Simple Linear Regression

William Sexton

Math 221C, BYU-I
July, 2019

Outline

Why bother with regression?

Why bother with regression?

Does your personality actually affect your life?

Why bother with regression?

Does your personality actually affect your life?

Pick one of the Big 5 Personality traits (extraversion, agreeableness, openness, conscientiousness, neuroticism)

Why bother with regression?

Does your personality actually affect your life?

Pick one of the Big 5 Personality traits (extraversion, agreeableness, openness, conscientiousness, neuroticism)

Consider a particular outcome (Career success, Life span, Mental health issues)

Public Health Significance of Neuroticism

- Predicting hours of TV viewed by children using hours of TV viewed by parents

- Predicting hours of TV viewed by children using hours of TV viewed by parents
- Marriage and divorce

- Predicting hours of TV viewed by children using hours of TV viewed by parents
- Marriage and divorce
- Predicting number of children using multiple factors

- Predicting hours of TV viewed by children using hours of TV viewed by parents
- Marriage and divorce
- Predicting number of children using multiple factors
- Predicting popular news articles

- Predicting hours of TV viewed by children using hours of TV viewed by parents
- Marriage and divorce
- Predicting number of children using multiple factors
- Predicting popular news articles
- Does the BYU-I learning model improve learning?

Does simple linear regression really matter?

Does simple linear regression really matter?

Possible criticisms:

- Complex real world phenomena cannot be explained by one variable.
- Most naturally occurring X and Y relationships are nonlinear. Linearity is an oversimplification.

Does simple linear regression really matter?

Possible criticisms:

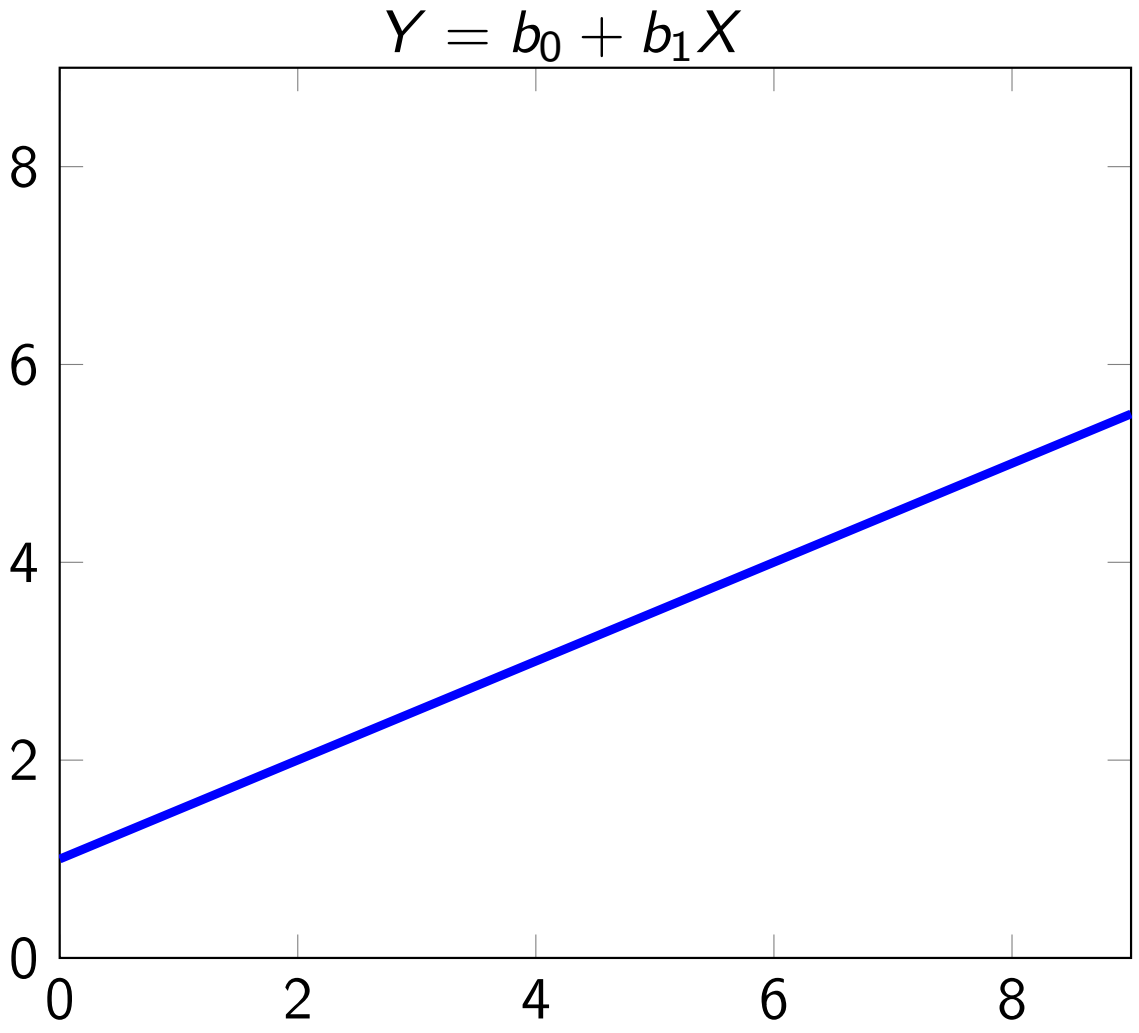
- Complex real world phenomena cannot be explained by one variable.
- Most naturally occurring X and Y relationships are nonlinear. Linearity is an oversimplification.

Counterpoints:

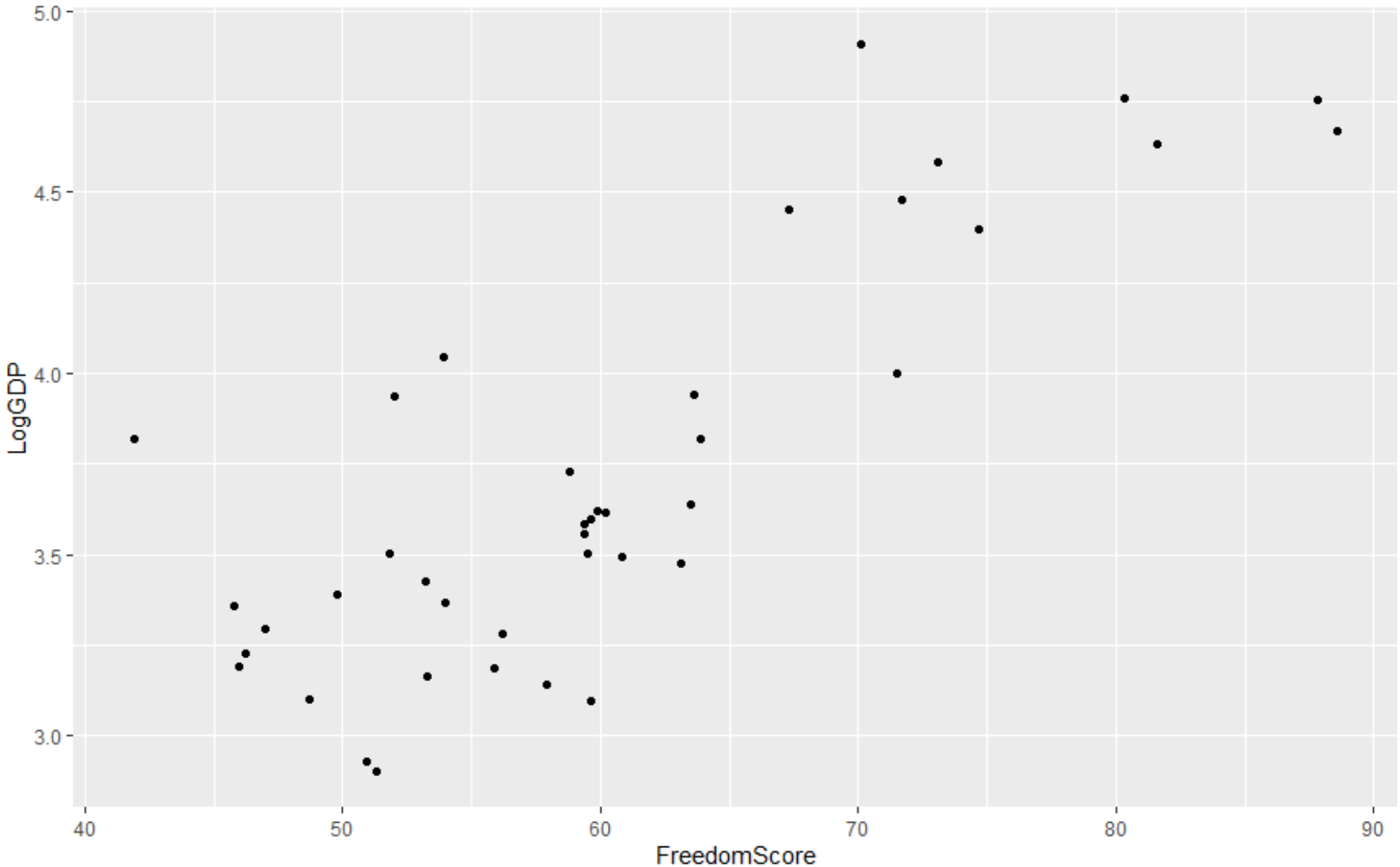
- Simple Linear regression is a surprisingly powerful tool.
- That being said, it is also foundational to other more sophisticated regression methods such as multiple regression.

Outline

Linear Equation Review



Scatterplot



Linear Equation Prepare Question

The first requirement of simple regression is that a linear equation can be used to describe the relationship between X and Y. To check this requirement *after* collecting data, you can make a scatterplot and look for a "hot dog" shape. How would you go about making an educated guess *before* collecting any data about whether this requirement is satisfied by X and Y?

Outline

Prepare Question

- Suppose you want to predict someones weight using height as the explanatory variable. You measure height in inches. You calculate the regression output and conduct a hypothesis test for the regression slope. You find sufficient evidence to reject the null hypothesis $\beta_1 = 0$ at a .05 level of significance. What would happen if you took your data and converted height from inches to feet and redid the regression analysis? What would change? What wouldn't change?

Prepare Question

- Suppose you want to predict someones weight using height as the explanatory variable. You measure height in inches. You calculate the regression output and conduct a hypothesis test for the regression slope. You find sufficient evidence to reject the null hypothesis $\beta_1 = 0$ at a .05 level of significance. What would happen if you took your data and converted height from inches to feet and redid the regression analysis? What would change? What wouldn't change?
- Would you expect anything to be different if you had *measured* height in feet to begin with?

Outline

Hypothesis Testing Assumptions

1. Linear Relation
2. Normal Error Term
3. Constant Variance
4. X's are known constants
5. Observations are independent

GitHub link

https://github.com/sextonw/Math-221C

- Click the green "clone or download" button, then "download zip"
- Extract zip contents to a location of your choice (e.g. Desktop)
- Open the ACS-PUMS folder

Group Activity

- Split into small groups of 5 or 6 people.
- Open up the data in excel (psam_p16.csv).
- Open up the data dictionary.
- Figure out what the variables are and design a hypothesis test with two of them.
- Copy the two variables you picked into your excel toolbox.
- Report your findings.

Outline

Prepare Question

Suppose you have a dataset with one response variable Y and many explanatory variables (e.g. $X_1, X_2, X_3, \ldots X_{60}$). You conduct a simple regression hypothesis test with Y and X_1, then do the same thing with Y and X_2, and then with Y and X_3, and so on until you find a statistically significant relationship at the .05 level with one of the explanatory variables, say X_{22}. You are very excited by this finding and share your results. Why might this be a bad idea?