

ORDINARY LEAST SQUARES

J. Alexander Branham

Fall 2016

INTRODUCTION TO ORDINARY LEAST SQUARES

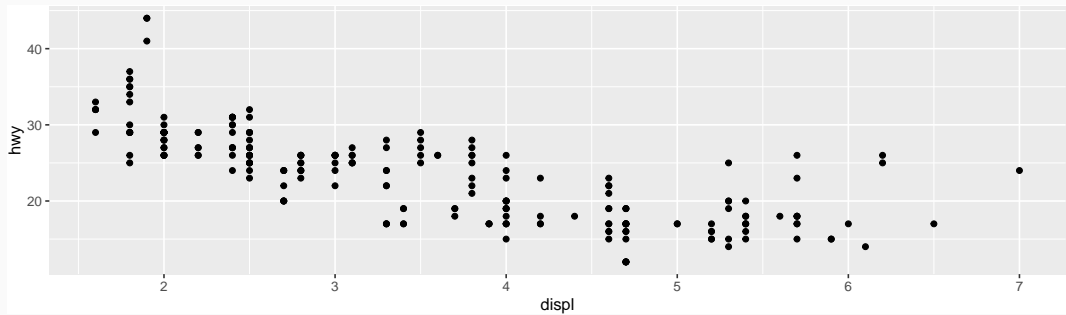
- Ordinary least squares regression (OLS) is probably the most widely-used model in political science

- Ordinary least squares regression (OLS) is probably the most widely-used model in political science
- At its core, it's all about drawing a line through data

- Ordinary least squares regression (OLS) is probably the most widely-used model in political science
- At its core, it's all about drawing a line through data
- This allows us to assess the effect of x on y

- Ordinary least squares regression (OLS) is probably the most widely-used model in political science
- At its core, it's all about drawing a line through data
- This allows us to assess the effect of x on y
- Dependent variable y must be continuous

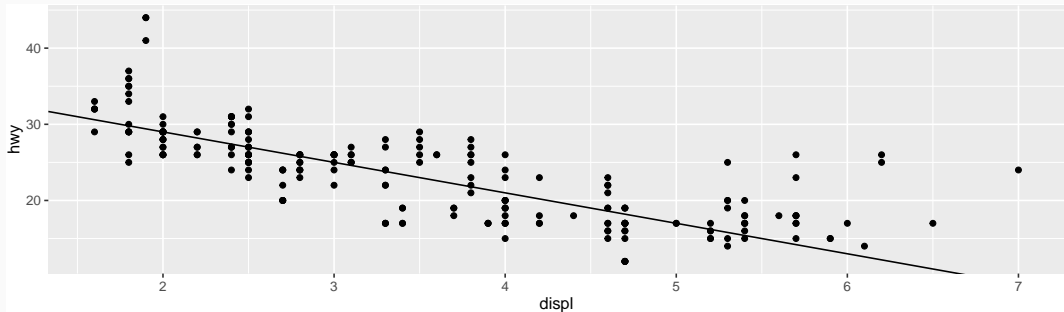
- Ordinary least squares regression (OLS) is probably the most widely-used model in political science
- At its core, it's all about drawing a line through data
- This allows us to assess the effect of x on y
- Dependent variable y must be continuous
 - OLS makes other assumptions you'll learn about in stats II



HOW TO DECIDE ON A LINE?

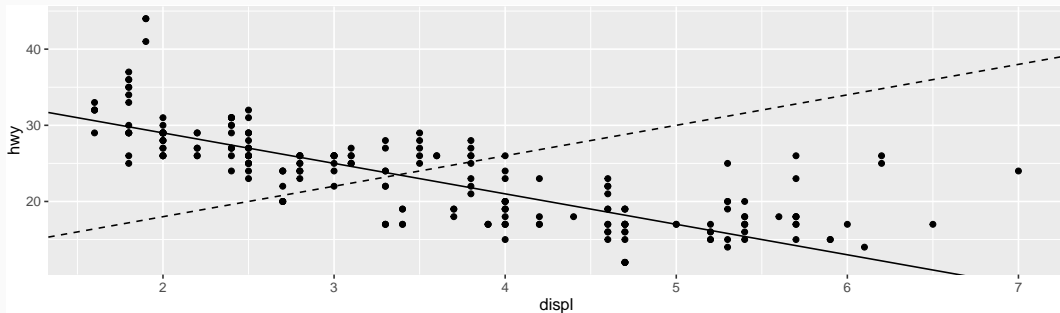
```
p <- p + geom_abline(slope = -4, intercept = 37)
```

p



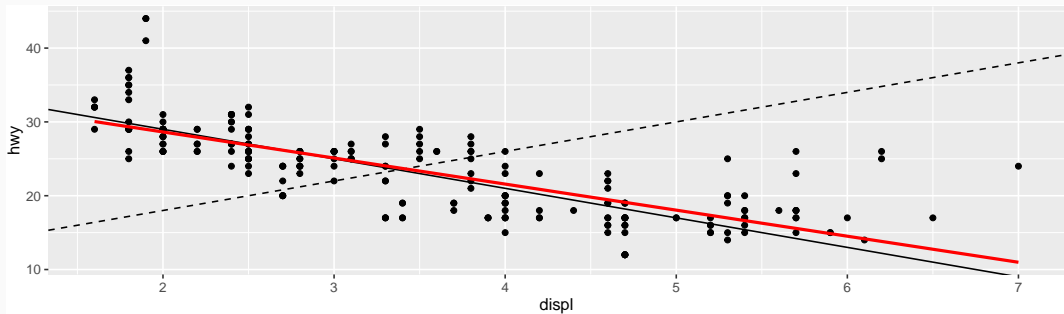
HOW TO DECIDE

```
p <- p + geom_abline(slope = 4, intercept = 10, linetype = "dashed")  
p
```



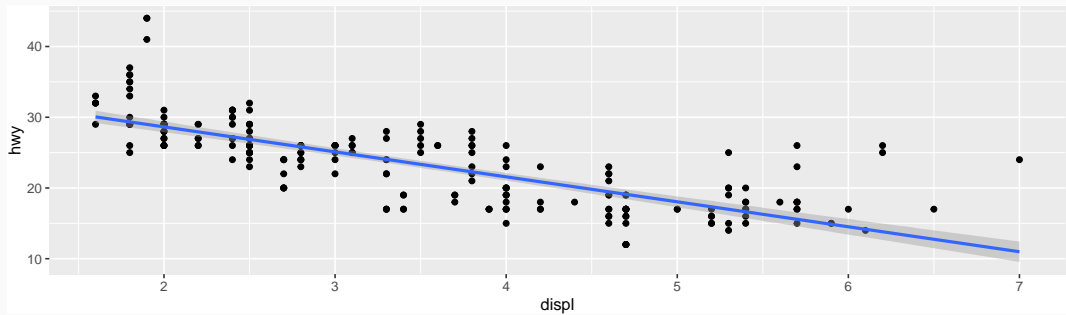
HOW TO DECIDE

```
p + geom_smooth(method = "lm", se = FALSE, color = "red")
```



```
lm(hwy ~ displ, data = mpg)

##
## Call:
## lm(formula = hwy ~ displ, data = mpg)
##
## Coefficients:
## (Intercept)      displ
##      35.698      -3.531
```



$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- Intercept ($\hat{\beta}_0$) - predicted y when $x = 0$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- Intercept ($\hat{\beta}_0$) - predicted y when $x = 0$
- Slope ($\hat{\beta}_1$) - a one unit change in x leads to a (slope) unit change in y , on average

- Notice that our line doesn't fit our data perfectly - we always make some error with our prediction

- Notice that our line doesn't fit our data perfectly - we always make some error with our prediction
- That's referred to as the *residual*

- Notice that our line doesn't fit our data perfectly - we always make some error with our prediction
- That's referred to as the *residual*
- If we refer to our predicted value as \hat{y} , then we can calculate the residual for each observation with $e_i = y_i - \hat{y}_i$

- OLS determines the “best” line by minimizing the sum of squared residuals

- OLS determines the “best” line by minimizing the sum of squared residuals
- How to find this?

- OLS determines the “best” line by minimizing the sum of squared residuals
- How to find this?
- One option: Plug in all the values for the slope & intercept and calculate the sum of squared residuals for these infinity combinations

- OLS determines the “best” line by minimizing the sum of squared residuals
- How to find this?
- One option: Plug in all the values for the slope & intercept and calculate the sum of squared residuals for these infinity combinations
- That's problematic...

How do we find the minimum sum of squared residuals?

Turns out we already know the solution - we learned it when we talked about *optimization*. We just need to *minimize* the sum of squared residuals with respect to the two coefficients:

$$\sum_{i=1}^n e_i^2$$

Rearrange above equation in terms of e_i :

$$e_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Rearrange above equation in terms of e_i :

$$e_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Substitute:

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

To find the minimum, we'll need to take the derivative with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$. Starting with $\hat{\beta}_0$:

To find the minimum, we'll need to take the derivative with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$. Starting with $\hat{\beta}_0$:

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

To find the minimum, we'll need to take the derivative with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$. Starting with $\hat{\beta}_0$:

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

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

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

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

The next step is to use the **chain rule** to take the derivative of the quantity in parentheses:

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

The next step is to use the **chain rule** to take the derivative of the quantity in parentheses:

$$\sum_{i=1}^n [-2(y_i - \hat{\beta}_0 - b_1 x_i)]$$

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

The next step is to use the **chain rule** to take the derivative of the quantity in parentheses:

$$\sum_{i=1}^n [-2(y_i - \hat{\beta}_0 - b_1 x_i)]$$

$$-2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - b_1 x_i)$$

Now let's take the partial derivative with respect to the slope ($\hat{\beta}_1$):

Now let's take the partial derivative with respect to the slope ($\hat{\beta}_1$):

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

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

Now let's take the partial derivative with respect to the slope ($\hat{\beta}_1$):

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

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

Using the chain rule again, we get:

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

How to find the *minimum* now that we have the partial derivatives of the sum of squared residuals?

How to find the *minimum* now that we have the partial derivatives of the sum of squared residuals?

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

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

How to find the *minimum* now that we have the partial derivatives of the sum of squared residuals?

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

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

((Solutions on board))

MULTIPLE VARIABLES

- What if as years pass, engine size increases *and* fuel efficiency increases?

MULTIPLE VARIABLES

- What if as years pass, engine size increases *and* fuel efficiency increases?
- Then the relationship we just observed might be *spurious*

MULTIPLE VARIABLES

- What if as years pass, engine size increases *and* fuel efficiency increases?
- Then the relationship we just observed might be *spurious*

MULTIPLE VARIABLES

- What if as years pass, engine size increases *and* fuel efficiency increases?
- Then the relationship we just observed might be *spurious*

```
lm(hwy ~ displ + year, data = mpg)
```

```
##
```

```
## Call:
```

```
## lm(formula = hwy ~ displ + year, data = mpg)
```

```
##
```

```
## Coefficients:
```

```
## (Intercept)      displ        year
```

```
##   -276.1544    -3.6110     0.1558
```

OLS IN MATRIX FORM

- Let's pretend that we know the **true** model

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector
 - X is $n \times k (+1)$ matrix

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector
 - X is $n \times k (+1)$ matrix
 - β is $k \times 1$ column vector

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector
 - X is $n \times k (+1)$ matrix
 - β is $k \times 1$ column vector
 - E is $n \times 1$ column vector

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector
 - X is $n \times k (+1)$ matrix
 - β is $k \times 1$ column vector
 - E is $n \times 1$ column vector
- Therefore, we have:

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector
 - X is $n \times k (+1)$ matrix
 - β is $k \times 1$ column vector
 - E is $n \times 1$ column vector
- Therefore, we have:

NOTATION

- Let's pretend that we know the **true** model
 - Y is $n \times 1$ column vector
 - X is $n \times k(+1)$ matrix
 - β is $k \times 1$ column vector
 - E is $n \times 1$ column vector
- Therefore, we have:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ 1 & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$Y = X\beta + E$$

OLS MINIMIZES THE SUM OF SQUARED RESIDUALS

- How to do that in matrix form?

OLS MINIMIZES THE SUM OF SQUARED RESIDUALS

- How to do that in matrix form?
- First, what is sum of squared residuals?

OLS MINIMIZES THE SUM OF SQUARED RESIDUALS

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

OLS MINIMIZES THE SUM OF SQUARED RESIDUALS

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

- Sum of squared residuals:

$$E'E$$

OLS MINIMIZES THE SUM OF SQUARED RESIDUALS

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

- Sum of squared residuals:

$$E'E$$

- (show why on board)

OLS MINIMIZES THE SUM OF SQUARED RESIDUALS

- How to do that in matrix form?
- First, what is sum of squared residuals?
- The residuals:

$$E = Y - X\hat{\beta}$$

- Sum of squared residuals:

$$E'E$$

- (show why on board)
- Alternatively,

$$\begin{aligned} E'E &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

TO MINIMIZE THE SUM OF SQUARES, WE TAKE THE DERIVATIVE

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

TO MINIMIZE THE SUM OF SQUARES, WE TAKE THE DERIVATIVE

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

- The first derivative with respect to $\hat{\beta}$

$$\frac{\partial E'E}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

TO MINIMIZE THE SUM OF SQUARES, WE TAKE THE DERIVATIVE

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

- The first derivative with respect to $\hat{\beta}$

$$\frac{\partial E'E}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

- To check that this is a minimum, we check to make sure that the second derivative is positive

TO MINIMIZE THE SUM OF SQUARES, WE TAKE THE DERIVATIVE

- Remember:

$$E'E = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

- The first derivative with respect to $\hat{\beta}$

$$\frac{\partial E'E}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

- To check that this is a minimum, we check to make sure that the second derivative is positive
- The second derivative is $2X'X$, which is positive definite so long as X is full rank

SOLVE FOR THE ESTIMATOR

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

SOLVE FOR THE ESTIMATOR

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

SOLVE FOR THE ESTIMATOR

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

- Premultiply each side by $(X'X)^{-1}$

$$(X'X)^{-1}X'Y = (X'X)^{-1}(X'X)\hat{\beta}$$

SOLVE FOR THE ESTIMATOR

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

- Premultiply each side by $(X'X)^{-1}$

$$(X'X)^{-1}X'Y = (X'X)^{-1}(X'X)\hat{\beta}$$

- We know that $(X'X)^{-1}(X'X) = I$

$$(X'X)^{-1}X'Y = I\hat{\beta}$$

SOLVE FOR THE ESTIMATOR

- Here ya go:

$$-2X'Y + 2X'X\hat{\beta} = 0$$

- Move things around and divide by two:

$$X'Y = X'X\hat{\beta}$$

- Premultiply each side by $(X'X)^{-1}$

$$(X'X)^{-1}X'Y = (X'X)^{-1}(X'X)\hat{\beta}$$

- We know that $(X'X)^{-1}(X'X) = I$

$$(X'X)^{-1}X'Y = I\hat{\beta}$$

- And I is (kinda) like multiplying by 1 so :

$$(X'X)^{-1}X'Y = \hat{\beta}$$