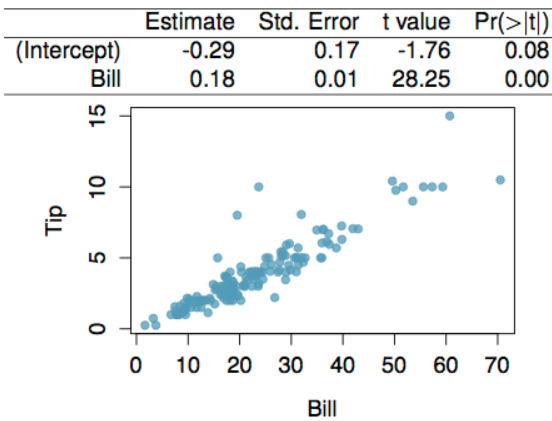
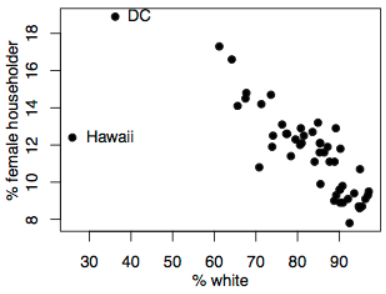
* Does meditation cure insomnia? Researchers randomly divided 400 people into 2 equal-sized groups, meditating daily for 30 minutes vs. attending a 2-hour info session on insomnia. At the study start, average difference between # of minutes slept between the 2 groups was ~0. After the study, average difference was ~32 minutes, + the meditation group had higher average # of minutes slept. To test whether an average difference of 32 minutes could be attributed to chance, we decided to conduct a randomization test. We wrote the # of minutes slept by each subject in the study on an index card, shuffled the cards together very well, + then dealt them into 2 equal-sized groups, representing those who meditated + those who attended the info session. Which of the following best describes the outcome of the randomization test?
* **Avg. difference between the 2 values on the 2 stacks of cards is expected to be ~0 minutes.**
* Simulating the experiment under the assumption of null being true + the null in this case is

*H*0 : *μmeditate* − *μinfosession* =0.

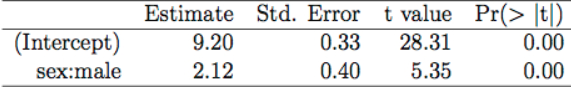
* The owner of a restaurant interested in studying tipping patterns of customers collected restaurant bills over a 2-week period that he believes provides a representative sample of his customers. Data recorded from 157 bills includes amount of the bill (in $) + the size of the tip (in $). The correlation coefficient of the relationship between these 2 variables = 0.915. Which of the following is true?



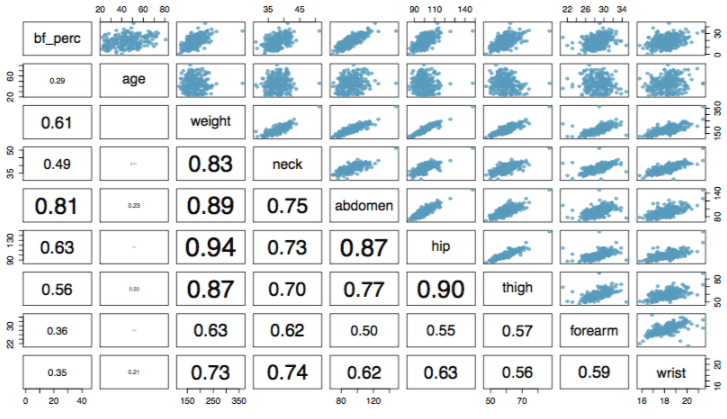
* **The intercept serves only to adjust the height of the regression line.**
* **The amount of the bills explains roughly 84% of the variability in the sizes of the tips.**
* **For each additional $ spent the tip is expected to be higher on average by $0.18 (18 cents).**
* **The regression model will yield a more reliable prediction for the size of the tip when the bill amount is $20 versus when the bill amount is $70.**
* T/F. Not all leverage points are influential.
* **True**
* If a leverage point is ON the trajectory of the regression line, it won’t affect the slope, + hence won’t be considered influential.
* **leverage point =** lies away from the center of the DP’s in the horizontal direction
* **influential point** = point that influences (changes) the slope of the regression line.
* usually a leverage point away from the trajectory of the rest of the data.
* See the relationship between % of white residents + % of households w/ a female head in all 50 US States + DC. Which of the below best describes the 2 points marked as DC + Hawaii?



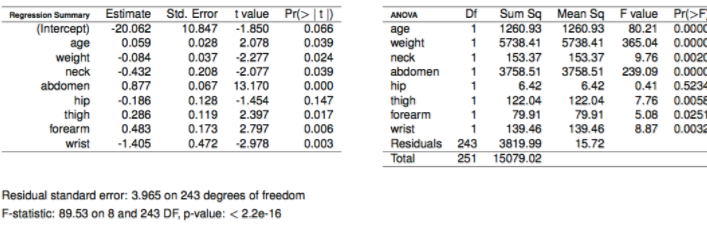
* **Hawaii has higher leverage and is more influential than DC.**
* See relationship between socioeconomic status (% of children in a neighborhood receiving reduced-fee lunches at school (**lunch**)) + bike helmet use (% of bike riders in the neighborhood wearing helmets (helmet)). Regression line: **helmet = 47.49 − 0.54\*lunch +** R2 = 72%. Which of the following is true?
* 72% of the % of children wearing bike helmets can be accurately predicted by the model.
* Increasing the % of children receiving reduced-fee lunches at school by 5% will increase % of bike riders wearing helmets in that neighborhood by 2.7%.
* The correlation coefficient is 0.85.
* See model predicting the heart weight (in g) of cats from gender. The coefficients are estimated using a dataset of 144 domestic cats. Which of the following is true?



* **Gender is a significant predictor of heart weight in cats.**
* **The expected heart weight for female cats is 9.2 grams, on average.**
* **On average, male cats are expected to have hearts 2.12 g higher than hearts of female cats.**
* **B/c Pr(>|t|) = 0 for `gender`, we can say gender IS a significant predictor of cat heart weight**
* Determine if I or II is higher, or if equal: *The uncertainty associated w/ the slope estimate when*
* I. there is a lot of scatter around the regression line
* II. there is very little scatter around the regression line
* **I is higher**
* Body fat % can be complicated to estimate, while variables such age, height, weight, + various body part measurements are easy to measure. Based on data on BF % + other various easy to obtain measurements, we develop a model to predict BF % based on the following variables:
* **Age** (yr.), **abdomen circumference** (cm), **forearm circumference** (cm), **weight** (lbs.), **hip circumference** (cm), **wrist circumference** (cm), **neck circumference** (cm), **thigh circumference** (cm)
* See the relationship between each of these variables + BF % (response variable) as well as the correlation coefficients between these variables:



* See the following model outputs associated with this analysis:



* Which of the following is supported by info provided in the model outputs above?
* **Abdomen circumference = most significant predictor of BF% since the p-value is the smallest**
* **All else held constant, people with wider hips tend to have lower body fat percentages.**
* **All else held constant, for each additional cm the forearm circumference is higher, body fat percentage is expected to be higher by 0.483 percentage points.**
* **The sample size is 252.**
* **The F-test for the significance of the model overall suggests that at least 1 of the slope coefficients is significantly different than 0**
* Do these data provide convincing evidence that age + FB% are significantly positively associated?
* **Yes, the p-value for testing for a positive correlation between age + BF% = 0.039. Since the p-value is small we reject the null = no relationship.**
* Construct a 95% CI for the slope of abdomen circumference and interpret it in context of the data.

> b0.bf.perc <- -20.062

> b1.abdomen.circum <- .877

> se <- .067

> null <- 0

> (t <- (b1.abdomen.circum) / se)

[1] 13.08955

>

> n <- 252

> dF <- n - 2

>

> pt.est <- b1.abdomen.circum

> (t.crit <- abs(qt(p = .025, df = 251)))

[1] 1.96946

> mOe <- t.crit\*se

> (lower <- pt.est - mOe)

[1] 0.7450462

> (upper <- pt.est + mOe)

[1] 1.008954

* (0.745, 1.009); All else held constant, for each additional cm in abdomen circumference, BF% is expected to be higher by 0.745 to 1.009 percentage points.
* T/F: Outliers should always be removed from the data set prior to final analysis.
* **False; only remove outliers w/ good justification that suggests removing them is appropriate.**