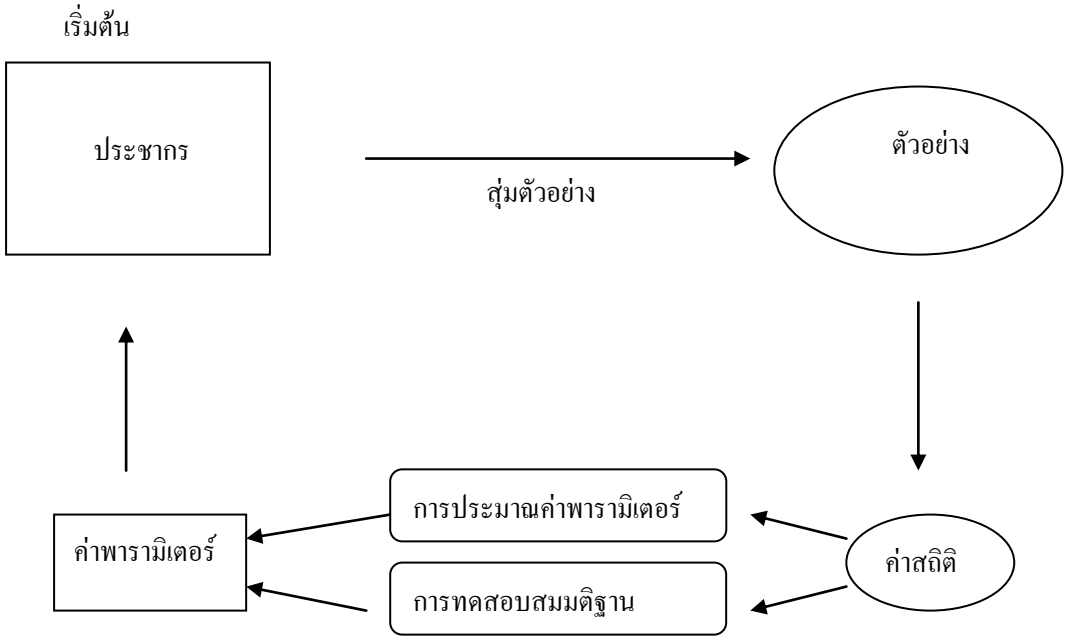
บทที่ 5
การอนุมานค่าเฉลี่ยของประชากร

5.1 การอนุมานทางสถิติ

ในการศึกษาคุณลักษณะของประชากรโดยตรงนั้น กระทำได้ค่อนข้างยาก เสียเวลาและค่าใช้จ่ายสูง ซึ่งประชากรทั่ว ๆ ไป มักจะมีขนาดใหญ่ จึงจำเป็นต้องอาศัยคุณลักษณะของตัวอย่างมาวิเคราะห์เพื่อการตัดสินใจ แล้วนำผลที่ได้จากตัวอย่างนั้นไปสรุปผลอ้างอิงเกี่ยวกับคุณลักษณะของประชากรว่าเป็นอย่างไร การศึกษาจากตัวอย่างซึ่งสุ่มมาจากประชากรนั้น ต้องสามารถแสดงคุณลักษณะของประชากรได้ดีอย่างแท้จริงด้วย วิธีการทางสถิติที่กล่าวมาข้างต้นนี้จึงเรียกว่าเป็นการอนุมานทางสถิติ สามารถแบ่งออกเป็น 2 ลักษณะดังนี้

1. การประมาณค่าพารามิเตอร์ (Estimation)
2. การทดสอบสมมติฐาน (Test of Hypothesis)

ดังรูป



1. การประมาณค่าพารามิเตอร์ (Estimation)

เป็นวิธีการที่ประมาณค่าพารามิเตอร์ หรือค่าที่ได้จากประชากร โดยอาศัยการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง แล้วใช้ค่าสถิติที่ได้จากกลุ่มตัวอย่างมาทำการประมาณค่าพารามิเตอร์นั้น โดยอาศัยทฤษฎีความน่าจะเป็น และการแจกแจงของกลุ่มตัวอย่าง ดังนั้น ในการประมาณค่าพารามิเตอร์จะต้องเกี่ยวข้องกับ ตัวประมาณ และค่าประมาณ โดยตัวประมาณคือ ตัวสถิติที่ใช้ประมาณค่าพารามิเตอร์ ส่วนค่าประมาณ หรือค่าสถิติ คือค่าที่เป็นตัวเลขของตัวสถิติ ที่คำนวณได้จากกลุ่มตัวอย่าง

คุณสมบัติของตัวประมาณที่ดี

1. เป็นตัวประมาณที่ไม่เอนเอียง (Unbias estimator)

คือค่าคาดหวังของตัวสถิติ ต้องเท่ากับ ค่าพารามิเตอร์ $E(\hat{\theta}) = \theta$

2. เป็นตัวประมาณที่มีประสิทธิภาพสูงที่สุด (Most efficient estimator)

ตัวประมาณที่ไม่เอนเอียงของค่าพารามิเตอร์ θ อาจมีได้หลายค่า แต่ตัวประมาณที่มีประสิทธิภาพมากที่สุด คือตัวประมาณไม่เอนเอียงแล้วยังมีความแปรปรวนน้อยที่สุด

3. เป็นตัวประมาณที่คงเส้นคงวา (Consistency)

ถ้าตัวอย่างมีขนาดใหญ่ขึ้น จะได้ว่า ค่าสถิติ จะมีค่าเข้าใกล้ค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ มีวิธีการประมาณได้ 2 แบบ คือ

1.1 การประมาณค่าแบบจุด (Point Estimation)

1.2 การประมาณค่าแบบช่วง (Interval Estimation)

1.1 การประมาณค่าแบบจุด (Point Estimation)

เป็นการประมาณค่าโดยใช้ ค่าสถิติ เพียงค่าเดียวเป็นตัวประมาณของค่าพารามิเตอร์ เช่นใช้ค่าเฉลี่ยของตัวอย่าง $\bar{x}$ เป็นค่าประมาณค่าเฉลี่ยของประชากร μ

ค่าสัดส่วนของตัวอย่าง $\hat{p}$ เป็นค่าประมาณค่าสัดส่วนของประชากร p เป็นต้น

การประมาณค่าแบบจุด เป็นการหาค่าประมาณที่คำนวณได้ค่อนข้างง่าย แต่ก็มีความคลาดเคลื่อนค่อนข้างสูง เนื่องจากเป็นการใช้ค่าเดียวในการประมาณ และเป็นค่าที่คำนวณจากข้อมูลของตัวอย่าง ซึ่งถือว่าเป็นตัวแปรสุ่มที่มีค่าผันไปตามกลุ่มของตัวอย่างที่สุ่มได้ สำหรับประชากร

หนึ่ง ๆ อาจจะมีกลุ่มตัวอย่างได้หลายชุด ดังนั้นในแต่ละชุดก็จะมีค่าประมาณของแต่ละกลุ่ม และค่าประมาณนี้ก็อาจจะมีค่าเท่ากันกับค่าพารามิเตอร์ หรือไม่เท่ากันก็ได้

1.2 การประมาณค่าแบบช่วง (Interval Estimation)

เป็นการประมาณค่าพารามิเตอร์ ที่อยู่ในรูปช่วง 2 ค่า โดยตัวเลขช่วงจะบอกค่าต่ำสุด และสูงสุดของค่าประมาณ เช่น $a < \theta < b$ หมายความว่า ค่าพารามิเตอร์ θ จะมีค่าอยู่ระหว่างa ถึง b

ช่วงที่หามาได้นั้น ต้องเป็นช่วงที่เชื่อมั่นได้ว่าค่าพารามิเตอร์ที่สนใจจะอยู่ในช่วงนี้ด้วยความน่าจะเป็นในระดับหนึ่ง โดยกำหนดร้อยละของความเชื่อมั่นเป็น $(1-\alpha)100\%$ และเรียกการประมาณแบบช่วงนี้ว่า **ช่วงความเชื่อมั่น** $(1-\alpha)100\%$ ของค่าพารามิเตอร์ θ(Confidence interval)

และ ช่วงความเชื่อมั่นที่ต้องการสำหรับการประมาณค่าพารามิเตอร์ในแต่ละครั้ง จะต้องเป็นช่วงที่แคบ แต่ต้องเชื่อมั่นได้ค่อนข้างมากว่าช่วงดังกล่าวนี้ต้องครอบคลุมพารามิเตอร์ได้ และเรียกค่าของ $1-\alpha$ เรียกว่า **สัมประสิทธิ์ความเชื่อมั่น** (Confidence coefficient) ส่วนค่าของ α เรียกว่า **ระดับนัยสำคัญ** (Significant level) เช่น ถ้า $\alpha = 0.05$ จะได้ช่วงแห่งความเชื่อมั่น คือ 95% และถ้า $\alpha = 0.10$ จะได้ช่วงแห่งความเชื่อมั่น คือ 90% โดยส่วนใหญ่ก็จะนิยมช่วงความเชื่อมั่นที่ 90% , 95% และ 99%

สมมุติให้ช่วงความเชื่อมั่น $P(a < \theta < b) = 1-\alpha = 0.90$

หมายความว่า ความน่าจะเป็นที่ค่าพารามิเตอร์ θ จะมีค่าอยู่ในระหว่างค่า a ถึงค่า b คือ 0.90 และจะเรียกค่า a ว่า ขีดจำกัดล่างของช่วงความเชื่อมั่น (Lower Confidence Limit) และเรียกค่า b ว่า ขีดจำกัดบนของช่วงความเชื่อมั่น (Upper Confidence Limit) ซึ่งค่าของ a และ b นั้นจะขึ้นอยู่กับตัวประมาณค่า $\hat{\theta}$ และการแจกแจงความน่าจะเป็นของ $\hat{\theta}$

2. การทดสอบสมมติฐาน (Test of Hypothesis)

เป็นวิธีการที่ต้องอาศัยข้อมูลจากตัวอย่างที่สุ่มจากประชากร มาช่วยในการตัดสินใจเกี่ยวกับข้อสันนิษฐานที่ว่า ค่าพารามิเตอร์ของประชากรดังกล่าวนั้น มีค่าเป็นดังที่ได้คาดไว้หรือไม่ โดยในกระบวนการทดสอบสมมติฐานก็จะมีทางเลือกให้สองทาง คือการตัดสินใจยอมรับสมมติฐาน และการตัดสินใจปฏิเสธสมมติฐาน ในการตัดสินใจเลือกทางใดทางหนึ่งนั้นก็จะต้องขึ้นอยู่กับข้อมูลที่ได้จากตัวอย่าง และวิธีการทางสถิติที่นำมาใช้

2.1 สมมติฐานทางสถิติ

เป็นข้อสมมุติ ความเชื่อ หรือเป็นข้อสรุปที่กล่าวอ้างเกี่ยวกับลักษณะของประชากรที่เราสนใจ ซึ่งก็หมายถึง ค่าพารามิเตอร์ โดยอาจจะมีกลุ่มเดียวหรือหลายกลุ่มขึ้นไป อาจจะเป็นความจริงหรือไม่เป็นความจริงก็ได้

2.2 การกำหนดสมมติฐานการทดสอบทางสถิติ

สมมติฐานที่จะต้องกำหนดเพื่อใช้ในการทดสอบประกอบด้วยสมมติฐาน 2 ชนิด คือ

1. สมมติฐานหลัก (Null Hypothesis) เป็นข้อสมมุติ เกี่ยวกับลักษณะของประชากรที่กำหนดขึ้นเพื่อทดสอบ ใช้สัญลักษณ์แทนด้วย H_0 โดยลักษณะของ H_0 จะเป็นการระบุค่าพารามิเตอร์ว่ามีค่าเท่ากับค่าใดค่าหนึ่งอย่างแน่นอน หรือมีค่าเท่ากันในระหว่างค่าพารามิเตอร์
2. สมมติฐานรอง (Alternative Hypothesis) เป็นข้อสมมุติ เกี่ยวกับลักษณะของประชากรที่กำหนดขึ้นเพื่อทดสอบ ใช้สัญลักษณ์แทนด้วย H_1 โดยลักษณะของ H_1 จะต้องเป็นพารามิเตอร์เดียวกับที่ H_0 แต่มีค่าของพารามิเตอร์ที่ตรงกันข้ามกับ H_0 โดยจะระบุค่าของพารามิเตอร์ที่ H_1 เป็นช่วงซึ่งอาจเป็นไม่เท่ากับ ($\neq$) มากกว่า (>) หรือน้อยกว่า (<)

