Haider Ali (100)

Probability & Statistics

Question-1: Short Questions (2012)

(i) Give two examples of permutation and combination form real life also mention mathematical formulae.

Two examples of permutation and combination from real life and their formulas are as follows:

Permutation:-

An arrangement of data/objects in order is called permutation.

Formula:-

$$^nP_r = \frac{n!}{(n-r)!}$$

Example:-

You have a social website account and this account has a password. Your account always opens with the arrangement of characters you set. If you use characters without order your account did not open. This is an example of permutation.

Combination:-

An arrangement of data/objects without regard to order is called combination.

Formula:-

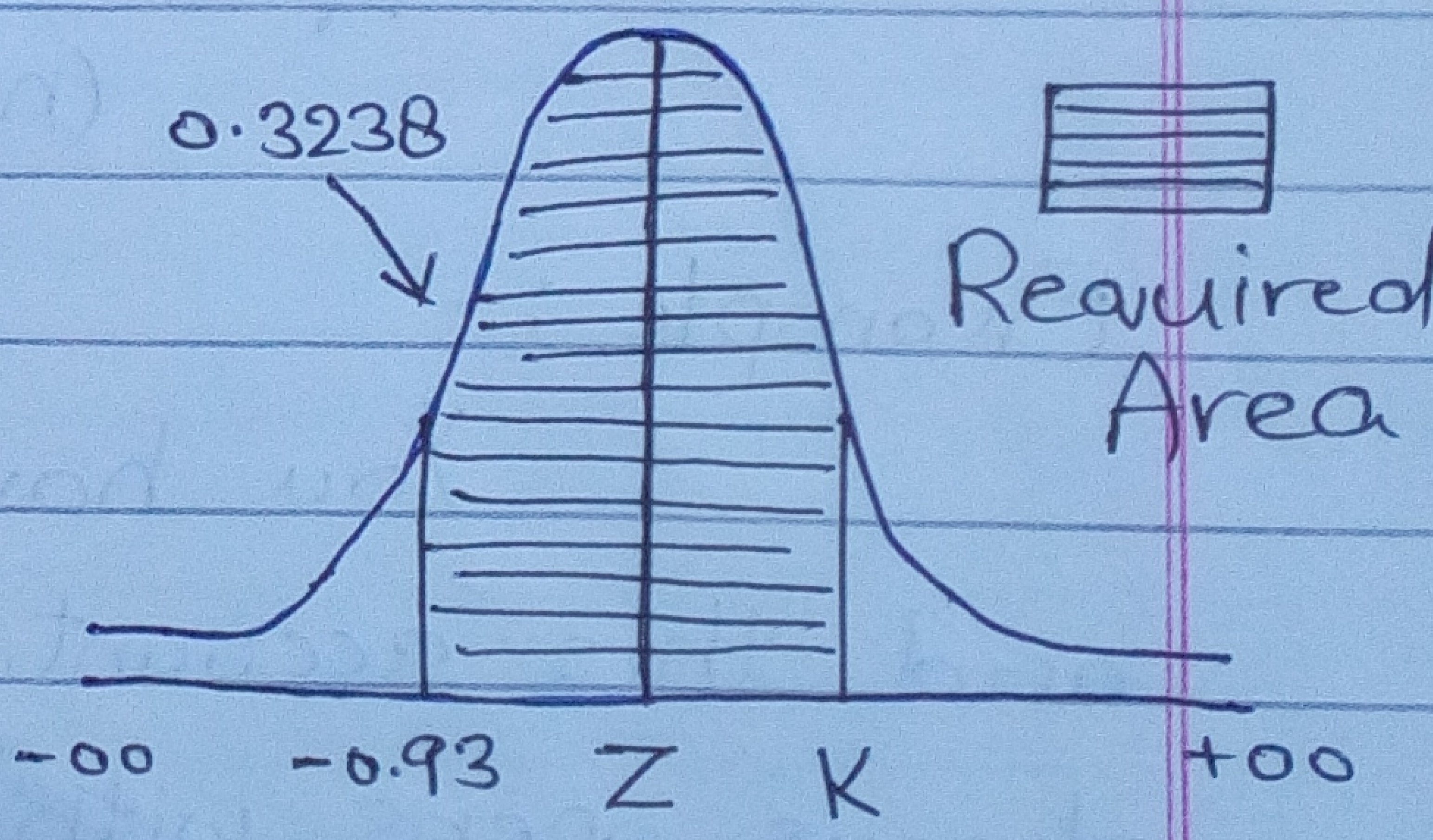
$$^nC_r = \frac{n!}{r!(n-r)!}$$

Example:-

If you have three fruits apple, mango and banana in breakfast you can eat either apple first, mango first or banana first. Their is no restriction of order so it is an example of combination.

(ii) Give a standard normal distribution. Find value of k such that

$$P(-0.93 < Z < k) = 0.7235$$



$$\text{table}(-0.93) + \text{table}(K) = 0.7235$$
$$0.3238 + \text{table}(K) = 0.7235$$
$$\text{table}(K) = 0.7235 - 0.3238$$
$$\text{table}(K) = 0.3997$$
$$K = 1.28$$

(iii) **Differentiate Qualitative and Quantitative data with examples.**

Qualitative Data:-

Qualitative data is a data that cannot be measured numerically rather it contains non-numeric quantities.
e.g; education, gender, eye colour, quality, satisfaction, intelligence etc.

Quantitative Data:-

Quantitative data is a data that can ~~not~~ be measured numerically.
e.g; height, weight, income or no of children etc.

