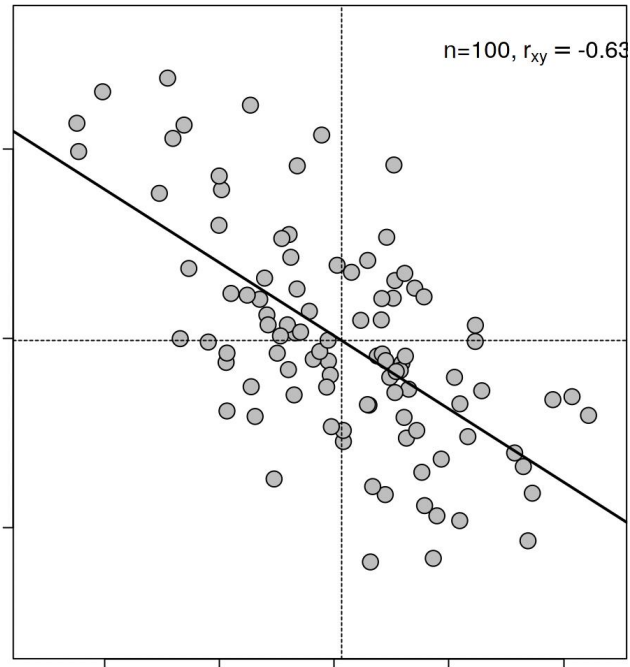


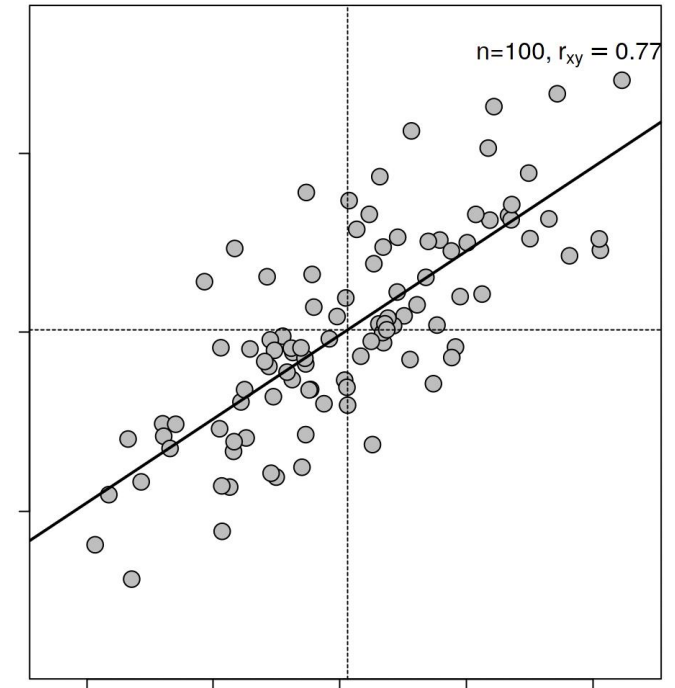
# Fundamentals of Econometrics Models



**Vicenç Soler**

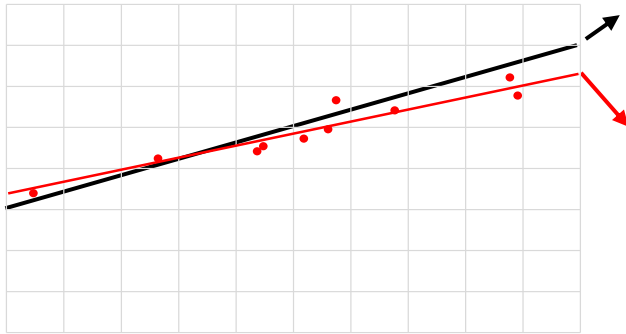
v.soler@tbs-education.org

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



# Fitted model and Theoretical model

Our data comes from a sample



Theor. model:

$$Y = 20 + 8X + \varepsilon$$

$$\varepsilon \sim N(0; 10)$$

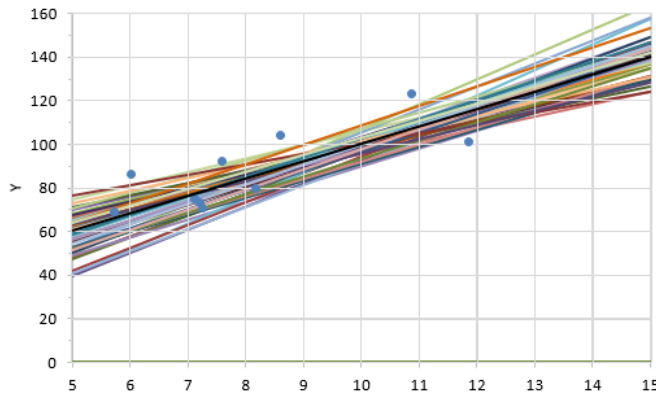
Unique but unknown

Fitted model:

sample of  $n = 10$

$$\hat{Y} = 41,9 + 5,6X + e$$

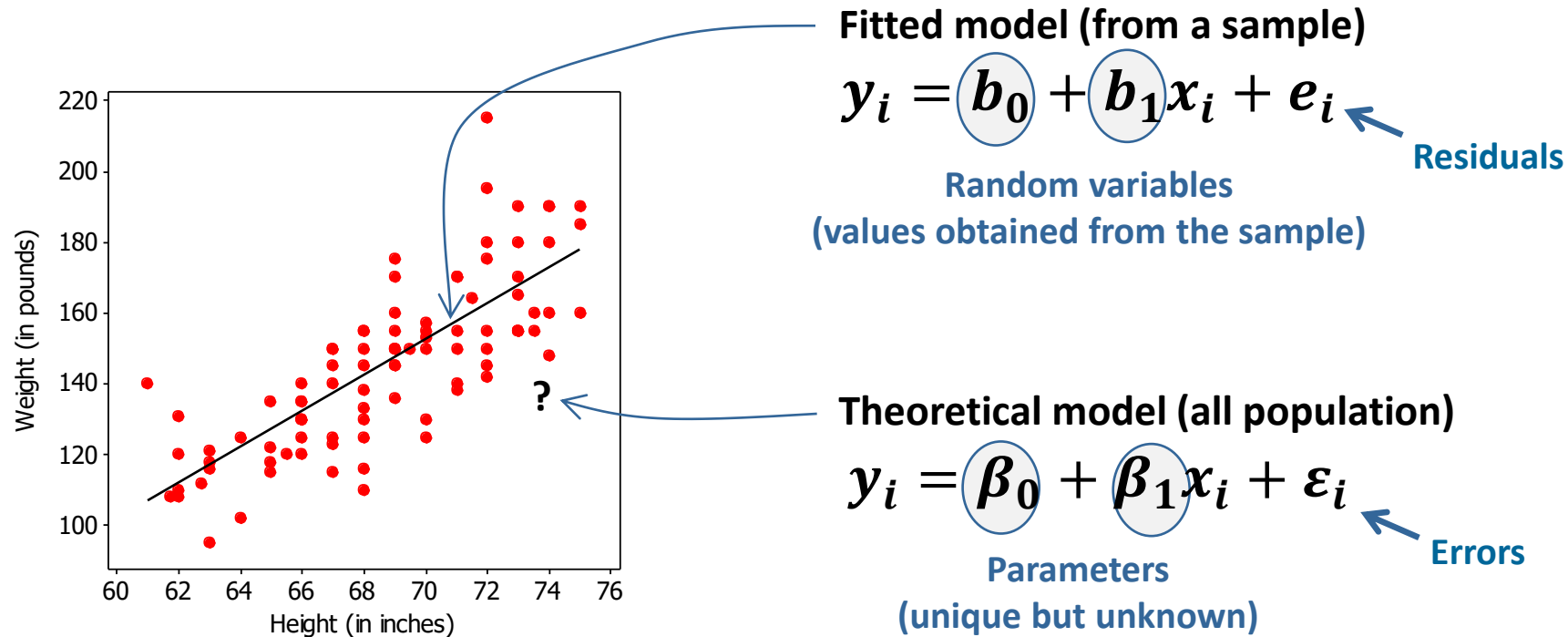
Known but they change  
from sample to sample



50 lines from 50 different samples

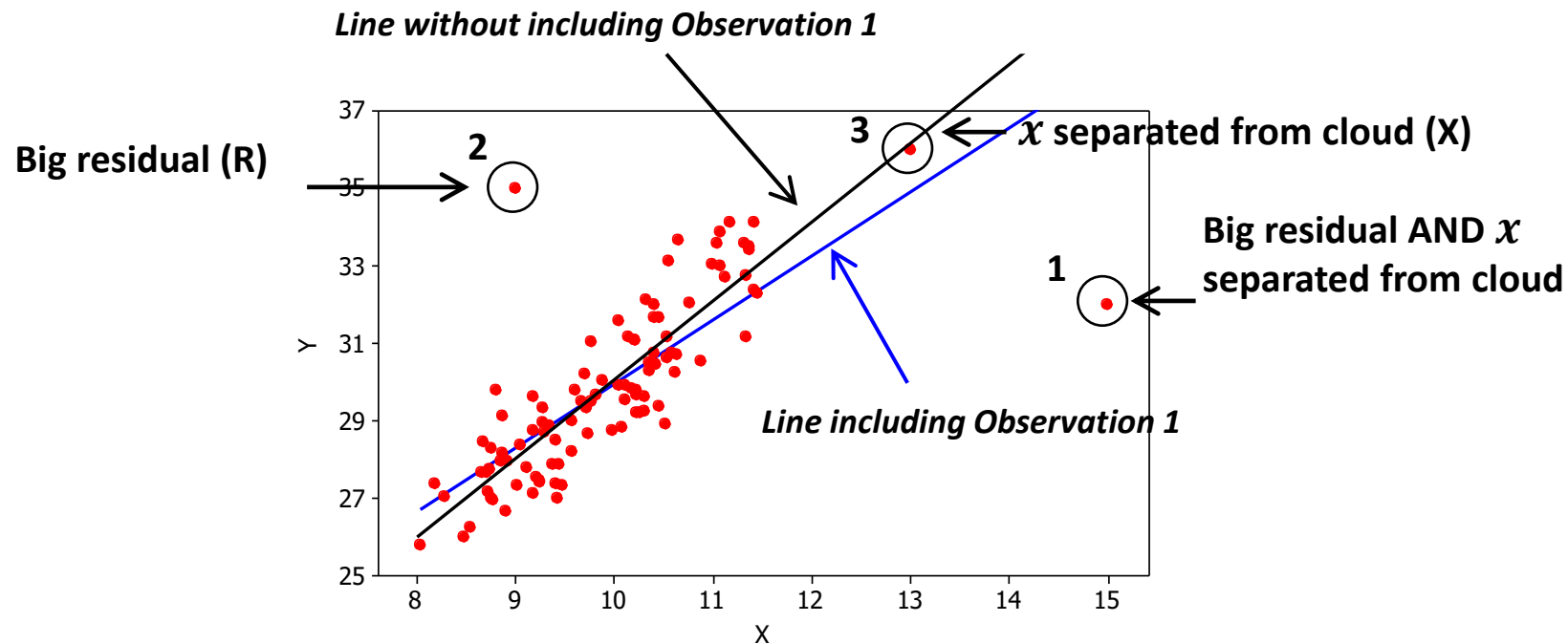
# Fitted model and Theoretical model

Our data comes from a sample



Sample vs. Population = Fitted model vs. Theoretical model

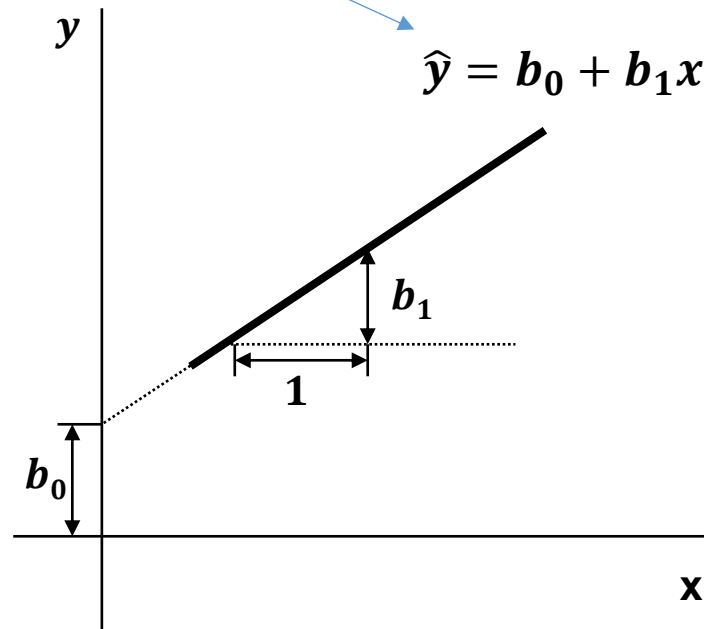
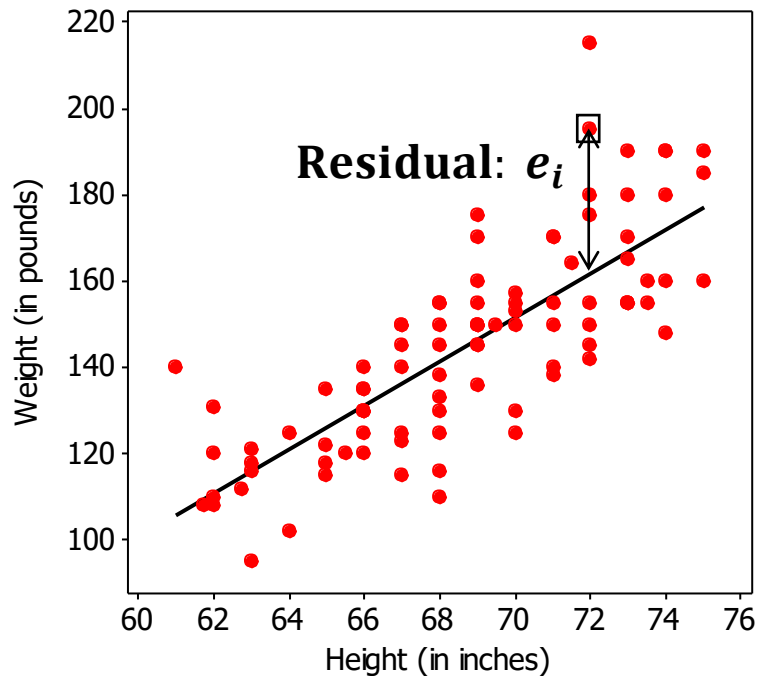
# Atypical observations



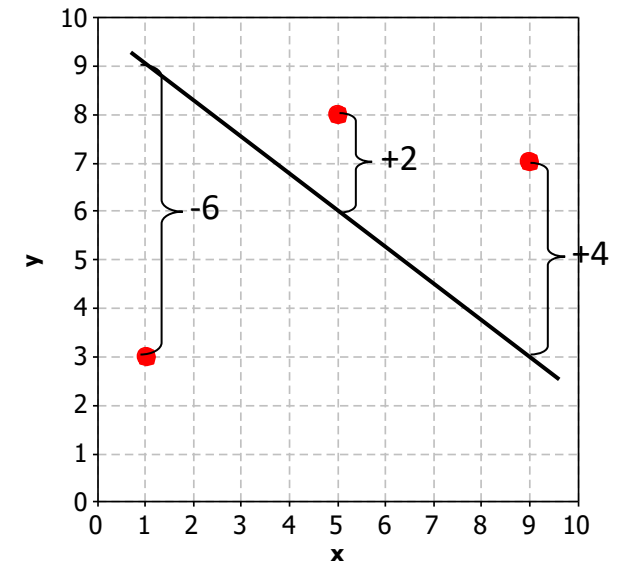
**As the goal is to generate a general model, it is always better to exclude the atypical values when generating the Regression Model**

# Residuals: Quick reminder

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



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



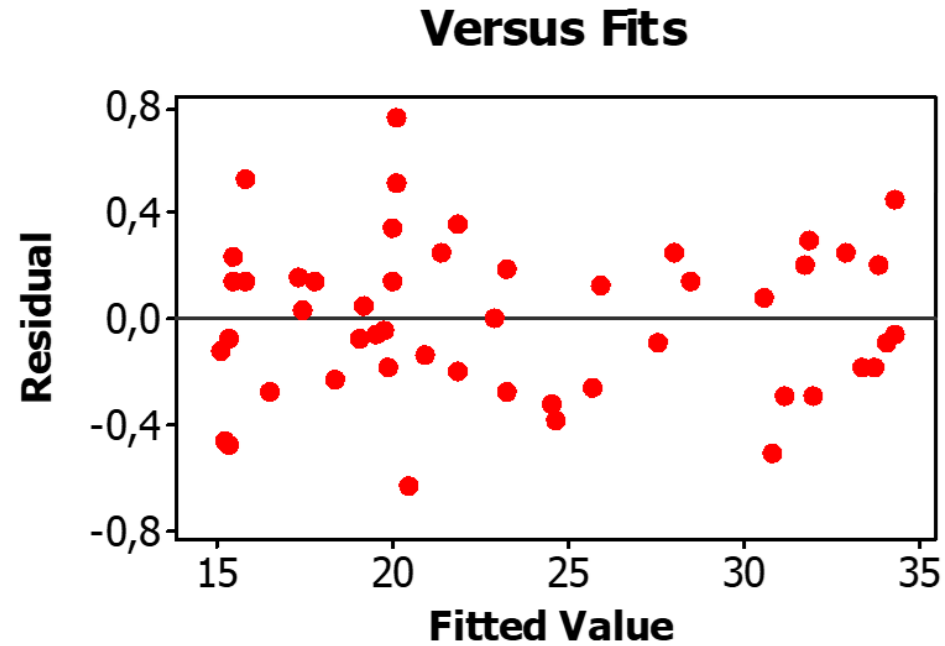
**GOAL:** To understand the relationship between  $y$  and  $x$  and/or to predict values of  $y$  given  $x$

# Model validation

**R<sup>2</sup> value**

**Analysis of the residuals**

# Analysis of the residuals

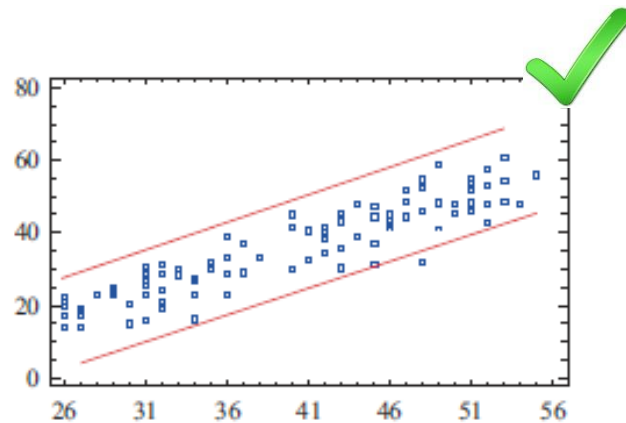


# Analysis of the residuals

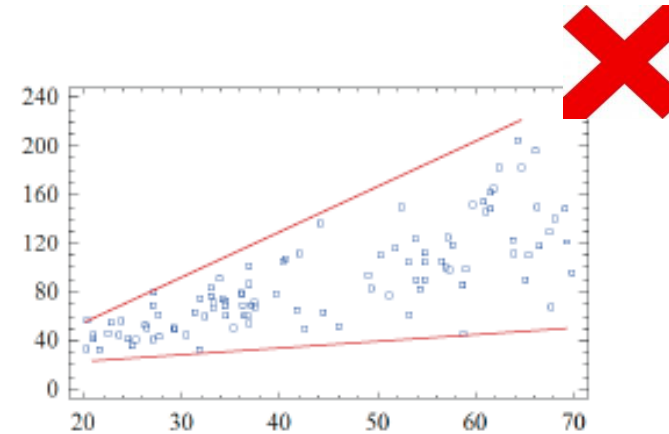
## Homoscedasticity vs Heteroscedasticity

***Constant variance in the residuals***

Dispersion must be constant



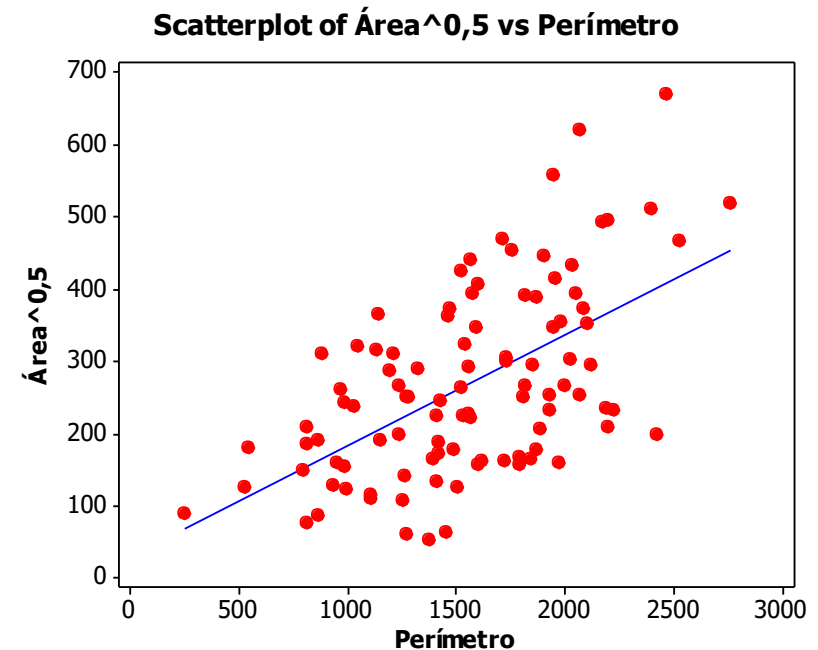
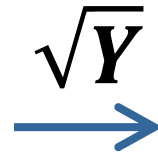
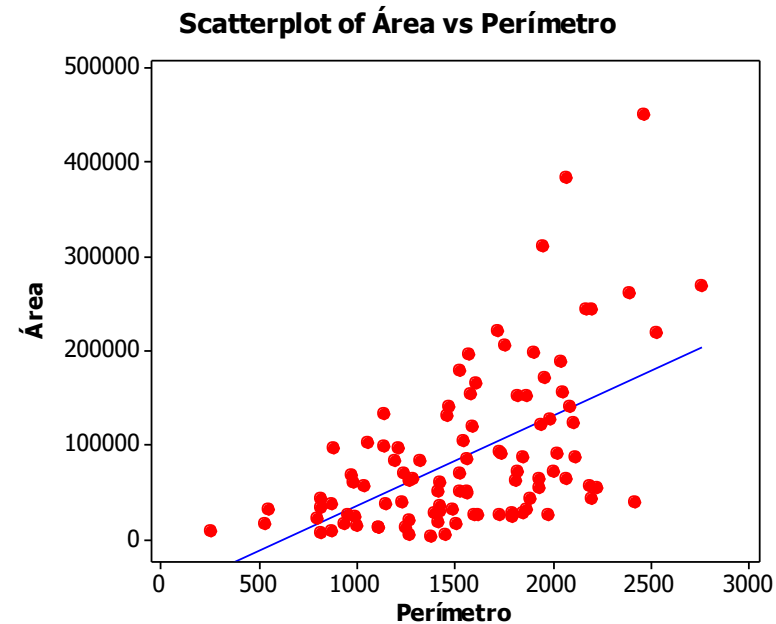
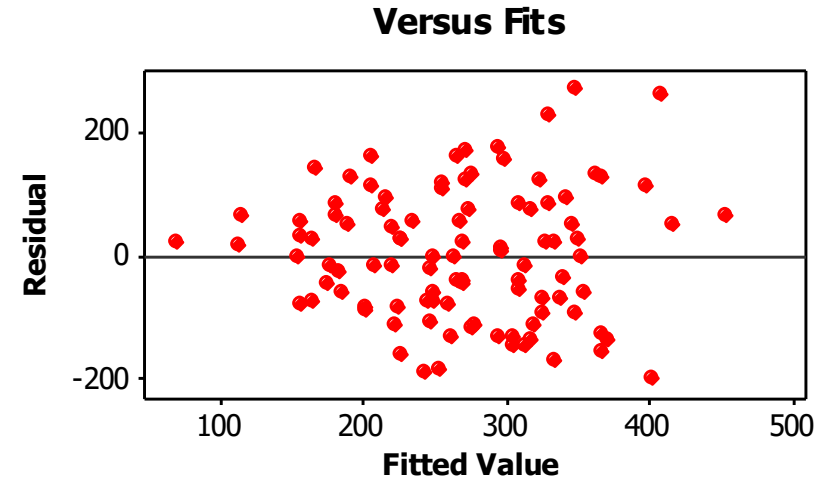
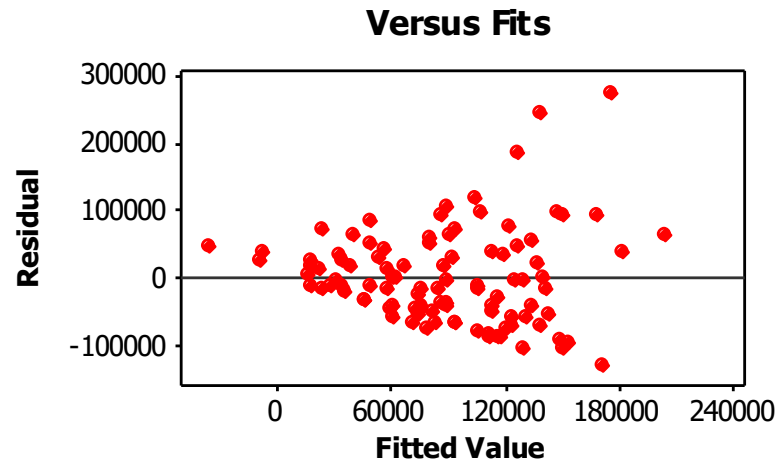
**Homoscedastic data**



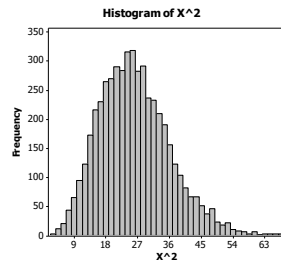
**Heteroscedastic data**



# Analysis of the residuals

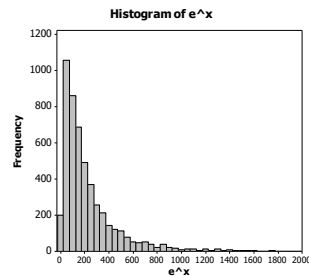


# We maybe need to transform data to convert it to linear

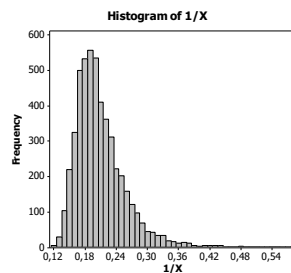


$$\sqrt{x}$$

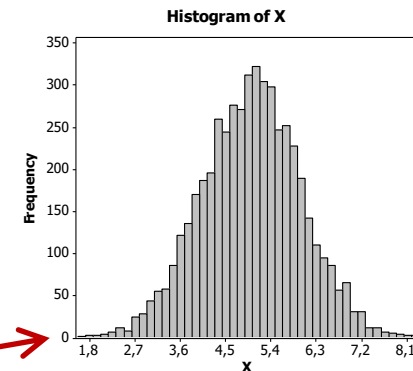
Transformation can be on  $x$ , on  $y$  or on both



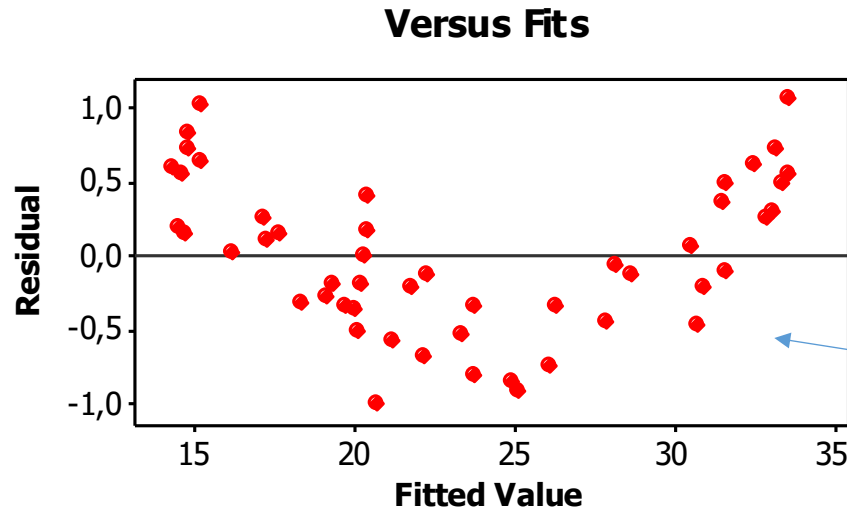
$$\ln(x)$$



$$1/x$$



# We maybe need to transform data to convert it to linear Example



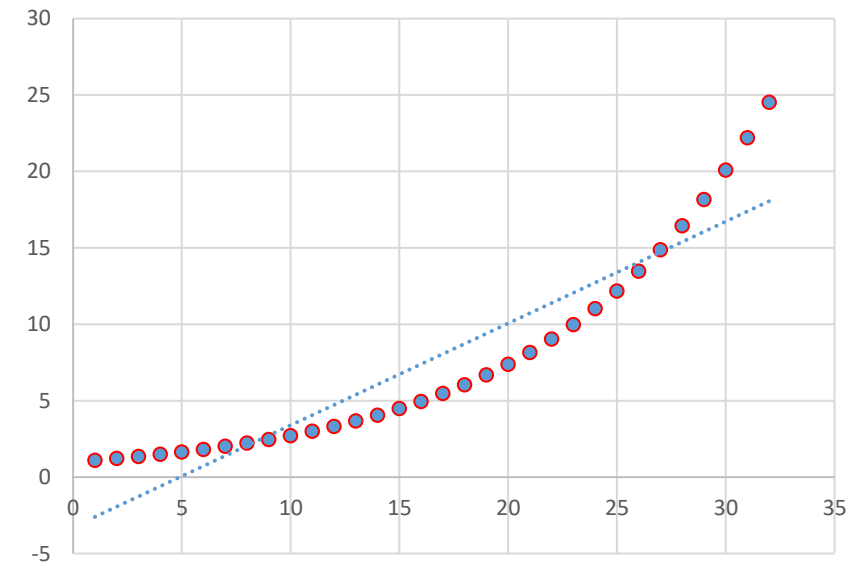
We should see a  
random pattern and  
this is not a random  
pattern

What about these residuals?

Positive->negative->positive

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

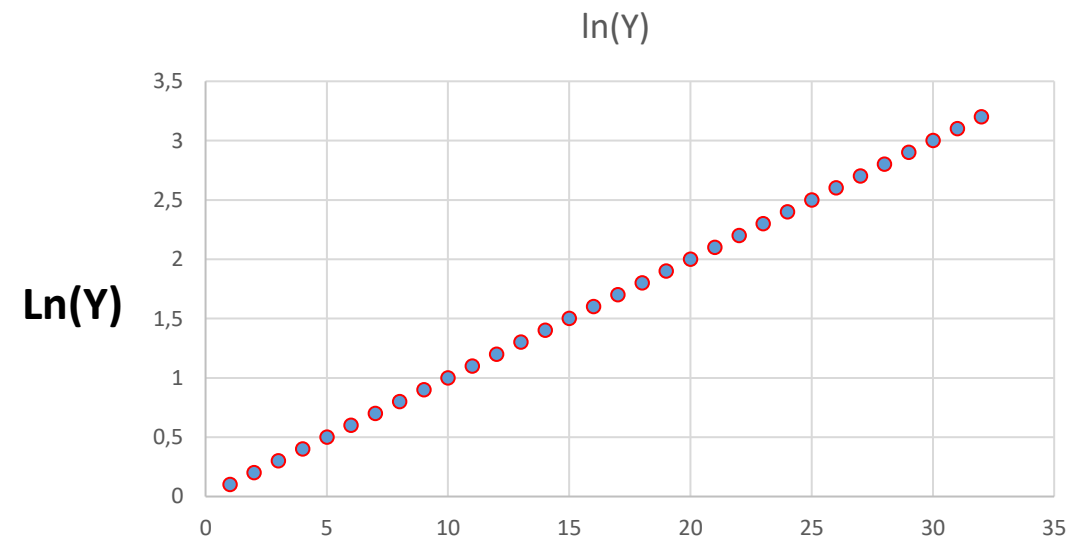
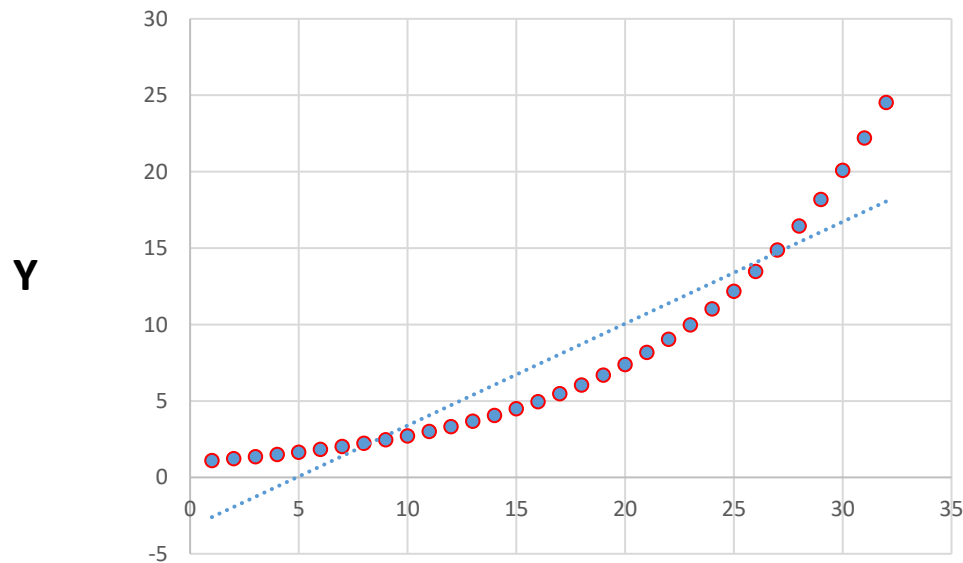
**Residuals = Actual - Predicted**



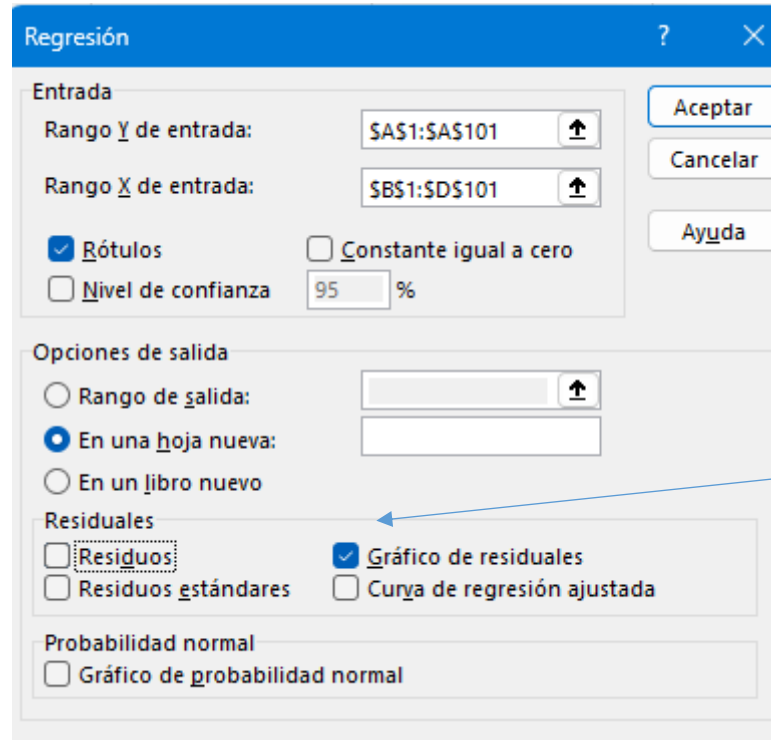
# We maybe need to transform data to convert it to linear

## 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.



# Residuals in Excel...



The image shows the 'Regresión' (Regression) dialog box in Excel. It is divided into several sections: 'Entrada' (Input), 'Opciones de salida' (Output options), 'Residuales' (Residuals), and 'Probabilidad normal' (Normal probability). In the 'Entrada' section, 'Rango Y de entrada' is '\$A\$1:\$A\$101' and 'Rango X de entrada' is '\$B\$1:\$D\$101'. Under 'Opciones de salida', 'En una hoja nueva' is selected. In the 'Residuales' section, 'Gráfico de residuales' (Residuals chart) is checked, while 'Residuos' (Residuals), 'Residuos estándares' (Standardized residuals), and 'Curva de regresión ajustada' (Adjusted regression curve) are unchecked. The 'Probabilidad normal' section has 'Gráfico de probabilidad normal' (Normal probability plot) unchecked. Buttons for 'Aceptar' (OK), 'Cancelar' (Cancel), and 'Ayuda' (Help) are on the right.

**Regresión**

**Entrada**

Rango Y de entrada: \$A\$1:\$A\$101

Rango X de entrada: \$B\$1:\$D\$101

☒ **R**ótulos ☐ **C**onstante igual a cero

☐ **N**ivel de confianza 95 %

**Opciones de salida**

☐ Rango de salida:

☒ **E**n una hoja nueva:

☐ **E**n un libro nuevo

**Residuales**

☐ **R**esiduos ☒ **G**ráfico de residuales

☐ **R**esiduos estándares ☐ **C**urva de regresión ajustada

**Probabilidad normal**

☐ **G**ráfico de probabilidad normal

Aceptar Cancelar Ayuda

Enable it to display  
graphs of residuals

# Analysis of Residuals

## Transformations ( $\ln$ , $\sqrt{x}$ , $x^2$ , 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

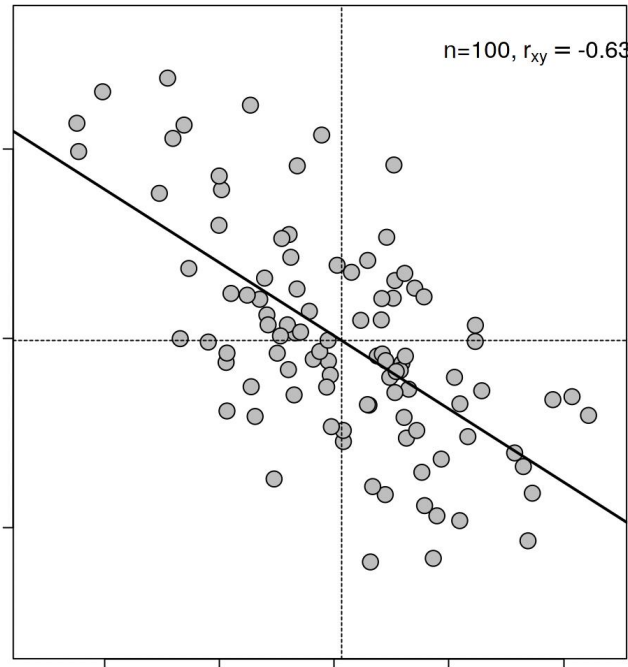
# Factor variables

- 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



# QUESTIONS?

# Fundamentals of Econometrics Models



**Vicenç Soler**

`v.soler@tbs-education.org`

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

