Stats 500: HW #4 Collinearity

Diana Liang

10/27/2019

**Full Model**

##   
## Call:  
## lm(formula = Employed ~ GNP.deflator + GNP + Unemployed + Armed.Forces +   
## Population + Year, data = longley)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.41011 -0.15767 -0.02816 0.10155 0.45539   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.482e+03 8.904e+02 -3.911 0.003560 \*\*   
## GNP.deflator 1.506e-02 8.492e-02 0.177 0.863141   
## GNP -3.582e-02 3.349e-02 -1.070 0.312681   
## Unemployed -2.020e-02 4.884e-03 -4.136 0.002535 \*\*   
## Armed.Forces -1.033e-02 2.143e-03 -4.822 0.000944 \*\*\*  
## Population -5.110e-02 2.261e-01 -0.226 0.826212   
## Year 1.829e+00 4.555e-01 4.016 0.003037 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3049 on 9 degrees of freedom  
## Multiple R-squared: 0.9955, Adjusted R-squared: 0.9925   
## F-statistic: 330.3 on 6 and 9 DF, p-value: 4.984e-10

**Condition Numbers (1)**

## [1] 1.000 17.855 25.153 60.785 1647.478 5751.216

Varation in predictors is 5000 times bigger between the largest and smallest variation. This means that some of the eigenvalues for the (X transpose)\*X are, relatively, very small and that some of the predictor variables are collinear.

**Correlation between Predictors (2)**

## GNP.deflator GNP Unemployed Armed.Forces Population Year  
## GNP.deflator 1.00 0.99 0.62 0.46 0.98 0.99  
## GNP 0.99 1.00 0.60 0.45 0.99 1.00  
## Unemployed 0.62 0.60 1.00 -0.18 0.69 0.67  
## Armed.Forces 0.46 0.45 -0.18 1.00 0.36 0.42  
## Population 0.98 0.99 0.69 0.36 1.00 0.99  
## Year 0.99 1.00 0.67 0.42 0.99 1.00

GNP.deflator, GNP, Population and Year predictor variables have a correlation of nearly 1, meaning that they are most likely collinear. These variables should be condensed down to one predictor variable.

**Variance Inflation Factors (3)**

## GNP.deflator GNP Unemployed Armed.Forces Population   
## 135.532 1788.513 33.619 3.589 399.151   
## Year   
## 758.981

GNP.deflator, GNP, Population and Year predictor variables had intensely greater variance inflation factors compared to Unemployed and Armed.Forces, meaning that those 4 variables are very highly correlated.

**Reduced Model**

##   
## Call:  
## lm(formula = Employed ~ GNP.deflator + Unemployed + Armed.Forces,   
## data = longley)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.81870 -0.46282 0.07278 0.15816 1.18427   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30.393412 1.864285 16.303 1.49e-09 \*\*\*  
## GNP.deflator 0.398076 0.030995 12.843 2.26e-08 \*\*\*  
## Unemployed -0.010725 0.003220 -3.330 0.0060 \*\*   
## Armed.Forces -0.008165 0.003829 -2.132 0.0543 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6776 on 12 degrees of freedom  
## Multiple R-squared: 0.9702, Adjusted R-squared: 0.9628   
## F-statistic: 130.3 on 3 and 12 DF, p-value: 2.02e-09

The full model was reduced to only include the predictor variables GNP.deflator, Unemployed, and Armed.Forces. The latter 2 did not show strong evidence of collinearity so they were kept in the reduced model. GNP, GNP.deflator, Population, and Year were heavily correlated, so only GNP.deflator was chosen to remain in the reduced model as the easiest to interpret of the four variables. The full model showed that the predictor variables Unemployed, Armed.Forces, and Year were significant, with a significance level of 0.05, in modeling the response variable Employed. The reduced model now only has predictor variables GNP.deflator and Unemployed as significant in modeling Employed.

**Reduced Model Condition Numbers**

## [1] 1.000 5.134 43.276

The reduced model shows relatively low eigenvalues for the remaining predictor variables, so there is little evidence of collinearity between them.

**Reduced Model Variance Inflation Factors**

## GNP.deflator Unemployed Armed.Forces   
## 3.655 2.959 2.320

The variance inflation factors are very similar and quite, which confirms there is little correlation between the remaining predictor variables.