What's Next

Multiple regression

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, i = 1, \ldots, n$$
$$\varepsilon_i \sim iidN(0, \sigma^2)$$

$$y = X\beta + \varepsilon$$

$$y = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}}_{n\times 1}, X = \underbrace{\begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nk} \end{bmatrix}}_{n\times k}, \beta = \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{k\times 1}, \varepsilon = \underbrace{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{n\times 1}$$

$X\beta$ is the mean

$$y = X\beta + \varepsilon$$
$$\varepsilon_i \sim iidN(0, \sigma^2)$$
$$y \sim N(X\beta, \sigma^2 I)$$

$$\mathcal{L}(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

Sum of squared errors

$$\sum \varepsilon_i^2 = \varepsilon'\varepsilon = \sum (y_i - X_i\beta)^2$$
$$= (y - X\beta)'(y - X\beta)$$

Derivative of inner product

$$f(b) = a'b = a_1b_1 + a_2b_2 + \cdots + a_kb_k$$
$$\frac{\partial f(b)}{\partial b_1} = a_1$$
$$\frac{\partial f(b)}{\partial b_2} = a_2$$
$$\frac{\partial f(b)}{\partial b} = \frac{\partial a'b}{\partial b} = \frac{\partial b'a}{\partial b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = a$$

Derivative of quadratic form

$f(b) = b'Ab$, where A is symmetric, example

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$
$$a_{11}b_1^2 + 2a_{12}b_1b_2 + a_{22}b_2^2$$
$$\frac{\partial b'Ab}{\partial b_1} = 2a_{11}b_1 + 2a_{12}b_2$$
$$\frac{\partial b'Ab}{\partial b_2} = 2a_{12}b_1 + 2a_{22}b_2$$
$$\frac{\partial b'Ab}{\partial b} = 2Ab$$

MLE

$$\mathcal{L}(\beta, \sigma^2)$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

$$\frac{\partial (y - X\beta)'I(y - X\beta)}{\partial \beta} = \frac{\partial (y - X\beta)}{\partial \beta} 2I(y - X\beta)$$
$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 = \frac{1}{\sigma^2}(-X')(y - X\beta)$$
$$\hat{\beta}_{mle} = (X'X)^{-1}X'y$$

MLE

$$\mathcal{L}(\beta, \sigma^2)$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0 = -\frac{n}{2\sigma^2} + \frac{1}{2{\sigma^2}^2}(y - X\beta)'(y - X\beta)$$
$$\hat{\sigma}^2_{mle} = \frac{\left(y - X\hat{\beta}_{mle}\right)'\left(y - X\hat{\beta}_{mle}\right)}{n}$$

Information matrix

$$\frac{\partial \mathcal{L}}{\partial \beta} = -\frac{1}{\sigma^2}(-X')(y - X\beta)$$
$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = -\frac{1}{\sigma^2}(-X')(-X) = -\frac{1}{\sigma^2}X'X$$
$$\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^{2^2}}(X')(y - X\beta)$$

Nb

$$-\mathrm{E}\left(\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \sigma^2}\right) = 0$$

Information matrix

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^{2^2}}(y - X\beta)'(y - X\beta)$$
$$\frac{\partial^2 \mathcal{L}}{(\partial \sigma^2)^2} = \frac{n}{2\sigma^{2^2}} - \frac{1}{\sigma^{2^3}}(y - X\beta)'(y - X\beta)$$
$$\mathrm{E}\left(\frac{\partial^2 \mathcal{L}}{(\partial \sigma^2)^2}\right) = \frac{n}{2\sigma^{2^2}} - \frac{n\sigma^2}{\sigma^{2^3}} = -\frac{n}{2\sigma^{2^2}}$$

Information matrix

$$I = \frac{1}{\sigma^2}\begin{bmatrix} X'X & 0 \\ 0 & \frac{n}{2\sigma^2} \end{bmatrix}$$
$$V(\beta, \sigma^2) = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^{2^2}}{n} \end{bmatrix}$$

General error structure

$$y = X\beta + \varepsilon$$
$$\varepsilon \sim N(0, \Sigma)$$
$$\mathrm{E}(\varepsilon\varepsilon') = \Sigma$$

$$y \sim N(X\beta, \Sigma)$$

Method of moments estimator

$$\frac{1}{n}X'\varepsilon = 0, \frac{1}{n}X'(y - X\beta) = 0$$
$$\beta_{mm} = (X'X)^{-1}X'y$$
$$= (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$$

$\mathrm{E}[(\beta_{mm}$

Maximum-likelihood

$$f_n(y; \mu, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right)$$

$$\mathcal{L} = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(y - X\beta)'\Sigma^{-1}(y - X\beta)$$
$$\frac{\partial\mathcal{L}}{\partial\beta} = -X'\Sigma^{-1}(y - X\beta) = 0$$
$$\beta_{mle} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

Σ not known

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon$$
$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$
$$\hat{\beta} \sim (\beta, \mathrm{E}[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}])$$
$$\hat{\beta} \sim (\beta, \mathrm{E}[(X'X)^{-1}X'\Sigma X(X'X)^{-1}])$$

Sandwich estimator

$$\hat{\beta} \sim (\beta, \mathrm{E}[(X'X)^{-1}X'\Sigma X(X'X)^{-1}])$$

If Σ is diagonal (heteroskedasticity), then use $\mathrm{var}(\hat{\beta})$

$$\approx \frac{1}{n}\left(\frac{1}{n}\sum X_i X_i'\right)^{-1}\left(\frac{1}{n}\sum e_i^2 X_i X_i'\right)\left(\frac{1}{n}\sum X_i X_i'\right)$$

If Σ is block diagonal, then the middle part gets replaced by clusters rather than individual squared errors to get clustered standard errors.

Projection matrix

$$P_X \equiv X(X'X)^{-1}X'$$

Note that P_X is symmetric and idempotent

$$[X(X'X)^{-1}X']'$$
$$X''[X(X'X)^{-1}]'$$
$$X''[(X'X)^{-1\prime}X']$$
$$X''[(X'X)'^{-1}X']$$
$$X''[(X'X'')^{-1}X']$$
$$X(X'X)^{-1}X'$$

Projection matrix

$$P_X \equiv X(X'X)^{-1}X'$$

Note that P_X is symmetric and idempotent

$$P_X P_X = X(X'X)^{-1}X'X(X'X)^{-1}X'$$
$$= X(X'X)^{-1}\cancel{X'X(X'X)^{-1}}X'$$

Projection matrix

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

Annihilation matrix

$$M_X \equiv I - X(X'X)^{-1}X' = I - P_X$$
$$P_X M_X = P_X - P_X P_X = 0$$
$$X'M_X = X'[I - X(X'X)^{-1}X']$$
$$X' - X'X(X'X)^{-1}X' = 0$$

$$rank(P_X) = k$$
$$rank(M_X) = n - k$$

residuals

$$e = y - \hat{y} = y - P_X y = [I - P_X]y$$

$$X'e = X'[I - P_X]y = 0$$

residuals

$$e = M_X y = M_X(X\beta + \varepsilon) = M_X \varepsilon$$

$$e'e = \varepsilon' M_X M_X \varepsilon = \varepsilon' M_X \varepsilon$$

$$e \sim N(0, M_X \sigma^2 I M_X) = N(0, \sigma^2 M_X)$$

Sum squared residuals

Theorem: If $x \sim N(0, V)$ then $x'Ax \sim \chi^2_{rank(A)}$ iff AV is idempotent.

$$\varepsilon/\sigma \sim N(0, I)$$

$$e'e = \varepsilon' M_X \varepsilon$$

$$\frac{e'e}{\sigma^2} = \frac{\varepsilon'}{\sigma} M_X \frac{\varepsilon}{\sigma}$$

$M_X I$ is idempotent

$$e'e \sim \chi^2_{n-k}$$

Identification

Definition: A parameter θ for a family of distributions is *identifiable* if distinct values of θ correspond to distinct values of the likelihood, i.e.,

$$\theta \neq \theta' \Rightarrow L(x|\theta) \neq L(x|\theta')$$

Identification example

$$y_i = \beta x_i + \varepsilon_i, i = 1, \ldots, n$$
$$\varepsilon_i = \lambda x_i + v_i, v \sim iidN(0, \sigma_v^2)$$

We can substitute in and write

$$y_i = (\beta + \lambda)x_i + v_i$$

The log likelihood is

$$\mathcal{L} = -\frac{n}{2}\log(2\pi\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum(y_i - (\beta + \lambda)x_i)^2$$

Identification example

The log likelihood is

$$\mathcal{L} = -\frac{n}{2}\log(2\pi\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum(y_i - (\beta + \lambda)x_i)^2$$

Identification example

The log likelihood is

$$\mathcal{L} = -\frac{n}{2}\log(2\pi\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum(y_i - (\beta + \lambda)x_i)^2$$

The first-order conditions are

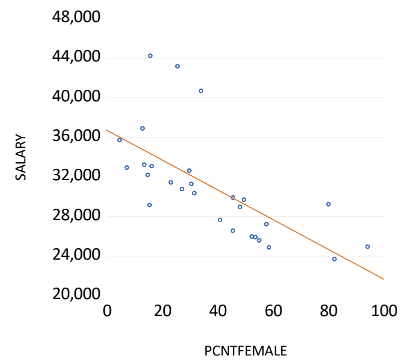
$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 = \frac{1}{\sigma_v^2}\sum(y_i - (\beta + \lambda)x_i)x_i$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 = \frac{1}{\sigma_v^2}\sum(y_i - (\beta + \lambda)x_i)x_i$$
$$\beta_{mle} + \lambda_{mle} = \frac{\sum y_i x_i}{\sum x_i^2}$$

OLS

$$y = \beta x + \varepsilon$$

$$\hat{\beta} = \frac{\sum yx}{\sum x^2} = \frac{\sum(\beta x + \varepsilon)x}{\sum x^2} = \beta + \frac{\sum \varepsilon x}{\sum x^2}$$
$$\text{plim}(\hat{\beta}) = \beta + \frac{\text{plim}\frac{1}{n}\sum \varepsilon x}{\text{plim}\frac{1}{n}\sum x^2}$$

	SALARY	PCNTFEMALE
Agriculture	36,879	12.9
Architecture	30,337	31.6
Art	27,198	57.6
Business	30,753	27.1
Chemistry	33,069	16.2
Dentistry	44,214	15.7
Drama	24,865	58.5
Economics	32,179	14.8
Education	28,952	48.1



Dependent Variable: SALARY
Method: Least Squares
Date: 12/03/17 Time: 13:02
Sample: 1 28
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36701.86	1478.344	24.82633	0.0000
PCNTFEMALE	-150.4170	33.27774	-4.520049	0.0001

R-squared	0.440027	Mean dependent var	30981.71
Adjusted R-squared	0.418490	S.D. dependent var	5302.816
S.E. of regression	4043.758	Akaike info criterion	19.51649
Sum squared resid	4.25E+08	Schwarz criterion	19.61164
Log likelihood	-271.2308	Hannan-Quinn criter.	19.54558
F-statistic	20.43084	Durbin-Watson stat	2.717705
Prob(F-statistic)	0.000119		

$$y = X\beta + Z\Gamma + \mathrm{v}$$
$$\hat{\beta} = \left(X'X\right)^{-1}X'y = \left(X'X\right)^{-1}X'[X\beta + Z\Gamma + \mathrm{v}]$$
$$\hat{\beta} = \beta + \left(X'X\right)^{-1}X'Z\Gamma + [\left(X'X\right)^{-1}X'\mathrm{v}]$$
$$\Lambda = \left(X'X\right)^{-1}X'Z$$

Bias

$$\hat{\beta} = \beta + \left(X'X\right)^{-1}X'Z\Gamma$$
$$\lambda = \frac{\sum zx}{\sum x^2}$$
$$\hat{\beta} = \beta + \lambda\gamma$$