เมื่อกำหนดสมมติฐานหลัก และสมมติฐานรองแล้ว ขั้นตอนการทดสอบสมมติฐานต่อมาก็คือ จะต้องใช้ข้อมูลจากตัวอย่าง เพื่อช่วยในการตัดสินใจเลือก ว่าจะยอมรับ H_0 หรือจะปฏิเสธ H_0 และในการสรุปผลการทดสอบ ไม่ว่าจะเป็นการยอมรับ H_0 หรือการปฏิเสธ H_0 ก็เป็นการตัดสินใจที่อาจจะผิดพลาด คลาดเคลื่อนได้ เนื่องจากการตัดสินใจดังกล่าวได้กระทำโดยอาศัยข้อมูลของตัวอย่าง เท่านั้น ดังนั้นการตัดสินใจอาจจะคลาดเคลื่อนจากความเป็นจริงของพารามิเตอร์จากประชากรได้

2.3 ความคลาดเคลื่อน จากการตัดสินใจในการทดสอบสมมติฐาน

ในการทดสอบสมมติฐาน ผลการทดสอบออกมาอาจยอมรับ H_0 หรือปฏิเสธ H_0 ก็ได้ แต่อย่างไรก็ดีผลของการทดสอบนี้ก็จะเป็นจริงที่ระดับความเชื่อมั่นที่กำหนดไว้เท่านั้น นั่นคือ ถ้าผลการทดสอบยอมรับ H_0 แสดงว่าผลการทดสอบนี้จะถูกต้อง ถ้า H_0 เป็นจริง แต่ถ้า H_0 เป็นเท็จ ก็ทำให้เกิดความคลาดเคลื่อนในการทดสอบนั้นขึ้น เช่นเดียวกันถ้าผลการทดสอบปฏิเสธ H_0 ก็แสดงว่าผลการทดสอบนี้จะถูกต้องถ้า H_0 เป็นเท็จ แต่ถ้า H_0 เป็นจริง ก็ทำให้เกิดความคลาดเคลื่อนในการ

ทดสอบขึ้นเช่นกัน ความคลาดเคลื่อนที่เกิดขึ้นในการทดสอบสมมติฐานสามารถแบ่งออกได้เป็น 2 ประเภท ดังนี้

ก. ความคลาดเคลื่อนชนิดที่ 1 (Type I error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธ H_0 ทั้ง ๆ ที่ H_0 เป็นจริง โดยกำหนดให้ α เป็นค่าความน่าจะเป็นที่ยอมให้เกิดความคลาดเคลื่อนชนิดที่ 1 และค่า α ที่นิยมกำหนดใช้ในการทดสอบกันมาก คือ $\alpha = 0.05$ และ $\alpha = 0.01$

ข. ความคลาดเคลื่อนชนิดที่ 2 (Type II error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการยอมรับ H_0 เมื่อ H_0 เป็นเท็จ โดยกำหนดให้ β เป็นค่าความน่าจะเป็นที่ยอมให้เกิดความคลาดเคลื่อนชนิดที่ 2

ความสัมพันธ์ของ ความคลาดเคลื่อนทั้งสองชนิด สรุปได้ดังตารางต่อไปนี้

ผลการทดสอบข้อเท็จจริง	ยอมรับ H_0	ปฏิเสธ H_0
H_0 เป็นจริง	ผลการทดสอบถูกต้อง	ความคลาดเคลื่อนชนิดที่ 1
H_0 เป็นเท็จ	ความคลาดเคลื่อนชนิดที่ 2	ผลการทดสอบถูกต้อง

