ภาคผนวก ก
สถิติที่ใช้ในการวิเคราะห์ข้อมูล

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ก. สถิติเชิงพรรณนา (Descriptive Statistics) สถิติเชิงพรรณนาที่ใช้ในการบรรยายจะถูกแบ่งออกเป็น 2 กลุ่มดังนี้

1. สถิติเชิงพรรณนาสำหรับข้อมูลเชิงกลุ่มหรือข้อมูลเชิงคุณภาพ ในที่นี้หมายถึง ข้อมูลสเกลแบ่งกลุ่ม (Nominal Scale) และสเกลอันดับ (Ordinal Scale) วิธีการที่ใช้สถิติเชิงพรรณนาในการบรรยายจะประกอบไปด้วย

- การแจกแจงความถี่
- ค่าร้อยละ (Percentage) คือการคำนวณหาสัดส่วนของข้อมูลในแต่ละตัวเทียบกับข้อมูลรวมทั้งหมด โดยให้ข้อมูลรวมทั้งหมดมีค่าเป็นร้อย (รศ.ธานินทร์ ศิลป์จารุ, 2552)

$$\text{ร้อยละ}(\%) = \frac{X}{N} \times 100$$

เมื่อ X แทน จำนวนข้อมูล (ความถี่) ที่ต้องการนำมาหาค่าร้อยละ
N แทน จำนวนข้อมูลทั้งหมด

2. สถิติเชิงพรรณนาสำหรับข้อมูลเชิงปริมาณ ในที่นี้หมายถึง ข้อมูลสเกลแบบช่วง (Interval Scale) และสเกลอัตราส่วน (Ratio Scale) ซึ่งเป็นข้อมูลที่จะต้องคำนวณค่าสถิติเพื่อสรุปลักษณะข้อมูล โดยค่าสถิติจะแสดงค่าการกระจายของข้อมูล

- ค่าเฉลี่ย (Average) เป็นค่ากลางที่ใช้กับข้อมูลสเกลแบบช่วงและสเกลอัตราส่วน (รศ. ธานินทร์ ศิลป์จารุ, 2552)

$$\overline{X} = \frac{\sum x}{N}$$

เมื่อ $\overline{X}$ แทน ค่าเฉลี่ย
$\sum x$ แทน ผลรวมของข้อมูลทั้งหมด
N แทน จำนวนข้อมูลทั้งหมด

ค่าการกระจายของข้อมูลเชิงปริมาณ เป็นค่าที่แสดงความแตกต่างของข้อมูลถ้าข้อมูลชุดใดมีการกระจายมาก แสดงว่าค่าของข้อมูลชุดนั้นมีความแตกต่างกันมากค่าการกระจายของข้อมูลเชิงปริมาณประกอบด้วย

- ค่าพิสัย (Range) คือ ค่าผลต่างของข้อมูลที่มีค่ามากที่สุด (Max) กับตัวที่มีค่าน้อยที่สุด (Min) ค่าพิสัยจะเป็นวิธีการวัดการกระจายของข้อมูลอย่างหยาบ (รศ.ธานินทร์ ศิลป์จารุ, 2552)

ค่าพิสัย (Range) = ค่าสูงสุด – ค่าต่ำสุด

- ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็นค่ารากที่สองของผลรวมของความแตกต่างระหว่างข้อมูลดิบกับค่าเฉลี่ยยกกำลังสอง (Sum of squares ของผลต่าง) หารด้วยจำนวนข้อมูลทั้งหมด (รศ.ธานินทร์ ศิลป์จารุ, 2552)

$$S = \sqrt{\frac{\sum(x + \bar{x})^2}{(n-1)}}$$

เมื่อ S แทน ค่าความเบี่ยงเบนมาตรฐาน

x แทน ข้อมูลแต่ละจำนวน

$\bar{x}$ แทน ค่าเฉลี่ย (Mean) ของข้อมูลในชุดนั้น

n แทน จำนวนข้อมูลจากกลุ่มตัวอย่าง

ข. สถิติเชิงอนุมานหรือเชิงสรุปอ้างอิง (Inferential Statistics) เป็นเทคนิคทางสถิติที่ศึกษาข้อมูลจากกลุ่มตัวอย่างเพื่อมุ่งสรุปอ้างอิงไปยังลักษณะของประชากร การทำให้ได้มาซึ่งกลุ่มตัวอย่างที่มีความเป็นตัวแทนของประชากร มีความ สำคัญยิ่งต่อการสรุปอ้างอิงผลจากค่าสถิติไปยังค่าประชากร ได้แก่

1. สถิติอ้างอิงแบบมีค่าพารามิเตอร์ (Parameter Inference) ใช้กับข้อมูลที่มีระดับการวัดตัวแปรเป็นระดับมาตราอันตรภาค หรือระดับมาตราอัตราส่วน ที่ลักษณะข้อมูลมีการแจกแจงแบบปกติหรือประชากรที่มีค่าแปรปรวนไม่แตกต่างกันมาก สถิติที่ใช้ทดสอบสมมุติฐานได้แก่

- Independent Sample T-test ใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่างว่ามีค่าแตกต่างกันหรือไม่ โดยที่ตัวแปรต้นเป็นตัวแปรเชิงคุณภาพ 2 กลุ่ม และตัวแปรตามเป็นตัวแปรเชิงปริมาณ การทดสอบสมมติฐานโดยกำหนดให้ค่านัยสำคัญทางสถิติที่ระดับ(Level Significance) 0.05

