

การวิเคราะห์เชิงเบส์สำหรับตัวแบบการทดสอบเชิงเส้นเชิงเดียว



นางสาววีรพา ฐานะปรัชญ์

วิทยานิพนธ์เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปฏิญาณคิติศาสตร์บัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2542

ISBN 974-334-078-5

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

BAYESIAN ANALYSIS FOR SIMPLE LINEAR REGRESSION MODEL

Miss Weerapa Thanapraph

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 1999

ISBN 974-334-078 -5

หน้าอีกที่นักวิทยานิพนธ์ การวิเคราะห์เชิงเบสสำหรับตัวแบบการทดสอบเชิงเส้นเชิงเดียว

โดย นางสาววีรพา ฐานะปัจจุบัน

ภาควิชา สถิติ

อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.ธีระพงษ์ วีระถาวร

คณะกรรมการคัดเลือกและประเมินคุณภาพของสถาบันอุดมศึกษา อนุมัติให้บัณฑิตวิทยานิพนธ์
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรบริบูรณ์ตามที่บันทึก

วีระพงษ์ คณะกรรมการคัดเลือกและประเมินคุณภาพ
(ผู้ช่วยศาสตราจารย์ ดร.วีระพงษ์ วีระถาวร)

คณะกรรมการสอนวิทยานิพนธ์

ธีระพงษ์ ประธานกรรมการ
(รองศาสตราจารย์ ศิริพงษ์ ชาเกทอง)

วีระพงษ์ อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.ธีระพงษ์ วีระถาวร)

คงสุข ธรรมรงค์ กรรมการ
(รองศาสตราจารย์ ผกวดต์ ศิรัชช์)

สุรัตน์ ธรรมรงค์ กรรมการ
(รองศาสตราจารย์ ดร.สุรัตน์ ธรรมรงค์)

จุฬาลงกรณ์มหาวิทยาลัย

**วิรพา ฐานะปรัชญา : การวิเคราะห์เชิงเบสสำหรับตัวแบบการ回帰โดยใช้เส้นเชิงเดียว
(BAYESIAN ANALYSIS FOR SIMPLE LINEAR REGRESSION MODEL)**
อ.ที่ปรึกษา : ดร.ธีระพงษ์ วีระถาวร , 222 หน้า. ISBN 974-334-078-5.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบค่าประมาณสัมประสิทธิ์การถดถอยของตัวแปร การถดถอยเชิงเส้นเชิงเดียวจาก 4 วิธี คือ วิธีก่อลังกลอนออยสุด (OLS) วิธีเชิงเบสเมื่อใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล (UNI) วิธีเชิงเบสเมื่อใช้การแจกแจงก่อนที่ให้ข้อมูล (NOR) และวิธีเชิงเบสเมื่อใช้การแจกแจงก่อนของเจฟเฟอร์ส (JEF) เกณฑ์ที่ใช้ในการตัดสินใจ คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (AMSE) และวัตถุประสงค์ของการที่จะเชื่อเพื่อศึกษาว่า NOR ในกรณีการหาค่า Z ที่เหมาะสมซึ่งทำให้วิธี NOR มีประสิทธิภาพใกล้เคียงกับวิธี OLS และ UNI มากที่สุด โดยที่ค่า Z เป็นค่าที่ทำให้ค่าเฉลี่ยก่อน (prior mean) เปี่ยมเบนจากค่าสัมประสิทธิ์การถดถอย Z เท่าของส่วนเบี่ยงเบนมาตรฐานก่อน (prior standard deviation) ซึ่งวัตถุประสงค์ทั้งสองคือการตั้งค่าไว้ที่ศึกษาเมื่อขนาดตัวอย่างเท่ากับ 10 30 50 และ 100 ความคลาดเคลื่อนมีการแจกแจงปกติซึ่งมีค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.3 0.5 0.7 และ 0.9 ตัวแปรอิสระเป็นตัวคงที่ซึ่งถูมารากการแจกแจงปกติตัวอย่างค่าเฉลี่ยเท่ากับ 1 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.1 0.15 0.2 0.25 0.3 0.5 0.7 และ 0.9 (เบอร์เรียนด์ของสัมประสิทธิ์ความแปรปัน (CV(X)) เท่ากับ 10% 15% 20% 25% 30% 50% 70% และ 90% ตามลำดับ) ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองด้วยเทคนิคคอมพิวเตอร์ซึ่งกระทำมา 500 ครั้งในแต่ละสถานการณ์

จากการพิสูจน์ทางพิชิตคณิตแสดงให้เห็นว่าการแจกแจงภายหลัง (posterior distribution) ที่ได้จากการวิธีเชิงเบส เมื่อใช้การแจกแจงก่อนของเจฟเฟอร์ส (Jeffreys's prior distribution) เนื่องจากนักการแจกแจงภายนอกที่ได้จากการวิธีเชิงเบสเมื่อใช้การแจกแจงก่อนที่ไม่ให้ข้อมูล (noninformative prior distribution) ดังนั้น ผู้วิจัยจึงเปรียบเทียบค่าประมาณสัมประสิทธิ์การถดถอยของ 3 วิธี คือ วิธี OLS วิธี UNI และวิธี NOR

ผลการวิจัยจากวัตถุประสงค์ของการทดลองว่า วิธี OLS และวิธี UNI มีประสิทธิภาพใกล้เคียงกัน โดยที่วิธี OLS จะมีประสิทธิภาพต่ำกว่าวิธี UNI เสมอ ซึ่งวิธี OLS มีประสิทธิภาพต่ำกว่าวิธี NOR มีค่าสูง ส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามมีค่าต่ำและขนาดตัวอย่างสูง ส่วนวิธี NOR มีประสิทธิภาพต่ำกว่าวิธี OLS มีค่าต่ำ ส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามมีค่าสูงและขนาดตัวอย่างต่ำ โดยประสิทธิภาพของตัวประมาณสัมประสิทธิ์การถดถอยของทั้งสามวิธีมีแนวโน้มเพิ่มขึ้นเมื่อ CV(X) และขนาดตัวอย่างเพิ่มขึ้น และมีแนวโน้มลดลงเมื่อส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามเพิ่มขึ้น

ผลการวิจัยจากวัตถุประสงค์ของการที่จะทราบว่า ค่า Z ที่เหมาะสมที่สุดคือส่วนเบี่ยงเบนมาตรฐานของตัวแปรตาม ထ้วย่อมต้องมีแนวโน้มเพิ่มขึ้น CV(X) และขนาดตัวอย่าง วิธี NOR มีแนวโน้มต่ำกว่าวิธี OLS และ UNI เมื่อค่า Z ที่เหมาะสมมีแนวโน้มเพิ่มขึ้น แต่วิธี NOR มีแนวโน้มไม่ต่ำกว่าวิธี OLS และ UNI เมื่อค่า Z ที่เหมาะสมมีแนวโน้มลดลง

ภาควิชา.....	สถิติ.....	ลายมือชื่อ	วิรพา ฐานะปรัชญา'
สาขาวิชา	สถิติ.....	ลายมือชื่ออาจารย์ที่ปรึกษา
ปีการศึกษา	2542.....	ลายมือชื่ออาจารย์ที่ปรึกษารวม

3971763526 : MAJOR STATISTICS

KEY WORD : Posterior Distribution / Noninformative Prior / Informative Prior / Jeffreys's Prior

WEERAPA THANAPRACH : BAYESIAN ANALYSIS FOR SIMPLE LINEAR
REGRESSION MODEL.

THESIS ADVISOR : ASSOC. PROF. THEERAPORN VERATHAWORN, Ph.D.

222 pp. ISBN 974-334-078 - 5.