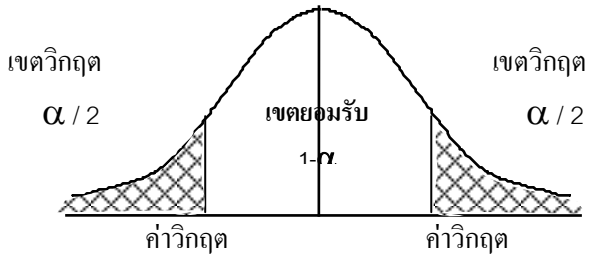
2.4 ขั้นตอนการทดสอบสมมติฐานทางสถิติ

1. กำหนดสมมติฐานหลัก H_0 และสมมติฐานรอง H_1
2. กำหนดระดับนัยสำคัญ α
3. เลือกค่าทดสอบสถิติที่เหมาะสม กับค่าพารามิเตอร์ที่ต้องการทดสอบ แล้วคำนวณค่าสถิติจากตัวอย่างที่สุ่มได้
4. กำหนดบริเวณอาณาเขตวิกฤต หรือเขตปฏิเสธ สมมติฐานหลัก H_0
5. สรุปผลการทดสอบ โดย ถ้าค่าทดสอบสถิติตกอยู่ในเขตวิกฤต (Critical Region) ซึ่งเป็นเขตที่ปฏิเสธ H_0 แสดงว่าผลการทดสอบจะปฏิเสธ H_0 แล้วจะยอมรับ H_1 แต่ถ้าค่าทดสอบสถิติตกอยู่นอกเขตวิกฤต แสดงว่าผลการทดสอบ จะยอมรับ H_0 (Acceptance Region) และเขตวิกฤตจะแสดงอยู่ภายใต้พื้นที่ใต้โค้งของการแจกแจงที่เกี่ยวข้องกับค่าสถิติ ที่เป็นตัวประมาณของพารามิเตอร์ที่ทดสอบ โดยเขตวิกฤตจะอยู่ที่ปลายโค้งด้านขวาหรือด้านซ้ายหรือทั้งสองด้านของการแจกแจงนั้นเสมอ โดยพื้นที่จะขึ้นอยู่กับค่า α ที่กำหนดใช้ในการทดสอบ

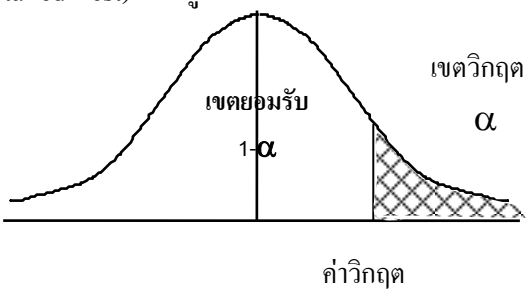
สำหรับลักษณะของเขตวิกฤตจะเป็นอย่างไรนั้นขึ้นอยู่กับการกำหนด H_1 โดยอยู่ที่ความต้องการของผู้วิจัยว่าต้องการทดสอบทางเดียว หรือสองทาง และการกำหนด H_1 ก็จะมีผลต่อการกำหนดเขตวิกฤต ดังนี้

ให้ θ หมายถึง พารามิเตอร์ใด ๆ ที่ต้องการทดสอบ และ θ_0 หมายถึง ค่าคงที่

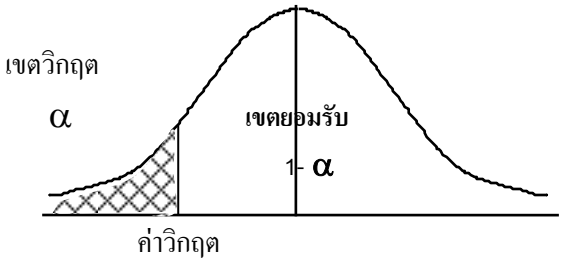
ถ้า $H_1 : \theta \neq \theta_0$ เขตวิกฤต หรือ บริเวณการปฏิเสธจะอยู่ส่วนปลายทั้งสองข้างของโค้งการแจกแจงโดยมีพื้นที่ด้านละเท่ากับ $\alpha/2$ ค่าวิกฤตมี 2 ค่า การทดสอบที่กำหนด H_1 ในลักษณะนี้จะเป็นการทดสอบแบบสองทาง (Two-tailed Test) ตามรูป



ถ้า $H_1 : \theta > \theta_0$ เขตวิกฤต หรือ บริเวณการปฏิเสธจะอยู่ส่วนปลายทั้งหมดด้านขวาของโค้งการแจกแจง มีพื้นที่เท่ากับ α มีค่าวิกฤต 1 ค่า การทดสอบที่กำหนด H_1 ในลักษณะนี้จะเป็นการทดสอบแบบทางเดียว (One-tailed Test) ตามรูป



ถ้า $H_1 : \theta < \theta_0$ เขตวิกฤต หรือ บริเวณการปฏิเสธ จะอยู่ส่วนปลายทั้งหมดทางด้านซ้ายของโค้งการแจกแจง มีพื้นที่เท่ากับ α มีค่าวิกฤต 1 ค่า การทดสอบที่กำหนด H_1ในลักษณะนี้จะเป็นการทดสอบแบบทางเดียว (One-tailed Test) ตามรูป



