Cause and Effect

Lesson objectives

- I can distinguish between correlation and causality
- I can identify the type of relationship between two variables

1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 399 #s 1 - 9 & 11 - 14

Definitions

Cause and Effect Relationship

- The correlation between two variables in which a change in one directly causes a change in the other

Common Cause Relationship

- The correlation between two variables in which both variables change as a result of a third common variable

Presumed Relationship

- A relationship that makes sense but does not seem to have a causation factor

Reverse Cause and Effect Relationship

- A relationship in which the independent and dependent variable are reversed

Accidental Relationship

- A relationship that is based purely on coincidence

Data analysis involves much more than fitting a line or a curve to a set of data points. Establishing a linear correlation between a dependent variable and an independent variable is just the first step in understanding the true nature of a relationship. Once you know there is a correlation, it is important to consider how and why such a correlation exists. Finding the meaning behind a linear correlation is what distinguishes true mathematical modelling from simply fitting a line to a set of data.

Example 1

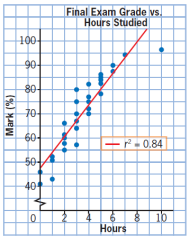
Analyse a Cause and Effect Relationship

The scatter plot and line of best fit show the relationship between the mark achieved and number of hours studied for a grade 12 data management final exam.

a) Does this correlation have a cause and effect relationship?

b) Interpret the line of best fit.

c) According to the linear model, for how many hours must a student study to achieve a perfect score of 100%? Comment on the validity of this answer.



a) r = √0.84 ≈ 0.92 suggesting a strong positive linear correlation. Given the widely held view that increased studying improves exam performance, there is likely to be a cause and effect relationship.

b) The equation of the line of best fit is y = 6.5x + 47. This would mean that if a student studied for 0 hours they would score 47 and for each extra hour of study, their mark would increase by 6.5.

c) Solving the equation for y = 100 gives x ≈ 8.2 meaning that if you studied for 8.2 hours you would score 100. There are limitations with this model though as one student studied for 10 hours and didn't score 100.

$y = 6.5x + 47$

$100 = 6.5x + 47$

$\frac{53}{6.5} = \frac{6.5x}{6.5}$

$8.154 = x$

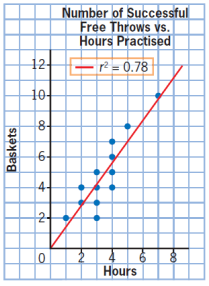
Your Turn

The scatter plot and line of best fit show the relationship between the number of successful free throws made out of 10 attempts and the number of hours spent practising for members of a basketball team.

a) Characterize this correlation with regard to cause and effect.

b) Interpret the line of best fit.

c) Discuss any limitations of this linear model.



a) r = √0.78 ≈ 0.88 suggesting a strong positive linear correlation. Given the widely held view that increased practise improves game performance, there is likely to be a cause and effect relationship.

b) The equation of the line of best fit is y = 1.4x. This would mean that if a player practised for 0 hours they would score 0 free throws and for each extra hour of practise, their number of free throws scored would increase by 1.4.

When analysing linear trends in data, it is important to distinguish between correlation and causality. Just because two variables share a strong linear correlation, it does not necessarily imply that a change in one variable is responsible for a change in the other. Inferring a cause and effect relation based strictly on correlational evidence is one of the most common errors in two-variable data analysis.

Sometimes two variables share a linear correlation because they both depend on another, third variable. This is known as a common cause relationship.

c) This model would suggest that it is possible to score more than 10 out of 10.

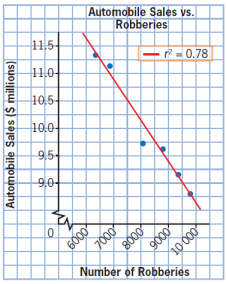
Example 2

Common Cause Relationships

The table and graph show a strong negative linear correlation between automobile sales and robberies from 1988 to 1993.

Year	Number of Robberies	Automobile Sales ($)
1988	6375	11 335 615
1989	6899	11 140 918
1990	8101	9 736 777
1991	9823	8 808 249
1992	9370	9 156 456
1993	8828	9 623 595

Source: CANSIM Table 079-0003, New motor vehicle sales, Canada, provinces and territories, Statistics Canada, April 11, 2014; CANSIM Table 252-0001, Crimes, by actual offences, Statistics Canada, November 15, 2001

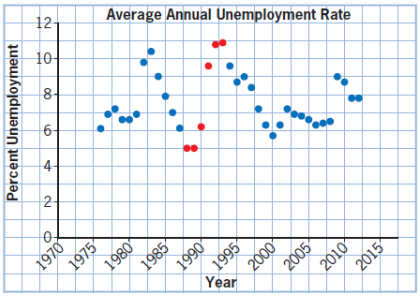


Is it likely that an increase in robberies would cause car sales to drop? Identify a possible common cause for these trends.

It doesn't seem likely that an increase in robberies has led to a decrease in automobile sales. There is likely to be a **COMMON CAUSE**. One possible answer is that during difficult economic times (recession or depression) there is a reduction in spending.

