PLSC 503 – Spring 2020
Cases and Variables

February 20, 2020

Under the Hood of $\mathbf{X}$

OLS (and regression methods more generally) requires:

- $\mathbf{X}$ is full column rank.
- $N > K$.
- "Sufficient" variability in $\mathbf{X}$.

“Perfect” Multicollinearity

Formally: There cannot be any set of λs such that:

$$\lambda_0 \mathbf{1} + \lambda_1 \mathbf{X}_1 + ... + \lambda_K \mathbf{X}_K = \mathbf{0}$$

If there was, it would imply

$$\mathbf{X}_j = \frac{-\lambda_0}{\lambda_j}\mathbf{1} + \frac{-\lambda_1}{\lambda_j}\mathbf{X}_1 + ... + \frac{-\lambda_K}{\lambda_j}\mathbf{X}_K$$

which means

$$\begin{aligned}
Y &= \beta_0 \mathbf{1} + \beta_1 \mathbf{X}_1 + ... + \beta_j \mathbf{X}_j + ... + \beta_K \mathbf{X}_K + \mathbf{u} \\
&= \beta_0 \mathbf{1} + \beta_1 \mathbf{X}_1 + ... + \beta_j \left(\frac{-\lambda_0}{\lambda_j}\mathbf{1} + \frac{-\lambda_1}{\lambda_j}\mathbf{X}_1 + ... + \frac{-\lambda_K}{\lambda_j}\mathbf{X}_K \right) + ... + \beta_K \mathbf{X}_K + \mathbf{u} \\
&= \left[\beta_0 + \beta_j \left(\frac{-\lambda_0}{\lambda_j} \right) \right] \mathbf{1} + \left[\beta_1 + \beta_j \left(\frac{-\lambda_1}{\lambda_j} \right) \right] \mathbf{X}_1 + ... + \left[\beta_K + \beta_j \left(\frac{-\lambda_K}{\lambda_j} \right) \right] \mathbf{X}_K + \mathbf{u} \\
&= \left(\beta_0 + \frac{\gamma_0}{\lambda_j} \right) \mathbf{1} + \left(\beta_1 + \frac{\gamma_1}{\lambda_j} \right) \mathbf{X}_1 + ... + \left(\beta_K + \frac{\gamma_K}{\lambda_j} \right) \mathbf{X}_K + \mathbf{u}
\end{aligned}$$

```
> Africa$newgdp<-(Africa$gdppppd-mean(Africa$gdppppd))*1000

> fit<-with(Africa, lm(adrate~gdppppd+newgdp+healthexp+subsaharan+
+                         muslperc+literacy))
> summary(fit)

Call:
lm(formula = adrate ~ gdppppd + newgdp + healthexp + subsaharan +
    muslperc + literacy)

Residuals:
    Min      1Q  Median      3Q     Max
-15.291  -4.329  -1.412   2.723  20.682

Coefficients: (1 not defined because of singularities)
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           -7.78020   10.33872  -0.753   0.4565
gdppppd                0.36142    0.58214   0.621   0.5385
newgdp                      NA         NA      NA       NA
healthexp              1.87001    0.75667   2.471   0.0182 *
subsaharanSub-Saharan  3.64354    4.54163   0.802   0.4275
muslperc              -0.07908    0.05967  -1.325   0.1932
literacy               0.12445    0.09867   1.261   0.2151
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.665 on 37 degrees of freedom
Multiple R-squared:  0.4782,Adjusted R-squared:  0.4077
F-statistic: 6.782 on 5 and 37 DF,  p-value: 0.0001407
```

So...

- Perfect multicollinearity is terrible, but

- Perfect multicollinearity not a problem at all.

$N > K$...

Statistically,

- we lack sufficient degrees of freedom to identify $\hat{\boldsymbol{\beta}}$.
- $\hat{\boldsymbol{\beta}}$ is "overdetermined."

Conceptually:

- Variables $>$ Cases means
- ...no unique conclusion about explanatory / causal factors.

$N = K$ in Practice

```
> smallAfrica<-subset(Africa,subsaharan=="Not Sub-Saharan")
> fit2<-with(smallAfrica,lm(adrate~gdppppd+healthexp+muslperc+
+                             literacy+war))
> summary(fit2)

Call:
lm(formula = adrate ~ gdppppd + healthexp + muslperc + literacy +
    war)

Residuals:
ALL 6 residuals are 0: no residual degrees of freedom!

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.12430         NA      NA       NA
gdppppd     -0.97906         NA      NA       NA
healthexp   -0.45166         NA      NA       NA
muslperc     0.01413         NA      NA       NA
literacy     0.09512         NA      NA       NA
war         -0.96429         NA      NA       NA

Residual standard error: NaN on 0 degrees of freedom
Multiple R-squared:      1,Adjusted R-squared:    NaN
F-statistic:   NaN on 5 and 0 DF,  p-value: NA
```

High (Non-Perfect) Multicollinearity

Recall that

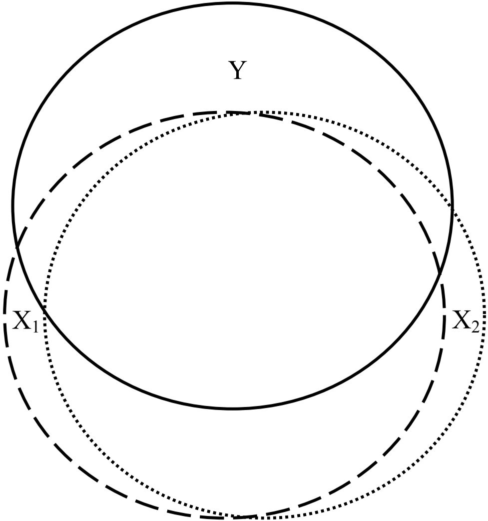
$$\widehat{\text{Var}(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

We can write the kth diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ as:

$$\frac{1}{(\mathbf{X}_k'\mathbf{X}_k)(1-\hat{R}_k^2)}$$

where $\hat{R}_k^2$ is the R^2 from the regression of $\mathbf{X}_k$ on all the other variables in $\mathbf{X}$.

The Obligatory Venn Diagram



High (Non-Perfect) Multicollinearity

Things to understand:

1. Multicollinearity is a *sample problem*.

2. Multicollinearity is a matter of *degree*.

Near-Perfect Collinearity: An Example

$$\text{HIV}_i = \beta_0 + \beta_1(\text{Civil War}_i) + \beta_2(\text{Intensity}_i) + u_i$$

```
> with(Africa, table(internalwar,intensity))

           intensity
internalwar  0  1  2  3
          0 30  0  0  0
          1  0  6  2  5
```

Table: Three Models

	Dependent variable:		
	adrate		
	(1)	(2)	(3)
internalwar	−4.459		−2.849
	(3.274)		(6.682)
intensity		−1.955	−0.837
		(1.481)	(3.018)
Constant	10.713***	10.502***	10.713***
	(1.800)	(1.734)	(1.821)
Observations	43	43	43
R^2	0.043	0.041	0.045
Adjusted R^2	0.020	0.017	−0.003
Residual Std. Error	9.860 (df = 41)	9.873 (df = 41)	9.973 (df = 40)
F Statistic	1.855 (df = 1; 41)	1.743 (df = 1; 41)	0.945 (df = 2; 40)
Note:			*p<0.1; **p<0.05; ***p<0.01

(Near-Perfect) Multicollinearity: Detection

1. *High R^2, but nonsignificant coefficients.*
2. *High pairwise correlations among independent variables.*
3. *High partial correlations among the* $\mathbf{X}$*s.*
4. *VIF and Tolerance.*

VIF / Tolerance

If $\hat{R}_k^2 = 0$, then

$$\widehat{\text{Var}(\hat{\beta}_k)} = \frac{\hat{\sigma}^2}{\mathbf{X}_k'\mathbf{X}_k};$$

So:

$$\text{VIF}_k = \frac{1}{1 - \hat{R}_k^2}$$

$$\text{Tolerance} = \frac{1}{\text{VIF}_k}$$

Rule of Thumb: VIF > 10 is a problem...

What To Do?

Don't:

- **Blindly drop covariates!!!**
- Restrict βs...

Do:

- **Add data**.
- **Transform the covariates**
    - Data reduction
    - First differences
    - Orthogonalize

What To Do? Shrinkage Methods

OLS is:

$$\begin{aligned}
\text{MSE} &= \mathsf{E}\{[\mathbf{Y} - \mathsf{E}(\mathbf{Y})]^2\} \\
&= \mathsf{E}[(Y_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^2] \\
&= [Y_i - \mathsf{E}(\mathbf{X}_i\hat{\boldsymbol{\beta}})]^2 + \{\mathsf{E}[(\mathbf{X}_i\hat{\boldsymbol{\beta}}) - \mathsf{E}(\mathbf{X}_i\hat{\boldsymbol{\beta}})]\}^2 \\
&= (\text{Bias})^2 + \text{Variance}
\end{aligned}$$

"Ridge regression":

$$\hat{\boldsymbol{\beta}}^R = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{Y}$$

- <u>Biases</u> $\hat{\boldsymbol{\beta}}$, but
- Increases the (perceived) independent variability in **X**
- Yields:

$$\widehat{\text{Var}(\hat{\boldsymbol{\beta}}_\ell^R)} = \frac{\hat{\sigma}^2}{(\mathbf{X}_\ell'\mathbf{X}_\ell + \lambda)(1 - R_\ell^2)}$$

What To Do? Lasso, Etc.

"LASSO" = "Least Absolute Shrinkage and Selection Operator."

- Formally:

$$\min_{\boldsymbol{\beta}} \left\{ \frac{1}{N} \sum_{i=1}^{N} (Y_i - \mathbf{X}_i \boldsymbol{\beta})^2 \right\} \text{ subject to } \sum_{j=1}^{p} |\beta_j| \leq t.$$

- Combines variable selection and shrinkage...
- Think ridge regression, but with some $\hat{\beta}$s set to zero
- Reduces overfitting + makes the model more interpretable

Example: Impeachment

```
> summary(impeachment)
     name              state             district     votesum    
 Length:433         Length:433         Min.   : 1   Min.   :0.00  
 Class :character   Class :character   1st Qu.: 3   1st Qu.:0.00  
 Mode  :character   Mode  :character   Median : 6   Median :2.00  
                                       Mean   :10   Mean   :1.85  
                                       3rd Qu.:13   3rd Qu.:4.00  
                                       Max.   :52   Max.   :4.00  
    pctbl96         unionpct         clint96       GOPmember          ADA98      
 Min.   : 0.0   Min.   :0.0257   Min.   :26.0   Min.   :0.000   Min.   :  0.0  
 1st Qu.: 2.0   1st Qu.:0.0930   1st Qu.:42.0   1st Qu.:0.000   1st Qu.:  5.0  
 Median : 5.4   Median :0.1690   Median :48.0   Median :1.000   Median : 30.0  
 Mean   :11.9   Mean   :0.1636   Mean   :50.3   Mean   :0.527   Mean   : 46.3  
 3rd Qu.:14.0   3rd Qu.:0.2150   3rd Qu.:57.0   3rd Qu.:1.000   3rd Qu.: 90.0  
 Max.   :74.0   Max.   :0.3733   Max.   :94.0   Max.   :1.000   Max.   :100.0  
```

Regression!

```
> fit<-lm(votesum~ADA98+GOPmember+clint96+pctbl96+unionpct)
> summary(fit)

Call:
lm(formula = votesum ~ ADA98 + GOPmember + clint96 + pctbl96 +
    unionpct)

Residuals:
   Min     1Q Median     3Q    Max
-3.271 -0.259  0.133  0.337  2.731

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.51785    0.23246   10.83   <2e-16 ***
ADA98       -0.02144    0.00238   -9.00   <2e-16 ***
GOPmember    1.59981    0.18043    8.87   <2e-16 ***
clint96     -0.00935    0.00433   -2.16    0.031 *
pctbl96      0.00347    0.00270    1.29    0.199
unionpct    -0.52544    0.48065   -1.09    0.275
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.629 on 427 degrees of freedom
Multiple R-Squared: 0.883,Adjusted R-squared: 0.882
F-statistic:  647 on 5 and 427 DF,  p-value: <2e-16
```

Assessing Collinearity

```
> idata=impeachment[c(-1,-2)]
> cor(idata)
          district  votesum  pctbl96 unionpct clint96 GOPmember    ADA98
district   1.00000 -0.03496 -0.06759  0.09155  0.1044  -0.02881  0.04988
votesum   -0.03496  1.00000 -0.28765 -0.26199 -0.6408   0.91977 -0.92795
pctbl96   -0.06759 -0.28765  1.00000 -0.09394  0.6165  -0.30911  0.30288
unionpct   0.09155 -0.26199 -0.09394  1.00000  0.3331  -0.19406  0.27563
clint96    0.10437 -0.64084  0.61651  0.33305  1.0000  -0.61196  0.67033
GOPmember -0.02881  0.91977 -0.30911 -0.19406 -0.6120   1.00000 -0.93918
ADA98      0.04988 -0.92795  0.30288  0.27563  0.6703  -0.93918  1.00000
```

```
> vif(fit)
    ADA98 GOPmember   clint96   pctbl96  unionpct
   10.292     8.878     3.313     1.998     1.371
```

Regression, again!

```
> fit2<-lm(votesum~ADA98+clint96+pctbl96+unionpct)
> summary(fit2)

Call:
lm(formula = votesum ~ ADA98 + clint96 + pctbl96 + unionpct)

Residuals:
   Min     1Q Median     3Q    Max
-3.300 -0.300  0.179  0.383  2.913

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.02775    0.17198   23.42   <2e-16 ***
ADA98       -0.04052    0.00111  -36.60   <2e-16 ***
clint96     -0.00658    0.00469   -1.40     0.16
pctbl96      0.00165    0.00293    0.56     0.57
unionpct     0.08300    0.51706    0.16     0.87
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.684 on 428 degrees of freedom
Multiple R-Squared: 0.862,Adjusted R-squared: 0.861
F-statistic:  667 on 4 and 428 DF,  p-value: <2e-16


> vif(fit2)
   ADA98  clint96  pctbl96 unionpct
   1.883    3.296    1.986    1.343
```

```
> ridge.vote<-lm.ridge(votesum~ADA98+GOPmember+clint96+pctbl96+unionpct,
  lambda=seq(0,5000,10))
> select(ridge.vote)
modified HKB estimator is 0.8365
modified L-W estimator is 0.4018
smallest value of GCV  at 10
```

Values of $\hat{\beta}_k^R$, by λ