Multivariate regression

Marius Ionescu

09/29/2011

Linear regression

Fact

Recall:

- *regression line studies the relationship between two variables.*
- *the independent variable is called X.*
- *the dependent variable is called Y.*
- *We can write the equation of the line as*

$$Y = a_0 + a_1 X.$$

Multivariate regression

Fact

- *Usually there are more than one independent variables that contribute to the behavior of Y.*
- *For example, to study the inflation in US, one should include the unemployment, GNP, and the deficit.*
- *We say that the unemployment, GNP, and the deficit are the independent or X variables.*
- *We label them X_1, X_2, and X_3.*
- *We label the dependent variable Y (in this example the inflation in US).*

Multivariate regression

Definition

If there is more than one explanatory variable (X_1, X_2, X_3 say) and one response variable (Y), it may be useful to model it as

$$y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3.$$

Example

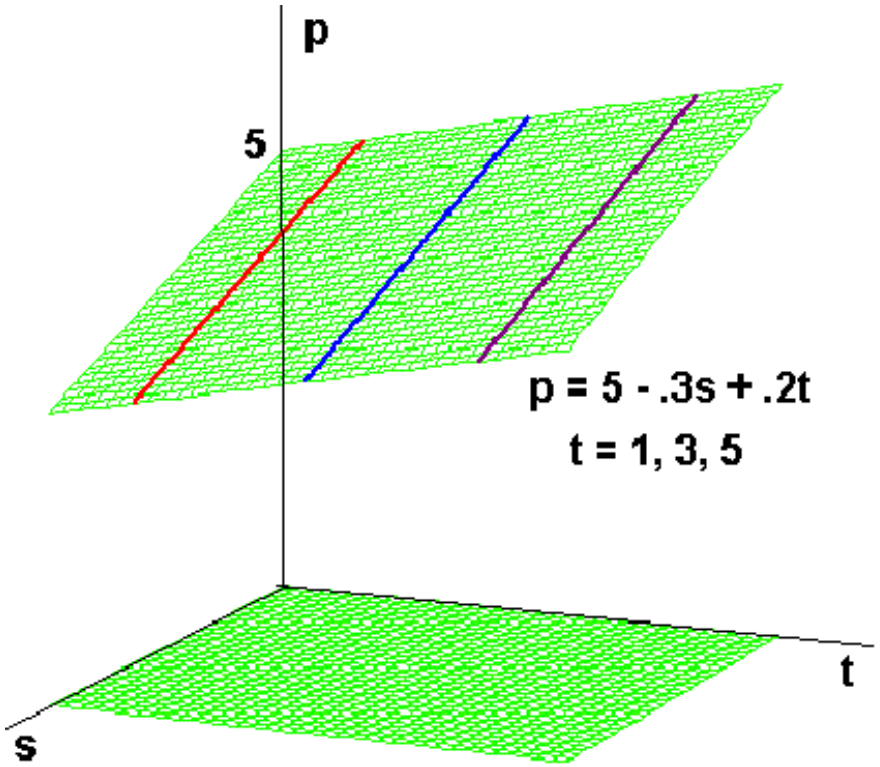
Example

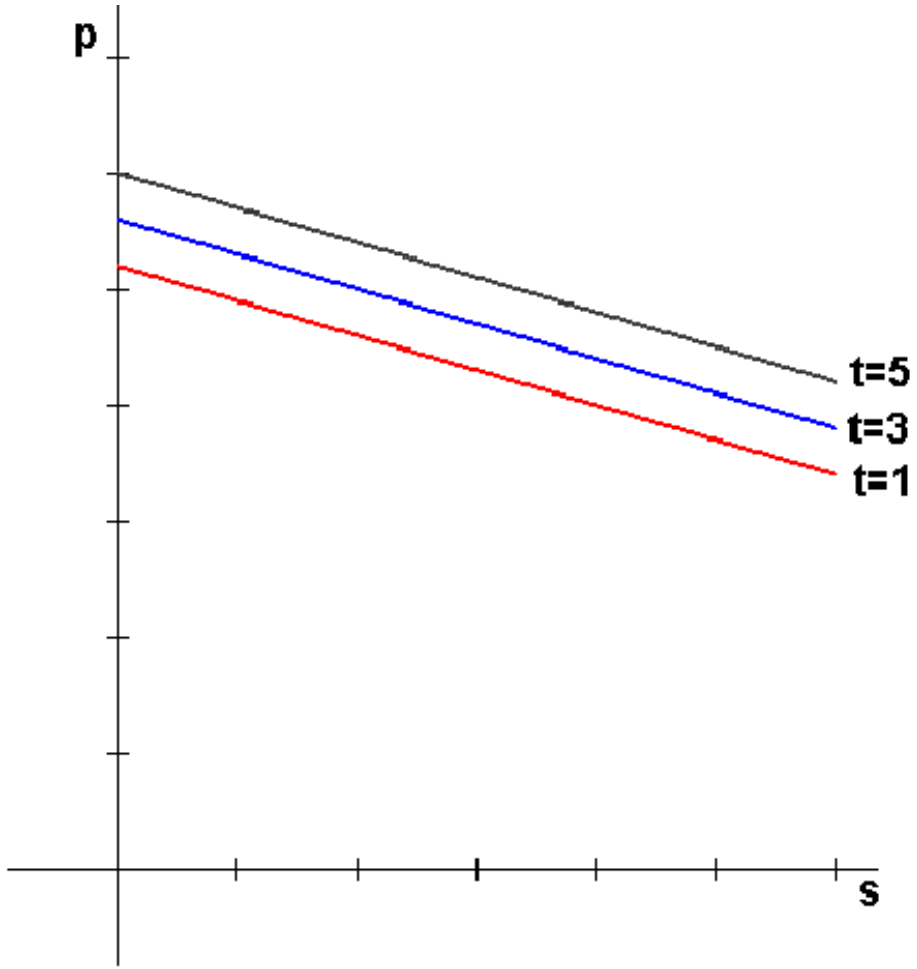
Aspirin is so acidic that it often upsets the stomach, so it is often administered with an antacid – which limits its effect. Suppose the pain, measured by the rating of headache sufferers, is given by

$$p = 5 - .3s + .2t,$$

where s is the aspirin dose and t is the antacid dose.

Graphs of the aspirin example





Correlation coefficient

Fact

- *As with simple regression, there is a (multiple) correlation R (independent of units) that measures how closely the data points (in 3-space or higher dimensions) follow a (hyper)plane.*
- *R is a number between 0 and 1.*
- *To interpret the direction of the relationship between variables, one looks at the signs (plus or minus) of the regression coefficients a_1, a_2, a_3.*
- *If a coefficient is positive, then the relationship of this variable with the dependent variable is positive;*
- *if the coefficient is negative then the relationship is negative.*

R^2

Fact

- *R^2 says how much better for predicting y is using regression line (i.e., using the y-value $\widehat{y}$ on the regression line at that point) than just always using Avg_Y.*
- *If $R^2 = 0.4$, say, "regression results in a 40% improvement in projection".*
- *It appears in the computer outputs as well.*
- *Let's do an example in Excel.*