STA 235H - Multiple Regression: Nonlinearity

Fall 2022

McCombs School of Business, UT Austin

Last week

- Reviewed more **multiple regression** models:
  - Interaction models
  - Logarithmic outcomes

Today

- Continue with nonlinearity:
  - Review of regressions with log variables
  - Polynomial terms in regressions
- Assessing issues with our data:
  - Outliers
  - Multicollinearity
- Binary response models:
  - Should we do something different than traditional OLS?

Logs, logs everywhere

Regressions and logarithms

- Every time we have a variable in a logarithm, we should think percentage change
- For a log-level regression $\log(Y) = \beta_1 + \beta_2 X + \varepsilon$:
  - Exact association: "For a one-unit increase in X, Y changes, on average, by $(\exp(\hat{\beta}_2) - 1) \times 100\%$"
  - Approximation: "For a one-unit increase in X, Y changes, on average by $\hat{\beta}_2 \times 100\%$"

What about if X is also in a logarithm?

How would we interpret coefficients now?

- There are also approximations that can be useful!

Model	Interpretation of β
Level-Level regression $y = \alpha + \beta x$	$\Delta y = \beta \Delta x$
Log-Level regression $\log(y) = \alpha + \beta x$	$\%\Delta y = 100 \cdot \beta \Delta x$
Level-Log regression $y = \alpha + \beta \log(x)$	$\Delta y = \frac{\beta}{100} \%\Delta x$
Log-Log regression $\log(y) = \alpha + \beta \log(x)$	$\%\Delta y = \beta \%\Delta x$

Let's practice!

$$\log(Revenue) = \beta_0 + \beta_1 Bechdel + \beta_2 Rating + \varepsilon$$

$$\log(Income) = \beta_0 + \beta_1 City + \beta_2 Profession + \beta_3 Education + \varepsilon$$

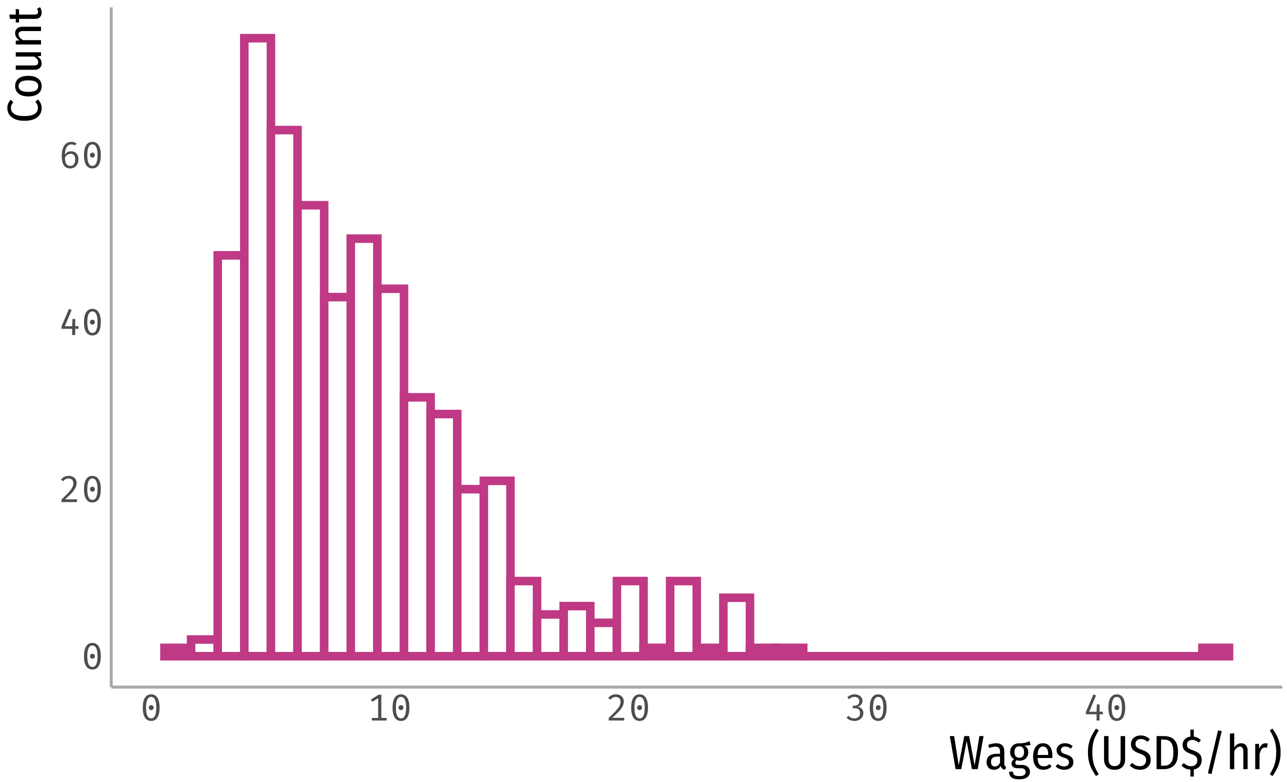
$$GPACollege = \beta_0 + \beta_1 SAT + \beta_2 \log(FamilyIncome) + \beta_3 GPAHS + \varepsilon$$