The first objective of this research is to compare four methods for estimating the regression coefficients of the simple linear regression model : Least Square method (OLS), Bayesian method using noninformative prior (UNI), Bayesian method using informative prior (NOR) and Bayesian method using Jeffreys's prior (JEF). The criterion of comparison is the ratio of the average value of the mean square error (AMSE). The second objective of this research is to study the NOR method in searching for the optimum z-value and whether the NOR method has the nearest efficiency to OLS and UNI where z-value is the value that makes prior mean deviate z-fold of prior standard deviation from the regression coefficient. Both of the objectives have studied sample sizes of 10, 30, 50 and 100. The parameters of the research has been the residuals of a normal distribution with a mean of 0, a standard deviation of 0.1, 0.3, 0.5, 0.7 and 0.9, and the independent variable is a scalar drawn from a normal distribution with a mean of 1, a standard deviation of 0.1, 0.15, 0.2, 0.25, 0.3, 0.5, 0.7 and 0.9 (percentage of coefficients of variation (CV(X)) are equal to 10%, 15%, 20%, 25%, 30%, 50%, 70% and 90%, respectively). The data has been obtained through simulation using Monte Carlo technique and repeated 500 times for each case.

By algebraic proof, this shows that the posterior distribution by using Jeffreys's prior distribution has the same distribution as the posterior distribution by using a noninformative prior distribution, so the researcher has compared the regression coefficient estimate of three methods : OLS, UNI and NOR.

The results of the first objective are that OLS has a nearer efficiency with UNI while OLS always has a greater efficiency than UNI. OLS is efficient when CV(X) is high, the standard deviation of the dependent variable is low, and the sample size is large. NOR is efficient when CV(X) is low, the standard deviation of the dependent variable is high, and the sample size is small. The efficiency of all regression coefficient estimates increased when CV(X) and the sample size increased. However the efficiency decreased when the standard deviation of the dependent variable increased.

The results of the second objective are that the optimum z-value follows the standard deviation of the dependent variable, whereas it is in converse to CV(X) and the sample size. NOR is better than OLS and UNI when the optimum z-value increases but NOR is worse than OLS and UNI when the optimum z-value decreases.

ภาควิชา.....	สถิติ.....	นายมีชัย วีระ ฐานะ-ปัจฉน์.....
สาขาวิชา	สถิติ.....	นายมีชัย อาจารย์ที่ปรึกษา
ปีการศึกษา	2542.....	นายมีชัย อาจารย์ที่ปรึกษาอ่วน

กิตติกรรมประกาศ

ขอขอบพระคุณรองศาสตราจารย์ ดร.ธีระพงษ์ วีระถาวร เป็นอย่างสูงในความกรุณาที่ให้
ความช่วยเหลือและแก้ไขข้อบกพร่องต่างๆ ในวิทยานิพนธ์เล่มนี้ให้สามารถสำเร็จสมบูรณ์ และใน
ความเมตตาที่มีต่อศิษย์ในการอนุมให้มีความคิดและปรับปรุงตนเองไปในทางที่ดีมาโดยตลอด
ขอขอบพระคุณผู้ช่วยศาสตราจารย์ ร้อยเอกมานพ วงศ์กัตต์ ที่ให้คำแนะนำทางด้านการเขียนของ
ปัญหาด้วยการให้โปรแกรมคอมพิวเตอร์ซึ่งเป็นส่วนสำคัญที่จะให้มาชื่นชมการวิจัยที่สมบูรณ์ และ
ขอขอบพระคุณคณะกรรมการตอบวิทยานิพนธ์ทุกท่านที่ได้กรุณาตรวจสอบและแก้ไขให้วิทยา
นิพนธ์เล่มนี้มีความสมบูรณ์ยิ่งขึ้น ตลอดจนอาจารย์ภาควิชาสถิติทุกท่านที่ได้ประสิทธิประสาทวิชา
ความรู้ให้แก่ศิษย์

