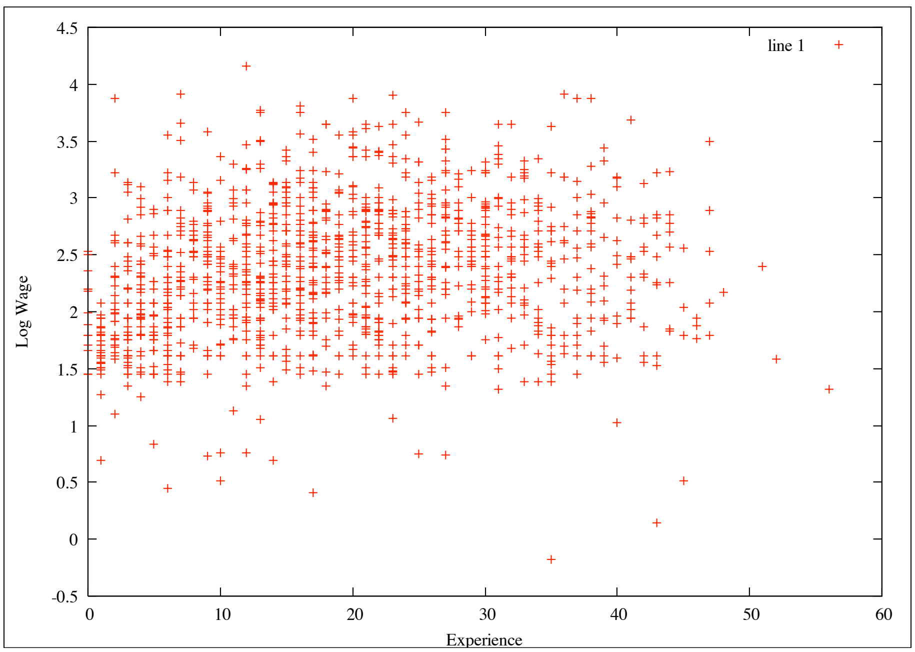
Economics 522A, Spring 2007

Lecture Note 6: Multivariate and Conditional Normal Model

1 Bivariate Normal Model

Suppose we want to develop a stastistical model for the CPS data examined in the previous lecture note. To simplify things, let us suppose that we first start by modeling the joint distribution of (y_i, x_i), where y_i is the log of wage, and x_i is experience. Here is a scatterplot:



Taking the log of wage has made the wage distribution a little more symmetric, and apart from the low experience observations, the data points look roughly elliptical. So, although the plot doesn't perfectly justify it, we might start by assuming that (y_i, x_i) are jointly multivariate normal:

$$z_i := \begin{pmatrix} y_i \\ x_i \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N(\mu, \Sigma), \quad i = 1, \ldots, n. \tag{1}$$

Here n is the sample size (in the CPS data, $n = 1289$.) The mean vector μ and variance matrix Σ can be written as

$$\mu = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{yy} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{xx} \end{pmatrix},$$

where μ_y is the marginal mean of y_i, σ_{yy} is the marginal variance of y_i, σ_{xy} is the covariance between y_i and x_i, etc.

Please remind yourself of the properties of the multivariate normal model, given in LN7 Addendum from Econ 520, and Ruud Ch. 10.5.1.

The density of z_i is:

$$f(z_i;\mu,\Sigma) = \det(2\pi\Sigma)^{-1/2}\exp\left(-\frac{1}{2}(z_i-\mu)'\Sigma^{-1}(z_i-\mu)\right).$$

So the joint likelihood is:

$$\begin{aligned}\mathcal{L}(\mu,\Sigma) &= f(z_1,\ldots,z_n;\mu,\Sigma)\\ &= \prod_{i=1}^{n} f(z_i;\mu,\Sigma)\\ &= \det(2\pi\Sigma)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(z_i-\mu)'\Sigma^{-1}(z_i-\mu)\right).\end{aligned}$$

The MLE can be solved analytically to get:

$$\hat{\mu} = \bar{z} = \frac{1}{n}\sum_i z_i = \begin{pmatrix} \frac{1}{n}\sum_i y_i \\ \frac{1}{n}\sum_i x_i \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{x} \end{pmatrix};$$

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{n}\sum_i (z_i-\hat{\mu})(z_i-\hat{\mu})'\\ &= \begin{pmatrix} \frac{1}{n}\sum_i (y_i-\bar{y})^2 & \frac{1}{n}\sum_i (y_i-\bar{y})(x_i-\bar{x}) \\ \frac{1}{n}\sum_i (y_i-\bar{y})(x_i-\bar{x}) & \frac{1}{n}\sum_i (x_i-\bar{x})^2 \end{pmatrix}\end{aligned}$$