(iv) **What will be the difference b/w Continous and Discrete Data give examples as well.**

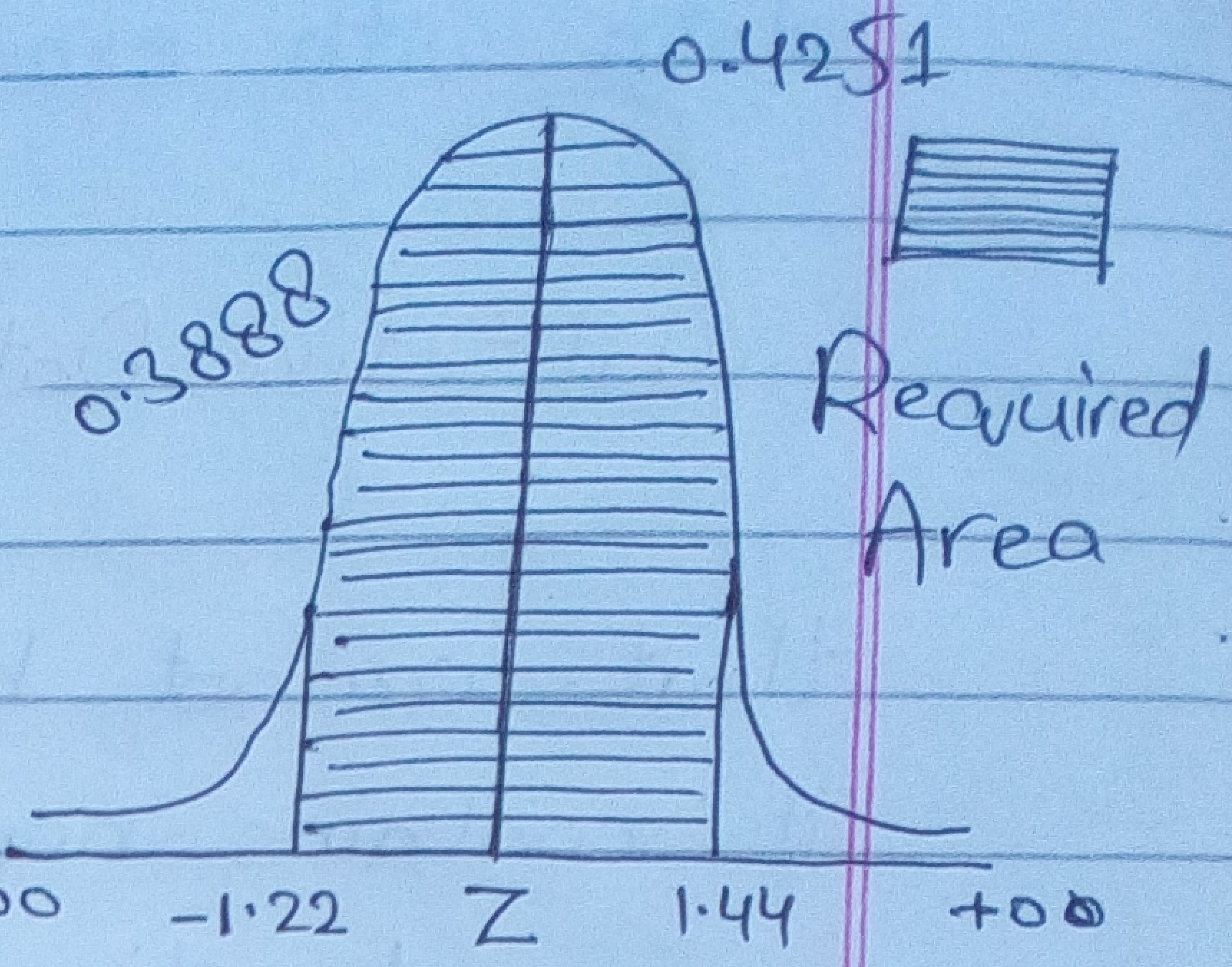
Continous Data:-

Continous data is a data that is in the form of fraction or integer within a given interval. e.g; 1, 1/5 etc.

Discrete Data:-

Discrete data is a data that is in the form of discrete set of integers or whole no. e.g; 1, 10 etc.

(v) Find the area b/w Z-scores -1.22 and 1.44



$P(-1.22 \leq Z \leq 1.44) = ?$

Required area is

$P(-1.22 \leq Z \leq 1.44) = 0.3888 + 0.4251$

$= 0.8139$

(vi) An experiment consists of tossing a die and and then flipping a coin once if the number on the die is even. If the number on the die is odd, the coin is flipped twice using appropriate notation to construct a Tree diagram.

