Cautions About Correlations and Causation

by Sophia

WHAT'S COVERED

This lesson discusses the cautions that exist between correlations and causation. You will come to know the limitations of the correlation coefficient. You will also know the difference between variables that are correlated and those that are causally related. This lesson covers:

1. Correlation
2. Causation and Correlation
3. Identifying Causation and Correlation

1. Correlation

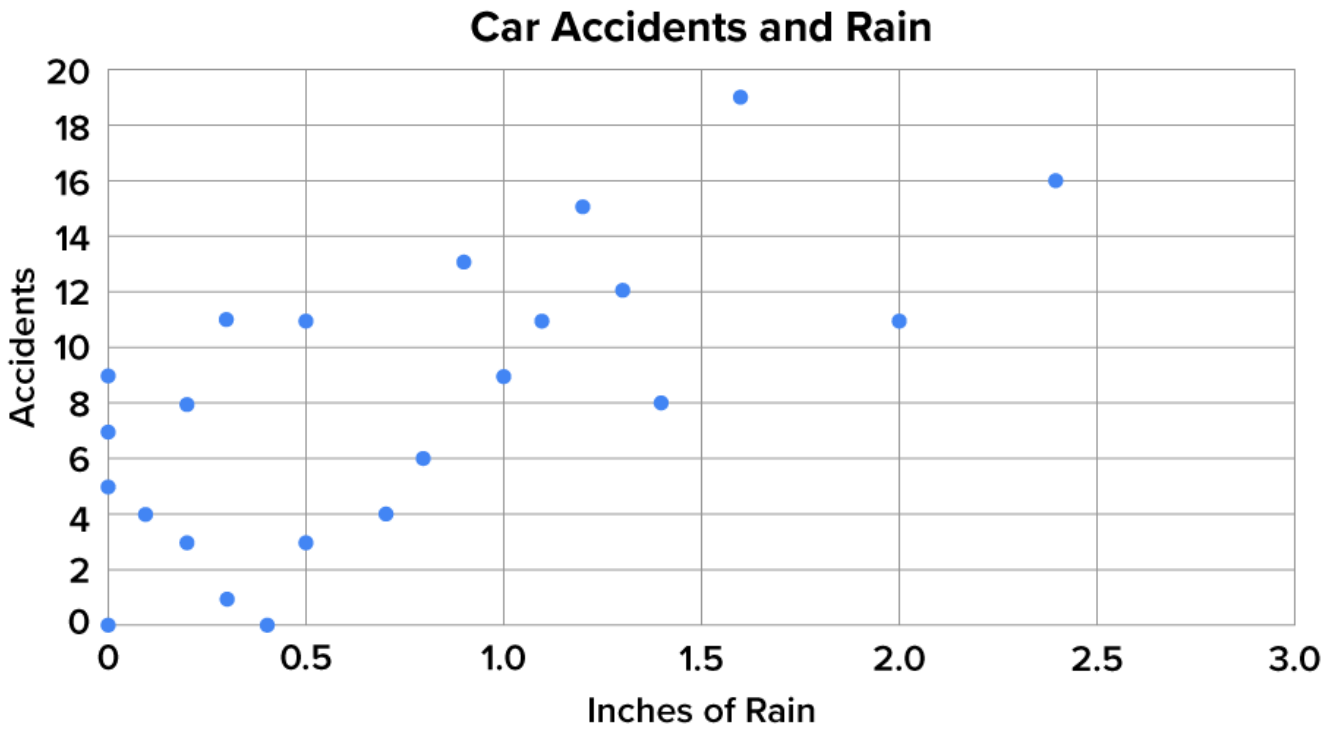
The correlation coefficient is a measurement that explains how well two variables are related in terms of how changes to one variable reflect changes in another. However, the correlation between two variables does not mean that a change in one variable *causes* changes in the other. Two variables are causally related if the change in one of them is responsible for the change in the other.

Correlation does not imply causation. The correlation coefficient indicates whether two quantities are associated, but not necessarily that they are causally linked.

2. Causation and Correlation

This scatterplot illustrates the relationship between how much rainfall a particular location receives on one day and the number of car accidents that occur on the same day.

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.



Can you tell if there's a correlation here? There is. When looking at the scatterplot, notice that while the data points are scattered quite a bit, there is a positive association here. This means the more rain, the higher the number of car accidents. The correlation coefficient is equal to 0.684, which tells you that there's a positive association between the two.

THINK ABOUT IT

Can you necessarily say that rain causes car accidents?

Well, not necessarily. Even though it may contribute to them, you don't know that it's going to be a causal relationship. It is important to recognize whether or not a relationship between variables is correlated or causal. Knowing the difference helps you establish accurate conclusions about a data set.

