

ผนวก ข

การพิสูจน์การแปลงตัวสถิติทดสอบ

โดยปกตินั้นในการดำเนินงานวิจัยมักจะไม่ทราบค่าพารามิเตอร์ ดังนั้นในการศึกษาครั้งนี้จึงทำการศึกษาในกรณีไม่ทราบค่าพารามิเตอร์ กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่อง ที่มี $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n ซึ่งพารามิเตอร์ในที่นี้คือค่าเฉลี่ยประชากร (μ) และความแปรปรวนประชากร (σ^2) โดยการใช้คุณสมบัติของตัวประมาณไม่เอนเอียงและความแปรปรวนต่ำสุด (Minimum Variance Unbiased Estimator) นั้น เป็นที่ทราบดีว่า

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

และ
$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

เป็นตัวประมาณพารามิเตอร์ที่ดีที่สุดของค่าเฉลี่ยประชากร และความแปรปรวนประชากรตามลำดับ โดยจะเรียก $\bar{X}$ ว่าค่าเฉลี่ยตัวอย่าง และเรียก S^2 ว่าความแปรปรวนตัวอย่าง

และในที่นี้จะทำการแปลงข้อมูล X_i ให้อยู่ในรูป X'_i เพื่อให้ X'_i มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 ดังนั้นจะได้

$$X'_i = \frac{X_i - \bar{X}}{S}$$

1. กรณีตัวสถิติทดสอบ Anderson-Darling (A^2)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่อง ที่มี $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n และมี $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ เป็นสถิติอันดับ

ตัวสถิติทดสอบ Anderson-Darling (A^2) คือ

$$A^2 = -n - \frac{1}{n} \left[\sum_{i=1}^n (2i-1) [\ln F(X_{(i)}) + \ln (1-F(X_{(n-i+1)}))] \right]$$

โดยที่ $F(X_{(i)}) =$ ความน่าจะเป็นสะสมของการแจกแจงปกติมาตรฐาน $= \Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right)$

การแปลงข้อมูล $X_{(i)}$ ให้อยู่ในรูป $X'_{(i)}$ ซึ่ง $X'_{(i)}$ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

$$\text{ให้ } X'_{(i)} = \frac{X_{(i)} - \bar{X}}{S}$$

$$\text{โดยที่ } \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\text{และ } S^2 = \frac{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}{n-1}$$

$$\begin{aligned} \text{พิจารณา } F(X'_{(i)}) &= \Phi\left(\frac{X'_{(i)} - \bar{X}'}{S'}\right) \\ &= \Phi\left(\frac{\frac{X_{(i)} - \bar{X}}{S} - 0}{1}\right) \\ &= \Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right) \end{aligned}$$

2. กรณีตัวสถิติทดสอบ Shapiro-Wilk (W)

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่อง ที่มี $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n และมี $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ เป็นสถิติอันดับ

ตัวสถิติทดสอบ Shapiro-Wilk (W) คือ

$$W = \frac{\left[\sum_{i=1}^k a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) \right]^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}$$

การแปลงข้อมูล $X_{(i)}$ ให้อยู่ในรูป $x'_{(i)}$ ซึ่ง $x'_{(i)}$ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

ให้ $x'_{(i)} = \frac{x_{(i)} - \bar{x}}{s}$

โดยที่ $\bar{x} = \frac{\sum_{i=1}^n x_{(i)}}{n}$

และ $s^2 = \frac{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}{n-1}$

พิจารณา $W' = \frac{\left[\sum_{i=1}^k a_{n-i+1} (x'_{(n-i+1)} - x'_{(i)}) \right]^2}{\sum_{i=1}^n (x'_{(i)} - \bar{x}')^2}$

$$= \frac{\left[\sum_{i=1}^k a_{n-i+1} \left(\frac{x_{(n-i+1)} - \bar{x}}{s} - \frac{x_{(i)} - \bar{x}}{s} \right) \right]^2}{\sum_{i=1}^n \left(\frac{x_{(i)} - \bar{x}}{s} - 0 \right)^2}$$
$$= \frac{\frac{1}{s^2} \left[\sum_{i=1}^k a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) \right]^2}{\frac{1}{s^2} \sum_{i=1}^n (x_{(i)} - \bar{x})^2} = W$$

3. กรณีตัวสถิติทดสอบ Shapiro-Francia (W')

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่อง ที่มี $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n และมี $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ เป็นสถิติอันดับ

ตัวสถิติทดสอบ Shapiro-Francia (W') คือ

$$W' = \frac{\left[\sum_{i=1}^n m_i X_{(i)} \right]^2}{\sum_{i=1}^n m_i^2 \cdot \sum_{i=1}^n (X_{(i)} - \bar{X})^2}$$

การแปลงข้อมูล $X_{(i)}$ ให้อยู่ในรูป $X'_{(i)}$ ซึ่ง $X'_{(i)}$ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

