Learn DU

SOLVED PYQ

Paper: Introductory Statistics for Economics

Course: B. A.(Hons.) Economics I Year

Year: 2022

Modified as Per New NEP Syllabus

ECON003: Introductory Statistics for Economics
Course : B. A.(Hons.) Economics I Year
Duration : 3 Hours **Maximum Marks : 90**

SEMESTER–1
2022

This paper contains four sections. Attempt all sections.

Q. 1. The marks of 21 students in a 50 marks mathematics test are given below:

18	**20**	**25**	**28**	**30**	**35**	**36**	**38**	**39**	**40**	**41**	**41**
41	**42**	**42**	**43**	**44**	**45**	**45**	**47**	**50**			

Calculate a 10% trimmed mean for the data above. Also calculate the median. It was later discovered that the student whose marks were recorded as 35, actually had 45 marks. How will this affect the median value? **(5)**

Ans.

$$n = 21$$

$$10\% \text{ of } n = 2.1 \approx 2$$

So, we remove first 2 and last 2 observations to calculate the trimmed mean.

$$\text{Hence, } 10\% \text{ trimmed mean} = \frac{\sum_{i=3}^{19} x_i}{17}$$

where x_i 's represent the marks of the students

$$= \frac{655}{17}$$

$$= 38.53$$

$$\text{Median term (s)} = \frac{21+1}{2}$$

$$= 11 \text{ th}$$

$$\text{Hence, median} = 41$$

Median would be one term after 41 (original) which in this case is still 41.

Q. 2. (a) An electronic machine has three major circuits-*X*, *Y* and *Z*. Circuits *Y* and *Z* are interdependent while circuit *X* operates independently of circuit *Y* and circuit *Z*. It is known that circuit *X* works properly 80% of the time, circuit *Y* 90% of the time and circuit *Z* 75% of the time. However if circuit *Z* fails there is a 60% chance that circuit Y will also fail. What is the probability that only circuit Y works?

Ans.

$$P(X) = 80\%$$

$$P(Y) = 90\%$$

$$P(Z) = 75\%$$

$$P\left(\frac{\bar{Y}}{\bar{Z}}\right) = 60\%$$

$$P\text{ (only } Y \text{ works)} = \left(Y \cap \bar{X} \cap \bar{Z}\right) = P\left(Y \cap \bar{Z}\right) \cap P\left(\bar{X}\right) \quad ...(i)$$

Now $$P\left(\frac{\bar{Y}}{\bar{Z}}\right) = 60\%$$

$$\Rightarrow \quad P\left(\frac{Y}{\bar{Z}}\right) = 40\%$$

$$\Rightarrow \quad \frac{P\left(Y \cap \bar{Z}\right)}{P\left(\bar{Z}\right)} = 40\%$$

$$\Rightarrow \quad P\left(Y \cap \bar{Z}\right) = 25\% \times 40\%$$

$$= 10\%$$

Putting this in (i), we get

$$P\text{ (only } \bar{Y} \text{ works) } P\left(Y \cap \bar{X} \cap \bar{Z}\right) = 10\,\% \times 20\,\%$$

$$= 2\%$$

(b) How many five lettered words can be made out of the 26 letters of the English alphabet if repetitions are allowed but not consecutive repetitions? This last sentence means that a letter may not follow itself in the same word.

Ans.

(a)		(b)		(c)		(d)		(e)
26	×	25	×	25	×	25	×	25
↓		↓		↓		↓		
Any of 26 letters		Any of 25 letters except one that is used in first place				Same logic as for (b)		

Hence, the number of words

$$= 1015\,6250$$

(c) If a machine is correctly set up it will produce 90% acceptable items. If it is incorrectly set up it will produce 30°% acceptable items. Past experience shows that 80% of set ups are correctly done. If after a certain set up, the first item produced is acceptable, what is the probability that the machine is correctly set up? **(4+3+3)**

Ans. W : – Working properly (machine is set up correctly)

A : – Acceptable items

$$P\left(A/W\right) = 90\%,\ P\left(A/\bar{W}\right) = 30\%$$

$$P(W) = 80\% \text{ Hence } P(\bar{W}) = 20\%$$

$$P\left(A/W\right) = P\left(\frac{W \cap A}{P(A)}\right)$$

$$= \frac{P(W) \cdot P(A/W)}{P(W) \cdot P(A/W) + P(\bar{W}) \cdot P(A/W)}$$

$$= \frac{80\% \times 90\%}{80\% \times 90\% + 20\% \times 30\%}$$

$$= 92.31\%$$

Q. 3. (a) A factory employs 10 workers in the production department, 8 workers in the packaging department and 7 workers in the delivery department. Out of these workers 5 are to be randomly selected for a training programme, (i) What is the probability that all selected workers will be from the same department? (ii) What is the probability that at least two different departments will be represented among the selected workers?

Ans. P (All selected workers are in the same department) = P (All are from production department) + P (All are from packaging department) + P (All are from the delivery department)

$$= \frac{^{10}C_5}{^{25}C_5} + \frac{^{8}C_5}{^{25}C_5} + \frac{^{7}C_5}{^{25}C_5}$$

$$= \frac{\frac{10!}{5151} + \frac{8!}{5131} + \frac{7!}{5121}}{\frac{251}{20!5!}}$$

$$= \frac{\frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1} + \frac{8 \times 7 \times 6}{3 \times 2} + \frac{7 \times 6}{2 \times 1}}{\frac{25 \times 24 \times 23 \times 22 \times 21}{5 \times 4 \times 3 \times 2 \times 1}}$$

$$= \frac{252 + 56 + 21}{53130}$$

$$= \frac{329}{53130}$$

(b) Let A be the event that a randomly selected individual likes vanilla flavour, B be the event that a randomly selected individual likes strawberry flavour and C be the event that a randomly selected individual likes chocolate flavour.

Suppose that

P(A) = 0.65 P(B) = 0.55 P (C) = 0.70

P(A∪ B) = 0.8 P(B ∩ C) = 0.3 P(A ∪ B ∪ C) = 0.9

(i) What is the probability that the individual likes both the vanilla and the strawberry flavours?

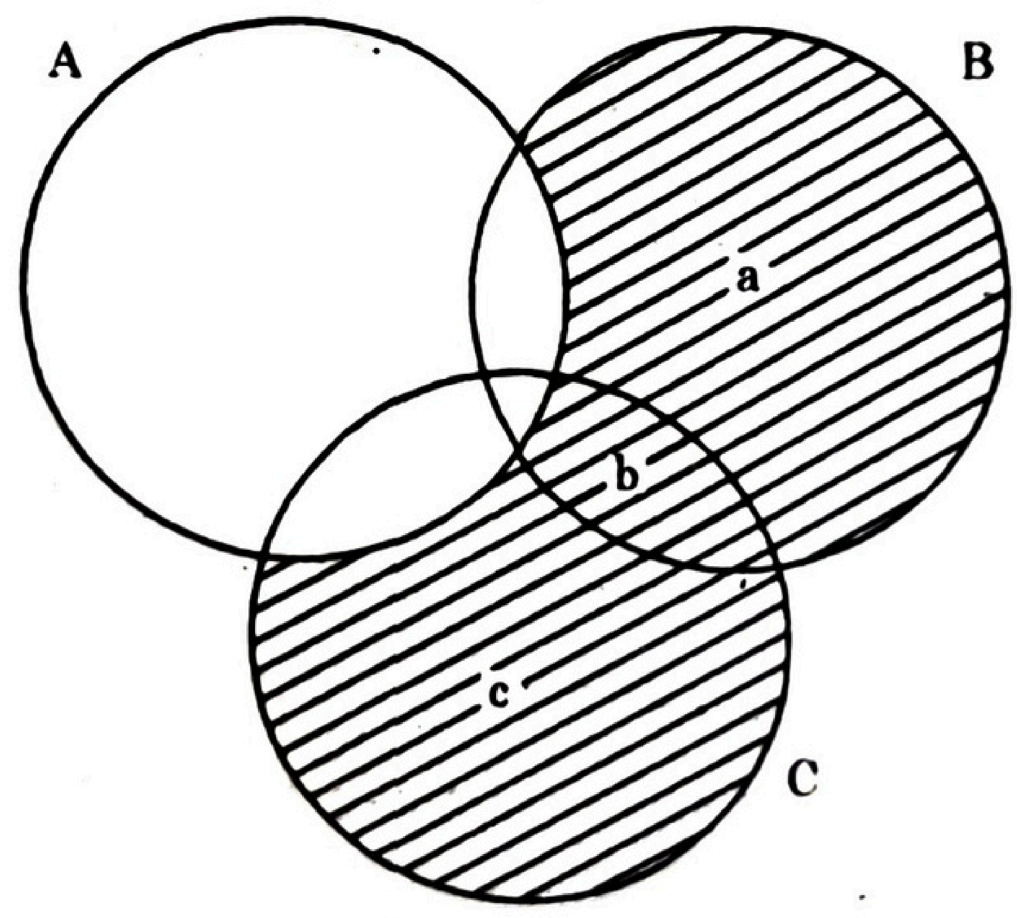
(ii) If it is known that the individual did not like vanilla, what now is the probability that the individual liked at least one of the other two flavours?

Ans. (i)

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$
$$= 0.65 + 0.55 -- 0.80$$
$$= 0.40$$

(ii)

$$B \cup C = P\left(\frac{B \cup C \cap \bar{A}}{P(\bar{A})}\right)$$



$$P\left(\frac{B \cup C}{\bar{A}}\right) = \frac{P((B \cap \bar{A}) \cup (C \cap \bar{A}))}{P(\bar{A})}$$

$$= \frac{P(A \cup B \cup C) - P(A)}{P(\bar{A})}$$

$$= \frac{0.9 - 0.65}{0.35}$$

$$= \frac{0.25}{0.35}$$

$$= \frac{5}{7}$$

(c) Show that for any three events A, B and C with P(C) > 0,
$P(A \cup B/C) = P(A/C) + P(B/C) - P(A \cap B/C)$ **(4+ 4+ 2)**

Ans. L.H.S $P(A \cup B/C) = \frac{P((A \cup B) \cap C)}{P(C)}$

$$= \frac{P((A \cap C) \cup (B \cap C))}{P(C)}$$

$$= \frac{P(A \cap C) + P(B \cap C) - P(A \cap C \cap B \cap C)}{P(C)}$$

Using $P(X \cap Y) = P(X) + P(Y) - P(X \cap Y)$

$\Rightarrow$ $P(A \cup B/C) = \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)} - \frac{P(A \cap B \cap C)}{P(C)}$

$$= P(A/C) + P(B/C) - P(A \cap B/C)$$

SECTION -II

Do any two out of Q. 4. Q. 5 and Q. 6.

Q. 4. (a) A bookseller estimates the probabilities for the number of eekly sales of a particular magazine as follows:

No. of sales(x)	0	1	2	3	4	5
Probability $p(x)$	0.12	0.16	0.31	.20	0.14	0.07

(i) Find the mean and standard deviation of the number of sales.

(ii) As weekly income, the salesperson receives a fixed amount of Rs 200, plus a commission of Rs 150 for each sale made. Find the mean nd standard deviation of this weekly income.

Ans. (i) $E(X) = 0 \times 0.12 + 1 \times 0.16 + 2 \times 0.32 + 3 \times 0.20 + 4 \times 0.14 + 5 \times 0.07$
$= 2.29$

$\text{Var}(X) = 0^2 \times 0.12 + 1^2 \times 0.16 + 2^2 \times 0.31 + 3^2 \times 0.20 + 4^2 \times 0.14 + 5^2 \times 0.07 - (2.29)^2$
$= 25.2659$

Hence, s. d. (X) = 5.0265

(ii) Let Y be the salesman income

Hence, $$Y = 150X + 200$$

$$\Rightarrow \quad E(y) = 150\,E(X)$$
$$= 150 \times 2.29$$
$$= ₹\ 343.50$$

(b) Suppose that the time between successive occurrences of an event follows an exponential distribution with mean I/A. minutes. Assume that an event occurs. What will be the probability that more than 6 minutes elapses before the occurrence of the next event?

Ans. Let x_i be the time between successive occurances

Now, $$x_i \sim \text{Exp}(\lambda)$$

$$\Rightarrow \quad E(X) = \frac{1}{\lambda}$$

$$P(X > 6) = \int_6^\infty \lambda\, e^{-\lambda x} dx$$

$$= \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_6^\infty$$

$$= -\left[e^{-\lambda x}\right]_6^\infty$$

$$= -\left[0 - e^{-6\lambda}\right] = \frac{1}{e^{6\lambda}}$$

(c) Obtain the mean and variance of the random variable X which has a uniform distribution on the interval A, B). **(4+3+3)**

Ans. $$X \sim U(A, B)$$

$$f(x) = \frac{1}{B - A}$$

$$\text{Mean} = E(X) = \int_A^B x\, f(x) dx$$

$$= \int_A^B x \cdot \frac{1}{B - A} dx$$

$$= \frac{1}{B-A} \cdot \left[\frac{x^2}{2}\right]_A^B$$

$$= \frac{B^2 - A^2}{2(B-A)}$$

$$= \frac{B+A}{2}$$

$$E(X^2) = \int_A^B x^2 \frac{1}{B-A} dx$$

$$= \frac{1}{B-A} \cdot \left[\frac{x^3}{3}\right]_A^B$$

$$= \frac{1}{B-A} \cdot \frac{B^3 - A^3}{3}$$

$$= \frac{(B-A)(B^2 + A^2 + AB)}{2(B-A)} = \frac{A^2 + B^2 + AB}{2}$$

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \frac{A^2 + B^2 + AB}{3} - \left(\frac{A+B}{2}\right)^2$$

$$= \frac{A^2 + B^2 + AB}{3} - \left(\frac{A^2 + B^2 + 2AB}{4}\right)$$

$$= \frac{4A^2 + 4B^2 + 4AB - 3A^2 - 3B^2 - 6AB}{12}$$

$$= \frac{A^2 + B^2 - 2AB}{12}$$

$$= \frac{(B-A)^2}{12}$$

Q. 5. (a) Certain random variable *X* has *cdf* *F*(*x*) given by:

$$\begin{cases} 0 & \text{for } x < -1 \\ \dfrac{x+1}{2} & \text{for } -1 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

(i) Obtain the *pdf* of *x*

(ii) Compute $P(-1/2 \le X \le 1/2)$

Ans. (i) $$f(x) = \frac{x+1}{2} \quad -1 < x < 1$$

(ii) $$P\left(\frac{-1}{2} \le X \le \frac{1}{2}\right) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{x+1}{2} dx$$

$$= \frac{1}{2}\left[\frac{x^2}{2} + x\right]_{-\frac{1}{2}}^{\frac{1}{2}}$$

$$= \frac{1}{2}\left[\frac{1}{8} + \frac{1}{2} - \frac{1}{8} + \frac{1}{2}\right] = \frac{1}{2}$$

(b) A research on crop yields suggests that daily rainfall in parts of India appears to be normally distributed with a mean of 2.2 inches during the rainy season. The standard deviation was determined to be 0.8 inches. What is the probability that it will rain more than 3.3 inches on any one day during the rainy season? How much rainfall must occur to exceed 10% of the daily precipitation?

Ans. Let xi be the daily rainfall during the rainy season.

