

## Statistical Variation

Roberto Puch-Solis

The idea is to motivate the need of statistics based on a question that a practitioner once asked me. The statistician working for the defence wrote: “...Mr Smith’s Y STR profile has also been observed in 40 profiles from a European database of 5,000 Euro-Asian profiles. Taking statistical variation into account, the latter figure is consistent with a Europe/West Asia frequency of the profile of about 1 man in every 100 ( $\approx$  upper 97.5% confidence limit)”. The question that the practitioner asked was: “What does *taking statistical variation into account* means?”

The contents of the talk will be:

- 1) Brief Y-STRs overview, to place question into context.
- 2) Discrete probability distribution – the proportion of people having a Y-STR profile follows a discrete probability distribution that has only two values. Discussion of the definition of discrete probability distributions would follow. If time permit, show more examples, for example: allele distributions in a population in standard DNA profiles.
- 3) Variation - introduction of statistical variation through a experiment. We know that when we toss a coin, the probability that we get a head, or a tail, is 0.5. However, if we toss a coin ten times, one may get 4 heads. If we do it again we can get 6 heads. The more tosses we do the less variable is the proportion. This leads to the question on how to quantify statistical variation. Connect this question to the statistical variation question in Y-STR’s.
- 4) More computer experiments to show the emerging normal distribution. This leads to the definition of continuous variable and the introduction of a normal distribution. Show other examples of continuous distributions if time permits, for example, Gamma, or Lognormal distribution, for representing peak heights from a process.
- 5) Confidence intervals. Use the previous knowledge to define what a confidence interval is and its connection to Y-STR question, in detail.
- 6) Probability distribution for proportion using Beta distribution.