บทที่ 2

ตัวสถิติและผลงานวิจัยที่เกี่ยวข้อง

การวิเคราะห์ข้อมูลเกี่ยวกับอายุการใช้งาน (life time) หรือช่วงเวลาของความล้มเหลว (failure time) ภายใต้การแจกแจงแบบเอกซ์โปเนนเชียล 2 พารามิเตอร์เป็นเรื่องที่มีความสำคัญเรื่องหนึ่ง ซึ่งมักจะเกี่ยวข้องกับงานทางด้านวิศวกรรมศาสตร์ วิทยาศาสตร์และทางการแพทย์ ข้อมูลลักษณะดังกล่าวมักจะเกิดการตัดทิ้งอันเนื่องมาจากข้อกำหนดของเวลาและข้อจำกัดอื่น ๆ ในบทนี้จึงกล่าวถึงรายละเอียดเกี่ยวกับชนิดของข้อมูลที่มีค่าถูกตัดทิ้ง (type of data censoring) ทฤษฎีของการแจกแจงแบบเอกซ์โปเนนเชียล 2 พารามิเตอร์พร้อมทั้งตัวอย่างการใช้สถิติทดสอบ และผลงานวิจัยที่เกี่ยวข้องซึ่งมีรายละเอียดต่าง ๆ ดังนี้

2.1 ชนิดของข้อมูลที่มีค่าถูกตัดทิ้ง (Type Of Data Censoring)

ก) ข้อมูลถูกตัดทิ้งประเภทที่ 1 (type I censoring) เป็นลักษณะของข้อมูลที่กำหนดระยะเวลาของข้อมูลที่ถูกตัดทิ้งไว้ล่วงหน้า เช่น ในการศึกษาเกี่ยวกับความคงทนของเครื่องจักรชนิดหนึ่งโดยศึกษาในระยะเวลา 8,000 ชั่วโมง ถ้าเครื่องจักรทำงานเกิน 8,000 ชั่วโมงจะถือว่าอายุการใช้งานของเครื่องจักรที่เกินมานี้เป็นข้อมูลที่มีค่าถูกตัดทิ้ง เพราะว่าเราไม่สามารถบันทึกข้อมูลได้

ข) ข้อมูลถูกตัดทิ้งประเภทที่ 2 (type II censoring) เป็นลักษณะของข้อมูลที่จะต้องกำหนดจำนวนข้อมูลที่ถูกตัดทิ้งไว้ล่วงหน้า เช่น ในการทดลองเกี่ยวกับประสิทธิภาพของเครื่องใช้ไฟฟ้าชนิดหนึ่งจำนวน 20 เครื่อง แทนที่เราจะทำการทดลองอย่างต่อเนื่องจนกระทั่งเครื่องใช้ไฟฟ้าที่นำมาทดลองทั้ง 20 เครื่องจะเสียหรือใช้งานไม่ได้ เราจะหยุดทำการทดลอง ณ เวลาหนึ่งของการเกิดเหตุการณ์ดังกล่าวครั้งที่ r เช่น หยุดทำการทดลองเมื่อเครื่องใช้ไฟฟ้าที่นำมาทำการทดลองจะเสียเป็นเครื่องที่ 15 ซึ่งการทดลองแบบนี้จะช่วยประหยัดทั้งเวลาและเงินของผู้ที่ทำการทดลอง

ค) **ข้อมูลถูกตัดทิ้งแบบสุ่ม (random censoring)** ข้อมูลที่มีค่าถูกตัดทิ้งในลักษณะเช่นนี้จะเป็นไปโดยสุ่มซึ่งส่วนใหญ่เราจะพบมากในทางการแพทย์ เช่น ในกรณีที่คนไข้บางรายขาดการติดต่อกับโรงพยาบาลหลังจากได้รับการรักษาจากโรงพยาบาล จึงทำให้ทางโรงพยาบาลไม่สามารถเก็บบันทึกข้อมูลได้ครบตามที่กำหนดไว้

