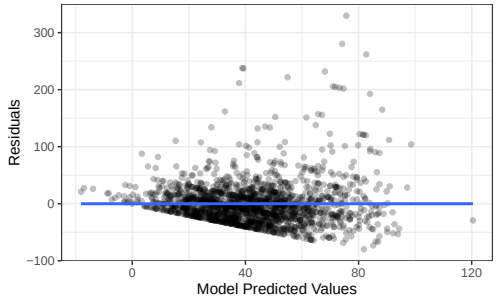
# Today's Agenda

## Practicing Multiple Linear Regressions

Justin Leinaweaver (Spring 2022)

| | Earnings (2021 USD) |
|---|---|
| Age | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |




| | age | education | exercise | height |
|---|---|---|---|---|
| age | 1 | -0.15 | -0.33 | -0.14 |
| education | -0.15 | 1 | 0.18 | 0.11 |
| exercise | -0.33 | 0.18 | 1 | 0.22 |
| height | -0.14 | 0.11 | 0.22 | 1 |

# For Today - Make Predictions (w/ 95% PIs)

Establish a baseline prediction for a hypothetical person who is 25 (Age), Education 13, Exercise 1, Height 67

Then calculate the predicted income if:

1. Finish college? (16)
2. Finish grad school? (18)
3. Increase exercise to 1x / week? (4)
4. Increase exercise to 3x / week? (7)

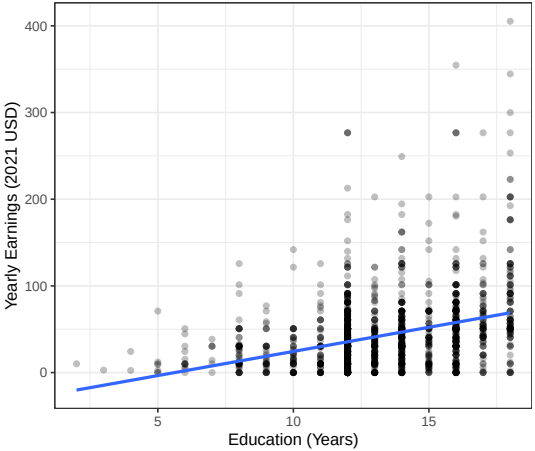
| | Earnings (2021 USD) |
|---|---|
| Age | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |

| Prediction | Low | Estimate | High |
|---|---|---|---|
| Baseline | -44.9 | 29.9 | 104.7 |
| Finish College | -28.7 | 46.1 | 120.9 |
| Finish Grad School | -17.9 | 56.9 | 131.7 |

| Prediction | Low | Estimate | High |
|---|---|---|---|
| Baseline | -44.9 | 29.9 | 104.7 |
| Exercise 1x / week | -42.2 | 32.6 | 107.4 |
| Exercise >3x | -39.5 | 35.3 | 110.1 |

| | 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 | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |

?

| | Earnings (2021 USD) |
|---|---|
| Age | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |

| Prediction | Low | Estimate | High |
|---|---|---|---|
| Baseline | -44.9 | 29.9 | 104.7 |
| Finish College | -28.7 | 46.1 | 120.9 |
| Finish Grad School | -17.9 | 56.9 | 131.7 |

| Prediction | Low | Estimate | High |
|---|---|---|---|
| Baseline | -44.9 | 29.9 | 104.7 |
| Exercise 1x / week | -42.2 | 32.6 | 107.4 |
| Exercise >3x | -39.5 | 35.3 | 110.1 |

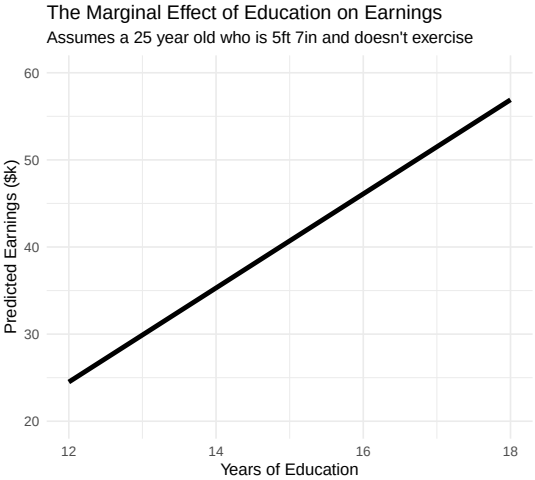
| | Earnings (2021 USD) |
|---|---|
| Age | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |

| Education | Prediction |
|---|---|
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | $46.1k |
| 17 | |
| 18 | $56.9k |

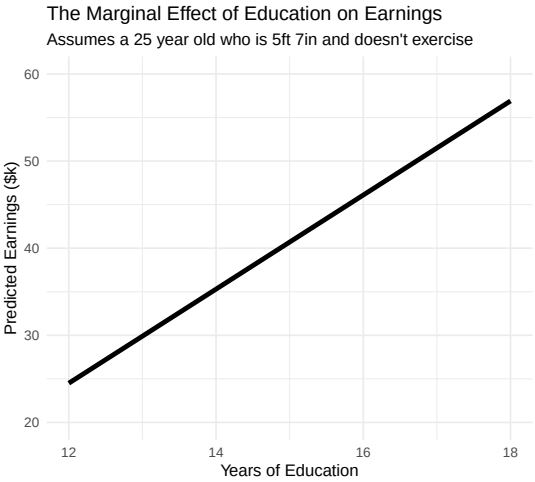
| | Earnings (2021 USD) |
|---|---|
| Age | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |

| Education | Prediction |
|---|---|
| 12 | $24.5k |
| 13 | $29.9k |
| 14 | $35.3k |
| 15 | $40.7k |
| 16 | $46.1k |
| 17 | $51.5k |
| 18 | $56.9k |

| Education | Prediction |
|---|---|
| 12 | $24.5k |
| 13 | $29.9k |
| 14 | $35.3k |
| 15 | $40.7k |
| 16 | $46.1k |
| 17 | $51.5k |
| 18 | $56.9k |

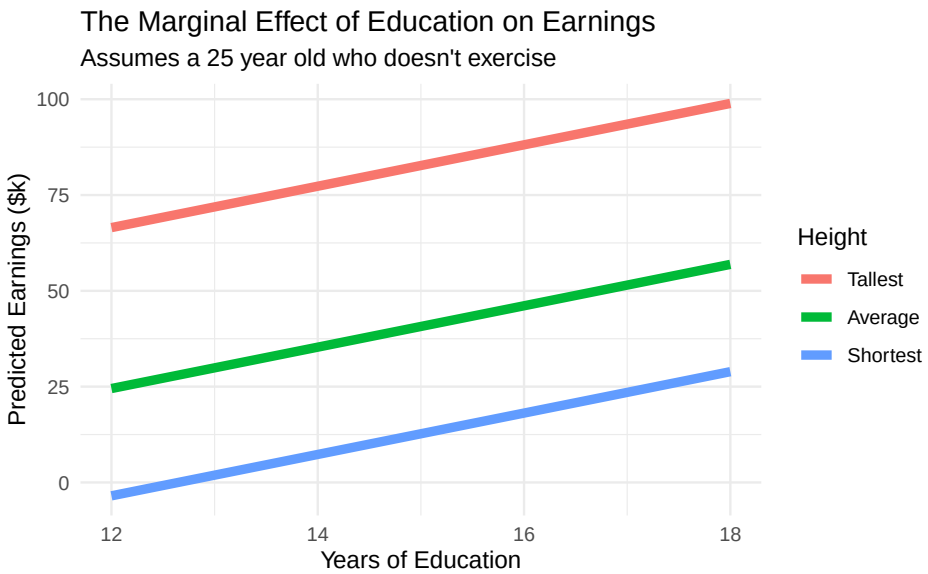


| | Earnings (2021 USD) |
|---|---|
| Age | 0.47* |
| | (0.05) |
| Education | 5.42* |
| | (0.35) |
| Exercise | 0.95* |
| | (0.41) |
| Height | 2.84* |
| | (0.24) |
| Constant | −241.27* |
| | (16.34) |
| Observations | 1,813 |
| Adjusted $R^2$ | 0.21 |
| Residual Std. Error | 37.43 (df = 1808) |
| F Statistic | 118.04* (df = 4; 1808) |
| *Note:* | *p < 0.05 |



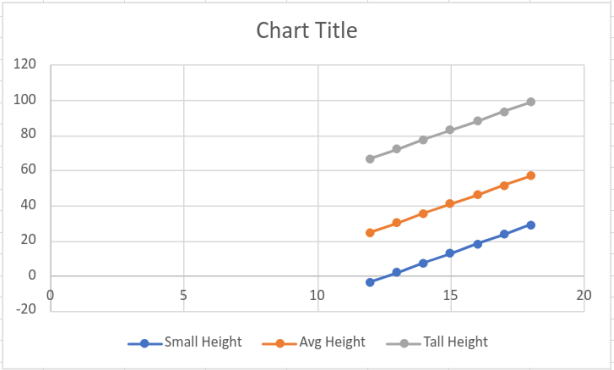
Create a marginal effects plot with THREE prediction lines

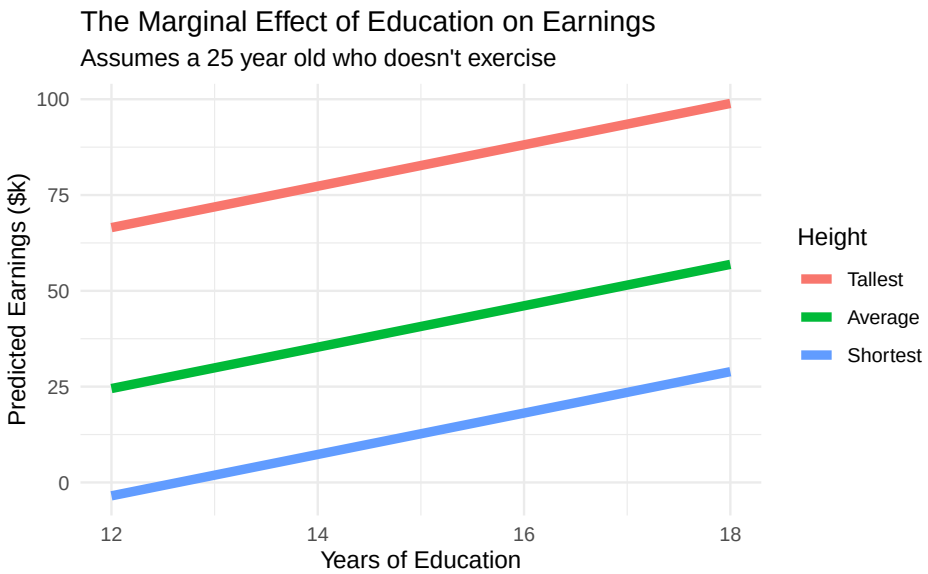
1. Predict the effect of education (12-18) for a 25 year old who doesn't exercise and is the shortest person in the sample (4'9")
2. Repeat above but assume a person of average height (5'7")
3. Repeat above but assume the tallest person in the sample (6'10")


The Marginal Effect of Education on Earnings
Assumes a 25 year old who doesn't exercise
Predicted Earnings ($k)
100
75
50
25
0
12
14
16
18
Years of Education
Height
Tallest
Average
Shortest

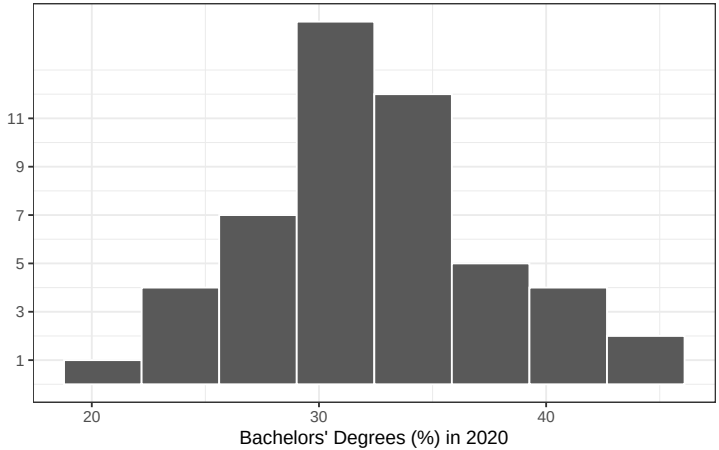
| | A | B | C | D |
|---|---|---|---|---|
| 1 | Education | Small Height | Avg Height | Tall Height |
| 2 | 12 | -3.5 | 24.5 | 66.5 |
| 3 | 13 | 1.9 | 29.9 | 71.9 |
| 4 | 14 | 7.3 | 35.3 | 77.3 |
| 5 | 15 | 12.7 | 40.7 | 82.7 |
| 6 | 16 | 18.1 | 46.1 | 88.1 |
| 7 | 17 | 23.5 | 51.5 | 93.5 |
| 8 | 18 | 28.9 | 56.9 | 98.9 |
| 9 | | | | |

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Education | Small Height | Avg Height | Tall Height |
| 2 | 12 | -3.5 | 24.5 | 66.5 |
| 3 | 13 | 1.9 | 29.9 | 71.9 |
| 4 | 14 | 7.3 | 35.3 | 77.3 |
| 5 | 15 | 12.7 | 40.7 | 82.7 |
| 6 | 16 | 18.1 | 46.1 | 88.1 |
| 7 | 17 | 23.5 | 51.5 | 93.5 |
| 8 | 18 | 28.9 | 56.9 | 98.9 |




The Marginal Effect of Education on Earnings
Assumes a 25 year old who doesn't exercise
Predicted Earnings ($k)
100
75
50
25
0
12
14
16
18
Years of Education
Height
Tallest
Average
Shortest

# 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)