

# Correlation: A Measure of Association

## OVERVIEW

---

In this lab, you will work with the method of statistical **correlation**. This lab is intended to enhance your basic knowledge of correlation. The concepts you have learned in Research Methods I lecture and Elementary Statistics regarding correlation will be practiced and applied in this lab.

### OBJECTIVES

By the end of the laboratory, you will be able to

- Understand the general concepts of statistical correlation.
- Create and interpret a scatterplot using *Minitab*.
- Compute by hand a Pearson's Correlation Coefficient.
- Perform and interpret a Pearson's Correlation Coefficient using *Minitab*.
- Understand the concepts of correlation as they apply to psychology.
- State conclusions in APA format.

### EQUIPMENT

- PC with *Minitab*
- Access to a printer
- Computer diskette to save files
- Scientific calculator

## BACKGROUND MATERIAL

---

Statistical Terms and Topics

- Positive correlation
- Negative correlation
- Perfect correlation
- Zero or near-zero correlation
  - Scatter plot
  - Correlation coefficient

Formula

<p><b>Pearson's Correlation Coefficient:</b></p>	$r = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$
--	---

Scenario

University students were asked in a survey questions such as their height, shoe size, number of children in their family, number of pets and how many dates they have been on in the past two weeks. From the data we have collected, we will determine if there is a relationship between any two of these variables.

**Data Set #1:**

Data set number one contains the number of dates in the past two weeks and the heights of 51 students surveyed. Step-by-step instructions for entering the data are given below.

	Column 1	Column 2			Column 1 continued	Column 2 continued
Participant Number	Dates in the past two weeks	Height (inches)		Participant Number	Dates in the past two weeks	Height (inches)
1	8	72		26	0	68
2	0	71		27	0	64
3	0	71		28	0	63
4	4	70		29	*	63
5	3	69		30	2	66
6	0	69		31	0	66
7	0	70		32	0	64
8	0	77		33	1	63
9	0	69		34	2	66
10	4	73		35	1	67
11	4	70		36	0	67
12	14	70		37	0	68
13	*	72		38	0	64
14	2	67		39	0	71
15	0	64		40	1	68
16	14	66		41	0	67
17	4	70		42	0	74
18	3	66.5		43	0	9
19	4	66		44	0	61

20	3	67		45	2	71
21	1	65		46	0	68
22	2	69		47	0	68
23	5	69		48	2	67
24	0	69		49	4	74
25	0	70		50	0	68
	0			51	0	71



## COMPUTER EXERCISE

1. Label the first column (C1) **dates** and the second column (C2) **height**.
2. Enter the data into *Minitab*.
3. **Save the data** as a project entitled *correlationdata*. (Continue saving often!)
4. Run the **descriptive statistics** for each variable. (Refer to the *Intro to Minitab Laboratory* as needed.)
5. Record your answers in the spaces provided below.

Dates in the past two weeks:

$n$  \_\_\_\_\_  $n^*$  \_\_\_\_\_ mean \_\_\_\_\_ median \_\_\_\_\_  
standard deviation \_\_\_\_\_

Height:

$n$  \_\_\_\_\_  $n^*$  \_\_\_\_\_ mean \_\_\_\_\_ median \_\_\_\_\_  
standard deviation \_\_\_\_\_

6. Create a Scatterplot using *Minitab*. The steps follow.
  - Go to **GRAPH>PLOT**
  - Choose column labeled dates for **Y** and column labeled height for **X**.
  - Click **OK**.
  - **Save** graph as *scatterplot1*. Go to **FILE > SAVE GRAPH AS**.
  - Print the scatterplot. Go to **FILE > PRINT GRAPH**.

7. Based on the Scatterplot, what type of relationship do you predict and why?

---



---

8. To gain a better understanding of the mathematical procedures behind a Pearson's Correlation Coefficient it is beneficial to work the formula by hand. Using the formula for a Pearson's Correlation Coefficient, compute by-hand  $r$  for the **first five data points (listed in the table below)** for dates and height.

Heights <b>x</b>	Dates <b>y</b>
72.0	8
71.0	0
71.0	0
70.0	4
69.0	3

<b>Pearson's Correlation Coefficient:</b>	$r = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$
---	---

In order to find the correlation coefficient, we see from the formula that we will need the mean and standard deviation of both  $x$  and  $y$ . It is possible to find the descriptive statistics both by hand or by using *Minitab*. For this exercise, however, we have listed the descriptive statistics for you. Use these descriptive statistics to complete the table.

$\bar{x} = \underline{70.6}$	$s_x = \underline{1.1}$
$\bar{y} = \underline{3}$	$s_y = \underline{3.3}$

The following is a table designed to demonstrate and simplify the mathematics of a Pearson's Correlation Coefficient. Round to the nearest tenths (one decimal place). Helpful explanations of each column are provided below the table.

		<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>x</b>	<b>y</b>	$x - \bar{x}$	$\frac{x - \bar{x}}{s_x}$	$y - \bar{y}$	$\frac{y - \bar{y}}{s_y}$	$\left( \frac{x - \bar{x}}{s_x} \right) \left( \frac{y - \bar{y}}{s_y} \right)$
72.0	8					
71.0	0					
71.0	0					
70.0	4					

69.0	3					
------	---	--	--	--	--	--

Total = \_\_\_\_\_ (The sum of column E equals "Total.")  
 $r = \text{Total} / n-1 = \underline{\hspace{2cm}}$

The following list explains in more detail the steps completed in the table.

- A consists of the  $x$  values.
- B consists of the  $y$  values.
- C is each individual  $x$  value minus  $\bar{x}$  / in other words,  $x - \bar{x}$ .
- B is A divided by the standard deviation for  $x$  ( $s_x$ ).
- C is each individual  $y$  value minus  $\bar{y}$  / in other words,  $y - \bar{y}$ .
- D is C divided by the standard deviation for  $y$  ( $s_y$ ).
- E is the product of B and D.

9. Listed below are the results from *Minitab* calculating the sample correlation coefficient for the five data points. Check the results of your "by-hand" computations from the table.

$r = \underline{0.264}$

10. Does your (by-hand)  $r$  - value agree with *Minitab's*  $r$  - value to the first decimal place?  
 \_\_\_\_\_

11. Choose a value for alpha; then use *Minitab* to compute a Pearson's Correlation Coefficient for the complete data set #1 for **dates** and **height**. The steps follow.

Go to **STAT > BASIC STATISTICS > CORRELATION**.

In the **Variables** box, select **dates** and **heights**.

Click **OK**.

Record the answers found in the session window.

$\alpha = \underline{\hspace{2cm}}$        $r = \underline{\hspace{2cm}}$        $p\text{-value} = \underline{\hspace{2cm}}$

12. What is the null hypothesis for Pearson's Correlation? *Hint: go to the help box under correlation.*  
 \_\_\_\_\_

13. What conclusions can you make based on the results? Remember, correlation does not mean causation.  
 \_\_\_\_\_  
 \_\_\_\_\_

14. What is the direction of the relationship?  
 \_\_\_\_\_

## Rest of the Data (Data sets 2-4)

Do not erase the data in columns 1 and 2. (You will need that data for the rest of the analysis.) Enter **shoe size in column 3**, number of **children** in family in **column 4** and **number of pets in column 5**.

	Column 3	Column 4	Column 5			Column 3 continued	Column 4 Continued	Column 5 continued
Participant Number	Shoe Size	Number of Children in family	Number of Pets		Participant Number	Shoe Size	Number of children in family	Number of Pets
1	12	2	2		26	9	2	3
2	11.5	3	1		27	7	2	2
3	11	2	1		28	7	3	1
4	9	2	0		29	8	2	3
5	8	8	6		30	7	2	0
6	11	2	6		31	8	1	2
7	10	2	5		32	9	2	1
8	13	3	7		33	9	3	1
9	11	4	5		34	8	3	2
10	11	1	2		35	8	3	2
11	11	2	1		36	9	3	0
12	9	2	1		37	9	2	2
13	13	3	*		38	6	2	1
14	9	2	0		39	12	2	1
15	8	2	7		40	8	2	6
16	9	1	0		41	10	3	0
17	9	3	1		42	11	3	0
18	8	2	1		43	10	3	7
19	9	5	3		44	8	3	2
20	10	4	1		45	10	2	2
21	8	3	11		46	10	3	0
22	10	2	1		47	11	4	2
23	9	3	2		48	9	4	0
24	11	3	1		49	13	2	1
25	11	4	5		50	11	2	1
					51	10	3	1

## INSTRUCTIONS

---

### Scenario

For college students, is there a relationship between a person's shoe size and the number of dates they have during a two-week period. Work through the steps below to help answer the question.

1. Use the data you have entered to complete the following exercise for **shoe size (C3)** and **dates (C1)**. (You can refer to this as data set #2).

1. Run the descriptive statistics for **shoe size** and **dates**.

Shoe Size:

$n$  \_\_\_\_\_  $n^*$  \_\_\_\_\_ mean \_\_\_\_\_ median \_\_\_\_\_  
standard deviation \_\_\_\_\_

Dates in the Past Two Weeks:

$n$  \_\_\_\_\_  $n^*$  \_\_\_\_\_ mean \_\_\_\_\_ median \_\_\_\_\_  
standard deviation \_\_\_\_\_

3. Create and print a Scatterplot using *Minitab*. Choose columns  $x$  and  $y$  as you feel they should be labeled. (Save as *scatterplot2*).

