

Association and Correlation

1. Let us consider a test of children's ability to write French text correctly from dictation. For simplicity we may assume that the spoken French is standardized by using a tape recorder, that the scoring has been made uniform, and that the scorers can decipher all the handwritings involved. Let

y = the score on this dictation test,

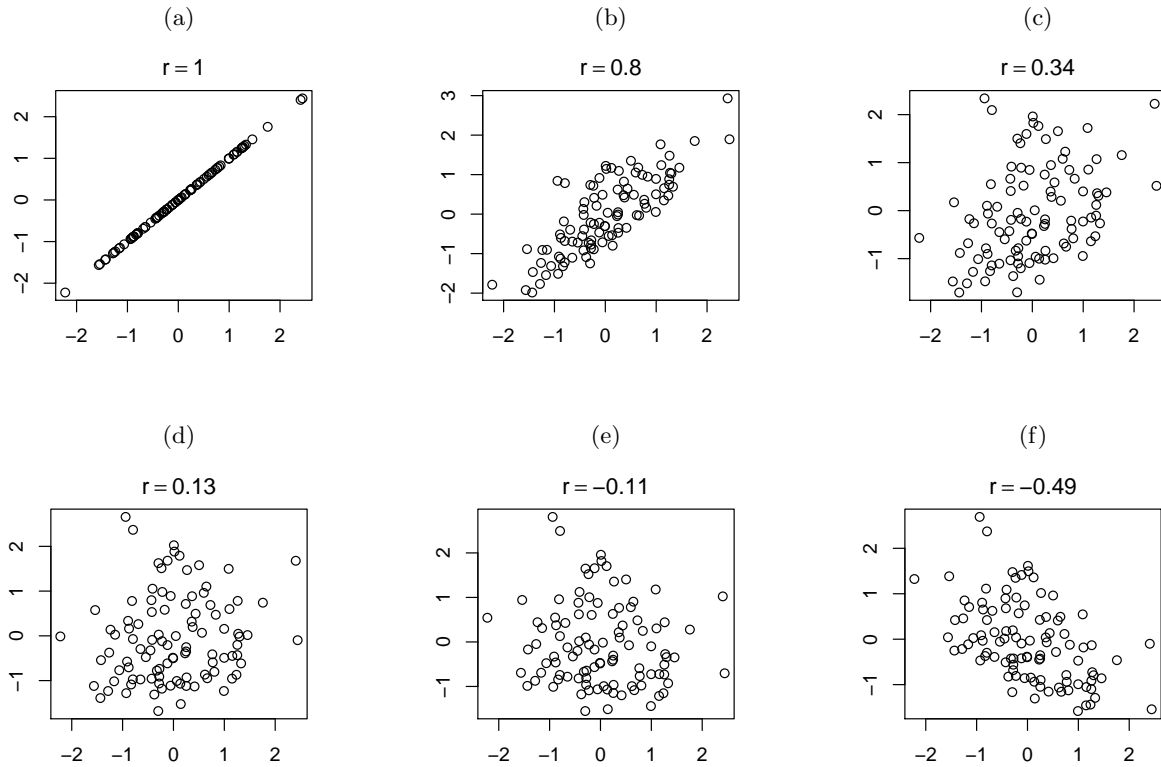
x = the weight of the child.

What is the relation between y and x ?

Solution: The relation depends on the context (population). Consider the following contexts:

- *The ages of the children vary over a wide range, say 5 to 15 years.* In this case, older children will have been in school longer, and so they will perform better on French dictation, at least where French is taught or spoken. Older children also weight more than younger children. Thus, there will likely be a strong positive relation between weight x and dictation score y .
- *The ages of the children are nearly constant, say 15 years plus or minus a few weeks.* Now, there will still be differences between the children, for instance between girls and boys. At this age, girls tend to have better language profficiency than boys do. They also tend to weight less than boys. Thus, it seems likely that there will be a weak negative relation between x and y .
- *The ages of the children are nearly constant, but they come from a mix of countries,* Let's say we have children from France, Holland, and the U. S. A. If the French tend to be lighter than the Dutch, who tend to be lighter than the Americans, then we would get a strong negative relation between weight x and French dictation y .

2. Consider two variables, x and y . In the scenarios below, we have collected $n = 100$ measurements of (x, y) pairs. There is a scatterplot of the 100 points, along with the computed correlation between the two variables. Is there an apparent association between the two variables? Is the relationship positive or negative? Weak or strong?



Solution: Looking at the scatterplots directly is the best way to get a sense of the association. The associations are a little subjective, but here are some plausible judgements: (a) extremely strong positive; (b) very strong positive; (c) moderate positive; (d) negligible; (e) negligible; (f) strong negative.

Causation

3. Which of the three requirements for causation (consistency, responsiveness, and a mechanism) apply to the relation between weight and performance in French dictation?

Solution:

As we established above, consistency likely does not apply: the magnitude and direction of the relationship between x and y likely depends on the context. Also, it doesn't seem like responsiveness applies: there is no way to intervene and change someone's weight. Even if we could (say, by regulating their diet), then there probably won't be a change in their French dictation ability. Finally, there does not seem to be a clear mechanism by which weight determines French dictation ability.

4. Consider a general x and y . How might one confirm the three requirements for causation?

- (a) Consistency?

Solution: Consistency can be confirmed by observation alone. We might look at a variety of different populations and see whether the relationship between x and y is consistent in direction and amount.

- (b) Responsiveness?

Solution: Responsiveness can sometimes be confirmed by experiment. If we can intervene and change x , we can see if y then changes. Note that it is not always possible to do this intervention.

- (c) A mechanism?

Solution: A mechanism can only be confirmed by explicitly describing the mechanism, and supporting the correspondence between each step in the mechanism and that in the process under study.

5. Interest rates and prices tend to be positively associated. Why might there not be a causal relationship between the two?

Solution: Inflation is a common cause of increased interest rates and increased prices.

6. During the industrial revolution in the United States (1750–1850), there was a strong positive association between the amount of scotch whisky imported into New York and the number of ministers there as well. Why might not there have been a causal relationship?

Solution: A common cause of both is the size of the population: as population increased, so did the amount of imports and the number of ministers.

7. There is a positive association between High school seniors' GPAs and their SAT scores. Why might not there be a causal relationship between the two?

Solution: It is implausible that someone's SAT score affects their GPA; the other direction is implausible as well. There are likely many common causes of both of these (e.g. stress level).

8. By now it has been established that smoking causes lung cancer, but establishing this took a long time. People knew that there was a positive association between the two, but this was not enough to establish causation. What are some potential common causes of both? Why is the other direction of the causality implausible?

Solution: There are many potential common causes: occupation, socioeconomic background, gender, ethnicity, and genetics, for example. As for the directionality, R. A. Fisher, one of the greatest statisticians of the twentieth century, proposed that lung cancer causes people's throats to be irritated, and that smoking soothes this irritation. Thus, Fisher proposed that lung cancer could cause smoking!

9. Many people remain unconvinced that there is a causal relationship between global temperature and air pollution. Why do some people dispute that there is a causal relationship?

Solution: Some people argue that despite there being a positive relationship in the last century, this relationship has not necessarily been consistent across this history of the Earth. We do not know for certain that this relationship will hold in the future.