

1.1 ค่าม่า

วิธีการทางสถิติช่วยให้สามารถคาดการณ์ต่าง ๆ ซึ่งเป็นประโยชน์ในการวางแผนล่วงหน้าได้อย่างใกล้เคียงความเป็นจริง เช่น การวิเคราะห์ความถดถอย ซึ่งเป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ของตัวแปรตั้งแต่สองตัวแปรขึ้นไป เพื่อใช้ประโยชน์ในการพยากรณ์แต้มกจะเกิดปัญหา กล่าวคือค่าสั่งเกตของตัวแปรบางค่าที่ต้องการศึกษาขาดหายไป ซึ่งการขาดหายไปนี้อาจเกิดขึ้นโดยไม่ตั้งใจหรือเนื่องจากไม่ได้เก็บค่าโดยจงใจหรืออาจเนื่องจากค่าใช้จ่ายในการเก็บค่าสั่งเกตมีจำนวนจำกัด หรืออาจเกิดจากเวลาหรือสภาวะแวดล้อมที่ห้องทำไว้ติดค่าสั่งเกตบางค่านั้นทึบไป เป็นต้น

ในการวิเคราะห์ความถดถอยนั้น ถ้าข้อมูลที่เก็บรวบรวมมาเพื่อกำกับการวิเคราะห์เป็นไปตามข้อตกลงเบื้องต้น และมีข้อมูลครบสมบูรณ์ทุกด้าน ก็ไม่เกิดปัญหาในการวิเคราะห์แต่ถ้าหากข้อมูลที่รวบรวมมาได้นั้นมีบางตัวสูญหายไปและไม่สามารถตามไปเก็บเพิ่มเติมได้ทำให้ข้อมูลของตัวอย่างไม่สมบูรณ์ทำกับการวิเคราะห์ไม่ได้ นอกจากว่าผู้วิจัยจะตัดค่าสั่งเกตดูนั้นทึบไป ซึ่งจะมีผลทำให้ขนาดตัวอย่างมีจำนวนน้อยลง และที่สำคัญยิ่งก็คือ ทำให้สูญเสียรายละเอียดบางอย่างไป ซึ่งอาจจะมีผลกระทบต่อผลสรุปของ การวิเคราะห์นั้น ๆ ได้โดยทั่วไปในการวิเคราะห์ความถดถอยนั้น วิธีที่นิยมใช้ในการคาดคะเนลัมปาราลิกซ์ความถดถอย (Regression coefficient) คือวิธีกำลังสองน้อยที่สุด (Ordinary least square) แต่เมื่อค่าสั่งเกตขาดหายไปจะไม่สามารถประมาณได้ตัววิธีดังกล่าว ใน การวิจัยนี้ได้ใช้ค่าประมาณวิธีต่าง ๆ แทนในค่าสั่งเกตที่ขาดหายไปเพื่อศึกษาเปรียบเทียบวิธีการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอย เนื่องข้อมูลนี้ลักษณะต่าง ๆ กัน

1.2 ที่มาของปัญหา

เนื่องจากปัญหาการสูญเสียของข้อมูล มีผลกระทำบ่อกรวิเคราะห์สมการลดคงด้วยวิธีกำลังสองน้อยที่สุด (Ordinary least square) และถ้าตัวค่าสัมภัยเก็ตที่สูญเสียก็อาจมีผลกระทำบ่อกรวิเคราะห์สมการลดคงอย่างไร้ความหมาย ดังนั้นจึงมีนักสถิติหลายท่านได้คิดวิธีประมาณสมการลดคงและประมาณค่าสัมภัยเก็ตที่สูญเสียแบบต่าง ๆ ขึ้น พอสรุปได้ดังนี้

ค.ศ. 1932 Wilks ได้ศึกษาวิธีประมาณสมการลดคงที่มีข้อมูลตัวแปรอิสระเพียงตัวเดียว ซึ่งมีการแจกแจงแบบปกติ และการสูญเสียของข้อมูลของตัวแปรอิสระเป็นไปอย่างสุ่ม และได้ทำการประมาณข้อมูลสูญเสียของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood

ค.ศ. 1951 Matthaai ได้ศึกษาวิธีประมาณสมการลดคงที่มีตัวแปรอิสระหลายตัวซึ่งมีการแจกแจงแบบปกติและมีตัวแปรอิสระเพียงตัวเดียวที่มีข้อมูลสูญเสียและการสูญเสียเป็นไปอย่างสุ่ม และได้ทำการประมาณข้อมูลสูญเสียของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood

ค.ศ. 1958 George L. Edgett ได้ศึกษาวิธีประมาณสมการลดคงกรณีที่มี 2 ตัวแปรอิสระซึ่งมีการแจกแจงแบบปกติ และมีตัวแปรอิสระเพียงตัวเดียวที่มีค่าสูญเสียและการสูญเสียเป็นไปอย่างสุ่ม และได้ทำการประมาณข้อมูลสูญเสียของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood

ค.ศ. 1957 T.W. Anderson ได้ทำการประมาณพารามิเตอร์รากที่มีหลายตัวแปรอิสระ ซึ่งมีการแจกแจงแบบปกติโดยทำการพิจารณาตัวแปรทั้งคู่ และประมาณข้อมูลสูญเสียของตัวแปรอิสระด้วยค่าประมาณ Maximum likelihood ซึ่งให้ผลสรุปเช่นเดียวกับวิธีการของ George L. Edgett

ค.ศ. 1972 - 1976 Linda S. Chan และ Olive Jean Dunn ได้ศึกษาเบรื้อง เทียนวิธีประมาณค่าสูญเสียในการวิเคราะห์ 4 วิธี คือไปร์โคตไช้ R^2 เป็นเกณฑ์ในการเบรื้องเทียน

ก. ศึกษาเมื่อไหร่มีข้อมูลสูญเสียแค่

๓. ศึกษาเมื่อมีข้อมูลสุ่มหายโดยตัดชุดข้อมูลที่มีค่าสุ่มหายออก
๔. ศึกษาเมื่อมีข้อมูลสุ่มหายโดยประมาณข้อมูลสุ่มหายด้วยค่าเฉลี่ย (Mean)
๕. ศึกษาเมื่อมีข้อมูลสุ่มหายโดยประมาณข้อมูลสุ่มหายด้วยวิเคราะห์สมการถดถอยเชิงเส้น (linear regression equation)

ซึ่งจากการศึกษาของ Linda S.Chan และ Olive Jean Dunn พบว่า วิธีค่าเฉลี่ยคือที่สุด

ในการวิจัยครั้งนี้ผู้วิจัยได้นำเอาวิธีการที่ T.W. Anderson ประมาณตัวพารามิเตอร์ ของค่าเฉลี่ยของตัวแปรไปแทนในข้อมูลสุ่มหายและได้ทำการประมาณสมการถดถอยด้วยวิธี OLS นอกจากผู้วิจัยได้นำเอาวิธีประมาณค่าสุ่มหายด้วยค่ามัธยฐานและด้วยวิธีของ Linda S.Chan และ Olive Jean Dunn แทนในค่าสัมเก็ตที่ขาดหายไป แล้วทำการประมาณค่าล้มประมาณประสิกความถดถอยด้วยวิธีกำลังสองน้อยที่สุด ทำการศึกษาด้วย และเปรียบเทียบสมการถดถอยเมื่อมีข้อมูลตัวแปรอิสระสุ่มหายอย่างสุ่มในสัดส่วนที่เท่ากัน ทั้งนี้เนื่องเป็นแนวทางให้ผู้ที่สนใจสามารถเลือกวิธีประมาณข้อมูลสุ่มหายในการวิเคราะห์การถดถอย เมื่อข้อมูลมีลักษณะต่าง ๆ กัน