This is also the method of moments estimator, since we are basically just replacing population expectations with sample averages. Using the CPS data we get:

$$\hat{\mu} = \begin{pmatrix} 2.34 \\ 18.79 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} 0.34 & 1.32 \\ 1.32 & 135.92 \end{pmatrix}.$$

2 Conditional Mean of y

Having estimated the joint distribution of (x_i, y_i), we might want to focus on certain aspects of that joint distribution, such as the conditional mean of y_i given x_i. Using standard results for the multivariate normal distribution:

$$y_i|x_i \sim N\left(\mu_y + \frac{\sigma_{xy}}{\sigma_{xx}}(x_i-\mu_x), \sigma_{yy} - (\sigma_{xy})^2/\sigma_{xx}\right).$$

Let

$$\begin{aligned}\beta_1 &= \mu_y - \frac{\sigma_{xy}}{\sigma_{xx}}\mu_x\\ \beta_2 &= \frac{\sigma_{xy}}{\sigma_{xx}}\\ \sigma^2 &= \sigma_{yy} - (\sigma_{xy})^2/\sigma_{xx}.\end{aligned}$$

Then we can write

$$y_i|x_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2).$$

We see that the conditional mean of y_i given x_i is linear in x_i:

$$E[y_i|x_i] = \beta_1 + \beta_2 x_i,$$

and the conditional variance is a constant that does not depend on the value of x_i:

$$Var[y_i|x_i] = \sigma^2.$$

Let's focus on the conditional mean parameters β_1 and β_2.

We can plug in the ML estimates to get corresponding estimates for the β parameters:

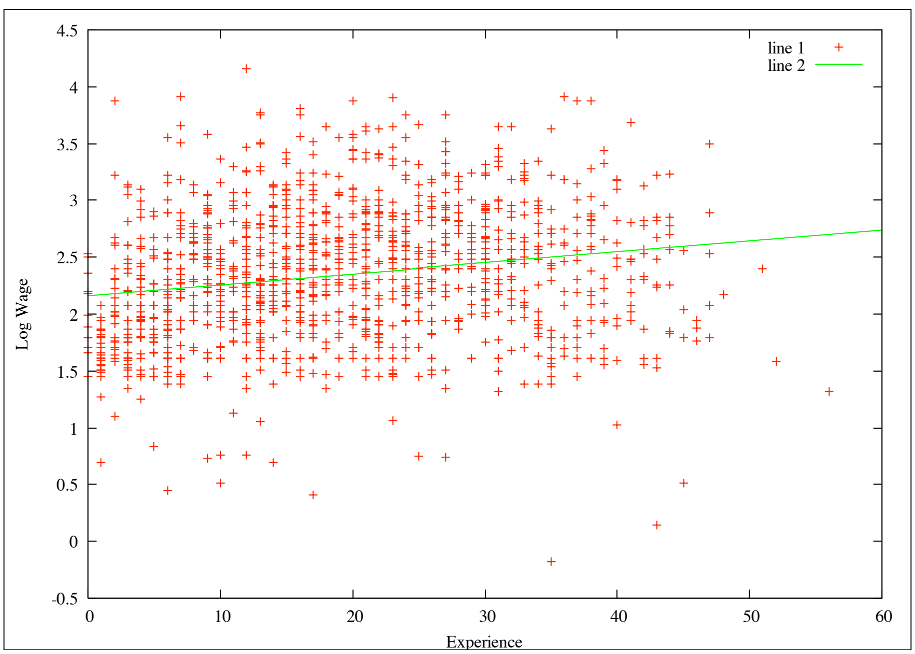
$$\hat{\beta}_2 = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_{xx}} = \frac{\frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_i (x_i - \bar{x})^2}.$$

$$\hat{\beta}_1 = \hat{\mu}_y - \hat{\beta}_2 \hat{\mu}_x = \bar{y} - \hat{\beta}_2 \bar{x}.$$

In our data:

$$\hat{\beta}_1 = 2.16, \quad \hat{\beta}_2 = 0.0097.$$

Here is the plot:



3 OLS

Now, consider the OLS coefficients. They solve the following problem:

$$\min_{\beta_1, \beta_2} \sum_i (y_i - \beta_1 - \beta_2 x_i)^2.$$

The first order conditions for a minimum are:

$$\begin{aligned} 2\sum_i (y_i - \beta_1 - \beta_2 x_i) &= 0; \\ 2\sum_i (y_i - \beta_1 - \beta_2 x_i)\cdot x_i &= 0. \end{aligned}$$

So the OLS coefficients $\hat{\beta}_1, \hat{\beta}_2$ satisfy what are sometimes called the "OLS Normal Equations":

$$\sum_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0; \tag{2}$$

$$\sum_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)\cdot x_i = 0. \tag{3}$$

