

การประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลีของข้อมูลแบบจับคู่

นายภูมิพงศ์ แก้วอำไพ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

INTERVAL ESTIMATION FOR THE DIFFERENCE BETWEEN TWO BERNOULLI  
PROPORTIONS OF PAIRED DATA

Mr. Poompong Kaewumpai

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Statistics Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2007

Copyright of Chulalongkorn University

501010

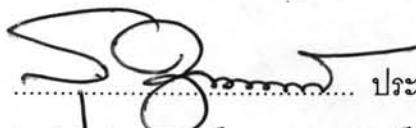
หัวข้อวิทยานิพนธ์      การประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลีของข้อมูล  
แบบจับคู่  
โดย                              นายภูมิพงศ์ แก้วอำไพ  
สาขาวิชา                      สถิติ  
อาจารย์ที่ปรึกษา              รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์


---

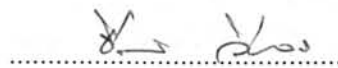
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย  
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี  
(รองศาสตราจารย์ ดร.อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

 ประธานกรรมการ  
(รองศาสตราจารย์ ดร. สุกุล ดุรงค์วัฒนา)

 อาจารย์ที่ปรึกษา  
(รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์)

 กรรมการ  
(รองศาสตราจารย์ ดร. ธีระพร วีระถาวร)

ภูมิพงศ์ แก้วอำไพ: การประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลีของข้อมูลแบบจับคู่.(INTERVAL ESTIMATION FOR THE DIFFERENCE BETWEEN TWO BERNOULLI PROPORTIONS OF PAIRED DATA) อ.ที่ปรึกษา: รศ. ร้อยเอก มานพ วรภักดี, 170 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบการประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลีของข้อมูลแบบจับคู่โดยการเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลอง และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น วิธีการประมาณที่ใช้ในการวิจัยในครั้งนี้คือ วิธีการประมาณของ Wald วิธีการประมาณของ Newcombe วิธีการประมาณของ May และ Johnson วิธีการประมาณของ Zhou และ Qin การเปรียบเทียบทำภายใต้สถานการณ์ค่าสัมประสิทธิ์สหสัมพันธ์ระดับสูง กลาง ต่ำ โดยที่ขนาดตัวอย่างของประชากรมีขนาดเท่ากันคือ 10,20,30,40,50,60,70,80 ค่าสัดส่วนประชากรที่ 1 และ 2 ( $p_0, p_1$ ) ให้มีค่าตั้งแต่ 0.1 ถึง 0.9 โดยเพิ่มค่าทีละ 0.1 ซึ่งจะให้ค่าผลต่างระหว่างค่าสัดส่วนของสองประชากร ( $p_1 - p_0$ ) มีความแตกต่างกันตั้งแต่ 0 ถึง 0.8 โดยเพิ่มค่าทีละ 0.1 กำหนดให้  $p_1 \geq p_0$  กำหนดระดับความเชื่อมั่น 3 ระดับคือ 90%, 95% และ 99% การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โลและทำการทดลอง 2000 ครั้ง ในแต่ละสถานการณ์ที่กำหนด ผลการวิจัยสรุปได้ดังนี้

เมื่อขนาดตัวอย่างของสองประชากรมีขนาดเล็กและค่าสัมประสิทธิ์สหสัมพันธ์ทุกระดับ วิธีการประมาณของ May และ Johnson ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดเมื่อ  $p_1 - p_0$  มีค่าน้อย(มีค่าเข้าใกล้ 0) และวิธีการประมาณของ Newcombe ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดเมื่อ  $p_1 - p_0$  มีค่าปานกลาง จนถึงมีค่ามาก(มีค่าเข้าใกล้ 1)

เมื่อขนาดตัวอย่างของสองประชากรมีขนาดกลางและค่าสัมประสิทธิ์สหสัมพันธ์ทุกระดับ วิธีการประมาณของ May และ Johnson ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดเมื่อ  $p_1 - p_0$  มีค่าน้อย(มีค่าเข้าใกล้ 0) แต่เมื่อขนาดตัวอย่างของสองประชากรมีขนาดกลางและค่าสัมประสิทธิ์สหสัมพันธ์ไม่สูง วิธีการประมาณของ Zhou และ Qin ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดเมื่อ  $p_1 - p_0$  มีค่าปานกลางหรือมาก(เข้าใกล้ 1)

เมื่อขนาดตัวอย่างของสองประชากรมีขนาดใหญ่และค่าสัมประสิทธิ์สหสัมพันธ์ทุกระดับ วิธีการประมาณของ Wald วิธีการประมาณของ May และ Johnson และ วิธีการประมาณของ Zhou และ Qin ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่ต่ำใกล้เคียงกันเมื่อ  $p_1 - p_0$  มีค่าน้อย(มีค่าเข้าใกล้ 0) แต่เมื่อขนาดตัวอย่างของสองประชากรมีขนาดใหญ่และค่าสัมประสิทธิ์สหสัมพันธ์ไม่สูง วิธีการประมาณของ Wald วิธีการประมาณของ Newcombe และ วิธีการประมาณของ Zhou และ Qin ให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่ต่ำใกล้เคียงกันเมื่อ  $p_1 - p_0$  มีค่าปานกลางหรือมาก(เข้าใกล้ 1)

ภาควิชา สถิติ

สาขาวิชา สถิติ

ปีการศึกษา 2550

ลายมือชื่อนิสิต.....

ภูมิพงศ์ แก้วอำไพ

ลายมือชื่ออาจารย์ที่ปรึกษา.....

Manop Worapakdi

##4882236126 : MAJOR STATISTICS

KEY WORD: PROPORTION/ BERNOULLI/ INTERVAL ESTIMATION/ PAIRED DATA

POOMPONG KAEWUMPAI: INTERVAL ESTIMATION FOR THE DIFFERENCE BETWEEN TWO BERNOULLI PROPORTIONS OF PAIRED DATA. THESIS ADVISOR :ASSOC. PROF. CAPT. MANOP VARAPHA KDI, 170 pp.

The objective of this thesis is to study and compare the interval estimation methods for the difference between two Bernoulli proportions of paired data by comparing their confidence coefficients and average confidence interval lengths. The estimation methods under consideration in this study are Wald's method, Newcombe's method, May and Johnson's method and Zhou and Qin's method. The comparison was done under condition of level of correlation coefficients(low, medium and high) and two sample sizes are 10,20,30,40,50,60,70,80 two population proportions  $(p_0, p_1)$  are ranging from 0 to 0.9 increasing by 0.1 and difference between two population proportions  $(p_1 - p_0)$  are ranging from 0 to 0.8 increasing by 0.1 and  $p_1 \geq p_0$ , all of which are considered at 90%, 95% and 99% confidence levels. The simulation of this research is repeated 2000 times in each situation by using the Monte Carlo Simulation method. The conclusion of this study are as follow :

If two sample sizes are small and all level of correlation coefficients, the confidence levels of May and Johnson's method are not lower than the given confidence levels and the average confidence interval lengths are shortest when  $p_1 - p_0$  is small(near 0) and the confidence levels of Newcombe's method are not lower than the given confidence levels and the average confidence interval lengths are shortest when  $p_1 - p_0$  is moderate to large(near 1).

If two sample sizes are medium and all level of correlation coefficients, the confidence levels of May and Johnson's method are not lower than the given confidence levels and the average confidence interval lengths are shortest when  $p_1 - p_0$  is small(near 0) but if two sample sizes are medium and level of correlation coefficient is low or medium, the confidence levels of Zhou and Qin's method are not lower than the given confidence levels and the average confidence interval lengths are shortest when  $p_1 - p_0$  is moderate or large(near 1).

If two sample sizes are large and all level of correlation coefficients, the confidence levels of Wald's method, May and Johnson's method and Zhou and Qin's method are not lower than the given confidence levels and the average confidence interval lengths are nearly the same short when  $p_1 - p_0$  is small(near 0) but if two sample sizes are large and level of correlation coefficient is low or medium, the confidence levels of Wald's method, Newcombe's method and Zhou and Qin's method are not lower than the given confidence levels and the average confidence interval lengths are nearly the same short when  $p_1 - p_0$  is moderate or large(near 1).

Department : Statistics  
Field of Study: Statistics  
Academic Year : 2007

Student's Signature.....  
Advisor's Signature.....

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือเป็นอย่างดีจาก รองศาสตราจารย์ ร้อยเอกมานพ วรภักดิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ปรึกษา ตลอดจนช่วยเหลือตรวจสอบแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดีจนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณด้วยความรู้สึกซาบซึ้ง เคารพและสำนึกในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สุกพล ตุงศ์วัฒนา และ รองศาสตราจารย์ ดร. ธีระพร วีระถาวร ประธานกรรมการและกรรมการสอบวิทยานิพนธ์ ที่กรุณาตรวจสอบแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบคุณคณาจารย์ทุกท่านที่กรุณาถ่ายทอดความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

สุดท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา ที่ให้การส่งเสริม สนับสนุน ด้านทุนการศึกษา รวมทั้งให้ความรัก และให้กำลังใจแก่ผู้วิจัยมาโดยตลอดจนสำเร็จการศึกษา คุณค่าและประโยชน์ของวิทยานิพนธ์ฉบับนี้ผู้วิจัยขอมอบแด่ บิดา มารดาและผู้มีพระคุณทุกท่าน

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	บ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์การวิจัย.....	5
1.3 สมมติฐานของการวิจัย.....	5
1.4 ขอบเขตเบื้องต้น.....	5
1.5 ขอบเขตของการวิจัย.....	6
1.6 เกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของช่วงความเชื่อมั่น.....	6
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
2 ทฤษฎีที่เกี่ยวข้องกับการวิจัย.....	8
2.1 การแจกแจงแบร์นูลลี.....	8
2.2 การแจกแจงทวินาม.....	8
2.3 ความน่าจะเป็นที่มีเงื่อนไข.....	9
2.4 ความแปรปรวนร่วมและสหสัมพันธ์.....	9
2.5 การแจกแจงแบร์นูลลีของข้อมูลแบบจับคู่.....	10
2.6 การประมาณค่าพารามิเตอร์ของการแจกแจงแบร์นูลลีของข้อมูลแบบจับคู่.....	11
2.7 การประมาณค่าแบบช่วง.....	14
2.8 ทฤษฎีบทลิมิตเข้าสู่ส่วนกลาง.....	14
2.9 วิธีการประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลีของข้อมูลแบบจับคู่.....	15
2.10 เกณฑ์การเปรียบเทียบ.....	18
3 วิธีดำเนินการวิจัย.....	21
3.1 ข้อกำหนดของการทดลอง.....	21

บทที่	หน้า
3.2	ขั้นตอนดำเนินการวิจัย..... 23
3.3	ขั้นตอนในการทำงานของโปรแกรม..... 30
4	ผลการวิจัย..... 32
4.1	การเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลองของช่วง ความเชื่อมั่น..... 33
4.2	การเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น..... 88
5	สรุปผลการวิจัยและข้อเสนอแนะ..... 152
5.1	สรุปผลการวิจัย..... 153
5.2	ปัจจัยที่มีอิทธิพลต่อความยาวเฉลี่ยของช่วงความเชื่อมั่นสำหรับค่า สัดส่วนเบร์นูลลีของข้อมูลแบบจับคู่..... 154
5.3	ข้อเสนอแนะ..... 155
	รายการอ้างอิง..... 160
	ภาคผนวก..... 161
	ประวัติผู้เขียนวิทยานิพนธ์..... 170































4.2.44 แสดงค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณ 4 วิธี ที่ระดับความเชื่อมั่น 95% ในแต่ละระดับขนาดตัวอย่าง เมื่อระดับผลต่างค่าสัดส่วน ของประชากร(Diff) มีค่าเท่ากับ 0.1 และสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ สูง $0.6 < r \leq 1$ .....	145
4.2.45 แสดงค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณ 4 วิธี ที่ระดับความเชื่อมั่น 95% ในแต่ละระดับขนาดตัวอย่าง เมื่อระดับผลต่างค่าสัดส่วน ของประชากร(Diff) มีค่าเท่ากับ 0.2 และสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ สูง $0.6 < r \leq 1$ .....	146
4.2.46 แสดงค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณ 4 วิธี ที่ระดับความเชื่อมั่น 99% ในแต่ละระดับขนาดตัวอย่าง เมื่อระดับผลต่างค่าสัดส่วน ของประชากร(Diff) มีค่าเท่ากับ 0 และสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ สูง $0.6 < r \leq 1$ .....	148
4.2.47 แสดงค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณ 4 วิธี ที่ระดับความเชื่อมั่น 99% ในแต่ละระดับขนาดตัวอย่าง เมื่อระดับผลต่างค่าสัดส่วน ของประชากร(Diff) มีค่าเท่ากับ 0.1 และสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ สูง $0.6 < r \leq 1$ .....	150
4.2.48 แสดงค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นที่คำนวณได้จากวิธีการประมาณ 4 วิธี ที่ระดับความเชื่อมั่น 99% ในแต่ละระดับขนาดตัวอย่าง เมื่อระดับผลต่างค่าสัดส่วน ของประชากร(Diff) มีค่าเท่ากับ 0.2 และสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ สูง $0.6 < r \leq 1$ .....	151

สารบัญภาพ

รูปที่	หน้า
3.1	แผนผังสำหรับขั้นตอนในการดำเนินงานวิจัย.....30
5.1	แผนภูมิผลสรุปการเลือกวิธีการประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลี ของข้อมูลแบบจับคู่เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ ต่ำ $0 \leq r < 0.4$ .....156
5.2	แผนภูมิผลสรุปการเลือกวิธีการประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลี ของข้อมูลแบบจับคู่เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับกลาง $0.4 \leq r \leq 0.6$ .157
5.3	แผนภูมิผลสรุปการเลือกวิธีการประมาณค่าแบบช่วงของผลต่างค่าสัดส่วนแบร์นูลลี ของข้อมูลแบบจับคู่เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ในระดับ สูง $0.6 < r \leq 1$ .....158