Aloha International Journal of Multidisciplinary Advancement (AIJMU)
ISSN 2622-3252
Volume 3 Number 1, February 2021 TIPS
http://journal.aloha.academy/index.php/aijmu DOI: http://dx.doi.org/10.33846/aijmu30105

What is a Good Way to Visualize the Positive Correlation between Two Numerical Variables?

Heru Santoso Wahito Nugroho[1] (corresponding author), Sanglar Polnok[2], Tanko Titus Auta[3], Ambo Dalle[4], Bahtiar[5], Pius Selasa[6], Koekoeh Hardjito[7], Atik Badi'ah[8], Fransiskus Salesius Onggang[9], I Putu Suiraoka[10], Alfian Helmi[11], Supriyanti[12], Azwarni[13]

[1]Department of Health, Poltekkes Kemenkes Surabaya, Indonesia; heruswn@gmail.com
[2]Faculty of Nursing, Naresuan University, Thailand; sanglarp@gmail.com
[3]Department of Nursing, Niger State Ministry of Health, Minna, Nigeria; tomatitus0@gmail.com
[4]Department of Nursing, Poltekkes Kemenkes Makassar, Indonesia; ambodallepoltekkesmks@gmail.com
[5]Department of Nursing, Poltekkes Kemenkes Makassar, Indonesia; bahtiar.poltekkes@gmail.com
[6]Department of Nursing, Poltekkes Kemenkes Kupang, Indonesia; piusselasa@gmail.com
[7]Department of Midwifery, Poltekkes Kemenkes Malang, Indonesia; koekoehhardjito@yahoo.co.id
[8]Department of Nursing, Poltekkes Kemenkes Yogyakarta, Indonesia; atik.cahyo@yahoo.com
[9]Department of Nursing, Poltekkes Kemenkes Kupang, Indonesia; frans.onggang@yahoo.com
[10]Department of Nutrition, Poltekkes Kemenkes Denpasar, Indonesia; suiraoka@gmail.com
[11]Department of Nursing, Poltekkes Kemenkes Aceh, Indonesia
[12]Department of Nursing, Poltekkes Kemenkes Aceh, Indonesia
[13]Department of Nursing, Poltekkes Kemenkes Aceh, Indonesia

Submitted: September 11, 2020 - Revised: January 30, 2021 - Accepted: February 15, 2021 - Published: February 28, 2021

ABSTRACT

Most of the results of correlation analysis between variables are not equipped with visualization, so it is deemed necessary to explain how best to visualize the results of correlation analysis, especially for numerical variables. In this case, the best way to present correlations between numerical variables is a scatter diagram. If the points on the diagram are closer to the straight line, the higher the correlation coefficient, meaning that the degree of correlation is getting stronger. In this case, the positive correlation is indicated by the line from the lower left to the upper right. This visualization helps clarify the reader's understanding of the results of the correlation analysis, as well as being a valuable lesson for their similar research projects.
Keywords: positive correlation; correlation coefficient; visualization; scatter diagram

INTRODUCTION

Correlation between two or more variables is something that is very popular in the world of research, especially in studies conducted with a quantitative approach. In scientific research, it is known that there are numerical variables (variables with interval or ratio scale data) and categorical variables (variables with nominal or ordinal scale data) [(1-3)].

In the application of correlation testing between variables, students and novice researchers are often confused about how to choose the appropriate statistical test, and especially how to visualize the results of the correlation test so that it is easily understood by readers, as stated by Taylor [(4)] that professionals as the reader must understand the results of the statistical analysis presented. Even in various journals, the majority of correlation tests between variables are not equipped with visualization. This gives rise to the obligation of senior researchers to share about how best to present the results of the correlation test between variables.

The description above is one of the problems in the application of research methods that must be resolved in order to create a better research report, which is easier for readers to understand. Furthermore, in order to ensure the success of solving this problem, the discussion is limited to visualizing the correlation between two numerical variables.

Based on the above problems, the question is formulated, namely: "What is a good way to visualize the correlation between two numerical variables?

Aloha International Journal of Multidisciplinary Advancement (AIJMU)
ISSN 2622-3252
Volume 3 Number 1, February 2021 TIPS
http://journal.aloha.academy/index.php/aijmu DOI: http://dx.doi.org/10.33846/aijmu30105

PRINCIPLES OF CORRELATION TEST

Correlation Coefficient

In general, the correlation between two numerical variables is analyzed in a linear correlation test, with the correlation coefficient as the degree of association between two variables [(4)]. There are two directions of correlation, namely positive correlation and negative correlation. In positive correlation, an increase in the score of the first numeric variable, followed by an increase in the score of the second numeric variable; vice versa, a decrease in the score of the first numeric variable, followed by a decrease in the score of the second numeric variable. In physics, this condition is often called "directly proportional to", meaning that the first numeric variable is directly proportional to the second numeric variable. Meanwhile, in negative correlation, the score increases for the first numeric variable, followed by a decrease in the score for the second numeric variable; vice versa, a decrease in the score of the first numeric variable, followed by an increase in the score of the second numeric variable. In physics, this condition is often called "inversely proportional to", meaning that the first numeric variable is inversely proportional to the second numeric variable.

The correlation coefficient (r) is a number that shows the degree of strength of the correlation between two variables. For positive correlation the maximum correlation coefficient is 1 and the minimum correlation coefficient is 0. Whereas for negative correlation, the maximum correlation coefficient is -1 and the minimum correlation coefficient is 0. In percentage terms, 1 = 100% which indicates that the first numerical variable is really fully correlated with the second numeric variable.

Case as Example

For example, the management of a pharmaceutical company provides a bonus of 2 USD (y) per month for each loyal customer recruited by the marketing staff (x). Thus, if tabulated, data on loyal customers recruited and bonuses for each marketing staff can be described in table 1.

Table 1. Distribution of loyal customers recruited and bonuses for marketing staff (by rank)

No	Marketing staff (initial)	Loyal customers recruited	Bonus (USD)
1	A	3	6
2	B	3	6
3	C	4	8
4	D	4	8
5	E	5	10
6	F	8	16
7	G	8	16
8	H	11	22
9	I	12	24
10	J	12	24
11	K	12	24
12	L	12	24
13	M	12	24
14	N	13	26

Based on table 1, it is known that for every 1 loyal customer who can be recruited by a marketing staff, the company management will give an reward in the form of a bonus of 2 USD. In other words, management provides a bonus (in USD) of 2 times the number of regular customers recruited by the marketing staff. This condition can be illustrated in the scatter diagram as shown in figure 1. It appears that the correlation between loyal customers recruited and bonuses forms a perfect linear line, which means that all the points are on a straight line. This linear line runs from the bottom left to the top right, which indicates that the direction of the correlation is positive. It can be said that the greater the x value, the greater the y value. In a case like this, there is a constant value increment, namely: the y value is 2 times the x value, so it can be written in a linear equation:

$$y = 2x$$

Information: y = bonus per month in USD; x = number of loyal customers recruited

Aloha International Journal of Multidisciplinary Advancement (AIJMU)
ISSN 2622-3252
Volume 3 Number 1, February 2021
http://journal.aloha.academy/index.php/aijmu
TIPS
DOI: http://dx.doi.org/10.33846/aijmu30105

Correlation that produces a linear equation like the example above is called perfect correlation, which is indicated by the correlation coefficient = 1. A positive value indicates that one variable is directly proportional to the other variable. Figure 2 is the output of calculating the correlation coefficient based on the Pearson correlation test using the PSPP program.

