

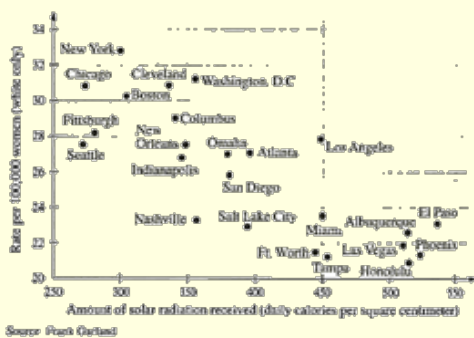
## S.8 Linear Correlation and Linear Regression

## S.8 Linear Correlation and Linear Regression

---

Consider each of these scatterplots. What story does each tell?

Amount of Sunlight  
vs. Breast Cancer Deaths




---



---

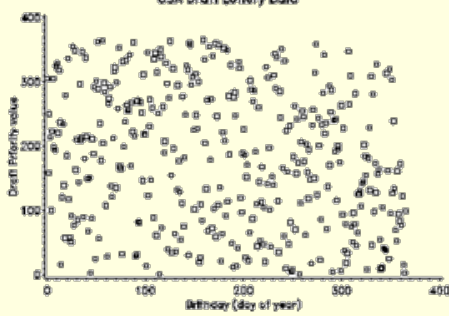


---



---

Military Draft Notices  
vs. Birth Day of the Year



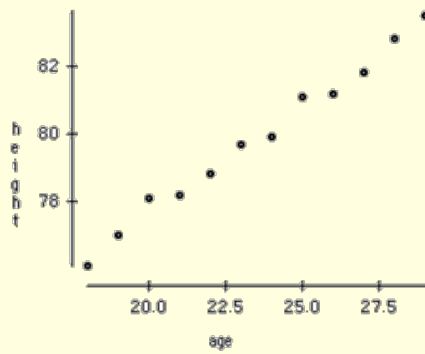
---

---

---

---

Age vs.  
Average Height in inches



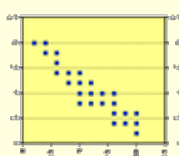
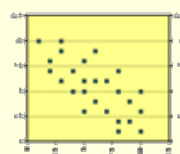
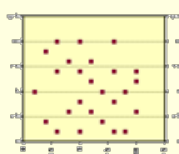
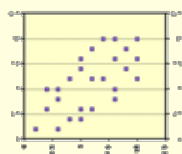
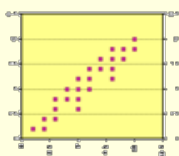
---

---

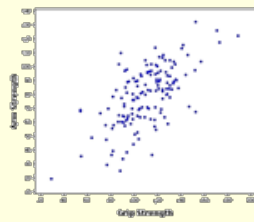
---

---

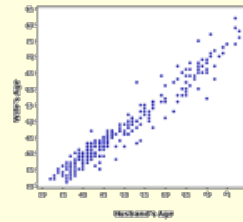
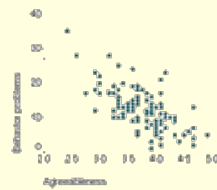
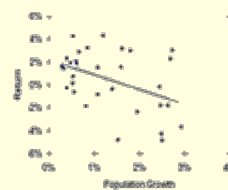
A scatterplot shows the correlation [relationship or association] between two variables. Correlations are identified by their direction [positive or negative] and their strength [weak or strong].



The scatter plot does not imply causation: one variable does not necessarily cause the other. It just shows whether they tend to move in the same or opposite directions.

Grip Strength  
and Arm Strength

Spouses Ages

Degree of Agreeableness  
vs. # of Behavior ProblemsPopulation Growth  
vs. GDP Growth

IB expects you to be able to:

- Draw a scatter gram from data
- Find the mean and standard deviation of the data
- Draw a line of best fit by eye in the general direction of the data:
  - through the mean [average] point of the data
  - balancing equal numbers of points above and below the line
- Predict an expected outcome using the line you have drawn

## IB Practice A

Eight students in Mr. O'Neil's Physical Education class did pushups and sit-ups. Their results are shown in the following table.

Student	1	2	3	4	5	6	7	8
number of pushups ( $x$ )	24	18	32	51	35	42	45	25
number of situps ( $y$ )	32	28	38	40	30	52	48	52

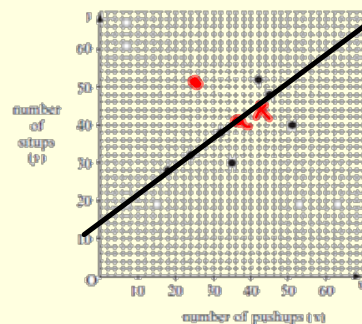
The graph to the right shows the results for the first seven students.

- a) What kind of correlation do the data show?

positive

What does this imply about the relationship between pushups and sit-ups?

↑ pushups, ↑ situps



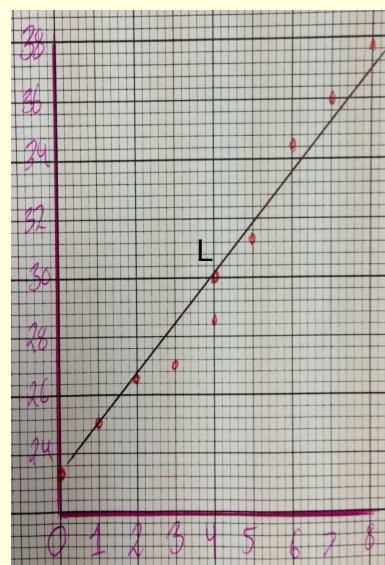
- b) Plot the results for the eighth student on the graph.
- c) Find the mean point  $(\bar{x}, \bar{y})$  of the data and plot it on the graph. Label it  $\bar{X}$ . (34, 40)
- d) Draw a line of best fit on the graph.
- e) A student can do 60 pushups. How many sit-ups can the student be expected to do?

## IB Practice B

In an experiment, a vertical spring was fixed at its upper end. It was stretched by hanging different weights on its lower end. The length of the string was then measured. The following readings were obtained.

Load (kg) $x$	0	1	2	3	4	5	6	7	8
Length (cm) $y$	23.5	25	26.5	27	28.5	31.5	34.5	36	37.5

- a) Plot these pairs of values on a scatter diagram, taking 1 cm to represent 1 kg on the horizontal axis and 1 cm to represent 2 cm on the vertical axis.
- b) i) Write down the mean value of the load ( $\bar{x}$ )  
 $\bar{x} = 4$
- ii) Write down the standard deviation of the load  
 $\sigma = 2.58$
- iii) Write down the mean value of the length ( $\bar{y}$ )  
 $\bar{y} = 30$
- iv) Write down the standard deviation of the length  
 $\sigma = 4.78$
- c) Plot the mean point ( $\bar{x}, \bar{y}$ ) on the scatter diagram. Label it  $L$ .
- d) Draw the line of best fit on your graph.
- e) Could you use your line to predict how far the spring will stretch with a load of 30 kg? Why or why not?



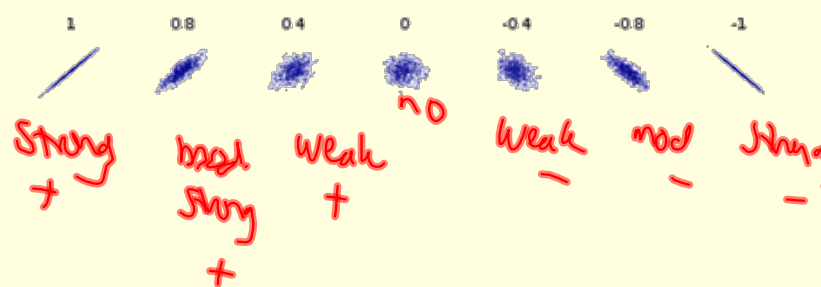
NO extrapolation.

### S.8 The Correlation Coefficient and the Regression Equation

We have been describing the correlation between two variables and sketching what seems to be the line of best fit. But as mathematicians, we want a more precise measure of correlation.

The Pearson's Product-Moment Correlation Coefficient gives us an exact number to compare correlations. It is designated by using the letter ' $r$ '. The diagrams below show some correlation coefficients for different sets of data.

Describe each correlation:



Your calculator can quickly give you the correlation coefficient [ $r$ ] and the regression line [ $y = ax + b$ ].

You will need to activate "DiagnosticOn" on your GDC:

2<sup>nd</sup> --> Catalog --> DiagnosticOn --> Enter --> Enter

The regression equation is found by using: Stat --> CALC --> LinReg(ax+b)

```

EDIT [2nd] TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg
  
```

```

LinReg(ax+b)
Xlist:L1
Ylist:L2
FreqList:
Store RegEQ:
Calculate
  
```



## IB Practice C:

Eight students in Mr. O'Neil's Physical Education class did pushups and sit-ups. Their results are shown in the following table.

Student	1	2	3	4	5	6	7	8
number of pushups ( $x$ )	24	18	32	51	35	42	45	25
number of situps ( $y$ )	32	28	38	40	30	52	48	52

- a) Find the Pearson's Product-Moment Correlation Coefficient. Describe the correlation in words.

$$r = 0.439, \text{ weak positive correlation}$$

- b) Find the equation of the regression line in the form  $y = mx + c$ . Indicate what the variables  $x$  and  $y$  stand for.

$$y = 0.373x + 27.3$$

# sit ups.  $\leftarrow$   $\rightarrow$  # p-ups

- c) Use your equation to predict how many sit-ups a student can be expected to do if she can do 60 pushups.

$$y = 0.373(60) + 27.3$$

$$y = 49.7 \text{ situps}$$

- d) A student does 100 pushups. Explain why you cannot use this equation to predict the number of sit-ups.

IB Practice D:

---

In an experiment, a vertical spring was fixed at its upper end. It was stretched by hanging different weights on its lower end. The length of the string was then measured. The following readings were obtained.

Load (kg) $x$	0	1	2	3	4	5	6	7	8
Length (cm) $y$	23.5	25	26.5	27	28.5	31.5	34.5	36	37.5

- Find the Pearson's Product-Moment Correlation Coefficient.  
Describe the correlation in words.
- Find the equation of the regression line in the form  $y = mx + c$ .  
Indicate what the variables  $x$  and  $y$  stand for.
- Use your equation to predict how long the spring will stretch with a load of 10 kg.
- Use your equation to predict how much mass is needed to stretch the spring 30 cm.

