

วิธีนอนนาราเวตริก สำหรับการประมาณค่าฟังก์ชันการอยู่รอด

โดยพิศุภา 2 ด้วอช่าว ที่มีค่าสังเกตไม่สมบูรณ์

นางสาว ลากันย์ สุวรรณการ

วิทยานิพนธ์นี้เป็นงานวิจัยของภาควิชาการศึกษาคณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

มหาวิทยาลัยศรีนครินทรวิโรฒ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2530

ISBN 974-568-515-1

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

014261

i 10299099

Nonparametric Methods for Estimating Survival Function in  
The Two-sample Problem with Incomplete Observations

Miss Sagun Suwannakarn

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Statistics

Graduate School

Chulalongkorn University

1997

ISBN 974-568-016-1



ลำกัณย์ สุวรรณการ : วิธีนอนพาราเมตริก สำหรับการประมาณค่าฟังก์ชันการอยู่รอด ใน ปัญหา 2 ตัวอย่าง ที่มีค่าสังเกตไม่สมบูรณ์(Nonparametric Methods for Estimating Survival Function in the Two-sample Problem with Incomplete Observations) อ.ที่ปรึกษา : ดร. สุปล ตุงศ์วัฒนา, 77 หน้า .

การศึกษาฟังก์ชันการอยู่รอด (survival function) นั้น ทำได้ทั้งวิธีการพาราเมตริก (parametric) และวิธีนอนพาราเมตริก (nonparametric) ในการวิจัยครั้งนี้ ได้ศึกษาวิธีนอนพาราเมตริก 3 วิธี คือ วิธี Product-Limit (PL), วิธี Life-table (actuarial) และวิธี Cox's regression model ในตัวอย่างที่มีค่าสังเกตไม่สมบูรณ์ที่เป็นแบบผสมของ Singly failure censored และ multiply time censored และในกลุ่มตัวอย่างประกอบด้วย 2 กลุ่มตัวอย่างย่อย ในขนาด เท่ากัน ให้ความสำคัญว่า ค่ามัธยฐานการอยู่รอด (median survival time) ที่ประมาณจากฟังก์ชันการอยู่รอดของแต่ละวิธี และผลการทดสอบการเปรียบเทียบการแจกแจงการอยู่รอด (survival distribution) ของ 2 กลุ่มตัวอย่างย่อย จะเปลี่ยนแปลงหรือไม่ เมื่อจำนวนค่าสังเกตไม่สมบูรณ์ (censored observations) ในตัวอย่างเพิ่มเป็น 10, 20 และ 30 เปอร์เซ็นต์ของขนาดตัวอย่าง

ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลที่สร้างขึ้น (random number generation) โดยใช้โปรแกรมคอมพิวเตอร์ โดยให้ข้อมูลมีการแจกแจงแบบเอ็กโปเนนเชียล (exponential distribution) ด้วยพารามิเตอร์  $\lambda = 0.10$  และค่ามัธยฐานเท่ากับ 6.93 และขนาดตัวอย่างที่ใช้ในการศึกษามี 3 ขนาด คือ 10, 30, 50 ในแต่ละตัวอย่างแบ่งเป็น 2 กลุ่มตัวอย่างย่อยกลุ่มละ 5, 15, 25 ตามลำดับ และในแต่ละตัวอย่างจะถูกกำหนดให้ค่าสังเกตไม่สมบูรณ์ในตัวอย่างเท่ากับ 10, 20 และ 30 เปอร์เซ็นต์ของขนาดตัวอย่าง โดยการสุ่ม

ผลการศึกษาพบว่า ค่าประมาณมัธยฐานการอยู่รอดมีค่ามากกว่าค่ามัธยฐานจริงในทุก ๆ วิธี และมีค่าสูงมากขึ้นเมื่อจำนวนค่าสังเกตไม่สมบูรณ์ในตัวอย่างเพิ่มมากขึ้น เมื่อเปรียบเทียบกัน ใน 3 วิธีในแต่ละกลุ่มตัวอย่าง พบว่าวิธี Cox's regression model เป็นวิธีที่ให้ค่าประมาณมัธยฐานใกล้เคียงกับค่ามัธยฐานจริงมากที่สุด และวิธี Life-table ให้ค่าประมาณแตกต่างจากค่าจริงมากที่สุด แต่ก็มีค่าใกล้เคียงกับวิธี Product-Limit และสำหรับการทดสอบการเปรียบเทียบการแจกแจงการอยู่รอดของ 2 กลุ่มตัวอย่างย่อย โดยใช้ Cox-Mantel test พบว่า ที่  $\alpha = 0.05$  ผลการทดสอบผลการเปลี่ยนแปลงเมื่อตัวอย่างมีจำนวนค่าสังเกตไม่สมบูรณ์เพิ่มมากขึ้น เฉพาะตัวอย่างขนาด 10 ทั้งนี้เพราะว่า นวัตกรรมนอนพาราเมตริกมาใช้กับข้อมูลที่ทราบการแจกแจง แต่การศึกษานี้ก็เป็นประโยชน์ในด้านทฤษฎีการวิเคราะห์การอยู่รอดเป็นอย่างมาก สำหรับการศึกษาดังกล่าวต่อไปควรจะศึกษากับข้อมูลจริง โดยเฉพาะข้อมูลทางด้านทางการแพทย์ ขนาดตัวอย่างในการศึกษาควรจะมีมากกว่าการศึกษาในครั้งนี้ (มากกว่า 50) และจำนวนค่าสังเกตไม่สมบูรณ์ในตัวอย่างก็ไม่ควรจะมีมากกว่า 30 เปอร์เซ็นต์ของขนาดตัวอย่าง

ภาควิชา ..... สติติ  
สาขาวิชา ..... สติติ  
ปีการศึกษา 2530 .....

ลายมือชื่อนิติติ .....  
ลายมือชื่ออาจารย์ที่ปรึกษา .....

SAG'IN SUWANNAKARN : NONPARAMETRIC METHODS FOR ESTIMATING SURVIVAL  
FUNCTION IN THE TWO-SAMPLE PROBLEM WITH INCOMPLETE OBSERVATIONS.

THESIS ADVISOR : DR. SUPOL DURONGWATANA, Ed.D. 77 pp.

Both nonparametric and parametric methods can be used to study the survival function. In this research, the three methods of nonparametric; Product-Limit (PL), Life-table (actuarial) and Cox's regression model were used to examine the two-sample problem with censored observations : singly failure and multiply time censored observations. This study aims to investigate the change of the estimated median survival time as well as to compare the survival distributions of two subgroups when the censored observations are increased by 10, 20, 30 percent of the sample size.

The 3 sets of data with the sizes of 10; 30 and 50 which have the exponential distribution with parameter  $\lambda = 0.10$  and median = 6.93 were generated by computer program. Each set was randomly divided into two equal subgroups ; subgroup 1 and subgroup 2. Thus each subgroup contains 5, 15 and 25 for the sizes of 10, 30 and 50, respectively. A sample in each group was randomly assigned to be censored data for 10, 20 and 30 percent of the sample size.

The result of the research indicates that all of the estimated medians were greater than the true median (6.93). In addition, they would be increased through the increasing of the censored data. Among the three methods, the Cox's regression model yielded the best estimated median for the true value in every group of data, while the life-table method gave the most different. However, the life-table estimated medians were closed to those of the Product-Limit. Cox-Mantel test of the two survival distributions showed that there was significantly change only in the sample size 10 for  $\alpha = 0.05$ . This is because the distribution of data is known.

This study could be useful for further study in real life especially in medical field. However, that the sample size should be greater than the one in this study ( $> 50$ ) but the censored data should not be greater than 30 percent of the sample size.

ภาควิชา ..... สสถิติ  
สาขาวิชา ..... สสถิติ  
ปีการศึกษา 2530 .....

ลายมือชื่อผู้สมัคร .....  
ลายมือชื่ออาจารย์ที่ปรึกษา .....

## กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ ดร. พนม ฑูร่งศ์วัฒนา อาจารย์ที่ปรึกษา ซึ่งได้ให้คำแนะนำ ข้อคิดเห็น และความช่วยเหลือในการปรับปรุงการเขียนวิทยานิพนธ์ฉบับนี้ ให้สำเร็จลุล่วงไปได้ด้วยดี

บุคคลที่มีส่วนช่วยในการวิจัยครั้งนี้ ได้แก่ อาจารย์ จักรพงษ์ วงศ์ละคร, ผู้ช่วยศาสตราจารย์ กุฑชวี กุณฑบุตร, คุณ สุรินทร์ กุณฑบุตร, คุณวันตา เวชชาภิรักษ์ และคุณจำนงค์ กิ่งแก้ว ซึ่งได้ให้ความช่วยเหลือทางเอกสารวิชาการ ชี้แนะงานอันเป็นประโยชน์ต่อการวิจัย ย้ำนวยความสะดวกในการใช้เครื่องคอมพิวเตอร์ และได้ให้กำลังใจแก่ผู้วิจัยเป็นอย่างดี ผู้วิจัยจึงขอขอบพระคุณในความกรุณาของทุกท่านไว้ ณ ที่นี้ด้วย

ขอขอบพระคุณ คุณปราณี นันตาและคุณเพชรณี ภูพันธ์ ที่ให้ความช่วยเหลือทางด้าน การพิมพ์และตรวจสอบแก้ไขวิทยานิพนธ์ฉบับนี้

สากัญย์ สุวรรณภากร

สารบัญ

หน้า

บทคัดย่อภาษาไทย	จ
บทคัดย่อภาษาอังกฤษ	ง
กิตติกรรมประกาศ	ฉ
สารบัญตารางประกอบ	ช

บทที่

1. บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 จุดประสงค์ของการวิจัย	6
1.3 สมมติฐานในการวิจัย	6
1.4 ขีดกลางเบื้องต้น	6
1.5 ขอบเขตของการวิจัย	6
1.6 ประโยชน์ที่คาดว่าจะได้รับ	7
2. ฟังก์ชันของเวลาการอยู่รอด (Function of Survival Time)	8
2.1 นิยาม	8
2.2 ความสัมพันธ์ของฟังก์ชันการอยู่รอด	14
3. วิธีนอนพาราเมตริกสำหรับการประมาณค่าฟังก์ชันการอยู่รอด สำหรับ ข้อมูลไม่สมบูรณ์ (Nonparametric Methods of Estimating Survival Function for Censored Data)	15
3.1 วิธี Product-Limit (PL)	17
3.1.1 นิยามและวิธีการคำนวณ	18
3.1.2 ความแปรปรวนของ $\hat{S}(t)$	20

3.1.3	การประมาณค่าเฉลี่ยและมัธยฐานการอยู่รอดโดยใช้ตัว ประมาณ PL	21
3.2	วิธี Life-table (actuarial) .....	24
3.2.1	การประมาณค่าฟังก์ชันการอยู่รอด $S(t)$ .....	24
3.2.2	การประมาณค่าฟังก์ชันการสูญเสีย $h(t)$ .....	25
3.2.3	การประมาณค่ามัธยฐานการอยู่รอดโดยใช้ตัวประมาณ Life-table	27
3.3	วิธี Cox's regression model .....	27
3.3.1	Cox's regression model .....	27
3.3.2	การประมาณค่าพารามิเตอร์ $\beta$ โดยวิธี Maximum- Likelihood (ML)	31
3.3.3	ปัญหา 2 ตัวอย่าง (two-sample problem) ....	33
4.	วิธีการดำเนินการวิจัย .....	35
4.1	การสร้างข้อมูลเพื่อใช้ในการศึกษา .....	35
4.2	ขนาดตัวอย่าง .....	36
4.3	การกำหนดข้อมูลให้เป็นข้อมูลไม่สมบูรณ์ ที่เป็นแบบผสมของ ... Singly failure และ multiply time censored	36
4.4	การวิเคราะห์ข้อมูล .....	39
4.5	ผลการวิเคราะห์ข้อมูล .....	41
5.	สรุปผลและข้อเสนอแนะ .....	53
5.1	สรุปผลการวิจัย .....	53
5.2	ข้อเสนอแนะ .....	54

	หน้า
บรรณานุกรม .....	55
ภาคผนวก ก. Cox-Mantel test .....	57
ข. Newton-Raphson Method .....	59
ค. โปรแกรมคอมพิวเตอร์สำหรับการสร้างข้อมูล (ภาษา Basic) .....	62
ง. คำสั่งที่ใช้ในโปรแกรมสำเร็จรูป BMDP .....	65
จ. ตัวอย่าง .....	71
ประวัติผู้วิจัย .....	77

สารบัญตาราง

ตารางที่		หน้า
1	เปรียบเทียบค่าประมาณมัธยฐานการอยู่รอดของแต่ละวิธี แยกตามกลุ่มตัวอย่างย่อยและจำนวนค่าสังเกตไม่สมบูรณ์ (ก) ตัวอย่างขนาด 10 (ข) ตัวอย่างขนาด 30 (ค) ตัวอย่างขนาด 50	45
2	เปรียบเทียบค่าประมาณมัธยฐานการอยู่รอดของแต่ละขนาดตัวอย่าง แยกตามกลุ่มตัวอย่างย่อย และจำนวนค่าสังเกตไม่สมบูรณ์ (ก) วิธี Product-Limit (PL) (ข) วิธี Life-table (actuarial) (ค) วิธี Cox's regression model	49
3	ผลการทดสอบ $H_0 : S_1(t) = S_2(t)$ (การแจกแจงการอยู่รอดของ 2 กลุ่มตัวอย่างย่อยไม่แตกต่างกัน) จากตัวอย่าง 100 ชุด โดย Cox-Mantel test (ก) $\alpha = 0.05$ (ข) $\alpha = 0.01$	51