PLSC 503 – Spring 2021
Variances and Collinearity

February 17, 2021

Variances: Why We Care

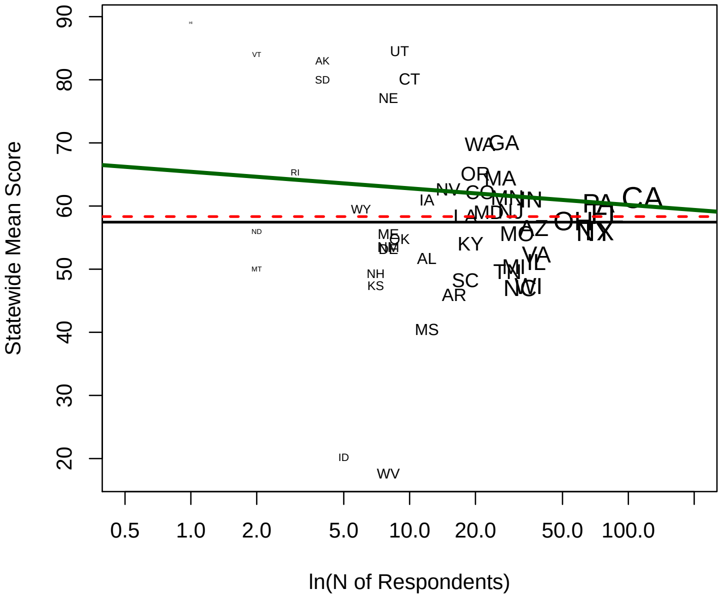
2016 ANES pilot study "feeling thermometer" toward gays and lesbians ($N = 1200$):

```
> summary(ANES$ftgay)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   40.50   54.00   57.45   88.50  100.00       1
```

Suppose we wanted to create aggregate measures, by state ($N = 51$). We would get:

```
> summary(StateFT)
    State               Nresp           meantherm
 Length:50          Min.   :  1.00   Min.   :17.62
 Class :character   1st Qu.:  8.00   1st Qu.:51.33
 Mode  :character   Median : 18.00   Median :57.11
                    Mean   : 24.00   Mean   :58.33
                    3rd Qu.: 30.75   3rd Qu.:62.55
                    Max.   :116.00   Max.   :89.00
```

Variances: Why We Care



Variances: A Generalization

Start with:

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + u_i$$

with:

$$\mathrm{Var}(u_i) = \sigma^2/w_i$$

with w_{iu} <u>known</u>.

Weighted Least Squares

WLS now minimizes:

$$\text{RSS} = \sum_{i=1}^{N} w_i (Y_i - \mathbf{X}_i \boldsymbol{\beta}).$$

which gives:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{WLS} &= [\mathbf{X}'(\sigma^2 \boldsymbol{\Omega})^{-1} \mathbf{X}]^{-1} \mathbf{X}'(\sigma^2 \boldsymbol{\Omega})^{-1} \mathbf{Y} \\
&= [\mathbf{X}' \mathbf{W}^{-1} \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}^{-1} \mathbf{Y}
\end{aligned}$$

where:

$$\mathbf{W} = \begin{bmatrix} \frac{\sigma^2}{w_1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma^2}{w_2} & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\sigma^2}{w_N} \end{bmatrix}$$

Getting to Know WLS

The variance-covariance matrix is:

$$\begin{aligned}\text{Var}(\hat{\beta}_{WLS}) &= \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &\equiv (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\end{aligned}$$

A common case is:

$$\text{Var}(u_i) = \sigma^2 \frac{1}{N_i}$$

where N_i is the number of observations upon which (aggregate) observation i is based.

"Robust" Variance Estimators

Recall that, if $\sigma_i^2 \neq \sigma_j^2 \forall i \neq j$,

$$
\begin{aligned}
\text{Var}(\beta_{\text{Het.}}) &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{Q}\,(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

where $\mathbf{Q} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})$ and $\mathbf{W} = \sigma^2\Omega$.

We can rewrite $\mathbf{Q}$ as

$$
\begin{aligned}
\mathbf{Q} &= \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X}) \\
&= \sum_{i=1}^{N} \sigma_i^2 \mathbf{X}_i \mathbf{X}_i'
\end{aligned}
$$

Huber's Insight

Estimate $\hat{\mathbf{Q}}$ as:

$$\widehat{\mathbf{Q}} = \sum_{i=1}^{N} \hat{u}_i^2 \mathbf{X}_i \mathbf{X}_i'$$

Yields:

$$\begin{aligned}\widehat{\mathrm{Var}(\beta)}_{\mathrm{Robust}} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\widehat{\mathbf{Q}}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\left[\mathbf{X}'\left(\sum_{i=1}^{N} \hat{u}_i^2 \mathbf{X}_i \mathbf{X}_i'\right)^{-1}\mathbf{X}\right](\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Practical Things

"Robust" VCV estimates:

- are heteroscedasticity-consistent, but
- are biased in small samples, and
- are less efficient than "naive" estimates when $\text{Var}(u) = \sigma^2 \mathbf{I}$.

