45.5/50

DANA 4800 Homework #3

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.

(R,B):(1,1),(1,2),...,(1,6),(2,1),...,(6,6)
Written out by rows:
(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]**

RV	Dices values	Number Outcomes	P(X)
-5	(1,6)	1	1/36
-4	(1,5), (2,6)	2	2/36
-3	(1,4), (2,5), (3,6)	3	3/36
-2	(1,3), (2,4), (3,5), (4,6)	4	4/36
-1	(1,2), (2,3), (3,4), (4,5), (5,6)	5	5/36
0	(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)	6	6/36
1	(2,1), (3,2), (4,3), (5,4), (6,5)	5	5/36
2	(3,1), (4,2), (5,3), (6,4)	4	4/36
3	(4,1), (5,2), (6,3)	3	3/36
4	(5,1), (6,2)	2	2/36
5	(6,1)	1	1/36

   c) Calculate E(X) and provide an interpretation of it. **[1+2 marks]**

$$E(X) = \sum x \cdot p(x)$$

$$\mathrm{E(X)} = -5(1/36) - 4(2/36) - 3(3/36) - 2(4/36) - 1(5/36) + 0 + 1(5/36) + 2(4/36) + 3(3/36) + 4(2/36) + 5(1/36) = 0$$

Calculate the expected value of the difference between the two dices (Red minus Black). Since both dice are fair and each outcome is equally likely, the distribution of positive and negative values is symmetric around zero. Therefore, the expected value of this difference will always be zero. This reflects the fact that neither die has an advantage over the other.

-.5

   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.

$$Var(X) = E(X^2) - (E(X))^2$$

$$\mathrm{E(X2)} = 25 \cdot \frac{1}{36} + 16 \cdot \frac{2}{36} + 9 \cdot \frac{3}{36} + 4 \cdot \frac{4}{36} + 1 \cdot \frac{5}{36} + 0 \cdot \frac{6}{36} + 1 \cdot \frac{5}{36} + 4 \cdot \frac{4}{36} + 9 \cdot \frac{3}{36} + 16 \cdot \frac{2}{36} + 25 \cdot \frac{1}{36}$$

Var(X) = 5.833- 0 = 5.833
SD(X) = $\sqrt{5.833}$≈ 2.415
Interpretation:
The standard deviation is about 2.41, meaning the difference between red and black typically varies by about

-.5

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) In the next three parts (a to c), we focus on betting $1 on even numbers and define the random variable X as the winning amount (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]**

E(X)	Probability
+1	18/38
-1	20/38

   b) Calculate the expected value of X. **[1 mark]**

$$E(X) = \sum x \cdot p(x) = +1(18/38) - 1(20/38) = -0.0526$$

   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]**

   The probability of winning favors the house, not the player, so the phrase "the house always wins" is reasonable in this context. Based on the expected value, the player can expect to lose approximately $0.0526 for every $1 bet in the long run.

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]**

   The answer of one employee does not affect the answer of another employee and 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 might not be independent. Because in this case, it is a small population, and the employees could get to know each other, and if someone expresses an opinion prior to the survey, they may be predisposed during the survey.

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]**

Yes, the trials might be independent. Because it is a huge population size, and the 15 employees could not get to know each other.

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.

d) Find the probability that none of the 15 employees will say that they are loyal to their company? **[2 marks]**

n=15

$$P(X = 0) = \frac{n!}{x!\,(n-x)!}\,\mathrm{p}^{x}(1-\mathrm{p})^{\mathrm{n-x}}$$

$$P(X = 0) = \frac{15!}{0!\,(15-0)!}\,(0.4)^{0}(1-0.4)^{15-0} = (0.6)^{15}$$

The probability that none of the 15 employees are loyal is 0.00047.

e) Find the probability that (exactly) 4 of the 15 employees will say that they are loyal to their company? **[2 marks]**

$$P(\mathrm{X} = 4) = \frac{15!}{4!\,(15-4)!}\,(0.4)^{4}(1-0.4)^{15-4}$$

$$P(X = 4) = 1365 \times 0.0256 \times 0.0036 \approx 0.127$$

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]**

$$\mathrm{P}(4 \le \mathrm{X} \le 10) = \mathrm{P}(\mathrm{X}=4) + \mathrm{P}(\mathrm{X}=5) + \mathrm{P}(\mathrm{X}=6) + \mathrm{P}(\mathrm{X}=7) + \mathrm{P}(\mathrm{X}=8) + +\mathrm{P}(\mathrm{X}=9) + \mathrm{P}(\mathrm{X}=10)$$

$$P(X = 4) = 1365 \times 0.0256 \times 0.0036 \approx 0.127$$

$$\mathrm{P}(\mathrm{X} = 5) = \frac{15!}{5!\,(15-5)!}(0.4)^{5}(1-0.4)^{15-5} = 0.1859$$

$$\mathrm{P}(\mathrm{X} = 6) = \frac{15!}{6!\,(15-6)!}(0.4)^{6}(1-0.4)^{15-6} = 0.2066$$

$$\mathrm{P}(\mathrm{X} = 7) = \frac{15!}{7!\,(15-7)!}(0.4)^{7}(1-0.4)^{15-7} = 0.1771$$

$$\mathrm{P}(\mathrm{X} = 8) = \frac{15!}{8!\,(15-8)!}(0.4)^{8}(1-0.4)^{15-8} = 0.1181$$

$$\mathrm{P}(\mathrm{X} = 9) = \frac{15!}{9!\,(15-9)!}(0.4)^{9}(1-0.4)^{15-9} = 0.0612$$

$$\mathrm{P}(\mathrm{X} = 10) = \frac{15!}{10!\,(15-10)!}(0.4)^{10}(1-0.4)^{15-1} = 0.025$$

$$\mathrm{P}(4 \le \mathrm{X} \le 10) = 0.9004$$

.9002

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 a typical 90-minute game is *independent** of the number of goals scored in any other 90-minute game or any other non-overlapping time interval. This means that the outcome in one period does not affect the outcome in another.

** As a soccer fan, I would note that in reality, the number of goals scored in one match is not always entirely independent from other matches—especially in tournaments or semifinals. Teams often adjust their strategies based on the outcomes or ongoing results of other games. For example, if a draw guarantees a better position (like not risking elimination or avoiding third place), teams may prefer to play defensively rather than take risks. Nowadays, with real-time communication, players and coaches are often aware of what's happening in parallel games, which can influence their approach during the match. So, while the independence assumption simplifies statistical analysis, it doesn't always fully reflect real-world behavior in sports contexts.*

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]**

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$$P(0) = \frac{e^{-3} \cdot 3^0}{0!} = 0.0498$$

The probability that there will be a "nil-nil draw" in a regular 90-minute game is 0.0498

c) Find the probability that there will be 4 or more goals scored in a regular 90-minute game. **[2 marks]**

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$$P(0) = \frac{e^{-3} \cdot 3^0}{0!} = 0.0498$$

$$P(1) = \frac{e^{-3} \cdot 3^1}{1!} = 0.1494$$

$$P(2) = \frac{e^{-3} \cdot 3^2}{2!} = 0.2240$$

$$P(3) = \frac{e^{-3} \cdot 3^3}{3!} = 0.2240$$

4. Add the probabilities for 0, 1, 2, 3 goals:

$$P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$P(X \le 3) \approx 0.0498 + 0.1494 + 0.2240 + 0.2240$$
$$P(X \le 3) = 0.6472$$

-2

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]**

$$\boldsymbol{e^{-3.5} \approx 0.030197}$$
$$\boldsymbol{P(0) = 0.030197 \times 1 = 0.030197}$$
$$\boldsymbol{P(1) = 0.030197 \times 3.5 = 0.105688}$$
$$\boldsymbol{P(2) = 0.030197 \times 3.5^2/2 = 0.184954}$$
$$\boldsymbol{P(3) = 0.030197 \times 3.5^3/6 = 0.215779}$$
$$\boldsymbol{P(x \leq 3.5) = 0.536617}$$
$$\boldsymbol{P(x > 3.5) = 1 - 0.536617 = 0.463382}$$

-.5

Under 3.5 ≈ 53.7% .6472
Over 3.5 ≈ 46.3% .3528

I would bet under the Vegas line (3.5 goals). The probability that the result would be over the Vegas line in this case is 4 goals or more is 53.7%, its means that the probability of a result under the Vegas line is 46.3%, which is bigger.

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.

a) Find the percentage of tenth-grade students who have an ISTEP score over 500. [2 marks]

$$Z = (X - \mu) / \sigma = (500 - 550) / 40 = -1.25$$

-1

b) Use the Z-table to find the percentage of tenth-grade students who have an ISTEP score between 600 and 630. **[2 marks]**

$$Z(600) = \frac{600 - 550}{40} = 1.25$$

$$Z(630) = \frac{630 - 550}{40} = 2$$

From Z-table, the probability for Z(630) is 0.9772 and the probability for Z(600) is 0.8944. Then the percentage of tenth-grade students who have an ISTEP score between 600 and 630 is 8.28%.

c) Suppose 36% of the Indiana tenth-grade students score below Y on the ISTEP exam. Find the value of Y. **[2 marks]**

$$\mathrm{X} = \mathrm{Z} * \sigma + \mu = (-0.36) * 40 + 550 = 535.6$$

Form Z-table, when the probability is 36% then equals to -0.36
The score is approximately 535.6 on the ISTEP exam.