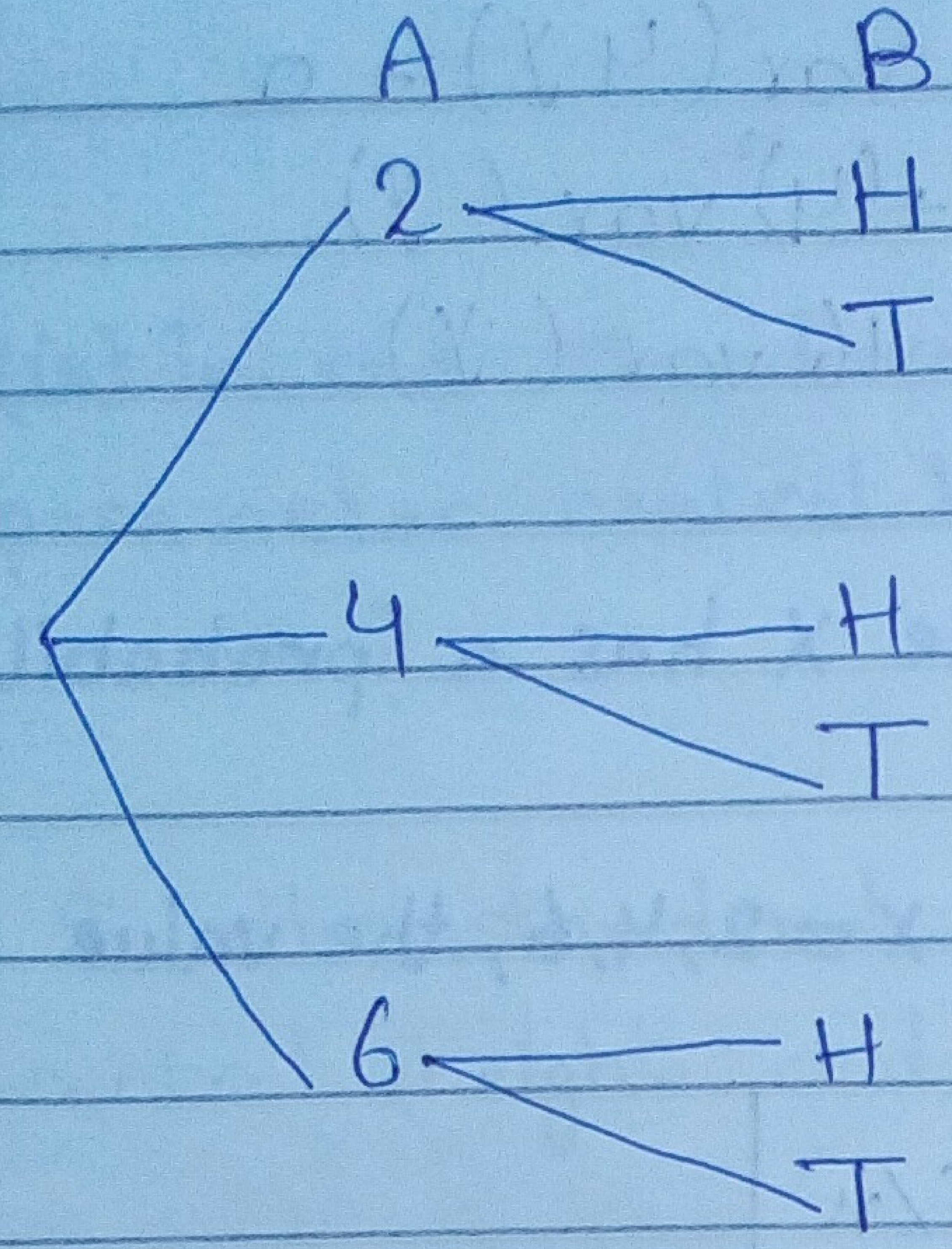
Let A be the sample space of tossing a die.

$A = \{1, 2, 3, 4, 5, 6\}$

Let B be the sample space of flipping a coin.

$B = \{H, T\}$

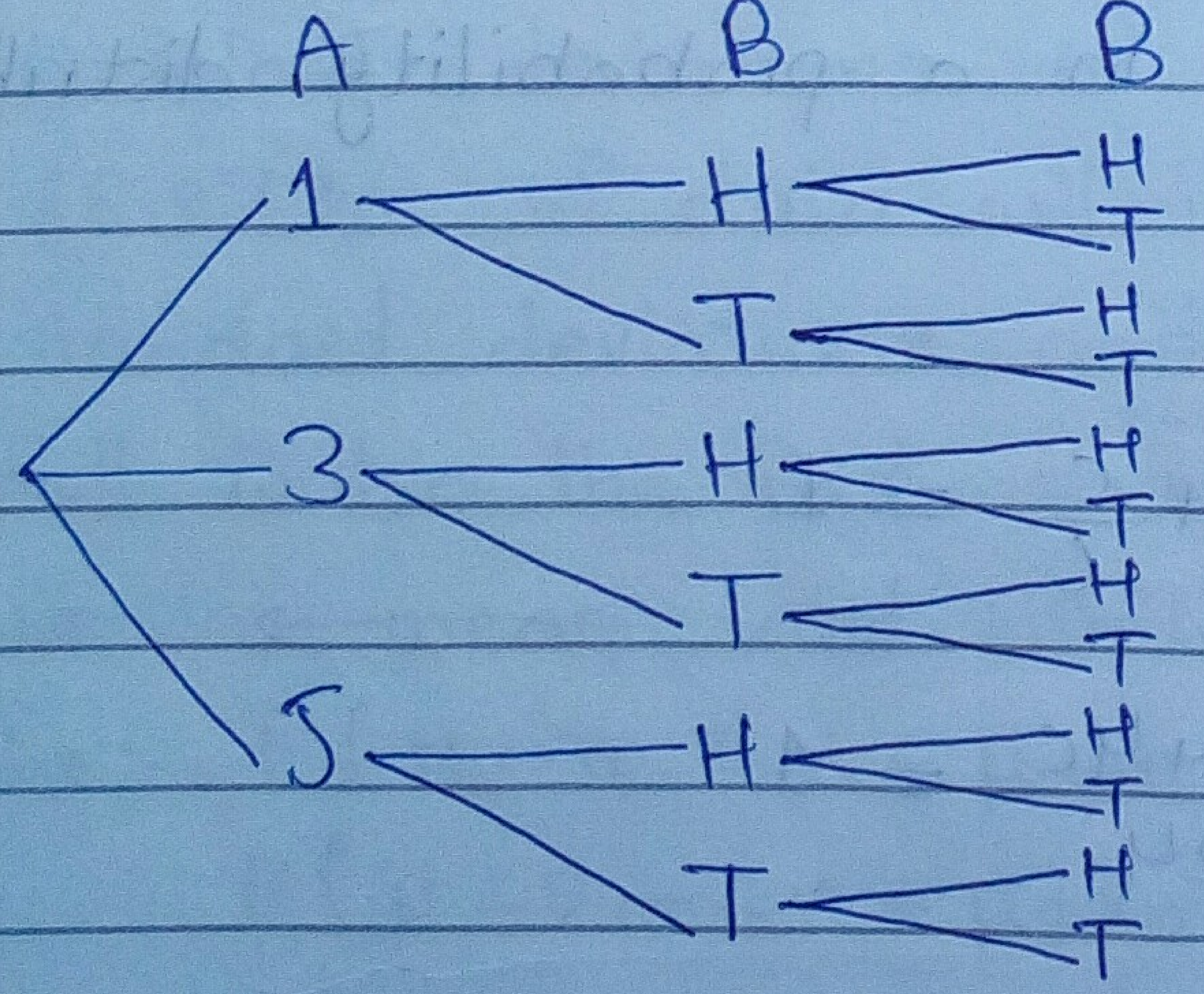
Tree diagram of tossing a die and then flipping a coin once if the number on the die is even.



A × B

(2, H)
(2, T)
(4, H)
(4, T)
(6, H)
(6, T)

Tree diagram of tossing a die and then flipping a coin twice if the number on the die is odd.



A × B × B

(1, H, H) (1, H, T)
(1, T, H) (1, T, T)
(3, H, H) (3, H, T)
(3, T, H) (3, T, T)
(5, H, H) (5, H, T)
(5, T, H) (5, T, T)

(vii) If X and Y are random variables then what will be the variance of the random variable ~~Z=3X~~
$Z = 3X - 4Y + 8$

$$Z = 3X - 4Y + 8$$
$$\text{var}(Z) = \text{var}(3X - 4Y + 8)$$

$var(Z) = var(3X) + Var(4Y) + 0$

$var(Z) = (3)^2 var(X) + (4)^2 var(Y)$

$var(Z) = 9\, var(X) + 16\, var(Y)$

(viii) The random variable X has a probability distribution given by

$P(X=x) = c/x$, $x = 2, 4, 6$, the value of c is......

x	$P(X=x) = c/x$
2	$c/2$
4	$c/4$
6	$c/6$
	$\Sigma P(x)$

We know that in a probability distribution $\Sigma P(x) = 1$, so

$$\frac{c}{2} + \frac{c}{4} + \frac{c}{6} = 1$$

$$\frac{c}{2} + \frac{6c+4c}{24} = 1$$

$$\frac{c}{2} + \frac{10c}{24} = 1$$

$$\frac{24c + 20c}{48} = 1$$

$$\frac{44c}{48} = 1$$

$$44c = 48$$

$$c = \frac{48}{44} = \frac{12}{11}$$

(ix) In the standard normal distribution, which parameter control height and width of Normal curve ?

In the standard normal distribution μ and σ control height and width of normal curve.

(x) State Empirical theorem.

The Empirical theorem states that for a normal distribution almost all data will fall within three standard deviations of the mean. It shows that 68.27 % of data will fall within the first ~~two~~ standard deviation~~s~~ and ~~99.73 %~~ 95.45 % of data will fall within the first ~~three~~ two standard deviations ~~of the mean.~~ and 99.73 % of data will fall

i.e; $P(\mu - \sigma \leq x \leq \mu + \sigma) = 68.27\ \%$

$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 95.45\ \%$

$P(\mu - 3\sigma \leq x \leq \mu + 2\sigma) = 99.73\ \%$

with-in the first three standard deviations of the mean.

Question -2: Short Questions (2013)

(a) Differentiate between of weighted Means and Geometric mean.

Weighted Mean:-

A mean in which each item being averaged is multiplied by a weight based on the item's relative importance is called weighted mean. The result is sumed and total is divided by the sum of the weights.

i.e; $\bar{x}_w = \dfrac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + w_3 + \cdots + w_n} = \dfrac{\sum_{i=0}^{n} x_i w_i}{\sum_{i=0}^{n} w_i}$

Geometric Mean:-

The geometric mean G of a set of n positive values $x_1, x_2, \ldots, x_n$ is defined as the positive nth root of their product.

i.e; $G = \sqrt[n]{x_1 \times x_2 \times x_3 \times \cdots \times x_n}$ where $x > 0$.

(b) Write properties of mean, median and mode (two properties of each).

The properties of mean, median and mode are following:

(i) Mean:-

(i) The sum of the square of deviations of the observations from mean is minimum.
i.e; $\Sigma (x-\bar{x})^2$ is minimum.

(ii) The sum of deviations of the observations from their mean is equal to zero.
i.e; $\Sigma (x-\bar{x}) = 0$

Median:-

(i) The sum of the absolute values of deviations of the observation from median is minimum.

(ii) It is positional average and is not influenced by the position of the observation.

Mode:-

(i) It is not affected by the presence of extremely large or small.

(ii) It can be determined for both the qualitative and quantitative data.

(c) Define primary and secondary data with two advantages disadvantages of each.

Primary Data:-

A data that is collected by first hand and did not undergo any statistical treatment is called primary data.

Advantages:-

(i) The investigator only collects data specific to the problem under study.

(ii) Their is no doubt about the quality of the data collected for the investigator.

Disadvantages:-

(i) It is expensive to collect the data.

(ii) The collection process is very time consuming.

Secondary Data:-

A data that is undergo from any statistical treatment or method is called secondary data.

Advantages:-

(i) It saves time.

(ii) It displays data in arranged order in the table form or graphical form.

Disadvantages:-

(i) There is a lack of control over data quality.

(ii) The data is collected by only self (primary data) so their is no surity of the appropriateness of the data.

(d) Define Correlation and Regression line with mathematical formula and graph?

Regression Line:-

A line which investigates the dependence of one variable called dependent variable or other variable called independent variable.

The regression line y on x is given as

$$y = a + bx$$

where a is intercept and b is slop. y is dependent and x is independent variable.

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$a = \bar{y} - b\bar{x}$$

The regression line x on y is given by

$$x = c + dy$$

where c is intercept and d is slope. x is independent and y is dependent variable.

$$d = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum y^2 - (\sum y)^2}$$

$$c = \bar{x} - d\bar{y}$$

Correlation:-

The measure of degree to which any two variables vary together is called correlation. Two variables are said to be

positively correlated if increase in one variable also causes increase in other variable.
Two variable are said to be negatively correlated if both variables tends to opposite direction to each other, if one variable decreases other increases. If one variable increases the other decreases

The correlation cofficient r is given by:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

(e) Differentiate b/w Variance, Standard Deviation and Cofficient of variation.

Variance:-

The variance of a set of observations is defined as the mean of the squares of deviations of the observations from their mean.

i.e; $$s^2 = \frac{\sum(x-\bar{x})^2}{n} \quad \text{(Ungrouped)}$$

$$s^2 = \frac{\sum f(x-\bar{x})^2}{\sum f} \quad \text{(Grouped)}$$

Standard Deviation:-

The positive square root of the variance is called standard deviation.

i.e; $$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \quad \text{(Ungrouped)}$$

$$S = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} \quad \text{(Grouped)}$$

Cofficient of Variation:-

The relative measure of variation which expressed the standard deviation as a percentage of the arithmetic mean of a data set is called a cofficient of variation.

i.e; $$C.V = \frac{S}{\bar{x}} \times 100$$

Question-2: Short Questions (2014)

(i) What is an average? Write the properties of a good average.