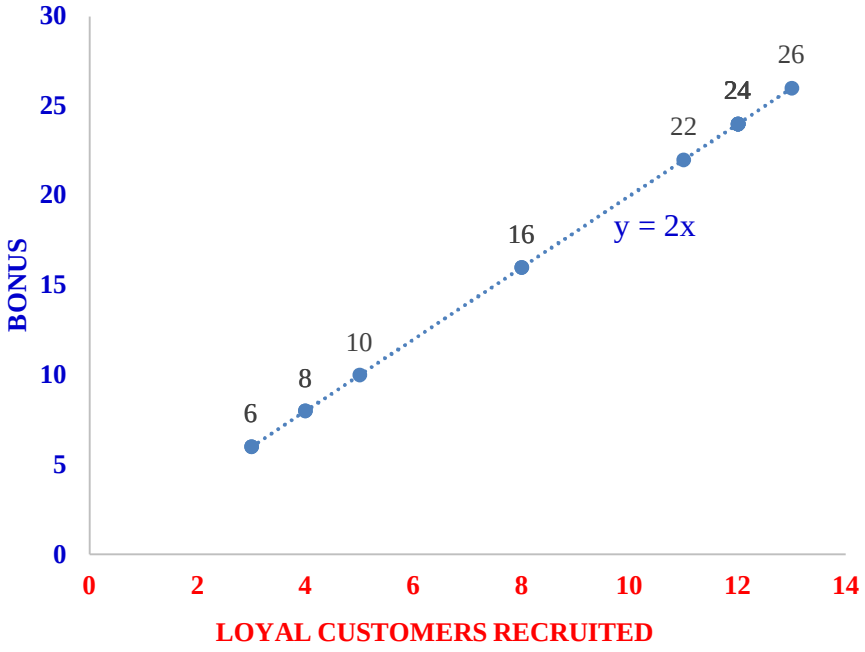


Figure 1. Perfect correlation between loyal customers recruited and bonus

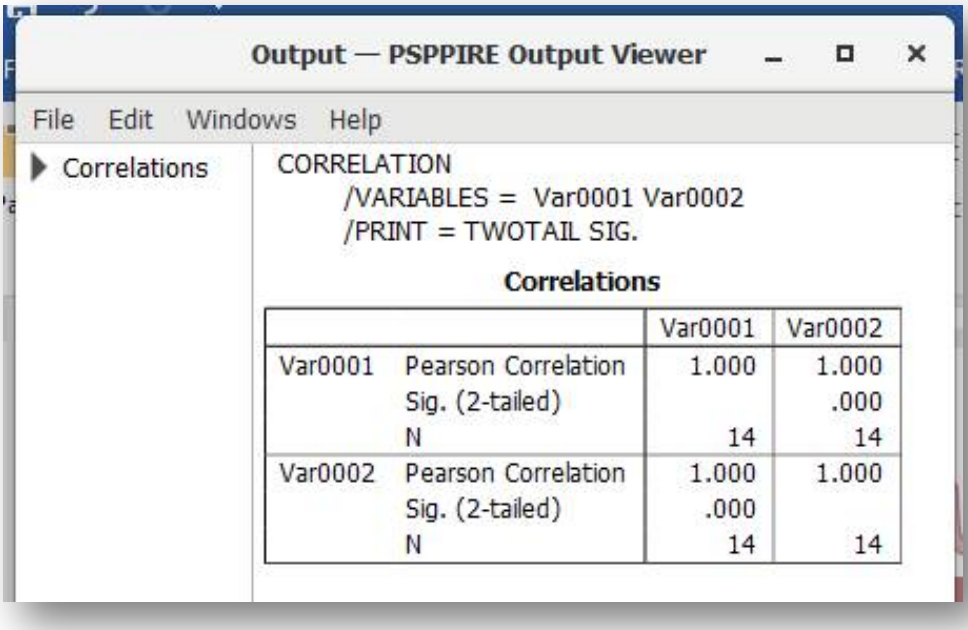


Figure 2. Output of Pearson correlation test using PSPP program (perfect correlation)

Aloha International Journal of Multidisciplinary Advancement (AIJMU)
ISSN 2622-3252
Volume 3 Number 1, February 2021
http://journal.aloha.academy/index.php/aijmu
TIPS
DOI: http://dx.doi.org/10.33846/aijmu30105

To be clearer, let's look at the changes that occur, if the management of the pharmaceutical company not only considers the number of loyal customers recruited, but also considers the discipline of the work of the marketing staff. For example, highly disciplined marketers get a bonus of 2 USD per month for each recruited loyal customer; while the lowly disciplined marketer gets a bonus of only USD 1 per month for each recruited loyal customer. Table 2 shows the changes that have occurred.

Table 2. Distribution of loyal customers recruited and bonuses for marketing staff

No	Marketing staff (initial)	Discipline	Loyal customers recruited	Bonus (USD)
1	A	Low	3	3
2	B	High	3	6
3	C	High	4	8
4	D	High	4	8
5	E	High	5	10
6	F	Low	8	8
7	G	High	8	16
8	H	High	11	22
9	I	Low	12	12
10	J	High	12	24
11	K	High	12	24
12	L	High	12	24
13	M	High	12	24
14	N	Low	13	13

Figure 3 shows that the correlation between recruited loyal customers and bonuses no longer forms a linear line, but there is still a tendency that the points follow the direction of the linear line, starting from the bottom left to the top right, which indicates that the direction of the correlation is positive. It can be said that the greater the x value, the greater the y value, which means that most (10) data points follow the direction of the linear line, and only a few (4) data points do not follow the linear line. In this second case, the increase in the bonus amount varies. It can be seen that the linear line changes to be slightly sloping, due to the decrease in bonus (y) for the first (A), sixth (F), ninth (I) and fourteenth (N) staff, which if it can be written as a linear equation:

$$y = 1.603x + 0.8033$$

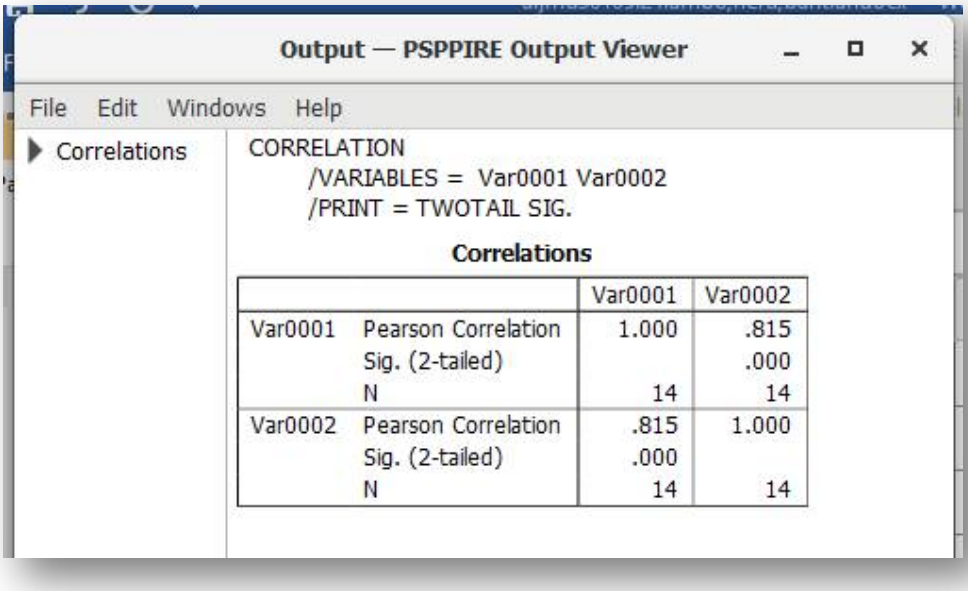
Note: y = bonus per month in USD; x = number of loyal customers recruited

