5.1 ความนำ

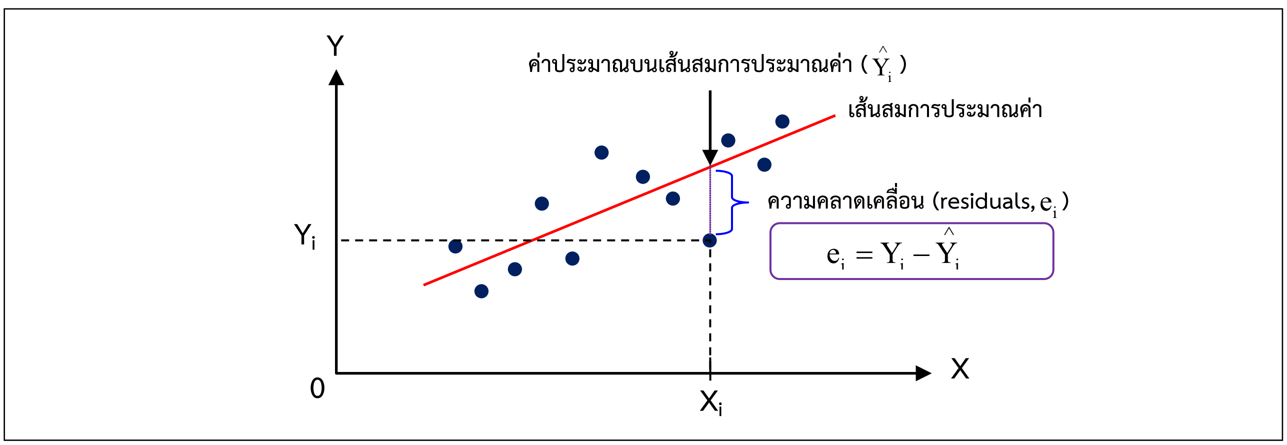
การวิเคราะห์ความแปรปรวนร่วม (ANCOVA) เป็นวิธีการวิเคราะห์ข้อมูลทางสถิติที่ถูกพัฒนาขึ้น ด้วยการนำเอาการวิเคราะห์ความแปรปรวนแบบทางเดียว (one-way ANOVA) มาใช้เปรียบเทียบค่าเฉลี่ยของตัวแปรผลลัพธ์ระหว่างกลุ่มและนำเอาการวิเคราะห์สมการถดถอยเชิงเส้น หรือ linear regression มาใช้ เพื่อปรับค่าความผันแปร หรือความคลาดเคลื่อนของตัวแปรผลลัพธ์ด้วยตัวแปรร่วมให้มีค่าลดน้อยลง ดังนั้นจึงส่งผลทำให้ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลมีความแม่นยำมากขึ้น หรือ กล่าวได้ว่า ***การวิเคราะห์ความแปรปรวนร่วม เป็นวิธีการผสมผสานระหว่างวิธีการ one-way ANOVA และวิธีการสมการถดถอยเชิงเส้นเข้าด้วยกัน*** นอกจากนี้ วิธีการทั้งสองดังกล่าวหากพิจารณาภายใต้โครงสร้างของตัวแบบทางสถิติ ยังพบว่า เป็นกรณีเฉพาะ (special cases) ที่อยู่ภายใต้ตัวแบบทางสถิติเช่นเดียวกัน เรียกว่า ***"ตัวแบบเชิงเส้นทั่วไป หรือ general linear model (GLM)"*** ซึ่งถือเป็นตัวแบบทางสถิติหนึ่งที่มีความสำคัญและถูกนำมาใช้ในการวิเคราะห์ข้อมูลสำหรับตัวแปรผลลัพธ์แบบต่อเนื่อง (continuous outcome) อย่างแพร่หลาย ***(von Eye and Wiedermann 2023)*** ***ดังนั้นในประเด็นการตรวจสอบข้อตกลงเบื้องต้นภายใต้วิธีการ ANCOVA ด้วยการนำเอาค่าความคลาดเคลื่อน หรือ residuals (e_i) มาพิจารณาจึงมีความสำคัญและจำเป็น*** โดยเฉพาะจากการศึกษางานวิจัยเกี่ยวข้องที่ผ่านมา พบว่า นักวิจัยส่วนใหญ่ ยังคงมีแนวคิด หรือ ความเข้าใจผิดค่อนข้างมากเกี่ยวกับการนำค่าความคลาดเคลื่อนมาใช้ในการพิจารณา เพื่อตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติและการเท่ากันของความแปรปรวน ซึ่งผลจากความเข้าใจผิดดังกล่าวยังอาจนำไปสู่แนวปฏิบัติในการประเมินและการตัดสินใจเกี่ยวกับการตรวจสอบข้อตกลงเบื้องต้นที่คลาดเคลื่อน หรือผิดไปจากความเป็นจริงได้ ***(Ernst and Albers 2017)*** ขณะเดียวกัน หากข้อตกลงเบื้องต้นดังกล่าว ถูกมองข้าม หรือมีการละเมิดเกิดขึ้น เช่น การแจกแจงแบบไม่ปกติ (non-normality) หรือ ความแปรปรวนแตกต่างกัน (heteroscedasticity) เป็นต้น ยังส่งผลกระทบโดยตรงต่อความถูกต้อง หรือ ความแม่นยำของค่าความคลาดเคลื่อนมาตรฐาน หรือ standard error (SE) ได้และทำให้ผลสรุปที่ได้ ทั้งจากการประมาณค่าพารามิเตอร์ด้วยค่า 95%CI และการทดสอบสมมติฐานทางสถิติด้วยค่า p-value ไม่ถูกต้องและนำไปสู่การสรุปผลที่ชี้นำไปในทางที่ผิดได้ ***(Flatt and Jacobs 2019)*** หรือ กล่าวได้ว่า ความถูกต้องของผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล ส่วนหนึ่งขึ้นกับการให้ความสำคัญเกี่ยวกับการละเมิดข้อตกลงเบื้องต้นทางสถิติของวิธีการที่ถูกเลือกนำมาใช้ ***(Knief and Forstmeier 2021)*** ดังนั้นเพื่อให้นักวิจัยสามารถเข้าใจแนวคิดและขั้นตอนในการนำค่าความคลาดเคลื่อน หรือ residuals (e_i) มาใช้พิจารณาข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนร่วม ภายใต้แบบแผนการทดลองสำหรับข้อมูลก่อน-หลังแบบวัดซ้ำสองกลุ่ม ด้วยโปรแกรม STATA ได้อย่างครอบคลุมและถูกต้อง ในบทนี้จึงได้นำเสนอเนื้อหาที่เกี่ยวข้องในประเด็นดังกล่าว ซึ่งประกอบด้วย แนวคิดการตรวจสอบข้อตกลงเบื้องต้นด้วยการนำค่าความคลาดเคลื่อนมาพิจารณา การตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ การตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับความแปรปรวนของความคลาดเคลื่อน รวมถึงบทสรุปและเอกสารอ้างอิง

5.2 แนวคิดการตรวจสอบข้อตกลงเบื้องต้นด้วยการนำค่าความคลาดเคลื่อนมาพิจารณา

การตรวจสอบข้อตกลงเบื้องต้นด้วยการนำค่าความคลาดเคลื่อนมาพิจารณา โดยเฉพาะสำหรับการวิเคราะห์สมการถดถอยเชิงเส้นตรงเกี่ยวกับประเด็นการแจกแจงแบบปกติและการเท่ากันของความแปรปรวน ส่วนใหญ่ยังคงพบว่า นักวิจัยมีแนวคิด หรือ ความเข้าใจผิด (misconceptions) ค่อนข้างมาก นั่นคือ ส่วนใหญ่นำค่าสังเกตของตัวแปรผลลัพธ์แบบต่อเนื่อง (Y_i) และตัวแปรอิสระ (X_i) แบบต่อเนื่อง มาใช้ในการตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติและการเท่ากันของความแปรปรวนโดยตรง ด้วยวิธีการใช้แผนภาพที่เกี่ยวข้อง มากกว่าการนำค่าความคลาดเคลื่อน หรือ residuals (e_i) มาพิจารณา *(Ernst and Albers 2017)* ทั้งที่ในความเป็นจริงแล้วแนวคิดเชิงทฤษฎีของการวิเคราะห์ถดถอยเชิงเส้นตรง ต้องการผลลัพธ์ ซึ่งทำให้ได้สมการประมาณค่าเชิงเส้น หรือเส้นประมาณค่าที่มีความพอดี (fitted line) โดยเป็นเส้นตรงที่ลากผ่านและเข้าใกล้ค่าข้อมูลแต่ละค่าได้ใกล้เคียงมากที่สุดและจากแนวคิดดังกล่าว จึงกลายเป็นที่มาของวิธีการประมาณค่าพารามิเตอร์ของการวิเคราะห์ถดถอยเชิงเส้นตรงที่เรียกว่า ***"วิธีกำลังสองน้อยที่สุด หรือ ordinary least square method (OLS)"*** และเป็นวิธีการที่มีการนำ "ค่าความคลาดเคลื่อน หรือ residuals (e_i)" มาพิจารณา ภายใต้แนวคิดที่ว่า

"ผลรวมกำลังสองของความคลาดเคลื่อน ต้องมีค่าน้อยที่สุด" หรือ $\sum e_i^2$ **มีค่าน้อยที่สุด**

นั่นคือ เมื่อกำหนดให้ **ค่าความคลาดเคลื่อน** หรือ **residuals** (e_i) ***หมายถึง ค่าระยะห่างที่ถูกวัดในแนวตั้ง หรือ แนวดิ่ง (vertical distances) ระหว่างค่าสังเกตของข้อมูล (Y_i) กับค่าประมาณบนเส้นสมการประมาณค่า ($\hat{Y}_i$)*** หรือ $e_i = Y_i - \hat{Y}_i$ *(Pandis 2016)* โดยสามารถแสดงได้ **ดังแสดงในแผนภาพที่ 5.1**



แผนภาพที่ 5.1 แสดงแนวทางการพิจารณาค่าความคลาดเคลื่อน หรือ residuals (e_i)

จากแนวคิดของการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด (OLS) ที่เกี่ยวข้องโดยตรงกับความคลาดเคลื่อน (residuals) จึงถือเป็นคุณลักษณะสำคัญภายใต้ตัวแบบเชิงเส้นทั่วไป (GLM) หรือการวิเคราะห์ถดถอยเชิงเส้นตรงที่นำไปสู่ประเด็นการตรวจสอบข้อตกลงเบื้องต้น ด้วยความคลาดเคลื่อน หรือ residuals ของวิธีการทางสถิติต่างๆ ที่อยู่ภายใต้ตัวแบบ GLM ซึ่งครอบคลุมทั้งวิธีการ t-test แบบสองกลุ่มอิสระ วิธีการวเคราะห์ความแปรปรวนแบบ one-way ANOVA วิธีการสมการถดถอยเชิงเส้นทั้งแบบง่ายและพหุคูณ รวมถึงการวิเคราะห์ความแปรปรวนร่วม (ANCOVA) ***(Burzykowski, Geubbelmans et al. 2023)*** ทั้งนี้เพื่อให้ค่าประมาณของพารามิเตอร์ที่ได้ มีคุณสมบัติเป็น ***"ตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด (best linear unbiased estimators, BLUE)" ตามทฤษฎีของ Gauss-Markov theorem ภายใต้เงื่อนไขเกี่ยวกับความคลาดเคลื่อน (residuals) ที่มีค่าเฉลี่ยเป็นศูนย์, ไม่มีความสัมพันธ์และมีความแปรปรวนเท่ากันในแต่ละค่าที่แตกต่างกันของตัวแปรอิสระ*** *(Schmidt and Finan 2018)* และจากเงื่อนไขภายใต้แนวคิดดังกล่าว จึงนำมาสู่ประเด็นการตรวจสอบข้อตกลงเบื้องต้นของความคลาดเคลื่อน (residuals) เกี่ยวกับการแจกแจงแบบปกติ (normality) การเท่ากันของความแปรปรวน (homoscedasticity) และความเป็นอิสระ (independence) ซึ่งจากประเด็นดังกล่าว หากนักวิจัยละเลย หรือ มีการละเมิดเกิดขึ้น อาจส่งผลกระทบตามมาต่อความแม่นยำของค่าประมาณที่ได้ รวมถึงความถูกต้องของผลลัพธ์ที่ได้จากการอนุมานค่าทางสถิติทั้งการทดสอบสมมติฐานทางสถิติด้วย p-value และการประมาณค่าพารามิเตอร์ด้วยช่วงเชื่อมั่น 95% CI ***(Ernst and Albers 2017)*** เนื่องจากความแม่นยำและความถูกต้องของการอนุมานค่าดังกล่าว ขึ้นกับความสอดคล้องที่เป็นไปตามเงื่อนไขของข้อตกลงเบื้องต้นภายใต้วิธีการทางสถิติของแต่ละวิธีที่ถูกเลือกนำมาใช้เป็นหลัก ***(Dombrowsky 2023)*** นั่นเอง

แต่อย่างไรก็ตาม ในประเด็นการตรวจสอบข้อตกลงเบื้องต้นของความคลาดเคลื่อน หรือ residuals เกี่ยวกับความเป็นอิสระ (independence) ดังกล่าวข้างต้น ในทางปฏิบัติ หากนักวิจัยมีการเลือกวิธีการสุ่มแบบง่าย (simple random sampling) หรือ แบบมีระบบ (systematic random sampling) มาใช้ในการสุ่มตัวอย่างสำหรับงานวิจัย ภายใต้แบบแผนการศึกษาที่ไม่ใช่การวัดซ้ำ (repeated measures) หรือ แบบกลุ่ม (cluster) ในทางปฏิบัตินักวิจัยสามารถการันตีความเป็นอิสระของค่าสังเกตและบ่งชี้ความสอดคล้องกับข้อตกลงเบื้องต้นของความคลาดเคลื่อนเกี่ยวกับความเป็นอิสระได้ โดยไม่จำเป็นต้องทำการตรวจสอบด้วยวิธีการอื่นเพิ่มเติม ***(Kim 2019)*** ดังนั้นจากความสำคัญของความคลาดเคลื่อน หรือ residuals ภายใต้แนวคิดของวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด (OLS) ดังกล่าวข้างต้น จึงน่าจะเป็นคำตอบที่ค่อนข้างชัดเจนสำหรับนักวิจัยซึ่งอาจมีข้อสงสัย หรือ คำถามว่า ***"ทำไม ? ในการตรวจสอบข้อตกลงเบื้องต้นสำหรับการวิเคราะห์ของวิธีการทางสถิติภายใต้ตัวแบบเชิงเส้นทั่วไป (GLM) รวมถึงการวิเคราะห์สมการถดถอยเชิงเส้นและการวิเคราะห์ความแปรปรวนร่วม ซึ่งถือเป็นกรณีเฉพาะของตัวแบบดังกล่าว จึงต้องนำเอาค่าความคลาดเคลื่อน หรือ residual มาพิจารณาเพื่อตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติและการเท่ากันของความแปรปรวน"*** โดยรายละเอียดสำหรับการตรวจสอบข้อตกลงเบื้องต้นในแต่ละประเด็น ด้วยการนำค่าความคลาดเคลื่อนมาพิจารณา จะได้กล่าวถึงในหัวข้อถัดไป

5.3 การตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ

การแจกแจงแบบปกติ (normal distribution) หรือ บางครั้งถูกเรียกว่า ***"Gaussian distribution"*** ถือเป็นการแจกแจงข้อมูลที่มีความสำคัญและถูกนำมาใช้ประกอบการพิจารณาทางสถิติ เมื่อตัวแปรผลลัพธ์มีลักษณะแบบต่อเนื่อง (continuous outcome) เนื่องจากมีคุณสมบัติเชิงทฤษฎีบางอย่างที่เที่ยงตรงและชัดเจน อีกทั้งยังสามารถนำไปใช้อธิบายลักษณะการแจกแจงของข้อมูลในสถานการณ์อื่นได้อย่างครอบคลุม โดยเฉพาะการมีคุณสมบัติการแจกแจงที่มีลักษณะแบบสมมาตรรอบค่าเฉลี่ย (symmetrical distribution around the mean) ซึ่งสามารถบ่งชี้ หรือ ระบุขอบเขตของการแจกแจงได้อย่างชัดเจนและครอบคลุม ทั้งในรูปแบบของค่าร้อยละ (%) และช่วงระหว่างค่าเฉลี่ยกับส่วนเบี่ยงเบนมาตรฐาน (SD) เช่น ประมาณร้อยละ 68.7 และร้อยละ 95.4 ของการแจกแจงข้อมูลเท่ากับ ค่าเฉลี่ย ± 1SD และค่าเฉลี่ย ± 2SD ตามลำดับ เป็นต้น *(Tsagris and Pandis 2021)* ดังนั้นจากคุณสมบัติดังกล่าว จึงทำให้การแจกแจงแบบปกติ ถูกนำมาใช้เป็นพื้นฐานทางทฤษฎีที่สำคัญภายใต้ตัวแบบ หรือ วิธีการทางสถิติสำหรับกรณีตัวแปรผลลัพธ์แบบต่อเนื่องเป็นส่วนใหญ่ โดยเฉพาะวิธีการทางสถิติที่อยู่ในกลุ่มที่เรียกว่า ***"สถิติทดสอบแบบพาราเมตริก หรือ parametric test"*** เช่น สถิติทดสอบ t-test แบบสองกลุ่มอิสระ สถิติทดสอบภายใต้วิธีการ ANOVA หรือ การวิเคราะห์สมการถดถอยเชิงเส้นตรง เป็นต้น *(Ghasemi and Zahediasl 2012)* ดังนั้นเพื่อให้นักวิจัยสามารถทำความเข้าใจและเรียนรู้แนวคิดและหลักการในการนำความคลาดเคลื่อน (residual) มาใช้ตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ รวมถึงวิธีการตรวจสอบข้อตกลงเบื้องต้นดังกล่าว ด้วยโปรแกรม STATA ได้อย่างเหมาะสมและครอบคลุม จึงสามารถแบ่งประเด็นการนำเสนอเป็นรายข้อได้ดังนี้

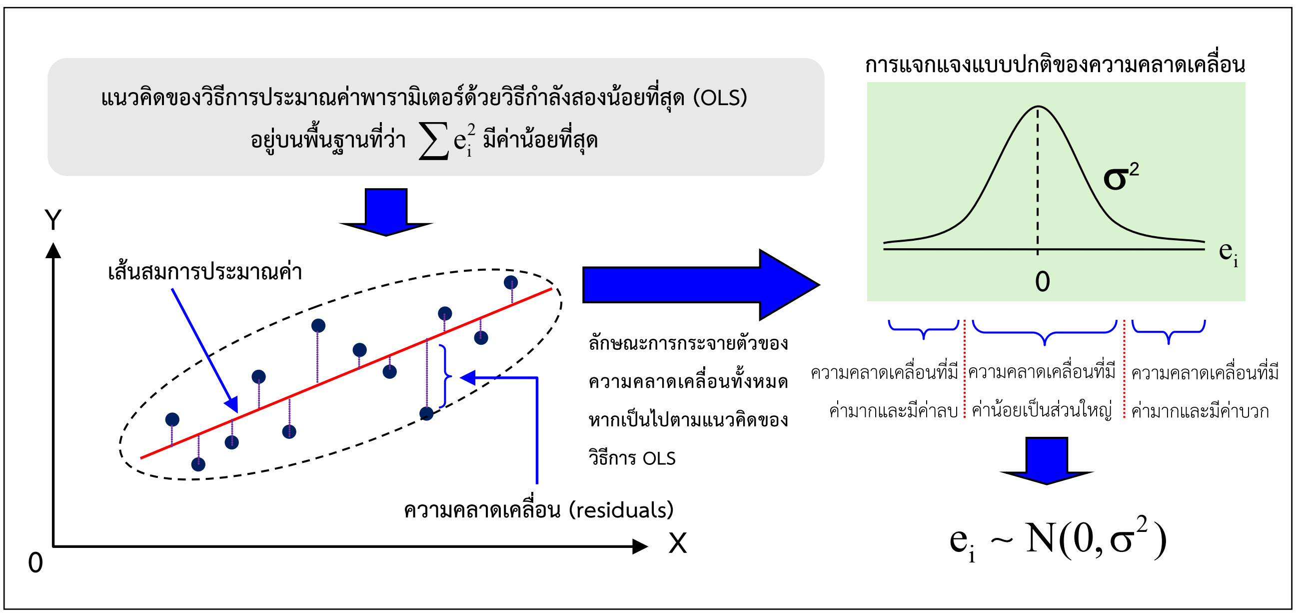
5.3.1 การนำค่าความคลาดเคลื่อนมาใช้ในการตรวจสอบรูปแบบการแจกแจงแบบปกติ
5.3.2 ผลกระทบจากการละเมิดข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ
5.3.3 วิธีการตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ
5.3.4 แนวทางพิจารณาเมื่อข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติถูกละเมิด
5.3.5 แนวทางจัดการข้อมูลที่ละเมิดข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ

5.3.1 การนำค่าความคลาดเคลื่อนมาใช้ในการตรวจสอบรูปแบบการแจกแจงแบบปกติ

เมื่อนักวิจัยเลือกตัดสินใจนำวิธีการวิเคราะห์ความแปรปรวนร่วม (ANCOVA) มาใช้ในการวิเคราะห์ข้อมูลภายใต้แบบแผนงานวิจัยเชิงทดลองสำหรับข้อมูลก่อน-หลังแบบวัดซ้ำสองกลุ่ม การตรวจสอบข้อตกลงเบื้องต้นในทางปฏิบัติ จึงจำเป็นต้องอาศัยแนวคิดและหลักการพื้นฐานตามโครงสร้างของตัวแบบเชิงเส้นทั่วไป (GLM) เนื่องจากวิธีการ ANCOVA ถือเป็นกรณีเฉพาะที่อยู่ภายใต้โครงสร้างของตัวแบบดังกล่าว *(Schober and Vetter 2021)* ดังนั้นประเด็นรูปแบบการแจกแจงแบบปกติของความคลาดเคลื่อน หรือ residuals จึงถูกนำมาพิจารณาเพื่อตรวจสอบเป็นข้อตกลงเบื้องต้นที่สำคัญ ตามแนวคิดเชิงทฤษฎีภายใต้วิธีการประมาณค่าด้วยวิธีกำลังสองน้อยที่สุด (OLS) ที่ว่า

หากสมการประมาณค่าเชิงเส้น หรือ เส้นประมาณค่าที่มีความพอดี (fitted line) เป็นเส้นตรงที่สามารถแสดงค่าเฉลี่ยของค่าสังเกตทั้งหมดได้อย่างแท้จริง นั่นหมายความว่า ***ความคลาดเคลื่อน หรือ residuals ที่มีค่าน้อย จะมีความถี่จำนวนมาก (หรือ กระจายอยู่รอบค่าศูนย์ ซึ่งเป็นค่าเฉลี่ย) ขณะที่ความคลาดเคลื่อนที่มีค่ามาก จะมีความถี่จำนวนน้อย (หรือ พบได้ในส่วนปลายของทั้งสองข้างจากค่าเฉลี่ย ซึ่งเท่ากับศูนย์)*** ***(Kim 2019)*** ซึ่งจากลักษณะการกระจายตัวของความคลาดเคลื่อนเมื่อเส้นประมาณค่าที่ได้มีความพอดีดังกล่าว จึงทำให้ความคลาดเคลื่อนที่ได้ มีแนวโน้มเข้าสู่การแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ^2 หรือ เขียนได้ว่า $e_i \sim N(0,\sigma^2)$

ซึ่งจากแนวคิดและรูปแบบการกระจายตัวของความคลาดเคลื่อนดังกล่าวข้างต้น จึงเป็นแหล่งที่มาว่า ***"ทำไม? เราจึงต้องทำการตรวจสอบข้อตกลงเบื้องต้นของความคลาดเคลื่อนเกี่ยวกับการแจกแจงแบบปกติ (normality)"*** โดยสามารถแสดงได้ **ดังแสดงในแผนภาพที่ 5.2**



แผนภาพที่ 5.2 แสดงการแจกแจงแบบปกติของความคลาดเคลื่อน หรือ residuals (e_i) ตามแนวคิดของวิธีการกำลังสองน้อยที่สุด (OLS)

5.3.2 ผลกระทบจากการละเมิดข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติ

อย่างไรก็ตาม นอกจากนักวิจัยจะต้องทำความเข้าใจเกี่ยวกับประเด็นที่มาของการนำรูปแบบการแจกแจงแบบปกติมาใช้ในการพิจารณา เพื่อตรวจสอบข้อตกลงเบื้องต้นด้วยความคลาดเคลื่อน ดังที่กล่าวไปข้างต้นแล้ว ผลกระทบที่เกิดขึ้นจากการละเมิดข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติดังกล่าว ยังถือเป็นอีกประเด็นที่มีความสำคัญ ซึ่งนักวิจัยจำเป็นต้องทราบและควรคำนึงถึง เนื่องจากมีผลโดยตรงต่อความแม่นยำและความน่าเชื่อถือของผลลัพธ์ที่ได้ *(Ghasemi and Zahediasl 2012)* โดยในทางปฏิบัติ ข้อตกลงเบื้องต้นของความคลาดเคลื่อน

เกี่ยวกับการแจกแจงแบบปกติ จะเกี่ยวข้องโดยตรงกับผลลัพธ์ที่ได้จากวิธีการอนุมานค่าทางสถิติ ซึ่งประกอบด้วยการประมาณค่าพารามิเตอร์ด้วยช่วงเชื่อมั่น 95%CI และการทดสอบสมมติฐานทางสถิติด้วยค่า p-value *(Flatt and Jacobs 2019)* เนื่องจากการได้มาของผลลัพธ์จากวิธีการอนุมานค่าทางสถิติทั้งสองดังกล่าว โดยเฉพาะวิธีการทางสถิติที่อยู่ภายใต้โครงสร้างของตัวแบบเชิงเส้นทั่วไป (GLM) จำเป็นต้องอาศัยการแจกแจงค่าสถิติของตัวอย่าง หรือ sampling distribution ภายใต้เงื่อนไขการนำรูปการแจกแจงแบบปกติมาใช้ในการพิจารณา *(Garg, Bhaskar et al. 2020)* ทั้งนี้เพื่อให้ได้ค่าความคลาดเคลื่อนมาตรฐาน หรือ standard error (SE) ซึ่งถือเป็นค่าประมาณของส่วนเบี่ยงเบนมาตรฐานในระดับประชากร ที่มีความแม่นยำและไม่ลำเอียง แต่อย่างไรก็ตาม แม้ว่าโดยทั่วไป แต่ละวิธีการทางสถิติ จะมีสถิติที่นำมาใช้ประมาณค่า อาทิ ค่าเฉลี่ย ($\overline{X}$) ผลต่างของค่าเฉลี่ย ($\overline{X}_1 - \overline{X}_2$) หรือ ค่าเฉลี่ยของผลต่าง ($\overline{X}_{diff}$) เป็นต้น ภายใต้สถิติทดสอบ เช่น t, z หรือ F เป็นต้น รวมถึงสูตรที่นำมาใช้ในการคำนวณค่า SE ที่แตกต่างกัน แต่นักวิจัยอาจสังเกตได้ว่า ***ทุกสูตรของการคำนวณค่า SE ในแต่ละวิธีการทางสถิติจะนำจำนวนขนาดตัวอย่าง (n) มาพิจารณาเป็นตัวหารเช่นเดียวกันเสมอ*** **ดังแสดงกรณีตัวอย่างในตารางที่ 5.1**

ตารางที่ 5.1 แสดงกรณีตัวอย่างของวิธีการทางสถิติ/สถิติที่ใช้ประมาณค่าและสถิติทดสอบ รวมถึงสูตรในการคำนวณค่าความคลาดเคลื่อนมาตรฐาน (SE)

วิธีการทางสถิติ/สถิติที่ใช้ประมาณค่าและสถิติทดสอบ	สูตรในการคำนวณค่าความคลาดเคลื่อนมาตรฐาน (SE)
1. ***วิธีการทางสถิติ*** : t-test แบบหนึ่งกลุ่ม ***สถิติที่ใช้ประมาณค่า*** : ค่าเฉลี่ย ($\overline{X}$) ***สถิติทดสอบ*** : t-test	$SE = \frac{S}{\sqrt{n}}$
2. ***วิธีการทางสถิติ*** : t-test แบบสองกลุ่มอิสระ (กรณี $\sigma_1^2 \neq \sigma_2^2$) ***สถิติที่ใช้ประมาณค่า*** : ผลต่างของค่าเฉลี่ย ($\overline{X}_1 - \overline{X}_2$) ***สถิติทดสอบ*** : t-test	$SE = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
3. ***วิธีการทางสถิติ*** : t-test แบบสองกลุ่มอิสระ (กรณี $\sigma_1^2 = \sigma_2^2$) ***สถิติที่ใช้ประมาณค่า*** : ผลต่างของค่าเฉลี่ย ($\overline{X}_1 - \overline{X}_2$) ***สถิติทดสอบ*** : t-test	$SE = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ เมื่อ $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$
4. ***วิธีการทางสถิติ*** : t-test แบบสองกลุ่มไม่อิสระ (paired t-test) ***สถิติที่ใช้ประมาณค่า*** : ค่าเฉลี่ยของผลต่าง ($\overline{X}_{diff}$) ***สถิติทดสอบ*** : t-test	$SE = \frac{S_{diff}}{\sqrt{n}}$

จากกรณีตัวอย่างดังแสดงในตารางที่ 5.1 ข้างต้น จึงพบว่า แม้นักวิจัยจะเลือกนำเอาวิธีการทางสถิติใด ๆ มาใช้ทำการวิเคราะห์ข้อมูลภายใต้วิธีการอนุมานค่าทางสถิติ เพื่อตอบโจทย์คำถามวิจัย หรือ วัตถุประสงค์ของการศึกษาที่แตกต่างกันและมีสูตรในการคำนวณค่าความคลาดเคลื่อนมาตรฐาน (SE) ของแต่ละวิธีการทางสถิติแตกต่างกันไป แต่ขณะเดียวกัน ยังพบว่า มีประเด็นที่คล้ายคลึงกัน นั่นคือ การนำขนาดตัวอย่าง (n) มาพิจารณาเป็นตัวหาร

ในทุก ๆ สูตรที่ถูกนำมาใช้ในการคำนวณหาค่า SE เช่นเดียวกันเสมอ ดังนั้นในทางปฏิบัติ จึงกล่าวได้ว่า ***"ค่าความคลาดเคลื่อนมาตรฐาน (SE) แปรผกผันโดยตรงกับขนาดตัวอย่าง (n)"*** หรือ เขียนได้ว่า

ค่าความคลาดเคลื่อนมาตรฐาน $(\mathrm{SE}) \sim \frac{1}{\mathrm{n}}$, เมื่อ n = ขนาดตัวอย่าง

จากการแปรผกผันระหว่างค่าความคลาดเคลื่อนมาตรฐานและขนาดตัวอย่างดังกล่าว จึงสามารถอธิบายได้ว่า ***"เมื่อขนาดตัวอย่าง มีจำนวนเพิ่มมากขึ้น จะทำให้ค่าความคลาดเคลื่อนมาตรฐานที่ได้ มีแนวโน้มลดน้อยลง ขณะเดียวกัน เมื่อขนาดตัวอย่าง มีจำนวนลดลง จะทำให้ค่าความคลาดเคลื่อนมาตรฐาน มีแนวโน้มเพิ่มมากขึ้น"*** ซึ่งการเพิ่มขึ้นและลดลงของค่าความคลาดเคลื่อนมาตรฐาน หรือ SE ดังกล่าว ยังส่งผลกระทบต่อเนื่องไปสู่ผลลัพธ์ที่ได้จากการประมาณค่าพารามิเตอร์และการทดสอบสมมติฐานทางสถิติ ทั้งช่วงเชื่อมั่น 95%CI และค่า p-value ตามลำดับอีกด้วย เนื่องจากในการพิจารณาเพื่อคำนวณค่าผลลัพธ์ที่ได้ดังกล่าว จำเป็นต้องอาศัยค่าความคลาดเคลื่อนมาตรฐานเข้าไปเกี่ยวข้องเสมอ นั่นคือ

- **ในการคำนวณ ค่า 95%CI** ค่าความคลาดเคลื่อนมาตรฐาน หรือ SE ***จะถูกพิจารณาเข้าไปเป็นตัวคูณในสูตรโดยตรง*** เพื่อบ่งชี้ผลที่เกิดขึ้นจากความคลาดเคลื่อนของการสุ่ม (sampling error) ก่อนถูกนำไปพิจารณาด้วยการบวกและลบ (±) กับค่าสถิติที่ใช้ประมาณค่า เพื่อกำหนดค่าออกมาเป็นช่วงที่มีค่าต่ำและค่าสูงภายใต้ระดับความเชื่อมั่นที่นักวิจัยกำหนด เช่น ช่วงเชื่อมั่นที่ระดับ 95% หรือ 95%CI เป็นต้น
- **ในการคำนวณค่า p-value** ค่าความคลาดเคลื่อนมาตรฐาน หรือ SE ***จะไม่ถูกพิจารณาโดยตรงเพื่อกำหนดค่า p-value แต่จะถูกพิจารณาทางอ้อมด้วยการเริ่มต้นเข้าไปเป็นตัวหารในสูตรของการคำนวณค่าสถิติทดสอบ เช่น t, z หรือ F เป็นต้น*** จากนั้นผลของค่าสถิติทดสอบที่ได้ จึงจะถูกนำไปพิจารณาเพื่อกำหนดค่า p-value ในรูปของความน่าจะเป็นอีกครั้งหนึ่ง

ซึ่งจากแนวทางการพิจารณาเพื่อคำนวณค่า 95%CI และค่า p-value ดังกล่าวข้างต้น จึงสามารถนำมาสรุปเป็นแผนภาพเพื่อแสดงการเชื่อมโยงและบ่งชี้ผลกระทบที่เกิดขึ้นต่อค่า 95%CI และ p-value จากขนาดตัวอย่าง (n) และค่าความคลาดเคลื่อนมาตรฐาน (SE) ที่เปลี่ยนแปลงไป ภายใต้กรณีตัวอย่างของวิธีการวิเคราะห์ ***ด้วยสถิติ t-test แบบหนึ่งกลุ่ม หรือ one sample t-test*** ได้ **ดังแสดงในแผนภาพที่ 5.3**

วิธีการวิเคราะห์ t-test แบบหนึ่งกลุ่ม หรือ one sample t-test

- **การแจกแจงค่าสถิติของตัวอย่าง :** *การแจกแจงค่าเฉลี่ยของตัวอย่าง*
- **สถิติที่ใช้ในการประมาณค่า :** *ค่าเฉลี่ยหนึ่งกลุ่ม* ($\overline{X}$)
- **สูตรความคลาดเคลื่อนมาตรฐาน (SE) :** $SE = \frac{S}{\sqrt{n}}$
- **สถิติทดสอบ :** *t-test* **, โดยมีสูตรดังนี้ :**

$$t = \frac{\overline{X} - \mu}{SE} \quad \text{หรือ} \quad t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

การประมาณค่าพารามิเตอร์ด้วย 95%CI

ผลที่เกิดขึ้นจากความคลาดเคลื่อนของการสุ่ม (sampling error)

- สูตรการคำนวณช่วงเชื่อมั่น : $\overline{X} \pm \boxed{t_{\frac{\alpha}{2}}(SE)}$ หรือ $\overline{X} \pm \boxed{t_{\frac{\alpha}{2}}(\frac{S}{\sqrt{n}})}$

สถิติที่ใช้ในการประมาณค่า

- จากสูตร จะสังเกตได้ว่า ความคลาดเคลื่อนมาตรฐาน (SE) *เป็นตัวคูณ*

  >>> ถ้า n *น้อย* จะได้ค่า SE *มาก* คำนวณตามสูตร จะได้ $\overline{X} \pm$ ค่ามาก ส่งผลให้ 95%CI *มีช่วงกว้าง*

  >>> ถ้า n *มาก* จะได้ค่า SE *น้อย* คำนวณตามสูตร จะได้ $\overline{X} \pm$ ค่าน้อย ส่งผลให้ 95%CI *มีช่วงแคบ*

การทดสอบสมมติฐานทางสถิติด้วย p-value

❶. การคำนวณสถิติทดสอบตามสูตร : $t = \frac{\overline{X} - \mu}{SE}$ หรือ $t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$

จากสูตร จะสังเกตได้ว่า ความคลาดเคลื่อนมาตรฐาน (SE) *เป็นตัวหาร*

>>> ถ้า n *น้อย* จะได้ค่า SE *มาก* คำนวณตามสูตร จะได้ $t = \frac{\overline{X} - \mu}{\text{ค่ามาก}}$ ส่งผลให้สถิติทดสอบ t *มีค่าน้อย*

>>> ถ้า n *มาก* จะได้ค่า SE *น้อย* คำนวณตามสูตร จะได้ $t = \frac{\overline{X} - \mu}{\text{ค่า}}$ ส่งผลให้สถิติทดสอบ t *มีค่ามาก*

❷. หลักการในการคำนวณค่า p-value :

>>> ถ้าค่าสถิติทดสอบ *มาก* ค่า p-value มีแนวโน้มได้ค่า *น้อย* โอกาสที่จะพบการมีนัยสำคัญ (sign.) *มีมาก*

>>> ถ้าค่าสถิติทดสอบ *น้อย* ค่า p-value มีแนวโน้มได้ค่า *มาก* โอกาสที่จะพบการมีนัยสำคัญ (sign.) *มีน้อย*

แผนภาพที่ 5.3 แสดงผลกระทบของ 95%CI และ p-value จากขนาดตัวอย่างและความคลาดเคลื่อนมาตรฐาน กรณีวิธีการวิเคราะห์ t-test แบบหนึ่งกลุ่ม