The measure which finds the center of a set of data is called average.

The averages are also known as measures of central tendencies or locations.

The properties of a good average are:

(i) It is regorously defined.

(ii) Based on all the observations made.

(iii) Simple to understand and easy to interpret.

(iv) Quickly and easily calculated.
(v) Amenable to mathematical treatment.
(vi) Relative stable in repeated sampling experiments.
(vii) Not influenced by abnormally large or small observations.

(ii) Define cofficient of variation, for what purpose, it is used?

The relative measure of variation which expresses the standard deviation as a percentage of a arithmetic mean of a data set is called cofficient of variation.

i.e; $C.V = \frac{S}{\bar{x}} \times 100$

It is used to compare the variation in two or more data sets or distributions that are measured in different units. e.g; one may be measured in hours and and the other in kilo-grams or rupees.

(iii) What is skewness? Write its forms.

The lack of symmetery of the values about some central value i.e; mean, median and mode is called skewness.

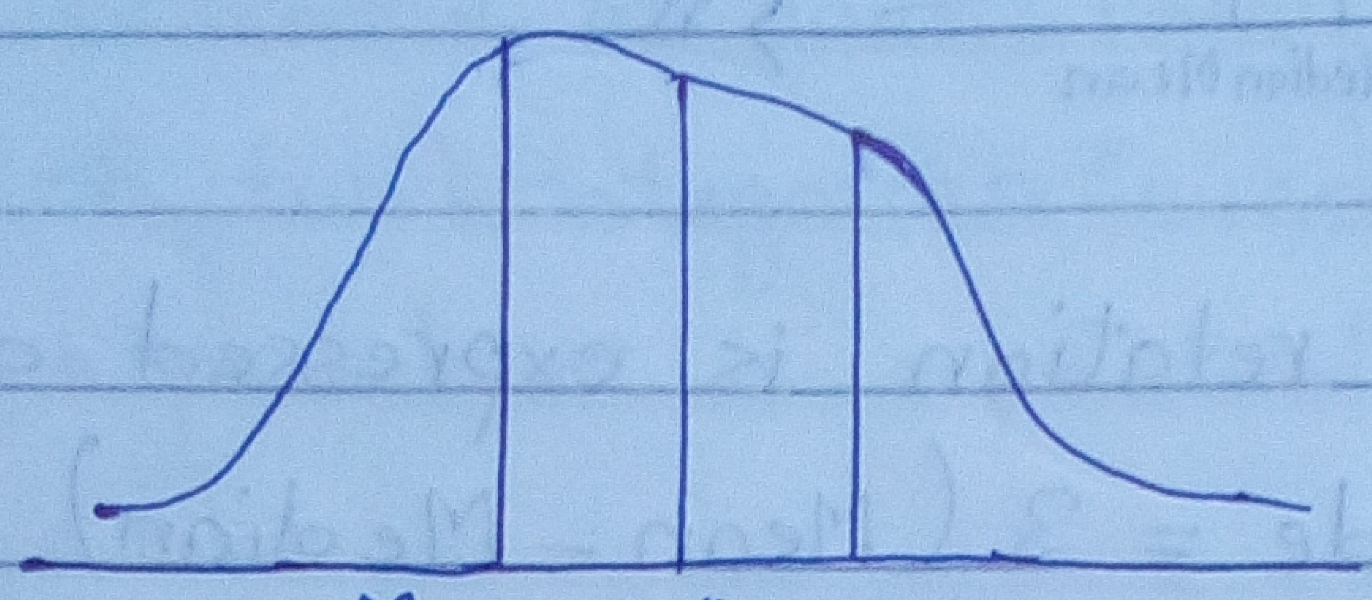
Its forms are:

(i) Positive Skewness (Mean > Median > Mode)

(ii) Negative Skewness (Mean < Median < Mode)

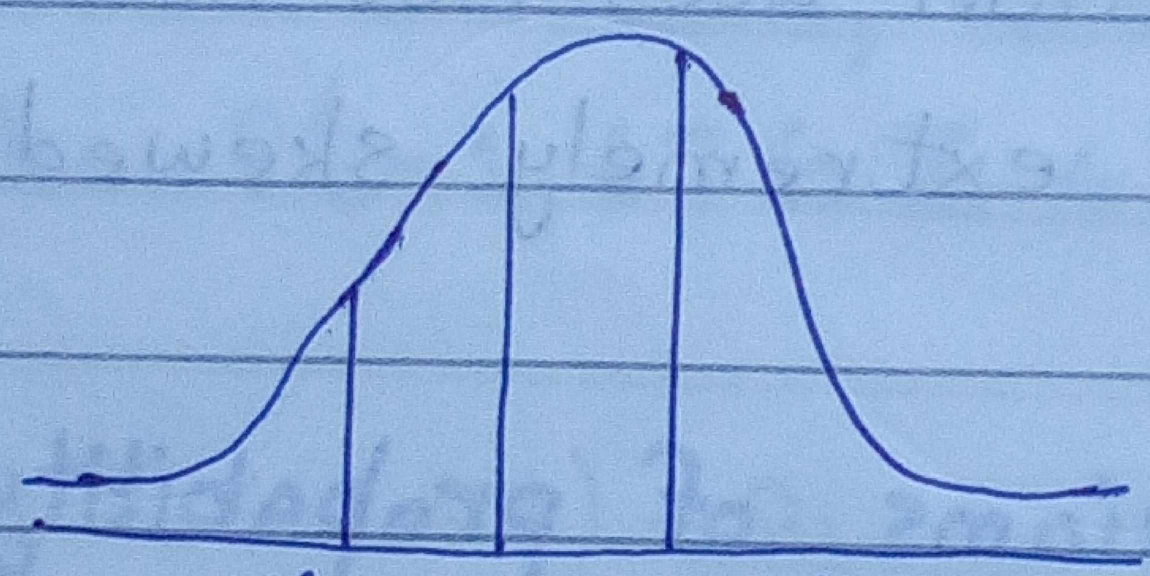
(iii) Symmeterical (Mean = Median = Mode)

Positive Skewness:-



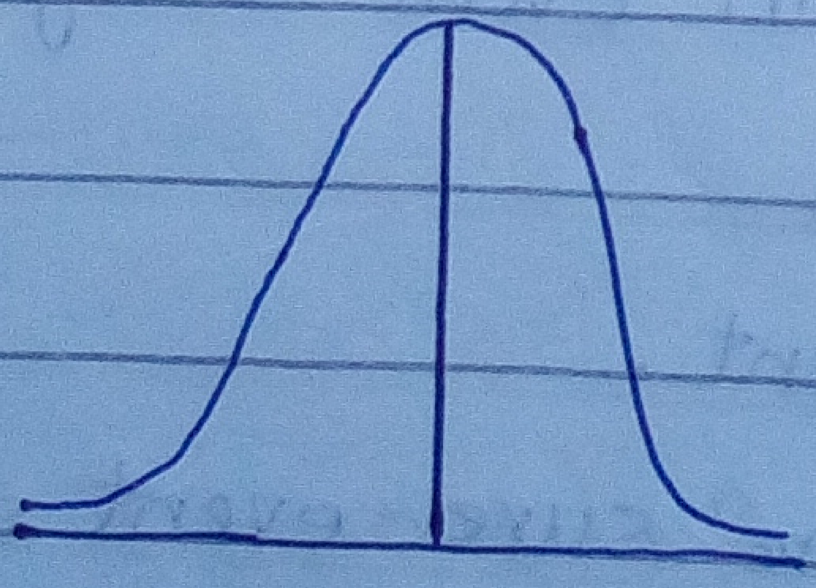
Mean Median Mode

Negative Skewness:-



Mean Median Mode

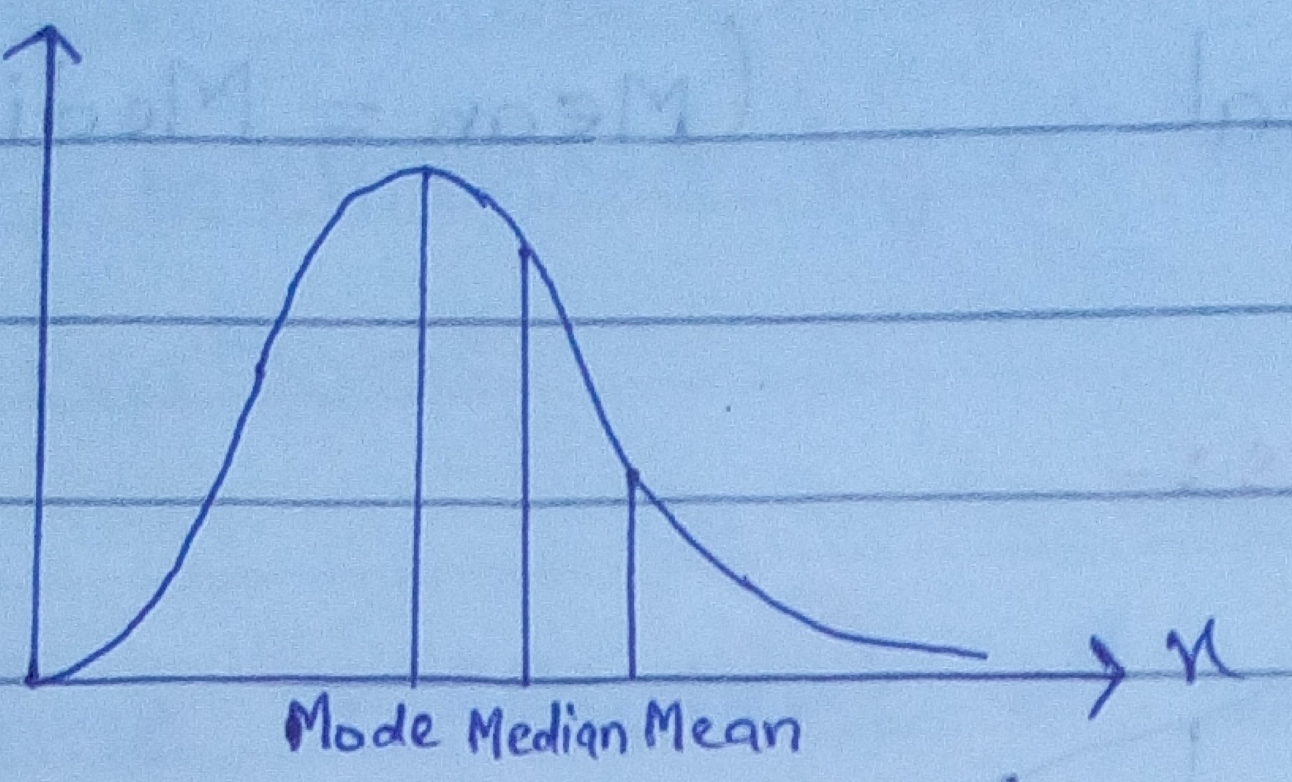
Symmetrical



Mean = Median = Mode

(iv) Write empirical relation b/w mean, median and mode.

Empirical relation between mean, median and mode is:



The empirical relation is expressed as:

Mean − Mode = 3 (Mean − Median)

or Mode = 3 Median − 2 Mean

The empirical relation does not hold in case of a J shaped or an extremely skewed distribution.

(v) Write three axioms of probability.

Three axioms of probability dist are:

(i) The probability of an event always lie b/w 0 and 1.

i.e; $0 \leq P(A_i) \leq 1$

where A_i is an event.

(ii) $P(S) = 1$ for the sure event S.

(iii) For any two mutually exculusive events such as A and B.

$$P(A \cup B) = P(A) + P(B)$$

(vi) Write the properties of a Binomial Probability Experiment.

The properties of Binomial probability experiment are:

(i) The probability of success remains constant for all successive trials.

(ii) The successive trials are independent.

(iii) The outcome of each trial may be classified into one of two categories "success" or "failure".

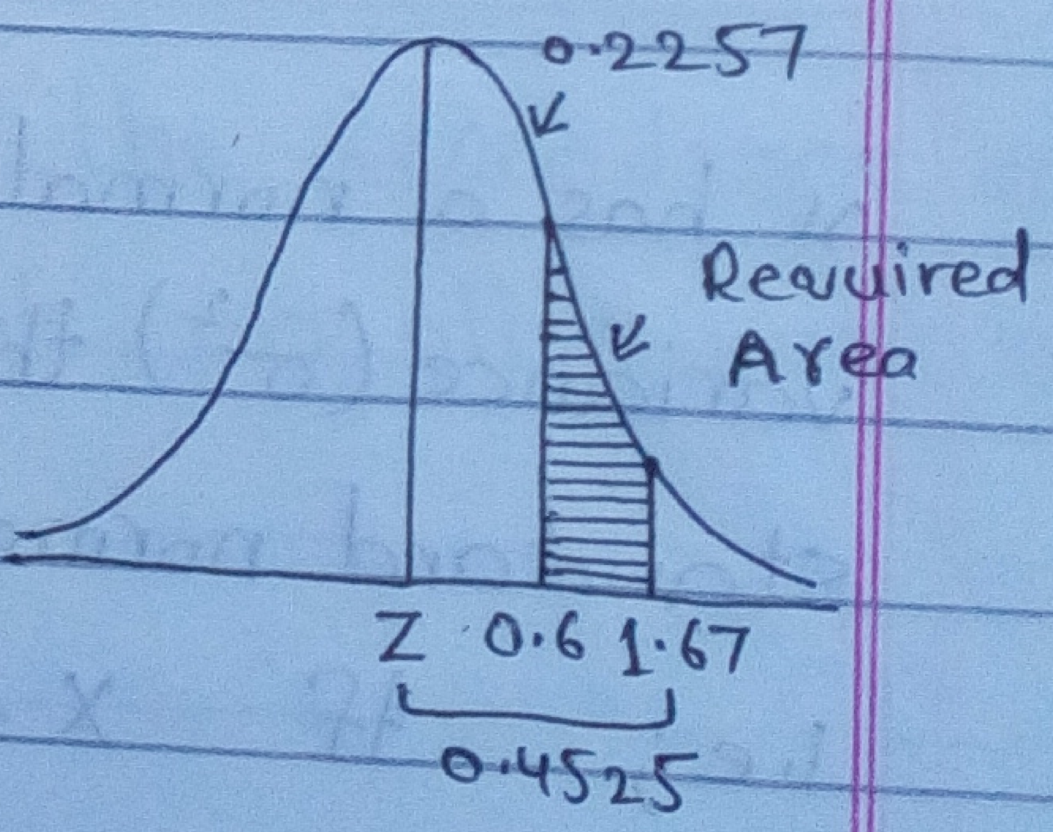
(iv) The experiment is repeated a fixed no of times, say n.

(vii) Let a random variable Z follow the Standard Normal Distribution, find $P(0.6 \leq Z \leq 1.67)$ and also $P(-1.67 \leq Z \leq -0.6)$.

$P(0.6 \leq Z \leq 1.67) = ?$

Required Area $= 0.4525 - 0.2257$

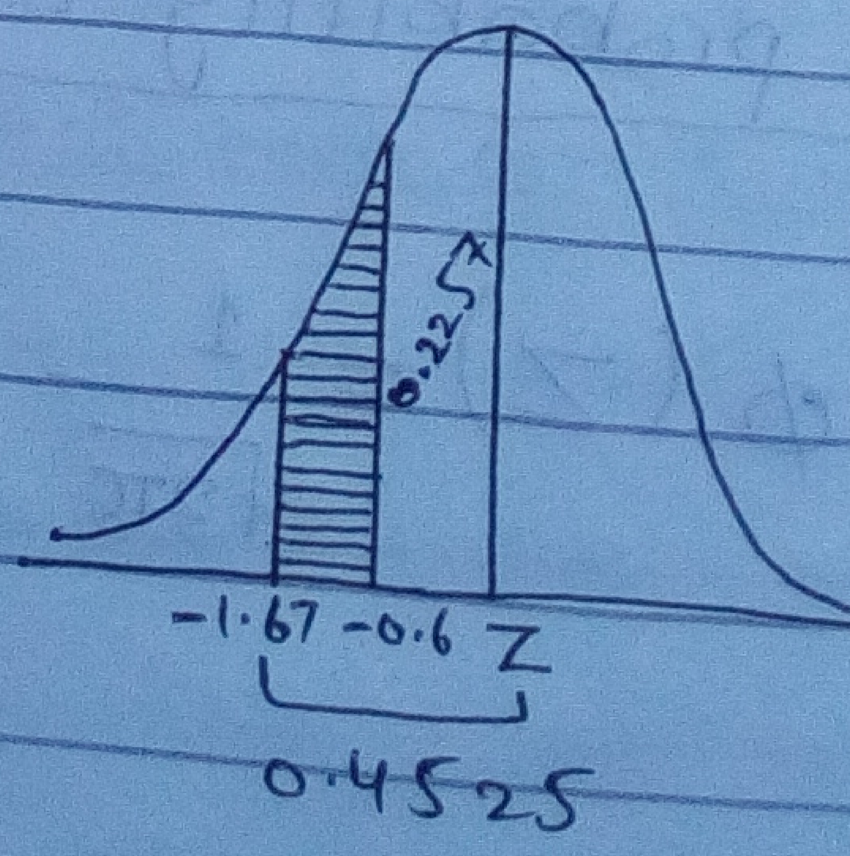
$P(0.6 \leq Z \leq 1.67) = 0.2268$



$P(-1.67 \leq Z \leq -0.6) = ?$

Required Area $= 0.4525 - 0.2257$

$P(-1.67 \leq Z \leq -0.6) = 0.2268$



(viii) Differentiate b/w Normal Probability Distribution and Standard Normal Probability Distribution.

