# Using Venn diagrams to calculate probability of joint events

Corinne Riddell
September 25, 2020

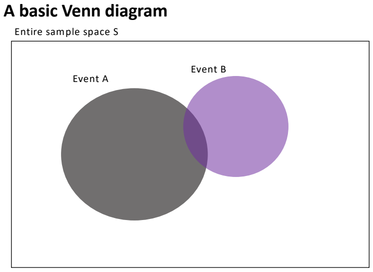
1

## Learning objectives for today

- How to make a Venn diagram
- How to use a Venn diagram to compute probabilities
- Determine whether two events are independent or dependent
- Conditional probability
- General addition rule for probability
- General multiplication rule for probability

2

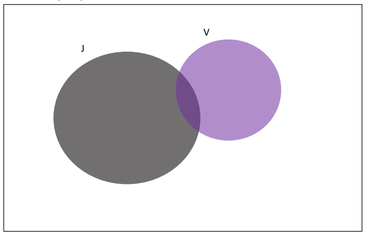
**A basic Venn diagram**




3

## A basic Venn diagram

Entire sample space S



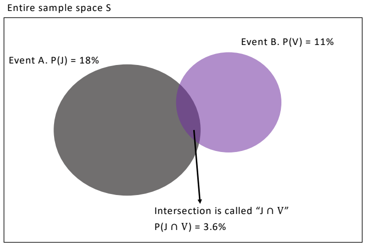

Suppose you had access to survey data about vaping and JUUL-related advertisements.

You could define the following random variables:
- J is the event "seen ad for JUUL"
- V is the event "vaped in the last 30 days"

4

## A basic Venn diagram

Entire sample space S

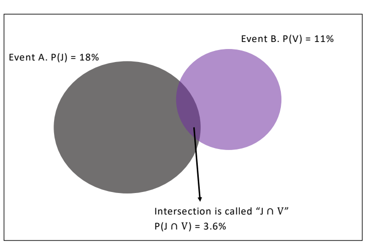



Think: ∩ stands for Intersection

You could also add the percentages you know from the survey data onto the Venn diagram.

5

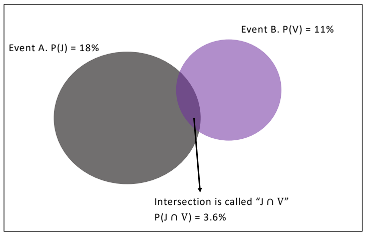
## Using a Venn diagram to calculate a probability




**What percent of individuals saw an ad for JUUL and do not vape?**

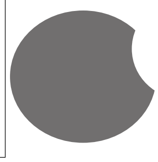
6

## Using a Venn diagram to calculate a probability



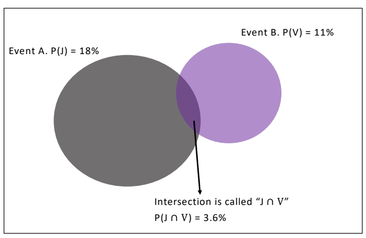

**What percent of individuals saw an ad for JUUL and do not vape?**

**This percent is represented by this area:**



7

## Using a Venn diagram to calculate a probability

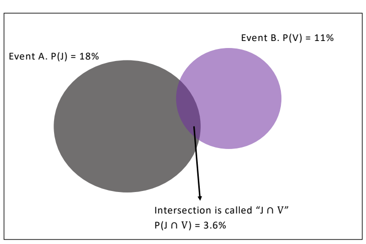



**What percent of individuals saw an ad for JUUL and do not vape?**

- Write this question as a probability statement
- Calculate the percentage

8

## Using a Venn diagram to calculate a probability




**What percent of individuals saw an ad for JUUL and do not vape?**

- Write this question as a probability statement

  Answer: $P(J \text{ and } V')$

- Calculate the percentage

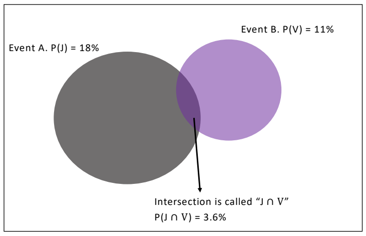
  Answer:

$P(J \text{ and } V')$
$= P(J) - P(J \cap V)$
$=0.18-0.036$
$= 0.144$
$=14.4\%$

- Conclude: 14.4% of individuals have seen a JUUL ad and do not vape.

9

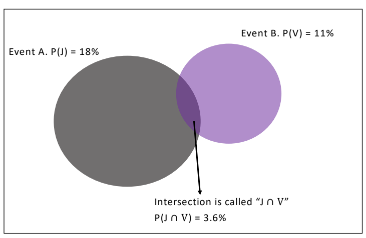
## Using a Venn diagram to calculate a probability



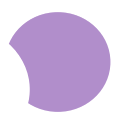

**Your turn: What percent of individuals vape but have not seen an ad for JUUL?**

10

## Using a Venn diagram to calculate a probability

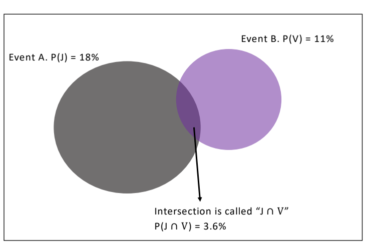



**Your turn: What percent of individuals vape but have not seen an ad for JUUL?**



11

## Using a Venn diagram to calculate a probability

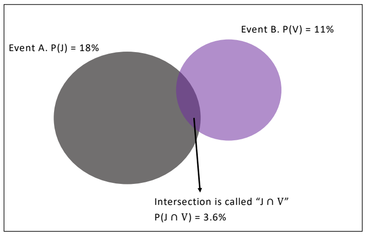



**Your turn: What percent of individuals vape but have not seen an ad for JUUL?**

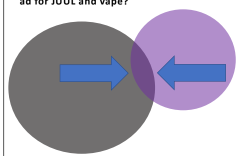
- Write this question as a probability statement
- Calculate the percentage
- Your answer:

12

## Generalized rule of addition




**What percent of individuals have seen an ad for JUUL and vape?**



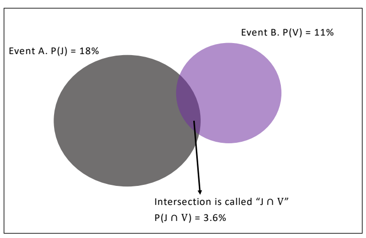
The "and" represents the intersection, the area where both events have occurred:

P(J ∩ V) = 3.6%

3.6% of individuals have seen an ad for JUUL and have vaped in past 30 days

13

## Generalized rule of addition




**What percent of individuals have seen an ad for JUUL or vaped in the last 30 days (or both)?**

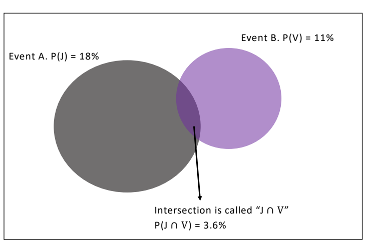
Think: **U** stands for **U**nion

P(J or V)
= P(J U V)
= P(J) + P(V) - P(J ∩ V)
= 0.18 + 0.11 – 0.036
= 0.256 = 25.6%

This is the **generalized rule for addition**. You need to subtract off the probability of the intersection so you do not double count the probability when two events are not disjoint!

14

## Conditional Probability

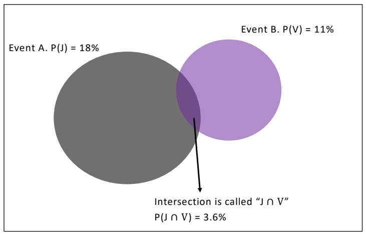



**Among individuals who have seen an ad for JUUL, what percent vaped in the past month?**

- This is a **conditional probability** question
- You can tell because of the way the question is phrased: "**Among individuals..**", that is, this probability is conditional on the event "seeing an ad for JUUL"
- We write conditional probabilities like this: **P(V|J)** and read this "Probability of V (occurring) **given** J (has occurred)"

15

## Conditional Probability




**Among individuals who have seen an ad for JUUL, what percent vaped in the past month?**

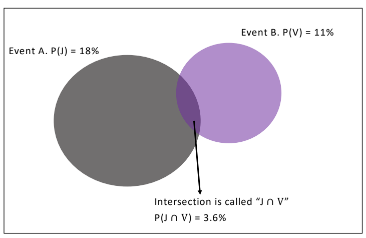
P(V|J) = P(V and J)/P(J)
= 0.036/0.18
= 0.2
= 20%

20% of individuals who have seen an ad for JUUL vaped in the last 30 days.

How does this 20% compare with the percent of individuals who vape among those who have not seen an ad for JUUL?

16

## Conditional Probability




**Among individuals who have not seen an ad for JUUL, what percent vaped in the past month?**

P(V|J')
= P(V and J')/P(J')

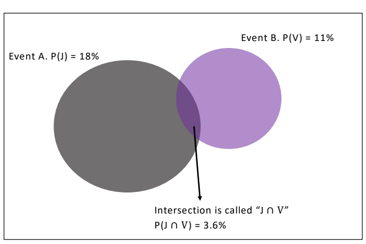
What is P(J')?
= 1 – 0.18 = 0.82

What is P(V and J')?
= 0.11 – 0.036
= 0.074

So:
P(V|J')
= P(V and J')/P(J')
= 0.074/0.82
= 0.0902439 … = 9.02%

17

## Conditional Probability




**Among individuals who have not seen an ad for JUUL, what percent vaped in the past month?**

P(V|J')
= P(V and J')/P(J')

What is P(J')?
= 1 – 0.18 = 0.82

What is P(V and J')?
= 0.11 – 0.036
= 0.074

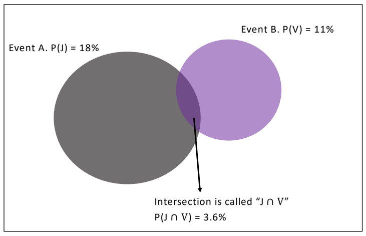
So:
P(V|J')
= P(V and J')/P(J')
= 0.074/0.82
= 0.0902439 … = 9.02%

From the last two slides: 9% of individuals who did not see an ad for JUUL vaped in the last month vs. 20% among those who did see an ad.

