10.34 Other things remaining the same, higher α (e.g., .05 rather than .01) gives higher power, but at the expense of a higher Type I error rate, so we are hardly better off. Not a good strategy for increasing power.

10.35 You ask what δ, what target ES, was used in the power calculation. If your friend doesn't know, or says it used *d* obtained in the study, explain about post hoc power, and advise your friend that the high value most likely means only that they obtained a large *d*. Before being excited about high power, your friend should calculate power using a chosen value of δ that's of research interest in the context.

10.36 You could say that you intend to use estimation rather than NHST to analyze the data, and that a precision for planning analysis is the corresponding way to justify choice of *N*. You could also comment about the advantages of estimation and the disadvantages of the dichotomous decision making that a power analysis assumes.

11

Correlation

How satisfied are you with your body? Are you concerned that advertisers and the media bombard us all with images of idealized but unrealistic bodies? Are there bad consequences for some young people? Researchers investigating such questions often study the relations between variables, starting with the *correlation* between two variables. Correlation is our topic in this chapter, and Figure 11.1 is a scatterplot, a revealing picture of the correlation between Body Satisfaction (*X*) and Well-being (*Y*) for a group of 106 college students. Each point represents one student, and is positioned to reflect that person's *X* and *Y* scores.

I'll say more about the measures in a moment, but for now simply notice that there's lots of scatter, and, overall, higher *Y* scores tend to go with higher *X* scores. Our measure of that tendency is $r = .47$, where *r* is a units-free measure of correlation that can take any value from −1 through 0 to 1. The *r* of .47 here indicates a moderately strong tendency for higher values of body satisfaction to be associated with higher well-being scores.

What do you think the scatter of points might be telling us?

Do you think greater body satisfaction tends to make people feel generally good?

Can you think of any other reason for the positive correlation?

Forgive me if once again I start the chapter with this encouragement to pause, reflect, and discuss.

Correlation helps us investigate numerous intriguing issues, and often enables meta-analysis. Correlation, in particular *r*, is one of the most important and widely used effect size measures in many disciplines. In addition, *r* leads to further highly useful statistical techniques, including regression, which we'll see in Chapter 12. So there are good reasons to know about *r*. Here's the agenda:

- The scatterplot, a revealing picture of correlation
- Calculating *r*
- The confidence interval on *r*
- Correlation and possible causation
- Interpreting values of *r*
- Using *r* for meta-analysis

The Body Satisfaction variable in Figure 11.1 is the mean rating of how satisfied a person is (from 1 = *very dissatisfied* to 5 = *very satisfied*) with various aspects of their body (e.g., satisfaction with one's face, one's muscle tone, one's weight…). It's a subscale of the Multidimensional Body Self-Relations Questionnaire (MBSRQ, Cash, 2000). The Well-being variable is the mean rating of strength of agreement (from 1 = *strongly disagree* to 7 = *strongly agree*) with

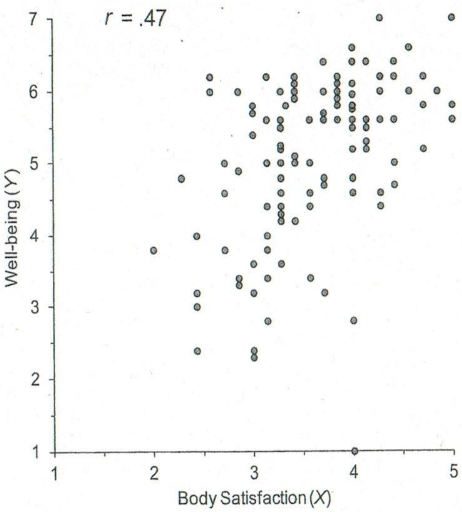


Figure 11.1. Scatterplot of Body Satisfaction (X) and Well-being (Y) scores for N = 106 college students.

a number of statements about a person's feeling of well-being (Diener, et al., 1985). For example, one item is "In most ways my life is close to ideal". In our data set, the two variables are correlated, $r = .47$, but with lots of scatter. One message of this chapter is that often there's considerable scatter, even when two variables show, overall, a fairly strong correlation. Note, for example, the one unhappy student, represented by the blue dot on the X axis, who had greater than average body satisfaction ($X = 4$), but minimum well-being ($Y = 1$).

Correlation does not necessarily imply causation.

I asked what the scatter of points might be telling us. You probably immediately thought that high body satisfaction causes improved well-being: X causes Y. However, perhaps a general feeling of well-being causes us to feel generally satisfied, in particular with our body: Perhaps Y causes X? Or each could influence the other. There could easily be other variables involved as well: For example, perhaps good health tends to increase both well-being and body satisfaction. These are all speculations about *causes* between variables, but another important message of this chapter is that correlation—the association pictured in a scatterplot—may not say anything about causes. Correlation does not necessarily imply causation. Much more on that later.

Correlation is driven by scatter. The focus is on spread within samples, not on means.

Before I say more about scatterplots, note a striking contrast with earlier chapters. So far in this book we have usually focused on the mean of a sample, and regarded the spread of points around that mean as a problem, a distraction. Here, however, we focus on that spread, rather than the means. We do this because correlation is driven by the scatter of data points away from their means. We're focusing on the *spread* of Well-being and Body Satisfaction scores; I haven't even reported means of those scores. Either spread or means can be important, depending on your perspective, but in this chapter and the next we're focusing on individuals and how they vary—we're focusing on difference, not sameness.

THE SCATTERPLOT, A REVEALING PICTURE OF CORRELATION

The key to correlation is the *scatterplot*, and another of the main messages of this chapter is that, for any (X, Y) relationship we care about, we need to see the scatterplot. Any single measure of correlation can capture only one aspect of such a relationship so, without the scatterplot, we may miss important things. There are other measures of correlation, but in this chapter we'll discuss the most widely used measure, Pearson's r, which can range from −1 through 0 to 1. Pearson's r measures the degree or strength of the *linear component* of an (X, Y) relationship, meaning the aspect of the relationship that can be represented by a straight line. We'll see many examples for which r does not tell the full story, and for which it's vital to see the scatterplot, perhaps in addition to calculating r. I'll come back to that idea shortly, but, first, let's consider some pictures to help build intuitions about r.

Figure 11.2 presents four scatterplots picturing different values of r. I'm following custom by labeling as X and Y the two variables, which I'm assuming have interval measurement. Each panel in the figure is a plot of $N = 50$ points, where each point is a pair of (X, Y) values. The first shows $r = .9$, which is close to the maximum of 1. The next two show lesser values of correlation, .6 and .3. On the right, X and Y are *uncorrelated* and so $r = 0$. A cloud of points as for $r = 0$, which roughly resembles a circular scatter, can be described as a shotgun blast. The cloud for $r = .3$ also looks rather like a shotgun blast. Do you think the scatter in Figure 11.1, for which $r = .47$, lies between the clouds for .6 and .3?

Pearson correlation, r, is a measure of the linear component of the relationship between two interval variables, X and Y. It takes values between −1 and 1, and is 0 if X and Y are uncorrelated.

Figure 11.3 presents three further scatterplots. The left plot illustrates the maximum correlation, $r = 1.0$. In the center plot, *Y decreases* as X increases, so the correlation is negative: $r = -.6$. Does the plot for −.6 look to have roughly the same amount of scatter as the plot for .6 in Figure 11.2? That's what we'd expect. Now look back at the $r = .9$ plot in Figure 11.2. Do you feel an urge to eyeball a straight line that seems to fit, or represent the points? If so, you may be thinking of something like the plot on the right in Figure 11.3, which shows the same data with $r = .9$ as in Figure 11.2, but displays also the *regression line* of Y on X. That line tells us what Y value is most likely associated with some particular X value we might be interested in.

Regression, or more particularly *linear regression*, is our topic in Chapter 12. Correlation as we discuss it in this chapter and linear regression are closely related, so it's a useful instinct to think of a straight line that may represent, at least to some extent, the relation between X and Y. As I mentioned, it's

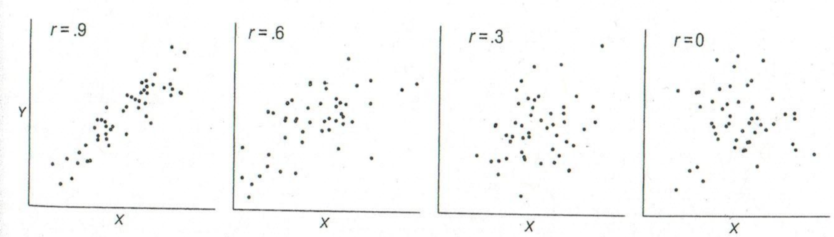


Figure 11.2. Scatterplots, each with N = 50, to illustrate the four indicated values of r, the Pearson correlation between X and Y.

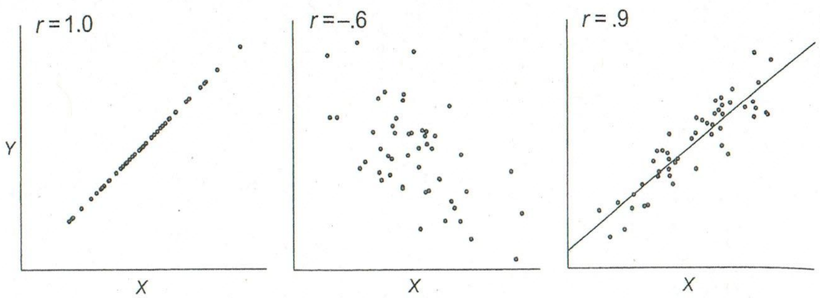


Figure 11.3. Three further scatterplots. On the left, $r = 1.0$, its maximum value. In the center the correlation is negative. The plot on the right is the same as that for $r = .9$ in Figure 11.2, but with the regression line of Y on X displayed.

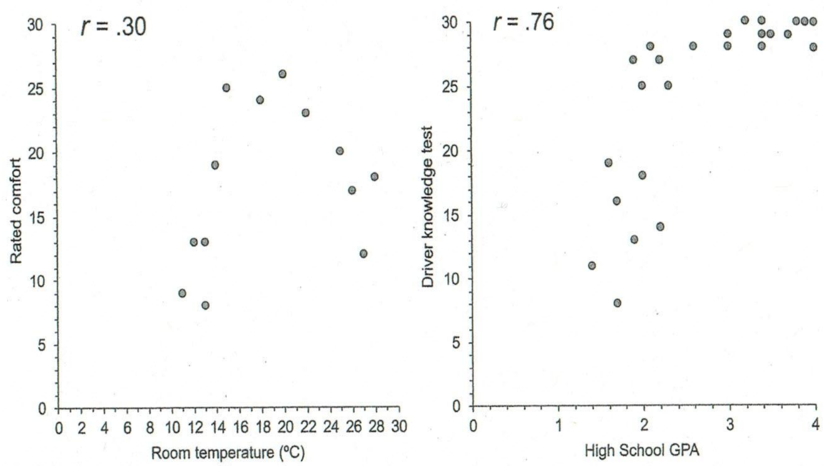


Figure 11.4. Scatterplots for two fictional data sets. On the left is an illustration of how rated room comfort (Y) might vary with room temperature (X). On the right is an illustration of how the score on an easy test—a driver knowledge test—(Y) might vary with High School GPA (X). Correlation values are shown, but in neither case does r represent the relationship well.

the strength of such a linear component in a relationship that's measured by Pearson's r. Incidentally, you may feel that the line for $r = .9$ in the figure should be a little steeper to fit the points well. That's a good observation, which we'll discuss in Chapter 12. For this chapter, just think informally of an eyeballed line that seems, approximately, to fit the points in a scatterplot.

A scatterplot may reveal that an (X, Y) relationship is *curvilinear*, or shows a *ceiling effect*.

Figure 11.4 shows two fictional but plausible cases in which there looks to be a relation between X and Y that's *nonlinear*—that can't be well represented by any straight line. On the left, there looks to be a *curvilinear* relation: Rated room comfort is greatest at a medium room temperature and decreases at both lower and higher temperatures. On the right, score on a driver knowledge test, required for obtaining a driver license, at first increases with High School GPA, then tops out at or near 30, the maximum possible score on the test. This pattern is typical of an easy test, and the topping out is called a *ceiling effect*: On a much harder test of driver knowledge, students with a high GPA may well do better, but the easy test denies them the opportunity. On an even harder

above 3 would scores rise: That would be an example of a *floor effect*, because scores can't go below zero.

For a full picture of an (X, Y) relationship, inspect and think about the scatterplot. The value of r may be informative, irrelevant, or misleading.

Correlation r can be calculated for any scatterplot of X and Y values, but it doesn't always make sense. On the left in Figure 11.4, $r = .30$ tells us little or nothing useful about the relationship, and may even mislead. Compare it with the $r = .30$ example in Figure 11.2. On the right, $r = .76$ does reflect the fact that, in general, Y increases as X increases, but it misses the topping out—the ceiling effect—which is probably one of the most important things to note. Yes, it's vital to see the scatterplot and to think carefully about what it tells us. The value of r may be highly informative, or irrelevant, or even misleading.

Estimating *r* From a Scatterplot

For high correlation, degree of tightness to the imagined line indicates r, as in Figure 11.3.

Leaving aside examples like those in Figure 11.4, it's often not difficult to make a rough estimate of r from a scatterplot. On the right in Figure 11.3, when $r = .9$ the points are quite tightly bunched to the line. In the center, with $r = -.6$ they are less tight to the imagined line, whereas on the left, for $r = 1.0$, they are on the line. For high correlations, tighter to the line means higher $|r|$, where $|r|$ is the absolute value of r, dropping the minus sign, if any. More specifically, for $|r|$ greater than around .7, degree of *tightness* to the imagined line helps when eyeballing r. Note that it's tightness that indicates r, not the slope of the line, which is easily changed just by changing the scale on the X or Y axis. Keep the Figure 11.3 examples in mind.

