

I'm not a robot



Karl pearson coefficient of correlation problems with solutions

It is denoted by the Greek letter ρ (rho). Short-cut method Let A = 68 , B = 69, c = 1 and d = 1 Note: The correlation coefficient computed by using direct method and short-cut method is the same. Introduction The correlation coefficient is an useful statistical tool for describing the type (positive or negative or uncorrelated) and intensity of linear relationship (such as moderately or highly) between two variables. It is denoted by 'r', where r is a pure number which means that r has no unit. By quantifying correlations. ... KARL PEARSON'S COEFFICIENT OF CORRELATION Interpreting correlation using a scatter plot shown above can be subjective. The students can also verify the results by using shortcut method. 2. Student who is best in Tamil is weakest in English subject and vice-versa. Regression analysis is widely used for prediction. Some of the non-linear relationships are displayed below. Correlation: Meaning, Significance, Types and Degree of Correlation Methods of measurements of Correlation Calculation of Correlation with Scattered Diagram Spearman's Rank Correlation Coefficient in Statistics A computer, while calculating the coefficient of correlation between X and Y, obtained the following results: N 25; 2X = 125; 2Y = 100; 2X = 650; 2Y? KARL PEARSON'S CORRELATION COEFFICIENT When there exists some relationship between two measurable variables, we compute the degree of relationship using the correlation coefficient. Galton observed that extreme characteristics (e.g., height) in parents are not passed on completely to their offspring. Yule was a prolific writer, and was active in Royal Statistical Society and received its Guy Medal in Gold in 1911, and served as its President in 1924-26.The concept of Association is due to him. In such cases the items are given the average of the ranks they would have received. Calculate the total of step deviations; i.e., [Tex]\sum\{dx\^{\prime}\prime}\ [/Tex]Take the deviations of Series Y from the assumed mean and divide them by Common Factor (C) to determine step deviation [Tex]\{dy\^{\prime}\prime}\prime}\ [/Tex]. Includes exercises and solutions for statistics students. 1. Subsequent efforts by Galton and Pearson brought many techniques of multiple regression and the product-moment correlation coefficient. He was educated at Winchester College and at University College London. NOTE 1. Solution: Repetitions of ranks In Commerce (X), 20 is repeated two times corresponding to ranks 3 and 4. The values between -1 and 0 indicate inverse relationship (association) between the attributes. Therefore, 4 is assigned for ranks 3,4 and 5 with m2=3. But there may exist non-linear relationship (curvilinear relationship). . know the uses of regression coefficients. Calculate their total; i.e., Σdy. Square the deviations of Y series and calculate their total; i.e., Σdy2. Multiply dx and dy and calculate their total; i.e., Σdxdy. Now, use the following formula to determine Coefficient of Correlation:[Tex]r=\frac{N\sum\{dxdy\}-\sum\{dx\}.\sum\{dy\}}{\sqrt{N\sum\{dx\^2\}-\sum\{dx\}\^2}}\sqrt{N\sum\{dy\^2\}-\sum\{dy\}\^2}}\} [/Tex] Where, N = Number of pair of observations Σdx = Sum of deviations of X values from assumed mean Σdy = Sum of deviations of Y values from assumed mean Σdx2 = Sum of squared deviations of X values from assumed mean Σdy2 = Sum of squared deviations of Y values from assumed mean Σdxdy = Sum of the products of deviations dx and dy Example:Use Assumed Mean Method and determine the coefficient of correlation for the following data: [Tex]r=\frac{N\sum\{dxdy\}-\sum\{dx\}.\sum\{dy\}}{\sqrt{N\sum\{dx\^2\}-\sum\{dx\}\^2}}\sqrt{N\sum\{dy\^2\}-\sum\{dy\}\^2}}\} [/Tex] [Tex]=\frac{420-(56\times21)}{\sqrt{(7\times420)-(28\times21)}}\sqrt{(7\times560)-(28\^2)}}\times\sqrt{(7\times315)-(21\^2)}}\} [/Tex] [Tex]=\frac{2,940-588}{\sqrt{(3,920-784)}\times\sqrt{(2,205-441)}}\} [/Tex] [Tex]=\frac{2,352}{\sqrt{3,136}\times\sqrt{1,764}}\}=\frac{2,352}{56\times42}\} [/Tex] [Tex]=\frac{2,352}{2,352}=1 [/Tex] Coefficient of Correlation = 1 It means that there is perfect positive correlation between the values of Series X and Series Y. [Tex]r=\frac{N\sum\{dxdy\}-\sum\{dx\}.\sum\{dy\}}{\sqrt{N\sum\{dx\^2\}-\sum\{dx\}\^2}}\sqrt{N\sum\{dy\^2\}-\sum\{dy\}\^2}}\} [/Tex] [Tex]=\frac{180\times150-(180\times105)-(180\times105)-445\times120}{\sqrt{(180\times105+445\times120)-189000+53400}}=\frac{136500}{24300}=0.561>0 [/Tex] Interpretation: There is a positive association between success in examination and attending coaching classes. Therefore, 3.5 is assigned for rank 2 and 3 with m1=2. Test for the association of attributes attending the coaching class and success in the examination. Solution: Interpretation: Degree of agreement between the referees 'A' and 'B' is 0.636 and they have "strong agreement" in evaluating the competitors. Co-variance Let (X,Y) be a bivariable normal random variable where V(X) and V(Y) exists. In Mathematics (Y), 30 is repeated three times corresponding to ranks 3, 4 and 5. Karl Pearson's coefficient of correlation When X and Y are linearly related and (X,Y) has a bivariate normal distribution, the co-efficient of correlation between X and Y is defined as This is also called as product moment correlation co-efficient which was defined by Karl Pearson. However the converse need not be true. The first person to give a mathematical formula for the measurement of the degree of relationship between two variables in 1890 was Karl Pearson. These benefits help market researchers / data analysts / data scientists to eliminate and evaluate the best set of variables to be used for building regression models for predictive purposes. Example 4.4 Calculate the Spearman's rank correlation