**Name: Classwork 11.1**

**Back to Robert Martin and Westvaco**

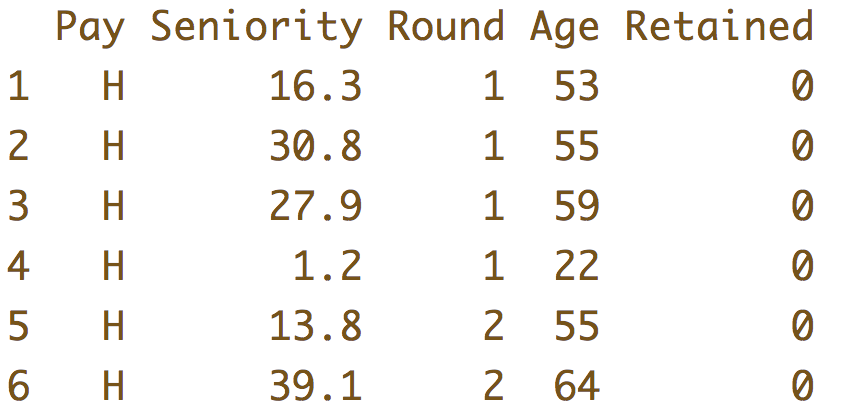
Before we looked at a tiny bit of the data (just 10 people in one department at Westvaco) because we knew so little about statistics. Now we have learned a lot. Let’s put it all together. There were 50 people in the engineering department at Westvaco that were part of the legal case brought by Martin [Source: *Martin v. Envelope Division of Westvaco Corp.,* CA, No. 92-03121-MAP, 850 Fed. Supp. 83 (1994)].

We are now going to do all three parts of statistical analysis, but now with R!

You can do this on the DataCamp Sandbox or by downloading today’s R notebook (see Classwork 11.1).

Westvaco <- read.csv("http://bit.ly/westvaco50", header=TRUE)

head(Westvaco)



Part I: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Explore the data. What visualization might help you explore whether there was age discrimination in the layoffs?
2. Check out the data but be careful not to jump to conclusions. What’s your opinion about the Westvaco data? Do you suspect discrimination? Consider both sides.
   1. What are some reasons (from the data) you have for suspecting discrimination?
   2. What are some reasons (from the data) you have for not suspecting discrimination?
   3. Is it possible to have gotten this pattern of data by firing people at random?

Part II: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Does retention (**Retained**) explain variation in **Age**? Let’s write \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, then \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, then \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ to answer this question.
2. Let’s create models of the variation in Age. Does adding the Retained variable explain more of the variation in Age? How much more?
3. What does the PRE mean? Does this mean that the data generating process (DGP) involved discrimination? Why or why not?
4. Could a DGP that did not involve discrimination have a PRE like this one?

Part III: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. So let’s simulate firing people at random.
   1. How many people were fired?
   2. What is the R code for randomly picking that many people?
   3. How would you calculate the mean age of these randomly fired people?
   4. Let’s hear from the class – what are a few of the mean ages of the people your R code randomly fired?
   5. The mean age of the 28 people that actually got fired was 49.86. Do you think that could happen by chance alone? Do you think it would be “likely”?
   6. How could we look at a bunch of the mean ages of randomly fired groups of people? Figure out the R code here.
   7. What’s the probability that we would get such an age (or older) by chance alone?
   8. Is “chance alone” more like the empty model or is it more like the complex model? Why?
   9. Remember the beginning of this class when we did something like this? How did we do it then? How is it similar to what we are doing today? How is it different?
2. Create a plan to test out whether we could have fired a group of people with an average age of 49.86 or higher *by chance alone*. Then talk to a partner about your plan. Then carry out your plan and make note of the R code.
3. If the probability of getting an average age of 49.86 or greater turns out to be small, does this favor Martin or Westvaco? Why or why not?
4. Summarize what we did today with the distribution triad.