Continue

Statistics.Education : By Vraj Prajapati Statistics.EducationIntroduction :In Karl Pearson’s Correlation Coefficient, we use covariance and the standard deviations of the variables to calculate the correlation in numerical form.The Correlation Coefficient is represented by “r” (small r).It’s a good idea to keep these formulas handy on a single page for quick reference while solving problems!Main formula used in Karl Pearson’s Correlation CoefficientFundamental Formula :The Basic formula of KP’s Correlation Coefficient (Product Moment Method) is = (,)/(x . y).There are two commonly used formulas for calculating both covariance and standard deviation, so be sure to take note as we go along!Fundamental Formula of KP’s Correlation CoefficientCovariance :Before learning about Covariance first look at these two graphs.Two Graph Showing Direction of RelationshipOne graph shows an upward trend, suggesting a positive relationship, while the other shows a downward trend, indicating a negative relationship.The direction of the relationship can be visually understood from these graphs, but not in numerical terms. To obtain direction of relationship in numerical form we use covariance.Definition of Covariance“Covariance measures the direction of the relationship in numerical form by calculating how much the deviations of two variables from their respective means correspond to each other.”ExampleExample : Price and Demand of SugarHere, we have the price of sugar and demand of sugar. To obtain the covariance we need to obtain deviations, to obtain deviations first we need to obtain the mean.For X, mean is 144.For Y, mean is 383.Obtaining DeviationsPrepare Three ColumnsNow let’s obtain deviations for variable X, by deducting mean 144 from each observations.Obtaining Deviations For XHere, the total of deviations is 0. Which indicates that the calculation is correct. In some very rare or special cases the sum of deviation from the mean may not be 0.Now let’s obtain deviations for variable Y, by deducting mean 383 from each observations.Obtaining Deviations For YThe sum of deviation is 0, so the calculation is correct.Obtaining Product of DeviationsProduct of Deviations of both the VariablesYou are not supposed to mention all these calculations in the table, it is just for demonstration purpose, you can write the values directly.Obtaining CovarianceCalculation of Covariance of X & YCovariance = -2292 which indicates there is a negative relationship between two variables. Which is in numerical form.Interpretation of CovarianceInterpretation of CovarianceCovariance is greater than 0, which is positive relationship between two variables, as one variable increase the other tends to increase.Covariance is less than 0, which is negative relationship between two variables, as one variable increase the other tends to decrease.Covariance is equals to 0, which is no relationship between two variables, as one variable increase the other remains unchanged.In our example Covariance is -2292, which indicates there is a negative relationship between two variables.Standard Deviation :Before learning about Standard Deviation first look at this graph.Graph Indicating Deviations From Mean LineHere, the green dotted lines show the deviation of each data points form the mean line but we can’t get numerical value from the graph. Thus to know the spread of data set in numerical form we obtain standard deviation.Definition“Standard deviation measures the spread of a dataset in numerical form by calculating how much individual data points deviate from the overall mean of the dataset.”Standard Deviation and Variance FormulaWhy this Formula ?We already have the deviations for both the variables.Table showing DeviationsNow, if we already have the deviations, you might wonder — why not just take their average?The reason is that the sum of deviations from the mean is always zero. So, if we try to calculate the average by dividing this sum by any number, we’ll still end up with zero.Why we cannot obtain Average of the Deviations?As a result, this method doesn’t give us a meaningful value for the average deviation.Thus, to overcome this problem, we square the deviations to eliminate negative values. Then, we find the average of the squared deviations and finally, we take the square root to bring the value back to the original scale. This process gives us the standard deviation.CalculationPrepare Columns for Square of DeviationsTo obtain square values and its sum at the same time follow this method.On your calculator follow this steps as mentioned below.For Variable X-44 × M+-24 × M+6 × M+16 × M+46 × M+Press Check button 3 Times to obtain individual square valuesPress MRC button to obtain Total of square valuesObtaining Square of Deviations for Variable XFor Variable Y117 × M+42 × M+-3 × M+-63 × M+-93 × M+Press Check button 3 Times to obtain individual square valuesPress MRC button to obtain Total of square valuesObtaining