Correlation Coefficient Types

Correlation or dependence is any linear statistical relationship, whether casual or not, between two random variables. When values of one variable increase as the values of other increases, this is known as Positive Correlation.

When values of one variable decrease as the values of another increase to form an Inverse relationship. It is known as Negative Correlation.

Where it is possible to predict, with a reasonably high level of accuracy, the values of one variable is describe as a Strong Correlation. A Weak Correlation is one where on average the values of one variable are related to other, but there are many exceptions.

There are many types of correlation coefficient used for various variables, each serving a different purposes.

* For 2 nominal variables – Contingency correlation Coeff.
* For 2 ordinal variables – Spearman /Kendall’s Tau Coeff.
* For 2 metric variables – Pearson’s Coeff.
* Both variables are dichotomas – Phi Coeff.

We commonly measure 4 types of Correlation Coefficient: -

1. Pearson Correlation Coeff.
2. Kendall Tau Correlation Coeff.
3. Spearman’s Rank Correlation Coeff.
4. Point-Biserial Correlation Coeff.
5. Cramer’s V Correlation Coeff.

*Pearson Correlation Coeff.*

It is used to measure degree of relationship b/w linearly related variables on a continuous scale.f

Assumptions –

* Both variables should be normally distributed
* No significant outliers
* Should be continuous variable
* The 2 variables should have a linear relationship.
* Observations should be paired observations

*Kendall’s Tau Correlation Coeff.*

* It is a non- parametric measure of relationship between columns of ranked data.
* Interpretation of Kendall’s Tau coeff. is in terms of probabilities of observing agreeable and non-agreeable pairs and very direct.
* Distribution of Kendall’s Tau has better statistical properties
* In most situations, interpretations of Kendall’s Tau and Spearman’s Rho correlation coeff. are very similar and thus lead to same inferences.
* Variables should be measured on Ordinal / Continuous scale.
* Determines whether there is a monotonic relationship between two variables.

Kendall’s Tau coeff. = Concordant Pair – Discordant Pair

Concordant Pair + Discordant Pair

*Spearman’s Rho Correlation Coeff.*

* Interpretation in Spearman’s coeff. is in terms of the difference between ranking in two ordinal variables.
* The difference in rank is calculated and then, the formula is applied.

Spearman’s Rho Coeff. => Rs = 6(Sum of d2)

n3 – n

*Point-Biserial Correlation Coeff.*

* It is a measure of association between one continuous variable and one dichotomous variable.
* In short, it is an extended version of Pearson’s coeff.
* No outliers should present for continuous variable for each dichotomous variable.
* Should be approximately normally distributed.
* Continuous variable should have equal variance for each category.

*Chi-Square Coeff.*

There are two tests,

1. Chi-square goodness of fit test – to determine if sample data matches to the population.
2. Chi-square test for independence – compares two variables in a contingency table to see if they are related. It tests whether distributions of categorical variables differ from each other.

X2 = (Observed - Expected)2

Expected

*Cramer’s V Correlation Coeff.*

* This correlation coeff indicates how strongly two categorical variables are associated.
* It is used as Post -test to determine strength of association after chi-square has determined the significance.
* Chi-square test says that there is a significant relationship between variables but it does not say how significant and important is, Cramer’s V indicates that.

1/2

V = X2

n(k-1)

Where, X2 = chi-square

K = no. of rows/columns