บทที่ 6
การประมาณค่า

การประมาณค่าเป็นวิธีการทางสถิติอย่างหนึ่ง ที่ใช้ตัวสถิติที่ได้จากตัวอย่างไปประมาณพารามิเตอร์ของประชากร ตัวสถิติที่สนใจนั้นเป็นค่าคงที่ที่คำนวณได้จากข้อมูลที่ได้จากตัวอย่างสุ่มแล้วนำไปประมาณค่าพารามิเตอร์ที่สอดคล้องกัน อันจะนำไปสู่การอ้างอิงถึงประชากร การทำความเข้าใจเกี่ยวกับเรื่องการประมาณค่า จะต้องอาศัยความรู้ทางด้านทฤษฎีที่เกี่ยวข้องกับการแจกแจงความน่าจะเป็นของตัวสถิติ ตัวประมาณค่าที่ดีและสอดคล้องกับตัวสถิติ และพารามิเตอร์ที่สนใจ เพื่อให้การประมาณค่าถูกต้องชัดเจน ใกล้เคียงกับความเป็นจริง โดยให้เกิดความผิดพลาดน้อยที่สุด

6.1 การประมาณค่า

การประมาณค่า (estimation) หมายถึงการนำเอาตัวสถิติที่ได้จากตัวอย่างไปประมาณพารามิเตอร์ของประชากร การประมาณค่าเป็นวิธีการหนึ่งในการหาข้อสรุปเกี่ยวกับประชากร โดยการอาศัยตัวอย่างที่ได้จากการชักตัวอย่างแบบสุ่มด้วยวิธีการต่าง ๆ ที่เหมาะสมแล้วนำข้อมูลที่ได้จากตัวอย่างนั้นไปคำนวณหาตัวสถิติเพื่อนำไปประมาณพารามิเตอร์ที่สนใจ อันได้แก่ ค่าเฉลี่ย (μ) ความแปรปรวน (σ^2) และสัดส่วน (P) ซึ่งเป็นตัวไม่ทราบค่า ถ้าให้ θ เป็นพารามิเตอร์ใด ๆ ที่สนใจ $\hat{\theta}$ เป็นตัวสถิติที่ใช้ประมาณพารามิเตอร์ θ เรียก $\hat{\theta}$ ว่าตัวประมาณ (estimator) ของ θ โดยที่ $\hat{\theta}$ จะเป็นตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นของตัวเอง ค่า $\hat{\theta}$ ที่คำนวณได้ เรียกว่าค่าประมาณ (estimate) ของ θ (Casella & Berger, 1990) ในการประมาณค่าพารามิเตอร์โดยอาศัยตัวสถิติจากตัวอย่างนั้น ผู้ทำการประมาณมีความมุ่งหวังที่จะประมาณให้มีความถูกต้องหรือต้องการให้ค่าประมาณกับค่าพารามิเตอร์นั้นเป็นค่าเดียวกัน แต่ในทางปฏิบัตินั้นทำได้ยาก เนื่องจากตัวอย่างที่เลือกมาเป็นเพียงส่วนหนึ่งของประชากรเท่านั้น อย่างไรก็ตามเราต้องการตัวประมาณที่เหมาะสมที่สุด ซึ่งควรมีสมบัติที่สำคัญ 2 ประการ คือ ความไม่เอนเอียง (unbiasedness) โดยที่ตัวประมาณค่า $\hat{\theta}$ ของ θ จะไม่เอนเอียง ถ้า $E(\hat{\theta})$ เท่ากับ θ และตัวประมาณนั้นควรจะมีความแปรปรวนต่ำสุด (minimum variance) โดยหากว่า θ_i เป็นตัวประมาณใด ๆ ของ θ แล้วจะได้ $V(\hat{\theta}) \leq V(\theta_i)$ (ประชุม สุวัตถี, 2527)

การประมาณค่าตัวพารามิเตอร์ของประชากรโดยใช้ตัวสถิติจากตัวอย่างเพื่ออธิบายลักษณะบางอย่างของประชากรโดยอาศัยตัวแทนจากตัวอย่างมี 2 แบบ คือการประมาณค่าแบบจุด (point estimation) และการประมาณค่าแบบช่วง (interval estimation) (ศิริชัย พงษ์วิชัย, 2547)

6.2 การประมาณค่าแบบจุด

การประมาณค่าแบบจุดเป็นการใช้ตัวสถิติที่คำนวณได้จากตัวอย่างไปประมาณพารามิเตอร์ซึ่งกระทำได้ทันที เช่น ใช้ $\overline{X}$ ประมาณ μ ใช้ s^2 ประมาณ σ^2 และใช้ p ประมาณ P วิธีการในการประมาณนั้น ทำได้โดยคำนวณค่าตัวสถิติได้เท่าใด ก็เอาค่าที่คำนวณได้นั้นไปประมาณพารามิเตอร์ได้ทันทีโดยไม่ต้องดำเนินการเพิ่มเติมแต่อย่างใด เรียกตัวประมาณที่ได้จากการประมาณแบบนี้ว่าตัวประมาณแบบจุด (point estimator) (ศิริชัย กาญจนวาสี, 2547; Dudewicz, 1988) จะเห็นได้ว่าการประมาณแบบจุดนั้นทำได้ง่ายแต่โอกาสผิดพลาดย่อมมีมาก อย่างไรก็ตาม นักวิจัยจะอาศัยตัวประมาณแบบจุดนี้ช่วยในการสร้างตัวประมาณแบบช่วงด้วย การประมาณค่าแบบจุดจะแบ่งเป็นสองกรณี คือ การประมาณค่าแบบจุดกรณีประชากรเดียวและการประมาณค่าแบบจุดกรณีสองประชากร

6.2.1 การประมาณค่าแบบจุดกรณีประชากรเดียว

การประมาณค่าแบบจุดสำหรับประชากรเดียวเป็นการประมาณค่าพารามิเตอร์ที่สนใจของหนึ่งประชากร โดยจะได้ค่าประมาณเป็นเลขโดดจำนวนเดียวไปประมาณพารามิเตอร์ซึ่งเป็นการยากที่จะทำให้ตัวประมาณมีค่าเท่ากับพารามิเตอร์ที่แท้จริง นอกจากจะทราบการแจกแจงค่าสถิติจากการสุ่มตัวอย่างของตัวประมาณที่สมบูรณ์ (ศิริชัย กาญจนวาสี, 2547) ในที่นี้จะนำเสนอการประมาณค่าแบบจุดของพารามิเตอร์สำหรับประชากรเดียวที่น่าสนใจและเป็นที่นิยมใช้กันเพียง 2 ตัว คือการประมาณค่าเฉลี่ยและการประมาณสัดส่วน

6.2.1.1 การประมาณค่าเฉลี่ยกรณีประชากรเดียว

การประมาณค่าแบบจุดของค่าเฉลี่ยกรณีประชากรเดียว ทำได้โดยการใช้ค่าเฉลี่ยที่คำนวณได้จากตัวอย่างซึ่งเป็นค่าคงที่ค่าเดียวนำไปประมาณค่าเฉลี่ยของประชากร

ตัวอย่างที่ 6.1 ในช่วง 1 สัปดาห์ก่อนเปิดภาคเรียนใหม่ ห้างสรรพสินค้าแห่งหนึ่งบันทึกรายการจำนวนรองเท้าที่ขายได้ในแต่ละวันเป็นดังนี้

วัน	อาทิตย์	จันทร์	อังคาร	พุธ	พฤหัสบดี	ศุกร์	เสาร์
จำนวนรองเท้า	34	11	15	22	28	35	40

จงหาค่าประมาณแบบจุดของจำนวนรองเท้าเฉลี่ยที่ขายได้ในช่วงนี้

วิธีทำ จำนวนรองเท้าที่ขายได้เฉลี่ยต่อวัน $= \dfrac{34+10+15+22+28+35+40}{7}$

$= 26.43$

ดังนั้น ค่าประมาณแบบจุดของจำนวนรองเท้าที่ขายต่อวันประมาณ 27 คู่

6.2.1.2 การประมาณค่าสัดส่วนกรณีประชากรเดียว

การประมาณค่าแบบจุดของสัดส่วนกรณีประชากรเดียว ทำได้โดยการใช้สัดส่วนที่คำนวณได้จากตัวอย่างซึ่งเป็นค่าคงที่ค่าเดียวประมาณสัดส่วนของประชากร

ตัวอย่างที่ 6.2 ในการสำรวจความต้องการที่อยู่อาศัยของประชาชนในชุมชนแห่งหนึ่ง สุ่มประชาชน 1,000 คน พบว่า 485 คน ต้องการมีที่อยู่เป็นของตนเองในเขตเทศบาล จงหาค่าประมาณแบบจุดของสัดส่วนของประชาชนที่ต้องการที่อยู่อาศัยในเขตเทศบาล

วิธีทำ สัดส่วนของประชาชนที่ต้องการที่อยู่อาศัยในเขตเทศบาลเป็น $\dfrac{485}{1000} = 0.485$

ดังนั้น ค่าประมาณแบบจุดของสัดส่วนของประชาชนที่ต้องการมีที่อยู่อาศัยในเขตเทศบาลเป็น 0.485

6.3 การประมาณค่าแบบจุดกรณีสองประชากร

การประมาณค่าแบบจุดในกรณีของสองประชากร เป็นการประมาณค่าแบบจุดของผลต่าง หรืออัตราส่วนของพารามิเตอร์ของสองประชากร โดยอาศัยค่าประมาณจากตัวอย่างที่สุ่มมาจากสองประชากร ในกรณีนี้จะพิจารณาการประมาณค่าแบบจุดกรณีสองประชากรของค่าเฉลี่ยและการประมาณค่าแบบจุดกรณีสองประชากรสัดส่วน

6.3.1 การประมาณค่าเฉลี่ยกรณีสองประชากร

การประมาณค่าแบบจุดของค่าเฉลี่ยกรณีสองประชากร เป็นการใช้ผลต่างของค่าเฉลี่ยที่ได้จากตัวอย่างที่มาจากสองประชากรที่สนใจ นำไปประมาณผลต่างของค่าเฉลี่ยของประชากร

ตัวอย่างที่ 6.3 นายแพทย์ของโรงพยาบาลแห่งหนึ่งต้องการทราบความแตกต่างระหว่างอายุเฉลี่ยของผู้ป่วยชายและผู้ป่วยหญิงที่เป็นโรคเบาหวาน ซึ่งเข้ามารับการรักษาพยาบาลที่โรงพยาบาลแห่งนี้ จึงเลือกผู้ป่วยชายและหญิงที่เข้ารับการรักษาที่โรงพยาบาลนี้มาเป็นตัวอย่าง 10 และ 12 ราย ตามลำดับ สอบถามอายุของผู้ป่วย (หน่วยเป็นปี) ปรากฏรายละเอียดดังนี้

ชาย	45	52	45	48	56	54	58	65	63	66		
หญิง	60	62	65	45	56	68	57	64	63	55	75	72

ค่าประมาณแบบจุดของผลต่างระหว่างอายุเฉลี่ยของผู้ป่วยชายและหญิงเป็นเท่าไร

วิธีทำ อายุเฉลี่ยของผู้ป่วยชาย

$$= \frac{45+52+45+48+56+54+58+65+63+66}{10}$$

= 55.2

อายุเฉลี่ยของผู้ป่วยหญิง

$$= \frac{60+62+65+45+56+68+57+64+63+55+75+72}{12}$$

= 61.8

ผลต่างของค่าเฉลี่ยของอายุผู้ป่วยชายและผู้ป่วยหญิง

= 61.8 – 55.2

= 6.6

ดังนั้น ค่าประมาณความแตกต่างของอายุเฉลี่ยของผู้ป่วยชายและผู้ป่วยหญิงเป็น 6.6 ปี

นั่นคือ อายุเฉลี่ยของผู้ป่วยโรคเบาหวานเพศชายกับเพศหญิงมีความแตกต่างกันประมาณ 6.6 ปี

6.3.2 การประมาณค่าสัดส่วนกรณีสองประชากร

การประมาณค่าแบบจุดของสัดส่วนกรณีสองประชากร เป็นการใช้ผลต่างสัดส่วนที่ได้จากตัวอย่างที่สุ่มมาจากสองประชากรที่สนใจ นำไปประมาณผลต่างของสัดส่วนของประชากร

ตัวอย่างที่ 6.4 จากการสำรวจความพึงพอใจการใช้เครื่องปรับอากาศยี่ห้อหนึ่งของพ่อบ้านกับแม่บ้านในชุมชนที่มีครอบครัวอาศัยอยู่ 300 ครอบครัว โดยสุ่มตัวอย่างมา 36 ครอบครัว พบว่ามีพ่อบ้านและแม่บ้านที่พึงพอใจเครื่องปรับอากาศยี่ห้อดังกล่าวจำนวน 9 คน และ 12 คน ตามลำดับ จงหาค่าประมาณแบบจุดของผลต่างระหว่างสัดส่วนของความพึงพอใจเครื่องปรับอากาศของพ่อบ้านและแม่บ้าน

วิธีทำ สัดส่วนความพึงพอใจการใช้เครื่องปรับอากาศของพ่อบ้าน $= \frac{9}{36}$

สัดส่วนความพึงพอใจการใช้เครื่องปรับอากาศของแม่บ้าน $= \frac{12}{36}$

ผลต่างสัดส่วนความพึงพอใจการใช้เครื่องปรับอากาศของพ่อบ้านและแม่บ้าน

$$= \frac{12}{36} - \frac{9}{36} = \frac{3}{36}$$

ดังนั้น ค่าประมาณแบบจุดของผลต่างระหว่างสัดส่วนของความพึงพอใจเครื่องปรับอากาศของพ่อบ้านและแม่บ้านเป็น $\frac{3}{36}$

นั่นคือ ประมาณได้ว่าความแตกต่างของสัดส่วนพ่อบ้านและแม่บ้านที่มีความพึงพอใจเครื่องปรับอากาศยี่ห้อนั้นเป็น $\frac{3}{36}$ หรือ $\frac{1}{12}$

6.4 การประมาณค่าแบบช่วง

การประมาณค่าแบบช่วงเป็นการใช้ช่วงตัวเลขประมาณพารามิเตอร์ของประชากร ช่วงตัวเลขซึ่งเรียกว่าช่วงความเชื่อมั่น (confidence interval) จะบอกค่าต่ำสุด ค่าสูงสุด ที่คาดว่าจะคลุมพารามิเตอร์นั้น ๆ และจะบอกระดับความเชื่อมั่นของการประมาณไว้ในรูป $P(L < \theta < U) = 1-\alpha$ ซึ่งหมายความว่าความน่าจะเป็นที่ θ จะอยู่ระหว่าง L และ U ด้วยความเชื่อมั่น $1-\alpha$ เมื่อ $0 < \alpha < 1$ นิยมเขียนระดับความเชื่อมั่นในรูป $100(1-\alpha)\%$ และเรียก $100(1-\alpha)\%$ ว่าระดับความเชื่อมั่นซึ่งเป็นตัวบ่งบอกถึงโอกาสที่ค่าประมาณที่คำนวณได้จะมีความเชื่อมั่นในการประมาณและบอกถึงโอกาสที่ค่าประมาณจะคลาดเคลื่อนไปจากค่าจริงของประชากรว่าเป็นเท่าใด เช่น ถ้า $1-\alpha = 0.90$ หมายความว่า ช่วงตัวเลขที่ใช้ประมาณมีโอกาสถูกต้อง 0.90 หรือ 90% หรือมีโอกาสผิดพลาด 0.10 หรือ 10% นั่นเอง ต่อไปนี้จะพิจารณาการประมาณค่าแบบช่วงกรณีประชากรเดียวและกรณีสองประชากร

บทนิยามที่ 6.1 ถ้าให้ L และ U เป็นฟังก์ชันของตัวอย่างสุ่ม $X_1, X_2, X_3, \ldots, X_n$ จากการแจกแจงแบบหนึ่งซึ่งไม่ทราบค่าพารามิเตอร์ ถ้า θ เป็นค่าที่ทำให้ $P(L \le \theta \le U) = 1-\alpha$, $0 < \alpha < 1$ แล้วจะเรียกช่วง (L, U) ว่าช่วงความเชื่อมั่นขนาด $100(1-\alpha)$ % ของพารามิเตอร์ และเรียก $1-\alpha$ ว่าสัมประสิทธิ์ความเชื่อมั่น (confidence coefficient)

6.4.1 การประมาณค่าแบบช่วงกรณีประชากรเดียว

การประมาณค่าแบบช่วงกรณีประชากรเดียว เป็นการใช้ตัวเลขสองตัวซึ่งเป็นค่าต่ำสุดและค่าสูงสุดของตัวเลขช่วงหนึ่งที่ต่อเนื่องกันไปประมาณพารามิเตอร์ของประชากรโดยระบุได้ว่าค่าพารามิเตอร์ที่สนใจจะอยู่ในช่วงตัวเลขที่กำหนดด้วยความเชื่อมั่นเท่าใด ในที่นี้จะกล่าวถึง การประมาณค่าแบบช่วงกรณีประชากรเดียวของค่าเฉลี่ยและสัดส่วนของประชากรเท่านั้น

6.4.1.1 การประมาณค่าแบบช่วงของค่าเฉลี่ยกรณีประชากรเดียว

การประมาณค่าแบบช่วงของค่าเฉลี่ยกรณีประชากรเดียว เป็นการใช้ช่วงตัวเลขช่วงหนึ่งที่คำนวณได้จากค่าเฉลี่ยของตัวอย่างไปประมาณพารามิเตอร์ของประชากร การประมาณแบบช่วงของค่าเฉลี่ยของประชากรมีหลายกรณีด้วยกัน ทั้งนี้ขึ้นอยู่กับว่าข้อมูลมีการแจกแจงปกติหรือไม่ และทราบค่าความแปรปรวนของประชากรหรือไม่ (เติมศรี ชำนิจารกิจ, 2544) ในการพิจารณาค่าประมาณแบบช่วงของค่าเฉลี่ยจะแบ่งเป็น 2 กรณี คือกรณีทราบค่าความแปรปรวนและกรณีไม่ทราบค่าความแปรปรวน

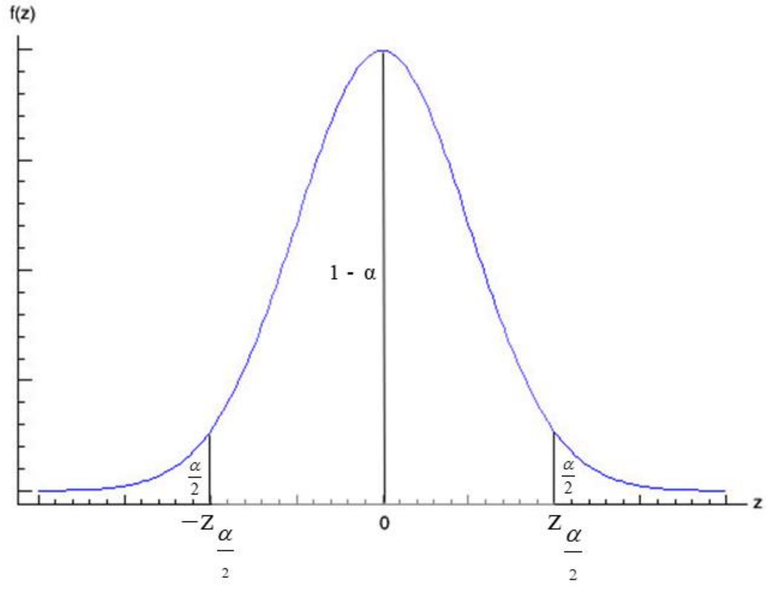
1) กรณีทราบค่าความแปรปรวน (σ^2) การประมาณค่าแบบช่วงของค่าเฉลี่ยกรณีประชากรเดียวกรณีนี้ จะใช้เมื่อตัวอย่างสุ่มมีขนาดใหญ่ ($n \geq 30$) หรือประชากรมีการแจกแจงแบบปกติแต่ตัวอย่างสุ่มมีขนาดเล็ก ($n < 30$)

ให้ $X_1, X_2, X_3, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปกติ $N(\mu, \sigma^2)$ จะได้ว่า $\overline{X}$ ซึ่งเป็นค่าเฉลี่ยของตัวอย่างจะมีการแจกแจงปกติที่มีค่าเฉลี่ย μ ความแปรปรวน $\frac{\sigma^2}{n}$

หรือกล่าวว่า $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ ดังนั้น $Z = \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

ให้ $0 \leq \alpha \leq 1$ และ $z_{\frac{\alpha}{2}}$ เป็นค่าที่ทำให้ $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$

ดังภาพที่ 6.1



ภาพที่ 6.1 ความน่าจะเป็น ที่ μ จะอยู่ระหว่าง $-z_{\frac{\alpha}{2}}$ กับ $z_{\frac{\alpha}{2}}$ เท่ากับ $1 - \alpha$

เมื่อแทนค่า $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ จะได้

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}} \right) = 1 - \alpha$$

คูณทุกพจน์ในวงเล็บด้วย $\dfrac{\sigma}{\sqrt{n}}$

$$P\left(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

บวกทุกพจน์ในวงเล็บด้วย $-\overline{X}$

$$P\left(-\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

คูณทุกพจน์ในวงเล็บด้วย –1 จะทำให้ เครื่องหมายอสมการเปลี่ยนเป็นตรงกันข้าม

$$P\left(\overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

เขียนใหม่เป็น

$$P\left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

ดังนั้น ช่วงความเชื่อมั่นขนาด 100(1 – α)% ของ μ คือ

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

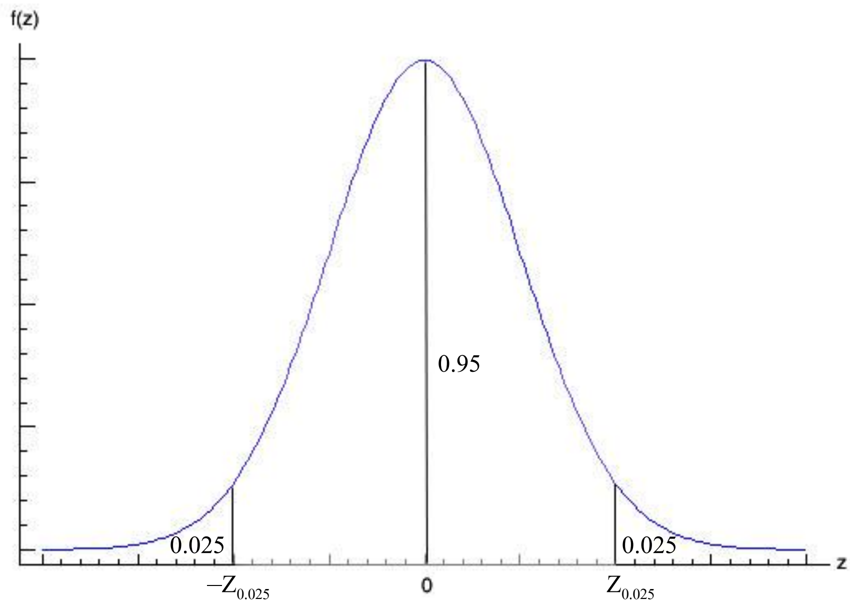
หรือ $$\overline{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

ตัวอย่างที่ 6.5 คะแนนสอบวิชาสถิติวิเคราะห์ของนักศึกษากลุ่มหนึ่งที่สุ่มมาจำนวน 49 คน มีค่าเฉลี่ยเป็น 63 ส่วนเบี่ยงเบนมาตรฐานเป็น 8 จากคะแนนเต็ม 100 คะแนน จงสร้างช่วงความเชื่อมั่นขนาด 95% และ 99% ของค่าเฉลี่ยของคะแนนสอบวิชานี้

วิธีทำ 1. สร้างช่วงความเชื่อมั่นขนาด 95%

จาก $\overline{X} = 63$ $s = 8$ $n = 49$

$1 - \alpha = 0.95$ $\alpha = 0.05$ $z_{0.025} = 1.96$



ภาพที่ 6.2 ความน่าจะเป็นที่ μ จะอยู่ระหว่าง $-z_{0.025}$ กับ $z_{0.025}$ เท่ากับ 0.95

ช่วงความเชื่อมั่นขนาด 95% ของ μ คือ

$$63 - 1.96\frac{8}{\sqrt{49}} < \mu < 63 + 1.96\frac{8}{\sqrt{49}}$$

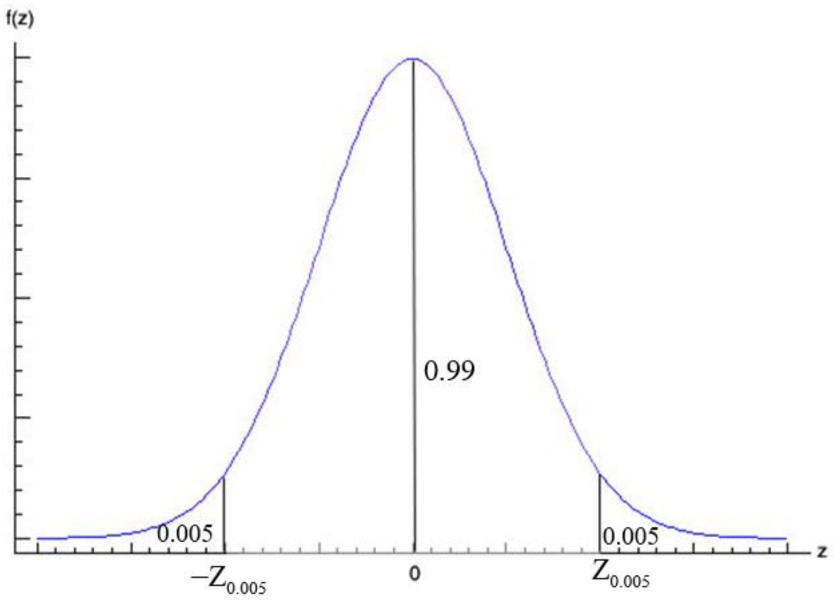
จะได้ $$63 - 2.24 < \mu < 63 + 2.24$$

$$60.76 < \mu < 65.24$$

2. สร้างความช่วงความเชื่อมั่นขนาด 99%

จาก $\overline{X}$ = 63 s = 8 n = 49

$1 - \alpha = 0.99$ $\alpha = 0.01$ $z_{0.005} = 2.575$



ภาพที่ 6.3 ความน่าจะเป็นที่ μ จะอยู่ระหว่าง $-z_{0.005}$ กับ $z_{0.005}$ เท่ากับ 0.99

ช่วงความเชื่อมั่นขนาด 99% ของ μ คือ

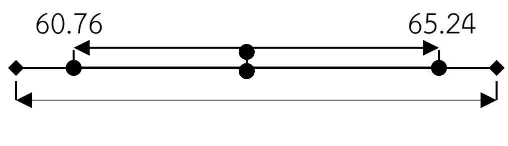
$$63 - 2.575\frac{8}{\sqrt{49}} < \mu < 63 + 2.575\frac{8}{\sqrt{49}}$$

จะได้

$$63 - 2.94 < \mu < 63 + 2.94$$

$$60.06 < \mu < 65.94$$

หากพิจารณาเปรียบเทียบช่วงความเชื่อมั่นขนาด 95% กับ 99% พบว่า ช่วงที่มีระดับความเชื่อมั่นสูงกว่า จะมีช่วงการประมาณกว้างกว่าซึ่งแสดงให้เห็นว่าการใช้ช่วงความเชื่อมั่นที่กว้างกว่าทำให้ระดับความเชื่อมั่นสูงกว่าและมีโอกาสคลาดเคลื่อนน้อยกว่า แสดงได้ดังภาพที่ 6.4



60.06 $\overline{X} = 63$ 65.94

ภาพที่ 6.4 ช่วงการประมาณขนาด 95% และ 99% ของคะแนนเฉลี่ย

จากตัวอย่างข้างต้นจะเห็นว่า ช่วงความเชื่อมั่นขนาด 100(1 – α)% จะคลุมค่าประมาณแบบจุด คือ $\overline{X} = 63$ หากค่า μ อยู่ ณ ตำแหน่งกึ่งกลางของช่วงพอดี จะได้ว่า $\mu = \overline{X}$ นั่นคือ ใช้ $\overline{X}$ ประมาณ μ โดยไม่มีค่าคลาดเคลื่อน ตามปกติแล้วเหตุการณ์อย่างนี้เกิดขึ้นได้ยาก จึงทำให้มีค่าคลาดเคลื่อนในการประมาณแบบจุดอยู่ แต่หากพิจารณาเทอม $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ จะพบว่าค่าคลาดเคลื่อนหรือความแตกต่างระหว่าง $\overline{X}$ กับ μ จะน้อยกว่า $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ และอาจกล่าวได้ว่าความแตกต่างระหว่าง $\overline{X}$ กับ μ จะขึ้นอยู่กับค่า n นั่นเอง

ทฤษฎีบทที่ 6.1 ถ้าใช้ $\overline{X}$ เป็นตัวประมาณค่าของ μ แล้ว เชื่อมั่นได้ 100(1 – α)% ว่าค่าคลาดเคลื่อน จะน้อยกว่า $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

จากตัวอย่างที่ 6.1 เชื่อมั่นถึง 95% ว่าค่าเฉลี่ย $\overline{X} = 63$ แตกต่างไปจากค่าเฉลี่ย μ แท้จริงโดยรวมแล้วน้อยกว่า 2.24 ในขณะเดียวกันจะแตกต่างไปจากค่า μ ที่แท้จริงน้อยกว่า 2.94 ด้วยความเชื่อมั่นถึง 99%

ทฤษฎีบทที่ 6.2 ถ้าให้ $\overline{X}$ เป็นตัวประมาณค่าของ μ แล้วเชื่อมั่นได้ 100(1 – α)% ว่าค่าคลาดเคลื่อนจะน้อยกว่า E ถ้าขนาดตัวอย่างเป็น $n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$

ตัวอย่างที่ 6.6 จากตัวอย่างที่ 6.5 ถ้าต้องการให้ค่าคลาดเคลื่อนในการประมาณไม่เกิน 2.5 ที่ระดับความเชื่อมั่น 95% และ 99% จะต้องใช้ตัวอย่างขนาดเท่าใด

วิธีทำ จาก $n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$

1. ที่ระดับความเชื่อมั่น 95%

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$\sigma = 8$

$E = 2.2$

$Z_{0.025} = 1.96$

$n = \left(\frac{(1.96)(8)}{2.5}\right)^2$

$= 39.34$

≈ 40

2. ที่ระดับความเชื่อมั่น 99%

$\alpha = 0.01$

$\frac{\alpha}{2} = 0.05$

$\sigma = 8$

$E = 2.94$

$Z_{0.005} = 2.575$

$n = \left(\frac{(2.575)(8)}{2.5}\right)^2$

$= 67.89$

≈ 68

นั่นคือ หากต้องการให้ค่าคลาดเคลื่อนในการประมาณไม่เกิน 2.5 ต้องใช้ขนาดตัวอย่างเป็น 40 ที่ระดับความเชื่อมั่น 95% และใช้ขนาดตัวอย่างขนาด 68 ที่ระดับความเชื่อมั่น 99%

2) กรณีไม่ทราบความแปรปรวน การประมาณค่าแบบช่วงสำหรับกรณีที่ไม่ทราบค่าความแปรปรวนของประชากร อาจเนื่องจากไม่มีการศึกษาในเรื่องนั้น ๆ มาก่อน การ

ประมาณค่ากรณีนี้ จะแบ่งเป็น 2 ประเด็น ได้แก่ ประเด็นตัวอย่างขนาดใหญ่ และประเด็นตัวอย่างขนาดเล็ก

2.1) ตัวอย่างขนาดใหญ่ ($n \geq 30$) โดยทฤษฎีขีดจำกัดสู่ส่วนกลาง สามารถใช้ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง (s) ประมาณส่วนเบี่ยงเบนมาตรฐานประชากร σ^2 จะได้ช่วงความเชื่อมั่นขนาด $(1 - \alpha)100\%$ ของค่าเฉลี่ย (μ) เป็น

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

ตัวอย่างที่ 6.7 ผลการวัดปริมาณการใช้กระแสไฟฟ้าของครัวเรือนในชุมชนหนึ่ง จำนวน 45 ครัวเรือน เป็นดังนี้

45	67	87	59	60	68	73	46	73	72	83	83	91	80	54
64	76	58	67	43	56	66	75	88	58	78	77	90	53	93
56	71	89	55	76	85	80	55	76	83	77	95	74	93	70

จงหาค่าประมาณช่วงความเชื่อมั่นขนาด 95 % ของค่าเฉลี่ยปริมาณการใช้กระแสไฟฟ้าในชุมชนนี้

วิธีทำ ต้องการประมาณค่าเฉลี่ย (μ) เมื่อตัวอย่างขนาดใหญ่ ($n > 30$) แต่ไม่ทราบค่าส่วนเบี่ยงเบนมาตรฐาน (σ) จากข้อมูลข้างต้น $n = 45$, $\overline{X} = 71.83$, $s = 13.98$ จะได้ว่า

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$71.83 - z_{0.025} \frac{13.98}{\sqrt{45}} < \mu < 71.83 + z_{0.025} \frac{13.98}{\sqrt{45}}$$

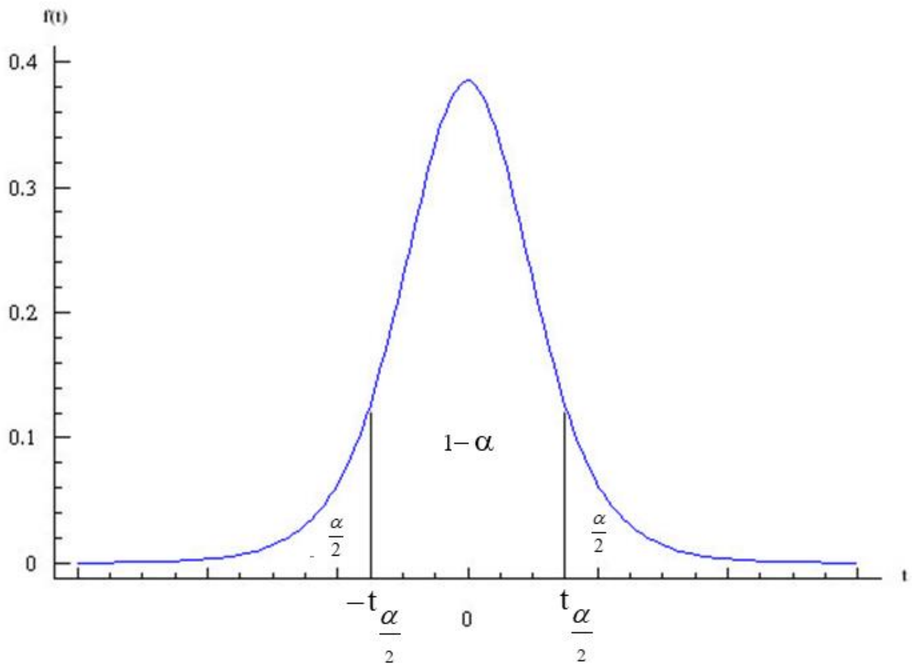
$$71.83 - 1.96 \frac{13.98}{\sqrt{45}} < \mu < 71.83 + 1.96 \frac{13.98}{\sqrt{45}}$$

$$71.83 - 4.08 < \mu < 71.83 + 4.08$$

$$67.75 < \mu < 75.46$$

ดังนั้น เชื่อมั่นได้ 95% ว่าค่าเฉลี่ยของปริมาณการใช้กระแสไฟฟ้ามีค่าอยู่ระหว่าง 67.75 ถึง 75.46 หน่วย

2.2) ตัวอย่างขนาดเล็ก (n < 30) การประมาณค่าแบบช่วงกรณีที่ไม่ทราบค่าความแปรปรวนและตัวอย่างมีขนาดเล็ก การสร้างช่วงความเชื่อมั่นจะอาศัยการแจกแจงของ T เมื่อ $T = \dfrac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ ซึ่งวิธีการนี้ดำเนินการเช่นเดียวกันกับกรณีตัวอย่างสุ่มมีขนาดใหญ่แต่จะใช้การแจกแจงแบบ t แทนการแจกแจงแบบปกติมาตรฐาน ดังภาพที่ 6.5



ภาพที่ 6.5 ความน่าจะเป็นที่ T จะอยู่ระหว่าง $-t_{\frac{\alpha}{2}}$ กับ $t_{\frac{\alpha}{2}}$ เท่ากับ $1 - \alpha$

จากภาพที่ 6.5 จะได้ $$P\left(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

เมื่อ $t_{\frac{\alpha}{2}}$ คือค่า t ที่ระดับขั้นความเสรี (ν) = n – 1 แทนที่ $T = \dfrac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ จะได้ว่า

$1-\alpha$

$\frac{\alpha}{2}$ $\frac{\alpha}{2}$

$-t_{\frac{\alpha}{2}}$ $t_{\frac{\alpha}{2}}$

$$P\left(-t_{\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} < t_{\frac{\alpha}{2}}\right) = 1-\alpha$$

คูณทุกพจน์ด้วย $\frac{s}{\sqrt{n}}$ และบวกทุกพจน์ด้วย $-\overline{X}$

$$P\left(-\overline{X}-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < -\overline{X}+t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

คูณทุกพจน์ด้วย -1

$$P\left(\overline{X}+t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} > \mu > \overline{X}-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

เขียนใหม่ได้เป็น

$$P\left(\overline{X}-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \overline{X}+t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

ดังนั้น ช่วงความเชื่อมั่นขนาด $100(1-\alpha)\%$ ของ μ คือ

$$\overline{X}-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} < \mu < \overline{X}+t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

หรือ $$\overline{X} \pm t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}$$

ตัวอย่างที่ 6.8 จากการสอบถามค่าใช้จ่ายเกี่ยวกับเอกสารประกอบการเรียนของนักศึกษา จำนวน 7 คน ในภาคการศึกษาหนึ่ง ปรากฏว่านักศึกษามีค่าใช้จ่าย ดังนี้

850 900 750 1200 1000 900 1500

จงสร้างช่วงความเชื่อมั่นขนาด 95% สำหรับค่าใช้จ่ายเฉลี่ยเกี่ยวกับเอกสารประกอบการเรียนของนักศึกษากลุ่มนี้ สมมติว่าค่าใช้จ่ายมีการแจกแจงปกติ

วิธีทำ

$$\bar{X} = \frac{\sum_{i=1}^{7} x_i}{n}$$

$$= \frac{850+900+750+1200+1000+900+1500}{7}$$

$$= 1014.286$$

$$s = 256.116$$

$$t_{\frac{\alpha}{2},\nu} = t_{0.025,\,6}$$

$$= 2.447$$

ช่วงความเชื่อมั่นขนาด 95% ของ μ คือ

$$1014.286 - 2.447\frac{256.116}{\sqrt{7}} < \mu < 1014.286 + 2.447\frac{256.116}{\sqrt{7}}$$

$$1014.286 - 236.876 < \mu < 1014.286 + 236.876$$

$$777.410 < \mu < 1251.162$$

ดังนั้น เชื่อมั่นได้ 95% ว่าค่าใช้จ่ายเฉลี่ยเกี่ยวกับเอกสารประกอบการเรียนของนักศึกษาอยู่ในช่วง 777.41 ถึง 1251.16 บาท

6.4.1.2 การประมาณค่าแบบช่วงของสัดส่วนกรณีประชากรเดียว

พิจารณาเมื่อชักตัวอย่างขนาด n จากประชากรที่มีการแจกแจงทวินาม เมื่อให้ X แทนการประสบความสำเร็จที่เกิดจากการชักตัวอย่าง สัดส่วนของความสำเร็จของตัวอย่างจะเป็น $\hat{P} = \frac{X}{n}$ ซึ่งเป็นค่าประมาณแบบจุดสำหรับพารามิเตอร์ P ในกรณีที่ตัวอย่างมีขนาดใหญ่ (n > 30) สามารถประมาณการแจกแจงทวินามด้วยการแจกแจงแบบปกติที่มีค่าเฉลี่ย $\mu_{\hat{P}}$ ความแปรปรวน $\sigma^2_{\hat{P}}$

โดยที่ $\mu_{\hat{p}} = E(\hat{P}) \quad = E\left(\frac{X}{n}\right) \quad = \frac{1}{n}E(X) \quad = \frac{1}{n}E(np) \quad = \quad p$

$$\sigma^2_{\hat{p}} = V\left(\frac{X}{n}\right) \quad = \frac{1}{n^2}V(X) = \frac{1}{n^2}(npq) = \frac{pq}{n} \quad = \frac{p(1-p)}{n}$$

แต่เนื่องจากไม่ทราบค่าของพารามิเตอร์ p จะแทน p ด้วย $\hat{p}$ จะได้ค่าเฉลี่ย $\mu_{\hat{p}} = \hat{p}$ และความแปรปรวน $\sigma^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n}$ และช่วงความเชื่อมั่นขนาด $(1-\alpha)100\%$ ของสัดส่วน p คือ

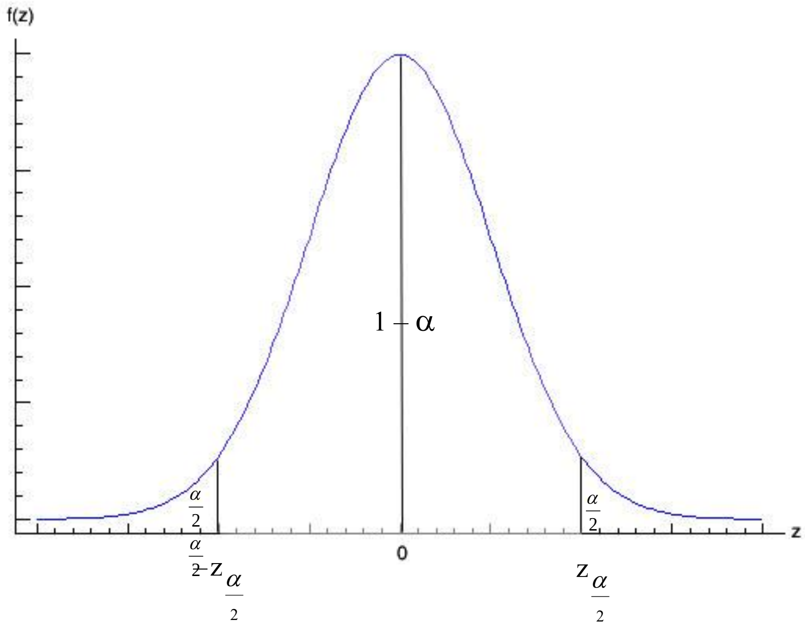
$$\hat{p} - Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ทฤษฎีบทที่ 6.3 ถ้าตัวอย่างสุ่มขนาด n ที่ได้จากประชากรที่มีการแจกแจงแบบทวินามด้วยค่าเฉลี่ย μ = np และความแปรปรวน σ^2 = npq แล้วการแจกแจงของการสุ่มตัวอย่างของ $\hat{p}$ ซึ่งแทนสัดส่วนของการประสบผลสำเร็จ จะประมาณได้ด้วยการแจกแจงแบบปกติด้วยค่าเฉลี่ย $\mu_{\hat{p}} = p$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_p = \sqrt{pq/n}$ ดังนั้น $z = \frac{\hat{p}-p}{\sqrt{pq/n}}$ เมื่อ z คือค่าของตัวแปรสุ่ม Z ที่มีการแจกแจงปกติมาตรฐาน

ตัวประมาณแบบจุดที่มีประสิทธิภาพที่สุดของสัดส่วน p ในการทดลองแบบทวินาม คือตัวสถิติ $\hat{p} = \frac{X}{n}$ ดังนั้น สัดส่วนของตัวอย่าง $\hat{p} = \frac{X}{n}$ จะใช้ในการประมาณพารามิเตอร์ p ถ้าสัดส่วน p ซึ่งเป็นตัวไม่ทราบค่า $0 < p < 1$ สามารถที่จะสร้างช่วงความเชื่อมั่นสำหรับ $\hat{p}$ โดยพิจารณาการแจกแจงการสุ่มตัวอย่างของ $\hat{p}$ ซึ่งจะมีการแจกแจงใกล้เคียงการแจกแจงปกติด้วยค่าเฉลี่ย $\mu_{\hat{p}} = p$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$ ดังนั้น สามารถยืนยันได้ว่า

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

เมื่อ $Z = \dfrac{\hat{P} - P}{\sqrt{\dfrac{pq}{n}}}$ และ $z_{\frac{\alpha}{2}}$ คือค่าของ Z ที่มีพื้นที่ใต้โค้งของการแจกแจงปกติมาตรฐานอยู่ทางขวาของค่านี้ $\dfrac{\alpha}{2}$ ดังภาพที่ 6.6



ภาพที่ 6.6 ความน่าจะเป็นที่ Z จะอยู่ระหว่าง $-z_{\frac{\alpha}{2}}$ กับ $z_{\frac{\alpha}{2}}$ เท่ากับ $1 - \alpha$

ถ้าแทนค่า Z ลงในอสมการจะได้

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\hat{P} - p}{\sqrt{\frac{pq}{n}}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

คูณทุกพจน์ในอสมการด้วย $\sqrt{\dfrac{pq}{n}}$

$$P\left(-z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < \hat{P} - p < z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

ลบทุกพจน์ในอสมการด้วย $\hat{P}$

$$P\left(-\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < -p < -\hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

คูณทุกพจน์ในอสมการด้วย -1

$$P\left(\hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} > p > \hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

และเขียนใหม่ได้เป็น

$$P\left(\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < p < \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

จะเห็นได้ว่า ไม่สามารถที่จะจัดเทอมต่าง ๆ ในอสมการให้ปราศจาก p ซึ่งเป็นตัวพารามิเตอร์ที่ไม่ทราบค่า แต่จะใช้ตัวประมาณแบบจุด $\hat{p} = \frac{x}{n}$ ในการประมาณ p ซึ่งเมื่อชักตัวอย่างขนาดใหญ่จะเกิดความคลาดเคลื่อนเพียงเล็กน้อยเท่านั้น (Walpole, 1971) ดังนั้น จะได้ว่า

$$P\left(\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 1 - \alpha$$

หากชักตัวอย่างขนาด n สามารถคำนวณค่าสัดส่วนของตัวอย่าง $\hat{p} = \frac{x}{n}$ และจะประมาณช่วงความเชื่อมั่นขนาด $100(1 - \alpha)\%$ สำหรับ p ได้ด้วย

$$\left(\hat{P} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

ดังนั้น ช่วงความเชื่อมั่นขนาด 100(1 – α)% สำหรับพารามิเตอร์ p จากการแจกแจงทวินามประมาณได้ด้วย

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

เมื่อ $\hat{p}$ คือสัดส่วนของการประสบความสำเร็จ $\hat{q} = 1 - \hat{p}$

$\frac{\alpha}{2}$ คือพื้นที่ใต้โค้งการแจกแจงปกติมาตรฐานที่อยู่ทางขวาของ $z_{\frac{\alpha}{2}}$

ตัวอย่างที่ 6.9 ในการผลิตโทรทัศน์แต่ละครั้งของโรงงานผลิตโทรทัศน์แห่งหนึ่ง พบโทรทัศน์มีตำหนิ 12% จึงทำการตรวจสอบโดยสุ่มโทรทัศน์มา 50 เครื่อง พบ 7 เครื่อง มีตำหนิ อยากทราบช่วงความเชื่อมั่นขนาด 95% สำหรับสัดส่วนโทรทัศน์ที่มีตำหนิ

วิธีทำ สัดส่วนของโทรทัศน์ที่มีตำหนิ $\hat{p} = \frac{7}{50} = 0.14$, $\hat{q} = 1 - 0.14 = 0.86$

$$z_{\frac{\alpha}{2}} = z_{0.25} = 1.96$$

ช่วงความเชื่อมั่นขนาด 95% สำหรับสัดส่วนโทรทัศน์ที่มีตำหนิ คือ

$$0.14 - 1.96\sqrt{\frac{(0.14)(0.86)}{50}} < p < 0.14 + 1.96\sqrt{\frac{(0.14)(0.86)}{50}}$$

$$0.14 - 0.10 < p < 0.14 + 0.10$$

$$0.04 < p < 0.24$$

ดังนั้น เชื่อมั่นได้ 95% ว่าสัดส่วนของโทรทัศน์ที่มีตำหนิจะอยู่ระหว่าง 0.04 ถึง 0.24

ตัวอย่างที่ 6.10 ในการสำรวจความคิดเห็นเกี่ยวกับการอนุญาตให้มีเครื่องบินส่วนตัว ผู้สำรวจได้สุ่มเลือกประชาชน 400 คน จากอาชีพ เพศ และวัยต่าง ๆ กัน ในจำนวนนี้มีอยู่ 140 คน ที่เห็นด้วยกับการอนุญาตให้มีเครื่องบินส่วนตัว จงหาช่วงความเชื่อมั่นขนาด 95% ของสัดส่วนของประชาชนที่เห็นด้วยกับการอนุญาตให้มีเครื่องบินส่วนตัว

วิธีทำ สัดส่วนของประชาชนที่เห็นด้วยกับการอนุญาตให้มีเครื่องบินส่วนตัว ($\hat{p}$)

$$\hat{p} = \frac{140}{400} = 0.35, \quad \hat{q} = 1 - \hat{p} = 1 - 0.35 = 0.65$$

ช่วงความเชื่อมั่นขนาด 95% ของ p เป็น

$$0.35 - 1.96\sqrt{\frac{0.35(0.65)}{400}} < p < 0.35 + 1.96\sqrt{\frac{0.35(0.65)}{400}}$$

$$0.35 - 1.96(0.0238) < p < 0.35+1.96(0.0238)$$

$$0.35 - 0.0467 < p < 0.35 + 0.0467$$

$$0.3033 < p < 0.3967$$

ดังนั้น เชื่อมั่นได้ 95% ว่าสัดส่วนของประชาชนที่เห็นด้วยกับการอนุญาตให้มีเครื่องบินส่วนตัวอยู่ระหว่าง 0.303 กับ 0.397

ตัวอย่างที่ 6.11 ในการสุ่มหลอดไฟจำนวน 75 หลอด พบว่าหลอดไฟจำนวน 5 หลอด มีอายุการใช้งานต่ำกว่าเกณฑ์มาตรฐาน จงหาช่วงความเชื่อมั่นขนาด 95% ของสัดส่วนหลอดไฟที่ไม่ได้มาตรฐาน

วิธีทำ สัดส่วนหลอดไฟที่ไม่ได้มาตรฐาน ($\hat{p}$)

ค่าประมาณแบบจุดของสัดส่วนหลอดไฟไม่ได้มาตรฐานเป็น

$$\hat{p} = \frac{x}{n} = \frac{5}{75} = 0.067$$

จะได้ช่วงความเชื่อมั่นขนาด 95% ของสัดส่วนหลอดไฟที่ไม่ได้มาตรฐานเป็น

$$0.067 - z_{0.025}\sqrt{\frac{0.067(1-0.067)}{75}} < p < 0.067 + z_{0.025}\sqrt{\frac{0.067(1-0.067)}{75}}$$

$$0.067 - 1.96\sqrt{\frac{0.063}{75}} < p < 0.067 + 1.96\sqrt{\frac{0.063}{75}}$$

$$0.067 - 1.96(0.028) < p < 0.067 + 1.96(0.028)$$

$$0.067 - 0.0548 < p < 0.067 + 0.0548$$

$$0.0122 < p < 0.1218$$

นั่นคือ สัดส่วนของหลอดไฟที่ไม่ได้มาตรฐานอยู่ในช่วง 0.012 ถึง 0.122

6.6 การประมาณค่าแบบช่วงกรณีสองประชากร

การประมาณค่าแบบช่วงกรณีสองประชากร เป็นการใช้ช่วงตัวเลขที่ประมาณได้จากตัวอย่างสุ่มไปประมาณผลต่างหรืออัตราส่วนของพารามิเตอร์ของประชากร (ศิริชัย กาญจนวาสี, 2547: Dudewicz, 1988) จะกล่าวถึงการประมาณค่าแบบช่วงกรณีสองประชากรของค่าเฉลี่ยและการประมาณค่าแบบช่วงกรณีสองประชากรของสัดส่วน

6.6.1 การประมาณผลต่างค่าเฉลี่ยระหว่างสองประชากร

การประมาณค่าแบบช่วงของค่าเฉลี่ยในกรณีนี้ เป็นการใช้ค่าเฉลี่ยจากตัวอย่างที่สุ่มมาจากสองประชากรไปประมาณค่าเฉลี่ยของประชากร ในการประมาณค่าแบบช่วงกรณีสองประชากรของค่าเฉลี่ยจะพิจารณาการประมาณค่าผลต่างระหว่างค่าเฉลี่ยซึ่งแบ่งเป็น 3 กรณี ได้แก่ 1) กรณีประชากรทั้งสองมีการแจกแจงปกติและทราบความแปรปรวน 2) กรณีประชากรทั้งสองมีการแจกแจงใด ๆ และตัวอย่างทั้งสองมีขนาดใหญ่ 3) กรณีประชากรทั้งสองมีการแจกแจงปกติหรือใกล้เคียงการแจกแจงปกติตัวอย่างทั้งสองมีขนาดเล็ก และไม่ทราบความแปรปรวน และ 4) กรณีการประมาณค่าผลต่างระหว่างค่าเฉลี่ยสำหรับข้อมูลคู่

6.6.1.1 ประชากรทั้งสองมีการแจกแจงปกติและทราบความแปรปรวน เป็นการประมาณช่วงผลต่างของค่าเฉลี่ยสองประชากร โดยทราบความแปรปรวนของทั้งสองประชากร ช่วงความเชื่อมั่นขนาด (1 – α) 100% ของผลต่างระหว่างค่าเฉลี่ย ($\mu_1 - \mu_2$) คือ

$$(\overline{X}_1 - \overline{X}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ตัวอย่างที่ 6.12 สุ่มเครื่องสำรองไฟยี่ห้อ A ขนาด 500 โวลต์ มา 100 เครื่อง มีอายุการใช้งานเฉลี่ย 8,640 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 220 ชั่วโมง และสุ่มเครื่องสำรองไฟยี่ห้อ B ขนาดเดียวกัน มา 200 เครื่อง มีอายุการใช้งานเฉลี่ย 8,440 ชั่วโมง ส่วนเบี่ยงเบนมาตรฐาน 180 ชั่วโมง เมื่อทราบว่าอายุการใช้งานของเครื่องสำรองไฟมีการแจกแจงปกติ จงหาช่วงความเชื่อมั่นขนาด 99% ของผลต่างอายุการใช้งานเฉลี่ยของเครื่องสำรองไฟยี่ห้อ A และยี่ห้อ B

วิธีทำ สุ่มเครื่องสำรองไฟยี่ห้อ A

จำนวน (n_A)	100	เครื่อง
อายุการใช้งานเฉลี่ย ($\overline{X}_A$)	8640	ชั่วโมง
ส่วนเบี่ยงเบนมาตรฐาน (σ_A)	220	ชั่วโมง

ความแปรปรวน (σ^2_A) $220^2 = 48400$ ชั่วโมง^2

สุ่มเครื่องสำรองไฟยี่ห้อ B

จำนวน (n_B)	200	เครื่อง
อายุการใช้งานเฉลี่ย ($\overline{X}_B$)	8440	ชั่วโมง
ส่วนเบี่ยงเบนมาตรฐาน (σ_B)	180	ชั่วโมง
ความแปรปรวน (σ^2_B)	$180^2 = 32400$	ชั่วโมง^2

ช่วงความเชื่อมั่นขนาด 99% ของผลต่างของอายุการใช้งานเฉลี่ยของเครื่องสำรองไฟยี่ห้อ A และยี่ห้อ B ($\mu_A - \mu_B$) เป็น

$$(\bar{X}_A - \bar{X}_B) - Z_{0.005}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} < \mu_A - \mu_B < (\bar{X}_A - \bar{X}_B) + Z_{0.005}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$(8640 - 8440) - 2.58\sqrt{\frac{48400}{100} + \frac{32400}{200}} < \mu_A - \mu_B < (8640 - 8440) + 2.58\sqrt{\frac{48400}{100} + \frac{32400}{200}}$$

$$200 - 2.58\sqrt{484 + 162} < \mu_A - \mu_B < 200 + 2.58\sqrt{484 + 162}$$

$$200 - 2.58(25.42) < \mu_A - \mu_B < 200 + 2.58(25.42)$$

$$200 - 65.57 < \mu_A - \mu_B < 200 + 65.57$$

$$134.43 < \mu_A - \mu_B < 265.57$$

ดังนั้น เชื่อมั่นได้ 99% ว่าช่วงความเชื่อมั่นของผลต่างของค่าเฉลี่ยของอายุการใช้งานของเครื่องสำรองไฟ A และ B อยู่ระหว่าง 134.43 ถึง 265.57 ชั่วโมง

6.6.1.2 ประชากรทั้งสองมีการแจกแจงใด ๆ และตัวอย่างทั้งสองมีขนาดใหญ่

พิจารณา 2 กรณี คือ กรณีทราบความแปรปรวน และกรณีไม่ทราบความแปรปรวน

1) ทราบความแปรปรวน ช่วงความเชื่อมั่นขนาด (1 – α) 100% ของผลต่างระหว่างค่าเฉลี่ย ($\mu_1 - \mu_2$) คือ

$$\left(\overline{X}_1 - \overline{X}_2\right) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2) ไม่ทราบความแปรปรวน ประชากรทั้งสองมีการแจกแจงใด ๆ และตัวอย่างทั้งสองมีขนาดใหญ่ โดยไม่ทราบ σ^2_1 และ σ^2_2 จะประมาณด้วย S^2_1 และ S^2_2 จะได้ช่วงความเชื่อมั่นขนาด (1-α)100% ของผลต่างระหว่างค่าเฉลี่ย ($\mu_1 - \mu_2$) คือ

$$(\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

ตัวอย่าง

6.6.1.3 ประชากรทั้งสองมีการแจกแจงปกติหรือใกล้เคียงการแจกแจงปกติตัวอย่างทั้งสองมีขนาดเล็ก และไม่ทราบความแปรปรวน... กรณีไม่ทราบค่าความแปรปรวน ~~การประมาณค่าแบบช่วงของค่าเฉลี่ยสองประชากรกรณีนี้~~ ประชากรมีการแจกแจงปกติ~~แต่~~สุ่มตัวอย่างขนาดเล็ก การประมาณช่วงความ เชื่อมั่นขนาด (1 – α) 100% ของผลต่างระหว่างค่าเฉลี่ย ($\mu_1 - \mu_2$) แบ่งเป็น 3 กรณี ได้แก่ 1กรณีประชากรมีความแปรปรวนเท่ากัน 2กรณีความแปรปรวนไม่เท่ากัน และ3การประมาณค่าผลต่างระหว่างค่าเฉลี่ยสำหรับข้อมูลคู่ (paired data)

1) กรณีไม่ทราบค่าความแปรปรวน แต่ประชากรมีความแปรปรวนเท่ากัน ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) ช่วงความเชื่อมั่นขนาด (1 – α) 100% ของผลต่างระหว่างค่าเฉลี่ย ($\mu_1 - \mu_2$) คือ

$$(\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2}, n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2}, n_1+n_2-2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

เมื่อ S_p คือส่วนเบี่ยงเบนมาตรฐานตัวอย่างร่วม (pooled sample standard deviation) โดยที่

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

ที่ระดับขั้นความเสรีเท่ากับ $n_1 + n_2 - 2$

ตัวอย่างที่ 6.13 ต้องการเปรียบเทียบสารตกค้างในน้ำยาล้างจาน 2 ชนิด ชนิดที่ 1 กับชนิดที่ 2 จึงได้ทำการสุ่มน้ำยาล้างจานชนิดที่ 1 จำนวน 10 ขวด และชนิดที่ 2 จำนวน 8 ขวด เพื่อตรวจหาปริมาณสารตกค้าง พบว่าปริมาณสารตกค้างในน้ำยาล้างจานชนิดที่ 1 มีค่าเฉลี่ยเป็น 3.1 มิลลิกรัม ส่วนเบี่ยงเบนมาตรฐานเป็น 0.5 มิลลิกรัม และชนิดที่ 2 มีค่าเฉลี่ยเป็น 2.4 มิลลิกรัม ส่วนเบี่ยงเบน

มาตรฐานเป็น 0.7 มิลลิกรัม จงหาช่วงความเชื่อมั่นขนาด 95% ของความแตกต่างระหว่างค่าเฉลี่ยของปริมาณสารตกค้างในน้ำยาล้างจานสองชนิดนี้

วิธีทำ น้ำยาล้างจานชนิดที่ 1

ขนาดตัวอย่าง (n_1) = 10 ขวด

ค่าเฉลี่ย ($\overline{X}_1$) = 3.1 มิลลิกรัม

ส่วนเบี่ยงเบนมาตรฐาน (S_1) = 0.5 มิลลิกรัม

น้ำยาล้างจานชนิดที่ 2

ขนาดตัวอย่าง (n_2) = 8 ขวด

ค่าเฉลี่ย ($\overline{X}_2$) = 2.4 มิลลิกรัม

ส่วนเบี่ยงเบนมาตรฐาน (S_2) = 0.7 มิลลิกรัม

ส่วนเบี่ยงเบนมาตรฐานตัวอย่างร่วม

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

$$S_p = \sqrt{\frac{(10-1)(0.5)^2 + (8-1)(0.7)^2}{10+8-2}}$$

$$= \sqrt{\frac{(9)(0.25)+(7)(0.49)}{16}}$$

$$= \sqrt{\frac{2.25+3.43}{16}}$$

$$= 0.60$$

ระดับขั้นความเสรีเท่ากับ

$$n_1 + n_2 - 2 = 10+8-2$$

$$= 16$$

ช่วงความเชื่อมั่นขนาด 95% ของความแตกต่างระหว่างค่าเฉลี่ยของปริมาณสารตกค้างในน้ำยาล้างจานชนิดที่ 1 และชนิดที่ 2 เป็น

$$(3.1-2.4) - t_{0.025,\,16}(0.60)\sqrt{\frac{1}{10}+\frac{1}{8}} < \mu_1 - \mu_2 < (3.1 - 2.4) + t_{0.025,\,16}(0.60)\sqrt{\frac{1}{10}+\frac{1}{8}}$$

$$(0.70) - (2.12)(0.60)(0.47) < \mu_1 - \mu_2 < (0.70) + (2.12)(0.60)(0.47)$$

$$(0.70) - (0.5978) < \mu_1 - \mu_2 < (0.70) + (0.5978)$$
$$(0.1022) < \mu_1 - \mu_2 < (1.2978)$$

นั่นคือ เชื่อมั่นได้ 95% ว่าความแตกต่างระหว่างปริมาณสารตกค้างเฉลี่ยในน้ำยาล้างจานชนิดที่ 1 และชนิดที่ 2 อยู่ในช่วง 0.10 มิลลิกรัม ถึง 1.30 มิลลิกรัม

2) **กรณีไม่ทราบค่าความแปรปรวน แต่ประชากรมีความแปรปรวนไม่เท่ากัน** $(\sigma_1^2 \neq \sigma_2^2)$ ช่วงความเชื่อมั่นขนาด $(1 - \alpha)$ 100% ของผลต่างระหว่างค่าเฉลี่ย $(\mu_1 - \mu_2)$ คือ

$$(\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

หรือเขียนในรูปแบบ

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

โดยระดับขั้นความเสรี (degree of freedom: df) คำนวณจาก

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

ตัวอย่างที่ 6.14 ถ้าสุ่มตัวอย่างขนาดเล็กจากสองประชากร จงหาช่วงความเชื่อมั่นขนาด 95 % ของ $\mu_1 - \mu_2$ เมื่อ $\bar{X}_1 = 3.1$ $s_1^2 = 0.25$ $n_1 = 10$ $\bar{X}_1 = 2.4$ $s_2^2 = 0.49$ $n_2 = 8$

วิธีทำ $\bar{X}_1 = 3.1$ $s_1^2 = 0.25$ $n_1 = 10$

$\bar{X}_2 = 2.4$ $s_2^2 = 0.49$ $n_2 = 8$

ระดับขั้นความเสรี คำนวณได้จาก

$$\nu = \frac{\left(\frac{02.5}{10}+\frac{0.49}{8}\right)^2}{\frac{\left(\frac{0.25}{10}\right)^2}{10-1}+\frac{\left(\frac{0.49}{10}\right)^2}{8-1}}$$

$$= \quad 18.037$$

จะได้ $\nu = 18$ โดยประมาณ

จากตาราง $t_{\frac{\alpha}{2},\nu} = t_{0.025,\,18} = 2.101$

ช่วงความเชื่อมั่นขนาด 95 % ของ $\mu_1 - \mu_2$ เป็น

$$\left(\overline{X}_1 - \overline{X}_2\right) - t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < \mu_1-\mu_2 < \left(\overline{X}_1 - \overline{X}_2\right) + t_{\frac{\alpha}{2},\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

หรือเขียนในรูปแบบ

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

จะได้

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{0.025,18}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \quad = \quad (3.1 - 2.4) \pm 2.101\sqrt{\frac{0.25}{10}+\frac{0.49}{8}}$$

$$= \quad 0.7 \pm 0.617$$

$$= \quad (0.083,\ 1.317)$$

หรือ $0.083 < \mu_1 - \mu_2 < 1.317$

ดังนั้น เชื่อมั่นได้ 95% ว่าความแตกต่างระหว่างค่าเฉลี่ยของประชากรที่กำหนดอยู่ในช่วง 0.083 ถึง 1.317

6.6.1.4 **กรณีการประมาณค่าผลต่างระหว่างค่าเฉลี่ยสำหรับข้อมูลคู่** เป็นการประมาณค่าผลต่างระหว่างค่าเฉลี่ยในกรณีที่ตัวอย่างสุ่ม 2 ชุด ที่ไม่เป็นอิสระต่อกัน เช่น อยากทราบว่าการใช้ยาลดน้ำหนักได้ผลมากน้อยเพียงใด โดยการเปรียบเทียบน้ำหนักก่อนและหลัง

การใช้ยา หรือการเปรียบเทียบผลการเรียนของนักศึกษาในปีสุดท้ายก่อนจบการศึกษากับปีแรกที่เข้าเป็นนักศึกษาว่าแตกต่างกันหรือไม่ การทดลองลักษณะนี้จะใช้ตัวอย่างขนาด n โดยแต่ละหน่วยตัวอย่างจะมีค่าสังเกต 2 ค่า เช่น น้ำหนักก่อนรับประทานยาลดน้ำหนักและน้ำหนักหลังรับประทานยาลดน้ำหนัก ค่าสังเกตที่ได้จากข้อมูลลักษณะนี้เรียกว่า ข้อมูลสังเกตชนิดคู่ (paired observations)

สมมติให้คู่อันดับ $(X_{11}, X_{21}), (X_{12}, X_{22}), \ldots ,(X_{1n}, X_{2n})$ เป็นตัวอย่างสุ่มขนาด n โดยที่ $X_{11}, X_{12}, X_{13}, \ldots, X_{1n}$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปกติที่มีค่าเฉลี่ย μ_1 และ $X_{21}, X_{22}, X_{23}, \ldots, X_{2n}$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปกติที่มีค่าเฉลี่ย μ_2 ตัวอย่างสุ่ม แต่ละคู่เป็นอิสระกับคู่อื่น ๆ กล่าวคือ (X_{1i}, X_{2i}) เป็นอิสระกับ (X_{1j}, X_{2j}) เมื่อ $i \neq j$ แต่ถ้าพิจารณาคู่ (X_{1i}, X_{2i}) ใด ๆ จะเห็นว่า X_{1i} ไม่เป็นอิสระต่อ X_{2i} ในแต่ละคู่ ถ้าให้

$$\begin{aligned} D_i &= X_{1i} - X_{2i} \quad , i = 1, 2, 3, \ldots, n \\ E(D_i) &= E(X_{1i}, X_{2i}) \\ &= E(X_{1i}) - E(X_{2i}) \\ &= \mu_1 - \mu_2 \end{aligned}$$

เนื่องจาก X_{1i} และ X_{2i} มีการแจกแจงปกติ ดังนั้น D_i จึงมีการแจกแจงปกติด้วย สำหรับความแปรปรวนของ D_i มีค่าเป็น

$$V(D_i) = V(X_{1i}) + V(X_{2i}) - 2\,\mathrm{Cov}(X_{1i} - X_{2i}) \text{ ซึ่งไม่ทราบค่า}$$

$$\mathrm{Cov}(X_{1i}, X_{2i}) = \text{ความแปรปรวนร่วมของ } X_{1i} \text{ และ } X_{2i}$$

แต่เนื่องจาก D_i เป็นผลต่างของ X_{1i} และ X_{2i} และเป็นตัวอย่างสุ่มชุดเดียวกัน จึงมีค่าเฉลี่ยที่เท่ากันและความแปรปรวนที่เท่ากันด้วย ดังนั้น D_i เมื่อ $i = 1, 2, 3, \ldots, n$ จึงมีค่าเฉลี่ยเท่ากันทั้งหมด และความแปรปรวนเท่ากันทั้งหมดด้วย (ทวีรัตนา ศิวดุลย์, 2539; ศุภชัย นาทะพันธ์, 2547) ดังนั้นจะให้ $E(D_i) = \mu_D$, $V(D_i) = \sigma_D^2$ จึงกล่าวได้ว่า $D_1, D_2, D_3, \ldots, D_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปกติ มีค่าเฉลี่ยของประชากรเป็น μ_D และไม่ทราบค่า σ_D^2 ซึ่งเป็นความแปรปรวนของประชากร ด้วยวิธีการเช่นเดียวกับการหาช่วงความเชื่อมั่นของค่าเฉลี่ย จะได้ว่าช่วงความเชื่อมั่นขนาด 95% ของ $\mu_D = \mu_1 - \mu_2$ เป็น

$$\overline{D} - t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}}$$

เมื่อ $$\overline{\mathbf{D}} = \frac{\sum_{i=1}^{n} \mathbf{D}_i}{\mathbf{n}},$$

$$S_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}$$

$$= \frac{\sum_{i=1}^{n} D_i^2 - n\overline{D}^2}{n-1}$$

$$S_D = \sqrt{s_D^2}$$

ตัวอย่างที่ 6.15 สมมติในการทดสอบว่ายาลดน้ำหนักชนิดหนึ่งจะมีผลต่อความดันโลหิตของผู้ได้รับยานี้มากน้อยเพียงใด ผู้ศึกษาได้วัดความดันโลหิตของกลุ่มทดลองก่อนได้รับยาลดน้ำหนักจำนวน 12 คน และทำการวัดความดันโลหิตอีกครั้งหนึ่งเมื่อได้รับยาลดน้ำหนักแล้ว ผลปรากฏดังตารางต่อไปนี้

คนที่	**ก่อนได้รับยา**	**หลังได้รับยา**	$D_i = X_{2i} - X_{1i}$
1	125	128	+3
2	120	128	+8
3	118	127	+9
4	134	141	+7
5	138	135	−3
6	125	129	+4
7	123	125	+2
8	126	124	−2
9	140	141	+1
10	133	138	+5
11	127	135	+8
12	131	137	+6
		รวม	**48**

ถ้ากำหนดให้ความดันโลหิตมีการแจกแจงปกติ เพื่อเปรียบเทียบผลกระทบของยาลดน้ำหนักที่มีต่อความดันโลหิต จงหาช่วงความเชื่อมั่นขนาด 95% ของผลต่างของความดันโลหิตก่อนและหลังการได้รับยาลดน้ำหนัก

วิธีทำ จากตาราง คำนวณหา $D_i = X_{2i} - X_{1i}$; $i = 1, 2, 3, ..., 12$

$$\overline{D} = \frac{48}{12}$$

$$= \ 4$$

$$S_D^2 = \frac{362-12(4)^2}{11}$$

$$= \frac{170}{11}$$

จากตารางในภาคผนวก จะได้ว่า $t_{0.05,\,11} = 1.796$

ช่วงความเชื่อมั่นขนาด 95% ของ μ_D มีค่าเป็น

$$\overline{D} - t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\frac{\alpha}{2},n-1}\frac{S_D}{\sqrt{n}}$$

$$1.962 < \mu_D < 6.038$$

ดังนั้น เชื่อมั่นได้ 95% ว่าความแตกต่างของความดันโลหิตก่อนรับยาลดน้ำหนักกับหลังรับยาลดน้ำหนักอยู่ในช่วง 1.962 ถึง 6.038

6.6.2 การประมาณค่าแบบช่วงกรณีสองประชากรของสัดส่วน

ในการประมาณค่าแบบช่วงกรณีความแตกต่างระหว่างสัดส่วนกรณีนี้ จะพิจารณาตัวประมาณแบบจุดของความแตกต่างระหว่างสัดส่วน $p_1 - p_2$ ดังนั้น $\hat{p}_1 - \hat{p}_2$ คือค่าประมาณแบบจุดของ $p_1 - p_2$ โดยจะเลือกตัวอย่างสุ่ม 2 ตัวอย่างที่เป็นอิสระกันขนาด n_1 และ n_2 จาก 2 ประชากรที่มีการแจกแจงแบบทวินามและคำนวณหาความแตกต่างของสัดส่วนของตัวอย่าง คือ $p_1 - p_2$ สามารถสร้างช่วงความเชื่อมั่นของ $p_1 - p_2$ โดยพิจารณาการแจกแจงการสุ่มตัวอย่างของ $\hat{p}_1 - \hat{p}_2$

ทฤษฎีบทที่ 6.4 ถ้าสุ่มตัวอย่างขนาด n_1 และ n_2 จากสองประชากรที่มีการแจกแจงแบบทวินาม มีค่าเฉลี่ย $\mu_1 = n_1p_1$ และ $\mu_2 = n_2p_2$ และความแปรปรวน $\sigma_1^2 = n_1p_1q_1$ และ $\sigma_2^2 = n_2p_2q_2$ ตามลำดับแล้ว การแจกแจงการสุ่มตัวอย่างของ $\hat{p}_1 - \hat{p}_2$ จะมีการแจกแจงแบบปกติโดยประมาณ

ด้วยค่าเฉลี่ย $\mu_{\hat{p}_1-\hat{p}_2}$ = $p_1 - p_2$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ ดังนั้น

$z = \dfrac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$ คือค่าของตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐาน Z

จากทฤษฎีบทที่ 6.4 หากสุ่มตัวอย่างขนาดโตเพียงพอเราคาดหวังว่าการแจกแจงการสุ่มตัวอย่างของ $\hat{p}_1 - \hat{p}_2$ จะใกล้เคียงกับการแจกแจงปกติด้วยค่าเฉลี่ย $\mu_{\hat{p}_1-\hat{p}_2}$ = $p_1 - p_2$ และความแปรปรวน $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ ดังนั้น สามารถยืนยันว่า

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

เมื่อ $Z = \dfrac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$

และ $z_{\frac{\alpha}{2}}$ คือค่าของ Z ที่มีพื้นที่ใต้โค้งของการแจกแจงปกติมาตรฐานอยู่ทางขวาของค่านี้อยู่ $\frac{\alpha}{2}$

$$P\left(-z_{\frac{\alpha}{2}} < \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

คูณแต่ละพจน์ของอสมการด้วย $\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ แล้วลบด้วย $\hat{p}_1 - \hat{p}_2$ และคูณด้วย -1 จะได้

$$P\left((\hat{P}_1 - \hat{P}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < p_1 - p_2 < (\hat{P}_1 - \hat{P}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} \right) = 1 - \alpha$$

ถ้า n_1 และ n_2 มีขนาดโตพอ สามารถแทนที่ p_1 และ p_2 ภายใต้เครื่องหมายรากด้วยตัวประมาณ $\hat{p}_1 = \frac{x_1}{n_1}$ และ $\hat{p}_2 = \frac{x_2}{n_2}$ แล้ว จะได้

$$P\left((\hat{P}_1 - \hat{P}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{P}_1 - \hat{P}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right) = 1 - \alpha$$

และจากตัวอย่างสุ่มที่เป็นอิสระกัน 2 ตัวอย่าง ขนาด n_1 และ n_2 ที่ถูกเลือกจากประชากรที่มีการแจกแจงทวินาม จะสามารถคำนวณ $\hat{P}_1 - \hat{P}_2$ ดังนั้น ช่วงความเชื่อมั่นขนาด 100(1 – α)% ของความแตกต่างระหว่างสัดส่วน $P_1 - P_2$ คือ

$$(\hat{P}_1 - \hat{P}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{P}_1 - \hat{P}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

เมื่อ $\hat{p}_1$ และ $\hat{p}_2$ เป็นสัดส่วนของการประสบความสำเร็จในตัวอย่างสุ่มขนาด n_1 และ n_2 ตามลำดับ $\hat{q}_1 = 1 - \hat{p}_1$ และ $\hat{q}_2 = 1 - \hat{p}_2$ และ $z_{\frac{\alpha}{2}}$ คือค่าของตัวแปรสุ่ม Z ที่ทำให้พื้นที่ใต้โค้งของการแจกแจงเท่ากับ $\frac{\alpha}{2}$ อยู่ทางขวาของ $z_{\frac{\alpha}{2}}$ และเปิดได้จากตารางในภาคผนวก

ตัวอย่างที่ 6.16 สุ่มตัวอย่างซึ่งประกอบด้วยผู้ใหญ่ 400 คน และอีกตัวอย่างหนึ่งประกอบด้วยวัยรุ่น 600 คน ที่ชมโทรทัศน์รายการหนึ่งทางช่อง 5 แล้วสอบถามเกี่ยวกับความชอบ ปรากฏว่ามีผู้ใหญ่ 100 คน และวัยรุ่น 300 คน ชอบรายการดังกล่าว จงสร้างช่วงความเชื่อมั่นขนาด 95% สำหรับผลต่างในสัดส่วนของผู้ใหญ่กับเด็กวัยรุ่นที่ชอบดูรายการนี้

วิธีทำ ให้ $\hat{p}_1$ แทนสัดส่วนของผู้ใหญ่ที่ชอบรายการโทรทัศน์

$\hat{p}_2$ แทนสัดส่วนของวัยรุ่นที่ชอบรายการโทรทัศน์

$\hat{p}_1 = \dfrac{100}{400} = 0.25 \qquad \hat{p}_2 = \dfrac{300}{600} = 0.50$

$\hat{p}_2 - \hat{p}_1 = 0.50 - 0.25 = 0.25$

$\hat{p}_1\hat{q}_1 = 0.25(1 - 0.25) = 0.25(0.75) = 0.1875$

$\hat{p}_2\hat{q}_2 = 0.50(1 - 0.50) = 0.50(0.50) = 0.2500$

$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

ดังนั้น ช่วงความเชื่อมั่นขนาด 95% ของความแตกต่างในสัดส่วนของผู้ใหญ่กับวัยรุ่นที่ชอบดูรายการนี้เป็น

$$0.25 - 1.96\sqrt{\frac{0.1875}{400}+\frac{0.2500}{600}} < p_1 - p_2 < 0.25 + 1.96\sqrt{\frac{0.1875}{400}+\frac{0.2500}{600}}$$

$$0.25 - 1.96(0.029) < p_1 - p_2 < 0.25 + 1.96(0.029)$$

$$0.25 - 0.0583 < p_1 - p_2 < 0.25 + 0.0583$$

$$0.19 < p_1 - p_2 < 0.31$$

ดังนั้น ความแตกต่างในสัดส่วนของผู้ใหญ่กับเด็กวัยรุ่นที่ชอบดูรายการนี้อยู่ระหว่าง 0.19 ถึง 0.31 ด้วยความเชื่อมั่น 95%

ตัวอย่างที่ 6.17 ในการสำรวจความคิดเห็นของประชาชนในจังหวัดนครศรีธรรมราชเกี่ยวกับสถานที่ในการสร้างศูนย์กีฬา ผู้ศึกษาได้ทำการศึกษาจากประชาชน 2 กลุ่ม กลุ่มแรกคือประชาชนในชุมชนในเมือง กลุ่มที่สองคือประชาชนในชนบท ในกลุ่มแรกมีผู้เห็นด้วยกับสถานที่ในการสร้างศูนย์กีฬา 44 คน จากผู้ตอบ 200 คน กลุ่มที่สองมีผู้เห็นด้วย 25 คน จากผู้ตอบ 250 คน ถ้าให้ p_1 และ p_2 เป็นสัดส่วนของคนในเขตชุมชนเมืองและชนบทที่เห็นด้วยกับสถานที่ในการสร้างศูนย์กีฬา จงสร้างช่วงความเชื่อมั่นขนาด 90% ของความแตกต่างระหว่างสัดส่วนของคน ในเขตชุมชนเมืองและชนบท

วิธีทำ ให้ $\hat{p}_1$ แทนสัดส่วนของประชาชนในชุมชนเมืองที่เห็นด้วยกับสถานที่ในการสร้างศูนย์กีฬา

$\hat{p}_2$ แทนสัดส่วนของประชาชนในชนบทที่เห็นด้วยกับสถานที่ในการสร้างศูนย์กีฬา

$\hat{p}_1 = \dfrac{44}{200} = 0.22 \qquad \hat{p}_2 = \dfrac{25}{250} = 0.10$

$$\hat{q}_1 = 1 - 0.22 = 0.78 \qquad \hat{q}_2 = 1 - 0.10 = 0.90$$

$$\hat{p}_1\hat{q}_1 = (0.22)(0.78) = 0.1716$$

$$\hat{p}_2\hat{q}_2 = (0.10)(0.90) = 0.09$$

$$\hat{p}_1 - \hat{p}_2 = 0.22 - 0.10 = 0.12$$

ช่วงความเชื่อมั่นขนาด 90% ของ $p_1 - p_2$ เป็น

$$0.12 - 1.645\sqrt{\frac{0.1716}{200} + \frac{0.0900}{250}} < p_1 - p_2 < 0.12 + 1.645\sqrt{\frac{0.1716}{200} + \frac{0.0900}{250}}$$

$$0.12 - 1.645(0.035) < p_1 - p_2 < 0.12 + 1.645(0.035)$$

$$0.12 - 0.057 < p_1 - p_2 < 0.12 + 0.057$$

$$0.063 < p_1 - p_2 < 0.177$$

ดังนั้น ช่วงความเชื่อมั่นขนาด 90% ของความแตกต่างระหว่างสัดส่วนของประชาชนในชุมชนเมืองและในชนบทที่เห็นด้วยกับการสร้างศูนย์กีฬาอยู่ในช่วง 0.063 ถึง 0.177

บทสรุป

การประมาณค่า เป็นการอนุมานทางสถิติที่นำข้อมูลจากตัวอย่างไปอ้างอิงถึงประชากร โดยการใช้ตัวสถิติจากตัวอย่างไปประมาณค่าพารามิเตอร์ของประชากรให้เกิดความผิดพลาดน้อยที่สุด การประมาณค่าที่ยอมรับกันโดยทั่วไปเป็นการประมาณค่าแบบช่วง โดยอาศัยค่าประมาณแบบจุดและช่วงของตัวเลขมากำหนดช่วงของการประมาณตามระดับความเชื่อมั่นที่ระบุไว้ การประมาณแบบช่วงทำให้เห็นว่าโอกาสของความคลาดเคลื่อนมีมากน้อยเพียงใด อันจะเป็นแนวทางสำคัญที่จะต้องเพิ่มความระมัดระวังในการสรุปผลของการประมาณค่าในแต่ละครั้งก่อนที่จะนำไปใช้ประโยชน์ต่อไป

แบบฝึกหัด

1. ในการสำรวจความต้องการศึกษาต่อที่มหาวิทยาลัยแห่งหนึ่ง โดยการสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 มาจำนวน 2,000 คน พบว่า 580 คน มีความต้องการที่จะศึกษาต่อในมหาวิทยาลัยแห่งนี้ จงหาค่าประมาณแบบจุดของสัดส่วนของนักเรียนที่ต้องการศึกษาต่อในมหาวิทยาลัยแห่งนี้
2. จากข้อมูลเกี่ยวกับจำนวนวันที่ผู้ป่วยจำนวน 60 คน เข้าพักรักษาตัวโรงพยาบาลมหาราชนครศรีธรรมราชเป็นเวลา 1 เดือน มีจำนวนวันเฉลี่ยเป็น 4.5 วัน ความแปรปรวนเป็น 1.125 วัน^2 จงประมาณจำนวนวันที่ผู้ป่วยคนหนึ่ง ๆ จะเข้ารักษาตัวในโรงพยาบาล
3. ส่วนสูงของนักศึกษาโปรแกรมวิชาพลศึกษา 15 คน ที่สุ่มมาเป็น 170 เซนติเมตร ส่วนเบี่ยงเบนมาตรฐาน 18 เซนติเมตร จงสร้างช่วงความเชื่อมั่น 95% ของส่วนสูงเฉลี่ย
4. จงสร้างช่วงความเชื่อมั่นขนาด 95% สำหรับความแตกต่างของรายได้ต่อวันของร้านค้าแห่งหนึ่งก่อนและหลังจากจัดให้มีโฆษณาทางสื่อต่าง ๆ เมื่อรายได้เฉลี่ยต่อวันเป็นดังนี้

	ก่อนโฆษณา	**หลักโฆษณา**
รายได้เฉลี่ย	1255	1330
ความแปรปรวน	215	238
ขนาดตัวอย่าง	50	30

5. สุ่มตัวอย่างทรานซิสเตอร์มา 300 เครื่อง เพื่อตรวจสอบปรากฏว่ามี 12 เครื่องที่ชำรุด จงสร้างช่วงความเชื่อมั่นขนาด 95% ของสัดส่วน p เมื่อ p คือสัดส่วนของเครื่องที่ชำรุด
6. สุ่มตัวอย่างขนาด 200 จาก 2 ประชากร ที่มีการแจกแจงแบบเบอร์นูลี จำนวนสมาชิกที่ประสบความสำเร็จ ในตัวอย่างที่ 1 และตัวอย่างที่ 2 เป็น 65 และ 74 ตามลำดับ

   6.1 จงคำนวณหาค่าคลาดเคลื่อนมาตรฐานของ $P_1 - P_2$ ถ้ากำหนดว่า $P_1 = P_2$

   6.2 จงคำนวณหาค่าประมาณแบบจุดสำหรับสัดส่วนจากประชากรที่ 1 และ 2
7. สาธารณสุขจังหวัดนครศรีธรรมราชต้องการทราบจำนวนแบคทีเรียต่อน้ำปริมาณน้ำ 1 หน่วยในแม่น้ำที่ไหลผ่านคลองนครน้อย จึงให้นักวิจัยเก็บตัวอย่างน้ำในน้ำดังกล่าว 10 จุด และนับจำนวนแบคทีเรียได้ดังนี้

   128 157 189 198 174 110 175 139 206 177

   จงสร้างช่วงความเชื่อมั่นขนาด 99% ของจำนวนแบคทีเรียต่อ 1 หน่วยปริมาตร
8. จากการสอบถามนักเรียน 500 คน พบว่า 220 คนเป็นโรคผิวหนัง 120 คน เป็นเป็นโรคท้องร่วง และ 45 คน เป็นทั้งโรคผิวหนังและท้องร่วง

   8.1 จงสร้างช่วงความเชื่อมั่น 95% ของสัดส่วนของนักเรียนที่เป็นโรคผิวหนัง

   8.2 จงสร้างช่วงความเชื่อมั่น 99% ของสัดส่วนของนักเรียนที่เป็นทั้งสองโรค

8.3 จงสร้างช่วงความเชื่อมั่น 99% ของสัดส่วนของนักเรียนที่ไม่เป็นทั้งสองโรค

9. ในการเลือกตั้งสมาชิกสภาเทศบาลของเทศบาลหนึ่ง มีผู้มาใช้สิทธิ์ออกเสียง 12,000 คน เมื่อถึงวันเลือกตั้ง ปรากฏว่ามีผู้ไปใช้สิทธิ์เพียงร้อยละ 72.5 จงสร้างความเชื่อมั่นขนาด 95% ของสัดส่วนของผู้ที่ไม่ใช้สิทธิ์ครั้งนี้

10. ผู้สร้างแบบทดสอบเชาว์ปัญญาเชื่อว่า ถ้าใช้แบบทดสอบนี้กับคนจำนวนมากจะมีคะแนนเฉลี่ยเท่ากับ 50 ผู้ใช้แบบทดลองใช้แบบทดสอบนี้กับนักศึกษา 15 ปรากฏว่าผลสอบเฉลี่ยเป็น 60.4 คะแนน ส่วนเบี่ยงเบนมาตรฐานเป็น 49 จงสร้างช่วงความเชื่อมั่นขนาด 95% ของความแปรปรวนของคะแนนสอบ

เอกสารอ้างอิง

(Casella & Berger, 1990

(ประชุม สุวัตถี, 2527)

(ศิริชัย พงษ์วิชัย, 2547)

(ศิริชัย กาญจนวาสี, 2547; Dudewicz, 1988)