



PRACTICE UNIT 8 – NON-LINEAR REGRESSION MODELS

OBJECTIVES

The objectives of this practice are:

- Analyse data obtained in a survey and strengthen the use of the Excel spreadsheet.
- Fit non-linear models to estimate the relationship between two variables. Measure the goodness of fit of these models and use them to make predictions.

EXERCISE U8_1. *Non-linear regression model with data from a survey* (Sheet '*Non-linear Height-Weight*')

From the survey conducted with the students of the course Statistics I carried out during the academic year 2014-2015 (same of previous practices) we have selected the values of their *Height* and *Weight*.

- A) Using the graphic presentation obtained in the Practice of Unit 7 estimate the exponential, logarithmic, polynomial (order 2) and power regression functions. Show in the same graph the measures of the goodness of these fits.
- B) Using the previous models, what would be the weight of a student 170 cm height?
- C) Which model would you select as the best one?

EXERCISE U8_2. *Linear and non-linear regression* (Sheet '*Machine*')

The manager of a company wants to analyse the relationship between the *experience*, in days, of the workers *in operating a machine* (X) and the *percentage of defective parts produced* (defaults rate) by the workers *when operating the machine* (Y). Sheet '*Machine*' shows the data gathered for this study.

- A) Obtain the least squares regression line that explains Y based on X. Interpret the parameters of the estimated regression line. Analyse the goodness of fit of this model.
- B) Estimate a power model that explains Y based on X ($Y = aX^b$) transforming the variables in the appropriate way to obtain a linear model and using the Excel functions.
- C) Analyse the goodness of fit of this last model using the general coefficient of determination.



Universidad
Zaragoza

- D)** Present the data graphically using a scatter plot and show the two previous regression models. Compare the value of the measure of the goodness of fit shown in the graph with the one calculated in the previous question.
- E)** Determine which of the two models fits the data better and use it to predict what the defaults rate would be for a worker with 100 days of experience in operating the machine. Is this prediction reliable?