STR Y haplotype 95% CI

X = 2 n= 1,000

- I made the x and n up AN EXAMPLE!
- frequency = 2/1,000 = 0.002
- (1-p) = 0.998
- What is the frequency at which we would see zero of the “p” haplotype in a data base of 1,000?

$n!/k!(n-k)!\{p\}^{0}\{1-p\}^{1000} = 1000!/0!1000!\{.998\}^{1000}$

$= 1\{.998\}^{1000} = 0.135$

The coefficient is 1, because there is only one way to get all non”p” haplotypes.

X = 2 n= 1,000

- frequency = 2/1,000 = 0.002
- (1-p) = 0.998
- What is the frequency at which we would see one of the “p” haplotype in a data base of 1,000?

  = $1000!/1!999!\{.998\}^{999}\{.002\}$

  = $1000*999!/999!1![\{.998\}^{999}\{0.002\}]$

  = 1000[{0.135}{0.002}]=0.2706

Why is the coefficent 1000?

Adding them up

- First we asked what is the expected frequency of zero observances of “p” haplotype
  =0.135
  Then we asked about one “p” haplotype
  = 0.2706
  = 0.41

If the frequency is 2/1,000 we would expect to see either none or one 41% of the time in a data base of 1,000.

$X = 2 \quad n = 1,000$

- frequency = 2/1,000 = 0.002
- (1-p) = 0.998
- What is the frequency at which we would see two of the “p” haplotype in a data base of 1,000?

$1000!/2!*998!\{.998\}^{998}\{0.002\}^2$

$= 1000*999/2!\{0.13561\}\{0.000004\}$

$=499500\{0.000000542\}=0.271$

Adding them up

- First we asked what is the expected frequency
  of zero observances of “p” haplotype
  =0.135
  Then we asked about one “p” haplotype
  = 0.2706
  Then we asked about two “p” haplotypes
  =0.2709
  = 0.681

If the frequency is 2/1,000 we would expect to see either none, one or two 68% of the time in a data base of 1,000.

So the upper 68% CONFIDENCE INTERVAL is 2/1000

X = 2 n= 1,000

- frequency = 2/1,000 = 0.002
- (1-p) = 0.998
- What is the frequency at which we would see three of the “p” haplotype in a data base of 1,000?

= $1000*999*998*997!/997!3![\{.998\}^{997}\{0.002\}^{3}]$

= $1000*999*998/3*2[\{0.13589\}\{8*10^{-9}\}]$

= $500*333*998\{1*10^{-9}\}=0.181$

Adding them up

- First we asked what is the expected frequency of zero observances of “p” haplotype
  =0.135
  Then we asked about one “p” haplotype
  = 0.2706
  Then we asked about two “p” haplotypes
  =0.2709
  Then we asked about three “p” haplotypes
  =0.181
  = 0.862

If the frequency is 2/1,000 we would expect to
see either none, one, two or three 86% of the time in a data base of 1,000.
So the upper 86% CI is 3/1000.

$X = 2 \quad n = 1{,}000$

- frequency = 2/1,000 = 0.002
- (1-p) = 0.998
- What is the frequency at which we would see four of the "p" haplotype in a data base of 1,000?

$= 1000*999*998*997*996!!/996!4![\{.998\}^{996}\{0.002\}^{4}$

$=250*333*499*997[2.18*10^{-12}] = 0.09$

Adding them up

- First we asked what is the expected frequency of zero observances of “p” haplotype
  =0.135
  Then we asked about one “p” haplotype
  = 0.2706
  Then we asked about two “p” haplotypes
  =0.2709
  Then we asked about three “p” haplotypes
  =0.181
  Then we asked about four “p” haplotypes
  = 0.090
  = 0.952

If the frequency is 2/1,000 we would expect to see either none, one, two, three or four “p” haplotypes 95% of the time in a data base of 1,000.

THE UPPER 95% CI is 4/1000.

How do we find the upper 99% CI?

GUESS WHAT IT IS?

Why is the frequency declining?