Module 3 Lecture - Probability Topics

Introduction to Statistical Methods

Quinton Quagliano, M.S., C.S.P
Department of Educational Psychology

Table of Contents

“I don’t mind not knowing. It doesn’t scare me.” — Richard P. Feynman

1 Overview and Introduction

1.1 Textbook Learning Objectives

- Understand and use the terminology of probability.
- Determine whether two events are mutually exclusive and whether two events are independent.
- Calculate probabilities using the Addition Rules and Multiplication Rules.
- Construct and interpret Contingency Tables.
- Construct and interpret Venn Diagrams.
- Construct and interpret Tree Diagrams.

1.2 Instructor Learning Objectives

- Begin to appreciate how probability connects back to the broader idea of statistics
- Understand the emphasis of chance and probability, not certainty, in statistics

1.3 Introduction

- Most decisions we make are a (hopefully educated) ____________.
  - E.g., Should I wear my boots if it might rain today?, What are the chances that my class will respond well to this activity?
  - Chance and probability is ____________, and, if we introspect closely, it plays a role in pretty much everything we do

Discuss: Try coming up with more examples of weighing probability in your own life

- Probability is a core part of ____________, and our results are often phrases in recognizing that level of uncertainty
  - This is why you never really hear good scientists say we have definitively ____________ something - there's always room left for uncertainty
- There is a lot of ____________ (and math) surrounding probability
  - You could take an entire semester long class just talking about probability and the ____________/math of it

  - We'll try to stay centered on exploring it as relevant to statistics, though this module may initially feel very removed from statistics

> Important
>
> Even though we won't be 'in the weeds' as much with probability calculations later on, it will remain an important background subject!

2 Terminology

2.1 Introduction

- Like with the last module, __________ makes perfect
  - Please use the various exercises in the book and visit GA and/or professor office hours!
  - This is also a great time to embrace flashcards, if that historically works well for you

2.2 Rapid Fire Vocabulary

- In the context of probability, an **experiment** is some __________ controlled and monitored action - similar to how we use the term in research design
  - If the experiment is not **pre-determined**, that is, it's outcome is not __________, then there is some amount of **chance**
  - So if we call something a __________ experiment, its some controlled action that we are not sure of the outcome
  - Classic example is a coin flip

> ? Example of how probability is intuitive - what is the chance that I flip a coin and it lands on 'tails'
>
> - (●) A) 25%
> - ( ) B) 75%
> - ( ) C) 50%
> - ( ) D) 0%
>
> Explanation:

- The final ______ of the experiment is called an **outcome**
  - The possibilities of ______ the different outcomes is called the **sample space** (notation = S)
  - Sample space is ______ as a *list*, a tree, or Venn diagram; more example of these later
  - Coin flip with possible heads (H)/tails (T) outcome example = $S = \{H, T\}$
- An **event** is some ______ of outcome(s)
  - Important to know that it could just be ______ outcome or more than on described outcome
  - ______ will be defined as upper case letters like A, B, C, etc.
- The **probability** of a certain ______ occurring is described as the **long-term relative frequency** of said event
  - What that means is that the probability describes how ______ a certain event would occur if the experiment were to be completed an infinite number of times; this hinges on the **law of large numbers**
  - The ______ of a certain event happening will be written as $P(x)$, where x is whatever event is being looked at
    - $P(x)$ will always be ≥ 0 and ≤ 1
  - When $P(x)$ of all possible ______ is equal to one another, we would say that this experiment is **equally likely**, sometimes also said to be **fair**
- Example of coin toss, where event A is the likelihood of flipping with an outcome of head on a single toss, and B is the likelihood of tails
  - If $P(A) = 0.50$, i.e., a 50% chance to land on heads
  - If $P(B) = 0.50$, i.e., a 50% chance to land on tails
  - If $P(A) = P(B)$, i.e., equally likely experiment

Discuss: Try writing out the same set of bullets that I just did, but with a six-sided equally likely die instead. We will calculate it out in a second, but try to just do this intuitively now.

2.3 Calculating Probability

- In the scenario that we aim to calculate the probability of a single ______ in an equally likely space:

- Steps to :
    1. Determine all outcomes from the experiment and list them out, counting the number of possible outcomes
    2. a specific event
    3. Assess how outcomes match that events
    4. the number of outcomes that match the event by the total number of outcomes; $P(x) = Matching/Total$
- Do a coin toss of a dime and nickel
    1. 4 outcomes in the sample space: $\{HH, TH, HT, TT\}$ where
        - H heads on a coin
        - T tails on a coin
        - In writing this would be: both coins land on heads, dime lands tail and nickel lands heads, dime lands heads and nickel lands tails, both coints land on tails
    2. Event A of getting at only one heads; could also be written as $A : n_H = 1$
    3. Two outcomes in the sample space match this criteria: $/TH, HT/$
    4. Thus the probability of event A is calculated as $P(A) = 2/4 = 0.50$

Discuss: Trying doing the same full calculation process now, but now for a 6-sided equally likely die.

- Practically, many events are not fair / equal - coin flips are affected by the person doing them, air resistance, etc.
    - This will be called **unfair or biased**

2.4 OR Events

- An **OR Event** is when an event is written in such a manner that an outcome could event A or B
    - This would be written as $A \cup B$ (that symbol is called a cup)
    - Applying the above, we would have a step process of determining the outcome match A OR B and would give the probability as $P(A \cup B)$

- To list the possible outcomes, we find the **set union** of the two sets of outcomes that match the ________________ events, example:
  - If A matches $\{1, 2\}$ and
  - If B matches $\{2, 3, 4\}$ then
  - $A \cup B$ matches $\{1, 2, 3, 4\}$
  - Then we proceed by using $A \cup B$ matches in step 4 of Calculating Probability

2.5 AND Events

- An **AND Event** is when an event is written in a manner that an outcome could match event A or B
  - This would be written as $A \cap B$ (this symbol is a cap, like a hat)
  - Thus probability would be written as $P(A \cap B)$
- To list the possible outcomes, we find the **set intersection** of the two sets of outcomes that match the ________________ events, example:
  - If A matches $\{1, 2, 3\}$ and
  - If B matches $\{2, 3, 4\}$ then
  - $A \cap B$ matches $\{2, 3\}$
  - Then we proceed by using $A \cap B$ matches in step 4 of Calculating Probability

> **Important**
>
> A mnemonic to remember caps and cups: cAps are for Ands

2.6 Conditionals and Complements

- **Complements** take the ________________, or everything but the matching outcomes of an event
  - This is written at the ________________ of an outcome, e.g., A' is the complement of A
  - Thus, probability of a conditional would be notated as $P(x')$
  - In such a scenario, instead of finding all the outcomes that DO match the event, we would list all the outcomes that ________________ match the event in step 3 of Calculating Probability
- **Conditionals** are the probability a certain event occurs ________________ that another event has already happened
  - It is written as $P(A|B)$, which translates to probability of A given B (the vertical bar is sometimes called a pipe)
  - This is calculated via the following formula, assuming $P(B) \neq 0$:

$$P(A|B) = \frac{A \cap B}{B}$$

- Once again, try working through the problems and solutions in your book for all of these scenarios!

3 Independent and Mutually Exclusive Events

3.1 Introduction

- We are often interested in how two possible events are ____________ related or unrelated
  - We can investigate both the independence and mutually exclusivity of events to determine this
  - These are both ____________ description

3.2 Independence

- Two events are considered to be **independent** if they are unrelated/do not ____________ one another
  - On the contrary, if they are somehow ____________ to one another, they are said to be **dependent**
- The following are the ____________ rules to show independence:
  - $P(A|B) = P(A)$
  - $P(B|A) = P(B)$
  - $P(A \cap B) = P(A)P(B)$ (The "Multiplication Rule")
  - Only one of these conditions needs to be met for establishing independence
- It is safest to ____________ dependence of events until independence has been established
  - Sort of counter-intuitive!

3.2.1 Connection to Sampling Process

- Recall that in a *random* sampling scenario, there is a theoretically ____________ chance for each member of the population of interest to be selected
  - Effectively, random sampling is sampling that is done *equally likely*
- Sampling can be done one of two ways, that impact whether sampling decisions were made ____________ of one another

- Sampling can be done **with replacement** which means the same individual could be chosen __________ times - this means the odds of being selected as a member of the population remain the same
- Sampling can also be done **without replacement**, meaning it is done in a manner that a person __________ be chosen multiple times, thus reducing the outcomes by one and changing the odds for each subsequent selection

3.3 Mutually Exclusivity

- Events are **mutually exclusive** when they cannot occur at the same time and follow the "Addition Rule":

$$P(A \cap B) = 0.00$$

- If unsure, assume __________ until this rule is met

4 Contingency Tables

4.1 Introduction

- The word "contingent" means to depend on, so **contingency tables** offer a nice view for displaying probabilities in such a way to __________ dependence

> **Important**
>
> Like with descriptive statistics, using graphical and tabular displays can make it much easier to navigate complex information, make sure you use them!

4.2 Example of a Contingency Table

	Speeding violation in the last year	No speeding violation in the last year	Total
Uses cell phone while driving	25	280	305
Does not use cell phone while driving	45	405	450
Total	70	685	755

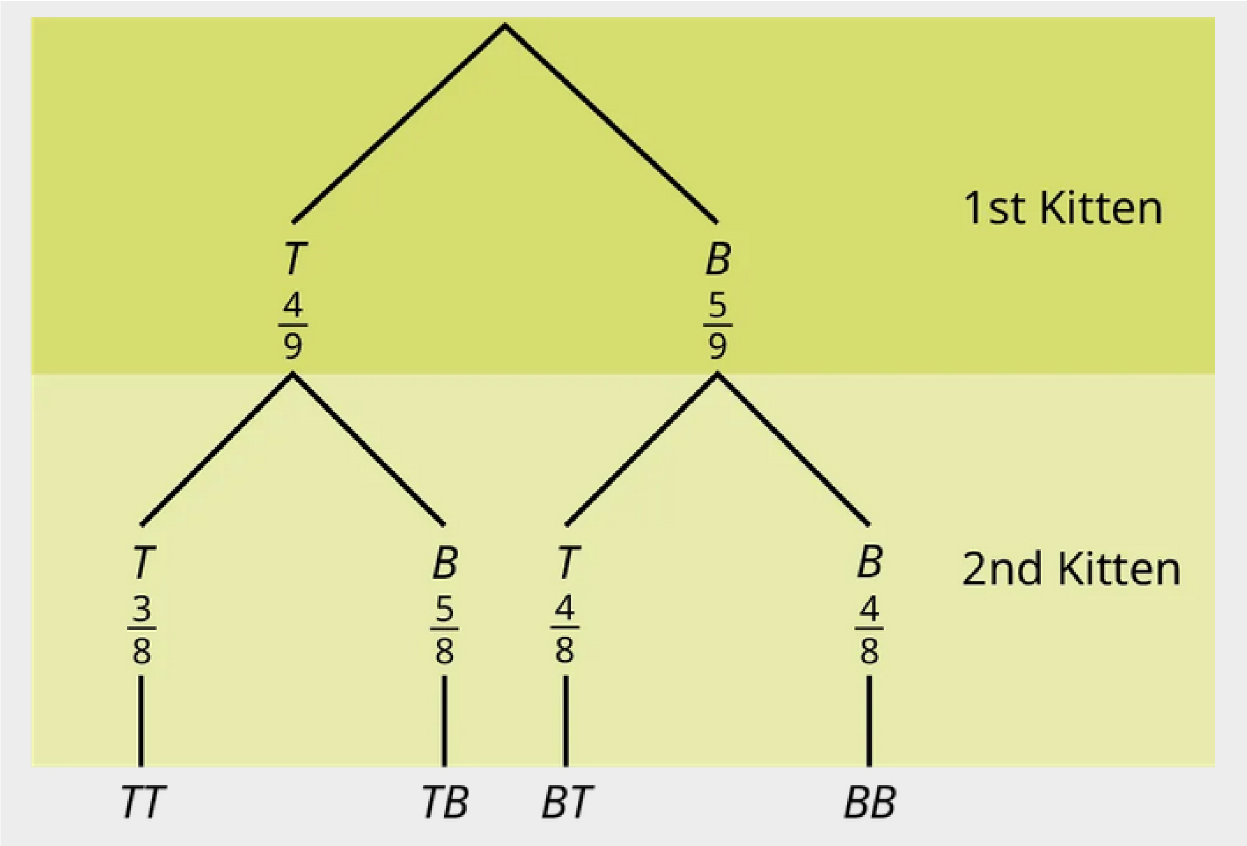
Table 3.3

5 Tree and Venn Diagrams

5.1 Introduction

- Just like with how descriptive statistics have several ways to graphically present, so do probabilities!
  - Contingency tables are a great start, but we can add to them with ________ diagrams and ________ diagrams
- For this section, I'll be a bit more brief only because these representations are not super ________ in most applied statistics

5.2 Tree Diagrams



5.3 Venn Diagrams

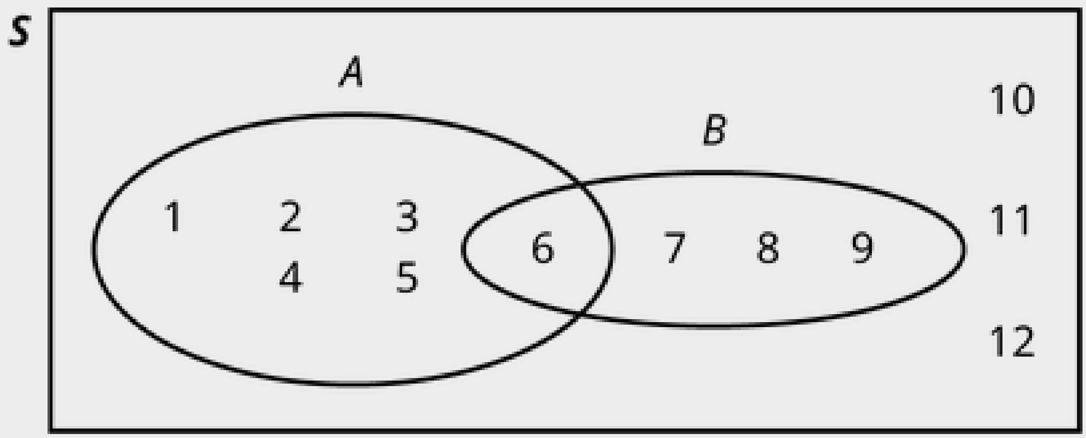


Figure 3.6

6 Conclusion

6.1 Recap

- Probability is a tough subject, but it is critical as a foundation topic prior to understanding inferential statistics!
- Understand that when we discuss statistical results going forward, it will always be in through the lens of probabilities and likelihood of outcomes.
- While "real" statistics won't often involve intensive hand calculation of probabilities, you should still have a foundation appreciation of how to do it - just like we did with standard deviation.

6.2 Lecture Check-in

- Make sure to complete and submit the lecture check-in
- Also make sure to attend to additional SPSS and practical assignment this week!