1. Consider an experiment of rolling two regular fair six-sided dice. One of the two dice is red colour and the other is black.
   a) List out all the possible outcomes of the above experiment. **[1 mark]**
   Note: You are not expected to draw any tree diagram here, but it would be nice if you were able to do so.

(1, 1)	**(1, 2)**	**(1, 3)**	**(1, 4)**	**(1, 5)**	**(1, 6)**
(2, 1)	**(2, 2)**	**(2, 3)**	**(2, 4)**	**(2, 5)**	**(2, 6)**
(3, 1)	**(3, 2)**	**(3, 3)**	**(3, 4)**	**(3, 5)**	**(3, 6)**
(4, 1)	**(4, 2)**	**(4, 3)**	**(4, 4)**	**(4, 5)**	**(4, 6)**
(5, 1)	**(5, 2)**	**(5, 3)**	**(5, 4)**	**(5, 5)**	**(5, 6)**
(6, 1)	**(6, 2)**	**(6, 3)**	**(6, 4)**	**(6, 5)**	**(6, 6)**

   b) Define X as the "Red minus Black". Construct a probability distribution table of X. **[3 marks]**

$\boldsymbol{X}$	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**
$\boldsymbol{f(x)}$	**1/36**	**2/36**	**3/36**	**4/36**	**5/36**	**6/36**	**5/36**	**4/36**	**3/36**	**2/36**	**1/36**

   c) Calculate $E(X)$ and provide an interpretation of it. **[1+2 marks]**
   E(X) = (-5)*(1/36) + (-4)*(2/36) + … (4)*(2/36) + (5)*(1/36) = 0 [1]
   Interpretation: When the experiment of "rolling two distinct dice" is repeated infinite number of times, the average difference between the Red die and Black die is 0. [2]

   d) Calculate $SD(X)$ and provide an interpretation of it. **[1+2 marks]**
   Note: Feel free to use either Excel or R to check the value. But make sure you are able to do the calculations if needed.
   $\boldsymbol{SD(X) = 2.4152}$ **(using Excel) [1]**
   Interpretation: When the experiment of "rolling two dice and recording the difference" is repeated infinite number of times, the typical difference from the average <u>difference between the Red die and Black die</u> is about 2.4. [2]

2. Consider the game of American Roulette. There are total of 38 slots where the ball can end up in when the roulette wheel stops: 36 winnable slots (number 1 to 36) and 2 non-winnable slots (0 and 00). In all American Roulette games, when the ball falls in the 0 or 00 slot, the house takes your money or you lose your bet.

a) <u>In the next three parts (a to c)</u>, we focus on <u>betting $1 on even numbers</u> and <u>define the random variable X as the winning amount</u> (from the gambler's point of view). If the ball falls in any of the even number slots (like 2, 4, 6, …, 34, and 36), you win $1 (or have a winning amount is +$1, or positive $1). Otherwise, you lose the $1 bet (or have a winning amount of -$1, or negative $1). Construct the probability distribution table of the random variable X. **[3 marks]**

Outcome	Win	Lose
Winning Amount or X (in $)	**+1**	**-1**
Probability or $f(x)$	**18/38**	**20/38**

b) Calculate the expected value of X. **[1 mark]**
$\boldsymbol{E(X) = (+1) \times \left(\frac{18}{38}\right) + (-1) \times \left(\frac{20}{38}\right) = \frac{-2}{38}}$ **or -0.053**

c) Provide an interpretation of the expected value from the gambler's point of view. Briefly explain to a casual gambler why they will always lose money in the long run. **[2+1 marks]**
Interpretation: When the experiment/game of "betting on even numbers in an American Roulette game" is repeated infinite number of times, the average winning amount (in $) is negative $0.053. [2]
In layperson's term, a gambler loses for about 5 cents per game of "betting on even numbers in an American Roulette game", in the long run. [1]

d) Suppose you work at the casino and want to create a brand-new game of "betting on the multiples of 4". What would be the reasonable payout so that the casino will have an edge on the gamblers in a long run. Please make sure you show some calculation to justify your answer. **[2+2 marks]**
Note: Feel free to take a look at this link, do some E(X) calculations for a few of those bets to fully understand how it works. Did yon notice any commonality of the E(X) of all bets (except the Sucker Bet or Five Bet)?
Source: https://www.casinonewsdaily.com/roulette-guide/american-roulette/
It's basically the same type of calculation of E(X) as above except it's backward.

$$\boldsymbol{E(X) = (+V) \times \left(\frac{9}{38}\right) + (-1) \times \left(\frac{29}{38}\right) = -0.053}$$

Solving for V gives 3. In other words, the payout is "3 to 1".

