Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 3249-3255
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6079

Spearman's correlation coefficient in statistical analysis

Khawla Ali Abd Al-Hameed[a]

[a]*Department of Soil Science and Water Resources, College of Agriculture, AL-Qasim Green University, Babil, Hilla, Iraq*

(Communicated by Madjid Eshaghi Gordji)

Abstract

In our research, we present a study to determine the required sample size (the water body) to estimate the Spearman correlation coefficient in terms of its strength and direction for the variables of the total annual evaporation and annual humidity so that we have a symbol for the total annual evaporation (a) and the annual total humidity (b) for three provinces in Iraq, which are Baghdad, Najaf and Babel for a period of ten Years 1991 to 2000, we know that the Simple Linear Correlation Coefficient (Spearman) is a tool for people who work in statistical analysis. The result of the correlation of total annual evaporation and annual total humidity in Baghdad governorate was negative and strong, and the result of the correlation of total annual evaporation with the humidity of Najaf Governorate was positive and weak, and the result of the correlation between total annual evaporation and annual humidity in Babil Governorate was positive and weak.

Keywords: Total annual evaporation data, Total annual humidity data, Spearman coefficient, Correlation.

1. Introduction

Spearman's correlation coefficient determines a simple linear relationship between two variables and measures without dimensions. Researchers often need to study this relationship. Also, the value varies from -1, called a simple and linear negative relationship, to +1, called a simple and linear positive relationship as well. The closer its value to zero, the degree of its simple linear correlation becomes smaller than the correlation coefficient of Spearman, and other statistics were calculated, such as determining the variables that can be used indirectly in [5] as well as the direct and indirect effects of the variables in path analysis, partial correlation and canonical correlation in [6].

Email address: khawlaali@agre.uoqasim.edu.iq (Khawla Ali Abd Al-Hameed)

Received: May 2021 *Accepted:* October 2021

We conclude that these statistics depended on the accuracy of Spearman's simple linear correlation coefficient. And the statistical result changes according to the size of the samples and their interpretation. As in the statistical result of a large sample, low volume coefficients are generated, and the relationship is important in practice. While the sample size is small, we conclude from this that the reliability of the estimates is low and there is no real relationship between the two variables as in ([6, 5, 2, 3, 4, 9, 1, 8, 7]).

As a result of the foregoing, the sample size must be accurate and appropriate. Finally, the goal of our study is to determine the sample size to estimate the simple linear correlation coefficient of Spearman between the sum of the variables of annual evaporation and humidity for three governorates (Baghdad, Najaf, Babylon). We know in this research that Baghdad governorate has a total area of 4,555 km^2, and Najaf governorate has a total area of 28, 824 km^2 and the area of water bodies have 200 km^2, and Babel Governorate has a total area of 5,119 km^2 and the area of water bodies have 250 km^2. Can we prove that the water bodies of Baghdad governorate are larger than Najaf and Babil, and the water bodies of Babel governorate are greater than Najaf governorate through the simple linear correlation coefficient of Spearman?

2. Definition

We know the simple linear correlation that there is a relationship between two variables or two phenomena, meaning that when one of the two variables changes, it will lead to the change of the other, whether by decrease or increase. When the two variables increase together and decrease together, the relationship will be positive between them. When one decreases with an increase in the other variable, the relationship between them will be negative. We will know a simplified definition for the Simplex Coefficient of Linear Correlation (Spearman): It is a coefficient that expresses the strength and direction of the relationship between two phenomena only. The relationship will be either negative or positive on the one hand, and weak or strong on the other hand. That Spearman is used for the correlation of ranks if we assume that the variable A has the rank (R_A) and that the variable B has the rank (R_B), and assuming that (d) represents the difference between the two ranks, meaning $(d = R_A - R_B)$. The Spearman coefficient of ranks correlation is given by the following formula:

$$r_s = 1 - 6\sum d^2/n(n^2 - 1)$$

where n is the number of ordered pairs.

