

Homework 1

Math 530/630 - Fall 2018

Install and load the `MASS` package in R so that you can use the `cats` dataset. (Note that if you have `dplyr` installed and wish to use `dplyr::select`, `MASS` also has a `select` function that will mask the `dplyr` version.) Using the `cats` dataset, we'll explore the difference between *covariance* and *correlation*.

1. Make a scatterplot of body weight versus heart weight in this sample of cats, and add the linear regression line. What do you see? (hint: comment on linearity of association and strength of that association) (See also: <http://guessthecorrelation.com>). Using the built-in R function `cov`, calculate the covariance of body weight and heart weight in this sample of cats. (hint: use `with(cats, cov(arg1, arg2))` or `cov(cats$arg1, cats$arg2)`)

```
install.packages("MASS")
library(MASS)
head(cats)
glimpse(cats)
select <- dplyr::select # if you want to override the MASS select function with the dplyr one
```

2. Convert the units of both body and heart weight from the metric system to the imperial system, using the code provided below, and re-make your scatterplot. What changed? What did not change? Also do the following:
 - Re-calculate the covariance between body and heart weight in pounds. What do you notice? Does it make sense?
 - Using the built-in R function `cor`, calculate two correlations between body and heart weight- one on the raw values in kg and g, respectively, and one using the same measurements after converting to pounds.
 - Compare and contrast the two correlations you calculated to the two covariances you have calculated.

```
# convert units for part c
cats_imperial <- cats %>%
  mutate(bwt_lbs = Bwt * 2.205,
         hwt_lbs = Hwt * 0.0022046)
```

3. Finally, go back to the original raw measures of body and heart weight (in kg and g, respectively).
 - Transform each variable into z-score form (you may wish to confirm for yourself that the mean = 0 and sd = 1 for each).
 - Calculate the covariance and the correlation between the two variables in z-score form.
 - Use this example to explain in words how a correlation is different from covariance (hint: do you think you want a measure of association between two variables that is sensitive to linear transformations? Why or why not? Do either of these two statistics appear to be insensitive to linear transformation? If so, look carefully at the formulas and try to explain in words why.)