บทที่ 4

“การประมาณค่าพารามิเตอร์และการทดสอบสมมุติฐาน”

บทที่ 4

ความหมายของคำว่าประชากรในทางสถิติก็คือ ผลรวมของหน่วยทั้งหลายในขอบเขตของเรื่องที่เราสนใจและศึกษา ตัวอย่างเช่น ถ้าเราต้องการจะศึกษาถึงความเห็นของคนในกรุงเทพฯเกี่ยวกับเรื่องของสงครามในเขมร ประชากรก็คือคนทุกคนที่อาศัยอยู่ในกรุงเทพฯ ถ้าเราสนใจถึงรายได้เฉลี่ยของชาวนาในจังหวัดอยุธยา ก็คือชาวนาทุกคนจังหวัดอยุธยา หรือถ้าเราสนใจศึกษาถึงสัดส่วนของผู้เป็นเจ้าของบ้านเองบ้านในเขตบางกะปิ ประชากรก็คือครอบครัวทั้งหมดที่อาศัยอยู่ในเขตบางกะปิ เป็นต้น สิ่งที่เราสนใจที่จะศึกษาจากประชากร เช่น ความคิดเห็นในเรื่องใดเรื่องหนึ่งเช่นวัดออกมาเป็นร้อยละของความคิดเห็นหรือค่าเฉลี่ยรายได้ของประชากร หรือสัดส่วนของสิ่งที่สนใจตลอดจนค่าความแปรปรวนของประชากรนั้นเราเรียกรวมๆว่าคุณลักษณะของประชากรหรือทางสถิติเราจะใช้คำว่าพารามิเตอร์ (parameter)

โดยปกติการที่จะศึกษาคุณลักษณะของประชากรได้นี้ผู้วิจัยหรือผู้ศึกษาจำเป็นต้องศึกษาจากทุกหน่วยในประชากร แต่ปัญหาที่ตามก็คือ ถ้าเราศึกษาจากทุกหน่วยในประชากรก็จะเกิดข้อเสียหลายประการ เช่น

1. ถ้าประชากรมีจำนวนมาก เราจะต้องเสียค่าใช้จ่ายมากและสิ้นเปลืองเวลามาก บางครั้งบางคราวถ้าเรามีงบประมาณจำกัดเราก็ไม่อาจจะทำได้หรือในกรณีที่มีเวลาจำกัดต้องรีบเร่งนำผลไปไปใช้งาน เช่น ถ้าศึกษาถึงการแพร่ระบาดของเพลี้ยในนาข้าวของพื้นที่เพาะปลูกในภาคกลาง ถ้าเราศึกษาจากทุกจังหวัดที่เพาะปลูกข้าวในภาคกลางก็จะทำให้ได้ผลไม่ทันต่อการปราบศัตรูพืชได้

2. ในกรณีของการศึกษาบางเหตุการณ์เช่นการทดสอบของโดยที่ของที่นำมาทดสอบแล้วเสียไปเลย เช่น ทดลองยิงลูกกระสุนปืน หรือตรวจสอบนมกระป๋องว่าได้คุณภาพหรือไม่โดยการใส่สารเคมีเข้าไป กรณีที่ยกตัวอย่างมานี้จะเห็นว่าเมื่อนำสิ่งของดังกล่าวมาตรวจสอบแล้ว สิ่งของดังกล่าวจะเสียไปเลยไม่สามารถนำไปใช้ได้อีก ดังนั้นเราจะทดสอบกับของทุกๆสิ่งคือลูกกระสุนทุกนัดว่าดีหมดเพื่อตรวจรับ หรือใช้นมทุกกระป๋องที่ผลิตได้เพื่อตรวจสอบมาตรฐานของสินค้าย่อมเป็นไปไม่ได้

ทั้งปัจจัยที่1 และปัจจัยที่2 จึงทำให้เกิดวิชาสถิติขึ้นมาเพื่อแก้ไขข้อเสียที่กล่าวมาแล้วกล่าวคือจะมีการใช้การสำรวจเพียงบางหน่วยในประชากรแทนที่จะศึกษาจากทุกหน่วยในประชากร

ตัวแทนของประชากรที่นำมาศึกษาจะเรียกว่า "กลุ่มตัวอย่าง" และวิธีการเลือกกลุ่มตัวอย่างก็มีแบบต่างๆ หลายแบบ ดังที่กล่าวมาแล้วในบทที่1 (ขอให้นักศึกษาพลิกกลับไปอ่านในตอนที่ว่าด้วยบทนี้) การดำเนินงานลำดับที่ถัดมาก็คือ เมื่อเราเลือกกลุ่มตัวอย่างได้แล้วเราจะดำเนินการกับกลุ่มอย่างไร เพื่อให้ได้ข้อมูลที่จะนำไปสรุปและอนุมานไปสู่ประชากรได้ ซึ่งการที่จะนำข้อมูลที่ได้จากกลุ่มตัวอย่างมาอนุมานกลับไปสู่คุณลักษณะของประชากร (parameter) ที่เราสนใจนั้น เราเรียกว่าการสร้างตัวประมาณค่า (Estimator) สำหรับคุณลักษณะของประชากร (parameter)

<u>การประมาณค่านั้นแบ่งเป็น 2 ส่วนคือ</u>

1. การประมาณค่าแบบจุด (หรือการใช้ค่าคงที่ใดๆเรียกว่า (Point Estimator)

การประมาณค่าแบบนี้ก็คือการพยายามสร้างค่าคงที่ (ตัวเลข) โดยใช้ข้อสนเทศจากกลุ่มตัวอย่างมาเป็นตัวแทนของคุณลักษณะของประชากร ตัวอย่างเช่น เราสุ่มตัวอย่างครัวเรือนในตำบลก.มา 5 ครัวเรือนจากประชากรทั้งหมด 50 ครัวเรือนเพื่อศึกษาถึงจำนวนสัตว์เลี้ยงของครัวเรือน ข้อมูลที่ได้จาก 5 ครัวเรือนเป็นดังนี้

ครัวเรือนที่1	1	2	3	4	5
จำนวนสัตว์ที่เลี้ยง	2	3	1	4	5

ในตัวอย่างนี้กลุ่มตัวอย่างจะมีขนาด 5 ครัวเรือน

ประชากรจะมีขนาด 50ครัวเรือน

เราจะนำข้อมูลจากกลุ่มตัวอย่างมารวบรวมและสร้างเป็นค่าคงที่ขึ้นมาง่ายโดยการหาค่าเฉลี่ยจากตัวอย่าง

ซึ่งในที่นี้ $\bar{X} = \sum_{i=1}^{5} X_i = (2+3+1+4+5)/5 = 15/5 = 3$

ถ้าเรากำหนดให้ค่าเฉลี่ยของจำนวนสัตว์เลี้ยงไว้ต่อครัวเรือนของประชากรมีค่าเท่ากับ μ

ดังนั้น μ ก็คือคุณลักษณะ (ค่าเฉลี่ย) ของประชากรที่เราไม่ทราบค่า (เราจะทราบได้ถ้าเราถามทุกครัวเรือนทั้ง 50 ครัวเรือน ว่าแต่ละครัวเรือนเลี้ยงสัตว์ไว้เท่าไรแล้วนำข้อมูลที่ได้มาหาค่าเฉลี่ย)

ในที่นี้เราจะประมาณค่า μ ได้ โดยการใช้ $\bar{X}$ เป็นตัวแทน นั่นก็คือเวลาให้คำตอบว่า มีค่าเท่าไรเราก็ตอบว่า มีค่าประมาณ 3 ตัว/ครัวเรือน หรือพูดว่าโดยเฉลี่ยแล้วครัวเรือนในตำบลก. จะเลี้ยงสัตว์ไว้ประมาณ 3 ตัว การที่เราประมาณ ด้วยเลข 3 นี้เราเรียกว่าการประมาณค่าด้วยจุด (ค่าคงที่ค่าหนึ่ง)

2. การประมาณค่าด้วยช่วงเชื่อมั่น (ช่วงเชื่อมั่น Interval Estimation)