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]**

$$P(X > 10) = e^{-\frac{k}{\mu}} = e^{-\frac{10}{6}} = 0.1889$$

   The probability that the time between two B-Line buses is more than 10 minutes is 0.1889.

   b) Find the probability that the time between two B-Line buses is less than 1 minute. **[2 marks]**

$$P(X < 1) = 1 - e^{-\frac{k}{\mu}} = e^{-\frac{1}{6}} = 0.1535$$

   The probability that the time between two B-Line buses is less than 1 minute is 0.1535

   c) Find the probability that the time between two B-Line buses is between 3 minutes and 6 minutes. **[2 marks]**

$$P(X > 3) = e^{-\frac{k}{\mu}} = e^{-\frac{3}{6}} = 0.6065$$
$$P(X > 6) = e^{-\frac{k}{\mu}} = e^{-\frac{6}{6}} = 0.3679$$
$$P(3 < X < 6) = 0.2387$$

   The probability that the time between two B-Line buses is between 3 minutes and 6 minutes is 0.2387

   d) Suppose 25% of the times between two B-Line buses is shorter than Y minutes. Find Y. **[2 marks]**

$$P(X) = e^{-\frac{k}{\mu}} = 1 - .25 = 0.75$$
$$\ln(P(X)) = \ln(e^{-\frac{k}{\mu}})$$
$$\ln\left(P(X)\right) = -\frac{k}{\mu}$$
$$\ln\left(P(X)\right) * \mu = -K$$

*k= Y

$$Y = -\ln\left(P(X)\right) * \mu$$

$$Y = -\ln(0.75) * \mu = -\ln(0.75) * 6 = 1.73$$

   Y is 1.73 minutes.

7. A survey has been conducted to determine what percentage of Canadian who have smoked Marijuana during their teenage years (13-19 years old). A random sample of 1,000 Canadians was drawn, and it is reported that 210 of them did smoke marijuana during their teenage years.

   a) Identify the parameter of interest. **[2 marks]**

   Subjects = Canadians

   Variable = whether or not they have smoked marijuana during their teenage years

   Type = categorical variable

   Answer = Define p as the proportion of all Canadians who smoked marijuana during their teenage years

   b) Estimate the above parameter using a 95% confidence interval. **[2+2 marks]**
   Note: Make sure you understand that checking assumptions is part of this procedure.

$$\text{Confidence Level } 95\% \Rightarrow \text{ Z} = 1.96$$

$$\bar{\text{p}} \pm \text{Z} * \sqrt{\frac{\bar{\text{p}}(1-\bar{\text{p}})}{n}}$$

$$\bar{\text{p}} = \frac{210}{1000} = 0.21$$

$$0.21 \pm 1.96 * \frac{\sqrt{0.21(1-0.21)}}{\sqrt{1000}}$$

$$0.21 \pm 0.0252$$

   Conditions checking:

   A. n = 1000, which is bigger than the requirement of 30,
   B. n*p = 1000*0.21 = 210, which is bigger than the requirement of 5,
   C. n*(1-p) = 1000*(1-0.21) = 790, which is bigger than the requirement of 5.

   Since all three conditions are met, the approximation is good.

8. One way to know the efficiency of international airports is to measure the percentage of flights that leave the airport on time on a daily basis. From a random sample of 100 days, it was found that an average of 81.5% of flights leave YVR (Vancouver International Airport) on time and a standard deviation of 4.5%. From past experience, it is known that the percentage of flights that leave YVR on time has a Normal distribution.

   a) Identify the subjects of interest. **[1 mark]**

   Subjects = days

   b) Provide a description of the parameter of interest in this study. **[2 marks]**

   Define μ as the average percentage (an average of 81.5%) of all flights that leave YVR on time on a daily basis.

c) Construct a 98% confidence interval of the parameter in part (a). **[3 marks]**

98% Confidence Interval of μ
98% CI means the middle-area is 0.98. That also means we have two 2% on either sides.
Using CA = 0.0099 (closest to 0.01) and look up the Z-table, we get a Z-score of 2.33.

$$\bar{p} \pm Z * s/\sqrt{n}$$

$$0.815 \pm 2.33 * \frac{0.045}{\sqrt{100}}$$

0.815 ±0.0105