PLSC 502 – Fall 2023
Linear Regression I

November 13, 2023

Linearity

Linearity means:

$$\frac{\partial Y}{\partial X} = m;$$

$$Y = mX + b$$

Other monotonic + "smooth" alternatives:

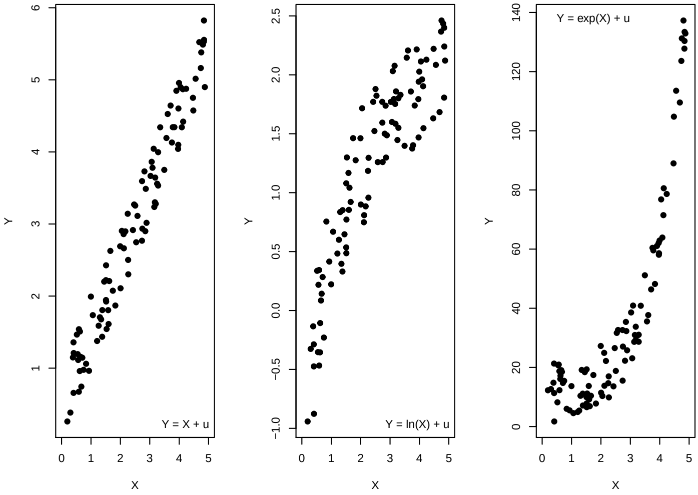
- *Logarithmic*:

$$\frac{\partial^2 Y}{\partial X \partial X} < 0$$

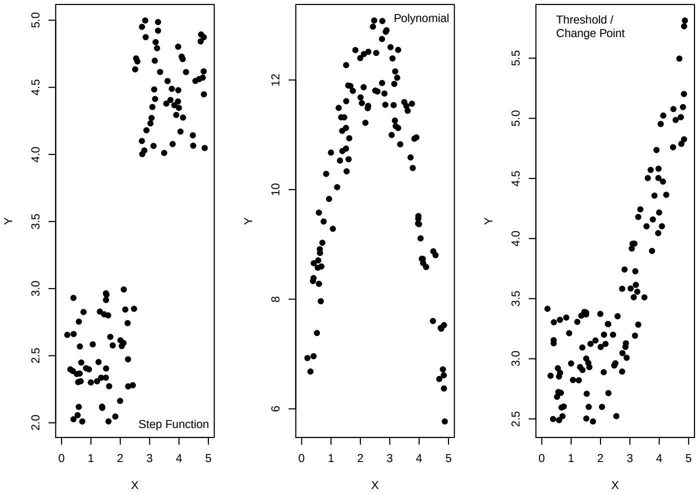
- *Exponential*:

$$\frac{\partial^2 Y}{\partial X \partial X} > 0$$

Linear, Logarithmic, Exponential



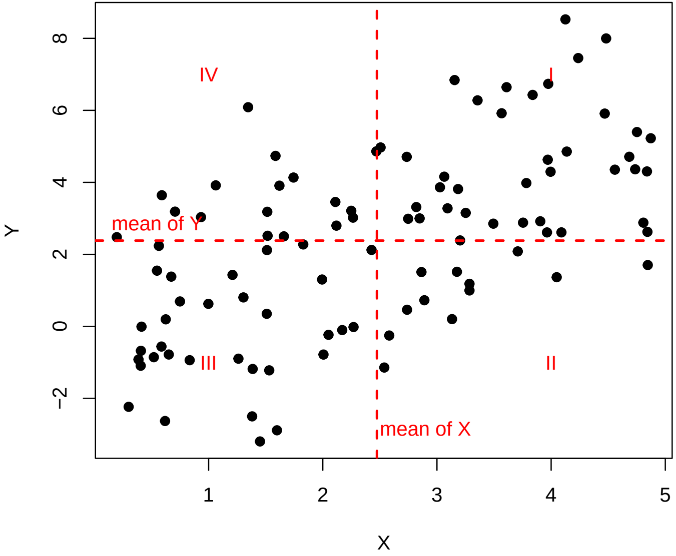
Other Possibilities



"Pearson's product-moment correlation coefficient":

$$
\begin{aligned}
r &= \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \\
&= \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \\
&= \frac{\sum_{i=1}^{N}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)}{N-1}
\end{aligned}
$$

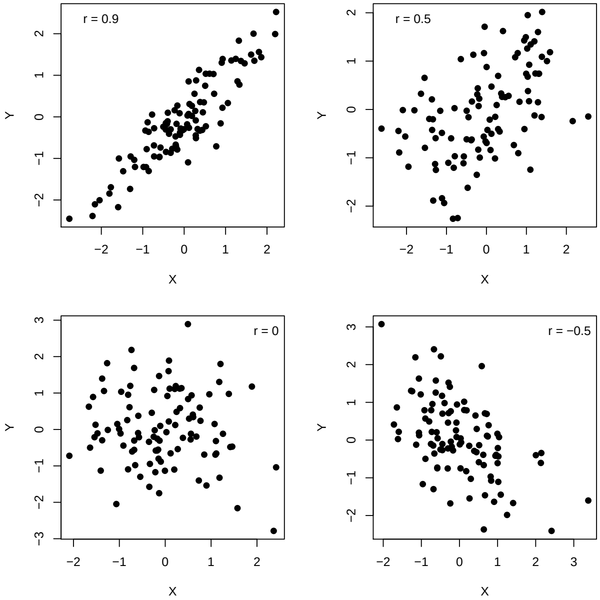
Pearson's r: Intuition



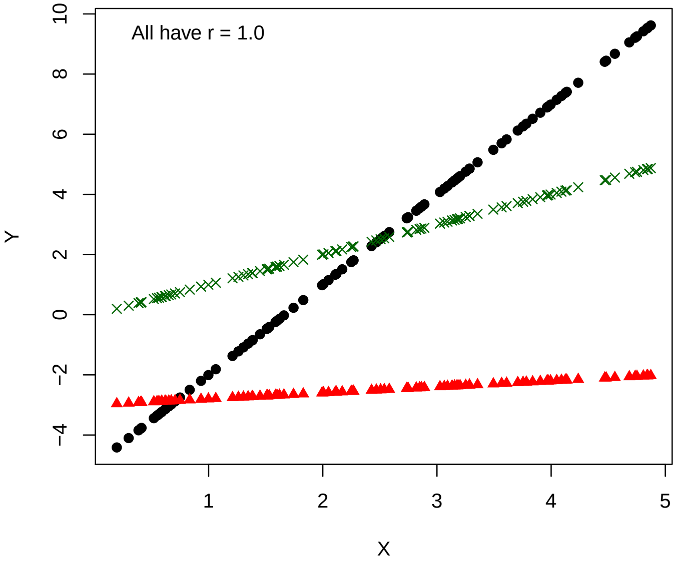
Pearson's r: Characteristics

- $r \in [-1, 1]$
- $r = 0 \leftrightarrow$ no *linear* association between Y and X.
- Sign(r) $\rightarrow$ "direction" of the *linear* association
- $|r| \rightarrow$ "strength" of the *linear* association
- In general:
    - $|r| < 0.3 \rightarrow$ "weak" linear association
    - $0.3 < |r| < 0.7 \rightarrow$ "moderate" linear association
    - $|r| > 0.7 \rightarrow$ "strong" linear association

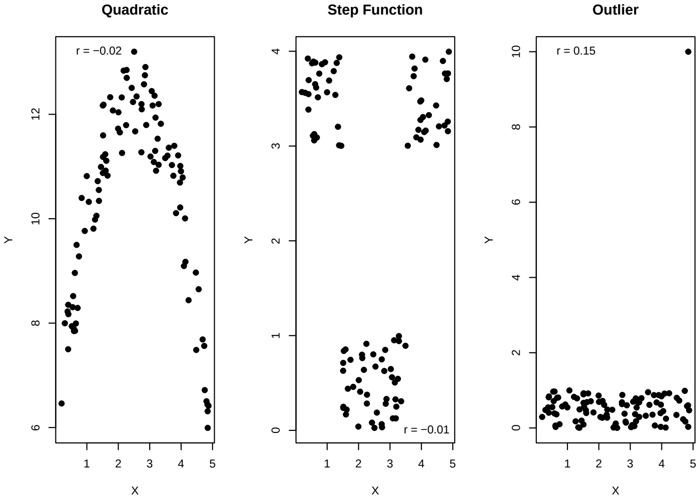
Examples



$r = \pm 1.0 \rightarrow ?$



Nonlinearity, etc.



The sampling distribution of r is:

- complex, and
- skewed as $|r| \to 1.0$.

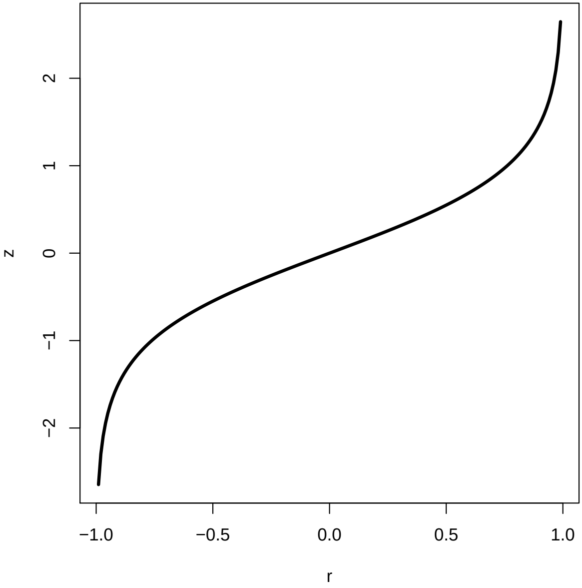
Fisher:

$$\hat{w} \equiv \frac{1}{2} \ln \left(\frac{1+\hat{r}}{1-\hat{r}} \right) \sim \mathcal{N} \left[\frac{1}{2} \ln \left(\frac{1+\hat{r}}{1-\hat{r}} \right), \frac{1}{\sqrt{N-3}} \right]$$

implying:

$$z_r = \frac{\frac{1}{2} \ln \left(\frac{1+\hat{r}}{1-\hat{r}} \right) - \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)}{\sqrt{\frac{1}{N-3}}} \sim \mathcal{N}(0,1)$$

Fisher's z Transformation of r



Alternative Approach (t)

Under $r = 0$, the standard error of $\hat{r}$ is:

$$\sigma_r = \sqrt{\frac{1 - r^2}{N - 2}}$$

This means that we can construct confidence intervals using a t distribution, as:

$$\frac{\hat{r}}{\sigma_r} = \frac{\hat{r}\sqrt{N-2}}{\sqrt{1-\hat{r}^2}} \quad \sim \quad t_{N-2}.$$

Note that this converges to z as $N \to \infty$.

Alternative Measure: Spearman's ρ

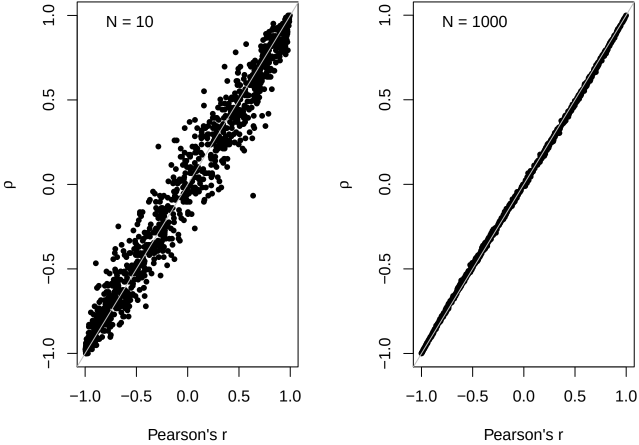
For sorted data on X and Y, where R_{Y_i} and R_{X_i} are the respective ranks,

$$\rho = 1 - \frac{6\sum_{i=1}^{N}(R_{Y_i} - R_{X_i})^2}{N(N^2 - 1)}$$

Characteristics:

- $\rho \in [-1, 1]$
- Same interpretation as r.
- Also appropriate for use with ordinal data; but
- When many "ties" occur, calculate Pearson's r on the ranks R_{Y_i} and R_{X_i}, and assign "partial" (or "half") ranks to tied individuals.

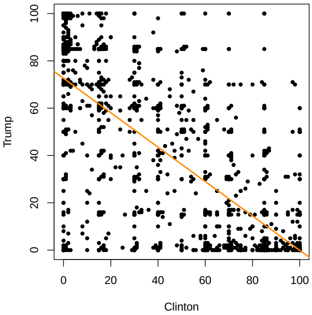
r vs. ρ Comparison (Simulation)



Real Data: ANES 2016 Feeling Thermometers

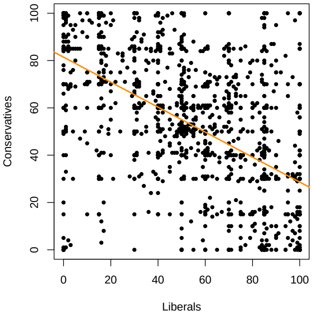
```
> describe(Therms,range=FALSE)
                           vars    n  mean    sd  skew kurtosis   se
Asian-Americans               1 2387 70.17 20.20 -0.38     0.02 0.41
Hispanics                     2 2387 69.35 20.91 -0.41     0.01 0.43
Blacks                        3 2387 69.00 21.19 -0.35    -0.24 0.43
Illegal Immigrants            4 2387 42.54 27.31  0.13    -0.71 0.56
Whites                        5 2387 71.63 19.40 -0.46     0.08 0.40
Dem. Pres. Candidate          6 2387 44.12 34.91  0.12    -1.42 0.71
GOP Pres. Candidate           7 2387 40.53 35.65  0.23    -1.43 0.73
Libertarian Pres. Candidate   8 2387 43.61 19.92 -0.58     0.25 0.41
Green Pres. Candidate         9 2387 43.20 20.87 -0.54     0.22 0.43
Dem. VP                      10 2387 48.24 25.91 -0.22    -0.44 0.53
GOP VP                       11 2387 49.59 33.42 -0.10    -1.21 0.68
John Roberts                 12 2387 53.75 18.39 -0.41     1.44 0.38
Pope Francis                 13 2387 69.55 25.17 -0.73     0.14 0.52
Christian Fundamentalists    14 2387 48.59 28.48 -0.07    -0.72 0.58
Feminists                    15 2387 56.94 26.65 -0.24    -0.47 0.55
Liberals                     16 2387 52.27 27.35 -0.24    -0.67 0.56
Labor Unions                 17 2387 56.70 24.74 -0.27    -0.29 0.51
Poor People                  18 2387 72.20 19.63 -0.36    -0.06 0.40
Big Business                 19 2387 49.34 22.52 -0.15    -0.18 0.46
Conservatives                20 2387 55.22 25.91 -0.24    -0.45 0.53
SCOTUS                       21 2387 59.34 19.38 -0.32     0.54 0.40
Gays & Lesbians              22 2387 62.83 26.86 -0.46    -0.20 0.55
Congress                     23 2387 41.17 22.32  0.02    -0.34 0.46
Rich People                  24 2387 53.53 20.69 -0.13     0.52 0.42
Muslims                      25 2387 55.80 25.64 -0.29    -0.23 0.52
Christians                   26 2387 74.40 23.80 -0.87     0.35 0.49
Jews                         27 2387 72.20 21.19 -0.45    -0.14 0.43
Tea Party                    28 2387 42.97 27.08 -0.06    -0.70 0.55
Police                       29 2387 75.57 22.50 -1.15     1.13 0.46
Transgender People           30 2387 57.29 26.88 -0.28    -0.31 0.55
Scientists                   31 2387 77.74 19.23 -0.77     0.39 0.39
BLM                          32 2387 48.26 32.66 -0.06    -1.15 0.67
```

Feeling Thermometers: Clinton vs. Trump



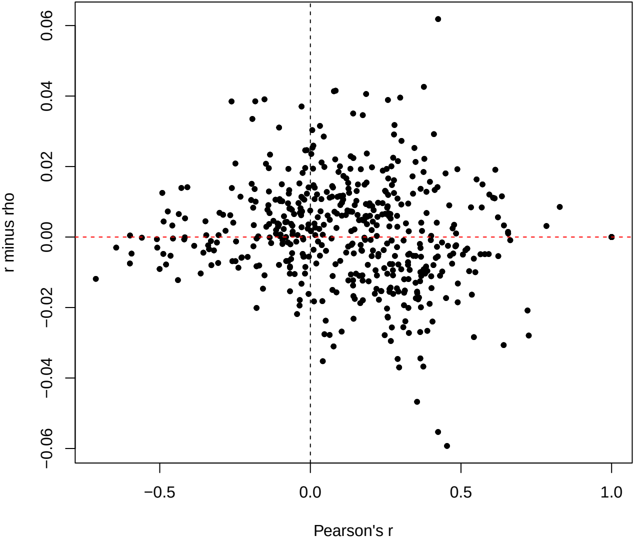

```
> rCT<-with(Therms, cor(‘Dem. Pres. Candidate‘,‘GOP Pres. Candidate‘))
> rCT
[1] -0.71227

> rCT2<-with(Therms, cor.test(‘Dem. Pres. Candidate‘,‘GOP Pres. Candidate‘))
> rCT2

Pearson’s product-moment correlation

data:  Dem. Pres. Candidate and GOP Pres. Candidate
t = -49.6, df = 2385, p-value <2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.73148 -0.69192
sample estimates:
     cor
-0.71227


> # Identical:
>
> (rCT*sqrt(nrow(Therms)-2)) / sqrt(1-(rCT^2))
[1] -49.557
```

Liberals and Conservatives

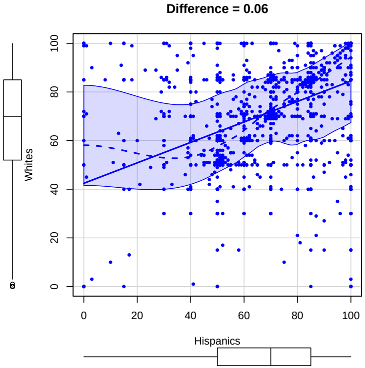


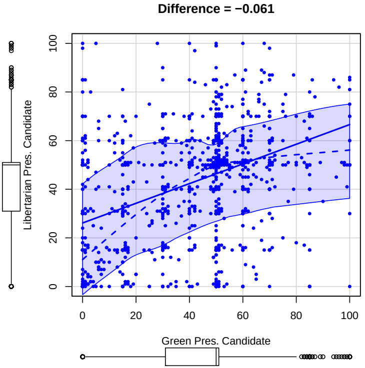

```
> rLC<-with(Therms, cor.test(Liberals,Conservatives))
> rLC

Pearson's product-moment correlation

data:  Liberals and Conservatives
t = -28.2, df = 2385, p-value <2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.52983 -0.46966
sample estimates:
     cor
-0.50035


> rhoLC<-with(Therms, SpearmanRho(Liberals,Conservatives))
> rhoLC
[1] -0.49128
```

Pairwise FT Differences between r and ρ



Biggest Differences Between r and ρ





Summary: Measures of Association

Which bivariate measure of association should I use?

		X			
		Nominal	Binary	Ordinal	Interval/Ratio
	Nominal	χ^2	χ^2	χ^2	t-test (and η)
Y	Binary	χ^2	ϕ, Q	γ, τ_c	t-test
	Ordinal	χ^2	γ, τ_c	γ, τ_a, τ_b	Spearman's ρ
	Interval / Ratio	t-test (and η)	t-test	Spearman's ρ	r

Linear Regression

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 μ 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 β_0 and β_1 (call them $\hat{\beta}_0$ and $\hat{\beta}_1$)
- Estimates of their *variability* $\rightarrow$ inference

Estimating β_0 and β_1

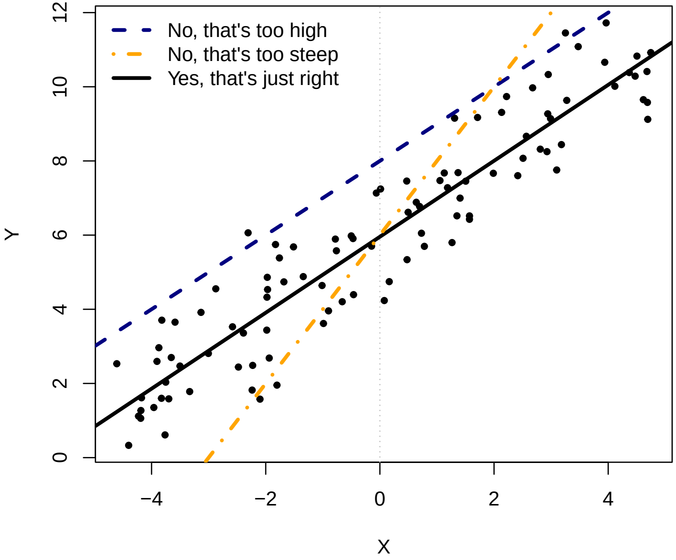
Suppose we have some estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. Then:

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

$\rightarrow$ estimated "residuals":

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

Intuition



No, that's too high
No, that's too steep
Yes, that's just right
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.

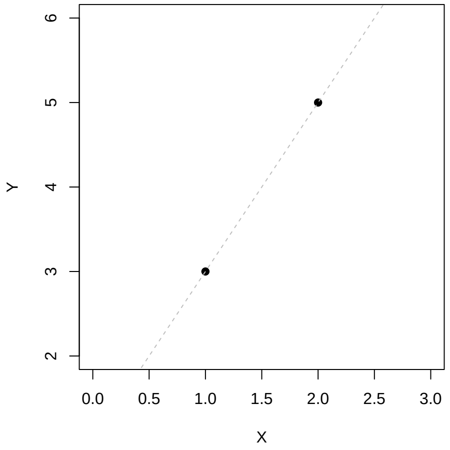
Possibilities:

- 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”)

$\rightarrow$ “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

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

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

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 $\hat{\beta}_0$:

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

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

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

(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:

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

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

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

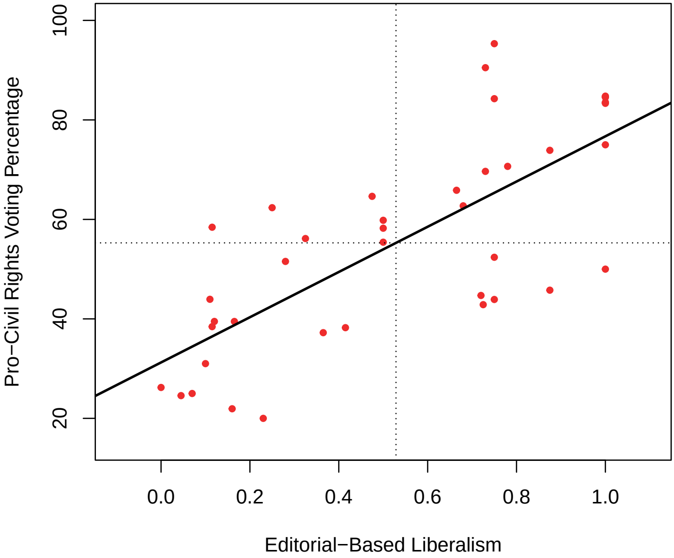
SCOTUS Data, OT1946-2022

Data from the Supreme Court Database and the justices' Segal-Cover scores...

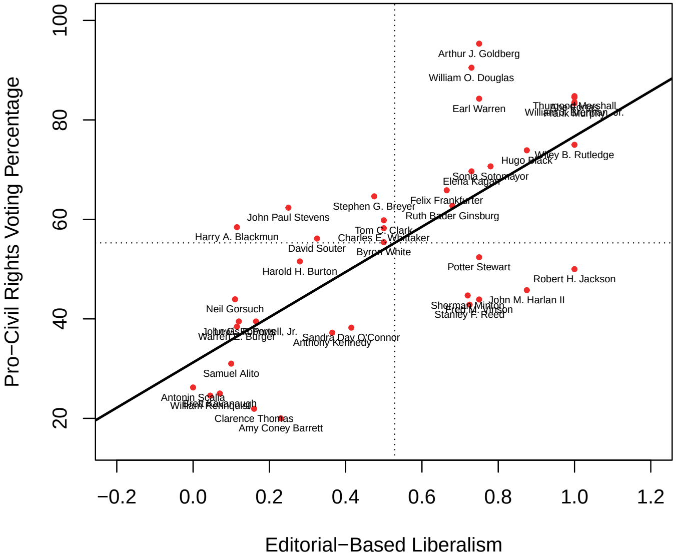
- Y is `CivLibs` = liberal voting percentage in civil rights & liberties cases
- X is `IdeologyScore` $\in [0, 1] \rightarrow$ SCOTUS justice liberalism

```
> describe(SCOTUS,skew=FALSE,trim=0)
                    vars  n    mean    sd  min     max range   se
justice                1 39   97.87 11.60   78  117.00 39.00 1.86
justiceName*           2 39   20.00 11.40    1   39.00 38.00 1.83
CivLibs                3 39   55.28 20.67   20   95.33 75.33 3.31
Nom.Order*             4 39   20.00 11.40    1   39.00 38.00 1.83
Nominee*               5 39   20.00 11.40    1   39.00 38.00 1.83
ChiefJustice*          6  4    1.00  0.00    1    1.00  0.00 0.00
SenateVote*            7 39   16.69  8.42    1   25.00 24.00 1.35
IdeologyScore          8 39    0.53  0.33    0    1.00  1.00 0.05
QualificationsScore*   9 39   16.38  7.82    1   25.00 24.00 1.25
Nominator (Party)*    10 39    6.92  3.72    1   13.00 12.00 0.60
Year                  11 39 1971.03 25.66 1937 2020.00 83.00 4.11
```

Le Scatterplot



Le Labeled Scatterplot



Estimating $\hat{\beta}$

```
> # Betas:

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

> Beta1
[1] 45.5



> Beta0 <- with(SCOTUS, mean(CivLibs) - (Beta1 * mean(IdeologyScore)))
> Beta0
[1] 31.2
```

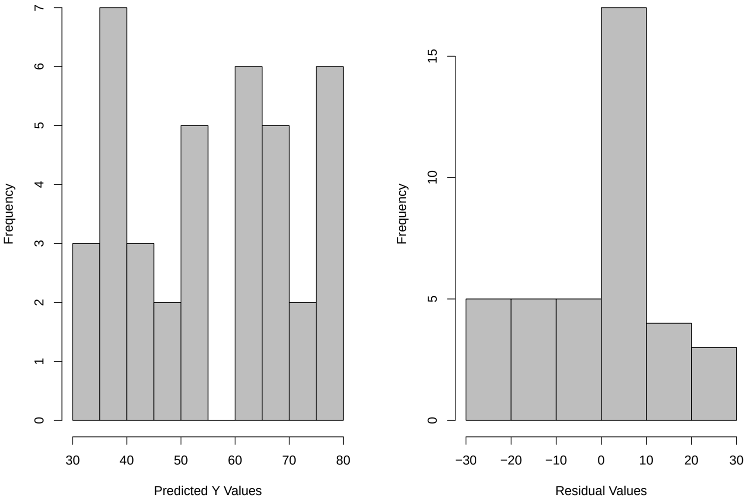
```
> # Residuals, etc.
>
> SCOTUS$Yhats <- with(SCOTUS, Beta0 + Beta1*IdeologyScore)
> SCOTUS$Uhats <- with(SCOTUS, CivLibs - Yhats)
```

```
> # Y itself:
> describe(SCOTUS$CivLibs)
   vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
X1    1 39 55.3 20.7   55.4    55.1 23.6  20 95.3  75.3 0.13    -1.03 3.31
```

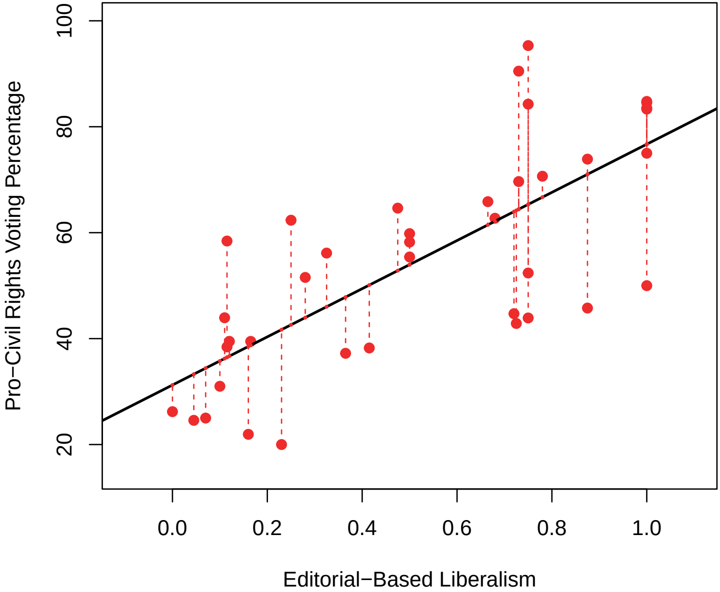
```
> # Predicted Ys:
> describe(SCOTUS$Yhats)
   vars  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
X1    1 39 55.3 15.1     54    55.4 18.2 31.2 76.7  45.5 -0.06    -1.47 2.42
```

```
> # Residuals:
> describe(SCOTUS$Uhats)
   vars  n mean   sd median trimmed  mad   min max range  skew kurtosis   se
X1    1 39    0 14.1   2.79   -0.13 11.2 -26.7  30  56.7 -0.07    -0.61 2.26
```

$\hat{Y}$s and $\hat{u}$s



Le Scatterplot, Again



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] 16235


> TotalUVar <- with(SCOTUS, sum((Uhats)^2))
> TotalUVar
[1] 7543


> TotalModelVar <- with(SCOTUS, sum((Yhats - mean(CivLibs))^2))
> TotalModelVar
[1] 8693


> RSE <- with(SCOTUS, sqrt(TotalUVar / (nrow(SCOTUS)-2)))
> RSE
[1] 14.3
```

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
-26.70 -10.01   2.79   7.64  29.99

Coefficients:
              Estimate Std. Error t value    Pr(>|t|)
(Intercept)      31.24       4.33    7.21 0.000000015 ***
IdeologyScore    45.47       6.96    6.53 0.000000121 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.3 on 37 degrees of freedom
Multiple R-squared:  0.535,Adjusted R-squared:  0.523
F-statistic: 42.6 on 1 and 37 DF,  p-value: 0.000000121
```

ANOVA with `lm`

```
> anova(fit)

Analysis of Variance Table

Response: CivLibs
              Df Sum Sq Mean Sq F value     Pr(>F)
IdeologyScore  1   8693    8693    42.6 0.00000012 ***
Residuals     37   7543     204
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```