Detecting Collinearity

1. Large F value, small t values. In other words, a small p value on the F statistic for the overall model and large p values on *all* the t values for the individual coefficients.

2. Large Variance Inflation Factors

Variance Inflation Factors Let $x_1, \ldots, x_K$ be the explanatory variables.

- Perform a regression with x_j as the response variable and all other variables as predictor variables.

- Obtain the R^2 value for each regression. Call it R_j^2.

- The variance inflation factor for x_j is

$$VIF_j = \frac{1}{1 - R_j^2}$$

Example For the case of 4 explanatory variables x_1, x_2, x_3, x_4, there are 4 variance inflation factors.

Response Variable	**Explanatory Variables**	R^2	VIF
x_1	x_2, x_3, x_4	R_1^2	$\frac{1}{1-R_1^2}$
x_2	x_1, x_3, x_4	R_2^2	$\frac{1}{1-R_2^2}$
x_3	x_1, x_2, x_4	R_3^2	$\frac{1}{1-R_3^2}$
x_1	x_1, x_2, x_3	R_4^2	$\frac{1}{1-R_4^2}$

Example Compute the variance inflation factors for the Cal Ripken Data.

- Run the regression of HITS vs ATBATS and HOMERUNS

- Run the regression of ATBATS vs HITS and HOMERUNS

- Run the regression of HOMERUNS vs HITS and ATBATS

Using SPSS to find VIFs Run the program **Analyze/Regression/Linear**, click on the **Statistics** button, and select the **Colinearity Diagnostics** checkbox.

Example Go back and compute the VIFs for the Cal Ripken data using SPSS

Using VIFs to determine multicollinearity

- Any VIF larger than 10 may indicate multicollinearity

- If the average of the VIFs is larger than 1, multicollinearity may be present. 5

- If the VIFs are less than $\frac{1}{1-R^2}$, where R^2 is for the overall model, multicollinearity is not an issue.