Week 3 Test Exercise

(a)

AIC: $\log(s^2) + 2k/n$

Small model: $AIC_0 = \log(s_0^2) + 2p_0/n$

Big model: $AIC_1 = \log(s_1^2) + 2p_1/n$

AIC: Choose small model if

$$AIC_0 < AIC_1$$

$$\log(s_0^2) + 2p_0/n < \log(s_1^2) + 2p_1/n$$

$$\log(s_0^2/ s_1^2) < 2/n*(p_1 - p_0)$$

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

(b)

Given:

$$e^x \approx 1 + x \textit{ (if x is small)}$$

If n is very large, $(2/n)(p_1 - p_0)$ is small. Therefore

$$e^{2/n*(p1 - p0)} \approx 1 + 2/n*(p_1 - p_0) \textit{ (if n is very large)}$$

Substituting this expression into the right side of the result from part (a) yields

$$s_0^2/ s_1^2 < 1 + 2/n*(p_1 - p_0) \textit{ (for n very large)}$$

$$s_0^2/ s_1^2 - 1 < 2/n*(p_1 - p_0)$$

$$\mathbf{(s_0^2 - s_1^2)/ s_1^2 < 2/n*(p_1 - p_0)} \textit{ (for n very large)}$$

(c)

$$s^2 = 1/(n-1) \sum_{i=1}^{n} e_i^2 = 1/(n-1)*e'e$$

$$\Rightarrow s_0^2 = 1/(n-1)*e_R'e_R,\ s_1^2 = 1/(n-1)*e_U'e_U$$

Substituting theses expressions into the result from part (b) yields

$$(1/(n-1)*e_R'e_R - 1/(n-1)*e_U'e_U)/ (1/(n-1)*e_U'e_U) < 2/n*(p_1 - p_0)$$

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

(d)

F-test (formula from Lecture 2.4.2 slides):

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

where k is the number of explanatory factors in the unrestricted model, and g is the number of explanatory factors removed from the unrestricted model to create the restricted model.

Under this test, we believe there is significant evidence to suggest that $\beta \neq 0$ (so the unrestricted model is preferred) if $F > F_{critical}$. Therefore a larger model is preferred if $F > F_{critical}$, and we stay with (prefer) a smaller model if $F < F_{critical}$.

Let $F_{critical} = 2$. Then a smaller model is preferred if $F < 2$:

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

In this case, with p_1 factors in the unrestricted model and p_0 in the restricted model, we get

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

$$(e_R'e_R - e_U'e_U)/(e_U'e_U) < 2(p_1 - p_0)/(n - p_1)$$

If n is very large, $n - p_1 \approx n$. Then we get the desired result:

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

is our condition for preferring a restricted model when doing an F-test with $F_{critical} = 2$ (and when n is very large), just as it was our condition for preferring a restricted model when using the AIC (when n is very large).