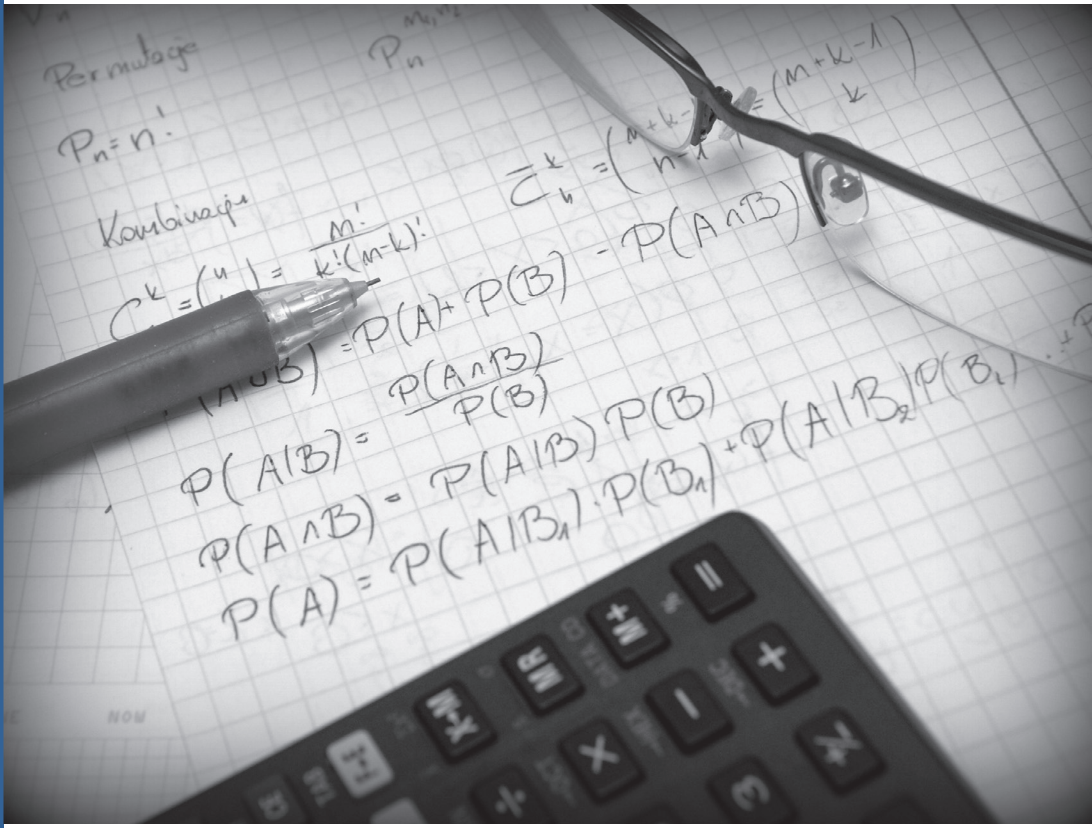
www.forensicCOE.org

Workshop on Statistics and Probability in Forensics Science

Cedric Neumann


Permutacje
$P_n = n!$
Kombinacje
$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(A|B) = \frac{P(A \cap B)}{P(B)}$
$P(A \cap B) = P(A|B)P(B)$
$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2)P(B_2)$
MR
M+
M−

May 17th–19th, 2016

TWO N'S FORENSICS

NIJ National Institute of Justice
STRENGTHEN SCIENCE. ADVANCE JUSTICE.

RTI INTERNATIONAL

Workshop on Statistics and Probability in Forensics Science

Cedric Neumann

DAY 1

MAY 17TH

TWO N'S FORENSICS

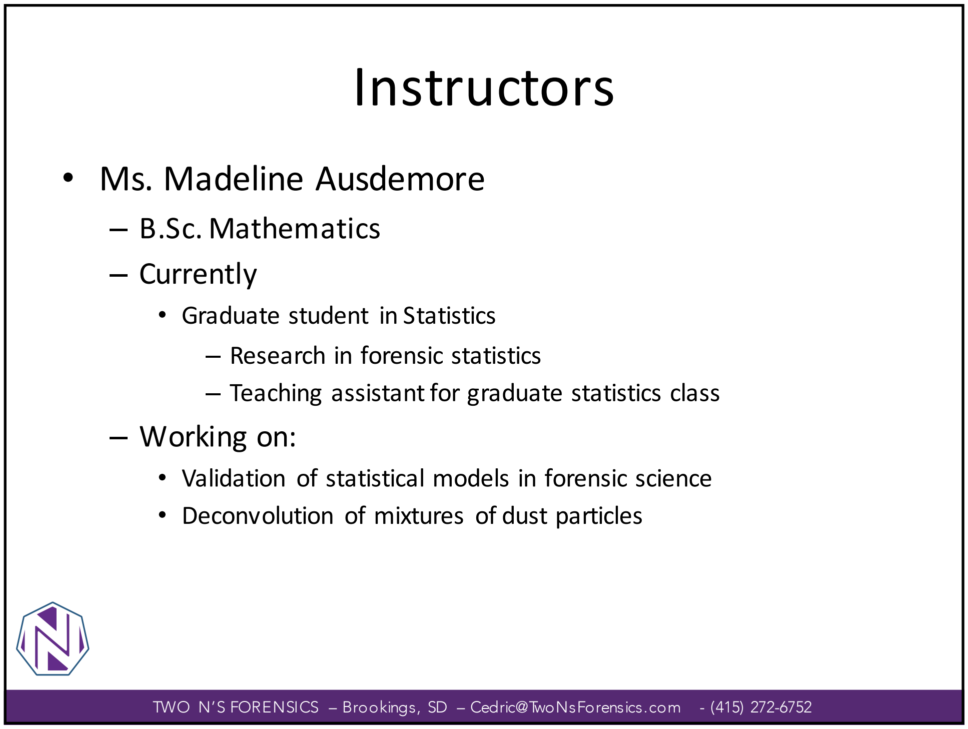
Workshop on Statistics and Probability in Forensics Science

Cedric Neumann

May 17th – 19th, 2016

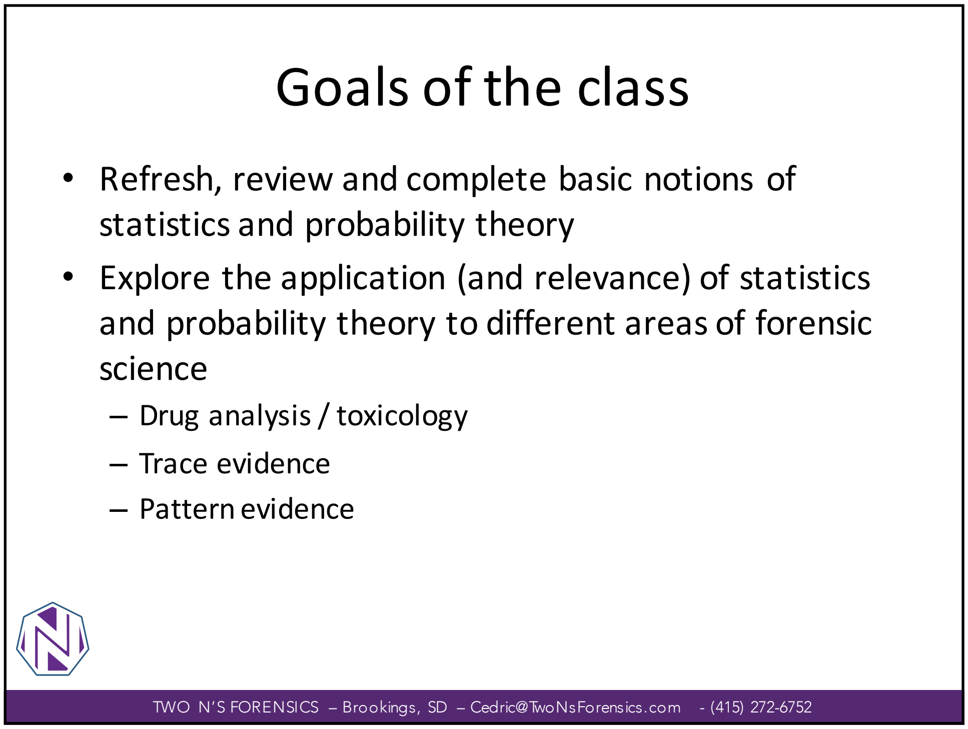
Instructors

- Dr. Cedric Neumann
  - Ph.D. in Forensic Science
    - Focused on pattern recognition and statistics
  - Assistant Professor of Statistics, SD State Univ.
  - Previously
    - Assistant Professor of Statistics and Forensic Science, PennState
    - Scientific Manager of R&D Statistics group at UK FSS
  - Working on statistical models to:
    - Quantify probative value of forensic evidence
    - Support decision-making during examination process



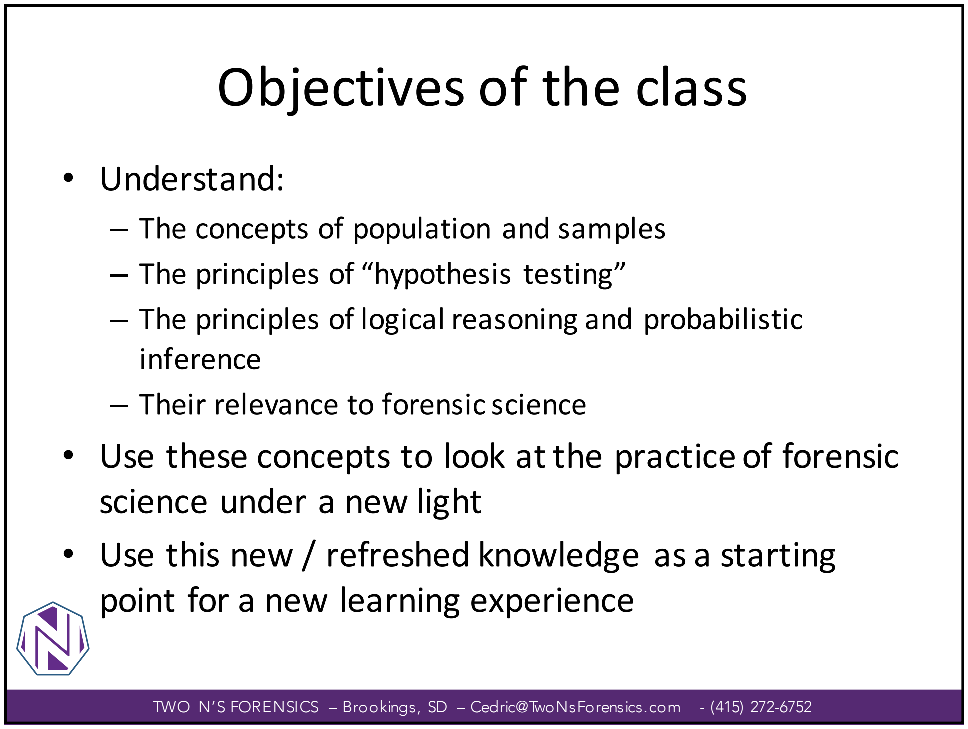
Instructors

- Ms. Madeline Ausdemore
  - B.Sc. Mathematics
  - Currently
    - Graduate student in Statistics
      - Research in forensic statistics
      - Teaching assistant for graduate statistics class
  - Working on:
    - Validation of statistical models in forensic science
    - Deconvolution of mixtures of dust particles



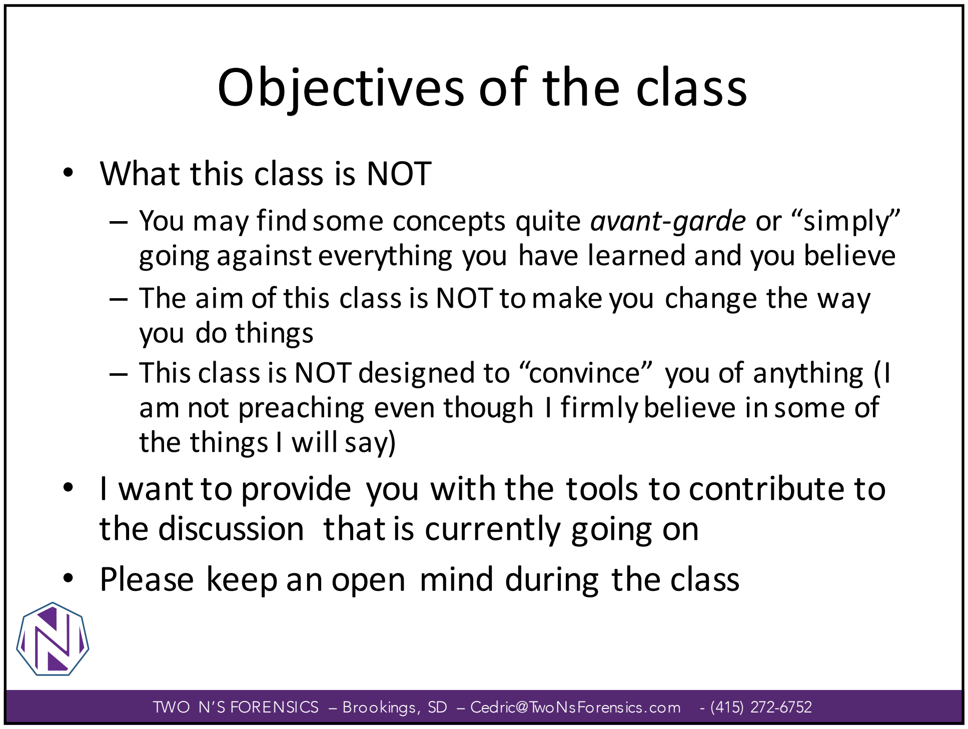
Goals of the class

- Refresh, review and complete basic notions of statistics and probability theory
- Explore the application (and relevance) of statistics and probability theory to different areas of forensic science
  - Drug analysis / toxicology
  - Trace evidence
  - Pattern evidence



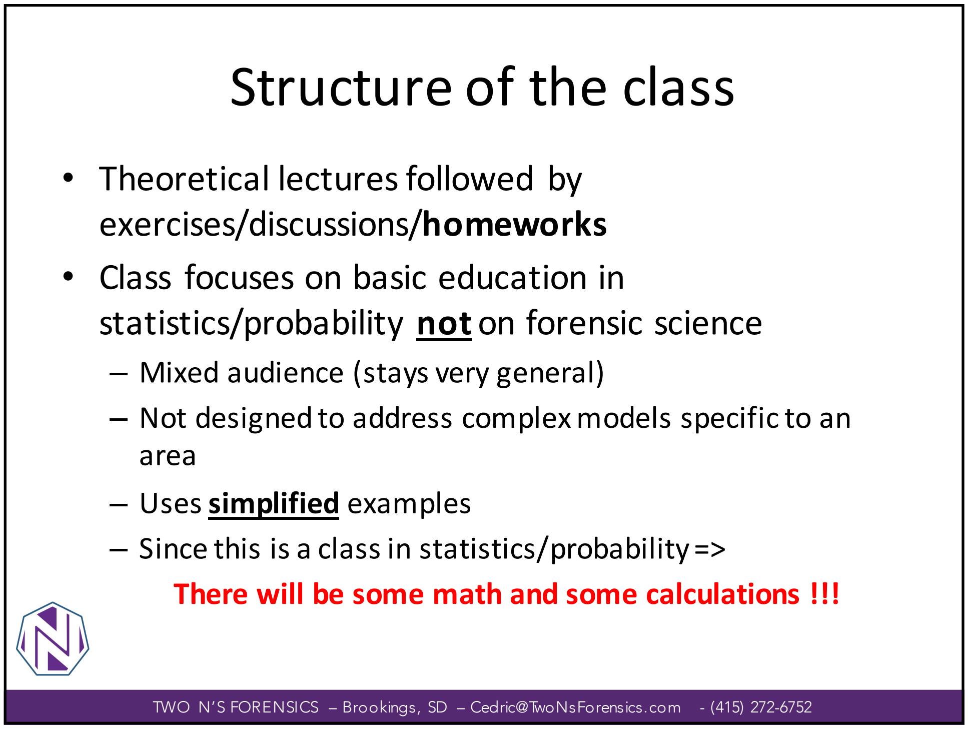
Objectives of the class

- Understand:
  - The concepts of population and samples
  - The principles of “hypothesis testing”
  - The principles of logical reasoning and probabilistic inference
  - Their relevance to forensic science
- Use these concepts to look at the practice of forensic science under a new light
- Use this new / refreshed knowledge as a starting point for a new learning experience



Objectives of the class

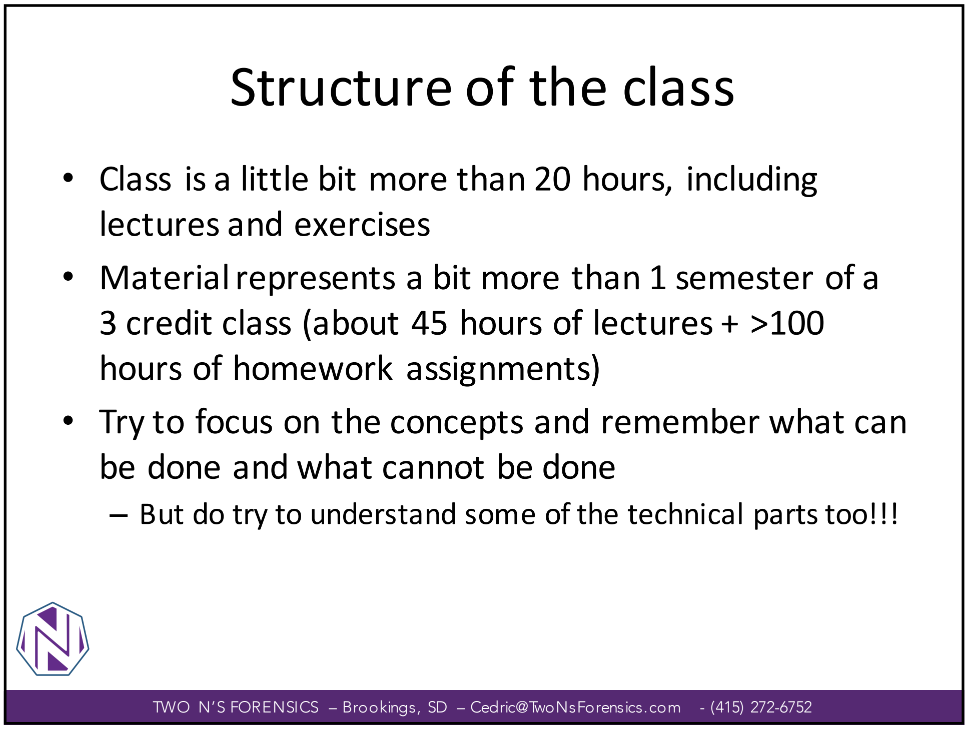
- What this class is NOT
  - You may find some concepts quite *avant-garde* or “simply” going against everything you have learned and you believe
  - The aim of this class is NOT to make you change the way you do things
  - This class is NOT designed to “convince” you of anything (I am not preaching even though I firmly believe in some of the things I will say)
- I want to provide you with the tools to contribute to the discussion that is currently going on
- Please keep an open mind during the class



Structure of the class

- Theoretical lectures followed by exercises/discussions/**homeworks**
- Class focuses on basic education in statistics/probability **not** on forensic science
  - Mixed audience (stays very general)
  - Not designed to address complex models specific to an area
  - Uses **simplified** examples
  - Since this is a class in statistics/probability =>

There will be some math and some calculations !!!



Structure of the class

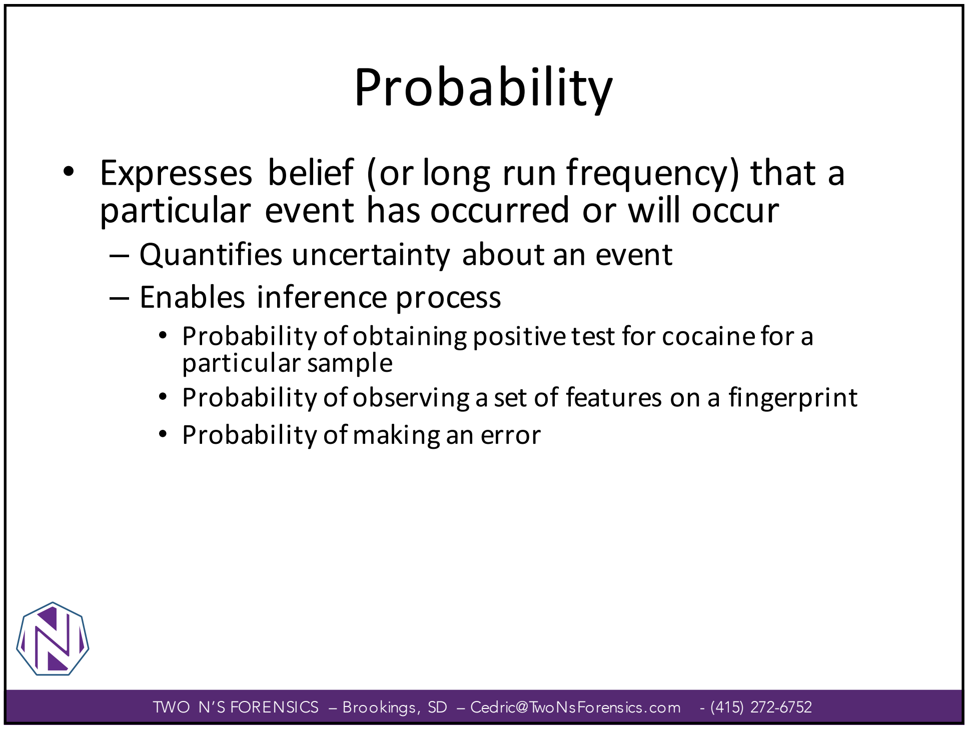
- Class is a little bit more than 20 hours, including lectures and exercises
- Material represents a bit more than 1 semester of a 3 credit class (about 45 hours of lectures + >100 hours of homework assignments)
- Try to focus on the concepts and remember what can be done and what cannot be done
  - But do try to understand some of the technical parts too!!!

Chapter 0

GENERAL DEFINITIONS

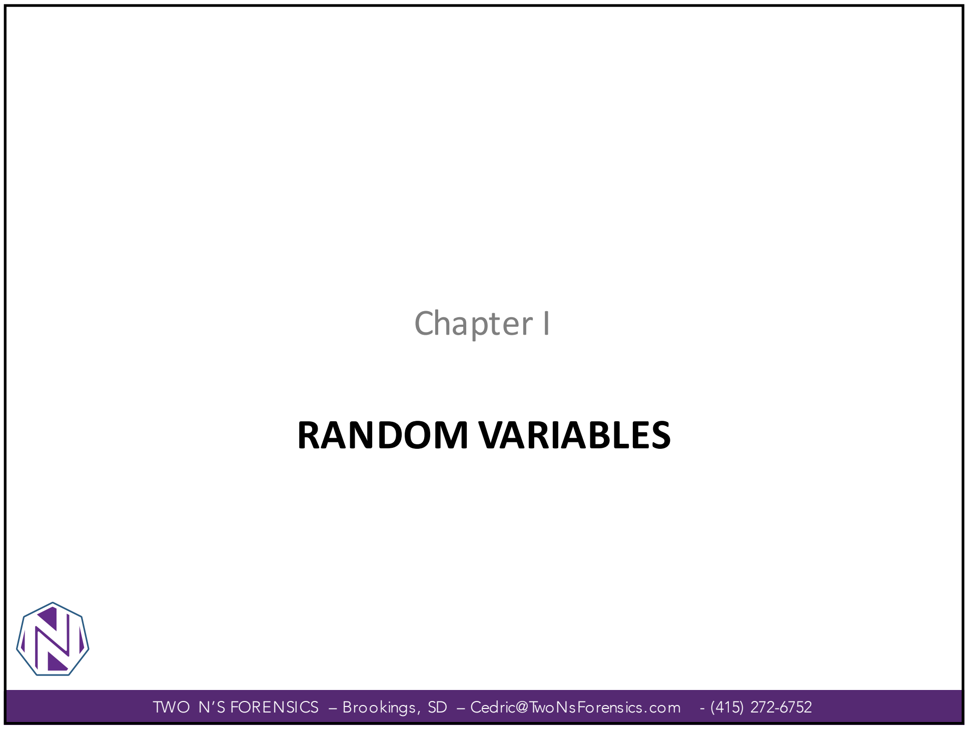
Statistics

- Pertains to the collection, analysis, interpretation and presentation of data
  - Data: observations made on objects and recorded in *variables*
    - *Calibration data for an analytical technique*
    - *Frequency of shoe sole patterns/sizes*
    - *Number of features in agreement used by examiners to form opinions*
    - *(Analytical, within-individual, between-individuals) variability of blood alcohol content under various conditions*



Probability

- Expresses belief (or long run frequency) that a particular event has occurred or will occur
  - Quantifies uncertainty about an event
  - Enables inference process
    - Probability of obtaining positive test for cocaine for a particular sample
    - Probability of observing a set of features on a fingerprint
    - Probability of making an error



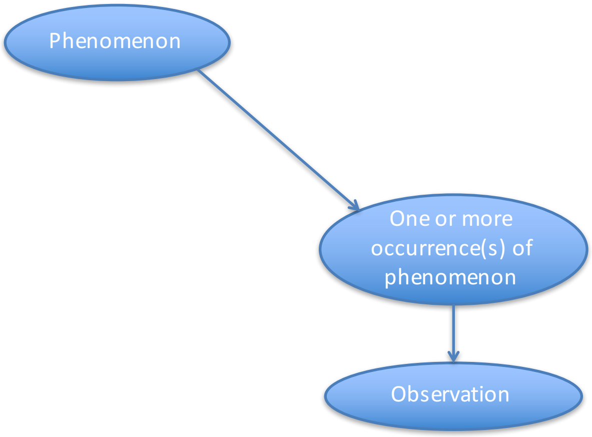
Chapter I

RANDOM VARIABLES

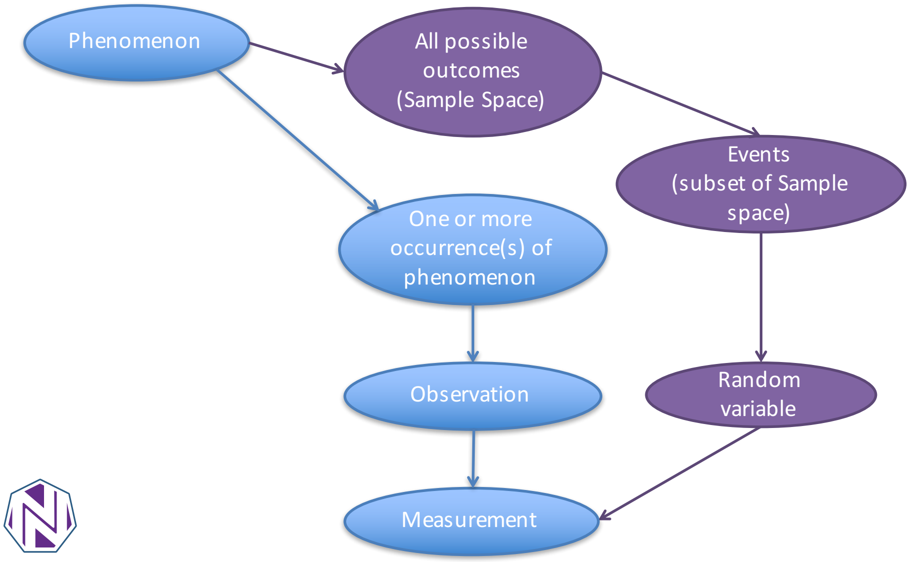
Science

- Scientists are interested in studying phenomena
  - They can be more or less complicated
    - Human height
    - Particles resulting from the collision of other particles
  - They can be theoretical
  - We usually want to do one of two things (sometimes both)
    - **Explain** a phenomenon by a set of observed outcomes
    - **Predict** a specific outcome of a phenomenon

Science

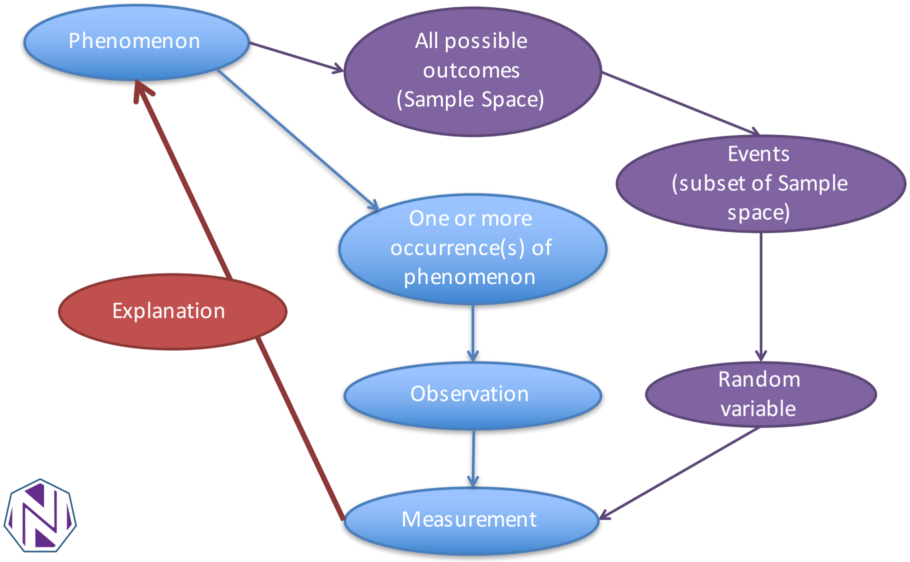


Science

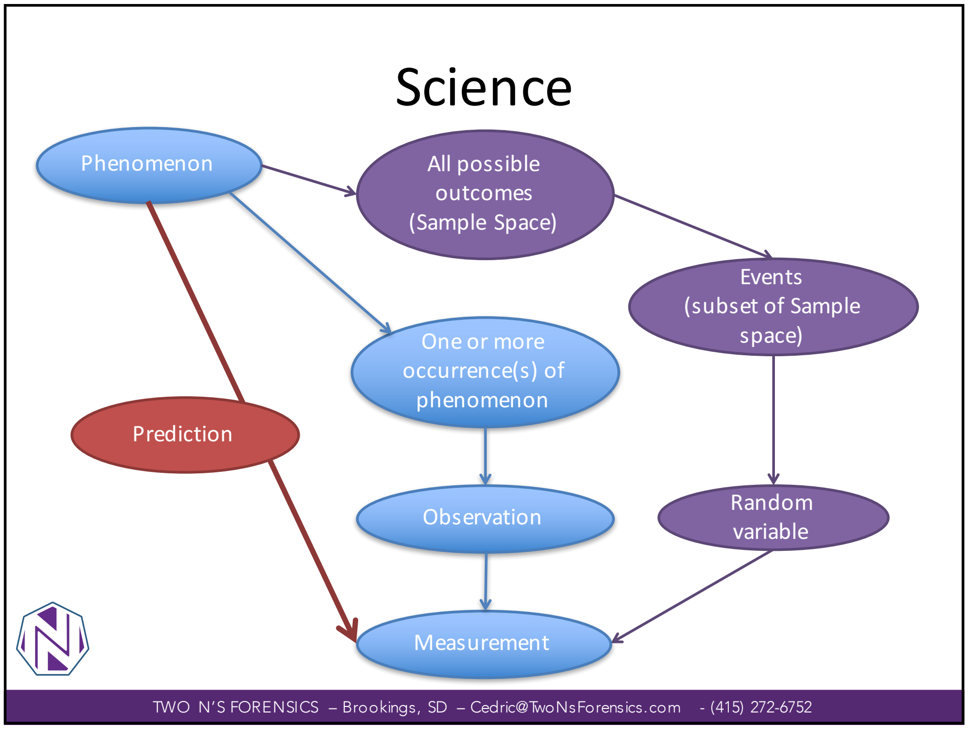

Phenomenon
All possible outcomes (Sample Space)
Events (subset of Sample space)
One or more occurrence(s) of phenomenon
Observation
Random variable
Measurement

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Science


Phenomenon
All possible outcomes (Sample Space)
Events (subset of Sample space)
One or more occurrence(s) of phenomenon
Explanation
Observation
Random variable
Measurement

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Science

Phenomenon

All possible outcomes (Sample Space)

Events (subset of Sample space)

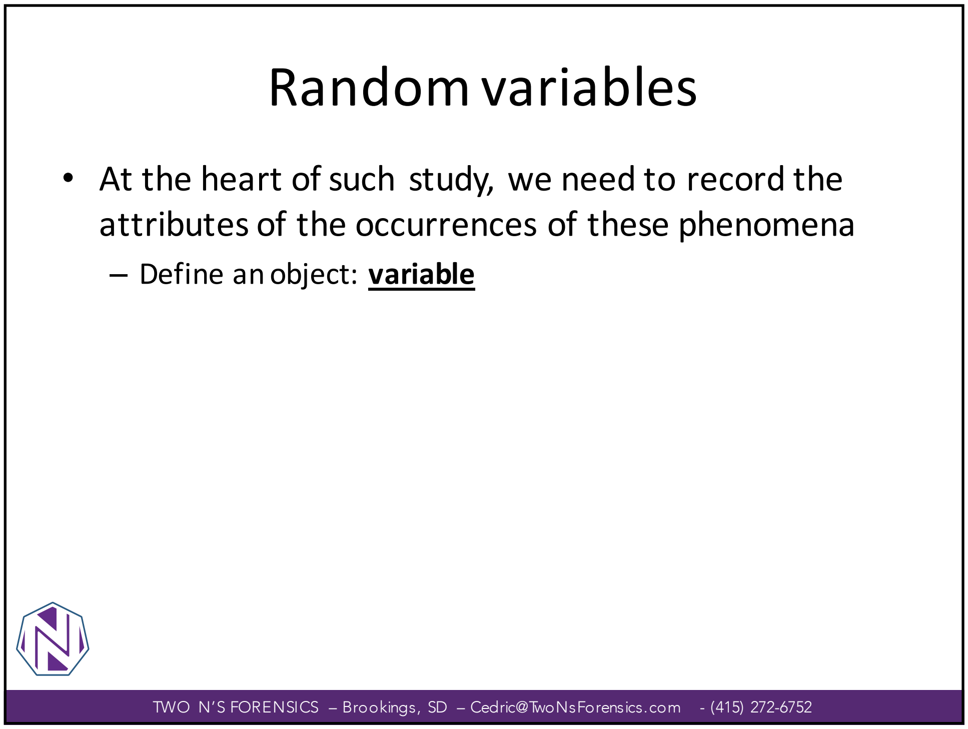
One or more occurrence(s) of phenomenon

Prediction

Observation

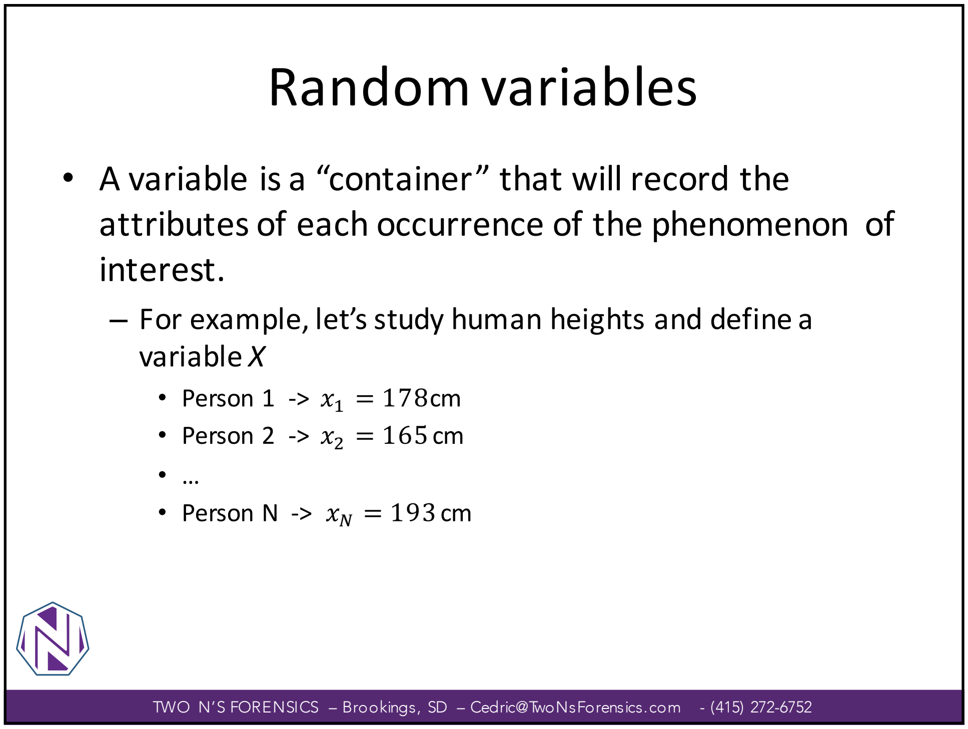
Random variable

Measurement



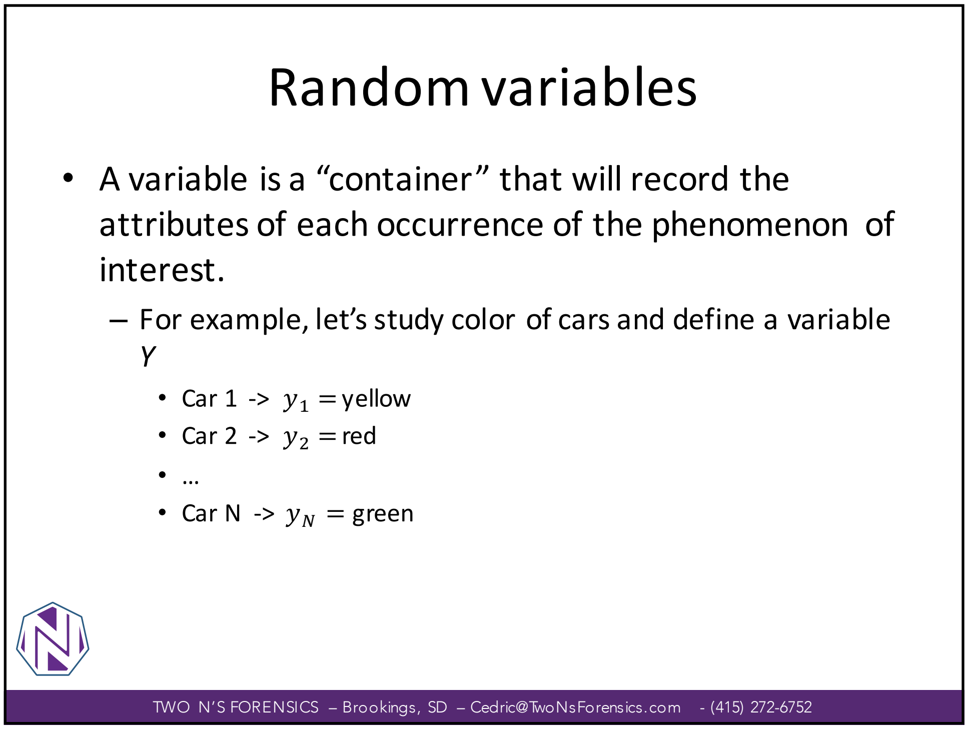
Random variables

- At the heart of such study, we need to record the attributes of the occurrences of these phenomena
  - Define an object: **variable**



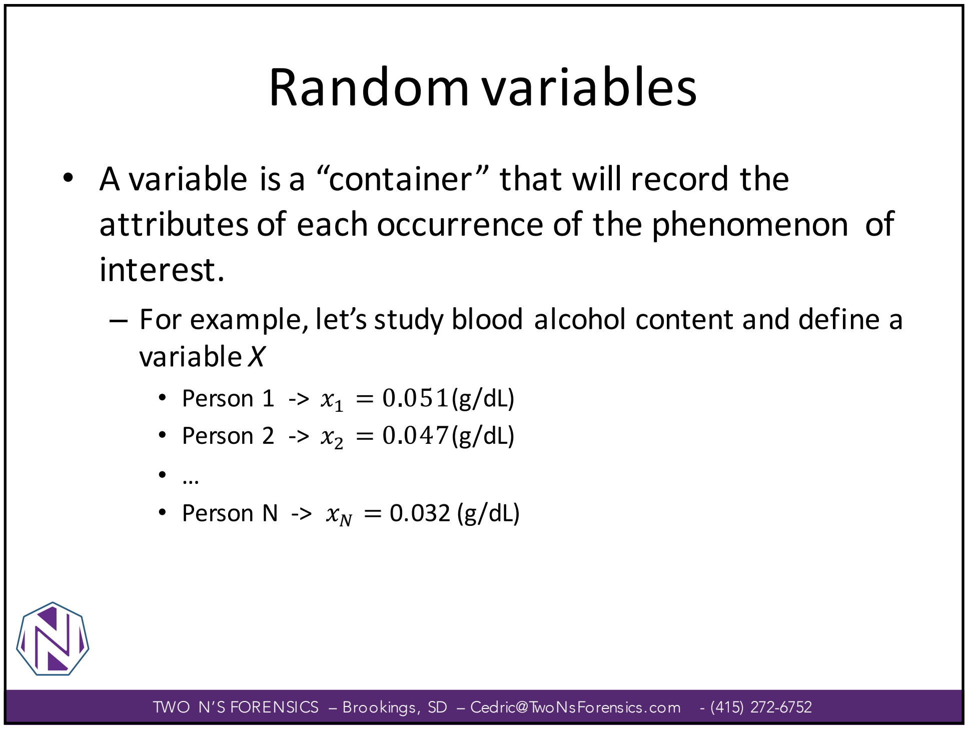
Random variables

- A variable is a “container” that will record the attributes of each occurrence of the phenomenon of interest.
  - For example, let’s study human heights and define a variable *X*
    - Person 1 -> $x_1 = 178$cm
    - Person 2 -> $x_2 = 165$ cm
    - …
    - Person N -> $x_N = 193$ cm



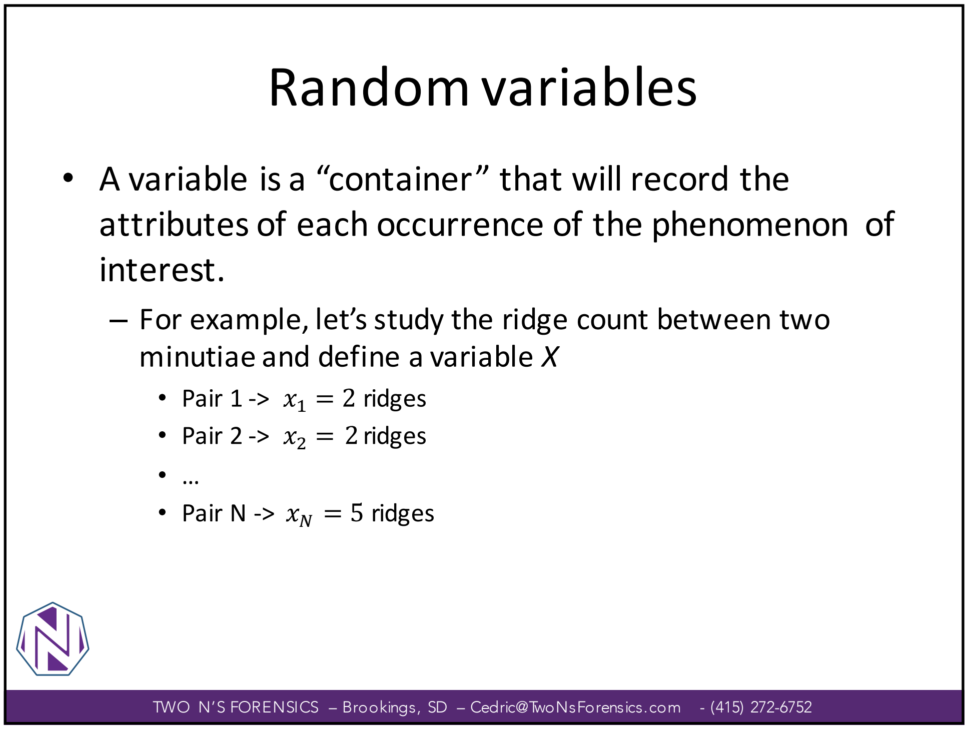
Random variables

- A variable is a “container” that will record the attributes of each occurrence of the phenomenon of interest.
  - For example, let’s study color of cars and define a variable *Y*
    - Car 1 -> $y_1 =$ yellow
    - Car 2 -> $y_2 =$ red
    - …
    - Car N -> $y_N =$ green



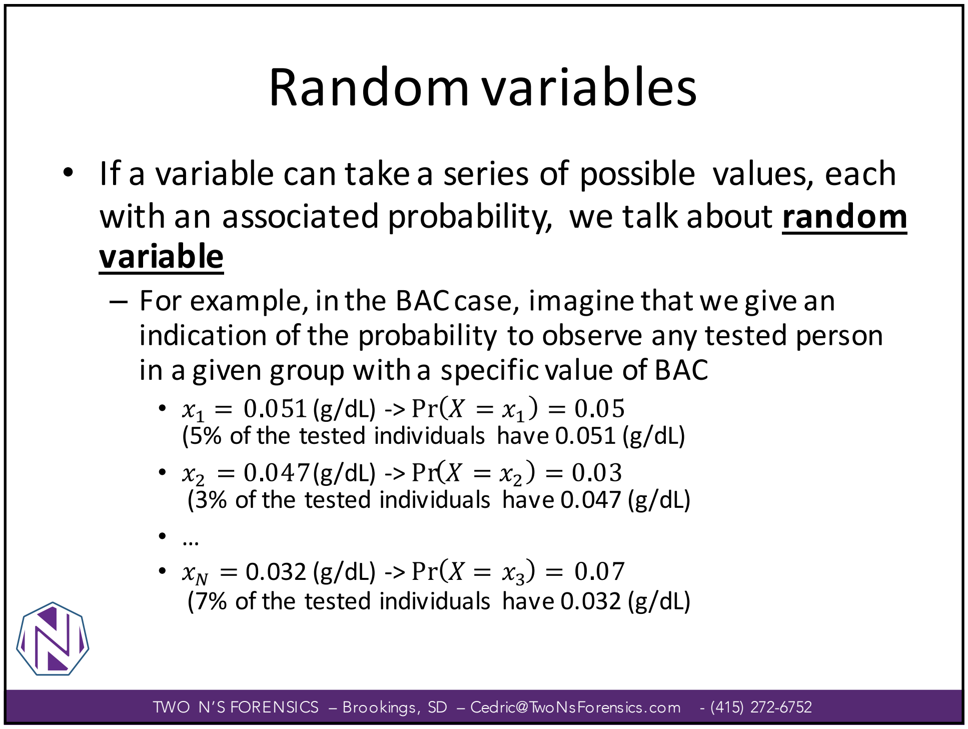
Random variables

- A variable is a “container” that will record the attributes of each occurrence of the phenomenon of interest.
  - For example, let’s study blood alcohol content and define a variable *X*
    - Person 1 -> $x_1 = 0.051$(g/dL)
    - Person 2 -> $x_2 = 0.047$(g/dL)
    - …
    - Person N -> $x_N = 0.032$ (g/dL)



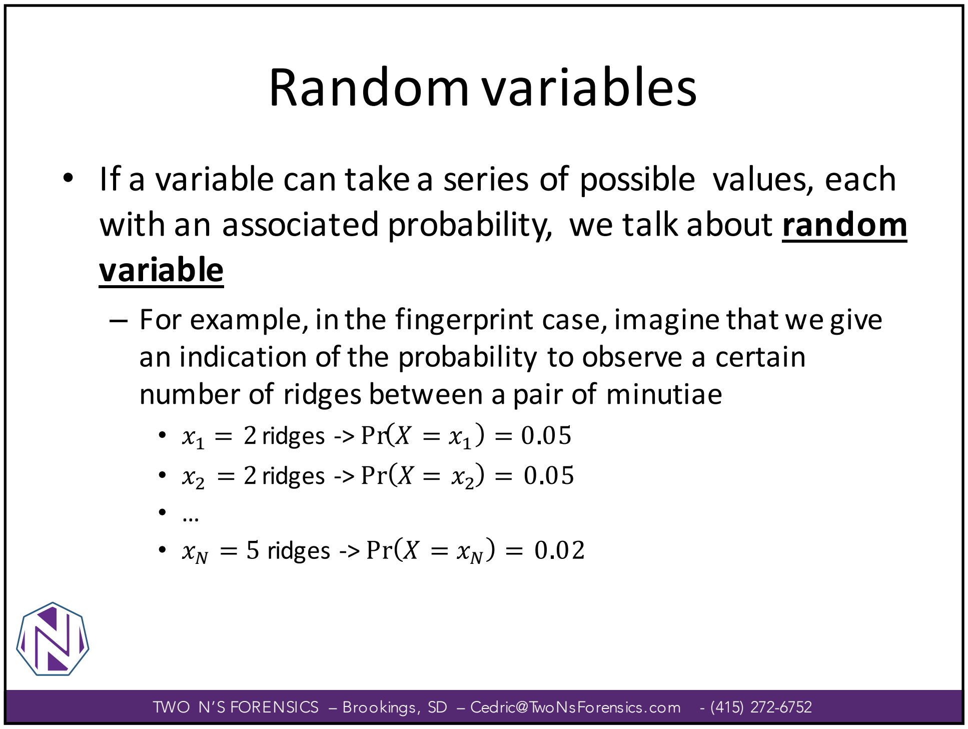
Random variables

- A variable is a “container” that will record the attributes of each occurrence of the phenomenon of interest.
  - For example, let’s study the ridge count between two minutiae and define a variable *X*
    - Pair 1 -> $x_1 = 2$ ridges
    - Pair 2 -> $x_2 = 2$ ridges
    - …
    - Pair N -> $x_N = 5$ ridges



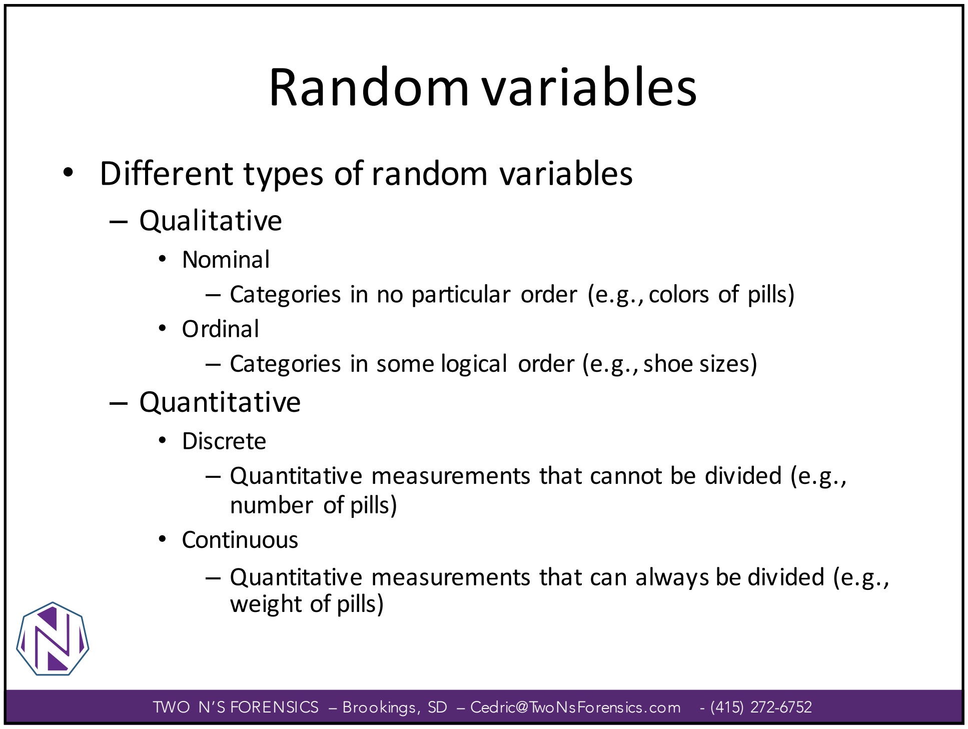
Random variables

- If a variable can take a series of possible values, each with an associated probability, we talk about **random variable**
  - For example, in the BAC case, imagine that we give an indication of the probability to observe any tested person in a given group with a specific value of BAC
    - $x_1 = 0.051$ (g/dL) -> $\Pr(X = x_1) = 0.05$
      (5% of the tested individuals have 0.051 (g/dL)
    - $x_2 = 0.047$(g/dL) -> $\Pr(X = x_2) = 0.03$
      (3% of the tested individuals have 0.047 (g/dL)
    - …
    - $x_N = 0.032$ (g/dL) -> $\Pr(X = x_3) = 0.07$
      (7% of the tested individuals have 0.032 (g/dL)



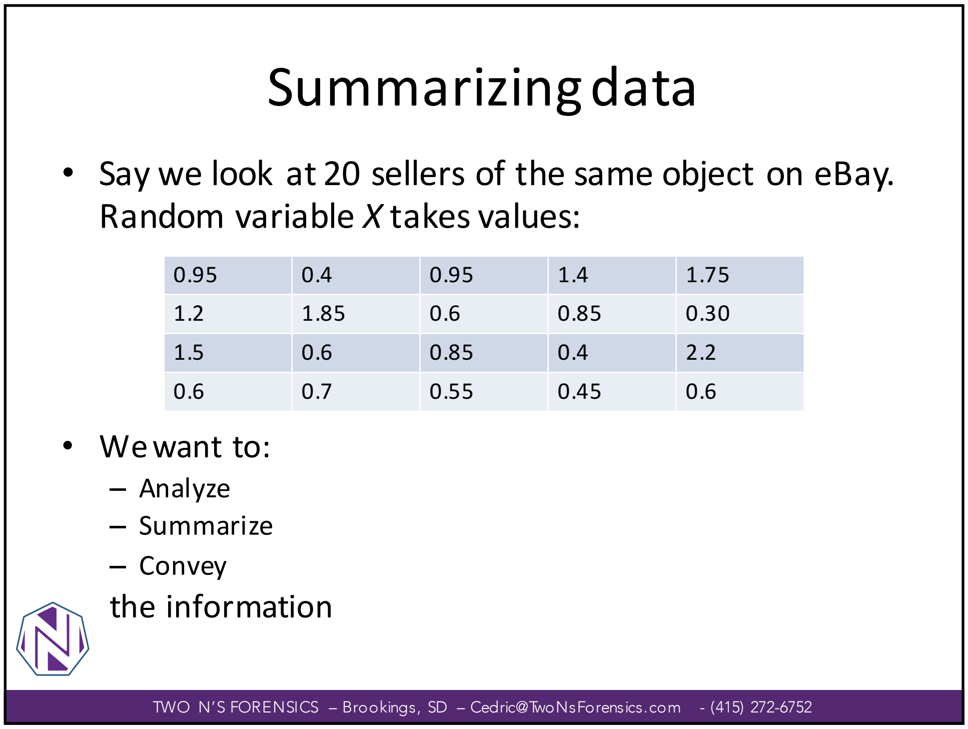
Random variables

- If a variable can take a series of possible values, each with an associated probability, we talk about **random variable**
  - For example, in the fingerprint case, imagine that we give an indication of the probability to observe a certain number of ridges between a pair of minutiae
    - $x_1 = 2$ ridges -> $\Pr(X = x_1) = 0.05$
    - $x_2 = 2$ ridges -> $\Pr(X = x_2) = 0.05$
    - …
    - $x_N = 5$ ridges -> $\Pr(X = x_N) = 0.02$



Random variables

- Different types of random variables
  - Qualitative
    - Nominal
      - Categories in no particular order (e.g., colors of pills)
    - Ordinal
      - Categories in some logical order (e.g., shoe sizes)
  - Quantitative
    - Discrete
      - Quantitative measurements that cannot be divided (e.g., number of pills)
    - Continuous
      - Quantitative measurements that can always be divided (e.g., weight of pills)



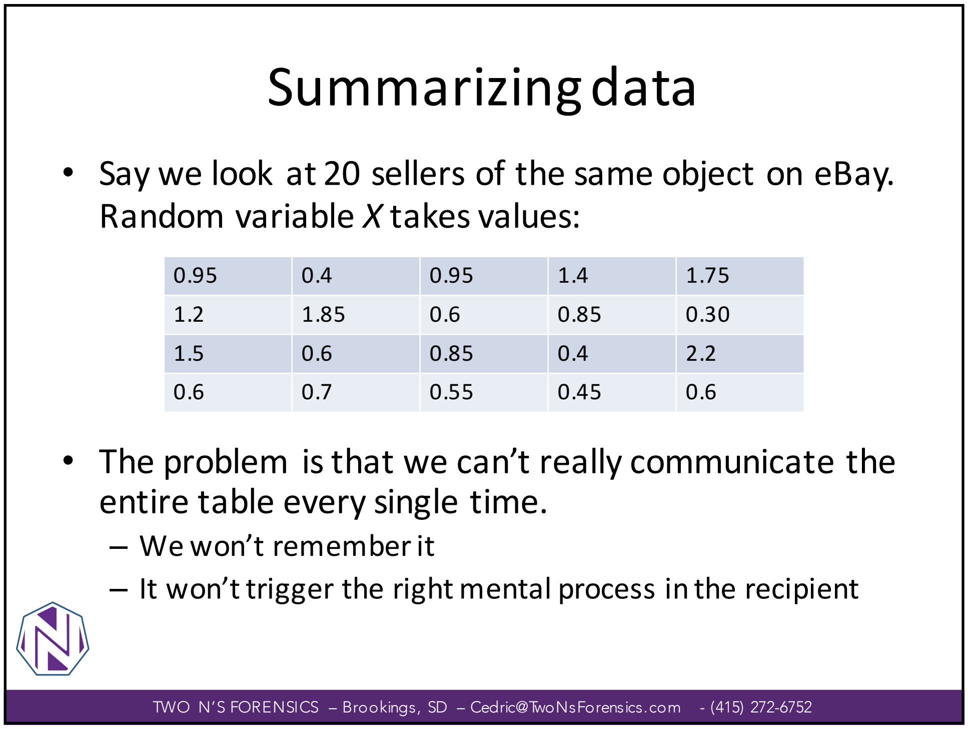
Summarizing data

- Say we look at 20 sellers of the same object on eBay. Random variable *X* takes values:

0.95	0.4	0.95	1.4	1.75
1.2	1.85	0.6	0.85	0.30
1.5	0.6	0.85	0.4	2.2
0.6	0.7	0.55	0.45	0.6

- We want to:
  - Analyze
  - Summarize
  - Convey

  the information

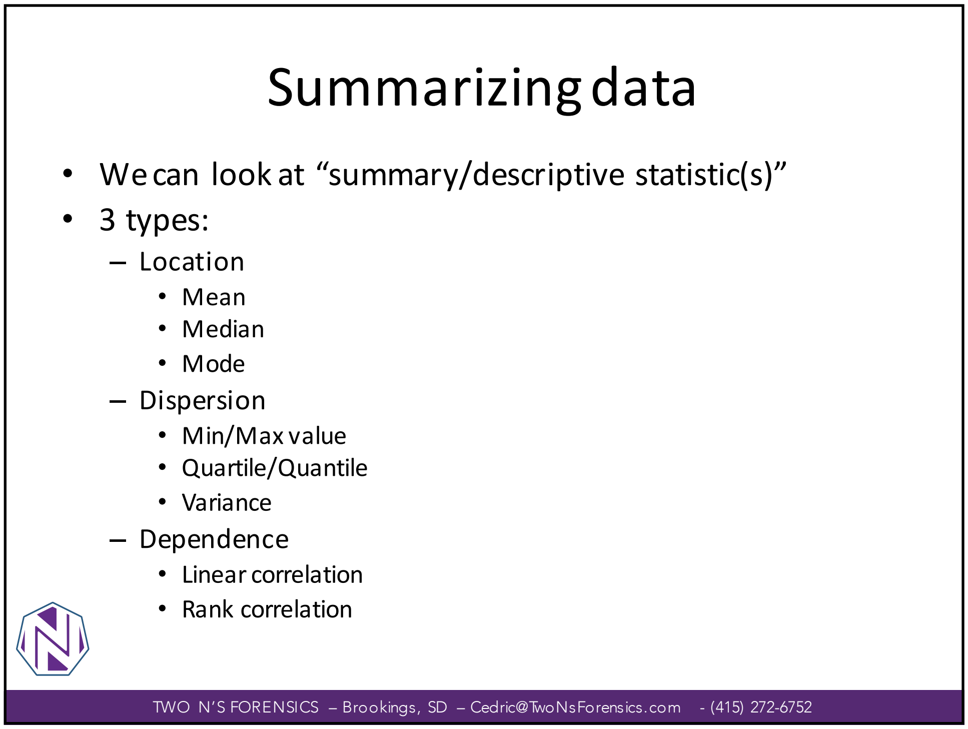


Summarizing data

- Say we look at 20 sellers of the same object on eBay. Random variable *X* takes values:

0.95	0.4	0.95	1.4	1.75
1.2	1.85	0.6	0.85	0.30
1.5	0.6	0.85	0.4	2.2
0.6	0.7	0.55	0.45	0.6

- The problem is that we can't really communicate the entire table every single time.
  - We won't remember it
  - It won't trigger the right mental process in the recipient



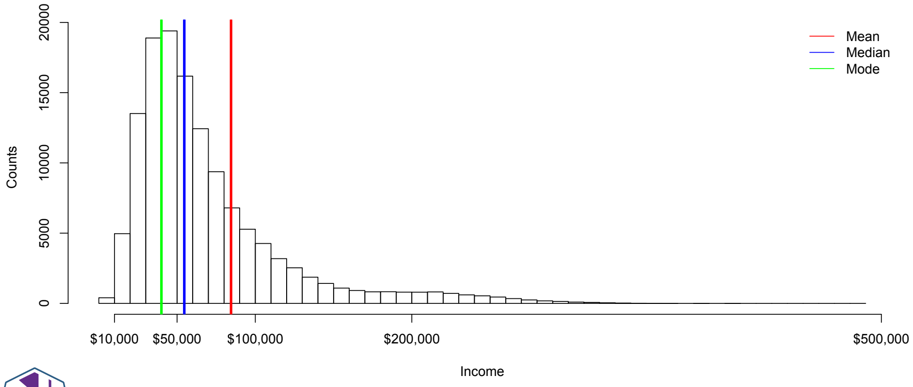
Summarizing data

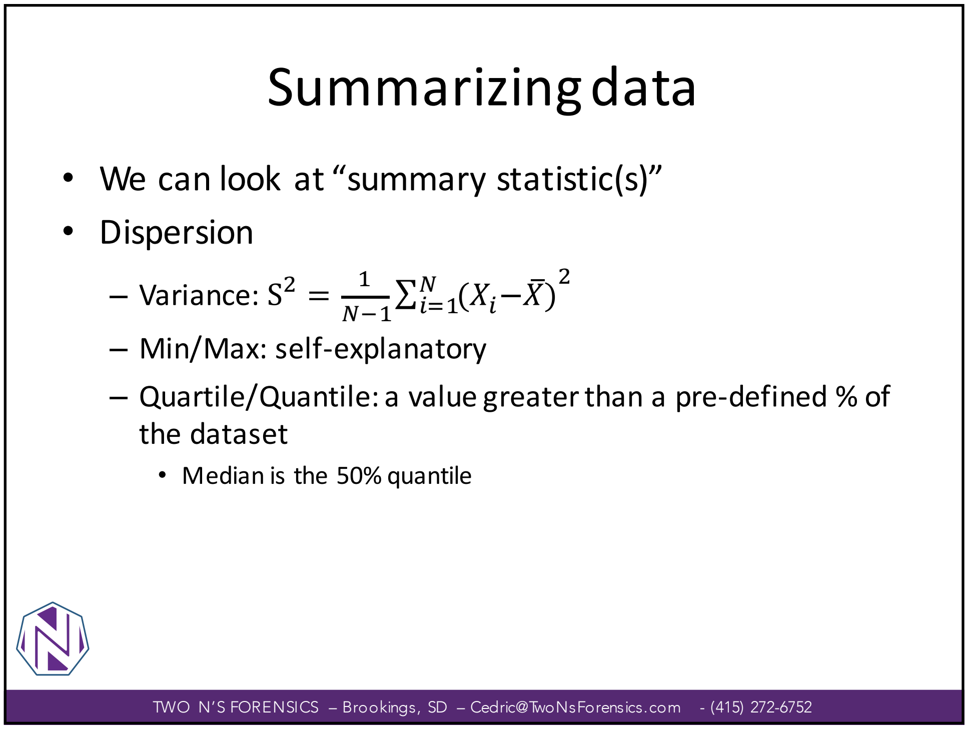
- We can look at "summary/descriptive statistic(s)"
- 3 types:
  - Location
    - Mean
    - Median
    - Mode
  - Dispersion
    - Min/Max value
    - Quartile/Quantile
    - Variance
  - Dependence
    - Linear correlation
    - Rank correlation

Summarizing data

- We can look at “summary statistic(s)”
- Location
  - Mean: $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$
  - Median: middle value
    - Sort all values from smallest to largest. Median is the middle one
  - Mode
    - Value of the random variable that appears most often in the dataset

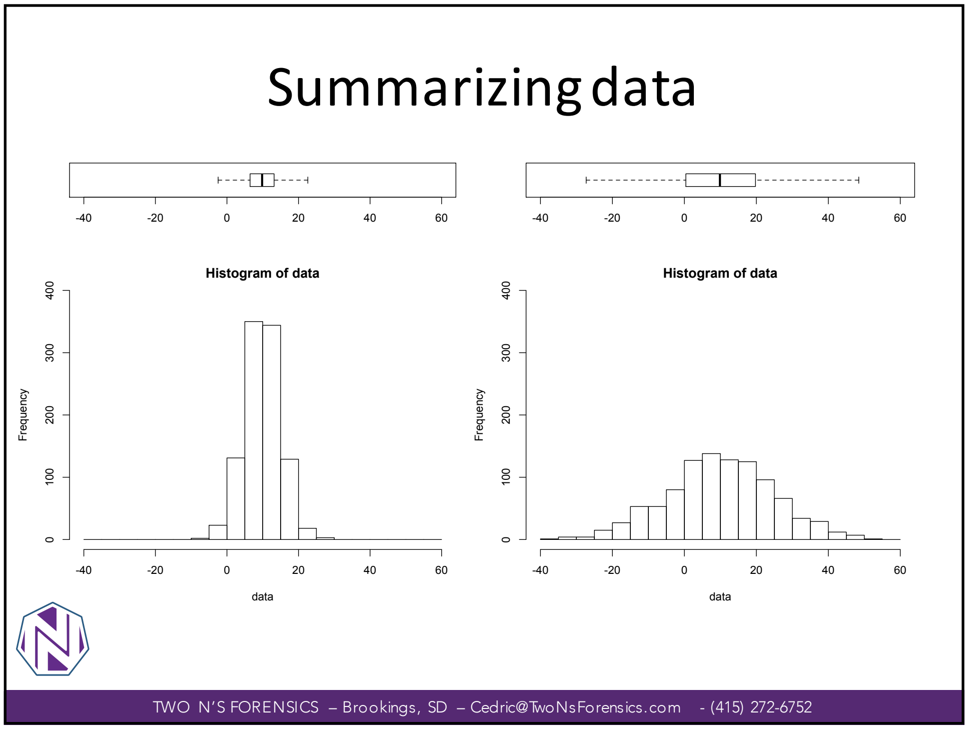
Summarizing data



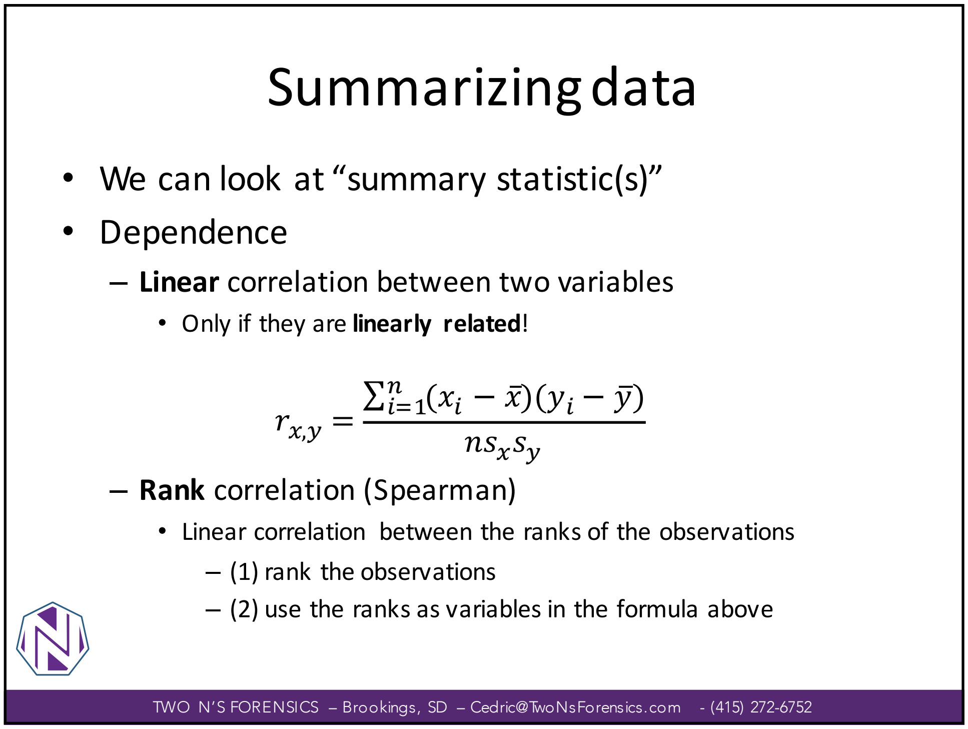


Summarizing data

- We can look at "summary statistic(s)"
- Dispersion
  - Variance: $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$
  - Min/Max: self-explanatory
  - Quartile/Quantile: a value greater than a pre-defined % of the dataset
    - Median is the 50% quantile



Summarizing data

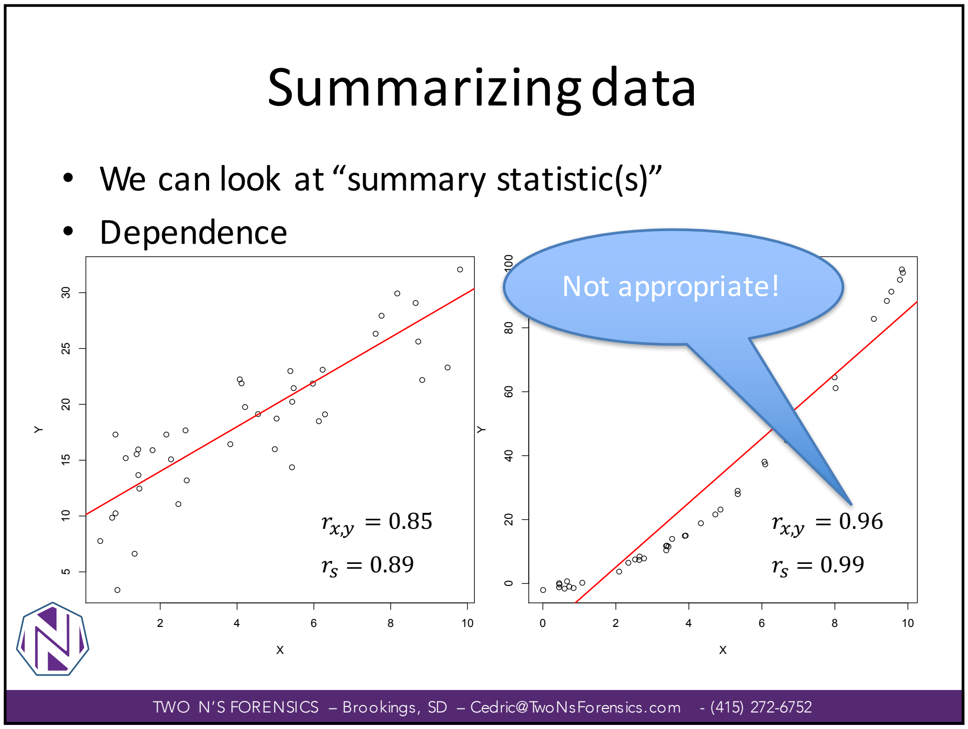


Summarizing data

- We can look at “summary statistic(s)”
- Dependence
  - **Linear** correlation between two variables
    - Only if they are **linearly related**!

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y}$$

  - **Rank** correlation (Spearman)
    - Linear correlation between the ranks of the observations
      - (1) rank the observations
      - (2) use the ranks as variables in the formula above



Summarizing data

- We can look at “summary statistic(s)”
- Dependence

$r_{x,y} = 0.85$

$r_s = 0.89$

Not appropriate!

$r_{x,y} = 0.96$

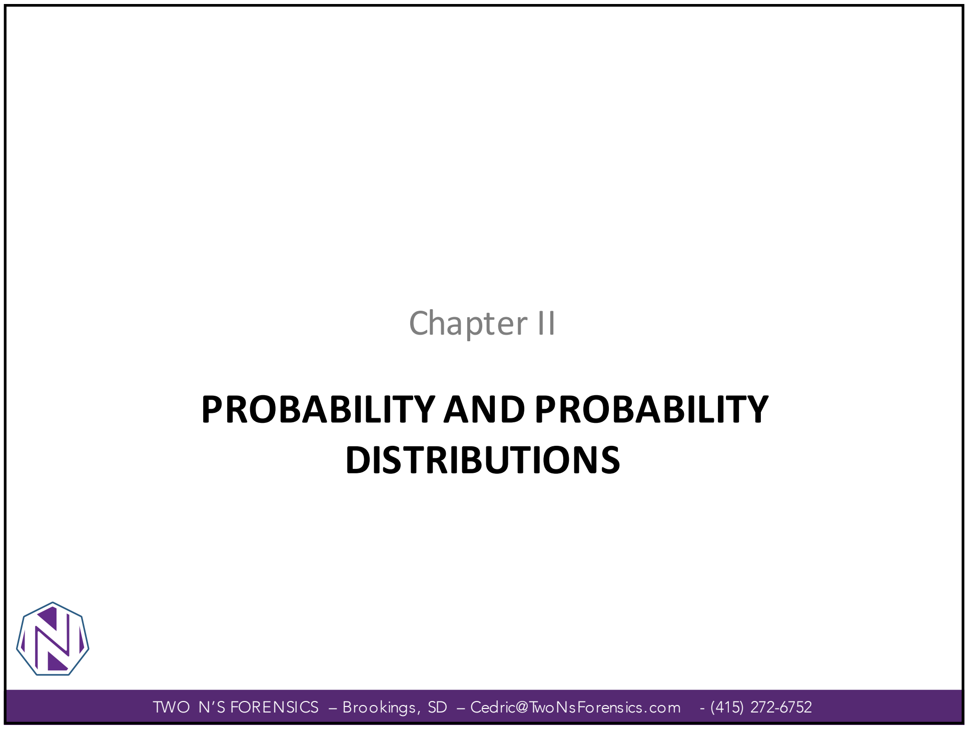
$r_s = 0.99$

Take home messages

- A random variable is a “container” that can store the results of a series of experiment
- We can express how often the random variable takes a certain value (i.e. how often the experiment results in a given observation) using a probability
- We can summarize these random variables using various summary statistics
  - But we lose information and we need to be careful

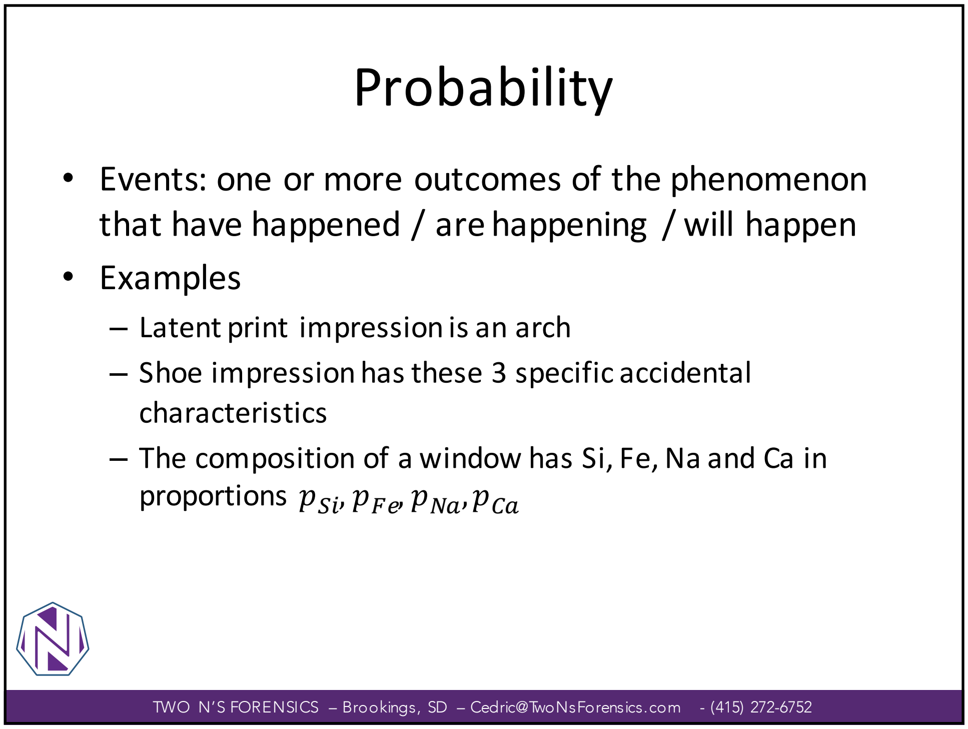
Chapter I

EXERCISES



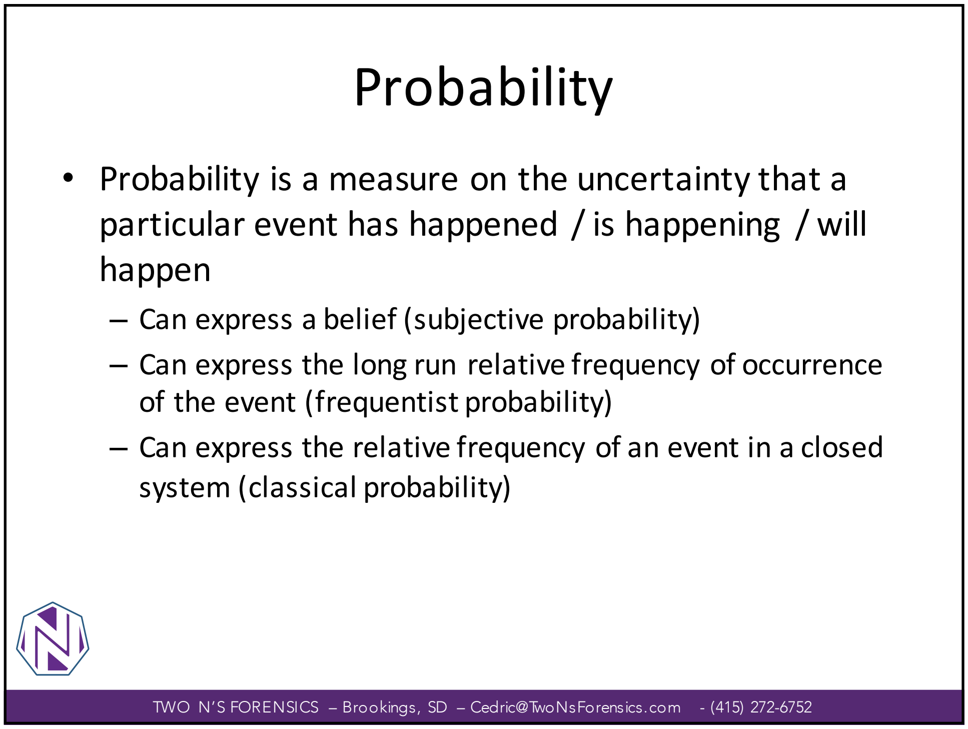
Chapter II

PROBABILITY AND PROBABILITY DISTRIBUTIONS



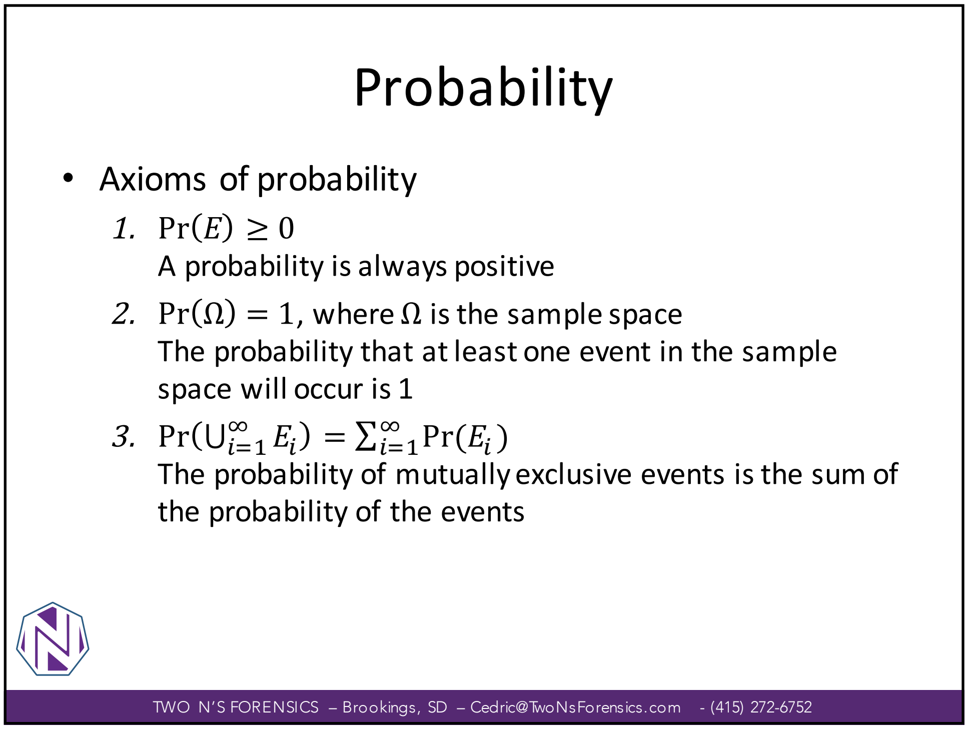
Probability

- Events: one or more outcomes of the phenomenon that have happened / are happening / will happen
- Examples
  - Latent print impression is an arch
  - Shoe impression has these 3 specific accidental characteristics
  - The composition of a window has Si, Fe, Na and Ca in proportions $p_{Si}, p_{Fe}, p_{Na}, p_{Ca}$



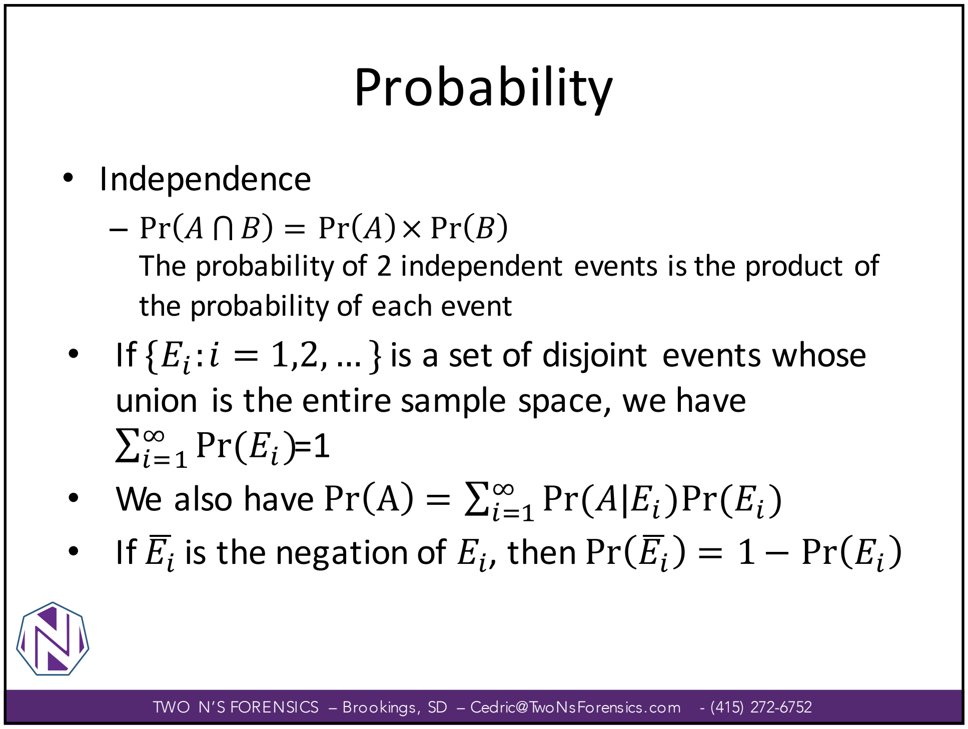
Probability

- Probability is a measure on the uncertainty that a particular event has happened / is happening / will happen
  - Can express a belief (subjective probability)
  - Can express the long run relative frequency of occurrence of the event (frequentist probability)
  - Can express the relative frequency of an event in a closed system (classical probability)



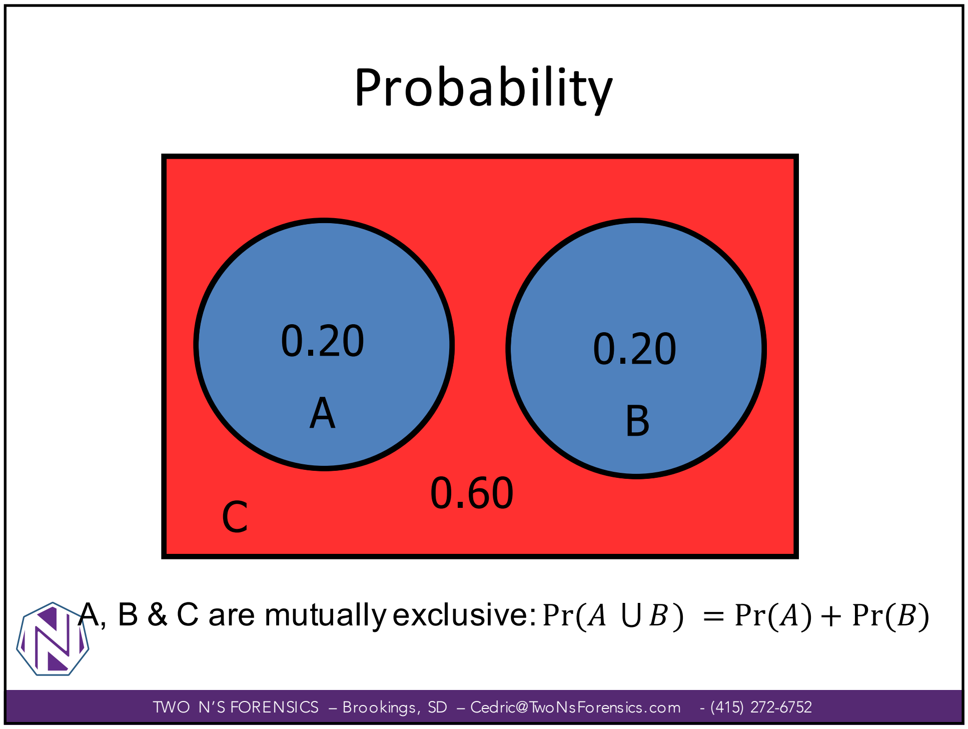
Probability

- Axioms of probability
  1. $\Pr(E) \geq 0$
     A probability is always positive
  2. $\Pr(\Omega) = 1$, where Ω is the sample space
     The probability that at least one event in the sample space will occur is 1
  3. $\Pr(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \Pr(E_i)$
     The probability of mutually exclusive events is the sum of the probability of the events



Probability

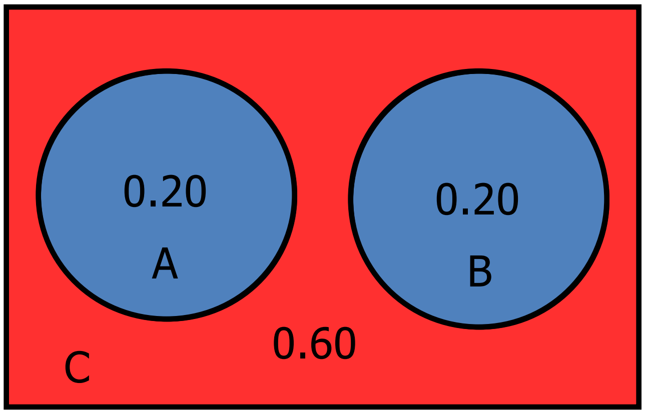
- Independence
  - $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$
    The probability of 2 independent events is the product of the probability of each event
- If $\{E_i : i = 1,2,\dots\}$ is a set of disjoint events whose union is the entire sample space, we have $\sum_{i=1}^{\infty} \Pr(E_i)$=1
- We also have $\Pr(\mathrm{A}) = \sum_{i=1}^{\infty} \Pr(A|E_i)\Pr(E_i)$
- If $\bar{E}_i$ is the negation of E_i, then $\Pr(\bar{E}_i) = 1 - \Pr(E_i)$



Probability

A, B & C are mutually exclusive: $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

Probability

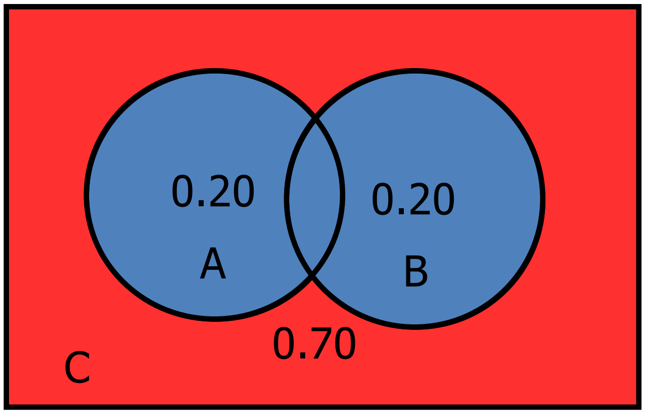


A, B & C are mutually exclusive

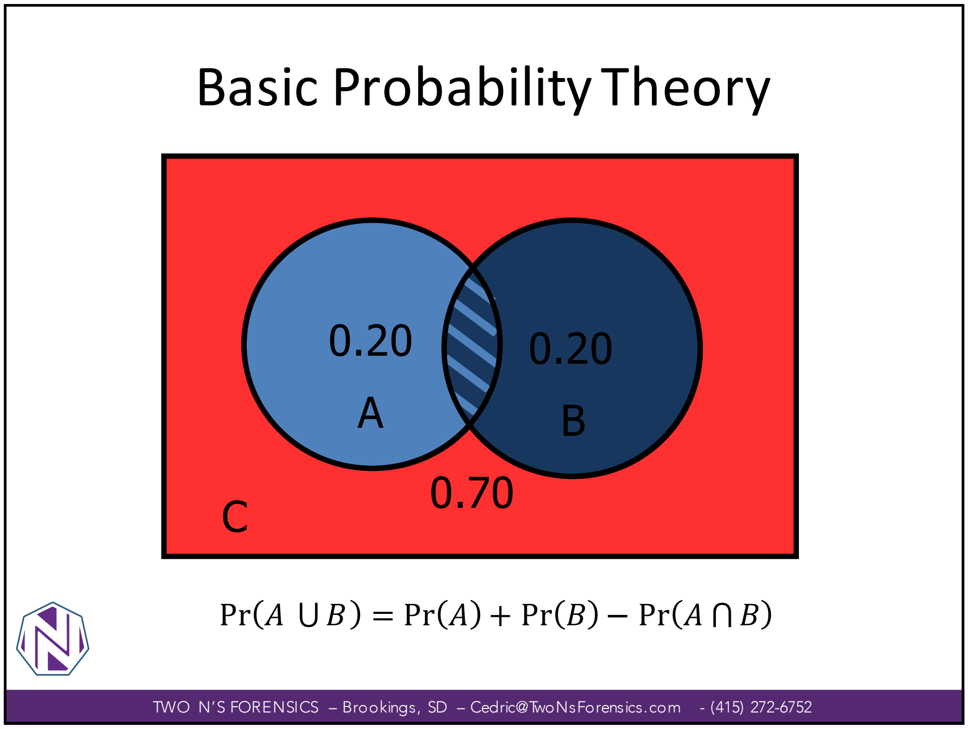
$\Pr(A \cup B \cup C) = 1$

$\Pr(A \cup B) = 1 - \Pr(C)$ or $\Pr(A) = 1 - \Pr(C \cup B)$

Basic Probability Theory

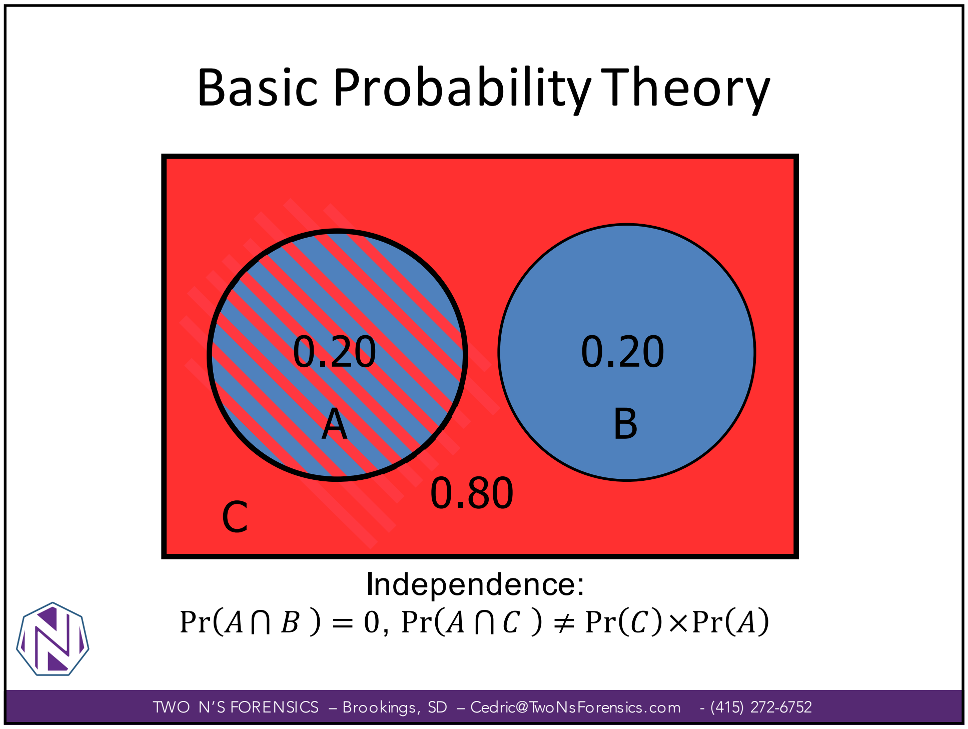


A & B are not mutually exclusive: $\Pr(A \cup B) \neq \Pr(A) + \Pr(B)$



Basic Probability Theory

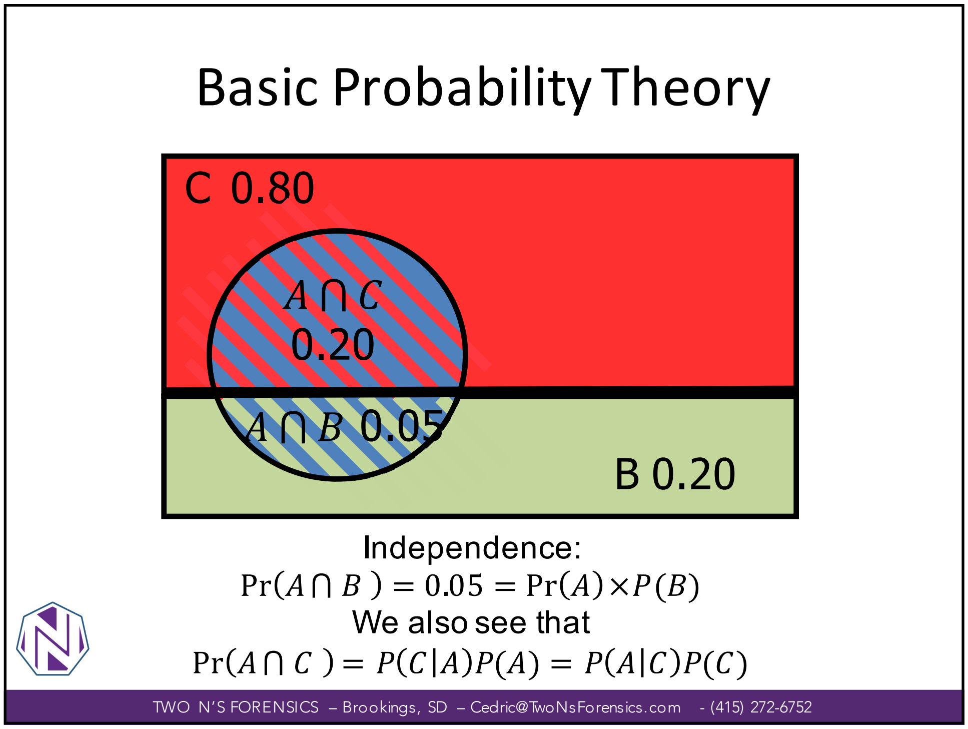
$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$



Basic Probability Theory

Independence:

$$\Pr(A \cap B) = 0,\ \Pr(A \cap C) \neq \Pr(C) \times \Pr(A)$$



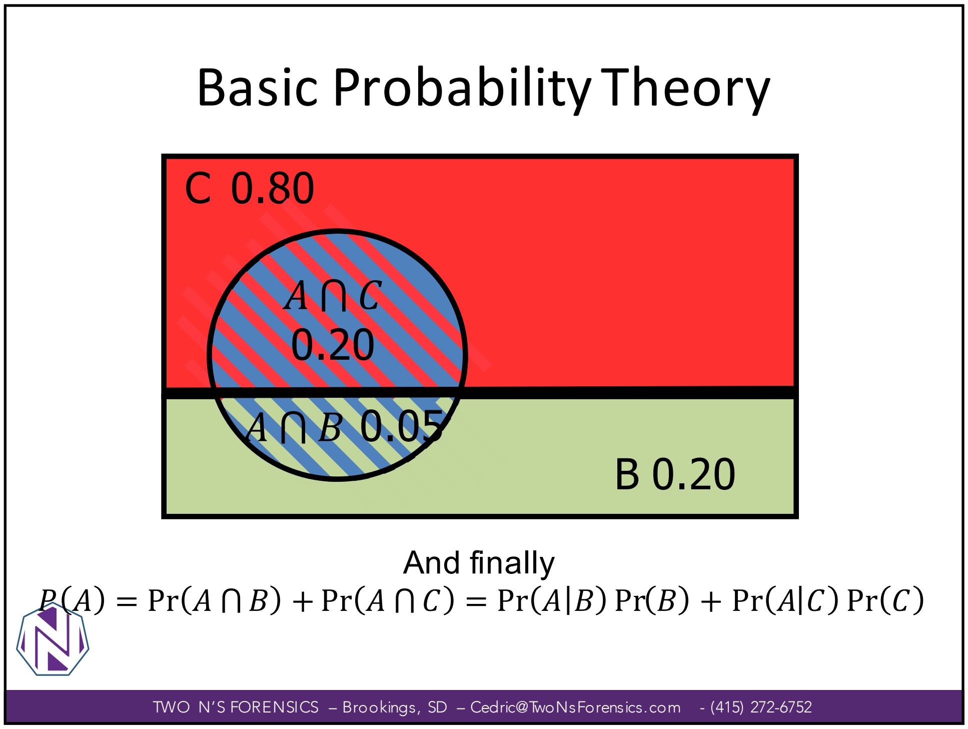
Basic Probability Theory

Independence:

