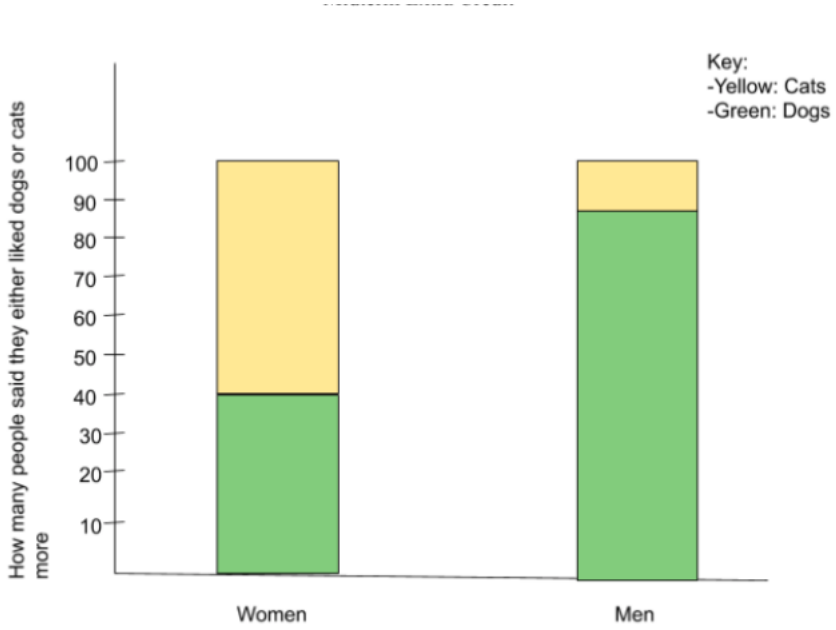
Summer 2019
Exam 1 (Student Practice Problems)



1. Use the plot to answer the following questions.
   (a) What is the probability that a woman likes cats more than dogs?
   Answer: 60/200=0.3
   (b) What is the probability that someone is a man and prefers dogs?
   Answer: 0.45. Being a man is one probability (1/2 because only two genders are shown) and liking dogs is another one (90/100);
   $P(\text{man and dogs}) = P(\text{dogs} \mid \text{man}) \times P(\text{man})$
   (c) What is the probability of you being a woman and liking both cats and dogs equally?
   Answer: 0 or 0%. While it may be true in your case or a friends case, it is not accounted for in this stacked barplot; therefore, we assume that there are not any and the probability would be zero.

2. Suppose that someone is trying to see if a town should make more gyms available to its people. The person conducting the study decides to randomly select a couple of people walking down the neighborhood of a busy street in the town and asks each of them the average amount of hours they work out in a week. What type of study would this be, and would we be able to establish a causal relationship based on the findings?
   Answer: This would be considered an observational study since the researcher is not implementing a treatment on the randomly selected people. Their answers are not being influenced by the researcher in any way. We would not be able to establish a causal relationship because this is an observational study.

3. You have six bingo balls in a hat numbered 1, 2, 3, 4, 5, and 6. What is the probability that you draw the 1, 3, and 5 (without replacement)?
   Answer: $1/6 \times 1/5 \times 1/4 = 1/120$

4. Polydactyly is a symptom of Ellis-van Creveld syndrome in which an individual has extra fingers or toes. The occurrence of this is higher in populations with low genetic diversity, such as the Amish community in which present in 7% of the population, compared to the US national average of 2.8%.

   Lancaster, PA has an Amish community consisting of 74,250 people. What is the expected occurrence of Polydactyly in this population? (Round up if its a decimal)

   A. 364

   B. 5198

   C. 2079

   D. 19

   Answer: B. $E(X) = 74250(0.07) = 5197.50 \rightarrow 5198$ individuals

5. There is a small aquarium filled with 52 small fish. The probability of getting a fish without spots given that it was black is 30.4%. The tank has 23 black fishes. What is the probability of getting a black and spotted fish from the tank?

   Answer:

$$\begin{aligned} P(\text{spotted given that its black}) &= 1 - P(\text{not spotted given that its black}) \\ &= 1 - 0.304 = 0.696 \\ P(\text{black fish}) &= 23/52 = 0.4423 \\ P(\text{black and spotted}) &= P(\text{spotted given that its black}) \times P(\text{black}) \\ &= 0.696 \times 0.4423 = 0.3078 \end{aligned}$$

   So there is about 31% chance of getting a black and spotted fish from the aquarium tank.

