Understanding Pearson Product-Moment Correlation Coefficient

EXERCISE 13

STATISTICAL TECHNIQUE IN REVIEW

Many studies are conducted to identify relationships between two or more variables. The correlational coefficient is the mathematical expression of the relationship or association studied. Two common analysis techniques are used to examine relationships: Spearman rank-order correlation coefficient, or *rho*, and the Pearson product-moment correlation coefficient, or r (see the algorithm in Exercise 12). The Spearman correlation coefficient (see Exercise 20) is a nonparametric analysis technique conducted to examine relationships when variables are measured at the ordinal level. The **Pearson correlation coefficient** is the most common correlational analysis technique conducted to examine the relationship between two continuous variables, which are measured at the interval or ratio level. The purpose of Pearson r is to examine associations or relationships and not to determine cause and effect between independent and dependent variables (Shadish, Cook, & Campbell, 2002).

Relationships are discussed in terms of direction and strength. The direction of the relationship is expressed as either positive or negative. A **positive or direct relationship** exists when one variable increases as the other variable increases or when one variable decreases as the other decreases. For example, an increase in calorie intake per day is related to increase in weight gain. Conversely, a **negative or inverse relationship** exists when one variable increases and the other variable decreases. For example, an increase in minutes of exercise per day is related to a decrease in weight.

The strength of a relationship is described as weak, moderate, or strong. Pearson r is never greater than −1.00 or +1.00, so an r value of −1.00 or +1.00 indicates the strongest possible relationship, either negative or positive, respectively. An r value of 0.00 indicates no relationship or association between two variables. To describe a relationship, the labels weak ($r < 0.3$), moderate ($r = 0.3$ to 0.5), and strong ($r > 0.5$) are used in conjunction with both positive and negative values of r. Thus, the strength of the negative relationships would be weak with $r < -0.3$, moderate with $r = -0.3$ to -0.5, and strong with $r > -0.5$ (Cohen, 1988; Plichta & Kelvin, 2013).

The significance of r values can be determined by examining the Table of Critical Values for r for the Pearson product-moment correlation coefficient in Appendix B at the back of this text. To use this table, you need to know the level of significance or alpha for the study, which is usually set at 0.05. The degrees of freedom (df) for Pearson r is the sample size minus 2 ($N - 2$). For example, if a study had $r = 0.36$, a sample size of 50, and $\alpha = 0.05$, is the r value statistically significant? The answer is yes because the $r = 0.36$ and $df = 48$ is larger than the critical table value 0.2787 at $\alpha = 0.05$ for a two-tailed test. The

significance of r values can also be determined with an independent samples t-test (see Exercise 28 for the process of calculating Pearson r and determining the significance).

Mirror-Image Table of Pearson *r* Results

A mirror-image table, as the name implies, has the same labels in the same order for both the x- and y-axes (see Exercise 7 for a discussion of x- and y-axes). Frequently, numbers or letters are assigned to each label, and only the letter or number designator is used to label one of the axes. To find the r value for a pair of variables, look both along the labelled or y-axis in Table 13-1 and then along the x-axis, using the number designator assigned to the variable you want to know the relationship for, and find the cell in the table with the r value. Table 13-1 is an example of a mirror-image table that displays the relationships among the variables hours of class attended, hours studying, and final grade as a percentage. The results in the table are intended as an example of a mirror-image table and are not based on research. If you were asked to identify the r value for the relationship between hours of class attended and the final grade as a percentage, the answer would be $r = 0.52$; for between hours studying and final grade as a percentage, the answer would be $r = 0.58$. The dash (–) marks located on the diagonal line of the table represent the variable's correlation with itself, which is always a perfect positive correlation, or $r = +1.00$. Since the results are the same in both sides of the mirror-image table, most researchers only include half of the table in their final report (Grove, Burns, & Gray, 2013; Plichta & Kelvin, 2013).

