# Workbook: Statistical Thinking for Forensic Practitioners

*Sam Tyner, Hal Stern, Alicia Carriquiry*

# Contents

# Chapter 1

# Introduction

This workbook is intended to accompany the Statistical Thinking for Forensic Practitioners workshop taught by members of the Center for Statistics and Applications in Forensic Evidence (CSAFE). The slides for this workshop were constructed initially by Dr. Hal Stern of UC-Irvine and Dr. Alicia Carriquiry of Iowa State University.

When taking the workshop, please follow along with the slides handout (if given) and this workbook. The workbook contains the same material as the slides, with room for you to take notes and to fill in the missing material.

# Chapter 2

# Statistical Preliminaries

Briefly, this section contains a broad review of probability concepts and of statistical inference concepts, with examples from the forensic science context. We will cover probability, data collection, statistical distributions, estimation, and hypothesis testing.

### 2.0.1 Definitions

- **population**: ______________________________________________
- **sample**: ______________________________________________
- **probability**: Using knowledge about the ________________ to make statements describing the ________________. Probability can loosely be thought of as a type of deductive reasoning, where we are applying general knowledge about the population of interest to make conclusions about a small part of that population.
- **statistics**: Using knowledge about the ________________ to make statements describing the ________________. Statistics can loosely be thought of as a type of inductive reasoning, where we are applying knowledge about a sample to state that something *may* be true about the population generally.

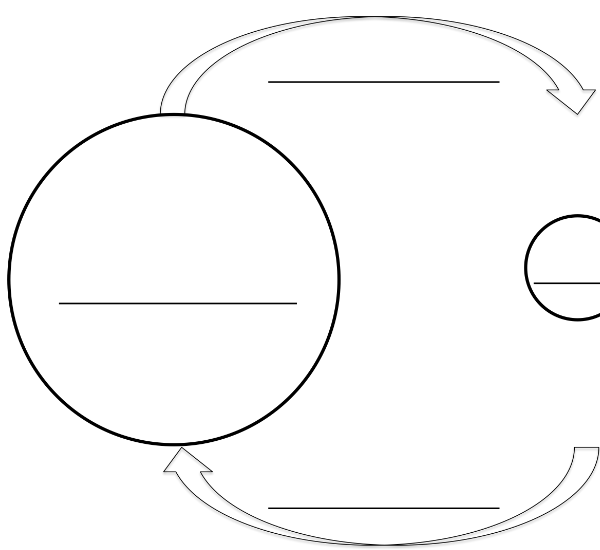

Figure 2.1: "The Big Picture"

### 2.0.2 Forensic Science Examples

- Suppose 100 1-pound bags of heroin are seized on the US-Mexico border, and the FBI want to know the chemical composition of the confiscated drugs to store in their database.
  - Population: ________________________________________
  - Sample: ________________________________________
- A window was broken in a robbery, and the suspect who was apprehended nearby had glass fragments lodged in the soles of their shoes. Do the fragments from the suspect's shoes have the same or similar chemical composition as the broken window?
  - Population 1: ________________________________________
  - Sample 1: ________________________________________
  - Population 2: ________________________________________
  - Sample 2: ________________________________________
- A city government employee is suspected of embezzling funds from the city's coffers. Forensic accountants examine a subset of the city's transactions to determine whether embezzling occurred and how much money was lost.
  - Population: ________________________________________
  - Sample: ________________________________________

How do you think this pertains to pattern evidence? List some possible relevant populations and samples below.

- Population 1: ________________________________________
- Sample 1: ________________________________________
- Population 2: ________________________________________
- Sample 2: ________________________________________
- Population 3: ________________________________________
- Sample 3: ________________________________________

## 2.1 Probability

Probability concerns the *uncertainty* of outcomes. The set of all possible outcomes is called the __________________ space, and a particular outcome or set of outcomes of interest is referred to as an ________________.

### 2.1.1 Examples

1. Footwear
   - Sample Space = All shoe sizes e.g. $\{6, 6.5, 7, 7.5, 8, 8.5, \ldots\}$
   - Event = Shoe of size 9

2. Footwear
   - Sample Space = Brand of shoe e.g. { Nike, Vans, Converse, ...}
   - Event = Nike sneaker
3. Firearms
   - Sample Space = CMS (consecutive matching striae) for a pair of bullets e.g. $\{0, 1, 2, 3, 4, \dots\}$
   - Event = CMS of 10 or more

### 2.1.2 Interpretation

The probability of observing an event in a sample space is a number less than or equal to 1 and greater than or equal to 0 that describes the ________________ that the event will occur.

There are two primary interpretations of probability:

1. The long run ________________ of occurrence of an event.
2. The _________________ belief of likelihood of an event occurring.

### 2.1.3 Basic Notation and Laws of Probability

Let an event of interest be denoted by _________. The probability of this event occurring is then denoted ______________. Recall that the probability of an event is always between 0 and 1. When $P(Y) = 0$, the event $Y$ will never happen. When $P(Y) = 1$, the event $Y$ will always happen. The sum of the probabilities of all possbile outcomes in the sample space always equal to _______.

The event of interest, $Y$, also has a complement event, $\overline{Y}$, which is read as "not $Y$". The complement, $\overline{Y}$, of an event, $Y$, is itself an event containing all outcomes in the sample space other than that initial event of interest, $Y$.

$$P(Y) + P(\overline{Y}) = \_\_\_\_$$

The above equation also gives us the following rules:

$$\begin{aligned} P(Y) &= 1 - P(\overline{Y}) \\ P(\overline{Y}) &= 1 - P(Y) \end{aligned} \tag{2.1}$$

### 2.1.4 Probability and Odds

The probability of an event defines the odds of the event. The odds *in favor* of an event $Y$ are defined as the probability of $Y$ divided by the probability of everything except $Y$ ("not $Y$"):

$$O(Y) = \frac{P(Y)}{P(\overline{Y})} = \frac{P(Y)}{1 - \_\_\_}.$$

Conversely, the odds *against* a event $Y$ are defined as the probability of everything except $Y$ ("not $Y$") divided by the probability of $Y$:

$$O(\overline{Y}) = \frac{P(\overline{Y})}{P(Y)} = \frac{1 - \_\_\_}{P(Y)}.$$

When we typically talk about odds, like in horse racing, the odds reported are the odds *against* the outcome of interest. Let's construct a horse race scenario using our probability notation to find the probability of a horse winning a race from the reported odds:

- Suppose you want to place a bet on a horse name Cleopatra winning the race. Odds for Cleopatra are reported as 4:1.
- $Y$ = Cleopatra wins the race
- $\overline{Y}$ = Any horse in the race *other than* Cleopatra wins the race.
- $O(\overline{Y}) = \dfrac{P(\overline{Y})}{P(Y)} = \frac{4}{1} = 4$
- We know that $P(Y)+P(\overline{Y}) = 1$. With this information, we can determine $P(Y)$, which is the probability that Cleopatra wins the race:

$$
\begin{aligned}
O(\overline{Y}) = \frac{P(\overline{Y})}{P(Y)} &= 4 \\
\Rightarrow \frac{P(\overline{Y})}{P(Y)} &= 4 \\
\Rightarrow \frac{1 - P(Y)}{P(Y)} &= 4 \qquad \textit{(See Equation 2.1)} \\
\Rightarrow \frac{1}{P(Y)} - 1 &= 4 \\
\Rightarrow \frac{1}{P(Y)} &= 5 \\
\Rightarrow P(Y) &= \frac{1}{5} = 0.2 \\
\Rightarrow P(\overline{Y}) &= 0.8
\end{aligned}
$$

- So, the odds for Cleopatra (4:1) mean that Cleopatra has a probability of 0.2 of winning the race. Because this outcome is not very likely (it will only happen in 1 race out of 5), you win money if Cleopatra wins simply because that is not a likely outcome.
- **Betting**: Suppose you bet \$1 on Cleopatra to win the race with 4:1 odds. You will win \$4 if Cleopatra wins, otherwise you've lost \$1.
- The amount you win (\$4) is determined so that you break even in the long run.
- Suppose 5 identical races are run. In 1 of those races, Cleopatra will win, and in the other 4, Cleopatra will lose. If you bet \$1 on Cleopatra in each race, you will lose that \$1 4 of 5 times. So, in order for you to break even, the designated amount you'll win when Cleopatra wins is \$4.
- This is a statistical concept known as *expected value*. Your expected value when placing the bet is \$0. We compute expected value by multiplying each possible outcome value by its probability and adding them all together:

$$
\begin{aligned}
\$4 \cdot P(Y) + (-\$1) \cdot P(\overline{Y}) &= 0 \\
\$4 \cdot 0.2 + (-\$1) \cdot 0.8 &= 0 \\
\$0.8 - \$0.8 &= 0
\end{aligned}
$$

### 2.1.5 Probability Math

Up until now, we have only considered one event, $Y$. Now, suppose we have another event that we are interested in, $Z$.

Let's consider the possibility of *either* of these two events, $Y$ or $Z$, occurring. We'd write this as $Y \cup Z$, which is mathematical notation for "$Y$ or $Z$ occurs". There are two scenarios that arise:

1. $Y$ and $Z$ cannot occur together: they are ____________ _____________
2. $Y$ and $Z$ can occur together.