A second eyeballing strategy is to use a cross through the means. In Figure 11.5, the second and fourth panels illustrate this idea. The vertical line is at the mean of the N values of X, and the horizontal line at the mean of the N values of Y. The second scatterplot labels the four quadrants defined by the cross: HH is high-high, meaning both X and Y are relatively high; HL is high-low, meaning X is high, Y is low, and so on. Now eyeball the numbers of points in the HH and LL quadrants, which I call the *matched* quadrants, and compare with the numbers in the LH and HL quadrants, the *mismatched* quadrants. The larger the proportion of matched quadrant points, compared with mismatched quadrant points, the larger the r. More mismatched than matched points indicates that r is negative. Actually, it's not just numbers of points that matter, but how far they are from the cross lines—we'll see that points far from the means have larger influence. However, looking at the whole pattern and noting relative numbers of points in the quadrants is often sufficient. Examine the first and third scatterplots: Eyeball the cross and estimate the r values. Look back to Figure 11.2 if that helps.

To eyeball r, consider the relative numbers of points in the quadrants defined by a cross through the X and Y means. More in HH, LL (matched quadrants) indicates r closer to 1; more in HL, LH (mismatched), r closer to -1.

They are $r_1 = .6$ and $r_2 = .3$. Examine the second and fourth panels, which show the same scatters of points, with cross. Do the quadrants help? For high correlations I find that thinking of tightness to the line helps. For any correlation, I find that imagining the cross and thinking in terms of relative numbers of points in the quadrants helps.

Figure 11.6 shows the two examples of Figure 11.4, but with crosses through the means. Check that considering the relative numbers of points in the pairs of quadrants does lead to a rough estimate of r. Do the crosses give extra insight into the relationships? Perhaps on the right the cross makes the topping out more salient but, in general, the crosses strategy helps the eyeballing of r rather than the insightful reading of a scatterplot. One last comment: The aim of eyeballing is not numerical accuracy, but to get a rough idea of r to help

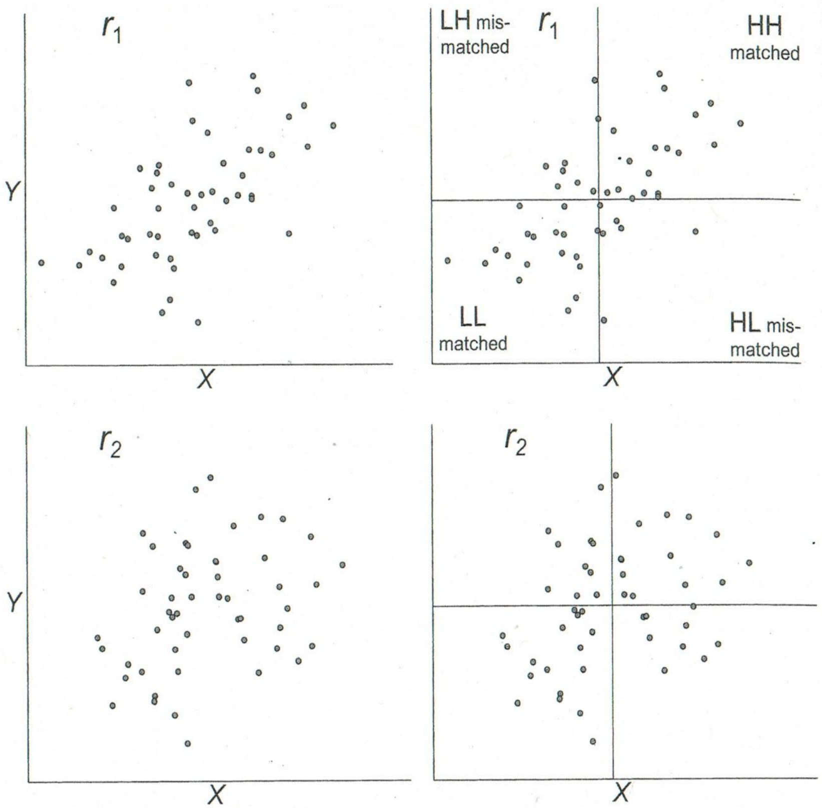


Figure 11.5. Two pairs of scatterplots, each with N = 50. The second of each pair is the same as the first, but shows a cross through the means of X and Y. Eyeball r for the first and third panels, then note how the cross can help. HH is high-high, LH is low-high, and so on. HH and LL are the matched quadrants, the other two the mismatched.

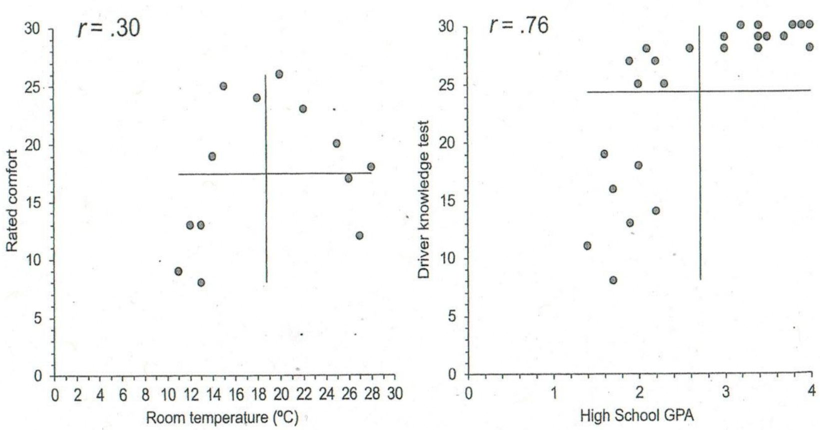


Figure 11.6. Same as Figure 11.4, but with a cross through the means of X and Y in each panel.

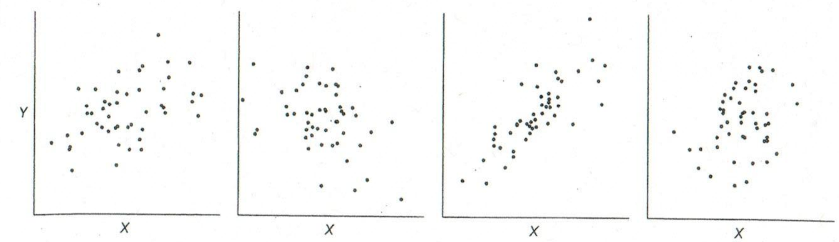


Figure 11.7. Four scatterplots, for eyeballing practice.

11.1 Eyeball r for each panel in Figure 11.7. In each case, compare the usefulness of tightness to the line and of quadrants.

esci **11.2** Open the **See r** page of **ESCI intro chapters 10–16**. At red 1, choose your own N. At red 2, click the radio button on, and use the slider to set r, the correlation you wish to see. ESCI generates a data set with your chosen N and r, and displays its scatterplot.

a. Choose, for example, $N = 30$ and $r = .3$. At red 5, click the first checkbox. Your screen should resemble Figure 11.8.

b. Vary N and r, and see a new data set when you change a value. Explore. What does $r = .1$ look like? What about $r = -1$? $r = .4$? $r = -.4$?

esci **11.3** At red 5 click to turn on the second checkbox, to see a cross through the means. Eyeball the relative numbers of points in the pairs of quadrants, for various values of r.

a. Click to turn off the first checkbox. Practice using the cross to help with eyeballing r.

b. Try negative as well as positive values. Try larger values of N.

esci **11.4** a. Turn off the cross, then eyeball the cross and estimate r. Rough is OK.

b. Make it a game: one person sets r, the other eyeballs r from the scatterplot. Try various values of r, and of N. Display the cross, then turn it off.

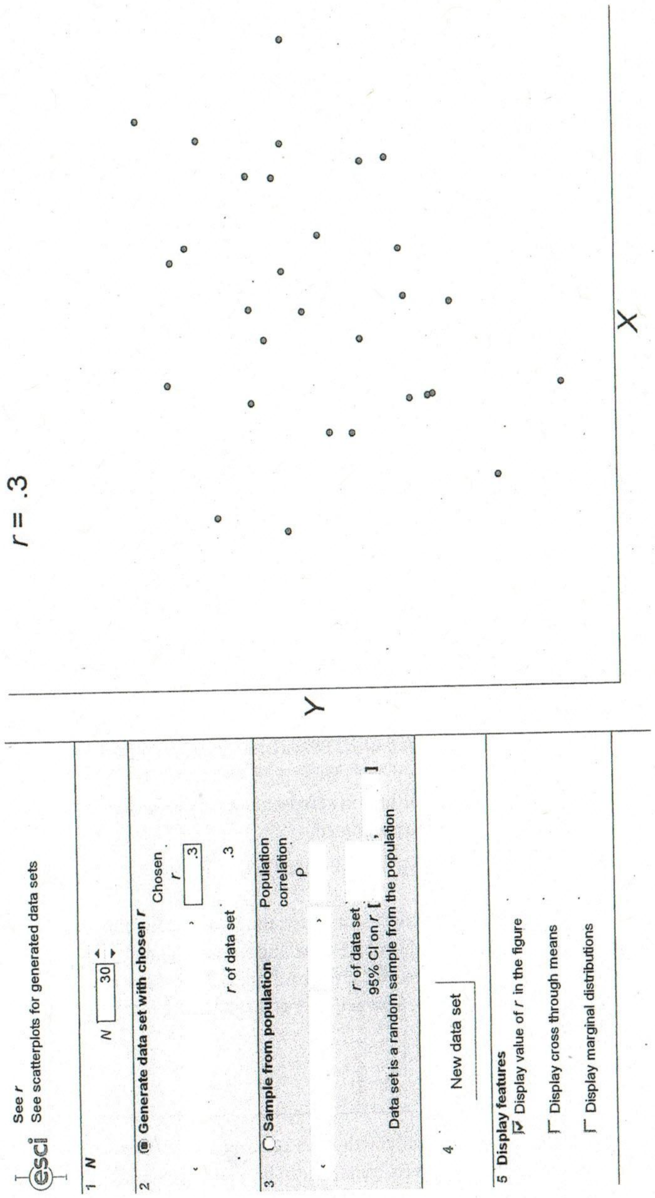
c. Keeping N and r the same, click the button at red 4 to see new examples. Note the variation in scatterplot appearance.

11.5 Eyeball r for each of the four scatterplots in Figure 11.9.

All four scatterplots in Figure 11.9 have $r = .4$, illustrating how scatterplot appearance can vary, even for a single value of r. Eyeballing is challenging, but as I mentioned we're aiming for intuitions, not precise numbers. Refer back to Figures 11.2, 11.3, and 11.5 if that helps. Now let's use an example we met in Chapter 8 to calculate r.

CALCULATING r

Thomason 1 is a small critical thinking study with $N = 12$ and a paired pretest–posttest design. It's a great example of how correlation can be vitally important. Back in Chapter 8, Figure 8.4 showed means and CIs for pretest and posttest LSAT logical reasoning scores, and the paired differences plotted on a difference axis. The CI on the mean difference was short, compared with the other CIs, which was great news that indicated a sensitive design. As I mentioned, it was



gure 11.8. A data set generated to have our chosen $N = 30$ and $r = .3$. From **See r**.

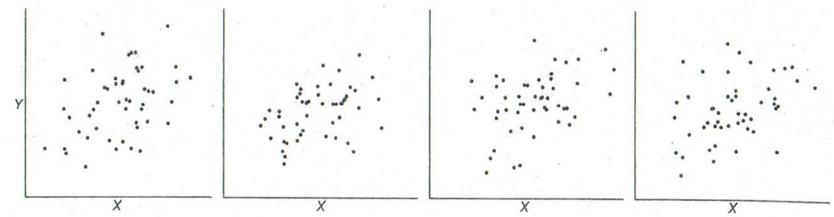


Figure 11.9. Four scatterplots, each with $N = 50$, for eyeballing practice.

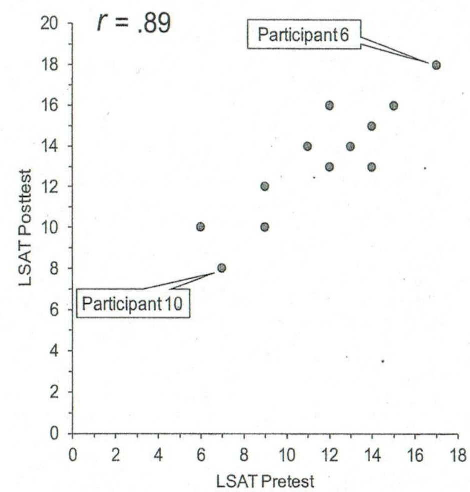


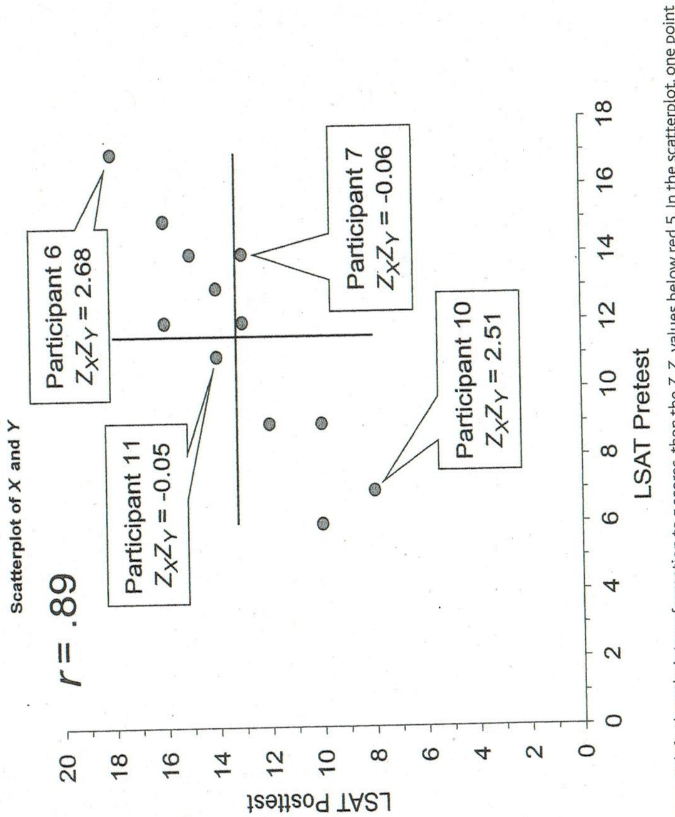
Figure 11.10. Scatterplot of the pretest and posttest data, from the paired design study Thomason 1, with $N = 12$. The data points for two of the participants are identified. From **Scatterplots**.

the high correlation of pretest and posttest that gave the short CI. In practice, the correlation of pretest and posttest scores is often high—at least .6 and sometimes up to .9—in which case we have a sensitive study, likely to give a relatively precise estimate. Let's calculate r for the Thomason 1 data.

If the two measures (e.g., pretest and posttest) in a paired design are highly correlated, the estimate of the mean difference is likely to be precise.

Figure 11.10 is a different picture of the same data: A simple scatterplot of posttest against pretest. Participant 6, for example, scored (17, 18) on the (pretest, posttest), so did well on both, and Participant 10 scored (7, 8), two low scores. Students high on pretest were generally high on posttest, and those low on pretest were also low on posttest, so the correlation is strong and positive: $r = .89$.

The value of r results from the *battle of the quadrants*: The matched team (HH, LL) pushes r up toward 1, whereas the mismatched team pushes it down toward −1. The more points in a team's quadrants, and the further the points are from the means of X and Y, the stronger the team. The battle is a quantitative version of our quadrant way to eyeball r. I'll use Figure 11.11 to explain. First, note the pretest (X) and posttest (Y) scores at left, and the corresponding points in the scatterplot. The data and points for four participants are highlighted. Next, see that the X and Y values have been transformed to z scores, which



	X LSAT Pretest	Y LSAT Posttest	Z_X	Z_Y	Z_XZ_Y
Mean	11.58333	13.25	0.000	0.000	
SD	3.315483	2.895922	1.000	1.000	
1	13	14	0.427	0.259	0.1107
2	12	13	0.126	-0.086	-0.0108
3	12	16	0.126	0.95	0.1193
4	9	12	-0.779	-0.432	0.3363
5	14	15	0.729	0.604	0.4405
6	17	18	1.634	1.64	2.6797
7	14	13	0.729	-0.086	-0.0629
8	9	10	-0.779	-1.122	0.8744
9	6	10	-1.684	-1.122	1.8899
10	7	8	-1.382	-1.813	2.5062
11	11	14	-0.176	0.259	-0.0456
12	15	16	1.031	0.95	0.9786
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					

Figure 11.11. The same data as in Figure 11.10. The pretest and posttest scores are shown at left, then their transformation to z scores, then the Z_XZ_Y values below red 5. In the scatterplot, one point in each of the four quadrants is identified, with its value of Z_XZ_Y and the data for those four participants are highlighted at left with heavy black lines. From **Scatterplots**.

it gives a simpler formula for *r*, and links more intuitively with the battle of the quadrants. The calculation uses Equation 3.4 for *z* scores from Chapter 3, which for our *X* and *Y* gives

$$Z_X = \frac{X - M_X}{s_X} \text{ and } Z_Y = \frac{Y - M_Y}{s_Y} \qquad (11.1)$$

z scores for *X* and *Y*.

The values of M_X and s_X, the mean and SD of *X*, are shown at the top left in the figure, similarly for *Y*. For Participant 6, for example, we get

$$Z_X = \frac{(17 - 11.58)}{3.32} = 1.63$$

which is the value reported in the Z_X column for this participant. It tells us that the point for Participant 6 is around 1.6 standard deviations, or around 5 LSAT points, to the right of the vertical line of the cross in the scatterplot, which looks about right. Similarly, $Z_Y = 1.64$, and the point is indeed considerably above the horizontal line of the cross.

Correlation *r* is calculated from the Z_XZ_Y values, which are positive in the matched (HH, LL), and negative in the mismatched quadrants.

The next step is, for each participant, to multiply the two *z* scores to get Z_XZ_Y, which is that participant's contribution to *r*. The values are shown in the column below red 5 and in the scatterplot for four of the points. For the matched (HH, LL) quadrants, Z_XZ_Y is positive, and for the mismatched quadrants, negative. To find *r*, we add all the Z_XZ_Y values and divide by the number of degrees of freedom, $(N - 1)$:

$$r = \frac{\sum Z_X Z_Y}{(N-1)} \qquad (11.2)$$

Pearson correlation, *r*.

Adding the Z_XZ_Y values is how the battle plays out, with the matched team contributing positive values and the mismatched team negative values. Incidentally, there's a parallel between why $(N - 1)$ appears in Equation 11.2 for *r*, and Equation 3.3 for standard deviation. (Ignore this remark if makes your eyes glaze over—you don't need to know—but 1 is subtracted because in both cases the *N* quantities added in the numerator are based on deviations from a sample mean, and calculating that mean uses one of the available *N* degrees of freedom, leaving $df = (N - 1)$ for the SD or correlation.)

11.6 Examine the Z_XZ_Y values below red 5 in Figure 11.11.

a. Identify the two that make the largest contribution to *r*.

b. Find the corresponding points in the scatterplot. What's special about them?

esci **11.7** Open **Scatterplots**. If you don't see the data in Figure 11.11, scroll right. Click the red 16 button to transfer the Thomason 1 data back left, replacing any data there. Click at red 4 to show *z* scores, and near red 8 to choose which display features you wish to see. Your screen should resemble Figure 11.11.

esci **11.8** Type in $X = 11$ and $Y = 11$ as Participant 13. (Use **Undo** or delete to remove them.)

a. Note the new Z_XZ_Y value, and where the new point falls. Does *r* change much? Explain.

b. Try different *X* and *Y* values for Participant 13. When does *r* change most? Explain.

The Influence of Outliers

> A point far from the mean of X and mean of Y has large (in absolute value) Z_X and Z_Y and therefore large Z_XZ_Y and large influence on r.

Figure 11.12 shows the difference made by the extreme additional outlier point (3, 30), for an imagined Participant 13 who supposedly goes from tiny pretest to huge posttest scores. In practice we would have serious doubts about such values—perhaps a data-recording error, or a student who for some reason made no effort at pretest? However, the example shows that a single outlier point can dramatically change r, even swinging it from large and positive, to negative. Focus on which quadrant a point lies in, and how far it is from the means of X and Y, to understand its Z_XZ_Y and, therefore, its influence on r. A point that's far from both cross lines has large (in absolute value) Z_X and Z_Y and therefore especially large Z_XZ_Y. Its contribution to r can be overwhelming, as in Figure 11.12 in which a single data point changed r from .89 to −.15.

Looking for outliers is yet another reason we need to see the scatterplot. You may be thinking, however, that journal articles rarely include scatterplots, so how can we be sure that the correlation analyses we read are appropriate? That's a good question. Unless the authors tell us that the scatterplots of their data show no outliers and no signs of departure from linearity, they are asking us to take their analyses and interpretation on trust. One advantage of open data, as we discussed in Chapter 10 as part of Open Science, is that even if the scatterplot is not shown in the report, anyone can go to the data and create the scatterplot. So, if you don't see a scatterplot for any (X, Y) relationship you care about, think carefully about what it might show, and perhaps try creating it yourself.

11.9 In the column of Z_XZ_Y values in Figure 11.12, find the value for the outlier point and compare with the other Z_XZ_Y values. Explain why this point is so influential on r.

esci 11.10 You can click at red 15 to reveal a second scatterplot, which is a graph of Z_Y against Z_X. It displays the same set of points, but on axes of Z_X and Z_Y rather than X and Y. It may help clarify how any point has Z_X and Z_Y values, which multiply to give Z_XZ_Y.

Quiz 11.1

1. Pearson's r indicates the degree of linear / nonlinear / quasi-linear / oblique relationship between two variables.
2. A negative correlation means that as one variable increases, the other ________.
3. If two variables are perfectly positively correlated, r will be ____; if the variables are perfectly negatively correlated, r will be ____; if the variables are not at all correlated, r will be ____.
4. When there is a strong positive relationship between two variables, a scatterplot will have most of its dots in the matched / mismatched / adjacent quadrants.
5. To calculate r, first transform the X and Y scores to ________.
6. When an X and a Y score are matched (both above the mean or both below the mean), then Z_XZ_Y will be positive / negative and will tend to increase / decrease the value of r. When an X and a Y score are mismatched (one above the mean, the other below the mean), then Z_XZ_Y will be positive / negative and will tend to increase / decrease the value of r.

SCATTERPLOTS THAT REVEAL

In Figure 11.4 the scatterplot of room comfort and room temperature suggested a curvilinear relation, and that of driver knowledge and High School GPA a

	X	Y			
Mean	10.92308	14.53846	Mean	0.000	0.000
SD	3.967819	5.410105	SD	1.000	1.000
2	LSAT Pretest	LSAT Posttest	Z_X	Z_Y	5 Z_XZ_Y
1	13	14	0.523	-0.1	-0.0521
2	12	13	0.271	-0.284	-0.0772
3	12	16	0.271	0.27	0.0733
4	9	12	-0.485	-0.469	0.2274
5	14	15	0.775	0.085	0.0662
6	17	18	1.532	0.64	0.9799
7	14	13	0.775	-0.284	-0.2205
8	9	10	-0.485	-0.839	0.4066
9	6	10	-1.241	-0.839	1.0408
10	7	8	-0.989	-1.209	1.1949
11	11	14	0.019	-0.1	-0.0019
12	15	16	1.027	0.27	0.2776
13	3	30	-1.997	2.858	-5.7068
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					

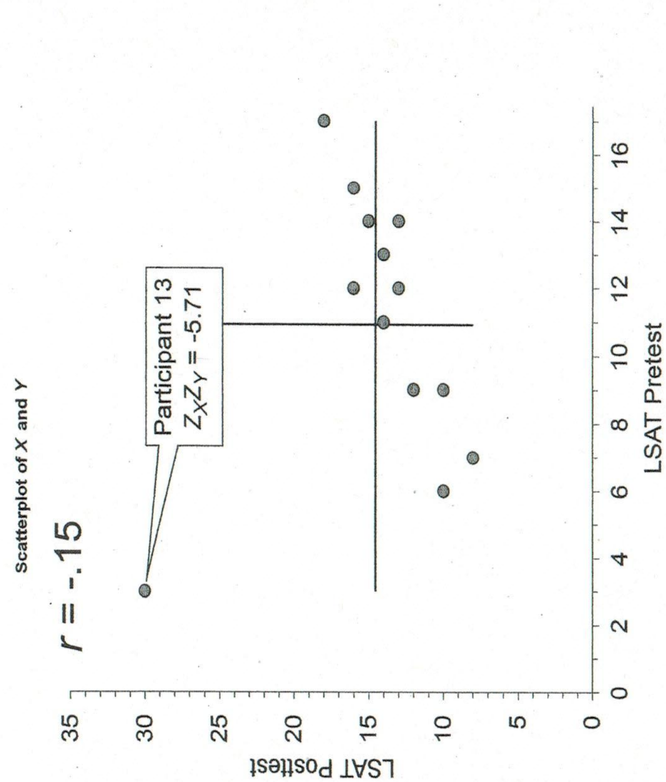


Figure 11.12. Same as Figure 11.11 except Participant 13 has been added, as an extreme outlier. Its value of Z_XZ_Y outweighs all the points in the matched (HH, LL) quadrants to give negative r.

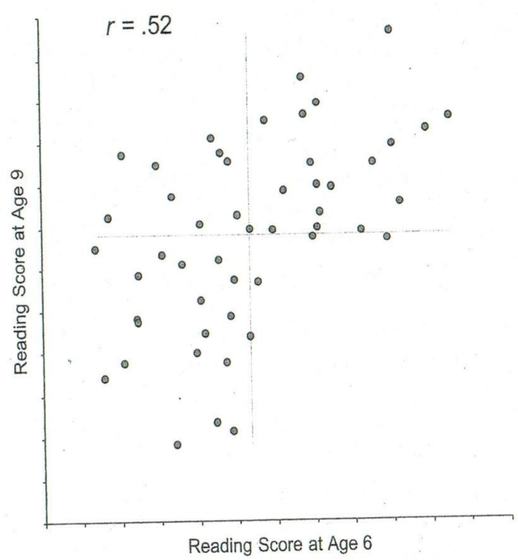


Figure 11.13. Data for N = 50 children showing a correlation of r = .52 between reading scores at ages 6 and 9. The lines mark the means. The bottom right quadrant is almost empty.

ceiling effect. These examples emphasize that we need to look carefully at the scatterplot because r measures only the degree of *linearity* in a relationship, and therefore often doesn't tell the whole story.

I learned that lesson many years ago, when some colleagues and I were studying children learning to read. We had reading scores for children at age 6, then again at 9. The data haven't survived, but Figure 11.13 is my reconstruction of the pattern. The correlation was about .5, but the scatterplot showed that virtually all the points fell in three quadrants, with the lower right quadrant virtually empty. A child scoring below the mean at age 6 might score above or below the mean at age 9—in the upper or lower left quadrant in the figure. However, a child scoring above the mean at age 6 was virtually guaranteed of scoring above the mean at age 9—in the HH, upper right quadrant. That's good news: If a child "gets it" (scores above the mean) by age 6, they will continue to progress. They can read. But if by age 6 they still don't quite get it (they score below the mean), then they may or may not get it by age 9. Only a few children slip considerably backward.

Examine the scatterplot for any (X, Y) relationship you care about.

Our r of about .5 was unsurprising, but didn't tell the whole story. Our main conclusion and guide for follow-up research was given not by r but by the pattern in the scatterplot. It's the same lesson we've been learning all along: Always plot your data in ways that are revealing, and think carefully about what the pictures might be telling you. Don't be blinded by calculations of r, p, CIs, or anything else, but use these *along with* pictures to seek the messages within your data. In particular, if you are reading a report of research that highlights one or more correlations, then you need to see the scatterplots to have the full story.

The Effect of Range Restriction on Correlation

Range restriction is another aspect of correlation best understood by consider-

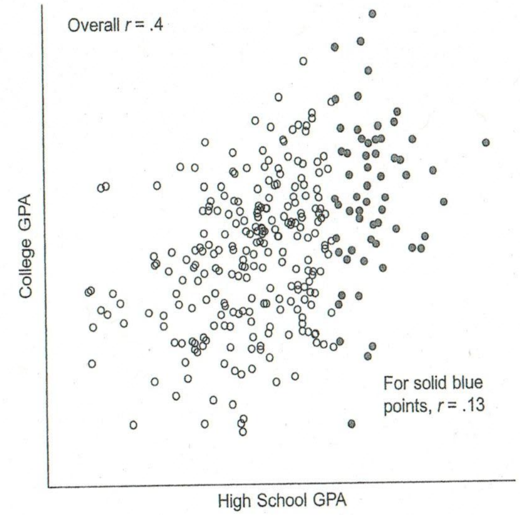


Figure 11.14. Randomly generated scatterplot with N = 300 and r = .40, which might represent college and high school GPA scores for 300 college students. Solid blue dots show students having the top 20% of high school GPA scores. For these 60 students, r = .13, the lower correlation reflecting range restriction of high school GPA.

investigated in a large meta-analysis by Richardson, Abraham, and Bond (2012), and found to average r = .40. Figure 11.14 shows what such a correlation might look like, for 300 college students. Suppose all 300 applied to enter a selective college, but the college accepted only students with a high school GPA in the top 20% of applicants—the solid blue points in Figure 11.14. For them, the correlation of high school and college GPA is r = .13. Inspecting those solid blue dots, does that look about right to you? The lower correlation reflects the restriction of the range of X, high school GPA.

If the range of possible X or Y values is restricted, we have ***range restriction*** and r is likely to be reduced.

In general, r can be strongly influenced by the ranges of possible X and Y values, especially for medium-sized r values, meaning values not close to 0, −1, or 1. If either range is restricted, the correlation is likely to be smaller. The more a range is restricted, the smaller the correlation is likely to be. The key lesson is that selection in either X or Y that restricts the X or Y range can yield a sample that badly underestimates the correlation in the whole population.

The X and Y axes should show the full ranges of possible X and Y values.

One problem is that graphing software often zooms the ranges on the X and Y axes to fit the data points. Always examine the range of values shown on each axis: If this doesn't cover the full range of possible values, perhaps the sample is range restricted and therefore r is misleadingly low? For example, a plot that showed high school GPA only from 3.0 to 4.0 would be suspect, and would need further investigation.

11.11 Suppose an elite college accepts only students with high school GPA in the top 10%. About what r would you expect for accepted students? Explain.

11.12 A survey found almost no correlation between IQ score and income, among college graduates. It concludes that a person's IQ hardly matters in the modern world. Is the conclusion justified? Explain.

11.13 In Figure 11.4, suppose you had data only for room temperature up to and including 18 °C.

a. Eyeball *r*.
b. Do the same if you had data only for 18 °C and above.
c. What can you conclude about range restriction?

What to Look for In a Scatterplot

I mentioned at the start that *r* is one of the most widely used and important effect size measures. That's true, even if much of the discussion has been about limitations and things to watch out for. Here I'll summarize by listing important things to bear in mind—or in eye—when inspecting a scatterplot.

- Pearson correlation *r* is a measure of the strength of the *linear* component of the (*X*, *Y*) relationship.
- Even for linear relationships, when |*r*| is around .3 or less, the scatterplot is close to a shotgun blast. Even for larger *r* there's considerable *scatter*.
- For large |*r*|, *tightness* to the line is helpful for eyeballing the value of *r*. For any *r*, eyeballing the balance of points in the matched and mismatched *quadrants* can help.
- Look for signs of departure from linearity, perhaps a *curvilinear* relationship, or a *ceiling* or *floor* effect.
- *Outliers* can have an enormous influence on *r*, which is especially sensitive to points far from the means of *X* and *Y*.
- Examine the scales on the *X* and *Y* axes for any sign of a range restriction. Also consider the sample. A *range restriction* can cause a sample to underestimate the correlation in the whole population, perhaps severely.
- Given *r* but no scatterplot, take care: Would the *scatterplot* reveal important aspects we're not being told?

INFERENCE: CONFIDENCE INTERVALS AND *p* VALUES

As a descriptive statistic, *r* measures the strength of the linear component of the relation between *X* and *Y* in a sample. For inference to the population, we need the CI on *r*, and I'll also mention *p* values. First comes the dance of the *r* values.

Dance of the *r* Values

Before discussing CIs on means back in Chapter 5, I used the dance of the means to illustrate the sampling variability that a CI describes. Let's do the same for correlations. I'll consider *r* for data sets that are random samples from a population that we assume is a very large, even infinite, collection of (*X*, *Y*) points. We assume *X* and *Y* have correlation ρ (rho) in the population—remember, don't confuse ρ with *p* as in *p* value.

Values of *r* in repeated random samples from a population with correlation ρ will vary because of sampling variability. This is the ***dance of the r values***.

With ρ = .5, I took 50 samples each of *N* = 10 and recorded the 50 *r* values. I did the same for *N* = 40 and *N* = 160. Table 11.1 reports some of my results: Because of sampling variability, the values in each column dance—this is the *dance of the r values*, or *dance of the correlations*.

We expect larger *N* to give less sampling variability and narrower dances, and the table shows that this happens for *r*. In the first column, with *N* = 10, *r* ranges from –.35 to .78, whereas in the last, with *N* = 160, the range is only

Table 11.1 Values of *r* to Illustrate the Dance of the *r* Values

N	10	40	160
	.78	.45	.59
	–.13	.28	.46
	.81	.59	.57
	.48	.31	.54
	.52	.71	.45
	.53	.39	.34
	.43	.53	.61
	.19	.50	.44
	–.35	.70	.47
	.41	.36	.52
	...	...	...
Mean of 50 *r* values	.46	.44	.51
SD of 50 *r* values	.30	.16	.06

from .34 to .61. There's *much* more variation with smaller *N*. At the bottom are the means and SDs for my full sets of 50 *r* values. As we'd expect, the three means are all fairly close to ρ = .5. The SDs measure the amount of bouncing and indicate that, as expected, there's more sampling variability (a larger SD) for smaller *N*. The SD decreases markedly as *N* goes from 10 to 40 to 160.

I used the **See r** page to take samples and obtain the *r* values in Table 11.1. You can see the dance for yourself.

esci 11.14 At red 3, click the radio button on, to select **Sample from population**. Use the slider at red 3 to select ρ. Check the top checkbox at red 5 to display the value of *r*. Click the button at red 4 to take a random sample of *N* points. Keep clicking, and watch how *r* and the patterns of points dance.

esci 11.15 Try various values of *N* and ρ.

a. For ρ = .5, compare *N* = 10 and *N* = 160. Can you see any difference in the amount of variability in *r*, meaning the width of the dance?
b. For *N* = 40, compare ρ = 0 and ρ = .9. Any difference in the width of the dance?

THE CONFIDENCE INTERVAL ON *r*

Just as we use sample mean, *M*, to estimate population mean μ, we can use *r* for our sample as an estimate of ρ, the correlation in the population it came from. So *r* is our point estimate of ρ, and we want a CI on *r* as our interval estimate of ρ. First, we need to consider the statistical model underlying the CI calculation—in other words, the assumptions that are required. The model is that our data set of *N* pairs of (*X*, *Y*) values is a random sample from a very large, even infinite, population of such (*X*, *Y*) data pairs in which *X* and *Y* have a ***bivariate normal distribution***, with correlation ρ. "Bivariate" means, as you probably guessed, that there are two variables, *X* and *Y*.

In a ***bivariate normal distribution***, *X* and *Y* are each normally distributed overall, each is normally distributed at any single value of the other, and the variance of each is homogeneous for all values of the other.

Think of a bivariate normal distribution as a smooth hill, flattish on top, with contours that are roughly elliptical in shape—in the shape of an ellipse, or oval. Panel A in Figure 11.15 illustrates the idea by showing the scatterplot for a random sample of *N* = 5,000 points from a bivariate normal distribution. The correlation is *r* = .6. Overall, both *X* and *Y* are normally distributed. In addition, if we choose a single value of *X*, the distribution of *Y* at that *X* value will be

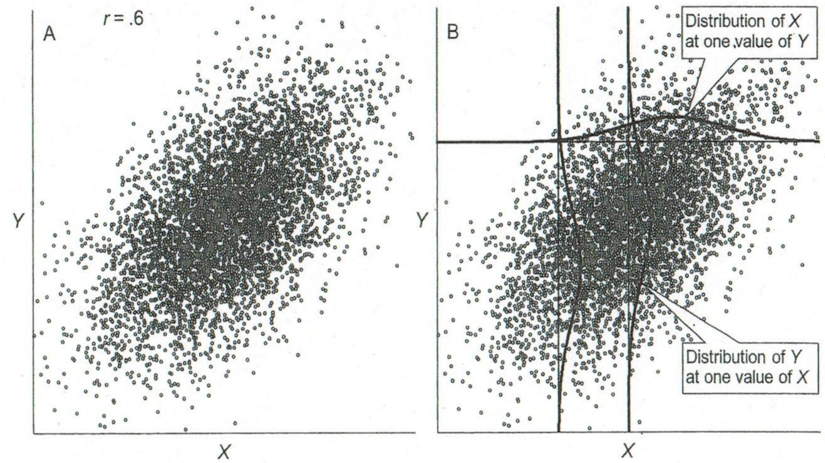


Figure 11.15. Random sample of $N = 5{,}000$ points from a bivariate normal distribution. Correlation is $r = .6$. Panel B shows the distributions of Y at two values of X, and the distribution of X at one value of Y.

normally distributed. Panel B illustrates such distributions of *Y* for two values of *X*. All such distributions are normal and, moreover, have the same standard deviation for every value of *X*. We say that the variance of *Y* is homogeneous for all *X*. The figure also shows one distribution of *X* values at a single value of *Y*. That's also normal, and also has the same standard deviation for all *Y*, so the variance of *X* is homogeneous for all *Y*. Scatterplots for samples from bivariate normal distributions look like those in Figures 11.2, 11.3, 11.5, 11.7, 11.8, and 11.9. The points are generally more closely clustered around the center, and become less close at positions farther from the center.

To calculate a CI on *r* we need to assume the data set is a random sample from a bivariate normal population. Is that a reasonable assumption? Often, but not always. It's usually close enough to reality for practical purposes, unless:

1. We have prior reason to suspect a nonlinear relationship. For example, we surely would have expected a curvilinear relationship between comfort and room temperature; or:
2. There are strong signs in the scatterplot of a nonlinear relationship, as with the two Figure 11.4 examples; or:
3. It doesn't make sense to think of our data set as a sample from a larger population. I'll say a word about this next.

If our data set is, for example, the number of points scored this season (*X*) and the income this year (*Y*) of the players in a particular basketball team, we could certainly calculate *r* for *X* and *Y*. However, for any kind of inference, not only for CIs on *r*, it needs to make sense to think of our data as a sample from a population. In this example, would it make sense to calculate a CI on the *r* for our team? If our interest is specifically in that team, then no, it wouldn't, because our data tell us about the whole team—the whole population we are interested in—and so we shouldn't calculate a CI. Perhaps if we were interested in basketball players in general we might think of our team as a sample,

If we do make the assumption of a bivariate normal population, all we need is the values of *N* and *r* and we can calculate a CI on *r* as our interval estimate of population correlation ρ. The details are complicated, so we'll simply let ESCI calculate the CI, and also the *p* value for any null hypothesis.

Calculation of a CI on *r*, or a *p* value to test a null hypothesis about ρ, requires the assumption that the data are a random sample from a bivariate normal distribution.

Figure 11.16 displays the 95% CI on $r = .4$ when $N = 30$, and near red 2 reports that the CI is [.05, .66], rounded to two decimal places. Note the CI is asymmetric, with upper MoE smaller than lower—the values are reported near red 2. The cat's-eye picture is also asymmetric—more bulge near the top, longer tail below. The null hypothesis value of $\rho_0 = 0$, as specified by the small slider near red 4, is marked by the red horizontal line. The corresponding *p* value of .028 is shown near 4. Do you agree that this *p* value looks about right? It's consistent with the lower limit of the CI being just a little above zero, so *p* is a little less than .05.

esci **11.16** Open the **One correlation** page and start with the checkboxes at red 3 and 4 *not* checked. Explore, using the spinner at red 1 to set *N* and the large vertical slider to set *r*.

a. For fixed *N*, try both positive and negative *r*.
b. Describe how the CI changes as you change *r*, with small *N*, then with large *N*. Any difference?

esci **11.17** For fixed *r*, describe how the CI changes as you change *N*. Describe how the pattern of change differs for small and large *r*.

esci **11.18** In Figure 11.1, $N = 106$ and $r = .47$. Find the CI on *r* and interpret.

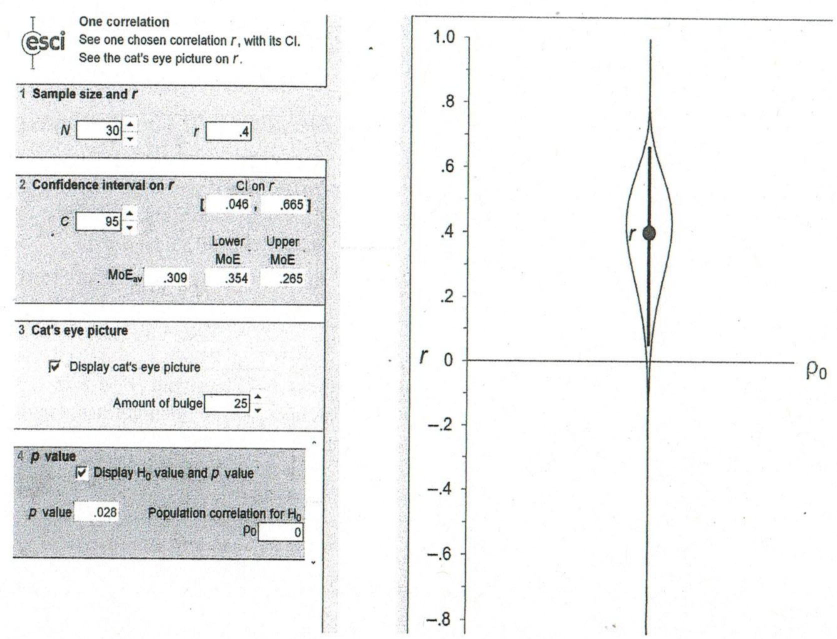


Figure 11.16. The 95% CI on $r = .4$, with $N = 30$. The cat's-eye picture is displayed, and the null hypothesis value $\rho_0 = 0$ is

11.19 In See r click **Sample from population** at red 3. Set $N = 75$ and $\rho = -.7$.

a. Take a few samples and note down the values of r.
b. Use **One correlation** to find the CI for each, and note whether each captures ρ.
c. If you were to take many further samples, what can you say about the CIs for those samples? What can you say about the capture of ρ?

11.20 If you knew the CIs for all the r values in Table 11.1, how would the CI lengths compare for the three columns? For each column, what can you say about capture of $\rho = .5$?

11.21 For a particular type of hybrid plant, a genetic model predicts a correlation of .5 between leaf length and width. You measure leaf length and width of 100 such plants and calculate $r = .36$. What do the data tell you about the model? Consider both the CI and the p value. *Hint*: For the p value, use the small slider near red 4 to set ρ_0.

11.22 A second genetic model predicts a correlation of .25.

a. What can you say about this second model?
b. Considering the two models, what would you recommend?

For fixed N, the CI on r gets shorter and more asymmetric as r moves farther from 0, especially when r gets close to 1 or −1.

Figure 11.17 illustrates what you may have discovered in Exercise 11.16. With fixed N, when $r = 0$ the CI is long and symmetric. The CI gets shorter and more asymmetric as r moves farther from 0, especially as it approaches 1 or −1. Of course, no CI can extend beyond those boundaries. For values of r that are equally far above and below zero, the CIs are mirror images. The lines in the figure at 1 and −1 are "fences" that "squash" any CI that approaches, especially the MoE nearer the fence.

For fixed r, if N is multiplied by four, the CI on r becomes roughly half as long.

Figure 11.18 illustrates what you may have discovered in Exercise 11.17. With fixed r, when N is small the CI is long. Increase N and the CI gets shorter, as we'd expect. Figure 11.18 illustrates this change with N for $r = 0$, and also

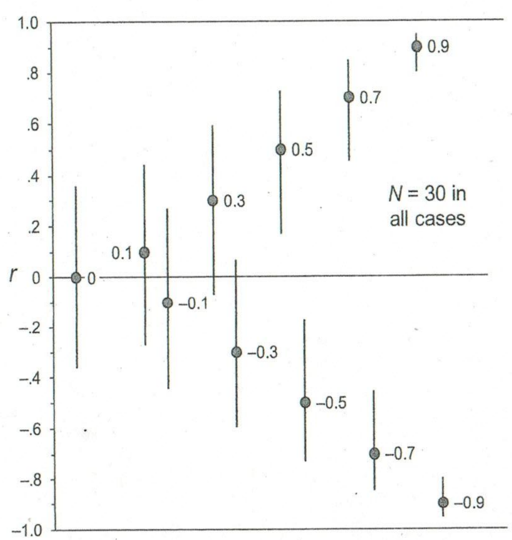


Figure 11.17. Examples of 95% CIs for various values of r, all with N = 30. The value of r is shown

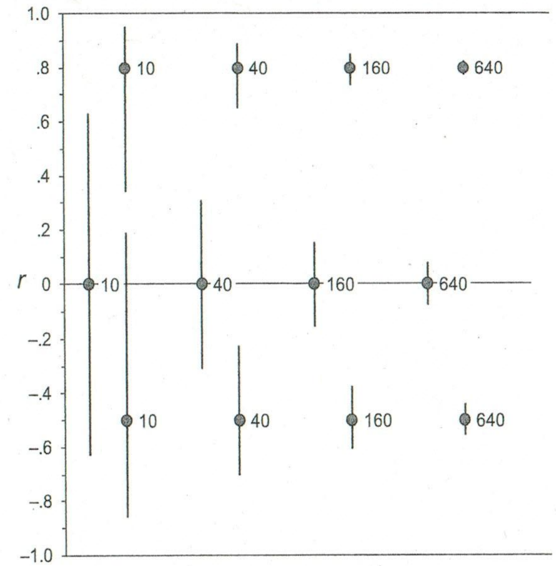


Figure 11.18. Examples of 95% CIs for various values of N and r. The value of N is shown next to each point. For the upper four points, $r = .8$, the middle four, $r = 0$, and the lower four, $r = -.5$.

for $r = -.5$ (lower four points) and $r = .8$ (upper four points). Recall the general guideline, which we first encountered in Chapter 1, that four times the sample size gives, approximately, a CI half as long. Does that hold for r? Does multiplying N by four give a CI on r that's half as long?

Yes it does, although only roughly, as you may have concluded from Figure 11.18.

11.23 Consider the SD values in Table 11.1.

a. Note that from the first to the second column, N is multiplied by 4 and the SD changes from .30 to .16, which is about half.
b. There are similar changes from the second to third column. Is that reasonable? Explain.

11.24 The length of a CI on a sample mean, M, depends on N, but it doesn't depend on M itself. What about a CI on r? Does it depend on N? Does it depend on r itself?

Now we've discussed CIs, I want to discuss planning a correlational study, then the interpretation of correlations and scatterplots. But first it's quiz time.

Quiz 11.2

1. Before using Pearson's r to describe a relationship, it is essential to first look at a scatterplot of the relationship. What should you be looking for?
   a. Whether the relationship is linear; if not, r should probably not be used.
   b. Whether there are outliers that are having a large influence on r.
   c. Whether the range of either the X or the Y variable is restricted, which can produce misleading values for r.
   d. All of the above.

2. The larger the sample size, the longer / shorter the CI on r. For a fixed N, the closer r is to 0, the longer / shorter the CI.
3. A CI on r depends on the assumption that the data set is a(n) _______ sample from a(n) _______ distribution.
4. Jaime measures IQ in 10 students and also obtains an IQ measure for each student's mother. He finds $r = 0.6$, 95% CI [−.05, .89]. This means that the population correlation could most likely be anywhere from _______ to _______.
5. The CI Jaime obtained is quite long. To obtain a CI half as long he would need to test about _____ participants.
6. As r gets closer to −1 or 1, the CI becomes _______ and more _______.

Planning a Correlational Study

In Chapter 10 we discussed how to choose N so our planned study would give a CI with our chosen target MoE, either on average or with 99% assurance. Planning a correlational study is more complicated because, as you saw in Figures 11.17 and 11.18, for r the CI length depends not only on N, but also on r itself. Exercise 11.24 emphasized the same point. To use precision for planning to choose N, when r is our effect size, we would need to specify a target r as well as target MoE. An added complication is that, unless $r = 0$, the CI on r is asymmetric, so upper MoE and lower MoE are different.

Rather than attempting a detailed discussion, or making an ESCI page, I've included Figure 11.19 to give the general idea. Use this to see what CI you would obtain if $r = .1$ or $.3$ or $.5$, and you use one of the values of N shown in the figure. The relationship of r, N, and CI length is a bit complicated, but the

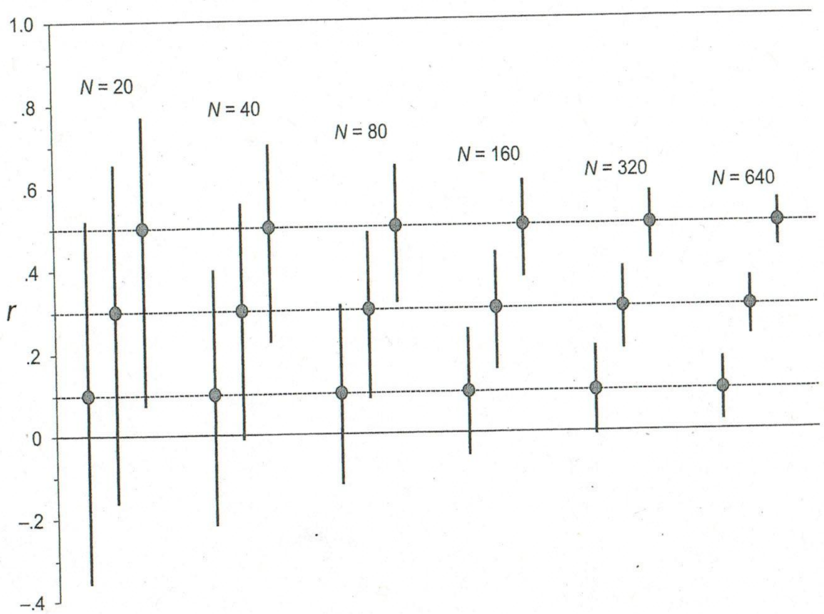


Figure 11.19. Examples of 95% CIs for $r = .1, .3,$ and $.5$, respectively, left to right, for each of various values of N, as indicated at the top.

figure can help you choose N likely to give a sufficiently short CI, especially if you have some idea of the size of population correlation you are investigating.

There's one important way in which the situation with r is less complicated and less uncertain than for means. The CI depends only on r and N, and there's no additional need to estimate σ or worry about assurance. For any particular r and N, the CI is determined, and any replication with the same N that happens to give the same r will give exactly the same CI. Any sample with $N = 20$, for example, which has $r = .1$ will give exactly the leftmost CI in the figure.

Studying Figure 11.19 suggests a few approximate guidelines. For correlations between about −.5 and .5:

- For N up to around 40, CIs are long, with MoE usually around .3 or more.
- To achieve MoE of around .2 or less, in most cases N of about 100 or more is required.
- To achieve MoE of around .1 or less, in most cases N of about 300 or more is required.

As Figure 11.17 illustrates, for larger correlations ($|r| > .5$) CIs are shorter, especially as $|r|$ approaches 1.

11.25 a. You are investigating a correlation you suspect is around .3 to .5, but can collect data only for a sample with $N = 80$. About what MoE can you expect?

b. You expect r around .1 to .2, and are keen for MoE not more than 0.1. With $N = 200$, are you likely to achieve that? About what N are you actually likely to need?

INTERPRETING CORRELATIONS AND SCATTERPLOTS

Correlation and Causation

"Can you believe, from 1999 to 2009 the correlation was $r = .95$ between annual U.S. oil imports from Norway and the annual number of drivers in the U.S. who died in collisions with trains!" At www.tylervigen.com/ you can see that and many other examples of bizarre correlations. They are good for a chuckle, but also alert us to the danger of seeing causation in a correlation. As I first mentioned when discussing Figure 11.1, if X and Y correlate, there may be interesting causal links, but the scatterplot can't tell us what they are. We might even be seeing an accidental lump in randomness: Examine a large number of possibly totally unrelated variables and you'll eventually find two that correlate, simply by chance. You would most likely be seeing a face in the clouds. We can't be sure, but I suspect that $r = .95$ for oil imports and train collision deaths is such a blip, such a face in the clouds, with no instructive underlying causes to be found. For correlations we encounter in research, however, there are often causal links to be investigated—although there may not be.

If X and Y are uncorrelated, with r close to zero and a shotgun blast scatterplot, then *most likely* there are no interesting causal links between X and Y. However, if r and a scatterplot suggest there is some relationship between X and Y, there are several possibilities:

- Simple causation: Either X causes Y, or Y causes X.
- A more complex pattern of causation, possibly including one or more other variables.
- There are no causal links, and we're seeing a face in the clouds.

Correlation does not necessarily imply causation.

Correlations can give valuable guidance as we investigate causality, but even a large r cannot immediately tell us which of the above possibilities applies. Hence the slogan "Correlation does not necessarily imply causation".

In Figure 11.4, the left scatterplot shows the curvilinear relation of Y (comfort) and X (room temperature). It's natural to assume that X causes Y, although if low comfort leads you to adjust the thermostat, Y would be influencing X. In plots of Thomason 1 posttest (Y) against pretest (X) scores, like Figures 11.10 and 11.11, we assume both X and Y are strongly influenced by a third variable, for example reasoning ability. Often we might consider various patterns of causation between X and Y, and perhaps other variables. Knowing just the correlation, however, *cannot* identify for us the pattern of causation. We need to be alert, because the slide from seeing correlation to assuming causation can be subtle and appealing, as the following example illustrates.

The slide from correlation to causation may be subtle and highly plausible, although unjustified. Be alert.

"Parental divorce leads to adolescent drug use." News reports like that would, no doubt, be based on data, but would the data come from experiments, with participants randomly assigned to different conditions? Of course not—it's impossible as well as totally unethical to assign families randomly to the Divorce or the Don't divorce condition, then come back a year or two later and note their children's drug use. Without that, however, it's very difficult to draw confident conclusions about causality. The news report is almost certainly based on a correlation: Researchers noted a tendency for children of divorced parents to use drugs more than those in intact families. Many causal links may be involved, or perhaps none at all—we may only be seeing a lump in the randomness. Perhaps cause runs in the opposite direction: Children's drug use causes marital problems. Perhaps there's an indirect link, with parental education, employment status, parenting skills, and housing situation all being important influences on both divorce and adolescent drug use. Researchers have developed advanced techniques to gather and analyze data to tease out such complex patterns of causation. These usually require investigation of many variables together, then analysis of a complex data set including many correlations.

The claim made by the news report may be highly plausible. However, it implies a direct causal link, but is based on a correlation. The causal link may or may not be correct, but most likely there are several, or many, underlying causal links that will be tricky to identify and untangle. Be alert to over-interpretation of correlations, a phenomenon which can be subtle and plausible.

11.26 "Eating vegetables linked to better progress in elementary school." Discuss that news report.

a. Is it plausible? What's it probably based on?
b. What causal links might there be? How might you investigate further?

11.27 Consider a couple whose marriage is violent and dysfunctional.

a. Does an overall strong correlation between divorce and adolescent drug use mean that the couple should necessarily avoid divorce for the sake of their adolescent children?
b. Explain how an overall strong correlation need not necessarily dictate what's best in an individual case.

Town Size and Walking Speed

Here's another interesting correlation to think about. Figure 11.20 is a scatterplot of data reported by Bornstein and Bornstein (1976). The researchers unobtrusively recorded the walking speed of individuals walking in the central city square, or main downtown street, in villages, towns, and cities in six countries. They found that, remarkably, walking speed tended to be more than twice as fast in large cities as in small villages. The correlation with population (expressed on a log scale) was $r = .90$, as Figure 11.20 reports.

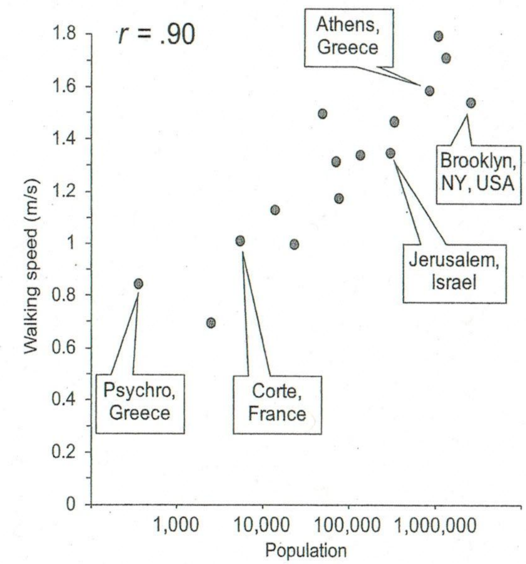
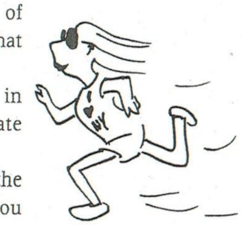


Figure 11.20. A scatterplot of average walking speed in a central public area, and population on a log scale, for towns and cities. Several of the points are identified. Data from Bornstein and Bornstein (1976).

11.28 How feasible would a close replication of the study reported in Figure 11.20 be? What would you expect it to find?

esci **11.29** Find the CI and the p value for the r in Figure 11.20. Is it reasonable to calculate those? Explain.

11.30 Suggest causal links that might underlie the correlation in Figure 11.20. How could you investigate your suggestions?



11.31 a. Search online for "walking speed and population size" or similar. I found several fascinating discussions about the pace of life in different countries, and different-sized towns and cities. Bornstein and Bornstein (1976) was identified as the starting point, and many more recent studies were mentioned.

b. Suggest an interesting further study.

11.32 My colleague Neil Thomason recalls chatting with a farmer from rural Bolivia amid the hubbub of New York City. The man asked: "Is everyone in New York important?" He explained that back home only important people walk fast. What further investigation does that observation suggest?

Reference Values for Correlation

I've been referring to $r = .1$ or $-.1$ as a small correlation, and $r = .9$ or $-.9$ as large. In relation to the possible range from -1 to 1 that's reasonable, but, just as with Cohen's *d*, any interpretation of particular *r* values should be a knowledgeable judgment in context. That's especially important here, because correlation is used in so many different ways, in so many different contexts, that it's not possible to specify any reference values with universal applicability.

Considering psychology, Cohen did suggest $r = .1$, .3, and .5 (or, equivalently, $-.1$, $-.3$, and $-.5$) as reference values for small, medium, and large correlations, but he emphasized that judgment in context should be preferred wherever possible. Others have suggested different reference values. Hinkle, Wiersma, and Jurs (2003), for example, labeled *r* values above .9 as "very high positive", values between .7 and .9 as "high positive", between .5 and .7 as "moderate positive", between .3 and .5 as "low positive", and between -.3 and .3 as "little if any correlation" (p. 109).

Interpret values of *r* in context. Correlation is used in such a variety of situations that reference values usually don't help.

Bosco et al. (2015) collected more than 140,000 values of *r* reported in organizational psychology research. This broad field studies people's attitudes, intentions, and performance, mostly in the context of employment. For that large set of *r* values, the first quartile was .07, the median was .16, and the third quartile .32. So researchers in this field study correlations most of whose scatterplots would fall between the two rightmost panels in Figure 11.2—they resemble shotgun blasts. Bosco et al. concluded that most research in organizational psychology focuses on correlations smaller than $r = .3$, and that $r = .2$ might often be a more realistic "medium" value. I conclude again that *r*, especially, needs to be interpreted in its particular context.

To interpret a value of *r*, consider also the CI, and any correlations reported by related past research. Also have in mind scatterplots like those in Figures 11.2 and 11.3. I'm always struck by how widely scattered the points are, even for *r* as large as .6. It's sobering to learn that many researchers are studying relationships between variables that have small values of *r* with scatterplots that look like shotgun blasts. Such relationships may be interesting and important—or they might not be—but, either way, it's still a shotgun blast.

Measuring Reliability and Validity

Test-retest reliability is usually assessed by a value of *r*.

In Chapter 2 I introduced reliability and validity as two important features of any measure. You may recall that reliability is repeatability or consistency, and it's often assessed by using *r* as a measure. To estimate the test-retest reliability of a new anxiety questionnaire, for example, you might calculate *r* for the anxiety scores of 100 people recorded on one day, with their scores on a second day, under conditions as similar as possible to those on the first day. I'm happy to report that many well-established psychological and educational tests give test–retest reliability correlations of .9 or even higher. That's great news—our best measures are highly reliable.

Correlation is used in various ways to assess validity.

Recall also that validity is the extent to which a measure actually measures what we want it to measure. Correlation is used in various ways to assess validity. I mentioned in Chapter 2 that we could, for example, correlate scores on the new anxiety questionnaire with scores on an already well-established measure of anxiety. A high correlation would suggest our measure of anxiety has reasonable validity. I also mentioned a test of job aptitude: The correlation between test scores and later job performance would be an estimate of the validity of the test. It's encouraging that many well-established psychological and educational tests often give validity correlations as high as .8 or .9. A good measure needs to have both good reliability and good validity. Assessing measures and developing new measures is likely to require much use of correlations.

With reliability and validity in mind, let's revisit the interpretation of *r* values. For reliability and validity, we might judge *r* values of .8 and .9 as good rather than very large, and .6 or .7 as small and inadequate rather than large. Here's an example. To study leadership, Chan (2007) obtained scores from 92 Chinese students in Hong Kong on an established rating scale for leadership, the SRBCSS, and a self-rating scale he was investigating, the RRSL. He found a correlation of $r = .38$ and reported that correlation to be statistically significant, $p < .01$, "suggesting that the Chinese RRSL has … validity when compared with Chinese SRBCSS leadership scores" (p. 160).

A *p* value for *r*, reported without mention of a null hypothesis value, almost always refers to a null hypothesis of zero, although we should be told that explicitly. Chan's conclusion was based on the *p* value indicating fairly strong evidence against the population value of the correlation being zero—absolutely no correlation whatsoever. But that's *not* good evidence that the validity correlation is usefully large. A correlation of only around .4 would usually be considered poor validity. For $N = 92$, the 95% CI on $r = .38$ is [.19, .54], so the validity correlation could plausibly have been as low as .2. It would have been more informative if Chan (2007) had reported *r* with its CI, then interpreted the point and interval estimates in terms of what they say about validity.

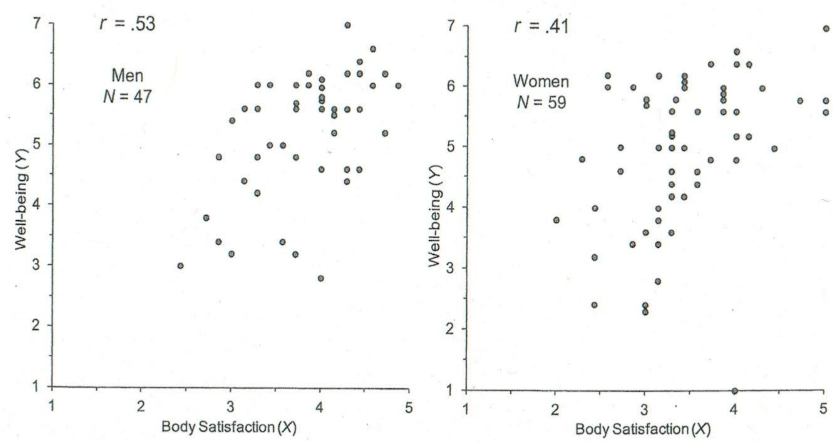


Figure 11.21. Scatterplots as in Figure 11.1, separately for women and men.

Beware any p value given with a value of r. What's the null hypothesis value, and is that, in the context, an appropriate reference point?

This example reminds us that a p value relates to a particular null hypothesis value. To interpret p we need to know that value, and consider its appropriateness as a reference point, in the context. Small p might lead us to reject the null hypothesis, but, for correlation, zero may be an inappropriate reference point, especially when assessing reliability or validity. Considering r and its CI is likely to lead to a better justified conclusion.

THE CI ON THE DIFFERENCE BETWEEN TWO INDEPENDENT CORRELATIONS

Now for some more inference. To compare two independent correlations we need the CI on the difference. By "independent" I mean the two correlations come from separate groups—for example the correlations for women and for men. The sample of 106 students whose data are shown in Figure 11.1 comprised 59 women and 47 men. Figure 11.21 shows the separate scatterplots, and that $r = .41$ for women and $r = .53$ for men. Is that surprising? Isn't body satisfaction considered more central to well-being by women? Before we get too excited by what may be an unexpected difference, we need to see the CI on that difference.

The CI on r is usually long, unless N is very large. The CI on the difference between two independent r values is usually especially long.

The CI on a difference between r values is tricky to calculate, but I use a good approximate method described by Zou (2007). Figure 11.22 displays $r_1 = .41$ for women, and $r_2 = .53$ for men, each with its CI. The difference between those correlations and the CI on the difference are displayed at right. The difference is shown at red 6 on the left to be $(r_2 - r_1) = 0.12$ [−0.19, 0.42] to two decimal places. As usual, I was surprised how long the CIs are, even with groups as large as 59 and 47. The CI on the difference tells us there's considerable uncertainty, because the population difference between the correlations for women and men could plausibly be as low as around −0.2 or as high as around 0.4. So there's no evidence of a population difference.

11.33 In Figure 11.22, at red 7 on the left, the p value is reported. What's the null hypothesis? Interpret and compare with the CI.

esci **11.34** a. At **Two correlations**, set the N values at red 1 and 2 and use the two large sliders to set the correlations for our example. Click at red 6 for the difference axis and red 7 for the p value. Your screen should resemble Figure 11.22.

b. Suppose the two groups had been four times the size (236 women, 188 men) and obtained the same r_1 and r_2 values. Find the CI on the difference and interpret.

esci **11.35** Using the original group sizes and r_1 value, how large would r_2 need to be for the p value to be less than .05? How does that strike you?

CORRELATION r FOR META-ANALYSIS

Is genius born or made? Could any of us be Michael Jordan, or Mozart, if we worked sufficiently hard to develop the requisite skills? Meta-analysis of correlations can help answer such questions. The issue here is the extent that practice and effort may be sufficient for achieving the highest levels of expertise. Ericsson, Krampe, and Tesch-Romer (1993) argued that years of effort is what matters most: "Many characteristics once believed to reflect innate talent are actually the

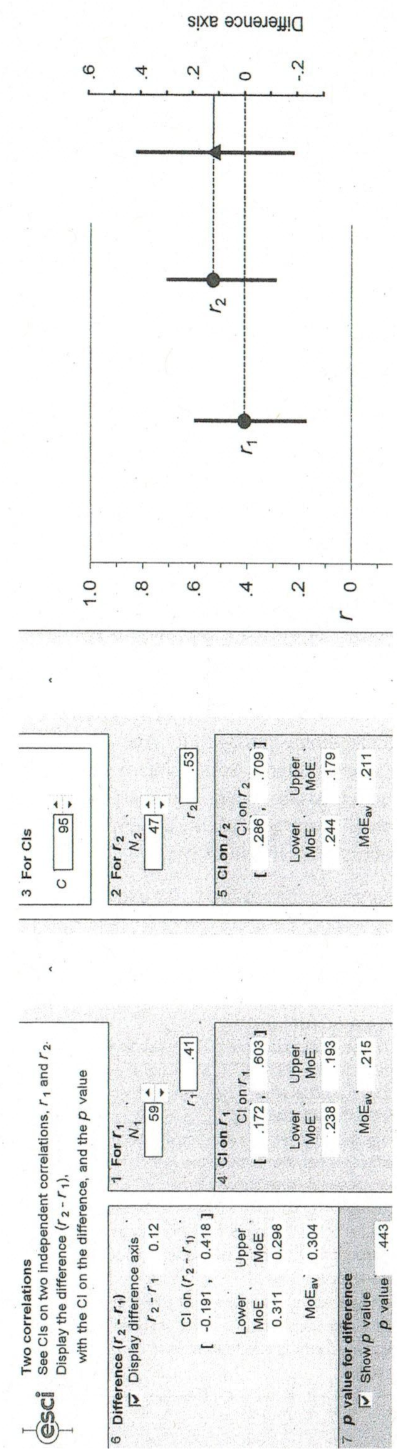


Figure 11.22. The difference between two correlations from Figure 11.21: $r_1 = .41$ with $N = 59$ for women, and $r_2 = .53$ with $N = 47$ for men. The difference is depicted by the triangle at right on the difference axis, with its CI. The limits of the CI on $(r_2 - r_1)$ are shown at left, near red 6. From **Two correlations**.

was enormously popularized by Malcolm Gladwell (2008), who argued in his book *Outliers* that 10,000 hours of focused practice is the key to achieving expertise.

r can be a useful effect size measure for meta-analysis.

However, this view is now being challenged, with one important contribution being a large meta-analysis of correlations between amount of intense practice and level of achievement: Macnamara et al. (2014) combined 157 correlations reported in a wide range of fields, from sports to music and education, and found $r = .35$ [.30, .39].

Figure 11.23 shows some of the data from Macnamara et al. (2014). As usual, to carry out meta-analysis we need to judge that all the studies are examining the same or sufficiently similar questions. The figure shows the random effects meta-analysis of the 16 main correlations for music. You can see at red 7 that the overall result is $r = .41$ [.28, .53], which is consistent with the main conclusion of the larger meta-analysis. The example shows that *r* can be useful for meta-analysis. Indeed correlation *r* and Cohen's *d* are probably the most widely used effect size measures for meta-analysis.

The researchers' conclusion was that, in many fields, the correlation between amount of practice and achievement is only modest, a result that conflicts with the established position of Ericsson and others. Search online for "10,000 hours and expertise", or "10,000 hours", and you should easily find articles describing the Ericsson–Gladwell position. You should also find recent articles with titles like "Scientists debunk the myth…" of 10,000 hours. No-one doubts that long effort and persistent focused practice is needed to achieve expertise, but the question is the extent of the contribution of other factors, notably innate talent. Thanks partly to meta-analysis of correlations, it seems that once again we can believe that genius is to some extent born, and not only made.

It's time for take-home messages. To write yours, you could think back to the pictures we've encountered. Scatterplots of course, but also quadrants and the battle, shotgun blasts, confidence intervals of course, and finally a forest plot. I couldn't think of any good picture of causation—maybe an arrow?—but causation probably springs to mind anyway.

Quiz 11.3

1. For a particular *N*, the CI for $r = .6$ is longer / shorter than the CI for $r = .3$. For a particular *r*, larger *N* gives a CI that is longer / shorter.
2. Correlation does not necessarily imply __________.
3. Messerli (2012) reported that countries in which people eat lots of chocolate also tend to win lots of Nobel prizes ($r = .79$, 95% CI [.55, .91], $N = 22$). Does this mean eating chocolate will cause you to win a Nobel prize? Suggest some other explanations for this association.
4. For *r*, Cohen's reference values are __________, __________, and __________ for small, medium, and large, but interpretation of *r* should depend on the __________.
5. The CI on the difference between two independent correlations is
   a. not helpful unless sample sizes are very large.
   b. sure to include zero.
   c. shorter than either of the CIs on the two separate correlations.
   d. longer than either of the CIs on the two separate correlations.
6. Considering meta-analysis,
   a. it's a disadvantage that the CI on *r* is almost always asymmetric.
   b. values of *r* need to be transformed to values of Cohen's *d*.
   c. it's advisable to use the fixed effect model when using *r*.
   d. None of the above.

2 ☑ Display by weights

Effect size, *r*, for each study

Study no.	Study name	*r* (1)	*N*	Study in?	LL (Correlation *r*)	UL (Correlation *r*)	Study weight % (11)
1	Violin, viola	.67	109	☑	[.551	, .762]	8.0
2	Strings	.51	55	☑	[.283	, .683]	6.8
3	Piano	.4	19	☑	[-.066	, .723]	4.0
4	Piano	.46	30	☑	[.12	, .704]	5.2
5	Piano	.47	19	☑	[.02	, .762]	4.0
6	Piano	.228	52	☑	[-.048	, .471]	6.6
7	Piano	-.224	24	☑	[-.575	, .197]	4.6
8	Piano	.104	52	☑	[-.174	, .366]	6.6
9	Piano	.322	16	☑	[-.207	, .705]	3.5
10	All	.231	97	☑	[.033	, .411]	7.8
11	Piano	.67	57	☑	[.496	, .792]	6.8
12	Piano	.41	107	☑	[.239	, .557]	8.0
13	Band	.34	178	☑	[.203	, .464]	8.6
14	Music majors	.31	64	☑	[.069	, .516]	7.1
15	Music majors	.54	19	☑	[.114	, .798]	4.0
16	All	.583	135	☑	[.459	, .685]	8.3
17				☑	[	,]	
30				☑	[	,]	

7 **Meta-analysis result**

Random effects .406 [.28 , .532]

8 Model for meta-analysis: ○ Fixed effect ◉ Random effects

9 Diamond ratio 2.03

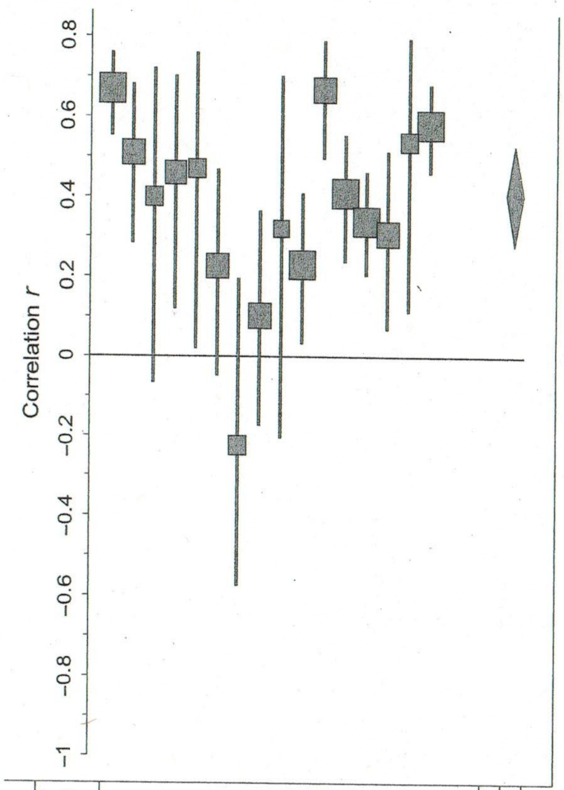


Figure 11.23. Random effects meta-analysis of the 16 main correlations of amount of practice with level of achievement, for music, from Macnamara et al. (2014). For each study the instrument or type of music training is shown under **Study name**, and *r* and *N* have been entered below red 1. From **Single r** in **ESCI intro Meta-Analysis**.

esci 11.36 Open the **Single r** page of **ESCI intro Meta-Analysis** and you should see the data shown in Figure 11.23.

a. At red 8, click between fixed effect and random effects meta-analysis, and note the diamond ratio.

b. What can you say about heterogeneity? Is your response consistent with the appearance of the forest plot? Explain.

11.37 Which CIs in the forest plot are most asymmetric? Is that what you expect?

esci 11.38 Click at red 4 to display p values for the individual studies and the result.

a. What null hypothesis do the p values relate to? What could you conclude?

b. Compare your conclusion with the conclusion of Macnamara et al. (2014).

c. Which approach to data analysis, the one using p you have just looked at or that of Macnamara et al., should we prefer and why?

11.39 Revise your take-home messages if you wish.

Reporting Your Work

Correlation may not imply causation, but identifying and interpreting associations between variables still serves as one of the most important tools available for scientific research. When reporting correlations, you should usually include:

- whether examining the correlation is planned or exploratory, unless this is already clear—be sure your research plan is more thought out than just "examine all possible correlations" or you'll likely be seeing faces in the clouds;
- basic descriptive statistics for both variables being correlated;
- a scatterplot if possible;
- the value of r and its CI. Remember that calculating a CI requires assuming the data are a sample from a bivariate normal population. If that seems problematic, make a comment and consider not reporting a CI;
- the sample size for calculating r (which, due to missing data, may not be the same as the sample size collected);
- if desired, the p value—be sure to state the null hypothesis, which is usually but not always that $\rho = 0$;
- an interpretation of the correlation that considers not only the point estimate but also the CI—consider what the full range of the CI means in terms of correlation strength, interpreted in the context, including any relevant past research; if range restriction is evident make your interpretation suitably tentative.

For correlations, report r, its CI, and sample size. Be sure to include a scatterplot and to discuss how well assumptions for correlation are met.

Typically, you will be reporting correlations for non-experimental research, where no variables have been manipulated. In this case, an essential guideline for reporting is to avoid causal language. That is, avoid phrases like "the effect of X on Y" or "X produced a change in Y". Instead, use language suggesting association without causation. For example:

- The relationship between X and Y was examined.
- Participants with high X also tended to have high Y.
- This shows that X is weakly related to Y.
- There was a strong association between X and Y.

Finally, report r values without a zero before the decimal point (e.g., $r = .36$). That's due to the persnickety APA style rule you've read about in previous chapters that there is no leading zero for statistics that cannot exceed 1 (APA, 2010, p. 113) in absolute value.

Here are some text-only examples:

The correlation between well-being and self-esteem was $r = .35$, 95% CI [.16, .53], $N = 95$. Relative to other correlates of well-being that have been reported, this is a fairly strong relationship. The CI, however, is somewhat long and consistent with anywhere from a weak positive to a very strong positive relationship.

The correlation between well-being and gratitude was $r = .35$, 95% CI [−.11, .69], $N = 20$. The CI is quite long. These data are only sufficient to rule out a strong negative relationship between these variables.

The correlation between well-being and GPA was $r = .02$, 95% CI [−.18, .22], $N = 95$. The CI suggests at most a weak positive or negative relationship between these variables.

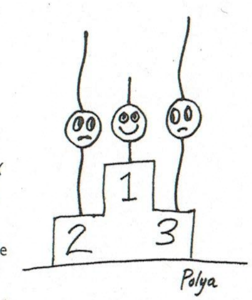
It is popular in journal articles to report large numbers of correlations using a correlation matrix. This is a very efficient way of reporting correlations, but be sure to include confidence intervals. As you can see in Table 11.2, in a correlation matrix each variable is represented along both rows and columns. Each cell reports the r value between its row and column variables. The diagonal cells are not reported because these are cells where the row and column variables are the same. (What would the correlation be between a variable and itself? Think about it or use ESCI to try for yourself.) Note that a correlation matrix is symmetric above and below the diagonal, so usually only one or the other is filled in. Although that may sound complicated when written out, in practice correlation matrices are easy to use. For example, in the matrix of three variables below, what is the correlation between well-being and negative affect? Between negative and positive affect? If this is making sense, your answers should be −.34 and −.10.

Table 11.2 Correlation Matrix for Different Components of Happiness

	Well-being	Positive Affect	Negative Affect
Well-being	—		
Positive Affect	.37 [.25, .47]	—	
Negative Affect	−.34 [−.44, −.22]	−.10 [−.22, .02]	—

Although you know that seeing a scatterplot is essential to interpreting the relationship between two variables, you'll notice that many journal articles report r values without accompanying scatterplots. This is a convention left over from a time when preparing figures was time consuming and expensive. Nowadays, figures are cheaper and there is a stronger emphasis on showing the data. Therefore, try to include scatterplots for key correlations whenever possible. Some statistical software even allows the creation of scatterplot matrices, in which each correlation value in a correlation matrix is represented by its scatterplot. This is an extremely powerful way to summarize lots of relationships at once. Follow this link: tiny.cc/spmatrix to an example of a scatterplot matrix. There are four variables, so six scatterplots, which are shown in the lower triangle of the matrix. The r values themselves are shown in the upper triangle.

Take-Home Messages



- Pearson correlation, r, is a measure of the strength of the linear component of the relation between two interval variables, X and Y. It can take values between −1 and 1.
- The scatterplot of X and Y is a picture of the N data points. The cross through the X and Y means can be helpful for eyeballing: r reflects the balance between points in the matched (HH, LL) quadrants and in the mismatched (LH, HL) quadrants.
- The value of r is the sum of the $Z_X Z_Y$ values for the points, divided by $(N - 1)$. It's the outcome of a battle between the matched and mismatched quadrants.
- Seeing the scatterplot is essential for understanding aspects of the relation between X and Y beyond the linear component measured by r. Watch for range restrictions, floor or ceiling effects, and curvilinear relationships.

- To calculate the CI on r, the CI on the difference between two independent correlations, or a p value, we must assume a bivariate normal population of (X, Y) values, with population correlation ρ. CIs are usually asymmetric and CI length depends on both N and r.
- The p value can be calculated using any ρ value as null hypothesis, but often $\rho_0 = 0$ is used without sufficient thought as to whether it's appropriate in the context.
- Correlation does not necessarily imply causality. An r value may be merely a face in the clouds, or there may be complex causal links amongst X, Y, and perhaps other variables.
- Values of r need to be interpreted in context. The Cohen reference values of .1, .3, and .5, and any other reference values, are only applicable in some contexts.
- Correlation is often used to assess the reliability and validity of measures, but beware p values and prefer CIs. Good measures can have values of $r = .8$ or higher.
- It can be useful to combine values of r by meta-analysis.

End-of-Chapter Exercises

1) For each of the scatterplots in Figure 11.24, eyeball an r value.

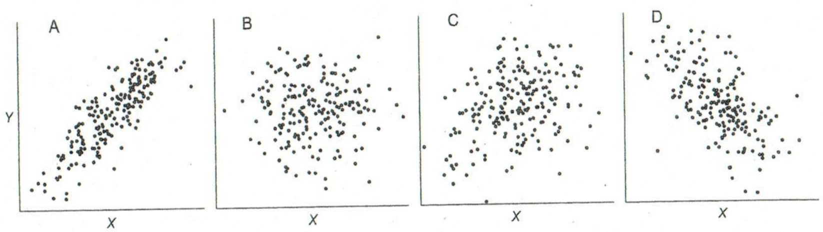


Figure 11.24. Four scatterplots for eyeballing practice.

2) To what extent does initial performance in a class relate to performance on a final exam? Table 11.3 lists first exam and final exam scores for nine students enrolled in an introductory psychology course. Exam scores are percentages, where 0 = no answers correct and 100 = all answers correct. You can load this data set (**Exam_Scores**) from the book website.

Table 11.3 Initial and Final Exam Scores for Nine Students

StudentID	Exam 1, X	Final Exam, Y	Z_X	Z_Y	$Z_X Z_Y$
1177	85.0	72.0			
1288	96.8	92.0			
1327	100.0	96.0			
1911	100.0	95.0			
1862	84.3	91.0			
1578	83.0	88.0			
1022	96.8	77.0			
1915	89.5	86.0			
1116	54.0	75.0			
Mean	87.71	85.78			
SD	14.35	8.97			
				Total $\sum Z_X Z_Y$	
				$(N-1)$	
				$r = \dfrac{\sum Z_X Z_Y}{(N-1)}$	

a. Enter the first and final exam scores into ESCI to generate a scatterplot. Just looking at the scatterplot, what r value do you eyeball?

b. Are there any cautions with using r to describe this data set? Does the relationship seem strongly nonlinear, is there restriction of range or a ceiling effect, or are there extreme outliers? Remember that to assess restriction of range it's best to set the scales of X and Y to show the full possible range of 0–100. You could edit the two axes to show that range.

c. Fill the blanks in Table 11.3 to calculate r, using a calculator or spreadsheet. Fill the Z_X and Z_Y columns, then multiply each student's pair of z scores to fill the $Z_X Z_Y$ column. Then complete the calculations to fill the bottom right cells and find r.

d. Using ESCI, you should find the same value of Pearson's r between Exam 1 and Final Exam scores.

3) Below are correlations reported in recent journal articles. For each, use the **One correlation** page of ESCI to calculate the 95% CI for ρ (rho). Then interpret. Assume that the authors verified that the relationships are reasonably linear and otherwise appropriate for description using r.

a. To what extent does income inequality relate to academic dishonesty? To investigate, Neville (2012) measured income inequality for each U.S. state, then measured search traffic for each state related to academic dishonesty (e.g. "buy term papers", "free college papers"). He found that states with higher income inequality also tend to have higher levels of search traffic related to academic dishonesty: $r = .45$, $N = 50$.

b. To what extent is early success in school related to income later in life? Ritchie et al. (2013) obtained annual income data for a large sample of British adults, then correlated these with reading scores from age 7. (For most participants, this was about 40 years prior to the income measurement.) Early reading scores were correlated with income, $r = .21$, $N = 1000$.

c. The full study of Ritchie et al. (2013) was much larger, with more than 14,000 participants. Would the CI for the full study be shorter or longer than you just calculated using $N = 1000$? Would it be greatly or only slightly different in length?

d. To what extent is your weight related to your ideas about diet and exercise? McFerran et al. (2013) conducted an online study in which participants were asked to report their Body Mass Index (BMI). Participants also rated the degree to which they believed exercise is more important than diet for controlling weight. The belief in exercise to control weight was correlated with BMI: $r = .25$, $N = 84$.

4) For each of the findings in Exercise 3, give at least two different causal explanations that could underlie the observed correlation.

5) For each of the findings in Exercise 3, use Figure 11.19 to select a sample size you consider reasonable for replicating the original study. Choosing N is a matter of judgment, but consider where a likely CI might fall in relation to 0. Also consider the CI on the original value or r that you found when answering Question 3 and remember that ρ could easily be anywhere within that interval, or even a little beyond it.

6) Is there really such a thing as beauty sleep? To investigate, researchers decided to examine the extent to which sleep relates to attractiveness. Each of 70 college students self-reported the amount of sleep they had the night before. In addition, a photograph was taken of each participant and rated for attractiveness on a scale from 1 to 10 by two judges of the opposite gender. The average rating score was used. You can load this data set (**Sleep_Beauty**) from the book website.

b. Like you, the researchers obtained a negative correlation. They concluded that those who sleep more are somewhat *less* attractive. You can see, however, that these researchers have made at least one serious error. Explain.

7) Clinton conducted a survey of college students to determine the extent to which well-being is related to campus involvement (**Campus_Involvement** data set on the book website). Participants completed a measure of subjective well-being (scale from 1 to 5) and a measure of campus involvement (scale from 1 to 5). Participants also reported gender (male or female) and commuter status (0 = resident, 1 = commuter).

a. What is the relationship between well-being and campus involvement for commuters? For residents? To what extent is this relationship different for these two groups of students? Interpret.

b. What is the relationship between well-being and campus involvement for men? For women? To what extent is this relationship different for these two groups of students? Interpret.

8) To what extent is analytic thinking incompatible with religious faith? Gervais & Norenzayan (2012) asked participants to complete a test of analytic thinking and a scale of religious belief. Scores on the analytic thinking task were negatively related to religious belief, $r = -.22$, $N = 179$. Later, Sanchez, Sundermeier, Gray, and Calin-Jageman (2016) conducted a close replication of this study using an online sample. They found $r = -.07$, $N = 454$. Use the **Single r** page of **ESCI intro Meta-Analysis** to integrate these two findings. Interpret, considering the result and its CI, and also the diamond ratio. What can you conclude?

? Answers to Quizzes

Quiz 11.1

1) linear; 2) decreases; 3) 1, −1, 0; 4) matched; 5) *z* scores; 6) positive, increase, negative, decrease.

Quiz 11.2

1) d; 2) shorter, longer; 3) random, bivariate normal; 4) very weakly negative to very strongly positive; 5) about four times as many participants: 4 × 10 = 40; 6) shorter, asymmetric.

Quiz 11.3

1) shorter, shorter; 2) causation; 3) No; wealth could be a common factor, enabling both larger chocolate consumption and larger expenditure on science education and research. Indeed Ortega (2013) pointed out that an indicator of wealth (GDP, or gross domestic product) is strongly correlated with both chocolate consumption and amount of Nobel prize success; 4) .1, .3, .5, context; 5) d; 6) d.

Answers to In-Chapter Exercises

11.1 .5, −.4, .8, .3. I prefer quadrants, except for .8, for which tightness to the line is useful.

11.2 b. .1: shotgun blast; −1: points exactly on a line sloping down to the right; Figure 11.9 illustrates $r = .4$; −.4: mirror image of $r = .4$.

11.5 They are all $r = .4$.

11.6 a. Participants 6 and 10, as marked in scatterplot; b. They are the farthest from the center of the cross, and from the two mean lines.

11.8 a. $Z_X Z_Y = 0.1238$; *r* decreases to .88, only a small change because the extra point fits with the previous general pattern; b. *r* changes most when $Z_X Z_Y$ is large and negative, because that influences the battle most; points at top left or bottom right give that large negative contribution.

11.9 It's large and negative, so has an enormous influence in the battle of the quadrants, because other values tend to be small and mainly positive.

11.11 The range of *X* is even more restricted, so it is likely that *r* is even smaller than .13.

11.12 By selecting only college graduates, the range of IQ scores is probably restricted, meaning the correlation between IQ scores and income would be reduced. In the whole population, that correlation may be large, meaning IQ would matter for income.

11.13 a. Up to and including 18 °C: I eyeballed .8 (calculated $r = .83$); b. For 18 °C and above, I eyeballed −.8 (calculated $r = -.85$); c. The effect of range restriction on *r* depends greatly on the relationship, especially if nonlinear.

11.15 a. For $\rho = .5$, in fact for any fixed ρ, larger *N* gives less variation in *r*, meaning a narrower dance.
b. For $N = .40$, in fact for any fixed *N*, the widest variation is for $\rho = 0$ and the dance becomes narrower for ρ approaching 1 or −1.

11.16 a. For a given *N* and *r*, the CI for −*r* is the mirror image of the CI for *r*.
b. For any given *N*, the CI is long and symmetric when $r = 0$, and becomes shorter and more asymmetric as *r* becomes closer to 1 or −1. For larger *N*, the CI is generally shorter, but the pattern of change with *r* is similar. See Figure 11.17.

11.17 The CI gets shorter as *N* increases, whatever the *r*. See Figure 11.18.

11.18 [.31, .61], so the population correlation is most likely around .4 to .5 but may plausibly be as low as around .3 or as high as around .6. The long CI signals considerable uncertainty.

11.19 a, b, c. The *r* values and CIs bounce around. In the long run, 95% of the CIs should include $\rho = -.7$.

11.20 In each column the CIs would bounce around. For $N = 10$ the intervals would be generally longest, and most variable in length. For $N = 160$, shortest and least variable. $N = 40$ would be intermediate. Long run capture rate should be 95% in all columns.

11.21 CI is [.18, .52]. Set $\rho_0 = .5$ and note that $p = .09$. Because $p > .05$, we can't reject that model. Correspondingly, the CI includes .5, but tells us that a wide range of values of ρ, roughly from .2 to .5, are plausibly consistent with the data.

11.22 a. CI is still [.18, .52]. Set $\rho_0 = .25$ and see $p = .23$. Because $p > .05$, we cannot reject the second model either, and .25 is within the CI; b. We need a more precise estimate of ρ to allow us to assess the two models better, so should collect a larger sample of data.

11.23 b. The SD values apply to the sampling distributions of *r* values, so they are standard errors. (Recall that SE is the standard deviation of a sampling distribution, e.g., the mean heap.) For means, SE = (population SD)/$\sqrt{N}$, which fits with the SD values in the table approximately halving as *N* is multiplied by 4. The pattern is reasonable, and the guideline seems to apply, at least roughly, for correlations as well as means.

11.24 The CI on *r* depends on *N* but also on *r* itself, as Figure 11.17 illustrates.

11.25 a. About 0.2; b. No, $N = 200$ gives MoE of roughly 0.15. Need *N* of over 300 (actually, more than 320) for MoE no more than 0.1.

11.26 a. The claim is almost certainly based on a correlation of reported vegetable consumption with school progress. It may be plausible, but quite likely reflects a number of causal relations; b. Perhaps a stable family, good parental education, and parental employment are all associated with both eating more vegetables and doing well at school. It may be possible to assign families randomly to low or high vegetable consumption conditions, then monitor school progress, but quite likely we couldn't run that study long enough to reveal any changes in performance. More complex methods would be needed to tease out multiple causal links.

11.27 a. No, for such a couple, divorce might bring calm to their children's lives and less pressure to use drugs; b. Any individual case is likely to have many aspects to consider, and an overall correlation is only one. Especially for a couple on the brink of divorce, the overall correlation may be of little relevance or even likely to mislead.

11.28 Very feasible; you could take further samples of towns and cities, and measure walking speeds. The results could easily be similar.

11.29 [.72, .97], $p < .001$ using $\rho_0 = 0$. It's reasonable to calculate those because it's reasonable to regard the 15 towns and cities studied as a random sample from the bivariate normal distribution of all towns and cities. The scatterplot in Figure 11.20 shows no strong signs of departure from what would be expected from such a distribution.

11.30 Perhaps people's general sense of urgency (the "pace of life") is greater in cities and prompts faster walking. To investigate, we could seek measures—perhaps of people's attitudes—that assessed the pace of life as directly as possible, and correlate those measures with log population.

11.31 b. We could ask people about their life goals, their immediate concerns, and their sense of urgency, and look for correlations with population size of town or city.

11.32 I would start by trying to understand what "important" might mean to the farmer. Wealthy? Powerful? Educated? From a Bolivian city rather than a village? Then think of correlations that might be informative to investigate.

11.33 Null hypothesis is no difference between the population correlations. The *p* of .44 means we can't reject the null hypothesis at the .05 level, which is consistent with the CI easily including 0 and the conclusion of no evidence of a population difference.

11.34 b. Difference is 0.12 [−0.03, 0.27], shorter but still a fairly long CI. Still no evidence of a difference. The *p* value is .12, so again we can't reject the null hypothesis.

11.35 r_2 = .681 for p = .05. That's a large difference between .41 for women and .68 for men, before obtaining even the weak evidence of p = .05 of a population difference.

11.36 a. Diamond ratio = 2.0, which is large and matches how the diamond appears to change between the two models of meta-analysis; b. There is considerable heterogeneity, which is consistent with very large bouncing of the CIs in the forest plot, including a number of CIs that do not overlap at all.

11.37 The CIs closest to 1 are most asymmetric, as we expect.

11.38 a. The popouts explain that p refers to $\rho_0 = 0$. Nine of 16 studies have $p < .05$, and the overall result is $p < .001$, so we confidently reject the null hypothesis that $\rho_0 = 0$ and conclude there's a statistically significant positive correlation of amount of practice with achievement; b. Macnamara et al. estimated the correlation quite precisely and interpreted the value as indicating a correlation considerably smaller than the theory of Ericsson predicts; c. The estimation approach is more informative, and was used by Macnamara et al. to arrive at a conclusion that corresponds closely with the research question they were investigating.

12

Regression

In Chapter 11 we saw that correlation is a measure of the relationship between X and Y. Like correlation, ***regression*** is based on a data set of (X, Y) pairs, but it's different from correlation in that it gives an estimate of Y for a value of X that we choose. So correlation is a number that summarizes, overall, how X and Y relate, whereas regression takes a chosen single value of X and provides an estimate of Y for that X. Recall that Figure 11.1 was a scatterplot of Well-being (the Y variable) and Body Satisfaction (the X variable), for 106 college students. Figure 11.21 showed the separate scatterplots for women and men. Suppose Daniel scores $X = 3.0$ for Body Satisfaction: What Well-being score would we expect for him, assuming he comes from the same population of college students? We can use regression to estimate Y (Daniel's Well-being score) for $X = 3.0$. There are two steps:

1. Calculate from the data the *regression line for Y on X*.
2. Use that line to calculate an estimate of Y for $X = 3.0$.

Regression focuses on what X can tell us about Y. Almost always, X can tell us part of the story of Y, but not all of it. Informally, the full story of Y divides into two parts:

$$\text{The story of } Y = \underbrace{\text{What } X \text{ can tell us about } Y}_{\text{First part, uses regression}} + \underbrace{\text{The remainder}}_{\text{Second part, what's left over}} \quad (12.1)$$

The informal story of Y.

Regression is thus different from correlation, but the two are intimately linked. We'll see that X makes its contribution to the Y story (the first part) via the regression line, but it's r that determines how large this contribution is. If the correlation is large, X and the regression line give considerable information about Y; if small, they tell only a small proportion of the Y story.

I said that correlation is an effect size measure that has long been routinely reported and interpreted by researchers, which is excellent. For regression the news is even better, because researchers not only report and interpret regression effect sizes, but quite often report regression CIs as well—meaning they are already largely using the new statistics.

Here's the agenda for this chapter:

- The regression line for Y on X: minimizing the standard deviation of residuals
- Regression, correlation, and the slope of the regression line
- The proportion of variance accounted for: r^2
- Regression reversed: the regression of X on Y
- Assumptions underlying simple linear regression
- Confidence intervals and the uncertainty of estimation of Y
- A possibly strange natural phenomenon: regression to the mean