$$xi \sim N\ (2.2, 0.8^2)$$

$$P(X > 3.3) = P\left(Z > \frac{3.3 - 2.2}{0.8}\right)$$

$$= P(Z > 1.375)$$

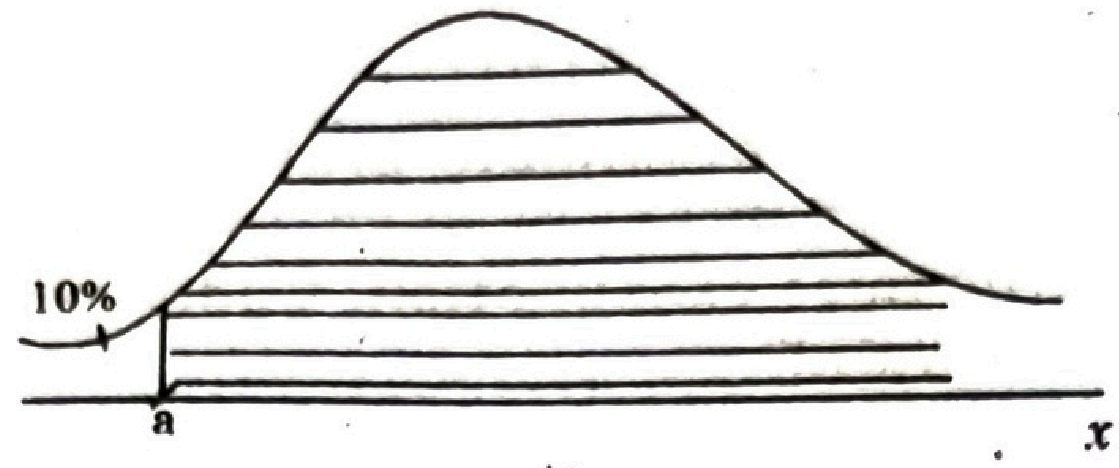
$$= 1 - 0.9147 = 0.0853$$

$$P(X > a) = 90\%$$

$$\Rightarrow \quad P\left(Z > \frac{a - 2.2}{0.8}\right) = 0.90$$

$$\Rightarrow \quad \frac{a - 2.2}{0.8} = 1.29$$

⇒ $$a = 1.168$$

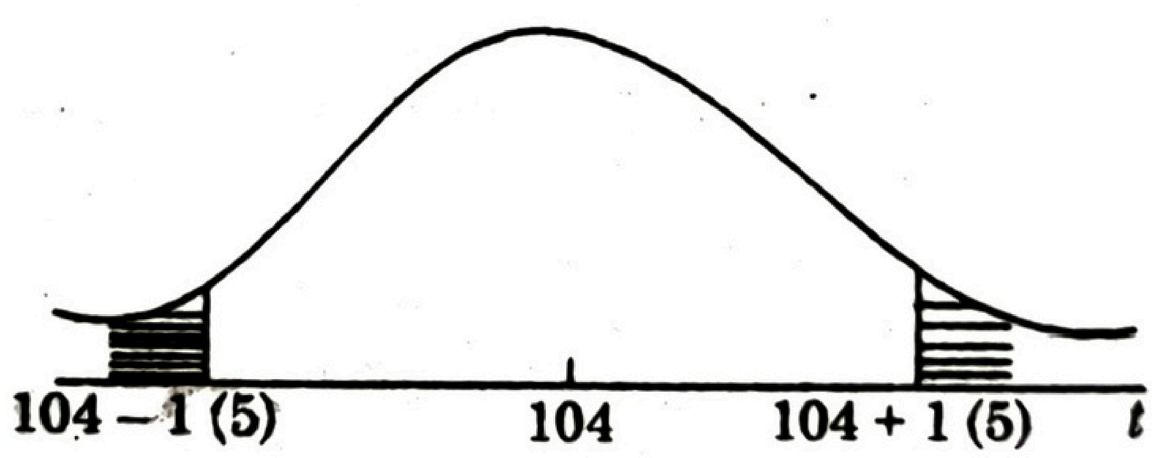


(c) Suppose that student participation in a competition that happens every year has a normal distribution with mean 104 students and standard deviation 5 students. What is the probability that student participation differs from mean by more than 1 standard deviation.

How would you characterize the top extreme 0.1% of the student participation values? (4+3+3)

Ans. Let X be the number of students participating in the competition every year.

So, $$X \sim N(104, 5^2)$$



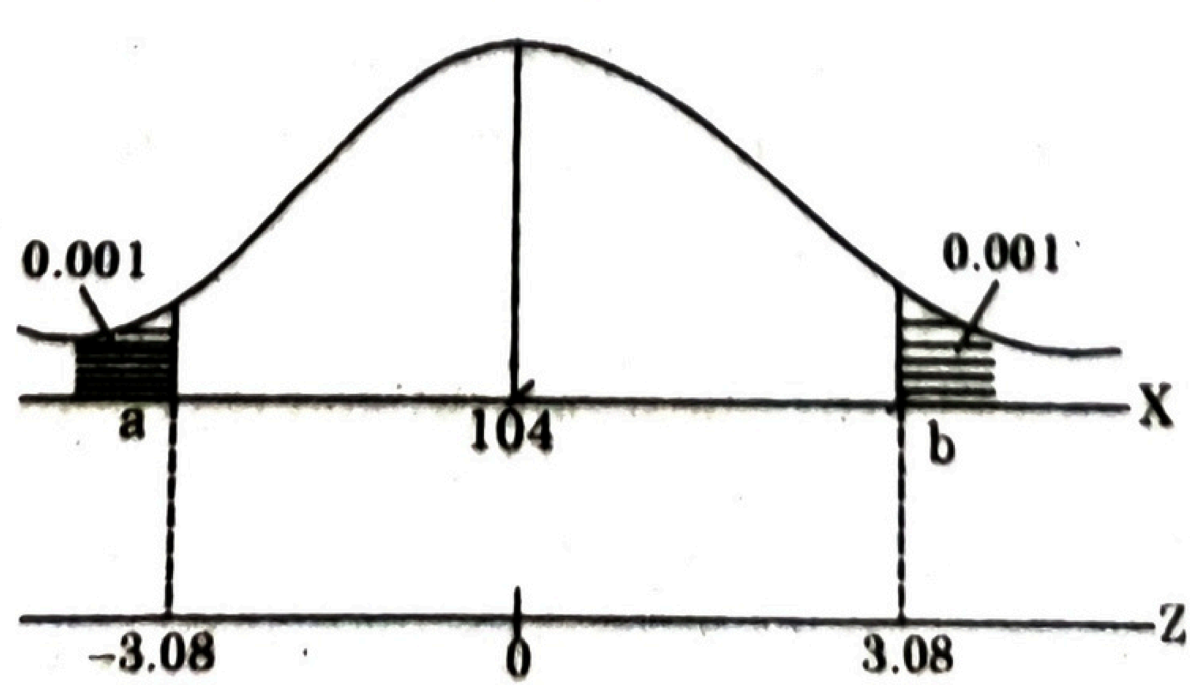
$$P(X > 109 \text{ or } X < 99) = 2P(X < 99)$$

$$= 2P\left(Z < \frac{99-104}{5}\right)$$

$$= 2P(Z < -1)$$

$$= 2 \times 0.1587$$

$$= 0.3174$$



$$-3.08 = \frac{a-104}{5}$$

⇒ $$a = 88.6$$

Now $$3.08 = \frac{b-104}{5}$$

$\Rightarrow$ $$b = 11.94$$

Q. 6. (a) Find the approximate probability that a student can correctly guess (i) 12 cr more out of 20, (ii) at the most 24 out of 40, questions on a true-false examination. Under what circumstances is this approximation valid?

Ans. P (probability of correct answer) $= \frac{1}{2}$

Let X be the number of correct answers.

(i) $$n = 20, P = \frac{1}{2}$$

$$X \sim N\,\text{Bin}\left(20, \frac{1}{2}\right)$$

$$X \sim N(10, 5)$$

$$P(X > 12) = P\left(Z > \frac{12-10}{\sqrt{5}}\right)$$

$$= P\left(Z > \frac{2}{\sqrt{5}}\right)$$

$$= P(Z > 0.8944)$$
$$= 1 - 0.8133$$
$$= 0.1867$$

(ii) $$n = 40, P = \frac{1}{2}$$

$$P(X \leq 24) = P\left(Z \leq \frac{24-20}{\sqrt{5}}\right)$$

$$= P(Z < 1.79)$$

Conditions under which the approximation is valid :

(a) When $n\,p > 5$

(b) $n\,(1-p) > 5$

In both of the cases in the questions (a) and (b) conditions are met. Hence, our approximation of the normal distribution to the binomial is valid.

(b) A repair team is responsible for a stretch of oil pipeline 2 miles long. The distance in miles along this stretch at which any crack can arise is represented by a uniformly distributed random variable, with

$$pdf\ f(x) = \begin{cases} 0.5 & \text{for } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the cumulative distribution function and the probability that any given crack occurs between 0.5 mile and 1.5 miles along this stretch of pipeline.

Ans. Cumulative Distribution function

$$F(x) = P(X < x)$$

$$= \int_{y=0}^{y=x} 0.5\, dy$$

$$= 0.5\,(x - 0)$$

$$= 0.5\,x$$

$$P(0.5 < X < 1.5) = \int_{0.5}^{1.5} 0.5\, dx$$

$$= 0.5\,(1.5 - 0.5)$$

$$= 0.5$$

(c) A random variable has a normal distribution with standard deviation equal to 10. If the probability that the random variable will take on a value less than 82.5 is 0.8212, what is the probability that it will take on a value greater than 58.3? **(4+3+3)**

Ans. Let the random variable be X

$$X \sim N(\mu, 10^2)$$

$$P(X < 82.5) = 0.8212$$

$$P\left(Z < \frac{(82.5 - \mu)}{10}\right) = 0.8212$$

$$\Rightarrow \quad \frac{82.5 - \mu}{10} = 0.92$$

$$\Rightarrow \quad \mu = 82.5 - 9.2$$

$$\mu = 73.3$$

Now $$P(X > 58.3) = P\left(Z < \frac{58.3 - 73.3}{10}\right)$$

$$= P(Z > -1.5)$$

$$= P(Z < 1.5)$$

$$= 0.9332$$

SECTION III

Do any two out of Q. 7, Q. 8 and Q. 9.

Q. 7. (a) Given the value of the joint probability distribution of X and Y shown in the table

Y		X	
		−1	1
	−1	1/8	1/2
	0	0	1/4
	1	1/8	0

(i) Find the conditional distribution of X given $Y = -1$

(ii) Are X and Y independent?

Ans. (i) $$P(X/Y=-1) = \frac{P(X \cap Y=-1)}{P(Y=-1)}$$

Now $$P(Y=-1) = \frac{1}{8}+\frac{1}{2}$$

$$= \frac{5}{8}$$

$$P(X=-1/Y=-1) = \frac{P(X=-1 \cap Y=-1)}{P(Y=-1)}$$

$$= \frac{1/8}{5/8} = \frac{1}{5}$$

$$P(X=1/Y=-1) = \frac{P(X=1 \cap Y=-1)}{P(Y=-1)}$$

$$= \frac{1/2}{5/8} = \frac{4}{5}$$

$X/Y=-1$	-1	1
$P(X/Y=-1)$	$\frac{1}{5}$	$\frac{4}{5}$

(ii) $$P(X=-1 \cap Y=0) = 0$$

and $$P(X=-1) = \frac{2}{8}$$

and $$P(Y=0) = \frac{1}{4}$$

Hence, $P(X=-1).P(Y=0) = \frac{1}{16} \neq P(X=-1 \cap Y=0)$

and X and Y are not independent.

(b) Let X be a normally distributed random variable with mean 10 and variance. A random sample of size n is chosen from this distribution. Let X be the sample mean. The standard deviation of X is found to be 0.4. What is the sample size n? (4+3+3)

Ans. If
$$X \sim N(10, 4)$$

then
$$\bar{X} \sim N\left(10, \frac{4}{n}\right)$$

So,
$$\frac{4}{n} = 0.4^2$$

$$\Rightarrow \quad n = \frac{4}{0.16}$$

$$\Rightarrow \quad n = 25$$

(Please note that there is a misprint in the question paper. $\bar{X}$ is the sample mean and not X)

Q. 8. (a) If E(XY) = 3, E(X) = E(Y) = 2

(i) What is the covariance between X and Y?

(ii) What is the covariance between U and V where $U = 3X + 1$ and $V = 3 - \frac{5}{3}Y$?

(iii) How is Correlation between X and Y related to Correlation between U and V?

Ans. (i)
$$\begin{aligned} \text{Cov}(x, y) &= E(XY) - E(X)E(Y) \\ &= 3 - 2 \times 2 \\ &= -1 \end{aligned}$$

(ii)
$$\begin{aligned} \text{Cov}\left(3X + 1, 3 - \frac{5Y}{3}\right) &= \text{Cov}\left(3x, -\frac{5y}{3}\right) \\ &= 3X\left(\frac{-5}{3}\right)\text{cov}(X, Y) \\ &= -5\,x\,(-1) \\ &= 5 \end{aligned}$$

(iii)
$$\text{Corr}(X, Y) = \text{corr}(U, Y)$$

Q. 9. (a) Suppose X and Y are two discrete random variables which have the joint probability mass function $p(x,y) = (x-2y)/18$, $(x, y) = (1,1), (1,2), (2,1), (2,2)$, 0 elsewhere. Determine the conditional mean of Y, given $X=2$. Also find the two marginal probability mass functions. Calculate the value of $E(3X-2Y)$.

Ans.
$$P(x, y) = \frac{x + 2y}{18}$$

		Y: 1	Y: 2
X	1	$\frac{3}{18}$	$\frac{5}{18}$
	2	$\frac{4}{18}$	$\frac{6}{18}$

$$P(Y=1/X=2) = \frac{P(Y=1\cap X=2)}{P(X=2)} = \frac{\frac{4}{18}}{\frac{4}{18}+\frac{6}{18}} = \frac{4}{10}$$

$$P(Y=2/X=2) = \frac{P(Y=2\cap X=2)}{P(X=2)} = \frac{\frac{6}{18}}{\frac{4}{18}+\frac{6}{18}} = \frac{6}{10}$$

Hence, conditional mean of Y given $X = 2$

$$= 1\times\frac{4}{10}+2\times\frac{6}{10}$$

$$= \frac{16}{10}$$

Marginal distribution of X

$$P(X=1) = \frac{3}{18}+\frac{5}{18}$$

$$= \frac{8}{18}$$

$$P(X=2) = \frac{4}{18}+\frac{6}{18}$$

$$= \frac{10}{18}$$

Hence,

X	1	2
$P(X)$	$\frac{8}{18}$	$\frac{10}{18}$

Marginal distribution of Y

$$P(Y=1) = \frac{3}{18} + \frac{4}{18}$$

$$= \frac{7}{18}$$

$$P(Y=2) = \frac{5}{18} + \frac{6}{18}$$

$$= \frac{11}{18}$$

Hence,

Y	1	2
P(Y)	$\frac{7}{18}$	$\frac{11}{18}$

Now
$$E(3X-2Y) = 3E(X) - 2E(Y)$$

$$= 3\left[\frac{1\times 8}{18} + \frac{2\times 10}{18}\right] - 2\left[\frac{1\times 7}{18} + \frac{2\times 11}{18}\right]$$

$$= 3\left[\frac{28}{18}\right] - 2\left[\frac{29}{18}\right]$$

$$= \frac{26}{18}$$

SECTION IV

Do any two out of Q. 10. Q. 11 and Q. 12.

Q. 10. (a) The average signing bonus for 10 players in a local hockey team is found to be ₹ 65890 with standard deviation equal to ₹ 12,300. Assume the signing bonus for all players to be normally distributed.

(i) Find a 98% confidence interval for the true signing bonus of a player

(ii) How can the interval you found be made more reliable? Is there any loss associated with an increase in reliability? Explain.

Ans.
$$\bar{X} = 65890$$
$$S = 12{,}300$$
$$n = 10$$

(i) 90% confidence interval for true avrage signing bonus (μ)

$$\bar{X} \pm t_{0.01,9}\frac{S}{\sqrt{n}}$$

$$65890 \pm 2.89 \times \frac{12,300}{\sqrt{10}}$$

$$= (54649.05, 77130.95)$$

(ii) We can increase the reliability of a confidence interval by increasing the level of significance. This way, the width of the interval will decrease. This is associated with higher probability of type-I error.

(b) Assume a random sample ($X_1, X_2, X_3, ... X_n$) from a population with mean μ and standard deviation σ^2.

(i) Show that $\hat{\mu} = \frac{(1 * X_1 + 2 * X_2 + 3 * X_3 + ... + n * X_n)}{(0.5n\ (n+1))}$ **is unbiased.**

(ii) Show that variance of $\hat{\mu} = \frac{2(2n+1)\sigma^2}{3n(n+1)}$

Ans. (i)

$$E(\hat{\mu}) = E\left(\frac{1 \cdot X_1 2 \cdot X_2 + 3X_3 + ... + nX_n}{0.5n(n+1)}\right)$$

$$= \frac{1}{0.5n(n+1)}\left[E(X_1) + 2E(X_2) + 3E(X_3) + ... + nE(x_n)\right]$$

$$= \frac{1}{0.5n(n+1)}\left[\mu + 2\mu + 3\mu + ...n\mu\right]$$

$$= \frac{\mu(1+2+3+...+n)}{0.5n(n+1)}$$

$$= \frac{\mu \cdot \frac{n(n+1)}{2}}{0.5n(n+1)}$$

$$\Rightarrow \quad E(\hat{\mu}) = \mu$$

Hence, $\hat{\mu}$ is unbiased.

(ii)

$$\text{Var}(\hat{\mu}) = \text{Var}\left[\frac{1 \cdot X_1 2 \cdot X_2 + 3X_3 + ... + nX_n}{0.5n(n+1)}\right]$$

$$= \frac{1}{0.5^2 n^2 (n+1)^2}\left[\text{Var } X_1 2^2 \text{Var.}(X_2) + 3^2 \text{Var } (X_3) + ... + n^2 \text{Var}(X_n)\right]$$

(Assuming X is independent)

$$\Rightarrow \quad Var(\hat{\mu}) = \frac{1}{0.25n^2(n+1)^2}\left[\sigma^2 + 2^2\sigma^2 + 3^2\sigma^2 + ... + n^2\sigma^2\right]$$

$$= \frac{\sigma^2}{0.25n^2(n+1)^2}\left[1^2 + 2^2 + 3^2 + ... + n^2\right]$$

$$\Rightarrow \quad = \frac{4\sigma^2}{n^2(n+1)^2}\left[\frac{n(n+1)(2n+1)}{6}\right]$$

$$= \frac{2\sigma^2(2n+1)}{3n(n+1)}$$

(c) What is Mean Squared error (MSE) of an estimator θ'?

Show that MSE $(\hat{\theta})$ = Variance $(\hat{\theta})$ + (Bias $(\hat{\theta})$)^2

Ans. Mean squared Error:

The mean square error (MSE) of an estimator $(\hat{\theta})$ is $E_\theta\left(\left[\hat{\theta} - \theta\right]^2\right)$

$$\text{Now } E_\theta\left[\hat{\theta} - \theta\right]^2 = E_\theta\left[\left(\hat{\theta} - E_\theta(\hat{\theta})\right) + \left(E_\theta(\hat{\theta}) - \theta\right)\right]^2$$

$$= E_\theta\left[\left(\hat{\theta} - E_\theta(\hat{\theta})\right)^2 + \left(E_\theta(\hat{\theta}) - \theta\right)^2 + 2\left(\hat{\theta} - E_\theta(\hat{\theta})\right)\left(E_\theta(\hat{\theta}) - \theta\right)\right]$$

$$= E_\theta\left(\hat{\theta} - E_\theta(\hat{\theta})\right)^2 + E_\theta\left(E_\theta(\hat{\theta}) - \theta\right)^2 + 2E_\theta\left(\hat{\theta} - E_\theta(\hat{\theta})\right)E_\theta\left(E_\theta(\hat{\theta}) - \theta\right)$$

$$\Rightarrow \quad MSE(\hat{\theta}) = Var(\hat{\theta}) + (Bias)^2 + 0$$

$$\left\{\text{Since } E_\theta\left(\hat{\theta} - E_\theta(\hat{\theta})\right) = 0\right\}$$

Q. 11. (a) Assume that elasticity of a rubber pipe (measured in ksi units) is normally distributed with standard deviation equals to 0.75ksi.

(i) Find a 95% confidence intenal for true pipe elasticity if a sample of 20 pipes yields sample average of 4.56 ksi.

(ii) What is the sample size needed if we want to be 99% confident that true average elasticity is within 0.2 ksi of the sample mean?

Ans. Let x_i be the elasticity of a rubber pipe

(i) $$x_i \sim N(\mu, 0.75^2)$$

$$\bar{X} = 4.56, n = 20$$

95% confidence interval for μ

$$= \left(\bar{X} \pm Z_{0.025} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$\Rightarrow \quad = \left(4.56 \pm 1.96 \times \frac{0.75}{\sqrt{20}}\right)$$

$$\Rightarrow \quad = (4.23, 4.89)$$

(ii)

$$|\bar{X} - \mu| = 0.2$$

$$Z_{0.005} = 2.58$$

$$\sigma = 0.75$$

$$n = \frac{\sigma^2 z^2}{(\bar{X} - \mu)^2}$$

$$= \frac{0.75^2 \cdot (2.58)^2}{(0.2)^2}$$

$$n = 93.6 \approx 94$$

(b) Suppose true average runs scored by Team *A* and Team *B* are equal in a college, equal to . The variance in runs scored team *A* is σ^2 whereas it is $4\sigma^2$ for team *B*. Let $\bar{A}$ denote average runs from a sample of *M* games played by *A*, while *B* denote average runs from a sample of *N* games played by *B*. Let the estimator for μ be $\hat{\mu} = \alpha\bar{A} + \beta\bar{B}$

(i) Under what conditions is $\hat{\mu}$ unbiased,

(ii) What is the variance of this estimator?

Ans.

$$\hat{\mu} = \alpha\bar{A} + \beta\bar{B}$$

(i)

$$E(\hat{\mu}) = \alpha E(\bar{A}) + \beta E(\bar{B})$$

$$\Rightarrow \quad E(\hat{\mu}) = \alpha\mu + \beta\mu$$

$$\Rightarrow \quad E(\hat{\mu}) = (\alpha + \beta)\mu$$

For $\hat{\mu}$ to be unbiased, $\alpha + \beta = 1$

(ii)

$$\text{Var}(\hat{\mu}) = \alpha^2 \text{Var}(\bar{A}) + \beta^2 \text{Var}(\bar{B})$$

(Assuming the team A and Team B runs are independent of each other)

$$\Rightarrow \quad \text{Var}(\hat{\mu}) = \alpha^2(\sigma^2) + \beta^2(4\sigma^2)$$

$$= \sigma^2(\alpha^2 + 4\beta^2)$$

(c) Consider a random sample $(X_1, X_2, X_3, \ldots X_n)$ from a probability mass function $P(x; \theta) = x^{\theta}(\theta + 1)$ where $0 \leq x \leq 1$.