In scenario #1, computing the probability of either $Y$ or $Z$ happening is easy: we just add their respective probabilities together:

$$Y, Z \text{ mutually exclusive } \Rightarrow P(Y \cup Z) = P(Y) + P(Z)$$

In scenario #2, computing the probability of either $Y$ or $Z$ happening is more complicated because we know there is a chance that $Y$ and $Z$ can happen together. We'd write this as $Y \cap Z$, which is mathematical notation for "$Y$ and $Z$ occurs". In scenario #1, this event never occurred, so $P(Y \cap Z) = 0$ there. To compute the probability of $Y$ or $Z$ occurring in scenario #2, we have to consider the probability of $Y$, the probability of $Z$, and the probability of $Y \cap Z$. If we just add $P(Y) + P(Z)$ as in scenario #1, we include the event $Y \cap Z$ twice, so we have to subtract one instance of it:

$$Y, Z \text{ not mutually exclusive } \Rightarrow P(Y \cup Z) = P(Y) + P(Z) - P(Y \cap Z).$$

This probability is much easier to think about when illustrated. In Figure 2.2, we consider human blood types. There are four groups: A, B, O, and AB, and there are two RH types: + and −. We first consider the blood types A and B, represented by the two non-overlapping circles. Define:

- Event $Y$ = a person has blood type A
- Event $Z$ = a person has blood type B
- Event $Y \cup Z$ = a person has blood type A or blood type B

These two events are *mutually exclusive* because one person cannot have both blood type A and blood type B. (The circles don't overlap in the venn diagram) So, the probability that a randomly selected person has blood type A or B is:

$$P(Y \cup Z) = \_\_\_\_ \quad + \quad \_\_\_\_$$

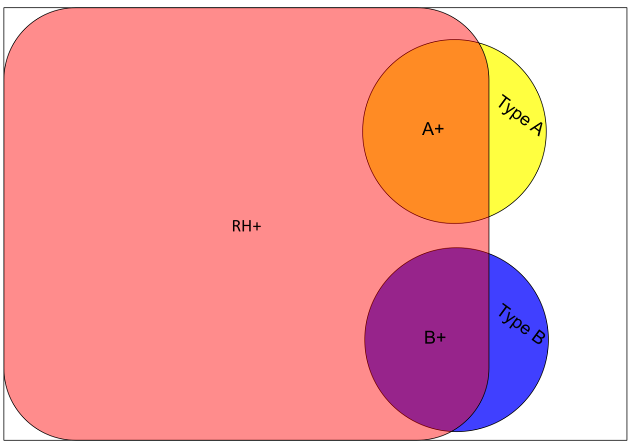


Figure 2.2: Probabilities of blood types in humans. Areas are approximate.

Return to Figure 2.2 and consider two other events: a person having blood type A or having the Rh factor (RH+). We see in Figure 2.2 that someone can have both type A blood and the Rh factor (blood type A+). Define:

- Event $Y$ = a person has blood type A
- Event $Z$ = a person has the Rh factor

- Event $Y \cup Z$ = a person has blood type A or the Rh factor
- Event $Y \cap Z$ = a person has blood type A and the Rh factor (they have A+ blood)

So, the probabilty that someone has either type A blood or has the Rh factor is the sum of probability of having type A blood (represented by the yellow circle) and the probability of having the Rh factor (represented by the red rectangle) minus the probability of having A+ blood (represented by the orange area of overlap that is counted twice) in Figure 2.2. So, the probability that a randomly selected person has blood type A or the Rh factor is:

$$P(Y \cup Z) = \_\_\_\_ \quad + \quad \_\_\_\_ \quad - \quad \_\_\_\_$$

### 2.1.6 Conditional Probability

Let's consider an event of interest $Y$ which has probability $P(Y)$. Then, suppose we learn of another event of interest $Z$ that has occurred. Knowing that $Z$ has occurred already may change our opinion about the likelihood of _______ occurring. The key idea here is that the probability of an event often depends on other information, leading us to the definition of *conditional probability*:

$$P(Y|Z),$$

which is the conditional ____________________ that $Y$ occurs given that we know $Z$ has occurred. Return to Figure 2.2. Suppose we want to know the probability of a person having type A blood, represented by the yellow circle. But, if we already know that a person has the Rh factor, we are only interested in the part of the type A circle that overlaps with the Rh+ rectangle. Thus the probability of having type A blood is different with different knowledge. The formula for calculating conditional probability is:

$$P(Y|Z) = \frac{P(Y \cap Z)}{P(Z)} \tag{2.2}$$

Returning to the venn diagram, the value $P(Y \cap Z)$ is represented by the overlap of the type A circle and the Rh+ rectangle, and the value $P(Z)$ is represented by the Rh+ rectangle. Then, the value $P(Y|Z)$ is the ratio of the overlap (A+) to the Rh+ rectangle.

Equation 2.2 also gives us a multiplication rule for computing probabilities:

$$P(Y \cap Z) = P(Y|Z) \cdot P(Z) \tag{2.3}$$

Philosophically speaking, it can be helpful to think of *all* probabilities as conditional. It is just a question of what information is assumed to be ________________.

#### 2.1.6.1 Examples

**Death Penalty Convictions**

A study of sentencing of 362 black people convicted of murder in Georgia in the 1980s found that 59 were sentenced to death (Baldus, Pulaski, and Woodworth (1983)). They also examined the race of the murder victim, either black or white, and found some disparities. In Table 2.1, DP means the defendant received the death penalty, NDP means the defendant did not receive the death penalty. The race of the victim (RV) is either black (B) or white (W).

Returning to Section 2.0.1, let's define the problem:

- **Population**: All black people convicted of murder in Georgia in the 1980s
- **Sample**: N/A (the whole population was studied)

| RV | DP | NDP | Total |
|---|---|---|---|
| W | 45 | 85 | 130 |
| B | 14 | 218 | 232 |
| Total | 59 | 303 | 362 |

Table 2.1: The results of the Baldus et al study for black defendants convicted of murder.

Using the numbers from Table 2.1, compute the following probabilities:

- $P(DP) = — = 0.\_\_\_\_\_$
- $P(DP|RV = W) = — = 0.\_\_\_\_\_$
- $P(DP|RV = B) = — = 0.\_\_\_\_\_$

Note: These numbers are selected from the study, and should not be considered a comprehensive summary of its results. There are a number of things not discussed here. The entire publication can be found online[1]

**Consecutive Matching Striae**

In firearms and toolmark analysis, the number of consecutive matching striae (CMS) between a crime scene sample and a lab sample is often used to help determine a match. Generally speaking, the higher the maximum number of CMS found in a pair, the more likely the two samples came from the same source. Several known match (KM) pairs and known non-match (KNM) pairs of bullets were examined, and the results are shown in Figure 2.3 (Hare, Hofmann, and Carriquiry (2017)). What is the probability of seeing two known matches (or two known non-matches) given the maximum number of CMS? Here, we condition on ________________________________________. Again, we briefly return to Section 2.0.1, let's define the problem:

- **Population**: All pairs of fired bullets from unknown sources
- **Sample**: A sample of pairs of known matches and known non-matches

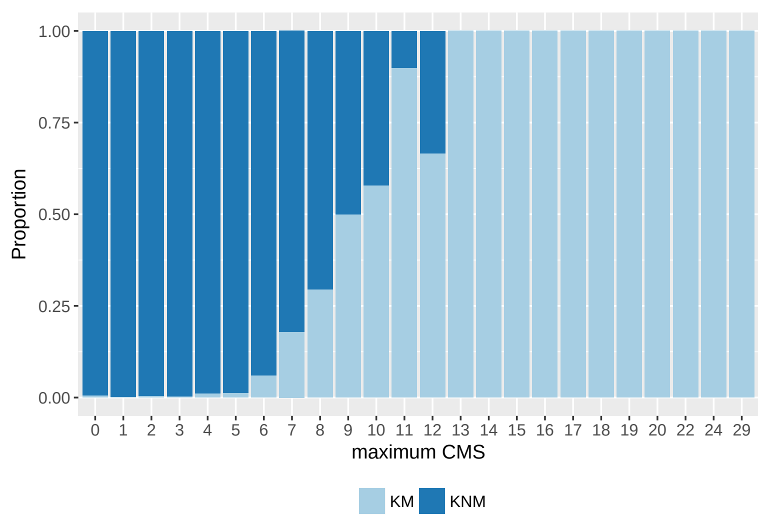


Figure 2.3: This bar chart represents the conditional probabilities of two bullets matching given the maximum number of CMS. The light blue represents known matches, while the dark blue represents known non-matches.

Generally, as seen in Figure 2.3, the probability of finding a match tends to increase with then number of maximum CMS. For ________ maximum CMS values is it much more likely that we have a ___________________________ pair.

[1]http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=6378&context=jclc.

### 2.1.7 Independence

If the likelihood of one event is *not* affected by knowing whether a second has occured, then the two events are said to be __________________________. For example, the region of the country where you live and what color car you drive are (probably) not related.

The death penalty example from the previous section demonstrates that defendants receiving the death penalty is *not* independent of the race of the victim. In other words, a black defendant found guilty of murder in Georgia in the 1980s received a different penalty depending on the race of the victim.

Another example from DNA analysis relies on on independence across chromosomes. By using loci on different chromosomes, there is independence between the allele counts, allowing for simple calculation of random match probabilities.

### 2.1.8 Probability Math... Again

Recall Equation 2.3, which gives us the probability of two events, $Y$ and $Z$ occurring together:

$$P(Y \cap Z) = P(Z) \cdot P(Y|Z) = P(Y) \cdot P(Z|Y)$$

If $Y$ and $Z$ are *independent*, there is a simple formula:

$$P(Y \cap Z) = \underline{\hspace{2em}} \cdot \underline{\hspace{2em}}$$

This is because $Z$ occurring does not effect the probability of $Y$ occurring, and vice versa. Thus,

$$P(Y|Z) = P(Y) \quad \text{and} \quad P(Z|Y) = P(Z)$$

For example, the probability of being left-handed and from Florida is equal to the probability of being left-handed times the probability of being from Florida, assuming the events “being left-handed” and “being from Florida” are independent.

Multiplying probabilities of events directly like this is *only* applicable when the events are independent. When *dependent* events are treated as independent events, things can go terribly wrong. An infamous example of this in the courts is the case *People v. Collins*[2]. This was a robbery trial, where eyewitnesses described the robbers a “black male with a beard and a moustache, and a white female with a blonde ponytail, fleeing in a yellow car”.

The prosecution provided estimated probabilities of each of these individual characteristic:

- P(black man with a beard) = __________________
- P(black man with a moustache) = __________________
- P(white woman with ponytail) = __________________
- P(white woman with blonde hair) = __________________
- P(yellow car) = __________________
- P(interratial couple in a car) = __________________

A mathematics “expert” talked about the so-called “multiplication rule for probability”, and directly multiplied the above probabilities together without considering that the events could be *dependent*. i.e. a man with a beard probably has a much higher chance of having a moustache than a man with no beard. Due to this faulty math, the conviction was set aside and the statistical reasoning criticized for ignoring dependence among the characteristics.

[2] *People v. Collins*, 68 Cal.2d 319, 438 P.2d 33 (1968)

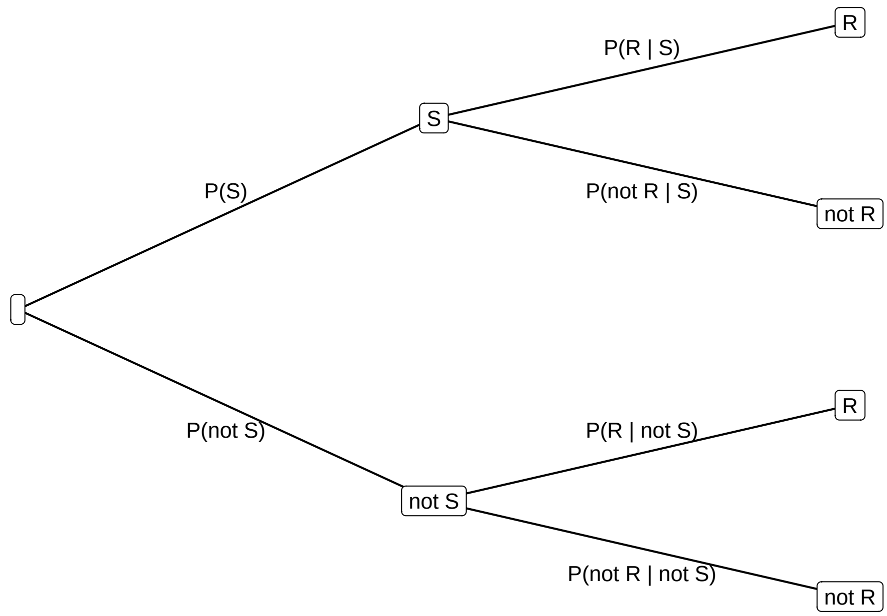


Figure 2.4: A probability tree showing the direction of flow when updating probabilities. Move from left to right on the tree through the events possible. Events are in boxes, probabilities are on the branches of the tree.

In a courtroom situation, let $S$ be the event that the suspect was present at the scene of the crime and $\overline{S}$ be the event that the suspect was not present at the scene. Assume that each juror has in mind an initial probability for the events $S$ and $\overline{S}$. Then, a witness says they saw a tall Caucasian male running from the scene, and the defendant is a tall Caucasian male. After hearing the witness' testimony, the jurors ______________ their probabilities. Next, an expert witness testifies that fragments from a window broken during the crime and fragments found on the defendant's clothing match. Again, the jurors update their __________________. This process continues throughout the trial. There are some key questions to consider:

- How should jurors update their probabilities?
- Do jurors *actually* think this way?

### 2.1.9 Bayes' Rule

*Bayes' Rule* provides an ______________ formula for probabilities. Like in the trial scenario above, suppose we have an initial estimate for the probability of event $S$, $P(S)$. Then, we learn that an event $R$ has occurred and we want to update or probability of event $S$. To do this, we need to know about the __________________ of $R$ and $S$. To update the probability of $S$, we can use Bayes' Rule, also called Bayes' _________________:

$$\begin{aligned} P(S|R) &= \frac{P(R \cap S)}{P(R)} = \frac{P(R|S)P(S)}{P(R)} \\ &= \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|\overline{S})P(\overline{S})} \end{aligned} \tag{2.4}$$

#### 2.1.9.1 Examples

Consider performing diagnostic tests for gunshot residue.

- Let $G$ denote the presence of gunshot residue
- Let $\overline{G}$ denote the ________________ of gunshot residue
- Let $T$ denote a __________________ diagnostic test
- Let $\overline{T}$ denote a negative diagnostic test

| Truth | $T$ | $\overline{T}$ |
|---|---|---|
| $G$ | True Positive | False Negative |
| $\overline{G}$ | False Positive | True Negative |

Table 2.2: All potential outcomes of a diagnostic test for gunshot residue.

The values in the table can also be thought of as conditional probabilities:

- The value $P(T|G)$ is the ______________________________ rate, also called *sensitivity* of the test
- The value $P(\overline{T}|\overline{G})$ is the ____________________________ rate, also called the *specificity* of the test
- The value $P(T|\overline{G})$ is the ______________________________ rate, the Type I error rate
- The value $P(\overline{T}|G)$ is the _________________________________ rate, the Type II error rate

Studies of the diagnostics test usually tell us $P(T|G)$, ________________, and $P(\overline{T}|\overline{G})$, ________________. Examiners may begin with some idea of $P(G)$, or the ________________ of gunshot residue in a similar situation. What is most relevent for the case is the *postitive predictive value*, or in probability notation, ________________. We can use _____________________________________ to obtain this value:

$$P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})}$$

Generally speaking, the most important thing to remember is that, in general, $P(T|G)$ ______ $P(G|T)$.

The careful application of Bayes' Rule can sometimes lead to surprising, non-intuitive results. Continuing with the gunshot residue test example, assume

- sensitivity is 98% ($P(\quad|\quad) = 0.98$)
- specificity is 96% ($P(\quad|\quad) = 0.96$)
- prevalence is 90% ($P(\quad) = 0.90$)
- Plug values into the Bayes' Rule formula to find $P(G|T)$:

$$\begin{aligned} P(G|T) &= \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})} \\ &= \frac{0.98 \cdot 0.9}{0.98 \cdot 0.9 + (1 - 0.96) \cdot (1 - 0.9)} \\ &= \frac{0.882}{0.882 + 0.004} \\ &= 0.995 \end{aligned} \tag{2.5}$$

- Now assume prevalence is 10% ($P(\quad) = 0.10$) and plug in the values again

$$
\begin{aligned}
P(G|T) &= \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})} \\
&= \frac{0.98 \cdot 0.1}{0.98 \cdot 0.1 + (1 - 0.96) \cdot (1 - 0.1)} \\
&= \frac{0.098}{0.098 + 0.036} \\
&= \frac{0.098}{0.134} \\
&= 0.731
\end{aligned} \tag{2.6}
$$

- So, even if there is a postive test, we are not really sure about whether gunshot residue is *actually* present.
- Why does this happen?? See Figure 2.5.

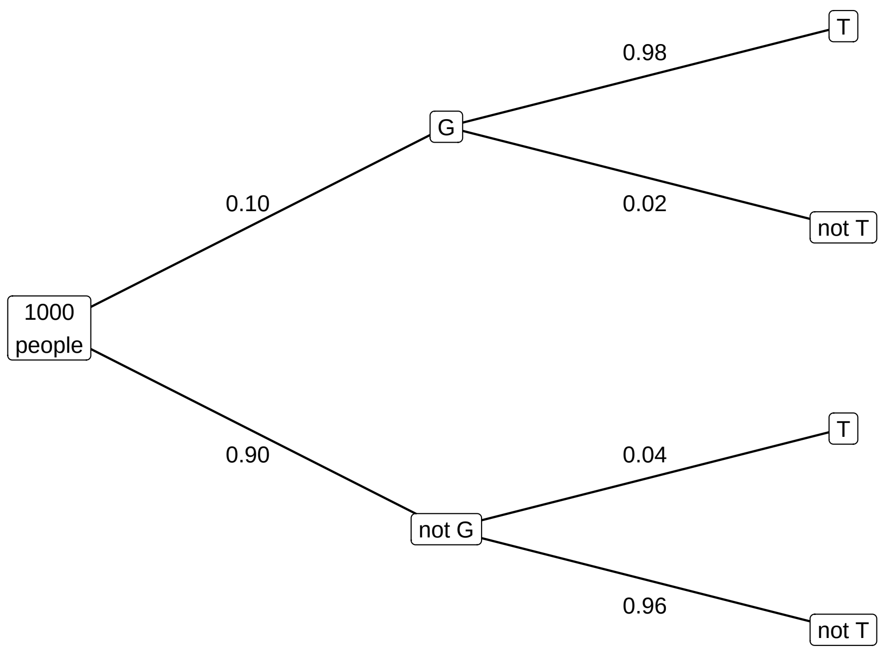


Figure 2.5: A probability tree showing the direction of flow when updating probabilities for the presence of gunshot residue. Suppose there are 1,000 people in the population you're considering. Write the number of people in the groups throughout the tree according to the probabilities indicated on the branches of the tree

### 2.1.10 Bayes' Rule to the Likelihood Ratio

In the general forensic setting, let $S$ denote the event that the evidence from the scene and comparison sample are from the same source. Let $E$ denote the evidence found at the scene. The formulation of Bayes' Rule for this situation is:

$$
P(S|E) = \frac{P(E|S)P(S)}{P(E|S)P(S) + P(E|\overline{S})P(\overline{S})}
$$

We can rewrite Bayes' Rule in terms of odds:

$$\frac{P(S|E)}{P(\overline{S}|E)} = \frac{P(E|S)}{P(E|\overline{S})}\frac{P(S)}{P(\overline{S})} \tag{2.7}$$

