บทที่ 2

ระเบียบวิธีที่ใช้ในการวิจัย

2.1 สถิติที่ใช้พารามิเตอร์ สถิติที่ไม่ใช้พารามิเตอร์ และแรงค์ทรานสฟอร์เมชั่น

การอ้างอิงทางสถิติส่วนมากหรือเกือบทั้งหมดตั้งอยู่บนข้อสมมติที่ว่าตัวอย่างที่สุ่มมานั้นมาจากประชากรที่ทราบว่า มีลักษณะการแจกแจงอย่างใดอย่างหนึ่ง ซึ่งฟังก์ชั่นของการแจกแจงขึ้นอยู่กับค่าพารามิเตอร์ค่าเดียว หรือหลายค่าของประชากรนั้น ๆ จึงเรียกวิธีการทางสถิติที่ใช้ประมาณ และทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์นั้น ๆ ว่าเป็นสถิติที่ใช้พารามิเตอร์ ซึ่งเป็นการประมาณ และทดสอบค่าพารามิเตอร์จากค่าสถิติ (Statistic) ที่คำนวณจากตัวอย่างที่เลือกมาจากประชากรนั้น ๆ โดยวิธีสุ่ม การใช้วิธีการทดสอบ แบบใช้พารามิเตอร์ในการประมาณ และทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์ของประชากร จะเป็นไปอย่างมีประสิทธิภาพ จะต้องขึ้นอยู่กับเงื่อนไขสองประการ ประการแรกลักษณะการแจกแจงของประชากร จะต้องเป็นแบบปกติ (Normality of Population) และประการที่สอง ค่าความแปรปรวนจะต้องคงที่ (Stability of variance) ซึ่งในการทดสอบโดยใช้ตัวสถิติ t และตัวสถิติเอฟ F จะต้องตั้งข้อสมมติว่าลักษณะของประชากรนั้น ๆ เป็นไปตามเงื่อนไขดังกล่าวทั้งสองประการ

ภายใต้สถานการณ์บางอย่าง หากลักษณะการแจกแจงของประชากรไม่เป็นแบบปกติอย่างเด่นชัด หรือผู้วิจัยไม่สามารถที่จะระบุลงไปอย่างแจ้งชัดว่าลักษณะการแจกแจงของประชากรเป็นแบบใด หรือไม่สามารถหาเหตุผลมาสนับสนุนการตั้งข้อสมมติได้ว่า ลักษณะการแจกแจงเป็นไปตามเงื่อนไขทั้งสองประการดังกล่าวมาแล้วข้างต้นในกรณีเช่นนี้ อาจใช้วิธีการทางสถิติที่เรียกว่า วิธีทดสอบแบบไม่ใช้พารามิเตอร์ หรือเรียกอีกอย่างหนึ่งว่า Distribution-Free Test ซึ่งเป็นวิธีการทดสอบ ซึ่งไม่จำเป็นต้องระบุ หรือกำหนดลักษณะการแจกแจงของข้อมูลในประชากรแต่อย่างใด เพียงแต่กำหนดว่าประชากรนั้นจะต้องมีฟังก์ชั่นของการแจกแจงเป็นแบบต่อเนื่อง (Continuous Probability Distribution Function) เท่านั้น แล้วใช้วิธีการทดสอบที่ได้มีผู้คิดค้นขึ้นมา โดยไม่ต้องกำหนดข้อสมมติเกี่ยวกับการแจกแจงของประชากรไว้อย่างเคร่งครัด

วิธีการทดสอบแบบไม่ใช้พารามิเตอร์ โดยปกติแล้ว จะต้องแปลงข้อมูล หรือค่าสังเกต เป็นค่าอันดับของค่าสังเกตนั้น ๆ ก่อนแล้วจึงนำค่าอันดับไปทดสอบ W.J. Conover (1980) ได้เสนอวิธีการที่นำค่าอันดับไปใช้ทดสอบกับวิธีการทดสอบแบบใช้พารามิเตอร์ นั่นคือ ไม่ว่าประชากรจะมีลักษณะการแจกแจงเป็นแบบใดหรือจะใช้การทดสอบแบบใดก็ตาม ผู้วิจัยสามารถแปลงข้อมูลให้เป็นค่าอันดับแล้วนำค่าอันดับไปวิเคราะห์ โดยใช้การทดสอบแบบใช้พารามิเตอร์ได้เสมอ วิธีการดังกล่าวเรียกว่า แรงค์ทรานส์ฟอร์เมชั่น ซึ่ง Conover และ Ronald Iman (1981) กล่าวว่า "แรงค์ทรานส์ฟอร์เมชั่น เป็นเสมือนตัวกลางหรือสะพานเชื่อมระหว่างวิธีการทดสอบแบบใช้พารามิเตอร์ และวิธีการทดสอบแบบไม่ใช้พารามิเตอร์ "

2.2 การทดสอบแบบแมน-วิทนีย์(Mann-Whitney Test)

การทดสอบแบบแมน-วิทนีย์ เป็นวิธีการที่ใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร 2 กลุ่มที่เป็นอิสระกัน ซึ่งแมน และวิทนีย์ เป็นผู้เสนอในปี พ.ศ. 2490 ตัวสถิติทดสอบแบบ แมน-วิทนีย์เป็นสถิติที่ไม่ใช้พารามิเตอร์

ข้อตกลงเบื้องต้นของตัวสถิติทดสอบ แมน-วิทนีย์ สำหรับการทดสอบสมมติฐาน การเท่ากันของค่าเฉลี่ยของประชากร 2 กลุ่ม (Gibbon 1971 : 140-149) มีดังนี้

1) ข้อมูลประกอบด้วยตัวอย่างสุ่ม X_1 , X_2, ..., X_m จากประชากรกลุ่มที่หนึ่ง และ ตัวอย่างสุ่ม Y_1 , Y_2, ... , Y_n จากประชากรกลุ่มที่สอง

2) ตัวอย่างสุ่มทั้งสองชุดเป็นอิสระกัน

3) ประชากรที่ศึกษามีการแจกแจงแบบต่อเนื่อง

4) สเกลที่ใช้วัดข้อมูลอย่างน้อยต้องเป็นสเกลแบบจัดเรียงอันดับ (Ordinal Scale)

5) ประชากรทั้งสองที่ศึกษามีรูปแบบการแจกแจงแบบเดียวกัน และความแปรปรวนเท่ากัน (Identical form)

ตัวสถิติทดสอบแบบแมน-วิทนีย์ คือ U หรือ U' ซึ่งคำนวณได้ดังนี้

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$$

โดยที่ $$D_{ij} = \begin{cases} 1 \text{ ถ้า } y_j < x_i & ; i = 1, 2, \ldots, m \\ 0 \text{ ถ้า } y_j > x_i & ; j = 1, 2, \ldots, n \end{cases}$$

หรือ

$$U' = \sum_{i=1}^{m} \sum_{j=1}^{n} (1-D_{ij})$$

โดยที่ U' = จำนวนครั้งที่ $x_i < y_j$

ถ้า m หรือ n มีค่ามาก การคำนวณค่า U และ U' ด้วยวิธีดังกล่าวข้างต้นอาจไม่ล่ะดวก ซึ่งอาจคำนวณค่า U และ U' ได้อีกวิธีหนึ่งดังนี้

$$U = mn + \frac{m(m+1)}{2} - R_x$$

$$U' = mn + \frac{n(n+1)}{2} - R_y$$

โดยที่ R_x = ผลรวมของค่าอันดับของ x จากการจัดอันดับร่วมกันของ x และ y

R_y = ผลรวมของค่าอันดับของ y จากการจัดอันดับร่วมกันของ x และ y

กรณีที่ต้องการทดสอบ 2 ด้าน (Two-sided test) จะปฏิเสธสมมติฐานว่าง ถ้าค่า U หรือ U' ที่คำนวณได้มีค่าน้อยกว่า $c_{\alpha/2}$ สำหรับการทดสอบแบบด้านเดียว (One-side test) จะปฏิเสธสมมติฐานว่า เมื่อ $U < c_\alpha$ หรือ $U' < c_\alpha$ โดยที่ c_α คือค่าวิกฤต (Critical value) ณ ระดับนัยสำคัญ α ซึ่งสามารถหาได้จากตารางการทดสอบแบบแมน-วิทนีย์ (ดูตารางที่ 1 ในภาคผนวก)

ในกรณีที่ตัวอย่างสุ่มมีขนาดใหญ่ ค่าสถิติ U จะมีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติ โดยที่ค่าคาดหวังและค่าความแปรปรวนของ U มีลักษณะดังนี้

$$E(U) = \frac{mn}{2}$$

$$Var(U) = \frac{mn(m+n+1)}{12}$$

ดังนั้นตัวสถิติที่ใช้ทดสอบเมื่อตัวอย่างที่สุ่มมามีขนาด ใหญ่คือ

$$Z = \frac{U - (mn/2)}{\left[\frac{mn(m+n+1)}{12}\right]^{\frac{1}{2}}}$$

ซึ่งการแจกแจงของ Z สามารถประมาณด้วยการแจกแจงแบบปกติมาตรฐาน (Standard Normal) ซึ่งค่าวิกฤตหาได้จากตารางการแจกแจงแบบปกติมาตรฐาน การประมาณนี้จะใช้ได้ดีแม้ว่าขนาดตัวอย่างทั้ง 2 ชุด มีขนาดเพียง 6 เท่านั้น (Gibbon 1971 : 145)

2.3 การทดสอบแบบครัสคัล-แวลิส (The Kruskal-Wallis Test)

เป็นสถิติที่ไม่ใช้พารามิเตอร์วิธีหนึ่ง ซึ่งใช้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรตั้งแต่ สามกลุ่มขึ้นไป ซึ่งครัสคัล-แวลิสได้พัฒนาขึ้นในปี พ.ศ. 2495 มีหลักการที่ใช้ในการทดสอบเป็นเช่นเดียวกับการทดสอบโดยใช้ตัวสถิติของ วิลค็อกซอน (Wilcoxon's Rank-sum test) ต่างกันที่ว่าตัวสถิติของ ครัสคัล-แวลิส ใช้ทดสอบกับประชากร ตั้งแต่ 3 กลุ่มขึ้นไป และมีลักษณะการวิเคราะห์ ทำนองเดียวกับ การวิเคราะห์ความแปรปรวนแบบทางเดียว

ตารางที่ 2.3.1 แสดงลักษณะของข้อมูลที่ใช้ในการทดสอบแบบครัสคัล-แวลิส

ทรีทเมนต์			
1	2	. . .	k
Y_{11}	Y_{21}		Y_{k1}
Y_{12}	Y_{22}		Y_{k2}
.	.		.
.	.		.
.	.		.
Y_{1n_1}	Y_{2n_2}		Y_{kn_k}

ให้ R_i แทนผลรวมของแรงค์ในกลุ่มที่ i

ดังนั้น $$R_i = \sum_{j=1}^{n_i} R(Y_{ij}) \quad ; \quad i = 1, 2, \ldots, k$$
$$j = 1, 2, \ldots, n_i$$

ให้ N แทนจำนวนค่าสังเกตทั้งหมด

ดังนั้น $$N = \sum_{i=1}^{k} n_i$$

$$E(R_i) = \left[\frac{n_i}{N}\right] \frac{N(N+1)}{2}$$

$$= \frac{n_i(N+1)}{2}$$

ตัวสถิติที่ใช้ทดสอบคือ $$S = \sum_{i=1}^{k} \left[R_i - E(R_i)\right]^2$$

$$= \sum_{i=1}^{k} \left[R_i - \frac{n_i}{2}(N+1)\right]^2$$

จะปฏิเสธสมมติฐานว่าง (Null hypothesis) เมื่อ S มีค่ามากกว่า S_α โดยที่ S_α เป็นค่าวิกฤตซึ่งหาได้จากตารางการทดสอบของครัสคัล-แวลิส (ดูตารางที่ 2 ในภาคผนวก)

$$R(Y_{ij}) = 1, 2 \ldots, N$$

$$E\left[R(Y_{ij})\right] = \frac{1}{N}\sum_{i=1}^{N} i = \frac{1}{2}(N+1)$$

$$V\left[R(Y_{ij})\right] = \frac{1}{N}\sum_{i=1}^{N}\left(i - \frac{N+1}{2}\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} i^2 - 2\,\frac{(N+1)}{2}\sum_{i=1}^{N} + \frac{N(N+1)^2}{4}$$

$$= \frac{1}{N}\;\frac{N}{6}(N+1)(2N+1) - (N+1)\frac{N}{2}(N+1) + N\,\frac{(N+1)^2}{4}$$

$$= \frac{1}{12}\;N^2 - 1$$

$$R_i = \sum_{j=1}^{n_i} R(Y_{ij})$$

$$\bar{R}_i = \frac{R_i}{n_i}$$

$$E(\bar{R}_i) = E\left(\frac{R_i}{n_i}\right)$$

$$= \frac{1}{n_i}\left[\frac{n_i(N+1)}{2}\right]$$

$$= \frac{1}{2}(N+1)$$

$$V(\bar{R}_i) = \frac{1}{12n_i}(N-n_i)(N+1)$$

ถ้า N มีขนาดใหญ่ เมื่อใช้ทฤษฎีลิมิตสู่ส่วนกลาง (Central limit theorem) จะได้

$$Z = \frac{\bar{R}_i - E(\bar{R}_i)}{\sqrt{V(\bar{R}_i)}} \sim N(0\ ,\ 1)$$

ซึ่ง Z_i^2 จะมีการแจกแจงซึ่งสามารถประมาณได้ด้วยการแจกแจงแบบไคสแควร์โดยมีองค่าแห่งความเป็นอิสระเท่ากับ 1

พิจารณา $H = \sum_{i=1}^{k} (1-f_i) Z_i^2 \quad ; \quad f_i = \frac{n_i}{N}$

$$= 12n_i \sum_{j=1}^{k}\left(\frac{N-n_i}{N}\right) \frac{\left(\bar{R}_i - \frac{N+1}{2}\right)^2}{(N-n_i)(N+1)}$$

$$= \frac{12}{N(N+1)}\left[\sum_{i=1}^{k} n_i \bar{R}_i^2 - (N+1)\sum_{i=1}^{k} n_i \bar{R}_i + \frac{(N+1)}{2}^2 \sum_{i=1}^{k} n_i\right]$$

$$= \frac{12}{N(N+1)}\left[\sum_{i=1}^{k}\frac{R_i^2}{n_i} - (N+1)\sum_{i=1}^{k} n_i \frac{R_i}{n_i} + \frac{N(N+1)^2}{4}\right]$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

H จะมีการแจกแจงซึ่งสามารถประมาณได้ด้วยการแจกแจงแบบไคสแควร์ โดยมีองศาแห่งความเป็นอิสระเท่ากับ k-1 เมื่อ N มีขนาดใหญ่ จะปฏิเสธสมมติฐานว่างเมื่อ H มีค่ามากกว่า $\chi^2_{(K-1),\alpha}$

ข้อตกลงเบื้องต้นสำหรับการทดสอบแบบครัสคัล-แวลลิส (Conover ; 1980 : 230) มีดังต่อไปนี้

1) กลุ่มตัวอย่างทั้ง k กลุ่ม เป็นตัวอย่างที่สุ่มมาจากแต่ละประชากร

2) ตัวอย่างสุ่มภายในกลุ่ม และระหว่างกลุ่มเป็นอิสระกัน

3) สเกลที่ใช้วัดข้อมูลอย่างน้อยต้องเป็นสเกลแบบจัดเรียงอันดับ (Ordinal Scale)

4) ประชากรที่นำมาทดสอบจะต้องมีการแจกแจงแบบเดียวกันและมีความแปรปรวนเท่ากัน

การคำนวณค่าสถิติ

1. กรณีที่ไม่มีอันดับเท่ากัน (Untied rank)

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

2. กรณีที่มีอันดับเท่ากัน (Ties rank) ให้ใช้ค่าเฉลี่ยของอันดับเป็นตัวแทนอันดับที่ซ้ำ

ให้ t_s = ความถี่ของการซ้ำในอันดับที่ s

ดังนั้นตัวสถิติที่ใช้ทดสอบคือ

$$H = \frac{\frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)}{1 - \frac{1}{N^3-N} \sum_{S=1}^{d} (t_s^3 - t_s)}$$

โดยที่ N = จำนวนค่าสังเกตทั้งหมด

n_i = จำนวนค่าสังเกตในแต่ละกลุ่มตัวอย่าง

R_i = ผลรวมของอันดับในแต่ละกลุ่มตัวอย่าง

d = จำนวนอันดับที่มีการซ้ำเกิดขึ้น

เมื่อจำนวนกลุ่มตัวอย่างเท่ากับ 3 (k=3) และในแต่ละกลุ่มมีจำนวนค่าสังเกตน้อยกว่าหรือเท่ากับ 5 จะปฏิเสธสมมติฐานว่างเมื่อ S มีค่ามากกว่า S_α เมื่อกลุ่มตัวอย่างเท่ากับ 3 จำนวนค่าสังเกตในแต่ละกลุ่มมีค่ามากกว่า 5 หรือขนาดตัวอย่างมากกว่า 15 จะปฏิเสธสมมติฐานว่าง เมื่อ H มีค่ามากกว่า $\chi^2_{(k-1)}$

2.4 การทดสอบแบบที

ในการทดสอบสมมติฐานเกี่ยวกับผลต่างระหว่างค่าเฉลี่ยของสองประชากร ถ้าให้ μ_1 และ μ_2 แทนค่าเฉลี่ยของประชากรที่หนึ่ง และประชากรที่สองที่นำทดสอบสมมติฐาน ตัวสถิติที่ใช้ทดสอบสมมติฐาน

$$H_o \quad : \quad \mu_1 = \mu_2$$

$$H_a \quad : \quad \mu_1 \neq \mu_2$$

ให้ $\bar{y}_1$ = ค่าเฉลี่ยของตัวอย่างขนาด n_1 ที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ย μ_1 และความแปรปรวน σ_1^2

$\bar{y}_2$ = ค่าเฉลี่ยของตัวอย่างขนาด n_2 ที่สุ่มมาจากประชากรที่มีการแจกแจงแบบปกติมีค่าเฉลี่ย μ_2 และความแปรปรวน σ_2^2

$\bar{y}_1 - \bar{y}_2$ จะมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเป็น $\mu_1-\mu_2$ และความแปรปรวนเป็น $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนเป็นหนึ่ง

เนื่องจากโดยทั่ว ๆ ไป ผู้ทดลองมักจะไม่ทราบค่าความแปรปรวนของประชากรทั้งสอง จึงต้องประมาณค่าความแปรปรวนของประชากรทั้งสองด้วย ค่าความแปรปรวนจากตัวอย่าง (S_1^2 และ S_2^2) แต่การประมาณค่าความแปรปรวนของประชากรด้วยความแปรปรวนจากตัวอย่างมีผลทำให้การแจกแจงของตัวสถิติ

$$\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^{\frac{1}{2}}}$$

มีการแจกแจงประมาณด้วยการแจกแจงแบบที (t) ที่มีองศาแห่งความอิสระ V โดยที่

$$\gamma = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2 \Bigg/ \left[\left[\frac{S_1^2}{n_1}\right]^2 \Big/ (n_1-1) + \left[\frac{S_2^2}{n_2}\right]^2 \Big/ (n_2-1)\right]$$

ในกรณีที่ความแปรปรวนของประชากรทั้งสองที่นำมาทดลองเท่ากันคือ σ^2 ตัวสถิติเพื่อการทดลองจะมีค่าเป็น

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

$$= \frac{(y_1 - y_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

เมื่อประมาณค่าความแปรปรวนของประชากร (σ^2) ด้วยค่าความแปรปรวนรวม (pooled variance) จากตัวอย่างที่สุ่มมาจากประชากรทั้งสอง ซึ่งเขียนแทนด้วยสัญลักษณ์ S_p^2

เมื่อ s_p^2 = ค่าเฉลี่ยถ่วงน้ำหนักของค่าความแปรปรวน จากตัวอย่างที่สุ่มมาจากประชากรทั้งสอง

$$= \frac{(n_1-1)\ s_1^2 + (n_2-1)\ s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

การแจกแจงของ $\dfrac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\ (\frac{1}{n_1} + \frac{1}{n_2})}}$ จะเป็นแบบที (t-distribution)

โดยมีองศาแห่งความเป็นอิสระเท่ากับ $n_1 + n_2 - 2$ นั่นคือ

$$T = \frac{(\bar{y}_1-\bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\ (\frac{1}{n_1} + \frac{1}{n_2})}} \sim t_{(n_1+n_2-2)}$$

ถ้าจำนวนตัวอย่างที่สุ่มมาจากแต่ละประชากรมีขนาดใหญ่ จะใช้การแจกแจงแบบปกติมาตรฐาน (Z) ประมาณการแจกแจงแบบ t

2.5 การทดสอบแบบเอฟ (F-test) ในการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว (One-Way Classification)

การวิเคราะห์ความแปรปรวน (Analysis of variance) เป็นวิธีการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร ตั้งแต่สามประชากรขึ้นไป หลักการสำคัญที่ใช้ในการทดสอบสมมติฐานด้วยวิธีนี้คือ การแยกความแปรปรวนหรือความแตกต่างของข้อมูลที่เกิดขึ้นทั้งหมดออกจากสาเหตุต่าง ๆ แล้วพิจารณาสัดส่วนของความแปรปรวนหรือความแตกต่างระหว่างประชากรและความแปรปรวน หรือความแตกต่างภายในประชากรเดียวกันว่ามีค่ามากน้อยเพียงใด ถ้าอัตราส่วนดังกล่าวมีมาก แสดงว่าความแปรปรวนหรือความแตกต่างระหว่างประชากรมีมาก เมื่อเทียบกับความแปรปรวนหรือความแตกต่างภายในประชากรเดียวกัน สามารถสรุปได้ว่าจำนวนประชากรทั้งหมดที่นำมาทดสอบความแตกต่างระหว่างค่าเฉลี่ย มีค่าเฉลี่ยของประชากรอย่างน้อยหนึ่งประชากรที่แตกต่างจากประชากรอื่น ๆ ที่นำมาทดสอบ ซึ่งวิธีวิเคราะห์ความแปรปรวนจะแตกต่างกันไปตาม

ลักษณะแผนการทดลอง แผนการทดลองที่ใช้ในการวิจัยนี้ใช้แผนการทดลองแบบสุ่มโดยสมบูรณ์ (Completely Randomized Design) หรือ ข้อมูลแบบแจกแจงทางเดียว ซึ่งเป็นแผนการทดลองแบบง่ายที่สุด การวิเคราะห์ความแปรปรวนจะแยกสาเหตุของความแปรปรวนของข้อมูลทั้งหมดว่า เนื่องมาจากอิทธิพลของทรีทเมนต์ (Treatment Effect) แต่เพียงอย่างเดียว ดังนั้นเพื่อให้แผนการทดลองนี้มีประสิทธิภาพมากที่สุด หน่วยทดลอง (Experimental Unit) ที่นำมาใช้ ควรมีลักษณะเหมือนกัน หรือคล้ายคลึงกันมากที่สุด (Homogeneous) หรือให้มีความแปรปรวนระหว่างหน่วยทดลองน้อยที่สุด

2.5.1 <u>การวิเคราะห์เมื่อมีจำนวนซ้ำเท่ากัน</u>

ข้อมูลที่ใช้วิเคราะห์ สามารถจัดให้อยู่ในรูปตารางที่ 2.5.1

ตารางที่ 2.5.1 แสดงลักษณะข้อมูลจากแผนการทดลองแบบสุ่มตลอด

	ทรีทเมนต์					
	1	2	3	. . .	t	
	y_{11}	y_{21}	y_{31}	. . .	y_{t1}	
	y_{12}	y_{22}	y_{32}	. . .	y_{t2}	
	.	.	.		.	
	.	.	.		.	
	.	.	.		.	
	y_{1r}	y_{2r}	y_{3r}	. . .	y_{tr}	
ผลรวมของทรีทเมนต์	$y_{1.}$	$y_{2.}$	$y_{3.}$	. . .	$y_{t.}$	$y_{..}$
ค่าเฉลี่ยของทรีทเมนต์	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$		$\bar{y}_{t.}$	$\bar{y}_{..}$

2.5.2 ตัวแบบเส้นตรงเชิงบวก (Linear Additive Model)

ค่า y ในตาราง 2.5.1 สามารถแสดงในรูปของผลบวกขององค์ประกอบต่าง ๆ ซึ่งกรณีนี้ได้แก่ ค่าเฉลี่ยโดยทั่วไป (General mean) คือ μ ผลกระทบจากวิธีทดลองที่แตกต่างกัน (Treatment effect) คือ τ_i ความคลาดเคลื่อนเชิงสุ่ม (Random error) ที่เกิดจากหน่วยทดลองที่ j ของสิ่งทดลองที่ i คือ ε_{ij} เมื่อรวมเขียนเป็นตัวแบบเส้นตรงเชิงบวก (Linear additive model)

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad ; \quad i = 1,2,\ldots,t. \quad j = 1,2,\ldots,r.$$

หรืออาจเขียนได้ในรูป

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

โดยที่ μ_i คือค่าเฉลี่ยจริงของสิ่งทดลองที่ i ซึ่งมีค่าเท่ากับ $\mu + \tau_i$ นั่นเอง

2.5.3 ข้อสมมติในการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว

1. ε_{ij} มีการแจกแจงอย่างอิสระแบบปกติ โดยมีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนร่วมกันคือ σ^2

2. τ_i มีข้อสมมติแยกเป็น 2 กรณีคือ

2.1 τ_i เป็นผลกระทบชนิดคงที่ (Fixed effect) หมายถึงสิ่งทดลองที่ใช้ในการทดลองถูกกำหนดให้คงที่ ทำให้ τ_i คงที่และ $\sum_{i=1}^{t} \tau_i = 0$

2.2 τ_i เป็นผลกระทบเชิงสุ่ม (Random effect) ในกรณีที่สิ่งทดลองที่ใช้ในการทดลองถูกเลือกอย่างสุ่ม จากประชากรของสิ่งทดลองทั้งหมด ค่า τ_i เป็นตัวอย่างสุ่มที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนเป็น σ_τ^2

สมมติฐานในการทดสอบคือ ไม่มีความแตกต่างระหว่างค่าเฉลี่ยของสิ่งทดลอง ซึ่งเขียนได้ดังนี้

$$H_0 : \mu_{T_1} = \mu_{T_2} = \ldots = \mu_{T_t}$$

ซึ่งถ้าสมมติฐานนี้เป็นจริง ย่อมหมายถึงว่า $\mu_i = \mu + \tau_i =$ ค่าคงที่ดังนั้นจึงอาจเขียนสมมติฐานได้อีกรูปแบบหนึ่งคือ

$$H_o : \tau_1 = \tau_2 = \ldots = \tau_t = o$$

$$H_a : \tau_i \text{ อย่างน้อยหนึ่งค่าที่ไม่เท่ากับศูนย์}$$

ถ้าสมมติฐานเป็นจริง ตัวแบบที่เหมาะสมสำหรับข้อมูลจะอยู่ในรูป $y_{ij} = \mu + \varepsilon_{ij}$

2.5.4 วิธีการวิเคราะห์ความแปรปรวน

ให้ y_{ij} = ค่าสังเกตจากหน่วยทดลองที่ได้รับทรีทเมนต์ i ซ้ำที่ j

y_i = ผลรวมของค่าสังเกตจากหน่วยทดลองที่ได้รับทรีทเมนต์ที่ i

$$= \sum_{j=1}^{r} y_{ij}$$

$\bar{y}_{i.}$ = ค่าเฉลี่ยของทรีทเมนต์ที่ i $= \frac{y_i}{r}$

$y_{..}$ = ผลรวมของค่าสังเกตทั้งหมด

$\bar{y}_{..}$ = ค่าเฉลี่ยของค่าสังเกตทั้งหมด

$$\text{Total SS} = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - y_{..})^2 = \sum_{i=1}\sum_{j=1} y_{ij}^2 - \frac{y_{..}^2}{t_r}$$

$$\text{Treatment SS} = \text{SST} = r\sum_{i=1}^{t}(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{1}^{t}\frac{y_{i.}^2}{r} - \frac{y_{..}^2}{tr}$$

$$\text{Error SS} = \text{SSE} = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \bar{y}_{i.})^2 = \text{Total SS} - \text{SST}$$

$\frac{y_{..}^2}{tr}$ เรียกว่า correction term (c)

ตารางที่ 2.5.3 การวิเคราะห์ความแปรปรวนจากแผนการทดลองแบบสุ่มตลอด

สาเหตุของความแปรปรวน	d.f.	S.S.	M.S.	F.
ระหว่างประชากร	t-1	$\sum_{i=1}^{t} \frac{y_{i.}^2}{r} - c$	$\frac{SST}{t-1} = MST$	$\frac{MST}{MSE}$
ภายในประชากร	t (r-1)	by subtraction	$\frac{SSE}{t(r-1)} = MSE$	
รวม	tr-1	$\sum_i \sum_j y_{ij}^2 - c$		

ค่าประมาณของ population variance (σ^2) คือ $S^2 = MSE$

Standard error ของ treatment mean ($S_{\bar{y}}$) $= \sqrt{S^2/r}$

Standard error ของผลต่างของ treatment ($S_{\bar{d}}$) $= \sqrt{2S^2/r}$

Coefficient of variation $= \frac{S}{\bar{y}} \times 100\%$

ปฏิเสธสมมติฐานว่าง เมื่อ F มีค่ามากกว่า F_α ที่องศาแห่งความเป็นอิสระ t-1 และ t(r-1)

2.6 แรงค์ทรานส์ฟอร์เมชั่น (Rank Transformation)

เป็นวิธีการสถิติวิธีหนึ่งซึ่ง Conover และ Iman (1981) กล่าวว่า เป็นเสมือนสะพานเชื่อมระหว่างสถิติที่ใช้พารามิเตอร์กับสถิติที่ไม่ใช้พารามิเตอร์ นั่นคือ ไม่ว่าจะเป็นการทดสอบใด ๆ หลังจากแปลงข้อมูลเป็นอันดับแล้ว สามารถใช้วิธีทดสอบแบบพารามิเตอร์กับอันดับของข้อมูลได้ ซึ่งในการนำแรงค์ทรานส์ฟอร์เมชั่นไปใช้และเป็นที่ยอมรับอย่างกว้างขวางคือ นำไปใช้ในการหาสัมประสิทธิ์สหสัมพันธ์อันดับ (Correlation Coefficient Based on Rank)

จากสัมประสิทธิ์สหสัมพันธ์ของ Pearson (Pearson's Product Moment Correlation Coefficient)

$$r = \frac{\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})}{\left[\sum_{i=1}^{n} (x_i-\bar{x})^2 \sum_{i=1}^{n} (y_i-\bar{y})^2\right]^{\frac{1}{2}}}$$

Spearman ได้แปลงค่าสังเกตเป็นค่าอันดับ แล้วหาสัมประสิทธิ์สหสัมพันธ์อันดับจากวิธีของ Pearson จะได้ตัวสถิติทดสอบคือ

$$\rho = \frac{\sum_{i=1}^{n} (R_i-\bar{R})\ (S_i-\bar{S})}{\left[\sum_{i=1}^{n} (R_i-\bar{R})^2 \sum_{i=1}^{n} (S_i-\bar{S})^2\right]^{\frac{1}{2}}}$$

โดย R_i = ค่าอันดับของกลุ่มตัวอย่าง x

S_i = ค่าอันดับของกลุ่มตัวอย่าง y

$\bar{R}$ = ค่าเฉลี่ยของอันดับของกลุ่มตัวอย่าง x

$$= \sum_{i=1}^{n} \frac{R_i}{n} = \frac{n+1}{2}$$

$\bar{S}$ = ค่าเฉลี่ยของอันดับของกลุ่มตัวอย่าง y

$$= \sum_{i=1}^{n} \frac{S_i}{n} = \frac{n+1}{2}$$

ในทำนองเดียวกัน แรงค์ทรานส์ฟอร์เมชั่น สามารถนำไปใช้ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร ซึ่งแบ่งได้เป็น 2 กรณีคือ

2.6.1 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร 2 ชุด

ในการทดสอบความแตกต่าง ระหว่างค่าเฉลี่ยของประชากร 2 ชุด เมื่อตัวอย่างสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ ผู้วิจัยจะใช้การทดสอบแบบที หรือใช้การทดสอบแบบแมน-วิทนีย์ เมื่อประชากรแจกแจงแบบอื่น ในการใช้แรงค์ทรานส์ฟอร์เมชั่นประยุกต์กับการทดสอบนี้กระทำได้โดยการใช้แบบทดสอบทีกับอันดับ จะได้ตัวสถิติ

$$t_R = \frac{(\bar{R}_1 - \bar{R}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

โดยที่ $\bar{R}_1$ = ค่าเฉลี่ยของอันดับของกลุ่มตัวอย่างที่ 1.

$\bar{R}_2$ = ค่าเฉลี่ยของอันดับของกลุ่มตัวอย่างที่ 2

n_1 = จำนวนตัวอย่างจากกลุ่มที่ 1

n_2 = จำนวนตัวอย่างจากกลุ่มที่ 2

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

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (R_{1i} - \bar{R}_1)^2}{n_1-1}$$

$$S_2^2 = \frac{\sum_{i=1}^{n_2} (R_{2i} - \bar{R}_2)^2}{n_2-2}$$

จะปฏิเสธสมมติฐานว่าง เมื่อ t_R ที่ได้จากการคำนวณน้อยกว่า $-t_{\alpha/2}$ หรือมากกว่า $t_{\alpha/2}$ ที่องศาแห่งความเป็นอิสระ n_1+n_2-2

2.6.2 การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรมากกว่า 2 ชุด

ในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากรมากกว่า 2 ชุด เมื่อลักษณะของข้อมูลเป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวน ใช้การทดสอบแบบเอฟ และใช้การทดสอบแบบครัสคัล-แวลลิส เมื่อลักษณะของข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น การใช้

แรงค์ทรานส์ฟอร์เมชั่นมาประยุกต์กับปัญหานี้ ทำได้โดยการจัดอันดับของข้อมูลร่วมกันแล้วใช้ การวิเคราะห์ความแปรปรวนกับอันดับของข้อมูลนั้น

ตารางที่ 2.6.2 ลักษณะข้อมูลในรูปอันดับที่ใช้ในการวิเคราะห์ความแปรปรวน

	ทรีทเมนต์				
	1	2	....	t	
	R_{11}	R_{21}		R_{t1}	
	R_{12}	R_{22}		R_{t2}	
	.	.		.	
	.	.		.	
	.	.		.	
	R_{1r}	R_{2r}		R_{tr}	
รวม	$R_{1.}$	$R_{2.}$		$R_{t.}$	$\frac{N(N+1)}{2}$

ให้ R = ค่าอันดับของหน่วยทดลองที่ได้รับทรีทเมนต์ i ซ้ำที่ j

R_i = ผลรวมของอันดับของหน่วยทดลองที่ได้รับจากทรีทเมนต์ i

$\bar{R}_i$ = ค่าเฉลี่ยของอันดับที่ได้รับจากทรีทเมนต์ i

$$\text{Total SS} = \sum_{i=1}^{t}\sum_{j=1}^{r}\left(R_{ij} - \frac{N+1}{2}\right)^2 = \sum_{i=1}^{t}\sum_{j=1}^{r} R_{ij}^2 - \frac{N}{4}(N+1)^2$$

$$\text{Treatment SS} = r\sum_{i=1}^{t}\left(R_i - \frac{N+1}{2}\right)^2 = \sum_{i=1}^{t}\frac{R_i^2}{r} - \frac{N}{4}(N+1)^2$$

$$\text{Error SS} = \sum_{i=1}^{t}\sum_{j=1}^{r}(R_{ij} - R.)^2 = \text{Total SS} - \text{Treatment SS.}$$

ตารางที่ 2.6.3 แสดงการวิเคราะห์ความแปรปรวนโดยใช้แรงค์ทรานส์ฟอร์เมชั่น

สาเหตุของความแปรปรวน	d.f.	S.S	M.S.	F_R
ระหว่างประชากร	t-1	$\sum_{i=1}^{t} \frac{R_i^2}{r} - \frac{N}{4}(N+1)^2$	SST/t-1=MST	$\frac{MST}{MSE}$
ภายในประชากร	t(r-1)	by subtraction	SSE/t(r-1)=MSE	
รวม	tr-1	$\sum_{i=1}^{t} \sum_{j=1}^{r} R_{ij}^2 - \frac{N}{4}(N+1)^2$		

จะปฏิเสธ สมมติฐานว่างเมื่อ F_R ที่คำนวณได้มีค่ามากกว่า F_α ที่องศาแห่งความเป็นอิสระ t-1 และ t(r-1)

2.7 การทดสอบสัดส่วนของผลการทดสอบที่ตรงกัน

หลังจากทดสอบความแตกต่างระหว่างค่าเฉลี่ยของประชากร โดยใช้แรงค์ทรานส์ฟอร์เมชั่นกับวิธีการทดสอบแบบต่าง ๆ แล้ว นำผลการทดสอบมาทดสอบสัดส่วนของผลการทดสอบที่ตรงกัน โดยใช้ Z-test

$$Z = \frac{P - P_o}{\sqrt{\frac{P_o(1-P_o)}{n}}}$$

โดยที่ P = สัดส่วนของผลการทดสอบที่ตรงกัน

P_o = สัดส่วนของผลการทดสอบที่คาดว่าจะตรงกัน

n = จำนวนข้อมูล

สมมติฐานที่ต้องการทดสอบ

$$H_o \quad : \quad P = P_o$$

$$H_a \quad : \quad P < P_o$$

ผลการทดสอบจะปฏิเสธสมมติฐานว่าง เมื่อ Z ที่ได้จากการคำนวณมีค่าน้อยกว่า $-Z_{\alpha/2}$