

Chapter 1 Bivariate Categorical Variables

When dealing with data in everyday situations more than one variable may be under investigation and when this occurs we say that the observations are multivariate.

Our investigations will be limited to situations involving two variables, observations involving two variables are referred to as bivariate.

When examining bivariate data we may be interested in examining each of the two individual data sets using statistical methods studied in UNIT 2. However, we may wish to examine how one variable relates to the other.

The study of bivariate data is concerned with the relationship between the two variables and whether or not there is any connection between them. If there is a connection between the two variables then as mathematicians we need to measure how they are related and on doing so use one of the variables to predict the other. In this section of our course, we shall investigate methods relevant to such questions.

VARIABLES OF BIVARIATE DATA DISTRIBUTIONS

The variables in bivariate data distributions can be both **categorical** or both **numerical** or one categorical and one numerical. In this course we shall deal with situations where the observed variables are either both categorical or both numerical.

Before identifying and describing associations between two variables we will clarify our understanding of variables and their classification.

A **variable** is any quantity or quality that can vary from one subject to another, that is, it can take on different values.

For example:

- the heights of 17 year old boys, the heights of these boys are going to be different for each of the boys under consideration
- mode of transport to work, people can travel by bus, car, train, bike, etc. to get to their place of work
- hotel star ratings, hotels can be rated as being one star, two star, three, four or five star

Variables can be classified as being **categorical** or **numerical**

CATEGORICAL VARIABLES

Variables that represent characteristics or qualities which are described in words are said to be categorical variables. Categorical variables take on values that are names or labels.

For example:

- mode of transport to work
- eye colour
- gender, male - female
- marital status
- the religion of a person
- hotel star rating

Categorical data is further classified as being **nominal** or **ordinal**.

Nominal Categorical Variables

A nominal categorical variable is one that has two or more categories but there is no natural or underlying sense of order to the categories.

Ordinal Categorical Variables

An ordinal categorical variable is one which has two or more categories that have a clear ordering of the variables.

In this chapter we shall focus our attention on bivariate data comprised of categorical variables and in the following chapter we will investigate numerical bivariate data distributions.

Explanatory variable and Response Variable

Before comparing two variables we must first identify each variable as either the explanatory variable or the response variable.

If examination of the variables shows that one variable is affected or influenced by the other variable then the two variables are correspondingly known as the response variable (or dependent variable) and explanatory variable (or independent variable).

A **response variable** is a variable that responds to a change and measures the outcome of a study. It is also called the dependent variable or predicted variable.

An **explanatory variable** is any factor that can influence or explain the response variable. It is often called the independent variable or predictor variable.

Example 1

Identify the explanatory variable and response variable in the following situation.

Does the amount of time spent on doing homework have an effect on the grade a student earns on an exam?

SOLUTION

In this case: Explanatory variable is the amount of time spent on homework.

Response variable is the grade the student earns.

This is because the grade a student earns depends on or is the result of the students studying.

Example 2

Identify the explanatory variable and response variable in the following situation.

Does the gender of a person determine whether or not the person is a vegetarian?

SOLUTION

In this case: Explanatory variable is the gender.

Response variable is whether a person is a vegetarian or not.

This is because whether or not a person is a vegetarian is dependent on the gender.

Example 3

Identify the explanatory variable and response variable in the following situation.

The daily temperature and the number of ice-creams sold.

SOLUTION

In this case: Explanatory variable is the temperature.

Response variable is the number of ice-creams sold.

This is because the number of ice-creams sold will depend the daily temperature.

To help determine which variable is the explanatory variable and which is the response variable if may help to consider one or more of the following:

Ask the question: Which variable explains changes or differences in the other?

Which variable would you use to predict the other?

Which variable occurs first and which occurs next?

Having asked the above questions and still not sure which variable is which you may find the following method useful.

Consider the statement: "The **response variable** depends on the **explanatory variable**."

Replace 'response variable' with one of the variables under consideration and replace 'explanatory variable' with the other variable.

If the statement makes sense you have identified which is the response variable and which is the explanatory variable.

If the statement does not make sense exchange the variables and then the statement should make sense.

Labelling Graph Axes

Before constructing graphs we need to identify the explanatory variable and the response variable for bivariate situations in order to label the graph axes correctly.

When **comparing categorical variables** using a graphical display the explanatory (or independent) variable is best placed on the horizontal axis. The response (or dependent) variable then forms part of the legend.

When comparing **numerical variables** using a graphical display the explanatory variable must be placed on the horizontal axis and the response variable is placed on the vertical axis.

Two-way Frequency Tables

A two-way frequency table deals with two variables, that is, bivariate data. Two-way tables show the relationships between **two categorical variables**. They are used extensively to analyse survey results as they provide a picture of the interrelation between the two variables and can help to find associations between them.

A two-way table shows the distribution of the explanatory variable in columns and the response variable in rows.

The entries in the cells of a two-way table can be frequency counts, relative frequencies, that is proportions of the total number or percentages of the totals.

Consider the following example.

The School Board is considering bringing change to the current school uniform and wishes to know if opinion about the proposed change depends on gender. A random survey of 95 staff and students gave the following results:

- of the 40 females surveyed, 22 were in favour of a change, and
- only 15 of the surveyed males were in favour of a change.

- (a) Construct a two-way frequency table to summarise and display these findings.

The two categorical variables involved in this situation are:

Gender: Male or Female

Opinion on change of uniform: In favour or Against

The School Board wish to know if opinion depends on the gender, hence the variable gender is the explanatory variable and the opinion variable is the response variable.

Now as gender is the explanatory variable it is assigned the columns and the opinions assigned the rows. The information given in the example can now be entered as shown below.

	Male	Female	Totals
In favour	15	22	
Against			
Totals		40	95

Using column and row totals the two-way frequency table can be completed as shown below.

	Male	Female	Totals
In favour	15	22	37
Against	40	18	58
Totals	55	40	95

The column and row totals represent the total number of observations in each column and row and are sometimes called column sums or row sums.

- (b) Give the two-way frequency table entries as percentages of the total.

	Male	Female	Totals
In favour	$\frac{15}{95} \times 100\% \approx 16\%$	$\frac{22}{95} \times 100\% \approx 23\%$	39%
Against	$\frac{40}{95} \times 100\% \approx 42\%$	$\frac{18}{95} \times 100\% \approx 19\%$	61%
Totals	$\frac{55}{95} \times 100\% \approx 58\%$	$\frac{40}{95} \times 100\% \approx 42\%$	100%

- (c) Give the two-way frequency table entries as a proportion of the total number.

	Male	Female	Totals
In favour	$\frac{15}{95} \approx 0.16$	$\frac{22}{95} \approx 0.23$	0.39
Against	$\frac{40}{95} \approx 0.42$	$\frac{18}{95} \approx 0.19$	0.61
Totals	$\frac{55}{95} \approx 0.58$	$\frac{40}{95} \approx 0.42$	1.00

- (d) What percentage of the respondents were male?

Answer: 58% of respondents were male

- (e) What proportion of the respondents were females against the a change in the uniform?

Answer: 0.19 of respondents were females against the change in the uniform.

- (f) What percentage of the respondents were in favour?

Answer: 39% of the respondents were in favour.

EXERCISE 1A

For each of the following situations numbered 1 to 11 state which is the explanatory variable and which is the response variable.

1. A survey was conducted to study whether church attendance can be explained by the gender of a person.
2. The age of a motor car and the asking price for the car.
3. A study investigated whether there was an association between cancer and smoking classification.
4. A survey investigating whether a person's favourite season (Summer, Autumn, Winter, Spring) can be explained by the gender of the person being asked to give their preference.
5. A study wanted to determine whether gender can explain whether a person is a frequent binge drinker.
6. Boys and girls in a Year 2 class were asked "with whom do you find it easiest to make friends?"
7. Performance in a team game and lack of sleep.
8. A survey into whether the type of flight (domestic or international) was associated with arrival time.
9. A survey investigated the age of people that listened to certain radio stations.
10. A survey of Year 12 students from ten colleges on whether or not they were on their college honour board and whether or not they played a sport.
11. A survey investigated whether a person's preferred choice of beverages can be explained in any way by the age of the person.
12. A survey of 300 Year 12 students found that 180 had a driver's licence and 90 had a job. One hundred said they had neither a driver's licence nor a job.
 - (a) Using the given information complete the following two-way table.

	Job	No Job	Total
Had licence			180
No licence		100	
Total	90		300

 - (b) How many of these students do not have a driver's licence?
 - (c) How many of these students do not have a job?
 - (d) What percentage of these students with a job have a driver's licence?
 - (e) What percentage of those without a licence have a job?
 - (f) What percentage of those with a licence have a job?
 - (g) What percentage of students do not have a licence?