Normal Probability Distribution:-

Normal probability distribution is the limiting form of binomial distribution when the no of trials i.e; n is very large and neither p (the probability of success) nor q (the probability of failure) is very small.

Its probability density function is given as:

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < +\infty$$

Standard Normal Probability Distribution:-

If the random variable x has a normal distribution with mean(μ) and variance (σ^2) then the random variable $Z = \frac{x-\mu}{\sigma}$ has a standard normal distribution with mean 0 and variance 1.

i.e; If $X \sim N(\mu, \sigma^2)$

then $Z \sim N(0, 1)$

Its probability density function is given as:

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \quad \text{for } -\infty < Z < +\infty$$

(ix) Write the principle of least squares.

The principle of least squares (LS) consists of determining the values of the unknown parameters that will minimize the sum of squares of errors (or residuals) where errors are defined as the difference between observed values and the corresponding values predicted or estimated by the fitted Model equation.
The parameter values thus determined will give the least sum of the squares and are known as least squares estimates.

(x) Differentiate b/w regression and correlation.

Regression:-

Regression investigates the dependence of one variable called dependent variable on other variable called independent variable.

Correlation:-

Correlation measures the degree of strength or degree of relationship between any two variables.

Question-4: Short Questions (2015)

(i) Define Event, Trial and Outcome.

Event:-

Any subset of the sample space is called event e.g; When a coin is tossed, the sample space is $S = \{H, T\}$ and the subset $A = \{H\}$ and the subset $B = \{T\}$ are the events.

Trial:-

The single performance of an experiment is called trial.

Outcome:-

The result obtained from a trial or random experiment is called an outcome.

(ii) Differentiate b/w Join and Marginal probabilities.

Marginal Probability:-

The probability of an event occuring [P(A)] is called marginal probability. It may be thought of an unconditional probability. It is not conditioned on another event.

e.g; The probability ~~of event A and event B occuring P(A and B)~~ that a card drawn is red $[P(\text{red}) = 0.5]$.

Joint Probability:-

The probability of event A and event B occuring P(A and B) is called joint probability.
It is the probability of the intersection of two or more events. The probability of the intersection of A and B may be written $P(A \cap B)$.
e.g; The probability that a card is a four and red
$= P(\text{four and red}) = \frac{2}{52} = \frac{1}{26}$ (There are two red fours in a deck of 52, the 4 of hearts and the 4 of diamonds).

(iii) Write one example of, each of the, Mutually exclusive event, Dependent event and independent event.

Mutually Exclusive Event:-

Two events are said to be mutually exclusive or disjoint if they cannot occur together
i.e; $A \cap B = \phi$
e.g; In a toss of single coin, we get either a head or tail, but both head and tail cannot occur together.

Dependent Event:-

Two events A and B are said to be said to be dependent if the occurance of one event depends on the occurance of other.
e.g; The result of two drawings of a ball from a bag are dependent, if the ball is not returned to the bag after the first draw.

When the reflection of events is without replacement, then the events are dependent events.

Independent Event:-

Two events A and B are said to be independent if the occurance of one event does not affect the occurance of other.

e.g; In two successive tosses of a fair coin, the outcome of the second toss does not depend on the outcome of the first toss.

When selection of events is with replacement, then the events are independent.

(iv) Write three axioms of probability.

(Same as Q-(v) of 2014)

Three axioms of probability are:

(i) The probability of an event always lie b/w 0 and 1.

i.e; $0 \leq P(A_i) \leq 1$

where A_i is an event.

(ii) $P(S) = 1$ for the sure event S.

(iii) For any two mutually exclusive events such as A and B,

$$P(A \cup B) = P(A) + P(B)$$

(v) **Write the properties of a Distribution Function for a continous variable.**

~~The~~

The properties of distribution function for a continous variable are:

(i) $f(x) \geq 0$ for all x.

(ii) $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

where x is the continous or random variable.

(vi) **Write the properties of a Binomial Probability Experiment.**

(Same as Q-(vi) of 2014)

The properties of Binomial Probability Experiment are:

(i) The probability of success remains constant for all successive trials.

(ii) The successive trials are independent.

(iii) The outcome of each trial may be classified into one of two categories "success" or "failure".

(iv) The experiment is repeated a fixed no of times, say n.

(vii) **Differentiate b/w Normal Probability Distribution and Standard Normal Probability Distribution.**

(Same as Q-(viii) of 2014)

(viii) **Let a random variable Z follow the Standard Normal Distribution, find $P(0.6 \leq Z \leq 1.67)$ and also $P(-1.67 \leq Z \leq -0.6)$.**

(Same as Q-(vii) of 2014)

(ix) Let X follow the uniform Distribution $f(x) = 1/(b-a)$, $a \le x \le b$ then prove that mean of X be equal to $(a+b)/2$.

$f(x) = 1/(b-a) \quad a \le x \le b$

We know that mean is

$$E(x) = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$$

$$E(x) = \int_a^b x \cdot \left(\frac{1}{b-a}\right) dx$$

$$E(x) = \frac{1}{b-a} \int_a^b x \cdot dx$$

$$E(x) = \frac{1}{b-a} \cdot \left| \frac{x^2}{2} \right|_a^b$$

$$E(x) = \frac{1}{2(b-a)} \left| x^2 \right|_a^b$$

$$E(x) = \frac{1}{2(b-a)} (b^2 - a^2)$$

$$E(x) = \frac{(b+a)(b-a)}{2(b-a)}$$

$$E(x) = \frac{b+a}{2}$$ Hence mean of x is equal to $(a+b)/2$

(x) A binomial distribution tends to become Normal distribution; discuss.

A binomial distribution tends to become normal distribution if the number of trials n are increased to a very large number for a fixed value of probability of success (p). The normal distribution is considered as a limiting form of binomial distribution.