Derivation of Equation 2.7 is shown in Equation 2.8. For now, just consider Equation 2.7:

- On the left, $\frac{P(S|E)}{P(\overline{S}|E)}$ are the odds in favor of $S$ given the evidence $E$.
- The last term on the right, $\frac{P(S)}{P(\overline{S})}$ are the odds in favor of $S$ before seeing the evidence $E$ (the "prior odds")
- The first term on the right $\frac{P(E|S)}{P(E|\overline{S})}$, is known as the ____________________ ratio
- The likelihood ratio (LR) is the factor by which we ________________ prior odds of two samples being from the same source to get ______________________ odds (after seeing evidence) of the same source.

$$\begin{aligned}
P(S|E) &= \frac{P(E|S)P(S)}{P(E|S)P(S) + P(E|\overline{S})P(\overline{S})} \\
\Rightarrow \frac{1}{P(S|E)} &= \frac{P(E|S)P(S) + P(E|\overline{S})P(\overline{S})}{P(E|S)P(S)} \\
&= 1 + \frac{P(E|\overline{S})P(\overline{S})}{P(E|S)P(S)} \\
\Rightarrow \frac{1}{P(S|E)} - 1 &= \frac{P(E|\overline{S})P(\overline{S})}{P(E|S)P(S)} \\
\frac{1}{P(S|E)} - \frac{P(S|E)}{P(S|E)} &= \\
\frac{1 - P(S|E)}{P(S|E)} &= \\
\frac{P(\overline{S}|E)}{P(S|E)} &= \frac{P(E|\overline{S})P(\overline{S})}{P(E|S)P(S)} \\
\Rightarrow \frac{P(S|E)}{P(\overline{S}|E)} &= \frac{P(E|S)P(S)}{P(E|\overline{S})P(\overline{S})}
\end{aligned} \tag{2.8}$$

#### 2.1.10.1 Examples

Return to the gunshot residue (GSR) test example. Define:

- $E$ = evidence = a positive test for (GSR)
- $S$ = suspect has GSR on them

$$LR = \frac{P(E|S)}{P(E|\overline{S})} = \frac{0.98}{0.04} = 24.5$$

In a high prevalence case ($P(G) = 0.9$), the prior odds are $\frac{0.9}{0.1} = 9$. The posterior odds are $LR \times$prior odds $= 24.5 \times 9 = 220.5 : 1$.

In a low prevalence case ($P(G) = 0.1$), the prior odds are $\frac{0.1}{0.9} = \frac{1}{9}$. The posterior odds are $LR \times$prior odds $= 24.5 \times \frac{1}{9} = 24.5 : 9 = 2.72 : 1$.

We can also compute the likelihood ratio if the evidence were a negative test. This value turns out to be $\frac{1}{48}$, which is **not** the reciprocal of the LR for the positive test.

### 2.1.11 Recap

- Probability is the ____________________ language of ____________________
- Provides a common scale, from ________ to ________, for describing the chance that an event will occur
- **Conditional** probability is a key concept! The probabilitity of an event depends on what ___________________ is available
- Independent events can be powerful! They allow us to ______________ probabilities of events *directly*, as is common in __________________________.
- ______________________________ is a mathematical result showing how we should _________________ our probabilities when available information changes.
  - This will later lead us to the likelihood ratio as a numerical __________________ of the evidence.
  - Bayes' Rule does not necessarily describe how people operate in practice.

### 2.1.12 Probability and the Courts

Sally Clark was the only person in the house when her first child died unexpectedly at 3 months old. The cause of death was determined to be SIDS, sudden infant death syndrome. One year later, Sally and her husband had a second child, who died at 2 months old under similar circumstances. Sally was convicted of murder.

During her trial, a pediatrician testified that the probability of a single SIDS death for a family like the Clarks (similar income, etc.) was $\frac{1}{8500} \approx 0.0001$, and thus the probability of two SIDS death in the family was $\frac{1}{8500^2} = \frac{1}{73\times10^6} \approx 1.37 \times 10^{-8}$. There are several problems with this approach to evidence. What do you think? Jot down a few ideas below:

____________________________________________________________

____________________________________________________________

____________________________________________________________

Issues with the evidence presented by the pediatrician:

1. Is the probability of a child dying of SIDS given, $\frac{1}{8500}$, correct for "families like the Clarks"?
2. The use of direct multiplication of probabilities assumes independence of the two deaths in the family. (Independence within the family is not a reasonable assumption.)
3. Alternative hypotheses (causes of death of the infants) were not considered. Did something else with perhaps a higher likelihood cause the children's deaths?

## 2.2 Probability to Statistical Inference

Probability is important, but it is only one tool in our toolbox. Another, more powerful tool is statistical inference.

### 2.2.1 Collecting Data

First, we consider data collection. Where do data come from? One data source is an *experiment*. An investigator designs a study and collects information on and maybe applies treatments to a *sample*, a subset of the population of interest. _____________ can tell us a great deal about how to design an ________________ or choose a _____________.

The area of statistics concerned with creating studies is called *experimental design*. The experimental design literature is extensive (see for example Morris (2011)). Here are a few crucial points:

- The goal of an experiment is to compare __________________
- Those __________________ must be ________________ assigned to units
- The _____________________ in the experiment must be large enough t obe able to make informed conclusions
- Blinding plays an important role in avoiding ___________. e.g. "double-blind" studies in medicine, where neither the patient nor the doctor administering the treatment know which treatment the patient is receiving

How is experimental design relevant to forensic science?

- Experiments are used to evaluate process improvements
- Blinding is used in "black box" studies, where examiners do not know ground truth

Experiments almost always involve *sampling* from the population of interest. Why?

- We sample because it is too _____________ or ___________________ to study the *entire* population
- A _____________ sample allows us to use the laws of ________________ to describe how certain we are that our __________________ answer reflects the __________________.
- There are many famous failures (cautionary tales) with __________________________ sampling. (See Figure 2.6.)

How is sampling relevant to forensic science?

- Sampling techniques used to determine which and how many bags of suspect powder collected from a crime scene to test.

All data collected can be divided into one of two groups: qualitative or quantitative.

- **Qualitative** data describe qualities about the observations. For example, the race of a suspect, or their level of education. There are two subcategories of qualitative data:
  - ____________________: the data belong to one of a discrete number of groups or categories. For example: blood type (A, B, AB, or O)
  - _____________________: the data belong to one group in a set of ordered values. For example, the evaluation of a teacher (poor, average, excellent). The categories have an inherent ordering, unlike in categorical data.
- **Quantitative** data describe quantities that can be measured on the observations. These are numerical data. There are also two subcategories of quantitative data:
  - ______________________: the values are distinct or separate. An easy-to-understand example is integer observations: $\{0, 1, 2, 3, 4, \ldots\}$. A forensic science example is consecutive matching striae on bullets or toolmarks. (See Figure 2.3)
  - ________________________: the values can take on any value in a finite or infinite interval. Continuous values fall anywhere on the number line. A forensic science example is the refractive index of a glass fragment.

### 2.2.2 Probability Distributions

Suppose we are to collect data on some characteristic for a sample of individuals or objects (e.g. weight, trace element concentration). A probability ________________ is used to describe these possible values and how


Figure 2.6: This picture from the US presidential election of 1948 shows President Harry Truman, who won the election, holding a newspaper that went to print with the headline "Dewey Defeats Truman!" The headline was based on biased sampling that favored typically Republican demographics. Image Source: https://blogs.loc.gov/loc/2012/11/stop-the-presses/

________________ each value is to occur. There are many, many possible probability distributions, but some of the most common are the Binomial, Poisson, Normal, and Lognormal distributions. The probabilities associated with each of these distributions and their possible outcomes are plotted in Figures 2.7-2.8.

- Discrete distributions
  - __________________________: counts the number of _______________ in a fixed number ($n$) of ________________. Possible values are $\{0, 1, 2, \ldots, n\}$.
  - _____________________: counts the number of _______________ occurring. Possible values are $\{0, 1, 2, 3, 4, 5, \ldots\}$
- Continuous distributions
  - ______________________: the famous, symmetric “bell-shaped” ______________. Possible values are all real numbers, $(-\infty, \infty)$
  - _____________________: the (natural) logarithm of observations from this distribution follow a _______________ distribution. Possible values are all positive real numbers, $(0, \infty)$

#### 2.2.2.1 Normal

You may already be familiar with this distribution with the bell-shaped curve. Measurement error is one example of something often assumed to follow a normal distribution. The normal distribution is described by two parameters: the ____________, denoted by $\mu$, and the ______________________________, denoted by $\sigma$. If we have a variable (say, something observed in our data like weight), we give the variable a capital letter, typically $X$. If this variable $X$ is normally distributed with mean $\mu$ and standard deviation $\sigma$, we write this as:

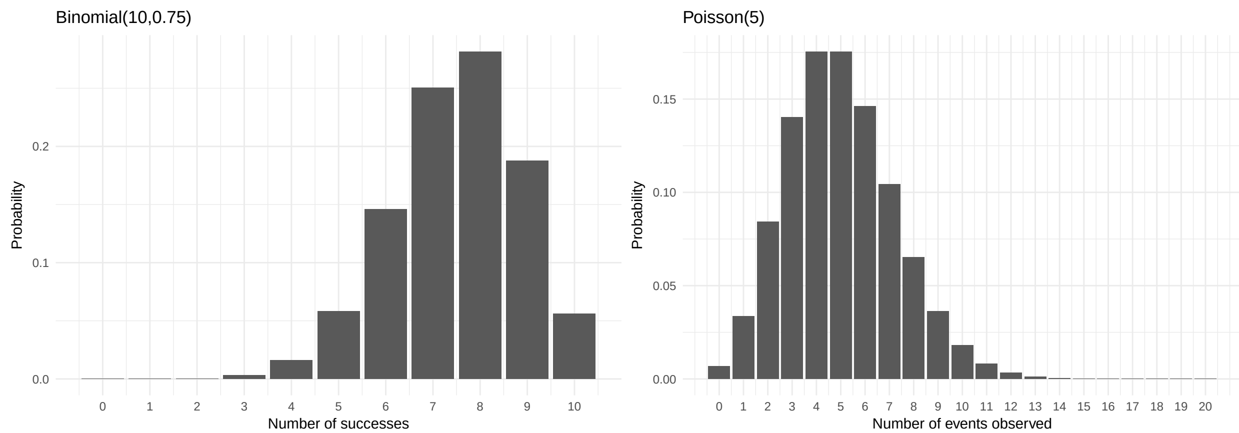


Figure 2.7: On the left, the probability of each possible outcome for variable with binomial distribution with 10 trials and probability of success 0.75. On the right, the probability of each possible outcome for variable with Poisson distribution with mean value 5.

