

450A Political Methodology I

Lecture 4: Simple Regression

1. Mechanics of Ordinary Least Squares

Stanford University

Mechanics of Ordinary Least Squares

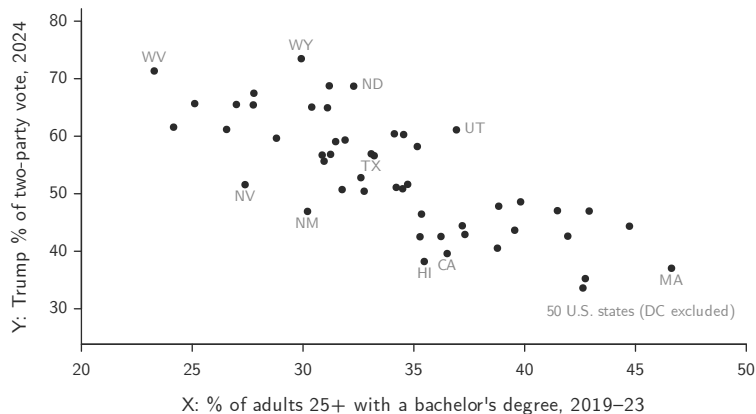
Principle of Least Squares

Deriving OLS Estimators

Analysis of Variance for Regression

Example: College Degrees and the 2024 Vote in the States

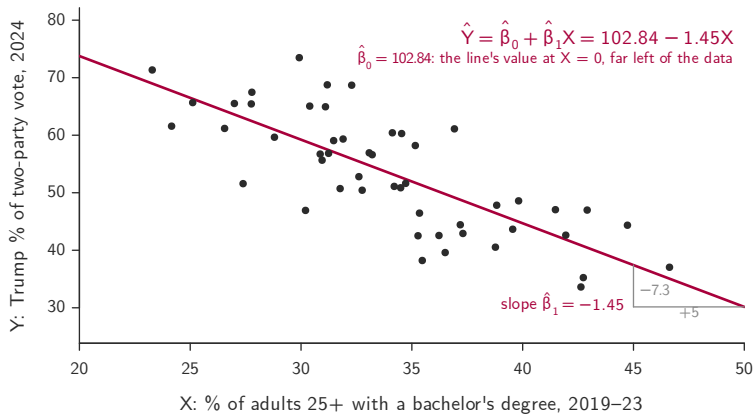
How do we fit the regression line $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ to the data?



Source: MIT Election Data and Science Lab, 2024 presidential returns by state; Census Bureau, American Community Survey 2019–23. DC excluded.

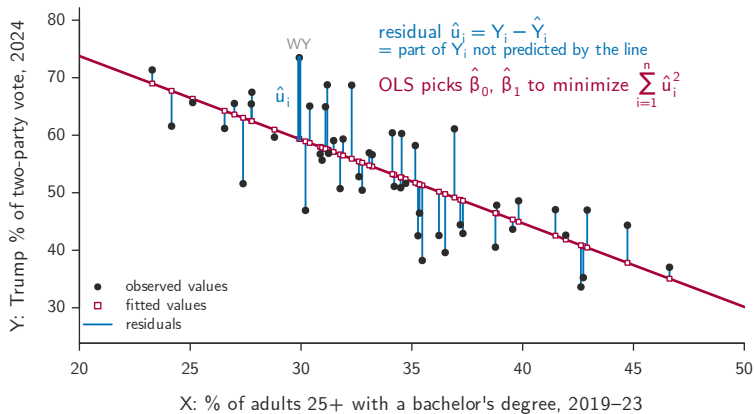
Ordinary Least Squares (OLS) Fit

How do we fit the regression line $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ to the data?



How OLS Fits a Regression Line to Data

Answer: We will minimize the sum of squared residuals



Ordinary Least Squares (OLS)

- Let $\{\tilde{\beta}_0, \tilde{\beta}_1\}$ be possible values for $\{\beta_0, \beta_1\}$
- *Ordinary Least Squares (OLS) regression* picks the following:

$$\{\hat{\beta}_0, \hat{\beta}_1\} = \underset{\tilde{\beta}_0, \tilde{\beta}_1}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \tilde{\beta}_0 - x_i \tilde{\beta}_1)^2$$

Why least squares (i.e. why minimize the sum of squared differences)?

- Easy to derive and analytically investigate
- Optimal under some assumptions (Lecture 4-2)

Mechanics of Ordinary Least Squares

Principle of Least Squares

Deriving OLS Estimators

Analysis of Variance for Regression

Univariate Least Squares Estimator

- Let's understand how to **calculate least squares** analytically
- Consider the simplest case: minimizing the sum of squares for a single variable Y
- n sample observations: $y_1, y_2, y_3, \dots, y_n$.
- What is a single number $\tilde{\mu}$ that summarizes all y 's?
- Find $\tilde{\mu}$ that minimizes the **sum of squared residuals** (SSR)

$$S(\tilde{\mu}) = \sum_{i=1}^n (y_i - \tilde{\mu})^2.$$

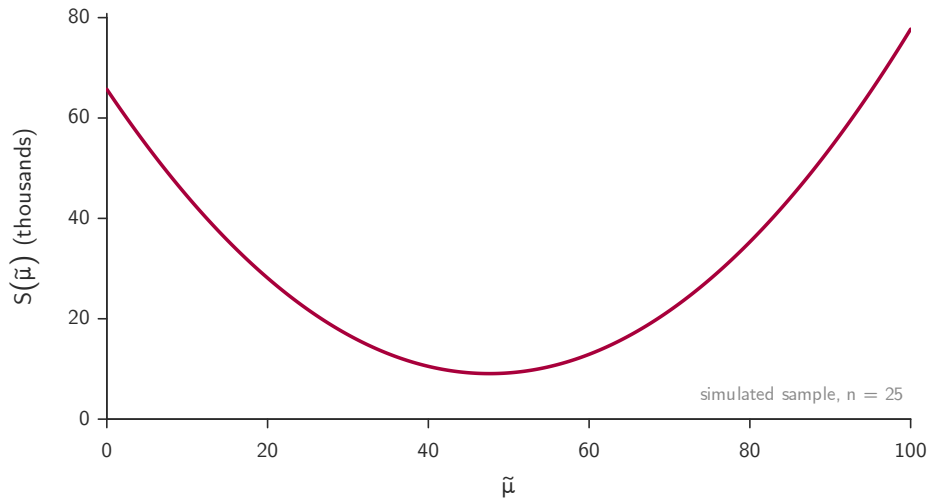
Worksheet Q1 (*OLS as sample mean*)

Show that the sample mean $\bar{Y}$ is the OLS estimator of μ in $Y_i = \mu + u_i$.

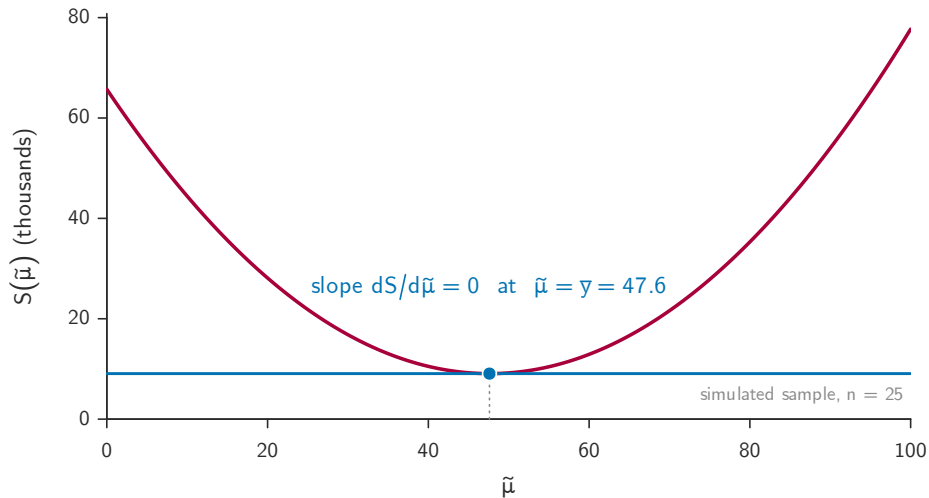
How do we solve this?

- 1 Calculate the derivative of S with respect to $\tilde{\mu}$
- 2 Set the derivative equal to 0
- 3 Solve for $\tilde{\mu}$ and replace $\tilde{\mu}$ with the solution

The Objective Function for SSR



The Slope of the Tangent Line for SSR



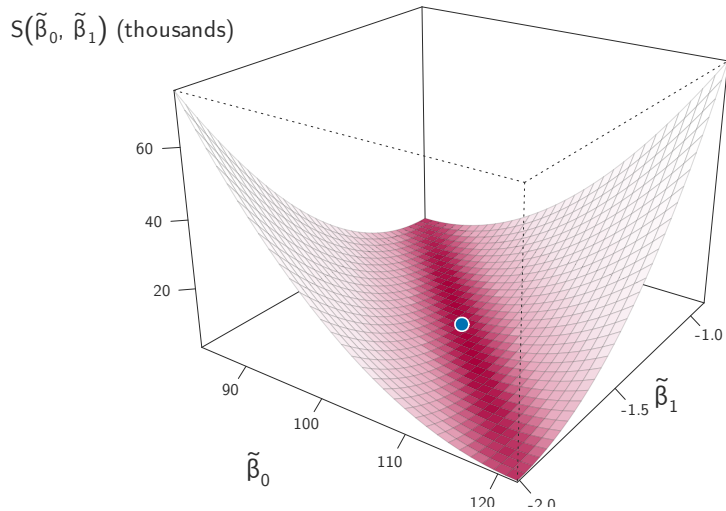
Deriving the Linear Least Squares Estimator

- Now we look at two variables, Y and X
- n pairs of sample observations: $\{y_1, x_1\}, \{y_2, x_2\}, \dots, \{y_n, x_n\}$
- Let $\{\tilde{\beta}_0, \tilde{\beta}_1\}$ be possible values for $\{\beta_0, \beta_1\}$
- Define the least squares objective function:

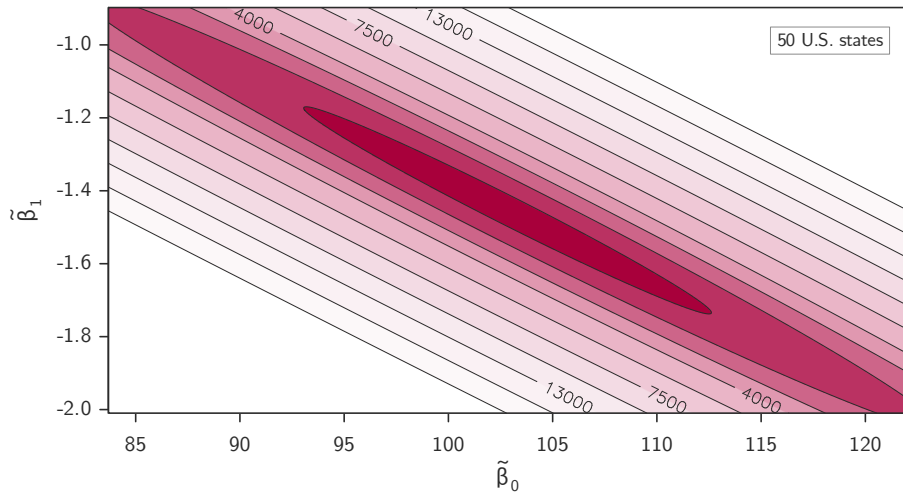
$$S(\tilde{\beta}_0, \tilde{\beta}_1) = \sum_{i=1}^n (y_i - \tilde{\beta}_0 - x_i \tilde{\beta}_1)^2.$$

- How do we derive the LS estimators for β_0 and β_1 ?
 - ① Take partial derivatives of S with respect to $\tilde{\beta}_0$ and $\tilde{\beta}_1$.
 - ② Set each of the partial derivatives to 0
 - ③ Solve for $\{\tilde{\beta}_0, \tilde{\beta}_1\}$ and replace them with the solutions

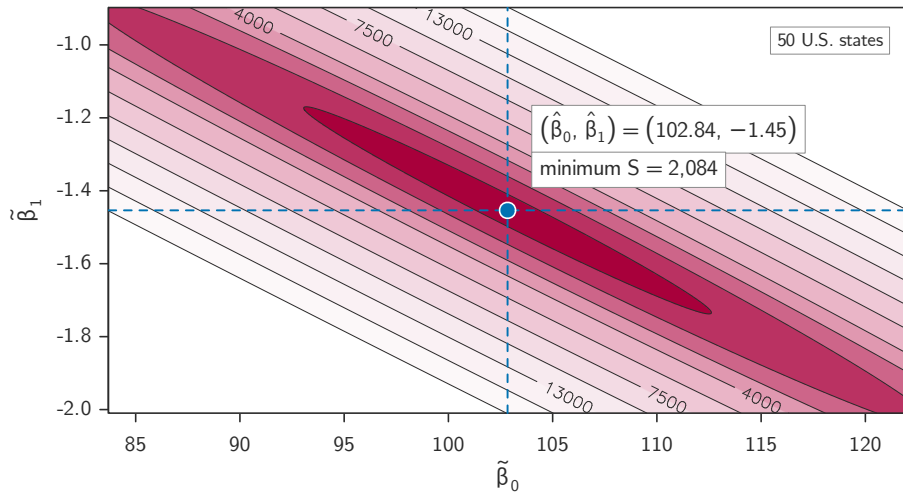
The Objective Function for OLS



The Objective Function for OLS: $S(\tilde{\beta}_0, \tilde{\beta}_1)$



The Objective Function for OLS



$$\begin{aligned}
 S(\tilde{\beta}_0, \tilde{\beta}_1) &= \sum_{i=1}^n (y_i - \tilde{\beta}_0 - x_i \tilde{\beta}_1)^2 \\
 &= \sum_{i=1}^n (y_i^2 - 2y_i \tilde{\beta}_0 - 2y_i \tilde{\beta}_1 x_i + \tilde{\beta}_0^2 + 2\tilde{\beta}_0 \tilde{\beta}_1 x_i + \tilde{\beta}_1^2 x_i^2)
 \end{aligned}$$

$$\frac{\partial S(\tilde{\beta}_0, \tilde{\beta}_1)}{\partial \tilde{\beta}_0} = \sum_{i=1}^n (-2y_i + 2\tilde{\beta}_0 + 2\tilde{\beta}_1 x_i)$$

and

$$\frac{\partial S(\tilde{\beta}_0, \tilde{\beta}_1)}{\partial \tilde{\beta}_1} = \sum_{i=1}^n (-2y_i x_i + 2\tilde{\beta}_0 x_i + 2\tilde{\beta}_1 x_i^2)$$

The first order conditions are

$$0 = \sum_{i=1}^n (-2y_i + 2\tilde{\beta}_0 + 2\tilde{\beta}_1 x_i)$$

and

$$0 = \sum_{i=1}^n (-2y_i x_i + 2\tilde{\beta}_0 x_i + 2\tilde{\beta}_1 x_i^2)$$

rearranging for $\tilde{\beta}_0$ and $\tilde{\beta}_1$ yields:

$$\hat{\beta}_0 n = \left(\sum_{i=1}^n y_i \right) - \hat{\beta}_1 \left(\sum_{i=1}^n x_i \right)$$

$$\hat{\beta}_1 \sum_{i=1}^n x_i^2 = \left(\sum_{i=1}^n x_i y_i \right) - \hat{\beta}_0 \left(\sum_{i=1}^n x_i \right)$$

and solving

$$\begin{aligned}\hat{\beta}_0 n &= \left(\sum_{i=1}^n y_i \right) - \hat{\beta}_1 \left(\sum_{i=1}^n x_i \right) \\ \hat{\beta}_1 \sum_{i=1}^n x_i^2 &= \left(\sum_{i=1}^n x_i y_i \right) - \hat{\beta}_0 \left(\sum_{i=1}^n x_i \right)\end{aligned}$$

yields the OLS estimators:

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

College Degrees and the 2024 Trump Vote, 50 States

R Code

```
> lm.out <- lm(trump ~ ba, data = states)
> summary(lm.out)

~~~~ output omitted ~~~~

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  102.8376      5.9818  17.192 < 2e-16 ***
ba           -1.4537      0.1738  -8.366 6.2e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.589 on 48 degrees of freedom
Multiple R-squared:  0.5932,    Adjusted R-squared:  0.5847
F-statistic: 69.99 on 1 and 48 DF,  p-value: 6.199e-11
```

Worksheet Q2 (*OLS slope as covariance-variance ratio*)

Prove that $\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(X, Y)}{\widehat{\text{Var}}(X)} = \frac{\text{Sample Covariance between } X \text{ and } Y}{\text{Sample Variance of } X}$.

The slope coefficient is the derivative of the **fitted regression line** with respect to X :

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$\frac{\partial \hat{Y}}{\partial X} = \hat{\beta}_1$$

$\hat{\beta}_1$ = how much the fitted value $\hat{Y}$ differs between two units whose X differs by one unit

In the example, $\hat{\beta}_1 = -1.45$. A state where the share of adults with a bachelor's degree is 1 point higher has a fitted Trump share 1.45 points lower. The slope compares states. It does not say what a degree does to anyone's vote.

Worksheet Q3

Show that the OLS residual vector $\hat{u}$ is orthogonal to the regressor vector x .

What does it mean?

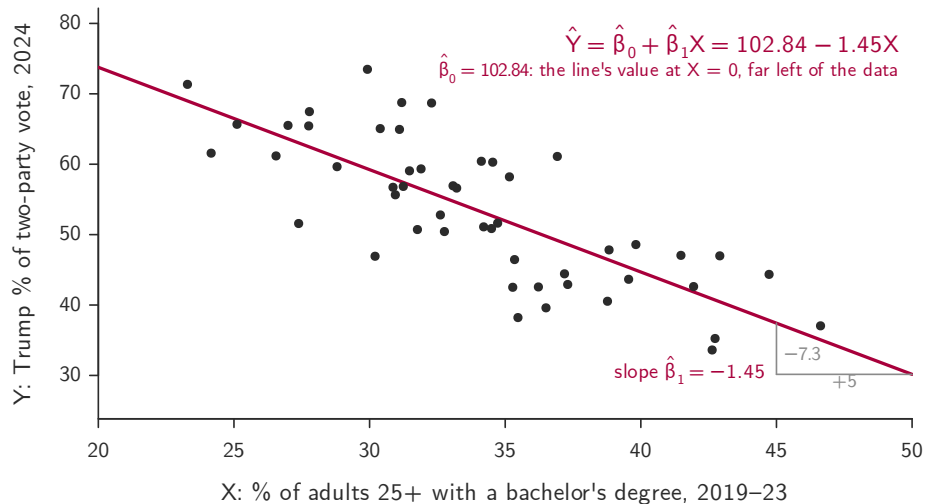
Mechanics of Ordinary Least Squares

Principle of Least Squares

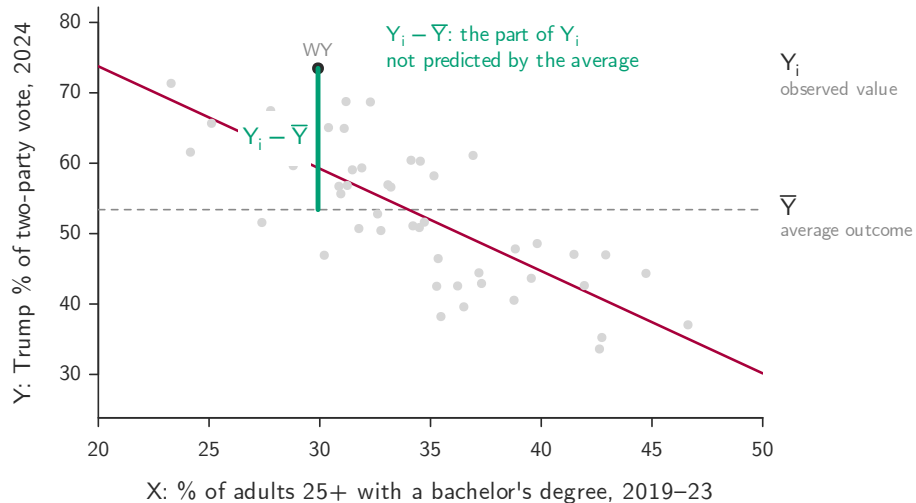
Deriving OLS Estimators

Analysis of Variance for Regression

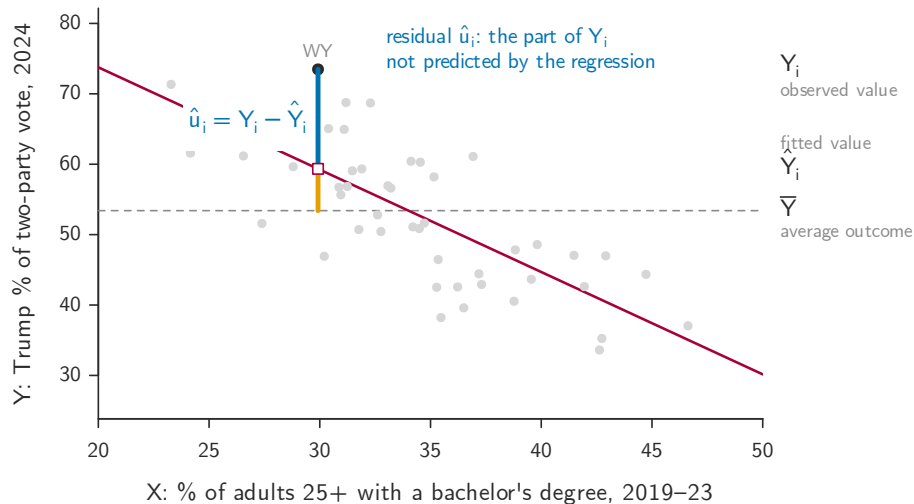
Ordinary Least Squares (OLS) Fit: $\hat{\beta}_0 \hat{\beta}_1$



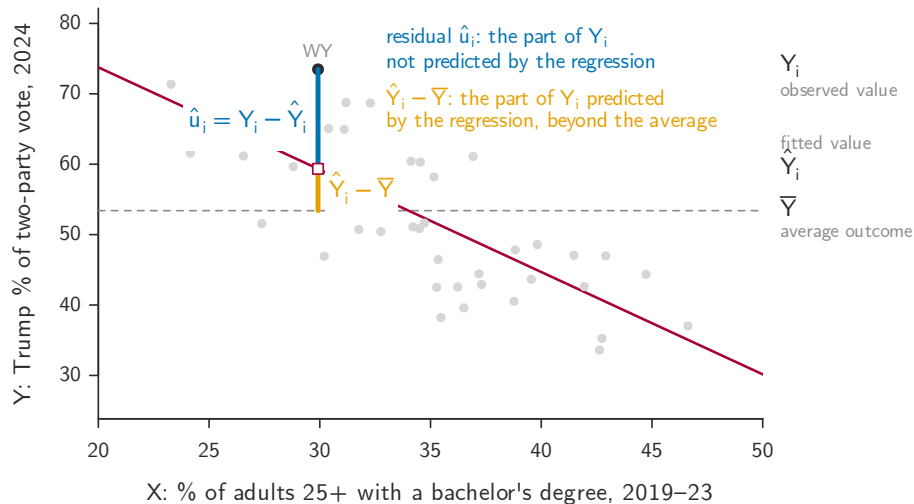
Pick a Y_i and the Average Outcome $\bar{Y}$



Decompose into Residual $\hat{u}_i$ and Predicted Part



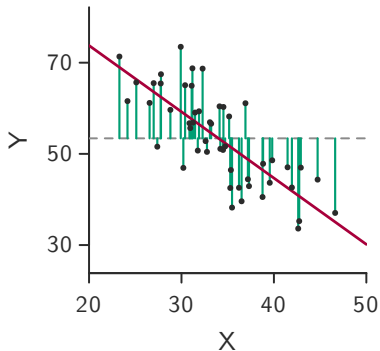
Decompose into Residual $\hat{u}_i$ and Predicted Part



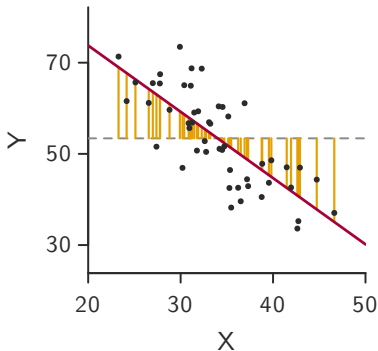
- $\sum_{i=1}^n (y_i - \bar{y})^2 = SST = (n - 1) \widehat{\text{Var}}(y)$ (Total Sum of Squares)
- $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SSE = (n - 1) \widehat{\text{Var}}(\hat{y})$ (Explained / Model Sum of Squares)
- $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \hat{u}_i^2 = SSR = (n - 1) \widehat{\text{Var}}(\hat{u})$ (Residual Sum of Squares)
- **Worksheet Q4** Show: $SST = SSE + SSR$

Visual Illustration of Sums of Squares

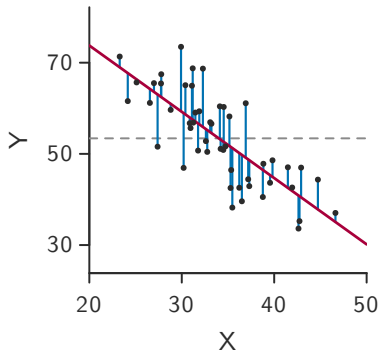
$$SST = \sum (y_i - \bar{y})^2$$
$$= 5,122$$



$$SSE = \sum (\hat{y}_i - \bar{y})^2$$
$$= 3,039$$



$$SSR = \sum (y_i - \hat{y}_i)^2$$
$$= 2,084$$



Coefficient of Determination R^2

Given $SST = SSE + SSR$ we can divide each side by the SST :

$$\begin{aligned}\frac{SST}{SST} &= \frac{SSE}{SST} + \frac{SSR}{SST} \\ \Leftrightarrow \frac{SSE}{SST} &= 1 - \frac{SSR}{SST} \equiv R^2\end{aligned}$$

Interpretation:

Share of the total variation in Y (i.e. SST) that is explained by X

Properties:

$$0 \leq R^2 \leq 1$$

- If $R^2 = 1$, all points are on a straight line (perfect fit)
- If $R^2 = 0$, the sample correlation between Y and X is zero and the fitted line is flat (X may still be related to Y in a nonlinear way)

College Degrees and the 2024 Trump Vote, 50 States

R Code

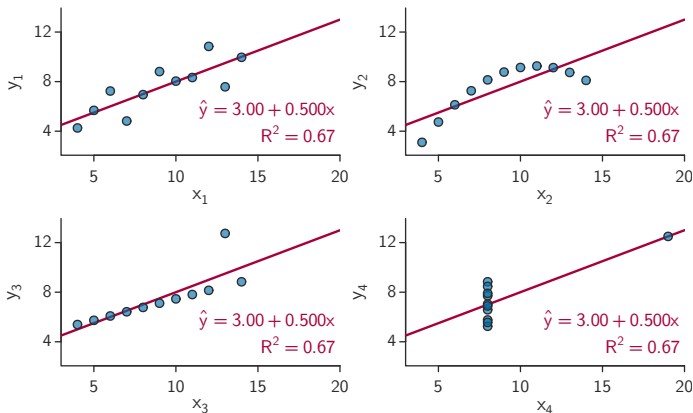
```
> lm.out <- lm(trump ~ ba, data = states)
> summary(lm.out)

~~~~ output omitted ~~~~

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  102.8376      5.9818  17.192 < 2e-16 ***
ba           -1.4537      0.1738  -8.366 6.2e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.589 on 48 degrees of freedom
Multiple R-squared:  0.5932,    Adjusted R-squared:  0.5847
F-statistic: 69.99 on 1 and 48 DF,  p-value: 6.199e-11
```

Does R^2 tell you how well a line describes the data? No! The same R^2 fits very different data:



Source: Anscombe (1973), "Graphs in Statistical Analysis," *The American Statistician* 27(1); R data set `anscombe`