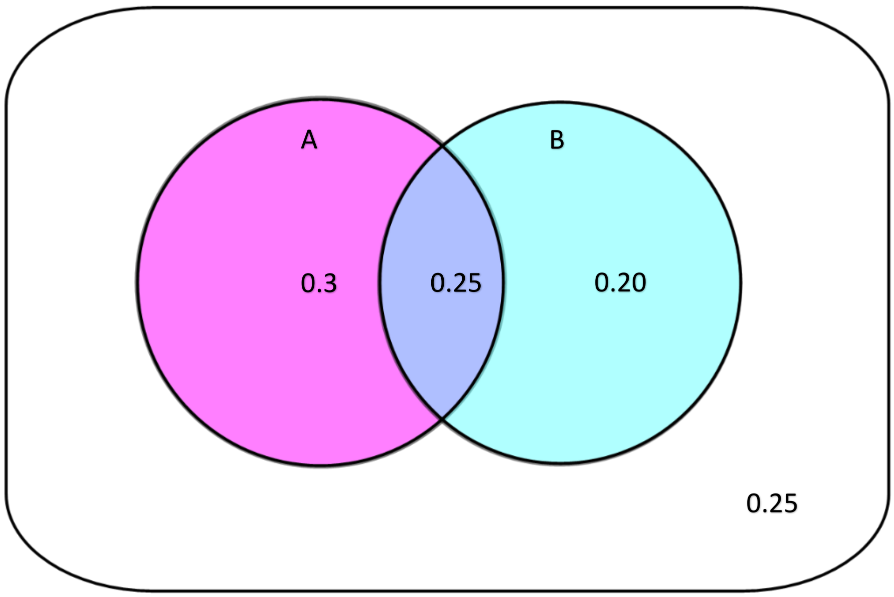
Venn Diagrams and Probability

Venn diagrams are a way of displaying the relationships between sets (or groups), and as such, can be useful for figuring out probabilities. Let's look at a couple of examples of how Venn diagrams can be useful for solving probability problems.

Example 1. Two Set Venn Diagram with Probabilities



First, check that all the probabilities add up to 1.

$$0.2 + 0.25 + 0.25 + 0.2 = 1.00$$

To read this diagram, all the probabilities corresponding to set A are $0.3 + 0.25 = 0.55$ (we are adding up the probabilities in A only, and that are also in B). All the probabilities in B are likewise the ones in B only and also in A: $0.20 + 0.25 = 0.45$. The remaining elements are in neither A nor B.

We can use the Venn diagram to answer some basic questions. Suppose that set A corresponds to the event that someone like Apples, and set B corresponds to the event that Blue is the person's favorite color.

a. What is the probability that a randomly selected person from the set likes Apples?

   The answer is the probability of being in set A, which we previously calculated to be 0.55.

b. What is the probability that a randomly selected person from the set doesn't select Blue as their favorite color?

   We previously found the probability associated with set B was 0.45, but this is probability of having Blue as a favorite color. So we calculate the probability of not liking blue to be the complement: $1 - 0.45 = 0.55$.

c. What is the probability of both liking Apples and having Blue as a favorite color?

This answer is where the two sets overlap, in the purple region: 0.25

d. What is the probability of either liking Apples OR having blue as a favorite color?

We can calculate this two ways. We can add up the things inside the sets: the pink region (for A only), the purple region (for both), and the blue region (for B only): $0.3 + 0.25 + 0.2 = 0.75$.

Or we can use the formula: $P(A) + P(B) - P(A \text{ } and \text{ } B) = 0.55 + 0.45 - 0.25 = 0.75$

We can see both ways give us the same answer. In the first method, we directly counted each section just once, but when we calculated the second way, we added in the purple region twice (since it's both in A and in B), so we had to subtract it off one time to get the correct value.

e. What is the probability of being neither an Apple lover nor someone who likes blue?