Square of Deviations for Variable YObtaining Standard DeviationsObtaining Standard Deviations of X & YA higher standard deviation means the data is more widely spread out, while a lower standard deviation indicates that the data points are closer together.There are actually three main ways to interpret standard deviation, but that’s a longer discussion. For now, we’ll stick to this basic understanding!Correlation :Now, let’s obtain correlation using the values of covariance and standard deviations.Table , Covariance and Standard Deviations CalculationsKarl Pearson’s Correlation Coefficient (Fundamental Formula)Calculation of “r” using Basic formula of KP’s Correlation CoefficientKarl Pearson’s Correlation Coefficient (Using Another Formula)Calculation of “r” using another formula of KP’s Correlation CoefficientHere, the value of r is -0.975. which indicates a negative correlation.No matter which formula of Karl Pearson method you use, the value of ‘r’ will be the same.Interpretation of Correlation CoefficientInterpretation of KP’s Correlation CoefficientAssignment :Obtain Correlation for the following Questions.Question 1Question 1Question 2Question 2Enjoying the content? Give it a clap Share it with your friends And don’t forget to check out my YouTube channel for more! Got questions or doubts? Drop them in the comments! Statistics Education YouTube✍ Copyright © 2025 Vraj Prajapati. These study notes and images are my original work. You’re welcome to reference and share the link, but please avoid direct copying of images or content without permission. Let’s respect creators! Sign Up Now &Daily Live Classes3000+ TestsStudy Material & PDFQuizzes With Detailed Analytics+ More BenefitsGet Free Access Now Statistics.Education : By Vraj Prajapati Statistics.EducationIntroductionThis post features a fully worked-out practical sum using Karl Pearson’s Product Moment Method for calculating correlation. Perfect for students looking for a clear, step-by-step example to understand the application.Practical SumQuestionFirst to solve this using Pearson's correlation method — we need to obtain X^2 , Y^2 & XY. For that we will prepare the following table.Table with X^2 , Y^2 & XY ColumnsNow to obtain the square values for X , we will use the following trick to get the values as well as the total of both X and X^2.On your calculator — do the following calculation.For Variable X7.3 M+ * =2 M+ * =2.1 M+ * =2.3 M+ * =3.1 M+ * =2.4 M+ * =2.6 M+ * =1.6 M+ * =1 M+ * =Press check button 3 times for individual valuesPress MRC button for total of X columnPress GT button for total of X^2 columnNow you will have table like thisValues of X and X^2 columnNow to obtain the square values for Y, we will use the following trick to get the values as well as the total of both Y and Y^2.On your calculator — do the following calculation.For Variable Y245.7 M+ * =106.1 M+ * =95.2 M+ * =34.2 M+ * =94.2 M+ * =46.2 M+ * =100.4 M+ * =33.4 M+ * =48.5 M+ * =Press check button 3 times for individual valuesPress MRC button for total of Y columnPress GT button for total of Y^2 columnNow you will have table like thisValues of Y and Y^2 columnNow to obtain the values for XY, we will use the following trick to get the values as well as the total of the column XY.On your calculator — do the following calculation.For column XY7.3 * 245.7 M+2 * 106.1 M+2.1 * 95.2 M+2.3 * 34.2 M+3.1 * 94.2 M+2.4 * 46.2 M+2.6 * 100.4 M+1.6 * 33.4 M+1 * 48.5 M+Press check button 3 times for individual valuesPress MRC button for total of XY columnNow you will have table like thisValues of XY columnNow lets put these values in formula.Formula of Karl Pearson’s Correlation Coefficient (Product Moment Method)Putting Values in the above formulaSimplifying the formulaFurther simplifying it by deducting valuesTaking square rootAnswerEnjoying the content? Give it a clap Share it with your friends And don’t forget to check out my YouTube channel for more! Got questions or doubts? Drop them in the comments! Statistics Education YouTube✍ Copyright © 2025 Vraj Prajapati.These study notes and images are my original work. You’re welcome to reference and share the link, but please avoid direct copying of images or content without permission. Let’s respect creators! The first person to give a mathematical formula for the measurement of the degree of relationship between two variables in 1890 was Karl Pearson. Karl Pearson's Coefficient of Correlation is also known as Product Moment Correlation or Simple Correlation Coefficient. This method of measuring the coefficient of correlation is the most popular and is widely used. It is denoted by 'r', where r is a pure number which means that r has no unit. 