coefficient for the following data. = 460; 2XY = 508 Later on it was ... Calculate Karl Pearson's coefficient of correlation between the length and weight and interpret the result. The youngest of nine children, he appeared as an intelligent child. Example: Age and health care are related. . A negative value of 'r' indicates negative correlation . If r = +1, then the correlation is perfect positive . If r = -1, then the correlation is perfect negative. It means that on the average, if fathers are tall then sons will probably tall and if fathers are short, probably sons may be short. In this case, a different formula is used when there is more than one item having the same value. In interpretation we use the adjective 'highly' . If X and Y are independent, then rxy = 0. Galton's progress in education was not smooth. More specifically, regression analysis helps one to understand how the typical value of the dependent variable (or 'response variable') changes when any one of the independent variables (regressors) or predictor(s) is varied, while the other independent variables are held fixed. Calculate coefficient of correlation from the following data and interpret. "Association" means that the attributes have some degree of agreement. i.e. -1 ≤ r ≤ 1 . A positive value of "r" indicates positive correlation. Pearson was beginning to work in statistics and Yule followed him into this new field. It helps to determine the impact of changes in the value(s) of the independent variable(s) upon changes in the value of the dependent variable. Example 4.7 Out of 1800 candidates appeared for a competitive examination 625 were successful; 300 had attended a coaching class and of these 180 came out successful. Interpretation The relation coefficient lies between -1 and +1. Solution: Let x denote height of father and y denote ... Explore all Karl Pearson's Correlation Coefficient related practice questions with solutions, important points to remember, 3D videos, & popular books. LEARNING OBJECTIVES The student will be able to . know the concept of regression, its types and their uses. . fit best line of regression by applying the method of least squares. . b+ bX + e Here, Y represents the dependent (response) variable, Xi represents the ith independent variable (regressor), a and bi are the regression coefficients and e is the error term. According to Karl Pearson, "Coefficient of Correlation is calculated by dividing the sum of products of deviations from their respective means by their number of pairs and their standard deviations." Formula of Karl Pearson's Coefficient of Correlation[Tex]Karl-Pearson's-Coefficient-of-Correlation(r)=\frac{Sum-of-Products-of-Deviations-from-their-respective-means}{Number-of-Pairs\times(Standard-Deviations-of-both-Series)}\} [/Tex] Or [Tex]r=\frac{sum\{xy\}}{N\times(\sigma_x\times\sigma_y)}\} [/Tex] Where, N = Number of Pair of Observations x = Deviation of X series from Mean [Tex](X-\bar{X})\} [/Tex] y = Deviation of Y series from Mean [Tex](Y-\bar{Y})\} [/Tex] [Tex]\sigma_x = Standard Deviation of X series [Tex]\sqrt{\frac{sum\{x\^2\}}{N}}\} [/Tex] [Tex]\sigma_y = Standard Deviation of Y series [Tex]\sqrt{\frac{sum\{y\^2\}}{N}}\} [/Tex] r = Coefficient of Correlation Actual Mean MethodDirect MethodShort-Cut MethodAssumed Mean MethodIndirect MethodStep-Deviation MethodThe steps involved in the calculation of coefficient of correlation by using Actual Mean Method are: The first step is to calculate the mean of the given two series (say X and Y). Now, take the deviation of X series from [Tex]\bar{X}\} [/Tex] and denote the deviations by x. Square the deviations of x and obtain the total; i.e., [Tex]\sum\{x\^2\}\} [/Tex] Take the deviation of Y series from [Tex]\bar{Y}\} [/Tex] and denote the deviations by y. Square the deviations of y and obtain the total; i.e., [Tex]\sum\{y\^2\}\} [/Tex] Multiply the respective deviations of Series X and Y and obtain the total; i.e., [Tex]\sum\{xy\}\} [/Tex]. Now, use the following formula to determine the Coefficient of Correlation:[Tex]r=\frac{sum\{xy\}}{\sqrt{sum\{x\^2\}}\times\sqrt{sum\{y\^2\}}}\} [/Tex] Example:Use Actual Mean Method and determine the coefficient of correlation for the following data: Solution:[Tex]\bar{X}=\frac{sum\{X\}}{N}=\frac{168}{7}=24 [/Tex] [Tex]\bar{Y}=\frac{sum\{Y\}}{N}=\frac{105}{7}=15 [/Tex] [Tex]r=\frac{sum\{xy\}}{\sqrt{sum\{x\^2\}}\times\sqrt{sum\{y\^2\}}}\} [/Tex] Σxy = 336, Σx2 = 448, Σy2 = 252 [Tex]r=\frac{336}{\sqrt{336}\sqrt{336}}=\frac{336}{336}=1 [/Tex] Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. This method of measuring the coefficient of correlation is the most popular and is widely used. Properties 1. Spurious correlation means an association extracted from correlation coefficient that may not exist in reality. 2x2 Contingency Table Yule's coefficient: Q = Note 1: The usage of the symbol α is not to be confused with level of significance. Solution: Let x denote marks in test-1 and y denote marks in test-2. He dabbled in medicine and then studied Mathematics at Cambridge. Actual Mean can sometimes come in fractions which can make the calculation of standard deviation complicated and difficult. Solution: Interpretation: This perfect negative rank correlation (- 1) indicates that scorings in the subjects, totally disagree. Change of Origin: If a constant is added or subtracted to the values then it will not have any effect on the value of correlation coefficient.Change of Scale: Similarly, if a constant is multiplied or divided by the values, then too it will not have any effect on the value of correlation coefficient. Example:Find the coefficient of correlation between the following figures; Solution:As the coefficient of correlation is not affected by the change in scale and origin, we will multiply the X series by 10 and divide the Y series by 100. Using the method of rank correlation determine the relationship between equity shares and preference shares prices. If the relationship is linear and there is only one independent variable, then the regression is called as simple linear regression. Types of 'Regression' Based on the kind of relationship between the dependent variable and the set of independent variable(s), there arises two broad categories of regression viz. , linear regression and non-linear regression. 4. After a year dong research in experimental physics, he returned to University College in 1893 to work as a demonstrator for Karl Pearson. Calculation Heights of father and son are positively correlated. Page 2SPEARMAN'S RANK CORRELATION COEFFICIENT If the data are in ordinal scale then Spearman's rank correlation coefficient is used. That is those who perform well in test-1 will also perform well in test-2 and those who perform poor in test-1 will perform poor in test-2. Simple Linear Regression It is one of the most widely known modeling techniques. Solution: Let x denote height of father and y denote height of son. Calculate their total; i.e., Σdx. Now, square the deviations of X series and calculate their total; i.e., Σdx2. Take the deviations of Y Series from the assumed mean and denote the values by dy. Spearman's correlation can be calculated for the subjectivity data also, like competition scores. Its interpretation is similar to that of Pearson's. Outliers (extreme observations) strongly influence the correlation coefficient. Correlation does not imply causal relationship. Spurious Correlation : The word 'spurious' from Latin means 'false' or 'illegitimate'. Therefore, Interpretation: Marks in Commerce and Mathematics are uncorrelated Page 3YULE'S COEFFICIENT OF ASSOCIATION This measure is used to know the existence of relationship between the two attributes A and B (binary complementary variables). Regression analysis is a vital statistical method for obtaining functional relationship between a dependent variable and one or more independent variables. If we see outliers in our data, we should be careful about the conclusions we draw from the value of r. 3. Step Deviation MethodThis method simplifies the calculation of coefficient of correlation as it was discovered by Karl Pearson. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete and nature of relationship is linear. An examination of publications of Sir Francis Galton and Karl Pearson revealed that Galton's work on inherited characteristics of sweet peas led to the initial conceptualization of linear regression. Spearman's rank correlation coefficient is given by the formula where Di = R1i - R2i R1i = rank of i in the first set of data R2i = rank of i in the second set of data and n = number of pairs of observations Interpretation Spearman's rank correlation coefficient is a statistical measure of the strength of a monotonic (increasing/decreasing) relationship between paired data. Coaching class is useful for success in examination. However, if we compute the linear correlation r for such data, it may be zero implying age and health care are uncorrelated, but non-linear correlation is present, where mi is the number of repetitions of ith rank Example 4.6 Compute the rank correlation coefficient for the following data of the marks obtained by 8 students in the Commerce and Mathematics. The steps involved in the calculation of coefficient of correlation by using Direct Method are: The first step is to calculate the sum of Series X (ΣX). Now, calculate the sum of Series Y (ΣY). Square the values of X Series and calculate their total; i.e., ΣX2. Square the values of Y Series and calculate their total; i.e., ΣY2. Multiply the values of Series X and Y and calculate their total; i.e., ΣXY. Now, use the following formula to determine Coefficient of Correlation: [Tex]r=\frac{N\sum(XY)-\sum(X).\sum(Y)}{\sqrt{N\sum(X\^2)-\sum(X)\^2}}\sqrt{N\sum(Y\^2)-\sum(Y)\^2}}\} [/Tex] Example:Use Direct Method and determine the coefficient of correlation for the following data: Solution:[Tex]r=\frac{N\sum(XY)-\sum(X).\sum(Y)}{\sqrt{N\sum(X\^2)-\sum(X)\^2}}\sqrt{N\sum(Y\^2)-\sum(Y)\^2}}\} [/Tex] [Tex]=\frac{7\times(2,856)-(168\times105)}{\sqrt{(7\times4,480)-(168\^2)}}\times\sqrt{(7\times1,827)-(105\^2)}}\} [/Tex] [Tex]=\frac{19,992-17,640}{\sqrt{31,360-28,224}}\times\sqrt{(12,789-11,025)}}\} [/Tex] [Tex]=\frac{2,352}{\sqrt{3,136}\times\sqrt{1,764}}\}=\frac{2,352}{56\times42}\} [/Tex] [Tex]=\frac{2,352}{2,352}=1 [/Tex] Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. Example 4.5 Quotations of index numbers of equity share prices of a certain joint stock company and the prices of preference shares are given below. Coefficient of Association Yule's Coefficient of Association measures the strength and direction of association. The general form of the simple linear regression equation is Y = a + bX + e, where 'X' is independent variable, 'Y' is dependent variable, 'a' is intercept, 'b' is slope of the line and 'e' is error term. In psychology, Pearson's coefficient aids in studying relationships between behavioral variables, enhancing understanding of psychological phenomena. Page 5 Regression analysis is a statistical method of determining the mathematical functional relationship connecting independent variable(s) and a dependent variable. This coefficient ranges from -1 to +1. Example 4.3 Two referees in a flower beauty competition rank the 10 types of flowers as follows: Use the rank correlation coefficient and find out what degree of agreement is between the referees. Suppose that price of a product (Y) depends mainly upon three promotional activities such as discount (X1), instalment scheme (X2) and free installation (X3). . calculate the regression coefficient and interpret the same. If the price of the product has linear relationship with each promotional activity, then the relationship among Y and X1, X2 and X3 may be expressed using the above general form as Y = a + b1 X 1 + b2 X 2 + b3 X 3 + e. Note 2: (AB): Number with attributes AB etc. That is, the closer to the ±1 means the stronger the monotonic relationship. The general form of the multiple linear regression equation is Y = a + b1X1 + b2X2 + b3X3 + ... The correlation coefficient is unaffected by change of scale and origin. For example, if two individuals are placed in the 8th place, they are given the rank (8+9) / 2 = 8.5 each, which is common rank to be assigned and the next will be 10; and if three ranked equal at the 8th place, they are given the rank (8 + 9 + 10) / 3 = 9 which is the common rank to be assigned to each; and the next rank will be 11. Example 4.1 The following data gives the heights(in inches) of father and his eldest son. Thus, if ui = [xi - A] /c and vi = [yi - B] /d with c ≠ 0 and d ≠ 0 ⇒ i=1,2, ..., n where A and B are arbitrary values. The correlation coefficient is free from the units or measurements of X and Y. . distinguish between correlation analysis and regression analysis. If the relationship between the dependent variable and the independent variable(s) is not linear, then the regression is called as non-linear regression . If r ≥ 0.7 then the correlation will be of higher degree. Then, covariance between X and Y is defined as cov(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y) If (xi,yi), i=1,2, ..., n is a set of n realisations of (X,Y), then the sample covariance between X and Y can be calculated from 1. Example 4.8 Verify whether the given data: N = 100, (A) = 75, (B) = 60 and (AB) = 15 is consistent. The steps involved in the calculation of coefficient of correlation by using Step Deviation Method are: First of all, take the deviations of Series X from the assumed mean and divide them by Common Factor (C) to determine step deviation [Tex]\{dx\^{\prime}\prime}\} [/Tex]. Udney Yule (1871 - 1951), was a British statistician. Repeated ranks When two or more items have equal values (i.e., a tie) it is difficult to give ranks to them. Construct a 2 × 2 contingency table for the given information. This equation can be used to estimate the value of response variable (Y) based on the given values of the predictor variable(s) X within its domain. Example 4.2 The following are the marks scored by 7 students in two tests in a subject. Children and elderly people need much more health care than middle aged persons as seen from the following graph. Karl Pearson's Coefficient of Correlation | Assumptions, Merits and Demerits Change of Scale and OriginCoefficient of Correlation does not depend upon the change of scale and origin. . If r = 0, then the variables are uncorrelated. There is a strong disagreement between equity shares and preference share prices. The outliers may be dropped before the calculation for meaningful conclusion. n the sample correlation co-efficient between X and Y can be calculated from or, equivalently 2. That a change in one variable causes a change in another. Remark: Consistency in the data using contingency table may be found as under. Calculate the total of step deviations; i.e., [Tex]\sum\{dy\^{\prime}\prime}\} [/Tex]Square the step deviation of Series Y and determine their total; i.e., [Tex]\sum\{dy\^{\prime}\prime\^2\}\} [/Tex]Square the step deviation of Series Y and determine their total; i.e., [Tex]\sum\{dy\^{\prime}\prime\^2\}\} [/Tex] Multiply [Tex]\{dx\^{\prime}\prime}\} [/Tex] and [Tex]\{dy\^{\prime}\prime}\} [/Tex], and determine their total; i.e., [Tex]\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\} [/Tex]Now, use the following formula to determine Coefficient of Correlation:[Tex]r=\frac{N\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\}-\sum\{dx\^{\prime}\prime\}\.\sum\{dy\^{\prime}\prime}\}}{\sqrt{N\sum\{dx\^{\prime}\prime\^2\}-\sum\{dx\^{\prime}\prime}\^2}}\sqrt{N\sum\{dy\^{\prime}\prime\^2\}-\sum\{dy\^{\prime}\prime}\^2}}\} [/Tex] Where, N = Number of pair of observations [Tex]\sum\{dx\^{\prime}\prime}\} [/Tex] = Sum of deviations of X values from assumed mean [Tex]\sum\{dy\^{\prime}\prime}\} [/Tex] = Sum of deviations of Y values from assumed mean [Tex]\sum\{dx\^{\prime}\prime\^2\}\} [/Tex] = Sum of squared deviations of X values from assumed mean [Tex]\sum\{dy\^{\prime}\prime\^2\}\} [/Tex] = Sum of squared deviations of Y values from assumed mean [Tex]\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\} [/Tex] = Sum of the products of deviations [Tex]\{dx\^{\prime}\prime}\} [/Tex] and [Tex]\{dy\^{\prime}\prime}\} [/Tex] Example:Use Step Deviation Method and determine the coefficient of correlation for the following data: Solution:[Tex]r=\frac{N\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\}-\sum\{dx\^{\prime}\prime\}\.\sum\{dy\^{\prime}\prime}\}}{\sqrt{N\sum\{dx\^{\prime}\prime\^2\}-\sum\{dx\^{\prime}\prime}\^2}}\sqrt{N\sum\{dy\^{\prime}\prime\^2\}-\sum\{dy\^{\prime}\prime}\^2}}\} [/Tex] [Tex]=\frac{180\times245-(180\times150)-(180\times150)-445\times120}{\sqrt{(180\times150+445\times120)-189000+53400}}=\frac{136500}{24300}=0.561>0 [/Tex] Interpretation: There is a positive association between success in examination and attending coaching classes. Therefore, 3.5 is assigned for rank 2 and 3 with m1=2. Test for the association of attributes attending the coaching class and success in the examination. Solution: Interpretation: Degree of agreement between the referees 'A' and 'B' is 0.636 and they have "strong agreement" in evaluating the competitors. Co-variance Let (X,Y) be a bivariable normal random variable where V(X) and V(Y) exists. In Mathematics (Y), 30 is repeated three times corresponding to ranks 3, 4 and 5. Karl Pearson's coefficient of correlation When X and Y are linearly related and (X,Y) has a bivariate normal distribution, the co-efficient of correlation between X and Y is defined as This is also called as product moment correlation co-efficient which was defined by Karl Pearson. However the converse need not be true. The first person to give a mathematical formula for the measurement of the degree of relationship between two variables in 1890 was Karl Pearson. These benefits help market researchers / data analysts / data scientists to eliminate and evaluate the best set of variables to be used for building regression models for predictive purposes. Example 4.4 Calculate the Spearman's rank correlation coefficient for the following data. = 460; 2XY = 508 Later on it was ... Calculate Karl Pearson's coefficient of correlation between the length and weight and interpret the result. The youngest of nine children, he appeared as an intelligent child. Example: Age and health care are related. . A negative value of 'r' indicates negative correlation . If r = +1, then the correlation is perfect positive . If r = -1, then the correlation is perfect negative. It means that on the average, if fathers are tall then sons will probably tall and if fathers are short, probably sons may be short. In this case, a different formula is used when there is more than one item having the same value. In interpretation we use the adjective 'highly' . If X and Y are independent, then rxy = 0. Galton's progress in education was not smooth. More specifically, regression analysis helps one to understand how the typical value of the dependent variable (or 'response variable') changes when any one of the independent variables (regressors) or predictor(s) is varied, while the other independent variables are held fixed. Calculate coefficient of correlation from the following data and interpret. "Association" means that the attributes have some degree of agreement. i.e. -1 ≤ r ≤ 1 . A positive value of "r" indicates positive correlation. Pearson was beginning to work in statistics and Yule followed him into this new field. It helps to determine the impact of changes in the value(s) of the independent variable(s) upon changes in the value of the dependent variable. Example 4.7 Out of 1800 candidates appeared for a competitive examination 625 were successful; 300 had attended a coaching class and of these 180 came out successful. Interpretation The relation coefficient lies between -1 and +1. Solution: Let x denote height of father and y denote ... Explore all Karl Pearson's Correlation Coefficient related practice questions with solutions, important points to remember, 3D videos, & popular books. LEARNING OBJECTIVES The student will be able to . know the concept of regression, its types and their uses. . fit best line of regression by applying the method of least squares. . b+ bX + e Here, Y represents the dependent (response) variable, Xi represents the ith independent variable (regressor), a and bi are the regression coefficients and e is the error term. According to Karl Pearson, "Coefficient of Correlation is calculated by dividing the sum of products of deviations from their respective means by their number of pairs and their standard deviations." Formula of Karl Pearson's Coefficient of Correlation[Tex]Karl-Pearson's-Coefficient-of-Correlation(r)=\frac{Sum-of-Products-of-Deviations-from-their-respective-means}{Number-of-Pairs\times(Standard-Deviations-of-both-Series)}\} [/Tex] Or [Tex]r=\frac{sum\{xy\}}{N\times(\sigma_x\times\sigma_y)}\} [/Tex] Where, N = Number of Pair of Observations x = Deviation of X series from Mean [Tex](X-\bar{X})\} [/Tex] y = Deviation of Y series from Mean [Tex](Y-\bar{Y})\} [/Tex] [Tex]\sigma_x = Standard Deviation of X series [Tex]\sqrt{\frac{sum\{x\^2\}}{N}}\} [/Tex] [Tex]\sigma_y = Standard Deviation of Y series [Tex]\sqrt{\frac{sum\{y\^2\}}{N}}\} [/Tex] r = Coefficient of Correlation Actual Mean MethodDirect MethodShort-Cut MethodAssumed Mean MethodIndirect MethodStep-Deviation MethodThe steps involved in the calculation of coefficient of correlation by using Actual Mean Method are: The first step is to calculate the mean of the given two series (say X and Y). Now, take the deviation of X series from [Tex]\bar{X}\} [/Tex] and denote the deviations by x. Square the deviations of x and obtain the total; i.e., [Tex]\sum\{x\^2\}\} [/Tex] Take the deviation of Y series from [Tex]\bar{Y}\} [/Tex] and denote the deviations by y. Square the deviations of y and obtain the total; i.e., [Tex]\sum\{y\^2\}\} [/Tex] Multiply the respective deviations of Series X and Y and obtain the total; i.e., [Tex]\sum\{xy\}\} [/Tex]. Now, use the following formula to determine the Coefficient of Correlation:[Tex]r=\frac{sum\{xy\}}{\sqrt{sum\{x\^2\}}\times\sqrt{sum\{y\^2\}}}\} [/Tex] Example:Use Actual Mean Method and determine the coefficient of correlation for the following data: Solution:[Tex]\bar{X}=\frac{sum\{X\}}{N}=\frac{168}{7}=24 [/Tex] [Tex]\bar{Y}=\frac{sum\{Y\}}{N}=\frac{105}{7}=15 [/Tex] [Tex]r=\frac{sum\{xy\}}{\sqrt{sum\{x\^2\}}\times\sqrt{sum\{y\^2\}}}\} [/Tex] Σxy = 336, Σx2 = 448, Σy2 = 252 [Tex]r=\frac{336}{\sqrt{336}\sqrt{336}}=\frac{336}{336}=1 [/Tex] Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. This method of measuring the coefficient of correlation is the most popular and is widely used. Properties 1. Spurious correlation means an association extracted from correlation coefficient that may not exist in reality. 