Correlation and Regression*

Worksheet 4

1. Observations made by the astronomer Edwin Hubble showed that the universe is expanding. If v is the galaxy's recession from the Milky Way and d is the distance to that galaxy, Hubble's law is the linear relationship.

$$v = H_0 d.$$

H_0 is known as *Hubble's constant*. To estimate H, we use the data below, Distance in measured in millions of light years. Velocity is measured in thousands of miles per second.

Cluster	Distance	Velocity
Virgo	22	0.8
Pegasus	68	2.4
Perseus	108	3.2
Coma Berenices	137	4.7
Ursa Major #1	255	9.3
Leo	315	12.0
Corona Borealis	390	13.4
Gemini	405	14.4
Bootes	685	24.5
Ursa Major #2	700	26.0
Hydra	1100	38.0

   (a) Provide a scatterplot of the data using d as the explanatory variable
   (b) Determine corresponding the regression line, and report your estimate of Hubble's constant. (Note that the y-intercept is 0.)
   (c) Give the sum of the residuals. Note that it is not equal to zero.
   (d) Determine the equation for the regression line using v as the explanatory variable.
   (e) Find the product of the slopes of the two regression lines.
   (f) How close are they are being the same line? Explain your answer.

2. Download the data set `mammals` by calling for `library("MASS")`
   (a) Enter `mammals` and describe the data set.
   (b) Plot the data with `plot(mammals)` and describe the plot.

*©2019 Joseph C Watkins

(c) Give appropriate transformations of one or both of the variables that gives a scatterplot suitable for linear regression. Make a scatterplot of the transformed data and describe the plot.

(d) Give the coefficients in the regression line of the transformed variables.

(e) On average, how does brain size change with a doubling of body size?

(f) Use `data.frame` with `mammals` and the residuals to give examples of mammal species with unusually large and unusually small brain size. (The `order` command can be used to order the residuals from small to large keeping the rows of the data intact.)