Georgia
Standards of Excellence
Curriculum Frameworks

Mathematics

Accelerated GSE Analytic Geometry B/Advanced Algebra

Unit 3: Applications of Probability

Georgia Department of Education

Richard Woods, Georgia's School Superintendent
"Educating Georgia's Future"

Unit 3
Applications of Probability

Table of Contents

OVERVIEW 3
STANDARDS ADDRESSED IN THIS UNIT 3
KEY STANDARDS 3
RELATED STANDARDS 4
ENDURING UNDERSTANDINGS 5
ESSENTIAL QUESTIONS 5
CONCEPTS/SKILLS TO MAINTAIN 5
SELECTED TERMS AND SYMBOLS 6
STRATEGIES FOR TEACHING AND LEARNING 7
EVIDENCE OF LEARNING 7
FORMATIVE ASSESSMENT LESSON (FAL) 8
SPOTLIGHT TASKS 8
TASKS 9
How Odd? 10
The Conditions are Right 35
Modeling Conditional Probabilities 1: Lucky Dip (FAL) 66
The Land of Independence 68
Modeling Conditional Probabilities 2 (FAL) 81
Medical Testing (FAL) 83
False Positives (Career and Technical Education Task) 86
Compound Confusion (FAL) 87
A Case of Possible Discrimination (Spotlight Task) 88
Culminating Task: Are You Positive? 103

OVERVIEW

In this unit, students will:

- take their previously acquired knowledge of probability for simple and compound events and expand that to include conditional probabilities (events that depend upon and interact with other events) and independence
- be exposed to elementary set theory and notation (sets, subsets, intersection and unions)
- use their knowledge of conditional probability and independence to make determinations on whether or not certain variables are independent

Although the units in this instructional framework emphasize key standards and big ideas at specific times of the year, routine topics such as estimation, mental computation, and basic computation facts should be addressed on an ongoing basis. Ideas related to the eight practice standards should be addressed constantly as well. This unit provides much needed content information and excellent learning activities. However, the intent of the framework is not to provide a comprehensive resource for the implementation of all standards in the unit. A variety of resources should be utilized to supplement this unit. The tasks in this unit framework illustrate the types of learning activities that should be utilized from a variety of sources. To assure that this unit is taught with the appropriate emphasis, depth, and rigor, it is important that the "Strategies for Teaching and Learning" in the Comprehensive Course Overview and the tasks listed under "Evidence of Learning" be reviewed early in the planning process.

STANDARDS ADDRESSED IN THIS UNIT

KEY STANDARDS

Understand independence and conditional probability and use them to interpret data

MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

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

MGSE9-12.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 context.

MGSE9-12.S.CP.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answers in context.

RELATED STANDARDS

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

MGSE7.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.

MGSE7.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.

MGSE7.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.

STANDARDS FOR MATHEMATICAL PRACTICE

Refer to the Comprehensive Course Overview for more detailed information about the Standards for Mathematical Practice.

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.

4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

SMP = Standards for Mathematical Practice

ENDURING UNDERSTANDINGS

- Use set notation as a way to algebraically represent complex networks of events or real world objects.
- Represent everyday occurrences mathematically through the use of unions, intersections, complements and their sets and subsets.
- Use Venn Diagrams to represent the interactions between different sets, events or probabilities.
- Find conditional probabilities by using a formula or a two-way frequency table.
- Understand independence as conditional probabilities where the conditions are irrelevant.
- Analyze games of chance, business decisions, public health issues and a variety of other parts of everyday life can be with probability.
- Model situations involving conditional probability with two-way frequency tables and/or Venn Diagrams.
- Confirm independence of variables by comparing the product of their probabilities with the probability of their intersection.

ESSENTIAL QUESTIONS

- How can I represent real world objects algebraically?
- How can I communicate mathematically using set notation?
- In what ways can a Venn diagram represent complex situations?
- How can I use a Venn diagram to organize various sets of data?
- How can two-way frequency tables be useful?
- How are everyday decisions affected by an understanding of conditional probability?
- What options are available to me when I need to calculate conditional probabilities?
- What connections does conditional probability have to independence?
- What makes two random variables independent?
- How do I determine whether or not variables are independent?

CONCEPTS/SKILLS TO MAINTAIN

It is expected that students will have prior knowledge/experience related to the concepts and skills identified below. It may be necessary to pre-assess in order to determine if time needs to be spent on conceptual activities that help students develop a deeper understanding of these ideas.

- Understand the basic nature of probability
- Determine probabilities of simple and compound events
- Organize and model simple situations involving probability
- Read and understand frequency tables

SELECTED TERMS AND SYMBOLS

The following terms and symbols are often misunderstood. These concepts are not an inclusive list and should not be taught in isolation. However, due to evidence of frequent difficulty and misunderstanding associated with these concepts, instructors should pay particular attention to them and how their students are able to explain and apply them.

The definitions below are for teacher reference only and are not to be memorized by the students. Students should explore these concepts using models and real life examples. Students should understand the concepts involved and be able to recognize and/or demonstrate them with words, models, pictures, or numbers.

http://intermath.coe.uga.edu/dictnary/homepg.asp
Definitions and activities for these and other terms can be found on the Intermath website.

- **Addition Rule:** $\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B}) - \mathrm{P}(\mathrm{A} \cap \mathrm{B})$

- **Complement:** Given a set A, the complement of A, denoted $\overline{A}$ 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 $P(A \mid B)$. The formula for a conditional probability is
$$\mathrm{P}(\mathrm{A} \mid \mathrm{B}) = \frac{\mathrm{P}(\mathrm{A} \cap \mathrm{B})}{\mathrm{P}(\mathrm{B})}$$

- **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 $\cap$.

- **Multiplication Rule for Independent Events**: $P(A \cap B) = P(A)P(B)$

- **Mutually Exclusive Events:** Two events that cannot occur simultaneously, meaning that the probability of the intersection of the two events is zero; also known as disjoint events

- **Outcome:** A possible result of an experiment

- **Overlapping Events:** Events that can occur simultaneously – they have an intersection.

- **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 $\cup$.

- **Venn Diagram:** A picture that illustrates the relationship between two or more sets.

STRATEGIES FOR TEACHING AND LEARNING

In this unit, students will be visiting the topic of probability for the first time since 7th grade. By unit's end they will have taken their previously acquired knowledge of probability for simple and compound events and expanded that to include conditional probabilities (events that depend upon and interact with other events) and independence. In order to communicate their learning on these topics, students will also be exposed to elementary set theory and notation (sets, subsets, intersection and unions).

EVIDENCE OF LEARNING

By the conclusion of this unit, students should be able to demonstrate the following competencies:

- use set notation to represent a set of events mathematically
- use the addition rule for two events and/or their probabilities
- read and interpret a two-way frequency table

- determine conditional probabilities given sufficient information
- use the formula for conditional probability
- use the formula for the probability independent events
- confirm whether or not two events are independent using probability

FORMATIVE ASSESSMENT LESSONS (FAL)

Formative Assessment Lessons are intended to support teachers in formative assessment. They reveal and develop students' understanding of key mathematical ideas and applications. These lessons enable teachers and students to monitor in more detail their progress towards the targets of the standards. They assess students' understanding of important concepts and problem solving performance, and help teachers and their students to work effectively together to move each student's mathematical reasoning forward.

More information on Formative Assessment Lessons may be found in the Comprehensive Course Overview.

SPOTLIGHT TASKS

A Spotlight Task has been added to each GSE mathematics unit in the Georgia resources for middle and high school. The Spotlight Tasks serve as exemplars for the use of the Standards for Mathematical Practice, appropriate unit-level Georgia Standards of Excellence, and research-based pedagogical strategies for instruction and engagement. Each task includes teacher commentary and support for classroom implementation. Some of the Spotlight Tasks are revisions of existing Georgia tasks and some are newly created. Additionally, some of the Spotlight Tasks are 3-Act Tasks based on 3-Act Problems from Dan Meyer and Problem-Based Learning from Robert Kaplinsky.

TASKS

The following tasks represent the level of depth, rigor, and complexity expected of all Analytic Geometry students. These tasks, or tasks of similar depth and rigor, should be used to demonstrate evidence of learning. It is important that all elements of a task be addressed throughout the learning process so that students understand what is expected of them. While some tasks are identified as a performance task, they may also be used for teaching and learning (learning/scaffolding task).

Task Name	Task Type *Grouping Strategy*	Content Addressed	SMPs Addressed
How Odd	Learning Task *Partner/Small Group Task*	Venn diagrams, set notation and the addition rule	1, 2, 4 – 7
The Conditions are Right	Learning Task *Partner/Small Group Task*	Conditional probability and frequency tables	1 – 8
Modeling Conditional Probabilities 1: Lucky Dip (FAL)	Formative Assessment Lesson	Understand conditional probability and represent subsets in multiple ways.	1-4
The Land of Independence	Performance Task *Individual/Partner/ Small Group Task*	Independence	1 – 8
Modeling Conditional Probabilities 2 (FAL)	Formative Assessment Lesson	Understand when conditional probabilities are equal,	1, 2, 4
Medical Testing (FAL)	Formative Assessment Lesson	Implement a strategy to solve conditional probabilities.	1 – 4
False Positives	Achieve CCSS- CTE Classroom Tasks	Exploring conditional probability using a variety of methods.	1 – 8
Compound Confusion (FAL)	Formative Assessment Lesson	Understand the meaning of independent, dependent, mutually exclusive, conditional, and overlapping probability.	1, 3
A Case of Possible Discrimination (Spotlight Task)	Performance Task *Individual/Partner Task*	Two-way table with joint, marginal, conditional probabilities and independence with simulation	1, 2, 3, 4, 6, 7
Culminating Task: Are You Positive?	Performance Task *Individual/Partner Task*	Conditional probability and frequency tables, independence, addition rule	1 – 8

How Odd?

Standards Addressed in this Task
MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answers in context.

Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

Common Student Misconceptions
1. Students may believe that multiplying across branches of a tree diagram has nothing to do with conditional probability.
2. Students may believe that independence of events and mutually exclusive events are the same thing.
3. Students may believe that the probability of A or B is always the sum of the two events individually.
4. Students may believe that the probability of A and B is the product of the two events individually, not realizing that one of the probabilities may be conditional.

In middle school mathematics, you took a first look at probability models. You most likely solved problems that involved selecting cards, spinning a spinner, or rolling die to find the likelihood that an event occurs. In this task you will build upon what you already know. You will start with an introduction to **set theory** (a way to algebraically represent different mathematical objects). This will allow you later on in this unit to better explore two branches of probability theory: conditional probability and independence. Through these topics you will be able to uncover how data analysis and probability can help inform us about many aspects of everyday life.

Comment(s):

The first job of this task is to allow students to become reacquainted with probability, a topic they have not studied since Grade 7. Their previous experience is that of compound events such as rolling two die or taking two marbles from a bag. Second, this task intends to introduce students to elementary set theory notation. This includes unions, intersections and complements. Finally, the task aims to allow students to represent probabilities with this newly learned notation so that they may be successful at reading, expressing and evaluating probabilities throughout this unit.

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.

Comment(s):

Students can work in groups of various sizes though it is suggested that they work in pairs. The social learning aspect of this task will allow for useful discussion among students. Having larger groups with only two die per group will allow some students to "sit on the sidelines" and not have a full learning experience. With the intent of the task to re-introduce and old topic it is vital that all students are engaged.

1. Roll your pair of dice 30 times, each time recording a success if one (or both) of the dice shows an odd number and a failure if the dice do not show an odd number.

Number of Successes	Number of Failures

Solution(s):

Answers on these trials will obviously vary. Students should have ratios that mimic the theoretical probability, successes being $3/4^{th}$ of all outcomes, failures being $1/4^{th}$.

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

Comment(s):

Although theoretically, students should arrive at 3/4 or 0.75, based on random chance, it is reasonable for a student to not have this answer A more advanced student may think of this solution as all the possible outcomes subtract out the number of outcomes with both die landing on an even number producing 1 – 3/6*3/6, but it is not expected.

3. You have just calculated an **experimental probability**. 30 trials are 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. A lattice diagram is useful in finding the theoretical probabilities for two dice thrown together. An incomplete lattice diagram is shown to the right. 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. Why does it have 36 spaces to be filled?

Dice Lattice

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

Comment(s):

Students may overlook the fact that rolling (1,2) and (2,1) are distinct outcomes. If students have trouble producing/understanding 36, a tree diagram is a strong way to convince them that 36 such outcomes exist. Another is to physically manipulate the dice. Holding the first at any number and rotate the second through all six digits. Then change the first die and again rotate the second through all six, and so on.

Solution(s):

Since each die has 6 sides, and an "and" operation is being used, there are 6 x 6 = 36 possible outcomes for the two dice.

b. Complete the lattice diagram 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. Give the subset that contains all elements that sum to 9.

Comment(s):

Students have never seen this notation before; they may be creative in their interpretation of the braces. Again, students who struggle with (3, 6) and (6, 3) may struggle here.

Solution(s):

{(3, 6), (6, 3), (4, 5), (5, 4)}

d. What is the probability that the sum of two die rolled will be 9?

Comment(s):

Students should remember that probabilities are the number of desired outcomes divided by the size of the sample space. These terms should be reviewed.