5.2 การอนุมานค่าเฉลี่ยของประชากร 1 กลุ่ม

5.2.1 การแจกแจงค่าเฉลี่ยของตัวอย่าง 1 กลุ่ม

จากประชากรขนาด N ซึ่งมีค่าเป็น $X_1, X_2, \ldots\ldots X_n$ ตามลำดับ โดยมี

ค่าเฉลี่ยของประชากร คือ $\mu = \dfrac{\sum_{i=1}^{n} X_i}{N}$

ความแปรปรวนของประชากร คือ $\sigma^2 = \dfrac{\sum_{i=1}^{n} (X_i - \mu)^2}{N}$

ถ้าทำการสุ่มตัวอย่างขนาด n จากประชากรข้างต้น จะได้

ค่าเฉลี่ยของตัวอย่าง คือ $\bar{x} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

ความแปรปรวนของตัวอย่าง คือ $S^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}$

จะพบว่า ค่าเฉลี่ยของตัวอย่างสามารถมีได้หลายค่า ขึ้นอยู่กับค่าสังเกตของชุดตัวอย่างที่สุ่มมา ซึ่งค่าเหล่านี้ก็จะเป็นตัวที่สามารถหาการแจกแจงของค่าเฉลี่ยของตัวอย่างได้ พิจารณาจากตัวอย่างต่อไปนี้

สุ่มตัวอย่างขนาด n = 2 จากประชากรขนาด N = 4 ที่มีค่าสังเกตเป็น 1 , 3 , 5 , 7

ให้ μ แทน ค่าเฉลี่ยของประชากร

$$\mu = \frac{1+3+5+7}{4} = 4$$

σ^2 แทนความแปรปรวนของประชากร

$$\sigma^2 = \frac{1^2+3^2+5^2+7^2}{4} - \left(\frac{16}{4}\right)^2 = 5$$

ถ้าสุ่มตัวอย่างแบบซ้ำกันได้ขนาด 2 จากประชากรนี้ ชุดของจำนวนตัวอย่างที่จะเป็นไปได้ทั้งหมด ก็คือ 16 ชุด คำนวณค่าเฉลี่ยตัวอย่างแต่ละชุดได้ดังนี้

ตัวอย่างชุดที่	**ค่าสังเกต**	$\bar{x}$	**ตัวอย่างชุดที่**	**ค่าสังเกต**	$\bar{x}$
1	1 , 1	1	9	5 , 1	3
2	1 , 3	2	10	5 , 3	4
3	1 ,5	3	11	5 , 5	5
4	1 , 7	4	12	5 , 7	6
5	3 , 3	3	13	7 , 1	4
6	3 , 5	4	14	7 , 3	5
7	3 , 7	5	15	7 , 5	6
8	3 , 1	2	16	7 , 7	7

นำค่า $\bar{x}$ ที่เป็นไปได้ทั้งหมดมาแจกแจงความถี่ได้ดังนี้

$\bar{x}$	ความถี่ (f)	$f\bar{x}$	$f\bar{x}^2$
1	1	1	1
2	2	4	8
3	3	9	27
4	4	16	64
5	3	15	75
6	2	12	72
7	1	7	49
รวม	16	64	296

ให้ $\mu_{\bar{x}}$ แทนค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง

$$\mu_{\bar{x}} = \frac{\sum f\bar{x}}{\sum f} = \frac{64}{16} = 4 = \mu$$

จะเห็นได้ว่า ค่าเฉลี่ยของ $\bar{x}$ เท่ากับ ค่าเฉลี่ยของประชากร

ให้ $\sigma^2_{\bar{x}}$ แทนความแปรปรวนของค่าเฉลี่ยของตัวอย่าง

$$\sigma^2_{\bar{x}} = \frac{\sum f\left(\bar{x}-\mu_{\bar{x}}\right)^2}{\sum f} = \frac{\sum f\bar{x}^2}{\sum f} - \mu_{\bar{x}}^2$$

$$= \frac{296}{16} - \left(\frac{64}{16}\right)^2 = 2.5$$

$$= \frac{\sigma^2}{n} = \frac{5}{2} = 2.5$$

จะเห็นว่าความแปรปรวนของค่าเฉลี่ยของตัวอย่าง

กรณีสุ่มตัวอย่างขนาด n จากประชากรขนาด N แบบใส่คืน

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

และกรณีสุ่มตัวอย่างขนาด n จากประชากรขนาด N แบบไม่ใส่คืน

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]$$

ถ้า N มีค่ามากๆ $\left[\frac{N-n}{N-1}\right]$ จะมีค่าเข้าใกล้ 1

ค่าเฉลี่ยของตัวอย่างจะมีค่าใกล้เคียงกับค่าเฉลี่ยของประชากร จึงเป็นเหตุผลที่ใช้ค่าเฉลี่ยของตัวอย่าง เป็นตัวประมาณค่าเฉลี่ยของประชากร

