

## ARMSPAN V'S HEIGHT

## This fortnight we are going to:

- review the statistical investigation process: for example, identifying a problem and posing a statistical question, collecting or obtaining data, analysing the data, interpreting and communicating the results.
- construct two-way frequency tables and determine the associated row and column sums and percentages
- use an appropriately percentaged two-way frequency table to identify patterns that suggest the presence of an association
- describe an association in terms of differences observed in percentages across categories in a systematic and concise manner, and interpret this in the context of the data
- construct a scatterplot, parallel dot plots, and divided bar charts to identify patterns in the data suggesting the presence of an association
- describe an association between two numerical variables in terms of direction (positive/negative), form (linear/non-linear) and strength (strong/moderate/weak)
- use a scatter plot to identify the nature of the relationship between variables


## Theoretical components

## Practical components

## Resources:

## PDF file: Week 1/2 Notes \& Exercises

The clip introduces bivariate data https://www.youtube.com/watch?v=dryFiDCZmd4 Dependent vs Independent variables
https://www.youtube.com/watch?v=aeH1FzqdQZ0
The clip introduces scatterplots
https://www.youtube.com/watch?v=G6Edu RybxA
A further explanation of scatter plots and correlation is given by
https://www.youtube.com/watch?v=CWnfwZRAuaY\& disable polymer=true

## Order

1. Read through the notes and examples
2. Work through the exercises
3. Complete the investigation at the end of the booklet.
4. Complete the reflection at the end of the booklet
5. Come and see your teacher and make sure you are up to date.

Work through the exercises and show the completed tasks to your teacher.

Be sure to ask for help as you need for the successful completion of all tasks.

Remember to regularly check Google Classroom for messages.

## Knowledge Checklist

- Bivariate data
- Two-way frequency table
- Parallel dot plots
- Divided bar charts
- Scatter plots


## Investigation

Complete the task at the end of the booklet and submit your work for checking. (:)

## MATHEMATICAL APPLICATIONS 3

## WEEK 1/2 NOTES \& EXERCISES

## Bivariate Data

A manager of a small ski resort has encountered a problem. She wants to be able to predict the number of skiers using her resort each weekend in advance, so that she can organise additional staffing catering if needed. She knows that good deep snow will attract lots of skiers, but shallow covering is unlikely to attract a crowd. To investigate the situation further, she collects the following data over twelve consecutive weekends at her resort.


As there are two types of data in this example, this is known as bivariate data. For each item (weekend), two variables are considered (depth of snow and number of skiers). When analysing bivariate data, we are interested in examining the relationship between the two variables.

In this case, the manager might be interested in answering the following questions:

1. Are visitor numbers related to depth of snow?
2. If there is a relationship between visitor numbers and depth of snow, is it always true or is it just a guide? In other words, how strong is the relationship?
3. How much confidence could be placed in the prediction?

In a relationship involving two variables, if the values of one variable 'depend' on the values of another variable, then the former variable is referred to as the dependent variable and the latter variable is referred to as the independent variable. When a relationship between two sets of variables is being examined, it is important to know
which one of the two variables depends on the other. Most often we can make a judgement about this, although sometimes it may not be possible.

Consider the case where a study compared the heights of company employees against their annual salaries. Common sense would suggest that the height of a company employee would not depend on the person's annual salary nor would the annual salary of a company employee depend on the person's height. In this case, it is not appropriate to designate one variable as independent and one as dependent.

In the case where the ages of company employees are compared with their annual salaries, you might reasonably expect that the annual salary of an employee would depend on the person's age. In this case, the age of the employee is the independent variable and the salary of the employee is the dependent variable. It is useful to identify the independent and dependent variables where possible since it is the usual practice when displaying data on a graph to place the independent variable on the horizontal axis and the dependent variable on the vertical axis.


The Australian Bureau of Statistics conducts real life statistics on different aspects of our lives to provide various government departments with information about the general population. We can also use this information to make informed decisions and predict what could happen in the future.

## Example

For each of the following pairs of variables, identify the independent variable and the dependent variable. If it is not possible to identify this, then write 'not appropriate'.

1. The number of visitors at a local swimming pool and the daily temperature
2. The blood group of a person and their favourite TV channel

## Solution

1. It is reasonable to expect the number of visitors at the swimming pool on any day will depend on the temperature on that day.

Daily temperature is the independent variable and number of visitors at a local swimming pool is the dependent variable
2. Common sense suggests that the blood type of a person does not depend on the person's TV channel preferences or vice versa.

Not appropriate

## EXERCISE 1

1. For each of the following pairs of variables, identify the independent variable and the dependent variable. If it is not possible to identify this, then write 'not appropriate'.

|  | Independent | Dependent |
| :--- | :--- | :--- |
| The age of an AFL <br> footballer and his annual <br> salary |  |  |
| The growth of a plant and <br> the amount of fertiliser it <br> receives |  |  |
| The number of books <br> read in a week and the <br> eye colour of the readers |  |  |
| The voting intentions of a <br> woman and her weekly <br> consumption of red meat |  |  |
| The month of the year <br> and the electricity bill for <br> that month |  |  |
| The mark obtained for a <br> Maths test and the <br> number of hours spent <br> preparing for the test |  |  |
| The mark obtained for a <br> Maths test and the mark <br> obtained for an English <br> test |  |  |
| The cost of grapes (in <br> dollars per kilogram) and <br> the season of the year |  |  |
| Ticket sales and revenue <br> of show |  |  |

## TWO-WAY FREQUENCY TABLE

When we are examining the relationship between two categorical variables, the twoway table is an excellent tool as it allows us to have a clear breakdown of the data.

## Example

At a local shopping centre, 34 seniors and 23 non-seniors were asked which of the two major political parties they preferred. Nineteen seniors and 12 nonseniors preferred Labor. Display these data in a two-way frequency table.

## Solution

Raw Data:

| Party <br> preference | Seniors | Non- <br> Seniors | Total |
| :--- | :---: | :---: | :---: |
| Labor | 19 | 12 | 31 |
| Liberal |  |  |  |
| Total | 34 | 23 | 57 |

Completed table:

| Party <br> preference | Seniors | Non- <br> Seniors | Total |
| :--- | :---: | :---: | :---: |
| Labor | 19 | 12 | 31 |
| Liberal | 15 | 11 | 26 |
| Total | 34 | 23 | 57 |

This shows a clear breakdown of the data in terms of numbers. It shows more seniors prefer Labor ( 19 seniors vs 12 non-seniors). However, only 23 nonseniors were surveyed comparted to 34 seniors. Using percentages helps overcome this.

The table is filled in by expressing the number in each cell as a percentage of the column's total. For example, to obtain the percentage of non-seniors who prefer Labor, divide the number of non-seniors who prefer Labor by the total number of non-seniors and multiply by 100. Percentage is $52.2 \%$ ( 1 decimal place).

| Party <br> preference | Seniors | Non- <br> Seniors |
| :--- | :---: | :---: |
| Labor | 55.9 | 52.2 |
| Liberal | 44.1 | 47.8 |
| Total | 100.0 | 100.0 |

Similar percentages of seniors and non-seniors preferred Labor.
However, we could also find the percentages of those who preferred Labor were senior or non-senior.

Senior: $\frac{19}{31} \times 100=61.3 \%$ of those preferring Labor were seniors.
Non-senior: $\frac{12}{31} \times 100=38.7 \%$ of those preferring Labor were non-seniors.
The general rule is that the independent variable (respondent's age category) is placed in the columns of the table and the percentages should be calculated in columns.

Comparing percentages in each row of a two-way table allows us to establish whether a relationship exists between the two categorical variables that are being examined. As we can see from the table, the percentage of seniors who preferred Labor is about the same as non-seniors. Likewise, the percentage of seniors and non-seniors preferring Liberal are almost equal. This indicates that for the group of people participating in the survey, party preference is not related to age-category.

## EXERCISE 2

1. Complete the table.

| Voted | Children | Adults | Total |
| :---: | :---: | :---: | :---: |
| Yes | 25 |  | 47 |
| No |  |  |  |
| Total | 51 |  | 92 |

2. Members of a gym club were asked their age and what kind of training they do. Each responder only did one kind of training. The table shows the results.

|  | Cardio | Weight |  |
| :---: | :---: | :---: | :---: |
| 45 or <br> over | 44 | 18 |  |
| Under <br> 45 | 12 | 26 |  |
|  |  |  |  |

a. How many gym members were asked altogether?
b. How many members do weight training?
c. What percentage of total members do weight training?
3. Ben surveyed all the students in Year 12 at his school and summarised the results in the following table.

|  | Play sports | Do not play <br> sports | Total |
| :---: | :---: | :---: | :---: |
| Height $<170$ <br> cm | 30 | 45 | 75 |
| Height $>170$ <br> cm | 46 | 73 | 119 |
| Total | 76 | 118 | 194 |

a. What percentage of Year 12 students whose height is less than 170 cm play sports?
b. What percentage of students from Year 12 do not play sports?
4. In a study, some people were asked how many times they lie in a day. 20 responders said they lie at least once a day, 5 of which were children. 13 children said they never lie, and 15 adults said they never lie.
a. Complete the table

|  | 0 times | 1 or more <br> times |  |
| :---: | :---: | :---: | :--- |
| Children |  |  |  |
| Adults |  |  |  |
|  |  |  |  |

b. What percentage of responders said they never lied?
c. What percentage of adults said they had lied at least once?
5. In a survey, 639 students and 51 teachers were asked whether they approved or disapproved of changing the timetable. 421 students and 36 teachers approved of the change. Display this data in a two-way table.
6. The data show the reactions of administrative staff and technical staff to an upgrade of the computer systems at a large corporation. (Multiple choice)

| Attitude | Administrative <br> staff | Technical <br> staff | Total |
| :--- | :---: | :---: | :---: |
| For | 53 | 98 | 151 |
| Against | 37 | 31 | 68 |
| Total | 90 | 129 | 219 |

a. From the table, we can conclude that:
i. $53 \%$ of administrative staff were for the upgrade
ii. $37 \%$ of administrative staff were for the upgrade
iii. $27 \%$ of administrative staff were against the upgrade
iv. $59 \%$ of administrative staff were for the upgrade
v. $54 \%$ of administrative staff were against the upgrade
b. From the table, we can conclude that:
i. $98 \%$ of technical staff were for the upgrade
ii. $65 \%$ of technical staff were for the upgrade
iii. $76 \%$ of technical staff were for the upgrade
iv. $31 \%$ of technical staff were against the upgrade
v. $14 \%$ of technical staff were against the upgrade

## PARALLEL DOT PLOTS

Dot plots are a quick way to represent data visually. Dot plots are usually used for small data sets because they can help you to quickly see the spread of data and identify any outliers.

Parallel dot plots help you compare two or more sets of data, which must be plotted against the same scale. This makes the comparison between the data sets easy and ensures it isn't misleading. Parallel dot plots must be in the same unit and same scale.

## EXERCISE 3

1. Elizabeth did an experiment to see how well plants grow in different conditions. She had 8 plants grow in the sunshine, and 8 plants grow in the shade. She measured how tall they grew in centimetres after 2 months and recorded the information as a parallel dot plot.

a. Which group of plants had a higher range of heights?
b. Which dot plot shows a positive skew?
c. How much higher is the median height of plants that grew in the sunshine than the median height of plants that grew in the shade?
2. Twenty people joined a group fitness class and over two weeks, they were tested on the number of pull ups they can do. The dot plots show the number of pull ups each person could do.

a. From the first to the second week, what was the increase in the median number of chin ups someone could do?
b. In week 1, the average number of chin ups was 2.25 . Did the average increase or decrease in the second week?
3. Mae and Amelia are goal shooters for their netball teams. The table shows the number of goals they score in each game of the season.

|  | Game | Game | Game | Game | Game | Game | Game | Game |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Mae | 2 | 5 | 5 | 4 | 3 | 2 | 2 | 4 |
| Amelia | 2 | 1 | 1 | 3 | 4 | 2 | 3 | 1 |

a. Construct parallel dot plots for this data.
b. How does the range of each set of scores compare?

## DIVIDED BAR CHARTS

When comparing two categorical variables, it can be useful to represent the results from a two-way table (in percentage form) graphically. We can do this using divided (segmented) bar charts. A segmented bar chart consists of two or more columns, each of which matches one column in the two-way table. Each column is subdivided into segments, corresponding to each cell in that column. The segmented bar chart is a powerful visual aid for comparing and examining the relationship between two categorical variables.

## Example

Sixty-seven primary and 47 secondary school students were asked about their attitude to the number of school holidays which should be given. They were asked whether there should be fewer, the same number, or more school holidays. 5 primary school students and 2 secondary students wanted fewer holidays, 29 primary and 9 secondary students thought they had enough holidays (that is, they chose the same number) and the rest thought they needed to be given more holidays. The data in percentage form is shown in two-way frequency tables and a segmented bar chart.

## Solution

Raw Data:

| Attitude | Primary | Secondary | Total |
| :--- | :---: | :---: | :---: |
| Fewer | 5 | 2 | 7 |
| Same | 29 | 9 | 38 |
| More | 33 | 36 | 69 |
| Total | 67 | 47 | 114 |

Percentages:

| Attitude | Primary | Secondary |
| :--- | :---: | :---: |
| Fewer | 7.5 | 4.3 |
| Same | 43.3 | 19.1 |
| More | 49.2 | 76.6 |
| Total | 100 | 100 |

Divided Bar Chart:


Secondary students were much keener on
having more holidays than were primary
students.

1. A survey was conducted asking students from Year $5,6,7$ and 8 what eye colour they have. Study the four divided bar graphs shown here and answer the following questions.

a. Which grade has the largest percentage of brown-eyed children?
b. Which grade has the largest percentage of blue-eyed children?
c. Which grade has the largest percentage of black-eyed children?
d. Which grade has no green-eyed children?
e. Which grade has the same percentage of green-eyed children as blueeyed children?
2. The local library constructed divided bar graphs showing the last six months of book loans by category. Assume that the total number of loans was the same for each month.

a. In which month did subscribers borrow the most children's books?
b. In which month did people borrow the same number of design books as environment books?
c. Which book category grew constantly during the 6 months?

## SCATTER PLOTS

We often want to know if there is a relationship between two numerical variables. A scatter plot, which gives a visual display of the relationship between two variables, provides a good starting point. Consider the data obtained from last year's class at Northbank College. Each student in the class of 29 students was asked to give an estimate of the average number of hours they studied per week during Year 12. They were also asked for the ENTER score they obtained.

| Average <br> hours <br> of study | ENTER <br> score |
| :---: | :---: |
| 18 | 59 |
| 16 | 67 |
| 22 | 74 |
| 27 | 90 |
| 15 | 62 |
| 28 | 89 |
| 18 | 71 |
| 19 | 60 |
| 22 | 84 |
| 30 | 98 |


| Average <br> hours <br> of study | ENTER <br> score |
| :---: | :---: |
| 14 | 54 |
| 17 | 72 |
| 14 | 63 |
| 19 | 72 |
| 20 | 58 |
| 10 | 47 |
| 28 | 85 |
| 25 | 75 |
| 18 | 63 |
| 19 | 61 |


| Average <br> hours <br> of study | ENTER <br> score |
| :---: | :---: |
| 17 | 59 |
| 16 | 76 |
| 14 | 59 |
| 29 | 89 |
| 30 | 93 |
| 30 | 96 |
| 23 | 82 |
| 26 | 35 |
| 22 | 78 |
|  |  |

The figure below shows the data plotted on a scatter plot.
It is reasonable to think that the number of hours of study put in each week by students would affect their ENTER scores and so the number of hours of study per week is the independent variable and appears on the horizontal axis. The ENTER score is the dependent variable and appears on the vertical axis.

In analysing the scatter plot, we look for a pattern in the way the points lie. Certain patterns tell us that certain relationships exist between the two variables. This is referred to as correlation. We look at what type of correlation exists and how strong it is.


In the figure above, we see some sort of pattern: the points are spread in a rough corridor from bottom left to top right. We refer to this data following such a direction as having a positive relationship. This tells us that as the average number of hours studied per week increases, the ENTER score also increases.

The point $(26,35)$ is an outlier. It stands out because it is well away from the other points and clearly is not part of the 'corridor' referred to previously. This outlier may have occurred because a student exaggerated the number of hours he or she worked in a week or perhaps there was a recording error. This needs to be checked.

We could describe the rest of the data as having a linear form as the straight line in the diagram below indicates.


When describing the relationship between two variables displayed on a scatter plot, we need to comment on:
a. The direction - whether it is positive or negative
b. The form - whether it is linear or non-linear
c. The strength - whether it is strong, moderate, or weak
d. Possible outliers.

Below is a gallery of scatter plots showing the various patterns we look for:


Weak, positive linear relationship


Weak, negative linear relationship


Perfect, negative linear relationship


Moderate, positive linear relationship


Moderate, negative linear relationship



Strong, positive linear relationship


Strong, negative linear relationship


## Example

The scatter plot on the right shows the number of hours people spend at work each week and the number of hours people get to spend on recreational activities during the week.

Decide whether a relationship exists between the variables and comment on whether it is positive or
 negative; weak, moderate or strong; and whether it has a linear form.

## Solution

- The points on the scatter plot are spread in a certain pattern, namely in a rough corridor from the top left to the bottom right corner. This tells us that as the work hours increase, the recreation hours decrease.
- The corridor is straight (that is, it would be reasonable to fit a straight line into it).
- The points are neither too tight nor too dispersed.
- The pattern resembles the central diagram in the gallery of scatter plots show previously.
There is a moderate, negative relationship between the two variables.


## EXERCISE 5

1. Use the wording in the gallery above to describe the correlation between the variables shown in the graphs below:



2. For each of the following pairs of variables, write down whether you would reasonably expect a relationship to exist between the pair and, if so, comment on whether it would be a positive or negative association.
a. Time spent in a supermarket and money spent
b. Income and value of car driven
c. Number of children living in a house and time spent cleaning the house
d. Age and number of hours of competitive sports played per week
e. Amount spent on petrol each week and distance travelled by car each week
f. Number of hours spent in front of a computer each week and time spent playing the piano each week
g. Amount spent on weekly groceries and time spent gardening each week
h. Amount of time spent studying and grade on the test
3. The population of a municipality (to the nearest ten thousand) together with the number of primary schools in that municipality is given below for 11 municipalities.

| Population <br> $(x 1000)$ | 110 | 130 | 130 | 140 | 150 | 160 | 170 | 170 | 180 | 180 | 190 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Number of <br> primary <br> schools | 4 | 4 | 6 | 5 | 6 | 8 | 6 | 7 | 8 | 9 | 8 |

Construct a scatter plot (use a ruler) for the data and use it to comment on the direction, form, and strength of any relationship between the population and the number of primary students. Note: The population is the independent variable and thus becomes the $x$ axis.
You may use Sheets or Excel to construct the scatter plot.
4. The table below contains data for the time taken to do a paying job and the cost of the job. Determine the independent variable. Construct a scatter plot for the data (you may use Sheets or Excel). Comment on whether a relationship exists between the time taken and the cost. If there is a relationship, describe it.

| Time taken <br> (hours) | Cost of job (\$) |
| :---: | :---: |
| 5 | 1000 |
| 7 | 1000 |
| 5 | 1500 |
| 8 | 1200 |
| 10 | 2000 |
| 13 | 2500 |
| 15 | 2800 |
| 20 | 3200 |
| 18 | 2800 |
| 25 | 4000 |
| 33 | 3000 |

## AST STYLE QUESTION

Students at different schools (W, X, Y and Z) were surveyed about the type of book they prefer to read. The results of the survey are shown in the graph below.

For example, 200 students from School W were surveyed. Of these students, one-eighth ( 25 students) prefer to read non-fiction books.


## School

Use the information above to answer the three multiple-choice questions on the following page (circle your choice of correct answer).

## Hint:

- Complete the table below, showing student preferences from each school.

|  | School W | School X | School Y | School Z |
| :--- | :---: | :---: | :---: | :---: |
| Mystery |  |  |  |  |
| Non-fiction | 25 |  |  |  |
| Science fiction |  |  |  |  |
| Humour |  | 0 |  |  |

At which two schools do the same number of students prefer to read science fiction books?
A School X and School Z
B School W and School X
C School X and School Y
D School W and School Z

Which graph correctly shows the total number of students that prefer to read each type of book?
$\square$ Mystery $\square$ Non-fiction $\square$ Science fiction $\quad \times$ Humour

A


B


C


D


Another 320 students were surveyed at School V about their reading preferences. The survey showed that:

- the same number of students at School W and School V preferred to read mystery books
- a quarter of the students at School V preferred to read non-fiction books
- at School V an equal number of students preferred to read science fiction and humour books.

How many students at School V preferred to read humour books?
A 70
B 90
C 120
D 160

## INVESTIGATION WEEK 1/2

Hypothesis: The length of a person's foot is approximately $15 \%$ of their height.
You are going to investigate to see whether it is true.

1. Complete the table below to collect data from nine other people. Yours will be the tenth. Make sure you keep both pieces of information together, as you will need to compare each person's foot length with their height.

|  | Foot length (cm) | Height (m) |
| :---: | :---: | :---: |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |

2. Draw a scatter plot using your data on the next page.Call this scatter plot 'our school'. Note: You may use Google Sheets to draw a scatter plot and print it.
3. Complete this table to show what people's foot length would be if they were $15 \%$ of their height:

| Foot length (cm) | Height (m) |
| :---: | :---: |
|  | 1.0 |
|  | 1.2 |
|  | 1.4 |
|  | 1.6 |
|  | 1.8 |

4. Plot these points on your scatter graph (in a different colour).
5. Use your scatter graph to evaluate whether the hypothesis was true for your school. Consider the following

- How close in shape were the two scatter plots?
- Did any of your school results fall within/near the hypothesis scatter plot "corridor"?
- Was the hypothesis reasonable or do you think you would need more information?


## MARKING RUBRIC

| CRITERIA | EXPECTATIONS | POSS | MULT | GIVEN | TOTAL |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Practical | Student completes practical work of the brief to an acceptable standard set by the teacher. | 2 | 3 |  | /6 |
| Investigation | Student completes the investigation of the brief to an acceptable standard set by the teacher. | 2 | 2 |  | /4 |
| Reasoning and communications | Student responses are accurate and appropriate in presentation of mathematical ideas in different contexts, with clear and logical working out shown. | 4 | - |  | 14 |
| Concepts and techniques | Student submitted work selects and applies appropriate mathematical modelling and problem solving techniques to solve practical problems, and demonstrates proficiency in the use of mathematical facts, techniques and formulae. | 4 | - |  | /4 |
|  | Submission Guidelines |  |  |  |  |
| Timeliness | Student submits the exercises and investigation by the set deadline. See scoring guidelines for specific details. | 2 | - |  | /2 |
|  |  | FINAL |  |  | /20 |

## Student Reflection:

How did you go with this week's work?

What was interesting?

What did you find easy?

What do you need to work on?

