STA 206 Homework 8

1. (a) True? — A "correct" model has unbiased estimates and expected value of residuals is 0. However, there may still be a lot of irreducible error/noise. We want to reduce MSE without overfitting as it reduces predictive capability

(b) False — Many nuisance X variables results in large variance, with low bias.

(c) False - R^2 is a poor model selection criterion as it does not decrease as more predictors are added on top of an existing model.

(d) True - all else held equal, a decrease in SSE_p will result in a lower C_p, AIC, and BIC (provided equivalent model complexity)

(e) True - This is true as its numerator value is the same while the denominator value is lower in $PRESS_p$, so $PRESS_p$ will always be greater.

(f) True — $(\log n)p > 2p$ for $\log n > 2$, so BIC penalizes more complex models for $n \geq 8$.

(g) False - The stepwise procedures are not guaranteed to find the best model and is only done because best subset selection is computationally expensive for large p and large n.