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 CorrelationKarl~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}} Or r=\frac{\sum{xy}}{N\times{\sigma_x}\times{\sigma_y}} Where, N = Number of Pair of Observations x = Deviation of X series from Mean (X-\bar{X}) y = Deviation of Y series from Mean (Y-\bar{Y}) \sigma_x = Standard Deviation of X series (\sqrt{\frac{\sum{x^2}}{N}}) \sigma_y = Standard Deviation of Y series (\sqrt{\frac{\sum{y^2}}{N}}) r = Coefficient of Correlation Methods of Calculating Karl Pearson's Coefficient of CorrelationActual Mean MethodDirect MethodShort-Cut Method/Assumed Mean Method/Indirect MethodStep-Deviation Method1. Actual Mean 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 \bar{X} and denote the deviations by x.Square the deviations of x and obtain the total; i.e., \sum{x^2} Take the deviation of Y series from \bar{Y} and denote the deviations by y.Square the deviations of y and obtain the total; i.e., \sum{y^2} Multiply the respective deviations of Series X and Y and obtain the total; i.e., \sum{xy} .Now, use the following formula to determine the Coefficient of Correlation:r=\frac{\sum{xy}}{\sqrt{\sum{x^2}\times{\sum{y^2}}}} Example:Use Actual Mean Method and determine the coefficient of correlation for the following data: Solution:\bar{X}=\frac{\sum{X}}{N}=\frac{168}{7}=24 \bar{Y}=\frac{\sum{Y}}{N}=\frac{105}{7}=15 r=\frac{\sum{xy}}{\sqrt{\sum{x^2}\times{\sum{y^2}}}} $\sum xy = 336$, $\sum x2 = 448$, $\sum y2 = 252$ r=\frac{336}{\sqrt{448\times252}}=\frac{336}{\sqrt{1,12,896}}=\frac{336}{336}=1 Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. 2. Direct MethodThe 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 ($\sum X$).Now, calculate the sum of Series Y ($\sum Y$).Square the values of X Series and calculate their total; i.e., $\sum X2$.Square the values of Y Series and calculate their total; i.e., $\sum Y2$.Multiply the values of Series X and Y and calculate their total; i.e., $\sum XY$.Now, use the following formula to determine Coefficient of Correlation: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}}} Example:Use Direct Method and determine the coefficient of correlation for the following data: Solution: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}}} =\frac{(7\times2,856)-(168\times105)}{\sqrt{(7\times4,480)-(168)^2}\times{\sqrt{(7\times1,827)-(105)^2}}} =\frac{19,992-17,640}{\sqrt{31,360-28,224}\times{\sqrt{12,789-11,025}}} =\frac{2,352}{\sqrt{3,136}\times{\sqrt{1,764}}}=\frac{2,352}{56\times42} =\frac{2,352}{2,352}=1 Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. 3. Short-Cut Method/Assumed Mean MethodActual Mean can sometimes come in fractions which can make the calculation of standard deviation complicated and difficult. In those cases, it is suggested to use Short-Cut Method to simplify the calculations. The steps involved in the calculation of coefficient of correlation by using Assumed Mean Method are: First of all, take the deviations of X Series from the assumed mean and denote the values by dx. Calculate their total; i.e., $\sum dx$.Now, square the deviations of X series and calculate their total; i.e., $\sum dx2$.Take the deviations of Y Series from the assumed mean and denote the values by dy. Calculate their total; i.e., $\sum dy$.Square the deviations of Y series and calculate their total; i.e., $\sum dy2$.Multiply dx and dy and calculate their total; i.e., $\sum dxdy$.Now, use the following formula to determine Coefficient of Correlation: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}}} Where, N = Number of pair of observations $\sum dx$ = Sum of deviations of X values from assumed mean $\sum dy$ = Sum of deviations of Y values from assumed mean $\sum dx2$ = Sum of squared deviations of X values from assumed mean $\sum dy2$ = Sum of squared deviations of Y values from assumed mean $\sum dxdy$ = Sum of the products of deviations dx and dy Example:Use Assumed Mean Method and determine the coefficient of correlation for the following data: Solution: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}}} =\frac{(7\times420)-(28\times21)}{\sqrt{(7\times560)-(28)^2}\times{\sqrt{(7\times315)-(21)^2}}} =\frac{2,940-588}{\sqrt{3,920-784}\times{\sqrt{2,205-441}}} =\frac{2,352}{\sqrt{3,136}\times{\sqrt{1,764}}}=\frac{2,352}{56\times42} =\frac{2,352}{2,352}=1 Coefficient of Correlation = 1 It means that there is perfect positive correlation between the values of Series X and Series Y. 4. Step Deviation MethodThis method simplifies the calculation of coefficient of correlation as the deviations are taken from assumed means and are divided by a common factor. 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 (dx^\prime) . Calculate the total of step deviations; i.e., \sum{dx^\prime} Take the deviations of Series Y from the assumed mean and divide them by Common Factor (C) to determine step deviation (dy^\prime) . Calculate the total of step deviations; i.e., \sum{dy^\prime} Square the step deviation of Series X and determine their total; i.e., \sum{dx^\prime{^2}} Square the step deviation of Series Y and determine their total; i.e., \sum{dy^\prime{^2}} Multiply (dx^\prime) and (dy^\prime) , and determine their total; i.e., \sum{dx^\prime{dy^\prime}} Now, use the following formula to determine Coefficient of Correlation:r=\frac{N\sum{dx^\prime{dy^\prime}}-\sum{dx^\prime}.\sum{dy^\prime}}{\sqrt{N\sum{dx^\prime{^2}}-(\sum{dx^\prime})^2}{\sqrt{N\sum{dy^\prime{^2}}-(\sum{dy^\prime})^2}}} Where, N = Number of pair of observations \sum{dx^\prime} = Sum of deviations of X values from assumed mean \sum{dy^\prime} = Sum of deviations of Y values from assumed mean \sum{dx^\prime{^2}} = Sum of squared deviations of X values from assumed mean \sum{dy^\prime{^2}} = Sum of squared deviations of Y values from assumed mean \sum{dx^\prime{dy^\prime}} = Sum of the products of deviations (dx^\prime) and (dy^\prime) Example:Use Step Deviation Method and determine the coefficient of correlation for the following data: Solution:r=\frac{N\sum{dx^\prime{dy^\prime}}-\sum{dx^\prime}.\sum{dy^\prime}}{\sqrt{N\sum{dx^\prime{^2}}-(\sum{dx^\prime})^2}{\sqrt{N\sum{dy^\prime{^2}}-(\sum{dy^\prime})^2}}} =\frac{(7\times35)-(7\times7)}{\sqrt{(7\times35)-(7)^2}\times{\sqrt{(7\times35)-(7)^2}}} =\frac{245-49}{\sqrt{245-49}\times{\sqrt{245-49}}} =\frac{196}{\sqrt{196}\times{\sqrt{196}}}=\frac{196}{14\times14} =\frac{196}{196}=1 Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series X and Series Y. 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. 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 it will not have any effect on the value of correlation coefficient.Example:Find the coefficient of correlation from the following figures: Solution:As the coefficient of correlation is not affected by the change in scale and origin of the variables, we will multiply the X Series by 10 and divide the Y series by 100. 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}}} =\frac{(8\times156)-[(-24)\times(-4)]}{\sqrt{(8\times1,584)-(-24)^2}\times{\sqrt{(8\times44)-(-4)^2}}} =\frac{1,248-96}{\sqrt{12,672-576}\times{\sqrt{352-16}}} =\frac{1,152}{\sqrt{12,096}\times{\sqrt{336}}}=\frac{1,152}{110\times18.3} =\frac{1,152}{2,013}=0.57 Coefficient of Correlation = 0.57 It means that there is a moderate degree of positive correlation between variables X and Y. 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 Pearson Correlation Coefficient: Correlation coefficients are used to measure how strong a relationship is between two variables. There are different types of formulas to get a correlation coefficient, one of the most popular is Pearson's correlation (also known as Pearson's r) which is commonly used for linear regression. The Pearson correlation coefficient, often symbolized as (r), is a widely used metric for assessing linear relationships between two variables. It yields a value ranging from –1 to 1, indicating both the magnitude and direction of the correlation. A change in one variable is mirrored by a corresponding change in the other variable in the same direction.This article provides detailed information on the Pearson Correlation Coefficient, its meaning, formula, interpretation, examples, and FAQs.What is the Pearson Correlation Coefficient?The Pearson Correlation Coefficient, denoted as r, is a statistical measure that calculates the strength and direction of the linear relationship between two variables on a scatterplot. The value of r ranges between -1 and 1, where:1 indicates a perfect positive linear relationship,-1 indicates a perfect negative linear relationship, and0 indicates no linear relationship between the variables.Pearson's Correlation Coefficient FormulaKarl Pearson's correlation coefficient formula is the most commonly used and the most popular formula to get the statistical correlation coefficient. It is denoted with the lowercase "r". The formula for Pearson's correlation coefficient is shown below:r = n(∑xy) - (∑x)(∑y) / √[n∑x²-(∑x)²][n∑y²-(∑y)² The full name for Pearson's correlation coefficient formula is Pearson's Product Moment correlation (PPMC). It helps in displaying the Linear relationship between the two sets of the data.Pearson's correlation helps in measuring the correlation strength (it's given by coefficient r-value between -1 and +1) and the existence (given by p-value) of a linear correlation relationship between the two variables and if the outcome is significant we conclude that the correlation exists.Cohen (1988) says that an absolute value of r of 0.5 is classified as large, an absolute value of 0.3 is classified as medium and an absolute value of 0.1 is classified as small.The interpretation of the Pearson's correlation coefficient is as follows:A correlation coefficient of 1 means there is a positive increase of a fixed proportion of others, for every positive increase in one variable. Like, the size of the shoe goes up in perfect correlation with foot length.If the correlation coefficient is 0, it indicates that there is no relationship between the variables.A correlation coefficient of -1 means there is a negative decrease of a fixed proportion, for every positive increase in one variable. Like, the