

# Investigating associations between categorical variables

Tuesday, 26 February 2019 5:55 PM

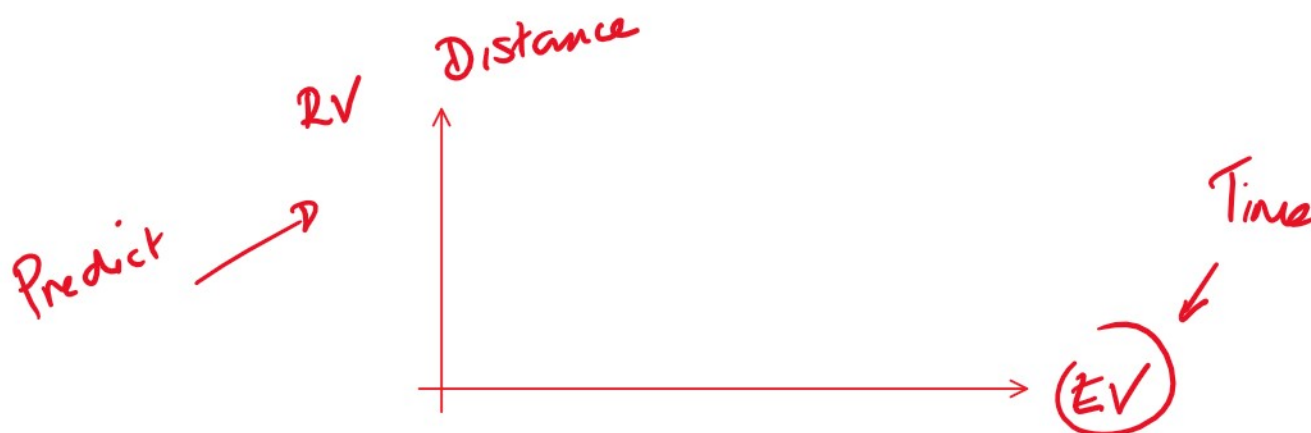
★ By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Know what it means by the term **associated**.
- Know what a two way table is
- Know how to use a two-way table to investigate an association
- Know how to write a short, concise report about associations
- Know how to analysing a situation where two categorical variables have multiple categories

## RECAP:

---

In the previous lesson we looked at the fact that when we have two data sets we might see that there is a link between the two. Time and distance, for example, might be linked when I'm in my car on a journey. Age and Height might also be connected. We looked at the idea of how a graph can be used to plot the data which might help us see a link.



There are other ways we can see if data is linked.

Barry wants us to use certain language to describe things in Mathematics.

## Association

---

When two variables are linked in some way, we say they are **associated**.

We can draw graphs to show associations.

We can also use something called a two-way table.

An example of a question which has been asked to create a two way table is shown below (extracted from the *Cambridge Further Mathematics Units 3 and 4 Textbook*):

**Question: Is a person's attitude to gun control associated with their sex?**

**Table 3.1**

Attitude	Sex	
	Male	Female
For	32	30
Against	26	12
Total	58	42

Total	58	42
-------	----	----

To know if there is an association, we need to decide what the **Explanatory Variable** is and what the **Response Variable** is.

It makes sense to make "attitude to gun control" the response variable and "sex" as the explanatory variable. The table then has strict rules for construction. Columns are Explanatory Variable and Rows are the Response Variable.

Best way to remember: **Row = Response**

**Table 3.1**

		Sex	
		Male	Female
RV	For	32	30
	Against	26	12
Total		58	42

EV

$$\frac{32}{58}$$

$$\frac{12}{42}$$

How do we now know if there is an association?

Currently, reading the raw numbers, we can't really make any sense of the data.

This is primarily because there were different numbers of males and females asked.

So, we use percentages to turn the data into something more relatable.

**Remember: Percent means out of 100**

**Table 3.2**

Attitude	Sex	
	Male	Female
For	55.2	71.4
Against	44.8	28.6
Total	100.0	100.0

Now, we need to interpret the result!

This is the hardest part of the course.

We can see, from the data, that 71.4% of the females are **for** gun control, whereas only 55.2% of males are **for** gun control.

Hence, we can state that attitude to gun control **is** associated with sex.

This can then be written as a "report".

From Table 3.2 we see that a much higher percentage of females were for gun control than males, 71.4% to 55.2%, indicating that attitude to gun control is associated with sex.

## Putting it into practice!

---

### Examples

Extracted from the Cambridge Further Mathematics Textbook Series

#### Example 1

A survey was conducted with 100 people.

As part of this survey, people were asked whether or not they supported banning mobile phones in cinemas. The results are summarised in the table.

EV

RV

Ban mobile phones	Sex	
	Male	Female
Yes	87.9%	65.8%
No	12.1%	34.2%
Total	100.0%	100.0%

Is there an association between support for banning mobile phones in cinemas and the sex of the respondent? Write a brief response quoting appropriate percentages.

#### Answer in the form of a report:

Yes; the percentage of males in support of banning mobile phones in cinemas (87.9%) was much higher than for females (65.8%).

#### Example 2

In the same survey people were asked whether or not they supported Sunday racing. The results are summarised in the table.

Is there an association between support for Sunday racing and the sex of the respondent? Write a brief response quoting appropriate percentages.

EV

RV

Sunday racing	Sex	
	Male	Female
For	55.6%	54.5%
Against	44.4%	45.5%
Total	100.0%	100.0%

#### Note!

We are looking for a significant difference in the percentages to assume there is an association.

#### Answer in the form of a report:

No; the percentage of males (55.6%) supporting Sunday racing is similar to the percentage of females supporting Sunday racing (54.5%).

## What if there are multiple categories?

---

Look at the following example:

Look at the following example:

EV

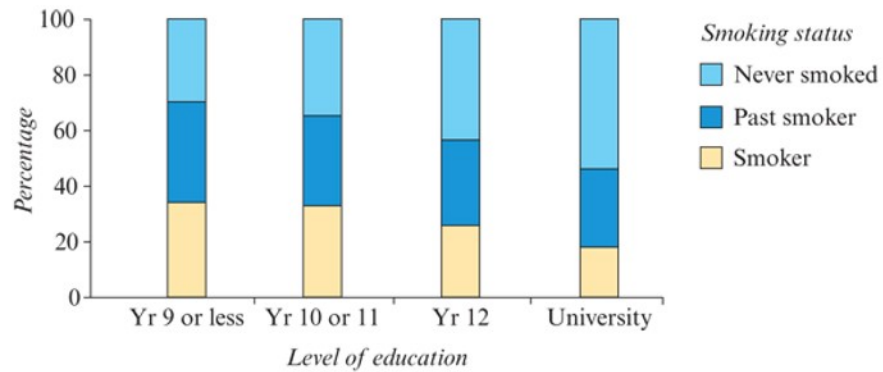
Smoking status	Education level (%)			
	Year 9 or less	Year 10 or 11	Year 12	University
Smoker	33.9	31.7	26.5	18.4
Past smoker	36.0	33.8	30.9	28.0
Never smoked	30.0	34.5	42.6	53.6
Total	99.9	100.0	100.0	100.0

EV

There are two categorical variables, but each variable has multiple categories. What do we do?

We can draw a segmented bar chart which will help us understand what the data is trying to show us. Or just read the data for what it is!

Is there an association between the education level and whether someone smokes?



**Answer in the form of a report:**

From the table, we see that the percentage of smokers steadily decreases with education level, from 33.9% for Year 9 or below to 18.4% for university. This indicates that smoking is associated with level of education.