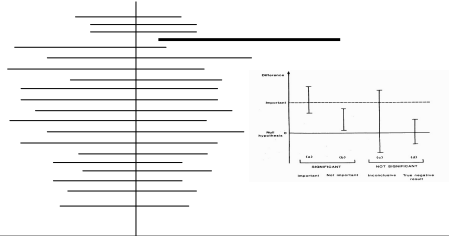


การประมาณค่า (Estimation)



ผู้ช่วยศาสตราจารย์นิคม ถนนอมเสียง
สาขาวิชาวิทยาการระบาดและชีวสถิติ
คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น
Email nikom@kku.ac.th
web <http://home.kku.ac.th/nikom>

สถิติในการวิเคราะห์ข้อมูล

1. สถิติเชิงพรรณนา (Descriptive Statistics)

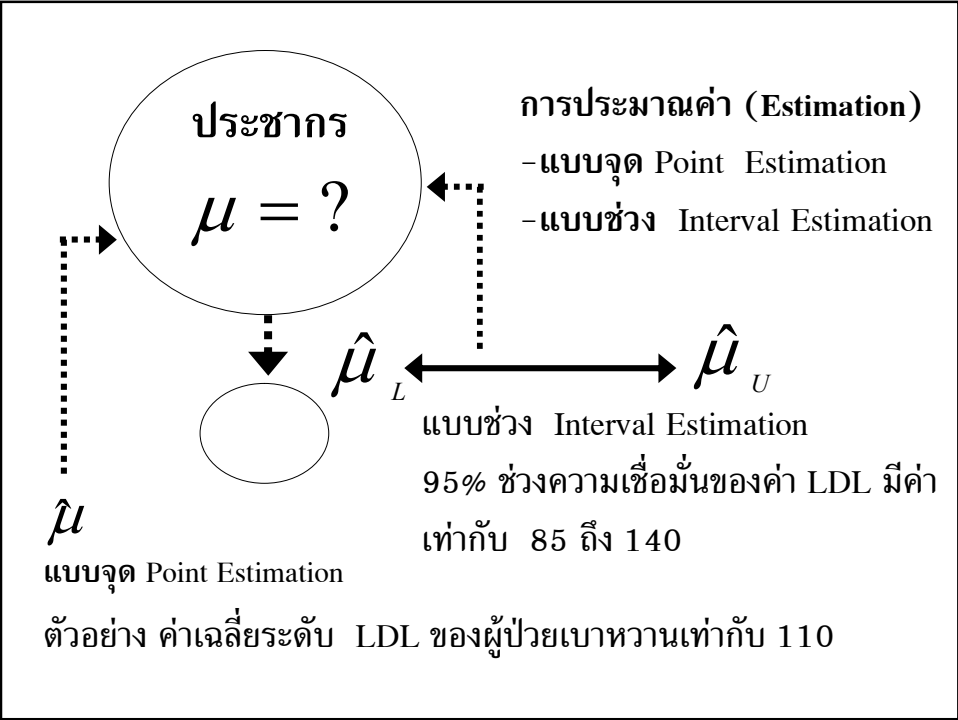
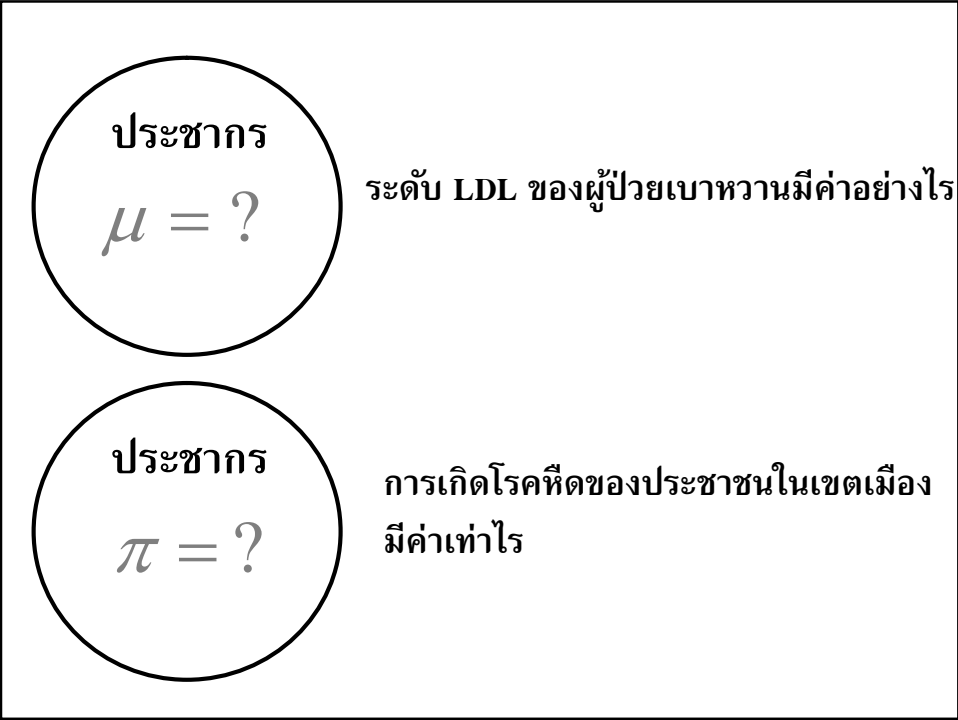
อธิบาย สรุป คุณลักษณะของตัวอย่าง

2. สถิติเชิงอนุมาน (Inferential statistics)

สรุป อ้างอิงไปถึงประชากร ประกอบด้วย

-การทดสอบสมมุติฐาน (Hypothesis testing)

-การประมาณค่า (Estimation)



การประมาณค่าพารามิเตอร์:

ช่วงความเชื่อมั่น (Confidence Interval: CI)

ถ้า พารามิเตอร์ที่สนใจ คือ ค่าเฉลี่ย (Mean)

กลุ่มเดียว - ประมาณ μ

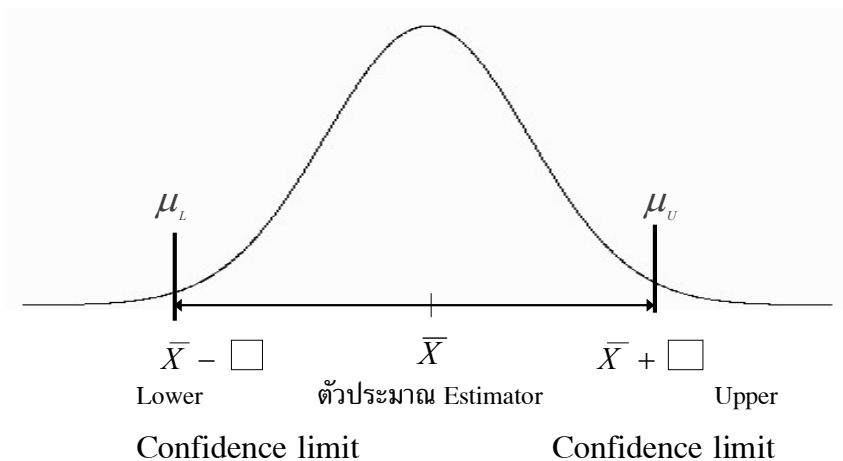
“ต้องการประมาณค่าเฉลี่ย blood glucose ของผู้ป่วยเบาหวาน”

2 กลุ่ม - ความแตกต่างค่าเฉลี่ย ($\mu_1 - \mu_2$)

“ประมาณความแตกต่าง ค่าเฉลี่ย ความดันโลหิต
ของคนสูบบุหรี่ และคนไม่สูบบุหรี่”

ถ้า พารามิเตอร์ที่สนใจ คือ ... ฯลฯ

การประมาณค่าแบบช่วงความเชื่อมั่น



ปัจจัยที่มีผลต่อช่วงความเชื่อมั่น

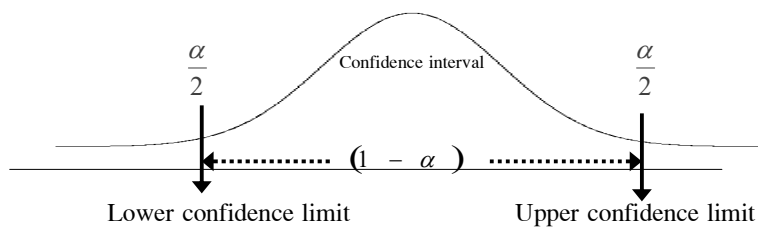
ความกว้าง แคม ของช่วงความเชื่อมั่น ขึ้นอยู่กับค่าอะไรบ้าง ?

ความผิดพลาดของช่วงความเชื่อมั่น alpha (α)

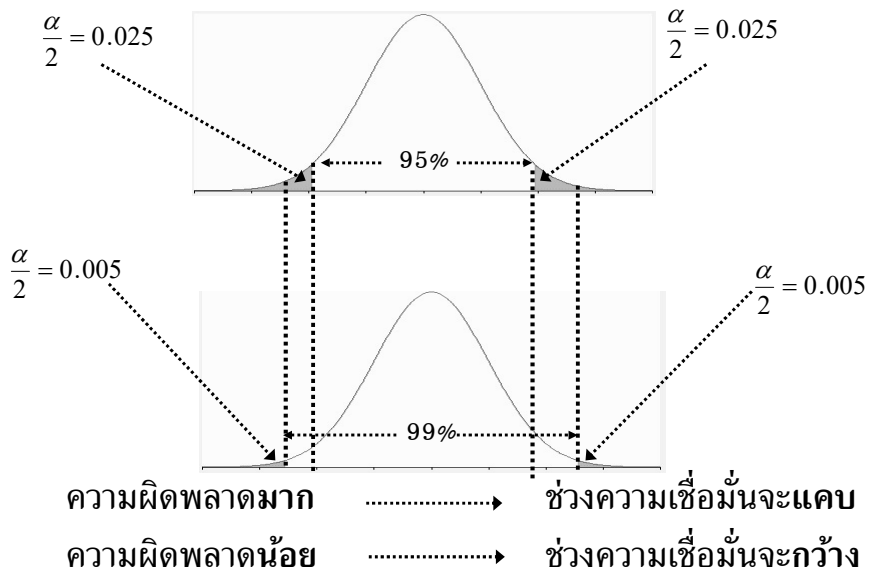
Alpha **น้อย** **มาก**
 ↓ ↓
ช่วงเชื่อมั่น กว้าง แคบ

ความผิดพลาดในการประมาณควรจะผิดพลาดข้างมากหรือข้างน้อย ?

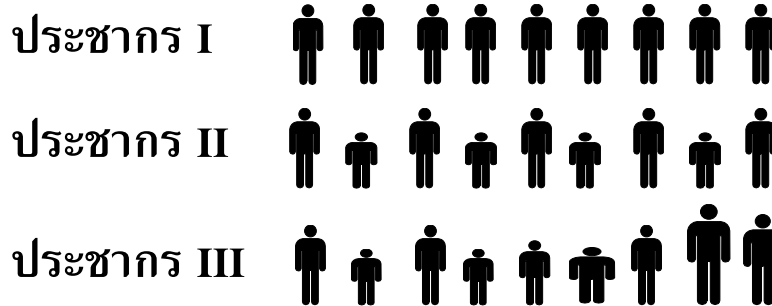
ถ้าให้ α = ความผิดพลาด



ถ้าให้ความผิดพลาดในการประมาณ = α



ความแปรปรวน



ความแปรปรวนน้อย> ช่วงความเชื่อมั่นจะแคบ

ความแปรปรวนมาก> ช่วงเชื่อความมั่นจะกว้าง

ขนาดตัวอย่าง

```
. cii 5 120 5
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]
	5	120	2.236068	113.7917 126.2083

$$se = \frac{5}{\sqrt{5}}$$

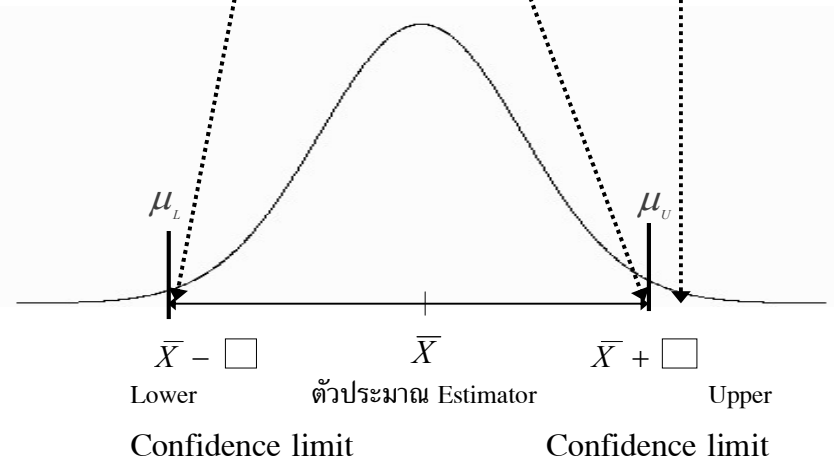
```
. cii 100 120 5
```

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]
	100	120	.5	119.0079 120.9921

$$se = \frac{5}{\sqrt{100}}$$

เมื่อ ความแปรปรวนและค่าเฉลี่ยเท่ากัน
การเปลี่ยนแปลงค่า n มีผลต่อ standard error

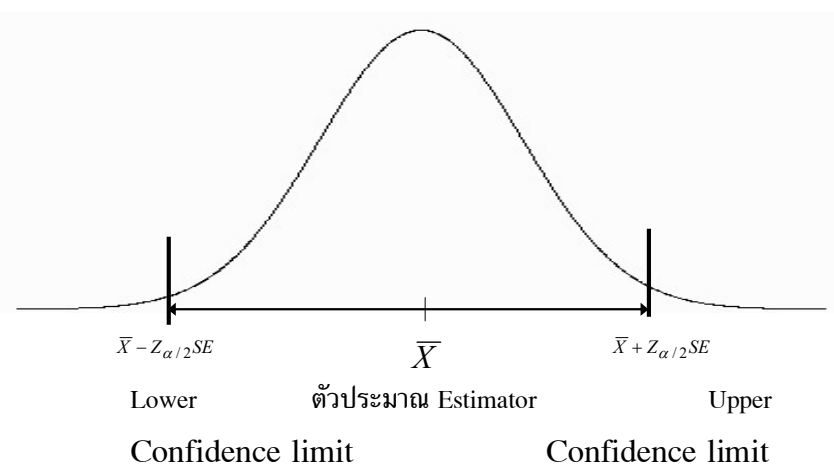
การประมาณค่าแบบช่วงความเชื่อมั่น



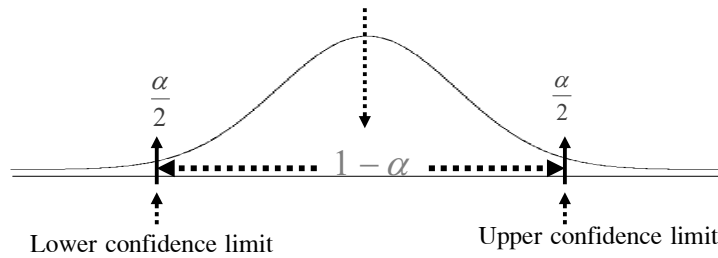
การประมาณค่าแบบช่วงความเชื่อมั่น

$$\bar{X} \pm Z_{\alpha/2} [\sigma / \sqrt{n}]$$

ตัวประมาณ สัมประสิทธิ์ความเชื่อมั่น x ค่าคาดเคลื่อน



ช่วงความเชื่อมั่น (confidence interval)



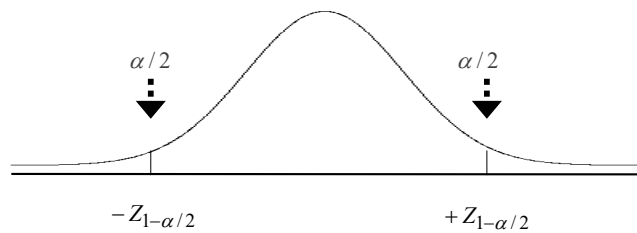
Estimator \pm Reliability Coefficient x Standard Error

การประมาณค่าแบบช่วงความเชื่อมั่น: ที่มา

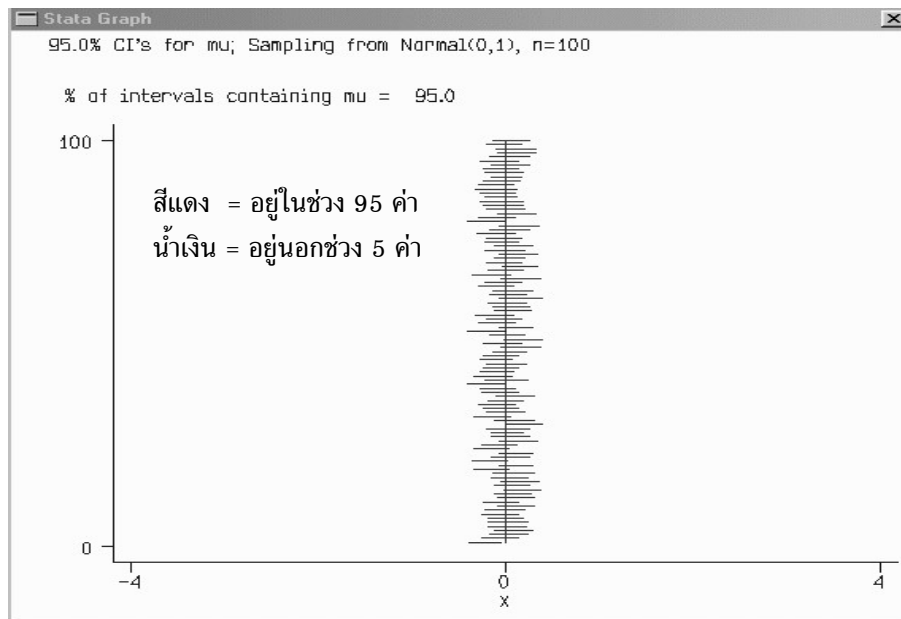
$$-Z_{\alpha/2} < Z < +Z_{\alpha/2}$$

$$-Z_{\alpha/2} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < +Z_{\alpha/2}$$

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



ความหมายของช่วงความเชื่อมั่น



ความหมายของช่วงความเชื่อมั่น

-ช่วงความเชื่อมั่นที่ $(1-\alpha)100\%$

เมื่อมีการสุ่มตัวอย่าง n จากประชากรเดียวกัน 100 ครั้ง

สร้างช่วงเชื่อมั่นได้ 100 ช่วง โอกาสที่จะพบว่าช่วงเชื่อมั่นอยู่ในช่วงที่กำหนด $(1-\alpha)100$ ช่วง (ไม่อยู่ในช่วงที่กำหนด

$(\alpha)100$ ช่วง)

-แต่ความเป็นจริงมีการศึกษา 1 ครั้ง ดังนั้น อธิบายได้ว่า

“(1- α)100% ช่วงความเชื่อมั่นจะมีค่าพารามิเตอร์อยู่ในช่วง”

เช่น “ด้วยความเชื่อมั่น 95% ค่าเฉลี่ยของค่า systolic BP

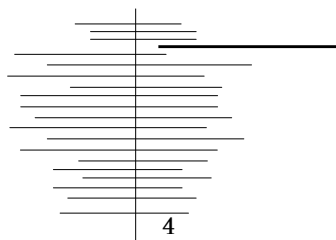
จะมีค่าอยู่ในช่วง 100.37 ถึง 139.63 mmHg”

ตัวอย่าง มีต้นส้ม 20 ต้น ต้องการทราบว่ามีส้มหล่นเท่าไร ?

- สุ่มต้นส้ม 1 ต้น วางตะกร้าใต้ต้นส้ม 5 ใบ
- 5, 3, 8, 2, 2
- คำนวณค่าเฉลี่ย $(5+3+8+2+2)/5 = 4$

```
. su or
Variable |      Obs      Mean  Std. Dev.   Min     Max
-----|-----
or |          5         4    2.54951     2     8
```

```
. ci or
Variable |      Obs      Mean  Std. Err.   [95% Conf. Interval]
-----|-----
or |          5         4    1.140175    .8343655    7.165634
```



95%CI ของค่าเฉลี่ย ของส้มที่หล่น
 หมายความว่า จากต้นส้ม 20 ต้น มี
 มีต้นส้ม 19 ต้นมีช่วงความเชื่อมั่น
 ครอบคลุมค่าเฉลี่ยที่เป็นจริง
 ต้นส้ม 1 ต้น มีช่วงความเชื่อมั่น
 ไม่ครอบคลุมค่าเฉลี่ยที่เป็นจริง

การประมาณช่วงความเชื่อมั่นค่าเฉลี่ยของประชากรเดียว
 กรณีทราบค่าความแปรปรวนของประชากร

-เมื่อทราบค่า σ จะใช้การแจกแจง Z

$$\bar{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

ตัวอย่าง นักวิจัยต้องการประมาณช่วงเชื่อมั่นค่าเฉลี่ยที่ระดับ
 ช่วงความเชื่อมั่นที่ 95% ของ Systolic Blood Pressure
 ข้อมูลประกอบด้วย 100 110 120 130 140
 (ถ้าทราบความแปรปรวนของประชากร = 225)

- ขั้นตอนที่ 1 หาดั้วประมาณ (ค่าเฉลี่ย = 120)
- ขั้นตอนที่ 2 หา SE = $\sigma/\sqrt{n} = 15/\sqrt{5}$
- ขั้นตอนที่ 3 กำหนดความผิดพลาด 0.05 หาสัมประสิทธิ์
ความเชื่อมั่น $Z_{1-\alpha/2} = Z_{1-0.05/2} = Z_{0.975} = 1.96$

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \dots \rightarrow 120 \pm 1.96 \frac{15}{\sqrt{5}} \dots \rightarrow 106.85, 133.15$$

ช่วงความเชื่อมั่นที่ 95% ของค่าเฉลี่ย Systolic Blood Pressure มีค่าระหว่าง 106.85 ถึง 133.15 mmHg

การประมาณช่วงความเชื่อมั่นค่าเฉลี่ยของประชากรเดียว
กรณีไม่ทราบค่าความแปรปรวนของประชากร

-เมื่อไม่ทราบค่า จะใช้การแจกแจง t

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \dots \rightarrow t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

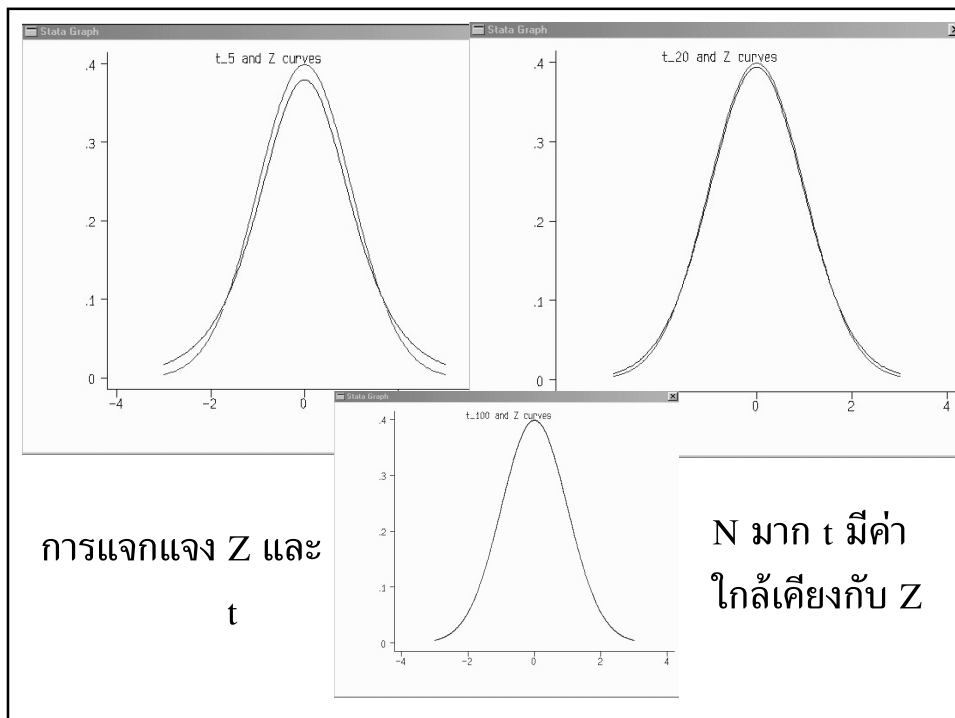
$$\bar{X} - t_{\frac{\alpha}{2}, df} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}, df} \cdot \frac{s}{\sqrt{n}}$$

เมื่อไม่ทราบความแปรปรวน

- ใช้ค่า S แทน σ
- ใช้การแจกแจง t แทนการแจกแจงแบบ Z

$(1-\alpha)100\%$ ช่วงเชื่อมั่นได้แก่

$$\bar{x} \pm t_{\frac{\alpha}{2}, df} \frac{s}{\sqrt{n}}$$



การแจกแจงแบบ T

คล้ายกับการแจกแจงแบบ Z การอ่านผลจากตาราง ต้องพิจารณาจากค่า **DF=Degree of Freedom** (ชั้นของการเป็นอิสระ)

ด้วย

Critical Values of Student's t						
df	.10	.05	.025	.01	.005	.0005
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
...						
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
35	1.306	1.690	2.030	2.438	2.724	3.591
...						
	1.282	1.645	1.960	2.326	2.576	3.291

ตัวอย่าง นักวิจัยต้องการประมาณช่วงความเชื่อมั่นค่าเฉลี่ยที่ระดับ 95% ของ Systolic Blood Pressure

ข้อมูลประกอบด้วย 100 110 120 130 140

-ขั้นตอนที่ 1 หาตัวประมาณ (ค่าเฉลี่ย = 120)

-ขั้นตอนที่ 2 หา SE = $s/\sqrt{n} = 15.81/\sqrt{5}$

-ขั้นตอนที่ 3 กำหนดความผิดพลาด 0.05 สามารถหา

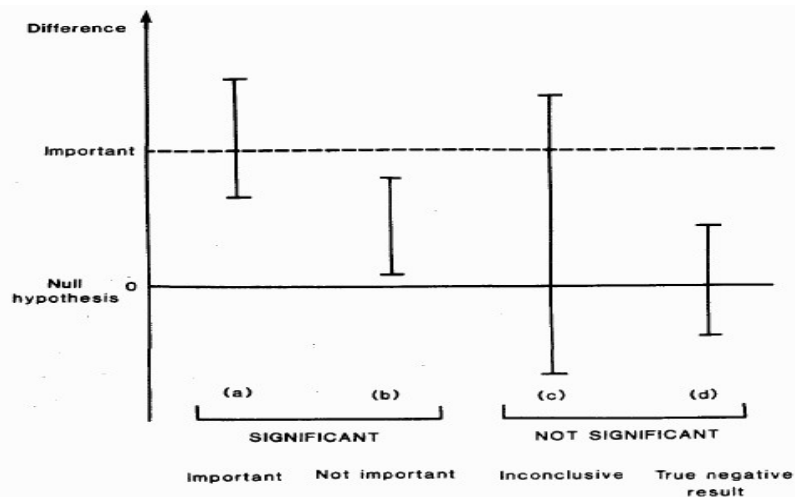
สัมประสิทธิ์ความเชื่อมั่น $t_{\alpha/2, n-1} = t_{0.025, 4} = 2.776$

```
. di invttail(4,.025)
2.7764451
```

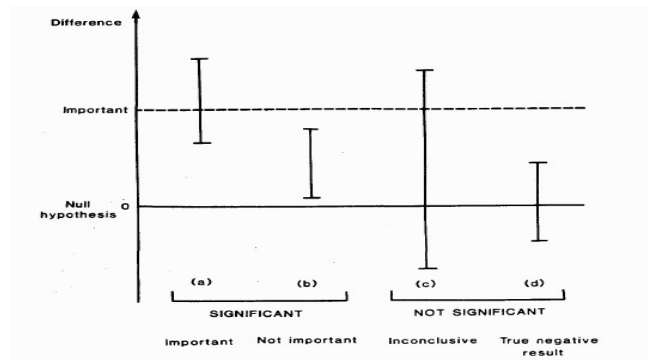
(1- α)100% ช่วงความเชื่อมั่นได้แก่

$$\bar{x} \pm t_{\frac{\alpha}{2}, df} \frac{s}{\sqrt{n}}$$
$$120 \pm 2.776 \frac{15.81}{\sqrt{5}}$$

95% ช่วงความเชื่อมั่นของค่าเฉลี่ย Systolic Blood Pressure
มีค่า 100.38 ถึง 139.63 mmHg

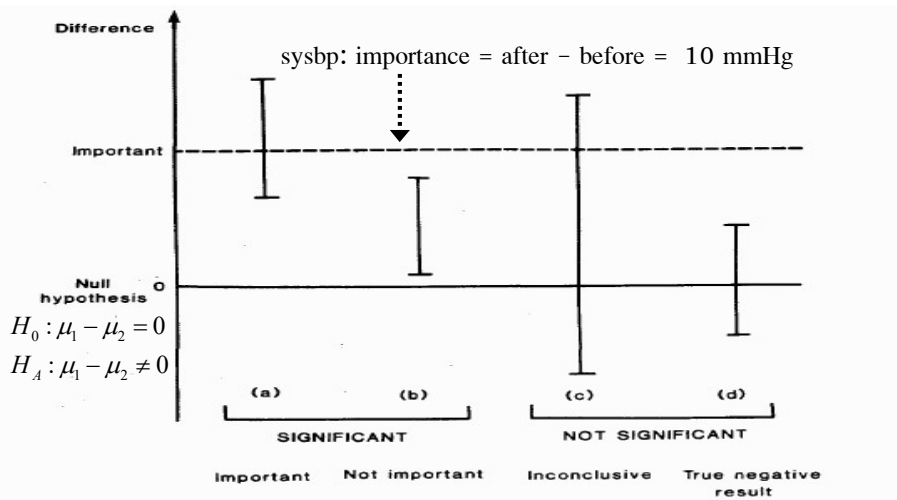


ช่วงความเชื่อมั่น 4 รูปแบบ สรุปรวมเกี่ยวกับ statistical significant และ practical importance (clinical significant)



- (a) the difference is significant and large enough to be of practical importance
- (b) the difference is significant but too small to be of practical importance
- (c) the difference is not significant but may be large enough to be of practical importance
- (d) the difference is not significant and also not large enough to be of practical importance

Berry, G. Statistical Significance and confidence intervals. The Medical Journal of Australia. 144(9); 1986.



ช่วงเชื่อมั่น 4 รูปแบบ สรุปลักษณะเกี่ยวกับ statistical significant และ practical importance (clinical significant)

ตัวอย่าง เป้าหมายการรักษาผู้ป่วยเบาหวาน ในการควบคุมความดันโลหิต
Systolic ให้ต่ำกว่า 130 mmHg ในการรักษาผู้ป่วยด้วยยาลด
ความดันโลหิต พบว่า ข้อมูลก่อนและหลัง มีดังนี้

```

sysbp   post sysbp
136     136
151     150
131     130
130     129
133     130
159     150
165     150
135     130
135     130
140     135

```

```

. ttest sysbp= psysbp

```

Paired t test						
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
sysbp	10	141.5	3.916489	12.38503	132.6403	150.3597
psysbp	10	137	2.928784	9.261629	130.3746	143.6254
diff	10	4.5	1.454877	4.600725	1.20884	7.79116

```

mean(diff) = mean(sysbp - psysbp)          t = 3.0930
Ho: mean(diff) = 0                          degrees of freedom = 9
Ha: mean(diff) < 0                          Ha: mean(diff) != 0
Pr(T < t) = 0.9936                          Pr(|T| > |t|) = 0.0129
                                           Ha: mean(diff) > 0
                                           Pr(T > t) = 0.0064

```

ตัวอย่าง เปรียบเทียบยาลดความดันโลหิต Systolic ของยา 2 ชนิด
ยา A และยา B ในการรักษาผู้ป่วยโรค Hypertension
ข้อมูลและการวิเคราะห์ มีดังนี้

(ยามีความสำคัญทางปฏิบัติหรือคลินิก เมื่อยามีค่าแตกต่างกันเกิน 5 mmHg
ผลข้างเคียง ราคา ยา ฯลฯ Systolic BP หลังได้รับยาไม่เกิน 150 mmHg)

id	sbp_be~e	sbp_af~r	tx	diff
1	180	130	1	50
2	185	135	1	50
3	179	128	1	51
4	181	132	1	49
5	184	134	1	50
6	180	135	2	45
7	185	137	2	48
8	179	130	2	49
9	181	133	2	48
10	184	135	2	49

```

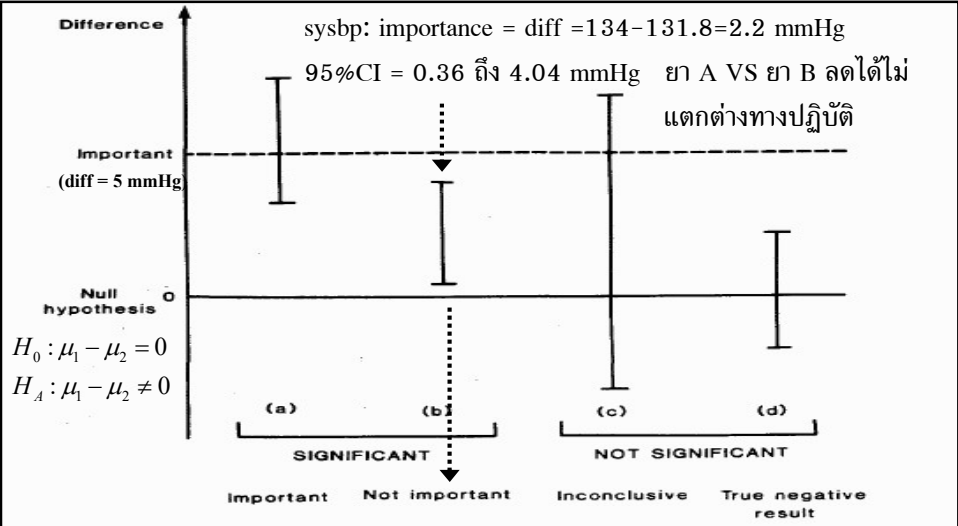
. ttest diff,by(tx)
Two-sample t test with equal variances
-----
Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]
-----+-----
1 | 5 50 .3162278 .7071068 49.12201 50.87799
2 | 5 47.8 .7348469 1.643168 45.75974 49.84026
-----+-----
combined | 10 48.9 .5259911 1.66333 47.71013 50.08987
-----+-----
diff | 2.2 .8 .3551967 4.044803
-----+-----
diff = mean(1) - mean(2) t = 2.7500
Ho: diff = 0 degrees of freedom = 8

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.9875 Pr(|T| > |t|) = 0.0251 Pr(T > t) = 0.0125

. su sbp_after if tx==1
Variable | Obs Mean Std. Dev. Min Max
-----+-----
sbp_after | 5 131.8 2.863564 128 135

. su sbp_after if tx==2
Variable | Obs Mean Std. Dev. Min Max
-----+-----
sbp_after | 5 134 2.645751 130 137

```



ช่วงเชื่อมั่น 4 รูปแบบ สรุปลักษณะเกี่ยวกับ statistical significant และ practical importance (clinical significant)