2x2 Contingency Table Yule's coefficient: Q = Note 1: The usage of the symbol α is not to be confused with level of significance. Solution: Let x denote marks in test-1 and y denote marks in test-2. He dabbled in medicine and then studied Mathematics at Cambridge. Actual Mean can sometimes come in fractions which can make the calculation of standard deviation complicated and difficult. Solution: Interpretation: This perfect negative rank correlation (- 1) indicates that scorings in the subjects, totally disagree. Change of Origin: If a constant is added or subtracted to the values then it will not have any effect on the value of correlation coefficient.Change of Scale: Similarly, if a constant is multiplied or divided by the values, then too it will not have any effect on the value of correlation coefficient. Example:Find the coefficient of correlation between the following figures; Solution:As the coefficient of correlation is not affected by the change in scale and origin, we will multiply the X series by 10 and divide the Y series by 100. Using the method of rank correlation determine the relationship between equity shares and preference shares prices. If the relationship is linear and there is only one independent variable, then the regression is called as simple linear regression. Types of 'Regression' Based on the kind of relationship between the dependent variable and the set of independent variable(s), there arises two broad categories of regression viz. , linear regression and non-linear regression. 4. After a year dong research in experimental physics, he returned to University College in 1893 to work as a demonstrator for Karl Pearson. Calculation Heights of father and son are positively correlated. Page 2SPEARMAN'S RANK CORRELATION COEFFICIENT If the data are in ordinal scale then Spearman's rank correlation coefficient is used. That is those who perform well in test-1 will also perform well in test-2 and those who perform poor in test-1 will perform poor in test-2. Simple Linear Regression It is one of the most widely known modeling techniques. Solution: Let x denote height of father and y denote height of son. Calculate their total; i.e., Σdx. Now, square the deviations of X series and calculate their total; i.e., Σdx2. Take the deviations of Y Series from the assumed mean and denote the values by dy. Spearman's correlation can be calculated for the subjectivity data also, like competition scores. Its interpretation is similar to that of Pearson's. Outliers (extreme observations) strongly influence the correlation coefficient. Correlation does not imply causal relationship. Spurious Correlation : The word 'spurious' from Latin means 'false' or 'illegitimate'. Therefore, Interpretation: Marks in Commerce and Mathematics are uncorrelated Page 3YULE'S COEFFICIENT OF ASSOCIATION This measure is used to know the existence of relationship between the two attributes A and B (binary complementary variables). Regression analysis is a vital statistical method for obtaining functional relationship between a dependent variable and one or more independent variables. If we see outliers in our data, we should be careful about the conclusions we draw from the value of r. 3. Step Deviation MethodThis method simplifies the calculation of coefficient of correlation as it was discovered by Karl Pearson. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete and nature of relationship is linear. An examination of publications of Sir Francis Galton and Karl Pearson revealed that Galton's work on inherited characteristics of sweet peas led to the initial conceptualization of linear regression. Spearman's rank correlation coefficient is given by the formula where Di = R1i - R2i R1i = rank of i in the first set of data R2i = rank of i in the second set of data and n = number of pairs of observations Interpretation Spearman's rank correlation coefficient is a statistical measure of the strength of a monotonic (increasing/decreasing) relationship between paired data. Coaching class is useful for success in examination. However, if we compute the linear correlation r for such data, it may be zero implying age and health care are uncorrelated, but non-linear correlation is present, where mi is the number of repetitions of ith rank Example 4.6 Compute the rank correlation coefficient for the following data of the marks obtained by 8 students in the Commerce and Mathematics. The steps involved in the calculation of coefficient of correlation by using Direct Method are: The first step is to calculate the sum of Series X (ΣX). Now, calculate the sum of Series Y (ΣY). Square the values of X Series and calculate their total; i.e., ΣX2. Square the values of Y Series and calculate their total; i.e., ΣY2. Multiply the values of Series X and Y and calculate their total; i.e., ΣXY. Now, use the following formula to determine Coefficient of Correlation: [Tex]r=\frac{N\sum(XY)-\sum(X).\sum(Y)}{\sqrt{N\sum(X\^2)-\sum(X)\^2}}\sqrt{N\sum(Y\^2)-\sum(Y)\^2}}\} [/Tex] Example:Use Direct Method and determine the coefficient of correlation for the following data: Solution:[Tex]r=\frac{N\sum(XY)-\sum(X).\sum(Y)}{\sqrt{N\sum(X\^2)-\sum(X)\^2}}\sqrt{N\sum(Y\^2)-\sum(Y)\^2}}\} [/Tex] [Tex]=\frac{7\times(2,856)-(168\times105)}{\sqrt{(7\times4,480)-(168\^2)}}\times\sqrt{(7\times1,827)-(105\^2)}}\} [/Tex] [Tex]=\frac{19,992-17,640}{\sqrt{31,360-28,224}}\times\sqrt{(12,789-11,025)}}\} [/Tex] [Tex]=\frac{2,352}{\sqrt{3,136}\times\sqrt{1,764}}\}=\frac{2,352}{56\times42}\} [/Tex] [Tex]=\frac{2,352}{2,352}=1 [/Tex] Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. Example 4.5 Quotations of index numbers of equity share prices of a certain joint stock company and the prices of preference shares are given below. Coefficient of Association Yule's Coefficient of Association measures the strength and direction of association. The general form of the simple linear regression equation is Y = a + bX + e, where 'X' is independent variable, 'Y' is dependent variable, 'a' is intercept, 'b' is slope of the