ดังนั้นในทางปฏิบัติเพื่อป้องกันมิให้เกิดปัญหาในการสร้างช่วงเชื่อมั่นหรือการทดสอบสมมุติฐานเกี่ยวกับพารามิเตอร์ P นักวิจัยพึงทำการสำรวจเบื้องต้นเพื่อกะประมาณค่าของ P คร่าว ๆ ก่อนเสมอ เมื่อทราบค่า p แล้วจึงนำค่า p มาเทียบกับตารางเพื่อหาขนาดของตัวอย่าง n ที่เหมาะสมสำหรับการสำรวจจริงต่อไป

4. สำหรับกรณีที่กลุ่มประชากรมีขนาดใหญ่มากเมื่อเทียบต่อขนาดตัวอย่าง คือ $N >> n$ โครงสร้างของสูตร $V(\hat{P})$ และ $\hat{V}(\hat{P})$ จะเปลี่ยนแปลงไปเล็กน้อย กล่าวคือ เมื่อ $N >> n$ แล้ว $\frac{N-n}{N} \cong 1$ และ $\frac{N-n}{N-1} \cong 1$

$$=> V(\hat{P}) = \frac{N-n}{N} \frac{NPQ}{n(N-1)} = \frac{N-n}{N-1} \frac{PQ}{n} \simeq \frac{PQ}{n}$$

และ

$$\hat{V}(\hat{P}) = \frac{N-n}{N} \frac{npq}{n(n-1)} = \frac{N-n}{N} \frac{pq}{n-1} \simeq \frac{pq}{n}$$

สำหรับกรณีเช่นนี้เหมาะสมกับวิธีการในทางปฏิบัติเพราะในทางปฏิบัตินั้นมีอยู่เสมอที่เราไม่สามารถทราบขนาดของประชากร N ทราบแต่เพียงว่าขนาดของประชากร N ใหญ่กว่าขนาดตัวอย่าง n มาก เมื่อพบสถานการณ์ทำนองนี้การประมาณค่าพารามิเตอร์ P และการทดสอบสมมุติฐานเกี่ยวกับพารามิเตอร์ P ให้ถือว่า $N >> n$ (หรือถือว่าสุ่มตัวอย่างแบบ with replacement ก็ได้เพราะ $N >> n$ กับสุ่มแบบ with replacement มีความหมายใกล้เคียงกันมาก) แล้วใช้สูตรความแปรปรวน $V(\hat{P})$ และ $\hat{V}(\hat{P})$ ข้างต้น แต่กรณีที่ไม่ทราบค่า N นั้นไม่เหมาะสมสำหรับการประมาณค่ายอดรวม T คือ $\hat{T} = Np$

ตัวอย่าง 2.4 ในการสำรวจภาวะเศรษฐกิจและสังคมของประชากรในเขตกรุงเทพฯ ซึ่งมุ่งที่จะกะประมาณสัดส่วนของประชากรที่มีที่อยู่อาศัยเป็นของตนเอง จากการสำรวจด้วยตัวอย่างขนาด n = 900 ครัวเรือน พบว่ามีอยู่ 360 ครัวเรือนเท่านั้นที่มีอยู่อาศัยเป็นของตนเอง จงกะประมาณสัดส่วน (หรือร้อยละ) ของประชากรในเขตกทม. ที่มีอยู่ที่มีที่อาศัยเป็นของตัวเองพร้อมทั้งช่วงเชื่อมั่น 99%

วิธีทำ

ก. $n = 900, a = \sum_{i}^{900} x_i = 360$

ดังนั้น $\hat{P} = p = \frac{360}{900} = 0.40 = 40\%$

นั่นคือครอบครัวในเขตกรุงเทพฯ จะมีที่อยู่อาศัยของตนเองประมาณ 40%

ข. $\hat{V}(\hat{P}) = \frac{pq}{n}$ (กรณีนี้เราถือว่า $N >> n$)

$$= \frac{(.40)(.60)}{900} = .0002666$$

$$s_{\hat{p}} \quad \sqrt{\hat{V}(\hat{p})} = .016$$

ดังนั้นช่วงเชื่อมั่น 99% ของ P คือ

$$\hat{P} + Z_{.005}s_{\hat{P}} < P < \hat{P} + Z_{.995}s_{\hat{P}}$$

$$.40-(2.58)(.016) < P < .40+(2.58)(.016)$$

$$.40-.041 < P < .40+.041$$

$$.359 < P < .441$$

หรือสามารถเชื่อมั่นได้ถึง 99% ว่าประมาณ 35.9% ถึง 44.1% ของครอบครัวในกรุงเทพฯ เท่านั้นที่มีที่อยู่อาศัยเป็นของตนเอง

2.5 การกำหนดขนาดตัวอย่าง (Determination of Sample Size)

ขนาดตัวอย่างที่เหมาะสม (Optimal Sample Size) นั้นโดยปกติสามารถกำหนดได้ 2 ลักษณะ ทั้งนี้ขึ้นอยู่กับเกณฑ์ที่ใช้พิจารณา ซึ่งอาจต้องใช้ทั้งสองวิธีถ้างานวิจัยเกี่ยวข้องอยู่กับหลักเกณฑ์ทั้งสองคือ

1. ขนาดตัวอย่างที่สอดคล้องกับระดับความเสี่ยง α และ β

2. ขนาดตัวอย่างที่สอดคล้องกับระดับความแม่นยำ (Precision) และความน่าเชื่อถือ (Reliability)

1. ขนาดตัวอย่างที่สอดคล้องกับระดับความเสี่ยง α และ β

วิธีนี้เหมาะสำหรับงานวิจัยในกรณีที่ต้องการสอบสมมุติฐาน[1] การกำหนดขนาดตัวอย่างโดยนัยนี้ นักวิจัยจะต้องมีความเข้าใจปัญหาทางทฤษฎีพอประมาณคือ

ก. ต้องทราบว่ากลุ่มประชากรมีฟังก์ชันความน่าจะเป็น (pdf) เป็นอย่างไรหรือตัวแปรสุ่มที่สนใจมีการแจกแจงแบบใด แบบปกติ ทวินาม อนุกรมเรขาคณิต เอกโพเนนเชียล พัวซอง ยูนิฟอร์ม เบต้า หรืออะไร