Year	Percent Unemployment	Year	Percent Unemployment	Year	Percent Unemployment
1976	6.1	1988	5.0	2000	5.7
1977	6.9	1989	5.0	2001	6.3
1978	7.2	1990	6.2	2002	7.2
1979	6.6	1991	9.5	2003	6.9
1980	6.6	1992	10.8	2004	6.8
1981	6.9	1993	10.9	2005	6.6
1982	9.8	1994	9.6	2006	6.3
1983	10.4	1995	8.7	2007	6.4
1984	9.0	1996	9.0	2008	6.5
1985	7.9	1997	8.4	2009	9.0
1986	7.0	1998	7.2	2010	8.7
1987	6.1	1999	6.3	2011	7.8
				2012	7.8

Source: Labour Force Survey, Annual Average Unemployment Rate, Statistics Canada



There is also an increase in unemployment which may then lead to an increase in robberies. We can see from the tables that there was an increase in unemployment during the given time period of 1988-1993.

Your Turn

While performing research for his data management project, Aidan discovers a strong linear correlation between the number of forest fires per year and the yield of tomato harvest for the same region. He wonders if growing more tomatoes causes more forest fires. Do you agree with Aidan's line of reasoning? If so, explain why. If not, offer a more likely explanation for this correlation.

Despite the strong linear correlation, it doesn't seem to make sense that growing more tomatoes is going to increase the number of forest fires.

It is more likely that there is a **COMMON CAUSE**. Climate change, and the associated increase in temperature, is a more likely reason for both of these occurrences.

It is important not to jump to a conclusion too quickly when deciding on the type of relationship that exists between two variables. In the example above, a cause and effect relationship was ruled out because it did not make sense. A reasonable case was made for a common cause relationship; however, this still would not necessarily constitute proof.

Variables can be related in different ways. A **presumed relationship** can exist when it seems to make sense that two variables are related, and yet no causality can be inferred and it is also difficult to identify a clear common cause factor. An example of this would be a positive correlation between the number of books in children's homes and their math scores.

A **reverse cause and effect relationship** occurs when the assumed causation becomes reversed. Consider the positive correlation between severe illness and depression. A researcher might hypothesize that being severely ill is very emotionally difficult and so it causes depression. However, in reality, depressed people struggle to take care of themselves, so they are more likely to become severely ill.

Sometimes two variables share a strong linear correlation for no logical reason at all. This is called an **accidental relationship**. Suppose that a positive correlation was found between the local kitten birth rate and the price of eggs. It would be sensible to consider such a correlation purely coincidental.

Example 3

Identifying Types of Relationships

Suggest the most likely type of relationship for each correlation.

a) The number of fire stations in a city is positively correlated with the number of parks.
b) The price of butter is positively correlated with fish population levels.
c) Seat belt infractions are positively correlated with traffic fatalities.
d) Self-esteem is positively correlated with vocabulary level.
e) Charged crimes is positively correlated with the size of the police force.

a) Common cause - Likely to do with an increase of population in the city. This would lead to both an increase in the number of parks as well as an increase in the number of fire stations.

b) Accidental - No obvious reason as to why the price of butter increases when the fish population increases.

c) Cause and effect - Seat belts are designed to save lives, so fewer people wearing seat belts will increase the number of traffic accidents that result with fatalities.

d) Presumed - It would make sense that someone with an increased vocabulary would have higher self-esteem. It is difficult to say one way or the other that one causes the other to happen, and it would also be difficult to find a common cause.

e) Reverse cause and effect - One might think that as crime rates increase, more police officers are recruited. One might also think that as more officers are recruited, more criminals are caught and charged.

Your Turn

Classify the relationships and justify your choice in each case.

a) A patient's stress level is negatively correlated with the amount of exercise performed.

b) Student math scores are positively correlated with English scores.

c) Pancake sales are negatively correlated with amount of rainfall.

d) Job interview success rate is positively correlated with number of years a person has been married.

a) Cause and effect - Studies have shown that an increase in exercise reduces a patient's stress levels.

b) Common cause - No direct relationship between the two although strong study habits will likely lead to stronger scores in both subjects.

c) Accidental - No logical reason as to why the two are connected.

d) Presumed - People in stable relationships tend to have good communication skills and are therefore likely to interview well.

Key Concepts

- A cause and effect relationship exists when one variable is directly responsible for a change in another variable.
- If two variables share a strong correlation, it does not imply that a cause and effect relationship exists.
- A common cause relationship exists when a common third variable is responsible for the correlation between two other variables.
- Several types of relationships can exist between two variables, including cause and effect, common cause, presumed, reverse cause and effect, and accidental.

R2. A positive correlation was discovered with each pair of the following variables:

- cases of the flu
- amount of severe winter weather
- tissue sales

What is the most likely type of relationship involved here? Explain.

Cases of the flu and the amount of severe winter weather most likely have a presumed relationship. Cold weather does not cause the flu.

Cases of the flu and tissue sales most likely have a reverse cause and effect relationship. More tissues are purchased when more people are sick.

The amount of severe winter weather and tissue sales most likely have a common cause effect relationship. More people are sick during the winter and sick people buy more tissues.

R1. Explain why correlational evidence alone does not imply a cause and effect relationship.

Establishing a linear correlation between a dependent variable and an independent variable lets you know that there is a correlation, but does not explain how and why such a correlation exists.

R3. A study found that caffeine intake is positively correlated with nervousness.

a) Suggest a cause and effect relationship that could explain this correlation.

b) Suggest a reverse cause and effect relationship that could be argued.

a) This could be a cause and effect relationship with caffeine intake as the independent variable and nervousness the dependent variable.

b) The reverse cause and effect relationship would be that nervous people are more likely to drink coffee.

R4. **a)** Make up an accidental relationship between two variables.

b) Explain why the correlation is likely to be accidental.

a) The number of females enrolled in undergraduate engineering programs and the number of "reality" shows on television both increased for several years.

b) They are likely to be coincidental.