Assignment on Distributions

1. Calculate frequency distribution for Age column in Titanic and also PMF for the same

    a.) Use pandas df tutorial.

Frequency distribution for the Age column with 10 bins

```
df = pd.read_csv(r'C:\Users\sailendraraja\Downloads\end to end data sets\titanic.csv')
```

```
age =df['Age']
print(f'frequency distribution for age with min age {min(age)}, max age {max(age)}')

pd.cut(age,bins=10).value_counts()
```

```
frequency distribution for age with min age 0.42, max age 80.0

(16.336, 24.294]    177
(24.294, 32.252]    169
(32.252, 40.21]     118
(40.21, 48.168]      70
(0.34, 8.378]        54
(8.378, 16.336]      46
(48.168, 56.126]     45
(56.126, 64.084]     24
(64.084, 72.042]      9
(72.042, 80.0]        2
Name: Age, dtype: int64
```

PMF for the Age column with 10 bins

```
pmf = age.value_counts()/len(age)
print(f'PMF with min age {round(min(pmf),5)}, max age {round(max(pmf),5)}')

pd.cut(pmf,bins=10).value_counts()
```

```
PMF with min age 0.00112, max age 0.03367

(0.00109, 0.00438]    37
(0.00438, 0.00763]    15
(0.0141, 0.0174]       8
(0.00763, 0.0109]      7
(0.0174, 0.0207]       6
(0.0272, 0.0304]       5
(0.0109, 0.0141]       5
(0.0239, 0.0272]       3
(0.0304, 0.0337]       1
(0.0207, 0.0239]       1
Name: Age, dtype: int64
```

2. The average monthly sales of 2,000 firms are normally distributed with mean Rs 38,000 and standard deviation Rs 10,000. Find :

    (i) The number of firms with sales of over Rs 50,000.

    (ii) The percentage of firms with sales between Rs 38,500 and Rs 41,000.

    (iii) The number of firms with sales between Rs 30,000 and Rs 50,000.

② $\mu = 38{,}000 \qquad \sigma = 10{,}000$

(i) $P(X > 50000)$

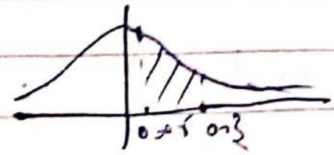
$$Z = \frac{X - \mu}{\sigma} = 1.2$$

$\Rightarrow P(X > 1.2) = 0.1151$

Firms = 230

(ii) $P(38500 < X < 41000)$

$= P(0.05 < X(z) < 0.3)$



% firms = 9.8%

(iii) $P(30000 < X < 50000)$

$= P(-0.8 < X(z) < 1.2)$

67.3%

Firms = 1346

3. A test is conducted which consists of 20 MCQs with every question having 4 options. Determine the probability of a person answering exactly 5 wrong answers.

(3) $n = 20$

$p = 3/4$

$q = 1 - p = 1/4$

$r = 5$

$$P(X=5) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

$$= {}^{20}C_{5} (0.75)^{5} (0.25)^{15}$$

$$= 0.0003426\ \%$$

4. In an observational astronomy experiment, let the average rate of photons reaching the telescope is 4 photons per second (Poisson random variable with mean of 4). Find the probability that no photon reaches the telescope in a given second.

(4) $\lambda = 4$, $r = 0$

$$P(X=0) = \frac{e^{-\lambda} \lambda^{r}}{r!}$$

$$= \frac{e^{-4}\, 4^{0}}{0!} = \cancel{[illegible]}$$

$$= 1.83\ \%$$

5. The number of calls coming per minute into a Customer support centre is Poisson random variable with mean 3.

(a) Find the probability that no calls come in a given 1-minute period.

(b) Assume that the number of calls arriving in two different minutes are independent. Find the probability that at least two calls will arrive in a given two-minute period.

(5) $\lambda = 3$

(a) $P(X=0) = \frac{e^{-3} 3^0}{0!} = 4.98\%$

$\lambda = 6$

(b) $P(X \geq 2) = 1 - P(X<2)$

$= 1 - \frac{e^{-6} 6^0}{0!} - \frac{e^{-6} 6^1}{1!} = $ ~~[illegible]~~ 98.26%

6. If a production line has a 20% defective rate. Calculate the probability of obtaining the first defected part after three good parts. What is the average number of inspections to obtain the first defective?

(6) $p = 0.2, \; q = 0.8$

$P(X=1) = {}^4C_1 (0.2)^1 (0.8)^3$

$= 40.96\%$

for every 2.5 ~~non-defective~~ inspection − 1 defective is found

or 5 inspection − 2 defects.

(59.04%)

7. The probability that a student is accepted to a prestigious college is 0.3. If 5 students from the same school apply, what is the probability that at most 2 are accepted?

