Math Objectives

- Students will explore both Pearson's Product Moment Correlation Coefficient (**PPMCC**) and Spearman's Rank Correlation Coefficient (**SRCC**), how to decide which will fit the given scenario better, and the appropriateness and limitations of each.

- Students will be asked to use their handhelds to discuss and analyze given sets of data, and describe their results in the context of the problems.

- Students will try to make a connection with how to understand these topics in IB Mathematics courses and on their final assessments.

Vocabulary

- Correlation
- Monotonic
- Causality

About the Lesson

- This lesson is aligning with the curriculum of IB Mathematics Applications and Interpretations and IB Mathematics Approaches and Analysis SL/HL

- This falls under the IB Mathematics Content Topic 4 Statistics and Probability:

  AI 4.10: (a) Spearman's rank correlation coefficient, r_s.

  (b) Awareness of the appropriateness and limitations of **PPMCC** and **SRCC**, and the effect of outliers on each.

  AA 4.10: (a) Equation of the regression line of x on y.

  (b) Use of this equation for prediction purposes.

  As a result, students will:

  - Apply this information to real world situations.

TI-Nspire™ Navigator™

- Transfer a File.
- Use Class Capture to examine patterns that emerge.
- Use Live Presenter to demonstrate.
- Use Teacher Edition computer software to review student documents.
- Use Quick Poll to assess students' understanding.

Tech Tips:

- This activity includes screen captures taken from the TI-Nspire CX II handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.

- Watch for additional Tech Tips throughout the activity for the specific technology you are using.

- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:
Student Activity
Nspire-PearsonsvsSpearmans-Student.pdf
Nspire-PearsonsvsSpearmans-Student.doc

Activity Materials

Compatible TI Technologies:

TI-Nspire™ CX Handhelds,

TI-Nspire™ Apps for iPad®,

TI-Nspire™ Software

In this activity, you will discuss and discover when and how to use both the Pearson's product moment correlation coefficient and the Spearman's rank correlation coefficient. You will learn about the benefits and limitations of these statistics and how they can be applied to real world situations.

First, a little background on both correlation coefficients. **Pearson's Product Moment Correlation Coefficient (PPMCC)** is a measure of the strength of a linear relationship between two variables that have been measured on interval or ratio scales. It is denoted by r. A **PPMC** attempts to draw a line of best fit through the data of two variables and the **PPMCC**, r, indicates how far away all these data points are to this line of best fit. The r can be a range of values from -1 to 1. A value of 0 indicates no correlation, when $r < 0$, there is a negative correlation, and when $r > 0$, there is a positive correlation. The closer r is to -1 or 1, the stronger the correlation.

Spearman's Rank Correlation Coefficient (SRCC) is a non-parametric measure of rank correlation, a statistical dependence between the rankings of two variables, and it assesses how well the relationship between two variables can be described using a monotonic function (data that is either all increasing or all decreasing). The **SRCC**, denoted by r_s, between two variables is equal to the **PPMCC** between the rank values of those two variables. **SRCC** will be high when observations have a similar (or identical for a correlation of 1) rank between two variables and will be low when the observations have a dissimilar rank between the two variables.

**Remember that both Pearson's and Spearman's measures strength and direction of the relationship between two variables, but Pearson's assesses the linear relationship and Spearman's evaluates the monotonic relationship. Pearson's works with raw data and Spearman's works with rank-ordered variables. The table below should be using when describing the strength of the r value.

r-value	Correlation
$0 < \lvert r \rvert \le 0.25$	Very Weak
$0.25 < \lvert r \rvert \le 0.50$	Weak
$0.50 < \lvert r \rvert \le 0.75$	Moderate
$0.75 < \lvert r \rvert \le 1$	Strong

education.ti.com

Teacher Tip: This activity will rely heavily on the students' capabilities to use the handheld efficiently and be able to navigate through the List & Spreadsheets page, the Data & Statistics page, and the menus for each. Some prep work may be needed to review these skills and the multiple ways students can analyze two variable statistics.

Problem 1 – Practicing Pearson's

The following table provides data collected by the mathematics and science departments at an IB Diploma school. The teachers have selected 10 students and recorded their semester 1 grades both Math and Physics.

Student	1	2	3	4	5	6	7	8	9	10
Maths Grade	67	77	27	97	47	37	37	67	57	57
Physics Grade	75	88	53	83	83	65	55	78	83	90

When dealing with data and using **PPMCC**, there is an appropriate order to analyzing the statistics. This problem will take you through that order, using the handheld for each step.

1. On your handheld, create a scatter plot of the data by entering it into a **List & Spreadsheets** page. Remember to title each column you use. Add a **Data & Statistics** page. Remember to assign both a horizontal and vertical variable to your scatterplot. Discuss with a classmate what you see. Describe any correlation you see.
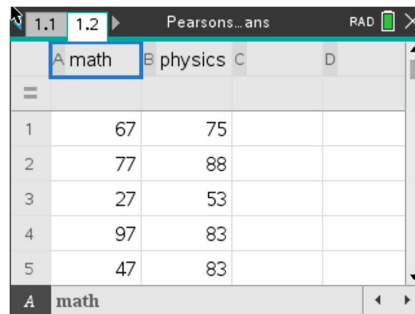
Possible Description: There seems to be a moderate positive linear correlation.
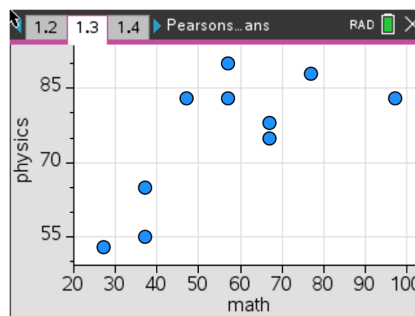
2. Now that you have discussed the correlation, it is time to find the statistics.
   (a) Confirm the strength and direction of your correlation by finding the **PPMCC** or r value on your handheld.

   Solution: $r = 0.712$ (which confirms that this is a moderate positive correlation)

(b) Explain to the class at least two ways to do this on the handheld.

Possible Explanation: Using **menu**, **Statistics**, **Stat Calculations**, **Two-Variable Statistics** or **Linear Regression**, enter the headings of the columns with your data. This can be done on the **List &Spreadsheet** page or on a **Calculator page**.

(c) Describe in the context of the problem, what this r value means.

Possible Description: Since the r-value is moderate and positive, as the Maths grades increase, it is moderately likely that the Physics grades will also increase.

3. Based on your results from number 2, discuss with a classmate if it would be appropriate to find a line of best fit for this Maths and Physics Grade data. Explain why or why not.

Possible Discussion: Since the r-value is above .5, it would be appropriate to find a line of best fit for this data.

4. If you and your classmate concluded that it would be appropriate to find the line of best fit:
(a) Find this line.
Solution: $y = 0.448x + 49.8$

(b) Explain to the class at least two ways to accomplish this on the handheld.
Possible Explanation: Using **menu**, **Statistics**, **Stat Calculations**, any of the **Linear Regression** commands, enter the headings of the columns with your data. This can be done on the **List &Spreadsheet** page or on a **Calculator page**. You can also create a scatter plot on **Data & Statistics** page and press **menu**, **Analyze**, **Regression** and select the line of your choice.

(c) In the context of the problem, explain what the gradient and y-intercept of the line of best fit represent.

Possible Explanation: The Physics grade increases 0.448 for every one point the Maths grade increases.

5. Using your line of best fit, make the following predictions:
(a) Find a student's Physics grade given that they received an 82 in their Maths class.
Solution: $y\,(Physics\ grade) = 0.448(82) + 49.5 = 86.2$

(b) Find a student's Maths grade given that they received a 41 in their Physics class.

Solution: $41 = 0.448x + 49.8$ $x \ (Maths \ grade) = -19.6$

(c) Discuss with a classmate, which, if either, of the answers to parts (a) and (b) are appropriate. Explain.

Possible Discussion: The input value (82) in "a" is within the given set of inputs (called interpolation) and is appropriate. The output value (41) in "b" is outside the given set of outputs (called extrapolation) and is not appropriate, especially since the grade was a negative value.

6. Discuss with a classmate the advantages and limitations for using Pearson's Product Moment Correlation Coefficient. Explain your conclusions.

Possible Explanation: Advantages – when comparing two items, knowing if there is a straightforward linear relationship could make understanding and describing easier with words like positive/negative or weak/moderate/strong. Limitations – this r-value is for a linear relationship only and if the relationship is not linear, this r-value is not help. Correlation does not imply causality, within this problem, a good Maths grade does not cause you to have a good Physics grade, there is just a correlation found. Also affected by outliers.

Problem 2 – Practicing Spearman's

Two judges at a county fair scored ten entries of the "best pie" competition in the table below out of 50 points.

Entry	A	B	C	D	E	F	G	H	J	K
Judge 1	43	25	45	37	21	35	30	28	34	29
Judge 2	32	27	46	47	28	36	40	24	30	33

1. Using the table below, fill in the remaining ranks for both Judge 1 and Judge 2.

Entry	A	B	C	D	E	F	G	H	J	K
Judge 1	2	9	1	3	10	4	6	8	5	7
Judge 2	6	9	2	1	8	4	3	10	7	5

Solution: In bold in the table.

2. Using the judge's rankings, find the r_s value by inputting the ranks on a **List & Spreadsheets** page and finding this value as you found the r value in problem 1.
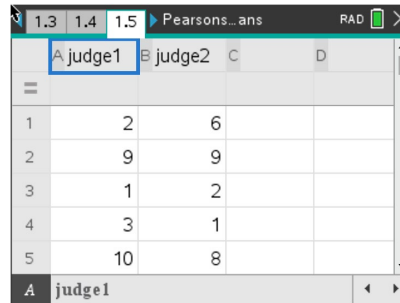
Solution: $r_s = 0.721$

3. Discuss this value with a classmate.
   (a) Describe the correlation.

   Solution: moderate and positive correlation

   (b) Describe what this r_s value means in the context of the problem.

   Solution: Judge 1 and Judge 2 moderately agree on the rankings of the pies. As Judge 1's rankings increase, Judge 2's rankings will likely increase.

Two more judges at the county fair scored the same ten entries of the "best pie" competition in the table below out of 50 points.

Entry	A	B	C	D	E	F	G	H	J	K
Judge 3	44	39	49	36	31	25	46	29	27	41
Judge 4	38	44	45	38	30	24	47	33	28	41

4. Using the table below, fill in the remaining ranks for both Judge 3 and Judge 4.

Entry	A	B	C	D	E	F	G	H	J	K
Judge 3	3	5	1	6	7	10	2	8	9	4
Judge 4	5.5	3	2	5.5	8	10	1	7	9	4

5. Discuss with a classmate what was different about this table as compared to the Judge 1 and 2 table. Explain.

Possible Explanation: Since there are two scores in Judge 4's group that are the same, you will take those two ranking spots and average them, thus giving you the 5.5's in both the 5th and 6th ranking spots.

6. Using the judge's rankings, find the r_s value by inputting the ranks on a **List & Spreadsheets** page and finding this value as you found the r value.

Solution: $r_s = 0.912$

7. Discuss this value with a classmate.
(a) Describe the correlation.

Possible Description: Very strong positive correlation.

(b) Describe what this r_s value means in the context of the problem.

Possible Description: Judge 3 and Judge 4 strongly agree on the rankings of the pies. As Judge 3's rankings increase, Judge 4's rankings will also increase.

8. Discuss with a classmate the advantages and limitations for using Spearman's Rank Correlation Coefficient. Explain your conclusions.

Possible Discussions: Advantages – Used on non-linear data, it can be used on data that has been ranked even if the original data is unknown or cannot be quantified, and not greatly affected by outliers. Limitations – It can handle only a limited number of observations or items, not useful for non-monotonic situations.

Further IB Application

One of the events at the Women's National Gymnastic Championship is the balance beam. The following judges awarded 8 participants the scores in the table below out of a maximum of 10 points.

Gymnast	A	B	C	D	E	F	G	H
Judge 1 (x)	8.2	5.6	7.1	6.7	8.4	9.3	9.0	8.5
Judge 2 (y)	8.9	6.9	7.6	7.9	9.1	9.5	9.2	7.9

a. (i) Write down the value of the Pearson's Product Moment Correlation Coefficient, r.

Solution: $r = 0.892$

(ii) Using the value of r, interpret the relationship between Judge 1's score and Judge 2's score.

Solution: Strong and Positive

b. Write down the equation of the regression line y on x.

Solution: $y = 0.651x + 3.268$

c. (i) Use your regression equation from part (b) to estimate Judge 2's score when Judge 1's score awards a perfect 10.

Solution: $y(Judge\ 2) = 9.78$

(ii) State whether this estimate is reliable. Justify your answer.

Solution: This score is reliable as it still falls under the highest available score of a perfect 10.

The head of the rules committee for the event would like to find the Spearman's Rank Correlation Coefficient.

d. Complete the information in the following table.

Solution: In bold.

Gymnast	A	B	C	D	E	F	G	H
Judge 1's Rank	**5**	8	**6**	7	**4**	1	**2**	3
Judge 2's Rank	**4**	8	7	**5.5**	3	1	2	**5.5**

e. (i) Find the value of the Spearman's Rank Correlation Coefficient, r_s.

Solution: $r_s = 0.862$

(ii) Comment on the result obtained for r_s.

Solution: Judge 1 and Judge 2 agree on the ranking of the gymnasts.

f. The head of the rules committee believes that Judge 2's score for gymnast B is too high and so decreases the score from 6.9 to 6.5. Explain why the value of the Spearman's Rank Correlation Coefficient, r_s, does not change.

Solution: Decreasing the score to 6.5, does not change the rank of gymnast B.

Teacher Tip: This activity might have to be adjusted for the IB AA courses as they do not have to know Spearman's for the exam.

Teacher Tip: Throughout this activity, the students are asked to discuss with classmates and explain how they achieved their answers. This is a wonderful opportunity to create a student led classroom. As you float around the room, listen to what they are saying, add to their discussions, and give them leading questions to see how they respond.

Teacher Tip: Basic understanding of the TI-Nspire CX II is needed to navigate through pages and menus, especially the **Calculator, List and Spreadsheets, and Data and Statistics** pages. It may be a good idea to review these before the activity.

TI-Nspire Navigator Opportunity: *Quick Poll (Open Response)*
Any part to any Problem in the activity would be a great way to quickly assess your student's understanding of finding and discussing both forms of Scientific Notation and Expanded Form.

**Note: This activity has been developed independently by Texas Instruments and aligned with the IB Mathematics curriculum, but is not endorsed by IB™. IB is a registered trademark owned by the International Baccalaureate Organization.