amount of water in a tank will decrease in a perfect correlation with the flow of a water tap.The Pearson correlation coefficient essentially captures how closely the data points tend to follow a straight line when plotted together. It's important to remember that correlation doesn't imply causation – just because two variables are related, it doesn't mean one causes the change in the other.Pearson Correlation Coefficient TablePearson Correlation Coefficient (r) RangeType of CorrelationDescription of RelationshipNew Illustrative Example0 < r ≤ 1PositiveAn increase in one variable associates with an increase in the other.Study Time vs. Test Scores: More hours spent studying tends to lead to higher test scores.r = 0NoneNo discernible relationship between the changes in both variables.Shoe Size vs. Reading Skill: A person's shoe size doesn't predict their ability to read.-1 ≤ r < 0NegativeAn increase in one variable associates with a decrease in the other.Outdoor Temperature vs. Home Heating Cost: As the outdoor temperature decreases, heating costs in the home increase.Pearson Correlation Coefficient OriginThe Pearson correlation coefficient, although named after statistician Karl Pearson, has a more interesting backstory. The concept of correlation itself can be traced back to Francis Galton, a 19th-century scientist and explorer. Galton was fascinated by inheritance and explored relationships between traits in families.While Galton planted the seed for the idea, the mathematical formula behind the coefficient actually came from French physicist Auguste Bravais in 1844. However, it was Karl Pearson who truly championed the concept in the late 1800s. He refined the mathematical treatment, explored its properties, and popularized its use in statistical analysis. For this reason, the coefficient bears his name, even though earlier contributions played a crucial role in its development.Types of Pearson Correlation CoefficientEach type of Pearson correlation coefficient offers unique insights and analytical tools for various research fields, from statistics and psychology to economics and engineering. Understanding these variations enhances the accuracy and depth of correlation analyses, enabling more informed decision-making and hypothesis testing.Adjusted Correlation CoefficientAdjusted correlation coefficient modifies the standard Pearson correlation coefficient to account for sample size and bias, especially when dealing with small sample sizes. It adjusts the correlation coefficient to provide a more accurate estimation of the population correlation.Weighted Correlation CoefficientWeighted correlation coefficient assigns different weights to individual data points based on their importance or reliability. This approach is useful when certain observations carry more significance or have different levels of precision.Reflective Correlation CoefficientReflective correlation coefficient evaluates the relationship between variables in a reflective model, commonly used in structural equation modeling (SEM) to analyze latent constructs. It assesses the relationship between observed variables and underlying constructs.Scaled Correlation CoefficientScaled correlation coefficient scales the correlation coefficient to a specific range or magnitude, facilitating comparison across different datasets or studies. It ensures consistency in interpretation by standardizing correlation values.Pearson's DistancePearson's distance measures the dissimilarity or similarity between two data points based on their correlation coefficient. It quantifies the extent of deviation from perfect correlation, providing insights into the relationship between variables.Circular Correlation CoefficientCircular correlation coefficient assesses the relationship between circular variables, such as angles or directions. It accounts for the cyclical nature of data and measures the degree of association between circular datasets.Partial CorrelationPartial correlation evaluates the relationship between two variables while controlling for the effects of one or more additional variables. It measures the unique association between variables after accounting for the influence of other factors, allowing researchers to isolate specific statistical relationships.Pearson Correlation Coefficient InterpretationPearson correlation coefficient (r) valueStrengthDirectionGreater than .5StrongPositiveBetween .3 and .5ModeratePositiveBetween 0 and .3WeakPositive0NoneNoneBetween 0 and –.3WeakNegativeBetween –.3 and –.5ModerateNegativeLess than –.5StrongNegativeFinding the Correlation Coefficient with Pearson Correlation Coefficient FormulaSteps to find the correlation coefficient with Pearson's correlation coefficient formula:Step 1: Firstly make a chart with the given data like subject, x, and y and add three more columns in it xy,x^2 and y^2.Step 2: Now multiply the x and y columns to fill the xy column. For example:- in x we have 24 and