3. Correlation

One of the simplest ways to study the relationship between the two phenomena or variables is the propagation method. When we have A and B variables, we collected these data from the pairs of values of these two phenomena, we represent them graphically in the diffusion method and have different models (Model 1) the points in which are scattered unconnected to a specific direction we conclude from this that there is no relationship between the two variables (A, B), (Model 2) in which the points are spread around a straight line so that the values of B decrease with increasing the values of A, so there is a simple inverse linear relationship between the variables (A, B), (Model 3) in which the points spread around a straight line where the values of A increase with the values of B, so there is a simple direct linear relationship between the variables (A, B), (Model 4) [7] in which the points spread around the curve, so there is a non-linear relationship Among the variables (A, B). We illustrate this graphically through the following four models:

Table 1: illustrates the diffusion models

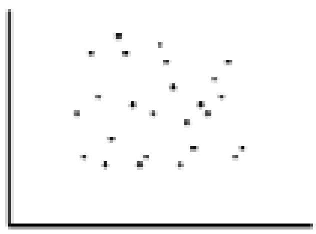
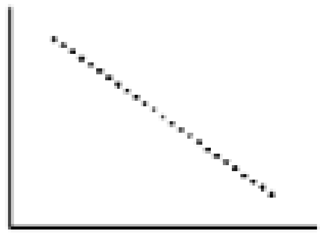
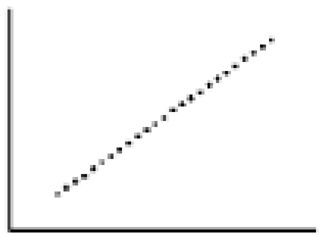
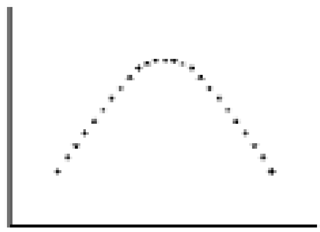
 **Model 1: In this model, scattered points that are not related to a specific direction, we conclude that there is no relationship between the two variables (A, B).**	 **Model 2: In which the points are spread around a straight line where the values of B decrease with increasing the values of A, so there is a simple and inverse linear relationship between the variables (A, B).**
 **Model 3: When the points are spread around a straight line, the A values will increase with the B values.**	 **Model 4: When the points are spread around the curve, then there is a non-linear relationship between the variables (A, B).**

Table 2: The following table shows the types of correlation, the direction of the relationship, and the diffusion models for each type:

value of correlation coefficient	**the meaning**
+1	Completely positive correlation
From 0.70 - 0.99	Strong positive association
From 0.50 - 0.69	Average positive correlation
From 0.01 - 0.49	Weak positive correlation
0	Not a positive relationship

Random correlation applies to reverse correlation (with a negative sign).

4. Examples

These examples show us how the annual total evaporation and humidity are related to the three governorates of Baghdad, Najaf, and Babylon from 1991 to 2000.

Example 4.1. *Find the strength and direction of the relationship between total annual evaporation (A) and moisture (B) annual for Baghdad governorate from 1991 to 2000.*

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
A	2373.4	3281.9	2721.2	2943.8	2840.2	2950.1	2927.2	3128.5	3084.6	2966.5
B	586	552	558	569	576	580	569	555	544	551

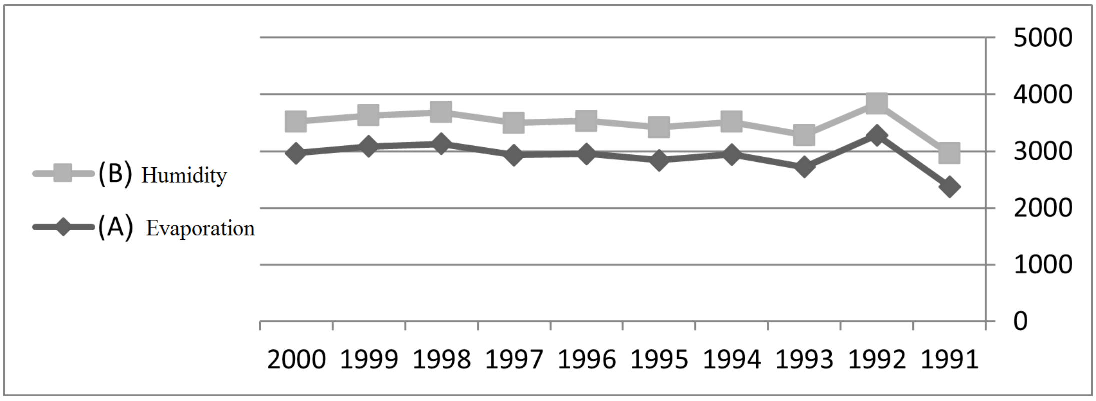
The solution:

Evaporation (A)	Relative Humidity (B)	Rank Evaporation(A)	Rank the relative humidity (B)	d	d^2
2373.4	586	1	10	-9	81
3281.9	552	10	3	7	49
2721.2	558	2	5	-3	9
2943.8	569	5	6.5	-1.5	2.25
2840.2	576	3	8	-5	25
2950.1	580	6	9	-3	9
2927.2	569	4	6.5	-2.5	6.25
3128.5	555	9	4	5	25
3084.6	544	8	1	7	49
2966.5	551	7	2	5	25
					$\sum$ 280.5

$$r_s = 1 - 6\sum d^2/n\left(n^2-1\right)$$

$$r_s = 1 - 6\,(280.5)/10\,(100-1) = 1 - 1683/990 = 1 - 1.7 = -0.7$$

The negative relationship is strong.
The following graph shows the relationship between total annual evaporation and humidity in Baghdad Governorate



Example 4.2. *Find the strength and direction of the relationship between total annual evaporation (A) and moisture (B) annual for Najaf governorate from 1991 to 2000.*

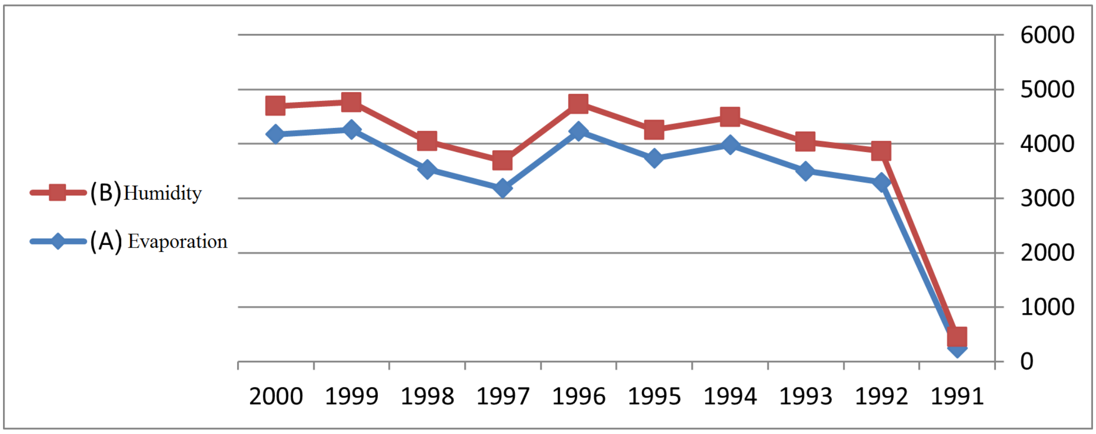
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
A	242.3	3295.9	3497.7	3982.1	3730.8	4228.2	3181.5	3530.2	4261.1	4175.5
B	205	569	537	504	528	506	505	516	505	521

The solution:

Evaporation (A)	Relative Humidity (B)	Rank Evaporation(A)	Rank the relative humidity (B)	d	d^2
242.3	205	1	1	0	0
3295.9	569	3	10	-7	49
3497.7	537	4	9	-5	25
3982.1	504	7	2	5	25
3730.8	528	6	8	-2	4
4228.2	506	9	5	4	16
3181.5	505	2	3.5	-1.5	2.25
3530.2	516	5	6	-1	1
4261.1	505	10	3.5	6.5	42.25
4175.5	521	8	7	1	1
					$\sum 164.5$

$$r_s = 1 - 6\sum d^2/n\left(n^2 - 1\right)$$
$$r_s = 1 - 6\,(164.5)/10\,(100 - 1) = 1 - 987/990 = 1 - 0.99 = 0.01$$

The positive relationship is weak.
The following graph shows the relationship between total annual evaporation and humidity in Baghdad Governorate



Example 4.3. *Find the strength and direction of the relationship between total annual evaporation (A) and moisture (B) annual for Babylon governorate from 1991 to 2000.*

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
A	1204.3	2012.2	2265.4	2507.1	2588.8	2440.7	2276.2	2214.7	2408.7	2558.4
B	306.3	346	367.8	368.6	366.4	380.7	367.5	382.3	387.8	385.6

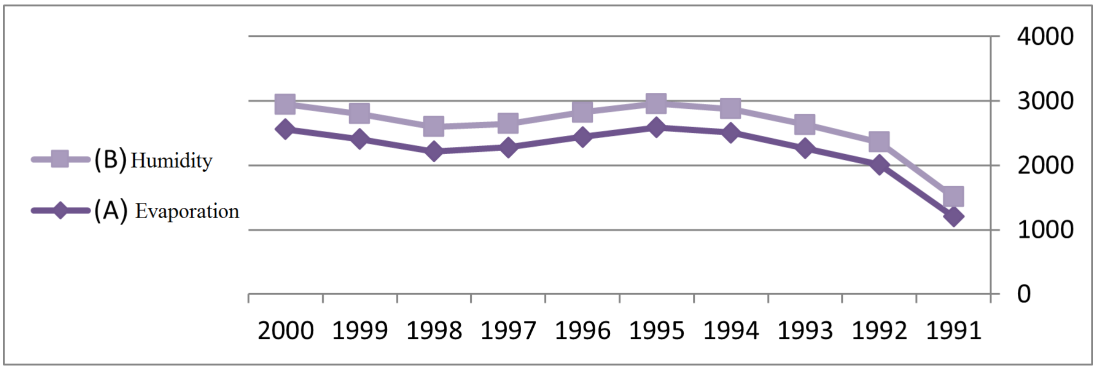
The solution:

Evaporation (A)	Relative Humidity (B)	Rank Evaporation(A)	Rank the relative humidity (B)	d	d^2
1204.3	306.3	1	1	0	0
2012.2	346	2	2	0	0
2265.4	367.8	4	5	-1	1
2507.1	368.6	8	6	2	4
2588.8	366.4	10	3	7	49
2440.7	380.7	7	7	0	0
2276.2	367.5	5	4	1	1
2214.7	382.3	3	8	-5	25
2408.7	387.8	6	10	-4	16
2558.4	385.6	9	9	0	0
					$\sum 96$

$$r_s = 1 - 6\sum d^2/n\left(n^2-1\right)$$

$$r_s = 1 - 6\,(96)/10\,(100-1) = 1 - 576/990 = 1 - 0.582 = 0.418$$

Relationship positive is weak.
The following graph shows the relationship between total annual evaporation and humidity in Babil Governorate



5. Conclusions

We conclude from our study that the simple linear correlation coefficient (Spearman) for the annual total evaporation and humidity data from 1991 to 2000 for the three governorates of Baghdad, Najaf and Babylon, it became clear to us that the water level of Babel Governorate is greater than Najaf and Baghdad.

References

[1] A. Göktas and Ö. Isçi, *A comparison of the most commonly used measures of doubly ordered square contingency tables via simulation*, Metod. Zvezki 8(1) (2011) 17–37.
[2] A. Cargnelutti Filho, M. Toebe, C. Burin, T.R.D. Silveira and G. Casarotto, *Sample size for estimating the Pearson correlation coefficient among corn characters*, Pesquisa Agropecuária Brasileira 45(12) (2010) 1363–1371.
[3] A. Cargnelutti Filho, S. J. Lopes, M. Toebe, T. R. da Silveira and I. A. Schwantes,*Sample size to estimate the Pearson correlation coefficient among characters of Crambe abyssinica*, Revista Ciência Agronômica, 42(1) (2011) 149–158.
[4] A. Cargnelutti Filho, S.J. Lopes, B. Brum, M. Toebe, T.R. da Silveira and G. Casarotto, *Sample size to estimate the Pearson correlation coefficient among characters of castor bean*, Semina: Ciências Agrárias 33(3) (2012) 953–962.
[5] C.D. Cruz and A.J. Regazzi, *Modelos biométricos aplicados ao melhoramento genético*, Viçosa: UFV, 1997, 390p.
[6] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson and R.L. Tatham, *Análise multivariada de dados*, Porto Alegre: Bookman, 2005. 593p.
[7] G. Keller and B. Waracck, *Statistics for Management and Economics*, 6th Edition Duxbury, 2001.
[8] J. Neeter and W. Waserman, *Applied Statistics*, 4th Edition, Louise Richardson, 1993.
[9] Santa Maria, Sari. et al. *Sample size for estimation of the Pearson correlation coefficient in cherry tomato tests*, Cienc. Rural 47 (10)(2017) 1–6.