Graphical user interface, text

Description automatically generated with medium confidence

Logo

Description automatically generated

**Abhilash Dikshit**

Course: ALY 6010

Instructor: Mohsen Soltanifar

Research: Spearman and Kendall

Date: 2022/12/11

**Spearman’s correlation coefficient**

Spearman’s correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data. In a sample it is denoted by and is by design constrained as follows

Arrow

Description automatically generated with low confidence

And its interpretation is similar to that of Pearson’s, e.g., the closer is to the stronger the monotonic relationship. Correlation is an effect size and so we can verbally describe the strength of the correlation using the following guide for the absolute value of :

•  .00-.19 “very weak”

•  .20-.39 “weak”

•  .40-.59 “moderate”

•  .60-.79 “strong”

•  .80-1.0 “very strong”

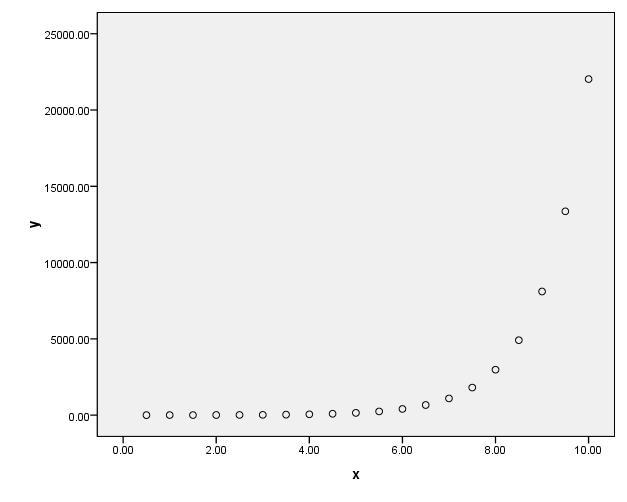
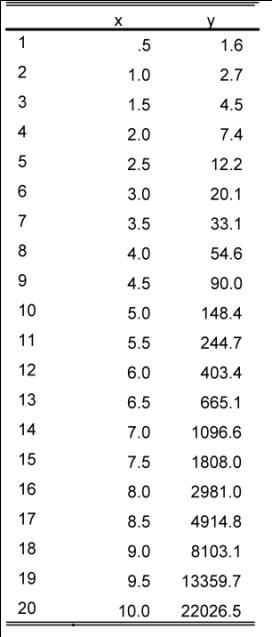
The calculation of Spearman’s correlation coefficient and subsequent significance testing of it requires the following data assumptions to hold:

•  interval or ratio level or ordinal.

•  monotonically related.

Note, unlike Pearson’s correlation, there is no requirement of normality and hence it is a nonparametric statistic.

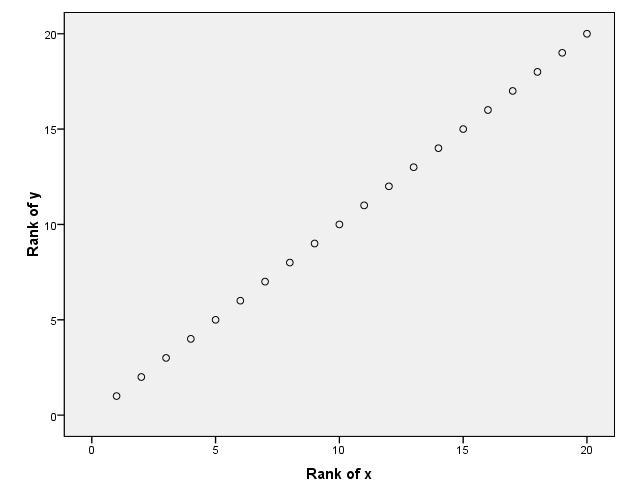
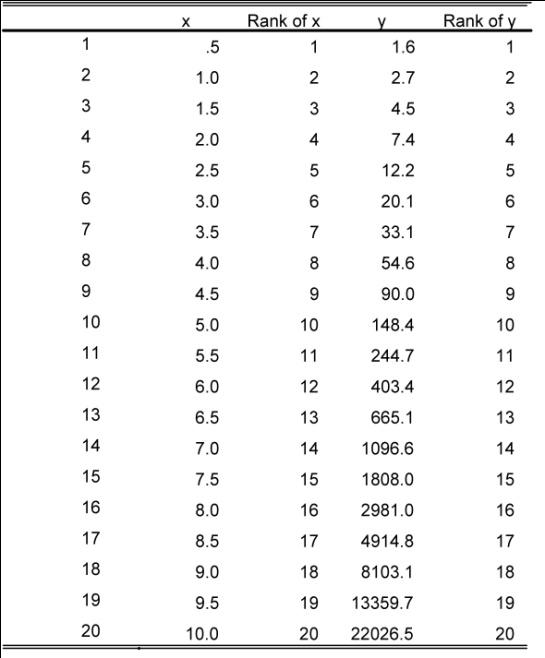
Let us consider some examples to illustrate it. The following table gives x and y values for the relationship. From the graph we can see that this is a perfectly increasing monotonic relationship.



The calculation of Pearson’s correlation for this data gives a value of .699 which does not reflect that there is indeed a perfect relationship between the data. Spearman’s correlation for this data however is 1, reflecting the perfect monotonic relationship.

Spearman’s correlation works by calculating Pearson’s correlation on the ranked values of this data. Ranking (from low to high) is obtained by assigning a rank of 1 to the lowest value, 2 to the next lowest and so on.

If we look at the plot of the ranked data, then we see that they are perfectly linearly related.



In the figures below various samples and their corresponding sample correlation coefficient values are presented. The first three represent the “extreme” monotonic correlation values of -1, 0 and 1:

Chart, scatter chart, box and whisker chart

Description automatically generated

Invariably what we observe in a sample are values as follows:

Chart, scatter chart

Description automatically generated

In the following scatterplot rs=0 which implies no (monotonic) correlation. However, there is a perfect quadratic relationship:

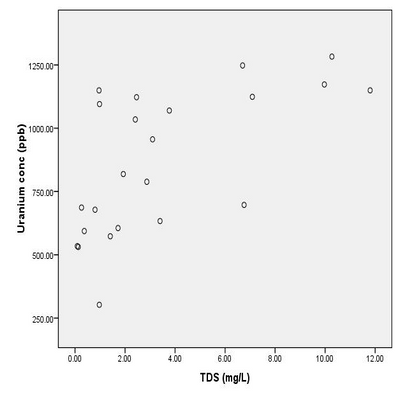
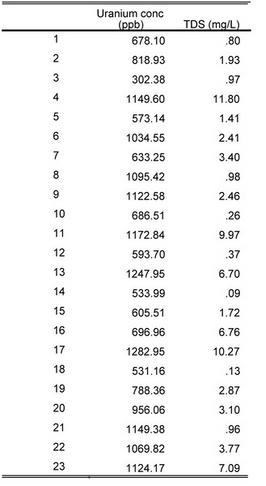
Chart, scatter chart

Description automatically generated

Example

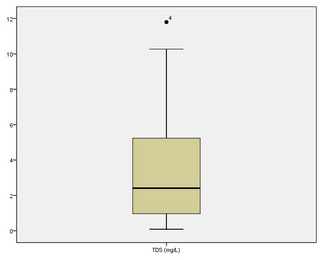
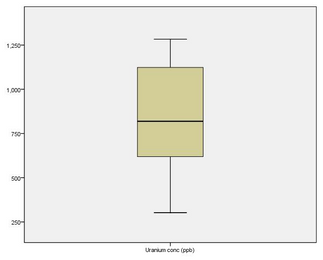
The following data comprises 23 groundwater samples that were collected recording the Uranium concentration (ppb) and the total dissolved solids (mg/L). It is of interest to know if the two variables are correlated.

We should initial consider if Pearson’s correlation is appropriate or whether we should resort to Spearman’s if there are assumption violations.



The scatterplot suggests a definite positive correlation between Uranium and TDS. However, there is possibly slight evidence of non-linearity for TDS values close to zero. However, this is debateable and so we shall move on and consider the other normality assumption.

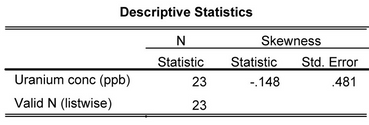
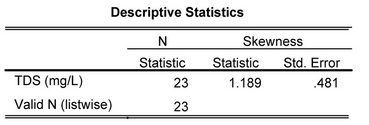
We need to perform some normality checks for the two variables. One simple way of doing this is to examine boxplots of the data. These are given below.



The boxplot for Uranium is fairly consistent with one from a normal distribution; the median is fairly close to the centre of the box and the whiskers are of approximate equal length.

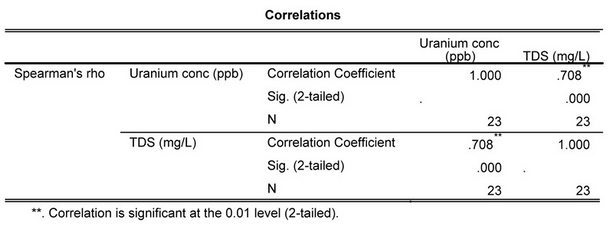
The boxplot for TDS is slightly disturbing in that the median is close to the lower quartile and the lower whisker is shorter than the upper one, which would be suggesting positive skewness. Also, there is an outlier and Pearson’s correlation is sensitive to these as well as skewness.

Since we have some doubts over normality, we shall examine the skewness coefficients to see if there is further evidence to suggest whether either of the variables is skewed.



A quick check to see if the skewness coefficients are not sufficiently large to warrant concern is to see if the absolute values of the skewness coefficients are less than two times their standard errors. Using this guide, the Uranium data’s skewness is consistent with the data being normal. However, the TDS skewness coefficient appears to be large enough to warrant concern that there is positive skewness present (1.189 > 2 x .481).

Hence, we do have concerns over the normality of our data and should continue with a Spearman’s correlation analysis. SPSS produces the following Spearman’s correlation output:



The significant Spearman correlation coefficient value of 0.708 confirms what was apparent from the graph; there appears to be a strong positive correlation between the two variables. Thus, large values of uranium are associated with large TDS values

However, we need to perform a significance test to decide whether based upon this sample there is any or no evidence to suggest that linear correlation is present in the population. To do this we test the null hypothesis, H0, that there is no monotonic

correlation in the population against the alternative hypothesis, H1, that there is monotonic correlation; our data will indicate which of these opposing hypotheses is most likely to be true. Let ρs be the Spearman’s population correlation coefficient then we can thus express this test as:

H0 : ρs =0

H1: ρs ≠0

i.e., the null hypothesis of no monotonic correlation present in population against the alternative that there is monotonic correlation present.

Since SPSS reports the p-value for this test as being .000 we can say that we have very strong evidence to believe H1, i.e., we have some evidence to believe that groundwater uranium and TDS values are monotonically correlated in the population.

This could be formally reported as follows:

"A Spearman's correlation was run to determine the relationship between 23 groundwater uranium and TDS values. There was a strong, positive monotonic correlation between Uranium and TDS ( rs= 0.71, n = 23, p < .001)."

# **Kendall’s Tau (Kendall Rank Correlation Coefficient)**

# In statistics, **correlation**refers to the strength and direction of a relationship between two variables. The value of a correlation coefficient can range from -1 to 1, with -1 indicating a perfect negative relationship, 0 indicating no relationship, and 1 indicating a perfect positive relationship.

**Kendall’s Tau** measures the relationship between two columns of ranked data.

The formula to calculate Kendall’s Tau, often abbreviated τ, is as follows:

τ = (C-D) / (C+D)

where:

C = the number of concordant pairs

D = the number of discordant pairs

The following example illustrates how to use this formula to calculate Kendall’s Tau rank correlation coefficient for two columns of ranked data.

## **Example:**

Suppose that two experts order four wines called {a, b, c, d}. The first expert gives the following order: O1 = [a,c,b,d], which corre- sponds to the following ranks R1 = [1,3,2,4]; and the second ex- pert orders the wines as O2 = [a,c,d,b] which corresponds to the following ranks R2 = [1,4,2,3]. The order given by the first expert is composed of the following 6 ordered pairs

P1 ={[a,c], [a,b], [a,d], [c,b], [c,d], [b,d]} .

The order given by the second expert is composed of the following 6 ordered pairs

P2 ={[a,c], [a,b], [a,d], [c,b], [c,d], [d,b]}

The set of pairs which are in only one set of ordered pairs is {[b,d] [d,b]} which gives a value of d∆ (P1,P2) = 2 . With this value of the sym- metric difference distance we compute the value of the Kendall rank correlation coefficient between the order given by these two experts as:

A picture containing text

Description automatically generated

This large value of τ indicates that the two experts strongly agree on their evaluation of the wines (in fact their agree about every- thing but one pair). The obvious question, now, is to assess if such a large value could have been obtained by chance or can be considered as evidence for a real agreement between the experts. This question is addressed in the next section.

**Kendall and Pearson coefficients of correlation**

Kendall coefficient of correlation can also be interpreted as a standard coefficient of correlation computed between two set of *N* (*N* − 1) binary values where each set represents all the possible pairs obtained from *N* objects and assigning a value of 1 when a pair is present in the order an 0 if not.

**References**

1. S. (2017, November 14). *Kendall’s Tau (Kendall Rank Correlation Coefficient)*. Statistics How To. <https://www.statisticshowto.com/kendalls-tau/>
2. Z. (2021, December 13). *Kendall’s Tau: Definition + Example*. Statology. <https://www.tology.org/kendalls-tau/>
3. *Spearman’s correlation*. (n.d.). Spearman’s Correlation. <https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf>