

# PLSC 502 – Fall 2024

## Linear Regression I

November 18, 2024

Recall that a (real-valued) *random variable*  $Y$  is:

$$Y_i = \underbrace{\mu}_{\text{"systematic"}} + \underbrace{u_i}_{\text{"stochastic"}}$$

Note that we typically require that:

$$\text{Cov}(\mu, u) = 0.$$

Allow  $\mu$  to vary *linearly* with some other variable  $X$ :

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

so that we get

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

Goals:

- *Point estimates* of  $\beta_0$  and  $\beta_1$  (call them  $\hat{\beta}_0$  and  $\hat{\beta}_1$ )
- Estimates of their *variability*  $\rightarrow$  inference

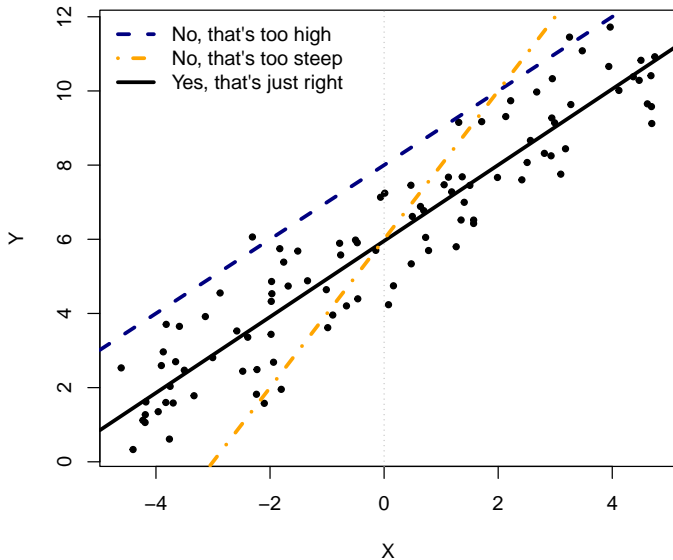
## Estimating $\beta_0$ and $\beta_1$

Suppose we have some estimates  $\hat{\beta}_0$  and  $\hat{\beta}_1$ . Then:

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

→ estimated “residuals”:

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



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

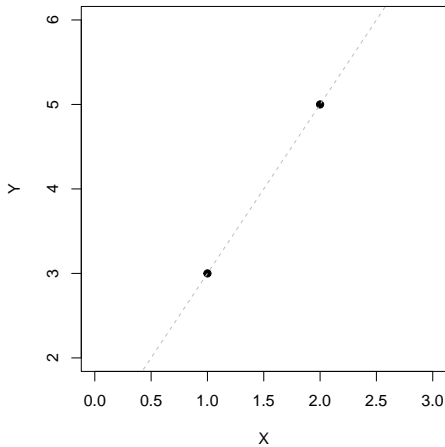
Possibilities:

- 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”)

→ “ordinary least squares” (“OLS”) regression...

# 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

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

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

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  $\hat{\beta}_1$ :

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

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

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

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

# World's Simplest Regression

So:

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

Note that, in this ( $N=2$ ) 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

Next, set:

$$-2 \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

and

$$-2 \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0$$

... and solve...

# Least Squares “Normal Equations”

(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: Solutions!

(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

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

The intuition:

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

# Parsing Variation in $Y$ : ANOVA

Note that the “total” variation in  $Y$  around its mean  $\bar{Y}$  is:

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

which comprises:

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

and:

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

# SCOTUS Data, OT1946-2023

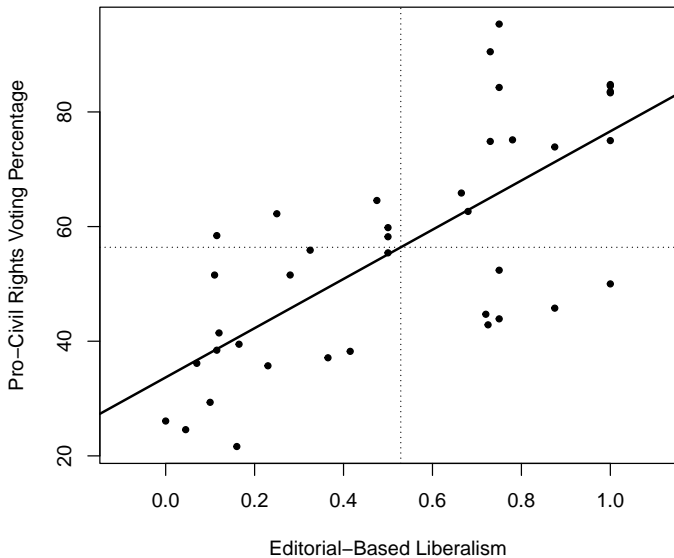
Data from the [Supreme Court Database](#) and the justices' Segal-Cover scores...

- $Y$  is CivLibs = liberal voting percentage in civil rights & liberties cases
- $X$  is IdeologyScore  $\in [0, 1] \rightarrow$  SCOTUS justice liberalism (based on post-nomination, pre-confirmation newspaper editorials)

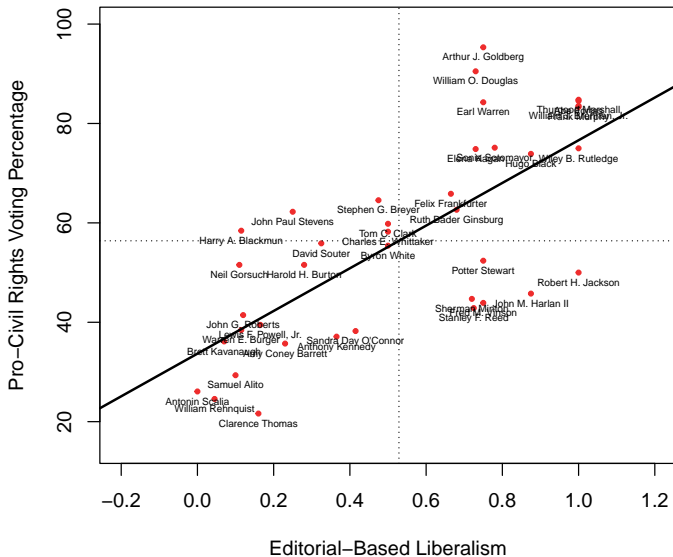
```
> describe(SCOTUS,skew=FALSE,trim=0)
```

	vars	n	mean	sd	median	min	max	range	se
justice	1	39	97.87	11.60	98.00	78.00	117.00	39.00	1.86
justiceName*	2	39	20.00	11.40	20.00	1.00	39.00	38.00	1.83
CivLibs	3	39	56.39	19.90	55.42	21.63	95.33	73.70	3.19
Nom.Order*	4	39	20.00	11.40	20.00	1.00	39.00	38.00	1.83
Nominee*	5	39	20.00	11.40	20.00	1.00	39.00	38.00	1.83
ChiefJustice*	6	4	1.00	0.00	1.00	1.00	1.00	0.00	0.00
SenateVote*	7	39	16.69	8.42	19.00	1.00	25.00	24.00	1.35
IdeologyScore	8	39	0.53	0.33	0.50	0.00	1.00	1.00	0.05
QualificationsScore	9	39	0.80	0.24	0.88	0.12	1.00	0.88	0.04
Nominator (Party)*	10	39	6.92	3.72	6.00	1.00	13.00	12.00	0.60
Year	11	39	1971.03	25.66	1967.00	1937.00	2020.00	83.00	4.11

# Le Scatterplot



# Le Labeled Scatterplot



```
> # Betas:

> Beta1 <- with(SCOTUS, (sum((IdeologyScore - mean(IdeologyScore)) *
+                             (CivLibs - mean(CivLibs)))) /
+                             sum((IdeologyScore - mean(IdeologyScore))^2)))
> Beta1
[1] 42.94

> Beta0 <- with(SCOTUS, mean(CivLibs) - (Beta1 * mean(IdeologyScore)))
> Beta0
[1] 33.69
```

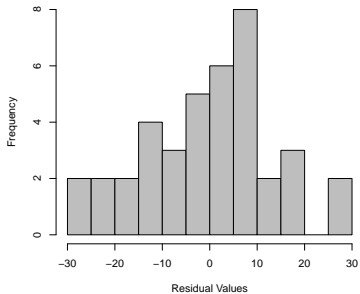
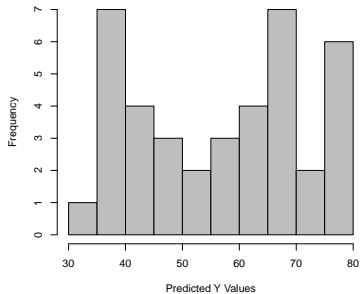
```
> SCOTUS$Yhats <- with(SCOTUS, Beta0 + Beta1*IdeologyScore)
> SCOTUS$Uhats <- with(SCOTUS, CivLibs - Yhats)

> # Y itself:
> describe(SCOTUS$CivLibs)
  vars  n mean   sd median trimmed  mad   min   max range skew kurtosis   se
X1    1 39 56.39 19.9  55.42   56.25 25.17 21.63 95.33  73.7 0.19   -1.06 3.19

> # Predicted Ys:
> describe(SCOTUS$Yhats)
  vars  n mean   sd median trimmed  mad   min   max range skew kurtosis   se
X1    1 39 56.39 14.28  55.16   56.46 17.19 33.69 76.62 42.94 -0.06   -1.47 2.29

> # Residuals:
> describe(SCOTUS$Uhats)
  vars  n mean   sd median trimmed  mad   min   max range skew kurtosis   se
X1    1 39   0 13.85   2.61   -0.02 11.66 -26.62 29.44 56.06 -0.11   -0.63 2.22
```

# $\hat{Y}$ and $\hat{u}$ Plots





## What's a “typical” residual?

Note that because

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

it's also true that:

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

Consider instead:

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

## Sums of Squares, RSE, etc.

```
> # Sums of squares:
>
> TotalYVar <- with(SCOTUS, sum((CivLibs - mean(CivLibs))^2))
> TotalYVar
[1] 15047

> TotalUVar <- with(SCOTUS, sum((Uhats)^2))
> TotalUVar
[1] 7294

> TotalModelVar <- with(SCOTUS, sum((Yhats - mean(CivLibs))^2))
> TotalModelVar
[1] 7753

> RSE <- with(SCOTUS, sqrt(TotalUVar / (nrow(SCOTUS)-2)))
> RSE
[1] 14.04
```

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

```
> fit<-lm(CivLibs~IdeologyScore,data=SCOTUS)
> summary(fit)
```

Call:

```
lm(formula = CivLibs ~ IdeologyScore, data = SCOTUS)
```

Residuals:

Min	1Q	Median	3Q	Max
-26.62	-9.84	2.61	8.05	29.44

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	33.69	4.26	7.91	1.8e-09	***
IdeologyScore	42.94	6.85	6.27	2.7e-07	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14 on 37 degrees of freedom

Multiple R-squared: 0.515, Adjusted R-squared: 0.502

F-statistic: 39.3 on 1 and 37 DF, p-value: 2.7e-07

```
> anova(fit)
```

Analysis of Variance Table

Response: CivLibs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
IdeologyScore	1	7753	7753	39.3	2.7e-07 ***
Residuals	37	7294	197		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Inference

For the linear model:

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

Note that we can rewrite the formula for  $\hat{\beta}_1$ :

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2} \\&= \frac{\sum_{i=1}^N (X_i - \bar{X})Y_i - \sum_{i=1}^N (X_i - \bar{X})\bar{Y}}{\sum_{i=1}^N (X_i - \bar{X})^2} \\&= \frac{\sum_{i=1}^N (X_i - \bar{X})Y_i - \bar{Y} \sum_{i=1}^N (X_i - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2} \\&= \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}(0))}{\sum_{i=1}^N (X_i - \bar{X})^2} \\&= \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - \bar{X})^2}\end{aligned}$$

## Variation in $\hat{\beta}_0$ and $\hat{\beta}_1$

$\hat{\beta}_0$  and  $\hat{\beta}_1$  are *random variables*...

- Q: Where does their variation come from?
- A: From the *stochastic* variation in  $Y$ ...
- ...that is, from  $u$ .

Next question: What does the random variation in  $Y$  “look like”?

## Getting To $\text{Var}(\hat{\beta}_1)$

An (as we will see, largely unimportant) assumption:

$$u_i \sim \text{i.i.d. } N(0, \sigma^2)$$

Implies:

$$\text{Var}(Y|X, \beta) = \sigma^2$$

so:

$$\begin{aligned}\text{Var}(\hat{\beta}_1) &= \text{Var} \left[ \frac{\sum_{i=1}^N (X_i - \bar{X}) Y_i}{\sum_{i=1}^N (X_i - \bar{X})^2} \right] \\&= \left[ \frac{1}{\sum (X_i - \bar{X})^2} \right]^2 \sum (X_i - \bar{X})^2 \text{Var}(Y) \\&= \left[ \frac{1}{\sum (X_i - \bar{X})^2} \right]^2 \sum (X_i - \bar{X})^2 \sigma^2 \\&= \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.\end{aligned}$$



$$\text{Var}(\hat{\beta}_0) \text{ and } \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

Similarly:

$$\text{Var}(\hat{\beta}_0) = \frac{\sum X_i^2}{N \sum (X_i - \bar{X})^2} \sigma^2$$

and :

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \sigma^2$$

## Note that:

- $\text{Var}(\hat{\beta}_0)$  and  $\text{Var}(\hat{\beta}_1) \propto \sigma^2$   
 $\hookrightarrow \text{Var}(\hat{\beta}_s)$  increases as  $Y$  gets “noisier” ...
- $\text{Var}(\hat{\beta}_0)$  and  $\text{Var}(\hat{\beta}_1) \propto -\sum(X_i - \bar{X})$   
 $\hookrightarrow \text{Var}(\hat{\beta}_s)$  decreases with greater variation in  $X$ ...
- $\text{Var}(\hat{\beta}_0)$  and  $\text{Var}(\hat{\beta}_1) \propto -N$   
 $\hookrightarrow \text{Var}(\hat{\beta}_s)$  decreases as  $N$  gets larger...
- $\text{sign}[\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)] = -\text{sign}(\bar{X})$   
 $\hookrightarrow$  The sign of the covariance between  $\hat{\beta}_0$  and  $\hat{\beta}_1$  is the opposite of the sign of the mean of  $X$

# The Gauss-Markov Theorem

*“Given the assumptions of the classical linear regression model, the least squares estimators are the **minimum variance estimators** among the class of unbiased linear estimators. (They are BLUE).”*

Imagine:

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Rewrite:

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

$k$  are “weights”:

$$\hat{\beta}_1 = \sum k_i Y_i$$

with  $k_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$ .

Alternative (non-LS) estimator:

$$\tilde{\beta}_1 = \sum w_i Y_i$$

Unbiasedness requires:

$$\begin{aligned} E(\tilde{\beta}_1) &= \sum w_i E(Y_i) \\ &= \sum w_i (\beta_0 + \beta_1 X_i) \\ &= \beta_0 \sum w_i + \beta_1 \sum w_i X_i \end{aligned}$$

Variance:

$$\begin{aligned}\text{Var}(\tilde{\beta}_1) &= \text{Var}\left(\sum w_i Y_i\right) \\&= \sigma^2 \sum w_i^2 \\&= \sigma^2 \sum \left[ w_i - \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} + \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right]^2 \\&= \sigma^2 \sum \left[ w_i - \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right]^2 + \sigma^2 \left[ \frac{1}{\sum (X_i - \bar{X})^2} \right]\end{aligned}$$

## Gauss-Markov (continued)

Because  $\sigma^2 \left[ \frac{1}{\sum (X_i - \bar{X})^2} \right]$  is a constant,  $\min[\text{Var}(\tilde{\beta}_1)]$  minimizes

$$\sum \left[ w_i - \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right]^2.$$

...which is minimized at:

$$w_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2},$$

implying:

$$\begin{aligned} \text{Var}(\tilde{\beta}_1) &= \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \\ &= \text{Var}(\hat{\beta}_1) \end{aligned}$$

# Gauss-Markov Requirements

For the Gauss-Markov theorem to hold, it must be the case that:

1.  $Y = f(X) + u$  with  $f(X)$  linear\*
2.  $E(u) = 0$
3.  $\text{Cov}(X, u) = 0$
- 4a.  $\text{Var}(u) = \sigma^2 \forall i$
- 4b.  $\text{Cov}(u_i, u_j) = 0$
5.  $\text{Rank}(\mathbf{X}) = k$

(...don't sweat these too much for now...)

\*But see below...



## BLUE vs. BUE:

- OLS has been BLUE since about 1821 (see, e.g., [Plackett 1949](#)).
- [Hansen \(2022\)](#): OLS is “BUE” – most efficient among *all* unbiased estimators, linear or otherwise...
- Challenged by others; resolved by [Portnoy \(2022\)](#): Any unbiased estimator *must* be linear (so “BLUE” = “BUE”).
- A pretty good nontechnical discussion of all this by Paul Allison is [here](#).

If  $u_i \sim N(0, \sigma^2)$ , then:

$$\hat{\beta}_0 \sim N[\beta_0, \text{Var}(\hat{\beta}_0)]$$

and

$$\hat{\beta}_1 \sim N[\beta_1, \text{Var}(\hat{\beta}_1)]$$

Means:

$$\begin{aligned} z_{\hat{\beta}_1} &= \frac{(\hat{\beta}_1 - \beta_1)}{\sqrt{\text{Var}(\hat{\beta}_1)}} \\ &= \frac{(\hat{\beta}_1 - \beta_1)}{\text{s.e.}(\hat{\beta}_1)} \\ &= \sim N(0, 1) \end{aligned}$$

## A Small Problem...

$$\sigma^2 = ???$$

Solution: use

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{N - k}$$

Gives:

$$\widehat{\text{Var}(\hat{\beta}_1)} = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2},$$

and

$$\widehat{\text{Var}(\hat{\beta}_0)} = \frac{\sum X_i^2}{N \sum (X_i - \bar{X})^2} \hat{\sigma}^2$$

Which means that:

$$\begin{aligned}
 \widehat{\text{s.e.}}(\hat{\beta}_1) &= \sqrt{\widehat{\text{Var}}(\hat{\beta}_1)} \\
 &= \sqrt{\frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2}} \\
 &= \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}
 \end{aligned}$$

and implies:

$$\begin{aligned}
 t_{\hat{\beta}_1} &\equiv \frac{(\hat{\beta}_1 - \beta_1)}{\widehat{\text{s.e.}}(\hat{\beta}_1)} = \frac{(\hat{\beta}_1 - \beta_1)}{\frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}} \\
 &= \frac{(\hat{\beta}_1 - \beta_1) \sqrt{\sum (X_i - \bar{X})^2}}{\hat{\sigma}} \\
 &\sim t_{N-k}
 \end{aligned}$$

# Predictions and Variance

Point prediction:

$$\hat{Y}_k = \hat{\beta}_0 + \hat{\beta}_1 X_k$$

$Y_k$  is unbiased:

$$\begin{aligned} E(\hat{Y}_k) &= E(\hat{\beta}_0 + \hat{\beta}_1 X_k) \\ &= E(\hat{\beta}_0) + X_k E(\hat{\beta}_1) \\ &= \beta_0 + \beta_1 X_k \\ &= E(Y_k) \end{aligned}$$

Variability:

$$\begin{aligned} \text{Var}(\hat{Y}_k) &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 X_k) \\ &= \frac{\sum X_i^2}{N \sum (X_i - \bar{X})^2} \sigma^2 + \left[ \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \right] X_k^2 + 2 \left[ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \sigma^2 \right] X_k \\ &= \sigma^2 \left[ \frac{1}{N} + \frac{(X_k - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] \end{aligned}$$

# Variability of Predictions

The variability of a prediction  $\hat{Y}_k$  is:

$$\text{Var}(\hat{Y}_k) = \sigma^2 \left[ \frac{1}{N} + \frac{(X_k - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]$$

which means that  $\text{Var}(\hat{Y}_k)$ :

- ...decreases in  $N$
- ...decreases in  $\text{Var}(X)$
- ...increases in  $|X_k - \bar{X}|$

*Standard error of the prediction:*

$$\widehat{\text{s.e.}}(\hat{Y}_k) = \sqrt{\sigma^2 \left[ \frac{1}{N} + \frac{(X_k - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]}$$

→ (e.g.) confidence intervals:

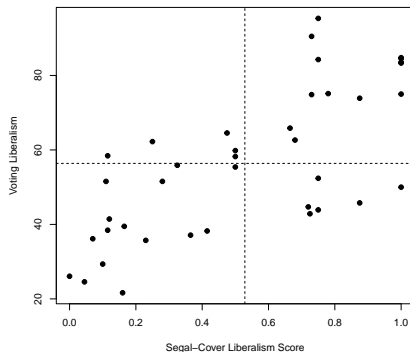
$$95\% \text{ c.i.}(\hat{Y}_k) = \hat{Y}_k \pm [1.96 \times \widehat{\text{s.e.}}(\hat{Y}_k)]$$

# Example: SCOTUS Liberalism

```
> with(SCOTUS, describe(CivLibs))
  vars  n mean   sd median trimmed  mad   min   max range skew kurtosis   se
X1     1 39 56.39 19.9  55.42   56.25 25.17 21.63 95.33  73.7 0.19   -1.06 3.19

> with(SCOTUS, describe(IdeologyScore))
  vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
X1     1 39 0.53 0.33   0.5   0.53 0.4  0  1  1 -0.06   -1.47 0.05
```

Scatterplot of SCOTUS Voting and Liberalism Scores





## Example, Continued

```
> SCLib<-lm(CivLibs~IdeologyScore,data=SCOTUS)
> summary(SCLib)    # regression
```

Call:

```
lm(formula = CivLibs ~ IdeologyScore, data = SCOTUS)
```

Residuals:

Min	1Q	Median	3Q	Max
-26.62	-9.84	2.61	8.05	29.44

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	33.69	4.26	7.91	1.8e-09	***
IdeologyScore	42.94	6.85	6.27	2.7e-07	***
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14 on 37 degrees of freedom

Multiple R-squared: 0.515, Adjusted R-squared: 0.502

F-statistic: 39.3 on 1 and 37 DF, p-value: 2.7e-07

## Example, Continued

```
> anova(SCLib)      # ANOVA
```

Analysis of Variance Table

Response: CivLibs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
IdeologyScore	1	7753	7753	39.3	2.7e-07 ***
Residuals	37	7294	197		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$\text{Var}(\hat{\beta})$ :

```
> vcov(SCLib)
```

	(Intercept)	IdeologyScore
(Intercept)	18.16	-24.79
IdeologyScore	-24.79	46.88

95 percent c.i.s:

```
> confint(SCLib)
```

	2.5 %	97.5 %
(Intercept)	25.05	42.32
IdeologyScore	29.06	56.81

99 percent c.i.s:

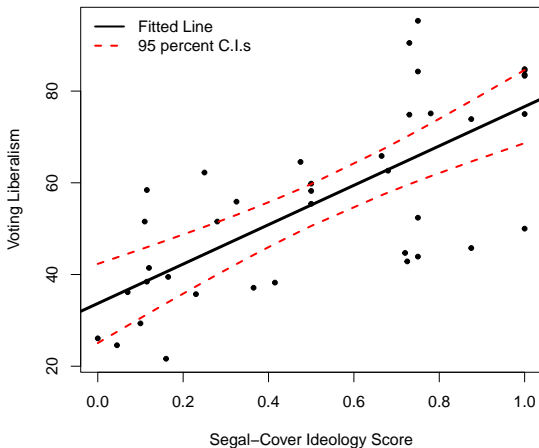
```
> confint(SCLib,level=0.99)
```

	0.5 %	99.5 %
(Intercept)	22.12	45.26
IdeologyScore	24.35	61.53

```
> SEs<-predict(SCLib,interval="confidence")
> SEs
      fit   lwr   upr
1  71.26 64.64 77.88
2  64.82 59.51 70.12
3  62.24 57.31 67.17
.
.
<rows omitted>
.
.
37 38.41 31.03 45.79
38 36.69 28.87 44.52
39 43.56 37.40 49.72
```

# A Plot, With CIs

Scatterplot of SCOTUS Voting and Ideology Scores, along with Least-Squares Line and 95% Prediction Confidence Intervals



## The code:

```
Sort<-order(SCOTUS$IdeologyScore)

pdf("SCLib-CI-24.pdf",6,5)
par(mar=c(4,4,2,2))
plot(SCOTUS$IdeologyScore,SCOTUS$CivLibs,pch=20,
      xlab="Segal-Cover Ideology Score",
      ylab="Voting Liberalism")
abline(SCLib,lwd=3)
lines(sort(SCOTUS$IdeologyScore),SEs[Sort,2],col="red",lwd=2,lty=2)
lines(sort(SCOTUS$IdeologyScore),SEs[Sort,3],col="red",lwd=2,lty=2)
legend("topleft",bty="n",lty=c(1,2),lwd=2,col=c("black","red"),
      legend=c("Fitted Line","95 percent C.I.s"))
dev.off()
```

# Surely There's An Easier Way?

```
> library(marginaleffects)
> plot_predictions(SCLib, condition="IdeologyScore", points=1)
  + theme_classic()
```

