Date _____ Topic Assignment No:- 1

Name - Aadarsh Kumar (L.E)
Roll No - 00676807223
Branch - C3E3

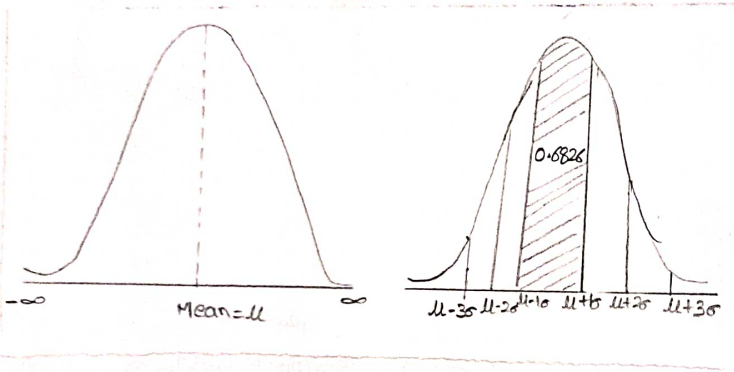
Q-1 :- Properties of Normal Distribution.

Normal distribution is a type of continous probability distribution that describes how random variable are distributed around their mean. It has the following properties

- The normal distribution curve is symmetric and bell-shaped, meaning that the left and right halves of the curve are Mirror images of each other.
- The mean, median, and mode of the normal distribution are equal and located at the center of the curve.
- The standard deviation of the normal distribution measure how spread out the date are from the mean. A smaller standard deviation mean that data are more clustered around the mean, while a larger standard deviation means the data are more dispersed.
- The total area under the normal distribution curve is equal to 1 or 100%. about 68% of the data are within one standard deviation of the mean, about 95% of the data are within two standard deviations of the mean, and about 99.7% of the data are within three standard deviations of the mean, and about 99.7% of the data are within three standard deviation of the mean. This is known as the empirical rule or the 68-95-99.7 rule.
- The normal distribution is defined by two parameter: The mean (μ) and the standard deviation (σ). The formula for the probability density function of the normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where x is the random variable, μ is the mean, and σ is the standard deviation.



Teacher's Signature _____

Date_____ Topic_____

Q-2 In a test on 2000 Electric bulbs. It was found that the life of a particular make was normally Distributed with an average life of 2040 hours and S.D. of 60 hours. Estimate the number of bulbs likely to burn for.

(i) More than 2150 hours.

(ii) less than 1950 hours.

(iii) More than 1920 hours and but less than 2160 hours.

Sl-2 Apply the formula for the standard score (Z-score):

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

where X is the raw score, μ is the mean, and σ is the standard deviation.

Using the given data, we can calculate the Z-scores of each of the cutoff points and then use a normal distribution table or a calculator to find the corresponding probabilities.

(i) For more than 2,150 hours, the Z-score is:

$$Z = \frac{2150 - 2040}{60} = 1.83$$

The probability of getting a Z-score less than or equal to 1.83 is 0.9664. Therefore the prob. of getting a Z-score grater than 1.83 is $1 - 0.9664 = 0.0336$. This means that about 3.36% of the bulbs are likely to burn for more than 2,150 hours. since there are 2,000 bulbs in total, we can estimate the number of bulbs by multiplying the prob. by the total number:- $0.0336 \times 2000 = 67.2$

There for, about 67 bulbs are likely to burn for more than 2,150 hours

(ii) For less than 1,950 hours, the Z-score is.

$$Z = \frac{1950 - 2040}{60} = -1.5$$

The probability of getting a Z-score less than or equal to -1.5 is

 Teacher's Signature_____

Date _____ Topic _____

This means that about 6.68% of the bulbs are likely to burn for less than 1,950 hours. Again, we can estimate the number of bulbs by multiplying the prob. by the total number:

$$0.0668 \times 2000 = \boxed{133.6}$$

about 134 bulbs are likely to burn for less than 1,950 hours

(iii) For more than 1,920 hours but less than 2,160 hours, the Z-score are:-

$$Z_1 = \frac{1920 - 2040}{60} = \boxed{-2}$$

$$Z_2 = \frac{2160 - 2040}{60} = \boxed{2}$$

The prob. of getting a Z-score between −2 and 2 is 0.9545. This means that about 95.45% of the bulbs are likely to burn for more than 1,920 hours but less than 2,160 hours.

$$0.9545 \times 2000 = \boxed{1909}$$

There fore, the about 1909 bulbs are likely to burn for more the 1,920 hours but less than 3,160 hours.

Q3 Let the Joint P.d.f of x and Y be

$$f(x,y) = \begin{cases} (x+y) & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Sol-3

Find $P[0 < x < \frac{1}{2},\ 0 < y < \frac{1}{4}]$, $E[X]$, $E[Y]$, $E[X4]$, $E[X+Y]$, $P[x,y]$

Teacher's Signature _____

Date _____ Topic _____

$$P[0<x<\tfrac{1}{2}, 0<y<\tfrac{1}{4}] = \int_0^{1/2}\left[\int_0^{1/4}(x+y)\,dy\right]dx = \int_0^{1/2}\left[xy+\tfrac{1}{2}y^2\right]_0^{1/4}dx$$
$$= \int_0^{1/2}\left(\tfrac{1}{4}x+\tfrac{1}{32}\right)dx = \tfrac{1}{4}\times\tfrac{1}{8}+\tfrac{1}{64} = \tfrac{3}{64}$$

Now $E[x] = \int_0^1\int_0^1 x f(x,y)\,dx\,dy = \int_0^1\int_0^1 x(x+y)\,dx\,dy$

$$= \int_0^1\left[x^2y+x\left(\tfrac{1}{2}y^2\right)\right]_0^1 dx = \int_0^1\left(x^2+\tfrac{1}{2}x\right)dx$$
$$= \tfrac{1}{3}+\tfrac{1}{4} = \tfrac{7}{12}$$

Similarly, $E[y] = \int_0^1\int_0^1 y f(x,y)\,dx\,dy = \tfrac{7}{12}$

Now $E[xy] = \int_0^1\int_0^1 xy\,f(x,y)\,dx\,dy \int_0^1\int_0^1 xy(x+y)\,dx\,dy$

$$= \int_0^1\left[x^2\left(\tfrac{1}{2}y^2\right)+x\left[\tfrac{1}{3}y^3\right)\right]_0^1 dx = \int_0^1\left(\tfrac{1}{2}x^2+\tfrac{1}{3}x\right)dx$$
$$= \tfrac{1}{6}+\tfrac{1}{6} = \tfrac{1}{3}$$

$$E[x+y] = \int_0^1\int_0^1 (x+y)\,f(x,y)\,dx\,dy = \int_0^1\int_0^1 (x+y)^2\,dx\,dy$$

$$= \int_0^1\int_0^1 (x^2+2xy+y^2)\,dx\,dy$$

$$= \int_0^1\left[x^2y+xy^2+\left(\tfrac{1}{3}y^3\right)\right]_0^1 dx$$

$$= \int_0^1\left(x^2+x+\tfrac{1}{3}\right)dx = \tfrac{1}{3}+\tfrac{1}{2}+\tfrac{1}{3} = \tfrac{7}{6}$$

We have $\operatorname{cov}[x,y] = E[xy]-E[x]E[y] = \tfrac{1}{3}-\left(\tfrac{7}{12}\right)^2 = -\tfrac{1}{144}$

Now $E[x^2] = \int_0^1\int_0^1 x^2 f(x,y)\,dx\,dy = \int_0^1\int_0^1 x^2(x+y)\,dx\,dy$

$$= \int_0^1\left[x^3y+x^2\left(\tfrac{1}{2}y^2\right)\right]_0^1 dx$$

$$= \int_0^1\left(x^3+\tfrac{1}{2}x^2\right)dx = \tfrac{1}{4}+\tfrac{1}{6} = \tfrac{5}{12}$$

Teacher's Signature _____

Date _____ Topic _____

similarly, $E[y^2] = \frac{5}{12}$

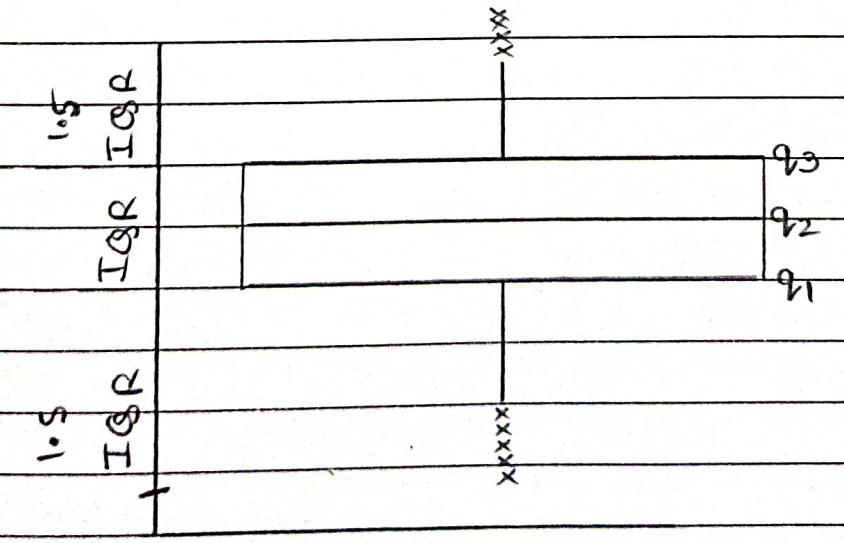
Now $var[x] = E[x^2] - (E(x))^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$

similarly, $var[y] = \frac{11}{144}$

Hence, $P[x,y) = \frac{conv[x,y]}{\sqrt{var[x]\,var[y]}} = \frac{\frac{-1}{144}}{\sqrt{\frac{11}{144} \times \frac{11}{144}}} = \frac{-1}{11}$

-4 BOX Plots

L For a given data set, we determine all the quarties q_1, q_2, q_3 the difference $q_3 - q_2$. (upper quartite lower quartile) is determined and is called IQR (Inner Quartile Range).
we draw a diagram with the data on the y-axis (in the increasing orders) we draw a box whose lower side is q_1 & upper side is q_3 and a middle line q_2.
ignoring the with we drown two straight lines above & below the box for the date within the 1.5 IQR range
The data beyond this is denoted as signal point and the colled outliners.

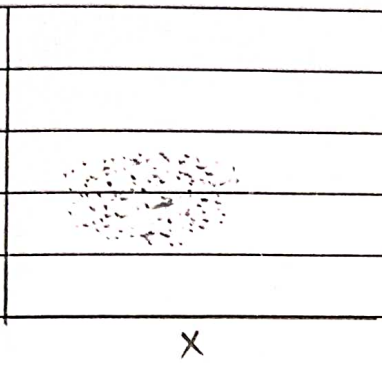


 Teacher's Signature _____

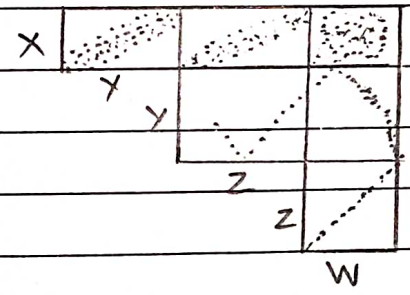
Date _____ Topic _____

Q-5 Scatter Diagrams

Sol-5 A scatter diagram is conducted by plotting each pairs of observations with one measurement in the pair on the vertical axis of the graph and the other measurement in the pair on the horizontal axis.



when two or more variables exists, the matrix of scatter diagram may be useful.



* Sample correlation coefficient.

$$P_{xy} = \frac{\sum \Delta x_i \Delta y_i}{\sqrt{\sum \Delta x_i \sum \Delta y^2 e}} \qquad \Delta x_i = x_i - \bar{x}, \quad \Delta y_i = y_i - \bar{y}$$

Q6 The mean of a certain normal population is Equal to the standard error of the mean of the standared sample of 100 form that distribution. Find the prob. that the mean of the sample of 25 from the distribution will be negative.

Sol-6 we need to use the mean of ~~the~~ s formula for the standard error of the mean:

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

Teacher's Signature _____

Date _____ Topic _____

where σ is the standard deviation and n is the sample size
∴ The standard deviation error of the mean is $\frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$
Given mean is equal to the standard error of the mean, therefore $\mu = \frac{\sigma}{10}$ —(1)

The z-test statistical formula is given by $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

using equation(1) in z-test statistical formula for the sample size 25,

$$Z = \frac{\bar{x} - \frac{\sigma}{10}}{\frac{\sigma}{\sqrt{25}}} = \frac{10\bar{x} - \sigma}{10} \times \frac{5}{\sigma} = \frac{(10\bar{x} - \sigma)5}{10\sigma} = \frac{50\bar{x}}{10\sigma} - \frac{\sigma}{10}$$

$$\boxed{Z = \frac{5\bar{x}}{\sigma} - 0.5}$$

Since, the population is normally distributed, and if $Z < 0.5$, $\bar{x}$ is negative. therefore the probability is given by
$P(Z < 0.5) = P(-\infty < Z < 0.5)$
Using the table of z-values,
$P(Z < 0.5) = 0.5 + P(0 < Z < 0.5)$

Q-7 Statistical Interval for a single sample
Sol-7 Statistical intervals provide a range of values within which a population parameter is estimated to lie with a certain level of confidence.
- for a single sample, intervals are constructed to estimated population parameters such as the mean, proportion, or variance.
- Type of Intervals.
① Confidence Intervals (CI) :- Estimate the range within which the population parameter is likely to fall.
mean :- $CI = \bar{x} \pm z^* (\sigma/\sqrt{n})$, proportion $CI = p \pm z^* \sqrt{p^*(1-p)}$

 Teacher's Signature _____

Date _____ Topic _____

Variance : $(I = [(n-1)*s^2 / \chi^2_{upper}, (n-1)*s^2 / \chi^2_{Lower}]$

2) prediction Intervals (PI) :- provide a range within which the population parameter value of a future observation

mean (μ) :- $PI = \bar{x} + (*(s*)(1+1/n))$

Individual (c) :- $\bar{x} + t*(s*)$

Teacher's Signature _____

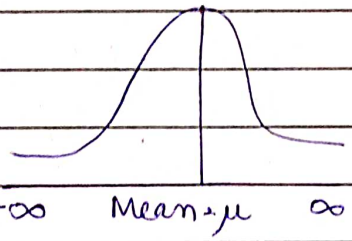
Topic.....Date.....

Quest Properties of Normal Distribution

(1) The normal probability curve with mean μ and standard deviation σ is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \leq X \leq \infty$$

(2) The graph of the curve is



(3) The curve is bell-shaped and symmetrical about the line $x = \mu$

(4) Mean, median and mode of the normal distribution coincide and the normal distribution is unimodal.

(5) $f(x)$ decreases rapidly as x increases.

(6) X-axis is an asymptote to the curve (the tangent to the curve at ∞ is X-axis)

(7) The maximum probability occurs at the point $x = \mu$ and is $\frac{1}{\sigma\sqrt{2\pi}}$

Teacher's Sign.....

Topic.....Date.....

(8) Mean deviation about mean $= \frac{4}{5}\sigma$

(9) Since $f(x)$, being the probability, can never be negative, so that no portion of the curve lies below the X-axis.

(10) A linear function of independent normal variates is also normal variate.

Ques 2 In normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

$z_1 = 0.31$

$z_2 = 0.92$ ($1 - 0.08 = 0.92$, since 8% are over 64, 92% are below 64)

$x_1 = 45^{th}$ percentile

$x_2 = 64^{th}$ percentile

From the standard normal distribution table

$z_1 \simeq -0.48$

$z_2 \simeq 1.41$

$\Rightarrow z = \frac{x - \mu}{\sigma}$

$x = \mu + z \times \sigma$

$\Rightarrow x_1 = \mu + (-0.48) \times \sigma$

$\Rightarrow x_2 = \mu + (1.41) \times \sigma$

$x_1 = 45, \; x_2 = 64$

Teacher's Sign.....

Topic.....Date.....