สุดท้ายนี้ ขอขอบพระคุณบุคลากรที่ได้โอกาสทางการศึกษา และขอบคุณสมาชิกทุกคน
ในครอบครัวและเพื่อนๆ ทุกคนที่เป็นกำลังใจเสมอมา

วีระพงษ์ วีระถาวร

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	๔
บทคัดย่อภาษาอังกฤษ	๕
กิตติกรรมประกาศ	๖
สารบัญ	๗
สารบัญตาราง	๘
สารบัญรูป	๙
สารบัญแผนผัง	๑๐
บทที่	
1 บทนำ	1
1.1 ความสำคัญและที่มาของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 สมมติฐานของการวิจัย	2
1.4 ขอบเขตของการวิจัย	2
1.5 เกณฑ์การตัดสินใจ	5
1.6 ประโยชน์ที่คาดว่าจะได้รับ	5
1.7 ตัวสถิติที่ใช้ในการวิจัย	5
2 ทฤษฎีและตัวแสตติที่เกี่ยวข้อง	8
2.1 ศักดิ์ประมาณกำลังสองน้อยตุด	8
2.2 การแยกแยะภายนลังเมื่อใช้การแยกแยะก่อนที่ให้ร้อมูลและการแยกแยะก่อนที่ไม่ให้ร้อมูล ในกรณีที่ σ^2 ทราบค่า	11
2.3 การแยกแยะภายนลังเมื่อใช้การแยกแยะก่อนที่ให้ร้อมูลและการแยกแยะก่อนที่ไม่ให้ร้อมูล ในกรณีที่ σ^2 ไม่ทราบค่า	16
2.4 การแยกแยะภายนลังเมื่อใช้การแยกแยะก่อนของเชฟเฟอร์ส	20
3 วิธีดำเนินการวิจัย	23
3.1 แผนกรากเดลลง	23
3.2 ขั้นตอนการวิจัย	24
3.3 การประมาณค่าสามประสิทธิ์การทดสอบจากแต่ละวิธี	24
3.4 การหาค่า Z ที่เหมาะสม	25

สารบัญ (ต่อ)

หน้า

บทที่

3.5 ค่าเฉลี่ยความคาดเดือนกำลังสองและการเปรียบเทียบประสิทธิภาพของตัวประเมินสัมประสิทธิ์การทดสอบ 26
4 ผลการวิจัย 31
4.1 การเปรียบเทียบตัวประมาณสัมประสิทธิ์การทดสอบ เมื่อตัวประมาณเบสที่ใช้การแยกแยะก่อนที่ให้รับมุลศึกษาในกรณีเฉพาะ 32
4.2 การหาตัวประมาณเบสที่ใช้การแยกแยะก่อนที่ให้รับมุลที่ไม่ดีที่สุด แต่ดีกว่าตัวประมาณกำลังสองน้อยสุดและตัวประมาณเบสที่ใช้การแยกแยะก่อนที่ไม่ให้รับมุล 105
5 สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ 171
5.1 สรุปผลการวิจัย 171
5.1.1 ผลสรุปจากการเปรียบเทียบตัวประมาณสัมประสิทธิ์การทดสอบ เมื่อตัวประมาณเบสที่ใช้การแยกแยะก่อนที่ให้รับมุลศึกษาในกรณีเฉพาะ 171
5.1.2 ผลสรุปจากการหาตัวประมาณเบสที่ใช้การแยกแยะก่อนที่ให้รับมุลที่ไม่ดีที่สุด แต่ดีกว่าตัวประมาณกำลังสองน้อยสุดและตัวประมาณเบสที่ใช้การแยกแยะก่อนที่ไม่ให้รับมุล 173
5.2 การอภิปรายผล 174
5.3 ข้อเสนอแนะ 175
รายงานยังคง 182
ภาคผนวก 183
ประวัติผู้วิจัย 222

สารบัญตราง

ສາທິປະໄຕ (ຕໍ່ປ)

ສາຂະບັດທາງ (ຕ່ອ)