ข. คุ้นเคยกับฟังก์ชันของ α และ β พอสมควร

ค. ทราบว่าสมมุติฐานที่สนใจนั้นควรใช้ test statistics ตัวใดเป็นตัวตัดสิน หรือสามารถสร้าง test สำหรับทดสอบสมมุติฐานที่สนใจนั้นได้[2]

การกำหนดขนาดตัวอย่างโดยนัยนี้ ผู้วิจัยจะต้องเป็นผู้กำหนดว่าตนต้องการให้งานทดสอบมีความเสี่ยง α และ β สูงเพียงใด ซึ่งเรื่องนี้นักวิจัยต้องกำหนดขึ้นเองโดยปกติไม่ควรกำหนดให้มีความเสี่ยงทั้งสองประเภทสูงเกินไป อาจกำหนดให้ $\alpha = 5\%, \beta = 1\%$ หรือ $\alpha = 1\%, \beta = 10\%$ หรืออื่น ๆ เป็นต้น ขนาดตัวอย่างที่ได้ตามวิธีนี้จะมีสูตรในการคำนวณแตกต่างกันไปตามลักษณะของประชากรและประเภทของสมมุติฐาน ไม่คงที่หรือเหมือนกัน นักวิจัยจึงจำเป็นต้องมีความรู้ทางทฤษฎีดีพอควร

ขอยกตัวอย่างวิธีกำหนดขนาดตัวอย่างสำหรับกรณีที่กลุ่มประชากรปกติคือ $X \sim N(\mu, \sigma^2)$ โดยที่ σ^2 เป็นพารามิเตอร์ที่เราทราบค่า และมีสมมุติฐานดังนี้

---

[1] สถิติอนุมานจำแนกออกเป็น 2 หมวดใหญ่ ๆ คือหมวดว่าด้วยการประมาณค่า (Estimation of Parameter) และหมวดที่ว่าด้วยการทดสอบสมมุติฐาน (Test of Hypothesis)

[2] อ่าน ทฤษฎีสถิติ 2

$$Ho: \mu = \mu_0 \text{ vs } H_1 : \mu = \mu_1 > \mu_0$$

สำหรับกรณีนี้ เราสามารถพิสูจน์ได้โดยง่ายว่า เราจะปฏิเสธสมมุติฐานหลัก (Ho) เมื่อ

$$\bar{x} > \mu_0 + Z_{1-\alpha_0/2} \frac{\sigma_0}{\sqrt{n}}$$

เมื่อกำหนดความเสี่ยงประเภทที่ 1 ให้เท่ากับ α_0

กำหนดให้ความเสี่ยงประเภทที่ 2 เท่ากับ β_0

ดังนั้น

β_0 = Pr (ยอมรับสมมุติฐานหลักเมื่อสมมุติฐานรองถูกต้อง)

$$= \Pr(\bar{x} < \mu_0 + Z_{1-\alpha_0/2} \frac{\sigma_0}{\sqrt{n}} \mid \mu = \mu_1)$$

$$= \Pr(\bar{x} - \mu_1 < \mu_0 - \mu_1 + Z_{1-\alpha_0/2} \frac{\sigma_0}{\sqrt{n}})$$

$$= \Pr\left(\frac{\bar{x} - \mu_1}{\sigma_0/\sqrt{n}} < \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + Z_{1-\alpha_0/2}\right)$$

$$= \Pr(Z < \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + Z_{1-\alpha_0/2})$$

$$=> \Pr(Z < \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + Z_{1-\alpha_0/2}) = \beta_0$$

เมื่อ $\Pr(Z < \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + Z_{1-\alpha_0/2}) = \beta_0$ แสดงว่า β_0 คือค่าความน่าจะเป็น

ซึ่งจากสมการแสดงว่า β_0 คือ พื้นที่ใต้โค้งด้านหางซ้าย (left hand tail) ของโค้ง N(0, 1)

ดังนั้น

$$\beta_0 = \Pr(Z < Z_{\beta_0})$$

นั่นคือ $\Pr(Z < \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + Z_{1-\alpha_0/2}) \quad = \quad \Pr(Z < Z_{\beta_0})$

$$=> \quad \frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + Z_{1-\alpha_0/2} \quad = \quad Z_{\beta_0}$$

นั่นคือ $n = \sigma_0^2 (Z_{\beta_0} - Z_{1-\alpha_0/2})^2/(\mu_0 - \mu_1)^2$

เช่น Ho : $\mu = 10$ vs H_1 : $\mu = 15$

เมื่อทราบว่า $\sigma^2 = 1600$ และกำหนดให้ $\alpha = 5\%$ $\beta = 1\%$

ดังนี้เราสามารถคำนวณหาขนาดตัวอย่างที่เหมาะสมที่สามารถควบคุมมิให้ความเสี่ยงประเภทที่ 1 เกินกว่า 5% และมิให้ความเสี่ยงประเภทที่ 2 เกินกว่า 1% ได้ดังนี้

$$\begin{aligned} n &= (1600)(Z_{.01} - Z_{.975})^2/(10-15)^2 \\ &= (1600)(-2.58 - 1.96)^2/(-5)^2 \\ &= 1{,}319.14 \\ &\simeq 1{,}319 \end{aligned}$$

นั่นคือ ถ้าสุ่มตัวอย่างมา 1,319 ชุด เพื่อทดสอบสมมุติฐาน Ho : $\mu = 10$ vs H_1 : $\mu = 15$ แล้ว ผลการทดสอบจะทำให้ความเสี่ยงของการตัดสินใจที่จะปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักเป็นจริงมีได้ไม่เกิน 5% และความเสี่ยงที่จะยอมรับสมมุติฐานหลักทั้ง ๆ ที่สมมุติฐานหลักไม่เป็นจริงมีได้ไม่เกิน 1%

2. ขนาดตัวอย่างที่สอดคล้องกับระดับความแม่นยำ (Precision) และความน่าเชื่อถือ (Reliability)

วิธีนี้เหมาะสำหรับกรณีที่ต้องการกะประมาณค่าของพารามิเตอร์ การกำหนดขนาดตัวอย่างโดยวิธีนี้นักวิจัยจำเป็นต้องมีความรู้ในสิ่งต่อไปนี้พอสมควรคือ

ก. ความแม่นยำ

ข. ความน่าเชื่อถือ

ค. ความแปรปรวนของตัวประมาณค่า (Error Variance) ของพารามิเตอร์ต่าง ๆ ในแผนสำรวจแบบต่าง ๆ กัน สำหรับข้อนี้นักศึกษาไม่จำเป็นต้องวิตกเพราะสูตรของความแปรปรวนของตัวประมาณค่าต่าง ๆ ที่นักศึกษาได้พบและศึกษามาโดยละเอียดเมื่อศึกษาถึงแผนสำรวจแต่ละแบบอยู่แล้ว

ความแม่นยำ คือ ความแตกต่างระหว่างค่าของตัวประมาณค่ากับค่าของพารามิเตอร์ ถ้าผลต่างมีค่าต่ำแสดงว่าตัวประมาณค่าให้ค่าที่แม่นยำ ถ้าผลต่างมีค่าสูงแสดงว่าตัวประมาณค่าให้ค่าที่ไม่แม่นยำ เรานิยามความแม่นยำดังนี้

$d = |\hat{\theta} - \theta|$ เมื่อ $\hat{\theta}$ คือ ค่าประมาณของค่าพารามิสเตอร์ $\hat{\theta}$ แต่ด้วยเหตุที่ θ เป็นพารามิเตอร์ซึ่งโดยปกติเราไม่อาจทราบค่าได้ ดังนั้นในทางปฏิบัตินักวิจัยและเป็นผู้ที่กำหนดผลต่าง d ขึ้นมาเอง โดยไม่ต้องคำนึงว่าค่าจริงของ θ จะเป็นเท่าไร เพราะถ้ากำหนดให้ d มีค่าต่ำย่อมแสดงว่านักวิจัยต้องการประมาณค่าให้ใกล้เคียงกับค่าของ θ มากที่สุดอยู่แล้ว

ความน่าเชื่อถือ (Reliability) หมายถึง ความน่าจะเป็นในรูป $Pr(T > t)$ โดยปกติคำว่าความน่าเชื่อถือ นิยมใช้ในงานที่เกี่ยวข้องกับอายุใช้งานของวัตถุ โดยวัดความน่าเชื่อถือในคุณภาพของวัตถุในรูปของความน่าจะเป็นที่วัตถุนั้นจะยังคงใช้งานได้ต่อเมื่อผ่านการใช้งานมาแล้วระยะหนึ่ง

ถ้าให้ T เป็นตัวแปรสุ่มแทนอายุใช้งานของวัตถุ

เราจะนิยามความน่าเชื่อถือได้ดังนี้ (R(t) = Reliability))

$R(t) = Pr(T > t)$

= ความน่าจะเป็นที่วัตถุจะยังคงใช้ได้เมื่อผ่านการใช้งานมาแล้ว t หน่วยเวลา หรือนัยหนึ่ง คือ ความน่าจะเป็นที่วัตถุจะมีอายุยาวนานกว่า t หน่วยเวลา

โดยอาศัยความหมายของคำทั้งสองขอให้นักศึกษาพิจารณาสมการความน่าจะเป็นต่อไปนี้

$$\Pr(|\hat{\theta} - \theta| < d) = 1 - \alpha$$ [1]

อ่านว่าความน่าจะเป็นที่ค่าประมาณจะคลาดเคลื่อนไปจากค่าจริงไม่ถึง d หน่วย มีค่าเท่ากับ $1 - \alpha$ หมายความว่า เราต้องการให้สามารถเชื่อมั่นได้ถึง $(1 - \alpha)100\%$ ว่าผลต่างระหว่างค่าจริงกับค่าประมาณนั้นมีอยู่ไม่ถึง d หน่วย

เมื่อพิจารณาสมการโดยละเอียดจะพบว่าสมการเขียนไว้ในรูปของ R(t) คือ $\Pr(|\hat{\theta} - \theta| > d) = \alpha$ แสดงว่า α คือ ค่าของ Reliability

และ d คือ ความแม่นยำ ดังนั้นสมการ $\Pr(|\hat{\theta} - \theta| > d) = \alpha$ จึงเป็นสมการที่ควบคุมทั้งระดับความแม่นยำและระดับความน่าเชื่อถือได้ในขณะเดียวกัน

ดังนั้น จากสมการ $\Pr(|\hat{\theta} - \theta| > d) \quad = \quad \alpha$

$$\Rightarrow \Pr\left(\left|\frac{\theta - \theta}{\sigma_{\hat{\theta}}}\right| > \frac{d}{\sigma_{\hat{\theta}}}\right) = \alpha$$

$$\Rightarrow \Pr\left(|Z| > \frac{d}{\sigma_{\hat{\theta}}}\right) = \alpha$$

$$\Rightarrow \Pr\left(|Z| > \frac{d}{\sigma_{\hat{\theta}}}\right) = \alpha = \Pr(|Z| > Z_{1-\alpha/2})$$

นั่นคือ $\dfrac{d}{\sigma_{\hat{\theta}}} = Z_{1-\alpha/2}$

ดังนั้น $d^2 = \sigma^2_{\hat{\theta}} Z^2_{1-\alpha/2}$

หรือ $(\text{ระดับความแม่นยำ})^2$ = (ความแปรปรวนของตัวประมาณค่า) × $(\text{สัมประสิทธิของระดับความน่าเชื่อถือ})^2$

---

[1] $\Pr(|\hat{\theta} - \theta| < d) = 1 - \alpha = \Pr(|\hat{\theta} - \theta| > d) = \alpha$

สมการ $d^2 = \sigma^2_{\hat{\theta}} Z^2_{1-\alpha/2}$ จะใช้เป็นสมการสำหรับกำหนดขนาดของตัวอย่างที่ควบคุมระดับความแม่นยำและความน่าเชื่อถือให้อยู่ในระดับที่นักวิจัยต้องการ ระดับความแม่นยำ d และระดับความน่าเชื่อถือ $1-\alpha$ เป็นค่าที่นักวิจัยจะต้องกำหนดขึ้นเองจะให้มีค่าเท่าไรก็ได้ โดยปกตินิยมกำหนดค่า d ไว้ต่ำ ๆ (คือแม่นยำมาก ๆ) และ $1-\alpha$ มีค่าสูง (คือน่าเชื่อถือมาก) แต่การกำหนดเช่นนี้ย่อมส่งผลสะท้อนสู่ขนาดตัวอย่าง n ด้วย เพราะอาจได้ขนาดตัวอย่าง n สูงเกินกำลังงบประมาณ กำลังงานและเวลา ทำให้ไม่อาจดำเนินงานได้ ในกรณีเช่นนี้ นักวิจัยก็สามารถลดระดับความแม่นยำและความน่าเชื่อถือลงจนได้ขนาดตัวอย่างที่พอเหมาะ

อนึ่งสูตร $d^2 = \sigma^2_{\hat{\theta}} Z^2_{1-\alpha/2}$ นี้ยังสามารถใช้คำนวณความถูกต้องแม่นยำและความน่าเชื่อถือของงานในภายหลังในกรณีที่กำหนดขนาดตัวอย่างโดยวิธีอื่น เช่น กำหนดเป็นร้อยละ หรือกำหนดเองตามความเหมาะสมได้อีกด้วย

ในกรณีของการกำหนดขนาดตัวอย่างด้วยร้อยละนั้น แม้จะไม่ใช่วิธีที่เหมาะสมในหลักทางทฤษฎีก็ตาม แต่วิธีนี้นิยมใช้กันมากในทางปฏิบัติเพราะสะดวกรวดเร็ว ถ้าจะใช้วิธีกำหนดขนาดตัวอย่างด้วยร้อยละนักวิจัยพึงปฏิบัติดังนี้

1. ควรสุ่มตัวอย่างมาประมาณ 5-60% ของกลุ่มประชากรหรืออย่างต่ำไม่ควรต่ำกว่า 30 หน่วย

2. ควรมีการวัดความถูกต้องแม่นยำและระดับความน่าเชื่อถือในภายหลัง (ในขั้นเสนอผล) ทุกครั้งโดยใช้สูตร

$$d^2 = \sigma^2_{\hat{\theta}} Z^2_{1-\alpha/2}$$

หรือจะใช้วิธีการในบทที่ 1 ก็ได้เพราะมีผลเสมอกัน

อนึ่งขอให้สังเกตว่าสมการ $d^2 = \sigma^2_{\hat{\theta}} Z^2_{1-\alpha/2}$ นี้ในทางปฏิบัติเราจะต้องใช้เป็น $d^2 = \hat{\sigma}^2_{\hat{\theta}} Z^2_{1-\alpha/2}$ หรือ $d^2 = \hat{V}(\hat{\theta}) Z^2_{1-\alpha/2}$ ซึ่งเมื่อใช้ไปในงานกำหนดขนาดตัวอย่าง $\hat{V}(\hat{\theta})$ จะมี n แฝงอยู่ ถ้าเรากำหนดค่า d และ $1-\alpha$ ก็จะคำนวณหาค่า n ได้ แต่ถ้าจะใช้ไปในการประเมินคุณภาพของขนาดตัวอย่างที่เลือกใช้ขึ้นเองด้วยร้อยละหรือวิธีอื่น ๆ ที่เห็นสมควร (เช่น

กำหนดตามประสบการณ์หรือตามงบประมาณ เวลา หรือข้อจำกัดอย่างอื่น ๆ) เราจะทราบค่า n และ $\hat{V}(\hat{\theta})$ แล้ว ปัญหาจึงอยู่ที่การคำนวณหา d และ $1-\alpha$ ซึ่งในกรณีนี้จะพบว่า เรามีสมการเดียวแต่มี 2 ตัวแปร วิธีปฏิบัติคือให้กำหนดค่าตัวแปรไว้ตัวหนึ่ง แล้วคำนวณหาค่าของอีกตัวแปรตัวหนึ่ง เช่น อาจกำหนดค่า $1-\alpha$ แล้วคำนวณหา d หรือกำหนดค่า d ก่อนแล้วจึงคำนวณหาค่าของ $1-\alpha$

อนึ่งสูตร $d^2 = \hat{V}(\hat{\theta})Z^2_{1-\alpha/2}$ สามารถเปลี่ยนรูปเป็น $d^2 = (CV)^2Z^2_{1-\alpha/2}$ ได้ เมื่อทราบค่า CV (Coefficient of Variation) โดยที่ $CV = \sigma/\bar{X}$ และ $CV = s/\bar{x}$ ซึ่งเรามักทราบค่า CV ได้จากประสบการณ์

การศึกษาเรื่องการกำหนดขนาดตัวอย่างในลำดับต่อไปจะศึกษาเฉพาะกรณีการกำหนดขนาดตัวอย่างให้สอดคล้องกับระดับความแม่นและความน่าเชื่อถือเท่านั้น กรณีกำหนดให้สอดคล้องกับ α และ β จะขอเว้นไว้เพราะจะได้กล่าวถึงในทฤษฎีสถิติ 2 โดยเฉพาะ

2.5.1 การกำหนดขนาดตัวอย่างเพื่อกะประมาณค่าเฉลี่ย $\bar{X}$

จาก $$\hat{V}(\bar{x}) = \frac{N-n}{N}\frac{s^2}{n} = (1-\frac{n}{N})\frac{s^2}{n} = \frac{s^2}{n} - \frac{s^2}{N}$$

กำหนดให้ $d = d_0$ และ $\alpha = \alpha_0$

ดังนั้น $$d_0^2 = (\frac{s^2}{n} - \frac{s^2}{N})Z^2_{1-\alpha_0/2}$$

$$= \frac{s^2}{n}Z^2_{1-\alpha_0/2} - \frac{s^2}{N}Z^2_{1-\alpha_0/2}$$

$$=> \frac{s^2}{n}Z^2_{1-\alpha_0/2} = d_0^2 + \frac{s^2}{N}Z^2_{1-\alpha_0/2}$$

$$= (Nd_0^2 + s^2Z^2_{1-\alpha_0/2})/N$$

ดังนั้น $$n = (Ns^2Z^2_{1-\alpha_0/2})/(Nd_0^2 + s^2Z^2_{1-\alpha_0/2})$$

s^2 คือ ค่าความแปรปรวนซึ่งสามารถทราบได้จากการสำรวจเบื้องต้นหรือจากประสบการณ์ เช่น จากงานวิจัยของบุคคลอื่นที่เคยทำมาแล้วในอดีต

2.5.2 การกำหนดขนาดตัวอย่างเพื่อกะประมาณยอดรวม T

จาก $\hat{V}(\hat{T}) = N^2 \frac{N-n}{N} \frac{s^2}{n} = N^2(1 - \frac{n}{N}) \frac{s^2}{n} = \frac{N^2s^2}{n} - Ns^2$

กำหนดให้ $d = d_0$ และ $\alpha = \alpha_0$

$$\text{ดังนั้น } d_0^2 = (\frac{N^2s^2}{n} - Ns^2)Z^2_{1-\alpha_0/2}$$

$$= \frac{N^2s^2}{n} Z^2_{1-\alpha_0/2} - Ns^2Z^2_{1-\alpha_0/2}$$

$$=> \quad \frac{N^2s^2}{n} Z^2_{1-\alpha_0/2} = d_0^2 + Ns^2Z^2_{1-\alpha_0/2}$$

$$n = (N^2s^2Z^2_{1-\alpha_0/2})/(d_0^2 + Ns^2Z^2_{1-\alpha/2})$$

ตัวอย่าง 2.5 ในเมืองหนึ่งสามารถจัดกรอบตัวอย่างได้เป็น 800 บล๊อก (block) นักวิจัยต้องการจะกะประมาณยอดรวมของจำนวนรถยนต์ที่จอดทิ้งไว้นอกบ้านในเวลากลางคืน คือ โดยทดลองทำการสำรวจเบื้องต้นด้วยการสุ่มตัวอย่างบล๊อกมา 30 บล๊อกแล้วทำการสำรวจ ปรากฏจำนวนรถยนต์ที่จอดทิ้งไว้นอกบ้านในแต่ละบล๊อกดังนี้

8,9,4,2,6,6,3,0,8,4

3,6,4,2,3,3,9,3,0,1

4,2,3,6,1,7,5,2,5,2

ก. อยากทราบว่าขนาดตัวอย่างที่กำหนดขึ้นเองนี้เหมาะสมหรือไม่

ข. ถ้าจะให้การกะประมาณยอดรวมคลาดเคลื่อนไปจากค่าจริง ± 400 คัน ณ. ระดับความน่าเชื่อถือ 95% อยากทราบว่าควรใช้ขนาดตัวอย่างเท่าไร

ค. ถ้าต้องการกะประมาณให้ค่าเฉลี่ย $\bar{x}$ (จำนวนรถยนต์ที่จอดทิ้งไว้นอกบ้านเฉลี่ยต่อ 1 บล๊อก) คลาดเคลื่อนไปจากค่าจริง ± 1 คัน ณ. ระดับความน่าเชื่อถือ 95% อยากทราบว่าควรสุ่มตัวอย่างมากี่บล๊อก

วิธีทำ จากข้อมูลพบว่า

$$s^2 = \frac{1}{n-1}\left\{\sum_{i}^{n} x_i^2 - \frac{(\Sigma x_i)^2}{n}\right\}$$

$$= \frac{1}{29}\left\{673 - \frac{(121)^2}{30}\right\}$$

$$= 184.97/29$$

$$s^2 = 6.378$$

$$s = 2.525$$

ก. จากการกำหนดขนาดตัวอย่าง $n = 30$ พบว่า $s^2 = 6.378$, $\bar{x} = 4.033$, $\hat{T} = N\bar{x} = (800)(4.033) = 3226.6$

$$\hat{V}(\hat{T}) = N^2 \frac{N-n}{N} \frac{s^2}{n}$$

$$= (800)^2 \frac{800-30}{800} \frac{6.378}{30}$$

$$= 130{,}961.6$$

จากสมการ $d^2 = \hat{V}(\hat{T})Z^2_{1-\alpha/2}$

$\Rightarrow \quad d^2 = (130{,}961.6)Z^2_{1-\alpha/2}$

(1) ถ้าจะให้งานนี้เชื่อถือได้ 95% จะพบว่า

$$d^2 = (130{,}961.6)(1.96)^2 = 503{,}102.05$$

$=> \quad d \quad \simeq \quad \pm 709$

หมายความว่า ถ้าใช้ขนาดตัวอย่าง $n = 30$ จะพบว่าการประมาณค่าแม้จะเชื่อมั่นในความถูกต้องถึง 95% แต่ค่าประมาณ $\hat{T}$ กับค่าจริงจะคลาดเคลื่อนกันถึง ± 709 คัน แต่ถ้าลดระดับความน่าเชื่อถือลง คือ กำหนดให้ $(1-\alpha) < .95$ ความแม่นยำจะสูงขึ้น

(2) ถ้าจะให้งานนี้มีความแม่นยำประมาณ ± 400 คัน จะพบว่า

$$\pm(400)^2 = (130{,}961.6)Z^2_{1-\alpha/2}$$

$$=> \quad Z^2_{1-\alpha/2} = 1.2217$$

$$=> \quad Z_{1-\alpha/2} = 1.105$$

แสดงว่า $\Pr(|Z| > 1.105) = \alpha$ ซึ่งเปิดตารางโค้งปกติจะพบว่า $\alpha = 26.70\%$ นั่นคือ ถ้าระบุให้งานมีความแม่นยำ โดยกำหนดให้ค่า $\hat{T}$ คลาดเคลื่อนไปจากค่าจริงเพียง ± 400 การประมาณค่าครั้งนี้ (โดยใช้ขนาดตัวอย่าง $n = 30$) จะมีความน่าเชื่อถือเพียง 73.30% เท่านั้น

หมายเหตุ ในทางปฏิบัติ นักวิจัยจะเปลี่ยนค่าของ d และ α ไปในลักษณะใดก็ได้ โดยปกตินิยมกำหนดค่า α ไว้ต่ำหรือถ้าจะกำหนด α นิยมกำหนด α ไว้เป็น 1% 5% 10%

ข. ถ้าต้องการให้ $d = \pm 400$ และ $(1-\alpha) = 95\%$ ควรกำหนดขนาดตัวอย่างดังนี้

$$\begin{aligned}
n &= (N^2s^2Z^2_{1-\alpha_0/2})/(d_0^2 + Ns^2Z^2_{1-\alpha_0/2}) \\
&= (800)^2(6.378)(1.96)^2/\{(\pm 400)^2 + (800)(6.378)(1.96)^2\} \\
&= 15{,}681{,}103/179{,}601.37 \\
&= 87.3 \\
&\simeq 87 \text{ บล๊อก}
\end{aligned}$$

นั่นคือ ถ้าต้องการให้ค่าประมาณของยอดรวมคลาดเคลื่อนไปจากค่าจริงประมาณ ± 400 และให้มีความน่าเชื่อถือได้ถึง 95% การสำรวจจริงควรใช้ตัวอย่างบล๊อกประมาณ 87 บล๊อก

ค. ต้องการประมาณค่าเฉลี่ย $\bar{x}$ ให้คลาดเคลื่อนไปจาก $\bar{X}$ เพียง ± 1 คัน และมีระดับความน่าเชื่อถือ 95%

ดังนั้น $d^2 = \hat{V}(\bar{x})Z^2_{1-\alpha/2}$

$$=> n = (Ns^2Z^2_{1-\alpha_0/2})/(Nd_0^2 + s^2Z^2_{1-\alpha_0/2})$$

$$= (800)(6.378)(1.96)^2/\{(800)(\pm 1)^2 + (6.378)(1.96)^2\}$$

$$= 19,601.378/(800 + 24.506)$$

$$= 19,601.379/824.506$$

$$= 23.77$$

$$= 24 \text{ บล๊อก}$$

นั่นคือ ถ้าสุ่มตัวอย่างบล๊อกมา 24 บล๊อก จะสามารถประมาณคำนวณค่าเฉลี่ยของรถยนต์ที่จอดทิ้งไว้นอกบ้านต่อบล๊อกได้คลาดเคลื่อนไปจากค่าเฉลี่ยจริงไม่เกิน ± 1 คัน และเชื่อมั่นในความถูกต้องได้ถึง 95%

2.5.3 การกำหนดขนาดตัวอย่างเพื่อกะประมาณอัตราส่วน R

จาก $\hat{V}(\hat{R}) \simeq \frac{1}{\bar{y}^2} \frac{N-n}{N} \frac{d_d^2}{n}$ เมื่อ $s_d^2 = \frac{1}{n-1} \sum_i^n (x_i - \hat{R}y_i)^2$

เมื่อกำหนดให้ $d = d_0$ และ $\alpha = \alpha_0$

$$=> d_0^2 = \left(\frac{1}{\bar{y}^2} \frac{N-n}{N} \frac{s_d^2}{n}\right)(Z^2_{1-\alpha_0/2})$$

$$\simeq \frac{1}{\bar{y}^2}(1-\frac{n}{N})\frac{s_d^2}{n}.Z^2_{1-\alpha_0/2}$$

$$\simeq \frac{s_d^2}{n\bar{y}^2}.Z^2_{1-\alpha_0/2} - \frac{s_d^2}{N\bar{y}^2}.Z^2_{1-\alpha_0/2}$$

$$=> \frac{s_d^2}{n\bar{y}^2}.Z^2_{1-\alpha_0/2} \simeq d_0^2 + \frac{s_d^2}{N\bar{y}^2}.Z^2_{1-\alpha_0/2}$$

$$\simeq (N\bar{y}^2 d_0^2 + s_d^2 Z^2_{1-\alpha_0/2})/N\bar{y}^2$$

$$=> \quad n \simeq (Ns_d^2 Z^2_{1-\alpha_0/2})/(N\bar{y}^2 d_0^2 + s_d^2 Z^2_{1-\alpha_0/2})$$

เมื่อ $s_d^2 = \frac{1}{n-1}\sum_i^n (x_i - \hat{R}y_i)^2$ เป็นค่าที่ได้รับจากงานสำรวจเบื้องต้น

2.5.4 การกำหนดขนาดตัวอย่างเพื่อกะประมาณสัดส่วน?

จาก $\hat{V}(\hat{P}) = \frac{N-n}{N}\frac{pq}{n-1} \simeq \frac{N-n}{N}\frac{pq}{n}$

$$\simeq (1-\frac{n}{N})\frac{pq}{n}$$

$$\simeq \frac{pq}{n} - \frac{pq}{N}$$

กำหนดให้ $d = d_0$ และ $\alpha = \alpha_0$

ดังนั้น

$$d_0^2 \simeq (\frac{pq}{n} - \frac{pq}{N})Z^2_{1-\alpha_0/2}$$

$$\simeq \frac{pq}{n}.Z^2_{1-\alpha_0/2} - \frac{pq}{N}.Z^2_{1-\alpha_0/2}$$

$$= \frac{pq}{n}.Z^2_{1-\alpha_0/2} \simeq d_0^2 + \frac{pq}{N}.Z^2_{1-\alpha_0/2}$$

$\simeq \quad (Nd_0^2 + pqZ^2_{1-\alpha_0/2})/N$

ดังนั้น $\quad n \quad \simeq \quad (NpqZ^2_{1-\alpha_0/2})/(Nd_0^2 + pqZ^2_{1-\alpha_0/2})$

เมื่อ p และ q เป็นค่าที่ได้มาจากการสำรวจเบื้องต้น

ตัวอย่าง 2.6 จากการทดลองสำรวจเบื้องต้นโดยสุ่มตัวอย่างนักศึกษามา 100 คน จากนักศึกษาในสถาบันการศึกษาแห่งหนึ่งในกรุงเทพมหานครซึ่งมีนักศึกษาทั้งสิ้น 5000 คน เพื่อกะประมาณสัดส่วนของนักศึกษาที่มีภูมิลำเนาอยู่ต่างจังหวัด ผลการสำรวจพบว่าเป็นนักศึกษาที่มีภูมิลำเนาอยู่ต่างจังหวัด 40 คน

อยากทราบว่า ถ้าจะทำการสำรวจจริง ควรสุ่มตัวอย่างนักศึกษามากี่คนจึงจะทำให้ค่าประมาณมีความคลาดเคลื่อนไปจากค่าจริง P เพียง $\pm 5\%$ ทั้งนี้ต้องมีระดับความน่าเชื่อถือสูงถึง 99%

วิธีทำ $n = 100 \quad a = \sum_{i}^{400} x_i = 40$

ดังนั้น $\hat{P} = 40/100 = .40$ และ $q = 1 - p = .60$

กำหนดให้ $d = \pm 5\% = \pm 5/100 = \pm .05$ และ $1 - \alpha = .99$

ดังนั้นขนาดตัวอย่างที่เหมาะสมคือ

$$
\begin{aligned}
n &\simeq (NpqZ^2_{1-\alpha_0/2})/(Nd_0^2 + pqZ^2_{1-\alpha_0/2}) \\
&\simeq (5000)(.40)(.60)(2.58)^2/\{(5000)(\pm .05)^2 + (.40)(.60)(2.58)^2\} \\
&\simeq 7987.68/(12.5 + 1.597) \\
&\simeq 7987.68/14.09 \\
&\simeq 566.9 \\
&\simeq 567
\end{aligned}
$$

นั่นคือ ถ้าต้องการให้มี $d = \pm 5\%$ และระดับความน่าเชื่อถือ 99% การสำรวจจริงควรสุ่มตัวอย่างนักศึกษามาเป็นตัวอย่างประมาณ 567 คน

2.5.5 คำแนะนำทั่วไป

1. ในทางปฏิบัตินั้น แบบสอบถามจะประกอบไปด้วยข้อถามมากมายหลายรายการและจำเป็นต้องประมาณค่าพารามิเตอร์หลายตัว เมื่อไม่มีข้อมูลในอดีตชี้ให้เห็นว่า S^2 มีค่าเท่าใด เราจำเป็นจะต้องทำการสำรวจเบื้องต้น (Preliminary Survey) ก่อนเสมอ ซึ่งจากการสำรวจเบื้องต้นทำให้เราสามารถคำนวณหาค่าความแปรปรวน s^2 และทำให้สามารถกำหนดขนาดตัวอย่างให้แก่รายการต่าง ๆ ได้มากมายหลายรายการซึ่งแปลว่าเราจะมีค่า-n หลายค่า ในกรณีเช่นนี้ให้เลือกค่า n ที่สูงที่สุดมาใช้เป็นขนาดตัวอย่างของงานสำรวจจริง

2. การกำหนดขนาดตัวอย่างด้วยร้อยละ คือ ใช้ตัวอย่างประมาณ 5-60% ของขนาดประชากรนั้น สะดวกรวดเร็วดีเพราะตัดขั้นตอนของการสำรวจเบื้องต้นออกไปจากแผนงานทำให้ประหยัดกำลังแรงงาน งบประมาณและเวลาได้มาก แต่วิธีนี้ไม่ให้ความมั่นใจแก่นักวิจัยในด้านความแม่นยำและความน่าเชื่อถือของค่าประมาณ ดังนั้นในทางปฏิบัติถ้าจำเป็นต้องใช้วิธีนี้ขอให้กำหนดขนาดตัวอย่างเผื่อไว้มาก ๆ คือ กำหนดค่า n ให้สูงไว้ก่อน และจะต้องคำนวณหาค่า d และ α ในภายหลังทุกครั้ง

3. นักศึกษาไม่จำเป็นต้องเสียเวลาจดจำสูตรสำหรับคำนวณหาค่า n สำหรับกะประมาณพารามิเตอร์แบบต่าง ๆ เพราะเป็นการสิ้นเปลืองสมองโดยไม่จำเป็นนักศึกษาสามารถคำนวณหาค่า n ได้จากสมการ $d^2 = \hat{V}(\hat{\theta})Z^2_{1-\alpha/2}$ ได้โดยตรง ขอเพียงให้ทราบหรือจำสูตร $\hat{V}(\hat{\theta})$ ได้ก็พอ ทั้งนี้แม้จะไม่ทราบค่าของ N ก็ยังคงทำได้ เพราะกรณีที่ไม่ทราบค่า N นั้นมักปรากฏเฉพาะกับกรณี N มีขนาดใหญ่มาก ซึ่งในกรณีเช่นนี้ให้ถือว่า $N >> n$ ซึ่งหมายความว่า $\frac{N-n}{N} \simeq 1$

เช่น ในตัวอย่าง 2.6 เราสามารถคำนวณค่า n ได้โดยตรงจากสูตร $d^2 = \hat{V}(\hat{P})Z^2_{1-\alpha/2}$ ดังนี้

$$d_0^2 \simeq \left(\frac{N-n}{N}\cdot\frac{pq}{n}\right)Z^2_{1-\alpha_0/2}$$

$$=> (\pm .05)^2 \simeq (1-n/5000)\ \frac{(.40)(.60)}{n}\cdot(2.58)^2$$

$$\simeq \frac{(.40)(.60)(2.58)^2}{n} - \frac{(.40)(.60)(2.58)^2}{5000}$$

$$=> .0025 \simeq 1.597/n - .00032$$

$$=> n \simeq 1.597/(.00032+.0025)$$

$$\simeq 566.9$$

$$\simeq 567$$

สำหรับกรณีของการประมาณค่า $\overline{X}$, T, R ก็สามารถกระทำได้ในทำนองเดียวกัน ขอเพียงทราบว่า $\hat{V}(x)$, $\hat{V}(\hat{T})$ และ $\hat{V}(\hat{R})$ มีสูตรเป็นอย่างไรเท่านั้นก็สามารถคำนวณหาขนาดตัวอย่าง n ได้ตามที่ต้องการ

4. ขนาดตัวอย่างที่ใช้สำหรับทดลองสำรวจเบื้องต้นนั้นจะใช้ขนาดเท่าไรก็ได้ ทั้งนี้ขึ้นอยู่กับความเหมาะสม เราอาจเดาเอาเองก็ได้ แต่โดยปกติไม่ควรให้ต่ำกว่า 30 หน่วย และในขณะสำรวจให้ดำเนินการสำรวจเสมือนขนาดตัวอย่างที่เดาไว้นั้นคือขนาดตัวอย่างจริง หมายความว่าให้เลือกตัวอย่าง (หน่วยสำรวจ) จากกรอบตัวอย่างมาโดยสุ่มตามแผน SRS ทุกประการ เพราะเราจะได้สามารถใช้ขนาดตัวอย่างนี้เป็นส่วนหนึ่งของการหาขนาดตัวอย่างสำหรับการสำรวจจริงได้ด้วย กล่าวคือ สมมุติว่า n' คือขนาดตัวอย่างที่ใช้ในการสำรวจเบื้องต้น และ n คือขนาดตัวอย่างที่คำนวณได้ตามสูตรเพื่อใช้สำหรับการสำรวจจริง

(1) ถ้า $n \leq n'$ ให้ถือว่า n' คือขนาดตัวอย่างที่ใช้สำรวจจริง นัยหนึ่งถือว่า การสำรวจเบื้องต้นนั้นก็คือการสำรวจจริง ไม่จำเป็นต้องสำรวจอีกต่อไป

(2) ถ้า $n > n'$ ให้สำรวจเพิ่มเติมอีกเพียง $(n-n')$ หน่วยเท่านั้น

(3) ถ้า $n > N$ ให้สำรวจทุกหน่วยในกลุ่มประชากร กรณีเช่นนี้มักไม่ปรากฏขึ้นบ่อยครั้งนัก เท่าที่พบมักปรากฏขึ้นในกรณี Stratified Random Sampling เมื่อมีการแบ่งสรรจำนวนตัวอย่าง (Allocation of Sample Size)

5. ในกรณีที่ต้องการกะประมาณค่าพารามิเตอร์สำหรับกรณีของประชากรย่อย (Estimation of Proportion over Subpopulation) ให้ใช้เทคนิคการประมาณค่าโดยยึดถือการแจกแจงพหุนาม (Multinomial Distribution) โดยที่ตัวแปรสุ่มในประชากรย่อยมีการแจกแจงแบบทวินาม และสัดส่วน p ในแต่ละประชากรย่อยมีการแจกแจงแบบปกติโดยประมาณ

สมมุติว่ากลุ่มประชากรขนาด N ประกอบขึ้นด้วยประชากรย่อย k กลุ่ม ๆ หนึ่ง ๆ มีขนาด $N_1, N_2, .., N_k$ ตามลำดับ และในกลุ่มตัวอย่างขนาด n ก็ประกอบไปด้วยหน่วยตัวอย่างจากกลุ่มประชากรย่อยทั้ง k กลุ่ม เป็นจำนวน $n_1, n_2, ..., n_k$ หน่วยตามลำดับ นั่นคือ $N = \sum_{i}^{k} N_i$ และ $n = \sum_{i}^{k} n_i$

ดังนั้นตัวแปรสุ่ม x_i; $i = 1,2,...,k$ จะมีการแจกแจงแบบทวินามคือ

$$x_i \sim B\,(N_iP_i,\ N_iP_iQ_i)\quad ;\ i = 1, 2, ..., k$$

และเมื่อ $n_i \to \infty$ ตัวแปรสุ่ม $p_i = \frac{1}{n} \sum_{j}^{n_i} x_{ij}$ จะมีการแจกแจงแบบปกติโดยประมาณคือ

$$p_i \sim N\,(P_i,\ \frac{N_i - n_i}{N_i - 1}\ \frac{P_iQ_i}{n_i});\ i = 1, 2, ..., k$$

โดยที่ $x_{ij} = 1$ เมื่อ $x_{ij} \in C_i$ และ $x_{ij} = 0$ เมื่อ $x_{ij} \in C_i'$ สำหรับ $i = 1, 2, ..., k$

นั่นคือสำหรับกลุ่มประชากรย่อยใด ๆ เราสามารถพิสูจน์ให้เห็นโดยง่ายว่า

(1) $\hat{P}_i = p_i = \frac{1}{n} \sum_{j}^{n} x_{ij}$ เป็นตัวประมาณค่าที่ปราศจากอคติของ $P_i = \frac{1}{N} \sum_{j}^{N} x_{ij}$

(2) $V(\hat{P}_i) = V(p_i) = \frac{N_i - n_i}{N_i - 1} \frac{P_i Q_i}{n}$; $i = 1, 2, \ldots, k$

(3) $\hat{V}(\hat{P}_i) = \hat{V}(p_i) = \frac{N_i - n_i}{N_i} \frac{p_i q_i}{n_i - 1}$; $i = 1, 2, \ldots, k$

และช่วงเชื่อมั่น $(1-\alpha)$ 100% ของการประมาณค่า P_i ปรากฏดังนี้

$$p_i + Z_{\alpha/2} \sqrt{\frac{N_i - n_i}{N_i} \frac{p_i q_i}{n_i - 1}} \leq P_i \leq p_i + Z_{1-\alpha/2} \sqrt{\frac{N_i - n_i}{N_i} \frac{p_i q_i}{n_i - 1}}$$

สำหรับทุก ๆ $i = 1, 2, \ldots, k$

สำหรับกรณีของการประมาณยอดรวม T_i ให้ปฏิบัติดังนี้คือ

ก. ถ้าสามารถทราบค่าของ N_i ; $i = 1, 2, \ldots, k$ เราสามารถประมาณยอดรวม T_i ; $i = 1, 2, \ldots, k$ ได้ดังนี้คือ

(1) $\hat{T}_i = N_i p_i$; $i = 1, 2, \ldots, k$

(2) $V(\hat{T}_i) = N_i^2 \frac{N_i - n_i}{N_i - 1} \frac{P_i Q_i}{n_i}$; $i = 1, 2, \ldots, k$

(3) $\hat{V}(\hat{T}_i) = N_i^2 \frac{N_i - n_i}{N_i} \frac{p_i q_i}{n_i - 1}$; $i = 1, 2, \ldots, k$

(4) $N_i p_i + Z_{\alpha/2} N_i \sqrt{\frac{N_i - n_i}{N_i} \frac{p_i q_i}{n_i - 1}} \leq \hat{T}_i \leq N_i p_i + Z_{1-\alpha/2} N_i \sqrt{\frac{N_i - n_i}{N_i} \frac{p_i q_i}{n_i - 1}}$

สำหรับทุก ๆ $i = 1, 2, \ldots, k$

ข. ถ้าไม่ทราบค่าของ N_i ; $i = 1, 2, \ldots, k$ เราสามารถกะประมาณยอดรวม T_i ; $i = 1, 2, \ldots, k$ ได้ดังนี้คือ

(1) $\hat{T}_i = \frac{N}{n} \sum_{j}^{n_i} x_{ij} = N p_i$

(2) $\hat{V}(\hat{T}_i) = V(N p_i) = N^2 \frac{(N-n)}{N} \frac{p' q'}{n-1}$