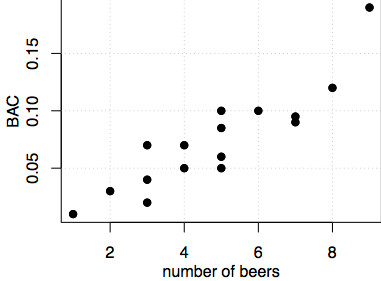
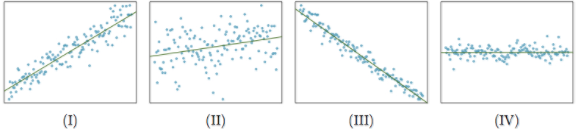
**Wk 1 Quiz**

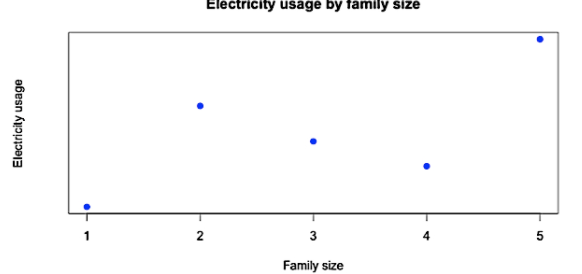
* A statistics instructor wants to use # of hours studied to predict exam scores in her class. She wants to use a linear regression model. Data from previous years shows the correlation between these 2 variables = 0.76. Which of the following is the best response for whether or not the instructor should use linear regression to predict exam scores for a student who studied 10 hours for the final?
* **Linear regression could be appropriate if the scatterplot shows a clear linear relationship.**
* Which of the following is false?
* **If the correlation coefficient is 1, then the slope must be 1 as well.**
* ~~Magnitude of correlation coefficient measures strength of linear association between 2 numerical variables.~~
* ~~Correlation coefficient and the slope always have the same sign (positive or negative).~~
* ~~Two numerical variables with a correlation of 0.01 have very weak linear association.~~
* 16 student volunteers at OSU drank a randomly assigned # of beers. 30 minutes later, a police officer measured their BAC in g of alcohol per dL of blood. The scatterplot displays the relationship between BAC + # of beers consumed. If the student who drank the highest # of beers (9) actually had a BAC of 0.15 g/deciliter, how would the strength of the association change?



* **Increase** (now closer to trend + possible regression line)
* R2 for the linear regression of 2 variables x + y = 0.60. The variables are negatively associated. Which of the following is the correct correlation coefficient? Choose the closest answer.
* **-0.77** (negative square root of .6)
* Of the 4 plots shown below, which appears to show the weakest relationship between 2 variables?



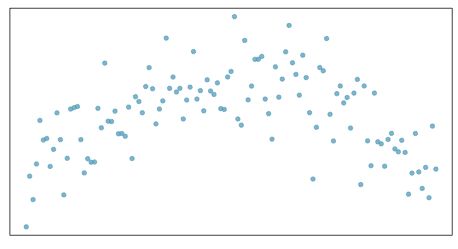
* **IV** (relationship is weakest b/c it is horizontal line = thus no direct relationship)
* Suppose the correlation coefficient R between an explanatory variable (size of a person’s home, measured in square ft.) + a response variable (amount spent on alcohol per year, measured in US $) is 0.34. If we change the units used to measure home size from square feet to square meters, which of the following will be true? Assume 1 square foot is roughly 0.09 square meters.
* **The new correlation coefficient should be the same: R=0.34** (R is unit-less)
* Which of the following is false?
* ~~Residuals of linear models should be distributed nearly normally around 0.~~
* ~~The residuals plot (residuals vs. x) should show a random scatter around 0.~~
* ~~A data point that has a negative residual is located below the regression line.~~
* **The variability of residuals should increase as x increases.**
* An young student collected data on household electricity usage for a few families, plotted the data, + observed that there did not appear to be a strong, positive linear relationship between the 2 variables as expected. The student still suspects such a relationship exists - which of the following is the best advice an experienced statistician could give to the girl in order to help her investigate whether there is a linear relationship?



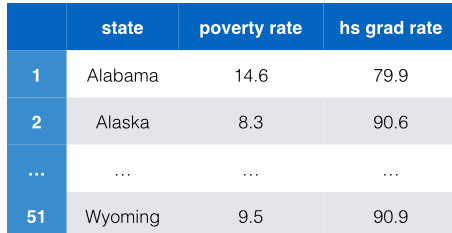
* **Collect electricity usage data for more families of sizes 1 through 5.**
* For a basketball team, 32% of variability in their PPG is explained by total salary of the opposing team. For this team, which of the following *could* be the correlation between their PPG + the salary of the opposing team?



* A colleague needs help w/ a stats problem: He brings you a plot w/ a correlation coefficient = .03, calculated himself. The plot shows 2 numerical variables which are obviously strongly related, + as a result your colleague is afraid he made a mistake calculating R (was surprised to get an answer so close to 0). Given only this info, which of the following responses is the best to give your colleague?



* **The correlation coefficient measures the strength of the *linear* relationship, therefore 2 variables that have a strong *non-linear* association might still have a low correlation coefficient.**
* A linear model is used for predicting poverty rate from HS graduation rate in the US (+ DC).
* **povertyˆ=64.68−0.62 HS grad rate**
* HS graduation rate for NC = 81.4% + poverty rate = 13.1%. What is the residual for this observation?
* A snippet of the data matrix is provided below 🡪 pay attention to the scale of the data



* **-1.1** (13.1 minus value below)

> int <- 64.68

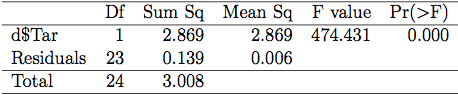
> HSgr <- 81.4

> slope <- -0.62

> (int + slope\*HSgr)

[1] 14.212

* The following ANOVA output is for the linear model predicting nicotine content (in mg) from tar content (in mg). Which of the following is R2? Choose the closest answer.



* **95%** 🡪 R2 = explained variability / total variability = SSt / SSe
* Based on a random sample of 170 married couples in Britain, a researcher finds the relationship between husbands’ + wives’ ages is described by:



* Which of the following is the best interpretation of the slope estimate?
* **For each additional year increase of husband’s age, we would expect the wife’s age to be 0.91 years higher, on average.**