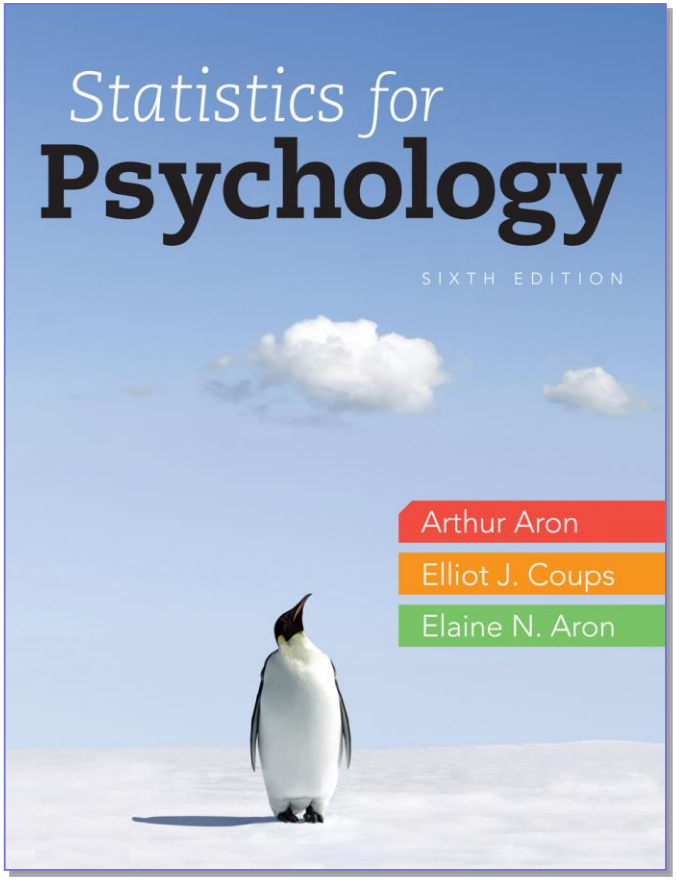
Statistics for **Psychology**

SIXTH EDITION



CHAPTER 11

Correlation

Correlation

- a statistical procedure that allows you to look at the relationship between two groups of scores.
- Some real-world examples.
  - Among students, there is a relationship between high school grades and college grades.
    - It isn’t a perfect relationship, but generally speaking students with better high school grades tend to get better grades in college.
  - There is a relationship between parents’ heights and the adult height of their children.
    - Taller parents tend to give birth to children who grow up to be taller than the children of shorter parents.
    - Again, the relationship isn’t perfect, but the general pattern is clear.

Correlation

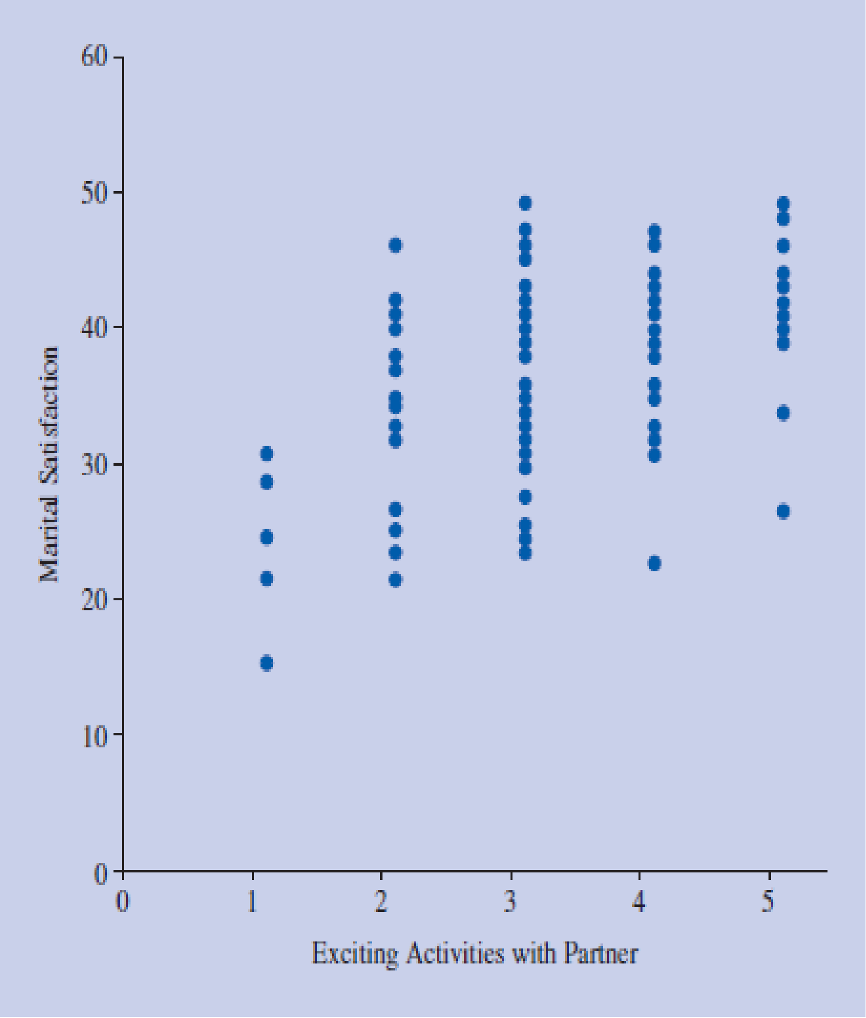
- The researchers were interested in finding out the relationship between *doing exciting things with a marital partner* and *the level of marital satisfaction people reported*.
- How exciting are the things you do together with your partner?
- In general, how often do you think that things between you and your partner are going well?”).
- a scale from 1 = *not exciting at all to* 5 = *extremely exciting.*

Correlation

- The relationship between these two groups of scores can be shown very clearly using a graph.
- The horizontal axis is for people’s answers to the question, “How exciting are the things you do together with your partner?”
- The vertical axis is for the marital satisfaction scores. Each person’s score on the two variables is shown as a dot.

Correlation

What is
the pattern?