6. A shoe company randomly sampled 500 customers asking what their favorite brand of shoe was and their preference for shoes with or without laces. The following data was obtained

	Nike	Adidas	Vans	Total
Shoelaces	210	50	60	320
No shoelaces	80	60	40	180
Total	290	110	100	500

   What is the probability that a customers favorite shoe is Nike without shoelaces?

   A. 0.360

   B. 0.420

   C. 0.580

   D. 0.160

   E. 0.940

   Answer: D. $80/500 = 0.16$

7. Determine if the statements below are true or false. If false, what is the correct response?

   (a) If I have a disc that is split in half with an option with an option mentioning I study at the library or study at the tutoring center. I spin the arrow 8 times and all 8 times the arrow lands on studying at the library. This must mean the probability of studying in the library is higher than studying in the tutoring center.
   Answer: FALSE. The arrow on the disc only has two options, each has a 1/2 chance no matter the outcome, so even if he spins it and it lands 8 times on the library it doesn't mean the probability is any less or higher. The probability will be the same. It is an independent trial.

   (b) I have a set of values, 1,2,4,4,6,7,8,8,5,3,2,4,6,7. If I wanted to find the mean then the result would be $\bar{x} = 5.98$.
   Answer: FALSE.

$$\frac{1+2+4+4+6+7+8+8+5+3+2+4+6+7}{14} = 4.783$$

   (c) 60% of Americans love the way Basketball is currently being played, we randomly sample 2 people that love this sport and where its going, we have the x representing the number of people that love this game and below is their probability. The calculated expected value was 1.2 and the variance is .50, while the Standard deviation is .69.

X	0	1	2
p(x)	0.16	0.48	0.36

   Answer: FALSE.

$$E(X) = (0 * 0.16) + (1 * 0.48) + (2 * 0.36) = 1.2$$

   and

$$Var(X) = 0.16 \times (0 - 1.2)^2 + 0.48 \times (1 - 1.2)^2 + 0.36 \times (2 - 1.2)^2 = 0.484$$

   (d) As I am watching the world cup (USA), I begin to immediately take down the number of goals each player makes to see there stats and confirm which player had the higher goal rate, what I am doing is called a retrospective study.
   Answer: FALSE. This is a prospective study.

   (e) It has been discovered, through research I had conducted with apple juice and regular water with color dye to make it look like the apple juice, acting as the placebo of the experiment, that if you drink apple juice during the afternoon if helps alleviate any pains within the stomach. I had 60 volunteers and 30 of them drank the apple juice and the other 30 drank what they also thought was apple juice. My explanatory variable is the amount of juice I poured in each cup and the response variable was if their stomach pain eased at noon.
   Answer: FALSE. The explanatory variable is the apple juice and the placebo which is water with dye that half of the volunteers drank while the other half drank the placebo. It has nothing to do with the amount, which was not even mentioned. The response variable is whether or not pain goes away at noon when drinking apple juice.

8. At Starbucks about 25% of the employees make below \$30,000, 50% make below \$40,000, 75% make below \$60,000, and only a few high executives make \$90,000. Determine whether the distribution is right skewed, left skewed, or symmetric. Specify whether the mean or median would best represent an observation in the data. Finally, determine if the variability of observation would be best represented using the standard deviation or IQR.
Answer: This distribution is symmetric (nobody makes above \$90k). Because it is symmetric, we want to use the mean and standard deviation.

9. Suppose you have 300 birds within a zoo. There are 85 male birds with green eyes and 75 female birds with green eyes. There are also 20 male birds with blue eyes and 120 female birds with blue eyes.
   (a) Draw a contingency table.
   Answer:

	Green	Blue	Total
Male	85	20	105
Female	75	120	195
Total	160	140	300

   (b) What is the probability that a bird is male given that the bird has green eyes?
   Answer: $\frac{85}{160} = 0.53$
   (c) What is the probability that a bird has blue eyes given that the bird is male?
   Answer: $\frac{25}{105} = 0.1904$
   (d) What is the probability that the bird is female?
   Answer: $\frac{195}{300} = 0.65$
   (e) What is the probability that the bird has green eyes given that the bird is female?
   Answer: $\frac{75}{195} = 0.3846$

10. Which calculation(s) will signs (+ or -) in linear combinations affect?
    A. expected value
    B. variance
    C. expected value and variance
    D. standard deviation

    Answer: A.