$\Pr(A \cap B\ ) = 0.05 = \Pr(A) \times P(B)$

We also see that

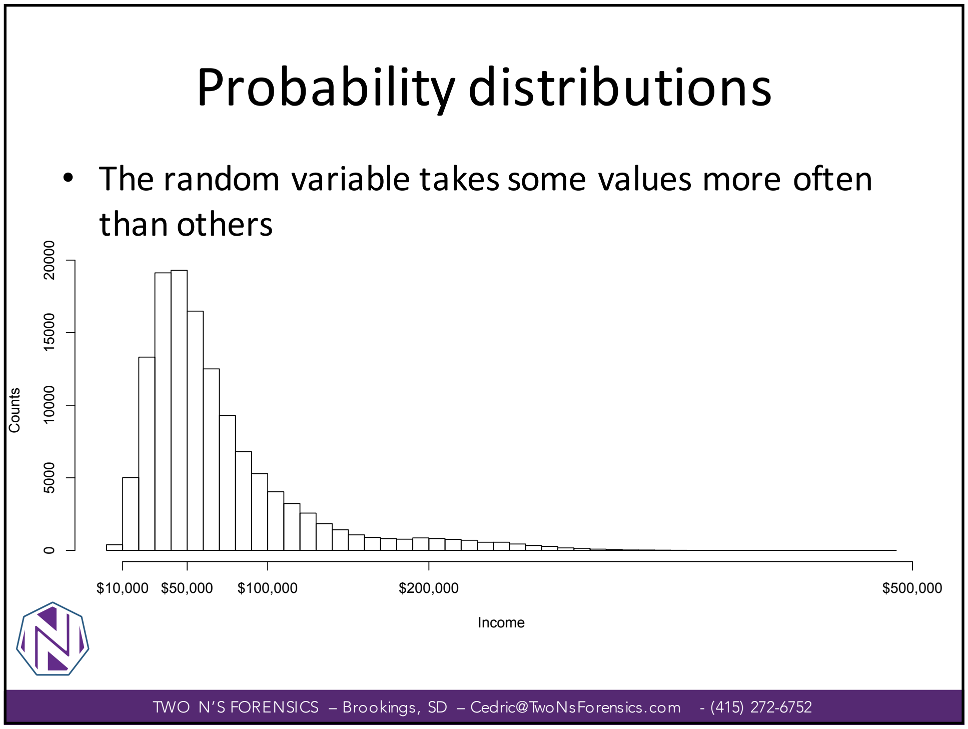
$\Pr(A \cap C\ ) = P(C|A)P(A) = P(A|C)P(C)$



Basic Probability Theory

And finally

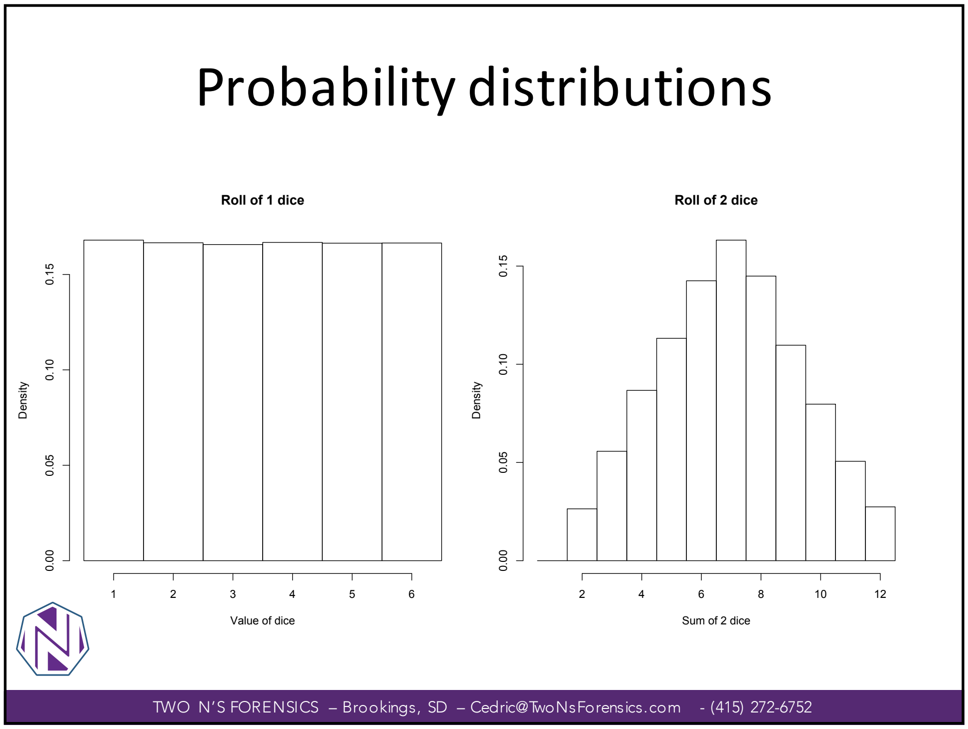
$P(A) = \Pr(A \cap B) + \Pr(A \cap C) = \Pr(A|B)\Pr(B) + \Pr(A|C)\Pr(C)$



Probability distributions

- The random variable takes some values more often than others

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Probability distributions

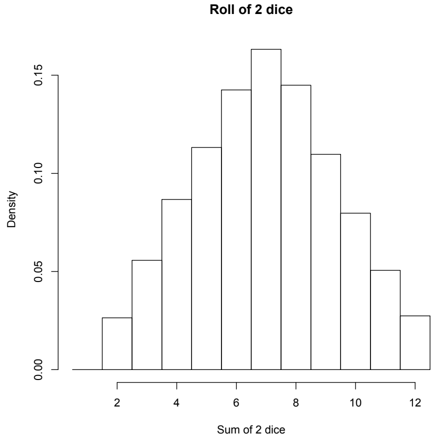
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

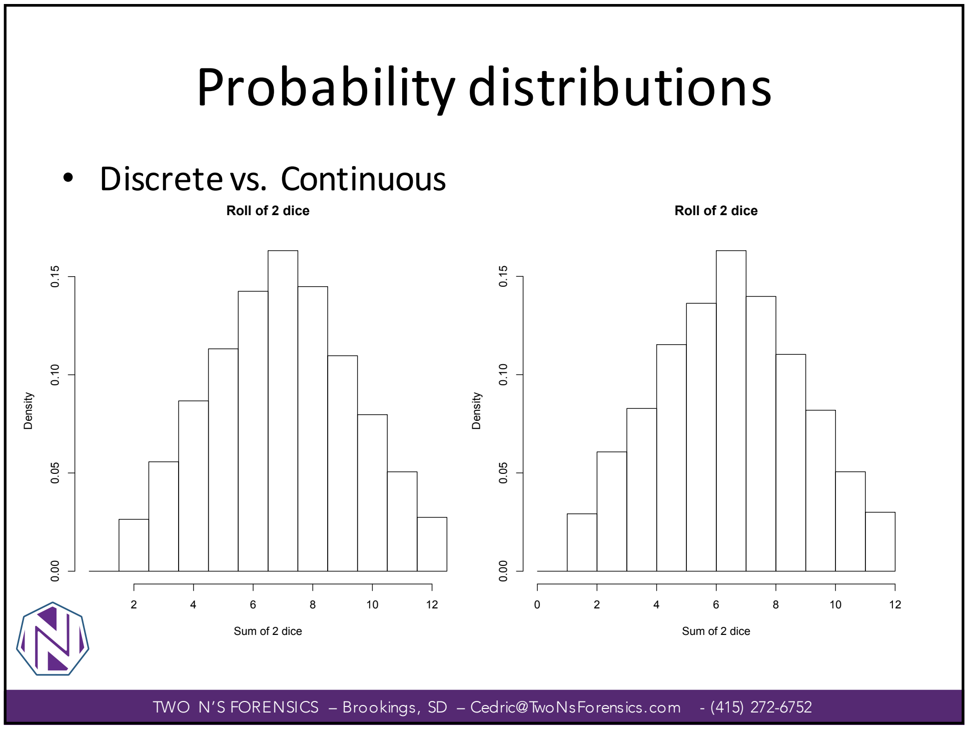
Probability distributions

- Different families of distributions
  - Based on:
    - The type of data
      - Discrete vs. continuous
    - The experiment that gave rise to the data
      - First observation
      - Number of successes
      - Natural (observational) experiment
      - …
    - **The presence of negative values**
      - Many probability distributions do not handle negative values
      - Conversely, if you are guaranteed to not have negative values, you cannot use some distributions

Probability distributions

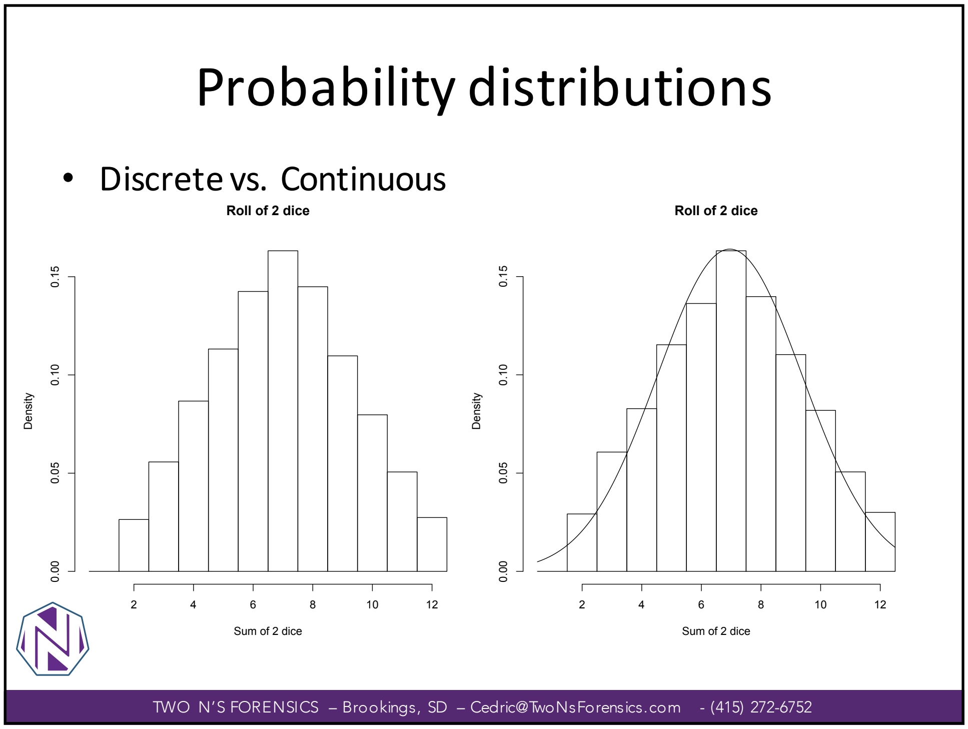
- Discrete vs. Continuous





Probability distributions

- Discrete vs. Continuous

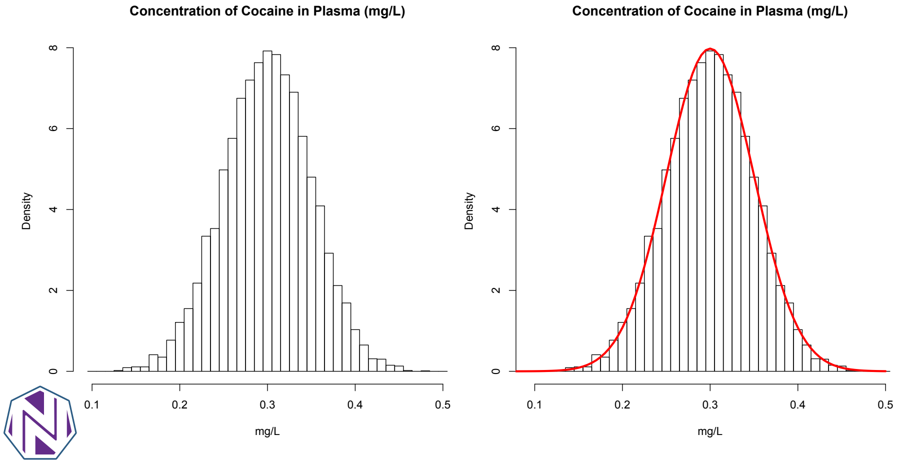


Probability distributions

- Discrete vs. Continuous

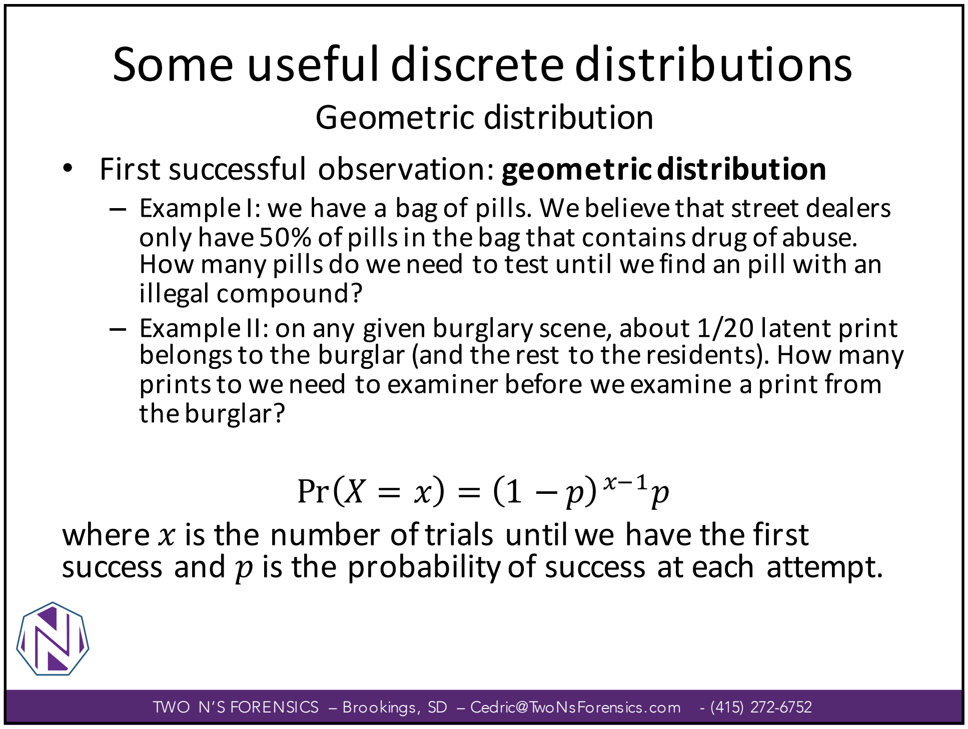
Probability distributions

- Discrete vs. Continuous



Probability distributions

- Each probability distribution is governed by a set of **parameters**
  - That we will assume to be known in this chapter
  - We will see how we estimate them next chapter



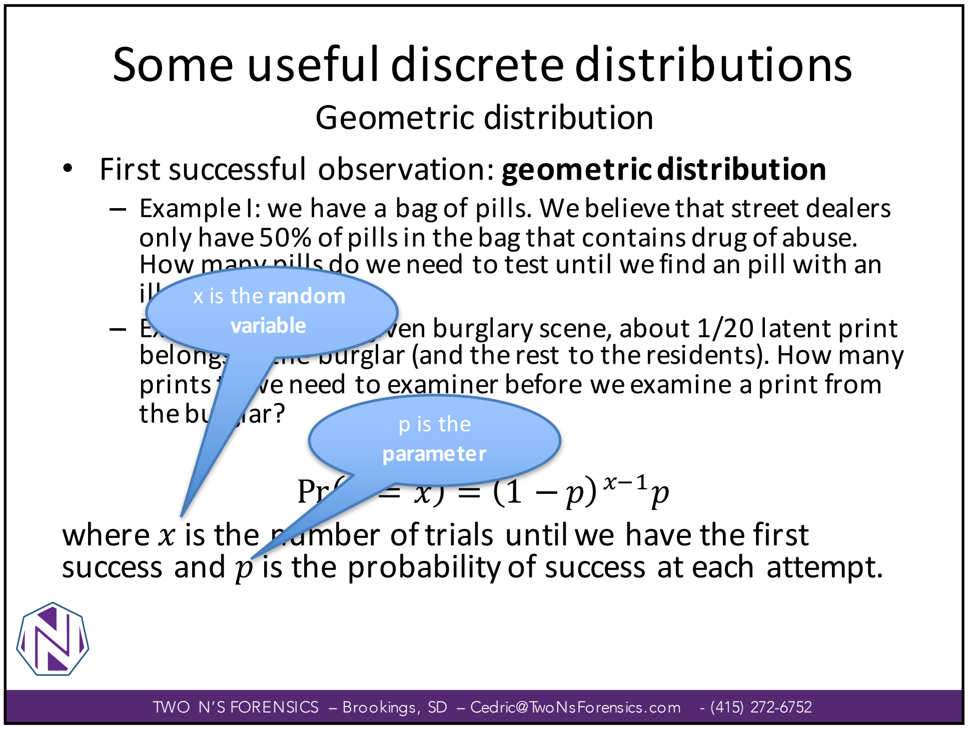
Some useful discrete distributions

Geometric distribution

- First successful observation: **geometric distribution**
  - Example I: we have a bag of pills. We believe that street dealers only have 50% of pills in the bag that contains drug of abuse. How many pills do we need to test until we find an pill with an illegal compound?
  - Example II: on any given burglary scene, about 1/20 latent print belongs to the burglar (and the rest to the residents). How many prints to we need to examiner before we examine a print from the burglar?

$$\Pr(X = x) = (1 - p)^{x-1} p$$

where x is the number of trials until we have the first success and p is the probability of success at each attempt.



Some useful discrete distributions

Geometric distribution

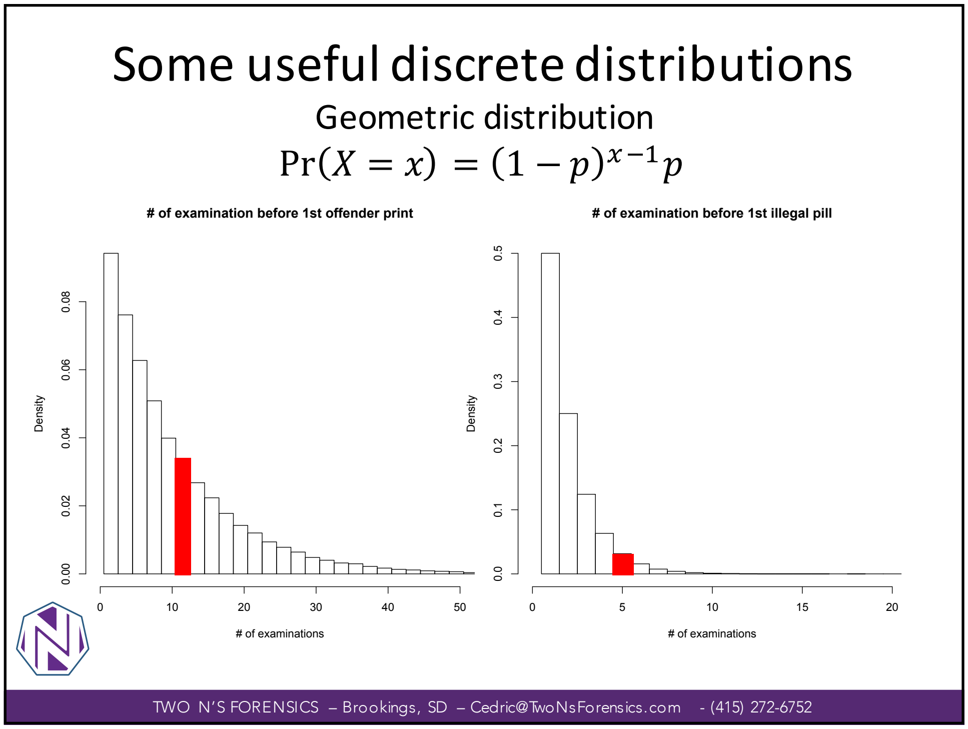
- First successful observation: **geometric distribution**
  - Example I: we have a bag of pills. We believe that street dealers only have 50% of pills in the bag that contains drug of abuse. How many pills do we need to test until we find an pill with an illegal compound?
  - Example II: on any given burglary scene, about 1/20 latent print belongs to the burglar (and the rest to the residents). How many prints to we need to examiner before we examine a print from the burglar?

$$\Pr(X = x) = (1 - p)^{x-1} p$$

where x is the number of trials until we have the first success and p is the probability of success at each attempt.

x is the **random variable**

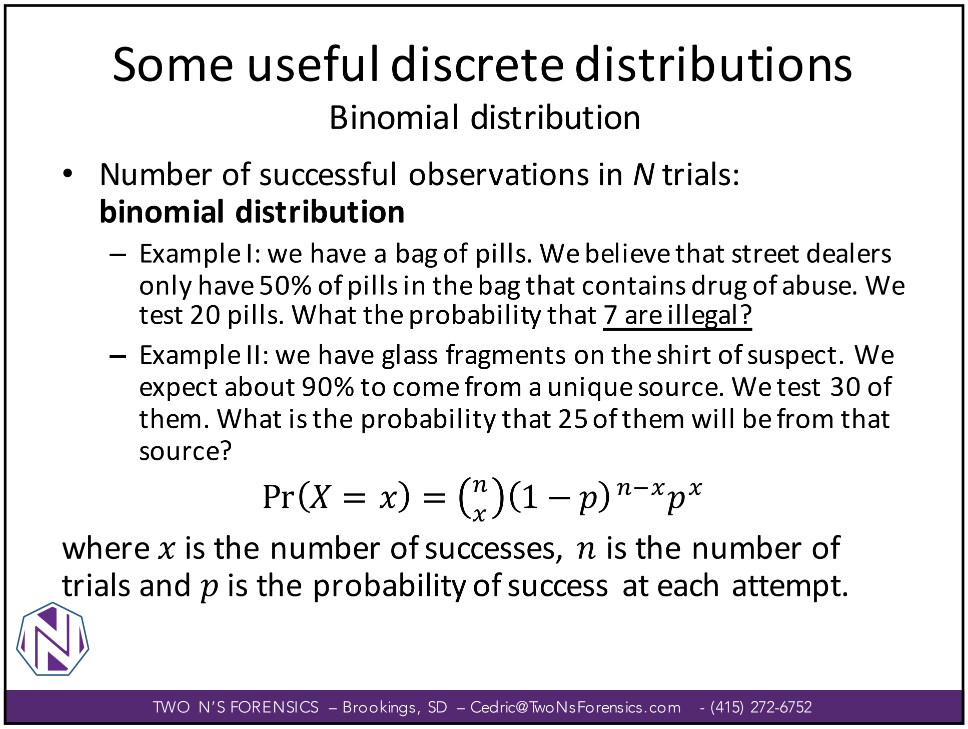
p is the **parameter**



Some useful discrete distributions

Geometric distribution

$$\Pr(X = x) = (1-p)^{x-1}p$$



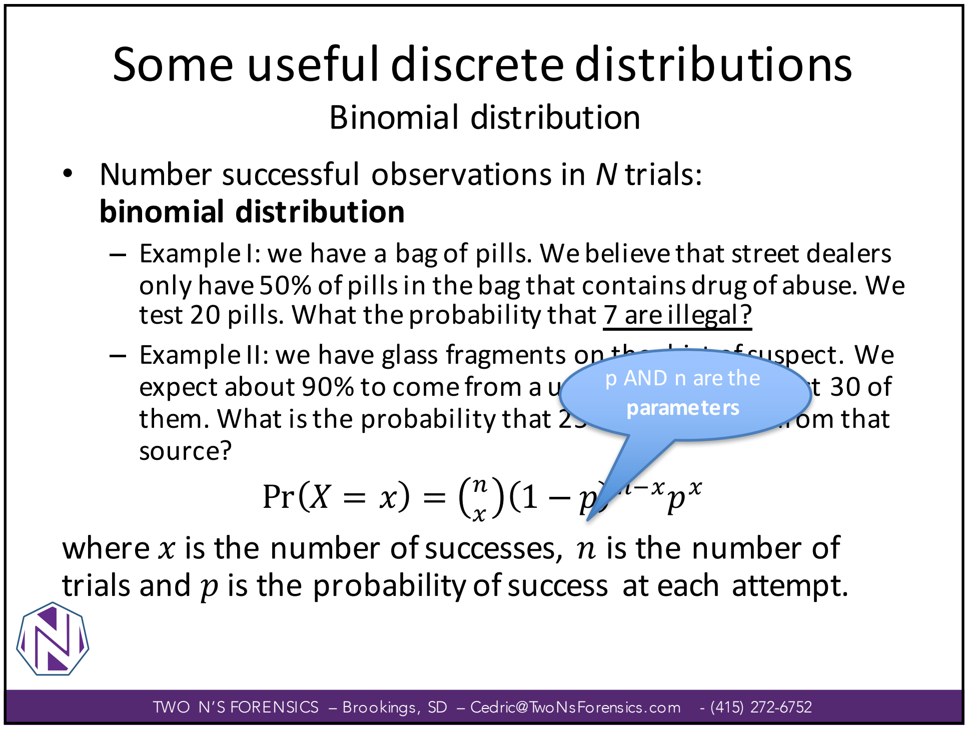
Some useful discrete distributions

Binomial distribution

- Number of successful observations in *N* trials: **binomial distribution**
  - Example I: we have a bag of pills. We believe that street dealers only have 50% of pills in the bag that contains drug of abuse. We test 20 pills. What the probability that 7 are illegal?
  - Example II: we have glass fragments on the shirt of suspect. We expect about 90% to come from a unique source. We test 30 of them. What is the probability that 25 of them will be from that source?

$$\Pr(X = x) = \binom{n}{x}(1-p)^{n-x}p^x$$

where x is the number of successes, n is the number of trials and p is the probability of success at each attempt.



Some useful discrete distributions

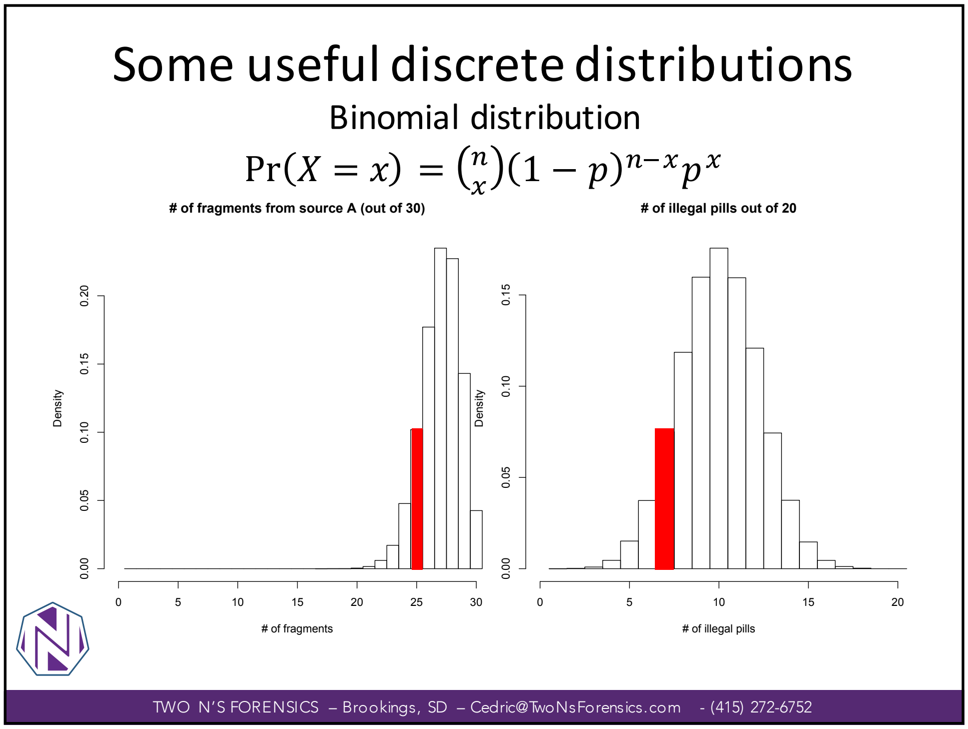
Binomial distribution

- Number successful observations in *N* trials: **binomial distribution**
  - Example I: we have a bag of pills. We believe that street dealers only have 50% of pills in the bag that contains drug of abuse. We test 20 pills. What the probability that 7 are illegal?
  - Example II: we have glass fragments on th suspect. We expect about 90% to come from a u t 30 of them. What is the probability that 2 rom that source?

p AND n are the **parameters**

$$\Pr(X = x) = \binom{n}{x}(1-p)^{n-x}p^x$$

where x is the number of successes, n is the number of trials and p is the probability of success at each attempt.



Some useful discrete distributions

Binomial distribution

$$\Pr(X = x) = \binom{n}{x}(1-p)^{n-x}p^x$$

Some useful discrete distributions

Hypergeometric distribution

- The issue with binomial distribution is that it assumes that we have an infinite amount of objects to sample from, and that we simply observe *N* of them
- Alternatively, it assumes that we are putting the object back in the pool before drawing another one
- When we have a finite sample, and that we do not want to replace the object back in the pool, we use the **hypergeometric distribution**

Some useful discrete distributions

Hypergeometric distribution

- Number successful observations in *n* draws (without replacement) from a finite population of size *N* that contains exactly *K* successes:
  hypergeometric distribution
  - Example I: we have a bag of 100 pills. We believe that street dealers only have 60 pills in the bag that contains drug of abuse. We test 20 pills. What the probability that 7 are illegal?
  - Example II: we have 50 glass fragments on the shirt of suspect. We expect about 45 to come from a unique source. We test 30 of them. What is the probability that 25 of them will be from that source?

$$\Pr(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

Some useful discrete distributions

Hypergeometric distribution

- Number successful observations in *n* draws (without replacement) from a finite population of size *N* that contains exactly *K* successes: **hypergeometric distribution**
  - Example I: we have a bag of 100 that street dealers only have 60 pills in the bag t use. We test 20 pills. What the probability that
  - Example II: we have 50 glass fragments he shirt of suspect. We expect about 45 to come from a unique rce. We test 30 of them. What is the probability that 25 of them w be from that source?

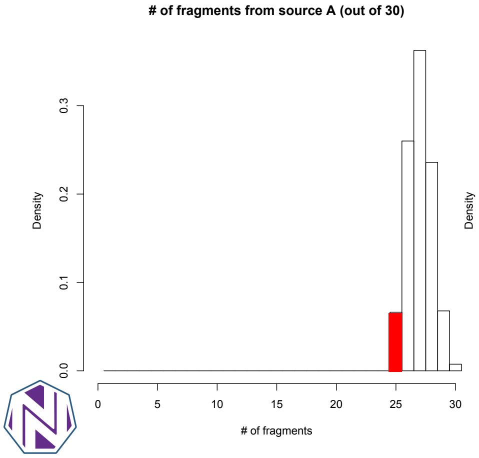
p, N and K are the **parameters**

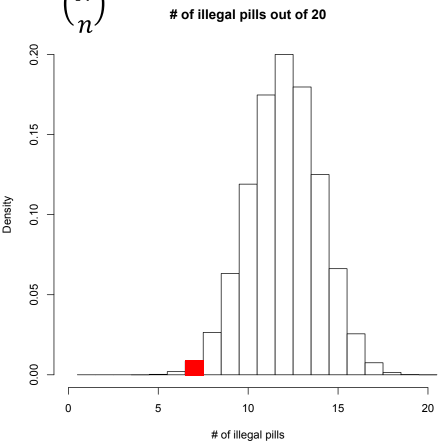
$$\Pr(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

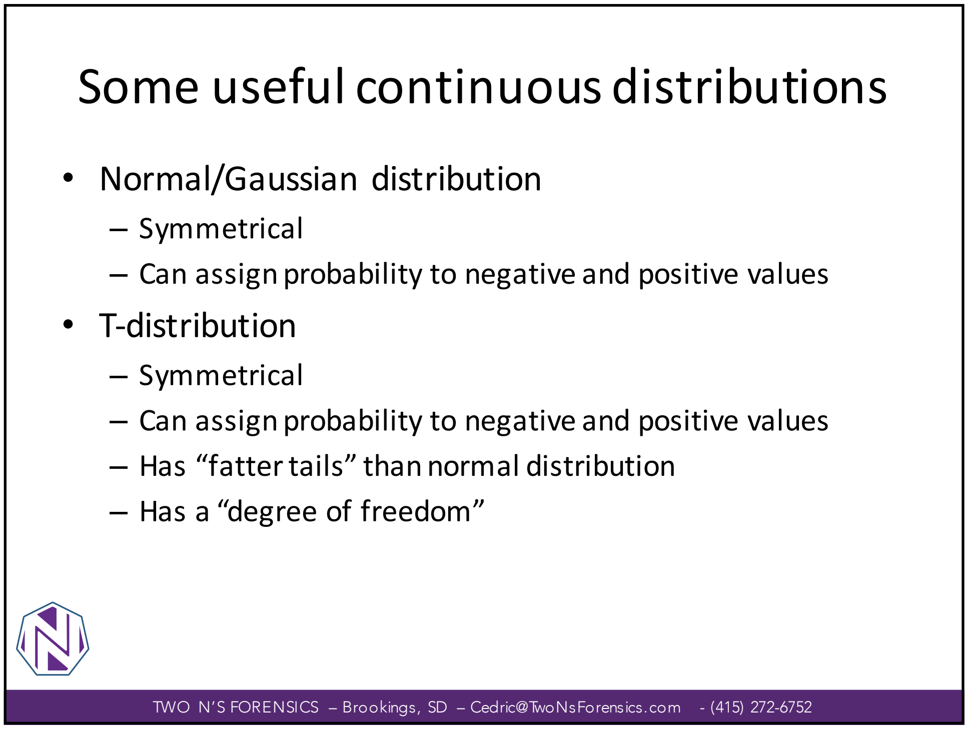
Some useful discrete distributions

Hypergeometric distribution

$$\Pr(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

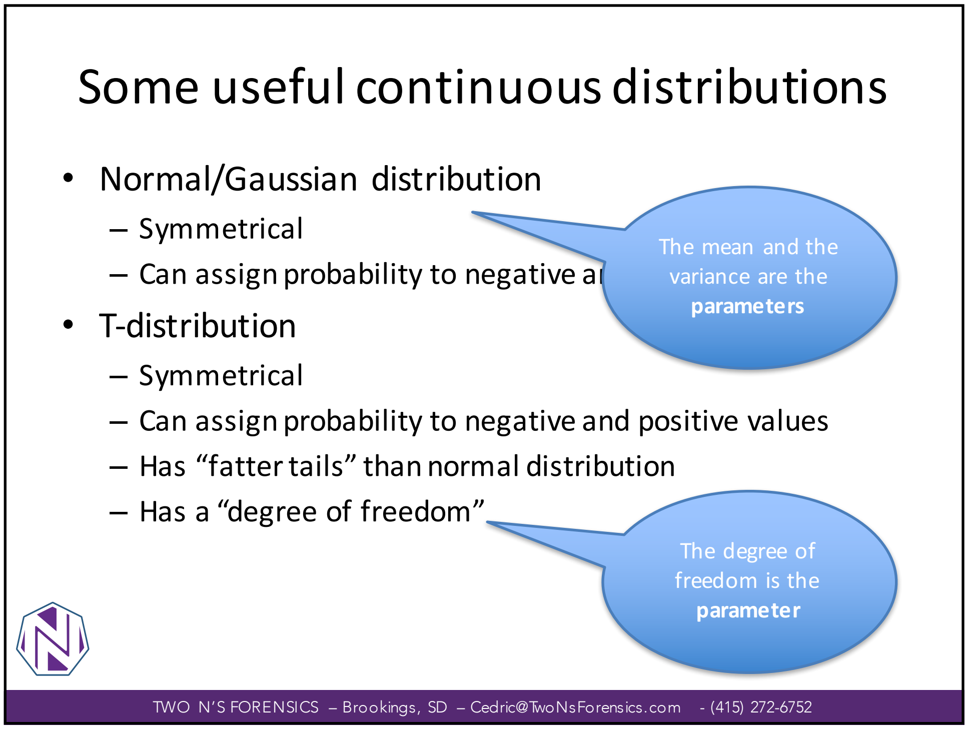






Some useful continuous distributions

- Normal/Gaussian distribution
  - Symmetrical
  - Can assign probability to negative and positive values
- T-distribution
  - Symmetrical
  - Can assign probability to negative and positive values
  - Has “fatter tails” than normal distribution
  - Has a “degree of freedom”

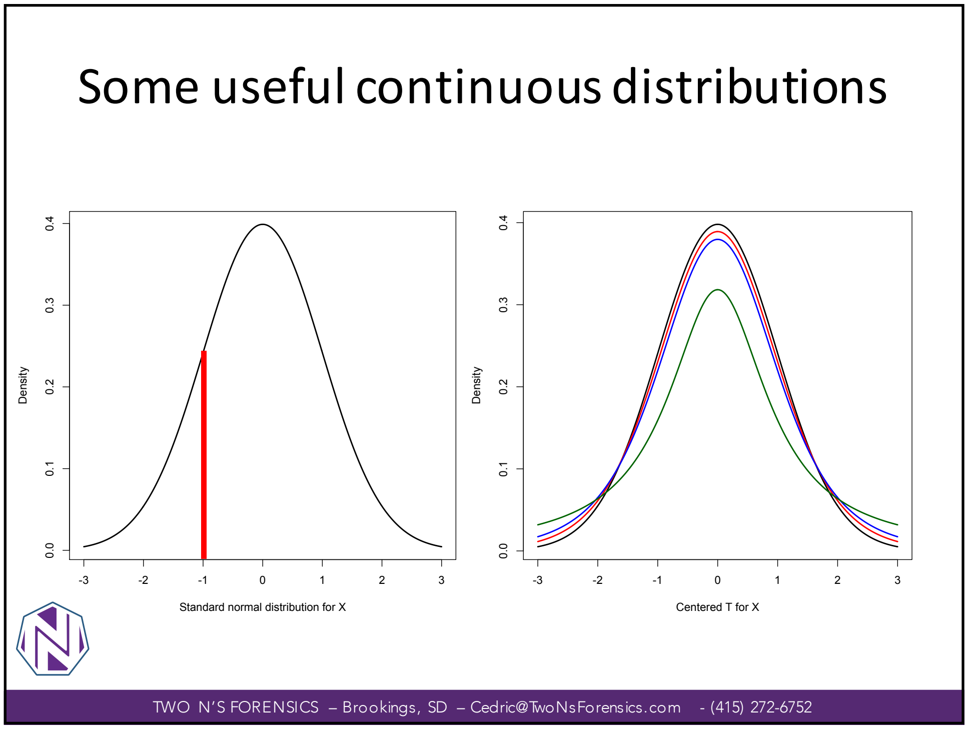


Some useful continuous distributions

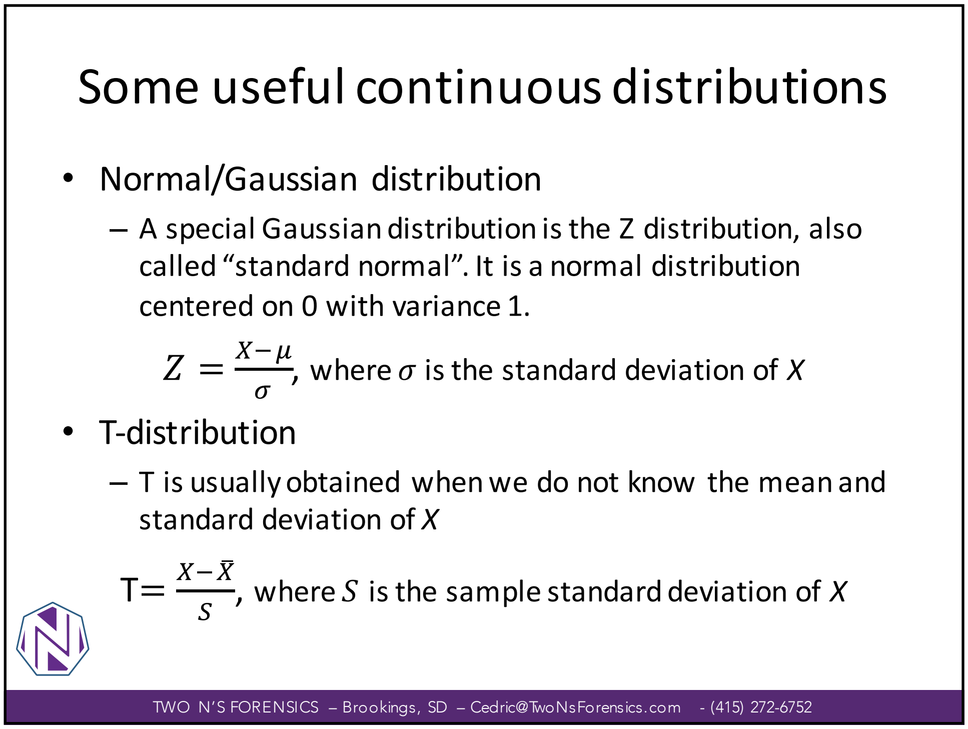
- Normal/Gaussian distribution
  - Symmetrical
  - Can assign probability to negative a
- T-distribution
  - Symmetrical
  - Can assign probability to negative and positive values
  - Has “fatter tails” than normal distribution
  - Has a “degree of freedom”

The mean and the variance are the **parameters**

The degree of freedom is the **parameter**



Some useful continuous distributions



Some useful continuous distributions

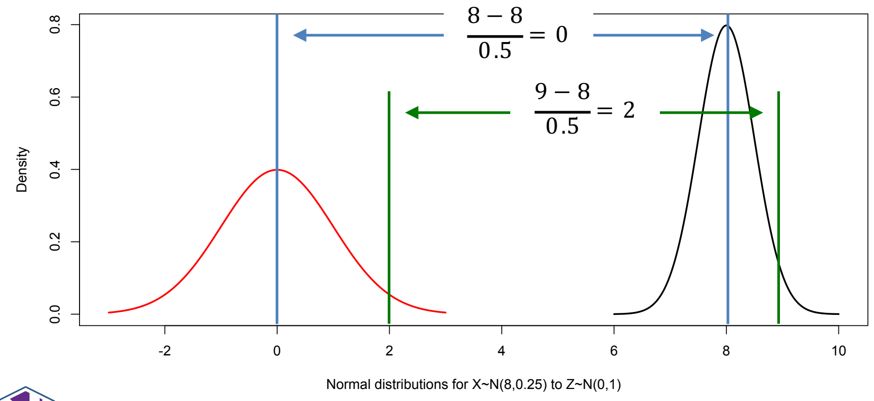
- Normal/Gaussian distribution
  - A special Gaussian distribution is the Z distribution, also called "standard normal". It is a normal distribution centered on 0 with variance 1.

$Z = \frac{X-\mu}{\sigma}$, where σ is the standard deviation of X

- T-distribution
  - T is usually obtained when we do not know the mean and standard deviation of X

T$= \frac{X-\bar{X}}{S}$, where S is the sample standard deviation of X

Some useful continuous distributions



$$\frac{8-8}{0.5}=0$$

$$\frac{9-8}{0.5}=2$$

Some useful continuous distributions

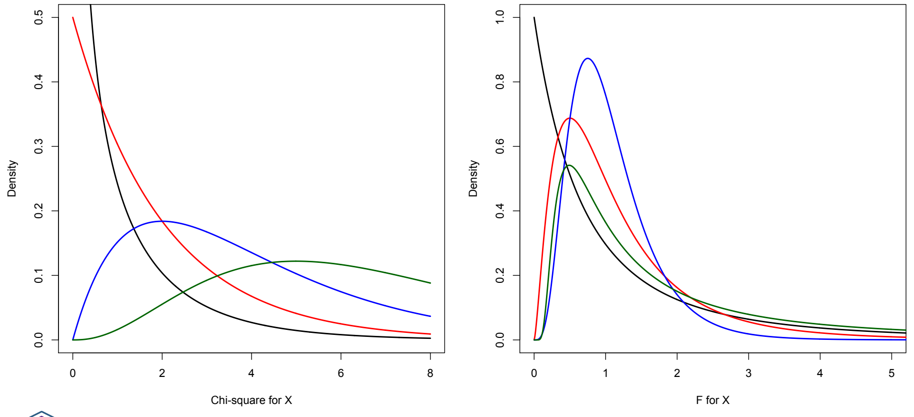
- χ^2 or chi-square (pronounced "ki square")
  - Non-symmetrical
  - Can assign probability <u>to positive values only</u>
  - Can be found when we sum squared measurements
  - Has a "degree of freedom"
- F-distribution
  - Non-symmetrical
  - Can assign probability <u>to positive values only</u>
  - Can be found when we have a ratio of some sort
  - Has <u>two</u> "degrees of freedom"

Some useful continuous distributions

- χ^2 or chi-square (pronounced "ki square")
  - Non-symmetrical
  - Can assign probability to positive values only
  - Can be found when we sum squared measurements
  - Has a "degree of freedom"
- F-distribution
  - Non-symmetrical
  - Can assign probability to positive values only
  - Can be found when we have a ratio of some sort
  - Has two "degrees of freedom"

The number of degrees of freedom are the **parameters**

Some useful continuous distributions

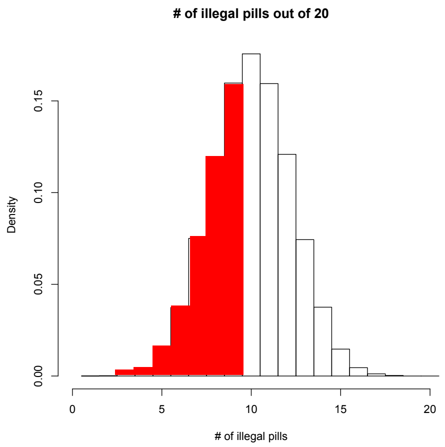


Probability of a range

- So far we have see that these distributions enable us to assign $\Pr(X = x)$
- But what about $\Pr(X \geq x)$ or $\Pr(X \leq x)$?
- Same concept

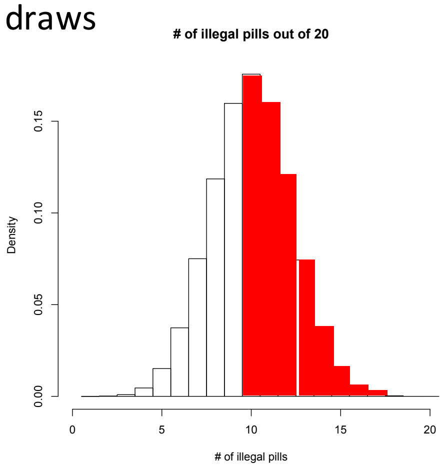
Probability of a range

- $\Pr(X \leq 9)$: Probability to have 9 or less illegal pills in 20 draws



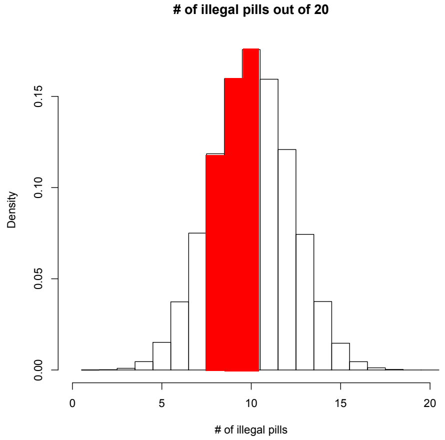
Probability of a range

- $\Pr(X \geq 10)$: Probability to have <u>10 or more</u> illegal pills in 20 draws



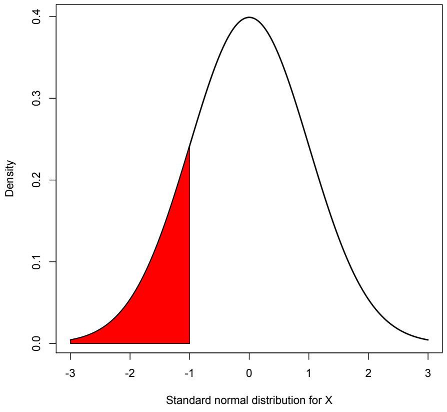
Probability of a range

- $\Pr(8 \leq X \leq 10)$: Probability to have <u>between 8 and 10</u> illegal pills in 20 draws



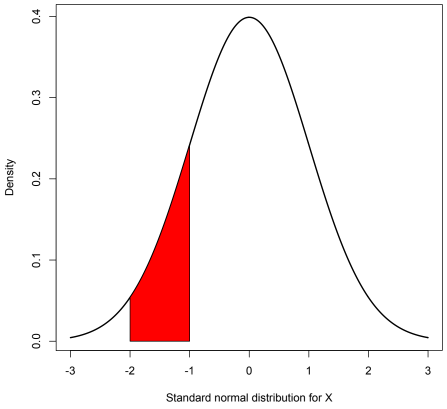
Probability of a range

- $\Pr(X \leq -1)$: Probability to observe a value of X smaller than -1



Probability of a range

- $\Pr(-2 \leq X \leq -1)$: Probability to observe a value in the range of -2 to -1



Tables for Z, T and χ^2

- In your binder, you will find tables to calculate various probabilities for Z, T and χ^2 distributions
- Here is how they work

Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where *z* is a constant from the table
- Let's say we want to calculate $\Pr(0 \leq Z \leq 1.15)$

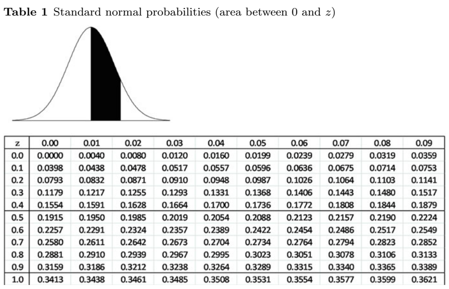


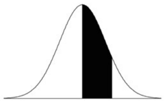
Table 1 Standard normal probabilities (area between 0 and z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

- The table is giving us $\Pr(0 \leq Z \leq \mathrm{z})$ where *z* is a constant from the table
- Let's say we want to calculate $\Pr(0 \leq Z \leq 1.15)$

Table 1 Standard normal probabilities (area between 0 and z)

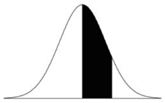


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

- The table is giving us $\Pr(0 \leq Z \leq \mathrm{z})$ where *z* is a constant from the table
- Let's say we want to calculate $\Pr(0 \leq Z \leq 1.15)$

Table 1 Standard normal probabilities (area between 0 and z)

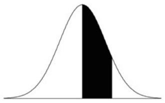


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where *z* is a constant from the table
- Let's say we want to calculate $\Pr(-\infty \leq Z \leq 1.15)$

Table 1 Standard normal probabilities (area between 0 and *z*)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

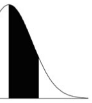
Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where *z* is a constant from the table
- Let's say we want to calculate $\Pr(-\infty \leq Z \leq 1.15)$

We are not accounting for the part below 0

Since the distribution is symmetrical, this is 50%

[illegible]ndard normal probabilities (area between 0 and *z*)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

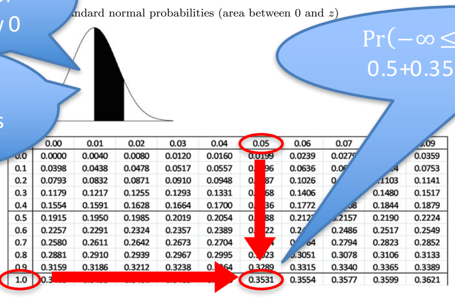
Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where z is a constant from the table
- Let's say we want to calculate $\Pr(-\infty \leq Z \leq 1.15)$

We are not accounting for the part below 0

Since the distribution is symmetrical, this is 50%

$\Pr(-\infty \leq Z \leq 1.15)$= 0.5+0.3531 = 0.8531



ndard normal probabilities (area between 0 and z)

[illegible]	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	[illegible]	[illegible]
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	[illegible]	[illegible]	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	[illegible]	[illegible]	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	[illegible]	0.1026	[illegible]	[illegible]	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	[illegible]	0.1406	[illegible]	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	[illegible]	0.1772	[illegible]	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	[illegible]	[illegible]	[illegible]	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	[illegible]	[illegible]	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	[illegible]	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	[illegible]	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where z is a constant from the table
- Let's say we want to calculate $\Pr(1.15 \leq Z)$

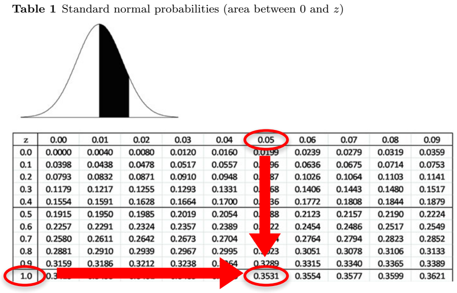


Table 1 Standard normal probabilities (area between 0 and z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

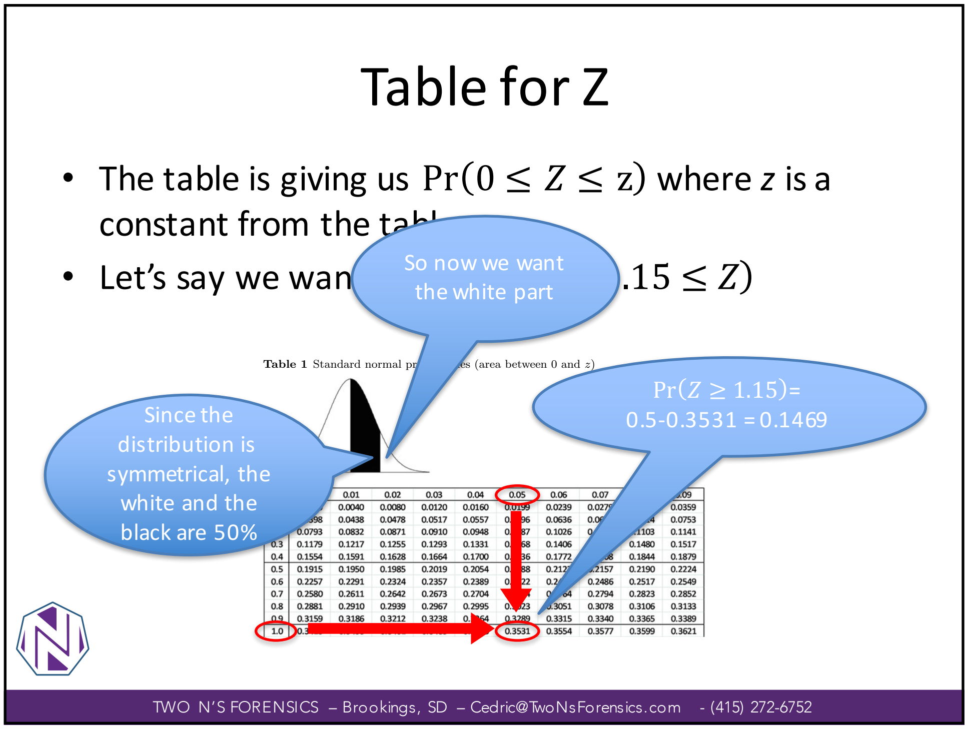


Table for Z

- The table is giving us $\mathrm{Pr}(0 \le Z \le \mathrm{z})$ where *z* is a constant from the tabl
- Let's say we wan $.15 \le Z)$

So now we want the white part

Since the distribution is symmetrical, the white and the black are 50%

$\mathrm{Pr}(Z \ge 1.15)$= 0.5-0.3531 = 0.1469

Table 1 Standard normal pr es (area between 0 and z)

[illegible]	[illegible]	0.01	0.02	0.03	0.04	0.05	0.06	0.07	[illegible]	.09
[illegible]	[illegible]	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.027[illegible]	[illegible]	0.0359
[illegible]	[illegible]398	0.0438	0.0478	0.0517	0.0557	0.[illegible]96	0.0636	0.06[illegible]	[illegible]4	0.0753
[illegible]	0.0793	0.0832	0.0871	0.0910	0.0948	0.[illegible]87	0.1026	0[illegible]	[illegible]1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.[illegible]68	0.1406	[illegible]	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.[illegible]36	0.1772	[illegible]08	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.[illegible]88	0.212[illegible]	[illegible]2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.[illegible]22	0.2[illegible]	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0[illegible]64	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.[illegible]23	[illegible]3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.[illegible]64	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

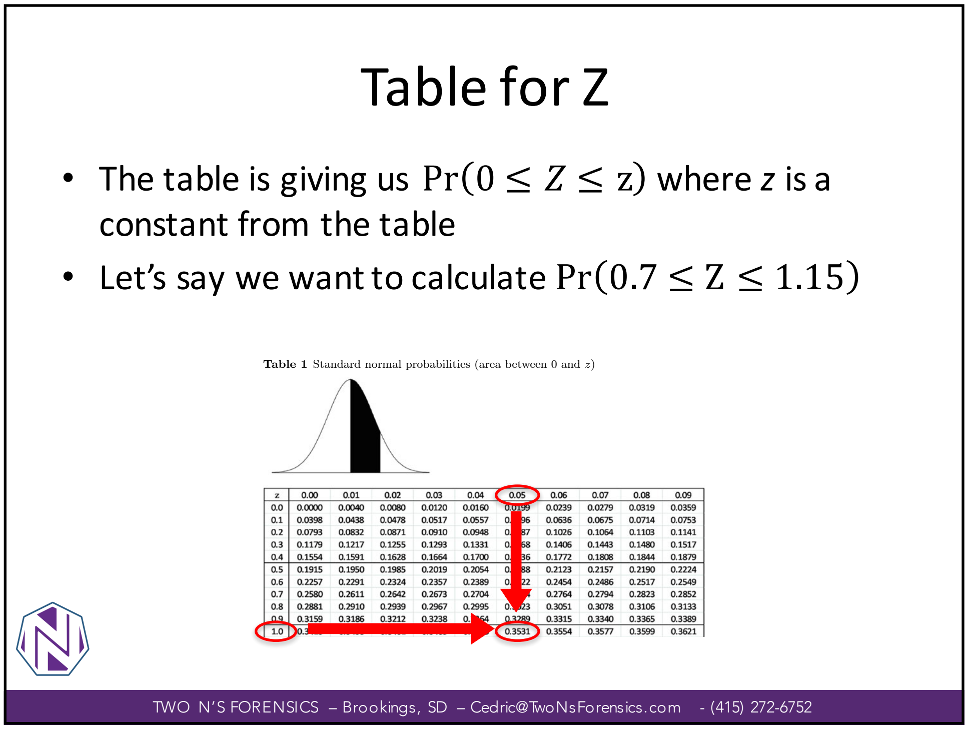


Table for Z

- The table is giving us $\mathrm{Pr}(0 \le Z \le \mathrm{z})$ where *z* is a constant from the table
- Let's say we want to calculate $\mathrm{Pr}(0.7 \le \mathrm{Z} \le 1.15)$

Table 1 Standard normal probabilities (area between 0 and z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.[illegible]96	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.[illegible]87	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.[illegible]68	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.[illegible]36	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.[illegible]88	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.[illegible]22	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.[illegible]23	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.[illegible]64	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

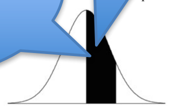
- The table i… $Z \leq z)$ where *z* is a consta…
- L… wa… …te $\Pr(0.7 \leq Z \leq 1.15)$

Now we want the black part, but it does not start at 0, it starts at 0.7

We need to **remove the part between 0 and 0.7** from the part that goes from 0 to 1.15

$\Pr(0.7 \leq Z \leq 1.15) =$
$\Pr(0.7 \leq Z) - \Pr(Z \leq 1.15)$
0.3531-0.2580 = 0.0951

…ndard… l probabilities (area…

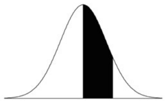


z	0.00	0.01	0.02	0.03	0.04	0.05	[illegible]	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	[illegible]	[illegible]	0.0279	0.0319	0.0359
0.1	[illegible]	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	[illegible]	0.0832	0.0871	0.0910	[illegible]	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	[illegible]	0.1217	0.1255	0.1293	[illegible]	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	[illegible]	0.1591	0.1628	0.1664	[illegible]	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	[illegible]	0.1950	0.1985	[illegible]	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	[illegible]	0.2291	0.2324	[illegible]	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where *z* is a constant from the table
- Let's say we want to calculate $\Pr(-0.7 \leq Z \leq 1.15)$

Table 1 Standard normal probabilities (area between 0 and *z*)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

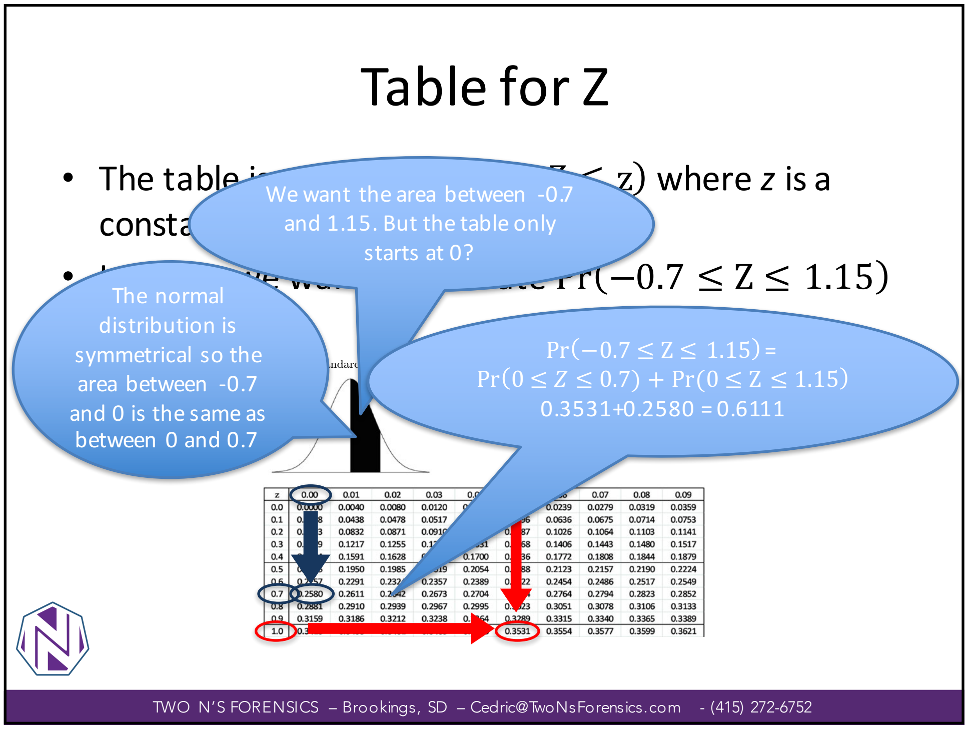


Table for Z

- The table i… Z ≤ z) where *z* is a consta…
- L… we wa… ate $\Pr(-0.7 \leq Z \leq 1.15)$

We want the area between -0.7 and 1.15. But the table only starts at 0?

The normal distribution is symmetrical so the area between -0.7 and 0 is the same as between 0 and 0.7

$\Pr(-0.7 \leq Z \leq 1.15) =$
$\Pr(0 \leq Z \leq 0.7) + \Pr(0 \leq Z \leq 1.15)$
0.3531+0.2580 = 0.6111

z	0.00	0.01	0.02	0.03	[illegible]	[illegible]	[illegible]	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	[illegible]	[illegible]	0.0239	0.0279	0.0319	0.0359
0.1	[illegible]	0.0438	0.0478	0.0517	[illegible]	[illegible]	0.0636	0.0675	0.0714	0.0753
0.2	[illegible]	0.0832	0.0871	[illegible]	[illegible]	[illegible]	0.1026	0.1064	0.1103	0.1141
0.3	[illegible]	0.1217	0.1255	[illegible]	[illegible]	[illegible]	0.1406	0.1443	0.1480	0.1517
0.4	[illegible]	0.1591	0.1628	[illegible]	0.1700	[illegible]	0.1772	0.1808	0.1844	0.1879
0.5	[illegible]	0.1950	0.1985	[illegible]	0.2054	[illegible]	0.2123	0.2157	0.2190	0.2224
0.6	[illegible]	0.2291	[illegible]	0.2357	0.2389	[illegible]	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	[illegible]	0.2673	0.2704	[illegible]	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	[illegible]	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	[illegible]	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3531	0.3554	0.3577	0.3599	0.3621

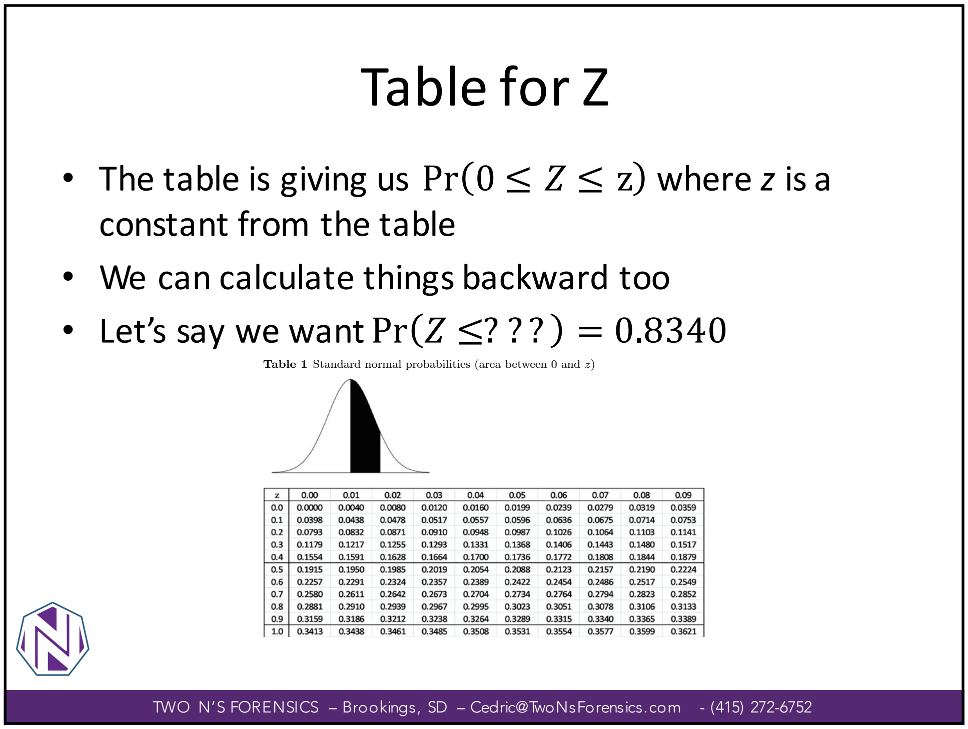


Table for Z

- The table is giving us $\Pr(0 \leq Z \leq z)$ where *z* is a constant from the table
- We can calculate things backward too
- Let's say we want $\Pr(Z \leq ???) = 0.8340$

Table 1 Standard normal probabilities (area between 0 and *z*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621

Table for Z

$\Pr(Z \leq z) = 0.8340$

$\Pr(Z \leq z) = \Pr(Z \leq 0) + \Pr(0 \leq Z \leq z) = 0.8340$

$\Pr(Z \leq z) = \Pr(Z \leq 0) + \Pr(0 \leq Z \leq z) = 0.5 + 0.3340$

constant from the ta

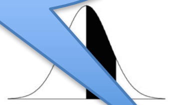
- We can calculate things ba rd too

$\Pr(Z \leq z) = 0.8340$

$z = 0.97$

r$(Z \leq ???) = 0.8340$

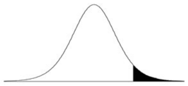
ormal probabilities (area between 0 and z)



z	0.00	0.01	0.0[illegible]	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	[illegible]0120	0.0160	0.0199	0.0239	0.[illegible]79	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.[illegible]17	0.0557	0.0596	0.0636	[illegible]	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091[illegible]	0.0948	0.0987	0.1026	0.[illegible]64	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.[illegible]43	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.[illegible]08	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.[illegible]57	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.[illegible]86	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.[illegible]94	0.2823	0.2852
0.8	0.2881	[illegible].2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.315[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]315	0.3340	0.3365	0.3389
1.0	0.3413	[illegible].3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621

Table for T

- The table is giving us $\Pr(T \leq t) = \alpha$ where t is a constant from the table and α is a pre-set probability
- You can also see the "df" column. This is the number of "degree(s) of freedom"
- This table works by row (one for each df)



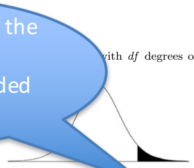
df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10

Table for T

- The table is giving us $\mathrm{Pr}(T \leq t) = \alpha$ where *t* is a constant from the table and α is a pre-set probability
- Let's say we want to have $\mathrm{Pr}(T \leq 6.965) = \alpha$ for *df=2*

Here the number is the value of α and represents the shaded area

vith df degrees of freedom. (*shaded area*

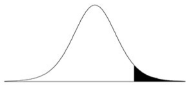


df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	[illegible]	[illegible]	[illegible]	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10

Table for T

- The table is giving us $\mathrm{Pr}(T \leq t) = \alpha$ where *t* is a constant from the table and α is a pre-set probability
- It works the other way too $\mathrm{Pr}(T \leq ???) = 0.025$ for *df=4*

Table 2 Values of t_α in a t distribution with df degrees of freedom. (*shaded area* $P(t > t_\alpha) = \alpha$)

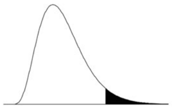


df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	[illegible]	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	[illegible]	[illegible]	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10

Table for χ^2

- The table is giving us $\Pr(\mathrm{X}^2 \leq \chi^2) = \alpha$ where χ^2 is a constant from the table and α is a pre-set probability
- Same concept as T table

Table 3 Values of $\chi^2_{\alpha,\mathrm{df}}$ in a chi-square distribution with *df* degrees of freedom (*shaded area* $P(\chi^2 > \chi^2_{\alpha,\mathrm{df}}) = \alpha$)

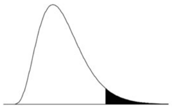


df	α = .995	α = .990	α = .975	α = .950	α = .050	α = .025	α = .010	α = .005	df
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15

Table for χ^2

- The table is giving us $\Pr(\mathrm{X}^2 \leq \chi^2) = \alpha$ where χ^2 is a constant from the table and α is a pre-set probability
- Say we want $\Pr(\mathrm{X}^2 \leq ???) = 0.05$ for *df=6*

Table 3 Values of $\chi^2_{\alpha,\mathrm{df}}$ in a chi-square distribution with *df* degrees of freedom (*shaded area* $P(\chi^2 > \chi^2_{\alpha,\mathrm{df}}) = \alpha$)



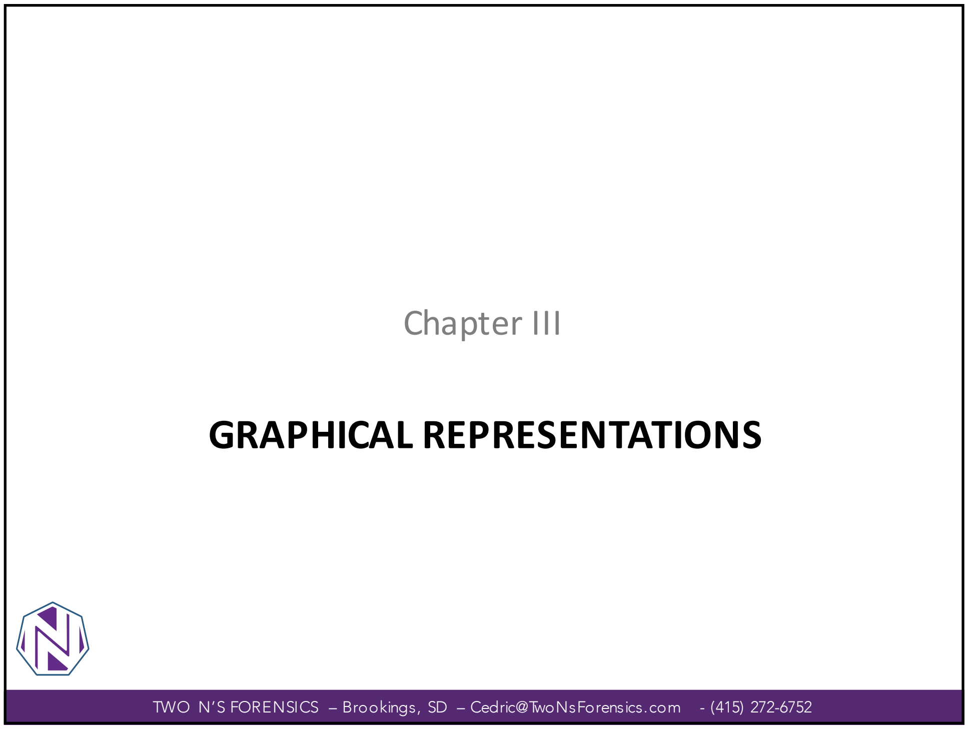
df	α = .995	α = .990	α = .975	α = .950	α = .050	α = .025	α = .010	α = .005	df
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	[illegible]	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	5
6	[illegible]	[illegible]	[illegible]	[illegible]	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15

Take home messages

- There are different types of probability distributions depending on the type of variable, and depending on what we want to model
- These probability distributions are governed by **parameters**
- We can use these probability distributions to assign the probability of a certain observation from an experiment
- We can assign these probabilities by calculating them or using pre-calculated tables

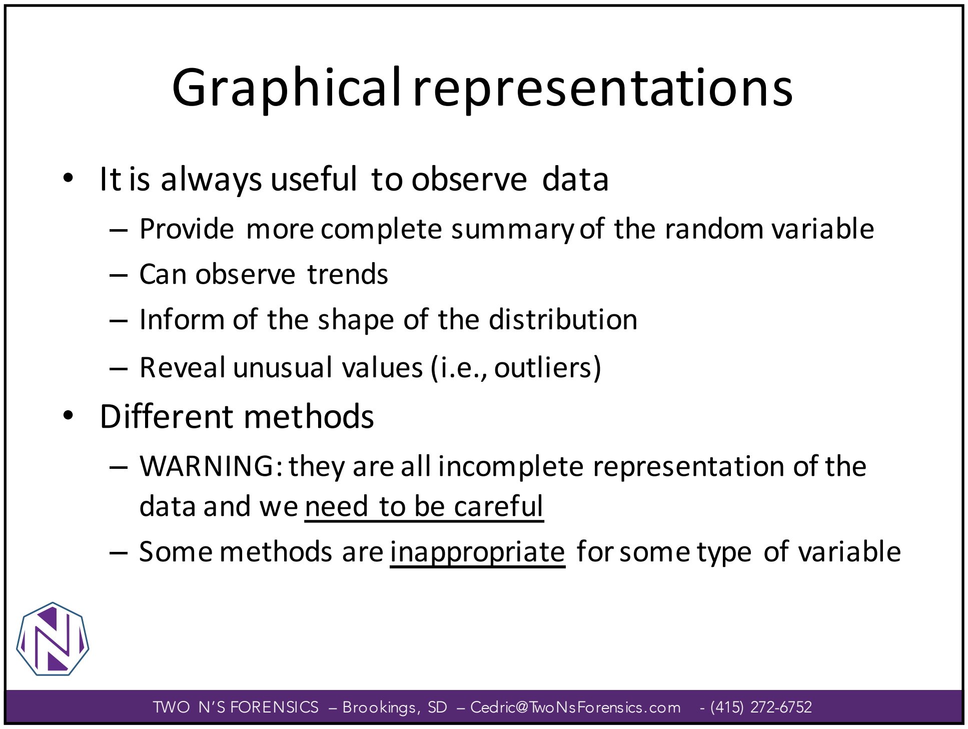
Chapter II

EXERCISES



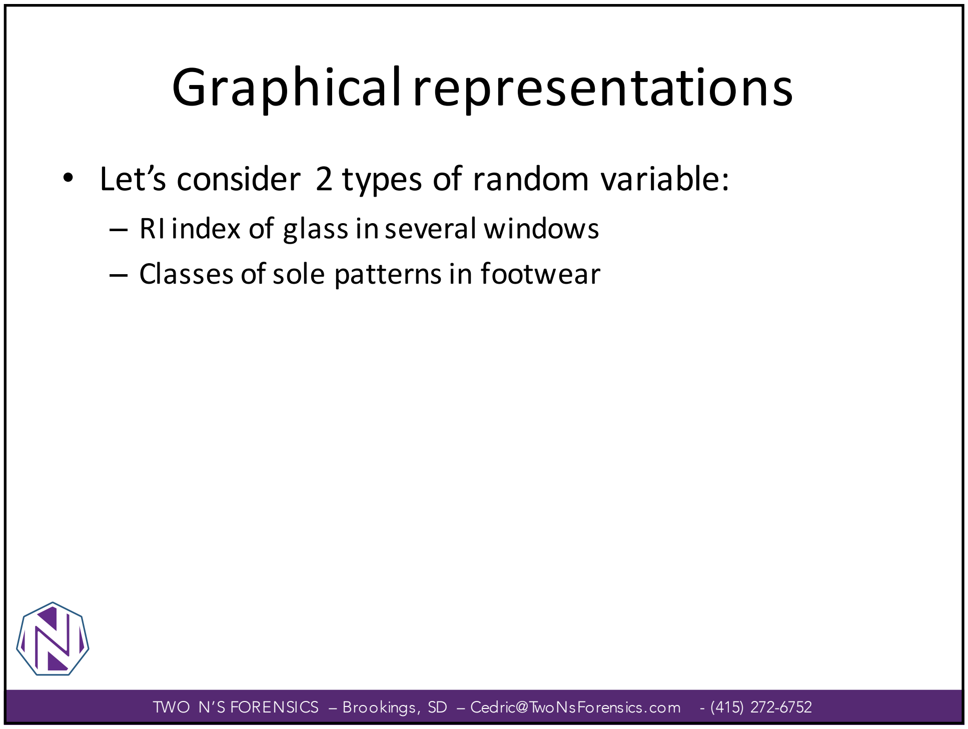
Chapter III

GRAPHICAL REPRESENTATIONS



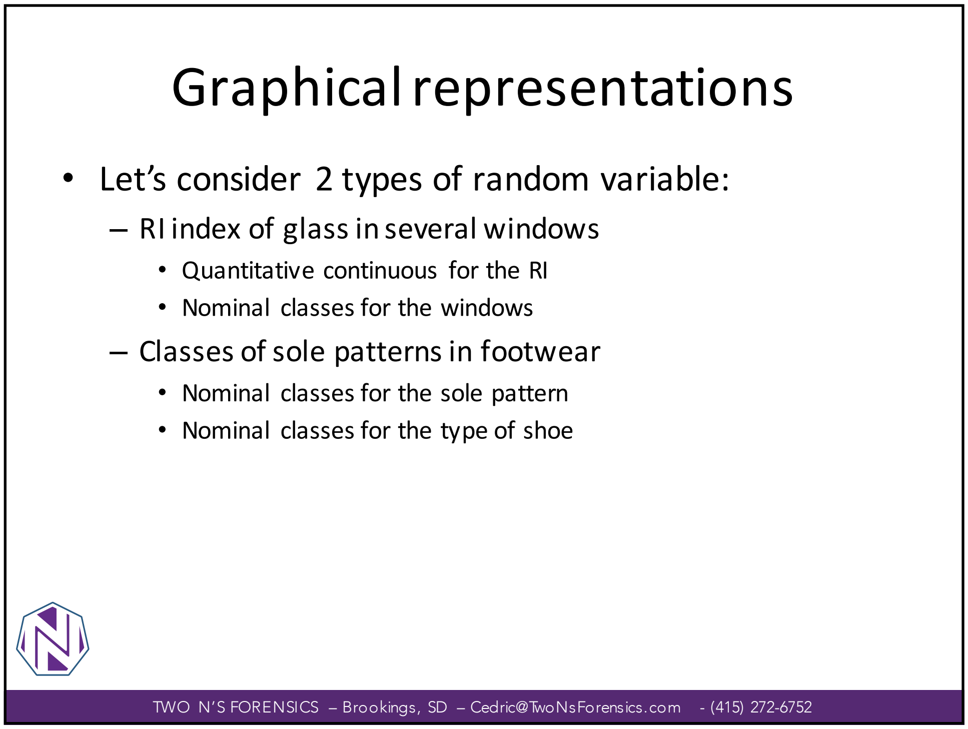
Graphical representations

- It is always useful to observe data
  - Provide more complete summary of the random variable
  - Can observe trends
  - Inform of the shape of the distribution
  - Reveal unusual values (i.e., outliers)
- Different methods
  - WARNING: they are all incomplete representation of the data and we need to be careful
  - Some methods are inappropriate for some type of variable



Graphical representations

- Let’s consider 2 types of random variable:
  - RI index of glass in several windows
  - Classes of sole patterns in footwear

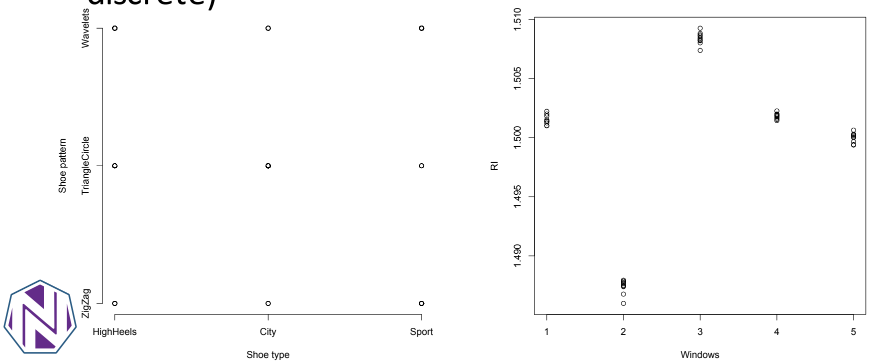


Graphical representations

- Let’s consider 2 types of random variable:
  - RI index of glass in several windows
    - Quantitative continuous for the RI
    - Nominal classes for the windows
  - Classes of sole patterns in footwear
    - Nominal classes for the sole pattern
    - Nominal classes for the type of shoe

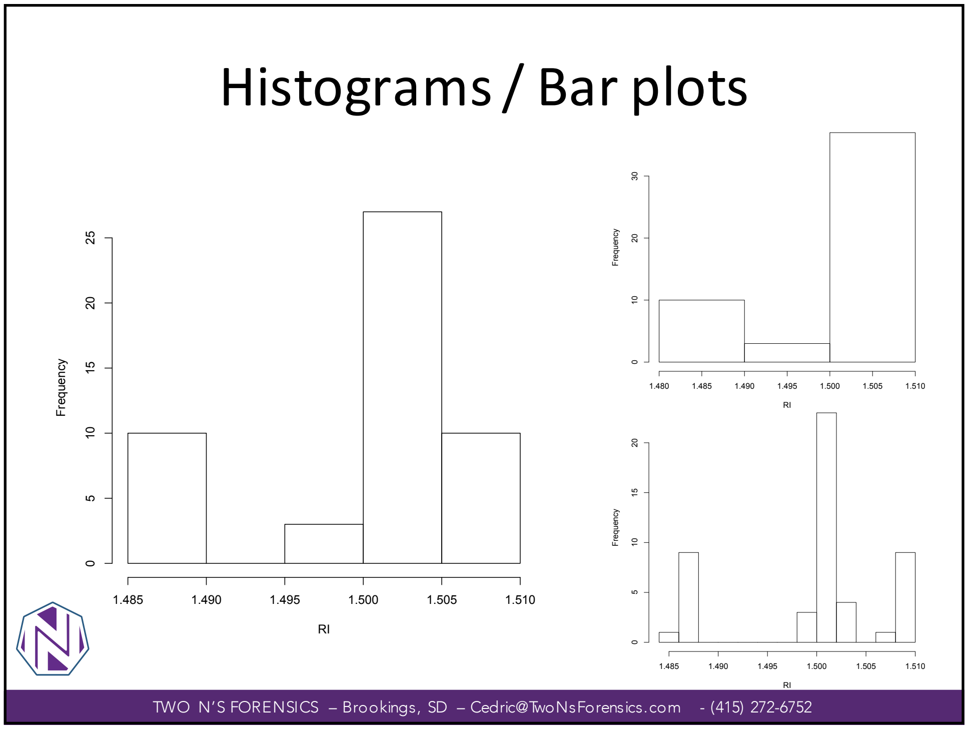
Scatter plots

- Plot 1 type of information (can be continuous, discrete, nominal or ordinal) against another type of information (usually continuous, or quantitative discrete)



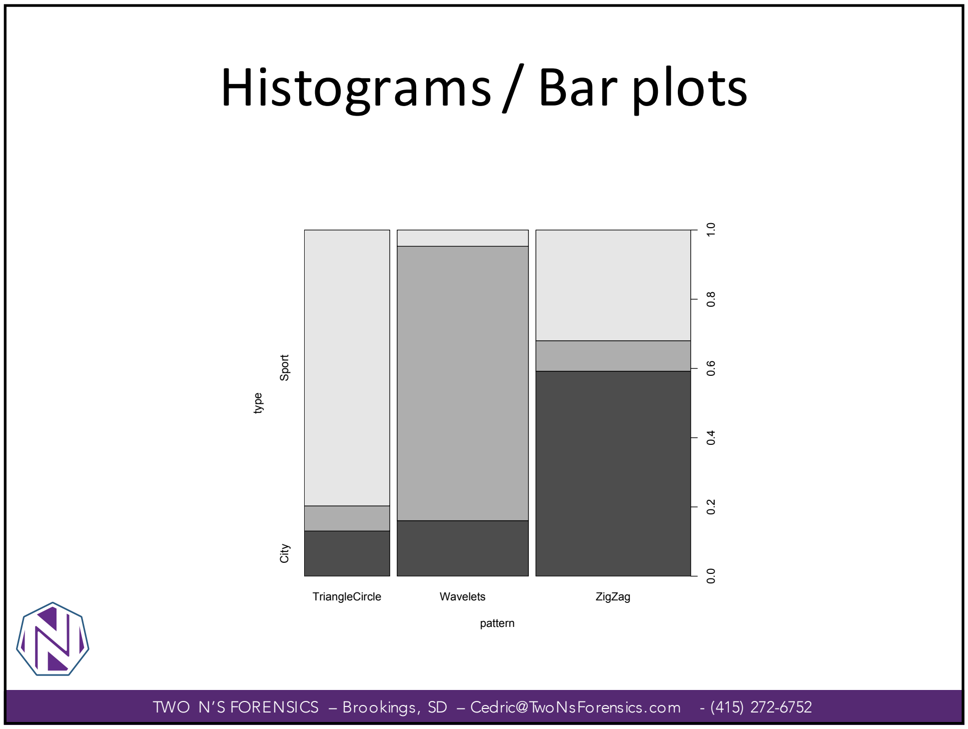
Histograms / Bar plots

- While scatter plot is informative, it is sometimes difficult to have a good feel for the distribution of the data.
- We can use a histogram
  - Plot the "counts" of for each considered numerical value
  - "Bin" the values when faced with a continuous variable
    - Size of the "bins" matter!
  - It is not appropriate when we have categorical data!!!
    - In that case, we use a "bar plot"



Histograms / Bar plots

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



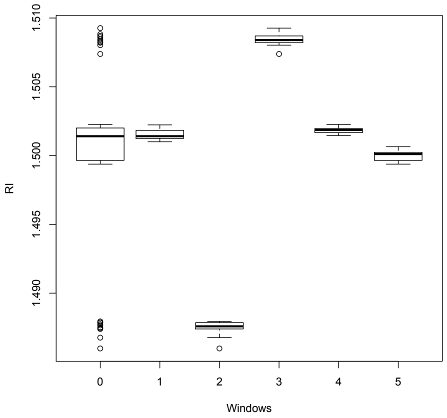
Histograms / Bar plots

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Boxplots

- Histograms are good to have a feel of the distributions but it is difficult to compare them
- We want to summarize the data a bit more
- Boxplots:
  - Plot a categorical variable against a **quantitative** variable
  - Show where the mass of the distribution is
  - Can be deceptive (if the original distribution is multi-modal)

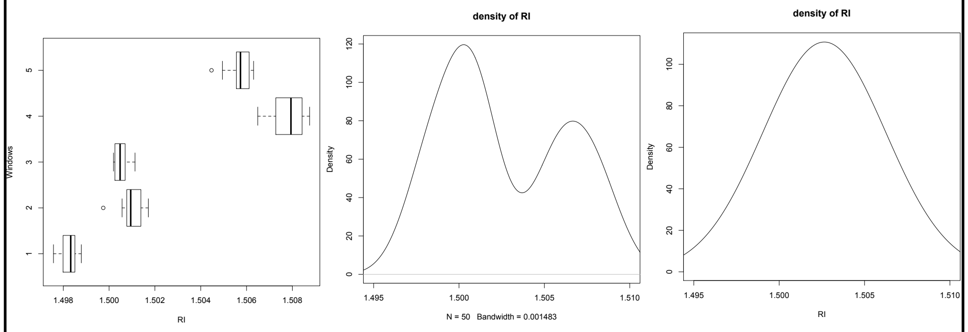
Boxplots



Density plots

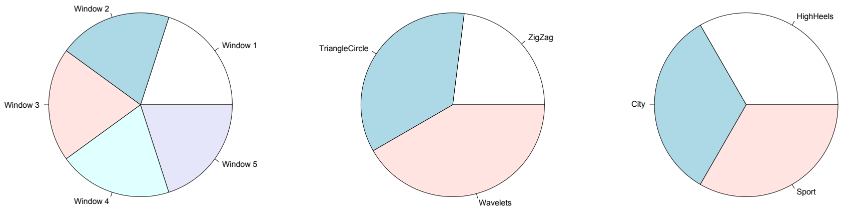
- Continuous version of the histogram
  - Variable needs to be continuous (or discrete but with small intervals and large range)
    - Can be parametric or non-parametric
    - Parametric: estimate the parameters of the distribution and then plot it
    - Non-parametric: find the best fit according to some constraints

Density plots



Pie charts

- Express proportions for categorical variables
  - Does not really work for continuous variables



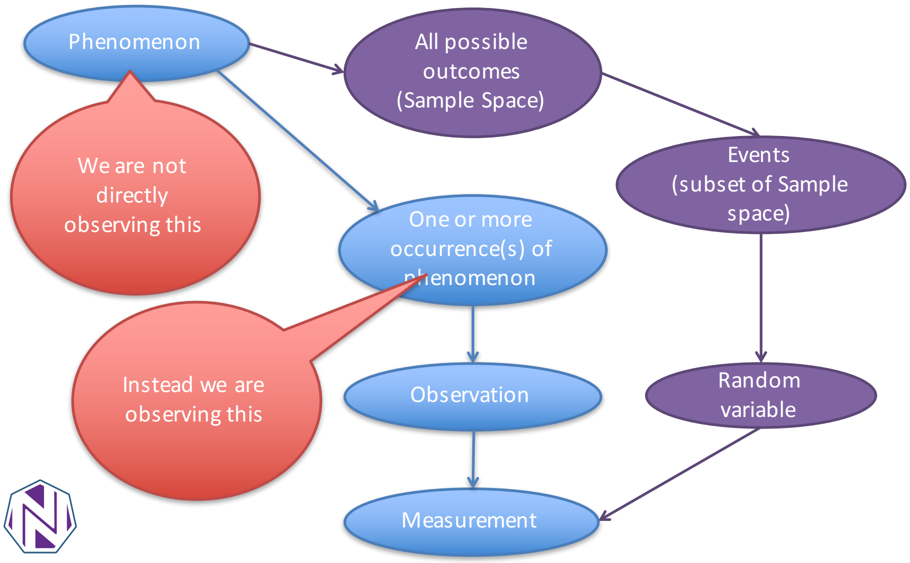
Take home messages

- Graphically representing data is useful
  - Quickly get a feel for the data and its distribution
  - Get a feel for the type of model needed, assumptions required and expected results
- Graphically representing data can be deceptive
  - It summarizes the data
  - Different ways of summarizing the data
  - Losing information!
- Need to use the appropriate type of plot

Chapter IV

POPULATION VS. SAMPLE PARAMETER ESTIMATES AND CONFIDENCE INTERVALS

Population vs. sample



Phenomenon

All possible outcomes (Sample Space)

Events (subset of Sample space)

We are not directly observing this

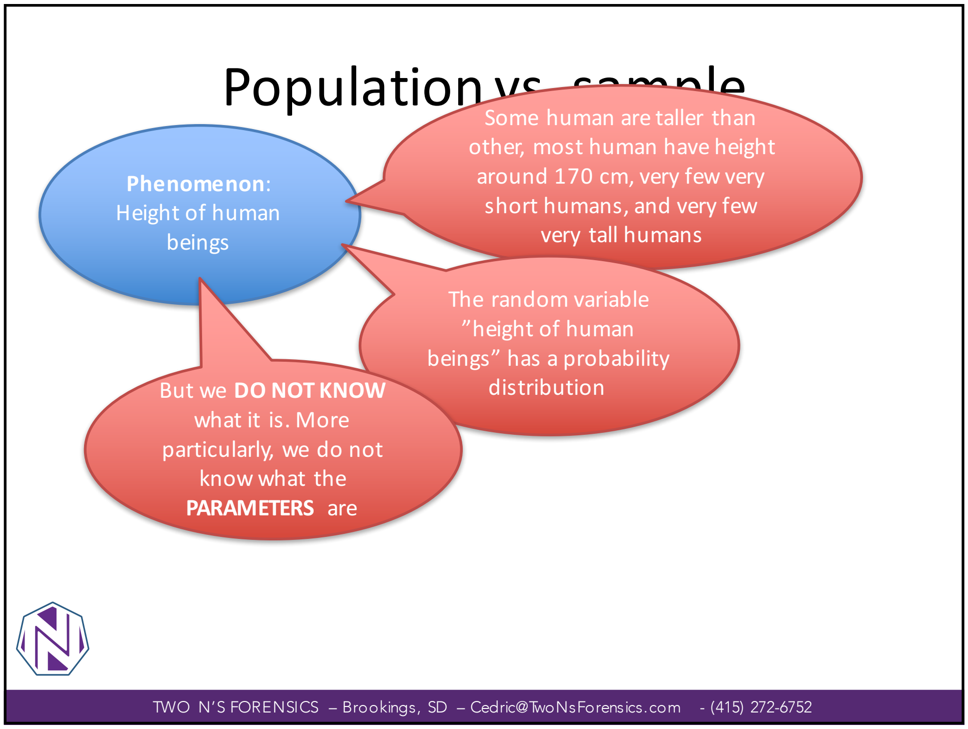
One or more occurrence(s) of phenomenon

Instead we are observing this

Observation

Random variable

Measurement



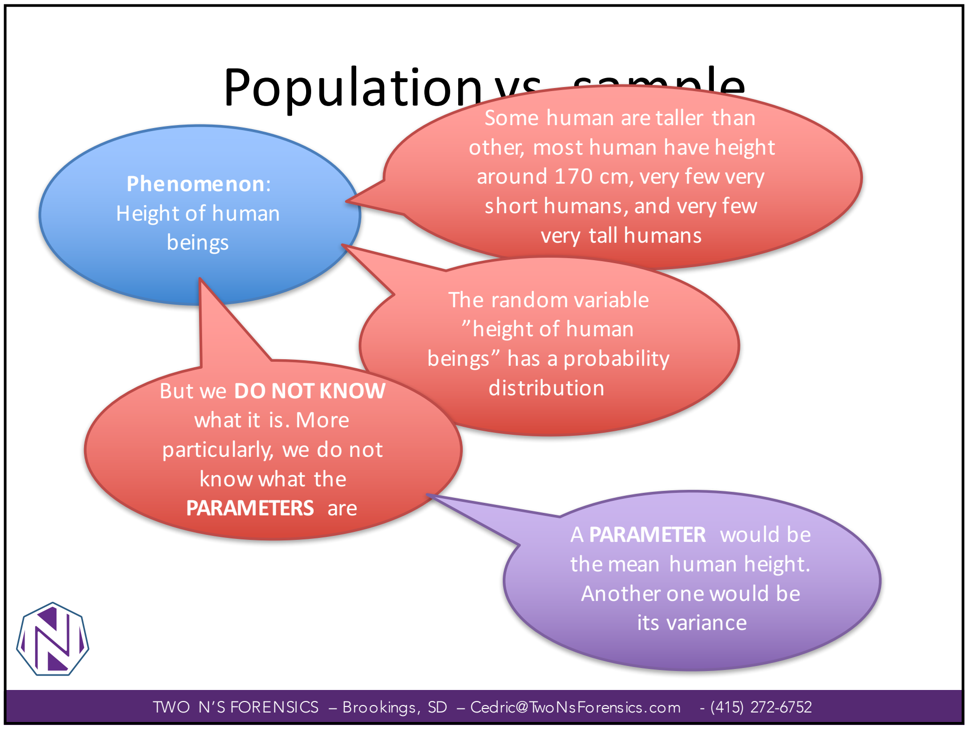
Population vs. sample

Phenomenon: Height of human beings

Some human are taller than other, most human have height around 170 cm, very few very short humans, and very few very tall humans

The random variable ”height of human beings” has a probability distribution

But we **DO NOT KNOW** what it is. More particularly, we do not know what the **PARAMETERS** are



Population vs. sample

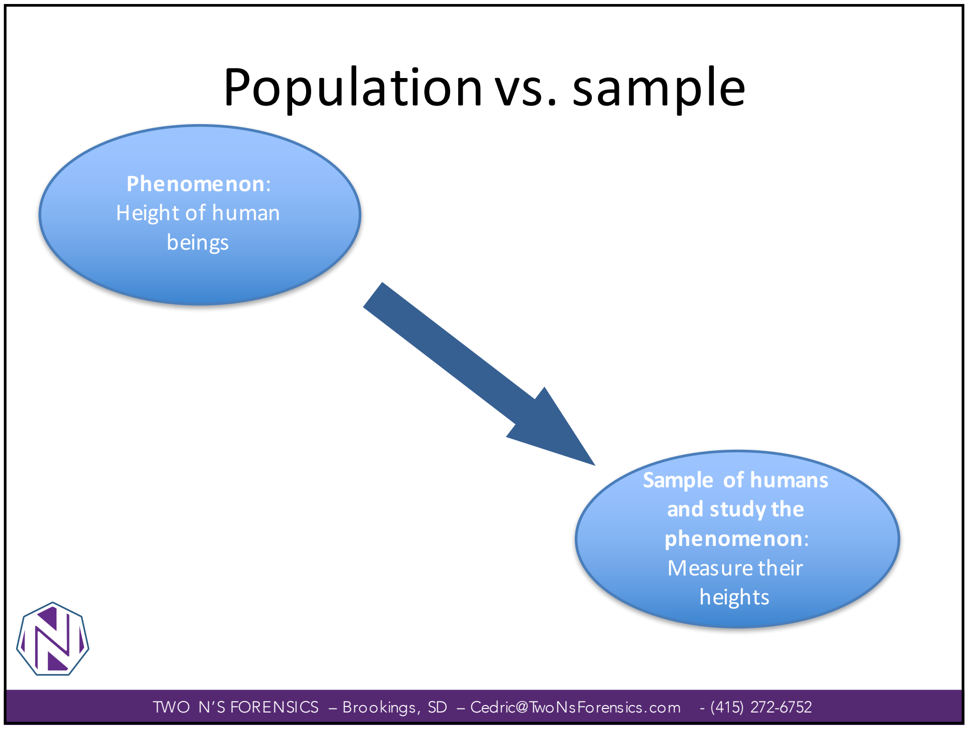
Phenomenon: Height of human beings

Some human are taller than other, most human have height around 170 cm, very few very short humans, and very few very tall humans

