Today's Agenda

Practice using OLS regressions to answer real world questions with data

Justin Leinaweaver (Spring 2022)

Our Work So Far

- Descriptive statistics (mean, median, IQR, etc.)
- Bar plots (categorical variables)
- Histograms (numerical variables)
- Facets (any viz on subsets of data)
- Box plots (numerical x categorical)
- Scatterplots (numerical x numerical)
- Fitting and evaluating simple OLS models

Which is a better model of personal income?

1. Mother's education level, or
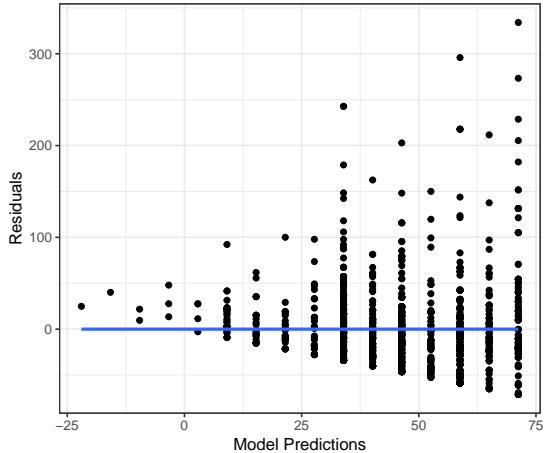
2. Personal exercise

	Earnings (2021)	
	(1)	(2)
Mother's Education	6.21*	
	(0.43)	
Exercise		1.88*
		(0.42)
Constant	−40.64*	36.69*
	(5.87)	(1.62)
Observations	1,570	1,815
Adjusted R^2	0.12	0.01
Residual Std. Error	40.45 (df = 1568)	41.75 (df = 1813)
F Statistic	212.09* (df = 1; 1568)	19.83* (df = 1; 1813)

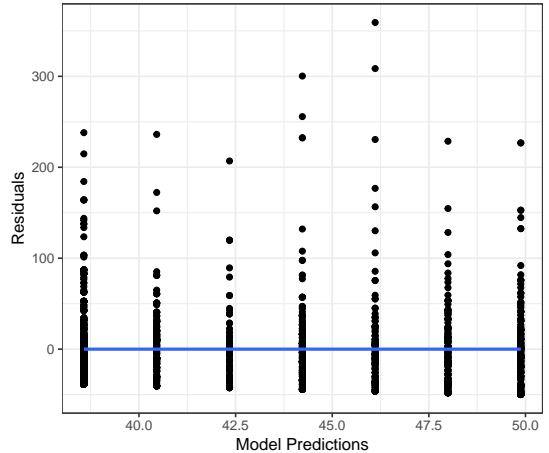*Note:*   *p < 0.05

Step 4 - Check the Residuals

Mother's Education Model

Strenuous Exercise Model

Make Predictions

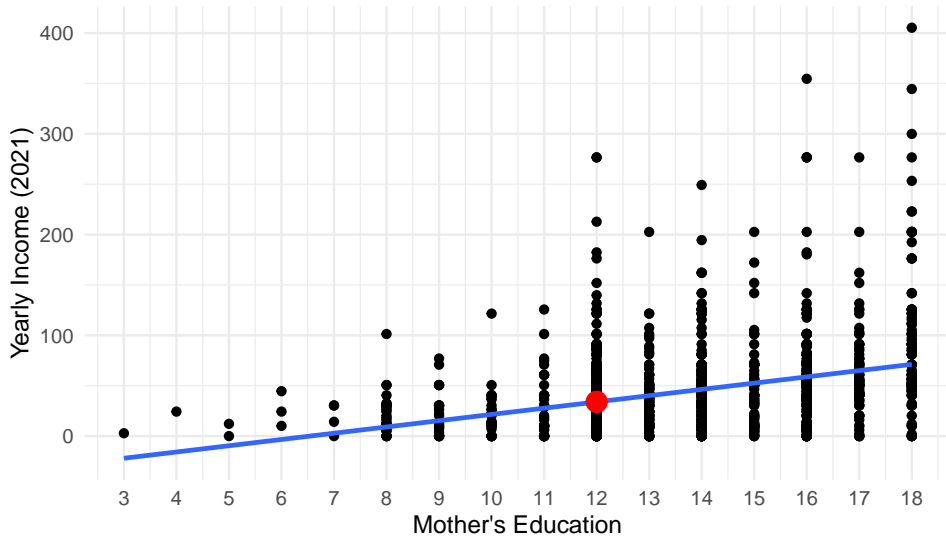| Prediction	Estimates
Mother completed high school (12)	$33.9k
Mother completed college (16)	$58.7k
Never does strenuous exercise (1)	$38.6k
Works out >3 times per week (7)	$49.8k

Make Predictions with 95% PIs

prediction interval

The approximate 95% ~~confidence interval~~ is found by taking your point estimate plus and minus some amount that you can calculate if you know the standard error of the estimate (SEE). The way this is calculated is:
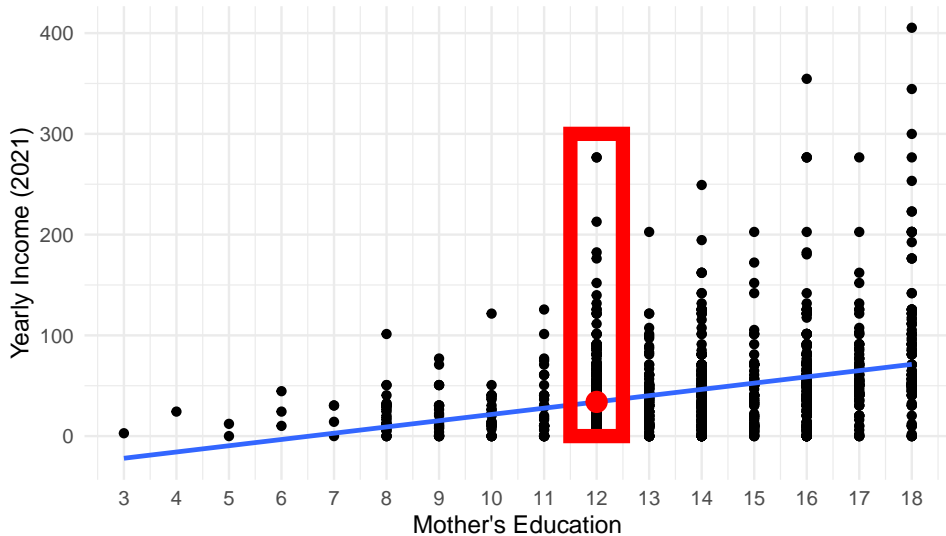
$$\text{Point Estimate} \pm 2(\text{Standard Error of the Estimate})$$

Source: Wilson, Keating, and Beal-Hodges (2012), p67

Why do we need a prediction interval?

Why do we need a prediction interval?

Point Estimate + 2 * Model Std Error

```
==================================
                    Earnings (2021)
----------------------------------
Mother's Education        6.21*
                         (0.43)

Constant                -40.64*
                         (5.87)

----------------------------------
Observations             1,570
Adjusted R2               0.12
Residual Std. Error   40.45 (df = 1568)
F Statistic         212.09* (df = 1; 1568)
==================================
Note:                    *p < 0.05
```

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.35
R Square	0.12
Adjusted R Square	0.12
Standard Error	40.45
Observations	1570

Make Predictions with 95% CIs

Prediction	Low	Estimates	High
Mother completed high school (12)	-47	$33.9k	114.8
Mother completed college (16)	-22.2	$58.7k	139.6
Never does strenuous exercise (1)	-44.9	$38.6k	122.1
Works out >3 times per week (7)	-33.7	$49.8k	133.3

Why do some states attract greater investment by companies and individuals than others?

Dataset 1

Based on the data for 2020 should states that want to grow the size of their economies focus on increasing college completion (bachelors) or increasing homeownership (homeowner rate)?
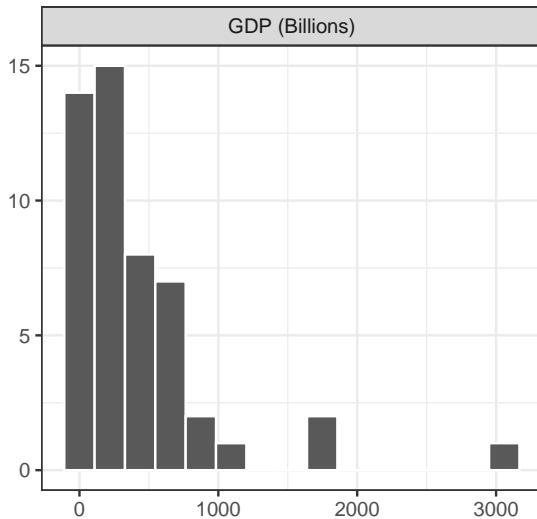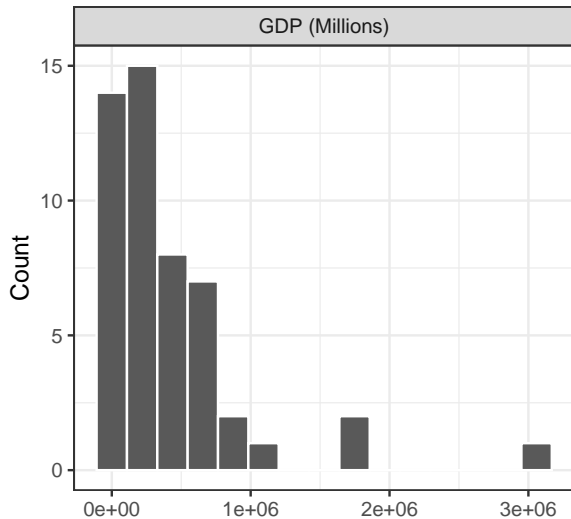
Based on the data for 2020 should states that want to grow the size of their economies focus on increasing college completion (bachelors) or increasing homeownership (homeowner rate)?

Fit and evaluate the two models

1. Model 1: Regress GDP (billions) on bachelors' degrees
2. Model 2: Regress GDP (billions) on homeownership rates
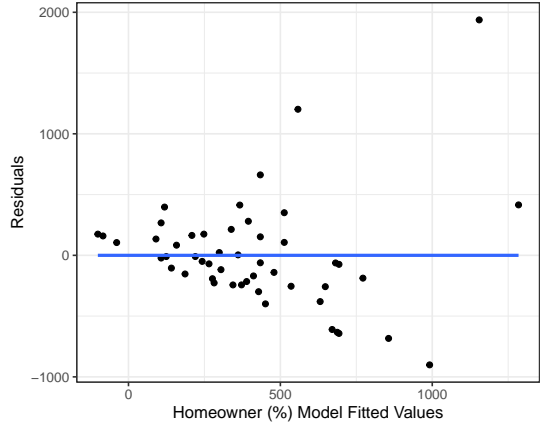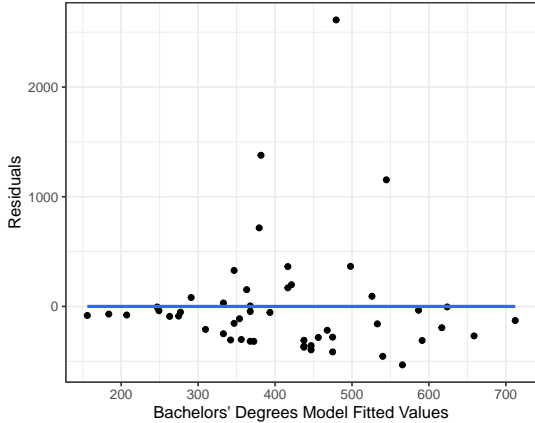
Calculate PEs with 95% CIs

1. Set Model 1 to the mean rate of bachelors' degrees
2. Set Model 2 to the mean rate of homeownership

	GDP (Billions USD)	
	(1)	(2)
Bachelors	23.27	
	(14.12)	
Homeownership		−56.30*
		(12.70)
Constant	−335.02	4,301.63*
	(460.39)	(879.84)
Observations	50	50
Adjusted R^2	0.03	0.28
Residual Std. Error (df = 48)	528.11	457.31
F Statistic (df = 1; 48)	2.71	19.64*
Note:		*p<0.05

Step 4: Check the Residuals

Making Predictions of GDP (Billions USD)

Prediction	Low	Estimate	High
Mean Bachelors' Degrees	-642.9	$413.3	1469.6
Mean Homeownership Rate	-501.1	$413.6	1328.2

Making Predictions of GDP (Billions USD)

Prediction	Low	Estimate	High
Mean Bachelors' Degrees	-642.9	$413.3	1469.6
Mean Homeownership Rate	-501.1	$413.6	1328.2

- What is effect of **increasing bachelors** from the mean by 10%?

- What is effect of **decreasing homeownership** from the mean by 10%?

Making Predictions of GDP (Billions USD)

Prediction	Low	Estimate	High
Mean Bachelors' Degrees	-642.9	$413.3	1469.6
Bachelors' Degrees + 10%	-567.9	$488	1544.5

Prediction	Low	Estimate	High
Mean Homeownership Rate	-501.1	$413.6	1328.2
Homeownership - 10%	-112.5	$802	1716.8