Economics 522A, Spring 2007

Lecture Note 7: Normal Linear Regression Model and OLS

Conditionally Normal Model

We want to extend our previous approach to the case where there are more than one "x" variable. In practice, these variables will be both continuous and discrete, and it is convenient to avoid having to fully specify their distributions if we can avoid it. So we will focus on conditional models for y given x.

Let

$$x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix}$$

be a $k \times 1$ vector of explanatory variables (also called "regressors" or "covariates" or "independent variables").[1]

Usually, the first element of x_i will be a constant: $x_{i1} = 1$.

Conditional normal model:

$$y_i | x_1, \ldots, x_n \sim N(x_i'\beta, \sigma^2),$$

and y_i independent of y_j conditional on $x_1, \ldots, x_n$, for $i \neq j$.

Let

$$y := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X := \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}.$$

Then we can write the model equivalently as:

$$y|X \sim N_n\left(X\beta, \sigma^2 I_n\right),$$

where

$$X\beta = \begin{bmatrix} x_1'\beta \\ \vdots \\ x_n'\beta \end{bmatrix}, \quad \sigma^2 I_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & \\ \vdots & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix}.$$

This expresses the distribution of the entire vector y as multivariate normal, conditional on X.

[1] The term "independent" variables is an unfortunate one, and doesn't relate to statistical independence in any meaningful way. But sometimes people use the term, so it's good to be aware of it. The variable y_i is also sometimes called the "dependent" variable.

Conditional Likelihood: since each y_i is conditionally normal and independent:

$$\begin{aligned} f(y_1, \ldots, y_n \mid x_1, \ldots, x_n; \beta, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\right). \end{aligned}$$

Alternatively, we could use the multivariate normal representation to write:

$$\begin{aligned} f(y_1, \ldots, y_n \mid x_1, \ldots, x_n; \beta, \sigma^2) &= f(y \mid X; \beta, \sigma^2) \\ &= \det(2\pi\sigma^2 I_n)^{-1/2} \exp\left(-\frac{1}{2}(y - X\beta)'[\sigma^2 I_n]^{-1}(y - X\beta)\right). \end{aligned}$$

Note that

$$\det(2\pi\sigma^2 I_n)^{-1/2} = \left((2\pi\sigma^2)^n\right)^{-1/2} = (2\pi\sigma^2)^{-n/2},$$

and

$$\begin{aligned} -\frac{1}{2}(y - X\beta)'[\sigma^2 I_n]^{-1}(y - X\beta) &= -\frac{1}{2}(y - X\beta)'\sigma^{-2} I_n (y - X\beta) \\ &= -\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta). \end{aligned}$$

So

$$\begin{aligned} f(y|X; \beta, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2\right). \end{aligned}$$

Detour: some linear algebra

It's useful to think of $X\beta$ is a slightly different way. Instead of focusing on rows of X, which correspond to "observations," or "individuals," we can think of X as being a collection of column vectors:

$$X = [\xi_1, \xi_2, \ldots, \xi_k],$$

where

$$\xi_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}, \quad \xi_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix},$$

and so on. So each ξ_j is the $n \times 1$ vector of all observations on the jth variable.

Viewed this way,

$$X\beta = \beta_1\xi_1 + \beta_2\xi_2 + \cdots \beta_k\xi_k,$$

so $X\beta$ is a linear combination of the column vectors $\xi_1, \ldots, \xi_k$.

We usually want to set up the variables so that the $\xi_1, \ldots, \xi_k$ are linearly independent, in other words that X has full column rank. Otherwise, there is some redundancy in $\xi_1, \ldots, \xi_k$.

A useful result:

Lemma: *rank*$(X) = k$ *if and only if* $X'X$ *is nonsingular.*

Proof: if rank$(X) < k$, then there is an $\alpha \in \mathbb{R}^k$ with $\alpha \neq 0$, such that $X\alpha = 0$. But then $X'X\alpha = 0$, so that $X'X$ is singular.

Conversely, suppose that $X\alpha \neq 0$ for all $\alpha \neq 0$. Then $X'X\alpha \neq 0$, and $X'X$ is nonsingular. $\square$

Notice that

$$X'X = \sum_{i=1}^{n} x_i x_i',$$

and is square $(k \times k)$. So if $X'X$ is nonsingular, then it is invertible, a fact that will come in handy.

Ordinary Least Squares

Let us assume that X has full column rank.

We will choose $\hat{\beta}$ to solve

$$\min_{\beta} (y - X\beta)'(y - X\beta).$$

We can also rewrite the objective function in various forms:

$$(y - X\beta)'(y - X\beta) = \sum_{i=1}^{n} (y_i - x_i'\beta)^2 = \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2.$$

The first order conditions can be written as:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) \\
0 &= \sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) \\
&\vdots \\
0 &= \sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})
\end{aligned}$$

Or more compactly:

$$0 = \sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}),$$

where 0 is now a $k \times 1$ vector of zeroes.

Or even more compactly:

<u>OLS Normal Equation:</u> $\quad 0 = X'(y - X\hat{\beta}).$

So the solution should have:

$$X'y = X'X\hat{\beta},$$

and if X has full column rank, then $X'X$ is nonsingular and invertible, so we can invert $X'X$ to get:

$$\hat{\beta} = (X'X)^{-1}X'y.$$

To estimate σ^2, we could form:

$$e_i := y_i - x_i'\hat{\beta}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

Then a natural estimator is

$$\hat{\sigma}^2 = \frac{1}{n}e'e = \frac{1}{n}\sum_{i=1}^{n} e_i^2.$$

An alternative estimator, often used, is

$$s^2 = \frac{1}{n-k}e'e.$$

Maximum Conditional Likelihood

We can write the log conditional likelihood as:

$$\log\left[(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_i(y_i - x_i'\beta)^2\right)\right] = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i(y_i - x_i'\beta)^2.$$

By the same argument as in the previous LN, the MLE $\hat{\beta}$ solves the OLS normal equations, so it is equal to the OLS coefficients, and

$$\hat{\sigma}^2 = \frac{1}{n}e'e.$$

Some Properties of OLS

Since $\hat{\beta} = (X'X)^{-1}X'y$, the OLS coefficients are linear in y.

Under our assumption that $y|X \sim N(X\beta, \sigma^2 I_n)$,

$$E[y|X] = X\beta,$$

so that

$$\begin{aligned}
E[\hat{\beta}|X] &= E[(X'X)^{-1}X'y|X] \\
&= (X'X)^{-1}X'E[y|X] \\
&= (X'X)^{-1}X'X\beta \\
&= \beta.
\end{aligned}$$

Also, $V[y|X] = \sigma^2 I_n$, so

$$\begin{aligned}
V[\hat{\beta}|X] &= V[(X'X)^{-1}X'y|X] \\
&= (X'X)^{-1}X'V[y|X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'[\sigma^2 I_n]X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned}$$

(Note that $X'X$ is symmetric, so $(X'X)^{-1}$ is symmetric.)

In addition, since $\hat{\beta}$ is a linear function of y, and y is conditionally multivariate normal,

$$\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1}).$$