# Module 10: Model Diagnostics and Remedial Measures in MLR

Jeffrey Woo

MSDS, University of Virginia

# Measures of Influence

| | Formula | Influential if |
|---|---|---|
| Cook's D, $D_i$ | $\frac{(\hat{\beta}_{(i)}-\hat{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\beta}_{(i)}-\hat{\beta})}{pMS_{res}}$; or $\frac{r_i^2}{p}\frac{h_{ii}}{1-h_{ii}}$ | $> F_{0.5,p,n-p}$ |
| $DFBETAS_{j,i}$ | $\frac{\hat{\beta}_j-\hat{\beta}_{j(i)}}{\sqrt{S^2_{(i)}C_{jj}}}$ | magnitude $> 2/\sqrt{n}$ |
| $DFFITS_i$ | $\frac{\hat{y}_i-\hat{y}_{(i)}}{\sqrt{S^2_{(i)}h_{ii}}}$; or $(\frac{h_{ii}}{(1-h_{ii})})^{1/2}t_i$ | magnitude $> 2\sqrt{p/n}$ |

Observations that have high leverage, and unusual combination of predictors and response tend to be influential.

# Influential Observations

Applet

# What to do with Influential Observations

- Influential observations usually have something interesting about them that make them "stand out" from the other observations.
- Fit the model with and without the influential observations and see how the models answer our questions of interest.
- Occasionally an observation is influential due to an error in the data entry.
- Rarely do I advocate deleting an influential data point. These observations must addressed.

# What to do with Influential Observations

- If dropping the influential data point(s) doesn't change the results, ok to drop, but you must note that you dropped outliers.
- If dropping the influential data point(s) changes the results, fit model with and without outliers, and clearly note how the results have changed. Try to characterize the influential data point(s).

# What to do with Influential Observations

- If you have influential observations due to small number of observations with large predictor and/or response, a log transformation on the variable can pull in the large values.
- Consider subsetting your data and create separate models for each subset; or focus on a subset and make it clear your analysis is for a subset.
- Knowing your data and context can help a lot in these decisions.

# Model Building Process

See document on Collab.

# Where Are We Headed?

- Modules 11 & 12: Logistic Regression. Binary response variable.