Chapter IX: Regression - Exercises

Bernardo D'Auria

Statistics Department

Universidad Carlos III de Madrid

GROUP 89 - COMPUTER ENGINEERING

December 14th, 2010

Exercise

The file *FrenosITV.sf3* has some properties of a sample of vehicles. We want to know if the age of the car and its power help to predict the effectiveness of braking. Build a multiple regression model that predicts the efficiency (variable "*Efficacia*") as a function of the kilometers traveled by car (variable "*KM* ") and its power (variable "*Potencia*").

Exercise

The file *FrenosITV.sf3* has some properties of a sample of vehicles. We want to know if the age of the car and its power help to predict the effectiveness of braking. Build a multiple regression model that predicts the efficiency (variable "*Efficacia*") as a function of the kilometers traveled by car (variable "*KM* ") and its power (variable "*Potencia*").

SOLUTION:

The variables *KM* and *Potencia* are not significant to predict the effectiveness of braking.

Exercise

Let denote by *PM10* the small sized solid or liquid particles that are dispersed in the atmosphere. High concentrations of PM10 may be harmful to health. The file *ConcentraPM10.sf3* contains a sample of 500 hourly observations of PM10 concentration together with other variables. These variables are:

- "*Coches*" (cars) = number of cars that passed in front of the concentration meter during the hour of measurement
- "*Temperatura*" (temperature) = Air temperature measured at a height of 2m from ground
- "*Viento*" (wind) = wind speed (m/s)
- "*Hora*" (hour) = day time

It calls for:

a) Build a model to explain the concentration of PM10 as a function of wind speed. It is assumed that the greater the wind speed, the cleaner the air is as PM10 particles get dispersed. The effect can be nonlinear and maybe if could be necessary to apply transformations to the variables.

b) Build a model to explain the concentration of PM10 as a function of car traffic, temperature and wind speed. Analyze the residuals. The model only has to contain significant variables.

c) Some analysts believe that sunlight reduces the concentration of PM10 by destroying some particles by biochemist effect. Other analysts say that the fact that during the night there is less PM10 is just due to the effect of less car traffic. Knowing that the daylight hours in the place where data were taken range from 7AM to 6PM, analyze which of two groups of analysts (or both) is right.

SOLUTION

a) Looking at the dispersion graph one can observe that when the wind is stronger the PM10 concentration is lower, but the relation is not linear especially for very high wind speeds. The relation looks quadratic and a good transformation is using $viento^c$ with $c \approx 0.3$ or as well $viento^2$. In this case one can just build a simple regression model.

b) Using the relation with the wind speed found in the answer above and adding temperature and cars we can notice that temperature is a no significant variable

c) Using a binary variable $sunlight = (Hora \geq 7) * (Hora \leq 18)$ we can see that indeed the *sunlight* is significant with respect the PM10 concentration. However its coefficient is negative such that concentration is higher during light hours and therefore the first analysts are wrong. On the other side adding the variable “Coches” to the model we see that concentration is not affected by the presence or absence of sunlight but only by the car traffic conditions. Therefore the second analysts seem to be right.

Exercise

Using the file *AlumnosIndustriales.sf3* contrast the following statements:

a) Boys are usual to carry more money than girls

b) Students who live farther away (takes longer to get to college) have more money with them

c) A boy and girl who are just as high, will have, on average, the same shoe size

Exercise

Using the file *AlumnosIndustriales.sf3* contrast the following statements:

a) Boys are usual to carry more money than girls

b) Students who live farther away (takes longer to get to college) have more money with them

c) A boy and girl who are just as high, will have, on average, the same shoe size

Solution

a) Doing a simple regression with $Y = dinero$ and $X = sexo$, it results no significant.

b) Doing a simple regression with $Y = dinero$ and $X = tiempo$, it results no significant.

c) Doing a multiple regression with $Y = zapato$, $X_1 = sexo$ and $X_2 = altura$, the variable $sexo$ keeps being significant and so we reject the hypothesis.

Exercise

The file *FrenosITV.sf3* contains data of a set of cars that come to the ITV station of Leganes.
The variable "*eficacia*" (effectiveness) takes values between 0 and 100 and measures the effectiveness of braking, so that better efficiency means better braking, in the sense that the pressure to brake is stronger.
The "*ABS*" variable has the value 1 if the vehicle has ABS and 0 if it has not ABS system.
Do vehicles with ABS have a braking more effective?

Exercise

The file *FrenosITV.sf3* contains data of a set of cars that come to the ITV station of Leganes.
The variable "*eficacia*" (effectiveness) takes values between 0 and 100 and measures the effectiveness of braking, so that better efficiency means better braking, in the sense that the pressure to brake is stronger.
The "*ABS*" variable has the value 1 if the vehicle has ABS and 0 if it has not ABS system.
Do vehicles with ABS have a braking more effective?

Solution

The variable "*ABS*" as explicative variable does not result to be significant.