การหาค่ารายได้ที่ขาดหายไปในการสำรวจตัวอย่าง

นางสาว รวีวรรณ สิริช่วยชูชัย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2526

ISBN 974-562-661-9

010496 | i 17079779

Imputation of Missing Income Data in Sample Survey

Miss Raweewan Sirichuaychuchai

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Statistics

Graduate School

Chulalongkorn University

1983

หัวข้อวิทยานิพนธ์	การหาค่ารายได้ที่ขาดหายไปในการสำรวจตัวอย่าง
โดย	นางสาว รวีวรรณ สิริช่วยชูชัย
ภาควิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. สุชาดา กีระนันทน์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.สุประดิษฐ์ บุนนาค)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร)

.....กรรมการ

(รองศาสตราจารย์ ดร. สุชาดา กีระนันทน์)

.....กรรมการ

(รองศาสตราจารย์ ชูศักดิ์ อุดมศรี)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ จลีพร โกลากุล)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

หัวข้อวิทยานิพนธ์ การหาค่ารายได้ที่ขาดหายไปในการสำรวจตัวอย่าง

ชื่อนิสิต นางสาวรวีวรรณ สิริช่วยชูชัย

อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. สุชาดา กีระนันทน์

ภาควิชา สถิติ

ปีการศึกษา 2526

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

รายได้มักได้รับการพิจารณาเป็นตัวแปรหลักที่สำคัญในการวิจัยทางสังคมศาสตร์ แต่จะประสบปัญหาของการไม่ได้รับคำตอบเสมอ การวิจัยนี้ศึกษาวิธีการหาค่ารายได้ที่ขาดหายไปในการสำรวจตัวอย่างโดยพิจารณากรณีที่ไม่สนใจกลไกของการตอบ กล่าวคือ ไม่พิจารณากรณีที่ความน่าจะเป็นของการตอบขึ้นอยู่กับระดับรายได้ สำหรับวิธีที่ใช้ในการศึกษาคือ วิธีการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้วและวิธีวิเคราะห์ความถดถอย

ผลการศึกษาวิจัยสำหรับวิธีการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้วพบว่าตัวแปรรายได้มีความสัมพันธ์กับฐานะทางเศรษฐกิจสูงกว่าตัวแปรอื่น ๆ อันได้แก่ รายจ่ายสำหรับครัวเรือน ระดับการศึกษาของหัวหน้าครัวเรือน อาชีพหลักของหัวหน้าครัวเรือน สถานภาพในการทำงานอาชีพหลัก จำนวนชั่วโมงทำงานต่อสัปดาห์ เขตที่ครัวเรือนอาศัยอยู่ระดับอายุของหัวหน้าครัวเรือน ประสบการณ์ในการทำงานอาชีพหลัก ดังนั้นจึงใช้ฐานะทางเศรษฐกิจเป็นตัวแปรแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้วโดยแบ่งออกเป็น 5 ชั้นภูมิ ตามลักษณะของตัวแปร เมื่อแบ่งชั้นภูมิแล้วได้ประมาณค่ารายได้ที่ไม่ได้รับคำตอบโดยพิจารณาว่าหน่วยนั้นอยู่ในชั้นภูมิใดแล้วประมาณค่ารายได้โดยใช้ค่าเฉลี่ยและ/หรือค่าสุ่มจากชั้นภูมินั้น ส่วนการประมาณค่ารายได้ที่ไม่ได้รับคำตอบด้วยวิธีวิเคราะห์ความถดถอยแบบขั้นตอนพบว่าตัวแปรรายได้ขึ้นอยู่กับฐานะทางเศรษฐกิจรายจ่ายสำหรับครัวเรือน จำนวนชั่วโมงทำงานต่อสัปดาห์สำหรับอาชีพหลัก การศึกษาของหัวหน้าครัวเรือน ประสบการณ์ในการทำงานอาชีพหลัก สถานภาพในการทำงานอาชีพหลักและอาชีพหลักของหัวหน้าครัวเรือน

จากการเปรียบเทียบวิธีการประมาณรายได้ที่ไม่ได้รับคำตอบระหว่างวิธีการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว กับวิธีวิเคราะห์ความถดถอยโดยใช้ร้อยละของผลต่างระหว่างค่าประมาณกับค่าจริงของรายได้ และใช้การทดสอบความแตกต่างระหว่างค่าประมาณกับค่าจริงของรายได้ ผลการวิเคราะห์

