Test3

2023-04-11

(a).

Consider the usual linear model, where $y = X\beta + \varepsilon$. We now compare two regressions, which differ in how many variables are included in the matrix X. In the full (unrestricted) model p1 regressors are included. In the restricted model only a subset of $p_0 < p_1$ regressors are included. Show that the smallest model is preferred according to the AIC if

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p1-p0)}.$$

(a). Answer

we need this :

$$AIC(p) = log(s_p^2) + \frac{2p}{n}$$

then,

$$AIC_0 = log(s_0^2) + \frac{2p_0}{n}$$

$$AIC_1 = log(s_1^2) + \frac{2p_1}{n}$$

and we know :

$$p_0 < p_1$$

so, $\frac{2p_0}{n} < \frac{2p_1}{n}$ -> $AIC_0 < AIC_1$.

$$log(s_0^2) + \frac{2p_0}{n} < log(s_1^2) + \frac{2p_1}{n}$$

$$log(s_0^2) - log(s_1^2) < \frac{2p_1}{n} - \frac{2p_0}{n}$$

$$log(\frac{s_0^2}{s_1^2}) < \frac{2(p_1 - p_0)}{n}$$

then we can get this

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p1-p0)}.$$

(b).

Argue that for very large values of n the inequality of (a) is equal to the condition

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p1 - p0).$$

Use that $e^x \approx 1 + x$ for small values of x.

(b). answer

for very large values of n,

$$\frac{2}{n}(p1 - p0) \approx 0$$

then,

$$e^{\frac{2}{n}(p1-p0)} \approx 1 + \frac{2}{n}(p1 - p0)$$

we know this before,

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p1-p0)}.$$

$$e^{\frac{2}{n}(p1-p0)} \approx 1 + \frac{2}{n}(p1 - p0)$$

so, it will be,

$$\frac{s_0^2}{s_1^2} < 1 + \frac{2}{n}(p1 - p0)$$

$$\frac{s_0^2}{s_1^2} - 1 < \frac{2}{n}(p1 - p0)$$

we can get

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p1 - p0).$$

(c).

Show that for very large values of n the condition in (b) is approximately equal to

$$\frac{e_R' eR - e_U' e_U}{e_U' e_U} < \frac{2}{n}(p1 - p0).$$

where e_R is the vector of residuals for the restricted model with p_0 parameters and e_U the vector of residuals for the full unrestricted model with p_1 parameters.

(c). answer

we already know

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p1-p0)}$$

$$e^{\frac{2}{n}(p1-p0)} \approx 1 + \frac{2}{n}(p1 - p0)$$

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p1 - p0)$$

so, we can get

$$s_0^2 = e_R' eR$$

$$s_1^2 = e_U' e_U$$

and another soloution,

$s = \sqrt{\frac{e'e}{n-k}}$ is standard error of the regression

for very large values of n,

$$\frac{1}{n-k} = 0$$

so,

$$s = \sqrt{e'e}$$

then we can get

$$s^2 = e'e$$

$$s_0^2 = e_R' eR$$

$$s_1^2 = e_U' e_U$$

(d).

Finally, show that the inequality from (c) is approximately equivalent to an F-test with critical value 2, for large sample sizes.

(d). answer

we know F,

$(e'e = e_U' e_U)$

$$F = \frac{(e_R' eR - e_U' e_U)/g}{e_U' e_U/(n-k)}$$

n is a very large values, and for the critical value = 2 so,

then,

$F < 2$

$$\frac{(e_R' eR - e_U' e_U)/g}{e_U' e_U/(n-k)} < 2$$

$g = (p_1 - p_0)$

$$\frac{(e_R' eR - e_U' e_U)/(p_1 - p_0)}{e_U' e_U/(n-k)} < 2$$

$(n - k) \approx n$

$$\frac{(e_R' eR - e_U' e_U)/(p_1 - p_0)}{e_U' e_U/n} < 2$$

$$\frac{(e_R' eR - e_U' e_U)}{e_U' e_U}\frac{n}{(p_1 - p_0)} < 2$$

so, we can get

$$\frac{e_R' eR - e_U' e_U}{e_U' e_U} < \frac{2}{n}(p1 - p0).$$