Correlation and linear regression:

Correlation is the strength of relationship between 2 characteristics in a population, to be obtained needs the following:

*** One population**

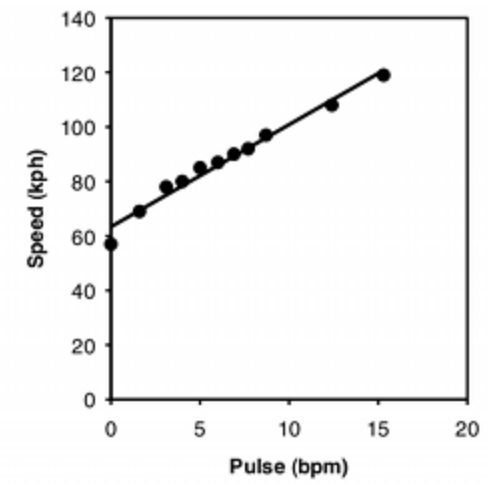
*** Two characteristics**

*** Both should be changing (variables are not constant)**

*** There must be some so sort of relationship between 2 in order to obtain the strength of this relationship.**

Need to determine which of 2 variables is X&Y according to following introduction

X	Y
Independent: the changing in X is independent on the change in Y	**Dependent: the change in Y is dependent on the change in X**
Less changing in a short period of time More constant	**More changing in a short period of time more changing**
As the cause	**As the effect**

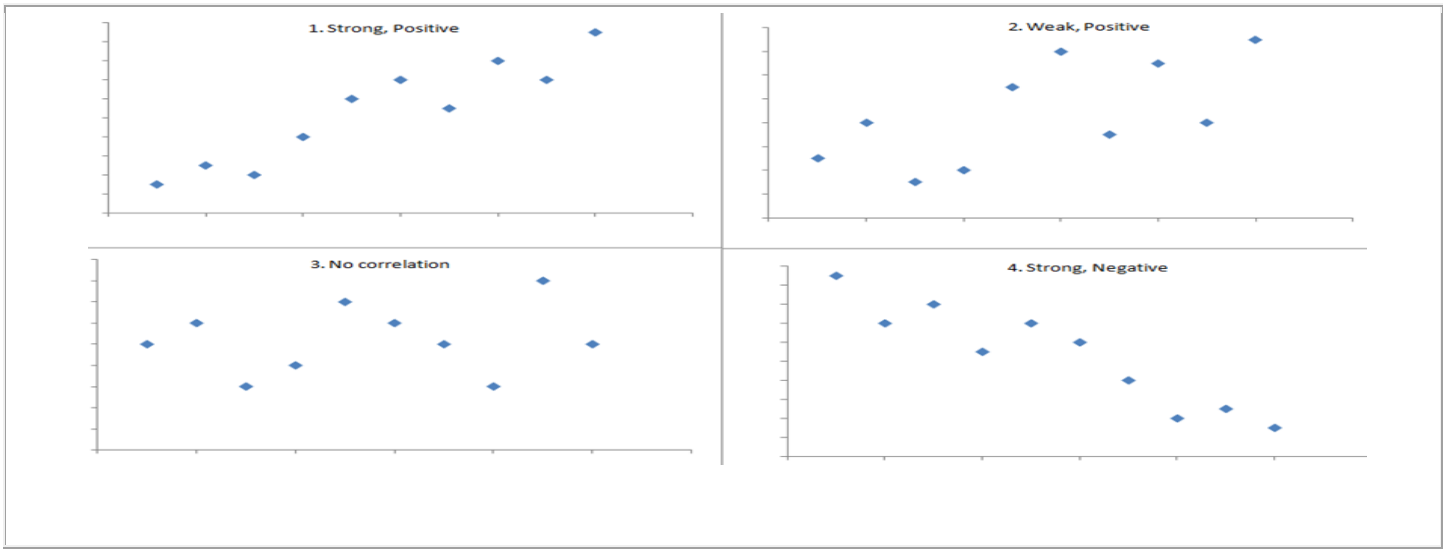


Graph of my pulse rate vs. speed on an elliptical exercise machine
(perfect direct positive correlation) r = 1

In this section we will first discuss correlation analysis, which is used to quantify the association between two continuous variables (e.g., between an independent and a dependent variable or between two independent variables). Regression analysis is a related technique to assess the relationship between an outcome variable and one or more risk factors or confounding variables. The outcome variable is also called the **response** or **dependent variable** and the risk factors and confounders are called the **predictors**, or **explanatory** or **independent variables**. In regression analysis, the dependent variable is denoted "*y*" and the independent variables are denoted by "*x*".

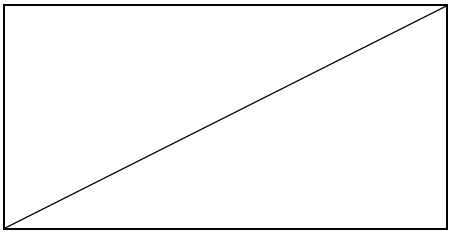
In correlation analysis, we estimate a sample **correlation coefficient**, more specifically the **Pearson Product Moment correlation coefficient**. The sample correlation coefficient, denoted **r**,Ranges between -1 and +1 and **quantifies the direction and strength of the linear association** between the two variables. The correlation between two variables can be positive (i.e., higher levels of one variable are associated with higher levels of the other) or negative (i.e., higher levels of one variable are associated with lower levels of the other).The sign of the correlation coefficient indicates the direction of the association. **The magnitude of the correlation coefficient indicates the strength of the association.**
For example, a correlation of r = 0.9 suggests a strong, positive association between two variables, where as a correlation of r = - 0.2 suggest a weak, negative association. A correlation close to zero suggests no linear association between two continuous variables.

If r < 3 no ---- correlation,
If r 0.3- < 0.5 ---- weak correlation,
If r 0.5 0.7 ---- moderate correlation,
If r 0.7 - 1 strong ve+ correlation,
If r - 1 strong ve- correlation

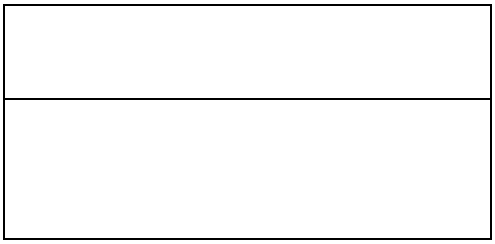


a strong positive association (r=0.9), similar to what we might see for the correlation between infant birth weight and birth length.

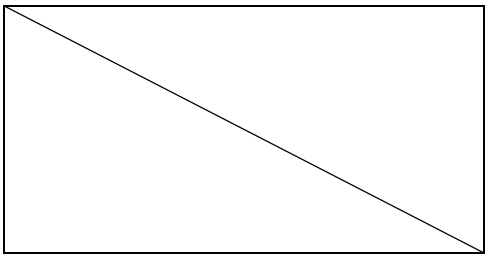
- a weaker association (r=0,2) that we might expect to see between age and body mass index (which tends to increase with age).
- the lack of association (r approximately 0) between the extent of media exposure in adolescence and age at which adolescents initiate sexual activity.
- the strong negative association (r= -0.9) generally observed between the number of hours of aerobic exercise per week and percent body fat.



Strong positive correlation



No correlation



Strong negative correlation

eg: the body weight and plasma volume of 8 healthy men are presented in this table: in general high plasma volume tends to be associated with high wt this relationship is measured by Pearson correlation :

No	Body wt kg		Plasma volume liter		
	X	X^2	Y	Y^2	X.Y
1	58	3364	2.75	7.56	159.50
2	70	4900	2.86	8.18	200.20
3	74	5476	3.37	11.36	249.38
4	63.5	4032.25	2.76	7.62	175.26
5	62	3844	2.62	6.86	162.44
6	70.5	497.25	3.49	12.18	246.05
7	71	541	3.05	9.30	216.55
8	66	4356	3.12	9.73	205.92
	$\Sigma x = 535$	Σx^2 =35983.5	Σy	Σy^2	$\Sigma x.y$ 1615.29

$$r = \frac{\mathbf{\Sigma}(\times - \bar{\times})\ (\mathbf{y} - \bar{\mathbf{y}})}{\sqrt{\mathbf{\Sigma}(\times - \times^{\mathbf{2}})\ .\times\ \mathbf{\Sigma}\ (\mathbf{y} - \mathbf{y^2})}}$$

$$\mathbf{r} = \frac{\mathbf{SPxy}}{\sqrt{(\mathbf{S}^2 \times . \mathbf{S}^2\mathbf{y})}}$$

SPxy = $\mathbf{\Sigma}(\times\ .\ \mathbf{y})\ -\ \mathbf{\Sigma}\ (\times).\mathbf{\Sigma}\ (\mathbf{y})/\boldsymbol{n}$ = 65.292 – (535x 24.02)/8 === 8.9545

$S^2 \times = \Sigma(\times^2) - (\Sigma \times)^2)/n$ = 35983.5 – $(535)^2/8$ = 205.375

$S^2y = \Sigma(y^2) - (\Sigma y)^2)/n$ = 72.798 – $(24.02)^2/8$ = 0.678

$$r = \frac{\Sigma(\times . y) - \Sigma(\times) . \Sigma(y)/n}{\Sigma(\times^2) - (\Sigma \times)^2)/n \,.\, \Sigma(y^2) - (\Sigma y)^2)/n}$$

= +0.759 ==== there is strong direct relationship between body wt & plasma volume