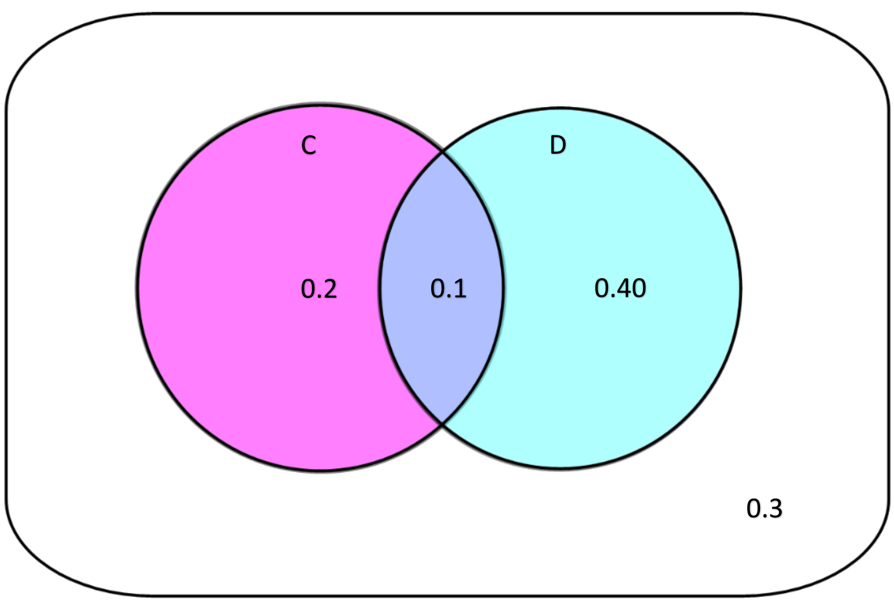
This is the complement of the event we calculated in the last question, so $1 - 0.75 = 0.25$, which corresponds to the probability of being outside either set (in the white region of the diagram).

f. Are A and B independent? To test this, we can check to see if $P(A) \cdot P(B) = P(A \text{ } and \text{ } B)$.

$$P(A) \cdot P(B) = 0.55 \cdot 0.45 = 0.2475$$
$$P(A \text{ } and \text{ } B) = 0.25$$

These value are really close, but $0.2475 \neq 0.25$ and so they can't be independent.

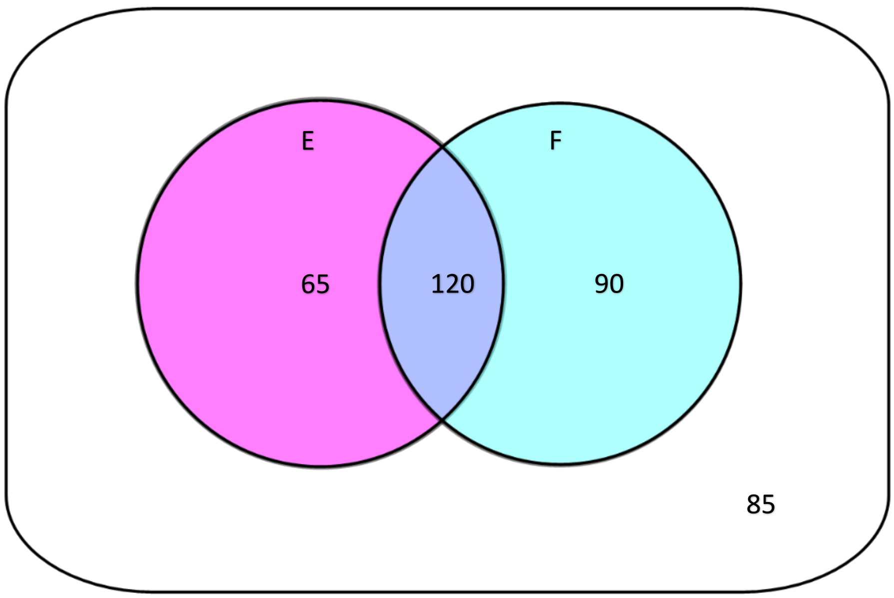
Practice Problem.



Suppose that set C corresponds to the event attends church every week, and set D corresponds to being a Democrat. Answer the following questions.

a. What is the probability a randomly selected person in the group attends church every week?
b. What is the probability a randomly selected person in the group is a Democrat?
c. What is the probability of a randomly selected person being both a regular churchgoer and a Democrat?
d. What is the probability of being either a regular churchgoer or a Democrat?
e. What is the probability of being neither a regular churchgoer nor a Democrat?
f. Are the events “regularly attends church” and “being a Democrat” independent events?

Example 2. Two Set Venn Diagrams with Counts



When dealing with counts, we can answer all the same kinds of questions we did above, as long as we do one thing first: find the total number of elements in the universal set. To do this, add up all the counts displayed: $65 + 120 + 90 + 85 = 360$. When we want to calculate probabilities, we now just add the counts we need, and then divide by this total.

Suppose that set E represents the event that the person likes a circus with elephants, and set F corresponds to the event that the person is in their 50s.

a. What is the probability that a randomly selected person likes elephants at the circus?

   First add up the total number of elements in set E: $65 + 120 = 185$, then divide that by the total: $\frac{185}{360} = \frac{37}{72} = 0.513\bar{8}$.

b. What is the probability that a randomly selected person is not in their 50s?

First, we can find out how many people ARE in their 50s: $120 + 90 = 210$, and then divide by the total: $\frac{210}{360} = \frac{7}{12} = 0.58\overline{3}$. And then find the complement: $1 - \frac{7}{12} = \frac{5}{12} = 0.41\overline{6}$. Or, we can just add up the counts not in set F: $65 + 85 = 150$, and then divide by the total: $\frac{150}{360} = \frac{5}{12}$.

c. What is the probability of being both a fan of elephants at circuses and also being in one's 50s?

This is the overlapping region of the sets: $\frac{120}{360} = \frac{1}{3}$

d. What is the probability of being either a fan of elephants at the circus or being in one's 50s?

$$65 + 120 + 90 = 275$$
$$\frac{275}{360} = \frac{55}{72} = 0.763\overline{8}$$

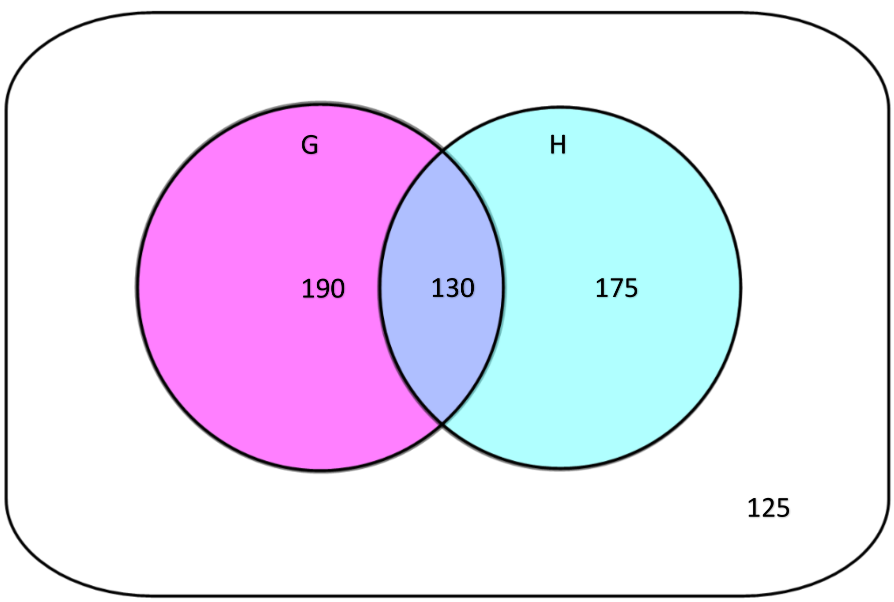
e. Are the events liking elephants at a circus and being in one's 50s independent?

