1.3 Notes: Linear Regression and Correlation Coefficient

Linear Regression- a straight line that attempts to predict the relationship between two points

Discrete Data - data that has distinguishable spaces between possible values.

Continuous Data - data that can be measured as finely as is practical (no spaces between values).

Correlation Coefficient - a number between −1 and +1 calculated so as to represent the linear dependence of two variables or sets of data.

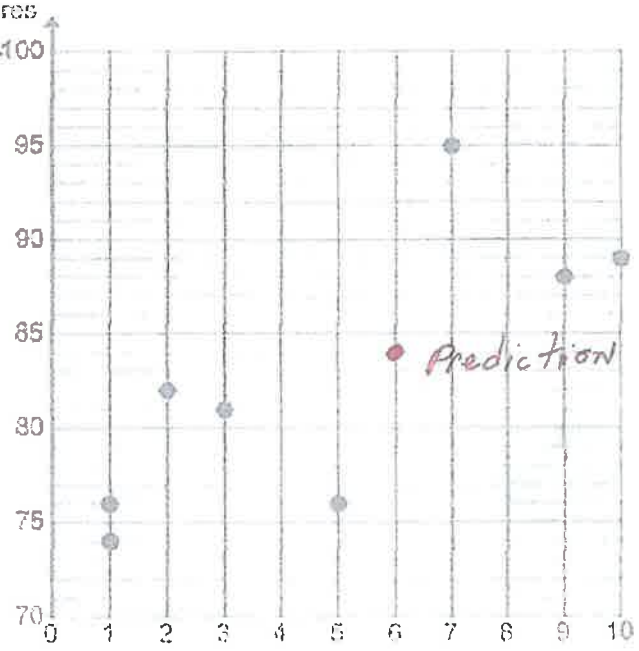
Learning Target: I can analyze the correlation coefficient to determine the strength of the equation for the line of best fit.

Example 1: Students collected the following data of studying and test scores.

1. Find the equation of the line that best fits the given data.

- Make a list (table) of the data
- STAT>EDIT>ENTER
- Fill L_1 (X List) and L_2 (Y List)
- MODE: Check to make sure STAT DIAGNOSTICS is set to ON
- STAT>CALC>LINREG (Click ENTER through the ~~list~~ 1.5t)
- a= 1.58 (slope) b= 74.9 (Y-Int)
- Equation: y = 1.58x + 74.9 (Y= form)
- r= .786 (correlation coefficient)

X	Y
1	74
1	76
2	82
3	81
5	76
7	94
9	88
10	89



2. Does this show a strong or a weak correlation? Explain. (See Correlation Chart)
Weak: all points are spread out. Not a tight line.

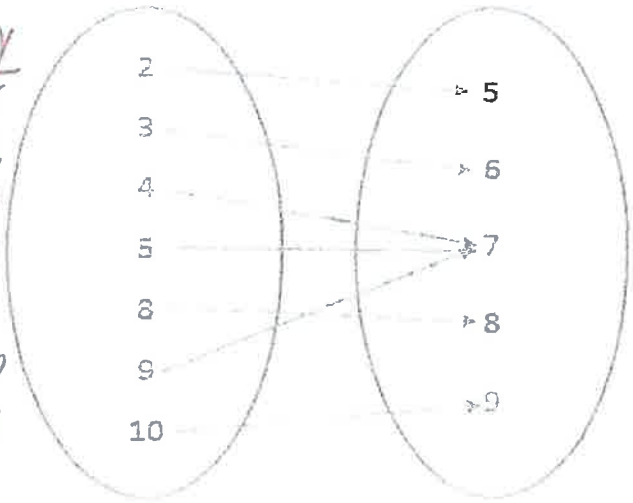
3. Is the data discrete or continuous?
Discrete
(AN argument might be made for continuous → you CAN measure ½ hours + pt point)

4. Use your equation to predict Sally's score if she studies for 6 hours.
y = 1.58(6) + 74.9
y = 84%

1.3 Notes: Linear Regression and Correlation Coefficient

Example 2: The data shows the number of tardies Mr. Pride recorded for his first period class on a given day of the school year.

y
5
6
7
7
8
7
9



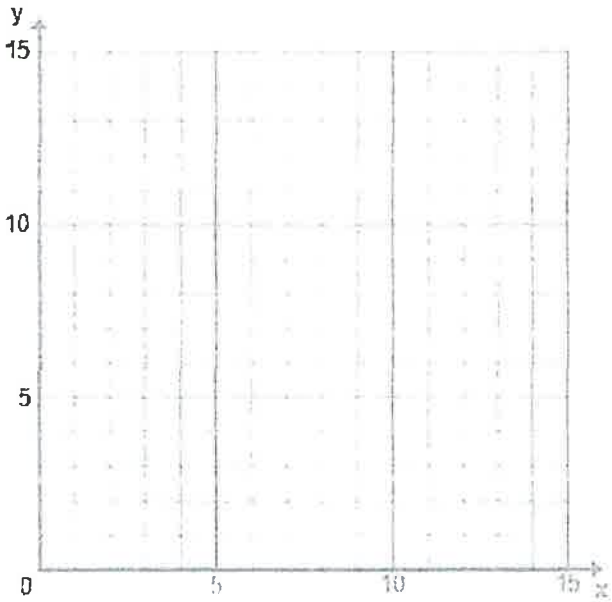
1. Find the equation of the line that best fits the data

   m= .356 b= 4.91

   Equation $y = .356x + 4.91$

2. What is the correlation coefficient?

   r= .86



3. Does this show a strong or weak correlation? Explain.

   Strong. .86 is close to 1.
   The graph shows points that are close together in almost a line.

4. Is the data discrete or continuous? Explain.

   Discrete → can't have 1/2 of a tardy.

5. Predict how many tardies Mr. Pride will have on day 100.

   $y = .356(100) + 4.91$

   $y = 40.5$ tardies

Textbook Reference: Pearson pp. 144-147