

Run L^AT_EX again to produce the table

For full credit you must SHOW ALL YOUR WORK. Please submit your solutions as a pdf file on Blackboard by Monday night.

1. (10 points) Suppose you are a teacher with 10 children in your class; 6 are boys and 4 are girls. You are going to *randomly* choose 3 of the 10 children to work together on a group project.

Calculate each of the following. You can use a spreadsheet to carry out and/or check your calculations, but you should **write out your calculations in detail**:

- (a) How many different such groups of 3 children are there? (Hint: this is a combinations calculation.)

Solution:

$$\binom{10}{3} = \frac{10 * 9 * 8}{3 * 2 * 1} = \frac{720}{6} = 120$$

- (b) How many different groups of 3 consist of *all boys*? Recall that there are 6 boys in the class.

(Hint: We can rephrase this question as, how many different ways are there of choosing 3 boys from the 6 boys in the class.)

Solution:

$$\binom{6}{3} = \frac{6 * 5 * 4}{3 * 2 * 1} = \frac{120}{6} = 20$$

- (c) Use your answers to parts (a) and (b) to calculate the *probability* that you will randomly choose a group with all boys. Express your answer first as a ratio of two integers, and then as a decimal (rounded to 3 decimal places).

Solution:

$$P(\text{all boys}) = \frac{\text{number of groups consisting of all boys}}{\text{total number of groups}} = \frac{20}{120} = \frac{1}{6} = 0.167$$

- (d) Use your answer to part (c) to calculate the probability that you will randomly select a group with *at least one girl* in it:

Solution:

$$P(\text{at least one girl}) = 1 - P(\text{all boys}) = 1 - \frac{1}{6} = \frac{5}{6} = 0.833$$

2. (10 points) Consider the random variable:

X = how long you wait for the subway after getting to the platform (rounded down to the nearest minute)

(a) Suppose you have collected data for X , summarized in the frequency table below. Fill in the table by computing the relative frequencies:

Solution:

X	Frequency, f	Relative frequency
0	1	$1/20 = 0.05$
1	3	$3/20 = 0.15$
2	5	$5/20 = 0.25$
3	4	$4/20 = 0.20$
4	2	$2/20 = 0.10$
5	2	$2/20 = 0.10$
6	0	$0/20 = 0$
7	1	$1/20 = 0.05$
8	0	$0/20 = 0$
≥ 9	2	$2/20 = 0.10$

(b) Sketch a graph of the relative frequency distribution (i.e., the relative frequency histogram). **Clearly label the horizontal and vertical axes.**

Solution:

(c) We can interpret the relative frequency distribution as a probability distribution. Use it to compute the following probabilities:

(i) What is the probability that you wait less than 5 minutes?

$$P(X < 5) =$$

Solution: $P(X < 5) = 0.05 + 0.15 + 0.25 + 0.20 + 0.10 = 0.75$

(ii) What is the probability that you wait between 5 and 9 minutes?

$$P(5 \leq X < 9) =$$

Solution: $P(5 \leq X < 9) = 0.1 + 0 + 0.05 + 0 = 0.15$

(iii) What is the probability that you wait more than 9 minutes?

$$P(X \geq 9) =$$

Solution: $P(X \geq 9) = 0.1$

3. (10 points) A bag contains 7 red marbles and 3 blue marbles. Five marbles are drawn from the bag *with replacement*. Consider the random variables:

R = the number of red marbles selected
 B = the number of blue marbles selected

- (a) What are the possible values of each of the following discrete random variables?

Solution: $R : \{0, 1, 2, 3, 4, 5\}, B : \{0, 1, 2, 3, 4, 5\}, R + B : \{5\}$

- (b) Explain why R is a binomial random variable with $n = 5$ and $p = 0.7$. (In particular, explain why it is essential that the marbles are drawn *with replacement*.)

Solution: Each of the 5 draws has two outcomes (red or blue). Since the marbles are drawn with replacement, the result of each draw is independent of every other draw, and the probability p of drawing a red marble remains constant. If the marbles were drawn without replacement, the probability of drawing a red marble does not remain constant.

- (c) Use the binomial distribution formula to down the probability distribution of R . You can use `=binomdist(i,n,p,false)` to check your answers, but show all of your calculations in the table below (as shown in the $i = 2$ row of the table):

Solution:

i	$P(R = i)$			
0	$C(5, 0) = 1$			
1				
2	$C(5, 2) = \frac{5 \cdot 4}{2 \cdot 1} = 10$	$p^2 = (0.7)^2 = 0.49$	$q^{5-i} = (0.3)^{5-2} = (0.3)^3 = 0.027$	$P(R = 2) = 10 * 0.49 * 0.027 = 0.1323$
3				
4				
5				

- (d) Calculate the expected value, variance and standard deviation of R , using the given formulas for the expected value, variance and standard deviation of a binomial random variable, in terms of n, p , and q (where $q = 1 - p$).