ในการวิจัยครั้งนี้ผู้วิจัยจะทำการศึกษาทั้งในกรณีที่ข้อมูลสมบูรณ์และมีค่าถูกตัดทิ้งทางขวา ซึ่งข้อมูลที่มีค่าถูกตัดทิ้งทางขวานี้เป็นกรณีเฉพาะของการที่มีข้อมูลถูกตัดทิ้งในลักษณะดังกล่าวข้างต้น การทดลองที่ทำให้ข้อมูลมีค่าถูกตัดทิ้งทางขวา เช่น การทดลองเกี่ยวกับอายุการใช้งานของเครื่องจักรที่ใช้ในการผลิตในโรงงานอุตสาหกรรม แล้วบันทึกระยะเวลาหรือจำนวนชั่วโมงที่เครื่องจักรนั้นจะเสียหรือใช้งานไม่ได้ ในระหว่างที่ดำเนินการทดลองอยู่เครื่องจักรที่เสียหรือใช้งานไม่ได้จะเป็นข้อมูลที่ไม่ถูกตัดทิ้ง (uncensored data) เมื่อสิ้นสุดการทดลองเครื่องจักรที่ยังคงอยู่ในสภาพที่ใช้งานได้ดีจะเป็นเครื่องจักรที่ไม่ทราบอายุการใช้งานที่แน่นอน ข้อมูลนี้จะเป็นข้อมูลที่มีค่าถูกตัดทิ้งทางขวา (right censored data)

2.2 การแจกแจงแบบเอกซ์โปเนนเชียล 2 พารามิเตอร์

จากที่ได้กล่าวมาแล้วในบทที่ 1 ว่าการแจกแจงแบบเอกซ์โปเนนเชียล 2 พารามิเตอร์เป็นการแจกแจงที่เรามักจะนำมาใช้เป็นตัวแบบในการศึกษาเกี่ยวกับอายุการใช้งาน โดยมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x_j) = \begin{cases} \dfrac{1}{\theta}\exp\left(-\dfrac{(x_j-\beta_j)}{\theta}\right) & , x_j \geq \beta_j \text{ , } \theta > 0 \\ & , j=1,2,..,k \\ 0 & \text{อื่น ๆ} \end{cases}$$

เมื่อ β_j เป็นพารามิเตอร์แสดงตำแหน่ง (location parameter)

และ θ_j เป็นพารามิเตอร์แสดงสเกล (scale parameter)

ให้ $X_1,X_2,,\ldots,X_r$ เป็นค่าสังเกตที่น้อยที่สุด r_j จากค่าสังเกต n ค่า ที่เป็นอิสระซึ่งกันและกันของประชากรใด ๆ มีฟังก์ชันความหนาแน่นร่วม (joint probability density function) ซึ่งอยู่ในรูปของ

$$g(x_1,x_2,..,x_r;\ \beta,\theta) = \frac{n!}{(n-r)!}\ \frac{1}{\theta^{r_j}} \exp\left[-\frac{1}{\theta}\ \sum_{i=1}^{r} (x_{ji}-\beta_j)+(n_j-r_j)(x_{rj}-\beta_j)\right]$$

$$,\theta \geq 0\ ,x_j > \beta_j$$

∴ ฟังก์ชันการแจกแจงร่วมของอายุการใช้งาน $X_{j1} \leq X_{j2} \leq \ldots \leq X_{jr}$ เมื่อ $j= 1,2\ldots,k$ คือ

$$f(x_{j1},x_{j2},..,x_{jr});\beta,\theta) = \prod_{j=1}^{k} g(x_1,x_2,\ldots,x_r;\beta,\theta)$$

$$= \prod_{j=1}^{k} \frac{n!}{(n-r)!}\ \theta^{-\sum_{j=1}^{k} r_j} \exp -\frac{1}{\theta}\left[\sum_{j=1}^{k}\sum_{i=1}^{r}(x_{ji}-\beta_j)+(n_j-r_j)(x_{jr}-\beta_j)\right]$$

$$,\theta \geq 0\ ,\ x_j > \beta_j$$

กำหนดให้ $y_j = \min(X_1,X_2,\ldots,X_{rj})$ เป็นตัวประมาณของพารามิเตอร์แสดงตำแหน่ง ซึ่ง y_j เป็นตัวสถิติอันดับที่หนึ่งของข้อมูลที่มีค่าสังเกตน้อยที่สุด r_j จากประชากร π_j เมื่อ $j = 1,2,\ldots,k$