in y we have 65 so xy will be 24×65=1560.Step 3: Now, take the square of the numbers in the x column and fill the x^2 column.Step 4: Now, take the square of the numbers in the y column and fill the y^2 column.Step 5: Now, add up all the values in the columns and put the result at the bottom. Greek letter sigma (Σ) is the short way of saying summation.Step 6: Now, use the formula for Pearson's correlation coefficient:-R = n(∑xy) - (∑x)(∑y) / √[n∑x²-(∑x)²][n∑y²-(∑y)² To know which type of variable we have either positive or negative.Assumptions of Pearson Correlation CoefficientLinear Relationship: Karl Pearson's correlation coefficient assumes a linear relationship between the two variables under consideration. It implies that as one variable changes, the other changes proportionally.Normality: The variables should follow a normal distribution. While Pearson's correlation coefficient is robust to deviations from normality, extreme departures may affect the validity of the correlation analysis.Homoscedasticity: This assumption suggests that the variability in one variable should be consistent across all levels of the other variable. In other words, the spread of data points around the regression line should remain constant.Interval or Ratio Scale: Pearson's correlation coefficient is appropriate for variables measured on an interval or ratio scale. These scales ensure meaningful numerical distances between observations.Independence: The observations used to compute the correlation coefficient should be independent of each other. Independence ensures that each data point contributes uniquely to the analysis without being influenced by other observations.Correlation Coefficient PropertiesCorrelation Coefficient Range: The correlation coefficient r ranges from -1 to +1, inclusive. A value of -1 indicates a perfect negative linear relationship, +1 denotes a perfect positive linear relationship, and 0 represents no linear relationship.Directionality: The sign of the correlation coefficient indicates the direction of the relationship between variables. A positive r indicates a positive association (both variables increase or decrease together), while a negative r suggests a negative association (one variable increases as the other decreases).Magnitude: The magnitude of the correlation coefficient represents the strength of the relationship between variables. Values closer to -1 or +1 indicate a stronger linear relationship, while values closer to 0 suggest a weaker relationship.No Causation: Pearson's correlation coefficient does not imply causation between variables. It only measures the degree of linear association and does not establish a cause-and-effect relationship.Symmetry: The correlation coefficient is symmetric, meaning the correlation between variables X and Y is the same as the correlation between Y and X.Invariance: The correlation coefficient remains unchanged under linear transformations of the variables (e.g., multiplication by a constant or addition of a constant), making it invariant to changes in scale and location.Pearson Correlation Coefficient InterpretationInterpreting the Pearson correlation coefficient (r) involves assessing the correlation strength, direction, and correlation significance of the relationship between two variables. Here's a guide to interpreting r:Strength of Relationship:Close to +1: Indicates a strong positive linear relationship. As one variable increases, the other tends to increase proportionally.Close to -1: Suggests a strong negative linear relationship. As one variable increases, the other tends to decrease proportionally.Close to 0: Implies a weak or no linear relationship. Changes in one variable do not consistently predict changes in the other.Direction of Relationship:Positive r: Both variables tend to increase or decrease together.Negative r: One variable tends to increase as the other decreases, and vice versa.Significance:Statistical significance indicates whether the observed correlation coefficient is likely to occur due to chance.Significance is typically assessed using a hypothesis test, such as the t-test for correlation coefficient, with the null hypothesis stating that the true correlation coefficient in the population is zero.If the p-value is less than the chosen significance level (e.g., 0.05), the correlation is considered statistically significant.Scatterplot Examination:Visual inspection of a scatterplot can provide additional insights into the relationship between variables.A scatterplot allows you to assess the linearity, directionality, and presence of outliers, complementing the numerical interpretation of r.Caution:Correlation does not imply causation. Even if a strong correlation is observed between two variables, it does not necessarily mean that changes in one variable cause changes in the other.Other factors, such as confounding variables or omitted variables, may influence the observed correlation.Sample