ก่อนที่จะกล่าวถึงการประมาณค่าแบบที่2 ให้นักศึกษาพิจารณาถึงวิธีการประมาณค่าแบบที่1 ดังตัวอย่างเช่นเราประมาณ μ ด้วยเลข 3(ค่า นี้เราหาได้จากกลุ่มตัวอย่าง) นักศึกษาอาจจะตั้งข้อสังเกตได้ว่าถ้าเกิด μ จริง(คุณลักษณะของประชากร) เป็น 4 ตัว ฉะนั้นก็แปลว่าการประมาณค่าแบบที่1 ผิดใช่ไหม ถ้าเช่นนั้นทำไมเราไม่ให้คำตอบค่าประมาณให้มากกว่านี้ เช่น ประมาณค่า μ ว่าอยู่ระหว่าง 1 ถึง 4 หรือจะเขียนคำตอบว่า ช่วงประมาณของ μ คือ(1,4) ความหมายก็คือค่า ที่เราประมาณอาจจะมีค่าได้ตั้งแต่1,2,3, หรือ 4 (เพื่อให้ง่ายในขั้นนี้จะขอยกตัวอย่างเป็นเลขจำนวนเต็มแต่จริงๆแล้วการเขียนค่าที่เป็นไปได้คือ 1-4 หมายถึงเลขใดๆในช่วง(real line) ตั้งแต่ 1-4

ค่าประมาณของ μ ซึ่งในที่นี้กำหนดให้เป็น(1-4) นี้เราเรียกว่าการประมาณค่า μ ด้วยช่วงเชื่อมั่น(ช่วง 1-4)

โดยที่ 1.หมายถึงขีดจำกัดล่าง

และ 4 หมายถึงขีดจำกัดบน

โดยทั่วๆไปเวลาจะเขียนสัญญลักษณ์แทนขีดจำกัดล่างของช่วงเชื่อมั่นสำหรับ μ เรามักใช้ $\hat{\mu}_L$ โดยที่ $\hat{\mu}$ คือตัวประมาณค่าและ L คือขีดจำกัดล่าง และจะเขียนขีดจำกัดบนของช่วงเชื่อมั่นด้วยสัญญลักษณ์ $\hat{\mu}_U$ เป็นต้น

<u>หมายเหตุ</u> ในหนังสือบางเล่มนิยมจะเขียนสัญญลักษณ์แทนคุณลักษณะของประชากรเป็น θ ดังนั้น ค่าประมาณแบบช่วงของ θ ก็คือ $(\hat{\theta}_L, \hat{\theta}_U)$

จะขอย้อนกลับไปยังตัวอย่างที่ว่าเรากำหนดให้ช่วงประมาณของค่าเฉลี่ยจากประชากรคือ

$\mu = (1,4)$

ดังนั้น $\hat{\mu}_L = 1$

และ $\hat{\mu}_U = 4$

ปัญหาที่ตามก็คือ เราจะสร้างช่วงค่าประมาณนี้ได้อย่างไรและมีเงื่อนไขในการประมาณค่าได้อย่างไร ความน่าเชื่อถือ(ความมั่นใจ)ที่ช่วงเชื่อมั่นดังกล่าวจะคลุมค่าของลักษณะประชากรที่เราสนใจนั้นได้สูงแค่ใด สิ่งเหล่านี้จะต้องศึกษาต่อไป แต่สิ่งหนึ่งที่เราคงพอจะมองประโยชน์ได้จากการสร้างช่วงค่าเชื่อมั่นก็คือ ความถูกต้องย่อมสูงกว่าการประมาณค่าด้วยจุดเป็นแน่

การสร้างช่วงเชื่อมั่นสำหรับคุณลักษณะต่างๆของประชากร

1. การประมาณค่าเฉลี่ยของประชากร

ดังที่เคยได้กล่าวมาแล้วว่า เรานิยมใช้สัญลักษณ์ μ แทนค่าเฉลี่ยของประชากรและเราใช้ $\bar{X}$ แทนค่าเฉลี่ยของตัวอย่างซึ่งเป็นค่าประมาณแบบจุดของ μ

ดังนั้นเราสามารถสร้างค่าช่วงเชื่อมั่นของ μ ณ.ระดับความเชื่อมั่น $(1-\alpha)100\%$ ได้คือ

$$\hat{\mu}_L = \bar{X} - Z_{\alpha/2}\,\sigma_{\bar{X}}$$

และ

$$\hat{\mu}_U = \bar{X} + Z_{\alpha/2}\,\sigma_{\bar{X}}$$

โดยที่

1. $Z_{\alpha/2}$ เป็นค่าของตัวแปรสุ่มที่มีการแจกแจงแบบปกติแบบมาตรฐานที่ทำให้เกิดพื้นที่ทางขวามือเป็น $\alpha/2$ เราเปิดค่า $Z_{\alpha/2}$ ได้จากตารางการแจกแจงปกติแบบมาตรฐาน ตัวอย่างเช่นกำหนดให้

$$\alpha = 1 - 95/100 = .05$$

ดังนั้น $\alpha/2 = .05/2 = .025$

หรือ ถ้าเราต้องการสร้างช่วงเชื่อมั่น ณ.ระดับ 99%

เราจะได้ว่า $\alpha = 1-99/100 = .01$

$$\alpha/_2 = .01/2 = .005$$

ถ้าเราต้องการสร้างช่วงเชื่อมั่นณ.ระดับ 90%

เราจะได้ว่า $\alpha = 1- 90/100 = .10$

$$\alpha/_2 = .10/2 = .05$$

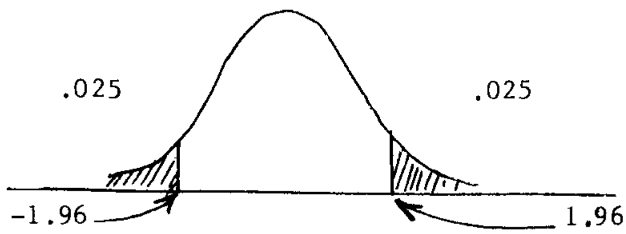
แต่โดยปกติแล้วเรามักนิยมใช้ณ.ระดับ 95% หรือ 90% หรือ 99% เมื่อเรากำหนดได้แล้ว เราก็เปิด $Z_{\alpha/2}$ จากตารางปกติได้ดังนี้ พิจารณาจากรูป เช่นกำหนดให้ $\alpha/2=.025$

.025 .025

Z = ? Z = ?

เปิดตารางหน้า 167 พิจารณาดูว่า Z=? ที่ให้พื้นที่ใต้โค้งเป็น .025 จะพบว่า Z= -1.96

หรือ $Pr(Z \leq -1.96) = .025$



ดังนั้นค่า $Pr(Z \geq 1.96) = .025$

(ให้กลับไปดูวิธีการเปิดตารางในบทที่3ของหนังสือเล่มนี้)

2. $\sigma_{\bar{X}}$ คือค่าของส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยจากตัวอย่าง $\bar{X}$

$\sigma_{\bar{X}} = \sqrt{\frac{N-n}{N-1}}\,(\sigma/\sqrt{n})$ ในเมื่อประชากรมีจำกัด

$\sigma_{\bar{X}} = \sigma/\sqrt{n}$ เมื่อประชากรมีขนาดไม่จำกัด

โดยที่ σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร โดยปกติโจทย์จะกำหนดค่า σ มาให้ แต่ถ้าไม่กำหนด σ มาให้เราก็สามารถหาตัวประมาณค่าของ σ ได้โดยใช้ S (แทนค่าส่วนเบี่ยงเบนมาตรฐานจากตัวอย่าง)แทนได้

โดยปกติแล้วเรามักใช้ $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

แต่ถ้าในกรณีที่ σ ไม่ทราบค่า เราจะใช้ S แทน

ดังนั้น ค่าประมาณที่ได้คือ $S_{\bar{X}} = S/\sqrt{n}$

<u>สรุป</u> เราสามารถหาค่าช่วงเชื่อมั่นของ $\hat{\mu}_L$ และ $\hat{\mu}_U$ ได้ดังนี้คือ

1. ถ้าทราบค่า σ

เราจะได้ว่า σ $\hat{\mu}_L = \bar{X} - Z_{\alpha/2}\sigma_{\bar{X}}$

แทนค่า $\sigma_{\bar{X}}$

$\hat{\mu}_L = \bar{X} - Z_{\alpha/2}\sigma/\sqrt{n}$

$\hat{\mu}_U = \bar{X} + Z_{\alpha/2}\sigma/\sqrt{n}$

2. ถ้าไม่ทราบค่า σ ใช้ S แทนก็จะได้

$\hat{\mu}_L = \bar{X} - Z_{\alpha/2}S/\sqrt{n}$

$\hat{\mu}_U = \bar{X} + Z_{\alpha/2}S/\sqrt{n}$

ในกรณีที่เราไม่ทราบค่า σ และเราใช้ตัวอย่างขนาดเล็กคือ $n \leq 30$ เราจะสร้างช่วงเชื่อมั่นได้ตามเงื่อนไขต่อไปนี้

$$\hat{\mu}_L = \bar{X} - t_{\alpha/2}^{(n-1)} \; S/\sqrt{n}$$

$$\hat{\mu}_U = \bar{X} + t_{\alpha/2}^{(n-1)} \; S/\sqrt{n}$$

ที่เราต้องเปลี่ยนแปลงเป็นรูปนี้แทนก็เพราะว่าจะได้ค่าประมาณที่ละเอียดและมีความแม่นยำสูงขึ้น ทั้งนี้เพราะในกรณีของตัวอย่างขนาดเล็กเราจะใช้ $t_{\alpha/2}^{(n-1)}$ แทน $Z_{\alpha/2}$ โดยที่ความหมายของ $t_{\alpha/2}^{(n1-)}$ หมายถึงค่าของ t ที่ให้พื้นที่ใต้โค้งเท่ากับ $\alpha/2$ เมื่อมีองค่าแห่งความเป็นอิสระ (degree of freedom (df))

วิธีการเปิดตารางค่าของ t จะเหมือนกับการเปิดตารางใช้ค่า Z ทุกประการเพียงแต่ตารางของ t ให้เปิดโดยใช้เงื่อนไขของ df เข้ามาช่วย วิธีการเปิดตาราง t ให้เปิดไปที่หน้า 314 (หนังสือสถิติเบื้องต้น) เช่น ถ้า เราใช้ขนาดตัวอย่างเป็น 16 และกำหนดให้ใช้ความเชื่อมั่น 90%

ดังนั้น $n = 16$

$df. = 16 - 1 = 15$

$\alpha = .01$

$\alpha/2 = .01/2 = .005$

ดูตารางที่ $Q = .005$ $df(\nu) = 15$

จะพบว่าให้ค่า $t = 2.947$ ประมาณ 2.95

นั่นคือ $t_{.005}^{(15)} = 2.947$

2. การประมาณค่าสัดส่วนของประชากร

สัดส่วนของประชากรใช้สัญลักษณ์คือ Π

ค่าประมาณด้วยจุดของ Π คือ p ซึ่งเราเป็นสัดส่วนของตัวอย่าง เช่นถ้าศึกษาสัดส่วนของผู้ครอบครองที่ดินในการเพราะปลูกของชาวนาในจังหวัดฉะเชิงเทรา ข้อมูลที่ศึกษาก็คือชาวนาแต่ละครอบครัวมีที่ดินที่อยู่ในครอบครองหรือไม่

ถ้าเราสุ่มตัวอย่างชาวนาในจังหวัดนี้มา 10 รายและถามว่ามีที่ดินในความครอบครองของตนเองหรือไม่ ปรากฏว่า ข้อมูลที่ได้มีดังนี้คือ

ชาวนาครอบครัวที่	1	2	3	4	5	6	7	8	9	10
ลักษณะการครอบครองที่ดิน	1	0	1	1	1	0	0	0	0	1

(โดยที่ 1 หมายถึง มีที่ดินของตนเอง 0 หมายถึง ไม่มีที่ดินเป็นของตนเอง)

ดังนั้นจากกลุ่มตัวอย่างที่ได้จะพบว่า

สัดส่วนของผู้ที่ครอบครองที่ดินจากกลุ่มตัวอย่างดังกล่าวคือ

$$p = a/n$$

โดยที่ $a = 1+0+1+\ldots+0+0+1 = 5$

$$p = a/n = 5/10$$

$$p = .5$$

ดังนั้นค่าประมาณของ Π ก็คือ $p = .5$

ถ้าจะสร้างช่วงเชื่อมั่นของ Π เราจะได้ค่า $\hat{\Pi}_L$ และ $\hat{\Pi}_U$ ดังนี้คือ

$$\hat{\Pi}_L = p - Z_{\alpha/2}\sqrt{p(1-p)/n}$$

$$\hat{\Pi}_U = p + Z_{\alpha/2}\sqrt{p(1-p)/n}$$

โดยที่ $(\hat{\Pi}_L , \hat{\Pi}_U)$ ที่ได้คือค่าประมาณ**ตัวแบบช่วงเชื่อมั่น**ของ Π **ณ.ระดับความเชื่อมั่น** **$(1-\alpha)100\%$**

3. การประมาณค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของประชากร

ความแปรปรวนของประชากรใช้สัญลักษณ์ก็คือ σ^2 ส่วนส่วนเบี่ยงเบนมาตรฐานเราใช้ σ เราประมาณค่าแบบจุดของ σ^2 คือ S^2 **ซึ่งเป็นความแปรปรวนของประชากร**

โดยที่ช่วงเชื่อมั่นของ σ^2 ก็คือ $\hat{\sigma}^2_L = (n-1)S^2/\chi^{2\,(n-1)}_{/2}$

และ $\hat{\sigma}^2_U = (n-1)S^2/\chi^{2\,(n-1)}_{\alpha/2}$

โดยที่ $\chi^{2\,(n-1)}_{\alpha/2}$ **คือค่าของตัวแปรสุ่มที่มีพื้นที่ใต้โค้งเป็น $\alpha/2$ และองศาแห่งความเป็นอิสระ**

(df) = n–1 การเปิดตาราง $\chi^{2\,(n-1)}_{\alpha/2}$ ก็ให้พิจารณาดูว่าองศาแห่งความเป็นอิสระ คือ n–1 และ α/2 เท่ากับเท่าไร ตัวอย่างเช่น ถ้าให้หาช่วงเชื่อมั่นของความแปรปรวน ณ.ระดับ 95% โดยการใช้ขนาดตัวอย่าง 10 ดังนั้นค่าที่จะเปิดจากตาราง χ^2(หน้า 312,313)คือ n – 1 = 10 – 1 = 9 และ α/2 = .05/2 = .025 เปิดตารางที่ Q = .025 และ ν = 9 ซึ่งจะให้ค่า χ^2 = 19.0228

ดังนั้น $\chi^{2\,(9)}_{.025} = 19.0228$

และ $\chi^{2\,(9)}_{1-\alpha/2} = \chi^{2\,(9)}_{1-.025} = \chi^{2\,(9)}_{.975}$

$$\chi^{2\,(9)}_{.975} = 2.70039$$

<u>แบบฝึกหัด</u>

๑. จงอธิบายความหมายของเทอมต่อไปนี้

ก. ประชากร	ข. พารามิเตอร์
ค. ตัวอย่าง	ง. ตัวสถิติ
จ. ตัวประมาณค่า	ฉ. การแจกแจงของการสุ่มตัวอย่าง

ก. ประชากรก็คือหน่วยทั้งหมดในขอบเขต(เรื่อง)ที่เราสนใจหรือสิ่งที่เราศึกษา

(ให้อ่านคำอธิบายในบทที่สรุปไว้แล้วในตอนต้นของบทนี้)

ข. พารามิเตอร์ คือคุณลักษณะของประชากรที่เราสนใจ เช่นรายได้เฉลี่ยของประชากรในประเทศไทย สัดส่วนของนักศึกษาหญิงในคณะนิติศาสตร์มหาวิทยาลัยรามคำแหง ผลผลิตข้าวที่ปลูกได้ในจังหวัดอยุธยา

ค. ตัวอย่าง ก็คือตัวแทน(บางส่วน)จากประชากร(สรุปสั้นๆก็คือหน่วยบางหน่วยในประชากร)

ง. ตัวสถิติ คือ ค่าที่ได้จากการศึกษาจากกลุ่มตัวอย่าง เช่น ค่าเฉลี่ยของกลุ่มตัวอย่าง สัดส่วนที่ได้จากกลุ่มตัวอย่าง เป็นต้น

จ. ตัวประมาณค่า คือตัวแทน(ตัวประมาณ) ของคุณลักษณะของประชากรนั่นเอง ตัวอย่างเช่น ถ้าเราต้องการศึกษาว่ามีประชากรในกรุงเทพฯคิดเป็นร้อยละเท่าไรที่ไม่เห็นด้วยกับการ ยืน

ราคารถเมล์ ในการศึกษาเรื่องนี้จะเห็นว่าเราไม่สามารถจะศึกษาโดยการถามจากคนทุกคนในกรุงเทพฯได้ทั้งนี้เพราะอาจเนื่องมาจากข้อจำกัดหลายประการด้วยกัน เช่นค่าใช้จ่าย ระยะเวลา กำลังคน เป็นต้น ดังนั้น เราจะดำเนินการศึกษาโดยการสุ่มตัวอย่างคนในกรุงเทพฯมา 500 คน แล้วสอบถามข้อคิดเห็นดังกล่าว นั่นหมายถึงคนในกรุงเทพฯ 500 คนนั้นคือตัวอย่าง(ตัวแทน) ขนาด 500 หน่วยนั่นเอง(ประชากรจริงของกรุงเทพฯ) คือ 5 ล้านคน เมื่อเราได้ข้อมูลจากคน 500 คนนี้แล้ว นำมาคิดเป็นเปอร์เซ็นต์ของคนที่ไม่เห็นด้วยกับหารขึ้นราคารถเมล์ตัวเลขที่คิดได้เราเรียกว่าตัวสถิติ เมื่อนำตัวเลขดังกล่าวไปใช้เป็นค่าประมาณของเปอร์เซ็นต์ของคนในกรุงเทพฯทั้งหมดเราจะเรียกค่าที่นำไปประมาณนี้ว่าตัวประมาณค่า(Estimator)

ฅ. จงอธิบายถึงกระบวนการในการประมาณพารามิเตอร์ประชากรต่อไปนี้ โดยอาศัยตัวอย่างสุ่ม

ก. รายได้เฉลี่ยต่อเดือนของรัฐมนตรี

ข. ค่าเช่าหอพักในย่านหัวหมาก

ค. เวลาที่ใช้เดินทางจากบ้านมาที่มหาวิทยาลัยในตอนเช้า ของนักศึกษารัฐศาสตร์

ง. เปอร์เซ็นต์ของนักศึกษาที่เห็นด้วยกับการเปลี่ยนแปลงโครงสร้างของสภานักศึกษา ม.ร.

จ. รายได้ต่อเดือนของครอบครัวชาวนาไทยในภาคอิสาน

ฉ. รายได้เริ่มต้นของนักศึกษา ม.ร. ที่เห็นด้วยกับรัฐบาลที่ต้องการให้ ม.ร. รับนักศึกษาจำนวนจำกัด

ก. รายได้เฉลี่ยของรัฐมนตรีต่อเดือน

เราจะหาค่าประมาณของรายได้ต่อเดือนของรัฐมนตรีได้โดยการใช้กระบวนการต่อไปนี้

1. สุ่มตัวอย่างรัฐมนตรีมาจำนวนหนึ่ง(เช่นสุ่มมา 5 คนหรือ 6 คนหรือ7คน หรืออื่นๆ) การกำหนดขนาดตัวอย่างนี้ขึ้นอยู่กับข้อจำกัดถึงที่เคยกล่าวมาแล้วแต่ถ้าเรายิ่งได้ขนาดตัวอย่างใหญ่มากเพียงใดเราจะได้ข้อสนเทศมากซึ่งยังผลให้เราสามารถประมาณค่าได้ใกล้เคียงกับความเป็นจริงของประชากรได้ ในที่นี้สมมุติว่า เราสุ่มตัวอย่างรัฐมนตรีมา 5 คน ถามว่ารายได้ต่อเดือนของแต่ละคนเป็นเท่าไรเช่นข้อมูลที่ได้เป็นดังนี้คือ

รายได้ของรัฐมนตรี	คนที่1	คนที่2	คนที่3	คนที่4	คนที่5
	15000	23000	18000	45000	27000

ดังนั้นรายได้เฉลี่ยจากตัวอย่าง =(15000+23000+18000+45000+37000)/5

= 27000 บาท

ดังนั้นค่าประมาณสำหรับรายได้ต่อเดือนของรัฐมนตรี 27600 บาท

<u>หมายเหตุ</u> เนื่องจากข-ฉ จะมีวิธีการดำเนินการคล้ายคลึงกับการดำเนินการในข้อก. จึงขอให้นักศึกษาลองแสดงความคิดเห็นในการดำเนินการข้อข-ฉเอง

๔. จงสร้างช่วงเชื่อมั่น ๙๕ สำหรับพารามิเตอร์ประชากร จากการศึกษาโดยอาศัยตัวอย่างสุ่ม ดังต่อไปนี้

ก. เจ้าหน้าที่กรมการค้าภายในได้สุ่มผงซักฟอกชนิด ๑ กก. มา ๓๖ กล่อง ได้น้ำหนักเฉลี่ย ๐.๙๕ กก. ส่วนเบี่ยงเบนมาตรฐาน ๐.๑๓๕ กก. (สมมติว่าน้ำหนักผงซักฟอกต่อกล่องมีการแจกแจงแบบปกติ)

ข. จากการสำรวจความคิดเห็นของนักศึกษา ๕๐๐ ราย พบว่า ๔๐๐ ราย เห็นด้วยกับการใช้รถจักรยานไปทำงาน

ค. จากการศึกษารายได้ต่อวันของครอบครัวชาวนาไทยในตำบลหนึ่งโดยอาศัยตัวอย่างครอบครัว ๒๕ ครอบครัว ได้รายได้รวม ๑๒๕๐ บาท ความแปรปรวนของรายได้เป็น ๑๒๕๐ บาท (สมมติว่ารายได้ต่อวันมีการแจกแจงแบบปกติ)

ง. จากการศึกษาความแตกต่างระหว่างรายได้ต่อวันของกรรมกร โดยอาศัยตัวอย่าง ได้ส่วนเบี่ยงเบนมาตรฐาน ๑๐ บาท จากกรรมกร ๒๑ ราย (สมมติว่ารายได้มีการแจกแจงแบบปกติ)

ก. สุ่มผงซักฟอกมา 36 กล่อง นั่นคือ $n = 36$

โดยที่ได้น้ำหนักเฉลี่ย 0.95 กก.นั่นคือ $\bar{X} = 0.95$

ส่วนเบี่ยงเบนมาตรฐาน 0.135 กก.นั่นคือ $S = 0.135$

ช่วงเชื่อมั่น 95% ของพารามิเตอร์ของประชากรในที่นี้คือค่าเฉลี่ยของประชากรซึ่งใช้สัญลักษณ์ จะมีค่าเท่ากับ

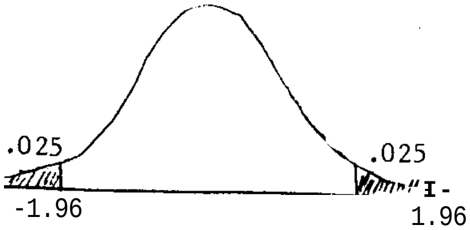
$$\hat{\mu}_L = \bar{X} - Z_{\alpha/2}S/\sqrt{n}$$

$$\hat{\mu}_U = \bar{X} + Z_{1-\alpha/2}S/\sqrt{n}$$

$\alpha = 1 - 95/100 = .05, \quad \alpha/2 = .05/2 = .025$

$z_{\alpha/2} = Z_{.025}$

เปิดตาราง $z_{\alpha/2} = 1.96$



ดังนั้น

$$\hat{\mu}_L = 0.95 - 1.96(0.135)/\sqrt{36} = .9059$$

$$\hat{\mu}_U = 0.95 + 1.96(0.135)/\sqrt{36} = .9941$$

ดังนั้นช่วง เชื่อมั่น 95% ของค่า เฉลี่ยของประชากรคือ $.9059 < \mu < .9941$

ข. $n = 500$, $p = 400/500 = 4/5$

ดังนั้นช่วงเชื่อมั่นของสัดส่วนของนักศึกษาทั้งหมดที่เห็นด้วยกับการใช้รถจักรยานไปทำงาน คือ

$$\Pi = p \pm Z_{.025}\sqrt{p(1-p)/n}$$

$$= \frac{4}{5} \pm 1.96\sqrt{\frac{4}{5}\left(1-\frac{4}{5}\right)/500}$$

$$= .8 \pm 1.96\sqrt{.8(.2)/500}$$

$$= .76, .82$$

ค. $n = 25$, $\Sigma X_i = 1250$, $\bar{X} = 1250/25 = 50$

$S^2 = 1250$ $S = 35.35$

(ความแปรปรวนคือค่าของ S^2)

