## Correlation

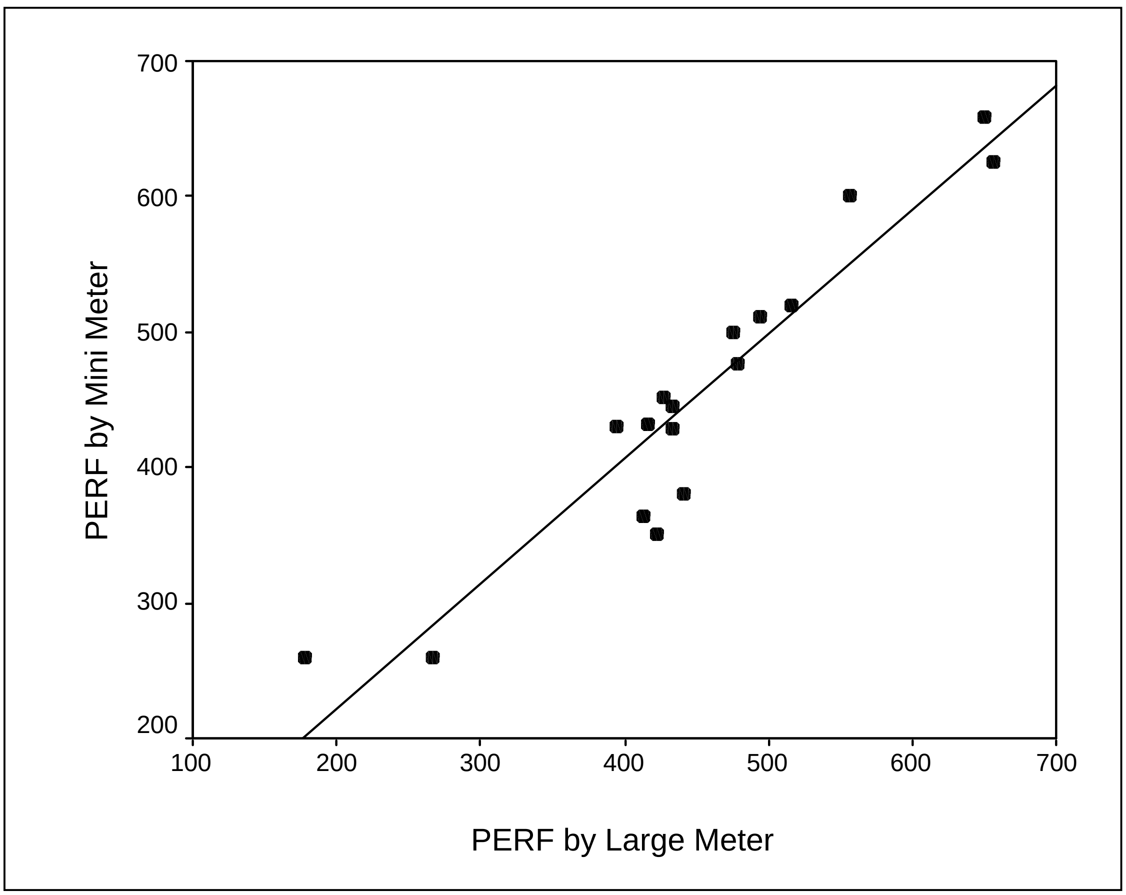
- Correlation is used to describe the linear relationship between two continuous variables (e.g., height and weight). In general, correlation tends to be used when there is no identified response variable. It measures the strength (qualitatively) and direction of the linear relationship between two or more variables.

- The Pearson correlation coefficient measures the strength of the ***linear*** association between two variables. It can be used to estimate the population correlation, $\rho$.

**Example**
**Correlating Two Lung Function Meters**
**(Wright vs. Mini-Wright)**

Peak expiratory flow rate (PERF) can be measured using either the “Wright Peak Flow Meter” or the “Mini-Wright Peak Flow Meter.” Do their measurements “correlate”?

r = 0.943




Note that this is a very different question than do their measurements “agree”.

## Properties of Correlations

- A correlation can range in value between –1 and 1:

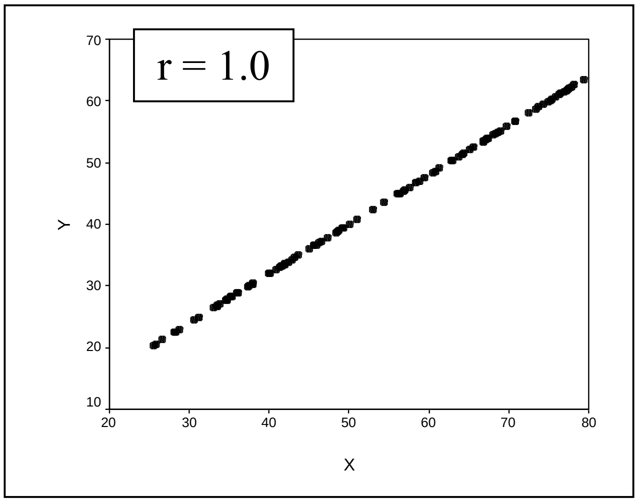
    - $r$ is a dimensionless quantity; that is, $r$ is independent of the units of measurement of $X$ and $Y$.

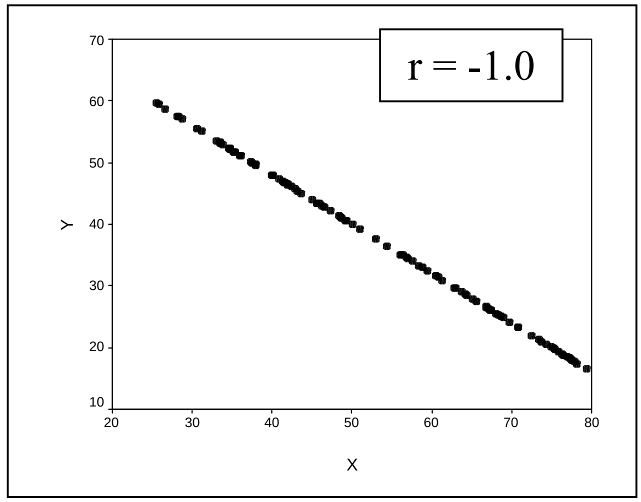
    - If the correlation is greater than 0, then as $X$ increases $Y$ increases and the two variables are said to be ***positively correlated***. An $r = 1$ is perfect positive correlation.

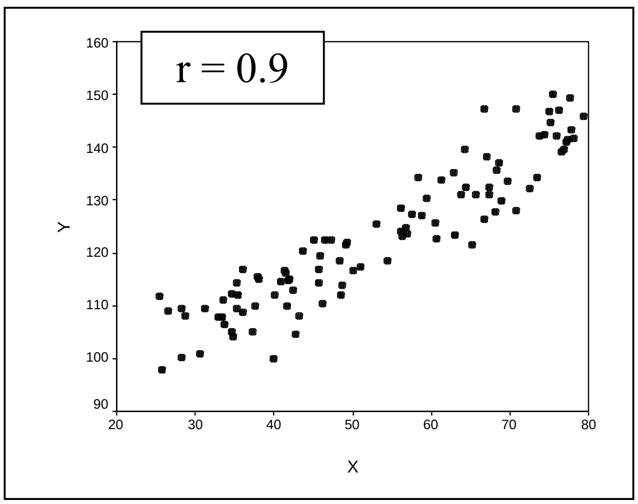
    - If the correlation is less than 0, then as $X$ increases $Y$ decreases and the two variables are said to be ***negatively correlated***. An $r = -1$ is perfect negative correlation.

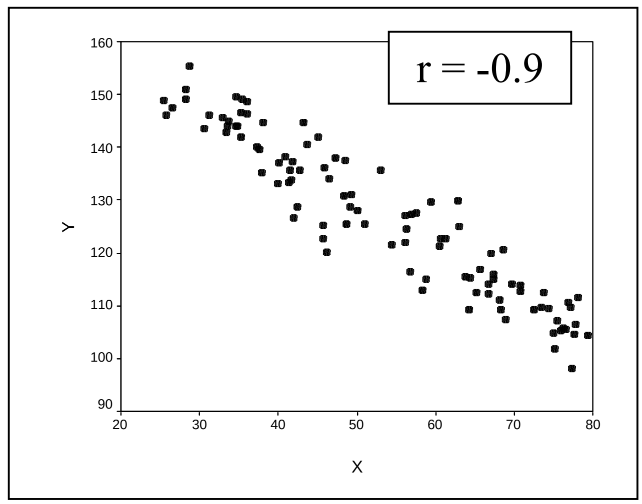
    - If the correlation is 0 then there is no *linear* relationship between $X$ and $Y$. The two variables are said to be ***uncorrelated***. The correlation is 0 when the covariance of X and Y is 0.

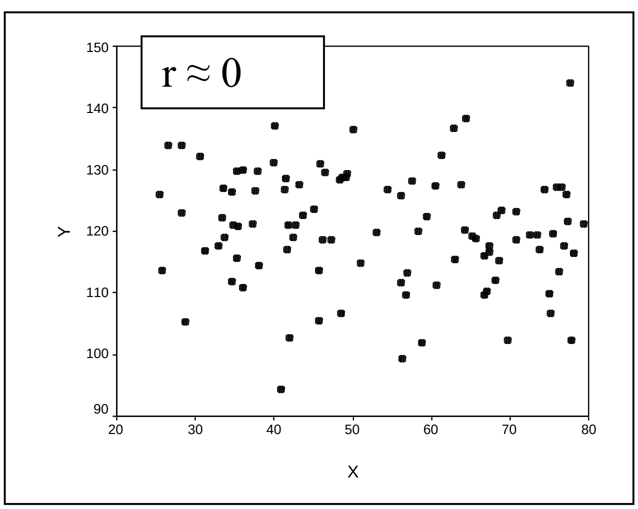
- The correlation coefficient is a measure of the strength of the linear trend relative to the variability of the data around that trend. Thus, it is dependent both on the magnitude of the trend and the magnitude of the variability in the data.


r = 1.0
Y
70 60 50 40 30 20 10
20 30 40 50 60 70 80
X

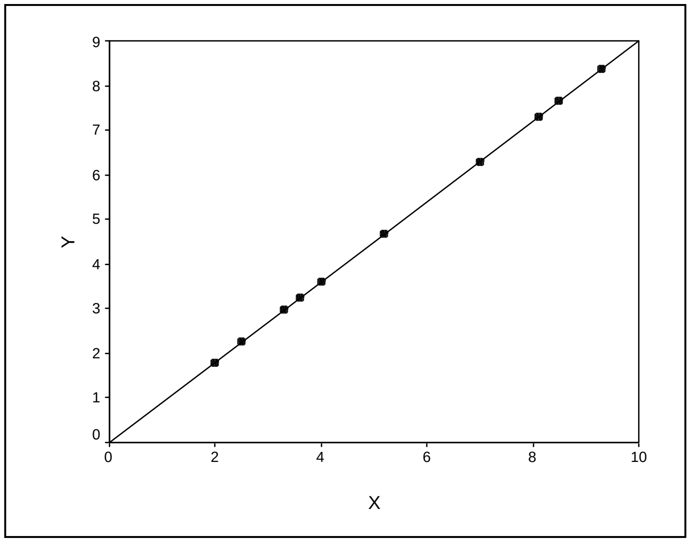

r = -1.0
Y
70 60 50 40 30 20 10
20 30 40 50 60 70 80
X

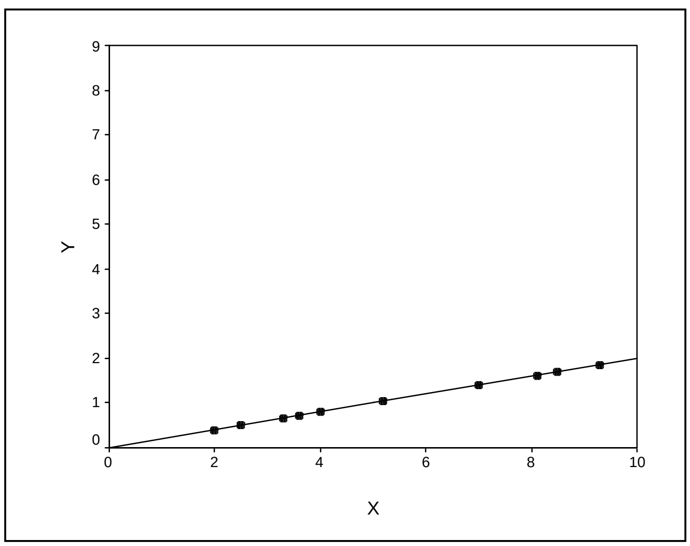

r = 0.9
Y
160 150 140 130 120 110 100 90
20 30 40 50 60 70 80
X


r = -0.9
Y
160 150 140 130 120 110 100 90
20 30 40 50 60 70 80
X


r ≈ 0
Y
150 140 130 120 110 100 90
20 30 40 50 60 70 80
X

Correlation is **not** a measure of the magnitude of the slope of the regression line.



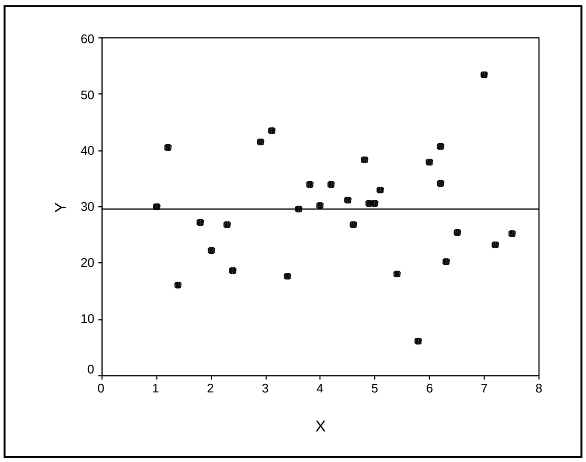

Both plots above have $r = 1$, but very different slopes.

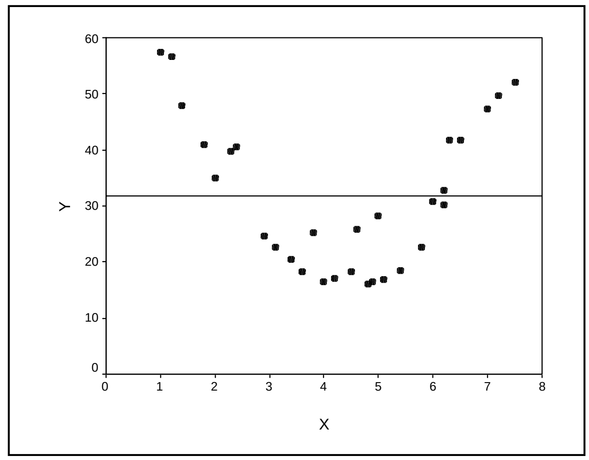
Note: In linear regression we can (and do) estimate the amount $Y$ increases/decreases with a one-unit change in $X$.

However, the slope does depend on the units of measurement of $X$ and $Y$, the correlation is dimensionless.

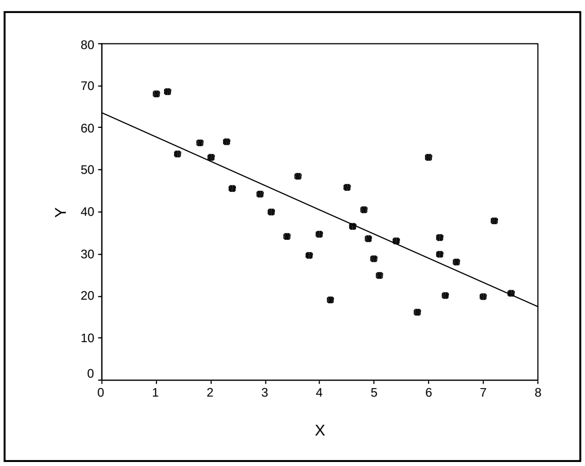
- Correlation is **not** a measure of the appropriateness of the straight-line model.

Examples where $r = 0$

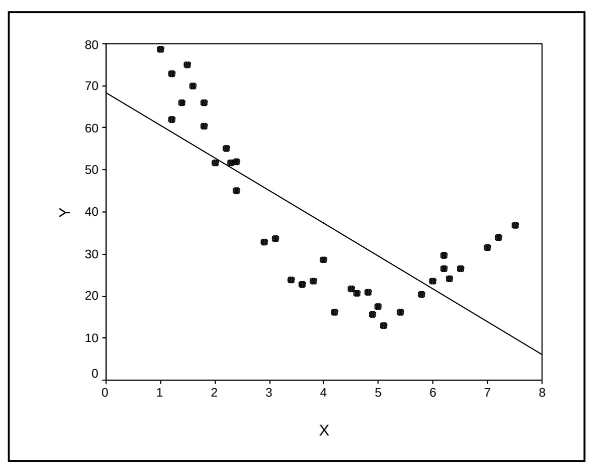



Examples when $r$ is high




$r = -0.766$




$r = -0.767$