นั้นคือสมการถดถอยที่สนใจนำมาเปรียบเทียบในวิถีวนินพนธ์ คือ

1. สมการถดถอย เมื่อมีข้อมูลสุ่มหายเลขโดดจะเรียกว่าวิธีวิเคราะห์สมบูรณ์
2. สมการถดถอย เมื่อมีข้อมูลสุ่มหายโดยตัดชุดข้อมูลที่มีค่าสุ่มหายออก โดยจะเรียกว่าวิธีสุ่มหาย
3. สมการถดถอย เมื่อมีข้อมูลสุ่มหายโดยประมาณข้อมูลสุ่มหายจากสมการถดถอยเชิงเส้นโดยจะเรียกว่าวิธีวิเคราะห์ความถดถอย
4. สมการถดถอย เมื่อใช้ตัวประมาณ maximum likelihood ประมาณข้อมูลสุ่มหายโดยจะเรียกว่าวิธี MAXIMUM LIKELIHOOD
5. สมการถดถอย เมื่อใช้ค่าเฉลี่ยประมาณข้อมูลสุ่มหายโดยจะเรียกว่าวิธีค่าเฉลี่ย
6. สมการถดถอย เมื่อใช้ค่ามัธยฐานประมาณข้อมูลสุ่มหาย โดยจะเรียกว่าวิธีค่ามัธยฐาน

1.3 วัดคุณภาพสังเคราะห์ของการวิจัย

เพื่อเปรียบเทียบวิธีการประมาณค่าสุญหายของตัวแปรอิสระในการวิเคราะห์การทดสอบโดยพิจารณาจากค่าค่าความเหลื่อมล้ำส่องเฉลี่ย (mean square error : MSE)

นอกจากนี้ยังได้พิจารณาถึงอัตราส่วนความคลาดเคลื่อนรวมของค่าประมาณที่ได้จากการทดสอบเมื่อแทนค่าข้อมูลสุญหายของตัวแปรอิสระด้วยวิธีการต่าง ๆ เมื่อเทียบกับวิธีที่ 1 ประกอบด้วย

1.4 สมมุติฐานการวิจัย

การประมาณค่าสุญหายในการวิเคราะห์การทดสอบเชิงช้อนด้วยวิธี MAXIMUM LIKELIHOOD ให้ผลลัพธ์ที่สูง

1.5 ขอบเขตการวิจัย

1. ในการศึกษาครั้งนี้จะเปรียบเทียบวิธีการประมาณค่าสุญหายในการวิเคราะห์การทดสอบโดยพิจารณาสุญหายของข้อมูลตัวแปรอิสระเป็นไปอย่างสุ่ม
2. ประชากรที่นำมาศึกษานี้การแจกแจงแบบปกติ $N(\mu, \sigma^2)$ เมื่อประชากรมีการกระจายตัวอย่างการกระจายปานกลางและการกระจายมาก ชี้งกากหรือ c.v. มีค่าเท่ากับ 0.05, 0.20, 1.00 เป็นค่าที่แสดงถึงการกระจายของข้อมูลตามลักษณะ
3. ค่าค่าความเหลื่อมล้ำของการแจกแจงปกติ $N(0, \sigma^2)$ โดยกำหนดให้ $\sigma = 5, 10, 20$ และ 25
4. ตัวแปรอิสระที่ทำการศึกษาจะพิจารณา 2, 3, 5 และ 7 ตัวแปร
5. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 30, 70 และ 100
6. ข้อมูลตัวแปรอิสระสุญหายในสัดส่วนที่เท่ากันทุกตัวแปรและจะศึกษาในกรณีที่มีข้อมูลสุญหายร้อยละ 5, 10 และ 15
7. การวิจัยครั้งนี้สร้างแบบจำลองข้อมูลใหม่สถานะการผู้คนที่ต้องการศึกษาโดยอาศัยเทคนิคคอมพิวเตอร์โล ชิมุเลชัน โดยจะกระทำข้า 700 ครั้ง

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการเลือกวิธีประมาณพั่งสุกหมายของคัวแบบปารอสระในการวิเคราะห์
การผลิตออก



ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย