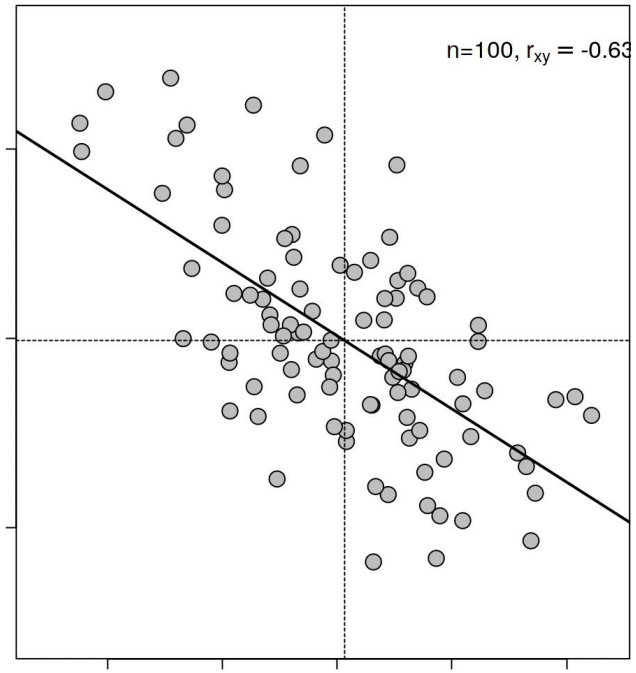
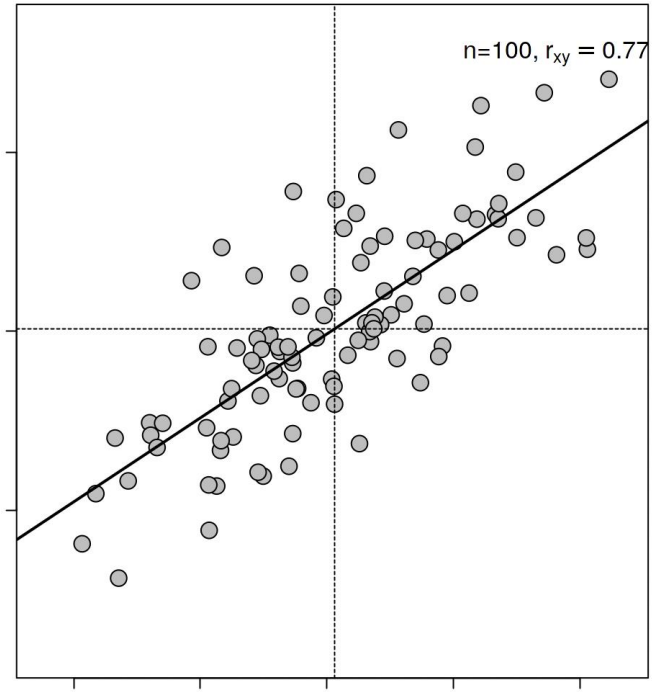
Fundamentals of Econometrics Models



Vicenç Soler

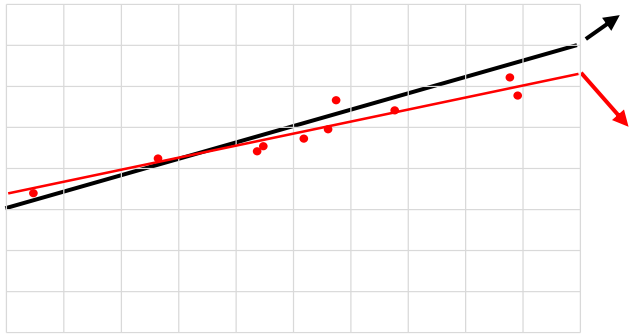
v.soler@tbs-education.org

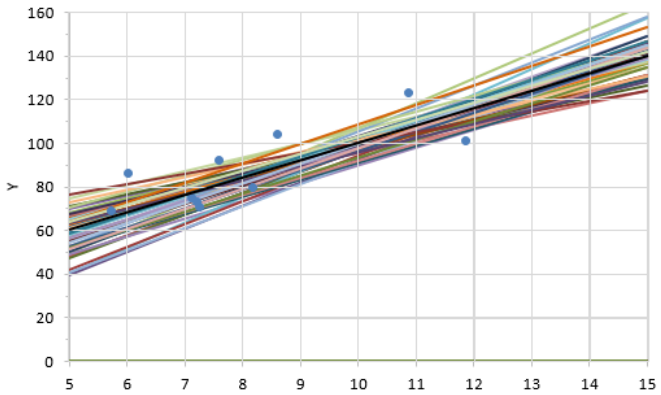
~~vincent.soler@tbs-education.org~~



Fitted model and Theoretical model

Our data comes from a sample





Theor. model:

$$Y = 20 + 8X + \varepsilon$$
$$\varepsilon \sim N(0; 10)$$

Unique but unknown

Fitted model:

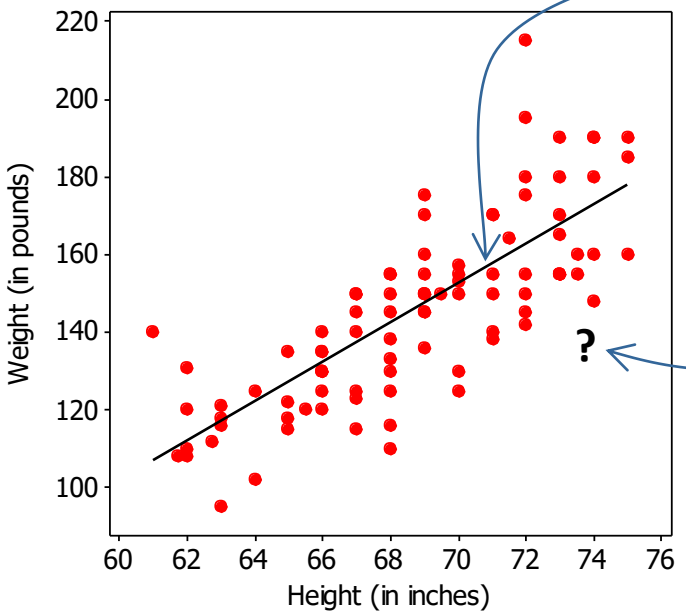
$$\text{sample of } n = 10$$
$$\hat{Y} = 41,9 + 5,6X + e$$

Known but they change from sample to sample

50 lines from 50 different samples

Fitted model and Theoretical model

Our data comes from a sample



Fitted model (from a sample)

$$y_i = b_0 + b_1 x_i + e_i$$

Residuals

Random variables
(values obtained from the sample)

Theoretical model (all population)

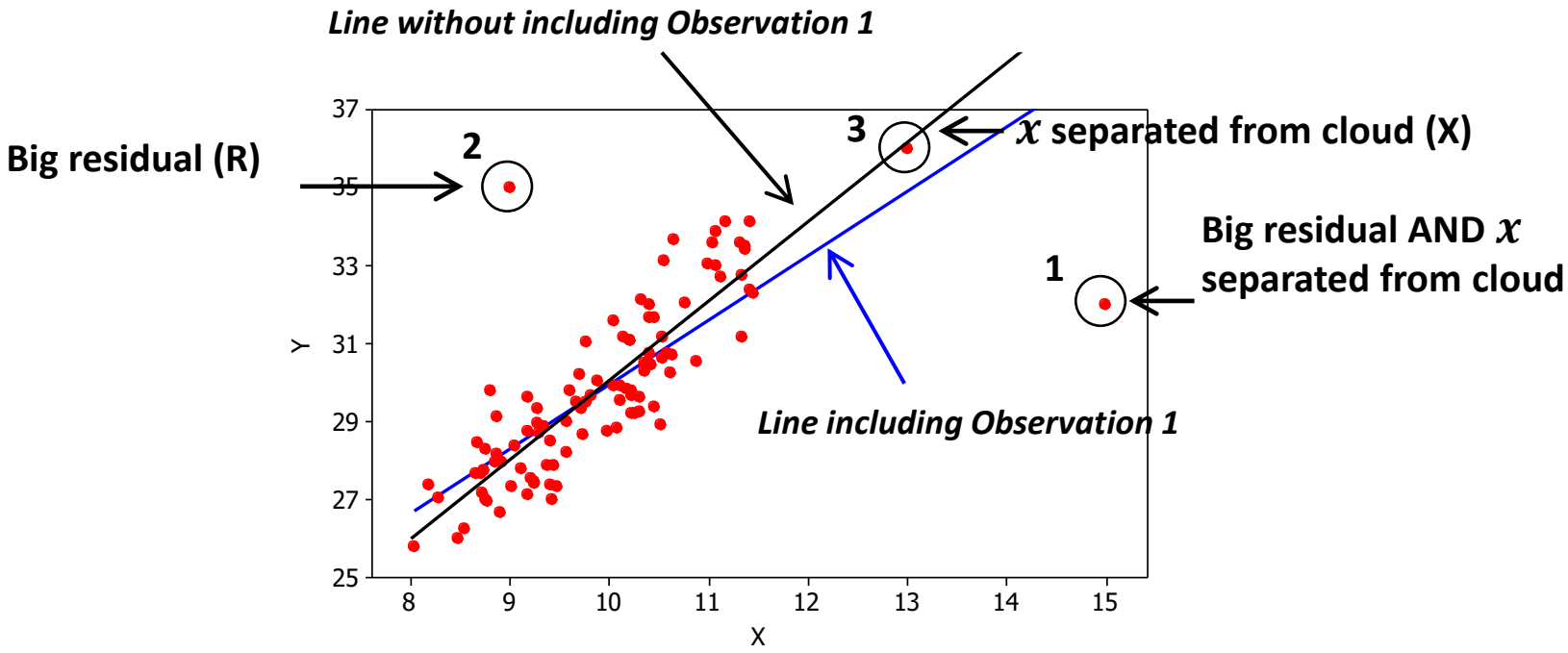
$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Errors

Parameters
(unique but unknown)

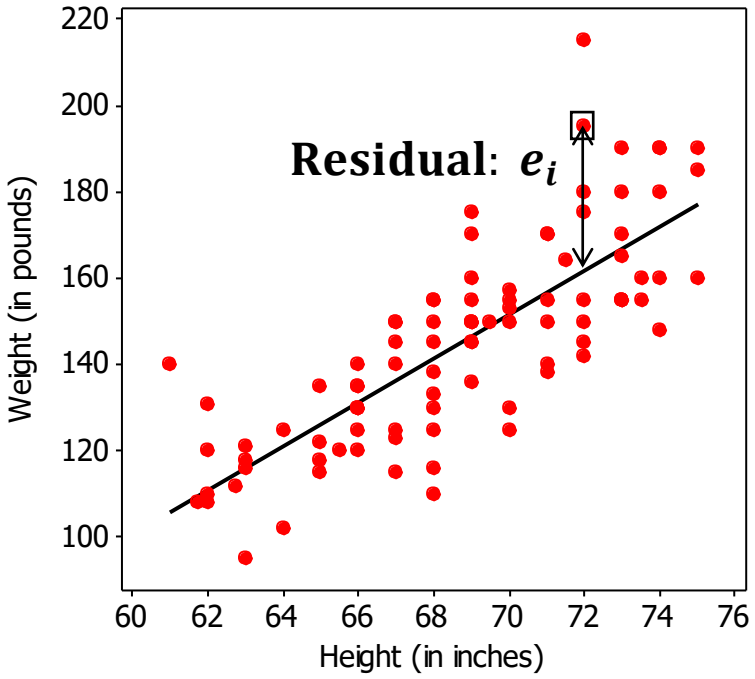
Sample vs. Population = Fitted model vs. Theoretical model

Atypical observations

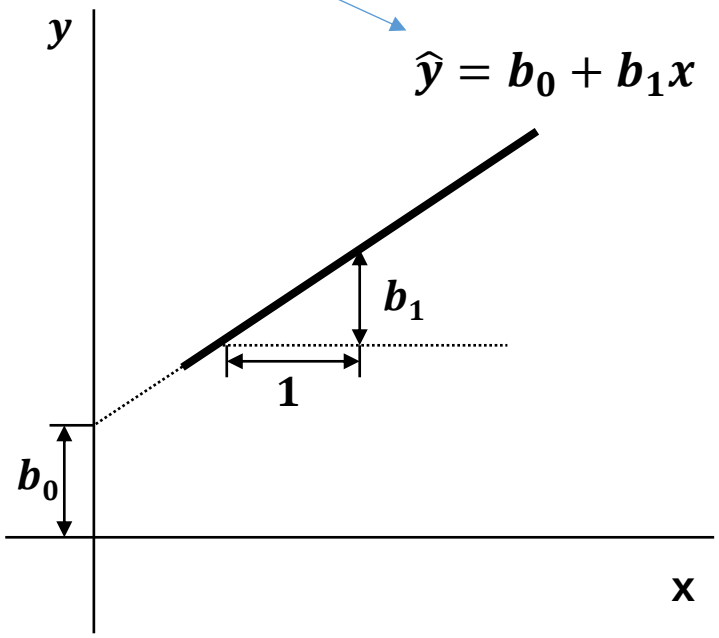


As the goal is to generate a general model, it is always better to exclude the atypical values when generating the Regression Model

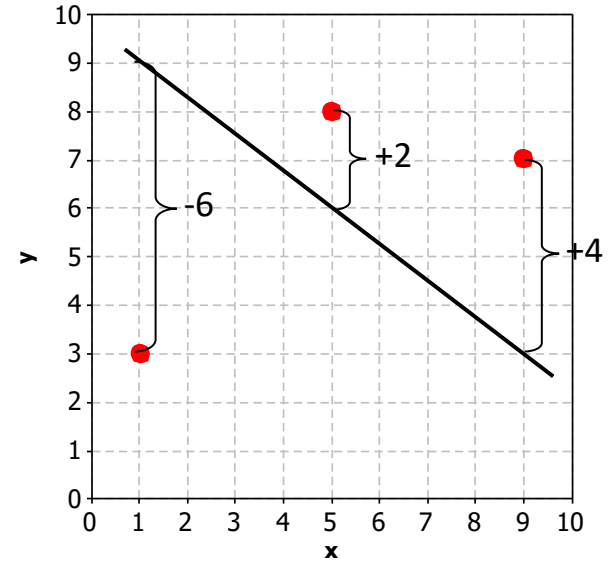
Residuals: Quick reminder



$\hat{y}$ = y estimated by the model



Residuals: actual Y – predicted Y ($\hat{y}$)



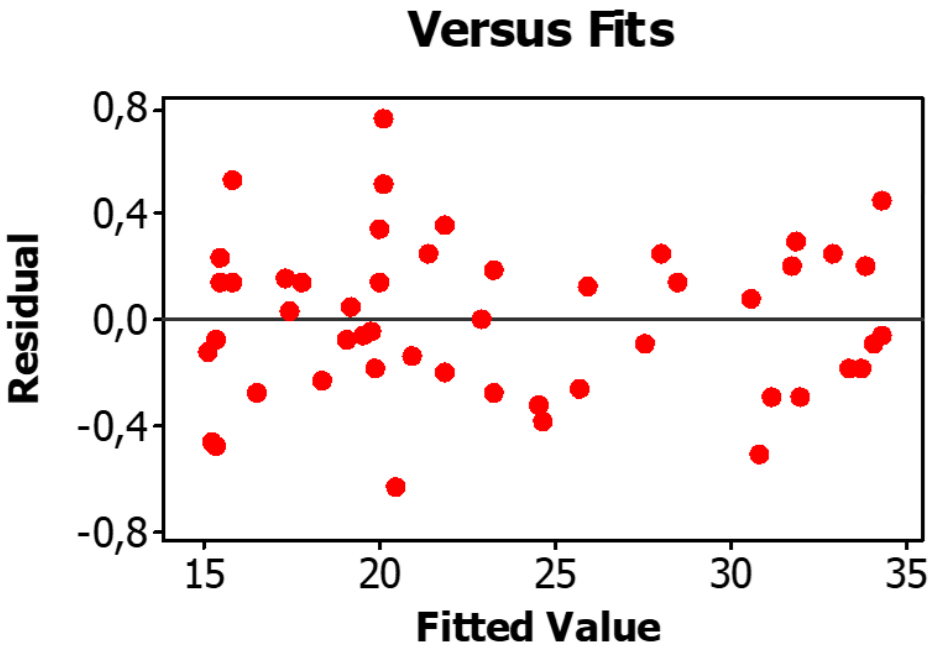
GOAL: To understand the relationship between y and x and/or to predict values of *y* given *x*

Model validation

R2 value

Analysis of the residuals

Analysis of the residuals

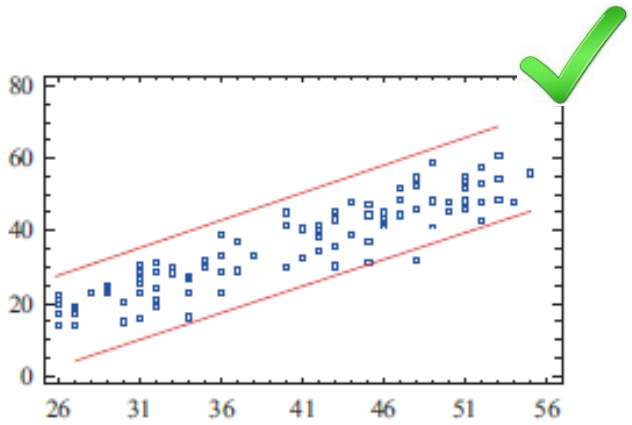


Analysis of the residuals

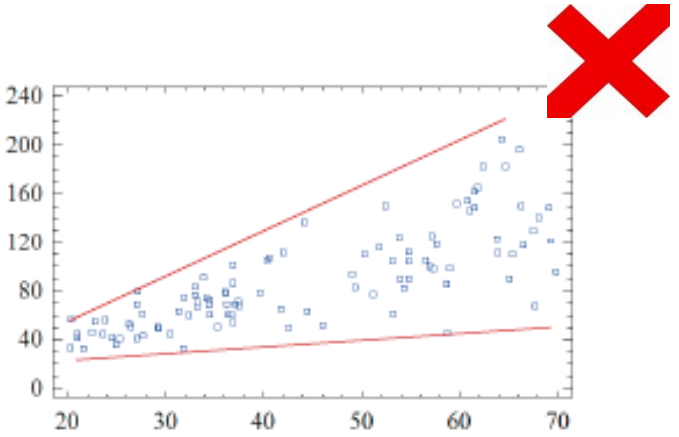
Homoscedasticity vs Heteroscedasticity

Constant variance in the residuals

Dispersion must be constant

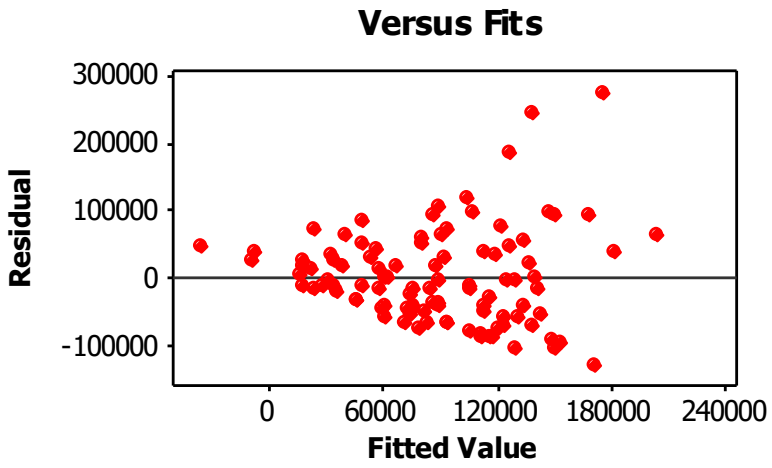


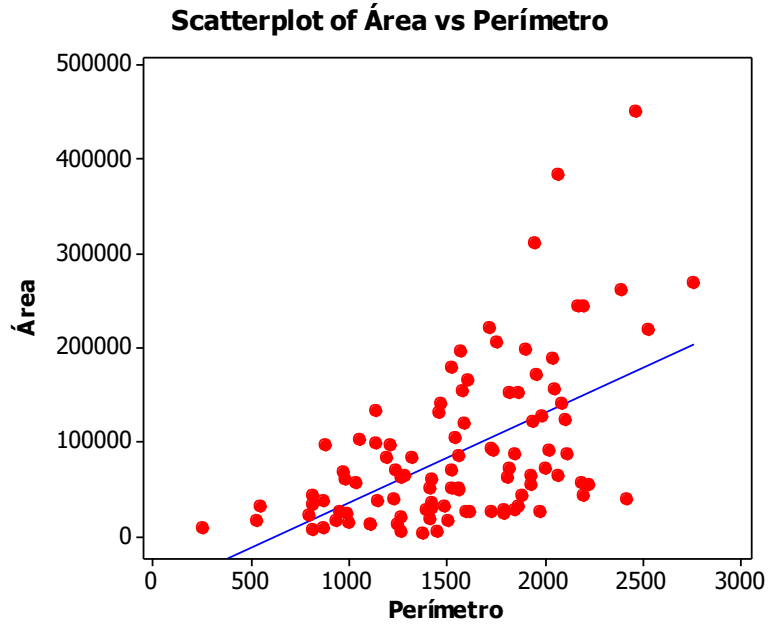
Homoscedastic data



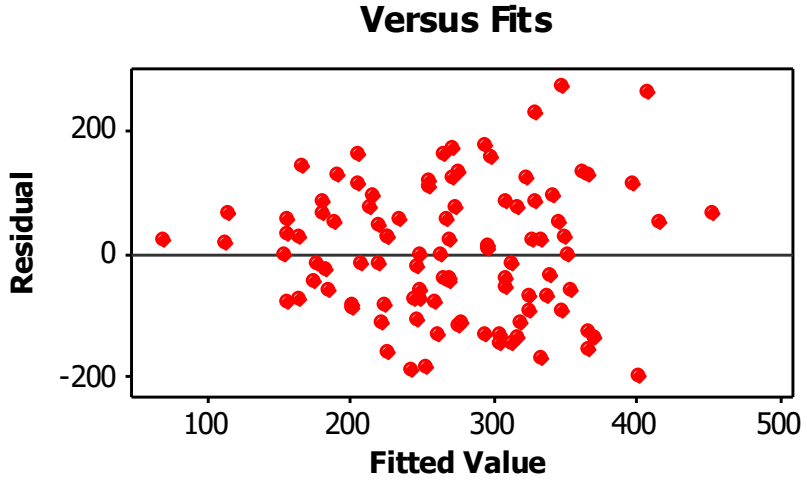
Heteroscedastic data

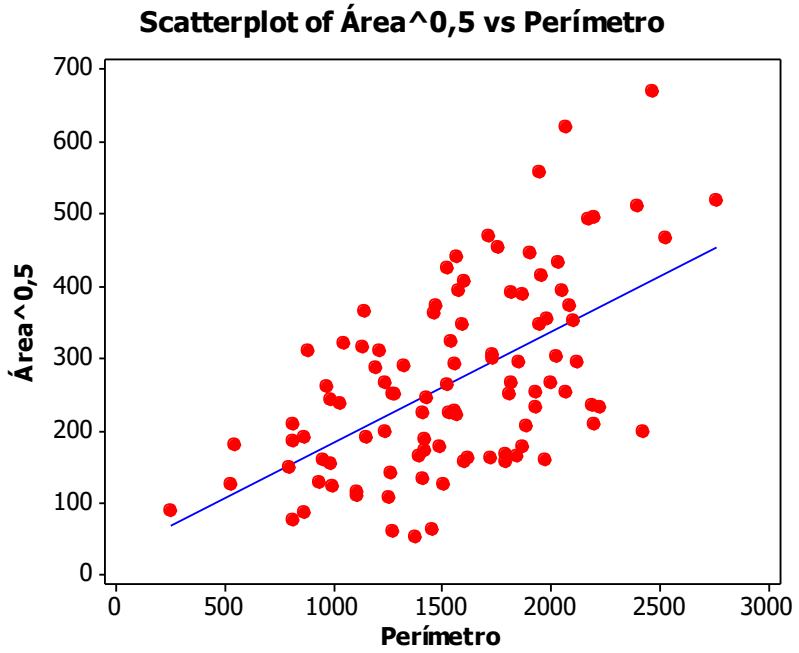
Analysis of the residuals



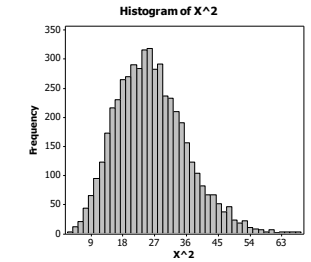


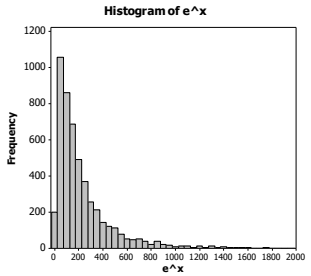
$\sqrt{Y}$

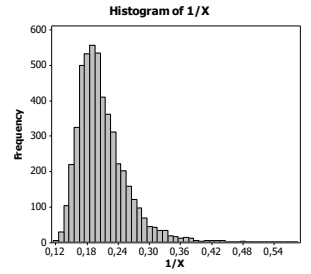




We maybe need to transform data to convert it to linear





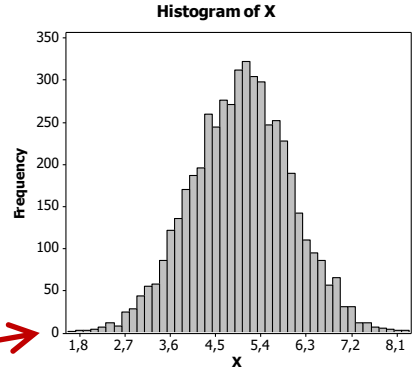


$\sqrt{x}$

$\ln(x)$

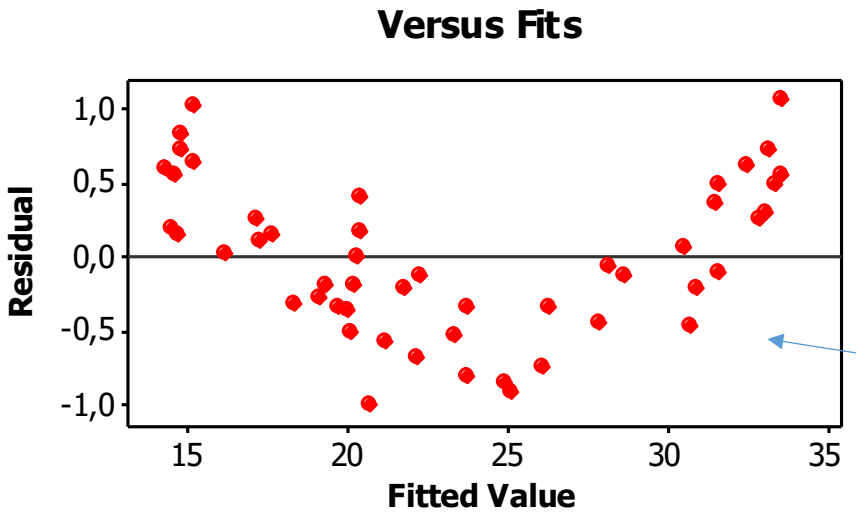
$1/x$

Transformation can be on x, on y or on both



We maybe need to transform data to convert it to linear

Example



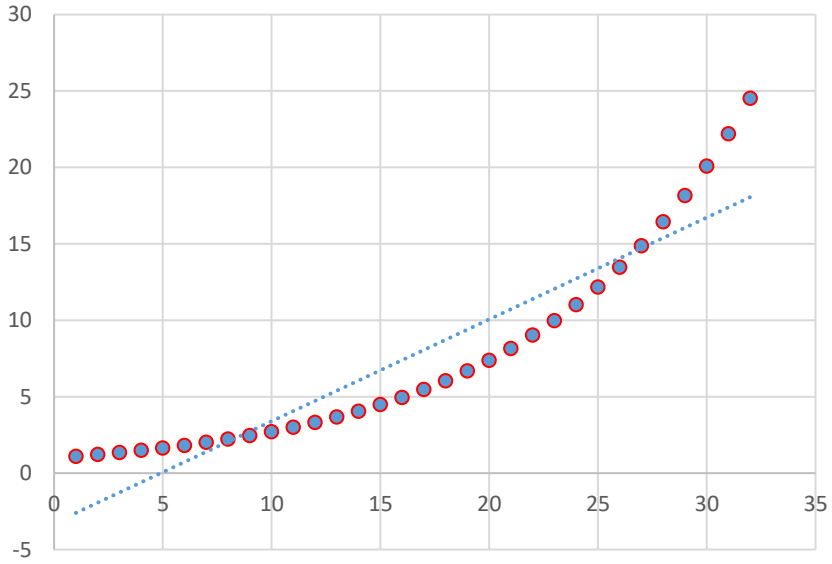
We should see a random pattern and this is not a random pattern