Size:Larger sample sizes tend to provide more reliable estimates of correlation coefficients, reducing the likelihood of obtaining spurious correlations.Context Dependence:The interpretation of r should consider the specific context and subject matter of the study. What is considered a strong or weak correlation may vary depending on the field of research and the variables under investigation.Bivariate CorrelationPearson's correlation coefficient is a statistical tool used to measure bivariate correlation. This refers to the strength and direction of the linear relationship between two variables. It assesses how much one variable tends to change along with the other. A positive correlation indicates that as one variable increases, the other tends to increase as well. Conversely, a negative correlation suggests that as one variable goes up, the other tends to go down. A value of zero indicates no linear relationship between the variables.Correlation MatrixThe Pearson correlation coefficient is particularly useful when analyzing datasets with multiple variables. In such cases, a correlation matrix can be constructed. This is a square table that summarizes the correlation coefficients between all possible pairs of variables within the data set. By looking at the correlation matrix, researchers can quickly identify which variables have strong positive, negative, or no linear relationship with each other. This helps them understand the overall structure of the data and identify potential relationships for further investigation.Pearson Correlation Coefficient Examples Example 1: There is some correlation coefficient that was given to tell whether the variables are positive or negative?0.69, 0.42, -0.23, -0.99Solution:The given correlation coefficient is as follows:0.69, 0.42, -0.23, -0.99Tell whether the relationship is negative or positive0.69: The relationship between the variables is a strong positive relationship0.42: The relationship between the variables is a strong positive relationship-0.23: The relationship between the variables is a weak negative relationship-0.99: The relationship between the variables is a very strong negative relationshipExample 2: Calculate the correlation coefficient for the following data by the help of Pearson's correlation coefficient formula:X = 10, 13, 15 ,17 ,19andY = 5,10,15,20,25.Solution:Given variables are,X = 10, 13, 15 ,17 ,19andY = 5,10,15,20,25.To, find the correlation coefficient of the following variables Firstly a table is to be constructed as follows, to get the values required in the formula also add all the values in the columns to get the values used in the formula.XYXYX² Y²10550100251310130169100151522522522517203402894001925475362625∑74∑75∑1103∑1144∑1375∑xy = 1103∑x = 74∑y = 75∑x² = 1144∑y² = 1375n = 5Put all the values in the Pearson's correlation coefficient formula:-R = n(∑xy) - (∑x)(∑y) / √ [n∑x²-(∑x)²][n∑y²-(∑y)²R= 5(1103) - (74)(75) / √ [5(1144)-(74)²][5(1375)-(75)²]R = -35 / √[244][1250]R = -35/552.26R = 0.0633The correlation coefficient is 0.064Example 3: Calculate the correlation coefficient for the following table with the help of Pearson's correlation coefficient formula:SUBJECTAGE XWeight Y1409922579322694548 9Solution:Make a table from the given data and add three more columns of XY, X^2, and Y^2. also add all the values in the columns to get $\sum xy$, $\sum x$, $\sum y$, $\sum x^2$, and $\sum y^2$ and n =4.SUBJECTAGE XWeight YXY $X^2Y^2$1409939601600980122579197562562413226915184844761454894806291679 21∑1513361225956252872 4∑xy = 12258∑x = 151∑y = 336∑x² = 5625∑y² = 28724n = 4Put all the values in the Pearson's correlation coefficient formula:-R = n(∑xy) - (∑x)(∑y) / √ [n∑x²-(∑x)²][n∑y²-(∑y)²R= 4(12258) - (151)(336) / √ [4(5625)-(151)²][4(28724)-(336)²]R = -1704 / √ [-301][-2000]R = -1704/775.886R = -2.1961The correlation coefficient is -2.196Example 4: Calculate the correlation coefficient for the following data with the help of Pearson's correlation coefficient formula:X = 5 ,9 ,14, 16andY = 6, 10, 16, 20 .Solution:Given variables are,X = 5 ,9 ,14, 16andY = 6, 10, 16, 20 .To, find the correlation coefficient of the following variables Firstly a table to be constructed as follows, to get the values required in the formula also add all the values in the columns to get the values used in the formula.XYXYX² Y²5630253691090811001416224196256162030256400∑ 44∑ 52∑ 664∑ 558∑ 792∑xy= 664∑x=44∑y=52∑x² =558∑y² =792n =4Put all the values in the Pearson's correlation coefficient formula:- R= n(∑xy) - (∑x)(∑y) / √ [n∑x²-(∑x)²][n∑y²-(∑y)²R= 4(664) - (44)(52) / √ [4(558)-(44)²][4(792)-(52)²]R= 368 / √[296][464] R=368/370.599R=0.994The correlation coefficient is 0.994Example 5: Calculate the correlation coefficient for the following data by the help of Pearson's correlation coefficient formula:X = 21,31,25,40,47,38andY = 70,55,60,78,66,80Solution:Given variables are,X = 21,31,25,40,47,38andY = 70,55,60,78,66,80To, find the correlation