Today's Agenda

1. Practice fitting and evaluating simple OLS regressions

2. Make and interpret point estimates with prediction intervals

Justin Leinaweaver (Spring 2022)

Let's check in on your new skills!

- Bar plots (categorical variables)
- Histograms (numerical variables)
- Box plots (numerical x categorical)
- Scatterplots (numerical x numerical)
- Descriptive statistics
  - mean, sd, median, min, max, range, IQR
- Simple OLS models
  - Table, scatterplot with regression line and residual plot

Which is a better model of personal income (earnk2021)?

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

Make Predictions with 95% CIs

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% 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

Confidence Intervals

How confident are you in the parameter estimates in your model?

- e.g. the regression coefficients

	Earnings (2021)
Mother's Education	6.21*
	(0.43)
Constant	−40.64*
	(5.87)
Observations	1,570
Adjusted R^2	0.12
Residual Std. Error	40.45 (df = 1568)
F Statistic	212.09* (df = 1; 1568)
Note:	*p < 0.05

95% Confidence Intervals

- $\beta \pm 2 * SE$

	Earnings (2021)
Mother's Education	6.21*
	(0.43)
Constant	−40.64*
	(5.87)
Observations	1,570
Adjusted R^2	0.12
Residual Std. Error	40.45 (df = 1568)
F Statistic	212.09* (df = 1; 1568)
Note:	*p < 0.05

95% Confidence Intervals

- $\beta \pm 2 * SE$
- $6.21 \pm 2 * (0.43)$

	Earnings (2021)
Mother's Education	6.21*
	(0.43)
Constant	−40.64*
	(5.87)
Observations	1,570
Adjusted R^2	0.12
Residual Std. Error	40.45 (df = 1568)
F Statistic	212.09* (df = 1; 1568)
Note:	*p < 0.05

95% Confidence Intervals

- $\beta \pm 2 * SE$
- $6.21 \pm 2 * (0.43)$
- Low ≈ 5.35

	Earnings (2021)
Mother's Education	6.21*
	(0.43)
Constant	−40.64*
	(5.87)
Observations	1,570
Adjusted R^2	0.12
Residual Std. Error	40.45 (df = 1568)
F Statistic	212.09* (df = 1; 1568)
Note:	*p < 0.05

95% Confidence Intervals

- $\beta \pm 2 * SE$
- $6.21 \pm 2 * (0.43)$
- Low ≈ 5.35
- High ≈ 7.07

	Earnings (2021)
Mother's Education	6.21*
	(0.43)
Constant	−40.64*
	(5.87)
Observations	1,570
Adjusted R^2	0.12
Residual Std. Error	40.45 (df = 1568)
F Statistic	212.09* (df = 1; 1568)
Note:	*p < 0.05

Estimating Uncertainty

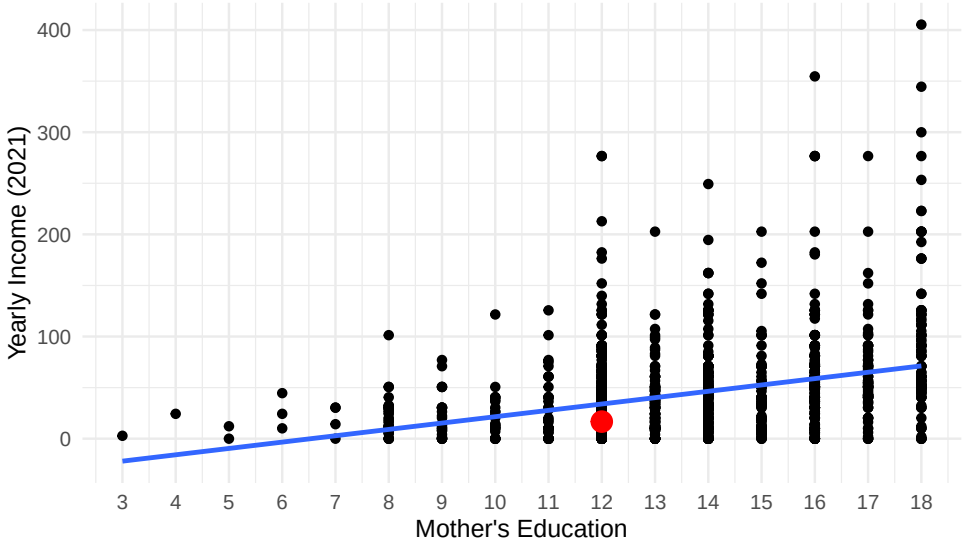
Confidence intervals

- How confident are you in the parameter estimates in your model?
- e.g. the regression coefficients

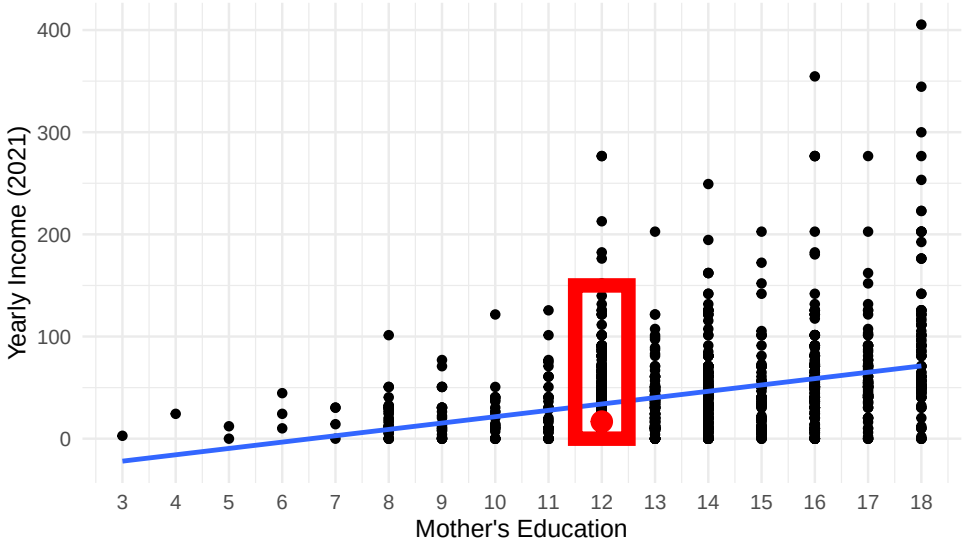
Prediction intervals

- How confident are you in the predictions made by your model?
- e.g. point estimates

Why do we need a prediction interval?



Why do we need a prediction interval?



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:

Point Estimate ± 2(Standard Error of the Estimate)

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

Make Predictions with 95% PIs

	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

What are the predicted earnings...

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

Dataset 1: The Motivating Problem

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)?

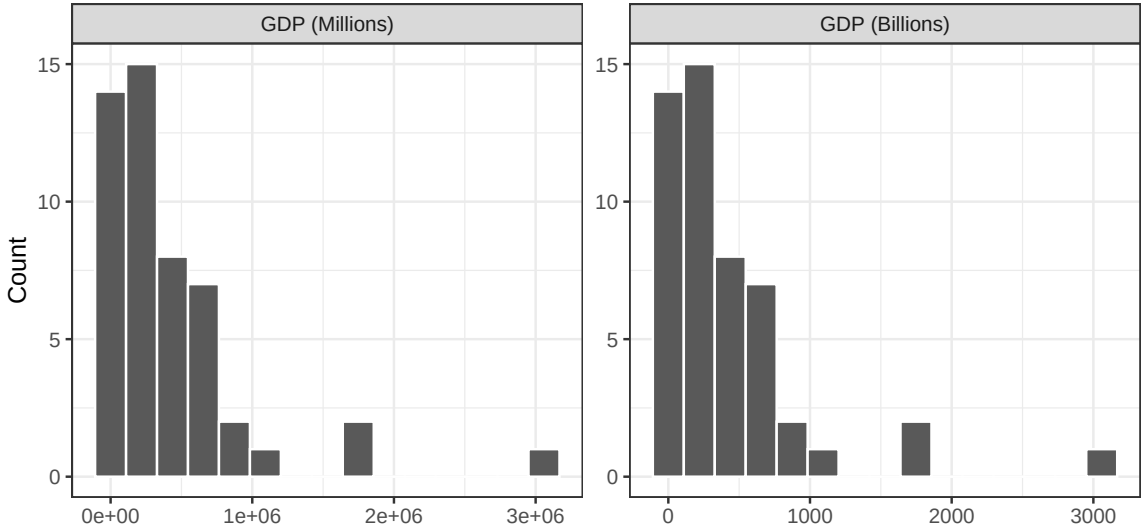
Fit and evaluate the two models

1. Model 1: Regress GDP on bachelors' degrees
2. Model 2: Regress GDP on homeownership rates

Calculate PEs with 95% CIs

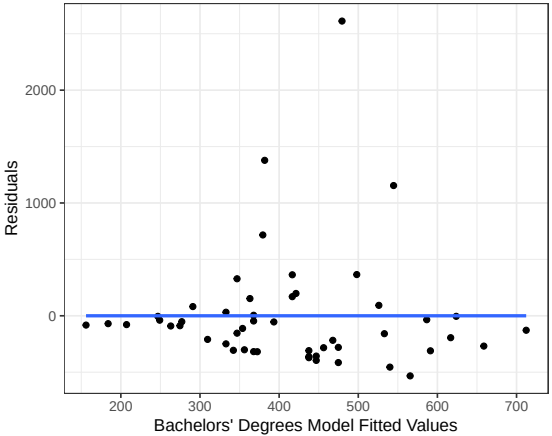
1. Set Model 1 to the mean rate of bachelors' degrees
2. Set Model 2 to the mean rate of homeownership

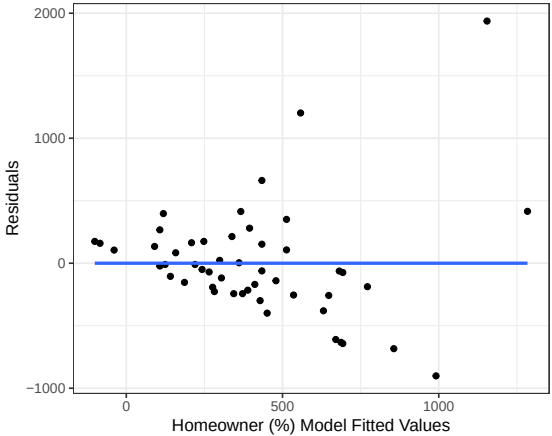
	GDP (Millions USD)	
	(1)	(2)
Bachelors	23,271.42	
	(14,124.28)	
Homeownership		−56,297.79*
		(12,704.98)
Constant	−335,020.10	4,301,632.00*
	(460,391.60)	(879,837.10)
Observations	50	50
Adjusted R^2	0.03	0.28
Residual Std. Error (df = 48)	528,114.80	457,308.00
F Statistic (df = 1; 48)	2.71	19.64*
Note:		*p<0.05


GDP (Millions)
GDP (Billions)
Count
15
10
5
0
0e+00
1e+06
2e+06
3e+06
0
1000
2000
3000

	GDP (Millions USD)		GDP (Billions USD)	
	(1)	(2)	(3)	(4)
Bachelors	23,271.42		23.27	
	(14,124.28)		(14.12)	
Homeownership		−56,297.79*		−56.30*
		(12,704.98)		(12.70)
Constant	−335,020.10	4,301,632.00*	−335.02	4,301.63*
	(460,391.60)	(879,837.10)	(460.39)	(879.84)
Observations	50	50	50	50
Adjusted R^2	0.03	0.28	0.03	0.28
Residual Std. Error (df = 48)	528,114.80	457,308.00	528.11	457.31
F Statistic (df = 1; 48)	2.71	19.64*	2.71	19.64*
Note:				*p<0.05

Step 4: Check the Residuals





Making Predictions of GDP (Billions USD)

Prediction	Low	Estimates	High
Mean Bachelors' Degrees			
Mean Homeownership Rate			

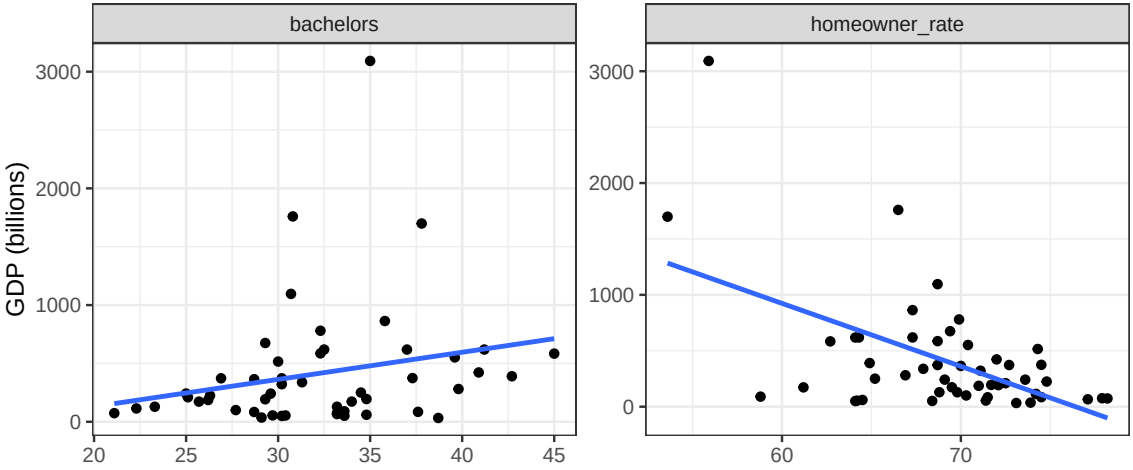
Making Predictions of GDP (Billions USD)

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

Making Predictions of GDP (Billions USD)

Prediction	Low	Estimate	High
Mean Bachelors' Degrees	-642.9	$413.3k	1469.6
Max Bachelors' Degrees	-344.09	$712k	1768.35

Prediction	Low	Estimate	High
Mean Homeownership Rate	-501.1	$413.6k	1328.2
Min Homeownership	369.33	$1284k	2198.57


bachelors
homeowner_rate
GDP (billions)
3000
2000
1000
0
20
25
30
35
40
45
60
70