line and 'e' is error term. In psychology, Pearson's coefficient aids in studying relationships between behavioral variables, enhancing understanding of psychological phenomena. Page 5 Regression analysis is a statistical method of determining the mathematical functional relationship connecting independent variable(s) and a dependent variable. This coefficient ranges from -1 to +1. Example 4.3 Two referees in a flower beauty competition rank the 10 types of flowers as follows: Use the rank correlation coefficient and find out what degree of agreement is between the referees. Suppose that price of a product (Y) depends mainly upon three promotional activities such as discount (X1), instalment scheme (X2) and free installation (X3). . calculate the regression coefficient and interpret the same. If the price of the product has linear relationship with each promotional activity, then the relationship among Y and X1, X2 and X3 may be expressed using the above general form as Y = a + b1 X 1 + b2 X 2 + b3 X 3 + e. Note 2: (AB): Number with attributes AB etc. That is, the closer to the ±1 means the stronger the monotonic relationship. The general form of the multiple linear regression equation is Y = a + b1X1 + b2X2 + b3X3 + ... The correlation coefficient is unaffected by change of scale and origin. For example, if two individuals are placed in the 8th place, they are given the rank (8+9) / 2 = 8.5 each, which is common rank to be assigned and the next will be 10; and if three ranked equal at the 8th place, they are given the rank (8 + 9 + 10) / 3 = 9 which is the common rank to be assigned to each; and the next rank will be 11. Example 4.1 The following data gives the heights(in inches) of father and his eldest son. Thus, if ui = [xi - A] /c and vi = [yi - B] /d with c ≠ 0 and d ≠ 0 ⇒ i=1,2, ..., n where A and B are arbitrary values. The correlation coefficient is free from the units or measurements of X and Y. . distinguish between correlation analysis and regression analysis. If the relationship between the dependent variable and the independent variable(s) is not linear, then the regression is called as non-linear regression . If r ≥ 0.7 then the correlation will be of higher degree. Then, covariance between X and Y is defined as cov(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y) If (xi,yi), i=1,2, ..., n is a set of n realisations of (X,Y), then the sample covariance between X and Y can be calculated from 1. Example 4.8 Verify whether the given data: N = 100, (A) = 75, (B) = 60 and (AB) = 15 is consistent. The steps involved in the calculation of coefficient of correlation by using Step Deviation Method are: First of all, take the deviations of Series X from the assumed mean and divide them by Common Factor (C) to determine step deviation [Tex]\{dx\^{\prime}\prime}\} [/Tex]. Udney Yule (1871 - 1951), was a British statistician. Repeated ranks When two or more items have equal values (i.e., a tie) it is difficult to give ranks to them. Construct a 2 × 2 contingency table for the given information. This equation can be used to estimate the value of response variable (Y) based on the given values of the predictor variable(s) X within its domain. Example 4.2 The following are the marks scored by 7 students in two tests in a subject. Children and elderly people need much more health care than middle aged persons as seen from the following graph. Karl Pearson's Coefficient of Correlation | Assumptions, Merits and Demerits Change of Scale and OriginCoefficient of Correlation does not depend upon the change of scale and origin. . If r = 0, then the variables are uncorrelated. There is a strong disagreement between equity shares and preference share prices. The outliers may be dropped before the calculation for meaningful conclusion. n the sample correlation co-efficient between X and Y can be calculated from or, equivalently 2. That a change in one variable causes a change in another. Remark: Consistency in the data using contingency table may be found as under. Calculate the total of step deviations; i.e., [Tex]\sum\{dy\^{\prime}\prime}\} [/Tex]Square the step deviation of Series X and determine their total; i.e., [Tex]\sum\{dx\^{\prime}\prime\^2\}\} [/Tex]Square the step deviation of Series Y and determine their total; i.e., [Tex]\sum\{dy\^{\prime}\prime\^2\}\} [/Tex] Multiply [Tex]\{dx\^{\prime}\prime}\} [/Tex] and [Tex]\{dy\^{\prime}\prime}\} [/Tex], and determine their total; i.e., [Tex]\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\} [/Tex]Now, use the following formula to determine Coefficient of Correlation:[Tex]r=\frac{N\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\}-\sum\{dx\^{\prime}\prime\}\.\sum\{dy\^{\prime}\prime}\}}{\sqrt{N\sum\{dx\^{\prime}\prime\^2\}-\sum\{dx\^{\prime}\prime}\^2}}\sqrt{N\sum\{dy\^{\prime}\prime\^2\}-\sum\{dy\^{\prime}\prime}\^2}}\} [/Tex] Where, N = Number of pair of observations [Tex]\sum\{dx\^{\prime}\prime}\} [/Tex] = Sum of deviations of X values from assumed mean [Tex]\sum\{dy\^{\prime}\prime}\} [/Tex] = Sum of deviations of Y values from assumed mean [Tex]\sum\{dx\^{\prime}\prime\^2\}\} [/Tex] = Sum of squared deviations of X values from assumed mean [Tex]\sum\{dy\^{\prime}\prime\^2\}\} [/Tex] = Sum of squared deviations of Y values from assumed mean [Tex]\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\} [/Tex] = Sum of the products of deviations [Tex]\{dx\^{\prime}\prime}\} [/Tex] and [Tex]\{dy\^{\prime}\prime}\} [/Tex] Example:Use Step Deviation Method and determine the coefficient of correlation for the following data: Solution:[Tex]r=\frac{N\sum\{dx\^{\prime}\prime\{dy\^{\prime}\prime}\}\}-\sum\{dx\^{\prime}\prime\}\.