11. At a beauty school, 15% of students dont wear makeup themselves. After taking a random sample of 100 students, how many students can be expected to not wear makeup?
Answer: 100(0.15)= 15 students are expected to not wear makeup.

12. Describe when we would use IQR or Standard Deviation to describe a variable and why.
Answer: We would use IQR when the data is right or left skewed. As for standard deviation, we would use that to describe data that is more symmetrical. This is because standard deviation is not as accurate describing skewed data due to standard deviation being sensitive to such skews. IQR on the other hand, is not affected by outliers as much and would then be used to describe skewed data.

13. For two events, A and B, $P(A) = 0.5$ and $P(B) = 0.2$, If A and B are independent, what is $P(A \text{ or } B)$?

    A. 0.4

    B. 0.07

    C. 1.1

    D. 0.6

    Answer: Since A and B are independent,

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ &= P(A) + P(B) - P(A)(B) \\ &= 0.5 + 0.2 - 0.1 \\ &= 0.6 \end{aligned}$$

14. Suppose you wanted to conduct an experiment to see if people prefer dogs or cats as pets. Which sample group would be the most accurate representation for the population of college students as a whole.

    A. 100 people selected from the UCR gym

    B. 100 people selected from the College of Natural & Agricultural Sciences

    C. 100 people selected randomly on the UCR campus

    D. 100 people selected from Bourns College of Engineering

    Answer: C. The student sample selected is the most general overall and not pertaining to a specific demographic.

15. I want to find out the association between weight and how quickly a car will travel one mile. I weigh 10 cars and then time them over one mile. Time over one mile is the

    A. Explanatory Variable

    B. Control Group

    C. Response Variable

    D. Independent Variable

    E. None of the Above

    Answer: C.

16. In a swimming competition, 66% of the participants were boys, 31% of the participants were from out of state schools, and 14% were boys from out of state schools. What is the probability that a randomly selected participant is a boy or from an out of state school?
    Answer: $0.66 + 0.31 - 0.14 = 0.83$

17. The students in STAT 100A scored the following midterm results:

$$98, 93, 85, 89, 90, 77, 88, 75, 71, 83, 90, 68, 95, 92, 80$$

Calculate Q1, Q3, Median, and IQR.
Hint: Order the test results from least to greatest.

A. Q1 = 89, Q3 = 95, Median = 88, IQR = 6
B. Q1 = 77, Q3 = 92, Median = 75, IQR = 15
C. Q1 = 77, Q3 = 92, Median = 88, IQR = 15
D. Q1 = 89, Q3 = 95, Median = 75, IQR = 6

Answer: C

18. On a given day, 44.36% of students at UCR drive to school. Of those students, 86.82% made it to class on time. Calculate $P(\text{on time to class} = \text{yes} \mid \text{drive to school} = \text{yes})$.
Answer:

$$\begin{aligned} P(\text{on time and drive}) &= P(\text{on time} \mid \text{drive}) \times P(\text{drive}) \\ &= 0.8682 \times 0.4436 \\ &= 0.3851 \end{aligned}$$

$$\begin{aligned} P(\text{drive and late}) &= P(\text{late} \mid \text{drive}) \times P(\text{drive}) \\ &= (1 - 0.8682) \times 0.4436 \\ &= 0.0585 \end{aligned}$$

So

$$\begin{aligned} P(\text{on time} \mid \text{drive}) &= \frac{P(\text{on time} \mid \text{drive}) \times P(\text{drive})}{P(\text{on time} \mid \text{drive}) \times P(\text{drive}) + P(\text{late} \mid \text{drive}) \times P(\text{drive})} \\ &= \frac{0.3851}{0.3851 + 0.0585} \\ &= 0.8681 \end{aligned}$$

19. Suppose researchers want to study a new flea medication for Siberian huskies. Researchers selected 100 Siberian huskies to test trail, the vet and owners are not told if the pet has received the medication or not. 50 received the treatment and 50 received a placebo.

(a) How many individuals were surveyed?
Answer: 100 Siberian huskies

(b) How was bias reduced in this study?
Answer: It's double blind, so none of the dogs/their owners or the vets know whether each dog is receiving treatment or placebo.