$45 = \mu - 0.48 \times \sigma \rightarrow (1)$

$64 = \mu + 1.41 \times \sigma \rightarrow (2)$

$64 = 45 + 0.48 \times \sigma + 1.41 \times \sigma$

$64 = 45 + 1.89 \times \sigma$

$19 = 1.89 \times \sigma$

$\boxed{\sigma \approx 10.05}$ Ans

$\mu = 45 + 0.48 \times 10.05$

$\boxed{\mu \approx 49.82}$ Ans

Ques3 X and Y are two random variable having joint density function $= \frac{1}{27}(2x+y)$, where x and y can assume only integer values 0, 1, 2. Find conditional distribution of y for $X = x$

To find the conditional distribution of Y given $X = x$, we need to calculate PMF of Y when $X = x$, denoted as $P(Y|X=x)$.

$$P(Y|X=x) = \frac{P(X=x, Y)}{P(X=x)}$$

$$f(x,y) = \frac{1}{27}(2x+y)$$

$$P(X=0) = \sum_{y=0}^{2} f(0,y) = \sum_{y=0}^{2} \frac{1}{27}(2(0)+y) = \frac{1}{27}(0+1+2) = \frac{3}{27}$$

$$P(X=1) = \sum_{y=0}^{2} f(1,y) = \sum_{y=0}^{2} \frac{1}{27}(2(1)+y) = \frac{1}{27}(2+3+4) = \frac{9}{27}$$

$$P(X=2) = \sum_{y=0}^{2} f(2,y) = \sum_{y=0}^{2} \frac{1}{27}(2(2)+y) = \frac{1}{27}(4+5+6) = \frac{15}{27}$$

Teacher's Sign.....

Topic.....Date.....

Now let's calculate $P(X=n, Y)$ for $X=0,1,2$;

- For $X=0$:
$$P(X=0, Y) = \frac{1}{27}(2(0)+Y) = \frac{1}{27}Y$$

- For $X=1$:
$$P(X=1, Y) = \frac{1}{27}(2(1)+Y) = \frac{1}{27}(2+Y)$$

- For $X=2$:
$$P(X=2, Y) = \frac{1}{27}(2(2)+Y) = \frac{1}{27}(4+Y)$$

$$\Rightarrow P(Y|X=0) = \frac{P(X=0,Y)}{P(X=0)} = \frac{1/27\, Y}{3/27} = \frac{Y}{3}$$

$$P(Y|X=1) = \frac{P(X=1,Y)}{P(X=1)} = \frac{1/27(2+Y)}{9/27} = \frac{2+Y}{9}$$

$$P(Y|X=2) = \frac{P(X=2,Y)}{P(X=2)} = \frac{1/27(4+Y)}{15/27} = \frac{4+Y}{15}$$

So, the conditional distribution of Y given $X=n$ is

- $P(Y|X=0) = \frac{Y}{3}$ for $Y=0,1,2$
- $P(Y|X=1) = \frac{2+Y}{9}$ for $Y=0,1,2$
- $P(Y|X=2) = \frac{Y+4}{15}$ for $Y=0,1,2$

Ques4 **Stem and Leaf Diagram**

A stem and leaf diagram is a simple and effective method for displaying a set of quantitative data. It allows for a quick visual representation of the distribution of the data while retaining the individual datapoints.

Teacher's Sign.....

Topic.....Date.....

(1) Stem: The stem consists of the leading digits of the data values. It represents the larger part of the number. For example, if the data values are 23, 27, 34, 36 the stems are 2, 2, 3, 3.

(2) Leaf :- The leaf consists of the last digit of each data value. It represents the finer part of the number. Using the same example data, the leaves would be 3, 7, 4, 6.

(3) Organization :- The stems are written vertically in ascending order on the left side of a vertical line, and the corresponding leaves are placed to the right of each stem. Leaves are typically arranged in ascending order with each stem.

(4) Example :- Using the data above, a stem and leaf plot might look like this:

2 | 3 7
3 | 4 6

This indicates that there are two data points in the 20s range (23 and 27) and two data points in the 30s range (34 and 36).

Ques5 Time Sequence Plots

A time sequence plot, known as a time series plot, is a graphical representation of data points collected or recorded over time. It's a fundamental tool in analyzing time based data in various fields such as economics, finance, weather forecasting etc.

(1) Representation of Time :- The horizontal axis of a time sequence plot represents time, typically with equally spaced

Teacher's Sign.....

Topic.....Date.....

intervals. Time can be measured in various units depending on the nature of the data, such as minutes, hours, days, months or years

(2) Vertical Axis :- The vertical axis represents the values of the data being measured or observed over time. This could include variables like temperature, stock prices, sales figures.

(3) Plotting Data Points :- Each data point is plotted on the graph at its corresponding time and value. Connecting these points with lines helps visualize trends and patterns in the data over time.

Ques6 What is point estimation?

Point estimation is a method used in statistics to estimate an unknown population parameter based on sample data. It involves using a single value, called a point estimator, to approximate the true value of the parameter of interest. The goal of point estimation is to provide a best guess or approximation of the parameter, given the available sample information.

Ques7 Central Limit Theorem

The central limit theorem (CLT) is one of the most important and fundamental concepts in statistics. It states that regardless of the distribution of the population from which a sample is drawn, the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, under certain conditions.

↳ The Central Limit Theorem holds under the following

Teacher's Sign.....

Topic.....Date.....

conditions:-

- The sample is drawn from a population with a finite mean (μ) and a finite standard deviation (σ)
- The sample size is sufficiently large. While there is no strict rule for what constitutes "sufficiently large", a commonly used guideline is that the sample size should be atleast 30. However, the CLT often starts to apply even for smaller sample sizes if the population distribution is not heavily skewed.

Teacher's Sign.....

Title _____ Date _____

_____ Page No. _____

Assignment

Q1 X is a normal variate with mean 30 and standard deviation 50, Find the probability that

(i) $26 \leq x \leq 40$

(ii) $x \geq 45$

(iii) $|x - 30| > 5$

Ans- Now, to find the probability the basic formula used is -

$$z = \frac{x - \mu}{\sigma}$$

where, μ is the mean and

σ is the standard deviation

(i) For $x = 26$:

$$z_1 = \frac{26 - 30}{50} = -0.08$$

For $x = 40$:

$$z_2 = \frac{40 - 30}{50} = 0.20$$

$$P(-0.08 \leq z \leq 0.20) = P(z \leq 0.20) - P(z \leq -0.08)$$

$P(-0.08 \leq z \leq 0.20)$ is approximately 0.4515.

(ii) Probability that $x \geq 45$:

here, $x = 45$:

$$z = \frac{45 - 30}{50} = 0.30$$

$P(z \geq 0.30)$ is approximately 0.3821.

DELTA®

Title _____ Date _____

Page No. _____

(iii) Probability that $|x-30| \geq 5$:

Here, we standardize $x = 25$ & $x = 35$:

$$z_1 = \frac{25-30}{50} = -0.10$$

$$z_2 = \frac{35-30}{50} = 0.10$$

$$P(z < -0.10) + P(z > 0.10)$$

$$= 0.4602 + 0.4602 = 0.9204$$

So, the probabilities obtained are -

$P(26 \leq x \leq 40) \approx 0.4515$ and $P(x \geq 45) \approx 0.3821$

$P(|x-30| > 5) \approx 0.9204$

Q2 In a referendum 60% of voters voted in favour. A random sample of 200 voters was selected what is the probability that in the sample -

(i) More than 130 voted in favour?

(ii) Between 105 & 130 inclusive voted in favour?

(iii) 120 voted in favour?

Ans- $P(x = K) = \binom{n}{K} \times p^K \times (1-p)^{n-K}$

- $P(x=K)$ is the probability of getting exactly K successes in n trials.
- $\binom{n}{K}$ is the number of combinations of n items taken K at a time.
- p is the probability of success in one trial.
- n is the no. of trials.

Also, $p = 0.60$ (probability of voting in favour)

$n = 200$

Title _____ Date _____

_____ Page No. _____

(i) Probability that more than 130 voted in favour:

$P(x > 130) = 1 - P(x \leq 130)$

Using the binomial probability formula:

$P(x \leq 130) = \sum_{x=0}^{130} \binom{200}{x} (0.60)^x (0.40)^{200-x}$

Calculating the sum:

$P(x \leq 130) \approx 0.0048$

$P(x > 130) = 1 - 0.0048 = 0.9952$

(ii) Probability that between 105 and 130 inclusive voted in favour -

$P(105 \leq x \leq 130) = \sum_{x=105}^{130} \binom{200}{x} (0.60)^x (0.40)^{200-x}$

calculating the sum -

$P(105 \leq x \leq 130) \approx 0.9609$

(iii) Probability that exactly 120 voted in favour:

$P(x = 120) = \binom{200}{120} (0.60)^{120} (0.40)^{80}$

Calculating the probability -

$P(x = 120) \approx 0.0650$

Q3 Frequency distribution and Histograms

Ans- Frequency Distribution → The frequency of a value is the number of times it occurs in a dataset. A frequency distribution is the pattern of frequencies of a variable. It's the no. of times each possible value of a variable occurs in a dataset.

There are different types of frequency distributions some of them are listed below -

(i) Grouped Frequency Distribution ⟹

The number of observations of each class interval of a variable class intervals are ordered groupings of a variables values.

Title _____ Date _____

Page No. _____

We can also use this type of frequency distribution for quantita variables.

Relative frequency distribution → The number of observations each class or the proportion of observation of each value or class interval of a variables.
We can use this type of frequency distribution for any type of variable when we are more interested in comparing frequencies tha the actual no. of observations.

Ungrouped frequency distribution → The number of observations of each value of a variable. We can use this type of frequency distribution for categorial variables.

Cummulative frequency distribution → The sum of the frequencies les than or equal to each value or class interval of a variable.
We can use this type of variable or frequency distribution for ordinal or quantitative variables when you want to understand how often observations fall below certain values.

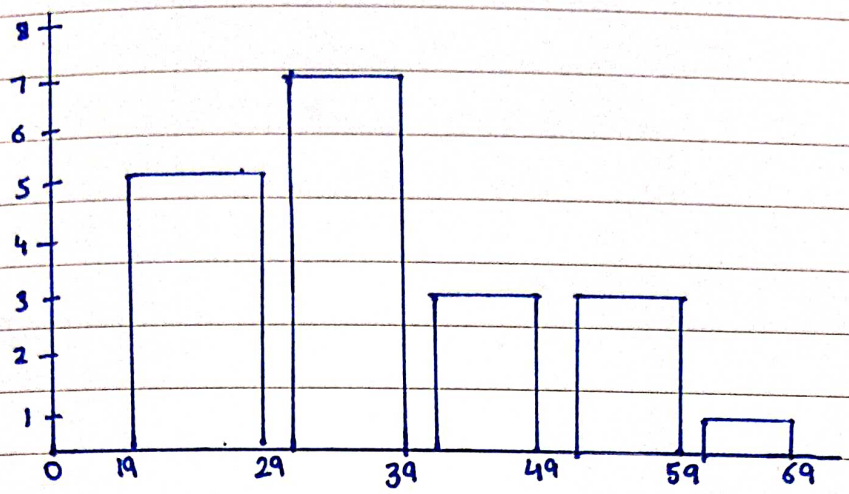
Histogram → A histogram is a graph that shows the frequency or relative frequency distribution of a quantitative variable. It look similar to a bar chart. The continuous variable is grouped into interval classes, just like a grouped frequency table. The y-axis of the bars shows the frequencies or relative frequencies, and the x-axis shows the interval classes. Each interval class is represented by a bar and the height of the bar shows the frequency or relative frequency of the interval class.

DELTA®

Title _____ Date _____

_____ Page No. _____

A histogram is an effective visual summary of several important characteristics of a variable.



Q4 Probability Plots

Ans- The probability plot is a way to visually comparing the data coming from different distributions. These data can be of empirical dataset or theoritical dataset. The probability plots can be of two types-

- P-P Plot → The p-p plot is the way to visualize the comparing of cumulative distribution function (CDF's) of the two distributions against each other.

- Q-Q Plot → The q-q plot is used to compare the quantities of two distributions. The quantities can be defined as continuous intervals with equal probabilities or dividing the samples between a similar way. The distributions maybe theoretical or sample distributions from a process etc. The normal probability plot is a case of the q-q plot.

Normal Probability plot → The normal probability plot is a way of knowing whether the dataset is normally distributed or not. In this plot, data is plotted against the theoretical normal distribution plot in a way such that if a given dataset is

