3 SETS

MODEL QUESTION SETS WITH ANSWERS

SET – 1

Time – 3 Hours **Full Marks – 70**

The figures in the right - hand margin indicate marks.

GROUP – A

1. Pick up the correct answer : [1 × 10 = 10

(a) If $r_{xy} = 0$ then which of the following is definitely true?

(i) x and y have no relation

(ii) x and y are not linearly related

(iii) x and y are independent

(iv) x and y are not independent

(b) If the slope of the line of regression of Y on X is -0.25 then the range for the regression coefficient of X on Y will be:

(i) $[0, 1]$ (ii) $(-1, 0)$

(iii) $(0, 4]$ (iv) $[-4, 0)$

(c) Which of the following index numbers does not satisfy TRT?

(i) Fisher's ideal index

(ii) Weighted GM of relatives index

(iii) Simple AM of relatives index

(iv) Simple GM of relatives index

(d) Which of the following methods can provide only linear trend ?

(i) Least square method

(ii) Semi average method

(iii) Graphical method

(iv) None of these

(e) What is the unbiased estimator of population total in simple random sampling

(i) Sample mean

(ii) Population size × Sample mean

(iii) Sample total (iv) None of these

(f) Which of the following is not a criterion for selecting a commodity ?

(i) should be stable in quality

(ii) should be a cheapercommodity

(iii) should be commonly used

(iv) should be a commodity of necessity

(g) If a binomial distribution has two modes, then which of the following is definitely true?

(i) $(n + 1)p$ is not an integer

(ii) np is a natural number

(iii) mean and variance are equal

(iv) $(n + 1)p$ is a natural number

(h) The area under the standard normal curve in the interval $[-1.96, 1.96]$ is_____

(i) 9.5% (ii) 100%

(iii) 95% (iv) 99%

(i) The description of all the sampling units is called________

(i) Sample size (ii) Sampling schedule

(iii) Sampling frame (iv) Stratified sampling

(j) In simple random sampling, what is the unbiassed estimate of population mean ?

(i) Population size (ii) Sample size

(iii) Sample mean (iv) Sampling fraction

2. (I) correct the statements if necessary [1 × 5 = 5

(a) If one regression coefficient is negative then the correlation coefficient is either positive or negative.

(b) Seasonal variation in a time series is the effect due to such factors which change only after a long time.

(c) If $p < ½$ then binomial distribution is symmetrical.

(d) Paasche's index number has an upward bias.

(e) Census provides a wider scope of study in comparison to sampling.

(II) Fill in the blanks: **[1× 5 = 5**

(f) The conditions under which a Binomial distribution tends to Poisson are __________

(g) _______ number of normal equations are to be solved while fitting a second degree trend.

(h) The price of a certain quality of rice in 2014 was Rs 20/- per KG which increased to Rs 28/- per KG in 2016. The price relative for the commodity is __________

(i) Two variables are related by only one equation $50x + 39y = 2765$. Coefficient of correlation between them is ______

(j) Error in tabulation causes _______ error.

GROUP – B

3. Answer any ten of the following questions **[2 × 10 = 20**

(a) Express the regression coefficients by using correlation coefficient.

(b) Write any two merits of moving average method in comparison to the semi-average method.

(c) Define a price relative and give one example.

(d) Derive the mode of Poisson distribution.

(e) Discuss the skewness of Poisson distribution.

(f) Define correlation coefficient and mention any two of its properties.

(g) Explain a business cycle using the suitable diagram.

(h) What is the simplest form of an index number and how is it computed ?

(i) Prove that the mean and variance of a standard normal distribution are 0 and 1 respectively.

(j) Examine TRT for Paasche's index.

(k) Mention the assumption under which simple random sampling is adopted.

(l) Mention at least four demerits of Census when compared to sampling.

4. Answer any three of the following questions. **[3 × 3 = 9**

(a) Explain the method of interpretation of the value of correlation coefficient.

(b) In simple random sampling, prove that sample mean is an unbiassed estimator of population mean.

(c) Derive Poisson distribution as a limiting case of Binomial distribution

(d) Examine TRT and FRT for simple GM of relatives method.

(e) With the help of suitable example explain sampling unit.

GROUP – C

Answer any three of the following questions **[7 × 3 = 21**

5. Define a time series giving suitable examples. Explain its various components.

6. Define an Index Number. Explain the various types of weighted and un-weighted index numbers and state their merits and demerits.

7. Derive the standard deviation of binomial distribution.

8. Explain what is meant by correlation. Describe the method of scatter diagram for studying correlation. Write its a advantages and disadvantages.

9. Explain the meaning of the sampling distribution of a statistic. Also construct the sampling distribution of mean for a simple random sample of 3 units selected without replacement from a population of 5 units.

ANSWERS TO SET – 1

GROUP – A

1. (a) (ii) x and y have no linear relation

(b) (iv) $[-4, 0)$

(c) (iii) Simple AM of relatives index

(d) (ii) Semi average method

(e) (ii) Population size × Sample mean

(f) (ii) Should be a cheaper commodity

(g) (iv) $(n+1)p$ is a natural number

(h) (iii) 95%

(i) (iii) Sampling frame

(j) (iii) Sample mean

2.(I) (a) If one regression coefficient is negative then the correlation coefficient is definitely negative.

(b) Trend component in a time series is the effect due to such factors which change only after a long time.

(c) If $p < ½$ then binomial distribution is symmetrical.

(d) Laspeyre's index number has an upward bias.

(e) Sampling provides a wider scope of study in comparison to Census.

(II) (f) The conditions under which a Binomial distribution tends to Poisson are (i) $n \to \infty$ (ii) $p \to 0$ (iii) $np = 1$ is a constant.

(g) Three number of normal equations are to be solved while fitting a second degree trend.

(h) The price of a certain quality of rice in 2014 was Rs 20/- per KG which increased to Rs 28/- per KG in 2016. The price relative for the commodity is 140 .

(i) Two variables are related by only one equation $50x + 39y = 2765$. Coefficient of correlation between them is -1 .

(j) Error in tabulation causes non-sampling error.

GROUP – B

3.(a) Express the regression coefficients by using correlation coefficient.

Ans. The normal equations for fitting the line of regression of Y on X of the form $y = a + bx$ are:

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \quad .... \text{(i)}$$

and $$\sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad .... \text{(ii)}$$

On solving these normal equations (i) and (ii) the value of b comes out to be :

$$b = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i x_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}}$$

$$= \frac{\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$= \frac{Cov(x,y)}{\sigma_x^2} \Rightarrow b_{yx} = \frac{Cov(x,y)}{\sigma_x^2}$$

and similarly $b_{xy} = \dfrac{Cov(x,y)}{\sigma_y^2}$

Further it is known that

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \Rightarrow Cov(x,y) = r_{xy}\sigma_x\sigma_y$$

Hence $b_{yx} = \dfrac{r_{xy}\sigma_x\sigma_y}{\sigma_x^2} = r_{xy}\dfrac{\sigma_y}{\sigma_x}$ and

$$b_{xy} = \frac{r_{xy}\sigma_x\sigma_y}{\sigma_y^2} = r_{xy}\frac{\sigma_x}{\sigma_y}$$

(b) Write any two merits of moving average method in comparison to the semi-average method.

Ans. The semi-average method pre-supposes the presence of linear trend in every time series and thus fails to provide non-linear trend. But moving average method can provide linear as well as non-linear trends.

Trend by the semi-average method has the drawback of being unduly influenced by higher values in the series but moving average method has no such demerit.

(c) Define a price relative and give one example.

Ans. A price relative for a commodity may be defined as the relative change in the price of the commodity with respect to the price in the base year.

Mathematically,

Price relative of a commodity =

$$\frac{\text{price of the commodity in the current year}}{\text{Price of the commodity in the base year}} \times 100$$

$$= \frac{p_1}{p_0} \times 100$$

(d) Derive the mode of Poisson distribution.

Ans. Mode of any probability distribution is that value of the random variable which has the maximum probability. Let r be the mode of Poisson distribution. Then r is a positive integer such that P (X = r) is the maximum. This means that

$$P(X = r) \geq P(X = r-1) \quad ......\text{(i)}$$

and $$P(X = r) \geq P(X = r+1) \quad ...... \text{(ii)}$$

From (i) $$\frac{P(X=r)}{P(X=r-1)} \geq 1 \Rightarrow \frac{e^{-\lambda}\lambda^r / r!}{e^{-\lambda}\lambda^{r-1} / (r-1)!} \geq 1$$

$$\Rightarrow \frac{\lambda}{r} \geq 1 \Rightarrow \lambda \geq r \Rightarrow r \leq \lambda \quad ....\text{(iii)}$$

From (ii) $\frac{P(X=r)}{P(X=r+1)} \geq 1 \Rightarrow \frac{e^{-\lambda}\lambda^r / r!}{e^{-\lambda}\lambda^{r+1} / (r+1)!} \geq 1$

$\Rightarrow \frac{r+1}{\lambda} \geq 1 \Rightarrow r+1 \geq \lambda \Rightarrow \lambda - 1 \leq r$

Combining (iii) and (iv), mode of Poisson distribution is that integral value of r which satisfies the condition $\lambda - 1 \leq r \leq \lambda$.

