Variation Inflation Factor

Definition

The **Variance Inflation Factor**, or **VIF** is a measure of multicollinearity in a linear regression. It provides a measurement of how much the regression coefficient variance is increased due to collinearity with other model variables.

Equations

$$\hat{var}(\hat{\beta}_j = \frac{s^2}{(n-1)\hat{var}(X_j)} * \frac{1}{1-R_j^2}$$

Example VIF Output

```
library(car)
# use swiss dataset
data(swiss)
model1 <- lm(Fertility ~ Agriculture + Education + Examination +
                  Catholic, data=swiss)
model2 <- update(model1, Fertility ~ Agriculture + Education + Examination)
model3 <- update(model1, Fertility ~ Agriculture + Education)
vif(model1)
```

```
## Agriculture   Education Examination    Catholic
##    2.147410    2.689400    3.675372    1.856475
```

```
vif(model2)
```

```
## Agriculture   Education Examination
##    2.089159    2.156227    2.410537
```

```
vif(model3)
```

```
## Agriculture   Education
##    1.692016    1.692016
```

Interpretation and Use

VIF is a fairly easy statistic to interpret. A value of 1 indicates no correlation between the variable and the other regression variables. A value $1 < VIF < 5$ suggests moderate correlation and a value $VIF > 5$ suggests strong correlation in the variables.

VIF	Indication
1	no correlation
$1 < VIF < 5$	moderate correlation
$VIF > 5$	strong correlation

Further Avenues

When **VIF** is high in a model for regression coefficients, it may be necessarily to remove or otherwise limit some model parameters. VIF can help identify if nested models may be a better choice for finding relationships. Other statistics that evaluate the value of nested models like AIC, BIC, and PRESS may help to find a better model.