Solution(s):

4/36 = 1/9

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

Comment(s):

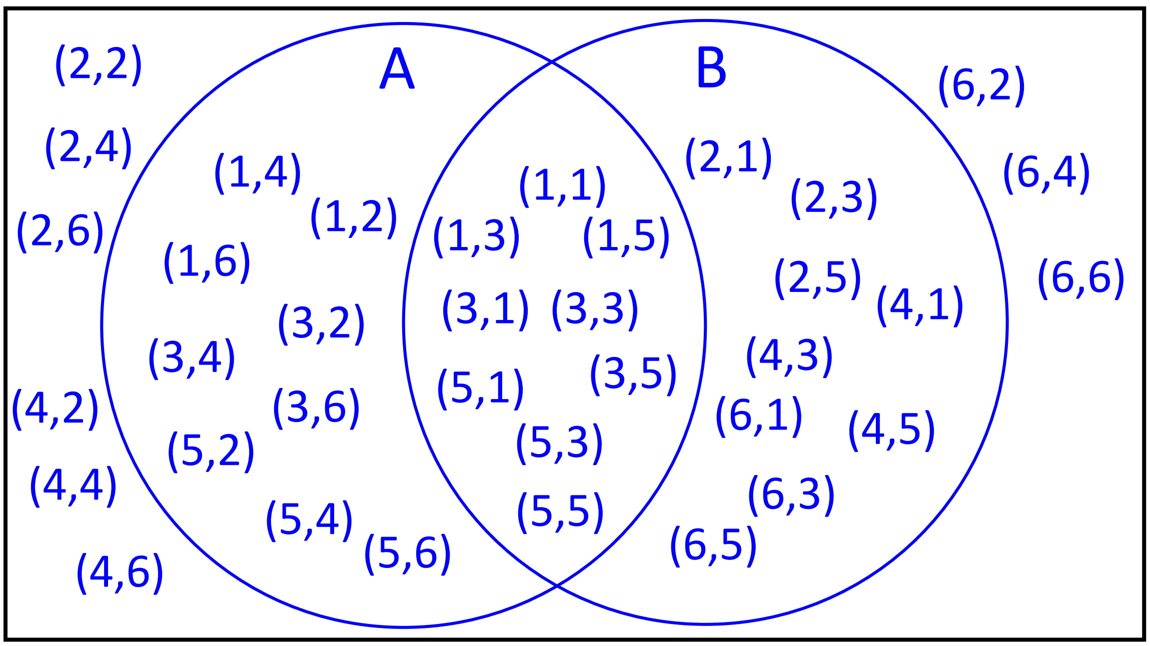
Students may simply get this incorrect by miscounting to 27.

Solution(s):

27/36 = 3/4

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 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).

Comment(s):

Despite the hint, students may answer 9.

Solution(s):
18

c. How many outcomes appear in circle B?

Comment(s):

Again, students may answer 9. Students may hesitate to put 18, as 18 + 18 = 36 and that is already the entire sample space without counting the (even, even) rolls. The teacher may need to clear up misconceptions about the overlap.

Solution(s):

18

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 $A \cap B$?

Comment(s):

At this point, students may wonder about a faster way to get the answers of 18 above and this answer of 9. Although counting techniques are not explicitly part of these standards, it is appropriate to show students the argument that circle A is the event (odd, any number) so there are 3 options for one dice and 6 options for the other dice which multiply to get 18. The same is true for circle B. The intersection is (odd, odd) so 3 options twice for 9 total outcomes.

Solution(s): 9

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?

Comment(s):
Students' answers may vary here, and wrong answers are a great source of information for misunderstandings. Students may take short cuts and add any combination of 18's and 9's. Students should be focusing on just counting the actual number of ordered pairs that appear somewhere in the circle, and by counting they should arrive at 27.

Solution(s): 27

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

Solution(s):

b. Circle A	c. Circle B	d. $A \cap B$	e. $A \cup B$
18	*18*	*9*	*27*

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

Comment(s):

Students may not pick up on the arithmetic relationship here. For instance, they may notice that all numbers are multiples of 9. When confused, the teacher may ask "How could you calculate e using the values in b, c, and d?" They may also be asked to look for a relationship such that a "quick" way to arrive at 27 may be developed.

Solution(s):

b + c – d = e.

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

Solution(s):

$A \cup B = A + B - A \cap B$.

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

Comment(s):

Students may get the set correct but have trouble wording why this set has not been used. Formal notation is needed for the answer, but the explanation of the events can be in layman's terms.

Solution(s):

$\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$ ***Because they contain only even numbers.***

In a Venn diagram the set of outcomes that are *not* included in some set is called the complement of that set. The notation used for the complement of set A is $\overline{A}$, read "A bar", or ~A, read "not A", or A'. For example, in the Venn diagram you completed above, the outcomes that are outside of $A \cup B$ are denoted $\overline{A \cup B}$.

Comment(s):
A' is another notation for the complement of Set A.

j. Which outcomes appear in $\overline{A} - B$?

Comment(s):

This set is the (even, even) set. Since the notation is new, students may be confused by this question. Displaying the Venn Diagram to the class and using a technique of shading regions is a way to show the desired outcomes.

Solution(s):

$\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$.

k. Which outcomes appear in $\overline{B} - \left(\overline{A \cup B}\right)$?

Comment(s):

This set is every outcome in A that is not in the intersection or (odd, even). Again, shading regions will help students who struggle. Students who have a good understanding of the possibilities for the dice but struggle with the notation may understand this argument: B is (any number, odd). B complement is (any number, even). The complement of A union B is (even, even). So taking out (even, even) from (any number, even) leaves only (odd, even).

Solution(s):

$\{(1,2),(1,4),(1,6),(3,2),(3,4),(3,6),(5,2),(5,4),(5,6)\}$.

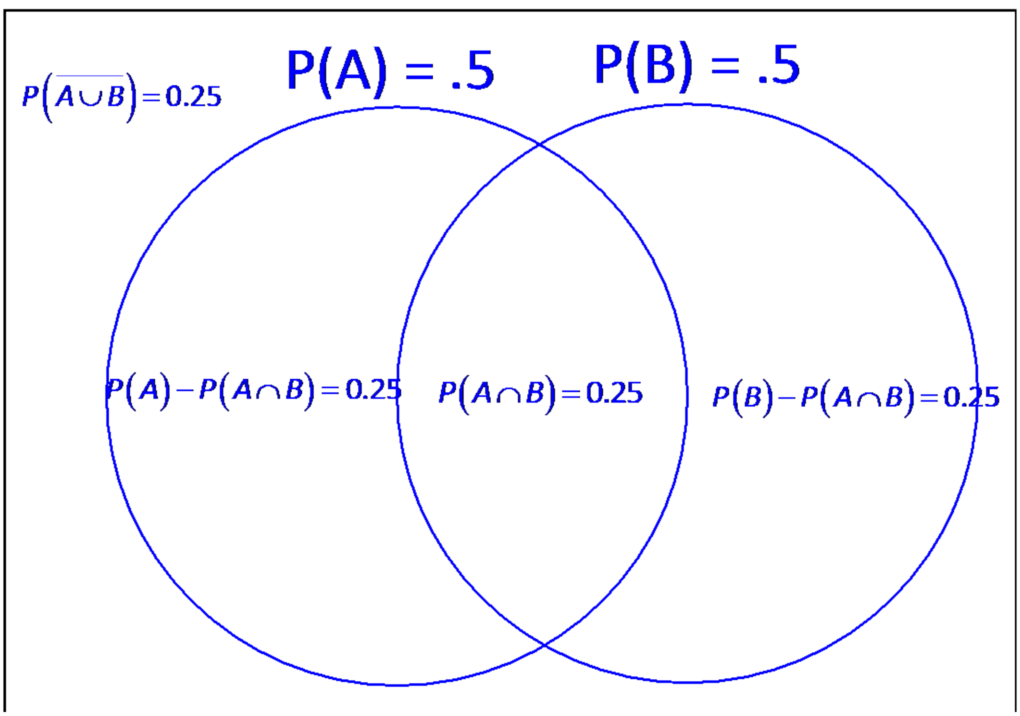
5. The investigation of the Venn diagram in question 4 should reveal a new way to see that the probability of rolling at least one odd number on two dice is $\frac{27}{36} = \frac{3}{4}$. How does the Venn diagram show this probability?
Comment(s):

Students should be able to discuss that A union B is the desired outcome and the entire diagram is the sample space. With 27 items in the union and 36 in the entire diagram, the probability is 27/36. Students who think this question is trivial may have trouble writing more than something such as "it's obvious."

Solution(s):

One possible solution: All of the outcomes in which at least one die is odd are in either A or B. There are 27 outcomes in A union B and 36 in the sample space, therefore the probability is 27/36.

6. 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.



a. Fill in the remaining probabilities in the Venn diagram.

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

Comment(s):

Students may find this question redundant, but the important connection to make is that the probabilities are found in the same way that they counted the outcomes earlier in 4e.

Solution(s):

Now that we know the probabilities, we can calculate 0.5 + 0.5 – 0.25 rather than count outcomes in the diagram.

c. Use the probabilities in the Venn diagram to find $P\left(\overline{B}\right)$.

Comment(s):

Students should identify the complement notation and determine that a probability and its complement add to 1, hence the answer is .5. This should be a concept they have seen in Grade 7 but may have lost.

Solution(s):

0.5

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

Comment(s):

For students who have trouble understanding, present the idea of a binomial: success vs. failure. Since P(B) is a success, the only other option is failure. Success and failure together comprise all options, hence 1 (100%).

Solution(s):

The probability of an event and its complement (opposite) always add to 1.

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.

Comment(s):

This part of the task is intended to allow students to test what they have learned. They will be asked to first produce a relatively simple Venn diagram, then a more complicated one. Each time the diagram will be built with real data from the classroom.

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?

Comment(s):

Naturally, there is no wrong answer, and students should embrace a chance to voice their opinion.

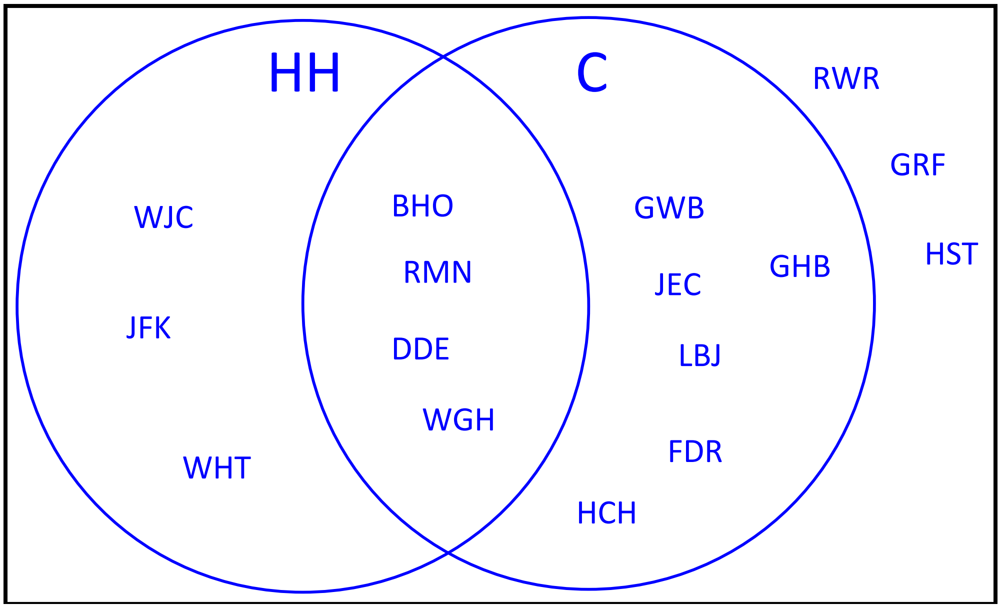
2. Each student should identify themselves by their 3 initials (first, middle, last). Any student who listens to both Country and Hip Hop may be listed in both categories. Record results of the class poll in the table.

Comment(s):

For the purpose of being able to provide answers in this part of the task, the following artificial data is being used.

Hip Hop (HH)	Country (C)	Neither (N)
BHO	***BHO***	***RWR***
WJC	***GWB***	***GRF***
RMN	***GHB***	***HST***
JFK	***JEC***	
DDE	***RMN***	
WGH	***LBJ***	
WHT	***DDE***	
	FDR	
	HCH	
	WGH	

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.)*



Comment(s):

This diagram is designed to transfer from the data table in a simple way. The highlighted initials appear in both Country and Hip Hop and therefore make up the intersection. Likewise, the "Neither" section is outside of the two circles. You may want to instruct students to circle or highlight the names that appear in both to avoid mistakes here.

Answers for c-f will vary based upon your class data. Again, the answers below are based on artificial data.

4. Find $P(\text{HH})$.

Comment(s):

Students may misinterpret exactly what is being asked here. The survey question was "What genre of music do you enjoy listening to?" so those names listed in both Hip Hop and Country should also be included. A likely incorrect answer may be $\frac{3}{16}$ ***if students fail to include names in the intersection.***

Solution(s):

$$\frac{7}{16}$$

5. Find $\mathbf{P}\left(\overline{\mathrm{C}}\right)$.

Solution(s):

$\frac{6}{16}=\frac{3}{8}$

6. Find $\mathrm{P}(\mathrm{HH}\cap \mathrm{C})$.

Comment(s):

Again, students should have first identified names that appear in both to correctly fill in the intersection.

Solution(s):

$\frac{4}{16}=\frac{1}{4}$

7. Find $\mathrm{P}(\mathrm{HH}\cup \mathrm{C})$.

Comment(s):

Hopefully by this point, some students will begin using the complement and understanding its usefulness when probabilities are large. In other words, they will identify that rather than counting all the outcomes in $HH\cup C$, ***they can simply calculate*** $1-{}^{3}\!/_{16}$.

