AS 91036

Mathematics and Statistics 1.11

Internally assessed
3 credits

Investigate bivariate numerical data using the statistical enquiry cycle

Introduction

A **variable** is a property that an individual has, e.g. their height. A variable may have different values for different individuals, or different values at different times for the same individual.

Bivariate data involves pairs of measurements, with *two* variables for the *one* subject.

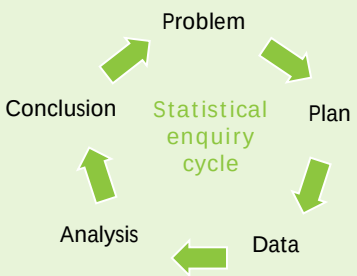
For example, you might measure the circumference of an egg at its widest point and the mass of that egg. Or you might record the time taken by a person to hop 100 metres and the time taken by that person to run 100 metres.

This achievement standard requires students to investigate bivariate numerical data using the statistical enquiry cycle. This will involve answering a question involving the relationship between the two variables.

The **relationship question** *will be supplied* in the assessment (there may be a selection of questions from which to choose). The data sets will *not* be supplied in the assessment and students will need to collect their own data.

Statistical enquiry cycle

The stages in the statistical enquiry cycle (PPDAC) should be followed during a statistical investigation.



Problem

The first stage involves defining the problem of interest, and posing a question.

In this achievement standard, the problem is posed using a given relationship question involving bivariate data. This question will give rise to an investigation into the relationship between two variables for the one subject.

Example

1. You may wonder if tall people have longer feet. The relationship question could be:

   Is there a relationship between the length of a student's right foot and their height?

   or

   How does the length of a student's right foot relate to their height?

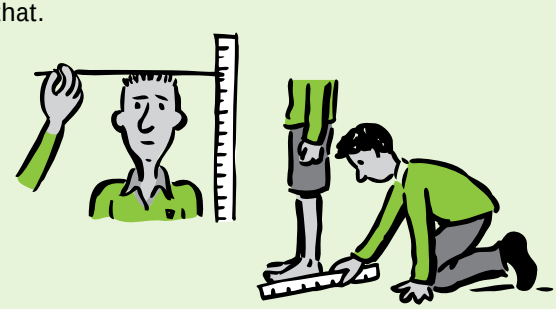
   Note: The question could just as easily have been about the length of a student's left foot.

2. The following question is *not* an appropriate relationship question.

   How does the length of a boy's right foot compare with the length of a girl's right foot?

   In this question only one variable (length of right foot) is being compared for two subjects (boys and girls).

You should give some thought to what answer you would predict for your question, and why you think that.



Ans. p. 29

Exercise A: Relationship questions

Which of the following questions is an appropriate relationship question for a bivariate data investigation? Remember there must be *two* variables for the *one* subject.

If a question is inappropriate, say why.

For each appropriate relationship question, name the subject and the two variables with suitable units, e.g. for 'Is there a relationship between the length of a carrot and its mass?' the subject is the carrot and the variables are length of carrot (cm or mm) and mass of carrot (g or kg).

1. Is there a relationship between a person's waist measurement and their weight?

2. How does the distance a car has travelled relate to its age?

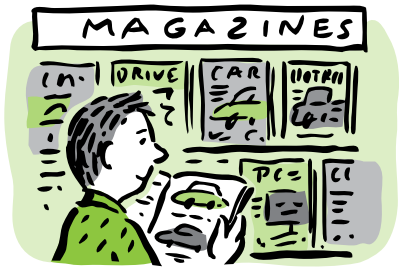
3. How does the birth weight of a labrador puppy compare with the birth weight of a poodle puppy?

4. How does the amount of rain in Auckland relate to the amount of rain in Wellington?

5. Is there a relationship between the number of pages in a book and its price?

6. How does the length of a person's arm relate to the length of their leg?

7. Is there a relationship between the daily number of customers in a shop and the total amount spent by these customers?



8. How do the times taken by girls to run 200 metres compare with the times taken by boys to run 200 metres?

9. Is there a relationship between the pulse rate of a person before and after exercise?

10. How does the height of a woman relate to the length of her armspan?

Plan

You will need to plan the collection of appropriate data, based on the relationship question.

Ask yourself the following questions:

- What data will need to be collected? This will depend on the variables in your relationship question.
- How much data will need to be collected? The usual minimum number of subjects is 30. Any **sample** of size less than 30 will make conclusions less reliable. The larger the sample size the better, but keep in mind how much time you have available for your investigation.
- Where will you get your sample from? For example, you might use a class of students.
- How will the data be collected? You will need to carefully explain how you will take your measurements so that you get consistent data (each person in the group should have a role in taking and recording measurements).
- What possible **sources of variation** are there in your data collection (i.e. what factors might affect the accuracy of your measures)? For example, when measuring heights of students, you would ensure no student is wearing shoes and that all student heights are measured in the same way using the same equipment.
- What other problems could arise in the collection of the data? Think hard about what could go wrong with your measurement process, or affect its accuracy.
- How can you manage these possible variations so that your data is as accurate as possible?

  What accuracy will you round measurements to?
- How will you record your results? Another person could check results. One person in your group could be the recorder.

Students will usually be working in small groups when planning and collecting the data, so each student will need to write down what he/she did during that process – what suggestions he/she made, what he/she did when collecting the data (each person in the group should be taking turns in each step of the data-gathering process).

The data

The measurements that were taken form the **data set** for the investigation. When you have collected the data, ask yourself:

- Does the data need **cleaning**? Are there any obvious errors that could be corrected, such as a data being recorded using the wrong units, e.g. height recorded as 1.62 (i.e. in metres) rather than 162 (i.e. in centimetres)? Are there some 'nonsense' answers (e.g. recording the time spent sleeping last night as 100 hours)? In this case the data would be removed as any other information from this subject is likely to be unreliable.
- Does the data need **sorting**? Remember that each pair of data values needs to be recorded as an **ordered pair** for that subject. Do not treat the two variables as values that can be sorted independently of each other. For example, in an investigation into the relationship between a student's height (in cm) and the length of their foot (in cm), the student's height measurement needs to stay paired with their foot length to make an ordered pair, e.g. Student 1 may have (height, foot length) = (167,42).

Bivariate data is best recorded in a **table** with three columns, as in the example below.

Student number	Height (cm)	Foot length (cm)
1	167	42
2	153	39

The first column is used to identify the subject number, e.g. Student 1, Student 2, etc.

The second column should be the measure for the first variable, e.g. Student 1's height, Student 2's height, etc.

The third column should be the measure for the second variable, e.g. Student 1's arm span, Student 2's arm span, etc.

Ans. p. 29

Exercise B: Planning a bivariate investigation

For each of the following relationship questions, you are part of a group of four students planning the investigation and collecting the bivariate data (pairs of measurements).

- Write down the variables involved and their units.
- Explain how the data could be collected.
- How much data is to be collected?
- Describe how the data is to be recorded.
- Discuss the role that each student in the group could take during the process of collecting the data.
- Explain how you could control variation in your measurements.

1. What is the relationship between a student's resting pulse rate and their pulse rate after a 2-minute run?

2. What is the relationship between the length of a person's index finger and their handspan?

3. What is the relationship between the length of a leaf and its maximum width?

4. What is the relationship between the length of a person's standing horizontal jump and the height of a standing vertical jump?

5. What is the relationship between the length of a person's right big toe and the length of their right thumb?

Analysis

Once the data has been collected (this may be done by a group) each student will need to analyse the data *individually*.

Scatter graphs

Any relationship between the two variables in bivariate data is best found by using a **scatter graph**. In a scatter graph, pairs of variables are plotted as points on a graph.

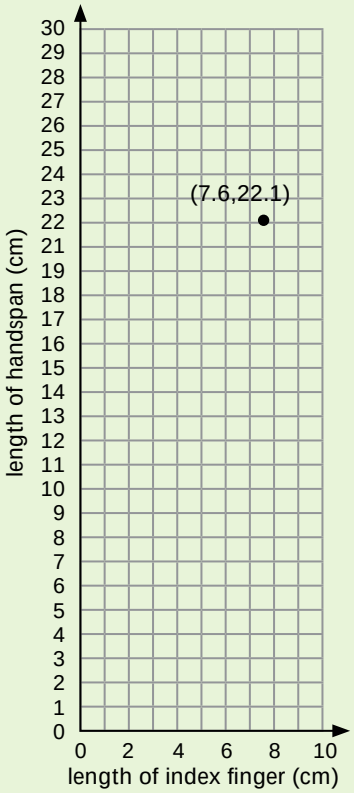
- The first variable is plotted on the ***x*-axis** (horizontal axis)
- The second variable is plotted on the ***y*-axis** (vertical axis)

Both axes should be labelled with the name of the variable and its unit.

Example

Miro is investigating the relationship between the length of a person's index finger and their handspan (how widely they can stretch their hand).