ให้ $X'_{(i)} = \frac{X_{(i)} - \bar{X}}{S}$

โดยที่ $\bar{X} = \frac{\sum_{i=1}^n X_{(i)}}{n}$

และ $S^2 = \frac{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}{n-1}$

พิจารณา $W' = \frac{\left[\sum_{i=1}^n m_i X'_{(i)} \right]^2}{\sum_{i=1}^n m_i^2 \cdot \sum_{i=1}^n (X'_{(i)} - \bar{X}')^2}$

$$= \frac{\left[\sum_{i=1}^n m_i \frac{X_{(i)} - \bar{X}}{S} \right]^2}{\sum_{i=1}^n m_i^2 \cdot \sum_{i=1}^n \left(\frac{X_{(i)} - \bar{X}}{S} - 0 \right)^2}$$

$$= \frac{\left[\sum_{i=1}^n m_i (X_{(i)} - \bar{X}) \right]^2}{\sum_{i=1}^n m_i^2 \cdot \sum_{i=1}^n (X_{(i)} - \bar{X})^2}$$

$$= \frac{\left[\sum_{i=1}^n m_i X_{(i)} \right]^2}{\sum_{i=1}^n m_i^2 \cdot \sum_{i=1}^n (X_{(i)} - \bar{X})^2}$$

(Ryan and Joiner (1973) อ้างจาก Steve Verrill and Richard A. Johnson)

ดังนั้น $W' = W$

4. กรณีตัวสถิติทดสอบของชาวและวู

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่อง ที่มี $X_1, X_2, \dots, X_n$ เป็นตัวอย่างสุ่มขนาด n และมี $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ เป็นสถิติอันดับ

ตัวสถิติทดสอบชาวและวู คือ

$$Z_A = - \sum_{i=1}^n \left[\frac{\log F_0(X_{(i)})}{n-i+0.5} + \frac{\log [1-F_0(X_{(i)})]}{i-0.5} \right]$$

$$Z_C = \sum_{i=1}^n \left[\log \left(\frac{\frac{F_0(X_{(i)})^{-1} - 1}{(n-0.5)} - 1}{(i-0.75)} \right) \right]^2$$

$$Z_K = \max_{1 \leq i \leq n} \left[(i-0.5) \log \left[\frac{i-0.5}{nF_0(X_{(i)})} \right] + [n-i+0.5] \log \frac{n-i+0.5}{n(1-F_0(X_{(i)}))} \right]$$

โดยที่ $F(X_{(i)})$ = ความน่าจะเป็นสะสมของการแจกแจงปกติมาตรฐาน

$$= \Phi \left(\frac{X_{(i)} - \bar{X}}{S} \right)$$

การแปลงข้อมูล $X_{(i)}$ ให้อยู่ในรูป $X_{(i)}'$ ซึ่ง $X_{(i)}'$ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

$$\text{ให้ } X'_{(i)} = \frac{X_{(i)} - \bar{X}}{S}$$

$$\text{โดยที่ } \bar{X} = \frac{\sum_{i=1}^n X_{(i)}}{n}$$

$$\text{และ } S^2 = \frac{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}{n-1}$$

$$\begin{aligned} \text{พิจารณา } F(X'_{(i)}) &= \Phi\left(\frac{X'_{(i)} - \bar{X}'}{S'}\right) \\ &= \Phi\left(\frac{\frac{X_{(i)} - \bar{X}}{S} - 0}{1}\right) \\ &= \Phi\left(\frac{X_{(i)} - \bar{X}}{S}\right) \\ &= F(X_{(i)}) \end{aligned}$$

ดังนั้น

$$\begin{aligned} Z'_A &= -\sum_{i=1}^n \left[\frac{\log F_0(X'_{(i)})}{n-i+0.5} + \frac{\log [1-F_0(X'_{(i)})]}{i-0.5} \right] \\ &= -\sum_{i=1}^n \left[\frac{\log F_0(X_{(i)})}{n-i+0.5} + \frac{\log [1-F_0(X_{(i)})]}{i-0.5} \right] \\ &= Z_A \end{aligned}$$

$$\begin{aligned}
 \text{และ } Z'_c &= \sum_{i=1}^n \left[\log \left(\frac{\frac{F_0(X'_{(i)})^{-1} - 1}{(n-0.5)}}{\frac{(i-0.75)}{(i-0.75)} - 1} \right) \right]^2 \\
 &= \sum_{i=1}^n \left[\log \left(\frac{F_0(X_{(i)})^{-1} - 1}{(n-0.5)} - 1 \right) \right]^2 \\
 &= Z_c
 \end{aligned}$$

$$\begin{aligned}
 \text{และ } Z'_K &= \max_{1 \leq i \leq n} \left[(i-0.5) \log \left[\frac{i-0.5}{nF_0(X'_{(i)})} \right] + [n-i+0.5] \log \frac{n-i+0.5}{n(1-F_0(X'_{(i)}))} \right] \\
 &= \max_{1 \leq i \leq n} \left[(i-0.5) \log \left[\frac{i-0.5}{nF_0(X_{(i)})} \right] + [n-i+0.5] \log \frac{n-i+0.5}{n(1-F_0(X_{(i)}))} \right] \\
 &= Z_K
 \end{aligned}$$