3. Identifying Causation and Correlation

If you begin fertilizing plants, you'd expect the use of fertilizer would influence the growth rate. The variables are causally dependent. If the data were shown in a scatterplot, there would be a clear trend illustrating a positive association between the use of fertilizer and how fast the plants grow.

You would expect a relationship between the frequency of weightlifting and the amount of added muscle mass to have some causality, and you would expect a positive correlation. If you were to look at a scatterplot, you would see a clear trend illustrating the positive association between the two.

Say you are interested in tracking the presence of germs on an operating room table at a hospital. You would expect a strong correlation between how often it was cleaned and how many germs were identified on it. You'd

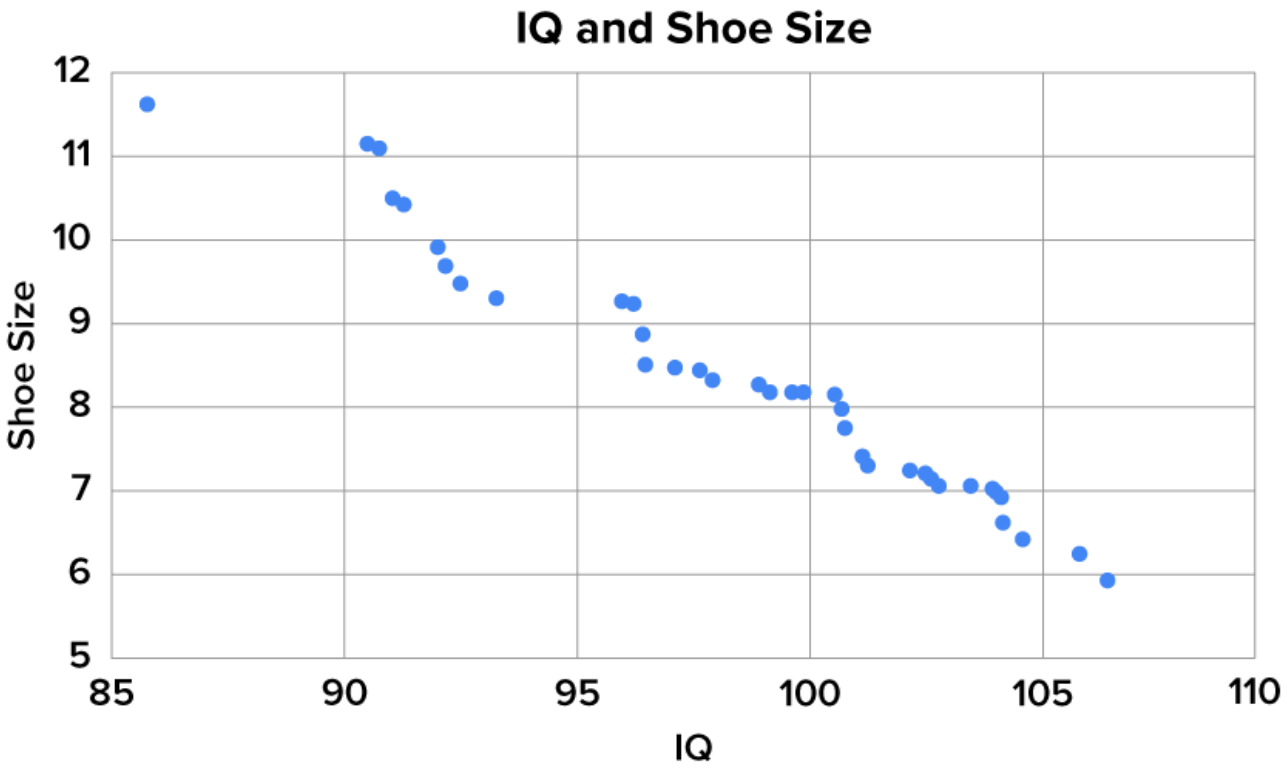
© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

probably expect there to be a causal relationship there as well.

One situation that might not necessarily imply causality would be how often somebody exercises and the amount of fast food that they consume. The results of the analysis of how often a person exercises and how much money he or she spends on fast food would have a clear connection. However, the quantities are not causally related. A change in one cannot be directly linked to the other.

If the data were shown in a scatterplot, a clear trend would indicate that the variables are correlated. In a case such as this, it is likely that a third variable is responsible for the observed correlation. This third variable is the level of the person's concern for his or her health. This would be considered an extraneous variable because it was not accounted for in the initial analysis.

Take a look at an example between two variables that are probably not causally related: shoe size and IQ score.



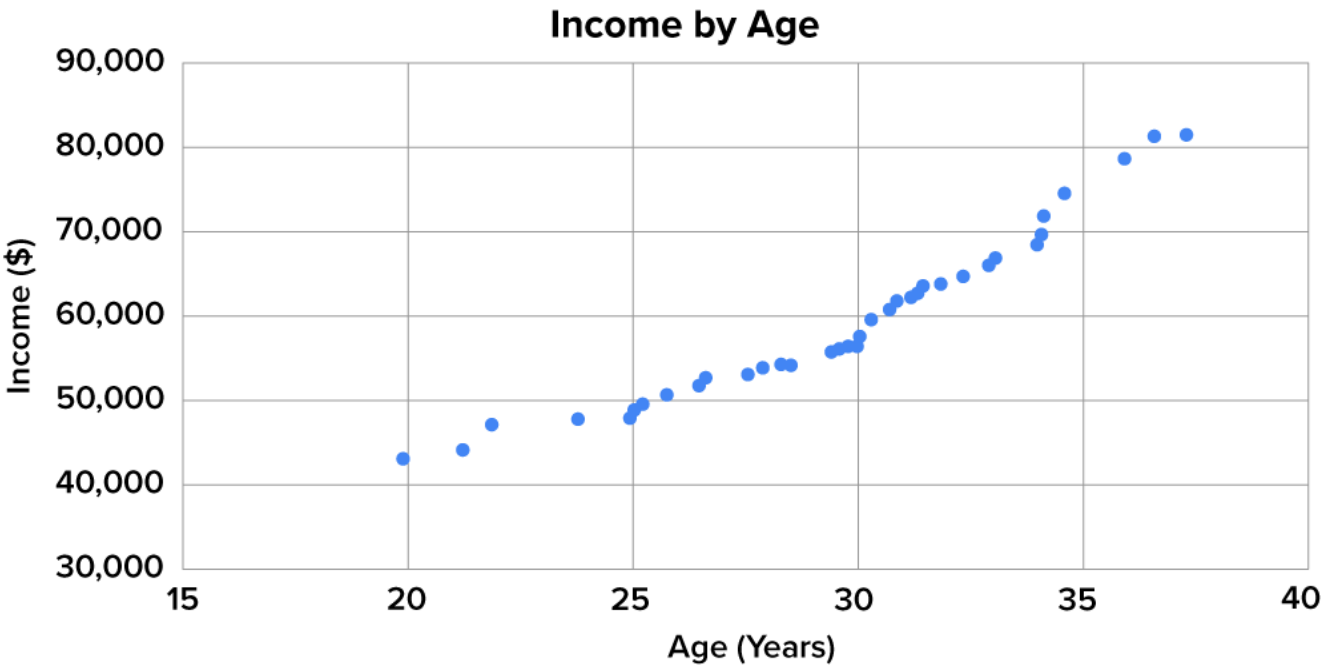
You would think that those with larger feet probably aren't going to have higher IQs, nor would those with smaller feet have higher IQs. However, in this particular instance of 40 random observations, hypothetically this shows a very strong negative correlation.

There's a downward-sloping line, and the points are clustered relatively close together. This tells you that the correlation coefficient is close to -1, indicating a strong negative correlation. But we know that correlation is one thing, and causality is something else.

You would hardly expect there to be a strong correlation between shoe size and IQ score. What you might have here would be an extraneous variable, or a third variable, that's responsible for such a correlation. It might be the testing conditions for the IQ test. That would be a good illustration of an extraneous variable that would affect IQ score, but not necessarily be due to shoe size.

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

The next example is something in which you would figure there *would* be some causation going on: age in relation to income. You would expect that the older somebody is, the more money they are going to be earning.



THINK ABOUT IT

Is it necessarily age that allows that to happen or are other factors at play here?

It's very possible that there are some other extraneous variables involved in this correlation. You might see that something such as education, work experience, or the amount of effort put into the job determines how much someone earns. Just because someone is older doesn't necessarily mean he or she has additional education.

In all likelihood, an older person is going to have a higher level of work experience, which might be the extraneous variable in a case like this. The scatterplot above shows that a positive correlation exists between age and income. The correlation coefficient is 0.958, which is very strong.

It's very possible that the causation is not due to age. It might be simply due to the fact that someone who is older has more work experience than a younger person does. It's important not to confuse causality and correlation because that could lead to wrong conclusions or poor predictions about the variables that you're looking at.

SUMMARY

In this lesson, you learned the limitations of the **correlation** coefficient. There's only so much the correlation coefficient can tell us: that there's a positive or negative association between variables, and the strength of that association. However, this does not necessarily mean that there is a relationship

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.

between **causation and correlation.** You then saw some examples to give you practice **identifying causation and correlation.**

Source: THIS TUTORIAL WAS AUTHORED BY DAN LAUB FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

© 2024 SOPHIA Learning, LLC. SOPHIA is a registered trademark of SOPHIA Learning, LLC.