13. A medical trial into the effectiveness of a new medication was carried out. 240 adults took part in the trial of which 150 were females. It was found that 80 females and 50 males responded positively to the medication.
- (a) Construct a two-way table showing the results of this medical trial.
- (b) How many men took part in the trial?
- (c) How many women responded negatively to the medication?
- (d) What percentage of these adults responded positively to the medication?
- (e) What percentage of the females in this trial responded positively to the medication?
- (f) What percentage of the adults that responded negatively were males?
- (g) What percentage of the males in the trial responded negatively to the medication?
14. A survey of undergraduate university students revealed that 67 of the 98 females and 88 of the 125 males worked part-time.
- (a) Construct a two-way table showing the results of this survey.
- (b) How many students participated in the survey?
- (c) What percentage of the female students worked part-time?
- (d) What percentage of these undergraduate students worked part-time?
- (e) What percentage of students that work part-time are female?
- (f) What percentage of the male students do not work part-time?
- (g) What proportion of students that do not work part-time are male?
15. A survey of 56 Year 10 students revealed that 48 had a bicycle, 26 have a scooter and 5 have neither a bicycle nor a scooter.
- (a) Construct a two-way table showing the results of this survey.
- (b) How many of these students have a bicycle and a scooter?
- (c) How many of these students only have a scooter?
- (d) What percentage of these students have only a bicycle?
- (e) What percentage of these students who have a bicycle also have a scooter?
- (f) What percentage of these students have neither a bicycle nor a scooter?
- (g) What percentage of students that have a bicycle do not have a scooter?

Percentage Two-way Frequency Tables

Two-way frequency tables can show percentages for the whole table, for columns or for rows.

Using the example considered at the start of this chapter.

	Male	Female	Totals
In favour	15	22	37
Against	40	18	58
Totals	55	40	95

Whole table percentage two-way frequency table:

	Male	Female	Total
In favour	$\frac{15}{95} \times 100\% \approx 16\%$	$\frac{22}{95} \times 100\% \approx 23\%$	39%
Against	$\frac{40}{95} \times 100\% \approx 42\%$	$\frac{18}{95} \times 100\% \approx 19\%$	61%
Total	58%	42%	100%

Row percentages two-way frequency table:

	Male	Female	Totals
In favour	$\frac{15}{37} \times 100\% \approx 41\%$	$\frac{22}{37} \times 100\% \approx 59\%$	100%
Against	$\frac{40}{58} \times 100\% \approx 69\%$	$\frac{18}{58} \times 100\% \approx 31\%$	100%

Column percentages two-way frequency table:

	Male	Female
In favour	$\frac{15}{55} \times 100\% \approx 27\%$	$\frac{22}{40} \times 100\% \approx 55\%$
Against	$\frac{40}{55} \times 100\% \approx 73\%$	$\frac{18}{40} \times 100\% \approx 45\%$
Totals	100%	100%

Each type of percentage table makes a different contribution to understanding the relationship between gender and opinion on changing the school uniform.

Exercise 1B

1. Consider the following two-way frequency table.

	Maths	Science	English
Females	65	72	104
Males	103	92	62

- (a) Use the grid below to construct a table showing each table entry as a percentage of the total.
Give percentages to the nearest tenth of a percent.

- (b) Use the grid below to construct a table showing a column percentaged two-way frequency table.
Give percentages to the nearest whole percent.

- (c) Use the grid below to construct a table showing a row percentage two-way frequency table.
Give percentages to the nearest whole percent.

2. Consider the following two-way frequency table.

	Chocolate	Vanilla	Strawberry	Mango
Children	55	16	62	12
Teenagers	50	29	45	32
Adults	43	57	28	68

- (a) Use the grid below to construct a table showing each table entry as a percentage of the total. Give percentages to the nearest whole percent.

- (b) Use the grid below to construct a table showing a column percentage two-way frequency table. Give percentages to the nearest whole percent.

- (c) Use the grid below to construct a table showing a row percentage two-way frequency table. Give percentages to the nearest whole percent.

3. Consider the following two-way frequency table.

Opinion	Age of respondent in years				
	18 to 29	30 to 41	42 to 53	54 to 65	Over 65
In favour	9	9	25	20	7
Not in favour	15	24	55	61	45
Undecided	20	20	19	16	5

- (a) Use the grid below to construct a table showing each table entry as a percentage of the total. Give percentages to the nearest 0.1%.

Opinion	Age of respondent in years				
	18 to 29	30 to 41	42 to 53	54 to 65	Over 65
In favour					
Not in favour					
Undecided					

- (b) Use the grid below to construct a table showing a column percentage two-way frequency table. Give percentages to the nearest whole percent.

Opinion	Age of respondent in years				
	18 to 29	30 to 41	42 to 53	54 to 65	Over 65
In favour					
Not in favour					
Undecided					

- (c) Use the grid below to construct a table showing a row percentage two-way frequency table. Give percentages to the nearest whole percent.

Opinion	Age of respondent in years				
	18 to 29	30 to 41	42 to 53	54 to 65	Over 65
In favour					
Not in favour					
Undecided					

Association between two categorical variables

Association between two variables means that there is a relationship or connection between them.

To determine if two categorical variables are associated we need to use an appropriately percentaged two-way frequency table.

Normally the explanatory variable categories should be entered in the columns of a two-way frequency table and if this is the case then we need to construct a column percentaged two-way frequency table.

However, if the explanatory variable uses the rows to show its different categories we need to construct a row percentaged two-way frequency table.

Having constructed the appropriately percentaged two-way frequency table we then move across the categories of the explanatory variable and see if the percentages of the categories in the response variable change or are about the same.

If the percentages in the response variable for the different categories across the explanatory variable do not change or are about the same then it is reasonable to say that there is no association between the two variables that define the two-way frequency table.

If the percentages are quite different for at least some of the categories then it is reasonable to say that there is an association between the two variables that define the two-way frequency table.

Constructing a segmented column graph makes it easier to determine if an association appears to exist between the variables.

Let us continue to consider the following example:

The School Board is considering bringing change to the current school uniform and wishes to know if opinion about the proposed change depends on gender. A random survey of 95 staff and students gave the following results:

- of the 40 females surveyed, 22 were in favour of a change, and
- only 15 of the surveyed males were in favour of a change.

As discussed earlier the explanatory variable is gender, the response variable is the opinion on a change of uniform and the two-way frequency table has been constructed with the explanatory variable counts entered in the columns and the response variable counts entered in the rows.

	Male	Female	Total
In favour	15	22	37
Against	40	18	58
Total	55	40	95

Let us now consider the question whether the results of the survey suggest that the opinion on school uniform change may be associated with (or explained by) gender?

In order to answer this question we need to consult the appropriately percentaged two-way frequency table which in this case is the column percentages two-way frequency table as we wish to see if the opinion percentages change as we move across gender.

Alternatively, as the explanatory variable uses columns to show its different categories we need to calculate column percentages.

Column percentages two-way frequency table:

	Male	Female
In favour	27%	55%
Against	73%	45%
Totals	100%	100%

Using the column percentages two-way frequency table and moving across from the male column to the female column the In favour – Against percentages (or proportions) change significantly.

The percentage of people who are in favour of change is much greater amongst the females surveyed (55%) than amongst the surveyed males (27%).

Alternatively, the percentage of male respondents decreases from 73% against the proposal to 45% against the proposal for the female respondents.

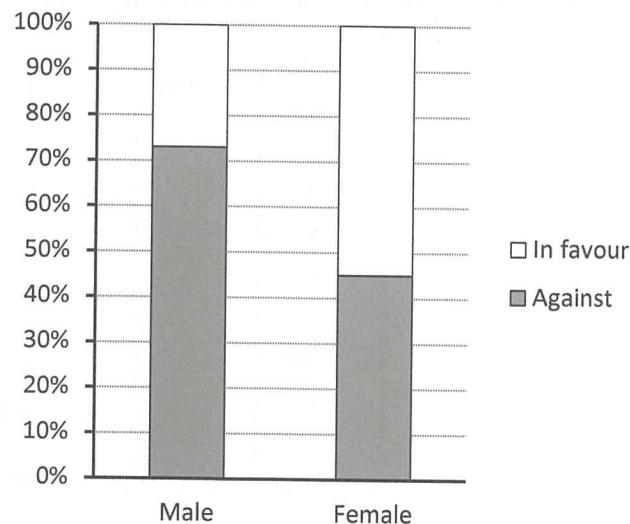
The changing proportions indicate that a strong association exists between the opinion on the proposal of bringing a change in the school uniform and gender.

Segmented Column Graphs or 100% Stacked Column Graphs

Relationships (associations) of the type discussed above are often easier to detect when the percentage two-way frequency table is displayed graphically using a **segmented column graph**.

A segmented column graph has a column for each category of the explanatory variable. For the situation under consideration the explanatory variable is gender and hence a column has been drawn for each category, that is one for males and one for females.

Each column is divided into "segments" such that the length of each segment indicates the proportion or percentage of observations in the response variable. Considering the male column, we can see it has been divided into two segments, In favour and Against and they have been drawn in proportion of the observations.



Examining of the segmented column graph it appears that a higher percentage of females are in favour of a change than males. More men are against the proposal. This suggests that the opinion on this issue may be associated with gender.

- NOTE:**
1. Two-way frequency tables are not always given with the explanatory variable categories shown in the columns. For example bivariate categorical data may have a large number categories in the response variable compared to the explanatory variable and in such cases it makes some sense to use the columns for the response variable and the rows for the explanatory variable categories as it makes better use of page space.
 2. Segmented column graphs or 100% Stacked Column Graphs are sometimes called proportional column graphs.

Example 4

In order to determine whether there was any evidence to suggest that natural hair colour was associated with gender a survey was conducted by hairdressers in a large city. The survey produced the following data regarding natural hair colour of 800 adults.

	Black	Brown	Blonde	Red	Totals
Male	56	136	112	16	320
Female	84	198	174	24	480
Totals	140	334	286	40	800

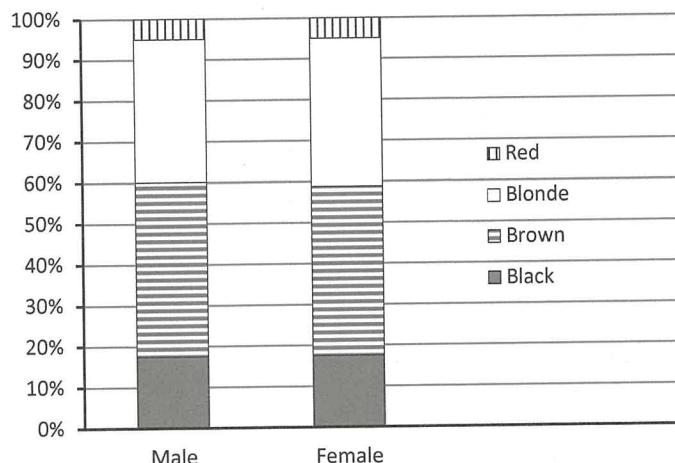
- Which is the explanatory variable and which is the response variable?
- Construct an appropriately percentaged two-way frequency table.
- Use your table to construct a proportional column graph.
- Comment on whether there seems to be an association between gender and natural hair colour.
Explain your reasoning and describe the association.

SOLUTION

- As we wish to see if natural hair colour can be explained by gender the explanatory variable is gender and the response variable is natural hair colour.
- As we wish to see if natural hair colour percentages change as we move across gender we will need to calculate row percentages. Alternatively, as the explanatory variable uses rows to show its different categories we need to calculate row percentages.
Another approach, in this case would be to redraw the table entering the explanatory variable categories in the columns and the response variable categories in the rows and then calculating the column percentages
In this example we shall calculate row percentages as shown below.

	Black (%)	Brown (%)	Blonde (%)	Red (%)	Totals
Male	17.5%	42.5%	35%	5%	100%
Female	17.5%	41.25%	36.25%	5%	100%

- (c) To construct a segmented column graph we label the horizontal axis using categories of the explanatory variable, percentages are entered on the vertical axis and the legend identifies the categories of the response variable.



- (d) As we move across the gender categories there is very little change in brown and blonde proportions and no change in black and red proportions suggesting that there is no association between gender and natural hair colour.

Example 5

Five hundred university students were classified in terms of the personality traits (introvert or extravert) and in terms of colour preference (red, green, yellow or blue) in order to see whether there is an association between these personality traits and colour preference. The data was summarised and is tabulated below.

Personality	Colour Preference				Totals
	Red	Green	Yellow	Blue	
Introvert	42	50		86	
Extravert			45		300
Totals	231			121	

- (a) Complete the table by determining the missing cell entries.
 (b) Which is the explanatory variable and which is the response variable?
 (c) Construct an appropriately percentaged two-way frequency table.
 (d) Comment on whether there seems to be an association between personality and preferred colour.
 Explain your reasoning and describe the association.

SOLUTION

(a)

Personality	Colour Preference				Totals
	Red	Green	Yellow	Blue	
Introvert	42	50	22	86	200
Extravert	189	31	45	35	300
Totals	231	81	67	121	500

- (b) In this situation we wish to consider if colour preference can be predicted or explained by the personality trait, hence the explanatory variable is personality trait and the response variable the colour preference.
 (c) As we wish to see if colour preference percentages change as we move across personality traits hence we will need to calculate row percentages. Alternatively, as the explanatory variable uses rows to show its different categories we need to calculate row percentages.

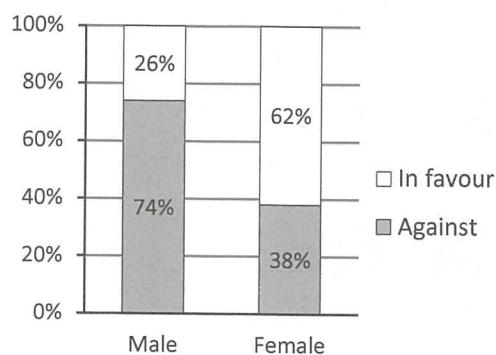
Personality	Colour Preference				Totals
	Red	Green	Yellow	Blue	
Introvert	21%	25%	11%	43%	100%
Extravert	63%	10.3%	15%	11.7%	100%

- (d) The change in the colour preference percentages as we move across the personality traits changes quite dramatically for red and blue and to a lesser but significant extent for green and yellow. These changes in the percentages as we move across the personality traits indicates that there is a strong association between personality and colour preference.

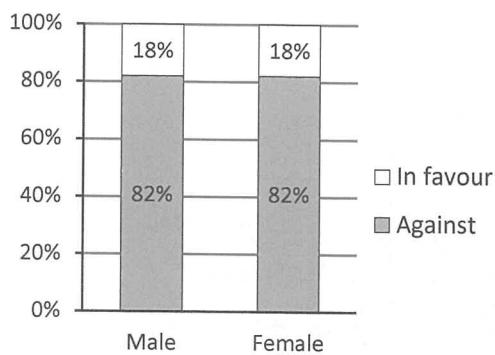
Exercise 1C

The segmented column graphs for questions 1 to 4 show the opinion percentages for different surveys. For each of these surveys comment on whether there seems to be an association between gender and opinion. Explain your reasons and describe the association.

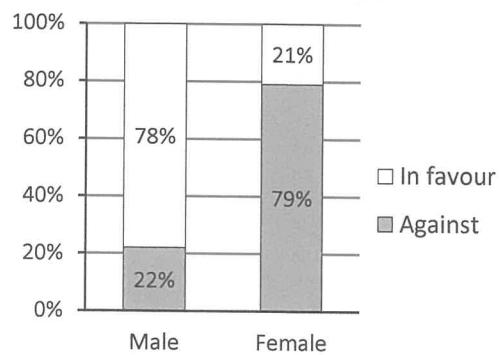
1.



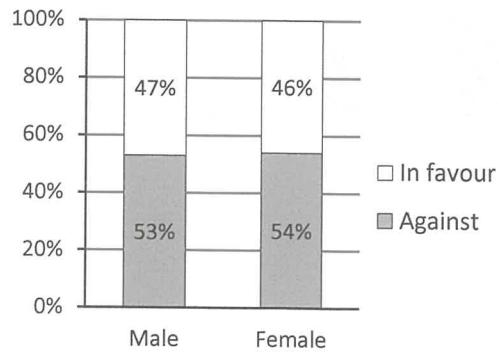
2.



3.

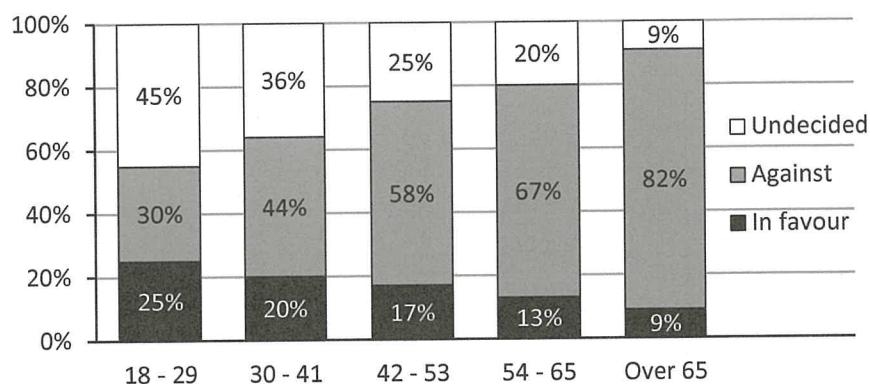


4.

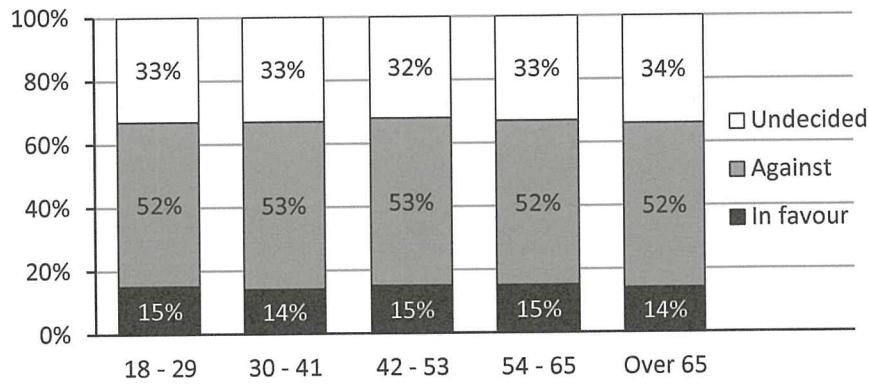


The segmented column graphs for questions 5 to 7 show the opinion percentages for different surveys. For each of these surveys comment on whether there seems to be an association between age bracket and opinion. Explain your reasons and describe the association.

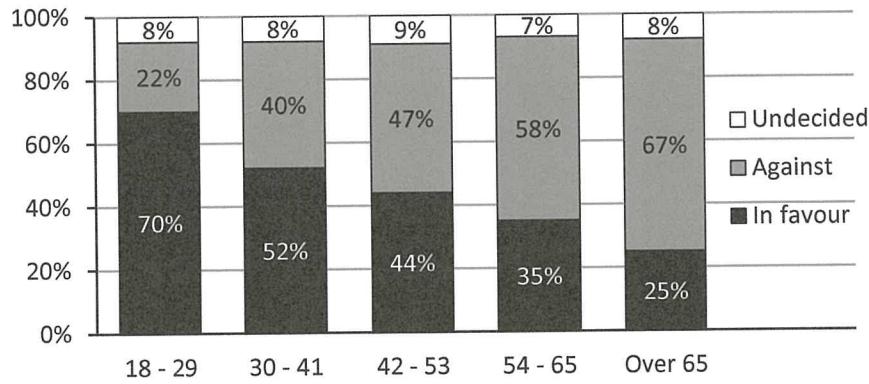
5.



6.



7.



8. A survey was conducted in which 1000 people aged 18 to 56 were asked which was their most preferred type of film out of the following film types: Comedy, Horror, Romance, Western. The survey data was processed and a row percentage two-way frequency table constructed which is shown below.

Age in years	Film type				Totals
	Comedy	Horror	Romance	Western	
18 - 30	21%	11%	10%	58%	100%
31 - 43	46%	13%	5%	36%	100%
44 - 56	63%	12%	20%	5%	100%

- (a) State the explanatory variable and response variable.
 (b) Discuss if there is an association between the two variable identified in (a).

9. The Public Transport Authority surveyed 300 adult train passengers on their opinions on train travel. The survey results are tabled below.

Gender	Positive opinion	Negative opinion	Neutral opinion
Male	61	51	23
Female	61	89	15

- (a) State the explanatory variable and the response variable.
 (b) Which variable in the table above is displayed using columns and which variable is displayed using rows?
 (c) Use the template below to show either the column or row percentages whichever is appropriate.

Gender	Positive opinion	Negative opinion	Neutral opinion	
Male				
Female				

- (d) Construct a 100% stacked column graph that compares the opinions of train travel by men and women.



- (e) Use your graph and/or table to determine whether gender may explain the opinions on train travel by these adults. Explain your reasons and give a description of the association.

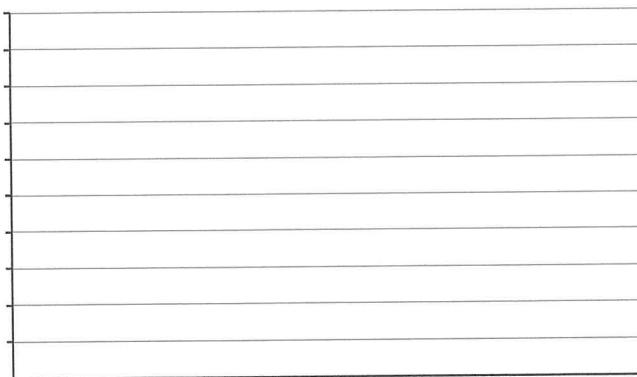
10. A survey was carried out to investigate whether there was any evidence to suggest that the importance of religion in a person's life was associated with gender. The results of the survey were tabulated and are shown below.

	Very important	Fairly important	Not important	Total
Male	43	65	42	150
Female	59	87	54	200
Totals	102	152	96	350

- (a) Determine and state the explanatory variable and the response variable for this survey.
- (b) What percentage of the surveyed males stated that religion was very important in their lives? Give your answer to the nearest tenth of a percent.
- (c) What percentage of the respondents were male and stated that religion was very important in their lives? Give your answer to the nearest tenth of a percent.
- (d) What percentage of the respondents indicated that religion was not important in their lives? Give your answer to the nearest tenth of a percent.
- (e) What percentage of the respondents that indicated that religion was fairly important in their lives were female? Give your answer to the nearest tenth of a percent.
- (f) Should column or row percentages be used to determine whether there is any evidence to suggest that the importance of religion in a person's life is associated with gender. Justify.
- (g) Construct an appropriately percentaged two-way frequency table for the this survey

	Very important	Fairly important	Not important	
Male				
Female				

- (h) Construct a proportional column graph that compares the importance of religion for these men and women.



- (i) Does the data support the theory that gender influences the importance of religion in the lives of these people? Justify you response

11. A public opinion survey explored the relationship between age and support for introducing daylight saving. The survey results are given in the table below.

		Age of respondents in years			
		21 - 40	41 - 60	Over 60	Totals
For	21 - 40	76	27	6	109
	Against	36	36	36	108
Undecided	12	7	4	23	
Totals	124	70	46	240	

- (a) What was the result of the survey?
- (b) What percentage of the 41 to 60 age group voted against the introduction of daylight saving?
- (c) What percentage of the respondents were over 60 and voted against the introduction of daylight saving?
- (d) The majority in which age group were against daylight saving?
- (e) What percentage of the undecided respondents were in the 21 to 40 age group?
- (f) Which is the explanatory variable and which is the response variable?
- (g) Using the template below to create a percentage two-way table that will enable you to identify patterns that may suggest the presence of an association between the variables.

		Age of respondents in years			
		21 - 40	41 - 60	Over 60	
For	21 - 40				
	Against				
Undecided					

- (h) Use your table to construct a segmented column graph.



- (i) Comment on whether there seems to be an association between the variables explaining your reasoning and describing the association.

12. A number of undergraduate university students were asked whether they worked part-time or not. The results by gender are tabled below.

	Male	Female
Worked part-time	70	85
Did not work part-time	32	38

Investigate if the data supports the theory that the gender of a person influences whether these students are engaged in part-time work while completing their studies. You must quote percentages to support your statement.

13. A survey was carried out which asked respondents if they own a gun and whether they are in favour or are not in favour of a law requiring all guns to be registered with a government authority. The survey results are tabled below.

	In favour of gun law	Not in favour of gun law
Owns a gun(s)	329	200
Does not own a gun	1586	281

Investigate if the data supports the conjecture that the ownership of a gun(s) influences whether the respondent is in favour or not in favour of a law requiring registration of guns. You must quote percentages to support your statement.

CHAPTER ONE REVIEW EXERCISE

1. Of the 350 children enrolled at ABC Primary School it was found that 60% had been immunised against contracting whooping cough. Also it was found that 30% of those who were immunised were boys and 50% of those who were not immunised were girls.
- (a) Identify and state the explanatory and response variables.

- (b) Display the given information in the following two-way frequency table.

	Boys	Girls	Totals
Immunised			
Not immunised			
Totals			

- (c) How many girls were enrolled at ABC Primary School?
- (d) How many girls were immunised against whooping cough?
- (e) What percentage of the boys were immunised against contracting whooping cough?
- (f) What percentage of these children were not immunised against contracting whooping cough?
- (g) Construct a percentaged two-way frequency table showing either column or row percentages as appropriate, rounding percentages to one decimal place

- (h) Construct a suitable graph to display the information stored in your percentaged two-way frequency table.



- (i) Is there an association between immunisation for whooping cough and gender.

2. High school students were surveyed with respect to the mode of transport used to get to school. It was found that of the 300 students surveyed 55% were female. Forty percent of the students surveyed came by car and of these 30 were male. Thirty percent used the bus and of these 70% were female. Twenty percent of these students rode a bike and only 4 females walked to school.

(a) Construct a two-way frequency table for the given information.

	Walk	Bike	Bus	Car	Totals
Male					
Female					
Totals					

(b) What percentage of students that came to school by car were female?

(c) What percentage of these students were females that used the bus to get to school?

(d) What percentage of the female students used the bus to get to school?

(e) Use the grid below to construct either a column or row percentaged two-way frequency table, whichever is appropriate to investigate if there is an association with gender and mode of transport used by these students to get to school.

(f) Use your table to construct a segmented column graph.



(g) Comment on whether the gender of these students might explain the mode of transport used to get to school. In your comment identify the explanatory and response variables.

3. Mr Andrews compared the examination results of all students studying Applications Unit 3 at his college with whether the students had completed last year's examination paper as part of their revision program. He prepared the following two-way table.

		Last year's examination paper	
Examination results		Completed paper	Did not complete paper
Passed		64	9
Did not pass		7	14

- (a) State the explanatory variable and the response variable.
 - (b) How many students studied Applications Unit 3?
 - (c) What percentage of the students completed last year's examination paper and passed the examination?
 - (d) What percentage of students that passed the exam did not complete last year's paper?
 - (e) What percentage of the students that completed last years paper did not pass the examination?
 - (f) Use the table template below to construct an appropriately percentaged two-way table to enable you to identify any patterns that may suggest the presence of an association between the variables.

Last year's examination paper			
Examination results	Completed paper	Did not complete paper	
Passed			
Did not pass			

- (g) Draw a segmented column graph to identify any patterns.



- (h) Comment on whether there seems to be an association between completing last year's examination and passing the examination.

(i) Given that an association between completing last year's examination paper and passing the examination exists, would this imply that completing last year's examination paper will improve examination scores for all students studying this subject? Explain.

4. A study was carried out to investigate whether there was any evidence to suggest that the likelihood of a woman developing cancer of the womb was associated with oestrogen use over time. The study produced the following results.

State of womb	Duration of oestrogen use			
	Never used	Under 1 year	1 to 5 years	More than 5 years
Healthy womb	345	112	67	21
Womb cancer	239	175	131	238

(a) State the explanatory variable and the response variable.

(b) Use the following template to construct a table showing either column or row percentages as appropriate to identify patterns that may suggest the presence of an association between the variables. Give percentages to the nearest whole percent.

State of womb	Duration of oestrogen use			
	Never used	Under 1 year	1 to 5 years	More than 5 years
Healthy womb				
Womb cancer				

(c) Construct a proportional column graph with equal height columns each representing an appropriate category.



(d) Comment on whether there seems to be an association between the variables.