i],m1)*choose(n-no[i],m2)}
d<-sum(b)
result1<-no*(b/d)
e<-sum(result1)
a<-mu/n
ybar<-a*e
powerno<-no^2
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

```
result2<-powerno*(b/d)
f<-sum(result2)
s2<-(1/(n^2))*((sigma2/m1)+(mu^2))*f
var<-s2-(ybar^2)
se<-sqrt(var)
return(c(ybar, se))
}
```

2. การหาความแปรปรวนและความคลาดเคลื่อนมาตรฐานของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาด้วยการสุ่มตัวอย่างแบบง่ายจากความแปรปรวนของการสุ่มตัวอย่างแบบชั้นภูมิเบย์

$$\hat{\sigma}_1^{\,2} = \frac{\sum_{i=1}^{m_1}(Y_i - \hat{\mu}_1)^2}{m_1 - 1}$$

$$\sum_{i=1}^{m_1}(Y_i - \hat{\mu}_1)^2 = (m_1 - 1)\hat{\sigma}_1^{\,2}$$

$$\sum_{i=1}^{m_1} Y_i^{\,2} - m_1\,\hat{\mu}_1^{\,2} = (m_1 - 1)\hat{\sigma}_1^{\,2}$$

$$\sum_{i=1}^{m_1} Y_i^{2} = (m_1 - 1)\hat{\sigma}_1^{\,2} + m_1\,\hat{\mu}_1^{\,2}$$

เพิ่ม $Y_i = 0$ ซึ่งเป็นค่าใช้จ่ายของบริษัทที่ไม่มีกิจกรรมวิจัยและพัฒนาอีก $m_2 = m - m_1$ ตัว ดังนั้นจะได้ค่าเฉลี่ยกำลังสองของข้อมูล m ตัว ดังนี้

$$\frac{\sum_{i=1}^{m_1} Y_i^{\,2} + 0+0+0+\cdots+0}{m} = \frac{(m_1 - 1)\hat{\sigma}_1^{\,2} + m_1\,\hat{\mu}_1^{\,2}}{m}$$

ดังนั้น

$$\hat{\sigma}^2_{\text{simple}} = \frac{m}{m-1}\left(\frac{(m_1 - 1)\hat{\sigma}_1^{\,2} + m_1\,\hat{\mu}_1^{\,2}}{m} - \left(\frac{m_1\hat{\mu}_1}{m}\right)^2\right)$$

โดย $\hat{\sigma}^2_{\text{simple}}$ = ความแปรปรวนของตัวประมาณค่าเฉลี่ยค่าใช้จ่ายด้านการวิจัยและพัฒนาด้วยการสุ่มตัวอย่างแบบง่าย

เป็นวิเคราะห์ข้อมูลด้วยโปรแกรม R ด้วยความเร็ว 3-8 วินาที ต่อ 1 ชุดข้อมูลด้วยเครื่องคอมพิวเตอร์ Intel® Core™ i3-2350M Processor 2.30 GHz ที่ติดตั้ง Windows® 10

```
simple <- function(m,m1,mu,sigma2){
g<-m/(m-1)
h<-(((m1-1)*sigma2)+(m1*mu^2))/m
j<-((m1*mu)/m)^2
varsimp<-g*(h-j)
sesimp<-sqrt(varsimp/m)
return(c(varsimp ,sesimp))
}
```

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียนวิทยานิพนธ์

นายสุวัฒน์ ปลื้มอารมณ์ เกิดเมื่อวันที่ 29 กุมภาพันธ์ พ.ศ. 2523 สำเร็จการศึกษาระดับปริญญาวิทยาศาสตรบัณฑิต (วท. บ.) สาขาวิทยาศาสตร์ทางทะเล ภาควิชาวิทยาศาสตร์ทางทะเล คณะประมง มหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2544 และสำเร็จการศึกษาระดับปริญญาวิทยาศาสตรมหาบัณฑิต (วท. ม.) สาขาวิทยาศาสตร์ทางทะเล ภาควิชาวิทยาศาสตร์ทางทะเล คณะประมง มหาวิทยาลัยเกษตรศาสตร์ ในปีการศึกษา 2552 และเข้ารับการศึกษาในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท. ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2556

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University