กำหนดให้ $\hat{\theta}$ เป็นตัวประมาณที่ไม่เอนเอียงและมีค่าความแปรปรวนต่ำสุด (UMVU) ของพารามิเตอร์แสดงสเกล ซึ่งแสดงถึงค่าเฉลี่ยของประชากร ตัวประมาณของ θ จะอยู่ในรูปของ

$$\hat{\theta} = R^{-1} \sum_{j=1}^{k}\left[\sum_{i=1}^{r} (x_{ji}-X_{j1}) + (n_j-r_j)(x_{jr}-X_{j1})\right]$$

โดยที่ $R = \sum_{j=1}^{k} r_j$

2.3 ตัวสถิติทดสอบที่ใช้ในการวิจัย

ก) ตัวสถิติทดสอบ IP

ในปี ค.ศ.1963 R.V.Hogg และ T.A.Tanis เสนอตัวสถิติทดสอบ IP โดยพัฒนามาจากตัวสถิติทดสอบ LR ซึ่งเสนอโดย Epstien และ Tsao เมื่อปี ค.ศ.1953 วิธีการดังกล่าวนี้มีหลักการของการกระทำซ้ำโดยเริ่มจากการนำตัวอย่างของประชากร 2 กลุ่มแรกที่ต้องการทดสอบมาทำการทดสอบสมมุติฐานก่อน ถ้าผลการทดสอบยอมรับสมมุติฐานจะนำตัวอย่างของประชากรทั้ง 2 กลุ่มนี้มารวมกันและนำไปทดสอบสมมุติฐานกับประชากรในกลุ่มที่ 3 ต่อไป ถ้าผลการทดสอบยอมรับสมมุติฐานของประชากรทั้ง 3 กลุ่มนี้ก็จะนำประชากรของกลุ่มที่ 3 ไปรวมกับประชากร 2 กลุ่มแรกและทำการทดสอบสมมุติฐานกับกลุ่มต่อไป เรากระทำเช่นนี้จนกระทั่งเกิดการปฏิเสธสมมุติฐานหรือยอมรับสมมุติฐานในทุกประชากรที่นำมาทดสอบ

ให้ $X_{i1} \le X_{i2} \le \dots \le X_{i,n-s}$, $i=1,2,\dots,k$ เป็นตัวสถิติอันดับ
เมื่อ s_i = จำนวนข้อมูลที่มีค่าถูกตัดทิ้งของกลุ่มที่ i

กำหนดให้ $Z_i = \min[X_{i1}]$, $i=1,2,\dots,k$

$$Z_1 = X_{11}$$

$$Z_2 = \min[X_{11},X_{21}]$$

$$\cdot$$

$$Z_5 = \min[X_{11},X_{21},X_{31},X_{41},X_{51}]$$

ให้
$$T_i = \sum_{j=1}^{i} \{ \sum_{h=1}^{r} X_{jh} - Z_i) + (n_j-r_j)(X_{jr}-Z_i)\}$$

$$V_i = \sum_{j=1}^{r} \{(X_{ij}-X_{i1}) + (n_j-r_j)(X_{ir}-X_{i1})\}$$

$$U_i = \left(\sum_{j=1}^{i-1} n_j \right) Z_{i-1} + n_i X_{i1} - \left(\sum_{j=1}^{i} n_j \right) Z_i$$

จะได้ว่า

$$J_i = \frac{U_i[2(r_1+r_2+\ldots+r_i)-4]}{2[T_{i-1} + V_i]}$$

เราจะปฏิเสธสมมุติฐานเมื่อ $J_i > F_\alpha(2,2(r_1+r_2\ldots r_i)-4)$

การคำนวณค่าสถิติของตัวสถิติทดสอบ IP แสดงได้ดังตัวอย่างที่ 1

ตัวอย่างที่ 1 วิศวกรผู้หนึ่งต้องการตรวจสอบอายุการใช้งานของเครื่องจักร 3 ชนิด ๆ ชนิดละ 10 เครื่องที่ใช้ในการผลิตสินค้าชนิดหนึ่ง โดยวิศวกรผู้นี้บันทึกเวลาที่เครื่องจักรดังกล่าวจะหยุดทำ-งานเป็นครั้งแรก (ชั่วโมง) กำหนดช่วงเวลาที่บันทึกข้อมูลภายใน 20 ชั่วโมง ปรากฎว่าหลังจาก 20 ชั่วโมง มีเครื่องจักรชนิดที่ 1 จำนวน 2 เครื่อง ชนิดที่ 2 จำนวน 3 เครื่อง และชนิดที่ 3 จำนวน 1 เครื่อง ซึ่งยังคงทำงานต่อไปและไม่สามารถบันทึกข้อมูลได้ ข้อมูลที่เก็บรวบรวมได้เป็นดังนี้

เครื่องจักรชนิดที่ 1 : 3.5, 4.7, 5.2, 5.5, 7.8, 8.3, 9.0, 10.2,_,_

เครื่องจักรชนิดที่ 2 : 3.0, 3.2, 6.3, 7.5, 9.1, 10.3, 11.2,_,_,_

เครื่องจักรชนิดที่ 3 : 3.6, 4.0, 4.2, 4.4, 5.9, 6.8, 8.1,10.8,12.6,_

ต้องการทดสอบ H_o : พารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 3 กลุ่มมีค่าเท่ากัน

เทียบกับ H_1 : พารามิเตอร์แสดงตำแหน่งของข้อมูลอย่างน้อย 2 กลุ่มมีค่าไม่เท่ากัน

จากข้อมูล $n_1=n_2=n_3=10$, $s_1=2, s_2=3, s_3=1$

$r_1=8$, $r_2=7$, $r_3=9$

$Z_1 = 3.5$

$Z_2 = \min[3.5, 3.0] = 3.0$

$Z_3 = \min[3.5, 3.0, 3.6] = 3.0$

ตารางที่ 2.1 แสดงการคำนวณค่าตัวสถิติทดสอบ IP

i	x_{1i}	x_{2i}	x_{3i}	$x_{1i}-Z_1$	$x_{2i}-Z_2$	$x_{3i}-Z_3$
1	3.5	3.0	3.6	-	-	-
2	4.7	3.2	4.0	1.2	0.2	0.4
3	5.2	6.3	4.2	1.7	3.3	1.6
4	5.5	7.5	4.4	2.0	4.5	1.8
5	7. 8	9.1	5.9	4.3	6.1	2.3
6	8.3	10.3	6.8	2.8	7.3	3.2
7	9.0	11.2	8.1	5.5	7.9	4.5
8	10.2	-	10.8	6.7	-	7.2
9	-	-	12.6	-	-	-
10	-	-	-	-	-	-
รวม	54.2	50.6	60.4	24.2	29.3	30.0

$$T_2 = \{(24.2 + (2)(10.2-3.5) + 29.3 + (3)(11.2-3.0)\}$$

$$= 91.5$$

$V_3 = \{30.0 + (1)(12.6-3.6)\} = 3.9$

$U_3 = (20)(3.0) + 10(3.6) -30(3.0) = 6.0$

จะได้ว่า $J_3 = \frac{6[2(24)-4]}{2[91.5+39.0]}$

$= 264/261$

$= 1.01$

จากตารางการแจกแจงเอฟถ้า $\alpha=0.05$ จะได้ว่า $F_{0.05}(2,44) = 3.20$ เนื่องจาก $J_3 < 3.20$ $\therefore$ ยอมรับสมมุติฐานที่ว่าค่าพารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 3 กลุ่มไม่แตกต่างกันอย่างมีนัยสำคัญ หรือ อาจกล่าวได้ว่าเครื่องจักรทั้ง 3 เครื่องจะหยุดทำงานเป็นครั้งแรกในช่วงเวลาเดียวกัน