กรณีนี้เราไม่ทราบความแปรปรวนของประชากร (σ^2) และเนื่องจากเราใช้ขนาดตัวอย่าง $n < 30$

ดังนั้น เราจึงใช้สูตร $\mu = \bar{X} \pm t_{\alpha/2}^{(n-1)} S/\sqrt{n}$

$$\mu = 50 \pm t^{(25-1)}_{.025} (35.35)/\sqrt{25}$$

$$= 50 \pm 2.064\ (35.35)/5$$

$$= 35.40\ ,\ 64.59$$

ดังนั้นช่วงเชื่อมั่น 95% ของรายได้เฉลี่ยต่อชั่นของครอบครัวชาวนาไทย ในตำบลที่ศึกษาจึงอยู่ในช่วง 35.40 บาท - 64.59 บาท

ง. $n = 21$ ราย $S = 10$ บาท

ประมาณค่า ความแปรปรวนของประชากร(σ)

$$\hat{\sigma}^2_L = (n-1)S^2/\chi^{2\,(n-1)}_{\alpha/2}$$

$$= (21-1)100/\chi^{2\,(21-1)}_{.025}$$

$$= 20(100)/34.1696$$

$$= 58.53$$

$$\hat{\sigma}^2_U = (n-1)/\chi^{2\,(21-1)}_{1-\alpha/2}$$

$$= (21-1)100/{}^{(21-1)}_{.975}$$

$$= 20(100)/9.59078 = 208.53$$

$$\hat{\sigma}_L = \sqrt{58.53} = 7.62$$

$$\hat{\sigma}_U = \sqrt{208.53} = 14.44$$

๕. จงสร้างช่วงเชื่อมั่น ๙๕ สำหรับผลต่างของพารามิเตอร์ประชากรของสองประชากร จากการศึกษาต่อไปนี้

ก. ในการศึกษารายได้ต่อวันของกรรมกรในโรงงานทอผ้า กับโรงงานยาสูบ ได้ข้อมูลสรุป ดังนี้

	ขนาดตัวอย่าง	รายได้เฉลี่ย	ความแปรปรวน
โรงงานทอผ้า	๒๕	๕๕	๑๐๐
โรงงานยาสูบ	๑๕	๗๐	๑๐๕

(สมมติว่ารายได้มีการแจกแจงแบบปกติ)

ข. ในการสำรวจทัศนคติของนักศึกษาชายและหญิง เกี่ยวกับมาตรการแบบหนึ่งที่รัฐบาลนำออกมาใช้เพื่อแก้ปัญหาจราจร ได้ข้อมูลสรุปดังนี้

	ขนาดตัวอย่าง	เห็นด้วย	ไม่เห็นด้วย
ชาย	๑๐๐	๗๐	๓๐
หญิง	๑๒๐	๗๐	๕๐

ก. การหาผลต่างของรายได้ต่อวันของกรรมกรในโรงงานทอผ้ากับโรงงานยาสูบ

	ขนาดตัวอย่าง	รายได้เฉลี่ย $\bar{X}$	ความแปรปรวน S^2
โรงงานทอผ้า	25	45	100
โรงงานยาสูบ	15	70	105

สร้างช่วงเชื่อมั่น 95% สำหรับผลต่างของพารามิเตอร์ของประชากร ในที่นี้มี 2 ประชากรคือ กรรมกรทอผ้า และกรรมกรโรงงานยาสูบ

ให้หาค่าเฉลี่ยของรายได้เฉลี่ยของประชากรของโรงงานยาสูบ = μ_1

ให้หาค่าเฉลี่ยของรายได้ของประชากรกรรมกรโรงงานทอผ้า = μ_2

เนื่องจากเราไม่ทราบค่า σ_1^2 และ σ_2^2 เราจึงใช้ S_p^2 แทนสูตรที่ใช้เป็น

$$\mu_1 - \mu_2 = (\bar{X}_1-\bar{X}_2) \pm t_{\alpha/2}^{(\nu)} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

สมมุติว่า $\sigma_1^2 = \sigma_2^2$ (ค่านี้เราไม่ทราบ)

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

$$= \frac{(15-1)105+(25-1)100}{(15-1)+(25-1)}$$

$$= 3870/38$$

$$= 101.84$$

$$S_p = 10.9$$

$$\nu = n_1+n_2-2 = 38$$

$$t^{38}_{.025} = 2.021 \quad (\text{ใช้ค่า } t^{40}_{025} \text{ แทน})$$

$$\mu_1 - \mu_2 = (70-45) \pm 2.021(10.09)\sqrt{\frac{1}{25}+\frac{1}{25}}$$

$$= (70-45) \pm 2.021(10.09)\sqrt{\frac{1}{25}+\frac{1}{25}}$$

$$= 25 \pm 2.01(10.09)(.106)$$

$$= 25 \pm 2.16$$

$$= 22.84, 27.16$$

ข. จากการสำรวจทัศนคติของนักศึกษาชายและหญิง เกี่ยวกับมาตรการแบบหนึ่งที่รัฐบาลนำออกมาใช้เพื่อแก้ปัญหาการจราจร ได้ข้อมูลสรุปดังนี้

	ขนาดตัวอย่าง	เห็นด้วย	ไม่เห็นด้วย
ชาย	100	70	30
หญิง	120	70	50

จงสร้างช่วงเชื่อมั่น 95% ของสัดส่วนสำหรับผลต่างของสัดส่วนของนักศึกษาชายและหญิงที่เห็นด้วยกับมาตรการแก้ปัญหาจราจร กำหนดให้ประชากรกลุ่มที่1คือนักศึกษาชาย สุ่มตัวอย่างมา $n_1 = 100$คน ได้ว่าสัดส่วนจากตัวอย่างสำหรับผู้ที่เห็นด้วยคือ $P_1 = 70/100 = .7$

ประชากรกลุ่มที่2 คือนักศึกษาหญิง สุ่มตัวอย่างมา $n_2 = 120$ คน ได้ว่าสัดส่วนของตัวอย่างที่เห็นด้วยคือ $P_2 = 70/120$

$$\Pi_1 - \Pi_2 = (P_1 - P_2) \pm Z_{\alpha/2}\sqrt{P_1Q_1/n_1 + P_2Q_2/n_2}$$

$$= \left(\frac{70}{100} - \frac{70}{120}\right) \pm 1.96\sqrt{\frac{70}{100}\cdot\frac{30}{100}\cdot\frac{1}{100} + \frac{70}{120}\cdot\frac{50}{120}\cdot\frac{1}{120}}$$

$$= .116 \pm 1.96\sqrt{.002 + .002}$$

$$= .116 \pm 0.1255$$

$$= -0.0095,\ 0.2415$$

การทดสอบสมมุติฐาน

สมมุติฐานก็คือข้อเสนอเกี่ยวกับลักษณะของประชากร (พารามิเตอร์) สมมุติฐานที่กล่าวมานั้น อาจจะเป็นจริงหรือเป็นเท็จก็ได้ โดยปกติแล้วการที่จะสรุปว่าสมมุติฐานเป็นจริงหรือเป็นเท็จนั้น เราจะต้องอาศัยข้อสนเทศจากตัวแทนของประชากร (ตัวอย่าง) เป็นเครื่องอ้างอิง ถ้าหากข้อมูลที่ได้จากตัวอย่างบ่งชี้ว่าสมมุติฐานเป็นจริง เราก็จะยอมรับสมมุติฐานนั้น แต่ถ้าหากข้อมูลจากตัวอย่างไม่สอดคล้องกับสมมุติฐานเราก็จะปฏิเสธสมมุติฐานนั้น