“Clustering”

Huber / White ????????? WLS / GLS

I know very little about my error variances...

I know a great deal about my error variances...

"Clustering"

A common case:

$$Y_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + u_{ij}$$

with

$$\sigma^2_{ij} = \sigma^2_{ik}.$$

"Robust, clustered" estimator:

$$\widehat{\text{Var}(\boldsymbol{\beta})}_{\text{Clustered}} = (\mathbf{X}'\mathbf{X})^{-1} \left\{ \mathbf{X}' \left[\sum_{i=1}^{N} \left(\sum_{j=1}^{n_j} \hat{u}^2_{ij} \mathbf{X}_{ij} \mathbf{X}'_{ij} \right) \right]^{-1} \mathbf{X} \right\} (\mathbf{X}'\mathbf{X})^{-1}$$

Robust / Clustered SEs: A Simulation

```
url_robust <- "https://raw.githubusercontent.com/IsidoreBeautrelet/economictheoryblog/master/robust_summary.R"
eval(parse(text = getURL(url_robust, ssl.verifypeer = FALSE)),
     envir=.GlobalEnv)

> set.seed(7222009)
> X <- rnorm(10)
> Y <- 1 + X + rnorm(10)
> df10 <- data.frame(ID=seq(1:10),X=X,Y=Y)
>
> fit10 <- lm(Y~X,data=df10)
> summary(fit10)

Residuals:
     Min       1Q   Median       3Q      Max
-1.12328 -0.65321 -0.05073  0.43937  1.81661

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8438     0.3020   2.794   0.0234 *
X             0.3834     0.3938   0.974   0.3588
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9313 on 8 degrees of freedom
Multiple R-squared:  0.1059,Adjusted R-squared:  -0.005832
F-statistic: 0.9478 on 1 and 8 DF,  p-value: 0.3588

> rob10 <- vcovHC(fit10,type="HC1")
> sqrt(diag(rob10))
(Intercept)           X
  0.2932735   0.2859552
```

Robust / Clustered SEs: A Simulation (continued)

```
> # "Clone" each observation 100 times
>
> df1K <- df10[rep(seq_len(nrow(df10)), each=100),]
> df1K <- pdata.frame(df1K, index="ID")
>
> fit1K <- lm(Y~X,data=df1K)
> summary(fit1K)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.84383    0.02704   31.20   <2e-16 ***
X            0.38341    0.03526   10.87   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.8338 on 998 degrees of freedom
Multiple R-squared:  0.1059,Adjusted R-squared:  0.105
F-statistic: 118.2 on 1 and 998 DF,  p-value: < 2.2e-16

> summary(fit1K, cluster="ID")

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8438     0.2766   3.050  0.00235 **
X             0.3834     0.2697   1.421  0.15551
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.8338 on 998 degrees of freedom
Multiple R-squared:  0.1059,Adjusted R-squared:  0.105
F-statistic:  2.02 on 1 and 9 DF,  p-value: 0.1889
```

“Real-Data” Example

```
> summary(Justices)

     name               score              civrts          econs      
 Length:31          Min.   :-1.0000   Min.   :19.80   Min.   :34.60  
 Class :character   1st Qu.:-0.4700   1st Qu.:35.90   1st Qu.:43.85  
 Mode  :character   Median : 0.3300   Median :43.70   Median :50.20  
                    Mean   : 0.1210   Mean   :51.42   Mean   :55.75  
                    3rd Qu.: 0.6250   3rd Qu.:75.55   3rd Qu.:66.65  
                    Max.   : 1.0000   Max.   :88.90   Max.   :81.70  

  Neditorials         eratio            scoresq          lnNedit      
 Min.   : 2.000   Min.   : 0.5000   Min.   :0.0000   Min.   :0.6931  
 1st Qu.: 4.000   1st Qu.: 0.7083   1st Qu.:0.1936   1st Qu.:1.3863  
 Median : 6.000   Median : 1.0000   Median :0.2500   Median :1.7918  
 Mean   : 8.742   Mean   : 2.0242   Mean   :0.4599   Mean   :1.8442  
 3rd Qu.:11.500   3rd Qu.: 2.5000   3rd Qu.:0.8281   3rd Qu.:2.4414  
 Max.   :47.000   Max.   :11.7500   Max.   :1.0000   Max.   :3.8501  
```

OLS...

```
> OLSfit<-with(Justices, lm(civrts~score))
> summary(OLSfit)

Call:
lm(formula = civrts ~ score)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   48.810      2.852  17.113  < 2e-16 ***
score         21.544      4.206   5.122 1.81e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.63 on 29 degrees of freedom
Multiple R-squared: 0.475,Adjusted R-squared: 0.4569
F-statistic: 26.24 on 1 and 29 DF,  p-value: 1.806e-05
```

WLS, Weighting by ln(N of Editorials)

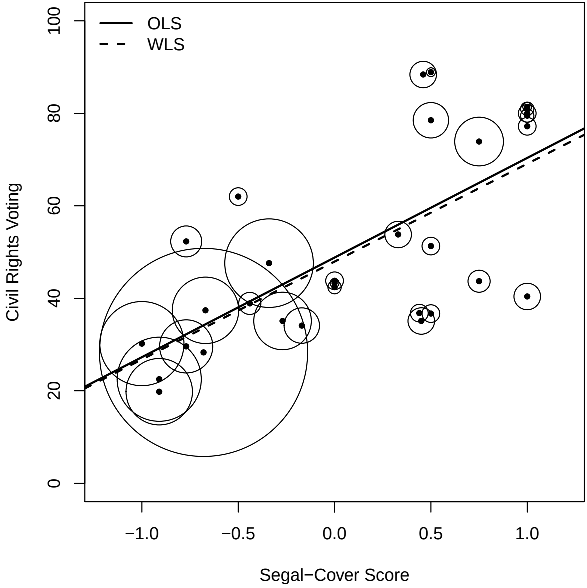
```
> WLSfit<-with(Justices, lm(civrts~score,weights=lnNedit))
> summary(WLSfit)

Call:
lm(formula = civrts ~ score, weights = lnNedit)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   47.936      2.600  18.439  < 2e-16 ***
score         21.158      3.797   5.572 5.18e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 19.59 on 29 degrees of freedom
Multiple R-squared: 0.5171,Adjusted R-squared: 0.5004
F-statistic: 31.05 on 1 and 29 DF,  p-value: 5.179e-06
```

Figure: Plot of `civrts` Against `score`, Weighted by `Neditorials`



“Robust” Standard Errors

```
> library(car)
> hccm(OLSfit, type="hc1")
            (Intercept)     score
(Intercept)    6.963921  2.929622
score          2.929622 13.931212


> library(rms)
> OLSfit2<-ols(civrts~score, x=TRUE, y=TRUE)
> RobSEs<-robcov(OLSfit2)
> RobSEs

Linear Regression Model

ols(formula = civrts ~ score, x = TRUE, y = TRUE)

         n Model L.R.       d.f.          R2      Sigma
        31      19.97          1       0.475      15.63

Residuals:
    Min      1Q  Median      3Q     Max
-29.954  -8.088  -2.120   9.396  29.680

Coefficients:
          Value Std. Error      t  Pr(>|t|)
Intercept 48.81      2.552 19.123 0.000e+00
score     21.54      3.610  5.968 1.739e-06

Residual standard error: 15.63 on 29 degrees of freedom
Adjusted R-Squared: 0.4569
```

Cases, Variables, and Collinearity

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

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

In Practice

```
> 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
```

- 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

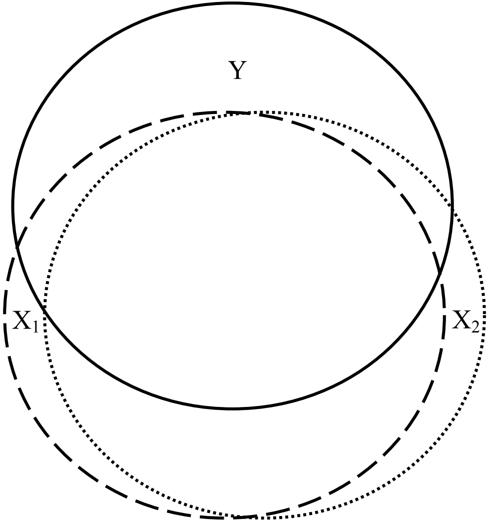
$$\widehat{\text{Var}(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

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

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

So:

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

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

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:

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

"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<-with(impeachment,
         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 Regression...

```
> 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