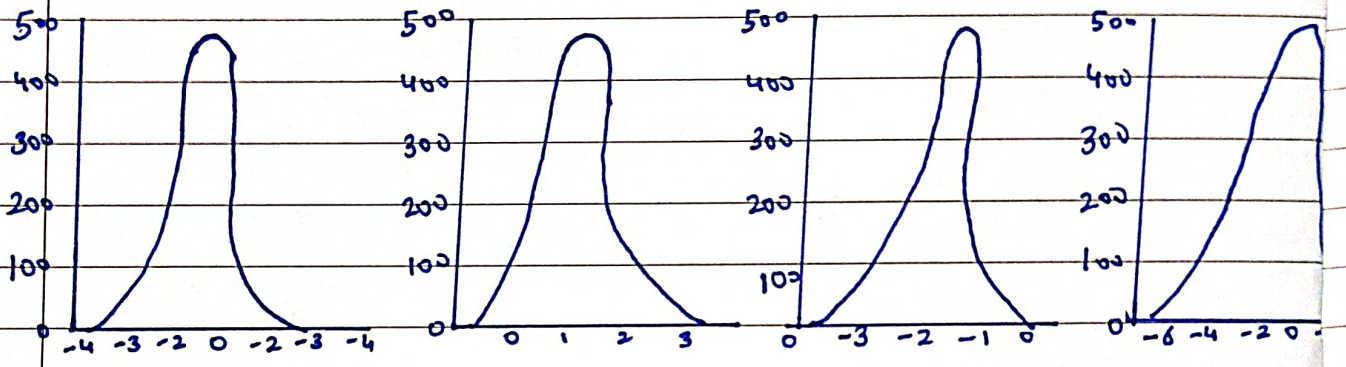
DELTA®

Title _____ Date _____

_____ Page No. _____

normally distributed it should form an approximate straight line. The normal probability case is a case of the probability plot. The normal probability plot has the following axis -

- Horizontal axis → Normal-order statistic medians.
- Vertical axis → order response values.



Standard Right skewed left skewed Heavy-tai

Q5 General concept of point estimation.

Ans- A point estimate is defined as a calculation where a sample statistic is used to estimate or approximate an unknown population parameter.

Point estimators are defined as functions that can be used to find the approximate value of a particular point from a given population parameter. The sample data of a population is used to find the point estimate or a statistic that can act as the best estimate of an unknown parameter that is given for a population.

Properties of point estimators →

- consistent
- unbiased
- most efficient

DELTA®

Methods used to calculate point estimators -

Title _____ Date _____

Page No. _____

The maximum likelihood method is a popularly used way to calculate point estimators. This method uses differential calculus to understand the probability function from a given no. of sample parameters.

Formulae that can be used to measure point estimators -
- Maximum likelihood estimation
- Jeffrey estimation
- Wilson estimation
- Laplace estimation

Point estimate for p statistics →
$p = x/n$, the proportion of succes in the sample, to be the point estimate of p.

point estimation of the mean →
It is determined by calculating the mean of a sample drawn from the population.

Q6 Sampling Distribution

Ans- A sampling distribution refers to the distribution of a statistic such as the mean, standard deviation, or proportion, calculated from multiple samples drawn from the same population. It provides insights into the behaviour & variability of the statistic under repeated sampling.

Example → Consider a population of heights of all adult males in a city. We want to estimate the average height of adult males in the city.

DELTA®

Title _____ Date _____

Page No. _____

(i) Sampling → We take multiple random samples of, say, 50 to adults males each from the population.

(ii) Population parameter → let, μ be the population mean height of adult males.

(iii) Sample statistic → For each sample, we calculate the sample mean height, denoted as $\bar{x}$.

(iv) Sampling Distribution → The distribution of these sample means ($\bar{x}$) across all samples forms the sampling distribution of the sample mean.

Q7 Central limit Theorem

Ans- The central limit theorem relies on the concept of a sampling distribution, which is the probability distribution of a statistic for a large no. of samples taken from a population.

The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough. Regardless of whether the population has a normal, poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

Formula → mean - $\mu_{\bar{x}} = \mu$

Standard deviation - $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Title _____ Date _____

_____ Page No. _____

where, $\bar{X}$ is the sampling distribution of the sample means.

$\sim$ means " follow the distribution "

μ is the mean of that population

σ is the standard deviation of the population

n is the sample size

Conditions of the central limit theorem →

(i) The sample size is sufficiently large. This condition is usually met if the sample size is $n \geq 30$.

(ii) The samples are independent & identically distributed random variables. This condition is usually met if the sampling is random.

(iii) The population's distribution has finite variance. Central limit theorem doesn't apply to distributions with infinite variance, such as the cauchy distribution. Most distribution have finite variance.