He drew a graph with the *x*-axis labelled 'length of index finger (cm)' and the *y*-axis labelled length of handspan (cm).

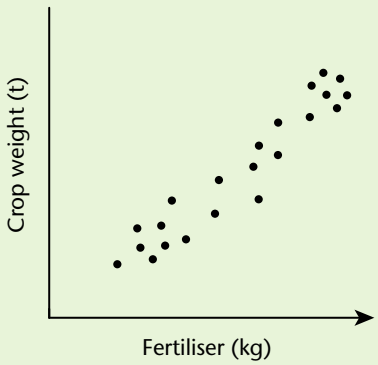


The first person he measured had an index finger length of 7.6 cm and a handspan of 22.1 cm, so he plotted the pair of measurements on the graph as the point (7.6,22.1), as shown.

Ans. p. 31

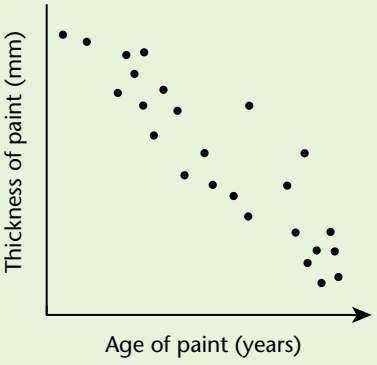
A scatter graph (or scatterplot) shows the relationship between two different variables. The scatter graphs below shows some types of relationship which can occur.

Positive linear relationship



The points lie approximately in a straight line with a positive gradient. As fertiliser use increases, the weight of the crop increases as well.

Negative linear relationship



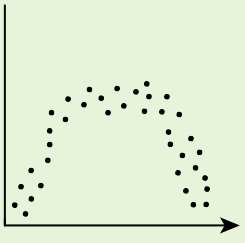
The points lie approximately in a straight line with a negative gradient.

As the age of the paint increases, its thickness decreases.

Other **non-linear** relationships can also occur, so that the points form a curve.

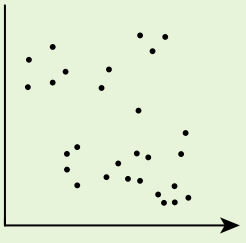
Sometimes there does not seem to be any apparent relationship between the variables.

Curved relationship



x and *y* have a parabolic relationship

No relationship



x and *y* appear unrelated

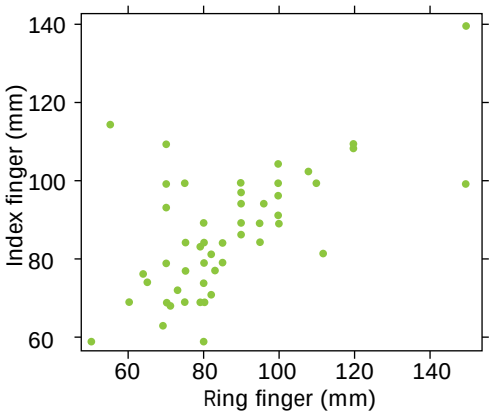
Exercise C: Scatter graphs

1. Discuss the relationship that you can see between the variables shown in the following scatter graphs.

   Use descriptions such as linear (with a positive or negative gradient), non-linear (curved), or no apparent relationship. In each case, comment on whether this would be the relationship (or lack of relationship) you would expect.

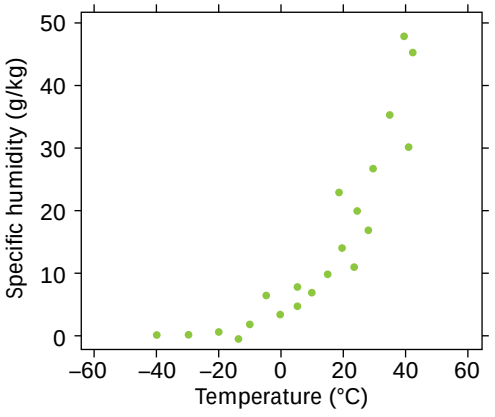
   a. Length of index finger and ring finger

