Section 3: Regression and Correlation

Objective: To see if there exists a linear relationship between **x** and **y** variables**.** **TI83_84 Calculator**

Steps to do:

1. Using Scatter Plot to see the possibility of any pattern and answering **4 key questions**
   a) To see if data exhibits a **linear pattern** or not
   b) To see if linear pattern is **positive or negative**
   c) To see how closely (**strongly or perfectly**) data are **clustered around** a **straight line.**
   d) To detect any **outlier** (a point that is lying far away from the other data points).
2. If a linear pattern exists, find **r = Correlation Coefficient** and explain it**.**

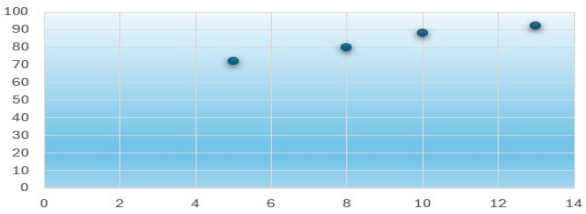
Coefficient of correlation	Correlation
0.0 - ±0.2	Little correlation
±0.2 - ±0.4	Weak correlation
±0.4 - ±0.7	Correlated
±0.7 - ±0.9	Strong correlation
±0.9 - ±1.0	Very strong correlation

3. Formulate the linear relationship by **equation y = ax + b**
4. Find **a** =slope, **b** = y-intercept by Formula/**TI83_84 Calculator**/Excel and make prediction.

Example 1: Is a relationship between **x** = hours of study and **y** = test score? **Online calculator**

1) Let's look at these data and plot them in X and Y axis called it as **scatter plot**

	x = Hours Study/week	y = Test Score
1	5	72
2	10	88
3	13	92
4	8	80



The pattern of data points **suggests** a strong linear positive correlation.

2) Find r = Correlation Coefficient by either Formula/Excel/ Ti
   r = **0.98** because it is close to 1, it **suggests** a very strong positive linear correlation
3) Find a =slope, b = y- intercept by either Formula/Excel/ Ti
   a = 2.59, b = 59.71 Regression equation **y = ax + b** **y = 2.59 x + 59.71**
4) Use this regression equation to make predictions

a) If someone studied **x = 12** hours, what will be his estimated test score
y = 2.59 x + 59.71 **y = 2.59 (12) + 59.71= 90.71**

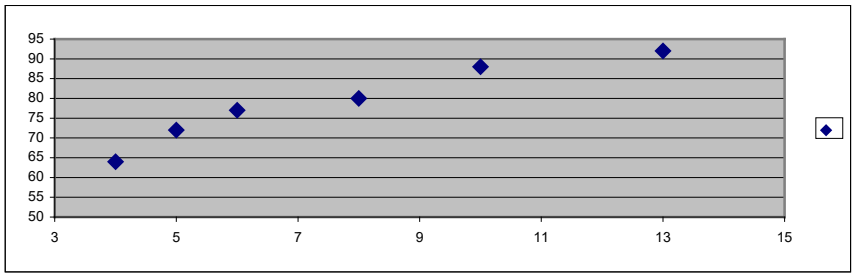
b) If someone studied **x = 20** hours, what will be his estimated test score
y = 2.59 x + 59.71 **y = 2.59 (20) + 59.71= 111.51 Impossible! Why?**

Because x = 20 is outside the scope of our data.

c) How many hours **x = ?** does one need to study to get **y = 85** on the test?
y = 2.59 x + 59.71 **85 = 2.59 (x) + 59.71** **25.29 =2.59x** **x = 9.76 hours**

Example 2: Is a relationship between **x** = hours of study and **y** = test score?

	x = Hours Study/week	y = Test Score	x^2	y^2	xy
1	5	72			
2	10	88			
3	13	92			
4	8	80			
5	6	77			
6	4	64			
	$\sum x = 46$	$\sum y = 473$	$\sum x^2 = 410$	$\sum y^2 = 37817$	$\sum xy = 3794$



Comment: A very strong positive linear correlation.

1. Compute the correlation coefficient and **<u>comment</u>** on that $r = 0.963$ Very strong...?

2. Compute the slope and y-intercept and write the equation of regression line. Slope = a = 2.92, y-itc = b = 56.41

$$y = a\,x + b = 2.92\,x + 56.41$$

3. Explain the slope based on the regression equation and the in relation of x and y variables.
In general, for every additional hour of study per week the score goes up by 2.92 points.

4. Compute average and standard deviation for both x and y variables. $\bar{x} = 7.67$, $\bar{y} = 78.83$, $s_x = 3.386$, $s_y = 10.28$
5. If one student studies 6 hours a week, use **Reg. Equ.** to estimate her test score. $x = 6, \quad y' = 73.93$

6. If one student has test score of 85, use **Reg. Equ.** to estimate number of hours he spends studying per week. $y = 85, \quad x' = 9.79$