Applications of Probability

Name: ________________

Date: ________________

Understand independence and conditional probability and use them to interpret data

MCC9-12.S.CP.1 Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events (“or,” “and,” “not”).

MCC9-12.S.CP.2 Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent.

MCC9-12.S.CP.3 Understand the conditional probability of A given B as P(A and B)/P(B), and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B.

MCC9-12.S.CP.4 Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities.

MCC9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations.

Use the rules of probability to compute probabilities of compound events in a uniform probability model

MCC9-12.S.CP.6 Find the conditional probability of A given B as the fraction of B’s outcomes that also belong to A, and interpret the answer in terms of the model.

MCC9-12.S.CP.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answer in terms of the model.

RELATED STANDARDS

Investigate chance processes and develop, use, and evaluate probability models.

MCC7.SP.5 Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.

MCC7.SP.8a Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs.

MCC7.SP.8b Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams. For an event described in everyday language (e.g., “rolling double sixes”), identify the outcomes in the sample space which compose the event.

Lesson 7.0 Definitions

- **Complement:** Given a set A, the complement of A, denoted A or A' , is the set of elements that are not members of A.
- **Conditional Probability:** The probability of an event A, given that another event, B, has already occurred; denoted $P(A \mid B)$.
- **Dependent Events:** Two or more events in which the outcome of one event affects the outcome of the other event or events.
- **Element:** A member or item in a set.
- **Independent Events:** Events whose outcomes do not influence each other.
- **Intersection of Sets:** The set of all elements contained in all of the given sets, denoted · .
- **Outcome:** A possible result of an experiment.
- **Sample Space:** The set of all possible outcomes from an experiment.

- **Set:** A collection of numbers, geometric figures, letters, or other objects that have some characteristic in common.
- **Subset:** a set in which every element is also contained in a larger set.
- **Union of Sets:** The set of all elements that belong to at least one of the given two or more sets denoted • .
- **Venn Diagram:** A picture that illustrates the relationship between two or more sets.

Lesson 7.1 "Dicey"

Part 1 – For this task you will need a pair of six-sided dice. In Part 1, you will be concerned with the probability that one (or both) of the dice show odd values.
1. Roll your pair of dice 30 times, each time recording a success if one (or both) of the dice show an odd number and a failure if the dice do not show an odd number.

Number of Successes	Number of Failures

2. Based on your trials, what would you estimate the probability of two dice showing at least one odd number? Explain your reasoning.

3. You have just calculated an *experimental probability*. 30 trials is generally sufficient to estimate the *theoretical probability*, the probability that you expect to happen based upon fair chance. For instance, if you flip a coin ten times you expect the coin to land heads and tails five times apiece; in reality, we know this does not happen every time you flip a coin ten times.

A lattice diagram is useful in finding the theoretical probabilities for two dice thrown together. An incomplete lattice diagram is shown below. Each possible way the two dice can land, also known as an **outcome**, is represented as an ordered pair. (1, 1) represents each die landing on a 1, while (4, 5) would represent the first die landing on 4, the second on 5.

(1,1)	(1,2)	(1,3)	(1,4)		
(2,1)	(2,2)				
(3,1)					

a. Why does it have 36 spaces to be filled?

b. Complete the lattice diagram of the previous page for rolling two dice.
The 36 entries in your dice lattice represent the *sample space* for two dice thrown. The sample space for any probability model is all the possible outcomes.

c. It is often necessary to list the sample space and/or the outcomes of a set using *set notation*. For the dice lattice above, the set of all outcomes where the first roll was a 1 can be listed as: {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)}. This set of outcomes is a **subset** of the set because all of the *elements* of the subset are also contained in the original set.
Using your lattice, give the subset that contains all elements that sum to 9.

d. Using your lattice, what is the probability that the sum of two die rolled will be 9?

e. Using your lattice, determine the probability of having at least one of the two dice show an odd number.

f. Using your lattice, determine the probability of having at least one of the two dice show the number 6.

g. Using your lattice, determine the probability of having both dices show the number 4.

h. Using your lattice, determine the probability of having any one of the dices showing the number 2.

i. Using your lattice, determine the probability of having one dice showing a 3 and the other showing a 4.

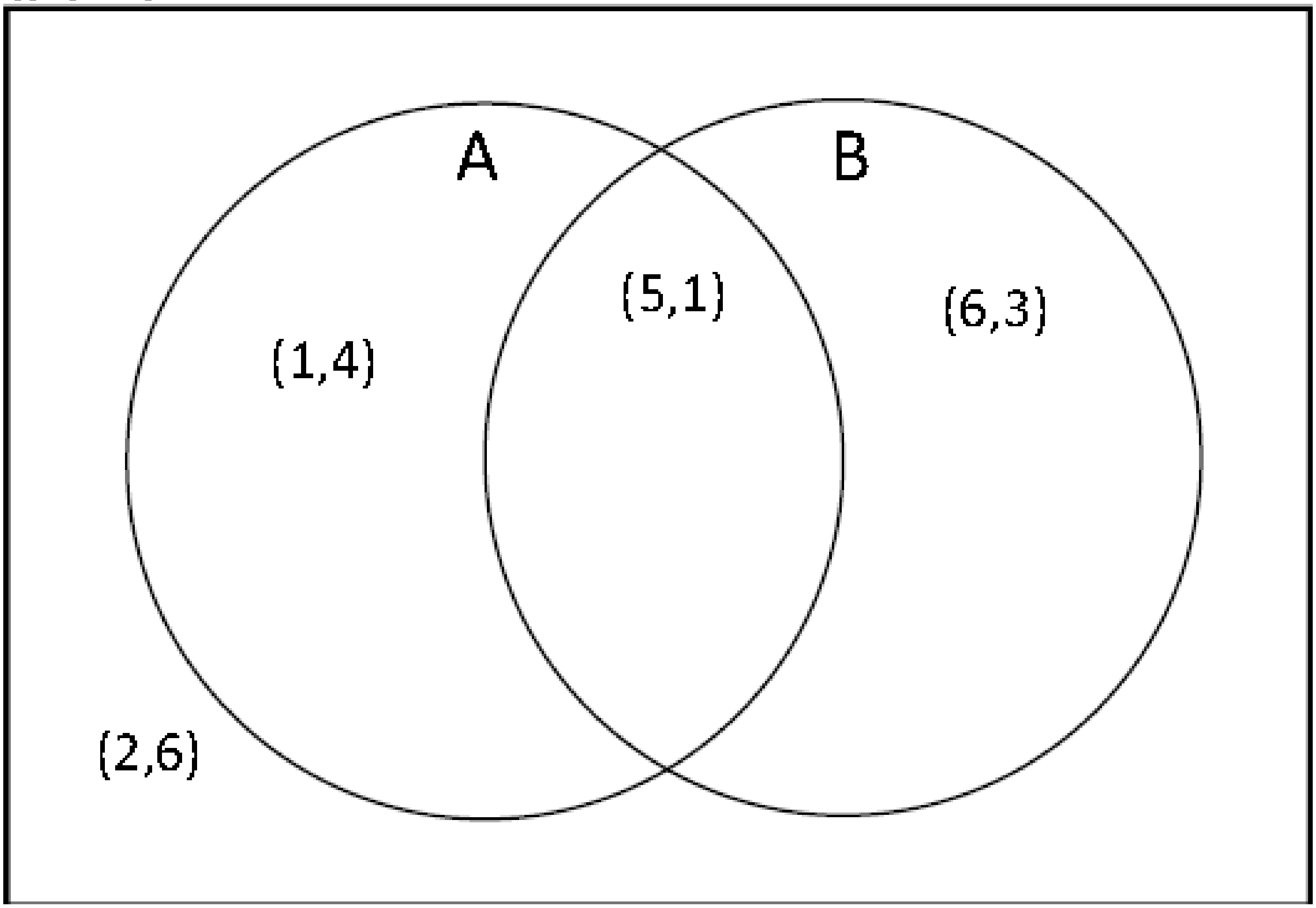
j. Using your lattice, determine the probability that the sum of the two die is 4 or less.

Using your lattice, determine the probability of having the compliment of (g). The **compliment of an event** is all the other possible outcomes of the respective experiment. In other words the complement of an event is all outcomes that are NOT the event. It refers to the probability that an event will **not** occur. It is easily calculated by subtracting the probability that the event will occur from 1. If the probability for an event is $P(A)$, then the compliment is denoted by $P(\overline{A})$. Formula: $P(\overline{A}) = 1 - P(A)$.

i. Using your lattice, determine the probability of having the compliment of (e).

4. The different outcomes that determine the probability of rolling odd can be visualized using a Venn Diagram, the beginning of which is seen below. Each circle represents the possible ways that each die can land on an odd number. Circle A is for the first die landing on an odd number and circle B for the second die landing on odd. The circles overlap because some rolls of the two dice are successes for both dice. In each circle, the overlap, and the area outside the circles, one of the ordered pairs from the lattice has been placed. (1,4) appears in circle A because the first die is odd, (6,3) appears in circle B because the second die is odd, (5,1) appears in both circles at the same time (the overlap) because each die is odd, and (2,6) appears outside of the circles because neither dice is odd.

a. Finish the Venn Diagram by placing the remaining 32 ordered pairs from the dice lattice of page 2 in the appropriate place.



b. How many outcomes appear in circle A? (Remember, if ordered pairs appear in the overlap, they are still within circle A).

c. How many outcomes appear in circle B?

d. The portion of the circles that overlap is called the **intersection**. The notation used for intersections is $\cap$. For this Venn Diagram the intersection of A and B is written $A \cap B$ and is read as "A intersect B" or "A and B." How many outcomes are in $A \cap B$?

e. When you look at different parts of a Venn Diagram together, you are considering the **union** of the two outcomes. The notation for unions is $\cup$, and for this diagram the union of A and B is written $A \cup B$ and is read "A union B" or "A or B." In the Venn Diagram you created, $A \cup B$ represents all the possible outcomes where an odd number shows. How many outcomes are in the union?

f. Record your answers to b, c, d, and e in the table below.

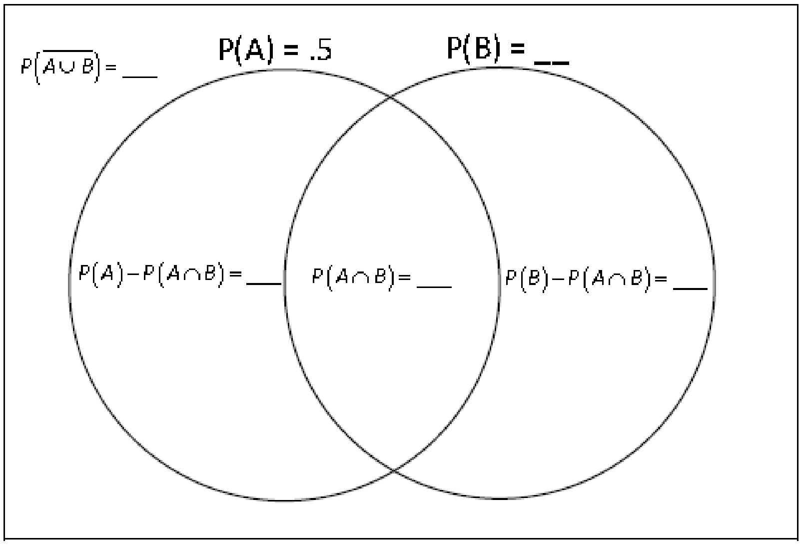
b. Circle A	c. Circle B	d. $A \cap B$	e. $A \cup B$

g. How is your answer to e related to your answers to b, c, and d?

h. Based on what you have seen, make a conjecture about the relationship of A, B, $A \cap B$ and $A \cup B$ using notation you just learned.

i. What outcomes fall outside of $A \cup B$ (outcomes we have not yet used)? Why haven't we used these outcomes yet?

j. Venn Diagrams can also be drawn using probabilities rather than outcomes. The Venn Diagram below represents the probabilities associated with throwing two dice together. In other words, we will now look at the same situation as we did before, but with a focus on probabilities instead of outcomes.



Fill in the remaining probabilities in the Venn Diagram.

k. Find $P(A \cup B)$ and explain how you can now use the probabilities in the Venn Diagram rather than counting outcomes.

l. Use the probabilities in the Venn Diagram to find $P(\overline{B})$.

m. What relationship do you notice between $P(B)$ and $P(\overline{B})$? Will this be true for any set and its complement?

Part 2 – Venn Diagrams can also be used to organize different types of data, not just common data sets like that generated from rolling two dice. In this part of the task, you'll have an opportunity to collect data on your classmates and use a Venn Diagram to organize it.

1. Music is a popular topic amongst high school students, but one in which not all can agree upon. Let's say we want to investigate the popularity of different genres of music in your math class, particularly, Hip Hop and Country music. What genre of music do you enjoy listening to: Hip Hop, Country, or Neither?

2. Imagine you take a poll in your classroom. Any student who listens to both Country and Hip Hop may be listed in both categories. The results of the class poll are listed in the table below.

HIP HOP	COUNTRY	NEITHER
X		
	X	
	X	
		X
X	X	
X	X	
X		
		X
	X	

3. Draw a Venn Diagram to organize your outcomes. (*Hint: Students listed in both the Hip Hop and Country categories should be identified first prior to filling in the diagram.)*

4. Find P(HH) .

5. Find P(C).

6. Find $P(\overline{C})$.

7. Find $P(HH \cap C)$.

8. Find $P(HH \cup C)$.

Part 3 – Now that you have had experience creating Venn Diagrams on your own and finding probabilities of events using your diagram, you are now ready for more complex Venn Diagrams.

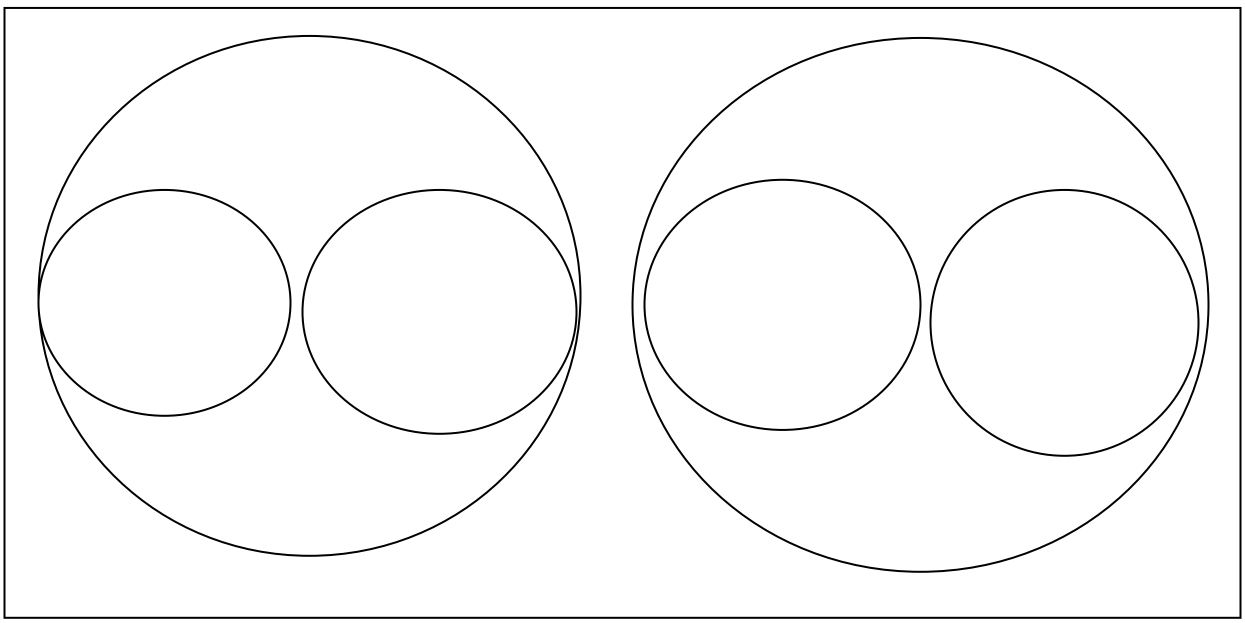
1. In this part of the task, you will be examining data on the preference of social networking sites based on gender.

Which social networking site do you prefer?

2. Imagine you have following results from a class poll.

	Twitter(T)	Facebook(FB)
Female(F)	13	15
Male(M)	8	5

3. Fill in the correct numbers for the Venn Diagram below to organize your outcomes. *(Hint: Notice that male and female will not overlap and neither will Twitter and Facebook).*



3. Find $P(\overline{T})$.

4. Find P(T $\cup$ M).

5. What is another way to write the probability of P(T $\cup$ M) using a complement?

6. Find $P(\overline{FB} \cap F)$.

7. Find $P(T \cup M) + P(\overline{T \cup M})$.

Let's summarize unions and intersections in probability

Suppose we are given an experiment with sample space S. Let A and B be events in S and let E be the event "either A occurs or B occurs". Then E occurs if the outcome of the experiment is either in A, or in B, or in both A and B.
If **either event A or event B occurs** we use the notation: $E = A \cup B$ (Formula: $A \cup B = A + B - A \cap B$)

Suppose we are given an experiment with sample space S. Let A and B be events in S and let F be the event "both A and B occur".
If **both A and B occur** we use the notation: $F = A \cap B$

Example
Let us roll a fair die. The sample space of equally likely simple events is:
S = {1, 2, 3, 4, 5, 6}.
Let A be the event "an odd number" and let B be the event "the number is divisible by 3".

(a) Find the probability of the event E = the number is odd **or** is divisible by 3.

Solution: $A = \{1,3,5\}$ = Event of an odd number
$B = \{3,6\}$ = Event of a number divisible by 3

The event A or event B to occur is: $E = A \cup B = \{1,3,5,6\}$

The probability for event A or event B to occur is $\dfrac{\text{eventA or event B to occur}}{\text{sample space of equally likely simple events}}$

Or using math and probability notation: $P(E) = \dfrac{A \cup B}{S} = \dfrac{\{1,3,5,6\}}{\{1,2,3,4,5,6\}} = \dfrac{4}{6} = \dfrac{2}{3}$

Therefore, the probability to roll an odd number **or** a number which is divisible by 3 is $\frac{2}{3}$.

(b) Find the probability of the event F = the number is odd **and** is divisible by 3.

Solution: $A = \{1,3,5\}$ = Event of an odd number
$B = \{3,6\}$ = Event of a number divisible by 3
The event A and event B to occur is: $F = A \cap B = \{3\}$

The probability for event A and event B to occur is $\dfrac{\text{eventA and event B to occur}}{\text{sample space of equally likely simple events}}$

Or using math and probability notation: $P(F) = \dfrac{A \cap B}{S} = \dfrac{\{3\}}{\{1,2,3,4,5,6\}} = \dfrac{1}{6}$

Therefore, the probability to roll an odd number **and** a number which is divisible by 3 is $\frac{1}{6}$.

Problems:

1. You roll a fair die. What is the probability that you roll an even number **or** a number greater than 3?

2. You roll a fair die. What is the probability that you roll an even number **and** a number greater than 3?

3. You roll a fair die. What is the probability that you roll an odd **or** a number greater than 2?

4. You roll a fair die. What is the probability that you roll an odd number **and** a number greater than 4?

5. You toss a coin 3 times. What is the probability getting heads 3 times?

6. You toss a coin 3 times. What is the probability getting heads 2 times?

7. You have a deck of cards (52 cards). If you pull a card what is the probability that it will be an ace.

8. You have a deck of cards (52 cards). If you pull a 2 cards what is the probability that one of them be an ace.

9. You roll a pair of fair dice. What is the probability that the sum of the numbers is 7 or 11?

10. You roll a pair of fair dice. What is the probability that both dice either turn up the same number or that the sum of the numbers is less than 5?

11. What is the probability that a number selected at random from the first 50 positive integers is exactly divisible by 3 or 4?

Lesson 7.2 Conditional Probability

Knowledge of **conditional probability** can inform us about how one event or factor affects another. Say-No-To-Smoking campaigns are vigilant in educating the public about the adverse health effects of smoking cigarettes. This motivation to educate the public has its beginnings in data analysis. Below is a table that represents a sampling of 500 people. Distinctions are made on whether or not a person is a smoker and whether or not they have ever developed lung cancer. Each number in the table represents the number of people that satisfy the conditions named in its row and column.

	Has been a smoker for 10+ years	Has not been a smoker
Has not developed lung cancer	202	270
Has developed lung cancer	23	5

1. How does the table indicate that there is a connection between smoking and lung cancer?

2. Using the 500 data points from the table, you can make reasonable estimates about the population at large by using probability. 500 data values is considered, statistically, to be large enough to draw conclusions about a much larger population. In order to investigate the table using probability, use the following outcomes:

S – The event that a person is a smoker
L – The event that a person develops lung cancer

Find each of these probabilities (write as percentages):

a. $P(S) =$ The probability that a person is a smoker: $P(S) = \frac{225}{500} = 0.45 = 45\%$

b. $P(\overline{S}) =$ The probability that a person is not a smoker: $P(\overline{S}) =$

c. $P(L) =$ The probability that a person develops lung cancer: $P(L) =$

d. $P(\overline{L}) =$ The probability that a person does not develop lung cancer: $P(\overline{L}) =$

e. $P(L \cap S) =$ ____: $P(L \cap S) = \frac{23}{500} = 0.046 = 4.6\%$

f. $P(\overline{S} \cap \overline{L}) =$ ____:

g. $P(\overline{S} \cap L) =$____:

h. $P(S \cap \overline{L}) =$____:

i. $P(S \cup L) =$____:

j. $P(\overline{S} \cup \overline{L}) =$____:

3. In order to use probability to reinforce the connection between smoking and lung cancer, you will use calculations of *conditional probability*.

a) By considering only those people who have been smokers, what is the probability of developing lung cancer?

b) Compare the value in 3a to the one for $P(L)$ in 2c. What does this indicate?

c) You should be able to confirm that a non-smoker is less likely to develop lung cancer. By considering only non-smokers, what is the probability of developing lung cancer?

4. When calculating conditional probability, it is common to use the term "given." In question 3a, you have calculated the probability of a person developing lung cancer given that they are a smoker. The condition (or, "given") is denoted with a single, vertical bar separating the probability needed from the condition. The probability of a person developing lung cancer given that they are a smoker is written $P(L / S)$.

a) Rewrite the question from 3c using the word "given."

b) Write the question from 3c using set notation:

5. Find the probability that a person was a smoker given that they have developed lung cancer and represent it with proper notation. (Hint: there are 23 smokers who had lung cancer out of a total of 225 smokers).

A Formula for Conditional Probability

The formulaic definition of conditional probability can be seen by looking at the different probabilities you calculated in the previous pages. The formal definition for the probability of event *A* given event *B* is the chance of both events occurring together with respect to the chance that *B* occurs.

Probability of *A* given *B*

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

Example: In (1a) you found that $P(S) = \frac{225}{500}$ and in (1e) you found $P(L \cap S) = \frac{23}{500}$.

Using the formula of conditional probability is another way to determine that

$$P(L/S) = \frac{P(L \cap S}{P(S)} = \frac{23/500}{225/500} = \frac{23/\cancel{500}}{225/\cancel{500}} = \frac{23}{225} = 10.2\%$$

which should yield the same answer as problem (5) of this section.

6. Find the probability that a person has not developed lung cancer given that they not been smoking and represent it with proper notation. Use the formula above.

7. Find the probability that a person has lung cancer given that they have **not** been smoking.

8. Based upon finding the conditional probabilities make an argument that supports the connection between smoking and lung cancer.

More problems on conditional probability:

9. For two events S and Q it is known that $P(Q)=.45$ and $P(S\cap Q)=.32$. Find $P(S/Q)$.

10. For two events X and Y it is known that $P(X)=\frac{1}{5}$ and $P(X\cap Y)=\frac{2}{15}$. Find $P(Y/X)$.

11. The probability of developing diabetes is 0.18 and the probability of being obese is 0.37. 8% of all people are obese and diabetic. What is the probability of being diabetic given that you are obese? Remember to change percent to a decimal.

12. The probability of developing cancer is 0.34 and the probability of someone being a smoker is 0.39. 36% of all people are smokers and developed cancer. What is the probability of getting cancer given that you are smoker? Remember to change percent to a decimal.

13. For two events B and C it is known that $P(C/B)=.61$ and $P(C\cap B)=.48$. Find $P(B)$.

14. For two events V and W it is known that $P(W)=\frac{2}{9}$ and $P(V/W)=\frac{5}{11}$. Find $P(V\cap W)$.

15. For two events G and H it is known that $P(H/G)=\frac{5}{14}$ and $P(H\cap G)=\frac{2}{3}$. Explain why you cannot determine the value of $P(H)$.

16. Referring back to the "smoker" problem, what is $P(S/\overline{L})$?

17. Referring back to the "smoker" problem, what is $P(\overline{S}/L)$?

Movie executives collect lots of different data on the movies they show in order to determine who is going to see the different types of movies they produce. This will help them make decisions on a variety of factors from where to advertise a movie to what actors to cast. Below is a two-way frequency table that compares the preference of *Harry Potter and the Deathly Hallows* to *Captain America: The First Avenger* based upon the age of the moviegoer. 200 people were polled for the survey.

	Prefers *Harry Potter*	Prefers *Captain America*
Under the age of 30	73	52
Age 30 or above	20	55

Define each event in the table using the following variables:
H – A person who prefers *Harry Potter and the Deathly Hallows*
C – A person who prefers *Captain America: The First Avenger*
Y – A person under the age of 30
E – A person whose age is 30 or above

1. By looking at the table, but without making any calculations, would you say that there is a relationship between age and movie preference? Why or why not?

2. Find the following probabilities. In terms of movie preference, explain what each probability—or probabilities together in the case of b, c, and d—would mean to a movie executive

a) $P(Y) =$

$P(E) =$

$P(C) =$

$P(H) =$

b) $P(C / Y) =$

c) $P(H / Y) =$

d) $P(E/C) =$

e) $P(Y/C) =$

Summarize what a movie executive can conclude about age preference for these two movies through knowing the probabilities that you have found.

The retail and service industries are another aspect of modern society where probability's relevance can be seen. By studying data on their own service and their clientele, businesses can make informed decisions about how best to move forward in changing economies. Below is a table of data collected over a weekend at a local ice cream shop, Frankie's Frozen Favorites. The table compares a customer's flavor choice to their cone choice.

Frankie's Frozen Favorites	Chocolate	Butter Pecan	Fudge Ripple	Cotton Candy
Sugar Cone	36	19	34	51
Waffle Cone	35	56	35	24

1. By looking at the table, but without making any calculations, would you say that there is a relationship between flavor and cone choice? Why or why not?

2. Find the following probabilities (write as percentages):

a) $P(W)$

b) $P(S)$

c) $P(C)$

d) $P(BP)$

e) $P(FR)$ **f)** $P(CC)$

3. In order to better investigate the correlation between flavor and cone choice, calculate the conditional probabilities for each cone given each flavor choice. A table has been provided to help organize your calculations.

Frankie's Frozen Favorites	Chocolate	Butter Pecan	Fudge Ripple	Cotton Candy
Sugar Cone	$P(S\|C)$	$P(S\|BP)$	$P(S\|FR)$	$P(S\|CC)$
Waffle Cone	$P(W\|C)$	$P(W\|BP)$	$P(W\|FR)$	$P(W\|CC)$

4. Compare and contrast the probabilities you found in question 2 with the conditional probabilities you found in question 3.
a. Which flavors actually affect cone choice?

b. Which do not?

c. How did you make this determination?

5. The relationship that you have observed between chocolate and cone choice (and fudge ripple and cone choice) is called **independence**. Multiple events in probability are said to be independent if the outcome of any one event does not affect the outcome of the others. The fact that $P(S/C)$ and $P(W/C)$ are approximately equal to each other indicates that the choice of cone is in no way affected by the choice of chocolate ice cream. The same is true for fudge ripple. When probabilities change depending on the situation, such as knowing sugar cones are more likely with cotton candy ice cream, the events have a **dependent** relationship. Answer the questions below to ensure you understand this new terminology:

a. Explain whether or not flipping a coin twice would be considered a set of two independent events.

b. A game is played where marbles are pulled from a bag, 8 of which are red and 2 are white. You score by pulling marbles from the bag, one at a time, until you pull a white marble. Are the events in this game independent or dependent? Why?

c. When you roll two dice, is the outcome of one dice dependent on the other? Why or why not?

d. You pull a card from deck of cards without replacing it. Then you pull a second card. Are these events dependent on the other? Why or why not?

e. You pull a card from deck of cards, look at it and put it back into the deck of cards. Then you pull a second card. Are these events dependent on the other? Why or why not?

6. Consider the statement, "the probability that a sugar cone is chosen given that chocolate ice cream is chosen." The desired probability relates to a sugar cone, but this choice is *independent* of the choice of chocolate. That is to say the statement "the probability that a sugar cone is chosen" is no different when "given that chocolate ice cream is chosen" is removed. Thus, we can say $P(S/C) = P(S)$. Which other parts of the table from question 3 can be written in a simpler way?

7. For independent events, the conditional probability formula,

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ becomes } P(A) = \frac{P(A \cap B)}{P(B)}.$$

Solve the equation on the right for $P(A \cap B)$ and place it in the box below.

PROBABILITY OF INDEPENDENT EVENTS A AND B
$P(A \cap B) =$

Notice that for *independent events*, the probability of two events occurring together is simply the product of each event's individual probability. For example when you flip a coin the chance to get tails is ½. The chance to get two tails in a row if you flip the coin twice is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Since both events are independent you just multiply their probabilities.

8. To conclude, let's go back and revisit Frankie's Frozen Favorites. In this problem, you discovered pairs of events that were independent of one another by comparing their conditional probabilities to the probabilities of single events. Because conditions do not need to be considered when calculating probabilities of two independent events, we arrived at the formula above.

a. Use the formula above, $P(A \cap B) = P(A) \times P(B)$, to verify that Fudge Ripple and Waffle Cone are independent events.

b. Explain in full why this formula will *not* accurately calculate $P(CC \cap W)$.

Lesson 7.3 Confirming Dependence

By developing a full picture of conditional probability in the previous task, you were able to conclude that events that occur without regard to conditions, independent events, are defined by the equation $P(A \cap B) = P(A) \cdot P(B)$. This equation is known as *necessary and sufficient*. It works exactly like a biconditional statement: two events *A* and *B* are independent if and only if the equation $P(A \cap B) = P(A) \cdot P(B)$ is true.

1. Based upon the definition of independence, determine if each set of events below are independent.

a. $P(A) = 0.45, P(B) = 0.30, P(A \cap B) = 0.75$

b. $P(A) = 0.12, P(B) = 0.56, P(A \cap B) = 0.0672$

c. $P(A) = \frac{4}{5}, P(B) = \frac{3}{8}, P(A \cap B) = \frac{7}{40}$

d. $P(A) = \frac{7}{9}, P(B) = \frac{3}{4}, P(A \cap B) = \frac{7}{12}$

2. Determine the missing values so that the events A and B will be independent.

a. $P(A) = 0.55, P(B) =$ __________, $P(A \cap B) = 0.1375$

b. $P(A) =$ __________, $P(B) = \frac{3}{10}, P(A \cap B) = \frac{1}{7}$

<u>Below is a recap of what we learned so far.</u>

Multiplication Rule (A∩B)

This region is referred to as 'A intersection B' and in probability; this region refers to the event that both **A and B** happen. When we use the word and we are referring to multiplication, thus A and B can be thought of as AxB or (using dot notation which is more popular in probability) A•B

If A and B are dependent events, the probability of this event happening can be calculated as shown below:

$$P(A \cap B) = P(A \cup B) - (P(A \text{ only}) + P(B \text{ only}))$$

Example: If you pick a card out of a deck of cards without replacing it and then pick another card. The outcome of the second card is affected by the outcome of the first card.

If A and B are independent events, the probability of this event happening can be calculated as shown below:

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

Example: If you roll two dice the outcome of one die is independent on the other die. The probability that you roll a 3 with one of the dice given that the other die is a 3 is still 1/6.

Conditional probability for two **independent events** can be redefined using the relationship above to become:

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

If the occurrence of event A in no way influences the occurrence of event B then the probability that event B occurs given that event A has occurred is the same as the probability of event B.

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

$$P(B|A) = P(B)$$

The above is consistent with the definition of independent events,

In other words **if P(B/A) is not equal to P(B) there is some dependency** between the probability of event B and the probability of event A.

Independence and Inference

With knowledge of probability and statistics, statisticians are able to make *statistical inferences* (conclusions) about large sets of data. Based upon what you have learned in this unit, you have the knowledge necessary to make basic inferences.

Much of the data collected every 10 years for the Census is available to the public. This data includes a variety of information about the American population at large such as age, income, family background, education history and place of birth. Below you will find three different samples of the Census that looks at comparing different aspects of American life. Your job will be to use your knowledge of conditional probability and independence to make conclusions about the American populace.

1. Gender vs. Income – Has the gender gap closed in the world today? Are men and women able to earn the same amount of money? The table below organizes income levels (per year) and gender.

	10- Under $10,00	10-40 Between $10,000 and $40,000	40-100 Between $40,000 and $100,000	100+ Over $100,000	
M Male	15	64	37	61	
F Female	31	73	14	58	

A. First of all find all of the probabilities below:

$P(M) =$

$P(F) =$

$P(10-) =$

$P(10-40) =$

$P(40-100) =$

$P(100+) =$

$P(F \cap 10-) =$

$P(F \cap 10-40) =$

$P(F \cap 40-100) =$

$P(F \cap 100+) =$

$P(M \cap 10-) =$

$P(M \cap 10-40) =$

$P(M \cap 40-100) =$

$P(M \cap 100+) =$

Now try to make a determination about whether or not income level is affected by gender by comparing corresponding income levels for women and men. Investigate whether your conclusion is true for all income levels. Show all the calculations you use and write a conclusion using those calculations. (Hint: Use $P(B|A) = \frac{P(A) \cdot P(B)}{P(A)}$ to show dependence between gender and income)

2. Bills vs. Education – When you grow up, do you think the amount of schooling you have had will be at all related to the amount of money you have to pay out in bills each month? Below is a table that compares two variables: the highest level of education completed (below a high school diploma, a high school diploma, or a college degree) and the amount paid for a mortgage or rent each month.

	Pays under $500	Pays between $500 and $1000	Pays over $1000	
Below high school	57	70	30	
High school diploma	35	47	11	
College degree	24	62	40	

By determining the probabilities of each education level and the probabilities of housing costs, you should be able to decide whether or not these two variables are independent. Show all the calculations you use, and write a conclusion about the *interdependence* of these two variables.

Gender vs. Commute – What else might gender affect? Is your commute to work related to whether or not you are male or female? The data below allows you to investigate these questions by presenting gender data against the minutes needed to commute to work each day.

	Under 30 minutes	Between 30 minutes and an hour	Over an hour	
Male	65	24	15	
Female	64	22	7	

By finding various probabilities from the table above, decide whether or not a person's gender is related to their commute time to work. Write your conclusion below and include any relevant calculations.