Intro to Linear Models and OLS

UIUC Department of Political Science Math Camp 2021

Miles D. Williams

milesdw2@illinois.edu

Where we're going. . .

Where we're going. . .

1. What are linear models?

Where we're going. . .

1. What are linear models?
2. How do we estimate them?

Where we're going. . .

1. What are linear models?
2. How do we estimate them?
3. How do we make inferences with them?

Where we're going. . .

1. What are linear models?
2. How do we estimate them?
3. How do we make inferences with them?
4. How do we extend them?

Where we're going. . .

1. What are linear models?
2. How do we estimate them?
3. How do we make inferences with them?
4. How do we extend them?
5. How do we know our models are good?

Some points of order

It's time for some ***real talk***. . .

Some points of order

- **Models** are not **estimators**!

Some points of order

- **Models** are not **estimators**!
  - *Estimators are a rule for solving the parameters of a model.*

Some points of order

- **Models** are not **estimators**!
  - *Estimators are a rule for solving the parameters of a model.*
- Be careful when using terms like "dependent" and "independent" variable.

Some points of order

- **Models** are not **estimators**!
  - *Estimators are a rule for solving the parameters of a model.*
- Be careful when using terms like “dependent” and “independent” variable.
  - Sometimes it’s better to say *response* or *outcome*

Some points of order

- **Models** are not **estimators**!
  - *Estimators are a rule for solving the parameters of a model.*
- Be careful when using terms like “dependent” and “independent” variable.
  - Sometimes it’s better to say *response* or *outcome*
  - Or *predictor* or *explanatory variable*

What is a linear model?

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

What is a linear model?

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

- It's a **statistical model** of the relationship between a set of ***explanatory variables*** and an ***outcome***.

What is a linear model?

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

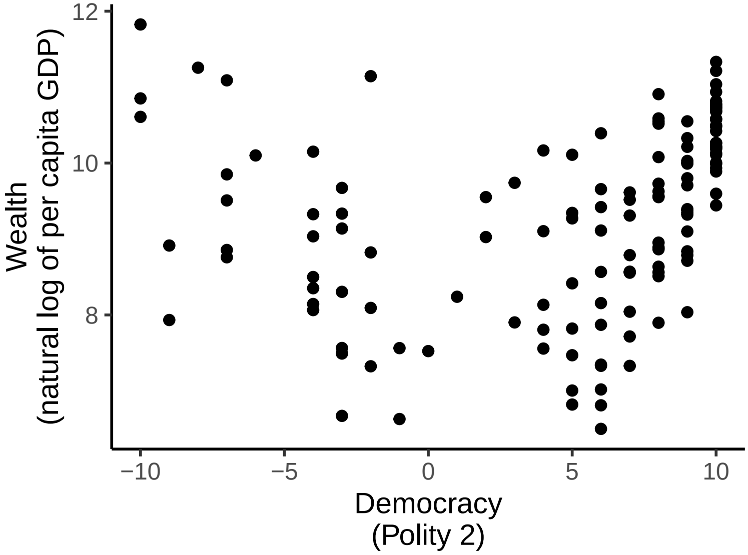
- It's a **statistical model** of the relationship between a set of ***explanatory variables*** and an ***outcome***.
- It models this relationship as an ***additive linear equation***.

An example using real-world data

QUESTION: What is the relationship between democracy and wealth?

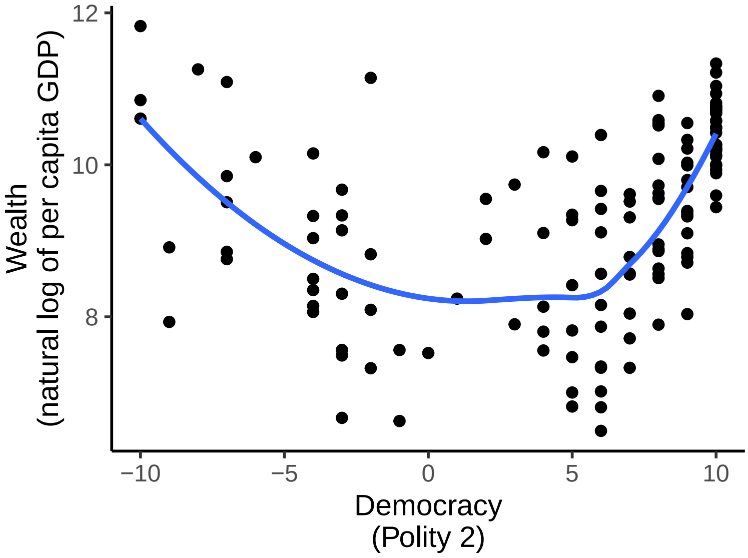
An example using real-world data

We have some data on democracy and wealth:



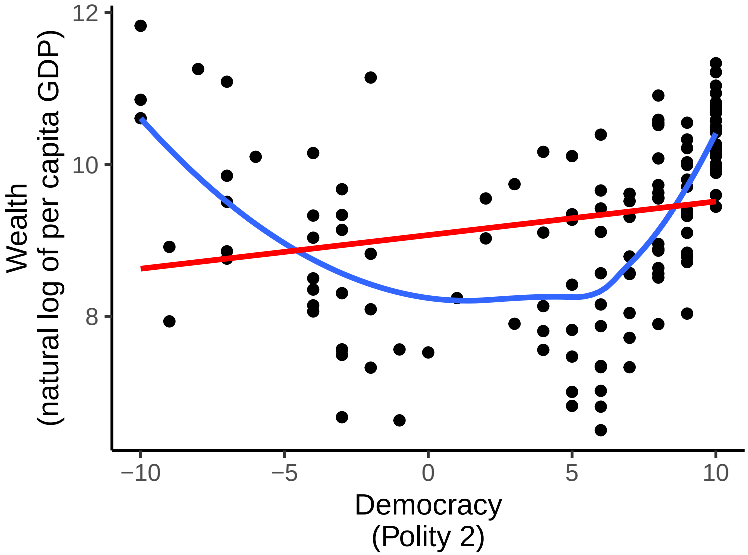
An example using real-world data

Do we model it like this?



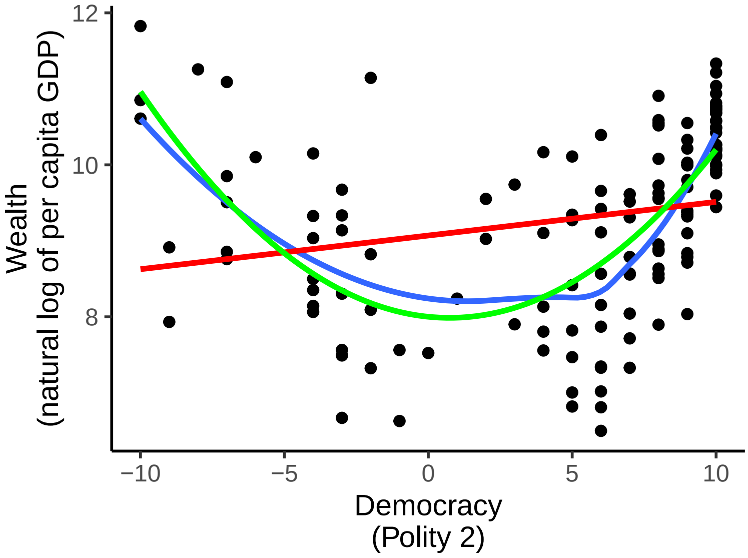
An example using real-world data

Or like this?



An example using real-world data

Or like this?!?!



An example using real-world data

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

- We can model the relationship as a linear equation.

An example using real-world data

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

- We can model the relationship as a linear equation.
- β_0 is average wealth when polity 2 is zero.

An example using real-world data

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

- We can model the relationship as a linear equation.
- β_0 is average wealth when polity 2 is zero.
- β_1 is the rate of change in wealth as polity 2 increase.

An example using real-world data

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

- We can model the relationship as a linear equation.
- β_0 is average wealth when polity 2 is zero.
- β_1 is the rate of change in wealth as polity 2 increase.
- ϵ_i is unexplained variation in wealth.

An example using real-world data

- We observe wealth and democracy...

An example using real-world data

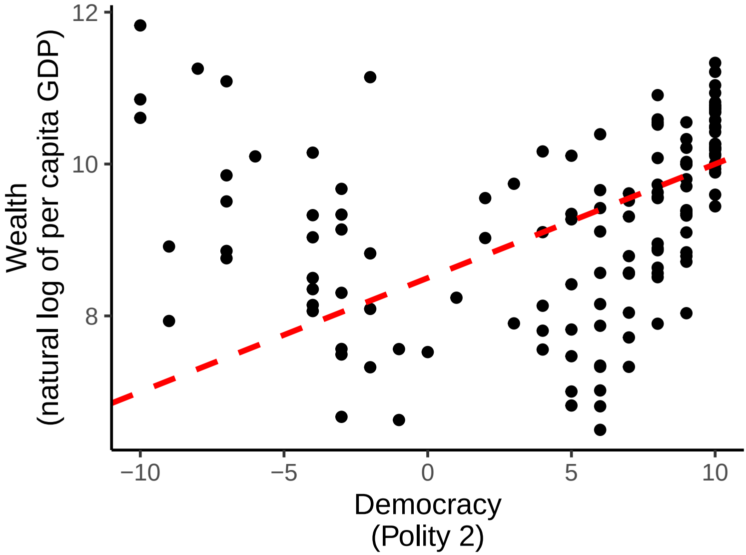
- We observe wealth and democracy...
- ...but not β_0, β_1, and ϵ_i.

An example using real-world data

- We observe wealth and democracy…
- …but not β_0, β_1, and ϵ_i.
- We need a way to select values of each.

An example using real-world data

My best guess (MBG): $\beta_0 = 8.5$ and $\beta_1 = 0.15$



An example using real-world data

- My best guess seem to do okay. . .

An example using real-world data

- My best guess seem to do okay…
- …but it's not very scientific.

An example using real-world data

- My best guess seem to do okay...
- ...but it's not very scientific.
- The values I chose for β_0 and β_1 look good to me.

An example using real-world data

- My best guess seem to do okay...
- ...but it's not very scientific.
- The values I chose for β_0 and β_1 look good to me.
- But ***you*** may have chosen differently.

An example using real-world data

- My best guess seem to do okay...
- ...but it's not very scientific.
- The values I chose for β_0 and β_1 look good to me.
- But ***you*** may have chosen differently.
- This solution is really subjective.

Enter ordinary least squares

- We most often use an estimator called “ordinary least squares” (OLS)

Enter ordinary least squares

- We most often use an estimator called “ordinary least squares” (OLS)
- OLS is a ***rule*** or ***criterion*** for selecting values of the unknown parameters.

Enter ordinary least squares

- We most often use an estimator called “ordinary least squares” (OLS)
- OLS is a ***rule*** or ***criterion*** for selecting values of the unknown parameters.
- What is this rule?

Enter ordinary least squares

OLS finds the βs that minimize the ***sum of the squared residuals*** (SSR):

$$\mathsf{SSR} = \sum_i \hat{\epsilon}_i^2$$

Enter ordinary least squares

The residuals are just the observed difference between our ***prediction*** for the response and the ***observed*** value of the response.

Enter ordinary least squares

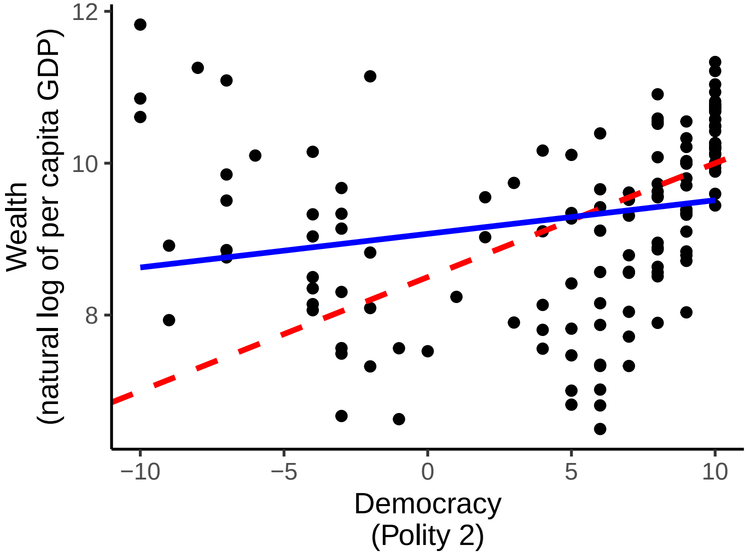
$\hat{\epsilon}_i = \text{wealth}_i - \widehat{\text{wealth}}_i$

$\widehat{\text{wealth}}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{democracy}_i$

$\hat{\beta}_0 =$ the selected value for the intercept.

$\hat{\beta}_1 =$ the selected value for the slope.

Enter ordinary least squares



$\hat{\beta}_0 = 9.07$ and $\hat{\beta}_1 = 0.044$

OLS is not the only solution. . . .

OLS is not the only solution. . . .

OLS is not the only solution. . . .

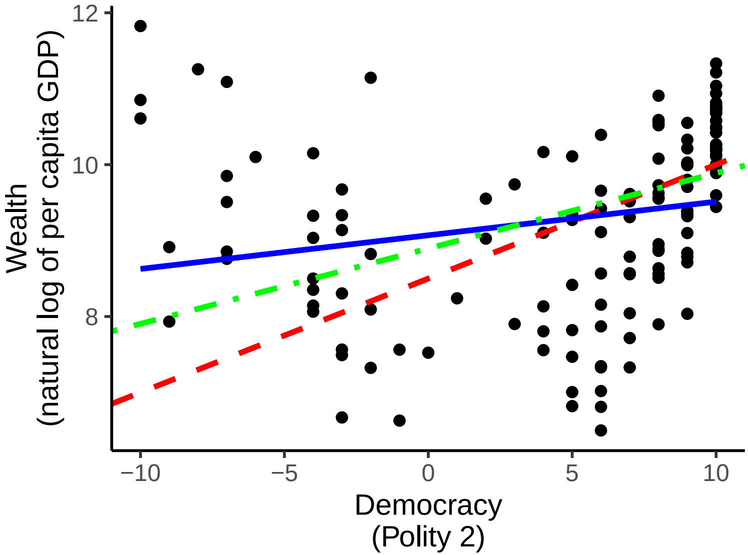
One alternative is ***least absolute deviations*** (LAD).

OLS is not the only solution. . . .

LAD finds the βs that minimize the ***sum of the absolute values of the residuals*** (SAVR).

$$\text{SAVR} = \sum_i |\hat{\epsilon}_i|$$

OLS is not the only solution. . . .



OLS is not the only solution. . . .

- We don't often use LAD

OLS is not the only solution. . . .

- We don't often use LAD
- It's robust to outliers

OLS is not the only solution. . . .

- We don't often use LAD
- It's robust to outliers
- But it doesn't always have a unique solution

OLS is not the only solution. . . .

- We don't often use LAD
- It's robust to outliers
- But it doesn't always have a unique solution
- We can't find its solution with an equation. . .

OLS is not the only solution. . . .

- We don't often use LAD
- It's robust to outliers
- But it doesn't always have a unique solution
- We can't find its solution with an equation. . .
- . . . which we ***can*** do with OLS.

Taking stock

Taking stock

1. What are linear models?

Taking stock

1. What are linear models?

- **Answer**: They are ***statistical models*** of the relationship between a set of explanatory variables and an outcome. They model this relationship as a ***linear additive equation***.

Taking stock

1. What are linear models?

- **Answer**: They are ***statistical models*** of the relationship between a set of explanatory variables and an outcome. They model this relationship as a ***linear additive equation***.

2. How do we estimate them?

Taking stock

1. What are linear models?

- **Answer**: They are ***statistical models*** of the relationship between a set of explanatory variables and an outcome. They model this relationship as a ***linear additive equation***.

2. How do we estimate them?

- **Answer**: We use ***estimators***, which are rules or criteria for selecting the values of unknown model parameters. Most often, we use an estimator called ***ordinary least squares*** (OLS).

More on ordinary least squares

More on ordinary least squares

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

More on ordinary least squares

$$\overbrace{\text{wealth}_i}^{\textit{Outcome}} = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

More on ordinary least squares

$$\overbrace{\text{wealth}_i}^{\textit{Outcome}} = \underbrace{\beta_0}_{\textit{Intercept}} + \beta_1 \text{democracy}_i + \epsilon_i$$

More on ordinary least squares

$$\overbrace{\text{wealth}_i}^{\textit{Outcome}} = \underbrace{\beta_0}_{\textit{Intercept}} + \overbrace{\beta_1}^{\textit{Slope}} \text{democracy}_i + \epsilon_i$$

More on ordinary least squares

$$\overbrace{\text{wealth}_i}^{\textit{Outcome}} = \underbrace{\beta_0}_{\textit{Intercept}} + \overbrace{\beta_1}^{\textit{Slope}} \underbrace{\text{democracy}_i}_{\textit{Predictor}} + \epsilon_i$$

More on ordinary least squares

$$\overbrace{\text{wealth}_i}^{\textit{Outcome}} = \underbrace{\beta_0}_{\textit{Intercept}} + \overbrace{\beta_1}^{\textit{Slope}} \underbrace{\text{democracy}_i}_{\textit{Predictor}} + \overbrace{\epsilon_i}^{\textit{Error}}$$

More on ordinary least squares

OLS finds the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize $\sum_i \hat{\epsilon}_i^2$.

More on ordinary least squares

Calculus to the rescue!!!

More on ordinary least squares

$$\begin{aligned}\hat{\beta}_1 &= \frac{cov(\text{democracy}_i,\ \text{wealth}_i)}{var(\text{democracy}_i)} \\ &= \frac{\sum_i(\text{democracy}_i - mean[\text{democracy}_i])(\text{wealth}_i - mean[\text{wealth}_i])}{\sum_i(\text{democracy}_i - mean[\text{democracy}_i])^2}\end{aligned}$$

More on ordinary least squares

$$\hat{\beta}_0 = \textit{mean}[\text{wealth}_i] - \hat{\beta}_1 \textit{mean}[\text{democracy}_i].$$

More on ordinary least squares

- The use of *mean*$[\cdot]$ should alert you to the fact that...

More on ordinary least squares

- The use of *mean*$[\cdot]$ should alert you to the fact that...
- ...our solution is a way of modeling the ***conditional mean*** of **wealth** given **democracy**.

More on ordinary least squares

- The use of *mean*$[\cdot]$ should alert you to the fact that...
- ...our solution is a way of modeling the ***conditional mean*** of **wealth** given **democracy**.
- *mean*$[\text{wealth}_i | \text{ democracy}_i]$.

More on ordinary least squares

- The use of *mean*$[\cdot]$ should alert you to the fact that...
- ...our solution is a way of modeling the ***conditional mean*** of **wealth** given **democracy**.
- *mean*[wealth$_i$| democracy$_i$].
- More generally, we express this as $\mathrm{E}(y_i|x_{ik})$.

More on ordinary least squares

- The use of *mean*$[\cdot]$ should alert you to the fact that...
- ...our solution is a way of modeling the ***conditional mean*** of **wealth** given **democracy**.
- *mean*[$\text{wealth}_i|$ democracy_i].
- More generally, we express this as $\mathrm{E}(y_i|x_{ik})$.
- That is, the ***expected value*** some response given the values of a set of predictors.

More on ordinary least squares

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

More on ordinary least squares

$$\begin{bmatrix} \text{wealth}_1 \\ \vdots \\ \text{wealth}_n \end{bmatrix} = \begin{bmatrix} 1 & \text{democracy}_1 \\ \vdots & \vdots \\ 1 & \text{democracy}_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

More on ordinary least squares

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$

More on ordinary least squares

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

More on ordinary least squares

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

More on ordinary least squares

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

OLS assumptions

What do we need to ensure OLS estimates are consistent?

OLS assumptions

1. The data-generating process underlying the observed data is additive and linear.

OLS assumptions

1. The data-generating process underlying the observed data is additive and linear.

- If an additive linear equation is not the best way to characterize our data, our model is ***misspecified***, and OLS estimates will not be reliable.

OLS assumptions

1. The data-generating process underlying the observed data is additive and linear.

- If an additive linear equation is not the best way to characterize our data, our model is ***misspecified***, and OLS estimates will not be reliable.

2. The explanatory variables are exogenous.

OLS assumptions

1. The data-generating process underlying the observed data is additive and linear.

- If an additive linear equation is not the best way to characterize our data, our model is ***misspecified***, and OLS estimates will not be reliable.

2. The explanatory variables are exogenous.

- If if there is some unobserved confounding variable that influences both the response and the predictors, OLS estimates may be biased.

OLS assumptions

- These are fundamentally unverifiable assumptions.

OLS assumptions

- These are fundamentally unverifiable assumptions.
- But, we can compare models to assess which is a ***better*** fit for the data.

OLS assumptions

- These are fundamentally unverifiable assumptions.
- But, we can compare models to assess which is a ***better*** fit for the data.
- And, we can rely on theory to make reasonable judgments about whether ***endogeneity*** (the opposite of exogeneity) is a problem.

OLS and statistical inference

So, now we have a linear model of wealth, and we know how to estimate it with OLS...

OLS and statistical inference

. . . we're done right?

OLS and statistical inference

WRONG!!!!!

OLS and statistical inference

- As ***scientists*** we're interested in testing hypotheses.

OLS and statistical inference

- As ***scientists*** we're interested in testing hypotheses.
- So, we're usually interested, not only in estimating linear models.

OLS and statistical inference

- As ***scientists*** we're interested in testing hypotheses.
- So, we're usually interested, not only in estimating linear models.
- We're also interested in making inferences about the estimated parameters of these models.

OLS and statistical inference

Does democracy predict greater wealth?

OLS and statistical inference

How do we know?

OLS and statistical inference

Enter standard errors!

OLS and statistical inference

While we use $\hat{\beta}$s to make predictions, we rely on $var(\hat{\beta})$s to know how ***precisely*** our $\hat{\beta}$s have been estimated.

OLS and statistical inference

We use the variance of the parameters to construct a number of quantities:

- Standard errors: $se(\hat{\beta}) = \sqrt{var(\hat{\beta})}$

OLS and statistical inference

We use the variance of the parameters to construct a number of quantities:

- Standard errors: $se(\hat{\beta}) = \sqrt{var(\hat{\beta})}$
- Test statistics: $t = \frac{\hat{\beta}}{se(\hat{\beta})}$

OLS and statistical inference

We use the variance of the parameters to construct a number of quantities:

- Standard errors: $se(\hat{\beta}) = \sqrt{var(\hat{\beta})}$
- Test statistics: $t = \frac{\hat{\beta}}{se(\hat{\beta})}$
- Confidence intervals: $\hat{\beta} \pm 1.96 \times se(\hat{\beta})$

OLS and statistical inference

We use the variance of the parameters to construct a number of quantities:

- Standard errors: $se(\hat{\beta}) = \sqrt{var(\hat{\beta})}$
- Test statistics: $t = \frac{\hat{\beta}}{se(\hat{\beta})}$
- Confidence intervals: $\hat{\beta} \pm 1.96 \times se(\hat{\beta})$
- p-values: $p = \Pr(|t| \geq t^*)$

OLS and statistical inference

So how do we calculate standard errors?

OLS and statistical inference

The equation for classical OLS standard errors is pretty simple.

OLS and statistical inference

$$\mathbf{V} = \begin{bmatrix} var(\hat{\beta}_0) & \cdots & cov(\hat{\beta}_0, \hat{\beta}_k) \\ \vdots & \ddots & \vdots \\ cov(\hat{\beta}_0, \hat{\beta}_k) & \cdots & var(\hat{\beta}_k) \end{bmatrix} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k}(\mathbf{X}'\mathbf{X})^{-1}$$

OLS and statistical inference

$$\begin{bmatrix} se(\hat{\beta}_0) & \cdots & se(\hat{\beta}_k) \end{bmatrix} = \sqrt{\text{diag}[\mathbf{V}]}$$

OLS and statistical inference

$$var(\hat{\beta}_0) = \frac{\sum_i \hat{\epsilon}_i}{n-k} \times \left[\frac{1}{n} + \frac{mean(\text{democracy}_i)}{n \times var(\text{democracy}_i)}\right]$$

$$var(\hat{\beta}_1) = \frac{\sum_i \hat{\epsilon}_i}{n-k} \times \left[\frac{1}{n \times var(\text{democracy}_i)}\right]$$

OLS and statistical inference

Simple enough, ***right***?

OLS and statistical inference

But we have a problem. . .

OLS and statistical inference

This solution for standard errors makes some **strong assumptions**. Namely. . .

OLS and statistical inference

1. That the errors (ϵ_i) of the model are ***identically*** and...

OLS and statistical inference

1. That the errors (ϵ_i) of the model are ***identically*** and...
2. ***independently*** distributed.

OLS and statistical inference

1. That the errors (ϵ_i) of the model are ***identically*** and...
2. ***independently*** distributed.

- (iid)

OLS and statistical inference

- These assumptions are quite heroic.

OLS and statistical inference

- These assumptions are quite heroic.
- The good news is that we don't have to make them.

OLS and statistical inference

- These assumptions are quite heroic.
- The good news is that we don't have to make them.
- We can calculate ***robust*** standard errors to avoid the identically distributed assumption.

OLS and statistical inference

- These assumptions are quite heroic.
- The good news is that we don't have to make them.
- We can calculate ***robust*** standard errors to avoid the identically distributed assumption.
- And we can ***cluster*** our standard errors if we think some observations are depedent.

OLS and statistical inference

- We refer to violations of the first assumption as ***heteroskedasticity*** (the opposite of ***homoskedasticity***).

OLS and statistical inference

- We refer to violations of the first assumption as ***heteroskedasticity*** (the opposite of ***homoskedasticity***).
- To deal with heteroskedasticity, we calculate a heteroskedasticity consistent (HC) variance-variance covariance matrix.

OLS and statistical inference

The solution for the HC0 variance-covariance matrix (the White estimator) is:

$$\Sigma = I \cdot \hat{\epsilon}\hat{\epsilon}'$$
$$\mathrm{HC}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

OLS and statistical inference

- This was named after Halbert White who published a paper about it in *Econometrica* in 1980.

OLS and statistical inference

- This was named after Halbert White who published a paper about it in *Econometrica* in 1980.
- This solution is robust to all kinds of heteroskedasticity.

OLS and statistical inference

- This was named after Halbert White who published a paper about it in *Econometrica* in 1980.
- This solution is robust to all kinds of heteroskedasticity.
- ***Which is great!***

OLS and statistical inference

- This was named after Halbert White who published a paper about it in *Econometrica* in 1980.
- This solution is robust to all kinds of heteroskedasticity.
- ***Which is great!***
- The reason is that it allows for individual-level variation in the residuals when computing parameter variances.

OLS and statistical inference

BUT

OLS and statistical inference

- HC0 errors suffer from finite sample bias.

OLS and statistical inference

- HC0 errors suffer from finite sample bias.
- For this reason, in practice we use a degrees of freedom multiplier to get us ***HC1 errors***.

OLS and statistical inference

This solution is just:

$$\mathsf{HC}_1 = \frac{n}{n-k} \times \mathsf{HC}_0$$

OLS and statistical inference

What about clustering?

OLS and statistical inference

- The clustered version of HC0 errors is called CR0 (CR = ***cluster robust***).

OLS and statistical inference

- The clustered version of HC0 errors is called CR0 (CR = ***cluster robust***).
- The solution for this is similar to that for HC0 errors.

OLS and statistical inference

- The clustered version of HC0 errors is called CR0 (CR = ***cluster robust***).
- The solution for this is similar to that for HC0 errors.
- The key difference is that the $\mathbf{X}'\Sigma\mathbf{X}$ part of the equation is set up by groups or clusters in the data, rather than by individual observations.

OLS and statistical inference

- The clustered version of HC0 errors is called CR0 (CR = ***cluster robust***).
- The solution for this is similar to that for HC0 errors.
- The key difference is that the $\mathbf{X}'\Sigma\mathbf{X}$ part of the equation is set up by groups or clusters in the data, rather than by individual observations.
- Like with HC0 errors, we need to use a degrees of freedom multiplier to account for finite sample bias.

OLS and statistical inference

- The clustered version of HC0 errors is called CR0 (CR = ***cluster robust***).
- The solution for this is similar to that for HC0 errors.
- The key difference is that the $\mathbf{X}'\Sigma\mathbf{X}$ part of the equation is set up by groups or clusters in the data, rather than by individual observations.
- Like with HC0 errors, we need to use a degrees of freedom multiplier to account for finite sample bias.
- We often need to account for clustering if observations occur within groups (like a classroom) or have repeated measures over time (like a panel time-series).

OLS and statistical inference

Let's get back to polity and wealth....

OLS and statistical inference

- In most applied settings, we'll use either HC1 or CR1 estimators of the variance-covariance matrix.

OLS and statistical inference

- In most applied settings, we'll use either HC1 or CR1 estimators of the variance-covariance matrix.
- For our model of wealth, we'll use HC1 errors.

OLS and statistical inference

- For the model intercept, the standard error is $se(\hat{\beta}_0) = 0.16$

OLS and statistical inference

- For the model intercept, the standard error is $se(\hat{\beta}_0) = 0.16$
- And for the estimated coefficient on democracy, $se(\hat{\beta}_1) = 0.02$

OLS and statistical inference

- With these standard errors, we next calculate a test statistic.

OLS and statistical inference

- With these standard errors, we next calculate a test statistic.
- This is the t-statistic.

OLS and statistical inference

- With these standard errors, we next calculate a test statistic.
- This is the t-statistic.
- $t = \hat{\beta}/se(\hat{\beta})$.

OLS and statistical inference

- With these standard errors, we next calculate a test statistic.
- This is the t-statistic.
- $t = \hat{\beta}/se(\hat{\beta})$.
- $t =$ *ratio* of **the parameter** to **it's standard error**.

OLS and statistical inference

- With these standard errors, we next calculate a test statistic.
- This is the t-statistic.
- $t = \hat{\beta}/se(\hat{\beta})$.
- $t =$ *ratio* of **the parameter** to **it's standard error**.
- Note that this means that the greater our precision in estimating $\hat{\beta}$, the larger our t-value will be.

OLS and statistical inference

- The nice thing about t is that is has a known theoretical distribution ***under the null hypothesis***.

OLS and statistical inference

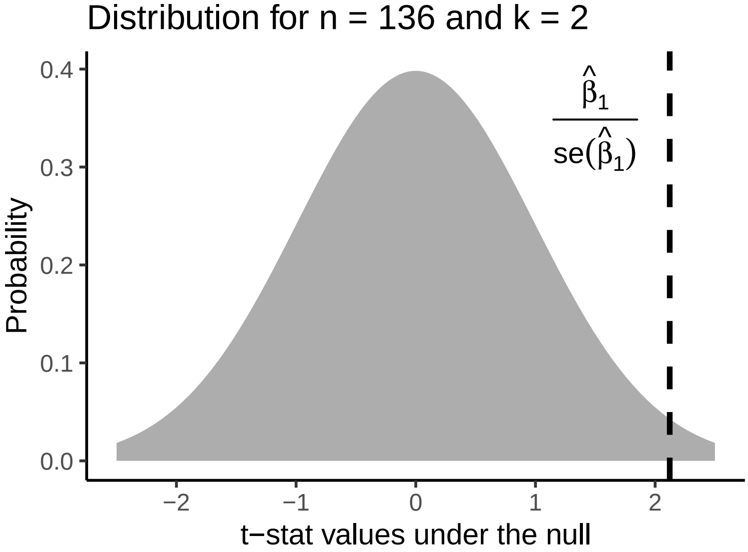
- The nice thing about t is that is has a known theoretical distribution ***under the null hypothesis***.
- We judge our t relative to the null to compute a p-value.

OLS and statistical inference

- The nice thing about t is that is has a known theoretical distribution ***under the null hypothesis***.
- We judge our t relative to the null to compute a p-value.
- This p-value tells us how surprised we should be to observe the t we calculated if the true linear relationship between a predictor and and outcome ***is zero***.

OLS and statistical infernece

Here's what this distribution looks like for $\hat{\beta}_1$...



OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.
- For a ***two-sided*** t-test this is:

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.
- For a ***two-sided*** t-test this is:
- $p = 1 - 2 \times F(|t|)$.

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.
- For a ***two-sided*** t-test this is:
- $p = 1 - 2 \times F(|t|)$.
- We ***reject the null hypothesis*** if $p \leq \alpha$.

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.
- For a ***two-sided*** t-test this is:
- $p = 1 - 2 \times F(|t|)$.
- We ***reject the null hypothesis*** if $p \leq \alpha$.
- The most common α-level we select is $\alpha = 0.05$.

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.
- For a ***two-sided*** t-test this is:
- $p = 1 - 2 \times F(|t|)$.
- We ***reject the null hypothesis*** if $p \leq \alpha$.
- The most common α-level we select is $\alpha = 0.05$.
- For democracy, our p-value is $p = 0.036$.

OLS and statistical inference

- It looks like our t-value for democracy is a pretty rare event if the true relationshp between democracy and wealth were zero.
- To quantify how rare it is, we need to calculate a p-value.
- For a ***two-sided*** t-test this is:
- $p = 1 - 2 \times F(|t|)$.
- We ***reject the null hypothesis*** if $p \leq \alpha$.
- The most common α-level we select is $\alpha = 0.05$.
- For democracy, our p-value is $p = 0.036$.
- That's less than our α-level, so we can reject the null hypothesis that the slope for democracy is zero.

OLS and statistical inference

We usually summarize all of this analysis with a ***regression table***.

OLS and statistical inference

Table 1: OLS Estimates with Robust S.E.s

	Model of Wealth
Constant	9.07 (0.16)***
Democracy (Polity 2)	0.04 (0.02)*
Num. obs.	136
RMSE	1.19

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Extensions of linear models

- We've covered a lot of ground.

Extensions of linear models

- We've covered a lot of ground.
- ***Don't worry***

Extensions of linear models

- We've covered a lot of ground.
- ***Don't worry***
- You'll have plenty of time get all of this straight

Extensions of linear models

But, before we wrap up I want to walk through a few more things.

Extensions of linear models

- The first is that linear models can accommodate ***nonlinear relationships***.

Extensions of linear models

- The first is that linear models can accommodate ***nonlinear relationships***.
- Take our model of wealth.

Extensions of linear models

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i$$

Extensions of linear models

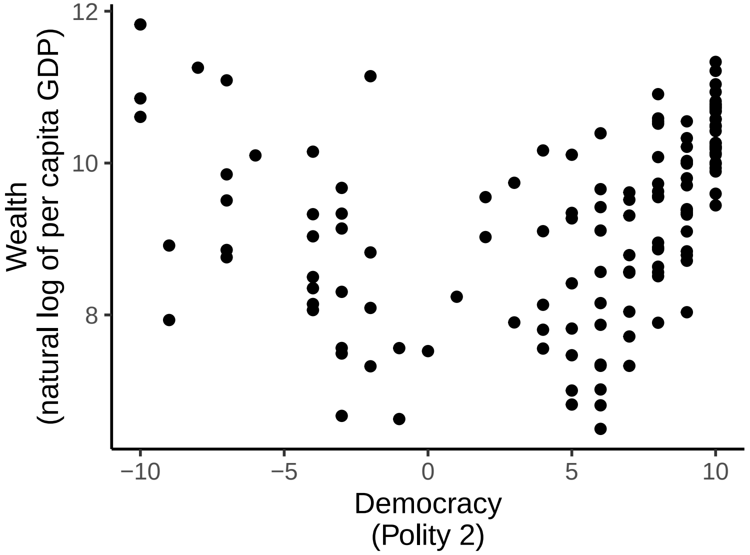
$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \text{democracy}_i^2 + \epsilon_i$$

Extensions of linear models

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \underbrace{\text{democracy}_i^2}_{\textit{Quadratic Term}} + \epsilon_i$$

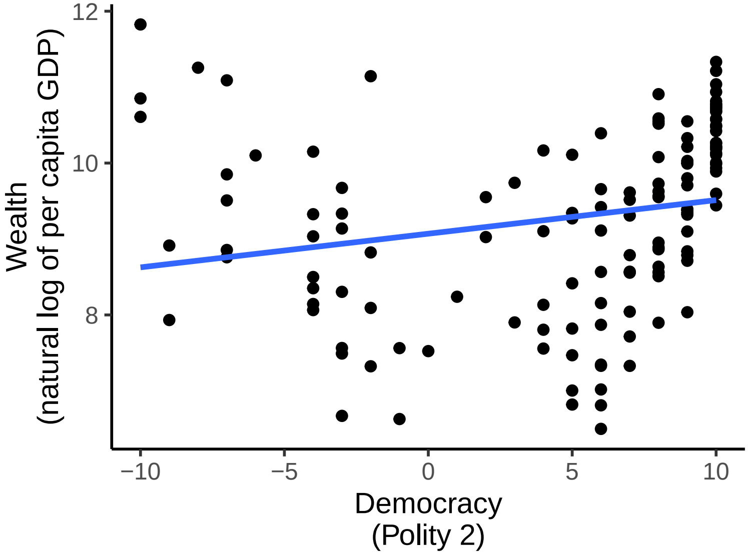
Extensions of linear models

Take our data on democracy and wealth:



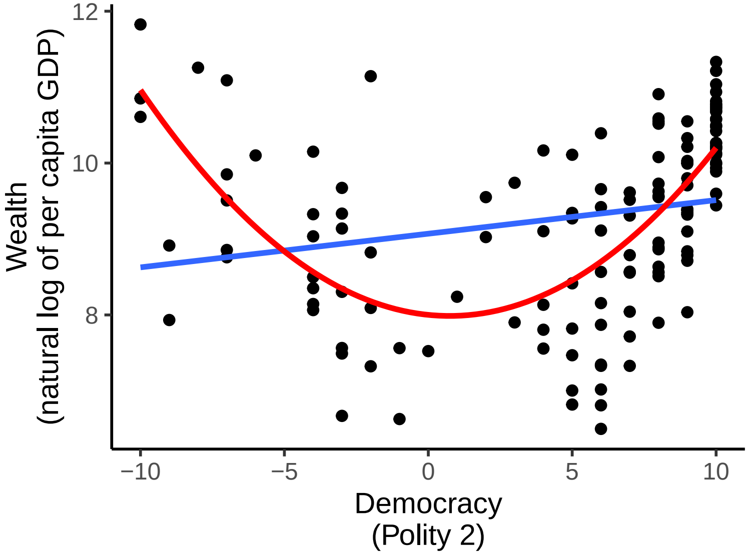
Extensions of linear models

We can model this relationship as a linear function....



Extensions of linear models

Or, we can model it as a quadratic function!



Multiple regression models

- Another extension of the basic linear model is a ***multiple regression model***.

Multiple regression models

- Another extension of the basic linear model is a ***multiple regression model***.
- These are models with multiple predictor variables on the right-hand side of the equation.

Multiple regression models

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \text{human capital}_i + \epsilon_i.$$

Multiple regression models

Our OLS estimate for β_1 will now reflect the ***residual linear relationship*** between democracy and wealth, ***after*** subtracting out variation in each explained as a linear function of ***human capital***.

Multiple regression models

- $\text{wealth}_i = \eta_0 + \eta_1 \text{human capital}_i + \upsilon_i$

Multiple regression models

- $\text{wealth}_i = \eta_0 + \eta_1 \text{human capital}_i + \upsilon_i$
- $\text{democracy}_i = \gamma_0 + \gamma_1 \text{human capital}_i + \varepsilon_i$

Multiple regression models

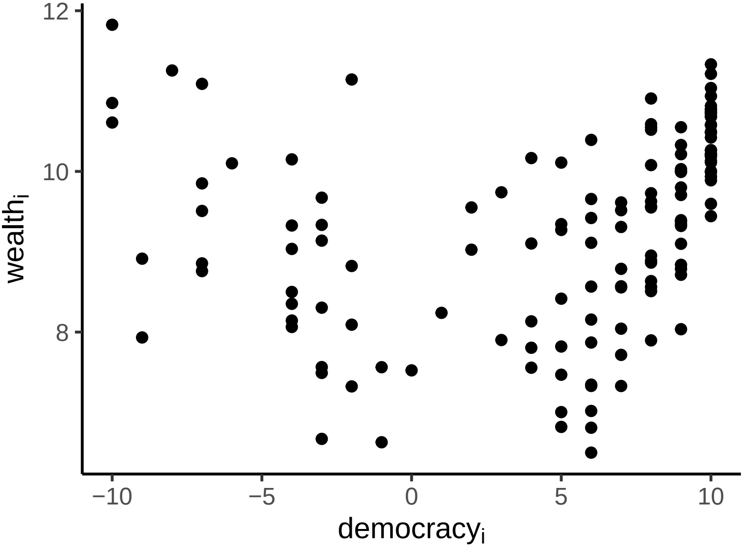
- $\text{wealth}_i = \eta_0 + \eta_1 \text{human capital}_i + \upsilon_i$
- $\text{democracy}_i = \gamma_0 + \gamma_1 \text{human capital}_i + \varepsilon_i$
- $\hat{\upsilon}_i = \text{wealth}_i - \widehat{\text{wealth}}_i$ ***and*** $\hat{\varepsilon}_i = \text{democracy}_i - \widehat{\text{democracy}}_i$

Multiple regression models

- $\text{wealth}_i = \eta_0 + \eta_1 \text{human capital}_i + \upsilon_i$
- $\text{democracy}_i = \gamma_0 + \gamma_1 \text{human capital}_i + \varepsilon_i$
- $\hat{\upsilon}_i = \text{wealth}_i - \widehat{\text{wealth}}_i$ ***and*** $\hat{\varepsilon}_i = \text{democracy}_i - \widehat{\text{democracy}}_i$
- $\hat{\upsilon}_i = \beta_0 + \beta_1 \hat{\varepsilon}_i + \epsilon_i$

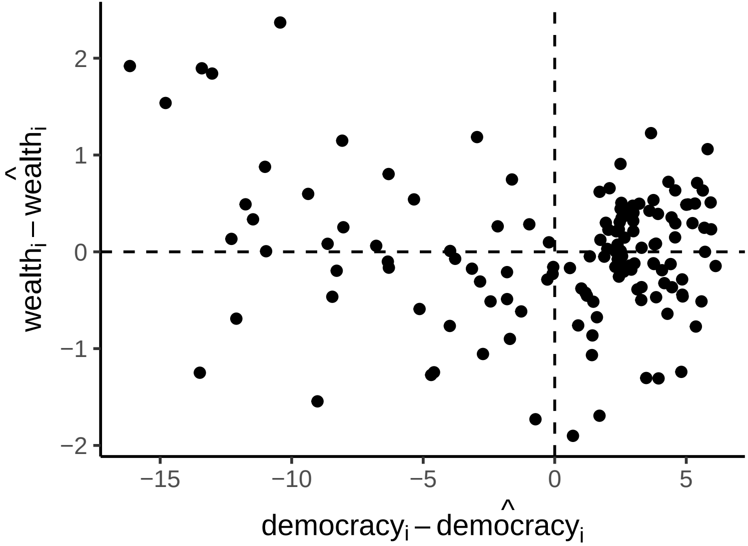
Multiple regression models

Here's our ***raw*** data:



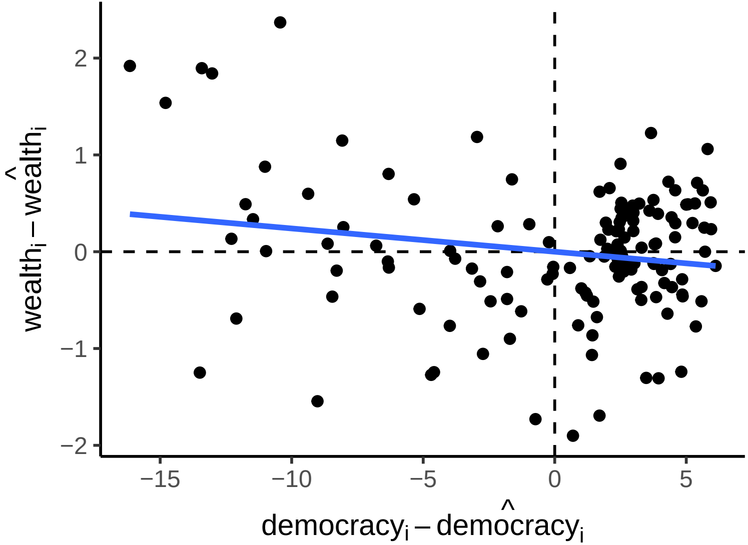
Multiple regression models

Here's what the ***residualized*** data looks like:



Multiple regression models

And this is now our estimated slope for democracy:



Multiple regression models

$$\begin{bmatrix} \text{wealth}_1 \\ \vdots \\ \text{wealth}_n \end{bmatrix} = \begin{bmatrix} 1 & \text{democracy}_1 & \text{human capital}_1 \\ \vdots & \vdots & \vdots \\ 1 & \text{democracy}_n & \text{human capital}_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

Multiple regression models

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$$

Multiple regression models

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Multiple regression models

Table 2: OLS Estimates with Robust S.E.s

	Model of Wealth
Constant	$5.50\ (0.21)^{***}$
Democracy (Polity 2)	$-0.02\ (0.02)$
Human Capital (HCI)	$1.47\ (0.09)^{***}$
Num. obs.	136
RMSE	0.70

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Goodness of fit

- We've covered a lot of ground so far.

Goodness of fit

- We've covered a lot of ground so far.
- Before we wrap up, I want to introduce one more thing...

Goodness of fit

- We've covered a lot of ground so far.
- Before we wrap up, I want to introduce one more thing...
- **Goodness of Fit** (GOF)

Goodness of fit

- These are metrics that tell us how ***good a fit*** a **model** is for the data.

Goodness of fit

- These are metrics that tell us how ***good a fit*** a **model** is for the data.
- Does the estimated model explain more variation in the response than we would expect it to by mere random chance?

Goodness of fit

- These are metrics that tell us how ***good a fit*** a **model** is for the data.
- Does the estimated model explain more variation in the response than we would expect it to by mere random chance?
- Is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ldots$

Goodness of fit

- These are metrics that tell us how ***good a fit*** a **model** is for the data.
- Does the estimated model explain more variation in the response than we would expect it to by mere random chance?
- Is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i \ldots$
- better than $y_i = \beta_0 + \epsilon_i$?

Goodness of fit

- One of the most common metrics we use is:

Goodness of fit

- One of the most common metrics we use is:
- $R^2 = 1 - \frac{SSR}{SST}$

Goodness of fit

- One of the most common metrics we use is:
- $R^2 = 1 - \frac{SSR}{SST}$
- adjusted-$R^2 = 1 - \frac{(n-1)SSR}{(n-k)SST}$

Goodness of fit

- One of the most common metrics we use is:
- $R^2 = 1 - \frac{SSR}{SST}$
- adjusted-$R^2 = 1 - \frac{(n-1)SSR}{(n-k)SST}$
- $SST = \sum_i (y_i - \bar{y})^2 = \sum_i (y_i - \hat{\beta}_0)$

Goodness of fit

- One of the most common metrics we use is:
- $R^2 = 1 - \frac{SSR}{SST}$
- adjusted-$R^2 = 1 - \frac{(n-1)SSR}{(n-k)SST}$
- $SST = \sum_i (y_i - \bar{y})^2 = \sum_i (y_i - \hat{\beta}_0)$
- ***The proportion variance explained in the response***

Goodness of fit

- We also use something called an F-statistic:

Goodness of fit

- We also use something called an F-statistic:
- $F = \frac{SST - SSR}{k-1} / \frac{SSR}{n-k}$

Goodness of fit

- We also use something called an F-statistic:
- $F = \frac{SST - SSR}{k-1} / \frac{SSR}{n-k}$
- This tells us how different our ***full model*** is from our ***reduced model***.

Goodness of fit

- We also use something called an F-statistic:
- $F = \frac{SST - SSR}{k-1} / \frac{SSR}{n-k}$
- This tells us how different our ***full model*** is from our ***reduced model***.
- It is proportional to the difference in the error between the two models.

Goodness of fit

- We also use something called an F-statistic:
- $F = \frac{SST - SSR}{k-1} / \frac{SSR}{n-k}$
- This tells us how different our ***full model*** is from our ***reduced model***.
- It is proportional to the difference in the error between the two models.
- The ***larger this value***, the better the full model does at reducing the unexplained variation in the response than the reduced model.

Goodness of fit

For our model of wealth:

$$\text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \text{human rights}_i + \epsilon_i$$

Goodness of fit

Our F-statistic lets us run an F-test:

$$\begin{aligned} \mathrm{H}_0 &: \beta_1 + \beta_2 = 0, \\ \mathrm{H}_A &: \beta_1 + \beta_2 \neq 0. \end{aligned}$$

Goodness of fit

- F-values, like t-values, have a known distribution under the ***null hypothesis***.

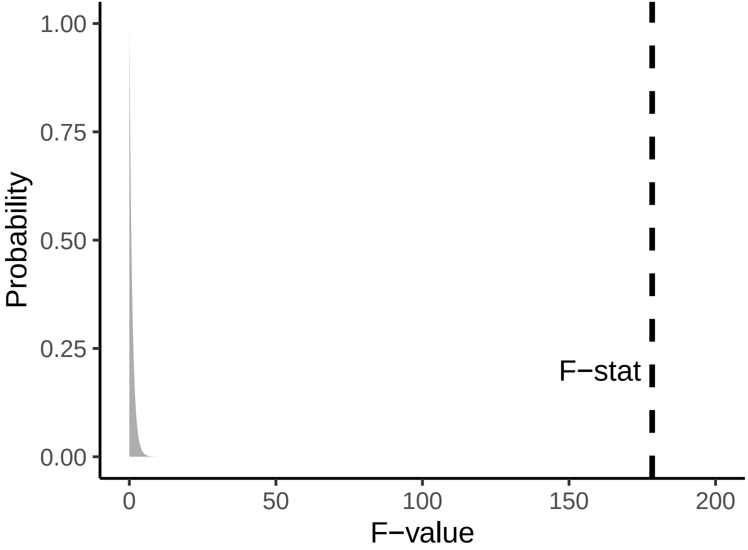
Goodness of fit

- F-values, like t-values, have a known distribution under the ***null hypothesis***.
- This means we can compute a p-value for the F-test.

Goodness of fit

- F-values, like t-values, have a known distribution under the ***null hypothesis***.
- This means we can compute a p-value for the F-test.
- As with the t-test, we use an α-level of $\alpha = 0.05$ as our threshold for rejecting the null hypothesis.

Goodness of fit



$p = 0$

Goodness of fit

So our model does pretty well.

Goodness of fit

BUT, is it the best?

Goodness of fit

We can use GOF to pick the best model!

Goodness of fit

- Note that I said "best" model.

Goodness of fit

- Note that I said “best” model.
- I did NOT say the “right” model.

Goodness of fit

- Note that I said “best” model.
- I did NOT say the “right” model.
- There is no way to prove that a model has been specified correctly.

Goodness of fit

- Note that I said “best” model.
- I did NOT say the “right” model.
- There is no way to prove that a model has been specified correctly.
- We can only judge whether it is a better fit for the data compared to an alternative.

Goodness of fit

$$
\begin{aligned}
(1) \quad & \text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \epsilon_i, \\
(2) \quad & \text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \text{human capital}_i + \epsilon_i, \\
(3) \quad & \text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \text{democracy}_i^2 + \epsilon_i, \\
(4) \quad & \text{wealth}_i = \beta_0 + \beta_1 \text{democracy}_i + \beta_2 \text{democracy}_i^2 \\
& \qquad\qquad + \beta_3 \text{human capital}_i + \epsilon_i.
\end{aligned}
$$

Goodness of fit

- Which of these is the best model of wealth?

Goodness of fit

- Which of these is the best model of wealth?
- We can use our GOF metrics to provide an informed answer to this question.

Goodness of fit

Table 3: OLS Estimates for Different Models of Wealth

	Model 1	Model 2	Model 3	Model 4
Contant	9.07^{***}	5.50^{***}	8.00^{***}	5.61^{***}
	(0.16)	(0.21)	(0.18)	(0.19)
Democracy (Polity 2)	0.04^{*}	-0.02	-0.04	-0.05^{**}
	(0.02)	(0.02)	(0.02)	(0.02)
Democracy^2			0.03^{***}	0.01^{***}
			(0.00)	(0.00)
Human Capital (HCI)		1.47^{***}		1.20^{***}
		(0.09)		(0.08)
Adj. R^2	0.04	0.67	0.41	0.74
Num. obs.	136	136	136	136
F statistic	4.50	178.37	48.15	154.24

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Goodness of fit

Table 4: Adjusted-R^2 for Different Models of Wealth

	Model 1	Model 2	Model 3	Model 4
Adj. R^2	0.04	0.67	0.41	**0.74**
Num. obs.	136	136	136	136

Goodness of fit

Table 5: Difference in Sum of Sq.

	Model 1	Model 2	Model 3	Model 4
Model 1	·	124.83*	74.79*	**140.05***
Model 2	·	·	-50.04	**15.22***
Model 3	·	·	·	**65.26***
Model 4	·	·	·	·

Goodness of fit

One the basis of these GOF metrics, it looks like the ***best*** fitting model of wealth models it as a quadratic function of polity and a linear function of human capital.

Wrapping up

So, in summary...

Wrapping up

- We've answered a number of questions:

Wrapping up

- We've answered a number of questions:

1. What are linear models?

Wrapping up

- We've answered a number of questions:

1. What are linear models?

- They are ***statistical models*** of the relationship between a set of ***explanatory variables*** an an ***outcome variable***.

Wrapping up

2. How do we estimate them?

Wrapping up

2. How do we estimate them?

- We use ***estimators*** to select appropriate values for the unknown parameters of a linear model.

Wrapping up

2. How do we estimate them?

- We use ***estimators*** to select appropriate values for the unknown parameters of a linear model.
- Specifically, we most often use a method called ***ordinary least squares*** (OLS).

Wrapping up

3. How do we make inferences with them?

Wrapping up

3. How do we make inferences with them?

- We can perform statistical inference with OLS estimates by calculating the variance of our model parameters.

Wrapping up

3. How do we make inferences with them?

- We can perform statistical inference with OLS estimates by calculating the variance of our model parameters.
- These capture the ***precision*** with which OLS has selected parameter values.

Wrapping up

3. How do we make inferences with them?

- We can perform statistical inference with OLS estimates by calculating the variance of our model parameters.
- These capture the ***precision*** with which OLS has selected parameter values.
- With these we calculate standard errors, t-values, and p-values.

Wrapping up

3. How do we make inferences with them?

- We can perform statistical inference with OLS estimates by calculating the variance of our model parameters.
- These capture the ***precision*** with which OLS has selected parameter values.
- With these we calculate standard errors, t-values, and p-values.
- We can also compute confidence intervals. . . but we'll save these for another time.

Wrapping up

4. How do we extend them?

Wrapping up

4. How do we extend them?

- Linear models can accommodate more complex relationships in data than linear relationships.

Wrapping up

4. How do we extend them?

- Linear models can accommodate more complex relationships in data than linear relationships.
- As long as the relationship can be expressed in additive terms with respect to model parameters, we can specify the model as a linear regression and estimate it with OLS.

Wrapping up

5. How do we know our models are good?

Wrapping up

5. How do we know our models are good?

- To judge the goodness of our models, we use ***goodness of fit*** metrics.

Wrapping up

5. How do we know our models are good?

- To judge the goodness of our models, we use ***goodness of fit*** metrics.
- These cannot prove whether our models are correct.

Wrapping up

5. How do we know our models are good?

- To judge the goodness of our models, we use ***goodness of fit*** metrics.
- These cannot prove whether our models are correct.
- But they can help us make an informed decision about whether a model is better than some alternative(s).

Wrapping up

5. How do we know our models are good?

- To judge the goodness of our models, we use ***goodness of fit*** metrics.
- These cannot prove whether our models are correct.
- But they can help us make an informed decision about whether a model is better than some alternative(s).
- We typically use the adjusted-R^2 and the F-value of a regression model to make this judgment.

Wrapping up

Now for some exercises!