AP Statistics
Unit 04 – Probability
Day 03 Notes – Conditional Probability

Name Key
Period

Conditional Probability: the probability that one event happens given that another event is already known to have happened.

Suppose we know that event A has happened. Then the probability that event B happens given that event A has happened is denoted by **P(B | A)**, which is the *probability of B given A*.

Calculating Conditional Probabilities:

The conditional probability P(B | A) is given by the formula:

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

Or, written in words rather than symbols:

$$\boxed{P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}}$$

Example: Pierced Ears

Let's look at the pierced ears problem again. This time, try answering questions about the conditional probability of a randomly selected student. Here is the data:

	Gender		
Pierced Ears?	**Male**	**Female**	**Total**
Yes	19	84	103
No	71	4	75
Total	90	88	**178**

Suppose we choose a student from the class at random. Find the probability that a student chosen at random…

a. Is a male given that they have pierced ears?

$$P(\text{male} \mid \text{pierced}) = \frac{P(\text{male and pierced})}{P(\text{pierced})} = \frac{19}{103} = 18.4\% \text{ + sentence}$$

b. Has pierced ears given that they are a male?

$$P(\text{pierced} \mid \text{male}) = \frac{P(\text{pierced and male})}{P(\text{male})} = \frac{19}{90} = 21.1\% \text{ + sentence}$$

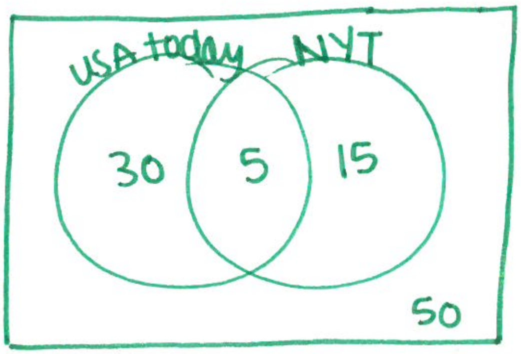
c. Is a female given that they do not have pierced ears?

$$P(\text{female} \mid \text{not pierced}) = \frac{P(\text{female and not pierced})}{P(\text{not pierced})} = \frac{4}{75} = 5.3\% \text{ + sentence}$$

Example: Who Reads the Paper?

The residents of a large apartment complex are classified based on their reading habits. 35% of the residents read *USA Today*, 20% read the *New York Times*, and 5% of residents read both.

a. Create a Venn diagram of the information given above.



b. Create a two-way table using the information given above.

USA Today \ NYT	yes	no	total
yes	5	30	35
no	15	50	65
total	20	80	100

c. What is the probability that a randomly selected resident who reads *USA Today* also reads the *New York Times*?

$$P(NYT \mid USA\ Today) = \frac{P(NYT \text{ and } USA\ Today)}{P(USA\ Today)} = \frac{5}{35} = 14.3\%$$

+ sentence

CHECK YOUR UNDERSTANDING:

Students at the University of New Harmony received 10,000 course grades last semester. The two-way table below breaks down these grades by which school of the university taught the course. The schools are Liberal Arts, Engineering and Physical Sciences, and Health and Human Services.

	Grade Level		
School	**A**	**B**	**Below B**
Liberal Arts	2142	1890	2268
EPS	368	432	800
HHS	882	630	588

College grades tend to be lower in engineering and the physical sciences (EPS) than in liberal arts and social studies (which includes Health and Human Services, known as HHS). Choose a University of New Harmony course grade at random. Consider the two events E: the grade comes from an EPS course and L: the grade is lower than a B.

1. Find P(L). Interpret this probability in context.
   $P(L) = 3656/10000 = 36.56\%$ → there is a 36.56% chance of selecting a course grade that is lower than a B.
2. Find P(E | L) and P(L | E). Which of these conditional probabilities tells you whether this college's EPS students tend to earn lower grades than students in liberal arts and social sciences? Explain.
   $P(E|L) = 800/3656 = 21.9\%$
   $P(L|E) = 800/1600 = 0.50 = 50\%$
   P(L|E) gives the prob. of getting a ↓ grade given the student studies EPS. Because 50% > P(L) = 36.56%, we can concude that grades are ↓ in EPS.

General Multiplication Rule:

The probability that events A and B both occur can be found using the general multiplication rule:

$$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B \mid A)$$

Where P(B | A) is the conditional Probability that event B occurs given that event A has already occurred.

In words, this rule says that for both of two events to occur, first one must occur, and then given that the first event has occurred, the second must occur.

Example: Teens with Online Profiles

The Pew Internet and American Life project find that 93% of teenagers (ages 12 to 17) use the Internet, and that 55% of online teens have posted a profile on a social-networking site.

Find the probability that a randomly selected teen sues the Internet and has posted a profile. Show your work.

P(online) = 0.93 and P(profile | online) = 0.55 so P(online and profile) = ?

$$P(\text{profile} \mid \text{online}) = \frac{P(\text{profile and online})}{P(\text{online})}$$

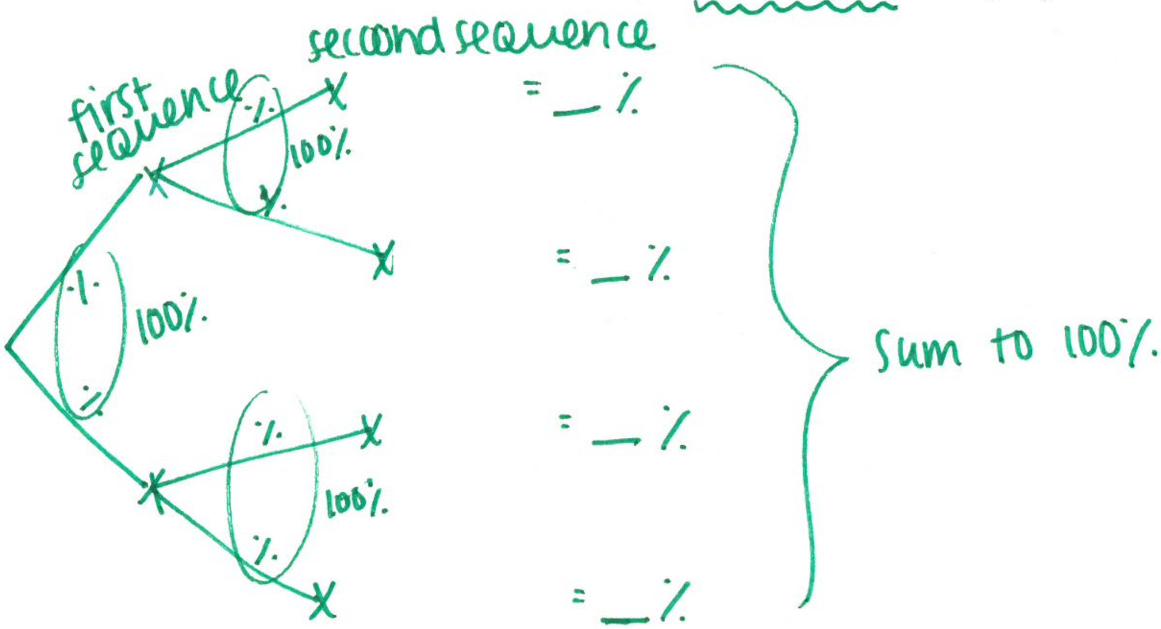
$$0.93 \cdot 0.55 = \frac{P(\text{profile and online})}{\cancel{0.93}} \cdot 0.93$$

P(profile and online) = 51.15%

+ sentence

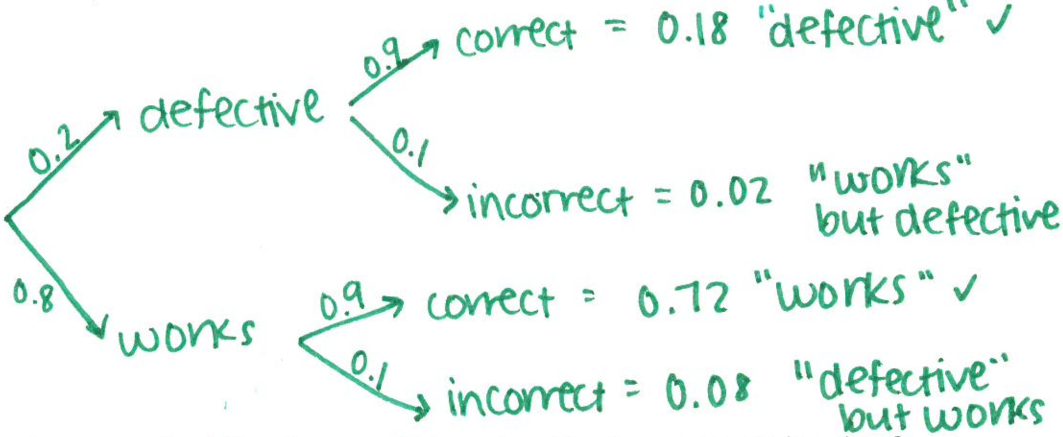
Tree Diagrams:

The general multiplication rule is especially useful when a chance process involves a sequence of outcomes. In such cases, we can use a tree diagram to display the sample space.



Example:

On a factory line, 20% of all items are defective. All items are checked before shipped but 10% of all the items are incorrectly classified as defective or non-defective. Draw a tree diagram to help solve the problem.



a. What percent of the items will be classified as non-defective?

$0.02 + 0.72 = 0.74$ or 74% of the items will be classified as non-defective.

b. What's the probability that the item works given that the item was correctly classified?

$$P(\text{works}|\text{correctly classified}) = \frac{P(\text{works and correctly classified})}{P(\text{correctly classified})}$$

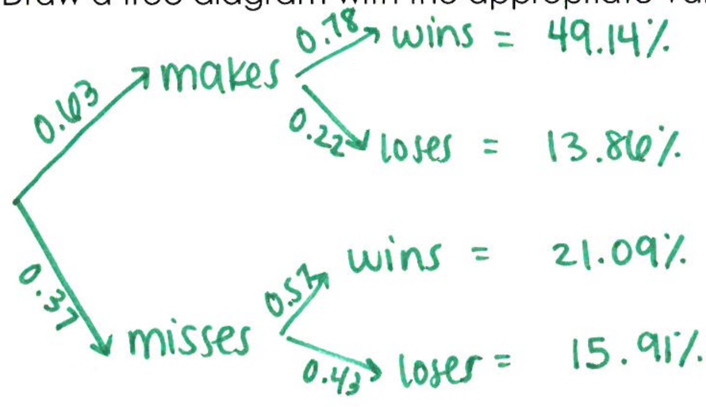
$$= \frac{0.72}{0.72+0.18} = 80\%$$

+sentence

Example: Serve It Up

Tennis great Roger Federer made 63% of his first serves in the 2011 season. When Federer made his first serve, he won 78% of the points. When Federer missed his first serve and had to serve again, he won only 57% of the points.

a. Draw a tree diagram with the appropriate values.



b. What's the probability that Federer makes the first serve and wins the point?

$P(\text{makes and wins}) = P(\text{makes}) \times P(\text{wins}) = 0.63 \cdot 0.78 = 49.14\%$

+sentence

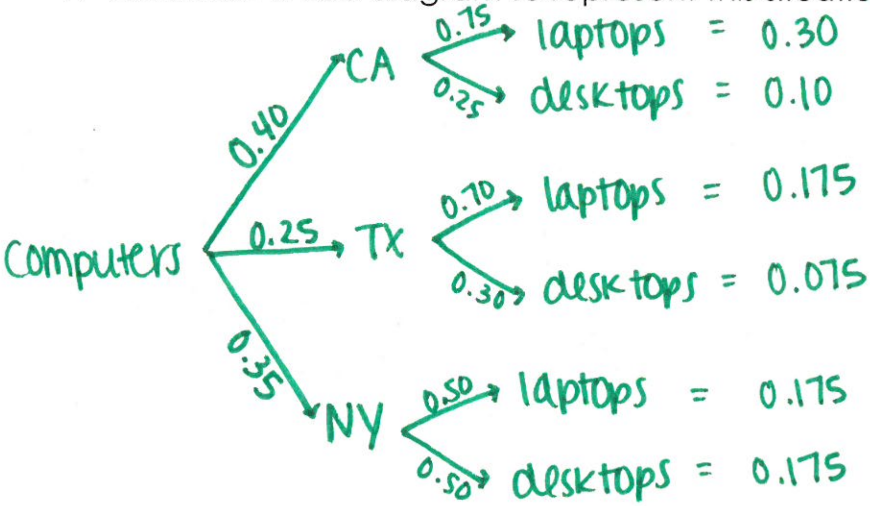
c. When Federer is serving, what's the probability that he wins the point?

$P(\text{wins}) = 0.4914 + 0.2109 = 70.23\%$ + sentence

CHECK YOUR UNDERSTANDING:

A computer company makes desktop and laptop computers at factories in three states: California, Texas, and New York. The California company produces 40% of the company's computers, the Texas factory makes 25%, and the remaining 35% are manufactured in New York. Of the computer made in California, 75% are laptops. Of those made in Texas and New York, 70% and 50%, respectively, are laptops. All computers are first shipped to a distribution center in Missouri before being sent out to stores. Suppose we select a computer at random from the distribution center.

1. Construct a tree diagram to represent this situation.



2. Find the probability that the computer is a laptop. Show your work.

   P(laptop) = 0.30 + 0.175 + 0.175 = 0.65 = 65% + sentence

3. Given that a laptop is selected, what is the probability that it was made in California.

   $P(CA \mid laptop) = \frac{0.30}{0.65} = 46.15\%$ + sentence

Independent Events:

Two events A and B are independent if the occurence of one event does not change the probability that the other event will happen.

In other words, **P(A | B) = P(A) and P(B | A) = P(B)**.

Example: Lefties Down Under

* Knowing that a student is male increases the probability that they are left-handed because 15.2% > 10%.

Is there a relationship between gender and handedness? To find out, we used CensusAtSchool's Random Data Selector to choose an SRS of 100 Australian high school students who completed a survey. The two-way table displays data on the gender and dominant hand of each student.

	Gender		
Dominant Hand	**Male**	**Female**	**Total**
Right	39	51	**90**
Left	7	3	**10**
Total	**46**	**54**	**100**

Are the events *male* and *left-handed* independent? Justify your answer.

$P(\text{left-handed} \mid \text{male}) = \dfrac{P(\text{left and male})}{P(\text{male})} \stackrel{?}{=} P(\text{left-handed})$ — be specific (more than this)

*sentence

$\dfrac{7}{46} = \dfrac{10}{100}$

15.2% ≠ 10%. So they are not independent

CHECK YOUR UNDERSTANDING:

For each chance process below, determine whether the events are independent. Justify your answer.

1. Shuffle a standard deck of cards, and turn over the top card. Put it back in the deck. Shuffle again, and turn over the top card. Define events A: first card is a heart and B: second card is a heart.

   Independent. Because we are replacing cards, knowing what the first card was will not help us predict what the second card will be.

2. Shuffle a standard deck of cards, and turn over the top two cards, one at a time. Define events A: first card is a heart and B: second card is a heart.

   Not independent. Once we know the suit of the first card, then the probability of getting a heart on the second card will change depending on what the first card was.

3. The 28 students in Mr. Tabor's AP Statistics class completed a brief survey. One of the questions asked whether each student was right- or left-handed. The two-way table summarizes the class data. Choose a student from the class at random. The events of interest are *female* and *right-handed*.

	21 Gender	7	
Dominant Hand	Female	Male	
Right	3	1	4
Left	18	6	24

Independent.

$P(\text{female}) \stackrel{?}{=} P(\text{female} \mid \text{right})$

$21/28 = 3/4$

$0.75 = 0.75$ ✓

Multiplication Rule for Independent Events:

If A and B are independent events, then the probability that A and B both occur is

$$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B)$$

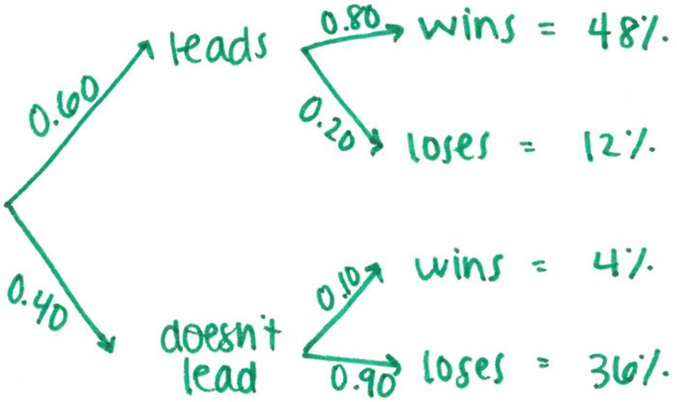
Because $P(B \mid A) = P(B)$ when A and B are independent.

Example: Sudden Infant Death

Assuming independence when it isn't true can lead to disaster. Several mothers in England were convicted of murder simply because two of their children had died in their cribs with no visible cause. An "expert witness" for the prosecution said that the probability of an unexplained crib death in a non-smoking middle-class family is 1/8500. He then multiplied 1/8500 by 1/8500 to claim that there is only a 1-in-72-million chance that two children in the same family could have died naturally. This is nonsense: it assumed that crib deaths are independent, and data suggest that they are not. Some common genetic or environmental cause, not murder, probably explains the deaths.

Example:

High school basketball team leads at half time in 60% of the games in the season. The team wins 80% of the time when they have the half time lead but only 10% when they don't.



a. What's the probability that the team wins a game during the season?

$P(\text{wins}) = 48\% + 4\% = 52\%$ + sentence

b. What's the probability that a team leads at half time given that they lose the game?

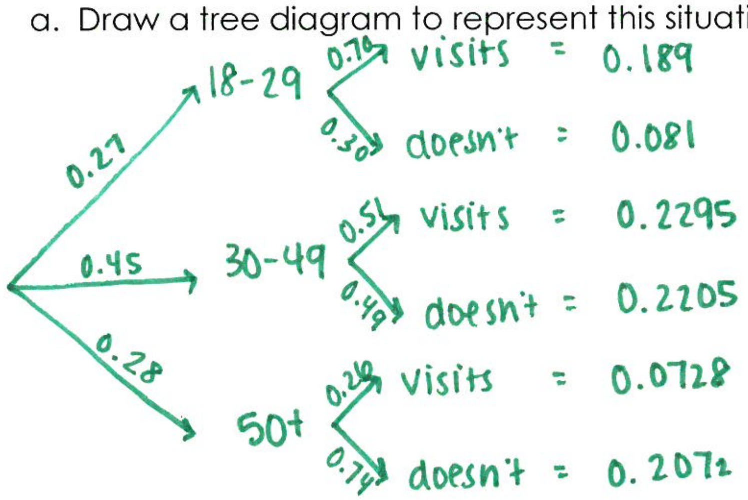
$P(\text{leads} \mid \text{lose}) = \dfrac{P(\text{leads and loses})}{P(\text{loses})} = \dfrac{0.12}{0.12+0.36} = \dfrac{0.12}{0.48} = 25\%$

+ sentence

Example:

Video-sharing sites, led by YouTube, are popular destinations on the Internet. Let's look only at adult Internet users, aged 18 and over. About 27% of adult Internet users are 18-29 years old, another 45% are 30-49 years old, and the remaining 28% are 50 and over. The Pew Internet and American Life Project finds that 70% of Internet users aged 18-29 have visited a video-sharing site, along with 51% of those aged 30-49 and 26% of those 50 or older. Do most Internet users visit YouTube and similar sites?

a. Draw a tree diagram to represent this situation.



b. Find the probability that this person has visited a video-sharing site? Show your work.

P(visits) = 0.189 + 0.2295 + 0.0728 = 0.4913

49.13% + sentence

c. Given that this person has visited a video-sharing site, find the probability that he or she is aged 18-29. Show your work.

$$P(18\text{-}29 \mid \text{visits}) = \frac{P(18\text{-}29 \text{ and visits})}{P(\text{visits})} = \frac{0.189}{0.4913} = 38.47\%$$

+sentence

d. Are being 30-49 and not visiting a video-sharing site independent of each other?

$$P(30\text{-}49 \mid \text{doesn't}) = \frac{P(30\text{-}49 \text{ and doesn't})}{P(\text{doesn't})} \stackrel{?}{=} P(30\text{-}49)$$

$$\frac{0.2205}{0.081 + 0.2205 + 0.2072} = 0.45$$

$$\frac{0.2205}{0.5087} = 0.45$$

$0.4334 \neq 0.45$ no!

the events being 30-49 years old and not visiting video sites are not independent.