Stat 20: A guide to estimating regression parameters

B. M. Bolstad, bolstad@stat.berkeley.edu

November 24, 2003

The goal of this document is to outline the steps that you should go through to estimate regression parameters in this class. This main text should be used in connection with the flow diagram which gives you a decision guide for the process of estimating regression coefficients.

1 Writing the regression model in the general form

1.1 Theory

We can write the general regression model in the form

$$\mu(\mathbf{x}_i) = \beta_0 g_0(\mathbf{x}_i) + \beta_1 g_1(\mathbf{x}_i) + \cdots + \beta_p g_p(\mathbf{x}_i) \tag{1}$$

the $\beta_0, \ldots, \beta_p$ are called the regression parameters. They are unknown and we will estimate them as part of the regression process. The $g_0(\mathbf{x}_i), \ldots, g_p(\mathbf{x}_i)$ are called the basis functions. They represent general functions of the $\mathbf{x}$ data. In practice we will have a regression model and will be able to identify the basis functions by looking at the model.

1.2 Some examples

Consider $\mathbf{x} = x$. The following table shows some regression models and identifies the basis functions

Regression Model	Basis functions
$\mu(x) = \beta_0 + \beta_1 x$	$g_0(x) = 1,\ g_1(x) = x$
$\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2$	$g_0(x) = 1,\ g_1(x) = x,\ g_2(x) = x^2$
$\mu(x) = \beta_0 + \beta_1 \frac{1}{x} + \beta_2 x^2$	$g_0(x) = 1,\ g_1(x) = \frac{1}{x},\ g_2(x) = x^2$
$\mu(x) = \beta_0 + \beta_1 \frac{1}{x^2} + \beta_2 \sin(x^2)$	$g_0(x) = 1,\ g_1(x) = \frac{1}{x^2},\ g_2(x) = \sin(x^2)$

Rather than just a single x value lets consider $\mathbf{x} = (x_1, x_2, x_3)$. The following table shows some regression models and identifies the basis functions.

Regression Model	Basis functions
$\mu(\mathbf{x}) = \beta_0 + \beta_1 x_1$	$g_0(\mathbf{x}) = 1,\ g_1(\mathbf{x}) = x_1$
$\mu(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	$g_0(\mathbf{x}) = 1,\ g_1(\mathbf{x}) = x_1,\ g_2(\mathbf{x}) = x_2,\ ,\ g_3(\mathbf{x}) = x_3$
$\mu(\mathbf{x}) = \beta_0 + \beta_1 x_1^2 + \beta_2 x_1 x_2 + \beta_3 \cos(x_3)$	$g_0(\mathbf{x}) = 1,\ g_1(\mathbf{x}) = x_1^2,\ g_2(\mathbf{x}) = x_1 x_2,\ ,\ g_3(\mathbf{x}) = \cos(x_3)$
$\mu(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_1 x_2 + \beta_3 x_1 x_2 x_3$	$g_0(\mathbf{x}) = 1,\ g_1(\mathbf{x}) = x_1,\ g_2(\mathbf{x}) = x_1 x_2,\ ,\ g_3(\mathbf{x}) = x_1 x_2 x_3$

2 Checking that the basis is orthogonal

2.1 Theory

Given any two basis function $g_j(\mathbf{x}), g_k(\mathbf{x})$ we say that they are orthogonal if and only if

$$\sum_{i=1}^{n} g_j(\mathbf{x}_i) g_k(\mathbf{x}_i) = 0 \tag{2}$$

We say a set of basis functions is orthogonal if $g_j(\mathbf{x}), g_k(\mathbf{x})$ are orthogonal for all possible j,k with $j \neq k$.

2.2 Some examples

Consider the basis function 1, x. To check whether the basis is orthogonal you need to check whether $\sum_{i=1}^{n}(1)(x_i) = \sum_{i=1}^{n} x_i = 0$ for your data set. For instance suppose that $x = (1, 0, -1, 2, 0, -2, 3, 0, -3)$ then $\sum x_i = 1 + 0 - 1 + 2 + 0 - 2 + 3 + 0 - 3 = 0$ so the basis 1 and x is orthogonal for our data. If our basis functions were 1 and x^2 then we would need to check that $\sum_{i=1}^{n}(1)(x_i^2) = \sum_{i=1}^{n} x_i^2 = 0$ assuming we had the same data then $\sum_{i=1}^{n} x_i^2 = 1 + 0 + 1 + 4 + 0 + 4 + 9 + 0 + 9 = 28$ which is not equal to 0 so the functions 1 and x^2 are not an orthogonal basis.

Now for a more complicated example. Consider the following data

x_1	x_2
1	-1
0	-1
-1	-1
1	0
0	0
-1	0
1	1
0	1
-1	1

If our basis functions are 1, x_1, x_2 then to check that the basis is orthogonal we check that $\sum_{i=1}^{n}(1)(x_{i1}) = \sum_{i=1}^{n} x_{i1} = 1+0+-1+1+0+-1+1+0-1 = 0$, $\sum_{i=1}^{n}(1)x_{i2} = \sum_{i=1}^{n} x_{i2} = -1-1-1+0+0+0+1+1+1 = 0$ and finally $\sum_{i=1}^{n} x_{i1}x_{i2} = 1*-1+0*1+-1*1+1*0+0*0+-1*0+1*1+0*1+-1*1 = 0$. So the basis is orthogonal.

Using the same data suppose instead that the basis functions are 1, x_1, $x_2 - 4$. then to check if the basis is orthogonal we check that $\sum_{i=1}^{n}(1)(x_{i1}) = \sum_{i=1}^{n} x_{i1} = 1 + 0 + -1 + 1 + 0 + -1 + 1 + 0 - 1 = 0$, $\sum_{i=1}^{n}(1)(x_{i2}-4) = (-1-4)+(-1-4)+(-1-4)+(0-4)+(0-4)+(0-4)+(1-4)+(1-4)+(1-4) = -5-5-5-4-4-4-3-3-3 = -36$ and so 1 and x_2 are not orthogonal therefore the basis is not orthogonal.

3 Estimating the parameters of a regression model: The Least squares method

3.1 Theory

The goal of the least squares method is to choose the values of $\beta_0, \beta_1, \ldots, \beta_p$ which minimize the error sum of squares (SSE)

$$\sum_{i=1}^{n} \left(Y_i - \beta_0 g_0\left(\mathbf{x}_i\right) - \beta_1 g_1\left(\mathbf{x}_i\right) - \cdots - \beta_p g_p\left(\mathbf{x}_i\right)\right)^2$$

by differentiating with respect to each of the $\beta_0, \beta_1, \ldots, \beta_p$ giving us a system of $p+1$ equations each of which we set equal to zero. This set of $p+1$ equations is called the *normal equations*. Note that in general you can show by differentiating the SSE above with respect to each of $\beta_0, \beta_1, \ldots, \beta_p$ that the general form for the normal equations is given by

$$\hat{\beta}_0 \sum_{i=1}^{n} g_j\left(\mathbf{x}_i\right) g_0\left(\mathbf{x}_i\right) + \cdots + \hat{\beta}_p \sum_{i=1}^{n} g_j\left(\mathbf{x}_i\right) g_p\left(\mathbf{x}_i\right) = \sum_{i=1}^{n} g_j\left(\mathbf{x}_i\right) Y_i \text{ for } j = 0, \ldots, p$$

The least squares estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ are then found by solving the system of equations.

3.2 Some examples

The least squares method was used to derive formula for the regression model $\mu(x) = \beta_0 + \beta_1 x$, see the notes from Nov 10.

Consider the regression model $\mu(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. The sum of squared errors for this model will be

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}\right)^2$$

we want the values of β_0, β_1 and β_2 which minimize this sum of squares. Call these values $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. To find these we must differentiate the sum of square errors with respect to β_0 and with respect to β_1 and with respect to β_2. This gives three equations which we set equal to zero (these are called the normal equations) and then solve for the $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. So differentiating the sum of squared error with respect to β_0 and setting equal to zero we get

$$\sum_{i=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right)(-1) = 0$$

(we put the hat on each β because we set the equation equal to 0) after simplifying a little we get the first normal equation

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} - \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} = 0$$

Differentiating the sum of squared error with respect to β_1 and setting equal to zero we get

$$\sum_{i=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right)(-x_{i1}) = 0$$

after simplifying a little we get the second normal equation

$$\sum_{i=1}^{n} x_{i1}y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_{i1} - \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 - \hat{\beta}_2 \sum_{i=1}^{n} x_{i1}x_{i2} = 0$$

Finally differentiating with respect to β_2 and setting equal to zero we get

$$\sum_{i=1}^{n} 2\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right)(-x_{i2}) = 0$$

after simplifying a little we get the third normal equation

$$\sum_{i=1}^{n} x_{i2}y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_{i2} - \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}x_{i2} - \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}^2 = 0$$

The three equations can then be solved for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. Rather than just solving these equations and getting general formula's for this model we should instead evaluate the summary statistics using the data, since this will make manipulating the equations much easier. Suppose that we have the following data:

y	x_1	x_2
21	1	-1
15	0	-1
13	-1	-1
20	1	0
25	0	1
18	-1	0
29	1	1
22	0	1
21	-1	1

And so we see that $n = 9$, $\sum_{i=1}^{n} y_i = 184$, $\sum_{i=1}^{n} x_{i1} = 0$, $\sum_{i=1}^{n} x_{i2} = 1$, $\sum_{i=1}^{n} x_{i1}y_i = 18$, $\sum_{i=1}^{n} x_{i2}y_i = 48$, $\sum_{i=1}^{n} x_{i1}x_{i2} = 0$, $\sum_{i=1}^{n} x_{i1}^2 = 6$ and $\sum_{i=1}^{n} x_{i2}^2 = 7$. Thus, the normal equations become

$$184 - 9\hat{\beta}_0 - \hat{\beta}_2 = 0$$

$$18 - 6\hat{\beta}_1 = 0$$

$$48 - \hat{\beta}_0 - 7\hat{\beta}_2 = 0$$

Now we just solve the three equations for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. The second equation gives $\hat{\beta}_1 = 18/6 = 3$. From the first equation we get $\hat{\beta}_2 = 184 - 9\hat{\beta}_0$. Substituting this into the third equation we get

$$48 - \hat{\beta}_0 - 7(184 - 9\hat{\beta}_0) = 0$$

and so we get

$$-1240 + 62\hat{\beta}_0 = 0$$

and so $\hat{\beta}_0 = 1240/62 = 20$ and $\hat{\beta}_2 = 184 - 9(20) = 4$.

3.3 The special case of the model $\mu(x) = \beta_0 + \beta_1 x$

For the special of the simple linear regression model $\mu(x) = \beta_0 + \beta_1 x$ we showed (see 10th Nov), using the least squares approach that for this particular model we could estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ using

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{3}$$

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{y}\bar{x}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Alternatively, we showed on Nov 12 that formula given by your textbook

$$\hat{\beta}_1 = r\frac{s_y}{s_x} \tag{4}$$

could be used in place of the formula above (and that the two agreed completely). Note that in this case r is the correlation, s_y is the sample standard deviation of the y data and s_x is the sample standard deviation of the x data.

4 Estimating the parameters of the regression model if the basis is orthogonal

4.1 Theory

If all the basis functions $g_0(\mathbf{x}), \ldots, g_p(\mathbf{x})$ are orthogonal, then we can use the following formula to estimate the regression parameters $\beta_0, \ldots, \beta_p$

$$\hat{\beta}_j = \frac{\sum_{i=1}^n g_j(\mathbf{x}_i) y_i}{\sum_{i=1}^n g_j^2(\mathbf{x}_i)} \text{ for } j = 0, \ldots, p$$

4.2 Some examples

Suppose that we have used our data and via the methods discussed in section 2 have shown that the basis 1, x, $x^2 - 5$ is orthogonal for our data. Then in this case when we fit the regression model $\mu(x) = \beta_0 + \beta_1 x + \beta_2(x^2 - 5)$ we get the following estimates parameter estimates:

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n (1) y_i}{\sum_{i=1}^n (1)^2} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n \left(x_i^2\right)^2}$$

Consider a different dataset, where we show that 1, x_1, x_2 and x_1x_2 is an orthogonal basis using the methods of section 2. Then when it comes to fit the regression model $\mu(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$ we would get the following parameter estimates

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n (1) y_i}{\sum_{i=1}^n (1)^2} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_{i1} y_i}{\sum_{i=1}^{n} x_{i1}^2}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_{i2} y_i}{\sum_{i=1}^{n} x_{i2}^2}$$

$$\hat{\beta}_3 = \frac{\sum_{i=1}^{n} x_{i1} x_{i2} y_i}{\sum_{i=1}^{n} \left(x_{i1} x_{i2}\right)^2}$$