ถ้าสุ่มตัวอย่างขนาด n จากประชากรที่เป็นประชากรอนันต์ ซึ่งมีค่าเฉลี่ย μ ค่าความแปรปรวน σ^2 จะได้ค่าเฉลี่ยของตัวอย่างมีการแจกแจงเข้าใกล้การแจกแจงแบบปกติด้วยค่าเฉลี่ย $\mu_{\bar{x}} = \mu$ ความแปรปรวน $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ หรือ $\bar{x} \sim N\left(\mu_{\bar{x}} = \mu, \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\right)$

จากที่ได้กล่าวมา **เมื่อสุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ สามารถสรุปการแจกแจงค่าเฉลี่ยของตัวอย่างได้เป็นกรณีต่าง ๆ ดังต่อไปนี้**

กรณีที่ 1 เมื่อทราบค่าความแปรปรวนของประชากร σ^2

ค่าเฉลี่ยของตัวอย่าง $\bar{x}$ จะมีการแจกแจงแบบปกติโดยมี

ค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง $\mu_{\bar{x}} = \mu$

และ ความแปรปรวนของค่าเฉลี่ยของตัวอย่าง $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

หรือ $\bar{x} \sim N\left(\mu_{\bar{x}} = \mu, \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\right)$

ดังนั้นจะได้ $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

โดย $Z \sim N(0, 1)$

กรณีที่ 2 เมื่อไม่ทราบค่าความแปรปรวนของประชากร σ^2 และสุ่มตัวอย่างขนาดใหญ่ ($n \geq 30$)

ค่าเฉลี่ยของตัวอย่าง $\bar{x}$ จะมีการแจกแจงแบบปกติโดยมี

ค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง $\mu_{\bar{x}} = \mu$

และ ความแปรปรวนของค่าเฉลี่ยของตัวอย่าง $\sigma^2_{\bar{x}} = \frac{S^2}{n}$

เนื่องจากไม่ทราบค่าความแปรปรวนของประชากร σ^2 จึงใช้ความแปรปรวนของตัวอย่าง S^2 มาเป็นตัวประมาณของ σ^2

ดังนั้นจะได้ $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{S / \sqrt{n}}$

โดย $Z \sim N(0, 1)$

กรณีที่ 3 เมื่อไม่ทราบค่าความแปรปรวนของประชากร σ^2 และสุ่มตัวอย่างขนาดเล็ก ($n < 30$)

ค่าเฉลี่ยของตัวอย่าง $\bar{x}$ จะมีการแจกแจงแบบที (Student's t–Distribution) โดยมีดีกรีความเป็นอิสระ (Degree of Freedom หรือ df.) เขียนแทนด้วย ν

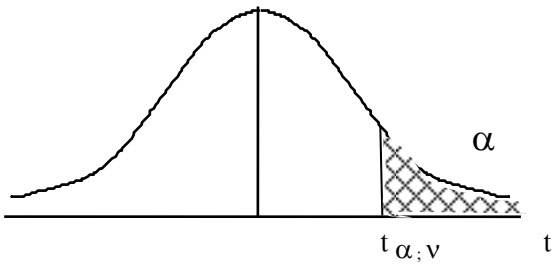
กำหนดให้ $\nu = n - 1$

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}} \sim t_{\nu} \quad , \quad t_{n-1}$$

คุณสมบัติของการแจกแจงแบบ ที

กำหนดให้ t เป็นตัวแปรสุ่มแบบต่อเนื่องที่มีการแจกแจงแบบ ที ดังนั้น การแจกแจงแบบที จึงมีลักษณะคล้ายกับการแจกแจงแบบปกติมาตรฐาน โดยมีเส้นโค้งของการแจกแจงเป็นรูประฆังคว่ำ ที่มีลักษณะสมมาตรที่มีค่าเฉลี่ยเป็น 0 เช่นเดียวกับเส้นโค้งของการแจกแจงแบบปกติมาตรฐาน แต่เส้นโค้งของการแจกแจงแบบทีจะมีหลายรูปแบบซึ่งขึ้นอยู่กับค่าของดีกรีความเป็นอิสระ ถ้าดีกรีความเป็นอิสระมีค่ามาก รูปโค้งของการแจกแจงแบบทีก็จะเป็นรูปโค้งที่โด่งกว่ารูปโค้งที่ดีกรีความเป็นอิสระน้อย ดังนั้นเมื่อ n มีค่ามาก ๆ เข้าสู่ ∞ การแจกแจงแบบทีจะมีค่าเข้าใกล้กับการแจกแจงแบบปกติมาตรฐานได้