Correlation

- Correlation describes the relationship between two variables.
- More precisely, the usual measure of a correlation describes the relationship between *two equal-interval numeric variables*
- Association between scores on two variables
  - Ex.: age and coordination skills in children, price and quality

Correlation

- Correlationcan be visualized from a scatter graph (scatterplot)
- *The pattern of the relationship b/w two variables can be seen in scatter graph*

Graphing Correlations: The Scatter Diagram -1

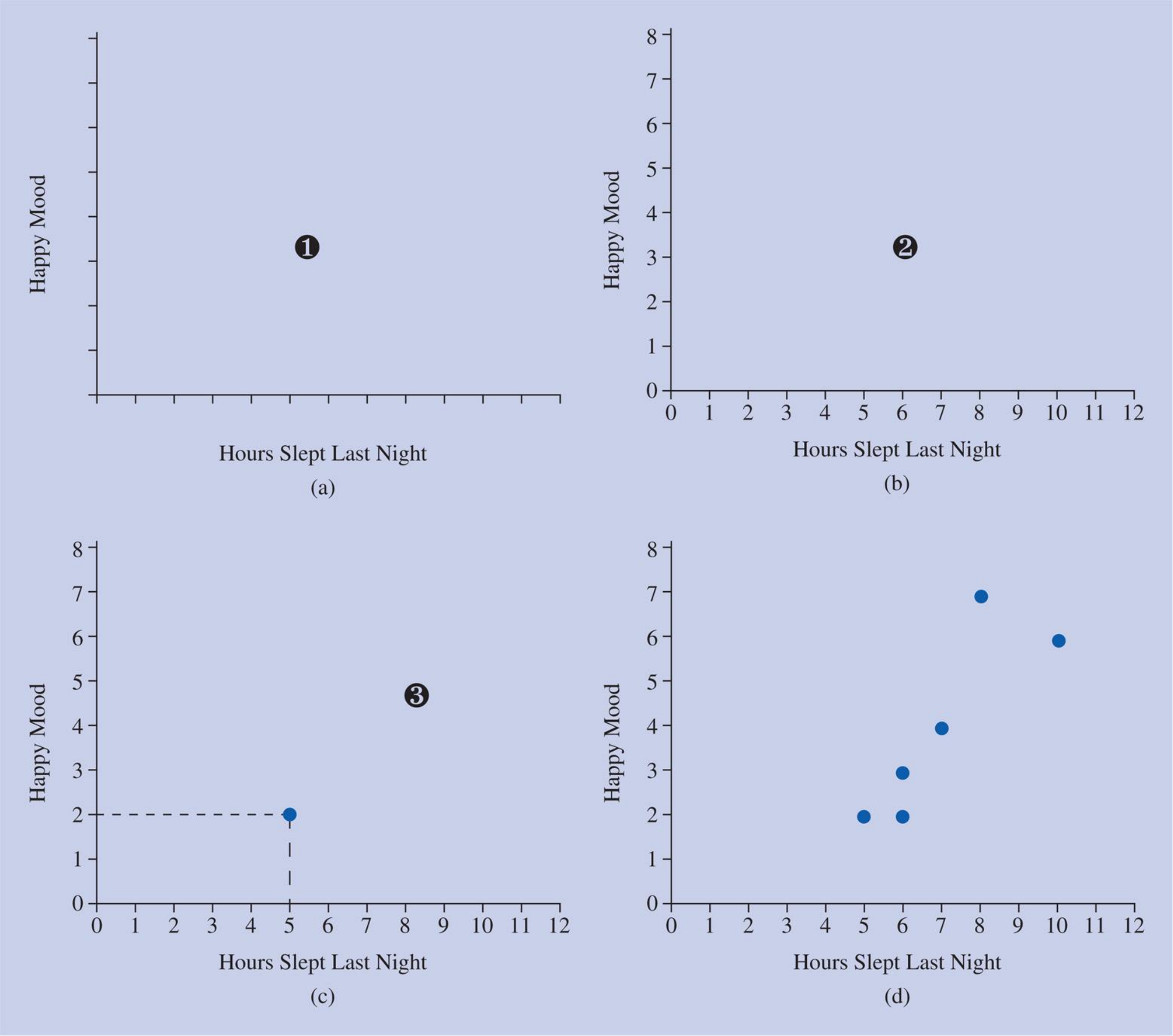
- Steps for making a scatter diagram
  1. Draw axes and assign variables to them
  2. Determine range of values for each variable and mark on axes
  3. Mark a dot for each person's pair of scores

Table 11-1 Hours Slept Last Night and Happy Mood Example (Fictional Data)

Hours Slept	Happy Mood
5	2
7	4
8	7
6	2
6	3
10	6

Table 11-1 Hours Slept Last Night and Happy Mood Example (Fictional Data)

Figure 11-2 Steps for making a scatter diagram. (a) (1) Draw the axes and decide which variable goes on which axis—the predictor variable (Hours Slept Last Night) on the horizontal axis, the other (Happy Mood) on the vertical axis. (b) (2) Determine the range of values to use for each variable and mark them on the axes. (c) (3) Mark a dot for the pair of scores for the first student. (d) (3) continued: Mark dots for the remaining pairs of scores.