3. A Randstad/Harris interactive survey reported that 40% of employees said they are loyal to their company. Suppose 15 employees are selected randomly from a company of N employees and they are asked about their loyalty to their company. Assuming that there are no issues with any biases.
   a) Explain the meaning of "independent trials", in the context of this question. **[2 marks]**
      Whether one employee says they are loyal to the company does not affect if another employee says that they are loyal to their company, or vice versa.

   b) If there are only N = 20 employees in the company (and all employees are working on site), do you think the "trials" are still independent to each other? Briefly justify your answer. **[0+2 marks]**
      No, the trials are not expected to be independent to each other. [0]
      It's mainly due to "peer pressure" and the fact that employees are very likely to know each other very well. In that case, the opinion of one employee might influence other employees. [2]

   c) What if there are N = 20,000 employees in the company (let's assume that all employees are working on site for the sake of argument), do you think the "trials" are independent to each other? Briefly justify your answer. **[0+2 marks]**
      It's hard to say, but it's definitely more likely that the "trials" are independent to each other than the previous part when N = 20.
      Note: There is a rule of thumb in most statistics textbook that when the population size (N) is at least 100 times bigger than the sample size (n), the trials can be safely assumed to be independent to each other.

   <u>Regardless of what the answer from previous parts, you can assume that the above situation satisfies all four Binomial conditions for the remaining of this question.</u>

   d) Find the probability that none of the 15 employees will say that they are loyal to their company? **[2 marks]**
      Please make sure that you are able to use the Binomial formula to do these calculations when you don't have access to R (like in midterms).
      Define X as the number of Successes or the number of employees says they are loyal to their company.
      With $n = 15, p = 0.4, X = 0$,
      Using formula: $p(0) = \frac{15!}{0!(15-0)!} 0.4^0 (1-0.4)^{15-0} = 0.00047$.
      Using R: dbinom(0,15,0.4) gives 0.00047.

   e) Find the probability that (exactly) 4 of the 15 employees will say that they are loyal to their company? **[2 marks]**
      With $n = 15, p = 0.4, X = 4$,
      Using formula: $p(4) = \frac{15!}{4!(15-4)!} 0.4^4 (1-0.4)^{15-4} = 0.1268$.
      Using R: dbinom(4,15,0.4) gives 0.1267758

f) Find the probability that at least 4 but at most 10 employees (among the 15 employees) will say that they are loyal to their company? **[2 marks]**
Here X could be 4, 5, …, 10.
Using formula (details not shown): p(4) + p(5) + … + P(10) = 0.9002
Using R: sum(dbinom(4:10,15,0.4)) gives 0.9001504

4. Sports betting generates entertainment as much as income for many. One such bet of a game is "over/under". Vegas or any betting sites will give you "a line" (or the number of expected goals of certain game). If you bet "over", that means you believe the total goals will be over this line. Suppose 3 goals are expected in a typical 90-minute game in the English Premier League (EPL) in the 2024-25 season. Use the Poisson distribution to answer the following questions.

a) Recall that Poisson distribution requires an "independence" condition. Briefly describe what it means here. **[2 marks]**
The number of goals scored in one 90-minute EPL game (in the same season) should not influence the number of goals scored in another 90-minute EPL game, and vice versa.

b) Find the probability that there will be a "nil-nil draw" in a regular 90-minute game, i.e. no goals from both teams. **[1 mark]**
Using formula:
$\boldsymbol{p(0) = \frac{e^{-3} \cdot 3^\wedge 0}{0!} = 0.0498.}$
Using R: ppois(0,3) gives 0.04978707.

c) Find the probability that there will be 4 or more goals scored in a regular 90-minute game. **[2 marks]**
Using formula:
Answer = p(4) + p(5) + … = 1 – p(0) – p(1) – p(2) – p(3) = 0.3528
Using R gives 0.3527681 (any of the following)
ppois(3,3,lower.tail = FALSE)
1-ppois(3,3)
1-sum(dpois(0:3,3))

d) If the "Vegas line" is 3.50 goals in a EPL (90-minute) game, would you bet "over" or "under" based on this Poisson model. Make sure you show some calculation to support your answer. Please only focus on probability. **[0+2 marks]**
We need to fin the 2 probabilities of "over" and "under".
Over: $\boldsymbol{P(X > 3.5) = P(X \geq 4) = 0.3528}$ **(from the previous part).**
Under: $\boldsymbol{P(X < 3.5) = P(X \leq 3) = 1 - 0.3528 = 0.6472.}$
I would bet on "under" because the probability is higher.