(c) What are the control and treatment groups?
Answer: The control group is made up of the 50 huskies receiving the placebo and the treatment is made up of the 50 huskies receiving the flea medication.

20. An ice cream place sampled 500 randomly selected customers and obtained the following table:

	Strawberry	Chocolate	Vanilla	Total
Syrup	420	100	120	640
No Syrup	160	120	80	360
Total	580	220	200	1000

What is the probability that a customer orders a strawberry flavor with no syrup?
Answer: P(strawberry and no syrup) = 160/1000 = 0.160

21. Select which statement(s) would indicate a lack of understanding of disjoint events and independence.

A. When rolling 2 6-sided dice, the roll of the first die has no effect on the roll of the second die. Therefore, each dice roll is considered independent

B. In independent events, if something is known about an event A, it provides no information about another event B

C. In independent events, if something is known about an event A, it provides information about another event B

D. In disjoint events, if an event A occurs we cannot be certain that an event B did not occur

E. In disjoint events, if an event A occurs we can be certain that an event B did not occur

Answer: C and D

22. A survey was sent out to students who attend the University of California Riverside. The question asked is On average how many hours total do you study a week? The survey randomly sampled 2564 students from all years. The average amount of time students spend studying was about 23 hours per week. Using this information answer the following questions.

(a) Which of the following is a case?

A. A student from UCR

B. The number of hours spent studying per week

C. 23 hours per week

D. The average number of hours students spend studying per week

Answer: A

(b) Which of the following is a variable?

A. A student from UCR

B. The number of hours spent studying per week

C. 23 hours per week

D. The average number of hours students spend studying per week

Answer: B

(c) Which of the following is a sample statistic?

A. A student from UCR

B. The number of hours spent studying per week

C. 23 hours per week

D. The average number of hours students spend studying per week

Answer: D. This will be the average calculated based on the sample.

23. Answer the following questions based on a standard 52 card deck.

(a) What is the probability that you draw a face card that is red?
Answer: $6/52$

(b) What is the probability of drawing a face card and then a 3 card?
Answer: $12/52 \times 4/51 = 0.0181$

(c) What is the probability of drawing a king of spades at random?
Answer: $1/52$

24. Jessica has a bag of marbles that contain 5 red marbles, 6 blue marbles, and 9 green marbles (20 marbles total). Assume she draws 3 marbles without replacement.

(a) If Jessica draws from the bag three times what is the probability that she chooses first red, then blue, and then a green marble?
Answer: $5/20 \times 6/19 \times 9/18$

(b) What is the probability of selecting just red and blue marbles?
Answer: $11/20 \times 10/20 \times 9/20$

(c) What is the probability of selecting at least one green marble?
Answer: $1 - P(\text{no green}) = 1 - 11/20 \times 10/19 \times 9/18$

(d) What is the probability of selecting 2 blue marbles and then a marble that is not blue?
Answer: $6/20 \times 5/19 \times 14/18$

25. Let A and B be independent events such that $P(A) = 0.5$ and $P(B) = 0.9$.

(a) What is $P(A \text{ and } B)$?
Answer: $P(A \text{ and } B) = P(A) \times (B) = 0.45$

(b) What is $P(A \text{ or } B)$?
Answer: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.9 + 0.5 - 0.45$

(c) What is $P(A|B)$?
Answer: $P(A|B) = P(A \text{ and } B)/P(B) = 0.45/0.9 = 0.5$.
Notice that, since A and B are independent, $P(A|B) = P(A)$.

26. Suppose you want to calculate the percentage of dog videos that are on Instagram. It is not possible for anyone to watch all dog videos on Instagram, so a random video picker is being used to select 2000 videos. You find that 4% of these videos are dog videos. Determine which of the following situations are a variable, an observation, a sample statistic, or a population parameter.

    (a) Percentage of all videos on Instagram that are dog videos.

        A. variable
        B. observation
        C. sample statistic
        D. population parameter

        Answer: D. Since in this situation we are talking about the whole population of dog videos in which they have to study, this is a population parameter.

    (b) 4% of total videos are dog videos

        A. variable
        B. observation
        C. sample statistic
        D. population parameter

        Answer: B. In this situation the 4% only represents the dog videos which are picked using the random video picker, hence it represents the sample of the total videos. Therefore, it is a sample statistic.