สรุปได้ว่า วิธีที่เหมาะสมสำหรับการประมาณรายได้ที่ขาดหายไปควรเป็นวิธีการแบ่งชั้นภูมิ เมื่อเลือกตัวอย่างแล้ว โดยอาจใช้ค่าเฉลี่ยรายได้จากชั้นภูมิหรือค่าสุ่มรายได้จากชั้นภูมิ เนื่องจากให้ค่าประมาณไม่แตกต่างจากค่าจริงที่ระดับนัยสำคัญ .05 นอกจากนี้ยังเป็นวิธีที่ง่าย สะดวกต่อการคำนวณ และให้ผลดีกว่าการประมาณรายได้ด้วยวิธีวิเคราะห์ความถดถอย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Thesis Tittle Imputation of Missing Income Data in Sample Survey

Name Miss Raweewan Sirichuaychuchai

Thesis Advisor Associate Professor Suchada Kiranandana, Ph.D

Department Statistics

Academic Year 1983

ABSTRACT

Income is regarded as a major variable in social science research. However, the responding unit frequently skips this item. This research studies the methods of imputation of missing income data in sample survey for the case where response mechanism is ignorable, i.e. where the probability of response for the income variable does not depend upon the value of that variable. The methods considered are poststratification method and regression analysis.

The study shows that, for the poststratification method, income has a stronger relationship to economic status than household's expenditure, household head's education, household head's major occupation, major occupation status, number of office hours per week, region, household head's age and experience of major occupation. Therefore, economic status was selected as poststratification variable, resulting in 5 strata. The missing income data for a unit classified as belonging to a stratum is estimated by the stratum mean income and by an income value randomly selected from the same stratum. As for the estimation of missing income by using stepwise multiple regression analysis, it is found that income is significantly determined by economic status, household's expenditure, number of office hours per week, household head's education, experience of major occupation, major occupation status and household head's major

occupation.

Comparison between the two methods used to estimate missing income was done using as criteria the percentage of difference between the estimated and true income and test of difference between the estimated and true income. The conclusion is that the poststratification method is better for estimating missing income and is a simpler and more practical method than the regression method.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.สุชาดา กีระนันทน์ หัวหน้าภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่กรุณาสละเวลาให้คำแนะนำในการวิจัยและให้ความช่วยเหลือในการปรับปรุงแก้ไขข้อบกพร่องต่าง ๆ อย่างดียิ่งตลอดมา จนทำให้การวิจัยครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี

และขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.เทียนฉาย กีระนันทน์ และคณาจารย์ผู้ร่วมวิจัยที่กรุณาให้ข้อมูลในงานวิจัย เรื่อง สภาวะทางเศรษฐกิจ สังคม และประชากรของกรุงเทพมหานคร ประกอบการทำวิทยานิพนธ์ครั้งนี้

รวีวรรณ สิริช่วยชูชัย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... ฉ

กิตติกรรมประกาศ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ฏ

บทที่

1. บทนำ 1

2. ระเบียบวิธีที่ใช้ในการวิจัย 9

3. ผลการวิเคราะห์ข้อมูล 17

4. สรุปผลการวิจัยและข้อเสนอแนะ 69

บรรณานุกรม 71

ภาคผนวก 73

ประวัติผู้เขียน 90

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
1	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและระดับอายุของหัวหน้าครัวเรือน.....	20
2	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและระดับการศึกษา.....	21
3	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและอาชีพหลัก.....	22
4	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและสถานภาพในการทำงานสำหรับอาชีพหลัก.....	23
5	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและฐานะทางเศรษฐกิจของหัวหน้าครัวเรือน.....	24
6	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและระดับรายจ่ายสำหรับครัวเรือน.....	25
7	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและจำนวนชั่วโมงทำงานต่อสัปดาห์สำหรับอาชีพหลัก.....	26
8	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและประสบการณ์ในการทำงานสำหรับอาชีพหลัก.....	27
9	จำนวนครัวเรือนจำแนกตามระดับรายได้หลักของหัวหน้าครัวเรือนและเขตที่ครัวเรือนอาศัยอยู่.....	28
10	แสดงค่าคราแมร์วีของตัวแปรที่มีความสัมพันธ์กับรายได้หลักของหัวหน้าครัวเรือน.....	32
11	แสดงค่าเฉลี่ยของรายได้ที่ทราบค่าหลังจากเลือกตัวแทนเพื่อแทนหน่วยที่ไม่ตอบเรื่องรายได้ โดยอาศัยปัจจัยพื้นฐานของหัวหน้าครัวเรือนและมีฐานะทางเศรษฐกิจเป็นตัวแปรแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว..	33

ตารางที่		หน้า
12	แสดงค่าเฉลี่ยของรายได้ที่ทราบค่าหลังจากเลือกตัวแทนอย่างสุ่มเพื่อแทนหน่วยที่ไม่ตอบเรื่องรายได้และมีฐานะทางเศรษฐกิจเป็นตัวแปรแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว.....	34
13	แสดงค่าจริงและค่าประมาณรายได้ที่ประมาณจากวิธีการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว โดยมีฐานะทางเศรษฐกิจเป็นตัวแปรแบ่งชั้นภูมิของตัวแทนหน่วยที่ไม่ตอบเรื่องรายได้ ซึ่งเลือกขึ้นมาโดยอาศัยปัจจัยของหัวหน้าครัวเรือน.....	36
14	แสดงค่าจริงและค่าประมาณรายได้ที่ประมาณจากวิธีการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว โดยมีฐานะทางเศรษฐกิจเป็นตัวแปรแบ่งชั้นภูมิของตัวแทนหน่วยที่ไม่ตอบเรื่องรายได้ซึ่งเลือกขึ้นมาโดยสุ่ม.....	39
15	แสดงการประมาณค่าพารามิเตอร์และค่าความคลาดเคลื่อนมาตรฐานที่ได้จากสมการถดถอย.....	49
16	แสดงค่าจริงและค่าประมาณรายได้ที่ประมาณจากสมการถดถอยของตัวแทนหน่วยที่ไม่ตอบเรื่องรายได้ซึ่งเลือกขึ้นมาโดยอาศัยปัจจัยของหัวหน้าครัวเรือน.....	51
17	แสดงค่าจริงและค่าประมาณรายได้ที่ประมาณจากสมการถดถอยของตัวแทนหน่วยที่ไม่ตอบเรื่องรายได้ซึ่งเลือกขึ้นมาโดยสุ่ม.....	54
18	แสดงค่าจริงกับค่าประมาณรายได้ที่ประมาณจากวิธีต่าง ๆ และร้อยละของผลต่างจากรายได้จริง ของตัวแทนหน่วยที่ไม่ตอบเรื่องรายได้ซึ่งเลือกขึ้นมาโดยอาศัยปัจจัยของหัวหน้าครัวเรือน.....	58
19	แสดงค่าจริงกับค่าประมาณรายได้ที่ประมาณจากวิธีต่าง ๆ และร้อยละของผลต่างจากรายได้จริงของตัวแทนหน่วยที่ไม่ตอบเรื่องรายได้ซึ่งเลือกขึ้นมาโดยสุ่ม.....	61
20	แสดงค่าสถิติ t ที่คำนวณได้จากการทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างรายได้ที่ประมาณได้จาก 3 วิธี กับค่าจริงของรายได้และระหว่างรายได้ที่ประมาณได้จากแต่ละวิธีของตัวแทนของครัวเรือนที่ไม่ตอบเรื่องรายได้ โดยอาศัยปัจจัยพื้นฐานของหัวหน้าครัวเรือน.....	66

ตารางที่ หน้า

21 แสดงค่าสถิติ t ที่คำนวณได้จากการทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างรายได้ที่ประมาณได้จาก 3 วิธีกับค่าจริงของรายได้และระหว่างรายได้ที่ประมาณได้จากแต่ละวิธีของตัวแทนครัวเรือนที่ไม่ตอบเรื่องรายได้ที่เลือกขึ้นมาโดยวิธีสุ่ม..... 67

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพที่		หน้า
1	แสดงแผนภาพการกระจายระหว่างรายได้หลักของหัวหน้าครัวเรือนและฐานะทางเศรษฐกิจของผู้ให้สัมภาษณ์.....	43
2	แสดงแผนภาพการกระจายระหว่างรายได้หลักของหัวหน้าครัวเรือนและรายจ่ายสำหรับครัวเรือน.....	44
3	แสดงแผนภาพการกระจายระหว่างค่า log ของรายได้หลักของหัวหน้าครัวเรือนและฐานะทางเศรษฐกิจของผู้ให้สัมภาษณ์.....	45
4	แสดงแผนภาพการกระจายระหว่างค่า log ของรายได้หลักของหัวหน้าครัวเรือนและรายจ่ายสำหรับครัวเรือน.....	46

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย