Unit 4: Correlation

Marius Ionescu

09/20/2011

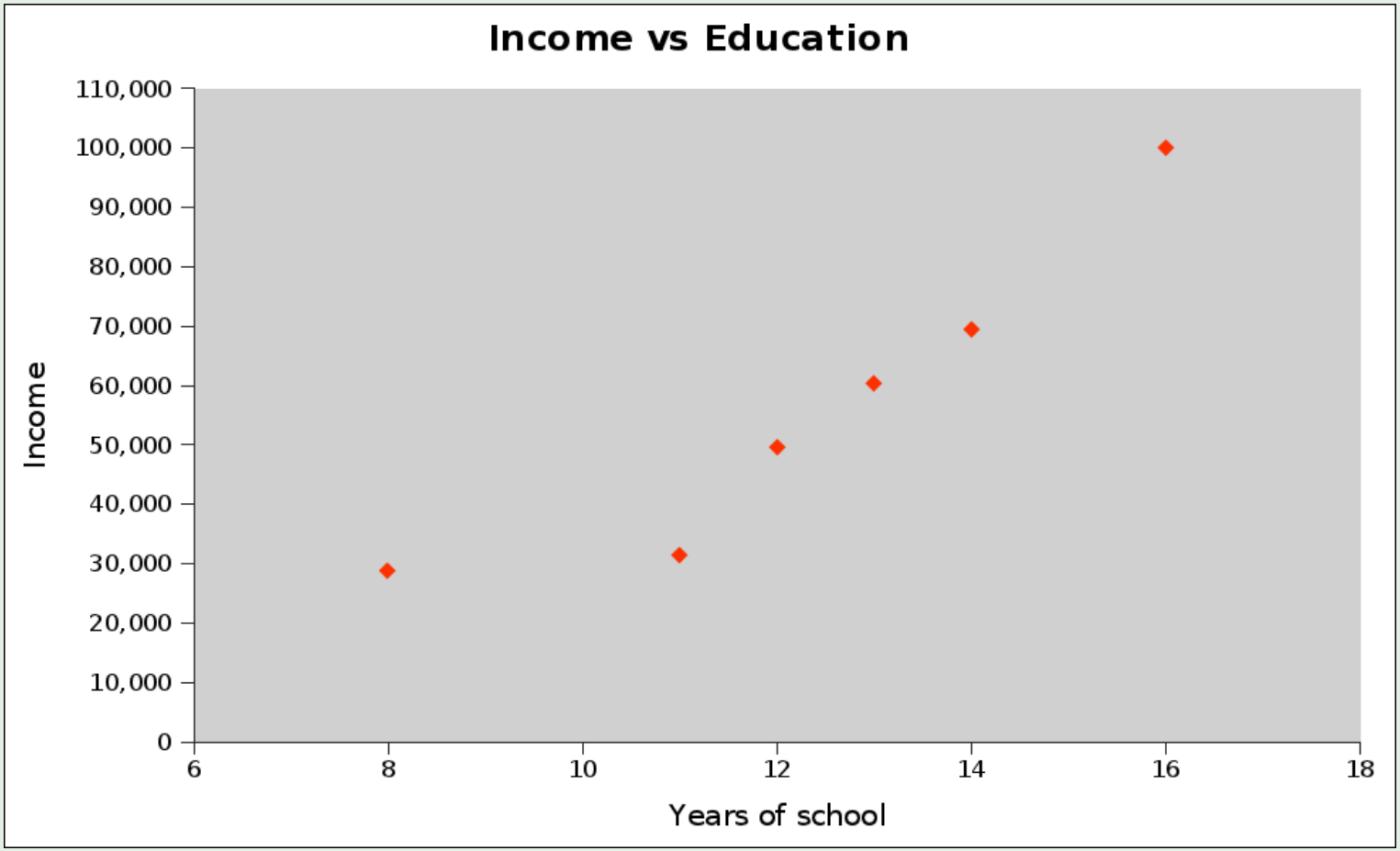
Fact

Correlation = the relationship between two variables

