***Association RQs and Correlation***

Association RQs can be answered with correlation analysis when the variables are interval or ratio level of measurement.

**Examples:**

Is time spent studying associated with GPA?

Is the font size for a website promotional discount associated with sales on the website?

Is the amount spent on a laptop associated with a person’s satisfaction with their laptop?

**Example Problem:**

Is the amount spent on a laptop associated with a person’s satisfaction with their laptop?

Positive association scatterplot:



Negative association scatterplot:



No association:



**Range of the correlation coefficient**:

• correlation can range from -1.0 to +1.0

• a correlation close to 0 means there is no association

**Strength of association:**

Strong association



Weaker association



**Nonlinear relationships**

We can only calculate a correlation coefficient if the relationship between the two variables is linear

Here’s an example of a nonlinear relationship:



Advanced statistical procedures are required to analyze this kind of relationship

**Interpreting correlations:**

1. correlation between income and fast food consumption

r = .44

r = -.52

r = .06

2. correlation between age and number of days per year out of town

r = -.55

r = .04

r = .61

3. correlation between the number of cats in household and frequency of air freshener purchase is .23

4. correlation between sociability and monthly streaming services expenditures is -.41

**Correlation is not causation**

**Example: self-service car wash advertising and sales**

Suppose we find a correlation of .54 between advertising for a self-service car and car wash sales

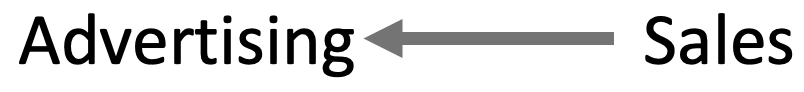
There are three possible explanations:

1. Advertising causes sales to increase



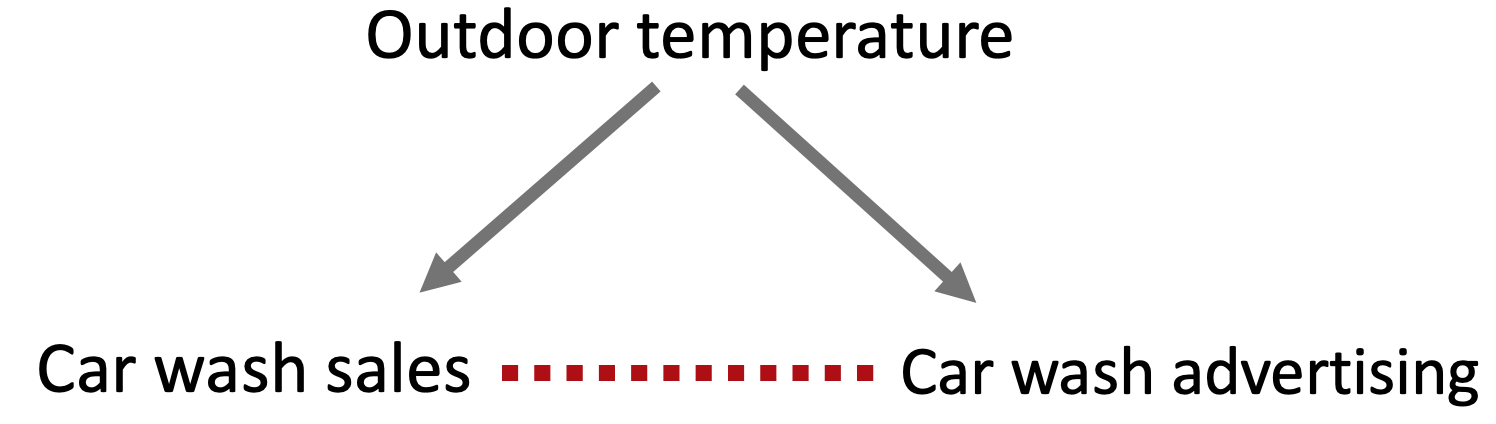
2. Sales cause the advertising to increase (reverse causation)

This might happen because the car wash owner wants to increase awareness through advertising, but doesn’t have enough money to increase advertising until sales go up



3. Some other factor is responsible for both an increase in advertising and an increase in sales—spurious (or fake) causation

In this situation, the car wash owner knows that people wash their cars more often when the weather is warm, so she increases advertising when the weather is warm (so people will think to use her car wash), and decreases advertising when the weather is cooler. This creates a correlation between sales and advertising, but not a causal relationship.



For any correlation analysis, we don’t know which of these three explanations is the correct one to explain an association between two variables. We need additional information.

An advantage of experimental (causal) marketing research is that it allows us to determine whether a causal relationship exists between two variables or not.

***Simple Regression***

Regression is another way to look at relationships between variables.

• if we’re looking at the relationship between two variables, it’s simple regression

• if we’re looking at more variables, it’s multiple regression

In simple regression, one variable is chosen to be the dependent variable; the other is the independent (“predictor”) variable

Both the independent and dependent variables must be interval or ratio measurement

Regression

* examines the association between variables, and
* builds a “prediction” model

**Example:**

A chain of yarn shops is interested in the relationship between advertising and sales

**RQ:** Are advertising expenditures on Friday associated with store traffic on Saturday?

Store traffic was measured on a Saturday in March in 20 randomly chosen stores in the chain. Advertising expenditures for the day before were also tabulated for these stores.

**Scatterplot depicting advertising expenditures and store traffic in our sample**



Regression draws a best-fitting line through the scatterplot

**Scatterplot with regression line**



“Best-fitting” means that the distance between the dots and the regression line is as small as possible

The regression line can be represented mathematically by

* where it intersects the y axis, and
* its slope

Y axis intercept = 65.76 people

slope coefficient = .35

Formula for the line (aka, the regression model):

Traffic = 65.71 people + .35 \* advertising $

Interpreting the regression model:

* When there is $0 advertising on Friday, how many people can we expect to be in the store on Saturday?
* For every $1 increase in advertising, how many more people come to the store?
* For a $10 increase in advertising, how many more people?

**“Prediction” equations**

In regression, the formula for the regression line is called the “prediction” equation

However, prediction equations don’t predict anything in the future

You cannot, based on the regression results, predict that if you increase your advertising by $10 on Friday, 3.5 more people will come into the store on Saturday

Instead, a prediction equation describes what has happened in the past

This is a big source of confusion for people who aren’t statisticians

But, just remember, ***prediction equations don’t predict anything!***

Also, remember that the association between advertising and store traffic revealed by this analysis is ***not a causal association***. (refer back to our original discussion that correlation is not causation)