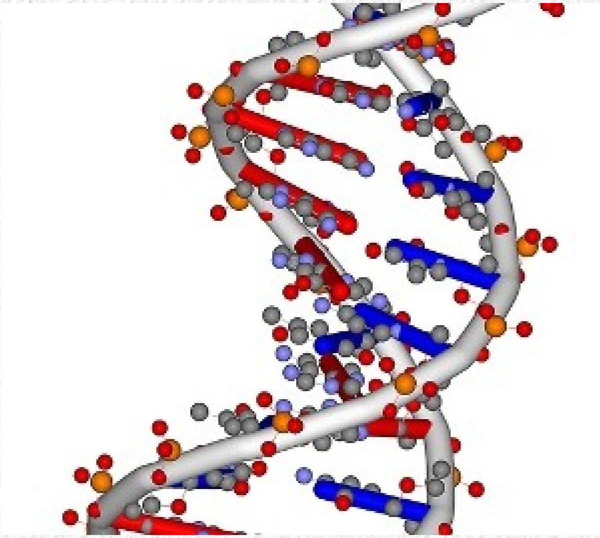
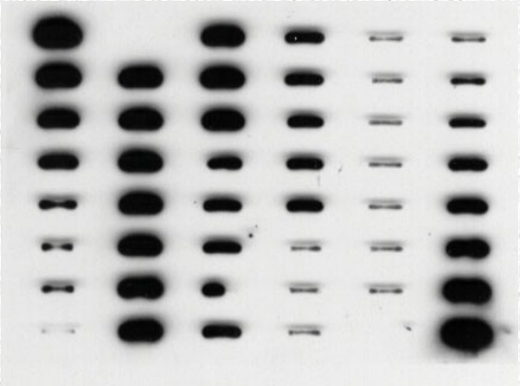
Likelihood





$$\mathrm{LR} = \frac{p(\mathrm{M}|\mathrm{P})}{p(\mathrm{M}|\mathrm{E})} = \frac{1}{p(\mathrm{M}|\mathrm{E})}$$

Likelihood

- Likelihood is a general approach to statistics that can be used for:
  - Estimating parameters, building confidence intervals
  - Testing hypotheses
  - Comparing hypotheses against one another
- Likelihood appears to be similar to probability, but has a very different interpretation

Definition of likelihood

- Likelihood is a measure of support for a particular estimated value (of a parameter, or from a model), given a set of data
- We need to assume a particular statistical distribution of deviations from the estimated value (such as normal, binomial, etc.) to use the approach
- We then use the formula for the assumed distribution to calculate the likelihood

Some nice features of likelihoods

- Likelihoods can be combined
  - Data can be added as it becomes available, such as adding observations until two treatment groups diverge
  - Big no-no with hypothesis testing
- No “sampling” distributions
  - Likelihoods of samples are just products of likelihoods of individual obserations
- Parameter estimates and confidence intervals
  - “Maximum likelihood estimates”
  - Even when analytical formulas aren't available

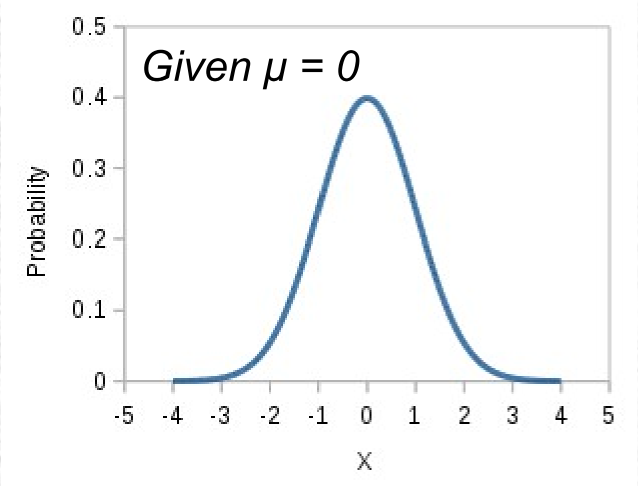
Likelihood and probability

- Probabilities treat parameters (μ, σ) as known, and calculate the chances of observing data values given the parameters → $p(x_i \mid \mu, \sigma)$
- Likelihoods invert this – they treat the data as known, and ask how likely a set of parameters is given the known data → $L(\mu, \sigma \mid x_i)$

Probability and likelihood – single observation

Probability

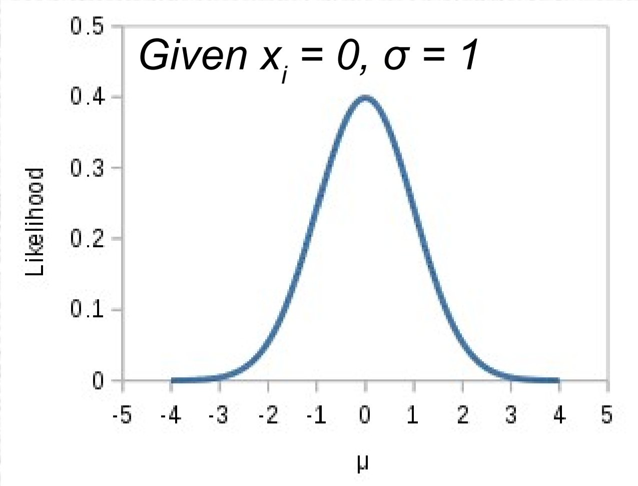
$p(x_i \mid \mu, \sigma)$



Use a probability distribution to represent a “random variable”

Likelihood

$L(\mu, \sigma \mid x_i)$

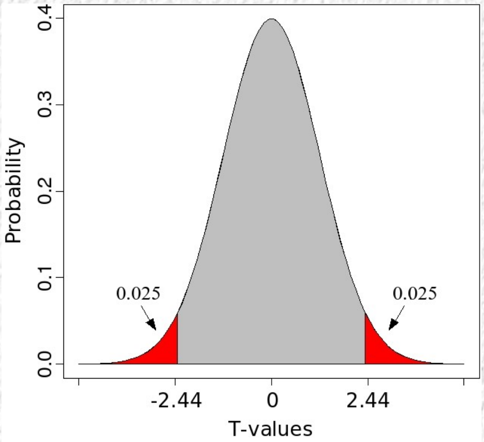


Likelihood of possible values of the mean given observed x

Probability and likelihood – a sample of data points

Probability

$$p(\bar{x} \mid \mu, s_{\bar{x}})$$



Use a "sampling distribution", such as the t

Likelihood

$$\prod L(\mu, \sigma \mid x_i)$$

Likelihood of a sample is the product of likelihoods of data points

Likelihood functions

- Derived from probability distributions
- Used to model differences between hypothetical values and observed data (residuals)
- Example: Normally distributed deviations

$$L\left(\mu \mid x_i\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{x_i - \mu}{\sigma}\right]^2}$$

Likelihood of parameters given a single data point - the normal probability distribution

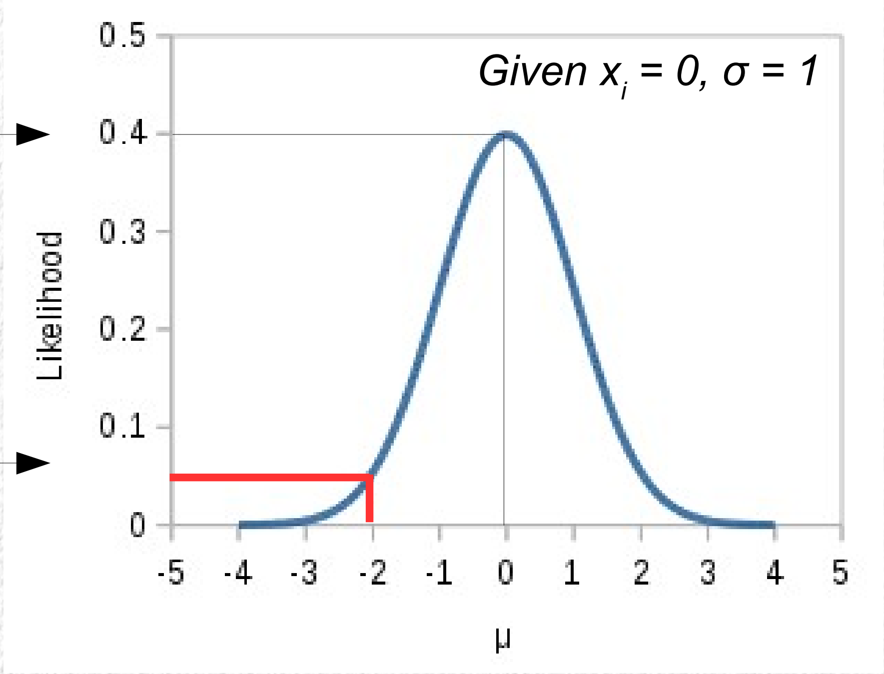
$$L\left(\mu \mid x_{i \ldots n}\right) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{x_i - \mu}{\sigma}\right]^2}$$

Likelihood of parameters given all the data – the product of all the likelihoods given each single data point

Example of likelihood function

Likelihood of $\mu = 0$

Likelihood of $\mu = -2$

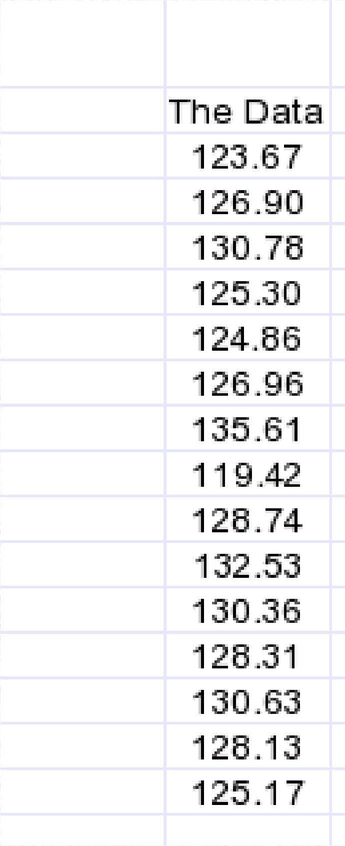


Given $x_i = 0$, $\sigma = 1$

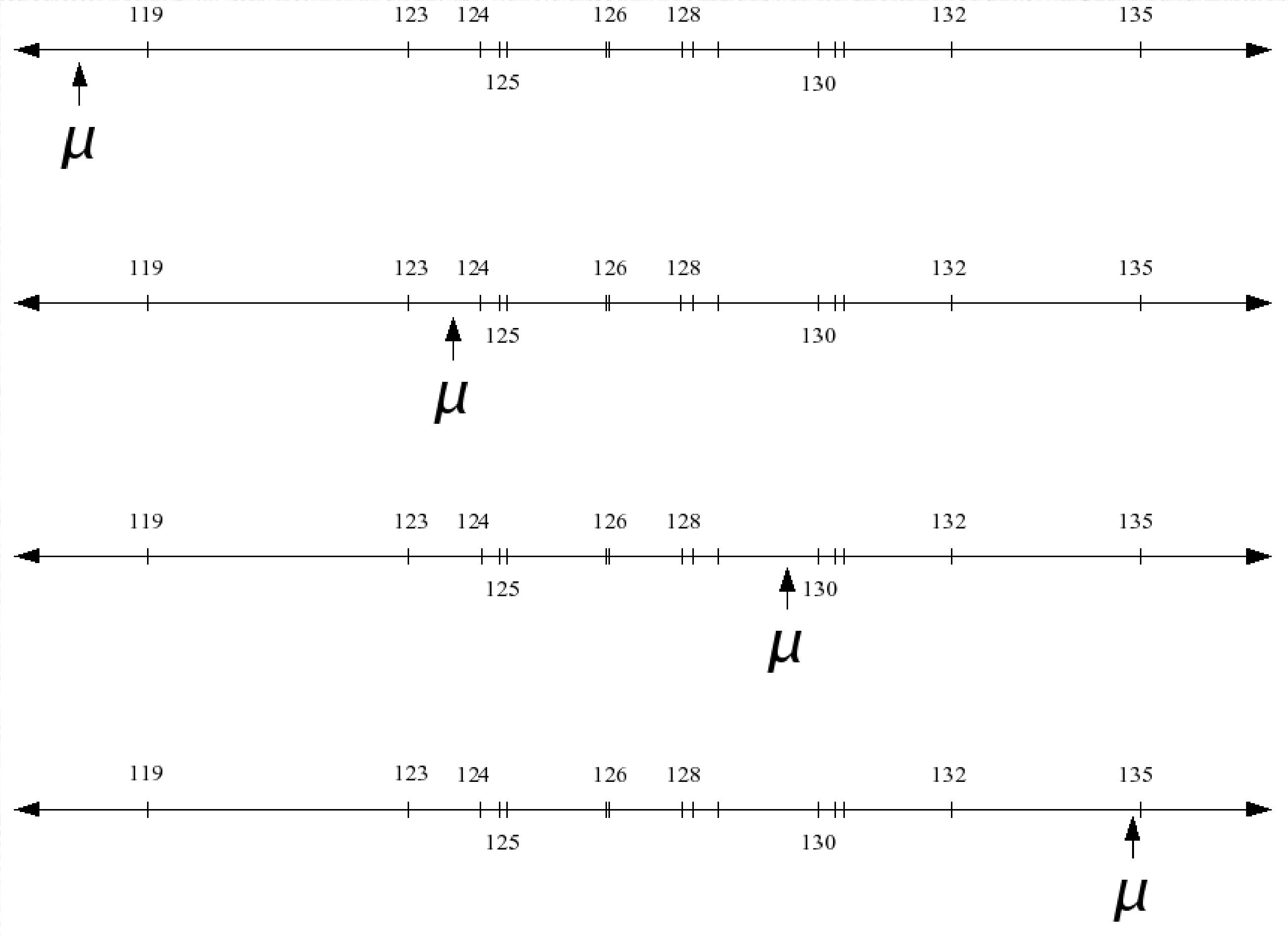
Using likelihood for estimation

- Most common use, well established
- Given some data, what is the best estimate for the mean, μ?
- Is it $\bar{x} = \frac{\sum x_i}{n}$?
- How do we know we have the best estimate?
- Maximum likelihood criterion: the parameter value with the highest likelihood given the data is the best estimate

Some data...



The Data
123.67
126.90
130.78
125.30
124.86
126.96
135.61
119.42
128.74
132.53
130.36
128.31
130.63
128.13
125.17



Infinite number of possible values of μ – which is best?

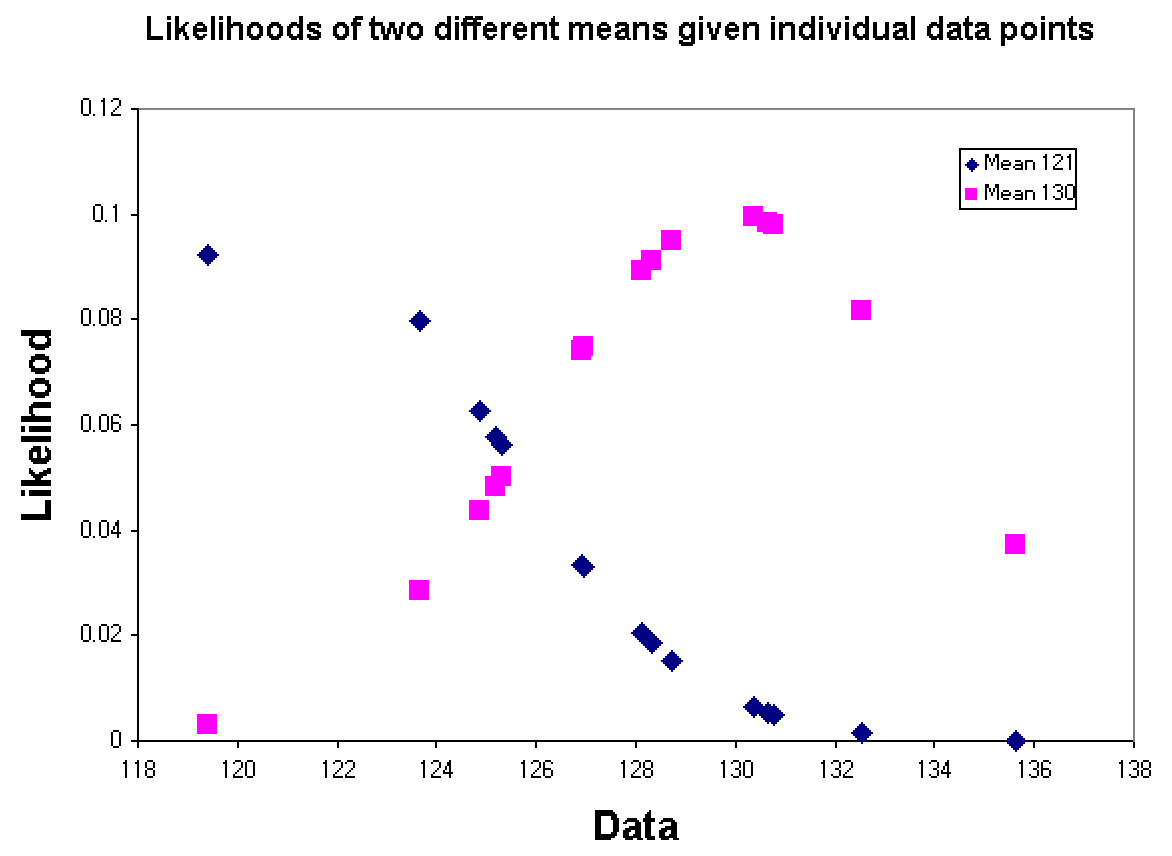
Estimating μ

- The data are what we know
- The best estimate of μ is not
- Need to specify a likelihood function to model deviations of estimates from data points
- We'll use the normal distribution

Likelihood of hypothetical means given the data

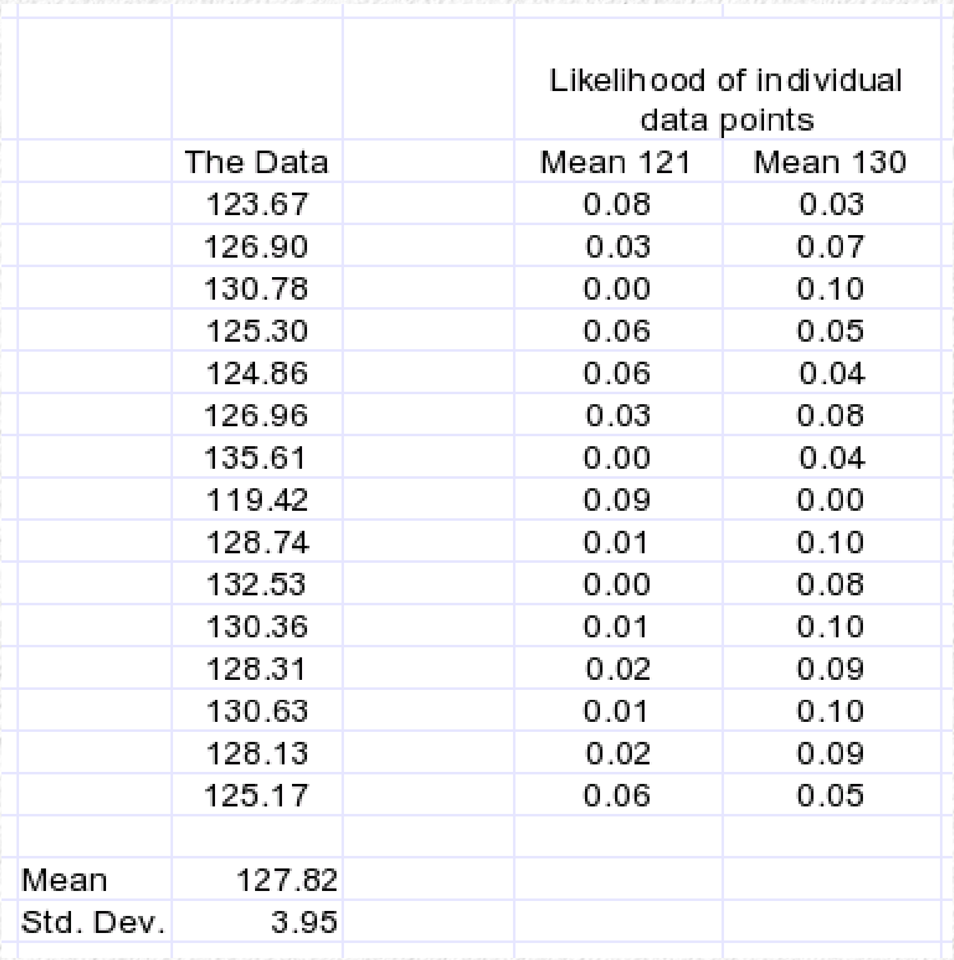
To start:
Two
possible
means

121 and 130



Symbols are likelihoods of either 121 or 130 given just one data point at a time

Example: What is the likelihood of a set of possible means given the data?



			Likelihood of individual data points	
	The Data		Mean 121	Mean 130
	123.67		0.08	0.03
	126.90		0.03	0.07
	130.78		0.00	0.10
	125.30		0.06	0.05
	124.86		0.06	0.04
	126.96		0.03	0.08
	135.61		0.00	0.04
	119.42		0.09	0.00
	128.74		0.01	0.10
	132.53		0.00	0.08
	130.36		0.01	0.10
	128.31		0.02	0.09
	130.63		0.01	0.10
	128.13		0.02	0.09
	125.17		0.06	0.05
Mean	127.82			
Std. Dev.	3.95			

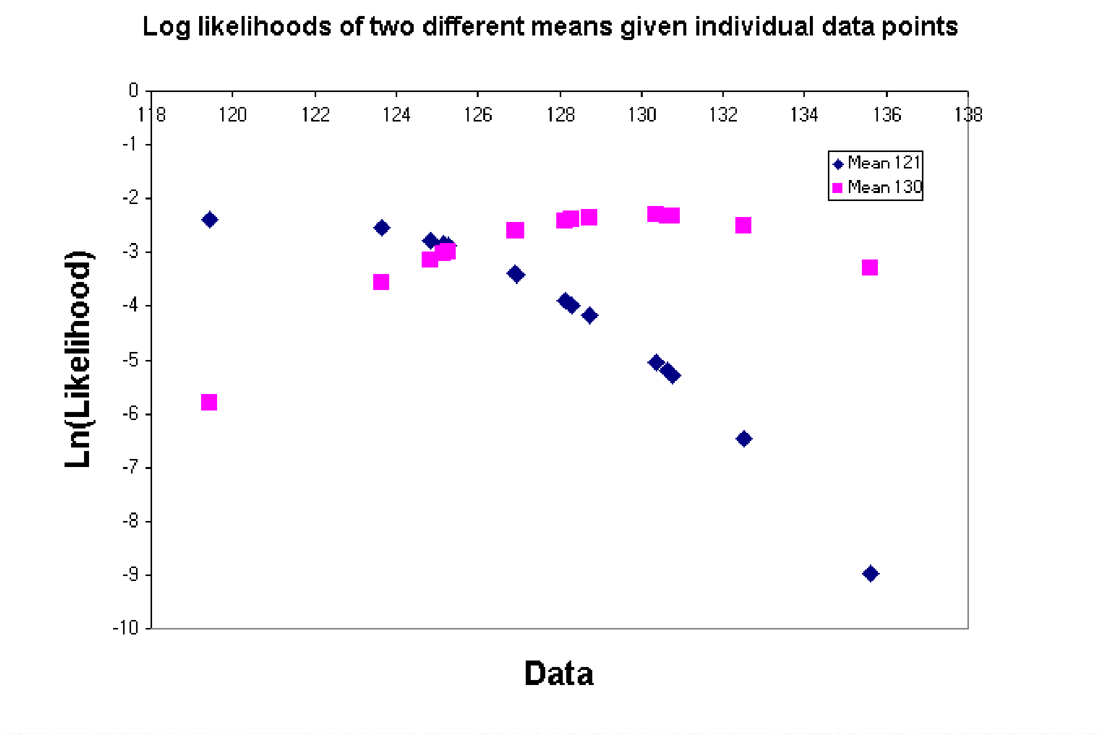
Likelihood of each mean given all of the data: multiply these individual likelihoods together

$$L(121|data) = 3.4 \times 10^{-28}$$

$$L(130|data) = 1.12 \times 10^{-19}$$

Which is bigger?

Ln(likelihood) changes the scale, makes likelihoods additive

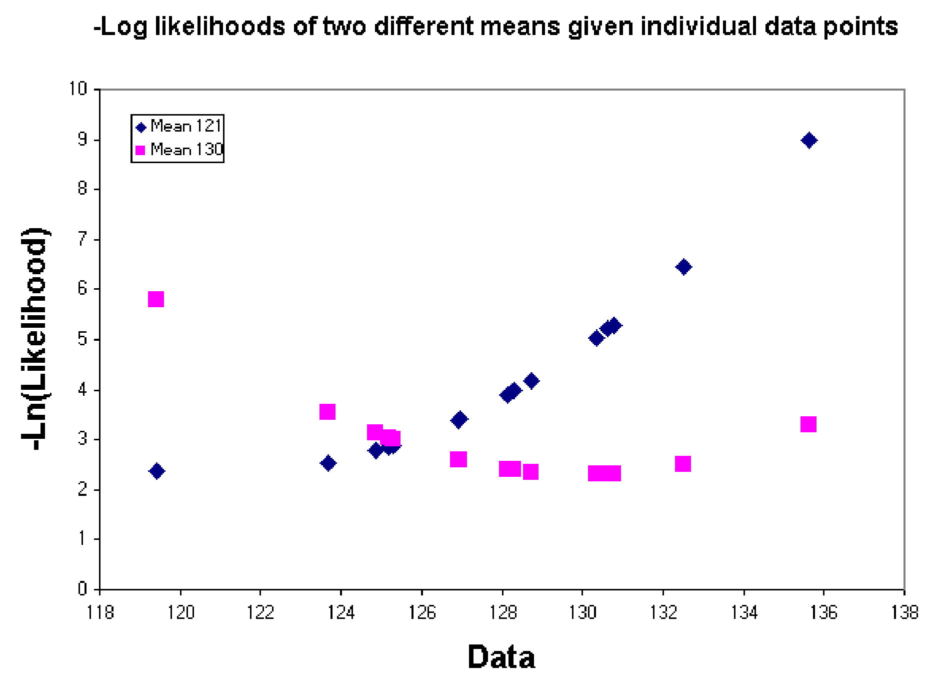


$$\ln(L(121|data)) = -63.79$$

$$\ln(L(130|data)) = -43.68$$

$$Loglik = -0.5\,n\ln(2\pi) - 0.5\,n\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$$

-Ln(likelihood) changes the direction



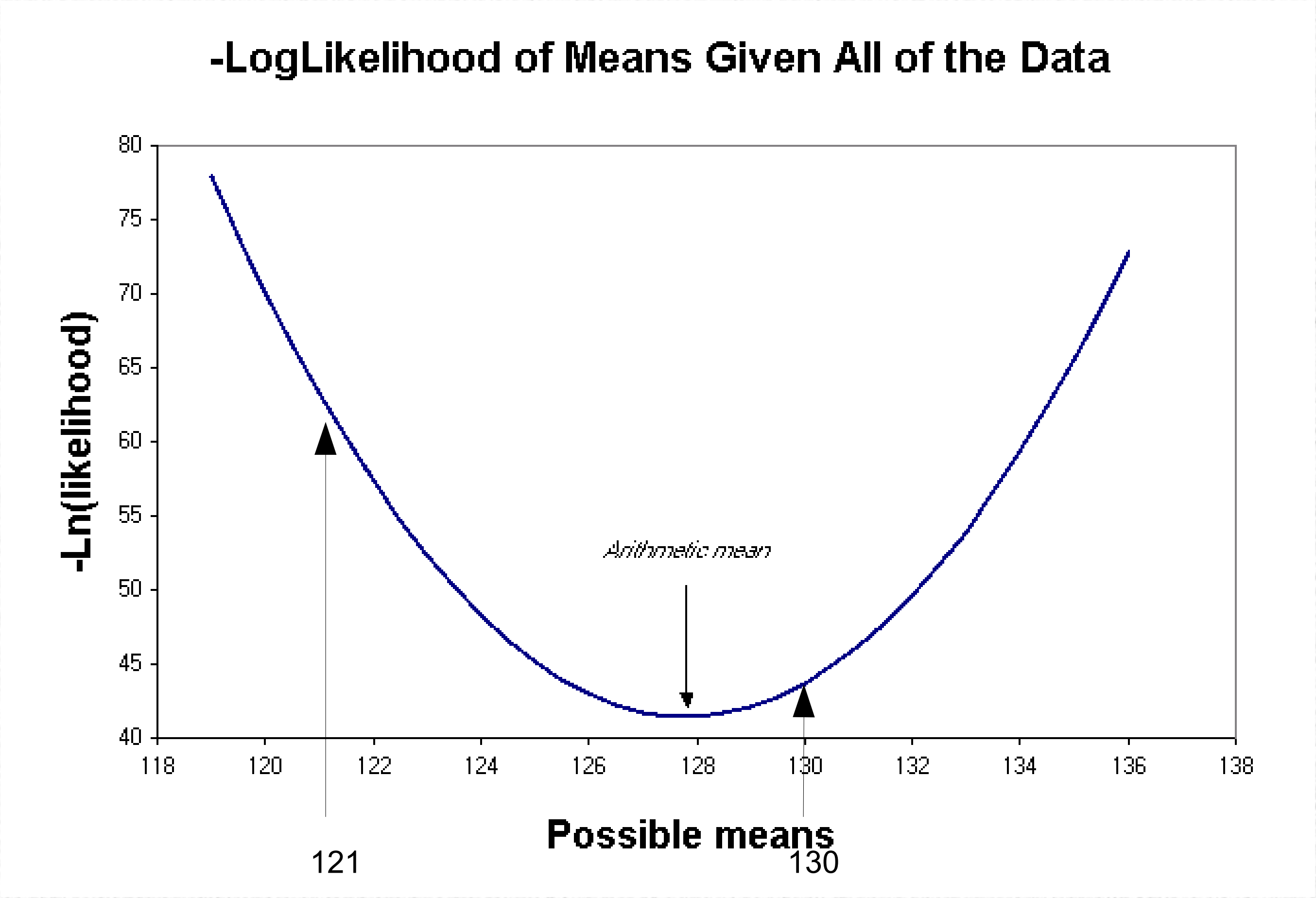
$$-\ln\left(L\left(121|data\right)\right)=63.79$$

$$-\ln\left(L\left(130|data\right)\right)=43.68$$

Why? Convenience (we are bad at negative numbers, -ln(L) has some nice properties we'll meet later)

*But, to find the **maximum likelihood estimate**, we need to find the minimum -logLikelihood value*

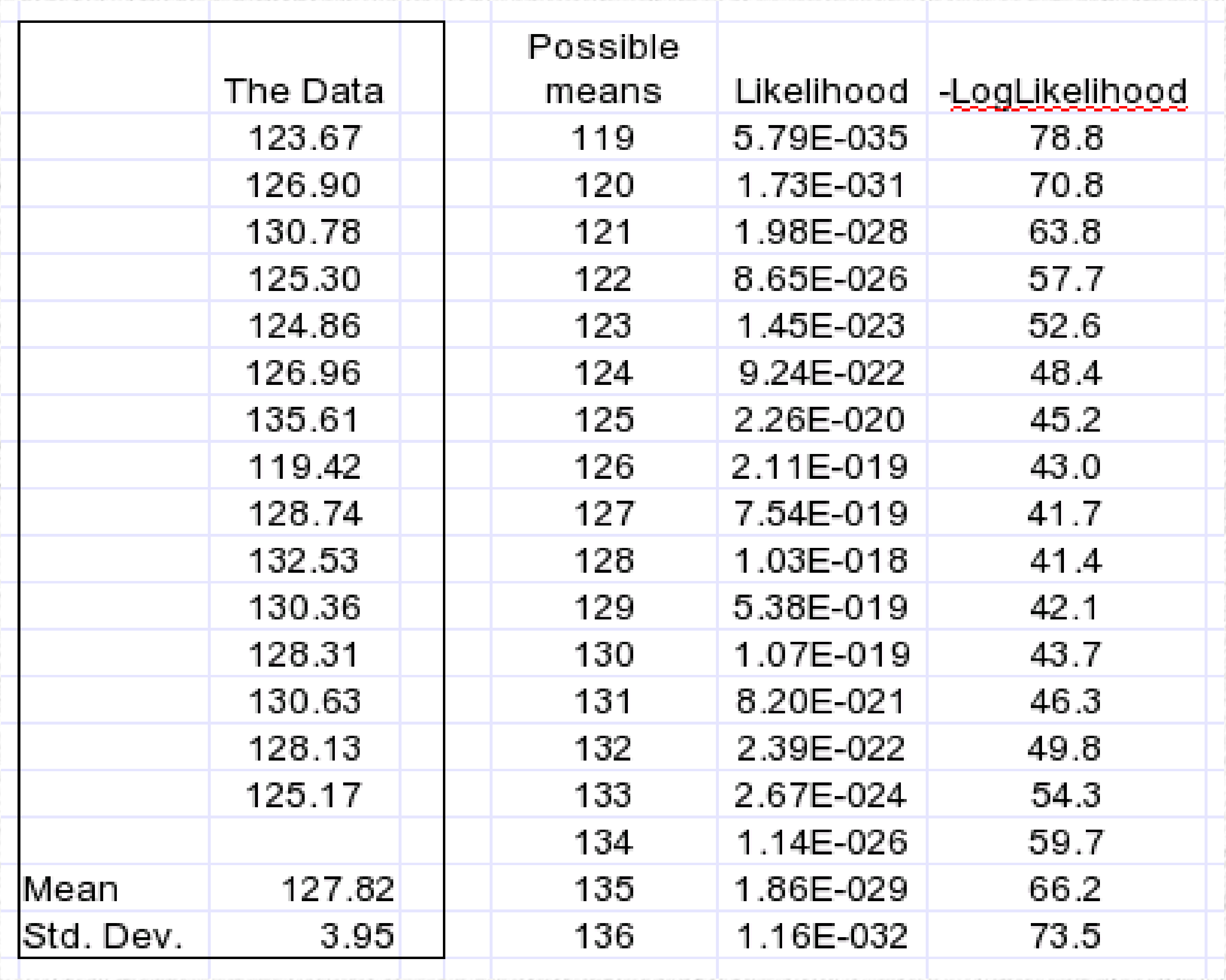
Likelihood of means given the data



Calculate -ln(L) for a range of possible means

***Maximum likelihood estimate** at the minimum of this function*

Calculations – comparing likelihoods among possible means



	The Data	Possible means	Likelihood	-LogLikelihood
	123.67	119	5.79E-035	78.8
	126.90	120	1.73E-031	70.8
	130.78	121	1.98E-028	63.8
	125.30	122	8.65E-026	57.7
	124.86	123	1.45E-023	52.6
	126.96	124	9.24E-022	48.4
	135.61	125	2.26E-020	45.2
	119.42	126	2.11E-019	43.0
	128.74	127	7.54E-019	41.7
	132.53	128	1.03E-018	41.4
	130.36	129	5.38E-019	42.1
	128.31	130	1.07E-019	43.7
	130.63	131	8.20E-021	46.3
	128.13	132	2.39E-022	49.8
	125.17	133	2.67E-024	54.3
		134	1.14E-026	59.7
Mean	127.82	135	1.86E-029	66.2
Std. Dev.	3.95	136	1.16E-032	73.5

Numerical solution – try different possible means, calculate -logLikelihood for each

Pick the one with the lowest -logLikelihood

Not an analytical solution! Only approximately correct (but often good enuf)

Simplifying the likelihood function

- We can drop any term that is a constant, or doesn't depend on the parameter(s) being estimated
- The values will be the same up to an additive constant – shapes will be the same, maximum will be at the same place

So, this

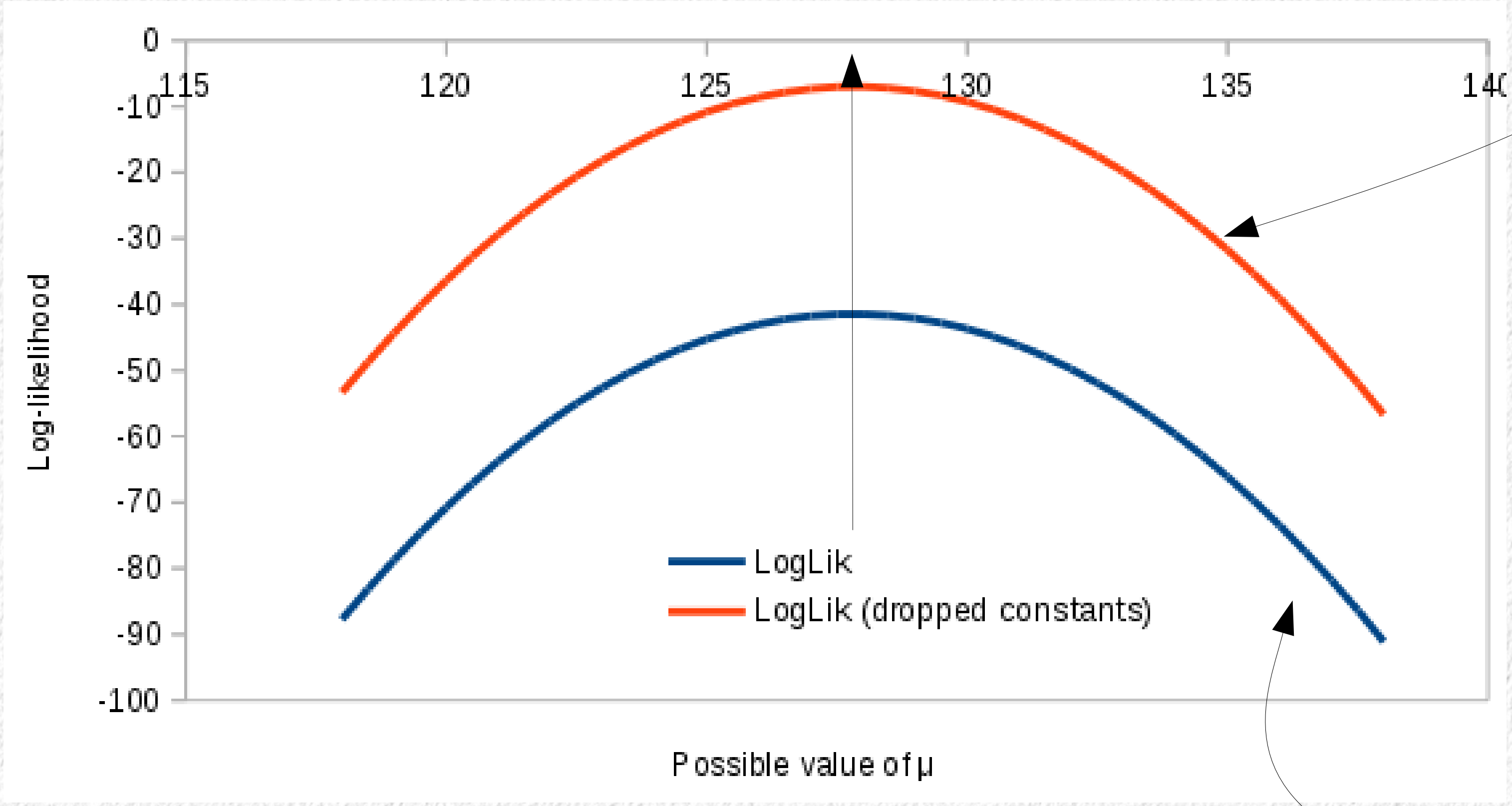
$$-0.5\,n\ln\left(2\pi\right)-0.5\,n\ln\left(\sigma^2\right)-\frac{1}{2\sigma^2}\sum\left(x_i-\mu\right)^2$$

and this

$$-\frac{1}{2\sigma^2}\sum\left(x_i-\mu\right)^2$$

will give the same answer

$$-0.5\,n\ln(2\pi)-0.5\,n\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$



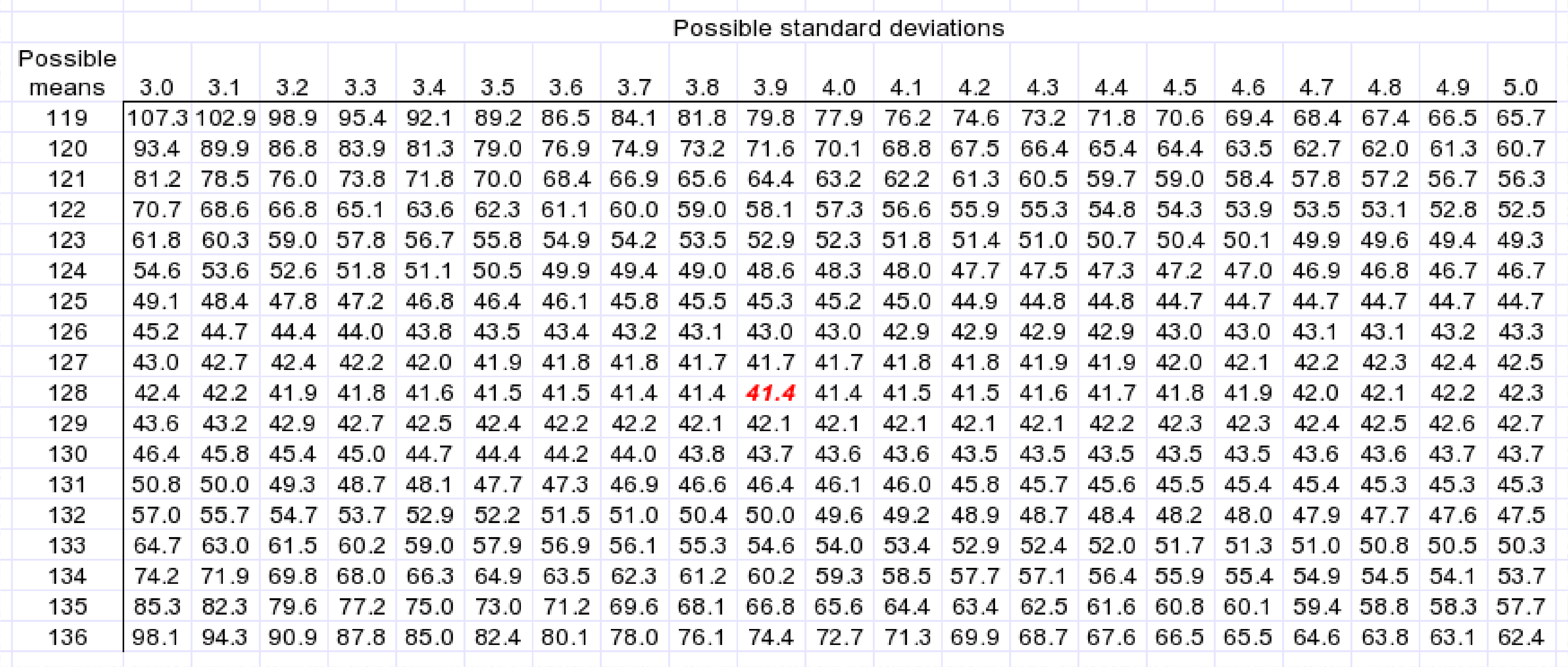
$$-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

Same shapes, differ by a constant amount across the whole curve, both identify the same best value for the estimate of μ

Both the mean and standard deviation can be estimated at once

- The likelihood function includes the standard deviation
- We can vary both the mean and the variance, and select values of both that jointly maximize the likelihood function

-Log likelihoods for combinations of mean and standard deviations

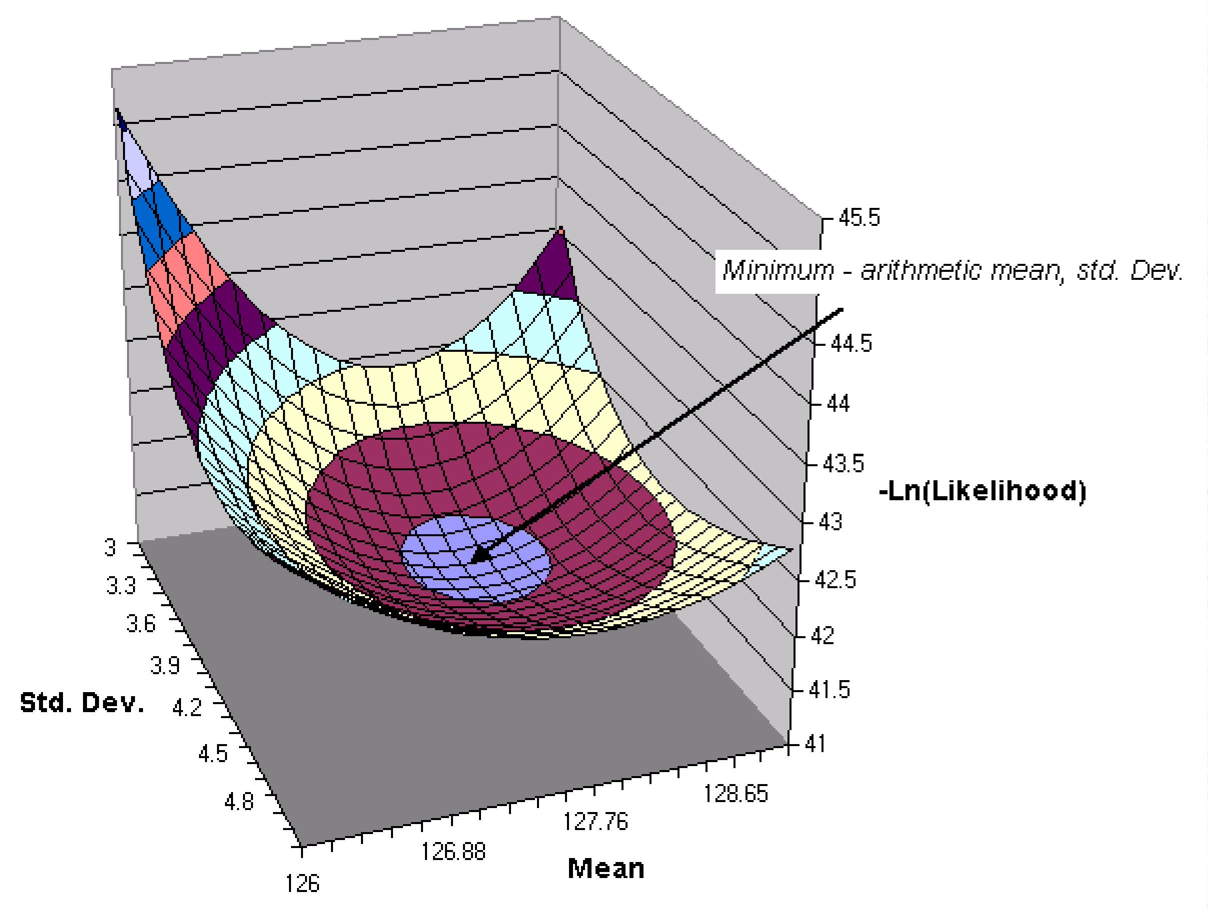


	Possible standard deviations																				
Possible means	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
119	107.3	102.9	98.9	95.4	92.1	89.2	86.5	84.1	81.8	79.8	77.9	76.2	74.6	73.2	71.8	70.6	69.4	68.4	67.4	66.5	65.7
120	93.4	89.9	86.8	83.9	81.3	79.0	76.9	74.9	73.2	71.6	70.1	68.8	67.5	66.4	65.4	64.4	63.5	62.7	62.0	61.3	60.7
121	81.2	78.5	76.0	73.8	71.8	70.0	68.4	66.9	65.6	64.4	63.2	62.2	61.3	60.5	59.7	59.0	58.4	57.8	57.2	56.7	56.3
122	70.7	68.6	66.8	65.1	63.6	62.3	61.1	60.0	59.0	58.1	57.3	56.6	55.9	55.3	54.8	54.3	53.9	53.5	53.1	52.8	52.5
123	61.8	60.3	59.0	57.8	56.7	55.8	54.9	54.2	53.5	52.9	52.3	51.8	51.4	51.0	50.7	50.4	50.1	49.9	49.6	49.4	49.3
124	54.6	53.6	52.6	51.8	51.1	50.5	49.9	49.4	49.0	48.6	48.3	48.0	47.7	47.5	47.3	47.2	47.0	46.9	46.8	46.7	46.7
125	49.1	48.4	47.8	47.2	46.8	46.4	46.1	45.8	45.5	45.3	45.2	45.0	44.9	44.8	44.8	44.7	44.7	44.7	44.7	44.7	44.7
126	45.2	44.7	44.4	44.0	43.8	43.5	43.4	43.2	43.1	43.0	43.0	42.9	42.9	42.9	42.9	43.0	43.0	43.1	43.1	43.2	43.3
127	43.0	42.7	42.4	42.2	42.0	41.9	41.8	41.8	41.7	41.7	41.7	41.8	41.8	41.9	41.9	42.0	42.1	42.2	42.3	42.4	42.5
128	42.4	42.2	41.9	41.8	41.6	41.5	41.5	41.4	41.4	***41.4***	41.4	41.5	41.5	41.6	41.7	41.8	41.9	42.0	42.1	42.2	42.3
129	43.6	43.2	42.9	42.7	42.5	42.4	42.2	42.2	42.1	42.1	42.1	42.1	42.1	42.1	42.2	42.3	42.3	42.4	42.5	42.6	42.7
130	46.4	45.8	45.4	45.0	44.7	44.4	44.2	44.0	43.8	43.7	43.6	43.6	43.5	43.5	43.5	43.5	43.5	43.6	43.6	43.7	43.7
131	50.8	50.0	49.3	48.7	48.1	47.7	47.3	46.9	46.6	46.4	46.1	46.0	45.8	45.7	45.6	45.5	45.4	45.4	45.3	45.3	45.3
132	57.0	55.7	54.7	53.7	52.9	52.2	51.5	51.0	50.4	50.0	49.6	49.2	48.9	48.7	48.4	48.2	48.0	47.9	47.7	47.6	47.5
133	64.7	63.0	61.5	60.2	59.0	57.9	56.9	56.1	55.3	54.6	54.0	53.4	52.9	52.4	52.0	51.7	51.3	51.0	50.8	50.5	50.3
134	74.2	71.9	69.8	68.0	66.3	64.9	63.5	62.3	61.2	60.2	59.3	58.5	57.7	57.1	56.4	55.9	55.4	54.9	54.5	54.1	53.7
135	85.3	82.3	79.6	77.2	75.0	73.0	71.2	69.6	68.1	66.8	65.6	64.4	63.4	62.5	61.6	60.8	60.1	59.4	58.8	58.3	57.7
136	98.1	94.3	90.9	87.8	85.0	82.4	80.1	78.0	76.1	74.4	72.7	71.3	69.9	68.7	67.6	66.5	65.5	64.6	63.8	63.1	62.4

-Ln(likelihood) for each combination of mean and standard deviation

Arithmetic mean = 127.82
Std. Dev. = 3.95

Estimating mean and standard deviation



Curve fitting

- The predicted value from a curve is the average of y expected for a given value of x
- We can calculate the likelihood of parameter values given the residuals around the line that they produce
- We will try this with the photosynthesis data shortly...

DNA evidence

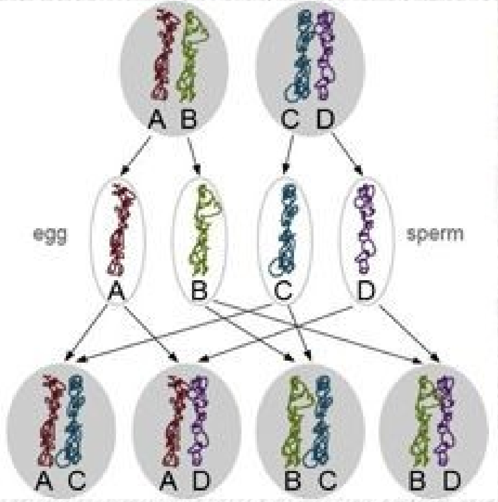
- Like fingerprints, but less subjective
  - Genes are discrete
  - Matches are either positive or negative
  - No judgment calls
- Yet DNA evidence is still subject to statistical uncertainty
  - If matches are either positive or negative, why do they talk about probabilities of a match?

Statistics and DNA evidence

- Probability statements in DNA cases involve some pretty small probabilities
  - Ex. “The chances that this sample belongs to somebody other than the defendant is 100 billion to one”
- What do statisticians mean when they say this?
- Where do these numbers come from?

Genes and alleles

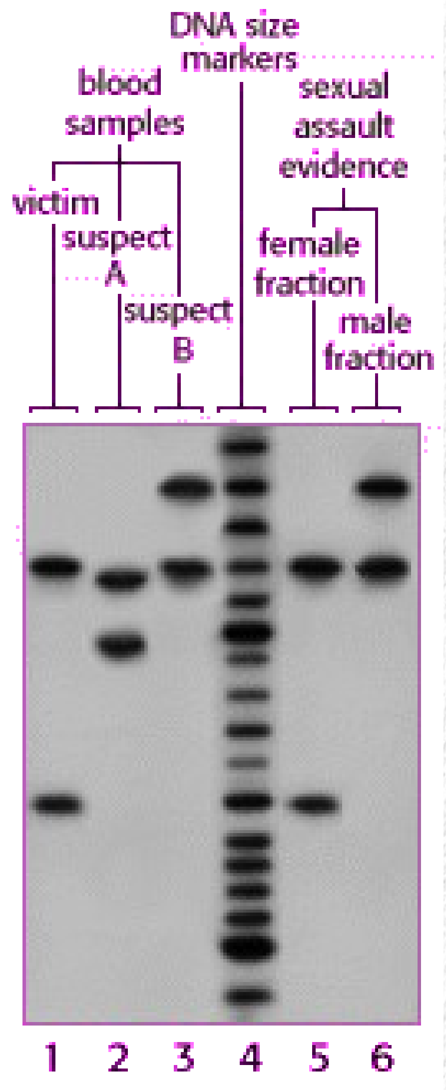
Mom Dad



Kids

- One copy from each parent of our 23 chromosomes
- Each chromosome in a pair has the same genes
- Genes can have different varieties, called “alleles”
- Can get the same alleles from mom and dad, or different ones
- If we get the same ones we're “homozygous”, if we get different ones we're “heterozygous”

Example of a DNA fingerprint - one locus



- Evidence is collected at a crime scene
- Two suspects, A and B, are apprehended
- Samples from the victim and the suspects suspects are compared with the crime scene evidence
- Alleles at a locus appear on a gel as bands
- A suspect that has both of the bands present in the evidence is a match

What's the probability of a genotype?

- If lots of people have the same genotype as the crime scene evidence, chances of a false positive are high – innocent people could be convicted
- We need to know two things:
  - What is the frequency of each allele in the population?

    ...which can be used to tell us...
  - What is the frequency of each genotype in the population?

Hardy-Weinberg with two alleles

Frequency of A in population = p

Frequency of B in population = q

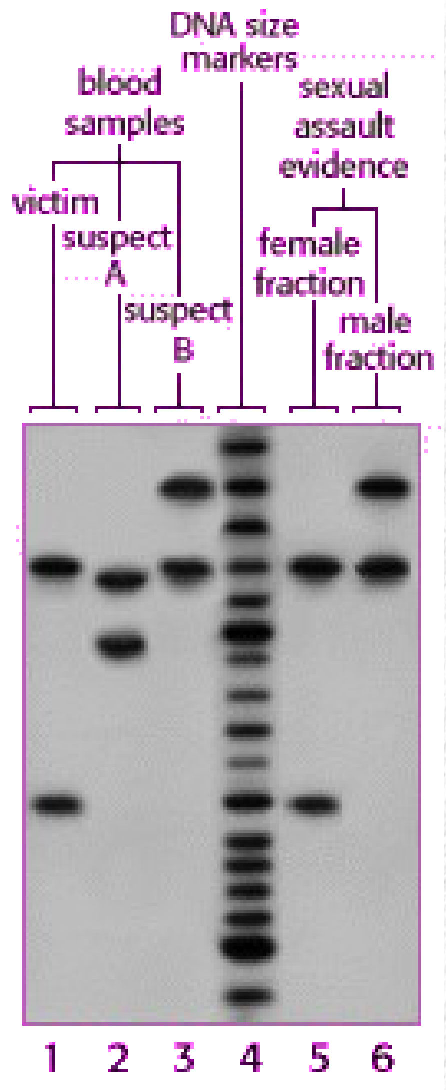
Frequency of AA homozygotes: p^2

Frequency of BB homozygotes: q^2

Frequency of heterozygotes: 2pq

	A (p)	B (q)
A (p)	AA (p^2)	AB (pq)
B (q)	AB (pq)	BB (q^2)

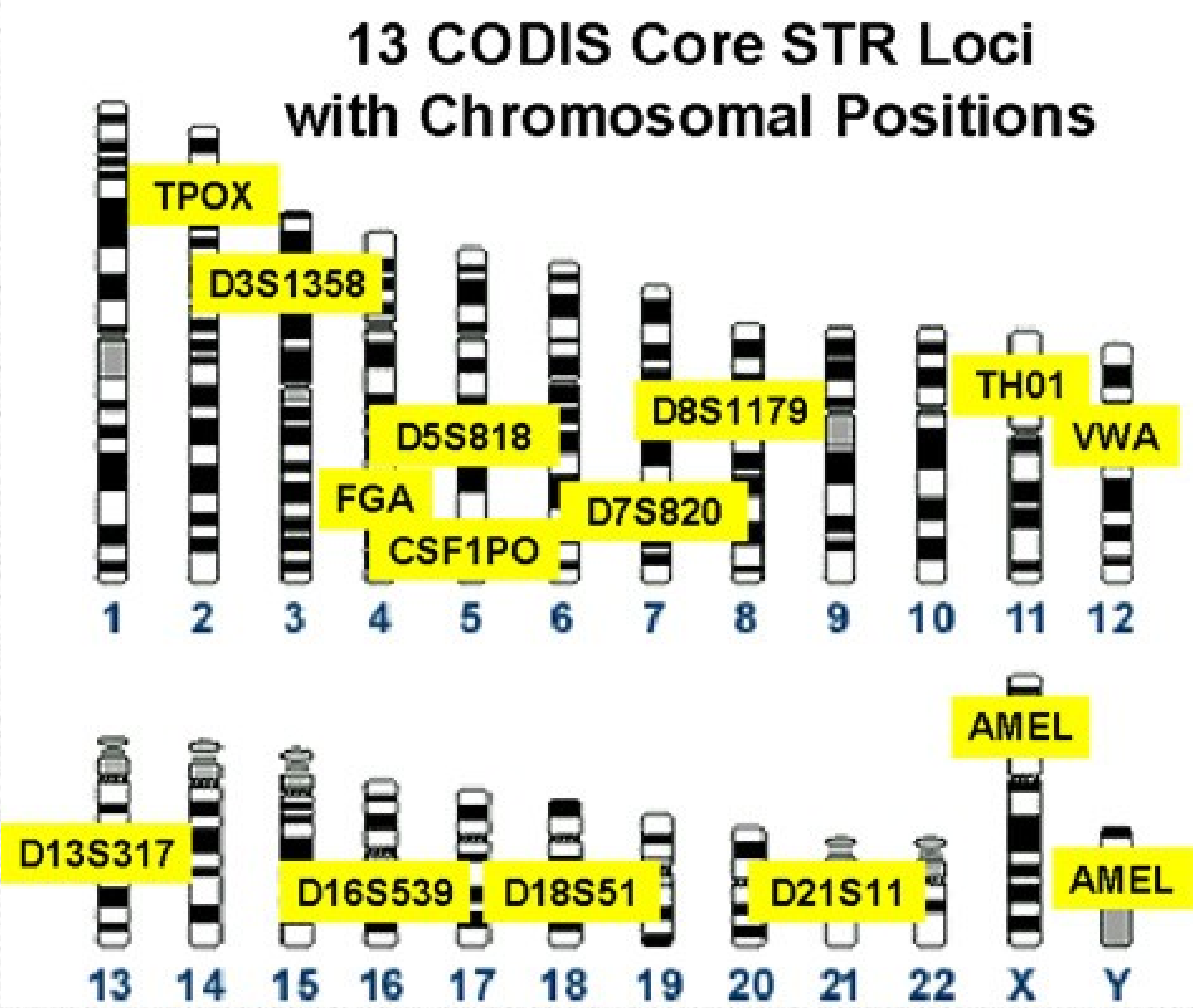
Probability of a match to our suspect



- Suspect B is a match
- If the frequency in the population of the higher band is 0.15, and of the lower band is 0.26...
- The genotype frequency of this heterozygous genotype would be:

  2(0.15)(0.26) = 0.078
- Therefore, we could sample from our population at random, and get a match to this genotype 7.8% of the time
- With San Diego County's population of 3 million, that would be 234,000 matches expected
- Not good enough! The solution? More genes.

13 loci used by FBI



The loci are either on different chromosomes, or are far apart on the same chromosome

Important?

An STR profile

STR Locus	Allele # (maternal)	Allele # (paternal)
D3S1358	16	16
VWA	15	16
FGA	19	20
D8S1179	12	12
D21S11	29	31.2
D18S51	12	17
D5S818	11	13
D13S317	10	11
D7S820	10	12
D16S539	8	11
THO1	6	7
TPOX	8	10
CSF1PO	8	12

Identifiers (this is allele #16)

Convert the profile to genotype frequencies

STR Locus	Allele 1	Allele 2	p	q	Genotype freq.
D3S1358	16	16	0.34	0.34	0.11621
VWA	15	16	0.15	0.26	0.07932
FGA	19	20	0.06	0.07	0.00839
D8S1179	12	12	0.13	0.13	0.01819
D21S11	29	31.2	0.18	0.05	0.01787
D18S51	12	17	0.05	0.18	0.01766
D5S818	11	13	0.24	0.23	0.10986
D13S317	10	11	0.02	0.31	0.01510
D7S820	10	12	0.34	0.12	0.08309
D16S539	8	11	0.04	0.3	0.02285
THO1	6	7	0.15	0.38	0.11535
TPOX	8	10	0.32	0.09	0.05720
CSF1PO	8	12	0.06	0.29	0.03283

Likelihood of genotype at each gene

Likelihood across all the genes is the product of the likelihoods for each = 3.03×10^{-19}

What can we do with this information?

- Likelihoods are usually not interpretable by themselves – need to compare them to other likelihoods
- The likelihood of a match if the sample didn't come from the suspect is tiny – 3.3×10^{-19}
- The likelihood of a match if the suspect is the source of the blood is 1
- The ratio of the likelihood if the suspect is the source of the blood to the likelihood if the suspect is not the source of the blood is a “likelihood ratio”
- Can be used to assess how much more likely one hypothesis is than the other

Likelihood ratio

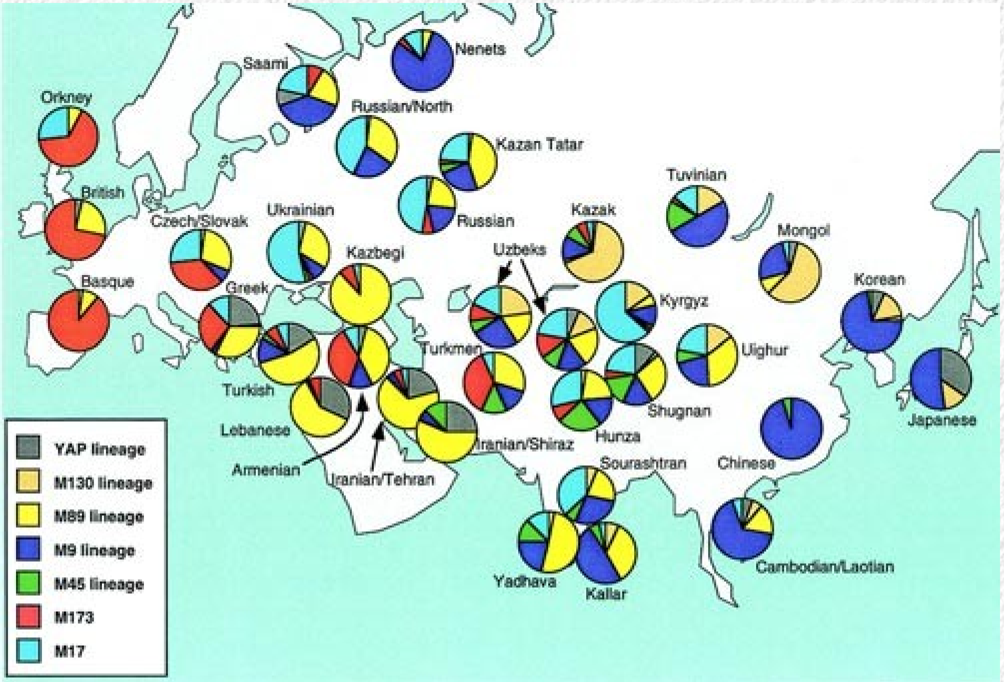
$$\mathrm{LR} = \frac{p\left(\mathrm{Match} \middle| \mathrm{Perp}\right)}{p\left(\mathrm{Match} \middle| \mathrm{Whoops}\right)} = \frac{1}{p\left(\mathrm{Match} \middle| \mathrm{Whoops}\right)}$$

$$\frac{1}{3.3 \times 10^{-19}} = 1 \times 10^{19}$$

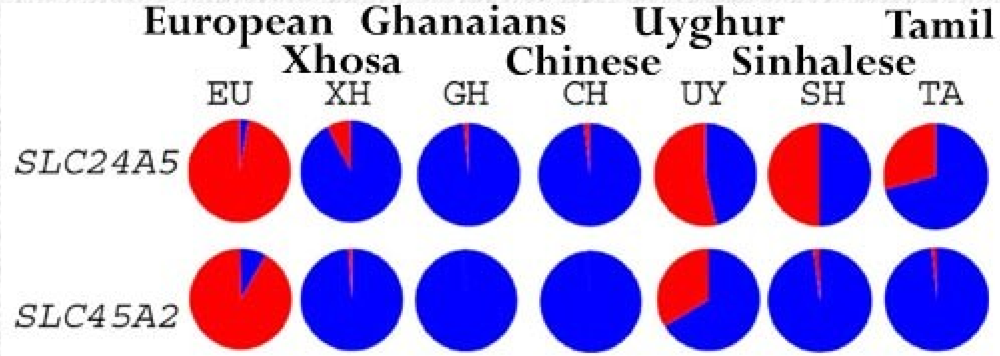
Finding your roots

- Allele frequencies vary among populations
- A DNA fingerprint can be used to assign a person to their most likely population of origin
- Need:
  - Gene frequency data from various populations
  - A DNA fingerprint

Allele frequencies vary among populations



We can make use of this variability to ask, which population is most likely to have produced the STR profile we have in hand?



Example: which population is this person most likely to have come from?

			Bahama African American			Navajo		
STR Locus	Allele 1	Allele 2	p	q	Genotype freq.	p	q	Genotype freq.
D3S1358	16	16	0.34	0.34	0.11621	0.14	0.14	0.02006
VWA	15	16	0.15	0.26	0.07932	0.02	0.43	0.01646
FGA	19	20	0.06	0.07	0.00839	0.19	0.09	0.03389
D8S1179	12	12	0.13	0.13	0.01819	0.11	0.11	0.01265
D21S11	29	31.2	0.18	0.05	0.01787	0.18	0.06	0.02219
D18S51	12	17	0.05	0.18	0.01766	0.09	0.12	0.02199
D5S818	11	13	0.24	0.23	0.10986	0.58	0.05	0.05766
D13S317	10	11	0.02	0.31	0.01510	0.15	0.22	0.06724
D7S820	10	12	0.34	0.12	0.08309	0.14	0.28	0.08088
D16S539	8	11	0.04	0.3	0.02285	0.01	0.15	0.00409
THO1	6	7	0.15	0.38	0.11535	0.17	0.61	0.20534
TPOX	8	10	0.32	0.09	0.05720	0.35	0.02	0.01151
CSF1PO	8	12	0.06	0.29	0.03283	0.02	0.29	0.00970

Populations differ in their gene frequencies and genotype frequencies

Likelihood of being a Bahama African American = 3.03×10^{-19}

Likelihood of being a Navajo = 2.03×10^{-21}

Likelihood of belonging to the Bahama African American population rather than the Navajo population

$$Likelihood\ ratio = \frac{L(profile|BAA)}{L(profile|Nav)} = \frac{3.03\times 10^{-19}}{2.03\times 10^{-21}} = 149.26$$

- 148.26 times more likely that this profile is from BAA rather than Navajo population
- Conclusion depends on comparison group used – think of these as competing hypotheses, and we are only considering two possible alternatives here