SM 2

Date:

Section: 7.1

Objective: Marginal and Conditional Distributions

Are you more likely to survive a boat crash if you have a first-class ticket? Do boys and girls like different colors? These are questions that can be answered by examining ***distributions*** of data.

Individuals: The objects described by a set of data. Individuals may be people, animals, or objects.

Categorical Variable: A characteristic of an individual that places the individual into one of several groups or categories. Examples: Eye color, favorite ice cream flavor, gender, whether or not the person did their homework…

Relative Frequency: The *fraction or percent* of a group who fall into a category.

Distribution of a Categorical Variable: Lists the different categories that the individuals in the data set fall into and states how many (or what percent) of individuals fall into each category.

Two-way Table: A table that is broken down into rows and columns. The values of one categorical variable go along the rows, and the values of another categorical variable go down the columns. A two-way table helps us see if there's any relationship between the two variables.

Marginal Distribution: Ignores the inside of the table, and just gives the *percent of all the individuals in the whole sample* who fall into each category. **Look at the totals in the margins!**

Conditional Distribution: Gives the *percent of individuals in just one sub-group* who fall into each category. (For example, the percent of just the boys who like each color instead of the percent of everyone in the entire sample who likes each color.)

Example: In 1912, the *Titanic* hit an iceberg on its first voyage across the Atlantic and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table gives information about adult passengers who lived and who died, based on what type of ticket they had.

	Survival Status		
Type of Ticket	**Survived**	**Died**	**Total**
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew	212	673	885
Total	711	1490	2201

a) Give the marginal distribution of survival status. (For everyone in the entire sample, what percent survived and what percent died?)

Survived: $\frac{711}{2201} = 32.3\%$

Died: $\frac{1490}{2201} = 67.7\%$

b) Give the marginal distribution of type of ticket. What does this distribution tell you about?

First class: $\frac{325}{2201} = 14.8\%$ Third class: $\frac{706}{2201} = 32.1\%$

Second class: $\frac{285}{2201} = 12.9\%$ Crew: $\frac{885}{2201} = 40.2\%$

(Tells what percent of everyone on the entire ship had each type of ticket)

c) Give the conditional distribution of survival status for people with 1st class tickets.

survived: $\frac{203}{325} = 62.5\%$ Died: $\frac{122}{325} = 37.5\%$

(what percent of 1st-class ticket holders survived & died)

d) Give the conditional distribution of survival status for people with 2nd class tickets.

survived: $\frac{118}{285} = 41.4\%$ died: $\frac{167}{285} = 58.6\%$

(what percent of 2nd-class ticket holders survived & died)

e) Give the conditional distribution of survival status for people with 3rd class tickets.

survived: $\frac{178}{706} = 25.2\%$ Died: $\frac{528}{706} = 74.8\%$

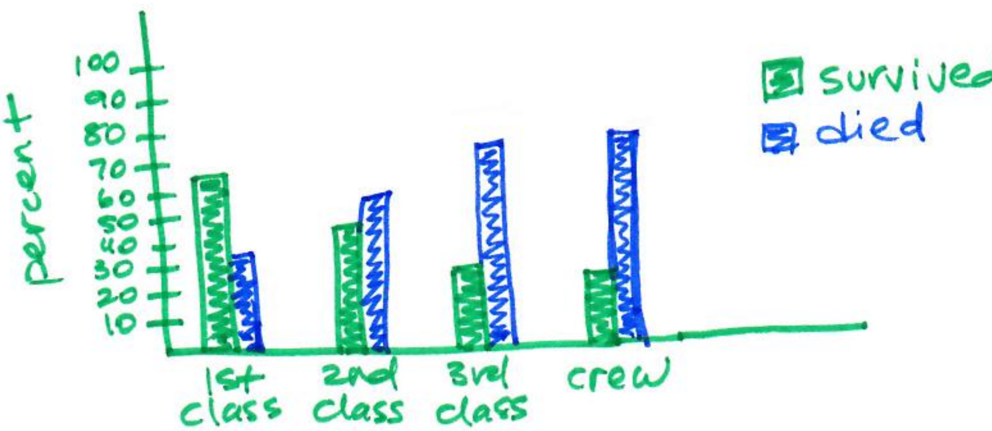
(what percent of 3rd-class ticket holders survived & died)

f) Give the conditional distribution of survival status for the crew.

survived: $\frac{212}{885} = 24.0\%$ Died: $\frac{673}{885} = 76.0\%$

(what percent of crew survived & died)

g) Draw side-by-side bar graphs to compare the distributions in parts c-f. Then write a few sentences comparing and contrasting the conditional distributions.



The conditional distributions of survival for 3rd class and crew are very similar. In both of those classes, nearly twice as many people died than survived. In second class, more died than survived, but the gap was not as extreme. In first class, over 60% survived.

h) One of your friends tries to argue that 3rd class tickets were actually better than 2nd class tickets by saying, "A higher number of 3rd class ticket holders survived the Titanic disaster than 2nd class ticket holders." Explain what is misleading about this statement.

The problem is that there were a higher number of people with 3rd class tickets than 2nd class tickets. For the comparison to be fair, you have to compare percentages. A lower percent of 3rd class ticket holders survived than 2nd class ticket holders.

i) What percent of the passengers in 1st class survived?

$\frac{203}{325} = 62.5\%$

j) What percent of the survivors were in 1st class?

$\frac{203}{711} = 28.6\%$

Take a survey in your class of favorite colors and fill in the following table:

	Favorite Color								
	Red	Orange	Yellow	Green	Blue	Purple	Pink	Other	Total
Boys	5	0	0	2	5	1	0	1	14
Girls	2	1	2	1	4	4	1	1	16
Total	7	1	2	3	9	5	1	2	30

a) Give the marginal distribution of favorite color for your class.

red: $7/30 = 23.3\%$
orange: $1/30 = 3.3\%$
yellow: $2/30 = 6.7\%$
green: $3/30 = 10\%$
blue: $9/30 = 30\%$
purple: $5/30 = 16.7\%$
pink: $1/30 = 3.3\%$
other: $2/30 = 6.7\%$

b) Give the marginal distribution of gender for your class.

Boys: $14/30 = 46.7\%$ Girls: $16/30 = 53.3\%$

c) Give the conditional distributions of favorite color for boys and girls.

Boys:
red: $5/14 = 35.7\%$
orange: $0/14 = 0\%$
yellow: $0/14 = 0\%$
green: $2/14 = 14.3\%$
blue: $5/14 = 35.7\%$
purple: $1/14 = 7.1\%$
pink: $0/14 = 0\%$
other: $1/14 = 7.1\%$

Girls:
red: $2/16 = 12.5\%$
orange: $1/16 = 6.25\%$
yellow: $2/16 = 12.5\%$
green: $1/16 = 6.25\%$
blue: $4/16 = 25\%$
purple: $4/16 = 25\%$
pink: $1/16 = 6.25\%$
other: $1/16 = 6.25\%$

d) Write a few sentences comparing and contrasting the conditional distributions of favorite color for boys and girls.

A much higher percentage of boys than girls chose red, and a much higher percentage of girls than boys chose yellow and purple. Both genders had high percentages that chose blue.

e) What percent of the girls in the class chose blue?

$\frac{4}{16} = 25\%$

f) What percent of the people who chose blue are girls?

$\frac{4}{9} = 44.4\%$

g) Was your answer to part f) part of a marginal distribution or a conditional distribution?

Conditional - asks about just the people who chose blue

h) What percent of the people in the class chose red, orange, or yellow as a favorite color?

$\frac{7+1+2}{30} = \frac{10}{30} = 33.3\%$

i) Was your answer to part h) part of a marginal distribution or a conditional distribution?

Marginal - asks about the whole class

SM 2

Date:

Section: 7.2

Objective: Venn Diagrams and Set Theory

Sample Space: The set of all possible outcomes for a chance process.

Event/Subset: An outcome or set of outcomes from the sample space.

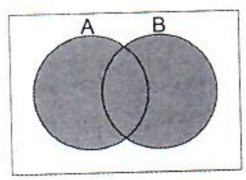
Complement $\left(A^C\right)$**:** "Not"

- All outcomes in the sample space that are not part of the event.

Chance Process	Sample Space	Event/Subset	Complement
Flip a coin	$S=\{\text{heads, tails}\}$	$B=\{\text{heads}\}$	$B^C=\{\text{tails}\}$
Roll a die	$S=\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$	even numbers $E=\{2,\ 4,\ 6\}$	$E^C=\{1,\ 3,\ 5\}$
Pick a letter in the word "probability"	$S=\{\text{P, R, O, B, A, I, L, T, Y}\}$	vowels $V=\{\text{O, A, I, Y}\}$	$V^C=\{\text{P, R, B, L, T}\}$

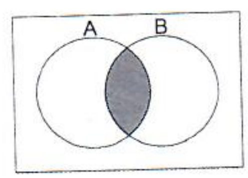
Union $(A\cup B)$**:** "Or", "Either"

- All of the elements that are in A or B or both.

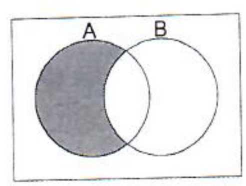


Intersection $(A\cap B)$**:** "And", "Both", "Overlap", "In common"

- All of the elements that are in *both* A and B.
- If the two sets don't have anything in common, the intersection is the "empty set", indicated by $\varnothing$ or $\{\ \}$.

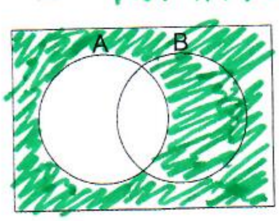


Note: If you want to write "everything in A that isn't in B," you can write either $A\cap B^C$ or $A-B$.

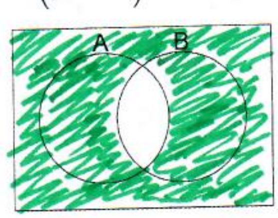


Examples: Shade the appropriate portion of the Venn diagram.

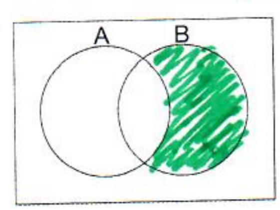
1. A^C not in A



2. $(A\cap B)^C$ not in both



3. $B-A$ In B, but not in A



Examples:

- Chance Process: Rolling a 10-sided die.
    - Event A: Rolling an odd number
    - Event B: Rolling a prime number

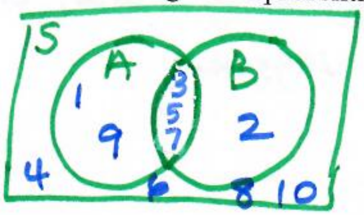
a. What is the sample space?

$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

b. List the outcomes in each event.

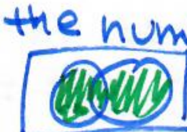
(odd) $A = \{1, 3, 5, 7, 9\}$
(prime) $B = \{2, 3, 5, 7\}$

c. Draw a Venn diagram representing the sample space with subsets A and B.



d. List all the outcomes in $A \cup B$.

$\{1, 2, 3, 5, 7, 9\}$ All the numbers that are odd or prime (or both)



e. List all the outcomes in $A \cap B$.

$\{3, 5, 7\}$ All the numbers that are both odd and prime.

f. List all the outcomes in A^C.

$\{2, 4, 6, 8, 10\}$ All the numbers that are not odd.

g. List all the outcomes in $(A \cup B)^C$.

$\{4, 6, 8, 10\}$ All the numbers that are not either odd or prime

h. List all the outcomes in $A - B$.

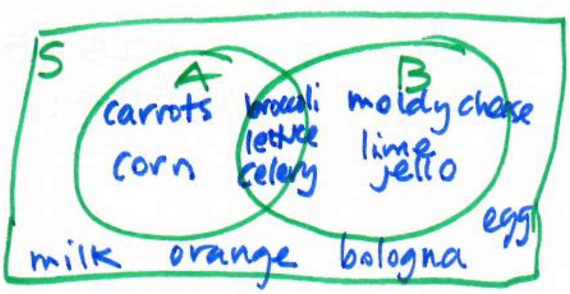
$\{1, 9\}$ All the numbers that are odd, but not prime.

- Chance Process: Reaching into a messy refrigerator and grabbing a food at random.
- Sample Space: S = {broccoli, carrots, moldy cheese, milk, orange, lettuce, lime jello, bologna, egg, corn, celery}
    - Event A: Picking a vegetable
    - Event B: Picking something green

a. List the outcomes in each event.

(veggies) A: {broccoli, carrots, lettuce, corn, celery}
(green) B: {broccoli, moldy cheese, lettuce, lime jello, celery}

b. Draw a Venn diagram representing the sample space with subsets A and B.



c. List all the outcomes in $A \cup B$. veggie or green

{broccoli, carrots, lettuce, corn, celery, moldy cheese, green jello}

d. List all the outcomes in $A \cap B$. veggie and green

{broccoli, lettuce, celery}

e. List all the outcomes in B^C. not green

{carrots, corn, milk, orange, bologna, egg}

f. List all the outcomes in $(A \cap B)^C$. not both veggie and green. (not in the overlap)

{carrots, corn, milk, orange, bologna, egg, moldy cheese, lime jello}

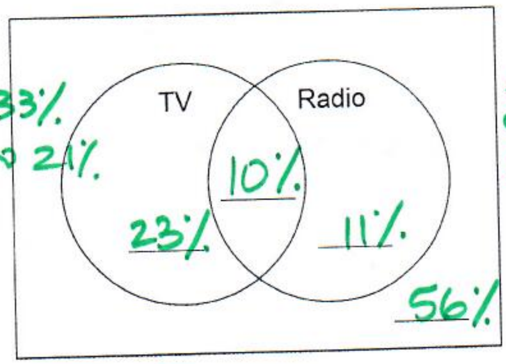
g. List all the outcomes in $B - A$. green, but not a veggie

{moldy cheese, lime jello}

Examples:

A political ad was run on TV and on radio.

- 33% of people saw it on TV. TV circle adds to 33%
- 21% heard it on the radio. Radio circle adds to 21%
- 10% of people both saw it on TV and heard it on the radio. overlap is 10%



entire diagram adds to 100%.

$23\% + 10\% + 11\% = 44\%$

$100\% - 44\% = 56\%$

Determine what percent:

a) only saw it 23%

b) only heard it 11%

c) neither heard it or saw it 56%

d) did not see it $11\% + 56\% = 67\%$

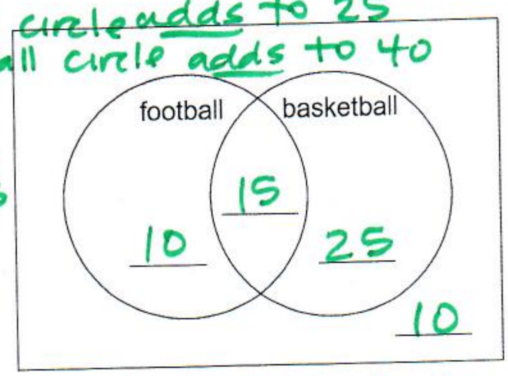
entire diagram adds to 60

A sample of 60 people are asked if they enjoy watching basketball and if they enjoy watching football.

- 25 people say they enjoy watching football football circle adds to 25
- 40 people say they enjoy watching basketball basketball circle adds to 40
- 15 people say they enjoy watching both overlap

(football) $25 - 15 = 10$

(basketball) $40 - 15 = 25$



$10 + 15 + 25 = 50$

$60 - 50 = 10$

Determine how many people:

a) enjoy football but not basketball 10

b) enjoy basketball but not football 25

c) don't enjoy either basketball or football 10

d) don't like football $25 + 10 = 35$

SM 2

Date:

Section: 7.3

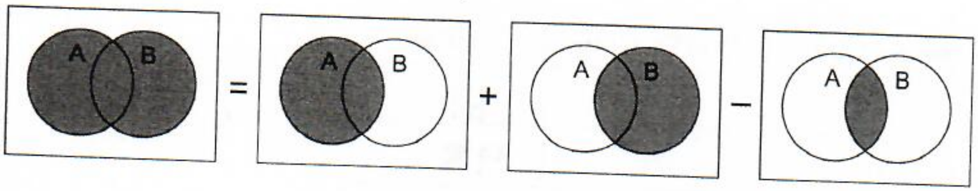
Objective: Probability from Venn Diagrams and Two-Way Tables

Probability: A value that represents the likelihood of an event. It can be expressed as a fraction, decimal, or a percentage. A probability of 0 means that the event is impossible and a probability of 1 (or 100%) means that the event is certain to occur.

$$\text{probability} = \frac{\text{total \# of favorable outcomes in the category of interest}}{\text{total \# of possible outcomes}}$$

Remember, $(A \cap B)$ means "A and B" and $(A \cup B)$ means "A or B (or both)". With "or" probabilities, makes sure you don't count the individuals who fall in both categories twice!

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



Example: In the Math Club, there are 34 students. Eleven of the students are seniors, including 7 of the 20 girls. A student is chosen at random from the club. Fill in the table and find the following probabilities:

a) $P(\text{boy})$ $\frac{14}{34} = \frac{7}{17}$

b) $P(\text{senior})$ $\frac{11}{34}$

c) $P(\text{boy} \cap \text{senior})$ (and) $\frac{4}{34} = \frac{2}{17}$

d) $P(\text{girl} \cup \text{non-senior})$ (or)

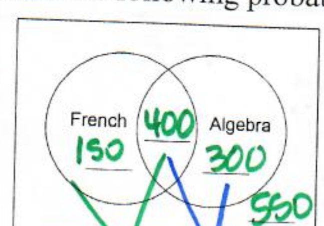
	Seniors	Non-Seniors	Total
Boys	4	10	14
Girls	7	13	20
Total	11	23	34

* Don't count the 13 non-senior girls twice!

$\frac{7+13+10}{34} = \frac{30}{34} = \frac{15}{17}$ or $P(\text{girl}) + P(\text{non-senior}) - P(\text{girl} \cap \text{non-senior})$

$\frac{20}{34} + \frac{23}{34} - \frac{13}{34} = \frac{30}{34} = \frac{15}{17}$

Example: The number of students in a high school is 1400. Of those students, 550 take French, 700 take algebra, and 400 take both French and algebra. Fill in the Venn diagram, then find the following probabilities.



adds to 550

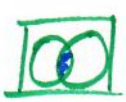
adds to 700

total = 1400

150 + 400 + 300 = 850

1400 − 850 = 550

a) $P(\text{does not take French})$ $\frac{300+550}{1400} = \frac{850}{1400} \approx 60.7\%$



b) $P(\text{algebra} \cap \text{French})$ (and) $\frac{400}{1400} \approx 28.6\%$

c) $P(\text{algebra, but not French})$ $\frac{300}{1400} \approx 21.4\%$

d) $P(\text{algebra} \cup \text{French})$ (or (or both))

$\frac{150+400+300}{1400} = \frac{850}{1400} \approx 60.7\%$

Conditional Probability: The probability of an event occurring when we already know that another event has occurred.

Example: $P(\text{lung cancer}|\text{smoke})$ would mean the probability of a person getting lung cancer given that the person smokes.

Conditional Probability Formula: $\boxed{P(A|B)=\dfrac{P(A\cap B)}{P(B)} \text{ or } \dfrac{\text{total \# in } A\cap B}{\text{total \# in } B}}$

denominator for $\cup$ or $\cap$ probabilities is overall total

★ **"And" and "or" probabilities are fractions of the entire sample, but with conditional probabilities, the condition becomes the denominator of the fraction!**

denominator for conditional is the part after the bar.

Examples:
An ice cream shop keeps track of whether people order vanilla or chocolate ice cream and whether they order a sugar cone or a waffle cone. Fill in the marginal totals and find the requested probabilities.

	Sugar Cone	Waffle Cone	Total
Vanilla	35	26	61
Chocolate	51	47	98
Total	86	73	159

a) $P(\text{vanilla})$ $\dfrac{61}{159}$

b) $P(\text{waffle})$ $\dfrac{73}{159}$

c) $P(\text{sugar})$ $\dfrac{86}{159}$

d) $P(\text{chocolate})$ $\dfrac{98}{159}$

e) $P(\text{vanilla} \cap \text{sugar})$ (and) $\dfrac{35}{159}$

f) $P(\text{vanilla} \cap \text{waffle})$ (and) $\dfrac{26}{159}$

g) $P(\text{chocolate} \cap \text{sugar})$ (and) $\dfrac{51}{159}$

h) $P(\text{chocolate} \cap \text{waffle})$ (and) $\dfrac{47}{159}$

i) $P(\text{vanilla} \cup \text{sugar})$ (or) $\dfrac{35+26+51}{159} = \dfrac{112}{159}$

j) $P(\text{vanilla} \cup \text{waffle})$ (or) $\dfrac{35+26+47}{159} = \dfrac{108}{59}$

k) $P(\text{chocolate} \cup \text{sugar})$ (or) $\dfrac{51+47+35}{159} = \dfrac{133}{159}$

l) $P(\text{chocolate} \cup \text{waffle})$ (or) $\dfrac{51+47+26}{159} = \dfrac{124}{159}$

m) $P(\text{vanilla}|\text{sugar})$ (vanilla ∩ sugar and) $\dfrac{35}{86}$

n) $P(\text{vanilla}|\text{waffle})$ $\dfrac{26}{73}$

o) $P(\text{chocolate}|\text{sugar})$ $\dfrac{51}{86}$

p) $P(\text{chocolate}|\text{waffle})$ $\dfrac{47}{73}$

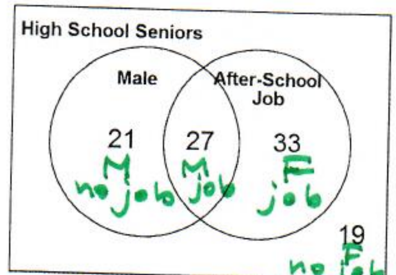
q) $P(\text{sugar}|\text{vanilla})$ $\dfrac{35}{61}$

r) $P(\text{sugar}|\text{chocolate})$ $\dfrac{51}{98}$

s) $P(\text{waffle}|\text{vanilla})$ $\dfrac{26}{61}$

t) $P(\text{waffle}|\text{chocolate})$ $\dfrac{47}{98}$

Use the Venn diagram to find the following probabilities.



a) $P(\text{job}|\text{male})$

$\frac{27}{48}$

$\frac{\text{males w/ jobs}}{\text{males}} = \frac{27}{21+27}$

b) $P(\text{female}|\text{job})$

$\frac{\text{female w/ job}}{\text{job}} \; \frac{33}{27+33} = \frac{33}{60}$

c) $P(\text{male}|\text{no job})$

$\frac{21}{21+19} = \frac{21}{40}$

d) $P(\text{no job}|\text{female})$

$\frac{19}{33+19} = \frac{19}{52}$

	job	no job	
M	27	21	48
F	33	19	52
	60	40	100

e) A student from the sample works at McTaco. What is the probability that the student is male?

has a job

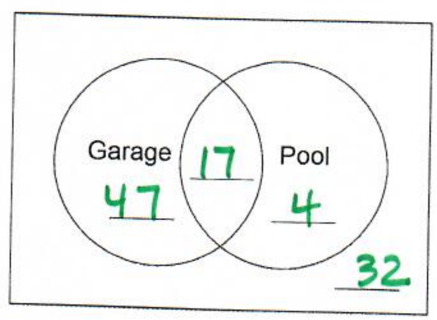
$P(\text{male}|\text{job}) = \frac{27}{60}$

f) Is a student from the sample more likely to have a job if he is a male? Justify your answer using conditional probability.

$P(\text{job}|\text{male})$ male: $\frac{27}{48}$ 56.25%   female $P(\text{job}|\text{female})$ $\frac{33}{52}$ 63.46%

No. In this sample, a female is more likely to have a job.

Real-estate ads suggest that 64% of homes have a garage, 21% have a pool, and 17% have both a garage and a pool. Fill in the Venn diagram, then answer the following questions.



a) Find $P(\text{garage} \cup \text{pool})$

$47+17+4=68$

68%

b) Find $P(\text{garage}|\text{pool})$

$\frac{17}{21}$ 81%

c) Find $P(\text{pool}|\text{garage})$

$\frac{17}{64}$ 26.6%

d) Find $P(\text{pool}|\text{no garage})$

$\frac{4}{36}$ 11.1%

e) Find $P(\text{no pool}|\text{garage})$

$\frac{47}{64} = 73.4\%$

f) Find $P(\text{no garage}|\text{no pool})$

$\frac{32}{79} = 40.5\%$