ข) ตัวสถิติทดสอบ LR

ในปี ค.ศ.1953 Epstien และ Tsao ได้เสนอตัวสถิติทดสอบ LR เพื่อใช้ในการทดสอบการเท่ากันของพารามิเตอร์แสดงตำแหน่งของการแจกแจงแบบเอกซ์โปเนนเชียล 2 พารามิเตอร์เป็นครั้งแรก แต่วิธีการดังกล่าวมีข้อจำกัดในการคำนวณค่าสถิติที่ซับซ้อนและใช้ได้กับประชากรที่มีเพียง 2 กลุ่มประชากรเท่านั้น ต่อมาในปี ค.ศ.1983 N.Singh ได้ทำการปรับปรุงตัวสถิติทดสอบ LR ให้ใช้ทดสอบกับประชากรที่มีมากกว่า 2 กลุ่มและมีค่าลู่เข้าสู่การแจกแจงเอฟ ซึ่งสามารถหาค่าวิกฤตได้จากตารางการแจกแจงเอฟโดยทั่วไป เราจะพิจารณาค่าตัวสถิติทดสอบ LR จากค่าอัตราส่วนของผลรวมของความแตกต่างของตัวสถิติอันดับที่ 1 ในกลุ่มประชากรที่กำลังพิจารณากับกลุ่มประชากรที่อยู่ถัดไปกับผลรวมของค่าความผิดพลาดที่ได้จากตัวประมาณภาวะน่าจะเป็นสูงสุด

ให้ $X_{j1} \leq X_{j2} \leq X_{j3} \ldots \leq X_{j,r}$,$j=1,2,\ldots,k$ เป็นตัวสถิติอันดับ เมื่อ s = จำนวนข้อมูลที่มีค่าถูกตัดทิ้งในกลุ่มที่ j โดยที่ $r_j = n_j - s_j$

ตัวประมาณภาวะน่าจะเป็นสูงสุดของ β_j จะอยู่ในรูปของ

$$\hat{\beta}_j = X_{j1}$$

กำหนดให้ m_j เป็นค่าเฉลี่ยของข้อมูลซึ่งอยู่ในรูปของ

$$m_j = r_j^{-1} \left\{ \sum_{i=1}^{r} (x_{ji} - X_{j1}) + (n_j - r_j)(x_{jr} - X_{j1}) \right\}$$

จะได้ว่า

$$M = (R-k)^{-1} \sum_{j=1}^{k} r_j m_j$$

และ

$$\theta = R^{-1} \sum_{j=1}^{k} \sum_{i=1}^{r_j} (x_{ji} - X_{j1}) + (n_j - r_j)(x_{jr} - X_{j1})$$

โดยที่ $R = \sum_{j=1}^{k} r_j$

ให้ $W_j = \{2(\sum_{i=1}^{k} n_i)(X_{j1} - X_{j-1,1})\}/\theta \quad ,j=2,..,k$

เมื่อ $X_{j1} > X_{j-1,1}$

ดังนั้น $V = (\sum_{j=2}^{k} W_j)\theta$

จะได้ว่าตัวสถิติ $U = V/2(k-1)M$ และเราจะปฏิเสธสมมุติฐานเมื่อ $U \geq F_{\alpha}(2(k-1),2(R-k))$