\sum\{dy\^{\prime}\prime}\}}{\sqrt{N\sum\{dx\^{\prime}\prime\^2\}-\sum\{dx\^{\prime}\prime}\^2}}\sqrt{N\sum\{dy\^{\prime}\prime\^2\}-\sum\{dy\^{\prime}\prime}\^2}}\} [/Tex] [Tex]=\frac{180\times245-(180\times150)-(180\times150)-445\times120}{\sqrt{(180\times150+445\times120)-189000+53400}}=\frac{136500}{24300}=0.561>0 [/Tex] Interpretation: There is a positive association between success in examination and attending coaching classes. Therefore, 3.5 is assigned for rank 2 and 3 with m1=2. Test for the association of attributes attending the coaching class and success in the examination. Solution: Interpretation: Degree of agreement between the referees 'A' and 'B' is 0.636 and they have "strong agreement" in evaluating the competitors. Co-variance Let (X,Y) be a bivariable normal random variable where V(X) and V(Y) exists. In Mathematics (Y), 30 is repeated three times corresponding to ranks 3, 4 and 5. Karl Pearson's coefficient of correlation When X and Y are linearly related and (X,Y) has a bivariate normal distribution, the co-efficient of correlation between X and Y is defined as This is also called as product moment correlation co-efficient which was defined by Karl Pearson. However the converse need not be true. The first person to give a mathematical formula for the measurement of the degree of relationship between two variables in 1890 was Karl Pearson. These benefits help market researchers / data analysts / data scientists to eliminate and evaluate the best set of variables to be used for building regression models for predictive purposes. Example 4.4 Calculate the Spearman's rank correlation coefficient for the following data. = 460; 2XY = 508 Later on it was ... Calculate Karl Pearson's coefficient of correlation between the length and weight and interpret the result. The youngest of nine children, he appeared as an intelligent child. Example: Age and health care are related. . A negative value of 'r' indicates negative correlation . If r = +1, then the correlation is perfect positive . If r = -1, then the correlation is perfect negative. It means that on the average, if fathers are tall then sons will probably tall and if fathers are short, probably sons may be short. In this case, a different formula is used when there is more than one item having the same value. In interpretation we use the adjective 'highly' . If X and Y are independent, then rxy = 0. Galton's progress in education was not smooth. More specifically, regression analysis helps one to understand how the typical value of the dependent variable (or 'response variable') changes when any one of the independent variables (regressors) or predictor(s) is varied, while the other independent variables are held fixed. Calculate coefficient of correlation from the following data and interpret. "Association" means that the attributes have some degree of agreement. i.e. -1 ≤ r ≤ 1 . A positive value of "r" indicates positive correlation. Pearson was beginning to work in statistics and Yule followed him into this new field. It helps to determine the impact of changes in the value(s) of the independent variable(s) upon changes in the value of the dependent variable. Example 4.7 Out of 1800 candidates appeared for a competitive examination 625 were successful; 300 had attended a coaching class and of these 180 came out successful. Interpretation The relation coefficient lies between -1 and +1. Solution: Let x denote height of father and y denote ... Explore all Karl Pearson's Correlation Coefficient related practice questions with solutions, important points to remember, 3D videos, & popular books. LEARNING OBJECTIVES The student will be able to . know the concept of regression, its types and their uses. . fit best line of regression by applying the method of least squares. . b+ bX + e Here, Y represents the dependent (response) variable, Xi represents the ith independent variable (regressor), a and bi are the regression coefficients and e is the error term. According to Karl Pearson, "Coefficient of Correlation is calculated by dividing the sum of products of deviations from their respective means by their number of pairs and their standard deviations." Formula of Karl Pearson's Coefficient of Correlation[Tex]Karl-Pearson's-Coefficient-of-Correlation(r)=\frac{Sum-of-Products-of-Deviations-from-their-respective-means}{Number-of-Pairs\times(Standard-Deviations-of-both-Series)}\} [/Tex] Or [Tex]r=\frac{sum\{xy\}}{N\times(\sigma_x\times\sigma_y)}\} [/Tex] Where, N = Number of Pair of Observations x = Deviation of X series from Mean [Tex](X-\bar{X})\} [/Tex] y = Deviation of Y series from Mean [Tex](Y-\bar{Y})\} [/Tex] [Tex]\sigma_x = Standard Deviation of X series [Tex]\sqrt{\frac{sum\{x\^2\}}{N}}\} [/Tex] [Tex]\sigma_y = Standard Deviation of Y series [Tex]\sqrt{\frac{sum\{y\^2\}}{N}}\} [/Tex] r = Coefficient of Correlation Actual Mean MethodDirect MethodShort-Cut MethodAssumed Mean MethodIndirect MethodStep-Deviation MethodThe steps involved in the calculation of coefficient of correlation by using Actual Mean Method are: The first step is to calculate the mean of the given two series (say X and Y). Now, take the deviation of X series from [Tex]\bar{X}\} [/Tex] and denote the deviations by x. Square the deviations of x and obtain the total; i.e., [Tex]\sum\{x\^2\}\} [/Tex] Take the deviation of Y series from [Tex]\bar{Y}\} [/Tex] and denote the deviations by y. Square the deviations of y and obtain the total; i.e., [Tex]\sum\{y\^2\}\} [/Tex] Multiply the respective deviations of Series X and Y and obtain the total; i.e., [Tex]\sum\{xy\}\} [/Tex]. Now, use the following formula to determine the Coefficient of Correlation:[Tex]r=\frac{sum\{xy\}}{\sqrt{sum\{x\^2\}}\times\sqrt{sum\{y\^2\}}}\} [/Tex] Example:Use Actual Mean Method and determine the coefficient of correlation for the following data: Solution:[Tex]\bar{X}=\frac{sum\{X\}}{N}=\frac{168}{7}=24 [/Tex] [Tex]\bar