การเปรียบเทียบการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงเมื่อข้อมูลถูกตัดปลาย

นางสาวชิวา กันยะมี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-14-2346-2

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

I18261393

A COMPARISON ON GOODNESS-OF-FIT TESTS FOR DISTRIBUTION WITH CENSORED DATA

Miss Tiwa Kanyamee

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 2005

ISBN 974-14-2346-2

หัวข้อวิทยานิพนธ์	การเปรียบเทียบการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงเมื่อข้อมูลถูกตัดปลาย
โดย	นางสาวธิวา กันยะมี
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี
(ผู้ช่วยศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.....อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์)

.....กรรมการ
(รองศาสตราจารย์ ดร.ธีระพร วีระถาวร)

.....กรรมการ
(อาจารย์ ดร.เสกสรร เกียรติสุไพบูลย์)

ธิวา กันยะมี : การเปรียบเทียบการทดสอบเทียบความกลมกลืนสำหรับการแจกแจงเมื่อข้อมูลถูกตัดปลาย. (A COMPARISON ON GOODNESS-OF-FIT TESTS FOR DISTRIBUTION WITH CENSORED DATA) อ. ที่ปรึกษา : รศ. ร.อ.มานพ วราภักดิ์, 199 หน้า. ISBN 974-14-2346-2.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบเทียบความกลมกลืนสำหรับการแจกแจงแกมมาเมื่อข้อมูลถูกตัดปลายประเภทที่ 2 ตัวสถิติที่ใช้ในการทดสอบ คือ ตัวสถิติทดสอบ K-S^C ตัวสถิติทดสอบ AD^C และตัวสถิติทดสอบ CVM^C โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ เมื่อประชากรมีการแจกแจงแกมมา การแจกแจงล็อกนอร์มอล การแจกแจงไวบูลล์ และการแจกแจงไคกำลังสอง ด้วยขนาดตัวอย่าง 10, 20, 25, 30, 40, 50 และ 60 เปอร์เซ็นต์การถูกตัดปลายข้อมูล 10%, 20% และ 30% ณ ระดับนัยสำคัญ 0.01, 0.05 และ 0.10 ทำการวิจัยโดยการจำลองสถานการณ์การทดลองด้วยเทคนิคมอนติคาร์โล โดยทำการทดลองซ้ำ 1,000 รอบ สำหรับแต่ละสถานการณ์ที่กำหนดในการคำนวณหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว ผลการวิจัยสรุปได้ดังนี้

1. **ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1**

ตัวสถิติทดสอบทั้ง 3 ตัว สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ในทุกกรณีที่ศึกษา

2. **อำนาจการทดสอบ**

เมื่อขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 25 ในทุกเปอร์เซ็นต์การตัดปลาย ตัวสถิติทดสอบ CVM^C มีอำนาจการทดสอบสูงสุด และเมื่อขนาดตัวอย่างมากกว่า 25 ที่เปอร์เซ็นต์การตัดปลาย 10% ตัวสถิติทดสอบ CVM^C และสถิติทดสอบ AD^C ให้อำนาจการทดสอบใกล้เคียงกัน แต่ที่เปอร์เซ็นต์การตัดปลาย 20% และ 30% ตัวสถิติทดสอบ CVM^C ให้อำนาจการทดสอบสูงสุด

อำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว แปรผันตามขนาดตัวอย่างและเปอร์เซ็นต์การถูกตัดปลายของข้อมูล

ภาควิชา สถิติ ลายมือชื่อนิสิต ธิวา กันยะมี

สาขาวิชา สถิติ ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2548

4582245526 : MAJOR STATISTICS

KEY WORD: GOODNESS-OF-FIT TEST / TYPE II CENSORED DATA

TIWA KANYAMEE : A COMPARISON ON GOODNESS-OF-FIT TESTS FOR DISTRIBUTION WITH CENSORED DATA. THESIS ADVISOR : ASSOC. PROF. CAPT. MANOP VARAPHAKDI, M.S. 199 pp. ISBN 974-14-2346-2.

The objective of this research is to compare power of the test of three statistics for testing goodness-of-fit test for gamma distribution based on type II censored data. The test statistics used in this research are $K\text{-}S^C$ test statistic, AD^C test statistic and CVM^C test statistic. By considering the ability of controlling the probability of type I error and the power of the test when the population are gamma distribution, lognormal distribution, Weibull distribution and chi-square distribution . Sample sizes are 10, 20, 25, 30, 40, 50 and 60. The analysis of data was performed in case of censored data for 10%, 20% and 30%. Signficance levels are 0.01, 0.05 and 0.10. The probability of type I error and the power of the test are computed through the Monte Carlo Simulation method and repeated 1,000 times for each situation. The results of this research can be summarized as follows:

1. Probability of type I error.

   The three test statistics can control the probability of type I error for all situations.

