assignment=42

August 1, 2023

1 Q1. Pearson correlation coefficient is a measure of the linear relationship between two variables. Suppose you have collected data on the amount of time students spend studying for an exam and their final exam scores. Calculate the Pearson correlation coefficient between these two variables and interpret the result.

To calculate the Pearson correlation coefficient between the amount of time students spend studying for an exam and their final exam scores, we would need to follow these steps: Collect the data on the two variables. We would need to record the amount of time each student spends studying and their corresponding exam scores. Calculate the mean and standard deviation for each variable. Calculate the covariance between the two variables using the formula: cov(X,Y) = Σ[(Xi - Xmean)*(Yi - Ymean)] / (n-1) where X and Y are the variables (time studying and exam scores, respectively) Xi and Yi are the individual data points Xmean and Ymean are the means of the variables n is the sample size. Calculate the Pearson correlation coefficient using the formula: r = cov(X,Y) / (SD(X) * SD(Y)) where r is the correlation coefficient cov(X,Y) is the covariance between the two variables SD(X) and SD(Y) are the standard deviations of the variables. The resulting correlation coefficient will be a number between -1 and 1. A positive correlation coefficient indicates a positive linear relationship between the two variables, meaning that as one variable increases, the other variable also tends to increase. A negative correlation coefficient indicates a negative linear relationship between the two variables, meaning that as one variable increases, the other variable tends to decrease. A correlation coefficient of zero indicates no linear relationship between the two variables. Analysis with example If the calculated Pearson correlation coefficient between time spent studying and exam scores is 0.8, this would indicate a strong positive linear relationship between the two variables. This means that as the amount of time spent studying increases, the exam scores tend to increase as well. If the correlation coefficient was -0.6, this would indicate a moderate negative linear relationship, meaning that as the amount of time spent studying increases, the exam scores tend to decrease.

2 Q2. Spearman's rank correlation is a measure of the monotonic relationship between two variables. Suppose you have collected data on the amount of sleep individuals get each night and their overall job satisfaction level on a scale of 1 to 10. Calculate the Spearman's rank correlation between these two variables and interpret the result.

To calculate the Spearman's rank correlation between the amount of sleep individuals get each night and their overall job satisfaction level on a scale of 1 to 10, we would need to follow these steps: Collect the data on the two variables. We would need to record the amount of sleep each individual gets and their corresponding job satisfaction level. Rank each variable separately. For example, we would rank each individual's amount of sleep from lowest to highest, assigning a rank of 1 to the individual with the lowest amount of sleep, 2 to the individual with the second-lowest amount of sleep, and so on. We would do the same for job satisfaction level, ranking from lowest to highest. Calculate the difference between the ranks for each individual for each variable. For example, if an individual had a rank of 4 for amount of sleep and a rank of 7 for job satisfaction, the difference would be 4-7 = -3. Calculate the Spearman's rank correlation coefficient using the formula: rs = 1 - (6 * Σd^2) / (n(n^2-1)) where rs is the rank correlation coefficient d is the difference between the ranks for each individual for each variable n is the sample size. The resulting correlation coefficient will be a number between -1 and 1. A positive correlation coefficient indicates a positive monotonic relationship between the two variables, meaning that as one variable increases, the other variable tends to increase. A negative correlation coefficient indicates a negative monotonic relationship between the two variables, meaning that as one variable increases, the other variable tends to decrease. A correlation coefficient of zero indicates no monotonic relationship between the two variables. Analysis with example If the calculated Spearman's rank correlation coefficient between amount of sleep and job satisfaction level is 0.6, this would indicate a moderate positive monotonic relationship between the two variables. This means that as the amount of sleep an individual gets increases, their job satisfaction tends to increase as well. If the correlation coefficient was -0.4, this would indicate a moderate negative monotonic relationship, meaning that as the amount of sleep an individual gets increases, their job satisfaction tends to decrease.

3 Q3. Suppose you are conducting a study to examine the relationship between the number of hours of exercise per week and body mass index (BMI) in a sample of adults. You collected data on both variables for 50 participants. Calculate the Pearson correlation coefficient and the Spearman's rank correlation between these two variables and compare the results.

To calculate the Pearson correlation coefficient and the Spearman's rank correlation between the number of hours of exercise per week and body mass index (BMI) in a sample of 50 participants, we would need to follow these steps: Collect the data on the two variables. We would need to record the number of hours of exercise per week and the BMI of each participant. Calculate the mean and standard deviation for each variable. Calculate the covariance between the two variables using the formula: cov(X,Y) = Σ[(Xi - Xmean)*(Yi - Ymean)] / (n-1) where X and

Y are the variables (hours of exercise and BMI, respectively) Xi and Yi are the individual data points Xmean and Ymean are the means of the variables n is the sample size. Calculate the Pearson correlation coefficient using the formula: r = cov(X,Y) / (SD(X) * SD(Y)) where r is the correlation coefficient cov(X,Y) is the covariance between the two variables SD(X) and SD(Y) are the standard deviations of the variables. Rank each variable separately using the same method described in the previous answer. Calculate the difference between the ranks for each individual for each variable. Calculate the Spearman's rank correlation coefficient using the formula: $rs = 1 - (6 * \Sigma d^2) / (n(n^2-1))$ where rs is the rank correlation coefficient d is the difference between the ranks for each individual for each variable n is the sample size. The Pearson correlation coefficient will indicate the strength and direction of the linear relationship between the two variables, while the Spearman's rank correlation coefficient will indicate the strength and direction of the monotonic relationship between the two variables. Example If the Pearson correlation coefficient between hours of exercise per week and BMI is 0.2, this would indicate a weak positive linear relationship between the two variables, meaning that as the number of hours of exercise per week increases, BMI tends to increase slightly. If the Spearman's rank correlation coefficient is 0.3, this would indicate a weak positive monotonic relationship between the two variables, meaning that as the number of hours of exercise per week increases, BMI tends to increase slightly, but not necessarily in a linear fashion. It is important to note that the Pearson correlation coefficient measures only linear relationships, while the Spearman's rank correlation coefficient can capture non-linear relationships as well. Therefore, the results of the two tests should be interpreted with caution, and further analysis may be needed to fully understand the relationship between the two variables.

