6.

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

where $H = X(X'X)^{-1}X'$

$\text{Var}(\hat{Y}|X) = \text{Var}(X(X'X)^{-1}X'Y|X)$

$= \text{Var}(HY|X)$

$= H \cdot \text{Var}(Y|X)$ where $\text{Var}(Y|X) = \sigma^2 I$

$= H \cdot \sigma^2 I = H \cdot \sigma^2$

(2) According to the description of the question, one assumption the model based is that there should not involve any kind of relation between predictors. In other words, it should not has any kind of multicollinearity. However, predictors X_{10} and X_{11}, Market size and per-capita income are somehow related to each other, therefore it violates the assumption of non-Multicollinearity of the model

Another assumption is Auto correlation, that the observations of predictors over response variable is correlated to themselves over different time. In this model each predictor that measures the team quality could be related to themselves over time. That is auto correlation could threute the validity of the model.

(3)

(a) The adjusted R square is 0.7751 which is reasonable to say that the model is fine. However the model looks skewed and has several leverage point, so maybe show consider do a transformation on the data.

(b) the residuals are not evenly spread, which means it is not non-patterned. We should possibly use some transformation on the model.

(c) point 222, point 223 are bad leverage points. (not all leverage points bad. in the model)

(b) The model looks better than the previous one. It has a great adj-R square of 0.8578, non-patterned residuals, evenly distributed variance. The only problem is there are still several bad leverage points. But overall, it is a valid model

(e) The F-statistics shows it is even more statistically significant. So it is a valid strategy

(f) we can create Dummy variables for manufacturer names and add it to the model.

4

(a) adj R-square is really high (0.9363), it is a valid model.

(b) The curved pattern shows a violation of the constant variance Assumption. Also the residuals also not strictly follow the normal distribution. Certain Degree of transformation should add to the model for adjustment.

(c) Standard deviation is more like explain the spreadness of our data. Since we have the assumption of normality of residuals, standard deviation can used to examine the normality of data.

F value is more likely to decide the statistically significance of a variable to the model. when deciding between models, we want to make sure all of our variables are statistically significant.

r measures the strength and direction of a liner relationship between two variables. we can derive the coefficient to measure the proportion of variance of data that explained by the model from r.

ratio of # of observation to # of decipters provides insight of vality of model. it tells you whether the model is right or wrong