Homework 6 – Intro. to Computational Statistics

For all problems, please show all your work. As described in the Homework Guidelines, use RMarkdown to write up your work as a .Rmd file, "knit" the result to a PDF file, and submit only that PDF file. Be sure to use R code for all your calculations, and the latex equation format to write up any math. See the Homework Guidelines for more formatting details.

You conduct an exit poll after an election, with the following results:

	18-29	30-44	45-59	60+
Democrat	86	72	73	71
Independent	52	51	55	54
Republican	61	74	70	73

Figure 1: Number of respondents per demographic category.

1.

a. Based on the exit poll results, is age independent of Party ID or not? Conduct a chi-squared test by hand, showing each step in readably-formatted latex.

b. Verify your results using R to conduct the test.

2.

a. Now test for independence using ANOVA (an F test). Your three groups are Democrats, Independents, and Republicans. The average age for a Democrat is 43.3, for an Independent it's 44.6, and for a Republican it's 45.1. The standard deviations of each are D: 9.1, I: 9.2, R: 9.2. The overall mean age is 44.2. Do the F test by hand, again showing each step.

b. Check your results in R using simulated data. Your simulated dataset should have two variables: age, and a factor indicating for each row whether it is Democrat, Independent, or Republican. There should be 302 Democrats drawn from a normal distribution with mean 43.3 and sd 9.1, and likewise for Independents and Republicans. One way to construct this dataset is to first use cbind to create the Democratic, Independent, and Republican datasets, and then use rbind to stack them together: eg, use `democrats <- cbind(rnorm(302,43.3,9.1),"democrat")` and likewise for the other two groups; then use rbind to glue the democrat, indpendent, and republican datasets into one data frame; and then be sure to name your variables and make sure that the first is numeric and the second a factor. Once this dataset has been constructed, conduct an F test using R's `aov` function on the ages and compare the results to 2a. Do your results match 2a? If not, why not?