|  |
| --- |
| **Unit IV**  **Unit Name: Basic Statistics (session 21 to session 29)** |
| **Overview:**  This unit includes Measures of Central tendency: Moments, skewness, Kurtosis. Moments, skewness and Kurtosis for Binomial distribution & Poisson distribution. Moments, skewness & kurtosis for Normal distribution. Evaluation of statistical parameters for Binomial, Poisson and Normal distributions, Correlation and regression, Rank correlation. |
| **Outcome:**  After completion of this unit, students would be able to:   1. identify suitable probability distribution to solve problems. 2. apply knowledge of random variables, probability distributions, measures of central tendency, correlation and regression to solve real life problems. |
| **Measure of central tendency**  A measure of central tendency is the value of the random variable which is representative of the entire distribution of the variable.  **Moments**  1. Raw moment: The non-central moment about any value  of random variable X is defined as  .  If  then we get raw moment or ordinary moment or simple moment.  2. Central moment: The central moment about the mean  of random variable X is defined as .  **Relation between raw and central moments**            **Skewness**  Skewness, which means lack of symmetry, is the property of a random variable or its distribution by which we get an idea about the shape of the probability curve of the distribution. If the probability curve is not symmetrical but has a longer tail on one side than on the other, the distribution is said to be skewed. If a distribution is skewed, then the averages mean, median and mode will take different values and the quartiles will not be equidistant from the median.  The measure of skewness used in common is the third order central moment ().  The moment coefficient of skewness is defined as .  **Kurtosis**  Even if we know the measures of central tendency, dispersion and skewness of a random variable (or its distribution). we cannot get a complete idea about the distribution. In order to analyze the distribution completely, another characteristic kurtosis is also required. Kurtosis means the convexity of the probability curve of the distribution. Using the measure of coefficient of kurtosis, we can have an idea about the flatness or peakedness of the probability curve near its top.  The only measure of kurtosis used is the fourth order central moment ().  The coefficient of kurtosis is defined as  .  **Note:** 1. Curve which is neither flat nor peaked is called a mesokurtic curve, for which  2. Curve which is flatter than the curve 1 is called platykurtic curve, for which  3. Curve which is more peaked than the curve 1 is called leptokurtic curve, for which  **Binomial Distribution**  If X is the discrete random variable such that its probability mass function is given by  where  Then X is said to follow a Binomial distribution. It is also denoted as .  Mean =  Variance =  Note: In general, any  order central moment for binomial distribution is given by  Skewness =  Kurtosis =  **Poisson Distribution**  If X is the discrete random variable such that its probability mass function is given by  Then X is said to follow a Poisson distribution with parameter .  Mean =  =  Variance =  =  In general, any  order central moment for Poisson distribution is given by  Skewness =  Kurtosis =  **Normal Distribution**  A continuous random variable X is said to follow normal distribution with parameter  (called mean) and  (called variance), if its probability density function is given by  where  It is also denoted as .  Note: If X is a normal variate with parameter  , then  is also a normal variate with mean  and standard deviation . It is called **Standard Normal Variate**.  Mean  Variance  In general, Central moments of odd powers of a normal distribution is zero.  Central moments of even powers is given by  Skewness = 0  Kurtosis =  **Correlation**  ***Correlation:*** Correlation is a statistical measure (expressed as a number) that describes the size and direction of a relationship between two or more variables. Two variables are said to be correlated if change in one variable affects the change in other variable, and the relation between them is known as correlation.  **Positive Correlation:**  Two variables are said to be positively correlated if they deviates in the same direction.  e.g. height & weight, income & expenditure      **Negative Correlation**  Two variables are said to be negatively correlated if they deviates in the opposite directions.  e.g. volume and pressure of a perfect gas, price and demand    **Un-correlation**  Two variables are said to be uncorrelated or statistically independent if there is no relation between them.  **Karl Pearson’s Product Moment coefficient of correlation:** Correlation coefficient between two random variables X and Y, usually denoted by  or  and defined as    Note:  If  then correlation is perfectly positive,  If  then correlation is perfectly negative,  If  then variables are uncorrelated.  **Spearman’s Rank Correlation:**  The method develop by Spearman is simpler than Karl Pearson’s method since, it depends upon ranks of the items and actual values of the items are not required. Hence this can be used to study correlation even when actual values are not known.  For instance, we can study correlation between intelligence and honesty by this method.  Spearman’s Rank Correlation coefficient is defined by      ***Regression:*** Regression can be defined as a method to estimate the value of one variable when that of other is known, when the variables are correlated. Regression analysis is a mathematical measure of average relationship between two or more correlated values.  **Equations of Lines of regression:**   1. Line of regression of is :     where regression coefficient of is given by   1. Line of regression of is :     where regression coefficient of y on x is given by  **Properties:** i) Lines of regression are passes through the point  ii)  iii) have same sign.  *References of the entire unit*  **1:** Probability, Statistics and Random Processes, T. Veerarajan, Tata McGraw Hill, 3rd edition.  **2:** Applied Mathematics, G.V. Khumbhojkar, C. Jamnadas & Co. |
| **Session 21**   1. Find the mean and standard deviation of the following probability distribution.  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X=xi | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | pi | 0.004 | 0.036 | 0.1 | 0.232 | 0.280 | 0.240 | 0.112 | 0.028 | 0.004 |   [Ans. mean=3.972, SD=1.410]   1. The first four moments of a distribution about the value 5 of the random variable X are 2, 20, 40 and 50. Compute a measure, each of central tendency, dispersion, skewness and kurtosis. Comment on the skewness and kurtosis of the distribution.   [Ans. mean=7, SD=4, skewness= - 64, Kurtosis= 162] |
| **Session 22**   1. The first three moments of a distribution about the value 2 of the random variable X are 1, 16 and - 40. Find the coefficient of skewness and Kurtosis of the distribution. 2. The first four moments of a distribution about X = 4 are 1, 4, 10, 45 respectively. Find the mean, variance, coefficient of skewness and coefficient of Kurtosis.   [Ans. 5; 3; 0; 26/9]   1. The distribution of a random variable X has mean 10, variance 16,  = 1 and . Obtain the first four simple moments of X.   [Ans. 10, 116, 1544, 23184] |
| **Session 23**   1. Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls, (ii) at least 1 boy, (iii) at most 2 girls and (iv) children of both sexes. Assume equal probabilities for boys and girls.   [Ans. 300,750,550,700]   1. Find the Binomial distribution if the mean is 4 and variance is 3. |
| **Session 24**   1. The probability that at any moment one telephone line out of 10 will be busy is 0.2. (i) What is the probability that 5 lines are busy? (ii) Find the expected number of busy lines and also find the probability of this number. (iii) What is the probability that all lines are busy? 2. Find the mean, variance, skewness and kurtosis of the probability distribution of the number of heads obtained in three flips of a balanced coin. 3. With usual notation find p of Binomial distribution if ,. Also find mean, variance, skewness and kurtosis. 4. Fit a binomial distribution for the following data:  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | x: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | Total | | f: | 5 | 18 | 28 | 12 | 7 | 6 | 4 | 80 |   Also find mean, variance, skewness and kurtosis. |
| **Session 25**   1. The number of monthly breakdowns of a computer is a RV having Poisson distribution with mean equal to 1.8. Find the probability that this computer will function for a month (a) without a breakdown, (b) with only one breakdown and (c) with at least one breakdown.   [Ans. 0.1653, 0.2975, 0.8347]   1. In a certain factory producing razor blades, there is a small chance for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing (i) no defective blade, (ii) at least 1 defective blade and (iii) atmost 1 defective blade in a consignment of 10,000 packets.   [Ans. 9802, 198, 9998]   1. A car hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as Poisson variate with mean 1.5. Calculate the proportion of days on which (i) neither car is used, (ii) some demand is refused.   [Ans. 0.2231, 0.1912] |
| **Session 26**   1. The number of accidents in a year attributed to taxi drivers in a city follows Poisson distribution with mean 3. Out of 1,000 taxi drivers, find approximately the number of drivers with (i) no accident in a year, (ii) more than 3 accidents in a year.   [Ans. 50, 577]   1. A radioactive source emits on the average 2.5 particles per second. Find the probability that 3 or more particles will be emitted in an interval of 4s.   [Ans. 0.0028]   1. If a random variable X follows Poisson distribution such that , find the mean, variance, skewness and kurtosis of the distribution. 2. Fit a Poisson distribution for the following data:  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | x: | 0 | 1 | 2 | 3 | 4 | 5 | Total | | f: | 142 | 156 | 69 | 27 | 5 | 1 | 400 |   Also find mean, variance, skewness and kurtosis. |
