

AP Statistics Extensions

Multiple Regression

Fall 2025

Matthew J. Madison
mjmadison@uga.edu

Multiple Regression

Agenda

- A little about me
- Question and answer
- Multiple regression
 - Review of simple linear regression
 - Conceptual overview
 - Examples using RStudio statistical software
 - Lab assignment

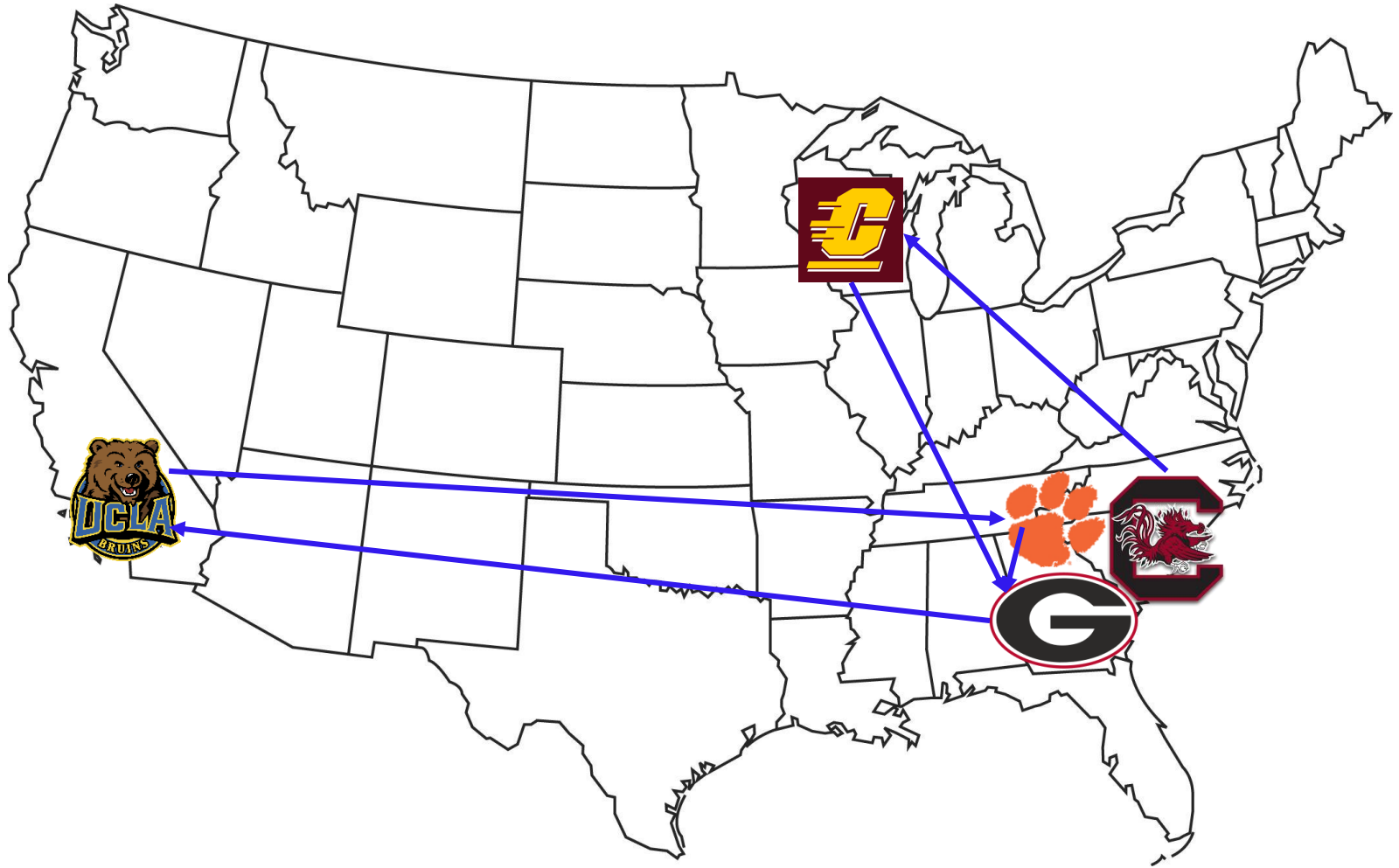


A Little About Me

- From Columbia, South Carolina
- Day job: professor at University of Georgia
- Hobbies: sports, true crime, cosmology



My Journey Summary



Questions

- Ask me questions about college, or graduate school, or my journey, or statistics, or psychometrics, or being a professor



Review of Simple Linear Regression

- Simple linear regression (SLR): goal is to predict a continuous DV with a single predictor variable

$$y_i = \beta_0 + \beta_1 x_1 + e_i$$

- Assumptions:
 - $e_i \sim N(0, \sigma^2)$
 - Homoscedasticity
 - Linearity
 - Independence
- Other considerations
 - Outliers and influential observations

Review of Simple Linear Regression

- Overall goal: explain variance in the DV
- **Interpret coefficients** for continuous predictors
- Intercept (β_0): the predicted y when predictor is zero
- Slope (β_i): the predicted change in y given a 1-unit change in x_i

Review of Simple Linear Regression

- Response = AP Statistics final exam (0 – 100 scale)
- Predictor: study time per week

$$\widehat{score}_i = 50 + 12(hours_i)$$

- Intercept interpretation: if someone studies 0 hours per week, predicted final exam score is 50
- Slope interpretation: for every 1 additional hour studied per week, predicted score increases by 12 points

Review of Simple Linear Regression

- **Model Evaluation**
- R^2 : proportion of variance explained by the predictor
- Predictors: study time per week

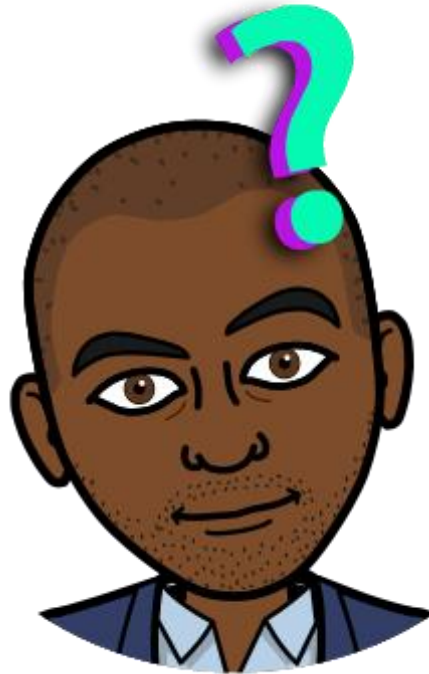
$$\hat{y}_i = 50 + 12(hours_i)$$

- Suppose $R^2 = .32$
- 32% of the variance in final exam scores is explained by hours studied
 - 68% is unexplained

Review of Simple Linear Regression

- **Model Evaluation**
- Significance of a predictor
 - A p -value $< .05$ for a predictor indicates that it is a significant predictor
 - I.e., it explains a significant proportion of variance in the response variable
 - A p -value $> .05$ for a predictor indicates that it does not predict the response variable

Questions



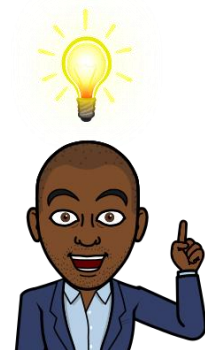
Multiple Regression

Multiple Regression

- Response = AP Statistics final exam (0 – 100 scale)
- Predictors: study time per week

$$\widehat{score}_i = 50 + 12(hours_i)$$

- $R^2 = .32$
- 32% of the variance in final exam scores is explained by hours studied
 - 68% is unexplained
- Let's add more predictors to explain more variance



Multiple Regression

- Multiple regression goal is to predict a continuous DV with **multiple predictor variables**:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e_i$$

- Assumptions:
 - $e_i \sim N(0, \sigma^2)$
 - Homoscedasticity
 - Linearity
 - Independence
- Other considerations
 - Multicollinearity
 - Outliers and influential observations

Multiple Regression

- Overall goal: explain variance in the DV **with multiple predictors**
- **Interpret coefficients** for continuous predictors
- Intercept (β_0): the predicted y when **all** predictors are zero
- Slope (β_i): the predicted change in y given a 1-unit change in x_i , **holding other variables constant**

Multiple Regression

- Response = AP Statistics final exam (0 – 100 scale)
- Predictors: study time per week, algebra EOC score

$$\widehat{score}_i = 26 + 7(hours_i) + 0.2(algebra_i)$$

- Intercept interpretation: if you study 0 hours per week **and** made a 0 on the algebra EOC, predicted score is 26
- Slope interpretation: for every 1 additional hour studied per week, predicted score increases by 7 points, **holding other variables constant**
- Slope interpretation: for every 1 additional point on the algebra EOC, predicted score increases by 0.2 points, **holding other variables constant**

Review of Simple Linear Regression

- **Model Evaluation**
- R^2 : proportion of variance explained by all the predictors
- Predictors: study time per week

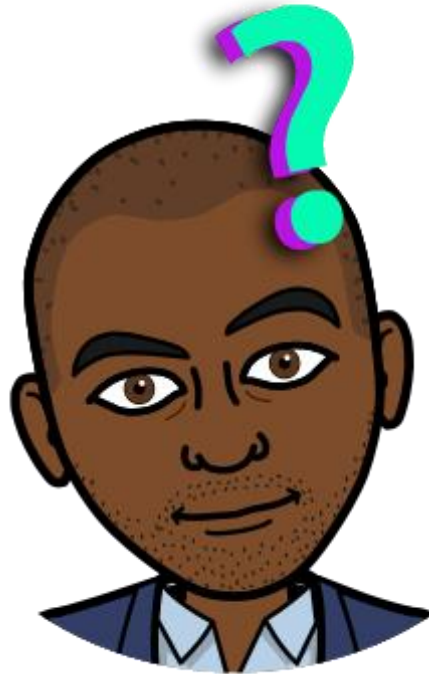
$$\widehat{score}_i = 26 + 7(hours_i) + 0.2(algebra)$$

- R^2 was .32 (68% unexplained)
- With the added predictor, R^2 increases to .48
 - Still 52% unexplained, but better than 68%

Review of Simple Linear Regression

- **Model Evaluation**
- Significance of all predictors
 - A p-value $< .05$ for a predictor indicates that it is a significant predictor
 - A p-value $> .05$ for a predictor indicates that it does not predict the response variable
- Model *parsimony*
 - Want a model that predicts well, but is efficient and doesn't have redundant or wasted space
 - Drop insignificant predictors until the model only has significant predictors

Questions



Examples in RStudio

- Example: predicting job performance
 - Open the software instructions document

Lab Assignment

- Complete the lab assignment on your own
 - Data: predicting freshman year college GPA
 - *fygpa*: first year college GPA
 - *hsgpa*: high school GPA
 - *sat*: SAT total score
 - *numap*: number of AP courses taken in high school
 - *study*: self-reported hours studied per week