

สถิติสำหรับงานตรวจพิสูจน์ดีเอ็นเอ

ทางนิติเวชศาสตร์



สุคนธ์ ประตุงกาญจนา
11 กันยายน 2557

ความรู้พื้นฐานทางสถิติ



ความน่าจะเป็น (Probability)

โยนลูกเต๋า 1 ลูก โอกาสที่
จะทอดออกมาได้ 1 หรือ

$P(1) = ?$

1/6

โยนลูกเต๋าที่ละ 2 ลูก พร้อม
กัน ให้ได้ หน้า 1 ทั้งสองลูก
มี โอกาสทำได้ $1/36$



โยนลูกเต๋าที่ละ 2 ลูก พร้อมกัน
ให้ได้ หน้า 1 และหน้า 6
มี โอกาสทำได้



$1/18$

โยนลูกเต๋าที่ละ 2 ลูก พร้อมกัน ให้
ลูกเต๋าลูกใหญ่ได้หน้า 1 และลูกเล็ก
ได้หน้า 6 มี โอกาสทำได้

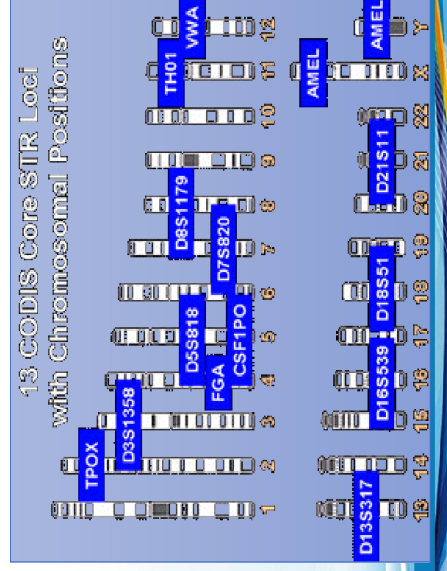


$1/36$

Microsatellite DNA

หากตรวจพิสูจน์ 15 ตำแหน่ง ผลที่ได้จะเหมือน

การโยนลูกเต๋าที่ละ 2 ลูก 15 ครั้ง



MtDNA

การตรวจ mtDNA ที่ region ที่ตาม ผลที่ได้จะเหมือนการโยน

ลูกเต๋าที่ละหน้า 1 ลูก เพียงครั้งเดียวเท่านั้น



Prior Probability

เครื่องบินตก มีคนตาย 100 คน



Prior Probability = ถ้า
ไม่ทำอะไรเลย แล้วมี
คนมาขอรับศพ 1 คน
โอกาสที่จะปล่อยศพได้
ถูกต้องมีเท่าไร ?

Prior Probability = 1/100

Posterior Probability

Posterior Probability = โอกาสที่จะปล่อยศพได้ถูกต้อง
หลังจากใช้ผลตรวจ DNA มาพิจารณาด้วย

$$= LR/[LR + (1 / \text{Prior Prob}) - 1]$$

* LR = likelihood ratio

Likelihood Ratio (LR)

- **Likelihood Ratio = สัดส่วนความน่าจะเป็นระหว่าง ใช่/ไม่ใช่**
- กรณีของการพิสูจน์บุคคล ถ้าทราบเลือดและผู้ต้องสงสัยมีรูปแบบดีเอ็นเอตรงกัน แล้วคำนวณ Likelihood Ratio ได้ = 1,000 แสดงว่า โอกาสที่ทราบเลือดนี้จะเป็นของผู้ต้องสงสัยรายนี้ เท่ากับ 1,000 เท่า เมื่อเทียบกับโอกาสที่ทราบเลือดนี้จะเป็นของบุคคลอื่นในกลุ่มประชากรเดียวกัน
- Likelihood Ratio = 1 / Match probability

Match probability

Match probability โอกาสที่คนสองคนที่ไม่ได้มี
ความสัมพันธ์เป็นญาติกัน จะมีรูปแบบดีเอ็นเอ
เหมือนกันโดยบังเอิญ

Match prob = ความถี่ของ DNA marker ใน
ฐานข้อมูล

ค่าความถี่ของอัลลีลต่าง ๆ ที่ตำแหน่ง D13S317 ในประชากรไทย

locus	allele	freq
D13S317	10	0.1227
D13S317	11	0.2304
D13S317	12	0.1416
D13S317	13	0.1316
D13S317	14	0.0068
D13S317	18	0.0005
D13S317	19	0.0005
D13S317	7	0.0011
D13S317	8	0.3219
D13S317	9	0.1405

Random Matching probability (RMP)

- RMP = โอกาสที่ตัวอย่างตรวจจะเป็นของบุคคลอื่นโดยบังเอิญ มีค่าเท่ากับ 1 ใน
- RMP = 1 / profile frequency
- ใช้อธิบายกรณีการตรวจเอกลักษณ์บุคคลเท่านั้น ไม่ใช่อธิบายการตรวจความสัมพันธ์ทางสายเลือด



เกณท์ตัดสินใจ



■ LR > 99	เชื่อว่า.....ใช่
■ Post Prob > 99%	
■ LR >1-99	ไม่แน่ใจ...น่าจะใช่
■ Post Prob 50-99%	
■ LR 0.01-1	ไม่แน่ใจ...น่าจะไม่ใช่
■ Post Prob 0.01-50%	
■ LR <0.01	ไม่ใช่
■ Post Prob <0.01%	

Statistic calculation for mtDNA typing

1. Haplotype frequency calculation

1.1 direct counting method

$$p = \frac{x}{n}$$

p = haplotype frequency
 x = matches in the database
 n = database size

In cases where a **mtDNA** profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

$$p = X/N$$

7 matches in 27,773

$$p = 7/27,773 = 0.000252 = 0.025\%$$

Statistic calculation for mtDNA typing

1. Haplotype frequency calculation

1.2 modified direct counting method

$$p = \frac{x+1}{n+1}$$

p = haplotype frequency
 x = matches in the database
 n = database size

In cases where a **mtDNA** profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

$$p = \frac{x+1}{n+1}$$

7 matches in 27,773

$$p = 8/27,774 = 0.000288 = 0.029\%$$

Statistic calculation for mtDNA typing

1. Haplotype frequency calculation

1.3 “pseudo-count” Estimate:

David Balding, Weight-of-Evidence for Forensic DNA Profiles, Wiley, 2005

$$p = \frac{x+2}{n+2}$$

p = haplotype frequency
 x = matches in the database
 n = database size

In cases where a **mtDNA** profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

$$p = \frac{x+2}{n+2}$$

7 matches in 27,773

$$p = 9/27,775 = 0.000324 = 0.032\%$$

Statistic calculation for mtDNA typing

2. Confidence interval estimation

2.1 normal distribution : Holland & Parsons (1999) Forensic Sci Rev

$$p + 1.96 \sqrt{\frac{(p)(1-p)}{n}}$$

p = haplotype frequency
 n = database size

$$p = X/N$$

7 matches in 27,773

$$p = 7/27,773 = 0.000252 = 0.025\%$$

An upper bound confidence interval can be placed on the profile's frequency using:

$$p + 1.96 \sqrt{\frac{(p)(1-p)}{N}}$$

$$\begin{aligned} &0.000252 + 1.96 \sqrt{\frac{(0.000252)(1-0.000252)}{27,773}} \\ &= 0.000252 + 0.000187 = 0.000439 \\ &= \mathbf{0.044\% (\sim 1 \text{ in } 2270)} \end{aligned}$$

Recommended by SWGDAM (2003)

Statistic calculation for mtDNA typing

2. Confidence interval estimation

2.1 exact binomial CI : Clopper and Pearson (1934)

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} = 0.05 \quad (x > 0)$$

p_0 = haplotype frequency
 n = database size
 X = matches in the database

$$p_0 = 1 - 0.05^{1/n} \quad (x = 0)$$

$$1 - \alpha^{1/N}$$

0 matches in 4,004

Recommended by SWGDAM (2013)

where α is the confidence coefficient (0.05 for a 95% confidence interval) and N is the number of individuals in the database.

$$1 - \alpha^{1/N} = 1 - (0.05)^{1/(4,004)} = 0.000748 = \mathbf{0.075\% (\sim 1 \text{ in } 1340)}$$

Statistic calculation for mtDNA typing

2. Confidence interval estimation

2.1 exact binomial CI : Clopper and Pearson (1934)

Input Values

Sample size:

Number positive:

Confidence level:

Decimal places in answer:

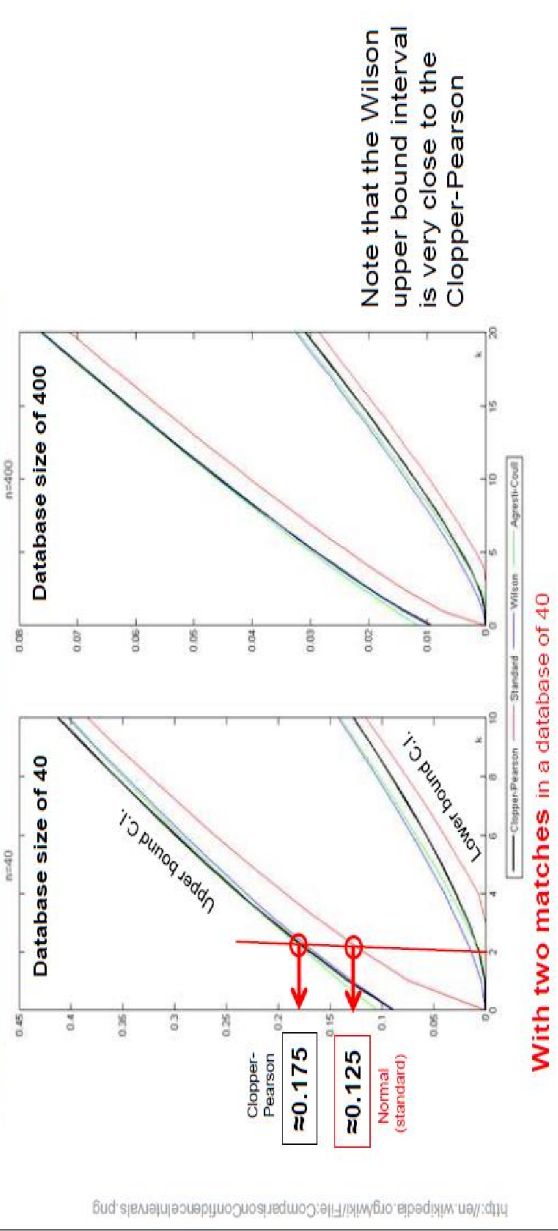
Results

Using 5 different methods

	Asymptotic (Wald)	Exact Binomial (Clopper-Pearson)	Wilson (Score)	Agresti-Coull (adjusted Wald)	Jeffreys
Number positive	7	7	7	7	7
Sample size	27773	27773	27773	27773	27773
Prevalence estimate	0.000252	0.000252	0.000252	0.000301	0.000252
Lower 90% confidence limit	9.5e-05	0.000118	0.000137	0.00013	0.000131
Upper 90% confidence limit	0.000409	0.000473	0.000465	0.000472	0.00045

Comparison of Clopper-Pearson to Normal (Standard) Confidence Intervals (C.I.)

Higher value is more conservative (favors the defendant)



Statistic calculation for mtDNA typing

3. Likelihood ratio calculation

$$LR = 1/\text{freq}$$

- Likelihood Ratio = สัดส่วนความน่าจะเป็นระหว่าง ใช้/ไม่ใช้
- กรณีของการพิสูจน์บุคคล ถ้าทราบเลือดและผู้ต้องสงสัยมีรูปแบบดีเอ็นเอตรงกัน แล้วคำนวณ Likelihood Ratio ได้ = 1,000 แสดงว่า โอกาสที่ทราบเลือดนี้จะเป็นของผู้ต้องสงสัยรายนี้ เท่ากับ 1,000 เท่า เมื่อเทียบกับโอกาสที่ทราบเลือดนี้จะเป็นของบุคคลอื่นในกลุ่มประชากรเดียวกัน
- Likelihood Ratio = 1 /Match probability

ตัวอย่าง กรณีตรวจความสัมพันธ์ที่สามารถตา

ลำดับที่	F003M-1	F002M-2
1	16167T	16129A
2	16223T	16162G
3	16274A	16172C
4	16320T	16304C
5	16362C	16396G
6	73G	16519C
7	94A	73C
8	263G	152C
9	309.1C	249D
10	315.1C	263C
11	469C	306.1C
12		315.1C
13		338T
14		523D
15		524D

Likelihood ratio	0	Probability	0.00000000%
จากการเปรียบเทียบลำดับดีเอ็นเอจากพันธุกรรมบริเวณ control region บนสายดีเอ็นเอและเคียวของภูมิภาคบนสายดีเอ็นเอที่ทราบแล้ว (CRS: NC012920; Andrews et al.,1990) พบว่า			
**** นาย วิชิตา สากุลวารี ใช้นี้ชื่อ(ผู้ต้องสงสัย)สามารถหาความสัมพันธ์ได้ กับ นางสาว ศรีลา ชาติสิน โดยมีค่าความน่าจะเป็นเท่ากับ 2 ล้านครั้งขึ้นไป			

ตัวอย่าง กรณีตรวจหาความสัมพันธ์ยัติสายมารดา

ลำดับที่	F003M-1	F003M-2
1	16108T	16108T
2	16129A	16129A
3	16162G	16162C
4	16172C	16172C
5	16304C	16304C
6	16519C	16519C
7	73G	73G
8	200G	200G
9	249D	249D
10	315.1C	263G
11	523D	315.1C
12	524D	523D
13		524D

Likelihood ratio 0 **Probability** 0.00000000%

จากการเปรียบเทียบลำดับเบสของสายพันธุ์กรรมสืบทอด control region บนสายพันธุ์คอนโทรลของภูมิภาคตรวจกับสายพันธุ์กรรมอ้างอิง (CRS: NC012920; Andrews et al.,1990) พบว่า

***ไม่สามารถระบุได้เกี่ยวกับสายลม เป็นยัติกรรมหรือยัติสายมารดาเดียวกัน เช่น ที่ -น้อง กับ นายแสงแดด เนื่องจากตำแหน่งที่เหมือนกันไม่ได้เพียง 1 ตำแหน่ง ซึ่งอาจเกิดจากกลายพันธุ์ (mutation) ได้

ตัวอย่าง กรณีตรวจหาความสัมพันธ์ยัติสายมารดา

ลำดับที่	F001M-1	F001M-2
1	16167T	16167T
2	16223T	16223T
3	16274A	16274A
4	16302G	16302G
5	16320T	16320T
6	16362C	16362C
7	73G	73G
8	94A	94A
9	263G	263G
10	309.1C	309.1C
11	315.1C	315.1C
12	489C	489C

Likelihood ratio 315 **Probability** 99.68382446%

จากการเปรียบเทียบลำดับเบสของสายพันธุ์กรรมสืบทอด control region บนสายพันธุ์คอนโทรลของภูมิภาคตรวจกับสายพันธุ์กรรมอ้างอิง (CRS: NC012920; Andrews et al.,1990) พบว่า

1. พบ ประจักษ์ พัสสูจน์ ปรากฏชื่อจากการเป็นญาติร่วมตระกูลสายมารดาเดียวกัน เช่น พี่ - น้อง กับ นาง น้อย พัสสูจน์
2. ความสัมพันธ์ นาย ประจักษ์ พัสสูจน์ เป็นญาติร่วมตระกูลสายมารดาเดียวกัน กับนาง น้อย พัสสูจน์ พบที่ rsacc 99.68382446 หมายความว่าพบยัติกรรมหรือยัติสายมารดาเดียวกัน ต่อกับประชากร mtDNA population Database (FBI, USA)

Table 2 Overview of database query settings, search results and LR values provided by 31 participating laboratories.

Lab	Database, population	Data source	Match type	Matches (x/N)	Frequency calculation	LR
1	EMPOP, Europe	F	Pattern	8/2636*	x+1/N+1	1/freq = 293.0
2	EMPOP, all	F+L	Pattern	35/8410	x/N	1/freq = 240.3 ^b
3	EMPOP, Europe	F+L	Pattern	26/2667	x+2/N+2	1/freq = 95.3 ^c
4	EMPOP, Europe	F+L	Pattern	26/2667	x/N	1/freq = 102.6 ^c
5	EMPOP, Europe	F+L	Pattern	26/2667	x+2/N+2	1/freq = 95.3 ^c
6	EMPOP, all	F+L	Pattern	35/7973	x/N	1/freq = 227.8 ^b
7	EMPOP, all	F+L	Pattern	43/8358	x+1/N+1	1/freq = 194.5
8	EMPOP, all	F+L	Pattern	35/8410	x+1/N+1	1/freq = 233.6 ^b
9	EMPOP, Europe	F	Pattern	26/2636	x+2/N+2	1/freq = 94.2
10	EMPOP, all	F+L	Pattern	35/7033	x/N	1/freq = 189.6
11	EMPOP, Europe	F+L	Pattern	26/2667	x+2/N+2	1/freq = 102.6 ^c
12	EMPOP, all	F+L	Literal	4/7057	x+1/N+1	1/freq = 106.5 ^c
13	EMPOP, all	F+L	Pattern	35/8410	x/N	1/freq = 240.3 ^b
14	EMPOP, Europe	F+L	Pattern	26/2636	x+1/N+1	1/freq = 97.7 ^c
15	EMPOP, Spain	F+L	Pattern	0/31	x+1/N+1	1/freq = 32.0
16	EMPOP, Europe	F+L	Pattern	26/2636	x+1/N+1	1/freq = 97.7 ^c
17	EMPOP, S-Europe	F+L	Pattern	8/2657*	x/N	1/freq = 123.7 ^c
18	EMPOP, S-Europe	F+L	Pattern	6/742	x/N	1/freq = 333.4
19	EMPOP, all	F+L	Pattern	35/8410	x+2/N+2	1/freq = 227.3 ^b
20	EMPOP, Europe	F+L	Pattern	26/2667	x+2/N+2	1/freq = 95.3 ^c
21	EMPOP, all	F+L	Pattern	35/8410	x/N	1/freq = 240.3 ^b
22	EMPOP, all	F+L	Pattern	26/2667	x+1/N+1	1/freq = 68.0
23	EMPOP, Europe	F+L	Pattern	26/2667	x+1/N+1 and Wilson interval	1/freq = 102.6 ^c
24	EMPOP, Europe	F+L	Pattern	2/1040	(x/N)+1.96*sqrt(pq/N)	1/freq = 218.0 ^b
25	OWN (Chile)	na	Pattern	35/8410	x+1/N+1	e ⁸⁹ /freq = 233.6 ^b
26	EMPOP, all	F+L	Pattern	26/2667	x/N	1/freq = 102.6 ^c
27	EMPOP, Europe	F+L	Pattern	24/2667	x+1/N+1	1/freq = 106.7 ^c
28	EMPOP, Europe	F+L	Literal	24/2667	x+1/N+1	1/freq = 97.7 ^c
29	EMPOP, Europe	F	Pattern	26/2636	x+1/N+1	1/freq = 98.8 ^c
30	EMPOP, Europe	F+L	Pattern	26/2667	x+1/N+1	1/freq = 98.8 ^c
31	EMPOP, Spain	F+L	Pattern	0/31	x+1/N+1	1/freq = 32.0

Combining Statistic values

1. mtDNA and autosomal result
 combine $LR = LR^*_{mt} \times LR^*_{aSTR}$
2. mtDNA and YSTR
 combine $LR = LR^*_{mt} \times LR^*_{YSTR}$

*All LR should be calculated from the same reference population

LocName	F003M-1	F003M-2	PI	FS	HS	FC	II
1 D8S1179	15,17	13,14	0.0020	0.2500	0.5000	0.7500	0.0020
2 D21S11	29,30	29,30	2.0020	3.2286	1.5010	1.2505	1.5077
3 D7S820	8,13	8,12	1.4262	0.9631	1.2131	1.1066	0.5000
4 CSF1PO	11,12	11,12	1.4700	2.0392	1.2350	1.1175	1.2235
5 D3S1358	16,18	16,18	4.0517	6.9108	2.5258	1.7629	1.6745
6 TH01	5,7	7,7	1.4593	0.9797	1.2297	1.1148	1.2512
7 D13S317	8,10	8,10	2.7154	4.5877	1.8577	1.4288	1.6246
8 D16S529	9,11	11,12	0.8240	0.6620	0.9120	0.9560	0.5000
9 D2S1328	19,19	19,19	4.0707	6.5579	2.5533	1.7677	2.7442
10 D19S433	13,15,2	15,2,15,2	2.6692	1.5946	1.8346	1.4173	1.9468
11 VWA	14,19	16,19	2.3268	1.4124	1.6634	1.3317	0.5000
12 TFOX	8,9	8,11	0.4416	0.4708	0.7208	0.8604	0.5000
13 D18S51	14,15	16,18	0.0032	0.2500	0.5000	0.7500	0.0032
14 D5S818	10,12	10,12	2.1795	3.6913	1.5898	1.2949	1.6047
15 FGA	19,19	19,22	5.9970	3.2485	3.4985	2.2492	0.5000
16 Amelogenin	X,Y	X,X	1.0000	1.0000	1.0000	1.0000	1.0000
17 Blood Group	A (1a)	A (1a)	1.0000	1.0000	1.0000	1.0000	1.0000
CPI	0.05242699	PostProbPI	4.981531925				
CFS	675.62	PostProbFS	99.85220704			Theta	0.010
CHS	91.76	PostProbHS	98.92190452			Prior Prob.	0.500
CFC	19.52	PostProbFC	95.1273494				
CI	0.00001088	PostProbII	.00108839			% (Theta = 0.0; Prior Prob = 0.500)	
Mutation	2	(D18S51, D8S1179)					