| **Session 27**   1. For a normal variate X with mean 25 and standard deviation 10, find the area between (i) X = 25, X = 35, (ii) X = 15, X = 35 and also the area such that, (iii)  , (iv) . Also find skewness and kurtosis. 2. The weights of 4000 students are found to be normally distributed with mean 50 kgs and standard deviation 5 kgs. Find the probability that a student selected at random will have weight a) less than 40 kgs b) between 45 and 65 kgs. 3. The incomes of a group of 10,000 persons were found to be normally distributed with mean Rs. 520 and standard deviation Rs. 60. Find (i) the number of persons having incomes between Rs. 400 and 550, (ii) the lowest income of the richest 500. |
| **Correlation**  **Session 28**   1. Calculate the correlation coefficient for the following heights (in inches) of fathers (X) and their sons(Y)  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 | | Y | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |   [**Ans :** r = 0.603**]**     1. Calculate the Karl Person’s correlation coefficient from the following data  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 28 | 45 | 40 | 38 | 35 | 33 | 40 | 32 | 36 | 33 | | Y | 23 | 34 | 33 | 34 | 30 | 26 | 28 | 31 | 36 | 35 |   [**Ans :** r = 0.5185**]**   1. Calculate the correlation coefficient from the following data  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 30 | 33 | 25 | 10 | 33 | 75 | 40 | 85 | 90 | 95 | 65 | 55 | | Y | 68 | 65 | 80 | 85 | 70 | 30 | 55 | 18 | 15 | 10 | 35 | 45 |   [**Ans :** r = - 0.9935**]**     1. A computer while calculating correction coefficient between two variables X and Y from 25 pairs of observations obtained the following results     [**Ans :** r = 0.67**]**   1. Calculate the rank correlation coefficient from the following data.  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 1 | 3 | 7 | 5 | 4 | 6 | 2 | 10 | 9 | 8 | | Y | 3 | 1 | 4 | 5 | 6 | 9 | 7 | 8 | 10 | 2 |   [**Ans :** R= 0.42**]**   1. The following table shows the marks obtained by 10 students in Accountancy and Statistics. Find the Spearman’s coefficient of rank correlation.  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Student No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | Accountancy | 45 | 70 | 65 | 30 | 90 | 40 | 50 | 57 | 85 | 60 | | Statistics | 35 | 90 | 70 | 40 | 95 | 40 | 60 | 80 | 80 | 50 |   [**Ans :** R= 0.90**]**   1. Find the coefficient of rank correlation between height of father and height of son from the following data.  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Height of father | 65 | 66 | 67 | 67 | 68 | 69 | 71 | 73 | | Height of son | 67 | 68 | 64 | 68 | 72 | 70 | 69 | 70 |  1. Ten competitors in a musical test were ranked by the three judges A, B, C in the following order-  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | A: | 1 | 6 | 5 | 10 | 3 | 2 | 4 | 9 | 7 | 8 | | B: | 3 | 5 | 8 | 4 | 7 | 10 | 2 | 1 | 6 | 9 | | C: | 6 | 4 | 9 | 8 | 1 | 2 | 3 | 10 | 5 | 7 |   Using rank correlation, discuss which pair of judge has the nearest approach in common likings in music.  **Ans: The pair of judges A and C has the nearest approach to common likings in music.** |
| **Regression**  **Session 29**   1. The following are the marks in Statistics (X) and Mathematics (Y) of ten students  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 56 | 55 | 58 | 57 | 56 | 60 | 54 | 59 | 57 | 58 | | Y | 68 | 67 | 67 | 65 | 68 | 70 | 66 | 68 | 66 | 70 |   Calculate the coefficient of correlation and estimate marks in Mathematics of a student who scored 62 marks in Statistics.  [**Ans :** r = 0.44 , Y = 69.5 **]**   1. It is given that the means of x and y are 5 and 10. If the line of regression of y on x is parallel to the line 20y = 9x + 40 , estimate the value of y at x = 30   [**Ans :** 20y = 9x + 155 and y = 175**]**   1. Find the two lines of regression from the following data  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 | | Y | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |   [**Ans :** x = 30.364 + 0.545 y and y = 23.667 + 0.667 x**]**   1. In partially destroyed laboratory record of an analysis of correlation data, the following results only are legible-   Variance of X = 9, regression equations are: 8X–10Y + 66 = 0 & 40X–18Y = 214  What was i) the mean of X and Y  ii) the correlation between X and Y  iii) the S.D. of Y  [**Ans:**  ]   1. Obtain the equation of the line of regression of **cost on age** from the following table giving the age of a car of certain make and the annual maintenance cost.  |  |  |  |  |  | | --- | --- | --- | --- | --- | | Age of car  (in years) | 2 | 4 | 6 | 8 | | Maintenance  (in thousands of Rs.) | 5 | 7 | 8.5 | 11 |   Also find maintenance cost of the car if its age is 9 years  [**Ans :** y = 3 + 0.975 x and y = Rs. 11775**]**   1. The two lines of regression are 5y-8x+17=0, 2y-5x+14=0, If  find .   Ans ( 4, 3, 4, 0.8)   1. The two lines of regression are x+2y-5=0, 2x+3y-8=0, If  find .   Ans ( 1, 2, 2, 0.87)   1. Find the correlation coefficient and hence find the two regression lines  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | x | 1 | 3 | 4 | 6 | 8 | 9 | 11 | 14 | | y | 1 | 2 | 4 | 4 | 5 | 7 | 8 | 9 |   Ans: r = 0.9770, y = 0.64x + 0.52, x = 1.5y - 0.5 |