(i) Find an estimator for θ using method of moments.

(ii) Use the above estimator to obtain a point estimate for 0 when the sample obtained is (0.1, 0.3, 0.5). (4+3+3)

Ans. (i)

$$E(x) = \int_0^1 x \cdot x^{\theta}(\theta+1)dx$$

$$= (\theta+1)\int_0^1 x^{\theta+1}\, dx$$

$$\Rightarrow \quad = (\theta+1)\left[\frac{x^{\theta+2}}{\theta+2}\right]_0^1 \Rightarrow \frac{(\theta+1)}{(\theta+2)}$$

Using method of moments

$$\frac{\hat{\theta}+1}{\hat{\theta}+2} = \bar{X}$$

$$\Rightarrow \quad \hat{\theta}+1 = \bar{X}\hat{\theta}+2\bar{X}$$

$$\Rightarrow \quad \hat{\theta}(1-\bar{X}) = 2\bar{X}-1$$

$$\Rightarrow \quad \hat{\theta} = \frac{2\bar{X}-1}{1-\bar{X}}$$

(ii).

$$\bar{X} = \frac{0.1+0.3+0.5}{3}$$

$$\Rightarrow \quad \bar{X} = \frac{0.9}{3}$$

$$\Rightarrow \quad \bar{X} = 0.3$$

Hence

$$\bar{X} = \frac{2(0.3)-1}{1-0.3}$$

$$= \frac{0.6-1}{0.7}$$

$$= \frac{-4}{7}$$

Q. 12. (a) For each of the following confidence intervals for population mean drawn from samples from normally distributed

populations, find the confidence level, width and mention the distribution associated with the statistic used:

(i) $\left(\bar{X} - 1.4\frac{\sigma}{\sqrt{n}}, \bar{X} + 2.05\frac{\sigma}{\sqrt{n}}\right)$ **where** $n = 49$

(ii) $\left(\bar{X} - 2.069\frac{s}{\sqrt{n}}, \bar{X} + 2.807\frac{s}{\sqrt{n}}\right)$ **where** $n = 24$

Ans. (i) Confidence interval width

$$= 2.05\frac{\sigma}{\sqrt{n}} + 1.4\frac{\sigma}{\sqrt{n}}$$

$$= 2.45\frac{\sigma}{\sqrt{n}}$$

Since σ is known, we use z –statistic

(ii) Confidence interval width

$$= 2.807\frac{s}{\sqrt{n}} + 2.069\frac{s}{\sqrt{n}}$$

$$= 4.876\frac{s}{\sqrt{n}}$$

Since σ is unknown, we use t-statistic

(b) Consider a random sample $(X_1, X_2, X_3 \ldots, X_n)$ from a probability distribution function $f(x; \theta) = x^{(\theta - 1)}$ where $0 \leq x \leq 1$. Find a maximum likelihood estimator for θ.

Ans. Maximum likelihood function (L)

$$L(x_i, \theta) = \theta x_1^{(\theta-1)} . \theta x_2^{(\theta-1)} \ldots\ldots\ldots \theta x_n^{\theta-1}$$

$$L(x_i, \theta) = \theta^n \cdot (x_1 \cdot x_2 \ldots\ldots\ldots x_n)^{\theta-1}$$

Taking log on both sides:

$$\text{Log } L = n \log\theta + (\theta-1)\log(x_1 . x_2 \ldots . x_n)$$

$$\Rightarrow \quad \log L = n \log\theta + (\theta-1)\int_{i=1}^{n} \log x_i$$

Taking the first order derivative w.r.t.θ

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta} + (-1)\sum_{i=1}^{n} \log x_i = 0$$

$$\Rightarrow \quad \frac{n}{\hat{\theta}} = \sum_{i=1}^{n} \log x_i$$

$$\Rightarrow \quad \hat{\theta} = \frac{n}{\sum_{i=1}^{n} \log x_i}$$

(c) Let $\hat{\theta}$, be an estimator for a population with mean θ , and variance σ^2. If a random sample of size 3 is drawn, and $\hat{\theta} = \frac{3X_1 - 2X_2 - kX_3}{k}$ .What value of k will give bias ($\hat{\theta}$) = 2θ ? (4+3+3)

Ans.

$$\hat{\theta} = \frac{3X_1 - 2X_2 - kX_3}{k}$$

$$\Rightarrow \quad E(\hat{\theta}) = \frac{3E(X_1) - 2E(X_2) - kE(X_3)}{k}$$

$$\Rightarrow \quad E(\hat{\theta}) = \frac{3\theta - 2\theta - k\theta}{k}$$

$$\Rightarrow \quad E(\hat{\theta}) = \frac{\theta(1-k)}{k}$$

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$\Rightarrow \quad 2\theta = \frac{\theta(1-k)}{k} - \theta$$

$$\Rightarrow \quad 2 = \frac{1-k}{k} - 1$$

$$\Rightarrow \quad 2 = \frac{1-k-k}{k}$$

$$\Rightarrow \quad 2k = 1 - 2k$$

$$\Rightarrow \quad 4k = 1$$

$$\Rightarrow \quad k = \frac{1}{4}$$