Chapter 5 Correlation

5 CORRELATION

It will be shown how the closeness of co-relation in any particular case admits of being expressed by a single number.
Sir Francis Galton

Today norm-based measurement is the dominant method in psychology for assessing individual differences. In fact, all of our most popular scales (e.g., the Wechsler intelligence tests, the Minnesota Multiphasic Personality Inventory, and the Strong Interest Inventory) use this type of measurement. But the success of this approach depended on additional methodological advances that enabled researchers to evaluate the scales and use them in empirical studies.

First, methods were needed for evaluating how well the scales measured what they were designed to measure—whether the Wechsler Intelligence Scale for Children (WISC) is a good measure of intelligence, for example. Second, investigators needed a method for studying individual differences. Variation between people in personality, social class, education, sex, political beliefs, and culture cannot be studied experimentally, because it is impossible to manipulate such characteristics. These needs were met with the development of a statistical method called "correlational analysis."

This highly mathematical technique was not devised, as you might expect, by a mathematician concerned with the abstract problem of describing the relationship between variables. The initial work on this method was done by Sir Francis Galton, the inventor of norm-based measurement. Galton, who was studying inheritance by breeding peas, was looking for a method to assess how similar parents and offspring are on different traits. The story of his research and of his subsequent invention of the correlation coefficient reveals the close association between correlation and norm-based measurement. Perhaps more than any other single event, the introduction of correlational analysis brought quantitative methods to the social sciences. With correlation, psychology for the first time had a powerful, objective method for observational research.

5.1 The Discovery of Regression

In his initial studies of heredity, Galton compared different generations of people on distributions of physical traits. The comparisons showed that the distributions were surprisingly constant across generations. There was little change in either the means or standard deviations of the scores. Further, fossil records of plants revealed constant distributions of characteristics over thousands of years.

This consistency was puzzling to Galton. He thought that if the physical traits he was observing were highly determined by heredity, they would *not* be constant over generations. If heredity is highly influential, he thought, there would be an increasing standard deviation in traits from generation to generation.

Let's look at Galton's reasoning using height in people as an example. Galton thought that height is determined mostly by heredity; so the children of tall parents should end up as tall adults, and the children of short parents should be short adults. But Galton knew that height is not completely determined by heredity; all children in the same family are not the same height even though they have the same parents.

Most likely, Galton supposed, about half the children in a family grow taller than their parents, and half end up shorter. But if this is true, there should be ever *increasing standard deviations* of the trait in the population. To understand why, consider, as Galton did, two tall people who have children; say, half their children are taller than they are as adults. The tall children marry other tall people and have children, half of whom grow taller than themselves. If this is repeated, generation after generation, we would end up with some very tall people, say, people 30 feet tall, having even taller children! The same process would occur for short parents; half of their children would end up shorter than them. As adults, the short people would have some shorter kids, and so on— until we would have some people, say, 1-foot-tall, having even shorter kids! The increasing numbers of very short and very tall people would dramatically increase the variability of height in the population.

But people have been on earth for thousands of generations and there are no gigantic or teeny-tiny people. Perhaps, thought Galton, children are not equally

likely to be taller or shorter than their parents. He wrote to his cousin Charles Darwin that he "was very desirous of ascertaining the facts of the case." In response, Darwin suggested that he investigate the question using sweet peas.

Galton then designed an experiment to determine the exact relationship between the sizes of parent peas and their offspring. Peas had the advantage over other plants of being self-fertilizing, so each offspring had only one parent.

Galton decided to study quantitative characteristics of peas (of course)—diameter and weight. (At about the same time as Galton's experiments, Gregor Mendel studied qualitative traits of peas, such as tall versus dwarf plants, by interbreeding plants and arrived at an entirely different theory of heredity from Galton's.) Galton selected seeds of seven different, evenly spaced sizes: three below average, one average, and three above average. (He was using Mill's method of concomitant variation here, but he would reach an entirely different type of conclusion than the method would reach.) He picked 10 seeds of each size, for a total of 70 seeds, to form a set. Nine sets were sent to friends in the country with explicit instructions on how to plant them. At harvest time, the plants were sent back to Galton.

5.2 The Concept of Regression

After measuring the sizes of the offspring, Galton classified them into the seven groups defined by the size of their parents. Table 1 shows the *average size* of the offspring with the sizes of their parents for all seven parent sizes. (The results are simplified here to clarify the relationship between parent and offspring.)

Put yourself in Galton's shoes and see if you can see a relationship between the parent sizes and the *average* offspring sizes. Can you state the relationship with a simple principle? Hint: Forget about the absolute size of the peas and think in terms of size measured as a *deviation* from the mean; the average pea had a diameter of 18. Table 1 shows these deviations in parentheses. Can you state the relationship now?

TABLE 1 PARENTS VERSUS AVERAGE OFFSPRING, SIZE IN HUNDREDS OF AN INCH (DEVIATION FROM THE MEAN OF 18 IN PARENTHESES)

	Parent Size		*Offspring Average Size*	
Biggest	21	(+3)	19	(+1)
	20	(+2)	182/3	*(+2/3)*
	19	**(+1)**	18 1/3	(+1/3)
Average	18	(0)	18	**(0)**
	17	**(-1)**	*17%*	*{-1/3)*
	16	(-2)	*17%*	*{-2/3)*
Smallest	15	(-3)	17	(-1)

Galton described the relationship as one of "regression to mediocrity" or *"regression to the mean."*

The average offspring was closer to the average-sized pea than its parent was. Big parents had smaller offspring; tiny parents had bigger offspring. Only average-sized parents had offspring who were the same size, on the average, as their parents.

