



บทที่ 1

บทนำ

1.1 คำนำ

วิธีการทางสถิติช่วยให้สามารถคาดการณ์ต่าง ๆ ซึ่งเป็นประโยชน์ในการวางแผนล่วงหน้าได้อย่างใกล้เคียงความเป็นจริง เช่น การวิเคราะห์ความถดถอย ซึ่งเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ของตัวแปรตั้งแต่สองตัวแปรขึ้นไป เพื่อใช้ประโยชน์ในการพยากรณ์แต่ก็มักจะมีปัญหา กล่าวคือค่าสังเกตของตัวแปรบางค่าที่ต้องการศึกษาขาดหายไป ซึ่งการขาดหายไปนี้อาจเกิดขึ้นโดยไม่ได้ตั้งใจหรือเนื่องจากไม่ได้เก็บค่าโดยจงใจหรืออาจเนื่องจากการใช้จำง่ายในการเก็บค่าสังเกตมีจำนวนจำกัด หรืออาจเกิดจากเวลาหรือสภาวะแวดล้อมที่ต้องทำให้ตัดค่าสังเกตบางค่าทิ้งกันไป เป็นต้น

ในการวิเคราะห์ความถดถอยนั้น ถ้าข้อมูลที่เก็บรวบรวมมาเพื่อทำการวิเคราะห์เป็นไปตามข้อตกลงเบื้องต้น และมีข้อมูลครบสมบูรณ์ทุกตัว ก็ไม่เกิดปัญหาในการวิเคราะห์ แต่ถ้าหากข้อมูลที่รวบรวมมาได้นั้นมีบางตัวสูญหายไปและไม่สามารถตามไปเก็บเพิ่มเติมได้ทำให้ข้อมูลของตัวอย่างไม่สมบูรณ์ทำการวิเคราะห์ไม่ได้ นอกจากนี้ผู้วิจัยจะตัดค่าสังเกตเหล่านั้นทิ้งไป ซึ่งจะมีผลทำให้ขนาดตัวอย่างมีจำนวนน้อยลง และที่สำคัญยิ่งก็คือ ทำให้สูญเสียรายละเอียดบางอย่างไป ซึ่งอาจจะมีผลกระทบต่อผลสรุปของการวิเคราะห์นั้น ๆ ได้โดยทั่วไปในการวิเคราะห์ความถดถอยนั้น วิธีที่นิยมใช้ในการคาดคะเนสัมประสิทธิ์ความถดถอย (Regression coefficient) คือวิธีกำลังสองน้อยที่สุด (Ordinary least square) แต่เมื่อค่าสังเกตขาดหายไปจะไม่สามารถประมาณได้ด้วยวิธีดังกล่าว ในการวิจัยนี้ได้ใช้ค่าประมาณวิธีต่าง ๆ แทนในค่าสังเกตที่ขาดหายไปเพื่อศึกษาเปรียบเทียบวิธีการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอย เมื่อข้อมูลมีลักษณะต่าง ๆ กัน

1.2 ที่มาของปัญหา

เนื่องจากปัญหาการสุ่มหาของข้อมูล มีผลกระทบต่อการใช้วิเคราะห์สมการถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Ordinary least square) และถ้าตัดค่าสังเกตที่สุ่มหาทั้งอาจมีผลกระทบต่อผลสรุปของการวิเคราะห์สมการถดถอยได้ ดังนั้นจึงมีนักสถิติหลายท่านได้คิดวิธีประมาณสมการถดถอยและประมาณค่าสังเกตที่สุ่มหาแบบต่าง ๆ ขึ้น พอสรุปได้ดังนี้

ค.ศ.1932 Wilks ได้ศึกษาวิธีประมาณสมการถดถอยเมื่อมีข้อมูลตัวแปรอิสระเพียงตัวเดียว ซึ่งมีการแจกแจงแบบปกติ และการสุ่มหาของข้อมูลของตัวแปรอิสระเป็นไปอย่างสุ่ม และได้ทำการประมาณข้อมูลสุ่มหาของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood

ค.ศ.1951 Matthai ได้ศึกษาวิธีประมาณสมการถดถอยที่มีตัวแปรอิสระหลายตัวซึ่งมีการแจกแจงแบบปกติและมีตัวแปรอิสระเพียงตัวเดียวที่มีข้อมูลสุ่มหาและการสุ่มหาเป็นไปอย่างสุ่ม และได้ทำการประมาณข้อมูลสุ่มหาของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood

ค.ศ.1958 George L. Edgett ได้ศึกษาวิธีประมาณสมการถดถอยกรณีที่มี 2 ตัวแปรอิสระซึ่งมีการแจกแจงแบบปกติ และมีตัวแปรอิสระเพียงตัวเดียวที่มีค่าสุ่มหาและการสุ่มหาเป็นไปอย่างสุ่ม และได้ทำการประมาณข้อมูลสุ่มหาของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood

ค.ศ.1957 T.W. Anderson ได้ทำการประมาณพารามิเตอร์กรณีที่มีหลายตัวแปรอิสระ ซึ่งมีการแจกแจงแบบปกติโดยทำการพิจารณาตัวแปรทีละคู่ และประมาณข้อมูลสุ่มหาของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood ซึ่งให้ผลสรุปเช่นเดียวกับวิธีการของ George L. Edgett

ค.ศ.1972 - 1976 Linda S.Chan และ Olive Jean Dunn ได้ศึกษาเปรียบเทียบวิธีประมาณค่าสุ่มหาในการวิเคราะห์ 4 วิธี คือไปนโดชใช้ R^2 เป็นเกณฑ์ในการเปรียบเทียบ

ก. ศึกษาเมื่อไม่มีข้อมูลสุ่มหาเลข

- ข. ศึกษาเมื่อมีข้อมูลสุทธหายโดยตัดชุดข้อมูลที่มีค่าสุทธหายออก
- ค. ศึกษาเมื่อมีข้อมูลสุทธหายโดยประมาณข้อมูลสุทธหายด้วยค่าเฉลี่ย (Mean)
- ง. ศึกษาเมื่อมีข้อมูลสุทธหายโดยประมาณข้อมูลสุทธหายด้วยวิธีวิเคราะห์สมการถดถอยเชิงเส้น (linear regression equation)

ซึ่งจากการศึกษาของ Linda S.Chan และ Olive Jean Dunn พบว่า วิธีค่าเฉลี่ยดีที่สุด

ในการวิจัยครั้งนี้ผู้วิจัยได้นำเอาวิธีการที่ T.W. Anderson ประมาณตัวพารามิเตอร์ของค่าเฉลี่ยของตัวแปรไปแทนในข้อมูลสุทธหายและได้ทำการประมาณสมการถดถอยด้วยวิธี OLS นอกจากนี้ผู้วิจัยได้นำเอาวิธีประมาณค่าสุทธหายด้วยค่ามัธยฐานและด้วยวิธีของ Linda S.Chan และ Olive Jean Dunn แทนในค่าสังเกตที่ขาดหายไป แล้วทำการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยที่สุด มาทำการศึกษาค้นคว้า และเปรียบเทียบสมการถดถอยเมื่อมีข้อมูลตัวแปรอิสระสุทธหายอย่างสุ่มในสัดส่วนที่เท่ากัน ทั้งนี้เพื่อเป็นแนวทางให้ผู้ที่สนใจสามารถเลือกวิธีประมาณข้อมูลสุทธหายในการวิเคราะห์การถดถอย เมื่อข้อมูลมีลักษณะต่าง ๆ กัน

นั่นคือสมการถดถอยที่สนใจนำมาเปรียบเทียบในวิทยานิพนธ์นี้ คือ

1. สมการถดถอย เมื่อไม่มีข้อมูลสุทธหายเลยโดยจะเรียกว่าวิธีวิเคราะห์สมบูรณ์
2. สมการถดถอย เมื่อมีข้อมูลสุทธหายโดยตัดชุดข้อมูลที่มีค่าสุทธหายออก โดยจะเรียกว่าวิธีสุทธหาย
3. สมการถดถอย เมื่อมีข้อมูลสุทธหายโดยประมาณข้อมูลสุทธหายจากสมการถดถอยเชิงเส้นโดยจะเรียกว่าวิธีวิเคราะห์ความถดถอย
4. สมการถดถอย เมื่อใช้ตัวประมาณ maximum likelihood ประมาณข้อมูลสุทธหาย โดยจะเรียกว่าวิธี MAXIMUM LIKELIHOOD
5. สมการถดถอย เมื่อใช้ค่าเฉลี่ยประมาณข้อมูลสุทธหายโดยจะเรียกว่าวิธีค่าเฉลี่ย
6. สมการถดถอย เมื่อใช้ค่ามัธยฐานประมาณข้อมูลสุทธหาย โดยจะเรียกว่า วิธีค่ามัธยฐาน

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการเลือกวิธีประมาณค่าสูญหายของตัวแปรอิสระในการวิเคราะห์
การถดถอย



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย