*ดร.อัจฉราวรรณ งามญาติ*
รองศาสตราจารย์ประจำภาควิชาการเงิน
คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์

# อันเนื่องมาแต่สูตรของยามาเน่*

## ABSTRACT

Thai students usually use 'Yamane's formula' for calculating the sample size, even when the students' measurement scales are not dichotomous variables. Yamane simplified the sample size formula using dichotomous variables that can be transformed proportionally, not continuously. In the proportional case the maximum variance of the sample will be 0.25, which is different from the case of continuous variables, when the maximum variance is not limited to any number. Yamane chose to simplify the original proportional sample size formula in a finite population by using 0.25, the maximum variance of the proportional case, thereby yielding 'Yamane's formula'. This new formula is a function of the population size with a random sampling error where the variance has disappeared from the formula. If the students have only two choices in their measurement scales, using 'Yamane's formula' is acceptable. However, more often the graduate students will use the five point likert scales in their research papers where scales are treated as continuous variables or metric variables. Therefore the sample size formula from the continuous variable case should be used instead of 'Yamane's formula'.

*ผู้เขียนขอขอบคุณ Mr.Barry Satz ที่กรุณาตรวจแก้ภาษาอังกฤษ ขอขอบคุณ ดร.นิติ รัตนปรีชาเวช และผู้ประเมินบทความของวารสารบริหารธุรกิจ ที่กรุณาช่วยอ่านและให้ข้อเสนอแนะในการปรับปรุงข้อเขียนนี้ให้มีความสมบูรณ์ยิ่งขึ้น ข้อผิดพลาดทั้งหมดที่เกิดขึ้นในบทความ ผู้เขียนขอน้อมรับไว้แต่เพียงผู้เดียว

# บทคัดย่อ

นักศึกษาไทยนิยมใช้ “สูตรของยามาเน่” ในการคำนวณหาขนาดของตัวอย่างทั้งที่มาตรวัดที่นักศึกษาใช้ไม่ใช่มาตรวัดแบบที่มีแค่สองตัวเลือก ในการสร้างสูตรของยามาเน่ที่นักศึกษาไทยชอบใช้ยามาเน่เลือกแปลงมาจากสูตรในการหาขนาดของตัวอย่างที่มาตรวัดมีสองตัวเลือกหรือแบบสัดส่วน ไม่ใช่สูตรที่ใช้กับตัวแปรที่มีมาตรวัดแบบต่อเนื่อง ทั้งนี้ค่าความแปรปรวนที่มากที่สุดของตัวอย่างกรณีของสัดส่วนสามารถแทนค่าด้วยค่า0.25 ซึ่งต่างจากกรณีที่ตัวแปรมีมาตรวัดแบบต่อเนื่องที่ไม่อาจระบุได้แน่ชัดว่าค่าความแปรปรวนที่มากที่สุดควรเป็นเท่าใด ยามาเน่เลือกแปลงสูตรในการหาขนาดของตัวอย่างกรณีสัดส่วนโดยสมมติให้ประชากรมีขนาดเล็ก และเลือกที่จะแทนค่าความแปรปรวนด้วย0.25 ซึ่งถือว่าให้ค่าความแปรปรวนที่สูงสุดของกรณีสัดส่วนแล้ว ดังนั้นภายใต้สูตรของยามาเน่ขนาดของตัวอย่างจึงขึ้นอยู่กับขนาดของประชากรและค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง โดยดูเหมือนกับว่าค่าความแปรปรวนของตัวอย่างหลุดหายไปจากสูตรของยามาเน่ อันที่จริงถ้ามาตรวัดที่นักศึกษาใช้เป็นแบบที่มีแค่สองตัวเลือกการจะใช้สูตรของยามาเน่ก็ไม่ได้ผิดอะไร แต่งานวิจัยของนักศึกษาปริญญาโทส่วนใหญ่ซึ่งวัดตัวแปรที่ไม่สามารถสังเกตได้นิยมใช้มาตรวัดลิเคิร์ทแบบห้าตัวเลือกซึ่งในทางการตลาดอนุโลมให้มาตรวัดนี้เป็นตัวแปรแบบต่อเนื่องหรือแบบเมตริก สูตรของยามาเน่ที่แปลงมาจากกรณีสัดส่วนจึงไม่เหมาะที่จะนำมาใช้ ดังนั้นหากต้องการหาขนาดของตัวอย่างแบบใช้สูตรเมื่อมีการเลือกตัวอย่างแบบสุ่มจึงต้องกลับไปใช้สูตรที่เหมาะกับตัวแปรแบบต่อเนื่องจึงจะถูกต้อง

# บทนำ

ที่มาของข้อเขียนนี้มาจากความรำคาญระคนความเหนื่อยที่จะต้องอธิบายให้นักศึกษารุ่นแล้วรุ่นเล่าที่นิยมขอยืมสูตรของยามาเน่ (Yamane's Formula) มาใช้โดยที่ไม่ได้คิดสักนิดว่าสูตรนี้มีที่มาที่ไปอย่างไร แค่เห็นมีคนอ้างมาใช้ก็ลอกๆ ตามเขาเท่านั้น ทั้งๆ ที่หลายครั้งมาตรวัดที่ใช้ไม่ได้เอื้อที่จะให้ใช้สูตรนี้เลย หากลองไปเปิดวิทยานิพนธ์ของนักศึกษาไทยดูจะพบการอ้างสูตรนี้มาใช้บ่อยมาก ครั้งแรกๆ ผู้เขียนก็พยายามอธิบายให้นักศึกษาที่บังเอิญมาเป็นนักศึกษาในที่ปรึกษาของผู้เขียนเข้าใจแม้ว่าจะต้องใช้เวลาตามสมควร แต่รู้สึกว่านักศึกษารุ่นต่อรุ่นก็มีความเข้าใจผิดเหมือนๆ กันซะทุกรุ่น จนผู้เขียนเบื่อที่จะต้องมานั่งบรรยายให้นักศึกษารุ่นแล้วรุ่นเล่าฟังซ้ำๆ กันเป็นรายคน นักศึกษาคงไม่รู้สึกอะไร แต่สำหรับคนสอนที่ต้องพูดอะไรซ้ำซากให้นักศึกษาฟังทีละคนมันสุดจะกล้ำกลืนเต็มทน จึงต้องพยายามนำมาเรียบเรียงเขียนให้เห็น เพื่อให้นักศึกษาหยิบสูตรนี้มาใช้ด้วยความระมัดระวัง เผื่อโอกาสหน้าถ้ามีนักศึกษาในที่ปรึกษาหยิบสูตรนี้มาใช้ผิดๆ อีกจะได้โยนข้อเขียนนี้ให้ไปอ่านก่อนเพื่อประหยัดเวลาในการอธิบายให้สั้นลง

แต่เดิมนั้น ผู้เขียนคิดว่าคุณยามาเน่เธอไม่ได้รู้เห็นด้วยกับสูตรของเธอที่มีคนนำไปอ้างผิดอยู่บ่อยๆ โดยไม่ได้สนใจกลับไปเปิดตำราสถิติที่เธออุตส่าห์เขียนออกมาเล่มใหญ่ว่าสูตรของเธอมีที่มาที่ไปอย่างไร และมีข้อสมมติอะไรบ้าง ก่อนที่จะวกเข้าสูตรของยามาเน่คงต้องเท้าความถึงสูตรที่ใช้ในการหาขนาดของตัวอย่างทั่วๆ ไปก่อน ซึ่งเป็นสิ่งที่พอดูได้จากตำราวิจัยตลาดภาษาอังกฤษแทบทุกเล่มซึ่งมักจะมีการพิสูจน์สูตรในการหาขนาดของตัวอย่างไว้ให้ดู แต่ไม่เข้าใจว่าทำไมนักศึกษาไทยไม่นิยมใช้ ถ้าจะให้เดาคงเป็นเพราะไม่ทราบว่าจะหาค่าความแปรปรวน (Variance) ของตัวแปรได้อย่างไร เพราะยังไม่ได้มีการเก็บข้อมูลรวมทั้งไม่รู้ว่ามาตรวัดเป็นเช่นไรก็คงจะจริง แต่จากประสบการณ์ที่ได้ให้คำปรึกษานักศึกษาปริญญาโทหลายๆ รุ่นพบว่า งานของนักศึกษาปริญญาโทมักจะวัดด้วยมาตรวัดลิเคิร์ท (Likert Scale) ซึ่งเป็นมาตรวัดแบบ 5 จุด เริ่มจากไม่เห็นด้วยอย่างยิ่ง (1) ถึงเห็นด้วยอย่างยิ่ง (5) เป็นส่วนใหญ่ จึงไม่น่ายากที่จะลองเดาดูว่าค่าถัวเฉลี่ย ($\bar{x}$) และค่าความแปรปรวน ($s^2$) ของกลุ่มตัวอย่างควรมีค่าอยู่ระหว่างเท่าไรจึงจะสมเหตุผล การที่นักศึกษาพยายามหลีกเลี่ยงที่จะหาค่าความแปรปรวนเพราะคิดไม่ออกว่าจะหามาจากไหน จึงเลี่ยงไปใช้สูตรของยามาเน่ซึ่งไม่มีค่าความแปรปรวนปรากฏให้เห็นอยู่ในสูตรเป็นทางเลือกแทน ความจริงแล้วการหาค่าความแปรปรวนของตัวแปรที่ใช้มาตรวัดแบบลิเคิร์ทเป็นเรื่องง่ายกว่าที่คาดคิด ไม่ได้ลึกลับซับซ้อนจนต้องพยายามเลี่ยงไปใช้สูตรของยามาเน่อย่างไรก็ตามก่อนที่จะสงสัยว่าทำไมผู้เขียนจึงดูเหมือนจะค้านสูตรของยามาเน่นัก ลองมาดูที่มาของสูตรที่ใช้ในการหาขนาดของตัวอย่างที่นิยมใช้กันโดยทั่วไปก่อนดีกว่า

จากการที่ใช้ตำราการวิจัยตลาดภาษาอังกฤษของผู้แต่งหลายๆ ท่าน สิ่งที่ผู้เขียนพอจะได้จากตำราที่ใช้ทุกเล่มว่าจะมีวิธีการพิสูจน์สูตรที่ใช้ในการหาขนาดของตัวอย่างคล้ายคลึงกัน และผู้เขียนก็ยังไม่พบว่าท่านใดพูดถึงสูตรของยามาเน่เลย แต่กลับพบว่าตำราวิจัยตลาดของผู้แต่งชาวไทยบางเล่มมักจะอ้างถึงสูตรนี้ โดยผู้เขียนเองก็ไม่แน่ใจว่าคุณยามาเน่เธอรู้เรื่องการอ้างชื่อเธอผิดๆ หรือไม่ ซึ่งการที่จะได้ขนาดของตัวอย่างที่เหมาะสมจะทำได้หลายวิธี อย่างไรก็ตามการใช้สูตรจะทำได้ก็ต่อเมื่อเงื่อนไขของตัวแปรมาจากการสุ่ม (Random) เลือกมาแบบสุ่มสี่สุ่มห้าสูตรที่ใช้ก็คงให้ค่าที่ไม่ถูกต้องเพราะหากไม่ใช้การเลือกตัวอย่างแบบสุ่มหรือที่นักศึกษาที่เคยเรียนวิจัยตลาดรู้จักกันว่าการเลือกตัวอย่างแบบใช้ความน่าจะเป็น (Probability Sampling) ก็จะไม่สามารถหาขนาดของตัวอย่างที่ให้ค่าระดับความเชื่อมั่นที่ต้องการโดยยอมให้ความคลาดเคลื่อนจากการสุ่มตัวอย่างอยู่ในระดับที่ยอมรับได้ด้วย

ก่อนอื่นต้องเท้าความถึงการแจกแจงของตัวแปรทางสถิติ ซึ่งมีอยู่มากมายหลายแบบ แต่เท่าที่ผู้เขียนเคยเห็นการอธิบายถึงที่มาของขนาดของตัวอย่างมักจะใช้ 2 แบบ คือ ตัวแปรที่มีการแจกแจงแบบปกติ (Normal Distribution) และตัวแปรที่มีการแจกแจงแบบ 2 ค่า (Binomial Distribution) ในทางวิจัยตลาดจะเรียกตัวแปรประเภทหลังนี้ว่าตัวแปรที่มี 2 ค่า (Dichotomous Variable) สำหรับตัวแปรที่จะใช้การแจกแจงแบบปกติได้จะต้องเป็นตัวแปรที่มีมาตรวัดแบบต่อเนื่อง (Continuous Variable) ซึ่งก็กินความถึงตัวแปรที่มีมาตรวัดแบบช่วง (Interval Scale) และตัวแปรที่มีมาตรวัดแบบสัดส่วน (Ratio Scale) เท่านั้น สำหรับตัวแปรที่มีมาตรวัดแบบแบ่งกลุ่ม (Nominal Scale) และที่มีมาตรวัดแบบเรียงลำดับ (Ordinal Scale) ไม่อาจจัดอยู่ในตัวแปรที่มีมาตรวัดแบบต่อเนื่อง (Continuous Variable) ได้ ยกเว้นตัวแปรนั้นมีแค่ 2 ค่า หรือที่สามารถแปลงให้อยู่ในรูปสัดส่วน (Proportion) ได้

ในกรณีของสัดส่วน ถ้าให้

โอกาสที่จะประสบความสำเร็จ = $\pi$

และ โอกาสที่จะไม่ประสบความสำเร็จ = $1-\pi$

นั่นหมายความว่าจากการสุ่มตัวอย่างแต่ละครั้ง สามารถหาสัดส่วนของโอกาสที่จะประสบความสำเร็จจากโอกาสทั้งหมดซึ่งเท่ากับหนึ่งได้ ในกรณีเช่นนี้มาตรวัดจะมีแค่ 2 ค่า คือ ถ้าใช่ มาตรวัด = 1 ถ้าไม่ใช่ มาตรวัด = 0 หรือที่รู้จักกันในทางสถิติคือการใช้ตัวแปรหุ่น (Dummy Variable) ด้วยการลงรหัสค่าทั้งสองของตัวแปรให้เท่ากับหนึ่งและศูนย์เท่านั้น ในกรณีเช่นนี้แม้ว่ามาตรวัดจะเป็นมาตรวัดแบบไม่ต่อเนื่องก็ตาม แต่การแทนค่ามาตรวัดด้วยหนึ่งและศูนย์จะสามารถหาค่าถัวเฉลี่ยและค่าความแปรปรวนของตัวแปรได้เหมือนกับตัวแปรที่มีมาตรวัดแบบต่อเนื่องเช่นกัน ทั้งนี้ค่าถัวเฉลี่ยของตัวแปรจะเท่ากับสัดส่วนของจำนวนผู้เลือกทางเลือกที่ใช่และลงรหัสเป็นหนึ่งต่อจำนวนตัวอย่างทั้งสิ้น จะทำให้ได้สัดส่วนซึ่งเป็นตัวเลขที่เปรียบเสมือนต่อเนื่อง จึงหาค่าประมาณการช่วง (Interval Estimate) ของสัดส่วนของประชากรได้เช่นเดียวกับตัวแปรแบบต่อเนื่องนั่นเอง ดังนั้นหากกำหนดระดับความเชื่อมั่นที่ต้องการและความคลาดเคลื่อนจากการสุ่มตัวอย่างที่พอรับได้ ก็จะนำไปสู่การกำหนดขนาดของตัวอย่างที่เหมาะสมได้

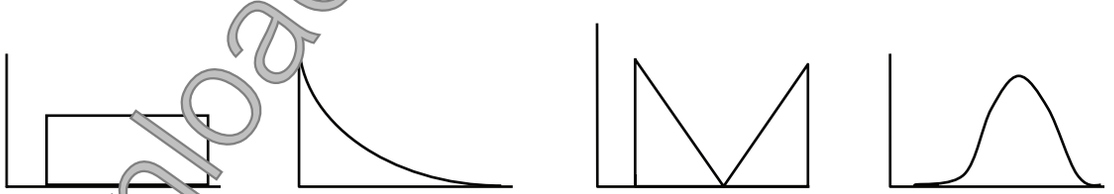
# การกระจายของการสุ่มตัวอย่าง กรณีตัวแปรต่อเนื่อง

ที่มาของทฤษฎีทางสถิติที่นักวิจัยใช้เป็นพื้นฐานในการศึกษาสถิติเบื้องต้น ได้แก่ Central Limit Theorem ตามทฤษฎีนี้ได้มาจากการวิจัยของนักสถิติในสมัยก่อนที่ทดลองสุ่มตัวอย่าง[1] ที่มีขนาดตัวอย่างเท่ากันหลายๆ ครั้ง แต่ละครั้งก็จะได้ค่าถัวเฉลี่ยของตัวอย่าง ($\bar{X}$) มาหนึ่งตัว และพบว่าเมื่อขนาดของตัวอย่าง ($n$) ใหญ่ขึ้นจะทำให้การกระจายของค่าถัวเฉลี่ยของตัวอย่าง ($\bar{X}$) จากการสุ่มตัวอย่างเข้าใกล้การแจกแจงแบบปกติ (Normal Distribution) ไม่ว่าประชากรจะมีการแจกแจงแบบใดก็ตาม โดยค่าที่คาดไว้ (Expected Value) ของค่าถัวเฉลี่ยของตัวอย่าง ($E(\bar{X})$) เท่ากับค่าถัวเฉลี่ยของประชากร ($\mu$) และมีค่าความแปรปรวนของค่าถัวเฉลี่ยของตัวอย่าง ($\sigma_{\bar{X}}^2$) เท่ากับค่าความแปรปรวนของประชากร ($\sigma_X^2$) หารด้วยขนาดของตัวอย่าง ($\mu$)

เพื่อความเข้าใจที่ตรงกันขอแยกประเภทของการกระจายที่เกี่ยวข้องดังนี้

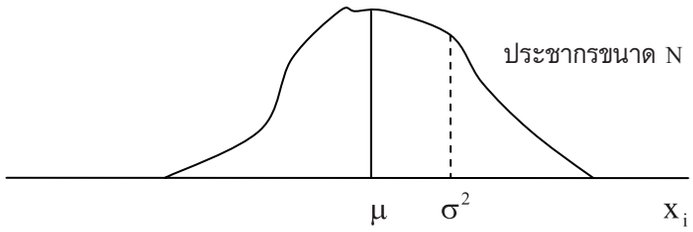
**1. การกระจายของประชากร (Population)** ขนาด N ตาม Central Limit Theorem จะมีการกระจายแบบใดก็ได้ ดังรูปต่อไปนี้ (Zinkmund, 2003, p.459)

ในทางสถิติอาจสรุปภาพรวมของประชากรไม่ว่าประชากรจะมีการกระจายแบบใดด้วยค่าบางค่าที่เรียกว่า Population Parameters ซึ่งสามารถอธิบายลักษณะภาพรวมของประชากรได้อย่างกะทัดรัด โดยทั่วไปจะมีอาทิเช่น ค่าถัวเฉลี่ยของประชากร ($\mu$) และค่าความแปรปรวนของประชากร ($\sigma^2$) เป็นต้น



**รูปที่ 1**: การกระจายของประชากรรูปแบบต่างๆ ทั้งแบบปกติและไม่ปกติ

[1] ส่วนคำว่าตัวอย่างในที่นี้ หมายความว่า สุ่มที่มาจากภาษาอังกฤษคือ Random ไม่ใช่การพูดถึงการเลือกตัวอย่างโดยทั่วไปซึ่งมีทั้งที่สุ่มและไม่สุ่มแต่มักมีผู้ใช้คำสุดรวมๆ กันว่าสุ่มตัวอย่าง แทนที่จะเรียกว่าการเลือกตัวอย่างซึ่งเป็นคำกว้างๆ และดูจะเหมาะสมกว่าและสามารถใช้ครอบคลุมได้ทั้งการเลือกตัวอย่างแบบสุ่มและไม่สุ่ม

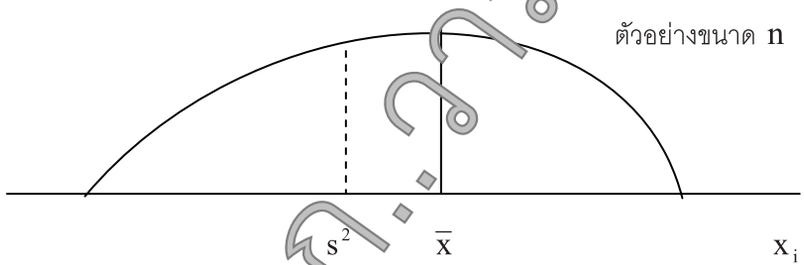



**รูปที่ 2**: การกระจายของประชากรและ Population Parameters

ค่าถัวเฉลี่ยของประชากร, $\mu = \frac{\sum_{i=1}^{N}(x_i)}{N}$

และมีความแปรปรวนของประชากร, $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$

**2. การกระจายของตัวอย่าง (Sample)** ขนาด $n$ ซึ่งนักสถิตินิยมสรุปให้เห็นภาพรวมด้วยค่า Sample statistics ได้แก่ค่าถัวเฉลี่ยของตัวอย่าง ($\bar{x}$) และค่าความแปรปรวนของตัวอย่าง ($s^2$) ทั้งนี้ตาม Central Limit Theorem ได้กล่าวบอกให้ทราบว่าการกระจายของตัวอย่างเป็นรูปร่างแบบเดียวกัน



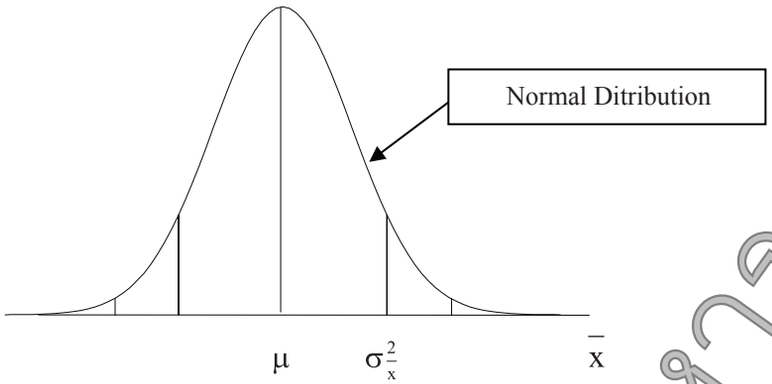

**รูปที่ 3**: การกระจายของกลุ่มตัวอย่างและ Sample Statistics

ค่าถัวเฉลี่ยของตัวอย่าง, $\bar{x} = \frac{\sum_{i=1}^{n}(x_i)}{n}$

และค่าความแปรปรวนของตัวอย่าง, $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

**3. การกระจายของค่าถัวเฉลี่ยของตัวอย่าง (Sample Mean, $\bar{x}$)** เมื่อมีการเลือกตัวอย่างที่มีขนาดเท่ากันซ้ำแล้วซ้ำอีกตามที่นักสถิติได้พิสูจน์ไว้แล้ว เป็นสิ่งที่ Central Limit Theorem ระบุไว้ว่าถ้าขนาดของตัวอย่าง ($n$) ใหญ่พอจะทำให้การกระจายของค่าถัวเฉลี่ยของตัวอย่างเข้าใกล้การกระจายแบบปกติ (Normal Distribution) โดยมีค่าที่คาดไว้ (Expected Value)[2] ของค่าถัวเฉลี่ยของตัวอย่าง ($E(\bar{x})$) เท่ากับค่าถัวเฉลี่ยของประชากร ($\mu$) และมีค่าความแปรปรวนของค่าถัวเฉลี่ยของตัวอย่างเท่ากับค่าความแปรปรวนของประชากร ($\sigma^2$) หารด้วยขนาดของตัวอย่าง ($n$)

[2] หรืออาจถือเสมือนหนึ่งว่าเป็นค่าถัวเฉลี่ยของค่าถัวเฉลี่ยของตัวอย่างอีกทีหนึ่ง




**รูปที่ 4**: การกระจายของค่าถัวเฉลี่ยของกลุ่มตัวอย่าง

ค่าที่คาดไว้ของค่าถัวเฉลี่ยของตัวอย่าง, $E(\bar{x}) = \mu$

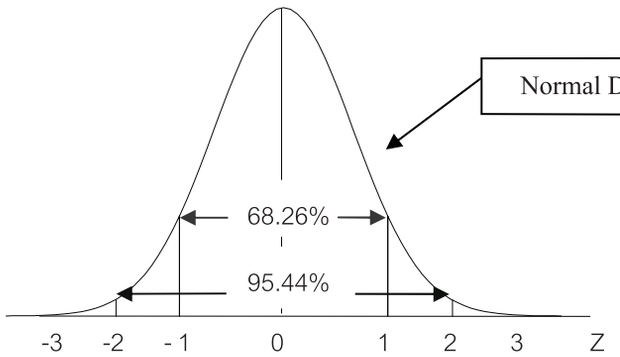
และมีค่าความแปรปรวนของค่าถัวเฉลี่ยของตัวอย่าง, $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

หรือ $= \frac{s^2}{n}$ ถ้าไม่ทราบค่า $\sigma^2$

จาก Central Limit Theorem นี้เอง มีข้อสังเกตว่าการที่จะมีการกระจายแบบปกติได้ ตัวแปรจะต้องเป็นตัวแปรต่อเนื่อง (Continuous Variable) และการที่ตัวแปรจะเป็นตัวแปรต่อเนื่องได้ก็หมายความว่า ณ จุดที่ $x_i = a$ จะหาค่าความน่าจะเป็นของตัวแปรไม่ได้ นั่นคือ Probability ( $x_i = a$) = 0 การจะหาค่าความน่าจะเป็นของตัวแปรต่อเนื่องจะต้องหาค่าความน่าจะเป็นระหว่างค่า $x_i$ ที่มีค่าระหว่าง 2 จุด เช่น Probability ($a < x_i < b$) = $\int_a^b f(x)\,dx$ = ซึ่งมากกว่าศูนย์ (Lindgren, 1968, p.67) ดังนั้น หากมีตัวแปรที่ไม่ต่อเนื่อง (Discrete Variable) การที่จะบอกว่ามีการกระจายแบบปกติแล้วหาค่าถัวเฉลี่ยและค่าความแปรปรวนตามสูตรที่กล่าวมาแล้วจึงทำไม่ได้ ดังนั้นสูตรในการหาขนาดของตัวอย่างจึงมีที่มาจาก Central Limit Theorem นั่นเอง

จากความรู้เรื่องการกระจายแบบปกติ (Normal Distribution) ทำให้ทราบว่าถ้าตัวแปรใดมีการกระจายแบบปกติก็จะสามารถแปลงให้อยู่ในรูปของคะแนนมาตรฐาน (Z-Scores)[3] ที่มีการกระจายแบบปกติ (Standardized Normal Distribution) ซึ่งมีค่าถัวเฉลี่ย (Mean) เท่ากับศูนย์ และค่าความคลาดเคลื่อนมาตรฐาน (Standard Deviation) เท่ากับ ±1 เช่นเดียวกับค่าความแปรปรวน (Variance) ซึ่งเท่ากับหนึ่งเช่นกัน และภายในช่วงของ Z-Scores ที่เท่ากับ 10 ±1 จะมีความน่าจะเป็นคงที่ซึ่งเท่ากับ 68.26% ส่วนภายในช่วงของ Z-Scores ที่เท่ากับ ±2 จะมีค่าความน่าจะเป็นเท่ากับ 95.44% และภายในช่วงของ Z-Scores ที่เท่ากับ ±3 จะมีค่าความน่าจะเป็นที่เท่ากับ 99.73% ไม่ว่าขนาดของตัวอย่างจะเป็นเท่าไรก็ตาม ตราบใดที่ขนาดของตัวอย่างนั้นใหญ่พอและที่ใหญ่พอก็แค่ 30 ตัวอย่างขึ้นไปก็สามารถใช้ Central Limit Theorem ได้แล้ว แต่เรามักจะเห็นการใช้ระดับความน่าจะเป็นที่เป็นตัวเลขกลมๆ คือ 95% ซึ่งเป็นขนาดของพื้นที่ของความน่าจะเป็นภายในช่วงของ Z-Scores ที่เท่ากับ ±1.96 นั่นเอง

[3] ต่อจากนี้เพื่อความสะดวกจะขอเรียกคะแนนมาตรฐานว่า Z-Scores หรือเรียกสั้นๆ ว่า Z แทนที่จะเรียกว่าคะแนนมาตรฐาน




**รูปที่ 5**: การกระจายแบบปกติที่แปลงให้อยู่ในรูปคะแนนมาตรฐาน

จากความรู้เรื่อง Central Limit Theorem และ Normal Distribution ทำให้ทราบว่าจากการสุ่มตัวอย่างเพียงครั้งเดียว หากขนาดของตัวอย่าง ($n$) มีค่าตั้งแต่ 30 ตัวอย่างขึ้นไป ค่าถัวเฉลี่ยของตัวอย่าง ($\bar{x}$) ไม่ควรอยู่ห่างจากค่าถัวเฉลี่ยของประชากร ($\mu$) เกินกว่าค่า Z ที่ ±1.96 คูณด้วยค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย ($\sigma_{\bar{x}}$) ด้วยระดับความเชื่อมั่น 95%

ดังนั้น หากระบุให้ $Z_\alpha$ สัมพันธ์กับระดับความเชื่อมั่น (1–$\alpha$) จะได้

$$\mu - Z_\alpha \sigma_{\bar{x}} \leq \bar{x} \leq \mu + Z_\alpha \sigma_{\bar{x}}$$

$$\mu - Z_\alpha \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$- Z_\alpha \frac{\sigma}{\sqrt{n}} \leq \bar{x} - \mu \leq + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

หรือ $$\left(\bar{x} - \mu\right)^2 = \frac{Z_\alpha^{\,2} \sigma^2}{n}$$

และ $$n = \frac{Z_\alpha^{\,2} \sigma^2}{\left(\bar{x} - \mu\right)^2}$$

เมื่อ $Z_\alpha$ เป็นค่า Z-scores ซึ่งได้จากการกำหนดระดับความเชื่อมั่นที่ (1–$\alpha$) อาทิเช่น $Z_{.05}$ = 1.96 หมายความว่ามีความเชื่อมั่น 95% ว่าค่า ($\bar{x}$) ควรจะอยู่ห่างจาก ($\mu$) ไม่เกิน ±1.96 เท่าของ $\sigma_{\bar{x}}$ ทั้งนี้แม้จะไม่ทราบค่าถัวเฉลี่ยของประชากร และการสุ่มตัวอย่างจะทำเพียงครั้งเดียว ซึ่งทำให้ได้ค่าถัวเฉลี่ยของกลุ่มตัวอย่าง ($\bar{x}$) เพียงค่าเดียว แต่จาก Central Limit Theorem ทำให้ทราบว่า 95 ครั้งจาก 100 ครั้งที่ทำการสุ่มตัวอย่าง ค่าถัวเฉลี่ยของประชากรควรตกอยู่ในช่วง $\bar{x}$ - (1.96) $\sigma_{\bar{x}}$ ถึง $\bar{x}$ + (1.96) $\sigma_{\bar{x}}$ นั่นเอง ซึ่งแน่นอนย่อมมีโอกาสพลาดประมาณ 5 ครั้งจาก 100 ครั้ง

$\sigma^2$ เป็นค่าความแปรปรวนของประชากร และหากไม่ทราบค่านี้สามารถใช้ค่าความแปรปรวนของตัวอย่าง ($s^2$) แทนที่ได้

$n$ เป็นขนาดของตัวอย่างซึ่งอย่างน้อยต้องไม่ต่ำกว่า 30 ตัวอย่าง เพื่อให้ค่าถัวเฉลี่ยของตัวอย่างมีการกระจายแบบปกติ

$\left(\bar{x} - \mu\right)$ เป็นค่าผลต่างระหว่างค่าถัวเฉลี่ยของตัวอย่างซึ่งเกิดขึ้นจากการสุ่มตัวอย่าง กับค่าถัวเฉลี่ยของประชากร เพราะหากเลือกตัวอย่างจากทั้งหมดของประชากรโดยที่การวัดทำอย่างถูกต้องค่า $\left(\bar{x} - \mu\right)$ จะต้องเท่ากับศูนย์ ทั้งนี้เพราะจะได้ค่าถัวเฉลี่ยของประชากร ($\mu$) แทนที่จะเป็นค่าถัวเฉลี่ยของตัวอย่าง ($\bar{x}$) แต่โดยทั่วไปขนาดของตัวอย่างจะน้อยกว่าขนาดของ ประชากรทำให้ค่าทั้งสองต่างกัน จึงอาจเรียกค่านี้ว่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (Random Sampling Error) หรือในตำราบางเล่ม (Malhotra, 2009, p.403) เรียกค่านี้ว่าระดับความแม่นยำ (Precision Level) นั่นเอง ถ้าค่าความแม่นยำสูง แสดงว่าค่าความคลาดเคลื่อนจากการสุ่มตัวอย่างมีขนาดเล็ก

และถ้าแทนค่า $\left(\bar{x} - \mu\right) = e$ ซึ่งได้แก่ความคลาดเคลื่อนจากการสุ่มตัวอย่าง จะได้สูตรในการหาขนาดของตัวอย่างในกรณีประชากรมีขนาดใหญ่ (Infinite Population) ดังนี้

$$n = \frac{Z_\alpha{}^2\sigma^2}{e^2} \qquad (1)$$

จากสูตรนี้ขนาดของตัวอย่างจะขึ้นอยู่กับค่า $Z_\alpha$ ที่สัมพันธ์กับระดับความเชื่อมั่น $(1-\alpha)$ และค่าความแปรปรวนของประชากร $(\sigma^2)$ ในทางเดียวกัน แต่ขนาดของตัวอย่างจะสัมพันธ์ในทางผกผันกับค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง $(e)$ ซึ่งคงไม่ใช่มีแต่ยามาเน่เท่านั้นที่จะได้สูตรนี้ ตำราวิจัยตลาดเท่าที่ผู้เขียนเคยใช้มาทุกเล่มก็ได้สูตรนี้เช่นกัน เป็นสูตรตั้งต้น และเชื่อว่านักศึกษาที่เรียนสถิติเบื้องต้นมาแล้วคงพิสูจน์สูตรนี้ได้ไม่ยาก

ภายใต้สูตรนี้มีสมมติฐานว่าประชากรมีขนาดใหญ่ (Infinite Population) ซึ่งในทางปฏิบัติแค่ 500,000 ตัวอย่างขึ้นไปก็น่าจะถือว่าใหญ่ (Infinity) ได้แล้ว แน่นอนเมื่อประชากรมีขนาดใหญ่ควรจะใช้สูตรนี้ แต่เมื่อประชากรมีขนาดเล็กจะต้องปรับสูตรอีกเล็กน้อย ในกรณีที่ประชากรมีขนาดเล็กค่าความแปรปรวนจะเล็กลงไปด้วย แต่ไม่เล็กลงตามสัดส่วนของประชากรที่ลดลง การคำนวณค่าความแปรปรวนของประชากรที่มีขนาดเล็กลง (Finite Population) จะต้องปรับด้วยตัวปรับค่าประชากรขนาดเล็ก (Finite Population Correction Factor; FPC) ซึ่งได้แก่ $\frac{N-n}{N-1}$ เมื่อ $N$ คือขนาดของประชากร และ $n$ คือขนาดของตัวอย่าง ซึ่งแน่นอน $\frac{N-n}{N-1}$ จะต้องน้อยกว่าหนึ่งตราบใดที่ $n$ มากกว่าศูนย์ นั่นคือความแปรปรวนของประชากรที่มีขนาดเล็กลงจะเล็กกว่าความแปรปรวนของประชากรที่มีขนาดใหญ่ด้วยตัวคูณ $\frac{N-n}{N-1}$ นั่นเอง

ดังนั้นในกรณีนี้ ค่าความแปรปรวนของประชากรที่มีขนาดเล็ก (Finite Population) จะเป็น $\sigma^2\frac{N-n}{N-1}$

และค่าความแปรปรวนของค่าถัวเฉลี่ยของตัวอย่าง

$$\sigma^2_{\bar{x}} = \left(\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}\right)$$

ค่าจำกัดสองของความคลาดเคลื่อนจากการสุ่มตัวอย่าง จะเป็น

$$(\bar{x}-\mu)^2 \quad \text{หรือ} \quad e^2 = \frac{Z_\alpha{}^2\sigma^2}{n}\cdot\frac{N-n}{N-1}$$

$$n(N-1)e^2 = Z_\alpha{}^2\sigma^2N - Z_\alpha{}^2\sigma^2n$$

$$n\left[(N-1)e^2 + Z_\alpha{}^2\sigma^2\right] = Z_\alpha{}^2\sigma^2N$$

ดังนั้นในกรณีประชากรที่มีขนาดเล็ก (Finite Population) ขนาดของตัวอย่างจะเป็น

$$n = \frac{Z_\alpha{}^2\sigma^2N}{(N-1)e^2 + Z_\alpha{}^2\sigma^2} \qquad (2)$$

ซึ่งหมายความว่านอกเหนือจากค่าระดับความเชื่อมั่น $(1-\alpha)$ ซึ่งเป็นตัวกำหนดค่า $Z_\alpha$ และค่าความแปรปรวนของประชากร และค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง $(e)$ ซึ่งอยู่ทางด้านขวาของสมการที่ (1) ที่มีอิทธิพลในการกำหนดขนาดของตัวอย่างแล้ว ขนาดของประชากร $(N)$ จะมีส่วนในการกำหนดขนาดของตัวอย่างในกรณีที่ประชากรมีขนาดเล็กด้วย

ตามสูตรข้างต้นก็ยังไม่เข้าถึงสูตรของยามาเน่ที่นักศึกษาและนักวิจัยไทยชอบใช้ เพราะในกรณีของสูตรข้างต้นมาตรวัดจำเป็นต้องมีแค่ 2 ค่า จะมีค่าเป็นอย่างไรก็ได้ ไม่จำกัดหน่วยที่ใช้ในการวัด แต่จะต้องเป็นมาตรวัดแบบช่วง (Interval Scale) หรือมาตรวัดแบบสัดส่วน (Ratio Scale) เท่านั้น เช่นเดียวกับค่าความแปรปรวนของประชากรไม่ได้จำกัดว่าจะเกิน 0.25 ไม่ได้ดังเช่นการกระจายของตัวแปรแบบสัดส่วน (Proportion)

## การกระจายของการสุ่มตัวอย่างกรณีสัดส่วน

ในกรณีที่มาตรวัดมีแค่ 2 ค่า สามารถแสดงให้เห็นในรูปของสัดส่วนซึ่งมีค่าความน่าจะเป็นรวมกันเท่ากับหนึ่ง ดังนั้นแทนที่ค่าเฉลี่ยของตัวอย่างจะเป็น $\bar{x}$ แต่จะมีค่าเท่ากับค่าของสัดส่วนของตัวอย่างซึ่งเท่ากับ $p$ แทนจาก Central Limit Theorem ค่าการกระจายของสัดส่วนของตัวอย่าง $(p)$ สำหรับตัวอย่างขนาด $n$ จะมีค่าที่คาดไว้ของสัดส่วนของตัวอย่าง $(E(p))$ และค่าความแปรปรวนของสัดส่วนของตัวอย่าง $(\sigma^2{}_p)$ ดังนี้

***อันเนื่องมาแต่****สูตรของยามาเน่*

$$E(p) = \pi$$

$$\sigma^2_p = \frac{\pi(1-\pi)}{n}$$ กรณีประชากรมีขนาดใหญ่ (Infinite Population)

หรือ $$\sigma^2_p = \frac{\pi(1-\pi)}{n}\frac{N-n}{N-1}$$ กรณีประชากรมีขนาดเล็ก (Finite Population)

เมื่อ $\pi$ คือ สัดส่วนของประชากร

$\sigma^2_p$ คือ ความแปรปรวนของสัดส่วนของตัวอย่าง

$n$ คือ ขนาดของตัวอย่าง

$N$ คือ ขนาดของประชากร

ซึ่งสามารถหาค่าขนาดของตัวอย่างได้เช่นเดียวกันกับในกรณีตัวแปรเป็นตัวแปรต่อเนื่อง (Continuous Variable) แม้จะสุ่มตัวอย่างเพียงครั้งเดียวและได้สัดส่วนของตัวอย่างเท่ากับ $p$ จาก Central Limit Theorem จะสรุปได้ว่าสัดส่วนของประชากร ($\pi$) จะอยู่ระหว่างค่า $p \pm Z_\alpha \sigma_P$

$$p - Z_\alpha \sigma_P \leq \pi \leq p + Z_\alpha \sigma_P$$

เมื่อ $p$ คือ สัดส่วนของตัวอย่าง

$\pi$ คือ สัดส่วนของประชากร

$\sigma_P$ คือ ค่าความคลาดเคลื่อนมาตรฐานของสัดส่วนของตัวอย่าง

$Z_\alpha$ คือ ค่า Z-scores ณ ระดับความเชื่อมั่น $(1-\alpha)$

หากทราบค่าสัดส่วนของประชากร ($\pi$), $\sigma^2_p = \frac{\pi(1-\pi)}{n}$

ค่าความคลาดเคลื่อนจากการสุ่มตัวอย่างของสัดส่วน จะเท่ากับ $(p-\pi)$ หรือสมมติให้เท่ากับ $e$

จะได้ $(p-\pi)^2 = e^2 = Z_\alpha{}^2 \sigma^2_p$

$$e^2 = Z_\alpha{}^2 \frac{\pi(1-\pi)}{n}$$

และ

$$\boxed{n = \frac{Z_\alpha{}^2 \pi(1-\pi)}{e^2}} \quad (3)$$

หรือหากไม่ทราบสัดส่วนของประชากร ($\pi$) ก็ใช้สัดส่วนของตัวอย่าง ($p$) แทน

ดังนั้น $$n = \frac{Z_\alpha{}^2 p(1-p)}{e^2}$$ กรณี Infinite Population

หรือ $$n = \frac{Z_\alpha{}^2 pq}{e^2}$$ เมื่อ $q = 1-p$

# ที่มาของสูตรของยามาเน่

จาก $e^2 = \frac{Z_\alpha{}^2 \pi(1-\pi)}{n}$ ในกรณีของสัดส่วนเมื่อประชากรมีขนาดใหญ่

ถ้าประชากรมีขนาดเล็ก (Finite Population) ค่าความแปรปรวนของสัดส่วนของตัวอย่างจะปรับด้วยตัวปรับค่าประชากรขนาดเล็ก (Finite Population Correction Factor; FPC) สมการข้างต้นควรจะเป็น

$$e^2 = \frac{Z_\alpha{}^2 \pi(1-\pi)}{n} \cdot \frac{N-n}{N-1} \qquad (4)$$

แต่ครั้งนี้ แทนที่ยามาเน่จะใช้ FPC ซึ่งเท่ากับ $\frac{N-n}{N-1}$ ยามาเน่กลับใช้ $\frac{N-n}{N}$ แทน ทั้งๆ ที่ก่อนหน้านี้ยามาเน่ใช้ $\frac{N-n}{N-1}$ ทั้งนี้อาจเป็นเพราะในกรณีที่ขนาดของประชากร N มีขนาดใหญ่มากสัดส่วนทั้งสองแทบจะไม่ต่างกัน จึงมีผลให้ยามาเน่ได้สมการนี้แทน

$$e^2 = \frac{Z_\alpha{}^2 \pi(1-\pi)}{n} \cdot \frac{N-n}{N} \quad \text{(Yamane, 1973, p.729)} \qquad (5)$$

$$nNe^2 = Z_\alpha{}^2 \pi(1-\pi)N - Z_\alpha{}^2 \pi(1-\pi)n$$

$$n\left[Ne^2 + Z_\alpha{}^2 \pi(1-\pi)\right] = Z_\alpha{}^2 \pi(1-\pi)N$$

$$\text{และ} \quad n = \frac{Z_\alpha{}^2 \pi(1-\pi)N}{Z_\alpha{}^2 \pi(1-\pi) + Ne^2}$$

และในกรณีของสัดส่วนค่า $\sigma^2_p$ จะสูงสุดเมื่อ $\pi = 5$ และ $1 - \pi = 0.5$ ซึ่งจะทำให้ค่า $n$ สูงสุดเช่นกัน

ยามาเน่แทนค่า $\pi = 5$ และ $1 - \pi = 0.5$

$$\text{ค่า} \quad n = \frac{Z_\alpha{}^2 (0.5)^2 N}{Z_\alpha{}^2 (0.5)^2 + Ne^2}$$

เมื่อ n คือ ขนาดของตัวอย่าง

N คือ ขนาดของประชากร

e คือ ความคลาดเคลื่อนจากการสุ่มตัวอย่าง

และถ้าให้ Z = 1.96 คือ ค่า Z-scores ณ ระดับความเชื่อมั่น 95%

$$\text{ดังนั้น} \quad n = \frac{(1.96)^2 (0.5)^2 N}{(1.96)^2 (0.5)^2 + Ne^2}$$

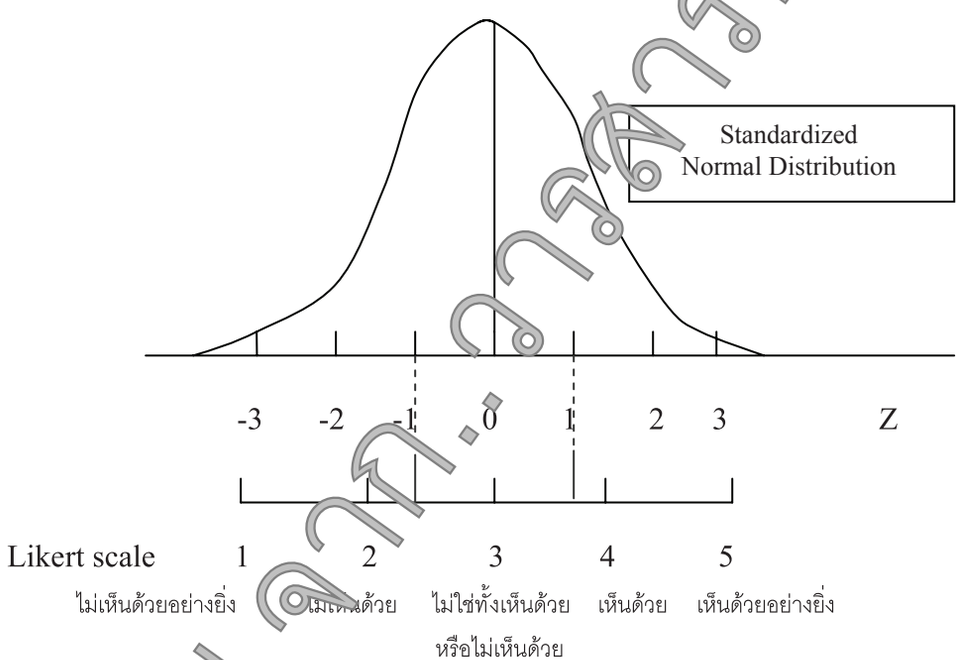
สำหรับค่า Z ที่เท่ากับ 1.96 ยามาเน่ ใช้เลขตัวกลม คือ 2 แทน จะได้

$$n = \frac{(2)^2 (0.5)^2 N}{(2)^2 (0.5)^2 + Ne^2}$$

$$\text{หรือ} \quad \boxed{n = \frac{N}{1 + Ne^2}} \qquad (6)$$

***อันเนื่องมาแต่****สูตรของยามาเน่*

นั่นหมายความว่ายามาเน่มีข้อสมมติอยู่ 2 ประการ คือ สมมติว่าสัดส่วน (Proportion) เท่ากับ 0.5 ซึ่งเป็นค่าที่ทำให้ $\sigma^2_p$ มีค่าสูงสุดที่ 0.25 และขนาดของตัวอย่าง (n) มีค่าสูงสุด และหากสมมติระดับความเชื่อมั่นที่ 95.44% ซึ่งให้ค่า Z เท่ากับ 2 จะได้สูตรที่นักศึกษาไทยชอบใช้แล้วเรียกว่าสูตรของยามาเน่ ซึ่งได้แก่สมการที่ (6) ทั้งๆ ที่มาตรวัดที่ท่านเหล่านั้นใช้ไม่ได้มีแค่ 2 ค่า จึงไม่อาจแปลงเป็นค่าสัดส่วน (Proportion) ได้ โดยเฉพาะนักศึกษาปริญญาโทที่นิยมวัดตัวแปรประเภทที่ไม่สามารถสังเกตได้ (Unobserved หรือ Latent Variable) โดยใช้มาตรวัดลิเคิร์ทแบบห้าตัวเลือก (5 Point Likert Scale) เป็นส่วนใหญ่ หรือใช้มาตรวัดที่เป็นตัวแปรต่อเนื่องที่มีมากกว่า 2 ค่า หรือวัตถุประสงค์ของงานวิจัยไม่ได้ต้องการพยากรณ์สัดส่วนของตัวแปรซึ่งมีแค่ 2 ค่าแต่อย่างใด ผู้เขียนจึงไม่เข้าใจว่าท่านเหล่านั้นจะนำสูตรของยามาเน่นี้มาใช้ได้อย่างไร เมื่อพิจารณามาตรวัดลิเคิร์ทแบบห้าตัวเลือกค่าต่ำสุดที่เป็นไปได้จะเป็น (1) และค่าสูงสุดที่เป็นไปได้จะเป็น (5) หากลองสมมติต่อว่าค่าตัวแปรที่ใช้มาตรวัดนี้มีการกระจายแบบปกติก็หมายความว่าสามารถแปลงมาตรวัดแบบลิเคิร์ทให้อยู่ในรูปของคะแนนมาตรฐาน (Z-Scores) ที่มีการกระจายแบบปกติ (Standardized Normal Distribution) ได้ ดังนั้นด้วยระดับความเชื่อมั่น 99.73% สามารถแบ่งค่า Z ออกได้มีค่าความคลาดเคลื่อนมาตรฐาน (Standard Deviation) ข้างละ 3 ส่วนโดยมีค่ากลางเท่ากับศูนย์ ในขณะที่มาตรวัดลิเคิร์ทแบบห้าตัวเลือกที่สัมพันธ์กับค่า Z-Scores จะมีค่ากลางที่ (3) และอาจแสดงได้ดังรูปต่อไปนี้




**รูปที่ 6**: คะแนนมาตรฐานของการกระจายแบบปกติเทียบกับคะแนนมาตรวัดลิเคิร์ทแบบห้าตัวเลือก

ลองคิดดูคร่าวๆ แล้ว หากตัวอย่างมีการกระจายแบบปกติ ค่าความแปรปรวนของมาตรวัดลิเคิร์ทแบบห้าตัวเลือกไม่น่าจะเกินหนึ่งภายในค่าความคลาดเคลื่อนมาตรฐาน ±1 แต่ถ้าการกระจายไม่เป็นแบบปกติค่านี้อาจเกินหนึ่งได้ จึงไม่มีเหตุผลใดเลยที่ค่าความแปรปรวนสูงสุดจะเท่ากับ $(0.5)^2$ หรือ 0.25 ตามสูตรของยามาเน่ ดังนั้นผู้เขียนจึงค่อนข้างไม่เห็นด้วยหากนักศึกษาใช้มาตรวัดทัศนคติลิเคิร์ทแบบห้าตัวเลือกเป็นส่วนใหญ่ แต่กลับนำสูตรของยามาเน่ตามสมการที่ (6) มาใช้เพียงเพื่อหลีกเลี่ยงการหาค่าความแปรปรวนของประชากรเท่านั้น เหตุที่ไม่เห็นด้วยเพราะข้อสมมติไม่ได้เป็นไปตามที่ยามาเน่ใช้ในการคำนวณสูตรข้างต้น สูตรของยามาเน่ไม่ใช่จะไม่มีความแปรปรวนติดอยู่ แต่ยามาเน่ได้แทนค่าความแปรปรวนด้วยตัวเลขที่คิดว่าเหมาะสมกับสถานการณ์การกระจายของสัดส่วนของตัวอย่าง (p) แต่ไม่เหมาะกับการกระจายของค่าเฉลี่ยของตัวอย่าง ($\bar{x}$) ที่เป็นตัวแปรต่อเนื่องเลย หากจะหาค่าความแปรปรวนของกลุ่มตัวอย่างต้องดูด้วยว่ามาตรวัดเป็นแบบใด ควรพยายามหาค่าความแปรปรวนตามมาตรวัดนั้นมาใช้ในการกำหนดขนาดของตัวอย่าง ซึ่งไม่ได้ยากอย่างที่นักศึกษากลัวเลย

นอกจากสูตรของยามาเน่ที่กล่าวข้างต้นแล้ว ยังมีอีกสูตรหนึ่งซึ่งนักศึกษานิยมอ้างถึง ได้แก่

$$n = \left( \frac{1}{4e^2 / Z_\alpha^{\ 2}} \right)$$

ซึ่งก็ไม่ได้ต่างจากสูตรของยามาเน่เท่าใด เพียงแต่มีที่มาจากประชากรที่มีขนาดใหญ่มากๆ (Infinite Population) แทนที่จะเป็นการสุ่มตัวอย่างจากประชากรที่มีขนาดเล็กลงมา (Finite Population) ที่มาของสูตรนี้จึงมาจากกรณีที่มาตรวัดมี 2 ค่าเช่นเดียวกัน และสามารถคำนวณสูตรให้ได้ค่าขนาดของตัวอย่างตามสมการที่ (3) ดังนี้

$$n = \frac{Z_\alpha^{\ 2} \pi(1 - \pi)}{e^2}$$

หากแทนค่า $\pi = 5$ และ $(1 - \pi) = 0.5$ ตามที่ยามาเน่เคยใช้ ก็จะได้

$$n = \frac{Z_\alpha^{\ 2} (0.5)^2}{e^2}$$

หรือ

$$n = \left( \frac{1}{4e^2 / Z_\alpha^{\ 2}} \right) \qquad (7)$$

## จะหาค่าความแปรปรวนของกลุ่มตัวอย่างก่อนเริ่มทำงานวิจัยได้อย่างไร

หากจะกลับไปตั้งต้นที่สูตรแรก ในกรณีที่มาตรวัดเป็นแบบช่วง (Interval Scale) หรือแบบสัดส่วน (Ratio Scale) ซึ่งถือเป็นตัวแปรต่อเนื่อง (Continuous Variable) ในกรณีที่ใช้การสุ่มตัวอย่างหรือมีการใช้การเลือกตัวอย่างแบบใช้ความน่าจะเป็น (Probability Sampling) และประชากรมีขนาดใหญ่ จะได้สูตรในการหาขนาดของตัวอย่าง ดังนี้

จาก (1) $n = \frac{Z_\alpha^{\ 2} \sigma^2}{e^2}$ หรือ $n = \frac{Z_\alpha^{\ 2} s^2}{e^2}$

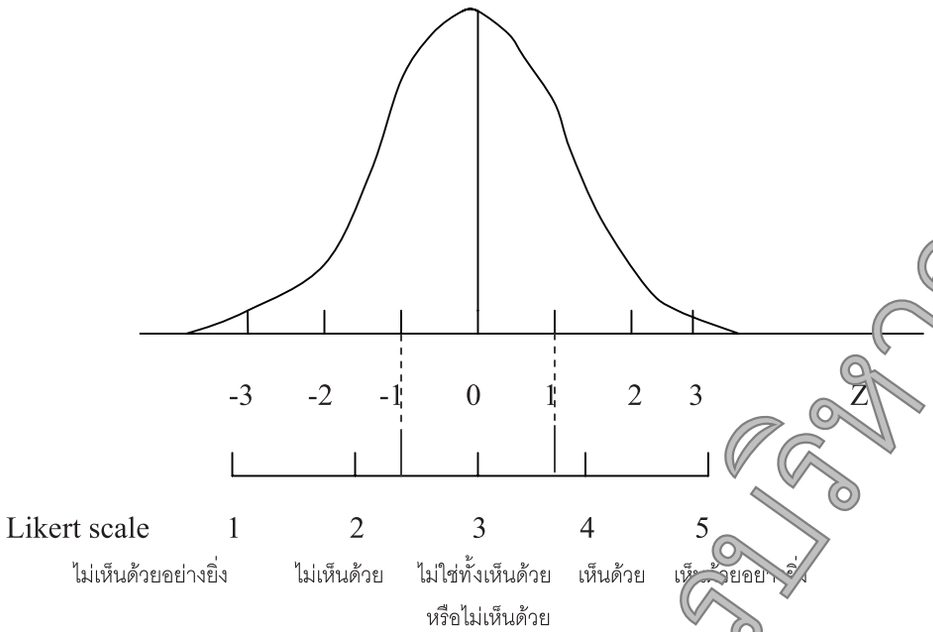
เมื่อ $n$ เป็นขนาดของตัวอย่าง

$Z_\alpha$ เป็นค่า Z-Scores ซึ่งสัมพันธ์กับระดับความเชื่อมั่น $(1 - \alpha)$

$\sigma^2$ เป็นค่าความแปรปรวนของประชากร และหากไม่ทราบค่านี้ สามารถใช้ค่าความแปรปรวนของกลุ่ม ตัวอย่าง ($s^2$) แทนที่ได้

$e$ เป็นค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง $\left(\bar{x} - \mu\right)$ เมื่อ $\bar{x}$ เป็นค่าถัวเฉลี่ยของกลุ่มตัวอย่างและ $\mu$ เป็นค่าถัวเฉลี่ยของประชากร

เหตุผลที่นักศึกษาไทยพยายามหลีกเลี่ยงที่จะใช้สูตรนี้คงเป็นเพราะไม่ทราบค่าจะหาค่าความแปรปรวนของประชากร ($\sigma^2$) หรือของกลุ่มตัวอย่าง ($s^2$) ก่อนที่จะมีการเลือกตัวอย่างจริงได้อย่างไร นอกจากนั้นก็ไม่รู้ว่าจะกำหนดค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) ให้เป็นเท่าไร ลองมาดูที่ค่า e กันก่อน จะเห็นว่าค่านี้จะขึ้นอยู่กับหน่วยที่ใช้ในการวัดค่าถัวเฉลี่ยของกลุ่มตัวอย่าง ($\bar{x}$) หรือค่าถัวเฉลี่ยของประชากร ($\mu$) ซึ่งก็มีหลากหลายแล้วแต่หน่วยวัดของตัวแปรในแบบสอบถามแต่ละข้อ ดังนั้นการที่จะหาค่า ซึ่งเป็นค่าสมบูรณ์ (Absolute Value) และขึ้นอยู่กับหน่วยวัดของแต่ละตัวแปรก่อนเริ่มการเก็บข้อมูลจึงค่อนข้างที่จะเป็นปัญหาของนักศึกษาส่วนใหญ่ จากประสบการณ์การสอนของผู้เขียนพบว่าในกรณีของนักศึกษาปริญญาตรีอาจทำได้ลำบาก เนื่องจากงานวิจัยของนักศึกษาปริญญาตรีมักเป็นงานวิจัยเชิงประยุกต์ (Applied Research) ซึ่งแบบสอบถามแต่ละข้อก็มีมาตรวัดที่หลากหลายเกินกว่าที่จะสรุปภาพรวมได้ว่าควรมีมาตรวัดเป็นแบบใด ดังนั้นเพียงแต่จะหาค่า e ก็ค่อนข้างมีปัญหาแล้ว แต่การที่จะหยิบตัวเลขตัวใดตัวหนึ่งมาใส่ก็ไม่ได้ผิดอะไรนัก แต่ต้องยอมให้ค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) เมื่อเทียบเป็นร้อยละของค่าถัวเฉลี่ยของกลุ่มตัวอย่างมีขนาดเล็กใหญ่ตามไปด้วย สำหรับงานวิจัยของนักศึกษาปริญญาโท ซึ่งส่วนใหญ่เป็นงานวิชาการ (Basic Research) โดยเฉพาะนักศึกษาที่ทำงานวิจัยทางสังคมศาสตร์ที่เกี่ยวข้องกับการวัดทัศนคติของกลุ่มเป้าหมาย มักจะต้องวัดตัวแปรที่ไม่สามารถสังเกตได้ (Unobserved Variable หรือ Latent Variable) และมีการใช้มาตรวัดแบบหลายรายการ (Multi-Item Measurement) ซึ่งผู้เขียนพบว่านักศึกษามักใช้มาตรวัดลิเคิร์ทแบบห้าตัวเลือกเป็นส่วนใหญ่ ในกรณีนี้ผู้เขียนมีความเห็นว่าจะสามารถประมาณการค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) ได้ดังนี้




**รูปที่ 7**: คะแนนมาตรฐานที่ ±1 เมื่อแปลงเป็นคะแนนของมาตรวัดลิเคิร์ทแบบห้าตัวเลือก

หากจะพิจารณามาตรวัดลิเคิร์ทแบบห้าตัวเลือก ค่าต่ำสุดที่เป็นไปได้จะเป็น (1) และค่าสูงสุดที่เป็นไปได้จะเท่ากับ(5) หากลองสมมติต่อว่ากลุ่มตัวอย่างมีการกระจายแบบปกติ (Normal Distribution) ค่าถัวเฉลี่ยของกลุ่มตัวอย่างจะเท่ากับ (3) ดังนั้นหากผู้วิจัยยอมให้ความคลาดเคลื่อนจากการสุ่มตัวอย่างมีได้ 5% ของค่าถัวเฉลี่ยจะได้ค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) เท่ากับ 5% ของ 3 หรือ 0.15 อย่างไรก็ตามการจะหวังให้การกระจายของกลุ่มตัวอย่างเป็นแบบปกติ (Normal) คงเป็นเรื่องยาก จากประสบการณ์ของผู้เขียนซึ่งเคยใช้ตัวอย่างถึง 2,000 ตัวอย่างในบางงานวิจัย การกระจายของหลายๆ ตัวแปรที่วัดยังเบ้ซ้ายบ้างเบ้ขวาบ้าง ไม่ได้มีการกระจายแบบปกติแต่อย่างใด สำหรับมาตรวัดแบบ 5 จุด ผู้เขียนลองมาวิเคราะห์ดูแล้ว โอกาสที่ค่าถัวเฉลี่ยของกลุ่มตัวอย่างจะเท่ากับ (1) หรือ (5) แทบไม่มีเลย แต่ถ้าจะได้ค่าถัวเฉลี่ยของกลุ่มตัวอย่างมีค่าค่อนไปทาง (2) หรือ (4) ยังมีโอกาสเป็นไปได้มากกว่า ดังนั้นอาจคำนวณค่าความคลาดเคลื่อนจากการสุ่มตัวอย่างที่ 5% ของค่าถัวเฉลี่ยได้คือ กรณีค่าถัวเฉลี่ยของกลุ่มตัวอย่างเท่ากับ 2 ค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) จะเท่ากับ 5% ของ 2 ซึ่งมีค่า 0.10 และในกรณีที่ค่าถัวเฉลี่ยของกลุ่มตัวอย่างเท่ากับ 4 ค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) จะเท่ากับ 5% ของ 4 ซึ่งมีค่า 0.20 และเมื่อพิจารณาสูตรในการหาขนาดของตัวอย่างข้างต้นจะพบว่าค่าขนาดของตัวอย่าง (n) จะแปรผกผันกับค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) ดังนั้น ถ้า e ยิ่งมีขนาดเล็ก n จะยิ่งมีขนาดใหญ่ หากจะขอยืมแนวคิดอนุรักษ์นิยม (Conservative) ของคุณยามาเน่ที่แทนค่า $\pi$ และ $(1-\pi)$ ด้วย 0.5 ซึ่งเป็นค่าที่ทำให้ค่าความแปรปรวนของสัดส่วนมีค่าสูงสุดที่ 0.25 ในกรณีนี้ก็น่าจะใช้ค่า e = .10 มากกว่าที่จะใช้ค่า e ที่เท่ากับ .15 หรือ .20 เพราะจะทำให้ขนาดของตัวอย่างซึ่งผกผันกับ e มีค่าสูง