Index finger versus ring finger

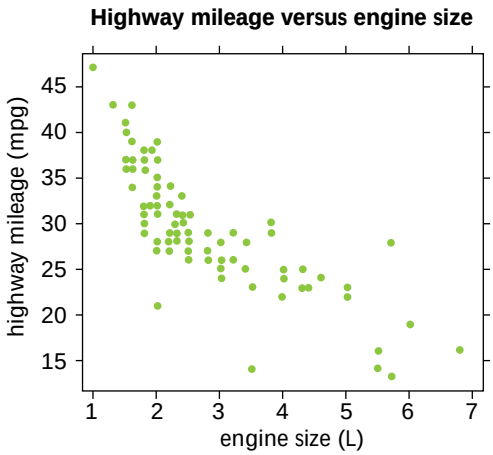


   b. Atmospheric temperature and humidity

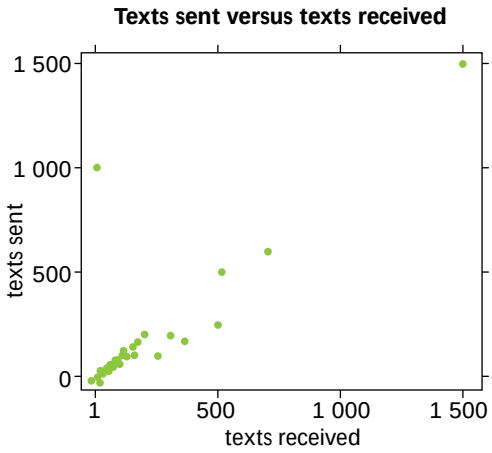
Humidity versus temperature in the atmosphere



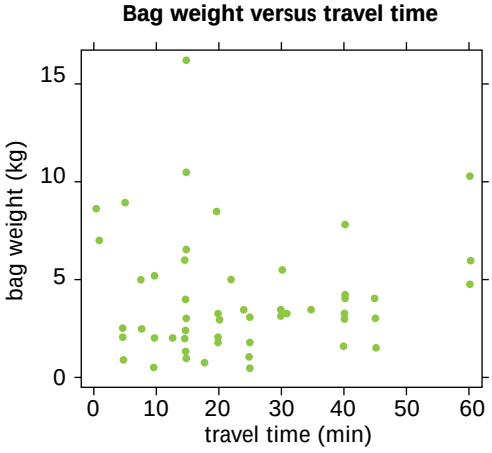
c. Engine size and fuel efficiency



d. Daily number of texts sent and received for Year 11 students



e. Weight of bag and time to travel to school for Year 11 students



2. The table shows the temperature at midday and the ice-cream sales at *Ices* for that day.

Day	Temperature at midday (°C)	Ice-cream Sales ($)
1	21.5	408
2	23.2	430
3	17.6	370
4	16.4	400
5	14.5	276
6	21.3	513
7	20.8	460
8	22.1	396
9	19.4	277
10	21.1	427
11	20.7	399
12	17	290
13	18.4	305
14	25.2	622
15	22.5	543
16	18.9	170
17	27.3	580
18	20.4	300
19	19.5	674
20	26.8	755

a. Draw a scatter graph for the data on the grid below. Give your graph a title.

b. Describe the relationship between temperature at midday and *Ices'* daily ice cream sales. Is this the relationship you would have expected? Give reasons.

3. Some cell phones were given an overall performance rating by a consumer magazine.

The table shows the costs and ratings of fourteen cell phones.

Phone	Price ($)	Rating (%)
A	699	81
B	399	76
C	999	74
D	399	73
E	599	73
F	299	69
G	299	68
H	599	68
I	249	67
J	349	65
K	169	63
L	149	61
M	249	59
N	199	58

Answers

Exercise A: Relationship questions (page 2)

1. Appropriate: subject: person; variables: waist measurement (cm) and weight (kg)
2. Appropriate: subject: car; variables: distance travelled (km) and age (years)
3. Inappropriate: one variable (birth weight) is being compared for two subjects: (labrador puppy and poodle puppy)
4. Inappropriate: one variable (rainfall) is being compared for two subjects: (Auckland and Wellington)
5. Appropriate: subject: book; variables: number of pages (whole number) and price ($)
6. Appropriate: subject: person; variables: arm length (cm) and leg length (cm)
7. Appropriate: subject: shop; variables: number of customers (whole number) and total amount spent ($)
8. Inappropriate: one variable (time to run 200 m) is being compared for two subjects: (boys and girls)
9. Appropriate: subject: person; variables: pulse rate before exercise (beats/min) and pulse rate after exercise (beats/min)
10. Appropriate: subject: woman; variables: height (cm) and armspan length (cm)

Exercise B: Planning a bivariate investigation (page 4)

Answers may vary, examples follow.

1. The variables are the student's resting pulse rate, in beats/min, as the *x*-variable and the pulse rate after the 2-minute run, in beats/min, as the *y*-variable.

   The resting pulse data could be collected in class with the pulse of a student taken for 30 seconds (using a stopwatch) and then doubled to give the number of beats per minute. The students would then go outside and run around a small area for a time of 2 minutes (using a stopwatch to time the run). Their pulse would be immediately recorded again, using the same procedure as above.

   The data would be recorded in a table with 3 columns – Student number/name, Resting pulse (beats/min), Run pulse (beats/min). Answers would be rounded to an agreed and consistent accuracy.

   Data would be collected from around 30 students (e.g. all the students in a class).

   In a group of four students, one could be responsible for the timing process (using a stopwatch), another for taking the pulse, a third supervising so that the procedures were carried out correctly and consistently, and the fourth recording the results.

   If students take their own pulse, there is the likelihood of variation among the results with students not following the procedure correctly.

   So students would need to practise how to take pulses correctly before the experiment began. Alternatively, the runners would need to be staggered (arriving at regular intervals) so that the person in the group assigned to taking pulses would be able to take each pulse immediately after the student finishes their 2-minute run.

   Other potential sources of variation include making sure that each student runs for the 2-minutes (some may walk) – possibly getting students to walk up and down a staircase, or jump on the spot may ensure that each student expends similar amounts of energy.

2. The variables are index finger length, in mm, as the *x*-variable and width of handspan, in mm, as the *y*-variable.

The index finger length could be measured by using a dress maker's tape, as could the handspan width. Starting and ending points for the length need to be defined carefully for consistent results. Students could work in pairs with one student measuring the length of the finger while another recorded the results. (Roles could be swapped for the measuring of the handspans.)

The data would be recorded in a table, with three columns with headings student number/name, Index finger length (mm) and Handspan width (mm). Answers would be rounded to an agreed and consistent accuracy.

Data would be collected from around 30 students (e.g. all the students in a class).

Variation can be controlled by very careful definition of the starting point for the measuring of an index finger (one student in the investigation group could mark this point with a pen on each person for consistency). Similarly handspans can vary quite a bit according to how much a person stretches out their hands, so this needs to be defined to be as wide a stretch as is possible, with this stretch being supervised while measurement is carried out.

3. Select a bush or a tree with reasonably sized leaves e.g. a camellia bush. The *x*-variable could be the length of the leaf, and the *y*-variable the width of the leaf at its widest point. Both measures would be in millimetres.

   Students could work in pairs with one student measuring the length of the leaf while another recorded the result in a table. Roles could be swapped with the measuring of the width of the leaf.

   The data would be recorded in a table, with three columns, with headings Leaf number, Leaf length (mm), Leaf width (mm). Answers would be rounded to an agreed and consistent accuracy.

   The number of leaves that would be measured is 30, and from different parts of the tree. (The leaves will not be pulled from the tree, but measured at the position in which they grow.)

   Starting and ending points for the length need to be defined carefully for consistent results. When measuring the width of the leaf, variability could arise as to which is the widest part of the leaf. Two measurements should be taken (by different group members) and the average recorded.

4. The variables are horizontal jump length, in cm, as the *x*-variable and vertical jump height, in cm, as the *y*-variable.

   Students would be asked to stand with their feet together on the ground and with their toes just behind a line marked on the floor or ground. They would be asked to leap as far forward as they can and the spot nearest to the start point where they touched the ground is noted, measured and recorded. (A measuring tape could be laid along the ground in order to make the measuring task quicker.) For the vertical jump, students would be asked to leap from a standing position (starting with both feet on the ground) next to a wall and reach as far up the wall as they can. The point that they touched the wall with their fingertips would be noted, measured and recorded. A tape could be fixed to the wall in order to make the task easier.

   The data would be recorded in a table, with three columns with headings: Student name/ number, Horizontal jump length (cm), Vertical jump length (cm). Answers would be rounded to an agreed and consistent accuracy.

   Members of the group would take turns in observing, taking the measures, supervising and recording the data in a table.

   The amount of data collected would be around 30 (e.g. all students in a class).

   The position of the feet before the horizontal jump needs to be carefully checked so that it is as close as possible behind the start line. The recording of the height of the jump could be difficult to record accurately, as the person only briefly touches the wall. Variation could be controlled by giving each student three jumps and taking the median length. Students could practise their jumps first so they master the techniques, then have their jump lengths recorded.

5. The variables are right thumb length, in mm, as the *x*-variable and right big toe length, in mm, as the *y*-variable.

   Both measures could be made using a

INDEX