The random variable ”height of human beings” has a probability distribution

But we **DO NOT KNOW** what it is. More particularly, we do not know what the **PARAMETERS** are

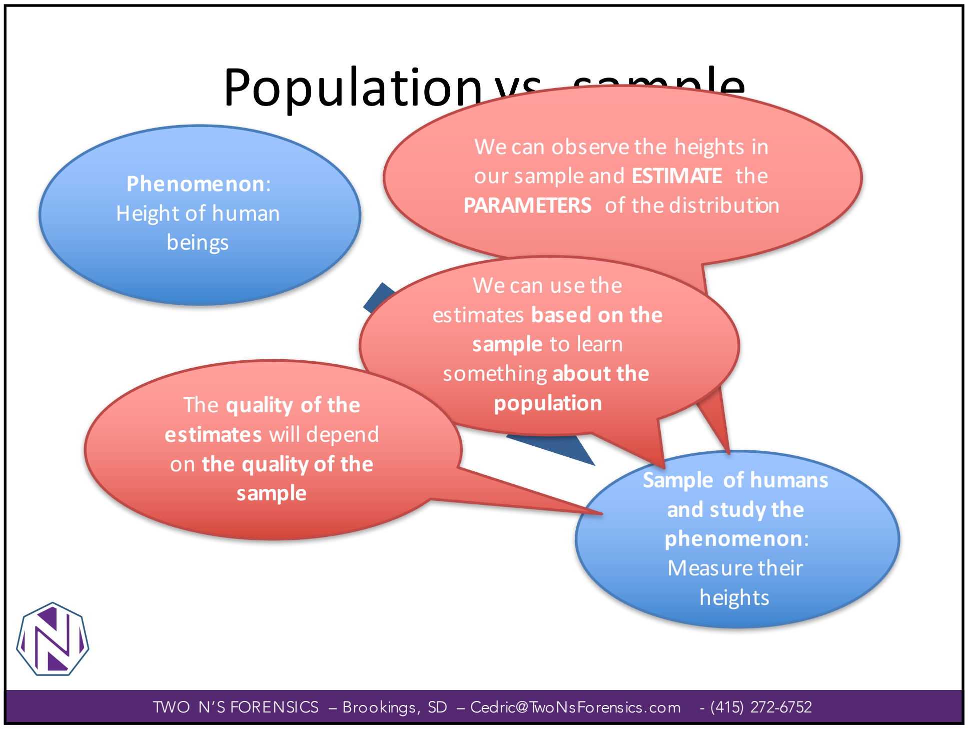
A **PARAMETER** would be the mean human height. Another one would be its variance



Population vs. sample

Phenomenon: Height of human beings

Sample of humans and study the phenomenon: Measure their heights



Population vs. sample

Phenomenon: Height of human beings

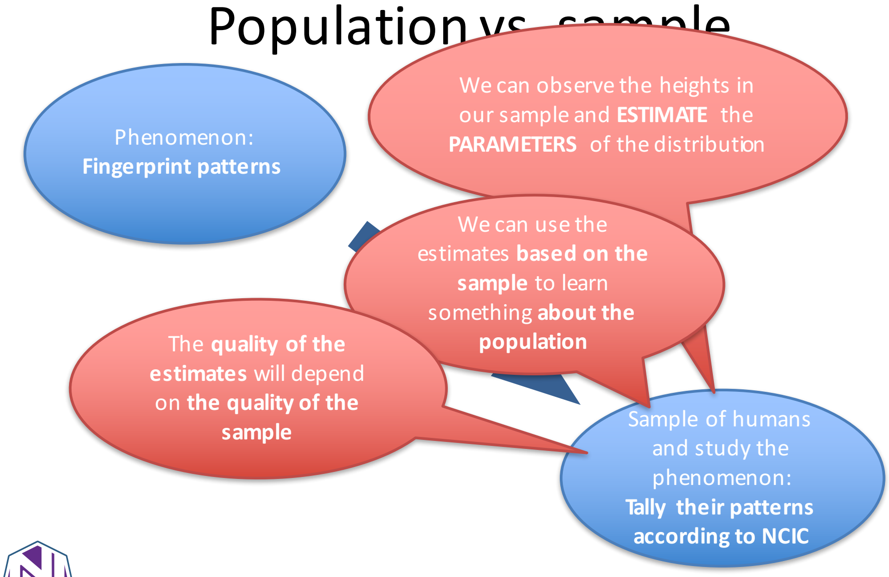
We can observe the heights in our sample and **ESTIMATE** the **PARAMETERS** of the distribution

We can use the estimates **based on the sample** to learn something **about the population**

The **quality of the estimates** will depend on **the quality of the sample**

Sample of humans and study the phenomenon: Measure their heights

Population vs. sample



Phenomenon:
Fingerprint patterns

We can observe the heights in our sample and **ESTIMATE** the **PARAMETERS** of the distribution

We can use the estimates **based on the sample** to learn something **about the population**

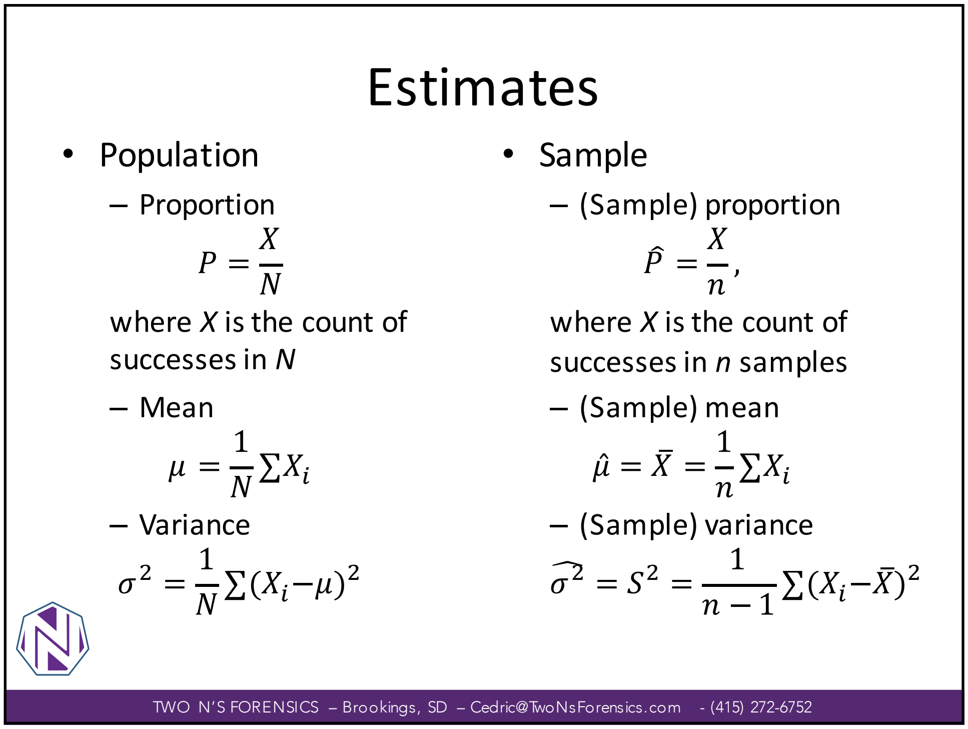
The **quality of the estimates** will depend on **the quality of the sample**

Sample of humans and study the phenomenon:
Tally their patterns according to NCIC

Population vs. sample

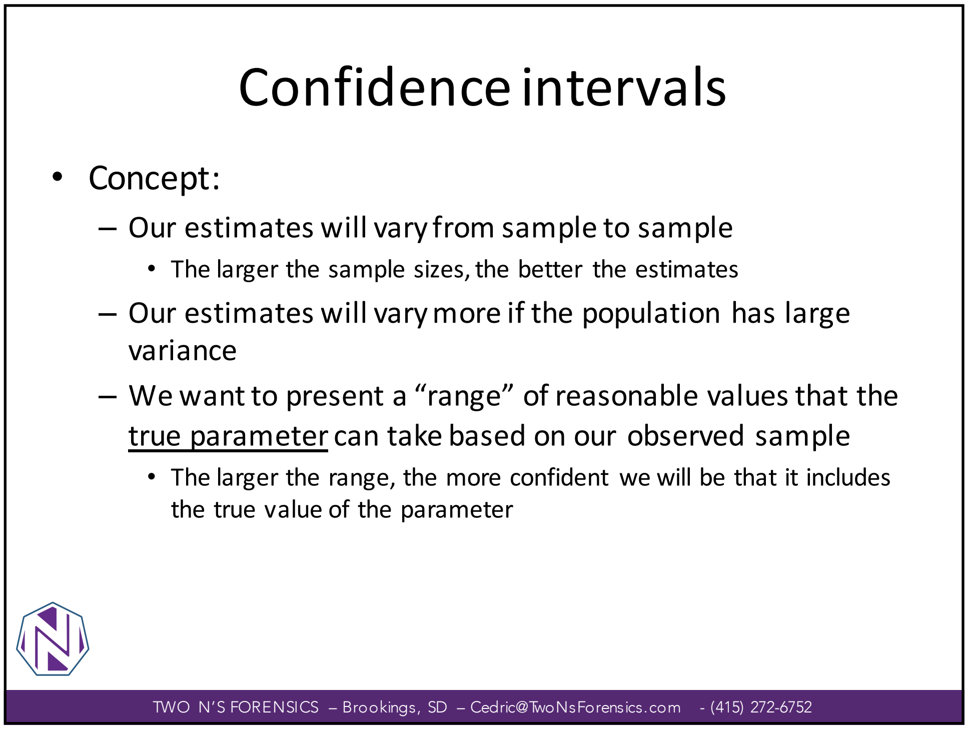
- Population
  - "True value" of the parameter(s)
  - Unknown
  - Is fixed
  - Denoted using Greek alphabet
    - $\mu, \sigma, \beta, \ldots$

- Sample
  - Estimate of the value of the parameter(s)
  - Can be calculated
  - Varies from sample to sample
  - Denoted using Latin alphabet
    - $\bar{X}, S, b, \ldots$
  - Or using ^
    - $\hat{\mu}, \hat{\sigma}, \hat{\beta}, \ldots$



Estimates

- Population
  - Proportion
    $$P = \frac{X}{N}$$
    where *X* is the count of successes in *N*
  - Mean
    $$\mu = \frac{1}{N}\Sigma X_i$$
  - Variance
    $$\sigma^2 = \frac{1}{N}\Sigma(X_i - \mu)^2$$
- Sample
  - (Sample) proportion
    $$\hat{P} = \frac{X}{n},$$
    where *X* is the count of successes in *n* samples
  - (Sample) mean
    $$\hat{\mu} = \bar{X} = \frac{1}{n}\Sigma X_i$$
  - (Sample) variance
    $$\widehat{\sigma^2} = S^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})^2$$

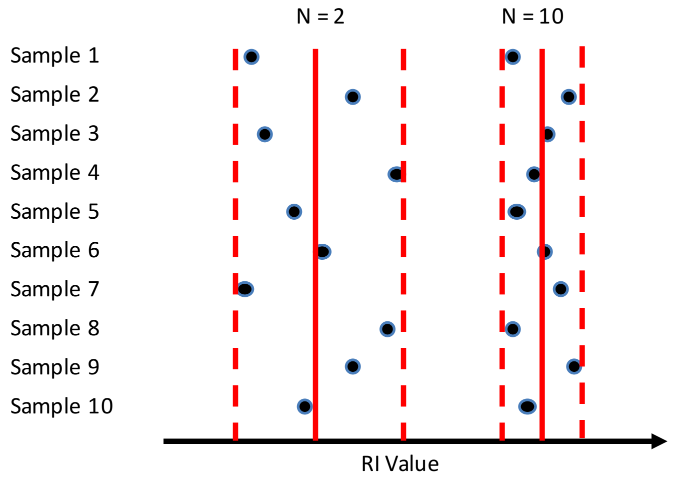


Confidence intervals

- Concept:
  - Our estimates will vary from sample to sample
    - The larger the sample sizes, the better the estimates
  - Our estimates will vary more if the population has large variance
  - We want to present a "range" of reasonable values that the true parameter can take based on our observed sample
    - The larger the range, the more confident we will be that it includes the true value of the parameter

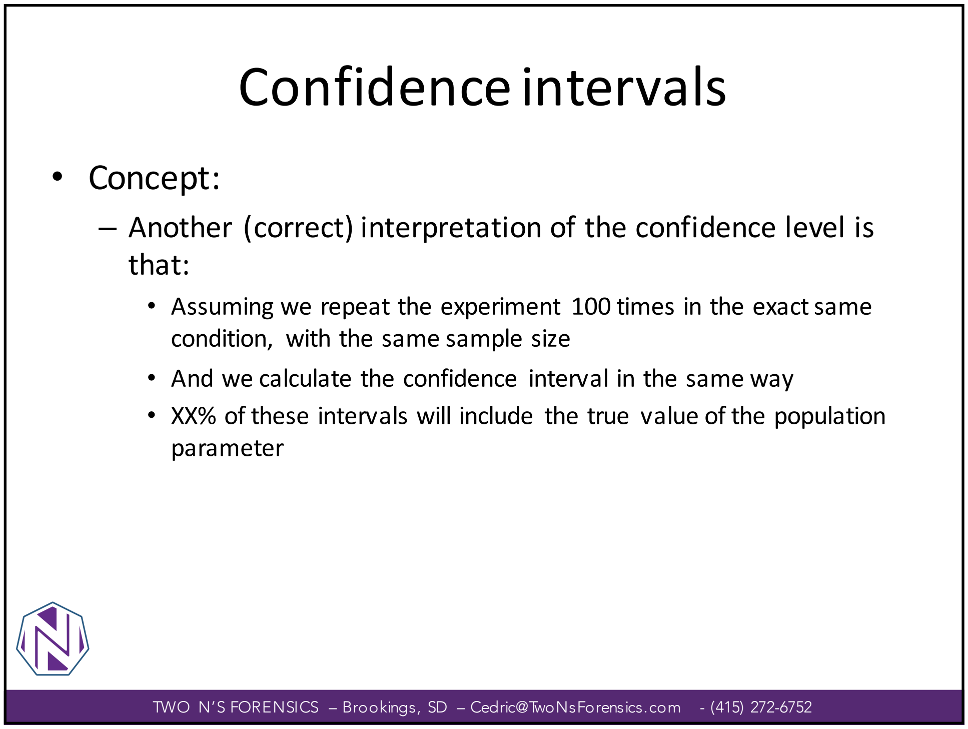
Confidence intervals

- Example (refractive index of glass):



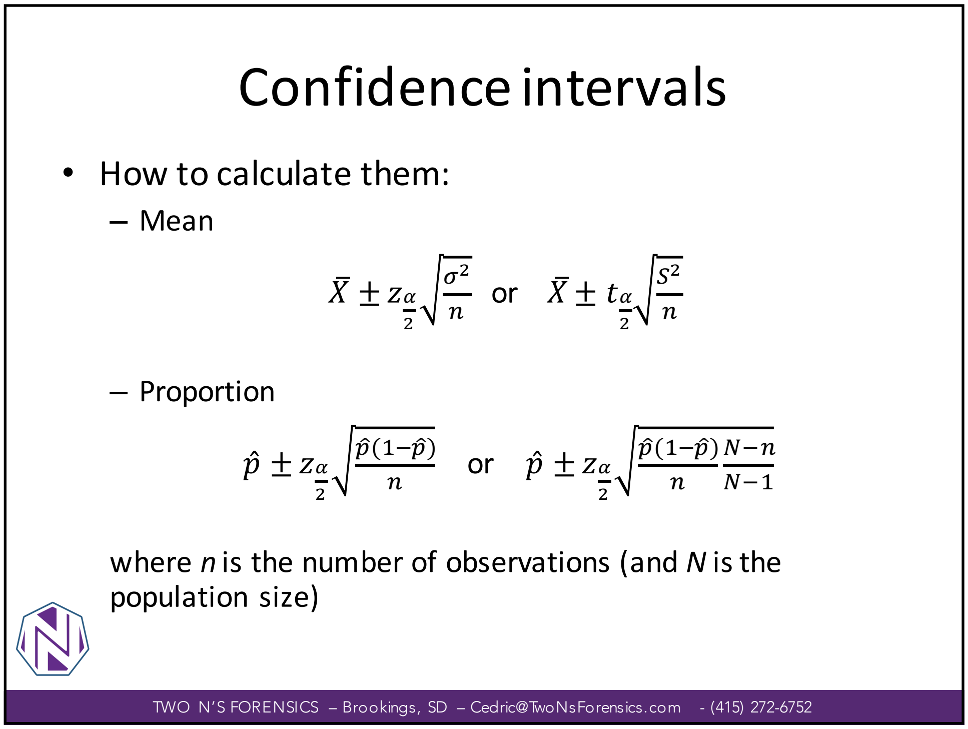
Confidence intervals

- Concept:
  - Ultimately, the range of value that we will propose is a function of:
    - A pre-defined level of confidence that we want to convey
    - Sample size
    - Variance
  - A confidence **level** represents *the confidence of the researcher that the <u>true value of the parameter</u> is included <u>within the reported range of values</u>*
    - It is not a measure of probability: **it is not the probability** that the true value of the parameter is within the range
    - **It is the confidence** of the researcher that it is



Confidence intervals

- Concept:
  - Another (correct) interpretation of the confidence level is that:
    - Assuming we repeat the experiment 100 times in the exact same condition, with the same sample size
    - And we calculate the confidence interval in the same way
    - XX% of these intervals will include the true value of the population parameter



Confidence intervals

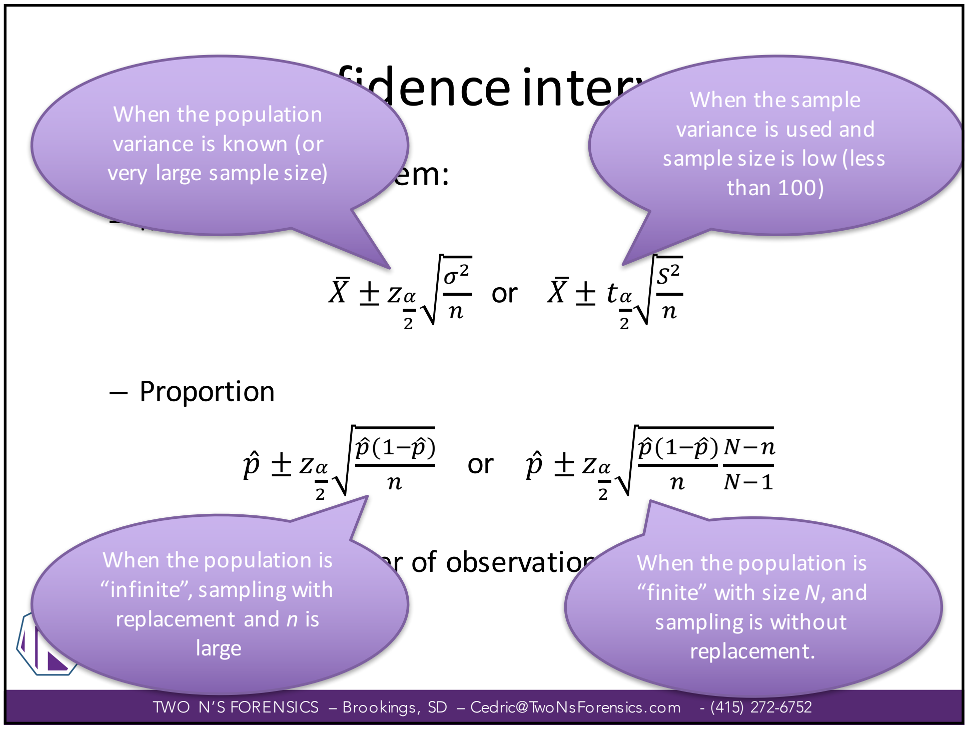
- How to calculate them:
  - Mean

$$\bar{X} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \quad \text{or} \quad \bar{X} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$$

  - Proportion

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{or} \quad \hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N-1}}$$

where *n* is the number of observations (and *N* is the population size)



fidence inter

When the population variance is known (or very large sample size)

When the sample variance is used and sample size is low (less than 100)

em:

$$\bar{X} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \quad \text{or} \quad \bar{X} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$$

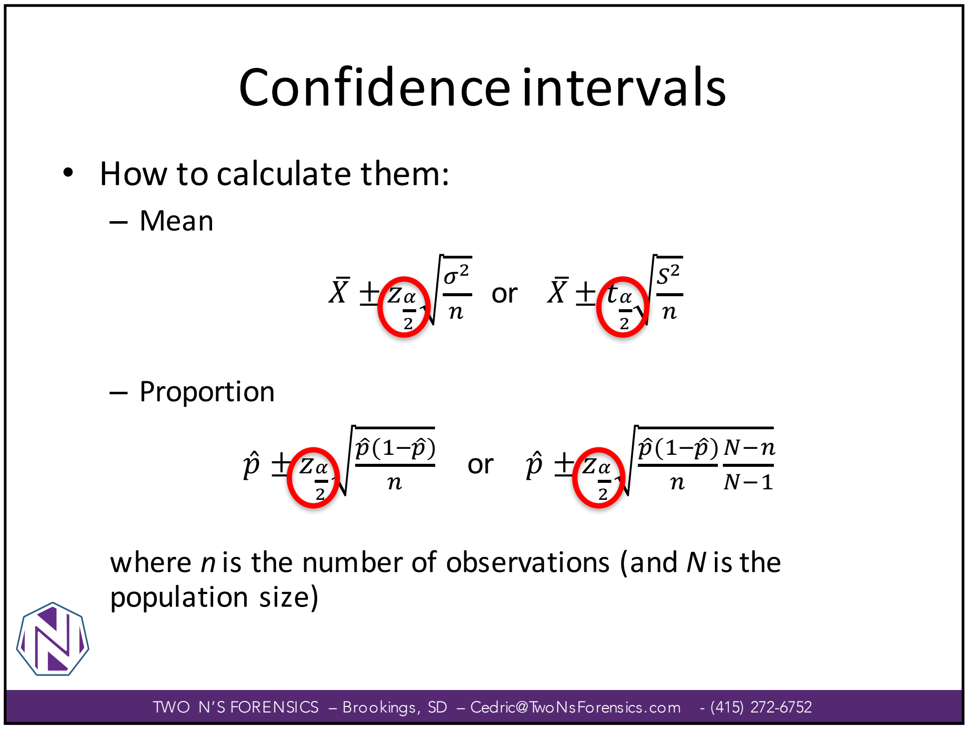
– Proportion

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{or} \quad \hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N-1}}$$

When the population is "infinite", sampling with replacement and *n* is large

r of observatio

When the population is "finite" with size *N*, and sampling is without replacement.



Confidence intervals

- How to calculate them:
  - Mean

$$\bar{X} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \quad \text{or} \quad \bar{X} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$$

  - Proportion

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{or} \quad \hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N-1}}$$

where *n* is the number of observations (and *N* is the population size)

Confidence interval

- Say we want to calculate the CI for the mean RI of a window based on a sample of fragments
  - Case 1 (many fragments: >100 or we “know” the variance of the RI in the window)

$$\sigma = \sqrt{\sigma^2} = 0.00005$$

$$\bar{X} = \frac{1}{200}\sum_{i=1}^{200} X_i = 1.5345$$

$$n = 200$$

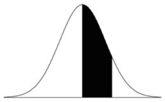
$$\bar{X} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}$$

We want to be 95% confident

-> this means that $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$

Confidence interval

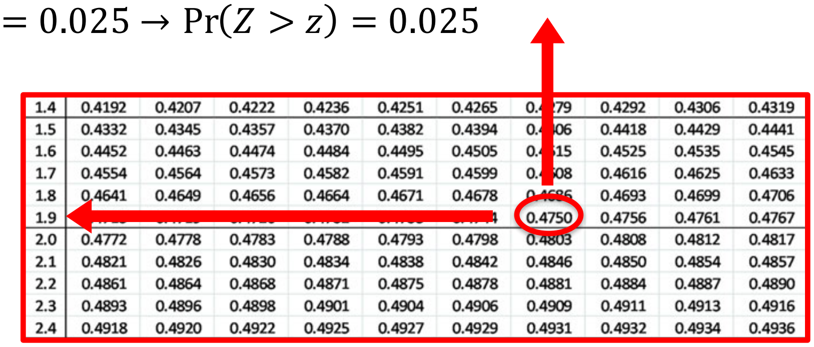
Table 1 Standard normal probabilities (area between 0 and z)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

Confidence interval

$$\frac{\alpha}{2} = 0.025 \rightarrow \Pr(Z > z) = 0.025$$



1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	[illegible]	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	[illegible]	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	[illegible]	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	[illegible]	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

$$\frac{\alpha}{2} = 0.025 \rightarrow \Pr(Z > z) = 0.025 \rightarrow z = 1.96$$

Confidence interval

- Say we want to calculate the CI for the mean RI of a window based on a sample of fragments
  - Case 1 (many fragments: >100 or we "know" the variance of the RI in the window)

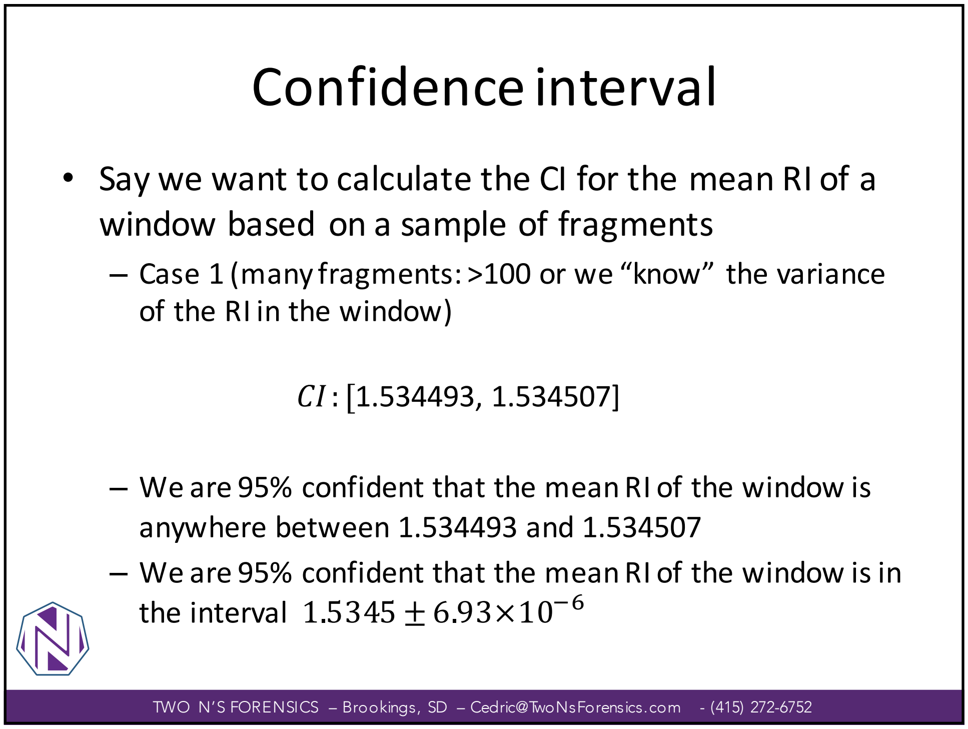
$$\sigma = \sqrt{\sigma^2} = 0.00005$$

$$\bar{X} = \frac{1}{200}\sum_{i=1}^{200} X_i = 1.5345$$

$$n = 200$$

$$1.5345 \pm 1.96 \frac{0.00005}{\sqrt{200}}$$

CI : [1.534493, 1.534507]

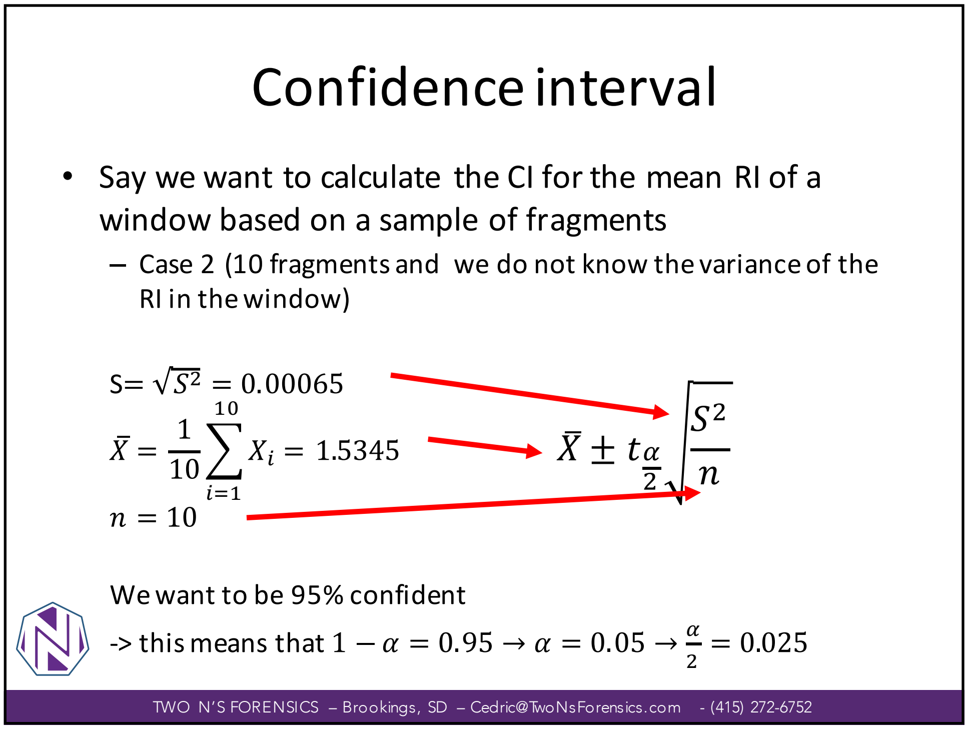


Confidence interval

- Say we want to calculate the CI for the mean RI of a window based on a sample of fragments
  - Case 1 (many fragments: >100 or we “know” the variance of the RI in the window)

CI : [1.534493, 1.534507]

  - We are 95% confident that the mean RI of the window is anywhere between 1.534493 and 1.534507
  - We are 95% confident that the mean RI of the window is in the interval $1.5345 \pm 6.93 \times 10^{-6}$



Confidence interval

- Say we want to calculate the CI for the mean RI of a window based on a sample of fragments
  - Case 2 (10 fragments and we do not know the variance of the RI in the window)

S= $\sqrt{S^2} = 0.00065$

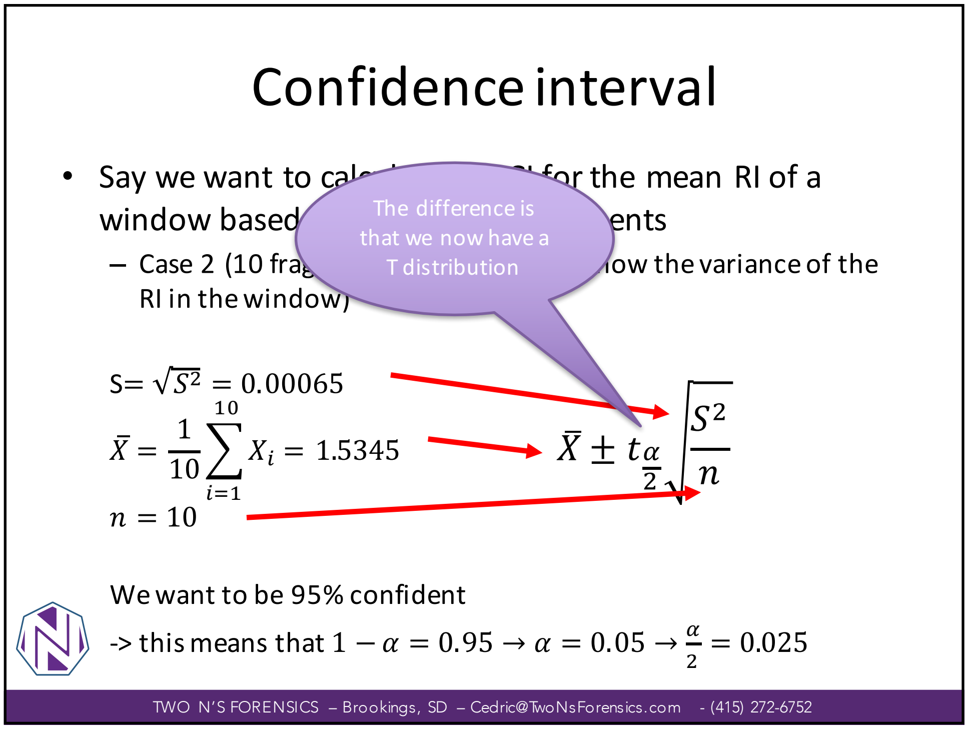
$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 1.5345$$

$n = 10$

$$\bar{X} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

We want to be 95% confident

-> this means that $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$



Confidence interval

- Say we want to cal… …I for the mean RI of a window based … …ents
  - Case 2 (10 frag… …now the variance of the RI in the window)

The difference is that we now have a T distribution

S= $\sqrt{S^2} = 0.00065$

$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 1.5345$

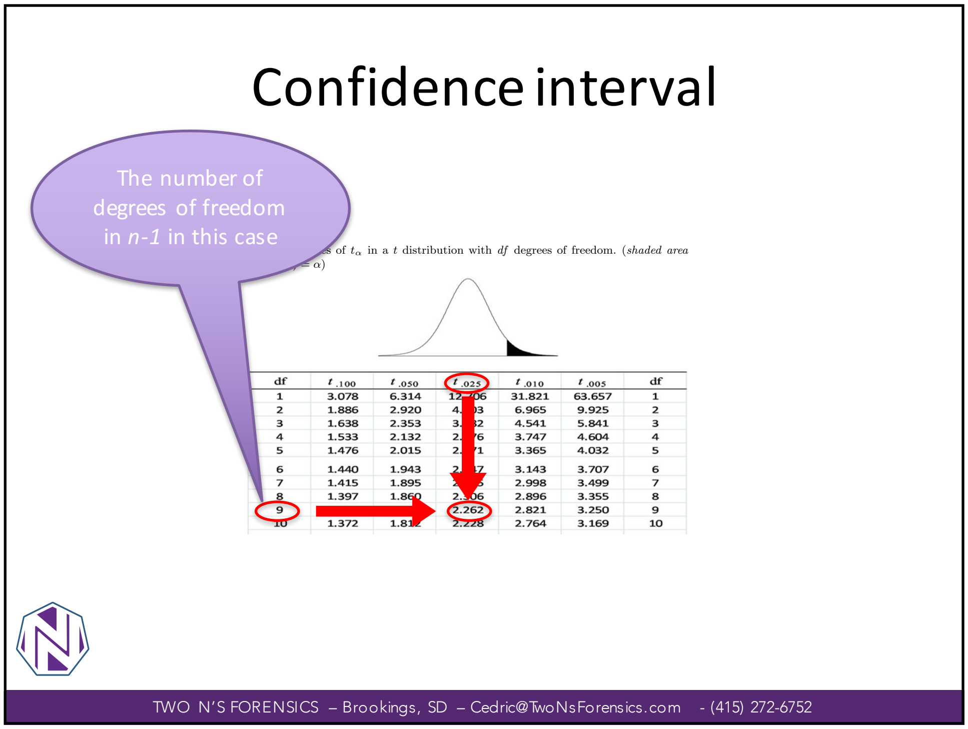
$n = 10$

$$\bar{X} \pm t_{\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}}$$

We want to be 95% confident

-> this means that $1 - \alpha = 0.95 \rightarrow \alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Confidence interval

The number of degrees of freedom in *n-1* in this case

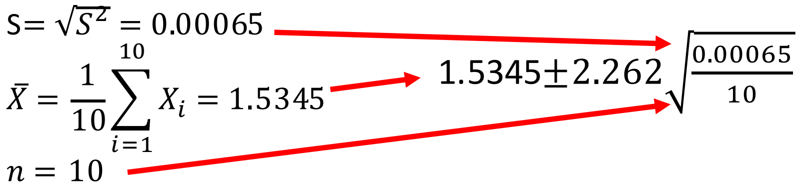
…s of t_α in a t distribution with df degrees of freedom. (*shaded area* …$= \alpha$)

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	[illegible]	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	[illegible]	[illegible]	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Confidence interval

- Say we want to calculate the CI for the mean RI of a window based on a sample of fragments
  - Case 2 (10 fragments and we do not know the variance of the RI in the window)



S= $\sqrt{S^2} = 0.00065$

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 1.5345$$

$n = 10$

$$1.5345 \pm 2.262\sqrt{\frac{0.00065}{10}}$$

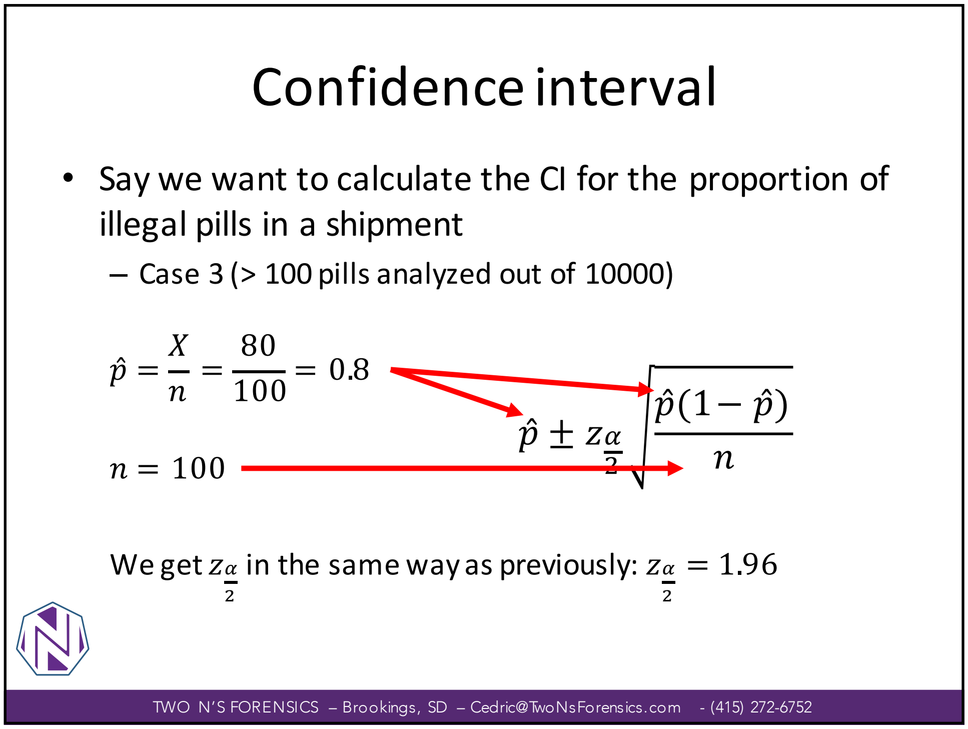
CI[1.534035,1.534965] vs. CI: [1.534493, 1.534507]

Confidence interval

- Say we want to calculate the CI for the mean RI of a window based on a sample of fragments
  - Case 2 (10 fragments and we do not know the variance of the RI in the window)

CI[1.534035,1.534965]

  - We are 95% confident that the mean RI of the window is anywhere between1.534035 and1.534965
  - We are 95% confident that the mean RI of the window is in the interval $1.5345 \pm 4.6\times10^{-4}$ (vs. 6.93×10^{-6} in 1)



Confidence interval

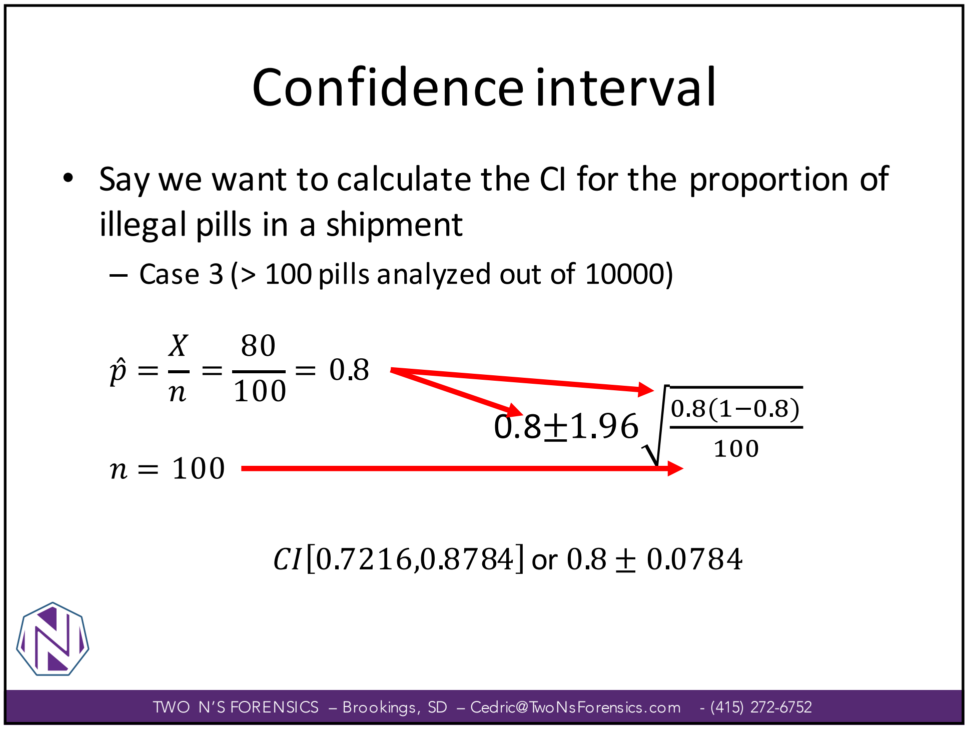
- Say we want to calculate the CI for the proportion of illegal pills in a shipment
  - Case 3 (> 100 pills analyzed out of 10000)

$$\hat{p} = \frac{X}{n} = \frac{80}{100} = 0.8$$

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$n = 100$$

We get $z_{\frac{\alpha}{2}}$ in the same way as previously: $z_{\frac{\alpha}{2}} = 1.96$



Confidence interval

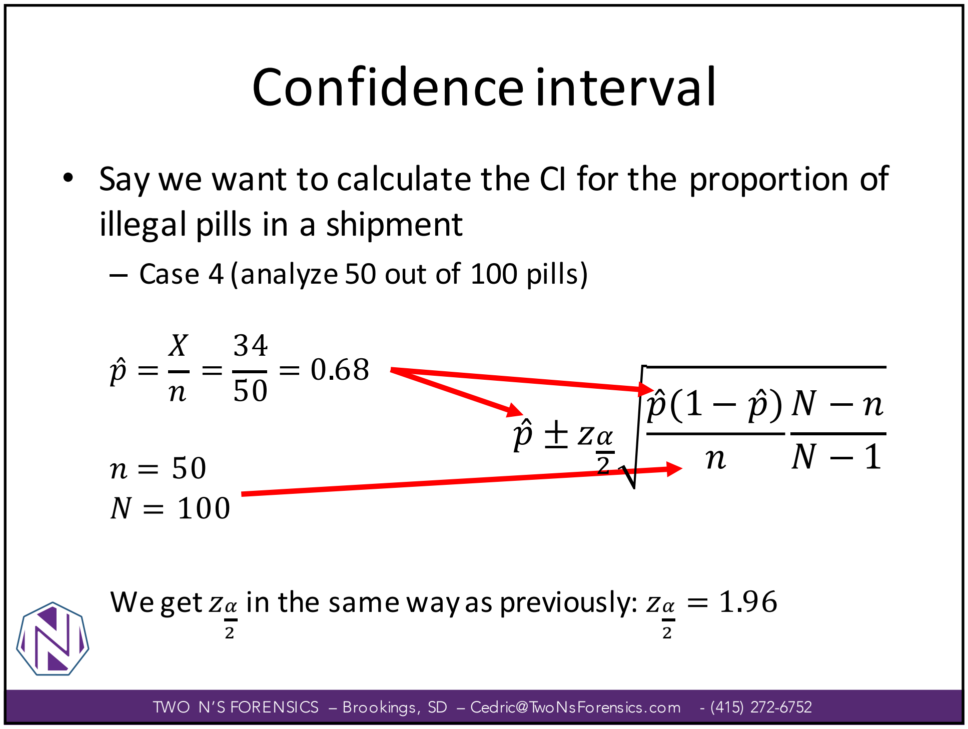
- Say we want to calculate the CI for the proportion of illegal pills in a shipment
  - Case 3 (> 100 pills analyzed out of 10000)

$$\hat{p} = \frac{X}{n} = \frac{80}{100} = 0.8$$

$$0.8 \pm 1.96 \sqrt{\frac{0.8(1-0.8)}{100}}$$

$$n = 100$$

$CI[0.7216{,}0.8784]$ or 0.8 ± 0.0784



Confidence interval

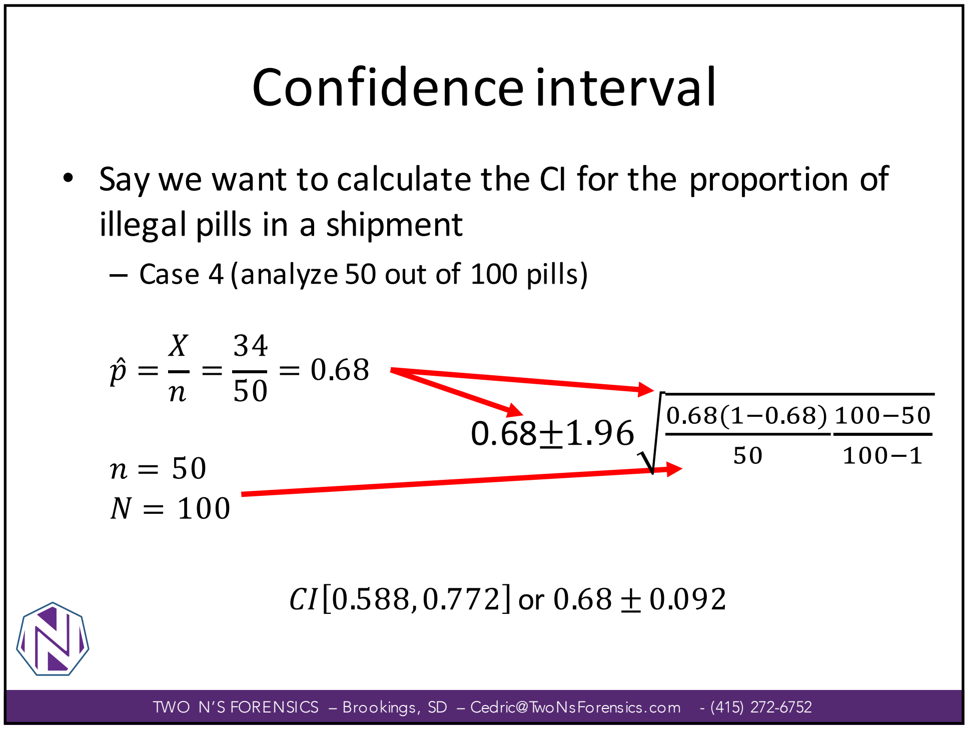
- Say we want to calculate the CI for the proportion of illegal pills in a shipment
  - Case 4 (analyze 50 out of 100 pills)

$$\hat{p} = \frac{X}{n} = \frac{34}{50} = 0.68$$

$$n = 50$$
$$N = 100$$

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}}$$

We get $z_{\frac{\alpha}{2}}$ in the same way as previously: $z_{\frac{\alpha}{2}} = 1.96$



Confidence interval

- Say we want to calculate the CI for the proportion of illegal pills in a shipment
  - Case 4 (analyze 50 out of 100 pills)

$$\hat{p} = \frac{X}{n} = \frac{34}{50} = 0.68$$

$$n = 50$$
$$N = 100$$

$$0.68 \pm 1.96 \sqrt{\frac{0.68(1-0.68)}{50} \frac{100-50}{100-1}}$$

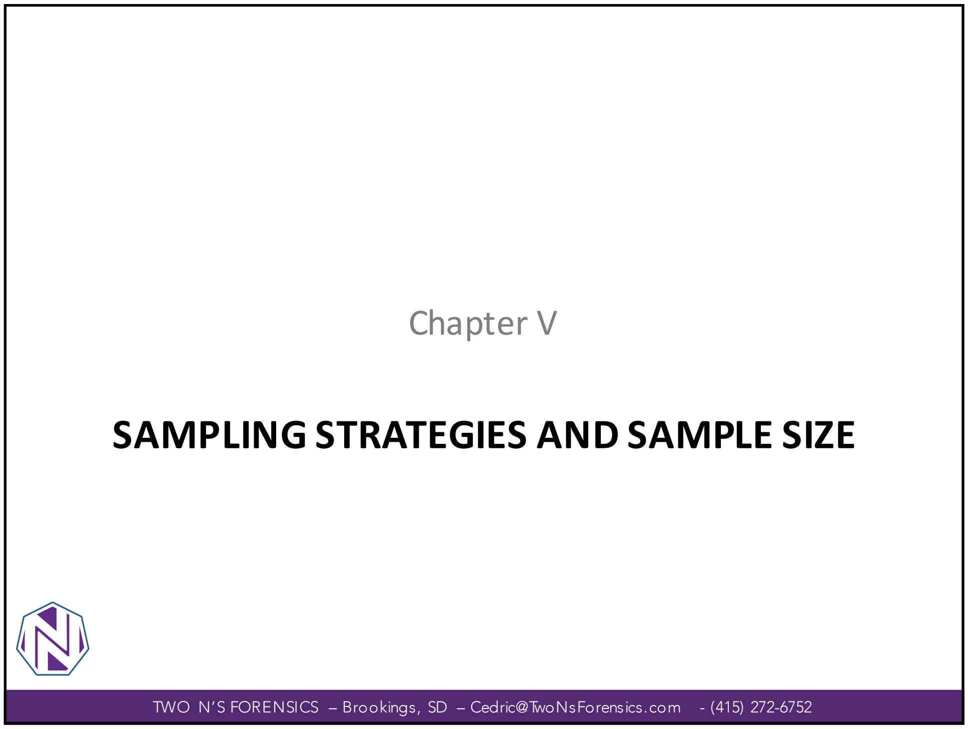
$CI[0.588, 0.772]$ or 0.68 ± 0.092

Take home messages

- There is always a population of results for an experiment
- We cannot observe all of them, so we are limited to a sample
- We use the sample to **estimate** the parameter(s) of the distribution of the observations in the population
- We can use an interval to express our confidence that the true value of a population parameter is within a certain range of value

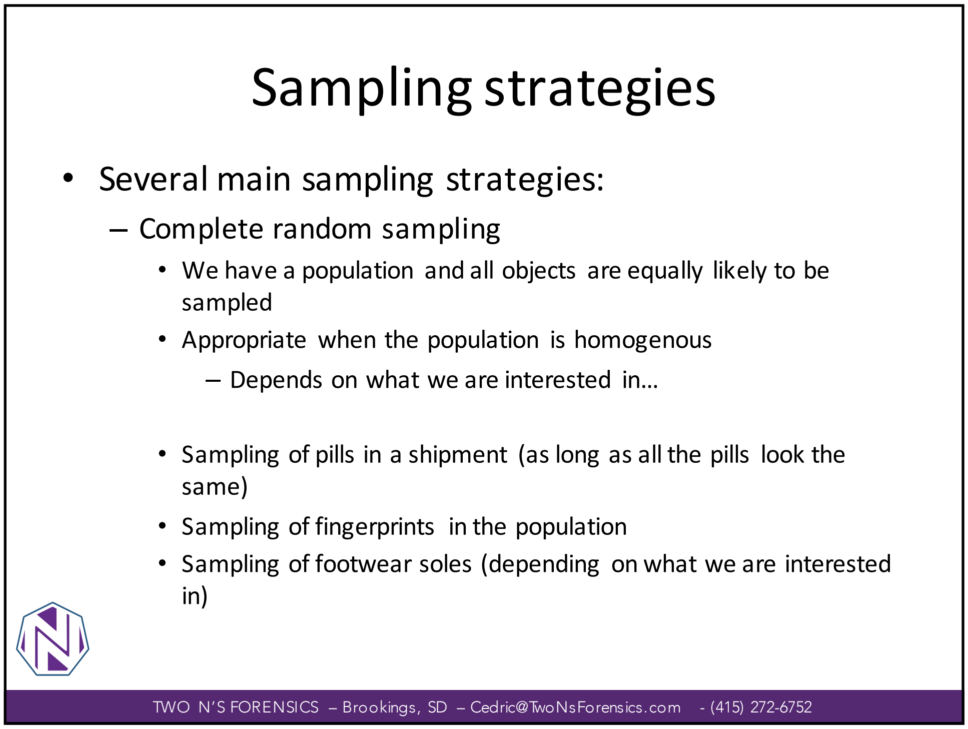
Chapter IV

EXERCISES



Chapter V

SAMPLING STRATEGIES AND SAMPLE SIZE

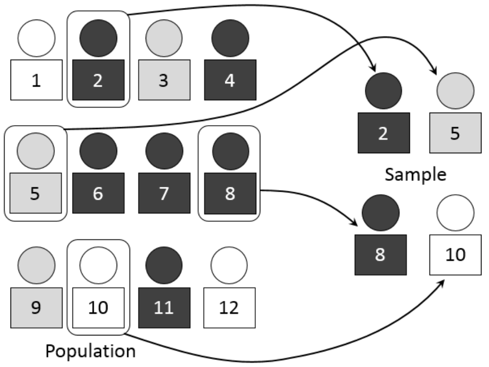


Sampling strategies

- Several main sampling strategies:
  - Complete random sampling
    - We have a population and all objects are equally likely to be sampled
    - Appropriate when the population is homogenous
      - Depends on what we are interested in…
    - Sampling of pills in a shipment (as long as all the pills look the same)
    - Sampling of fingerprints in the population
    - Sampling of footwear soles (depending on what we are interested in)

Sampling strategies

- Several main sampling strategies:
  - Complete random sampling



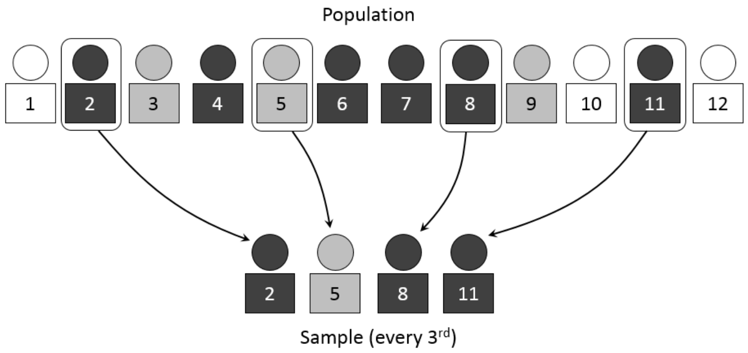
Source: https://commons.wikimedia.org/wiki/File:Simple_random_sampling.PNG

Sampling strategies

- Several main sampling strategies:
  - Systematic sampling
    - The population is organized (somehow) and objects are selected at regular intervals
    - ”phone book” sampling: take every 10th person in the book
    - Issue: you need to make sure that the object selected is measurable (e.g., the 10th person may refuse to provide material)
    - Sampling of pills in a shipment

Sampling strategies

- Several main sampling strategies:
  - Systematic sampling



Source: https://upload.wikimedia.org/wikipedia/commons/c/c4/Systematic_sampling.PNG

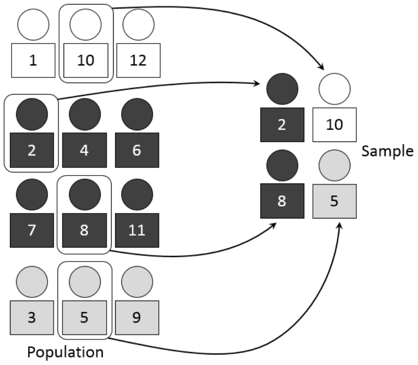
Sampling strategies

- Several main sampling strategies:
  - Stratified sampling
    - The population has classes
    - We sample each class separately with a number of sample proportional to the size of the class
    - The classes need to be disjoint

    - Sampling of DNA profile
    - Sampling of footwear sole design

Sampling strategies

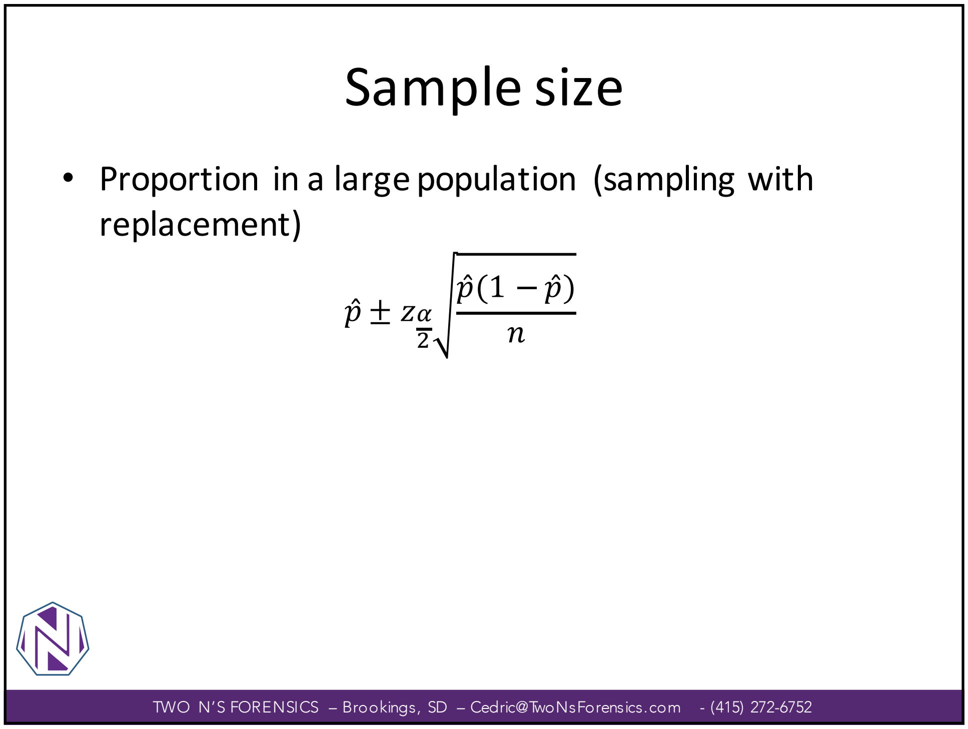
- Several main sampling strategies:
  - Stratified sampling



Source: https://upload.wikimedia.org/wikipedia/commons/f/fa/Stratified_sampling.PNG

Sample size

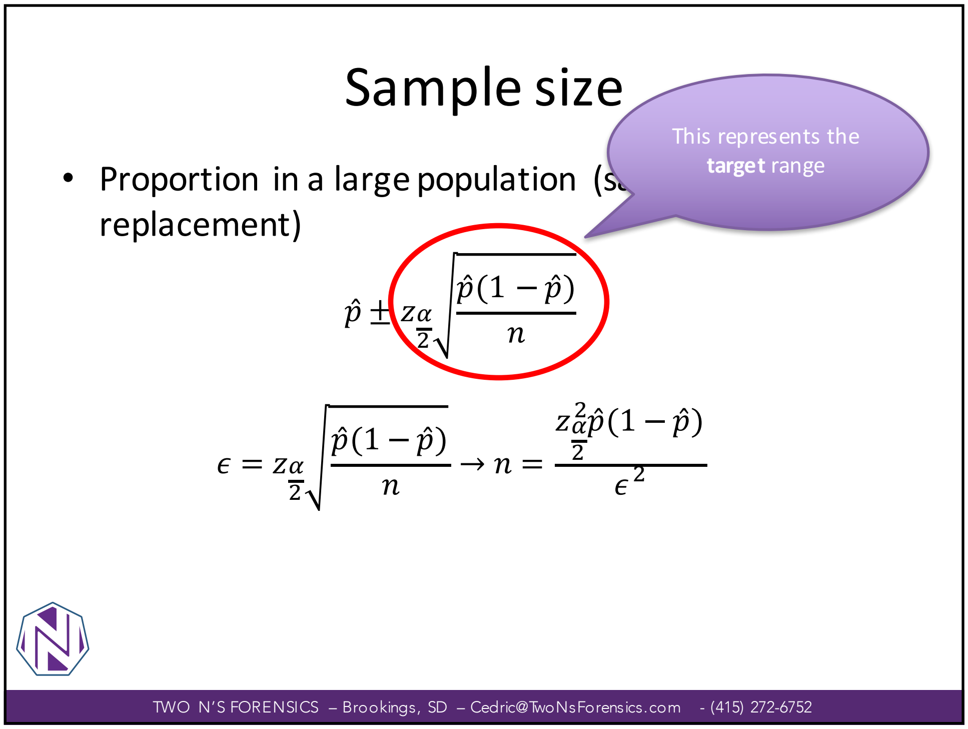
- We have seen that the confidence interval for the population parameter depends on:
  - Sample size
  - Confidence level
  - Variance of population (estimated by sample variance in most cases)
- Determination of sample size is basically a "pick 3, get the fourth one" problem
  - Pick confidence level
  - Pick range
  - Use variance
  - Get sample size



Sample size

- Proportion in a large population (sampling with replacement)

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



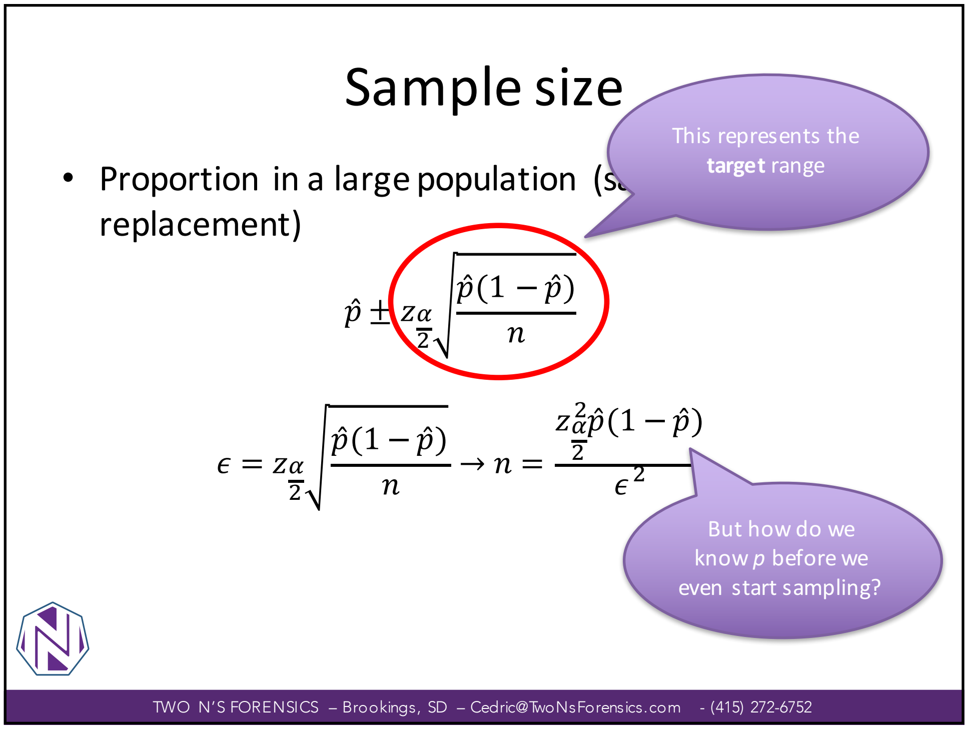
Sample size

This represents the **target** range

- Proportion in a large population (s replacement)

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow n = \frac{z_{\frac{\alpha}{2}}^{2}\hat{p}(1-\hat{p})}{\epsilon^{2}}$$



Sample size

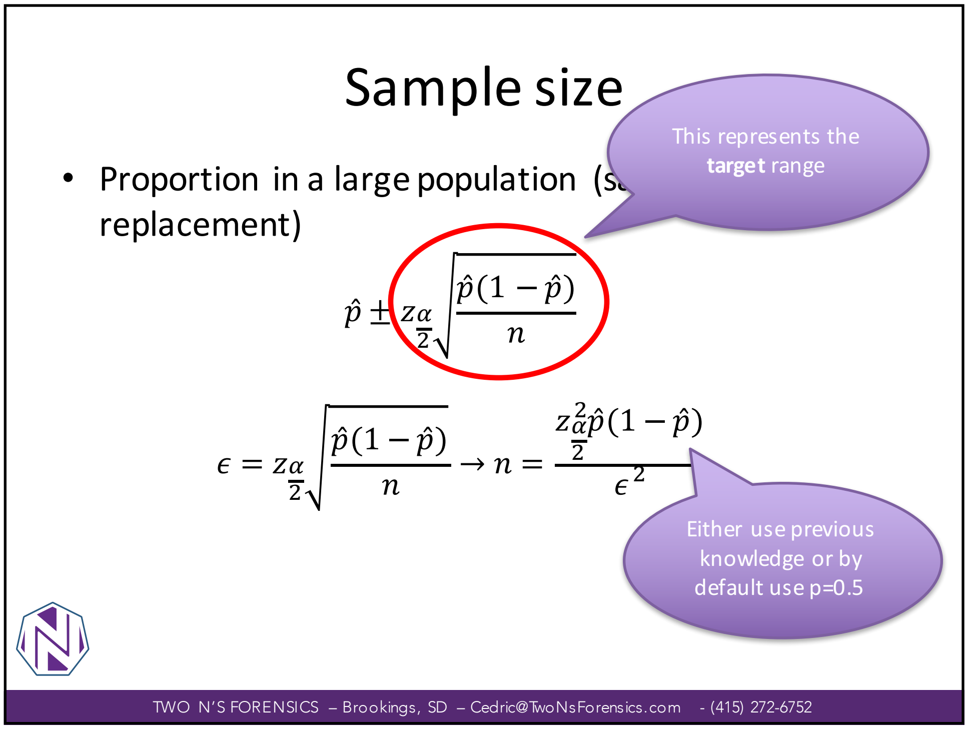
- Proportion in a large population (s replacement)

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

This represents the **target** range

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\epsilon^2}$$

But how do we know *p* before we even start sampling?



Sample size

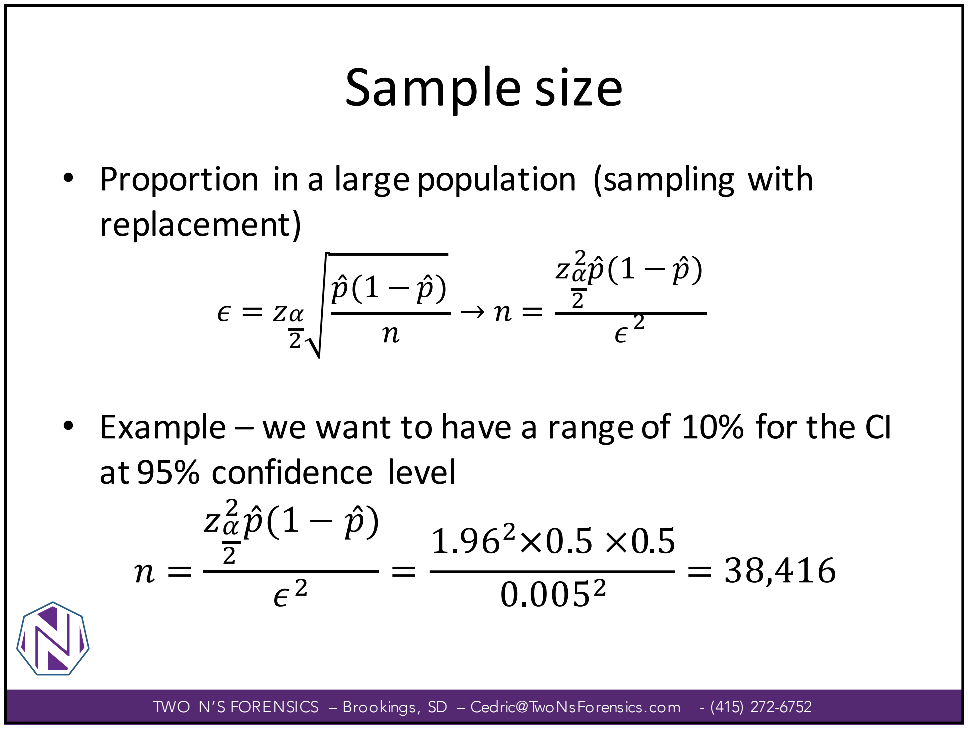
- Proportion in a large population (s replacement)

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

This represents the **target** range

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\epsilon^2}$$

Either use previous knowledge or by default use p=0.5



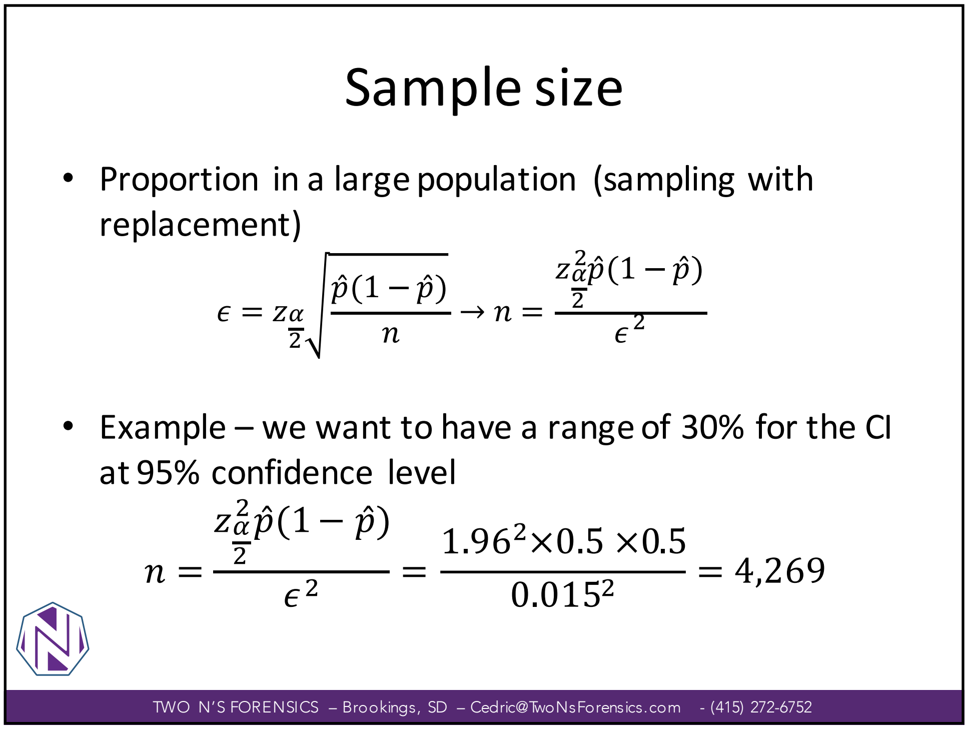
Sample size

- Proportion in a large population (sampling with replacement)

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\epsilon^2}$$

- Example – we want to have a range of 10% for the CI at 95% confidence level

$$n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\epsilon^2} = \frac{1.96^2 \times 0.5 \times 0.5}{0.005^2} = 38{,}416$$



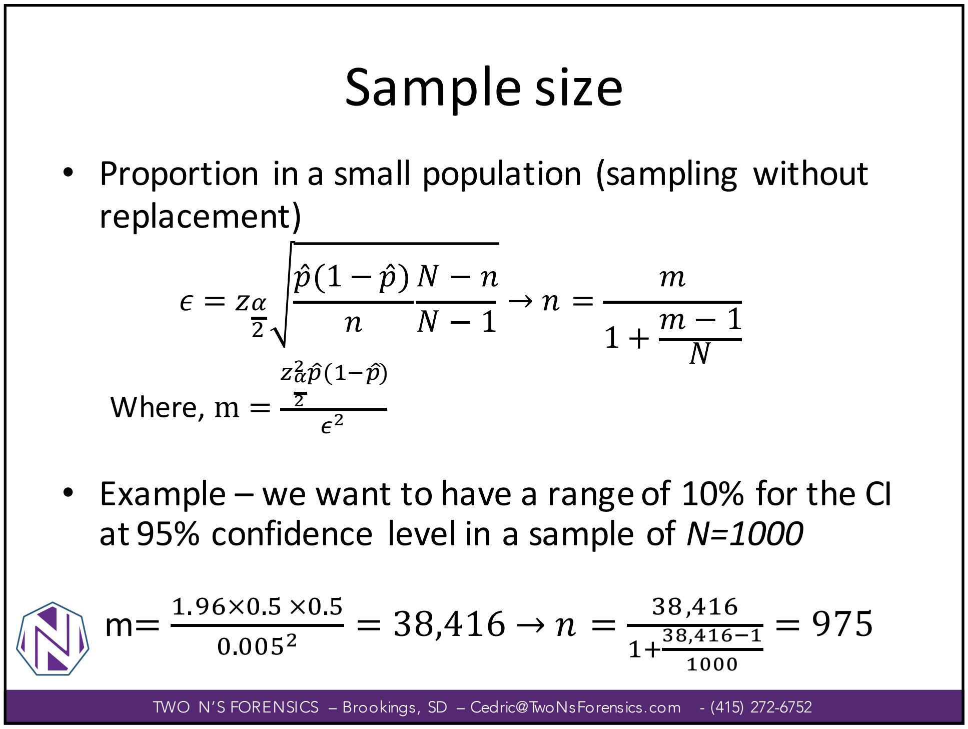
Sample size

- Proportion in a large population (sampling with replacement)

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\epsilon^2}$$

- Example – we want to have a range of 30% for the CI at 95% confidence level

$$n = \frac{z^2_{\frac{\alpha}{2}}\hat{p}(1-\hat{p})}{\epsilon^2} = \frac{1.96^2 \times 0.5 \times 0.5}{0.015^2} = 4{,}269$$



Sample size

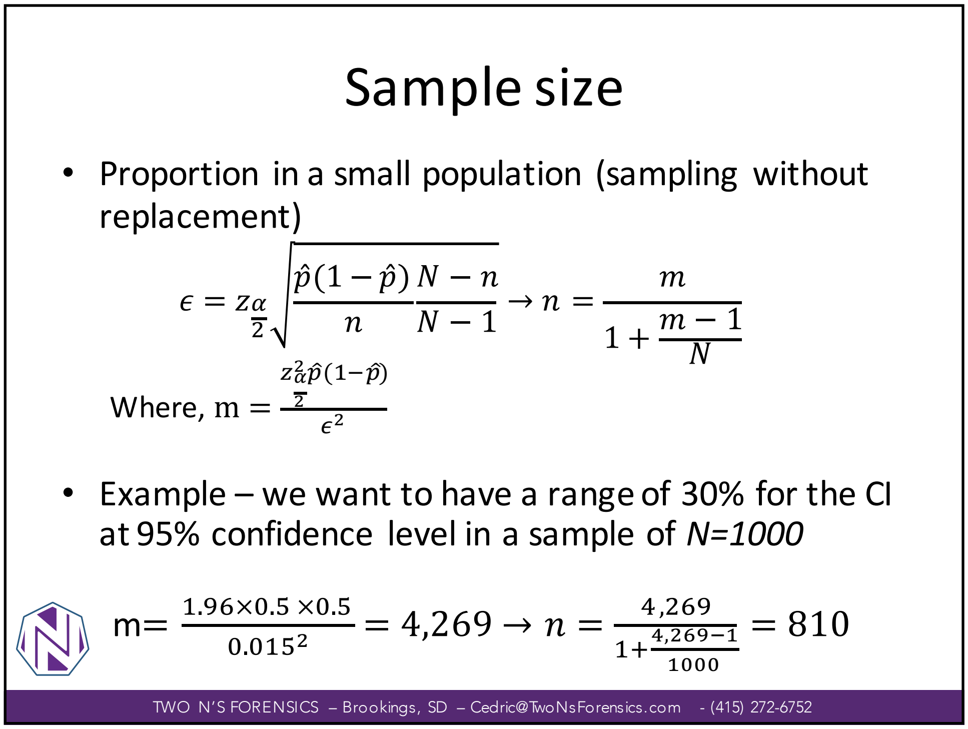
- Proportion in a small population (sampling without replacement)

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N-1}} \rightarrow n = \frac{m}{1+\frac{m-1}{N}}$$

Where, $\mathrm{m} = \frac{z_{\frac{\alpha}{2}}^2\hat{p}(1-\hat{p})}{\epsilon^2}$

- Example – we want to have a range of 10% for the CI at 95% confidence level in a sample of *N=1000*

$$\mathrm{m}= \frac{1.96\times 0.5\ \times 0.5}{0.005^2} = 38{,}416 \rightarrow n = \frac{38{,}416}{1+\frac{38{,}416-1}{1000}} = 975$$



Sample size

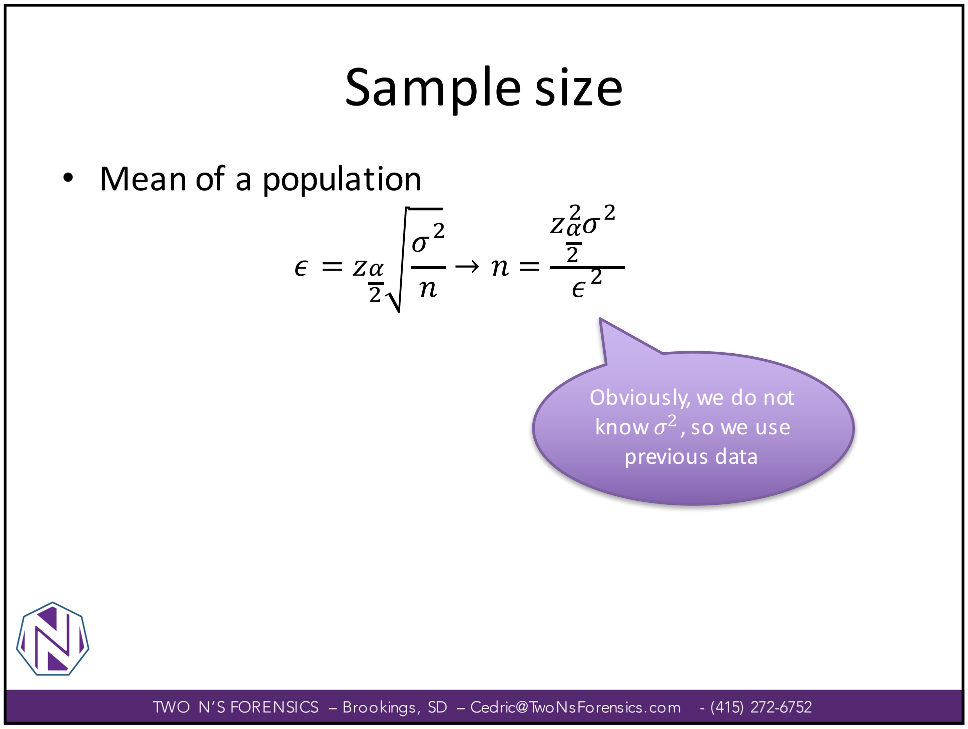
- Proportion in a small population (sampling without replacement)

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N-1}} \rightarrow n = \frac{m}{1+\frac{m-1}{N}}$$

Where, $\mathrm{m} = \frac{z_{\frac{\alpha}{2}}^2\hat{p}(1-\hat{p})}{\epsilon^2}$

- Example – we want to have a range of 30% for the CI at 95% confidence level in a sample of *N=1000*

$$\mathrm{m}= \frac{1.96\times 0.5\ \times 0.5}{0.015^2} = 4{,}269 \rightarrow n = \frac{4{,}269}{1+\frac{4{,}269-1}{1000}} = 810$$

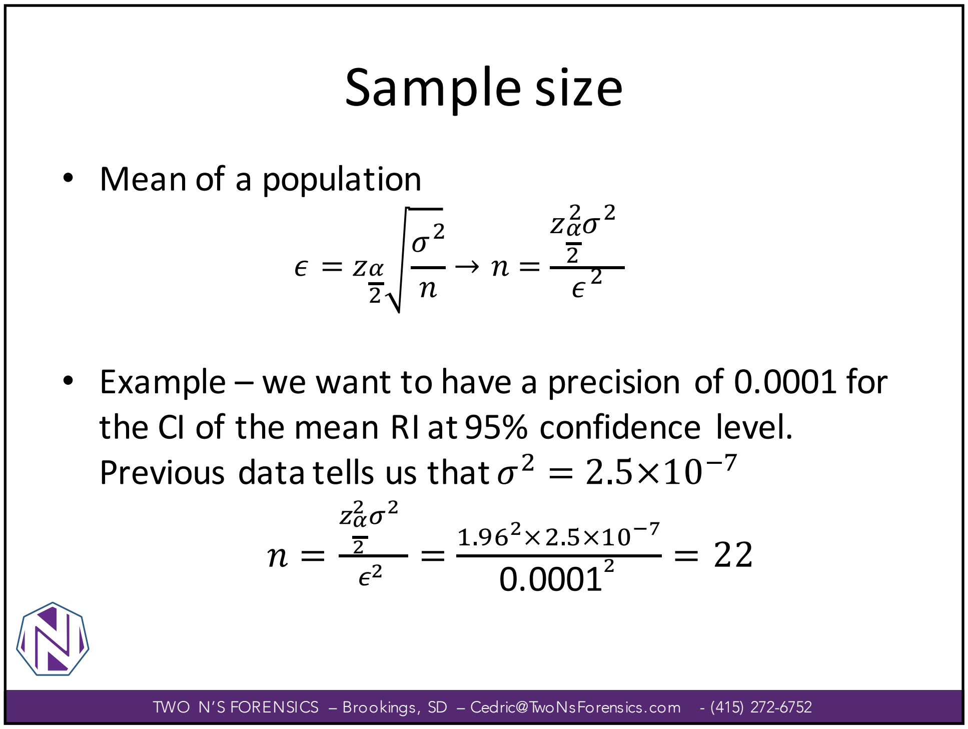


Sample size

- Mean of a population

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \to n = \frac{z^2_{\frac{\alpha}{2}}\sigma^2}{\epsilon^2}$$

Obviously, we do not know σ^2, so we use previous data



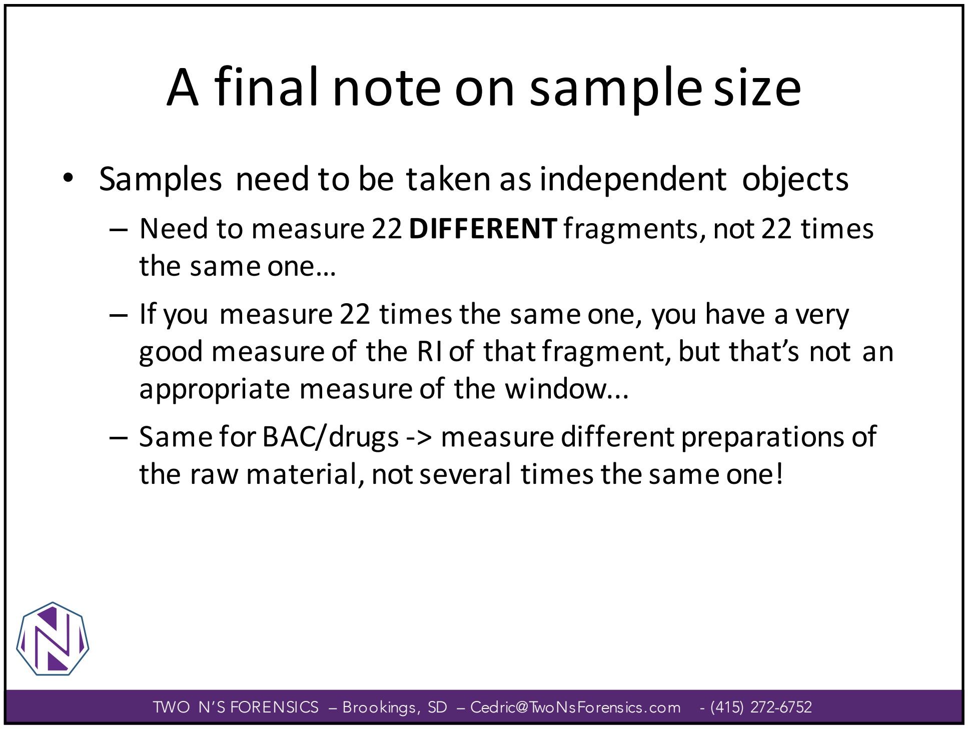
Sample size

- Mean of a population

$$\epsilon = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \to n = \frac{z^2_{\frac{\alpha}{2}}\sigma^2}{\epsilon^2}$$

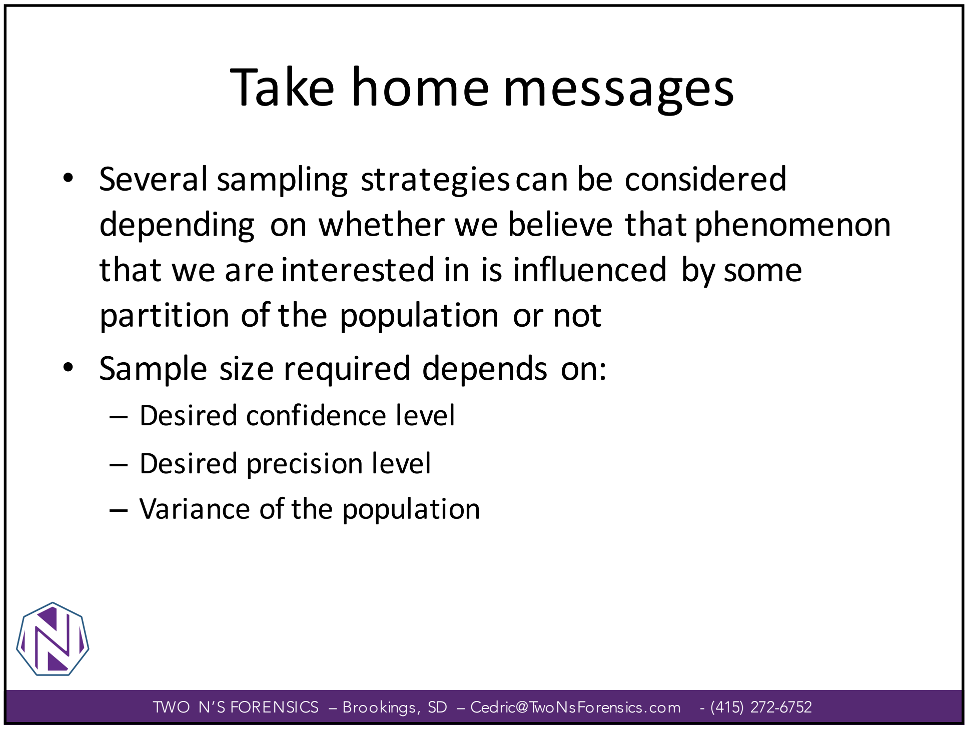
- Example – we want to have a precision of 0.0001 for the CI of the mean RI at 95% confidence level. Previous data tells us that $\sigma^2 = 2.5\times10^{-7}$

$$n = \frac{z^2_{\frac{\alpha}{2}}\sigma^2}{\epsilon^2} = \frac{1.96^2\times2.5\times10^{-7}}{0.0001^2} = 22$$



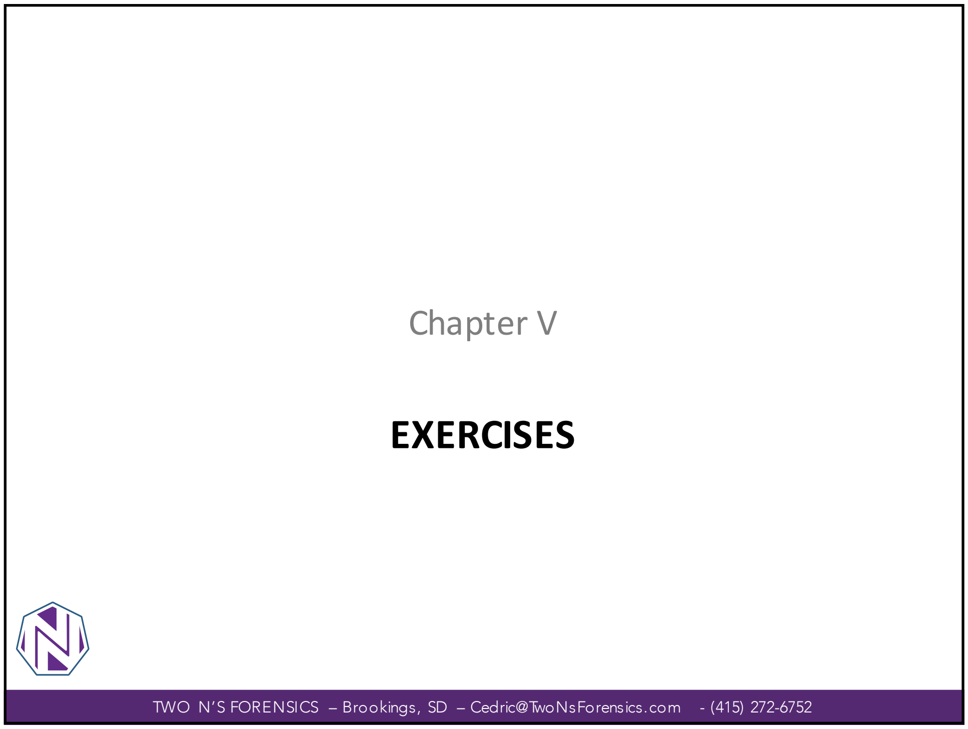
A final note on sample size

- Samples need to be taken as independent objects
  - Need to measure 22 **DIFFERENT** fragments, not 22 times the same one…
  - If you measure 22 times the same one, you have a very good measure of the RI of that fragment, but that's not an appropriate measure of the window...
  - Same for BAC/drugs -> measure different preparations of the raw material, not several times the same one!



Take home messages

- Several sampling strategies can be considered depending on whether we believe that phenomenon that we are interested in is influenced by some partition of the population or not
- Sample size required depends on:
  - Desired confidence level
  - Desired precision level
  - Variance of the population



Chapter V

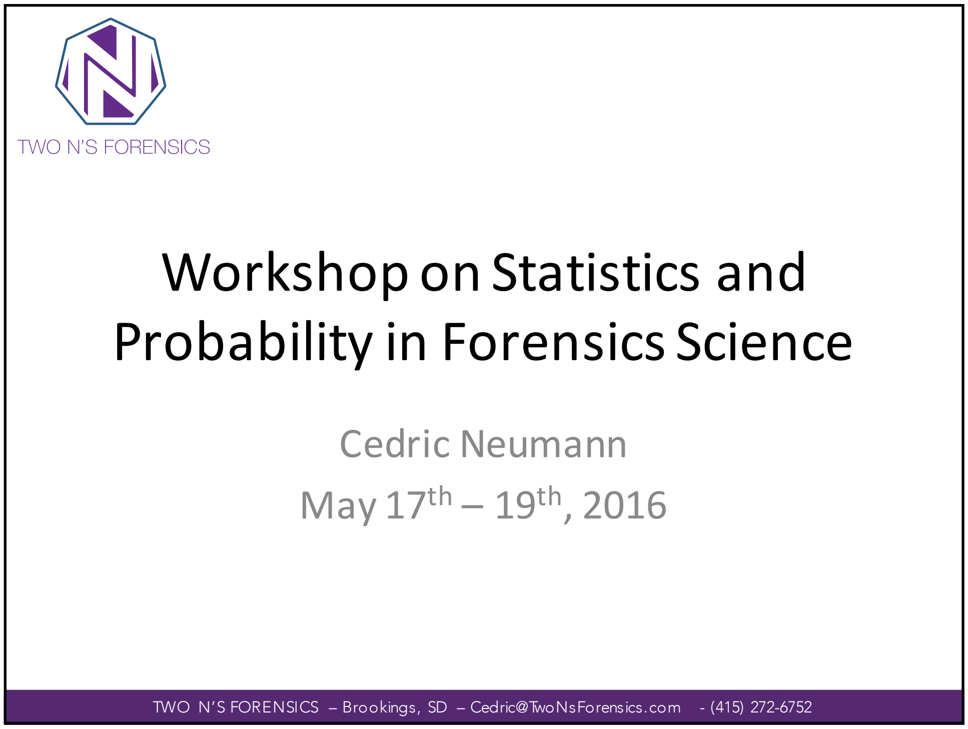
EXERCISES

Workshop on Statistics and Probability in Forensics Science

Cedric Neumann

DAY 2

MAY 18TH

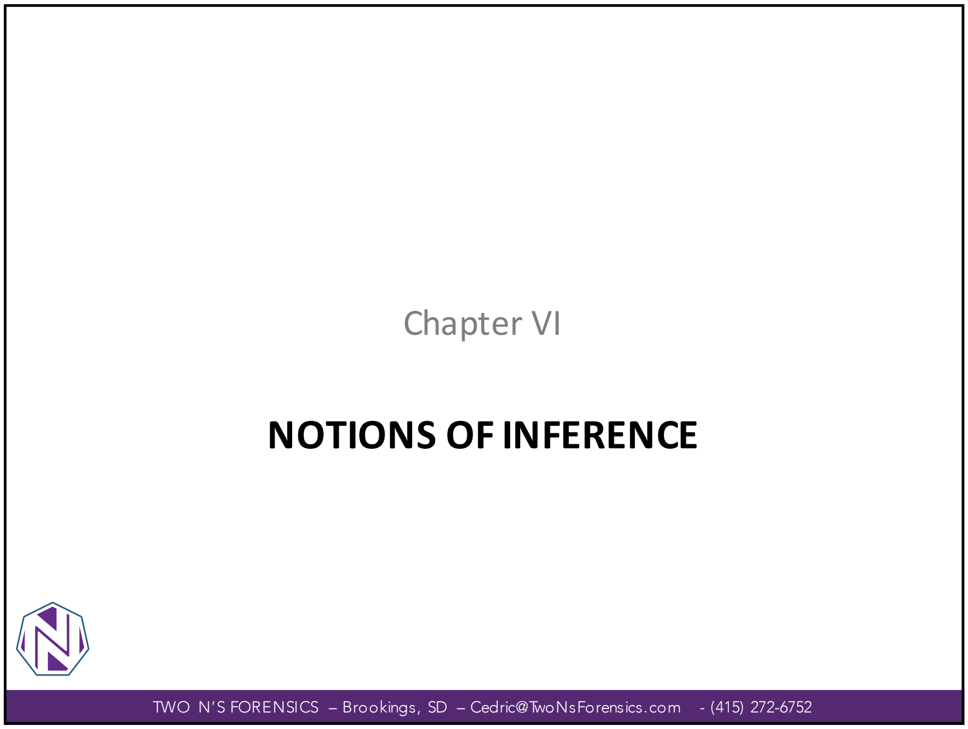


TWO N'S FORENSICS

Workshop on Statistics and Probability in Forensics Science

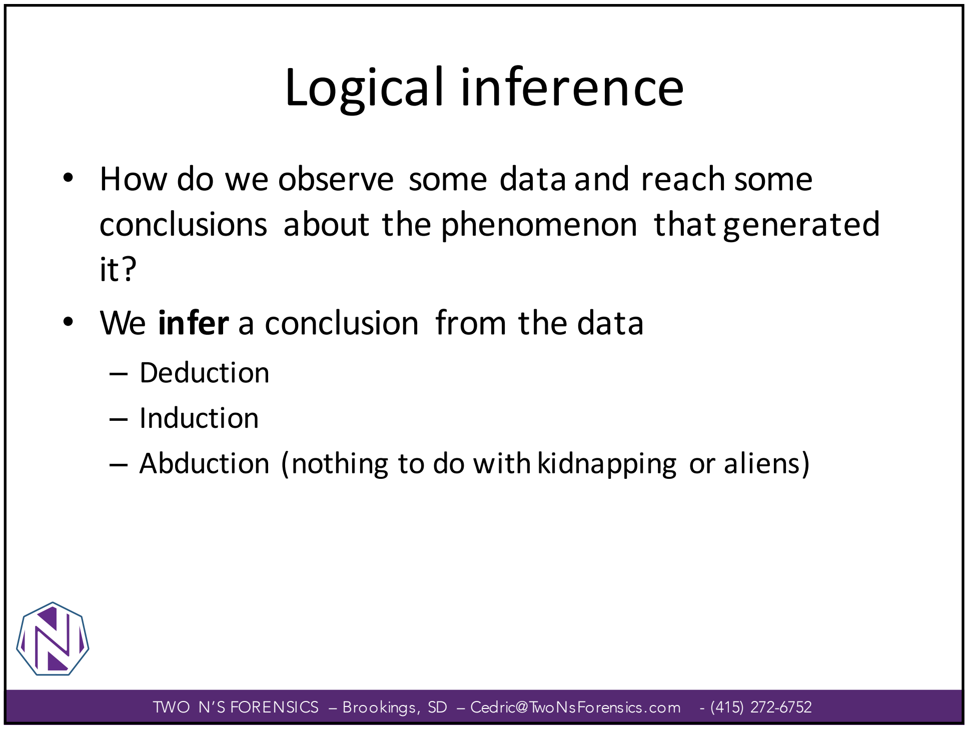
Cedric Neumann

May 17th – 19th, 2016



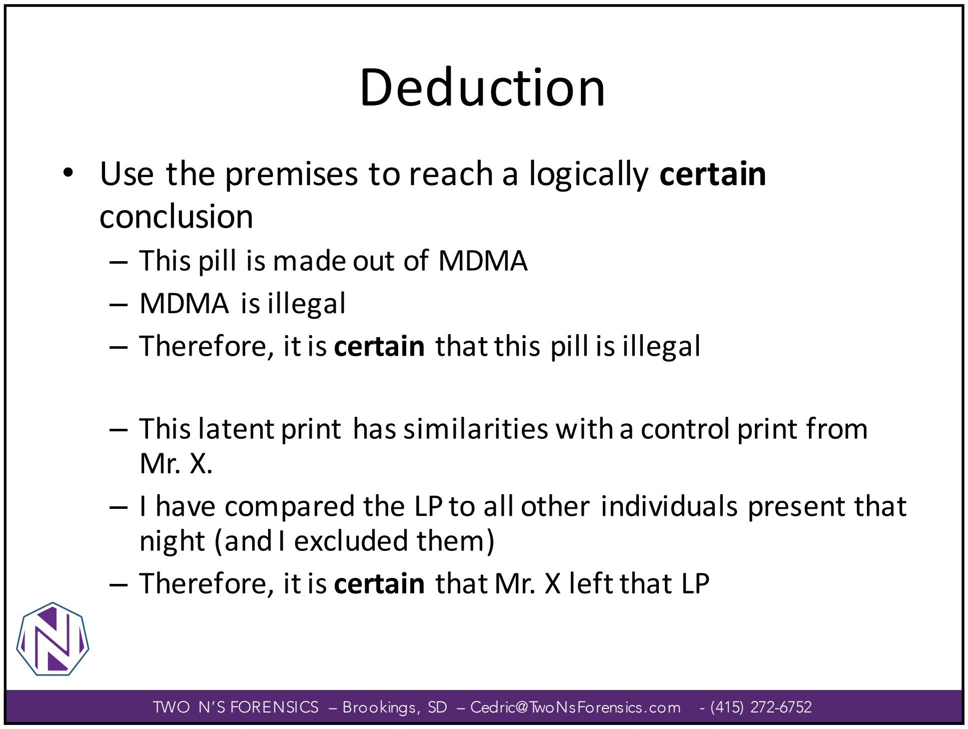
Chapter VI

NOTIONS OF INFERENCE



Logical inference

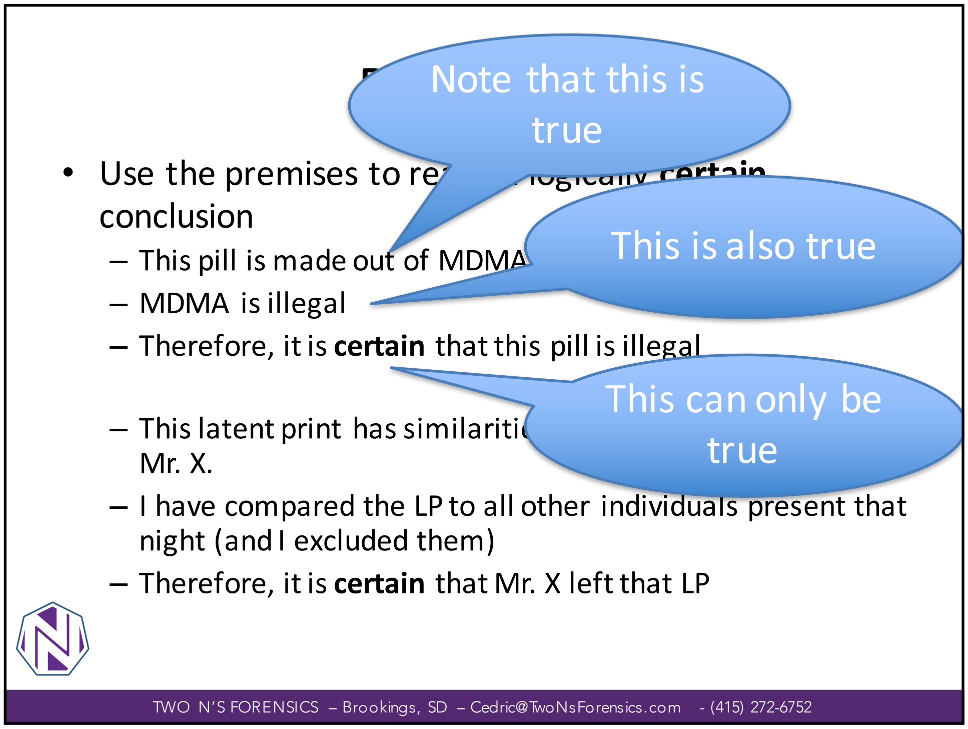
- How do we observe some data and reach some conclusions about the phenomenon that generated it?
- We **infer** a conclusion from the data
  - Deduction
  - Induction
  - Abduction (nothing to do with kidnapping or aliens)



Deduction

- Use the premises to reach a logically **certain** conclusion
  - This pill is made out of MDMA
  - MDMA is illegal
  - Therefore, it is **certain** that this pill is illegal

  - This latent print has similarities with a control print from Mr. X.
  - I have compared the LP to all other individuals present that night (and I excluded them)
  - Therefore, it is **certain** that Mr. X left that LP

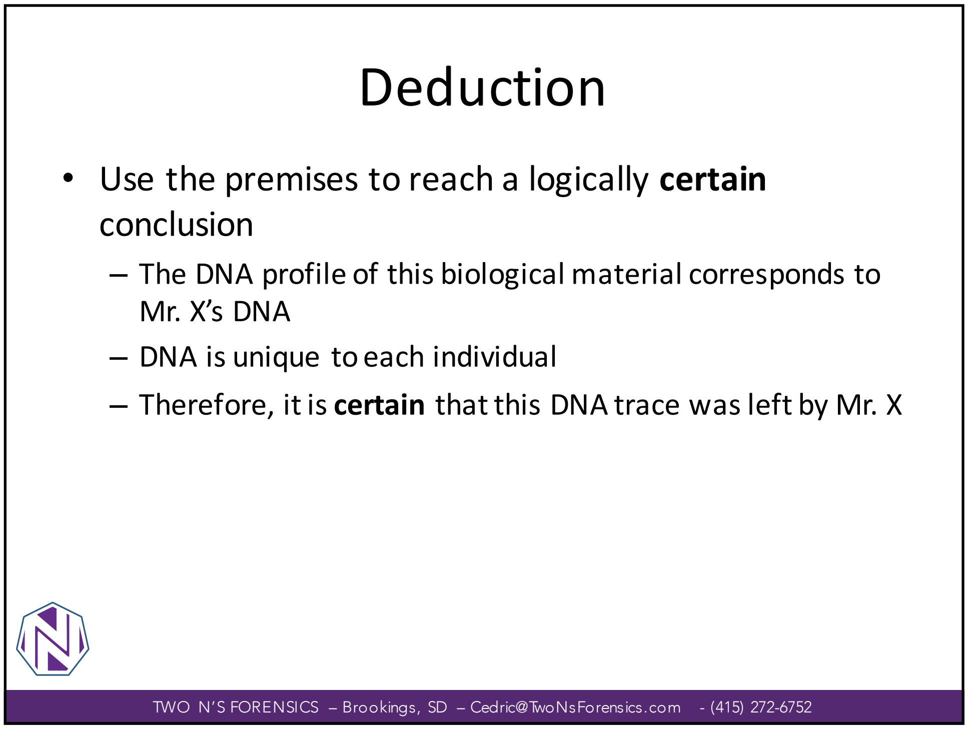


Note that this is true

- Use the premises to re **logically certain** conclusion
  - This pill is made out of MDMA
  - MDMA is illegal
  - Therefore, it is **certain** that this pill is illegal
  - This latent print has similariti Mr. X.
  - I have compared the LP to all other individuals present that night (and I excluded them)
  - Therefore, it is **certain** that Mr. X left that LP

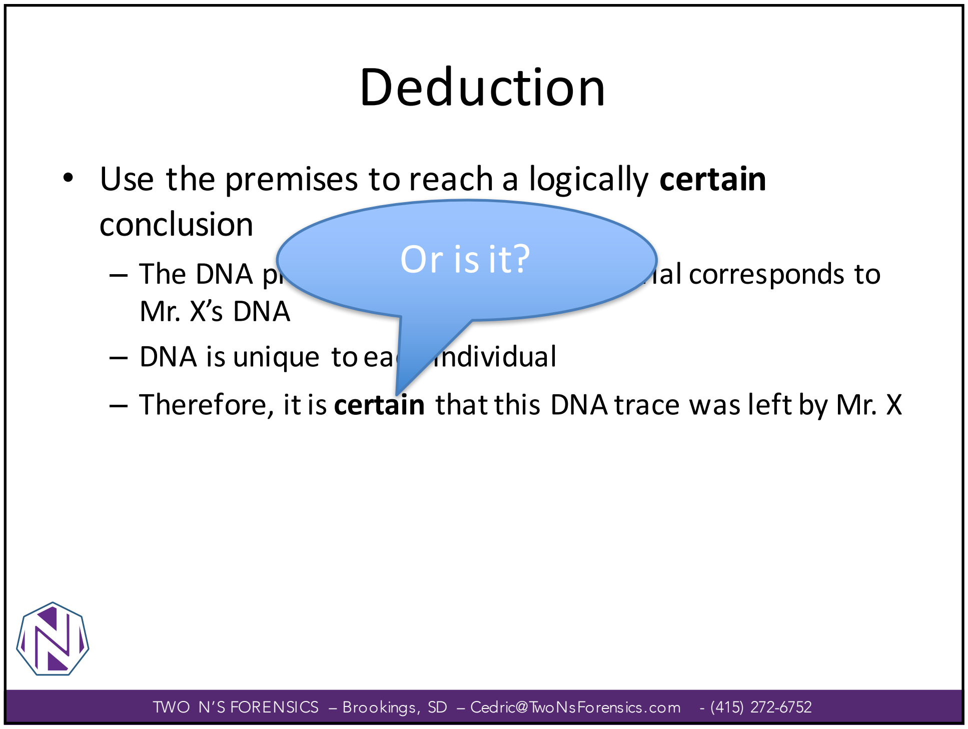
This is also true

This can only be true



Deduction

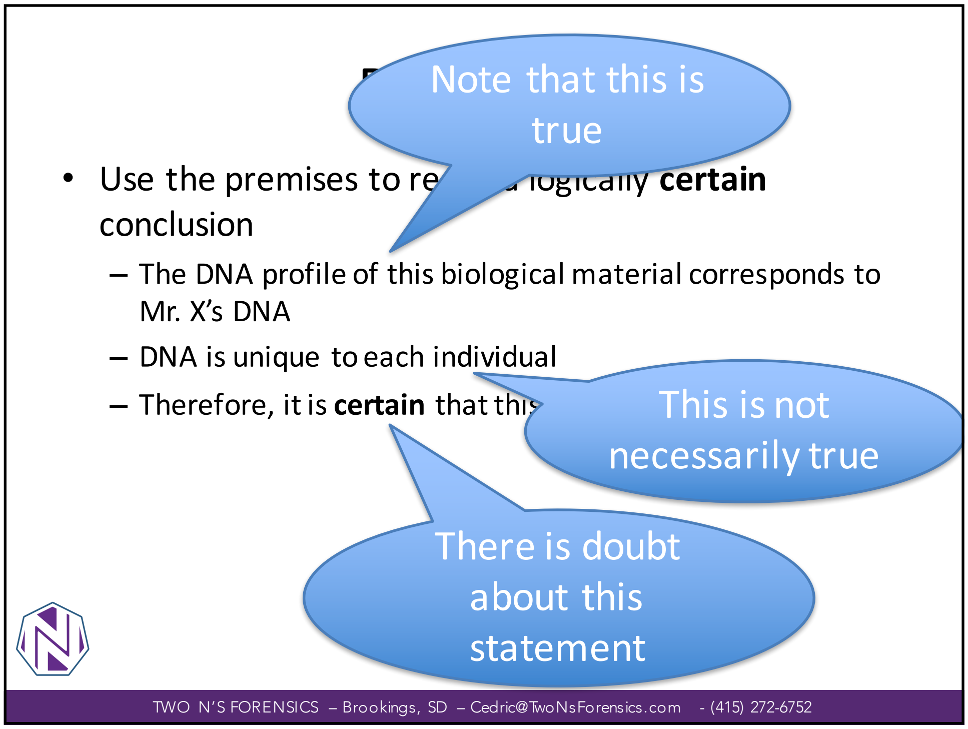
- Use the premises to reach a logically **certain** conclusion
  - The DNA profile of this biological material corresponds to Mr. X's DNA
  - DNA is unique to each individual
  - Therefore, it is **certain** that this DNA trace was left by Mr. X



Deduction

- Use the premises to reach a logically **certain** conclusion
  - The DNA pr ial corresponds to Mr. X's DNA
  - DNA is unique to ea individual
  - Therefore, it is **certain** that this DNA trace was left by Mr. X

Or is it?

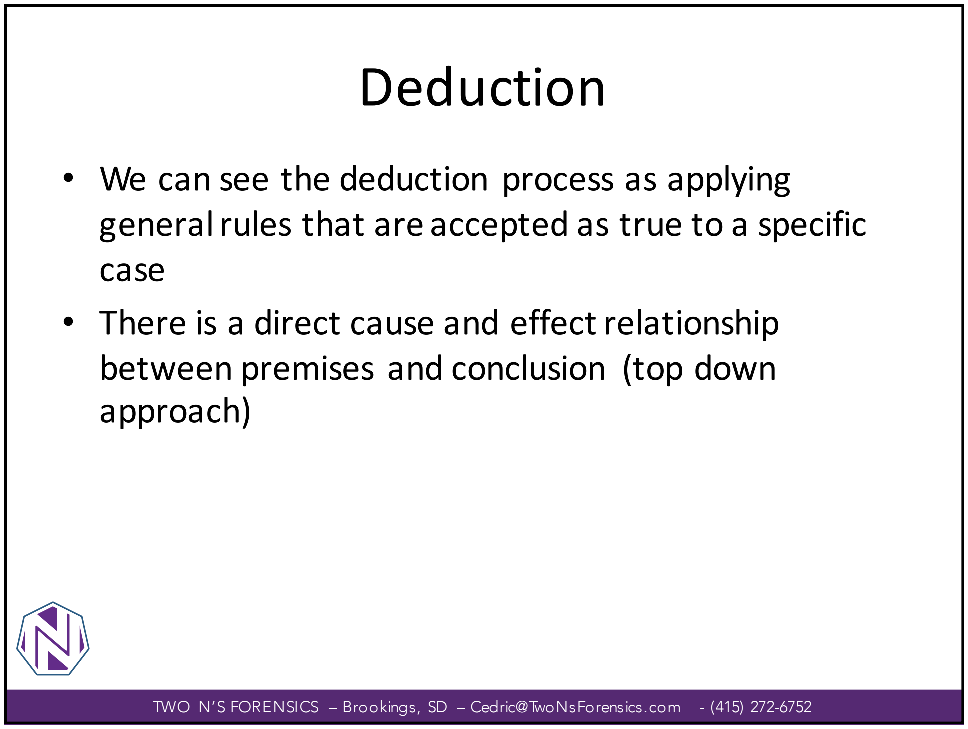


Note that this is true

- Use the premises to re logically **certain** conclusion
  - The DNA profile of this biological material corresponds to Mr. X's DNA
  - DNA is unique to each individual
  - Therefore, it is **certain** that this

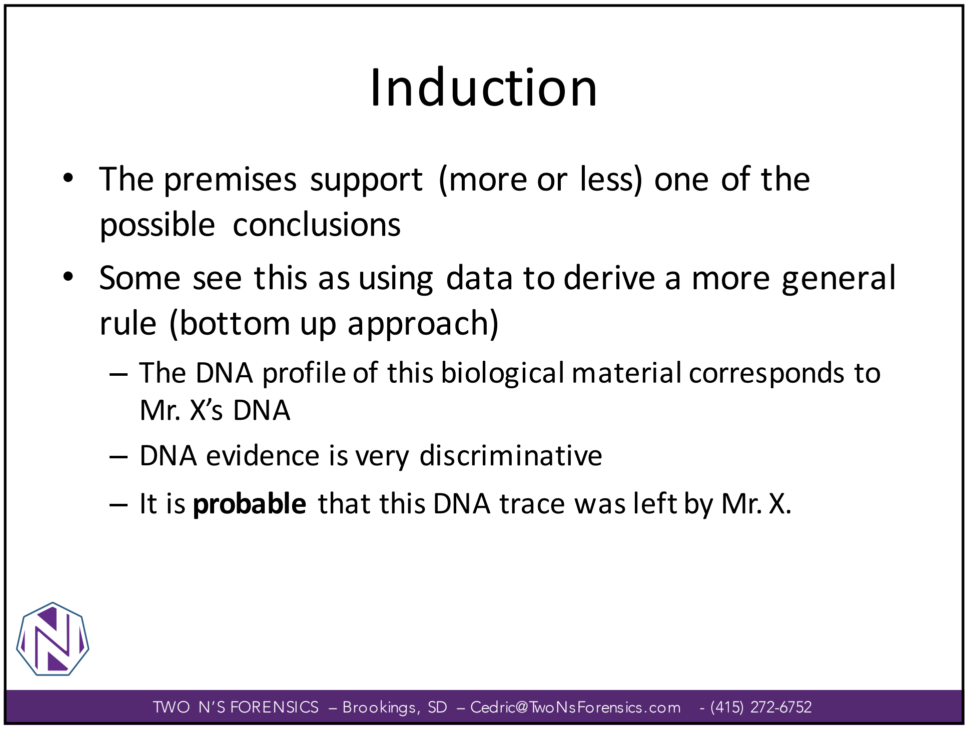
This is not necessarily true

There is doubt about this statement



Deduction

- We can see the deduction process as applying general rules that are accepted as true to a specific case
- There is a direct cause and effect relationship between premises and conclusion (top down approach)



Induction

- The premises support (more or less) one of the possible conclusions
- Some see this as using data to derive a more general rule (bottom up approach)
  - The DNA profile of this biological material corresponds to Mr. X's DNA
  - DNA evidence is very discriminative
  - It is **probable** that this DNA trace was left by Mr. X.

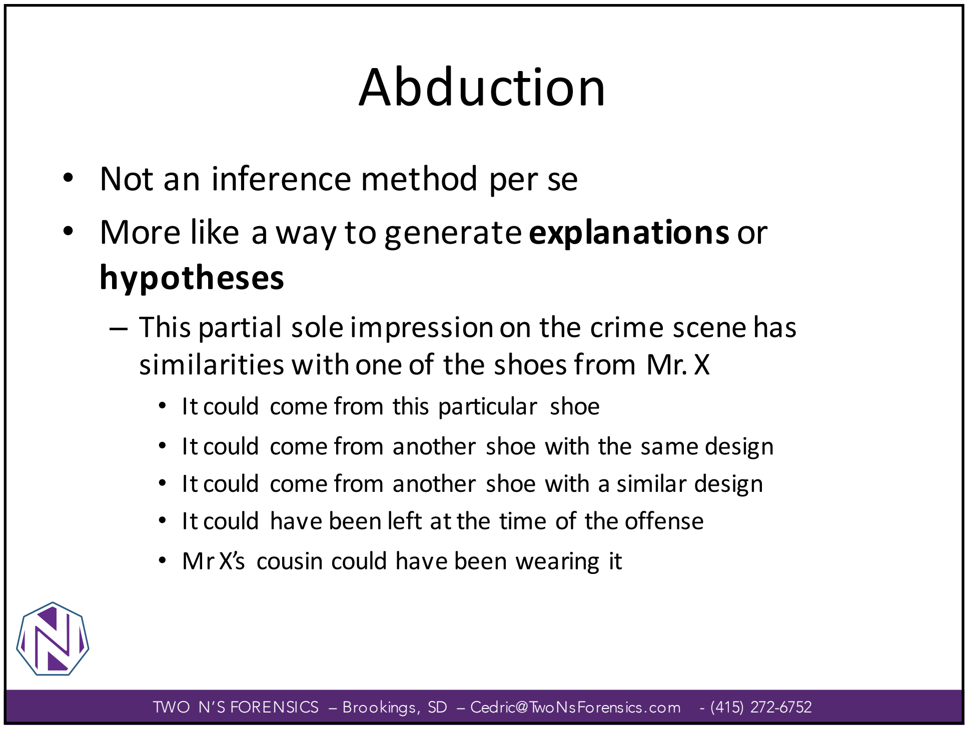
Induction

- The premises su
  possible concl
- Some see this
  rule (bottom u
  - The GC/MS spect e
    found in cocaine s
  - Cocaine spectrum is fairly specific
  - It is **probable** that this substance contains cocaine

We are not totally excluding that some other substances could have a similar mass breakdown at that retention time

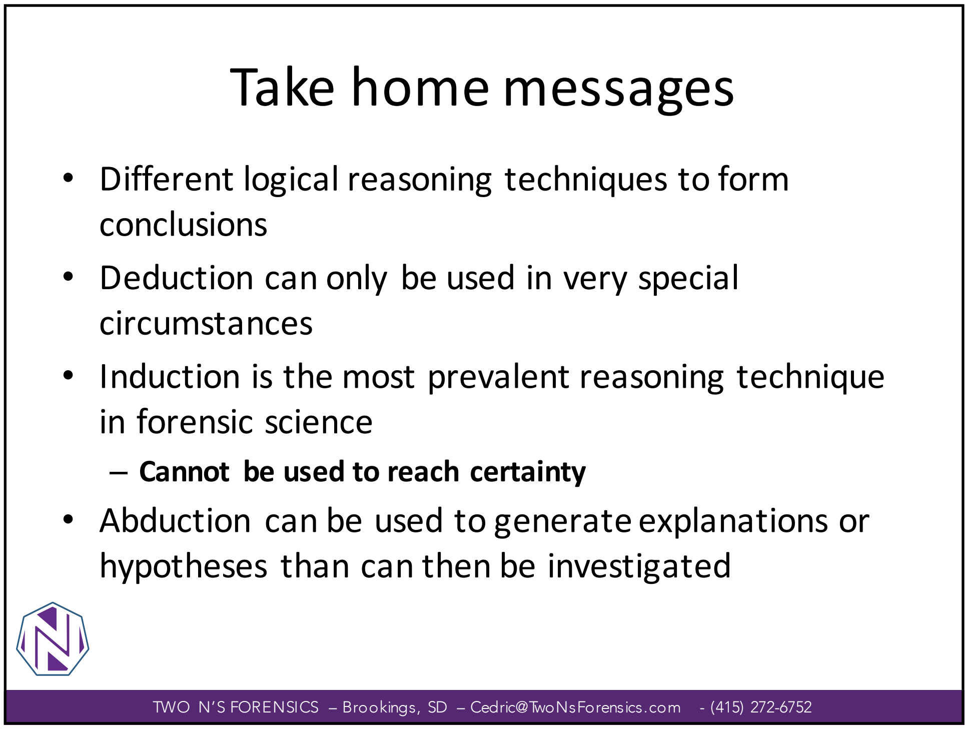
Abduction

- Not an inference method per se
- More like a way to generate **explanations** or **hypotheses**
  - Explanations don't have to follow any premise and they don't even have to be reasonable
  - They will then be tested formally (by induction or deduction)



Abduction

- Not an inference method per se
- More like a way to generate **explanations** or **hypotheses**
  - This partial sole impression on the crime scene has similarities with one of the shoes from Mr. X
    - It could come from this particular shoe
    - It could come from another shoe with the same design
    - It could come from another shoe with a similar design
    - It could have been left at the time of the offense
    - Mr X's cousin could have been wearing it



Take home messages

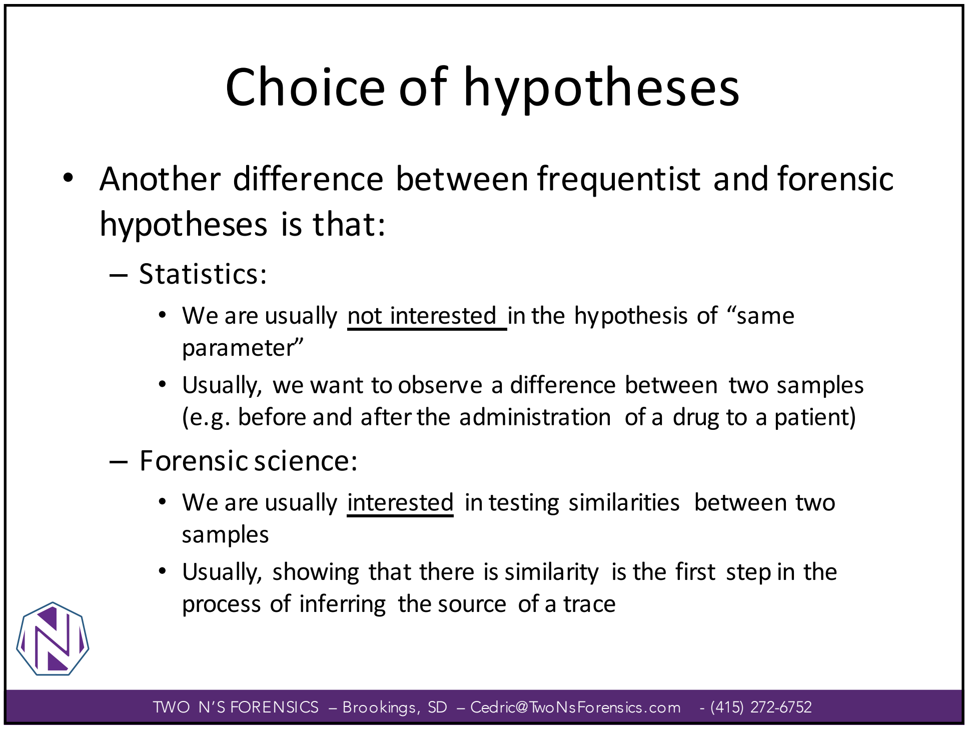
- Different logical reasoning techniques to form conclusions
- Deduction can only be used in very special circumstances
- Induction is the most prevalent reasoning technique in forensic science
  - **Cannot be used to reach certainty**
- Abduction can be used to generate explanations or hypotheses than can then be investigated

Chapter VII

HYPOTHESIS TESTING

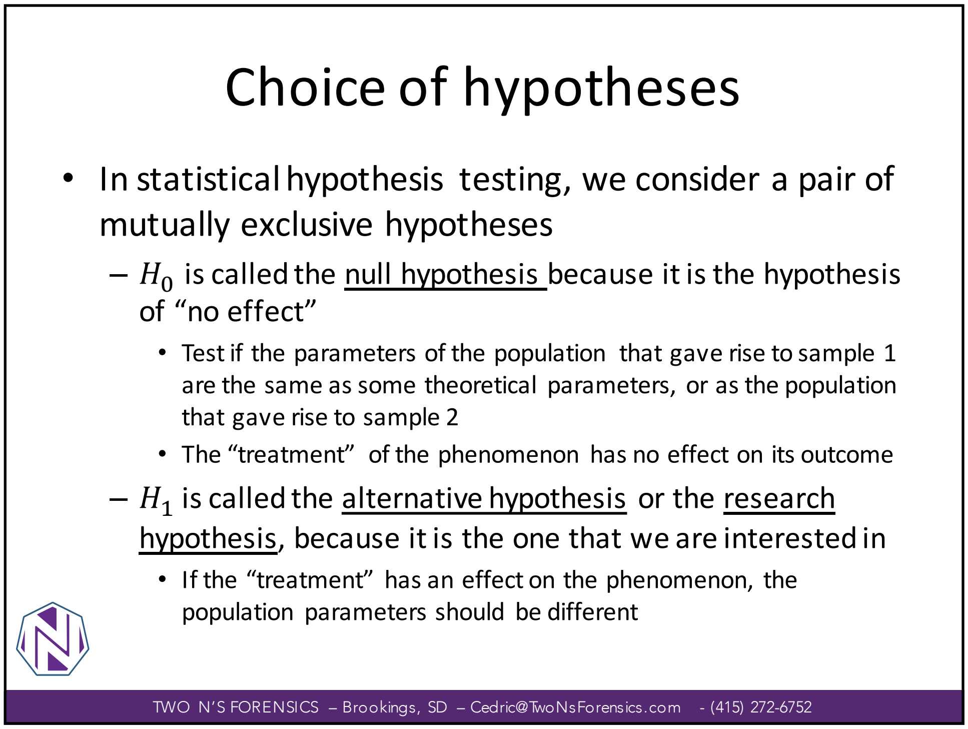
Hypothesis testing

- This chapter relates to classic frequentist hypothesis testing, not to the determination of the source of a forensic evidence. Not the same hypotheses!!!
- Frequentist hypothesis:
  - One or more sets of observations arise from **populations with the same parameters** (might be two different but identical sources)
- Forensic hypothesis:
  - Two sets of observations arise from the **same population** (a single source)



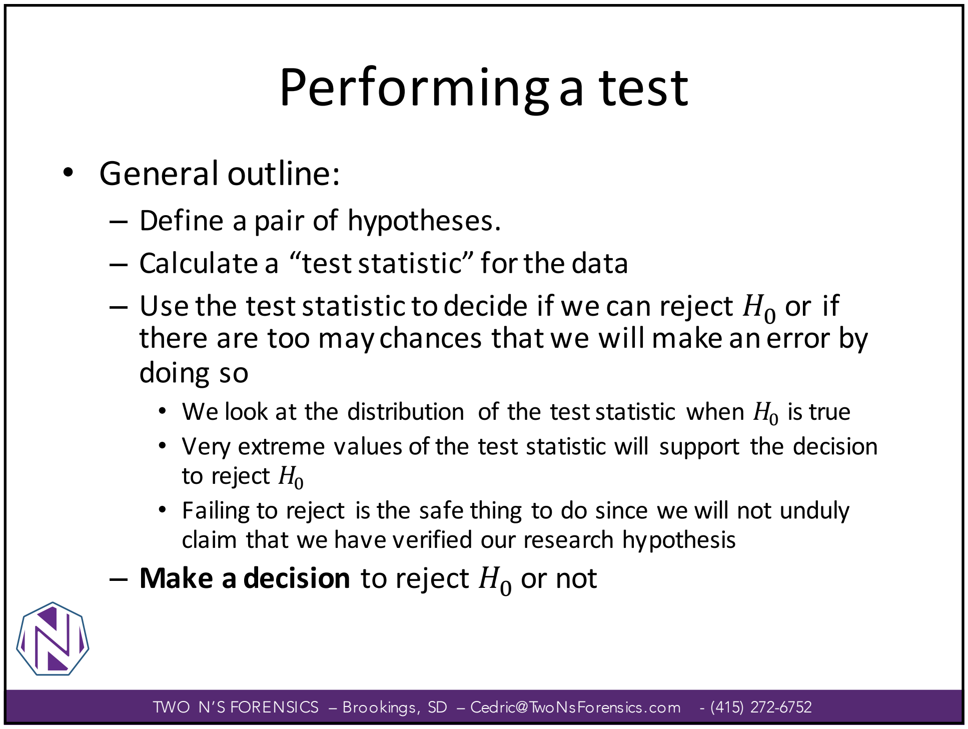
Choice of hypotheses

- Another difference between frequentist and forensic hypotheses is that:
  - Statistics:
    - We are usually not interested in the hypothesis of "same parameter"
    - Usually, we want to observe a difference between two samples (e.g. before and after the administration of a drug to a patient)
  - Forensic science:
    - We are usually interested in testing similarities between two samples
    - Usually, showing that there is similarity is the first step in the process of inferring the source of a trace



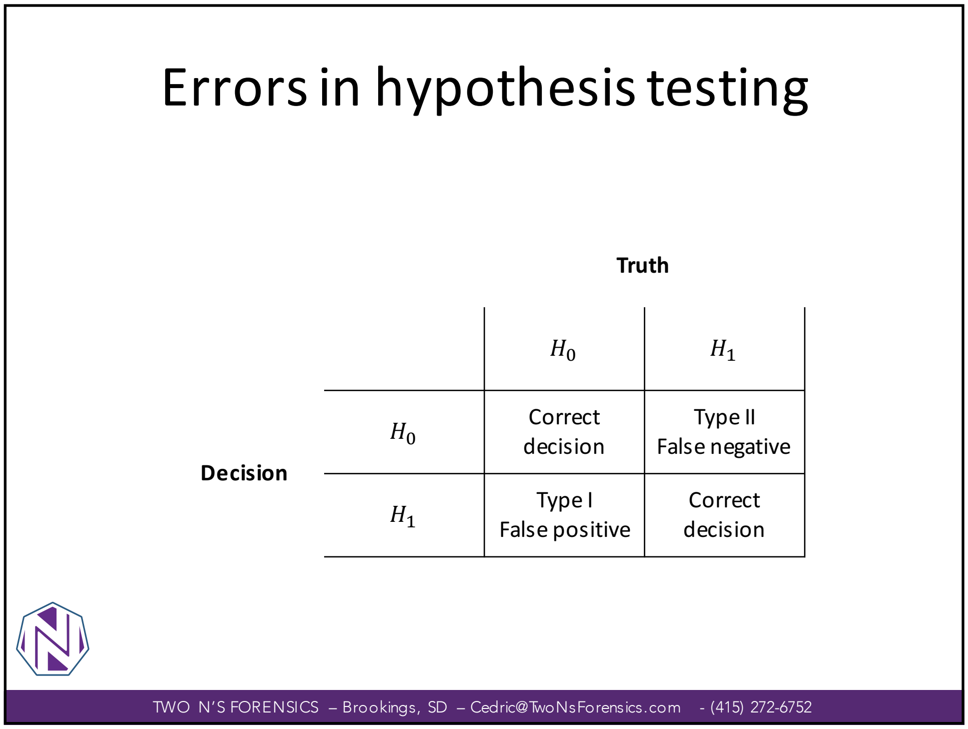
Choice of hypotheses

- In statistical hypothesis testing, we consider a pair of mutually exclusive hypotheses
  - H_0 is called the null hypothesis because it is the hypothesis of "no effect"
    - Test if the parameters of the population that gave rise to sample 1 are the same as some theoretical parameters, or as the population that gave rise to sample 2
    - The "treatment" of the phenomenon has no effect on its outcome
  - H_1 is called the alternative hypothesis or the research hypothesis, because it is the one that we are interested in
    - If the "treatment" has an effect on the phenomenon, the population parameters should be different



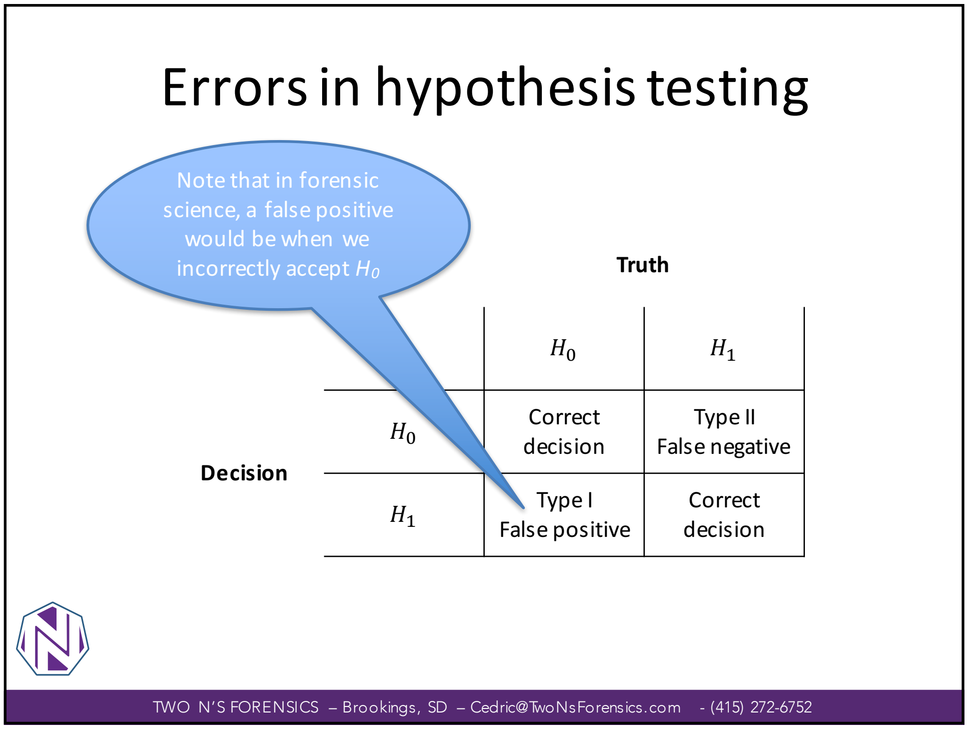
Performing a test

- General outline:
  - Define a pair of hypotheses.
  - Calculate a “test statistic” for the data
  - Use the test statistic to decide if we can reject H_0 or if there are too may chances that we will make an error by doing so
    - We look at the distribution of the test statistic when H_0 is true
    - Very extreme values of the test statistic will support the decision to reject H_0
    - Failing to reject is the safe thing to do since we will not unduly claim that we have verified our research hypothesis
  - **Make a decision** to reject H_0 or not



Errors in hypothesis testing

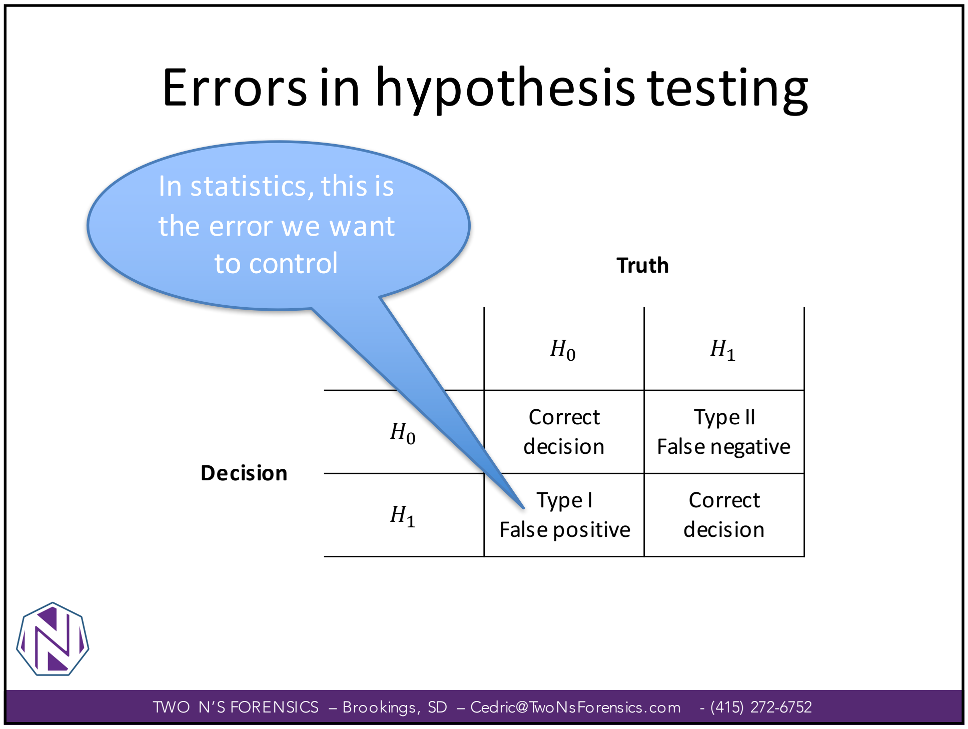
		Truth	
		H_0	H_1
Decision	H_0	Correct decision	Type II False negative
	H_1	Type I False positive	Correct decision



Errors in hypothesis testing

Note that in forensic science, a false positive would be when we incorrectly accept H_0

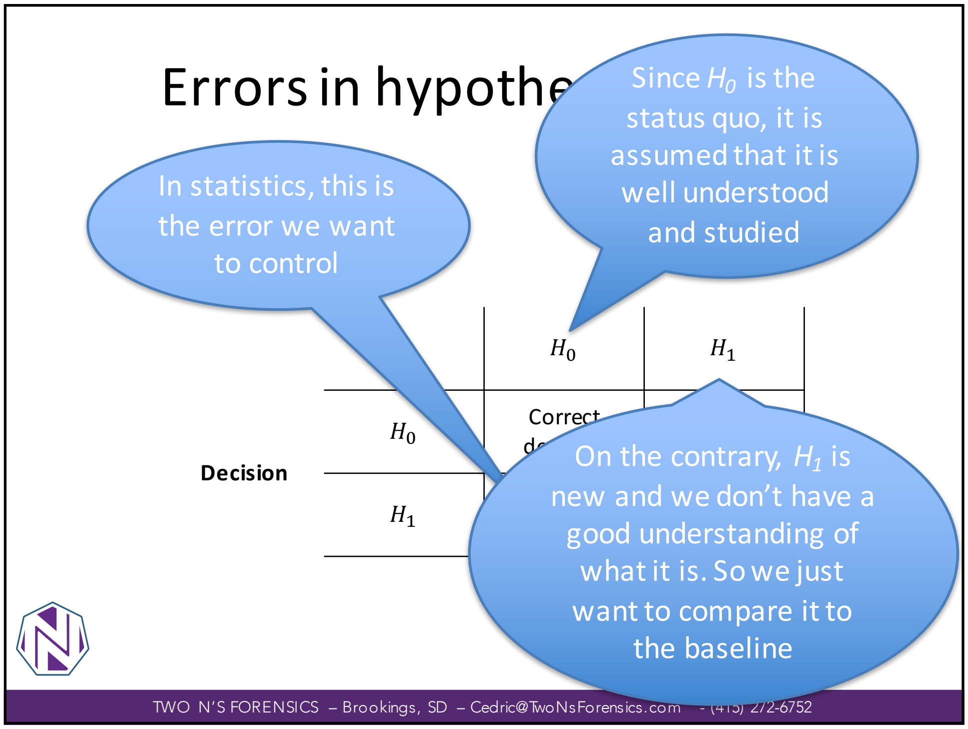
		Truth	
		H_0	H_1
Decision	H_0	Correct decision	Type II False negative
	H_1	Type I False positive	Correct decision



Errors in hypothesis testing

In statistics, this is the error we want to control

		Truth	
		H_0	H_1
Decision	H_0	Correct decision	Type II False negative
	H_1	Type I False positive	Correct decision



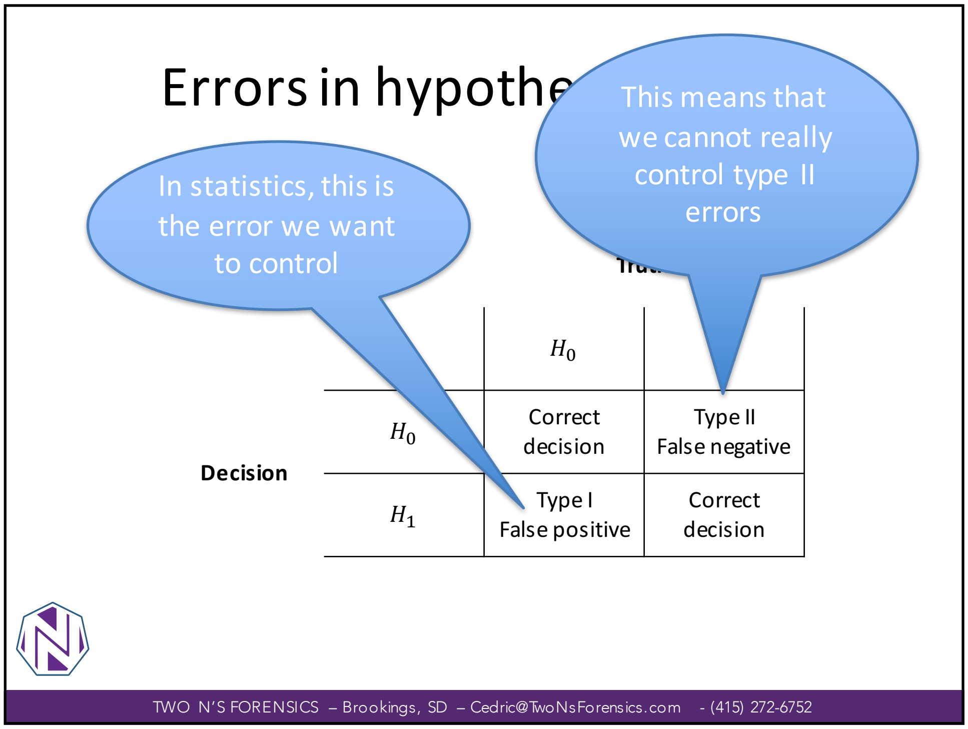
Errors in hypothe

In statistics, this is the error we want to control

Since H_0 is the status quo, it is assumed that it is well understood and studied

		H_0	H_1
Decision	H_0	Correct de	
	H_1		

On the contrary, H_1 is new and we don't have a good understanding of what it is. So we just want to compare it to the baseline



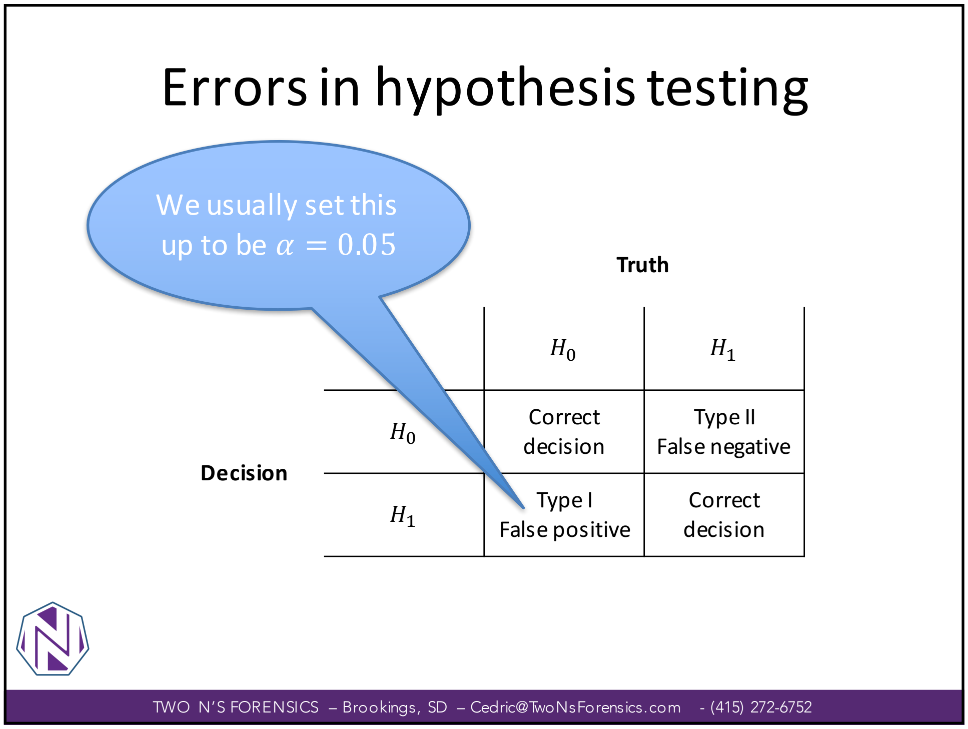
Errors in hypothe

In statistics, this is the error we want to control

This means that we cannot really control type II errors

Trut

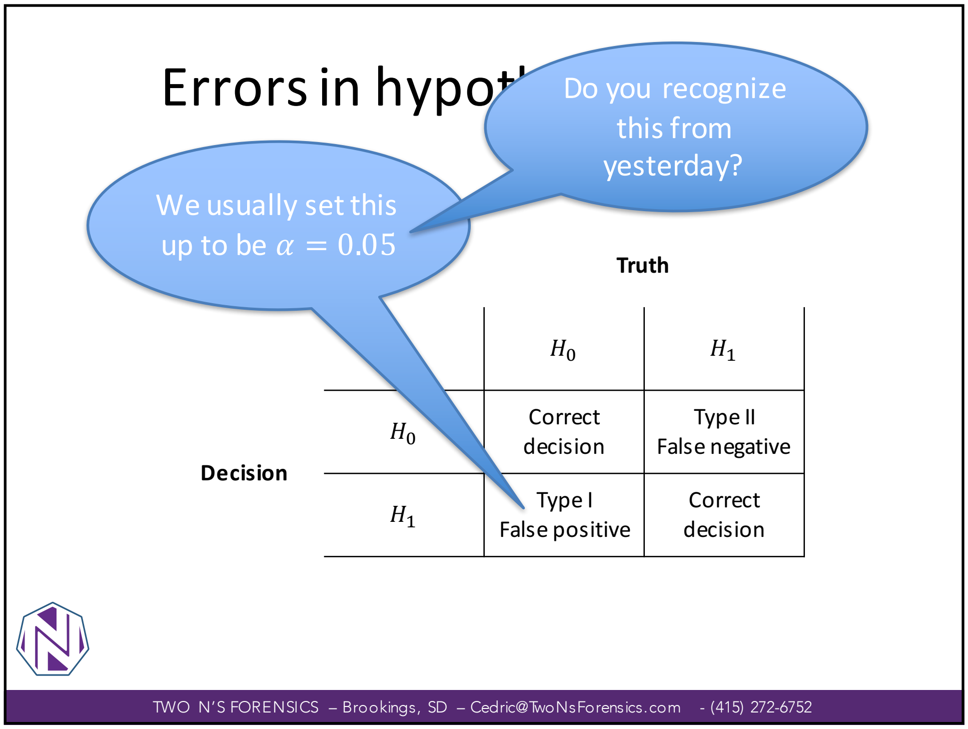
		H_0	
Decision	H_0	Correct decision	Type II False negative
	H_1	Type I False positive	Correct decision



Errors in hypothesis testing

We usually set this up to be $\alpha = 0.05$

		Truth	
		H_0	H_1
Decision	H_0	Correct decision	Type II False negative
	H_1	Type I False positive	Correct decision



Errors in hypot

Do you recognize this from yesterday?

We usually set this up to be $\alpha = 0.05$

		Truth	
		H_0	H_1
Decision	H_0	Correct decision	Type II False negative
	H_1	Type I False positive	Correct decision

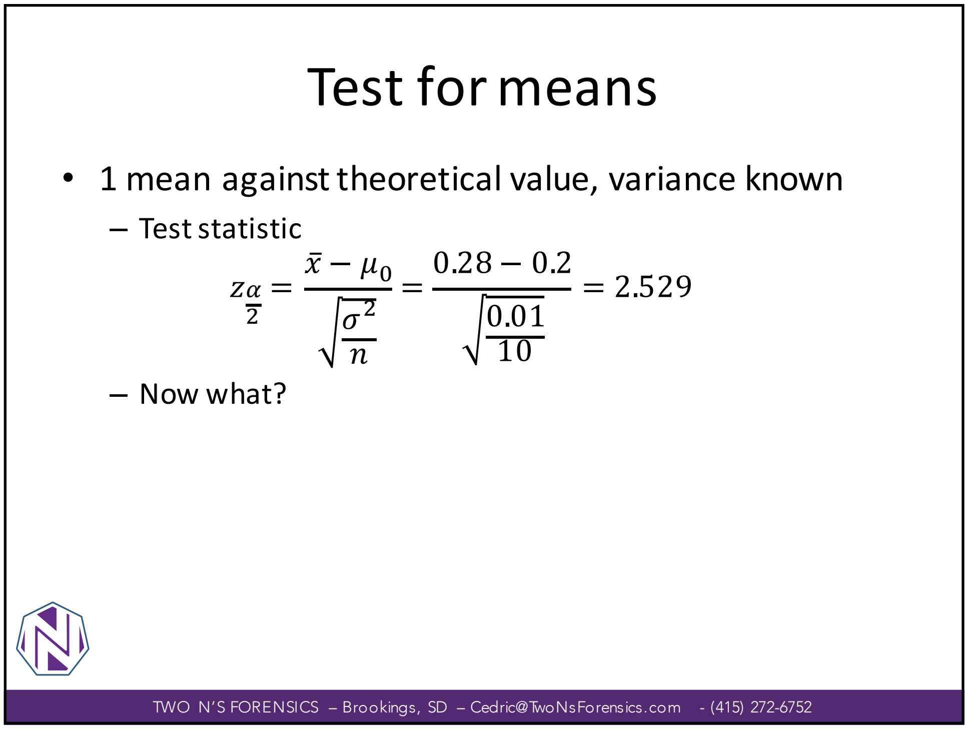
Test statistic(s)

- Different test statistics depending on what we are trying to test, and depending on the type of variables
  - Continuous
    - Comparing means: Z-test; T-test
    - Comparing proportions: Z-test
  - Categorical
    - Comparing proportions across multiple categories: chi-square test
    - Testing independence between different categorical variables

Test for means

- 1 mean against theoretical value, variance known
  - Example: we want to test if the average tolerance to cocaine of tolerant individuals is different than 0.2mg/L found in recreational users. We take a sample of tolerant individuals (say 10) and we measure their cocaine concentration ($\bar{x} = 0.28$). We also ”know” based on past experience that the variance of the concentration in recreational users is 0.01
  - Hypotheses:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

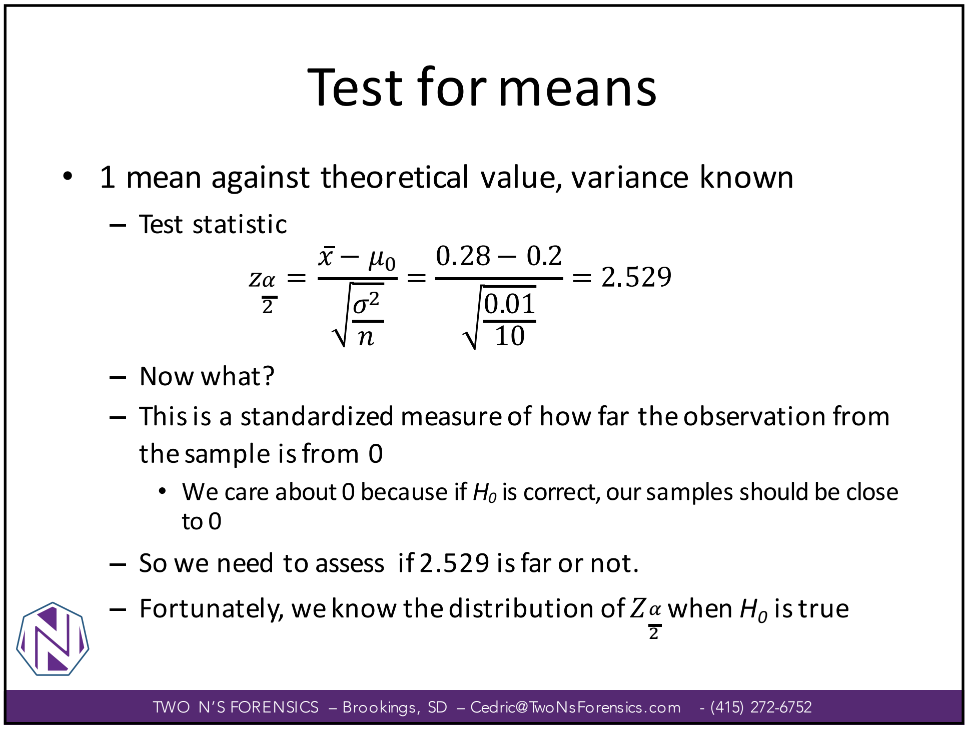


Test for means

- 1 mean against theoretical value, variance known
  - Test statistic

$$z_{\frac{\alpha}{2}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} = \frac{0.28 - 0.2}{\sqrt{\frac{0.01}{10}}} = 2.529$$

  - Now what?

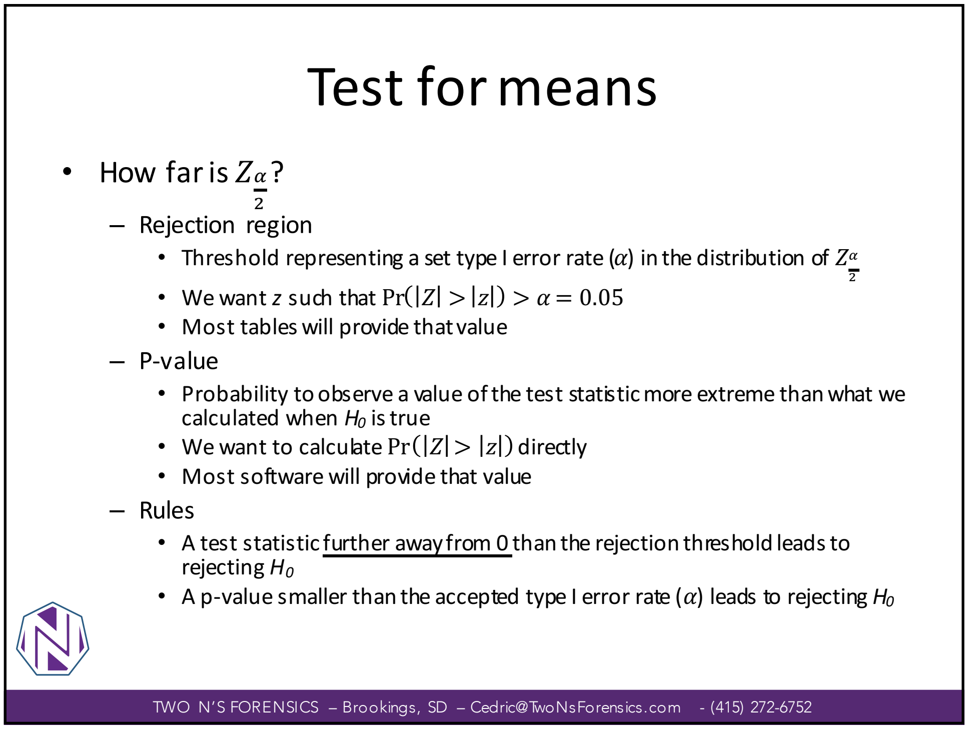


Test for means

- 1 mean against theoretical value, variance known
  - Test statistic

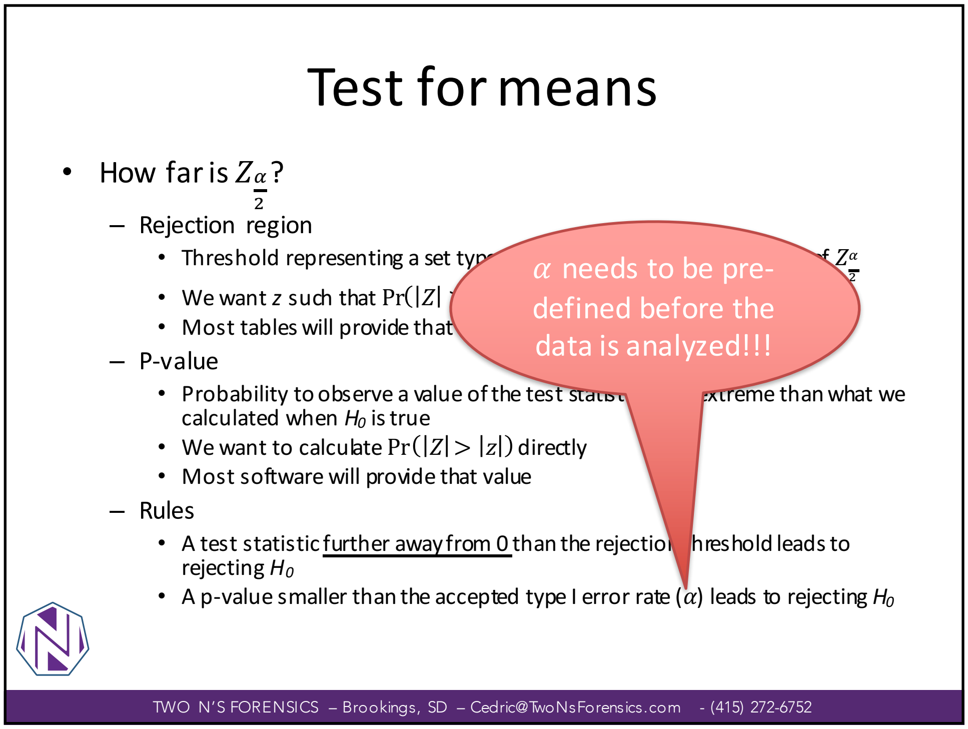
$$z_{\frac{\alpha}{2}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} = \frac{0.28 - 0.2}{\sqrt{\frac{0.01}{10}}} = 2.529$$

  - Now what?
  - This is a standardized measure of how far the observation from the sample is from 0
    - We care about 0 because if H_0 is correct, our samples should be close to 0
  - So we need to assess if 2.529 is far or not.
  - Fortunately, we know the distribution of $Z_{\frac{\alpha}{2}}$ when H_0 is true



Test for means

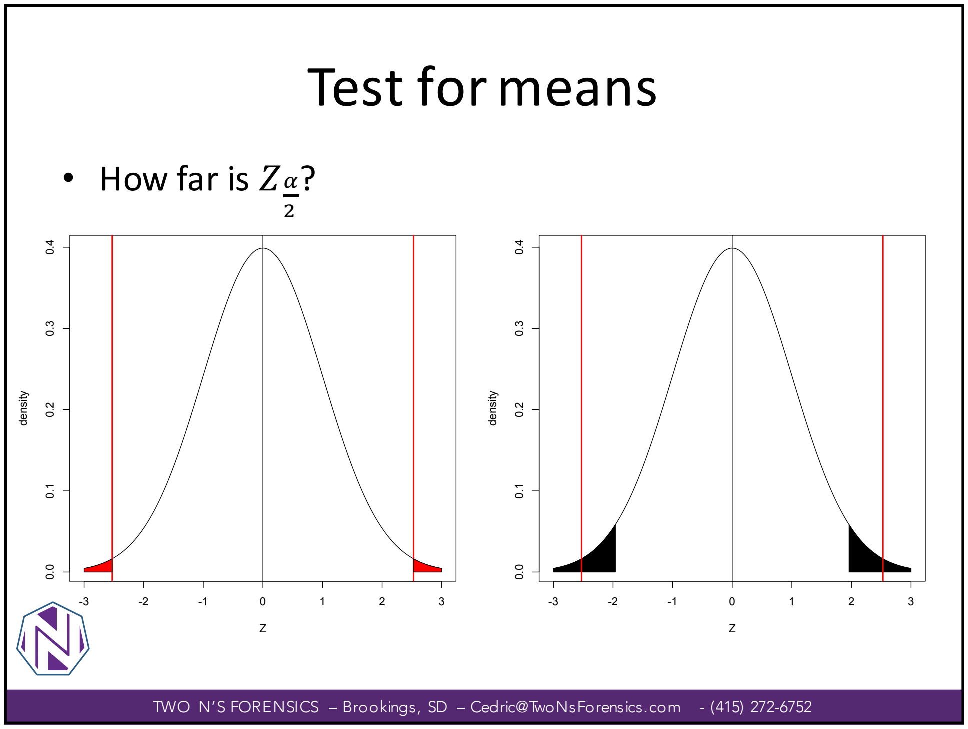
- How far is $Z_{\frac{\alpha}{2}}$?
  - Rejection region
    - Threshold representing a set type I error rate (α) in the distribution of $Z_{\frac{\alpha}{2}}$
    - We want z such that $\Pr(|Z| > |z|) > \alpha = 0.05$
    - Most tables will provide that value
  - P-value
    - Probability to observe a value of the test statistic more extreme than what we calculated when H_0 is true
    - We want to calculate $\Pr(|Z| > |z|)$ directly
    - Most software will provide that value
  - Rules
    - A test statistic further away from 0 than the rejection threshold leads to rejecting H_0
    - A p-value smaller than the accepted type I error rate (α) leads to rejecting H_0



Test for means

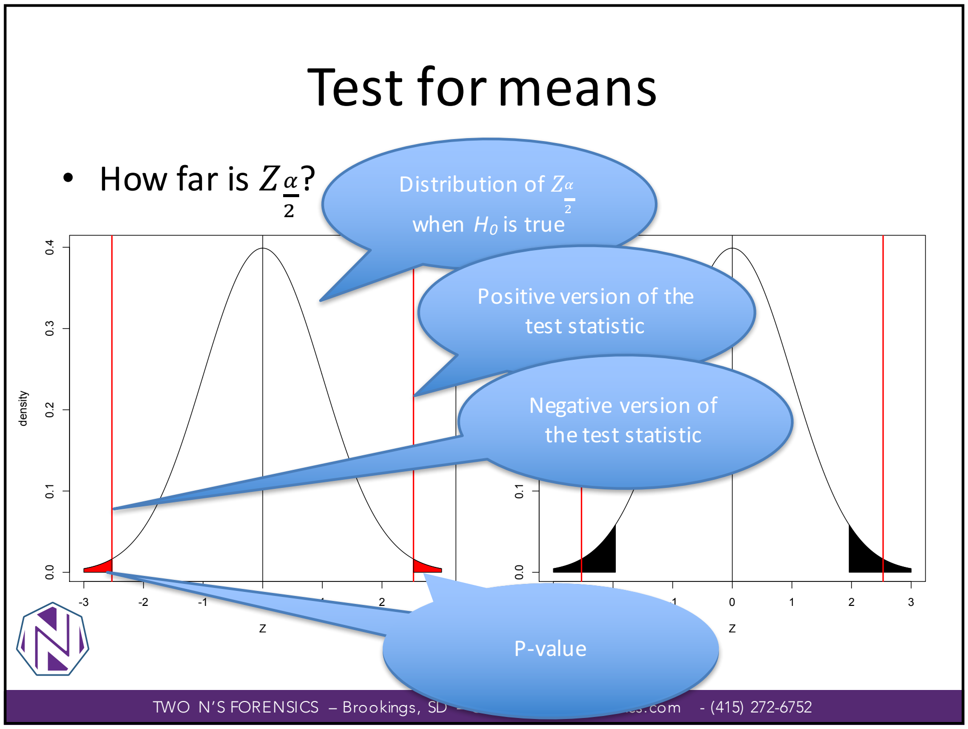
- How far is $Z_{\frac{\alpha}{2}}$?
  - Rejection region
    - Threshold representing a set typ f $Z_{\frac{\alpha}{2}}$
    - We want z such that $\Pr(|Z|$
    - Most tables will provide that
  - P-value
    - Probability to observe a value of the test statis xtreme than what we calculated when H_0 is true
    - We want to calculate $\Pr(|Z| > |z|)$ directly
    - Most software will provide that value
  - Rules
    - A test statistic further away from 0 than the rejectio hreshold leads to rejecting H_0
    - A p-value smaller than the accepted type I error rate (α) leads to rejecting H_0

α needs to be pre-defined before the data is analyzed!!!



Test for means

- How far is $Z_{\frac{\alpha}{2}}$?



Test for means

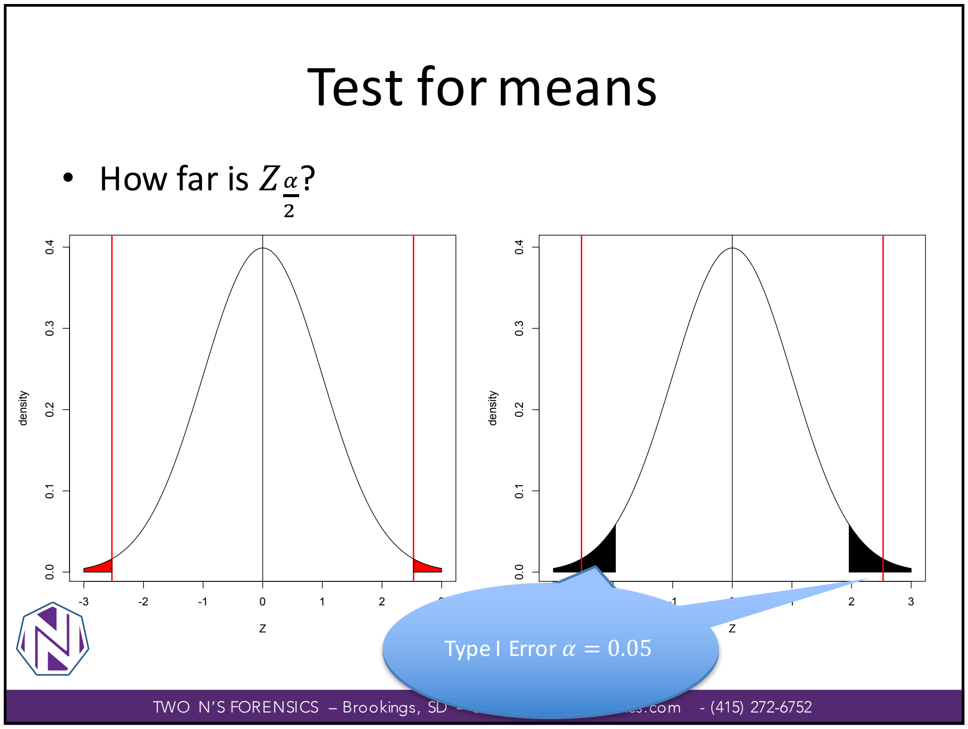
- How far is $Z_{\frac{\alpha}{2}}$?

Distribution of $Z_{\frac{\alpha}{2}}$ when H_0 is true

Positive version of the test statistic

Negative version of the test statistic

P-value

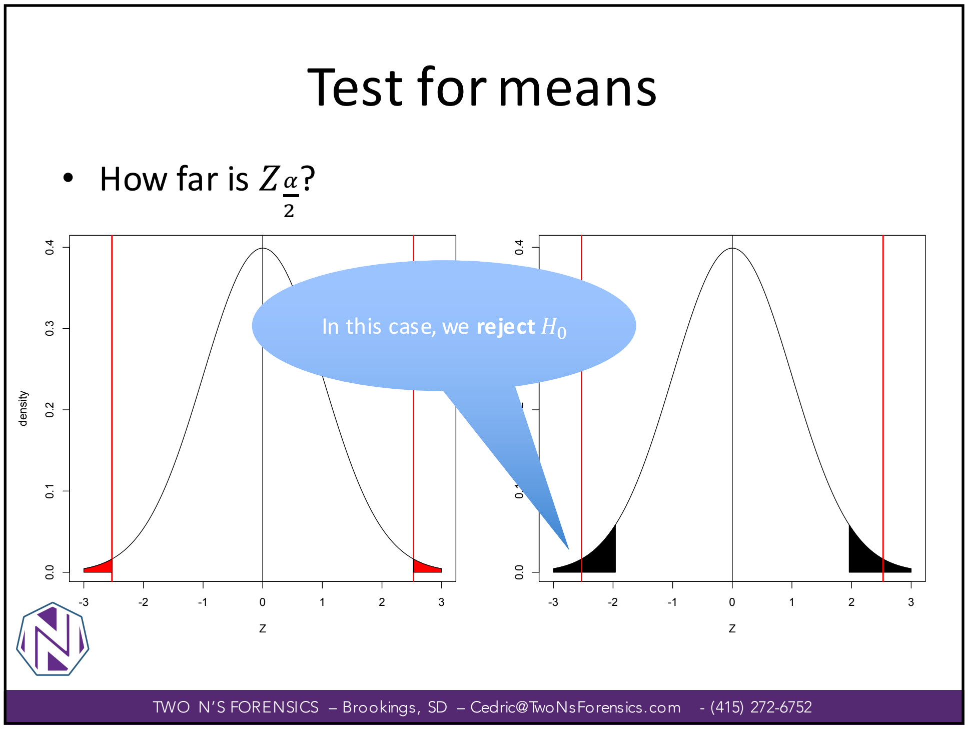


Test for means

- How far is $Z_{\frac{\alpha}{2}}$?

Type I Error $\alpha = 0.05$

TWO N'S FORENSICS – Brookings, SD – .com - (415) 272-6752

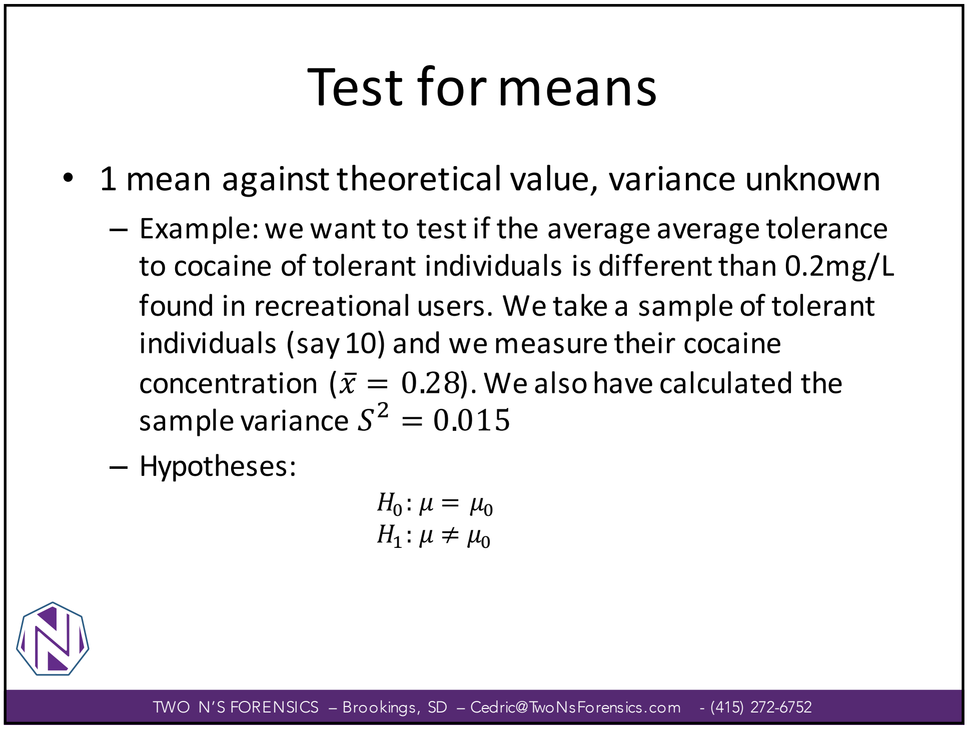


Test for means

- How far is $Z_{\frac{\alpha}{2}}$?

In this case, we **reject** H_0

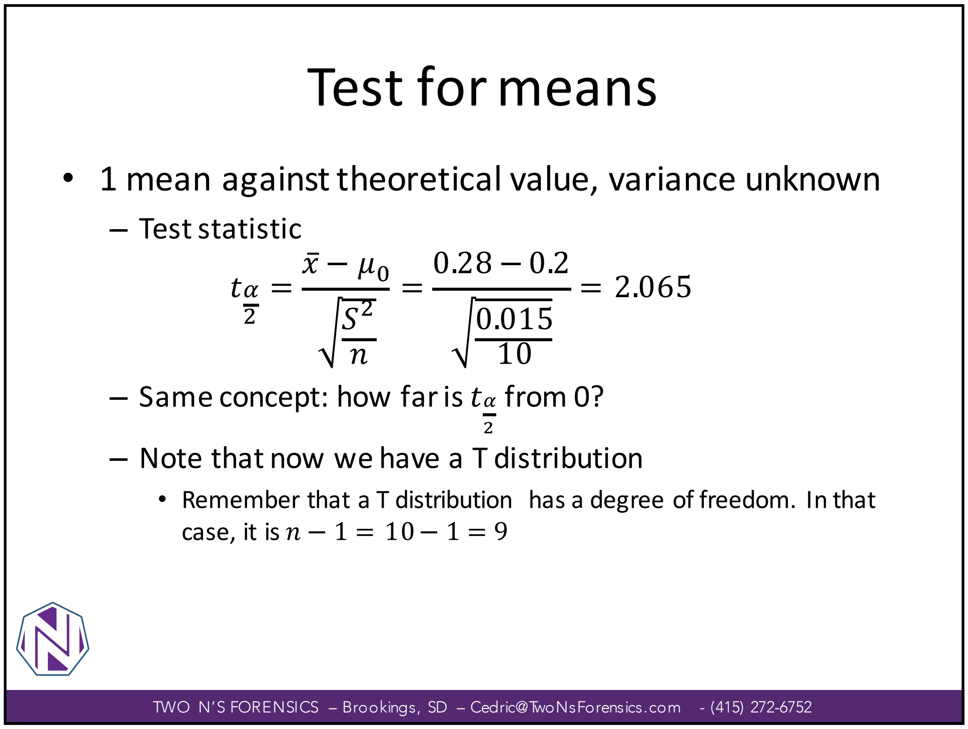
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Test for means

- 1 mean against theoretical value, variance unknown
  - Example: we want to test if the average average tolerance to cocaine of tolerant individuals is different than 0.2mg/L found in recreational users. We take a sample of tolerant individuals (say 10) and we measure their cocaine concentration ($\bar{x} = 0.28$). We also have calculated the sample variance $S^2 = 0.015$
  - Hypotheses:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

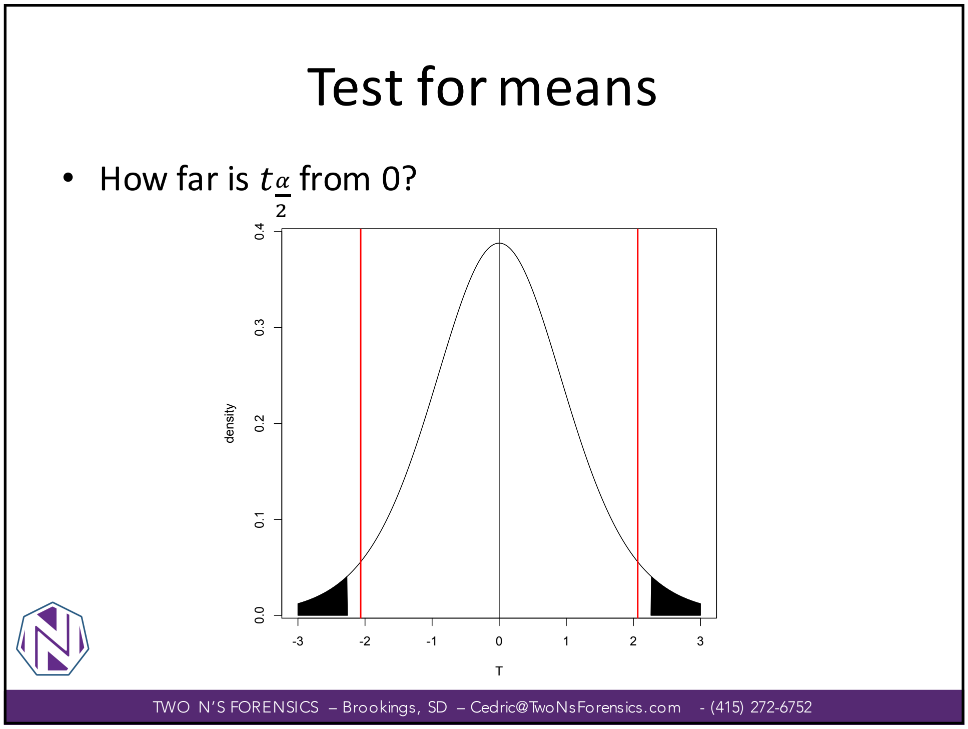


Test for means

- 1 mean against theoretical value, variance unknown
  - Test statistic

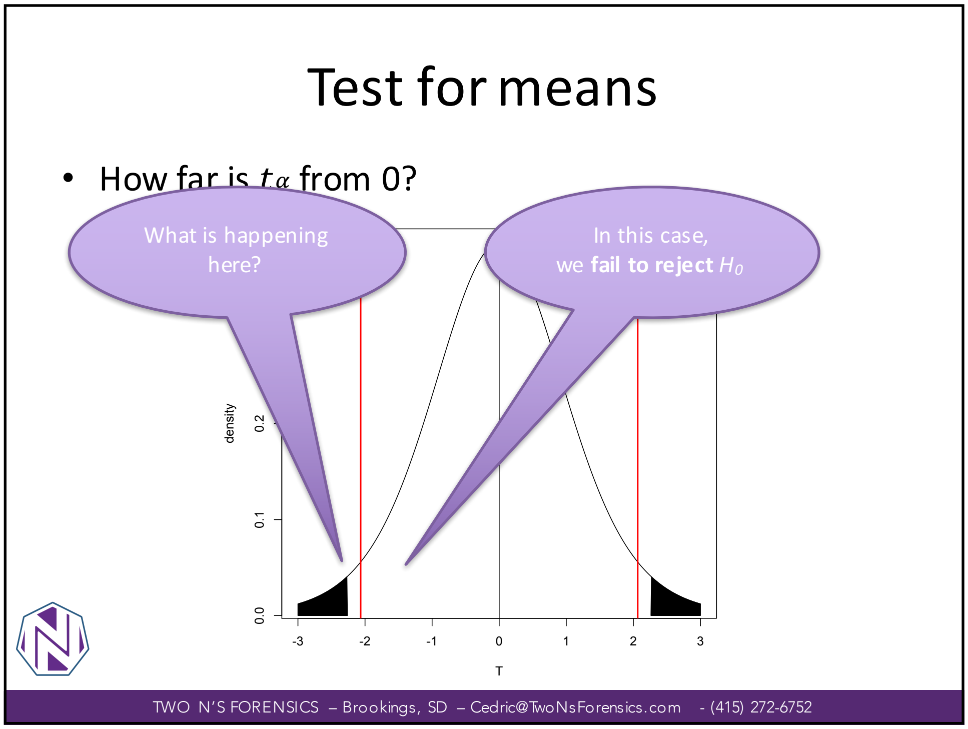
$$t_{\frac{\alpha}{2}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{S^2}{n}}} = \frac{0.28 - 0.2}{\sqrt{\frac{0.015}{10}}} = 2.065$$

  - Same concept: how far is $t_{\frac{\alpha}{2}}$ from 0?
  - Note that now we have a T distribution
    - Remember that a T distribution has a degree of freedom. In that case, it is $n - 1 = 10 - 1 = 9$



Test for means

- How far is $t_{\frac{\alpha}{2}}$ from 0?



Test for means

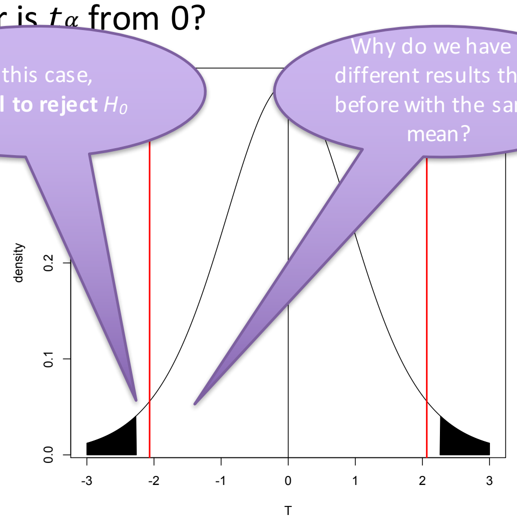
- How far is $t_{\frac{\alpha}{2}}$ from 0?

What is happening here?

In this case, we **fail to reject** H_0

Test for means

- How far is t_α from 0?



In this case, we **fail to reject** H_0

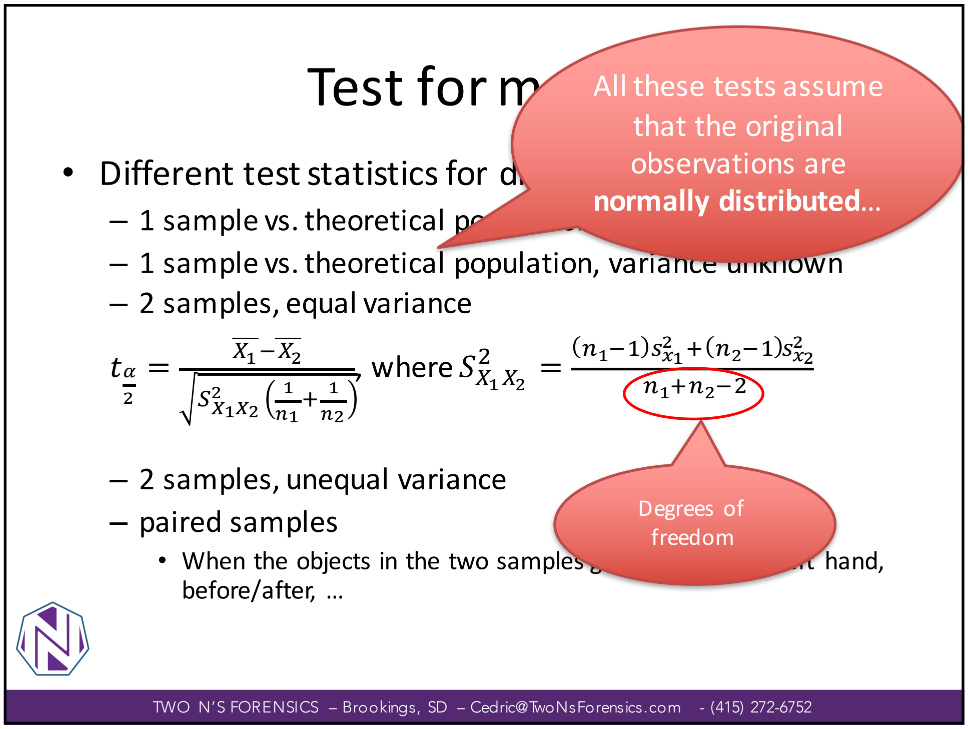
Why do we have a different results than before with the same mean?

Test for means

- Different test statistics for different situations:
  - 1 sample vs. theoretical population, variance known
  - 1 sample vs. theoretical population, variance unknown
  - 2 samples, equal variance

$$t_{\frac{\alpha}{2}} = \frac{\overline{X_1}-\overline{X_2}}{\sqrt{S^2_{X_1X_2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \text{ where } S^2_{X_1X_2} = \frac{(n_1-1)s^2_{x1}+(n_2-1)s^2_{x2}}{n_1+n_2-2}$$

  - 2 samples, unequal variance
  - paired samples
    - When the objects in the two samples go by pairs: right/left hand, before/after, …



Test for m

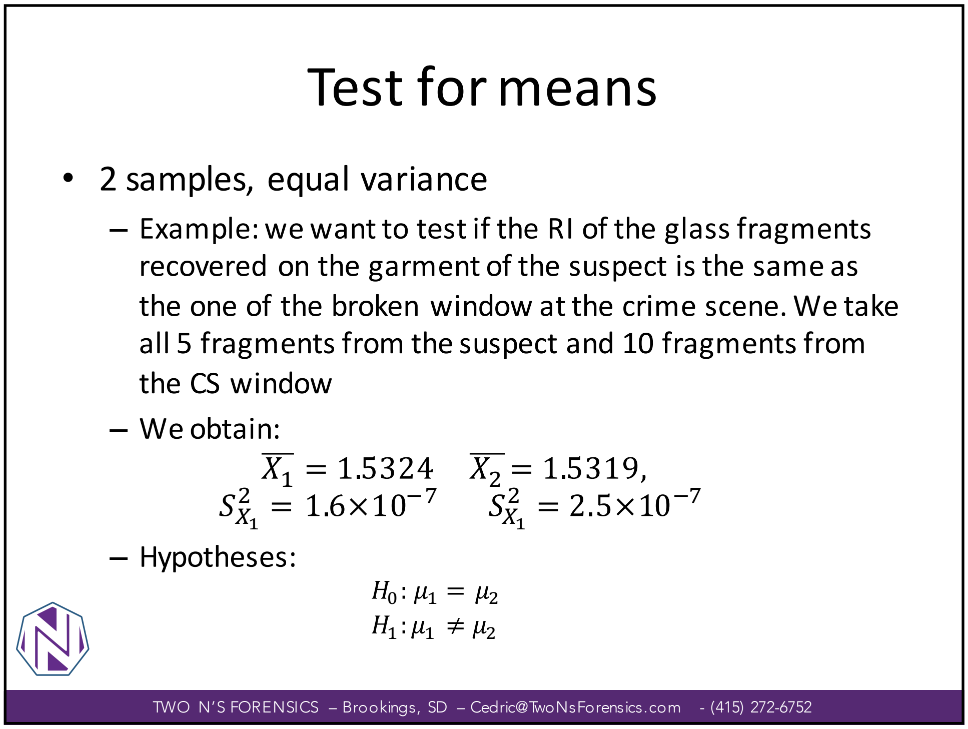
All these tests assume that the original observations are **normally distributed…**

- Different test statistics for d
  - 1 sample vs. theoretical po
  - 1 sample vs. theoretical population, variance unknown
  - 2 samples, equal variance

$$t_{\frac{\alpha}{2}} = \frac{\overline{X_1}-\overline{X_2}}{\sqrt{S^2_{X_1X2}\left(\frac{1}{n_1}+\frac{1}{n2}\right)}},\ \text{where } S^2_{X_1X_2} = \frac{(n_1-1)s^2_{x1}+(n_2-1)s^2_{x2}}{n_1+n_2-2}$$

Degrees of freedom

  - 2 samples, unequal variance
  - paired samples
    - When the objects in the two samples g t hand, before/after, …



Test for means

- 2 samples, equal variance
  - Example: we want to test if the RI of the glass fragments recovered on the garment of the suspect is the same as the one of the broken window at the crime scene. We take all 5 fragments from the suspect and 10 fragments from the CS window
  - We obtain:

$$\overline{X_1} = 1.5324 \quad \overline{X_2} = 1.5319,$$
$$S^2_{X_1} = 1.6\times10^{-7} \quad S^2_{X_1} = 2.5\times10^{-7}$$

  - Hypotheses:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Test for means

- 2 samples, equal variance

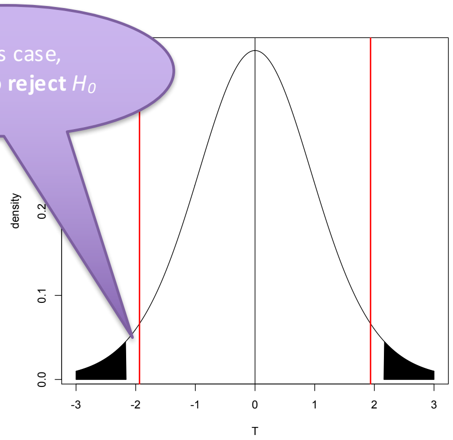
$$S^2_{X_1X_2} = \frac{(n_1-1)s^2_{x_1} + (n_2-1)s^2_{x_2}}{n_1+n_2-2} =$$

$$\frac{(5-1)1.6e-7+(10-1)2.5e-7}{10+5-2} = 2.22\times 10^{-7}$$

$$t_{\frac{\alpha}{2}} = \frac{\overline{X_1}-\overline{X_2}}{\sqrt{S^2_{X_1X_2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{1.5324-1.5319}{\sqrt{2.22\times 10^{-7}\left(\frac{1}{5}+\frac{1}{10}\right)}} = 1.936$$

Df = 10+5-2=13

Test for means



In this case, we **fail to reject** H_0

Test for proportions

- Same concept, different test statistics
  - Good thing is that they are all Z's
  - One proportion test: $Z_{\frac{\alpha}{2}} = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}$
  - Two proportions, pooled variance

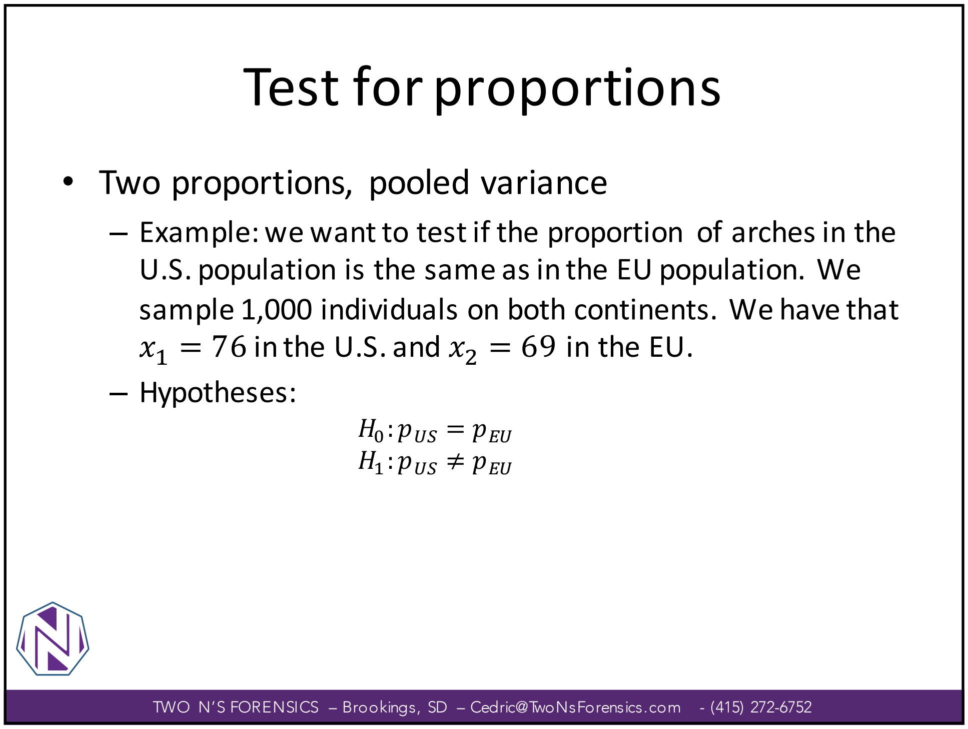
$$Z_{\frac{\alpha}{2}} = \frac{(\widehat{p_1}-\widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \text{ where } \hat{p} = \frac{x_1+x_2}{n_1+n_2}$$

Test for proportions

- Same concept, different test statistics
  - Good thing is that they are all Z's
  - One proportion test: $Z_{\frac{\alpha}{2}} = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}$
  - Two proportions, pooled variance

$$Z_{\frac{\alpha}{2}} = \frac{(\widehat{p_1}-\widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \text{ where } \hat{p} = \frac{x_1+x_2}{n_1+n_2}$$

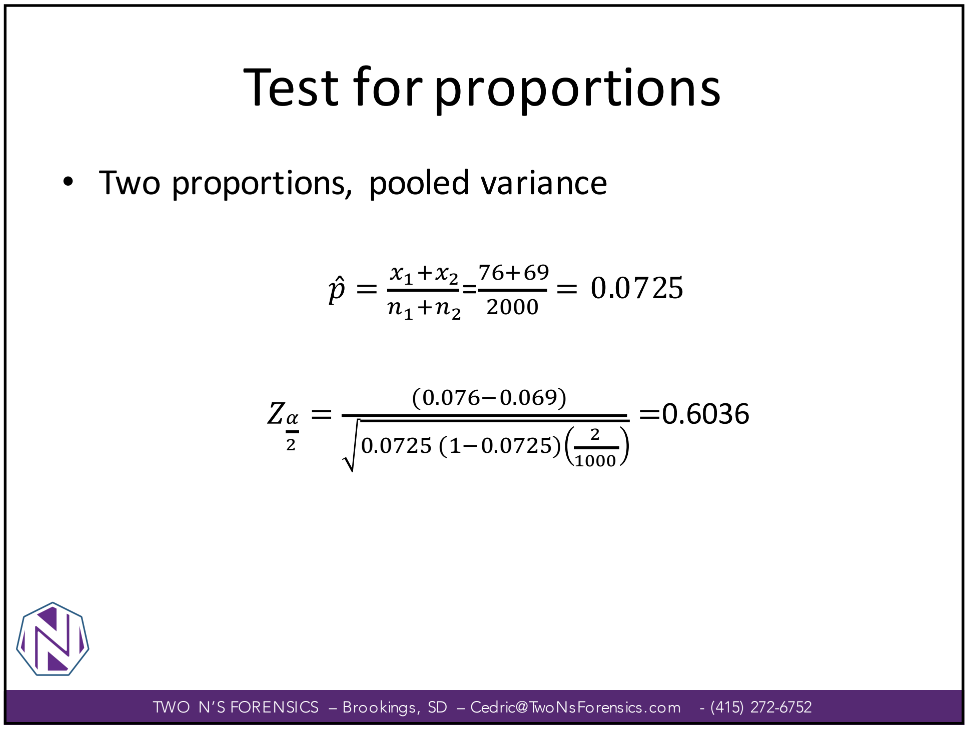
To guarantee that these tests will perform correctly, we need to have at least 10 "successes" (i.e., np=10)



Test for proportions

- Two proportions, pooled variance
  - Example: we want to test if the proportion of arches in the U.S. population is the same as in the EU population. We sample 1,000 individuals on both continents. We have that $x_1 = 76$ in the U.S. and $x_2 = 69$ in the EU.
  - Hypotheses:

$$H_0: p_{US} = p_{EU}$$
$$H_1: p_{US} \neq p_{EU}$$



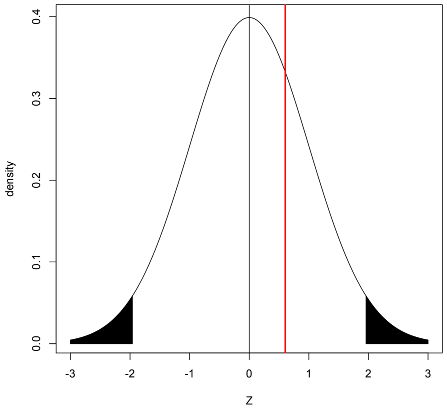
Test for proportions

- Two proportions, pooled variance

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{76+69}{2000} = 0.0725$$

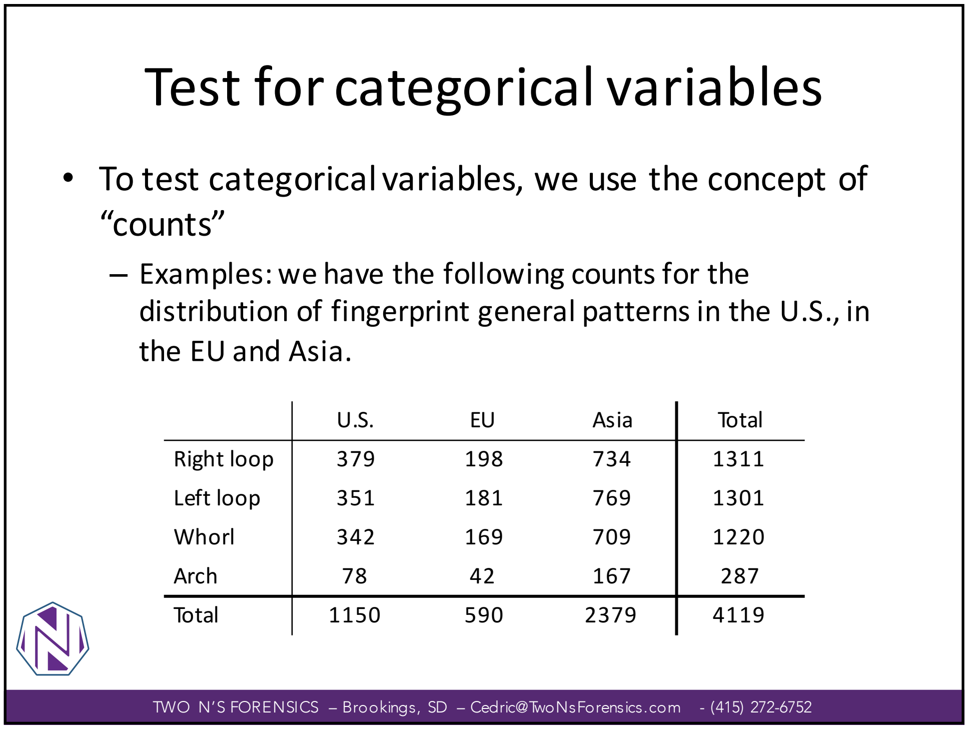
$$Z_{\frac{\alpha}{2}} = \frac{(0.076-0.069)}{\sqrt{0.0725\,(1-0.0725)\left(\frac{2}{1000}\right)}} = 0.6036$$

Test for proportions



Test for categorical variables

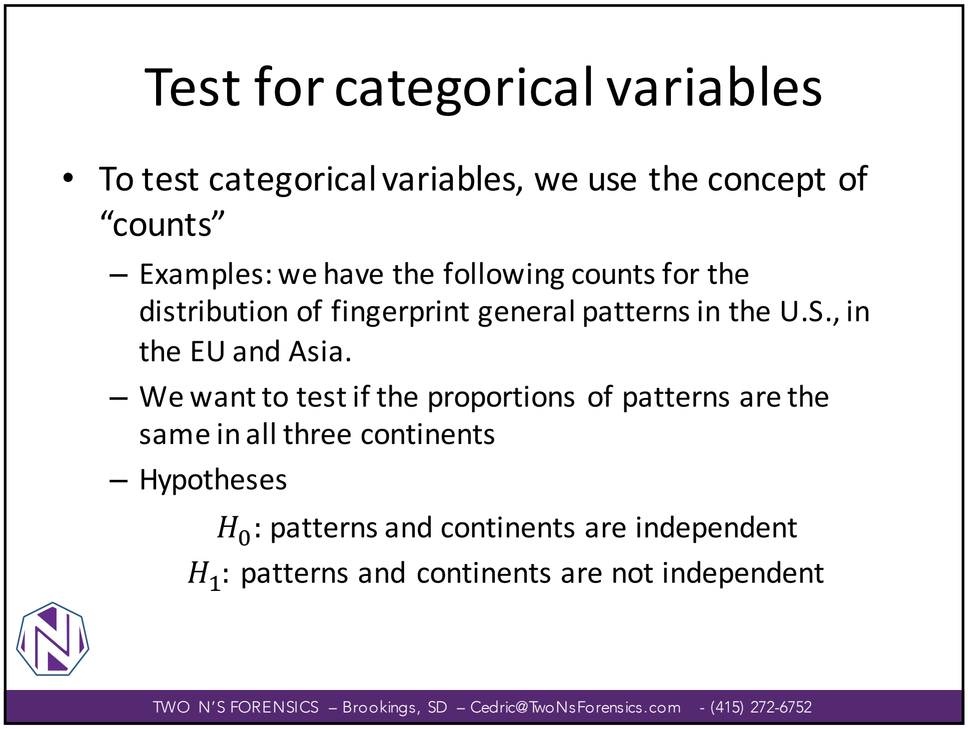
- To test categorical variables, we use the concept of "counts"
  - How many objects have we observed in category *i*?
  - Example:
    - How many brands of shoes with sole pattern *i*?
    - How many pills with Popeye design?
- We compare the observed counts, with the expected counts under H_0
- We use a chi-square statistics $\chi^2 = \sum_{i=1}^{k} \left(\frac{(O_k - E_k)^2}{E_k} \right)$, where $k = r \times c$ is the number of categories
  - χ^2 has degrees of freedom $(r-1)(c-1)$



Test for categorical variables

- To test categorical variables, we use the concept of "counts"
  - Examples: we have the following counts for the distribution of fingerprint general patterns in the U.S., in the EU and Asia.

	U.S.	EU	Asia	Total
Right loop	379	198	734	1311
Left loop	351	181	769	1301
Whorl	342	169	709	1220
Arch	78	42	167	287
Total	1150	590	2379	4119

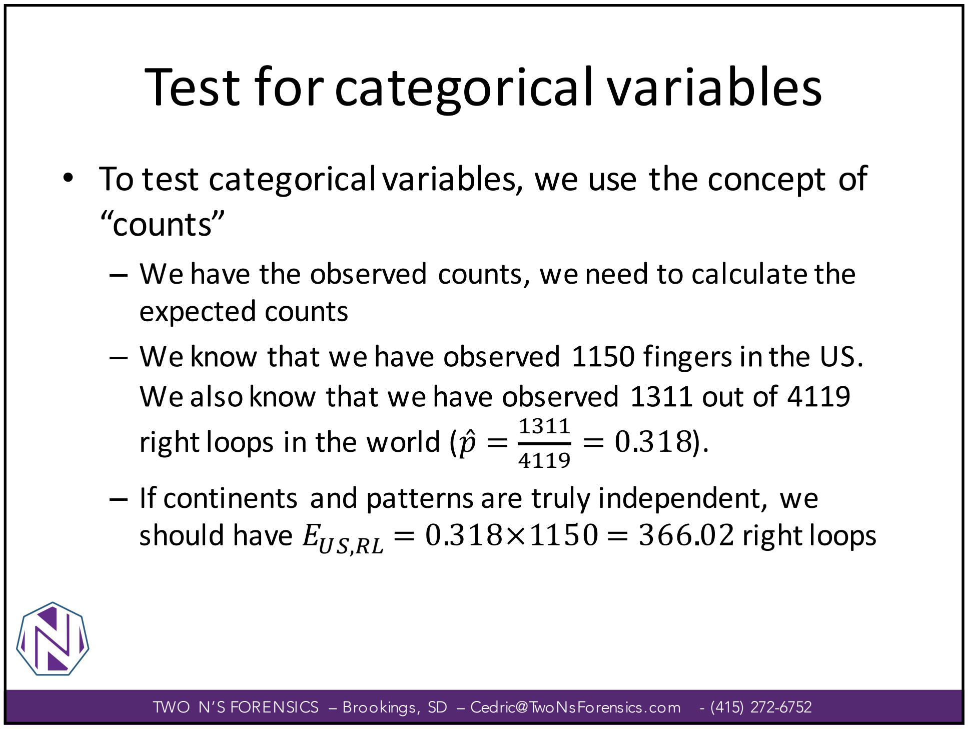


Test for categorical variables

- To test categorical variables, we use the concept of "counts"
  - Examples: we have the following counts for the distribution of fingerprint general patterns in the U.S., in the EU and Asia.
  - We want to test if the proportions of patterns are the same in all three continents
  - Hypotheses

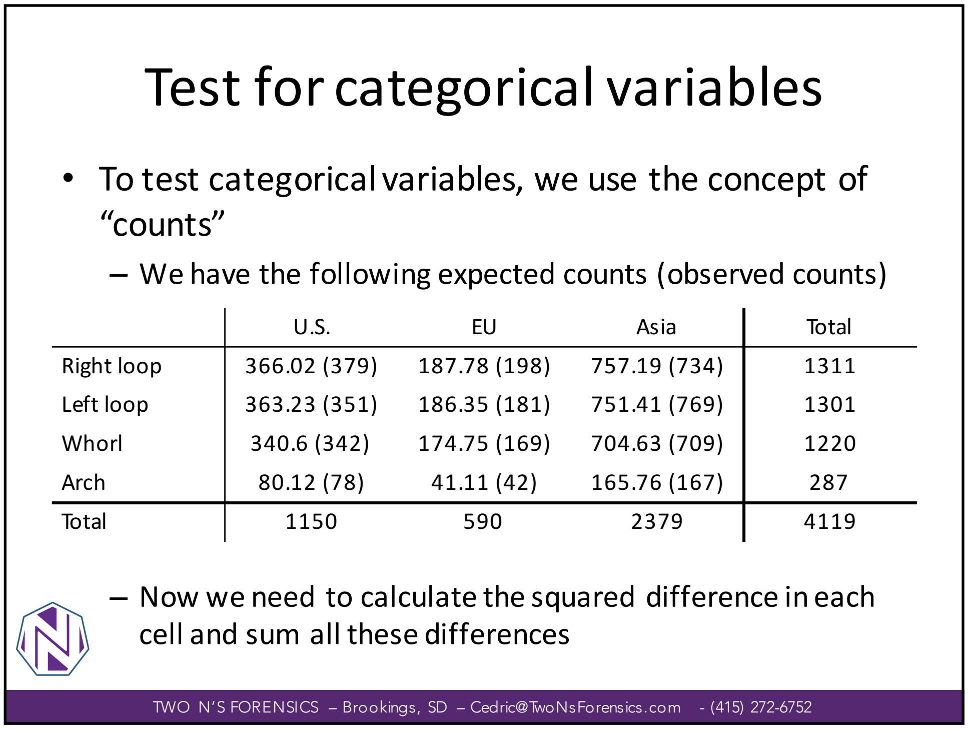
H_0: patterns and continents are independent

H_1: patterns and continents are not independent



Test for categorical variables

- To test categorical variables, we use the concept of "counts"
  - We have the observed counts, we need to calculate the expected counts
  - We know that we have observed 1150 fingers in the US. We also know that we have observed 1311 out of 4119 right loops in the world ($\hat{p} = \frac{1311}{4119} = 0.318$).
  - If continents and patterns are truly independent, we should have $E_{US,RL} = 0.318 \times 1150 = 366.02$ right loops



Test for categorical variables

- To test categorical variables, we use the concept of "counts"
  - We have the following expected counts (observed counts)

	U.S.	EU	Asia	Total
Right loop	366.02 (379)	187.78 (198)	757.19 (734)	1311
Left loop	363.23 (351)	186.35 (181)	751.41 (769)	1301
Whorl	340.6 (342)	174.75 (169)	704.63 (709)	1220
Arch	80.12 (78)	41.11 (42)	165.76 (167)	287
Total	1150	590	2379	4119

  - Now we need to calculate the squared difference in each cell and sum all these differences

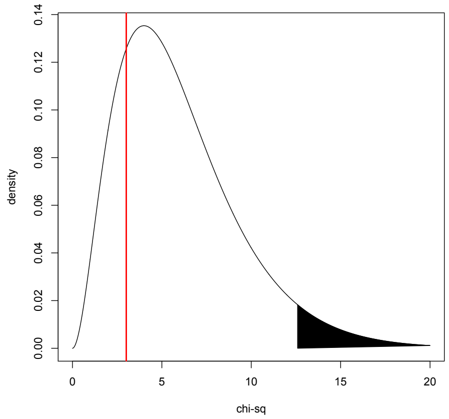
Test for categorical variables

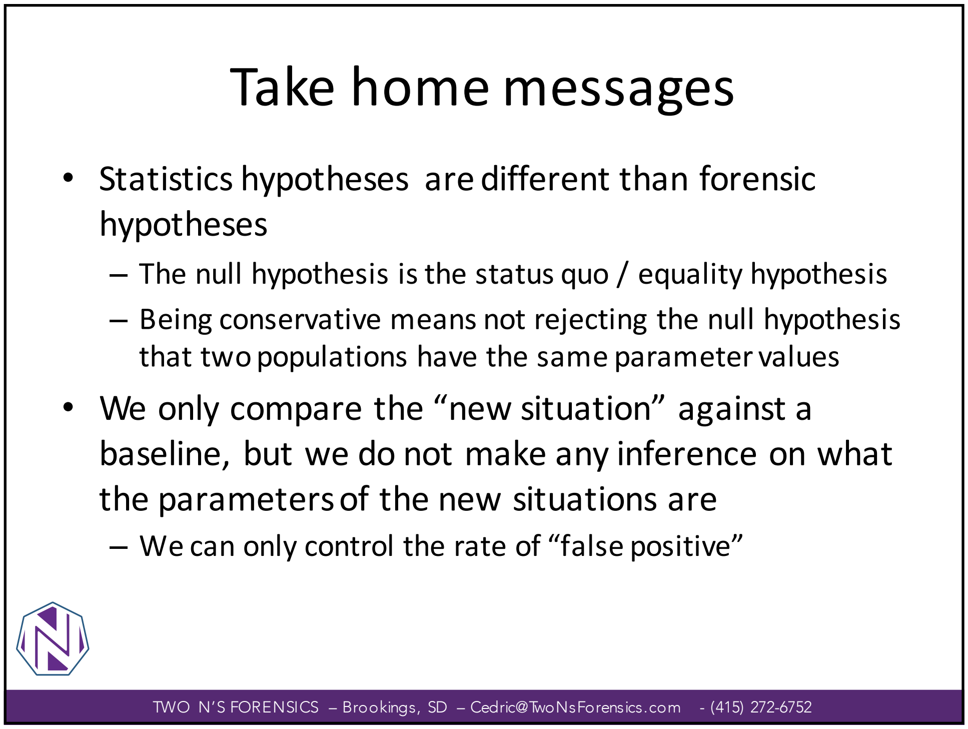
- To test categorical variables, we use the concept of "counts"
  - We have the following expected counts (observed counts)
  - Now we need to calculate the squared difference in each cell and sum all these differences

$$\chi^2 = \frac{(379 - 366.02)^2}{366.02} + \frac{(351 - 363.23)^2}{363.23} + \cdots + \frac{(167 - 165.76)^2}{165.76} = 3.01$$

$$df = (4 - 1)\times(3 - 1) = 6$$

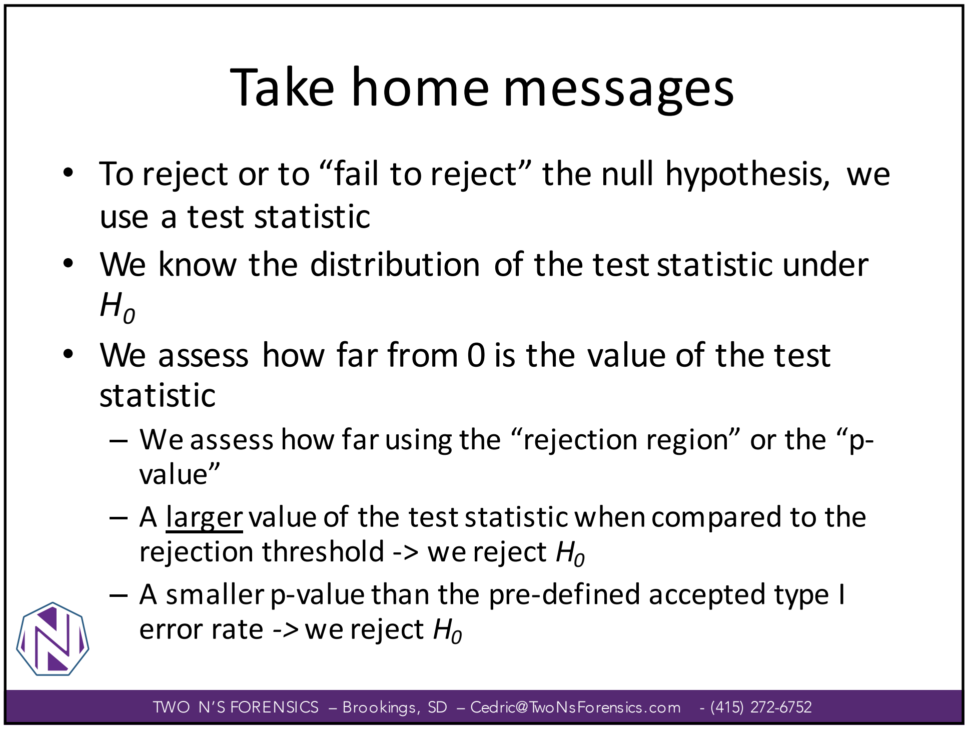
Test for categorical variables





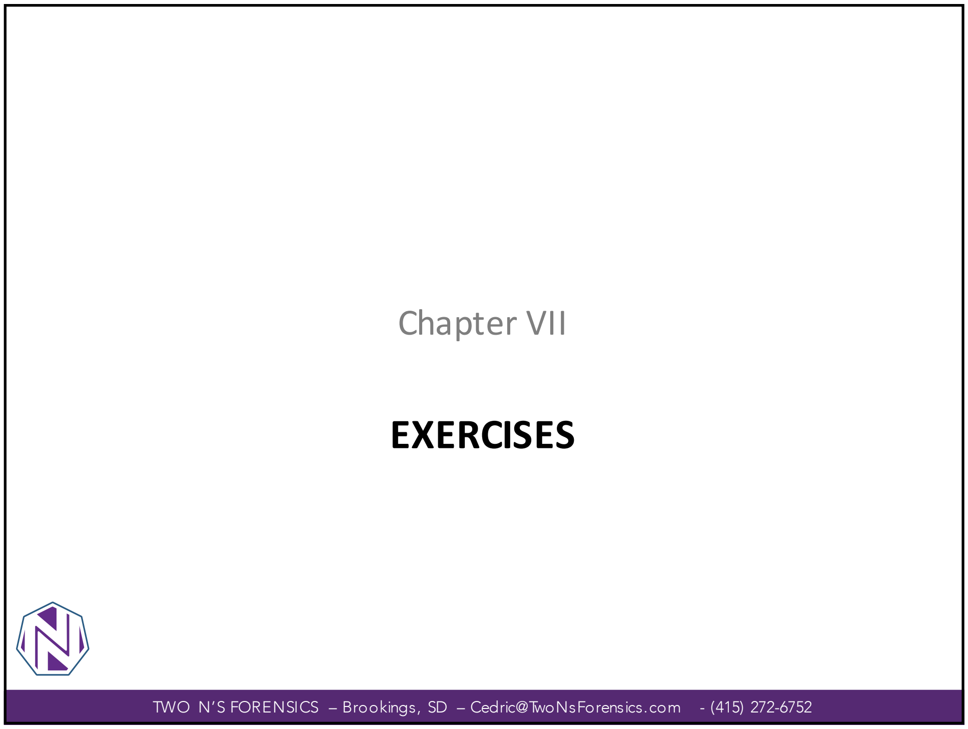
Take home messages

- Statistics hypotheses are different than forensic hypotheses
  - The null hypothesis is the status quo / equality hypothesis
  - Being conservative means not rejecting the null hypothesis that two populations have the same parameter values
- We only compare the "new situation" against a baseline, but we do not make any inference on what the parameters of the new situations are
  - We can only control the rate of "false positive"



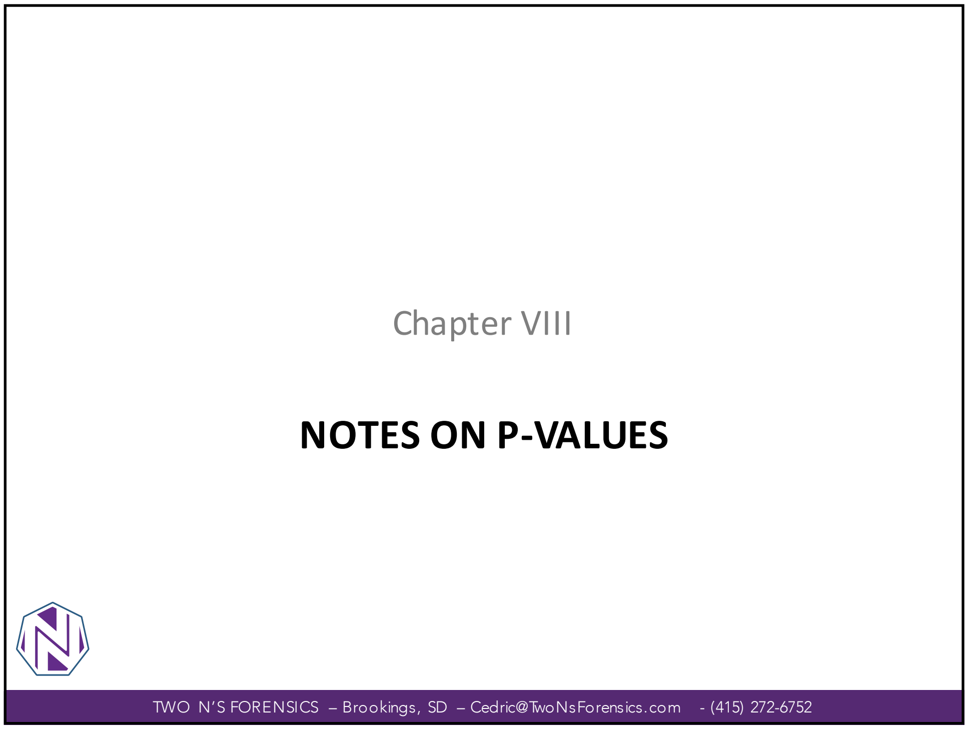
Take home messages

- To reject or to "fail to reject" the null hypothesis, we use a test statistic
- We know the distribution of the test statistic under H_0
- We assess how far from 0 is the value of the test statistic
  - We assess how far using the "rejection region" or the "p-value"
  - A larger value of the test statistic when compared to the rejection threshold -> we reject H_0
  - A smaller p-value than the pre-defined accepted type I error rate -> we reject H_0



Chapter VII

EXERCISES

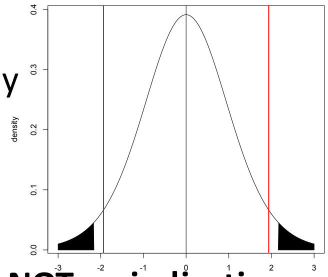


Chapter VIII

NOTES ON P-VALUES

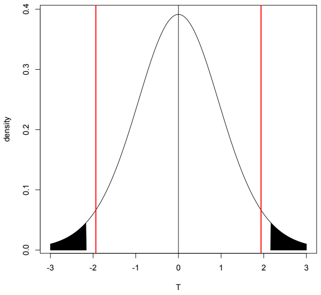
Notes on p-values

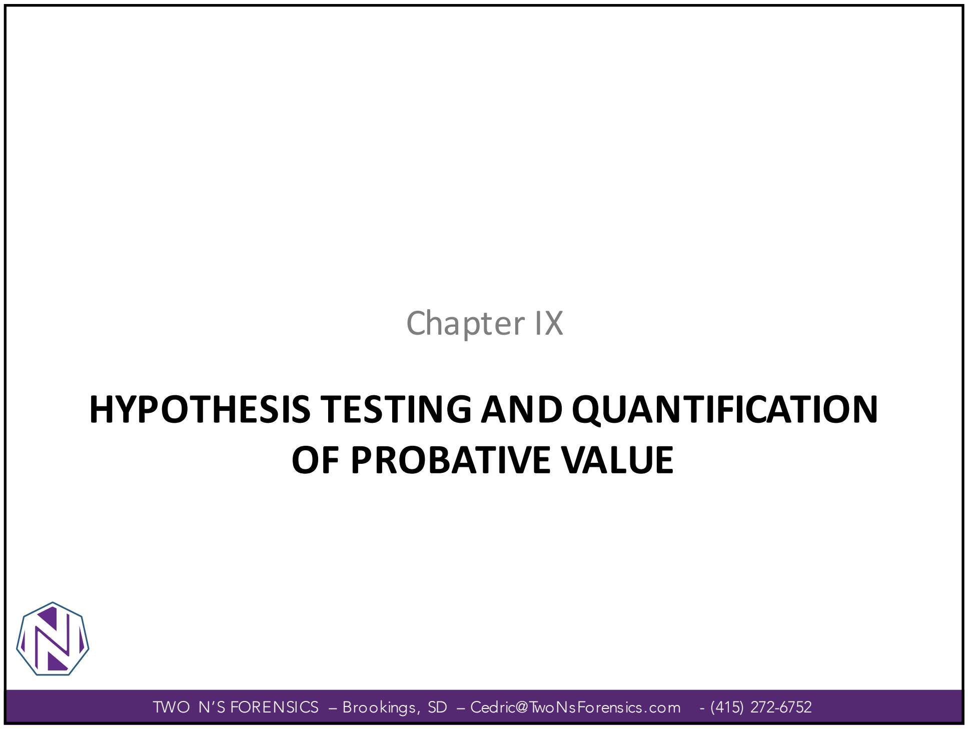
- A p-value is a probability
- It is the probability to observe a value of the test statistic that is more extreme than the one we calculated when H_0 is true
- It is the probability of erroneously rejecting H_0
- **It is NOT the probability that H_0 or H_1 is true**
- **The magnitude of the p-value is NOT an indication of the strength of the (lack of) association**



Notes on p-values

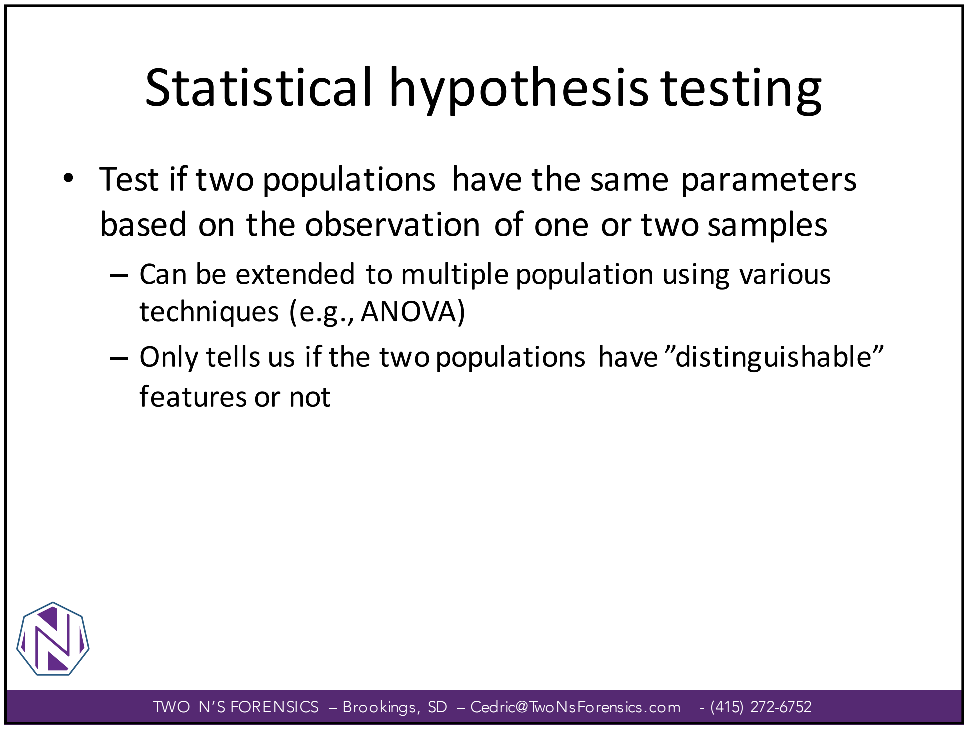
- A p-value is criteria for decision.
- It is a value that needs to be compared to a pre-defined threshold
- Hard and fast rule!





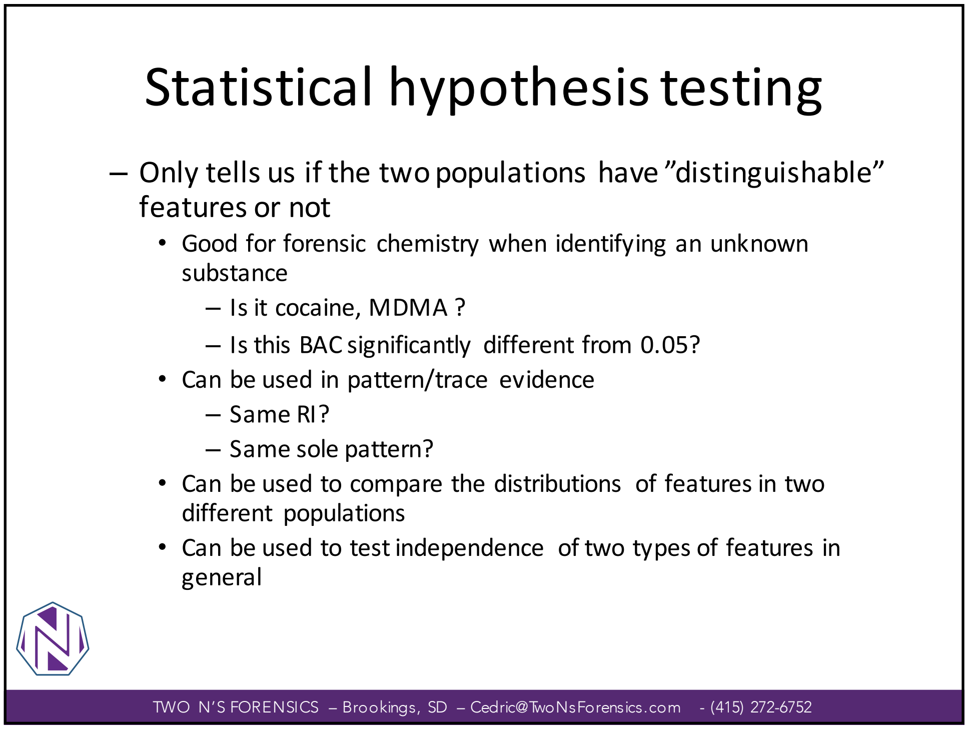
Chapter IX

HYPOTHESIS TESTING AND QUANTIFICATION OF PROBATIVE VALUE



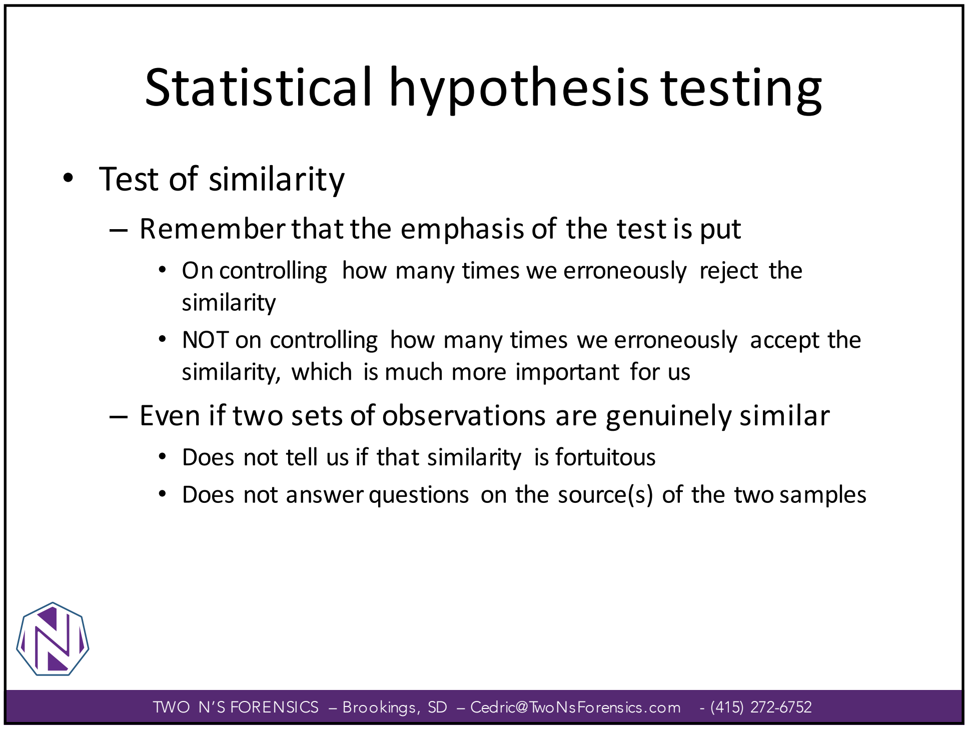
Statistical hypothesis testing

- Test if two populations have the same parameters based on the observation of one or two samples
  - Can be extended to multiple population using various techniques (e.g., ANOVA)
  - Only tells us if the two populations have ”distinguishable” features or not



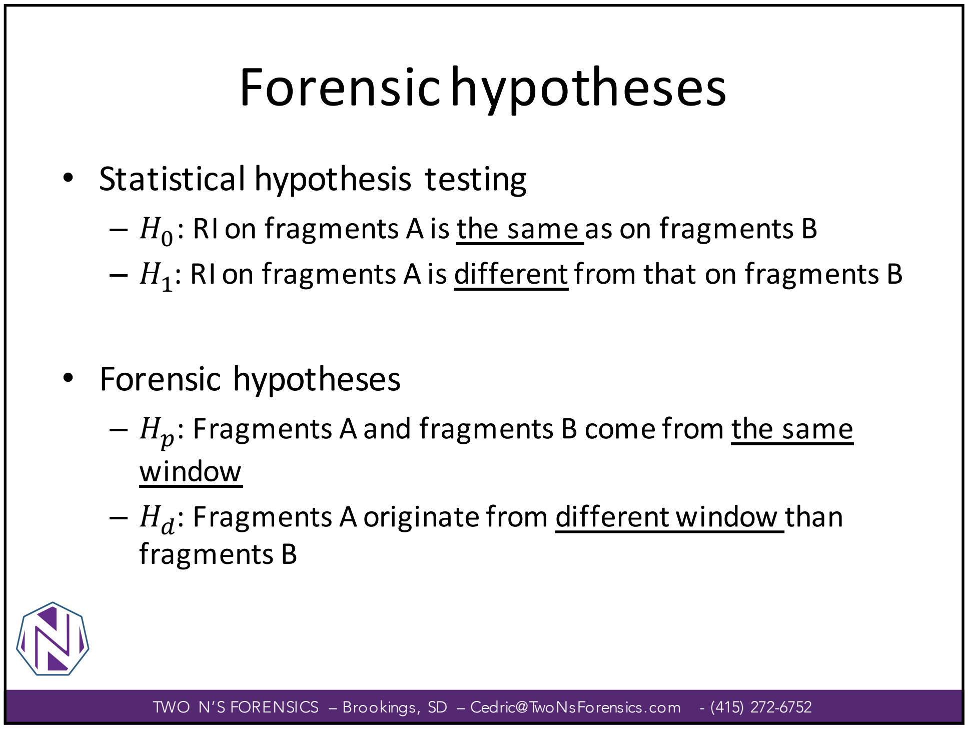
Statistical hypothesis testing

- Only tells us if the two populations have ”distinguishable” features or not
  - Good for forensic chemistry when identifying an unknown substance
    - Is it cocaine, MDMA ?
    - Is this BAC significantly different from 0.05?
  - Can be used in pattern/trace evidence
    - Same RI ?
    - Same sole pattern?
  - Can be used to compare the distributions of features in two different populations
  - Can be used to test independence of two types of features in general



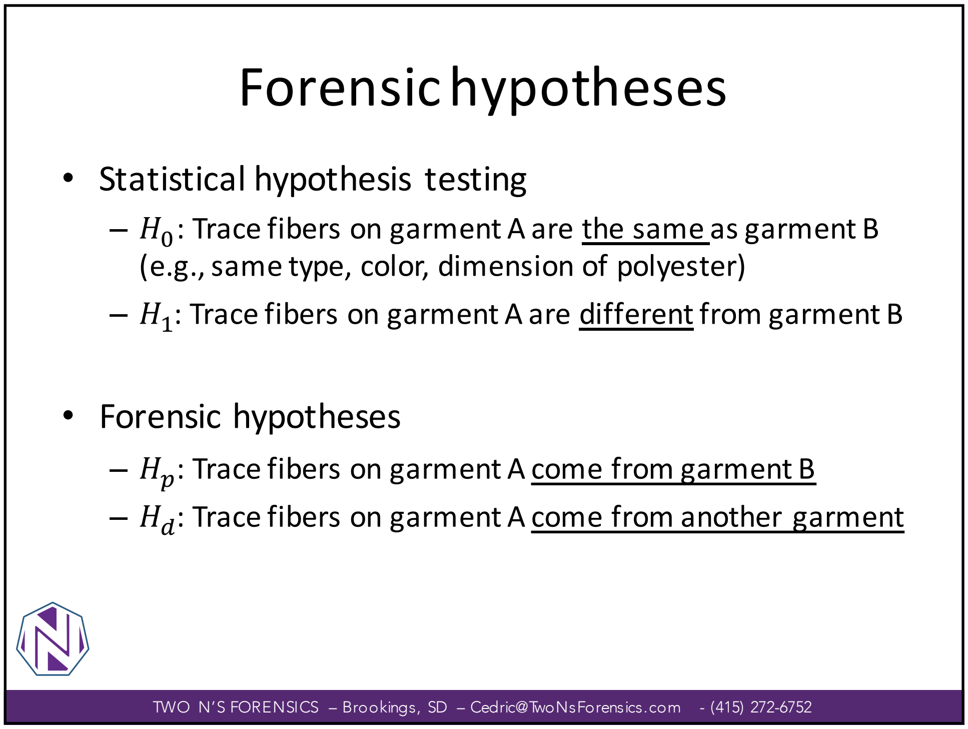
Statistical hypothesis testing

- Test of similarity
  - Remember that the emphasis of the test is put
    - On controlling how many times we erroneously reject the similarity
    - NOT on controlling how many times we erroneously accept the similarity, which is much more important for us
  - Even if two sets of observations are genuinely similar
    - Does not tell us if that similarity is fortuitous
    - Does not answer questions on the source(s) of the two samples



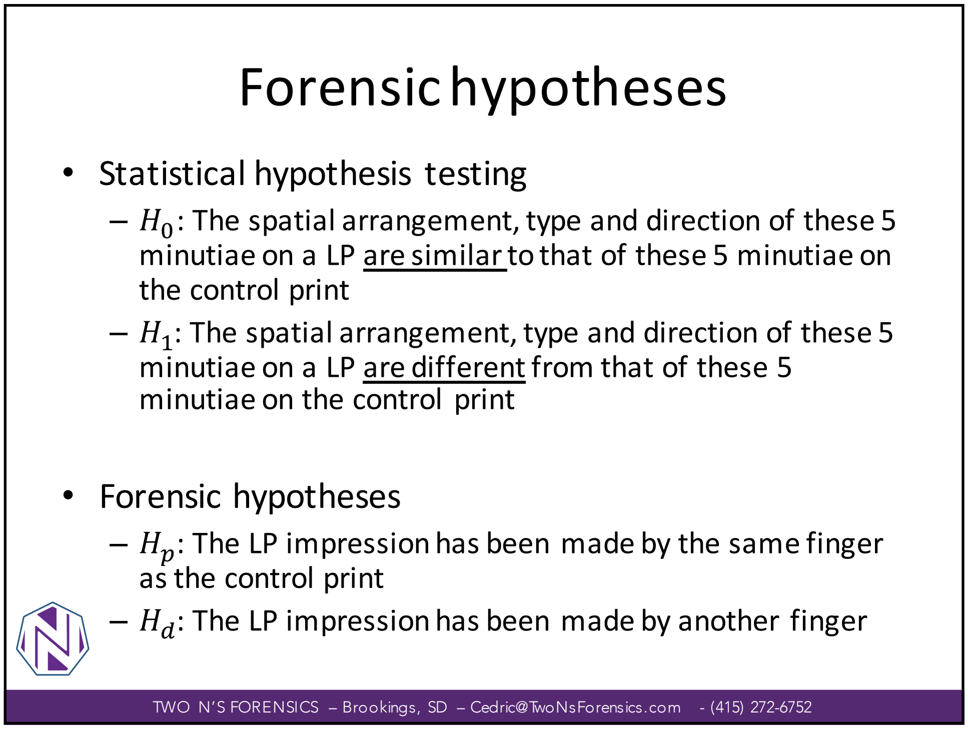
Forensic hypotheses

- Statistical hypothesis testing
  - H_0: RI on fragments A is the same as on fragments B
  - H_1: RI on fragments A is different from that on fragments B
- Forensic hypotheses
  - H_p: Fragments A and fragments B come from the same window
  - H_d: Fragments A originate from different window than fragments B



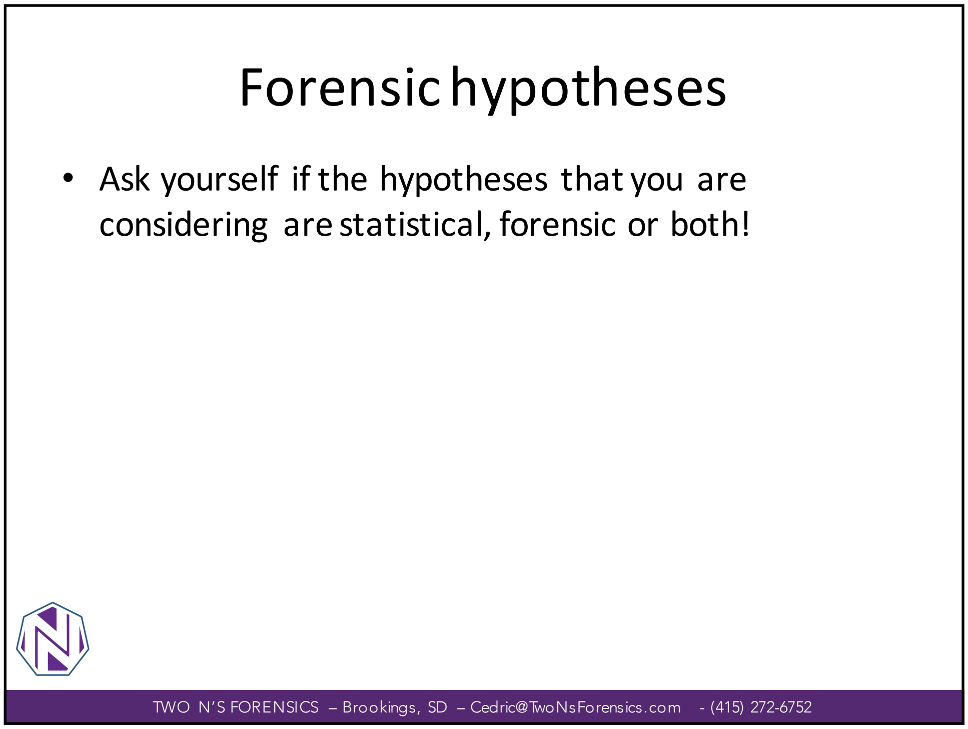
Forensic hypotheses

- Statistical hypothesis testing
  - H_0: Trace fibers on garment A are the same as garment B (e.g., same type, color, dimension of polyester)
  - H_1: Trace fibers on garment A are different from garment B
- Forensic hypotheses
  - H_p: Trace fibers on garment A come from garment B
  - H_d: Trace fibers on garment A come from another garment



Forensic hypotheses

- Statistical hypothesis testing
  - H_0: The spatial arrangement, type and direction of these 5 minutiae on a LP are similar to that of these 5 minutiae on the control print
  - H_1: The spatial arrangement, type and direction of these 5 minutiae on a LP are different from that of these 5 minutiae on the control print
- Forensic hypotheses
  - H_p: The LP impression has been made by the same finger as the control print
  - H_d: The LP impression has been made by another finger



Forensic hypotheses

- Ask yourself if the hypotheses that you are considering are statistical, forensic or both!

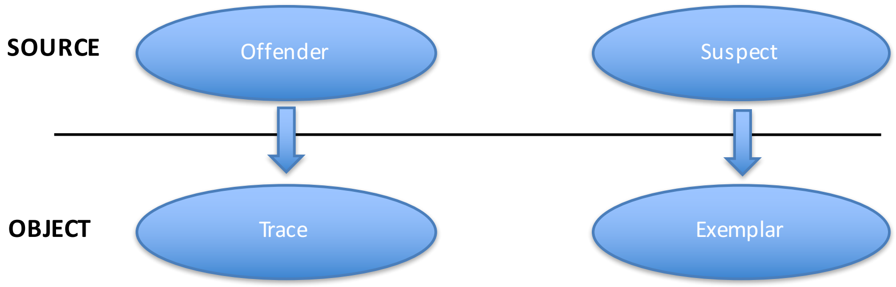
Forensic inference

SOURCE

Offender

Suspect

Forensic inference



SOURCE

Offender

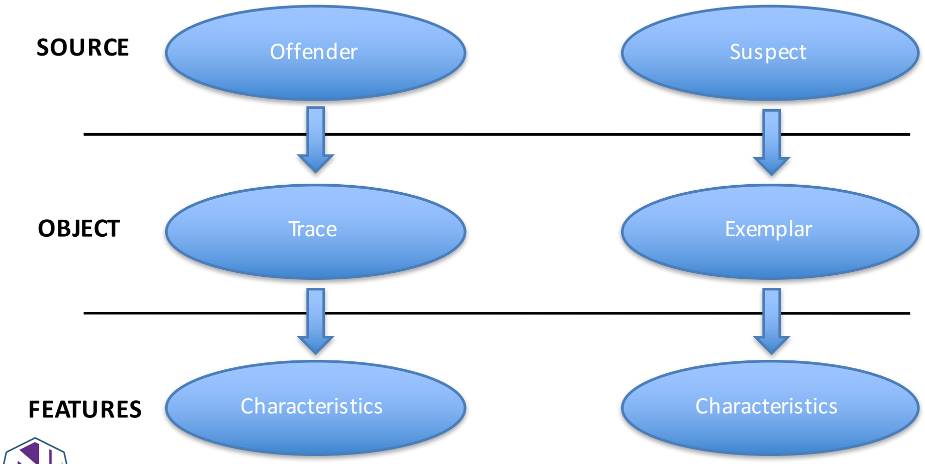
Suspect

OBJECT

Trace

Exemplar

Forensic inference



SOURCE

Offender

Suspect

OBJECT

Trace

Exemplar

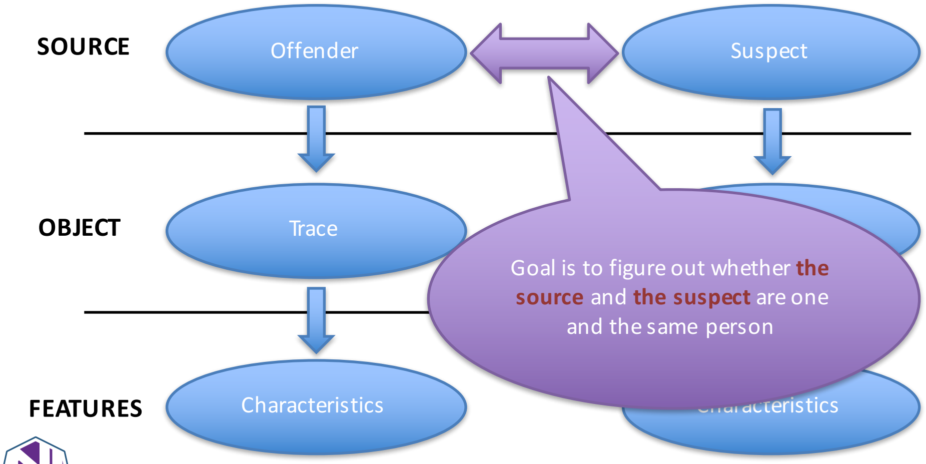
FEATURES

Characteristics

Characteristics

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Forensic inference



SOURCE

Offender

Suspect

OBJECT

Trace

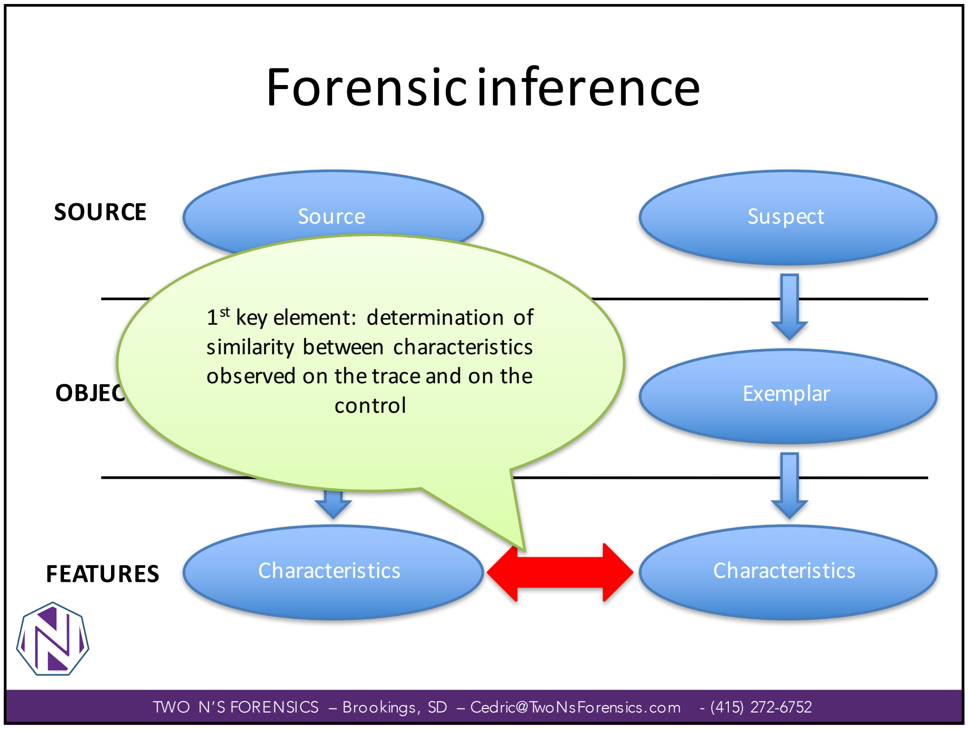
FEATURES

Characteristics

Characteristics

Goal is to figure out whether **the source** and **the suspect** are one and the same person

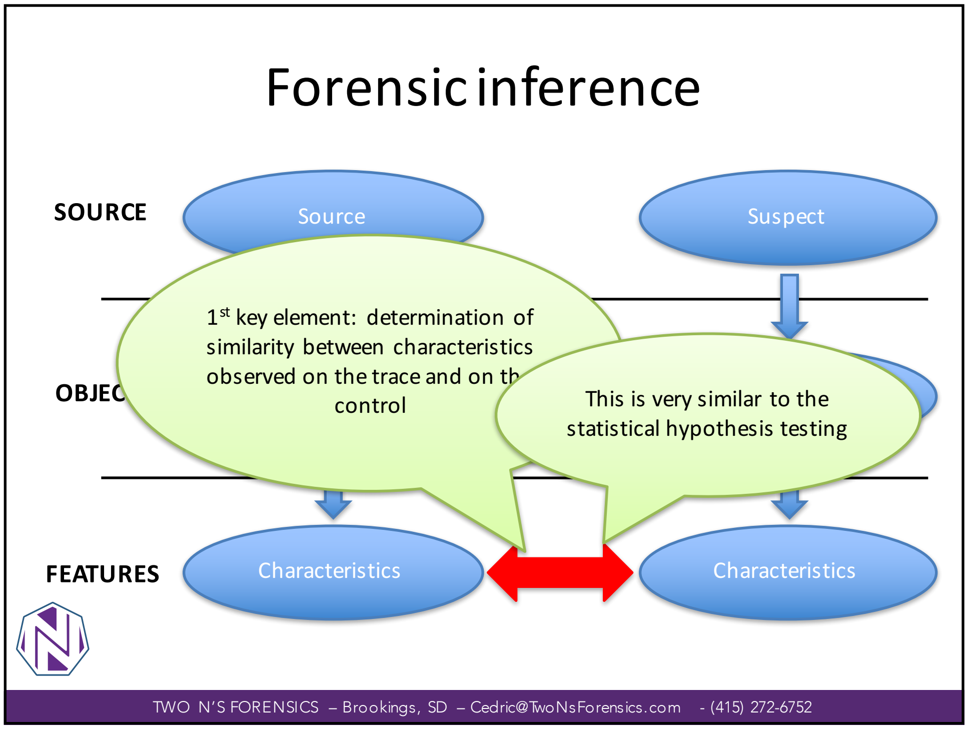
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Forensic inference

1st key element: determination of similarity between characteristics observed on the trace and on the control

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



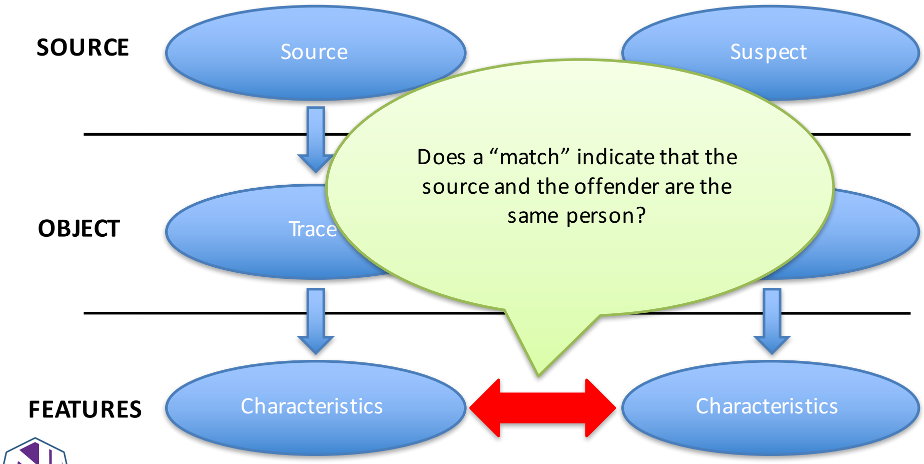
Forensic inference

1st key element: determination of similarity between characteristics observed on the trace and on th control

This is very similar to the statistical hypothesis testing

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

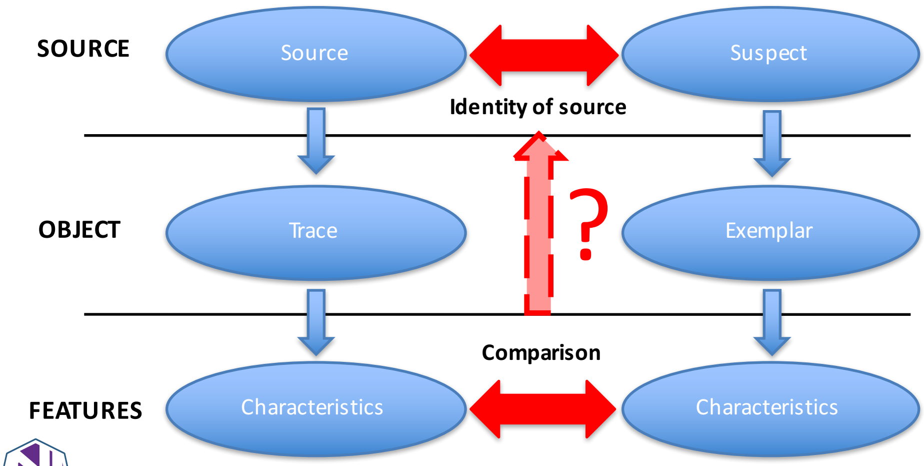
Forensic inference



Does a “match” indicate that the source and the offender are the same person?

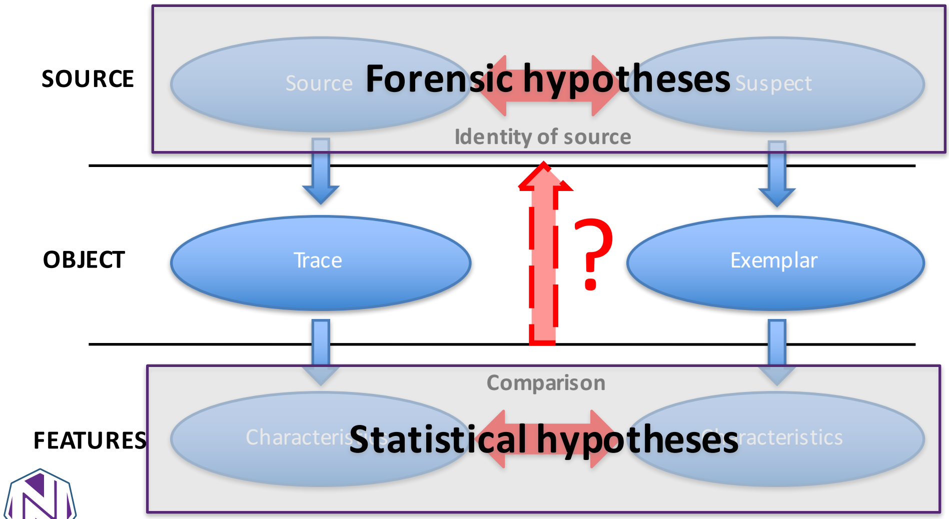
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Forensic inference



TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Forensic inference



SOURCE

Source

Forensic hypotheses

Suspect

Identity of source

OBJECT

Trace

?

Exemplar

Comparison

FEATURES

Characteris

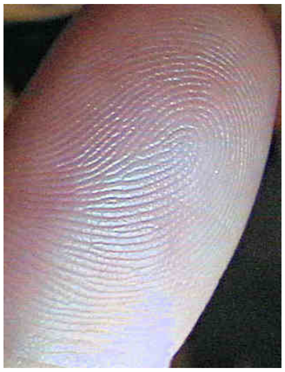
Statistical hypotheses

Characteristics

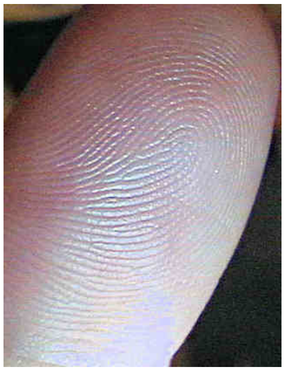
Forensic inference

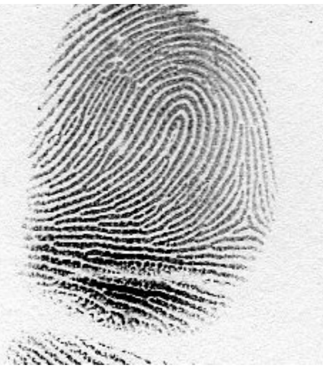
- So how do we justify the jump?
  - Uniqueness
  - Discriminating power
  - Earth population

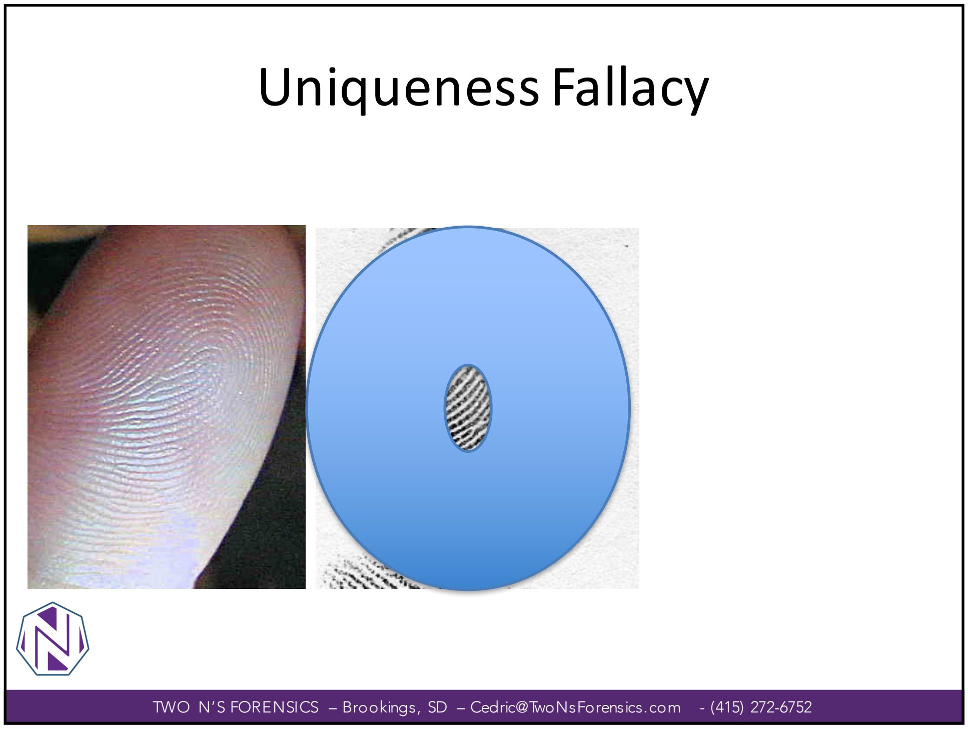
Uniqueness Fallacy



Uniqueness Fallacy

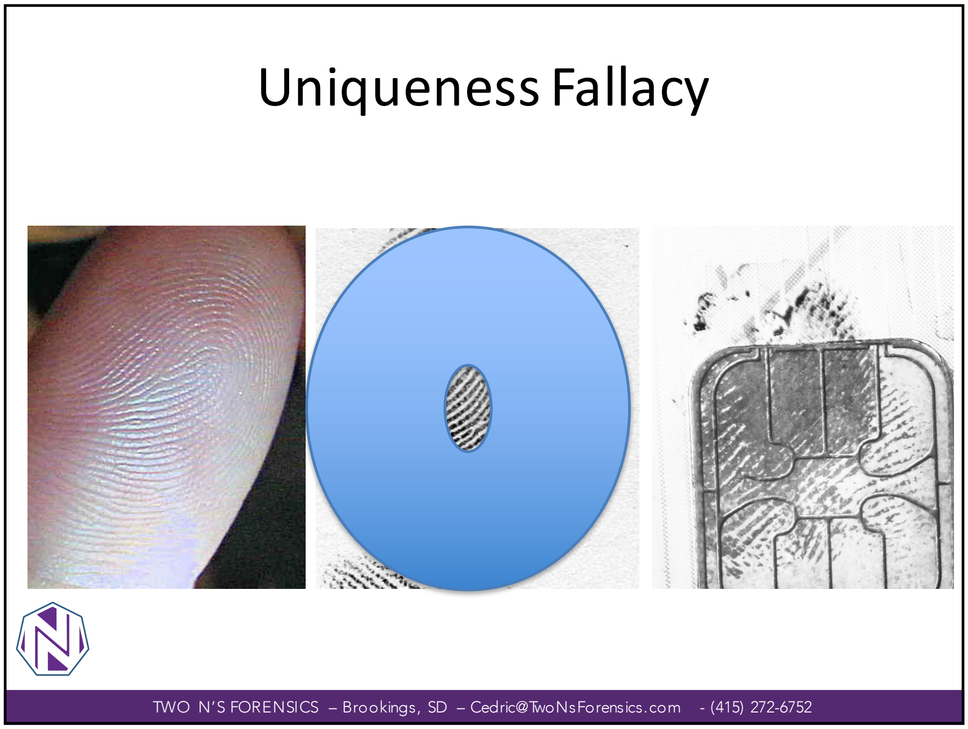






Uniqueness Fallacy

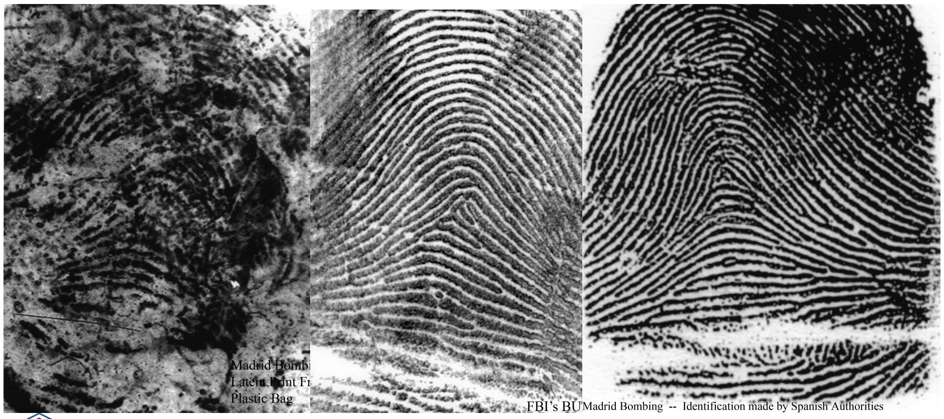
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



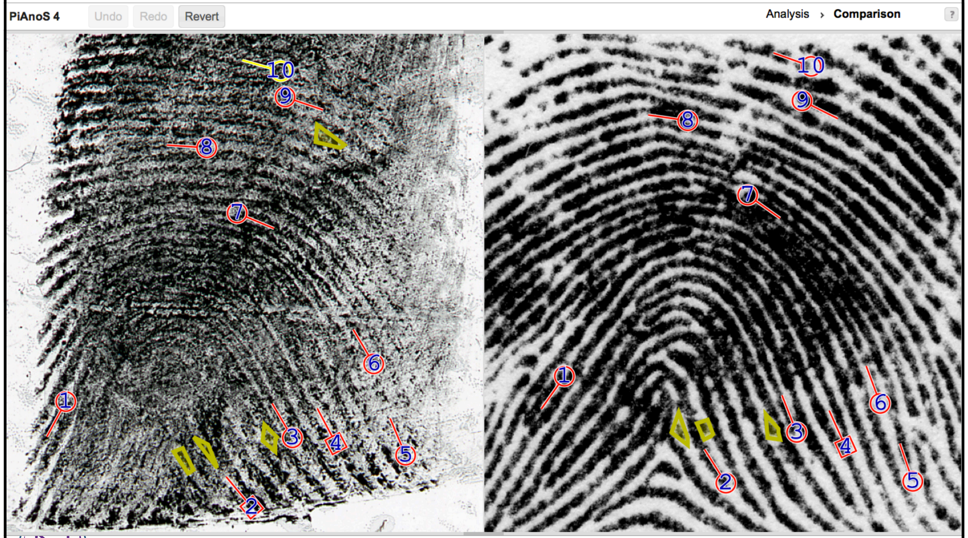
Uniqueness Fallacy

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

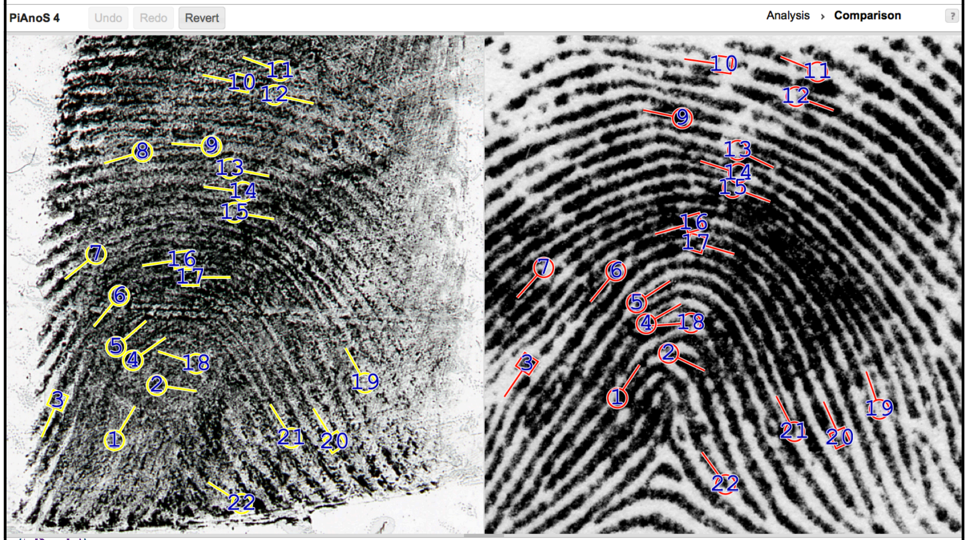
Uniqueness Fallacy


Madrid Bombi
Latent Print Fr
Plastic Bag
FBI's BU
Madrid Bombing -- Identification made by Spanish Authorities

Example 1

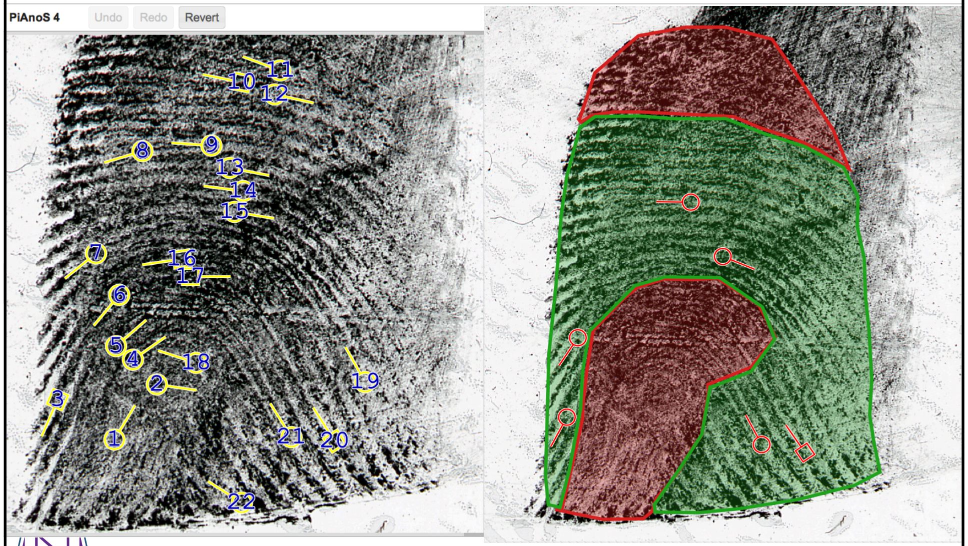


Example 2



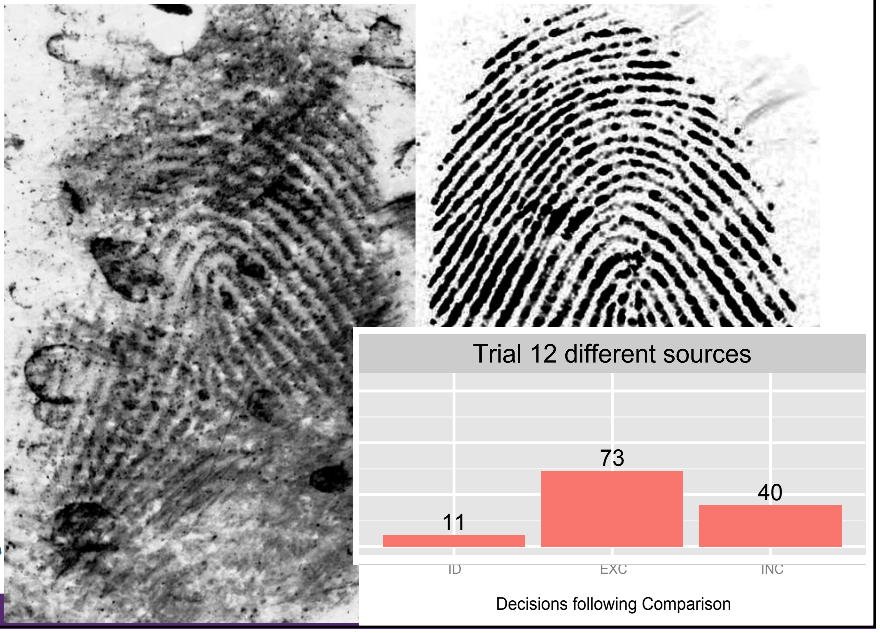
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Example 2

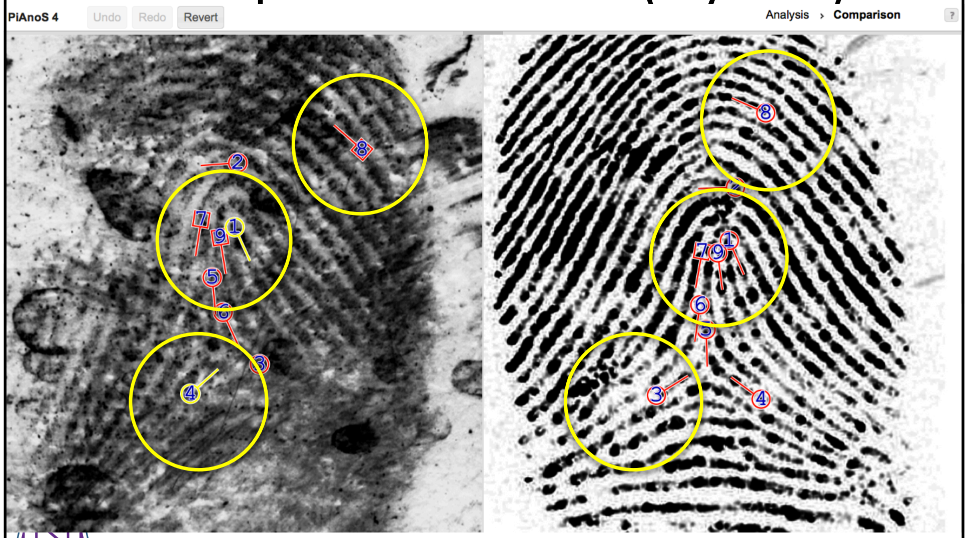


TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

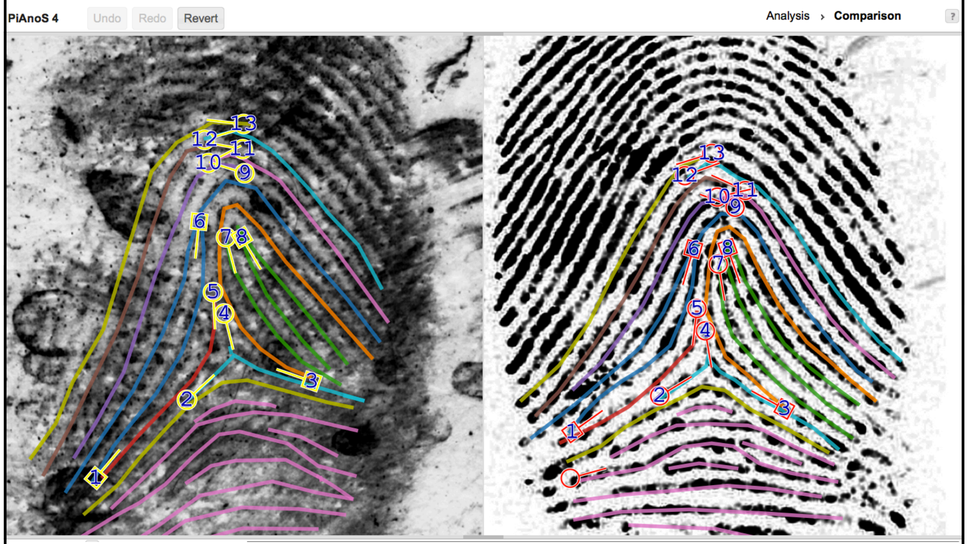
Trial 12



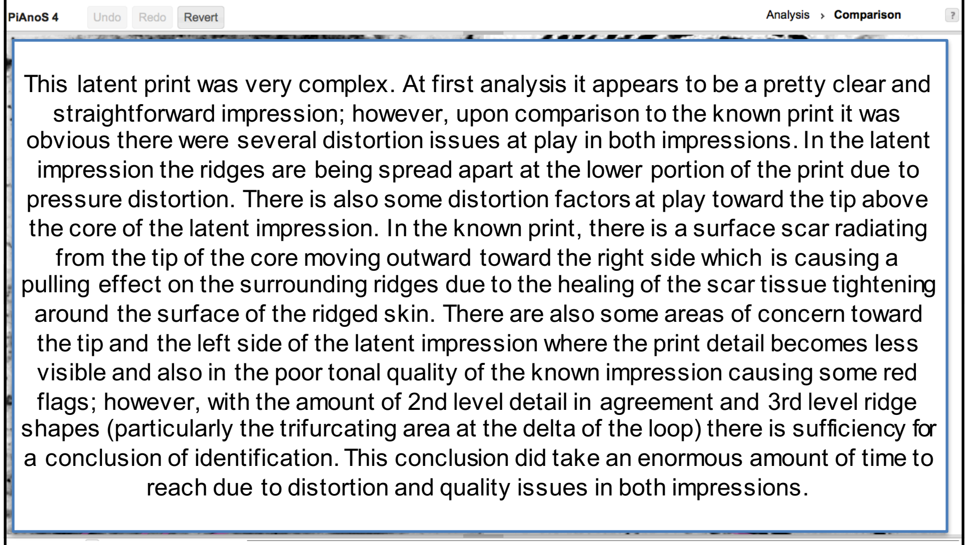
Example 3 – Certified (5 years)



Example 4 – Certified (7 years)

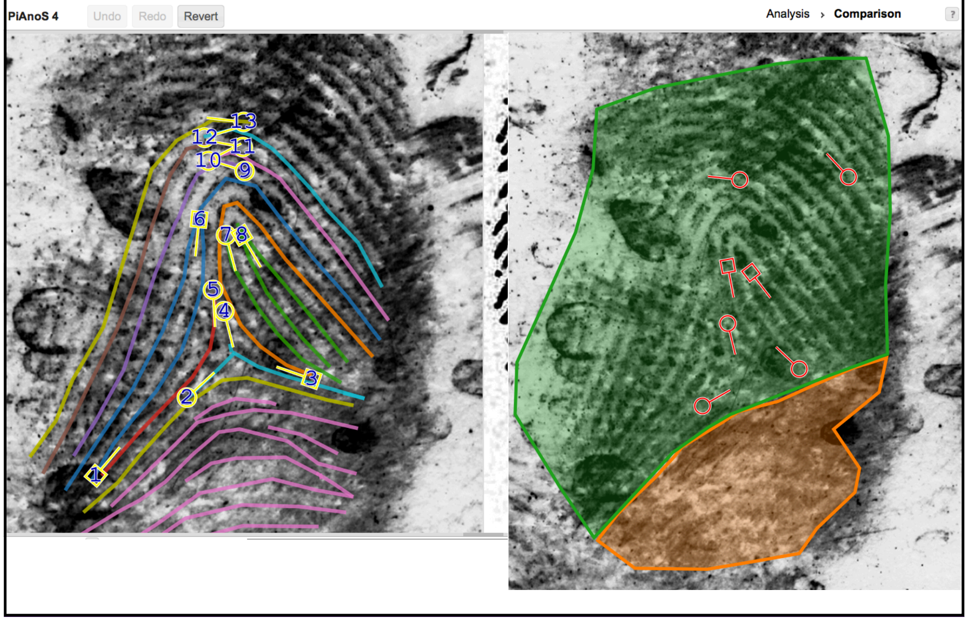


Example 4 – Certified (7 years)



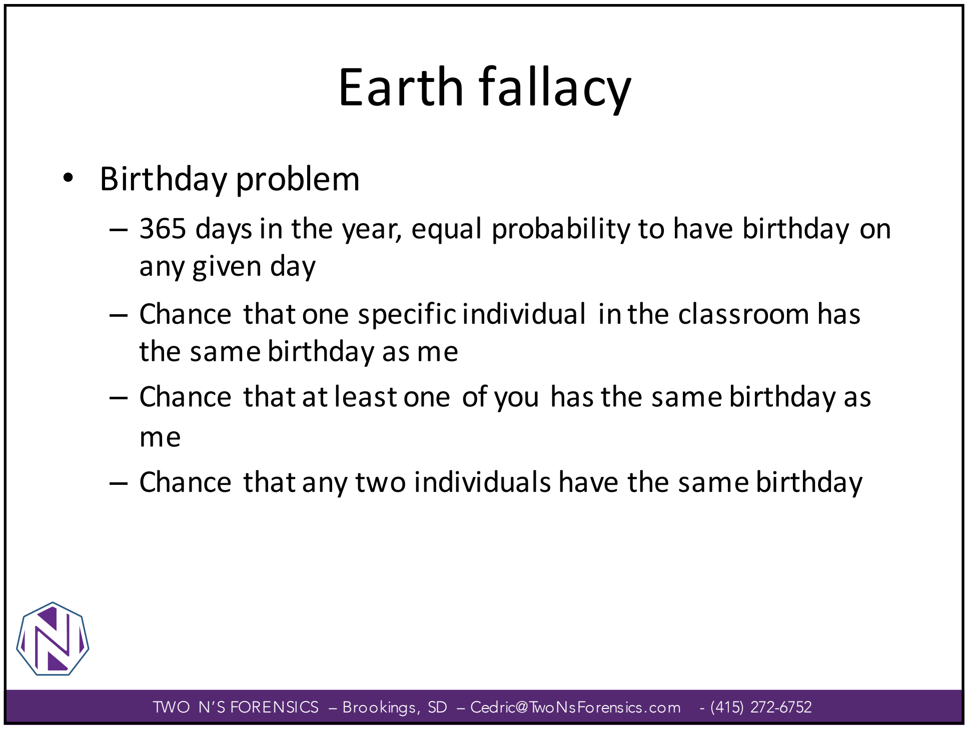
This latent print was very complex. At first analysis it appears to be a pretty clear and straightforward impression; however, upon comparison to the known print it was obvious there were several distortion issues at play in both impressions. In the latent impression the ridges are being spread apart at the lower portion of the print due to pressure distortion. There is also some distortion factors at play toward the tip above the core of the latent impression. In the known print, there is a surface scar radiating from the tip of the core moving outward toward the right side which is causing a pulling effect on the surrounding ridges due to the healing of the scar tissue tightening around the surface of the ridged skin. There are also some areas of concern toward the tip and the left side of the latent impression where the print detail becomes less visible and also in the poor tonal quality of the known impression causing some red flags; however, with the amount of 2nd level detail in agreement and 3rd level ridge shapes (particularly the trifurcating area at the delta of the loop) there is sufficiency for a conclusion of identification. This conclusion did take an enormous amount of time to reach due to distortion and quality issues in both impressions.

Example 4 – Certified (7 years)



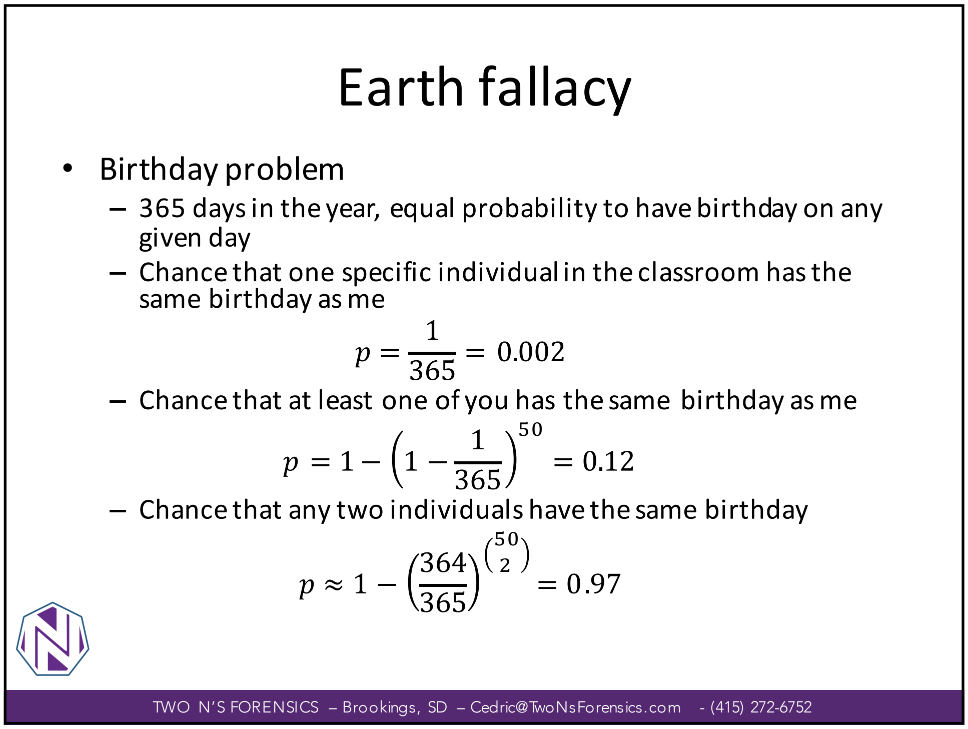
Earth fallacy

- If we have a type of feature that is very very discriminating (not unique, but close)
  - Say a good quality arrangement of 25 minutiae
  - Say the probability to observe any given set of 25 minutiae by chance is $\frac{1}{7{,}000{,}000{,}000}$
  - If we observe a LP and CP with the same arrangement of 25 minutiae, it has to be him, right?



Earth fallacy

- Birthday problem
  - 365 days in the year, equal probability to have birthday on any given day
  - Chance that one specific individual in the classroom has the same birthday as me
  - Chance that at least one of you has the same birthday as me
  - Chance that any two individuals have the same birthday



Earth fallacy

- Birthday problem
  - 365 days in the year, equal probability to have birthday on any given day
  - Chance that one specific individual in the classroom has the same birthday as me

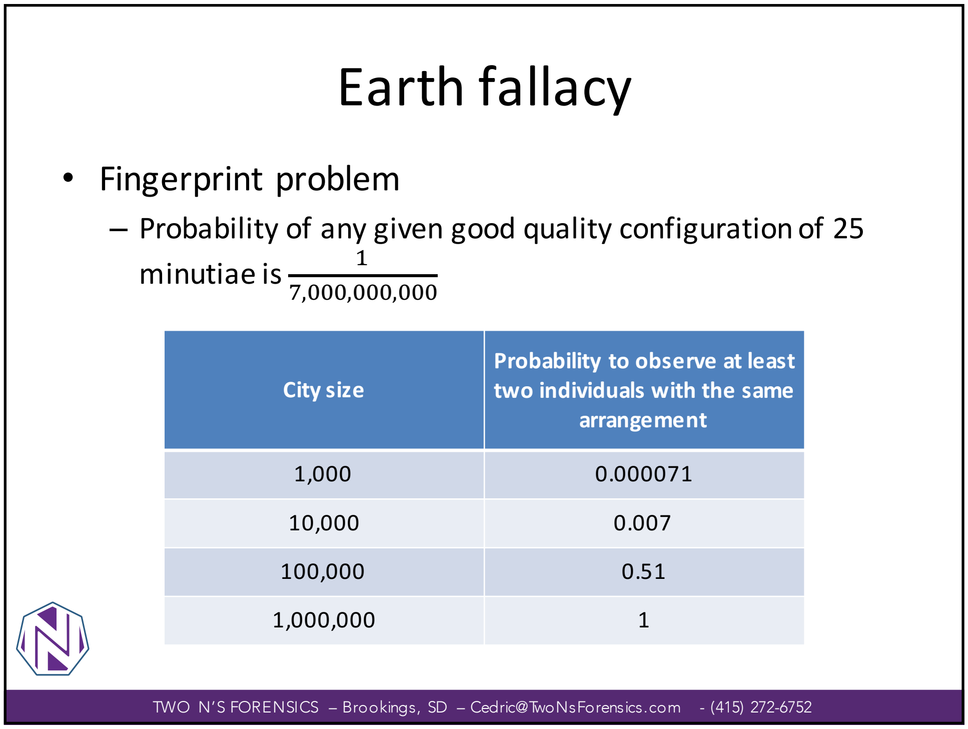
$$p = \frac{1}{365} = 0.002$$

  - Chance that at least one of you has the same birthday as me

$$p = 1 - \left(1 - \frac{1}{365}\right)^{50} = 0.12$$

  - Chance that any two individuals have the same birthday

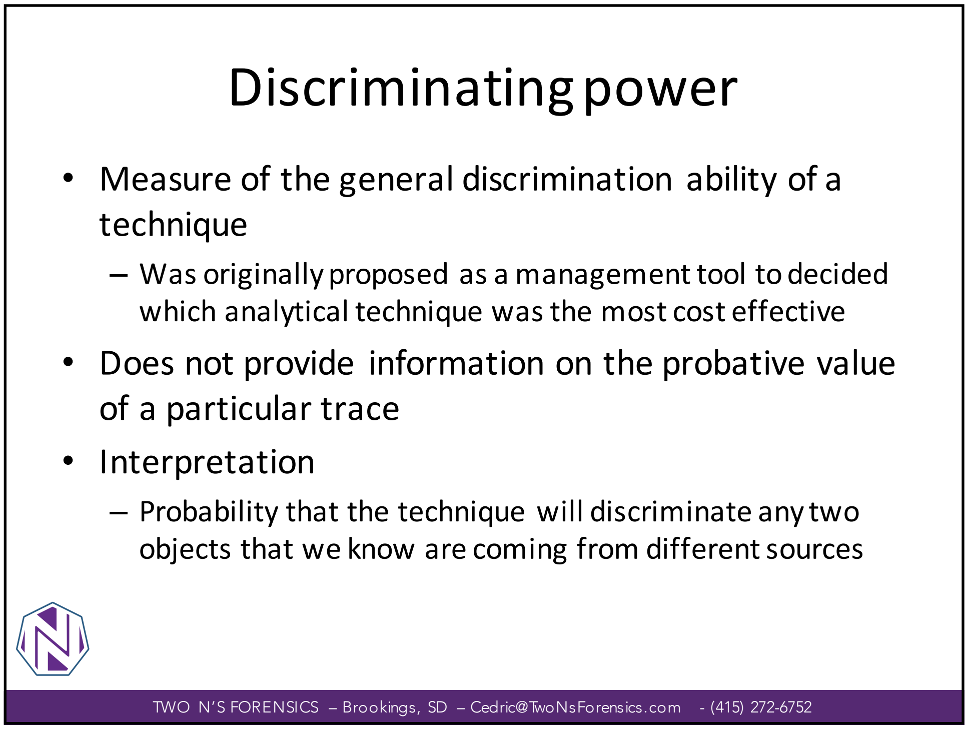
$$p \approx 1 - \left(\frac{364}{365}\right)^{\binom{50}{2}} = 0.97$$



Earth fallacy

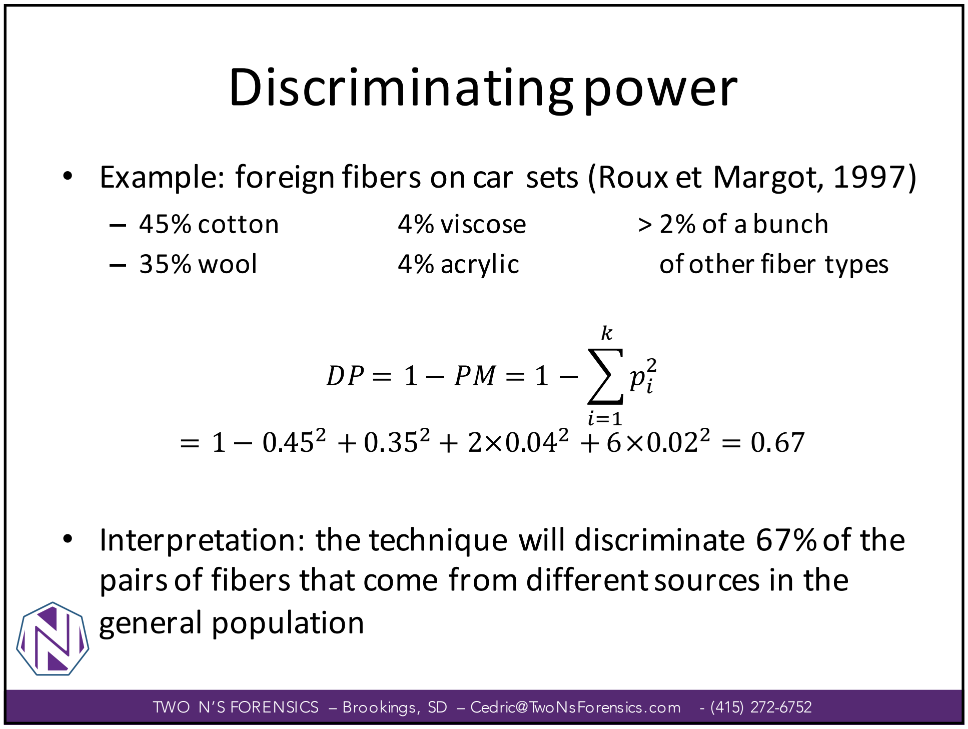
- Fingerprint problem
  - Probability of any given good quality configuration of 25 minutiae is $\frac{1}{7{,}000{,}000{,}000}$

City size	Probability to observe at least two individuals with the same arrangement
1,000	0.000071
10,000	0.007
100,000	0.51
1,000,000	1



Discriminating power

- Measure of the general discrimination ability of a technique
  - Was originally proposed as a management tool to decided which analytical technique was the most cost effective
- Does not provide information on the probative value of a particular trace
- Interpretation
  - Probability that the technique will discriminate any two objects that we know are coming from different sources



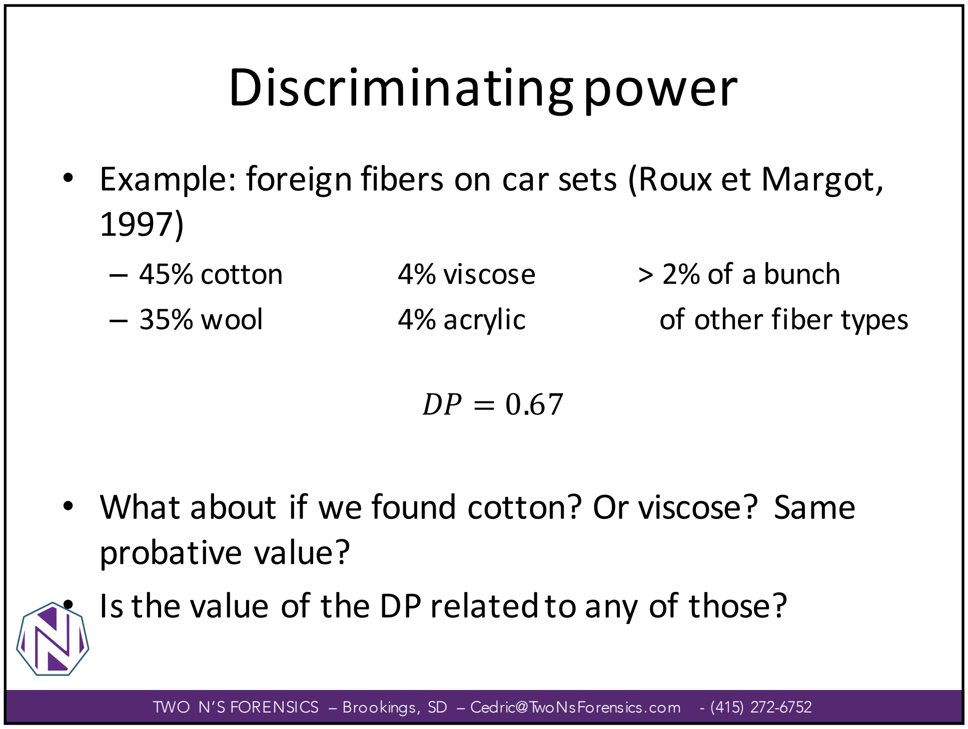
Discriminating power

- Example: foreign fibers on car sets (Roux et Margot, 1997)
  - 45% cotton 4% viscose > 2% of a bunch
  - 35% wool 4% acrylic of other fiber types

$$DP = 1 - PM = 1 - \sum_{i=1}^{k} p_i^2$$

$$= 1 - 0.45^2 + 0.35^2 + 2\times 0.04^2 + 6\times 0.02^2 = 0.67$$

- Interpretation: the technique will discriminate 67% of the pairs of fibers that come from different sources in the general population

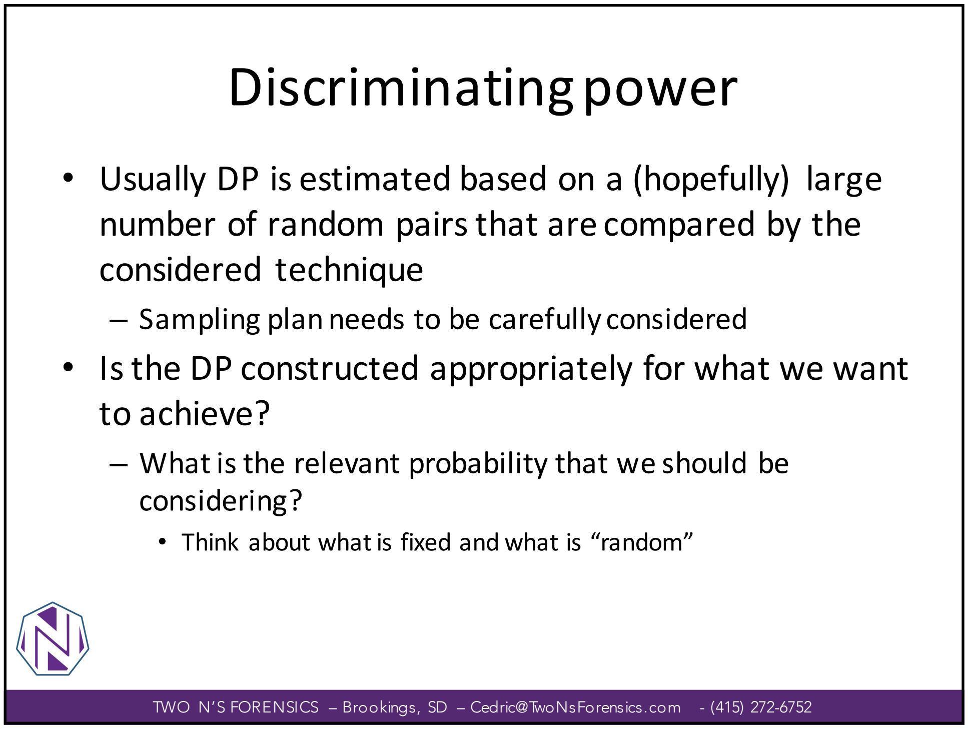


Discriminating power

- Example: foreign fibers on car sets (Roux et Margot, 1997)
  - 45% cotton 4% viscose > 2% of a bunch
  - 35% wool 4% acrylic of other fiber types

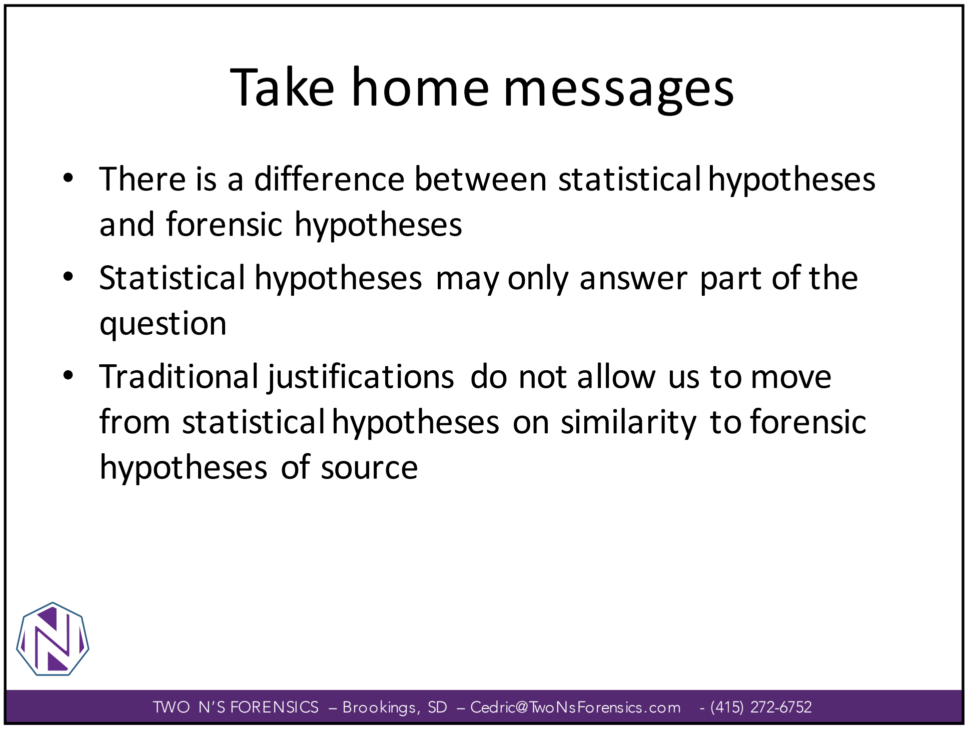
$$DP = 0.67$$

- What about if we found cotton? Or viscose? Same probative value?
- Is the value of the DP related to any of those?



Discriminating power

- Usually DP is estimated based on a (hopefully) large number of random pairs that are compared by the considered technique
  - Sampling plan needs to be carefully considered
- Is the DP constructed appropriately for what we want to achieve?
  - What is the relevant probability that we should be considering?
    - Think about what is fixed and what is “random”



Take home messages

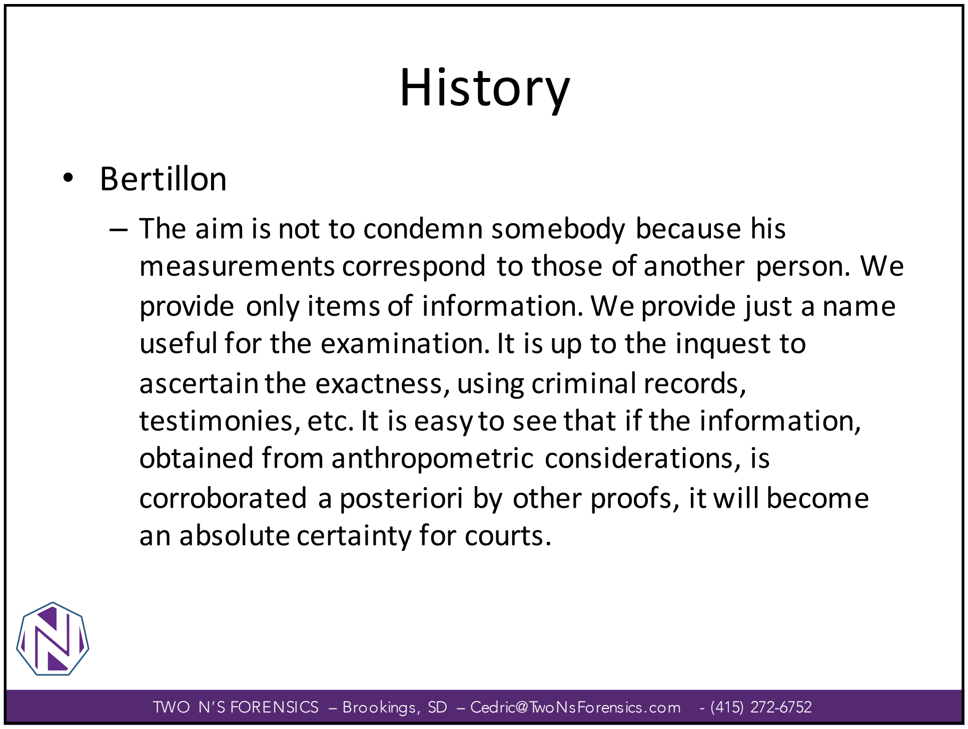
- There is a difference between statistical hypotheses and forensic hypotheses
- Statistical hypotheses may only answer part of the question
- Traditional justifications do not allow us to move from statistical hypotheses on similarity to forensic hypotheses of source

Chapter X

QUANTIFICATION OF THE PROBATIVE VALUE IN THE PAST

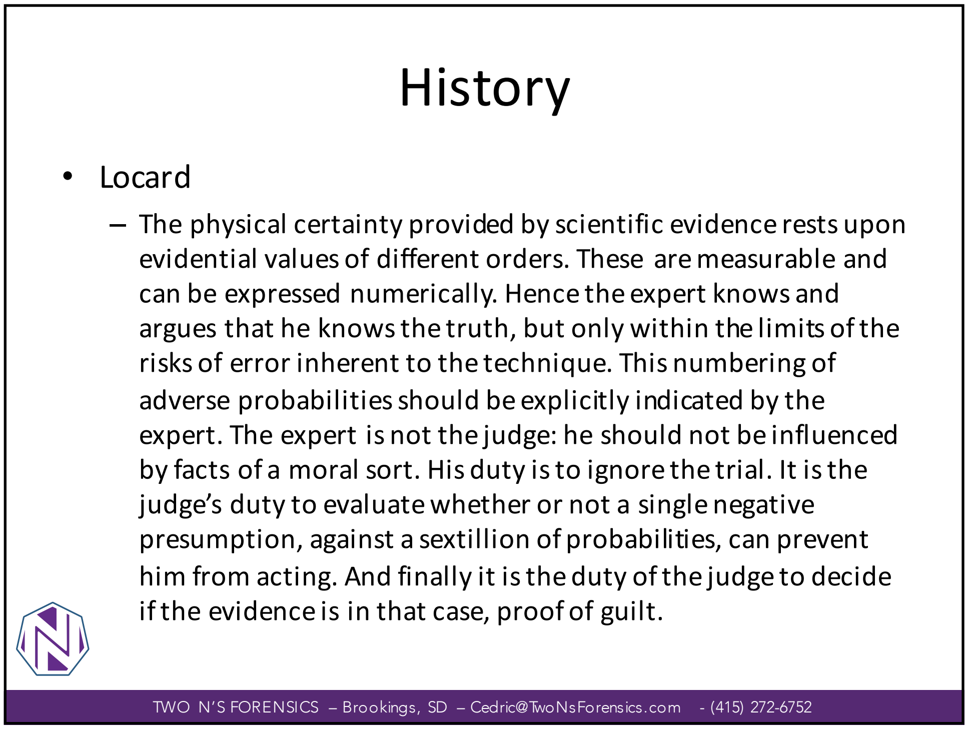
History

- Not a new concern
- Galton (1892)
  - Our problem is this: given two finger prints, which are alike in their minutiae, what is the chance that they were made by different persons?



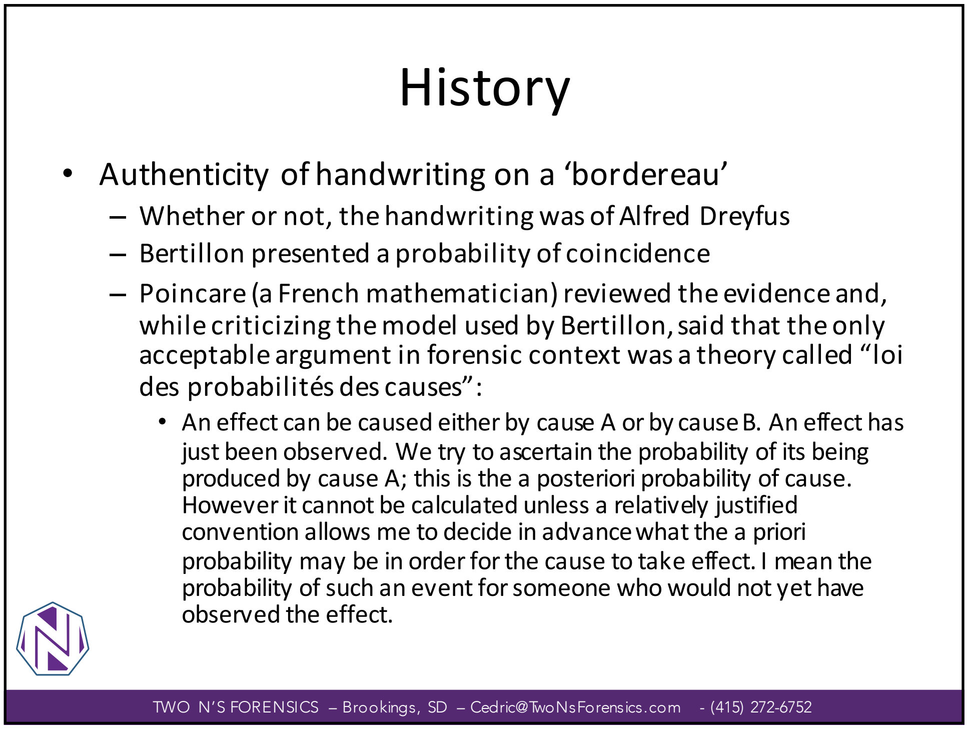
History

- Bertillon
  - The aim is not to condemn somebody because his measurements correspond to those of another person. We provide only items of information. We provide just a name useful for the examination. It is up to the inquest to ascertain the exactness, using criminal records, testimonies, etc. It is easy to see that if the information, obtained from anthropometric considerations, is corroborated a posteriori by other proofs, it will become an absolute certainty for courts.



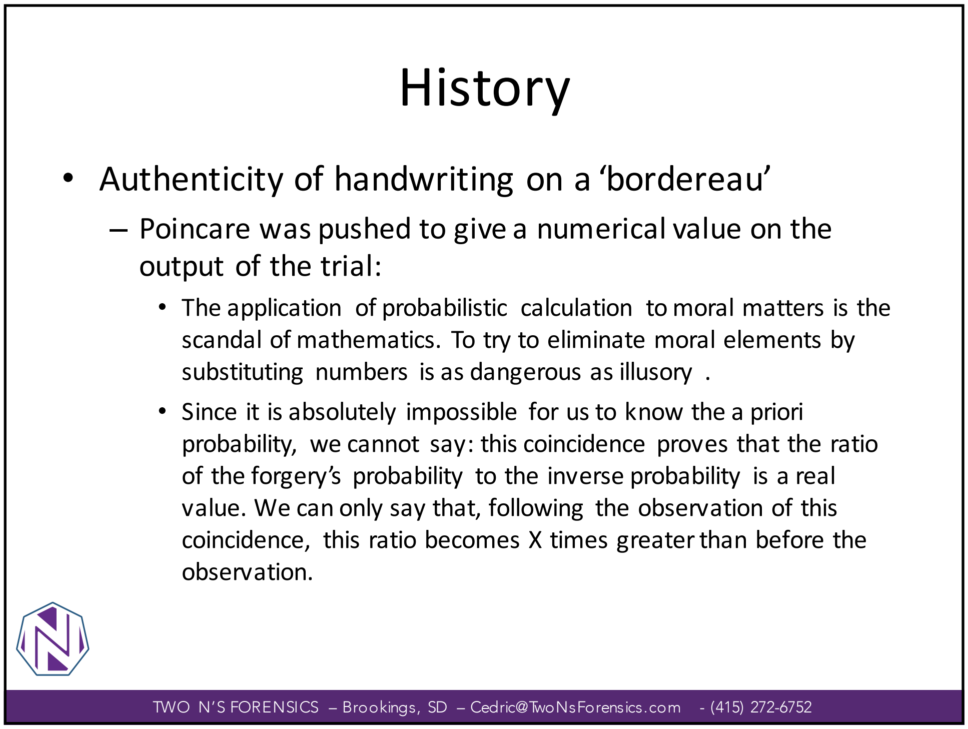
History

- Locard
  - The physical certainty provided by scientific evidence rests upon evidential values of different orders. These are measurable and can be expressed numerically. Hence the expert knows and argues that he knows the truth, but only within the limits of the risks of error inherent to the technique. This numbering of adverse probabilities should be explicitly indicated by the expert. The expert is not the judge: he should not be influenced by facts of a moral sort. His duty is to ignore the trial. It is the judge's duty to evaluate whether or not a single negative presumption, against a sextillion of probabilities, can prevent him from acting. And finally it is the duty of the judge to decide if the evidence is in that case, proof of guilt.



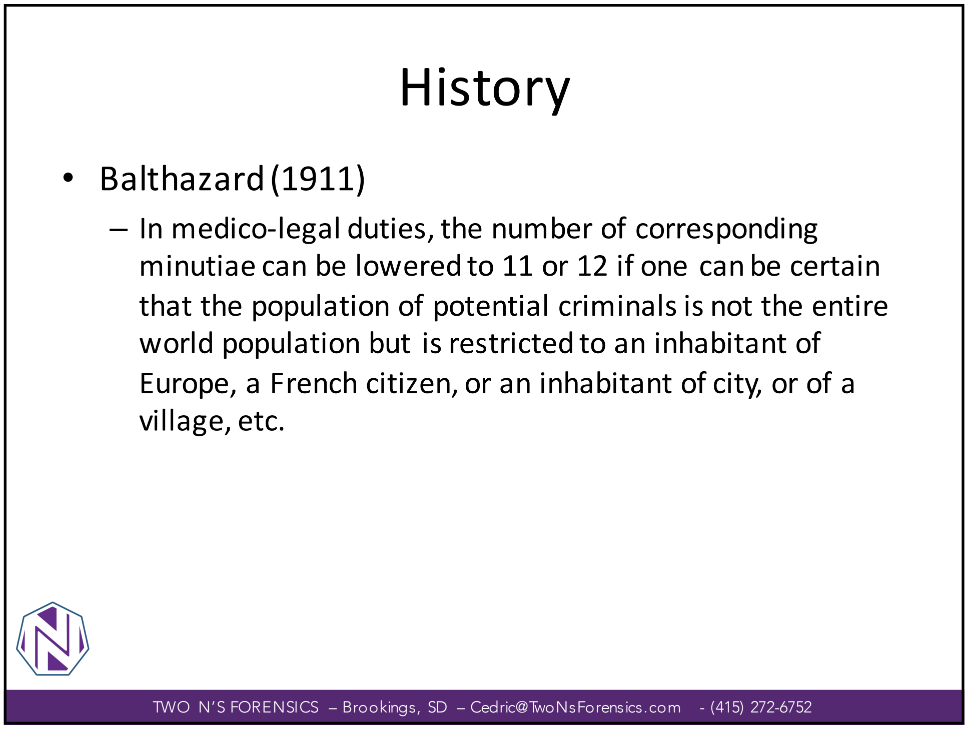
History

- Authenticity of handwriting on a ‘bordereau’
  - Whether or not, the handwriting was of Alfred Dreyfus
  - Bertillon presented a probability of coincidence
  - Poincare (a French mathematician) reviewed the evidence and, while criticizing the model used by Bertillon, said that the only acceptable argument in forensic context was a theory called “loi des probabilités des causes”:
    - An effect can be caused either by cause A or by cause B. An effect has just been observed. We try to ascertain the probability of its being produced by cause A; this is the a posteriori probability of cause. However it cannot be calculated unless a relatively justified convention allows me to decide in advance what the a priori probability may be in order for the cause to take effect. I mean the probability of such an event for someone who would not yet have observed the effect.



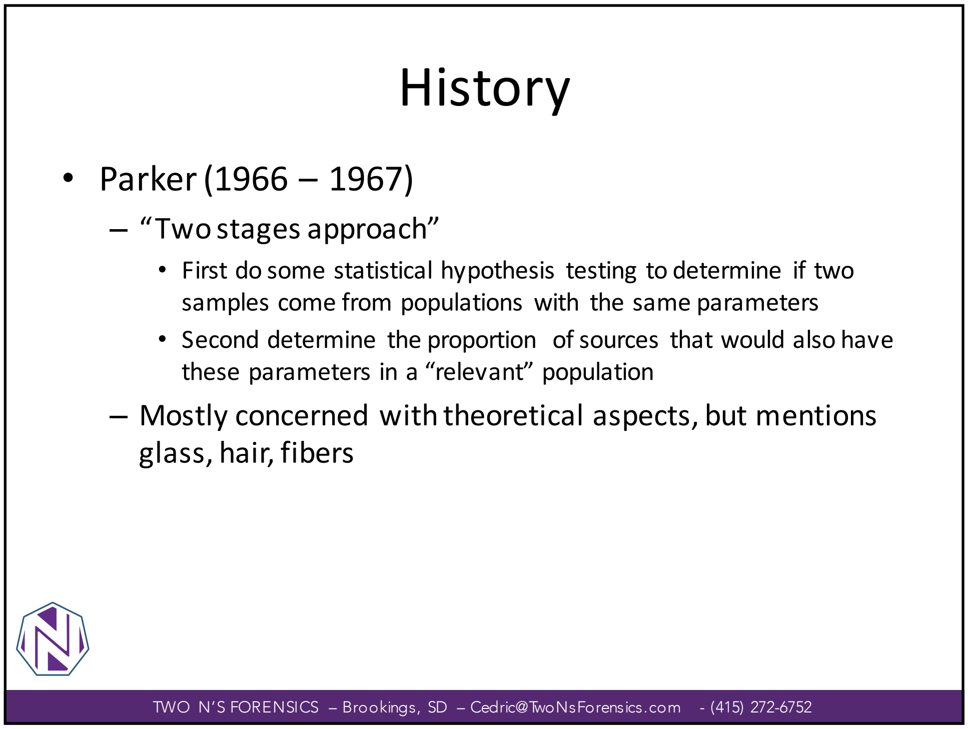
History

- Authenticity of handwriting on a ‘bordereau’
  - Poincare was pushed to give a numerical value on the output of the trial:
    - The application of probabilistic calculation to moral matters is the scandal of mathematics. To try to eliminate moral elements by substituting numbers is as dangerous as illusory .
    - Since it is absolutely impossible for us to know the a priori probability, we cannot say: this coincidence proves that the ratio of the forgery’s probability to the inverse probability is a real value. We can only say that, following the observation of this coincidence, this ratio becomes X times greater than before the observation.



History

- Balthazard (1911)
  - In medico-legal duties, the number of corresponding minutiae can be lowered to 11 or 12 if one can be certain that the population of potential criminals is not the entire world population but is restricted to an inhabitant of Europe, a French citizen, or an inhabitant of city, or of a village, etc.



History

- Parker (1966 – 1967)
  - "Two stages approach"
    - First do some statistical hypothesis testing to determine if two samples come from populations with the same parameters
    - Second determine the proportion of sources that would also have these parameters in a "relevant" population
  - Mostly concerned with theoretical aspects, but mentions glass, hair, fibers

History

- Finkelstein et Fairley (1970)
  - First description of the use of Bayes theorem in the legal literature
  - Enables to combine various pieces of information and update probability of guilt

History

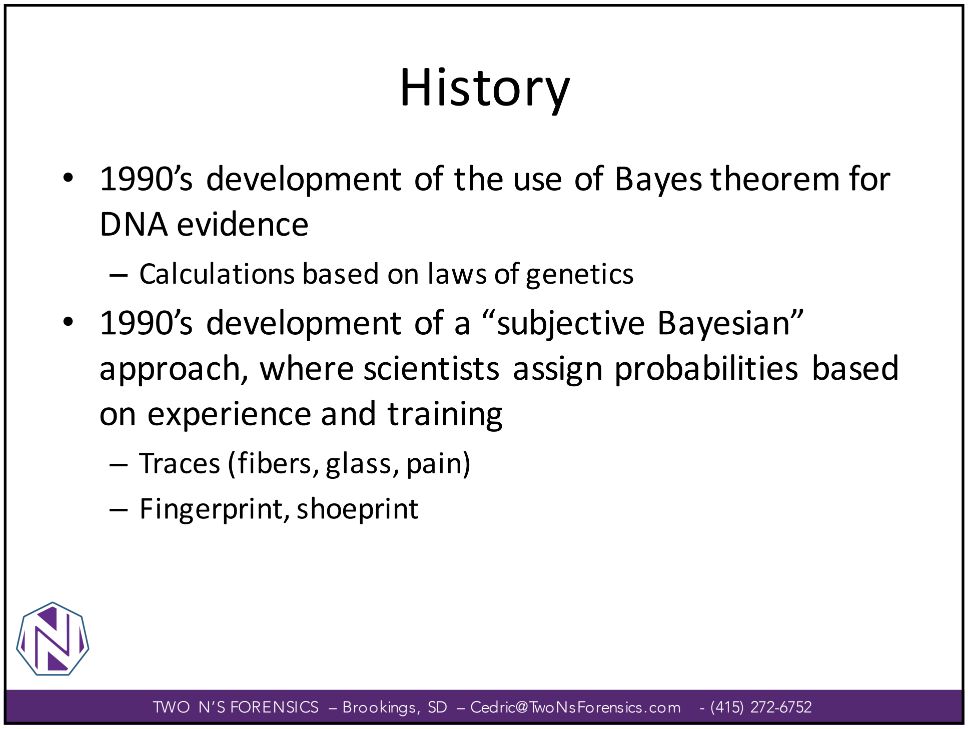
- Evett et al. (1977 – 1986)
  - Applies method of Parker to glass evidence
  - Slowly moves towards a Bayesian approach
  - Crude at the beginning -> more sophisticated
  - Starts applying the technique to
    - Fiber evidence
    - Bloodstain typing
  - Creates the bases for the quantification of the probative value of DNA evidence

History

- Smalldon et Moffat (1973)
  - Introduce the discriminating power (at least in forensic science)
  - As a management tool to optimize cost-efficiency
- Kwan 1977
  - PhD thesis on the Inference of Identity of Source
  - Review/introduce many different concepts
    - Qualitative identity
    - Quantitative identity
    - Definition of source
    - Feature selection
    - Hypothetical-deductive method (and associated “metrics”)

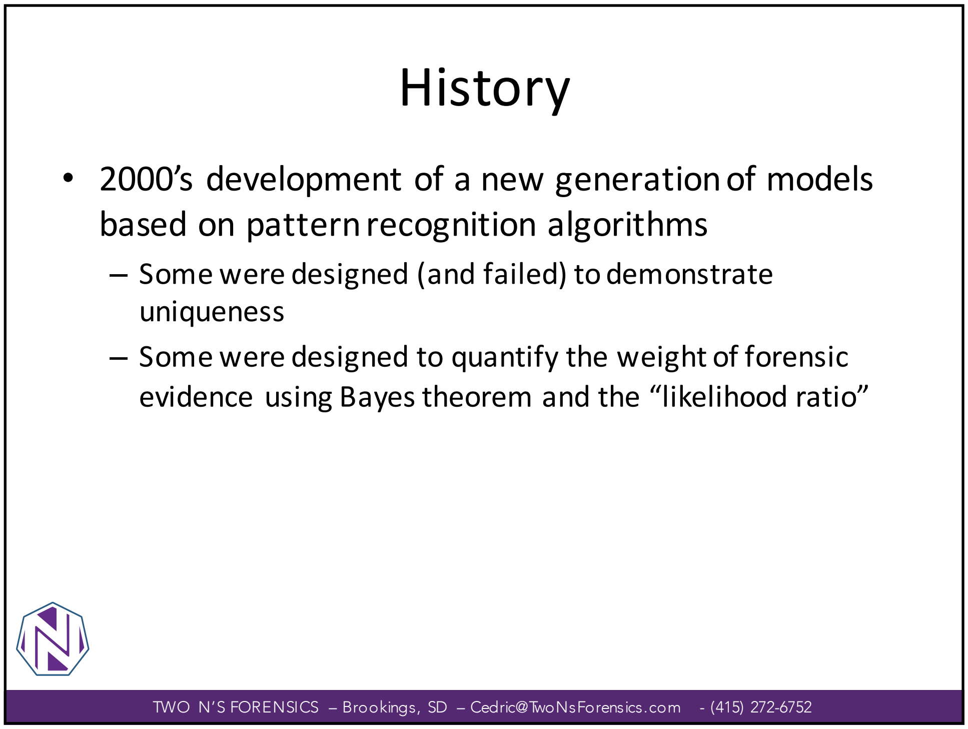
History

- Lindley (1977)
  - Propose the first fully “Bayesian” approach for the quantification of the weight of forensic evidence
  - Applied to glass evidence



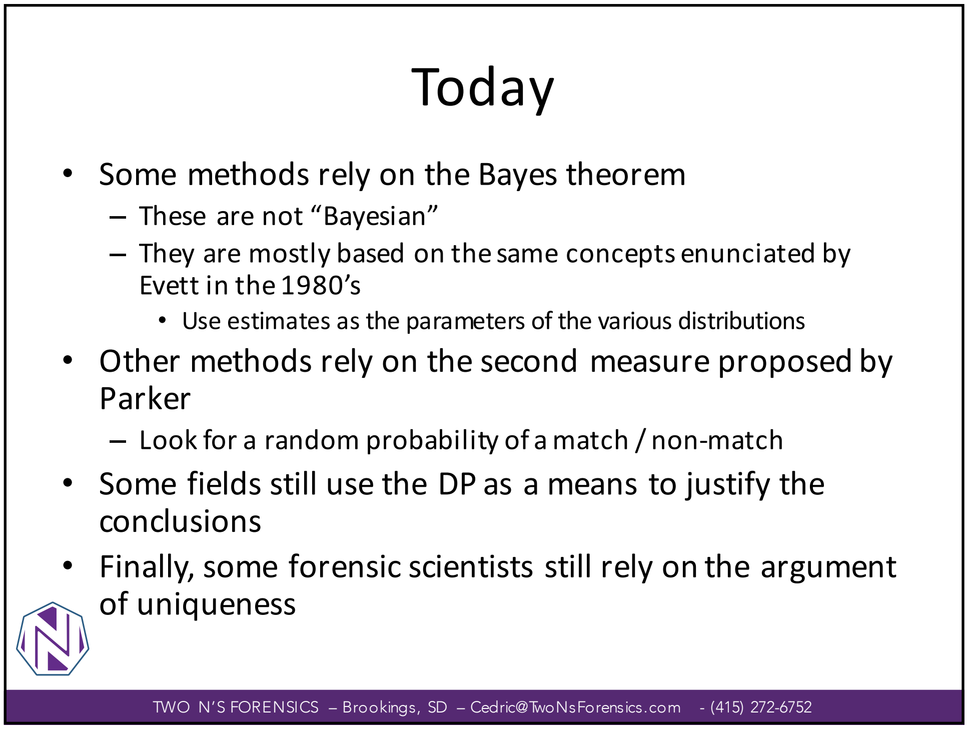
History

- 1990's development of the use of Bayes theorem for DNA evidence
  - Calculations based on laws of genetics
- 1990's development of a "subjective Bayesian" approach, where scientists assign probabilities based on experience and training
  - Traces (fibers, glass, pain)
  - Fingerprint, shoeprint



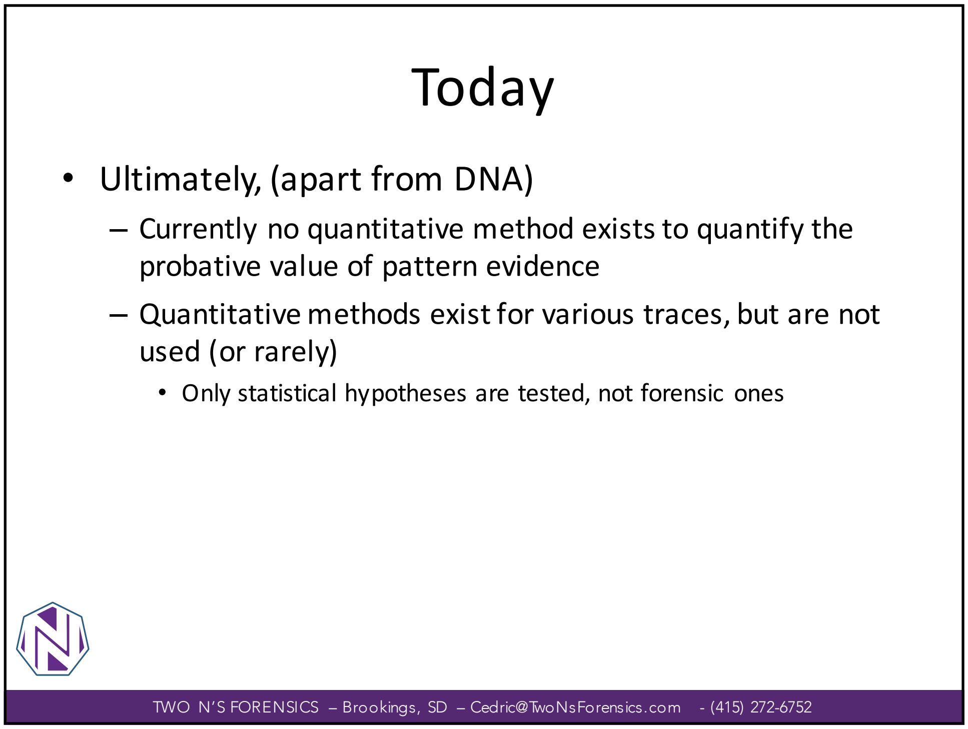
History

- 2000's development of a new generation of models based on pattern recognition algorithms
  - Some were designed (and failed) to demonstrate uniqueness
  - Some were designed to quantify the weight of forensic evidence using Bayes theorem and the "likelihood ratio"



Today

- Some methods rely on the Bayes theorem
  - These are not "Bayesian"
  - They are mostly based on the same concepts enunciated by Evett in the 1980's
    - Use estimates as the parameters of the various distributions
- Other methods rely on the second measure proposed by Parker
  - Look for a random probability of a match / non-match
- Some fields still use the DP as a means to justify the conclusions
- Finally, some forensic scientists still rely on the argument of uniqueness



Today

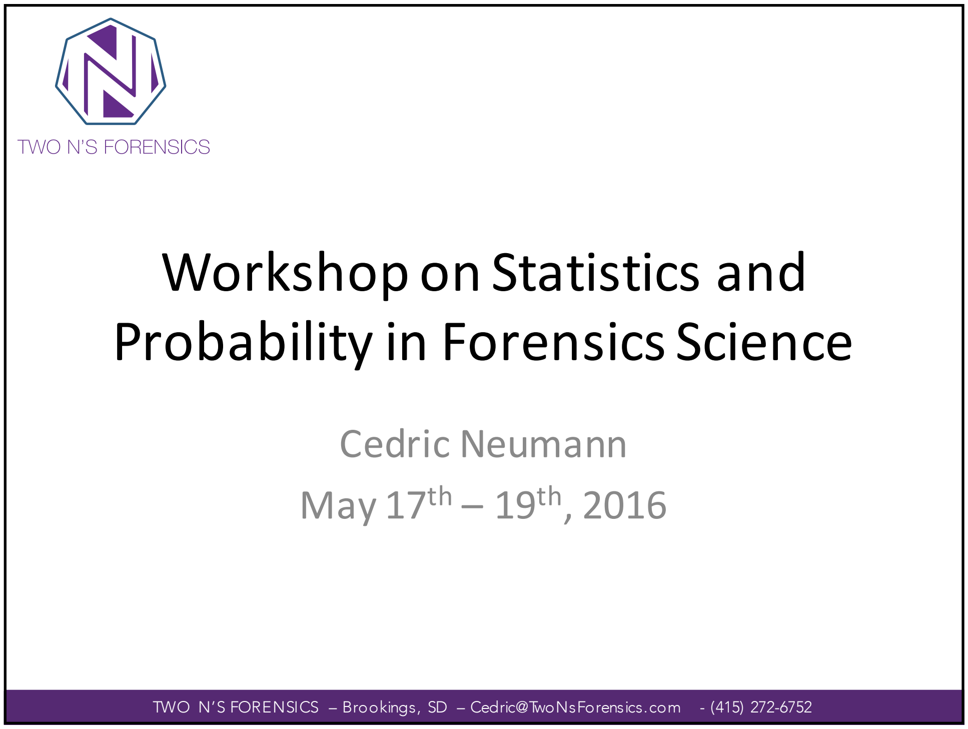
- Ultimately, (apart from DNA)
  - Currently no quantitative method exists to quantify the probative value of pattern evidence
  - Quantitative methods exist for various traces, but are not used (or rarely)
    - Only statistical hypotheses are tested, not forensic ones

Workshop on Statistics and Probability in Forensics Science

Cedric Neumann

DAY 3

MAY 19TH

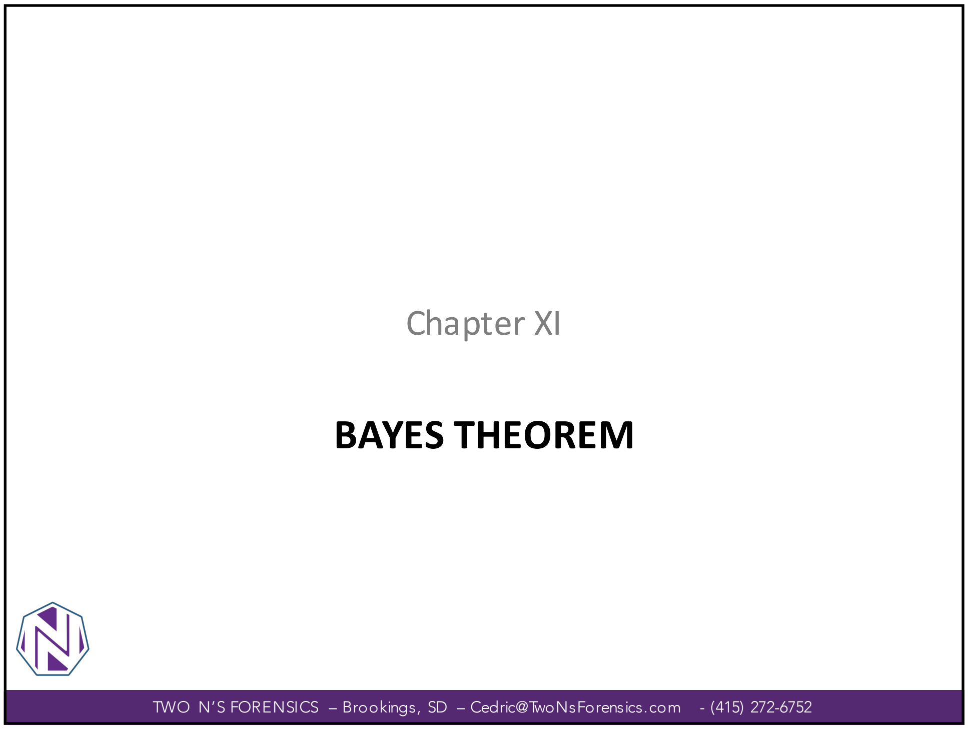


TWO N'S FORENSICS

Workshop on Statistics and Probability in Forensics Science

Cedric Neumann

May 17th – 19th, 2016



Chapter XI

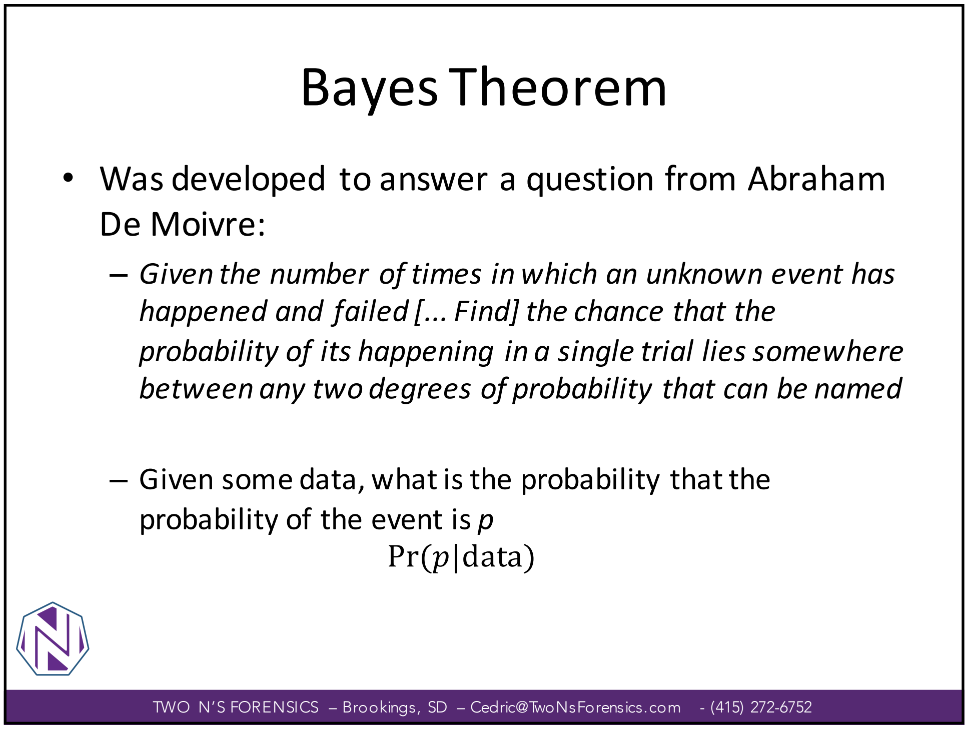
BAYES THEOREM

Rev. Thomas Bayes

- c 1701 – 1761
- Wrote 2 books (1731 and 1736)
- Was a Minister until 1752
- Became interested in probabilities in 1755
- “Bayes Theorem” was read posthumously at the Royal Society in 1763.

Rev. Thomas Bayes

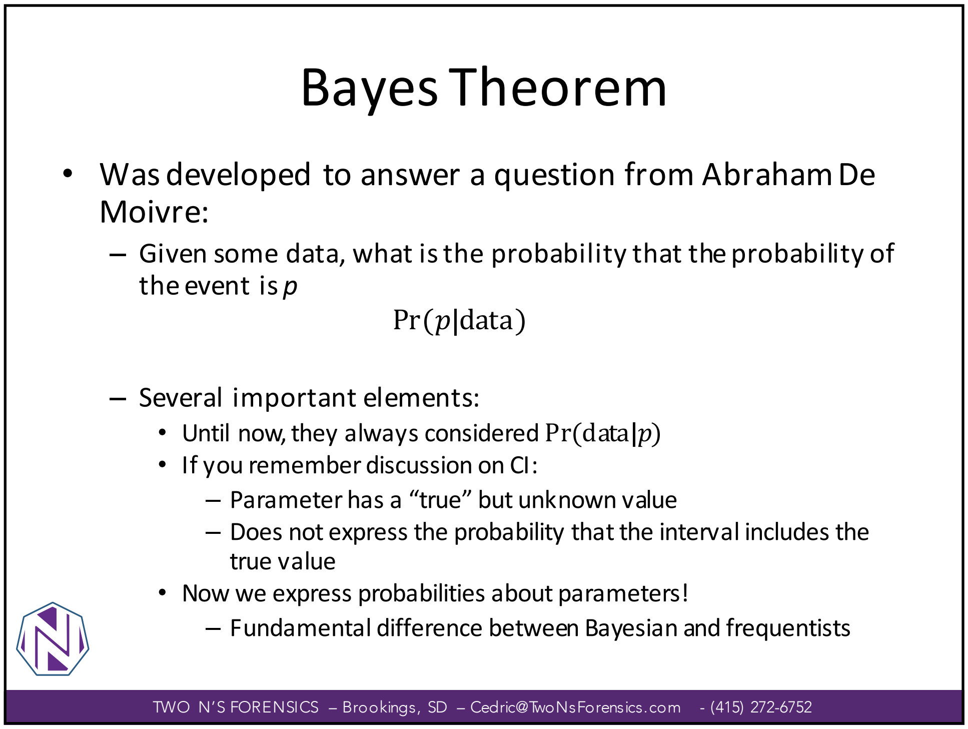
Rev. Thomas Bayes Tomb in London – 500m from Royal Statistical Society
Cedric’s pilgrimage in 2011



Bayes Theorem

- Was developed to answer a question from Abraham De Moivre:
  - *Given the number of times in which an unknown event has happened and failed [... Find] the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named*
  - Given some data, what is the probability that the probability of the event is *p*

$$\Pr(p|\text{data})$$



Bayes Theorem

- Was developed to answer a question from Abraham De Moivre:
  - Given some data, what is the probability that the probability of the event is *p*

$$\Pr(p|\text{data})$$

  - Several important elements:
    - Until now, they always considered $\Pr(\text{data}|p)$
    - If you remember discussion on CI:
      - Parameter has a "true" but unknown value
      - Does not express the probability that the interval includes the true value
    - Now we express probabilities about parameters!
      - Fundamental difference between Bayesian and frequentists

Bayes Theorem

- Was de to answer a question from Abraham De M
  he probability t f
  data)
  - Seve elements:
    - Until now, th always considered Pr(da
    - If you remember discussion on CI:
      - Parameter has a “true” but unknown valu
      - Does not express the probability that the i erval includes the true value
    - Now we express probabilities about parameters!
      - Fundamental difference between Bayesian and frequentists

In the frequentist world, the parameter is fixed and the interval varies

In the Bayesian world, the interval is fixed and the parameter varies

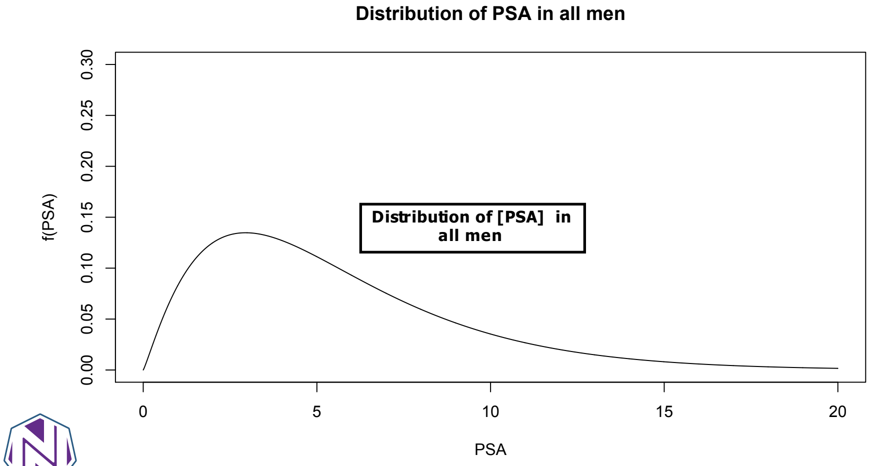
Bayes Theorem

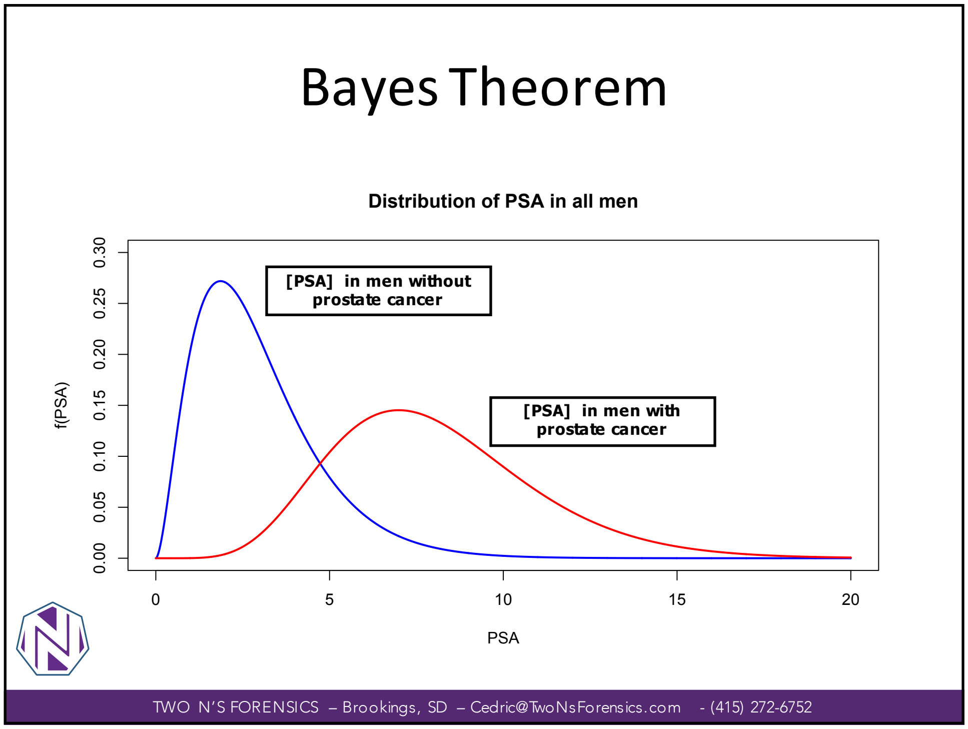
- Theorem was presented in 1763 by Richard Price
  - They were attempting to demonstrate existence of God from a series of observations
    - They were concerned with
      - Pr(God exists | order in nature)
    - But they could only observe
      - Pr(order in nature | God exists) vs.
      - Pr(order in nature | chance)
    - Bayes theorem is a way to connect these two probabilities

Bayes Theorem

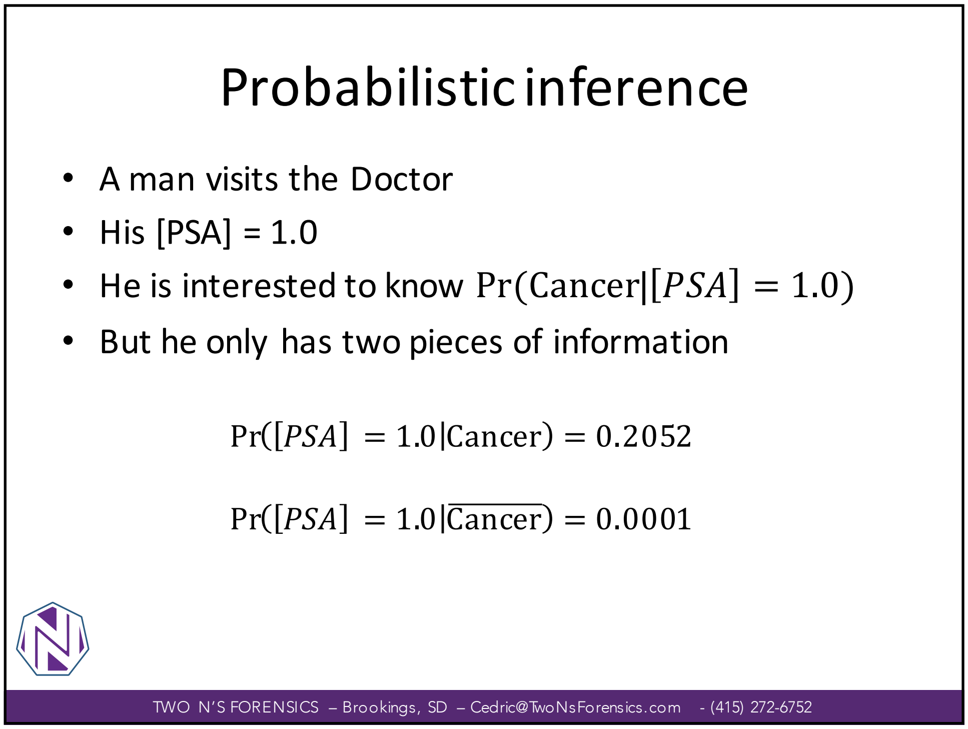
- Testing for prostate cancer in men
- Concentration of Prostate Specific Antigen
  - [PSA]
  - 0 to 5.99 ng/mL is low
  - 6 to 19.9 ng/mL is moderately elevated
  - 20 ng/mL or more is significantly elevated

Bayes Theorem





Bayes Theorem

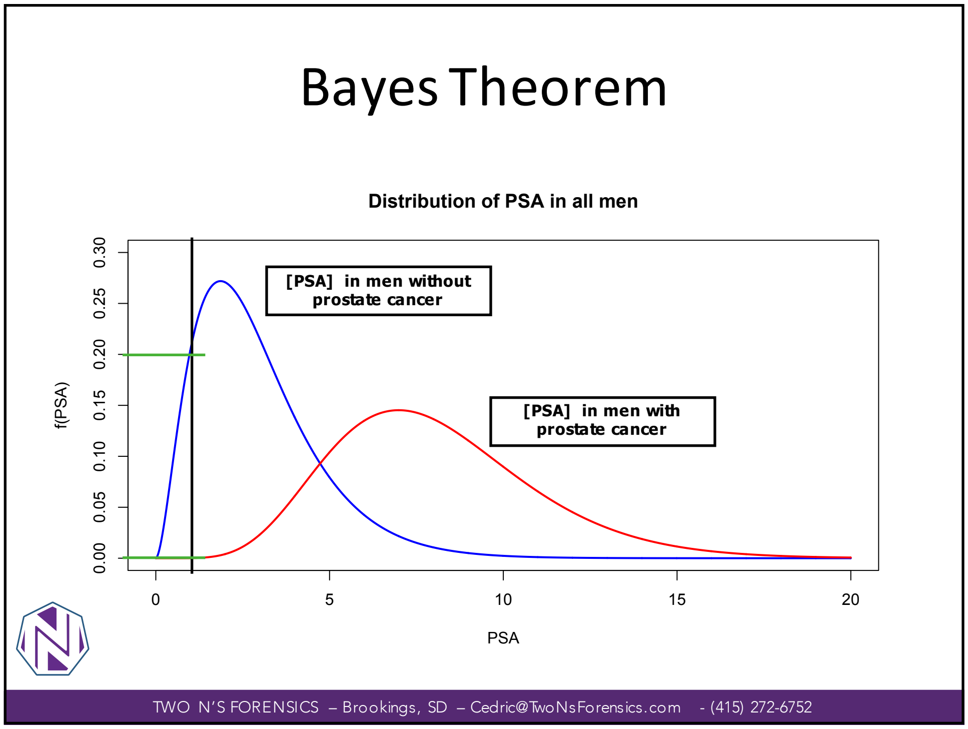


Probabilistic inference

- A man visits the Doctor
- His [PSA] = 1.0
- He is interested to know $\Pr(\text{Cancer}|[PSA] = 1.0)$
- But he only has two pieces of information

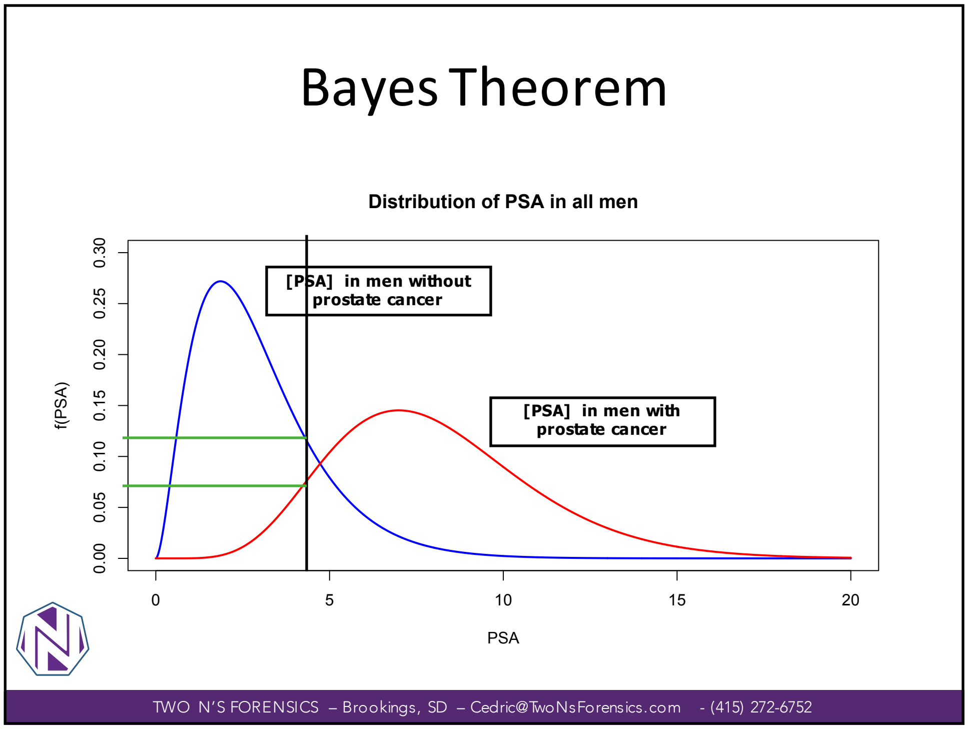
$$\Pr([PSA] = 1.0|\text{Cancer}) = 0.2052$$

$$\Pr([PSA] = 1.0|\overline{\text{Cancer}}) = 0.0001$$



Bayes Theorem

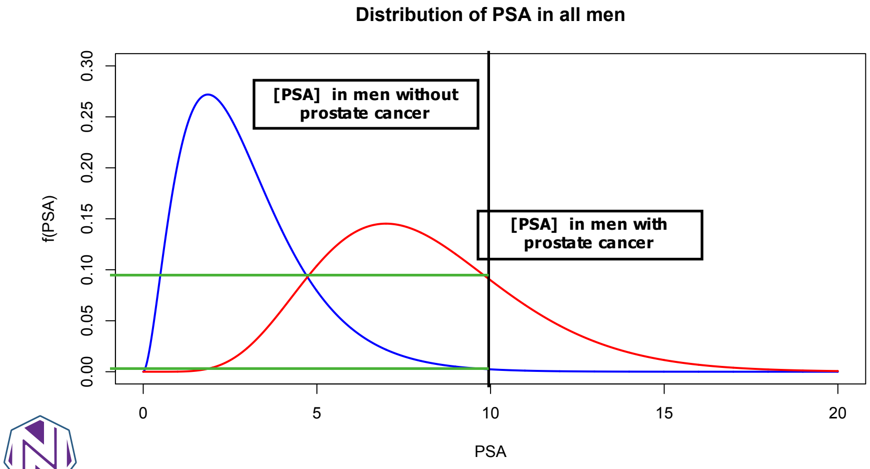
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Bayes Theorem

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Bayes Theorem



Bayes Theorem

- A man visits the Doctor
- His [PSA] = 1.0
- He is interested to know $\Pr(\text{Cancer}|[PSA] = 1.0)$

- Likelihood ratio

$$\text{LR} = \frac{\Pr([PSA] = 1.0|\text{Cancer})}{\Pr([PSA] = 1.0|\overline{\text{Cancer}})} = \frac{0.2052}{0.0001} = 2052$$

- It is 2052 times more likely to have $[PSA]$=1 when you don't have cancer than when you do

Bayes Theorem

- A man visits the Doctor
- His [PSA] = 1.0
- He is interested to know $\Pr(\text{Cancer}|[PSA] = 1.0)$
- Li

$$\frac{|\text{Cancer})}{\text{Cancer})} = \frac{0.2052}{0.0001} = 2052$$

Still not answering the question

- It is 2052 times more likely to have $[PSA]=1$ when you don't have cancer than when you do

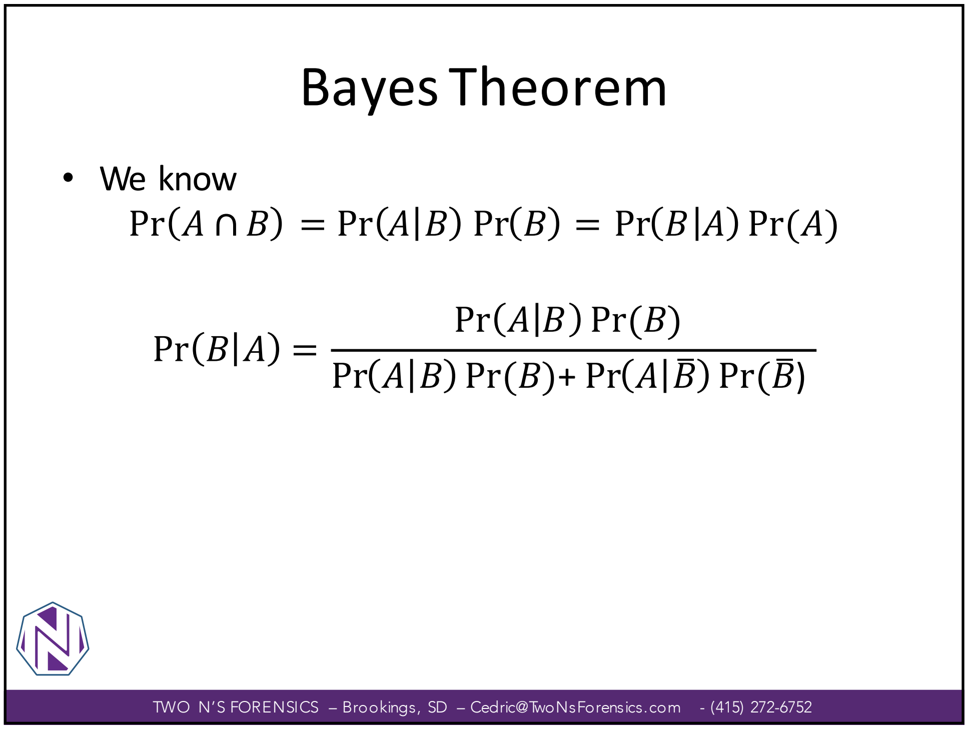
Bayes Theorem

- We know

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$

But if B and $\bar{B}$ are disjoint and exhaustive partitions of sample space, $\Pr(A) = \Pr(A|B)\Pr(B)+\Pr(A|\bar{B})\Pr(\bar{B})$

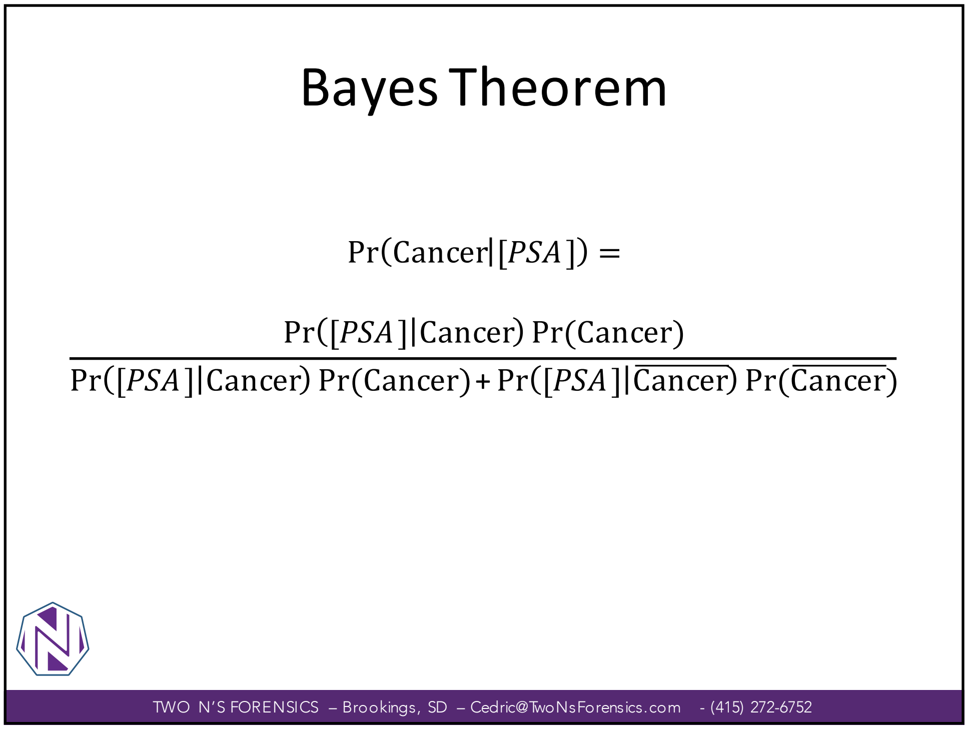


Bayes Theorem

- We know

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

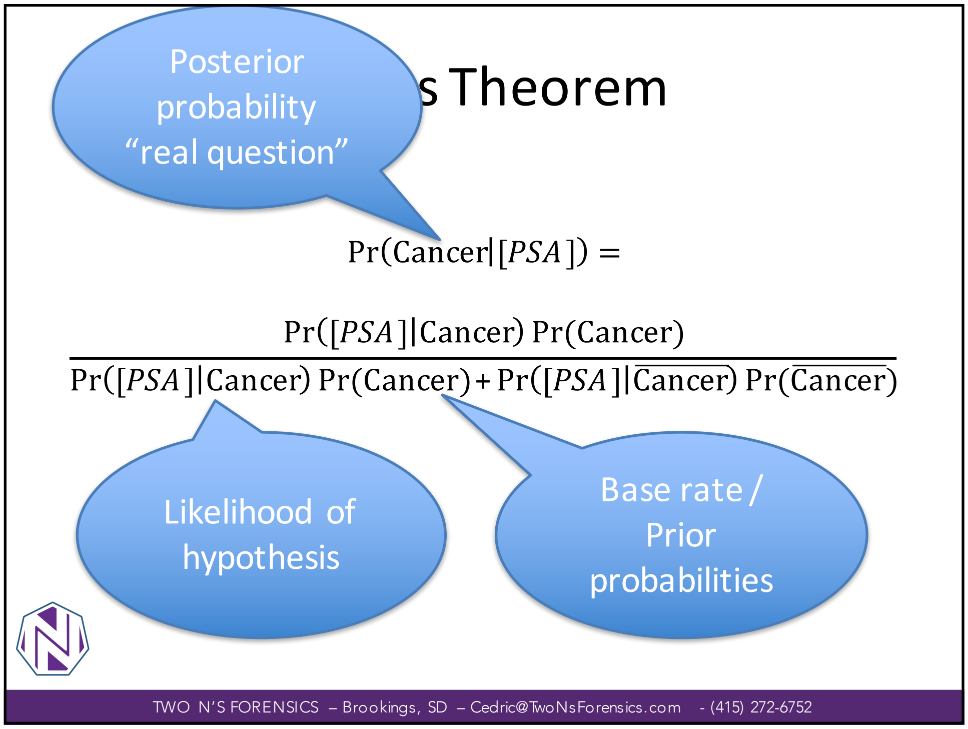
$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A|B)\Pr(B) + \Pr(A|\bar{B})\Pr(\bar{B})}$$



Bayes Theorem

$$\Pr(\text{Cancer}|[PSA]) =$$

$$\frac{\Pr([PSA]|\text{Cancer})\Pr(\text{Cancer})}{\Pr([PSA]|\text{Cancer})\Pr(\text{Cancer}) + \Pr([PSA]|\overline{\text{Cancer}})\Pr(\overline{\text{Cancer}})}$$



s Theorem

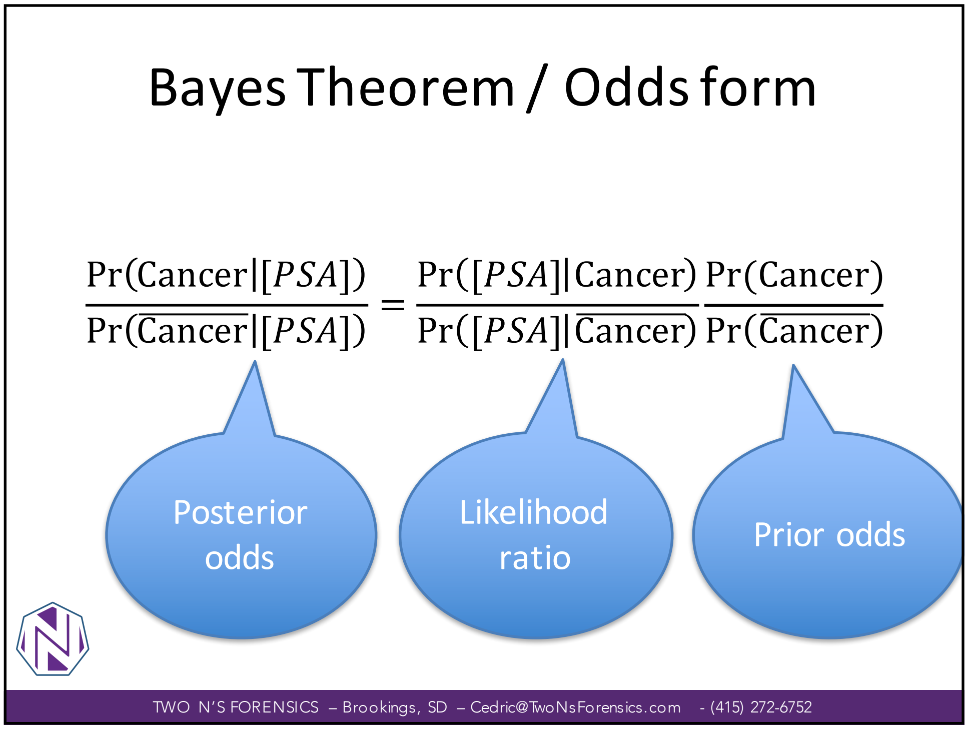
Posterior probability “real question”

$$\Pr(\text{Cancer}|[PSA]) =$$

$$\frac{\Pr([PSA]|\text{Cancer})\Pr(\text{Cancer})}{\Pr([PSA]|\text{Cancer})\Pr(\text{Cancer})+\Pr([PSA]|\overline{\text{Cancer}})\Pr(\overline{\text{Cancer}})}$$

Likelihood of hypothesis

Base rate / Prior probabilities



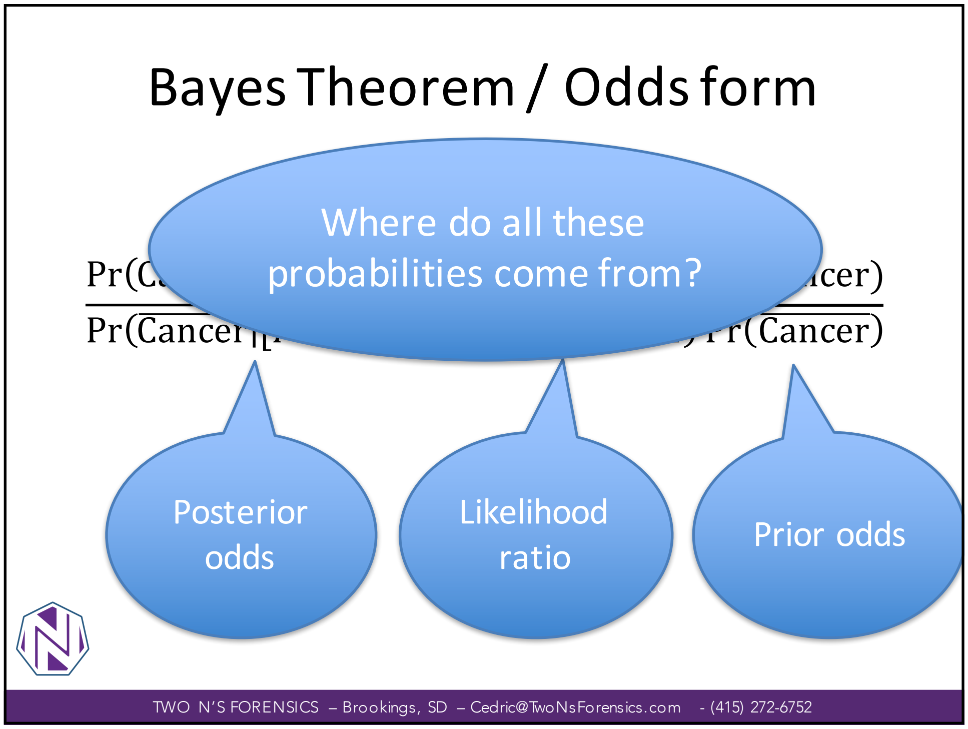
Bayes Theorem / Odds form

$$\frac{\Pr(\text{Cancer}|[PSA])}{\Pr(\overline{\text{Cancer}}|[PSA])}=\frac{\Pr([PSA]|\text{Cancer})}{\Pr([PSA]|\overline{\text{Cancer}})}\frac{\Pr(\text{Cancer})}{\Pr(\overline{\text{Cancer}})}$$

Posterior odds

Likelihood ratio

Prior odds



Bayes Theorem / Odds form

Where do all these probabilities come from?

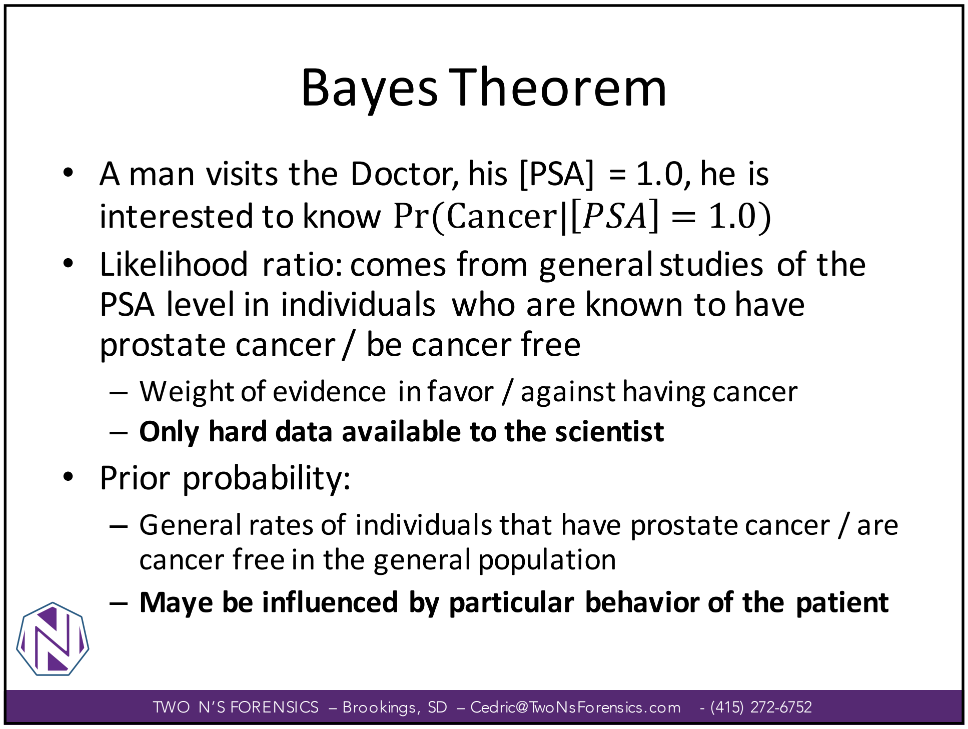
Pr(C… …cer)

Pr(Cancer|[… …)Pr(Cancer)

Posterior odds

Likelihood ratio

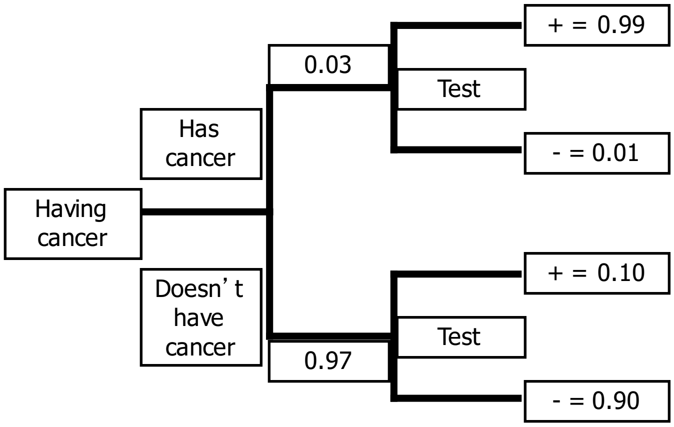
Prior odds



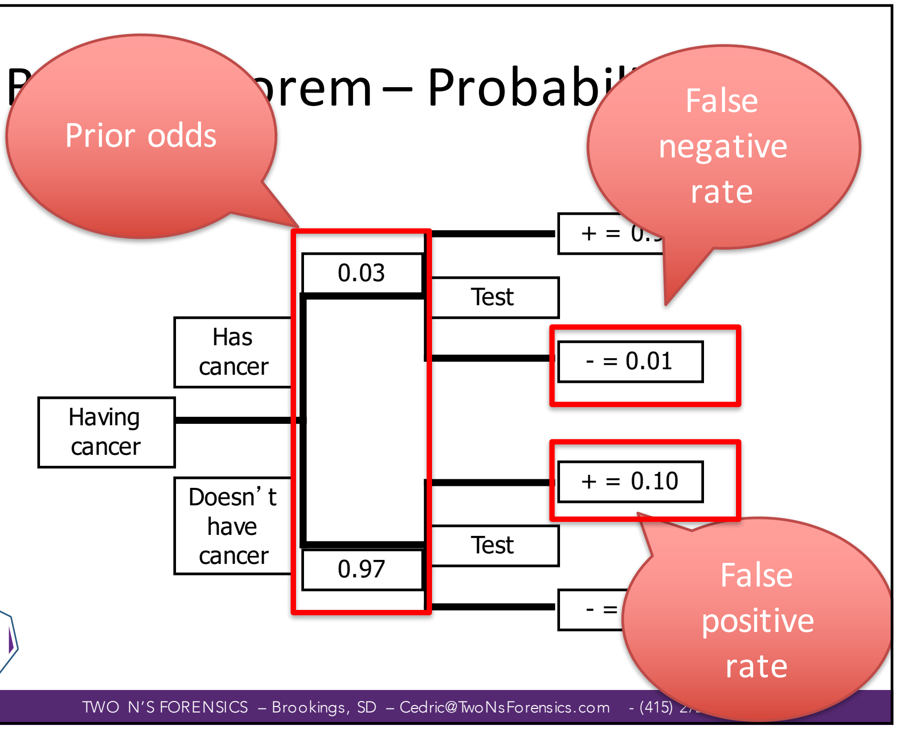
Bayes Theorem

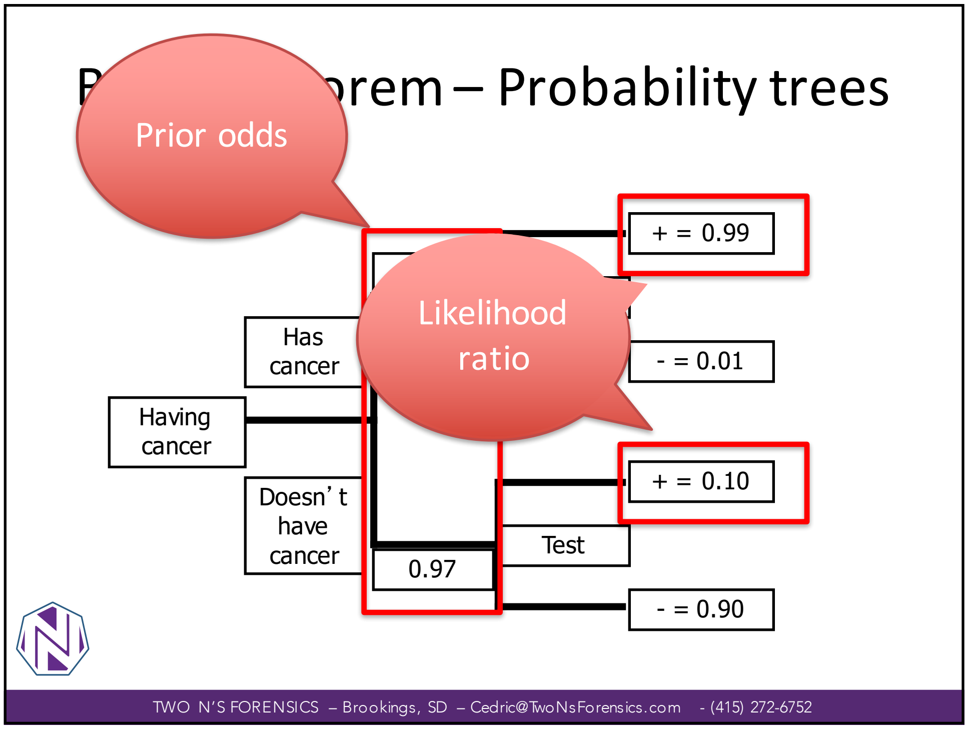
- A man visits the Doctor, his [PSA] = 1.0, he is interested to know $\Pr(\text{Cancer}|[PSA] = 1.0)$
- Likelihood ratio: comes from general studies of the PSA level in individuals who are known to have prostate cancer / be cancer free
  - Weight of evidence in favor / against having cancer
  - **Only hard data available to the scientist**
- Prior probability:
  - General rates of individuals that have prostate cancer / are cancer free in the general population
  - **Maye be influenced by particular behavior of the patient**

Bayes Theorem – Probability trees



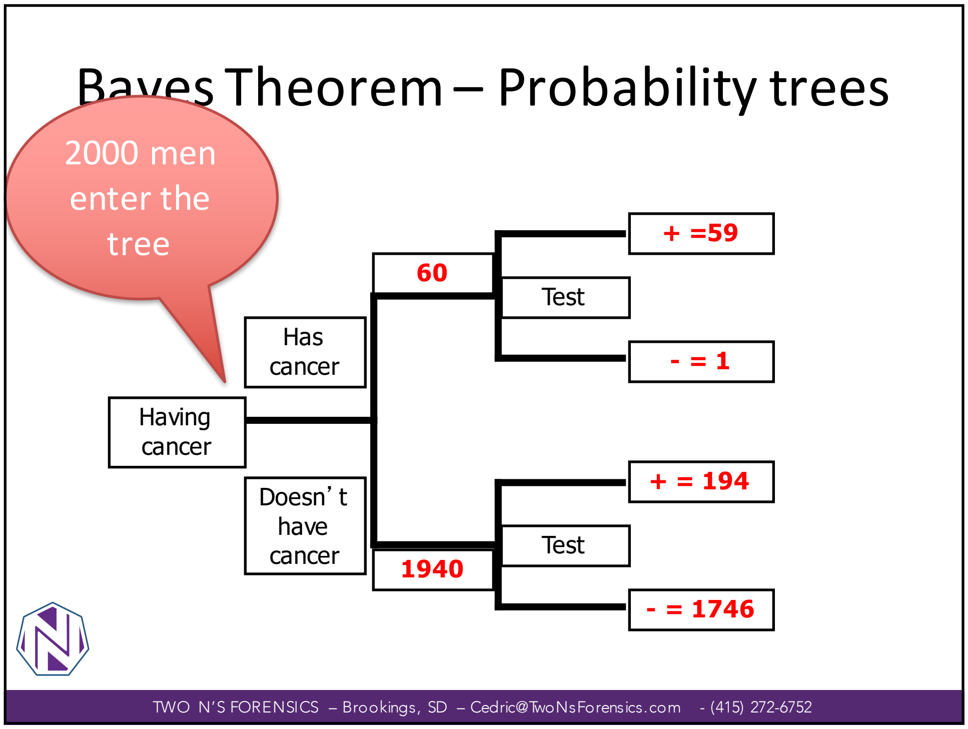
P orem – Probabil





B orem – Probability trees

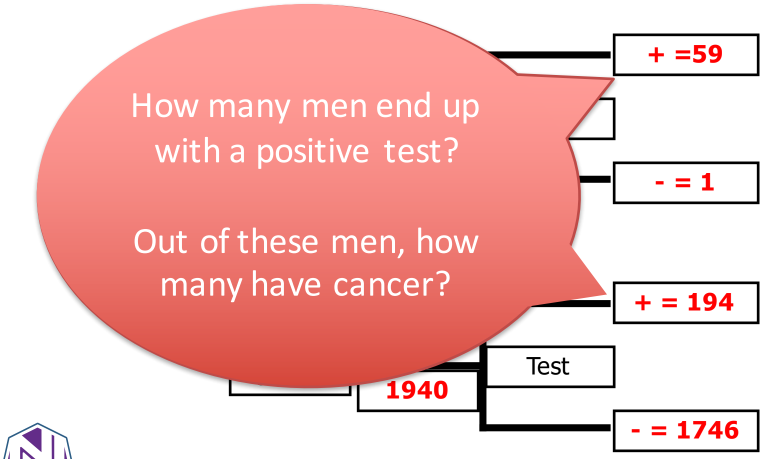
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Bayes Theorem – Probability trees

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752

Bayes Theorem – Probability trees



How many men end up with a positive test?

Out of these men, how many have cancer?

Bayes Theorem

- A woman takes a pregnancy test, she is interested to know $\Pr(\text{Pregnant}|T = +)$
- From validation studies (**hard data**):

$$\Pr(T = +|\text{Pregnant}) = 0.99$$
$$\Pr(T = +|\overline{\text{Pregnant}}) = 0.01$$

- Prior probability:
  - Is it rate of pregnant women in the population?
  - **Or is it sexual behavior of that particular woman?**

Bayes Theorem

- Prior probability:
  - PSA:
    - Cancer rate in the population
    - Risk factor of that particular man
  - Pregnancy:
    - Sexual behavior of that particular woman
- What happens if we cannot assign a probability to these events?

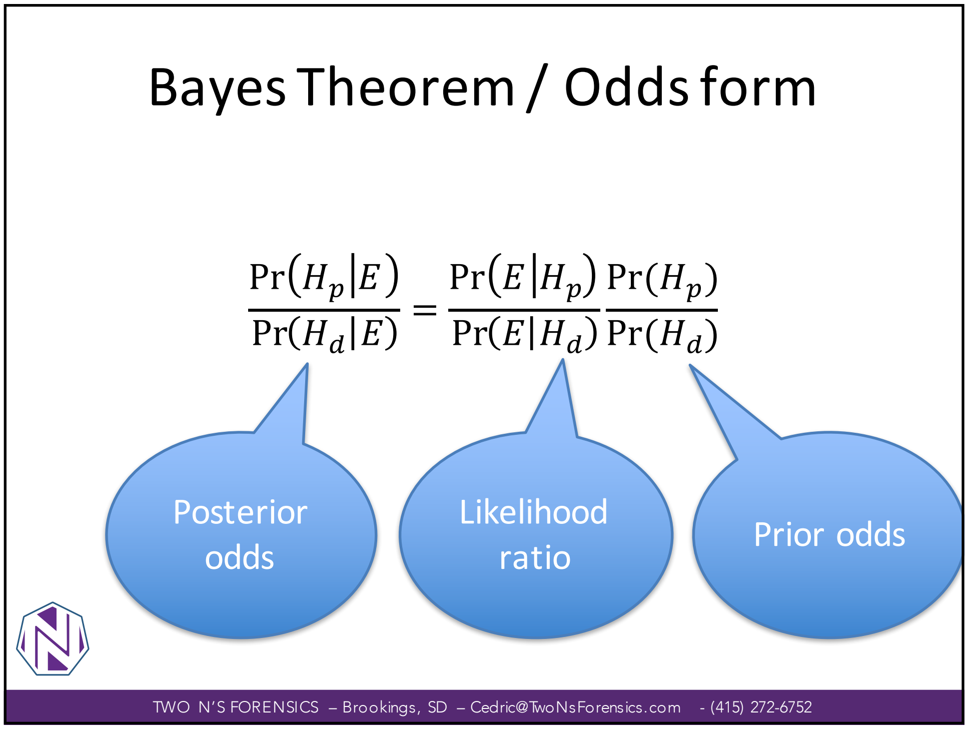
Bayes Theorem

- Prior probability:
  - Sexual be... of that particular wom...
- What happens if we cannot assign a probability to these events?

We cannot calculate posterior probabilities/odds

Then we cannot answer the question…

So what do we do?



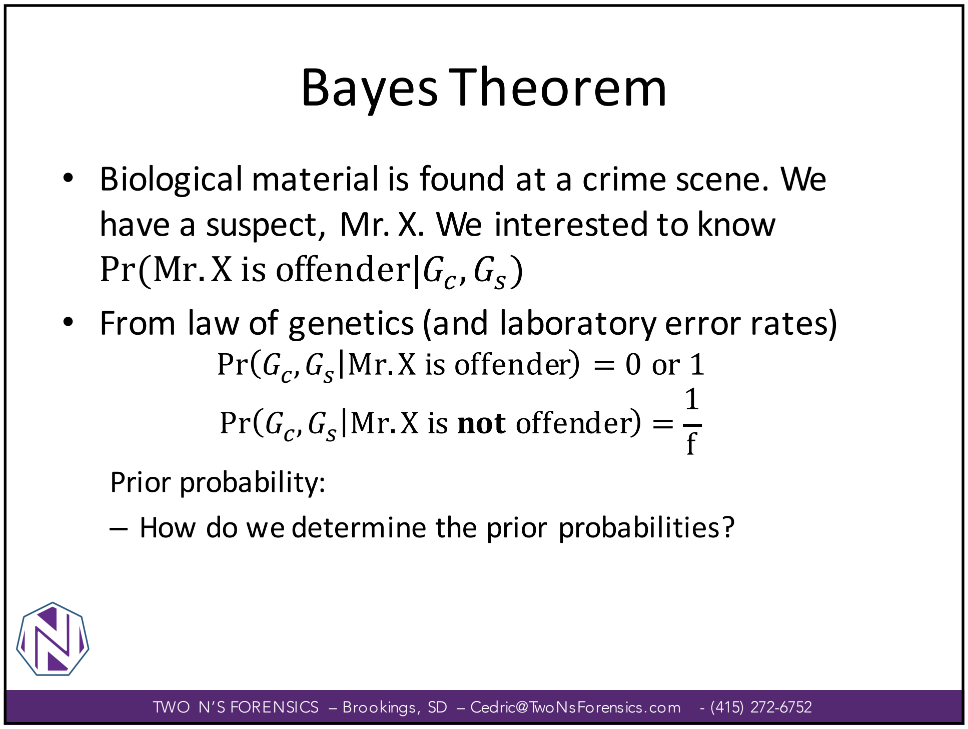
Bayes Theorem / Odds form

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \frac{\Pr(H_p)}{\Pr(H_d)}$$

Posterior odds

Likelihood ratio

Prior odds



Bayes Theorem

- Biological material is found at a crime scene. We have a suspect, Mr. X. We interested to know $\Pr(\text{Mr. X is offender}|G_c, G_s)$
- From law of genetics (and laboratory error rates)

$$\Pr(G_c, G_s|\text{Mr. X is offender}) = 0 \text{ or } 1$$

$$\Pr(G_c, G_s|\text{Mr. X is } \textbf{not} \text{ offender}) = \frac{1}{\mathrm{f}}$$

Prior probability:

– How do we determine the prior probabilities?

Bayes Theorem

- Biological material is found at a crime scene. We have a suspect, Mr. X. We interested to know $\Pr(\text{Mr. X is offender}|G_c, G_s)$
- From law of genetics (and laboratory error rates)

$$\Pr(G_c, G_s|\text{Mr. X is offende}\quad = 0 \text{ or } 1$$

$$\Pr(G_c, G_s|\text{Mr. X is}\quad \text{ffender}) = \frac{1}{\text{f}}$$

P

robabilities?

What is more likely: an adventitious match or a laboratory error?

Bayes Theorem

- Shoe impression is found at a crime scene. We have a suspect, Mr. X. We interested to know $\Pr(\text{Mr. X is offender}|E_u, E_s)$
- From modeling class and acquired characteristics on shoe print (and some other sources of variability)

$$\Pr(E_u, E_s|\text{Mr. X is offender}) \approx 1$$

$$\Pr(E_u, E_s|\text{Mr. X is } \textbf{not} \text{ offender}) = \frac{1}{\text{f}}$$

Prior probability:

– How do we determine the prior probabilities?

Bayes Theorem

- LP impression is found at a crime scene. We have a suspect, Mr. X. We have a "match" between E_u, E_s We interested to know $\Pr(\text{Mr. X is offender}|\text{match})$
- From looking at similarities between ridge pattern and error rates data

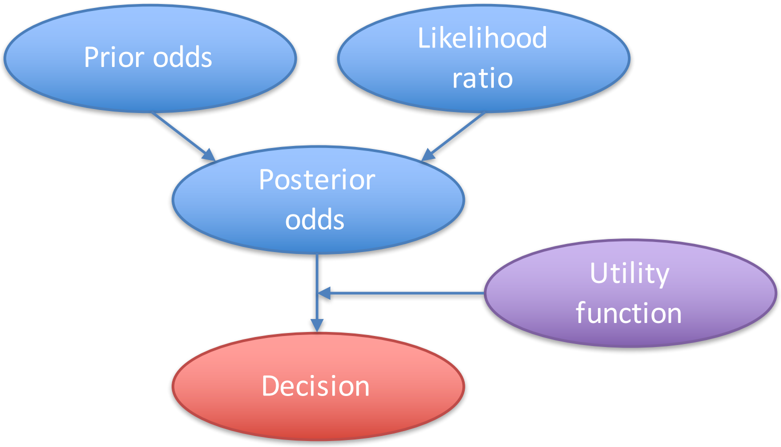
$$\Pr(\text{match}|\text{Mr. X is offender}) \approx 1$$

$$\Pr(\text{match}|\text{Mr. X is } \textbf{not} \text{ offender}) = \frac{1}{\text{f}}$$

Prior probability:

– How do we determine the prior probabilities?

A final note on inference



Prior odds

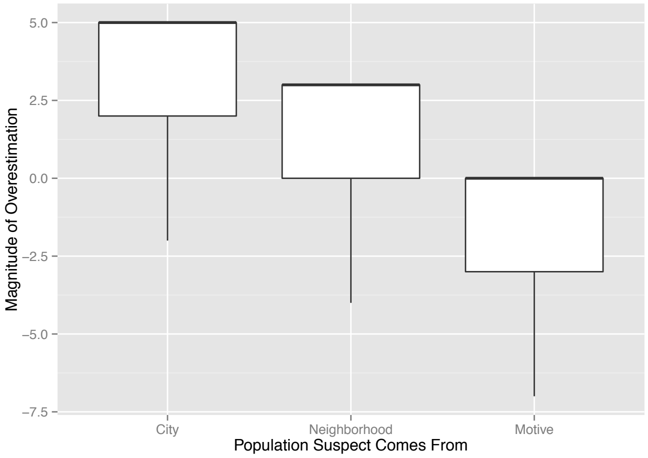
Likelihood ratio

Posterior odds

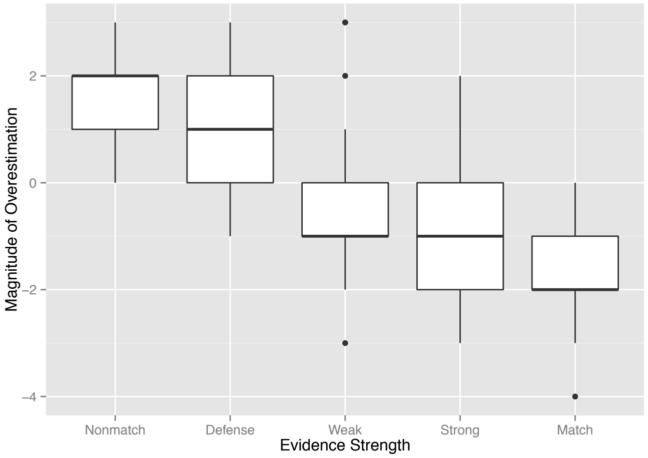
Utility function

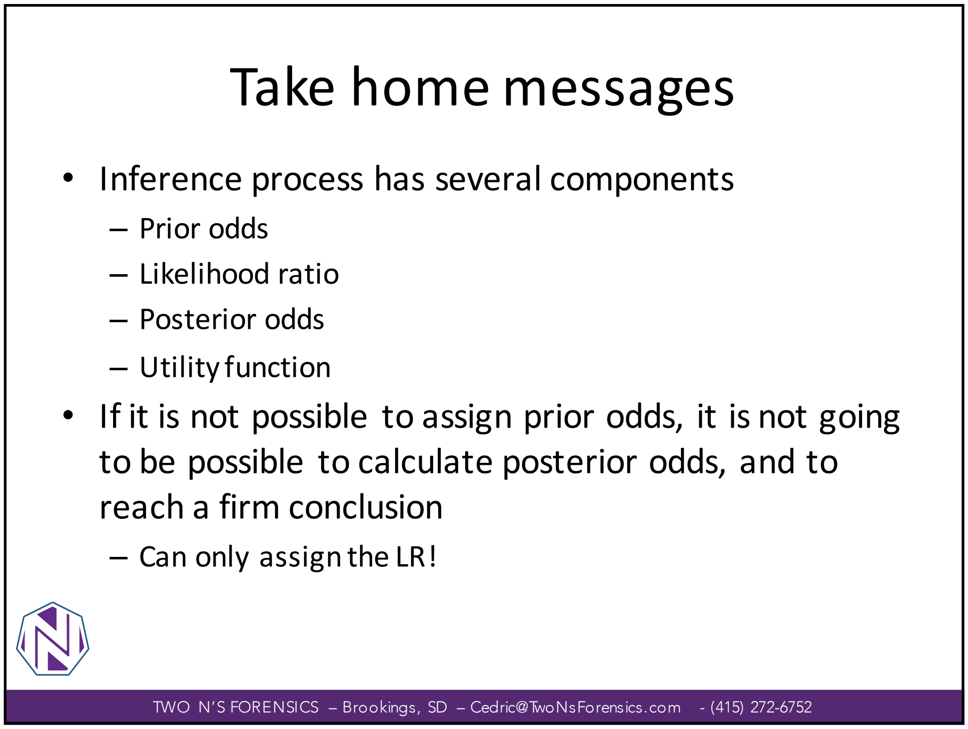
Decision

Pilot study



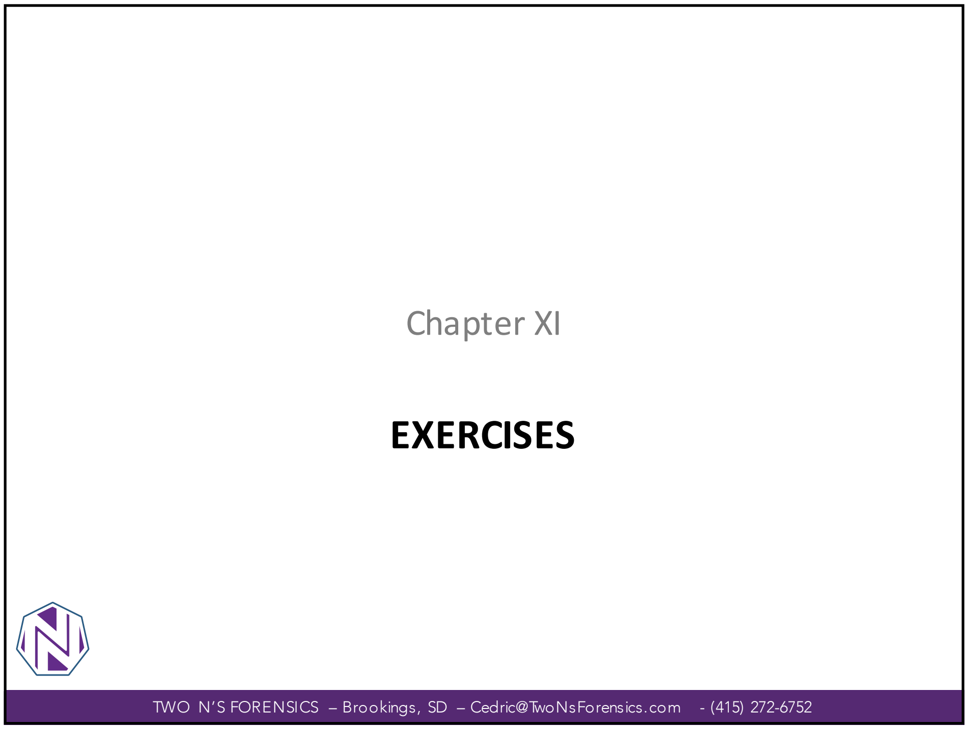
Pilot study





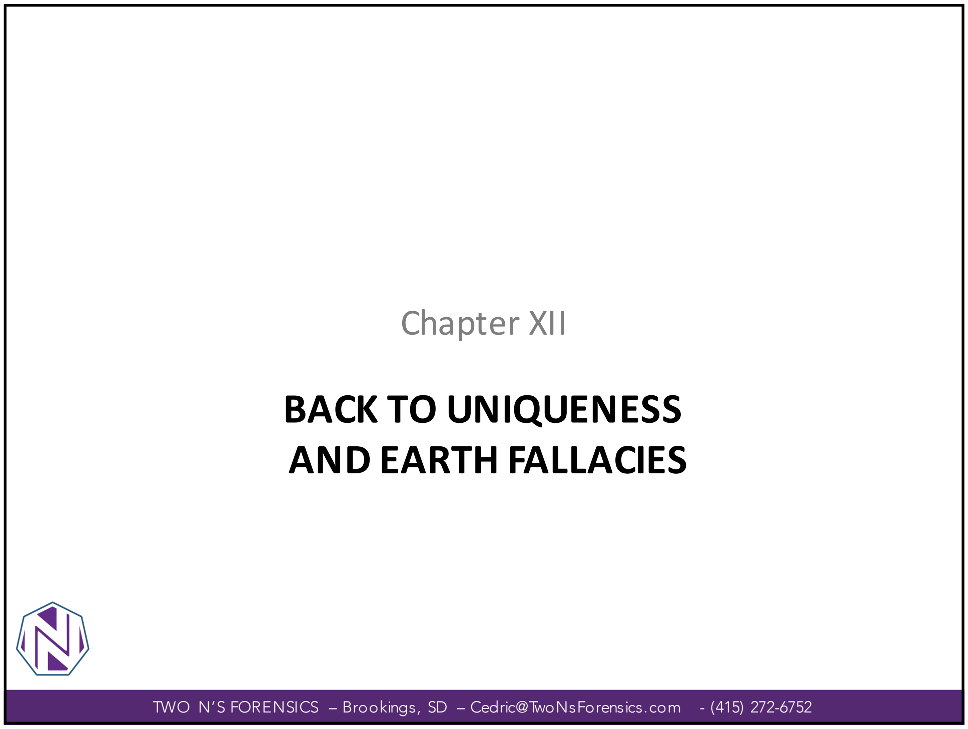
Take home messages

- Inference process has several components
  - Prior odds
  - Likelihood ratio
  - Posterior odds
  - Utility function
- If it is not possible to assign prior odds, it is not going to be possible to calculate posterior odds, and to reach a firm conclusion
  - Can only assign the LR!



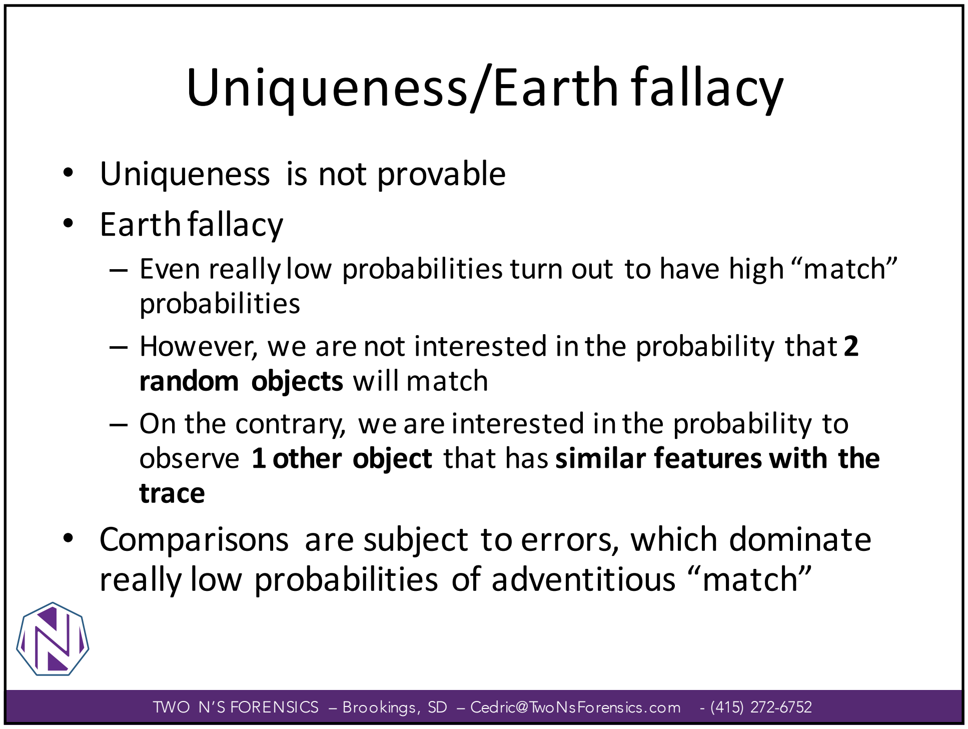
Chapter XI

EXERCISES



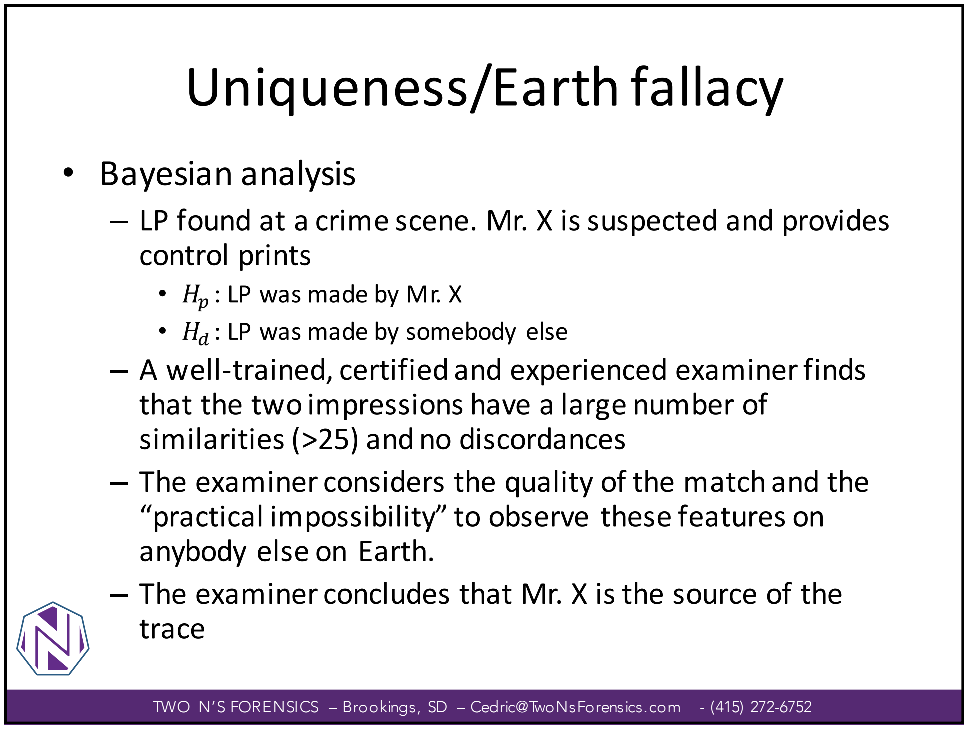
Chapter XII

BACK TO UNIQUENESS AND EARTH FALLACIES



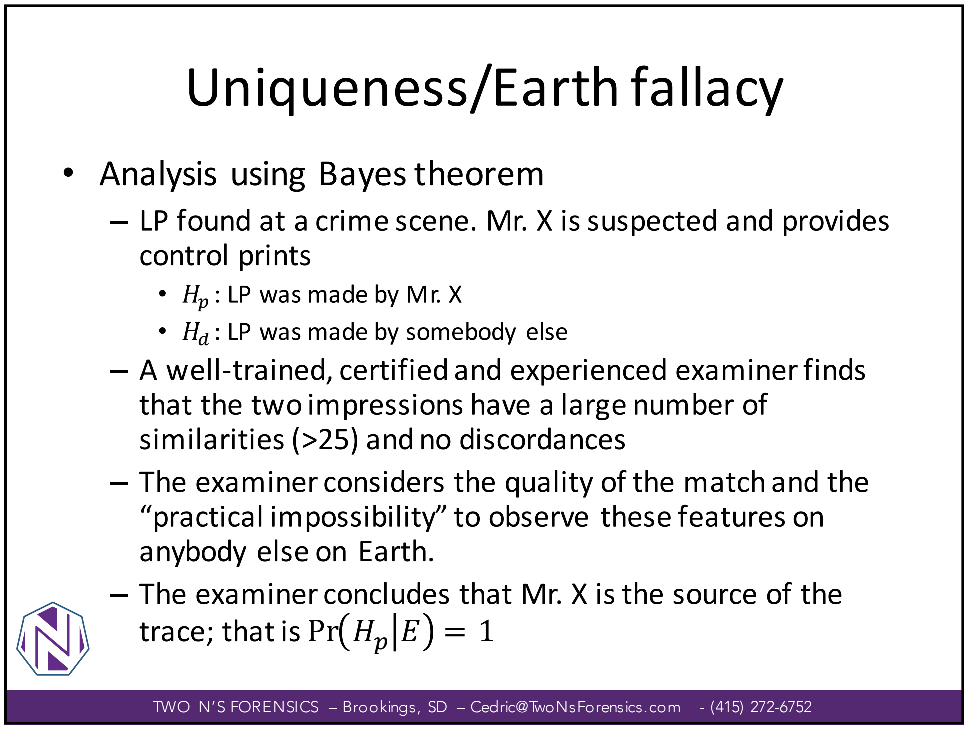
Uniqueness/Earth fallacy

- Uniqueness is not provable
- Earth fallacy
  - Even really low probabilities turn out to have high "match" probabilities
  - However, we are not interested in the probability that **2 random objects** will match
  - On the contrary, we are interested in the probability to observe **1 other object** that has **similar features with the trace**
- Comparisons are subject to errors, which dominate really low probabilities of adventitious "match"



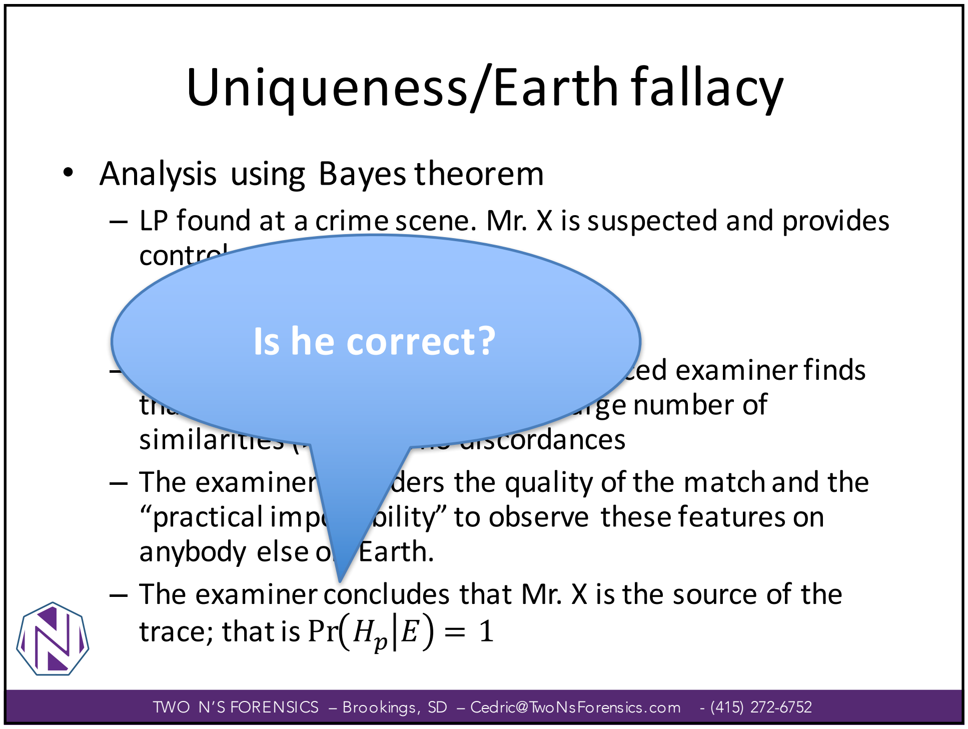
Uniqueness/Earth fallacy

- Bayesian analysis
  - LP found at a crime scene. Mr. X is suspected and provides control prints
    - H_p : LP was made by Mr. X
    - H_d : LP was made by somebody else
  - A well-trained, certified and experienced examiner finds that the two impressions have a large number of similarities (>25) and no discordances
  - The examiner considers the quality of the match and the "practical impossibility" to observe these features on anybody else on Earth.
  - The examiner concludes that Mr. X is the source of the trace



Uniqueness/Earth fallacy

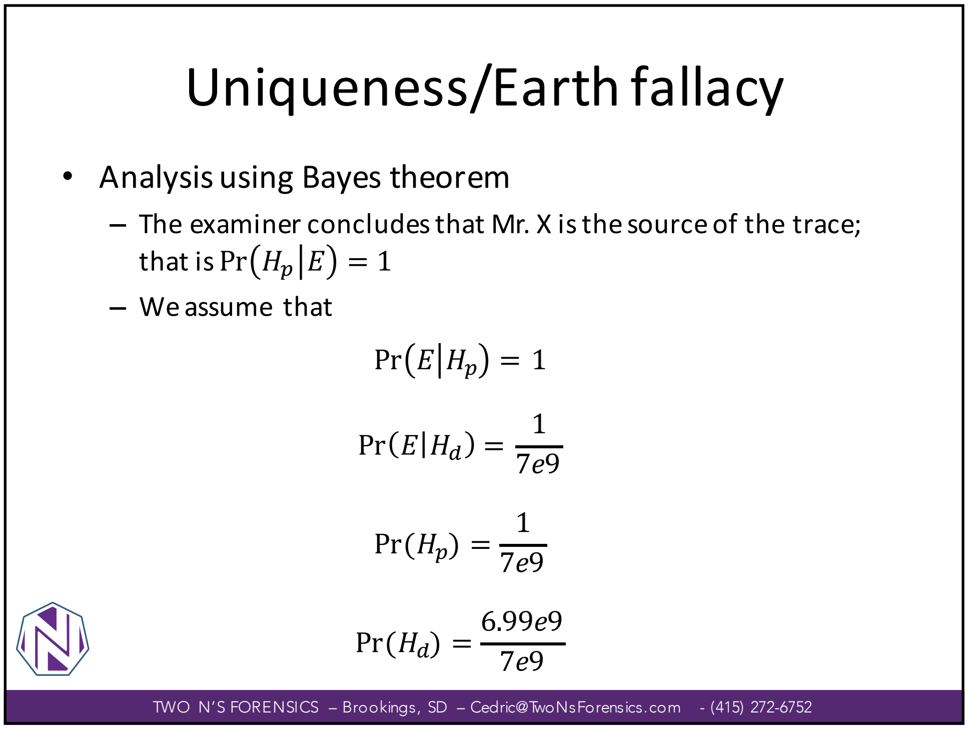
- Analysis using Bayes theorem
  - LP found at a crime scene. Mr. X is suspected and provides control prints
    - H_p : LP was made by Mr. X
    - H_d : LP was made by somebody else
  - A well-trained, certified and experienced examiner finds that the two impressions have a large number of similarities (>25) and no discordances
  - The examiner considers the quality of the match and the "practical impossibility" to observe these features on anybody else on Earth.
  - The examiner concludes that Mr. X is the source of the trace; that is $\Pr(H_p|E) = 1$



Uniqueness/Earth fallacy

- Analysis using Bayes theorem
  - LP found at a crime scene. Mr. X is suspected and provides contr
  - ed examiner finds tha rge number of similarities discordances
  - The examiner ders the quality of the match and the "practical imp bility" to observe these features on anybody else o Earth.
  - The examiner concludes that Mr. X is the source of the trace; that is $\Pr(H_p|E) = 1$

Is he correct?



Uniqueness/Earth fallacy

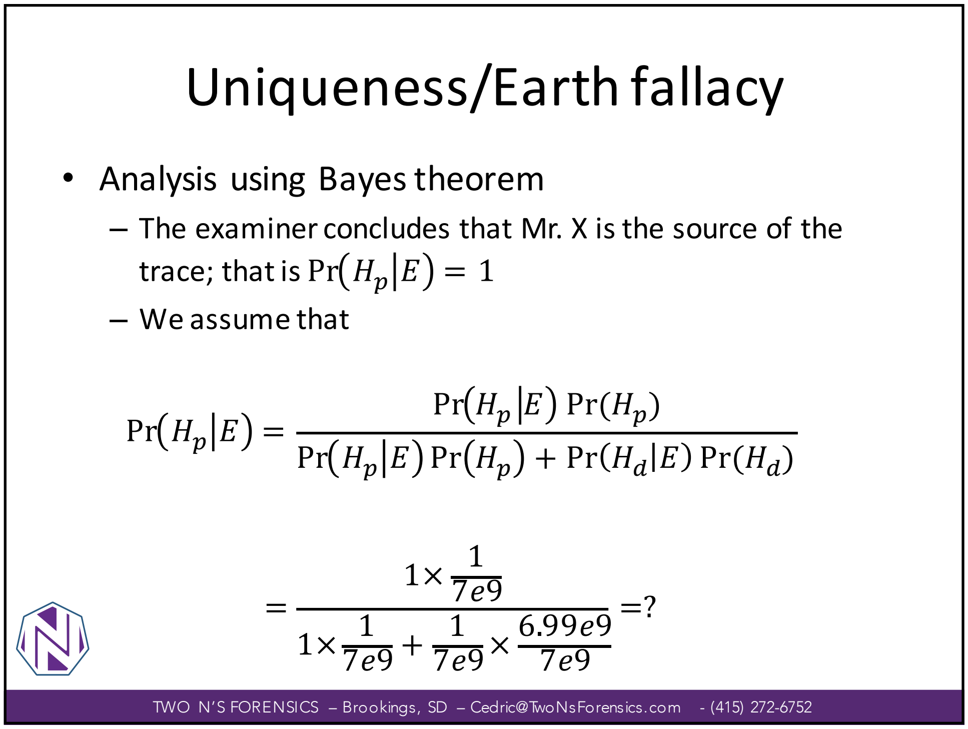
- Analysis using Bayes theorem
  - The examiner concludes that Mr. X is the source of the trace; that is $\Pr(H_p|E) = 1$
  - We assume that

$$\Pr(E|H_p) = 1$$

$$\Pr(E|H_d) = \frac{1}{7e9}$$

$$\Pr(H_p) = \frac{1}{7e9}$$

$$\Pr(H_d) = \frac{6.99e9}{7e9}$$

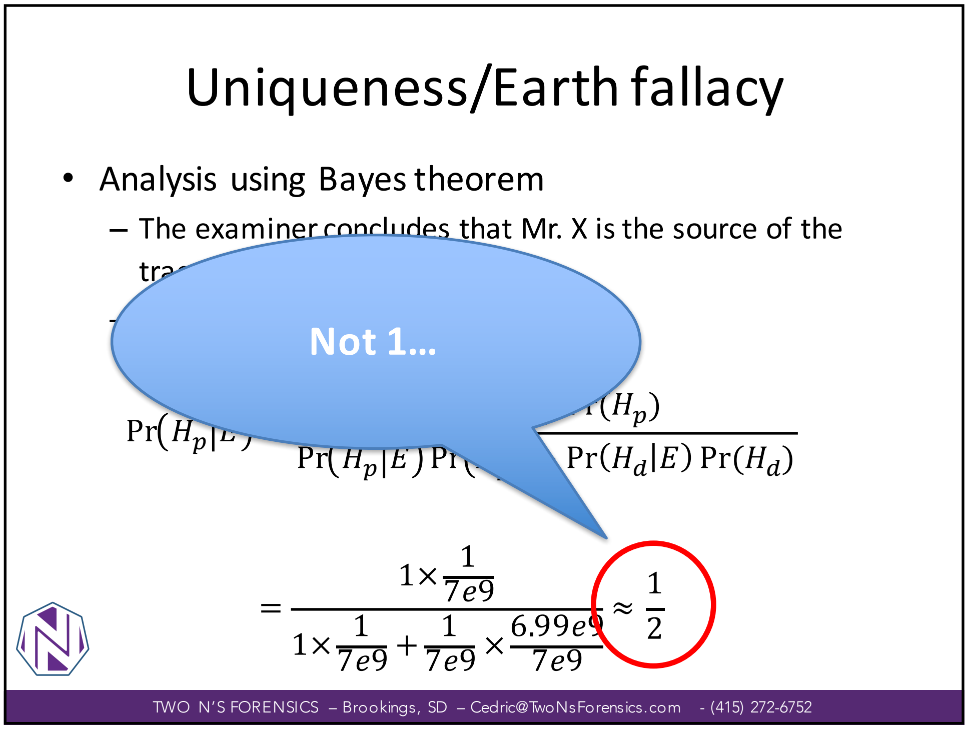


Uniqueness/Earth fallacy

- Analysis using Bayes theorem
  - The examiner concludes that Mr. X is the source of the trace; that is $\mathrm{Pr}(H_p|E) = 1$
  - We assume that

$$\mathrm{Pr}(H_p|E) = \frac{\mathrm{Pr}(H_p|E)\,\mathrm{Pr}(H_p)}{\mathrm{Pr}(H_p|E)\,\mathrm{Pr}(H_p) + \mathrm{Pr}(H_d|E)\,\mathrm{Pr}(H_d)}$$

$$= \frac{1 \times \frac{1}{7e9}}{1 \times \frac{1}{7e9} + \frac{1}{7e9} \times \frac{6.99e9}{7e9}} =?$$



Uniqueness/Earth fallacy

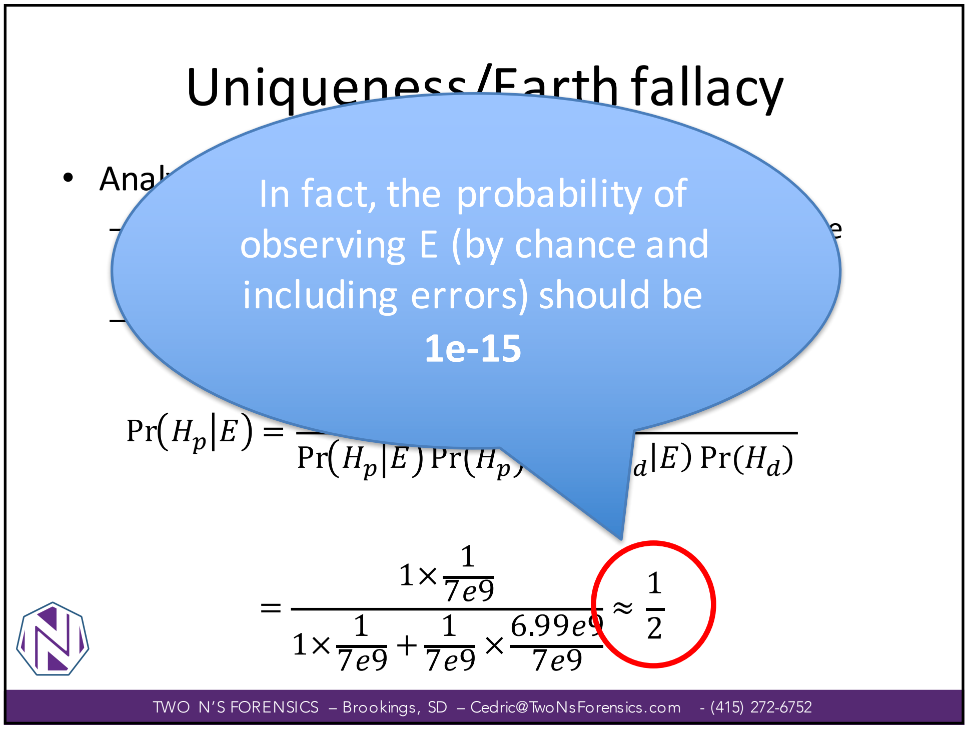
- Analysis using Bayes theorem
  - The examiner concludes that Mr. X is the source of the tra
  -

Not 1…

$\mathrm{Pr}(H_p|E$ … (H_p)

$\mathrm{Pr}(H_p|E)\,\mathrm{Pr}($ … $\mathrm{Pr}(H_d|E)\,\mathrm{Pr}(H_d)$

$$= \frac{1 \times \frac{1}{7e9}}{1 \times \frac{1}{7e9} + \frac{1}{7e9} \times \frac{6.99e9}{7e9}} \approx \frac{1}{2}$$



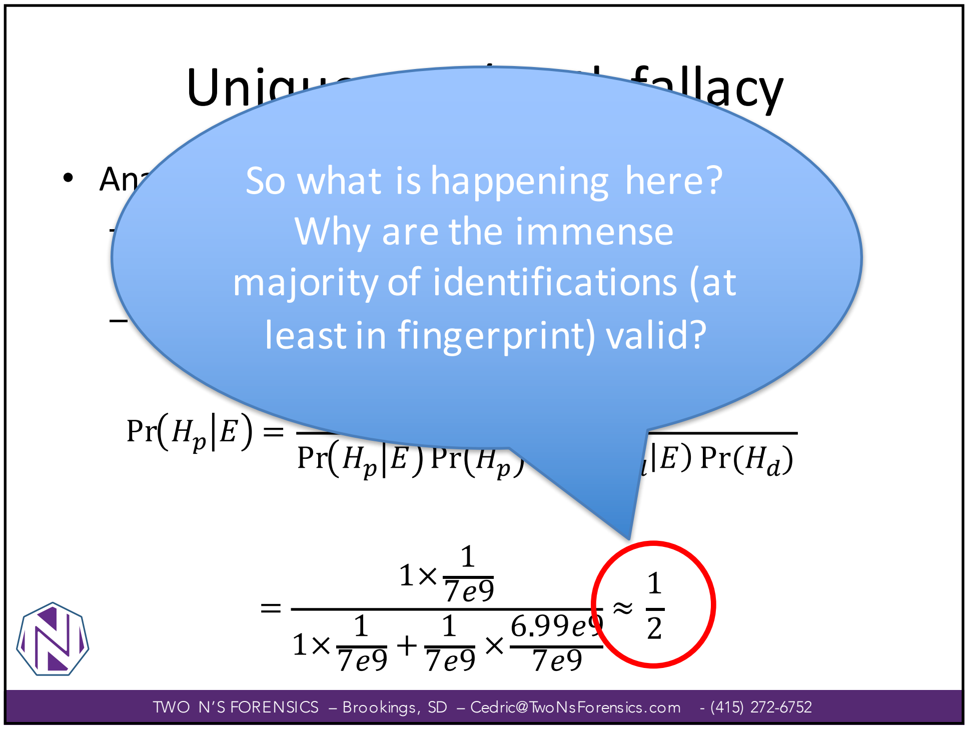
Uniqueness/Earth fallacy

- Anal

In fact, the probability of observing E (by chance and including errors) should be **1e-15**

$$\Pr(H_p|E) = \frac{}{\Pr(H_p|E)\Pr(H_p) \quad {}_d|E)\Pr(H_d)}$$

$$= \frac{1\times\frac{1}{7e9}}{1\times\frac{1}{7e9}+\frac{1}{7e9}\times\frac{6.99e9}{7e9}} \approx \frac{1}{2}$$



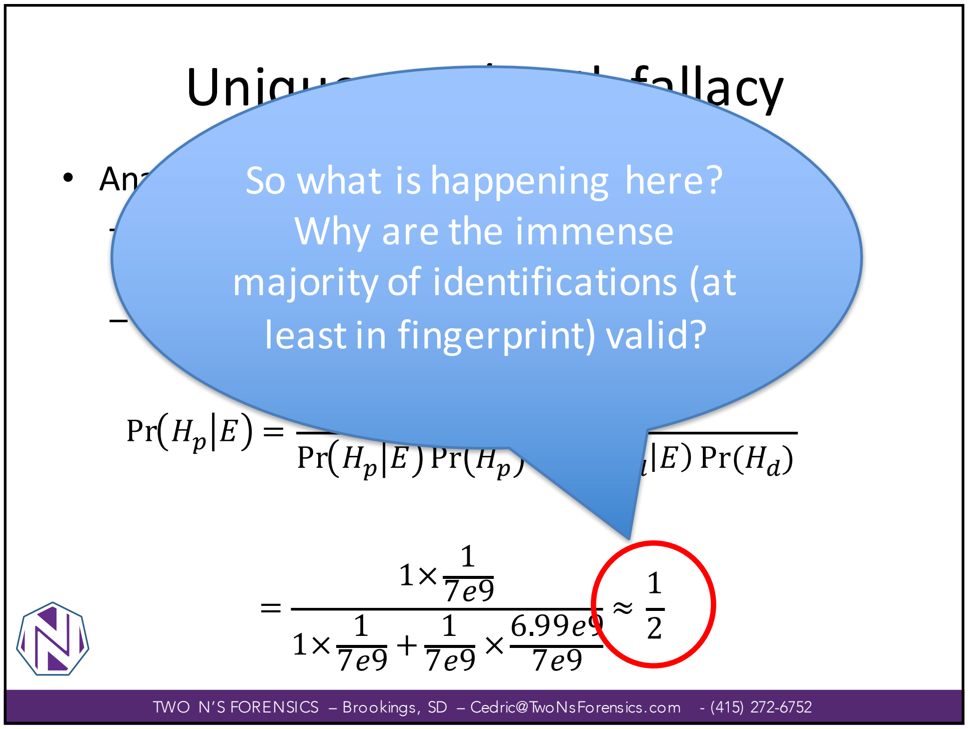
Uniq fallacy

- An

So what is happening here? Why are the immense majority of identifications (at least in fingerprint) valid?

$$\Pr(H_p|E) = \frac{}{\Pr(H_p|E)\Pr(H_p) \quad |E)\Pr(H_d)}$$

$$= \frac{1\times\frac{1}{7e9}}{1\times\frac{1}{7e9}+\frac{1}{7e9}\times\frac{6.99e9}{7e9}} \approx \frac{1}{2}$$



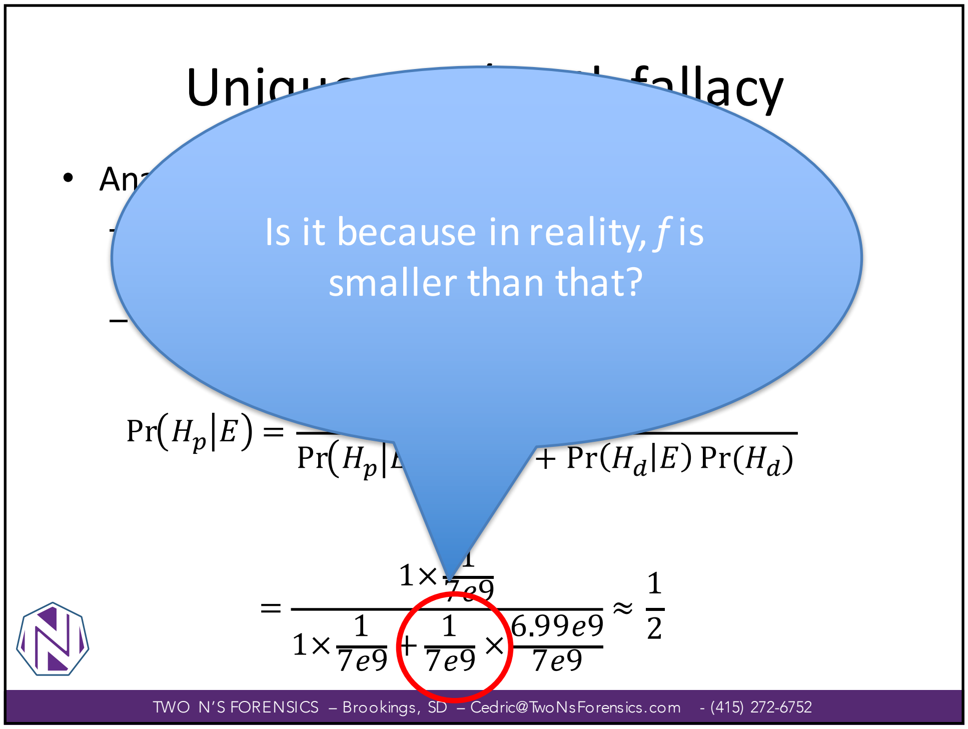
Uniqu fallacy

- An
  -
  -

So what is happening here? Why are the immense majority of identifications (at least in fingerprint) valid?

$$\Pr(H_p|E) = \frac{}{\Pr(H_p|E)\Pr(H_p) \quad |E)\Pr(H_d)}$$

$$= \frac{1\times\frac{1}{7e9}}{1\times\frac{1}{7e9}+\frac{1}{7e9}\times\frac{6.99e9}{7e9}} \approx \frac{1}{2}$$



Uniqu fallacy

- An
  -
  -

Is it because in reality, *f* is smaller than that?

$$\Pr(H_p|E) = \frac{}{\Pr(H_p|E \quad + \Pr(H_d|E)\Pr(H_d)}$$

$$= \frac{1\times\frac{1}{7e9}}{1\times\frac{1}{7e9}+\frac{1}{7e9}\times\frac{6.99e9}{7e9}} \approx \frac{1}{2}$$

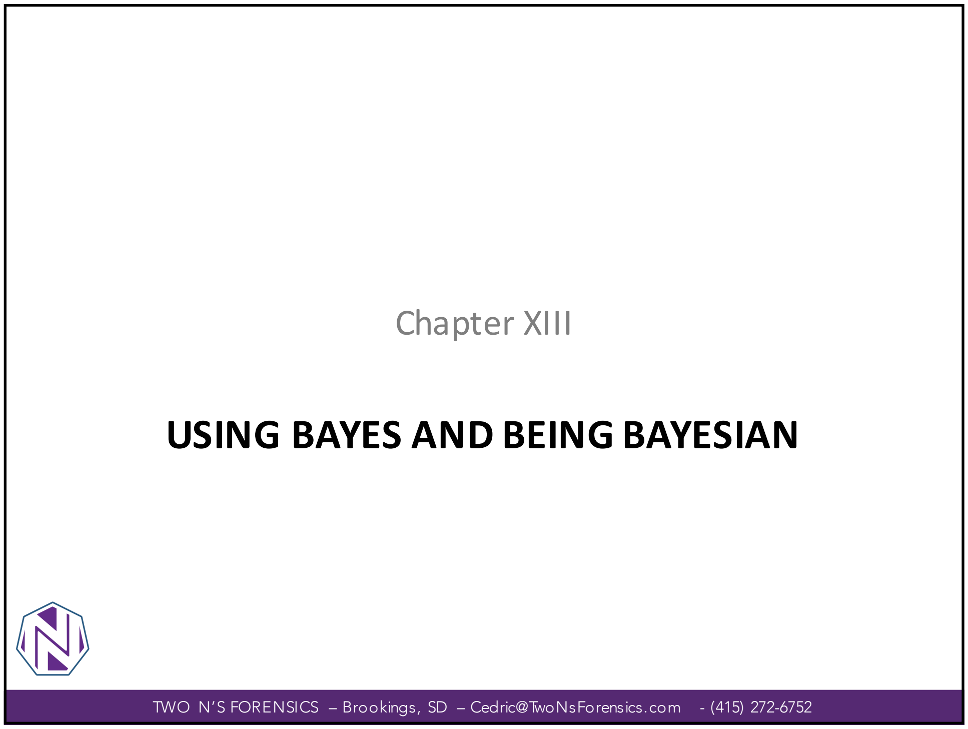
Uniqueness/Earth fallacy

- Analysis using Bayes theorem
  - The examiner concludes that Mr. X is the source of the trace; that is $\Pr(H_p|E) = 1$
  - We assume that

$$\Pr(H_p|E) = \frac{\Pr(H_p|E)\Pr(H_p)}{\Pr(H_p|E)\Pr(H_p) + [RMP + FPP(1 - RMP)]\Pr(H_d)}$$

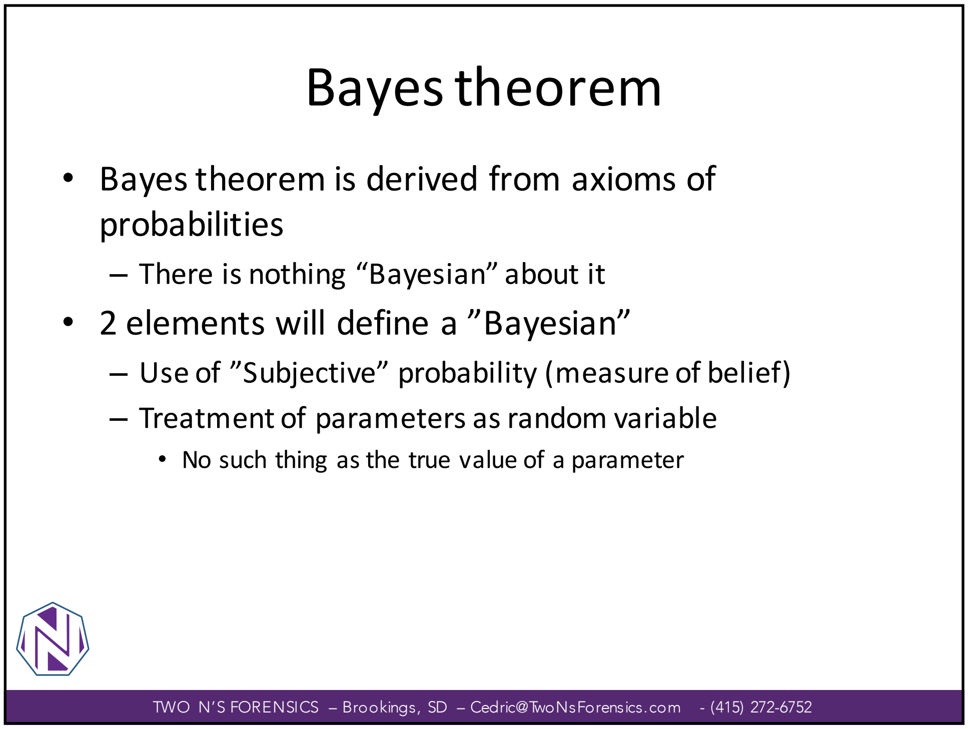
Take home messages

- We can apply Bayes theorem to analyze the Earth/Uniqueness fallacy
  - Shows that we really need to have a ridiculously rare sets of feature (and no error) to have a near certain posterior probability
  - Shows that in most cases the Earth population is not truly considered
  - Shows that the utility function to map a posterior probability to a categorical conclusion plays an important role



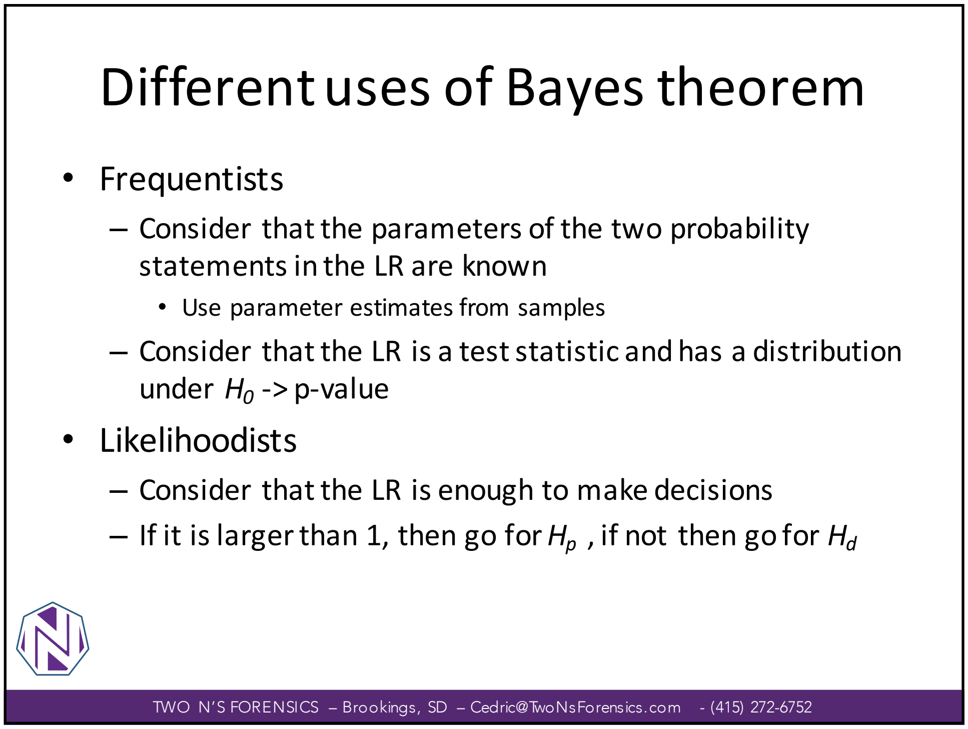
Chapter XIII

USING BAYES AND BEING BAYESIAN



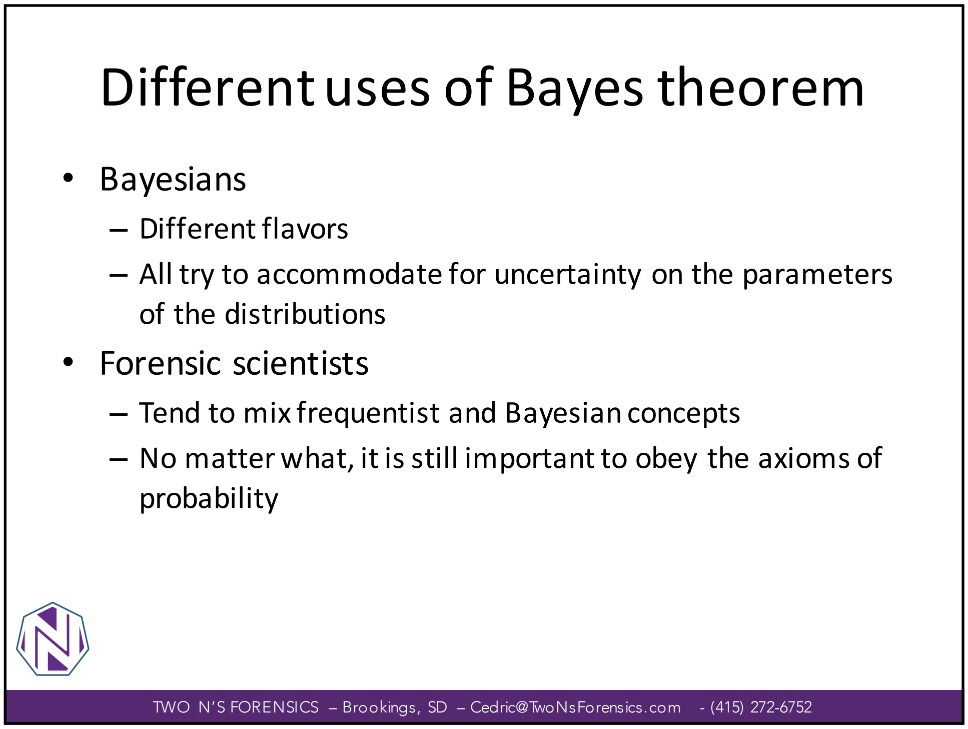
Bayes theorem

- Bayes theorem is derived from axioms of probabilities
  - There is nothing “Bayesian” about it
- 2 elements will define a ”Bayesian”
  - Use of ”Subjective” probability (measure of belief)
  - Treatment of parameters as random variable
    - No such thing as the true value of a parameter



Different uses of Bayes theorem

- Frequentists
  - Consider that the parameters of the two probability statements in the LR are known
    - Use parameter estimates from samples
  - Consider that the LR is a test statistic and has a distribution under H_0 -> p-value
- Likelihoodists
  - Consider that the LR is enough to make decisions
  - If it is larger than 1, then go for H_p , if not then go for H_d



Different uses of Bayes theorem

- Bayesians
  - Different flavors
  - All try to accommodate for uncertainty on the parameters of the distributions
- Forensic scientists
  - Tend to mix frequentist and Bayesian concepts
  - No matter what, it is still important to obey the axioms of probability

Pitfall of Bayesian approach

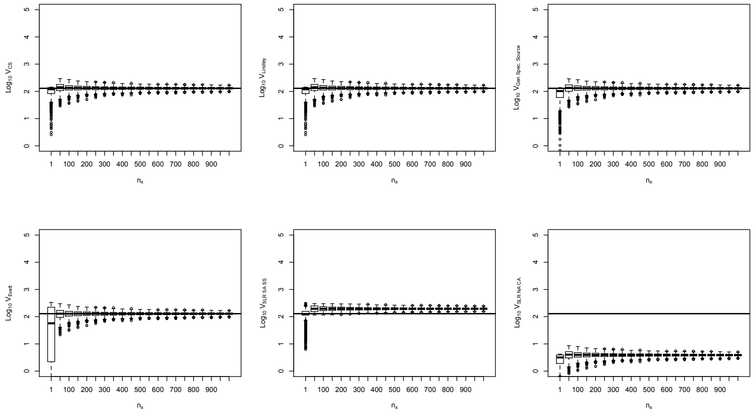
- Subjective probability
- Formally capturing the uncertainty on the parameters

$$BF = \frac{\int f(E_u|\theta) f(\theta|E_s)\pi(\theta)d\theta}{\int f(E_u|\theta) f(\theta|E_a)\pi(\theta)d\theta}$$

  - Some researchers propose to have CI on the LR
  - Some researchers propose to have PI on the LR
  - Some researchers just propose to report 1 value
    - It's "my LR" – "it's the best I can do"
- Developing LR in high dimensions (i.e., for complex variables such as fingerprint, shoeprint, DNA mixtures)

Other approaches

- Data dimension reduction
  - "Score based" approach



Other approaches

- Two stages approach (as Parker 1966)
  - Statistical hypothesis test for the numerator
  - Repeated statistical hypothesis tests for the denominator
    - Tally how many random source fail to reject hypothesis of similarity
- Random man non-excluded / Probability of inclusion
  - DNA
  - Evaluate how many individual could not be excluded based on the trace
  - Essentially related to the denominator of the LR
    - But can have issues when the numerator is not 1

Other approaches

- Completely subjective approach
  - R v T

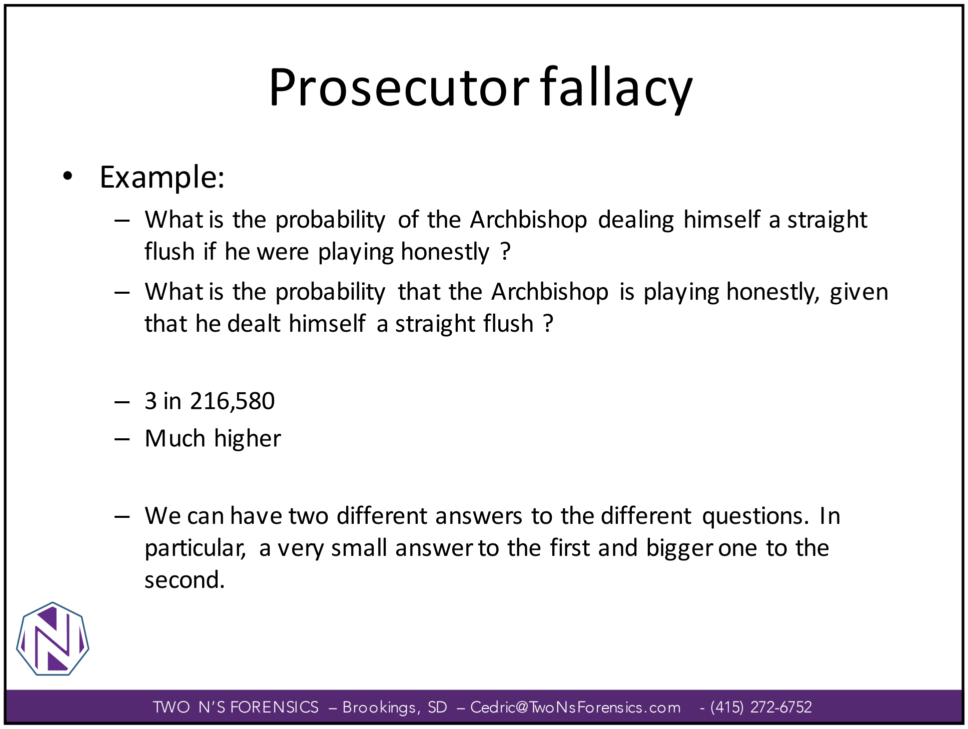
Other approaches

- Posterior probabilities
  - QD
- Categorical
  - Fingerprint
- Consistent with / cannot exclude

Prosecutor fallacy

- D. J. Balding and P. J. Donnelly: The prosecutor's fallacy and DNA evidence. Criminal Law Review, 1994, 711-721.

- Two different questions
  - What is the probability that the defendant fingerprint match the latent on the crime scene, given that he is innocent ?
  - What is the probability that the defendant is innocent, given that his fingerprint match the latent on the crime scene ?

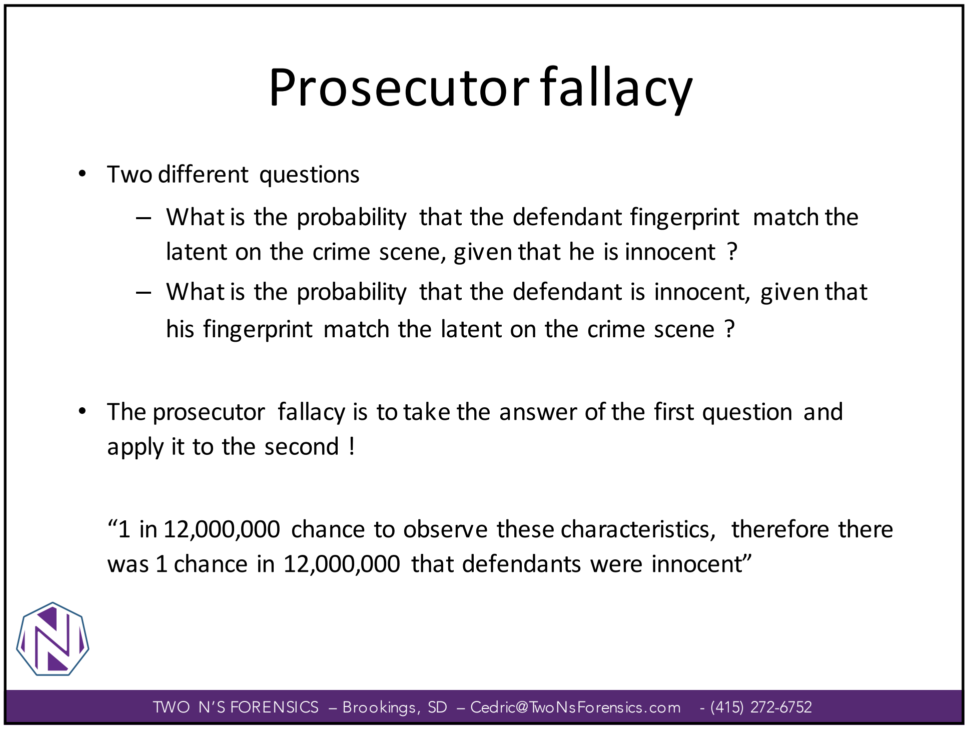


Prosecutor fallacy

- Example:
  - What is the probability of the Archbishop dealing himself a straight flush if he were playing honestly ?
  - What is the probability that the Archbishop is playing honestly, given that he dealt himself a straight flush ?

  - 3 in 216,580
  - Much higher

  - We can have two different answers to the different questions. In particular, a very small answer to the first and bigger one to the second.

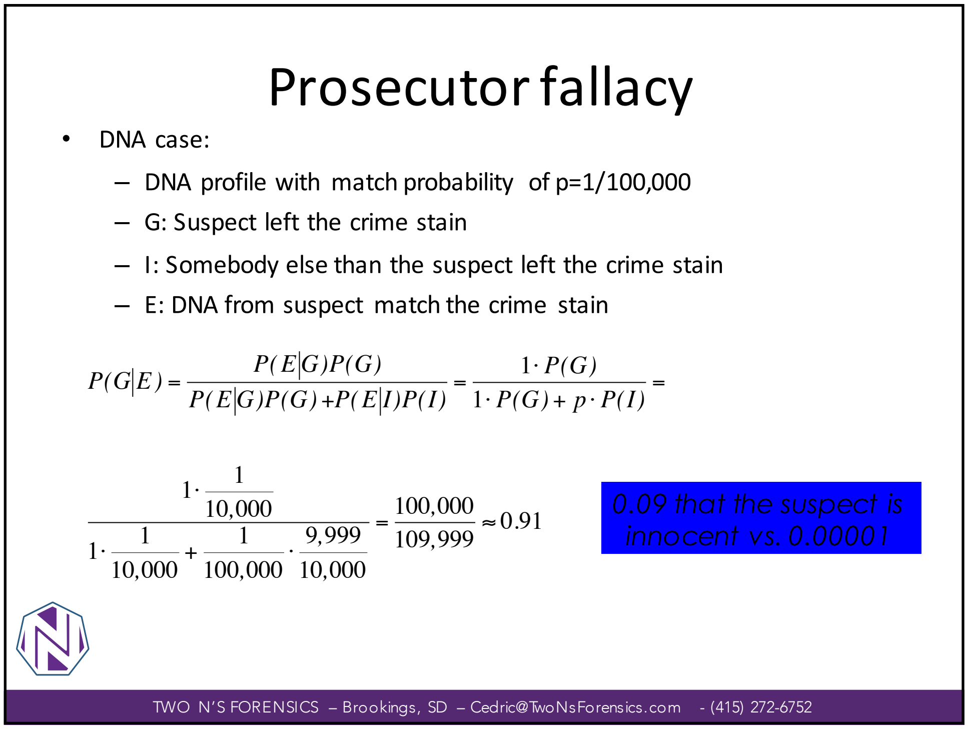


Prosecutor fallacy

- Two different questions
  - What is the probability that the defendant fingerprint match the latent on the crime scene, given that he is innocent ?
  - What is the probability that the defendant is innocent, given that his fingerprint match the latent on the crime scene ?

- The prosecutor fallacy is to take the answer of the first question and apply it to the second !

"1 in 12,000,000 chance to observe these characteristics, therefore there was 1 chance in 12,000,000 that defendants were innocent"



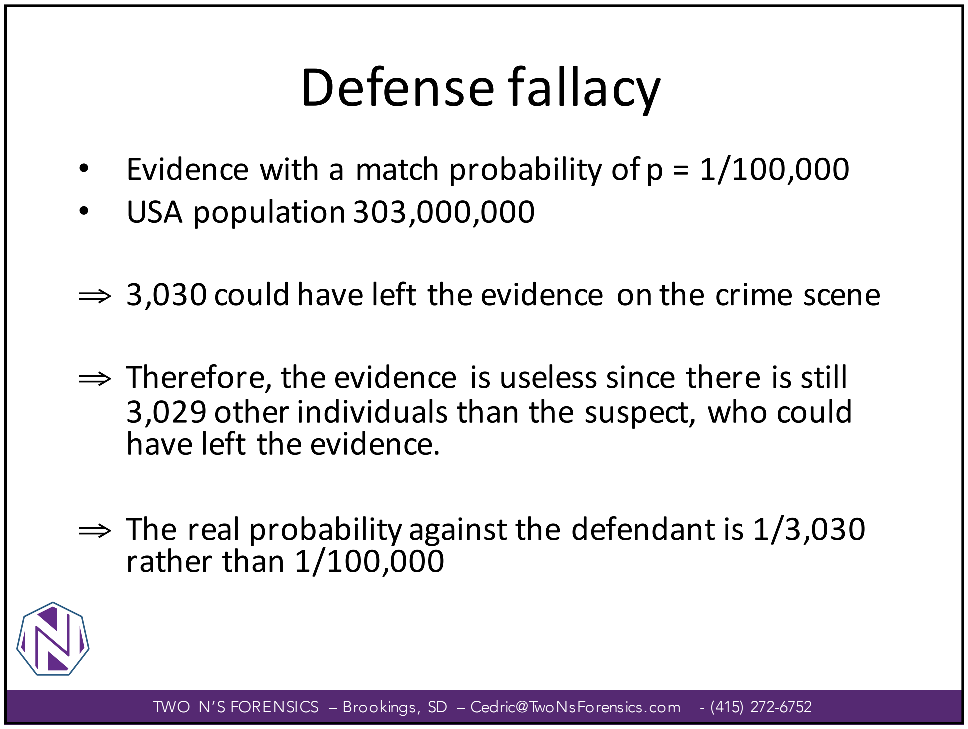
Prosecutor fallacy

- DNA case:
  - DNA profile with match probability of p=1/100,000
  - G: Suspect left the crime stain
  - I: Somebody else than the suspect left the crime stain
  - E: DNA from suspect match the crime stain

$$P(G|E)=\frac{P(E|G)P(G)}{P(E|G)P(G)+P(E|I)P(I)}=\frac{1\cdot P(G)}{1\cdot P(G)+p\cdot P(I)}=$$

$$\frac{1\cdot\frac{1}{10,000}}{1\cdot\frac{1}{10,000}+\frac{1}{100,000}\cdot\frac{9,999}{10,000}}=\frac{100,000}{109,999}\approx 0.91$$

0.09 that the suspect is innocent vs. 0.00001



Defense fallacy

- Evidence with a match probability of p = 1/100,000
- USA population 303,000,000

⇒ 3,030 could have left the evidence on the crime scene

⇒ Therefore, the evidence is useless since there is still 3,029 other individuals than the suspect, who could have left the evidence.

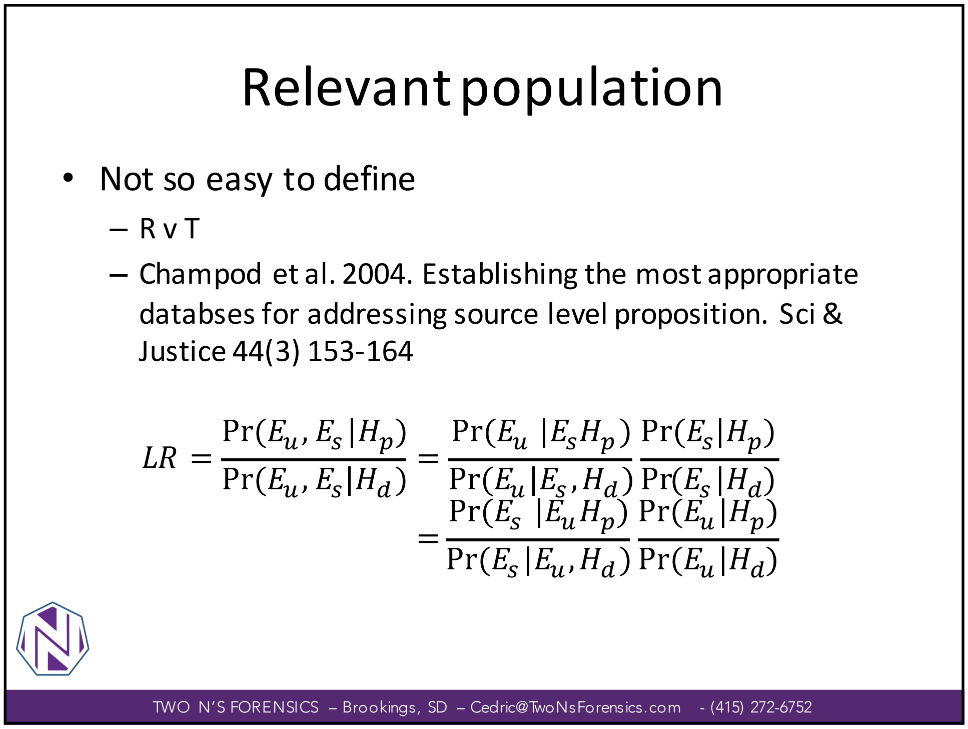
⇒ The real probability against the defendant is 1/3,030 rather than 1/100,000

Take home messages

- Different types of conclusions
  - They all relate to the Bayes theorem
  - Provided that we have the right kind of data and arguments, we can justify most of them
- Truly Bayesian approach is extremely complex to implement
- Most likelihood ratio methods proposed in forensic science are hybrid methods
- We need to be careful when expressing probabilities in court

Chapter XIV

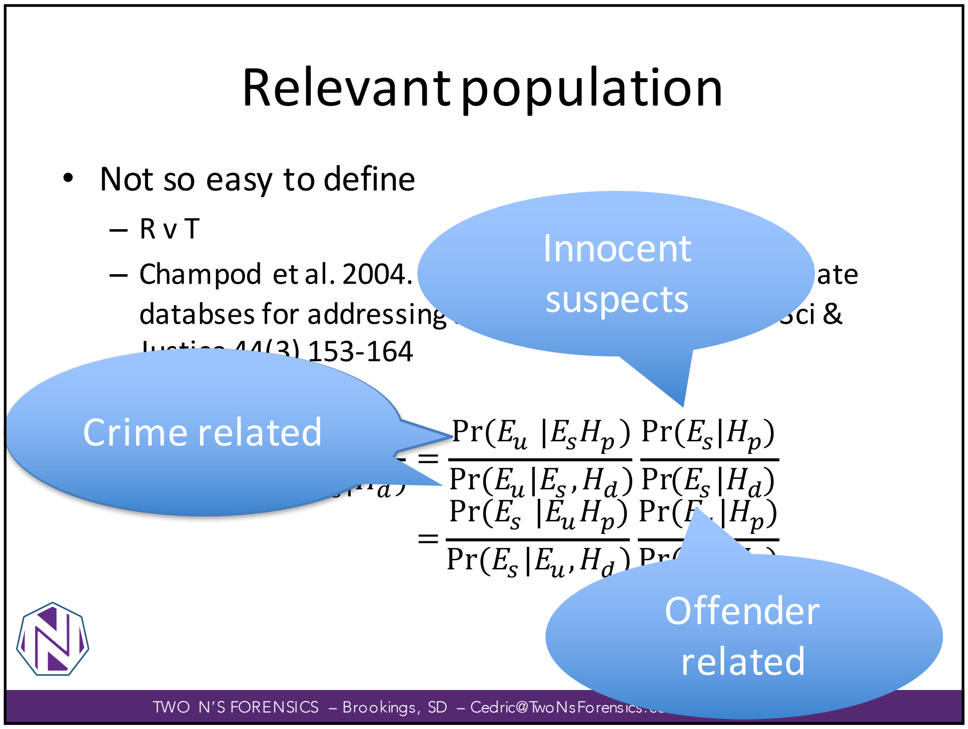
RELEVANT POPULATION AND DATABASES



Relevant population

- Not so easy to define
  - R v T
  - Champod et al. 2004. Establishing the most appropriate databses for addressing source level proposition. Sci & Justice 44(3) 153-164

$$LR = \frac{\Pr(E_u, E_s | H_p)}{\Pr(E_u, E_s | H_d)} = \frac{\Pr(E_u \, | E_s H_p)}{\Pr(E_u | E_s, H_d)} \frac{\Pr(E_s | H_p)}{\Pr(E_s | H_d)}$$
$$= \frac{\Pr(E_s \, | E_u H_p)}{\Pr(E_s | E_u, H_d)} \frac{\Pr(E_u | H_p)}{\Pr(E_u | H_d)}$$



Relevant population

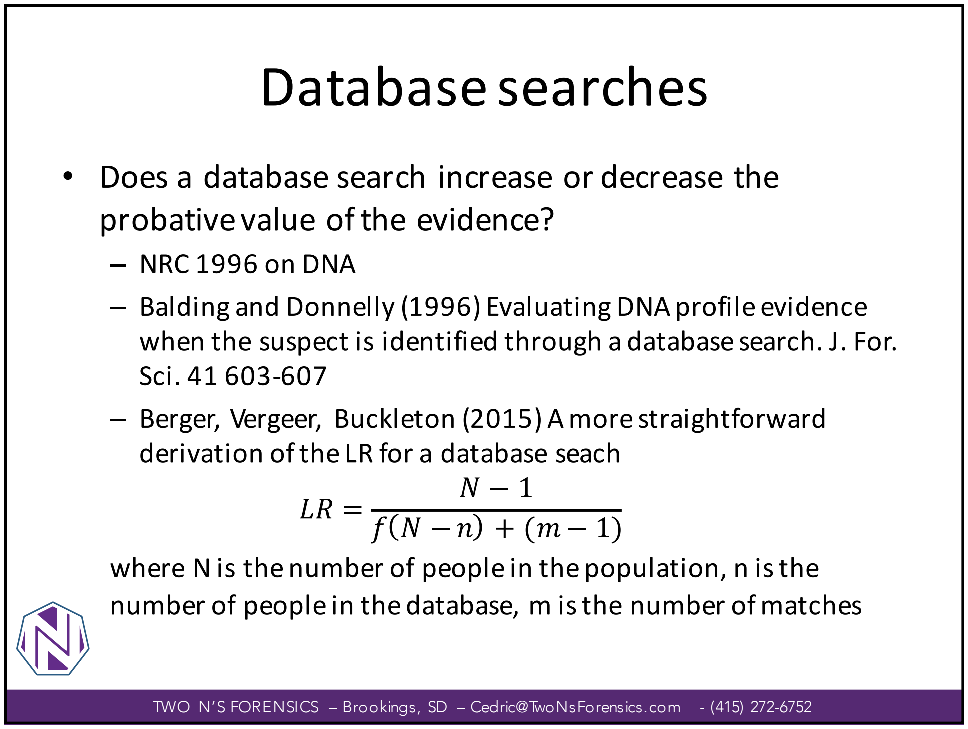
- Not so easy to define
  - R v T
  - Champod et al. 2004. ate databses for addressing ci & Justice 44(3) 153-164

Innocent suspects

Crime related

$$= \frac{\Pr(E_u \, | E_s H_p)}{\Pr(E_u | E_s, H_d)} \frac{\Pr(E_s | H_p)}{\Pr(E_s | H_d)}$$
$$= \frac{\Pr(E_s \, | E_u H_p)}{\Pr(E_s | E_u, H_d)}$$

Offender related

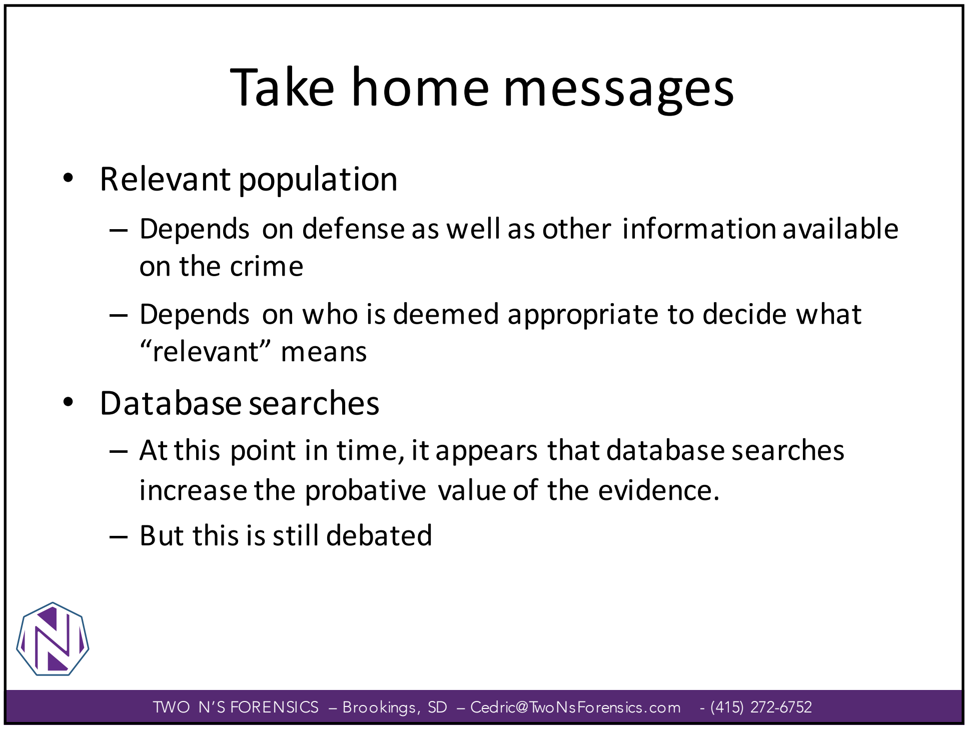


Database searches

- Does a database search increase or decrease the probative value of the evidence?
  - NRC 1996 on DNA
  - Balding and Donnelly (1996) Evaluating DNA profile evidence when the suspect is identified through a database search. J. For. Sci. 41 603-607
  - Berger, Vergeer, Buckleton (2015) A more straightforward derivation of the LR for a database seach

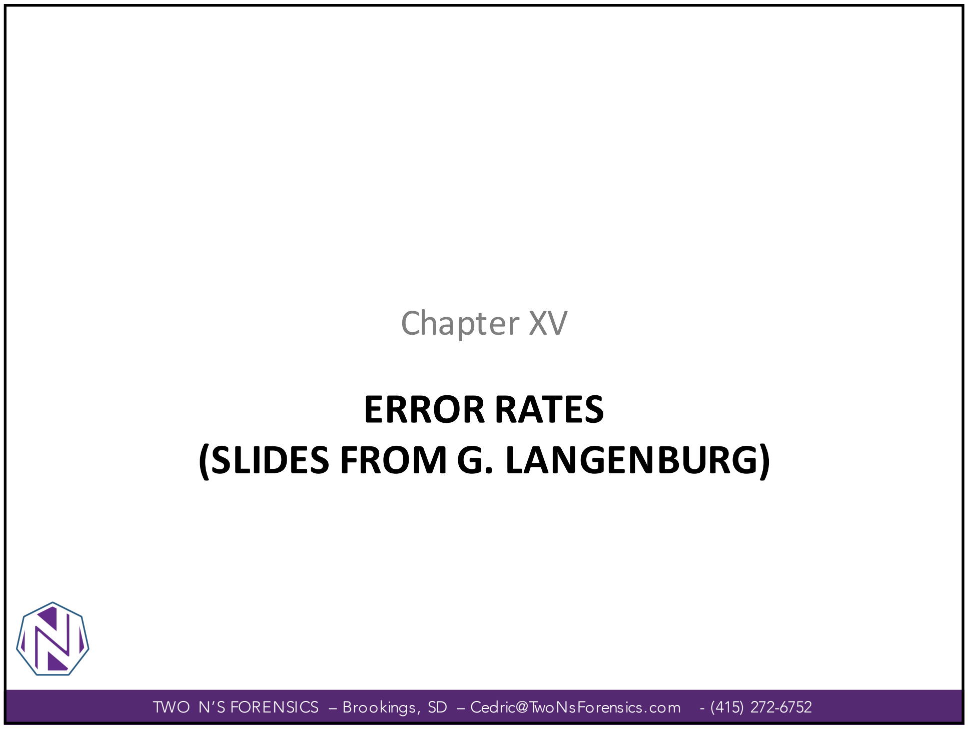
$$LR = \frac{N-1}{f(N-n)+(m-1)}$$

where N is the number of people in the population, n is the number of people in the database, m is the number of matches



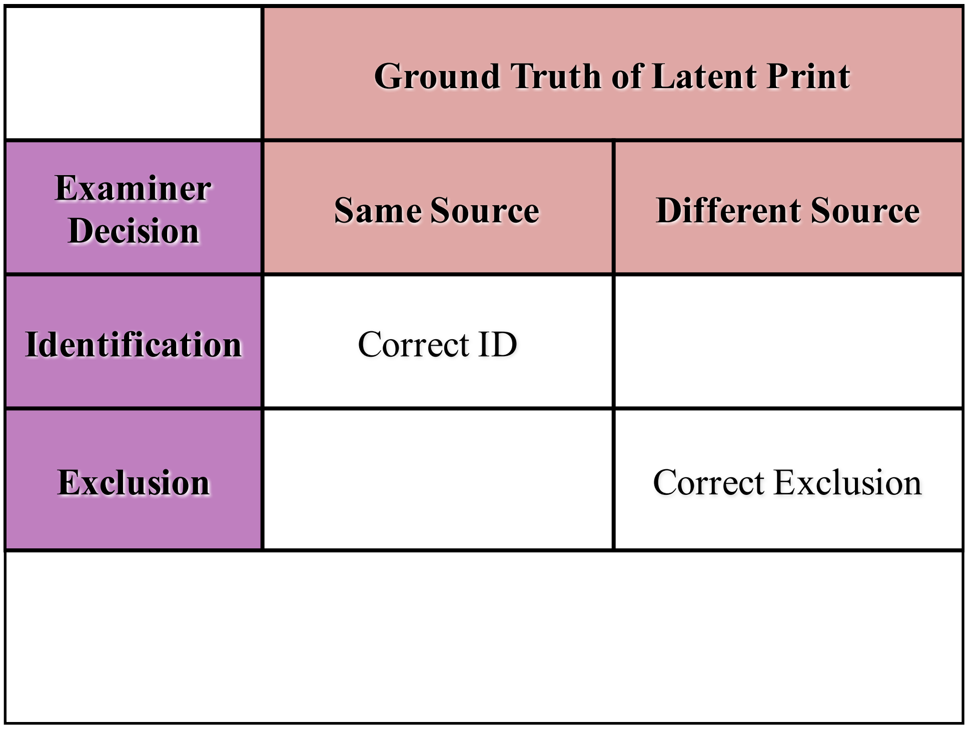
Take home messages

- Relevant population
  - Depends on defense as well as other information available on the crime
  - Depends on who is deemed appropriate to decide what "relevant" means
- Database searches
  - At this point in time, it appears that database searches increase the probative value of the evidence.
  - But this is still debated

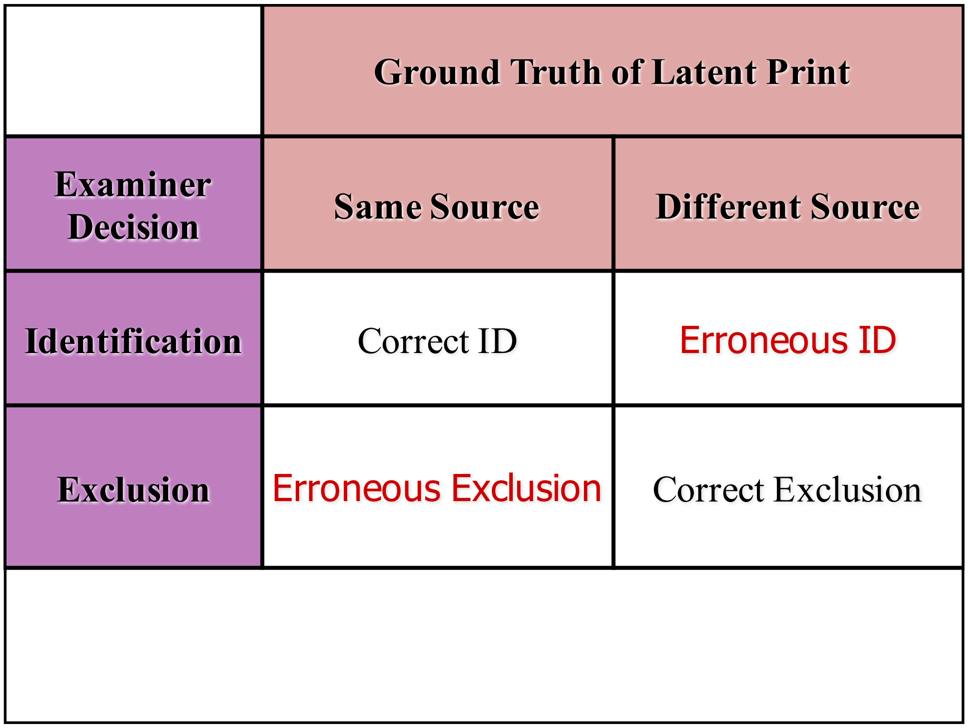


Chapter XV

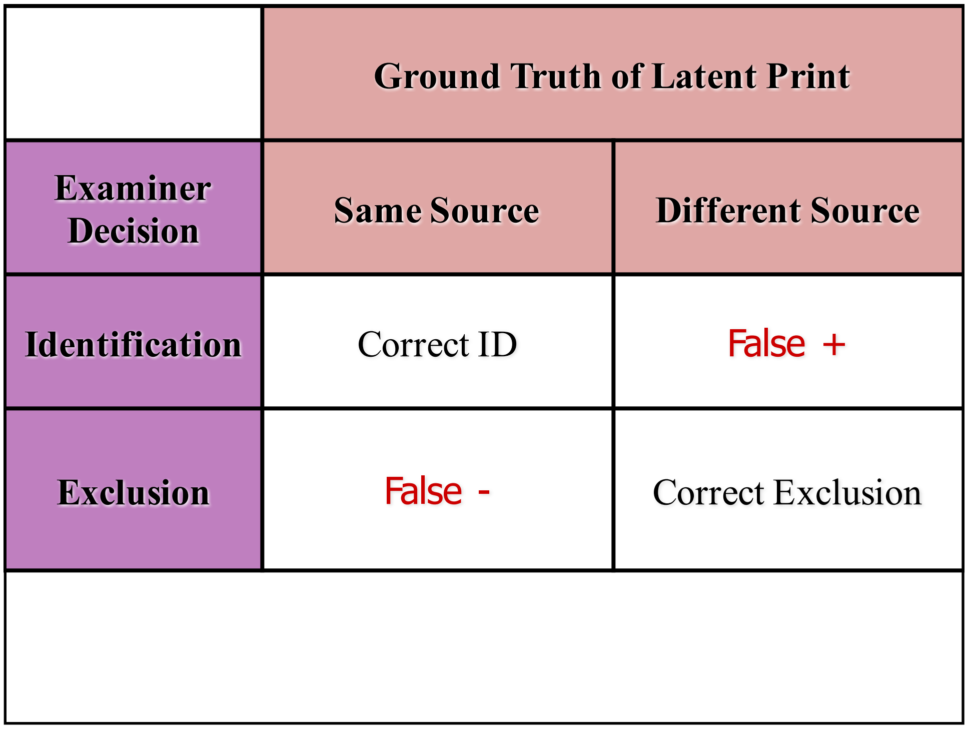
ERROR RATES
(SLIDES FROM G. LANGENBURG)



	Ground Truth of Latent Print	
Examiner Decision	**Same Source**	**Different Source**
Identification	Correct ID	
Exclusion		Correct Exclusion



	Ground Truth of Latent Print	
Examiner Decision	**Same Source**	**Different Source**
Identification	Correct ID	Erroneous ID
Exclusion	Erroneous Exclusion	Correct Exclusion



	Ground Truth of Latent Print	
Examiner Decision	**Same Source**	**Different Source**
Identification	Correct ID	False +
Exclusion	False -	Correct Exclusion

Error rates

	Ground Truth of Latent Print	
Examiner Decision	Same Source	Different Source
Identification	A Correct ID	B False +
Exclusion	C False -	D Correct Exclusion

- False negative (erroneous exclusion):

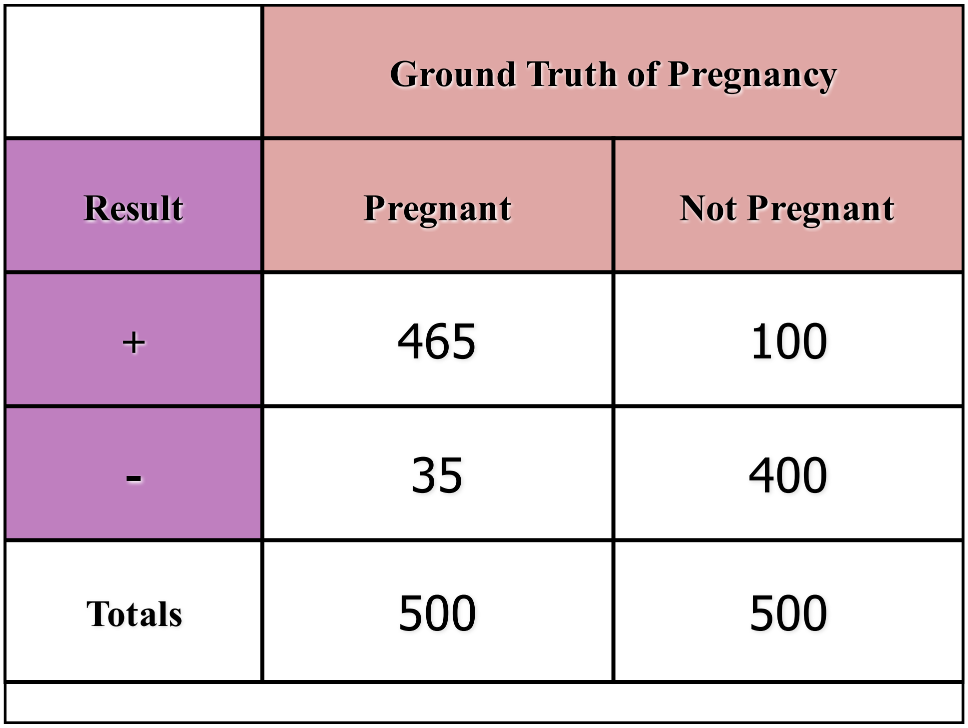
$$\Pr(-|\text{mate}) = \frac{C}{A+C}$$

- False positive (erroneous identification):

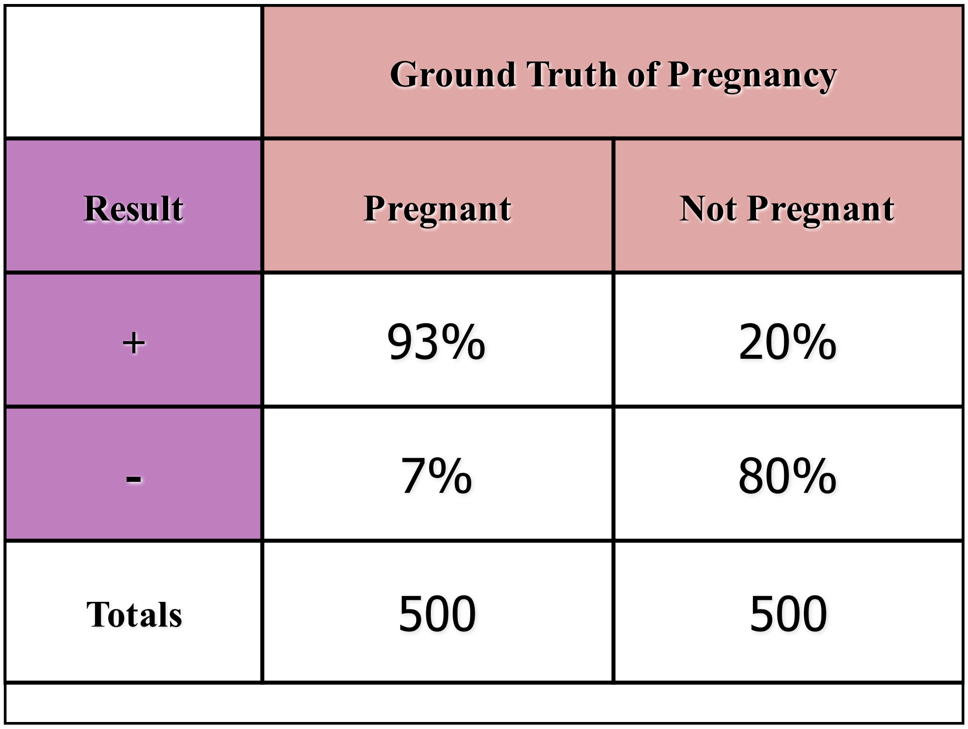
$$\Pr(+|\text{non mate}) = \frac{B}{B+D}$$

Error Rate - Example 1

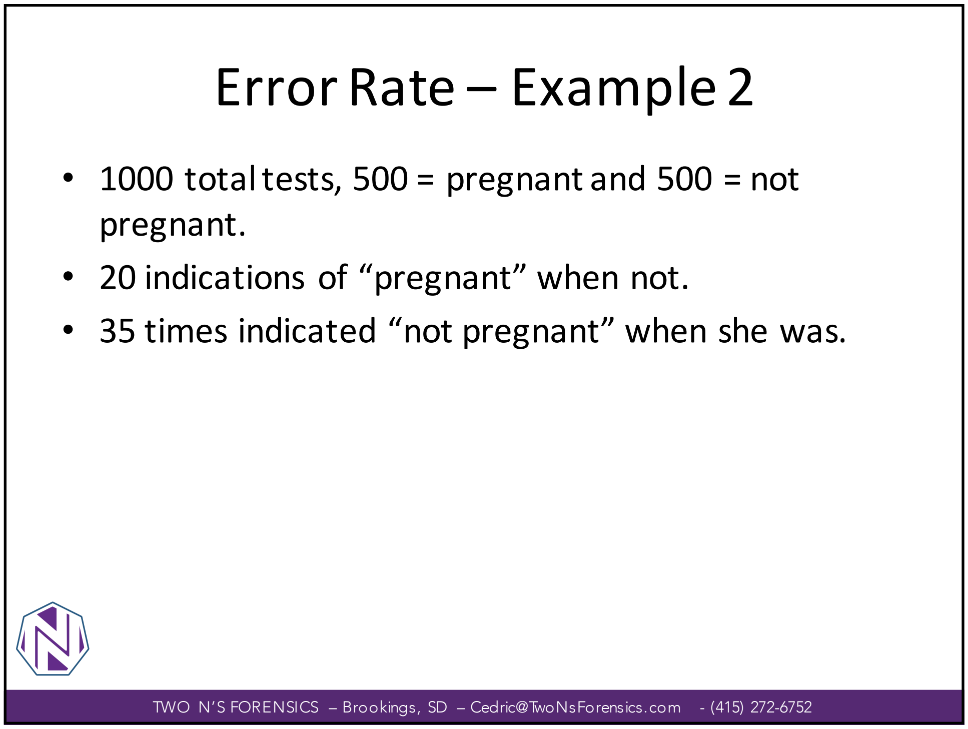
- 1000 total tests, 500 = pregnant and 500 = not pregnant.
- 100 indications of "pregnant" when not.
- 35 times indicated "not pregnant" when she was.



	Ground Truth of Pregnancy	
Result	Pregnant	Not Pregnant
+	465	100
-	35	400
Totals	500	500

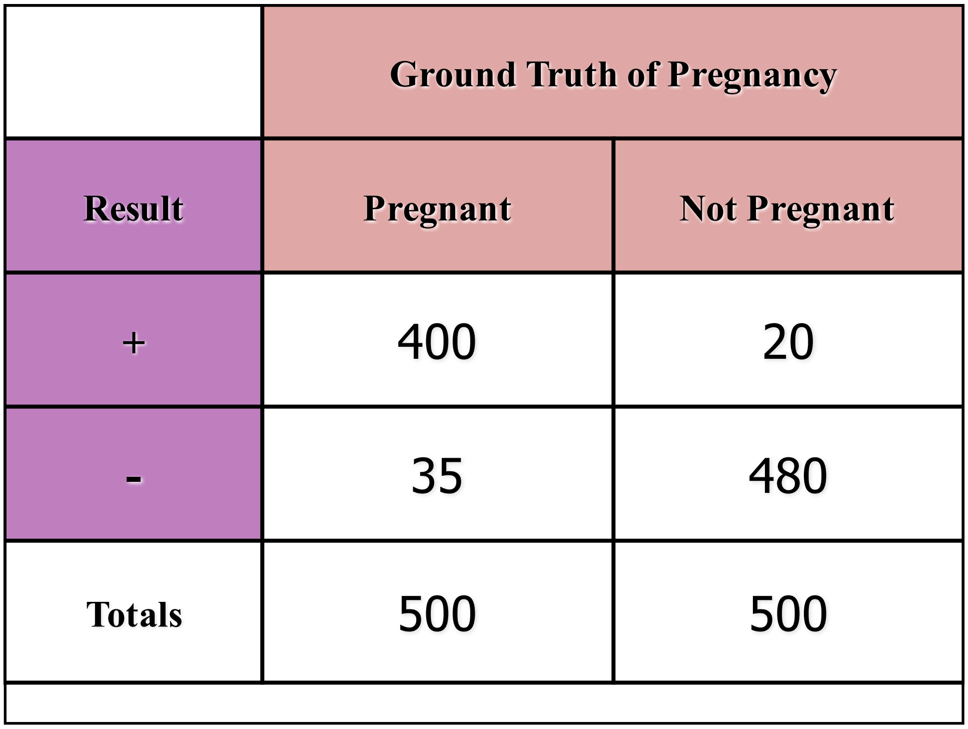


	Ground Truth of Pregnancy	
Result	Pregnant	Not Pregnant
+	93%	20%
-	7%	80%
Totals	500	500

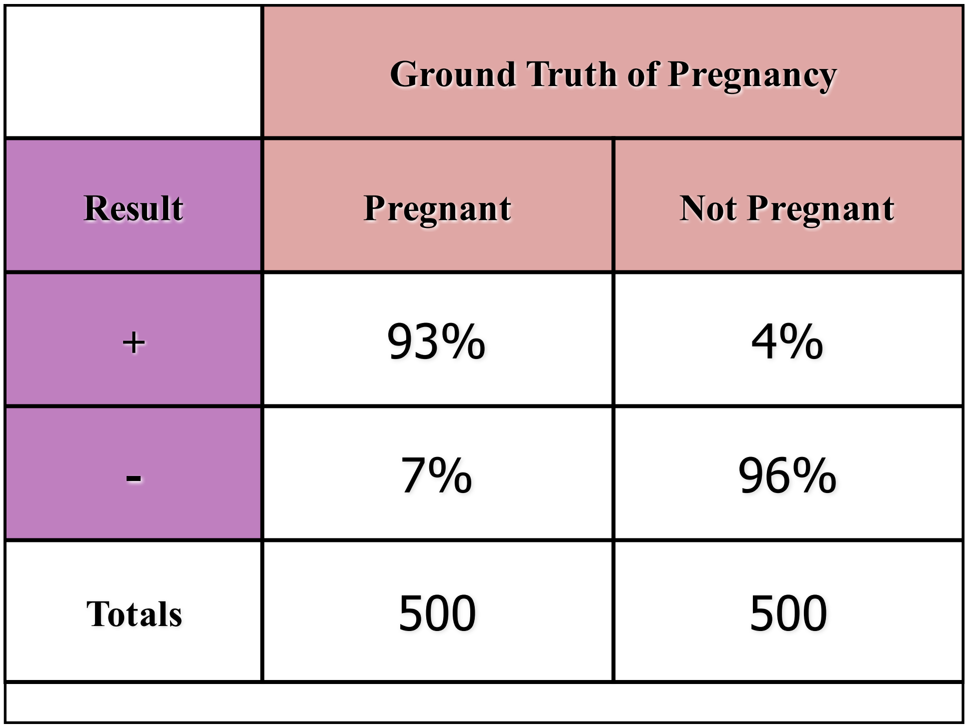


Error Rate – Example 2

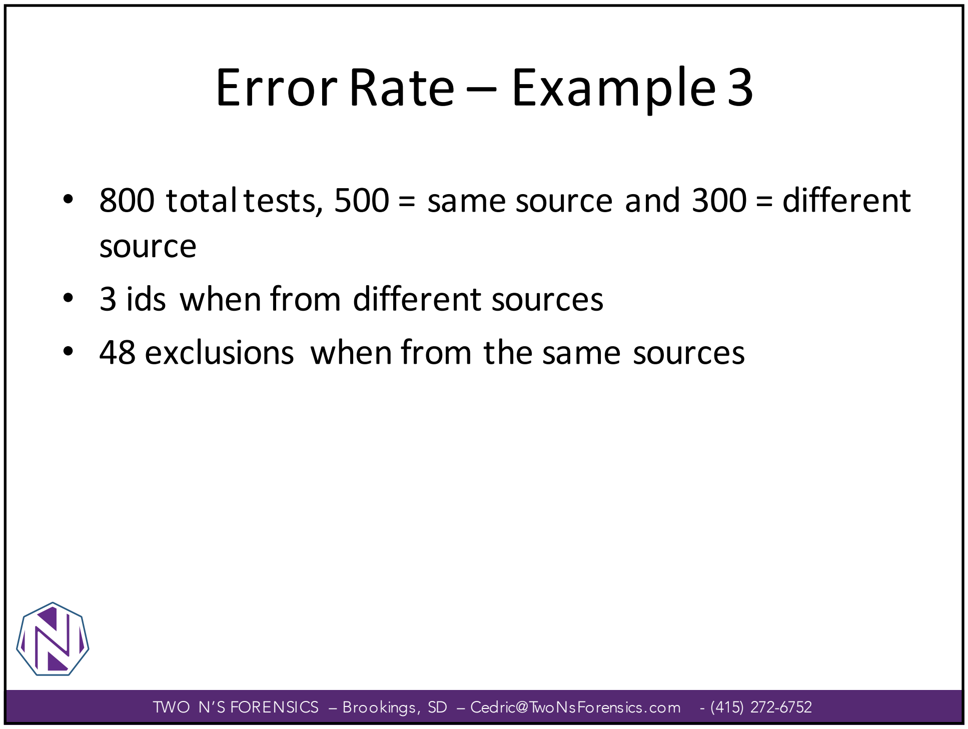
- 1000 total tests, 500 = pregnant and 500 = not pregnant.
- 20 indications of “pregnant” when not.
- 35 times indicated “not pregnant” when she was.



	Ground Truth of Pregnancy	
Result	**Pregnant**	**Not Pregnant**
+	400	20
-	35	480
Totals	500	500



	Ground Truth of Pregnancy	
Result	Pregnant	Not Pregnant
+	93%	4%
-	7%	96%
Totals	500	500



Error Rate – Example 3

- 800 total tests, 500 = same source and 300 = different source
- 3 ids when from different sources
- 48 exclusions when from the same sources

	Ground Truth of Source	
Result	Same Source	Different Source
ID	452	3
EXC	48	297
Totals	500	300

	Ground Truth of Source	
Result	Same Source	Different Source
ID	90.4%	1%
EXC	9.6%	99%
Totals	500	300

Error Rate – Example 4

- 2112 total tests, 1232 = same source and 880 = different source
- 23 ids when from different sources
- 70 exclusions when from the same sources
- 322 inconclusives when from the same source
- 92 inconconclusives when from different sources

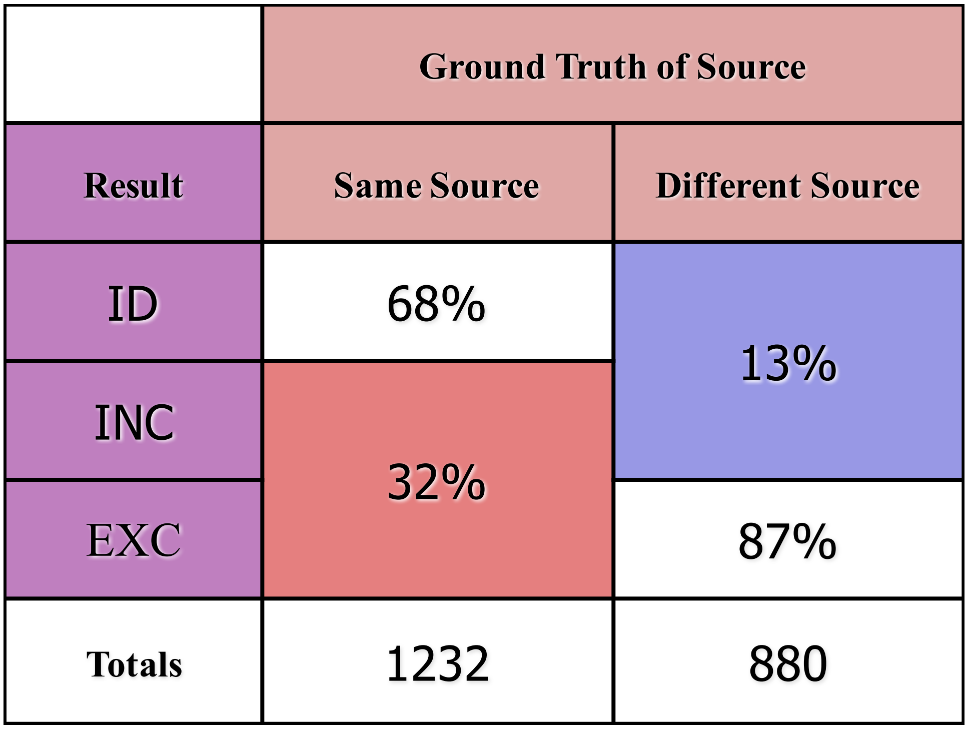
	Ground Truth of Source	
Result	**Same Source**	**Different Source**
ID	840	23
INC	322	92
EXC	70	765
Totals	1232	880

	Ground Truth of Source	
Result	Same Source	Different Source
ID	68%	2.6%
INC	26%	10.4%
EXC	5.7%	87%
Totals	1232	880

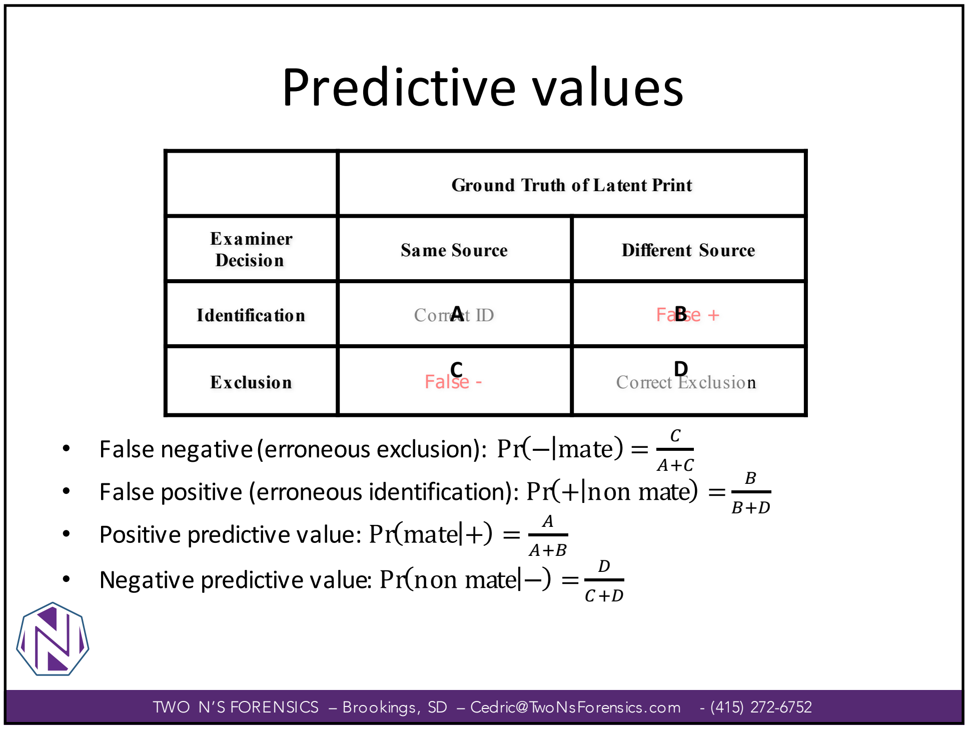
	Ground Truth of Source	
Result	Same Source	Different Source
ID	840	23
EXC	70	765
Totals	910	788

	Ground Truth of Source	
Result	**Same Source**	**Different Source**
ID	92%	3%
EXC	8%	97%
Totals	910	788

	Ground Truth of Source	
Result	**Same Source**	**Different Source**
ID	840	23
INC	322	92
EXC	70	765
Totals	1232	880



	Ground Truth of Source	
Result	Same Source	Different Source
ID	68%	13%
INC	32%	
EXC		87%
Totals	1232	880

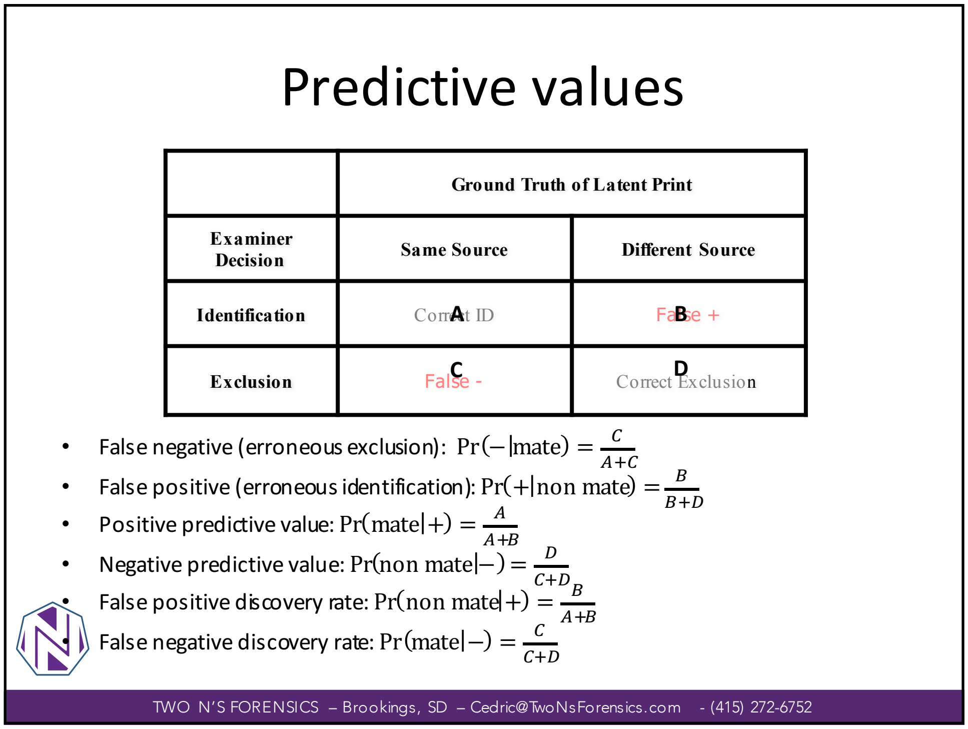


Predictive values

	Ground Truth of Latent Print	
Examiner Decision	Same Source	Different Source
Identification	A Correct ID	B False +
Exclusion	C False -	D Correct Exclusion

- False negative (erroneous exclusion): $\Pr(-|\text{mate}) = \frac{C}{A+C}$
- False positive (erroneous identification): $\Pr(+|\text{non mate}) = \frac{B}{B+D}$
- Positive predictive value: $\Pr(\text{mate}|+) = \frac{A}{A+B}$
- Negative predictive value: $\Pr(\text{non mate}|-) = \frac{D}{C+D}$

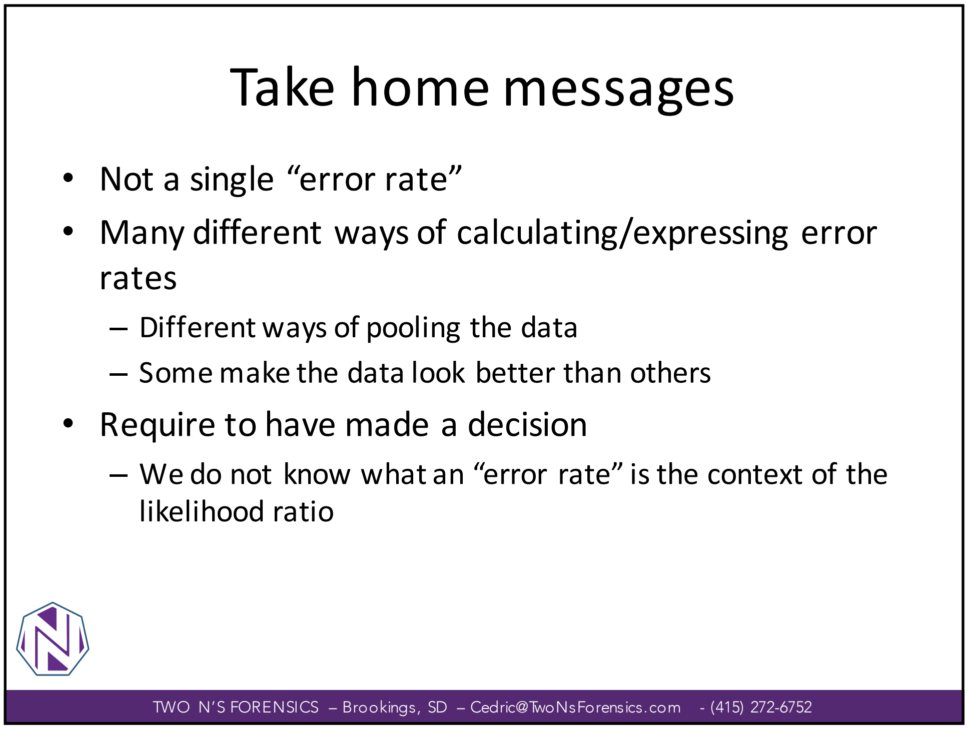
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



Predictive values

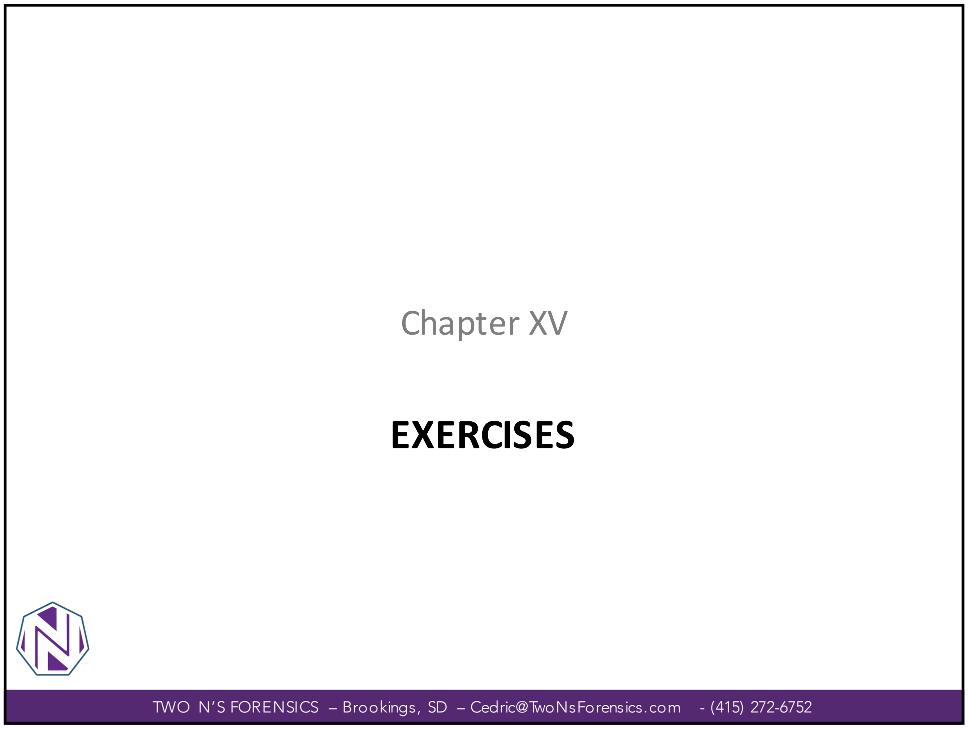
	Ground Truth of Latent Print	
Examiner Decision	Same Source	Different Source
Identification	A Correct ID	B False +
Exclusion	C False -	D Correct Exclusion

- False negative (erroneous exclusion): $\Pr(-|\text{mate}) = \frac{C}{A+C}$
- False positive (erroneous identification): $\Pr(+|\text{non mate}) = \frac{B}{B+D}$
- Positive predictive value: $\Pr(\text{mate}|+) = \frac{A}{A+B}$
- Negative predictive value: $\Pr(\text{non mate}|-) = \frac{D}{C+D}$
- False positive discovery rate: $\Pr(\text{non mate}|+) = \frac{B}{A+B}$
- False negative discovery rate: $\Pr(\text{mate}|-) = \frac{C}{C+D}$



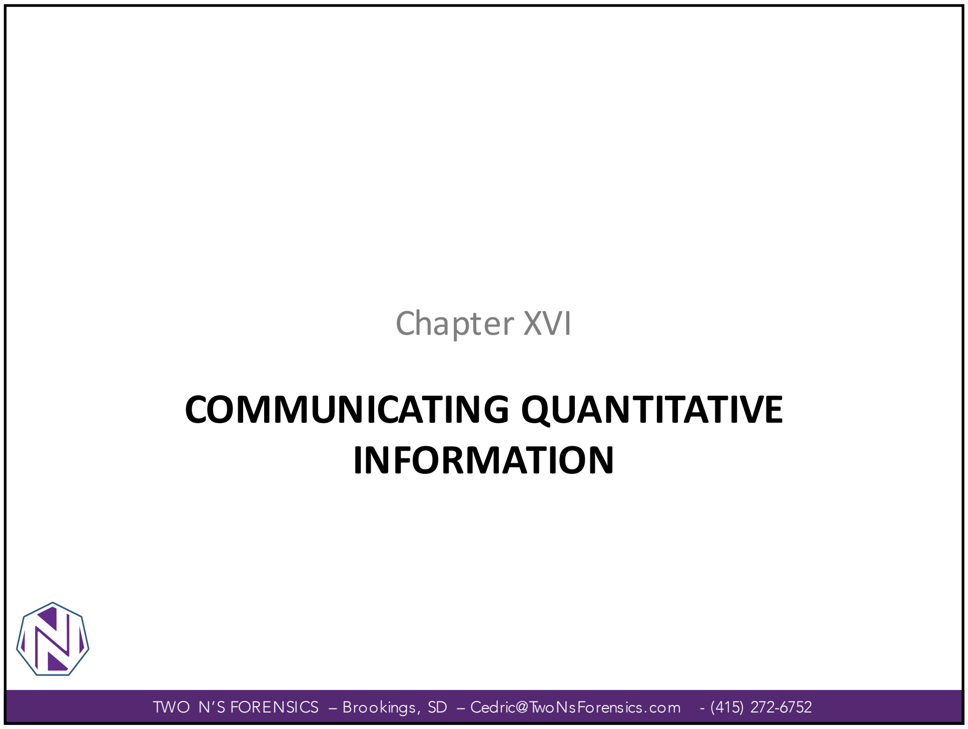
Take home messages

- Not a single “error rate”
- Many different ways of calculating/expressing error rates
  - Different ways of pooling the data
  - Some make the data look better than others
- Require to have made a decision
  - We do not know what an “error rate” is the context of the likelihood ratio



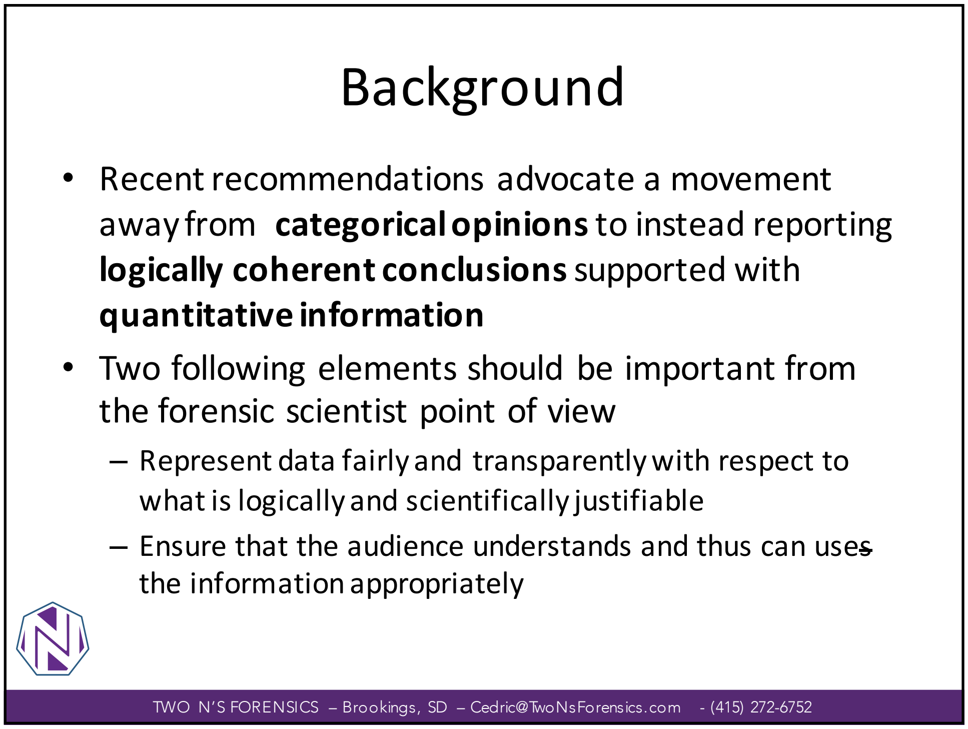
Chapter XV

EXERCISES



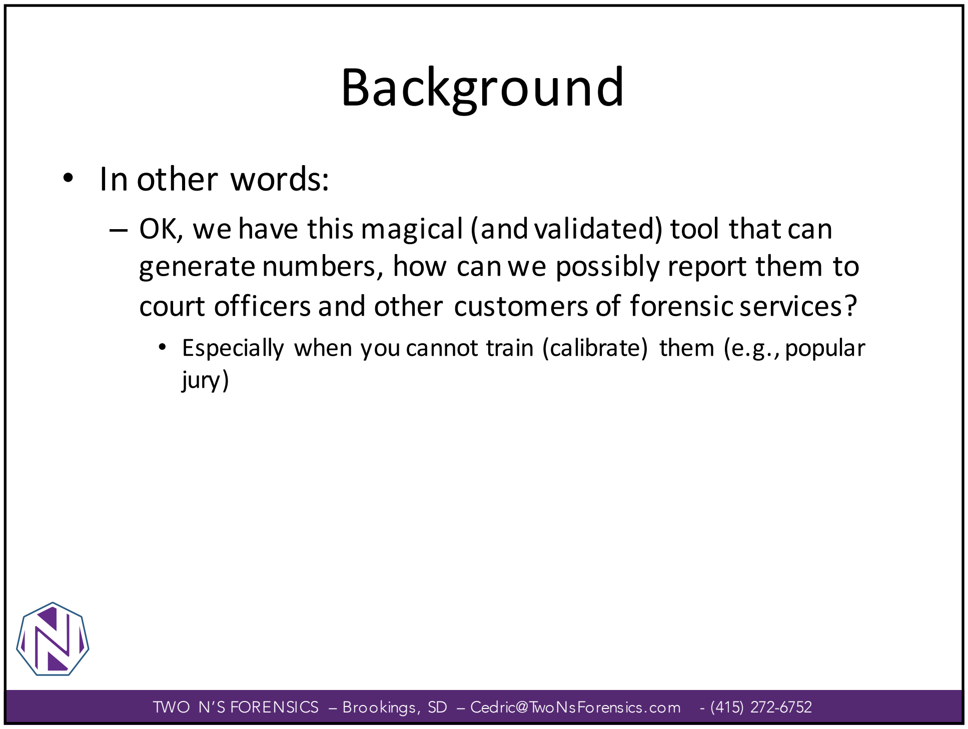
Chapter XVI

COMMUNICATING QUANTITATIVE INFORMATION



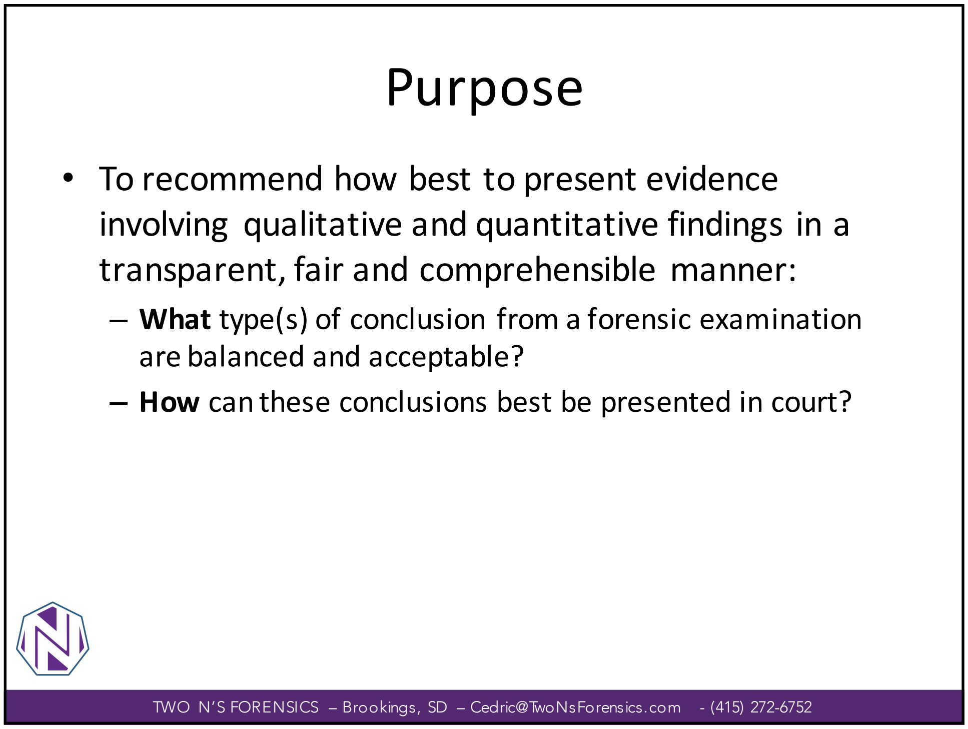
Background

- Recent recommendations advocate a movement away from **categorical opinions** to instead reporting **logically coherent conclusions** supported with **quantitative information**
- Two following elements should be important from the forensic scientist point of view
  - Represent data fairly and transparently with respect to what is logically and scientifically justifiable
  - Ensure that the audience understands and thus can use~~s~~ the information appropriately



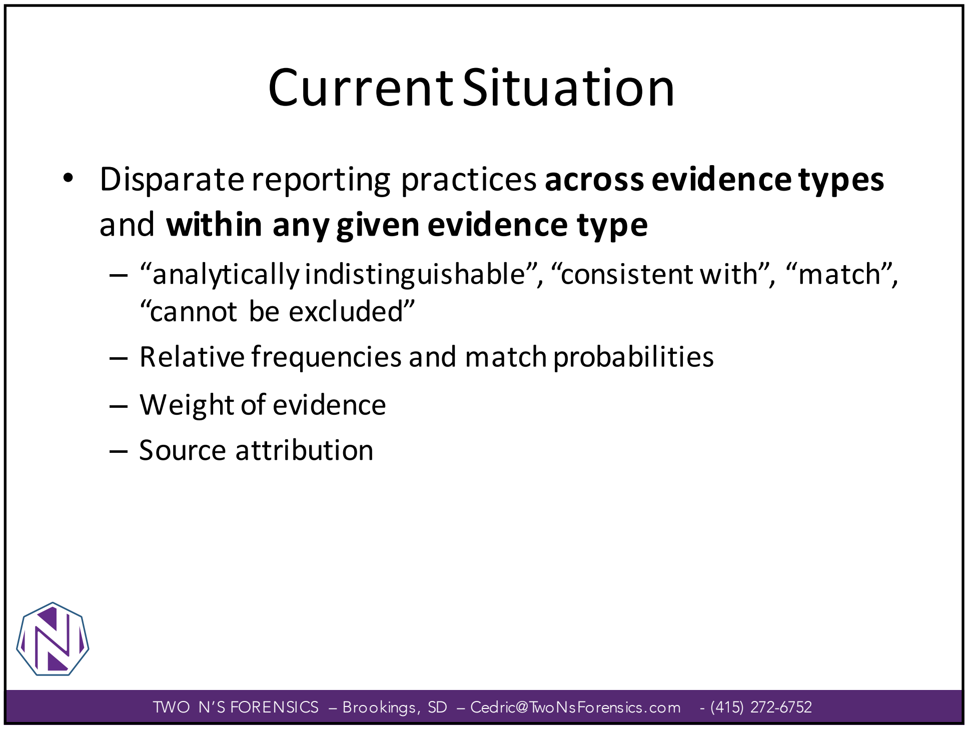
Background

- In other words:
  - OK, we have this magical (and validated) tool that can generate numbers, how can we possibly report them to court officers and other customers of forensic services?
    - Especially when you cannot train (calibrate) them (e.g., popular jury)



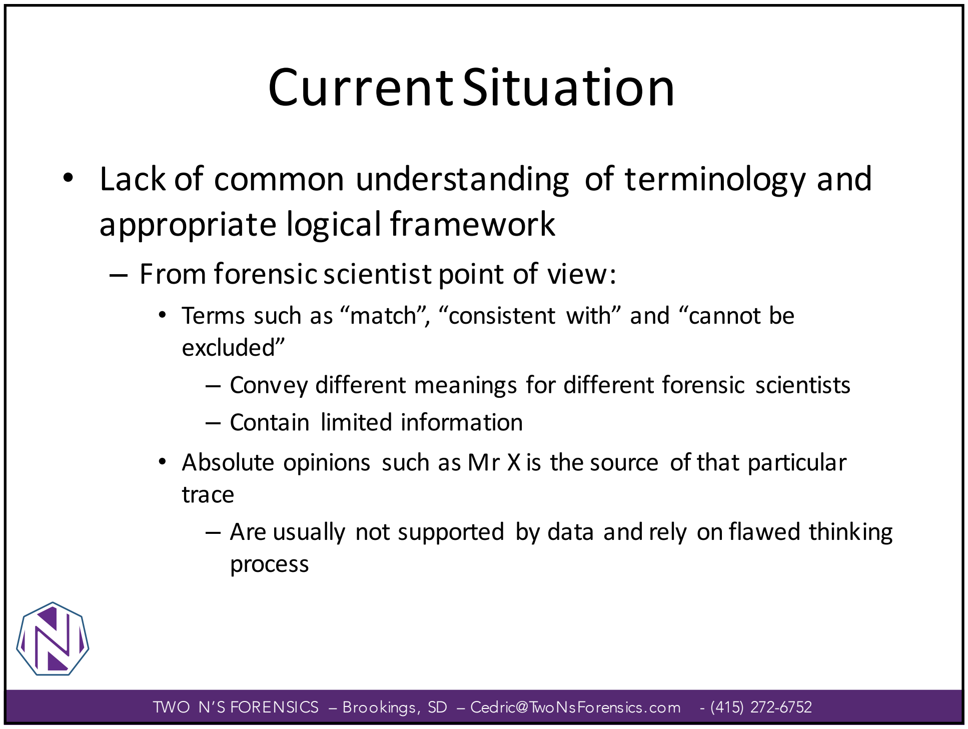
Purpose

- To recommend how best to present evidence involving qualitative and quantitative findings in a transparent, fair and comprehensible manner:
  - **What** type(s) of conclusion from a forensic examination are balanced and acceptable?
  - **How** can these conclusions best be presented in court?



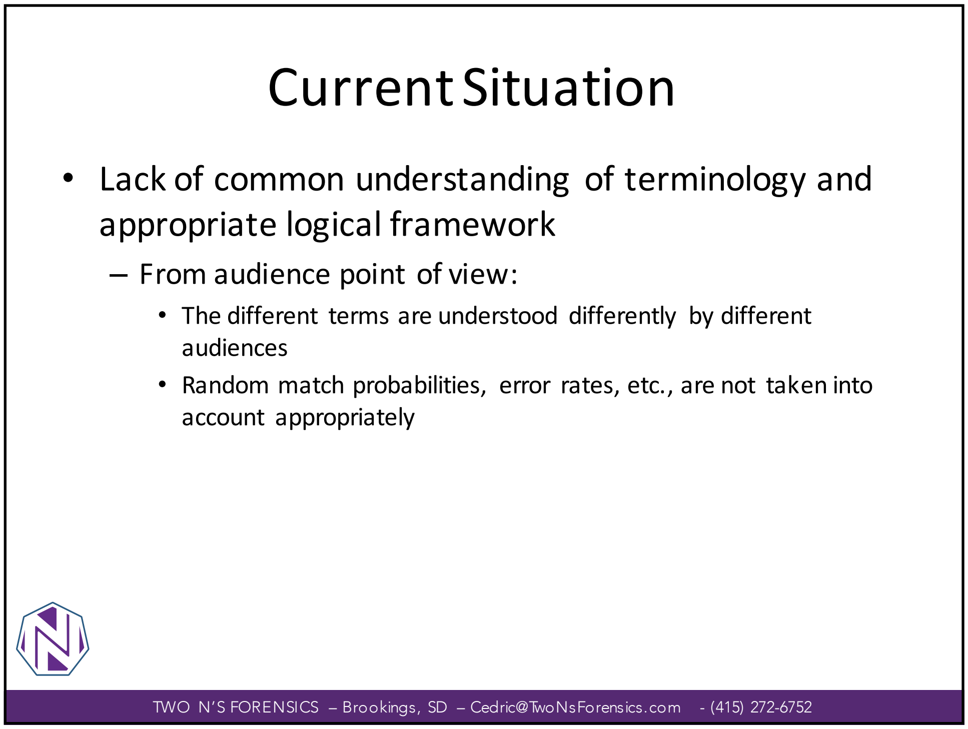
Current Situation

- Disparate reporting practices **across evidence types** and **within any given evidence type**
  - “analytically indistinguishable”, “consistent with”, “match”, “cannot be excluded”
  - Relative frequencies and match probabilities
  - Weight of evidence
  - Source attribution



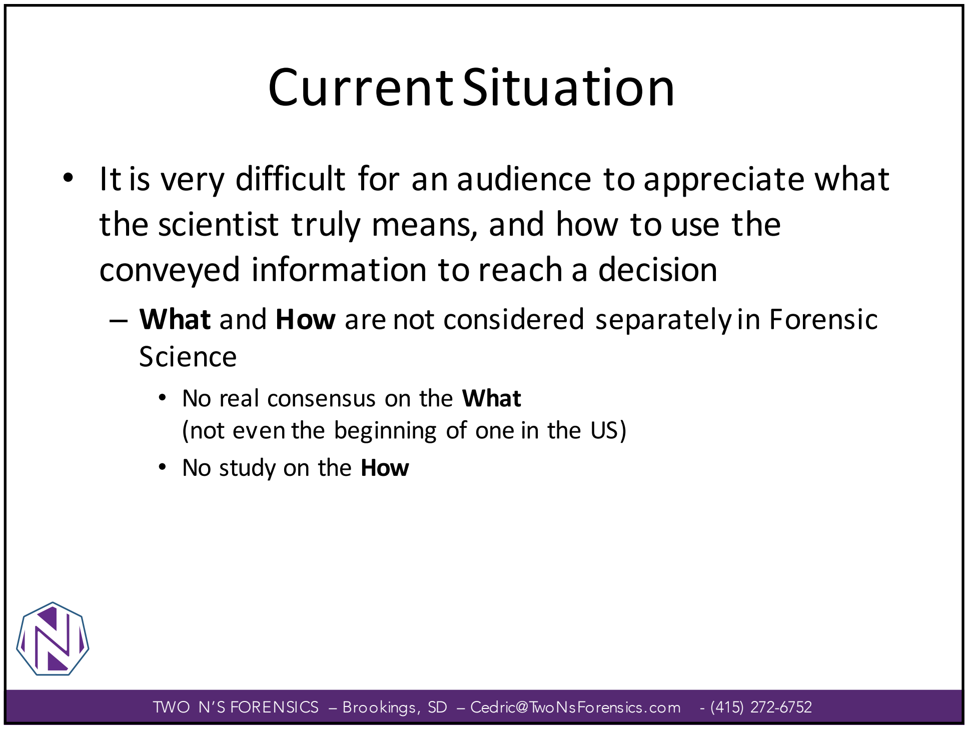
Current Situation

- Lack of common understanding of terminology and appropriate logical framework
  – From forensic scientist point of view:
    - Terms such as “match”, “consistent with” and “cannot be excluded”
      – Convey different meanings for different forensic scientists
      – Contain limited information
    - Absolute opinions such as Mr X is the source of that particular trace
      – Are usually not supported by data and rely on flawed thinking process



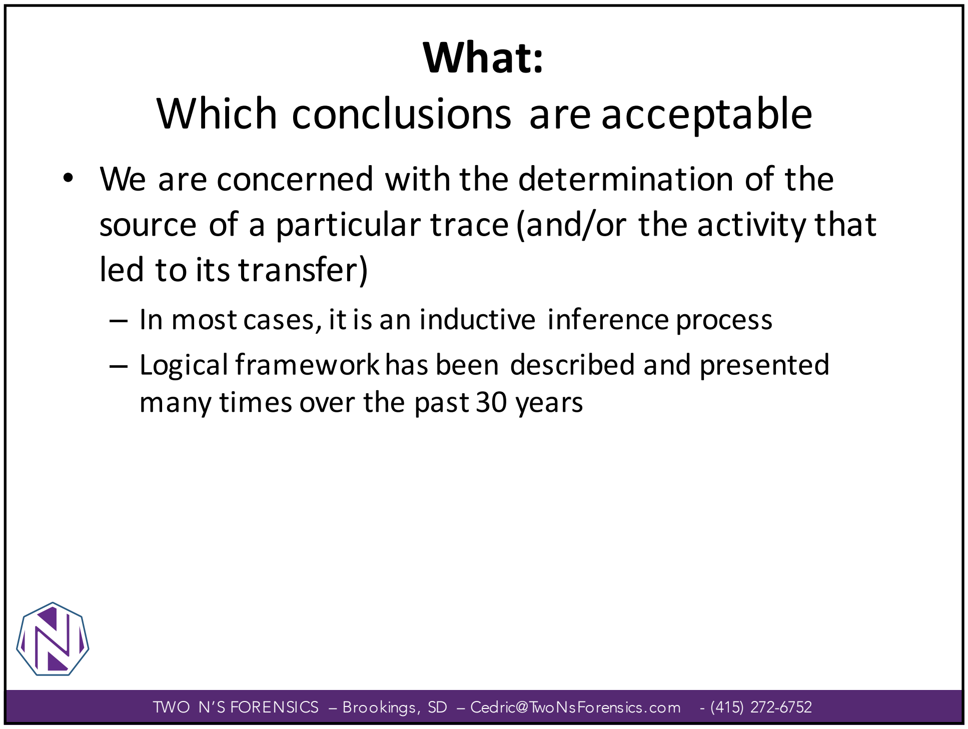
Current Situation

- Lack of common understanding of terminology and appropriate logical framework
  – From audience point of view:
    - The different terms are understood differently by different audiences
    - Random match probabilities, error rates, etc., are not taken into account appropriately



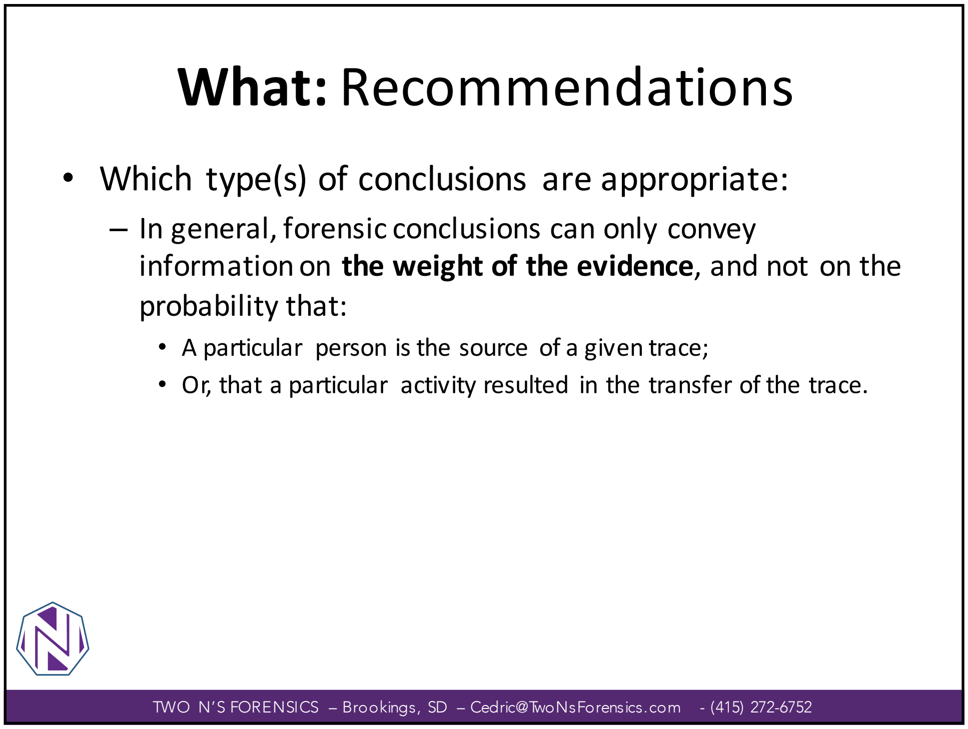
Current Situation

- It is very difficult for an audience to appreciate what the scientist truly means, and how to use the conveyed information to reach a decision
  – **What** and **How** are not considered separately in Forensic Science
    - No real consensus on the **What**
      (not even the beginning of one in the US)
    - No study on the **How**



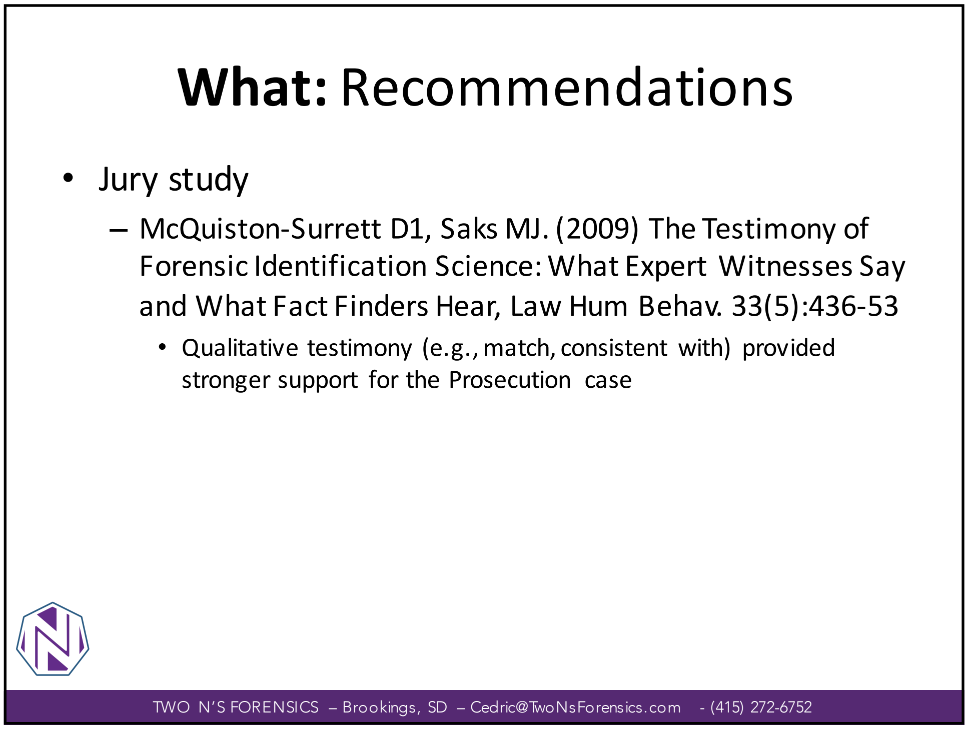
What:
Which conclusions are acceptable

- We are concerned with the determination of the source of a particular trace (and/or the activity that led to its transfer)
  – In most cases, it is an inductive inference process
  – Logical framework has been described and presented many times over the past 30 years



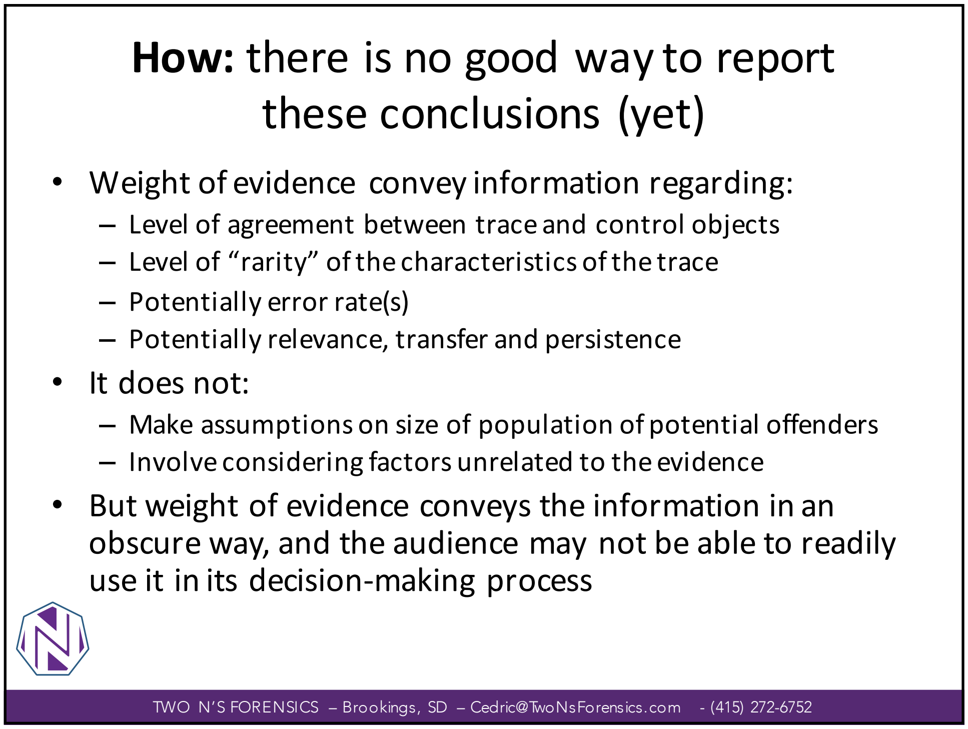
What: Recommendations

- Which type(s) of conclusions are appropriate:
  - In general, forensic conclusions can only convey information on **the weight of the evidence**, and not on the probability that:
    - A particular person is the source of a given trace;
    - Or, that a particular activity resulted in the transfer of the trace.



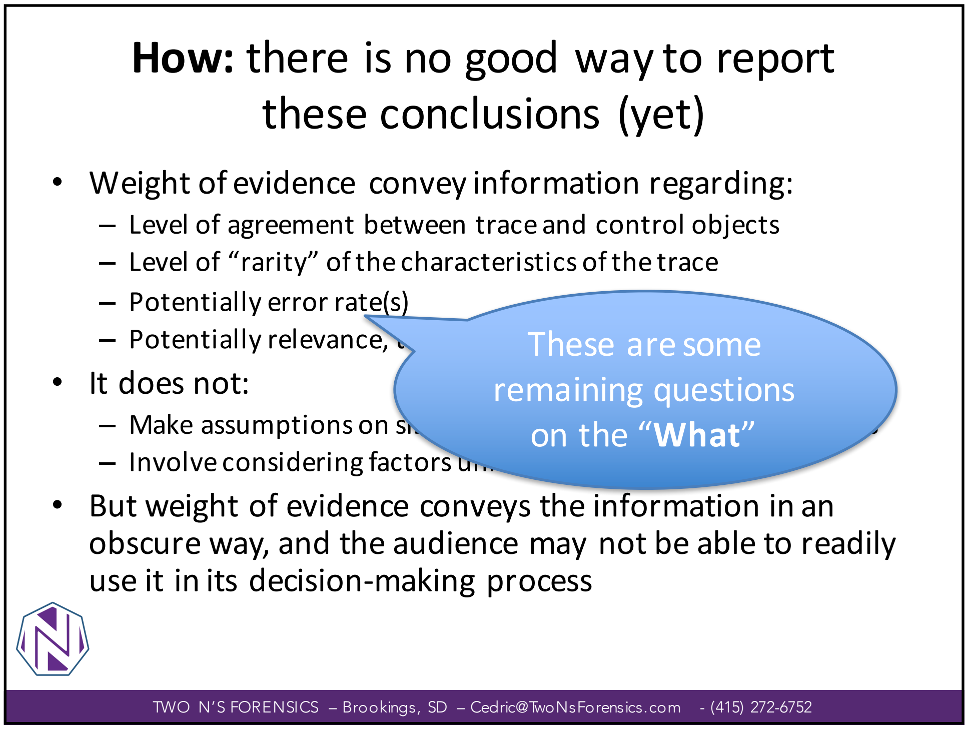
What: Recommendations

- Jury study
  - McQuiston-Surrett D1, Saks MJ. (2009) The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Fact Finders Hear, Law Hum Behav. 33(5):436-53
    - Qualitative testimony (e.g., match, consistent with) provided stronger support for the Prosecution case



How: there is no good way to report these conclusions (yet)

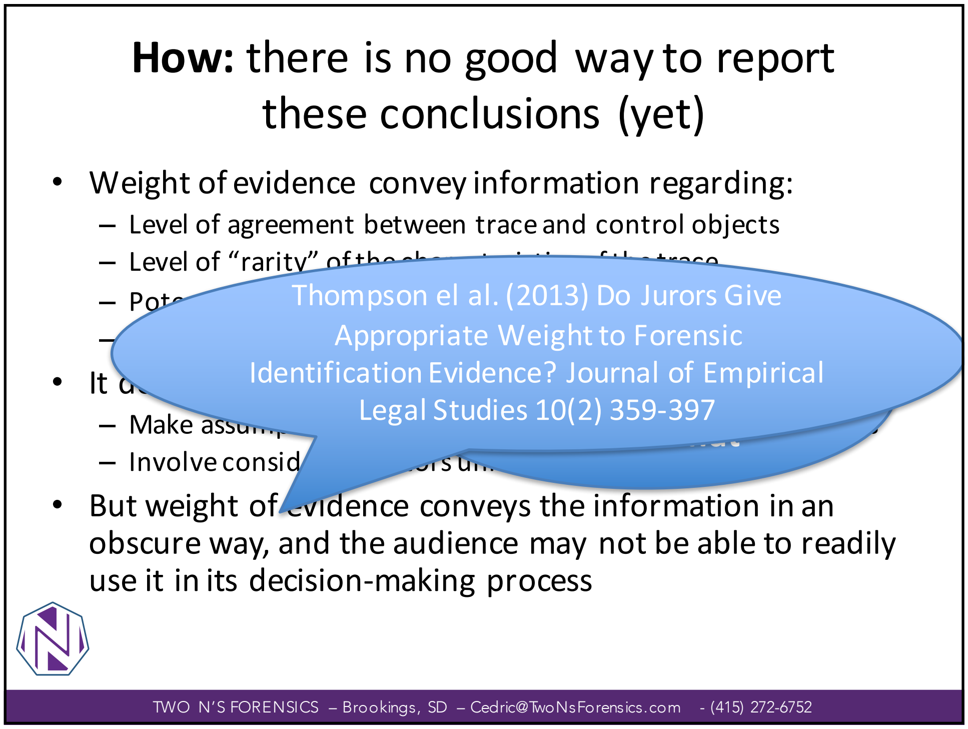
- Weight of evidence convey information regarding:
  - Level of agreement between trace and control objects
  - Level of "rarity" of the characteristics of the trace
  - Potentially error rate(s)
  - Potentially relevance, transfer and persistence
- It does not:
  - Make assumptions on size of population of potential offenders
  - Involve considering factors unrelated to the evidence
- But weight of evidence conveys the information in an obscure way, and the audience may not be able to readily use it in its decision-making process



How: there is no good way to report these conclusions (yet)

- Weight of evidence convey information regarding:
  - Level of agreement between trace and control objects
  - Level of "rarity" of the characteristics of the trace
  - Potentially error rate(s)
  - Potentially relevance, t
- It does not:
  - Make assumptions on si
  - Involve considering factors un
- But weight of evidence conveys the information in an obscure way, and the audience may not be able to readily use it in its decision-making process

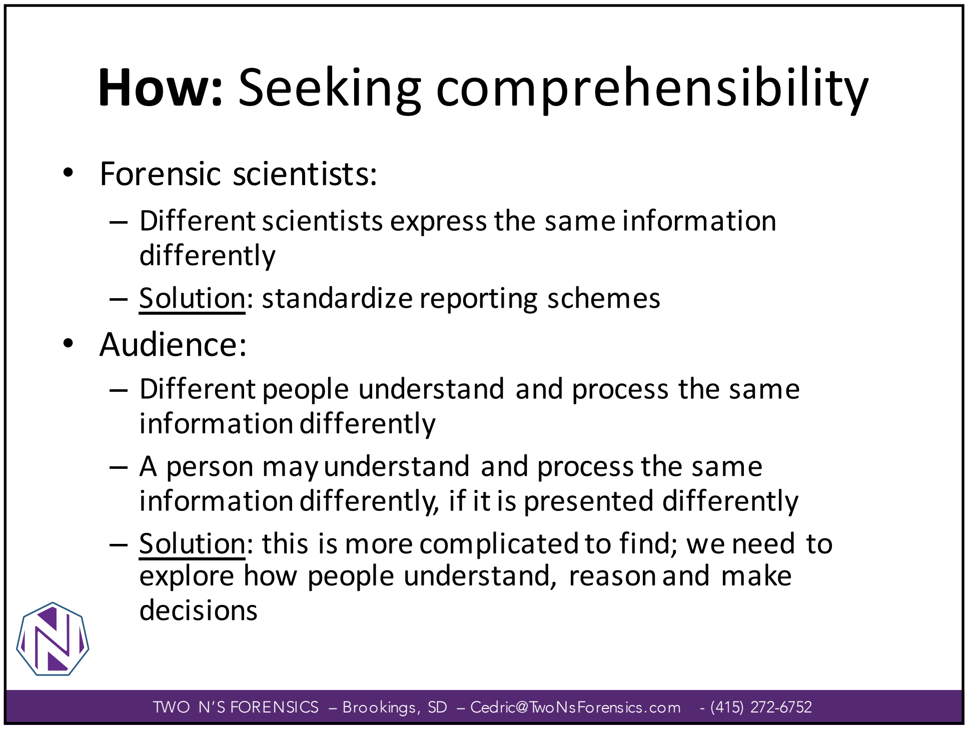
These are some remaining questions on the "**What**"



How: there is no good way to report these conclusions (yet)

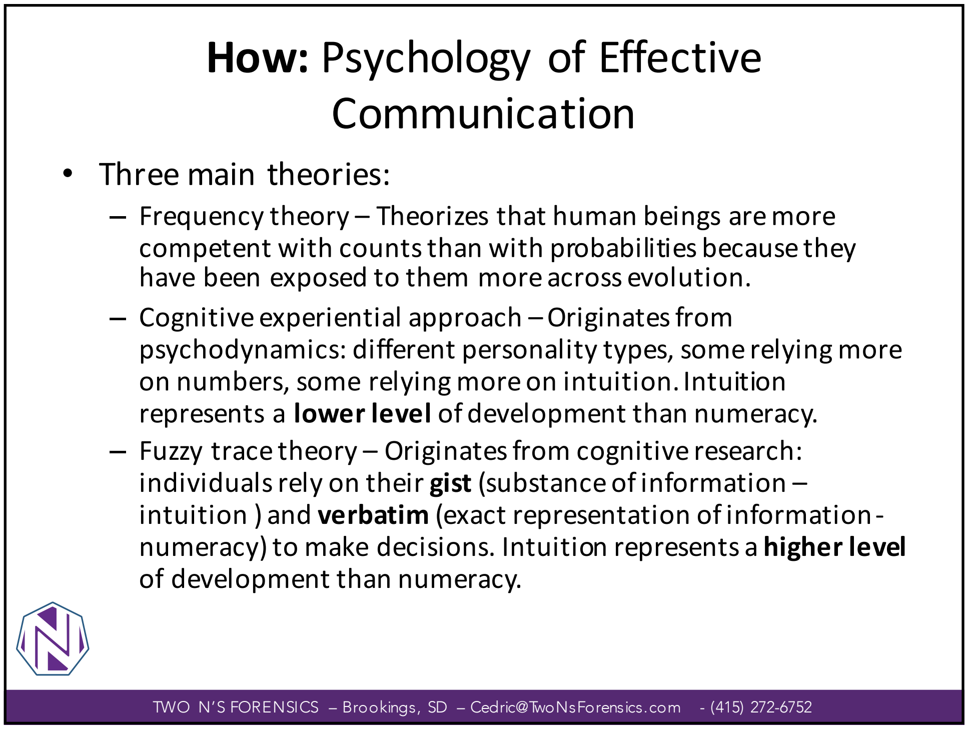
- Weight of evidence convey information regarding:
  - Level of agreement between trace and control objects
  - Level of "rarity" of the characteristics of the trace
  - Pote
  -
- It d
  - Make assum
  - Involve consid ors u
- But weight of evidence conveys the information in an obscure way, and the audience may not be able to readily use it in its decision-making process

Thompson el al. (2013) Do Jurors Give Appropriate Weight to Forensic Identification Evidence? Journal of Empirical Legal Studies 10(2) 359-397



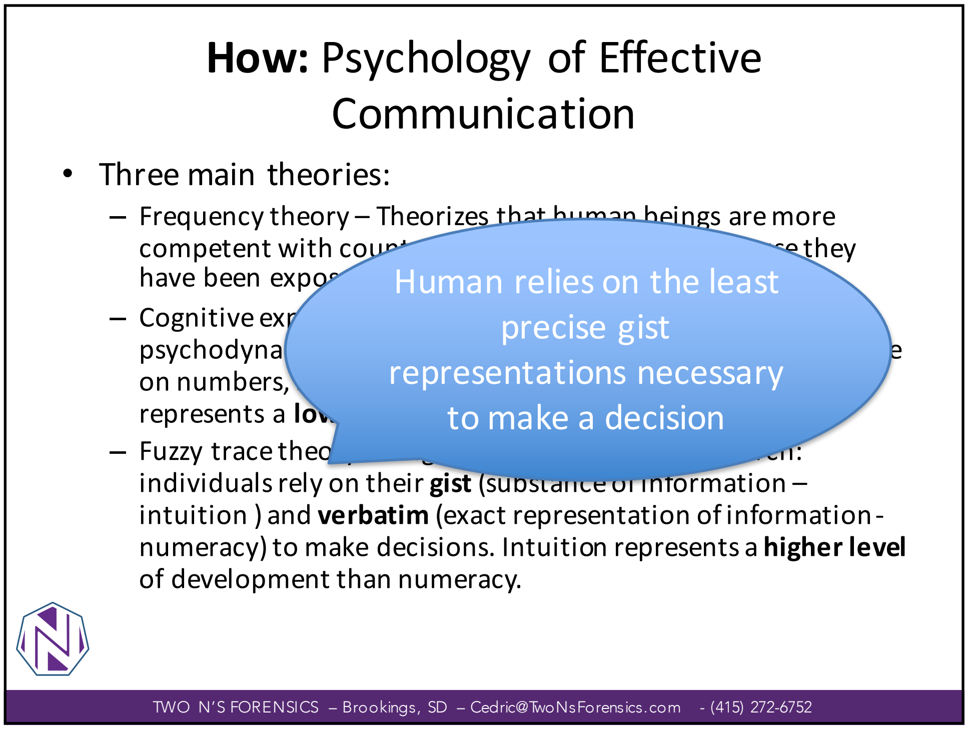
How: Seeking comprehensibility

- Forensic scientists:
  - Different scientists express the same information differently
  - Solution: standardize reporting schemes
- Audience:
  - Different people understand and process the same information differently
  - A person may understand and process the same information differently, if it is presented differently
  - Solution: this is more complicated to find; we need to explore how people understand, reason and make decisions



How: Psychology of Effective Communication

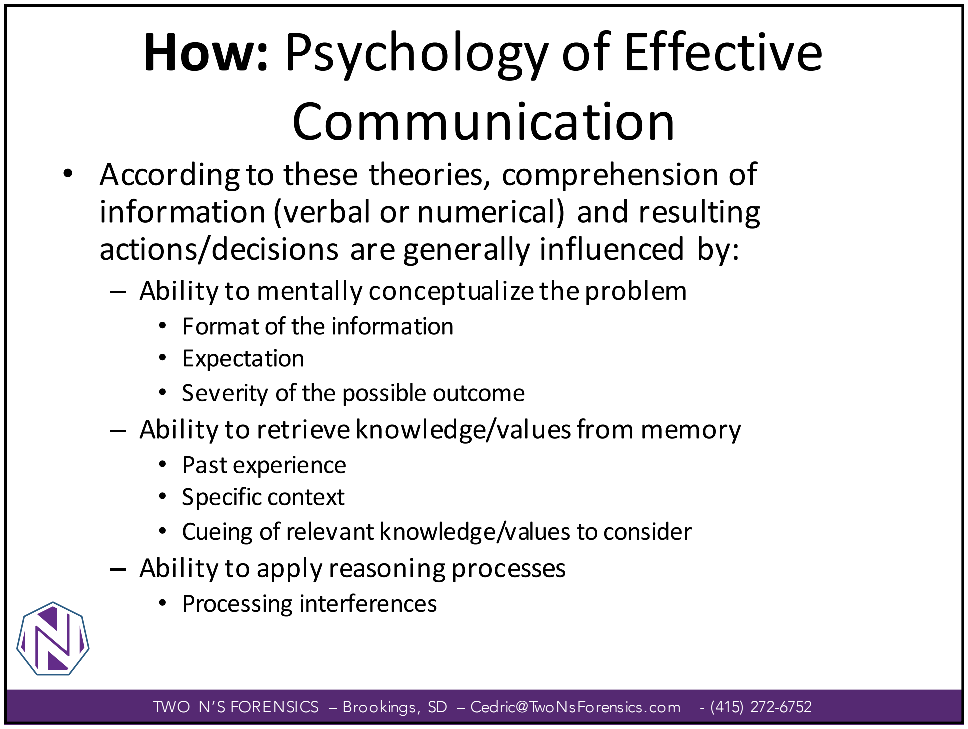
- Three main theories:
  - Frequency theory – Theorizes that human beings are more competent with counts than with probabilities because they have been exposed to them more across evolution.
  - Cognitive experiential approach – Originates from psychodynamics: different personality types, some relying more on numbers, some relying more on intuition. Intuition represents a **lower level** of development than numeracy.
  - Fuzzy trace theory – Originates from cognitive research: individuals rely on their **gist** (substance of information – intuition) and **verbatim** (exact representation of information - numeracy) to make decisions. Intuition represents a **higher level** of development than numeracy.



How: Psychology of Effective Communication

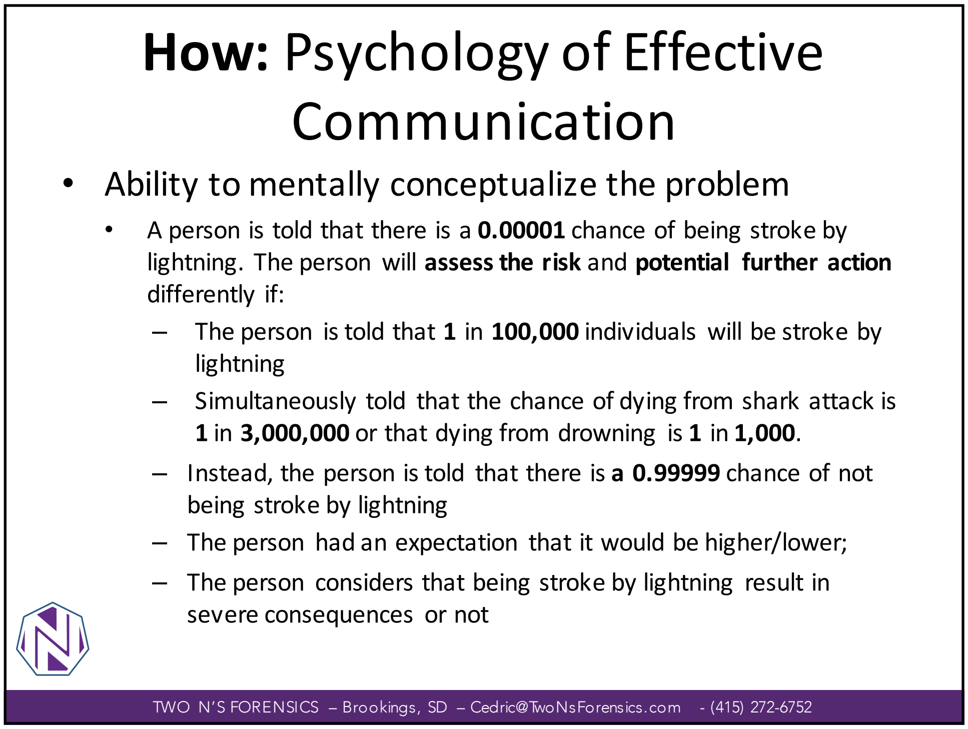
- Three main theories:
  - Frequency theory – Theorizes that human beings are more competent with coun se they have been expos
  - Cognitive exp psychodyna e on numbers, represents a **lo**
  - Fuzzy trace theo ch: individuals rely on their **gist** (substance of information – intuition) and **verbatim** (exact representation of information - numeracy) to make decisions. Intuition represents a **higher level** of development than numeracy.

Human relies on the least precise gist representations necessary to make a decision



How: Psychology of Effective Communication

- According to these theories, comprehension of information (verbal or numerical) and resulting actions/decisions are generally influenced by:
  - – Ability to mentally conceptualize the problem
    - Format of the information
    - Expectation
    - Severity of the possible outcome
  - – Ability to retrieve knowledge/values from memory
    - Past experience
    - Specific context
    - Cueing of relevant knowledge/values to consider
  - – Ability to apply reasoning processes
    - Processing interferences



How: Psychology of Effective Communication

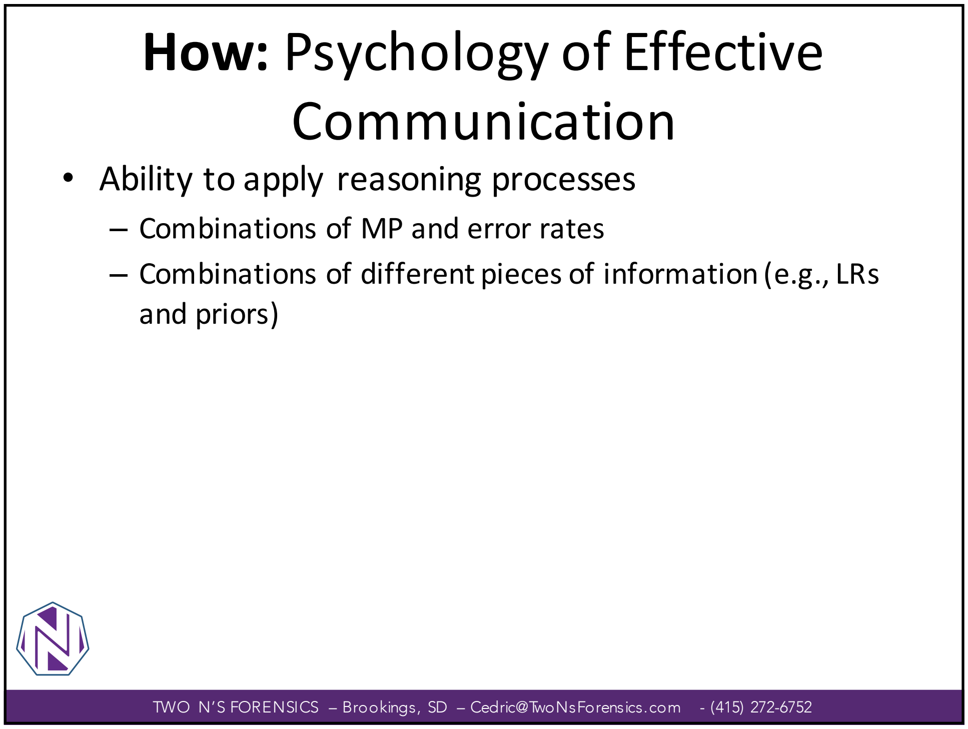
- Ability to mentally conceptualize the problem
  - A person is told that there is a **0.00001** chance of being stroke by lightning. The person will **assess the risk** and **potential further action** differently if:
    - – The person is told that **1** in **100,000** individuals will be stroke by lightning
    - – Simultaneously told that the chance of dying from shark attack is **1** in **3,000,000** or that dying from drowning is **1** in **1,000**.
    - – Instead, the person is told that there is **a 0.99999** chance of not being stroke by lightning
    - – The person had an expectation that it would be higher/lower;
    - – The person considers that being stroke by lightning result in severe consequences or not

How: Psychology of Effective Communication

- Ability to mentally conceptualize the problem
  - A person is told that there is a **very low** chance of being stroke by lightning. The person will **assess the risk** and **potential further action** differently if:
    - The person is told that **few** individuals will be stroke by lightning – or that **a few** individuals will be stroke by lightning
    - Simultaneously told that there is **less chance** of dying from shark attack or **more chance** of dying from drowning.
    - Instead, the person is told that there is **an extremely high** chance of **not** being stroke by lightning
    - The person had an expectation that it would be **moderate/high**;
    - The person considers that being stroke by lightning result in severe consequences or not

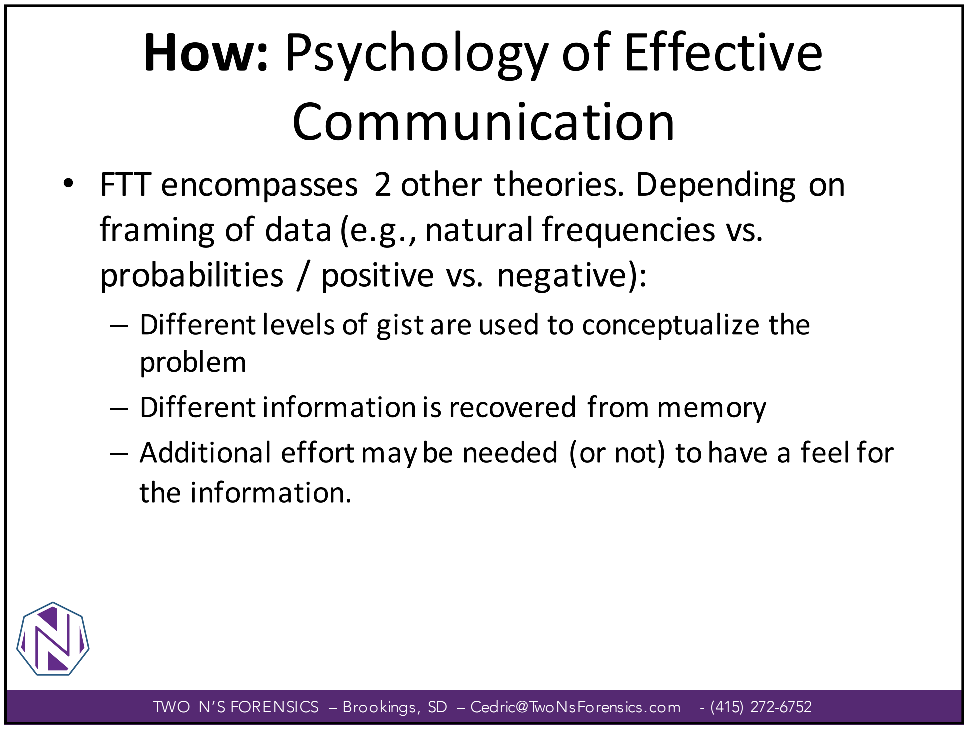
How: Psychology of Effective Communication

- Ability to retrieve knowledge/values from memory
  - A person is told that there is a **0.00001** chance of being stroke by lightning. The person will **assess the risk** and **potential further action** differently if:
    - The person does (not) know anybody who had been stroke
    - The person knows/is informed that most individuals survive
    - The person realise that he/she never walks on golf courses (or swim) during thunderstorms.



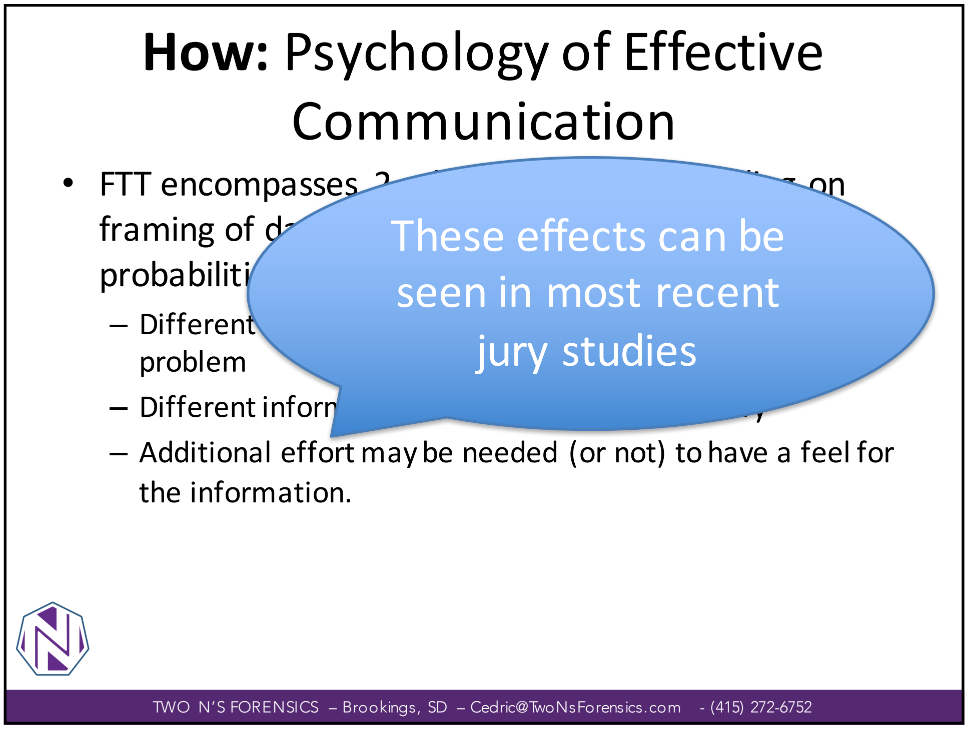
How: Psychology of Effective Communication

- Ability to apply reasoning processes
  - Combinations of MP and error rates
  - Combinations of different pieces of information (e.g., LRs and priors)



How: Psychology of Effective Communication

- FTT encompasses 2 other theories. Depending on framing of data (e.g., natural frequencies vs. probabilities / positive vs. negative):
  - Different levels of gist are used to conceptualize the problem
  - Different information is recovered from memory
  - Additional effort may be needed (or not) to have a feel for the information.

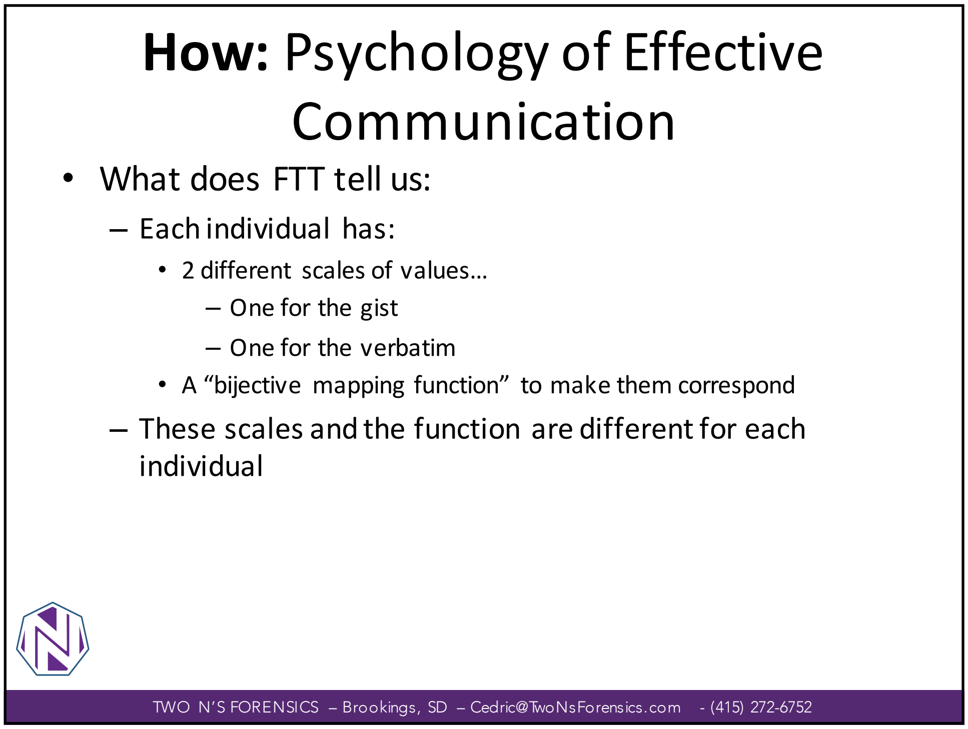


How: Psychology of Effective Communication

- FTT encompasses 2 … on framing of d… probabiliti…
  - Different … problem
  - Different inforn…
  - Additional effort may be needed (or not) to have a feel for the information.

These effects can be seen in most recent jury studies

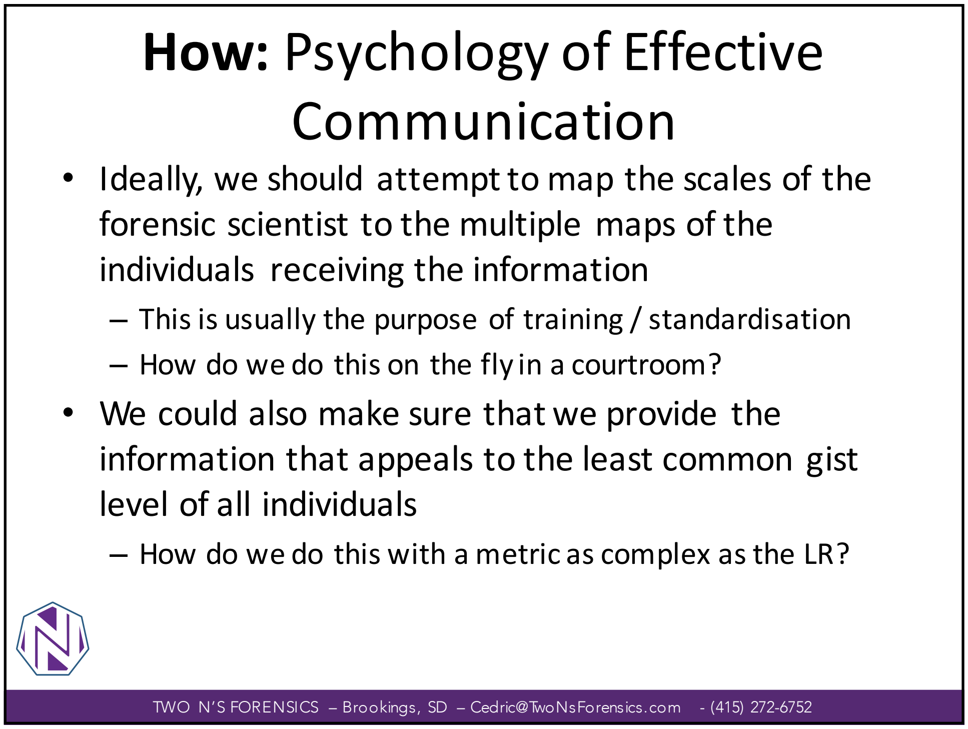
TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



How: Psychology of Effective Communication

- What does FTT tell us:
  - Each individual has:
    - 2 different scales of values…
      - One for the gist
      - One for the verbatim
    - A "bijective mapping function" to make them correspond
  - These scales and the function are different for each individual

TWO N'S FORENSICS – Brookings, SD – Cedric@TwoNsForensics.com - (415) 272-6752



How: Psychology of Effective Communication

- Ideally, we should attempt to map the scales of the forensic scientist to the multiple maps of the individuals receiving the information
  - This is usually the purpose of training / standardisation
  - How do we do this on the fly in a courtroom?
- We could also make sure that we provide the information that appeals to the least common gist level of all individuals
  - How do we do this with a metric as complex as the LR?

Workshop on Statistics and Probability in Forensics Science

Cedric Neumann

EXERCISES AND NOTES

Exercises

Chapter I – Random variables

1. Calculate the mean, the median and the variance / standard deviation of the following dataset on the observed dose of MDMA in 10 pills (in mg)

$$X = \{55,40,52,55,47,54,49,49,60,46\}$$

2. Define the type of the following variables
   a. The number of minutiae in friction ridge impressions
   b. The dose of MDMA in pills
   c. The design of the face of pills
   d. The size of shoes
   e. Blood alcohol content
   f. The color of fibers
   g. The number of glass fragments transferred on a garment
   h. The size of garments

Chapter II – Probability and probability distributions

1. The probability of observing an arch on any given person is 7%. The probability of observing a certain spatial arrangement of 4 minutiae is 8%. The probability to observe the same spatial arrangement of 4 minutiae on arches is 9%.
   a. Check if spatial arrangement and friction ridge pattern are independent
   b. Calculate the probability of observing the spatial arrangement given that you are looking at an arch
   c. Calculate the probability of observing something else than an arch

d. Calculate the probability to observe the spatial arrangement on something else than an arch

e. Calculate the probability of observing an arch or the spatial arrangement

2. The probability to observe red viscose fibers on a garment is 3%.

   a. Calculate the probability that we observe red viscose on the first garment we process

   b. Calculate the probability that we observe red viscose on one of the first three garments that we process

   c. Calculate the probability that we need to process more than three garments to observe red viscose

3. The probability to observe a counterfeit penny is about 5%. We observe a sample of 100 pennies from a much larger population of pennies.

   a. Calculate the probability that we observe 4 (repeat for 5 and 6) counterfeit pennies in the sample of 100.

   b. Is the result in a surprising?

   c. Calculate the probability to observe between 4 and 6 counterfeit pennies in the sample.

4. A sample of 100 white pills contains 60 pills composed of MDMA. You sample 50 pills out of the 100. What is the probability that 30 of them contain MDMA?

5. Solve the following equations:

$\Pr(Z \leq 2.58) =$

$\Pr(Z \leq -1.25) =$

$\Pr(Z \geq 1.96) =$

$\Pr(Z \leq z) = 0.7190$

$\underset{\mathrm{df}=12}{\Pr}(T \geq t) = 0.01$

$\underset{\mathrm{df}=12}{\Pr}(T \leq -t) = 0.01$

$\underset{\mathrm{df}=18}{\Pr}(T \leq t) = 0.995$

$\underset{\mathrm{df}=11}{\Pr}(\mathrm{X}^2 \geq \chi^2) = 0.975$

$\underset{\mathrm{df}=11}{\Pr}(\mathrm{X}^2 \leq \chi^2) = 0.025$

$\underset{\mathrm{df}=10}{\Pr}(\mathrm{X}^2 \geq \chi^2) = 0.01$

6. Solve the following equation for $\mu = 15$ and $\sigma^2 = 4$

$\Pr(X \leq 17) =$

$\Pr(X \geq 11.7) =$

$\Pr(12.5 \leq X \leq 16.5) =$

Chapter IV – Parameter estimates and confidence intervals

1. The purity of a shipment of 100 bags of cocaine is believed to be normally distributed. The purity of 10 bags has been measured.
    a. Estimate the purity of the shipment using a 95% confidence interval.

$x = \{0.7599, 0.7582, 0.7291, 0.7475, 0.7530, 0.7482, 0.7596, 0.7705, 0.7434, 0.7410\}$

   b. What is the probability that the CI includes the true value of 0.75?
   c. What would have happened if we were to analyze another 10 samples?

2. A random sample of 100 individuals are tested for blood alcohol content. After having tested the 30 first individuals, it turns out that 12 of them have a BAC larger than the legal limit.
    a. Estimate the proportion of individuals that have a BAC larger than the legal limit using a 90% confidence interval

   b. Estimate the proportion of individuals that have a BAC larger than the legal limit using a 95% confidence interval
   c. Repeat a and b, knowing that 27 out of 60 individuals have a BAC larger than the legal limit.
   d. What can you observe by comparing a, b and c.

Chapter V – Sample size

1. We want to characterize the proportion of individuals with arch friction ridge pattern in the general population.
   a. Calculate the sample size that we need to estimate that proportion with a precision of $\pm\, 0.01$ and a confidence of 95%
   b. What would happen if you want to determine the same proportion (with the same precision and confidence) in a finite population of 1,000 people?
   c. What would happen if you want to redo b but you use the information that the proportion should be around 5%?

Chapter VII – Hypothesis testing

1. A study shows that 80 out of 120 fingerprint "identifications" were made based on more than 12 minutiae in common between the trace and control impressions. Test the hypothesis that more than 65% of "identifications" are made based on more than 12 minutiae.
2. Two garments are processed for foreign fibers. On the first garment, 190 foreign fibers (out of 336) are pink nylon, while on the second garment, 482 (out of 773) are pink nylon. Test whether the proportion of foreign pink nylon fibers is the same on both garments.
3. The refractive indices of fragments from 2 different windows are compared to determine if the average refractive index of both windows is the same. Use the following data to perform the test

$$n_a = 19;\ \bar{X}_a = 1.748421; S_a^2 = 0.579314$$

$$n_b = 28;\ \bar{X}_b = 1.386429; S_b^2 = 0.1651646$$

4. A researcher is interested in comparing the rates of different shoe designs in different sub-populations. Test whether the distributions of patterns are different from one sub-population to another.

	Sport shoes	City shoes	Hiking shoes	Casual shoes
Design A	56	83	43	55
Design B	25	44	18	11
Design C	23	53	21	33
Design D	45	89	38	60
Design E	28	37	17	17

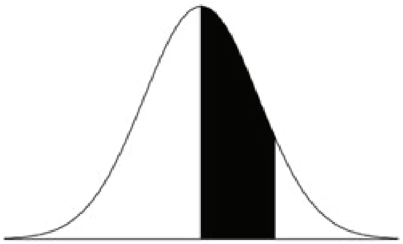
Chapter XI – Bayes theorem

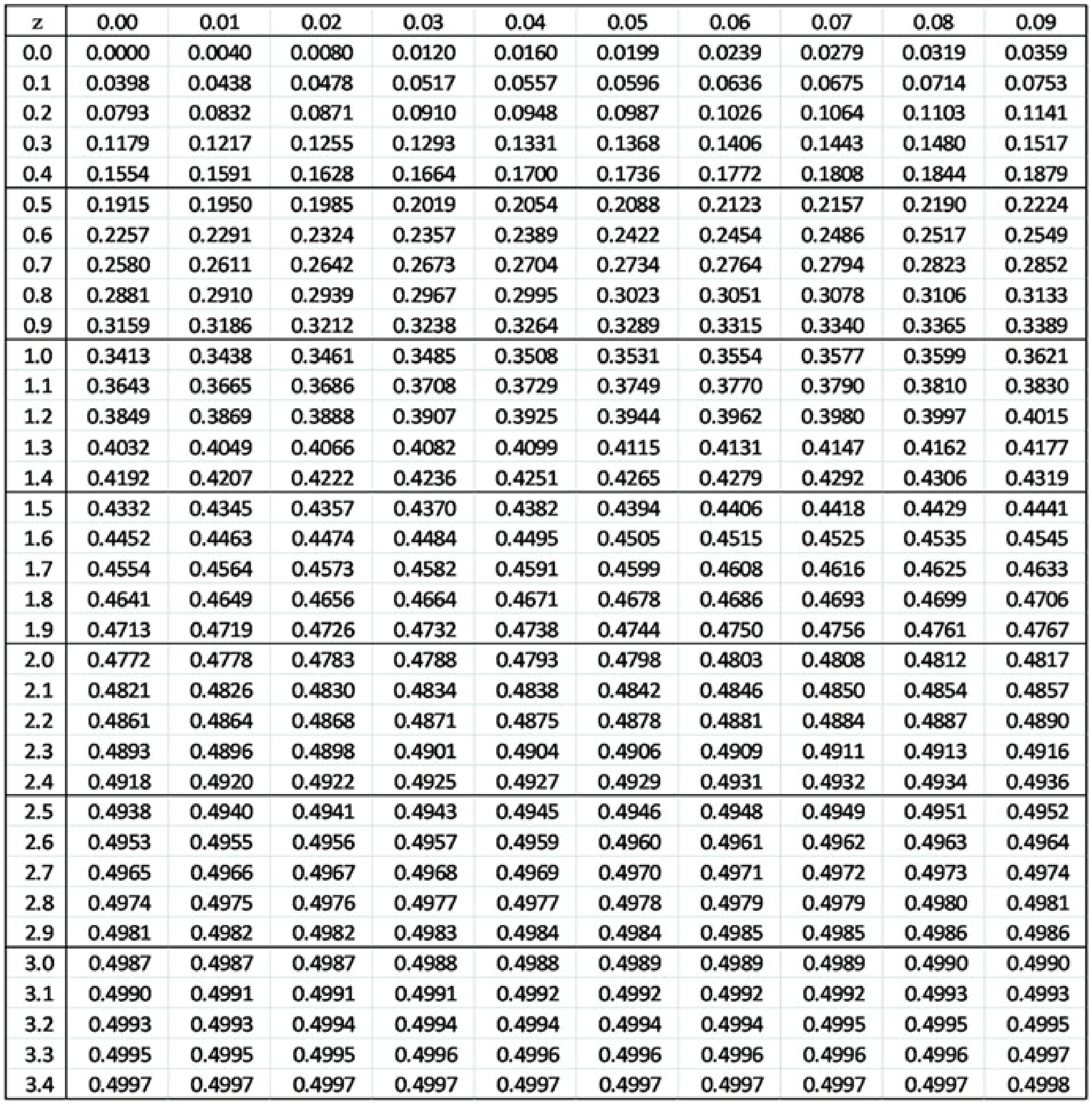
1. A partial DNA profile is found at a crime scene and compared with that of Mr. X. The probability of observing the partial DNA profile at the crime scene given that the biological material was left by Mr. X. is 0.67. The probability to observe the partial DNA profile if Mr. X. is not the source of the biological material is 0.0001.
   a. Calculate the LR
   b. Calculate the probability that Mr. X is the source of the partial DNA profile if the population of potential offender is 10,000
   c. What would happen if it is 1,000,000?
   d. Does the LR change between b and c?

2. A finger impression is found at a crime scene and compared with a control impression from Mr. X by an examiner in laboratory A. The examiner declares that they “match”. Examiners of laboratory A are known to be very good at correctly declaring matches when the donors of the control impression are also the donors of the trace. Examiners from laboratory A are known to have an error rate of 1 in 100,000 cases.
    a. Calculate the LR
    b. Calculate the probability that Mr. X is the source of the trace if the population of potential offenders is 100,000?
    c. What would happen if one considers that police detectives propose the correct source (using non-fingerprint evidence) in about 80% of the cases?
    d. What would happen if we assume prior odds that Mr. X is the source are “50/50”?

Source: B. Lawal, Applied Statistical Methods in Agriculture, Health and Life Science, Springer,

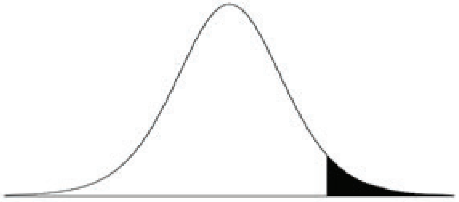
Table 1 Standard normal probabilities (area between 0 and z)

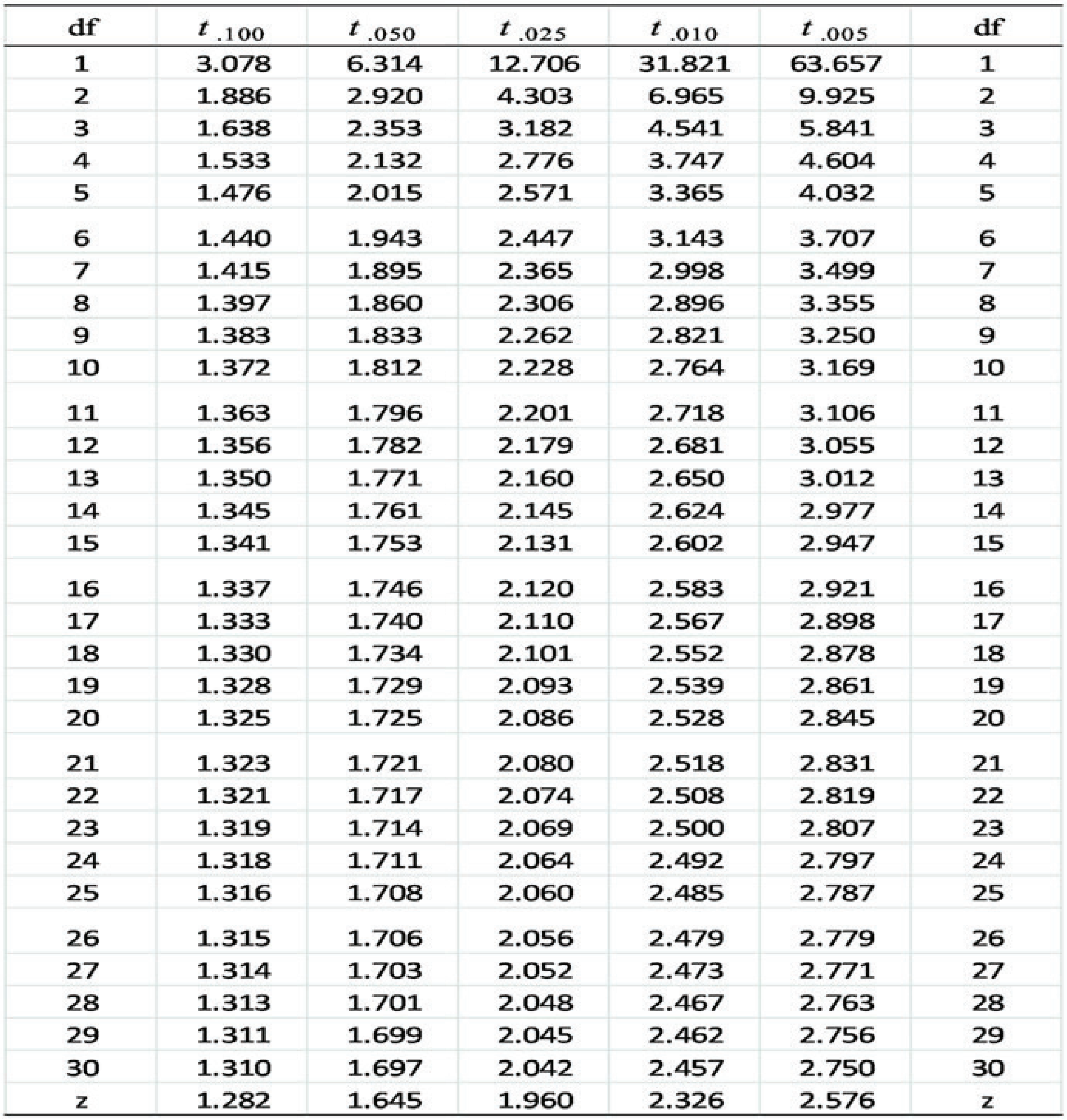




z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

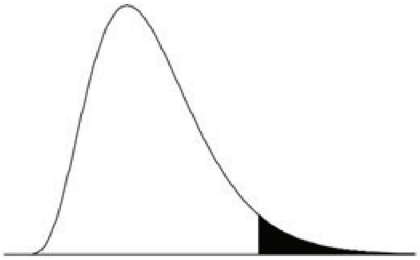
Table 2 Values of t_α in a t distribution with df degrees of freedom. (*shaded area* $P(t > t_\alpha) = \alpha$)

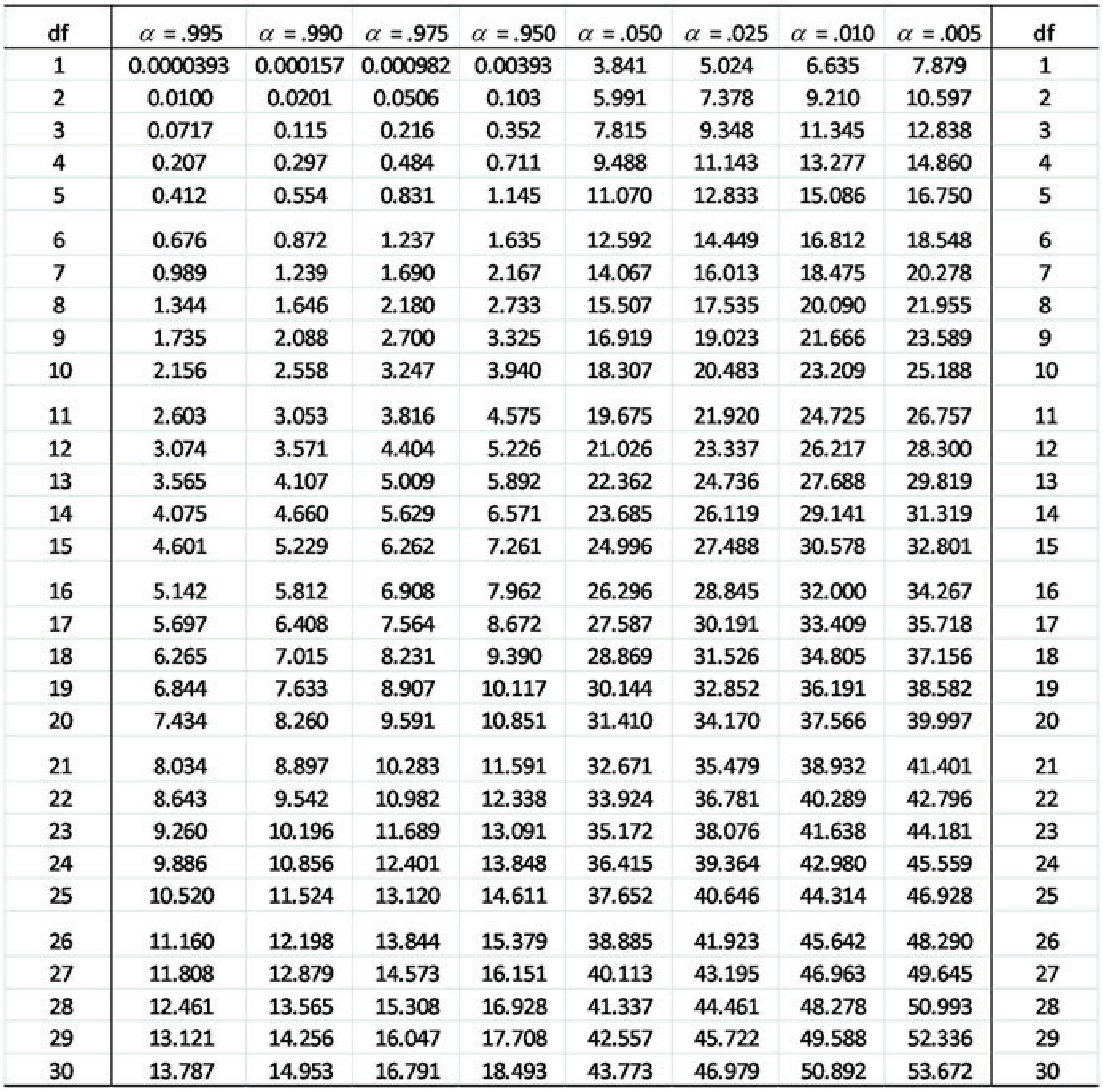




df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
30	1.310	1.697	2.042	2.457	2.750	30
z	1.282	1.645	1.960	2.326	2.576	z

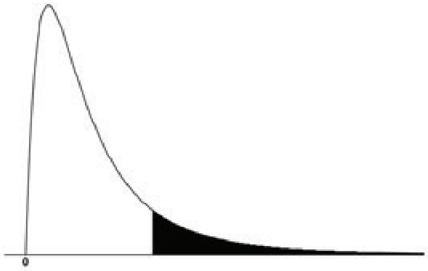
Table 3 Values of $\chi^2_{\alpha,\mathrm{df}}$ in a chi-square distribution with *df* degrees of freedom (*shaded area* $P(\chi^2 > \chi^2_{\alpha,\mathrm{df}}) = \alpha$)

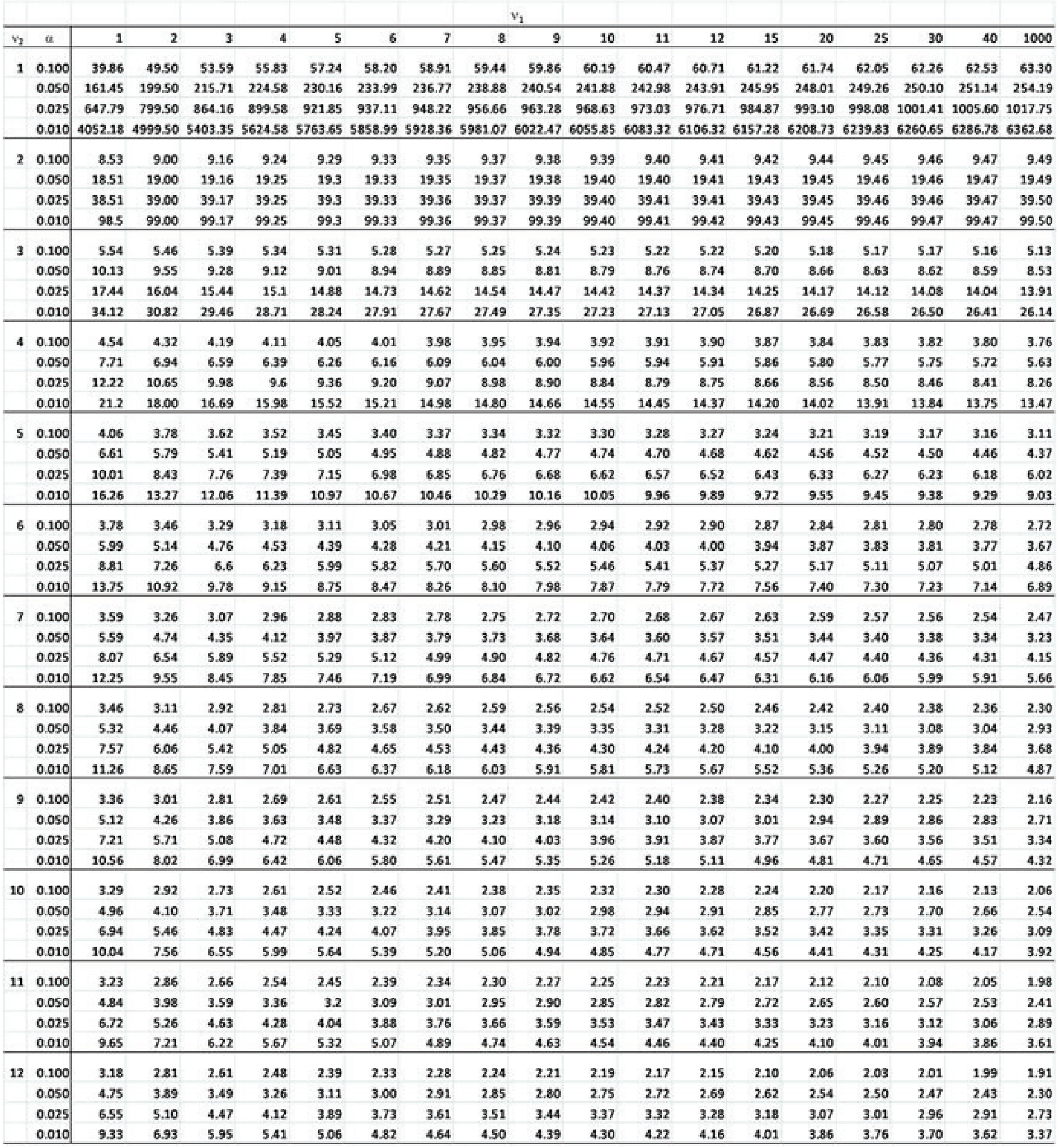




df	α = .995	α = .990	α = .975	α = .950	α = .050	α = .025	α = .010	α = .005	df
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.195	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	30

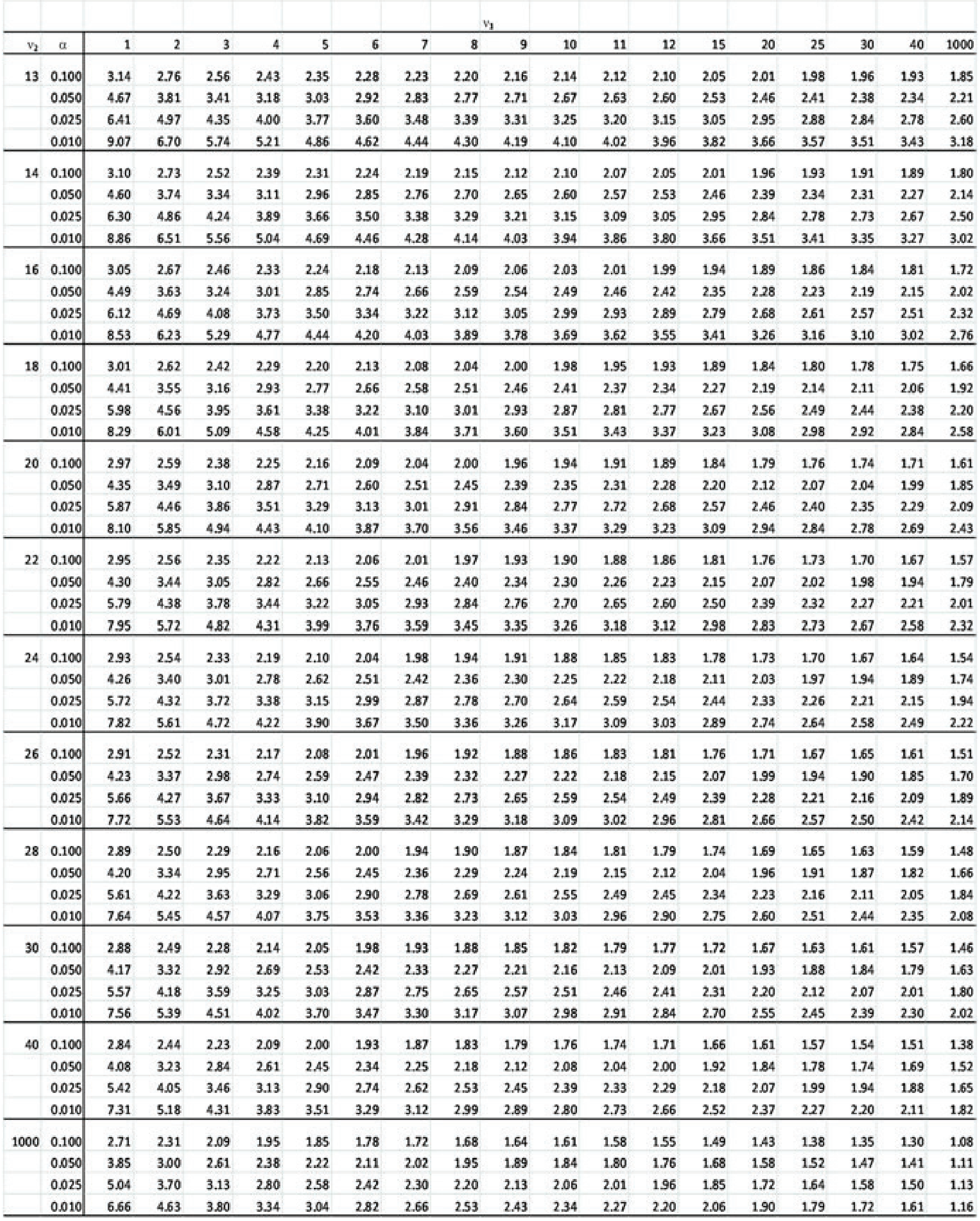
Table 4 Values of f_{α,ν_1,ν_2} in an F distribution (*shaded area* $P(F > f_{\alpha,\nu_1,\nu_2}) = \alpha$). Numerator degrees of freedom is ν_1 and denominator degrees of freedom is ν_2.





		ν_1																	
ν_2	α	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	1000
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.47	60.71	61.22	61.74	62.05	62.26	62.53	63.30
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	245.95	248.01	249.26	250.10	251.14	254.19
	0.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	973.03	976.71	984.87	993.10	998.08	1001.41	1005.60	1017.75
	0.010	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6083.32	6106.32	6157.28	6208.73	6239.83	6260.65	6286.78	6362.68
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41	9.42	9.44	9.45	9.46	9.47	9.49
	0.050	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.49
	0.025	38.51	39.00	39.17	39.25	39.3	39.33	39.36	39.37	39.39	39.40	39.41	39.41	39.43	39.45	39.46	39.46	39.47	39.50
	0.010	98.5	99.00	99.17	99.25	99.3	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47	99.47	99.50
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22	5.20	5.18	5.17	5.17	5.16	5.13
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.63	8.62	8.59	8.53
	0.025	17.44	16.04	15.44	15.1	14.88	14.73	14.62	14.54	14.47	14.42	14.37	14.34	14.25	14.17	14.12	14.08	14.04	13.91
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05	26.87	26.69	26.58	26.50	26.41	26.14
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90	3.87	3.84	3.83	3.82	3.80	3.76
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75	5.72	5.63
	0.025	12.22	10.65	9.98	9.6	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75	8.66	8.56	8.50	8.46	8.41	8.26
	0.010	21.2	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.20	14.02	13.91	13.84	13.75	13.47
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.24	3.21	3.19	3.17	3.16	3.11
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.52	4.50	4.46	4.37
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.43	6.33	6.27	6.23	6.18	6.02
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.72	9.55	9.45	9.38	9.29	9.03
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90	2.87	2.84	2.81	2.80	2.78	2.72
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.94	3.87	3.83	3.81	3.77	3.67
	0.025	8.81	7.26	6.6	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37	5.27	5.17	5.11	5.07	5.01	4.86
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.56	7.40	7.30	7.23	7.14	6.89
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67	2.63	2.59	2.57	2.56	2.54	2.47
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.51	3.44	3.40	3.38	3.34	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67	4.57	4.47	4.40	4.36	4.31	4.15
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.31	6.16	6.06	5.99	5.91	5.66
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50	2.46	2.42	2.40	2.38	2.36	2.30
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.22	3.15	3.11	3.08	3.04	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20	4.10	4.00	3.94	3.89	3.84	3.68
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.52	5.36	5.26	5.20	5.12	4.87
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38	2.34	2.30	2.27	2.25	2.23	2.16
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.01	2.94	2.89	2.86	2.83	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87	3.77	3.67	3.60	3.56	3.51	3.34
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	4.96	4.81	4.71	4.65	4.57	4.32
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.24	2.20	2.17	2.16	2.13	2.06
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.73	2.70	2.66	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62	3.52	3.42	3.35	3.31	3.26	3.09
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.31	4.25	4.17	3.92
11	0.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21	2.17	2.12	2.10	2.08	2.05	1.98
	0.050	4.84	3.98	3.59	3.36	3.2	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.72	2.65	2.60	2.57	2.53	2.41
	0.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43	3.33	3.23	3.16	3.12	3.06	2.89
	0.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.25	4.10	4.01	3.94	3.86	3.61
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15	2.10	2.06	2.03	2.01	1.99	1.91
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.62	2.54	2.50	2.47	2.43	2.30
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28	3.18	3.07	3.01	2.96	2.91	2.73
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.01	3.86	3.76	3.70	3.62	3.37

Table 4 Values of f_{α,ν_1,ν_2} in an F distribution (continued)



		ν_1																	
ν_2	α	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	1000
13	0.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10	2.05	2.01	1.98	1.96	1.93	1.85
	0.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.53	2.46	2.41	2.38	2.34	2.21
	0.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15	3.05	2.95	2.88	2.84	2.78	2.60
	0.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.82	3.66	3.57	3.51	3.43	3.18
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.07	2.05	2.01	1.96	1.93	1.91	1.89	1.80
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.46	2.39	2.34	2.31	2.27	2.14
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05	2.95	2.84	2.78	2.73	2.67	2.50
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.66	3.51	3.41	3.35	3.27	3.02
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99	1.94	1.89	1.86	1.84	1.81	1.72
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.35	2.28	2.23	2.19	2.15	2.02
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.93	2.89	2.79	2.68	2.61	2.57	2.51	2.32
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.41	3.26	3.16	3.10	3.02	2.76
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.95	1.93	1.89	1.84	1.80	1.78	1.75	1.66
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.27	2.19	2.14	2.11	2.06	1.92
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.81	2.77	2.67	2.56	2.49	2.44	2.38	2.20
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.23	3.08	2.98	2.92	2.84	2.58
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.84	1.79	1.76	1.74	1.71	1.61
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.07	2.04	1.99	1.85
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68	2.57	2.46	2.40	2.35	2.29	2.09
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.84	2.78	2.69	2.43
22	0.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86	1.81	1.76	1.73	1.70	1.67	1.57
	0.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.15	2.07	2.02	1.98	1.94	1.79
	0.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.65	2.60	2.50	2.39	2.32	2.27	2.21	2.01
	0.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	2.98	2.83	2.73	2.67	2.58	2.32
24	0.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83	1.78	1.73	1.70	1.67	1.64	1.54
	0.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.11	2.03	1.97	1.94	1.89	1.74
	0.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54	2.44	2.33	2.26	2.21	2.15	1.94
	0.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.89	2.74	2.64	2.58	2.49	2.22
26	0.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.83	1.81	1.76	1.71	1.67	1.65	1.61	1.51
	0.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.07	1.99	1.94	1.90	1.85	1.70
	0.025	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.54	2.49	2.39	2.28	2.21	2.16	2.09	1.89
	0.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.81	2.66	2.57	2.50	2.42	2.14
28	0.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79	1.74	1.69	1.65	1.63	1.59	1.48
	0.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.04	1.96	1.91	1.87	1.82	1.66
	0.025	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.49	2.45	2.34	2.23	2.16	2.11	2.05	1.84
	0.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.75	2.60	2.51	2.44	2.35	2.08
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.72	1.67	1.63	1.61	1.57	1.46
	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.88	1.84	1.79	1.63
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41	2.31	2.20	2.12	2.07	2.01	1.80
	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.45	2.39	2.30	2.02
40	0.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.66	1.61	1.57	1.54	1.51	1.38
	0.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.78	1.74	1.69	1.52
	0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29	2.18	2.07	1.99	1.94	1.88	1.65
	0.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.27	2.20	2.11	1.82
1000	0.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.64	1.61	1.58	1.55	1.49	1.43	1.38	1.35	1.30	1.08
	0.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.68	1.58	1.52	1.47	1.41	1.11
	0.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	2.13	2.06	2.01	1.96	1.85	1.72	1.64	1.58	1.50	1.13
	0.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.27	2.20	2.06	1.90	1.79	1.72	1.61	1.16

NOTES