ส่วน $\sigma^2$ นั้นคงต้องใช้ $s^2$ ซึ่งได้จากกลุ่มตัวอย่างแทนที่อยู่แล้ว เพราะปกติก็จะไม่ทราบค่าความแปรปรวนของประชากร ($\sigma^2$) อย่างไรก็ตามแม้จะต้องการใช้ค่าความแปรปรวนของกลุ่มตัวอย่าง ($s^2$) แทน หากยังไม่ได้เริ่มงานวิจัยจะไม่ทราบค่านี้เช่นกัน หากจะย้อนกลับไปดูข้อมูลงานวิจัยเก่าที่ใช้มาตรวัดแบบ 5 จุด ค่า $s^2$ ที่ได้จากงานวิจัยจะหลากหลายมาก แต่ผู้วิจัยคงพอสรุปค่า $s^2$ ส่วนใหญ่ที่ได้จากงานวิจัยที่ผ่านมาได้ แม้ว่างานวิจัยที่ต่างกันจะให้ค่า $s^2$ ที่ต่างกันด้วย หากผู้วิจัยมีการอ้างอิงที่มาของค่า $s^2$ ที่ใช้ว่าได้มาจากงานของนักวิจัยใดน่าจะใช้ค่า $s^2$ นั้นเป็นจุดตั้งต้นในการคำนวณขนาดของตัวอย่างที่ยอมรับได้ สำหรับงานวิจัยของผู้วิจัยที่ผ่านมาที่ใช้มาตรวัดแบบนี้จะขึ้นอยู่กับกลุ่มเป้าหมายว่าเป็นใคร ซึ่งพอสรุปได้ว่าค่า $s^2$ ส่วนใหญ่จะอยู่ระหว่าง 1 ถึง 2 หากจะมองกลุ่มตัวอย่างที่มีการกระจายแบบปกติแล้วจะสามารถแบ่งมาตรวัดลิเคิร์ทแบบห้าตัวเลือกออกได้เป็น 4 ช่วงดังรูปที่ 7 โดยมีค่ากลางอยู่ที่ (3) ดังนั้นทางด้านซ้ายของค่า (3) จะมีอยู่ 2 ช่วง และทางด้านขวาจะมี 2 ช่วง หากจะแปลงมาตรวัดลิเคิร์ทแบบห้าตัวเลือกออกเป็นค่า Z-Scores ที่มีค่ากลางอยู่ที่ 0 และมีค่าความคลาดเคลื่อนมาตรฐาน (Standard Deviation) ที่ ±1 ภายใน Z-Scores ที่เท่ากับ ±3 โอกาสที่ค่าถัวเฉลี่ยของประชากรจะตกอยู่ภายในช่วงนี้ถึง 99.73% ซึ่งหมายความว่าสามารถแบ่งค่า Z ออกเป็นข้างละ 3 ช่วง เท่ากับว่าหากหารช่วงของมาตรวัด (1) ถึง (5) ด้วย 6 ค่า Z ที่เท่ากับ +1 หรือ –1 จะเทียบเป็นมาตรวัดลิเคิร์ทแบบห้าตัวเลือกได้ประมาณ 0.67 หน่วย [4/6 = 0.667] ซึ่งคำนวณเป็นค่าความแปรปรวนของกลุ่มตัวอย่าง ($s^2$) ได้ประมาณ 0.45 หน่วยเท่านั้น [$(0.67)^2$=0.45] อย่างไรก็ตามโอกาสที่ผู้วิจัยจะได้ลักษณะการกระจายของกลุ่มตัวอย่างเป็นแบบปกติมีน้อยมาก ดังนั้นผู้เขียนขอใช้วิจารณญาณที่ได้จากบทสรุปงานวิจัยที่เคยใช้มาตรวัดลิเคิร์ทแบบห้าตัวเลือกในการประมาณการค่าความแปรปรวนของกลุ่มตัวอย่าง ($s^2$) ว่าจะอยู่ระหว่าง 1 ถึง 2 ดังนั้นหากจะใช้ระดับความเชื่อมั่นที่ 95% จะได้ค่า Z = ±1.96 และถ้าจะคิดในแนวอนุรักษ์ (Conservative) คือให้ขนาดของตัวอย่างใหญ่พอที่จะรองรับระดับความเชื่อมั่นที่ต้องการโดยยอมให้ความคลาดเคลื่อนจากการสุ่มตัวอย่างเท่ากับ 5% ของค่าเฉลี่ยของตัวอย่าง ผู้เขียนขอใช้ e = .10 ด้วยเหตุผลที่กล่าวมาแล้ว จะสามารถหาขนาดของตัวอย่างได้ดังนี้

จากสูตร $$n = \frac{Z_\alpha^2 s^2}{e^2}$$

จะได้ $$n = \frac{(1.96)^2(1)}{(0.10)^2} = 384$$ เมื่อ $s^2 = 1$ และ

$$n = \frac{(1.96)^2(2)}{(0.10)^2} = 768$$ เมื่อ $s^2 = 2$

ในกรณีที่งบประมาณในการสุ่มตัวอย่างมีน้อย ถ้ายอมให้ความคลาดเคลื่อนจากการสุ่มตัวอย่างเพิ่มเป็น 10% ของค่าถัวเฉลี่ยของกลุ่มตัวอย่าง[4] เมื่อสมมติให้ค่าถัวเฉลี่ยของกลุ่มตัวอย่างอยู่ใกล้ๆ 2 จะได้ค่า e = .20 ซึ่งจะได้ค่าขนาดของตัวอย่างดังนี้

$$n = \frac{(1.96)^2(1)}{(0.20)^2} = 96$$ เมื่อ $s^2 = 1$ และ

$$n = \frac{(1.96)^2(2)}{(0.20)^2} = 192$$ เมื่อ $s^2 = 2$

---

[4] ความจริงหากจะพิจารณาสูตรของยามาเน่ หากแทนค่า e = 0.05 เท่ากับว่ามีการใช้ค่าความคลาดเคลื่อนจากการสุ่มตัวอย่างที่ 10% ของค่าถัวเฉลี่ย [e/p = 0.05/0.5 = 10%]

***อันเนื่องมาแต่****สูตรของยามาเน่*

อย่างไรก็ตามในกรณีที่คำนวณขนาดของตัวอย่างได้น้อยกว่า 100 ตัวอย่าง ผู้เขียนมีความเห็นว่าควรปรับขนาดของตัวอย่างให้ได้อย่างน้อย 100 ตัวอย่างสำหรับงานวิจัยเชิงปริมาณ เพราะในกรณีที่ประมวลผลออกมาเป็นค่าร้อยละขนาดของตัวอย่างขั้นต่ำไม่ควรต่ำกว่า 100 ตัวอย่าง อันที่จริงมีหลายครั้งที่ผู้เขียนไม่ได้ใช้สูตรในการหาขนาดของตัวอย่างเลย โดยเฉพาะอย่างยิ่งเมื่อมีการเลือกตัวอย่างแบบไม่อาศัยความน่าจะเป็น (Non-Probability Sampling) จากประสบการณ์งานวิจัยที่เคยทำ จากการพูดคุยกับนักวิจัยด้วยกัน และจากการอ่านตำราวิจัยตลาดภาษาอังกฤษหลายๆ เล่ม หากไม่มีความจำเป็นต้องวิเคราะห์กลุ่มย่อย พอสรุปได้ว่าควรเลือกตัวอย่างประมาณ 200 ตัวอย่างเป็นอย่างน้อย อย่างไรก็ตามเมื่อมีการเก็บตัวอย่างของจริงเสร็จแล้ว การคำนวณค่าความแปรปรวนสูงสุดของกลุ่มตัวอย่าง ($s^2$) ที่ได้จากทุกคำถามในแต่ละงานวิจัยจะทำได้ง่ายอยู่แล้ว แม้ว่าค่าความแปรปรวนของกลุ่มตัวอย่าง ($s^2$) นั้นจะแตกต่างจากสิ่งที่ผู้วิจัยคาดเดาไว้ตั้งแต่ตอนจัดทำข้อเสนองานวิจัย สิ่งที่ผู้วิจัยต้องยอมเสียก็คือค่าความคลาดเคลื่อนจากการสุ่มตัวอย่าง (e) ที่ต่างไปจากค่าที่ระบุไว้ตอนเขียนข้อเสนอโครงการวิจัย ทั้งนี้ค่านี้อาจเพิ่มขึ้นหรือลดลงได้เนื่องจากค่าความแปรปรวนของกลุ่มตัวอย่าง ($s^2$) ที่สามารถหาได้จริงจากสูงหรือต่ำกว่าค่าที่ใช้ในการคำนวณขนาดของตัวอย่างตามสูตรที่ใช้ในช่วงก่อนการเก็บตัวอย่าง แต่ผู้เขียนยังไม่เคยเห็นผู้อ่านท่านใดตามไปตรวจสอบจุดนี้เลย

ในการกำหนดขนาดของตัวอย่าง นอกจากการใช้สูตรหรือการใช้วิจารณญาณกำหนดขนาดของตัวอย่างตามประสบการณ์ที่มีแล้ว ประเภทของสถิติที่ใช้ในการวิเคราะห์ข้อมูลก็มีส่วนในการกำหนดขนาดของตัวอย่างเช่นกัน ยกตัวอย่างเช่น ในกรณีที่ผู้วิจัยใช้สถิติการวิเคราะห์องค์ประกอบ (Exploratory Factor Analysis) มีผู้ระบุว่า “หลักปฏิบัติคร่าวๆ คือ ควรใช้ขนาดของตัวอย่างอย่างน้อย 4-5 เท่าของจำนวนตัวแปรที่ใช้ในการวิเคราะห์” (Basilesky, 1994; อ้างถึงใน Malhotra, 2010, p.639) นั่นก็หมายความว่า ถ้าต้องใช้ตัวแปรในการวิเคราะห์องค์ประกอบ 30 ตัวแปร ขนาดของตัวอย่างที่เก็บไม่ควรน้อยกว่า 150 ตัวอย่างโดยประมาณนั่นเอง

## เอกสารอ้างอิง


Lindgren, Bernard W. *Statistical Theory*. 3rd ed. New York: Macmillan Publishing Co., Inc., 1976.

Malhotra, Naresh K. *Marketing Research: An Applied Orientation*. 6th ed. Upper Saddle River: Pearson Education, Inc., 2010.

Yamane, Taro. Statistics: An *Introductory Analysis*. 3rd ed. New York: Harper & Row, Publishers, Inc., 1973.

Zinkmund, William G. *Exploring Marketing Research*. 8th ed. Mason: South-Western, 2003.