PLSC 502 – Autumn 2016
Bivariate Regression I

November 10, 2016

Random Variables

$$Y_i = \mu + u_i$$

$$\mu_i = \beta_0 + \beta_1 X_i$$

so that we get

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Goals:

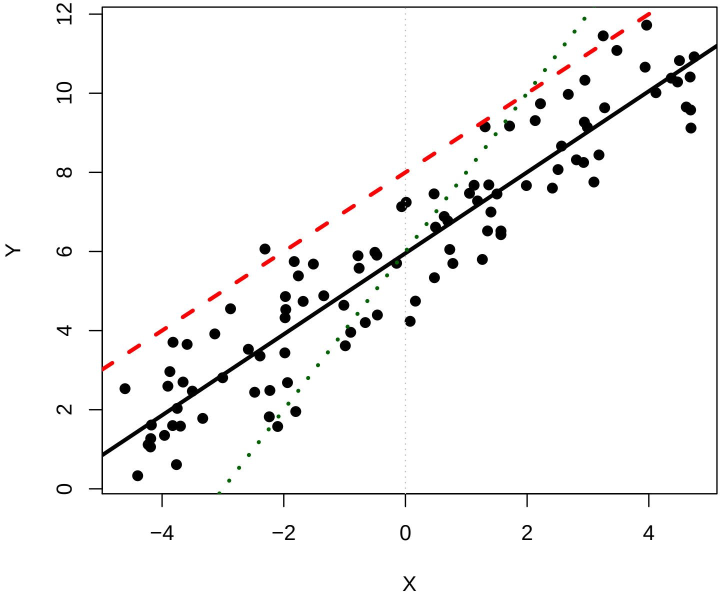
- *Point estimates* of β_0 and β_1
- Estimates of variability

Estimating β_0 and β_1

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

"Residuals":

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \end{aligned}$$



Y
X
0
2
4
6
8
10
12
−4
−2
0
2
4

"Loss Function"

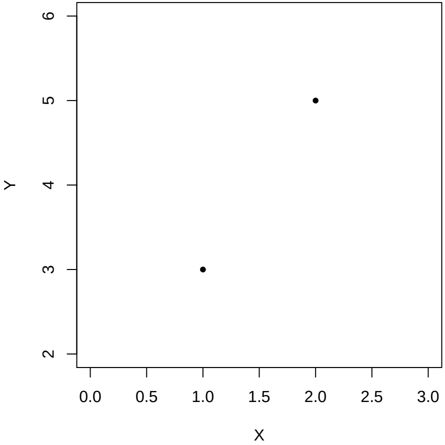
Key Idea: Select $\hat{\beta}_0$ and $\hat{\beta}_1$ to make the $\hat{u}_i$s as small as possible.

Possibilities:

- Pick $\hat{\beta}_0$ and $\hat{\beta}_1$ so as to minimize $\sum_{i=1}^{N} \hat{u}_i$
- Pick $\hat{\beta}_0$ and $\hat{\beta}_1$ so as to minimize $\sum_{i=1}^{N} |\hat{u}_i|$ ("MAD")
- Pick $\hat{\beta}_0$ and $\hat{\beta}_1$ so as to minimize $\sum_{i=1}^{N} \hat{u}_i^2$ ("least squares")

The Simplest Regression In Human History

```
> d
  x y
1 1 3
2 2 5
```



World's Simplest Regression

Recall:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

So, for $i = 1$

$$\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1(1)$$

and for $i = 2$

$$\hat{Y}_2 = \hat{\beta}_0 + \hat{\beta}_1(2)$$

Means:

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= 3 - \hat{\beta}_0 + \hat{\beta}_1(1) \text{ for } i = 1, \text{and} \\ &= 5 - \hat{\beta}_0 + \hat{\beta}_1(2) \text{ for } i = 2 \end{aligned}$$