27. Suppose you want to select 4 balls from a basket containing 24 balls. What is the probability that a particular ball is not picked?
    Answer: There are 24 balls in the basket. For the first ball, you have a 23/24 chance of selecting the ball of interest. For the second ball, you have a 22/23 chance of not picking it. For the third ball you have a 21/22 chance of not picking it. For the fourth ball, you have a 20/21 chance of not picking it.

$$\frac{23}{24} \times \frac{22}{23} \times \frac{21}{22} \times \frac{20}{21} = 0.8333$$

28. At a middle school 48% of the students have access to internet at home. The rest only have access at school. A group of 10 students are chosen at random. Find the probability that

    (a) a student only uses the internet at school.
        Answer: P(internet only at school) = 1-P(internet at home)$= 1 - 0.48 = 0.52$

    (b) at least 1 student has access to the internet at home.
        Answer: P(at least one has home internet) = 1-P(none) $= 1 - 0.52^{10} = 0.999$

29. There is a new teaching assistant in Mrs. Bedfords class. The TA planned a small game of probability in order to bond with the students. The TA brought a sack with 50 different colored beanies: 25 red, 7 blue, 8 pinks, and 10 whites. The game is that 4 random students will be chosen and each will pick a beanie blindly. The TA will then guess what color each student actually got. (Since 50% of the beanies are red, the TA will always guess red.)

    (a) What is the probability of the TA guessing the correct color for the first student who picks out a beanie?
    Answer: P(correct guess)= .5

    (b) What is the probability of the first two students getting red beanies and the 3rd student picking a pink beanie?
    Answer: P(red then red then pink) $= \frac{25}{50} \times \frac{24}{49} \times \frac{8}{48}$

    (c) Say that the TA really wanted to impress the students, and actually placed 35 red beanies, 10 blue beanies, 3 pinks, and 2 whites in order to have a better chance at guessing that the students picked a red beanie. What is the probability of all 4 students NOT getting a red beanie and the TA guessing incorrectly?
    Answer: P(all students not getting red)

$$\frac{15}{50} \times \frac{14}{49} \times \frac{13}{48} \times \frac{12}{47}$$

30. After your stay at a hotel, you are asked to fill out a brief questionnaire. You are asked the following questions:

    1. What is your gender?
        (a) Male
        (b) Female
    2. How would you rate the service?
        (a) Excellent
        (b) Very Good
        (c) Good
        (d) Poor
    3. What is your annual household income?
        (a) 15,000 - 30,000 USD
        (b) 30,000 - 45,000 USD
        (c) 45,000 - 60,000 USD
        (d) 60,000+ USD

    Label these questions according to their interval, nominal, or ordinal scale.
    Answer: (1) Gender is of nominal scale. (2) Quality of service is of ordinal scale. (3) Income is of interval scale. All three are categorical variables.

31. You roll fair a 6-sided die three times in a row, betting your friend 10 dollars each time that it will land between 4 to 6. You've lost the bet three times in a row, what is the probability that you lose the fourth time?

    A. 25%

    B. 50%

    C. 66%

    D. 42.5%

    Answer: B. Each roll is independent of the rest.

32. For the second time, a band is offering its website's members its new album for download at a discounted rate using their "Pay whatever, fam" model. Members can pay \$1, \$2, \$3, or \$0 for the album before it's released to the wider public at an increased price. For the band's previous discounted album sale, 27.34% of its members paid \$1, 23.57% paid \$2, 15% paid \$3, and 34% paid \$0. This time, the band expects to sell 700 albums. If the proportions from the first sale remain the same for the second sale, by how much should the band expect a purchase price to vary?
    Answer:

$$E(X) = (3)(0.15) + (2)(0.2357) + (1)(0.2743) + (0)(0.34) = \$1.1957$$

$$\begin{aligned} Var(X) &= 0.15 \times (3 - 1.1957)^2 + 0.2357 \times (2 - 1.1975)^2 + 0.2734 \times (1 - 1.1957)^2 + 0.34 \times (0 - 1.1957)^2 \\ &= 1.136685 \end{aligned}$$

and the standard deviation will be

$$SD(X) = \sqrt{Var(X)} = \sqrt{1.136685} = 1.0661$$