2. Power of the test.

   For the sample size is less than or equal to 25 $(n \leq 25)$ and for all percents of censored data, CVM^C test statistic has highest power. When the sample size is more than 25 $(n > 25)$, CVM^C test statistic and AD^C test statistic have nearly same power at 10% of censored data, but CVM^C test statistic has highest power at 20% and 30% of censored data.

Power of the test of three statistics varies directly to sample size and percent of censored data.

Department Statistics
Field of study Statistics
Academic year 2005

Student's signature ติวา กัลยะมี
Advisor's signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงลงได้ด้วยความกรุณาของ รองศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษาและคำแนะนำ ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่างๆเป็นอย่างดี ผู้วิจัยจึงใคร่ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

นอกจากนี้ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา ในฐานะประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.ธีระพร วีระถาวร และอาจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ ในฐานะคณะกรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาตรวจสอบและให้คำแนะนำอันเป็นประโยชน์ในการแก้ไขวิทยานิพนธ์ให้เสร็จสมบูรณ์ยิ่งขึ้น

ท้ายนี้ ผู้วิจัยใคร่ขอน้อมรำลึกถึงพระคุณครูอาจารย์ทุกท่านที่ได้ให้ความรู้แก่ผู้วิจัย และกราบขอบพระคุณ บิดา มารดา และพี่ชาย ที่ช่วยส่งเสริมและสนับสนุนให้ผู้วิจัยได้มีโอกาสทางการศึกษาเสมอมาจนสำเร็จการศึกษา สุดท้ายนี้ขอขอบคุณเพื่อนๆ และน้องนิสิตปริญญาโท สาขาสถิติทุกท่านที่ให้ความช่วยเหลือ คำแนะนำ และกำลังใจในการทำวิทยานิพนธ์เป็นอย่างดีตลอดมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ฎ

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของปัญหา 2

1.2 วัตถุประสงค์ของการวิจัย 3

1.3 สมมติฐานการวิจัย 3

1.4 ข้อตกลงเบื้องต้น 3

1.5 ขอบเขตการวิจัย 4

1.6 เกณฑ์การประเมินตัวสถิติทดสอบ 4

1.7 คำจำกัดความ 5

1.8 ประโยชน์ที่คาดว่าจะได้รับ 5

บทที่ 2 ทฤษฎีที่เกี่ยวข้องกับการวิจัย

2.1 ชนิดของข้อมูลที่ถูกตัดปลาย 6

2.2 สถิติทดสอบที่ใช้ในการวิจัย 7

2.2.1 สถิติทดสอบ $K\text{-}S^C$ 7

2.2.2 สถิติทดสอบ AD^C 8

2.2.3 สถิติทดสอบ CVM^C 9

2.3 การแจกแจงของข้อมูลที่ใช้ในการวิจัย 15

2.3.1 การแจกแจงแกมมา (Gamma Distribution) 16

2.3.2 การแจกแจงล็อกนอร์มอล (Lognormal Distribution) 17

2.3.3 การแจกแจงไวบูลล์ (Weibull Distribution) 18

2.3.4 การแจกแจงไคกำลังสอง (Chi-square Distribution) 19

หน้า

บทที่ 3 วิธีการดำเนินการวิจัย

3.1 แผนการทดลอง 19

3.2 ขั้นตอนในการทดลอง..... 23

3.3 การสร้างข้อมูลให้มีการแจกแจงตามลักษณะที่ต้องการศึกษา..... 23

3.4 การหาค่าวิกฤตของตัวสถิติทดสอบ AD^C และ CVM^C..... 29

3.5 การหาความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 30

3.6 การหาค่าอำนาจการทดสอบ 32

3.7 ขั้นตอนในการทำงานของโปรแกรม 33

บทที่ 4 ผลการวิจัย

4.1 ค่าเปอร์เซ็นต์ไทล์ของตัวสถิติทดสอบ AD^c และ CVM^c..... 37

4.2 การเปรียบเทียบค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 41

4.3 การเปรียบเทียบค่าอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 วิธี 48

บทที่ 5 สรุปผลการวิจัย และข้อเสนอแนะ

5.1 สรุปผลการวิจัย 169

5.2 ข้อเสนอแนะ..... 170

รายการอ้างอิง..... 174

ภาคผนวก..... 175

ภาคผนวก ก..... 176

ภาคผนวก ข..... 196

ประวัติผู้เขียนวิทยานิพนธ์..... 199

สารบัญตาราง

ตาราง		หน้า
3.1	ค่าพารามิเตอร์ α และ λ และค่าสัมประสิทธิ์ความเบ้และความโด่งต่างๆ เมื่อประชากรมีการแจกแจงแกมมา	20
3.2	ค่าพารามิเตอร์ α และ λ และค่าสัมประสิทธิ์ความเบ้และความโด่งต่างๆ เมื่อประชากรมีการแจกแจงแกมมาที่มิใช่พารามิเตอร์เดียวกันกับสมมติฐานว่าง	20
3.3	ค่าพารามิเตอร์ μ และ σ และค่าสัมประสิทธิ์ความเบ้และความโด่งต่างๆ เมื่อประชากรมีการแจกแจงล็อกนอร์มอล	21
3.4	ค่าพารามิเตอร์ α และ β และค่าสัมประสิทธิ์ความเบ้และความโด่งต่างๆ เมื่อประชากรมีการแจกแจงไวบูลล์	21
3.5	ค่าพารามิเตอร์ n และค่าสัมประสิทธิ์ความเบ้และความโด่งต่างๆ เมื่อประชากรมีการแจกแจงไคกำลังสอง	22
3.6	แสดงจำนวนข้อมูลที่ถูกตัดปลาย	22
4.1	ความคลาดเคลื่อนในการทดสอบสมมติฐานทางสถิติ	37
4.2	ค่าเปอร์เซ็นต์ไทล์ที่ 0.90, 0.95 และ 0.99 สำหรับตัวสถิติทดสอบ AD^C และ CVM^C เมื่อพารามิเตอร์ $\alpha = 16, \lambda = 1$	40
4.3	ค่าเปอร์เซ็นต์ไทล์ที่ 0.90, 0.95 และ 0.99 สำหรับตัวสถิติทดสอบ AD^C และ CVM^C เมื่อพารามิเตอร์ $\alpha = 3.3, \lambda = 1$	41
4.4	ค่าเปอร์เซ็นต์ไทล์ที่ 0.90, 0.95 และ 0.99 สำหรับตัวสถิติทดสอบ AD^C และ CVM^C เมื่อพารามิเตอร์ $\alpha = 0.76, \lambda = 1$	42
4.5	ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทั้ง 3 ตัว สำหรับการแจกแจงแกมมาที่พารามิเตอร์ $\alpha = 16, \lambda = 1$	45
4.6	ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทั้ง 3 ตัว สำหรับการแจกแจงแกมมาที่พารามิเตอร์ $\alpha = 3.3, \lambda = 1$	46
4.7	ค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติทั้ง 3 ตัว สำหรับการแจกแจงแกมมาที่พารามิเตอร์ $\alpha = 0.76, \lambda = 1$	47
4.8	ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงแกมมา ($\alpha = 17.4, \lambda = 1$) มีสัมประสิทธิ์ความเบ้ 0.48 และสัมประสิทธิ์ความโด่ง 3.34	50
4.9	ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงล็อกนอร์มอล ($\mu = 2.8, \sigma = 0.16$) มีสัมประสิทธิ์ความเบ้ 0.49 และสัมประสิทธิ์ความโด่ง 3. 43	60

ตาราง หน้า

4.10 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไวบูลล์ ($\alpha = 2.02 , \beta = 18$) มีสัมประสิทธิ์ความเบ้ 0.6 และสัมประสิทธิ์ความโด่ง 3.22..... 70

4.11 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไคกำลังสอง $\chi^2_{(19)}$ มีสัมประสิทธิ์ความเบ้ 0.6 และสัมประสิทธิ์ความโด่ง 3.63..... 80

4.12 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงแกมมา ($\alpha = 4 , \lambda = 1$) มีสัมประสิทธิ์ความเบ้ 1.0 และสัมประสิทธิ์ความโด่ง 4.50 90

4.13 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงล็อกนอร์มอล ($\mu = 1.3 , \sigma = 0.32$) มีสัมประสิทธิ์ความเบ้ 1.02 และสัมประสิทธิ์ความโด่ง 4.91 100

4.14 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไวบูลล์ ($\alpha = 1.48 , \beta = 3.5$) มีสัมประสิทธิ์ความเบ้ 1.1 และสัมประสิทธิ์ความโด่ง 4.47 110