Sum of Squared Residuals

$$
\begin{aligned}
\hat{S} &= u_1^2 + u_1^2 \\
&= [3 - \hat{\beta}_0 + \hat{\beta}_1(1)]^2 + [5 - \hat{\beta}_0 + \hat{\beta}_1(2)]^2 \\
&= (9 + \hat{\beta}_0^2 + \hat{\beta}_1^2 - 6\hat{\beta}_0 - 6\hat{\beta}_1 + 2\hat{\beta}_0\hat{\beta}_1) + \\
&\qquad (25 + \hat{\beta}_0^2 + 4\hat{\beta}_1^2 - 10\hat{\beta}_0 - 20\hat{\beta}_1 + 4\hat{\beta}_0\hat{\beta}_1) \\
&= 2\hat{\beta}_0^2 + 5\hat{\beta}_1^2 - 16\hat{\beta}_0 - 26\hat{\beta}_1 + 6\hat{\beta}_0\hat{\beta}_1 + 34
\end{aligned}
$$

Choose $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize this...

Minimizing...

For:

$$\hat{S} = 2\hat{\beta}_0^2 + 5\hat{\beta}_1^2 - 16\hat{\beta}_0 - 26\hat{\beta}_1 + 6\hat{\beta}_0\hat{\beta}_1 + 34$$

We have:

$$\begin{aligned}
\frac{\partial \hat{S}}{\partial \hat{\beta}_0} &= 4\hat{\beta}_0 + 6\hat{\beta}_1 - 16 \\
\frac{\partial \hat{S}}{\partial \hat{\beta}_1} &= 6\hat{\beta}_0 + 10\hat{\beta}_1 - 26
\end{aligned}$$

So for $\hat{\beta}_1$:

$$\begin{aligned} 4\hat{\beta}_0 + 6\hat{\beta}_1 - 16 = 0 \quad &\Rightarrow \quad 2\hat{\beta}_0 = -3\hat{\beta}_1 + 8 \\ &\Rightarrow \quad \hat{\beta}_0 = -3/2\hat{\beta}_1 + 4 \end{aligned}$$

$$\begin{aligned} 6\hat{\beta}_0 + 10\hat{\beta}_1 - 26 = 0 \quad &\Rightarrow \quad 5\hat{\beta}_1 - 3(-3/2\hat{\beta}_1 + 4) - 13 = 0 \\ &\Rightarrow \quad 5\hat{\beta}_1 - 9/2\hat{\beta}_1 + 12 - 13 = 0 \\ &\Rightarrow \quad \frac{1}{2}\hat{\beta}_1 - 1 = 0 \\ &\Rightarrow \quad \boldsymbol{\hat{\beta}_1 = 2} \end{aligned}$$

And for $\hat{\beta}_0$:

$$\begin{aligned} 4\hat{\beta}_0 + 6(2) - 16 = 0 \quad &\Rightarrow \quad 4\hat{\beta}_0 = 4 \\ &\Rightarrow \quad \boldsymbol{\hat{\beta}_0 = 1} \end{aligned}$$

World's Simplest Regression

So:

$$Y_i = 1 + 2X_i + u_i$$

Note that, in this case:

$$\begin{aligned}\hat{\beta}_1 &= (5-3)/(2-1) \\ &= 2, \text{ and}\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= -2(2) + 5 \\ &= 1\end{aligned}$$

Least Squares with > 2 Observations

$$
\begin{aligned}
\hat{S} &= \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \\
&= \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \\
&= \sum_{i=1}^{N} (Y_i^2 - 2Y_i\hat{\beta}_0 - 2Y_i\hat{\beta}_1 X_i + \hat{\beta}_0^2 + 2\hat{\beta}_0\hat{\beta}_1 X_i + \hat{\beta}_1^2 X_i^2)
\end{aligned}
$$

Least Squares with > 2 Observations

Then:

$$\begin{aligned}
\frac{\partial \hat{S}}{\partial \hat{\beta}_0} &= \sum_{i=1}^{N} (-2Y_i + 2\hat{\beta}_0 + 2\hat{\beta}_1 X_i) \\
&= -2\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \\
&= -2\sum_{i=1}^{N} \hat{u}_i
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \hat{S}}{\partial \hat{\beta}_1} &= \sum_{i=1}^{N} (-2Y_i X_i + 2\hat{\beta}_0 X_i + 2\hat{\beta}_1 X_i^2) \\
&= -2\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i \\
&= -2\sum_{i=1}^{N} \hat{u}_i X_i
\end{aligned}$$

Least Squares with > 2 Observations

(Algebra happens...):

$$\sum_{i=1}^{N} Y_i = N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_i$$

and

$$\sum_{i=1}^{N} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{N} X_i + \hat{\beta}_1 \sum_{i=1}^{N} X_i^2$$

Least Squares with > 2 Observations

(More algebra...):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

and

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Intuition

$$\hat{\beta}_1 = \frac{\text{Covariance of } X \text{ and } Y}{\text{Variance of } X}$$

Parsing Variation in Y

Note that the "total" variation in Y is:

$$\text{SS}_{Total} = \sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

which comprises:

$$\begin{aligned} \text{SS}_{Residual} &= \sum_{i=1}^{N}(\hat{u}_i)^2 \\ &= \sum_{i=1}^{N}(Y_i - \hat{Y})^2 \end{aligned}$$

and:

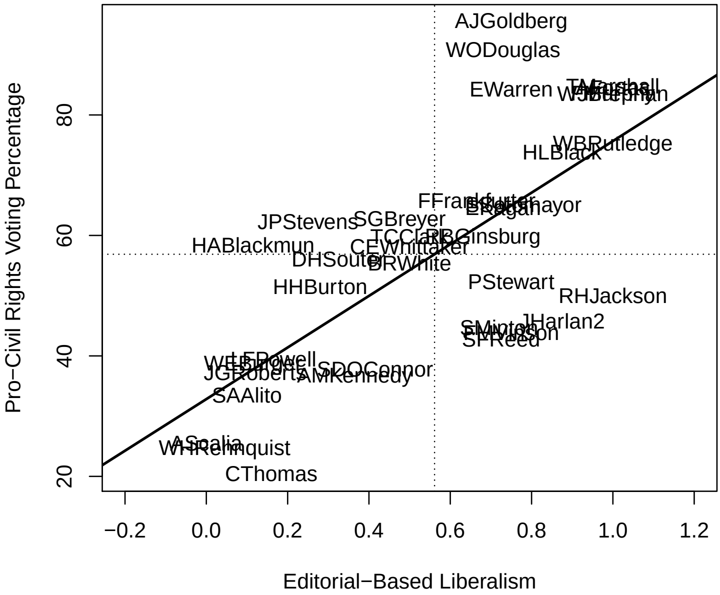
$$\text{SS}_{Model} = \sum_{i=1}^{N}(\hat{Y}_i - \bar{Y})^2$$

SCOTUS Example

- `ideology_score` $\in [0, 1] \rightarrow$ SCOTUS justice liberalism
- `civlibs` = liberal voting percentage

```
> summary(SCOTUS)

    justice      justiceName           civlibs     ideology_score
 Min.   : 78   Length:36          Min.   :21   Min.   :0.00
 1st Qu.: 88   Class :character   1st Qu.:42   1st Qu.:0.27
 Median : 96   Mode  :character   Median :57   Median :0.67
 Mean   : 96                      Mean   :57   Mean   :0.56
 3rd Qu.:105                      3rd Qu.:68   3rd Qu.:0.76
 Max.   :114                      Max.   :95   Max.   :1.00
```


AJGoldberg
WODouglas
EWarren
WBRutledge
HLBlack
JPStevens
SGBreyer
HABlackmun
DHSouter
BRWhite
HHBurton
PStewart
RHJackson
JHarlan2
SFReed
SDOConnor
AMKennedy
SAAlito
CThomas
20
40
60
80
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Pro−Civil Rights Voting Percentage
Editorial−Based Liberalism

