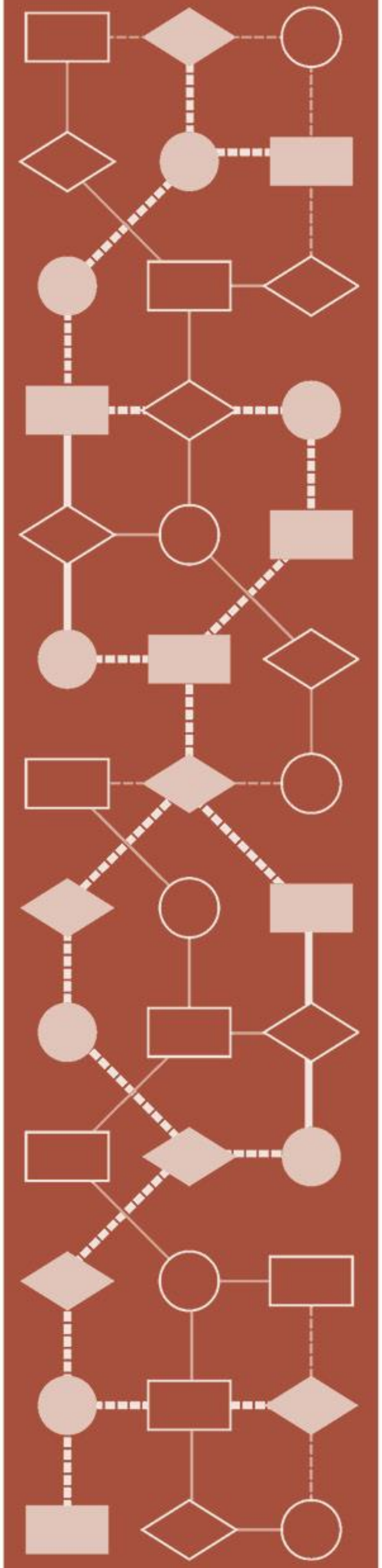
PROBLEM SOLVING



Mathematics Assessment Project
CLASSROOM CHALLENGES
A Formative Assessment Lesson

Devising a Measure: *Correlation*

Mathematics Assessment Resource Service
University of Nottingham & UC Berkeley

For more details, visit: http://map.mathshell.org

Devising a Measure: *Correlation*

MATHEMATICAL GOALS

This lesson unit is intended to help you assess how well students understand the notion of positive correlation. In particular this unit aims to identify and help students who have difficulty in:

- Understanding correlation as the degree of fit between two variables.
- Making a mathematical model of a situation.
- Testing and improving the model.
- Communicating their reasoning clearly.
- Evaluating alternative models of the situation.

COMMON CORE STATE STANDARDS

This lesson relates to **all** the *Mathematical Practices* in the *Common Core State Standards for Mathematics,* with a particular emphasis on Practices 2, 3, 5, and 6:

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

This lesson gives students the opportunity to apply their knowledge of the following *Standards for Mathematical Content* in the *Common Core State Standards for Mathematics*:

S-ID: Summarize, represent, and interpret data on two categorical and quantitative variables.
N-Q: Reason quantitatively and use units to solve problems.

INTRODUCTION

- Before the lesson, students work individually on an assessment task designed to reveal their current understanding and difficulties. You then review their work and create questions for students to answer in order to improve their methods.
- At the start of the lesson, students work alone answering your questions, then work collaboratively in small groups to produce, in the form of a poster, a better solution to the task than they did individually. In a whole-class discussion students compare and evaluate the different methods they have used.
- Then, working in the same small groups, students analyze sample responses to the task. In a whole-class discussion students explain and compare the alternative methods.
- In a follow-up lesson, students review what they have learnt.

MATERIALS REQUIRED

- Each student will need a copy of the *Drive-in Movie Theater* task and *Scatter Graphs A, B,* and *C*.
- Each small group of students will need a large sheet of paper, a felt-tipped pen, copies of the *Sample Responses to Discuss*, and a blank sheet of paper. If possible, use a data projector and computer with spreadsheet software to demonstrate the spreadsheet *Changing correlations.xls*. You may also need extra copies of *Scatter Graphs A*, *B*, and *C,* extra paper, and calculators.

TIME NEEDED

20 minutes before the lesson, a 90-minute lesson (or two 50-minute lessons), and 10 minutes in a follow-lesson. Timings are only approximate and depend on the needs of the class.

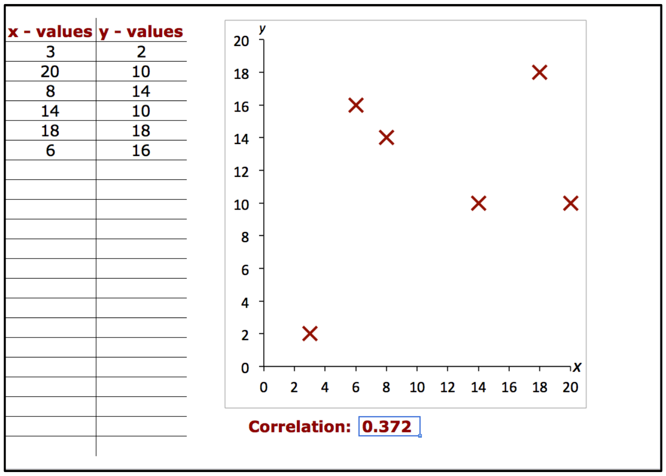
BEFORE THE LESSON

Whole-class introduction (10 minutes)

The task for this lesson involves students working with correlation. The focus of the lesson is, however, exploring different ways to calculate correlations. In the lesson students will make decisions on what math to apply to the problem, interpret and test their chosen method, and decide if their method makes sense or needs improving. Students will also critique the reasoning of others.

Have students complete the assessment task, in class, a few days before the formative assessment lesson. This will give you an opportunity to assess the work and to find out the kinds of difficulties students have with it. You should then be able to target your help more effectively in the follow-up lesson.

First remind the class of what a correlation means quantitatively by showing the students the excel file called *Changing correlations.xls*. If you do not have access to a computer and data projector then either sketch the sheet on the board or use Slide P-11 from the projector resource with an overhead projector, simultaneously running the spreadsheet on a laptop to calculate the correlations.



x - values	y - values
3	2
20	10
8	14
14	10
18	18
6	16

Correlation: 0.372

This sheet allows students to see how the correlation between the x- and y-values changes as more data is incorporated into the graph. You can add up to 14 more pairs of x- and y-values.

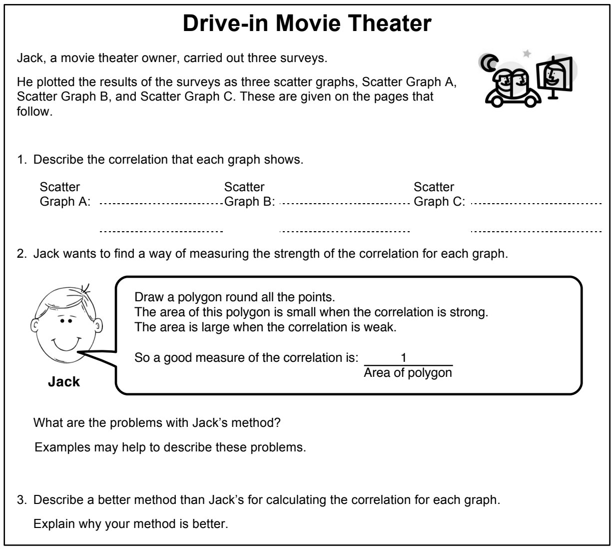
> *The correlation coefficient between the six x-values and the six y-values is now 0.372.*
>
> *Give me some values for x and y that will increase this positive correlation.*

Enter these values into the table.

> *Why do these values make the positive correlation greater?*
>
> *What is the **greatest possible** value for the correlation coefficient?*
>
> *Now give me some values for x and y that will keep the correlation positive, but make it closer to zero.*
>
> *Now give me some values for x and y that will result in a negative correlation.*

Remind students that correlations range from +1 to -1 and that both of these extremes show a strong correlation. Correlations near to zero show no association between the variables.

Assessment task: *Drive-in Movie Theater* (10 minutes)



Drive-in Movie Theater

Jack, a movie theater owner, carried out three surveys.

He plotted the results of the surveys as three scatter graphs, Scatter Graph A, Scatter Graph B, and Scatter Graph C. These are given on the pages that follow.

1. Describe the correlation that each graph shows.

Scatter Graph A: .... Scatter Graph B: .... Scatter Graph C: ....

2. Jack wants to find a way of measuring the strength of the correlation for each graph.

Draw a polygon round all the points.
The area of this polygon is small when the correlation is strong.
The area is large when the correlation is weak.

So a good measure of the correlation is: $\frac{1}{\text{Area of polygon}}$

Jack

What are the problems with Jack's method?

Examples may help to describe these problems.

3. Describe a better method than Jack's for calculating the correlation for each graph.

Explain why your method is better.

Give each student a copy of the assessment task *Drive-in Movie Theater* and *Scatter Graph A, B,* and *C*. Some students may need an extra sheet of paper.

First make sure all students understand the context. You can use the projector resource Slides P-1 to P-5 to help.

Has anyone visited a drive-in movie theater?

Did you go to the theater in the summer or winter?

Did you buy anything to eat or drink at the theater?

Then ask students to:

Read through the questions and try to answer them as carefully as you can.

Try to present your work in an organized and clear manner, so everyone can understand it.

It is important that, as far as possible, students are allowed to answer the questions without your assistance.

Students should not worry too much if they cannot understand or do everything because in the next lesson they will engage in work that should help them. Explain to students that, by the end of the next lesson, they should expect to answer questions such as these confidently. This is their goal.

Students who sit together often produce similar answers so that when they come to compare their work they have little to discuss. For this reason we suggest that, when students do the task individually, you ask them to move to different seats. Then at the beginning of the formative assessment lesson, allow them to return to their usual seats. Experience has shown that this produces more profitable discussions.

Assessing students' responses

Collect students' responses to the task. Make some notes on what their work reveals about their current levels of understanding and their different problem solving approaches.

We suggest that you do not score students' work. The research shows that this will be counterproductive, as it will encourage students to compare their scores and will distract their attention from what they can do to improve their mathematics.

Instead, help students to make further progress by summarizing their difficulties as a list of questions. Some suggestions for these are given in the *Common issues* table on pages T-3 and T-4. We suggest that you make a list of your own questions, based on your students' work, using the ideas on the following page. We recommend that you:

- write one or two questions on each student's work, or
- give each student a printed version of your list of questions and highlight the questions for each individual student.

If you do not have time to do this, you could select a few questions that will be of help to the majority of students and write these questions on the board when you return the work to the students at the beginning of the lesson.

Common issues	**Suggested questions and prompts**
Incorrect terminology used for describing the correlation between two sets of data (Q1)	- Use the correct mathematical terms to describe the correlations.
Has difficulty getting started on Q2	- Sketch two scatter graphs with different correlations. How can you use Jack's method to figure out the two correlations? Does Jack's method work for your two scatter graphs? Can you think of instances when Jack's method will not work?
Insufficient explanation provided (Q2) For example: The student simply states Jack's method does not work. Or: The student does not consider the size of the area of the polygon. Or: The student does not consider the shape of the polygon. Or: The student does not consider the sample size. Or: The student does not consider outliers. Or: The student does not consider if the method is easily repeatable. Or: The student does not consider the scale.	- Can you describe an instance when Jack's method does not work? - To support your work compare two different scatter graphs. - What size areas will not work? [The correlation will be greater than 1 for areas less than 1.] - Can you draw two polygons of approximately the same area, but different correlations? What does this tell you about Jack's method? - What would happen to the correlation if the sample size increased but the area of the polygon remained the same? What does this tell you about Jack's method? - What would happen to the correlation if there were an outlier? What does this tell you about Jack's method? - What would happen to the correlation if the scale changed? What does this tell you about Jack's method? - Suppose someone else used Jack's method on the same data, would they get the same correlation coefficient?
Has difficulty getting started on Q3	- What do you notice about each graph when there is a strong/weak correlation? - What do you notice about the data in each table when there is a strong/weak correlation? - Approximately what is the correlation for each graph?
Unproductive approach used (Q3)	- Read the question again. What are you trying to figure out? Is your method helping you to get there? - Are there any other ways of approaching the problem that might be productive?

Common issues	Suggested questions and prompts
The use of a calculator is suggested to obtain the correlation (Q3)	• You now need to figure out a method that gives approximately the same answer as the calculator.
A descriptive answer is provided but not supported by detailed math (Q3) For example: The student suggests adding a line of best fit and then measuring the distance between the data points and this line.	• Now use math to describe your method. • Now test your method on one of the scatter graphs.
The method described could result in incorrect correlations (Q3) For example: The student's method could result in a correlation outside the permissible range of zero to one. Or: The student's method does not allow for correlations equal to one or zero. Or: The student's method will produce a small correlation coefficient for strong correlations and a large correlation coefficient for weak correlations.	• Now test your method. • Can your method give a correlation bigger than one? • Can your method give a correlation of one or zero? • Will strong correlations result in a correlation coefficient close to one? Will weak correlations result in a correlation coefficient close to zero?
The method used would be difficult to accurately repeat (Q3) For example: The student draws by eye the line of best fit.	• Is your method accurate? • Suppose someone else used your method on the same data, would they get the same correlation coefficient? • Can you use a calculator to help you accurately draw the line of best fit?
The method is not tested on a scatter graph or a range of circumstances (Q3)	• Now test your method on one of the scatter graphs. • Now consider outliers. • Now consider plotting the scatter graph on a graph with a different scale. • Now consider using your method to plot more data/less data. • Does your method work for correlations that are very steep/very flat/that start high up the *y*-axis? • Does your method work for data that forms a near straight line but is spread over a wider/smaller range of values?
Approach is difficult to understand (Q3)	• Read through your work again. Would someone unfamiliar with this type of work understand what you have done?

SUGGESTED LESSON OUTLINE

Improve individual solutions to the assessment task (10 minutes)

Give each student a mini-whiteboard, pen, and eraser. Begin the lesson by briefly reintroducing the problem. You may want to show the class Slides P-1 to P-5 of the projector resource.

> *Recall what we were looking at in a previous lesson. What was the task about?*
>
> *Today we are going to work together to try to improve your initial attempts at this task.*
>
> *First, I have looked at your work and have some questions I would like you to think about.*
>
> *On your own, carefully read through the questions I have written. I would like you to use the questions to help you to think about ways of improving your own work.*
>
> *Use your mini-whiteboards to make a note of anything you think will help improve your work.*

Return the *Drive-in Movie Theater* assessment to the students.

If you have not added questions to individual pieces of work or highlighted questions on a printed list of questions then write your list of questions on the board. Students should select from this list only those questions they think are appropriate to their own work.

Collaborative small-group work (25 minutes)

Organize the class into small groups of two or three students. To encourage mathematical discussion you may want to group together students who have used different methods to calculate correlation.

Give each group a large sheet of paper and a felt-tipped pen.

Have copies of *Scatter Graph A*, *B*, *C*, and calculators for students who request them.

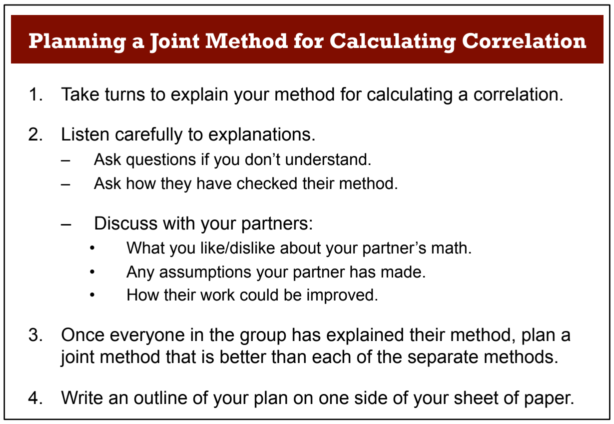
Deciding on a Strategy

Ask students to share their ideas about the task and plan a joint method.

> *I want you to share your method for calculating a correlation with your group.*
>
> *You are then going to come up with a joint method for calculating correlations that is better than your separate ideas.*

Once students have evaluated the relative merits of each approach, ask them to write an outline of their method on one side of their sheet of paper.

Slide P-6 of the projector resource summarizes how students should work together:



Planning a Joint Method for Calculating Correlation

1. Take turns to explain your method for calculating a correlation.
2. Listen carefully to explanations.
   - Ask questions if you don't understand.
   - Ask how they have checked their method.
   - Discuss with your partners:
     - What you like/dislike about your partner's math.
     - Any assumptions your partner has made.
     - How their work could be improved.
3. Once everyone in the group has explained their method, plan a joint method that is better than each of the separate methods.
4. Write an outline of your plan on one side of your sheet of paper.

Implementing the Strategy

As students finish planning their strategy, give them a copy of *Scatter Graph B,* and a glue stick.

> *Glue Scatter Graph B in the center of the reverse side of your large sheet of paper.*
>
> *Write your method next to or on the scatter graph.*

While students work in small groups you have two tasks: to note different student approaches to the task and to support student problem solving.

Support student approaches to the task

In particular, note any common mistakes. Are students making any incorrect assumptions? Are students testing their method under a range of circumstances? Do they use a calculator and if so how? You can then use this information to focus a whole-class discussion at the end of the lesson. Attend also to students' mathematical decisions. Do they track their own progress? Do they notice if they have chosen a method that does not seem to be productive? If so, what do they do?

Support student problem solving

Try not to make suggestions that move students towards a particular approach to this task. Instead, ask questions that help students clarify their thinking, promote further progress and encourage students to check their work and detect errors. You may want to use some of the questions in the *Common issues* table to support your own questioning or, if the whole-class is struggling on the same issue, write relevant questions on the board and hold a brief whole-class discussion. You could also give any struggling students one of the *Sample Responses to Discuss*.

> *What do you already know? What do you need to know?*
>
> *When the positive correlation between two sets of data is strong, what does the graph look like? What do you notice about the data? How can you use this information?*
>
> *Is there a connection between steepness of the line of best fit and the strength of the correlation? [No.]*
>
> *How can you test your method?*
>
> *How can you improve your method as a result of your tests?*

Whole-class discussion (10 minutes)

When students have had sufficient time to work on their posters hold a whole-class discussion to review how students have worked. Have students solved the problem using a variety of methods? Or have you noticed some interesting ways of working or some incorrect methods. If so, you may want to focus the discussion on these. Equally, if you have noticed different groups using similar strategies but making different assumptions, you may want to compare solutions.

Ask other students to comment on:

- As the correlation gets stronger will the correlation coefficient increase?
- Did students test their method for a range of data?
- Could their method be accurately repeated using different scales?
- Could their method result in impossible correlation coefficients, for example coefficients bigger than 1?
- Was the reasoning easy to understand and follow?

You may want to also draw on the questions in the *Common issues* table to support your own questioning.

Extending the lesson over two days

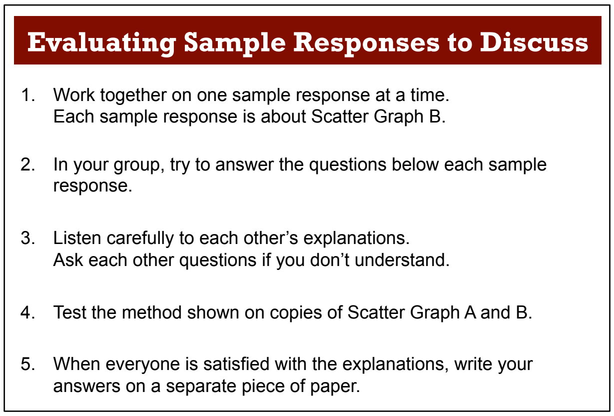
If you are taking two days to complete the unit then you may want to end the first lesson here. At the start of the second day, briefly remind students of the problem, before moving on to the collaborative analysis of sample responses.

Collaborative analysis of *Sample Responses to Discuss* (30 minutes)

Depending on your class and time, give each group 1, 2, or all 3 *Sample Responses to Discuss* and a blank sheet of paper and ask for written comments.

Be selective about what you hand out. For example, groups that have successfully completed the task using one method will benefit from looking at different approaches. Other groups that have struggled with a particular approach may benefit from seeing a student version of a similar method. This task gives students the opportunity to evaluate a variety of possible approaches to the task.

Show Slide P-7 of the projector resource, which summarizes how students should work together:



Evaluating Sample Responses to Discuss

1. Work together on one sample response at a time.
   Each sample response is about Scatter Graph B.
2. In your group, try to answer the questions below each sample response.
3. Listen carefully to each other's explanations.
   Ask each other questions if you don't understand.
4. Test the method shown on copies of Scatter Graph A and B.
5. When everyone is satisfied with the explanations, write your answers on a separate piece of paper.

> *I have copies of Scatter Graph A and C for you to use.*

Encourage students to focus on the math of the student work, not simply to check to see if the answer is correct or whether the student has neat writing etc.

During the small-group work, support the students in their analysis. As before, try to help students develop their thinking, rather than resolve difficulties for them.

Again encourage students to test each method. Students may want to sketch a graph showing perfect or weak correlations using, say, five data points.

Note similarities and differences between the sample approaches and those approaches students took in the small-group work. Also check to see which methods students have difficulties in understanding. This information can help you focus the next activity, a whole-class discussion.

Sam has drawn a line of best fit on the graph. He has then measured the distance each cross is from this line. He has figured out a total for all these distances.

Are Sam's two statements about the distances correct? [Yes.]

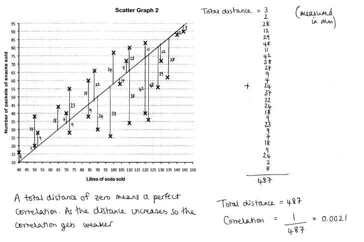
What would happen to the correlation if a different scale were used? [The correlation would change.]

How could Sam overcome the problem of scale?
[Measure each distance using the scale on the axis.]

How do you think Sam drew the line of best fit?
[If he drew it by eye then his method is inaccurate.]

Suppose the total of all distances is zero. What is the correlation?
[Infinity.]

How could you improve Sam's method?



Total distance = 3
2
28
12
29
48
11
42
38
27
9
4
+ 24
37
22
24
18
9
23
9
7
18
9
24
2
8
487

(measured in mm)

A total distance of zero means a perfect correlation. As the distance increases so the correlation gets weaker

Total distance = 487

Correlation = $\frac{1}{487}$ = 0.0021

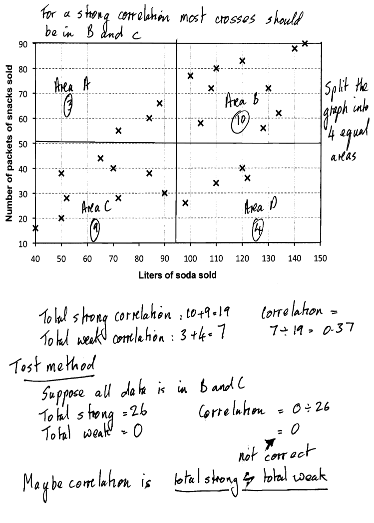
Nina has split the graph into four equal areas. Nina's statement about a strong positive correlation is correct. 19 and 7 represent the total number of values in areas B and C and areas A and D respectively. Nina's correlation calculation is incorrect.

Is Nina's statement about a strong correlation correct? [Yes.]

What is incorrect about the correlation calculation? [One problem with the correlation = total strong ÷ total weak *is that there could be a correlation greater than 1 (including infinity!).]*

*What would be a better correlation calculation? [A better calculation for the correlation is 19 ÷ 26 (*total strong total ÷ total number of data points*.]*

Could this improved method ever give a correct correlation of one/zero/above one? [Yes/Yes/No.]



For a strong correlation most crosses should be in B and C

Split the graph into 4 equal areas

Total strong correlation : 10+9=19
Total weak correlation : 3+4=7

Correlation =
7 ÷ 19 = 0.37

Test method

Suppose all data is in B and C
Total strong = 26
Total weak = 0

Correlation = 0 ÷ 26
= 0
not correct

Maybe correlation is total strong ÷ total weak

Judith uses just the table to figure out a method. She correctly notices that for a strong correlation the rank order of the x and y values should be similar.

Judith's first correlation calculation is incorrect because it would result in a higher rank difference, producing a higher correlation figure.

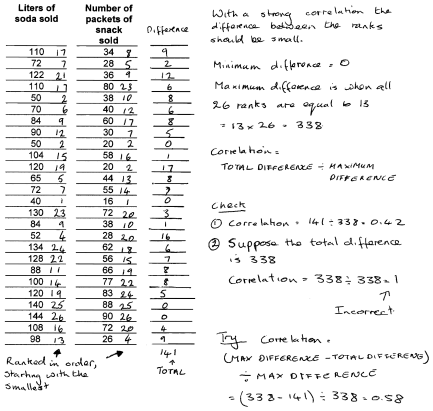
You may find some students have difficulty understanding Judith's method. These questions may help:

Using Judith's method, when would there be perfect correlation between the two sets of data?

Sketch a graph of a perfect correlation, using say six data points.
Now rank separately the two values for each data point.
What do you notice about the difference between the ranking for each data point?
What would you notice if the correlation was very weak?

You may want to ask further questions about Judith's work:

Is Judith's statement about a strong correlation correct?



Liters of soda sold		Number of packets of snack sold		Difference
110	17	34	8	9
72	7	28	5	2
122	21	36	9	12
110	17	80	23	6
50	2	38	10	8
70	6	40	12	6
84	9	60	17	8
90	12	30	7	5
50	2	20	2	0
104	15	58	16	1
120	19	20	2	17
65	5	44	13	8
72	7	55	14	7
40	1	16	1	0
130	23	72	20	3
84	9	38	10	1
52	4	28	20	16
134	24	62	18	6
128	22	56	15	7
88	11	66	19	8
100	14	77	22	8
120	19	83	24	5
140	25	88	25	0
144	26	90	26	0
108	16	72	20	4
98	13	26	4	9
				141

Ranked in order, starting with the smallest

TOTAL

With a strong correlation the difference between the ranks should be small.

Minimum difference = 0

Maximum difference is when all 26 ranks are equal to 13

= 13 × 26 = 338

Correlation =
TOTAL DIFFERENCE ÷ MAXIMUM DIFFERENCE

Check

① Correlation = 141 ÷ 338 = 0.42

② Suppose the total difference is 338

Correlation = 338 ÷ 338 = 1

Incorrect

Try Correlation =
(MAX DIFFERENCE − TOTAL DIFFERENCE)
÷ MAX DIFFERENCE
= (338 − 141) ÷ 338 = 0.58

Why did Judith rank the data?

Is Judith's figure of 338 for the largest total possible difference in ranking correct?

Why is Judith's first correlation formula incorrect?

Does Judith's first correlation calculation produce a high figure for a strong correlation? [No.]

Could Judith's method ever give a correlation of one/zero/above one? [Yes/Yes/No.]

Is Judith's second correlation formula correct? Explain your answer.

To answer the question "Is Judith's figure of 338 for the total possible largest difference in ranking correct?" students may write all the values in a long ordered sum:

Total difference: $(26 - 1) + (25 - 2) + (24 - 3)\ldots = 338$.

Or students may check Judith's method by using guess and check.
For example:

4 data points:

Rank soda	Rank snacks	Difference
1	4	3
2	3	1
3	2	1
4	1	3
	Total:	8

Or $4 \times 2 = 8$

6 data points:

Rank soda	Rank snacks	Difference
1	6	5
2	5	3
3	4	1
4	3	1
5	2	3
6	1	5
	Total:	18

Or $6 \times 3 = 18$

Students may then assume that for 26 data points the total difference in ranking is $26 \times 13 = 338$.

This method works for data sets containing an even number of data points.

This works because the total difference in rank is the sum of two equal sets of consecutive odd numbers.

Suppose n is the even number of data points. Then:

$$\text{Sum of one set of odd numbers} = \left(\frac{n}{2}\right)^2$$

$$\text{Sum of the difference in rank} = 2 \times \left(\frac{n}{2}\right)^2 = 2 \times \frac{n^2}{4} = \frac{n^2}{2} = n \times \frac{n}{2}$$

Students should not be expected to provide an algebraic explanation for the total difference in rank.

Whole-class discussion: *comparing different approaches* (15 minutes)

Hold a whole-class discussion to consider the different approaches used in the sample work. Focus the discussion on parts of the task students found difficult. Ask the students to compare the different methods.

Which method did you like best? Why?

Which method did you find most difficult to understand? Why?

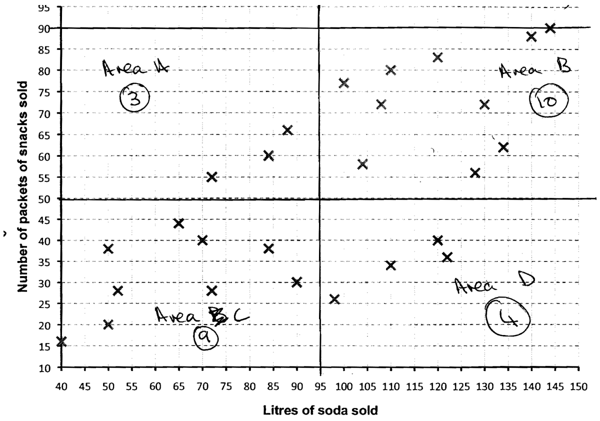
How could the student improve his/her answer?

Try to focus the discussion on any common misconceptions you noticed in the collaborative work. You may want to draw on the questions in the *Common issues* table to support your own questioning. Try to resist simply explaining the mistakes students have made.

To support the discussion, you may want to use Slides P-8, P-9, and P-10 of the projector resource.

In another lesson you may want to discuss with your class a statistical method for calculating correlation coefficients.

For example, this method is similar to Nina's.



Suppose:

a = number of crosses in area A.

b = number of crosses in area B.

c = number of crosses in area C.

d = number of crosses in area D.

$$\text{Correlation} = \frac{(bc - ad)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$= \frac{(10 \times 9 - 3 \times 4)}{\sqrt{(3+10)(9+4)(3+9)(10+4)}}$$

$$= \frac{78}{\sqrt{13 \times 13 \times 12 \times 14}} = 0.46$$

Does this formula allow the correlation to vary between 0 and 1?

Test it on your graphs.

Follow-up lesson: review of solutions (10 minutes)

Give a copy of the *How Did You Work?* questionnaire to each student.

Ask students to complete the review questionnaire.

Read through your first solution. Think about all the work you've done on this problem: the first time you tried it, when you used my questions working alone, when you worked with your partner.

Fill in the questions as you reflect on your experience.

If you have time you may also want to ask students to use what they have learned to attempt the task again.

Some teachers give this as a homework task.

SOLUTIONS

Assessment task: *Drive-in Movie Theater*

1. Scatter Graph A: Strong positive correlation.

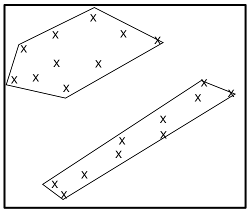
   Scatter Graph B: Weak positive correlation.

   Scatter Graph C: No correlation.

2. Here are some of the problems with Jack's method:

   Jack's method does not take into account the shape of the polygon.

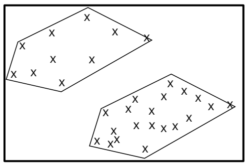
   For example: Two different scatter graphs, with the same number of data points and equal areas enclosing these points will have the same correlation.

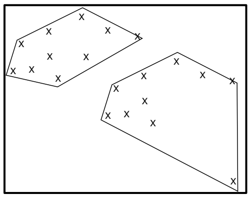
   Area is dependent on the scale of the graph. The measure of the correlation will change when the scale changes.

   Jack's method does not take into account the number of points in the data set.

   For example: Adding points that lie within the area enclosing the data points should increase the measure of the correlation. This does not happen with Jack's method.

   An outlier data point can have a disproportional effect on the measure of the correlation.

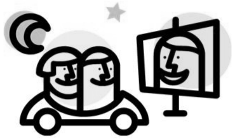
   

   The measure of the correlation cannot be zero and can be greater than one.

3. Students are not expected to devise a statistical method that is commonly used to measure correlation.

   The emphasis of the task is for students to provide a method that has been tested, revised and explained clearly.

Drive-in Movie Theater



Jack, a movie theater owner, carried out three surveys.

He plotted the results of the surveys as three scatter graphs, Scatter Graph A, Scatter Graph B, and Scatter Graph C. These are given on the pages that follow.

1. Describe the correlation that each graph shows.

   Scatter Graph A: ------------------------------

   Scatter Graph B: ------------------------------

   Scatter Graph C: ------------------------------

2. Jack wants to find a way of measuring the strength of the correlation for each graph.

Jack

Draw a polygon round all the points.
The area of this polygon is small when the correlation is strong.
The area is large when the correlation is weak.

So a good measure of the correlation is: $\dfrac{1}{\text{Area of polygon}}$

What are the problems with Jack's method?

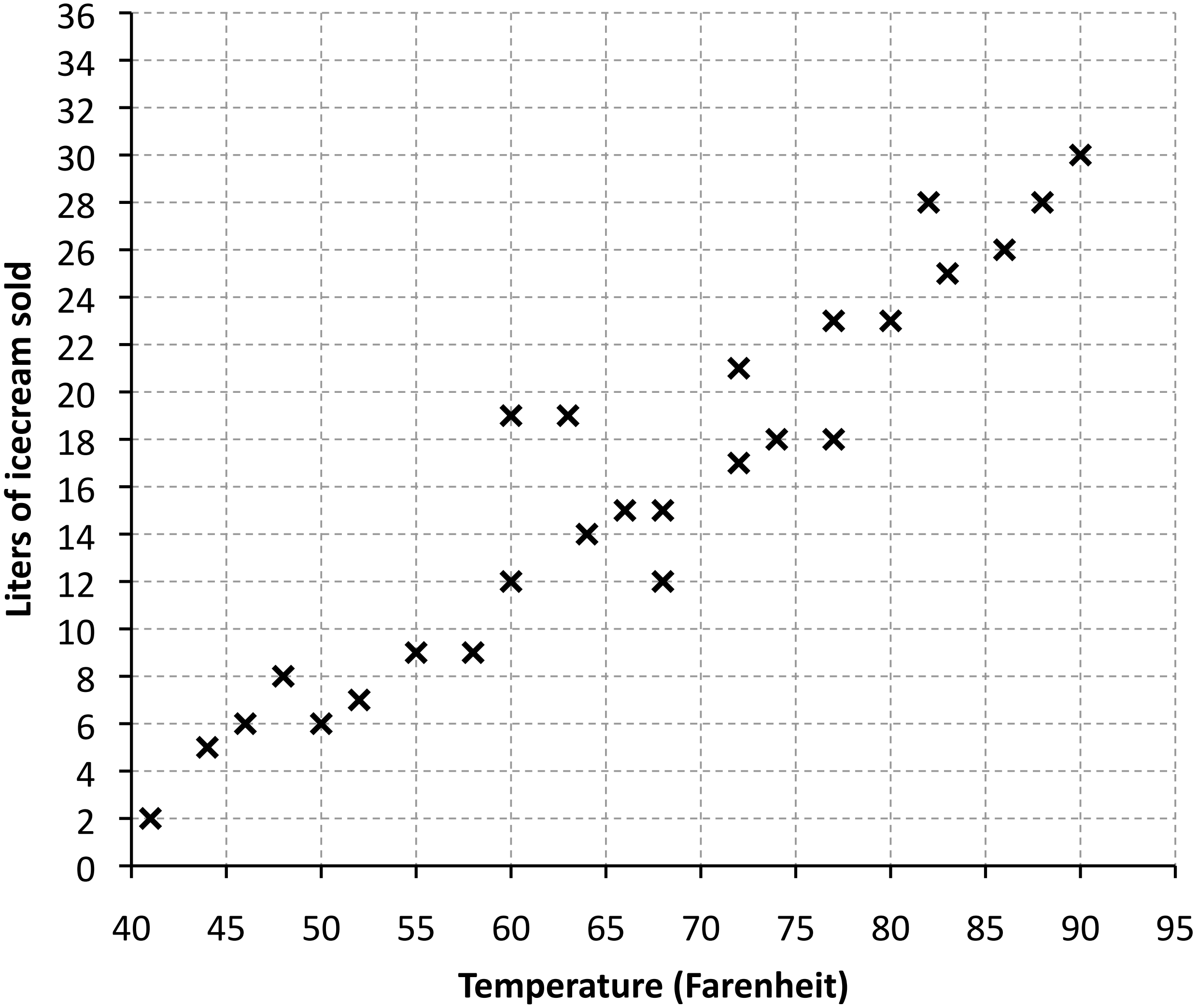
Examples may help to describe these problems.

3. Describe a better method than Jack's for calculating the correlation for each graph.

   Explain why your method is better.

Scatter Graph A

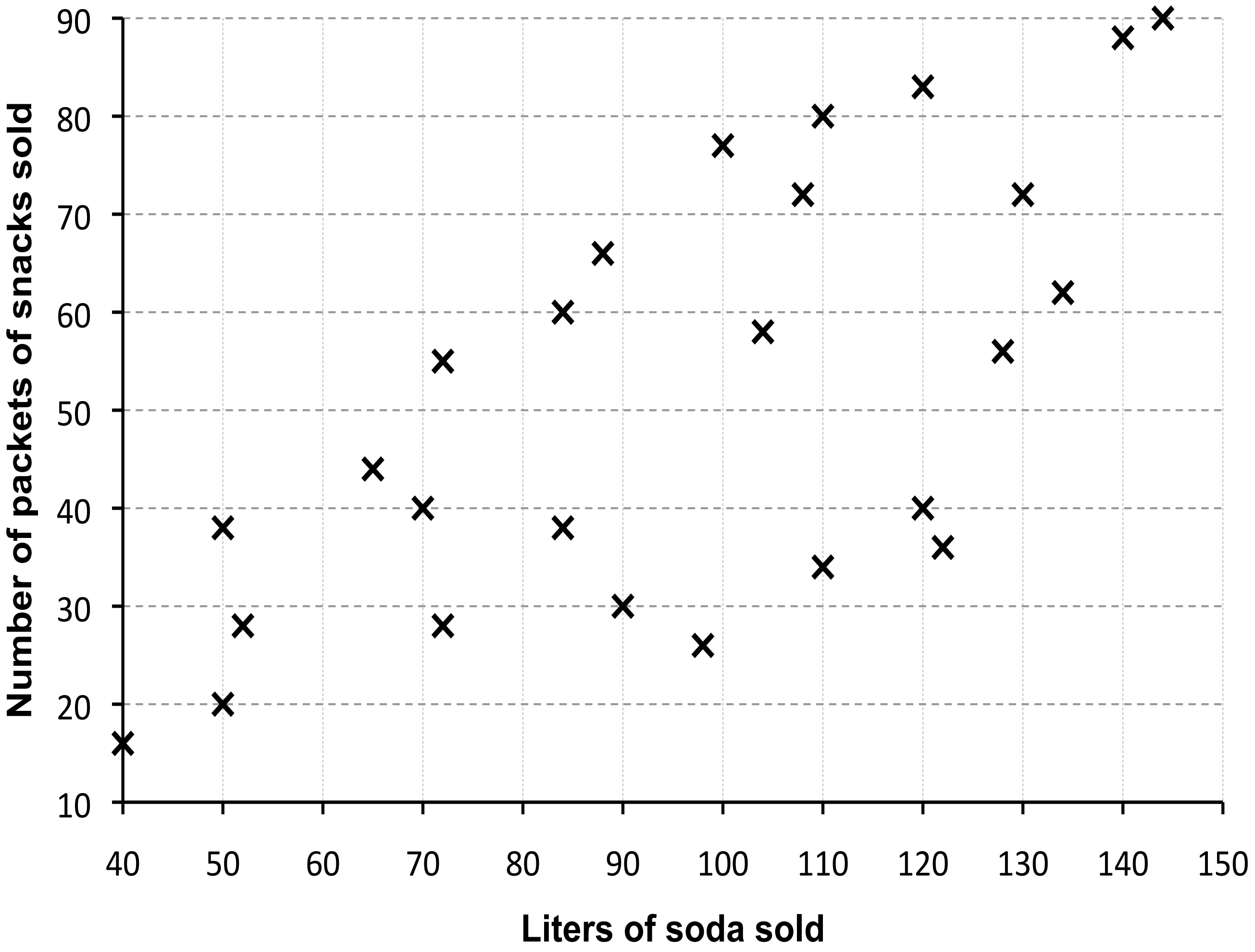
Ice cream sold v Outside temperature



Temperature (Fahrenheit)	Liters of ice cream sold
41	2
83	25
64	14
52	7
72	17
68	15
50	6
55	9
74	18
66	15
68	12
82	28
72	21
88	28
86	26
60	19
90	30
80	23
77	23
44	5
46	6
58	9
63	19
48	8
60	12
77	18

Scatter Graph B

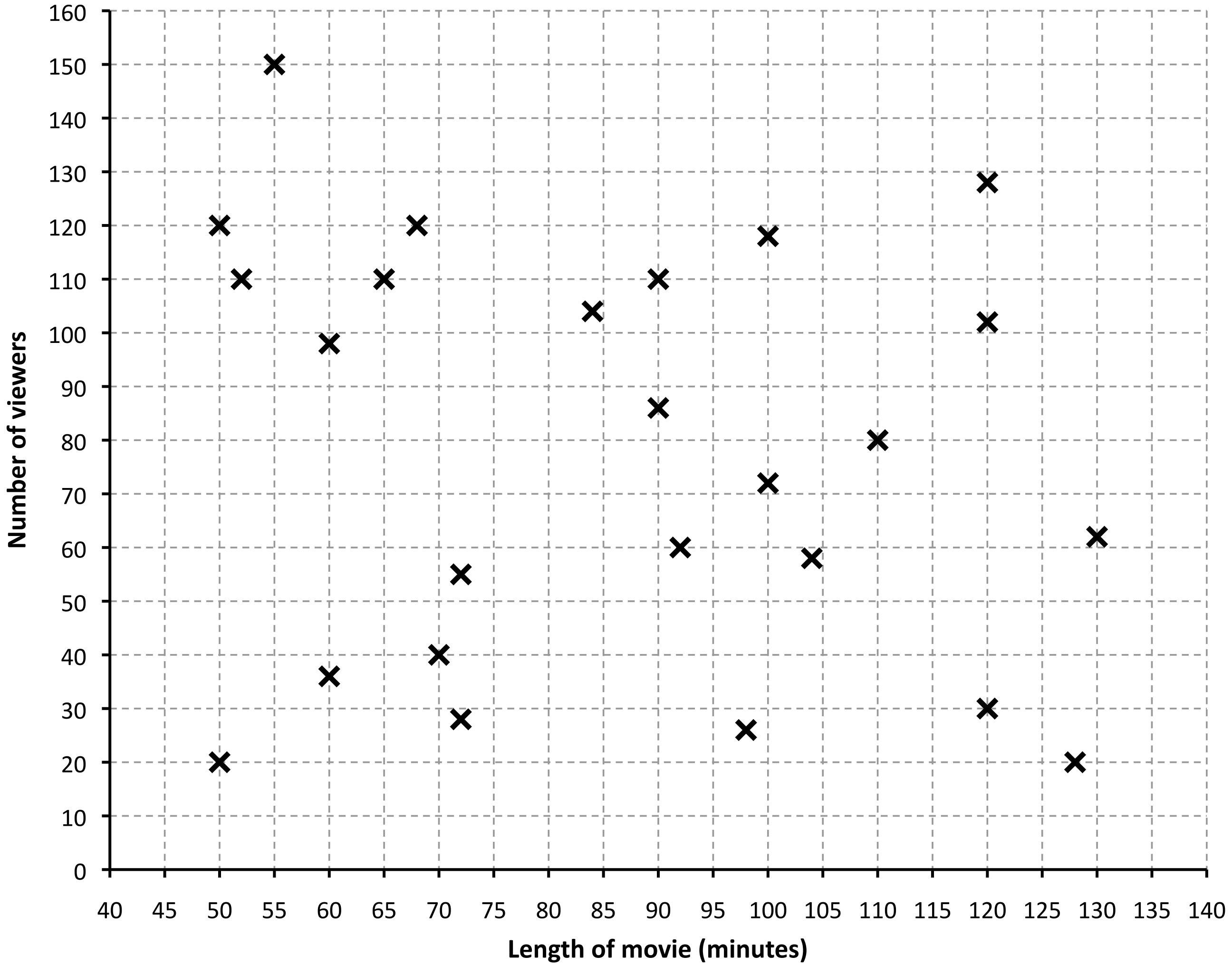
Snacks sold v Soda sold



Liters of soda sold	Number of packets of snack sold
110	34
72	28
122	36
110	80
50	38
70	40
84	60
90	30
50	20
104	58
120	20
65	44
72	55
40	16
130	72
84	38
52	28
134	62
128	56
88	66
100	77
120	83
140	88
144	90
108	72
98	26

Scatter Graph C

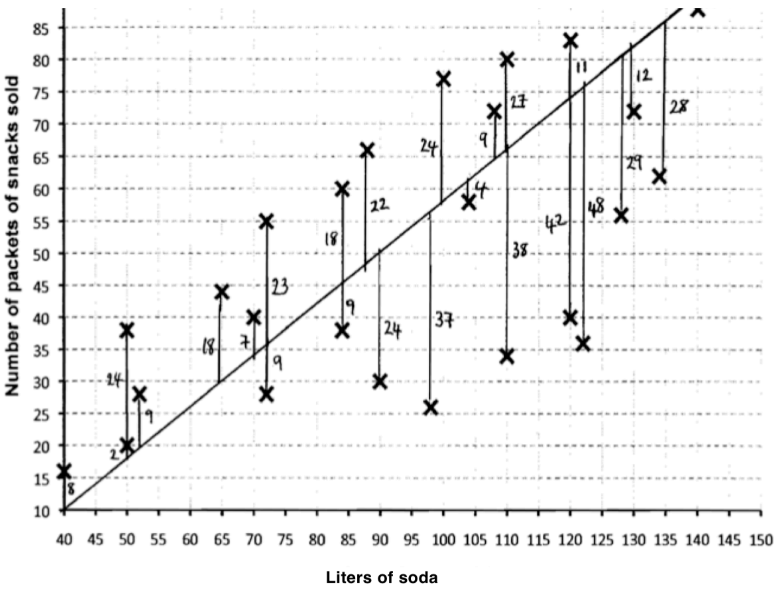
Number of viewers v Length of a movie



Length of movie (minutes)	**Number of viewers**
55	150
72	28
60	36
110	80
50	120
70	40
92	60
90	110
50	20
104	58
68	120
65	110
72	55
60	98
150	72
84	104
52	110
130	62
128	20
90	86
100	118
120	128
120	102
120	30
100	72
98	26

Sample Responses to Discuss: Sam

Scatter Graph B



Total distance = 3
2
28
12
29
48
11
42
38
27
9
4
24
37
22
24
18
9
23
9
7
18
9
24
2
8
+

487

A total distance of zero means a perfect correlation. As the distance increases so the correlation gets weaker

(measured in mm)

Total distance = 487

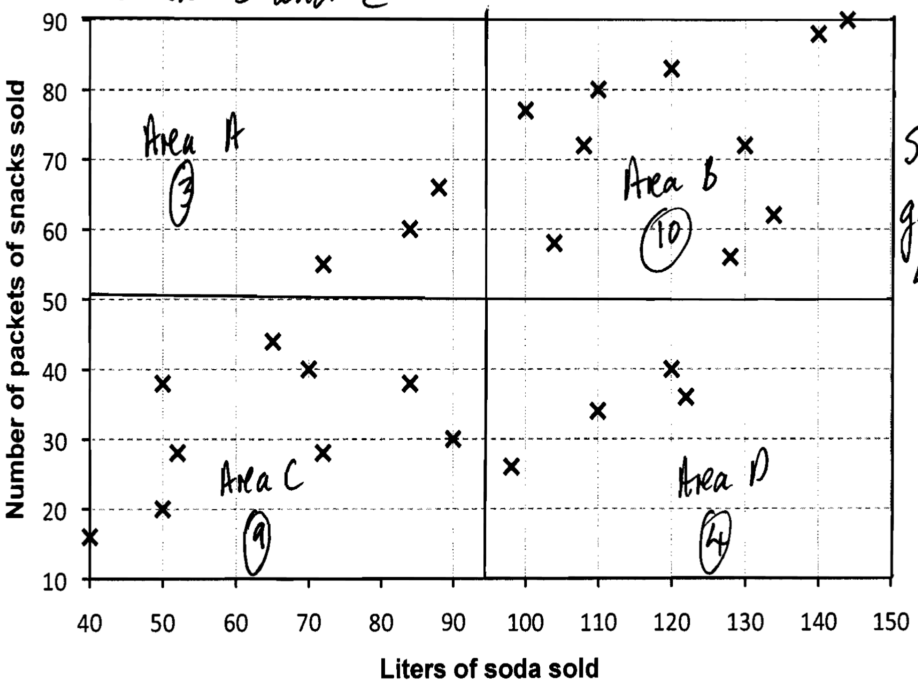
$$\text{Correlation} = \frac{1}{487} = 0.0021$$

1. Are Sam's two statements about the distances correct?
2. What would happen to the correlation if a different scale were used?
3. Suppose the total of all distances is zero. What is the correlation?
4. How could you improve Sam's method?

Sample Responses to Discuss: Nina

Scatter Graph B

For a strong correlation most crosses should be in B and C



Split the graph into 4 equal areas

Total strong correlation : 10+9=19

Total weak correlation : 3+4=7

correlation = 7÷19 = 0.37

Test method

Suppose all data is in B and C

Total strong = 26

Total weak = 0

Correlation = 0÷26

= 0

not correct

Maybe correlation is total strong ÷ total weak

1. Is Nina's statement about a strong correlation correct?
2. What is incorrect about the correlation calculation?
3. How could Nina improve her method?

Sample Responses to Discuss: Judith

Scatter Graph B

Liters of soda sold	Number of packets of snack sold	Difference
110 17	34 8	9
72 7	28 5	2
122 21	36 9	12
110 17	80 23	6
50 2	38 10	8
70 6	40 12	6
84 9	60 17	8
90 12	30 7	5
50 2	20 2	0
104 15	58 16	1
120 19	20 2	17
65 5	44 13	8
72 7	55 14	7
40 1	16 1	0
130 23	72 20	3
84 9	38 10	1
52 4	28 20	16
134 24	62 18	6
128 22	56 15	7
88 11	66 19	8
100 14	77 22	8
120 19	83 24	5
140 25	88 25	0
144 26	90 26	0
108 16	72 20	4
98 13	26 4	9
		141

↑ ↗ Ranked in order, starting with the smallest

↑ TOTAL

With a strong correlation the difference between the ranks should be small.

Minimum difference = 0

Maximum difference is when all 26 ranks are equal to 13

$= 13 \times 26 = 338$

Correlation = TOTAL DIFFERENCE ÷ MAXIMUM DIFFERENCE

Check

① Correlation = $141 \div 338 = 0.42$

② Suppose the total difference is 338

Correlation = $\underset{\uparrow \text{ Incorrect}}{338 \div 338 = 1}$

Try Correlation = (MAX DIFFERENCE − TOTAL DIFFERENCE) ÷ MAX DIFFERENCE

$= (338 - 141) \div 338 = 0.58$

1. Is Judith's statement about a strong correlation correct?
2. Is Judith's figure of 338 for the largest total possible difference in ranking correct?
3. Why is Judith's first correlation formula incorrect?
4. Is Judith's second correlation formula correct? Explain your answer.

How Did You Work?

1. Compare the sample responses and your group response. What are the advantages and disadvantages of each approach?

	Advantages	Disadvantages
Sam		
Nina		
Judith		
Our group work		

2. Now that you have seen Sam's, Nina's, and Judith's work, what would you do if you started the task again?

3. What do you think are the difficulties someone new to the task will face?

Drive-in Movie Theater

Jack, a movie theater owner, carried out three surveys.

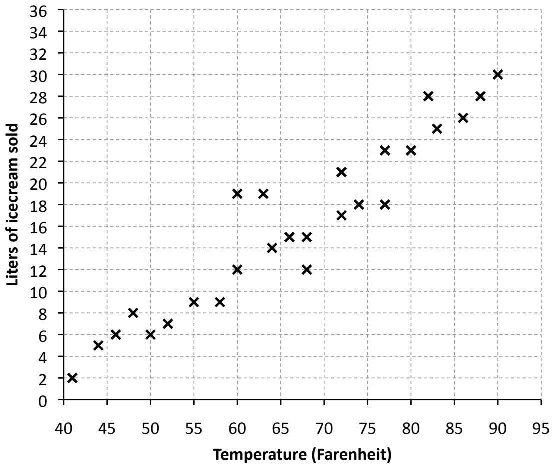
He plotted the results of the surveys as three scatter graphs, Scatter Graph A, Scatter Graph B, and Scatter Graph C. These are given on the pages that follow.

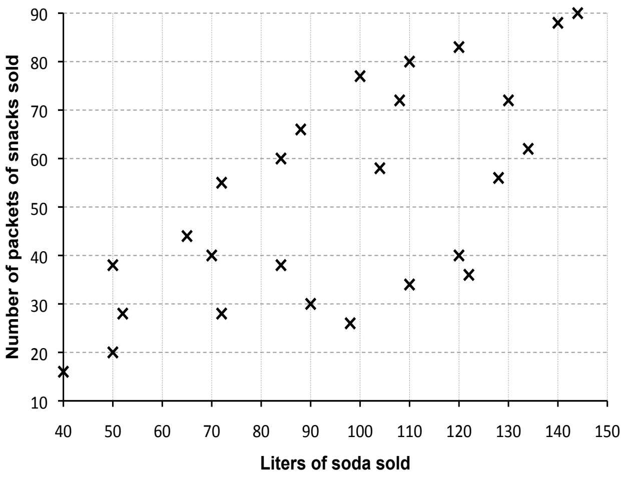
1. Describe the correlation that each graph shows.

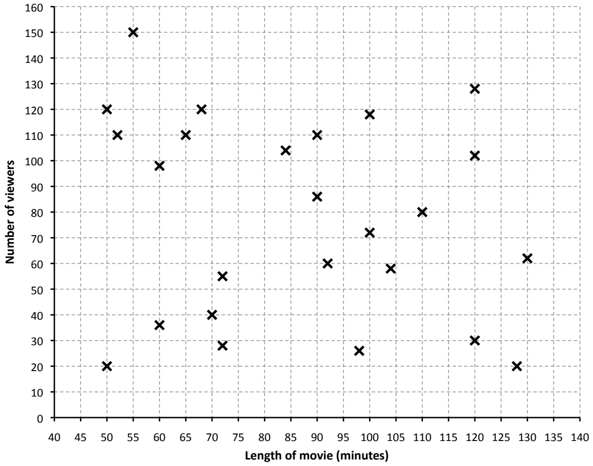
Scatter Graph A: ____________________

Scatter Graph B: ____________________

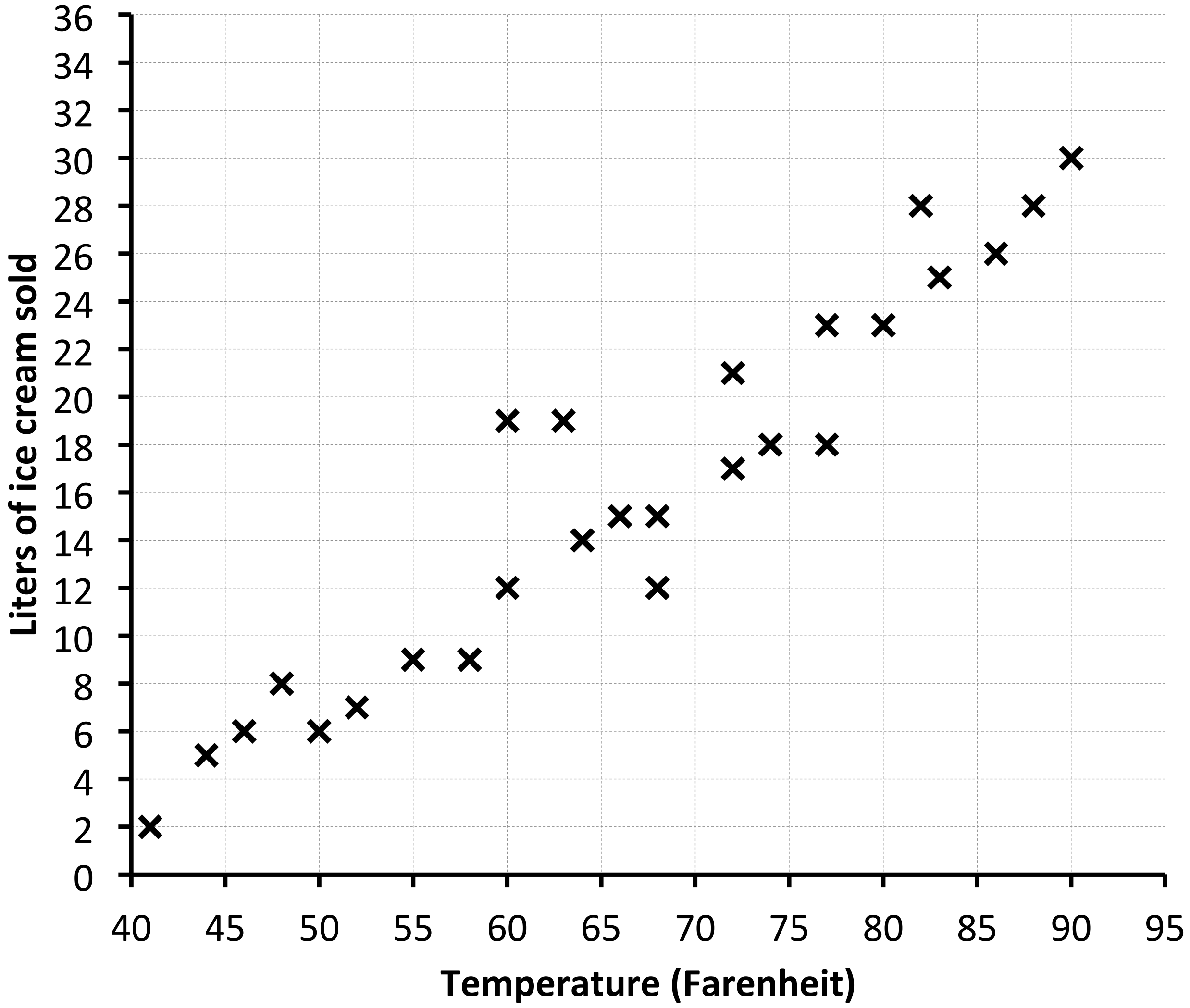
Scatter Graph C: ____________________





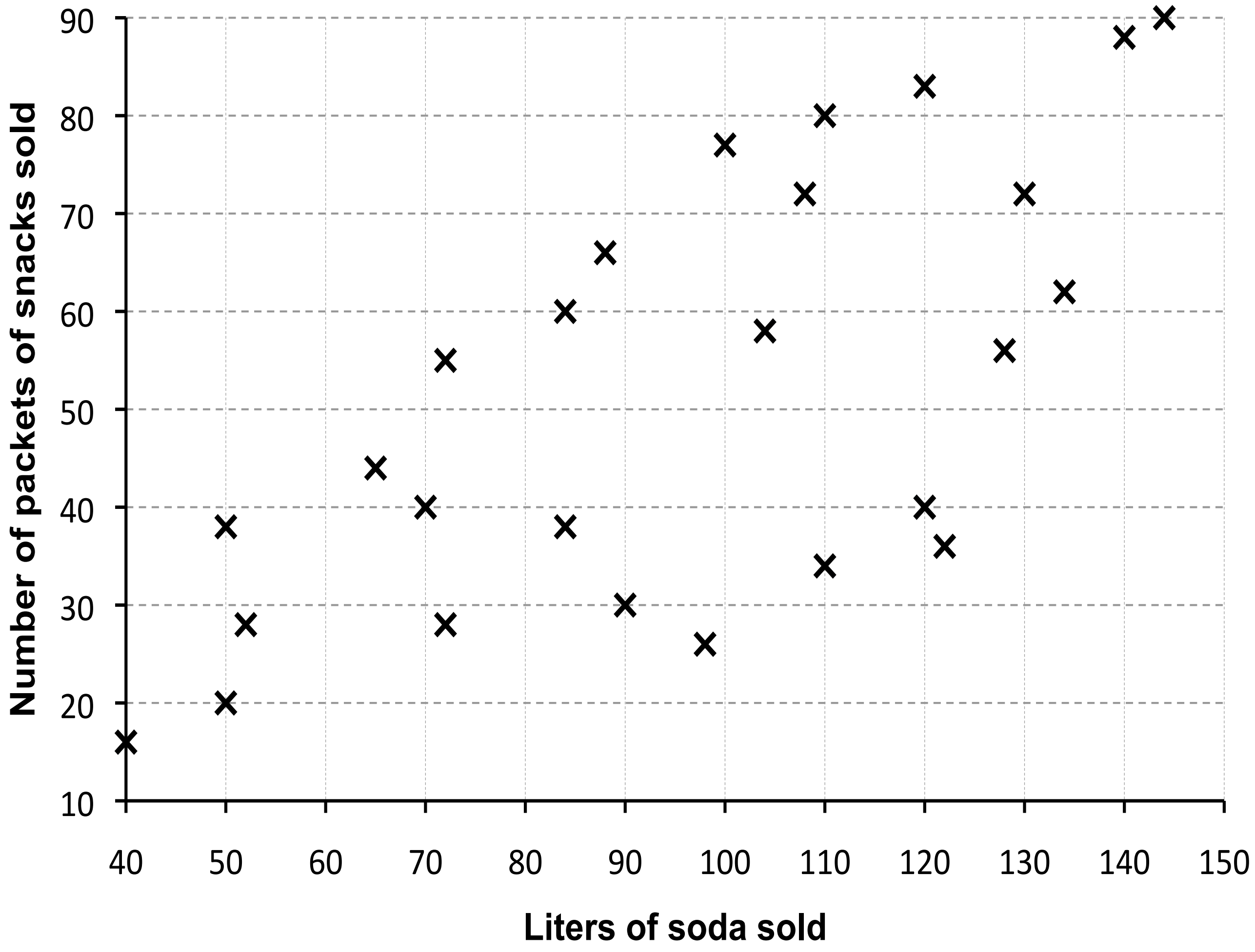


A: Ice cream sold v Outside temperature



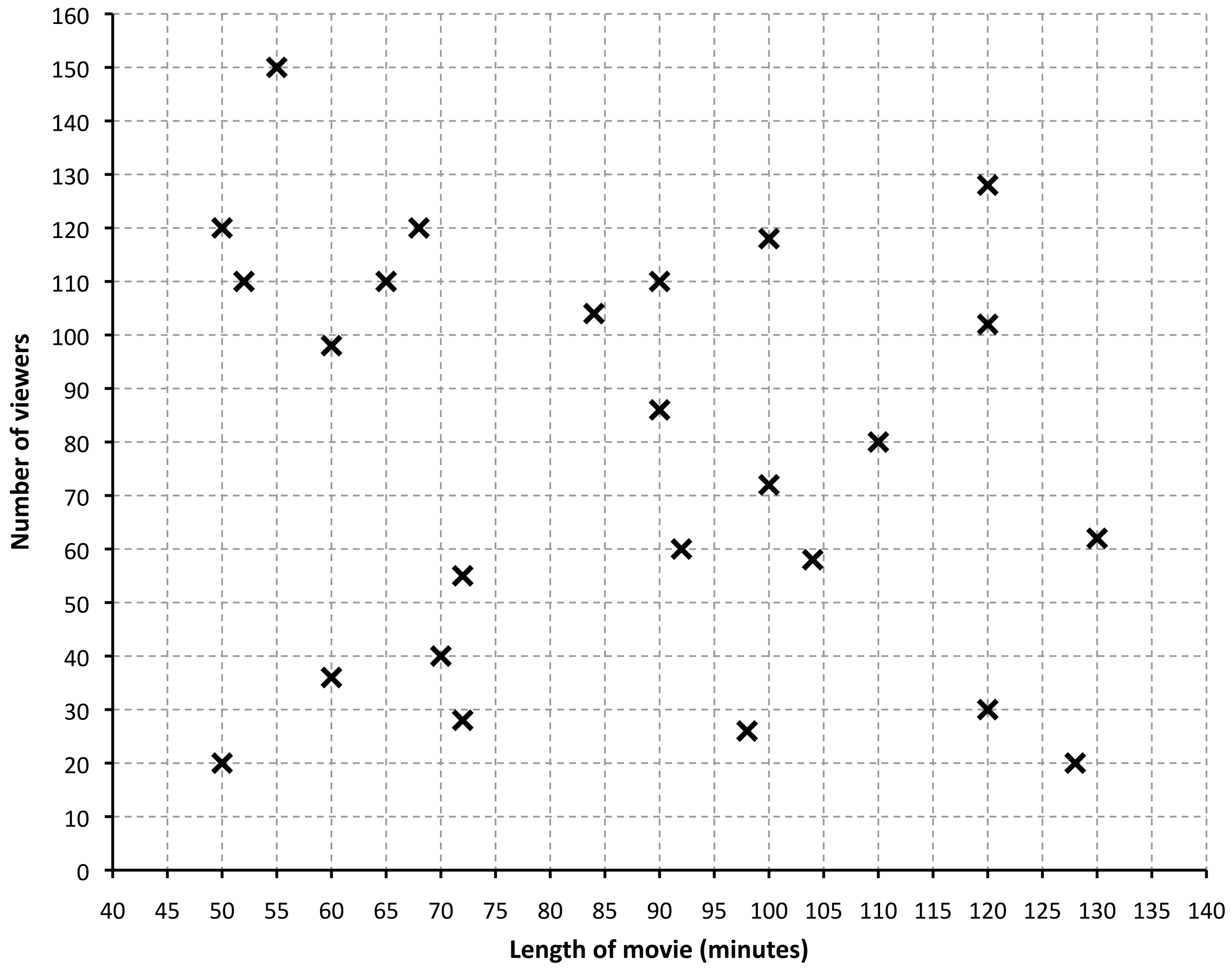
Temperature (Fahrenheit)	Liters of ice cream sold
41	2
83	25
64	14
52	7
72	17
68	15
50	6
55	9
74	18
66	15
68	12
82	28
72	21
88	28
86	26
60	19
90	30
80	23
77	23
44	5
46	6
58	9
63	19
48	8
60	12
77	18

B: Snacks sold v Soda sold



Liters of soda sold	Number of packets of snack sold
110	34
72	28
122	36
110	80
50	38
70	40
84	60
90	30
50	20
104	58
120	20
65	44
72	55
40	16
130	72
84	38
52	28
134	62
128	56
88	66
100	77
120	83
140	88
144	90
108	72
98	26

C: Number of viewers v Length of movie



Length of movie (minutes)	Number of viewers
55	150
72	28
60	36
110	80
50	120
70	40
92	60
90	110
50	20
104	58
68	120
65	110
72	55
60	98
150	72
84	104
52	110
130	62
128	20
90	86
100	118
120	128
120	102
120	30
100	72
98	26

Drive-in Movie Theater

2. Jack wants to find a way of measuring the strength of the correlation for each graph.

Jack

Draw a polygon round all the points.
The area of this polygon is small when the correlation is strong.
The area is large when the correlation is weak.

So a good measure of the correlation is: $\dfrac{1}{\text{Area of polygon}}$

What are the problems with Jack's method?

Examples may help to describe these problems.

3. Describe a better method than Jack's for calculating the correlation for each graph.

Explain why your method is better.

Planning a Joint Method for Calculating Correlation

1. Take turns to explain your method for calculating a correlation.

2. Listen carefully to explanations.
   - Ask questions if you don't understand.
   - Ask how they have checked their method.

   - Discuss with your partners:
     - What you like/dislike about your partner's math.
     - Any assumptions your partner has made.
     - How their work could be improved.

3. Once everyone in the group has explained their method, plan a joint method that is better than each of the separate methods.

4. Write an outline of your plan on one side of your sheet of paper.

Evaluating Sample Responses to Discuss

1. Work together on one sample response at a time.
   Each sample response is about Scatter Graph B.

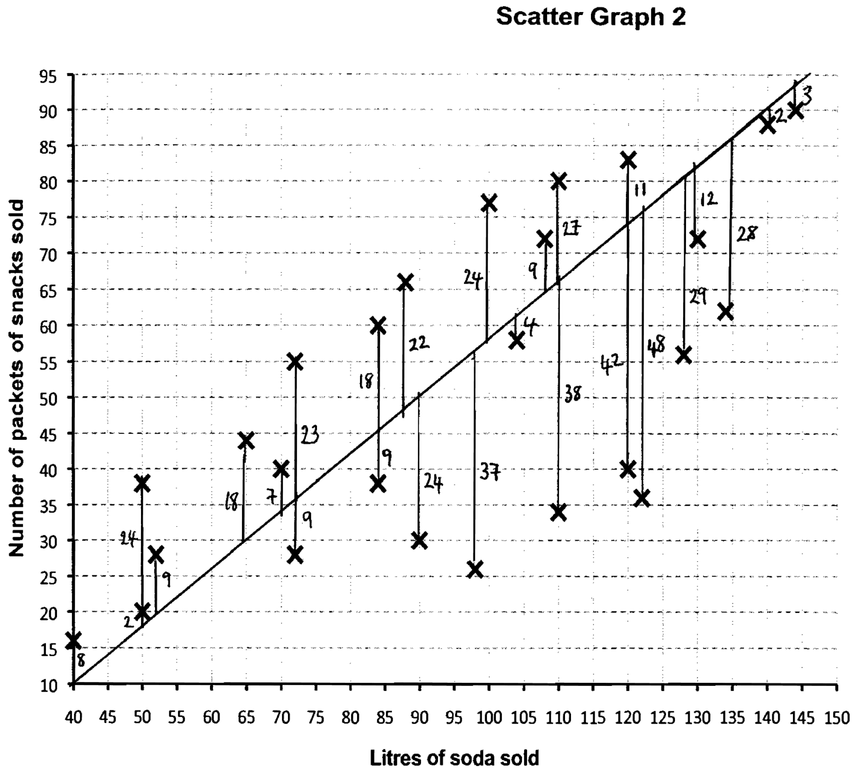
2. In your group, try to answer the questions below each sample response.

3. Listen carefully to each other's explanations.
   Ask each other questions if you don't understand.

4. Test the method shown on copies of Scatter Graph A and B.

5. When everyone is satisfied with the explanations, write your answers on a separate piece of paper.

Sample Responses to Discuss: Sam



Total distance = 3
2
28
12
29
48
11
42
38
27
9
4
\+ 24
37
22
24
18
9
23
9
7
18
9
24
2
8

487

(measured in mm)

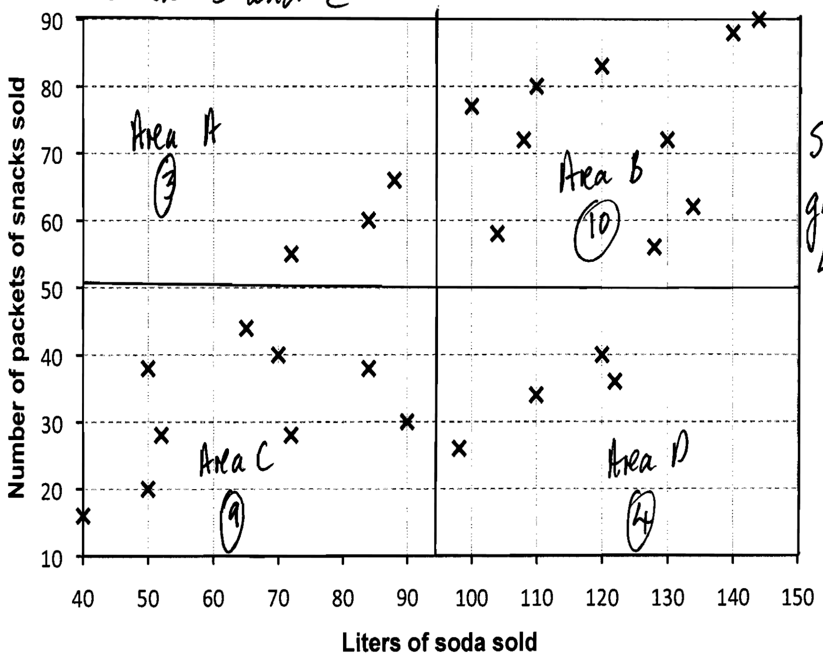
A total distance of zero means a perfect correlation. As the distance increases so the correlation gets weaker

Total distance = 487

Correlation = $\frac{1}{487}$ = 0.0021

Sample Responses to Discuss: Nina

For a strong correlation most crosses should be in B and C



Split the graph into 4 equal areas

Total strong correlation : 10+9=19
Total weak correlation : 3+4=7

Correlation =
7÷19 = 0.37

Test method

Suppose all data is in B and C

Total strong = 26
Total weak = 0

Correlation = 0÷26
= 0

not correct

Maybe correlation is total strong ÷ total weak

Sample Responses to Discuss: Judith

Liters of soda sold		Number of packets of snack sold		Difference
110	17	34	8	9
72	7	28	5	2
122	21	36	9	12
110	17	80	23	6
50	2	38	10	8
70	6	40	12	6
84	9	60	17	8
90	12	30	7	5
50	2	20	2	0
104	15	58	16	1
120	19	20	2	17
65	5	44	13	8
72	7	55	14	7
40	1	16	1	0
130	23	72	20	3
84	9	38	10	1
52	4	28	20	16
134	24	62	18	6
128	22	56	15	7
88	11	66	19	8
100	14	77	22	8
120	19	83	24	5
140	25	88	25	0
144	26	90	26	0
108	16	72	20	4
98	13	26	4	9
				141

Ranked in order, starting with the smallest

TOTAL

With a strong correlation the difference between the ranks should be small.

Minimum difference = 0

Maximum difference is when all 26 ranks are equal to 13

= 13 × 26 = 338

Correlation =

TOTAL DIFFERENCE ÷ MAXIMUM DIFFERENCE

Check

① Correlation = 141 ÷ 338 = 0.42

② Suppose the total difference is 338

Correlation = 338 ÷ 338 = 1

Incorrect

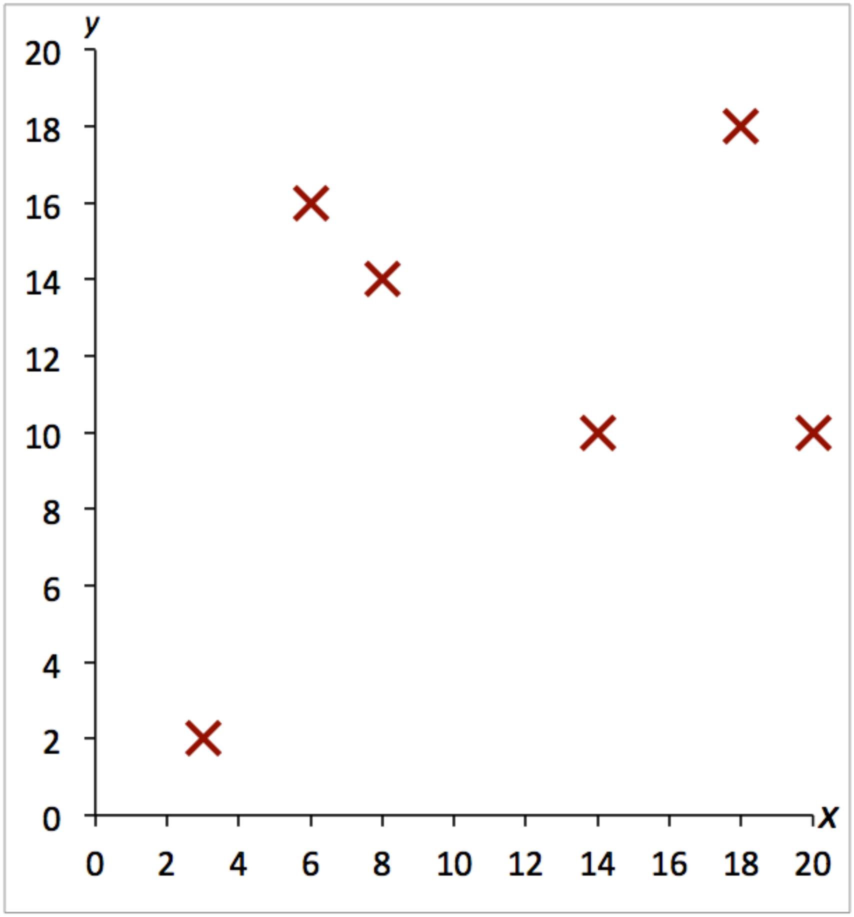
Try Correlation =

(MAX DIFFERENCE − TOTAL DIFFERENCE) ÷ MAX DIFFERENCE

= (338 − 141) ÷ 338 = 0.58

Spreadsheet Example

x - values	**y - values**
3	2
20	10
8	14
14	10
18	18
6	16



Correlation: 0.372

Mathematics Assessment Project

Classroom Challenges

These materials were designed and developed by the
Shell Center Team at the Center for Research in Mathematical Education
University of Nottingham, England:

Malcolm Swan,
Nichola Clarke, Clare Dawson, Sheila Evans, Colin Foster, and **Marie Joubert**
with
Hugh Burkhardt, Rita Crust, Andy Noyes, and **Daniel Pead**

We are grateful to the many teachers and students, in the UK and the US,
who took part in the classroom trials that played a critical role in developing these materials

The classroom observation teams in the US were led by
David Foster, Mary Bouck, and **Diane Schaefer**

This project was conceived and directed for
The Mathematics Assessment Resource Service (MARS) by
Alan Schoenfeld at the University of California, Berkeley, and
Hugh Burkhardt, Daniel Pead, and **Malcolm Swan** at the University of Nottingham

Thanks also to Mat Crosier, Anne Floyde, Michael Galan, Judith Mills, Nick Orchard, and Alvaro Villanueva who contributed to the design and production of these materials

This development would not have been possible without the support of
Bill & Melinda Gates Foundation

We are particularly grateful to
Carina Wong, Melissa Chabran, and **Jamie McKee**

The full collection of Mathematics Assessment Project materials is available from

http://map.mathshell.org