(7) $P = 0.3 \quad n = 5$

$P(X \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$= {}^5C_0 (0.3)^0 (0.7)^5 + 5C_1 (0.3)^1 (0.7)^4$
$+ 5C_2 (0.3)^2 (0.7)^3$

$= 83.69\ \%$

8. **The maximum weight that an elevator in an apartment complex can accommodate is 800kg. The average adult weight be about 70 kg with a variance of 200. What is the probability that the lift safely reaches the ground when there are 10 different adults in the lift? What if there are 12 adults?**

(8)

Max Weight $= 800$

$\mu = 70$ $\quad$ Var $= 200$ $\quad$ S.d $= \sigma = \sqrt{200} = 14.14$

$$P(X=10) = P(Z < 0.707) \qquad z = \frac{800-700}{141.4} = 0.707$$

$$= 77.94\%$$

$$P(X=12) = P(Z < -0.282) \qquad z = \frac{800-840}{141.4} = -0.282$$

$$= 38.9\%$$

9. **A student, to test his luck, went to an examination unprepared. It was a MCQ type examination with two choices for each questions. There are 50 questions of which at least 20 are to be answered correctly to pass the test. What is the probability that he clears the exam? If each question has 4 choices instead of two. What is the probability that he clears the exam?**

(9) $n = 50$

$p = 1/2, \ q = 1/2$

$$P(X \geq 20) = \sum_{i=20}^{50} {}^{50}C_i (0.5)^i (0.5)^{50-i}$$

$$= 94.05\%$$

$p = 1/4, \ q = 1/4$

$$P(X \geq 20) = 1.39\%$$

10. A company manufactures LED bulbs with a faulty rate of 30%. If I randomly select 6 bulbs, what is the probability that exactly 2 are faulty?

(10) $p = 0.3$ $n = 6$
$q = 0.7$

$P(x=2) = {}^6C_2 (0.3)^2 (0.7)^4$

$= 32.4\%$

11. For a writer, the efficiency of typing is 6 errors per hour entering 77 words per minute. What is the probability of error 2 errors in 322 word report?

(11) ~~Ans~~ 6 errors/hr 77 wpm = 77x60 wp/hr

x 322 wp/hr

$x \times 77 \times 60 = 322 \times 6$

$x = 0.418$ Error/hr

$P(X=2) = \frac{e^{-\lambda} \lambda^r}{r!} = \frac{e^{-0.418} \lambda^2 (0.418)}{2!}$

$= 5.75\%$

12. Executives in the New Zealand Forestry Industry claim that only 5% of all old sawmills sites contain soil residuals of dioxin (an additive previously used for anti-sap-stain treatment in wood) higher than the recommended level. If Environment Canterbury randomly selects 20 old saw mill sites for inspection, assuming that the executive claim is correct:

a) Calculate the probability that less than 1 site exceeds the recommended level of dioxin.

b) Calculate the probability that less than or equal to 1 site exceed the recommended level of dioxin.

c) Calculate the probability that at most (i.e., maximum of) 2 sites exceed the recommended level of dioxin.

(12) $P = 0.05,\ q = 0.95$

$n = 20$

(a) $P(x < 1) = {}^{20}C_0 (0.05)^0 (0.95)^{20}$

$= 35.84\%$

(b) $P(x \leq 1) = {}^{20}C_0 (0.05)^0 (0.95)^{20} + {}^{20}C_1 (0.05)^1 (0.95)^{19}$

$= 73.58\%$

(c) $P(P \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$

$= 92.45\%$

13. Inland Revenue audits 5% of all companies every year. The companies selected for auditing in any one year are independent of the previous year's selection.

    a) What is the probability that the company 'Ross Waste Disposal' will be selected for auditing exactly twice in the next 5 years?

    b) What is the probability that the company will be audited exactly twice in the next 2 years?

    c) What is the exact probability that this company will be audited at least once in the next 4 years?

(13) $p = 0.05 \quad q = 0.95$

$n = 5 \quad r = 2$

(a) $P(X=2) = {}^{5}C_2 (0.05)^2 (0.95)^3$

$= 2\%$

(b) $P(X=2) = {}^{2}C_2 (0.05)^2 (0.95)^0$

$= \cancel{0.25}\ 0.25\%$

(c) $P(X \geq 1) = \sum_{i=1}^{4} {}^{4}C_i\, (0.05)^i (0.95)^{4-i}$

$= 18.54\%$

14. The probability that a driver must stop at any one traffic light coming to Supervised Learning
is 0.2. There are 15 sets of traffic lights on the journey.

a) What is the probability that a student must stop at exactly 2 of the 15 sets of traffic lights?

b) What is the probability that a student will be stopped at 1 or more of the 15 sets of traffic lights?

(14) $p = 0.2$, $q = 0.8$
$n = 15$

(a) $P(x=2) = {}^{15}C_2 (0.2)^2 (0.8)^{13}$

$= 23.08\%$

(b) $P(x \geq 1) = \sum_{i=0}^{15} {}^{15}C_i (0.2)^i (0.8)^{15-i} = 1 - P(x<1)$

$= 96.48\%$