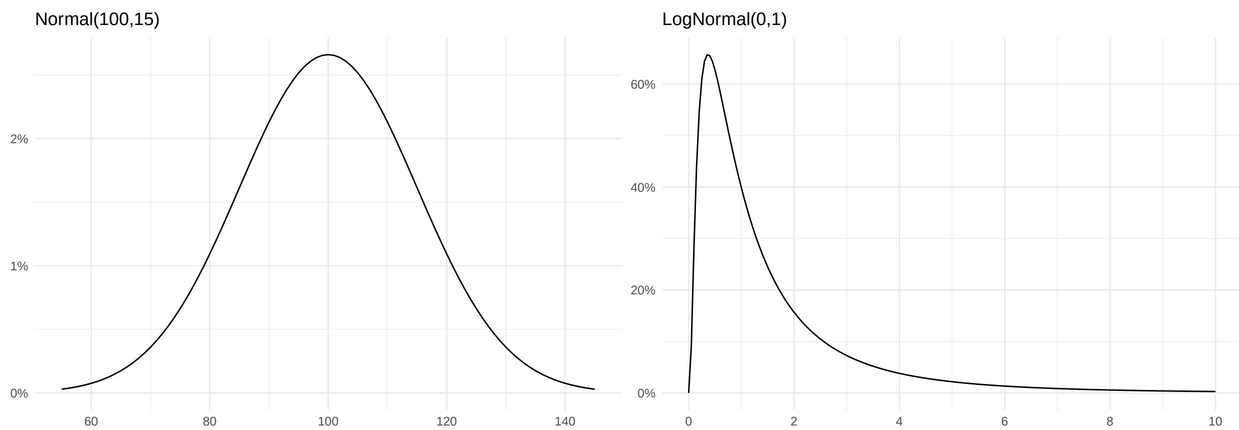


Figure 2.8: On the left, the probability distribution curve of possible outcomes for a variable with Normal distribution with mean value 100 and standard deviation 15. On the right, the probability distribution curve of possible outcomes for a variable with Lognormal distribution with mean value 0 and standard deviation 1.

$$X \sim N(\_\_\_\_, \_\_\_\_)$$

In measurement error, for example, we typically assume that the mean is 0. So, if $X$ represents measurement error, we'd write $X \sim N(0, \sigma)$.

There are many nice properties of the normal distribution. For instance, we know that _______% of observable values lie within _______ standard deviations of the mean ($\mu \pm 2\sigma$), and also that _______% of observable values lie within _______ standard deviations of the mean ($\mu \pm 3\sigma$). When working with the normal distribution, we use software (such as Excel, Matlab, R, SAS, etc.), tables[3], or websites like onlinestatbook.com or stattrek.com to compute probabilities of events.

#### 2.2.2.2 Lognormal

We often act as if everything is normally distributed, but of course this is not true. For instance, a quantity that is certain to be _______________ (greater than or equal to zero) cannot possible be normally distributed.

[3]See for example http://www.stat.ufl.edu/~athienit/Tables/Ztable.pdf

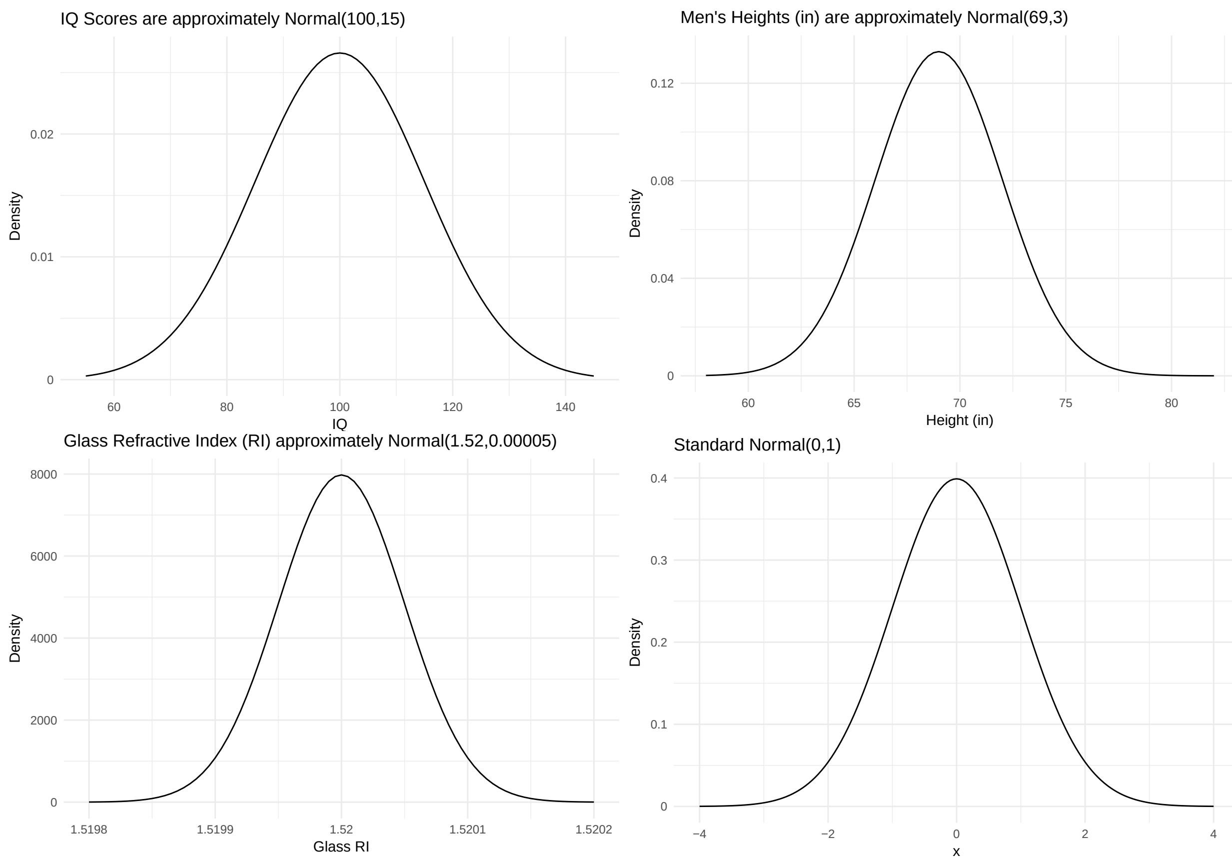


Figure 2.9: Four examples of the probability distribution functions for normally distributed variables

Consider trace element concetration: either none is detected, or there is some amount greater than 0 detected.

In cases where nonnegative values are not possible, we may believe that the (natural) _____________ of the quantity is normal, which gives us a _______________ distribution for the quantity itself. The lognormal distribution, like the normal, has two parameters: mean (on the log scale), denoted _______, and standard deviation (on the log scale), denoted _______.

#### 2.2.2.3 Discrete

Some quantities take on very few possible values. These are *discrete* data.

Recall the two common discrete distributions from section 2.2.2:

- Binomial:
  - Data are ____________ (two categories: "success" or "failure")
  - Data are a result of $n$ independent _____________
  - $P(\text{success}) = p$ on each trial. (Same _____________ of success each time)
  - Expected number of successes you expect to see out of $n$ trials: _______ $\times$ _______
  - Example: Suspect a student of cheating on an exam, response is the number of correct answers.
- Poisson:
  - Data are counts: number of events occurring in a _____________ time

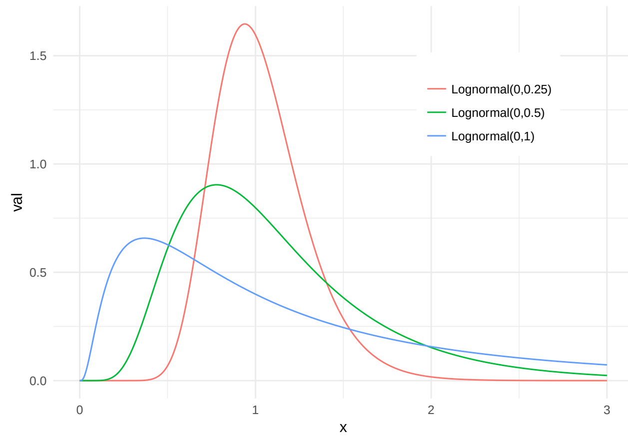


Figure 2.10: Three lognormal distributions with the same mean (on the log scale) and different standard deviations (on the log scale)

  - The mean and the ___________________ of this distribution are the same, so the variablility in responses increases as the ______________ increases.
  - Example: number of calls to 911 between 10:00 and midnight on Friday nights. See Figure 2.11 for a forensics example.

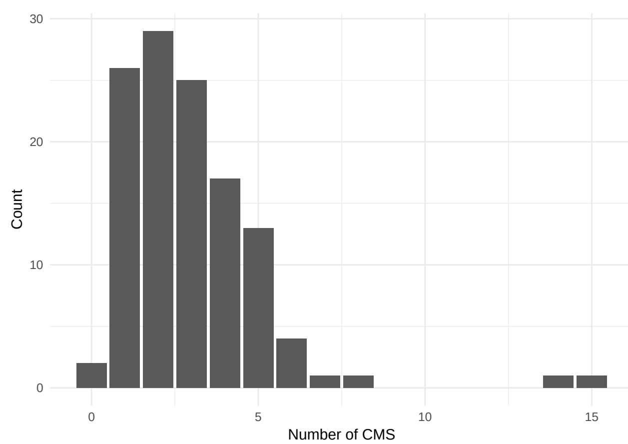


Figure 2.11: Distribution of the maximum number of CMS for a randomly selected bullet compared to 118 known lands approximately follows a Poisson distribution.

## 2.3 Statistical Inference - Estimation

Recall from Section 2.0.1:

- The ____________________ is the universe of objects of interest.
- The ____________________ is comprised of the objects available for study.
- ____________________ is deductive: use knowledge about the population to make statements describing the sample
- ____________________ is inductive: use knowledge about the sample to make statements describing the population

- Probability and statistics are used together!
    1. Build or assume a ____________ for a population
    2. Assess the ______________ using the model
    3. Refine the model, return to step 2.

### 2.3.1 Background

A __________________ is a numerical characteristic of the population, e.g. the population mean. Statistical methods are usually concerned with learning about population parameters from _______________________________.

Note: The mean of a *sample* and the mean of a *population* are differenct concepts. The mean of a sample can be calculated exactly, while the mean of a population is (usually) unknown, because there are too many objects in the population to record and calculate the mean.

The idea underlying statistical inference is that we can apply laws of probability to draw _______________ about a population from a sample. This process is briefly summarized below:

- Observe ______________ mean
- If we have a "good" sample this sample mean should be close to the _________________ mean.
- The laws of __________________ tell us how close we can expect them to be.

For example, suppose we are interested in the average height of the adult population in the U.S.

- Population: _______________________
- Sample: _______________________
- We can take the average height of everyone here and use this sample _________________ to make _________________ about the __________________ mean of all U.S. adults.
- Note: This approach will work if our sample is a ______________ sample from the population. This assumption may be questionable, so it should be verified.

The *goal* of statistical inference is __________________ about a _____________________. Different possible parameters are:

- Mean
- Variance
- Proportion

We can also make different types of inferential statements, depending on what question we are trying to answer and how we are going to report our results. We will talk about:

- __________________ estimate: an estimate of a parameter value
- __________________ estimate: a range of plausible values for a parameter
- Hypothesis __________________: examine a specific hypothesis about the true value of a parameter

When you want to do statistical inference, it is always inportant to look at your sample data before proceeding directly to inference. We do this because we want to

1. See general ____________________ in the data
2. Get an idea of the _________________ of the distribution of the data
3. Identify __________________ values and/or errors.

How we look at our data to check for these three things? If our data are ___________________, we look at a table of frequencies or a bar chart of the different outcomes. If our data are __________________, we look at histograms of the values, or numerical summaries such as mean, median, standard deviation, or percentiles.

A quick example shows why it can be important to examine your data before a formal statistical analysis:

- Suppose the data are (19,20,21,22,23). Then, the mean is $\frac{19+20+21+22+23}{5} = 21$, the median is 21, and the standard deviation is 1.58.
- But what if the data you receive are (19,20,21,22,93)? Then, the mean is $\frac{19+20+21+22+93}{5} = 35$, the median is 21, and the standard deviation is 32.4.
- There could have been a typo, or someone interpreted some handwriting wrong, etc. The moral of the story is ALWAYS look at your data first!

### 2.3.2 Point Estimation

An _______________ is a rule for estimating a population _______________ from a sample. We evaluate the quality of the estimator by considering two key properties:

- Bias: how close ________________________ an estimator is to the true population mean
- Variability: how _______________ is the estimate?

For the population mean, we might use sample mean as an estimator because it has ________ bias and ____________ variability is the sample is ______________. There are other possible estimators for the mean such as:

- The median (good for skewed data or data with outliers)
- The midrange ($\frac{\max + \min}{2}$)
- 47 (obviously this is just guessing and is not advised)

Let $\theta$ denote an unknown population parameter that we wish to estimate. The letter $\theta$ represents the true value of the parameter. In Figure 2.12, we see what would happen in many repeated attempts to estimate $\theta$ using estimators with different properties.

### 2.3.3 Standard Errors

One limitation of just providing a point estimate is that it doesn't give us any indication of ____________________. As we saw in Figure 2.12, a point estimate alone can be very different from the true mean. We can do better than this!

The __________________________________________ of an estimator measures the uncertainty in our estimate. When looking at a summary statistic, like mean, median, or percentiles, that statistic is also a __________________________ quantity. This means that if we had observed a different set of sample values, we would observe different values of the summary statistics. The idea of standard error is similar to the idea of standard deviation. Both are measures of spread, or variability. The difference is that standard deviation is a measure of variability of a sample or population, while standard error is a measure of variability of an ______________________.

Consider a population with that is normally distributed with mean 100 and standard deviation 15. Recall from Section 2.2.2.1 that 95% of observations from a normal distribution fall within two standard deviations of the mean. So, in this example we expect 95% of observations to be between 70 and 130. This distribution is shown on left in Figure 2.13. Now suppose we want to look at the distribution of the *estimates* of the population mean from several samples. This will demonstrate the idea of standard error, using sample size of $n = 25$. The formula to compute standard error is:

$$se = \frac{\quad}{\sqrt{\quad}}$$

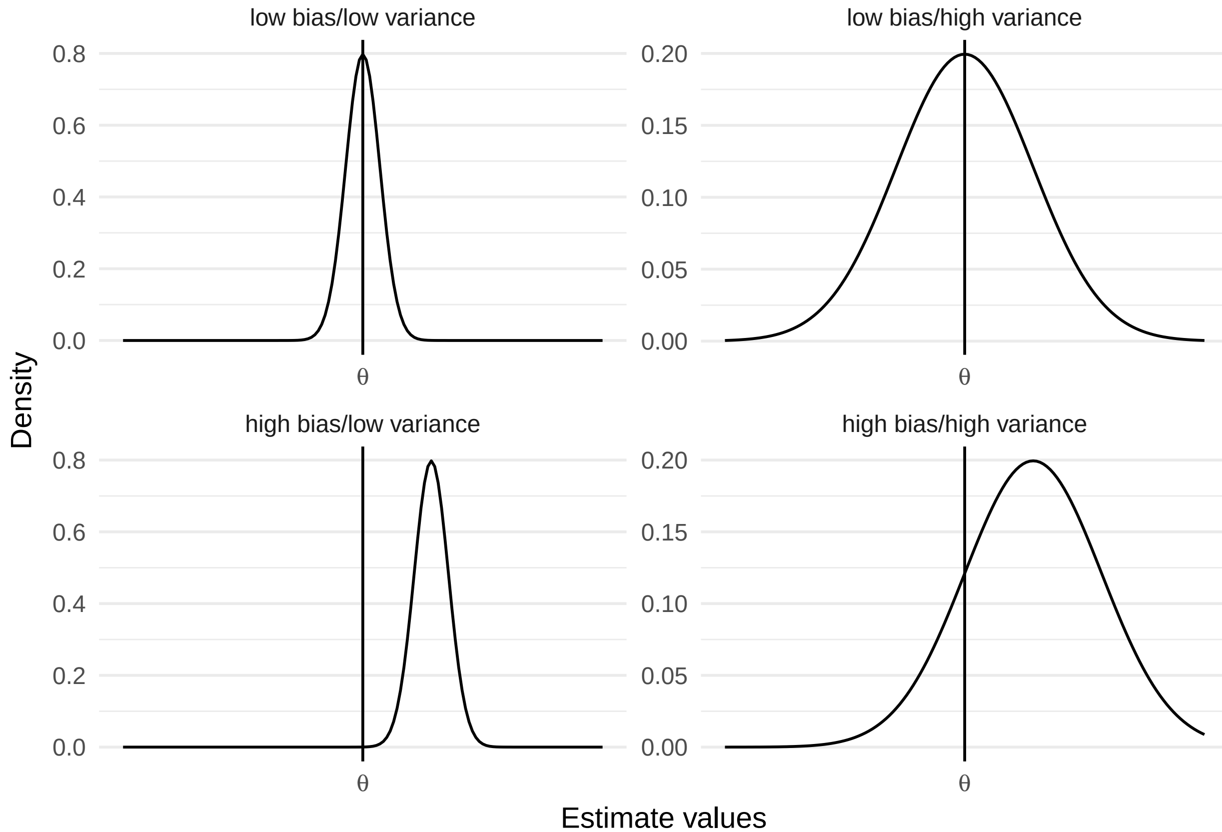


Figure 2.12: The curve in each plot shows the distribution of estimates we would see under each condition shown.

The standard error for this example is $\frac{15}{\sqrt{25}} = 3$. The mean of a sample of size 25 from this population should be about ____________. And about 95% of the time, the sample mean will be between ________ and ________. This distribution is shown on right in Figure 2.13.

### 2.3.4 Sample Size

The size of a sample plays a *critical* role in determining how accurate we can be. Again, consider a population with distribution $N(100, 15)$. We can use _______________ to examine the effect of sample size. We simulate samples from a normal distribution with mean 100 and standard deviation 15. We use four different sample sizes: 10, 25, 50, and 100. We take 500 samples of each size and compute the mean for each sample, leaving us with 2,000 means that we have calculated. We show histograms of the means of these samples for each sample size in Figure 2.14.

### 2.3.5 Interval Estimation

A _________________ interval is an interval based on sample data that, with some specified confidence ______________, contains a population parameter. Essentially, a confidence interval takes a _________________ estimate and then adds some information about ______________________. Typically, we get an approximate ________% confidence interval for a quantity by taking a point estimate and adding and subtracting 2 standard errors from it.

The most well-established procedures for finding confidence intervals are those related to drawing conclusion

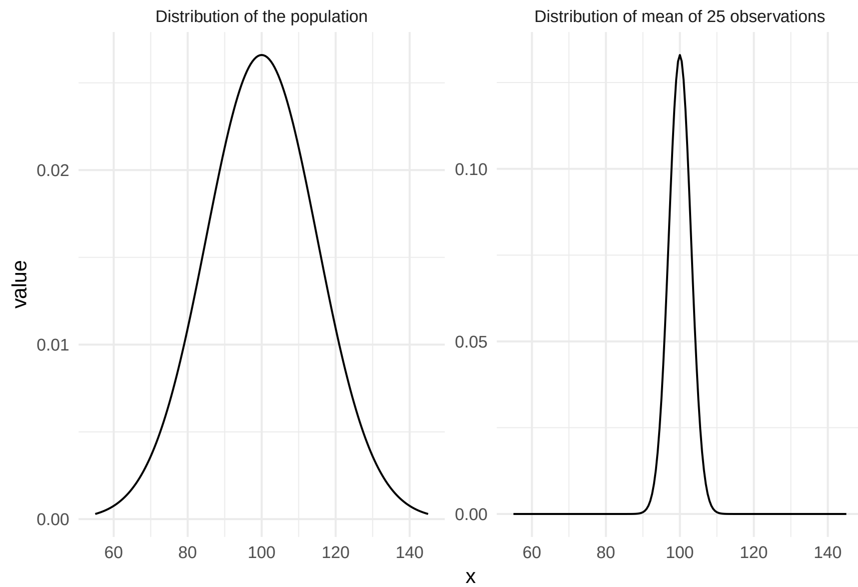


Figure 2.13: On the left, the distribution of the population, which is distributed N(100,15). On the right, the distribution of the sample mean for samples of size 25 from the populations, which is distributed N(100,3).

about the mean of a ________________ population. Suppose that we have acquired a random sample of $n$ observations from a normal population.

- **Point** estimate: the natural point estimate of the population mean is the sample mean, as we've seen already. The sample mean is often denoted by ________. To calculate the sample mean, simple add up all the values and divide by the number of values you have, $n$. In mathematical notation this is written:

$$\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

- **Interval** estimate: denote the standard error of the sample mean by $SE(\overline{X})$, and the standard deviation of the population by $SD$(population). Then, compute the standard error by:

$$SE(\overline{X}) = \frac{SD(\text{population})}{\sqrt{n}}$$

Then, an approximate 95% confidence interval is computed:

$$\overline{X} \pm 2 \cdot SE(\overline{X}) \tag{2.9}$$

A key result is that these procedures for point and interval estimation work well even if the population does **not** follow a ________________ distribution **as long as the sample is** ___________________.

For example, suppose there are 10 glass fragments found at a crime scene, and the concentration of aluminum in each one is measured. The mean aluminum concentration of the sample was 0.73 and the standard deviation was 0.04. The standard error is thus:

$$SE(\overline{X}) = \frac{\quad\quad}{\sqrt{\quad\quad}} = 0.013$$

The approximate 95% confidence interval for the mean aluminum concentration in the crime scene window is:

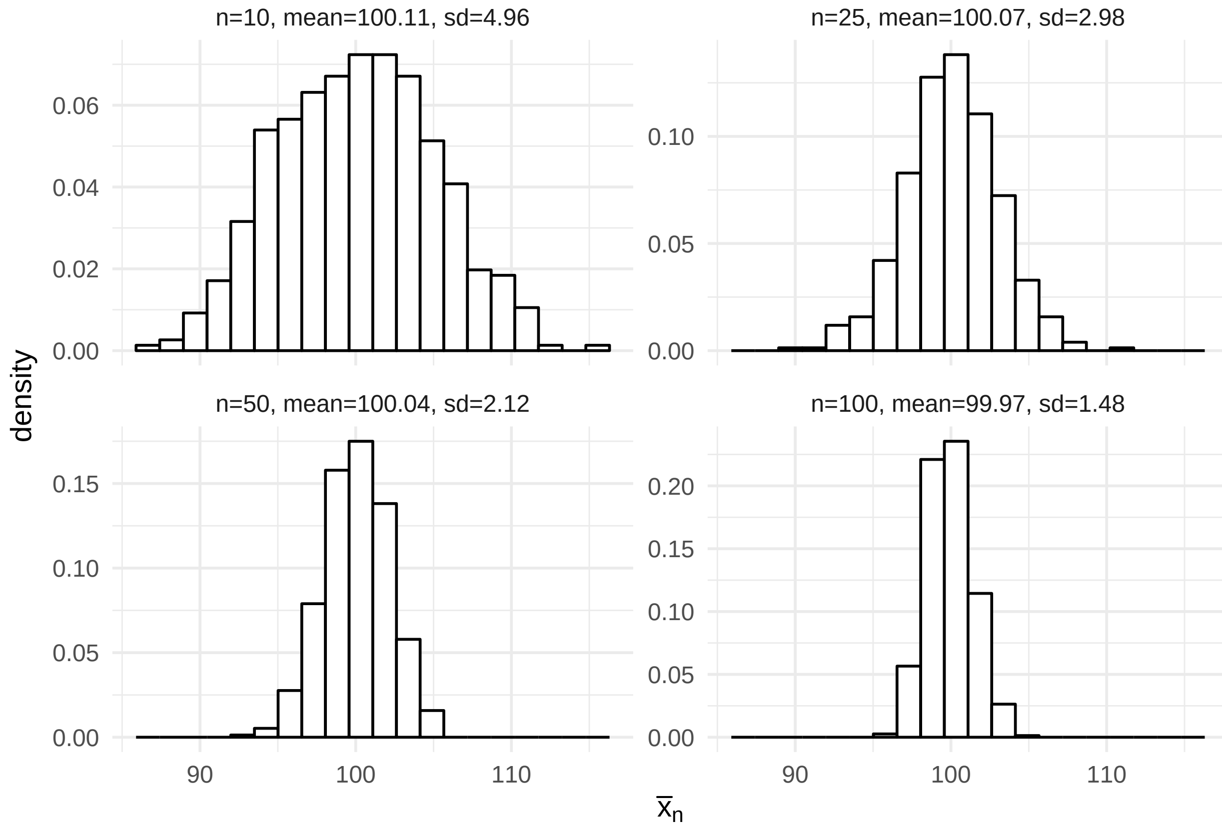


Figure 2.14: How does sample size affect the sampling distribution? As sample size increases, the standard error decreases, so the distribution of sample means becomes more narrow.

$$0.73 \pm 1.96 \cdot 0.013 = (\_\_\_\_, \_\_\_\_)$$

The *interpretation* of the confidence interval is important: 95% of the intervals ______________ in this way will contain the _________ population parameter.

## 2.4 Statistical Inference - Hypothesis Testing

Sometimes we wish to formally ______________ a hypothesis about a population parameter. The hypothesis to be evaluated is known as the ___________ hypothesis. This hypothesis is usually the status quo, or what is assumed to be true. When hypothesis testing we look for evidence ______________ the null. There is also a an ___________________ or research hypothesis that helps us design the test. If we ______________ the null hypothesis, then we say that we have a statistically ___________________ result.

As with anything in life, errors are possible in hypothesis testing. There are two main typs of errors that we care about:

1. Type I: _______________ the null hypothesis when it is ______________. (false positive)
2. Type II: ________________________________ the null hypothesis when it is ______________. (false negative)

Of these two errors, Type I error is often considered more serious. We only want to ________________ the null hypothesis is there is *strong* evidence against it. These statistical testing ideas are closely related to concepts in the __________________________________:

- The null hypothesis: the defendant is ________________
- The alternative hypothesis: the defendant is _______________
- In court, a Type I error is to find guilty when the defendant is __________________
- In court, a Type II error is to find innocent when the defendant is __________________

Ultimately, the basic idea of hypothesis testing is to compute a __________________________ that measures "distance" between the _________ we have collected and what we expect under the _________ hypothesis. Typically, we use a test statistic of the form:

$$\frac{\text{point estimate} - \text{null hypothesis value}}{SE(\text{estimate})} \tag{2.10}$$

This test statistic can be interpreted as the number of _________________________ the sample estimate is from the ___________________ value under the null hypothesis.

A common way to to summarize hypothesis tests is by attaching a ________________ to the test statistic. This probability is called a __________________. The $p$-value of a hypothesis test gives the probability that we would get data like that in our sample (or something even more ______________), given our assumption that the null hypothesis is ___________. This idea is demonstrated in Figure 2.15.

Small $p$-values mean that we have observed ____________ data that lead us to question the ____________ hypothesis, which we have *assumed to be true*. Small $p$-values tell us that the sample data are unlikely to happen by chance under the null. A $p$-value, however, *only* addresses the ____________ hypothesis. It does *not* speak to the likelihood of the _______________ hypothesis being true.

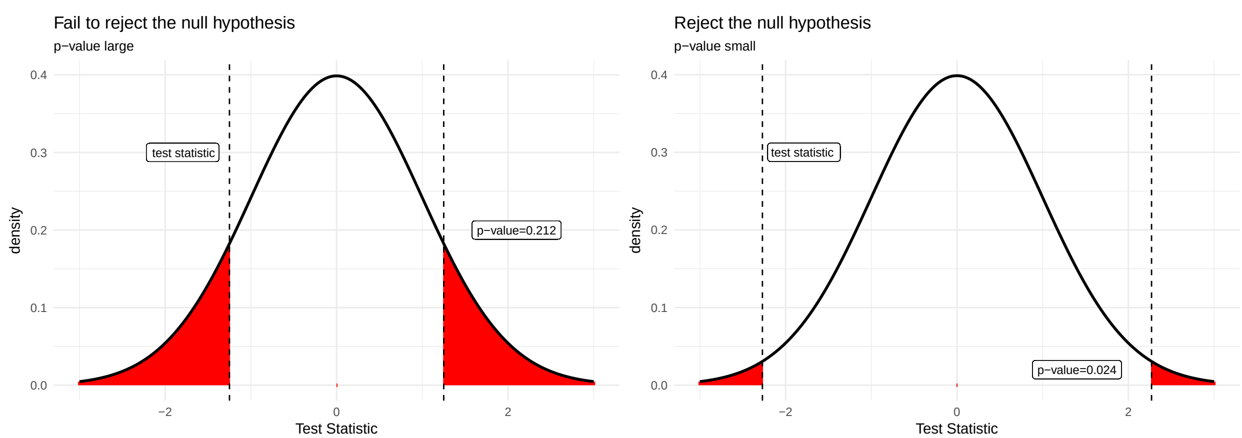


Figure 2.15: On the left, an example of a $p$-value that is large (test statistic small), leading us to fail to reject the null hypothesis. On the right, an example of a $p$-value that is small (test statistic large), leading us to reject the null hypothesis.

### 2.4.1 Normal Data

The most well-established hypothesis testing procedure is for testing a hypothesis about the _____________ of a __________________ population. This population parameter is denoted $\mu$. We can test the hypothesis that the population mean, $\mu$, is equal to some specified values $\mu_0$. The hypotheses for this test are:

- Null hypothesis - $H_0 : \mu = \mu_0$
- Alternative hypothesis - $H_A : \mu \neq \mu_0$

The test statistic, call it $T$, to perform this hypothesis follows the form of Equation 2.10:

$$t = \frac{\_\_\_\_ - \_\_\_\_}{SE(\_\_\_\_)} \tag{2.11}$$

The $p$-value for this hypothesis test is obtained from a $t$ distribution (see Figure 2.16) using software, a table of values, or an online calculator. These hypothesis testing procedures work well even if the population is *not* normally distributed **as long as the sample size is ____________________**.

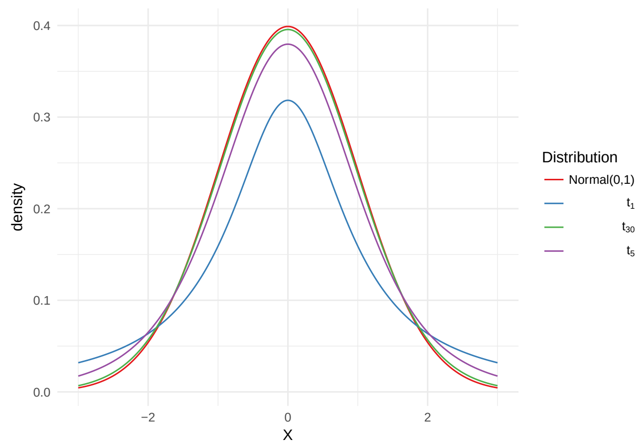


Figure 2.16: The $t$ distribution is simlar to the Normal distribution (red line) is shape, but the tails are higher. As the sample size increases, the $t$ distribution approaches the the Normal distribution.

#### 2.4.1.1 Example

Suppose we want to estimate the mean amount of a trace element for the population of all bullets in Iowa. We get a random sample of 400 bullets from the state.

- The sample mean concentration is $\overline{X} = 55.5$
- The sample standard deviation is $s = 22.0$
- The standard error of the mean is $SE(\overline{X}) = \frac{22}{\sqrt{400}} = \frac{22}{20} = \_\_\_\_\_\_$

Suppose we have reason to believe that Remington (mean = 58) is the main producer in this area. We can check this idea with a hypothesis test.

- Null hypothesis - $H_0 : \mu = 58$
- Alternative hypothesis - $H_A : \mu \neq 58$
- Test statistic - $t = \frac{\overline{X} - \mu_0}{SE(\overline{X})} = \frac{55.5 - 58}{1.10} = -2.27$

The value of the test statistic is more than two standard errors away from the mean under the null hypothesis. The exact $p$-value is 0.023. This means that if the null hypothesis is true, then observing a value 2.27 standard errors or more away from the mean happens only 2.3% of the time. So, we reject our assumption that the population mean concentration is 58.

We can also calculate a 95% confidence interval for the mean using Equation 2.9:

$$55.5 \pm 1.96 \cdot 1.10 = (53.3, 57.7)$$

The hypothesized value (58) is is not in the 95% confidence interval, which also suggests that population mean concentration equal to 58 is not possible.

### 2.4.2 Confidence Intervals

There is a very close relationship between tests and interval estimates. Recall that a confidence interval (CI) gives a range of plausible values for the true population parameter, which here is the mean. A hypothesis test evaluates whether a specified ($\mu_0$) is a ________________ value for the mean. A CI collects all values of $\mu_0$ that we would find plausible in a test.

Statistical hypothesis test are *very* popular in practice. Sometimes, they address the scientific question of interest, but often they do not.[4]

### 2.4.3 Comparing Two Means

In section 2.4.1, we discussed hypothesis testing methods for one sample. In practice, we are often interested in comparing ________ samples from _______ different populations. For now, assume we have random samples from each of the two populations that we are interested in. The test we want to do is a test for ___________________ of parameters in the two populations.

#### 2.4.3.1 Example

Suppose, for example, that we have collected broken glass at a crime scene, and glass fragments on a suspect. Define $\mu_{scene}$ to be the mean trace element level for population of glass at the scene. Define $\mu_{suspect}$ to be the mean element level for the population of glass on the suspect. We can compare the means to address the question of whether or not the glass fragments on the suspect could plausibly have come from the crime scene.

Hypotheses:

- $H_0 : \mu_{scene}$ ____ $\mu_{suspect}$
- $H_A : \mu_{scene}$ ____ $\mu_{suspect}$

Suppose 10 glass fragments are taken from the glass found at the scene (denote these by $Y$), and 9 fragments are found on the suspect (denote these by $X$). Concentrations of a trace element were measured in each fragment of glass. Summary values from the samples are:

- $\overline{X} = 5.3$
- $s_X = 0.9$
- $SE(\overline{X}) = \frac{s_X}{\sqrt{n_X}} = \frac{0.9}{\sqrt{10}} = 0.28$
- $\overline{Y} = 5.9$
- $s_Y = 0.85$
- $SE(\overline{Y}) = \frac{s_Y}{\sqrt{n_Y}} = \frac{0.85}{\sqrt{9}} = 0.28$
- Obeserved difference $= \overline{X} - \overline{Y} =$ ____ $-$ ____ $= 0.6$
- The standard error for the *difference*, $\overline{X} - \overline{Y}$, is

$$SE(\{\overline{X} - \overline{Y}\}) = \sqrt{SE(\overline{X})^2 + SE(\overline{Y})^2} = \sqrt{\_\_\_\_ + \_\_\_\_} = 0.4$$

[4]For more reading on this topic, consider this article from *Nature*: http://www.nature.com/news/psychology-journal-bans-p-values-1.17001

- The test statistic for this hypothesis test is:

$$t = \frac{(\overline{X} - \overline{Y}) - 0}{SE(\{\overline{X} - \overline{Y}\})} = \frac{0.6}{0.4} = 1.5$$

- The corresponding *p*-value for this statistic is 0.15.
- So, we fail to reject the null hypothesis that the two glass population means are equal.
- Interpretation is a *key* issue. When we say we fail to reject the null, we are saying there is a possibility of a common source.

### 2.4.4 Discussion

There are three key points for you to take away:

1. Hypothesis testing does *not* treat the two hypotheses ____________________. The null hypothesis is given priority. This is appropriate when there is reason to ________________ the null hypothesis until there is significant evidence _____________________ it. We don't necessarily always want this to be the case. (We will discuss this more later on in a forensic context.)

2. The *p*-values that result from hypothesis tests depend heavily on the sample size. If you have the same ___________ and standard deviation, but _________________ the sample size, the result will me more significant, due to the sample size alone.

3. Interpreting the results of the hypothesis test can be tricky. If we _______________ the null hypothesis, this does not necessarily mean that we have found an important difference in the context of our problem. In addition ________________________________________ the null hypothesis does not necessarily mean the null hypothesis is true.

## 2.5 Overview of Statistical Preliminaries

- We reviewed the basics of probability
  - Probability is the language of uncertainty
  - It is important to understand what is being assumed when talking about probability
  - For instance, the probability of having disease given a positive test is different than the probability of having a positive test given the disease
  - Probability distributions describe the variability in a population or in a series of measurements
- We reviewed basics of statistical inference
  - Statistical inference uses sample data to draw conclusions about a population
  - Point estimation, interval estimation, and hypothesis tests are main tools
  - It is critical that our procedures account for variation that could be observed due to chance

# Chapter 3

# Statistics for Forensic Science

## 3.1 Brief Review of Probability and Statistics

## 3.2 The Forensic Examination

## 3.3 Common Approaches to Assessing Forensic Evidence

### 3.3.1 Significance Testing / Coincidence Probability

### 3.3.2 Likelihood Ratio

# References

Baldus, David C., Charles Pulaski, and George Woodworth. 1983. "Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience." *Journal of Criminal Law and Criminology*.

Hare, Eric, Heike Hofmann, and Alicia Carriquiry. 2017. "Automatic Matching of Bullet Land Impressions." *The Annals of Applied Statistics* Upcoming.

Morris, Max D. 2011. *Design of Experiments: An Introduction Based on Linear Models*. Chapman; Hall.