This could be the beginning of providing evidence that JUUL marketing affects the chance that an individual will vape.

18

## Conditional Probability and the Generalized Multiplication Rule




We can rewrite the formula for conditional probability to derive the **generalized multiplication rule**:

Formula for conditional probability
P(A|B) = P(A and B)/P(B)

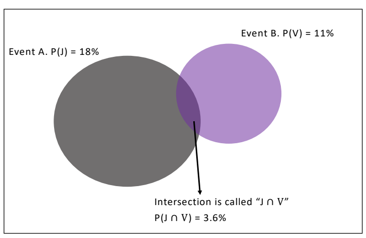
Rearrange for P(A and B):
**P(A and B) = P(A|B)×P(B)**
**P(A ∩ B) = P(A|B)×P(B)**

Thus, the probability that both A and B occur is equal to P(A|B)×P(B).

This is called the **General Multiplication Rule** for any two events.

19

## Independence




Two events A and B that both have a positive probability are independent if:

**P(B|A) = P(B)**

Under independence, we can rewrite the probability of A and B:

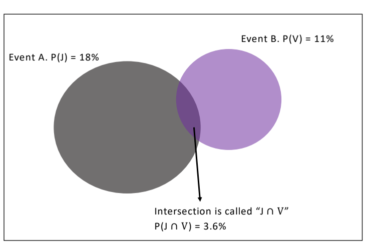
**P(A ∩ B) = P(A|B)×P(B)**
**Rewrite: P(A ∩ B) = P(A) X P(B)**

You can use either rule to verify independence.

**Are J and V independent?**

20

## Examine independence using rule #1




If two events are **independent** then the following statement is true:

**P(B|A) = P(B)**

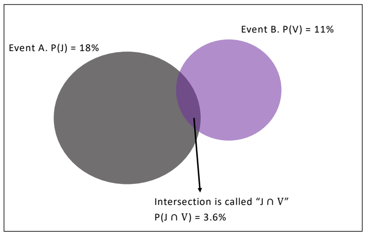
**Are J and V independent?**

1. We know that P(V) = 11%
2. From earlier questions, we also know that P(V|J) = 20%

Thus, P(V|J) does not equal P(V), so these events are not independent.

21

**Examine independence using rule #2**




If two events are **independent** then the following statement is true:

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

**Are J and V independent?**

1. $P(J \cap V) = 3.6\%$

2. $P(J) \times P(V) = 0.18 \times 0.11 = 0.0198 = 1.98\%$

Since these quantities are not equal, these two events are not independent.

22

# Additional slides on independence

23

## Independence

- What sorts of events are independent? Two events are independent if knowing that one event occurred does not change the probability that the other occurred

24

## Example 1a

- Down syndrome is a genetic disorder caused when abnormal cell division results in an extra full or partial copy of chromosome 21.[1]
- The largest risk factor for having a child with Down syndrome is advanced maternal age.[1]
- Suppose that Martha is 40 and her baby has been diagnosed with Down syndrome. Martha's best friend Jane, also 40, is hoping to conceive. Is her baby's risk of Down syndrome independent of Martha's baby's risk?

Reference: https://www.mayoclinic.org/diseases-conditions/down-syndrome/symptoms-causes/syc-20355977

25

## Example 1b

- The risk of having a baby with Down syndrome is 1/100 among 40 year olds. Suppose that Jenny and Samantha are two 40-year old women. What is the probability that they both have babies with Down syndrome?

26

## Example 2a

- Breast Cancer is the most common cancer diagnosed in women in the US.
- If a woman's mother, sister, or daughter was diagnosed with breast cancer, her estimated risk of breast cancer is doubled (compared to a woman whose mother/sister/daughter has not been diagnosed).[1]
- Jacqueline was diagnosed with breast cancer. Is her daughter's chance of developing breast cancer independent or dependent on Jacqueline's diagnosis?

Reference. https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470

27

## Example 2b

- A white woman aged 40 who had her first period at age 12 and first child at age 27. Her mother had Breast Cancer. Given this information, she has an estimated 5-year risk of breast cancer of 1.1% and lifetime risk of breast cancer of 18.8%.
- Had the woman's mother not had breast cancer, here 1-year risk would have been 0.6% and lifetime risk 11.1%. Based on these risk estimates, are the events "mother has breast cancer" and "daughter has breast cancer" independent? Why?
- (See here to calculate breast cancer risk under a variety of settings. For the above risk estimates I set the variables as specified and selected "Unknown" for having the BRCA1/2 gene, and ever having a breast biopsy. I selected "No" for every having a medical history of any breast cancer.)

28