Scatter Diagram

Example

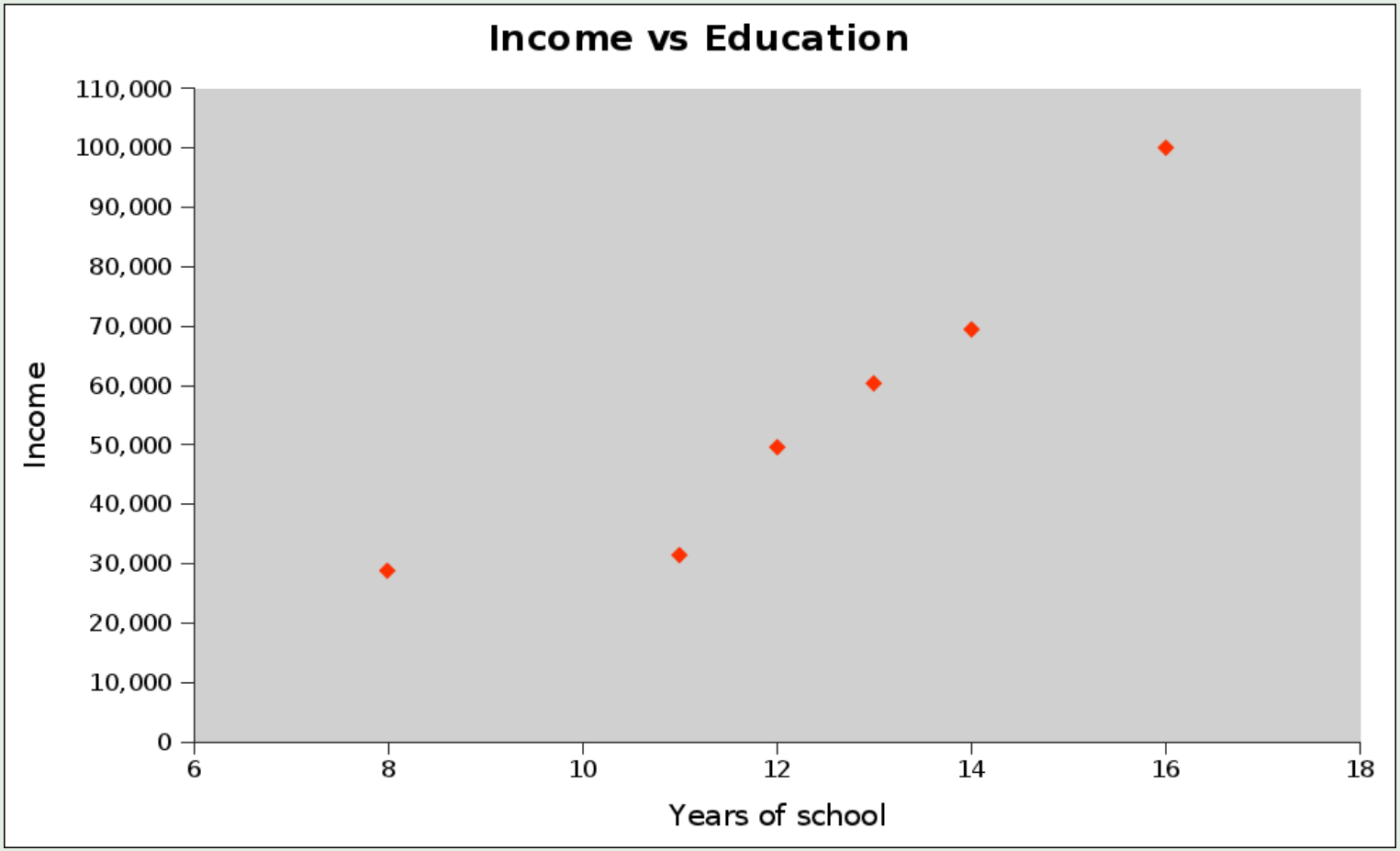
Education pays:



Scatter Diagram

Example

Education pays:

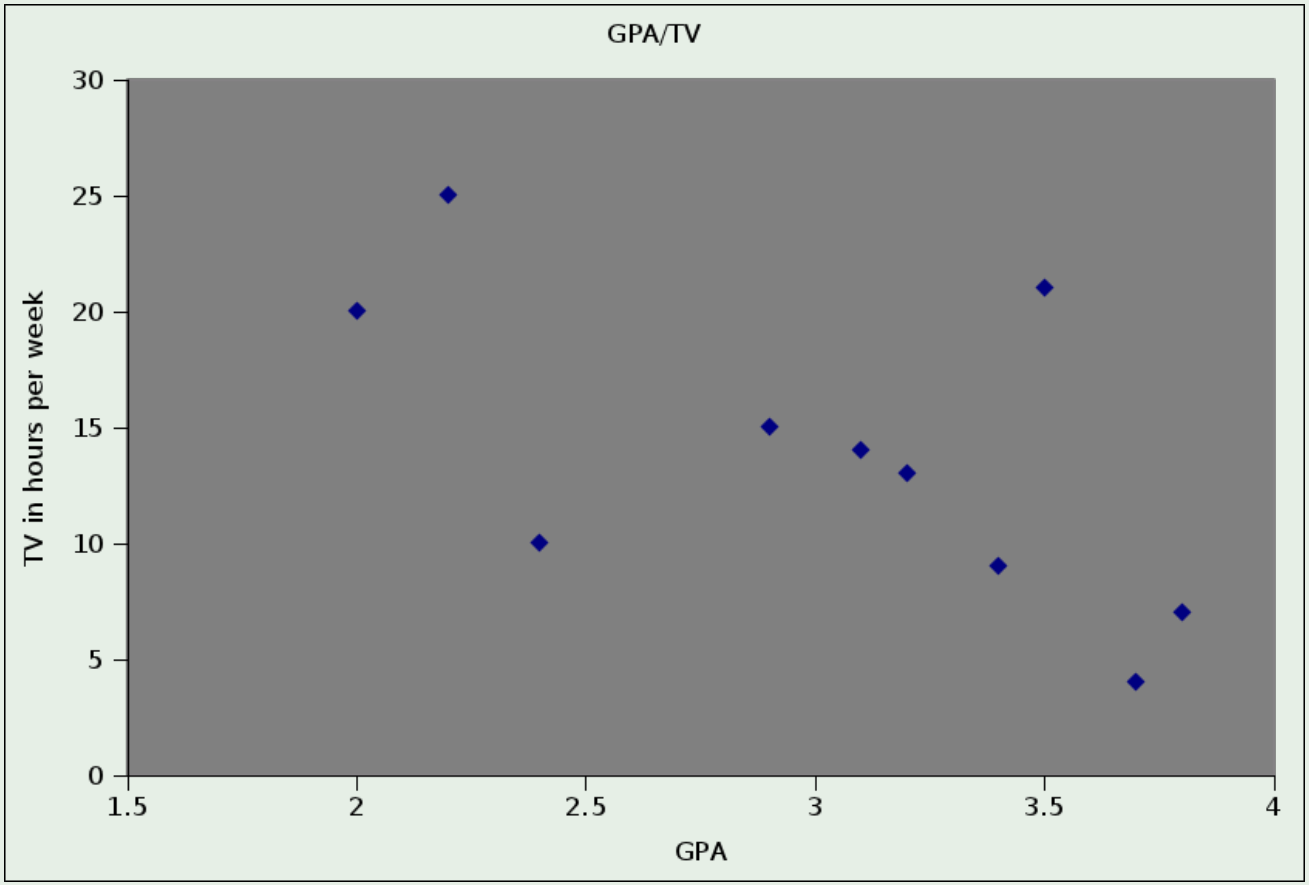


We see a positive correlation

Another scatter diagram

Example

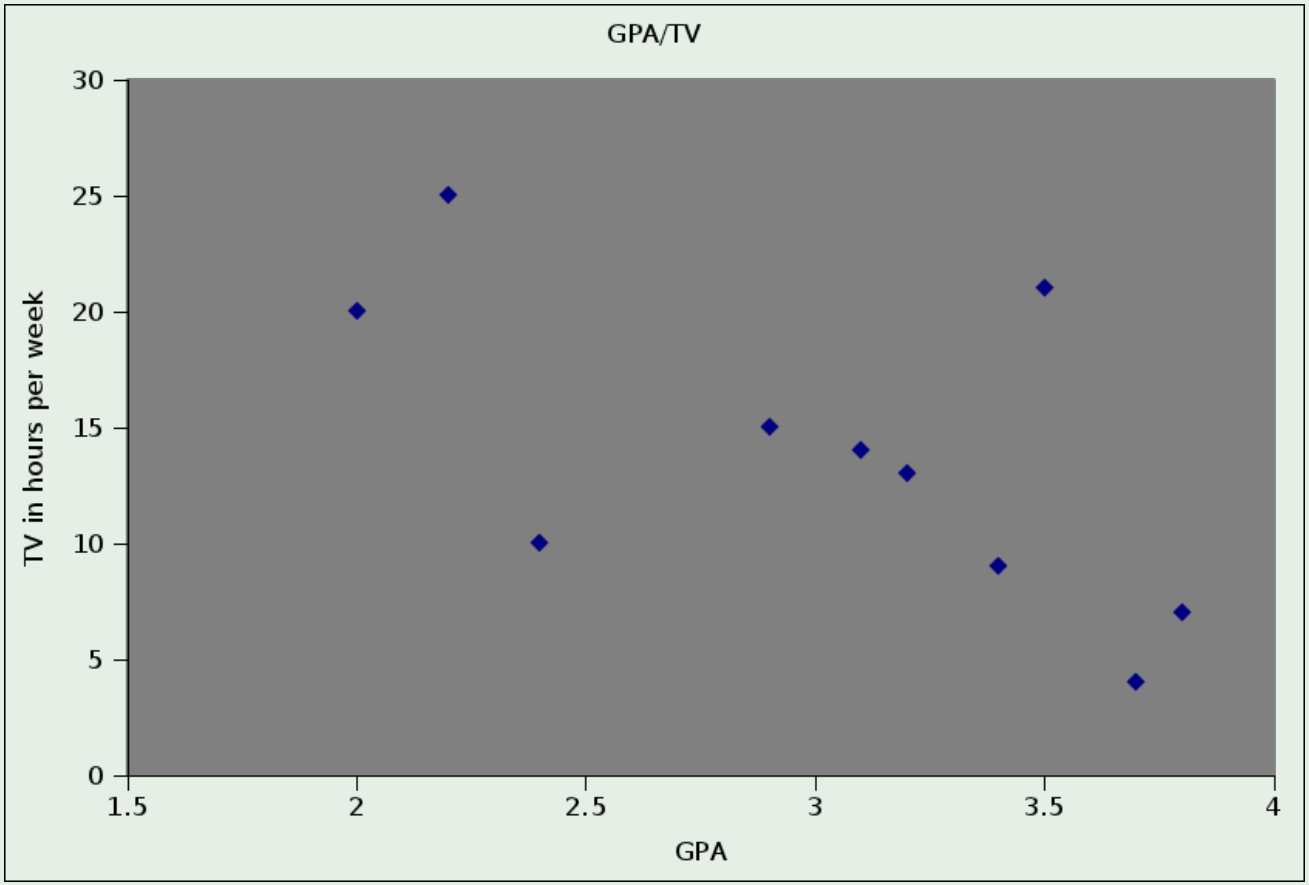
GPA/TV hourse per week:



Another scatter diagram

Example

GPA/TV hourse per week:



We see a negative correlation.

Scaling a scatter diagram

Fact

Scaling a scatter diagram

Fact

- *The best way to choose a scale for a scatter diagram is to use standard units*

Scaling a scatter diagram

Fact

- *The best way to choose a scale for a scatter diagram is to use standard units*
- *If x and y_i is your data point and the average of x is $\overline{x}$ and the average of y is $\overline{y}$ with SD_x and SD_y then convert to standard units as follows*

$$\mathrm{X} = \frac{x_i - \overline{x}}{SD_x} \text{ and } \mathrm{Y} = \frac{y_i - \overline{y}}{SD_y}.$$

How do we measure the correlation?

Definition

How do we measure the correlation?

Definition

- The **correlation coefficient** is

$$r = \text{average}(XY).$$

How do we measure the correlation?

Definition

- The **correlation coefficient** is

$$r = \text{average}(\mathrm{XY}).$$

- That is, the correlation coefficient equals the average of (X in standard units times Y in standard units).

Properties

Fact

Properties

Fact

- *The correlation coefficient is always between* -1 *and* 1.

Properties

Fact

- *The correlation coefficient is always between* -1 *and* 1.
- *We standardize it so that* ± 1 *is perfectly linear, and* 0 *is void of any linear relation.*

Example

Find the correlation of

x	y
1	4
2	5
6	4
11	15

Example

Find the correlation of

x	y
1	4
2	5
6	4
11	15

Exceptional Cases

Fact

The r-value can be misleading when we have

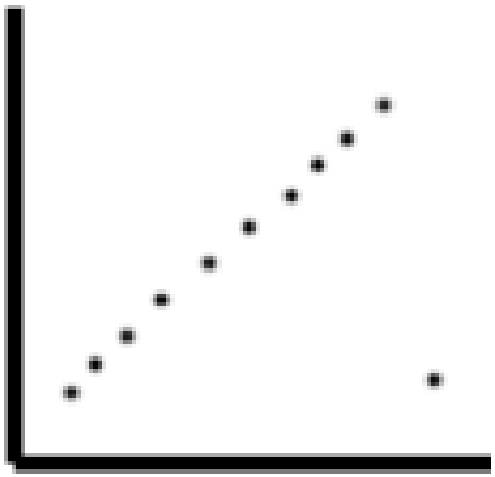
- *outliers*
- *nonlinear association*

Exceptional Cases

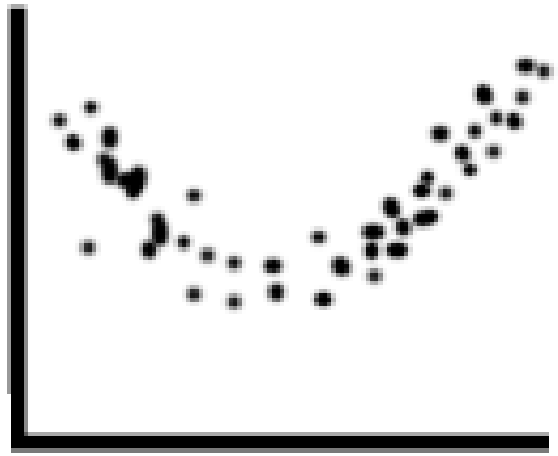
Fact

The r-value can be misleading when we have

- *outliers*
- *nonlinear association*



CLEAR ASSOCIATION
BUT R = 0



NONLINEAR ASSOC.
BUT R = 0

Important

Fact

r is valid when the scatterplot is football-shaped!

Fact

Fact

- *Are based on rates (e.g., averages).*

Fact

- *Are based on rates (e.g., averages).*
- *Used often in Poly Sci and Sociology.*

Fact

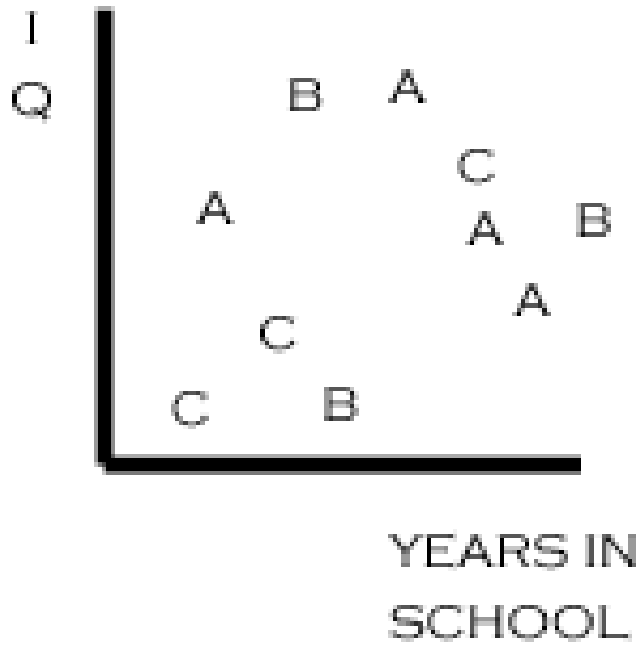
- *Are based on rates (e.g., averages).*
- *Used often in Poly Sci and Sociology.*
- *They tend to overstate the strength of an association.*

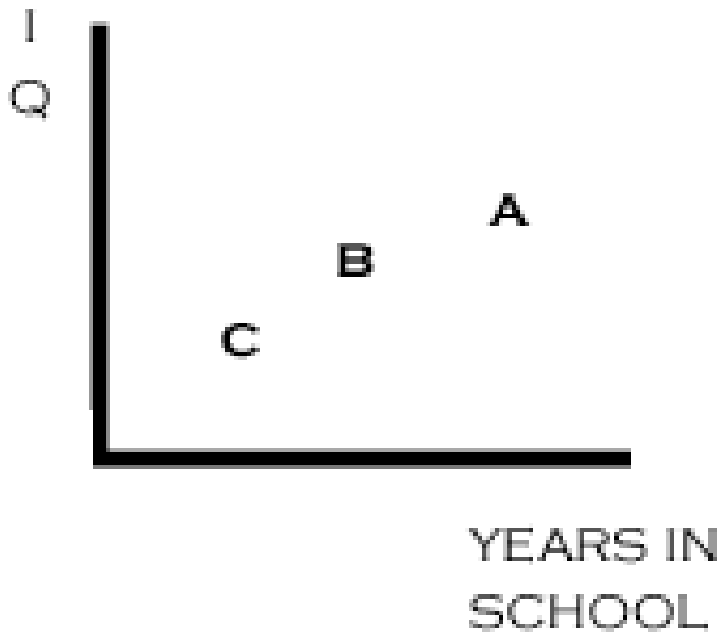
Fact

- *Are based on rates (e.g., averages).*
- *Used often in Poly Sci and Sociology.*
- *They tend to overstate the strength of an association.*
- *Reason: When we average we get rid of variations.*