Getting squared

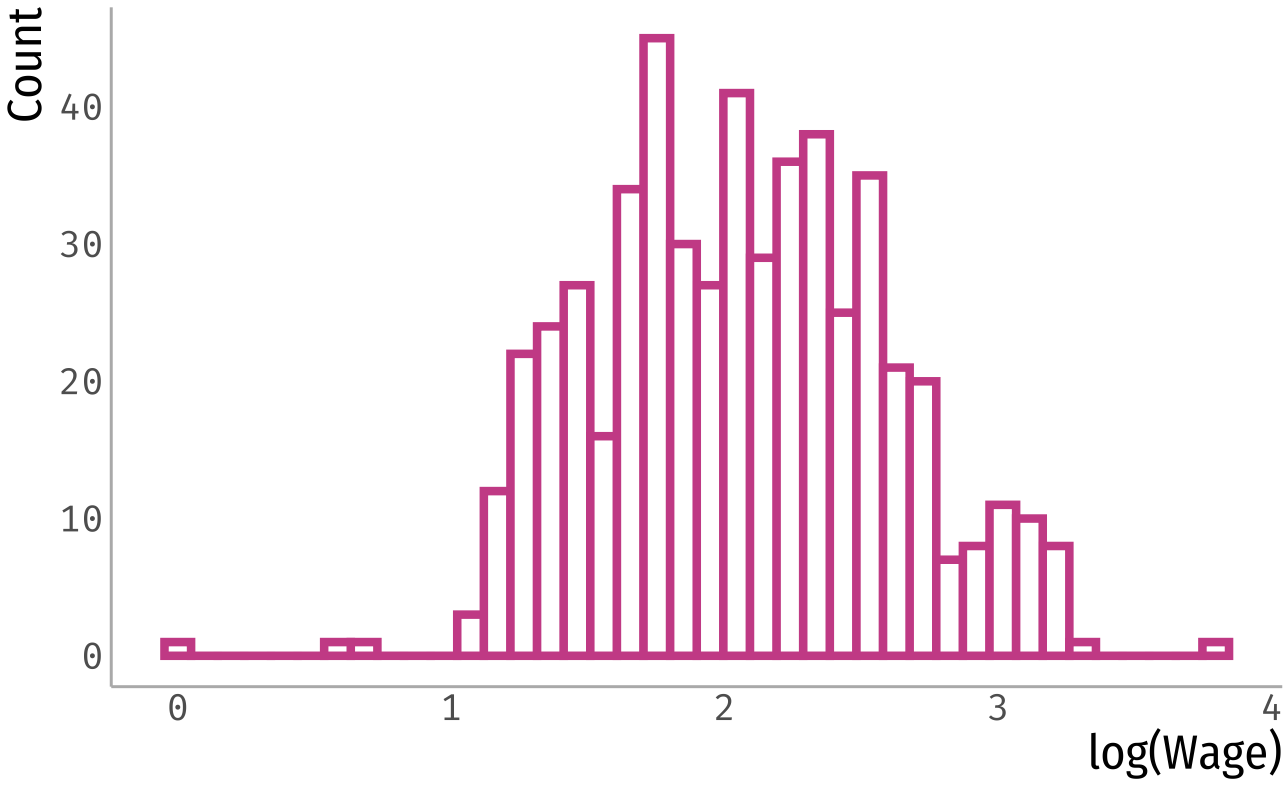
Adding polynomial terms

- Another way to capture **nonlinear associations** between the outcome (Y) and covariates (X) is to include **polynomial terms**:
    - e.g. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$
- Let's look at an example!

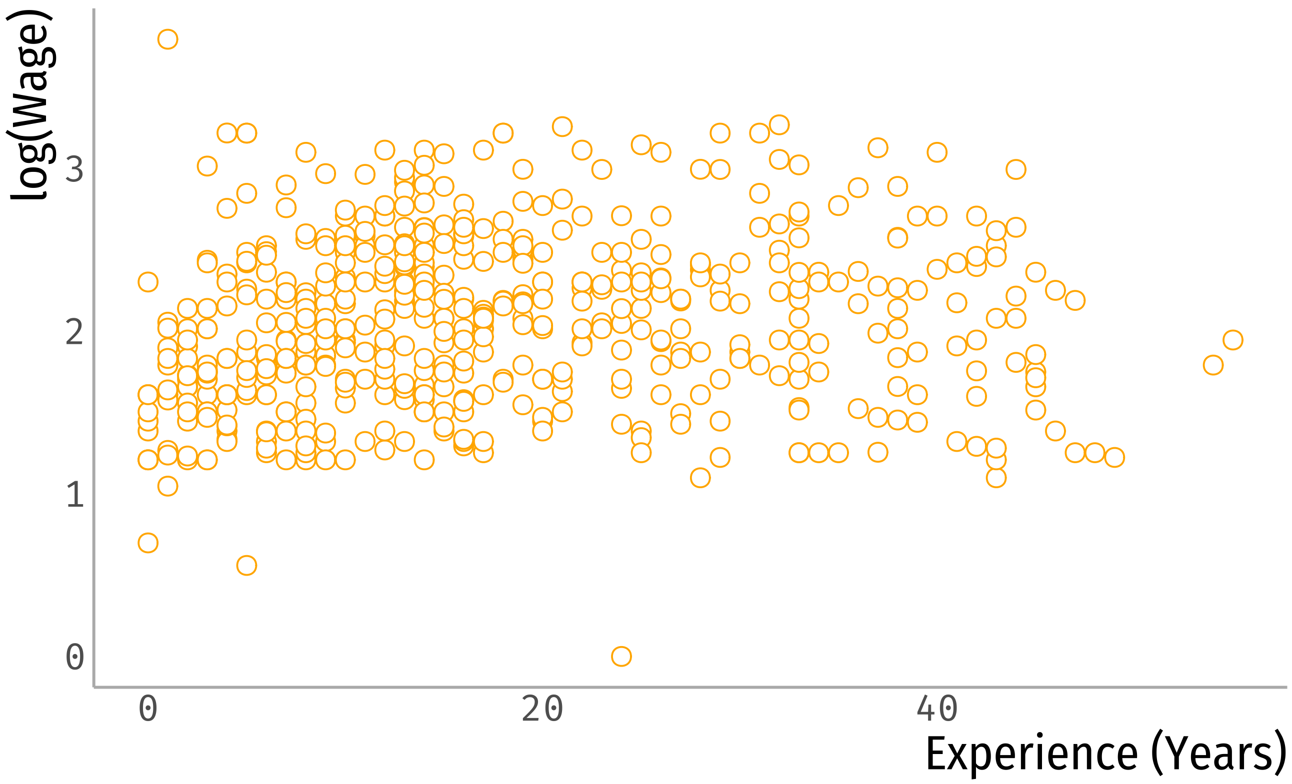
Determinants of wages: CPS 1985



Determinants of wages: CPS 1985

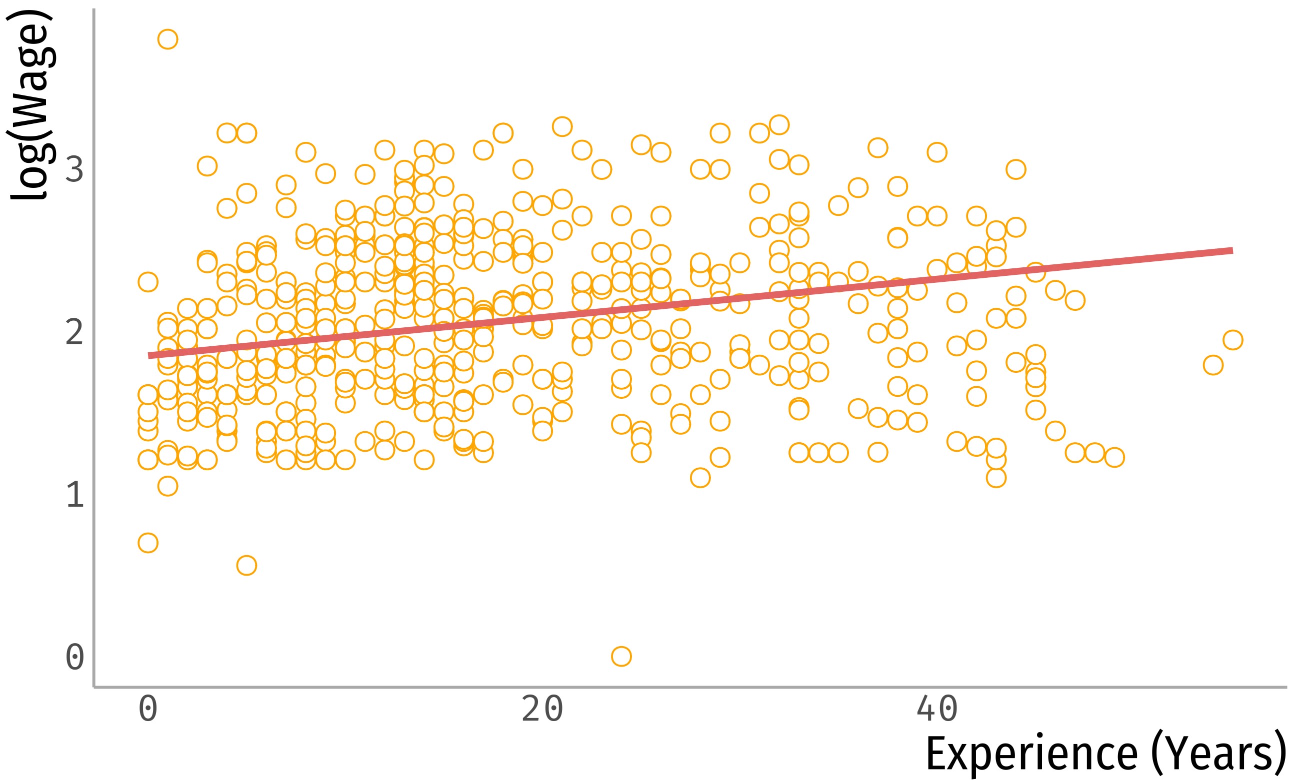


Experience vs wages: CPS 1985



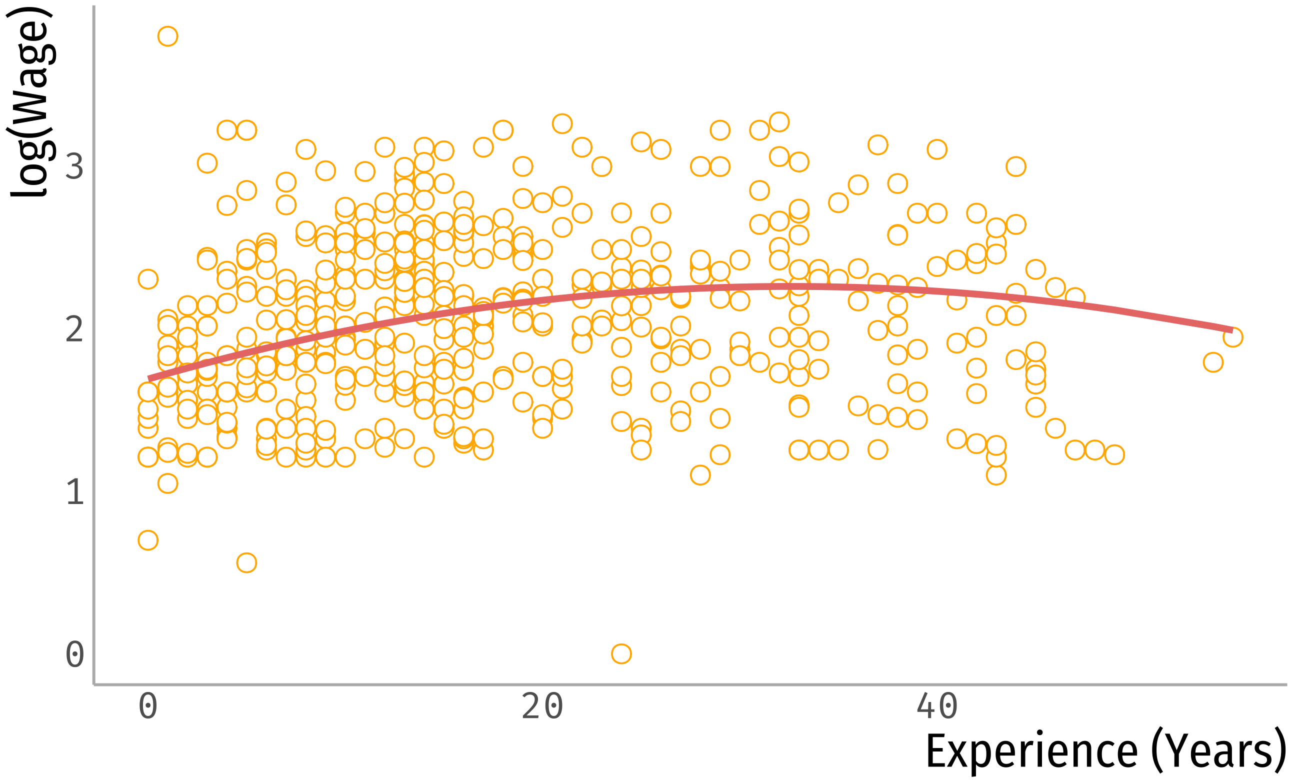
Experience vs wages: CPS 1985

$$\log(Wage) = \beta_0 + \beta_1 Educ + \beta_2 Exp + \varepsilon$$



Experience vs wages: CPS 1985

$$\log(Wage) = \beta_0 + \beta_1 Educ + \beta_2 Exp + \beta_3 Exp^2 + \varepsilon$$



Mincer equation

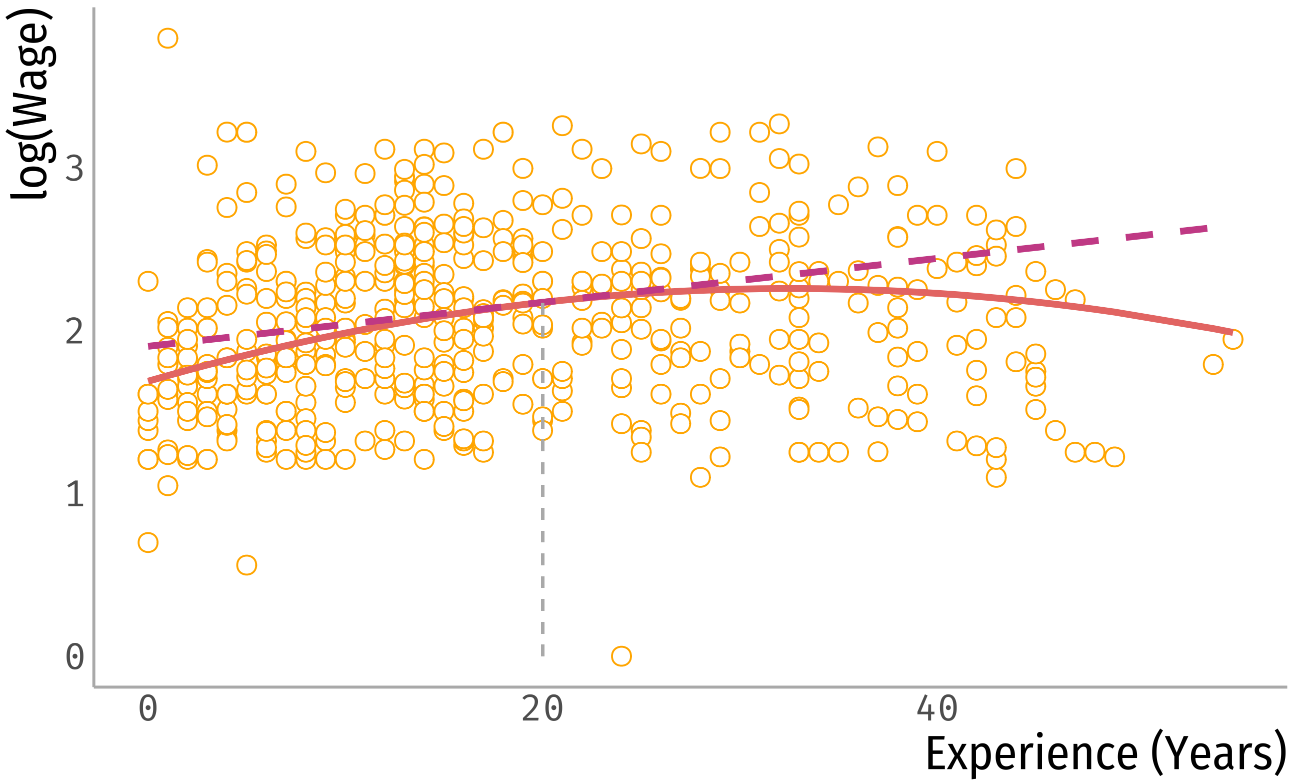
$$\log(Wage) = \beta_0 + \beta_1 Educ + \beta_2 Exp + \beta_3 Exp^2 + \varepsilon$$

- Interpret the coefficient for **education**

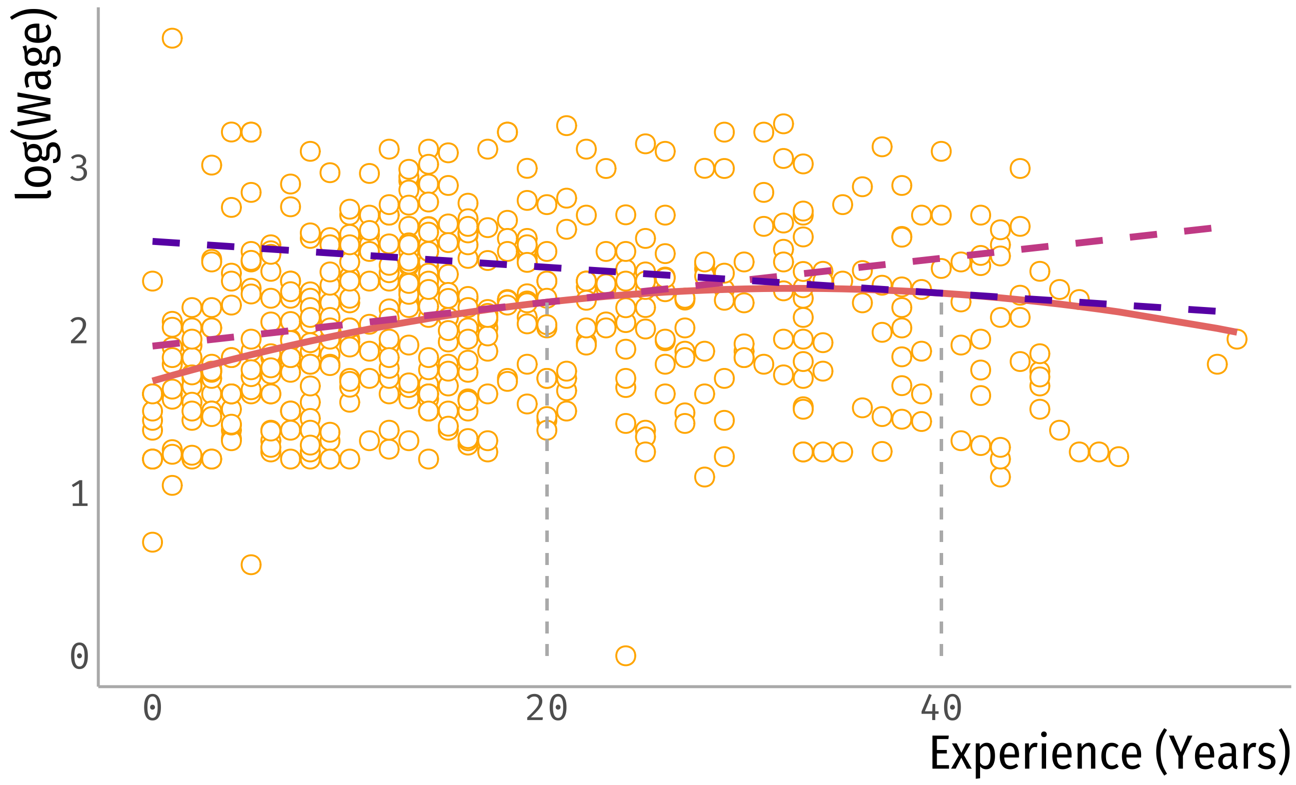
One additional year of education is associated, on average, to $\hat{\beta}_1 \times 100\%$ *increase in hourly wages, holding experience constant*

- What is the association between experience and wages?

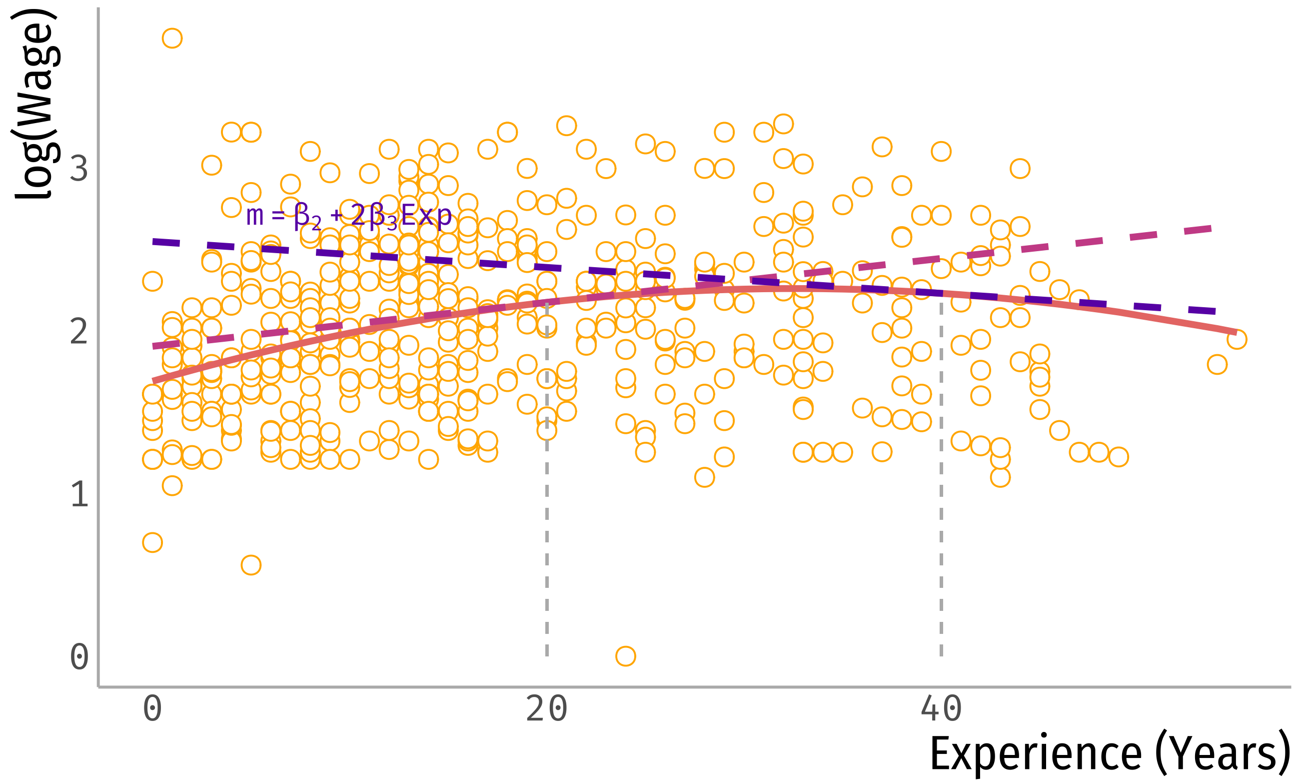
Interpreting coefficients in quadratic equation



Interpreting coefficients in quadratic equation



Interpreting coefficients in quadratic equation



Interpreting coefficients in quadratic equation

$$\log(Wage) = \beta_0 + \beta_1 Educ + \beta_2 Exp + \beta_3 Exp^2 + \varepsilon$$

What is the association between experience and wages?

- Pick a value for Exp_0 (e.g. mean, median, one value of interest)

Increasing work experience from 20 to 21 years is associated, on average, to a $(\hat{\beta}_2 + 2\hat{\beta}_3 \cdot 20)100\%$ *increase on hourly wages, holding education constant*

Let's go to R!

References

- Ismay, C. & A. Kim. (2021). "Statistical Inference via Data Science". Chapter 6 & 10.