การหาความน่าจะเป็นของการแจกแจงแบบที สามารถหาได้โดยการเปิดจากตารางการแจกแจงแบบที จากภาคผนวกหรือตารางสถิติได้ โดยมีค่าความน่าจะเป็นดังรูป



โดย $P(t > t_{\alpha;\nu}) = \alpha$

สำหรับกรณีที่ประชากรไม่ได้มีการแจกแจงแบบปกติ หรืออาจจะไม่ทราบการแจกแจงของประชากรว่ามีการแจกแจงเป็นแบบใด แต่ถ้าสุ่มตัวอย่างที่มีขนาดใหญ่ การแจกแจงของค่าเฉลี่ยของตัวอย่างก็สามารถหาได้โดยอาศัยทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorem)

ทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorem)

ถ้าสุ่มตัวอย่างขนาด n จากประชากรขนาด N ใด ๆ ที่ไม่ได้มีการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเป็น μ และมีความแปรปรวนเป็น σ^2 แล้ว ถ้าตัวอย่างนั้นมีขนาดใหญ่ $n \rightarrow \infty$ การแจกแจงของค่าเฉลี่ยของตัวอย่าง $\bar{x}$ ก็จะมีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติ

โดยมี ค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง คือ $\mu_{\bar{x}} = \mu$

และ ความแปรปรวนของค่าเฉลี่ยของตัวอย่าง คือ $\sigma^2_{\bar{x}} = \dfrac{\sigma^2}{n}$

หรือ $\bar{x} \sim N\left(\mu_{\bar{x}} = \mu, \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\right)$

ดังนั้น $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ กรณีที่ทราบความแปรปรวนของประชากร

$Z = \frac{\bar{x} - \mu}{S / \sqrt{n}}$ กรณีไม่ทราบความแปรปรวนของประชากร

โดย $Z \sim N(0, 1)$

5.2.2 การประมาณค่าเฉลี่ยของประชากร 1 กลุ่ม

ถ้าสุ่มตัวอย่างขนาด n จากประชากรขนาด N ซึ่งมีค่าเฉลี่ย μ ค่าความแปรปรวน σ^2 จะได้ค่าเฉลี่ยของตัวอย่างมีการแจกแจงเข้าใกล้การแจกแจงแบบปกติด้วยค่าเฉลี่ย $\mu_{\bar{x}} = \mu$ ความแปรปรวน $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$ หรือ $\bar{x} \sim N\left(\mu_{\bar{x}} = \mu, \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\right)$

และในการประมาณค่าเฉลี่ยของประชากร μ ค่าสถิติที่ใช้เป็นตัวประมาณของ μ คือค่าเฉลี่ยของตัวอย่าง $\bar{x}$ จากที่ได้กล่าวมา **เมื่อสุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติสามารถสรุปการประมาณแบบช่วงของค่าเฉลี่ยของประชากร 1 กลุ่ม ได้เป็นกรณีต่าง ๆ ดังต่อไปนี้**

กรณีที่ 1 เมื่อทราบค่าความแปรปรวนของประชากร σ^2

ค่าเฉลี่ยของตัวอย่าง $\bar{x}$ จะมีการแจกแจงแบบปกติโดยมี

ค่าเฉลี่ยของค่าเฉลี่ยของตัวอย่าง $\mu_{\bar{x}} = \mu$

และ ความแปรปรวนของค่าเฉลี่ยของตัวอย่าง $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

หรือ $\bar{x} \sim N\left(\mu_{\bar{x}} = \mu, \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\right)$

ดังนั้นจะได้ $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

โดย $Z \sim N(0, 1)$

พิจารณาความน่าจะเป็นที่ Z มีค่าอยู่ระหว่าง $Z = -Z_{\alpha/2}$ และ $Z = Z_{\alpha/2}$ ดังรูป