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

เมื่อ t แทน ค่าสถิติที่ใช้พิจารณาใน t-distribution

$\overline{X_1}$, $\overline{X_2}$ แทน ค่าเฉลี่ยของกลุ่มตัวอย่างที่ 1 และ 2

S_1^2 , S_2^2 แทน ความแปรปรวนของกลุ่มตัวอย่างที่ 1 และ 2

n_1 , n_2 แทน ขนาดของกลุ่มตัวอย่างที่ 1 และ 2

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

df แทน ค่าคงที่ที่อาจมีค่าเป็นศูนย์ หรือมีค่าเป็นบวก หรือลบก็ได้

- One-Way ANOVA หรือ F-test ใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มตัวอย่าง โดยที่ตัวแปรต้นเป็นตัวแปรเชิงคุณภาพมากกว่า 2 กลุ่ม และตัวแปรตามเป็นตัวแปรเชิงปริมาณ การทดสอบสมมติฐานโดยกำหนด ให้ค่านัยสำคัญทางสถิติที่ระดับ (Level Significance) 0.05

$$F = \frac{S_1^2}{S_2^2}$$

เมื่อ F แทน ค่าการแจกแจงที่ใช้ในการพิจารณาใน F-Distribution

S_1^2 แทน ค่าการกระจายของข้อมูลกลุ่มที่ 1

S_1^2 แทน ค่าการกระจายของข้อมูลกลุ่มที่ 2

กรณีพบความแตกต่างอย่างมีนัยสำคัญทางสถิติ จะทำการทดสอบความแตกต่างกันเป็นรายคู่ ที่ระดับนัยสำคัญ 0.05 หรือระดับความเชื่อมั่น 95% โดยใช้สูตรตามวิธี Scheffe เพื่อเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่าง

$$Sf = \sqrt{(k-1)F_{(\alpha,k-1,N-k)}}\sqrt{MSw\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

เมื่อ MSw แทน ค่าเฉลี่ยของความแปรปรวนภายในกลุ่ม

n_i , n_j แทน จำนวนตัวอย่างในกลุ่มที่ I และ j

N แทน จำนวนตัวอย่างทั้งหมด

k แทน จำนวนกลุ่ม

ค่า $F_{(\alpha,k-1,N-k)}$ ได้มาจากการเปิดตาราง F ที่ α = 0.05, df_1 = k-1, df_2 = N-k

2. สถิติอ้างอิงแบบไม่มีพารามิเตอร์ (Non Parameter Inference) ใช้กับข้อมูล หรือตัวแปรที่ไม่สามารถใช้กับสถิติอ้างอิงแบบมีพารามิเตอร์ได้ มีลักษณะระดับมาตราเรียงลำดับ และลักษณะข้อมูลไม่ได้มีการแจกแจงแบบปกติ สถิติที่ใช้ในการทดสอบสมมุติฐานได้แก่

- การวิเคราะห์การถดถอยเชิงเส้นตรงพหุคูณ (Multiple Linear Regression) เป็นการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรอิสระมากกว่า 1 ตัว และตัวแปรตาม 1 ตัว ผลที่ได้จากการวิเคราะห์สามารถสรุปได้เป็นความสัมพันธ์ อยู่ในรูปของสมการเส้นตรง และสามารถอธิบายและเปรียบเทียบความสัมพันธ์ของตัวแปรอิสระแต่ละตัว ว่าตัวแปรใดมีอิทธิพลต่อการเปลี่ยนแปลง Y มากที่สุด สามารถเขียนอยู่ในรูปแบบความสัมพันธ์ในรูปของสมการได้ดังนี้

<u>สมการพยากรณ์ในรูปคะแนนดิบ</u>

ในการวิเคราะห์การถดถอยพหุคูณ จะทำให้ได้สมการณ์เชิงเส้นตรงในรูปคะแนนดิบดังนี้

$$Y' = a + b_1x_1 + b_2x_2 + ... + b_kx_k$$

เมื่อ Y' แทน คะแนนพยากรณ์ของตัวเกณฑ์ (ตัวแปรตาม)

a แทน ค่าคงที่ของสมการพยากรณ์ในรูปแบบคะแนนดิบ

$b_1,....b_k$ แทน น้ำหนักคะแนนหรือสัมประสิทธิ์การถดถอยของตัวพยากรณ์ตัวที่ 1 ถึง ตัวที่ k ตามลำดับ

$X_1,....X_k$ แทน คะแนนของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ 1 ถึง ตัวที่ k ตามละดับ

k แทน จำนวนตัวพยากรณ์ (ตัวแปรอิสระ)

<u>สมการณ์พยากรณ์ในรูปคะแนนมาตรฐาน</u>

ถ้าคะแนนตัวพยากรณ์ทุกตัวเปลี่ยนเป็นรูปคะแนนมาตรฐาน และต้องการพยากรณ์เกณฑ์ในรูปของคะแนนมาตรฐานแล้ว สมการพยากรณ์จะเขียนได้ดังนี้

$$Z'_Y = \beta_1 Z_1 + \beta_2 Z_2 + ... + \beta_k Z_k$$

เมื่อ Z'_Y แทน คะแนนพยากรณ์ในรูปของคะแนนมาตรฐานของตัวเกณฑ์ (ตัวแปรตาม)

$\beta_1 , ...\beta_k$ แทน สัมประสิทธิ์การถดถอยในรูปของคะแนนมาตรฐานของตัวพยากรณ์ตัวที่ 1 ถึง ตัวที่ k ตามลำดับ

$Z_1 , ...Z_k$ แทน คะแนนมาตรฐานของตัวพยากรณ์ (ตัวแปรอิสระ) ตัวที่ 1 ถึง ตัวที่ k ตามลำดับ

k แทน จำนวนตัวพยากรณ์