Theme : Pearson Correlation Coefficient

Teacher:
prof. G.Shadmanova

Plan:

1. **What is Pearson Correlation Coefficient**

2. **Pearson Correlation Coefficient's Formulas**

3. **Examples**

1. What is Pearson Correlation Coefficient

* The **Pearson correlation coefficient** is a very helpful statistical formula that measures the strength between variables and relationships. In the field of statistics, this formula is often referred to as the **Pearson R test**.

* . When conducting a statistical test between two variables, it is a good idea to conduct a Pearson correlation coefficient value to determine just how strong that relationship is between those two variables.

Pearson Correlation Coefficient's Formulas

* In order to determine how strong the relationship is between two variables, a formula must be followed to produce what is referred to as the **coefficient value**. The coefficient value can range between -1.00 and 1.00.

* If the coefficient value is in the negative range, then that means the relationship between the variables is negatively correlated, or as one value increases, the other decreases. If the value is in the positive range, then that means the relationship between the variables is positively correlated, or both values increase or decrease together. Let's look at the formula for conducting the Pearson correlation coefficient value.

* Step one: Make a chart with your data for two variables, labeling the variables (x) and (y), and add three more columns labeled (xy), (x^2), and (y^2). A simple data chart might look like this:

* More data would be needed, but only three samples are shown for purposes of example.

Person	Age (x)	Score (y)	(xy)	(x^2)	(y^2)
1					
2					
3					

* Step two: Complete the chart using basic multiplication of the variable values.

Person	Age (x)	Score (y)	(xy)	(x^2)	(y^2)
1	20	30	600	400	900
2	24	20	480	576	400
3	17	27	459	289	729

Step three: After you have multiplied all the values to complete the chart, add up all of the columns from top to bottom.

Person	Age (x)	Score (y)	(xy)	(x^2)	(y^2)
1	20	30	600	400	900
2	24	20	480	576	400
3	17	27	459	289	729
Total	61	77	1539	1265	2029

Step four: Use this formula to find the Pearson correlation coefficient value.

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

Where:

N	=	number of pairs of scores
Σxy	=	sum of the products of paired scores
Σx	=	sum of x scores
Σy	=	sum of y scores
Σx^2	=	sum of squared x scores
Σy^2	=	sum of squared y scores

* Step five: Once you complete the formula above by plugging in all the correct values, the result is your coefficient value! If the value is a negative number, then there is a negative correlation of relationship strength, and if the value is a positive number, then there is a positive correlation of relationship strength. Note: The above examples only show data for three people, but the ideal sample size to calculate a Pearson correlation coefficient should be more than ten people

Examples

* Let's say you were analyzing the relationship between your participants' age and reported level of income. You're curious as to if there is a positive or negative relationship between someone's age and their income level.

After conducting the test, your Pearson correlation coefficient value is +0.20. Therefore, you would have a slightly positive correlation between the two variables, so the strength of the relationship is also positive and considered strong. You could confidently conclude there is a strong relationship and positive correlation between one's age and their income. In other words, as people grow older, their income tends to increase as well.

* Perhaps you were interested in learning more about the relationship strength of your participants' anxiety score and the number of hours they work each week. After conducting the test, your Pearson correlation coefficient value is -0.80. Therefore, you would have a negative correlation between the two variables, and the strength of the relationship would be weak.

You could confidently conclude there is a weak relationship and negative correlation between one's anxiety score and how many hours a week they report working. Therefore, those who scored high on anxiety would tend to report less hours of work per week, while those who scored lower on anxiety would tend to report more hours of work each week.