4.15 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไคกำลังสอง $\chi^2_{(5)}$ มีสัมประสิทธิ์ความเบ้ 1.27 และสัมประสิทธิ์ความโด่ง 5.4..... 120

4.16 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงแกมมา ($\alpha = 1.1 , \lambda = 1$) มีสัมประสิทธิ์ความเบ้ 1.9 และสัมประสิทธิ์ความโด่ง 8.42 130

4.17 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงล็อกนอร์มอล ($\mu = 0 , \sigma = 0.56$) มีสัมประสิทธิ์ความเบ้ 2.04 และสัมประสิทธิ์ความโด่ง 11.25 140

4.18 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไวบูลล์ ($\alpha = 1.03 , \beta = 1$) มีสัมประสิทธิ์ความเบ้ 1.9 และสัมประสิทธิ์ความโด่ง 8.44 150

4.19 ค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไคกำลังสอง $\chi^2_{(2)}$ มีสัมประสิทธิ์ความเบ้ 2.0 และสัมประสิทธิ์ความโด่ง 9.0..... 160

สารบัญภาพ

ภาพประกอบ หน้า

3.1 แผนผังแสดงขั้นตอนการคำนวณหาค่าวิกฤต..... 34

3.2 แผนผังแสดงขั้นตอนหาค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1..... 35

3.3 แผนผังแสดงขั้นตอนหาค่าอำนาจการทดสอบ..... 36

4.1 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงแกมมา ($\alpha = 17.4, \lambda = 1$) มีสัมประสิทธิ์ความเบ้ 0.48 และสัมประสิทธิ์ความโด่ง 3.34..... 51

4.2 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงล็อกนอร์มอล ($\mu = 2.8, \sigma = 0.16$) มีสัมประสิทธิ์ความเบ้ 0.49 และสัมประสิทธิ์ความโด่ง 3. 43..... 61

4.3 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไวบูลล์ ($\alpha = 2.02, \beta = 18$) มีสัมประสิทธิ์ความเบ้ 0.6 และสัมประสิทธิ์ความโด่ง 3.22..... 71

4.4 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไคกำลังสอง $\chi^2_{(19)}$ มีสัมประสิทธิ์ความเบ้ 0.6 และสัมประสิทธิ์ความโด่ง 3.63 81

4.5 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงแกมมา($\alpha = 4, \lambda = 1$) มีสัมประสิทธิ์ความเบ้ 1.0 และสัมประสิทธิ์ความโด่ง 4.50..... 91

4.6 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงล็อกนอร์มอล ($\mu = 1.3, \sigma = 0.32$) มีสัมประสิทธิ์ความเบ้เท่ากับ 1.02 และสัมประสิทธิ์ความโด่ง 4.91..... 101

4.7 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไวบูลล์ ($\alpha = 1.48, \beta = 3.5$) มีสัมประสิทธิ์ความเบ้ 1.1 และสัมประสิทธิ์ความโด่ง 4.47..... 111

4.8 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไคกำลังสอง $\chi^2_{(5)}$ มีสัมประสิทธิ์ความเบ้ 1.27 และสัมประสิทธิ์ความโด่ง 5.4..... 121

4.9 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงแกมมา($\alpha = 1.1, \lambda = 1$) โดยมีสัมประสิทธิ์ความเบ้ 1.9 และสัมประสิทธิ์ความโด่ง 8.42..... 131

ภาพประกอบ หน้า

4.10 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 เมื่อประชากรมีการแจกแจงล็อกนอร์มอล ($\mu = 0, \sigma = 0.56$) มีสัมประสิทธิ์ความเบ้ 2.04 และสัมประสิทธิ์ความโด่ง 11.25..... 141

4.11 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไวบูลล์ ($\alpha = 1.03\,1, \beta = 1$)มีสัมประสิทธิ์ความเบ้ 1.9 และสัมประสิทธิ์ความโด่ง 8.44..... 151

4.12 เปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบทั้ง 3 ตัว เมื่อประชากรมีการแจกแจงไคกำลังสอง $\chi^2_{(2)}$ มีสัมประสิทธิ์ความเบ้ 2.0 และสัมประสิทธิ์ความโด่ง 9.0 161

5.1 แผนผังการเลือกใช้สถิติทดสอบเมื่อทราบขนาดตัวอย่างและเปอร์เซ็นต์การตัดปลายข้อมูล..... 172