การแจกแจงความน่าจะเป็น

(Probability Distribution)

ผศ. นิคม ถนอมเสียง
ภาควิชาชีวสถิติและประชากรศาสตร์
คณะสาธารณสุขศาสตร์ ม.ขอนแก่น
Email: nikom@kku.ac.th

After studying this chapter,

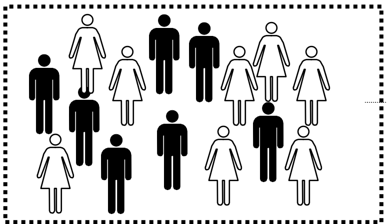
1. understand selected discrete distributions and how to use them to calculate probabilities in real-world problems.
2. understand selected continuous distributions and how to use them to calculate probabilities in real-world problems.
3. be able to explain the similarities and differences between distributions of the discrete type and the continuous type and when the use of each is appropriate.

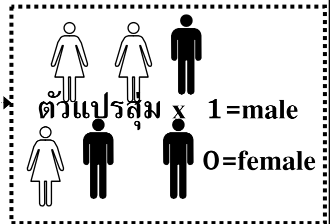
การแจกแจงความน่าจะเป็น

-การแสดงความน่าจะเป็นของค่าที่เป็นไปได้ทุกค่า
ของตัวแปรสุ่ม

-แสดงในรูป ตาราง, กราฟ, ฟังก์ชันทางคณิตศาสตร์

ตัวแปรที่สนใจของหน่วยสังเกตุ เช่นเพศของบุคคล
เมื่อ หน่วยสังเกตถูกสุ่ม ตัวแปรที่สนใจเรียกว่า “**ตัวแปรสุ่ม**”





การแสดงความน่าจะเป็นเช่น

สมมุติ มีประชากร 2 คน เป็นโรคฟันผุ 1 คน D+
ไม่ผุ 1 คน D-
-สุ่มแล้วใส่คืน 2 ครั้ง

โอกาสที่สุ่มแต่ละครั้ง ได้คนฟันผุ =1/2 ฟันไม่ผุ =1/2

เมื่อสุ่ม 2 ครั้ง โอกาสพบฟันผุ, ไม่ผุ ได้แก่

D+D+, D+D-, D-D+, D-D-

กฎความน่าจะเป็น
Pr(A and B) = Pr(A) x Pr(B)
Pr(A or B) = Pr(A) + Pr(B)

ความน่าจะเป็นในการเกิดเหตุการณ์

ตัวแปรสุ่มที่สนใจ คือ ฟันผุ (X)

ถ้าให้การสุ่มได้ผู้ป่วยฟันผุทั้งสองคน =2 (X=2)

หนึ่งคน=1 (X=1) ศูนย์คน=0 (X=0)

ความน่าจะเป็น

P(X=2) = P(D+D+) $= \frac{1}{2} x \frac{1}{2} = \frac{1}{4}$

กฎความน่าจะเป็น

Pr(A and B) = Pr(A) x Pr(B)

Pr(A or B) = Pr(A) + Pr(B)

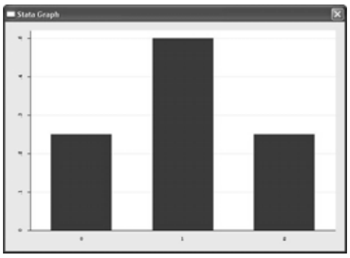
P(X=1) = P(D+D-)+P(D-D+) $= \left(\frac{1}{2} x \frac{1}{2}\right) + \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{2}{4}$

P(X=0) = P(D-D-) $= \frac{1}{2} x \frac{1}{2} = \frac{1}{4}$

แสดงในรูปตาราง

X	ความน่าจะเป็น
0	1/4
1	2/4
2	1/4

กราฟ



ฟังก์ชันทางคณิตศาสตร์

$$f(x) = \begin{cases} \frac{x+1}{4}; & x = 0,1 \\ \frac{x-1}{4}; & x = 2 \end{cases}$$

การแจกแจงความน่าจะเป็น

1. การแจกแจงความน่าจะเป็นไม่ต่อเนื่อง

ได้แก่ การแจกแจงแบบทวินาม (binomial distribution)

การแจกแจงแบบปัวซงค์ (Poisson distribution)

การแจกแจงแบบ Multinomial distribution

ฯลฯ

2. การแจกแจงความน่าจะเป็นต่อเนื่อง

เช่น การแจกแจงแบบปกติ (Normal distribution)

การแจกแจงแบบที (t distribution)

การแจกแจงแบบเอฟ (F distribution)

การแจกแจงแบบไคร์สแควร์ (Chi-Square distribution)

ฯลฯ

ประเภทการแจกแจงความน่าจะเป็นการเกิดเหตุการณ์

1. การแจกแจงความน่าจะเป็นตัวแปรไม่ต่อเนื่อง
   เช่น ป่วย ไม่ป่วย, หาย ไม่หาย

2. การแจกแจงความน่าจะเป็นตัวแปรต่อเนื่อง
   เช่น อายุ น้ำหนัก ความดันโลหิต sysBP

การแจกแจงความน่าจะเป็นตัวแปรไม่ต่อเนื่อง

คุณสมบัติ

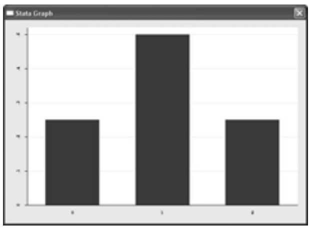
1. f(x) = P(X=x)

   ค่าฟังก์ชันของตัวแปรสุ่ม x มีค่าเท่ากับ

   ความน่าจะเป็นที่ X มีค่าเท่ากับ x

2. $f(x) \geq 0$ ทุกค่าของ x

3. $\sum_{all\,x} f(x) = \sum_{all\,x} P(X=x) = 1$



การแจกแจงความน่าจะเป็นตัวแปรไม่ต่อเนื่อง

-การแจกแจงแบบทวินาม

$$f(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}; x = 0,1,2,\ldots,n$$

-การแจกแจงแบบปัวซงค์

$$f(x) = P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!};\ x = 0,1,2,\ldots$$

การแจกแจงความน่าจะเป็นตัวแปรไม่ต่อเนื่อง

1. f(x) = P(X=x) ค่าฟังก์ชันของตัวแปรสุ่ม x มีค่าเท่ากับ ความน่าจะเป็นที่ X มีค่าเท่ากับ x

ตัวอย่าง อุบัติการณ์ของฟันผุในเด็กมีค่าเท่ากับ 10% บริษัทยาสีฟันต้องการทดสอบยาสีฟันว่าได้ผลในการป้องกันฟันผุ โดยให้ทดลองใช้ยาสีฟันกับเด็กกลุ่มหนึ่ง 20 คนหลังจากนั้น 1 ปี พบเด็กฟันผุ 1 คน บริษัทยาสีฟันจะโฆษณาว่ายาสีฟันได้ผลดีหรือไม่

-การเกิดฟันผุ มีการแจกแจงแบบทวินาม

$$f(x) = P(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x}; x = 0,1,2,...,n$$

-การเกิดฟันผุ มีการแจกแจงแบบทวินาม

$$f(x) = P(X \le 1) = P(x = 0) + P(x = 1)$$

$$= \binom{20}{0}(.10)^{0}(1-.10)^{20-0} + \binom{20}{1}(.10)^{1}(1-.10)^{20-1}$$

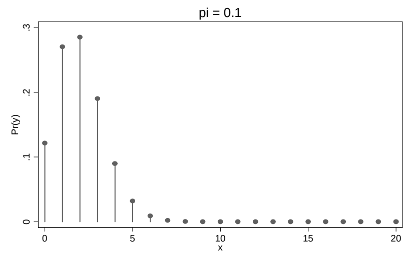
$$= .90^{20} + \binom{20}{1}(.10)(.90)^{12}$$

$$= .122 + .270 = .392$$

การเกิดฟันผุ หลังใช้ยาสีฟัน 1 ปีมีความน่าจะเป็นเท่ากับ .39 หรือ 39% (มากกว่าการเกิดฟันผุ ตามปกติ)

X	f(x)=P(X=x)
0	.12157
1	.27017
2	.28518
3	.19012
4	.089779
5	.031921
6	.008867
7	.0019705
8	.0003558
9	.0000527
10	6.44e-06
11	6.51e-07
12	5.42e-08
13	3.71e-09
14	2.06e-10
15	9.15e-12
16	3.18e-13
17	8.22e-15
18	2.22e-16
19	0
20	0
Total	**1**

1. f(x) = P(X=x)
   ค่าฟังก์ชันของตัวแปรสุ่ม x มีค่าเท่ากับ ความน่าจะเป็นที่ X มีค่าเท่ากับ x
2. $f(x) \geq 0$ ทุกค่าของ x
3. $\sum_{all\,x} f(x) = \sum_{all\,x} P(X=x) = 1$



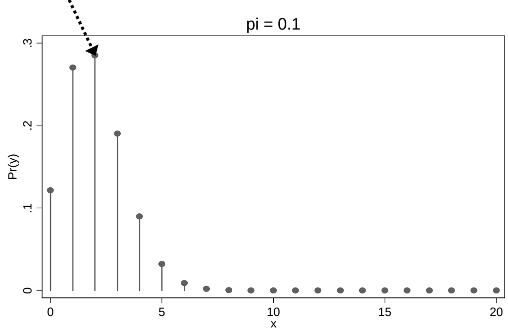
X	f(x)=P(X=x)
0	.12157
1	.27017
2	.28518
3	.19012
4	.089779
5	.031921
6	.008867
7	.0019705
8	.0003558
9	.0000527
10	6.44e-06
11	6.51e-07
12	5.42e-08
13	3.71e-09
14	2.06e-10
15	9.15e-12
16	3.18e-13
17	8.22e-15
18	2.22e-16
19	0
20	0
Total	**1**

```
. di binomial(20,0,.1)
.12157665

. di  binomial(20,1,.1)-binomial(20,0,.1)
.27017034

. di  binomial(20,2,.1)-binomial(20,1,.1)
.28517981

. di  binomial(20,3,.1)-binomial(20,2,.1)
.19011987
```



วิเคราะห์โดย STATA

- ความน่าจะเป็นที่พบฟันผุ อย่างน้อย 1 คน

$$f(x) = P(X \le 1) = P(x = 0) + P(x = 1)$$

$$= \binom{20}{0}(.10)^0(1-.10)^{20-0} + \binom{20}{1}(.10)^1(1-.10)^{20-1}$$

$$= .90^{20} + \binom{20}{1}(.10)(.90)^{12}$$

$$= .122 + .270 = .392$$

```
. bitesti 20 1 0.10
        N   Observed k   Expected k   Assumed p   Observed p
-------------------------------------------------------------
       20          1            2       0.10000      0.05000
  Pr(k >= 1)           = 0.878423  (one-sided test)
  Pr(k <= 1)           = 0.391747  (one-sided test)
  Pr(k <= 1 or k >= 3) = 0.714820  (two-sided test)
```

การแจกแจงความน่าจะเป็นตัวแปรไม่ต่อเนื่อง

การแจกแจงทวินาม (binomial)

1. การเกิดเหตุการณ์เป็นอิสระต่อกัน
2. การเกิดเหตุการณ์ ให้ผลลัพธ์ อย่างใดอย่างหนึ่ง
   สำเร็จหรือไม่สำเร็จ
3. ความน่าจะเป็นของความสำเร็จมีค่าคงที่เท่ากับ p
   ความน่าจะเป็นของความไม่สำเร็จมีค่าคงที่เท่ากับ 1- p

ฟังก์ชันการแจกแจงทวินาม (binomial)

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \; ; x = 0,1,2,\ldots n$$

เขียนฟังก์ชัน ด้วยสัญลักษณ์ X ~ b(n,p) หรือ X ~ b(x; n, p)

ลักษณะการแจกแจงทวินาม (binomial)

1. มีพารามิเตอร์ 2 ตัวได้แก่ N, π
2. ค่าเฉลี่ย = $N\pi$
3. ความแปรปรวน = $N\pi(1-\pi)$
4. ส่วนเบี่ยงเบนมาตรฐาน = $\sqrt{N\pi(1-\pi)}$
5. ลักษณะการแจกแจงขึ้นอยู่กับ N, π

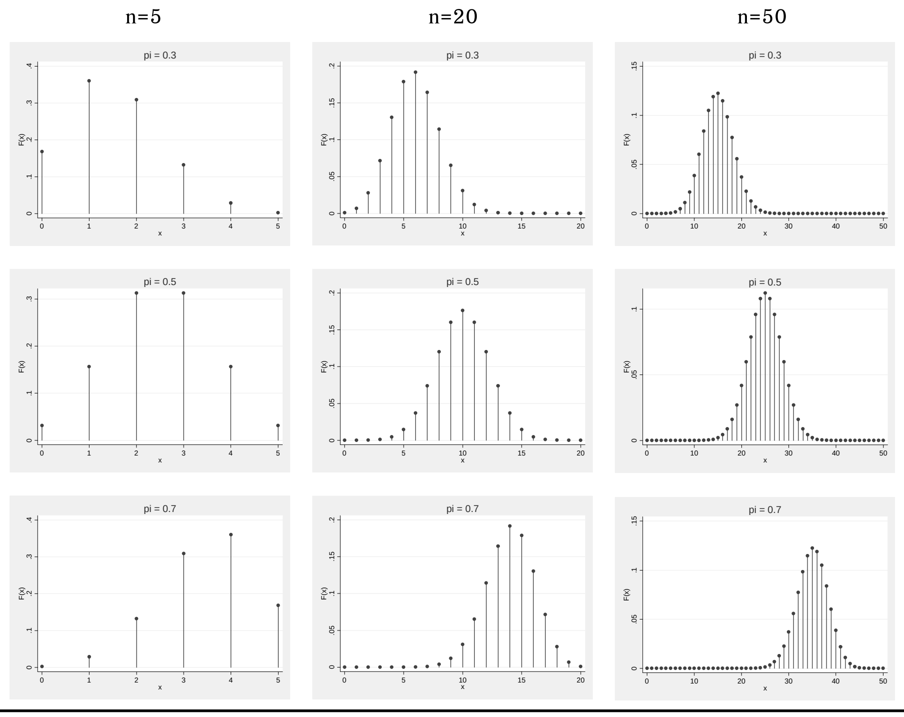
เมื่อ n น้อย ลักษณะการแจกแจงที่พบ

π น้อย การแจกแจง เบ้ทางบวก

π = .5 การแจกแจงสมมาตร

π มาก การแจกแจง เบ้ทางลบ

เมื่อ N มาก ลักษณะการแจกแจงแบบสมมาตร



การแจกแจงปัวซอง (Poisson Distribution)

-การแจกแจงปัวซอง ตั้งชื่อตามนักคณิตศาสตร์ชาวฝรั่งเศส ชื่อ Simeon Denis Poisson

ถ้าให้ x เป็นจำนวนเหตุการณ์ที่เกิดขึ้นในช่วงเวลาใดเวลาหนึ่ง ความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้นได้แก่

$$f(x) = \frac{e^{-\lambda}\lambda^{n}}{x!} ; x = 0,1,2,\ldots ; e = 2.7182$$

λ = ค่าเฉลี่ยของจำนวนครั้งของเหตุการณ์ที่เกิดขึ้นในช่วงเวลา

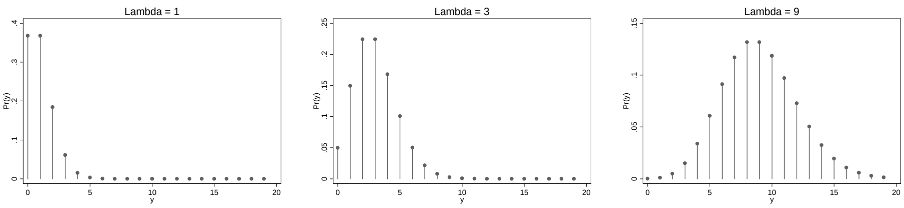
เขียนสัญลักษณ์ X ~ p(λ) หรือ X ~ P(x; λ)

ลักษณะของการแจกแจงแบบปัวซอง

1. เหตุการณ์ที่เกิดขึ้นเป็นอิสระต่อกัน โดยที่เหตุการณ์ที่เกิดขึ้นในช่วงใดช่วงหนึ่งหรือเวลาใดเวลาหนึ่ง จะไม่มีผลต่อความน่าจะเป็นของการเกิดเหตุการณ์ในช่วงอื่น ๆ หรือเวลาอื่น ๆ
2. ในช่วงใดช่วงหนึ่งมีจำนวนเหตุการณ์ที่เกิดขึ้นอย่างไม่จำกัด
3. ความน่าจะเป็นของการเกิดเหตุการณ์ในช่วงใด ๆ เป็นสัดส่วนกับความยาวของช่วงทั้งหมด
4. ในช่วงเวลาสั้น ความน่าจะเป็นของการเกิดเหตุการณ์จะมีค่าน้อย

ลักษณะของการแจกแจงแบบปัวซอง

1. มีพารามิเตอร์ 1 ตัวได้แก่ λ
2. ค่าเฉลี่ย = ความแปรปรวน = λ
3. ส่วนเบี่ยงเบนมาตรฐาน = $\sqrt{\lambda}$



ตัวอย่าง การเกิดอุบัติเหตุถนนมิตรภาพ ช่วงระหว่างประตูเข้ามหาวิทยาลัยขอนแก่น กับโรงพยาบาลศรีนครินทร์ โดยรถจักรยานยนต์ในช่วง 1 สัปดาห์ พบว่าเกิดอุบัติเหตุ 2 ครั้ง

1. ให้หาค่าความน่าจะเป็นของการไม่เกิดอุบัติเหตุในช่วง 1 สัปดาห์

วิธีทำ

จากโจทย์มีค่า $\lambda = 2$

1. ให้หาค่าความน่าจะเป็นของการไม่เกิดอุบัติเหตุในช่วง 1 สัปดาห์

$$p(x) = \frac{e^{-\lambda}\lambda^{n}}{x!}$$

$$p(0) = \frac{2^{0}e^{-2}}{0!} = .1353$$

```
.display 1-gammap(0+1,2) หรือ
.display poissonp(2,0)
 .13533528
```

2. ให้หาค่าความน่าจะเป็นการเกิดอุบัติเหตุ 3 ครั้งในช่วง 2 สัปดาห์

ช่วง 2 อาทิตย์มีการเกิดอุบัติเหตุเท่ากับ 2(2) = 4

ดังนั้นความน่าจะเป็นเท่ากับ $p(x \leq 3) = p(0) + p(1) + p(2) + p(3)$

$$p(0) = \frac{4^0 e^{-4}}{0!} = .0183\,;\; p(1) = \frac{4^1 e^{-4}}{1!} = .0733$$

$$p(2) = \frac{4^2 e^{-4}}{2!} = .1465;\; p(3) = \frac{4^3 e^{-4}}{3!} = .1954$$

ดังนั้น $p(x \leq 3) = p(0) + p(1) + p(2) + p(3)$

$= .0183 + .0733 + .1465 + .1954$

$= .4335$

```
.display 1-gammap(4,4) หรือ
.display poisson(4,3)
.43347012
```

การแจกแจงความน่าจะเป็นตัวแปรต่อเนื่อง

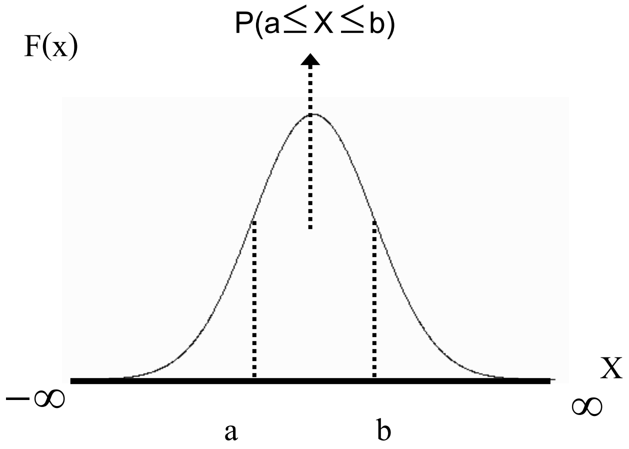
คุณสมบัติ

1. $f(x) \geq 0$ ทุกค่าของ x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ พื้นที่ใต้โค้ง f(x) ทั้งหมดคือ ความน่าจะเป็นของทุกค่าของ x มีค่าเท่ากับ 1
3. $P(X=a) = 0$ เมื่อ a=ค่าคงที่

   ความน่าจะเป็นของ X ที่มีค่าเท่ากับค่าคงที่ มีค่าเท่ากับ 0

$$P(a \leq X \leq b) = P(a < X < b)$$

$$P(a \leq X < b) = P(a \leq X < b)$$

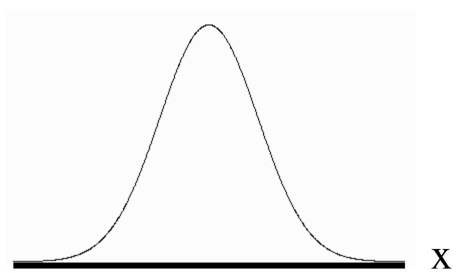


พื้นที่ใต้โค้ง f(x) ทั้งหมดคือความน่าจะเป็นของทุกค่าของ x มีค่าเท่ากับ 1 ($-\infty$ ถึง ∞)

การแจกแจงความน่าจะเป็นของตัวแปรต่อเนื่อง

-การแจกแจงแบบปกติ (Normal Distribution)

- คิดโดย Abraham De Moivre
- Karl Friedrick Gauss (1774-1855) เผยแพร่และใช้อย่างกว้างขวาง เรียกชื่อ Gaussian



การแจกแจงความน่าจะเป็นของตัวแปรต่อเนื่อง

ฟังก์ชันการแจกแจงแบบปกติ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

π และ e เป็นค่าคงที่มีค่า 3.1459 และ 2.71828

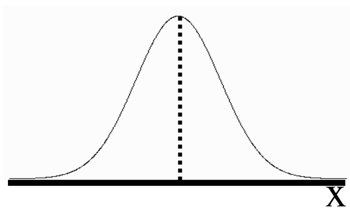
μ ค่าเฉลี่ย

σ ส่วนเบี่ยงเบนมาตรฐาน

เขียนสัญญลักษณ์ $X \sim n(\mu, \sigma^2)$

คุณสมบัติการแจกแจงแบบปกติ

- รูปโค้งระฆังคว่ำด้านซ้ายและขวาของ μ มีความเท่ากัน (mirror image) (สมมาตรรอบค่าเฉลี่ย)
- ปลายทั้งสองข้างของโค้งค่อย ๆ ลาดลงสู่แกน x จรดแกน x ที่อนันท์
- มีจุดเปลี่ยนเว้าที่ $\mu \pm \sigma$
- ค่าเฉลี่ย มัธยฐาน ฐานนิยม มีค่าเท่ากัน เมื่อลากเส้นจากยอดโค้งตั้งฉากกับแกน x
- พื้นที่ทั้งหมดใต้โค้งเหนือแกน x หรือค่าความน่าเป็นของพื้นที่เท่ากับ 1

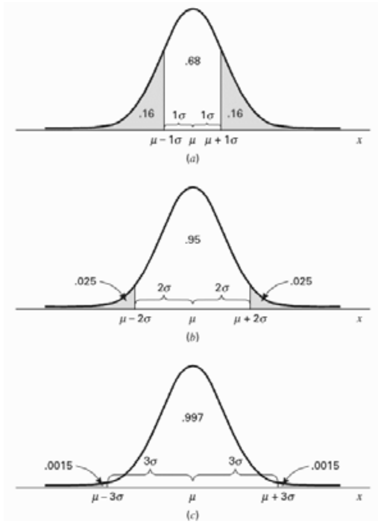


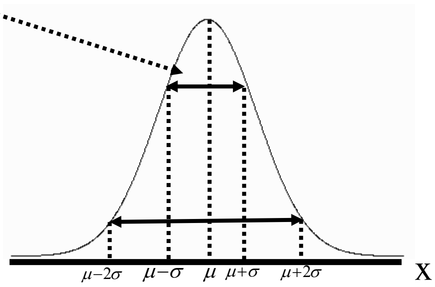
-เมื่อแบ่งโค้งโดยลากเส้นตั้งฉากจากยอดถึงแกน x
ระยะห่างจากค่าเฉลี่ยทั้งสองข้างเป็นดังนี้

1 หน่วย SD พื้นที่ใต้โค้งเท่ากับ 68.26%

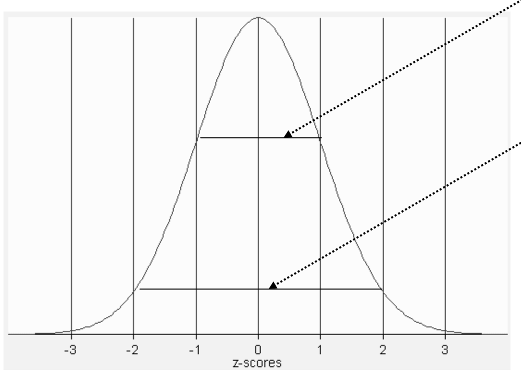
2 หน่วย SD พื้นที่ใต้โค้งเท่ากับ 95.45%

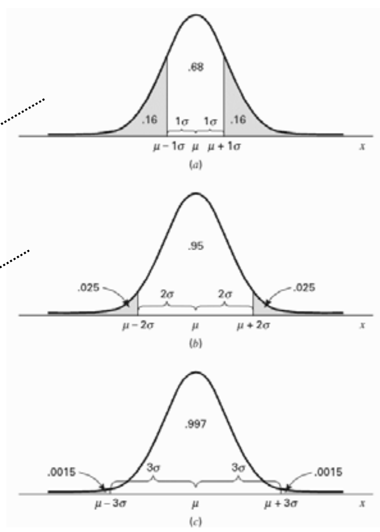
3 หน่วย SD พื้นที่ใต้โค้งเท่ากับ 99.73%





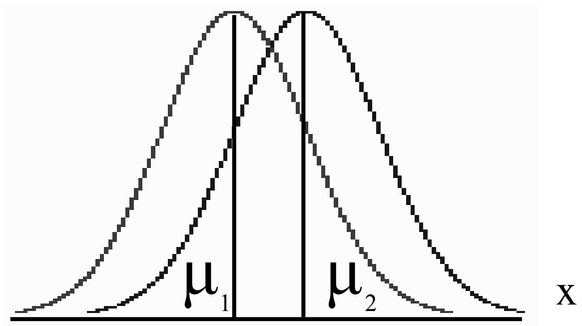
.zdemo -1 -2 -3 0 +1 +2 +3

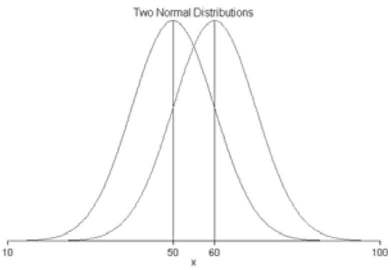




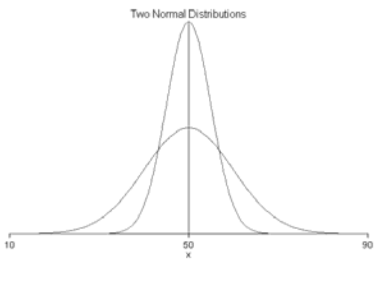
-ความแตกต่างของการแจกแจงแบบปกติ ขึ้นอยู่กับ ค่า μ และ σ

เมื่อค่า μ แตกต่างกัน ตำแหน่งบนแกน x แตกต่าง

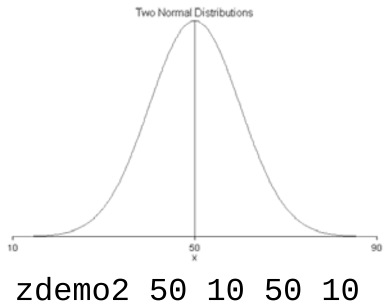




zdemo2 50 10 60 10

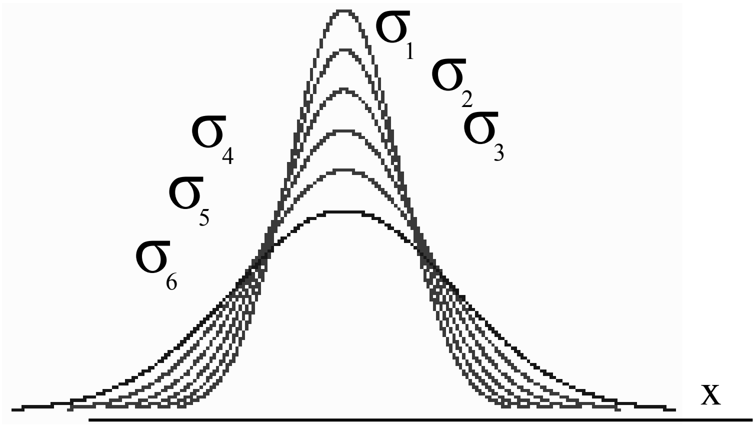


zdemo2 50 10 50 5



zdemo2 50 10 50 10

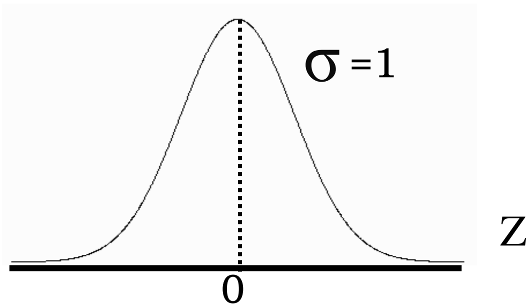
ถ้าค่า σ ต่างกัน ความแบนราบและความโด่งแตกต่าง



การแจกแจงปกติมาตรฐาน

Standard Normal Distribution

- เป็นการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1
- บางครั้งเรียก unit normal distribution

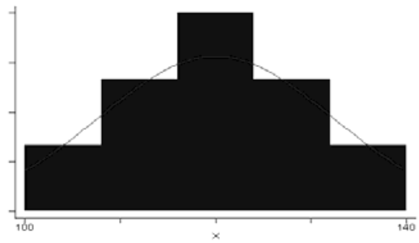


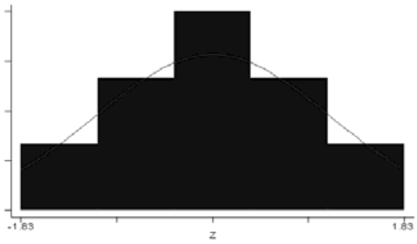
Z~N(0,1)

การแจกแจงปกติมาตรฐาน แปลงได้จาก ค่าตัวแปรสุ่มดังนี้

$$z = \frac{x_i - \mu}{\sigma}$$

X	Z
50	-1.55
60	-0.77
70	0.00
80	0.77
90	1.55
70.00	0.00 ⇦ μ
12.91	1 ⇦ σ





ฟังก์ชันการแจกแจงปกติมาตรฐาน Z~N(0,1)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

π และ e เป็นค่าคงที่มีค่า 3.1459 และ 2.71828

Z ค่ามาตรฐาน

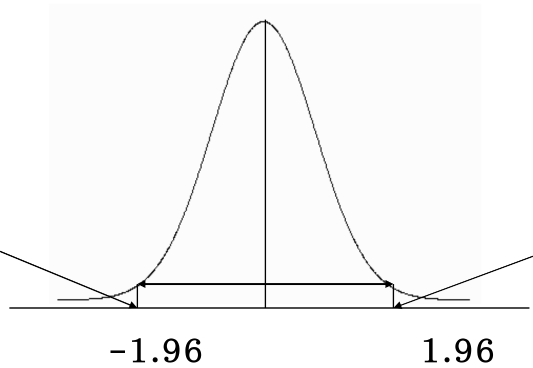
ใช้คำนวณความน่าจะเป็นใต้โค้งปกติ

ค่า **Z** มีความสำคัญ เพราะเมื่อทราบค่า Z

สามารถหา**<u>ค่าความน่าจะเป็น</u>**ได้

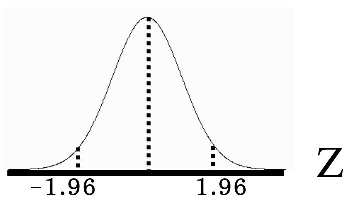
เช่น Z = 1.96 มีค่าความน่าจะเป็นเท่ากับ 0.025

Z = -1.96 มีค่าความน่าจะเป็นเท่ากับ 0.025



การเปิดหาค่าความน่าจะเป็นจากตาราง

Normal Distribution



(A)	(B)	(C)
Z	Area Between Mean & Z	Area Beyond Z
1.9500	0.4744	0.0256
1.9600	0.4750	0.0250
1.9700	0.4756	0.0244

การใช้โปรแกรม STATA หาค่าความน่าจะเป็น

Normal Distribution

. display normprob(1.96) หรือ

. di normal(1.96)

```
di normal(0)
.5
```

.9750021

. display 1-normprob(1.96) หรือ

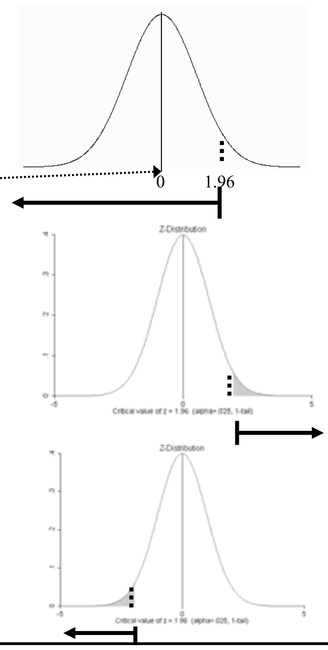
. di 1-normal(1.96)

.0249979

. display normprob(-1.96) หรือ

. di normal(-1.96)

.0249979



ตัวอย่าง ค่า CHOL ประชากร ~N(0,1)

มี μ = 200 mg/100ml σ = 20 mg/100ml

จงหาความน่าจะเป็นที่คน คนหนึ่งคนจากประชากร

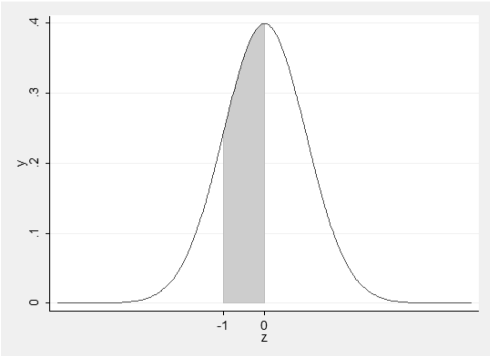
มีค่า CHOL

1. x ระหว่าง 180 ถึง 200
2. x>225
3. x<150

1. ระหว่าง 180 ถึง 200

P(180<X<200) = P[(180-200)/20]<Z< [(200-200)/20]

=P(-1<Z<0) = 0.3414



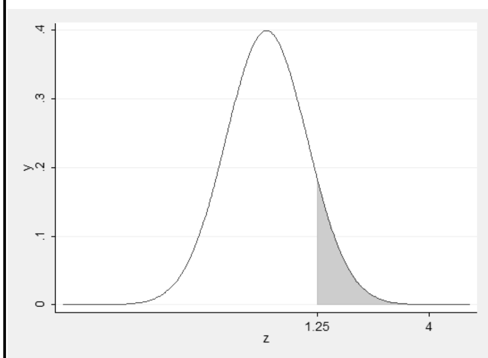

```
. di normal((200-200)/2)
 .5
. di  normal((180-200)/20)
 .15865525
. di  .5-.15865525
 .34134475
```
นำมาลบกัน

```
. di normal((200-200)/20)-normal((180-200)/20)
 .34134475
```

2. x> 225 mg/100ml

P(X>225) = P[Z> [(225-200)/20]

=P(Z>1.25) = 0.1056

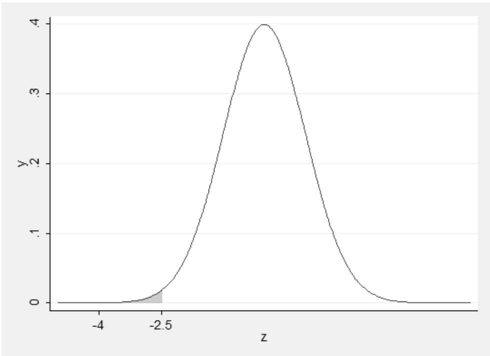



```
. di 1-normal((225-200)/20)
.10564977
```

3. X< 150 mg/100ml

$P(X<150) = P[Z< [(150-200)/20]$

$=P(Z<-2.5) = 0.0062$



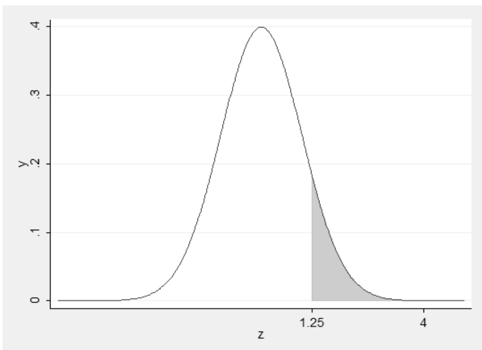

```
. di normal((150-200)/20)
.00620967
```

การหาค่า Z เมื่อทราบค่า Probability Value

1. $P(Z>Z_1) = 0.1056$
2. $P(Z<Z_1) = 0.025$
3. $P(Z>Z_1) = 0.05$

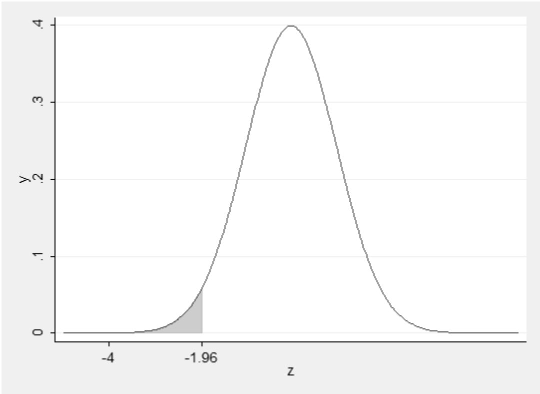
1. $P(Z>Z_1) = 0.1056$

```
. di invnormal(1-.1056)
1.2502726
```



2. $P(Z<Z_1) = 0.025$

```
. di invnormal(.025)
-1.959964
```



3. $P(Z>Z_1) = 0.05$

```
. di invnormal(1-0.05)
1.6448536
```