Example

- Please make a sctterdiagram for the below data

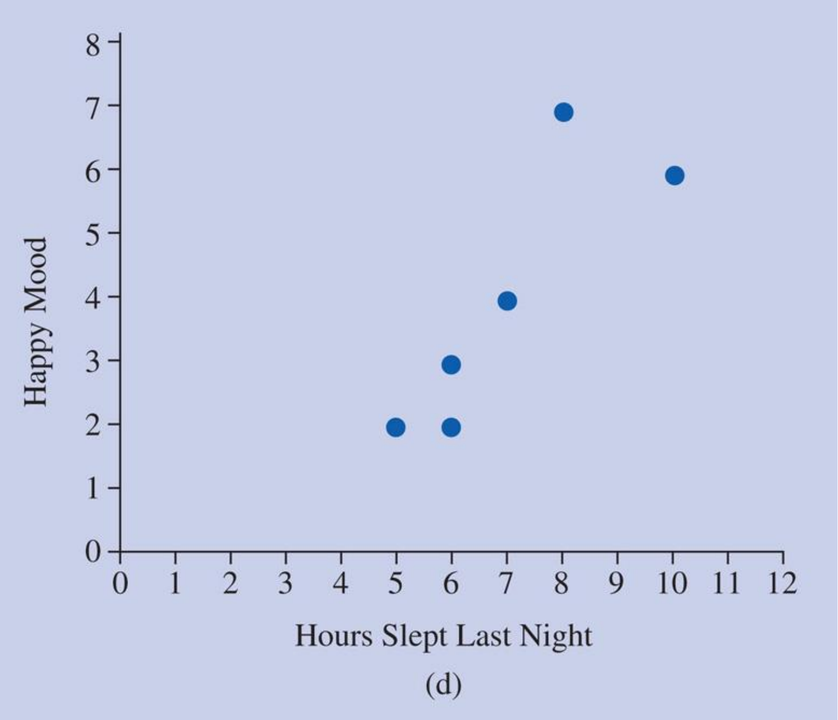
Person	X	Y
A	3	4
B	6	7
C	1	2
D	4	6

Patterns of Correlation

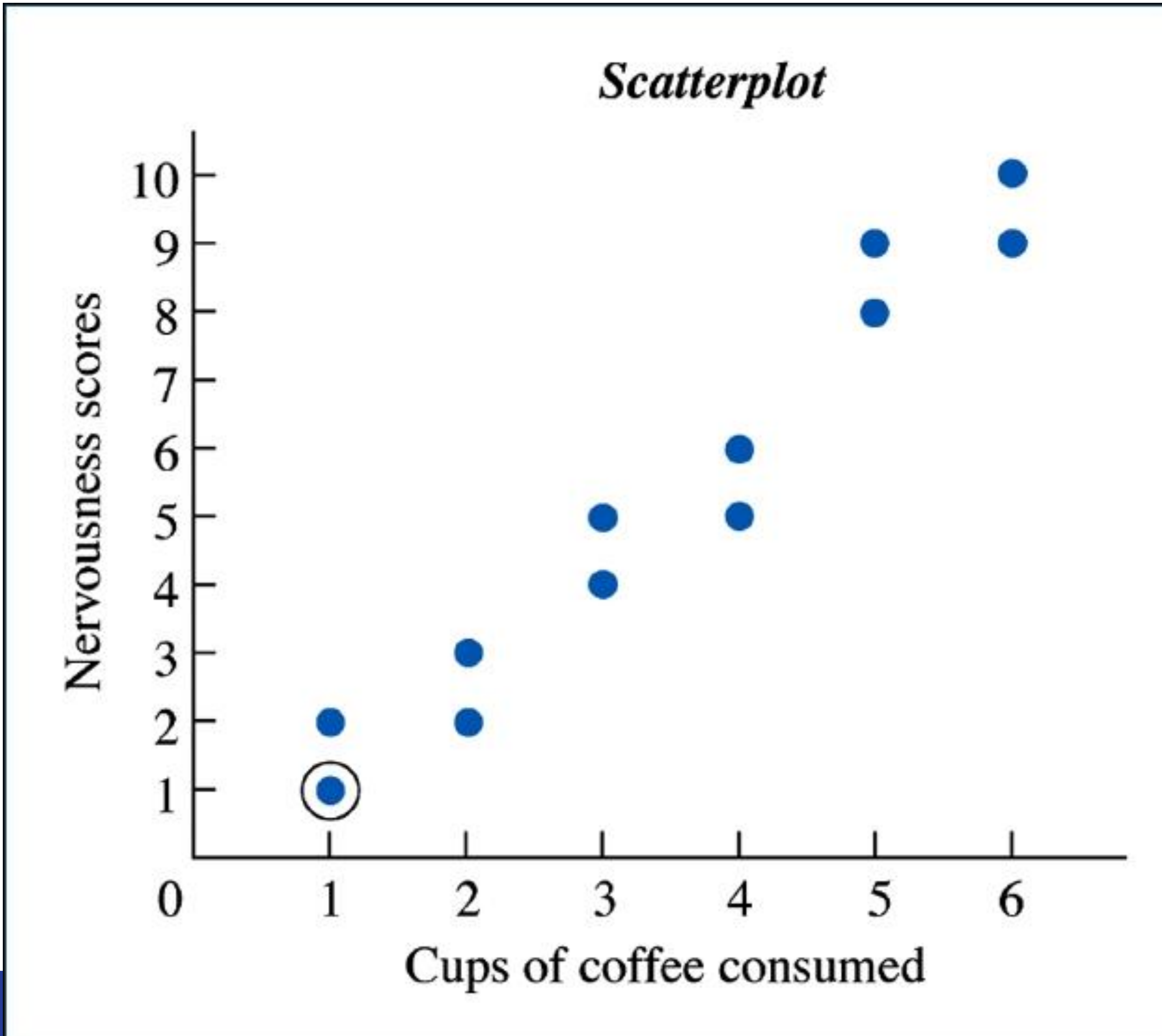
- Linear correlation
- Curvilinear correlation
- No correlation
- Positive correlation
- Negative correlation

Linear Correlation

- Describes a situation where the pattern of dots falls roughly in a straight line



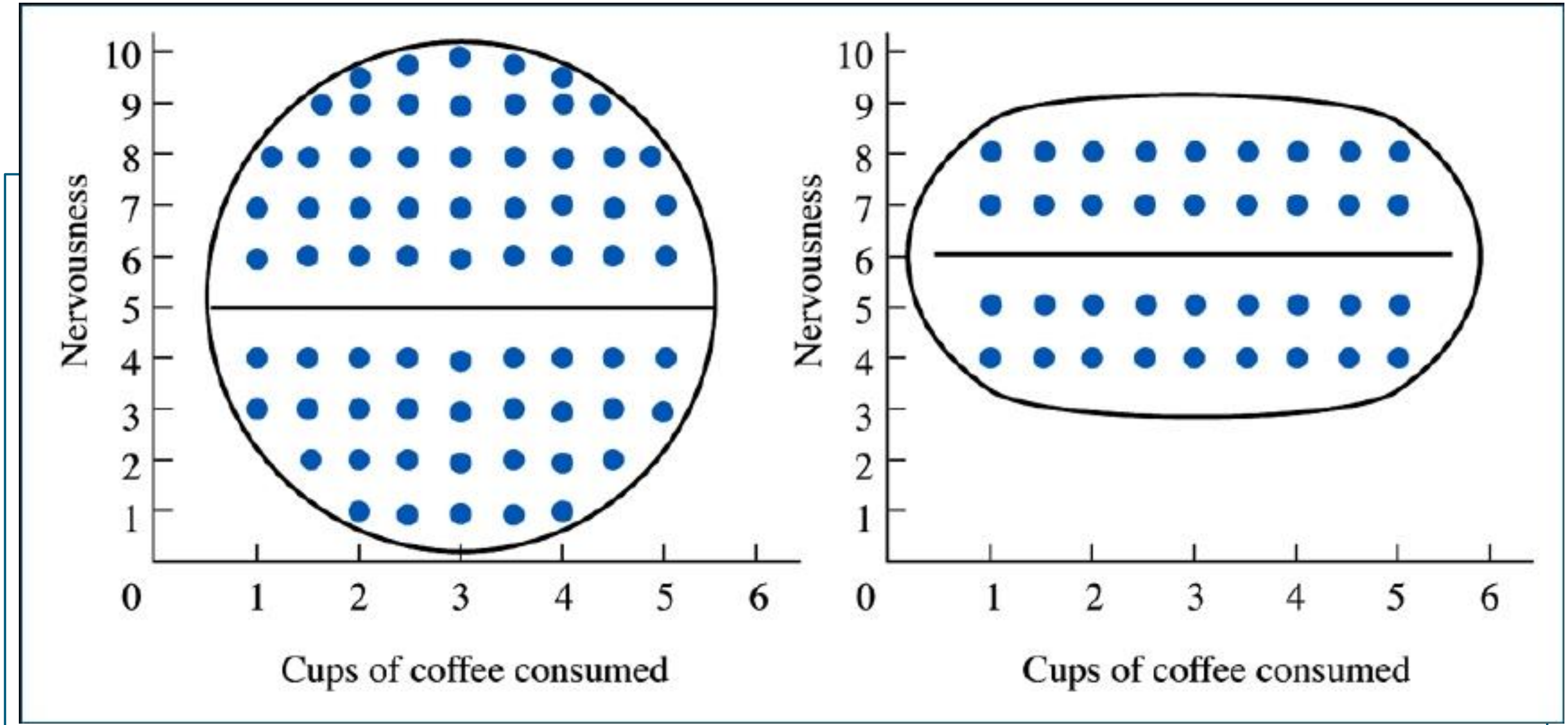
Strong Relationship



No correlation

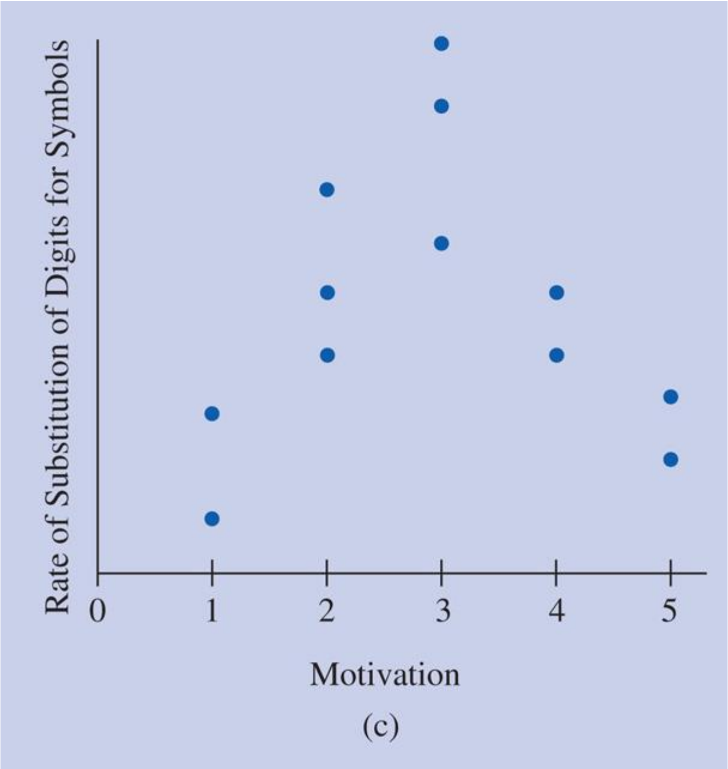
- is also possible for two variables to be essentially unrelated to each other
- Ex. income and shoe size

No Relationship

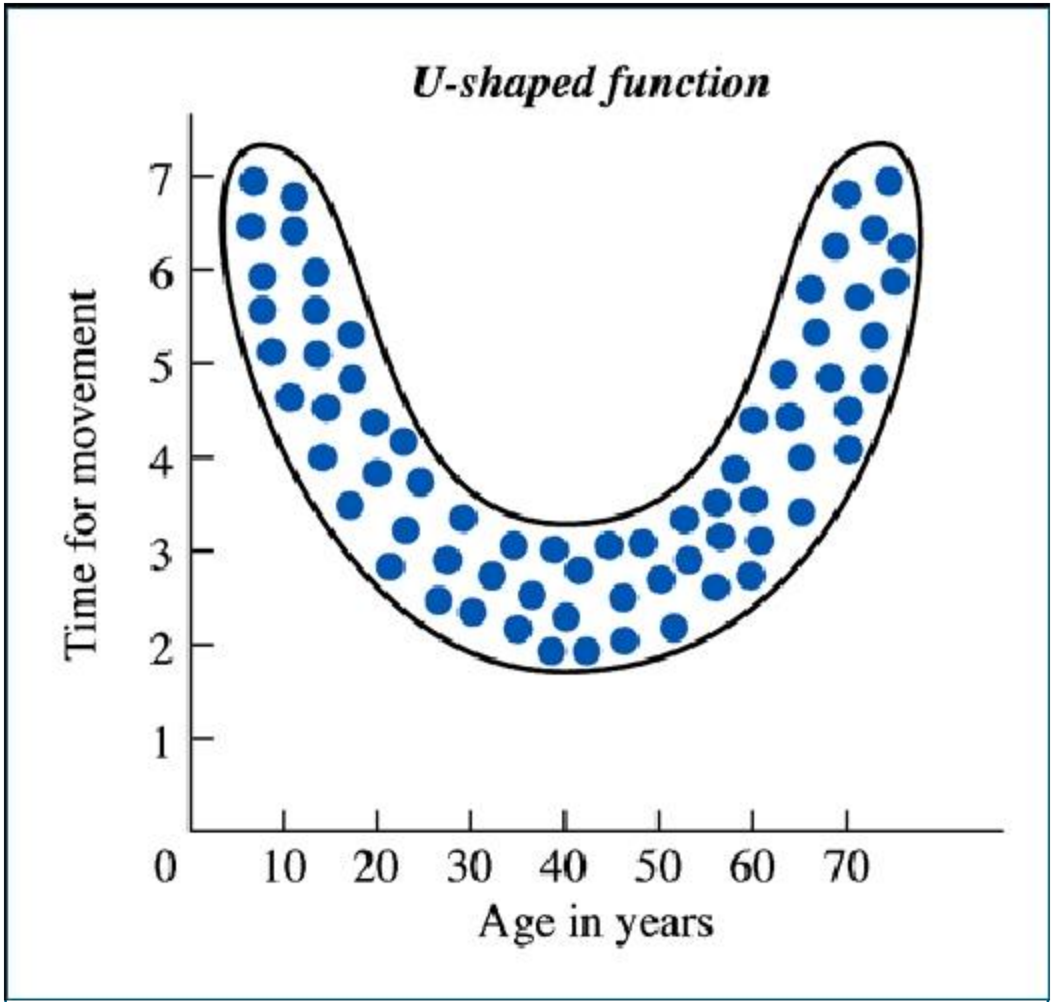


Curvilinear Correlation

- Pattern of dots is curved, not in a straight line
- U-shaped
- n-shaped



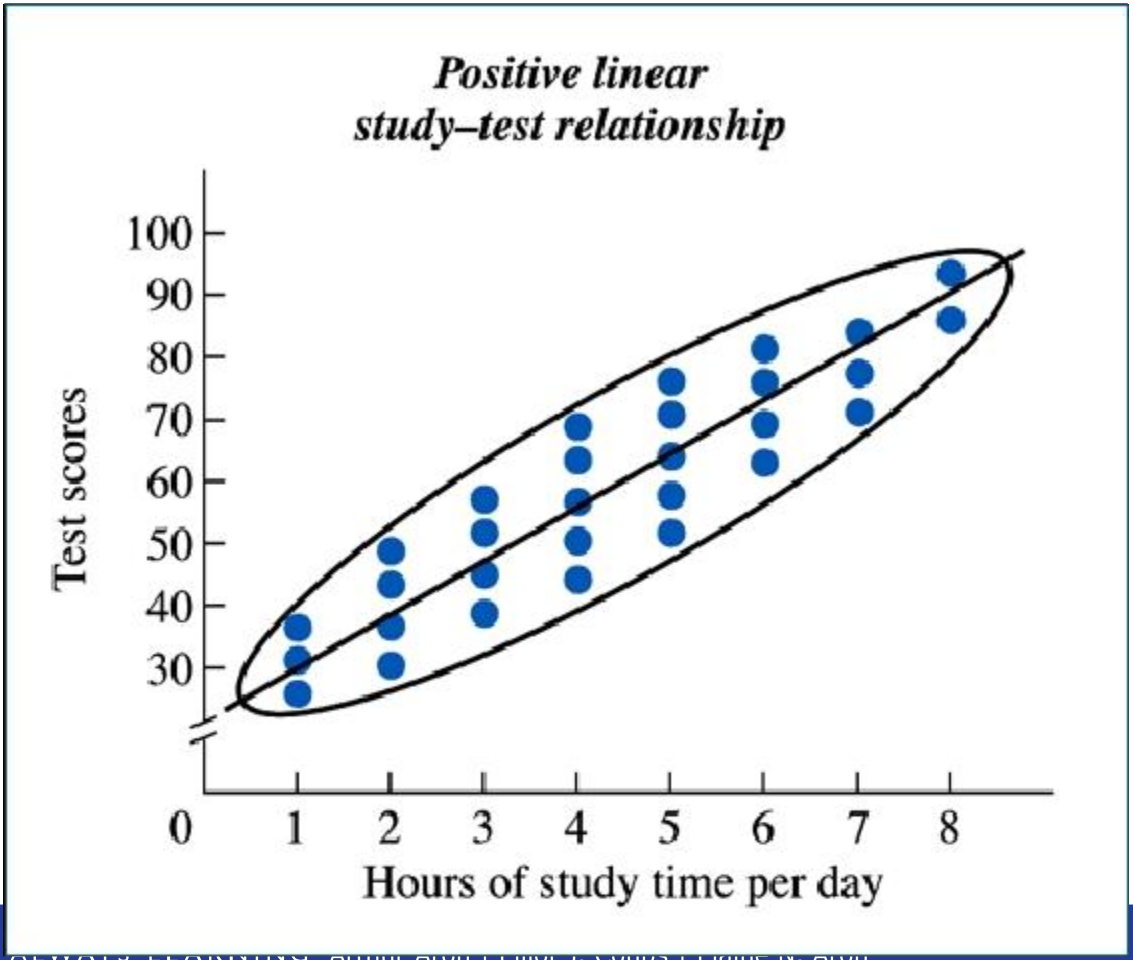
A Scatterplot of a Nonlinear Relationship



Positive Correlation

- Pattern of dots goes up from left to right
  - High scores on one variable go with high scores on the other variable
  - Low scores on one variable go with low scores on the other variable
  - Middle scores go with middle scores

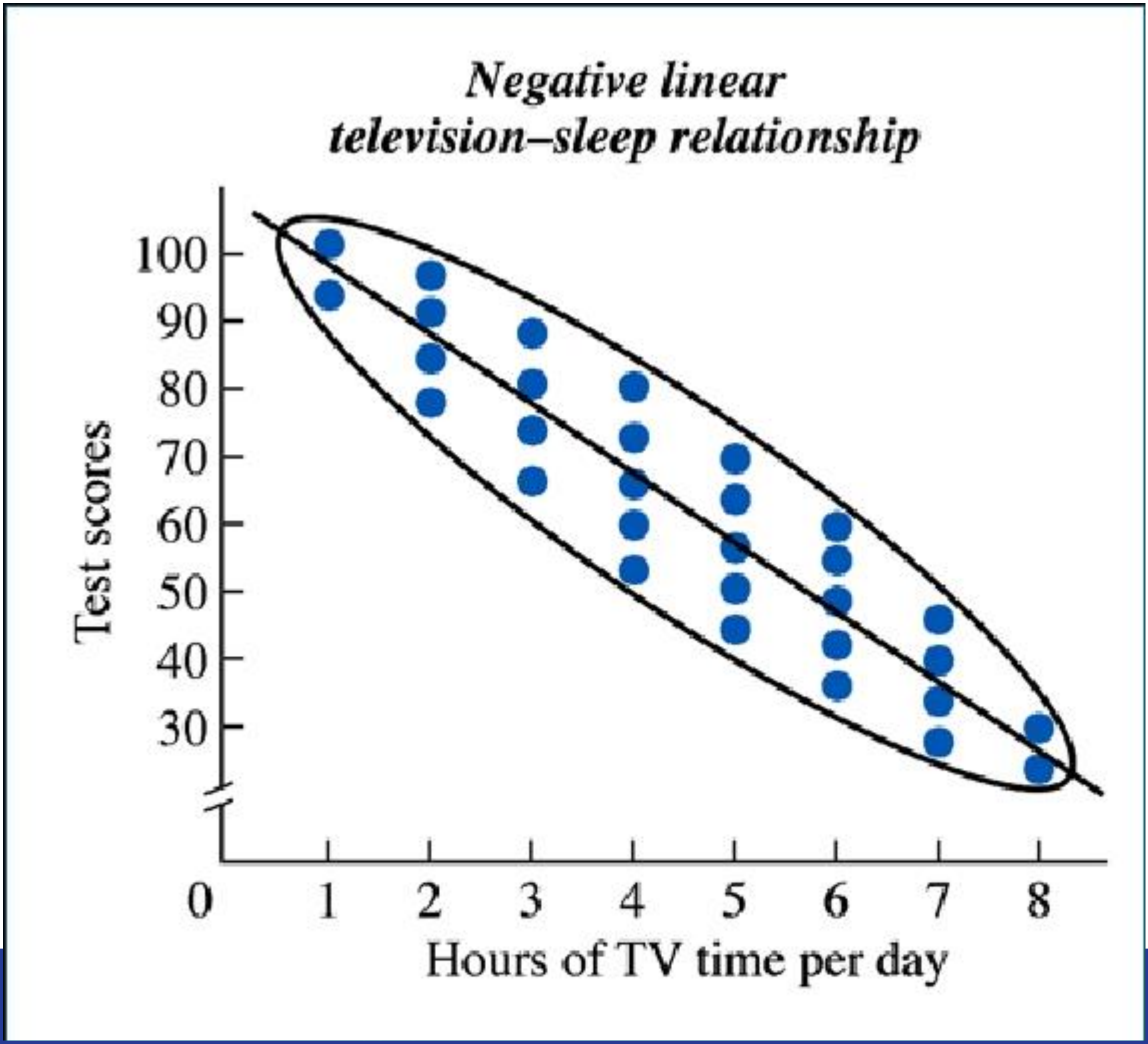
A Scatterplot of a Positive Linear Relationship



Negative Correlation

- Pattern of dots goes down from left to right
  - High scores on one variable go with low scores on the other variable
  - Low scores on one variable go with high scores on the other variable
  - Middle scores go with middle scores

A Scatterplot of a Negative Linear Relationship



Question...

- Students with higher math scores tend to obtain higher reading scores. Likewise, students with lower math scores tend to obtain lower reading scores. Is the relationship between math and reading scores direct or inverse?

Direct Inverse

Question...

- Students with high math scores tend to have higher reading scores. Likewise, students with low math scores tend to have lower reading scores. Is the correlation coefficient between students' math and reading scores negative or positive?

– +

Correlation Coefficients

- Looking at a scatter diagram gives you a rough idea of the relationship between two variables, but it is not a very precise approach.
- What you need is a number that gives the exact correlation
- Correlation coefficients describe the degree and direction of linear correlation
- Remember Z score?
- *Z scores are so useful when figuring the exact* correlation. It has to do with what happens if you multiply a score on one variable by a score on the other variable, which is called a *cross-product. When using Z scores, this is called a* ***cross-product of Z scores***

Correlation Coefficients

- Low scores give –Z scores; high scores give + Z scores
- If a low Z score in one variable goes along w/ a low Z score in another variable for the same person; the result would be positive (- x- = +)
- If a high Z score in one variable goes along w/ a high Z score in another variable for the same person; the result would be positive, again (+ x + = +)
- These all form positive correlation coefficients

Correlation Coefficients

- Low scores give –Z scores; high scores give + Z scores
- If a low Z score in one variable goes along w/ a high Z score in another variable for the same person; the result would be negative (- x + = -)
- If a high Z score in one variable goes along w/ a low Z score in another variable for the same person; the result would be negative, again (+ x - = -)
- These all form negative correlation coefficients

Correlation Coefficients

- The sign of the correlation coefficient (-/+) shows the **direction** of the linear correlation between two variables
- The actual value of the correlation coefficient shows the **strenght** of the linear correlation
  - from a low of 0 to a high of 1 (ignoring the sign)
  - Correlation coefficient of +.85 represents a larger linear correlation than a correlation of +.45
  - Likewise, correlation coefficient of -.90 represents a larger linear correlation than a correlation of +.85
    - Because .90 > .85 (ignore the sign to evaluate the strenght of the correlation)

Correlation Coefficients

Table 11-2 The Effect on the Correlation of Different Patterns of Raw Scores and Z Scores

Pair of Raw Scores		Pair of Z Scores		Cross-Product of Z Scores	
X	Y	Z_X	Z_Y	$Z_X Z_Y$	Effect on Correlation
High	High	+	+	+	Contributes to positive correlation
Low	Low	−	−	+	Contributes to positive correlation
High	Low	+	−	−	Contributes to negative correlation
Low	High	−	+	−	Contributes to negative correlation
Middle	Any	0	+, −, or 0	0	Makes correlation near 0
Any	Middle	+, −, or 0	0	0	Makes correlation near 0

Note: + indicates a positive number, − indicates a negative number.

Correlation Coefficients

- However, you are still left with the problem of figuring the *strength of a positive* or negative correlation on some standard scale.
- The larger the number (either positive or negative), the bigger the correlation.
- But how large is large, and how large is not very large?
- You can't judge the strength of the correlation from the sum of the cross-products alone, *because it gets bigger just by adding the cross-products of more people together*

Correlation Coefficients

- **The solution is to divide this sum of the cross-products by the number of people in the study**.
- You figure the *average of the cross-products of Z scores.*
- *It turns out that because of the nature of Z scores, this average can never* be more than +1, which would be a positive linear perfect correlation.
- It can never be less than -1, which would be a negative linear perfect correlation.
- In the situation of no linear correlation, the average of the cross-products of *Z scores is 0.*

Degree of Linear Correlation: The Correlation Coefficient

- Figure correlation using *Z* scores
- Cross-product of *Z* scores
  - Multiply *Z* score on one variable by *Z* score on the other variable
- Correlation coefficient
  - Average of the cross-products of *Z* scores

Degree of Linear Correlation: The Correlation Coefficient

- General formula for the correlation coefficient:

$$r = \frac{\sum Z_X Z_Y}{N}$$

- Positive perfect correlation: $r = +1$
- No correlation: $r = 0$
- Negative perfect correlation: $r = -1$

Example 1

People	Sleep hours	Happy mood
1	5	2
2	7	4
3	8	7
4	6	2
5	6	3
6	10	6

Example 1

Table 11-3 Figuring the Correlation Coefficient for the Sleep and Mood Study (Fictional Data)

Number of Hours Slept (X)				Happy Mood (Y)				Cross-Products
Deviation		Dev Squared	Z Scores	Deviation		Dev Squared	Z Scores	
X	$X - M$	$(X - M)^2$	Z_X ❶	Y	$Y - M$	$(Y - M)^2$	Z_Y ❶	$Z_X Z_Y$
5	−2	4	−1.23	2	−2	4	−1.04	❷ 1.28
7	0	0	0	4	0	0	0	0
8	1	1	.61	7	3	9	1.56	.95
6	−1	1	−.61	2	−2	4	−1.04	.63
6	−1	1	−.61	3	−1	1	−.52	.32
10	3	9	1.84	6	2	4	1.04	1.91
$\Sigma = 42$		$\Sigma(X - M)^2 = 16$		$\Sigma = 24$		$\Sigma(Y - M)^2 = 22$		❸ $\Sigma Z_X Z_Y = 5.09$
$M = 7$		$SD^2 = 16/6 = 2.67$		$M = 4$		$SD^2 = 22/6 = 3.67$		$r = 5.09/6 = .85$
		$SD = 1.63$				$SD = 1.92$		❹

Example 2

Number of exposure	Number of words recalled
1	3
2	2
3	6
4	4
5	5
6	5
7	6
8	9

Example 2

Table 11-5 Figuring the Correlation Coefficient for the Effect of Number of Exposures to Each Word on the Number of Words Recalled (Fictional Data)

Number of Exposures (X)				Number of Words Recalled (Y)				Cross-Products
Deviaition		Dev Squared	Z Scores	Deviation		Dev Squared	Z Scores	
X	$X - M$	$(X - M)^2$	Z_X ❶	Y	$Y - M$	$(Y - M)^2$	Z_Y ❶	$Z_X Z_Y$ ❷
1	−3.5	12.25	−1.53	3	−2	4	−1.00	1.53
2	−2.5	6.25	−1.09	2	−3	9	−1.50	1.64
3	−1.5	2.25	−.66	6	1	1	.50	−.33
4	−.5	.25	−.22	4	−1	1	−.50	.11
5	.5	.25	.22	5	0	0	.00	.00
6	1.5	2.25	.66	5	0	0	.00	.00
7	2.5	6.25	1.09	6	1	1	.50	.55
8	3.5	12.25	1.53	9	4	16	2.00	3.06
$\Sigma = 36$		$\Sigma(X - M)^2 = 42$		$\Sigma = 40$		$\Sigma(Y - M)^2 = 32$		❸ $\Sigma Z_X Z_Y = 6.56$
$M = 4.50$		$SD^2 = 42/8 = 5.25$		$M = 5.00$		$SD^2 = 32/8 = 4.00$		❹ $r = 6.56/8 = .82$
		$SD = 2.29$				$SD = 2.00$		

Example 3

- For the following data set of interval/ratio scores, calculate the Pearson correlation coefficient.

X	Y
1	8
2	6
3	6
4	5
5	1
6	3

Example 3

- For the following data set of interval/ratio scores, calculate the Pearson correlation coefficient.

X	Y
1	8
2	6
3	6
4	5
5	1
6	3

r = -.88

Example 4

- Find the correlation between spelling and vocabulary

	X *Spelling*	*Y* *Vocabulary*
Sandra	*8*	*10*
Neil	*5*	*6*
Laura	*4*	*7*
Jerome	*1*	*3*

Example 4

- Find the correlation between spelling and vocabulary

$r = .96$

	X Spelling	*Y* Vocabulary
Sandra	*8*	*10*
Neil	*5*	*6*
Laura	*4*	*7*
Jerome	*1*	*3*

Interpreting a Correlation

- A correlation is strong and positive if highs on one variable go with highs on the other, and lows with lows
- A correlation is strong and negative if lows go with highs, and highs with lows
- There is no correlation if sometimes highs go with highs and sometimes with lows

Significance of the Correlation Coefficient

- The correlation coefficient is a descriptive statistic, like the mean or standard deviation. The correlation coefficient describes the linear relationship between two variables.
- However, in addition to *describing* this relationship, we may also want to test whether it is statistically significant.
- In the case of a correlation, the question is usually whether it is significantly different from zero.
- That is, the null hypothesis in hypothesis testing for a correlation is usually that in the population the true relation between the two variables is no correlation (r = 0

Significance of the Correlation Coefficient

- *t* is used to determine the significance of a correlation coefficient

$$t = \frac{r}{\sqrt{(1 - r^2)/(N-2)}}$$

- with *df* = *N*-2

Assumptions for the Significance Test of a Correlation Coefficient

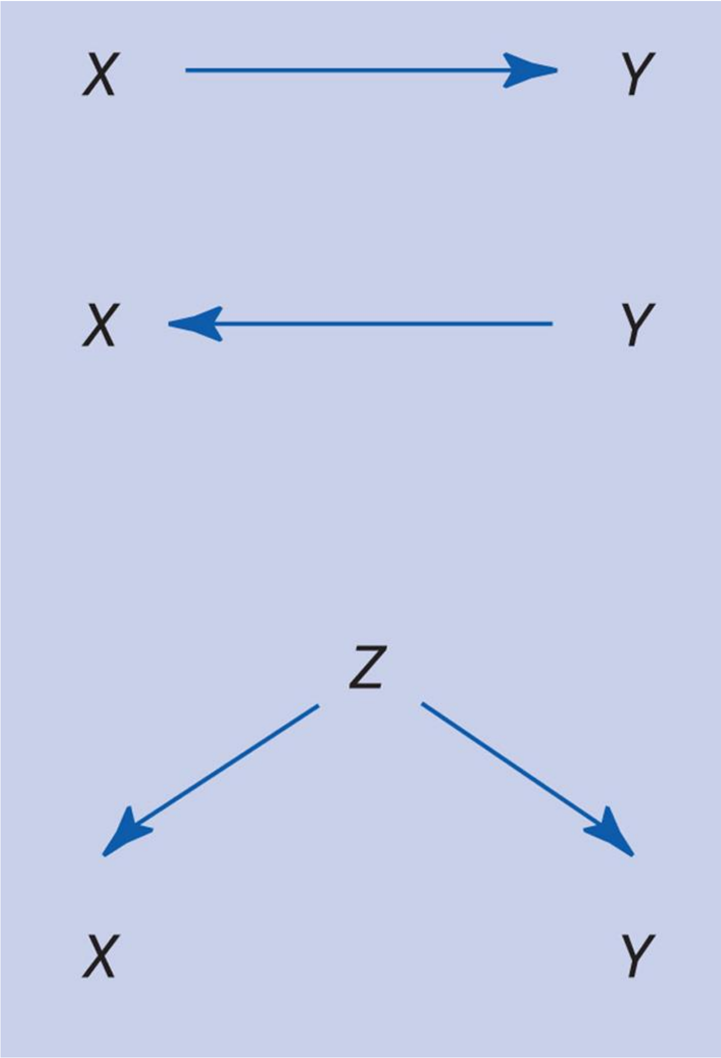
- 1. **The population of each variable (*X* and *Y*) follows a normal distribution**
- 2. **There is an equal distribution of each variable at each point of the other variable**

Correlation and Causality

- If two variables have a significant linear correlation, we normally assume that there is something causing them to go together.
- However, you can't know the **direction of causality** (what is causing what) just from the fact that the two variables are correlated.
- **Correlation does not mean causality; it just shows two variables are related with each other**

The principle is that for any correlation between variables *X* and *Y*, there are at least three possible directions of causality:

1. *X* could be causing *Y.*
2. *Y* could be causing *X.*
3. Some third factor could be causing both *X* and *Y.*

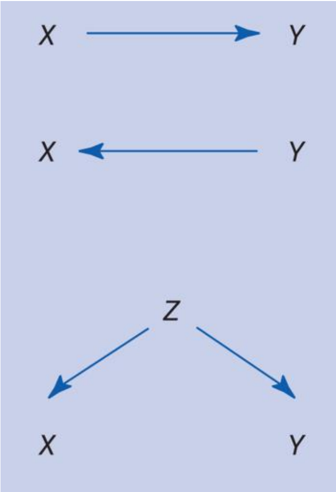


Correlation and Causality -1

Ex. the correlation between doing exciting activities with your partner and satisfaction with the relationship.

There are three possible directions of causality for these two variables:

1. It could be that doing exciting activities together causes the partners to be more satisfied with their relationship.
2. It could also be that people who are more satisfied with their relationship choose to do more exciting activities together.
3. Another possibility is that something like having less stress (versus more stress) at work makes people happier in their marriage and also gives them more time and energy to do exciting activities with their partner.



Correlation and Causality -1

Correlation and Causality

- Correlational research design
  - Correlation as a statistical procedure
  - Correlation as a research design

Issues in Interpreting Correlation Coefficients

- Statistical significance
- Proportionate reduction in error
  - r^2
  - Used to compare correlations
  - Ex. A correlation of .60 is actually four times as large as one of .30
  - ($.60^2 = .36$; $.30^2 = .09$; $.36 = .09 \times \mathbf{4}$)

Issues in Interpreting Correlation Coefficients

- Restriction in range
  - The correlation is based on people who include only a limited range of the possible values on one of the variables.
- Unreliability of measurement
- Curvilinearity: Make each variables scores rank order and calculate Spearman’s rho (advanced stats)

Correlation in Research Articles

- Scatter diagrams occasionally shown
- Correlation matrix

Table 11-9 Pearson Correlations Between Temperature and Recorded Behaviors

Variable	Temperature	Panting	Stretching	Wing venting	Yawning
Temperature	—				
Panting	.740**	—			
Stretching	.069	–.059	—		
Wing venting	.214*	.166	.182*	—	
Yawning	.333**	.365**	.175	.312*	—

Note: Yawning and stretching were measured as the number of yawns and stretches recorded for all four birds in a group during each 2-min interval; panting and wing venting were measured as the number of birds observed engaging in these behaviors during each trial.

* $p < .05$. ** $p < .01$.

Source: Gallup, A. C., Miller, M. L., & Clark, A. B. (2010). The direction and range of ambient temperature change influences yawning in budgerigars *(Melopsittacus undulatus)*. *Journal of Comparative Psychology*, *124*, 133–138. Published by the American Psychological Association. Reprinted with permission.

Table 11-9 Pearson Correlations Between Temperature and Recorded Behaviors