CLARK UNIVERSITY

Multiple linear regression

Math 218, Mathematical Statistics

D Joyce, Spring 2016

This is a very common method used to understand numerical relations among numerical factors. For this method, we assume that we have k independent variables $x_1, \ldots, x_k$ that we can set, then they probabilistically determine an outcome Y. Furthermore, we assume that Y is linearly dependent on the factors according to

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

where ϵ is a normal error. This is just like we had for simple linear regression except k doesn't have to be 1.

In an experiment, we have n observations, n typically being much more than k. For the i^{th} observation we set the independent variables to the values

$$x_{i1}, x_{i2} \ldots, x_{ik}$$

and measure a value y_i for the random variable Y_i. Thus, the model can be described by the equations

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i$$

for $i = 1, 2, \ldots, n$, where the errors ϵ_i are independent normal variables each with mean 0 and the same unknown variance σ^2.

Altogether this model for multiple linear regression has $k+2$ unknown parameters: $\beta_0, \beta_1, \ldots, \beta_k$, and σ^2.

When k was equal to 1, we found the least squares line $y = \hat\beta_0 + \hat\beta_1 x$. It was a line in the plane $\mathbf{R}^2$. Now with $k \geq 1$, we'll have a least squares hyperplane

$$y = \hat\beta_0 + \hat\beta_1 x_1 + \hat\beta_2 x_2 + \cdots + \hat\beta_k x_k$$

in $\mathbf{R}^{k+1}$. The way to find the estimators $\hat\beta_0, \hat\beta_1, \ldots$, and $\hat\beta_k$ is exactly the same, namely, take the partial derivatives of the squared error

$$Q = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}))^2$$

with respect to the $k+1$ variables $\beta_0, \ldots, \beta_k$, set them all to 0, and solve to find the critical point. There will only be one critical point except in exceptional situations, like when $n < k$, and it will give the minimum least squared error Q. To find that value, $k+1$ linear equations need to be solved simultaneously for $k+1$ unknowns, so methods of linear algebra are needed, but not very advanced methods since it's just solving a simultaneous system of linear equations.

When that system is solved we have fitted values

$$\hat y_i = \hat\beta_0 + \hat\beta_1 x_{i1} + \hat\beta_2 x_{i2} + \cdots + \hat\beta_k x_{ik}$$

for $i = 1, \ldots, n$ that should be close to the actual values y_i. The differences are the residuals

$$e_i = y_i - \hat y_i.$$

The error sum of squares, SSE, is defined as before, namely as

$$\text{SSE} = \sum e_i^2$$

and it is used along with the total sum of squares,

$$\text{SST} = \sum (y_i - \overline{y})^2$$

and the regression sum of squares

$$\text{SSR} = \sum (\hat y_i - \overline{y})^2$$

exactly as in the case when $k = 1$, so that

$$\text{SST} = \text{SSE} + \text{SSR}$$

and

$$r^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}.$$

The only difference is that we don't have a correlation coefficient r. Instead, we define the positive square root of r^2 to be r and we call it the *multiple correlation coefficient*.

We'll look at example 11.1 in the text next.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`