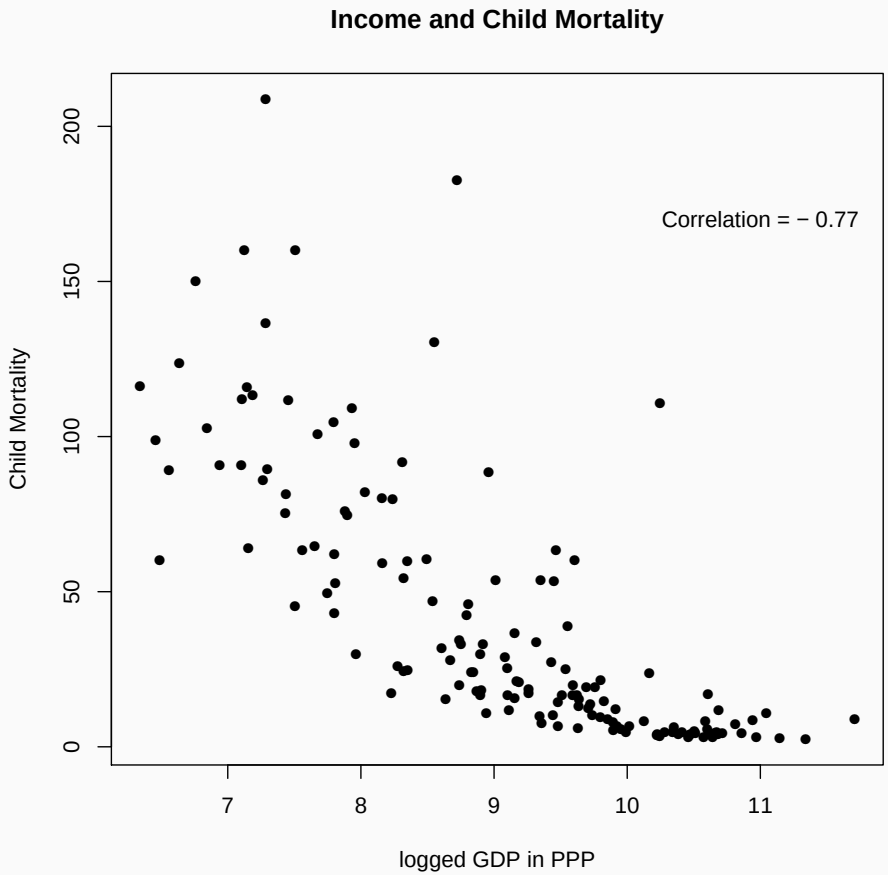
Political Science 209 - Fall 2018

Linear Regression

Florian Hollenbach

9th October 2018

Recall Correlation & Scatterplot



What is the correlation?

Recall the definition of correlation

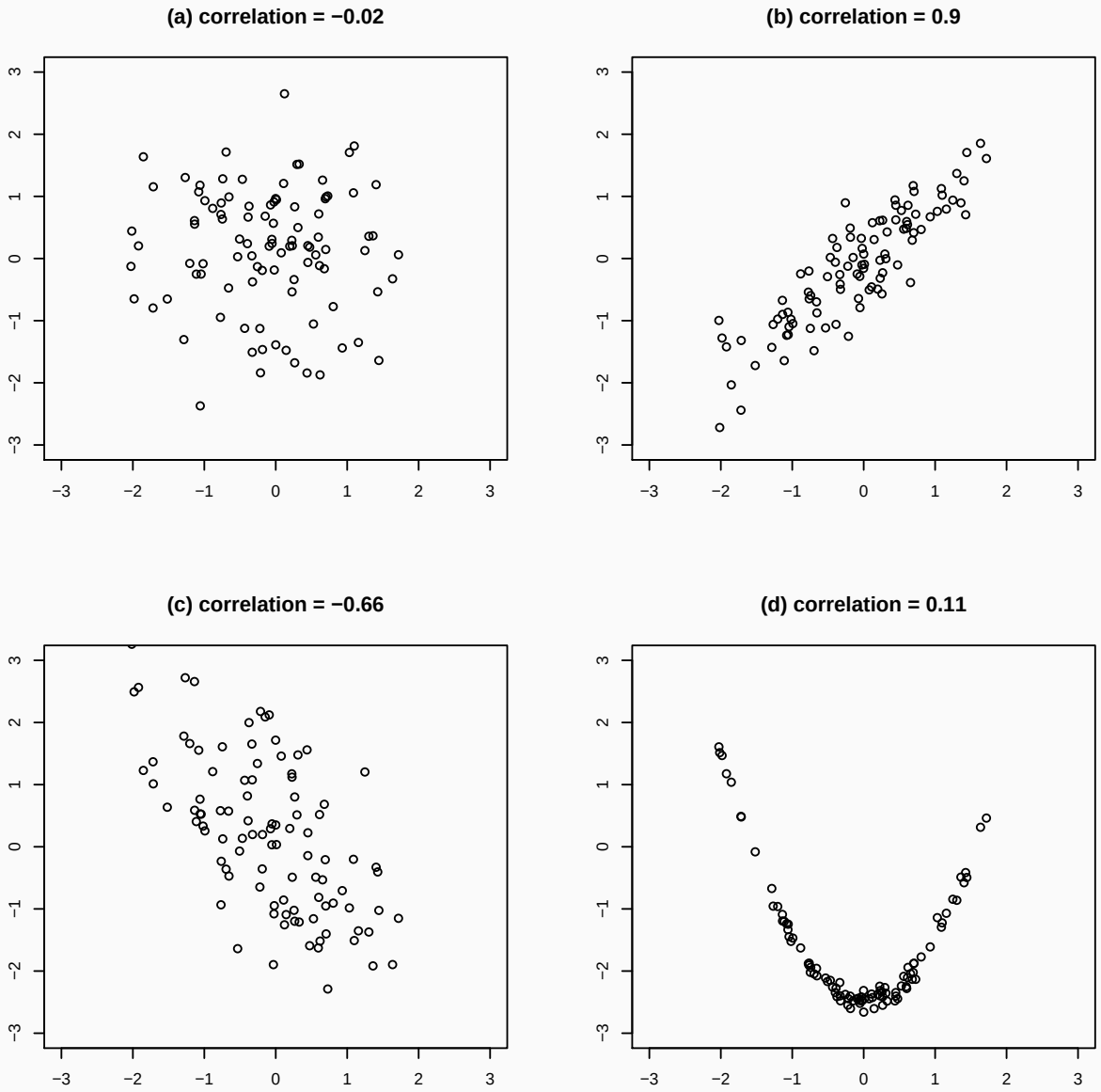
$$\text{Correlation (x,y)} = \frac{1}{N}\sum_{i=1}^{N} \text{z-score of } x_i \times \text{ z-score of } y_i$$

$$\text{Correlation (x,y)} = \frac{1}{N}\sum_{i=1}^{N} \frac{x_i - \bar{x}}{sd_x} \times \frac{y_i - \bar{y}}{sd_y}$$

Correlations & Scatterplots/Data points

1. positive correlation $\rightsquigarrow$ upward slope
2. negative correlation $\rightsquigarrow$ downward slope
3. high correlation $\rightsquigarrow$ tighter, close to a line
4. correlation cannot capture nonlinear relationship

Correlations & Scatterplots/Data points



Moving from Correlation to Linear Regression

Preview:

- linear regression allows us to create predictions
- linear regression specifies direction of relationship
- linear regression allows us to examine more than two variables at the same time (*statistical control*)

Linear Regression

- regression has one dependent (y) and *for now* one independent (x) variable
- regression is a statistical method to estimate the linear relationship between variables

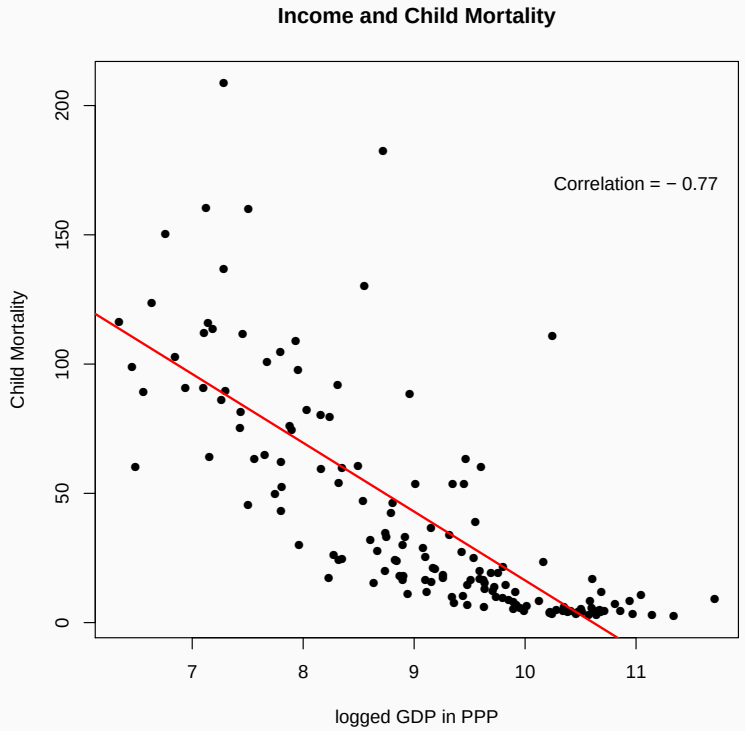
Linear Regression

- goal of regression is to approximate the (linear) relationship between X and Y as best as possible

Linear Regression

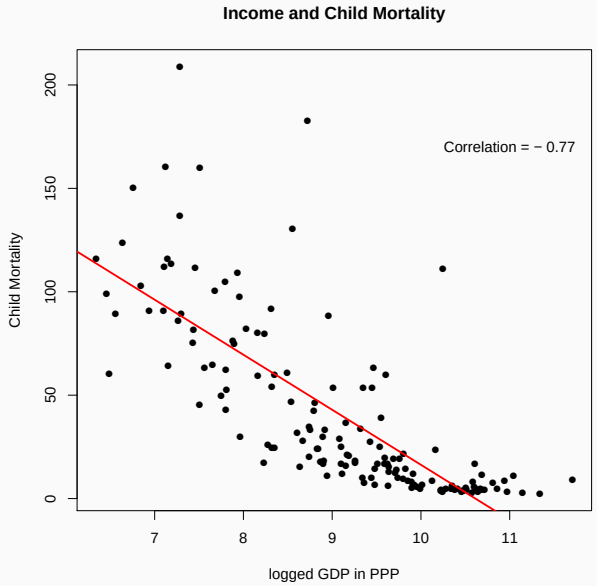
- goal of regression is to approximate the (linear) relationship between X and Y as best as possible
- regression is the mathematical model to draw best fitting line through cloud of points

Linear Regression



Linear regression is the mathematical model to draw best fitting line through cloud of points

Linear Regression



- regression line is an estimate of the (for now bivariate) relationship between x and y
- for each x we have a prediction of y: what would we expect y to be given the value of x?

What is the equation of a line?

Equation of a line?

What is the equation of a line?

Equation of a line? $y = mx + b$

$\rightarrow$ b? m?

What is the equation of a line?

Equation of a line?

$y = mx + b$

b $\rightarrow$ y-intercept

m $\rightarrow$ slope

What is the equation of a line?

Equation of a line?

$y = mx + b$

b $\rightarrow$ y-intercept

m $\rightarrow$ slope

regression equation:

$Y = \alpha + \beta X + \epsilon$

$\rightarrow \alpha?\ \beta?\ \epsilon?$

What is the equation of a line?

Equation of a line?

$y = mx + b$

b $\rightarrow$ y-intercept

m $\rightarrow$ slope

regression equation:

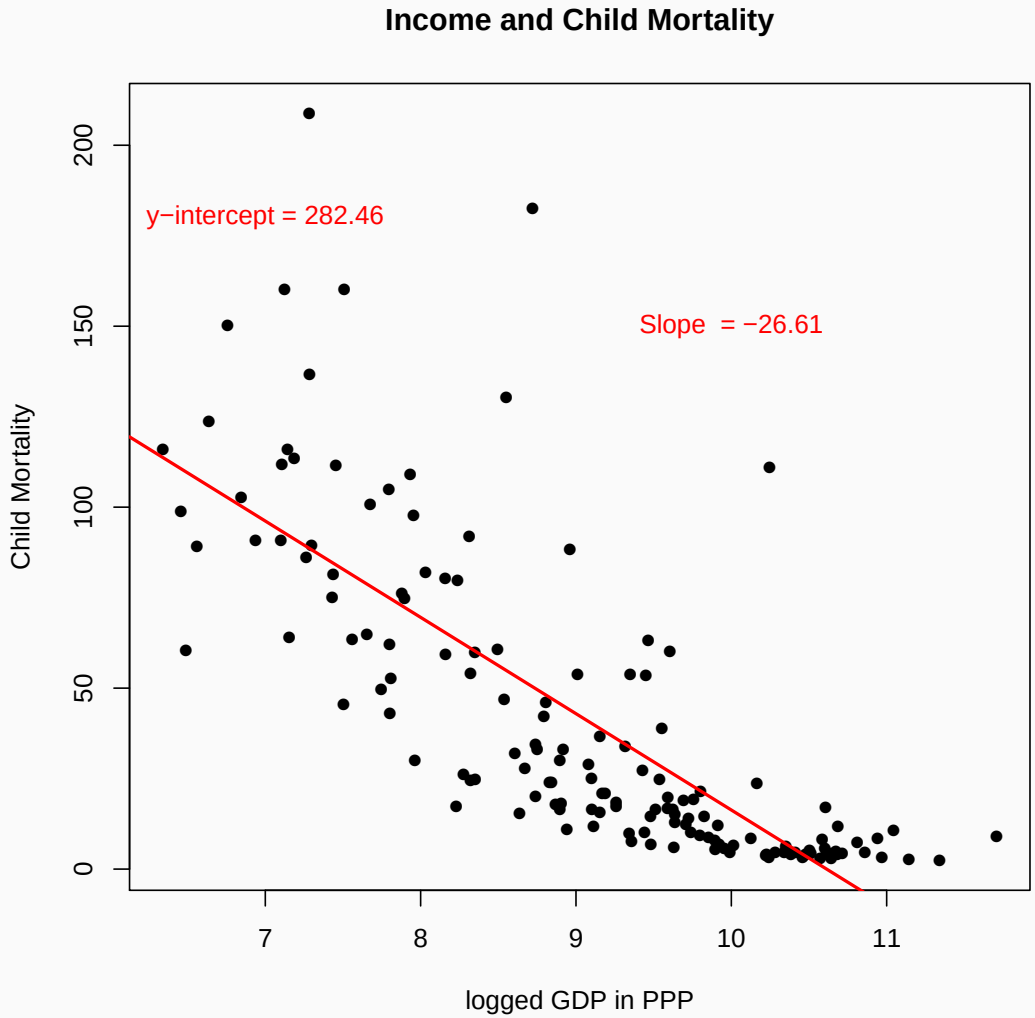
$Y = alpha + \beta X + \epsilon$

$\alpha \rightarrow$ y-intercept

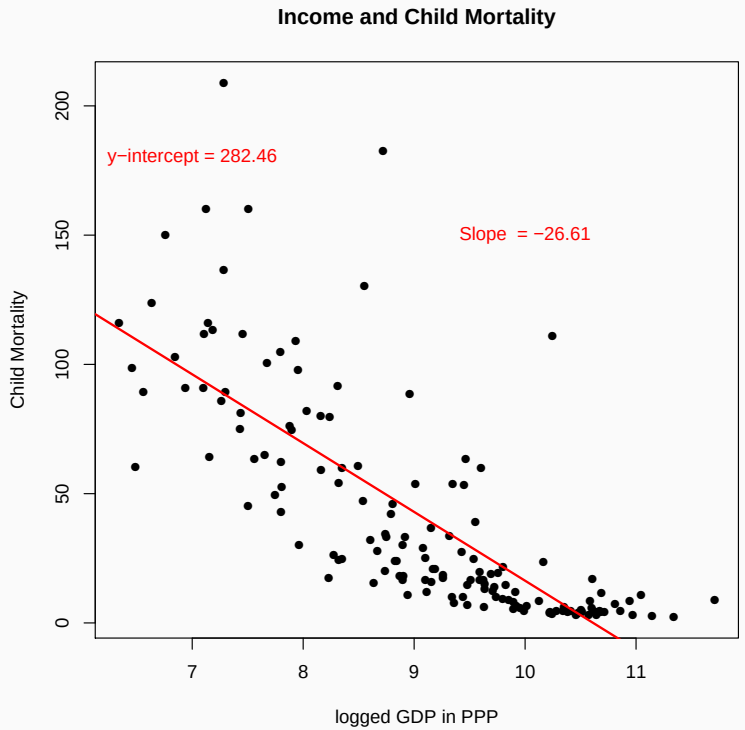
$\beta \rightarrow$ slope

$\epsilon \rightarrow$ error

Regression equation



Regression equation



$$Y = 282.46 + -26.61X + \epsilon$$

Regression equation

- but, we don't know the equation that generates the data
- our regression line is an estimate, based on the collected data

Regression equation

- but, we don't know the equation that generates the data
- our regression line is an estimate, based on the collected data

- estimates are denoted with little hats: $\hat{\beta}$ and $\hat{\alpha}$

Regression equation

- but, we don't know the equation that generates the data
- our regression line is an estimate, based on the collected data

- estimates are denoted with little hats: $\hat{\beta}$ and $\hat{\alpha}$

- ϵ is an estimate of how good/bad our approximation is