5.3 THE REGRESSION COEFFICIENT

Galton found that a single number, a fraction, described the regression to the mean. For each of the seven groups of peas, the average offspring deviation from the mean was 1/3 the deviation of its parent. Table 1 shows that if you divide the parent deviation from the mean by 3, in each case you get the average offspring deviation. Galton called this number the *regression coefficient,* and gave it the symbol r, for regression.

This simple result could only be discovered using norm-based measurements on the peas. The regression coefficient relates the *deviation* of the parent from the *mean* to the *deviation* of the offspring from the *mean.*

This result was astounding. Could it really be true that a single number is sufficient to describe the relationship between relatives? Galton wrote:

> *This curious result was based on so many plantings . . . that I could entertain no doubt of the truth of my conclusions. (Galton, 1886, p. 246)*

But people were another story. There were no good data available that could be used to test regression to the mean for people.

Galton immediately began to collect family records of height for two generations. After five years, he had collected about 300 cases of parents' and their children's heights as adults, enough, he judged, for a fair test of regression.

The results confirmed his understanding that the distribution of height changes little from generation to generation. The mean and standard deviation for fathers and sons were virtually identical; the same result was found for mothers and daughters. Now for the critical point. Was there regression to the mean?

5.4 The Scatterplot: The Graph of Correlation

The test for regression followed the same design as the pea study, with one exception: People have two parents, not one; and the offspring are of two types—men and women, each with a different distribution of height; men are taller, on the average, than women.

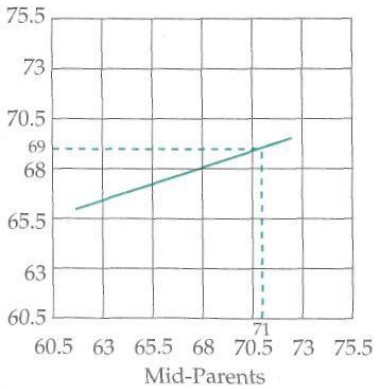
To solve this problem, Galton multiplied the heights of all the women in the study by 1.08, a factor chosen to equalize the average heights of men and women. Then he averaged the parents' heights to obtain a single number called the mid-parent value.

Before describing what Galton found, let's look at the range of logical possibilities for the regression. We can better appreciate his specific result if we first see the full range of possibilities.

Let's start with the degree of regression Galton found for peas, $r = 1/3$ or .33. For this value of r, the children's deviation from the mean would be 1/3 of their

parents' deviation from the mean. Figure 1 shows this regression for height. The graph shows the heights expected for the *average* child, if *r* = .33. The straight line in the graph is called the *regression line.* Using this line, you can determine the *average* height of the children for any height mid-parent (assuming here that *r* = .33). For example, mid-parents who are 71 inches tall can expect children who average 69 inches. This case is shown on the graph.

The regression line shows only the results for the *average* child. It doesn't show the variability in height among the children in the same family (or among children with the same sized parents). This variability can be seen if we plot the individual parent/child cases on the graph with the regression line, as in Figure 2. Sixty cases are plotted there, less than Galton used, but enough to clearly show the results. Each diamond in the graph, 0, is one case and is located by the height of the mid-parent (on the horizontal axis) and the height of their child (on the vertical axis). This type of graph, which is known by a variety of



URE 1
.33

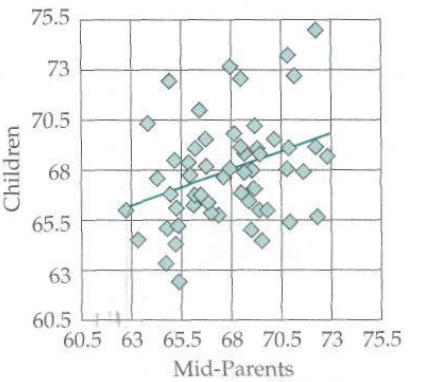


FIGURE 2
r = .33

different names—scattergram, dot diagram, X-Y chart, or *scatterplot*—is still the best technique we have for displaying the relationship between two measures.

> A *scatterplot* is a graph showing, for a group of subjects, the values for each subject on two measures. Each subject is plotted as a point; the point is located by the values of the subject on the two measures. The measures are plotted on the horizontal and vertical axes of the graph.

The scatterplot for *r* = .33 shows considerable variability, or *scatter,* around the regression line. Pick a mid-parent height and look at the wide range of heights of the children. With *r* = .33, there is only weak similarity in the heights of parents and children.

Figure 3 shows the regression line and individual cases for *r* = .66. There is more similarity between parent and child here than for *r* = .33. The plot shows that very tall parents do not have very short children, and vice versa; very short parents do not have very tall kids. There also is less scatter around the regression line than for *r* - .33. Children in the same family would not vary widely in height if *r* = .66.

The next plot (Figure 4) shows how parent and child heights would be related if r = .90. The children would be very similar in height to their parents here; the child deviation from the mean is 9/10 of the parents' deviation, so there is only a small regression to the mean. There also is little scatter around the regression line.