Solution(s):

$\frac{13}{16}$

8. In part 1, you found the relationship between A, B, $\mathrm{A}\cup\mathrm{B}$, and $\mathrm{A}\cap\mathrm{B}$ to be $\mathrm{A}\cup\mathrm{B}=\mathrm{A}+\mathrm{B}-\mathrm{A}\cap\mathrm{B}$. In a similar way, write a formula for $\mathrm{P}(\mathrm{A}\cup\mathrm{B})$.

Comment(s):

While students will understand the main idea here, they may struggle with writing it with the proper notation. Drawing their attention to the similar line of reasoning that exists here will help produce the correct statement.

Solution(s):

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

9. Now find $P(HH \cup C)$ using the formula instead of the Venn diagram. Did you get the same answer as you did in f above?

Comment(s):

Students should arrive at the same answer.

Solution(s):

Yes. $P(HH) + P(C) - P(HH \cap C) = \frac{7}{16} + \frac{10}{16} - \frac{4}{16} = \frac{13}{16}$

10. In what situation might you be forced to use the formula instead of a Venn Diagram to calculate the union of two sets?

Comment(s):

This question is intentionally open-ended, and it is intended for students to think critically about the purpose of a Venn diagram. Responses may include the fact that larger sets are hard to enumerate in a Venn diagram (for instance, rolling three dice) or that using probabilities is a "faster" way to communicate the important information.

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. Again, you will collect data on students in your class, record the data in a *two-way frequency table*, and then create a Venn diagram to organize the results of the poll. Which social networking site do you prefer?

Comment(s):

Like part 2, this part is designed to allow students to build a Venn Diagram with some guidance.

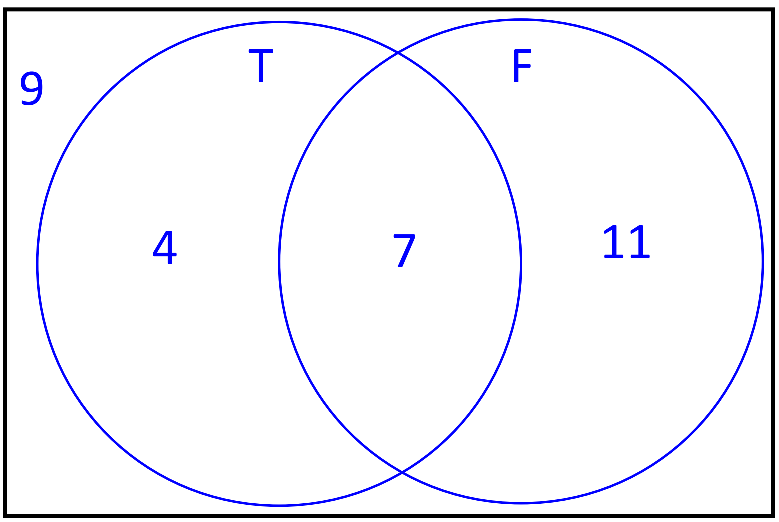
2. Record results from the class poll in the table.

	Twitter (T)	Facebook (FB)
Female (F)	7	11
Male (M)	4	9

Comment(s):

The artificial data above is given for the sake of providing solutions and is given in the event that your students do not have access to social networking sites and therefore would not have a preference.

3. Draw a Venn diagram to organize your outcomes. *(Hint: Notice that male and female will not overlap and neither will Twitter and Facebook).*



Comment(s):

The overlap between two circles should be Male or Female overlapping with Facebook or Twitter (for instance, Male and Twitter overlapping or Female and Facebook). It is not expected that every student will produce the same diagram, but every diagram should lead to correct probabilities below.

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

Comment(s):

Answers will vary depending on classroom data.

Solution(s):

$^{20}\!/_{31}$

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

Comment(s):

Like some questions from the end of part 1, notation will be the large hang-up here. Students who have trouble getting this answer should be first asked to describe this set of outcomes. More specifically, "Which students are in $(T \cup M)$? Which students are not? How can you use the complement to express those students that are not in $(T \cup M)$?"

Solution(s):

$P(\overline{FB \cap F})$

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

Comment(s):
Answers will vary depending on your classroom data.

Solution(s):
$\frac{7}{31}$

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

Comment(s):

Answers will vary depending on your classroom data.

Solution(s):

$\frac{4}{31} + \frac{11}{31} = \frac{15}{31}$

How Odd?

Standards Addressed in this Task
MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answers in context.

Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

In middle school mathematics, you took a first look at probability models. You most likely solved problems that involved selecting cards, spinning a spinner, or rolling die to find the likelihood that an event occurs. In this task you will build upon what you already know. You will start with an introduction to **set theory** (a way to algebraically represent different mathematical objects). This will allow you later on in this unit to better explore two branches of probability theory: conditional probability and independence. Through these topics you will be able to uncover how data analysis and probability can help inform us about many aspects of everyday life.

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 are 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. A lattice diagram is useful in finding the theoretical probabilities for two dice thrown together. An incomplete lattice diagram is shown to the right. 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. Why does it have 36 spaces to be filled?

Dice Lattice

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

b. Complete the lattice diagram 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.

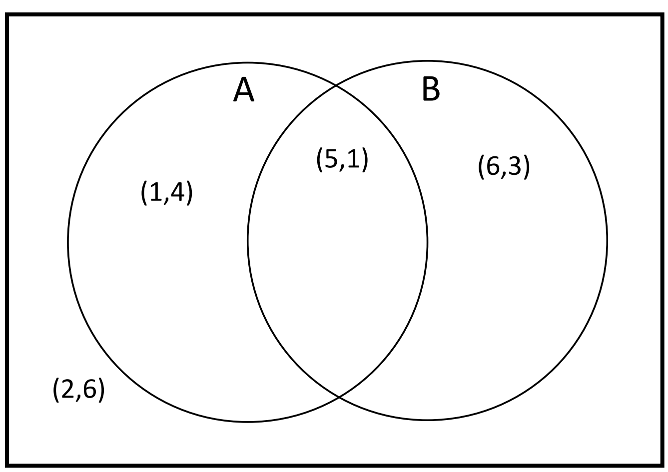
c. It is often necessary to list the sample space and/or the outcomes of a set using **set notation.** For the dice lattice above, the set of all outcomes where the first roll was a 1 can be listed as: {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)}. This set of outcomes is a **subset** of the set because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

d. 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.

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 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, $A \cup B$ and $A \cap B$ 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?

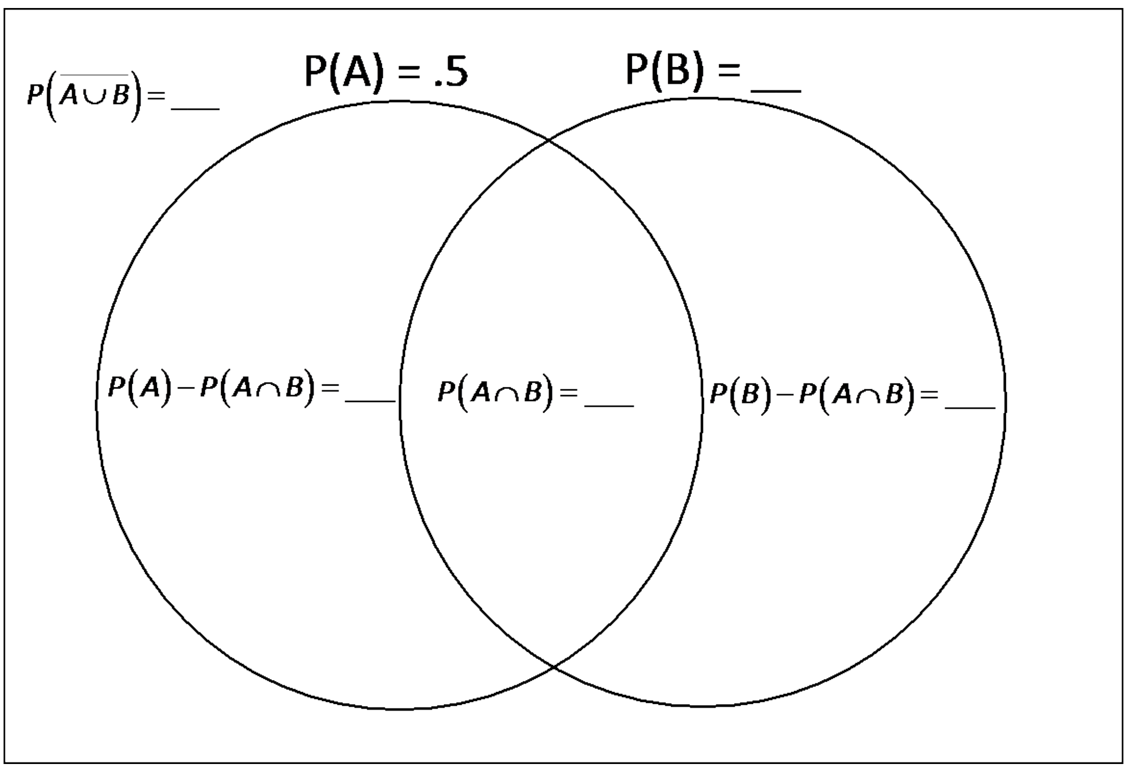
In a Venn diagram the set of outcomes that are *not* included in some set is called the complement of that set. The notation used for the complement of set A is $\overline{A}$, read "A bar", or ~A, read "not A". For example, in the Venn diagram you completed above, the outcomes that are outside of $A \cup B$ are denoted $\overline{A \cup B}$.

j. Which outcomes appear in $\overline{A} - B$?

k. Which outcomes appear in $\overline{B}-\left(\overline{A\cup B}\right)$?

5. The investigation of the Venn Diagram in question 4 should reveal a new way to see that the probability of rolling at least one odd number on two dice is $\frac{27}{36}=\frac{3}{4}$. How does the Venn diagram show this probability?

6. 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.



a. Fill in the remaining probabilities in the Venn diagram.

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

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

d. 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. Each student should identify themselves by their 3 initials (first, middle, last). Any student who listens to both Country and Hip Hop may be listed in both categories. Record results of the class poll in the table.

Hip Hop (HH)	Country (C)	Neither (N)

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 $P(\overline{C})$.

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

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

8. In part 1, you found the relationship between A, B, $A \cup B$, and $A \cap B$ to be $A \cup B = A + B - A \cap B$. In a similar way, write a formula for $P(A \cup B)$.

9. Now find $P(HH \cup C)$ using the formula instead of the Venn diagram. Did you get the same answer as you did in f above?

10. In what situation might you be forced to use the formula instead of a Venn Diagram to calculate the union of two sets?

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. Again, you will collect data on students in your class, record the data in a **two-way frequency table**, and then create a Venn diagram to organize the results of the poll. Which social networking site do you prefer?

2. Record results from the class poll in the table.

	Twitter (T)	Facebook (FB)
Female (F)		
Male (M)		

3. Draw a Venn diagram to organize your outcomes. *(Hint: Notice that male and female will not overlap and neither will Twitter and Facebook).*

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 $P\left(\overline{FB} \cap F\right)$.

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

The Conditions are Right

Standards Addressed in this Task

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

MGSE9-12.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 context.

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

8. Look for and express regularity in repeated reasoning by expecting students to understand broader applications and look for structure and general methods in similar situations.

Common Student Misconceptions

1. Students may believe that multiplying across branches of a tree diagram has nothing to do with conditional probability.
2. Students may believe that independence of events and mutually exclusive events are the same thing.
3. Students may believe that the probability of A or B is always the sum of the two events individually.
4. Students may believe that the probability of A and B is the product of the two events individually, not realizing that one of the probabilities may be conditional.

Imagine the last time you entered to win a raffle at a fair or carnival. You look at your ticket, 562104. As they begin to call off the winning ticket, you hear 562, but everyone has the same first 3 digits. Then 1 and 0 are called off. You know that excited feeling you get? Did you know there is a lot of math behind that instinct you feel that you might just win the prize? Now imagine those times when you are waiting to get your latest grade back on your English test. You're really not sure how you did, but as your teacher starts to talk about test results, her body language just isn't positive. She keeps saying things like "well, you guys tried hard." Again, there is significant math happening behind that sinking feeling you now have. In this task, you will be investigating how probability can be used to formalize the way real-life conditions change the way we look at the world.

Part 1 - A Game of Pig

To begin this task, you and your team members will compete in a dice game called Pig. The object of the game is to score the most points after 10 rounds of dice rolls. Your score is equal to the sum of all the dice that you roll. If you roll 5 then 5 then 3 then 2 your score is at 15. Your turn starts with a single die roll. You are allowed to keep rolling with the following restrictions:

- If you roll 6 at any time, another die is added to your pool. After the first 6, you will have two die to roll, after the second 6, you will have three to roll. Keep in mind if you roll more than one 6, more than one die is added.
- If you roll 1 at any time, your turn is immediately over, and your score for that turn is 0. It does not matter if it is the first roll or the twentieth.
- You may stop your turn after any single roll, record your score, and pass play to the next player.

You can keep score below. Play a few games, and while you play try to take note of successful strategies.

Comment(s):

Although most students will be able to understand the game by reading the rules, some may need a demonstration. The game is designed to move fairly quickly, so allowing time to play the game a couple of times through is appropriate.

The purpose of playing this game is to motivate the ideas inherent in conditional probability. As situations change, the probabilities tied to those situations change as well. In this variation of the game Pig, students should find that it is more and more likely to lose your score as the number of dice increase.

Players / Round				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Total				

1. Regardless of whom won, what kind of strategies were most successful? Least successful? Explain why you think so.

Comment(s):

Students should have responses that resonate with the idea that the chance of losing your score increases as you roll additional 6's. Actual determinations on when it is "too dangerous" to keep rolling will depend upon the student's own treatment of risk versus reward. More advanced students may think the strategy is "obvious." This does not mean that they know how to best describe this obvious strategy, and they should be pushed to do so.

Solution(s):

The most successful strategy is first to always reroll the first dice. Beyond the first dice, it depends on how many points you have earned. With two dice it is not too risky to take the chance if you are at 10 points; your chance of getting a 1 is low enough that scoring more points is vital to winning. Rolling the dice after you have 4 or more is always a bad strategy.

2. How does your strategy change as you roll more 6's? How many dice is too dangerous to keep rolling?

Comment(s):

Students should not be expected to do too many calculations here (although, they are included for your benefit below). Instead, demonstrating their understanding of the fact that the probabilities vary as the dice increase is what is important here

Solution(s):

The chance of losing increases each time you add dice. Adding a single dice decreases your odds of keeping your turn alive from 5/6 = 0.833 to 25/36 = 0.694. The third die makes your chance of staying alive only 125/216 = 0.579. A fourth changes your chance to stay alive 625/1296 = 0.482. Although risk-loving students will argue that this is a good chance to keep your turn alive, most will concede that 4 dice and above is too dangerous to keep rolling. The chance of your turn ending for any k amount of dice is $(5/6)^k$.

3. How would your strategy change if you only lost if you rolled at least two 1's at the same time?

Comment(s):

This question is intentionally open-ended and harder to make a conclusion about than the original game. Again, in order to motivate conditional probability, it is more important that students recognize the probabilities changing with the new rule. Precise calculations are not expected or required.

Solution(s):

You should always keep rolling if you have two dice. The chance of your turn ending with two dice is only 1/36. With three dice, the chance of your turn ending is still only 2/27 = 0.074. Students should not easily produce this exact calculation, but they should still feel safe rolling with three dice. Four dice have a 19/144 = 0.131 chance of landing at least two 1's. Five dice: 763/3888 = 0.196. Six dice: 0.263. Seven dice: 0.330. Students may not think seven dice still presents a good chance to score points. Allowing them to try a few trials of seven dice may be appropriate.

Part 2 – An Introduction to Conditional Probability

As you were able to see by playing Pig, the fact that the probability in a given situation can change greatly affects how a situation is approached and interpreted. This sort of idea is prevalent across society, not just in games of chance. 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.

Comment(s):
Having played a game that draws upon conditional probability, students will now be moved to make those calculations and then formalize the notation for it. This part of the task accomplishes this by having students investigate a phenomenon that they should be familiar with.

	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?

Comment(s):

Students should be familiar with the connection here, but answers need to reference the table, specifically. Students should not necessarily be making calculations here.

Solution(s):

The table has fewer smokers, but the smokers have more cases of 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):

Comment(s):

These answers should only begin to give students a sense of conditional probabilities. Students may respond to e), the chance that you have lung cancer and smoke as 0.046 by saying that it isn't actually likely that smoking causes lung cancer. This probability just determines that a person chosen at random is both a smoker and has lung cancer. Since you are looking for two very specific qualities from a random person (rather than looking for causality in one from the other) the probability remains small.

Solution(s):

a) $P(S)$ **0.45 = 45%**

b) $P(\overline{S})$**0.55 = 55%**

c) $P(L)$ **0.056 = 5.6%**

d) $P(\overline{L})$ **0.944 = 94.4%**

e) $P(L \cap S)$ **0.046 = 4.6%**

f) $P(\overline{S} \cap \overline{L})$ **0.54 = 54%**

g) $P(\overline{S} \cap L)$ **0.01 = 1%**

h) $P(S \cap \overline{L})$**0.404 = 40.4%**

i) $P(S \cup L)$ **0.46 = 46%**

j) $P(\overline{S} \cup \overline{L})$ **0.954 = 95.4%**

3. In order to use probability to reinforce the connection between smoking and lung cancer, you will use calculations of *conditional probability*.
Comment(s):
Students should be using only portions of the table to answer these questions. If a student struggles with a particular condition, showing them how to cover up (ignore) the row or column not being considered will help. With all three questions students should see that by narrowing your view to a condition (a single row on the table), your probabilities can change considerably.

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

Solution(s):

23/225 = 0.102

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

Solution(s):

The value is about double that in 2c, indicating that you are twice as likely to develop lung cancer if you are a smoker.

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?

Solution(s):

5/275 = 0.0182

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 $P\left(L \mid S\right)$.

Comment(s):

Here, students are asked to become familiar with formal notation. Writing out the probability as a sentence need not be verbatim as listed below. However, it is important that students name the proper condition with the "given."

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

Solution(s):

"What is the probability of developing lung cancer given that you are not a smoker."

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

Solution(s):

$P(L \mid \overline{S})$

5. Find the probability that a person was a smoker given that they have developed lung cancer and represent it with proper notation.

Solution(s):

$P(S \mid L) = 23/28 = 0.821$

6. Find the probability that a given cancer-free person was not a smoker and represent it with proper notation.

Comment(s):

Students may have trouble interpreting this question and may need to try and reword it before working.

Solution(s):

$P(\overline{S} \mid \overline{L}) = 270/472 = 0.572$

7. How does the probability in number 6 compare to $P\left(\overline{L} \mid \overline{S}\right)$? Are they the same or different and how so?

Comment(s):

This is an opportunity to reiterate with students the importance of correctly identifying the condition versus the probability they seek. Had they misinterpreted number 6 above as $P\left(\overline{L} \mid \overline{S}\right)$ ***rather than*** $P(\overline{S} \mid \overline{L})$***, their solution would have been very different.***

Solution(s):

Number 6 is asking for $P(\overline{S} \mid \overline{L}) = 270/472 = 0.572$ ***which is the probability that a person was a non-smoker given that the individual is cancer-free. This is very different from*** $P\left(\overline{L} \mid \overline{S}\right) = {}^{270}\!/_{275} = 0.982$ ***which is the probability that a person did not have lung cancer given that the person did not smoke.***

8. Based upon finding the conditional probabilities make an argument that supports the connection between smoking and lung cancer.
Comment(s):

Students will most likely point to the answer from 3a and 3b in that the chance of having lung cancer doubles if you are a smoker. Students may also point to question 5 as it shows that 82% of those diagnosed with lung cancer were smokers, or they may point out that you are much less likely to have lung cancer if you do not smoke from 3c. Students should be encouraged to answer this question using multiple calculations as well as clear and concise mathematical language in order to make a solid argument.

Solution(s):

The probabilities in question 3 show that cancer is more likely to occur if we know that a person already is a smoker, and they show that the probability of not having cancer is more likely if we know the person is* not *a smoker. Further, in question 5, it is shown that if you have lung cancer, there is a 82% chance you were a smoker.

Part 3 – A Formula for Conditional Probability

Comment(s):

The standard here asks for students to make good sense of the conditional probability formula. During this portion of the task, students work on their own and together with the teacher as a class.

Students commonly overlook the fact that the condition in the equation is always the denominator in the right hand side. Having students write the RHS for $P(B \mid A)$ correctly is a good test of how well they have read and understood the formula.

The formulaic definition of conditional probability can be seen by looking at the different probabilities you calculated in part 2. 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. As a formula,
.

Probability of *A* given *B*

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

In part 2 you found that $P(S) = \frac{225}{500}$ and $P(L \cap S) = \frac{23}{500}$. Using the formula for conditional probability is another way to determine that $P(L \mid S) = \frac{23}{225}$:

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

1. Using the same approach that is shown above, show that the conditional probability formula works for $P(\overline{S} \mid \overline{L})$.

Solution(s):

$$P(\overline{S} \mid \overline{L}) = \frac{P(\overline{S} \cap \overline{L})}{P(\overline{L})} = \frac{270/500}{472/500} = \frac{270}{472} = 0.572$$

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

Solution(s):

0.711

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

Solution(s):

0.667

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

Comment(s):

The unknown is the denominator of the conditional probability formula, so the formula can be manipulated to $P(B) = \frac{P(A \cap B)}{P(A \mid B)}$.

Solution(s):

0.787

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

Comment(s):

The unknown is the numerator in the conditional probability formula, so the formula can be manipulated to $P(A \cap B) = P(A \mid B)P(B)$.

Solution(s):

0.101

6. For two events G and H it is known that $P(H \mid G) = \frac{5}{14}$ and $P(H \cap G) = \frac{1}{3}$. Explain why you cannot determine the value of $P(H)$.

Comment(s):

This is a very common mistake students make with this topic.

Solution(s):

The conditional probability formula is specific in that the denominator of the right hand side will always be the given. Since G is the given in this situation, only P(G) can be determined, not P(H).

Part 4 – Box Office

Comment(s):

The introduction here is transparent. The intention is for students to see data that, even in a lighter context, can be seen as applicable to their learning.

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?

Comment(s):

Students may overlook that age does not seem to impact Captain America; the fact that Harry Potter is affected by the age bands does not mean all movies are.

Solution(s):

There is sometimes a relationship. The Harry Potter column is skewed toward the younger viewers, as 73 is more than triple 20, but the Captain America column is split evenly. For Harry Potter, age is clearly a factor in terms of who watches, but for Captain America age is not a factor.

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.

Comment(s):

Students may interpret "would mean to a movie executive" differently. The solutions presented are interpreting it as if a movie executive would like to understand how to better promote and sell more tickets for movies of the same type.

a. $P(E)$

Solution(s):

77/200 = 0.385. Around 39% of moviegoers are over the age of 30. This would tell an executive that they are more likely to be selling tickets to younger people. Knowing this may tell executives that more money is to be made on movies geared toward younger generations.

b. $P(H)$ and $P(C)$

Solution(s):

P(H) = 93/200 = 0.465 and P(C) = 107/200 = 0.535. This would indicate that Captain America is a more popular movie, but not by a lot.

c. $P(C\,|\,Y)$ and $P(H\,|\,Y)$

Solution(s):

P(C|Y) = 52/125 = 0.416 and P(H|Y) = 73/125 = 0.584. This indicates that young people are about 17% more likely to go see Harry Potter (or a movie like it) than Captain America. This confirms to movie executives that Harry Potter is a movie for younger generations.

d. $P(E\,|\,C)$ and $P(Y\,|\,C)$

Solution(s):

P(E|C) = 55/107 = .514 and P(Y|C) = 52/107 = 0.486. This indicates that a person who has gone to see Captain America is not likely to be a certain age, which would indicate that this movie can be catered to crowds of any age.

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

Comment(s):

Classroom discussions of this can be geared toward what the executives would do next with this information: How would this information help executives decide how to advertise for these films? Should they advertise more or less to older audiences for Harry Potter? Do you think movie executives expect these sorts of probabilities to be true? Why or why not?

Solution(s):

By looking at conditional probabilities, executives can conclude that certain movies, like Harry Potter, are more likely to have viewers who are under the age of 30. Other fun movies, such as Captain America, appeal to all ages.

Part 5 – ICE CREAM

Comment(s):

In this final portion of the task, students are investigating data that has both dependent and independent variables in the same setting. The purpose is to motivate the independence formula.

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?

Comment(s):

Being that students have now seen many two-way tables like this, it is expected that they should pick up on the 50/50 split for chocolate and fudge ripple rather quickly and the discrepancy with the other two flavors. If students have trouble understanding what this question is asking of them, a guiding question such as "which is more likely: a sugar cone with fudge ripple or a sugar cone with cotton candy?" will help.

Solution(s):

Cone choice can depend on flavor choice, but it is not a strict rule. Chocolate and fudge ripple each have equal amounts of cones sold, and this indicates that there is no relationship between cone and flavor choice if we are speaking about chocolate and fudge ripple. However, cotton candy and butter pecan have data that reflect something different: for these flavors, cone choice may vary. A cotton candy lover may be more likely to order a sugar cone, while butter pecan fans are more likely to order a waffle cone.

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

Solution(s):

a. $P(W)$ ***150/290 = 0.517 = 51.7%***

b. $P(S)$ ***140/290 = 0.483 = 48.3%***

c. $P(C)$ ***71/290 = 0.245 = 24.5%***

d. $P(BP)$ ***75/290 = 0.259 = 25.9%***

e. $P(FR)$ ***69/290 = 0.238 = 23.8%***

f. $P(CC)$ ***75/290 = 0.259 = 25.9%***

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	$P(S\mid C)$	$P(S\mid BP)$	$P(S\mid FR)$	$P(S\mid CC)$
	36/71 = 0.507	***19/75 = 0.253***	***34/69 = 0.493***	***51/75 = 0.68***
Waffle Cone	$P(W\mid C)$	$P(W\mid BP)$	$P(W\mid FR)$	$P(W\mid CC)$
	35/71 = 0.493	***56/75 = 0.747***	***35/69 = 0.507***	***24/75 = 0.32***

4. Compare and contrast the probabilities you found in question 2 with the conditional probabilities you found in question 3. Which flavors actually affect cone choice? Which do not? How did you make this determination?

Comment(s):

The question asks students to verbalize again what they said in question 1, but this time with the reinforcement of calculations. If students are not reaching the same conclusion as question 1, they will need redirection in order to make the connection.

Solution(s):

Each probability in the table is calculated with the condition "given FLAVOR." For chocolate and fudge ripple, the probabilities that a certain cone is chosen is about 50% each time, indicating that it is just as likely to choose either cone. For butter pecan and cotton candy this is not the case. After knowing that cotton candy is chosen it is much more likely that sugar cone is chosen, and after knowing that butter pecan is chosen it is much more likely that a waffle cone is chosen.

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. Since $P(S\mid C)$ and $P(W\mid C)$ are approximately the same probability, this indicated that based on our sample that the choice of cone appears to not be affected by the choice of ice cream; it appears the relationship is independent. When probabilities change depending on the conditional category, such as knowing sugar cones are more likely with cotton candy ice cream, the events are said to have a **dependent** relationship. Answer the questions below to ensure you understand this new terminology:

Comment(s):

An important digression here: the fact that students are seeing the term independence for the first time should be a barrier in terminology only. They have experienced the idea of independence before, it has just never been designated as such. If they struggle with the definition, talk them through the questions below. They should not have trouble grasping the fact that a coin is not depending upon another. Then return to the formal definition.

The solutions below can be explained in a variety of ways. No formal calculations are necessary, sound arguments should be the focus of a correct answer.

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

Solution(s):

These are independent events. A coin landing heads or tails on one flip never has any impact on a further flip. The same is true for two separate coins. You are no more or less likely to flip heads or tails based upon how another coin lands.

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?

Solution(s):

These events are dependent. The chance of pulling a white marble out of the bag changes as the game progresses. For instance, if you begin by pulling out three red marbles in a row, there are now only 5 red marbles in the bag. Since there are 2 white marbles, the ratio of white to red is now 2:5 making it more likely to pull a white now, 2/7, than at the beginning of the game when the chance of drawing a white marble was 2/10.

c. Explain whether or not the dice game of Pig you played represents independent or dependent events.

Solution(s):

The game of Pig is a game of dependent probabilities. Whenever a 6 is rolled, the chance that your score is erased on future rolls increases, as you have more opportunity to roll a 1. While any individual die is independent of another, the probability of losing is related to the dice as a group.

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?

Comment(s):

Students may try to be too particular about the probabilities being exactly equal. It is important to discuss that the probabilities obtained in this task are built from data tables, and they will therefore never be "ideal" in the same respect that theoretical probabilities are. There is no definite rule to indicate whether or not two experimental probabilities can be considered equal. It helps to think of the question of "independent or dependent" as a question with three answers instead of two; the third answer being inconclusive. For students who prefer a definite answer, it is good to point out that when you are in the gray area between dependent and independent, it is safer to err on the side of the events being dependent. If it turns out that the condition is irrelevant, then $P(A|B)=P(A)$ ***so there is, essentially, no harm done.***

Solution(s):

$P(W|C)=P(W)$, $P(S|FR)=P(S)$ and $P(W|FR)=P(W)$

7. For independent events, the conditional probability formula, $P(A|B)=\dfrac{P(A\cap B)}{P(B)}$, becomes $P(A)=\dfrac{P(A\cap B)}{P(B)}$. Solve this equation for $P(A\cap B)$ and place it in the box below.

Probability of Independent Events *A* and *B*
$P(A\cap B)=P(A)\cdot P(B)$

Notice that for *independent events*, the probability of two events occurring together is simply the product of each event's individual probability.

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.

Comment(s):

This question is designed to do two things: formatively assess the students' understanding of the independence formula and begin to move into the idea that independence can be tested for by using the formula.

a. Use the formula to verify that Fudge Ripple and Waffle Cone are independent events.

Solution(s):

$P(W) = 0.517$
$P(FR) = 0.238$
$P(W \cap FR) = 0.121$
$P(W) \cdot P(FR) = 0.123$

Since the product is the same as the intersection, both are about 12%, the events are independent.

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

Solution(s):

We have already seen that the choices of cotton candy and waffle cone are not independent. The formula is only proper to use when two events are known to be independent.

The Conditions are Right

Standards Addressed in this Task

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

MGSE9-12.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 context.

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

8. Look for and express regularity in repeated reasoning by expecting students to understand broader applications and look for structure and general methods in similar situations.

Imagine the last time you entered to win a raffle at a fair or carnival. You look at your ticket, 562104. As they begin to call off the winning ticket, you hear 562, but everyone has the same first 3 digits. Then 1 and 0 are called off. You know that excited feeling you get? Did you know there is a lot of math behind that instinct you feel that you might just win the prize? Now imagine those times when you are waiting to get your latest grade back on your English test. You're really not sure how you did, but as your teacher starts to talk about test results, her body language just isn't positive. She keeps saying things like "well, you guys tried hard." Again, there is significant math happening behind that sinking feeling you now have. In this task, you will be investigating how probability can be used to formalize the way real-life conditions change the way we look at the world.

Part 1 - A Game of Pig

To begin this task, you and your team members will compete in a dice game called Pig. The object of the game is to score the most points after 10 rounds of dice rolls. Your score is equal to the sum of all the dice that you roll. If you roll 5 then 5 then 3 then 2 your score is at 15. Your turn starts with a single die roll. You are allowed to keep rolling with the following restrictions:

- If you roll 6 at any time, another die is added to your pool. After the first 6, you will have two die to roll, after the second 6, you will have three to roll. Keep in mind if you roll more than one 6, more than one die is added.
- If you roll 1 at any time, your turn is immediately over, and your score for that turn is 0. It does not matter if it is the first roll or the twentieth.
- You may stop your turn after any single roll, record your score, and pass play to the next player.

You can keep score below. Play a few games, and while you play try to take note of successful strategies.

Players / Round				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Total				

1. Regardless of who won, what kind of strategies were most successful? Least successful? Explain why you think so.

2. How does your strategy change as you roll more 6's? How many dice is too dangerous to keep rolling?

3. How would your strategy change if you only lost if you rolled at least two 1's at the same time?

Part 2 – An Introduction to Conditional Probability

As you were able to see by playing Pig, the fact that the probability in a given situation can change greatly affects how a situation is approached and interpreted. This sort of idea is prevalent across society, not just in games of chance. 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) $P(S)$

b) $P(\overline{S})$

c) $P(L)$

d) $P(\overline{L})$

e) $P(L \cap S)$

f) $P\left(\overline{S} \cap \overline{L}\right)$

g) $P\left(\overline{S} \cap L\right)$

h) $P\left(S \cap \overline{L}\right)$

i) $P\left(S \cup L\right)$

j) $P\left(\overline{S} \cup \overline{L}\right)$

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 $P\left(L \mid S\right)$.

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.

6. Find the probability that a given cancer-free person was not a smoker and represent it with proper notation.

7. How does the probability in number 6 compare to $P\left(\overline{L} \mid \overline{S}\right)$? Are they the same or different and how so?

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

Part 3 – A Formula for Conditional Probability

The formulaic definition of conditional probability can be seen by looking at the different probabilities you calculated in part 2. 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. As a formula,

$$\boxed{\begin{array}{c}\textbf{Probability of } \boldsymbol{A} \textbf{ given } \boldsymbol{B} \\[1em] P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}\end{array}}$$

In part 2 you found that $P(S) = \frac{225}{500}$ and $P(L \cap S) = \frac{23}{500}$. Using the formula for conditional probability is another way to determine that $P(L \mid S) = \frac{23}{225}$:

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

1. Using the same approach that is shown above, show that the conditional probability formula works for $P(\overline{S} \mid \overline{L})$.

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

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

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

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

6. For two events *G* and *H* it is known that $P(H \mid G) = \frac{5}{14}$ and $P(H \cap G) = \frac{1}{3}$. Explain why you cannot determine the value of $P(H)$.

Part 4 – Box Office

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. $P(E)$

b. $P(H)$ and $P(C)$

c. $P(C \mid Y)$ and $P(H \mid Y)$

d. $P(E \mid C)$ and $P(Y \mid C)$

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

Part 5 – ICE CREAM

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	$P(S \mid C)$	$P(S \mid BP)$	$P(S \mid FR)$	$P(S \mid CC)$
Waffle Cone	$P(W \mid C)$	$P(W \mid BP)$	$P(W \mid FR)$	$P(W \mid CC)$

4. Compare and contrast the probabilities you found in question 2 with the conditional probabilities you found in question 3. Which flavors actually affect cone choice? Which do not? 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 $P(S \mid C)$ and $P(W \mid C)$ 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. Explain whether or not the dice game of Pig you played represents independent or dependent events.

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 $P(S \mid C) = P(S)$. Which other parts of the table from question 3 can be written in a simpler way?

7. For independent events, the conditional probability formula, $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, becomes $P(A) = \dfrac{P(A \cap B)}{P(B)}$. Solve this equation 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.

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 to verify that Fudge Ripple and Waffle Cone are independent events.

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

Formative Assessment Lesson: Modeling Conditional Probabilities 1: Lucky Dip

Source: Formative Assessment Lesson Materials from Mathematics Assessment Project
http://map.mathshell.org/materials/download.php?fileid=1215

ESSENTIAL QUESTIONS:

- How do you understand conditional probability?
- How do you represent events as a subset of a sample space using tables and tree diagrams?
- How do you communicate your reasoning clearly?

TASK COMMENTS:

Tasks and lessons from the Mathematics Assessment Project are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding. For more information access the MAP website:
http://www.map.mathshell.org/materials/background.php?subpage=formative

The task, *Modeling Conditional Probabilities 1: Lucky Dip*, is a Formative Assessment Lesson (FAL) that can be found at the website: http://map.mathshell.org/materials/lessons.php?taskid=409&subpage=problem

The FAL document provides a clear lesson design, from the opening of the lesson to the closing of the lesson.

The PDF version of the task can be found at the link below:
http://map.mathshell.org/materials/download.php?fileid=1215

STANDARDS ADDRESSED IN THIS TASK:

Understand independence and conditional probability and use them to interpret data

MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or*, *and*, *not*).

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results*.

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

Standards for Mathematical Practice
This lesson uses all of the practices with emphasis on:

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

The Land of Independence

Standards Addressed in this Task
MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

8. Look for and express regularity in repeated reasoning by expecting students to understand broader applications and look for structure and general methods in similar situations.

Common Student Misconceptions

1. Students may believe that multiplying across branches of a tree diagram has nothing to do with conditional probability.
2. Students may believe that independence of events and mutually exclusive events are the same thing.

Part 1 – Confirming Independence

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.

Comment(s):

The introduction is emphasizing the necessary and sufficient nature of the equation $P(A \cap B) = P(A) \cdot P(B)$. It does so in order to get students ready to use that equation to confirm whether or not two events are independent when their probabilities are known. The task begins by having students confirm independence using the formula with raw numbers. As these are not tied to any data, the answers will be clean and the product of the two probabilities will always be exactly that of the intersection. Please note that this will not be the case for the real data examples.

With the U.S. Census data, there is room for some interpretation of the results. It is rare in any real world settings for two variables to ever be truly independent. Of course, this is true for any math once it is taken off the page and turned into engineering or science. The answers

presented here are typical, but not absolute, answers. The task was written to avoid the gray areas, but do not be surprised when some students, especially more creative thinkers, go beyond what was intended on the page. The same is true for your own interpretations. If the data present new ideas to you, it is by no means necessary to go "by the book." A student's exposure to the interpretive side of mathematics is an active part of the Standards for Mathematical Practice.

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

Comment(s):

Students should simply be substituting into the formula here to confirm that the products are equal to the intersection.

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

Solution(s):

a) not independent
b) independent
c) not independent
d) independent

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

Comment(s):

Students should again be substituting into the formula to solve for the missing piece. Students may also notice that they need to just divide the intersection by the individual event given to determine the missing event's probability.

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

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

Solution(s):

a) *0.25*
b) *10/21 = 0.476*

Part 2 – Independence and Inference

With knowledge of probability and statistics, statisticians are able to make *statistical inferences* 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.

Comment(s):

The data presented in these tables is data from the 2000 US Census. It was sampled using Fathom, a statistical software package. For each data set, some values were excluded based on perceived relevance, and the excluded points are explained before each data set. Each sampling began with a sample size of 500 before being modified to the sample sizes that appear.

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.

Comment(s):

In this data set, those data points that did not have an income listed or were listed as zero were not included. These included data points that were clearly too young to have income, could possibly have been retired with no income (based upon the age listed), or simply not have any income.

	Under $10,00	Between $10,000 and $40,000	Between $40,000 and $100,000	Over $100,000	
Male	15	64	37	61	***177***
Female	31	73	14	58	***176***
	46	***137***	***51***	***119***	***353***

By finding different probabilities from the table above, make a determination about whether or not income level is affected by gender. Investigate whether your conclusion is true for all income levels. Show all the calculations you use and write a conclusion using those calculations.

Comment(s):

The types of solutions here will vary greatly upon the interpretation of the question. Looking at the full picture of all incomes will lead students to the conclusion that gender affects income, as the table is skewed toward males as income increases. The exception is the highest income bracket, which is where differing opinions may occur. The consistency between male and female in this bracket may allow some students to think with less constraint than just the numbers provide: if the highest paying jobs are split evenly amongst males and females, does the inequality really exist? What, then, about the lowest income bracket? How can there be inequity in one income bracket but not another? If this data set seems to raise more questions than answers with students then they are doing the critical thinking required of them by the Standards for Mathematical Practice. The idea that mathematical analysis opens a door for a way to encounter the world is a great discussion to have with students. However, in order to temper the discomfort of different answers, it will be useful to, as a class, re-frame the question so that everyone's approach to the calculations can be more streamlined.