4. Based on the Scatterplot, what type of relationship do you predict and why?

---

---

---

5. Choose a value for alpha, and compute a Pearson's Correlation Coefficient using *Minitab*.

$\alpha =$  \_\_\_\_\_  $r =$  \_\_\_\_\_  $p$ -value = \_\_\_\_\_

6. Based on the results, will you reject the null at the .05 significance level?

---

---

7. What conclusions can you make based on the results?

---

---

---

8. What is the direction of the relationship?

---

Use *Minitab* to print scatterplots and compute Pearson's Correlation Coefficients for data sets 3 and 4. (Save these as *scatterplot3* and *scatterplot4*). Be sure to choose alpha and record it BEFORE using *Minitab* to find the correlation coefficient and the *p*-value. Record alpha, correlation coefficients and *p*-values. What conclusions can be made from the results of these data sets?

### **Data Set #3: Number of children in family and number of pets**

#### Scenario

Measure the association between the number of children in a family and the number of pets. Do you expect there to be a strong association? \_\_\_\_\_ would you expect the association to be positive or negative? \_\_\_\_\_ For data set #3 use the data in columns 4 and 5 to find the correlation coefficient for number of children in family and pets.

$\alpha =$  \_\_\_\_\_                       $r =$  \_\_\_\_\_                       $p - \text{value} =$  \_\_\_\_\_

Conclusions:

---

---

---

---

### **Data Set #4: Dates in the past two weeks and number of pets**

## Scenario

Measure the association between the number of dates a person has in the last two weeks and the number of pets they have. Do you expect there to be a strong association? \_\_\_\_\_ Would you expect the association to be positive or negative? \_\_\_\_\_ For data set #4, use the data in columns 1 and 5 to find the correlation coefficient for dates in the past two weeks and number of pets.

$\alpha =$  \_\_\_\_\_

$r =$  \_\_\_\_\_

$p$ -value = \_\_\_\_\_

Conclusions:

---

---

---

---

## Application to Psychology

In psychology, it is extremely important that research results are communicated to other psychologists as well as to the community. To ensure that research is communicated effectively and consistently, the format for research reports is standardized. There are specific steps and formats to follow when publishing research. The format for a research paper is found in the APA Manual. In a research report, there is a specific section for *Introduction, Methods, Procedures, Results and Discussion*. The Results section contains the statistical results, and as the rest of the paper, follows a specific format. When presenting the results of a correlation, The American Psychological Association (APA) requires that you include the correlation coefficient ( $r$ ), degrees of freedom (df), and the  $p$ -value. (When doing research you will probably always use a computer to calculate results such as the correlation coefficient and  $p$ -value. However, a  $p$ -value can be calculated by hand using the degrees of freedom and a table. When writing in APA format it is important to know the degrees of freedom for a Pearson's Correlation. Although it is likely that a computer will be used to generate the  $p$ -value, you must still record the degrees of freedom. **The number of degrees of freedom for a Pearson's  $r$  is  $n-2$ .**)

Example:  $r = .75$ ;  $df = 59$ ;  $p$ -value = .03 would be written as:

$$r(59) = .75, p < .05$$

The following is an excerpt from an abstract on the relationship of positive and negative perfectionism to anxiety and obsessive compulsive behavior. Note how the statistics are recorded.

There was a positive correlation between negative perfectionism (MEC) and anxiety,  $r(40) = .32, p < .05$ . There was also a positive correlation between

negative perfectionism (MEC) and obsessive compulsive behavior,  $r(40) = .26, p < .05$ . The totals of the Maudsley Obsessive Compulsive Inventory and the totals of the State Trait Anxiety Inventory also revealed a positive correlation,  $r(40) = .44, p < .05$ . There was a very weak correlation between positive perfectionism (POS) and anxiety,  $r(40) = -.05, p > .05$ . There was also a weak correlation between positive perfectionism (POS) and obsessive compulsive behavior,  $r(40) = -.12, p > .05$ .

Brummett, R. & Yandell, L. *Positive and Negative Perfectionism: Relationships to Anxiety and Obsessive Compulsive Behavior*. (1999).

- What is the correlation coefficient for negative perfectionism and anxiety?
- $r =$  \_\_\_\_\_
- What is the correlation coefficient for negative perfectionism and obsessive compulsive behavior?  $r =$  \_\_\_\_\_
- What are the degrees of freedom for this project?  $df =$  \_\_\_\_\_
- What was the sample size of this study?  $n =$  \_\_\_\_\_
- In what direction is the correlation between positive perfectionism and obsessive compulsive behavior?  
 \_\_\_\_\_
- Which correlation(s) are significant at the .05 significance level?  
 \_\_\_\_\_  
 \_\_\_\_\_
- What conclusions can be drawn from the results given in this abstract?  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_



## Assignment 1

**Record the results from Data Sets 3 and 4 in APA format below.**

### **Data Set 3: Number of children in family and the number of pets**

---



---



---

## **Data Set 4: Number of Dates in the past two weeks and the number of pets**

---

---

---

### Ethics Application

When working with surveys, participants' identities must be kept confidential. To ensure confidentiality, participant identities are always kept separate from raw data. In the scenario used in this lab however, participant identities were never obtained.