Solution:

$E[R] = np = 5(0.7) = 3.5$
 or we can use the probability distribution:
 $E[R] = 0 * 0.00243 + 1 * 0.02835 + 2 * 0.13230 + 3 * 0.30870 + 4 * 0.36015 + 5 * 0.16807 = 0 + 0.02835 + 0.2646 + 0.9261 + 1.4406 + 0.84035 = 3.5$

$\text{Var}(R) = npq = 5(0.7)(0.3) = 1.05$

$\text{SD}(R) = \sqrt{npq} = \sqrt{1.05} \approx 1.025$

4. (10 points) Advanced Placement exams are scored on a 5-point scale, where 1 is the lowest possible score and 5 is the highest possible. Two students Alice and Bob are preparing to take the statistics AP exam. Alice and Bob come up with subjective probability distributions to represent how they think they will do on the exam.

Alice's subjective probability distribution for her exam score A :

$i:$	1	2	3	4	5
$P(A = i):$	0.05	?	0.15	0.4	0.3

Bob's subjective probability distribution for his exam score B :

$i:$	1	2	3	4	5
$P(B = i):$?	0.45	0.3	0.15	0.05

- (a) Calculate the missing values in the distributions:

(i) $P(A = 2) =$

Solution: $P(A = 2) = 1 - (0.05 + 0.15 + 0.4 + 0.3) = 1 - 0.9 = 0.1$

(ii) $P(B = 1) =$

Solution: $P(B = 1) = 1 - (0.45 + 0.3 + 0.15 + 0.05) = 1 - 0.95 = 0.05$

- (b) According to Alice's probability distribution, what is the probability that she will get a score of 3 or higher? What about for Bob?

$P(A \geq 3) =$

$P(B \geq 3) =$

Solution:

$P(A \geq 3) = 0.15 + 0.4 + 0.3 = 0.85$

$P(B \geq 3) = 0.3 + 0.15 + 0.05 = 0.5$

- (c) Find the expected value of each probability distribution. Show your calculations:

(i) $E[A] =$

Solution: $E[A] = 1 * (0.05) + 2 * (0.1) + 3 * (0.15) + 4 * (0.4) + 5 * (0.3) = 0.05 + 0.2 + 0.45 + 1.6 + 1.5 = 3.8$

(ii) $E[B] =$

Solution: $E[B] = 1 * (0.05) + 2 * (0.45) + 3 * (0.3) + 4 * (0.15) + 5 * (0.05) = 0.05 + 0.9 + 0.9 + 0.6 + 0.25 = 2.7$

5. (10 points) Airlines typically overbook flights (i.e., sell more tickets than there are seats on the plane) in order to maximize revenue. For example, suppose an airline sells 240 tickets for a flight on a plane with 200 seats. The airline estimates that each customer who has bought a ticket has an 80% probability of showing up for the flight.

- (a) How might the airline come up with the estimate of 80% for the probability that a given customer will show up for the flight?

Solution: Based on data from previous flights

- (b) Let the random variable X represent the number of passengers who actually show up for the flight. Explain why this can be interpreted as a binomial random variable with $n = 240$ and $p = 0.8$.

Solution: two outcomes, constant probability, independent trials (each passenger)

- (c) Assuming X is a binomial random variable with $n = 240$ and $p = 0.8$, what is the number of passengers that will show up for the flight on average; i.e., what is the expected value of X ?

Solution:

$$E(X) = np = 240(0.8) = 192$$

- (d) What is the probability that more than 200 passengers will show up for the flight? I.e., what is $P(X > 200)$? Calculate this probability in two ways:

- (i) Create a spreadsheet with the entire binomial probability distribution for X (you can use `=binomdist(i,n,p,false)` for this, i.e., you don't need to use the binomial distribution formula). Then within your spreadsheet calculate the sum $P(X = 201) + P(X = 202) + P(X = 203) + \dots + P(X = 239) + P(X = 240)$.
- (ii) Use the spreadsheet function `=binomdist(i,n,p,true)`. (Recall that `=binomdist(i,n,p,true)` returns the cumulative probability $P(X \leq i)$)

Highlight the two cells in your spreadsheet containing the formulas calculating $P(X > 200)$ by these two different methods, and submit the link to your spreadsheet along with your written solutions. Also write the probability below:

$$P(X > 200) =$$

Solution: `=1 - binomdist(200, 240, 0.8, true) = 0.0826`

- (e) Why would the airline be particularly interested in $P(X > 200)$? Recall that they have overbooked the flight and that the plane has 200 seats!