We have to test here: $P(E) \cdot P(F) = P(E\ and\ F)$.

$$P(E) = \frac{185}{360} = \frac{37}{72},\ \ P(F) = \frac{210}{360} = \frac{7}{12}$$
$$P(E) \cdot P(F) = \frac{37}{72} \cdot \frac{7}{12} = \frac{259}{864} \approx 0.299769 \ldots$$
$$P(E\ and\ F) = \frac{1}{3} = 0.\overline{3}$$

These last two values are not the same, so they cannot be independent.

Practice Problem.



Suppose that set G represents the event that the person plays multi-player games online, and set H be the event that the person considers their marriage a happy one.

a. What is the probability that a randomly selected person has a happy marriage?
b. What is the probability that a randomly selected person does not play multi-player games online?
c. What is the probability that a randomly selected person has both a happy marriage and plays multi-player games online?
d. What is the probability of being either a multi-player online gamer, or has a happy marriage?
e. What is the probability that the randomly selected person neither plays multi-player games online nor has a happy marriage?
f. Are these two events independent?

Example 3. Converting a Two-Way Table to a Venn Diagram

Venn diagrams and two-way tables can often tell us some of the same things, so we can represent the same type of information in both forms. In this example we are going to convert the two-way table shown below into a two-set Venn diagram.

	Like Skateboards	Do Not Like Skateboards	Totals
Like Snowmobiles	80	25	105
Do not like Snowmobiles	45	10	55
Totals	125	35	160

So let's define our two sets. Let set K be the set of people who like skateboards. And let set N be the set of people who like snowmobiles.

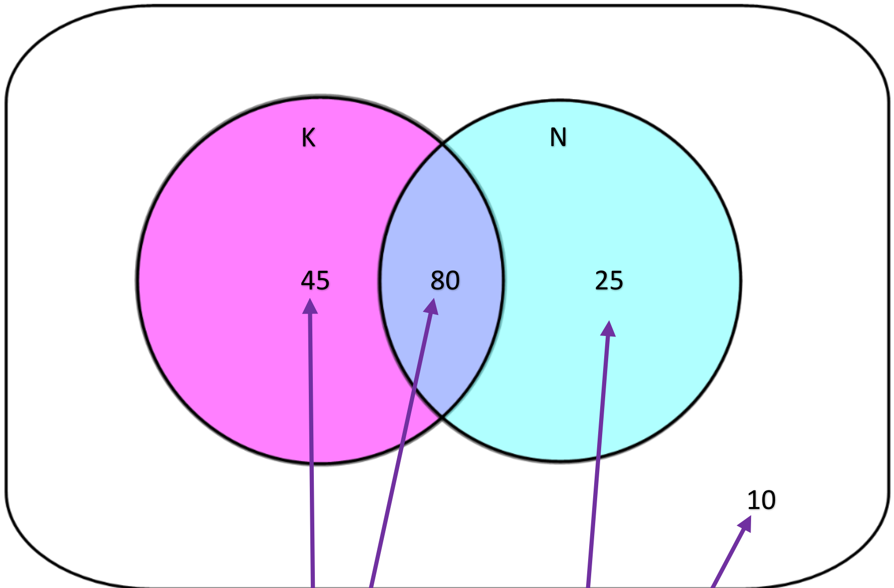
The number of people who both like skateboards and snowmobiles will go in the intersection (overlap in purple) of the two sets in the Venn diagram. (80)

The people who like skateboards but don't like snowmobiles will go in set K where the two sets don't overlap. (45)

The people who do like snowmobiles but not skateboards go into set N where there is no overlap. (25)

And then the number of people that like neither (10) go into the white space outside both sets.

Let's see what this looks like in the Venn diagram.



	Like Skateboards	Do Not Like Skateboards	Totals
Like Snowmobiles	80	25	105
Do not like Snowmobiles	45	10	55
Totals	125	35	160

You can now answer all the same probability questions about the diagram and the table using either representation.

Practice Problem.

Convert the following two-way table to a Venn diagram. Be sure to state what each set represents. Depending on how you define your sets, your Venn diagram may look slightly different, so the labels are extremely important.

Biking to School

	Can Bike	Can't Bike	Total
Boys	7	4	11
Girls	9	10	19
Total	16	14	30