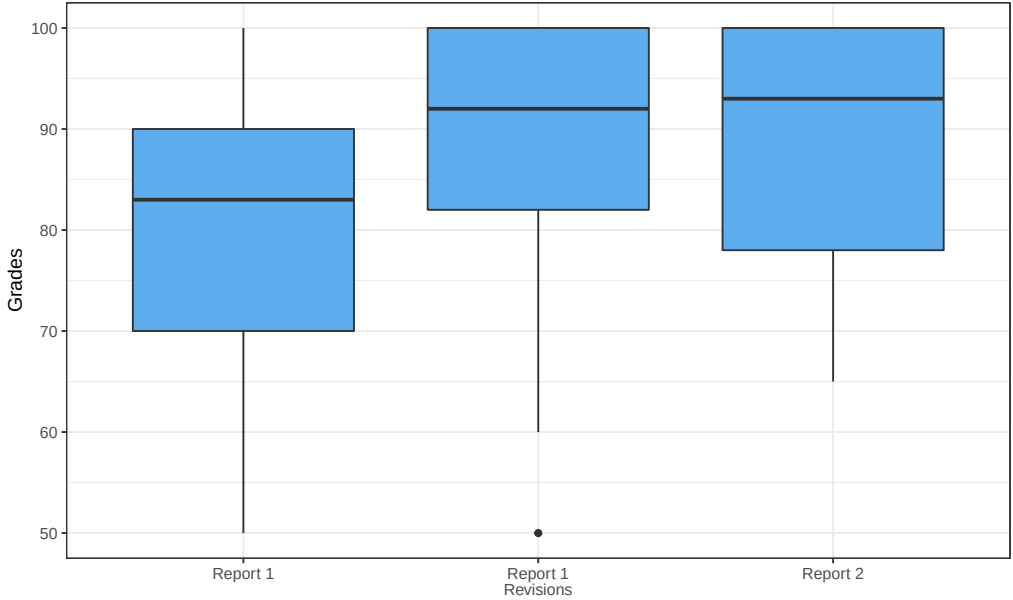
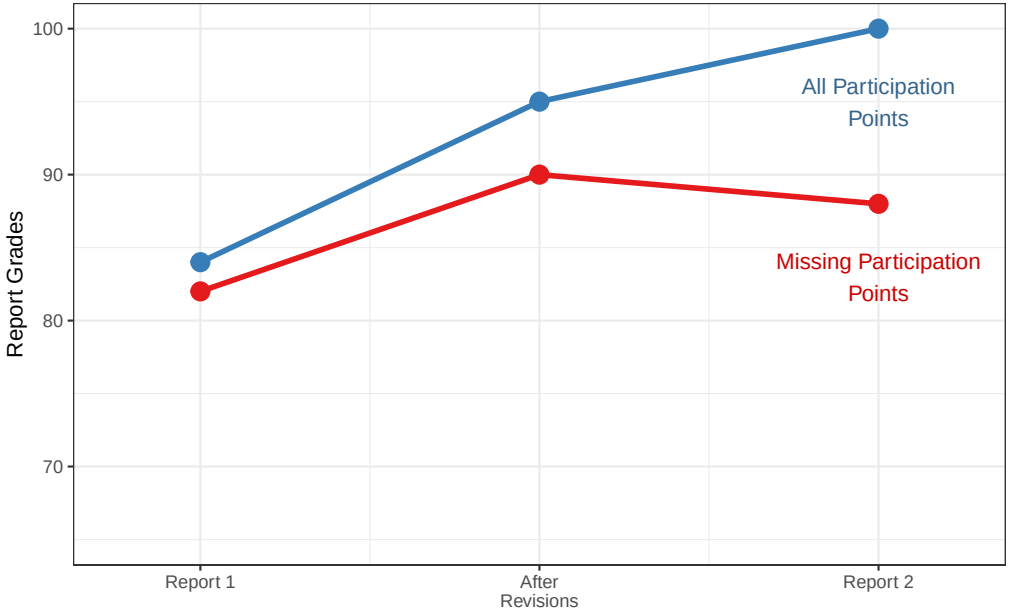
Today's Agenda

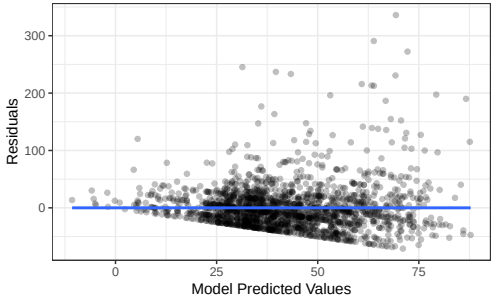
Practicing Multiple Linear Regressions (Adding marginal effects plots)

Justin Leinaweaver (Spring 2022)


Grades
100
90
80
70
60
50
Report 1
Report 1
Revisions
Report 2


100
90
80
70
Report Grades
All Participation Points
Missing Participation Points
Report 1
After Revisions
Report 2

	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05



	age	education	exercise
age	1.00	-0.15	-0.33
education	-0.15	1.00	0.18
exercise	-0.33	0.18	1.00

	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

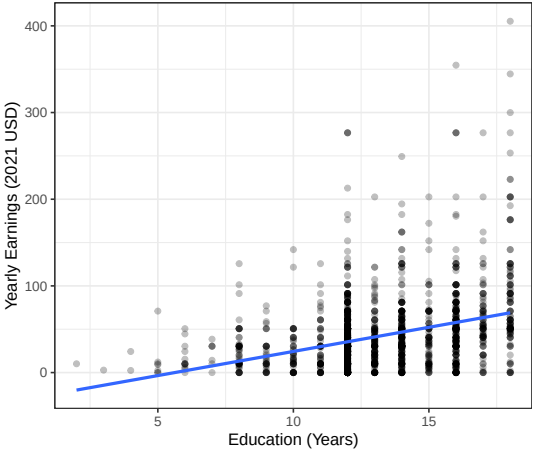
Baseline: Age 20, Education 13, Exercise 1

Prediction	Low	Estimate	High
Baseline	-50.8	27	104.8
Finish College	-33.7	44.1	121.9
Finish Grad School	-22.3	55.5	133.3

Prediction	Low	Estimate	High
Baseline	-50.8	27	104.8
Exercise 1x / week	-45.4	32.4	110.2
Exercise >3x	-40	37.8	115.6

	Earnings (2021 USD)
Education	5.57*
	(0.36)
Constant	−31.34*
	(4.89)
Observations	1,813
Adjusted R^2	0.11
Residual Std. Error	39.50 (df = 1811)
F Statistic	235.79* (df = 1; 1811)
Note:	*p < 0.05

	Earnings (2021 USD)
Education	5.57*
	(0.36)
Constant	−31.34*
	(4.89)
Observations	1,813
Adjusted R^2	0.11
Residual Std. Error	39.50 (df = 1811)
F Statistic	235.79* (df = 1; 1811)
Note:	*p < 0.05



	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

?

	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

Prediction	Estimate
Baseline	$27k
Finish College	$44.1k
Finish Grad School	$55.5k

Prediction	Estimate
Baseline	$27k
Exercise 1x / week	$32.4k
Exercise >3x	$37.8k

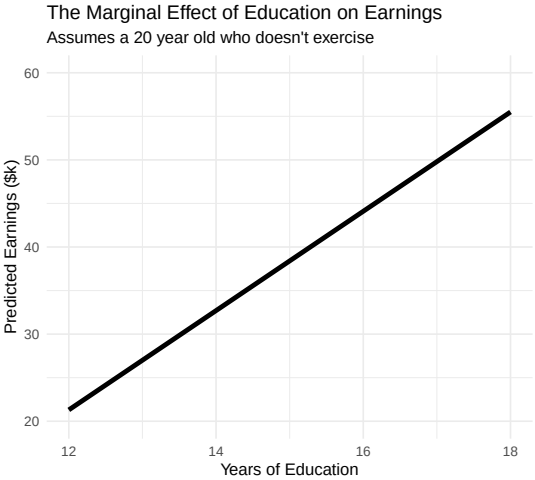
	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

Education	Prediction
12	
13	
14	
15	
16	$44.1k
17	
18	$55.5k

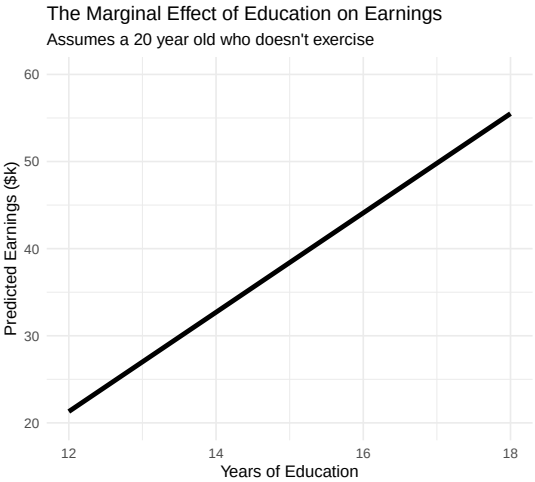
	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

Education	Prediction
12	$21.3k
13	$27k
14	$32.7k
15	$38.4k
16	$44.1k
17	$49.8k
18	$55.5k

Education	Prediction
12	$21.3k
13	$27k
14	$32.7k
15	$38.4k
16	$44.1k
17	$49.8k
18	$55.5k



	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05



	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

Practice

Make a marginal effects plot of exercise for a 22 year old who has finished their undergraduate degree (education = 16)

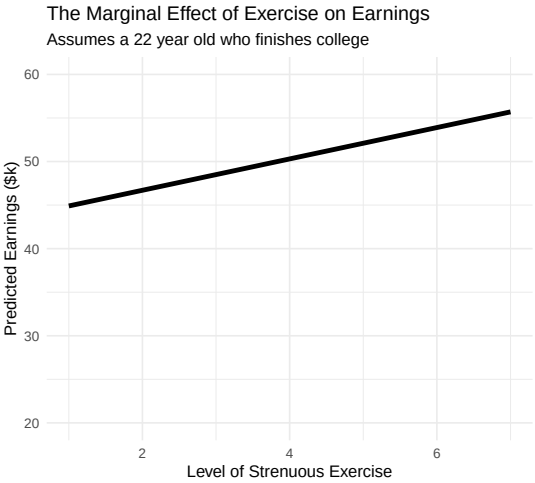
	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

Exercise	Prediction
1	
2	
3	
4	
5	
6	
7	

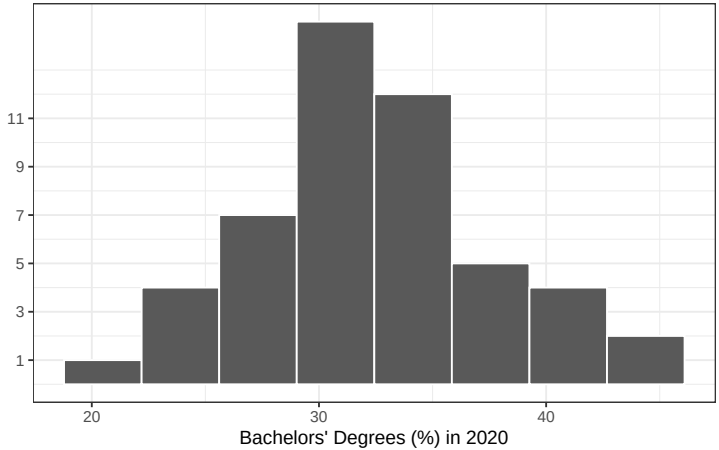
	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05

Exercise	Prediction
1	$44.9k
2	$46.7k
3	$48.5k
4	$50.3k
5	$52.1k
6	$53.9k
7	$55.7k

	Earnings (2021 USD)
Age	0.42*
	(0.06)
Education	5.71*
	(0.36)
Exercise	1.80*
	(0.42)
Constant	−56.86*
	(5.86)
Observations	1,813
Adjusted R^2	0.14
Residual Std. Error	38.88 (df = 1809)
F Statistic	101.17* (df = 3; 1809)
Note:	*p < 0.05



What is the "best" model of bachelor's degree completion in dataset 1?



What is the "best" model of bachelor's degree completion in dataset 1?

1. Choose the logical predictors
2. Fit a simple OLS regression to each predictor
3. Fit a multiple regression with the "best" of those
4. Evaluate the model using all five steps
5. Use the model to make predictions

What is the "best" model of bachelor's degree completion in dataset 1?

Outcome

- Bachelors' Degrees

Predictors to Consider

- GDP (Billions), GDP (Rate), Homeownership, Manufacturing employment, Minimum wage, Population, Rental Vacancy Rate, State Tax Rate on Wages, Unemployment

Common Regression Mistakes (Wheelan ch12)

1. Linear regression on nonlinear relationships
2. Correlation does not equal causation
3. Reverse causality
4. Omitted variable bias (too few variables)
5. Highly correlated explanatory variables (multicollinearity)
6. Extrapolating beyond the data
7. Data mining (too many variables)