The maximum value for *r* is 1 (Figure 5). (If r were greater than 1, the children would be farther from the mean than their parents; consequently, the standard deviation of height also would *increase* from generation to generation. But since the standard deviation is equal for the parents and children, *r* cannot be greater than 1.) With *r* = 1, every case falls on the regression line, so

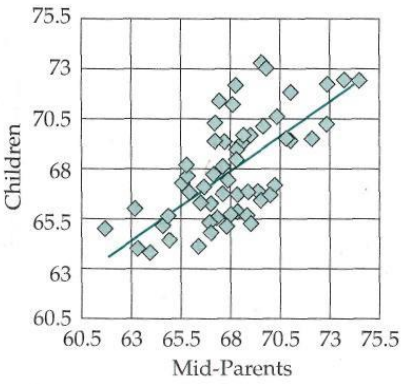


FIGURE 3
$r = .66$

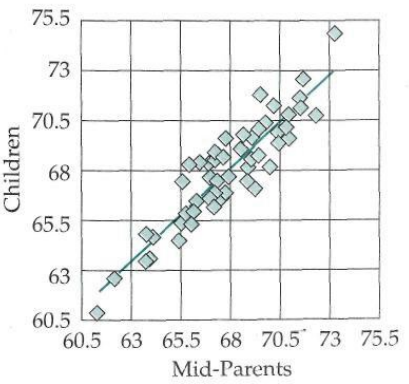
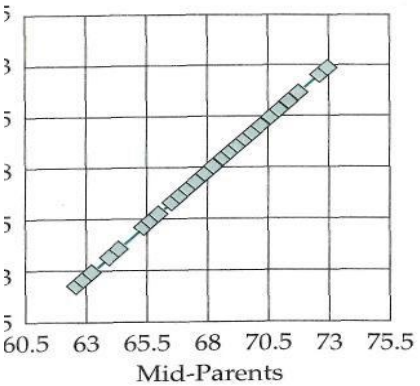


FIGURE 4
$r = .90$



:E 5

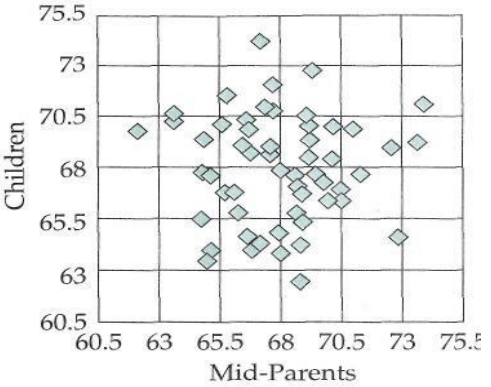


FIGURE 6
$r = 0$

all children are identical in height to their mid-parent. Galton was positive he would *not* find $r = 1$, since he knew that all sons (or daughters) of the same parents are not the same height.

There is one more possibility to examine, which Galton also thought would not happen, $r = 0$ (Figure 6). The regression line is horizontal at the mean in this case; with $r = 0$, the heights of the children are not related to the heights of the parents. Tall parents are just as likely to have short, average, or tall children; the same expectations hold for the children of tall parents as for those of short parents. There is simply no similarity in height between parent and child. Galton was fairly sure he would not find $r = 0$, because casual observation indicated some degree of similarity between parent and child. Now, the actual result. Galton found the regression coefficient was $r = .66$.

On the average, children have 2/3 of their mid-parent deviation. The plot, Figure 3, shows a marked similarity between parent and child. This result was better than Galton could have hoped for. First, his paradox was explained. Regression to the mean of $r = .66$ allows the parent and child to be quite similar in height and still have constancy in the mean and standard deviation across generations. Second, the usefulness of the regression coefficient was confirmed for people. The index was clearly a measure of similarity between relatives; $r - 1$ indicated perfect similarity; $r = 0$ indicated no similarity; values between 0 and 1 could be interpreted as degrees of similarity. The higher the value of r, the less the regression to the mean and the greater the similarity between relatives. Since Galton considered the effects of the environment negligible, for him *r* was an index of biological inheritance.

These results suggested a lifetime of research. Galton and his followers could find the degree of inheritance for all major human faculties and characteristics, for all possible pairings of relatives—child versus parent, grandparent versus grandchild, uncle versus nephews, etc. The research would be time-consuming (and pretty dull), but the results would form the empirical foundation for eugenics.

Galton did not realize the full importance of his regression analysis yet. He thought its use was limited to studies of inheritance—studies examining the relationship between relatives on the same trait. It was only several years later, while he was working on a project unrelated to inheritance, that the real significance of his method suddenly struck him. It was a moment of joy for Galton.

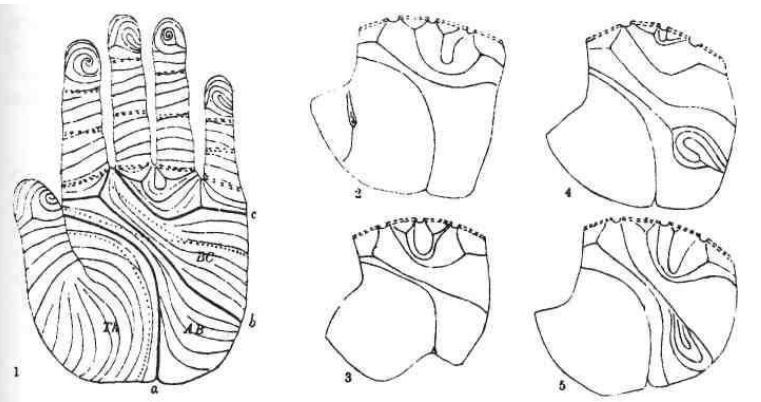
5.5 The Correlation Coefficient

5.5.1 The Independence of Variables

Galton became interested in the problem of personal identity when he was invited to give a lecture on the system of criminal identification developed by Alphonse Bertillon. "Bertillonage" was the only systematic method of establishing personal identity at the time. The method involved careful measurement of different characteristics of the body, such as height, foot length, and head size. Criminologists believed that this set of measurements was sufficient to accurately establish a person's unique identity and guard against false

impersonation. (Fingerprinting, the modern system for establishing identity, was developed by Galton as an alternative to Bertillon's system.)

Fingerprints were a sidelight of Galton's main work on heredity, but a sidelight that gave the occasion for Galton's most fruitful insight—the invention of correlation analysis. This insight occurred while Galton was considering a criticism of Bertillonage.



Galton's fingerprinting system, showing ridges and the creases of the palm. (The Granger Collection, New York.)

According to Galton, the claimed high accuracy of Bertillonage was based on the presumed "independence of the variables measured."

> Two variables are said to be *independent* of each other if the variation in scores on one variable is in no way related to the variation in scores on the other variable.

Galton thought that the accuracy of the method was not as high as claimed because the variables used by Bertillon were *not independent* of each other. He thought, for example, that tall people would most likely have big feet and that short people would have small feet. If this were true, the two measures would not be independent, and including foot size in the system would add little information over knowing a person's height. But how could Galton demonstrate this lack of independence?

5.5.2 z Scores

The problem of comparing variables was related to the heredity problem that Galton had been working on already. Galton had demonstrated that the regression coefficient, r, is basically a measure of similarity; *r* could index the degree of similarity between the heights of fathers and sons, for example. But could *r* be computed between different measures taken on the same person? Does it make sense to ask what is the similarity between, say, head length and height? One measure varies around a mean of 67 inches, while the other varies around a mean of 7.5 inches. How can you say a person's height is identical, or slightly different, or very different from his head size?

This was the question Galton was thinking about while visiting the grounds of Naworth Castle when, in his words:

> *A temporary shower drove me to seek refuge in a reddish recess in the rock by the side of the pathway. There the idea flashed across me, and I forgot everything else for a moment in my great delight. (Galton, 1909, p. 300)*

He had figured out how to compare different measures and determine an index of correlation.

The solution was based on an extension of Galton's norm-based measurement scheme of describing a person's score as a *deviation* from the mean. For example, let's say that Big Joe is 72 inches tall and the average man's height is 67 inches. Then Joe is 5 inches above the mean. If the standard deviation of height is 2.5 inches, then Joe is 5/2.5, or 2, standard deviations above the mean. Next consider Big Joe's head. Let's say it is 8.1 inches long, and the mean head length is 7.5 inches, and the standard deviation of this measure is 0.3 inches. Joe's head length is .6 inches above average; this is 2 standard deviations above average, since the standard deviation is 0.3 (.6/.3 = 2). Now we can compare Joe's height with his head size: Both are 2 standard deviations above average, so Joe's height is identical with his head size! In this special sense, Joe is as tall as his head is long! This comparison of a person's scores on two different measures is made by transforming the scores to a new scale, a scale where the unit of

measure is based on the standard deviation of the measure. Today these new scores are called z *scores:*

> *A z score* is equal to the difference between a score and the mean divided by the standard deviation. A z score expresses how far a score is from the mean in units of the standard deviation.

The formula for computing a z score from an original or *raw* score is

$$z = \frac{X - M}{SD}$$

where, X is the original score, M is the mean of the X scores, and *SD* is the standard deviation of the X scores.

A z score of zero occurs when the X score is equal to the mean; positive z scores occur when X scores are above the mean, and negative z scores when the X scores are below the mean. If a person, like Big Joe, has the same z score on two different measures, then he falls at the same percentile on both measures. Two standard deviations above the mean is at the 98th percentile; so Joe is taller than all but 2% of people, and his head also is longer than all but 2% of other heads.

5.5.3 r, an Index of Correlation

With the z score, Galton could determine the similarity of head size and height. Transforming the original scores to z scores puts both measures on the same scale so that the scores can be directly compared. The value of *r* can then be computed on these scores just as if it were a problem in heredity. (The computation of *r* will be discussed in the next section.) If *r* turned out to be equal to zero, the variables would be independent. Positive values of *r* would indicate similarity between the z scores on the two measures, that is, a lack of independence. Galton found for head size versus height r = +.35, indicating lack of independence. This finding confirmed his criticism of Bertillonage.

Galton called *r* an index of co-relation or *correlation* (the second spelling caught on). In his own words:

> *Two variable organs are said to be co-related when the variation of the one is accompanied on the average by more or less variation of the other, and in the same direction. Thus the length of the arm is said to be co-related with that of the leg, because a person with a long arm has usually a long leg, and conversely. If the co-relation be close then a person with a very long arm would usually have a very long leg; if it be moderately close then the length of his leg would only be long, not very long; and if there were no co-relation at all then the length of his leg would on the average be mediocre. . . .Between these two extremes are an endless number of intermediate cases, and it will be shown how the closeness of co-relation in any particular case admits of being expressed by a simple number. (Galton, in Pearson, 1930, p. 50)*

This passage, the first public presentation of correlation, was read at a meeting of the British Royal Society on December 20, 1888. This was the start of a revolution in research methods.

5.5.4 Calculating r

Galton computed *r* by a graphical method. Today the easiest, most accurate computation is by computer. The computer is programmed to follow the computational formula developed by Karl Pearson, the inventor of the chi-square test we discussed in the last chapter. Pearson developed and extended the mathematical basis of correlation and derived the modern formula for *r.* Because of this work, *r* is now known as the *Pearson correlation coefficient.* In a strange twist of history, Galton's name is no longer linked with correlation, and many scientists now must think it was developed by Pearson!

5.5.5 r and the Reliability and Validity of Measures

The correlation coefficient provides an important tool for evaluating the reliability and validity of norm-based scales. A good example of the application of this statistic is found in the manual for the latest

revision of the Wechsler Intelligence Scale for Children (WISC-III, Wechsler, 1991).

The reliability of the WISC was reported for children in age groups from 6 to 16 years old. For each age group, 200 children were tested. Reliability was studied using the split-half method: The test, composed of numerous items with differing content, was split in half to form, in effect, two IQ tests with similar content. Each subject was scored on both halves of the test and these scores were correlated. A high correlation would indicate that scores on the two halves of the scale are consistent with each other. If the correlation is close to zero, the implication would be that the test is inconsistent and does not even correlate with itself. The split-half correlation was high for the WISC. For example, for 11-year-olds, the correlation was .90. This high value places the test among the most reliable psychological tests available.

The validity of the WISC was studied by correlating WISC scores with other established measures of intelligence and determining the correlations of WISC scores with other variables that should correlate with intelligence for theoretical reasons. If the WISC is valid, the correlation of the WISC and other intelligence tests should be high, close to the reliability of the WISC. The correlation between WISC scores and school grades was expected to be positive but lower than the correlation with other intelligence tests because grades depend on more than intelligence.

The WISC manual reports that the WISC correlates with the Stanford-Binet Intelligence test, $r = .83$, and with mathematics and English grades in school, $r = .41$ and $r = .40$, respectively. These correlations support the validity of the WISC. We will discuss the reliability and validity of measures further in Chapter 12, Planning the Study.

5.6 Correlation and Prediction: The Method of Least Squares

As we have discussed, Galton saw the correlation coefficient as an index of similarity—a measure of the degree of co-relation between different variables. His interpretation still is in common use today, for example, whenever we compute the correlation between two

measures to determine the degree of similarity between them. But there also is a different interpretation of correlation today. This interpretation was discovered by George Yule (1897). Yule, a mathematician working in Pearson's laboratory saw that Galton's and Pearson's work on correlation was a specific case of a general method of analysis that astronomers and physicists had been using for almost a century— the *method of least squares.*

Scientists had used this method to develop mathematical models *to predict* events such as the movement of the moon and planets and the occurrence of high and low tides. The method involved calculating equations to minimize the error in predicting one variable, the dependent variable, from a set of independent variables. Astronomers would take a set of observations of the position of the moon at different times during the year and with the method of least squares calculate an equation to predict the moon's position in the future. The method calculated the equation so it would be the best possible fit to the observations. The method got the name *least squares* because it guaranteed that the fit had the least squared error. The error is the difference between the value predicted by the equation and the observed value.

Yule thought that Galton's problem of describing the relationship between 7 parents and offspring also could be considered as a problem in prediction: How well can you predict offspring characteristics from parent characteristics?

When Yule worked out the least squares solution to this problem, the result was astonishing. The equation for prediction calculated by the method of least squares was the exact equation for Galton's regression line. Yule's calculations also showed that the correlation coefficient, *r,* was a measure of the accuracy of the predictions: r = 1. was characteristic of perfect prediction (zero error) and r = 0 was the worst possible prediction (maximum error).
The equation from the method of least squares for predicting height (based on Galton's data) was

$$Y = 26.8 + .6(X),$$

where Y was the predicted height of the offspring and X was the height of the parent. This is the equation for a straight line, a *linear relationship* between the variables Y and X.

Two variables, *Y* and *X,* are said to have a *linear relationship* if they are related by an equation of the type

Y = a + *bX.*

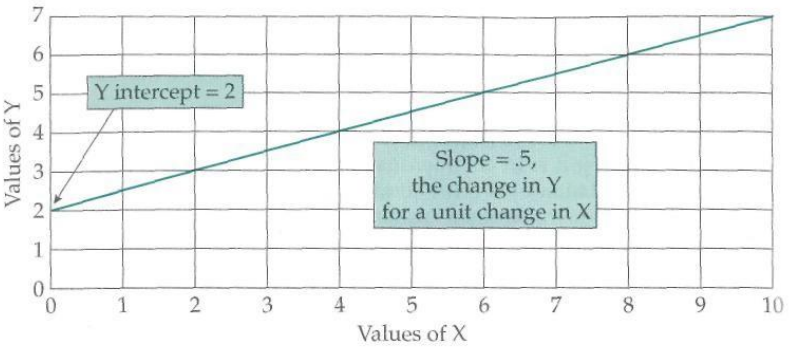


FIGURE 7
Graph of the linear relationship $Y = 2 + .5X$.

The constant *b* is the slope of the line and the constant *a* is the Y intercept of the line. The slope of the line is equal to the change in Y for a change in X; the Y intercept is the value of Y for X = 0, the point where the line crosses the Y axis. Figure 7 shows the slope and Y intercept for the equation Y - *2* + 0.5X.

Yule's work demonstrated that what Galton was doing when he calculated a correlation coefficient was predicting one variable from another variable assuming a linear relationship. If you examine the scatterplots shown earlier in this chapter, you will see that the higher the value of *r,* the closer the data points fall to the regression line.

The Pearson correlation coefficient, r, is a measure of how well one variable, Y, can be predicted from another variable, X, using the linear relationship *Y = a + bX. A* value of *r* = 1 indicates perfect, error-free prediction. A value of *r* = 0 indicates prediction at a chance level.

Yule's insight linked Galton and Pearson's work to an established area of mathematics and led to the development of two methods of enormous importance in the social sciences: (1) methods for studying *nonlinear* relationships, and (2) a new method of *statistical control.*

5.6.1 Nonlinear Relationships

The method of least squares is not limited to linear relationships. With a simple change of the model, nonlinear relationships can be studied. Let's consider the relationship between anxiety and test performance in school. Optimal performance is expected at a middle level of anxiety—anxiety that is neither too high to hinder performance nor too low to motivate the student to do well. This curvilinear relationship is shown in Figure 8. The curve follows the equation $Y = a + bX + cX^2$, where *a, b,* and *c are* constants calculated to minimize errors in prediction. This equation differs from the equation for the straight line by the addition of the term cX^2.

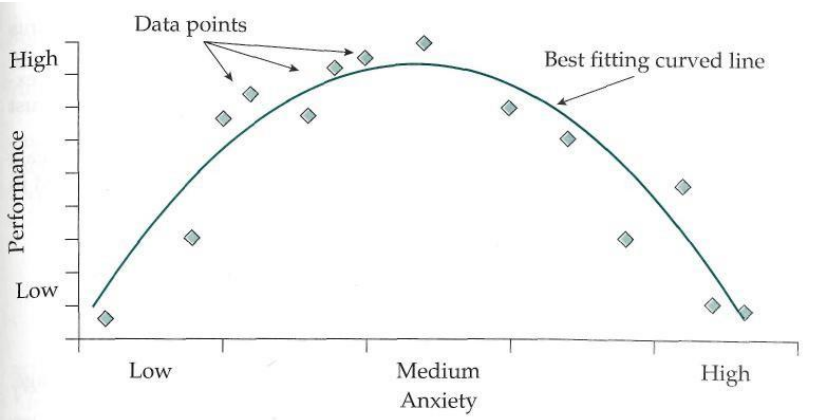


FIGURE 8
Curvilinear relationship between anxiety and performance

With the method of least squares, experimenters can determine the extent to which observed scores match what is expected for a curvilinear relationship. The degree of fit of this model is measured by a generalization of the correlation coefficient that is called the *multiple correlation coefficient.*

This development was important for psychology because nonlinear relationships are common in psychological research. Growth curves, which show how physical and psychological characteristics change

over time, are typically nonlinear. Psychophysical relationships, which show how sensory acuity changes as a function of the magnitude of a physical stimulus, also are nonlinear.

5.6.2 Statistical Control

In Mill's methods, variables are controlled by holding them constant over different conditions in an experiment. But such control is not always possible. In studies of diseases, like cancer, or disorders, like depression or schizophrenia, experiments to find the causes are impossible. Such research must be correlational. The method of least squares provides a powerful tool for analysis in such cases.

Let's take breast cancer as an example. A recent theory holds that exposure to sunshine *reduces* the risk of a woman developing breast cancer (Cowley, 1992). According to the theory, the vitamin D produced by sunshine helps the body absorb calcium which, in turn, controls the growth of cancerous cells. But how can this suspected relationship be studied? One way would be to compare rates of breast cancer in groups of women who get different amounts of sunshine, say, women in Seattle versus women in Florida. But how can we control other variables that we already know are linked with cancer? For example, we know that the risk of cancer increases directly with age, so age must be accounted for in studies of cancer.

Age can be "controlled" in such studies by using the method of least squares in a two-stage process. In the first stage, age is measured, then used in a mathematical model to predict the incidence of cancer:

$$\text{Model 1: } Y = a + b\,(\text{Age})$$

where *Y* is the incidence of cancer, the dependent variable, and Age is the independent variable.

We can fit this model to the data to determine how well cancer is predicted by age. In the second stage, we fit the model:

$$\textit{Model 2:}Y = a + b(\text{Age}) + c(\text{Sunshine})$$

and calculate how much better Model 2 fits the data compared to Model 1. If Model 2 fits the results significantly better, this constitutes evidence that the amount of sunshine is related to the risk of cancer.

Note that the logic here is a variation of the logic of Mill's method of difference. Models 1 and 2 differ in only one independent variable, Sunshine. So if Model 2 fits the data better than Model 1, it must be due to this variable. The variable Age is *statistically controlled* since it is present in both equations.

The sunshine-breast cancer hypothesis was supported by correlational research that used statistical controls for age and other variables linked to cancer. The hypothesis currently is undergoing a more rigorous experimental test by the National Institute of Health. Sixty thousand women are receiving either vitamin D and calcium supplements or a placebo in a 9-year experiment. Because this type of research does not expose subjects to harmful agents, there are no ethical problems to prevent its use with people.

Yule's contribution to research methods was enormous. Least squares is now the most popular method of analysis in psychology. It is used, for example, in studies evaluating how well SAT scores can predict college grades, in studies predicting what kinds of patients will benefit most from psychotherapy, in studies predicting what types of prisoners will violate parole, and in studies of the link between stress and susceptibility to colds. Its applications are limitless.

5.7 CORRELATION'S BAD REPUTATION

We have seen that the correlation coefficient can be interpreted as a measure of similarity between two variables and as a measure of the degree of accuracy in predicting one variable from another assuming a linear relationship. However, sometimes scientists wish to go beyond these interpretations to infer cause-and-effect relationships between the variables being correlated. Scientists can run into problems here.

To understand how, let's try to state the logic of correlation in the language of Mill's methods of induction. As you remember from Chapter 2, Mill's methods have an *if. . . then . . .* format. *If* specific observations are made, *then* you can reach a specific conclusion. Let's try this for correlation:

> *If* you observe each of *n* subjects on two variables, V1: and V2*,* and compute the correlation coefficient, *r,* and find that *r*

> is not zero, *then* you can conclude that V1 and V2 are related by . . .

How should we finish the conclusion? If V1 and V2 are correlated, what exactly does this mean? It is possible that V1 is the cause or part of the cause of V2*;* or vice versa, V2 might be the cause of V1*;* or V1 and V2 both might be caused by a third variable, V3; or V1 and V2 might not be linked by any fact of causation. There is no single conclusion that follows logically from the fact that two variables are correlated. We cannot write a simple ending for our method of induction. This ambiguity in what correlation implies about the relationship between variables is the reason for correlation's bad reputation.

To better understand correlation, we need to be familiar with the variety of relationships between variables that can lead to correlation. Let's start our discussion by looking at how Galton used the relation of causality to interpret his own results on heredity and Bertillonage.

5.7.1 Causal Relationships

Common cause. Galton's first correlations were computed comparing different relatives on the same variable, height for parents, V1 and height for their children, V2*;* and comparing two variables, the sizes of different body parts for the same people, for example, length of the right arm, *V1,* and length of the left arm, V2. In both cases, Galton interpreted the correlations as being due to a common cause; V1 and V2 correlate because both are caused, in part, by another factor that they share. Parents and children share the same genetic material, which determines height. The lengths of both of a person's arms result from the same biological process of growth. This process is shared by both arms.

Common cause is a popular explanation for why variables correlate. IQ test scores correlate with school grades, it is claimed, because both are influenced by the individual's intellectual skills. Two measures of extraversion correlate because both measure the trait of extraversion. However, variables may correlate for other reasons. Instead of sharing a common cause, one variable may directly cause, or at least be involved in the cause, of the second variable.

Direct causation. Suppose we visit a New Year's Eve party. Just after midnight, we interview the revelers to find out how many drinks they have had, V1, and administer a blood test for alcohol, V2, No doubt the correlation between V1 and V2 will be high, maybe r = +.90. People who have had many drinks will have a high concentration of alcohol in their blood; teetotalers will have no alcohol in their blood. Does this correlation indicate a common cause? In other words, is there a third factor that causes people to drink and also causes alcohol to form in the blood? No, not at all—the relationship is simpler; the alcohol you drink goes into your bloodstream. This direct causation is the explanation for the high correlation.

Partial causation. Correlations also are found when the causation is not as direct as in the alcohol example. "Partial causation" or "indirect causes" frequently occur in studies of risk factors for disease and psychopathology Several studies have shown, for example, that separation from a parent before age 11 increases a child's risk of developing depression as an adult. The separation may result from the death of a parent, divorce, or a temporary circumstance.

Separation is called a "risk factor" here because separation before age 11 (yes or no), V1 is correlated with adult depression (yes or no), V2. Approximately 40% of depressed adults report separation from a parent as a child (Roy, 1981); many children separated from a parent do not develop depression, though, and many depressed adults were not separated from their parents as youngsters.

Although separation and depression are related, the correlation in this case does not reflect a common cause or direct causation; separation does not cause depression in itself. The relationship is understood instead as one of partial causation: For some children, separation initiates a chain of events that eventually leads to depression. But why this occurs for some children and not others is a mystery; circumstances other than the separation must be involved.

5.7.2 Noncausal Relationships

Common cause, direct, and indirect causation involve cause-and-effect relationships between variables; but other noncausal relationships can result in substantial correlations as well.

Correlation by chance. Correlations can occur between variables simply by *chance.* Let's say a scientist measures the visual acuity and height of 20 people and calculates *r* as +.60. For this group, the taller people have better eyesight than the shorter people. This correlation could result from bias in the selection of participants for the study; by chance, the group may include a disproportionate number of tall, eagle-eyed, and short, nearsighted people. With a larger, more representative sample of subjects, the correlation might be zero.

This interpretation of a correlation as due to chance can never be entirely discredited, since there is always some possibility, even if very small, of getting a biased sample of subjects. Replicating the result with another group of subjects helps to discredit this explanation. In addition, there is a statistical test that allows the experimenter to investigate the credibility of the chance interpretation. This test calculates the probabilities of getting correlations by chance and uses these results to help decide if the observed correlation is due to chance or reflects a systematic relationship between the variables.

Correlation by custom. Go out on a busy street corner and note for each of the men and women who walk by, $V1_t$*:* M or F, the number of earrings worn, V2: 0, 1, 2 earrings, or more. Make perhaps 200 observations. You can anticipate a substantial correlation between V1 and V2. Men will usually be wearing 0 or 1 earring, women 0 or 2 earrings. Few woman wear 1 earring and few men wear 2 earrings. Is this a causal relationship? Hardly.

This is a correlation due to fashion or custom. Since our culture makes sharp distinctions between men and women, there are many correlations between gender and other variables, such as interests, skills, and values; these correlations reflect the different experiences of men and women in our culture. We can call these examples correlations due to *custom.* Often they are easy to spot as noncausal because the variables form no logical causal chain or have no common cause. How, for example, can gender cause the number of earrings worn? But sometimes it is difficult to identify whether the relationship is causal or noncausal, especially in cases that involve a *common correlate,* as in the next example.

Common correlates. The psychiatrist Alfred Adler maintained that a person's personality was determined, in part, by birth order. According to his theory, firstborn children develop different personalities than later borns, because they have different relationships with their siblings and parents. Adler thought that firstborns were more ambitious, more responsible, more organized, and less peer oriented than people of other birth orders. One prediction, based on these ideas, is that there will be a correlation between birth order, V, and the occupational prestige of a person's job, V_2, with firstborns having the higher prestige positions.

This prediction has been supported. If you were to survey, say, physicians and car salesmen in your town, you would probably find a higher percent of firstborn physicians than firstborn salesmen. Adlerians would interpret this correlation as one of partial causality; in their view, experiences associated with different birth orders set up a chain of events terminating in the individual's employment. But there is another plausible interpretation of the correlation, one that has nothing to do with birth order or personality.

The alternative explanation concerns money, the money it takes to become a physician or lawyer. It is known that families with more money have fewer children than families with less money; families with money, therefore, have a higher proportion of firstborn children than poorer families. (If you have just 2 children, 50% are firstborn; if you have ten children, only 10% are firstborn.) It takes money to go to college, and then on to medical or law school. Put these two facts together and you would expect to find more children of well-to-do parents in medical and law school, and lots of them should be firstborn! Since it takes less money to become a car salesman, there should be fewer firstborns in the showroom.

This alternative hypothesis explains the correlation between V1 and V2 by evoking money as a *common correlate.* If we statistically control money, the correlation between V1 and V2 should be reduced to zero.

Which explanation is right? We don't know. We would have to do further research to find out. And that is precisely the problem with correlational findings. After the fact, you usually can create several plausible, and distinctly different, explanations for any correlation you

find. Does the correlation reflect *common cause, partial cause, chance, custom,* or a *common correlate?* The answer will come only with more studies. This ambiguity in interpretation is behind the derogatory phrase "just a correlational study."

But don't let these problems prejudice you against correlational studies. As we will see in the next chapters, all types of studies have their problems and all require additional research to provide further evidence in support of their conclusions. Although the history of correlational research is short, its successes have been great. Determining the severe consequences of smoking (over 1,000 deaths per day in the United States alone), discovering the role of fluoride in fighting cavities, and demonstrating that rapid eye movements are an index of dreaming, are all classic results of correlational research.

5.8 KEY TERMS

Regression to the mean

Regression coefficient, r

Regression line

Scatterplot

Bertillonage

Independence of variables

z scores

Pearson correlation coefficient

Method of least squares

Linear versus curvilinear relationship

Statistical control

Multiple correlation coefficient

Common cause

Direct causation

Partial causation

Correlation by chance

Correlation by custom

Common correlates

5.9 Key People

Francis Galton

Alphonse Bertillon

David Wechsler

George Yule

Alfred Adler

5.10 Review Questions

1. Why was the constancy of the distribution of physical traits of plants and animals across generations puzzling to Galton?

2. Describe Galton's pea study and summarize its results.

3. Galton was excited about discovering regression to the mean, but he was troubled about its implications for his program of eugenics. Why would Galton find regression troublesome?

4. What were the differences between Galton's study of peas and his study of people's heights? Did Galton find regression to the mean for people?

5. Sketch a scatterplot showing the relationship between parents' and children's heights (as adults) that Galton discovered. Draw the regression line on your plot.

6. Why did Galton think the accuracy of Bertillonage may have been inflated?

7. What would the value of the correlation coefficient be for two independent variables?

8. Describe how Galton used z scores to demonstrate that height and head size are not independent.

9. Explain why Galton's correlation analysis can be considered as a special case of the method of least squares.

10. Describe the difference between a linear and a curvilinear relationship. Give an example of each.

11. Describe ow age was statistically controlled in the study on sunshine and the risk of breast cancer.

12 Explain how statistical control follows the logic of Mill's method of difference.

13. Describe five different relationships that can exist between variables that are correlated. Give an example of each.