ขั้นตอนในการทดสอบสมมุติฐาน

1. ให้กำหนดสมมุติฐานเชิงสถิติขึ้น สมมุติฐานนั้นจะประกอบด้วย

   สมมุติฐานหลัก (H_o) และสมมุติฐานรอง (H_a)

สมมุติฐานหลัก (H_o) มีอยู่ 3 แบบคือ

$$1.\ H_o: \theta = \theta_o \quad 2.\ H_o: \theta \le \theta_0 \quad 3.\ H_0: \theta \ge \theta_o$$

โดยที่ θ คือคุณลักษณะของประชากรที่เราสนใจที่จะทดสอบ เช่น θ อาจจะเป็นค่าเฉลี่ยของประชากร (μ) ค่าความแปรปรวนของประชากร (σ^2) หรือค่าสัดส่วนของสิ่งที่เราสนใจในประชากร (Π) ส่วน Θ_0 คือค่าคงที่ (ตัวเลข) ซึ่งเราตั้งขึ้นเพื่อศึกษาคุณลักษณะของประชากร เช่น "ประชากรคือคนไทยมีรายได้เฉลี่ย 3500 บาท/ปี" นั่นคือ $H_o: \mu = 3500$

หรือ ถ้าเรากล่าวว่า ประชากรไทยมีรายได้เฉลี่ยมากกว่า 3500 บาท สมมุติฐานจะเป็น

$H_o: \quad \mu \geq 3500$ บาทเป็นต้น

การกำหนดสมมุติฐานหลักได้ย่อมส่งผลไปถึงสมมุติฐานรอง (H_a) ด้วย โดยที่ สมมุติฐานรองก็คือสมมุติบานที่ตรงกันข้ามกับสมมุติฐานหลักนั่นเอง เช่น

ถ้า	$H_o: \mu = 3500$	$Ha: \mu \neq 3500$	(1)
หรือ	$H_o: \mu \geq 3500$	$Ha: \mu < 3500$	(2)
หรือ	$H_o: \mu \leq 3500$	$Ha: \mu > 3500$	(3)

การตั้งสมมุติฐานในลักษณะที่(1) เรียกว่า Two-sided Test หรือสมมุติฐาน 2 ทาง

สมมุติฐานในลักษณะที่(2) เรียกว่า One-sided Test หรือสมมุติฐานทางเดียว (ทางขวา) และการตั้งสมมุติฐานแบบที่3 เรียกว่า One-sided Test หรือสมมุติฐานทางเดียว (ทางด้านซ้าย)

การทราบว่าเป็นชนิดหางเดียวหรือ 2 หางอยู่ทางด้านซ้ายหรือทางด้านขวาจะส่งผลไปถึงการเลือกค่าวิกฤติเพื่อจะใช้ในการทดสอบที่จะกล่าวถึงในข้อ2

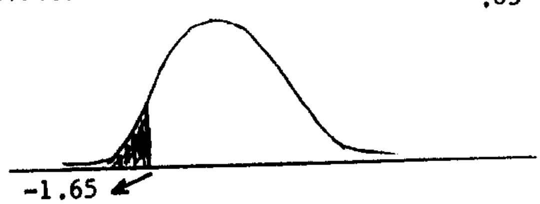
2. ให้เลือกกำหนดระดับนัยสำคัญ(ค่า α)

โดยทั่วไป α ที่นิยมใช้ก็คือ(ค่า α) α = .01 , α= .05

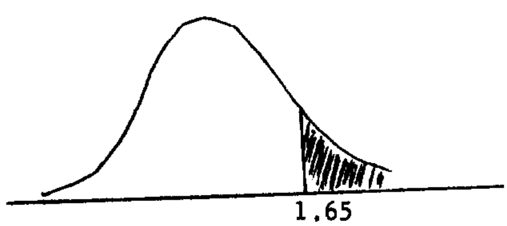
การกำหนดขนาดของ α จะทำให้เราสามารถหาค่าวิกฤติ(ค่าที่ใช้ตัดสินใจ) ได้ ค่าวิกฤติก็คือค่าที่ได้จากตาราง เช่นค่า Z, t, χ^2 เป็นต้น

ตัวอย่างเช่นถ้าใช้ตาราง Z ในการแก้ปัญหาของเรา จะต้องพิจารณาว่าสมมุติฐานชนิด One-sided Testหรือ Two-sided Test ถ้าสมมุติฐานเป็นOne-sided test ค่าของเขตวิกฤติของ Z ก็คือ $|Z_\alpha|$ ตัวอย่างเช่น ถ้าเลือก α=.05 และสมมุติฐานเป็นชนิด One-sided Test $|Z_{.05}|$ = 1.65

หรือ ถ้าพิจารณารายละเอียดต่อไป ถ้าเป็นกรณีหางซ้าย $Z_{.05}$ = -1.65



และถ้าเป็นกรณีของหางขวา $Z_{.05}$ = 1.65



3. ให้เลือกตัวสถิติที่จะทดสอบ ให้ใช้หลักเกณฑ์ในการเลือกโดยดูจากลักษณะของสมมุติฐานที่จะทดสอบ เช่นใช้ 1. $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

5. Z =

6. $$t = \frac{(X_1-X_2) - D_o}{S_p\sqrt{1/n_1+1/n_2}}$$

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

แบบฝึกหัด

๖. จงกำหนดสมมติฐาน (Ho, Ha) จากคำกล่าวต่อไปนี้

ก. ค่าใช้จ่ายเฉลี่ยรายเดือนของนักศึกษา ม.ร. มากกว่า ๑๐๐๐ บาท

ข. รายได้ต่อวันของครอบครัวชาวนาไทยในภาคอิสาณน้อยกว่า ๑๐ บาท

ค. นักศึกษารัฐศาสตร์ ม.ร. เห็นด้วยกับนโยบายของรัฐบาลในการแก้ปัญหาเศรษฐกิจมากกว่าครึ่งหนึ่ง

ง. ปุ๋ยยี่ห้อ ก ให้ผลผลิตมากกว่าปุ๋ยยี่ห้อ ข

ก. ค่าใช้จ่ายเฉลี่ยรายเดือนของนักศึกษามร.มากกว่า 1000 บาท

H_o: $\mu \leq 1000$ vs Ha: $\mu > 1000$

ข. รายได้ต่อวันของครอบครัวชาวนาไทยในภาคอิสาณน้อยกว่า 10 บาท

H_o: $\mu \geq 20$ vs Ha: $\mu < 20$

ค. นักศึกษารัฐศาสตร์มร.เห็นด้วยกับนโยบายของรัฐบาลในการแก้ปัญหาเศรษฐกิจมากกว่าครึ่งหนึ่ง

H_o: $\Pi \leq \frac{1}{2}$ vs Ha: $\Pi > \frac{1}{2}$

ง. ปุ๋ยยี่ห้อก. ให้ผลผลิตมากกว่าปุ๋ยยี่ห้อข.

H_o: $\mu_D \leq D_o$ vs Ha: $\mu_D > D_o$

$$\mu_D = \mu_1 - \mu_2$$

๗. จากการสำรวจในอุตสาหกรรมเฉพาะกรุงเทพมหานคร เมื่อปี ๒๕๑๐ เราได้ค่าจ้างเฉลี่ยต่อวันเป็น ๔๔ บาท นักปกครองคนหนึ่งคาดว่า ปัจจุบันนี้ค่าจ้างเฉลี่ยจะเพิ่มขึ้นอีกประมาณ ๔ บาท

ก. จงกำหนดสมมติฐานในการทดสอบ (Ho, Ha)

ข. ในการสรุปผลการตรวจสอบนั้น จะเกิดความคลาดเคลื่อนแบบ ๑ หรือ

ก. H_o: $\mu_{2523} \leq$ 50 บาท vs H_1: $\mu_{2523} > 50$

ข. ความคลาดเคลื่อนแบบที่ 1 ที่เกิดขึ้นก็คือถ้าในปัจจุบันค่าจ้างเฉลี่ยไม่ถึง 50 บาทต่อวันเป็นจริง แล้วเราจะปฏิเสธสมมุติฐานหลัก(ทั้งๆที่จริงเราควรจะยอมรับ) กรณีเช่นนี้จะเกิดความคลาดเคลื่อนชนิดที่1(α) ส่วนความคลาดเคลื่อนชนิดที่2(β) ก็คือ ถ้าปรากฏว่าในภาวะการณ์ปัจจุบันนี้จริงๆค่าจ้างเฉลี่ยมีค่าเท่ากับ 50 บาทต่อวัน แต่เราไม่ยอมรับสมมุติฐานที่ว่า รายได้เฉลี่ยที่เพิ่มขึ้นจากปี2520 ไม่ถึง 5 บาท กรณีเช่นนี้เป็นความคลาดเคลื่อนชนิดที่2

ค. ตัวสถิติที่เลือกมาใช้ก็คือ ตัว Z โดยที่ Z จะมีการแจกแจงแบบปกติ

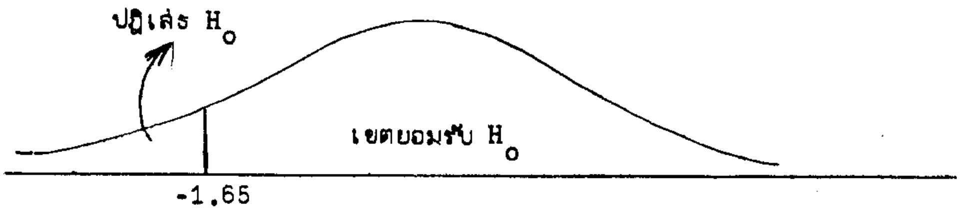
ง. จากโจทย์ $n = 1600$, $\bar{X} = 47$, $S = 20$

$$Z = \frac{47-50}{20/\sqrt{1600}}$$

$$= -3/(20/40) = -1$$

$$Z_\alpha = Z_{.05}$$

จากค่าวิกฤติที่กำหนดให้



พิจารณาค่าสถิติที่ได้ $Z = -6$ แต่ค่าวิกฤติเท่ากับ -1.65 ดังนั้นเราจึงปฏิเสธสมมุติฐานหลักที่ว่า รายได้เฉลี่ยของคนงานเท่ากับ 45บาทในปี 2523 และยอมรับว่า ปี 2523 รายได้เฉลี่ยของคนงานเพิ่มขึ้นจากปี2510 อีก 5 บาท

๘. บัณฑิตคนหนึ่งกล่าวว่า" เงินเดือนเริ่มต้นของบัณฑิตมร.จะขึ้นอยู่กับสาขาวิชาที่จบ" เพื่อที่จะสนับสนุนคำกล่าวนี้ เขาจึงทำการศึกษาโดยใช้ตัวอย่างของบัณฑิตที่ทำงานและจบปีที่แล้วจำนวน ๒๐๐ ราย ได้ข้อมูลสรุปดังนี้

	เงินเดือนเริ่มต้น	น้อยกว่า ๒๕๐๐	๒๕๐๐-๓๕๐๐	มากกว่า ๓๕๐๐
สาขาวิชาที่จบ	นิติศาสตร์	๓๐	๖๐	๒๐
	เศรษฐศาสตร์	๑๐	๒๐	๒๐
	รัฐศาสตร์	๒๐	๔๐	๑๐
	ศึกษาศาสตร์	๒๐	๑๐	๐

ข้อนี้เป็นการทดสอบเกี่ยวกับการทดสอบความเป็นอิสระของ 2 ตัวแปร(ดูรายละเอียดหน้า -229

1. โดยที่ตัวแปรทั้งสองคือ คณะที่จบและรายได้ที่ได้รับ เราจะตั้งสมมุติฐานได้ดังนี้คือ

H_o: คณะที่จบและรายได้เป็นอิสระกัน(หรืออาจจะกล่าวได้ว่ารายได้ไม่ขึ้นอยู่กับคณะที่จบ)

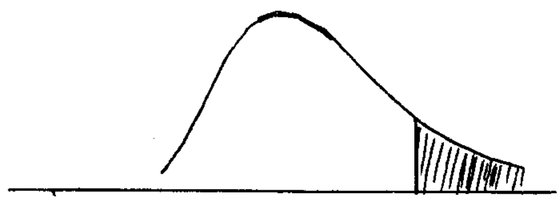
Ha: รายได้ขึ้นอยู่กับคณะที่จบ

สาขาวิชาที่จบ \ เงินเดือนเริ่มต้น	น้อยกว่า 2500	2500-3500	มากกว่า 3500	รวม
นิติศาสตร์	(33.85) 30	(55) 60	(21.15) 20	110
รัฐศาสตร์	(21.54) 10	(35) 40	(13.46) 10	70
ศึกษาศาสตร์	(9.23) 20	(15) 10	(5.77) 0	30
รวม	80	130	50	260

2. กำหนดให้ =.05 n= 260

$$\chi^2 = \sum_{i,j}^{c,k} (O_{ij}-E_{ij})^2/E_{ij}$$

ปฏิเสธ H_o: ถ้า χ^2 (ที่คำนวณ) > $\chi^{2\,(3-1)(4-1)}_{\cdot 08}$ = 12.5916



ตัวอย่างในการคิดคำนวณหาค่า E_{ij} เช่น

$$E_{11} = (80x110)/260 = 33.85$$

$$= \sum_{i,j}^{r,c} (O_{ij}-E_{ij})^2/E_{ij}$$

$$\chi^2 = \frac{(30-33.85)^2}{33.85}+\frac{(10-15.38)^2}{15.38}+\ldots+\frac{(0-5.77)^2}{5.77}$$

$$= (r-1)(c-1) = (4-1)(3-1) = 6$$

$$\chi^2 = 28.7320$$

5. สรุปผล เราจะปฏิเสธสมมุติฐานหลักที่ว่า คณะที่จบและรายได้เป็นอิสระต่อกัน นั่นคือ รายได้มีผลเนื่องมาจากการศึกษาในสาขาวิชาต่าง (คณะ)

1. H_o : ชาวนาในแต่ละตำบลมีระดับรายได้เป็นเอกภาพกัน

   Ha : ชาวนาในแต่ละตำบลมีระดับรายได้ไม่เป็นเอกภาพกัน

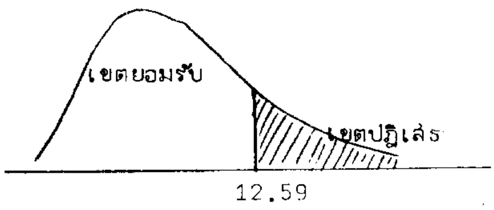
2. ระดับนัยสำคัญ $\alpha = .05$

3. ตัวสถิติที่จะทดสอบ

$$\chi^2 = \sum_{i,j}^{r,c} (O_{ij}-E_{ij})^2/E_{ij}$$

เกณฑ์การตัดสินใจคือ ปฏิเสธ H_o ถ้า $\chi^2_{\text{คำนวณ}} > \chi^{2\,(3-1)(4-1)}_{.05}$

โดยที่ $\chi^{2\;\;6}_{.05} = 12.59$



4.

ตำบล \ ระดับรายได้	สูง	ปานกลาง	ต่ำ	ขนาดตัวอย่าง
ก.	10%(100)	50%(500)	40%(400)	1000
ข.	10%(120)	30%(360)	60%(720)	1200
ค.	20%(100)	50%(250)	30%(150)	500
ง.	20%(60)	40%(120)	40%(120)	300
รวม	380	1230	1390	3000

<u>หมายเหตุ</u> ค่าที่อยู่ในวงเล็บคือค่าที่คิดออกมาเป็นขนาดตัวอย่างแล้ว

ระดับรายได้		สูง	ปานกลาง	ต่ำ	รวม
ตำบล	ก.	(126.66) 100	(410) 500	(463.33) 400	1000
	ข.	(152) 120	(492) 360	(556) 720	1200
	ค.	(63.33) 100	(205) 250	(231.66) 150	500
	ง	(38) 60	(123) 120	(139) 120	300
รวม		380	1230	1390	3000

$$E_{11} = \frac{1000 \times 380}{3000} = 126.66$$

$$E_{12} = \frac{1000 \times 1230}{3000} = 410$$

$$\vdots$$

$$E_{33} = \frac{300 \times 1390}{3000} = 139$$

$$\chi^2 = \sum_{i,j}^{r,c} (O_{ij}-E_{ij})^2/E_{ij}$$

$$= \frac{(100-126.66)^2}{126.66} + \frac{(500-410)^2}{410} + \ldots + \frac{(120-139)^2}{139}$$

$$= 199.852$$

5. สรุปผล เนื่องจาก χ^2 ที่คำนวณได้มีค่ามากกว่า χ^2 จากตาราง เราจึงปฏิเสธสมมุติฐานหลัก และยอมรับสมมุติฐานรองที่ว่า ชาวนาในแต่ละตำบลมีระดับรายได้เป็นเอกภาพกัน

๑๐. จงกำหนดสมมุติฐานที่จะทดสอบ (Ho,Ha) ตัวสถิติทดสอบ เกณฑ์การตัดสินใจ และสรุปผลการทดสอบสมมุติฐานต่อไปนี้

ก. ในการศึกษาโดยอาศัยตัวอย่างสุ่ม เพื่อทดสอบคำกล่าวที่ว่า "รายได้ต่อปีของครอบครัวเกษตรกรไทยในภาคอิสาณน้อยกว่า ๓๐๐๐ บาท" นั้นได้ข้อมูลสรุปดังนี้

ขนาดตัวอย่าง	๑๐๐๐๐ ครอบครัว
รายได้เฉลี่ย	๒๕๐๐ บาท
ส่วนเบี่ยงเบนมาตรฐาน	๒๐๐๐ บาท

ข. ในการทดสอบคำกล่าวที่ว่า "รายได้เฉลี่ยต่อเดือนของครอบครัวที่อาศัยอยู่ในสลัมแห่งหนึ่งจะน้อยกว่า ๑๐๐๐ บาท" ได้ใช้ตัวอย่างสุ่มของครอบครัว ๑๐๐ ราย ปรากฏว่ามีครอบครัวที่มีรายได้ต่ำกว่า ๑๐๐ บาท ถึง ๗๐ ครอบครัว

ค. ในการศึกษาครอบครัวตัวอย่างของเกษตรกรไทยในคลองจินดาจำนวน ๑๐๐ ราย เพื่อทดสอบคำกล่าวที่ว่า " รายได้เฉลี่ยต่อปีของเกษตรกรในปี ๒๕๑๕ และปี ๒๕๒๐ เป็น ๓๐๐๐๐ และ ๓๑๕๐๐ บาทตามลำดับ และได้ส่วนเบี่ยงเบนมาตรฐานของการเปลี่ยนแปลงในรายได้เป็น ๑๐๐๐ บาท

ง. ครูป.๑ ก มีความสนใจในการแจกแจงอายุของนักเรียนในชั้น โดยเชื่อว่า "นักเรียนในชั้นจะมีส่วนเบี่ยงเบนมาตรฐานของอายุน้อยกว่า ๑ ปี" จากตัวอย่างนักเรียน ๒๑คน ได้ส่วนเบียงเบนมาตรฐานของอายุเป็น ๐.๙๕ ปี

ก. ในการศึกษาโดยอาศัยตัวอย่างสุ่ม เพื่อทดสอบคำกล่าวที่ว่า "รายได้เฉลี่ยต่อปีของครอบครัวไทยในภาคอิสาณน้อยกว่า ๓๐๐๐ บาท" นั้นได้ข้อมูลสรุปดังนี้

ขนาดตัวอย่าง ๑๐๐๐๐ ครอบครัว

รายได้เฉลี่ย ๒๕๐๐ บาท

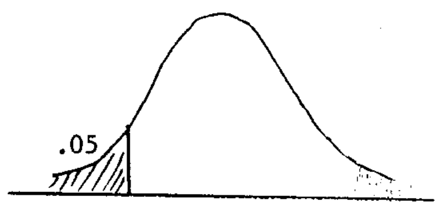
ส่วนเบี่ยงเบนมาตรฐาน ๒๐๐๐ บาท

1. $H_o: \mu \geq 3000$ vs $Ha: \mu < 3000$

2. $\alpha = .05$ $n = 10000$

$\bar{X} = 2500$ $s = 2000$

3. $Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$



$$Z = \frac{2500-3000}{2000/\sqrt{10000}}$$

$$= (-500)/(2000/100)$$

$$= -25$$

4. จะพบว่า Z ที่คำนวณได้ ตกอยู่ในเขตปฏิเสธ Ho นั่นคือ เราจะปฏิเสธHo และยอมรับ Haสรุปได้ว่า "รายได้ต่อปีของครอบครัวเกษตรกรไทยในภาคอิสาณน้อยกว่า 3000บาท"

ข.

1. $Ho: \Pi \geq .5$ vs $Ha: \Pi < .5$

(Π_0 คือรายได้เฉลี่ยต่อเดือนที่น้อยกว่า 1000 บาท)

2. $\alpha = .05$, $n = 100$ $p = 70/100 = .7$

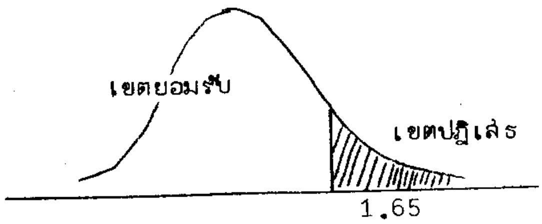
3. ตัวสถิติที่จะใช้ทดสอบ ก็คือ $Z = \frac{P-\Pi_0}{\sqrt{\Pi_0(1-\Pi_0)/n}}$

4. $Z = \frac{.7 - .5}{\sqrt{.5(1-.5)/100}}$

$= .2/(\sqrt{.25/100})$

$= 20/.5 = 40$

5.



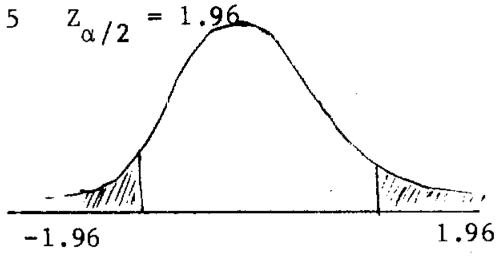
ที่คำนวณได้มีค่ามากกว่า Z จากตาราง ดังนั้นเราจึงปฏิเสธสมมุติฐานหลัก นั่นคือ "รายได้เฉลี่ยต่อเดือนของครอบครัวที่อาศัยอยู่ในลำสัมแห่งหนึ่งจะน้อยกว่า 1000 บาท"

ค. ในการศึกษาครอบครัวตัวอย่างของเกษตรกรไทยในคลองจินดาจำนวน ๑๐๐ ราย เพื่อทดสอบคำกล่าวที่ว่า "รายได้เฉลี่ยต่อปี ของเกษตรกรในปี ๒๕๑๕ และปี ๒๕๒๐ เป็น ๓๐๐๐๐ และ ๓๑๕๐๐ บาท ตามลำดับ และได้ส่วนเบี่ยงเบนมาตรฐานของการเปลี่ยนแปลงในรายได้เป็น ๑๐๐๐ บาท

1. $H_o: \mu_{20} - \mu_{15} \quad 31500-3000 = 1500$

$H_a: \mu_{20} - \mu_{15} \neq 1500$

2. $\alpha = .5 \quad Z_{\alpha/2} = 1.96$



$\mu_{20} - \mu_{15} = D_o = 1500$

$n_1 = n_2 = 100$

$\sigma = 1000$

3. เนื่องจากโจทย์ข้อนี้มิได้กำหนดค่าเฉลี่ยของรายได้ที่ได้จากกลุ่มตัวอย่างในปี 2515 และปี 2520 คือ $\bar{X}$, $\bar{Y}$