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.2.12	การศึกษาตัวประมาณแบบที่ใช้การแยกแจงก่อนที่ให้ข้อมูล เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสูมมาจาก การแยกแจงปากติโดยมี $CV(X) = 25\%$ เมื่อไม่ทราบค่าความแปรปรวน โดยศึกษาในกรณีการหาค่า Z ที่เหมาะสม	147
4.2.13	การศึกษาตัวประมาณแบบที่ใช้การแยกแจงก่อนที่ให้ข้อมูล เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสูมมาจาก การแยกแจงปากติโดยมี $CV(X) = 30\%$ เมื่อไม่ทราบค่าความแปรปรวน โดยศึกษาในกรณีการหาค่า Z ที่เหมาะสม	150
4.2.14	การศึกษาตัวประมาณแบบที่ใช้การแยกแจงก่อนที่ให้ข้อมูล เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสูมมาจาก การแยกแจงปากติโดยมี $CV(X) = 50\%$ เมื่อไม่ทราบค่าความแปรปรวน โดยศึกษาในกรณีการหาค่า Z ที่เหมาะสม	153
4.2.15	การศึกษาตัวประมาณแบบที่ใช้การแยกแจงก่อนที่ให้ข้อมูล เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสูมมาจาก การแยกแจงปากติโดยมี $CV(X) = 70\%$ เมื่อไม่ทราบค่าความแปรปรวน โดยศึกษาในกรณีการหาค่า Z ที่เหมาะสม	156
4.2.16	การศึกษาตัวประมาณแบบที่ใช้การแยกแจงก่อนที่ให้ข้อมูล เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสูมมาจาก การแยกแจงปากติโดยมี $CV(X) = 90\%$ เมื่อไม่ทราบค่าความแปรปรวน โดยศึกษาในกรณีการหาค่า Z ที่เหมาะสม	159
5.1.1	วิธีที่ให้ประสิทธิภาพดีที่สุด ในแต่ละสถานการณ์ ในกรณีที่ σ^2 ทราบค่า	172
5.1.2	วิธีที่ให้ประสิทธิภาพดีที่สุด ในแต่ละสถานการณ์ ในกรณีที่ σ^2 ไม่ทราบค่า	173

**สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย**

ສາບັດງານ

ສາທິປະລຸງ (ຕ່ອ)

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.2.13 การเปลี่ยนแปลงของค่า Z เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มมาจากการแจก- แจงปกติ โดยมี $CV(X) = 30\%$ โดยที่ไม่ทราบค่าความแปรปรวน 168	168
4.2.14 การเปลี่ยนแปลงของค่า Z เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มมาจากการแจก- แจงปกติ โดยมี $CV(X) = 50\%$ โดยที่ไม่ทราบค่าความแปรปรวน 168	168
4.2.15 การเปลี่ยนแปลงของค่า Z เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มมาจากการแจก- แจงปกติ โดยมี $CV(X) = 70\%$ โดยที่ไม่ทราบค่าความแปรปรวน 169	169
4.2.16 การเปลี่ยนแปลงของค่า Z เมื่อตัวแปรอิสระเป็นค่าคงที่ซึ่งสุ่มมาจากการแจก- แจงปกติ โดยมี $CV(X) = 90\%$ โดยที่ไม่ทราบค่าความแปรปรวน 169	169

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญแผนผัง

แผนผังที่	หน้า
1 การเขียนโปรแกรมในการหาค่าประมาณสัมประสิทธิ์การลดด้อยและค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง.....	29
2 การเลือกวิธีประมาณค่าสัมประสิทธิ์การลดด้อยที่ให้ค่า AMSE ต่ำสุดในสถานการณ์ต่างๆ เมื่อ σ^2 ทราบค่า.....	175
3 การเลือกวิธีประมาณค่าสัมประสิทธิ์การลดด้อยที่ให้ค่า AMSE ต่ำสุดในสถานการณ์ต่างๆ เมื่อ σ^2 ไม่ทราบค่า.....	178

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย