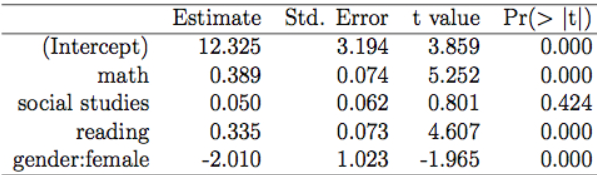
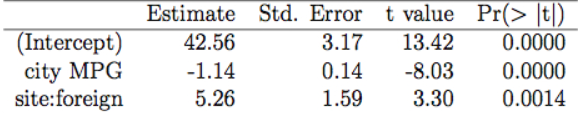
***Week 3 Quiz***

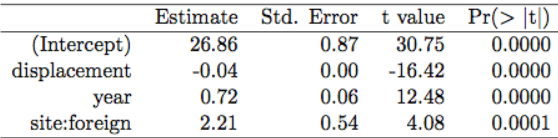
* Data were collected on 200 HS students’ scores on various tests, including science, math, reading + social studies + gender of the students. The model output for predicting science scores from the rest of the variables is shown below. Which of the following is the linear model for female students?



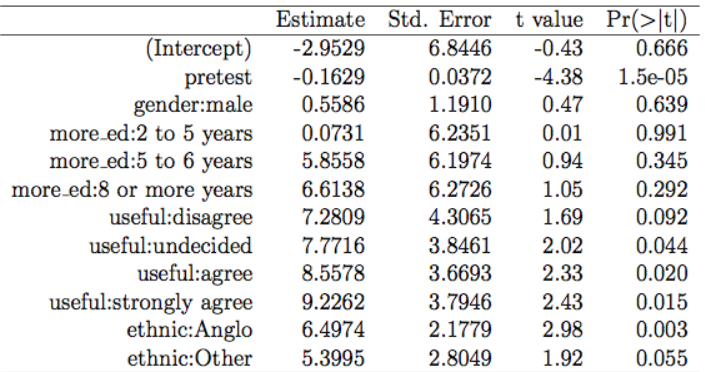
* **scienceˆ = 10.315 + 0.389 math + 0.050 social studies + 0.335 reading**
* We modeled prices of 93 cars (in $1,000s) using city MPG, manufacturing site (foreign or domestic). The regression output is provided below. Note that domestic is the reference level for manufacturing site. Data are outdated so the prices may seem low.



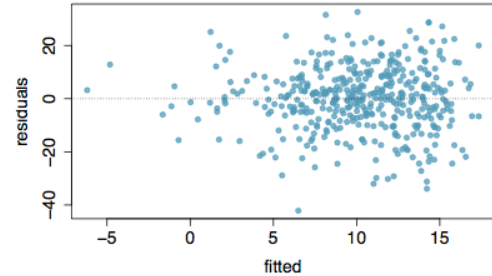
* Which of the following is the correct predicted price (in $1,000s) of a foreign car that gets 26 MPG?
* **42.56 – (26\*1.14) + 5.26** = $18,180
* Which of the following is true?
* **The 95% CI for slope of city MPG can be calculated as −1.14 ± (1.96∗0.14).**
* **Manufacturing site = a significant predictor of car price, given info on the city MPG of the car.**
* **City MPG = significant predictor of car price, given info on the manufacturing site of the car.**
* **If we add another variable to the model, say highway MPG, the p-values associated w/ city MPG and manufacturing site may change.**
* We modeled gas mileage of 398 cars built in the 70’s + early 80’s using engine displacement (in cubic in.), year of manufacture in relation to 1970 (e.g. 4 = car was built in 1974; 12 = built in 1982, etc.), + manufacturing site (domestic to USA = 0; foreign to USA = 1). The regression output is provided below. Note that domestic = reference level for manufacturing site.



* Which of the following is the best interpretation of the slope of year?
* **All else held constant, the model predicts that later model cars will get an average of 0.72 additional mpg for each year difference in the date of manufacture.**
* estimate for intercept (b0) = expected value of y when all predictors = 0, on average.
* estimate for a slope (say b1) = All else held constant, for each unit increase in x1, expect y to be higher/lower on average by b1
* Which of the following is true?
* **Given info on the manufacturing site of the car + the year of manufacture, engine displacement is a significant predictor of gas mileage.**
* **The 95% CI for slope of the displacement coefficient can be calculated as −0.04 +/- (1.96∗0.00)**
* CI = b(i) +/- t\*\_(n – k – 1) \* SE\_b(i)
* **Given engine displacement + manufacture site, manufacture year = significant predictor of mpg**
* **If we add another variable to the model, like price of the car, the p-values associated with year of manufacture, site of manufacture, and engine displacement may change.**
* Which of the following is the degrees of freedom associated with the p-value for site?
* **394** = 398 – 3 - 1
* You’re considering adding an explanatory variable to an existing multiple linear regression model. What is generally true regarding R2 + Adj. R2 as a result of adding a variable?
* **If the variable is a meaningful predictor, R2 and adjusted R2 will both increase.**
* R2 always increases w/ more predictors, Adj. R2 only does so w/ meaningful predictors whose contributions overcome penalty of being added
* Adj. R2 = 1 – (SSE / (n – k – 1))/(SST / (n – 1))
* Which of the following is true?
* **A parsimonious model = model containing highest possible predictive power w/ least amount of predictors.**
* **Adjusted R2 applies a penalty for the number of predictors included in the regression model.**
* Which of the following is true about conditions for multiple linear regression?
* **The residuals plot should show constant variability of residuals around 0.**
* **It is ideal for there to be no strong relationships between any of the explanatory variables.**
* **When residuals are plotted in a histogram, they should appear normally distributed around 0.**
* **With multiple predictors in the model, it’s necessary for each of the numerical explanatory variables to have a linear relationship with the response variable.**
* Which of the following is true?
* **R2 is always greater than or equal to adjusted R2.**
* **When dealing with collinearity, a bad strategy = add more predictors to the model, 1 at a time, until the bad effects of collinearity disappear from the analysis.**
* **In backwards model selection using adjusted R2 as the criterion, we drop variables from the model, one at a time, until adjusted R2 is maximized.**
* True / False. If the F-test assessing the overall significance of a multiple linear regression yields a significant p-value, all variables included in the model must be significant predictors.
* True
* **False** 🡪 At least 1
* The data in this question come from the Second International Mathematics Study on 8th graders from randomly sampled classrooms in the US who completed mathematics achievement tests at the beginning + end of the academic year. Students also answered questions regarding their attitudes toward mathematics. The linear model output below is for predicting the gain score in this test (posttest - pretest score) using the following explanatory variables:
* pretest: score on the exam taken at the beginning of the semester
* gender: male or female
* more\_ed: expected # of years for continued education (up to 2 years, 2 to 5 years, 5 to 6 years, 8 or more years)
* useful: Math = useful in everyday life (strongly disagree, disagree, undecided, agree, strongly agree)
* ethnic: ethnicity of student (African American, Anglo, Other)



* What does the intercept in this model represent?
* **An African American female student who scored 0 on the pretest, expects to continue their education for up to 2 years, who strongly disagrees with the statement on usefulness of math.**
* Which of the following conditions can this residuals plot for this model be used to check?



* **Constant variability of residuals** (want random scatter around 0)
* R2 will never decrease when a predictor is added to a linear model.
* **True**
* Which of the following is true?
* **In backwards model selection using p-value criterion, start w/ full model, + drop variable w/ highest p-value, 1 at a time, until all variables in model = significant given the others**
* **1 consequence of collinearity in multiple regression = biased estimates on slope coefficients**
* **R2 is always greater than or equal to adjusted R2.**
* Which of the following is the best definition of a parsimonious model?
* **The simplest model with the highest predictive power.**
* A high correlation between 2 explanatory variables such that the 2 variables contribute redundant information to the model is known as
* **collinearity**
* Suppose you have performed forward selection using Adj. R2 criterion + have chosen a model w/ 6 predictors. Based on studies of model selection, which of the following is most likely to be true?
* **All 6 predictors will not necessarily be significant in the model** (just raise Adj. R2)
* **The model you’ve arrived at is not necessarily the most parsimonious model.**
* ***Your final model has a higher adjusted R2 than any of the smaller models tried*.**