5. Many states have programs for assessing the skills of students in various grades. The Indiana Statewide Testing for Educational Progress (ISTEP) is one such program. In a recent year, all tenth-grade Indiana students took the English/language arts exam. The average score was 550 and the standard deviation was 40. Assuming the ISTEP score follows a Normal distribution, answer the following questions.
   Subjects: Indiana tenth-grade students
   # of variable: 1
   Variable: ISTEP score (no units)
   Type: numerical variable
   a) Find the percentage of tenth-grade students who have an ISTEP score over 500. [2 marks]
      Please make sure that you are able to use the Z-table to find the answers when you don't have access to R (like in midterms).
      Using Z-table:
      $\boldsymbol{Z_1 = \frac{500-550}{40} = -1.25}$ **-> the cumulative area is 0.1056.**
      Hence, the answer is 1 – 0.1056 = 0.8944.
      Using R: Either one of the following gives 0.8943502
      pnorm(500,550,40,lower.tail = FALSE)
      1-pnorm(500,550,40)

   b) Use the Z-table to find the percentage of tenth-grade students who have an ISTEP score between 600 and 630. **[2 marks]**
      Using Z-table:
      $Z_1 = \frac{600-550}{40} = +1.25$ and $Z_2 = \frac{630-550}{40} = +2.00$
      The two cumulative areas are 0.8944 and 0.9772 respectively.
      Hence, the answer is the difference of the two = 0.9772 – 0.8944 = 0.0828.
      Using R: diff(pnorm(c(600,630),550,40)) gives 0.08289964

   c) Suppose 36% of the Indiana tenth-grade students score below Y on the ISTEP exam. Find the value of Y. **[2 marks]**
      Using Z-table:
      The closest cumulative area to 0.36 is 0.3594 which has the corresponding Z-score of -0.36.
      $Y = 550 - 0.36 \times 40 = 535.6$
      Using R: qnorm(0.36,550,40) gives 535.6616.

6. Suppose the times between the "99 UBC B-Line" bus arriving at the Broadway station follows an Exponential distribution with average of 6 minutes.
   a) Find the probability that the time between two B-Line buses is more than 10 minutes. **[2 marks]**
      Given: $\boldsymbol{\cancel{\mu = 8}}$ $\boldsymbol{\mu = 6}$ **(mean parameter)**
      Using formula:
      $\boldsymbol{P(X > 10) = e^{-\frac{10}{6}} = 0.1889}$
      Using R: 1-pexp(10,1/6) gives 0.1888756.

b) Find the probability that the time between two B-Line buses is less than 1 minute. **[2 marks]**
Using formula:
$\boldsymbol{P(X < 1) = 1 - e^{-\frac{1}{6}} = 0.1535}$
Using R: pexp(1,1/6) gives 0.1535183.

c) Find the probability that the time between two B-Line buses is between 3 minutes and 6 minutes. **[2 marks]**
Using formula:
1) Easier (with no calculus):
$\boldsymbol{P(X > 6) = e^{-\frac{6}{6}} = 0.36788}$
$\boldsymbol{P(X > 3) = e^{-\frac{3}{6}} = 0.60653}$
Answer = $\boldsymbol{P(X > 3) - P(X > 6) = 0.60653 - 0.36788 = 0.23865}$
2) Harder (using calculus):
$\boldsymbol{P(3 < X < 6) = \int_3^6 \frac{1}{6} e^{-\frac{x}{6}}\, dx = e^{-\frac{3}{6}} - e^{-\frac{6}{6}} = .60653 - 0.36788 = 0.23865}$
Using R: either one of the following gives 0.2386512.
pexp(6,1/6)-pexp(3,1/6)
diff(pexp(c(3,6),1/6))

d) Suppose 25% of the times between two B-Line buses is shorter than Y minutes. Find Y. **[2 marks]**
Given: left hand area = 0.25 or right hand area = 0.75.
$\boldsymbol{P(X > y) = 0.75 = e^{-\frac{y}{6}}}$ **->** $\boldsymbol{y = -6 \cdot ln(0.75) = 1.726}$
Or 25% of the times between two B-Line buses is less than 1.726 minutes.