4 Q4. A researcher is interested in examining the relationship between the number of hours individuals spend watching television per day and their level of physical activity. The researcher collected data on both variables from a sample of 50 participants. Calculate the Pearson correlation coefficient between these two variables.

To calculate the Pearson correlation coefficient between the number of hours individuals spend watching television per day and their level of physical activity in a sample of 50 participants, we would need to follow these steps: Collect the data on the two variables. We would need to record the number of hours individuals spend watching television per day and their level of physical activity for each participant. Calculate the mean and standard deviation for each variable. Calculate the covariance between the two variables using the formula: cov(X,Y) = Σ[(Xi - Xmean)*(Yi - Ymean)] / (n-1) where X and Y are the variables (hours of television watched and level of physical activity, respectively) Xi and Yi are the individual data points Xmean and Ymean are the means of the variables n is the sample size. Calculate the Pearson correlation coefficient using the formula: r = cov(X,Y) / (SD(X) * SD(Y)) where r is the correlation coefficient, cov(X,Y) is the covariance between the two variables SD(X) and SD(Y) are the standard deviations of the variables. The resulting Pearson correlation coefficient will be a value between -1 and 1. A value of -1 indicates a perfect negative correlation (as one variable increases, the other decreases). A value of 0 indicates no correlation. A value of 1 indicates a perfect positive correlation (as one variable increases, the other also increases). For example, if the Pearson correlation coefficient between hours of television watched per day and level of physical activity is -0.2, this would indicate a weak negative correlation between the two variables, meaning that as the number of hours of television watched

per day increases, the level of physical activity tends to decrease slightly. It is important to note that correlation does not necessarily imply causation, and further analysis may be needed to fully understand the relationship between the two variables.

5 Q5. A survey was conducted to examine the relationship between age and preference for a particular brand of soft drink. The survey results are shown below:

Age(Years) Soft drink Preference 25_____Coke 42_____Pepsi 37_____Mountain Dew 19_____Coke 31_____Pepsi 28_____Coke The data given shows the ages and soft drink preferences of six individuals. To examine the relationship between age and soft drink preference, we can create a contingency table with age as the row variable and soft drink preference as the column variable: _____Coke_____Pepsi_____Mountain Dew 19_____X 25_____X 28_____X 31_____X 37_____X 42_____X From this table, we can see that there are 3 individuals who prefer Coke, 2 who prefer Pepsi, and 1 who prefers Mountain Dew. In terms of age, there is 1 individual in the 19-year-old age group, 2 in the 25-28 age group, 1 in the 31-year-old age group, 1 in the 37-year-old age group, and 1 in the 42-year-old age group. It is difficult to assess the relationship between age and soft drink preference based on this small sample size. However, we can calculate the marginal frequencies and proportions to provide some descriptive statistics of the sample. The marginal frequency for each row and column variable can be calculated as follows: For the row variable (age): 1 person in the 19-year-old age group, 2 people in the 25-28 age group, 1 person in the 31-year-old age group, 1 person in the 37-year-old age group, and 1 person in the 42-year-old age group. For the column variable (soft drink preference): 3 people prefer Coke, 2 people prefer Pepsi, and 1 person prefers Mountain Dew. The marginal proportions for each variable can be calculated by dividing the marginal frequency for each category by the total sample size (which is 6 in this case): For the row variable (age): 19 years old: 1/6 = 0.17, 25-28 years old: 2/6 = 0.33, 31 years old: 1/6 = 0.17, 37 years old: 1/6 = 0.17, 42 years old: 1/6 = 0.17. For the column variable (soft drink preference): Coke: 3/6 = 0.5, Pepsi: 2/6 = 0.33, Mountain Dew: 1/6 = 0.17.

Based on this information, we can see that Coke is the most popular soft drink among this small sample, and that the age groups are relatively evenly distributed. We cannot draw any strong conclusions about the relationship between age and soft drink preference without further analysis or a larger sample size.

6 Q6. A company is interested in examining the relationship between the number of sales calls made per day and the number of sales made per week. The company collected data on both variables from a sample of 30 sales representatives. Calculate the Pearson correlation coefficient between these two variables.

To calculate the Pearson correlation coefficient between the number of sales calls made per day and the number of sales made per week, we need to first calculate the covariance and standard deviations for both variables. Let X be the number of sales calls made per day and Y be the

number of sales made per week. We have data for X and Y for 30 sales representatives. The sample mean of X, denoted as , can be calculated by adding up all the X values and dividing by the sample size: image.png Similarly, we can calculate the sample mean of Y, denoted as : image.png Next, we calculate the sample covariance between X and Y, denoted as S_{XY}: image.png Finally, we calculate the sample standard deviation of X, denoted as S_X, and the sample standard deviation of Y, denoted as S_Y: image.png With these values, we can now calculate the Pearson correlation coefficient, denoted as r: image.png Once we plug in the values we have, we can calculate r. It is important to note that r is a measure of the strength and direction of the linear relationship between X and Y. If r is close to 1, it indicates a strong positive linear relationship, while if r is close to -1, it indicates a strong negative linear relationship. If r is close to 0, it indicates a weak or no linear relationship between the variables. Without the actual data, we cannot calculate the Pearson correlation coefficient for this specific scenario.

[ ]: