Kevin Su

# STA 206 Quiz 3

1. (a) False, $X_1, X_2$, and/or $X_3$ may be significant but with low t-value due to effects from multicollinearity. We should consider dropping a variable to see if it improves our selected, though unspecified model criterion (stepwise selection?)

(b) False, if there exists unusually high leverage points in one model but not another, $PRESS_p$ can be higher with smaller SSE. If the X variables are identical, then it is true as there should be no disparity in leverage.

(c) True. Adding X variables will not increase model bias, but it will increase variance.

2. (a) Model 1 seems to have larger variance than Model 2

(b) Model 1 seems to have larger bias than Model 2

(c) Model 1 seems to have larger MSEE

(d) Model 1 has smaller coefficient of multiple determination.

3. (a) Red square has the highest $X_1$ and $X_2$ and is thus the outlier in both $X_1$/$X_2$, but Blue has the more influential case as Red lies close to the fit.

(b) Blue triangle is the most influential and red square is the least influential. Blue has the highest residual and does not have as many observations in X to offset its Y observation. Red has many observations close in X and would have the lowest residual of the three. Green has high residual but many observations close in X.