Example

Consider IQ versus years in school, separated by age group: A = 30 year olds; B = 40 year olds; C = 50 year olds





Association is not Causation

Fact

Important:

"Correlation does not imply causation"

Example

Example

Are people rich because they go to school or they go to school because they are rich?

Example

Example (Phillips Curve)

Example

Example (Phillips Curve)

- Phillips curve is a historical inverse relationship between the rate of unemployment and the rate of inflation in an economy.

Example

Example (Phillips Curve)

- Phillips curve is a historical inverse relationship between the rate of unemployment and the rate of inflation in an economy.
- Stated simply, the lower the unemployment in an economy, the higher the rate of inflation.

Example

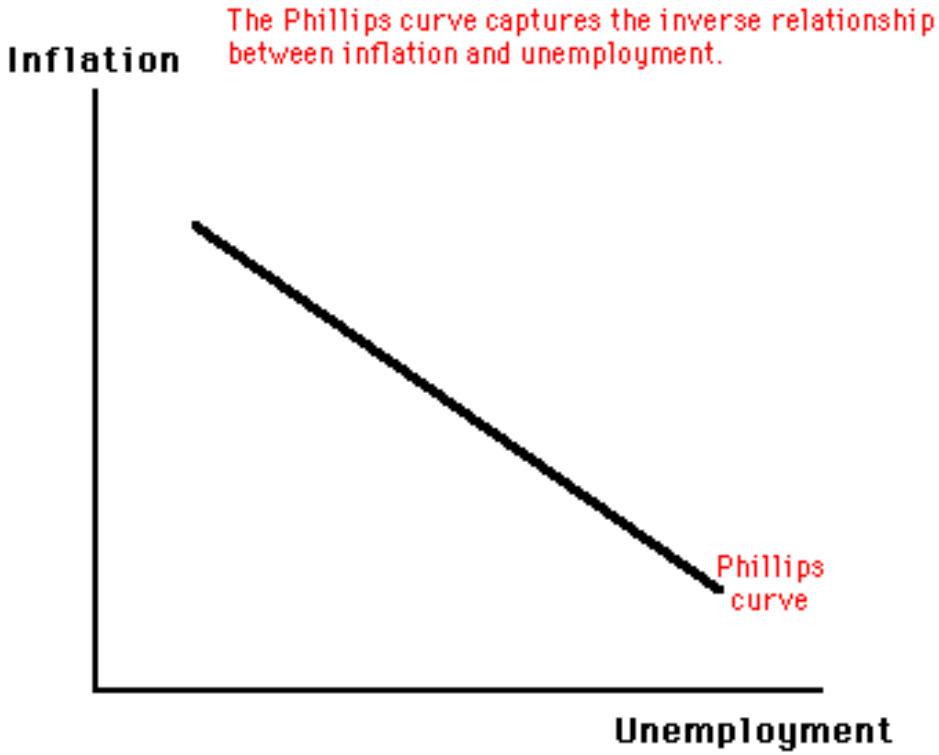
Example (Phillips Curve)

- Phillips curve is a historical inverse relationship between the rate of unemployment and the rate of inflation in an economy.
- Stated simply, the lower the unemployment in an economy, the higher the rate of inflation.

Example

Example (Phillips Curve)

- Phillips curve is a historical inverse relationship between the rate of unemployment and the rate of inflation in an economy.
- Stated simply, the lower the unemployment in an economy, the higher the rate of inflation.



Example

Example (Phillips Curve)

Example

Example (Phillips Curve)

- The data fit a line very well from 1940 to late 1960's.

Example

Example (Phillips Curve)

- The data fit a line very well from 1940 to late 1960's.
- It was used in policy and we got high inflation and high unemployment.

Example

Example

Studies showed that there is a strong correlation between crime rate and temperature in Chicago.

Example

Example

Studies showed that there is a strong correlation between crime rate and temperature in Chicago.
If the mayor puts 200 more police officers in the street will the temperature go down?