Multiple Regression

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$i = 1:n$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & & x_{pn} \end{pmatrix}$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$\dim(X) = n \times (p+1)$

$\dim(\beta) = (p+1) \times 1$

$\dim(y) = n \times 1$

$\dim(\varepsilon) = n \times 1$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

$$y = X\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

1) LSE: Find $\hat{\beta}$ that minimizes

$$S(\beta) = \| y - X\beta \|^2 = (y - X\beta)'(y - X\beta)$$

2) MLE

$$y \sim N(X\beta, \sigma^2 I_n)$$

$$L(\beta,\sigma^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{ -\frac{(y_i-\beta_0-\beta_1 x_{1i}-\cdots-\beta_p x_{pi})^2}{2\sigma^2} \right\}$$

or, equivalently

$$L(\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{(y-X\beta)'(y-X\beta)}{2\sigma^2} \right\}$$

Maximizing $L(\beta,\sigma^2)$ is the same as maximizing

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

---

1) For LSE we find

$$\frac{\partial S(\beta)}{\partial \beta_0} = 0, \ldots, \frac{\partial S(\beta)}{\partial \beta_p} = 0$$

If X is a full rank matrix with $n > (p+1)$ and $\text{rank}(X) = p+1$ then we can show that the LSE is unique and given by

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

2) For MLE we also find

$$\frac{\partial \log L(\beta,\sigma^2)}{\partial \beta_0} = 0, \ldots, \frac{\partial \log L(\beta,\sigma^2)}{\partial \beta_p} = 0$$

If, say, σ^2 is known/fixed
The MLE with X a full rank matrix and $n > (p+1)$
is also
$$\hat{\beta} = (X'X)^{-1}X'y$$
But we also have σ^2
so we also find
$$\frac{\partial \operatorname{Log} L(\beta, \sigma^2)}{\partial \sigma^2} = 0$$
and this gives us
$$\hat{\sigma}^2_{MLE} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n} = $$
$$= \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} \quad \text{(not unbiased)}$$
so we use
$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-(p+1)} \quad \text{instead}$$

Properties of $\hat{\beta}$

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

$$\cdot\ Var(\hat{\beta}) = \underbrace{(X'X)^{-1}}X' \underbrace{Var(y)}_{\sigma^2 I_n} X(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

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

F-test for nested models

$$\begin{cases} M_1 : y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i \\ M_2 : y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \beta_{p+1} x_{(p+1)i} + \\ \qquad \cdots + \beta_{p+q} x_{(p+q)i} + \varepsilon_i \end{cases}$$

H_0: $\beta_{(p+1)} = \cdots = \beta_{(p+q)} = 0$

H_a: At least one of $\beta_{(p+1)}, \cdots, \beta_{(p+q)}$ is $\neq 0$

Fit M_1 and M_2 and get

$\underline{SSE_1}$ and $\underline{SSE_2}$

X_1 $\quad X_2$

$\hat{\beta}^{(1)}$ $\quad \hat{\beta}^{(2)}$

$$SSE_1 = (y - X_1 \hat{\beta}^{(1)})'(y - X_1 \hat{\beta}^{(1)})$$

$$SSE_2 = (y - X_2 \hat{\beta}^{(2)})'(y - X_2 \hat{\beta}^{(2)})$$

Note that $SSE_1 > SSE_2$

$$F = \frac{(SSE_1 - SSE_2)/q}{SSE_2/(n-(p+q+1))}$$

Under H_0 $F \sim F$-distribution with q d.f. in the numerator and $n-(p+q+1)$ d.f. in the denominator

What about comparing models that are not nested?

- AIC : Akaike's information criterion

If you have a model M

$$AIC(M) = -2\log L(M) + 2\underline{p(M)}$$

$L(M)$: likelihood of model M evaluated at the MLE

$p(M)$: # of parameters in model M

• BIC : Bayesian information criterion

$$BIC(M) = -2 \log L(M) + \log(n) \cdot p(M)$$

If you have models $M_1, M_2, \ldots, M_K$ you can choose the one that minimizes AIC / BIC

Adjusted R^2

$$R^2_{adj} = 1 - \left[\frac{SSE}{n-(p+1)}\right] \Big/ \left[\frac{\Sigma(y_i - \bar{y})^2}{n-1}\right]$$

$$SST = \Sigma(y_i - \bar{y})^2$$