TABLE 13-1 MIRROR-IMAGE PEARSON *r* CORRELATION TABLE

Variables	1	2	3
1. Hours of class attended	–	0.34	0.52
2. Hours studying	0.34	–	0.58
3. Final grade as a percentage	0.52	0.58	–

Effect Size of an *r* Value

The Pearson r value is equal to the effect size (ES) or the strength of a relationship between two variables. In the previous table, the association between hours of class attended and hours of studying is $r = 0.34$; thus the $r = ES = 0.34$ for this relationship. The ES is used in power analysis to determine sample size and examine the power of studies. (Exercise 24 describes the elements of power analysis, and Exercise 25 includes the steps for conducting a power analysis.) The strength of the ES is the same as that for the r values, with a weak effect size as one <0.3 or <-0.3, a moderate effect size 0.3 to 0.5 or -0.3 to -0.5, and a strong effect size >0.5 or >-0.5. The smaller the ES, the greater the sample size needed to detect significant relationships in studies. A large ES or association between two variables usually requires a smaller sample size to identify a significant relationship (Cohen, 1988). Correlational studies usually involve examining relationships among a number of variables, so a large sample size is important.

Percentage of Variance Explained in a Relationship

The **percentage of variance explained** is calculated to increase understanding about the relationship between two variables in terms of clinical importance. The percentage of variance explained is calculated with the Pearson r value. To calculate the percentage of variance explained, square the r value and multiply by 100% to determine a percentage (Cohen, 1988).

Formula: % variance explained = $r^2 \times 100\%$

Example: $r = 0.58$ (correlation between hours studying and final grade as a percentage)
$(0.58)^2 \times 100\% = 0.3364 \times 100\% = 33.64\%$ variance explained

In this example, the hours studying explains 33.64% of the variance in the final course grade. Calculating the percentage of variance explained helps researchers and consumers of research understand the practical implications of reported results. The stronger the r value, the greater the percentage of variance explained. For example if $r = 0.5$, then 25% of the variance in one variable is explained by another variable; if $r = 0.6$, then 36% of the variance is explained. All Pearson r values need to be examined for clinical importance, but the potential for clinical importance increases with moderate and strong relationships, where $r \geq 0.3$ and yields a 9% or higher variance explained. Keep in mind that a result may be statistically significant ($p \leq 0.05$), but it may not represent a clinically important finding. The r values in a study need to be examined for statistical significance and clinical importance.

RESEARCH ARTICLE

Source

Ulrich, C. M., Zhou, Q., Hanlon, A., Danis, M., & Grady, C. (2014). The impact of ethics and work-related factors on nurse practitioners' and physician assistants' views on quality of primary healthcare in the United States. *Applied Nursing Research, 27*(3), 152–156.

Introduction

Ulrich and colleagues (2014) noted that many primary care services are provided by nurse practitioners (NPs) and physician assistants (PAs), and it is important for them to have ethical knowledge in managing challenging practice issues. However, little is known about the relationships between ethics and other work-related factors with the delivery of quality care. Thus the purpose of this study was "to describe how NP and PA perceptions of their ethics preparedness, confidence, physician collegiality, and autonomy, along with the patient demands they encountered, influence perceived quality of care in their clinical practice" (Ulrich et al., 2014, p. 153). The findings from this correlational study were as follows: "Ethics preparedness and confidence were significantly associated with perceived quality of care ($p < 0.01$) as were work-related characteristics such as percentage of patients with Medicare and Medicaid, patient demands, physician collegiality, and practice autonomy ($p < 0.01$).... The researchers concluded that providing ethics education and addressing restrictive practice environments might improve teamwork, collaborative practice, and quality of care" (Ulrich et al., 2014, p. 152).

Relevant Study Results

"Three-quarters of respondents reported that they could provide high quality care to all their patients. However, 33.4% also indicated that cost concerns influence the degree to which they can provide quality care, and one of four (25.3%) did not feel that they could make clinical decisions without cost pressures. Higher levels of ethics preparedness, ethics confidence, practice autonomy, and physician collegiality were significantly associated with better perceived quality care ($p < 0.01$; Table 2). Two-thirds of practitioners (65.9%) reported a high level of physician support, reporting that their physician colleagues

TABLE 2 DESCRIPTIVE STATISTICS OF PRIMARY VARIABLES AND ASSOCIATIONS AMONG THEM (PEARSON *r*) (*N* = 1371)

	Mean (*SD*)	1	2	3	4	5
1. Perceived quality of care	35.80 (5.58)	1.00				
2. Ethics preparedness	24.13 (4.14)	0.37**	1.00			
3. Ethics confidence	23.23 (4.57)	0.29**	0.57**	1.00		
4. Patient care demand	11.80 (2.81)	−0.42**	−0.27**	−0.22**	1.00	
5. Practice autonomy	16.68 (3.23)	0.53**	0.24**	0.13**	−0.32**	1.00
6. Physician collegiality†	15.93 (3.02)	0.25**	0.19**	0.02	−0.12**	0.13**

**$p < 0.01$.
†Negatively skewed. This variable was then dummy coded with low and high physician collegiality, respectively (less than or equal to 12 = low; 12.1%), greater than 12 = high; 87.9%.

Ulrich, C. M., Zhou, Q., Hanlon, A., Danis, M., & Grady, C. (2014). The impact of ethics and work-related factors on nurse practitioners' and physician assistants' views on quality of primary healthcare in the United States. *Applied Nursing Research, 27*(3), p. 155.

valued their unique perspective, were a good source of professional stimulation, and provided an important source of personal support. Nonetheless, 31.6% of respondents reported feeling overwhelmed by the needs of their patients, and 40.8% described patient relationships as becoming more adversarial than they used to be. Overall, those respondents who reported higher patient demands (e.g., adversarial patient relationships, patients requesting unnecessary treatments) in their practice also tended to report lower quality care ($r = -0.42$, $p < 0.01$). Most respondents felt prepared to handle the ethical issues that might arise in their primary care practice (63%); and both ethics preparedness and respondents' degree of ethics confidence were positively associated with perceptions of quality care" (Ulrich et al., 2014, p. 154).

STUDY QUESTIONS

1. What is the value of the Pearson *r* for the relationship between perceived quality of care and ethics preparedness?

2. Is the relationship between perceived quality of care and ethics preparedness positive or negative? What is the strength of the relationship? Provide a rationale for your answers.

3. Is the relationship between perceived quality of care and ethics preparedness statistically significant? Provide a rationale for your answer.

4. What do the 1.00 values along the diagonal of Table 2 mean?

5. Discuss the quality of the sample size in this study. Document your answer.

6. Which two variables have the strongest positive relationship? Is this an expected finding? Provide a rationale for your answer.

7. What is the Pearson *r* and effect size (*ES*) for the association between perceived quality of care and practice autonomy? Describe this relationship using words rather than numbers.

8. What percentage of variance is explained by the relationship between the variables perceived quality of care and practice autonomy? Show your calculations and round your answer to two decimal places.

9. Is the relationship between perceived quality of care and practice autonomy statistically significant? Is this relationship clinically important? Provide rationales for your answers.

10. Do the results in Table 2 support the statement that practice autonomy causes perceived quality of care? Provide a rationale for your answer.

Answers to Study Questions

1. $r = 0.37^{**}$. The r value is listed in Table 2 and can be identified by locating where variable 1 perceived quality of care intersects with variable 2, ethics preparedness.

2. $r = 0.37^{**}$ is a positive relationship, where an increase in ethics preparedness is associated with or related to an increase in perceived quality of care. This is a moderate relationship since it lies between 0.30 and 0.50 (Cohen, 1988; Grove et al., 2013).

3. Yes, the relationship between perceived quality of care and ethics preparedness is statistically significant. The ** indicate that the r value is statistically significant since its probability or $p < 0.01$ as identified in the footnote for Table 2. Most nurse researchers set their level of significance or alpha (α) = 0.05. Since $p < 0.01$ is smaller than α, the relationship is significant.

4. When a variable is correlated with itself the $r = 1.00$, a perfect positive relationship. The $r = 1.00$ runs along the diagonal of Table 2 and represents each variable's relationship or association with itself.

5. The sample size, $N = 1{,}371$, is extremely strong for the number of variables studied (Grove et al., 2013). Correlational studies need large sample sizes to detect small associations between variables, which is accomplished in this study (Cohen, 1988; Plichta & Kelvin, 2013).

6. $r = 0.57^{**}$ is the strongest positive relationship, which is the association between ethics preparedness and ethics confidence. This is an expected finding because healthcare providers who have increased ethics preparedness would probably have increased ethics confidence, which is supported in this study with a strong r value.

7. $r = ES = 0.53^{**}$ (see Table 2). Pearson $r = 0.53$ represents a strong, positive relationship between perceived quality of care and practice autonomy since the $r > 0.50$ (Cohen, 1988; Grove et al., 2013). This correlation indicates that as practice autonomy increases, perceived quality of care increases, or as practice autonomy decreases, perceived quality of care decreases.

8. $r^2 \times 100\%$ = % variance explained by a relationship. Calculation: $(0.53)^2 \times 100\% = 0.2809 \times 100\% = 28.09\%$ variance explained by the relationship between perceived quality of care and practice autonomy.

9. $r = 0.53^{**}$ is statistically significant as indicated by ** described in the footnote of Table 2 as $p < 0.01$. Most nurse researchers set their level of significance or alpha (α) = 0.05. Since $p < 0.01$ is smaller than α, the relationship is statistically significant. The relationship is also clinically important since the percent of variance explained by this relationship is high, 28.09% (see the answer to Question 8). This means that 28.09% of the variance in perceived quality of care is explained by practice autonomy, which is important for NPs, PAs, and administrators to know in setting guidelines for practice sites.

10. No. Pearson *r* is calculated to examine a relationship or association between two variables and does not determine causality between the variables. Relationship indicates that two variables are linked to each other but not that one variable causes the other. Causality indicates a strong relationship between two variables, but one of the variables needs to precede the other in time and be present when the effect occurs (Shadish et al., 2002).

Questions to Be Graded

EXERCISE 13

Name: _____ Class: _____

Date: _____

Follow your instructor's directions to submit your answers to the following questions for grading. Your instructor may ask you to write your answers below and submit them as a hard copy for grading. Alternatively, your instructor may ask you to use the space below for notes and submit your answers online at http://evolve.elsevier.com/Grove/Statistics/ under "Questions to Be Graded."

1. What is the value of the Pearson r for the relationship between variables 1 and 6? Identify the variables correlated in this relationship.

2. Describe the correlation $r = -0.42^{**}$ using words. Is this a significant correlation? Provide a rationale for your answer.

3. What percentage of variance is explained by the relationship between the variables perceived quality of care and patient care demands? Show your calculations and round your answer to two decimal places.

4. Is the relationship between the variables perceived quality of care and patient care demands clinically important? Provide a rationale for your answer.

5. What Pearson r value was calculated for the relationship between ethics confidence and physician collegiality? What are the effect size (ES) and strength of this relationship?

6. Is the relationship between perceived ethics confidence and physician collegiality statistically significant? Provide a rationale for your answer.

7. Is the relationship between perceived ethics confidence and physician collegiality clinically important? Provide a rationale for your answer.

8. In a hypothetical study, perceived stress and perceived quality of life were correlated at Pearson $r = 0.39$. The sample size was $N = 62$, $\alpha = 0.05$, and the data were examined with a two-tailed test. Using the Table of Critical Values for Pearson r in Appendix B, determine if $r = 0.39$ is statistically significant. Provide a rationale for your answer.

9. How did the study participants describe the patient care demands?

10. How might the Ulrich et al. (2014) study results be used in practice?