(e) Discuss the skewness of Poisson distribution.

Ans. For Poisson distribution

$\mu_2 = \lambda,\ \mu_3 = \lambda,$

So, $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda}$

Since λ is always positive, $\gamma_1 = +\sqrt{\beta_1} = +\frac{1}{\sqrt{\lambda}}$

Thus for Poisson distribution, γ_1 is always positive and it is concluded that Poisson distribution is always positively skewed.

(f) Define correlation coefficient and mention any two of its properties.

Ans. Correlation coefficient may be defined as the degree of linear relationship between two variables.

It is computed by the formula: $r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$

Properties:

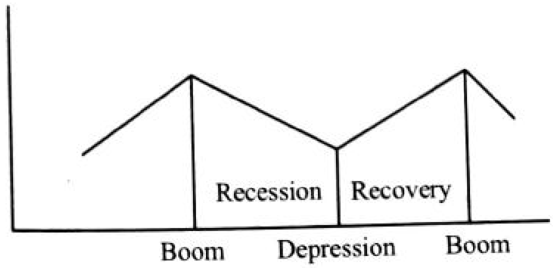
(i) Correlation coefficient is independent of change of origin and scale.

(ii) The limits of the correlation coefficient are ±1 i.e. $-1 \leq r \leq 1$ or $|r| \leq 1$.

(g) Explain a business cycle using the suitable diagram.

Ans. Every business passes through 4 stages Boom – Recession – Depression – Recovery – Boom which is called a business cycle. Boom refers to the period in which the business experiences the maximum profit level and Depression is the period having the minimum level of profit. The period from Boom to Depression in which the profit level goes on decreasing is called the stage of Recession and the period from Depression to the next Boom where the profit level again moves on increasing to reach a highest level is termed as the Recovery stage. The time interval between two consecutive booms is called the period of the cycle provided it is more than one year.

The diagram given below presents a business cycle.



(h) What is the simplest form of an index number and how is it computed ?

Ans. The simplest form of an index number is a price relative. It is a univariate index which indicates the relative change in the price of a commodity with respect to time or space or due to any other characteristic. It is computed by the formula:

$$= \frac{\text{Price of commodity in the current period}}{\text{Price of the commodity in the base period}} \times 100$$

It is a pure number free from units of measurement.

(i) Prove that the mean and variance of a standard normal distribution are 0 and 1 respectively.

Ans. If $X \sim N(\mu, 6)$ i.e X is a normal variate having mean = μ and standard deviation = σ, then by definition, a standard normal variable is given by: $= \frac{X-\mu}{\sigma}$.

$Z\frac{X-\mu}{\sigma} \Rightarrow X = \mu + Z\sigma$ Hence $E(X) = E(\mu + Z\sigma)$

$\Rightarrow \mu = E(\mu) + \sigma\ E(Z)$

$\Rightarrow \mu = \mu + \sigma E(Z) \Rightarrow \sigma E(Z) = 0$

$\Rightarrow E(Z) = 0$ i.e. Mean of $Z = 0$

Further $V(X) = V(\mu + Z\sigma) \Rightarrow \sigma^2 = \sigma^2 V(Z)$

$\Rightarrow V(Z) = 1$

So the mean and variance of a standard normal distribution are 0 and 1 respectively.

(j) Examine TRT for Paasche's index.

Ans. For Paasche's index,

$$P_{01}^{Pa} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \text{ and } P_{10}^{Pa} = \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}$$

So, $P_{01}^{Pa} \times P_{10}^{Pa} \neq 1$

Hence Paasche's index does not satisfy TRT.

(k) Mention the assumption under which simple random sampling is adopted.

Ans. Simple random sampling is adopted under the following assumptions:

(i) All the units in the population are homogeneous i.e. every unit of the population bears all the characteristics of the population.

(ii) Every unit has equal chance of being included in the sample.

(l) Mention at least four demerits of Census when compared to sampling.

Ans. The demerits of census as compared to sampling are as follows:

(i) In census method data are collected from every unit of the population but in sampling, data are collected from a subset of the population consisting of only some selected units. Hence in census data are to be collected from more units as compared to that in sampling.

(ii) Census requires more time as compared to that needed for sampling.

(iii) Census is more expensive in comparison to sampling.

(iv) Census provides a limited scope of study in comparison to sampling.

4.(a) Explain the method of interpretation of the value of correlation coefficient.

Ans. The interpretation of the value of correlation coefficient can be done by using the probable error.

Probable error of r = PE (r) = 0.6745 × Standard error of r

Standard error of r = SE (r) $\frac{1-r^2}{\sqrt{n}}$

Interpretation:

(i) If | r | > 6 × PE (r) then r is highly significant i.e. there is a strong association between the variables.

(ii) If | r | < PE (r) then r is insignificant i.e. there is a weak association between the variables.

(iii) If PE (r) ≤ | r | ≤ 6 × PE (r) then the variables are moderately correlated.

(b) In simple random sampling, prove that sample mean is an unbiassed estimator of population mean.

Ans. For simple random sampling :

N = Population size

n = Sample size

$Y_i = i^{th}$ unit of the sample (i = 1, 2, ..., N)

$y_i = i^{th}$ unit of the sample (i = 1, 2, ..., n)

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \text{Population mean}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \text{Sample mean}$$

$$P(Y_i \text{ is included in the sample}) = \frac{n}{N}$$

$$\alpha_i = \begin{cases} 1, \text{ it } y_i \text{ is included in the sample} \\ 0, \text{ Otherwise} \end{cases}$$

$$\therefore P(\alpha_i = 1) = \frac{n}{N} \text{ and } P(\alpha_i = 0) = \left(1 - \frac{n}{N}\right)$$

Further $\sum_{i=1}^{n} y_i = \sum_{i=1}^{N} \alpha_i Y_i$

Need to prove that $E(\bar{y}) = \bar{Y}$.

Proof : $E(\bar{y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{N}\alpha_i y_i\right)$

$$= \frac{1}{n}\sum_{i=1}^{N} Y_i E(\alpha_i) = \frac{1}{n}\sum_{i=1}^{N} Y_i\left[1 \times \frac{n}{N} + 0 \times \left(1 - \frac{n}{N}\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} Y_i \cdot \frac{n}{N} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \bar{Y}$$

(c) Derive Poisson distribution as a limiting case of Binomial distribution.

Ans. Binomial distribution tends to a Poisson distribution under the conditions:

(i) $n \to \infty$, (ii) $p \to 0$ and

(iii) np is a constant denoted by λ.

Thus the probability generating function of Poisson distribution can be obtained by applying these conditions on the probability generating function of Binomial distribution.

Hence for a Poisson distribution,

$$P(X = r) = \lim_{n \to \infty} \binom{n}{r} p^r q^{n-r}$$

$$= \lim_{\substack{n \to \infty \\ p \to 0}} \left[\frac{n!}{r!(n-r)!} p^r q^{n-r}\right]$$

$$= \lim_{\substack{n\to\infty \\ p\to 0}} \left[\frac{n(n-1)(n-2)...(n-r+1)(n-r)!}{r!(n-r)!} \left(\frac{np}{n}\right)^r \left(1-\frac{np}{n}\right)^{n-r} \right]$$

$$= \lim_{\substack{n\to\infty \\ p\to 0}} \left[\frac{n(n-1)(n-2)...(n-r+1)}{r!} \left(\frac{\lambda^r}{n^r}\right)\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-r} \right]$$

$$= \frac{\lambda^r}{r!}\lim_{n\to\infty}\left[\frac{n}{n}\times\frac{(n-1)}{n}\times\frac{(n-2)}{n}\times....\times\frac{(n-r+1)}{n}\right] \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \times \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-r}$$

$$= \frac{\lambda^r}{r!}\lim_{n\to\infty}\left[1\times\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right).....\left(1-\frac{r-1}{n}\right)\right] e^{-\lambda}\times 1 = \frac{e^{-\lambda}\lambda^r}{r!}$$

Thus the probability generating function of Poisson distribution is $P(X=r) = \frac{e^{-\lambda}\lambda^r}{r!}(r=0,1,2,3,.....)$

(d) Examine TRT and FRT for simple GM of relatives method.

Ans. The formula for index number by the simple GM of relatives method is given by:

$$P_{01} = \text{Antilog}\left[\frac{1}{n}\Sigma\log\left(\frac{p_1}{p_0}\right)\right]$$

$$\text{So } P_{10} = \text{Antilog}\left[\frac{1}{n}\Sigma\log\left(\frac{p_0}{p_1}\right)\right]$$

$$\text{Hence } P_{01}\times P_{10} = \text{Antilog}\left[\frac{1}{n}\Sigma\log\left(\frac{p_1}{p_0}\right)\right] \times \text{Antilog}\left[\frac{1}{n}\Sigma\log\left(\frac{p_0}{p_1}\right)\right]$$

$$= \text{Antilog}\left[\frac{1}{n}\Sigma\log\left(\frac{p_1}{p_0}\right)+\frac{1}{n}\Sigma\log\left(\frac{p_0}{p_1}\right)\right]$$

$$= \text{Antilog}\left[\frac{1}{n}\Sigma\log\left\{\left(\frac{p_1}{p_0}\right)+\log\left(\frac{p_0}{p_1}\right)\right\}\right]$$

$$= \text{Antilog}\left[\frac{1}{n}\Sigma\log\left(\frac{p_1}{p_0}\times\frac{p_0}{p_1}\right)\right]$$

$$= \text{Antilog}\left[\frac{1}{n}\Sigma 0\right] = \text{Antilog } (0) = 1$$

Since $P_{01}\times P_{10} = 1$

the index numbet formula based on simple geometric mean of relatives satisfies Time Reversal Test.

(e) With the help of suitable example explain sampling unit.

Ans. At the time of dealing with large population spread over a wide area, it becomes difficult and involves more time and money for collecting data if units are selected at random directly from the population units. Hence in such cases, the population is divided into disjoint subsets on some basis such as geographical location, in terms of institutions etc. Each such subset of the population is termed as a sampling unit and the required sample is obtained by selecting some of these sampling units.

The concept of sampling units can be clear from the following example:

Suppose a total of 500, 000 students passed +2 examinations of CHSE, Odisha in 2018 and it is required to get a 1% sample i.e a sample of 5000 pass out students from about 1500 colleges spread over the 30 districts of odisha. If the 5000 students are selected at random directly out of the 5 lakh students, there is every possibility that a very small number of students may be selected from different colleges. As a result for collection of data, extensive touring is necessary which is likely to consume more money as well as more time and also a large manpower if time needs to be minimized.

On the other hand the entire population can be divided into 1500 colleges and 1% i.e 15 colleges may be selected at random and data are collected from all the students passed out from these colleges, then the process becomes more convenient as well as likely to be economical and time saving.

In such a situation each college is called a sampling unit.

GROUP – C

5. Define a time series giving suitable examples. Explain its various components.

Ans. A time series may be defined as the chronological arrangement of occurrences of an event presented preferably at equal intervals of time. Thus a time series is a bivariate data in which time is the independent variable.

Examples of time series are: (i) Annual production of rice in Odisha for a period of 15 years. (ii) Monthly sales of a departmental store for three consecutive years. (iii) Quarterly profits of a business house for 5 consecutive years. (iv) Daily sales of newspaper for six months etc.

The value of a time series is the combined effect due to a large number of factors. Such factors can be classified into four major categories called the components of the time series. The components of a time series are: (i) Trend or long term movement, (ii) Seasonal variations (iii) Cyclical fluctuations (iv) Random component or Irregular fluctuations. The two components Seasonal variations and Cyclical fluctuations are combined together and named as Short term fluctuations or periodic changes. A time series may contain any one or more of the four components.

Trend or long term movement: It is the component of a time series which presents the general tendency of the values of the time series to increase or decrease with respect to time. So trend of a time series can be studied only if the data is observed over a long period of time. Due to this reason, trend is also called the long term movement. Trend in a time series mainly arises out of the effect due to such factors which either do not change or show a slow and gradual change which can be marked only if the data is studied over a long period of time. Examples of factors due to which trend occurs are: Soil fertility, Changes in the taste and habit of a population, Changes caused from scientific inventions and discoveries etc. Sometimes it is observed that the trend shows the behaviour of remaining more or less constant or fluctuating between two fixed values for example the average temperature, average rainfall of a specific place remain almost constant every year.

Seasonal variations: In the analysis of time series, a season means a specific time interval of less than one year time. Popularly used seasons during the analysis of time series are Months and Quarters. A year is divided into 12 months and a quarter means a period of three consecutive months.

It is observed that time series data presented season wise is likely to show a uniform change periodically occurring at a regular interval of one year or less time. Such variations in the time series are referred to as seasonal variations. Seasonal variations are precise, definite and are predictable for future. So its study is highly essential for running a business successfully.

Seasonal variations in a time series mainly occur due to two types of forces namely (i) Natural forces, (ii) Social customs and traditions.

Examples of seasonal variation caused from natural forces are: (a) increase in the sale of ice-cream, cold drinks etc. during summer season. (b) Increase in the sale of woolen clothes during winter season. (c) Decrease in the price of agricultural commodities during their respective harvesting seasons etc.

Examples of seasonal variation caused from social customs and traditions are: (a) increase in sale of fire crackers during Diwali, (b) Increase in the sale of greeting cards during period of New year, (c) Increase in the sale of copies, books etc during admission season etc.

Cyclical fluctuations: These are the variations observed in a time series at more or less regular interval but their period of occurrence is more than one year. In other words these are oscillatory movements in the time series with period of oscillation more than one year. A complete period of oscillation is called the period of cycle. Cyclical fluctuations in trade occur due to the business cycle. So the study of cyclical fluctuation helps in running a business smoothly. The stages of a business cycle are: Boom – Recession – Depression – Recovery – Boom.

Boom is the period showing the maximum profit and Depression is the period showing the minimum profit level. The time interval between Boom and Depression is called Recession and the time interval between Depression and the next Boom is called Recovery. The period of time between two consecutive Booms is called the period of the cycle. Some times the period of the cycle is not perfectly regular; so in such cases the average of all the periods of the cycles experienced by the time series is considered as the period of the cycle.

Examples of factors responsible for cyclical variations are: (i) changes in fashion, (ii) changes caused from scientific and technological developments etc.

Irregular fluctuations: The fluctuations in a time series which are purely random, erratic, uncertain and unpredictable are termed as irregular fluctuations. Such fluctuations are caused due to two types of forces namely (i) Natural forces such as flood, earthquake, cyclone etc., (ii) socio-economic and anti-social activities such as war, strike, lock-out, terrorist activities, communal riot etc. Since these fluctuations are unpredictable it is very difficult to isolate them and so their study is not possible.

These fluctuations sometimes become very effective and may give rise to seasonal as well as cyclical fluctuations. For example flood in the coastal districts of Odisha is experienced in the rainy season of every year. So it becomes a seasonal variation.

6. Define an Index Number. Explain the various types of weighted and un-weighted index numbers and state their merits and demerits.

Ans. An index number may be defined as the relative change in the level of a phenomenon or a group of phenomena with respect to time or place or due to any other characteristic. Index numbers are popularly used to measure the relative change in economic phenomena such as price, quantity etc. The simplest form of an index number is the price relative which may be defined as the relative change in the price level of a commodity with respect to time.

Mathematically,

$$= \frac{p_1}{p_0} = \frac{\text{Price of the commodity in the curent year}}{\text{Price of the commodity in the base year}}$$

Index numbers are broadly classified into two categories namely

(i) Un-weighted index numbers and

(ii) Weighted index numbers

An un-weighted index number is the index number computed by assuming that all the commodities are of equal importance. So no specific weight is assigned to any commodity. On the other hand a weighted index number is that which is computed by assigning rational weights to various commodities in accordance with their relative importance.

Notations:

P_{01} = Price index for the current year denoted by 1 with respect to the base year denoted by 0

p_{0i} = Price of the i^{th} commodity in the base year

p_{1i} = Price of the i^{th} commodity in the current year

q_{0i} = Quantity of the i^{th} commodity in the base year

q_{1i} = Quantity of the i^{th} commodity in the current year

w_i = Number of commodities

Un-weighted index numbers and their merits, demerits:

(i) Simple aggregative method :

$$P_{01} = \frac{\sum_{i=1}^{n} p_{1i}}{\sum_{i=1}^{n} p_{oi}} \times 100$$

Merit: It is simple to understand and easy to calculate. It satisfies Time Reversal Test. It satisfies Circular Test.

Demerit: It is highly affected by the units in which price and quantity are measured i.e. it does not satisfy unit test. It does not satisfy Factor Reversal Test.

(ii) Simple arithmetic mean of relatives :

$$P_{01} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{p_{1i}}{p_{0i}}\right)\right] \times 100$$

Merit: It is simple to understand and easy to calculate. It satisfies unit test.

Demerit: It does not satisfy Time Reversal Test, Factor Reversal Test and Circular Test.

(iii) Simple geometric mean of relatives :

$$P_{01} = \left[\text{Antilog}\left\{\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{p_{1i}}{p_{0i}}\right)\right\}\right] \times 100$$

Merit: It satisfies unit test. It uses geometric mean which is the best average for construction of index numbers.

Demerit: It satisfies Time Reversal Test and Circular Test.

Weighted index numbers and their merits, demerits:

(i) Weighted aggregative method:

$$P_{01} = \frac{\sum_{i=1}^{n} w_i p_{1i}}{\sum_{i=1}^{n} w_i p_{oi}} \times 100$$

Merit: It satisfies unit test.

Demerit: It does not satisfy Time Reversal Test, Factor Reversal Test and Circular Test.

(ii) Weighted arithmetic mean of relatives:

$$P_{01} = \left[\frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \left(\frac{p_{1i}}{p_{oi}}\right)\right] \times 100$$

Merit: It satisfies unit test.

Demerit: It does not satisfy Time Reversal Test, Factor Reversal Test and Circular Test.

(iii) Weighted geometric mean of relatives:

$$P_{01} = \left[\text{Antilog}\left\{\frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \log\left(\frac{p_{1i}}{p_{0i}}\right)\right\}\right] \times 100$$

Merit: It satisfies unit test and Time Reversal Test.

Demerit: It does not satisfy Factor Reversal Test and Circular Test.

Some special cases and their merits, demerits:

(i) Laspeyre's index number:

$$P_{01}^{La} = \frac{\sum_{i=1}^{n} p_{1i} q_{0i}}{\sum_{i=1}^{n} p_{0i} q_{0i}} \times 100$$

Merit: It satisfies unit test.

Demerit: It does not satisfy Time Reversal Test, Factor Reversal Test and Circular Test. Under increasing trend of the prices of commodities, it is likely to over estimate the relative change.

(ii) Paasche' index number :

$$P_{01}^{Pa} = \frac{\sum_{i=1}^{n} p_{1i} q_{0i}}{\sum p_{0i} q_{1i}} \times 100$$

Merit: It satisfies unit test.

Demerit: It does not satisfy Time Reversal Test, Factor Reversal Test and Circular Test. Under increasing trend of the prices of commodities, it is likely to under estimate the relative change.

(iii) Fisher's Ideal index number:

$$P_{01}^{Id} = \sqrt{\frac{\sum_{i=1}^{n} p_{1i} q_{0i}}{\sum_{i=1}^{n} p_{0i} q_{0i}} \times \frac{\sum_{i=1}^{n} p_{1i} q_{1i}}{\sum p_{0i} q_{1i}}} \times 100$$

Merit: It satisfies unit test, Time Reversal Test and Factor Reversal Test. It is the geometric mean of Laspeyre's and Paasche's index numbers. So the over estimation due to Laspeyre and under estimation due to Paasche are likely to cancel out each other giving the true level of relative change.

Demerit: It does not satisfy Circular Test.

7. Derive the standard deviation of binomial distribution.

Ans. Variance of Binomial distribution is given by: $V(X) = E(X^2) - [E(X)]^2$

$$E(X) = \sum_{x=0}^{n} x.{}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x.\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!} p^x q^{n-x}$$

$$= np\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$$

$$= np(q+p)^{n-1} = np \ (\because p+q=1)$$

$$E(X^2) = \sum_{x=0}^{n} x^2.{}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} [x(x-1)+x].\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} x(x-1)\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$+ \sum_{x=1}^{n} x\frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x}$$

$$+ \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$$

$$= n(n-1)p^2(q+p)^{n-2} + np(q+p)^{n-1}$$

$$= n^2p^2 - np^2 + np \ (\because p+q=1)$$

Hence $V(X) = E(X^2) - [E(X)]^2$

$$= n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

So standard deviation of Binomial distribution

$$= \sqrt{V(X)} = \sqrt{npq}$$

8. Explain what is meant by correlation. Describe the method of scatter diagram for studying correlation. Write its a advantages and disadvantages.

Ans : Correlation between two variables may be expressed as the average effect on the value of one variable due to an average change in the value of the variable. In other words if a change in the value of a variable x results at a change in the value of another variable y which is may not be true for every individual cases but is true on an average basis then the two variables are said to be correlated. Thus correlation need not be necessarily a functional relationship but every functional relationship is a case of correlation. Hence in order to study the correlation between two variables x and y, it is necessary to get data on x and y for a set of n observations. Such a data presenting the values of each observation as an order paid of the form (x_i, y_i) is called a bivariate data. In a bivariate data, x_i = value of the variable x for the i^{th} observation and y_i = value of the variable y for the i^{th} observation.

Correlation between two variables can be studied by three methods such as (i) Graphical method, (ii) Product moment method and (iii) Method of rank correlation.

According to the grapical method the ordered pairs (x_i, y_i) (where i = 1, 2, 3 ..., n) are plotted on a coordinate plane. The set of points so obtained on the plane is called a scatter diagram. The correlation between the variables is obtained by examining the general direction of movement of the scatter diagram observed from left to right.

Positive Correlation : If the scatter diagram a general behavior of movement from the left hand bottom corner towards the right hand top corner, the variables are positively correlated i.e. increase in the value of one variable results at an increase in the value of the other variable.

Negative Correlation : If the scatter diagram shows a general behavior of movement from the left hand top corner towards the right hand bottom corner of the plane, the variables are negatively correlated i.e increase in the value of one variable results at a decrease in the value of the other variable.

Zero Correlation : If the scatter diagram does not show any general behavior of movement either from the left hand bottom corner towards the right hand top corner or from the left hand top corner towards the right hand bottom corner of the plane, then the variables have Zero correlation which indicates that the variables have no linear relationship.

Perfect Positive Correlation : If the points of the scatter diagram lie on a straight line making an angle of 45° with the positive X-axis then the variables are perfectly positively correlated i.e. increase in the value of one variable result at an equal amount of increase in the value of the other variable.

Perfect Negative correlation : If the points of the scatter diagram lie on a straight line making an angle of 135° with the positive X-axis then the variables are perfectly negatively correlated i.e. increase in the value of one variable results at an equal amount of decrease in the value of the other variable.

Advantages :

The most important advantage of the method is that it is easy to understand and simple to apply because it does not involve any mathematical calculation.

Disadvantages :

(i) It is an eye-inspection method, so it fails to provide a numerical value for the degree of relationship between the variables. As a result no comparison can be made between the degree of relationship between two pairs of variables.

(ii) It can not guarantee zero correlation between the variables.

9. Explain the meaning of the sampling distribution of a statistic. Also construct the sampling distribution of mean for a simple random sample of 3 units selected without replacement from a population of 5 units.

Ans. The sampling distribution of a statistic may be defined as the set of the values of the statistic computed from all possible samples of some particular type from a population. For example let the population size be N and a simple random sample of n units is to be drawn from it without replacement. Then the number of possible sample that can be drawn = ${}^NC_n$. Let the statistic be sample mean. Then sample mean has ${}^NC_n$ possible values. This set of ${}^NC_n$ possible values of the sample mean is called the sampling distribution of mean.

Let the population units be Y_1, Y_2, Y_3, Y_4, Y_5. Taking simple random samples from the population without replacement, the number of possible samples = ${}^5C_3$= 10. The following table presents the sampling distribution of mean.

Sampling Distribution of Mean

Sample Number	Units included	Sample mean
1	Y_1, Y_2, Y_3	$\frac{Y_1+Y_2+Y_3}{3}$
2	Y_1, Y_2, Y_4	$\frac{Y_1+Y_2+Y_4}{3}$
3	Y_1, Y_2, Y_5	$\frac{Y_1+Y_2+Y_5}{3}$
4	Y_1, Y_3, Y_4	$\frac{Y_1+Y_3+Y_4}{3}$
5	Y_1, Y_3, Y_5	$\frac{Y_1+Y_3+Y_5}{3}$
6	Y_1, Y_4, Y_5	$\frac{Y_1+Y_4+Y_5}{3}$
7	Y_2, Y_3, Y_4	$\frac{Y_2+Y_3+Y_4}{3}$
8	Y_2, Y_3, Y_5	$\frac{Y_2+Y_3+Y_5}{3}$
9	Y_2, Y_4, Y_5	$\frac{Y_2+Y_4+Y_5}{3}$
10	Y_3, Y_4, Y_5	$\frac{Y_3+Y_4+Y_5}{3}$

SET – 2

Time – 3 Hours **Full Marks – 70**

The figures in the right - hand margin indicate marks.

GROUP – A

1. Pick up the correct answer : **[1 × 10 = 10**

(a) If the regression lines are perpendicular to each other then what is the value of the correlation coefficient ?

(i) – 1 (ii) 0
(iii) 1 (iv) ∞

(b) Which component of the time series repeats at an interval of more than one year ?

(i) Irregular fluctuation
(ii) Cyclical fluctuation
(iii) Seasonal variation (iv) None of these

(c) Under the method of regression, the limits for the product of the regression coefficients are _____

(i) – 1 & 1 (ii) 0 & 1
(iii) – ∞ & ∞ (iv) None of these

(d) Which of the following pairs can not represent the mean and variance of a binomial distribution respectively?

(i) (3, 9/5) (ii) (12, 6)
(iii) (5,10/3) (iv) (5, 15/4)

(e) A downward bias in the relative change is observed in the index number computed by _____ method?

(i) Simple Geometric mean of relatives method
(ii) Fisher's Ideal method
(iii) Laspeyre's method
(iv) Paasche's method

(f) If the standard deviation of Poisson distribution is 2.3 then what is its mode ?

(i) 2 (ii) 2 & 3
(iii) 3 (iv) 5

(g) In simple random sampling without replacement, the number of samples, each of size n, that can be formed from a population of size N is:

(i) N^n (ii) n^N
(iii) $\binom{N}{n}$ (iv) Nn

(h) If the total area under a normal curve with mean 2000 and SD 480 is 100%, then what is the probability of the random variable taking a value larger than the median?

(i) 50 % (ii) 25 %
(iii) cannot be determined without table
(iv) 100 %

(i) The formula that over estimates an Index Number is:

(i) Laspeyre's formula (ii) Paasche's formula
(iii) Fisher's formula
(iv) Arithmetic mean of Laspeyre's and Paasche's formulae

(j) The description of all the sampling units is called________

(i) Sample size (ii) Sampling schedule
(iii) Sampling frame (iv) Stratified sampling

2. (I) correct the statements if necessary [1 × 5 = 5

(a) In time series, cyclical variation means long-term movement.

(b) In India the unit of Wholesale Price Index Number is Rupees.

(c) The value of rank correlation coefficient lies between – 1 and + 1.

(d) Binomial distribution is always positively skewed.

(e) Sampling cannot contain Non-sampling error.

(II) Fill in the blanks: [1× 5 = 5

(f) For a sample of size *n* units from a population of N sampling units, the sampling fraction is ______.

(g) If mean of a Poisson distribution is 16 then its standard deviation is ______.

(h) The type of correlation between yearly number of births and annual production of electricity in India ______

(i) Increase in the sale of fire crackers during Diwali is an example of ______.

(j) To find index number of prices by using Laspeyre's formula the weights are ______

GROUP – B

3. Answer any ten of the following questions [2 × 10 = 20

(a) Mention any two properties of regression coefficients

(b) Write the normal equations for fitting the line of regression of y on x.

(c) What are the requirements for selection of a base year in fixed base method ?

(d) Write a short note on irregular fluctuation as a component of time series.

(e) Write the probability mass function of a binomial distribution with parameters (n=10, $p = \frac{1}{4}$) and give its standard deviation.

(f) Give one merit and one demerit of moving averages method.

(g) Explain circular test.

(h) What is the mathematical condition for TRT ?

(i) Prove that the probability of selection of any unit of the population in simple random sampling with replacement and that without replacement remains constant.

(j) Describe the area property of Normal distribution.

(k) Prove that the mean and standard deviation of a standard normal distribution are 0 and 1 respectively.

(l) Examine Fisher's ideal index for TRT and FRT.

4. Answer any three of the following questions. [3 × 3 = 9

(a) Explain how trend values can be obtained by graphical method ?

(b) Index numbers are called "Economic Barometers". Give reasons.

(c) Derive the formula for Spearman's rank correlation coefficient.

(d) Derive the mode of the distribution with probability mass function, B(X; n, p).

(e) Describe the advantages of sampling over complete enumeration.

GROUP – C

Answer any three of the following questions [7 × 3 = 21

5. Discuss the semi average method for fitting trend to a time series.

6. What is meant by a Base Year? How the Base Year can be selected for construction of an index number?

7. Derive Poisson distribution as a limiting case of binomial distribution.

8. What is the need for rank correlation method? Derive the formula for Spearman's Rank correlation coefficient for n pairs of observations without any repetition of ranks.

9. Derive the variance of sample mean in simple random sampling.

ANSWERS TO SET – 2

GROUP – A

1. **(a)** (ii) 0
   (b) (iii) Seasonal variation
   (c) (ii) 0 & 1
   (d) (i) (3, 9/5)
   (e) (iv) Paasche's method
   (f) (iv) 5
   (g) (iii) $\binom{N}{n}$
   (h) (i) 50 %
   (i) (i) Laspeyre's formula
   (j) (iii) Sampling frame

2.(I) (a) In time series, trend means long-term movement.

(b) Wholesale Price Index Number has no unit.

(c) The value of rank correlation coefficient lies between – 1 and + 1 (Correct statement)

(d) Poisson distribution is always positively skewed.

(e) Sampling cannot contain both sampling error as well as non-sampling error.

(II) (f) For a sample of size n units from a population of N sampling units, the sampling fraction is $\frac{n}{N}$

(g) If mean of a Poisson distribution is 16 then its standard deviation is 4

(h) The type of correlation between yearly number of births and annual production of electricity in India is Spurious correlation or non-sense correlation .

(i) Increase in the sale of fire crackers during Diwali is an example of seasonal variation.

(j) To find index number of prices by using Laspeyre's formula the weights are Quantity of the commodities consumed in the base period.

GROUP – B

3.(a) Mention any two properties of regression coefficients

Ans. Two properties of regression coefficients are: (i) Both the regression coefficients and the correlation coefficient are always of the same sign. (ii) The product of the regression coefficients lies between 0 and 1 i.e $0 \le b_{yx} b_{xy} \le 1$.

(b) Write the normal equations for fitting the line of regression of y on x.

Ans: The normal equations for fitting the line of regression of y on x are:

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2$$

(c) What are the requirements for selection of a base year in fixed base method ?

Ans. The requirements of a base year in fixed base method are as follows :

(i) It should be a normal year from economic point of view that is the price level in the base year should be neither high nor low.

(ii) The base year should not be too distant from the current year.

(iii) The base year should be free from natural calamities and activities such as war, strike etc.

(d) Write a short note on irregular fluctuation as a component of time series.

Ans. In analysis of time series, it is observed that the value at some point(s) of time become unusually high or low as compared to other observations. Such phenomena are due to factors which have erratic behaviour or due to unexpected events. These fluctuations in the time series are termed as Irregular fluctuation or Random component.

Irregular fluctuation in a time series occur due to two types of causes namely Natural causes and Socio-economic causes.

Examples of natural factors that may give rise to irregular fluctuation are: Earthquake, Cyclone, Flood etc.

The socio-economic factors responsible for irregular fluctuation are: War, Strike, Lock-out, Communal riot etc.

(e) Write the probability mass function of a binomial distribution with parameters (n=10, $p = \frac{1}{4}$) and give its standard deviation.

Ans. The probability mass function of a binomial distribution with parameters (n=10, p = $\frac{1}{4}$) is given by:

$B(X; 10, \frac{1}{4}) = {}^{10}C_r (\frac{1}{4})^r (\frac{3}{4})^{10-r}$. The standard deviation of the distribution is $\sqrt{10 \times \frac{1}{4} \times \frac{1}{3}} = \frac{\sqrt{30}}{4}$

(f) Give one merit and one demerit of moving averages method.

Ans. One merit of moving average method is that it completely eliminates the effect due to cyclical fluctuation when the cycles occur at regular intervals.

One demerit of the moving average method is that it does not use data completely that is it fails to provide trend values for all the points of time given in the series.

(g) Explain circular test.

Ans. Circular test is conducted by considering more than two periods of comparison. Considering three periods of comparison denoted by a, b and c, an index number formula is said to satisfy circular test whenever it satisfies the condition, . Similarly if the periods considered are 0, 1, 2, 3, . . . , n then the condition becomes

(h) What is the mathematical condition for TRT ?

Ans. According to TRT,

$$P_{01} \times P_{10} = 1$$

(i) Prove that the probability of selection of any unit of the population in simple random sampling with replacement and that without replacement remains constant.

Ans. Let Population size = N and the size of the simple random sample to be selected = n

Let y_i be the i^{th} unit of the population.

y_i will be included in the sample only if it gets selected in any one of the 1^{st}, 2^{nd}, . . . , n^{th} draws.

Case – I: Simple Random Sampling without replacement (SRSWOR)

Probability that y_i is selected in the 1^{st} draw $= \frac{1}{N}$

Probability that y_i is selected in the 2^{nd} draw = P (yi is selected in the 2^{nd} draw given that some other it was not selected in the 1^{st} draw)

$$\left(1-\frac{1}{N}\right)\frac{1}{N-1} = \frac{1}{N}$$

Similarly in each of the n draws, the probability of selection of y_i becomes $\frac{1}{N}$

So the probability that y_i is included in the sample

$$= \frac{1}{N} + \frac{1}{N} + \frac{1}{N} + ...n \text{ times} = \frac{n}{N}$$

Case – II: Simple Random Sampling with replacement (SRSWR)

Probability that y_i is selected in the 1^{st} draw $= \frac{1}{N}$

Probability that y_i is selected in the 2^{nd} draw $= \frac{1}{N}$

Similarly in each of the n draws, the probability of selection of y_i becomes $\frac{1}{N}$

So the probability that y_i is included in the sample

$$= \frac{1}{N} + \frac{1}{N} + \frac{1}{N} + ...n \text{ times} = \frac{n}{N}$$

Thus it is proved that the probability of selection of any unit of the population into the sample remains constant as $\frac{n}{N}$

(j) Describe the area property of Normal distribution.

Ans. (1) Total area nuder the normal curve
$= P(-\infty < X < \infty) = 1, X \sim N(\mu, \sigma)$
Total area under the standard normal curve

$$= P(-\infty < Z < \infty) = 1 \left(Z = \frac{x-\mu}{\sigma}\right)$$

(2) $P(-\infty < X \le \mu) = P(\mu \le X < \infty) = 0.5$
$P(-\infty < Z \le 0) = P(0 \le Z < \infty) = 0.5$

(3) $P(X \le t) = 1 - P(X \ge t)$

(4) $P(X \le t) = P(X \ge t)$ (where $t \ge 0$)

(5) $P(\mu - \sigma \le X \le \mu + \sigma) = P(-1 \le Z \le 1) = 0.6826$

(6) $P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(-2 \le Z \le 2) = 0.9544$

(7) $P(\mu - 3\sigma \le X \le \mu + 3\sigma) = P(-3 \le Z \le 3) = 0.9973$

(8) $P(\mu - 1.96\sigma \le X \le \mu + 1.96\sigma)$
$= P(-1.96 \le Z \le 1.96) = 0.95$

(9) $P(\mu - 2.58\sigma \le X \le \mu + 2.58\sigma)$
$= P(-2.58 \le Z \le 2.58) = 0.99$

(k) Prove that the mean and standard deviation of a standard normal distribution are 0 and 1 respectively.