The first equation can be rearranged to get:

$$\sum_i y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum_i x_i = 0$$

$$\Rightarrow \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}.$$

Plug this into the second normal equation:

$$\sum_i y_i x_i - (\bar{y} - \hat{\beta}_2 \bar{x}) \sum_i x_i - \hat{\beta}_2 \sum_i x_i^2 = 0.$$

$$\Rightarrow \sum_i y_i x_i - \sum_i \bar{y} x_i = \hat{\beta}_2 \left[\sum_i x_i^2 - \bar{x} \sum_i x_i \right].$$

$$\Rightarrow \frac{1}{n} \sum_i x_i (y_i - \bar{y}) = \hat{\beta}_2 \left[\frac{1}{n} \sum_i x_i^2 - (\bar{x})^2 \right].$$

$$\Rightarrow \hat{\beta}_2 = \frac{\frac{1}{n}\sum_i x_i (y_i - \bar{y})}{\frac{1}{n}\sum_i x_i^2 - (\bar{x})^2}.$$

A bit of algebra shows that this is equal to:

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_i (x_i - \bar{x})^2}.$$

So the OLS coefficients are identical to the MLE estimates.

4 Conditional Modeling

Recall that we started by assuming joint normality for (y_i, x_i). A nice feature of the multivariate normal distribution is that the marginal distribution of x_i is normal, and the conditional distribution of $y_i|x_i$ is normal. We focused on the parameters of the conditional distribution of $y_i|x_i$, and showed that MLE gave the same result as OLS.

Suppose we *only* make the assumption that y_i is conditionally normally distributed:

$$y_i|x_i \sim N(\beta_1 + \beta_2 x_i, \sigma^2). \tag{4}$$

We allow the distribution of x_i to be arbitrary—it could be nonnormal, discrete, even degenerate. In order to be precise about some of our later arguments, we will assume that the above conditional distribution holds conditional on *all* the x_is:

$$y_i|x_1, \ldots, x_n \sim N(\beta_1 + \beta_2 x_i, \sigma^2),$$

and that conditional on all the x_i, the y_i are independent.[1] Then we can write the joint conditional density of the ys given the xs as:

$$\begin{aligned} f(y_1, \ldots, y_n | x_1, \ldots, x_n; \beta_1, \beta_2, \sigma^2) &= \prod_{i=1}^{n} f(y_i | x_i; \beta_1, \beta_2, \sigma^2) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_i (y_i - \beta_1 - \beta_2 x_i)^2 \right). \end{aligned}$$

In *conditional maximum likelihood*, we treat this conditional density as the (conditional) likelihood, and maximize it with respect to the parameters:

$$\max_{\beta_1, \beta_2, \sigma^2} (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_i (y_i - \beta_1 - \beta_2 x_i)^2 \right).$$

Equivalently, maximize the log of the conditional likelihood:

$$\max_{\beta_1, \beta_2, \sigma^2} -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (y_i - \beta_1 - \beta_2 x_i)^2.$$

First order conditions:

$$\begin{aligned} -\frac{1}{2\sigma^2}(-2) \sum_i (y_i - \beta_1 - \beta_2 x_i) &= 0 \\ -\frac{1}{2\sigma^2}(-2) \sum_i (y_i - \beta_1 - \beta_2 x_i) \cdot x_i &= 0 \\ -\frac{n}{2} \cdot \frac{1}{2\pi\sigma^2}(2\pi) - \frac{1}{2}\left[-\frac{1}{\sigma^4} \right] \sum_i (y_i - \beta_1 - \beta_2 x_i)^2 &= 0 \end{aligned}$$

Simplify a bit to get:

$$\begin{aligned} \sum_i (y_i - \hat\beta_1 - \hat\beta_2 x_i) &= 0 \\ \sum_i (y_i - \hat\beta_1 - \hat\beta_2 x_i) x_i &= 0 \\ -\frac{n}{2\hat\sigma^2} + \frac{1}{2\hat\sigma^4} \sum_i (y_i - \hat\beta_1 - \hat\beta_2)^2 &= 0 \end{aligned}$$

[1] Suppose we assume that the (y_i, x_i) are IID, and that the conditional model in (4) holds. Then these further assumptions will hold.

Notice that the first two equations above are exactly the same as the normal equations in the OLS problem. The third equation above can be rearranged to get:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2.$$

So the solution for $\hat{\beta}_1$ and $\hat{\beta}_2$ are the same as in OLS, and to get the estimate for $\hat{\sigma}^2$ we can form the "OLS residuals"

$$e_i := y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$$

and then calculate

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i e_i^2.$$

(This last expression should seem sensible! It is like a sample variance.)

5 Some Properties of the OLS coefficients

Since our model is conditional on the x_is, let's see if we can derive some simple properties of the OLS coefficients.

First, consider the slope coefficient $\hat{\beta}_2$.

$$\begin{aligned} E[\hat{\beta}_2 | x_1, \ldots, x_n] &= E\left[\frac{\frac{1}{n}\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n}\sum_i (x_i - \bar{x})^2} \mid x_1, \ldots, x_n\right] \\ &= \frac{E\left[\sum_i (y_i - \bar{y})(x_i - \bar{x}) | x_1, \ldots, x_n\right]}{\sum_i (x_i - \bar{x})^2}. \end{aligned}$$

Useful trick: define

$$\epsilon_i := y_i - E[y_i | x_i] = y_i - \beta_1 - \beta_2 x_i.$$

So

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i,$$

and

$$\epsilon_i | x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

Note: ϵ_i is *NOT* the same as e_i, which is the residual using the OLS coefficients. Then

$$\begin{aligned} (y_i - \bar{y}) &= \beta_1 + \beta_2 x_i + \epsilon_i - \frac{1}{n}\sum_i (\beta_1 + \beta_2 x_i + \epsilon_i) \\ &= \beta_2 (x_i - \bar{x}) + \epsilon_i + \frac{1}{n}\sum_i \epsilon_i \end{aligned}$$

So

$$\sum_i (y_i - \bar{y})(x_i - \bar{x}) = \beta_2 \sum_i (x_i - \bar{x})^2 + \sum_i (\epsilon_i - \bar{\epsilon})(x_i - \bar{x}).$$

And

$$
\begin{aligned}
E\left[\sum_i (y_i-\bar{y})(x_i-\bar{x})|x_1,\ldots,x_n\right] &= \beta_2\sum_i (x_i-\bar{x})^2 + E\left[\sum_i (\epsilon_i-\bar{\epsilon})(x_i-\bar{x}) \mid x_1,\ldots,x_n\right] \\
&= \beta_2\sum_i (x_i-\bar{x})^2.
\end{aligned}
$$

Therefore,

$$E[\hat{\beta}_2|x_1,\ldots,x_n] = \beta_2.$$

So $\hat{\beta}_2$ is conditionally unbiased. Also, by the law of iterated expectations,

$$E[\hat{\beta}_2] = E\left[E[\hat{\beta}_2|x_1,\ldots,x_n]\right] = \beta_2.$$

By similar arguments, we can show that

$$E[\hat{\beta}_1|x_1,\ldots,x_n] = \beta_1.$$

6 Conditional vs. Joint Modeling

We are going to work with conditional models for a little while, so it is worth stopping to think about the general relationship between, say, unconditional MLE and conditional MLE.

Return to the joint normal model given in (1). We decomposed the model into a marginal model for x_i:

$$x_i \overset{\text{i.i.d.}}{\sim} N(\mu_x, \sigma_{xx}),$$

and a conditional model for y_i given x_i:

$$y_i|x_i \sim N(\beta_1+\beta_2 x_i, \sigma^2).$$

Note that there is a 1-1 mapping between the original parameters $(\mu_x, \mu_y, \sigma_{xx}, \sigma xy, \sigma_{yy})$ and the parameters $(\mu_x, \sigma_{xx}, \beta_1, \beta_2, \sigma^2)$. Under the reparametrizion, we have a set of parameters related to the marginal distribution of x_i: $\theta_1 = (\mu_x, \sigma_{xx})$, and a set of parameters for the conditional distribution: $\theta_2 = (\beta_1, \beta_2, \sigma^2)$.

So, generalizing a bit, we have a joint model and a marginal-conditional decomposition:

$$f(x_i, y_i; \theta_1, \theta_2) = f(x_i;\theta_1) f(y_i|x;\theta_2).$$

The joint likelihood can be written

$$
\begin{aligned}
f(x_1, y_1, \ldots, x_n, y_n; \theta_1, \theta_2) &= \prod_{i=1}^n f(x_i, y_i; \theta_1, \theta_2) \\
&= \prod_{i=1}^n f(x_i;\theta_1) f(y_i|x_i;\theta_2) \\
&= \prod_{i=1}^n f(x_i;\theta_1) \times \prod_{i=1}^n f(y_i|x_i;\theta_2).
\end{aligned}
$$

The joint MLE solves:

$$\max_{(\theta_1,\theta_2)\in\Theta} f(x_1, y_1, \ldots, x_n, y_n; \theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta_1) \times \prod_{i=1}^{n} f(y_i|x_i; \theta_2),$$

where Θ is the joint parameter space for (θ_1, θ_2). The conditional MLE solves:

$$\max_{\theta_2\in\Theta_2} \prod_{i=1}^{n} f(y_i|x_i; \theta_2),$$

where Θ_2 is the parameter space for θ_2.

If θ_1 only enters the marginal density of x_i, and θ_2 only enters the conditional density of $y_i|x_i$, and the joint parameter space is a Cartesian product:

$$\Theta = \Theta_1 \times \Theta_2,$$

then conditional MLE will give the same result for $\hat{\theta}_2$ as unconditional MLE.