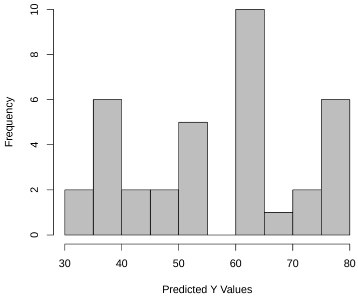
Estimating $\hat{\beta}$

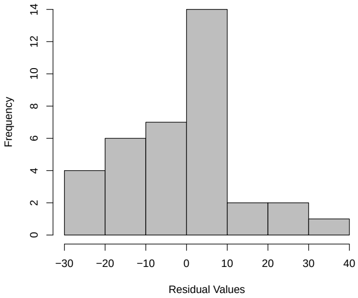
```
> Beta1 <- with(SCOTUS, (sum((ideology_score - mean(ideology_score)) *
                              (civlibs - mean(civlibs))) /
                          sum((ideology_score - mean(ideology_score))^2)))
> Beta1
[1] 43

> Beta0 <- with(SCOTUS, mean(civlibs) - (Beta1 * mean(ideology_score)))
> Beta0
[1] 33
```

$\hat{Y}$, $\hat{u}$, etc.

```
> SCOTUS$Yhats <- with(SCOTUS, Beta0 + Beta1*ideology_score)
> SCOTUS$Uhats <- with(SCOTUS, civlibs - Yhats)
> describe(SCOTUS$civlibs)
   vars  n mean sd median trimmed mad min max range skew kurtosis  se
X1    1 36   57 20     57      57  23  21  95    75 0.14    -0.94 3.3
> describe(SCOTUS$Yhats)
   vars  n mean sd median trimmed mad min max range  skew kurtosis  se
X1    1 36   57 14     62      57  18  33  76    43 -0.21     -1.4 2.3
> describe(SCOTUS$Uhats)
   vars  n mean sd median trimmed mad min max range skew kurtosis  se
X1    1 36    0 14   0.44   -0.21  13 -26  30    56 0.02    -0.58 2.3
```





What's a "typical" residual?

Recall that:

$$\bar{\hat{u}} = \frac{\sum_{i=1}^{N} \hat{u}_i}{N} = 0$$

Consider:

$$\text{RSE} = \sqrt{\left(\frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-1}\right)}$$

Sums of Squares, RSE, etc.

```
> TotalYVar <- with(SCOTUS, sum((civlibs - mean(civlibs))^2))
> TotalYVar
[1] 13651

> TotalUVar <- with(SCOTUS, sum((Uhats)^2))
> TotalUVar
[1] 6877

> TotalModelVar <- with(SCOTUS, sum((Yhats - mean(civlibs))^2))
> TotalModelVar
[1] 6774

> RSE <- with(SCOTUS, sqrt(TotalUVar / (nrow(SCOTUS)-2)))
> RSE
[1] 14
```

Estimating $\hat{\beta}$ via `lm`

```
> with(SCOTUS, summary(lm(civlibs~ideology_score)))

Call:
lm(formula = civlibs ~ ideology_score)

Residuals:
    Min      1Q  Median      3Q     Max
-25.662 -10.715   0.437   8.139  30.374

Coefficients:
              Estimate Std. Error t value    Pr(>|t|)
(Intercept)      32.83       4.78    6.86 0.000000067 ***
ideology_score   42.84       7.40    5.79 0.000001628 ***
---
Signif. codes:  0 *** 0.001  **  0.01  *  0.05  .  0.1     1

Residual standard error: 14 on 34 degrees of freedom
Multiple R-squared:  0.496,Adjusted R-squared:  0.481
F-statistic: 33.5 on 1 and 34 DF,  p-value: 0.00000163
```