coefficient of the following variables Firstly a table is to be constructed as follows, to get the values required in the formula also add all the values in the columns to get the values used in the formula.XYXYX² Y²2170147044149003155170596130252560150062536004078312016006084476631022209435638803040144640 0∑202∑409∑13937∑7280∑28265∑xy= 13937∑x=202∑y=409∑x² =7280∑y² =28265n =6Put all the values in the Pearson's correlation coefficient formula:-R= n(∑xy) - (∑x)(∑y) / √ [n∑x²-(∑x)²][n∑y²-(∑y)²R= 6(13937) - (202)(409) / √ [6(7280)-(202)²][6(28265)-(409)²]R= 1004 / √[2876][2909]R=1004 / 2892.452938R=-0.3471The correlation coefficient is -0.3471Example 6: Calculate the correlation coefficient for the following data by the help of Pearson's correlation coefficient formula:SUBJECTHeight XWeight Y1437822468326854356 7Solution:Make a table from the given data and add three more columns of XY , X^2 and Y^2 and add all the values in the columns to get $\sum xy$, $\sum x$, $\sum y$, $\sum x^2$ and $\sum y^2$ and n =4.SUBJECTHeight XWeight YXYX²Y²1437833541849608422468163257646243268522106767225435672345122544 89∑1282989541431722422∑xy= 9541∑x=128∑y=298∑x² =4317∑y² 22422n =4Put all the values in the Pearson's correlation coefficient formula:-R= n(∑xy) - (∑x)(∑y) / √ [n∑x²-(∑x)²][n∑y²-(∑y)²R= 4(9541) - (128)(298) / √ [4(4317)-(128)²][4(22422)-(298)²]R= 20 / √ [884][884]R=20/884R=0.02262The correlation coefficient is 0.02262People Also Read:Pearson Correlation Coefficient Practice Problems1. Given a Pearson correlation coefficient of r = 0.85 between the amount of time students spent studying and their score on a math test, interpret the strength and direction of the relationship.2. You have data on the number of ice creams sold and the outdoor temperature. After calculating, you find r = −0.62. What does this say about the relationship between the temperature and ice cream sales?3. Consider the following small dataset representing hours studied (X) and test scores (Y):Hours Studied (X)Test Score (Y)1502553654705 80Calculate the Pearson correlation coefficient (r) for the data.The Pearson Correlation Coefficient (r) is a statistical measure of the strength and direction of a linear relationship between two variables on a scatterplot. It ranges from -1 to 1, with 1 indicating a perfect positive relationship, -1 indicating a perfect negative relationship, and 0 indicating no linear relationship. The formula involves summing products of paired scores and dividing by the square root of the product of the sums of squared scores. While r quantifies the degree of linear association, it doesn't imply causation. Developed by Francis Galton, Auguste Bravais, and Karl Pearson, it's foundational in fields like psychology and economics, aiding in the analysis of linear relationships under certain assumptions about the data. The first person to give a mathematical formula for the measurement of the degree of relationship between two variables in 1890 was Karl Pearson. Karl Pearson's Coefficient of Correlation is also known as Product Moment Correlation or Simple Correlation Coefficient. This method of measuring the coefficient of correlation is the most popular and is widely used. It is denoted by 'r', where r is a pure number which means that r has no unit. 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 CorrelationKarl~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}} Or r=\frac{\sum{xy}}{N\times{\sigma_x}\times{\sigma_y}} Where, N = Number of Pair of Observations x = Deviation of X series from Mean (X-\bar{X}) y = Deviation of Y series from Mean (Y-\bar{Y}) \sigma_x = Standard Deviation of X series (\sqrt{\frac{\sum{x^2}}{N}}) \sigma_y = Standard Deviation of Y series (\sqrt{\frac{\sum{y^2}}{N}}) r = Coefficient of Correlation Methods of Calculating Karl Pearson's Coefficient of CorrelationActual Mean MethodDirect MethodShort-Cut Method/Assumed Mean Method/Indirect MethodStep-Deviation Method1. Actual Mean 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 \bar{X} and denote the deviations by x.Square the deviations of x and obtain the total; i.e., \sum{x^2} Take the deviation of Y series from \bar{Y} and denote the deviations by y.Square the deviations of y and obtain the total; i.e., \sum{y^2} Multiply the respective deviations of Series X and Y and obtain the total; i.e., \sum{xy} .Now, use the following formula to determine the Coefficient of Correlation:r=\frac{\sum{xy}}{\sqrt{\sum{x^2}\times{\sum{y^2}}}} Example:Use Actual Mean Method and determine the coefficient of correlation for the following data: Solution:\bar{X}=\frac{\sum{X}}{N}=\frac{168}{7}=24 \bar{Y}=\frac{\sum{Y}}{N}=\frac{105}{7}=15 r=\frac{\sum{xy}}{\sqrt{\sum{x^2}\times{\sum{y^2}}}} $\sum xy = 336$, $\sum x2 = 448$, $\sum y2 = 252$ r=\frac{336}{\sqrt{448\times252}}=\frac{336}{\sqrt{1,12,896}}=\frac{336}{336}=1 Coefficient of Correlation = 1 It means that there is a perfect positive correlation between the values of Series