Figure 3. Strong correlation between loyal customers recruited and bonus

Aloha International Journal of Multidisciplinary Advancement (AIJMU)
ISSN 2622-3252
Volume 3 Number 1, February 2021 TIPS
http://journal.aloha.academy/index.php/aijmu DOI: http://dx.doi.org/10.33846/aijmu30105

Figure 4 is the output of calculating the correlation coefficient based on the Pearson correlation test using the PSPP program. With the change in 4 data points, the correlation coefficient has decreased from 1,000 to 0.815. Referring to the classification of the degree of correlation strength according to Ratner [5], the correlation coefficient still shows a strong correlation. The following is a breakdown of the degree of correlation strength based on the correlation coefficient according to Ratner [5]:

1. Correlation coefficient between 0 and 0.3 (0 and -0.3) indicate a weak correlation via a shaky linear rule.
2. Correlation coefficient between 0.3 and 0.7 (-0.3 and -0.7) indicate a moderate correlation via a fuzzy-firm linear rule.
3. Correlation coefficient between 0.7 and 1.0 (-0.7 and -1.0) indicate a strong correlation via a firm linear rule.



		Var0001	Var0002
Var0001	Pearson Correlation	1.000	.815
	Sig. (2-tailed)		.000
	N	14	14
Var0002	Pearson Correlation	.815	1.000
	Sig. (2-tailed)	.000	
	N	14	14

Figure 4. Output of Pearson correlation test using PSPP program (strong correlation)

DISCUSSION

In the previous section it has been suggested that researchers and professionals from various fields of science must understand well the results of correlation analysis [4], which of course also applies to numerical variables. This implies that novice researchers and students as prospective researchers and professionals must be prepared well in advance so that they can early understand the correlation test results they get from scientific journals, proceedings, research reports and other sources. The skills to understand the results of correlation analysis can build and sharpen their ability to perform correlation analysis well in their research projects, especially the ability to present or visualize correlation test results, so that they can be easily understood by readers.

Based on the explanation of the principles of the correlation test as described above, we provide several suggestions, especially for research that uses correlation analysis between numerical variables as a method to answer research statements. First, researchers should formulate a specific hypothesis from the start, namely the presence of a positive correlation or a negative correlation. An example is "there is a positive correlation between loyal customers recruited with bonuses received by pharmaceutical company marketing staff". Of course, to be able to formulate a specific hypothesis, the researcher must have a strong argument based on the justifiable literature. Second, the results of the correlation analysis between numerical variables should be presented visually in the form of a scatter diagram so that it is easier for readers to understand. With a scatter diagram, researchers will be able to explain the results of correlation analysis more easily, and of course this will be very useful for readers to understand its meaning, and also apply it to their studies with similar analyzes.

CONCLUSION

Visualization of the results of correlation analysis between numerical variables in the form of a scatter diagram can help clarify readers' understanding of the results of the analysis, as well as be valuable lessons for readers to later apply them in their research projects.

REFERENCES

1. Nugroho HSW, Badi'ah A. Descriptive Data Analysis for Interval or Ratio Scale Data. Aloha International Journal of Multidisciplinary Advancement (AIJMU). 2019;1(5):121-123.
2. Nugroho HSW. Descriptive Data Analysis for Categorical Data (Analisis Data Secara Deskriptif untuk Data Kategorik). Ponorogo: FORIKES; 2014.
3. Nugroho HSW. Biostatistics for Midwifery Students (Biostatistika untuk Mahasiswa Kebidanan). Ponorogo: FORIKES; 2014.
4. Taylor R. Interpretation of the Correlation Coefficient: A Basic Review. Journal of Diagnostic Medical Sonography. 1990:6(1):35-39.
5. Ratner B. The Correlation Coefficient: Definition [Internet]. DM Stat-1. 2020 [cited 2020 Dec 31]. Available from: http://www.dmstat1.com/res/TheCorrelationCoefficientDefined.html