Solution(s):

$P(M) = 177/353 = 0.501$
$P(F) = 176/353 = 0.499$

$P(lowI) = 46/353 = 0.130$
$P(lowI \cap M) = 15/353 = 0.024$
$P(lowI) \cdot P(M) = 0.0651$
$P(highI) = 119/353 = 0.337$
$P(highI \cap F) = 58/353 = 0.164$
$P(highI) \cdot P(F) = 0.168$

$P(midI) = 51/353 = 0.144$
$P(midI \cap F) = 14/353 = 0.040$
$P(midI) \cdot P(F) = 0.00576$

1st possible answer: The calculations made for confirming independence indicate that income and gender are dependent upon one another. When looking at the intersection and products of gender with different levels of income, the relationship $P(A \cap B) = P(A) \cdot P(B)$ does not necessarily hold. It seems to hold for upper incomes, as each probability is about 16.5%, but it does not hold for the other income levels. The lowest income is off by 4% (2.4% and 6.5%). The same is true for the middle income and female probabilities, as they compare at 4% and 0.6%. These do not lead to a strong indication that gender and income act independently, otherwise these probabilities would be much closer.

2nd possible answer: The calculations made for confirming independence vary in their consistency. However, there is evidence to show that gender and income act independently of one another with respect to the highest income brackets. When looking at the probabilities for the intersection of these events compared to the product of the individual probabilities, they are virtually the same, 16.4% and 16.8%, respectively. This is not the case for the lower income brackets, but the question of income and gender inequality is usually one that asks whether or not women can achieve as much financial success as men. And for the highest income bracket (and* only *the highest income bracket) the probabilities indicate that gender and income are independent.

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.

***Comment(s)*:**
In this data set, only those data points that had a positive amount of income were used (no zeroes). Many of these data points were of a child's age. Some were of adult age, and the interpretation is that they do not have a job. Those data points were not included.

	Pays under $500	Pays between $500 and $1000	Pays over $1000	
Below high school	57	70	30	***157***
High school diploma	35	47	11	***93***
College degree	24	62	40	***126***
	116	***179***	***81***	***376***

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.

Comment(s):
The solution shown below is not necessarily exhaustive in terms of calculations, and it need not be. Students may think to calculate many more probabilities than the ones shown below. Guide students to calculate only probabilities that are relevant to the question. Here, the question of the connection between housing expenditure and education is approached by looking at the chances of having a large housing expenditure and having a good education,

having a low housing expenditure and having inadequate education, and having average housing expenditure and having an average amount of education. It would certainly be just as relevant to have calculations that compare the chance of living in an upper class home with having little education; these calculations should similarly show that there is a dependent relationship. For instance, looking at just one level of income and calculating all the probabilities for that income at different levels of education would be more than enough to show the variables are dependent.

Solution(s):

$P(Col) = 126/476 = 0.338$

$P(HighHE) = 81/376 = 0.215$

$P(Col \cap HighHE) = 40/376 = 0.106$

$P(Col) \cdot P(HighHE) = 0.0727$

$P(HS) = 93/376 = 0.247$

$P(MidHE) = 179/376 = 0.476$

$P(HS \cap MidHE) = 47/376 = 0.125$

$P(HS) \cdot P(MidHE) = 0.118$

$P(BelH) = 157/376 = 0.418$

$P(LowHE) = 116/376 = 0.309$

$P(BelH \cap LowHE) = 57/376 = 0.152$

$P(BelH) \cdot P(LowHE) = 0.129$

1st possible answer: The relationship between education and income is a dependent relationship. If you compare the chances of having a college degree to the chances of living with high housing expenses you see that the product of the probabilities (about 7%) is not equal to the chance of being both college educated and living with low housing expenditures (about 10%). This variation seems to indicate these two variables are not independent. The classical idea that a good education leads to a more comfortable life, financially, seems to be true here. Further, there is a similar discrepancy in the probabilities of having little education and lower housing expenses. The product shows around 13% while the intersection shows around 15%, again indicating that these variables are not independent of one another.

2nd possible answer: (This concedes some of the points made in the first interpretation, but focuses on the calculations for the middle tier of housing expenses) The dependence of housing expenses on education is shown from the calculations for high cost housing compared with college education and for low cost housing compared with less education (again, this argument should include some of the particulars from above). However, when looking at the middle tier housing calculations, the probabilities for having a high school diploma and living in median cost housing seem to confirm independence as the intersection and product are each about 12%. This indicates that mid-level housing acts independently of having what we normally think of as a mid-level education.

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.

Comments:

This table compiles only those data points that had a travel time listed. Other than the young or the retired, the other group that may be overlooked are those that work at home, and therefore have no commute time. The Census does not distinguish between these three possibilities, they are all listed with a commute time as zero.

	Under 30 minutes	Between 30 minutes and an hour	Over an hour	
Male	65	24	15	***104***
Female	64	22	7	***93***
	129	***46***	***22***	***197***

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.

Comments:

Of the three tables, this is the table that seems most likely to reflect independence because of the first two columns as each is split about 50/50, slightly favoring males. It is expected that students will notice this, as this idea of a 50/50 split appears briefly in task 2 as well. If students do get hung up on anything, it will be the commute of over an hour. There are twice as many males as females with a long commute, and the fact that the values here are small may allow students to draw quick conclusions without making any calculations. However, the fact that the sample size is so small causes the probabilities to be no more off than the under 30 minute column. Below you will find a solution that allows for a student to conclude dependence based upon these two columns being slightly off, 1.5%, but this solution is considered fringe, at best. Most students should be come to a conclusion more like the 1st suggested solution.

Solutions:

$P(M) = 104/197 = 0.528$

$P(F) = 93/197 = 0.472$

$P(short) = 129/197 = 0.655$ $P(med) = 46/197 = 0.234$

$P(short \cap F) = 64/197 = 0.325$ $P(med \cap M) = 24/197 = 0.122$

$P(short) \cdot P(F) = 0.309$ $P(med) \cdot P(M) = 0.124$

$P(long) = 0.112$

$P(long \cap M) = 15/197 = 0.0761$

$P(long) \cdot P(M) = 0.0591$

***1st possible solution**: By using the relationship for independent events, $P(A \cap B) = P(A) \cdot P(B)$, and considering the two variables gender and commute time, it can be concluded that these to variables act independently of one another. The probabilities for different commute lengths when compared with gender tend to show that the product of a commute time with a gender is equal to the intersection of the two. For commutes below an hour long, the independence equation holds especially well. The probabilities for females and short commutes have a product of 31% and an intersection of 32.5%, and the probabilities for a 30-60 minute commute and male have a product of 12.4% and an intersection of 12.4% and 12.2%.*

***2nd possible solution**: The data in the table, combined with knowledge of $P(A \cap B) = P(A) \cdot P(B)$ for independent probabilities indicate that there may be some dependence between the variables commute time and gender. For shorter and longer commutes, the probabilities of gender and commute times have products and intersections that are off by as much as 1.5%. The fact that this inconsistency happens twice indicates that these variables are not acting completely independently.*

The Land of Independence

Standards Addressed in this Task

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.
8. Look for and express regularity in repeated reasoning by expecting students to understand broader applications and look for structure and general methods in similar situations.

Part 1 – Confirming Independence

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

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

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. $P(A) =$ ______ $, P(B) = \dfrac{3}{10}, P(A \cap B) = \dfrac{1}{7}$

Part 2 – Independence and Inference

With knowledge of probability and statistics, statisticians are able to make *statistical inferences* 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.

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.

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

By finding different probabilities from the table above, make a determination about whether or not income level is affected by gender. Investigate whether your conclusion is true for all income levels. Show all the calculations you use and write a conclusion using those calculations.

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.

Formative Assessment Lesson: Modeling Conditional Probabilities 2

Source: Formative Assessment Lesson Materials from Mathematics Assessment Project
http://map.mathshell.org/materials/download.php?fileid=1200

ESSENTIAL QUESTIONS:

- How do you represent events as a subset of a sample space using tables and tree diagrams?
- How do you understand when conditional probabilities are equal for particular and general situations?

TASK COMMENTS:

Tasks and lessons from the Mathematics Assessment Project are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding. For more information access the MAP website:
http://www.map.mathshell.org/materials/background.php?subpage=formative

The task, *Modeling conditional Probabilities 2*, is a Formative Assessment Lesson (FAL) that can be found at the website: http://map.mathshell.org/materials/lessons.php?taskid=405&subpage=concept

The FAL document provides a clear lesson design, from the opening of the lesson to the closing of the lesson.

The PDF version of the task can be found at the link below:
http://map.mathshell.org/materials/download.php?fileid=1200

STANDARDS ADDRESSED IN THIS TASK:

Understand independence and conditional probability and use them to interpret data

MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent**.**

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results*.

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

Standards for Mathematical Practice
This lesson uses all of the practices with emphasis on:

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

Formative Assessment Lesson: Medical Testing

Source: Formative Assessment Lesson Materials from Mathematics Assessment Project
http://map.mathshell.org/materials/download.php?fileid=1275

ESSENTIAL QUESTIONS:

- How do you make sense of a real life situation and decide what math to apply to the problem?
- How do you understand and calculate the conditional probability of an event A, given an event B, and interpret the answer in terms of a model?
- How do you represent events as a subset of a sample space using tables, tree diagrams, and Venn diagrams?
- How do you interpret the results and communicate their reasoning clearly?

TASK COMMENTS:

Tasks and lessons from the Mathematics Assessment Project are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding. For more information access the MAP website:
http://www.map.mathshell.org/materials/background.php?subpage=formative

The task, *Medical Testing*, is a Formative Assessment Lesson (FAL) that can be found at the website: http://map.mathshell.org/materials/lessons.php?taskid=438&subpage=problem

The FAL document provides a clear lesson design, from the opening of the lesson to the closing of the lesson.

The PDF version of the task can be found at the link below:
http://map.mathshell.org/materials/download.php?fileid=1275

STANDARDS ADDRESSED IN THIS TASK:

Understand independence and conditional probability and use them to interpret data
MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results*.

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

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

MGSE9-12.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 context.

MGSE9-12.S.CP.7 Apply the Addition Rule, P(A or B) = P(A) + P(B) – P(A and B), and interpret the answers in context.

Standards for Mathematical Practice
This lesson uses all of the practices with emphasis on:
1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

FALSE POSITIVES (Career and Technical Education Task)

Source: National Association of State Directors of Career Technical Education Consortium
PDF: http://www.achieve.org/files/CCSS-CTE-Task-False-Positives-FINAL.pdf
Word: http://www.achieve.org/files/CCSS-CTE-Task-False-Positives-FINAL.doc

Introduction
Students will investigate the accuracy of a cancer test.

Standard Addressed in this Task
MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.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 context.

Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them by requiring students to make sense of the problem and determine an approach.

2. Reason abstractly and quantitatively by requiring students to reason about quantities and what they mean within the context of the problem.

4. Model with mathematics by asking students to use mathematics to model a situation by identifying important quantities.

6. Attend to precision by expecting students to attend to units as they perform calculations. Rounding and estimation are a key part.

Common Student Misconceptions
1. Students may believe that the probability of A and B is the product of the two events individually, not realizing that one of the probabilities may be conditional.

Formative Assessment Lesson: Compound Confusion

Source: Georgia Mathematics Design Collaborative

This lesson is intended to help you assess how well students are able to:

- Utilize what they already know about compound probability in the context of real world situations
- Reason qualitatively, choose and apply the correct type of compound probability to the given situation
- Understand the meaning of independent, dependent, mutually exclusive, conditional, and overlapping probability
- Solve scenarios using independent, dependent, mutually exclusive, conditional, and overlapping probability

STANDARDS ADDRESSED IN THIS TASK:

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

STANDARDS FOR MATHEMATICAL PRACTICE:

This lesson uses all of the practices with emphasis on:

1. Make sense of problems and persevere in solving them
4. Model with mathematics

TASK DESCRIPTION, DEVELOPMENT AND DISCUSSION:

Tasks and lessons from the Georgia Mathematics Design Collaborative are specifically designed to help teachers effectively formatively assess their students. The way the tasks and lessons are designed gives the teacher a clear understanding of what the students are able to do and not do. Within the lesson, teachers will find suggestions and question prompts that will help guide students towards understanding.

The task, *Compound Confusion* is a Formative Assessment Lesson (FAL) that can be found at: http://ccgpsmathematics9-10.wikispaces.com/Georgia+Mathematics+Design+Collaborative+Formative+Assessment+Lessons

A Case of Possible Discrimination (Spotlight Task)

This activity is adapted by one of the authors, Christine Franklin, from *Navigating through Data Analysis in Grades 9-12*, Burrill, Gail, Christine Franklin, Landy Godbold, and Linda Young; p. 29-41, NCTM, Reston, VA. 2003.

Georgia Standards of Excellence Addressed:
MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results*.

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

Standards of Mathematical Practice Addressed:
1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.
7. Look for and make use of structure.

A Case of Possible Discrimination

Statisticians are often asked to look at data from situations where an individual or individuals believe that discrimination has taken place. A well-known study of possible discrimination was reported in the *Journal of Applied Psychology.* The scenario of this study is given below. The study motivates statistical reasoning of a context that is consistently relevant (discrimination cases that often end up in court). This particular study examined whether discrimination was being practiced against women by male supervisors in the banking industry.