การคำนวณค่าสถิติของตัวสถิติทดสอบ LR แสดงไว้ในตัวอย่างที่ 2

ตัวอย่างที่ 2 เภสัชกรผู้หนึ่งได้ทำการวิจัยเกี่ยวกับประสิทธิภาพในการรักษาโรคของยาบรรเทาปวด 2 ชนิด โดยทำการทดลองกับคนไข้ 2 กลุ่ม ๆ ละ 10 คน คนไข้ในกลุ่มที่ 1 ได้รับยาบรรเทาปวดชนิดที่ 1 และคนไข้กลุ่มที่ 2 จะได้รับยาบรรเทาปวดชนิดที่ 2 ตามลำดับ โดยสังเกตช่วงเวลาในการบรรเทาอาการปวดของยาทั้ง 2 ชนิด ซึ่งบันทึกระยะเวลาเป็นชั่วโมง ข้อมูลที่เก็บรวบรวมได้เป็นดังนี้

คนไข้กลุ่มที่ 1 : 1.03, 1.09, 1.12, 1.35, 1.54, 1.85, 2.15, 2.42, 2.64, 2.70

คนไข้กลุ่มที่ 2 : 1.10, 1.17, 1.25, 1.39, 1.50, 1.59, 1.99, 2.47, 2.75, 2.80

ต้องการทดสอบ H_0 : พารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 2 กลุ่มมีค่าเท่ากัน

เทียบกับ H_1 : พารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 2 กลุ่มมีค่าไม่เท่ากัน

จากข้อมูล $n_1=n_2=10$, $s_1=0$, $s_2=0$, $r_1=10$, $r_2=10$ พารามิเตอร์แสดงตำแหน่งของประชากรในกลุ่มที่ 1 (β_1) = 1.03 และพารามิเตอร์แสดงตำแหน่งของประชากรในกลุ่มที่ 2 (β_2) = 1.10

ดังนั้นค่าของ

$$m_1 = r_1^{-1}\{ \sum_{i=1}^{r1} (x_{1i}-X_{11}) + (n_1-r_1)(x_{1.10}-X_{11})\}$$

และ

$$m_2 = r_2^{-1}\{ \sum_{i=2}^{r2} (x_{2i}-X_{21}) + (n_2-r_2)(x_{2.10}-X_{21})\}$$

สามารถคำนวณได้ดังตารางข้างล่างนี้

ตารางที่ 2.2 แสดงการคำนวณค่าตัวสถิติทดสอบ LR

i	x_{1i}	x_{2i}	$x_{1i}-X_{1i}$	$x_{2i}-X_{2i}$
1	1.03	1.10	-	-
2	1.09	1.17	0.06	0.07
3	1.12	1.25	0.09	0.15
4	1.35	1.39	0.32	0.29
5	1.54	1.50	0.51	0.40
6	1.85	1.59	0.82	0.49
7	2.15	1.99	1.12	0.89
8	2.42	2.47	1.39	1.37
9	2.64	2.75	1.61	1.65
10	2.70	2.80	1.67	1.70
รวม	17.89	18.01	7.59	7.01

แทนค่า $m_1 = 1/10\ \{7.59 + (10-10)(2.70-1.03)\} = 0.759$

$m_2 = 1/10\ \{7.01 + (10-10)(2.80-1.10)\} = 0.701$

จะได้ว่า

$M = 1/18\ (7.59 + 7.01) = 0.811$

และ $\theta = (7.59 + 7.01)/20 = 0.73$

เนื่องจาก j=2 ดังนั้น

$W_2 = \{2(10)(1.10-1.03)/0.73\} = 1.918$

และ $V = 1.918 \times 0.73 = 1.40$

จะได้ค่าของตัวสถิติทดสอบ

$$U = 1.40/2(1)(0.811) = 0.863$$

จากตารางเอฟ จะได้ $F_{0.05}(2,36) = 3.28$ ซึ่งมากกว่า 0.863

∴ เราจะยอมรับสมมุติฐานที่ว่าค่าพารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 2 กลุ่มมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ หรืออาจกล่าวได้ว่าระยะเวลาที่สั้นที่สุดในการบรรเทาอาการปวดของยาแก้ปวดทั้ง 2 ชนิดอยู่ในช่วงเวลาเดียวกันเท่ากัน

ค) ตัวสถิติทดสอบ TIKU

ในปี ค.ศ.1985 Kambo และ Awad นักสถิติชาวจอร์แดน ได้พัฒนาตัวสถิติทดสอบนี้ขึ้น โดยมีแนวความคิดมาจากตัวสถิติทดสอบ TIKU ซึ่ง M.L TIKU ได้เสนอในปี ค.ศ. 1981 หลักการของตัวสถิติทดสอบนี้จะพิจารณาค่าอัตราส่วนของผลต่างของค่าพารามิเตอร์แสดงตำแหน่งกับตัวประมาณร่วมของค่าเฉลี่ย (θ) ซึ่งตัวสถิติทดสอบนี้นำค่าของผลต่างของค่าพารามิเตอร์แสดงตำแหน่งมาพิจารณาโดยตรง ถ้าค่าพารามิเตอร์แสดงตำแหน่งของกลุ่มประชากรมีค่าแตกต่างกันมากกว่าค่าเฉลี่ยของกลุ่มประชากรที่นำมาทดสอบก็จะนำไปสู่การปฏิเสธสมมุติฐานได้ง่าย ซึ่งส่งผลให้ตัวสถิติทดสอบดังกล่าวนี้มีค่าอำนาจการทดสอบสูง

ให้ $X_{i1} \le X_{i2} \le X_{i3} \le \ldots \le X_{i,ni-si}$ เป็นตัวสถิติอันดับ

เมื่อ s_i = จำนวนของข้อมูลที่มีค่าถูกตัดทิ้งในกลุ่มที่ i

กำหนดให้ θ^* เป็นตัวประมาณร่วมของ θ ซึ่งอยู่ในรูปของ

$$\theta^* = \frac{1}{d} \sum_{i=1}^{k} \left[\sum_{j=1}^{ni-si} X_{ij} + s_i X_{i,ni-si} - n_i(X_{i1}) \right]$$

เมื่อ $d = \sum_{i=1}^{k} (n_i - s_i - 1)$

กำหนด $Y_1 = \min_{1 \le i \le k} X_{i1}$ $\quad Y_k = \max_{1 \le i \le k} X_{i1}$

จะได้ว่า

$$T = \left| Y_k - Y_1 \right| / \theta^*$$

เราจะปฏิเสธสมมุติฐานเมื่อ $T > V_\alpha$

เมื่อ $V_\alpha = -\ln(1-(1-\alpha)^{1/k-1})/n$

ตัวอย่างที่ 3 จากข้อมูลในตัวอย่างที่ 1 , $n_1=n_2=n_3=10$, $s_1=2$, $s_2=3$, $s_3=1$, $r_1=8$, $r_2=7$, $r_3=9$ และ $d=21$ เราจะทดสอบสมมุติฐานโดย

ต้องการทดสอบ H_0 : พารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 3 กลุ่มมีค่าเท่ากัน

เทียบกับ H_1 : พารามิเตอร์แสดงตำแหน่งของข้อมูลอย่างน้อย 2 กลุ่มมีค่าไม่เท่ากัน

เราคำนวณค่า θ^* จาก

$$\theta^* = \frac{1}{d} \sum_{i=1}^{k} \left[\sum_{j=1}^{ni-si} X_{ij} + s_i X_{i,ni-si} - n_i(X_{i1}) \right]$$

$= 1/21\ [165.2 + 66.6 - 101] = 130.8/21$

$= 6.228$

เราจะเปรียบเทียบค่าพารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 3 กลุ่มโดย

$$Y_1 = \min [3.5,\ 3.0,\ 3.6] = 3.0$$

$$Y_k = \max [3.5,\ 3.0,\ 3.6] = 3.6$$

จะได้ว่า

$$T = \left| 3.6-3.0 \right| / 6.228$$

$$= 0.096$$

เราปฏิเสธสมมุติฐานเมื่อ $T > V_\alpha$ ณ ระดับนัยสำคัญ $\alpha=0.05$

เนื่องจาก $V_{0.05} = -\ln (1-(1-0.05)^{1/2})/10 = 0.36$ ซึ่งมากกว่า 0.096

∴ เรายอมรับสมมุติฐานที่ว่าค่าพารามิเตอร์แสดงตำแหน่งของข้อมูลทั้ง 3 กลุ่มไม่แตกต่างกันอย่างมีนัยสำคัญ หรืออาจจะกล่าวได้ว่าเครื่องจักรทั้ง 3 ชนิดจะหยุดทำงานเป็นครั้งแรกในช่วงเวลาเดียวกัน

2.4 ผลงานวิจัยที่เกี่ยวข้อง

ผลงานวิจัยของนักสถิติที่เกี่ยวข้องกับการทดสอบการเท่ากันของพารามิเตอร์แสดงตำแหน่งของการแจกแจงแบบเอกซ์โปเนนเชียล 2 พารามิเตอร์ ที่สำคัญพอสรุปได้ดังนี้

ในปี ค.ศ.1971 S.Kumar และ H.I.Patel เสนอตัวสถิติทดสอบเพื่อใช้ทดสอบการเท่ากันของพารามิเตอร์แสดงตำแหน่งของ 2 ประชากร โดยใช้ทดสอบในกรณีที่มีตัวอย่างขนาดเล็ก ซึ่งภายหลังเรียกว่าตัวสถิติทดสอบ KUMAR-PATEL (KP) ต่อมาในปี ค.ศ.1973 D.G.Weinman และคณะได้ศึกษาอำนาจการทดสอบของตัวสถิติทดสอบ KP และตัวสถิติทดสอบ LR โดยนำมาเปรียบเทียบกันในกรณีที่มีตัวอย่างขนาดเล็กพบว่าโดยส่วนใหญ่แล้วตัวสถิติทดสอบ LR จะมีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ KP ยกเว้นในกรณีที่ความแตกต่างระหว่างค่าพารามิเตอร์แสดงตำแหน่งมากกว่า 2 เท่าของค่าเฉลี่ยของประชากรที่นำมาทดสอบเท่านั้น ในปี ค.ศ. 1981 H.K.Hsieh ได้เสนอผลงานวิจัยในการเปรียบเทียบอำนาจการทดสอบของตัวสถิติทดสอบ LR,IP,และ PERNG(1979) ในกรณีที่มีประชากร 2 กลุ่มพบว่าเมื่อพารามิเตอร์แสดงสเกล (θ) มีขนาดเล็ก ตัวสถิติทดสอบ IP จะมีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ LR เล็กน้อย แต่เมื่อพารามิเตอร์แสดงสเกลมีค่าสูงขึ้นตัวสถิติทดสอบ LR จะมีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ

IP และ PERNG ในทุกกรณี หลังจากนั้นในปี ค.ศ.1989 B.N.Shetty และ P.C.Joshi เสนอผลงานวิจัยในการศึกษาอำนาจการทดสอบของตัวสถิติทดสอบ LR กับตัวสถิติทดสอบ TIKU โดยศึกษาเฉพาะในกรณีที่มีประชากร 2 กลุ่มกับตัวอย่างที่มีขนาดเล็ก ($n\leq10$) ทั้งในกรณีที่มีข้อมูลสมบูรณ์และกรณีที่มีข้อมูลมีค่าถูกตัดทิ้งทั้งทางซ้ายและทางขวา ผลการวิจัยพอที่จะสรุปได้ดังนี้

1. กรณีที่มีข้อมูลสมบูรณ์ตัวสถิติทดสอบ TIKU จะให้อำนาจการทดสอบใกล้เคียงกับตัวสถิติทดสอบ LR

2. กรณีที่มีข้อมูลถูกตัดทิ้งทั้งทางซ้ายและทางขวาตัวสถิติทดสอบ TIKU จะมีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบ LR