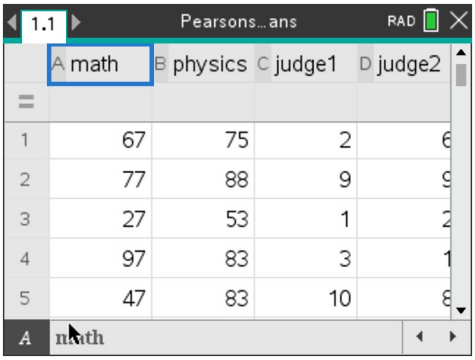
Pearson's vs. Spearman's
Student Activity

Name _____

Class _____

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.

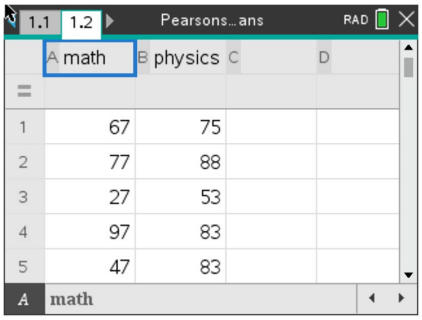
r-value	Correlation
$0 < \|r\| \leq 0.25$	Very Weak
$0.25 < \|r\| \leq 0.50$	Weak
$0.50 < \|r\| \leq 0.75$	Moderate
$0.75 < \|r\| \leq 1$	Strong

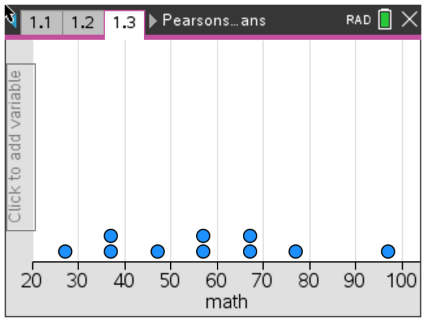
Problem 1 – Practicing Pearson's

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.

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.

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

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

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.

4. If you and your classmate concluded that it would be appropriate to find the line of best fit:
   (a) Find this line.

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

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

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.

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

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

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

Problem 2 – Practicing Spearman's

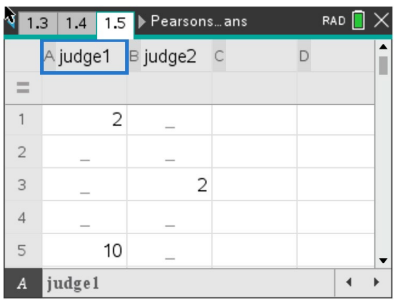
Two judges at a county fair scored ten entries of the "best pie" competition in the table below out of 50 points.

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				10					
Judge 2			2					10		

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.



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

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

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								9	
Judge 4		3							9	

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

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.

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

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

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

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.

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

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

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

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

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.

Gymnast	A	B	C	D	E	F	G	H
Judge 1's Rank		8				1		
Judge 2's Rank		8				1		

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

(ii) Comment on the result obtained for r_s.

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.