What about these residuals?

Positive->negative->positive

It looks as if it is needed a change in the model to better fit

Residuals = Actual - Predicted



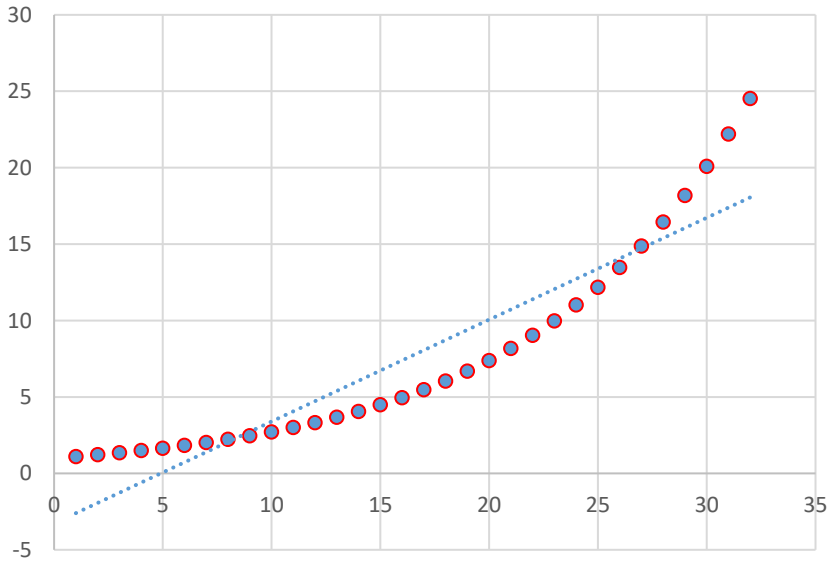
Actual data and its regression line

We maybe need to transform data to convert it to linear

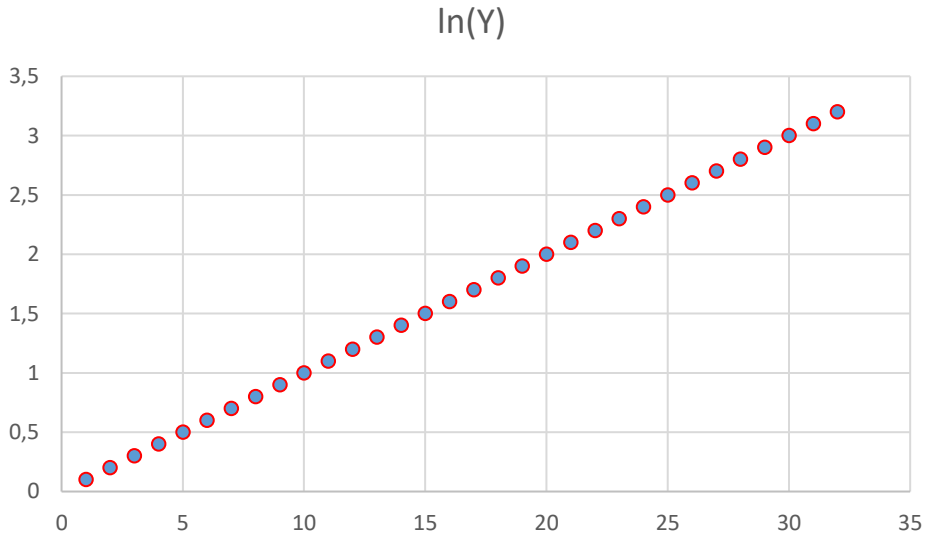
Example

Next step: As it seems data follows an exponential function, transform Y variable to **ln(Y)** and then generate a new Regression line using ln(Y) as the dependent variable.

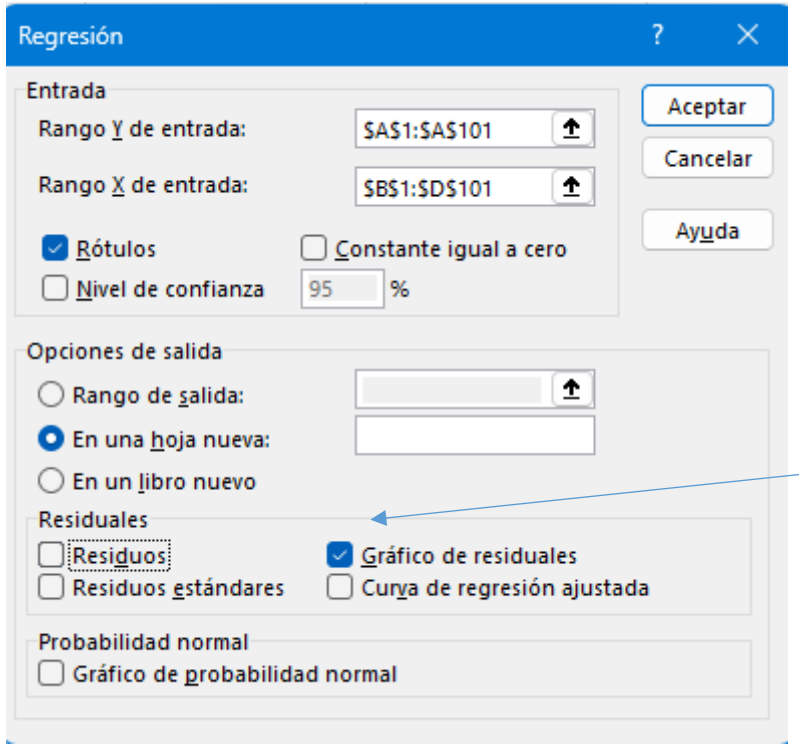
Y



Ln(Y)



Residuals in Excel…



Enable it to display graphs of residuals

Analysis of Residuals

Transformations (ln, sqrt, x^2, etc.):

- Allow us to obtain a better linear model when data is not linear, and predict better.
- It is more difficult to interpret the model and the relationships between variables.

Factor Variables in R

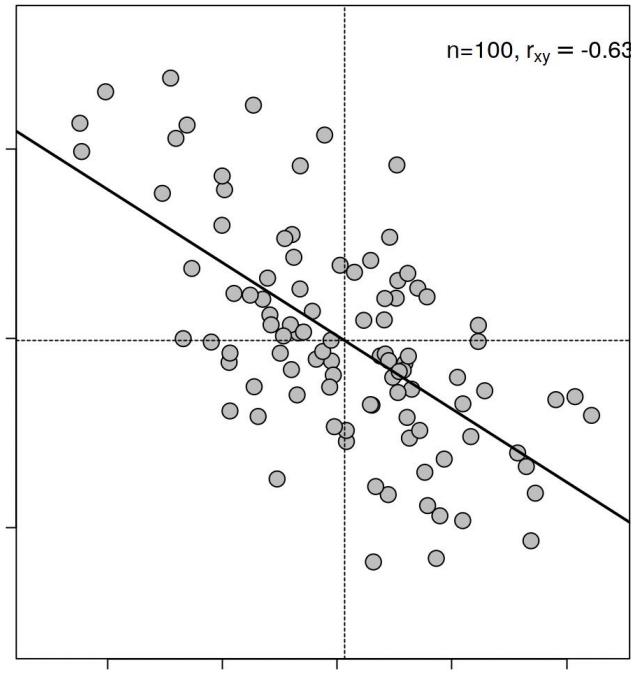
Factor variables

- Linear Regressions in R allow to work with factor variables.
- The variables display in the Table of Coefficients will always display all the categories but one.
- And it is important when interpreting the equation

Analysis of Residuals

QUESTIONS?

Fundamentals of Econometrics Models



Vicenç Soler

v.soler@tbs-education.org

~~vincent.soler@tbs-education.org~~