Ans. If $X \sim N(\mu, \sigma)$ then $Z = \frac{X-\mu}{\sigma}$ is called a standard normal variate.

Here mean of $X = \mu$ and variance of $X = \sigma^2$

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

$\Rightarrow E(X) = E(\mu + Z\sigma) \Rightarrow \mu = E(\mu) + \sigma E(Z)$

$\Rightarrow \mu = \mu + \sigma E(Z) \Rightarrow \sigma E(Z) = 0 \Rightarrow E(Z) = 0$

Further $X = \mu + Z\sigma \Rightarrow V(X) = V(\mu + Z\sigma)$

$\Rightarrow \sigma^2 = \sigma^2 V(Z) \Rightarrow V(Z) = 1$

Hence standard deviation of Z = 1.

So Mean and standard deviation of a standard normal variate are 0 and 1 respectively.

(l) Examine Fisher's ideal index for TRT and FRT.

Ans. Test for TRT :

$$P_{01}^{Id} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$P_{10}^{Id} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$\therefore P_{01}^{Id} \times P_{10}^{Id} = 1$$

So Fisher's ideal index satisfies TRT.

Test for FRT :

$$P_{01}^{Id} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$Q_{01}^{Id} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$\therefore P_{01}^{Id} \times Q_{01}^{Id} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \text{True value ratio}$$

Hence Fisher's ideal index satisfies FRT.

4. (a) Explain how trend values can be obtained by graphical method ?

Ans. According to graphical method, the given values of the time series are plotted against the respective time periods and the points plotted are joined serially by straight lines. The diagram so obtained is called the historigram. The trend for the given time series can be determined by drawing a free hand smooth curve drawn through the historigram in order to represent the general behavior of the time series data to increase or decrease with respect to time. A suitable trend curve should satisfy the following properties.

(i) The trend curve should be drawn in such a way that there should be equal number of points of the historigram lying on both sides of the trend curve.

(ii) It should be such that the sum of deviations of the points of the historigram taken parallel to the y-axis should be zero.

(iii) For the trend curve, the sum of squares of deviations parallel to the y-axis should be the minimum.

After drawing a suitable trend curve, the y-coordinates of the points on the trend curve corresponding to various points of time provide the required trend values for the given time series.

(b) Index numbers are called "Economic Barometers". Give reasons.

Ans. Index Numbers are popularly used to measure the relative changes in characteristics like Price, Sales, Income etc which are economic characteristics. Using Index Numbers, the wage, DA and other allowances are adjusted so that people shall become relaxed from economic pressure. It is known that Barometer is a device that measures the atmospheric pressure. Similarly, Index Numbers are used to indicate the pressure on the population due to the economic policies and the economic conditions of the state. So Index Numbers can be rightly called Economic Barometers.

(c) Derive the formula for Spearman's rank correlation coefficient.

Ans. Spearman's rank correlation coefficient is defined as the simple correlation coefficient between the sets of ranks assigned to various observations of a bivariate data. If a bivariate data has n observations of the form $P_i(A_i, B_i)$ and no two values in any of the two series are identical, then the ranks assigned to the observations in a fixed order (ascending or descending) corresponding to each characteristic are 1, 2, 3. ..., n. However the series of ranks may not be identical. Let us denote the ranks of P_i according to characteristics A and B as x_i and y_i respectively.

Then

$$\sum_{i=1}^{n} x_i = 1+2+3+...+n = \sum_{i=1}^{n} y_i = \frac{n(n+1)}{2} \quad .... \text{(i)}$$

$$\sum_{i=1}^{n} x_i^2 = 1^2+2^2+3^3+...+n^2$$

$$= \sum_{i=1}^{n} y_i^2 = \frac{n(n+1)(2n+1)}{6} \quad ... \text{(ii)}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2} = \overline{y} \quad ... \text{(iii)}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2 = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \left(\frac{n+1}{2}\right)\left(\frac{2n+1}{3} - \frac{n+1}{2}\right)$$

$$= \left(\frac{n+1}{2}\right)\left(\frac{n-1}{6}\right)\left(\frac{n^2-1}{12}\right) = \sigma_y^2 \quad ... \text{(iv)}$$

Let the difference between the ranks of P_i be denoted by $d_i = x_i - y_i$

$$d_i^2 = (x_i - y_i)^2 \Rightarrow d_i^2 = x_i^2 + y_i^2 - 2x_i y_i$$

$$\Rightarrow \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} x_i y_i$$

$$\Rightarrow 2\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} d_i^2$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \frac{\sum_{i=1}^{n} x_i^2}{2} + \frac{\sum_{i=1}^{n} y_i^2}{2} - \frac{\sum_{i=1}^{n} d_i^2}{2}$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \frac{1}{2}$$

$$\times\left[\frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)(2n+1)}{6}\right] - \frac{\sum_{i=1}^{n} d_i^2}{2}$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \frac{n(n+1)(2n+1)}{6} - \frac{\sum_{i=1}^{n} d_i^2}{2} \quad ... \text{(v)}$$

Hence Spearman's Rank Correlation Coefficient

$$= R = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

$$= \frac{\frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6} - \frac{\sum_{i=1}^{n} d_i^2}{2}\right] - \left(\frac{n+1}{2}\right)\left(\frac{n+1}{2}\right)}{\sqrt{\left(\frac{n^2-1}{12}\right)\left(\frac{n^2-1}{12}\right)}}$$

$$= \frac{\left(\frac{n+1}{2}\right)\left[\frac{2n+1}{3} - \frac{n+1}{2}\right] - \frac{\sum_{i=1}^{n} d_i^2}{2n}}{\left(\frac{n^2-1}{12}\right)}$$

$$= \frac{\left(\frac{n+1}{2}\right)\left(\frac{n-1}{6}\right) - \frac{\sum_{i=1}^{n} d_i^2}{2n}}{\left(\frac{n^2-1}{12}\right)}$$

$$= \frac{\left(\frac{n^2-1}{12}\right)}{\left(\frac{n^2-1}{12}\right)} - \frac{12\sum_{i=1}^{n} d_i^2}{2n(n^2-1)} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

(d) Derive the mode of the distribution with probability mass function, B(X; n, p).

Ans. The probability mass function B(X; n, p) is given by: $P(X = x) = \binom{n}{r} p^r q^{n-r}$

Where p + q = 1 and r = 0, 1, 2, . . . , n

Mode of any probability distribution may be defined as the value of the random variable having the maximum probability.

So let k be the mode of the given probability distribution. Then k is a positive integer such that $0 \leq k \leq n$ and $P(X = k)$ is the maximum.

Hence $P(X = k) \geq P(X = k+1)$ .... (i)

and $P(X = k) \geq P(X = k-1)$ .... (ii)

From (i) we have $P(X = k) \geq P(X = k+1)$

$$\Rightarrow \frac{P(X=k)}{P(X=k+1)} \geq 1 \Rightarrow \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k+1} p^{k+1} q^{n-k-1}} \geq 1$$

$$\Rightarrow \frac{\frac{n!}{k!(n-k)!} p^k q^{n-k}}{\frac{n!}{(k+1)!(n-k-1)!} p^{k+1} q^{n-k-1}} \geq 1$$

$$\Rightarrow \left(\frac{k+1}{n-k}\right) \times \frac{q}{p} \geq 1$$

$$\Rightarrow kq + q \geq np - kp$$

$$\Rightarrow kq + kp \geq np - q$$

$$\Rightarrow k(q+p) \geq np - 1 + p$$

$$\Rightarrow k \geq (n+1)p - 1$$

$$\Rightarrow (n+1)p - 1 \leq k \quad .... \text{(iii)}$$

Further from (ii) we have

$$P(X=k) \geq P(X=k-1)$$

$$\Rightarrow \frac{P(X=k)}{P(X=k-1)} \geq 1$$

$$\Rightarrow \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k-1} p^{k-1} q^{n-k+1}} \geq 1$$

$$\Rightarrow \frac{\frac{n!}{k!(n-k)!} p^k q^{n-k}}{\frac{n!}{(k-1)!(n-k+1)!} p^{k-1} q^{n-k+1}} \geq 1$$

$$\Rightarrow \left(\frac{n-k+1}{k}\right) \times \frac{p}{q} \geq 1$$

$$\Rightarrow np - kp + p \geq kp$$

$$\Rightarrow (n+1)p \geq kq + kp$$

$$\Rightarrow (n+1)p \geq k(q+p)$$

$$\Rightarrow k \leq (n+1)p \quad .... \text{(iv)}$$

Combining (iii) and (iv) it is clear that mode of the given distribution is that positive integer k which satisfies the condition $\Rightarrow (n+1)p - 1 \leq k \leq (n+1)p$

Hence if $(n+1)p$ is an integer, then the distribution has two modes $(n+1)p$ and $(n+1)p - 1$.

On the other hand if $(n+1)$p is not an integer, then the distribution has only one mode which is the integral part of $(n+1)p$.

(e) Describe the advantages of sampling over complete enumeration.

Ans. The advantages of sampling over complete enumeration are :

(i) In sampling, data are collected from lesser number of units.

(ii) Sampling is cheaper as compared to census.

(iii) Sampling can be done in less time.

(iv) Sampling provides a wider scope of study.

(v) Sampling can be undertaken with a smaller manpower.

(vi) Sampling can provide the precision of the estimate.

(vii) In the case of having the chance of damage of a machine or loss of a life during experimentation, the study can be conducted only by the method of sampling.

GROUP – C

5. Discuss the semi average method for fitting trend to a time series.

Ans. According to this method, the observations in the time series are divided into two groups each containing equal number of consecutive observations. However if the series contains odd number of observations, the middle observation is excluded and the rest of the series is divided into two equal groups. The arithmetic means of the two groups are determined separately. These two averages are called the semi averages. The semi average of each group is plotted against the mid point of time of the corresponding group. The two points so obtained are joined by a straight line to form the trend of the given time series. The trend values of different points of time are the ordinates of the points corresponding to the respective points of time given in the series.

Merits:

(i) The method does not require complicated mathematical calculations so it is easy to understand and simple to apply.

(ii) It is an objective method because every one gets the same trend for the same time series.

(iii) It is capable of being used for forecasting the trend values because the trend is a straight line and so can be extended further.

(iv) It uses data completely i.e. it can provide trend values for each point of time in the series.

Demerits:

(i) The method presupposes the existence of a linear trend in every time series which may not be true in all cases.

(ii) It is not capable of giving non-linear trends.

(iii) It has the demerit of being unduly influenced by larger values in the time series. So the trend is more likely to contain effects due to irregular fluctuations.

(iv) Removing the middle observation in case of data having odd number of observations is purely arbitrary.

Uses:

This method can be used to measure the trend of a time series with minimum effort particularly when it is known to have a linear trend.

6. What is meant by a Base Year? How the Base Year can be selected for construction of an index number ?

Ans. The base year is that year from among the past years with whose price level the current year's price level is to be compared. Index numbers can be constructed by two methods on the basis of the selection of the base year. They are (i) Fixed base method and (ii) Chain base method.

According to the fixed base method, a specific year from among the past years is taken as the base year and the price level of all other years are compared with the price level of that base year. Hence an ideal base year under fixed base method should have the following characteristics.

(a) It should be a normal year from economic point of view i.e the price level should be neither very high nor very low during the base year.

(b) It should be free from irregular fluctuations such as flood, cyclone, earthquake, war etc which are likely to cause economic instability.

(c) The base year should not be far away from the current year because in such cases some commodities might have become or some new commodities might have entered into use duerto scientific developments.

Thus fixed base method can be used to compare the price level of a year with that of a recent past year.

According to chain base method, the year preceding the current year is taken as the base year. Hence this method can be useful for comparing the price level of a year with a distant past year.

7. Derive Poisson distribution as a limiting case of binomial distribution.

Ans. Poisson distribution is a limiting case of binomial distribution under the following conditions:

(i) $n \to \infty$ i.e the number of independent Bernoullian trials is large.

(ii) $p \to 0$ i.e the constant probability of success at each trial is very small.

(iii) $np = \lambda$ is a constant

Hence the probability generating function of a Poisson distribution will be:

$$P(X;\lambda) = \lim_{\substack{n\to\infty \\ p\to 0}} B(X;n,p) = \lim_{\substack{n\to\infty \\ p\to 0}} {}^nC_x p^x q^{n-x}$$

$$= \lim_{\substack{n\to\infty \\ p\to 0}} \frac{n!}{x!(n-x)!}\left(\frac{np}{n}\right)^x\left(1-\frac{np}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!}\lim_{n\to\infty}\left[\frac{n}{n}\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-x+1}{n}\right)\right]$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!}\lim_{n\to\infty}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\right]$$

$$e^{-\lambda}\times 1 = e^{-\lambda}\frac{\lambda^x}{x!}$$

Thus the probability generating function of a Poisson distribution is given by:

$$P(X;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}(x = 0,1,2,\cdots)$$

8. What is the need for rank correlation method? Derive the formula for Spearman's Rank correlation coefficient for n pairs of observations without any repetition of ranks.

Ans. The method of rank correlation is useful for determination of correlation between attributes which are characteristics not capable of being measured directly. For example qualification and work efficiency are expected to be related but neither qualification nor work efficiency is measurable. So in such cases the correlation between these two attributes can be studied through the method rank correlation.

Spearman's rank correlation coefficient may be defined as the simple correlation coefficient between two sets of ranks assigned two the observations on the basis of the attributes under consideration. Assuming that no two observations have the same level of any attribute, the ranks of the n given observations will be 1, 2, 3, …, n for each attribute. However these ranks of the observations may not be in the same order.

Let x_i = rank of the i^{th} observation according to attribute x and y_i = rank of the i^{th} observation according to attribute y. Then Spearman's rank correlation coefficient will be the simple correlation coefficient between the pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

It is clear that

$$\sum_{i=1}^{n} x_i = 1+2+3+\cdots+n = \frac{n(n+1)}{2}$$

$$= \sum_{i=1}^{n} y_i \Rightarrow \bar{x} = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{(n+1)}{2} = \bar{y}$$

$$\sum_{i=1}^{n} x_i^2 = 1^2 + 2^2 + 3^2 + \cdots + n^2$$

$$= \frac{n(n+1)(2n+1)}{6} = \sum_{i=1}^{n} y_i^2$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \left(\frac{n+1}{2}\right)\left(\frac{4n+2-3n-3}{6}\right)$$

$$= \left(\frac{n+1}{2}\right)\left(\frac{n-1}{6}\right) = \left(\frac{n^2-1}{12}\right) = \sigma_y^2$$

Let $d_i = x_i - y_i$

= rank difference for the i^{th} observation

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (x_i - y_i)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} x_i y_i$$

$$= \frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)(2n+1)}{6} - 2\sum_{i=1}^{n} x_i y_i$$

$$\Rightarrow 2\sum_{i=1}^{n} x_i y_i = 2 \times \frac{n(n+1)(2n+1)}{6} - \sum_{i=1}^{n} d_i^2$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \frac{n(n+1)(2n+1)}{6} - \frac{1}{2}\sum_{i=1}^{n} d_i^2$$

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}$$

$$= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \frac{1}{2n}\sum_{i=1}^{n} d_i^2 - \left(\frac{n+1}{2}\right)^2$$

$$= \left(\frac{n+1}{2}\right)\left(\frac{4n+2-3n-3}{6}\right)$$

$$- \frac{1}{2n}\sum_{i=1}^{n} d_i^2 = \left(\frac{n^2-1}{12}\right) - \frac{1}{2n}\sum_{i=1}^{n} d_i^2$$

Hence Spearman's rank correlation coefficient

$$= R = \frac{\text{cov}(x,y)}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{\left(\frac{n^2-1}{12}\right) - \frac{1}{2n}\sum_{i=1}^{n} d_i^2}{\sqrt{\left(\frac{n^2-1}{12}\right)\left(\frac{n^2-1}{12}\right)}}$$

$$= \frac{\left(\frac{n^2-1}{12}\right)}{\left(\frac{n^2-1}{12}\right)} - \frac{\frac{1}{2n}\sum_{i=1}^{n} d_i^2}{\left(\frac{n^2-1}{12}\right)} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

9. Derive the variance of sample mean in simple random sampling.

Ans. In simple random sampling

Sample mean = $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and

Population mean = $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Further, it is known that sample mean is an unbiased estimate of the population mean i.e.

Sample mean = E $(\bar{y}) = \bar{Y}$

Let us define

$$\alpha_i = \begin{cases} 1 \textit{ if } Y_i \textit{ is included in the sample} \\ 0 \textit{ if } Y_i \textit{ is not included in the sample} \end{cases}$$

So Variance of Sample mean = V$(\bar{y})$

= E $[(\bar{y}) - E(\bar{y})^2]$ = E $[(\bar{y}) - \bar{Y}]^2$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n} y_i - \bar{Y}\right]^2 = \frac{1}{n^2} E\left[\sum_{i=1}^{n} y_i - n\bar{Y}\right]^2$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} (y_i - \bar{Y})\right]^2$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2\right] + \frac{1}{n^2} E\left[\sum_{i=1}^{n}\sum_{j \neq i}^{n} (y_i - \bar{Y})(y_j - \bar{Y})\right]$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{N} \alpha_i (y_i - \bar{Y})^2\right]$$

$$+\frac{1}{n^2}E\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})E(\alpha_i\alpha_j)\right]$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2E(\alpha_i)\right]$$

$$+\frac{1}{n^2}E\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})E(\alpha_i\alpha_j)\right]$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2\left\{1\times\frac{n}{N}+0\times\left(1-\frac{n}{N}\right)\right\}\right]$$

$$+\frac{1}{n^2}E\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})\right.$$

$$\left.\left\{1\times\frac{n(n-1)}{N(N-1)}+0\times\left(1-\frac{n(n-1)}{N(N-1)}\right)\right\}\right]$$

$$=\frac{1}{nN}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2\right]+\frac{1}{n^2}$$

$$\times\frac{n(n-1)}{N(N-1)}\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})\right]$$

$$=\frac{1}{nN}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2\right]+\frac{(n-1)}{nN(N-1)}$$

$$\left[\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})\right]$$

$$=\frac{1}{nN}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{(n-1)}{nN(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})\right]$$

$$=\frac{1}{nN}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2-\frac{(n-1)}{(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2\right.$$

$$\left.+\frac{(n-1)}{(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{(n-1)}{(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}^{N}(y_i-\bar{Y})(y_j-\bar{Y})\right]$$

$$=\frac{1}{nN}\left[\left(1-\frac{n-1}{N-1}\right)\sum_{i=1}^{N}(y_i-\bar{Y})^2\right.$$

$$\left.+\frac{n-1}{N-1}\left\{\sum_{i=1}^{N}(y_i-\bar{Y})^2+\sum_{i=1}^{N}\sum_{j\neq 1}^{N}(y_i-\bar{Y})(y_j-\bar{Y})\right\}\right]$$

$$=\frac{1}{nN}\left[\left(\frac{N-n}{N-1}\right)\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{n-1}{N-1}\left\{\sum_{i=1}^{N}(y_i-\bar{Y})\right\}^2\right]$$

$$=\frac{1}{nN}\left[\left(\frac{N-n}{N-1}\right)\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{n-1}{N-1}\left\{\sum_{i=1}^{N}y_i-\sum_{i=1}^{N}\bar{Y}\right\}^2\right]$$

$$=\frac{1}{nN}\left[\left(\frac{N-n}{N-1}\right)\sum_{i=1}^{N}(y_i-\bar{Y})^2+\frac{n-1}{N-1}\{N\bar{Y}-N\bar{Y}\}^2\right]$$

$$=\frac{N-n}{nN}\times\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

$$=\left(\frac{N-n}{nN}\right)S^2=\frac{S^2}{n}\left(\frac{N-n}{N}\right)$$

$$=\frac{S^2}{n}\left(\frac{N}{N}-\frac{n}{N}\right)=\frac{S^2}{n}(1-f)$$

SET – 3

Time – 3 Hours **Full Marks – 70**

The figures in the right - hand margin indicate marks.

GROUP – A

1. Pick up the correct answer : **[1 × 10 = 10**

(a) Which of the following is true?

(i) Secular trend refers to cyclical component

(ii) The multiplicative model assumes that the value of the original data of a time series is the sum of the four components

(iii) The most important factor causing seasonal variations is change in population

(iv) Floods is an example of random component

(b) A lock-out in a factory for a month is associated with the component of a time series called as:

(i) Trend (ii) Seasonal variation

(iii) Cyclical variation (iv) Irregular variation

(c) Which of the following index numbers over estimates the relative change ?

(i) Fisher's index (ii) Laspeyre's index

(iii) Simple aggregative index

(iv) None of these

(d) Under simple random sampling, what is the unbiased estimate of population mean square?

(i) s^2 (ii) S^2

(iii) σ^2 (iv) None of these

(e) Which of the following is true?
   (i) Laspeyre's formula satisfies Time Reversal Test
   (ii) Paasche's formula satisfies unit test
   (iii) Fisher's ideal formula satisfies Circular Test
   (iv) No formula satisfies Circular Test

(f) What is the standard deviation of Binomial distributions ?
   (i) n and p (ii) npq
   (iii) $\sqrt{npq}$ (iv) $\sqrt{np}$

(g) If the correlation coefficient between X and Y is r then the correlation coefficient between 2X and 2Y will be:
   (i) 2r   (ii) r^2
   (iii) r + 2   (iv) None of these

(h) If one of the regression coefficient is larger than 1 then the other regression coefficient is smaller than 1.
   (i) True   (ii) False
   (iii) Cannot be said   (iv) None of these

(i) If the total area under a normal curve with mean 2000 and SD 480 is 100%, then what is the probability of the random variable taking a value larger than the median?
   (i) 50 %
   (ii) 25 %
   (iii) cannot be determined without table
   (iv) 100 %

(j) If X is a Poisson variate with parameter λ, then the value of $E(X^2)$ is:
   (i) λ (ii) λ^2
   (iii) $\lambda(\lambda + 1)$ (iv) $\lambda(\lambda - 1)$

2. (I) correct the statements if necessary [1 × 5 = 5

   (a) The random component of time series repeats at an interval of 12 months.
   (b) Laspeyre's index number formula satisfies Time Reversal Test.
   (c) Binomial distribution is always positively skewed.
   (d) For a homogeneous population, simple random sampling is the most suitable method.
   (e) If two variables have perfect positive correlation, then its scatter diagram will be a straight line making an angle of 135° with the positive X-axis.

(II) Fill in the blanks: [1× 5 = 5

   (f) To find index number of prices by using Laspeyre's formula the weights are ______
   (g) The formula for probable error of correlation coefficient is ______
   (h) Correlation coefficient is ______ of the regression coefficients.
   (i) If one regression coefficient is 0 then the correlation coefficient is ______.
   (j) If mean of a Poisson distribution is 16 then its standard deviation is ______.

GROUP – B

3. Answer any ten of the following questions [2 × 10 = 20

   (a) Name the components of time series.
   (b) Name the methods for measuremtn of trend.
   (c) Show that, the sum of all binomial probabilities is 1.
   (d) Write any two uses of index number.
   (e) Prove that correlation coefficient is the geometric mean of the regression coefficients.
   (f) Define Population and Sample.
   (g) What are the parameters of Normal distribution ?
   (h) Mention two requirements for selection of a base year in fixed base method.
   (i) Write the normal equations for fitting a linear trend to a time series.
   (j) An unbiased die is tossed 5 times in succession. Getting a 6 is considered a success. Find the probability that success occurs at least once.
   (k) Prove that the probability of selection of any unit of the population in simple random sampling with replacement and that without replacement remains constant.
   (l) What is the mathematical condition for FRT ?

4. Answer any three of the following questions. [3 × 3 = 9

   (a) Write the normal equations for fitting of a quadratic trend to a time series data.

(b) Differentiate between random sampling and non-random sampling.

(c) Derive Poisson distribution as a limiting case of Binomial distribution.

(d) Explain the necessity of assigning weights to commodities for the construction of index number. Also explain the different types of weights used in the construction of index number.

(e) Discuss the effect of change of origin and scale on correlation coefficient.

GROUP – C

Answer any three of the following questions **[7 × 3 = 21**

5. Define a time series giving suitable examples. Explain its various components.

6. Derive the variance of binomial distribution. Also if $X \sim B\left(9, \frac{2}{3}\right)$, find the mean and standard deviation..

7. What is Simple Random Sampling and what are the different methods by which a simple random sample can be drawn.

8. Define Index Number. Distinguish between weighted and un-weighted index numbers. Give the formulas commonly used for their computation.

9. Define regression coefficients and discuss their properties.

ANSWERS TO SET – 3

GROUP – A

1. (a) (iv) Floods is an example of random component

(b) (iv) Irregular variation

(c) (ii) Laspeyre's index

(d) (ii) Paasche's formula satisfies unit test

(e) (iii) $\sqrt{npq}$ (f) (i) s^2

(g) (iv) None of these (h) (i) True

(i) (i) 50 % (j) (iii) $\lambda(\lambda + 1)$

2. (I)(a) The seasonal component of time series repeats at an interval of 12 months.

(b) Fisher's index number formula satisfies Time Reversal Test.

(c) Poisson distribution is always positively skewed.

(d) For a homogeneous population, simple random sampling is the most suitable method. (correct statement)

(e) If two variables have perfect positive correlation, then its scatter diagram will be a straight line making an angle of 45° with the positive X-axis.

(II) (f) To find index number of prices by using Laspeyre's formula the weights are <u>Quantity of the commodities consumed in the base period.</u>

(g) The formula for probable error of correlation coefficient is PE $(r) = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$

(h) Correlation coefficient is <u>geometric mean</u> of the regression coefficients.

(i) If one regression coefficient is 0 then the correlation coefficient is <u>0</u>.

(j) If mean of a Poisson distribution is 16 then its standard deviation is <u>4</u>

GROUP – B

3. (a) Name the components of time series.

Ans. The components of time series are: (i) Trend, (ii) Seasonal variation, (iii) Cyclical fluctuation and (iv) Irregular fluctuation.

(b) Name the methods for measuremtn of trend.

Ans: The methods for measurement of trend are :

(i) Free hand curve method

(ii) Semi average method

(iii) Moving average method

(iv) Method of least squares

(c) Show that, the sum of all binomial probabilities is 1.

Ans. The probability generating function is given by: $P(X=r) = \binom{n}{r} p^r q^{n-r}$