SCENARIO

Researcher conducted a study at a banking conference where 48 male bank supervisors were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as "routine' or whether the person's file should be held and other applicants interviewed. The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female. Of the 48 files reviewed, 35 were promoted. (B.Rosen and T. Jerdee (1974), "Influence of sex role stereotypes on personnel decisions," *J. Applied Psychology,* 59:9-14.)

TEACHER COMMENTARY TO PRELIMINARY QUESTIONS

This activity attempts to guide the students through the statistical investigative process: (1) Formulate a statistical question (2) Design a study to collect data that will be used to answer the statistical question (3) Analyze the data using appropriate numerical and graphical representations (4) Draw conclusions by interpreting the results and connecting to the original statistical question.

For question 1, it is important to emphasize that the variables being measured are categorical. We are considering a bivariate scenario where we are investigating whether an association exists between gender and promotion status. We call gender the explanatory variable being used to explain or predict the response variable, promotion status. We summarize these categorical variables with a statistic. Here we can use counts or percentages. The statistic is quantitative. Help the student not to view the counts or percentages as the data – the data is being either male or female or either promoted or not.

For question 2, there are two treatments for the explanatory variable of gender (male or female). To be an experiment, the researcher has control over assigning subjects to certain experimental treatments and then observes the subject's outcome for the response variable. In an observational study, the research simply observes the subject's outcome on a response variable without imposing a treatment on the subject (the treatment pre-exists for the subject).

It is important for the researcher to randomly assignment the banking supervisors to a treatment (either a male or female folder) to help control for bias; i.e., to balance the two treatment groups with respect to lurking variables.

For question 3, illustrate for the student that taking half (or 50%) of 35 is because there was the same number of male folders as female folders at 24 each. If the number of male folders had been different (let's say of the 48 folders distributed, 20 were labeled male and 28 were *labeled female, than we would expect (20/48) or 42% of the 35 promoted folders to be male for an expected count of about 15.*

For question 6, because the actual study results fall into the 'gray area', we need to think about how to statistically further investigate whether 21 would be considered unusual under a model of random variation (no discrimination). The students have two benchmarks – what they expect presuming no discrimination and what they would expect where in their minds, discrimination is clearly being practiced.

PRELMINARY QUESTIONS

1. What is the statistical question being asked by the researchers? What are the two variables of interest that the researchers will measure to answer this question? Are these variables categorical or quantitative?

   Is there an association between gender of an applicant and the promotion status of the applicant by bank supervisors? The variables of interest are gender (male or female) and promotion status (yes or no). These variables are categorical.

2. Why would this study be classified as an experiment and not an observational study?
   What would need to be assumed about the manner in which the files were distributed to the banking supervisors in order to infer that gender is the cause of the apparent differences?

   The researchers had control over the file labeled male or female that a banking supervisor received. To make cause and effect statements the researchers need to randomly distribute the files.

3. Suppose there was no discrimination involved in the promotions. Enter the expected numbers of males promoted and females promoted for this case in Table 1. Explain your choice of numbers. Explain your choice of numbers.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 1 No discrimination

Expect either 17 or 18 for the number of males promoted. Mathematically, we expect half of 35 or 17.5. But for the observed count in the table, we need a whole number (17 or 18).

4. Suppose there was strong evidence of discrimination against the women in those recommended for promotion. Create a table that would show this case.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 2 Strong case of discrimination against the women

Most students will write 23 or 24 for the number of males promoted although may also see the count of 22. Choosing one of these numbers is selecting all or almost males being promoted.

5. Suppose the evidence of discrimination against the women falls into a 'gray area'; i.e., a case where discrimination against the women is not clearly obvious without further investigation. Create a table that would show this case.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 3. 'Gray' case of discrimination against the women

Most students will enter 20 or 21 although a few may enter 19 for the number of males promoted. They will select these numbers since they are somewhere between the number of males they selected for the scenario of no discrimination versus the scenario of obvious discrimination.

Returning to the study reported earlier in the activity scenario, the results were reported that of the 24 "male" files, 21 were recommended for promotion. Of the 24 "female" files, 14 were recommended for promotion.

6. Enter the data from the actual discrimination study in Table 4.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 4. Actual discrimination study

21 will be entered for the observed count of males promoted.

7. Given a file was recommended for promotion, what percentage was male?
Given a file was recommended for promotion, what percentage was female?

For male, 21/35 = 60% ; for female, 14/35 = 40%.

8. Without exploring the data any further, do you think there was discrimination by the bank supervisors against the females? How certain are you?

Answers will vary – listen closely for how students justify their choice and help the student understand the justification must be statistical.

9. Could the smaller number of recommended female applicants for promotion be attributed to variation due to randomness or chance? What is your sense of how likely the smaller number of recommended females could have occurred by chance? Your benchmark for comparing what was expected for the number of females is contained in Table 1.

Presuming there is no discrimination we expect 17 or 18 males to be promoted out of 35. However, we know that from one experiment to the next there can be random (chance) variation in the total number of males promoted. We can't expect 17 or 18 each time. The question becomes, " What are typical counts we expect to observe under random variation (i.e., no discrimination in this context)"? Would observing 19 or 20 counts be typical? Would observing 21 be typical or would it be unusual?

10. Suppose that the bank supervisors looked at files of actual female and male applicants. Assume that all of the applicants were identical with regard to their qualifications and we use the same results as before (21 males and14 females). If a lawyer retained by the female applicants hired you as a statistical consultant, how would you consider obtaining evidence to make a decision about whether the observed results were due to expected variation (random process) or if the observed results were due to an effect; i.e., discrimination against the women?

Statisticians would formalize the overriding question by giving two statements, a ***null hypothesis*** *which represents no discrimination so that any departure from Table 1 is due solely to a chance process, and an* ***alternative hypothesis*** *which represents discrimination against the women*

I would consider designing a simulation that would model the experiment presuming no discrimination; i.e., carrying out a simulation where there are 48 file folders with 24 male and 25 female and then randomly promoting 35 of the file folders. For each simulation, I will count the number of male folders promoted. That is one statistic for a simulated experiment. I would them repeat the simulation several times. I will create a distribution of the different counts and observe which statistics (counts) are typical and which counts would be considered unusual under the model of random variation (no discrimination).

SIMULATION OF THE DISCRIMINATION CASE

TEACHER COMMENTARY TO SIMULATION QUESTIONS

For the questions to consider before creating the dot plot, note that with only 20 simulations, some of the simulated sampling distributions may appear skewed and not centered at 17.5 but we anticipate most of the simulated will begin to take on a symmetric shape and centered near

17 or 18. If the number of male folders was different from the number of female folders (for example, 20 male (42%) and 28 female (58%), then we would expect the simulated sampling distribution to be centered at a mean of 35(.42) = 14.7 and the shape to be right skewed.

For question 6, select 3 students to display the simulated sampling distribution they constructed. Attempt to select one student where in the 20 simulations, no value is larger than 20. Select another student where only once does a count occur that is 21 or greater. Select a third student where a count of 21 or higher occurs twice. The different P-values will be 0%, 5%, and 10%. Then have the class discuss how small is 'small enough' to reject the presumption of no discrimination.

So that the students can better visualize how the sampling distribution for this scenario will behave in the long run, collect all 20 simulations from each student and then combine all the simulations into one class simulated sampling distribution. Then have the students answer simulation questions 6-9 using the class sampling distribution.

Have the students reflect on the 4 parts of the statistical investigative process with this scenario. What is the role of a sampling distribution as it relates to statistical inference; i.e., making conclusion to answer a statistical question. What is the role of the probability we call a P-value?

For more detailed information on the statistical components of this activity, refer to the reference for this activity from the NCTM Navigating through Data Analysis in Grades 9 -12.

An article for students to consider can be found at the following link:
http://www.pnas.org/content/early/2012/09/14/1211286109.full.pdf+html

This recent research study used a similar research design to the study discussed in this task.

Using a deck of cards, let 24 black cards represent the males, and 24 red cards represent the females (remove 2 red cards and 2 black cards from the deck). This will simulate the 48 folders, half of which were labeled male and the other half female.

For the questions to consider before creating the dot plot, student answers may vary. We expect the simulated sampling distribution in the long run to be symmetric centered around a mean (or expected value) of 17.5

1: Shuffle the 48 cards at least 6 or 7 times to insure that the cards counted out are from a random process. You are simulating what can happen with random variation where no discrimination is being practiced.

2: Count out the top 35 cards. These cards represent the applicants recommended for promotion to bank manager. The simulation could be conducted more efficiently by dealing out 13 "not promoted" cards, which would be equivalent to dealing out

35 “promoted” cards.

3: Out of the 35 cards, count the number of black cards (representing the males).

4: On the number line provided, Figure 1, create a dot plot by placing a blackened circle or X above the number of black cards counted. The range of values for possible black cards is 11 to 24.

5: Repeat steps 1 – 4 nineteen more times for a total of 20 simulations.

Before you perform the 20 simulations, what would you expect for the shape of the distribution you create in the dot plot? At what value would you expect the distribution to be centered?

FIGURE 1

DOT PLOT TO BE USED TO GRAPH THE 20 SIMULATED RESULTS

11 12 13 14 15 16 17 18 19 20 21 22 23 24

Number of Men Promoted

You have created a sampling distribution that is the distribution of possible values of a statistic for repeated samples of the same size from a population. For the scenario under consideration, the number of black cards (number of males promoted) from each of the simulations is the statistic.

6: Using the results (the counts) plotted on the number line, estimate the likelihood that 21 or more black cards (males) out of 35 will be selected if the selection process is random; that is, if there is no discrimination against the women in the selection process.

Answers will most likely be 0%, 5%, or 10%.

*The probability of observing 21 or more black cards if the selection process is due to expected variation from a random process is called the **P-value.*** Typically a P-value less than 0.05 is considered *statistically significance* to where a researcher would reject the assumption of the results being due to random variation and concluding there is strong evidence to support that the results indicate discrimination against women.

7: Based on you P-value, what would you conclude? How would you answer your original statistical question?

Have students compare their P-value to a significance level of 5%. If P-value is less than or equal to 5%, reject no discrimination and conclude strong evidence to support discrimination is being practice by male supervisors against women for promotions.

8. Look at the dot plot and comment on the shape, center, and variability of distribution of the counts by answering the following.
   (a) Is the distribution somewhat symmetric, pulled (skewed) to the right, or pulled to the left?
   (b) Do you observe any unusual observation(s)? Does 21 or higher appear to be unusual observations?
   (c) Where on the dot plot is the lower 50% of the observations?
   (d) Estimate the mean of the distribution representing the number of black cards obtained out of the 20 simulations.
   (e) Comment on the variability of the data.

Answers will vary depending upon the individual simulated sampling distributions.

9. Is the behavior of this distribution what you might expect? Compare your answers to question 8 to your answers to the questions earlier before performing the 20 simulations.

Answers will vary depending upon the individual simulated sampling distributions. Expect most simulated sampling distributions appear similar to what was expected.

A Case of Possible Discrimination (Spotlight Task)

This activity is adapted by one of the authors, Christine Franklin, from *Navigating through Data Analysis in Grades 9-12*, Burrill, Gail, Christine Franklin, Landy Godbold, and Linda Young; p. 29-41, NCTM, Reston, VA. 2003.

Georgia Standards of Excellence Addressed:
MGSE9-12.S.CP.1 Describe categories of events as subsets of a sample space using unions, intersections, or complements of other events (*or, and, not*).

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results*.

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

Standards of Mathematical Practice Addressed:
1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
6. Attend to precision.
7. Look for and make use of structure.

Spotlight Task: A Case of Possible Discrimination

Statisticians are often asked to look at data from situations where an individual or individuals believe that discrimination has taken place. A well-known study of possible discrimination was reported in the *Journal of Applied Psychology.* The scenario of this study is given below. The study motivates statistical reasoning of a context that is consistently relevant (discrimination cases that often end up in court). This particular study examined whether discrimination was being practiced against women by male supervisors in the banking industry.

SCENARIO
Researcher conducted a study at a banking conference where 48 male bank supervisors were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as "routine' or whether the person's file should be held and other applicants interviewed. The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female. Of the 48 files reviewed, 35 were promoted. (B.Rosen and T. Jerdee (1974), "Influence of sex role stereotypes on personnel decisions," *J. Applied Psychology,* 59:9-14.)

PRELMINARY QUESTIONS

1. What is the statistical question being asked by the researchers? What are the variables of interest that the researchers will measure to answer this question? Are these variables categorical or quantitative?

2. Why would this study be classified as an experiment and not an observational study?

   What would need to be assumed about the manner in which the files were distributed to the banking supervisors in order to infer that gender is the cause of the apparent differences?

3. Suppose there was no discrimination involved in the promotions. Enter the expected numbers of males promoted and females promoted for this case in Table 1. Explain your choice of numbers.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 1 No discrimination

4. Suppose there was strong evidence of discrimination against the women in those recommended for promotion. Create a table that would show this case. Explain your choice of numbers.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 2 Strong case of discrimination against the women

5. Suppose the evidence of discrimination against the women falls into a 'gray area'; i.e., a case where discrimination against the women is not clearly obvious without further investigation. Create a table that would show this case. Explain your choice of numbers.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 3. 'Gray' case of discrimination against the women

Returning to the study reported earlier in the activity scenario, the results were reported that of the 24 "male" files, 21 were recommended for promotion. Of the 24 "female" files, 14 were recommended for promotion.

6. Enter the data from the actual discrimination study in Table 4.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 4. Actual discrimination study

7. Given a file was recommended for promotion, what percentage was male? Given a file was recommended for promotion, what percentage was female?

8. Without exploring the data any further, do you think there was discrimination by the bank supervisors against the females? How certain are you?

9. Could the smaller number of recommended female applicants for promotion be attributed to variation due to randomness or chance? What is your sense of how likely the smaller number of recommended females could have occurred by chance? Your

benchmark for comparing what was expected for the number of females is contained in Table 1.

10. Suppose that the bank supervisors looked at files of actual female and male applicants. Assume that all of the applicants were identical with regard to their qualifications and we use the same results as before (21 males and14 females). If a lawyer retained by the female applicants hired you as a statistical consultant, how would you consider obtaining evidence to make a decision about whether the observed results were due to expected variation (random process) or if the observed results were due to an effect; i.e., discrimination against the women?

Statisticians would formalize the overriding question by giving two statements, a ***null hypothesis*** *which represents no discrimination so that any departure from Table 1 is due solely to a chance process, and an* ***alternative hypothesis*** *which represents discrimination against the women*

SIMULATION OF THE DISCRIMINATION CASE

Using a deck of cards, let 24 black cards represent the males, and 24 red cards represent the females (remove 2 red cards and 2 black cards from the deck). This will simulate the 48 folders, half of which were labeled male and the other half female.

1: Shuffle the 48 cards at least 6 or 7 times to insure that the cards counted out are from a random process. You are simulating what can happen with random variation where no discrimination is being practiced.

2: Count out the top 35 cards. These cards represent the applicants recommended for promotion to bank manager. The simulation could be conducted more efficiently by dealing out 13 “not promoted” cards, which would be equivalent to dealing out 35 “promoted” cards.

3: Out of the 35 cards, count the number of black cards (representing the males).

4: On the number line provided, Figure 1, create a dot plot by placing a blackened circle or X above the number of black cards counted. The range of values for possible black cards is 11 to 24.

5: Repeat steps 1 – 4 nineteen more times for a total of 20 simulations.

Before you perform the 20 simulations, what would you expect for the shape of the distribution you create in the dot plot? At what value would you expect the distribution to be centered?

FIGURE 1

DOT PLOT TO BE USED TO GRAPH THE 20 SIMULATED RESULTS

11 12 13 14 15 16 17 18 19 20 21 22 23 24

Number of Men Promoted

You have created a sampling distribution that is the distribution of possible values of a statistic for repeated samples of the same size from a population. For the scenario under consideration, the number of black cards (number of males promoted) from each of the simulations is the statistic.

6: Using the results (the counts) plotted on the number line, estimate the likelihood that 21 or more black cards (males) out of 35 will be selected if the selection process is random; that is, if there is no discrimination against the women in the selection process.

The probability of observing 21 or more black cards if the selection process is due to expected variation from a random process is called the ***P-value.*** Typically a P-value less than 0.05 is considered *statistically significance* to where a researcher would reject the assumption of the results being due to random variation and concluding there is strong evidence to support that the results indicate discrimination against women.

7: Based on you P-value, what would you conclude? How would you answer your original statistical question?

8. Look at the dot plot and comment on the shape, center, and variability of distribution of the counts by answering the following.
   (a) Is the distribution somewhat symmetric, pulled (skewed) to the right, or pulled to the left?
   (b) Do you observe any unusual observation(s)? Does 21 or higher appear to be unusual observations?
   (c) Where on the dot plot is the lower 50% of the observations?

(d) Estimate the mean of the distribution representing the number of black cards obtained out of the 20 simulations.

(e) Comment on the variability of the data.

9. Is the behavior of this distribution what you might expect? Compare your answers to question 8 to your answers to the questions earlier before performing the 20 simulations.

Culminating Task: Are You Positive?

Standards Addressed in this Task

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

MGSE9-12.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 context.

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

8. Look for and express regularity in repeated reasoning by expecting students to understand broader applications and look for structure and general methods in similar situations.

The following task was adapted from an Op-Ed piece for the NYTimes.com entitled "Chances Are" by Steven Strogatz. http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

We know that modern medicine is rich with biology, chemistry and countless other branches of science. While mathematics does not make a lot of headlines for changing the way we think about our health and well-being, there are many ways in which mathematics is informing and improving the way scientists and doctors approach the world of medicine.

Consider the following data for a group of 1000 women. Of these women, 8 are known to have breast cancer. All 1000 undergo the same test to determine whether or not they have breast cancer, and 77 test positive. 7 of the 8 that have breast cancer test positive and 70 of those that do not have breast cancer test positive.

1. Based on the data presented above, do you think the test for breast cancer is effective at identifying women who have breast cancer? Why or why not?

   Comment(s):

   This question is designed for students to make an initial hypothesis regarding this test for breast cancer based simply on a first look at the data.

   Solution(s):

   Answers will vary. Students may draw some initial conclusions on the validity of the test given that 7 of the 8 women who have breast cancer tested positive. They may also notice the discrepancy between the number of women with breast cancer that tested positive and the number of women that tested positive who did not have breast cancer.

2. Organize the data using the table below.

Women with Breast Cancer *8*	Test Positive *7*	
	Test Negative *1*	
Women without Breast Cancer *992*	Test Positive *70*	
	Test Negative *922*	

3. Find the following probabilities:

a. The probability that a woman tests positive given that she has breast cancer.

Solution(s):

$$P(+\ test \mid cancer) = {}^{7}\!/_{8} = .875$$

b. The probability that a woman tests positive given that she does not have breast cancer.

Solution(s):

$$P(+\ test \mid no\ cancer) = {}^{70}\!/_{992} = .071$$

4. Determine whether or not the events of having breast cancer and testing positive for breast cancer are independent. Show all relevant calculations.

Solution(s):

$$P(cancer) = {}^{8}\!/_{1000} = .008$$

$$P(+\ test) = {}^{77}\!/_{1000} = .077$$

$$P(+\ test \mid cancer) = {}^{7}\!/_{8} = .875$$

$$P(cancer) \times P(+\ test) = {}^{8}\!/_{1000} \times {}^{77}\!/_{1000} = .0006$$

$$P(+\ test \mid cancer) \times P(cancer) = {}^{7}\!/_{8} \times {}^{8}\!/_{1000} = .007$$

$$P(cancer) \times P(+\ test) \neq P(+\ test \mid cancer) \times P(cancer)$$

Therefore, having breast cancer and testing positive for breast cancer are not independent.

The analysis you have done so far should seem straightforward. Testing for breast cancer is successful at identifying those women who have it. It is reassuring to know that testing positive for breast cancer is not independent of having breast cancer, as this seems to indicate that screening for breast cancer is an effective way to identify when a woman actually has breast cancer.

There is one aspect of this analysis that needs further inspection. The probabilities that you have found have the condition that a woman does or does not have breast cancer. In reality, a woman knowing this before getting tested is highly unlikely. The point of getting tested is to find out! While the results you have found seem sound, it will be good to find the probabilities from a more realistic standpoint.

5. Now you will look through the lens of a woman who tests positive for breast cancer.
   a. Find the probability that a woman has breast cancer given that her test result is positive.

Comment(s):

Students may take the more formulaic approach here and first have (7/1000) / (77/1000). Students may also think this answer is incorrect at first and may need assurance that they are correct.

Solution(s):

7/77 = 0.091 = 9.1%

b. What seems strange about this result to you?

Solution(s):

After seeing that the test is fairly successful at identifying women that have breast cancer, it now seems like we are saying that the test is only 9% successful, totally contradicting our earlier results.

c. Compare this probability to what you calculated in question 3a. What is causing these probabilities to be so different?

Comment(s):

Expect students to be able to come to the conclusion that the numerators are the same while the denominators are vastly difference. The more nuanced part of the answer will vary based upon the ability level of the student.

Solution(s):

This probability is much lower than was seen in question 3. In question 3, the sample space is only the 8 women who are already known to have breast cancer. In this question, the sample space is all 77 women that test positive. In both cases, the event is the same 7 women who both test positive and have breast cancer. The fact that there are 992 women that don't have breast cancer allow for many more women to be identified incorrectly. It isn't that the medical test is consistently wrong; it is that the test is given more opportunity to be wrong.

6. Let's also look through the lens of a woman who tests negative.

a. Find the probability that a woman does not have breast cancer given that her test result is negative.

Comment(s):

Again, students may arrive at this result using the formulaic approach of (992/1000) / (993/1000). Unlike #5, students should expect a result like this and find it more agreeable.

Solution(s):

922/923 = 0.9989

b. What does this result indicate?

Solution(s):

This indicates that a negative test result is almost always correct: a person who tests negative is very likely to be cancer-free.

7. This task focused specifically on medical tests for breast cancer. It is not a stretch to say that the efficacy of most medical tests is similar to that of what you have investigated here. Write a paragraph that discusses the use and effectiveness of medical tests in regards to the probability theory that underlies them.

Solution(s):

What this task has shown us is that test results are not always as cut and dry as they seem. In our investigation, we saw that a person who takes a certain test for breast cancer and tests positive is actually only about 9% likely to actually have cancer. This result certainly makes us wonder about how useful such a test really is. Why 9%? The fact is that most women do not have breast cancer, so the many women that are cancer free sometimes are falsely identified. Since this pool of women is so great, it overshadows those women that actually have cancer. However, we can still trust the effectiveness of the test. 7 of 8 women who have cancer are correctly identified by the test, and nearly 100% of women who do not have cancer are correctly identified as such. Rather than disregard test results, we must instead be wary of what a positive result tells us. When testing for cancer or other disease, a positive result should result in further testing and concern rather than all out panic.

Culminating Task: Are You Positive?

Standards Addressed in this Task

MGSE9-12.S.CP.2 Understand that if two events A and B are independent, the probability of A and B occurring together is the product of their probabilities, and that if the probability of two events A and B occurring together is the product of their probabilities, the two events are independent.

MGSE9-12.S.CP.3 Understand the conditional probability of A given B as P (A and B)/P(B). Interpret independence of A and B in terms of conditional probability; that is, 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.

MGSE9-12.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. *For example, use collected data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

MGSE9-12.S.CP.5 Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

MGSE9-12.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 context.

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them by requiring students to interpret and make meaning of a problem and find a logical starting point, and to monitor their progress and change their approach to solving the problem, if necessary.

2. Reason abstractly and quantitatively by requiring students to make sense of quantities and their relationships to one another in problem situations.

3. Construct viable arguments and critique the reasoning of others by engaging students on discussion of why they agree or disagree with responses, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. Model with mathematics by expecting students to apply the mathematics concepts they know in order to solve problems arising in everyday situations, and reflect on whether the results are sensible for the given scenario.

5. Use appropriate tools strategically by expecting students to consider available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a compass, a calculator, software, etc.

6. Attend to precision by requiring students to calculate efficiently and accurately; and to communicate precisely with others by using clear mathematical language to discuss their reasoning.

7. Look for and make use of structure by expecting students to apply rules, look for patterns and analyze structure.

8. Look for and express regularity in repeated reasoning by expecting students to understand broader applications and look for structure and general methods in similar situations.

The following task was adapted from an Op-Ed piece for the NYTimes.com entitled "Chances Are" by Steven Strogatz. http://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

We know that modern medicine is rich with biology, chemistry and countless other branches of science. While mathematics does not make a lot of headlines for changing the way we think about our health and well-being, there are many ways in which mathematics is informing and improving the way scientists and doctors approach the world of medicine.

Consider the following data for a group of 1000 women. Of these women, 8 are known to have breast cancer. All 1000 undergo the same test to determine whether or not they have breast cancer, and 77 test positive. 7 of the 8 that have breast cancer test positive and 70 of those that do not have breast cancer test positive.

1. Based on the data presented above, do you think the test for breast cancer is effective at identifying women who have breast cancer? Why or why not?

2. Organize the data using the table below.

<table>
<tr><td rowspan="2">Women with Breast Cancer</td><td>Test Positive</td></tr>
<tr><td>Test Negative</td></tr>
</table>

Women without Breast Cancer	Test Positive
	Test Negative

3. Find the following probabilities:

a. The probability that a woman tests positive given that she has breast cancer.

b. The probability that a woman tests positive given that she does not have breast cancer.

4. Determine whether or not the events of having breast cancer and testing positive for breast cancer are independent. Show all relevant calculations.

The analysis you have done so far should seem straightforward. Testing for breast cancer is successful at identifying those women who have it. It is reassuring to know that testing positive for breast cancer is not independent of having breast cancer, as this seems to indicate that screening for breast cancer is an effective way to identify when a woman actually has breast cancer.

There is one aspect of this analysis that needs further inspection. The probabilities that you have found have the condition that a woman does or does not have breast cancer. In reality, a woman knowing this before getting tested is highly unlikely. The point of getting tested is to find out! While the results you have found seem sound, it will be good to find the probabilities from a more realistic standpoint.

5. Now you will look through the lens of a woman who tests positive for breast cancer.

a. Find the probability that a woman has breast cancer given that her test result is positive.

b. What seems strange about this result to you?

c. Compare this probability to what you calculated in question 3a. What is causing these probabilities to be so different?

6. Let's also look through the lens of a woman who tests negative.

a. Find the probability that a woman does not have breast cancer given that her test result is negative.

b. What does this result indicate?

7. This task focused specifically on medical tests for breast cancer. It is not a stretch to say that the efficacy of most medical tests is similar to that of what you have investigated here. Write a paragraph that discusses the use and effectiveness of medical tests in regards to the probability theory that underlies them.